

OLASILIK VE İSTATİSTİK

1. HAFTA

İSTATİSTİK VE VERİ ANALİZİNE GİRİŞ

1.1 GENEL BAKIŞ: İSTATİSTİKSEL ÇIKARIM, ÖRNEKLER, POPÜLASYONLAR VE OLASILIĞIN ROLÜ

1980'lerden başlayarak ve 21. yüzyıla kadar devam ederek, dünya endüstrisinde kalitenin iyileştirilmesine aşırı miktarda dikkatle odaklanılmıştır. İstatistiksel yöntemlerin kullanılmasıyla ve yönetim personeli arasında istatistiksel düşüncenin gerçekleştirilmesiyle ülkelerin endüstri anlamındaki başarısı gittikçe artmıştır.

A. Bilimsel Verilerin Kullanımı

Üretimde, gıda ürünlerinin geliştirilmesinde, bilgisayar yazılımlarında, enerji kaynaklarında, ilaçlarda ve diğer birçok alanda istatistiksel yöntemlerin kullanılması, bilgi veya bilimsel verilerin toplanmasını içermektedir. Elbette, verilerin toplanması yeni bir şey değildir. Bin yılı aşkın süredir yapılmaktadır. Veriler toplandı, özetlendi, raporlandı ve incelenmek üzere saklandı. Bununla birlikte, bilimsel bilgilerin toplanması ile çıkarımsal istatistikler arasında derin bir ayrım vardır. Son yıllarda haklı olarak dikkat çeken ikincisidir.

Çıkarımsal istatistiklerin ürünü, istatistik pratisyenleri tarafından kullanılan istatistiksel yöntemlerden oluşan geniş bir "araç kutusu" olmuştur. Bu istatistiksel yöntemler, belirsizlik ve değişkenlik karşısında bilimsel yargılarda bulunma sürecine katkıda bulunmak için tasarlanmıştır. Bir üretim sürecinden belirli bir malzemenin ürün yoğunluğu her zaman aynı olmayacaktır. Gerçekten de söz konusu süreç sürekli olmaktan ziyade toplu bir süreçse, hattan çıkan partiler arasında yalnızca malzeme yoğunluğunda bir değişiklik olmayacak (partiden partiye değişim), aynı zamanda parti içinde de varyasyon olacaktır. Sürecin kalitesini iyileştirmek için sürecin neresinde değişiklik yapılabileceği konusunda daha fazla fikir edinmek için bunun gibi bir süreçten gelen verileri analiz etmek için istatistiksel yöntemler kullanılır. Bu süreçte kalite, bu yakınlık kriterinin zamanın hangi kısmında karşılandığı ile uyumlu olarak hedef yoğunluk değerine yakınlık ile ilgili olarak iyi tanımlanabilir.

Bir mühendis, kirlilik çalışmaları sırasında havadaki kükürt monoksiti ölçmek için kullanılan belirli bir aletle ilgilenebilir. Mühendisin aletin etkinliği hakkında şüpheleri varsa, ele alınması gereken iki varyasyon kaynağı vardır. Birincisi, aynı gün ve aynı yerde bulunan

kükürt monoksit değerlerinin değişimidir. İkincisi, gözlemlenen değerler ile o sırada havada bulunan gerçek kükürt monoksit miktarı arasındaki farktır. Bu iki varyasyon kaynağından herhangi biri aşırı derecede büyükse (mühendis tarafından belirlenen bazı standartlara göre), aletin değiştirilmesi gerekebilir.

Hipertansiyonu azaltan yeni bir ilaçla ilgili bir biyomedikal çalışmada, hastaların % 85'i rahatlama yaşadı, oysa mevcut ilacın veya "eski" ilacın kronik hipertansiyonu olan hastaların %80'inde rahatlama sağladığı genel olarak kabul ediliyor. Bununla birlikte, yeni ilacın yapımı daha pahalıdır ve bazı yan etkilere neden olabilir. Yeni ilaç benimsenmeli mi? Bu, ilaç firmalarının sıklıkla karşılaştığı (çoğunlukla çok daha karmaşık olan) bir sorundur. Yine, varyasyonun dikkate alınması gerekir. “% 85” değeri, çalışma için seçilen belirli hasta sayısına dayanmaktadır. Belki de çalışma yeni hastalarla tekrarlansaydı, gözlemlenen “başarı” sayısı % 75 olurdu! Karar sürecinde dikkate alınması gereken, çalışmadan çalışmaya doğal farklılıktır. Açıktır ki, bu değişkenlik önemlidir, çünkü hastadan hastaya değişiklik, soruna özgüdür.

B. Bilimsel Verilerdeki Değişkenlik

Yukarıda tartışılan problemlerde kullanılan istatistiksel yöntemler, değişkenliği ele almayı içerir ve her durumda incelenecek değişkenlik, bilimsel verilerde karşılaşılan değişkenliktir. Proseste gözlenen ürün yoğunluğu hep aynı olsaydı ve hep hedefte olsaydı istatistiksel yöntemlere gerek kalmazdı. Sülfür monoksiti ölçen cihaz her zaman aynı değeri veriyorsa ve değer doğruysa (yani doğruysa), istatistiksel analize gerek yoktur. İlaça verilen yanıtın doğasında hastadan hastaya değişkenlik olmasaydı (yani ya her zaman rahatlama getirir ya da getirmez), ilaç firmalarındaki bilim adamları için hayat basit olurdu ve kararda hiçbir istatistikçiye ihtiyaç duyulmazdı. İstatistik araştırmacıları, yukarıda açıklananlar gibi sistemlerden gelen verilerin analizine izin veren çok sayıda analitik yöntem ürettirler. Bu, çıkarımsal istatistik dediğimiz bilimin gerçek doğasını yansıtır, yani sadece verileri raporlamanın ötesine geçerek bilimsel sistem hakkında sonuçlar (veya çıkarımlar) çıkarmamıza izin veren teknikleri kullanır. İstatistikçiler, bilimsel sistemler hakkında sonuçlar çıkarmak için temel olasılık yasalarını ve istatistiksel çıkarımı kullanırlar. Bilgi, örnekler veya gözlem koleksiyonları şeklinde toplanır. Örneklem süreci 2. Bölüm'de tanıtılmaktadır ve tartışma tüm ders boyunca devam etmektedir.

Numuneler, tüm bireylerin koleksiyonları olan popülasyonlardan veya belirli bir türdeki bireysel öğelerden toplanır. Bazen bir popülasyon bilimsel bir sistemi ifade eder. Örneğin, bir bilgisayar kartı üreticisi kusurları ortadan kaldırmak isteyebilir. Bir örneklem süreci, süreçten

rastgele örneklenen 50 bilgisayar kartı hakkında bilgi toplamayı içerebilir. Burada popölasyon, firma tarafından belirli bir süre boyunca üretilen tüm bilgisayar kartlarıdır. Bilgisayar kartı sürecinde bir iyileştirme yapılırsa ve ikinci bir kart örneği toplanırsa, süreçteki değışikliğin etkinliğine ilişkin varılan herhangi bir sonuç, "geliştirilmiş süreç" kapsamında üretilen bilgisayar kartlarının tüm popölasyonunu kapsamalıdır. Bir ilaç deneyinde hastalardan bir numune alınır ve her birine kan basıncını düşürmek için özel bir ilaç verilir. İlgi, hipertansiyondan muzdarip olanların popölasyonu hakkında sonuçlar çıkarmaya odaklanmıştır.

Çoğu zaman, bilimsel verilerin sistematik bir şekilde toplanması çok önemlidir ve planlama gündemin üst sıralarında yer alır. Zaman zaman planlama, zorunlu olarak, oldukça sınırlıdır. Genellikle popölasyondaki öğelerin veya nesnelerin yalnızca belirli özelliklerine veya özelliklerine odaklanılır. Her bir özelliğin, nüfus hakkında bilgi edinmek isteyen bilim adamı veya mühendis olan "müşteri" için belirli bir mühendisliği veya sözgelimi biyolojik önemi vardır. Örneğin, yukarıdaki resimlerden birinde, işlemin kalitesi, bir işlemin çıktısının ürün yoğunluğu ile ilgiliydi. Bir mühendisin proses koşullarının, sıcaklığın, nemin, belirli bir bileşenin miktarının vb. etkilerini incelemesi gerekebilir. Bu faktörleri, arzu edilen reçete veya deneysel tasarıma göre önerilen seviyelere sistematik olarak taşıyabilir. Bununla birlikte, belirli bir tür ağaçta odun yoğunluğunu etkileyen faktörlerin araştırılmasıyla ilgilenen bir orman bilimcisi mutlaka bir deney tasarlayamaz. Bu durum, verilerin sahada toplandığı ancak faktör düzeylerinin önceden seçilemediği gözlemsel bir çalışmayı gerektirebilir. Bu tür çalışmaların her ikisi de istatistiksel çıkarım yöntemlerine uygundur. İlkinde, çıkarımların kalitesi, deneyin uygun şekilde planlanmasına bağlı olacaktır. İkincisinde, bilim adamı toplanabilecek olanın insafına kalmıştır. Örneğin, bir agronomistin yağışın bitki verimi üzerindeki etkisini incelemekle ilgilenmesi ve verilerin bir kuraklık sırasında toplanması üzücüdür.

Yöneticiler tarafından istatistiksel düşünmenin önemi ve bilimsel personel tarafından istatistiksel çıkarımın kullanılması yaygın olarak kabul edilmektedir. Araştırma bilim adamları bilimsel verilerden çok şey kazanırlar. Veriler, bilimsel fenomenlerin anlaşılmasını sağlar. Ürün ve süreç mühendisleri, süreci iyileştirmek için çevrimdışı çabalarında çok şey öğrenirler. Ayrıca düzenli olarak üretim verilerini toplayarak (çevrimiçi izleme) değerli bilgiler edinirler. Bu, süreci istenen kalitede tutmak için gerekli değışiklikleri belirlemelerini sağlar.

Bir bilim adamının yalnızca örnekte temsil edilen bir dizi verinin bir tür özetini elde etmek istediği zamanlar vardır. Başka bir deyişle, çıkarımsal istatistik gerekli değildir. Bunun yerine, bir dizi tek rakamlı istatistik veya tanımlayıcı istatistik yararlıdır. Bu sayılar, verilerin konumu, verilerdeki değışkenlik ve örneklemedeki gözlemlerin dağılımının genel doğası

hakkında bir merkez duygusu verir. İstatistiksel çıkarıma yol açan hiçbir özel istatistiksel yöntem dahil edilmemiş olsa da, çok şey öğrenilebilir. Zaman zaman, tanımlayıcı istatistiklere grafikler eşlik eder. Modern istatistiksel yazılım paketleri ortalamaların, medyanların, standart sapmaların ve diğer tek sayı istatistiklerinin hesaplanmasına ve ayrıca örneğin doğasının bir "ayak izini" gösteren grafiklerin üretilmesine izin verir. Histogramlar, gövde-yaprak çizimleri, saçılma çizimleri, nokta çizimleri ve kutu çizimleri dahil olmak üzere tek sayılı istatistiklerin ve grafiklerin tanımları ve çizimleri takip eden bölümlerde verilecektir.

C. Olasılığın Rolü

Bu derste, ileriki haftalar temel olasılık kavramlarını ele alıyor. Bu kavramlarda kapsamlı bir temellendirme, okuyucunun istatistiksel çıkarımı daha iyi anlamasına olanak tanır. Olasılık kuramının biçimselliği olmadan öğrenci, modern istatistiksel yöntemlerle veri analizinden elde edilen gerçek yorumu takdir edemez. İstatistiksel çıkarımı incelemeden önce olasılığı incelemek oldukça doğaldır. Olasılık unsurları, sonuçlarımızdaki gücü veya "güveni" ölçmemize izin verir. Bu anlamda, olasılık kavramları, istatistiksel yöntemleri tamamlayan ve istatistiksel çıkarımın gücünü ölçmemize yardımcı olan önemli bir bileşen oluşturur. O halde olasılık disiplini, tanımlayıcı istatistikler ile çıkarımsal yöntemler arasındaki geçişi sağlar. Olasılık unsurları, sonucun bilim veya mühendislik uygulayıcılarının ihtiyaç duyduğu dile getirilmesine izin verir. Okuyucunun, istatistiksel yöntemlerin kullanımından elde edilen sonuçların yorumlanmasında genellikle "sonuç" sağlayan bir P-değeri kavramını anlamasını sağlayacak bir örnek aşağıdadır.

Örnek 1.1: Bir mühendisin, 100 parçadan numune alınan ve 10 tanesinin kusurlu olduğu bir üretim sürecinden gelen verilerle karşılaştığını varsayalım. Zaman zaman kusurlu ürünlerin olması beklenir ve tahmin edilir. Açıkçası bu 100 madde örneklemini temsil ediyor. Ancak uzun vadede şirketin süreçteki kusurları ancak %5 oranında tolere edebileceği tespit edilmiştir. Şimdi, olasılık unsurları, mühendisin, sürecin doğası ile ilgili olarak numune bilgisinin ne kadar kesin olduğunu belirlemesine izin verir. Bu durumda, popülasyon kavramsal olarak süreçteki tüm olası öğeleri temsil eder. Eğer süreç kabul edilebilirse, yani %5'ten fazlası kusurlu olan ürünler üretmiyorsa, süreçten 100 üründen oluşan rastgele bir örnekleme 10 veya daha fazla kusurlu ürün elde etme olasılığının 0,0282 olduğunu öğrendiğimizi varsayalım. . Bu küçük olasılık, sürecin gerçekten de uzun vadede %5'i aşan kusurlu ürün oranına sahip olduğunu gösteriyor. Başka bir deyişle, kabul edilebilir bir süreç koşulu altında, elde edilen numune bilgisi nadiren ortaya çıkar. Ancak gerçekleşti! Açıktır ki, proses kusurlu oranı önemli bir miktarda %5'i aşarsa, çok daha yüksek bir olasılıkla gerçekleşecektir.

Bu örnekten, olasılık unsurlarının, örnek bilgilerin bilimsel sistem hakkında kesin veya sonuçsuz bir şeye dönüştürülmesine yardımcı olduğu açıkça ortaya çıkıyor. Aslında, öğrenilenler muhtemelen mühendis veya yönetici için endişe verici bilgilerdir. Aslında Bölüm 10'da detaylandıracağımız istatistiksel yöntemler, 0,0282'lik bir P değeri üretti. Sonuç, sürecin büyük olasılıkla kabul edilemez olduğunu gösteriyor. P-değeri kavramı sonraki bölümlerde uzun uzadıya ele alınmaktadır. Aşağıdaki örnek, ikinci bir açıklama sağlar.

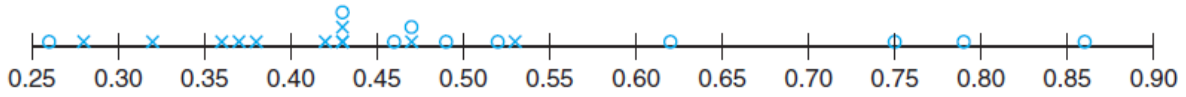
İstatistik çalışmasında, temel olarak planlı bir çalışmada veya bilimsel araştırmada ortaya çıkan tesadüfi sonuçların sunumu ve yorumlanmasıyla ilgilenilmektedir. Örneğin, bir trafik ışığının yerleştirilmesini gerekçelendirmeyi umarak, herhangi bir X kavşağında meydana gelen aylık kaza sayısını kaydedebiliriz; bir montaj hattından çıkan ürünleri "kusurlu" veya "kusursuz" olarak sınıflandırabiliriz veya bir asidin konsantrasyonu değiştiğinde kimyasal bir reaksiyonda salınan gazın hacmiyle ilgilenebiliriz. Bu nedenle, istatistikçi genellikle sayımları veya ölçümleri temsil eden sayısal verilerle veya bazı kriterlere göre sınıflandırılabilen kategorik verilerle uğraşmaktadır.

Örnek 1.2: Genellikle bilimsel çalışmanın doğası, istatistiksel çıkarımda olasılık ve tündengelimli akıl yürütmenin oynadığı rolü belirleyecektir. Sayfa 314'teki Alıştırma 9.40, ağaçların kökleri ile bir mantarın eylemi arasındaki ilişkinin geliştirilmesi üzerine Virginia Politeknik Enstitüsü ve Eyalet Üniversitesi'nde yürütülen bir çalışmayla ilgili verileri sağlar. Mineraller mantardan ağaçlara, şeker ise ağaçlardan mantara aktarılır. 10 kuzey kırmızı meşe fidanından iki numune bir seraya dikildi, biri nitrojenle işlenmiş fideler ve diğeri nitrojensiz fideler içeriyordu. Diğer tüm çevre koşulları sabit tutuldu. Tüm fideler mantar *Pisolithus tinctorus* içeriyordu. Gövde ağırlıkları gram cinsinden 140 günün sonunda kaydedilmiştir. Veriler Tablo 1.1'de verilmiştir.

Tablo1.1: Örnek 1.2 için Veri Kümesi

No Nitrogen	Nitrogen
0.32	0.26
0.53	0.43
0.28	0.47
0.37	0.49
0.47	0.52
0.43	0.75
0.36	0.79
0.42	0.86
0.38	0.62
0.43	0.46

Bu örnekte, iki ayrı popülasyondan iki örnek vardır. Deneyin amacı, nitrojen kullanımının köklerin büyümesi üzerinde bir etkisinin olup olmadığını belirlemektir. Çalışma, karşılaştırmalı bir çalışmadır (yani, iki popülasyonu belirli bir önemli özellik açısından karşılaştırmaya çalışıyoruz). Verileri Şekil 1.1'deki nokta çiziminde gösterildiği gibi çizmek öğreticidir. $\circ$ değerleri “nitrojen” verilerini temsil eder ve $\times$ değerleri “nitrojen içermeyen” verileri temsil eder.



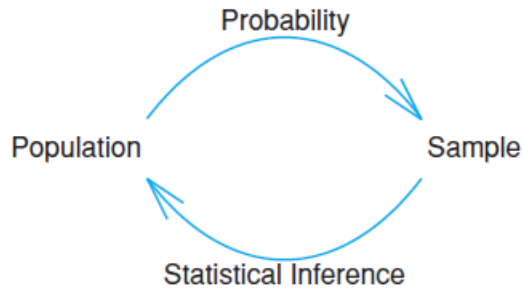
Şekil 1.1: Gövde ağırlığı verilerinin nokta grafiği.

Verilerin genel görünümünün okuyucuya ortalama olarak nitrojen kullanımının gövde ağırlığını artırdığını düşündürebileceğine dikkat edin. Dört nitrojen gözlemi, nitrojen içermeyen gözlemlerin herhangi birinden önemli ölçüde daha büyüktür. Azot içermeyen gözlemlerin çoğu, verilerin merkezinin altında görünüyor. Veri setinin görünümü nitrojenin etkili olduğunu gösteriyor gibi görünüyor. Ancak bu nasıl ölçülebilir? Görünen görsel kanıtların tümü bir anlamda nasıl özetlenebilir? Önceki örnekte olduğu gibi, olasılığın temelleri kullanılabilir. Sonuçlar bir olasılık ifadesinde veya P-değerinde özetlenebilir. Burada özet olasılığı üreten istatistiksel çıkarımı göstermeyeceğiz. Örnek 1.1'de olduğu gibi, bu yöntemler Bölüm 10'da tartışılacaktır. Nitrojen etkisinin olmadığı, diğer bir deyişle her iki numunenin de aynı popülasyondan üretildiği düşünüldüğünde, konu “buna benzer verilerin gözlemlenebilme olasılığı” etrafında dönmektedir. . Bu olasılığın küçük olduğunu varsayalım, diyelim ki 0.03. Bu kesinlikle nitrojen kullanımının kırmızı meşe fidelerinin ortalama gövde ağırlığını gerçekten etkilediğine (görünüşe göre artırdığına) dair güçlü bir kanıt olacaktır.

D. Olasılık ve İstatistiksel Çıkarım Birlikte Nasıl Çalışır?

Okuyucunun kendi başına bir bilim olan olasılık disiplini ile çıkarımsal istatistik disiplini arasındaki net farkı anlaması önemlidir. Daha önce belirttiğimiz gibi, olasılık kavramlarının kullanımı veya uygulanması, istatistiksel çıkarım sonuçlarının gerçek hayatta yorumlanmasına izin verir. Sonuç olarak istatistiksel çıkarımın olasılık kavramlarından yararlandığı söylenebilir. Yukarıdaki iki örnekten, örnek bilgilerinin analistin kullanımına sunulduğu ve istatistiksel yöntemlerin ve olasılık öğelerinin yardımıyla, popülasyonun bazı özellikleri hakkında sonuçlar çıkarıldığı (süreç kabul edilebilir görünmüyor) anlaşılabilir. Örnek 1.1 ve nitrojenin Örnek 1.2'deki ortalama gövde ağırlıklarını etkilediği görülmektedir).

Bu nedenle, istatistiksel bir problem için, çıkarımsal istatistiklerle birlikte örnek, popülasyon hakkında sonuçlar çıkarmamıza izin verir; çıkarımsal istatistikler, olasılık unsurlarını net bir şekilde kullanır. Bu akıl yürütme doğası gereği endüktiftir. Şimdi Bölüm 2'ye ve ötesine geçerken, okuyucu burada iki örneğimizde yaptığımızın aksine istatistiksel problemleri çözmeye odaklanmadığımızı fark edecektir. Hiçbir numunenin yer almadığı birçok örnek verilecektir. Bilinen popülasyonun tüm özellikleri ile açıkça tanımlanmış bir popülasyon olacaktır. Ardından, önemli sorular, popülasyondan varsayımsal olarak çıkarılabilecek verilerin doğasına odaklanacaktır. Böylece, olasılıktaki öğelerin, popülasyonun bilinen özelliklerine dayanarak, popülasyondan alınan varsayımsal verilerin özellikleri hakkında sonuçlar çıkarmamıza izin verdiği söylenebilir. Bu tür akıl yürütme doğası gereği tümdengelimlidir. Şekil 1.2, olasılık ve çıkarımsal istatistikler arasındaki temel ilişkiyi göstermektedir.



Şekil 1.2: Olasılık ve çıkarımsal istatistikler arasındaki temel ilişki.

Şimdi, genel olarak, hangisi daha önemli, olasılık alanı mı yoksa istatistik alanı mı? Her ikisi de çok önemlidir ve açıkça tamamlayıcıdır. İki disiplinin pedagojisine ilişkin tek kesinlik, eğer istatistiklerin yalnızca bir "yemek kitabı" seviyesinden daha fazla öğretilmesi gerekiyorsa, o zaman önce olasılık disiplininin öğretilmesi gerektiği gerçeğinde yatmaktadır. Bu kural, analist o örnekteki belirsizliğin temellerini öğrenene kadar bir örnekten bir popülasyon hakkında hiçbir şey öğrenilemeyeceği gerçeğinden kaynaklanmaktadır. Örneğin, Örnek 1.1'i ele alalım. Soru, süreç tarafından tanımlanan popülasyonun %5'ten fazla kusurlu olup olmadığı etrafında toplanıyor. Başka bir deyişle, varsayım, ortalama olarak 100 öğeden 5'inin kusurlu olmasıdır. Şimdi, numune 100 öğe içeriyor ve 10 tanesi kusurlu. Bu varsayımı destekliyor mu yoksa çürütüyor mu? Yüzeyde varsayımın çürütülmesi gibi görünebilir çünkü 100 kişiden 10'u "biraz fazla" görünüyor. Ancak olasılık unsurları olmadan nasıl bilebiliriz? Sadece ileriki bölümlerde materyali inceleyerek sürecin kabul edilebilir olduğu (%5 kusurlu) koşulları öğreneceğiz. 100 kişilik bir örnekleme 10 veya daha fazla kusurlu ürün elde etme olasılığı 0,0282'dir.

Olasılık unsurlarının, bilim adamı veya mühendisin bir karar oluşturmak için kanıt olarak kullanabileceği bir özet sağladığı iki örnek verdik. Veriler ve sonuç arasındaki köprü, elbette, istatistiksel çıkarım, dağılım teorisi ve gelecek bölümlerde tartışılan örnekleme dağılımlarının temellerine dayanmaktadır.

1.2 ÖRNEK (NUMUNE) ALMA İŞLEMLERİ; VERİLERİN TOPLANMASI

Bölüm 1.1'de örnekleme kavramını ve örnekleme sürecini çok kısaca tartıştık. Örnekleme basit bir kavram gibi görünse de evren ya da kitleler hakkında cevaplanması gereken soruların karmaşıklığı, örnekleme sürecinin zaman zaman çok karmaşık olmasını zorunlu kılmaktadır. Örnekleme kavramı Bölüm 8'de teknik bir şekilde tartışılırken, burada bazı sağduyulu örnekleme kavramlarını vermeye çalışacağız. Bu, değişkenlik kavramının tartışılmasına doğal bir geçiştir.

A. Basit Rastlantı (Rastgele/Rassal) Örnekleme

Uygun örneklemenin önemi, analistin sorulan soruları yanıtlayabildiği güven derecesi etrafında döner. Problemde sadece tek bir popülasyon olduğunu varsayalım. Örnek 1.2'de iki popülasyonun yer aldığını hatırlayın. Basit rasgele örnekleme, belirli bir örneklem büyüklüğündeki herhangi bir belirli örneğin, aynı büyüklükteki diğer herhangi bir örneklemle aynı seçilme şansına sahip olduğu anlamına gelir. Örnek boyutu terimi, basitçe örnekteki öğelerin sayısı anlamına gelir. Açıkçası, birçok durumda örnek seçiminde bir rasgele sayılar tablosu kullanılabilir. Basit rasgele örneklemenin erdemi, örneğin, çıkarımların yapılması gerekenden farklı (muhtemelen daha sınırlı) bir popülasyonu yansıtmaması sorununun ortadan kaldırılmasına yardımcı olmasıdır. Örneğin, Amerika Birleşik Devletleri'ndeki belirli bir eyaletteki siyasi tercihlerle ilgili belirli soruları yanıtlamak için bir örneklem seçilecektir. Örnek, diyelim ki 1000 ailenin seçimini içeriyor ve bir anket yapılacaktır. Şimdi, rastgele örneklemenin kullanılmadığını varsayalım. Aksine, seçilen 1000 ailenin tamamı veya neredeyse tamamı kentsel bir ortamda yaşıyor. Kırsal alanlardaki siyasi tercihlerin kentsel alanlardakilerden farklı olduğuna inanılmaktadır. Başka bir deyişle, çekilen örneklem aslında popülasyonu sınırlandırmıştır ve bu nedenle çıkarımların “sınırlı popülasyon” ile sınırlandırılması gerekir ve bu durumda sınırlama istenmeyebilir. Gerçekten de, bir bütün olarak durum hakkında çıkarımların yapılması gerekiyorsa, burada açıklanan 1000 büyüklüğündeki örneğe genellikle önyargılı bir örnek denir.

Daha önce ima ettiğimiz gibi, basit rasgele örnekleme her zaman uygun değildir. Hangi alternatif yaklaşımın kullanılacağı problemin karmaşıklığına bağlıdır. Çoğu zaman, örneğin,

örnekleme birimleri homojen değildir ve doğal olarak kendilerini homojen, örtüşmeyen gruplara ayırırlar. Bu gruplara tabaka denir ve tabakalı rasgele örnekleme adı verilen bir prosedür, her bir tabaka içinde bir numunenin rasgele seçimini içerir. Amaç, her bir katmanın ne fazla ne de az temsil edildiğinden emin olmaktır. Örneğin bir ilde yapılacak tahvil referandumuna ilişkin ön görüş almak için örnek bir anket yapıldığını varsayalım. Şehir, doğal tabakaları temsil eden çeşitli etnik gruplara bölünmüştür. Herhangi bir grubu göz ardı etmemek veya aşırı temsil etmemek için, her gruptan ayrı rastgele aile örnekleri seçilebilir.

B. Deneysel Tasarım

Rastgelelik veya rastgele atama kavramı, Bölüm 1.1'de çok kısaca tanıtılan ve mühendislik veya deneysel bilimin hemen hemen her alanında önemli bir temel olan deneysel tasarım alanında büyük bir rol oynar. Bu, 13 ila 15. Bölümlerde uzun uzadıya tartışılacaktır. Ancak, burada rastgele örnekleme bağlamında kısa bir sunum yapmak öğretici olacaktır. Bir dizi sözde tedaviler veya tedavi kombinasyonları, bir anlamda incelenecek veya karşılaştırılacak popülasyonlar haline gelir. Bir örnek, Örnek 1.2'deki nitrojene karşı nitrojen içermeyen işlemlerdir. Başka bir basit örnek, "plasebo" ve "aktif ilaç" olabilir veya bir korozyon yorgunluğu çalışmasında, kaplanmış veya kaplanmamış örneklerin yanı sıra örneklerin maruz kaldığı düşük veya yüksek nem koşullarını içeren tedavi kombinasyonlarımız olabilir. Aslında, dört tedavi veya faktör kombinasyonu (yani, 4 popülasyon) vardır ve birçok bilimsel soru sorulabilir ve istatistiksel ve çıkarımsal yöntemlerle yanıtlanabilir. Önce Örnek 1.2'deki durumu ele alalım. Denemede 20 adet hastalıklı fidan bulunmaktadır. Verilerin kendilerinden fidelerin birbirinden farklı olduğunu görmek kolaydır. Azot grubu (veya azotsuz grup) içinde gövde ağırlıklarında önemli değişkenlik vardır. Bu değişkenlik, genellikle deneysel birim olarak adlandırılan şeyden kaynaklanmaktadır. Bu, çıkarımsal istatistikte çok önemli bir kavramdır ve aslında açıklaması bu bölümde bitmeyecek. Değişkenliğin doğası çok önemlidir. Deneysel birimlerdeki aşırı homojen olmama durumundan kaynaklanan çok büyükse, değişkenlik iki popülasyon arasındaki saptanabilir herhangi bir farkı "yıklar". Bu durumda gerçekleşmediğini hatırlayın.

Şekil 1.1'deki nokta grafiği ve P-değeri, bu iki koşul arasında net bir ayrım olduğunu göstermiştir. Bu deneysel birimler, veri toplama sürecinin kendisinde nasıl bir rol oynuyor? Sağduyulu ve aslında oldukça standart yaklaşım, 20 fideyi veya deney birimini iki muamele veya koşula rastgele atamaktır. İlaç çalışmasında, bir anlamda açıkça farklı olacak olan toplam 200 mevcut hastayı kullanmaya karar verebiliriz. Onlar deneysel birimlerdir. Bununla birlikte, hepsi, ilacın potansiyel bir tedavi olduğu aynı kronik duruma sahip olabilir. Daha sonra

tamamen rastgele olarak adlandırılan bir tasarımda, 100 hasta rastgele plaseboya ve 100 hasta aktif ilaca atanır. Yine, veri sonuçlarındaki değişkenliği (yani, ölçülen sonuçtaki değişkenliği) üreten, bir grup veya tedavi içindeki bu deneysel birimlerdir; örneğin kan basıncı veya ilaç etkinlik değeri önemli olan her neyse. Korozyon yorulma çalışmasında deneysel birimler korozyona konu olan numunelerdir.

C. Neden Deneysel Birimler Rastgele Atanmaktadır?

Deney birimlerini işlemlere veya işlem kombinasyonlarına rastgele atamamanın olası olumsuz etkisi nedir? Bu, ilaç çalışmasında en açık şekilde görülmektedir. Sonuçlarda değişkenlik yaratan hastaların özellikleri arasında yaş, cinsiyet ve kilo bulunmaktadır. Sadece tesadüfen, plasebo grubunun, tedavi grubundakilerden ağırlıklı olarak daha ağır olan bir insan örneğini içerdiğini varsayalım. Belki de daha kilolu bireylerin daha yüksek kan basıncına sahip olma eğilimi vardır. Bu, sonucu açıkça saptırır ve aslında, istatistiksel çıkarım uygulaması yoluyla elde edilen herhangi bir sonucun ilaçla çok az ilgisi olabilir ve daha çok iki hasta örneği arasındaki ağırlık farklılıklarıyla ilgisi olabilir.

Değişkenlik terimine verilen önemin altını çizmeliyiz. Deney birimleri arasındaki aşırı değişkenlik, bilimsel bulguları "kamufle eder". Gelecek bölümlerde, değişkenlik ölçülerini karakterize etmeye ve ölçmeye çalışacağız. Takip eden bölümlerde, örneklerde hesaplanabilecek belirli miktarları tanıtıyor ve tartışıyoruz; miktarlar, verilerin konum merkezine ve verilerdeki değişkenliğe göre numunenin doğası hakkında bir fikir verir. Bu tek sayı ölçümlerinin birkaçının tartışılması, hangi istatistiksel bilgilerin ileriki bölümlerde kullanılacak istatistiksel yöntemlerin önemli bileşenleri olacağına ilişkin bir önizleme sağlamaya hizmet eder. Veri setinin doğasını karakterize etmeye yardımcı olan bu ölçümler, tanımlayıcı istatistikler kategorisine girer. Bu materyal, veri setinin karakterizasyonunda daha da ileri giden resimsel ve grafik yöntemlerin kısa bir sunumunun başlangıcıdır. Okuyucu, burada gösterilen istatistiksel yöntemlerin metin boyunca kullanılacağını anlamalıdır. Okuyucuya deneysel tasarım çalışmalarının neleri kapsadığına dair daha net bir resim sunmak için Örnek 1.3'ü sunuyoruz.

Örnek 1.3: Bir alüminyum metalinin korozyon geciktirici bir madde ile kaplanması korozyon miktarını azaltıp azaltmadığını belirlemek için bir korozyon çalışması yapılmıştır. Kaplama, bu tür malzemelerde yorulma hasarını en aza indirmek için reklamı yapılan bir koruyucudur. Ayrıca nemin korozyon miktarı üzerindeki etkisi de ilgi çekicidir. Bir korozyon ölçümü, arızaya kadar binlerce döngü olarak ifade edilebilir. Kaplamasız ve kimyasal korozyon

kaplamalı olmak üzere iki kaplama seviyesi kullanılmıştır. Ayrıca, iki bağıl nem seviyesi %20 bağıl nem ve %80 bağıl nemdir.

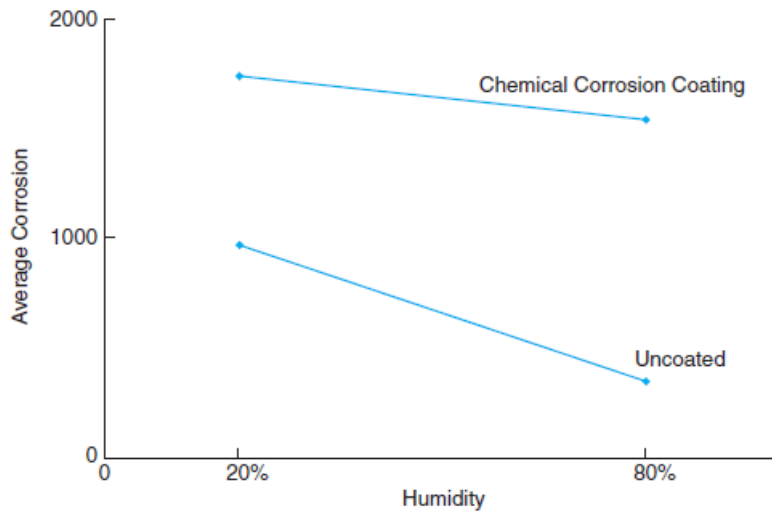
İster sayısal ister kategorik olsun, herhangi bir bilgi kaydına gözlem olarak atıfta bulunacağız. Böylece, geçen yıl boyunca Ocak'tan Nisan'a kadar her ay X kavşağında meydana gelen kazaların sayısını temsil eden 2, 0, 1 ve 2 sayıları bir dizi gözlem oluşturmaktadır. Benzer şekilde, beş öge incelendiğinde kusurlu veya hatasız bulunan ögeleri temsil eden N, D, N, N ve D kategorik verileri gözlem olarak kaydedilir.

Deney, aşağıdaki tabloda listelenen dört tedavi kombinasyonunu içerir. Kullanılan sekiz deney birimi vardır ve bunlar hazırlanmış alüminyum numunelerdir; ikisi, dört tedavi kombinasyonunun her birine rastgele atanır. Veriler Tablo 1.2'de sunulmuştur.

Tablo 1.2: Örnek 1.3 için Veriler

Coating	Humidity	Average Corrosion in Thousands of Cycles to Failure
Uncoated	20%	975
	80%	350
Chemical Corrosion	20%	1750
	80%	1550

Korozyon verileri iki numunenin ortalamasıdır. Ortalamaların grafiği Şekil 1.3'te gösterilmektedir. Nispeten büyük bir arıza döngüsü değeri, az miktarda korozyonu temsil eder. Beklenebileceği gibi, nemdeki bir artışın korozyonu daha da kötüleştirdiği görülmektedir. Kimyasal korozyon kaplama prosedürünün kullanılmasının korozyonu azalttığı görülmektedir.



Şekil 1.3: Örnek 1.3 için korozyon sonuçları.

Bu deneysel tasarım çiziminde, mühendis dört işlem kombinasyonunu sistematik olarak seçmiştir. Bu durumu okuyucunun bu noktaya kadar maruz kaldığı kavramlarla ilişkilendirmek için, dört işlem kombinasyonunu temsil eden koşulların dört ayrı popülasyon olduğu ve her popülasyon için gözlemlenen iki korozyon değerinin önemli bilgiler olduğu varsayılmalıdır. . Popülasyondaki belirli özellikleri yakalama ve özetlemede ortalamanın önemi Bölüm 1.3'te vurgulanacaktır. Şekilden nemin rolü ve numuneleri kaplamanın etkisi hakkında sonuçlar çıkarabilirsek de, ortalama etrafındaki değişkenliği hesaba katmadan sonuçları analitik bir bakış açısından gerçekten değerlendiremeyiz. Yine, daha önce belirttiğimiz gibi, her işlem kombinasyonu için iki korozyon değeri birbirine yakınsa, Şekil 1.3'teki resim doğru bir tasvir olabilir. Ancak şekildeki her aşınma değeri, geniş ölçüde dağılmış iki değerin ortalamasıysa, o zaman bu değişkenlik, gerçekten de, yalnızca ortalamalar gözlemlendiğinde ortaya çıkan herhangi bir bilgiyi gerçekten "yıkabilir". Yukarıdaki örnek bu kavramları göstermektedir:

- (1) tedavi kombinasyonlarının (kaplama, nem) deneysel birimlere (numuneler) rastgele atanması
- (2) numune bilgilerinin özetlenmesinde numune ortalamalarının (ortalama korozyon değerleri) kullanılması
- (3) herhangi bir numunenin veya numune setlerinin analizinde değişkenlik ölçülerinin dikkate alınması ihtiyacı

Bu örnek, Bölüm 1.3 ve 1.4'te takip edenlere, yani bir veri kümesindeki konum merkezi ölçümlerini gösteren tanımlayıcı istatistiklere ve değişkenliği ölçenlere ihtiyaç olduğunu göstermektedir.

1.3 KONUM ÖLÇÜLERİ: ÖRNEKLEM ORTALAMA VE MEDYAN

Konum ölçümleri, analiste veri merkezinin veya başka bir konumun nerede bulunduğuna dair bazı nicel değerler sağlamak için tasarlanmıştır. Örnek 1.2'de, nitrojen numunesinin merkezi, nitrojen içermeyen numunenin merkezini açıkça aşıyor gibi görünmektedir. Bariz ve çok yararlı bir ölçü, örnek ortalamadır. Ortalama, basitçe sayısal bir ortalamadır.

Tanım 1.1: Bir örneklemdaki gözlemlerin, $x_1, x_2, \dots, x_n$ olduğunu varsayalım. $\bar{x}$ ile gösterilen örneklem ortalaması,

$$\bar{x} = \sum_{i=1}^n \frac{x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Gelecekteki bölümlerde ayrıntılı olarak tartışılan başka merkezi eğilim ölçüleri de vardır. Önemli bir ölçü, örnek medyandır. Örneklem medyanının amacı, örneğin merkezi eğilimini, aşırı değerlerden veya aykırı değerlerden etkilenmeyecek şekilde yansıtmaktır.

Tanım 1.2: Artan büyüklük sırasına göre düzenlenmiş, bir örneklemdaki gözlemler $x_1, x_2, \dots, x_n$ için örnek medyan

örneklemdaki gözlemlerin, olduğunu varsayalım. $\bar{x}$ ile gösterilen örneklem ortalaması,

$$\bar{x} = \begin{cases} x_{(n+1)/2} \\ \frac{1}{2}(x_{n/2} + x_{(n/2)+1}) \end{cases}$$

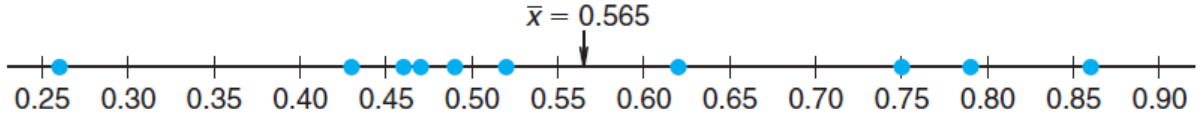
Örnek olarak, veri setinin şu olduğunu varsayalım: 1.7, 2.2, 3.9, 3.11 ve 14.7. Örnek ortalama ve medyan, sırasıyla,

$$\bar{x} = 5.12 \text{ ve } \tilde{x} = 3.11$$

Açıkçası, ortalama, aşırı gözlem olan 14.7'nin varlığından önemli ölçüde etkilenirken, medyan veri setinin gerçek "merkezine" vurgu yapar. Örnek 1.2'deki iki örneklemlili veri seti durumunda, bireysel numuneler için iki merkezi eğilim ölçüsü şu şekildedir:

$$\begin{aligned} \bar{x} \text{ (no nitrogen)} &= 0.399 \text{ gram,} \\ \tilde{x} \text{ (no nitrogen)} &= \frac{0.38 + 0.42}{2} = 0.400 \text{ gram,} \\ \bar{x} \text{ (nitrogen)} &= 0.565 \text{ gram,} \\ \tilde{x} \text{ (nitrogen)} &= \frac{0.49 + 0.52}{2} = 0.505 \text{ gram.} \end{aligned}$$

Açıkça ortalama ve medyan arasında kavram olarak bir fark vardır. Numune ortalamasının, numunedeki verilerin ağırlık merkezi olması, mühendislik geçmişi olan okuyucunun ilgisini çekebilir. Bir anlamda, bireysel verilerin konumları olan bir "ağırlıklar" sistemini dengelemek için bir dayanak noktasının yerleştirilebileceği noktadır. Bu, nitrojenli numune ile ilgili olarak Şekil 1.4'te gösterilmiştir.



Şekil 1.4: Azotlu sap ağırlığının ağırlık merkezi olarak numune ortalaması.

Gelecek bölümlerde, $\bar{x}$ 'in hesaplanmasının temeli, popülasyon ortalamasının bir tahminidir. Daha önce belirttiğimiz gibi, istatistiksel çıkarımın amacı, popülasyon özellikleri veya parametreleri hakkında sonuçlar çıkarmaktır ve tahmin, istatistiksel çıkarımın çok önemli bir özelliğidir.

Medyan ve ortalama birbirinden oldukça farklı olabilir. Bununla birlikte, gövde ağırlığı verileri durumunda, nitrojen içermeyen numune ortalama değerinin medyan değere oldukça benzer olduğuna dikkat edin.

A. Konumların Diğer Ölçüleri

Numunedeki verilerin konum merkezini ölçmenin birkaç başka yöntemi vardır. Bu noktada onlarla ilgilenmeyeceğiz. Çoğunlukla, örnek ortalamanın alternatifleri, ortalama ve medyan arasındaki uzlaşmaları temsil eden değerler üretmek için tasarlanmıştır. Bu diğer önlemleri nadiren kullanırız. Bununla birlikte, bir tahmin edici sınıfını, yani kırılmış ortalamalar sınıfını tartışmak öğretici olacaktır. Kırılmış bir ortalama, hem en büyük hem de en küçük değer kümesinin belirli bir yüzdesinin "kırılmasıyla" hesaplanır. Örneğin, %10 kırılmış ortalama, en büyük %10'luk ve en küçük %10'luk kısım çıkarılarak ve kalan değerlerin ortalaması hesaplanarak bulunur. Örneğin, gövde ağırlığı verileri söz konusu olduğunda, her örnek için örnek boyutu 10 olduğundan, en büyük ve en küçük eleriz. Yani nitrojensiz grup için %10 kırılmış ortalama şu şekilde verilir:

$$\bar{x}_{tr(10)} = \frac{0.32 + 0.37 + 0.47 + 0.43 + 0.36 + 0.42 + 0.38 + 0.43}{8} = 0.39750,$$

ve sahip olduğumuz nitrojenli grup için %10 kırılmış ortalama için

$$\bar{x}_{tr(10)} = \frac{0.43 + 0.47 + 0.49 + 0.52 + 0.75 + 0.79 + 0.62 + 0.46}{8} = 0.56625.$$

Bu durumda, beklendiği gibi, kırılmış ortalamaların bireysel numuneler için hem ortalamaya hem de ortancaya yakın olduğuna dikkat edin. Kırılmış ortalama, elbette, aykırı değerlere örnek ortalamasından daha duyarsızdır, ancak medyan kadar duyarsız değildir. Öte yandan,

kırpılmış ortalama yaklaşımı, örnek ortancadan daha fazla bilgi kullanır. Örnek ortancanın aslında, ortadaki bir veya iki gözlem dışında tüm örnek verilerin çıkarıldığı kırpılmış ortalamanın özel bir durumu olduğuna dikkat edin.

1.4 DEĞİŞKENLİK ÖLÇÜLERİ

Örnek değişkenliği, veri analizinde önemli bir rol oynar. Süreç ve ürün değişkenliği, mühendislik ve bilimsel sistemlerde hayatın bir gerçeğidir: Süreç değişkenliğinin kontrolü veya azaltılması genellikle büyük bir zorluk kaynağıdır. Giderek daha fazla süreç mühendisi ve yöneticisi, ürün kalitesinin ve bunun sonucunda üretilen ürünlerden elde edilen kârın süreç değişkenliğinin bir fonksiyonu olduğunu öğreniyor. Sonuç olarak, 9'dan 15'e kadar olan Bölümlerin çoğu, örneklem değişkenliğinin önemli bir rol oynadığı veri analizi ve modelleme prosedürleriyle ilgilidir. Küçük veri analizi problemlerinde bile, belirli bir istatistiksel yöntemin başarısı, örneklemdeki gözlemler arasındaki değişkenliğin büyüklüğüne bağlı olabilir. Bir örneklemdeki konum ölçümleri, bir veri setinin doğasının uygun bir özetini sağlamaz. Örneğin, Örnek 1.2'de, numune değişkenliğini hesaba katmadan nitrojen kullanımının büyümeyi arttırdığı sonucuna varamayız.

Bu tür veri setinin analizinin ayrıntıları Bölüm 9'a ertelenmiş olsa da, Şekil 1.1'den nitrojen içermeyen gözlemler arasındaki değişkenliğin ve nitrojen gözlemleri arasındaki değişkenliğin kesinlikle bazı sonuçları olduğu açık olmalıdır. Aslında, nitrojen numunesindeki değişkenliğin nitrojen içermeyen numuneninkinden daha büyük olduğu görülmektedir. Belki de nitrojenin dahil edilmesiyle ilgili olarak, sadece gövde yüksekliğini artırmakla kalmayan (azot içermeyen numune için $\bar{x}$ 0,399 grama kıyasla $\bar{x}$ 0,565 gram) ama aynı zamanda gövde yüksekliğindeki değişkenliği de artıran bir şeyler vardır (örn. gövde yüksekliği daha tutarsız).

Başka bir örnek olarak, aşağıdaki iki veri setini karşılaştırın. Her biri iki örnek içerir ve ortalamalardaki fark iki örnek için kabaca aynıdır, ancak B veri seti örneklerin alındığı iki popülasyon arasında çok daha keskin bir kontrast sağlıyor gibi görünmektedir. Böyle bir deneyin amacı iki popülasyon arasındaki farkları tespit etmekse, B veri seti durumunda görev başarılmıştır. Ancak, veri seti A'da iki örnek içindeki büyük değişkenlik zorluk yaratır. Aslında, iki popülasyon arasında bir ayrım olup olmadığı net değildir.

Data set A:	X	X	X	X	X	X	0	X	X	0	0	X	X	X	0	0	0	0	0	0	0
							↑									↑					
							$\bar{x}_x$									$\bar{x}_0$					
Data set B:	X	X	X	X	X	X	X	X	X	X	0	0	0	0	0	0	0	0	0	0	0
							↑									↑					
							$\bar{x}_x$									$\bar{x}_0$					

A. Konumların Diğer Ölçüleri

Pek çok merkezi eğilim veya konum ölçüsü olduğu gibi, birçok yayılma veya değişkenlik ölçüsü de vardır. Belki de en basiti, $X_{\max} - X_{\min}$ örnek aralığıdır. Aralık çok yararlı olabilir ve istatistiksel kalite kontrolüyle ilgili Bölüm 17'de uzun uzadıya tartışılmıştır. En sık kullanılan örnek yayılım ölçüsü, örnek standart sapmasıdır. Yine $x_1, x_2, \dots, x_n$ örnek değerleri gösterir.

Tanım 1.3: s^2 ile gösterilen örneklem varyansı şu şekilde verilir:

$$s^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n - 1}$$

s ile gösterilen örneklem standart sapması, s^2 'nin pozitif kareköküdür, yani,

$$s = \sqrt{s^2}$$

Örnek standart sapmanın aslında bir değişkenlik ölçüsü olduğu okuyucu için açık olmalıdır. Bir veri setindeki büyük değişkenlik, nispeten büyük $(x - \bar{x})^2$ değerleri ve dolayısıyla büyük bir örneklem varyansı üretir. $n - 1$ miktarı genellikle varyans tahminiyle ilişkili serbestlik dereceleri olarak adlandırılır. Bu basit örnekte, serbestlik dereceleri, hesaplama değişkenliği için mevcut olan bağımsız bilgi parçalarının sayısını göstermektedir. Örneğin, veri setinin (5, 17, 6, 4) örneklem varyansını ve standart sapmasını hesaplamak istediğimizi varsayalım. Örnek ortalama $\bar{x} = 8$ 'dir. Varyansın hesaplanması şunları içerir:

$$(5 - 8)^2 + (17 - 8)^2 + (6 - 8)^2 + (4 - 8)^2 = (-3)^2 + 9^2 + (-2)^2 + (-4)^2.$$

Parantez içindeki miktarların toplamı sıfırdır. Genel olarak, $\sum_{i=1}^n (x_i - \bar{x}) = 0$ 'dır. O zaman bir örneklem varyansının hesaplanması, ortalama $\bar{x}$ 'ten n bağımsız kareli sapma içermez. Aslında, $x - \bar{x}$ 'in son değeri ilk $n - 1$ tanesi tarafından belirlendiği için, bunların s^2 'yi üreten $n - 1$ "bilgi parçası" olduğunu söylüyoruz. Bu nedenle, bir örneklem varyansını hesaplamak için n serbestlik derecesi yerine $n - 1$ serbestlik derecesi vardır.

Örnek 1.4: Bir mühendis bir pH metrede "eğimi" test etmekle ilgileniyor. Nötr bir maddenin pH'ı (pH = 7.0) ölçülerek ölçüm cihazında veriler toplanır. 10 adet örneklem alınır ve sonuçlar şu şekilde verilir:

7.07 7.00 7.10 6.97 7.00 7.03 7.01 7.01 6.98 7.08.

Örneklem ortalaması $\bar{x}$ şu şekilde verilir:

$$\bar{x} = \frac{7.07 + 7.00 + 7.10 + \dots + 7.08}{10} = 7.0250.$$

Örneklem varyansı s^2 şu şekilde verilir:

$$s^2 = \frac{1}{9}[(7.07 - 7.025)^2 + (7.00 - 7.025)^2 + (7.10 - 7.025)^2 + \dots + (7.08 - 7.025)^2] = 0.001939.$$

Sonuç olarak, örneklem standart sapması şu şekilde verilir:

$$s = \sqrt{0.001939} = 0.044.$$

Dolayısıyla, $n - 1 = 9$ serbestlik derecesi ile örneklem standart sapması 0,0440'tır.

İstatistikçiler, bir veri kümesi oluşturan herhangi bir işlemi tanımlamak için deney kelimesini kullanırlar. İstatistiksel bir deneyin basit bir örneği, yazı tura atmaktır. Bu deneyde, yalnızca iki olası sonuç vardır, tura veya yazı. Başka bir deney, bir füzenin fırlatılması ve belirli zamanlarda hızının gözlemlenmesi olabilir. Seçmenlerin yeni bir satış vergisiyle ilgili görüşleri de bir deney gözlemi olarak değerlendirilebilir. Deneyi birkaç kez tekrarlayarak elde edilen gözlemlerle özellikle ilgilenilmektedir. Çoğu durumda, sonuçlar şansa bağlıdır ve bu nedenle kesin olarak tahmin edilemez. Bir kimyager aynı koşullar altında birkaç kez bir analiz yaparsa, deneysel prosedürde bir şans unsuru olduğunu gösteren farklı ölçümler elde edecektir. Bir yazı tura tekrar tekrar atıldığında bile, belirli bir atışın tura geleceğinden emin olamayız. Ancak, her atış için tüm olasılıklar bilinmektedir.

B. Standart Sapma ve Varyans Birimleri

Tanım 1.3'ten, varyansın ortalama $\bar{x}$ 'ten ortalama karesel sapmanın bir ölçüsü olduğu açık olmalıdır. Tanım, n yerine $n - 1$ serbestlik derecesine göre bir bölmeyi kullansa da, ortalama karesel sapma terimini kullanıyoruz. Tabii ki, n büyükse, paydadaki fark önemsizdir. Sonuç olarak, örneklem varyansı, gözlenen verilerdeki birimlerin karesi olan birimlere sahipken, örneklem standart sapması doğrusal birimlerde bulunur. Örnek olarak, Örnek 1.2'deki verileri ele alalım. Gövde ağırlıkları gram cinsinden ölçülür. Sonuç olarak, numune standart sapmaları gram cinsindendir ve varyanslar gram² cinsinden ölçülür. Aslında, bireysel standart sapmalar, nitrojensiz durum için 0,0728 gram ve nitrojen grubu için 0,1867 gramdır. Standart

sapmanın nitrojen örneğinde önemli ölçüde daha büyük değişkenliği gösterdiğine dikkat edin. Bu durum Şekil 1.1'de gösterilmiştir.

C. Hangi Değişkenlik Ölçüsü Daha Önemlidir?

Daha önce belirttiğimiz gibi, numune aralığının istatistiksel kalite kontrol alanında uygulamaları vardır. Okuyucu, hem numune varyansının hem de numune standart sapmasının kullanımının gereksiz olduğunu düşünebilir. Her iki ölçüm de değişkenliği ölçmede aynı kavramı yansıtır, ancak numune standart sapması değişkenliği lineer birimlerle ölçerken, numune varyansı kare birimleriyle ölçülür. Her ikisi de istatistiksel yöntemlerin kullanımında büyük rol oynar. İstatistiksel çıkarım bağlamında gerçekleştirilenlerin çoğu, popülasyonların özellikleri hakkında sonuçlar çıkarmayı içerir. Bu özellikler arasında popülasyon parametreleri adı verilen sabitler vardır. İki önemli parametre popülasyon ortalaması ve popülasyon varyansıdır. Örnek varyansı, popülasyon varyansı hakkında çıkarımlar yapmak için kullanılan istatistiksel yöntemlerde açık bir rol oynar. Kitle ortalaması hakkında yapılan çıkarımlarda, örneklem ortalamasıyla birlikte örneklem standart sapması da önemli bir yere sahiptir. Genel olarak çıkarımsal teoride varyans daha çok dikkate alınırken uygulamalarda standart sapma daha çok kullanılır.

1.5 KESİRLİ (AYRIK) VE SÜREKLİ VERİ

Gözlemsel çalışmaların veya tasarlanmış deneylerin analizi yoluyla istatistiksel çıkarım, birçok bilimsel alanda kullanılmaktadır. Toplanan veriler, uygulama alanına bağlı olarak kesikli veya sürekli olabilir. Örneğin, bir kimya mühendisi, verimi en üst düzeye çıkaran koşullara yol açacak bir deney yapmakla ilgilenebilir. Elbette burada verim, bir süreklilik üzerinde ölçülen yüzde veya gram/pound cinsinden olabilir. Öte yandan, bir kombinasyon ilaç deneyi yürüten bir toksikolog, doğası gereği ikili olan verilerle karşılaşabilir (yani, hasta yanıt verir veya vermez).

İstatistiksel çıkarımlar yapmamızı sağlayan olasılık teorisinde kesikli ve sürekli veriler arasında büyük ayrımlar yapılır. Genellikle istatistiksel çıkarım uygulamaları, veriler sayım verileri olduğunda bulunur. Örneğin, bir mühendis, örneğin 1 milisaniyede bir sayaçtan geçen radyoaktif parçacıkların sayısını incelemekle ilgilenebilir. Bir liman tesisinin verimliliğinden sorumlu personel, belirli bir liman kentine her gün gelen petrol tankerlerinin sayısı ile ilgilenebilir. Bölüm 5'te, sayım verilerinin olduğu durumlar için verileri işlemenin farklı yollarına yol açan birkaç farklı senaryo tartışılmaktadır.

Dersin bu ilk aşamasında bile ikili verilerle ilgili bazı ayrıntılara özel dikkat gösterilmelidir. İkili verilerin istatistiksel analizini gerektiren uygulamalar çoktur. Genellikle analizde kullanılan ölçü örnek oranıdır. Açıkçası, ikili durum iki kategori içerir. Verilerde yer alan n birim varsa ve x , kategori 1'e giren sayı olarak tanımlanıyorsa, o zaman $n - x$, kategori 2'ye girer. Böylece, x/n , kategori 1'deki örneklem oranıdır ve $1 - x/n$ kategori 2'deki numune oranıdır. Biyomedikal uygulamada, 50 hasta numune birimlerini temsil edebilir ve eğer 50 hastadan 20'si (50'sinde ortak olan) mide rahatsızlığında ilaç verildikten sonra bir iyileşme yaşıyorsa, o zaman $20/50 = 0,4$, ilacın başarılı olduğu örnek oranıdır ve $1 - 0,4 = 0,6$, ilacın başarılı olmadığı örnek oranıdır. Aslında, ikili veriler için temel sayısal ölçüm genellikle 0 veya 1 ile gösterilir. Örneğin, tıbbi örneğimizde, başarılı bir sonuç 1 ve başarısız bir sonuç 0 ile gösterilir. Sonuç olarak, örneklem oranı aslında birler ve sıfırların bir örneklem ortalamasıdır. Başarılı kategori için,

$$\frac{x_1 + x_2 + \dots + x_{50}}{50} = \frac{1 + 1 + 0 + \dots + 0 + 1}{50} = \frac{20}{50} = 0.4.$$

A. İkili Veri Durumlarında Ne Tür Problemler Çözülür?

İkili verilerle uğraşan bilim adamlarının ve mühendislerin karşılaştığı problem türleri, sürekli ölçümlerin ilgi çekici olduğu durumlarda görülenlerden pek farklı değildir. Bununla birlikte, örneklem oranlarının istatistiksel özellikleri, sürekli popülasyonlardan alınan ortalamalardan kaynaklanan örneklem ortalamalarından oldukça farklı olduğu için farklı teknikler kullanılmaktadır. Sayfa 33'teki Alıştırma 1.6'daki örnek verileri ele alalım. Bu çizimin altında yatan istatistiksel problem, bir müdahalenin, örneğin sertleştirme sıcaklığındaki bir artışın, silikon kauçuk işlemiyle ilişkili popülasyon ortalama gerilme mukavemetini değiştirip değiştirmeyeceğine odaklanır. Öte yandan, bir kalite kontrol alanında, bir otomobil lastiği üreticisinin, süreçten rastgele seçilen 5000 lastiğin sevkiyatında 100 tanesinde kusur olduğunu bildirdiğini varsayalım. Burada örneklem oranı $100/5000 = 0.02$ 'dir. Lekeleri azaltmak için tasarlanan süreçte değişiklik yapıldıktan sonra 5000 adetlik ikinci bir numune alınarak 90 adet lastik aşındırılır. Örnek oranı $90/5000 = 0,018$ 'e düşürülmüştür. Şu soru ortaya çıkıyor: "Numune oranındaki 0,02'den 0,018'e düşüş, popülasyon oranında gerçek bir iyileşme önermek için yeterince önemli mi?" Bu çizimlerin her ikisi de örnek ortalamalarının istatistiksel özelliklerinin kullanılmasını gerektirir - biri sürekli bir popülasyondan alınan örneklerden, diğeri ayrı (ikili) bir popülasyondan alınan örneklerden. Her iki durumda da, numune ortalaması, bir popülasyon parametresinin bir tahminidir, ilk resimde bir popülasyon ortalaması (yani, ortalama gerilme mukavemeti) ve ikinci durumda bir popülasyon oranıdır (yani,

popülasyondaki kusurlu lastiklerin oranı). Yani burada popülasyon parametreleriyle ilgili bilimsel sonuçlar çıkarmak için kullanılan örnek tahminlerimiz var. Bölüm 1.3'te belirttiğimiz gibi, istatistiksel çıkarım kullanan birçok pratik problemde genel tema budur.

1.6 İSTATİSTİKSEL MODELLEME, BİLİMSEL İNCELEME VE GRAFİK TEŞHİS

Genellikle istatistiksel bir analizin nihai sonucu, varsayılan bir modelin parametrelerinin tahminidir. Genellikle modelleme ile uğraştıkları için bilim adamları ve mühendisler için bu doğaldır. İstatistiksel bir model deterministik değildir, bunun yerine bazı olasılıksal yönler içermelidir. Bir model formu genellikle analist tarafından yapılan varsayımların temelidir. Örneğin, Örnek 1.2'de bilim insanı, numune bilgileri yoluyla nitrojen ve nitrojen olmayan popülasyonlar arasında bir düzeyde ayırım yapmak isteyebilir. Analiz, veriler için belirli bir model gerektirebilir, örneğin, iki örneğin normal veya Gauss dağılımlarından gelmesi.

Açıkçası, istatistiksel yöntemlerin kullanıcısı, popülasyonu tamamen karakterize etmek için yeterli bilgi veya deneysel veri üretemez. Ancak veri kümeleri genellikle popülasyonun belirli özelliklerini öğrenmek için kullanılır. Bilim adamları ve mühendisler veri kümeleriyle uğraşmaya alışkındır. Veri koleksiyonlarının doğasını karakterize etmenin veya özetlemenin önemi açık olmalıdır. Çoğu zaman, bir grafik gösterim yoluyla bir veri koleksiyonunun özeti, verilerin alındığı sistem hakkında fikir verebilir. Örneğin, Bölüm 1.1 ve 1.3'te nokta grafiklerini gösterdik.

Bu bölümde, örneklemenin rolü ve istatistiksel çıkarımın geliştirilmesi için verilerin görüntülenmesi ayrıntılı olarak incelenmektedir. Biz sadece, istatistiksel popülasyonların incelenmesini tamamlayan bazı basit ama genellikle etkili gösterimleri tanıtıyoruz.

A. Dağılım Grafiği

Zaman zaman öne sürülen model biraz karmaşık bir biçim alabilir. Örneğin, çeşitli yüzdelerde pamuk içeren kumaş numunesinin üretildiği bir deney tasarlayan bir tekstil imalatçısını ele alalım. Tablo 1.3'teki verileri göz önünde bulundurun.

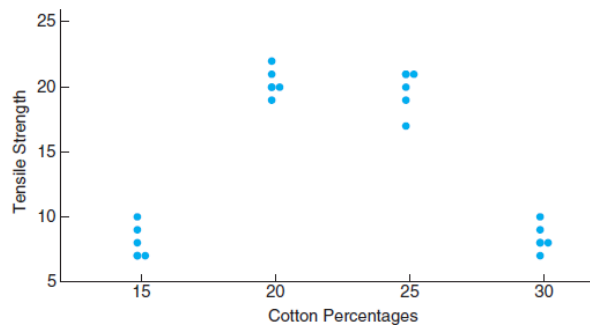
Tablo 1.3: Çekme Dayanımı

Cotton Percentage	Tensile Strength
15	7, 7, 9, 8, 10
20	19, 20, 21, 20, 22
25	21, 21, 17, 19, 20
30	8, 7, 8, 9, 10

Dört pamuk yüzdesinin her biri için beş kumaş numunesi üretilir. Bu durumda, hem deney modeli hem de kullanılan analiz türü, deneyin amacını ve tekstil bilimcisinden önemli girdileri dikkate almalıdır. Bazı basit grafikler, örnekler arasındaki net ayrıma önemli ışık tutabilir. Şekil 1.5'te örnek araçlar ve değişkenlik, dağılım grafiğinde güzel bir şekilde tasvir edilmiştir. Bu deneyin olası bir amacı, basitçe hangi pamuk yüzdelерinin diğerlerinden gerçekten farklı olduğunu belirlemektir. Başka bir deyişle, nitrojen/nitrojen içermeyen verilerde olduğu gibi, hangi pamuk yüzdeleri için popülasyonlar arasında veya daha spesifik olarak popülasyon ortalamaları arasında net ayrımlar vardır? Bu durumda, belki de makul bir model, her numunenin normal bir dağılımdan gelmesidir. Buradaki amaç, daha fazla numunenin dahil edilmesi dışında nitrojen/nitrojen içermeyen verilere çok benzer. Analizin biçimselliği, Bölüm 10'da tartışılan hipotez testi kavramlarını içerir. Bu arada, bu formalite, tanılama planının ışığında belki de gerekli değildir. Ancak bu, deneyin gerçek amacını ve dolayısıyla veri analizine uygun yaklaşımı tanımlıyor mu? Bilim adamının, deneyde pamuk konsantrasyonu aralığında maksimum popülasyon ortalama gerilme mukavemetinin varlığını tahmin etmesi muhtemeldir. Burada verilerin analizi, popülasyon ortalama gerilme mukavemetini pamuk konsantrasyonu ile ilişkilendiren bir yapı tipini öne süren farklı bir model tipi etrafında dönmelidir. Başka bir deyişle, şu şekilde bir model yazılabilir

$$\mu_{t,c} = \beta_0 + \beta_1 C + \beta_2 C^2,$$

$\mu_{t,c}$, C ürünündeki pamuk miktarına göre değişen popülasyon ortalama gerilme mukavemetidir. Bu modelin anlamı, sabit bir pamuk seviyesi için, bir popülasyon çekme mukavemeti ölçümleri olduğu ve popülasyon ortalamasının $\mu_{t,c}$ olduğudur. İşlevsel biçim bilim adamı tarafından seçilir. Bazen veri analizi modelin değiştirilmesini önerebilir. Ardından veri analisti, bazı analizler yapıldıktan sonra değiştirilebilecek bir modeli "eğlendirir". Ampirik bir modelin kullanımına, β_0 , β_1 ve β_2 'nin verilerle tahmin edildiği tahmin teorisi eşlik eder. Ayrıca, model yeterliliğini belirlemek için istatistiksel çıkarım kullanılabilir.



Şekil 1.5: Kopma mukavemeti ve pamuk yüzdelерinin dağılım grafiği.

Buradaki iki veri örneğinden iki nokta belirginleşiyor: (1) Verileri tanımlamak için kullanılan modelin türü genellikle deneyin amacına bağlıdır; ve (2) modelin yapısı istatistiksel olmayan bilimsel girdilerden yararlanmalıdır. Bir model seçimi, ortaya çıkan istatistiksel çıkarımın dayandığı temel bir varsayımı temsil eder. Grafiklerin ne kadar önemli olabileceği kitap boyunca ortaya çıkacak. Grafikler genellikle, resmi istatistiksel çıkarımın sonuçlarının bilim insanına veya mühendise daha iyi iletilmesini sağlayan bilgileri gösterebilir. Bazen, çizimler veya keşifsel veri analizi, analiste resmi analizden elde edilemeyen bir şey öğretebilir. Neredeyse tüm resmi analizler, veri modelinden gelişen varsayımları gerektirir. Grafikler, aksi takdirde fark edilmeyecek olan varsayımların ihlalini güzel bir şekilde vurgulayabilir. Kitap boyunca, resmi veri analizini desteklemek için grafikler yaygın olarak kullanılmıştır. Aşağıdaki bölümler, keşfedici veya tanımlayıcı veri analizinde yararlı olan bazı grafik araçları göstermektedir.

B. Gövde-Yaprak Grafiği

Büyük kütleler halinde üretilen istatistiksel veriler, gövde-yaprak çizimi adı verilen birleşik tablo ve grafik gösterimde sunulursa, dağılımın davranışını incelemek için çok yararlı olabilir.

Bir gövde-yaprak grafiğinin yapısını göstermek için, bir yılın en yakın onda biri olarak kaydedilen 40 benzer araba aküsünün "ömrünü" belirten Tablo 1.4'teki verileri göz önünde bulundurun. Piller 3 yıl garantilidir. İlk olarak, her gözlemi bir kök ve bir yapraktan oluşan iki kısma ayırın, böylece gövde ondalıktan önceki basamağı temsil eder ve yaprak sayısının ondalık kısmına karşılık gelir. Başka bir deyişle, 3.7 sayısı için 3 rakamı gövdeyi, 7 rakamı ise yaprağı belirtir. Verilerimiz için dört gövde 1, 2, 3 ve 4, Tablo 1.5'te sol tarafta dikey olarak listelenmiştir; yapraklar, uygun gövde değerinin karşısındaki sağ tarafa kaydedilir. Böylece, 1.6 sayısının yaprağı 6, gövdenin 1 karşısında kaydedilir; 2.5 sayısının yaprak 5'i gövde 2'nin karşısına kaydedilir. Her gövdenin karşısında kaydedilen yaprak sayısı, frekans sütununda özetlenmiştir.

Tablo 1.4: Araç Pil Ömrü

2.2	4.1	3.5	4.5	3.2	3.7	3.0	2.6
3.4	1.6	3.1	3.3	3.8	3.1	4.7	3.7
2.5	4.3	3.4	3.6	2.9	3.3	3.9	3.1
3.3	3.1	3.7	4.4	3.2	4.1	1.9	3.4
4.7	3.8	3.2	2.6	3.9	3.0	4.2	3.5

Tablo 1.5: Pil Ömrünün Gövde ve Yaprak Grafiği

Stem	Leaf	Frequency
1	69	2
2	25669	5
3	0011112223334445567778899	25
4	11234577	8

Tablo 1.5'teki gövde-yaprak grafiği yalnızca dört gövde içerir ve sonuç olarak dağılımın yeterli bir resmini sağlamaz. Bu sorunu çözmek için arsamızdaki sap sayısını artırmamız gerekiyor. Bunu başarmanın basit bir yolu, her gövde değerini iki kez yazmak ve ardından ilk kez görüldüğü uygun kök değerinin karşısındaki 0, 1, 2, 3 ve 4 yapraklarını ve 5, 6, 7 yapraklarını kaydetmektir. 8 ve 9, ikinci kez görüldüğü aynı gövde değerinin karşısında. Bu değiştirilmiş çift gövde ve yaprak grafiği, Tablo 1.6'da gösterilmektedir; burada, 0 ile 4 numaralı yapraklara karşılık gelen gövdeler, sembolle ve 5 ile 9 numaralı yapraklara karşılık gelen gövdeler, $\cdot$ simgesiyle kodlanmıştır.

Tablo 1.6: Pil Ömrünün Çift Gövde ve Yaprak Grafiği

Stem	Leaf	Frequency
1 $\cdot$	69	2
2 $\star$	2	1
2 $\cdot$	5669	4
3 $\star$	001111222333444	15
3 $\cdot$	5567778899	10
4 $\star$	11234	5
4 $\cdot$	577	3

Herhangi bir problemde, uygun kök değerlere karar vermeliyiz. Örneğimizin büyüklüğü bize rehberlik etse de, bu karar biraz keyfi olarak veriliyor. Genellikle 5 ile 20 gövde arasından seçim yaparız. Mevcut veri sayısı ne kadar azsa, kök sayısı için tercihimiz o kadar az olur. Örneğin, veriler 1'den 21'e kadar rastgele seçilen 40 iş gününde bir kafeterya hattındaki kişi sayısını temsil eden sayılardan oluşuyorsa ve çift gövde ve yaprak grafiği seçersek, gövdeler 0, 0 $\cdot$, 1 olacaktır. , 1 $\cdot$ ve 2 böylece en küçük gözlem 1'de gövde 0 ve yaprak 1, 18 sayısında gövde 1 $\cdot$ ve yaprak 8 ve en büyük gözlem 21'de gövde 2 ve yaprak 1 bulunur. Öte yandan, veriler, belirli bir bayiden 100 yeni otomobil için mümkün olan en iyi anlaşmaları temsil eden 18.800 ile 19.600 \$ arasındaki rakamlardan oluşuyorsa ve tek gövdeli bir arsa seçersek, gövdeler 188, 189, 190, olacaktır. . . , 196 ve yaprakların her biri artık iki basamak içerecektir. 19.385 dolara satılan bir arabanın gövde değeri 193 ve iki basamaklı yaprak 85 olacaktır. Aynı

gövdeye ait çok basamaklı yapraklar genellikle gövde-yaprak çiziminde virgülle ayrılır. Ondalık basamağın sağındaki tüm basamaklar yaprağı temsil ettiğinde, verilerdeki ondalık noktalar genellikle göz ardı edilir. Tablo 1.5 ve 1.6'daki durum böyledir. Bununla birlikte, veriler 21.8 ile 74.9 arasında değişen sayılardan oluşuyorsa, 2, 3, 4, 5, 6 ve 7 rakamlarını kök olarak seçebiliriz, böylece 48.3 gibi bir sayının gövde değeri 4 ve a 8.3 yaprağı.

Gövde ve yaprak grafiği, verileri özetlemenin etkili bir yolunu temsil eder. Başka bir yol, farklı sınıflar veya aralıklar halinde gruplanan verilerin, her bir gövdeye ait yaprakların sayılması ve her bir gövdenin bir sınıf aralığını tanımladığına dikkat edilerek oluşturulabileceği frekans dağılımının kullanılmasıdır. Tablo 1.5'te, 2 yapraklı gövde 1, 2 gözlem içeren 1.0–1.9 aralığını tanımlar; 5 yapraklı gövde 2, 5 gözlem içeren 2.0–2.9 aralığını tanımlar; 25 yapraklı gövde 3, 25 gözlemleri 3.0–3.9 aralığını tanımlar; ve 8 yapraklı gövde 4, 8 gözlem içeren 4.0–4.9 aralığını tanımlar. Tablo 1.6'daki çift gövde ve yaprak grafiği için gövdeler yedi sınıf aralığını tanımlar: 1.5–1.9, 2.0–2.4, 2.5–2.9, 3.0–3.4, 3.5–3.9, 4.0–4.4 ve sırasıyla 2, 1, 4, 15, 10, 5 ve 3 frekansları.

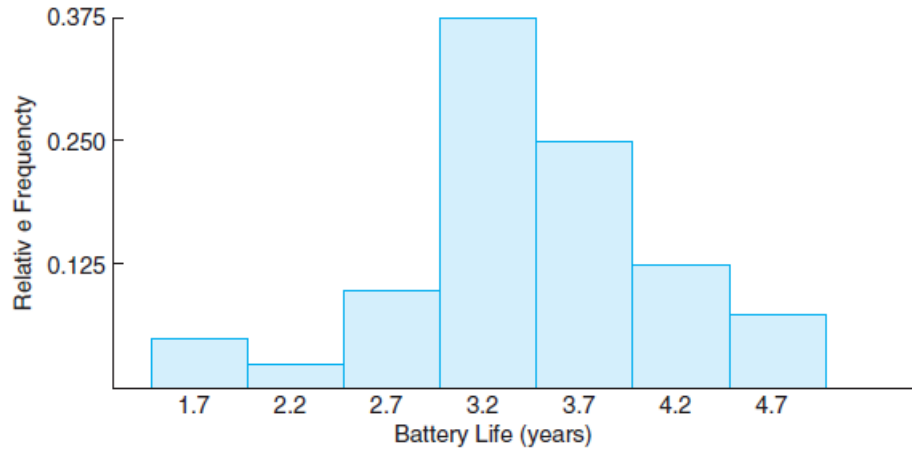
C. Histogram

Her sınıf frekansını toplam gözlem sayısına bölerek, her bir sınıftaki gözlem setinin oranını elde ederiz. Göreceli frekansları listeleyen bir tabloya göreli frekans dağılımı denir. Her bir sınıf aralığının orta noktasını gösteren Tablo 1.4'teki verilere ilişkin göreli frekans dağılımı Tablo 1.7'de verilmiştir.

Tablo 1.7: Pil Ömrünün Bağlı Frekans Dağılımı

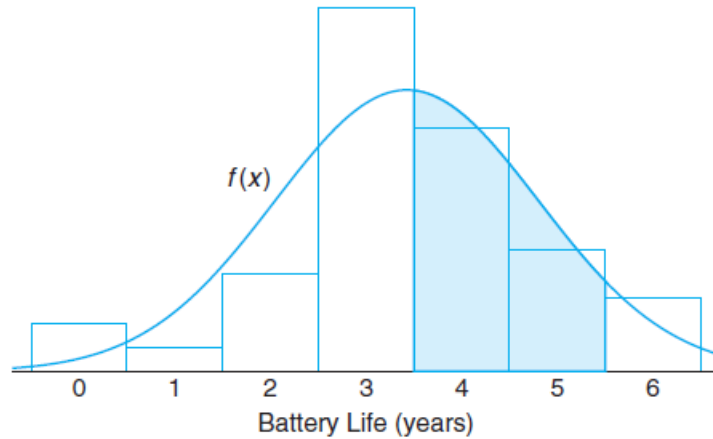
Class Interval	Class Midpoint	Frequency, f	Relative Frequency
1.5–1.9	1.7	2	0.050
2.0–2.4	2.2	1	0.025
2.5–2.9	2.7	4	0.100
3.0–3.4	3.2	15	0.375
3.5–3.9	3.7	10	0.250
4.0–4.4	4.2	5	0.125
4.5–4.9	4.7	3	0.075

Tablo şeklinde göreceli bir frekans dağılımı tarafından sağlanan bilgilerin, grafik olarak sunulduğunda anlaşılması daha kolaydır. Her aralığın orta noktasını ve karşılık gelen bağlı frekansı kullanarak, bir göreli frekans histogramı oluştururuz (Şekil 1.6).



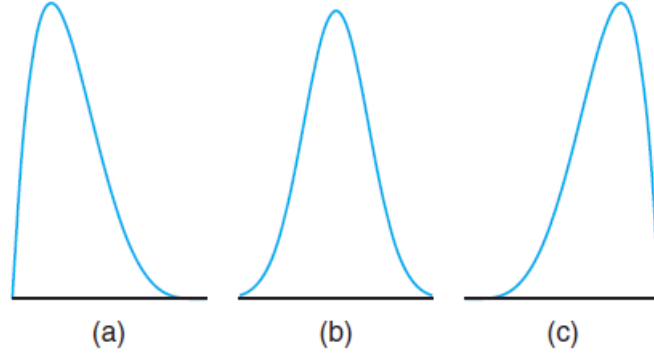
Şekil 1.6: Bağlı frekans histogramı.

Birçok sürekli frekans dağılımı, Şekil 1.7'deki karakteristik çan şeklindeki eğri ile grafiksel olarak temsil edilebilir. Şekil 1.6 ve 1.7'de gördüğümüz gibi grafik araçlar, nüfusun doğasının tanımlanmasına yardımcı olur. 5. ve 6. Bölümlerde, popülasyonun dağılımı adı verilen bir özelliğini tartışıyoruz. Bir dağılımın veya olasılık dağılımının daha kesin bir tanımı metnin ilerleyen kısımlarında verilecek olsa da, bu noktada örnek boyutu büyüdükçe sınırdaki Şekil 1.7'de görüleceği gibi görülebilir.



Şekil 1.7: Frekans dağılımını tahmin etme.

Bir dağılımın, iki kenarı çakışacak şekilde dikey bir eksen boyunca katlanabilmesi durumunda simetrik olduğu söylenir. Dikey bir eksene göre simetrisi olmayan bir dağılımın çarpık olduğu söylenir. Uzun bir sağ kuyruğa ve çok daha kısa bir sol kuyruğa sahip olduğundan, Şekil 1.8(a)'da gösterilen dağılımın sağa çarpık olduğu söylenmektedir. Şekil 1.8(b)'de dağılımın simetrik olduğunu, Şekil 1.8(c)'de ise sola çarpık olduğunu görüyoruz.



Şekil 1.8: Verilerin çarpıklığı.

Bir gövde-yaprak grafiğini 90°'lik bir açıyla saat yönünün tersine döndürürsek, ortaya çıkan yaprak sütunlarının histograma benzer bir resim oluşturduğunu gözlemleriz. Sonuç olarak, verilere bakmadaki birincil amacımız dağılımın genel şeklini veya biçimini belirlemekse, göreceli bir frekans histogramı oluşturmak nadiren gerekli olacaktır.

D. İki Çubuk Çizişi veya Kutu Grafiği

Bir numunenin özelliklerini yansıtmaya yardımcı olan başka bir ekran, kutu çizimidir. Bu çizim, verilerin çeyrekler arası aralığını, medyanın içinde görüntülendiği bir kutuya alır. Çeyrekler arası aralığın uç noktaları olarak 75. yüzdelik dilim (üst çeyrek) ve 25. yüzdelik dilim (alt çeyrek) vardır. Oldukça büyük numuneler için, ekran konumun merkezini, değişkenliği ve asimetri derecesini gösterir.

Ek olarak, kutu çizimi adı verilen bir varyasyon, izleyiciye hangi gözlemlerin aykırı değerler olabileceği konusunda bilgi sağlayabilir. Aykırı değerler, verilerin büyük bir kısmından alışılmadık şekilde uzak olduğu düşünülen gözlemlerdir. Aykırı değerleri tespit etmek için tasarlanmış birçok istatistiksel test vardır. Teknik olarak, bir aykırı değer "nadir bir olayı" temsil eden bir gözlem olarak görülebilir (veri yığınının bu kadar uzakta bir değer elde etme olasılığı düşüktür). Aykırı değerler kavramı, Bölüm 12'de regresyon analizi bağlamında yeniden ortaya çıkıyor.

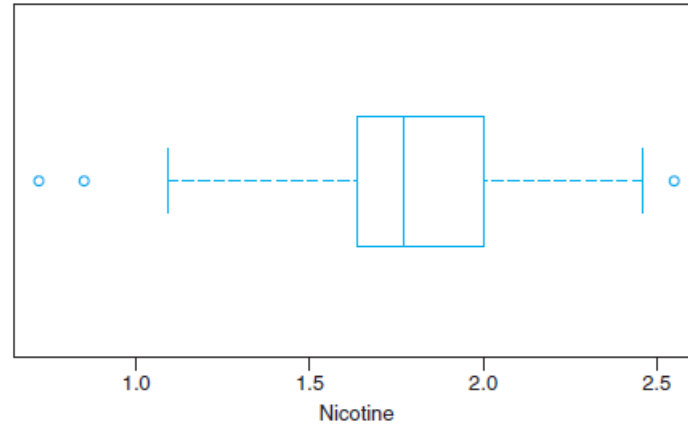
Kutu grafiğindeki veya kutu grafiğindeki görsel bilgilerin, aykırı değerler için resmi bir test olması amaçlanmamıştır. Aksine, bir teşhis aracı olarak görülüyor. Hangi gözlemlerin aykırı olduğunun belirlenmesi, kullanılan yazılımın türüne göre değişirken, yaygın bir prosedür, çeyrekler arası aralığın katlarını kullanmaktır. Örneğin, kutudan uzaklık çeyrekler arası aralığın 1,5 katını aşarsa (her iki yönde), gözlem bir aykırı değer olarak etiketlenebilir.

Örnek 1.5: Nikotin içeriği rastgele 40 sigara örneğinde ölçüldü. Veriler Tablo 1.8'de gösterilmektedir.

Tablo 1.8: Örnek 1.5 için Nikotin Verileri

1.09	1.92	2.31	1.79	2.28	1.74	1.47	1.97
0.85	1.24	1.58	2.03	1.70	2.17	2.55	2.11
1.86	1.90	1.68	1.51	1.64	0.72	1.69	1.85
1.82	1.79	2.46	1.88	2.08	1.67	1.37	1.93
1.40	1.64	2.09	1.75	1.63	2.37	1.75	1.69

Şekil 1.9, 0.72 ve 0.85 gözlemlerini alt kuyrukta hafif aykırı değerler olarak, 2.55 gözlemi ise üst kuyrukta hafif aykırı değerler olarak gösteren, verilerin kutu ve bıyık grafiğini göstermektedir. Bu örnekte, çeyrekler arası aralık 0,365'tir ve çeyrekler arası aralığın 1,5 katı 0,5475'tir. Öte yandan Şekil 1.10, bir dal-yaprak çizimi sağlar.



Şekil 1.9: Örnek 1.5 için kutu grafiği.

```

The decimal point is 1 digit(s) to the left of the |
 7 | 2
 8 | 5
 9 |
10 | 9
11 |
12 | 4
13 | 7
14 | 07
15 | 18
16 | 3447899
17 | 045599
18 | 2568
19 | 0237
20 | 389
21 | 17
22 | 8
23 | 17
24 | 6
25 | 5

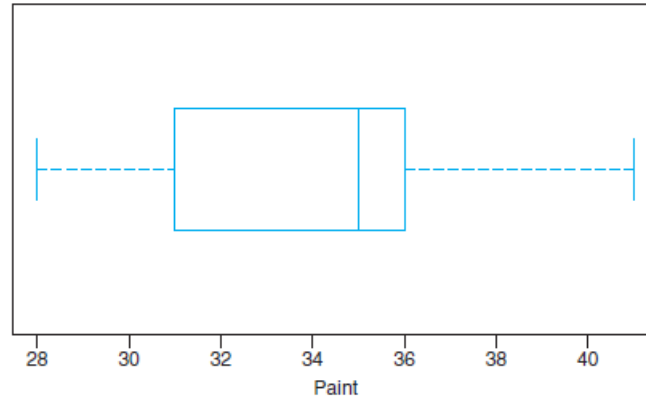
```

Şekil 1.10: Nikotin verileri için gövde ve yaprak grafiği.

Örnek 1.6: Boya kutusunun "kulaklar" kalınlığını ölçen 30 örnekten oluşan Tablo 1.9'daki verileri göz önünde bulundurun (Kaynakça'da Hogg ve Ledolter, 1992'nin çalışmasına bakın). Şekil 1.11, bu asimetrik veri seti için bir kutu ve bıyık grafiğini göstermektedir. Sol bloğun sağdaki bloktan önemli ölçüde daha büyük olduğuna dikkat edin. Medyan 35'tir. Alt çeyrek 31, üst çeyrek ise 36'dır. Ayrıca sağdaki aşırı gözlemin soldaki aşırı gözleminden daha uzak olduğuna dikkat edin. Bu veri setinde aykırı değerler yok.

Tablo 1.9: Örnek 1.6 için veriler

Sample	Measurements	Sample	Measurements
1	29 36 39 34 34	16	35 30 35 29 37
2	29 29 28 32 31	17	40 31 38 35 31
3	34 34 39 38 37	18	35 36 30 33 32
4	35 37 33 38 41	19	35 34 35 30 36
5	30 29 31 38 29	20	35 35 31 38 36
6	34 31 37 39 36	21	32 36 36 32 36
7	30 35 33 40 36	22	36 37 32 34 34
8	28 28 31 34 30	23	29 34 33 37 35
9	32 36 38 38 35	24	36 36 35 37 37
10	35 30 37 35 31	25	36 30 35 33 31
11	35 30 35 38 35	26	35 30 29 38 35
12	38 34 35 35 31	27	35 36 30 34 36
13	34 35 33 30 34	28	35 30 36 29 35
14	40 35 34 33 35	29	38 36 35 31 31
15	34 35 38 35 30	30	30 34 40 28 30



Şekil 1.11: “Kulaklar” boyasının kalınlığı için kutu grafiği.

Kutu grafiklerinin ve diğer grafiksel gösterimlerin analiste yardımcı olabileceği başka yollar da vardır. Birden çok örnek grafiksel olarak karşılaştırılabilir. Veri grafikleri, değişkenler arasındaki ilişkileri önerebilir. Grafikler, örneklerde anormalliklerin veya aykırı gözlemlerin saptanmasına yardımcı olabilir.

E. İki Çubuk Çizişi veya Kutu Grafiği

Dağılımın veya numunenin, doğasını daha da tanımlayan konum merkezi ve değişkenlik ölçülerinden başka özellikleri vardır. Örneğin, medyan verileri (veya dağılımı) iki kısma ayırırken, dağılımın kısımlarını veya parçalarını bölen ve çok yararlı olabilecek başka ölçüler de vardır. Üçüncü çeyrek verinin üst çeyreğini diğerlerinden, ikinci çeyrek ortancayı ve birinci çeyrek verinin alt çeyreğini diğerlerinden ayıracak şekilde kartillere göre dört parçaya ayrılmıştır. Dağılım, dağılımın hesaplama yüzdelikleri ile daha da ince bir şekilde bölünebilir. Bu nicelikler, analiste dağılımın sözde kuyukları hakkında bir fikir verir (yani, nispeten aşırı, küçük veya büyük değerler). Örneğin, 95. yüzdelik dilim, en yüksek %5'i en alttaki %95'ten ayırır. Dağılımın alt tarafındaki veya alt kuyruğundaki uç noktalar için benzer tanımlar geçerlidir. 1. yüzdelik dilim, en alttaki %1'i dağılımın geri kalanından ayırır. Yüzdelikler kavramı, gelecek bölümlerde ele alınacak olan birçok konuda önemli bir rol oynayacaktır.

1.7 GENEL İSTATİSTİKSEL ÇALIŞMA TÜRLERİ: TASARLANMIŞ DENEY, GÖZLEMSEL ÇALIŞMA VE RETROSPEKTİF ÇALIŞMA

Önceki bölümlerde, bir popülasyondan örnekleme kavramını ve popülasyon hakkında önemli bilgileri öğrenmek veya belki de doğrulamak için istatistiksel yöntemlerin kullanımını vurguladık. Bu istatistiksel yöntemlerin kullanımıyla aranan ve öğrenilen bilgiler, birçok önemli bilim ve mühendislik alanında karar vermede ve problem çözmede sıklıkla etkili olabilir. Örnek olarak, Örnek 1.3, korozyona karşı tehlikeli bir hassasiyete sahip olabilecek belirli bir alüminyum alaşımının kullanılmasının tavsiye edilmediği koşulların belirlenmesinde sonuçların yardımcı olabileceği basit bir deneyi açıklamaktadır. Sonuçlar, yalnızca alaşımı üretenler için değil, aynı zamanda onu kullanmayı düşünen müşteriler için de faydalı olabilir. Bu resim ve 13'ten 15'e kadar olan Bölümlerde yer alan daha pek çoğu, bazı özellikler veya ölçümler (korozyon seviyesi) hakkında bilgi edinmek için ilgilenilen deneysel koşulların (kaplama koşulları ve nem kombinasyonları) tasarlanması veya kontrol edilmesi kavramını vurgulamaktadır. bu şartlar. Korozyon ölçüsünde merkezi eğilim ölçülerini ve değişkenlik ölçülerini kullanan istatistiksel yöntemler kullanılır. Okuyucunun metnin ilerleyen kısımlarında gözlemleyeceği gibi, bu yöntemler genellikle Bölüm 1.6'da tartışılan gibi bir istatistiksel modele yol açar. Bu durumda model, nemin ve kullanılan kaplama tipinin bir fonksiyonu olarak korozyon ölçüsünü tahmin etmek (veya tahmin etmek) için kullanılabilir. Yine bu tür bir model geliştirirken, merkezi eğilim ve değişkenliği vurgulayan betimsel istatistikler çok faydalı olmaktadır.

Örnek 1.3'te verilen bilgiler, tasarlanmış bir deney aracılığıyla kullanılan ve bu metinde sunulan istatistiksel yöntemler kullanılarak sorulan ve yanıtlanan mühendislik sorularının türlerini güzel bir şekilde göstermektedir. Bunlar

- (i) Bu deneyde bağıl nemin alüminyum alaşımının bağıl nem aralığındaki korozyonu üzerindeki etkisinin doğası nedir?
- (ii) Kimyasal korozyon kaplaması, korozyon seviyelerini azaltır mı ve etki bir şekilde ölçülebilir mi?
- (iii) Kaplama tipi ile bağıl nem arasında alaşımın korozyonu üzerindeki etkilerini etkileyen etkileşim var mı? Eğer öyleyse, bunun yorumu nedir?

A. Etkileşim Nedir?

Alaşımın hem üreticileri hem de kullanıcıları için önemli olan konuları ele aldıklarından (i) ve (ii) sorularının önemi okuyucu için net olmalıdır. Peki ya soru (iii)? Etkileşim kavramı, 14. ve 15. Bölümlerde uzun uzadıya tartışılacaktır. Şekil 1.3'teki grafiği ele alalım. Bu, basit tasarlanmış bir deneyde iki faktör arasındaki etkileşimin saptanmasının bir örneğidir. Numune araçlarını bağlayan hatların paralel olmadığına dikkat edin. Paralellik, bağıl nemin etkisinin (çizgilerin eğiminin bir sonucu olarak görülen) aynı olduğunu, yani hem kaplanmamış bir durum hem de kimyasal korozyon kaplaması için olumsuz bir etki olduğunu gösterirdi. Negatif eğimin, nem arttıkça korozyonun daha belirgin hale geldiği anlamına geldiğini hatırlayın. Paralelliğin olmaması, kaplama tipi ile bağıl nem arasında bir etkileşim olduğunu gösterir. Kaplanmamış durum için daha dik bir eğimin aksine, korozyon kaplaması için neredeyse "düz" çizgi, yalnızca kimyasal korozyon kaplamasının faydalı olmadığını (çizgiler arasındaki yer değiştirmeye dikkat edin), ayrıca kaplamanın varlığının nemin etkisini oluşturduğunu gösterir. ihmal edilebilir Açıkçası tüm bu sorular, iki ayrı faktörün etkisi ve varsa etkileşimin yorumlanması için çok önemlidir.

İstatistiksel modeller, verilerin tasarlanmış bir deneyden geldiği (i), (ii) ve (iii)'de listelenenler gibi soruları yanıtlamada son derece yararlıdır. Ancak, tasarlanmış bir deneyi kullanmak için her zaman lüks veya kaynaklara sahip olunmaz. Örneğin, bilim insanı veya mühendisin ilgilendiği koşulların, yalnızca önemli faktörler kontrol edilemediği için uygulanamadığı birçok durum vardır. Örnek 1.3'te, bağıl nem ve kaplama türünün (veya kaplama eksikliğinin) kontrol edilmesi oldukça kolaydır. Bu, elbette tasarlanmış bir deneyin tanımlayıcı özelliğidir. Pek çok alanda, çalışılması gereken faktörler çeşitli nedenlerle kontrol

edilememektedir. Örnek 1.3'teki gibi sıkı kontrol, analistin bulunan herhangi bir farklılığın (örneğin korozyon seviyelerinde) kontrol altındaki faktörlerden kaynaklandığından emin olmasını sağlar. İkinci bir örnek olarak, sayfa 33'teki Alistırma 1.6'yı ele alalım. Bu durumda 24 silikon kauçuk örneğinin seçildiğini ve kürlleme sıcaklık seviyelerinin her birine 12 atandığını varsayalım. Sıcaklıklar dikkatli bir şekilde kontrol edilir ve dolayısıyla bu, kürlleme sıcaklığı olan tek bir faktörün olduğu tasarlanmış bir deney örneğidir. Ortalama gerilme mukavemetinde bulunan farklılıkların, farklı kürlleme sıcaklıklarına atfedildiği varsayılacaktır.

B. Faktörler Kontrol Edilmezse Ne Olur?

Kontrol edilen hiçbir faktör olmadığını ve sabit tedavilerin deneysel birimlere rastgele atanmadığını ve yine de bir veri setinden bilgi toplamaya ihtiyaç olduğunu varsayalım. Bir örnek olarak, ilginin kandaki kolesterol seviyeleri ile kanda ölçülen sodyum miktarı arasındaki ilişkiye odaklandığı bir çalışmayı düşünün. Bir grup birey, hem kan kolesterolü hem de sodyum için zaman içinde izlendi. Elbette böyle bir veri setinden bazı yararlı bilgiler toplanabilir. Bununla birlikte, kan sodyum seviyelerinin kesinlikle sıkı bir şekilde kontrol edilemeyeceği açık olmalıdır. İdeal olarak, denekler rastgele iki gruba ayrılmalıdır; bir gruba belirli bir yüksek kan sodyum seviyesi ve diğerine belirli bir düşük kan sodyum seviyesi atanmalıdır. Açıkçası bu yapılamaz. Açıkça kolesteroldeki değişiklikler, kontrol edilmeyen bir dizi başka faktörden birindeki değişiklikler nedeniyle yaşanabilir. Faktör kontrolünün olmadığı bu tür çalışmalara gözlemsel çalışma denir. Çoğu zaman, öznelere zaman içinde gözlemlendiği bir durumu içerir.

Biyolojik ve biyomedikal çalışmalar genellikle zorunlu olarak gözlemsel çalışmalardır. Ancak, gözlemsel çalışmalar bu alanlarla sınırlı değildir. Örneğin, ortam sıcaklığının bir kimya fabrikası tarafından tüketilen elektrik gücü üzerindeki etkisini belirlemek için tasarlanmış bir çalışmayı ele alalım. Açıktır ki, ortam sıcaklığı seviyeleri kontrol edilemez ve bu nedenle veri yapısı, tesisten yalnızca zaman içindeki verilerin izlenmesi olabilir.

İyi tasarlanmış bir deney ile gözlemsel çalışmalar arasındaki çarpıcı farkın, ikincisiyle gerçek neden ve sonucu belirlemedeki zorluk olduğu açık olmalıdır. Ayrıca, temel tepkide bulunan farklılıklar (örneğin, korozyon seviyeleri, kan kolesterolü, tesis elektrik enerjisi tüketimi), kontrol edilmeyen diğer temel faktörlerden kaynaklanıyor olabilir. İdeal olarak, tasarlanmış bir deneyde rahatsız edici faktörler, rastgeleleştirme işlemi yoluyla eşitlenir. Kan kolesterolündeki değişikliklerin nedeni kesinlikle yağ alımı, egzersiz aktivitesi vb. olabilir. Elektrik enerjisi tüketimi, üretilen ürün miktarından ve hatta üretilen ürünün saflığından etkilenebilir.

Dikkatlice tasarlanmış deneylerle karşılaştırıldığında, gözlemsel bir çalışmanın genellikle göz ardı edilen diğer bir dezavantajı, ikincisinden farklı olarak, gözlemsel çalışmaların, ilgilenilen faktörlerin aralıklarını etkileyen doğanın, çevresel veya diğer kontrolsüz koşulların insafına kalmasıdır. Örneğin, kan sodyum seviyelerinin kan kolesterolü üzerindeki etkisine ilişkin biyomedikal çalışmada, gerçekten de güçlü bir etkinin olması mümkündür, ancak kullanılan belirli veri seti, deneklerin doğası gereği sodyum seviyelerinde yeterince gözlenen varyasyon içermemektedir. seçildi. Elbette, tasarlanmış bir deneyde, analist bir dizi faktör seçer ve kontrol eder.

Çok yararlı olabilen ancak tasarlanmış bir deneyle karşılaştırıldığında bariz dezavantajları olan üçüncü bir istatistiksel çalışma türü, retrospektif bir çalışmadır. Bu tür bir çalışma, kesinlikle tarihsel verileri, belirli bir zaman diliminde alınan verileri kullanır. Geriye dönük verilerin belirgin bir avantajı, veri toplama maliyetinin azalmasıdır. Bununla birlikte, tahmin edilebileceği gibi, açık dezavantajlar vardır.

(i) Tarihsel verilerin geçerliliği ve güvenilirliği genellikle şüphelidir.

(ii) Zaman, verilerin yapısının önemli bir yönüyse, eksik veriler olabilir.

(iii) Verilerin toplanmasında bilinmeyen hatalar olabilir.

(iv) Yine, gözlemsel verilerde olduğu gibi, ölçülen değişkenlerin (bir çalışmadaki faktörler) aralıkları üzerinde herhangi bir kontrol yoktur. Aslında, geçmiş verilerde bulunan aralıklar mevcut çalışmalar için uygun olmayabilir.

Bölüm 1.6'da, değişkenler arasındaki ilişkilerin modellenmesine biraz dikkat edilmiştir. Bölüm 11 ve 12'de ele alınan ve Bölüm 14 ve 15'te tartışılan tasarlanmış deneyler için bir veri analizi biçimi olarak gösterilen regresyon analizi kavramını tanıttık. Bölüm 1.6'da, kumaşın popülasyon ortalama gerilme mukavemetini yüzdelere ilişkilendiren bir model Deneyisel birimleri temsil eden 20 kumaş örneğinde, örnek olarak pamuk kullanılmıştır. Bu durumda veriler, bilim adamı tarafından bireysel pamuk yüzdelerinin seçildiği basit tasarlanmış bir deneyden geldi.

Genellikle, Bölüm 11 ve 12'de tartışılan model oluşturma prosedürleri yoluyla değişkenler arasındaki ilişkileri gözlemlemek amacıyla hem gözlemsel veriler hem de geriye dönük veriler kullanılır. Amaç istatistiksel model oluşturmak olduğunda, tasarlanmış deneylerin avantajları kesinlikle geçerli olsa da, birçok alan vardır. Hangi deneylerin tasarımı mümkün değildir. Bu nedenle, gözlemsel veya tarihsel veriler kullanılmalıdır. Burada, sayfa

472'deki Alıştırma 12.5'te bulunan tarihsel bir veri setine atıfta bulunuyoruz. Amaç, tüketilen aylık elektrik gücünü ortalama ortam sıcaklığı x_1 , gün sayısı ile ilişkilendiren bir denklem veya ilişki ile sonuçlanacak bir model oluşturmaktır. x_2 , ortalama ürün saflığı x_3 ve üretilen ürün tonu x_4 . Veriler, geçen yılın geçmiş verileridir.

OLASILIK

2.1 ÖRNEKLEM UZAYI

İstatistik çalışmasında, temel olarak planlı bir çalışmada veya bilimsel araştırmada ortaya çıkan tesadüfi sonuçların sunumu ve yorumlanmasıyla ilgilenilmektedir. Örneğin, bir trafik ışığının yerleştirilmesini gerekçelendirmeyi umarak, herhangi bir X kavşağında meydana gelen aylık kaza sayısını kaydedebiliriz; bir montaj hattından çıkan ürünleri "kusurlu" veya "kusursuz" olarak sınıflandırabiliriz veya bir asidin konsantrasyonu değiştiğinde kimyasal bir reaksiyonda salınan gazın hacmiyle ilgilenebiliriz. Bu nedenle, istatistikçi genellikle sayımları veya ölçümleri temsil eden sayısal verilerle veya bazı kriterlere göre sınıflandırılabilen kategorik verilerle uğraşmaktadır.

İster sayısal ister kategorik olsun, herhangi bir bilgi kaydına gözlem olarak atıfta bulunacağız. Böylece, geçen yıl boyunca Ocak'tan Nisan'a kadar her ay X kavşağında meydana gelen kazaların sayısını temsil eden 2, 0, 1 ve 2 sayıları bir dizi gözlem oluşturmaktadır. Benzer şekilde, beş öge incelendiğinde kusurlu veya hatasız bulunan öğeleri temsil eden N , D , N , N ve D kategorik verileri gözlem olarak kaydedilir.

İstatistikçiler, bir veri kümesi oluşturan herhangi bir işlemi tanımlamak için deney kelimesini kullanırlar. İstatistiksel bir deneyin basit bir örneği, yazı tura atmaktır. Bu deneyde, yalnızca iki olası sonuç vardır, tura veya yazı. Başka bir deney, bir füzenin fırlatılması ve belirli zamanlarda hızının gözlemlenmesi olabilir. Seçmenlerin yeni bir satış vergisiyle ilgili görüşleri de bir deney gözlemi olarak değerlendirilebilir. Deneyi birkaç kez tekrarlayarak elde edilen gözlemlerle özellikle ilgilenilmektedir. Çoğu durumda, sonuçlar şansa bağlıdır ve bu nedenle kesin olarak tahmin edilemez. Bir kimyager aynı koşullar altında birkaç kez bir analiz yaparsa, deneysel prosedürde bir şans unsuru olduğunu gösteren farklı ölçümler elde edecektir. Bir yazı tura tekrar tekrar atıldığında bile, belirli bir atışın tura geleceğinden emin olamayız. Ancak, her atış için tüm olasılıklar bilinmektedir.

Tanım 2.1: İstatistiksel bir deneyin tüm olası sonuçlarının oluşturduğu kümeye “örneklem uzay” denir ve S sembolü ile gösterilir.

Bir örneklem uzayındaki her sonuç, örneklem uzayının bir elemanı veya üyesi ya da sadece bir örneklem noktası olarak adlandırılır. Örnek uzayın sonlu sayıda elemanı varsa, elemanları virgülle ayırarak ve parantez içinde listeleyebiliriz. Böylece, yazı tura atıldığında olası sonuçların örnek uzayı $S = \{Y, T\}$ olarak yazılabilir; burada Y ve T , sırasıyla yazı ve turaya karşılık gelmektedir.

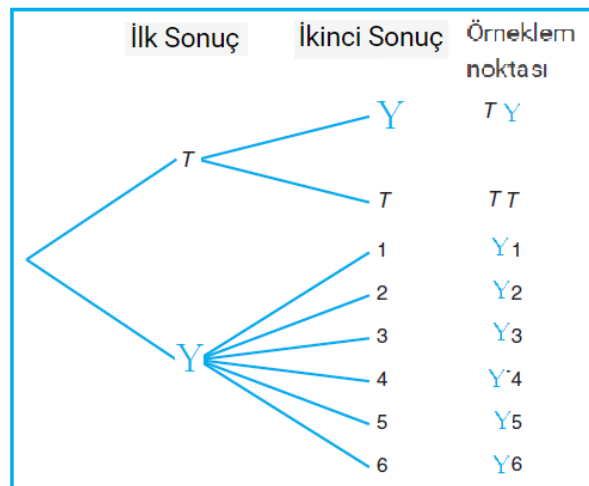
Örnek 2.1: Bir zar atma deneyini düşünün.

Cevap 2.1: Üst yüzde görünen sayı ile ilgileniyorsanız, örnek uzay $S_1 = \{1,2,3,4,5,6\}$ 'dir. Yalnızca sayının çift mi yoksa tek mi olduğuyla ilgileniyorsak, örnek uzay basitçe $S_2 = \{tek, çift\}$ olur.

Örnek 2.1, bir deneyin sonuçlarını açıklamak için birden fazla örnek uzayın kullanılabileceğini göstermektedir. Bu durumda S_1 , S_2 'den daha fazla bilgi sağlar. S_1 'de hangi ögenin oluştuğunu bilirse, S_2 'de hangi sonucun oluştuğunu söyleyebiliriz; bununla birlikte, S_2 'de ne olduğuna dair bir bilgi, S_1 'de hangi ögenin oluştuğunu belirlemede çok az yardımcı olur. Genel olarak, deneyin sonuçlarıyla ilgili en fazla bilgiyi veren örnek uzayın kullanılması arzu edilir. Bazı deneylerde, örnek uzayın elemanlarını bir ağaç diyagramı aracılığıyla sistematik olarak listelemek yararlıdır.

Örnek 2.2: Bir yazı tura atmayı deneyi gerçekleştirilmektedir. Para tura gelirse ikinci kez havaya atılmaktadır. Yazı gelirse, zar bir kez atılmaktadır. Örneklem uzayını oluşturunuz.

Cevap 2.2: Örnek uzayın en çok bilgiyi sağlayan öğelerini listelemek için Şekil 2.1'deki ağaç diyagramı oluşturulur.

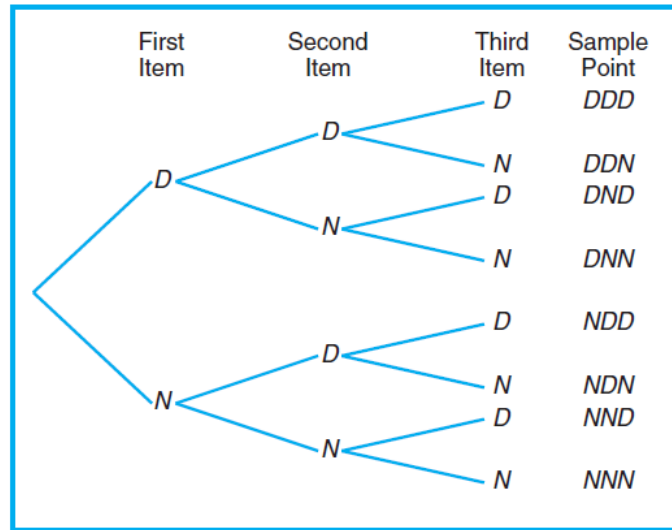


Şekil 2.1: Örnek 1.2 için ağaç diyagramı.

Ağacın dalları boyunca uzanan çeşitli yollar, farklı örnek noktalarını vermektedir. Sol üst daldan başlayarak ve ilk yol boyunca sağa doğru hareket ederek, madeni paranın art arda iki atışında tura gelme olasılığını gösteren TY örnek noktasını elde ederiz. Benzer şekilde, Y3 örnekleme noktası, madeni paranın zar atıldığında yazı ve ardından 3 gelme olasılığını göstermektedir.

Örnek 2.3: Bir üretim sürecinden rastgele üç öge seçildiğini varsayalım. Her öge incelenir ve hatalı ise D veya hatasız ise N olarak sınıflandırılmaktadır. Örnek uzayın en fazla bilgiyi sağlayan ögelerini listeleyiniz.

Cevap 2.3: Bunun için Şekil 2.2'deki ağaç diyagramını oluştururuz. Bu durumda, ağacın dalları boyunca uzanan çeşitli yollar farklı örnek noktaları vermektedir. İlk yoldan başlayarak, incelenen üç ögenin de kusurlu olma olasılığını gösteren DDD örnekleme noktasını elde ederiz.



Şekil 2.2: Örnek 1.3 için ağaç diyagramı.

Diğer yollar boyunca ilerledikçe, örnek uzayın

$S = \{DDD, DDN, DND, DNN, NDD, NDN, NND, NNN\}$ olduğunu görürüz.

2.2 OLAYLAR

Herhangi bir deney için, örnek uzayda belirli bir ögenin ortaya çıkmasından ziyade belirli olayların meydana gelmesiyle ilgilenebiliriz. Örneğin, bir zar atıldığında sonucun 3'e bölünebilir olduğu A olayıyla ilgilenebiliriz. Bu, sonuç, Örnek 2.1'deki S_1 örnek uzayının $A = \{3,6\}$ alt kümesinin bir ögesi ise gerçekleşmektedir.

Daha ileri bir örnek olarak, Örnek 2.3'teki kusur sayısının 1'den büyük olduğu B olayıyla ilgilenebiliriz. Sonuç, S örnek uzayının $B = \{DDN, DND, NDD, DDD\}$ altkümesinin bir elemanı ise bu gerçekleşir.

Her olaya, örnek uzayın bir alt kümesini oluşturan örnek noktaların bir koleksiyonunu atarız. Bu alt küme, olayın doğru olduğu tüm öğeleri temsil eder.

Tanım 2.2: Bir olay, bir örnek uzayın alt kümesidir.

Tanım 2.3: Bir A olayının S 'ye göre tümleyeni, S 'nin A 'da olmayan tüm örneklemelerinin oluşturduğu altkümesidir ve A' ile gösterilmektedir.

Örnek 2.4: $S = \{\text{kitap, cep telefonu, mp3, kâğıt, kırtasiye, dizüstü bilgisayar}\}$ örneklem uzayını ele alalım.

Cevap 2.4: $A = \{\text{kitap, kırtasiye, dizüstü bilgisayar, kâğıt}\}$ olsun. O halde A 'nın tümleyeni $A' = \{\text{cep telefonu, mp3}\}$ olur.

Şimdi yeni olayların oluşmasına neden olacak olaylarla belirli operasyonları ele alalım. Bu yeni olaylar, verilen olaylarla aynı örnek uzayın altkümeleri olacaktır. A ve B 'nin bir deneyle ilişkili iki olay olduğunu varsayalım. Başka bir deyişle, A ve B , aynı örnek uzay S 'nin altkümeleridir. Örneğin, bir zarın atılmasında, A 'nın çift bir sayının olması olayı ve B 'nin 3'ten büyük bir sayının göstermesi olayını kabul edebiliriz. O zaman $A = \{2, 4, 6\}$ ve $B = \{4, 5, 6\}$ altkümeleri, aynı $S = \{1, 2, 3, 4, 5, 6\}$ örnek uzayının altkümeleridir. Sonuç $\{4, 6\}$ alt kümesinin bir öğesiyse, bu yalnızca A ve B 'nin kesişimidir.

Tanım 2.4: $A \cap B$ sembolü ile gösterilen A ve B olaylarının kesişimi, A ve B için ortak olan tüm öğeleri içeren olaydır.

Tanım 2.5: $A \cap B = \emptyset$ ise, yani A ve B 'nin hiçbir ortak elemanı yoksa, A ve B olaylarına ayrık olaylar denir.

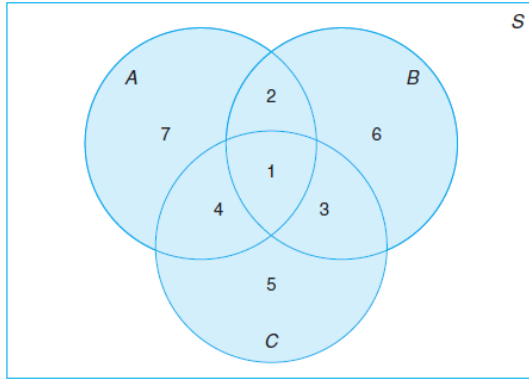
Tanım 2.6: $A \cup B$ sembolü ile gösterilen A ve B olaylarının birleşimi, A veya B 'ye veya her ikisine ait tüm öğeleri içeren olaydır.

Örnek 2.5: $A = \{a, b, c\}$ ve $B = \{b, c, d, e\}$ olsun; o zaman $A \cup B = \{a, b, c, d, e\}$.

Örnek 2.6: $M = \{x \mid 3 < x < 9\}$ ve $N = \{y \mid 5 < y < 12\}$, ise $M \cup N = \{z \mid 3 < z < 12\}$.

Örnek 2.7: Olaylar ve karşılık gelen örnek uzay arasındaki ilişki, Venn diyagramları aracılığıyla grafiksel olarak gösterilebilir. Bir Venn diyagramında, örnek uzayın bir dikdörtgen olmasına izin veririz ve olayları dikdörtgenin içine çizilen dairelerle temsil ederiz.

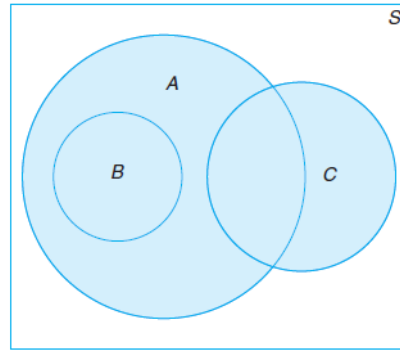
Böylece, Şekil 2.3'te görüyoruz ki



$$\begin{aligned} A \cap B &= 1, 2, \\ B \cap C &= 1, 3, \\ A \cup C &= 1, 2, 3, 4, 5, 7, \\ B' \cap A &= 4, 7, \\ A \cap B \cap C &= 1, \\ (A \cup B) \cap C' &= 2, 6, 7 \end{aligned}$$

Şekil 2.3: Çeşitli bölgeler tarafından temsil edilen olaylar.

Örnek 2.8: Şekil 1.4'te A, B ve C olaylarının hepsinin S örnek uzayının altkümeleri olduğunu görüyoruz.



Şekil 2.4: S örneklem uzayının olayları.

B olayının A olayının bir altkümesi olduğu da açıktır; $B \cap C$ olayının elemanı yoktur ve bu nedenle B ve C birbirini dışlar yani ayrıktır; $A \cap C$ olayının en az bir elemanı vardır; ve

$A \cup B = A$ olayıdır. Bu nedenle, Şekil 2.4, 52 oyun kağıdından oluşan sıradan bir desteden rastgele bir kart seçtiğimiz ve aşağıdaki olayların gerçekleşip gerçekleşmediğini gözlemlediğimiz bir durumu tasvir edebilir:

A: Kart kırmızı gelmesi,

B: Kartın, karo kızı, valesi veya papazı olması,

C: Kartın as olması.

Cevap 2.8: Venn diyagramları aracılığıyla kolayca doğrulanabilecek birkaç sonuç aşağıdaki gibidir:

- | | |
|---------------------------|---------------------------------|
| 1. $A \cap \phi = \phi$. | 6. $\phi' = S$. |
| 2. $A \cup \phi = A$. | 7. $(A')' = A$. |
| 3. $A \cap A' = \phi$. | 8. $(A \cap B)' = A' \cup B'$. |
| 4. $A \cup A' = S$. | 9. $(A \cup B)' = A' \cap B'$. |
| 5. $S' = \phi$. | |

2.3 ÖRNEKLEMLERİ SAYMA

İstatistikçinin dikkate alması ve değerlendirmeye çalışması gereken sorunlardan biri, bir deney yapıldığında belirli olayların meydana gelmesiyle ilgili şans unsurudur. Bu problemler, Bölüm 2.4'te tanıtılacak bir konu olan olasılık alanına aittir. Pek çok durumda, bir olasılık problemini, her elemanı gerçekten listelemeden, örnek uzaydaki nokta sayısını sayarak çözebileceğiz. Genellikle çarpma kuralı olarak adlandırılan temel sayma ilkesi Kural 2.1'de belirtilmiştir.

Kural 2.1: Bir işlem n_1 farklı yolla yapılabilir ve bu yolların her biri için n_2 farklı yolla ikinci bir işlem yapılabilir, o zaman iki işlem birlikte $n_1 \cdot n_2$ yolla yapılabilir.

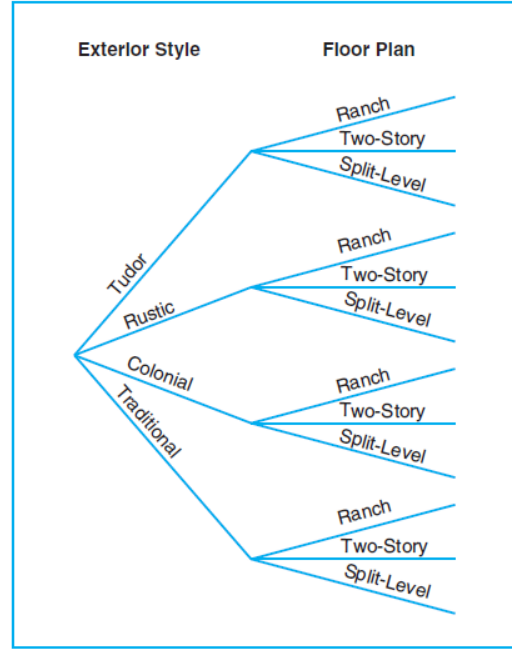
Örnek 2.9: Bir çift zar bir kez atıldığında, örnek uzayda kaç tane örnek nokta vardır?

Cevap 2.9: İlk zar $n_1 = 6$ yoldan herhangi birinde yüzü yukarı gelecek şekilde gelebilir. Bu 6 yolun her biri için, ikinci zar da $n_2 = 6$ şekilde yüzü yukarı gelebilir. Bu nedenle, bir çift zar $n_1 n_2 = (6)(6) = 36$ olası yoldan gelebilir.

Örnek 2.10: Bir mimar, potansiyel ev alıcılarına çiftlik (Ranch), iki katlı (Two-Story,) ve iki seviyeli (Split-Level) kat planlarında Tudor (Tudor), rüstik (Rustic), kolonyal (Colonial) ve geleneksel (Traditional) dış tasarım seçenekleri sunar. Bir alıcı bu evlerden birini kaç farklı şekilde sipariş edebilir?

Cevap 2.10: $n_1 = 4$ ve $n_2 = 3$ olduğundan, bir alıcı $n_1 n_2 = (4)(3) = 12$ olası ev arasından seçim yapmalıdır.

Önceki iki örneğe verilen yanıtlar, ağaç diyagramları oluşturularak ve dallar boyunca çeşitli yollar sayılarak doğrulanabilir. Örneğin, Örnek 1.10'da, farklı dış stillere karşılık gelen $n_1 = 4$ dal olacaktır ve sonra bu 4 dalın her birinden farklı kat planlarını temsil edecek şekilde uzanan $n_2 = 3$ dal olacaktır. Bu ağaç diyagramı, Şekil 2.5'te gösterildiği gibi, dallar boyunca yolların verdiği $n_1 n_2 = (4)(3) = 12$ ev seçeneğini verir.



Şekil 2.5: Örnek 1.10 için ağaç diyagramı.

Örnek 2.11: 22 üyeli bir kulübün bir başkan ve bir sayman seçmesi gerekiyorsa, bu ikisi kaç farklı şekilde seçilebilir?

Cevap 2.11: Sandalye pozisyonu için toplam 22 olasılık vardır. Bu 22 olasılığın her biri için, saymanı seçmek için 21 olasılık vardır. Çarpma kuralını kullanarak $n_1 \times n_2 = 22 \times 21 = 462$ farklı şekilde elde ederiz.

Kural 2.2: Bir işlem n_1 yolla yapılabilirse ve bunların her biri için n_2 yolla ikinci bir işlem yapılabilirse ve ilk ikisinin her biri için n_3 yolla üçüncü bir işlem yapılabilirse, o zaman sıra k işlem sayısı $n_1 n_2 \cdots n_k$ şekilde gerçekleştirilebilir.

Örnek 2.12: 0, 1, 2, 5, 6 ve 9 rakamlarından her rakam sadece bir kez kullanılabilirse kaç tane 4 basamaklı çift sayı oluşturulabilir?

Cevap 2.12: Sayının çift olması gerektiğinden, birler basamağı çift olmalı yani 0, 2 ve 6. Bununla birlikte, dört basamaklı bir sayı için binler basamağı 0 olamaz. Bu nedenle, birler basamağına gelecek durumu 0 veya 0 değil olarak iki kısımda ele almalıyız.

Birler basamağı 0 ise (yani, $n_1 = 1$), binler basamağına $n_4 = 5$ farklı rakam, yüzler basamağına $n_3 = 4$ ve onlar basamağına $n_2 = 3$ farklı rakam gelebilir. Dolayısıyla bu durumda toplam $n_1 n_2 n_3 n_4 = (1)(3)(4)(5) = 60$ tane dört basamaklı çift sayı elde ederiz.

Öte yandan, birler basamağı 0 değilse (yani 2 ve 6 ise, $n_1 = 2$), binler basamağına $n_4 = 4$ farklı rakam, yüzler basamağına $n_3 = 4$ ve onlar basamağına $n_2 = 3$ seçeneğimiz vardır. Bu durumda toplam $n_1 n_2 n_3 n_4 = (2)(3)(4)(4) = 96$ tane dört basamaklı çift sayı vardır.

Dolayısıyla dört basamaklı çift sayıların toplamı $60 + 96 = 156$ olarak hesaplanabilir.

Tanım 2.7: Bir nesne grubunun bir kısmının veya hepsinin herhangi bir sıralanışı (dizilimi) permütasyon olarak adlandırılır.

Sıklıkla, bir grup nesnenin tüm olası düzenlerini veya düzenlemelerini öğeler olarak içeren bir örnek uzayla ilgileniriz. Örneğin, 6 kişinin bir masa etrafında oturması için kaç farklı düzenlemenin mümkün olduğunu bilmek isteyebiliriz veya toplam 20 bileten 2 çekiliş için kaç farklı sıralamanın mümkün olduğunu sorabiliriz. Farklı düzenlemelere permütasyon denir.

Tanım 2.8: Negatif olmayan herhangi bir n tamsayısı için, $n!$, “ n faktöriyel” olarak adlandırılır ve şu şekilde tanımlanır:

$$n! = n(n-1) \cdots (2)(1),$$

özel durum ise $0! = 1$.

TEOREM 2.1: n nesnenin permütasyon sayısı $n!$ 'dir.

TEOREM 2.2: n farklı nesneden r tanesini seçen permütasyonların toplam sayısı

$${}_nP_r = \frac{n!}{(n-r)!} \text{ 'dir.}$$

Örnek 2.13: Bir yıl içinde, bir istatistik bölümündeki 25 lisansüstü öğrenciden oluşan bir sınıfa üç ödül (araştırma, öğretim ve hizmet) verilecektir. Her öğrenci en fazla bir ödül alabilirse, kaç olası seçim vardır?

Cevap 2.13: Ödüller ayırt edilebilir olduğundan, bir permütasyon problemidir. Örnek noktalarının toplam sayısı

$${}_{25}P_3 = \frac{n!}{(n-r)!} = \frac{25!}{(25-3)!} = \frac{25!}{22!} = (25)(24)(23) = 13800.$$

Örnek 2.14: 50 kişilik öğrenci kulübünden bir başkan ve bir sayman seçilecektir. Buna göre aşağıdaki şartlara göre kaç farklı yönetim seçimi mümkündür?

- a.) Herhangi bir kısıtlama yoktur;
- b.) A, ancak başkan olduğu durumda yönetimde görev yapacak;
- c.) B ve C birlikte görev yapacak veya hiç hizmet vermeyecek;
- d.) D ve E birlikte görev yapmayacak.

Cevap 2.14:

- a.) Herhangi bir kısıtlama olmaksızın, yönetim seçimlerinin toplam sayısı,

$${}_{50}P_2 = \frac{n!}{(n-r)!} = \frac{50!}{48!} = (50)(49) = 2450.$$

- b.) A sadece başkan olduğu zaman yönetimde görev yapacağı için burada iki durumumuz var: (i) A'nın başkan olarak seçilmesi, bu da saymanın pozisyonu için geri kalan 49 olası sonuç verir ve $(1)(49)=49$. (ii) Yönetim, A'sız kalan 49 kişi arasından seçilir. ${}_{49}P_2=(49)(48) = 2352$ seçenek sayısına sahiptir. Bu nedenle, toplam seçenek sayısı $49+2352 = 2401$ 'dir.
- c.) B ve C birlikte hizmet verdiğinde seçim sayısı 2'dir. Hem B hem de C seçilmediğinde seçim sayısı ${}_{48}P_2=(48)(47)=2256$ 'dır. Dolayısıyla bu durumda toplam seçenek sayısı $2+2256 = 2258$ 'dir.

d.) D'nin memur olarak görev yaptığı ancak E'nin görev yapmadığı seçim sayısı $(2)(48) = 96$ 'dır; burada 2, D'nin görev alabileceği pozisyon sayısıdır ve 48 ise E hariç diğerlerinin görevde kalan kişilerden seçim sayısıdır. E'nin yönetimde görev yaptığı ancak D'nin görev yapmadığı seçim sayısı da $(2)(48) = 96$ 'dır. Hem D hem de E seçilmediğinde seçim sayısı ${}_{48}P_2 = (48)(47) = 2256$ 'dır. Bu nedenle, toplam sayı seçenek sayısı $(2)(96) + 2256 = 2448$ 'dir.

***Bu problemin başka bir kısa çözümü daha vardır: D ve E birlikte ancak 2 şekilde hizmet edebildiğinden, cevap 2450 (kısıt olmadan toplam seçim) $- 2 = 2448$ 'dir.

TEOREM 2.3: Bir daire içinde düzenlenmiş n nesnenin permütasyon sayısı $(n-1)!$ 'dir.

Nesnelerin bir daire içinde düzenlenmesiyle oluşan permütasyonlara dairesel permütasyonlar denir. Saat yönünde ilerlerken, iki düzenlemedeki karşılık gelen nesnelerin önünde veya arkasında farklı bir nesne olmadıkça, iki dairesel permütasyon farklı kabul edilmez. Örneğin, 4 kişi birliğin oynuyorsa, hepsi bir konumu saat yönünde hareket ettirirse yeni bir permütasyonumuz olmaz. Bir kişiyi sabit bir pozisyonda düşünüp diğer üçünü $3!$ 'te düzenleyerek yollar, birliğin oyunu için 6 farklı düzenleme olduğunu görüyoruz.

TEOREM 2.4: n permütasyon içerisinde n_1 tanesi birbirine benzeyen bir tür, n_2 tanesi birbirine benzeyen ikinci bir tür, n_k tanesi birbirine benzeyen k . başka bir tür olmak üzere, permütasyonlarının toplam sayısı

$$\frac{n!}{(n_1)! (n_2)! \dots (n_k)!}$$

Örnek 2.15: Kolej futbolu antrenman seansında, savunma koordinatörünün arka arkaya 10 oyuncuya sahip olması gerekir. Bu 10 oyuncu içerisinde 1 tane 1. sınıf, 2 tane 2. sınıf, 4 tane genç sınıf ve 3 tane son sınıf öğrencisi bulunmaktadır. Sadece sınıf seviyeleri ayırt edilecekse, arka arkaya kaç farklı şekilde sıralanabilirler?

Cevap 2.15: Doğrudan Teorem 2.4'ü kullanarak, toplam düzenleme sayısı

$$\frac{n!}{(n_1)! (n_2)! \dots (n_k)!} = \frac{10!}{(1)! (2)! (4)! (3)!} = 12600.$$

TEOREM 2.5: Bir n nesne kümesini, birinci hücrede n_1 öge, ikinci hücrede n_2 öge vb. olacak şekilde r hücreye bölme yollarının sayısı,

$$\binom{n}{n_1, n_2, \dots, n_r} = \frac{n!}{(n_1)! (n_2)! \dots (n_r)!}$$

Burada $n_1 + n_2 + \dots + n_r = n$

Çoğu zaman, bir n nesne kümesini hücre adı verilen r alt kümeye bölmenin yollarının sayısı ile ilgileniriz. r alt kümelerinin olası her çiftinin kesişimi boş küme $\varnothing$ ise ve tüm alt kümelerin birleşimi orijinal kümeyi veriyorsa, bir bölme elde edilmiştir. Bir hücre içindeki elemanların sırasının hiçbir önemi yoktur. Örneğin $\{a, e, i, o, u\}$ kümesini ele alalım. Birinci hücrenin 4 ögeyi ve ikinci hücrenin 1 ögeyi içerdiği olası bölümler

$\{(a, e, i, o), (u)\}, \{(a, i, o, u), (e)\}, \{(e, i, o, u), (a)\}, \{(a, e, o, u), (i)\}, \{(a, e, i, u), (o)\}$ olmaktadır.

4 ögeden oluşan bir kümeyi, ilk hücrede 4 öge ve ikincide 1 öge içeren iki alt kümeye veya hücrelere bölmenin 5 yolu olduğunu görüyoruz. Teorem 2.5'i kullanarak bu sonucu benzer şekilde elde ederiz:

$$\binom{5}{4, 1} = \frac{5!}{(4)! (1)!} = 5$$

Örnek 2.16: Bir konferansta 7 lisansüstü öğrenci 1 tane üç kişilik ve 2 tane çift kişilik otel odasına kaç farklı şekilde yerleştirilebilir?

Cevap 2.16: Olası bölümlerin toplam sayısı

$$\binom{7}{3, 2, 2} = \frac{7!}{(3)! (2)! (2)!} = 210$$

TEOREM 2.6: n farklı nesneden r tanesini seçen kombinasyonların sayısı,

$$\binom{n}{r} = \frac{n!}{r! (n - r)!}$$

Pek çok problemde, sıra gözetmeksizin n nesneden r nesneyi seçme yollarının sayısı ile ilgileniyoruz. Bu seçimlere kombinasyon denir. Kombinasyon aslında iki hücreli bir bölümdür, bir hücre seçilen r nesneyi içerir ve diğer hücre kalan (n-r) nesneleri içerir.

Örnek 2.17: Küçük bir çocuk, annesinden 10 macera oyunu ve 5 spor oyunundan oluşan koleksiyonundan 5 tanesini vermesini istemektedir. Annesi kaç değişik şekilde 3 macera ve 2 spor oyunu verebilir?

Cevap 2.17: 10 macera oyunundan 3 macera oyunu seçmenin yol sayısı

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} = \binom{10}{3} = \frac{10!}{3!(10-3)!} = 120$$

5 spor oyunundan 2 spor oyunu seçmenin yol sayısı

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} = \binom{5}{2} = \frac{5!}{2!(5-2)!} = 10$$

$n_1 = 120$ ve $n_2 = 10$ ile çarpma kuralını (Kural 2.1) kullanarak, $(120)(10) = 1200$ değişik şekilde yolumuz olur.

Örnek 2.18: STATISTICS kelimesindeki harflerden kaç farklı harf düzenlemesi yapılabilir?

Cevap 2.18: Teorem 2.6 tartışmasındaki argümanın aynısını kullanarak, bu örnekte Teorem 2.5'i şu şekilde uygulayabiliriz:

$$\binom{10}{3, 3, 2, 1, 1} = \frac{10!}{(3)!(3)!(2)!(1)!(1)!} = 50400$$

Burada toplam 10 harf var, her biri 3 kez görünen 2 harf (S, T), 2 kez görünen I harfi ve 1 kez görünen A ve C harfleri. Öte yandan, bu sonuç doğrudan Teorem 2.4 kullanılarak elde edilebilir.

OLASILIK VE İSTATİSTİK

2. HAFTA

OLASILIK

2.1 ÖRNEKLEM UZAYI

İstatistik çalışmasında, temel olarak planlı bir çalışmada veya bilimsel araştırmada ortaya çıkan tesadüfi sonuçların sunumu ve yorumlanmasıyla ilgilenilmektedir. Örneğin, bir trafik ışığının yerleştirilmesini gerekçelendirmeyi umarak, herhangi bir X kavşağında meydana gelen aylık kaza sayısını kaydedebiliriz; bir montaj hattından çıkan ürünleri "kusurlu" veya "kusursuz" olarak sınıflandırabiliriz veya bir asidin konsantrasyonu değiştiğinde kimyasal bir reaksiyonda salınan gazın hacmiyle ilgilenebiliriz. Bu nedenle, istatistikçi genellikle sayımları veya ölçümleri temsil eden sayısal verilerle veya bazı kriterlere göre sınıflandırılabilen kategorik verilerle uğraşmaktadır.

İster sayısal ister kategorik olsun, herhangi bir bilgi kaydına gözlem olarak atıfta bulunacağız. Böylece, geçen yıl boyunca Ocak'tan Nisan'a kadar her ay X kavşağında meydana gelen kazaların sayısını temsil eden 2, 0, 1 ve 2 sayıları bir dizi gözlem oluşturmaktadır. Benzer şekilde, beş öge incelendiğinde kusurlu veya hatasız bulunan öğeleri temsil eden N , D , N , N ve D kategorik verileri gözlem olarak kaydedilir.

İstatistikçiler, bir veri kümesi oluşturan herhangi bir işlemi tanımlamak için deney kelimesini kullanırlar. İstatistiksel bir deneyin basit bir örneği, yazı tura atmaktır. Bu deneyde, yalnızca iki olası sonuç vardır, tura veya yazı. Başka bir deney, bir füzenin fırlatılması ve belirli zamanlarda hızının gözlemlenmesi olabilir. Seçmenlerin yeni bir satış vergisiyle ilgili görüşleri de bir deney gözlemi olarak değerlendirilebilir. Deneyi birkaç kez tekrarlayarak elde edilen gözlemlerle özellikle ilgilenilmektedir. Çoğu durumda, sonuçlar şansa bağlıdır ve bu nedenle kesin olarak tahmin edilemez. Bir kimyager aynı koşullar altında birkaç kez bir analiz yaparsa, deneysel prosedürde bir şans unsuru olduğunu gösteren farklı ölçümler elde edecektir. Bir yazı tura tekrar tekrar atıldığında bile, belirli bir atışın tura geleceğinden emin olamayız. Ancak, her atış için tüm olasılıklar bilinmektedir.

Tanım 2.1: İstatistiksel bir deneyin tüm olası sonuçlarının oluşturduğu kümeye “örneklem uzay” denir ve S sembolü ile gösterilir.

Bir örneklem uzayındaki her sonuç, örneklem uzayının bir elemanı veya üyesi ya da sadece bir örneklem noktası olarak adlandırılır. Örnek uzayın sonlu sayıda elemanı varsa, elemanları virgülle ayırarak ve parantez içinde listeleyebiliriz. Böylece, yazı tura atıldığında olası sonuçların örnek uzayı $S = \{Y, T\}$ olarak yazılabilir; burada Y ve T , sırasıyla yazı ve turaya karşılık gelmektedir.

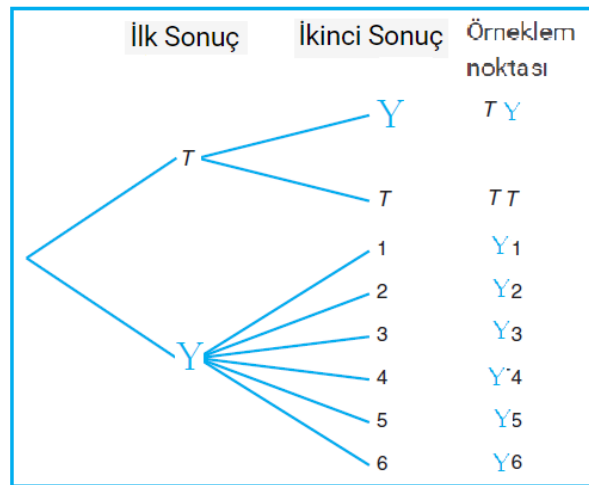
Örnek 2.1: Bir zar atma deneyini düşünün.

Cevap 2.1: Üst yüzde görünen sayı ile ilgileniyorsanız, örnek uzay $S_1 = \{1,2,3,4,5,6\}$ 'dir. Yalnızca sayının çift mi yoksa tek mi olduğuyla ilgileniyorsak, örnek uzay basitçe $S_2 = \{tek, çift\}$ olur.

Örnek 2.1, bir deneyin sonuçlarını açıklamak için birden fazla örnek uzayın kullanılabileceğini göstermektedir. Bu durumda S_1 , S_2 'den daha fazla bilgi sağlar. S_1 'de hangi ögenin oluştuğunu bilirse, S_2 'de hangi sonucun oluştuğunu söyleyebiliriz; bununla birlikte, S_2 'de ne olduğuna dair bir bilgi, S_1 'de hangi ögenin oluştuğunu belirlemede çok az yardımcı olur. Genel olarak, deneyin sonuçlarıyla ilgili en fazla bilgiyi veren örnek uzayın kullanılması arzu edilir. Bazı deneylerde, örnek uzayın elemanlarını bir ağaç diyagramı aracılığıyla sistematik olarak listelemek yararlıdır.

Örnek 2.2: Bir yazı tura atmayı deneyi gerçekleştirilmektedir. Para tura gelirse ikinci kez havaya atılmaktadır. Yazı gelirse, zar bir kez atılmaktadır. Örneklem uzayını oluşturunuz.

Cevap 2.2: Örnek uzayın en çok bilgiyi sağlayan öğelerini listelemek için Şekil 2.1'deki ağaç diyagramı oluşturulur.

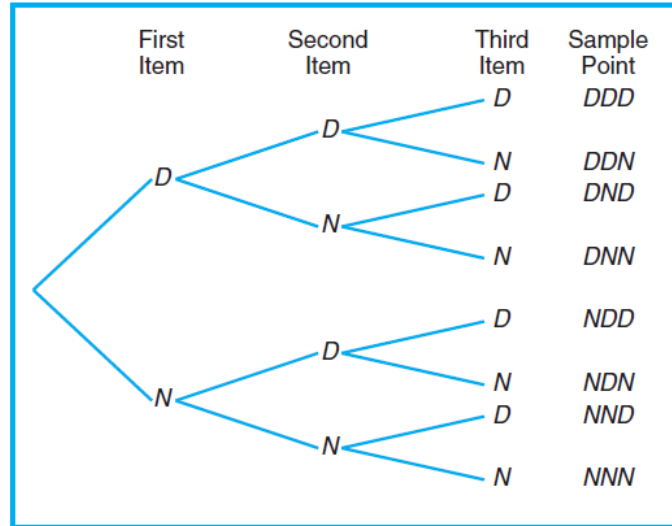


Şekil 2.1: Örnek 1.2 için ağaç diyagramı.

Ağacın dalları boyunca uzanan çeşitli yollar, farklı örnek noktalarını vermektedir. Sol üst daldan başlayarak ve ilk yol boyunca sağa doğru hareket ederek, madeni paranın art arda iki atışında tura gelme olasılığını gösteren TY örnek noktasını elde ederiz. Benzer şekilde, Y3 örnekleme noktası, madeni paranın zar atıldığında yazı ve ardından 3 gelme olasılığını göstermektedir.

Örnek 2.3: Bir üretim sürecinden rastgele üç öge seçildiğini varsayalım. Her öge incelenir ve hatalı ise D veya hatasız ise N olarak sınıflandırılmaktadır. Örnek uzayın en fazla bilgiyi sağlayan ögelerini listeleyiniz.

Cevap 2.3: Bunun için Şekil 2.2'deki ağaç diyagramını oluştururuz. Bu durumda, ağacın dalları boyunca uzanan çeşitli yollar farklı örnek noktaları vermektedir. İlk yoldan başlayarak, incelenen üç ögenin de kusurlu olma olasılığını gösteren DDD örnekleme noktasını elde ederiz.



Şekil 2.2: Örnek 1.3 için ağaç diyagramı.

Diğer yollar boyunca ilerledikçe, örnek uzayın

$S = \{DDD, DDN, DND, DNN, NDD, NDN, NND, NNN\}$ olduğunu görürüz.

2.2 OLAYLAR

Herhangi bir deney için, örnek uzayda belirli bir ögenin ortaya çıkmasından ziyade belirli olayların meydana gelmesiyle ilgilenebiliriz. Örneğin, bir zar atıldığında sonucun 3'e bölünebilir olduğu A olayıyla ilgilenebiliriz. Bu, sonuç, Örnek 2.1'deki S_1 örnek uzayının $A = \{3,6\}$ alt kümesinin bir ögesiye gerçekleşmektedir.

Daha ileri bir örnek olarak, Örnek 2.3'teki kusur sayısının 1'den büyük olduğu B olayıyla ilgilenebiliriz. Sonuç, S örnek uzayının $B = \{DDN, DND, NDD, DDD\}$ altkümesinin bir elemanı ise bu gerçekleşir.

Her olaya, örnek uzayın bir alt kümesini oluşturan örnek noktaların bir koleksiyonunu atarız. Bu alt küme, olayın doğru olduğu tüm öğeleri temsil eder.

Tanım 2.2: Bir olay, bir örnek uzayın alt kümesidir.

Tanım 2.3: Bir A olayının S'ye göre tümleyeni, S'nin A'da olmayan tüm örneklemelerinin oluşturduğu altkümesidir ve A' ile gösterilmektedir.

Örnek 2.4: $S = \{\text{kitap, cep telefonu, mp3, kağıt, kırtasiye, dizüstü bilgisayar}\}$ örneklem uzayını ele alalım.

Cevap 2.4: $A = \{\text{kitap, kırtasiye, dizüstü bilgisayar, kağıt}\}$ olsun. O halde A'nın tümleyeni $A' = \{\text{cep telefonu, mp3}\}$ olur.

Şimdi yeni olayların oluşmasına neden olacak olaylarla belirli operasyonları ele alalım. Bu yeni olaylar, verilen olaylarla aynı örnek uzayın altkümeleri olacaktır. A ve B'nin bir deneyle ilişkili iki olay olduğunu varsayalım. Başka bir deyişle, A ve B, aynı örnek uzay S'nin altkümeleridir. Örneğin, bir zarın atılmasında, A'nın çift bir sayının olması olayı ve B'nin 3'ten büyük bir sayının göstermesi olayını kabul edebiliriz. O zaman $A = \{2, 4, 6\}$ ve $B = \{4, 5, 6\}$ altkümeleri, aynı $S = \{1, 2, 3, 4, 5, 6\}$ örnek uzayının altkümeleridir. Sonuç $\{4, 6\}$ alt kümesinin bir ögesiye, bu yalnızca A ve B'nin kesişimidir.

Tanım 2.4: $A \cap B$ sembolü ile gösterilen A ve B olaylarının kesişimi, A ve B için ortak olan tüm öğeleri içeren olaydır.

Tanım 2.5: $A \cap B = \emptyset$ ise, yani A ve B'nin hiçbir ortak elemanı yoksa, A ve B olaylarına ayrık olaylar denir.

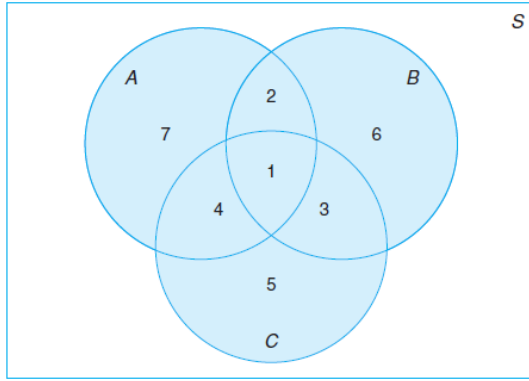
Tanım 2.6: $A \cup B$ sembolü ile gösterilen A ve B olaylarının birleşimi, A veya B'ye veya her ikisine ait tüm öğeleri içeren olaydır.

Örnek 2.5: $A = \{a, b, c\}$ ve $B = \{b, c, d, e\}$ olsun; o zaman $A \cup B = \{a, b, c, d, e\}$.

Örnek 2.6: $M = \{x \mid 3 < x < 9\}$ ve $N = \{y \mid 5 < y < 12\}$, ise $M \cup N = \{z \mid 3 < z < 12\}$.

Örnek 2.7: Olaylar ve karşılık gelen örnek uzay arasındaki ilişki, Venn diyagramları aracılığıyla grafiksel olarak gösterilebilir. Bir Venn diyagramında, örnek uzayın bir dikdörtgen olmasına izin veririz ve olayları dikdörtgenin içine çizilen dairelerle temsil ederiz.

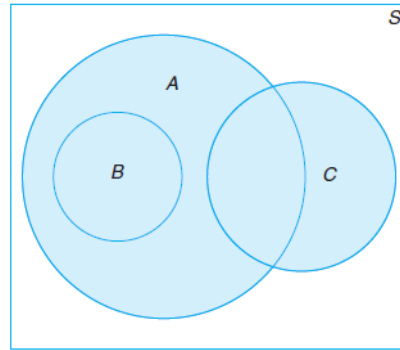
Böylece, Şekil 2.3'te görüyoruz ki



$$\begin{aligned} A \cap B &= 1, 2, \\ B \cap C &= 1, 3, \\ A \cup C &= 1, 2, 3, 4, 5, 7, \\ B' \cap A &= 4, 7, \\ A \cap B \cap C &= 1, \\ (A \cup B) \cap C' &= 2, 6, 7 \end{aligned}$$

Şekil 2.3: Çeşitli bölgeler tarafından temsil edilen olaylar.

Örnek 2.8: Şekil 1.4'te A, B ve C olaylarının hepsinin S örnek uzayının altkümeleri olduğunu görüyoruz.



Şekil 2.4: S örneklem uzayının olayları.

B olayının A olayının bir altkümesi olduğu da açıktır; $B \cap C$ olayının elemanı yoktur ve bu nedenle B ve C birbirini dışlar yani ayrıktır; $A \cap C$ olayının en az bir elemanı vardır; ve

$A \cup B = A$ olayıdır. Bu nedenle, Şekil 2.4, 52 oyun kağıdından oluşan sıradan bir desteden rastgele bir kart seçtiğimiz ve aşağıdaki olayların gerçekleşip gerçekleşmediğini gözlemlediğimiz bir durumu tasvir edebilir:

A: Kart kırmızı gelmesi,

B: Kartın, karo kızı, valesi veya papazı olması,

C: Kartın as olması.

Cevap 2.8: Venn diyagramları aracılığıyla kolayca doğrulanabilecek birkaç sonuç aşağıdaki gibidir:

- | | |
|---------------------------|---------------------------------|
| 1. $A \cap \phi = \phi$. | 6. $\phi' = S$. |
| 2. $A \cup \phi = A$. | 7. $(A')' = A$. |
| 3. $A \cap A' = \phi$. | 8. $(A \cap B)' = A' \cup B'$. |
| 4. $A \cup A' = S$. | 9. $(A \cup B)' = A' \cap B'$. |
| 5. $S' = \phi$. | |

2.3 ÖRNEKLEMLERİ SAYMA

İstatistikçinin dikkate alması ve değerlendirmeye çalışması gereken sorunlardan biri, bir deney yapıldığında belirli olayların meydana gelmesiyle ilgili şans unsurudur. Bu problemler, Bölüm 2.4'te tanıtılacak bir konu olan olasılık alanına aittir. Pek çok durumda, bir olasılık problemini, her elemanı gerçekten listelemeden, örnek uzaydaki nokta sayısını sayarak çözebileceğiz. Genellikle çarpma kuralı olarak adlandırılan temel sayma ilkesi Kural 2.1'de belirtilmiştir.

Kural 2.1: Bir işlem n_1 farklı yolla yapılabilir ve bu yolların her biri için n_2 farklı yolla ikinci bir işlem yapılabilir, o zaman iki işlem birlikte $n_1 \cdot n_2$ yolla yapılabilir.

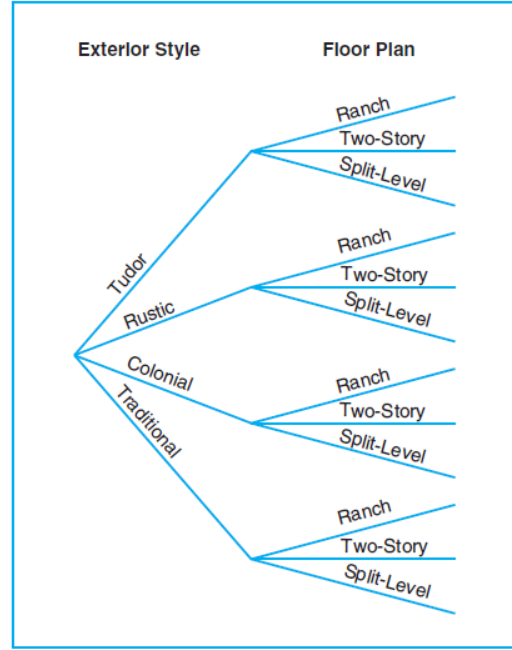
Örnek 2.9: Bir çift zar bir kez atıldığında, örnek uzayda kaç tane örnek nokta vardır?

Cevap 2.9: İlk zar $n_1 = 6$ yoldan herhangi birinde yüzü yukarı gelecek şekilde gelebilir. Bu 6 yolun her biri için, ikinci zar da $n_2 = 6$ şekilde yüzü yukarı gelebilir. Bu nedenle, bir çift zar $n_1 n_2 = (6)(6) = 36$ olası yoldan gelebilir.

Örnek 2.10: Bir mimar, potansiyel ev alıcılarına çiftlik (Ranch), iki katlı (Two-Story,) ve iki seviyeli (Split-Level) kat planlarında Tudor (Tudor), rüstik (Rustic), kolonyal (Colonial) ve geleneksel (Traditional) dış tasarım seçenekleri sunar. Bir alıcı bu evlerden birini kaç farklı şekilde sipariş edebilir?

Cevap 2.10: $n_1 = 4$ ve $n_2 = 3$ olduğundan, bir alıcı $n_1 n_2 = (4)(3) = 12$ olası ev arasından seçim yapmalıdır.

Önceki iki örneğe verilen yanıtlar, ağaç diyagramları oluşturularak ve dallar boyunca çeşitli yollar sayılarak doğrulanabilir. Örneğin, Örnek 1.10'da, farklı dış stillere karşılık gelen $n_1 = 4$ dal olacaktır ve sonra bu 4 dalın her birinden farklı kat planlarını temsil edecek şekilde uzanan $n_2 = 3$ dal olacaktır. Bu ağaç diyagramı, Şekil 2.5'te gösterildiği gibi, dallar boyunca yolların verdiği $n_1 n_2 = (4)(3) = 12$ ev seçeneğini verir.



Şekil 2.5: Örnek 1.10 için ağaç diyagramı.

Örnek 2.11: 22 üyeli bir kulübün bir başkan ve bir sayman seçmesi gerekiyorsa, bu ikisi kaç farklı şekilde seçilebilir?

Cevap 2.11: Sandalye pozisyonu için toplam 22 olasılık vardır. Bu 22 olasılığın her biri için, saymanı seçmek için 21 olasılık vardır. Çarpma kuralını kullanarak $n_1 \times n_2 = 22 \times 21 = 462$ farklı şekilde elde ederiz.

Kural 2.2: Bir işlem n_1 yolla yapılabilirse ve bunların her biri için n_2 yolla ikinci bir işlem yapılabilirse ve ilk ikisinin her biri için n_3 yolla üçüncü bir işlem yapılabilirse, o zaman sıra k işlem sayısı $n_1 n_2 \cdots n_k$ şekilde gerçekleştirilebilir.

Örnek 2.12: 0, 1, 2, 5, 6 ve 9 rakamlarından her rakam sadece bir kez kullanılabilirse kaç tane 4 basamaklı çift sayı oluşturulabilir?

Cevap 2.12: Sayının çift olması gerektiğinden, birler basamağı çift olmalı yani 0, 2 ve 6. Bununla birlikte, dört basamaklı bir sayı için binler basamağı 0 olamaz. Bu nedenle, birler basamağına gelecek durumu 0 veya 0 değil olarak iki kısımda ele almalıyız.

Birler basamağı 0 ise (yani, $n_1 = 1$), binler basamağına $n_4 = 5$ farklı rakam, yüzler basamağına $n_3 = 4$ ve onlar basamağına $n_2 = 3$ farklı rakam gelebilir. Dolayısıyla bu durumda toplam $n_1 n_2 n_3 n_4 = (1)(3)(4)(5) = 60$ tane dört basamaklı çift sayı elde ederiz.

Öte yandan, birler basamağı 0 değilse (yani 2 ve 6 ise, $n_1 = 2$), binler basamağına $n_4 = 4$ farklı rakam, yüzler basamağına $n_3 = 4$ ve onlar basamağına $n_2 = 3$ seçeneğimiz vardır. Bu durumda toplam $n_1 n_2 n_3 n_4 = (2)(3)(4)(4) = 96$ tane dört basamaklı çift sayı vardır.

Dolayısıyla dört basamaklı çift sayıların toplamı $60 + 96 = 156$ olarak hesaplanabilir.

Tanım 2.7: Bir nesne grubunun bir kısmının veya hepsinin herhangi bir sıralanışı (dizilimi) permütasyon olarak adlandırılır.

Sıklıkla, bir grup nesnenin tüm olası düzenlerini veya düzenlemelerini öğeler olarak içeren bir örnek uzayla ilgileniriz. Örneğin, 6 kişinin bir masa etrafında oturması için kaç farklı düzenlemenin mümkün olduğunu bilmek isteyebiliriz veya toplam 20 bileten 2 çekiliş için kaç farklı sıralamanın mümkün olduğunu sorabiliriz. Farklı düzenlemelere permütasyon denir. Yani, sıralamanın kaç farklı biçimde gerçekleşeceğini hesaplamada permütasyon kullanılmaktadır.

Tanım 2.8: Negatif olmayan herhangi bir n tamsayısı için, $n!$, “ n faktöriyel” olarak adlandırılır ve şu şekilde tanımlanır:

$$n! = n(n-1) \cdots (2)(1),$$

özel durum ise $0! = 1$.

TEOREM 2.1: n nesnenin permütasyon sayısı $n!$ 'dir.

TEOREM 2.2: n farklı nesneden r tanesini seçen permütasyonların toplam sayısı

$${}_nP_r = \frac{n!}{(n-r)!} \text{ 'dir.}$$

Örnek 2.13: Bir yıl içinde, bir istatistik bölümündeki 25 lisansüstü öğrenciden oluşan bir sınıfa üç ödül (araştırma, öğretim ve hizmet) verilecektir. Her öğrenci en fazla bir ödül alabilirse, kaç olası seçim vardır?

Cevap 2.13: Ödüller ayırt edilebilir olduğundan, bir permütasyon problemidir. Örnek noktalarının toplam sayısı

$${}_{25}P_3 = \frac{n!}{(n-r)!} = \frac{25!}{(25-3)!} = \frac{25!}{22!} = (25)(24)(23) = 13800.$$

Örnek 2.14: 50 kişilik öğrenci kulübünden bir başkan ve bir sayman seçilecektir. Buna göre aşağıdaki şartlara göre kaç farklı yönetim seçimi mümkündür?

- a.) Herhangi bir kısıtlama yoktur;
- b.) A, ancak başkan olduğu durumda yönetimde görev yapacak;
- c.) B ve C birlikte görev yapacak veya hiç hizmet vermeyecek;
- d.) D ve E birlikte görev yapmayacak.

Cevap 2.14:

- a.) Herhangi bir kısıtlama olmaksızın, yönetim seçimlerinin toplam sayısı,

$${}_{50}P_2 = \frac{n!}{(n-r)!} = \frac{50!}{48!} = (50)(49) = 2450.$$

- b.) A sadece başkan olduğu zaman yönetimde görev yapacağı için burada iki durumumuz var: (i) A'nın başkan olarak seçilmesi, bu da saymanın pozisyonu için geri kalan 49 olası sonuç verir ve $(1)(49)=49$. (ii) Yönetim, A'sız kalan 49 kişi arasından seçilir. ${}_{49}P_2=(49)(48) = 2352$ seçenek sayısına sahiptir. Bu nedenle, toplam seçenek sayısı $49+2352 = 2401$ 'dir.
- c.) B ve C birlikte hizmet verdiğinde seçim sayısı 2'dir. Hem B hem de C seçilmediğinde seçim sayısı ${}_{48}P_2 =(48)(47)=2256$ 'dır. Dolayısıyla bu durumda toplam seçenek sayısı $2+2256 = 2258$ 'dir.

d.) D'nin memur olarak görev yaptığı ancak E'nin görev yapmadığı seçim sayısı $(2)(48) = 96$ 'dır; burada 2, D'nin görev alabileceği pozisyon sayısıdır ve 48 ise E hariç diğerlerinin görevde kalan kişilerden seçim sayısıdır. E'nin yönetimde görev yaptığı ancak D'nin görev yapmadığı seçim sayısı da $(2)(48) = 96$ 'dır. Hem D hem de E seçilmediğinde seçim sayısı ${}_{48}P_2 = (48)(47) = 2256$ 'dır. Bu nedenle, toplam sayı seçenek sayısı $(2)(96) + 2256 = 2448$ 'dir.

***Bu problemin başka bir kısa çözümü daha vardır: D ve E birlikte ancak 2 şekilde hizmet edebildiğinden, cevap 2450 (kısıt olmadan toplam seçim) $- 2 = 2448$ 'dir.

TEOREM 2.3: Bir daire içinde düzenlenmiş n nesnenin permütasyon sayısı $(n-1)!$ 'dir.

Nesnelerin bir daire içinde düzenlenmesiyle oluşan permütasyonlara dairesel permütasyonlar denir. Saat yönünde ilerlerken, iki düzenlemedeki karşılık gelen nesnelerin önünde veya arkasında farklı bir nesne olmadıkça, iki dairesel permütasyon farklı kabul edilmez. Örneğin, 4 kişi briç oynuyorsa, hepsi bir konumu saat yönünde hareket ettirirse yeni bir permütasyonumuz olmaz. Bir kişiyi sabit bir pozisyonda düşünüp diğer üçünü $3!$ 'te düzenleyerek yollar, briç oyunu için 6 farklı düzenleme olduğunu görüyoruz.

TEOREM 2.4: n permütasyon içerisinde n_1 tanesi birbirine benzeyen bir tür, n_2 tanesi birbirine benzeyen ikinci bir tür, n_k tanesi birbirine benzeyen k . başka bir tür olmak üzere, permütasyonlarının toplam sayısı

$$\frac{n!}{(n_1)! (n_2)! \dots (n_k)!}$$

Örnek 2.15: Kolej futbolu antrenman seansında, savunma koordinatörünün arka arkaya 10 oyuncuya sahip olması gerekir. Bu 10 oyuncu içerisinde 1 tane 1. sınıf, 2 tane 2. sınıf, 4 tane genç sınıf ve 3 tane son sınıf öğrencisi bulunmaktadır. Sadece sınıf seviyeleri ayırt edilecekse, arka arkaya kaç farklı şekilde sıralanabilirler?

Cevap 2.15: Doğrudan Teorem 2.4'ü kullanarak, toplam düzenleme sayısı

$$\frac{n!}{(n_1)! (n_2)! \dots (n_k)!} = \frac{10!}{(1)! (2)! (4)! (3)!} = 12600.$$

TEOREM 2.5: Bir n nesne kümesini, birinci hücrede n_1 öge, ikinci hücrede n_2 öge vb. olacak şekilde r hücreye bölme yollarının sayısı,

$$\binom{n}{n_1, n_2, \dots, n_r} = \frac{n!}{(n_1)! (n_2)! \dots (n_r)!}$$

Burada $n_1 + n_2 + \dots + n_r = n$

Çoğu zaman, bir n nesne kümesini hücre adı verilen r alt kümeye bölmenin yollarının sayısıyla ilgileniriz. r alt kümelerinin olası her çiftinin kesişimi boş küme $\emptyset$ ise ve tüm alt kümelerin birleşimi orijinal kümeyi veriyorsa, bir bölme elde edilmiştir. Bir hücre içindeki elemanların sırasının hiçbir önemi yoktur. Örneğin $\{a, e, i, o, u\}$ kümesini ele alalım. Birinci hücrenin 4 ögeyi ve ikinci hücrenin 1 ögeyi içerdiği olası bölümler

$\{(a, e, i, o), (u)\}, \{(a, i, o, u), (e)\}, \{(e, i, o, u), (a)\}, \{(a, e, o, u), (i)\}, \{(a, e, i, u), (o)\}$ olmaktadır.

4 ögeden oluşan bir kümeyi, ilk hücrede 4 öge ve ikincide 1 öge içeren iki alt kümeye veya hücrelere bölmenin 5 yolu olduğunu görüyoruz. Teorem 2.5'i kullanarak bu sonucu benzer şekilde elde ederiz:

$$\binom{5}{4, 1} = \frac{5!}{(4)! (1)!} = 5$$

Örnek 2.16: Bir konferansta 7 lisansüstü öğrenci 1 tane üç kişilik ve 2 tane çift kişilik otel odasına kaç farklı şekilde yerleştirilebilir?

Cevap 2.16: Olası bölümlerin toplam sayısı

$$\binom{7}{3, 2, 2} = \frac{7!}{(3)! (2)! (2)!} = 210$$

TEOREM 2.6: n farklı nesneden r tanesini seçen kombinasyonların sayısı,

$$\binom{n}{r} = \frac{n!}{r! (n-r)!}$$

Pek çok problemde, sıra gözetmeksizin n nesneden r nesneyi seçme yollarının sayısıyla ilgileniyoruz. (Yani, n tane farklı nesneden r tanesinin kaç farklı biçimde seçilme durumu ile ilgilenilmektedir.) Bu seçimlere kombinasyon denir. Kombinasyon aslında iki hücreli bir bölümdür, bir hücre seçilen r nesneyi içerir ve diğer hücre kalan (n-r) nesneleri içerir. **Kısaca, seçim işleminin kaç farklı biçimde yapılacağını belirlemede kombinasyon kullanılır.**

Örnek 2.17: Küçük bir çocuk, annesinden 10 macera oyunu ve 5 spor oyunundan oluşan koleksiyonundan 5 tanesini vermesini istemektedir. Annesi kaç değişik şekilde 3 macera ve 2 spor oyunu verebilir?

Cevap 2.17: 10 macera oyunundan 3 macera oyunu seçmenin yol sayısı

$$\binom{n}{r} = \frac{n!}{r! (n-r)!} = \binom{10}{3} = \frac{10!}{3! (10-3)!} = 120$$

5 spor oyunundan 2 spor oyunu seçmenin yol sayısı

$$\binom{n}{r} = \frac{n!}{r! (n-r)!} = \binom{5}{2} = \frac{5!}{2! (5-2)!} = 10$$

$n_1 = 120$ ve $n_2 = 10$ ile çarpma kuralını (Kural 2.1) kullanarak, $(120)(10) = 1200$ değişik şekilde yolumuz olur.

Örnek 2.18: STATISTICS kelimesindeki harflerden kaç farklı harf düzenlemesi yapılabilir?

Cevap 2.18: Teorem 2.6 tartışmasındaki argümanın aynısını kullanarak, bu örnekte Teorem 2.5'i şu şekilde uygulayabiliriz:

$$\binom{10}{3, 3, 2, 1, 1} = \frac{10!}{(3)! (3)! (2)! (1)! (1)!} = 50400$$

Burada toplam 10 harf var, her biri 3 kez görünen 2 harf (S, T), 2 kez görünen I harfi ve 1 kez görünen A ve C harfleri. Öte yandan, bu sonuç doğrudan Teorem 2.4 kullanılarak elde edilebilir.

2.4 BİR OLAYIN OLASILIĞI

Olasılık teorisinin bu gelişiminin bir sonucu olarak, istatistiksel çıkarım, tüm tahminleri ve genellemeleriyle birlikte, şans oyunlarının çok ötesine geçerek, politika, iş dünyası, hava tahmini ve bilimsel araştırma gibi şans olayları ile ilgili diğer birçok alanı kapsayacak şekilde genişledi. Bu tahminlerin ve genellemelerin makul ölçüde doğru olması için, temel olasılık teorisinin anlaşılması esastır.

"Ahmet muhtemelen tenis maçını kazanacak" ifadesini söylediğimizde ya da "Bir zar atıldığında çift sayı gelme olasılığım yüzde elli" veya "Üniversitenin bu akşamki futbol maçını kazanması pek olası değil" ya da "Mezun olan sınıfımızın çoğu muhtemelen 3 yıl içinde evlenecek" dediğimizde ne demek istiyoruz? Her durumda, kesin olmadığımız bir sonucu ifade

ediyoruz, ancak geçmiş bilgiler veya deneyin yapısını anlamamız nedeniyle, ifadenin geçerliliğine bir dereceye kadar güveniyoruz.

Bu bölümün geri kalanında, yalnızca örnek uzayının sonlu sayıda eleman içerdiği deneyleri ele alacağız. Böyle bir istatistiksel deneyden kaynaklanan bir olayın meydana gelme olasılığı, 0 ile 1 arasında değişen ağırlıklar veya olasılıklar adı verilen bir dizi gerçek sayı aracılığıyla değerlendirilir. **Örnek uzaydaki her noktaya, öyle bir olasılık atarız ki tüm olasılıkların toplamı 1'dir.** Deney yapıldığında belirli bir örnekleme noktasının olma olasılığının yüksek olduğuna inanmak için nedenimiz varsa, atanan olasılık 1'e yakın olmalıdır. Öte yandan, gerçekleşmesi muhtemel olmayan bir numune noktasına 0'a yakın bir olasılık atanır. Madeni para veya zar atmak gibi birçok deneyde, tüm numune noktalarının oluşma şansı aynıdır ve eşit olasılıklar atanır. Örnek uzayın dışındaki noktalar için yani gerçekleşmesi mümkün olmayan basit olaylar için 0 olasılık atarız.

Tanım 2.9: Bir A olayının olasılığı, A 'daki tüm numune noktalarının ağırlıklarının toplamıdır. Bu yüzden

$$0 \leq P(A) \leq 1, P(\emptyset) = 0, \text{ ve } P(S) = 1.$$

Ayrıca, eğer $A_1, A_2, A_3, \dots$ birbirini dışlayan olaylar dizisi ise, o zaman

$$P(A_1 \cup A_2 \cup A_3 \cup \dots) = P(A_1) + P(A_2) + P(A_3) + \dots.$$

Bir A olayının olasılığını bulmak için A 'daki örnek noktalara atanan tüm olasılıkları toplarız. Bu toplam A 'nın olasılığı olarak adlandırılır ve $P(A)$ ile gösterilir.

Örnek 2.19: Bir madeni para iki kez atılıyor. En az 1 yazı gelme olasılığı nedir?

Cevap 2.19: Bu deney için örnek uzay

$$S = \{YY, YT, TY, TT\}.$$

Madeni para dengeliyse (hilesiz ise), bu sonuçların her birinin gerçekleşmesi eşit derecede olasıdır. Bu nedenle, her numune noktasına bir ω olasılığı atarız. O zaman $4\omega = 1$ veya $\omega = 1/4$. A , meydana gelen en az 1 yazı olayını temsil ediyorsa, o zaman

$$A = \{YY, YT, TY\} \text{ ve } P(A) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}.$$

Örnek 2.20: Bir zar, çift bir sayının tek bir sayının ortaya çıkma olasılığının iki katı olacak şekilde ayarlanmıştır (hileli zar). E, zarın bir kerelik atışında 4'ten küçük bir sayının gelmesi olayı ise, $P(E)$ 'yi bulunuz.

Cevap 2.20: Örnek uzay $S = \{1, 2, 3, 4, 5, 6\}$ şeklindedir. Her tek sayıya w olasılığı ve her çift sayıya $2w$ olasılığı olarak belirtirsek. Çünkü hileli olarak ayarlanmış bir zar vardır. Olasılıkların toplamı 1 olması gerektiğinden, $9w = 1$ veya $w = 1/9$ elde ederiz. Bu nedenle, sırasıyla her bir tek ve çift sayının gelme olasılıkları $1/9$ ve $2/9$ olarak elde edilir. Öyleyse,

$$E = \{1, 2, 3\} \text{ ve } P(E) = \frac{1}{9} + \frac{2}{9} + \frac{1}{9} = \frac{4}{9}.$$

Örnek 2.21: Örnek 2.20'deki şartlara göre çift sayı gelme olayı A, 3'e bölünebilen sayı gelme olayı B olsun. $P(A \cup B)$ ve $P(A \cap B)$ 'yi bulunuz.

Cevap 2.21: $A = \{2, 4, 6\}$ ve $B = \{3, 6\}$ olayları için $A \cup B = \{2, 3, 4, 6\}$ ve $A \cap B = \{6\}$ olarak belirlenir.

Her tek sayıya $1/9$ ve her çift sayı $2/9$ olasılığına sahip olduğu Örnek 1.20'de belirlenmişti. Dolayısıyla,

$$A \cup B = \{2, 3, 4, 6\} \text{ ve } P(A \cup B) = \frac{2}{9} + \frac{1}{9} + \frac{2}{9} + \frac{2}{9} = \frac{7}{9}.$$

$$A \cap B = \{6\} \text{ ve } P(A \cap B) = \frac{2}{9}.$$

Kural 2.3: Bir deney N farklı eşit derecede olası sonuçtan herhangi biriyle sonuçlanabiliyorsa ve bu sonuçların tam olarak n 'si A olayına karşılık geliyorsa, o zaman A olayının olasılığı $P(A) = \frac{n}{N}$ 'dir.

Örnek 2.22: Mühendisler için bir istatistik sınıfı 25 endüstri, 10 makine, 10 elektrik ve 8 inşaat mühendisliği öğrencisinden oluşmaktadır. Dersin sorumlusu tarafından bir soruyu yanıtlaması için rastgele bir kişi seçilirse, seçilen öğrencinin (a) endüstri mühendisliği bölümü ve (b) inşaat mühendisliği veya elektrik mühendisliği bölümü olma olasılığını bulunuz.

Cevap 2.22: Endüstri, makine, elektrik ve inşaat mühendisliği bölümlerinde okuyan öğrencileri sırasıyla I, M, E ve C ile gösterelim. Sınıftaki toplam öğrenci sayısı 53 olup, hepsinin seçilme olasılığı eşittir.

(a) 53 öğrenciden 25'i endüstri mühendisliği alanında eğitim aldığından, seçilen öğrencinin rastgele bir endüstri mühendisliği bölümünde olması I'nın olasılığı,

$$P(I) = \frac{25}{53}.$$

(b) 53 öğrenciden toplam 18'i inşaat veya elektrik mühendisliği mezunu olduğundan,

$$P(C \cup E) = \frac{18}{53}.$$

Örnek 2.23: Toplam 52 tane oyuncak arabası bulunan çocuğun 4 tane A marka oyuncak arabası ve 4 tane B marka oyuncak arabası vardır. Annesi çocuğuna rastgele olarak 5 oyuncak araba dağıtırsa bu 5 tane oyuncak arabasından 2 tane A marka oyuncak arabası ve 3 tane B marka oyuncak arabası olma olasılığını bulunuz.

Cevap 2.23: 4 tane A marka oyuncak arabadan 2 tane A marka oyuncak arabanın dağıtılma şeklinin sayısı

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} = \binom{4}{2} = \frac{4!}{2!(4-2)!} = 6$$

4 tane B marka oyuncak arabadan 3 tane B marka oyuncak arabanın dağıtılma şeklinin sayısı ise

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} = \binom{4}{3} = \frac{4!}{3!(4-3)!} = 4.$$

Çarpma kuralına göre (Kural 2.1), çocuğun elinde $n = (6)(4) = 24$ tane 2 A marka ve 3 B marka oyuncak olmasının yolu vardır.

Hepsi eşit olasılığa sahip 5 tane oyuncak arabanın dağıtılmasının toplam sayısı ise

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} = \binom{52}{5} = \frac{52!}{5!(52-5)!} = 2598960.$$

Bu nedenle, 5 tane oyuncak arabanın dağıtıldığı bir durumda 2 tanesi A ve 3 tanesi B marka oyuncak gelme olasılığı

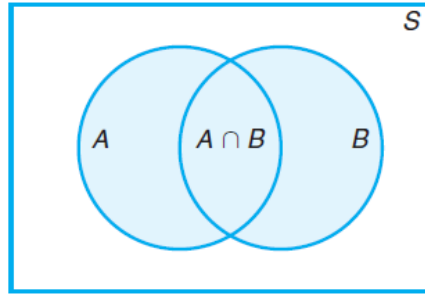
$$P(C) = \frac{24}{2598960} = 0.9 \cdot 10^{-5}.$$

2.5 TOPLAMA ÖNERMELERİ

Çoğu zaman, bazı olayların olasılığını diğer olayların bilinen olasılıklarından hesaplamak en kolay yoldur. Söz konusu olay, diğer iki olayın birleşimi veya bazı olayların tamamlayıcısı olarak temsil edilebiliyorsa, bu tanım doğru olabilir. Olasılıkların hesaplanmasını sıklıkla basitleştiren birkaç önemli yasa bunu takip eder.

TEOREM 2.7: A ve B iki olay ise, o zaman $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

ISPAT: Şekil 2.6'daki Venn diyagramını ele alalım. $P(A \cup B)$, $A \cup B$ 'deki örnek noktaların olasılıklarının toplamıdır. Şimdi $P(A) + P(B)$, A'daki tüm olasılıklar ile B'deki tüm olasılıkların toplamıdır. Bu nedenle $(A \cap B)$ içindeki olasılıkları iki kez ekledik. Bu olasılıkların toplamı $P(A \cap B)$ olduğundan, $A \cup B$ 'deki olasılıkların toplamını elde etmek için bu olasılığı bir kez çıkarmalıyız.



Şekil 2.6: Toplama önermesi

Sonuç 2.1: A ve B olayları ayrık ise (birbirini dışlıyorsa), $P(A \cup B) = P(A) + P(B)$ olur.

Sonuç 2.1, Teorem 2.7'nin hemen bir sonucudur, çünkü eğer A ve B karşılıklı olarak birbirini dışlıyorsa, $A \cap B = \emptyset$ ve sonra $P(A \cap B) = P(\emptyset) = 0$. Genel olarak, Sonuç 2.2'yi yazabiliriz.

Sonuç 2.2: $A_1, A_2, \dots, A_n$ olayları ayrık olaylar ise, o zaman $P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$.

$A_1, A_2, \dots, A_n$ olayları ayrık ve $A_1 \cup A_2 \cup \dots \cup A_n = S$ ise, $\{A_1, A_2, \dots, A_n\}$ olaylarının tümüne S örneklem uzayının **bölüntülenmesi** denir.

Sonuç 2.3: $A_1, A_2, \dots, A_n$ olayları S örneklem uzayının bir bölüntülenmesi ise, o zaman $P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n) = P(S) = 1$.

Teorem 2.7'yi genişletirsek;

TEOREM 2.8: A, B , ve C üç olayı için;
 $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$.

Örnek 2.24: Mehmet, dönem sonunda bir üniversitenin elektrik elektronik mühendisliği bölümünden mezun olacak bir öğrencidir. Beğendiği iki şirketle görüştüktan sonra A şirketinden teklif alma olasılığının 0.8, B şirketinden teklif alma olasılığının ise 0.6 olduğunu değerlendirmektedir. Her iki şirketten de teklif alma olasılığının 0.5 olduğuna inanıyorsa, bu iki şirketten en az birisinden teklif alma olasılığı nedir?

Cevap 2.24: Toplam önermesini kullanarak,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.8 + 0.6 - 0.5 = 0.9.$$

Örnek 2.25: Bir çift zar atıldığında toplamının 7 veya 11 gelme olasılığı nedir?

Cevap 2.25: 7'nin meydana geldiği olay A ve 11'in geldiği olay B olsun. Şimdi, toplam 36 tane örneklem noktasından 6 tanesinde $\{\{1,6\}, \{2,5\}, \{3,4\}, \{4,3\}, \{5,2\}, \{6,1\}\}$ zarların toplamı 7 olur ve sadece tanesinde $\{\{5,6\}, \{6,5\}\}$ zarların toplamı 11 olur. Tüm örneklem noktaları eşit derecede olası olduğundan, $P(A) = 6/36 = 1/6$ ve $P(B) = 2/36 = 1/18$ 'e sahibiz. A ve B olayları birbirini dışlar (yani A ve B olayları ayrıktır), çünkü toplam 7 ve 11 aynı atışta gerçekleşemez. Öyleyse,

$$P(A \cup B) = P(A) + P(B) = \frac{1}{6} + \frac{1}{18} = \frac{2}{9}.$$

Bu sonuç Kural 1.3 ile de hesaplanabilirdi. Şöyle ki toplam örneklem noktası 36, istenilen durum ise toplamda yani zarların toplamının 7 ve 11 gelmesinin durumları 8 örneklem noktasında elde edilmektedir. Bu durumda

$$P(A \cup B) = \frac{n}{N} = \frac{8}{36} = \frac{2}{9}.$$

Örnek 2.26: Yeni bir otomobil satın alan bir kişinin yeşil, beyaz, kırmızı ve mavi rengi seçme olasılığı sırasıyla 0.09, 0.15, 0.21 ve 0.23 ise, bu renklerdeki otomobillerden herhangi birini satın alma olasılığı nedir?

Cevap 2.26: Bir alıcının sırasıyla yeşil, beyaz, kırmızı veya mavi bir otomobili seçtiği olaylar G, W, R ve B olsun. Bu dört olay birbirini dışlamaktadır (yani ayrık olaylardır), dolayısıyla

$$P(G \cup W \cup R \cup B) = P(G) + P(W) + P(R) + P(B) = 0.09 + 0.15 + 0.21 + 0.23 = 0.68.$$

Genellikle bir olayın olma olasılığını hesaplamak, o olayın olmama olasılığını hesaplamaktan daha zordur. A olayı için durum böyleyse, basitçe önce $P(A')$ 'yi buluruz ve sonra Teorem 2.7'yi kullanarak çıkarma yoluyla $P(A)$ 'yi buluruz.

TEOREM 2.9: A ve A' birbirini tamamlayan olaylar ise, o zaman

$$P(A) + P(A') = 1.$$

Örnek 2.27: Bir otomobil tamircisinin herhangi bir iş gününde 3, 4, 5, 6, 7, 8 veya daha fazla araca bakım yapma olasılıkları sırasıyla 0.12, 0.19, 0.28, 0.24, 0.10 ve 0.07'dir. Bu durumda ertesi gün iş yerinde en az 5 araba tamir etme olasılığı nedir?

Cevap 2.27: En az 5 aracı tamir etme olayı E olsun. Şimdi, $P(E) = 1 - P(E')$, burada E' , 5'ten az araca hizmet verilmesi olayıdır. Böylece,

$$P(E') = 0.12 + 0.19 = 0.31,$$

Teorem 2.9'dan $P(E) = 1 - 0.31 = 0.69$ sonucu çıkar.

Örnek 2.28: Belirli bir bilgisayar kablosunun uzunluğu için üreticinin belirtilmelerinin 2000 ± 10 milimetre olduğunu varsayalım. Bu sektörde, küçük kablonun büyük kablo kadar arızalı (özellikleri karşılamayan) olduğu bilinmektedir. Yani, uzunluğu 2010 milimetreyi aşan bir kabloyu rastgele üretme olasılığı, uzunluğu 1990 milimetreden küçük bir kablo üretme olasılığına eşittir. Üretim prosedürünün spesifikasyonları karşılama olasılığı 0,99 olarak bilinmektedir.

(a) Rastgele seçilen bir kablonun çok büyük olma olasılığı nedir?

(b) Rastgele seçilen bir kablonun 1990 milimetreden büyük olma olasılığı nedir?

Cevap 2.28: M, bir kablonun spesifikasyonları karşılaması olayı olsun. Sırasıyla S ve L, kablonun çok küçük ve çok büyük olduğu olaylar olsun. Sonra

(a) $P(M) = 0.99$ ve $P(S) = P(L) = (1 - 0.99)/2 = 0.005$.

(b) Rastgele seçilmiş bir kablonun uzunluğunu X ile göstererek,

$P(X \geq 1990) + P(X < 1990) = 1$. Yani kablonun 1990 milimetreden küçük ve büyük olma olasılıklarının toplamı olasılığın tanımı gereği 1'dir. Teorem 2.9 kullanılarak,

$$P(X \geq 1990) = 1 - P(S) = 1 - 0.005 = 0.995.$$

PROBLEM 2.1: Bir deney, biri yeşil biri kırmızı olan bir çift zar atmayı ve gelen sayıları kaydetmeyi içermektedir.

- a) Sayıların toplamı 8'den büyük olan A olayına karşılık gelen öğeleri listeleyiniz;
- b) Her iki zardan birinde 2'nin geldiği B olayına karşılık gelen elemanları listeleyin;
- c) Yeşil zarda 4'ten büyük bir sayının geldiği C olayına karşılık gelen elemanları listeleyin;
- d) $A \cap C$ olayına karşılık gelen elemanları listeleyin;
- e) $A \cap B$ olayına karşılık gelen elemanları listeleyin;
- f) $B \cap C$ olayına karşılık gelen elemanları listeleyin;
- g) A , B ve C olaylarının kesişimlerini ve birleşimlerini göstermek için bir Venn diyagramı oluşturunuz.

PROBLEM 2.2:

- (a) 6 kişi bir otobüse binmek için kaç farklı şekilde dizilebilir?
- (b) 6 kişiden 3'ü birbirini takip etmekte ısrar ederse, kaç farklı şekilde dizilebilir?
- (c) 6 kişiden 2'si birbirini takip etmeyi reddederse, kaç farklı şekilde dizilebilir?

PROBLEM 2.3: Bölgesel heceleme yarışmasında 8 finalist 3 erkek ve 5 kızdan oluşmaktadır. Yarışmanın sonunda S örnek uzayındaki örneklem noktalarının sayısını

- a.) 8 finalistin tümü için
- b.) İlk 3 pozisyon için bulunuz.

PROBLEM 2.4: 100 öğrencilik bir lise mezuniyet sınıfında 54 öğrenci matematik, 69 öğrenci tarih ve 35 öğrenci hem matematik hem de tarih okudu. Bu öğrencilerden biri rasgele seçilirse,

- (a) öğrencinin matematik veya tarih dersi alma olasılığını bulunuz.
- (b) öğrencinin bu derslerden hiçbirini almamış olma olasılığını bulunuz.
- (c) öğrencinin tarih aldığı ama matematik almadığı olasılığını bulunuz.

OLASILIK VE İSTATİSTİK

3. HAFTA

OLASILIK (DEVAM)

2.6 KOŞULLU OLASILIK, BAĞIMSIZLIK VE ÇARPIM KURALI

Olasılık teorisinde çok önemli bir kavram koşullu olasılıktır. Bazı uygulamalarda, uygulayıcı belirli kısıtlamalar altında olasılık yapısıyla ilgilenir. Örneğin, epidemiyolojide, genel popülasyondan bir kişinin diyabet olma olasılığını incelemek yerine, 35 ila 50 yaş aralığındaki Asyalı kadınlar veya 40 ila 60 yaş aralığında İspanyol erkekler gibi farklı bir grup için bu olasılığı bilmek daha ilgi çekici olabilir. Bu tür olasılığa koşullu olasılık denir.

A. Koşullu Olasılık

Bir A olayının meydana geldiği bilindiğinde bir B olayının olma olasılığına koşullu olasılık denir ve $P(B|A)$ ile gösterilir. $P(B|A)$ sembolü genellikle "A'nın olması durumunda B'nin olma olasılığı" veya basitçe "A verildiğinde B'nin olasılığı" olarak okunur.

Bir zar atıldığında tam kare elde etme B olayını ele alalım. Zar, çift sayıların tek sayıların ortaya çıkma olasılığının iki katı olacak şekilde yapılandırılmıştır. Sırasıyla tek ve çift sayılara atanan $1/9$ ve $2/9$ olasılıklarıyla $S = \{1, 2, 3, 4, 5, 6\}$ örnek uzayına dayalı olarak, B'nin oluşma olasılığı $1/3$ 'tür. Şimdi, zarın atılmasının 3'ten büyük bir sayıyla sonuçlandığının bilindiğini varsayalım. Şimdi, S'nin bir altkümesi olan $A = \{4, 5, 6\}$ örnek uzayıyla uğraşıyoruz. Olasılığı bulmak için B'nin A uzayına göre meydana gelmesi için, önce A'nın elemanlarına, toplamaları 1 olacak şekilde orijinal olasılıklarıyla orantılı yeni olasılıklar atamalıyız. A'daki tek sayıya w olasılığı ve B'ye $2w$ olasılığı atamak iki çift sayı, elimizde $5w = 1$ veya $w = 1/5$ var. A uzayına göre, B'nin tek eleman 4'ü içerdiğini buluruz. Bu olayı $B|A$ sembolü ile göstererek, $B|A = \{4\}$ yazarız ve dolayısıyla $P(B|A) = 2/5$ olur.

Bu örnek, olayların farklı örnek uzaylara göre düşünüldüğünde farklı olasılıklara sahip olabileceğini göstermektedir.

Ayrıca $P(B|A) = \frac{2}{5} = \frac{2/9}{5/9} = \frac{P(A \cap B)}{P(A)}$ yazabiliriz, burada $P(A \cap B)$ ve $P(A)$ orijinal

örneklem uzayı S'den bulunmaktadır. Başka bir deyişle, S'nin bir alt uzayı A'ya göre koşullu bir olasılık, doğrudan orijinal örneklem uzayı S'nin elemanlarına atanan olasılıklardan hesaplanabilir.

Tanım 2.10: A verildiğinde, $P(B|A)$ ile gösterilen B'nin koşullu olasılığı, $P(A) > 0$ olması koşuluyla $P(B|A) = \frac{P(A \cap B)}{P(A)}$ ile tanımlanır.

Ek bir örnek olarak, örnek uzayımızın S küçük bir kasabada üniversite derecesi için gereklilikleri tamamlamış yetişkin nüfusu olduğunu varsayalım. Bunları cinsiyete ve çalışma durumuna göre sınıflandıracamız. Veriler Tablo 2.1'de verilmiştir.

Tablo 2.1: Küçük Bir Kasabadaki Yetişkinlerin Sınıflandırılması

	Çalışan	İşsiz	Toplam
Erkek	460	40	500
Kadın	140	260	400
Toplam	600	300	900

Bu kişilerden biri, kasabada yeni endüstriler kurmanın avantajlarını tanıtmak için ülke çapında bir tur için rastgele seçilecek. Aşağıdaki olaylarla ilgileneceğiz:

M: Rastgele seçilen birinin erkek olma olayı seçilir,

E: Rastgele seçilen birinin çalışan olma olayı.

İndirgenmiş örnek uzay E'yi kullanarak, $P(M|E) = 460/600 = 23/30$ olduğunu buluruz. Yani seçilen birinin çalışan olduğu biliniyorsa erkek olma olasılığını bulmuş oluruz.

$n(A)$ herhangi bir A kümesindeki eleman sayısını gösterebilir. Bu gösterimi kullanarak, her yetişkinin seçilme şansı eşit olduğundan,

$$P(M|E) = \frac{n(E \cap M)}{n(E)} = \frac{n(E \cap M)/n(S)}{n(E)/n(S)} = \frac{P(E \cap M)}{P(E)} \quad \text{yazabiliriz, burada}$$

$P(E \cap M)$ ve $P(E)$ orijinal örnek uzay S'den bulunmaktadır.

Bu sonucu doğrulamak için:

$$P(E) = \frac{600}{900} = \frac{2}{3} \text{ ve } P(E \cap M) = \frac{460}{900} = \frac{23}{45}. \text{ Böylece } P(M|E) = \frac{23/45}{2/3} = \frac{23}{30}$$

Örnek 2.28: Düzenli tarifeli bir uçuşun zamanında kalkma olasılığı $P(D) = 0.83$ 'tür; zamanında varma olasılığı $P(A) = 0.82$ 'dir; ve zamanında kalkıp zamanında varma olasılığı $P(D \cap A) = 0.78$ 'dir. Bir uçağın

(a) zamanında kalktıysa zamanında varma olasılığını bulunuz

(b) zamanında vardıysa zamanında kalkması olasılığını bulunuz.

Cevap 2.28: (a) Zamanında kalktığı düşünülürse, bir uçağın zamanında varma olasılığı

$$P(A|D) = P(D \cap A)/P(D) = 0.78/0.83 = 0.94 \text{ 'tür.}$$

(b) Bir uçağın zamanında vardığına göre, zamanında kalkma olasılığı

$$P(D|A) = P(D \cap A)/P(A) = 0.78/0.82 = 0.95 \text{ 'tir.}$$

Koşullu olasılık kavramı, ek bilgiler ışığında, yani başka bir olayın meydana geldiği bilindiğinde, bir olayın olasılığı fikrini yeniden değerlendirme yeteneği sağlar. $P(A|B)$ olasılığı, B olayının meydana geldiği bilgisine dayalı olarak $P(A)$ 'nın güncellenmesidir. Örnek 2.28'de, uçağın zamanında varma olasılığını bilmek önemlidir. Uçağın zamanında kalkmadığı bilgisi veriliyor. Bu ek bilgiyle donanmış olarak, daha uygun olasılık $P(A|D')$, yani zamanında ayrılmadığı göz önüne alındığında, zamanında varma olasılığı hesaplanabilir. Çoğu durumda, daha önemli olan koşullu olasılığın gözlemlenmesinden elde edilen sonuçlar tabloyu tamamen değiştirir. Bu örnekte, $P(A|D')$ hesaplaması şu şekildedir:

$$P(A|D') = \frac{P(A \cap D')}{P(D')} = \frac{P(A) - P(A \cap D)}{1 - P(D)} = \frac{0.82 - 0.78}{0.17} = 0.24$$

Sonuç olarak, ek bilgilerin varlığında zamanında varma olasılığı ciddi şekilde azalır.

- $(A \cap D) \cup (A \cap D') = A \cap (D \cup D') = A$
- $P(A \cap D) + P(A \cap D') = P(A)$

Örnek 2.29: Koşullu olasılık kavramının hem endüstriyel hem de biyomedikal uygulamalarda sayısız kullanımı vardır. Belirli bir kumaş türünün şeritlerinin üretildiği tekstil endüstrisindeki endüstriyel bir süreci düşünün. Bu şeritler, uzunluk ve doku yapısı olmak üzere iki şekilde kusurlu olabilir. İkincisi durumunda, tanımlama süreci çok karmaşıktır. Süreçle ilgili tarihsel bilgilerden, şeritlerin %10'unun uzunluk testinde, %5'inin doku testinde başarısız olduğu ve yalnızca %0.8'inin her iki testte de başarısız olduğu bilinmektedir. Süreçten rastgele bir şerit seçilirse ve hızlı bir ölçüm bunun uzunluk testinde başarısız olduğunu belirlerse, bunun doku kusurlu olma olasılığı nedir?

Cevap 2.29: Olayları göz önünde bulundurun

L: uzunluk kusurlu, T: doku kusurlu.

Şeridin uzunluğunda kusurlu olduğu göz önüne alındığında, bu şeridin doku kusurlu olma olasılığı şu şekilde verilir

$$P(T|L) = \frac{P(T \cap L)}{P(L)} = \frac{0.008}{0.1} = 0.08$$

Bu nedenle, koşullu olasılığı bilmek, yalnızca $P(T)$ 'yi bilmekten çok daha fazla bilgi sağlar.

B. Bağımsız Olaylar

Şimdi, sıradan bir desteden arka arkaya 2 kartın değiştirilerek çekildiği bir deneyi düşünün. Olaylar şu şekilde tanımlanır:

A: ilk kart bir as,

B: ikinci kart bir maçadır.

Birinci kart değiştirildiği için hem birinci hem de ikinci çekiliş için örnek alanımız 4 as ve 13 maça içeren 52 karttan oluşuyor. Dolayısıyla, $P(B|A) = 13/52 = 1/4$ ve $P(B) = 13/52 = 1/4$.

Yani, $P(B|A) = P(B)$ olduğunda, A ve B olaylarının **bağımsız** olduğu söylenir.

Koşullu olasılık, ek materyaller ışığında bir olayın olasılığının değiştirilmesine izin verse de, aynı zamanda çok önemli olan bağımsızlık kavramını veya mevcut bağlamda bağımsız olayları daha iyi anlamamızı sağlar. Örnek 2.28'deki havaalanı sorusunda, $P(A|D)$ $P(A)$ 'dan farklıdır. Bu, D'nin ortaya çıkmasının A'yı etkilediğini gösterir ve bu çizimde kesinlikle bu beklenir. Ancak, A ve B olaylarının ve $P(A|B) = P(A)$ olduğu durumu ele alalım.

Başka bir deyişle, B'nin meydana gelmesinin A'nın olma ihtimali üzerinde hiçbir etkisi olmamıştır. Burada A'nın meydana gelmesi B'nin meydana gelmesinden bağımsızdır. Bağımsızlık kavramının önemi ne kadar vurgulansa azdır. Bu dersin hemen hemen tüm konularında ve uygulamalı istatistiğin tüm alanlarında önemli bir rol oynar.

Tanım 2.11: Koşullu olasılıkların var olduğu varsayılarak, A ve B iki olayı ancak ve ancak $P(B|A) = P(B)$ veya $P(A|B) = P(A)$ ise bağımsızdır. Aksi takdirde, A ve B bağımlıdır.

$P(B|A) = P(B)$ koşulu, $P(A|B) = P(A)$ anlamına gelir ve bunun tersi de geçerlidir. $P(B|A) = P(B) = 1/4$ olduğunu gösterdiğimiz kart destesi deneylerinde, $P(A|B) = P(A) = 1/13$ olduğunu da görebiliriz.

C. Çarpma Önergeleri/Çarpım Kuralı

Tanım 2.10'daki formülü $P(A)$ ile çarparak, iki olayın birden olma olasılığını hesaplamamızı sağlayan aşağıdaki önemli çarpma kuralını (veya çarpım kuralını) elde ederiz.

TEOREM 2.10: Bir deneyde A ve B olaylarının her ikisi de gerçekleşebilirse, o zaman $P(A) > 0$ olması koşuluyla $P(A \cap B) = P(A) \cdot P(B|A)$.

Böylece, hem A'nın hem de B'nin olma olasılığı, A'nın olma olasılığı ile A'nın olması durumunda B'nin olma koşullu olasılığının çarpımına eşittir. $A \cap B$ ve $B \cap A$ olayları eşdeğer olduğundan, Teorem 2.10'dan $P(A \cap B) = P(B \cap A) = P(B) \cdot P(A|B)$ yazabileceğimiz sonucu çıkar.

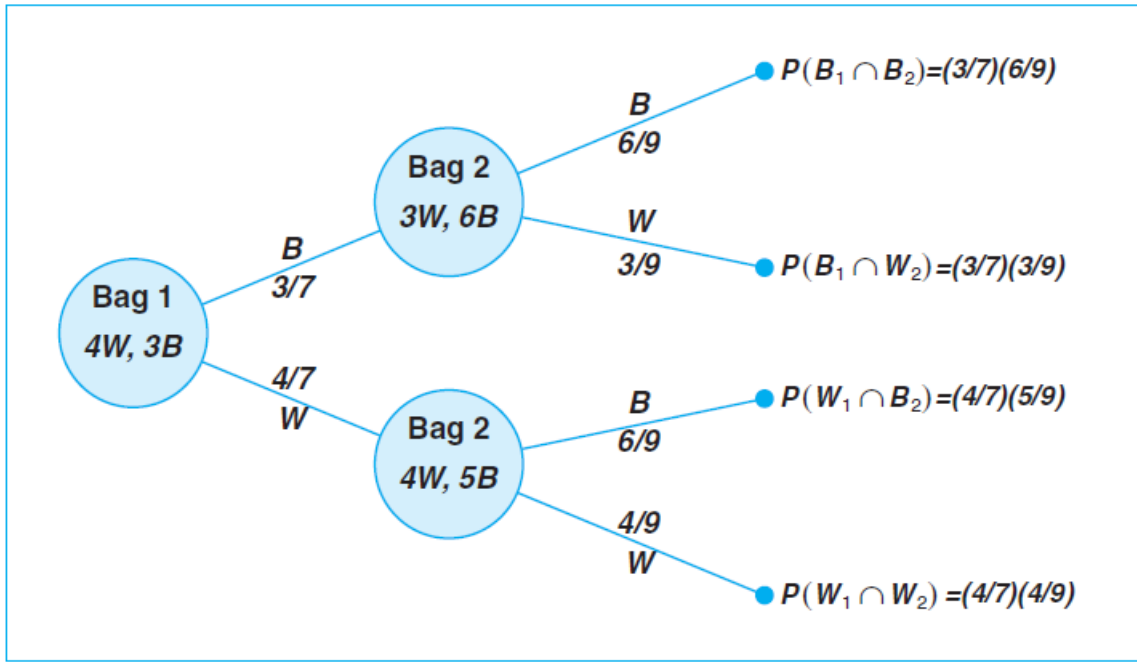
Yani hangi olayın A, hangi olayın B olarak adlandırıldığı önemli değildir.

Örnek 2.30: 5'i arızalı 20 sigorta içeren bir sigorta kutumuz olduğunu varsayalım. Rastgele 2 sigorta seçilir ve ilki değiştirilmeden arka arkaya kutudan çıkarılırsa, her iki sigortanın da arızalı olma olasılığı nedir?

Cevap 2.30: Birinci sigortanın arızalı olması durumunda A, ikinci sigortanın arızalı olması durumunda ise B diyelim; sonra $A \cap B$ 'yi A'nın ve ardından B'nin A'nın gerçekleşmesinden sonra meydana geldiği olay olarak yorumlarız. Önce arızalı bir sigortayı çıkarma olasılığı $1/4$ 'tür; o zaman kalan 4 tane arızalı sigortadan ikinci bir arızalı sigortayı çıkarma olasılığı $4/19$ 'dur. Dolayısıyla, $P(A \cap B) = (1/4) \cdot (4/19) = 1/19$.

Örnek 2.31: Birinci torbada 4 beyaz ve 3 siyah bilye vardır ve ikinci torbada 3 beyaz ve 5 siyah bilye vardır. Birinci torbadan bir bilye çekiliyor ve ikinci torbaya görülmeyecek şekilde atılıyor. Şimdi ikinci torbadan çekilen bir bilyenin siyah olma olasılığı kaçtır?

Cevap 2.31: Sırasıyla B_1 , B_2 ve W_1 , birinci torbadan siyah bir topun, ikinci torbadan siyah bir topun ve birinci torbadan bir beyaz topun çekilmesini temsil etsin. Birbirini dışlayan $B_1 \cap B_2$ ve $W_1 \cap B_2$ olaylarının birleşimiyle ilgileniyoruz. Çeşitli olasılıklar Şekil 2.7'de gösterilmektedir.



Şekil 2.7: Örnek 2.31 için ağaç diyagramı.

$$\begin{aligned}
 P[(B_1 \cap B_2) \text{ or } (W_1 \cap B_2)] &= P(B_1 \cap B_2) + P(W_1 \cap B_2) \\
 &= P(B_1)P(B_2|B_1) + P(W_1)P(B_2|W_1) \\
 &= \left(\frac{3}{7}\right) \left(\frac{6}{9}\right) + \left(\frac{4}{7}\right) \left(\frac{5}{9}\right) = \frac{38}{63}.
 \end{aligned}$$

Örnek 2.30'da, birinci sigorta değiştirilirse ve ikincisi çıkarılmadan önce sigortalar tamamen yeniden düzenlenirse, ikinci seçimdeki sigortanın arızalı olma olasılığı hala $1/4$ 'tür; yani, $P(B|A) = P(B)$ ve A ve B olayları bağımsızdır. Bu doğru olduğunda, aşağıdaki özel çarpma kuralını elde etmek için Teorem 2.10'da $P(B|A)$ yerine $P(B)$ koyabiliriz.

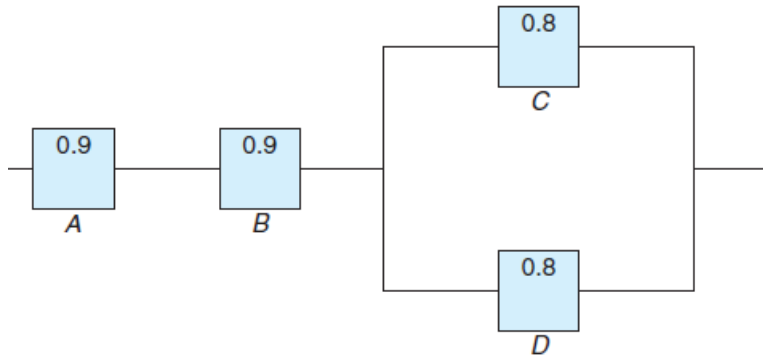
TEOREM 2.11: İki olay A ve B bağımsızdır ancak ve ancak $P(A \cap B) = P(A) \cdot P(B)$ ise.

Bu nedenle, iki bağımsız olayın meydana gelme olasılığını elde etmek için, sadece bunların bireysel olasılıklarının çarpımını buluruz.

Örnek 2.32: Küçük bir kasabada acil durumlar için bir itfaiye aracı ve bir ambulans bulunur. İtfaiyenin ihtiyaç duyulduğunda hazır olma olasılığı 0.98 ve çağrıldığında ambulansın hazır olma olasılığı 0.92'dir. Yanan bir binadan kaynaklanan bir yaralanma durumunda, bağımsız olarak çalıştıklarını varsayarak hem ambulansın hem de itfaiyenin hazır bulunma olasılığını bulunuz.

Cevap 2.32: A ve B, itfaiye ve ambulansın mevcut olduğu ilgili olayları temsil etsin. O zaman $P(A \cap B) = P(A) \cdot P(B) = (0.98) \cdot (0.92) = 0.9016$.

Örnek 2.33: Bir elektrik sistemi, Şekil 2.8'de gösterildiği gibi dört bileşenden oluşur. A ve B bileşenleri çalışırsa ve C veya D bileşenlerinden biri çalışırsa sistem çalışır. Her bir bileşenin güvenilirliği (çalışma olasılığı) da Şekil 2.8'de gösterilmiştir. (a) Tüm sistemin çalışma olasılığını bulun, (b) tüm sistem çalışırken C bileşeninin çalışmama olasılığını bulun. Dört bileşenin bağımsız olarak çalıştığını varsayalım.



Şekil 2.8: Örnek 2.33 için bir elektrik sistemi.

Cevap 2.33: Sistemin bu konfigürasyonunda, A, B ve alt sistem C ve D bir seri devre sistemi oluştururken, alt sistem C ve D'nin kendisi bir paralel devre sistemidir.

(a) Tüm sistemin çalışma olasılığı açıkça şu şekilde hesaplanabilir:

$$\begin{aligned} P[A \cap B \cap (C \cup D)] &= P(A)P(B)P(C \cup D) = P(A)P(B)[1 - P(C' \cap D')] \\ &= P(A)P(B)[1 - P(C')P(D')] \\ &= (0.9)(0.9)[1 - (1 - 0.8)(1 - 0.8)] = 0.7776. \end{aligned}$$

Yukarıdaki eşitlikler, dört bileşen arasındaki bağımsızlık nedeniyle geçerlidir.

(b) Bu durumda koşullu olasılığı hesaplamak için şuna dikkat edilir:

$$P = \frac{P(\text{sistem çalışırken } C' \text{'nin çalışmaması})}{P(\text{tüm sistemin çalışması})}$$
$$P = \frac{P(A \cap B \cap C' \cap D)}{P(\text{tüm sistemin çalışması})} = \frac{(0.9) * (0.9) * (1 - 0.8) * (0.8)}{(0.776)} = 0.1667.$$

Çarpım kuralı, iki olaydan daha fazla duruma genişletilebilir.

TEOREM 2.12: Bir deneyde, $A_1, A_2, \dots, A_k$ olayları gerçekleşebilirse, o zaman $P(A_1 \cap A_2 \cap \dots \cap A_k) = P(A_1)P(A_2 | A_1)P(A_3 | A_1 \cap A_2) \dots P(A_k | A_1 \cap A_2 \cap \dots \cap A_{k-1})$.
 $A_1, A_2, \dots, A_k$ olayları bağımsızdır, o halde $P(A_1 \cap A_2 \cap \dots \cap A_k) = P(A_1)P(A_2) \dots P(A_k)$.

Tanım 2.12: $A = \{A_1, A_2, \dots, A_n\}$, A 'nın herhangi bir alt kümesinde $A_{i_1}, \dots, A_{i_k}$, $k \leq n$ için ise karşılıklı olarak bağımsızdır, bu durumda $P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k})$.

2.7 BAYES ÖNERMELERİ/KURALI

Bayes istatistiği, bilim ve mühendislikteki birçok pratik durumda deneysel verilerin analizinde uygulanan özel bir istatistiksel çıkarım biçiminde kullanılan bir araçlar topluluğudur. Bayes kuralı, olasılık teorisindeki en önemli kurallardan biridir. **Genel olarak Bayes Teorimi; bir olasılık değerini, bildiğimiz diğer bazı olasılık değerlerini kullanarak hesaplama fikridir.**

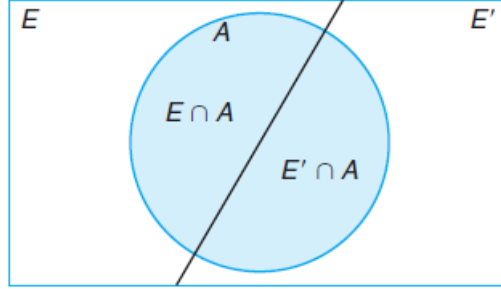
A. Toplam Olasılık

Şimdi, küçük bir kasabanın yetişkinleri arasından rastgele bir kişinin ülkeyi gezmek ve kasabada yeni endüstriler kurmanın avantajlarını duyurmak üzere seçildiği koşullu olasılıkta bahsedilen resme geri dönelim. Diyelim ki istihdam edilenlerden 36'sı ve işsiz olanlardan 12'si Tenis Kulübü üyesidir. A; rastgele seçilen birisinin Tenis Kulübü üyesi olma olasılığını bulmak istiyoruz. Bu durumda E; rastgele seçilen birisinin çalışan olması olasılığı olsun. Şekil 2.9'a atıfta bulunarak, A'yı birbirini dışlayan iki olay olan $(E \cap A)$ ve $(E' \cap A)$ 'nın birleşimi olarak

yazabiliriz. Dolayısıyla, $A = (E \cap A) \cup (E' \cap A)$ ve Teorem 2.7'nin Sonuç 2.1'in ve ardından Teorem 2.10'a göre

$$\begin{aligned} P(A) &= P[(E \cap A) \cup (E' \cap A)] = P(E \cap A) + P(E' \cap A) \\ &= P(E)P(A|E) + P(E')P(A|E'). \end{aligned}$$

yazabiliriz.



Şekil 2.9: A, E ve E' olayları için Venn diyagramı.

Bölüm 2.6'daki koşullu olasılık kısmının başlangıcındaki veriler, A kümesi için yukarıda verilen ek verilerle birlikte hesaplamamızı sağlar.

$$P(E) = \frac{600}{900} = \frac{2}{3}, \quad P(A|E) = \frac{36}{600} = \frac{3}{50},$$

$$P(E') = \frac{1}{3}, \quad P(A|E') = \frac{12}{300} = \frac{1}{25}.$$

Böylece $P(A) = P(E)P(A|E) + P(E')P(A|E') = \left(\frac{2}{3}\right)\left(\frac{3}{50}\right) + \left(\frac{1}{3}\right)\left(\frac{1}{25}\right) = \frac{4}{75}.$

Örnek uzayın k altkümeye bölündüğü durum için yukarıdaki gösterimin genelleştirilmesi, bazen toplam olasılık teoremi veya eleme kuralı (yok etme kuralı) olarak adlandırılan aşağıdaki teorem tarafından kapsanır.

TEOREM 2.13: $i = 1, 2, \dots, k$ için $P(B_i) = 0$ olsun. $B_1, B_2, \dots, B_k$ S örneklem uzayının bir bölüntülenmesi ise

O zaman S'nin herhangi bir A olayı için A'nın olasılığı;

$$P(A) = \sum_{i=1}^k P(B_i \cap A) = \sum_{i=1}^k P(B_i)P(A|B_i)$$

Örnek 2.34: Belirli bir montaj fabrikasında, B_1, B_2 ve B_3 olmak üzere üç makine, ürünlerin sırasıyla %30, %45 ve %25'ini oluşturmaktadır. Her bir makinenin ürettiği ürünlerin sırasıyla %2, %3 ve %2 oranında kusurlu olduğu geçmiş deneyimlerden bilinmektedir. Şimdi, bitmiş bir ürünün rastgele seçildiğini varsayalım. Bu ürünün arızalı olma olasılığı nedir?

Cevap 2.34: Aşağıdaki olayları göz önünde bulundurun:

A: ürün kusurlu,

B_1 : Ürün B_1 makinesi tarafından yapılır,

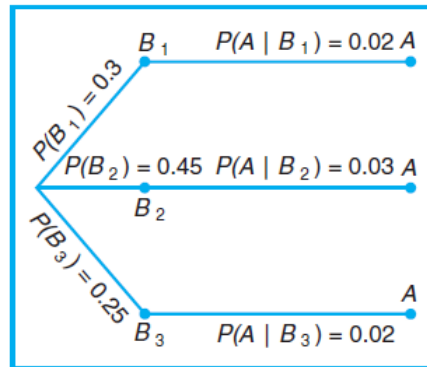
B_2 : Ürün B_2 makinesi tarafından yapılır,

B_3 : Ürün B_3 makinesi tarafından yapılmıştır.

Yok etme kuralını uygulayarak yazabiliriz.

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3).$$

Şekil 2.10'daki ağaç diyagramına bakarak, üç dalın olasılıkları verdiğini görüyoruz.



Şekil 2.10: Örnek 2.7 için Venn diyagramı.

$$P(B_1)P(A|B_1) = (0.3)(0.02) = 0.006,$$

$$P(B_2)P(A|B_2) = (0.45)(0.03) = 0.0135,$$

$$P(B_3)P(A|B_3) = (0.25)(0.02) = 0.005,$$

$$P(A) = 0.006 + 0.0135 + 0.005 = 0.0245.$$

B. Bayes' Kuralı/Baye's Önermesi

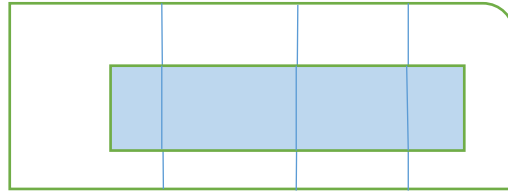
Örnek 2.34'te eleme kuralına göre $P(A)$ 'yı istemek yerine, şimdi koşullu olasılığı $P(B_i|A)$ bulma problemini ele aldığımızı varsayalım. Diğer bir deyişle, bir ürünün rastgele seçildiğini ve kusurlu olduğunu varsayalım. Bu ürünün B_i makinesi tarafından yapılmış olma olasılığı nedir? Bu tür sorular, Bayes kuralı adı verilen aşağıdaki teorem kullanılarak yanıtlanabilir:

TEOREM 2.14: Bayes Kuralı: $i = 1, 2, \dots, k$ için $P(B_i) = 0$ olsun. $B_1, B_2, \dots, B_k$ olayları S örneklem uzayının bir bölüntülenmesi olsun. $P(A) \neq 0$ ve A ; S örneklem uzayının herhangi bir olayı ise $r=1,2,\dots,k$ için;

$$P(B_r|A) = \frac{P(B_r \cap A)}{\sum_{i=1}^k P(B_i \cap A)} = \frac{P(B_r)P(A|B_r)}{\sum_{i=1}^k P(B_i)P(A|B_i)}$$

Bu teorem şu şekilde daha sade halde yazılabilir:

$$P(B|A) = \frac{P(B)P(A|B)}{P(A)}$$



Örnek 2.35: Örnek 2.34'e göre, eğer bir ürün rasgele seçilirse ve kusurlu bulunursa, bunun B_3 makinesi tarafından yapılmış olma olasılığı nedir?

Cevap 2.35: Bayes kuralı uygulanırsa;

$$P(B_3|A) = \frac{P(B_3)P(A|B_3)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3)},$$

$$P(B_3|A) = \frac{0.005}{0.006 + 0.0135 + 0.005} = \frac{0.005}{0.0245} = \frac{10}{49}.$$

Örnek 2.36: Bir imalat firması, belirli bir ürünün tasarımı ve geliştirilmesi için üç analitik plan kullanmaktadır. Maliyet nedenleriyle, üçü de değişen zamanlarda kullanılır. Aslında, ürünlerin sırasıyla %30, %20 ve %50'sinde plan 1, 2 ve 3 kullanılmaktadır. Kusur oranı, aşağıdaki gibi üç prosedür için farklıdır:

$$P(D|P_1) = 0.01, P(D|P_2) = 0.03, P(D|P_3) = 0.02,$$

burada $P(D|P_j)$, verilen j planındaki kusurlu ürün olasılığıdır. Rastgele bir ürün gözlemlenir ve kusurlu bulunursa, büyük olasılıkla hangi plan kullanılmış ve dolayısıyla sorumlu tutulmuştur?

Cevap 2.36: Sorunun ifadesinden

$$P(P_1) = 0.30, P(P_2) = 0.20, \text{ ve } P(P_3) = 0.50,$$

$j = 1, 2, 3$ için $P(P_j | D)$ 'yi bulmalıyız. Bayes kuralı (Teorem 2.14) şunu gösterir:

$$\begin{aligned} P(P_1|D) &= \frac{P(P_1)P(D|P_1)}{P(P_1)P(D|P_1) + P(P_2)P(D|P_2) + P(P_3)P(D|P_3)} \\ &= \frac{(0.30)(0.01)}{(0.3)(0.01) + (0.20)(0.03) + (0.50)(0.02)} = \frac{0.003}{0.019} = 0.158. \end{aligned}$$

$$P(P_2|D) = \frac{(0.03)(0.20)}{0.019} = 0.316$$

$$P(P_3|D) = \frac{(0.02)(0.50)}{0.019} = 0.526.$$

Plan 3 verilen bir kusurun koşullu olasılığı, üçünün en büyüğüdür; bu nedenle, rastgele bir ürün için kusurlu olması, büyük olasılıkla plan 3'ün kullanılmasının sonucudur.

OLASILIK VE İSTATİSTİK

4. HAFTA

RASTLANTI DEĞİŞKENLERİ VE OLASILIK DAĞILIMLARI

3.1 RASTLANTI DEĞİŞKENİ KAVRAMI

İstatistik, popülasyonlar ve popülasyon özellikleri hakkında çıkarımlar yapmakla ilgilidir. Deneyler şansa bağlı sonuçlarla yapılır. Bir dizi elektronik bileşenin test edilmesi, çeşitli şans gözlemlerinin üretildiği herhangi bir süreci tanımlamak için kullanılan bir terim olan istatistiksel deneye bir örnektir. Sonuca sayısal bir açıklama atamak genellikle önemlidir. Örneğin, üç elektronik bileşen test edildiğinde olası her sonucun ayrıntılı bir tanımını veren örnek uzay $S = \{NNN, NND, NDN, DNN, NDD, DND, DDN, DDD\}$ şeklinde yazılabilir, N, kusurlu olmadığını ve D kusurlu olduğunu belirtmektedir. Kişi doğal olarak ortaya çıkan kusurların sayısı ile ilgilenir. Böylece, örneklem uzayındaki her noktaya 0, 1, 2 veya 3 gibi sayısal bir değer atanacaktır. Bu değerler, elbette deneyin sonucuna göre belirlenen rastgele niceliklerdir. Bunlar, üç elektronik bileşen test edildiğinde arızalı öğelerin sayısı olan rastlantı değişkeni X tarafından varsayılan değerler olarak görülebilir. Yani;

S	NNN	NND	NDN	DNN	NDD	DND	DDN	DDD
Kusurlu olan	0	1	1	1	2	2	2	3

Tanım 3.1: Rastlantı değişkeni, gerçek bir sayıyı örneklem uzaydaki her elemanla ilişkilendiren bir fonksiyondur.

Bir **rastlantı değişkenini** belirtmek için büyük bir harf, **X** diyelim ve bu durumda değerlerinden biri için buna **karşılık gelen** küçük **x** harfini kullanacağız. Yukarıdaki elektronik bileşen testi örneğinde, X rastlantı değişkeninin S örnek uzayının $E = \{DDN, DND, NDD\}$ alt kümesindeki tüm öğeler için 2 değerini aldığını fark ettik. Yani, X'in her olası değeri, verilen deney için örneklem uzayın bir alt kümesi olan bir olayı temsil etmektedir.

Örnek 3.1: İçinde 4 kırmızı ve 3 siyah top bulunan bir torbadan geri konulmadan arka arkaya iki top çekiliyor. Y'nin kırmızı topların sayısı olduğu rastlantı değişkeni Y'nin olası sonuçları ve y değerleri şöyledir:

Sample Space	y
RR	2
RB	1
BR	1
BB	0

Cevap 3.1:

Örnek 3.2: Bir depo memuru, daha önce kontrol etmiş olan üç çelik fabrikası çalışanına rastgele üç koruyucu kask verir. Sinan, Jale ve Buğra bu sırayla üç şapkadan birini alırlarsa, kaskları iade etmenin olası sıraları için örnek noktaları listeleyiniz ve doğru eşleşmelerin sayısını temsil eden M rastlantı değişkeninin m değerini bulunuz.

Cevap 3.2: S, J ve B sırasıyla Sinan'ın, Jale'nin ve Buğra'nın kasklarını temsil ediyorsa, kaskların iade edilebileceği olası düzenlemeler ve doğru eşleşme sayısı:

Sample Space	m
SJB	3
SBJ	1
BJS	1
JSB	1
JBS	0
BSJ	0

Önceki iki örneğin her birinde, örneklem uzay sonlu sayıda öge içermektedir. Öte yandan, 5 gelene kadar bir zar atıldığında, sonsuz bir eleman dizisine sahip bir örnek uzay elde ederiz, $S = \{F, NF, NNF, NNNF, \dots\}$, burada F ve N sırasıyla 5'in gelip gelmediğini temsil etmektedir. Ancak bu deneyde bile, elemanların sayısı tam sayıların sayısına eşitlenebilir.

Rastlantı değişkenin doğası gereği kategorik olduğu durumlar vardır. Genellikle kukla değişkenler olarak adlandırılan değişkenler kullanılır. İyi bir örnek, aşağıdaki örnekte gösterildiği gibi, rastlantı değişkenin doğada ikili olduğu durumdur:

Örnek 3.3: Bileşenlerin üretim hattından geldiği ve kusurlu veya kusurlu olmadıklarının şart koşulduğu basit durumu göz önünde bulundurun. X rasgele değişkenini,

$$X = \begin{cases} 1, & \text{bileşen arızalıysa} \\ 0, & \text{bileşen arızalı değilse} \end{cases} \text{ olarak tanımlayın.}$$

Cevap 3.3: Açıkça 1 veya 0 ataması oldukça uygun olsa da keyfidir. Bu daha sonraki bölümlerde netleşecektir. İki olası değeri açıklamak için 0 ve 1'in seçildiği rastlantı değişkenine

Bernoulli rastlantı değişkeni denir.

Örnek 3.4: İstatistikçiler, parçaları veya çok sayıda malzemeyi kabul etmek veya reddetmek için örnekleme planlarını kullanır. Bu örnekleme planlarından birinin, 12'si kusurlu olan 100 ögeli bir parçadan 10 ögeyi bağımsız olarak örneklemeyi içerdiğini varsayalım.

Cevap 3.4: X , 10'luk örnekleme kusurlu bulunan madde sayısı olarak tanımlanan rastlantı değişkeni olsun. Bu durumda rastlantı değişkeni 0, 1, 2, . . . , 9, 10 değerlerini alır.

Örnek 3.5: Bir numune alma planının, bir kusur gözlemlenene kadar bir süreçten numune alınmasını içerdiğini varsayalım.

Cevap 3.5: Sürecin değerlendirilmesi, ardışık kaç ögenin gözlemlendiğine bağlı olacaktır. Bu bağlamda, X , bir kusur bulunmadan önce gözlemlenen ögelerin sayısı ile tanımlanan bir rastlantı değişkeni olsun. N kusurlu değil ve D kusurlu olduğunda, örnek uzaylar $X = 1$ verildiğinde $S = \{D\}$, $X = 2$ verildiğinde $S = \{ND\}$, $X = 3$ verildiğinde $S = \{NND\}$ vb.

Tanım 3.2: Bir örneklem uzayı, sonlu sayıda olasılık içeriyorsa veya tam sayılar kadar eleman içeren bitmeyen bir dizi içeriyorsa, buna kesikli (ayrık) örneklem uzayı denir.

Bazı istatistiksel deneylerin sonuçları ne sonlu ne de sayılabilir olabilir. Örneğin, belirli bir otomobil markasının 5 litre benzinle önceden belirlenmiş bir test rotasında kat edeceği mesafeleri ölçen bir araştırma yapıldığında durum böyledir. Mesafenin herhangi bir doğruluk derecesinde ölçülen bir değişken olduğunu varsayarsak, o zaman açıkça örnek uzayda tam sayıların sayısına eşitlenemeyen sonsuz sayıda olası mesafeye sahibiz. Ya da, bir kimyasal reaksiyonun meydana gelme süresinin uzunluğu kaydedilecek olsaydı, bir kez daha örnek uzayımızı oluşturan olası zaman aralıklarının sayısı sonsuz ve sayılamayacak kadar çok olurdu. Artık tüm örnek uzayların kesikli olması gerekmediğini görüyoruz.

Tanım 3.3: Bir doğru parçası üzerindeki nokta sayısına eşit sonsuz sayıda olasılık içeren bir örneklem uzayına sürekli örneklem uzayı denir.

3.2 KESİKLİ (AYRIK) OLASILIK DAĞILIMLARI

Kesikli bir rastlantı değişkeni, değerlerinin her birini belirli bir olasılıkla varsaymaktadır. Örnek 3.2'deki basit olaylar için eşit ağırlıklar alınır, hiçbir çalışanın doğru kaskı geri almama olasılığı, yani M'nin 0 değerini alma olasılığı $1/3$ 'tür. Çünkü toplam 6 tane örneklem uzayı içerisinde her birinin olasılığı $1/6$ 'dır. Bu durumda M'nin 0 değerini aldığı 2 durum vardır. Dolayısıyla olasılığı ise $1/6+1/6=1/3$ 'tür.

Sample Space	m
<i>SJB</i>	3
<i>SBJ</i>	1
<i>BJS</i>	1
<i>JSB</i>	1
<i>JBS</i>	0
<i>BSJ</i>	0

M'nin olası değerleri m ve olasılıkları;

m	0	1	3
$P(M = m)$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$

m değerlerinin tüm olası durumları içerdiğine ve dolayısıyla olasılıkların toplamının 1 olduğuna dikkat edin.

Sıklıkla, bir X rastlantı değişkeninin tüm olasılıklarını bir formülle temsil etmek mümkündür. Böyle bir formül zorunlu olarak $f(x)$, $g(x)$, $r(x)$ vb. ile göstereceğimiz x sayısal değerlerinin bir fonksiyonu olacaktır. Bu nedenle $f(x) = P(X = x)$; yani $f(3) = P(X = 3)$ şeklinde yazabiliriz. **Sıralı çiftler kümesine $(x, f(x))$ kesikli rastlantı değişkeni X'in olasılık fonksiyonu, olasılık kütle fonksiyonu veya olasılık dağılımı denir.**

Tanım 3.4: $(x, f(x))$ sıralı çiftlerinin oluşturduğu bir küme, olası her bir sonuç x için,

1.) $f(x) \geq 0$

2.) $\sum_x f(x) = 1$

3.) $P(X = x) = f(x)$

Koşullarını sağlıyorsa, kesikli rastlantı değişkeni X'in olasılık fonksiyonu veya olasılık dağılımıdır.

Örnek 3.6: Bir perakende satış noktasına 3 adet arızalı bilgisayar içeren 20 adet benzer dizüstü bilgisayar sevkiyatı yapmaktadır. Bir okul bu bilgisayarlardan rastgele 2 tane satın alırsa, kusurlu sayısı için olasılık dağılımını bulunuz.

Cevap 3.6: X, değerleri okul tarafından satın alınan arızalı bilgisayarların olası sayısı olan bir rastlantı değişkeni olsun. O halde x yalnızca 0, 1 ve 2 sayılarını alabilir. Çünkü ya okulun aldığı 2 bilgisayarın hiç biri kusurlu değildir, ya bir tanesi kusurludur ya da iki tanesi kusurludur. Dolayısıyla;

$$f(0) = P(X = 0) = \frac{\binom{3}{0}\binom{17}{2}}{\binom{20}{2}} = \frac{68}{95}, \quad f(1) = P(X = 1) = \frac{\binom{3}{1}\binom{17}{1}}{\binom{20}{2}} = \frac{51}{190},$$
$$f(2) = P(X = 2) = \frac{\binom{3}{2}\binom{17}{0}}{\binom{20}{2}} = \frac{3}{190}.$$

Böylece, X'in olasılık dağılımı şu şekildedir:

x	0	1	2
f(x)	$\frac{68}{95}$	$\frac{51}{190}$	$\frac{3}{190}$

Örnek 3.7: Bir araba acentesi, yan hava yastıkları olan belirli bir yabancı arabanın envanterinin %50'sini satarsa, yan hava yastıkları olan arabaların acente tarafından satılan sonraki 4 araba arasındaki olasılık dağılımı için bir formül bulunuz. (Not: Yani araba firmasının sattığı arabaların yarısı hava yastıklıdır. Satılacak olan 4 araba arasından hava yastıklı olanların sayısının olasılık dağılımı için bir formül bulunuz.)

Cevap 3.7: Yan hava yastıklı bir otomobilin satılma olasılığı 0.5 olduğundan, örneklem uzayında $2^4 = 16$ noktanın oluşma olasılığı eşittir. (1. Araba için 2 olasılık var ya hava yastıklı ya hava yastıksız, 2. Araba için de, 3. Araba için de ve 4. Araba için de aynı olasılık olduğundan $2*2*2*2=16$ tane örneklem uzayının sayısı olmuş olur.) Bu nedenle, tüm olasılıkların ve ayrıca fonksiyonumuzun paydası 16'dır. Yan hava yastığı olan 3 arabayı satma yollarının sayısını elde etmek için, bir hücreye yan hava yastığı olan 3 araba ve diğerine atanan yan hava yastığı olmayan model olmak üzere 4 sonucu iki hücreye ayırmanın yollarının sayısını düşünmemiz gerekir. Bu $\binom{4}{3} = 4$ şekilde yapılabilir. Genel olarak, yan hava yastıklı x modelin ve yan hava yastıksız 4-x modelin satılması olayı, x'in 0, 1, 2, 3 veya 4 olabileceği $\binom{4}{x}$ şekilde gerçekleşebilir. Böylece, olasılık dağılımı $f(x) = P(X = x)$,

$$f(x) = \frac{1}{16} \binom{4}{x} \text{ 'tir, } x = 0, 1, 2, 3, 4 \text{ için.}$$

Bir rastlantı X değişkeninin gözlenen değerinin bir x gerçek sayısından küçük veya ona eşit olma olasılığını hesaplamak isteyebileceğimiz birçok problem vardır. Her x gerçek sayısı için $F(x) = P(X \leq x)$ yazarak, **$F(x)$ 'i X rastlantı değişkeninin kümülatif (birikimli/yığılmı) dağılım fonksiyonu olarak tanımlarız.**

Tanım 3.5: Olasılık dağılımı $f(x)$ olan kesikli rastlantı değişkeni X 'in kümülatif dağılım fonksiyonu $F(x)$ şu şekildedir:

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t), -\infty < x < \infty$$

Rastlantı değişkeni M için, Örnek 3.2'deki doğru eşleşmelerin sayısı,

$$F(2) = P(M \leq 2) = f(0) + f(1) = \frac{1}{3} + \frac{1}{2} = \frac{5}{6}.$$

M 'nin kümülatif dağılım işlevi şu şekildedir:

$$F(m) = \begin{cases} 0, & \text{for } m < 0, \\ \frac{1}{3}, & \text{for } 0 \leq m < 1, \\ \frac{5}{6}, & \text{for } 1 \leq m < 3, \\ 1, & \text{for } m \geq 3. \end{cases}$$

Kümülatif dağılım fonksiyonunun, yalnızca verilen rasgele değişken tarafından varsayılan değerler için değil, tüm gerçek sayılar için tanımlanan monoton azalmayan bir fonksiyon olduğu gerçeğine özellikle dikkat edilmelidir.

Örnek 3.8: Örnek 3.7'deki rastlantı değişkeni X 'in kümülatif dağılım fonksiyonunu bulunuz. $F(x)$ kullanarak, $f(2) = 3/8$ olduğunu doğrulayınız.

Cevap 3.8: Örnek 3.7'deki olasılık dağılımının doğrudan hesaplamaları $f(0) = 1/16$, $f(1) = 1/4$, $f(2) = 3/8$, $f(3) = 1/4$ ve $f(4) = 1/16$ verir. Öyleyse,

$$F(0) = f(0) = \frac{1}{16},$$

$$F(1) = f(0) + f(1) = \frac{5}{16},$$

$$F(2) = f(0) + f(1) + f(2) = \frac{11}{16},$$

$$F(3) = f(0) + f(1) + f(2) + f(3) = \frac{15}{16},$$

$$F(4) = f(0) + f(1) + f(2) + f(3) + f(4) = 1.$$

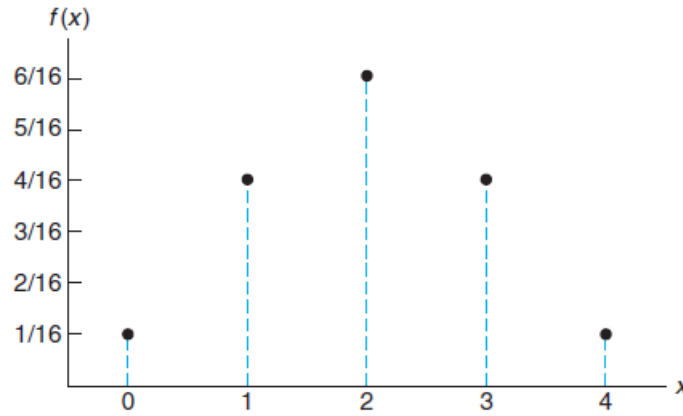
Buradan

$$F(x) = \begin{cases} 0, & \text{for } x < 0, \\ \frac{1}{16}, & \text{for } 0 \leq x < 1, \\ \frac{5}{16}, & \text{for } 1 \leq x < 2, \\ \frac{11}{16}, & \text{for } 2 \leq x < 3, \\ \frac{15}{16}, & \text{for } 3 \leq x < 4, \\ 1 & \text{for } x \geq 4. \end{cases}$$

Dolayısıyla,

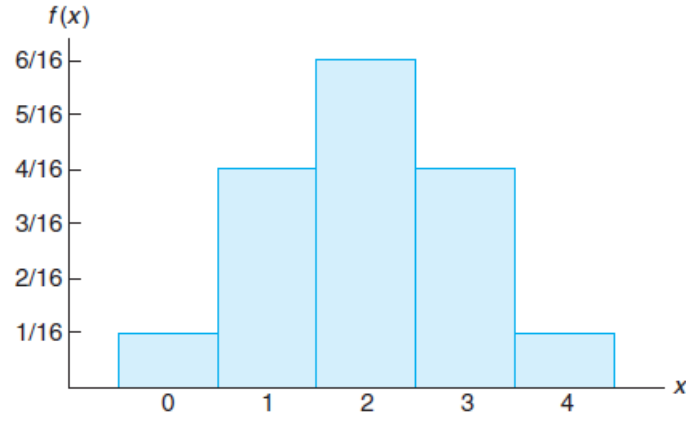
$$f(2) = F(2) - F(1) = \frac{11}{16} - \frac{5}{16} = \frac{3}{8}.$$

Bir olasılık dağılımına grafik biçiminde bakmak genellikle yararlıdır. Şekil 3.1'i elde etmek için Örnek 3.7'deki $(x, f(x))$ noktaları çizilebilir. Noktaları x eksenine kesikli veya düz bir çizgi ile birleştirerek, bir olasılık kütle fonksiyonu grafiği elde ederiz. Şekil 3.1, X 'in hangi değerlerinin ortaya çıkma olasılığının yüksek olduğunu görmeyi kolaylaştırır ve ayrıca bu durumda mükemmel bir simetrik durumu gösterir.



Şekil 3.1: Olasılık kütle fonksiyonu grafiği.

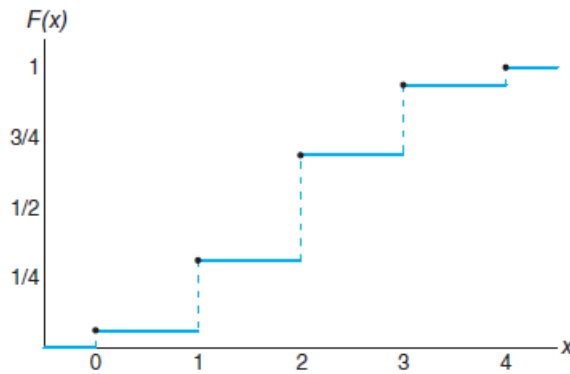
$(x, f(x))$ noktalarını çizmek yerine, Şekil 3.2'deki gibi daha sık olarak dikdörtgenler oluştururuz. Burada dikdörtgenler, eşit genişlikteki tabanları her x değerinde ortalananak ve yükseklikleri $f(x)$ tarafından verilen karşılık gelen olasılıklara eşit olacak şekilde oluşturulmuştur. Kaideler dikdörtgenler arasında boşluk kalmayacak şekilde yapılmıştır. Şekil 3.2'ye olasılık histogramı denir.



Şekil 3.2: Olasılık histogramı.

Şekil 3.2'deki her taban birim genişliğe sahip olduğundan, $P(X = x)$, x merkezli dikdörtgenin alanına eşittir. Tabanlar birim genişlikte olmasa bile, x değerlerinden herhangi birini varsayarak X 'in olasılıklarına eşit olacak alanları vermek için dikdörtgenlerin yüksekliğini ayarlayabiliriz. Olasılıkları temsil etmek için alanların kullanılması kavramı, sürekli bir rastlantı değişkenin olasılık dağılımını ele almamız için gereklidir.

Şekil 3.3'te bir adım fonksiyonu olarak görünen Örnek 3.7'deki kümülatif dağılım fonksiyonunun grafiği, $(x, F(x))$ noktalarının çizilmesiyle elde edilir.



Şekil 3.3: Kesikli kümülatif dağılım fonksiyonu.

3.3 SÜREKLİ OLASILIK DAĞILIMLARI

Bir sürekli rastlantı değişkeni, değerlerinden herhangi birini tam olarak kabul etme olasılığı 0'dır. Sonuç olarak, olasılık dağılımı tablo şeklinde verilemez. **Sürekli olasılık dağılımları parçalı fonksiyon ile gösterilmektedir.**

İlk başta bu şaşırtıcı görünebilir, ancak belirli bir örneği düşündüğümüzde daha makul hale gelir. Değerleri 21 yaştan üzerindeki tüm insanların boyları olan rastgele bir değişkeni tartışalım. Herhangi iki değer arasında, örneğin 163,5 ve 164,5 santimetre, hatta 163,99 ve 164,01 santimetre arasında, biri 164 santimetre olan sonsuz sayıda yükseklik vardır. Tam olarak 164 santimetre boyunda olan ve 164 santimetreye çok yakın olan ve aradaki farkı insani olarak ölçemeyeceğiniz kadar büyük boy kümelerinden biri olmayan bir kişiyi rastgele seçme olasılığı uzaktır ve bu nedenle olaya 0 olasılık atarız. Ancak, boyu en az 163 santimetre olan ancak 165 santimetreden fazla olmayan bir kişinin seçilme olasılığından bahsederek durum böyle değildir. Şimdi rastlantı değişkenimizin bir nokta değerinden ziyade bir aralıkla uğraşyoruz.

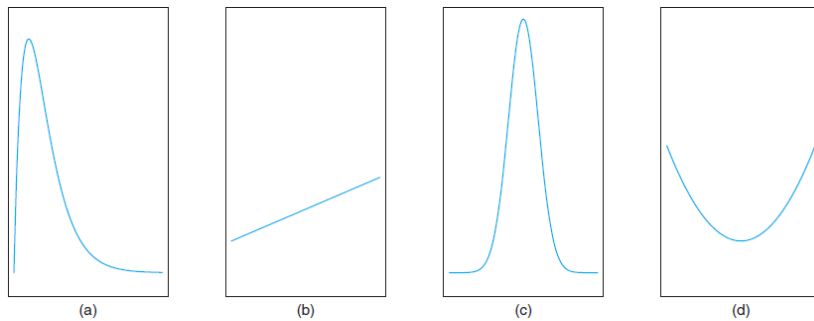
$P(a < X < b)$, $P(W \geq c)$ vb. gibi sürekli rastlantı değişkenlerinin çeşitli aralıkları için hesaplama olasılıklarıyla ilgileneceğiz.

X sürekli olduğunda,

$$P(a < X \leq b) = P(a < X < b) + P(X = b) = P(a < X < b)$$

Yani aralığın bitiş noktasını dâhil edip etmememizin bir önemi yoktur. **Ancak, X ayrık olduğunda bu doğru değildir.**

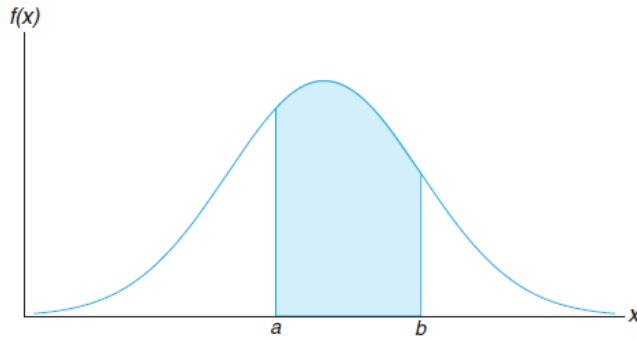
Bir sürekli rastlantı değişkeninin olasılık dağılımı tablo şeklinde gösterilemese de formül olarak ifade edilebilir. Böyle bir formül zorunlu olarak sürekli rastlantı değişkeni X'in sayısal değerlerinin bir fonksiyonu olacaktır ve bu haliyle $f(x)$ fonksiyonel notasyonu ile temsil edilecektir. **Sürekli değişkenlerle uğraşırken, $f(x)$ genellikle X'in olasılık yoğunluk fonksiyonu veya basitçe yoğunluk fonksiyonu olarak adlandırılır.** X, sürekli bir örnek uzay üzerinde tanımlandığından, $f(x)$ 'in sonlu bir sayıda süreksizliğe sahip olması mümkündür. Bununla birlikte, istatistiksel verilerin analizinde pratik uygulamaları olan yoğunluk fonksiyonlarının çoğu sürekli ve grafikleri, bazıları Şekil 3.4'te gösterilen çeşitli şekillerde olabilir. Alanlar olasılıkları temsil etmek için kullanılacağından ve olasılıklar pozitif sayısal değerler olduğundan, yoğunluk fonksiyonu tamamen x ekseninin üzerinde olmalıdır.



Şekil 3.4: Tipik yoğunluk fonksiyonları.

Bir olasılık yoğunluk fonksiyonu, eğrisinin altındaki alan $f(x)$ 'in tanımlandığı X aralığı üzerinden hesaplandığında 1'e eşit olacak şekilde oluşturulur. X 'in bu aralığı sonlu bir aralık ise, $f(x)$ 'i aralığın genişletilmiş bölümlerindeki tüm noktalarda sıfır olarak tanımlayarak aralığı tüm gerçek sayılar kümesini içerecek şekilde genişletmek her zaman mümkündür. Şekil 3.5'te, X 'in a ve b arasında bir değer alma olasılığı $x = a$ ve $x = b$ 'deki ordinatlar arasındaki yoğunluk fonksiyonu altındaki taralı alana eşittir ve integral hesabından

$$P(a < X < b) = \int_a^b f(x)dx.$$



Şekil 3.5: $P(a < X < b)$.

Tanım 3.6: $f(x)$ fonksiyonu aşağıdaki koşulları sağlıyorsa, gerçek sayılar kümesi üzerinde tanımlanan sürekli rastlantı değişkeni X için bir olasılık yoğunluk fonksiyonudur (pdf).

1. $f(x) \geq 0$, tüm $x \in R$.
2. $\int_{-\infty}^{\infty} f(x)dx = 1$.
3. $P(a < X < b) = \int_a^b f(x)dx$. (NOT: Sürekli olasılık dağılımlarındaki gösterim aralığı $\leq$ veya $\geq$ şeklinde de gösterilebilir.)

Örnek 3.9: Kontrollü bir laboratuvar deneyi için $^{\circ}C$ cinsinden reaksiyon sıcaklığındaki hatanın olasılık yoğunluk fonksiyonu

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

olan sürekli rastlantı değişkeni X olduğunu varsayalım.

- a.) $f(x)$ 'in bir yoğunluk fonksiyonu olduğunu doğrulayınız.
b.) $P(0 < X \leq 1)$ bulunuz.

Cevap 3.9: Tanım 3.6'yı kullanıyoruz.

- a.) Açıkçası, $f(x) \geq 0$. Tanım 3.6'daki koşul 2'yi doğrulamak için,

$$\int_{-\infty}^{\infty} f(x) dx = \int_{-1}^2 \frac{x^2}{3} dx = \frac{x^3}{9} \Big|_{-1}^2 = \frac{8}{9} + \frac{1}{9} = 1.$$

- b.) Tanım 3.6'daki formül 3'ü kullanarak şunu elde ederiz:

$$P(0 < X \leq 1) = \int_0^1 \frac{x^2}{3} dx = \frac{x^3}{9} \Big|_0^1 = \frac{1}{9}.$$

Tanım 3.7: Yoğunluk fonksiyonu $f(x)$ olan sürekli rastlantı değişkeni X 'in kümülatif dağılım fonksiyonu $F(x)$ şu şekildedir:

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt, \quad -\infty < x < \infty.$$

Tanım 3.7'nin hemen bir sonucu olarak, iki sonuç yazılabilir.

$$. P(a < X < b) = F(b) - F(a)$$

$$. f(x) = \frac{dF(x)}{dx} \text{ (Eğer } F(x) \text{'in türevi varsa)}$$

Örnek 3.10: Örnek 3.9'daki yoğunluk fonksiyonu için, $F(x)$ 'i bulun ve $P(0 < X \leq 1)$ 'i değerlendirmek için kullanın.

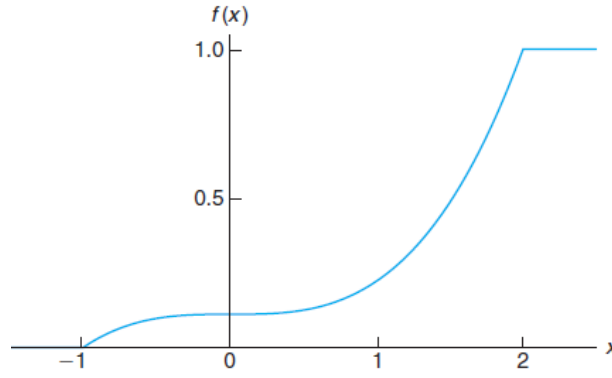
Cevap 3.10: $-1 < x < 2$ için,

$$F(x) = \int_{-\infty}^x f(t) dt = \int_{-1}^x \frac{t^2}{3} dt = \frac{t^3}{9} \Big|_{-1}^x = \frac{x^3 + 1}{9}.$$

Öyleyse,

$$F(x) = \begin{cases} 0, & x < -1, \\ \frac{x^3+1}{9}, & -1 \leq x < 2, \\ 1, & x \geq 2. \end{cases}$$

Kümülatif dağılım fonksiyonu $F(x)$, Şekil 3.6'da ifade edilmiştir.



Şekil 3.6: Sürekli kümülatif dağılım fonksiyonu.

Dolayısıyla,

$$P(0 < X \leq 1) = F(1) - F(0) = \frac{2}{9} - \frac{1}{9} = \frac{1}{9},$$

Örnek 3.11: Enerji Bakanlığı (EB) projeleri ihaleye çıkarır ve genellikle makul bir teklifin ne olması gerektiğini tahmin etmektedir. Tahmini "b" olarak adlandırınız. EB, kazanan (düşük) teklifin yoğunluk fonksiyonunun

$$f(y) = \begin{cases} \frac{5}{8b}, & \frac{2}{5}b \leq y \leq 2b, \\ 0, & \text{elsewhere.} \end{cases}$$

olduğunu belirlemiştir.

F(y)'yi bulun ve kazanan teklifin EB'nin ön tahmini "b"den düşük olma olasılığını belirlemek için kullanın.

Cevap 3.11: $\frac{2}{5}b \leq y \leq 2b$ için,

$$F(y) = \int_{\frac{2}{5}b}^y \frac{5}{8b} dy = \frac{5t}{8b} \Big|_{\frac{2}{5}b}^y = \frac{5y}{8b} - \frac{1}{4}.$$

Böylece,

$$F(y) = \begin{cases} 0, & y < \frac{2}{5}b, \\ \frac{5y}{8b} - \frac{1}{4}, & \frac{2}{5}b \leq y < 2b, \\ 1, & y \geq 2b. \end{cases}$$

Kazanan teklifin "b" ön teklif tahmininden düşük olma olasılığını belirlemek için,

$$P(Y \leq b) = F(b) = \frac{5}{8} - \frac{1}{4} = \frac{3}{8}.$$

PROBLEM 4.1: Sarıkaya Mağaralarına giren bir aracın Düzce plakalı olma olasılığı 0.12'dir; kampçı olma olasılığı 0.28; ve Düzce plakalı bir kampçı olma olasılığı ise 0.09'dur.

- a) Sarıkaya Mağaralarına giren bir kampçının Düzce plakalı olma olasılığı nedir?
- b) Sarıkaya Mağaralarına giren Düzce plakalı bir aracın kampçı olma olasılığı nedir?
- c) Sarıkaya Mağaralarına giren bir aracın Düzce plakalı olmaması veya kampçı olmaması olasılığı nedir?

PROBLEM 4.2: Bir emlakçının birkaç yeni evi müşterilerine göstermek için 8 ana anahtarı vardır. Herhangi bir evi yalnızca 1 ana anahtar açmaktadır. Bu evlerin %40'ı genellikle açık bırakılmaktadır. Bu durumda, emlakçı ofisten ayrılmadan önce yanına rastgele 3 ana anahtar alması durumunda emlakçının müşteriye göstermek için bir evi açabilme olasılığı nedir?

PROBLEM 4.3: Ülkenin belirli bir bölgesinde, 40 yaş üstü bir yetişkinin kanserli olma olasılığının 0.05 olduğu geçmiş deneyimlerden bilinmektedir. Bir doktorun kanserli bir kişiye doğru teşhis koyma olasılığı 0.78 ve kanser olmayan bir kişiye yanlış teşhis koyma olasılığı 0.06 ise, 40 yaş üstü bir yetişkine kanser teşhisi konma olasılığı kaçtır?

PROBLEM 4.4: Bir çikolata fabrikasındaki dört müfettişin, montaj hattının sonunda her bir çikolata paketinin üzerine son kullanma tarihini damgalaması gerektiğini varsayalım. Paketlerin %20'sine damga vuran Mehmet, her 200 pakette bir son kullanma tarihini damgalayamıyor; Paketlerin %60'ına damga vuran Hasan, her 100 pakette bir son kullanma tarihini damgalamay başaramıyor; Paketlerin %15'ine damga vuran Enes, her 90 pakette bir kez son kullanma tarihini damgalayamıyor ve paketlerin %5'ine damga vuran Polat, her 200 pakette bir son kullanma tarihini damgalayamıyor. Bir müşteri çikolata paketinin son kullanma tarihinin görünmediğinden şikayet ederse, bu çikolatanın Mehmet tarafından incelenmiş olma olasılığı nedir?

OLASILIK VE İSTATİSTİK

5. HAFTA

RASTLANTI DEĞİŞKENLERİ VE OLASILIK DAĞILIMLARI (DEVAM)

3.4 BİRLEŞİK (ORTAK) OLASILIK DAĞILIMLARI

Önceki bölümlerde rastlantı (rasgele) değişkenler ve bunların olasılık dağılımları üzerine yaptığımız çalışma, bir deneyin sonuçlarını tek bir rastlantı değişkeni tarafından varsayılan değerler olarak kaydettiğimiz için, tek boyutlu örnek uzaylarla sınırlıdır. Bununla birlikte, birkaç rastlantı değişkenin eşzamanlı sonuçlarını kaydetmeyi arzu edebileceğimiz durumlar olacaktır. Örneğin, kontrollü bir kimyasal deneyden salınan gazın V hacmini ve çökelti miktarını ölçerek sonuçlardan (p, v) oluşan iki boyutlu bir numune uzayı oluşturabiliriz veya sertlikle ilgilenebiliriz. Soğuk çekilmiş bakırın H ve gerilme mukavemeti T, sonuçta elde edilen sonuçlar (h, t) . Lise verilerine dayalı olarak üniversitede başarı olasılığını belirlemeye yönelik bir çalışmada, üç boyutlu bir örneklem uzayı kullanabilir ve her bireyin yetenek testi puanını, lisedeki sınıf sıralamasını ve üniversitedeki birinci sınıfın sonundaki not ortalamasını kaydedebiliriz.

X ve Y iki ayrı rastlantı değişkeni ise, bunların eşzamanlı oluşum olasılık dağılımı, X ve Y rastlantı değişkenleri aralığındaki herhangi bir değer çifti (x, y) için $f(x, y)$ değerlerine sahip bir fonksiyonla temsil edilebilir. Bu fonksiyona X ve Y'nin birleşik (ortak) olasılık dağılımı olarak atıfta bulunmak alışılmış bir durumdur.

Dolayısıyla, kesikli (ayrık) durumda

$$f(x, y) = P(X = x, Y = y);$$

Yani $f(x, y)$ değerleri, x ve y sonuçlarının aynı anda meydana gelme olasılığını verir. Örneğin, 18 tekerlekli bir aracın lastiklerinin bakımı yapılacaksa ve X bu lastiklerin kat edildiği mil sayısını ve Y değiştirilmesi gereken lastik sayısını temsil ediyorsa, o zaman $f(30000, 5)$ olasılığı: lastikler 30.000 milin üzerinde kullanılmış ve kamyonun 5 yeni lastiğe ihtiyacı var olarak nitelendirilir.

Tanım 3.8: $f(x, y)$ fonksiyonu aşağıdaki koşulları sağlıyorsa; X ve Y kesikli (ayrık) rastlantı değişkenlerinin birleşik (ortak) olasılık dağılımı veya olasılık kütle fonksiyonudur.

1. $f(x, y) \geq 0$, tüm (x, y) için,
2. $\sum_x \sum_y f(x, y) = 1$,
3. $P(X = x, Y = y) = f(x, y)$.

xy düzleminde herhangi bir A bölgesi için;

$$P[(X, Y) \in A] = \sum_A \sum f(x, y)$$

Örnek 3.12: 3 mavi, 2 kırmızı ve 3 yeşil kalemin bulunduğu bir kutudan rastgele iki tükenmez kalem seçiliyor. X, seçilen mavi kalemlerin sayısı ve Y, seçilen kırmızı kalemlerin sayısıysa,

- a.) Birleşik olasılık fonksiyonu $f(x, y)$ bulunuz,
- b.) $P[(X, Y) \in A]$ bulunuz, burada A, $\{(x, y) | x + y \leq 1\}$ bölgesidir.

Cevap 3.12: Olası değer çiftleri (x, y) ; $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$, $(0, 2)$ ve $(2, 0)$ şeklindedir.

- a.) Şimdi, örneğin $f(0, 1)$, kırmızı ve yeşil kalemlerin seçilme olasılığını temsil eder. 8 kalemde herhangi 2 kalemi seçmenin eşit olasılıklı yollarının toplam sayısı $\binom{8}{2} = 28$ 'dir. $f(0, 1)$ demek; mavi kalemlerden hiçbirini seçilemedi, 1 tane kırmızı seçildi yani 1 yeşil 1 kırmızı kalem seçilmiş demektir. Dolayısıyla; 2 kırmızı kalemde 1 kırmızıyı ve 3 yeşil kalemde 1 yeşili seçmenin yol sayısı $\binom{2}{1} * \binom{3}{1} = 6$. Sonuç olarak $f(0, 1)$ 'in olasılığı $= 6/28 = 3/14$ olmaktadır. Benzer hesaplamalar, Tablo 3.1'de sunulan diğer durumlar için olasılıkları verir. Olasılıkların toplamının 1 olduğuna dikkat edin.

$f(x, y)$		x			Row Totals
		0	1	2	
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
Column Totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

Tablo 3.1: Örnek 3.12 için Ortak Olasılık Dağılımı.

İlerleyen bölümlerde, Tablo 3.1'deki ortak olasılık dağılımının $x = 0, 1, 2$; $y = 0, 1, 2$; ve $0 \leq x + y \leq 2$ için

$$f(x, y) = \frac{\binom{3}{x} \binom{2}{y} \binom{3}{2-x-y}}{\binom{8}{2}},$$

formülüyle temsil edilebileceği bir yöntem de bulunmaktadır.

b.) (X, Y) 'nin A bölgesine düşme olasılığı;

$$\begin{aligned} P[(X, Y) \in A] &= P(X + Y \leq 1) = f(0, 0) + f(0, 1) + f(1, 0) \\ &= \frac{3}{28} + \frac{3}{14} + \frac{9}{28} = \frac{9}{14}. \end{aligned}$$

X ve Y sürekli rastlantı değişkenleri olduğunda, birleşik yoğunluk fonksiyonu $f(x, y)$ xy düzleminin üzerinde uzanan bir yüzeydir.

A , xy düzleminde herhangi bir bölgede olduğu varsayımı altında, $P[(X, Y) \in A]$ olasılığı; A bölgesinin üzerinde ve $f(x, y)$ yüzeyinin altında kalan cismin hacmine eşittir.

Tanım 3.9: $f(x, y)$ fonksiyonu aşağıdaki koşulları sağlıyorsa; X ve Y sürekli rastlantı değişkenlerinin birleşik (ortak) yoğunluk fonksiyonudur.

1. $f(x, y) \geq 0$, tüm (x, y) için,
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$,
3. $P[(X, Y) \in A] = \int \int_A f(x, y) dx dy$.

xy düzleminde herhangi bir A bölgesi için;

$$P[(X, Y) \in A] = \sum \sum_A f(x, y)$$

Örnek 3.13: Özel sektöre ait bir işletme, hem arabalı bir tesis hem de yürüyerek girilen bir tesis işletiyor. Rastgele seçilmiş bir günde, sırasıyla tesislere X ve Y arabalı ve yürüyerek girilen kullanımda olduğu zamanın oranları olsun ve bu rastlantı değişkenlerin ortak yoğunluk fonksiyonu aşağıdaki gibidir

$$f(x, y) = \begin{cases} \frac{2}{5}(2x + 3y), & 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

a.) Tanım 3.9'daki koşul 2'yi doğrulayınız.

b.) $P[(X, Y) \in A]$ bulunuz, burada A , $\{(x, y) | 0 < x < 1/2; 1/4 < y < 1/2\}$ bölgesidir.

Cevap 3.13:

a.) $f(x, y)$ 'nin bütün bölge üzerindeki integrali

$$\begin{aligned}
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dx \, dy &= \int_0^1 \int_0^1 \frac{2}{5} (2x + 3y) \, dx \, dy \\
&= \int_0^1 \left(\frac{2x^2}{5} + \frac{6xy}{5} \right) \Big|_{x=0}^{x=1} dy \\
&= \int_0^1 \left(\frac{2}{5} + \frac{6y}{5} \right) dy = \left(\frac{2y}{5} + \frac{3y^2}{5} \right) \Big|_0^1 = \frac{2}{5} + \frac{3}{5} = 1.
\end{aligned}$$

b.) Olasılığı hesaplamak için;

$$\begin{aligned}
P[(X, Y) \in A] &= P\left(0 < X < \frac{1}{2}, \frac{1}{4} < Y < \frac{1}{2}\right) \\
&= \int_{1/4}^{1/2} \int_0^{1/2} \frac{2}{5} (2x + 3y) \, dx \, dy \\
&= \int_{1/4}^{1/2} \left(\frac{2x^2}{5} + \frac{6xy}{5} \right) \Big|_{x=0}^{x=1/2} dy = \int_{1/4}^{1/2} \left(\frac{1}{10} + \frac{3y}{5} \right) dy \\
&= \left(\frac{y}{10} + \frac{3y^2}{10} \right) \Big|_{1/4}^{1/2} \\
&= \frac{1}{10} \left[\left(\frac{1}{2} + \frac{3}{4} \right) - \left(\frac{1}{4} + \frac{3}{16} \right) \right] = \frac{13}{160}.
\end{aligned}$$

Kesikli rastlantı değişkenleri X ve Y'nin birleşik olasılık dağılımı $f(x, y)$ verildiğinde, yalnızca X'in olasılık dağılımı $g(x)$, $f(x, y)$ 'nin Y değerleri üzerinden toplanmasıyla elde edilir. Benzer şekilde, tek başına Y'nin olasılık dağılımı $h(y)$, $f(x, y)$ 'nin X'in değerleri üzerinden toplanmasıyla elde edilir. $g(x)$ ve $h(y)$ 'yi sırasıyla X ve Y'nin marjinal dağılımları olarak tanımlarız.

X ve Y **sürekli rastlantı değişkenler** olduğunda, toplamalar integrallerle değiştirilir. Artık aşağıdaki genel tanımları yapabiliriz:

Tanım 3.10: Yalnızca X'in ve yalnızca Y'nin marjinal dağılımları kesikli durum için;

$$g(x) = \sum_y f(x, y) \text{ ve } h(y) = \sum_x f(x, y)$$

Yalnızca X'in ve yalnızca Y'nin marjinal dağılımları sürekli durum için;

$$g(x) = \int_{-\infty}^{\infty} f(x, y) dy \text{ ve } h(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

Burada marjinal terimi kullanılmaktadır çünkü kesikli durumda $g(x)$ ve $h(y)$ değerleri; ilgili sütun ve satırların marjinal toplamalarıdır.

Yani; bir ortak olasılık dağılımını oluşturan X ve Y rastlantı değişkenlerinin sadece kendilerinin olasılıklarından oluşan dağılıma marjinal olasılık dağılımı denir.

Örnek 3.14: Tablo 3.1'deki sütun ve satır toplamalarının yalnızca X'in ve yalnızca Y'nin marjinal dağılımını verdiğini gösteriniz.

Cevap 3.14: X rastlantı değişkeni için şunu görüyoruz:

$$g(0) = f(0, 0) + f(0, 1) + f(0, 2) = \frac{3}{28} + \frac{3}{14} + \frac{1}{28} = \frac{5}{14},$$

$$g(1) = f(1, 0) + f(1, 1) + f(1, 2) = \frac{9}{28} + \frac{3}{14} + 0 = \frac{15}{28},$$

$$g(2) = f(2, 0) + f(2, 1) + f(2, 2) = \frac{3}{28} + 0 + 0 = \frac{3}{28},$$

Yani Tablo 3.1'e bakıldığında $g(x)$ için sütundaki toplamalar baz alınmıştır. $h(y)$ için ise satırdaki toplamalar baz alınmıştır.

$f(x, y)$		x			Row Totals
		0	1	2	
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
Column Totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

Tablo 3.1: Örnek 3.12 için Ortak Olasılık Dağılımı.

Tablo biçiminde, bu marjinal dağılımlar aşağıdaki gibi yazılabilir:

x	0	1	2
$g(x)$	$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$

y	0	1	2
$h(y)$	$\frac{15}{28}$	$\frac{3}{7}$	$\frac{1}{28}$

Örnek 3.15: Örnek 3.13'teki birleşik yoğunluk fonksiyonu için $g(x)$ ve $h(y)$ 'yi bulunuz.

Cevap 3.15: Tanımdan,

$$g(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_0^1 \frac{2}{5}(2x + 3y) dy = \left(\frac{4xy}{5} + \frac{6y^2}{10} \right) \Big|_{y=0}^{y=1} = \frac{4x + 3}{5}, \quad 0 \leq x \leq 1 \text{ için.}$$

$$h(y) = \int_{-\infty}^{\infty} f(x, y) dx = \int_0^1 \frac{2}{5}(2x + 3y) dx = \frac{2(1 + 3y)}{5}, \quad 0 \leq y \leq 1 \text{ için.}$$

Tanım 3.11: X ve Y, kesikli veya sürekli iki rastlantı değişkeni olsun. $X = x$ olduğu göz önüne alındığında Y rastlantı değişkeninin koşullu dağılımı;

$$f(y|x) = \frac{f(x,y)}{g(x)}, g(x) > 0$$

Benzer şekilde, $Y = y$ verildiğinde X'in koşullu dağılımı;

$$f(x|y) = \frac{f(x,y)}{h(y)}, h(y) > 0$$

Kesikli rastlantı değişkeni X'in, kesikli değişken $Y = y$ olduğu bilindiğinde, a ile b arasına düşme olasılığını bulmak istersek, şunu değerlendiririz:

$$P(a < X < b | Y = y) = \sum_{a < x < b} f(x|y),$$

burada toplam, X'in a ve b arasındaki tüm değerleri üzerinde uzanır. X ve Y sürekli olduğunda,

$$P(a < X < b | Y = y) = \int_a^b f(x|y) dx.$$

Örnek 3.16: Örnek 3.12'ye bakarak, $Y = 1$ verildiğinde X'in koşullu dağılımını bulunuz ve bunu $P(X = 0 | Y = 1)$ belirlemek için kullanınız.

Cevap 3.16: $y = 1$ olmak üzere $f(x|y)$ 'yi bulmamız gerekiyor. İlk önce şunu buluyoruz:

$$h(1) = \sum_{x=0}^2 f(x, 1) = \frac{3}{14} + \frac{3}{14} + 0 = \frac{3}{7}.$$

Sonra

$$f(x|1) = \frac{f(x, 1)}{h(1)} = \left(\frac{7}{3}\right) f(x, 1), \quad x = 0, 1, 2.$$

$f(x, y)$		x			Row Totals
		0	1	2	
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
Column Totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

Tablo 3.1: Örnek 3.12 için Ortak Olasılık Dağılımı.

Böylece

$$f(0|1) = \left(\frac{7}{3}\right) f(0,1) = \left(\frac{7}{3}\right) \left(\frac{3}{14}\right) = \frac{1}{2}, \quad f(1|1) = \left(\frac{7}{3}\right) f(1,1) = \left(\frac{7}{3}\right) \left(\frac{3}{14}\right) = \frac{1}{2},$$

$$f(2|1) = \left(\frac{7}{3}\right) f(2,1) = \left(\frac{7}{3}\right) (0) = 0,$$

ve X'in koşullu dağılımı, $Y = 1$ olduğu göz önüne alındığında,

x	0	1	2
$f(x 1)$	$\frac{1}{2}$	$\frac{1}{2}$	0

Sonuç olarak

$$P(X = 0 | Y = 1) = f(0|1) = \frac{1}{2}.$$

Bu nedenle, seçilen 2 kalem yedeğinden 1'inin kırmızı olduğu biliniyorsa, diğerinin mavi olmama olasılığı 1/2'ye eşittir.

Örnek 3.17: X'in birim sıcaklık değişimi ve Y'nin belirli bir atomik parçacığın ürettiği spektrum kaymasının oranı olduğu rasgele değişkenler (X, Y) için ortak yoğunluk

$$f(x, y) = \begin{cases} 10xy^2, & 0 < x < y < 1, \\ 0, & \text{elsewhere.} \end{cases} \quad \text{'dir.}$$

- Marjinal yoğunlukları $g(x)$, $h(y)$ ve koşullu yoğunluğu $f(y|x)$ bulunuz.
- Sıcaklık 0.25 birim arttığında, spektrumun toplam gözlemlerin yarısından fazla (1/2'den büyük) kayma olasılığını bulunuz.

Cevap 3.17: a.) Tanımdan

$$g(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_x^1 10xy^2 dy$$

$$= \frac{10}{3} xy^3 \Big|_{y=x}^{y=1} = \frac{10}{3} x(1 - x^3), \quad 0 < x < 1,$$

$$h(y) = \int_{-\infty}^{\infty} f(x, y) dx = \int_0^y 10xy^2 dx = 5x^2 y^2 \Big|_{x=0}^{x=y} = 5y^4, \quad 0 < y < 1.$$

Dolayısıyla

$$f(y|x) = \frac{f(x, y)}{g(x)} = \frac{10xy^2}{\frac{10}{3} x(1 - x^3)} = \frac{3y^2}{1 - x^3}, \quad 0 < x < y < 1.$$

b.) Böylece

$$P\left(Y > \frac{1}{2} \mid X = 0.25\right) = \int_{1/2}^1 f(y \mid x = 0.25) dy = \int_{1/2}^1 \frac{3y^2}{1 - 0.25^3} dy = \frac{8}{9}.$$

Örnek 3.18: Birleşik (ortak) yoğunluk fonksiyonu

$$f(x, y) = \begin{cases} \frac{x(1+3y^2)}{4}, & 0 < x < 2, 0 < y < 1, \\ 0, & \text{elsewhere,} \end{cases}$$

verildiğinde $g(x)$, $h(y)$, $f(x|y)$ 'yi bulunuz ve

$P(1/4 < X < 1/2 \mid Y = 1/3)$ değerini hesaplayınız.

Cevap 3.18: a.) Marjinal yoğunluğun tanımı gereği

$$\begin{aligned} g(x) &= \int_{-\infty}^{\infty} f(x, y) dy = \int_0^1 \frac{x(1+3y^2)}{4} dy \\ &= \left(\frac{xy}{4} + \frac{xy^3}{4} \right) \Big|_{y=0}^{y=1} = \frac{x}{2}, \end{aligned}$$

$0 < y < 1$ için.

$$\begin{aligned} h(y) &= \int_{-\infty}^{\infty} f(x, y) dx = \int_0^2 \frac{x(1+3y^2)}{4} dx \\ &= \left(\frac{x^2}{8} + \frac{3x^2y^2}{8} \right) \Big|_{x=0}^{x=2} = \frac{1+3y^2}{2}. \end{aligned}$$

$0 < x < 2$ için.

Bu nedenle, koşullu yoğunluk tanımını kullanarak

$$f(x|y) = \frac{f(x, y)}{h(y)} = \frac{x(1+3y^2)/4}{(1+3y^2)/2} = \frac{x}{2},$$

$$P\left(\frac{1}{4} < X < \frac{1}{2} \mid Y = \frac{1}{3}\right) = \int_{1/4}^{1/2} \frac{x}{2} dx = \frac{3}{64}.$$

3.5 İSTATİSTİKSEL BAĞIMSIZLIK

Tanım 3.12: X ve Y, sırasıyla birleşik olasılık dağılımı $f(x, y)$ ve marjinal dağılımları $g(x)$ ve $h(y)$ olan, kesikli veya sürekli iki rastlantı değişkeni olsun. X ve Y rastlantı değişkenlerinin, ancak ve ancak, aralıklarındaki tüm (x, y) için

$$f(x, y) = g(x) \cdot h(y)$$

olması durumunda istatistiksel olarak bağımsız olduğu söylenir.

Örnek 3.18'deki sürekli rastlantı değişkenleri istatistiksel olarak bağımsızdır, çünkü iki marjinal dağılımın çarpımı ortak yoğunluk fonksiyonunu verir. Bununla birlikte, Örnek 3.17'deki sürekli değişkenler için durum açıkça böyle değildir. Kesikli rastlantı değişkenlerinin istatistiksel bağımsızlığını kontrol etmek, (x, y) 'nin tüm kombinasyonları için olmasa da bazı kombinasyonları için ortak olasılık dağılımına eşit marjinal dağılımların çarpımına sahip olmak mümkün olduğundan, daha kapsamlı bir araştırma gerektirir. $f(x, y) \neq g(x) \cdot h(y)$ olacak şekilde $f(x, y)$ 'nin tanımlandığı herhangi bir nokta (x, y) bulabilirsiniz, X ve Y kesikli değişkenleri istatistiksel olarak bağımsız değildir.

Örnek 3.19: Örnek 3.12'deki rastlantı değişkenlerinin istatistiksel olarak bağımsız olmadığını gösteriniz.

Cevap 3.19: $(0, 1)$ noktasını ele alalım. Tablo 3.1'den

$f(x, y)$		x			Row
		0	1	2	Totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
Column Totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

Tablo 3.1: Örnek 3.12 için Ortak Olasılık Dağılımı

$f(0, 1)$, $g(0)$ ve $h(1)$ üç olasılığı;

$$f(0, 1) = \frac{3}{14},$$

$$g(0) = \sum_{y=0}^2 f(0, y) = \frac{3}{28} + \frac{3}{14} + \frac{1}{28} = \frac{5}{14},$$

$$h(1) = \sum_{x=0}^2 f(x, 1) = \frac{3}{14} + \frac{3}{14} + 0 = \frac{3}{7}.$$

Açıkça,

$$f(0, 1) \neq g(0)h(1), \text{ ve bu nedenle } X \text{ ve } Y \text{ istatistiksel olarak bağımsız değildir.}$$

İki rastlantı değişkeniyle ilgili önceki tüm tanımlar, n rastlantı değişkeni durumuna genelleştirilebilir. $f(x_1, x_2, \dots, x_n)$, $X_1, X_2, \dots, X_n$ rastlantı değişkenlerinin ortak olasılık fonksiyonu olsun. X_1 'in marjinal dağılımı, örneğin, kesikli durum için

$$g(x_1) = \sum_{x_2} \cdots \sum_{x_n} f(x_1, x_2, \dots, x_n)$$

ve sürekli durum için

$$g(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_n) dx_2 dx_3 \cdots dx_n$$

şeklindedir.

Artık $g(x_1, x_2)$ gibi ortak marjinal dağılımlar elde edebiliriz, burada

$$g(x_1, x_2) = \begin{cases} \sum_{x_3} \cdots \sum_{x_n} f(x_1, x_2, \dots, x_n) & (\text{discrete case}), \\ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_n) dx_3 dx_4 \cdots dx_n & (\text{continuous case}). \end{cases}$$

Çok sayıda koşullu dağılımı düşünebiliriz. Örneğin, $X_4 = x_4, X_5 = x_5$, ve $X_n = x_n$ verildiğinde X_1, X_2, X_3 'ün birleşik koşullu dağılımı,

$$f(x_1, x_2, x_3 \mid x_4, x_5, \dots, x_n) = \frac{f(x_1, x_2, \dots, x_n)}{g(x_4, x_5, \dots, x_n)},$$

olarak yazılır. Burada $g(x_4, x_5, \dots, x_n)$,

X_4, X_5, X_n rastlantı değişkenlerinin ortak marjinal dağılımıdır.

Tanım 3.12'nin genelleştirilmesi, X_1, X_2 ve X_n değişkenlerinin karşılıklı istatistiksel bağımsızlığı için aşağıdaki tanıma götürür.

Tanım 3.13: $X_1, X_2, \dots, X_n$, birleşik olasılık dağılımı $f(x_1, x_2, \dots, x_n)$, ve marjinal dağılımları $f_1(x_1), f_2(x_2), \dots, f_n(x_n)$ olan kesikli veya sürekli n rastlantı değişkeni olsun.

Rastlantı değişkenleri $X_1, X_2, \dots, X_n$ tanımlanan aralıklarındaki bütün $(x_1, x_2, \dots, x_n)$ değerleri için ancak ve ancak

$$f(x_1, x_2, \dots, x_n) = f_1(x_1) * f_2(x_2) * f_3(x_3) \cdots f_n(x_n)$$

ise birbirleriyle (ikili olarak/karşılıklı) istatistiksel olarak bağımsız olduğu söylenir,

Örnek 3.20: Karton kaplarda paketlenmiş bozulabilir bir gıda ürününün yıl cinsinden raf ömrünün, olasılık yoğunluk fonksiyonu

$$f(x) = \begin{cases} e^{-x}, & x > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

ile verilen bir rastlantı değişkeni olduğunu varsayalım. X_1, X_2 ve X_3 bu kaplardan bağımsız olarak seçilen üçünün raf ömürlerini temsil etmek üzere

$P(X_1 < 2, 1 < X_2 < 3, X_3 > 2)$ 'yi bulunuz.

Cevap 3.20: Kaplar bağımsız olarak seçildiğinden, X_1, X_2 ve X_3 rastlantı değişkenlerinin istatistiksel olarak bağımsız olduğunu ve ortak olasılık yoğunluğuna sahip olduğunu varsayabiliriz.

$$f(x_1, x_2, x_3) = f(x_1)f(x_2)f(x_3) = e^{-x_1}e^{-x_2}e^{-x_3} = e^{-x_1-x_2-x_3}, \quad x_1 > 0, x_2 > 0, x_3 > 0, \text{ için.}$$

$f(x_1, x_2, x_3) = 0$ diğer durumlar için. Böylece

$$\begin{aligned} P(X_1 < 2, 1 < X_2 < 3, X_3 > 2) &= \int_2^\infty \int_1^3 \int_0^2 e^{-x_1-x_2-x_3} dx_1 dx_2 dx_3 \\ &= (1 - e^{-2})(e^{-1} - e^{-3})e^{-2} = 0.0372. \end{aligned}$$

PROBLEM 5.1: 7 televizyonluk bir sevkiyatta 2 kusurlu televizyon vardır. Bir otel rastgele 3 televizyon satın alıyor. Otel tarafından satın alınan kusurluların sayısı x ise, X 'in olasılık dağılımını bulunuz. Sonuçları bir olasılık histogramı olarak grafiksel olarak ifade ediniz.

PROBLEM 5.2: Sürekli bir rasgele değişken X ; $x = 1$ ile $x = 3$ arasında değerler alabilen, $f(x)=1/2$ ile verilen bir yoğunluk fonksiyonuna sahiptir.

(a) Eğrinin altındaki alanın 1'e eşit olduğunu gösterin.

(b) $P(2 < X < 2.5)$ 'i bulunuz.

(c) $P(X \leq 1.6)$ 'yi bulunuz.

PROBLEM 5.3: X ve Y'nin ortak olasılık dağılımı şu şekilde verilirse:

$$f(x, y) = \frac{x + y}{30}, \quad \text{for } x = 0, 1, 2, 3; y = 0, 1, 2;$$

- a.) $P(X \leq 2, Y = 1)$
- b.) $P(X > 2, Y \leq 1)$
- c.) $P(X > Y)$
- d.) $P(X + Y = 4)$

PROBLEM 5.4: X, belirli bir sayısal kontrol makinesinin arızalanma sayısını gösterebilir: herhangi bir günde 1, 2 veya 3 kez. Y, bir acil durum çağrısında bir teknisyenin aranma sayısını gösterebilir. Ortak olasılık dağılımları şu şekilde verilmektedir:

$f(x, y)$		x		
		1	2	3
y	1	0.05	0.05	0.10
	3	0.05	0.10	0.35
	5	0.00	0.20	0.10

- (a) X'in marjinal dağılımını hesaplayınız.
- (b) Y'nin marjinal dağılımını değerlendiriniz.
- c) $P(Y = 3 | X = 2)$ 'yi bulunuz.

OLASILIK VE İSTATİSTİK

6. HAFTA

MATEMATİKSEL BEKLENTİ

4.1 BİR RASTLANTI DEĞİŞKENİNİN ORTALAMASI

İki madeni para 16 kez atılıyorsa ve X , her atışta oluşan tura sayısı ise, o zaman X 'in değerleri 0, 1 ve 2'dir. Deneyin sırasıyla 4, 7 ve 5 kez olmak üzere sırasıyla hiç tura gelmediğini, bir tura ve iki tura geldiğini varsayalım. İki madeni paranın atış başına ortalama tura sayısı (X rastlantı değişkeninin ortalaması/ X 'in olasılık dağılımının ortalaması),

$$\frac{(0)(4) + (1)(7) + (2)(5)}{16} = 1.06.$$

Bu, verilerin ortalama bir değeridir ve yine de $\{0, 1, 2\}$ 'nin olası bir sonucu değildir. Bu nedenle, bir ortalama, deney için mutlaka olası bir sonuç değildir. Örneğin, bir satıcının ortalama aylık geliri muhtemelen aylık maaş çeklerinden herhangi birine eşit olmayacaktır.

Şimdi ortalama tura sayısı hesaplamamızı aşağıdaki eşdeğer forma sahip olacak şekilde yeniden yapılandıralım:

$$(0) \left(\frac{4}{16} \right) + (1) \left(\frac{7}{16} \right) + (2) \left(\frac{5}{16} \right) = 1.06.$$

4/16, 7/16 ve 5/16 sayıları, sırasıyla 0,1 ve 2 tura ile sonuçlanan toplam atışların kesirleridir. Bu kesirler aynı zamanda deneyimizdeki farklı X değerleri için göreceli frekanslardır. Aslında o zaman, veri kümemizdeki toplam gözlem sayısını bilmeden meydana gelen farklı değerleri ve bunların göreceli sıklıklarını bilerek bir veri kümesinin ortalamasını veya ortalamasını hesaplayabiliriz.

Bu nedenle, atışların 4/16'sı hiç tura ile sonuçlanmazsa, atışların 7/16'sı bir tura ile sonuçlanırsa ve atışların 5/16'sı iki tura ile sonuçlanırsa, atış başına ortalama tura sayısı toplam atış sayısı 16, 1000 ve hatta 10.000 olsun, 1.06 olacaktır.

Bu bağıl frekans yöntemi, uzun vadede bekleyebileceğimiz iki madeni paranın her atışında ortalama tura sayısını hesaplamak için kullanılır. **Bu ortalama değere X rastlantı değişkeninin ortalaması veya X 'in olasılık dağılımının ortalaması olarak atıfta bulunacağız** ve hangi rastlantı değişkene atıfta bulunduğumuz açık olduğunda bunu μ_x veya

basitçe μ olarak yazacağız. İstatistikçiler arasında bu ortalamaya matematiksel beklenti veya X rastlantı değişkeninin beklenen değeri olarak atıfta bulunmak ve onu $E(X)$ olarak belirtmek de yaygındır.

Bir hilesiz madeni paranın iki kez atıldığını varsayarsak, deneyimizin örnek uzayının $S = \{TT, TY, YT, YY\}$ olduğunu buluruz. 4 örnek noktasının tümü eşit derecede olası olduğundan,

$$P(X = 0) = P(YY) = 1/4, P(X = 1) = P(YT) + P(TY) = 1/2 \text{ ve } P(X = 2) = P(TT) = 1/4,$$

burada tipik bir öge, diyelim ki YT, birinci atışın bir yazı ve ardından ikinci atışta bir tura ile sonuçlandığını göstermektedir. Şimdi, bu olasılıklar, uzun vadede verilen olayların göreceli frekanslarıdır. Bu nedenle,

$$\mu = E(X) = (0) \left(\frac{1}{4}\right) + (1) \left(\frac{1}{2}\right) + (2) \left(\frac{1}{4}\right) = 1.$$

Bu sonuç, arka arkaya 2 madeni para atan bir kişinin ortalama olarak her atışta 1 tura alacağı anlamına gelir.

2 madeni paranın her atışında beklenen tura sayısını hesaplamak için yukarıda açıklanan yöntem, herhangi bir ayırık rasgele değişkenin ortalama veya beklenen değerinin, X rastlantı değişkeninin $x_1, x_2, \dots, x_n$ değerlerinin her birinin karşılık gelen olasılık $f(x_1), f(x_2), \dots, f(x_n)$ ile çarpılmasıyla ve bunların toplanmasıyla elde edilebileceğini önerir. Ancak bu, yalnızca rastlantı değişkeni kesikli ise doğrudur. Sürekli rastlantı değişkenler söz konusu olduğunda, beklenen bir değer tanımı, toplam yerine integral alınmasıyla sonuçlanmaktadır.

Tanım 4.1: X, $f(x)$, olasılık dağılımına sahip bir rastlantı değişkeni olsun. X'in ortalama veya beklenen değeri

X kesikli ise;

$$\mu = E(x) = \sum_x x f(x)$$

X sürekli ise;

$$\mu = E(x) = \int_{-\infty}^{\infty} x f(x) dx$$

Matematiksel beklentide, beklenen değer olasılık dağılımı kullanılarak hesaplanır.

Örnek 4.1: 4 sağlam, 3 kusurlu ürün içeren bir parti maldan, kalite kontrol görevlisi 3 tanesini test etmek için alıyor. Seçilen ürünler içerisinde sağlam olanlarının sayısının beklenen değerini bulunuz.

Cevap 4.1: X, numunedeki sağlam olanların sayısını temsil etsin. X'in olasılık dağılımı.

$$f(x) = \frac{\binom{4}{x} \binom{3}{3-x}}{\binom{7}{3}}, \quad x = 0, 1, 2, 3.$$

Hesaplamalar sonunda $f(0) = 1/35$, $f(1) = 12/35$, $f(2) = 18/35$ ve $f(3) = 4/35$ verir. Öyleyse,

$$\mu = E(X) = (0) \left(\frac{1}{35} \right) + (1) \left(\frac{12}{35} \right) + (2) \left(\frac{18}{35} \right) + (3) \left(\frac{4}{35} \right) = \frac{12}{7} = 1.7.$$

Yani 4 sağlam, 3 kusurlu ürün içeren bir parti maldan rastgele seçilen 3 tanesinden 1.7 tanesi sağlam olacaktır.

Örnek 4.2: Bir tıbbi cihaz şirketinin satış görevlisinin belirli bir günde iki randevusu vardır. İlk randevuda, başarılı olursa 1000\$ komisyon kazanabileceği anlaşmayı yapmak için %70 şansı olduğuna inanıyor. Öte yandan, ikinci randevuda anlaşma yapmak için yalnızca %40 şansı olduğunu ve başarılı olursa 1500\$ kazanabileceğini düşünüyor. Kendi olasılık tahminine göre beklenen komisyon değeri nedir? Randevu sonuçlarının birbirinden bağımsız olduğunu varsayalım.

Cevap 4.2: İlk olarak, satış görevlisinin iki randevu için 4 olası komisyon toplamına sahip olabileceğini biliyoruz: 0\$, 1000\$, 1500\$ ve 2500\$. Daha sonra ilişkili olasılıklarını hesaplamamız gerekir. Bağımsızlıkla,

$$\begin{aligned} f(\$0) &= (1 - 0.7)(1 - 0.4) = 0.18, & f(\$2500) &= (0.7)(0.4) = 0.28, \\ f(\$1000) &= (0.7)(1 - 0.4) = 0.42, & \text{and } f(\$1500) &= (1 - 0.7)(0.4) = 0.12. \end{aligned}$$

elde ederiz.

Bu nedenle, satış elemanı için komisyonun beklenen değeri

$$\begin{aligned} E(X) &= (\$0)(0.18) + (\$1000)(0.42) + (\$1500)(0.12) + (\$2500)(0.28) \\ &= \$1300. \end{aligned}$$

Örnek 4.1 ve 4.2, okuyucunun bir rastlantı değişkeninin beklenen değeri ile ne demek istediğimizi anlamasına olanak sağlamak için tasarlanmıştır. Her iki durumda da rastlantı değişkenleri kesiklidir. Bir mühendisin belirli bir tür elektronik aygıtın ortalama ömrüyle

ilgilendiği, sürekli bir rastlantı değişkeni içeren bir örnekle devam edeceğiz. Bu, pratikte sıklıkla ortaya çıkan bir arızaya kadar geçen süre probleminin bir örneğidir. Bir cihazın ömrünün beklenen değeri, onun değerlendirilmesi için önemli bir parametredir.

Örnek 4.3: Belirli bir elektronik cihazın ömrünü saat cinsinden ifade eden rastlantı değişkeni X olsun. Olasılık yoğunluk fonksiyonu

$$f(x) = \begin{cases} \frac{20,000}{x^3}, & x > 100, \\ 0, & \text{elsewhere.} \end{cases}$$

'tir. Bu tür bir cihazın beklenen ömrünü bulunuz.

Cevap 4.3: Tanım 4.1'i kullanarak,

$$\mu = E(X) = \int_{100}^{\infty} x \frac{20,000}{x^3} dx = \int_{100}^{\infty} \frac{20,000}{x^2} dx = 200.$$

Bu nedenle, bu tür bir cihazın ortalama 200 saat dayanmasını bekleyebiliriz.

Şimdi X 'e bağlı yeni bir rastlantı değişkeni $g(X)$ düşünelim; yani, $g(X)$ 'in her değeri, X 'in değeri tarafından belirlenir. Örneğin, $g(X)$, X^2 veya $3X-1$ olabilir. X , 2 değerini aldığı anda, $g(X)$, $g(2)$ olduğu varsayılmaktadır. Özellikle, X ; $x = -1, 0, 1, 2$ ve $g(X) = X^2$ için $f(x)$ olasılık dağılımına sahip kesikli bir rastlantı değişkeni ise, o zaman

$$\begin{aligned} P[g(X) = 0] &= P(X = 0) = f(0), \\ P[g(X) = 1] &= P(X = -1) + P(X = 1) = f(-1) + f(1), \\ P[g(X) = 4] &= P(X = 2) = f(2), \end{aligned}$$

ve böylece $g(X)$ 'in olasılık dağılımı

$g(x)$	0	1	4
$P[g(X) = g(x)]$	$f(0)$	$f(-1) + f(1)$	$f(2)$

yazılabilir.

Bir rastlantı değişkenin beklenen değerinin tanımıyla şunu elde ederiz:

$$\begin{aligned} \mu_{g(X)} &= E[g(x)] = 0f(0) + 1[f(-1) + f(1)] + 4f(2) \\ &= (-1)^2 f(-1) + (0)^2 f(0) + (1)^2 f(1) + (2)^2 f(2) = \sum_x g(x)f(x). \end{aligned}$$

Bu sonuç Teorem 4.1'de hem kesikli hem de sürekli rastgele değişkenler için genelleştirilmiştir.

TEOREM 4.1: $X, f(x)$, olasılık dağılımına sahip bir rastlantı değişkeni olsun. X kesikli ise, rastlantı değişkeni $g(x)$ 'in beklenen değeri;

$$\mu_{g(x)} = E[g(x)] = \sum_x g(x)f(x)$$

X sürekli ise;

$$\mu_{g(x)} = E[g(x)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

Örnek 4.4: Herhangi bir güneşli Cuma günü saat 16:00 ile 17:00 arasında araba yıkamasından geçen arabalarının sayısı X 'in aşağıdaki olasılık dağılımına sahip olduğunu varsayalım:

x	4	5	6	7	8	9
$P(X = x)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

$g(X) = 2X - 1$, yönetici tarafından görevliye ödenen para miktarını dolar olarak gösterebilir. Görevlinin bu belirli zaman dilimi için beklenen kazancını bulunuz.

Cevap 4.4: Teorem 4.1'e göre,

$$\begin{aligned} E[g(X)] &= E(2X - 1) = \sum_{x=4}^9 (2x - 1)f(x) \\ &= (7) \left(\frac{1}{12} \right) + (9) \left(\frac{1}{12} \right) + (11) \left(\frac{1}{4} \right) + (13) \left(\frac{1}{4} \right) \\ &\quad + (15) \left(\frac{1}{6} \right) + (17) \left(\frac{1}{6} \right) = \$12.67. \end{aligned}$$

Örnek 4.5: X ,

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

yoğunluk fonksiyonuna sahip rastlantı değişkeni olsun.

$g(X) = 4X + 3$ 'ün beklenen değerini bulunuz.

Cevap 4.5: Teorem 4.1'e göre

$$E(4X + 3) = \int_{-1}^2 \frac{(4x + 3)x^2}{3} dx = \frac{1}{3} \int_{-1}^2 (4x^3 + 3x^2) dx = 8.$$

Şimdi, matematiksel beklenti kavramımızı, $f(x, y)$ birleşik olasılık dağılımına sahip iki rastlantı değişkeni X ve Y durumuna genişleteceğiz.

Tanım 4.2: X ve Y birleşik olasılık dağılımı $f(x, y)$ olan rastlantı değişkenleri olsun.

$g(x, y)$ rastlantı değişkeninin ortalaması veya beklenen değeri

X ve Y kesikli ise;

$$\mu_{g(X,Y)} = E[g(X,Y)] = \sum_x \sum_y g(x,y)f(x,y)$$

X ve Y sürekli ise;

$$\mu_{g(X,Y)} = E[g(X,Y)] = \iint_{-\infty}^{\infty} g(x,y)f(x,y)dxdy$$

Çeşitli rastlantı değişkenli fonksiyonların matematiksel beklentilerinin hesaplanması için Tanım 4.2'nin genelleştirilmesi basittir.

Örnek 4.6: X ve Y , 3. Haftadaki Tablo 3.1'de belirtilen birleşik olasılık dağılımına sahip rastlantı değişkenleri olsun. $g(X,Y) = X*Y$ 'nin beklenen değerini bulunuz. (Tablo, kolaylık sağlamak için burada yeniden gösterilmiştir.)

$f(x, y)$		x			Row
		0	1	2	Totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
Column Totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

Tablo 3.1: Örnek 3.12 için Ortak Olasılık Dağılımı

Cevap 4.6: Teorem 4.2'ye göre,

$$\begin{aligned}
 E(XY) &= \sum_{x=0}^2 \sum_{y=0}^2 xyf(x,y) \\
 &= (0)(0)f(0,0) + (0)(1)f(0,1) \\
 &\quad + (1)(0)f(1,0) + (1)(1)f(1,1) + (2)(0)f(2,0) \\
 &= f(1,1) = \frac{3}{14}.
 \end{aligned}$$

Örnek 4.7: Aşağıdaki yoğunluk fonksiyonu

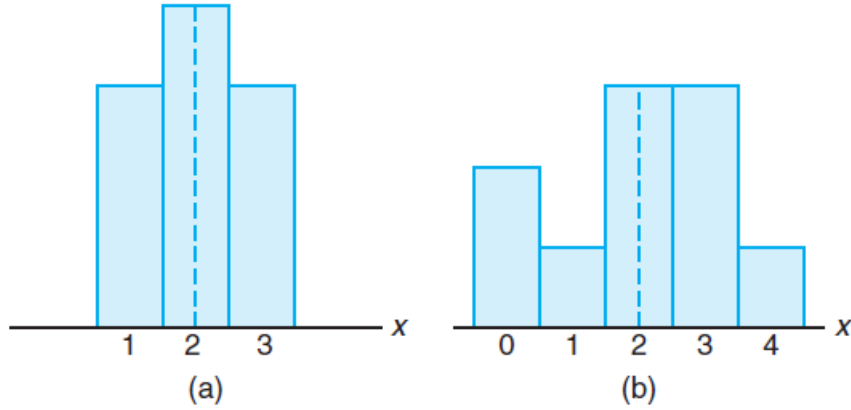
$$f(x, y) = \begin{cases} \frac{x(1+3y^2)}{4}, & 0 < x < 2, 0 < y < 1, \\ 0, & \text{elsewhere.} \end{cases} \quad \text{için } E\left(\frac{Y}{X}\right)\text{'i bulunuz.}$$

Cevap 4.7: Teorem 4.2'ye göre,

$$E\left(\frac{Y}{X}\right) = \int_0^1 \int_0^2 \frac{y(1+3y^2)}{4} dx dy = \int_0^1 \frac{y+3y^3}{2} dy = \frac{5}{8}.$$

4.2 RASTLANTI DEĞİŞKENLERİNİN VARYANSI VE KOVARYANSI

Rastgele bir X değişkeninin ortalama veya beklenen değeri, olasılık dağılımının nerede merkezlendiğini açıkladığı için istatistikte özel bir öneme sahiptir. Bununla birlikte, tek başına ortalama, dağılımın şeklinin yeterli bir tanımını vermez. Ayrıca dağılımdaki değişkenliği karakterize etmemiz gerekiyor. Şekil 4.1'de, aynı ortalamaya ($\mu = 2$) sahip, ancak değişkenlik veya ortalama hakkındaki gözlemlerinin dağılımı bakımından önemli ölçüde farklılık gösteren iki ayrı olasılık dağılımının histogramları gösterilmiştir.



Şekil 4.1: Eşit ortalamalı dağılımlar ve eşit olmayan dağılımlar.

Bir rastlantı X değişkeninin değişkenliğinin en önemli ölçüsü, Teorem 4.1'in

$g(X) = (X - \mu)^2$ ile uygulanmasıyla elde edilir. **Sonuç, X rastlantı değişkeninin varyansı veya X'in olasılık dağılımının varyansı olarak adlandırılır.** $\text{Var}(X)$ veya σ_X^2 sembolü veya herhangi bir rastlantı değişkenine atıfta bulunduğumuzda basitçe σ^2 ile gösterilir.

Tanım 4.3: X, olasılık dağılımı $f(x)$ ve ortalaması μ olan bir rastlantı değişkeni olsun.

X'in varyansı

X kesikli ise;

$$\sigma^2 = E[(X - \mu)^2] = \sum_x (x - \mu)^2 f(x)$$

X sürekli ise;

$$\sigma^2 = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

Varyansın pozitif karekökü σ ; X'in standart sapması olarak adlandırılır.

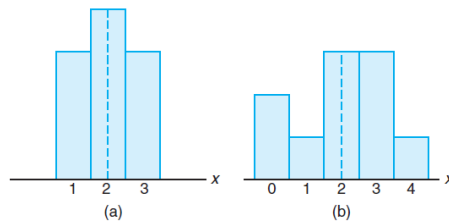
Örnek 4.8: Rastlantı değişkeni X, herhangi bir iş gününde resmi iş amaçları için kullanılan otomobil sayısını temsil etmektedir. A şirketi için olasılık dağılımı [Şekil 4.1(a)] şu şekildedir:

x	1	2	3
$f(x)$	0.3	0.4	0.3

ve B şirketi için [Şekil 4.1(b)]

x	0	1	2	3	4
$f(x)$	0.2	0.1	0.3	0.3	0.1

B şirketi için olasılık dağılımının varyansının A şirketi için olandan daha büyük olduğunu gösteriniz.



Şekil 4.1: Eşit ortalamalı dağılımlar ve eşit olmayan dağılımlar.

Cevap 4.8: A şirketi için şunu buluruz:

$$\mu_A = E(X) = (1)(0.3) + (2)(0.4) + (3)(0.3) = 2.0,$$

$$\sigma_A^2 = \sum_{x=1}^3 (x - 2)^2 f(x) = (1 - 2)^2 (0.3) + (2 - 2)^2 (0.4) + (3 - 2)^2 (0.3) = 0.6.$$

B şirketi için

$$\mu_B = E(X) = (0)(0.2) + (1)(0.1) + (2)(0.3) + (3)(0.3) + (4)(0.1) = 2.0,$$

$$\begin{aligned}\sigma_B^2 &= \sum_{x=0}^4 (x - 2)^2 f(x) \\ &= (0 - 2)^2(0.2) + (1 - 2)^2(0.1) + (2 - 2)^2(0.3) \\ &\quad + (3 - 2)^2(0.3) + (4 - 2)^2(0.1) = 1.6.\end{aligned}$$

Açıkçası, resmi iş amaçları için kullanılan otomobil sayısındaki varyans, B şirketi için A şirketinden daha fazladır.

Genellikle hesaplamaları basitleştiren σ^2 'yi bulmak için alternatif ve tercih edilen bir formül aşağıdaki teoremden belirtilmiştir.

TEOREM 4.2: X rastlantı değişkeninin varyansı

$$\sigma^2 = E(X^2) - \mu^2$$

Örnek 4.9: Bir üretim hattından 3 parça numune alındığında ve test edildiğinde, rastlantı değişkeni X bir makine için kusurlu parça sayısını temsil etmektedir. Aşağıdakiler, X'in olasılık dağılımıdır.

x	0	1	2	3
$f(x)$	0.51	0.38	0.10	0.01

Teorem 4.2'yi kullanarak σ^2 'yi hesaplayınız.

Cevap 4.9: İlk önce;

$$\mu = (0)(0.51) + (1)(0.38) + (2)(0.10) + (3)(0.01) = 0.61.$$

Sonra

$$E(X^2) = (0)(0.51) + (1)(0.38) + (4)(0.10) + (9)(0.01) = 0.87.$$

Sonuç olarak

$$\sigma^2 = 0.87 - (0.61)^2 = 0.4979.$$

Örnek 4.10: Yerel bir verimlilik mağazaları zincirinden binlerce litre içme suyu ürünü için haftalık talep miktarı ton cinsinden X sürekli rastlantı değişkeni olsun ve aşağıdaki olasılık yoğunluğuna sahip olsun.

$$f(x) = \begin{cases} 2(x-1), & 1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases} \quad X\text{'in ortalamasını ve varyansını bulunuz.}$$

Cevap 4.10: $E(X)$ ve $E(X^2)$ 'nin hesaplanması;

$$\mu = E(X) = 2 \int_1^2 x(x-1) dx = \frac{5}{3}$$

$$E(X^2) = 2 \int_1^2 x^2(x-1) dx = \frac{17}{6}. \quad \text{Sonra}$$

$$\sigma^2 = \frac{17}{6} - \left(\frac{5}{3}\right)^2 = \frac{1}{18}.$$

Bu noktada, varyans veya standart sapma, yalnızca aynı ölçü birimlerine sahip iki veya daha fazla dağılımı karşılaştırdığımızda anlam kazanır. Bu nedenle, iki şirketten alınan portakal suyu şişelerinin litre cinsinden ölçülen içerik dağılımlarının varyanslarını karşılaştırabiliriz ve daha büyük değer, ürünü daha değişken veya daha az tekdüze olan şirketi gösterir. Boy dağılımının varyansını yetenek puanlarının dağılımının varyansı ile karşılaştırmak anlamlı olmayacaktır. İlerleyen bölümde, standart sapmanın tek bir gözlem dağılımını tanımlamak için nasıl kullanılabileceği gösterilecektir.

Şimdi bir X rastlantı değişkeninin varyansı kavramımızı X ile ilgili rastlantı değişkenleri içerecek şekilde genişleteceğiz. Rastlantı değişkeni $g(X)$ için varyans $\sigma_{g(X)}^2$ ile gösterilir ve aşağıdaki teorem aracılığıyla hesaplanır.

TEOREM 4.3: X , $f(x)$ olasılık dağılımına sahip rastlantı değişkeni olsun.

Rastlantı değişkeni $g(X)$ 'in varyansı

X kesikli ise;

$$\sigma_{g(X)}^2 = E\{(g(X) - \mu_{g(X)})^2\} = \sum_x [g(X) - \mu_{g(X)}]^2 f(x)$$

X sürekli ise

$$\sigma_{g(X)}^2 = E\{(g(X) - \mu_{g(X)})^2\} = \int_{-\infty}^{\infty} [g(X) - \mu_{g(X)}]^2 f(x) dx$$

Örnek 4.11: X aşağıdaki olasılık dağılımına sahip rastlantı değişkeni olduğuna göre

x	0	1	2	3
$f(x)$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$

$g(X) = 2X + 3$ 'ün varyansını hesaplayınız.

Cevap 4.11: İlk olarak, rastlantı değişkeni $2X+3$ 'ün ortalamasını buluyoruz. Teorem 4.1'e göre,

$$\mu_{2X+3} = E(2X + 3) = \sum_{x=0}^3 (2x + 3)f(x) = 6.$$

$$\begin{aligned}\sigma_{2X+3}^2 &= E\{[(2X + 3) - \mu_{2X+3}]^2\} = E[(2X + 3 - 6)^2] \\ &= E(4X^2 - 12X + 9) = \sum_{x=0}^3 (4x^2 - 12x + 9)f(x) = 4.\end{aligned}$$

Şimdi, Teorem 4.3'ü kullanarak,

Örnek 4.12: X, Örnek 4.5'te verilen yoğunluk fonksiyonuna sahip bir rastlantı değişkeni olsun.

Rastlantı değişkeni $g(X) = 4X + 3$ 'ün varyansını bulunuz.

Cevap 4.12: Örnek 4.5'te $\mu_{4X+3} = 8$ olarak bulunmuştur. Şimdi, Teorem 4.3'ü kullanarak;

$$\begin{aligned}\sigma_{4X+3}^2 &= E\{[(4X + 3) - 8]^2\} = E[(4X - 5)^2] \\ &= \int_{-1}^2 (4x - 5)^2 \frac{x^2}{3} dx = \frac{1}{3} \int_{-1}^2 (16x^4 - 40x^3 + 25x^2) dx = \frac{51}{5}.\end{aligned}$$

PROBLEM 6.1: Ayırık (kesikli) rasgele değişken X'in olasılık dağılımı

$$f(x) = \binom{5}{x} \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{5-x}, \quad x = 0, 1, 2, 3, 4, 5.$$

'tir. X'in ortalamasını bulunuz.

PROBLEM 6.2: X, aşağıdaki olasılık dağılımına sahip rastgele bir değişken olsun:

x	-3	6	9
$f(x)$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$

$\mu_{g(x)}$ 'yi bulun, burada $g(X) = (2X + 1)^2$.

PROBLEM 6.3: X ve Y'nin aşağıdaki ortak olasılık fonksiyonuna $f(x, y)$ sahip olduğunu varsayalım:

y	x	
	2	4
1	0.10	0.15
3	0.20	0.30
5	0.10	0.15

(a) $g(X, Y) = XY^2$ 'nin beklenen değerini bulunuz.

(b) μ_X ve μ_Y 'yi bulunuz.

OLASILIK VE İSTATİSTİK

7. HAFTA

MATEMATİKSEL BEKLENTİ (DEVAM)

$\mu_X = E(X)$ ve $\mu_Y = E(Y)$ olmak üzere $g(X, Y) = (X - \mu_X)(Y - \mu_Y)$ ise, Tanım 4.4, X ve Y'nin kovaryansı olarak adlandırılan beklenen bir değeri verir. $\sigma_{X,Y}$ veya $Cov(X, Y)$ ile gösterilir.

Tanım 4.4: X ve Y, birleşik olasılık dağılımı $f(x, y)$ olan rastlantı değişkenleri olsun.

X ve Y'nin kovaryansı

X ve Y kesikli ise;

$$\sigma_{X,Y} = E[(X - \mu_X)(Y - \mu_Y)] = \sum_x \sum_y (x - \mu_X)(y - \mu_Y) f(x, y)$$

X ve Y sürekli ise;

$$\sigma_{X,Y} = E[(X - \mu_X)(Y - \mu_Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f(x, y) dx dy$$

İki rastlantı değişken arasındaki kovaryans, ikisi arasındaki ilişkinin doğasının bir ölçüsüdür. X'in büyük değerleri genellikle büyük Y değerleri veya küçük X değerleri Y'nin küçük değerleri ile sonuçlanırsa, pozitif $X - \mu_X$ genellikle pozitif $Y - \mu_Y$ ile sonuçlanır ve negatif $X - \mu_X$ genellikle negatif $Y - \mu_Y$ ile sonuçlanır. Böylece, $(X - \mu_X)(Y - \mu_Y)$ çarpımı pozitif olma eğiliminde olacaktır. Öte yandan, büyük X değerleri genellikle küçük Y değerleri ile sonuçlanırsa, $(X - \mu_X)(Y - \mu_Y)$ çarpımı negatif olma eğiliminde olacaktır. Kovaryansın işareti, iki bağımlı rastgele değişken arasındaki ilişkinin pozitif mi yoksa negatif mi olduğunu gösterir. X ve Y istatistiksel olarak bağımsız olduğunda, kovaryansın sıfır olduğu gösterilebilir. Bununla birlikte, sohbet genellikle doğru değildir. İki değişken sıfır kovaryansa sahip olabilir ve yine de istatistiksel olarak bağımsız olmayabilir. Kovaryansın yalnızca iki rasgele değişken arasındaki doğrusal ilişkiyi tanımladığını unutmayın. Bu nedenle, X ve Y arasındaki bir kovaryans sıfırsa, X ve Y'nin doğrusal olmayan bir ilişkisi olabilir, bu da onların zorunlu olarak bağımsız olmadıkları anlamına gelir.

$\sigma_{X,Y}$ için alternatif ve tercih edilen formül Teorem 4.4'te belirtilmiştir.

TEOREM 4.4: μ_X ve μ_Y ortalamasına sahip iki rastlantı değişkeni X ve Y'nin kovaryansı

$\sigma_{X,Y} = E(XY) - \mu_X\mu_Y$ ile hesaplanmaktadır.

Örnek 4.13: 3.Haftadaki Örnek 3.12'de, mavi kalemli sayısının X ve kırmızı kalemli sayısının Y sayısını içeren bir durumu açıklamaktadır. Bir tükenmez kalem için iki tanesi, belirli bir kutudan rastgele seçilmektedir. Aşağıdaki birleşik olasılık dağılımı verilmiştir:

$f(x,y)$		x			$h(y)$
		0	1	2	
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
$g(x)$		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

X ve Y'nin kovaryansını bulunuz.

Cevap 4.13: Örnek 4.6'dan $E(XY) = 3/14$ olduğu bulunmuştur. Şimdi;

$$\mu_X = \sum_{x=0}^2 xg(x) = (0) \left(\frac{5}{14} \right) + (1) \left(\frac{15}{28} \right) + (2) \left(\frac{3}{28} \right) = \frac{3}{4},$$

$$\mu_Y = \sum_{y=0}^2 yh(y) = (0) \left(\frac{15}{28} \right) + (1) \left(\frac{3}{7} \right) + (2) \left(\frac{1}{28} \right) = \frac{1}{2}.$$

Böylece

$$\sigma_{X,Y} = E(XY) - \mu_X\mu_Y = \frac{3}{14} - \left(\frac{3}{4} \right) \left(\frac{1}{2} \right) = -\frac{9}{56}.$$

Örnek 4.14: Maraton yarışlarında yarışan erkek koşucuların oranını X ve kadın koşucuların oranını Y rastlantı değişkenleri temsil etmektedir. X ve Y'nin birleşik yoğunluk fonksiyonu

$$f(x,y) = \begin{cases} 8xy, & 0 \leq y \leq x \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

ile tanımlanır. X ve Y'nin kovaryansını bulunuz.

Cevap 4.14: İlk önce ortalama değerler hesaplanır. Öncesinde sınırlar net olarak belirlenmelidir:

$y \leq x \leq 1$ ve $0 \leq y \leq x$ olarak iki durumda bakabiliriz.

Eğer ki integralin içini **dydx** sıralamasında yaparsak bu durumda önce y'ye göre çözeceğimizden dolayı y'nin aralığı $0 \leq y \leq x$ olduğundan en içteki integralin aralığı 0-x olmalıdır. Bu durumda en dıştaki integral x'e bağlı olmak durumunda olduğundan ve x'in sınırları $y \leq x \leq 1$ olduğundan integralin sınırları y-1 olmalıdır. Fakat en dıştaki çözüme

geldiğimizden burada y'ye en küçük değeri olan 0 verilir. Bu durumda ortalama aşağıdaki gibi yazılır:

$$\mu_X = \underbrace{\int_0^1 \int_0^x (x)f(x,y)dydx}_{\substack{\text{En dıştaki integralin aralığı } x' \text{ in sınırlarıdır} \\ \text{En içteki integralin aralığı } y' \text{ nin sınırlarıdır} \\ \text{Böyle bir durumda } dydx \text{ olmuş olur}}}$$

Eğer ki integralin içini **dx dy** sıralamasında yaparsak bu durumda önce x'ye göre çözeceğimizden dolayı x'nin aralığı $y \leq x \leq 1$ olduğundan en içteki integralin aralığı y-1 olmalıdır. Bu durumda en dıştaki integral y'ye bağlı olmaktadır. $0 \leq y \leq x$ y'nin sınırları olduğundan integralin sınırları 0-x olmalıdır. Fakat en dıştaki çözüme geldiğimizden burada x'e en küçük değeri olan 0 verilir. Bu durumda ortalama aşağıdaki gibi yazılır:

$$\mu_X = \underbrace{\int_0^1 \int_y^1 (x)f(x,y)dx dy}_{\substack{\text{En dıştaki integralin aralığı } x' \text{ in sınırlarıdır} \\ \text{En içteki integralin aralığı } y' \text{ nin sınırlarıdır} \\ \text{Böyle bir durumda } dydx \text{ olmuş olur}}}$$

Dolayısıyla iki çözümde aynı sonucu verecektir.

$\mu_X = \int_0^1 \int_y^1 (x)f(x,y)dx dy$ çözümü ile devam edilirse;

$$\int_0^1 \int_y^1 (x)(8xy)dx dy = \int_0^1 \int_y^1 8x^2y dx dy$$

Bu durumda ilk önce $\int_y^1 8x^2y dx$ integrali çözülmüşse çözüm $\frac{8x^3y}{3}$ olur. Bu sonuç dıştaki integralin içine aktarılır:

$\int_0^1 \frac{8x^3y}{3} dy$ olur. Bu integralin sonucu da $\frac{4}{5}$ 'e eşit olmuş olur. Yani $\mu_X = \int_0^1 \int_y^1 (x)(8xy)dx dy = \frac{4}{5}$. Aynı mantıkla

$\mu_Y = \int_0^1 \int_y^1 (y)(8xy)dx dy$ işleminin sonucu $\frac{8}{15}$ olarak bulunur.

Yukarıda verilen birleşik yoğunluk fonksiyonundan,

$E(XY) = \int_0^1 \int_y^1 (xy)(8xy)dx dy = \frac{4}{9}$ bulunur.

Böylece

$$\sigma_{XY} = E(XY) - \mu_X \mu_Y = \frac{4}{9} - \left(\frac{4}{5}\right) \left(\frac{8}{15}\right) = \frac{4}{225}.$$

İki rastlantı değişkeni arasındaki kovaryans, ilişkinin doğası hakkında bilgi sağlasa da, $\sigma_{X,Y}$ ölçeksiz olmadığı için $\sigma_{X,Y}$ 'nin büyüklüğü ilişkinin sağlamlığı hakkında hiçbir şey göstermez. Genlik, hem X hem de Y'yi ölçmek için kullanılan birimlere bağlı olacaktır. **İstatistikte yaygın olarak kullanılan korelasyon katsayısı adı verilen kovaryansın ölçeksiz bir versiyonu vardır.**

Tanım 4.5: X ve Y, standart sapmaları sırasıyla sırasıyla σ_X ve σ_Y olan, kovaryansları $\sigma_{X,Y}$ olan rastlantı değişkenleri olsun. X ve Y'nin korelasyon (ilinti) katsayısı

$$\rho_{XY} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}$$

$$*-1 \leq \rho_{XY} \leq 1$$

* ρ_{XY} X ve Y'nin ölçü birimlerinden bağımsızdır

*X ve Y arasında doğrusal bir ilişki varsa;

$$y = ax + b \text{ ve } b > 0 \text{ ise } \rho_{XY} = 1$$

$$y = ax + b \text{ ve } b < 0 \text{ ise } \rho_{XY} = -1$$

Örnek 4.15: Örnek 4.13'te X ve Y arasındaki korelasyon katsayısını bulunuz.

Cevap 4.15: Teorem 4.2'den $\sigma^2 = E(X^2) - \mu^2$ Tanım 4.1'den de $\mu = E(x) = \sum_x x f(x)$

Dolayısıyla;

$$E(X^2) = (0^2) \left(\frac{5}{14}\right) + (1^2) \left(\frac{15}{28}\right) + (2^2) \left(\frac{3}{28}\right) = \frac{27}{28}$$

$$E(Y^2) = (0^2) \left(\frac{15}{28}\right) + (1^2) \left(\frac{3}{7}\right) + (2^2) \left(\frac{1}{28}\right) = \frac{4}{7},$$

$$\sigma_X^2 = \frac{27}{28} - \left(\frac{3}{4}\right)^2 = \frac{45}{112}$$

$$\sigma_Y^2 = \frac{4}{7} - \left(\frac{1}{2}\right)^2 = \frac{9}{28}.$$

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{-9/56}{\sqrt{(45/112)(9/28)}} = -\frac{1}{\sqrt{5}}.$$

Örnek 4.16: Örnek 4.14'teki X ve Y'nin korelasyon katsayısını bulunuz.

Cevap 4.16: Örnek 4.14'te aşağıdaki ifadeler bulunmuştu:

$$E(XY) = \int_0^1 \int_y^1 (xy)(8xy) dx dy = \frac{4}{9}$$

$$\sigma_{XY} = E(XY) - \mu_X \mu_Y = \frac{4}{9} - \left(\frac{4}{5}\right) \left(\frac{8}{15}\right) = \frac{4}{225}.$$

Kovaryansı

$$\mu_X = \int_0^1 \int_y^1 (x)(8xy) dx dy = \frac{4}{5}$$

$$\mu_Y = \int_0^1 \int_y^1 (y)(8xy) dx dy = \frac{8}{15}$$

$$\sigma_X^2 = E(X^2) - \mu_X^2$$

$\sigma_Y^2 = E(Y^2) - \mu_Y^2$ işlemlerinin sonuçları için $E(X^2)$ ve $E(Y^2)$ hesaplanmalıdır. Dolayısıyla,

$$E(X^2) = \int_0^1 \int_y^1 (x^2)(8xy) dx dy = \frac{2}{3}$$

$$E(Y^2) = \int_0^1 \int_y^1 (y^2)(8xy) dx dy = \frac{1}{3}$$

Böylece

$$\sigma_X^2 = \frac{2}{3} - \left(\frac{4}{5}\right)^2 = \frac{2}{75}$$

$$\sigma_Y^2 = \frac{1}{3} - \left(\frac{8}{15}\right)^2 = \frac{11}{225}.$$

$$\rho_{XY} = \frac{4/225}{\sqrt{(2/75)(11/225)}} = \frac{4}{\sqrt{66}}.$$

Örnek 4.15'teki kovaryansın büyüklüğü (işareti dikkate alınmadan) Örnek 4.16'dakinden daha büyük olmasına rağmen, bu iki örnekteki korelasyon katsayılarının büyüklükleri arasındaki ilişkinin tam tersi olduğuna dikkat edin. Bu, ilişkinin ne kadar güçlü olduğuna karar vermek için kovaryansın büyüklüğüne bakamayacağımızın kanıtıdır.

4.3 RASTLANTI DEĞİŞKENLERİN DOĞRUSAL (LİNEER) BİRLEŞİMLERİNİN ORTALAMALARI VE VARYANSLARI

Bu bölümde, daha sonraki bölümlerde yer alacak olan rastlantı değişkenlerinin ortalama ve varyans hesaplamalarını basitleştirecek bazı faydalı özellikler geliştiriyoruz.

TEOREM 4.5: a ve b sabit ise, o zaman

$$E(ax + b) = aE(x) + b$$

Sonuç 4.1: $a = 0$ olarak ayarlayarak, $E(b) = b$ olduğunu görüyoruz.

Sonuç 4.2: $b = 0$ olarak ayarlayarak, $E(aX) = aE(X)$ olduğunu görüyoruz.

Örnek 4.17: Teorem 4.5'i kesikli rastlantı değişken $f(X) = 2X - 1$ 'e uygulayarak, Örnek 4.4'ü buna göre yeniden çözünüz.

Cevap 4.17: Teorem 4.5'e göre,

$$E(2X - 1) = 2E(X) - 1. \text{ Dolayısıyla,}$$

$$\begin{aligned} \mu &= E(X) = \sum_{x=4}^9 xf(x) \\ &= (4) \left(\frac{1}{12} \right) + (5) \left(\frac{1}{12} \right) + (6) \left(\frac{1}{4} \right) + (7) \left(\frac{1}{4} \right) + (8) \left(\frac{1}{6} \right) + (9) \left(\frac{1}{6} \right) = \frac{41}{6}. \end{aligned}$$

$$\mu_{2X-1} = (2) \left(\frac{41}{6} \right) - 1 = \$12.67,$$

Örnek 4.18: Teorem 4.5'i sürekli rastlantı değişkeni $g(X) = 4X + 3$ 'e uygulayarak, Örnek 4.5'i yeniden çözünüz.

Cevap 4.18: Örnek 4.5 için, $E(4X + 3) = 4E(X) + 3$ olarak yazmak için Teorem 4.5'i kullanabiliriz. Dolayısıyla,

$$E(X) = \int_{-1}^2 x \left(\frac{x^2}{3} \right) dx = \int_{-1}^2 \frac{x^3}{3} dx = \frac{5}{4}.$$

$$E(4X + 3) = (4) \left(\frac{5}{4} \right) + 3 = 8,$$

TEOREM 4.6: X rastlantı değişkeninin iki veya daha fazla fonksiyonunun toplamının veya farkının beklenen değeri, fonksiyonların beklenen değerlerinin toplamı veya farkıdır. Yani,

$$E[g(X) \pm h(X)] = E[g(X)] \pm E[h(X)]$$

Örnek 4.19: X , aşağıdaki gibi olasılık dağılımına sahip rastlantı değişkeni olsun:

x	0	1	2	3
$f(x)$	$\frac{1}{3}$	$\frac{1}{2}$	0	$\frac{1}{6}$

$Y = (X - 1)^2$ 'nin beklenen değerini bulunuz.

Cevap 4.19: Teorem 4.6'yı $Y = (X - 1)^2$ fonksiyonuna uygulayarak yazabiliriz:

$$E[(X - 1)^2] = E(X^2 - 2X + 1) = E(X^2) - 2E(X) + E(1).$$

Sonuç 4.1'den, $E(1) = 1$ ve doğrudan hesaplama ile,

$$E(X) = (0)\left(\frac{1}{3}\right) + (1)\left(\frac{1}{2}\right) + (2)(0) + (3)\left(\frac{1}{6}\right) = 1.$$

$$E(X^2) = (0)\left(\frac{1}{3}\right) + (1)\left(\frac{1}{2}\right) + (4)(0) + (9)\left(\frac{1}{6}\right) = 2.$$

$$E[(X - 1)^2] = 2 - (2)(1) + 1 = 1.$$

Örnek 4.20: Bir market zincirinde belirli bir içeceğe yönelik haftalık talep, ton cinsinden sürekli bir rasgele değişkendir $g(X) = X^2 + X - 2$, burada X , aşağıdaki $f(x)$ yoğunluk fonksiyonuna sahiptir.

$$f(x) = \begin{cases} 2(x - 1), & 1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

İçecek için haftalık talebin beklenen değerini bulun.

Cevap 4.20: Teorem 4.6'ya göre,

$$E(X^2 + X - 2) = E(X^2) + E(X) - E(2).$$

Sonuç 4.1'den, $E(2) = 2$ olarak integral alırsak,

$$E(X) = \int_1^2 2x(x-1) dx = \frac{5}{3} \text{ and } E(X^2) = \int_1^2 2x^2(x-1) dx = \frac{17}{6}.$$

Sonuç olarak:

$$E(X^2 + X - 2) = \frac{17}{6} + \frac{5}{3} - 2 = \frac{5}{2},$$

Yani bu mağazalar zincirinden içeceğe haftalık ortalama talep 2500 litre.

Birleşik olasılık dağılımı $f(x, y)$ olan iki rastlantı değişkenimiz X ve Y olduğunu varsayalım. Sonraki bölümlerde çok yararlı olacak iki ek özellik, bu iki rastlantı değişkeninin toplamının, farkının ve çarpımının beklenen değerlerini içermektedir. Ancak önce, verilen değişkenlerin fonksiyonlarının toplamının veya farkının beklenen değeri üzerine bir teorem aşağıda belirtilmiştir. Bu, elbette, Teorem 4.6'nın yalnızca bir uzantısıdır.

TEOREM 4.7: X ve Y rastlantı değişkenlerinin iki veya daha fazla fonksiyonunun toplamının veya farkının beklenen değeri, fonksiyonların beklenen değerlerinin toplamı veya farkıdır. Yani,

$$E[g(X, Y) \pm h(X, Y)] = E[g(X, Y)] \pm E[h(X, Y)]$$

Sonuç 4.3: $g(X, Y) = g(X)$ ve $h(X, Y) = h(Y)$ ise,

$$E[g(X) \pm h(Y)] = E[g(X)] \pm E[h(Y)]$$

Sonuç 4.4: $g(X, Y) = X$ ve $h(X, Y) = Y$ ise,

$$E[X \pm Y] = E[X] \pm E[Y]$$

X , A makinesinden bir parçanın günlük üretimini ve Y , B makinesinden aynı türden bir öğenin günlük üretimini temsil ediyorsa, $X + Y$, her iki makine tarafından günlük olarak üretilen toplam parça sayısını temsil eder. Sonuç 4.4, her iki makine için ortalama günlük üretimin, her makinenin ortalama günlük üretiminin toplamına eşit olduğunu belirtir.

TEOREM 4.8: X ve Y iki bağımsız rastlantı değişken olsun.

O zaman $E(XY) = E(X)E(Y)$.

Teorem 4.8, bir yeşil zar ve bir kırmızı zar atma deneyi dikkate alınarak kesikli değişkenler için gösterilebilir. Rastlantı değişkeni X'in yeşil zardaki sonucu ve Y rastlantı değişkeninin ise kırmızı zardaki sonucu temsil ettiği varsayalım. Daha sonra XY, zar çiftinde oluşan sayıların çarpımını temsil etmektedir. Uzun zaman döneminde, sayıların çarpımlarının ortalaması, yeşil zarda oluşan ortalama sayı ile kırmızı zarda oluşan ortalama sayının çarpımına eşittir.

Sonuç 4.5: X ve Y iki bağımsız rastlantı değişkeni olsun.

O zaman $\sigma_{XY} = 0$

Örnek 4.21: X ve Y aşağıdaki birleşik olasılık dağılımına sahip bağımsız rastlantı değişkenleridir.

$$f(x, y) = \begin{cases} \frac{x(1+3y^2)}{4}, & 0 < x < 2, 0 < y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Teorem 4.8'in önerdiği gibi $E(XY) = E(X)E(Y)$

olduğunu gösteriniz.

Cevap 4.21: Teoremden;

$$E(XY) = \int_0^1 \int_0^2 \frac{x^2 y (1 + 3y^2)}{4} dx dy = \frac{5}{6}, \quad E(X) = \frac{4}{3}, \quad \text{and} \quad E(Y) = \frac{5}{8}. \quad \text{Böylece} \quad E(X)E(Y) = \left(\frac{4}{3}\right)\left(\frac{5}{8}\right) = \frac{5}{6} = E(XY).$$

TEOREM 4.9: X ve Y birleşik olasılık dağılımı f(x, y) olan rastlantı değişkenleri ve a, b ve c sabit olmak üzere;

$$\sigma_{aX+bY+c}^2 = a^2 \sigma_X^2 + b^2 \sigma_Y^2 + 2ab \sigma_{XY}$$

Sonuç 4.6: b=0 ise;

$$\sigma_{aX+c}^2 = a^2 \sigma_X^2 = a^2 \sigma^2$$

Sonuç 4.7: a=1 ve b=0 ise;

$$\sigma_{X+c}^2 = \sigma_X^2 = \sigma^2$$

Sonuç 4.8: b=0 ve c=0 ise;

$$\sigma_{aX}^2 = a^2 \sigma_X^2 = a^2 \sigma^2$$

Örnek 4.22: X ve Y, varyansları $\sigma_X^2 = 2$ ve $\sigma_Y^2 = 4$ ve kovaryansı $\sigma_{XY} = -2$ olan rastlantı değişkenleri ise, $Z = 3X - 4Y + 8$ rastlantı değişkeninin varyansını bulunuz.

Cevap 4.22: Sonuç 4.6'dan ve Teorem 4.9'dan

$$\begin{aligned}\sigma_Z^2 &= \sigma_{3X-4Y+8}^2 = \sigma_{3X-4Y}^2 \\ &= 9\sigma_X^2 + 16\sigma_Y^2 - 24\sigma_{XY} \\ &= (9)(2) + (16)(4) - (24)(-2) = 130.\end{aligned}$$

Örnek 4.23: X ve Y, belirli bir kimyasal ürün grubundaki iki farklı türde safsızlığın miktarlarını gösterebilir. X ve Y'nin varyansları $\sigma_X^2 = 2$ ve $\sigma_Y^2 = 3$ olan bağımsız rastlantı değişkenleri olduğunu varsayalım. $Z = 3X - 2Y + 5$ rastlantı değişkeninin varyansını bulunuz.

Cevap 4.23: Sonuç 4.6'dan ve Teorem 4.10'dan

$$\begin{aligned}\sigma_Z^2 &= \sigma_{3X-2Y+5}^2 = \sigma_{3X-2Y}^2 \\ &= 9\sigma_X^2 + 4\sigma_Y^2 \\ &= (9)(2) + (4)(3) = 30.\end{aligned}$$

Sonuç 4.6 ve 4.7, bir rastlantı değişkene bir sabit eklenirse veya ondan çıkarılırsa varyansın değişmediğini belirtmektedir. Bir sabitin eklenmesi veya çıkarılması, X'in değerlerini basitçe sağa veya sola kaydırır, ancak değişkenliğini değiştirmez. Bununla birlikte, rastlantı değişkeni bir sabitle çarpılır veya bölünürse, Sonuç 4.6 ve 4.8, varyansın sabitin karesiyle çarpıldığını veya bölündüğünü belirtmektedir.

Sonuç 4.9: X ve Y bağımsız rastlantı değişkenleri ise, o zaman

$$\sigma_{aX+bY}^2 = a^2\sigma_X^2 + b^2\sigma_Y^2$$

Sonuç 4.10: X ve Y bağımsız rastlantı değişkenleri ise, o zaman

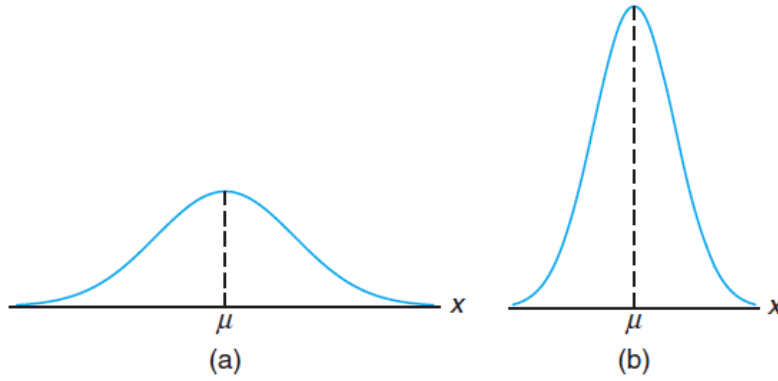
$$\sigma_{aX-bY}^2 = a^2\sigma_X^2 + b^2\sigma_Y^2$$

Sonuç 4.11: $X_1, X_2, \dots, X_n$ bağımsız rastlantı değişkenleri ise, o zaman

$$\sigma_{a_1X_1+a_2X_2+\dots+a_nX_n}^2 = a_1^2\sigma_{X_1}^2 + a_2^2\sigma_{X_2}^2 + \dots + a_n^2\sigma_{X_n}^2$$

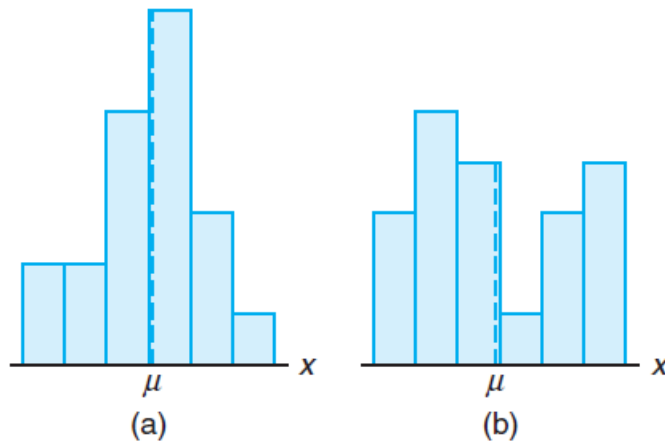
4.4 CHEBYSHEV TEOREMİ

Bölüm 4.2'de bir rastlantı değişkeninin varyansının bize ortalama hakkındaki gözlemlerin değişkenliği hakkında bir şeyler söylediğini belirtmiştik. Rastlantı değişkeninin küçük bir varyansı veya standart sapması varsa, değerlerin çoğunun ortalama etrafında gruplanmasını bekleriz. **Bu nedenle, rastlantı değişkeninin ortalamaya göre belirli bir aralıkta bir değer alma olasılığı, daha büyük standart sapmaya sahip benzer bir rastlantı değişkenine göre daha yüksektir. Olasılığı alan açısından düşünürsek, büyük bir σ değerine sahip sürekli bir dağılımın daha büyük bir değişkenliği göstermesini bekleriz** ve bu nedenle, Şekil 4.2(a)'daki gibi, alanın daha fazla yayılmasını beklemeliyiz. Küçük bir standart sapmaya sahip bir dağılımın alanının çoğu, Şekil 4.2(b)'deki gibi μ 'ye yakın olmalıdır.



Şekil 4.2: Sürekli gözlemlerin ortalamaya göre değişkenliği.

Kesikli bir dağılım için de aynı şekilde tartışabiliriz. Şekil 4.3(b)'deki olasılık histogramındaki alan, Şekil 4.3(a)'dakinden çok daha fazla yayılmıştır ve ölçümlerin veya sonuçların daha değişken dağılımını göstermektedir.



Şekil 4.3: Kesikli gözlemlerin ortalamaya göre değişkenliği.

Rus matematikçi P. L. Chebyshev (1821–1894), ortalamaya göre simetrik olan herhangi iki değer arasındaki alanın kesirinin standart sapma ile ilişkili olduğunu keşfetti. Bir olasılık dağılım eğrisinin altındaki alan veya bir olasılık histogramındaki alan 1'e eklendiğinden, herhangi iki sayı arasındaki alan, bu sayılar arasında bir değer alan rasgele değişkenin olasılığıdır.

Chebyshev'den kaynaklanan aşağıdaki teorem, bir rastlantı değişkenin herhangi bir k gerçek sayısı için ortalamasının k standart sapması dahilinde bir değer alma olasılığının ölçülü bir tahminini vermektedir.

TEOREM 4.10: CHEBYSHEV TEOREMİ: Herhangi bir X rastlantı değişkeninin, ortalamasının k standart sapması içinde bir değer alma olasılığı en az $(1 - 1/k^2)$ 'dir.

Yani,

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$

$k = 2$ için teorem, X rastlantı değişkeninin ortalamasının iki standart sapması içine düşme olasılığının en az $1 - 1/2^2 = 3/4$ olduğunu belirtmektedir. Yani, herhangi bir dağılımın gözlemlerinin dörtte üçü veya daha fazlası $\mu \pm 2\sigma$ aralığındadır. Benzer şekilde teorem, herhangi bir dağılımın gözlemlerinin en az dokuzda sekizinin $\mu \pm 3\sigma$ aralığına düştüğünü söyler.

Örnek 4.24: Bir rastlantı değişkeni X , $\mu = 8$ ortalamasına, $\sigma^2 = 9$ varyansına ve bilinmeyen bir olasılık dağılımına sahiptir. Buna göre

a) $P(-4 < X < 20)$ **b)** $P(|X - 8| \geq 6)$ 'yı bulunuz.

Cevap 4.24:

(a) $P(-4 < X < 20) = P[8 - (4)(3) < X < 8 + (4)(3)] \geq \frac{15}{16}.$

(b) $P(|X - 8| \geq 6) = 1 - P(|X - 8| < 6) = 1 - P(-6 < X - 8 < 6)$
 $= 1 - P[8 - (2)(3) < X < 8 + (2)(3)] \leq \frac{1}{4}.$

Chebyshev teoremi herhangi bir gözlem dağılımı için geçerlidir ve bu nedenle sonuçlar genellikle zayıftır. Teoremin verdiği değer yalnızca bir alt sınırdır. Yani, bir rastlantı değişkenin ortalamasının iki standart sapması içine düşme olasılığını biliyoruz, ancak gerçekte ne kadar fazla olabileceğini asla bilemeyiz. Yalnızca olasılık dağılımı bilindiğinde kesin olasılıkları belirleyebiliriz. Bu nedenle teoremi dağılımsız sonuç olarak adlandırıyoruz.

PROBLEM 7.1: Bir uçağın belirli bir havaalanında kalkış izni elde etmesi için geçen dakika cinsinden süre, rastgele bir değişken $Y = 3X - 2$ 'dir, burada X ,

$$f(x) = \begin{cases} \frac{1}{4}e^{-x/4}, & x > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

yoğunluk fonksiyonuna sahiptir. Y rastgele değişkeninin ortalamasını ve varyansını bulunuz.

PROBLEM 7.2: X ve Y 'nin aşağıda gösterildiği gibi ortak olasılık dağılımına sahip bağımsız rasgele değişkenler olduğunu varsayalım.

y	x	
	2	4
1	0.15	0.10
3	0.25	0.25
5	0.15	0.10

a.) $E(2X - 3Y)$;

b.) $E(XY)$.

PROBLEM 7.3: X ve Y , varyansları $\sigma_X^2 = 5$ ve $\sigma_Y^2 = 3$ olan bağımsız rasgele değişkenlerse, $Z = -2X + 4Y - 3$ rasgele değişkeninin varyansını bulunuz.

PROBLEM 7.4: Bir otomobil üretim fabrikasında 120 yeni işin açıldığını ve 70 pozisyon için 1000 kişinin başvurduğunu varsayalım. Başvuranlar arasından en iyi 120'yi seçmek için şirket, mekanik beceri, el becerisi ve matematik becerisini kapsayan bir test uyguluyor. Bu testin ortalama notu 60'tır ve puanların standart sapması 6'dır. 84 puan alan bir kişi işlerden birine girebilir mi? [İpucu: Chebyshev teoremini kullanın.] Dağılımın ortalamaya göre simetrik olduğunu varsayalım.

OLASILIK VE İSTATİSTİK

8. / 9. HAFTA

BAZI KESİKLİ (AYRIK) OLASILIK DAĞILIMLARI

5.1 GİRİŞ

Kesikli bir olasılık dağılımı ister grafiksel olarak bir histogramla, tablo biçiminde veya bir formül aracılığıyla temsil edilsin, bir rastlantı değişkeninin davranışı açıklanır. Genellikle, farklı istatistiksel deneyler tarafından üretilen gözlemler aynı genel davranış tipine sahiptir. Sonuç olarak, bu deneylerle ilişkili kesikli rastlantı değişkenleri, esas olarak aynı olasılık dağılımıyla tanımlanabilir ve bu nedenle tek bir formülle temsil edilebilir. Aslında, pratikte karşılaşılan kesikli rastlantı değişkenlerinin çoğunu tanımlamak için yalnızca bir avuç önemli olasılık dağılımına ihtiyaç vardır.

Bu kadar az sayıda dağılım, gerçek hayattaki birkaç rastlantı olgusunu tanımlar. Örneğin, yeni bir ilacın etkinliğinin test edilmesini içeren bir çalışmada, ilacı kullanan tüm hastalar arasında iyileşen hastaların sayısı yaklaşık olarak bir binom dağılımını izler (Bölüm 5.2). Endüstriyel bir örnekte, bir üretim partisinden seçilen bir ürün numunesi test edildiğinde, numunedeki kusurlu ürün sayısı genellikle bir hipergeometrik rastlantı değişkeni olarak modellenir (Bölüm 5.3). İstatistiksel bir kalite kontrol probleminde, deneyi yapan kişi, gözlemsel veriler belirli sınırları aştığında, işlem ortalamasının bir kaymasını işaret edecektir. Yanlış alarm üretmek için gereken örnek sayısı, negatif binom dağılımının özel bir durumu olan geometrik bir dağılımı takip eder (Bölüm 5.4). Öte yandan, bir bireyin kan örneğinin sabit bir miktarındaki beyaz küre sayısı genellikle rastlantıdır ve bir Poisson dağılımı ile açıklanabilir (Bölüm 5.5). Bu bölümde, yaygın olarak kullanılan bu dağılımları çeşitli örneklerle sunuyoruz.

5.2 BİNOM (İKİ TERİMLİ) VE ÇOK TERİMLİ (MULTİNOM/MULTİNOMİNAL) DAĞILIMLAR

Bir deney genellikle, her biri başarı veya başarısızlık olarak etiketlenebilecek iki olası sonucu olan tekrarlanan denemelerden oluşmaktadır. En bariz uygulama, her bir denemenin kusurlu veya kusurlu olmayan bir parçayı gösterebileceği, bir montaj hattından çıkan parçaların test edilmesiyle ilgilidir. Her iki sonucu da başarı olarak tanımlamayı seçebiliriz. Süreç, **Bernoulli süreci olarak adlandırılır. Her denemeye Bernoulli denemesi denir.**

A. Bernoulli Süreci

Açıkçası, Bernoulli süreci aşağıdaki özelliklere sahip olmalıdır:

1. Deney, tekrarlanan denemelerden oluşur.
2. Her deneme, başarılı veya başarısız olarak sınıflandırılabilir bir sonuçla sonuçlanır.
3. p ile gösterilen başarı olasılığı denemeden denemeye sabit kalır.
4. Tekrarlanan denemeler bağımsızdır.

Üç ögenin bir üretim sürecinden rastgele seçildiği, incelendiği ve kusurlu veya kusursuz olarak sınıflandırıldığı Bernoulli deneylerini düşünün. Arızalı bir öge başarılı olarak belirlenir. Başarıların sayısı, 0'dan 3'e kadar tamsayı değerleri varsayan rastgele bir X değişkenidir. Sekiz olası sonuç ve X 'in karşılık gelen değerleri şunlardır:

Outcome	NNN	NDN	NND	DNN	NDD	DND	DDN	DDD
x	0	1	1	1	2	2	2	3

Öğeler bağımsız olarak seçildiğinden ve sürecin %25 kusurlu ürettiğini varsaydığımızdan,

$$P(NDN) = P(N)P(D)P(N) = \left(\frac{3}{4}\right) \left(\frac{1}{4}\right) \left(\frac{3}{4}\right) = \frac{9}{64}.$$

Benzer hesaplamalar, diğer olası sonuçlar için olasılıkları verir. X 'in olasılık dağılımı bu nedenle

x	0	1	2	3
$f(x)$	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$

B. İki Terimli (Binom) Dağılım

n Bernoulli denemesindeki başarıların sayısı X , iki terimli rasgele (rastlantı) değişken olarak adlandırılır. Bu kesikli (ayrık) rasgele değişkenin olasılık dağılımına binom dağılımı denir ve değerleri, deneme sayısına ve belirli bir denemede başarı olasılığına bağlı olduklarından $b(x; n, p)$ ile gösterilecektir. Bu nedenle, X 'in olasılık dağılımı için, kusurların sayısı

$$P(X = 2) = f(2) = b\left(2; 3, \frac{1}{4}\right) = \frac{9}{64}.$$

Şimdi $b(x; n, p)$ için bir formül elde etmek için yukarıdaki örneği genelleştirelim. Yani, iki terimli bir deney için n denemede x başarı olasılığını veren bir formül bulmak istiyoruz. İlk olarak, belirli bir sırayla x başarı ve $n - x$ başarısızlık olasılığını göz önünde bulundurun. Denemeler bağımsız olduğundan, farklı sonuçlara karşılık gelen tüm olasılıkları çarpabiliriz. Her başarı p olasılıkla ve her başarısızlık $q = 1 - p$ olasılıkla gerçekleşir. Bu nedenle, belirtilen sıra için olasılık $p^x q^{n-x}$ 'dir. Şimdi, deneyde x başarılı ve $n - x$ başarısız olan örnek noktaların toplam sayısını belirlemeliyiz. Bu sayı, bir grupta x ve diğerinde $n - x$ olmak üzere n sonucun iki gruba ayrılma sayısına eşittir ve Bölüm 2.3'te tanıtıldığı gibi $\binom{n}{x}$ olarak yazılır. Bu bölümler birbirini dışladığından, genel formülü elde etmek için tüm farklı bölümlerin olasılıklarını toplarız veya basitçe $p^x q^{n-x}$ 'i $\binom{n}{x}$ ile çarpabiliriz.

Bir Bernoulli denemesi, p olasılıkla başarılı ve $q = 1 - p$ olasılıkla başarısızlıkla sonuçlanabilir. Daha sonra, n bağımsız denemede başarı sayısı olan binom rastlantı (rastgele) değişkeni X 'in olasılık dağılımı şu şekildedir:

$$b(x; n, p) = \binom{n}{x} p^x q^{n-x}, x = 0, 1, 2, \dots, n$$

Örnek 5.1: Belirli bir tür cihazın bir çarpışma testinden sonra çalışıyor olma olasılığı $3/4$ 'tür. Test edilen 4 cihazdan 2 tanesinin çalışıyor olma olasılığını bulunuz.

Cevap 5.1: Testlerin bağımsız olduğunu ve 4 testin her biri için $p = 3/4$ olduğunu varsayarsak, şunu elde ederiz:

$$b(x; n, p) = \binom{n}{x} p^x q^{n-x} = b\left(2; 4, \frac{3}{4}\right) = \binom{4}{2} \left(\frac{3}{4}\right)^2 \left(\frac{1}{4}\right)^2 = \frac{27}{128}$$

C. Binom (İki Terimli) Adı Nereden Geliyor?

Binom dağılımı adını, $(q + p)^n$ 'nin iki terimli açılımındaki $n + 1$ teriminin $x = 0, 1, 2, \dots, n$ için $b(x; n, p)$ 'nin çeşitli değerlerine karşılık gelmesinden almaktadır. Yani,

$$\begin{aligned} (q + p)^n &= \binom{n}{0} q^n + \binom{n}{1} p q^{n-1} + \binom{n}{2} p^2 q^{n-2} + \dots + \binom{n}{n} p^n \\ &= b(0; n, p) + b(1; n, p) + b(2; n, p) + \dots + b(n; n, p). \end{aligned}$$

$q + p = 1$, herhangi bir olasılık dağılımı için tutması gereken bir koşul olan;

$$\sum_{x=0}^n b(x; n, p) = 1,$$

Sıklıkla $P(X < r)$ veya $P(a \leq X \leq b)$ bulmanın gerekli olduğu problemlerle ilgileniriz.
Binom toplamları

$$B(r; n, p) = \sum_{x=0}^r b(x; n, p)$$

0.1'den 0.9'a kadar seçilen p değerleri $n = 0, 1, 2, \dots, 20$ için Ek'te Tablo A.1'de verilmiştir. Tablo A.1'in kullanımını aşağıdaki örnekle açıklıyoruz.

Örnek 5.2: Bir hastanın nadir görülen bir kan hastalığından kurtulma olasılığı 0.4'tür. 15 kişinin bu hastalığa yakalandığı biliniyorsa,

- (a) en az 10 kişinin yaşama,
- (b) 3 ila 8 kişinin yaşama ve
- (c) tam olarak 5 kişinin yaşama olasılığı nedir?

Cevap 5.2: Hayatta kalan insan sayısı X olsun.

- (a) $P(X \geq 10) = 1 - P(X < 10) = 1 - \sum_{x=0}^9 b(x; 15, 0.4) = 1 - 0.9662 = 0.0338$
- (b) $P(3 \leq X \leq 8) = \sum_{x=3}^8 b(x; 15, 0.4) = \sum_{x=0}^8 b(x; 15, 0.4) - \sum_{x=0}^2 b(x; 15, 0.4)$
 $= 0.9050 - 0.0271 = 0.8779$
- (c) $P(X=5) = b(5; 15, 0.4) = \sum_{x=0}^5 b(x; 15, 0.4) - \sum_{x=0}^4 b(x; 15, 0.4)$
 $= 0.4032 - 0.2173 = 0.1859$

Örnek 5.3: Büyük bir zincir perakendeci, bir üreticiden belirli bir tür elektronik cihaz satın almaktadır. Üretici, cihazın kusurlu olma oranının %3 olduğunu belirtmektedir.

- (a) Müfettiş, bir gönderiden rastgele 20 ürün seçer. Bu 20 üründen en az birinin kusurlu olma olasılığı nedir?,
- (b) Perakendecinin ayda 10 gönderi aldığını ve müfettişin gönderi başına rastgele 20 cihazı test ettiğini varsayalım. Sevkiyattan seçilen ve test edilen 20 cihaz arasından her biri en az bir kusurlu cihaz içeren tam 3 sevkiyat olma olasılığı nedir?

Cevap 5.3: 20 cihaz içindeki arızalı cihaz sayısını X ile gösterelim. Daha sonra X , $b(x; 20, 0.03)$ dağılımını takip eder. Buradan,

a.) $P(X \geq 1) = 1 - P(X = 0) = 1 - b(0; 20, 0.03) = 1 - \binom{20}{0} (0.03)^0 (1 - 0.03)^{20-0} = 0.4562$

b.) Bu durumda, her gönderi en az bir kusurlu ürün içerebilir veya içermeyebilir. Bu nedenle, her sevkiyatın testi, (a) bölümünden $p = 0,4562$ olan bir Bernoulli denemesi olarak görülebilir. Sevkiyattan sevkiyata bağımsızlık varsayılarak ve en az bir kusurlu ürün içeren sevkiyatların sayısını Y ile ifade ederek, Y başka bir binom dağılımını $b(y; 10, 0.4562)$ takip eder. Öyleyse,

$$P(Y=3) = \binom{10}{3} (0.4562)^3 (1 - 0.4562)^{10-3} = 0.1602$$

D. Uygulama Alanları

Örnek 5.1'den 5.3'e kadar, iki terimli dağılımın birçok bilimsel alanda uygulama bulduğu açık olmalıdır. Bir endüstri mühendisi, endüstriyel bir süreçteki "orantılı kusurlu" ile yakından ilgilenir. Süreçler için kalite kontrol önlemleri ve örnekleme şemaları genellikle iki terimli dağılıma dayalıdır. Bu dağılım, bir sürecin sonucunun ikili olduğu ve sürecin sonuçlarının bağımsız olduğu ve başarı olasılığının denemeden denemeye sabit olduğu herhangi bir endüstriyel durum için geçerlidir. Binom dağılımı, tıbbi ve askeri uygulamalar için de yaygın olarak kullanılmaktadır. Her iki alanda da başarı veya başarısızlık sonucu önemlidir. Örneğin, ilaç işinde "tedavi" veya "tedavi yok" önemlidir ve "vurmak" veya "ıskalamak" genellikle güdümlü bir füze ateşlemenin sonucunun yorumlanmasıdır.

Herhangi bir iki terimli rasgele (rastlantı) değişkenin olasılık dağılımı yalnızca n , p ve q parametreleri tarafından varsayılan değerlere bağlı olduğundan, bir iki terimli rasgele değişkenin ortalamasının ve varyansının da bu parametreler tarafından varsayılan değerlere bağlı olduğunu varsaymak makul görünmektedir. Aslında bu doğrudur ve Teorem 5.1'in ispatında n , p ve q 'nin fonksiyonları olarak herhangi bir iki terimli rasgele değişkenin ortalamasını ve varyansını hesaplamak için kullanılabilecek genel formüller elde ederiz.

TEOREM 5.1: Binom dağılımı $b(x; n, p)$ 'nin ortalaması ve varyansı:

$$\mu = n * p \text{ ve } \sigma^2 = n * p * q$$

Örnek 5.4: Belirli bir kırsal topluluktaki tüm içme kuyularının %30'unda bir kirlilik olduğu tahmin edilmektedir. Sorunun gerçek boyutunun anlaşılması için bazı testlerin gerekli olduğu belirlenmiştir. Bölgedeki tüm kuyuları test etmek çok pahalıdır, bu nedenle test için rastgele 10 kuyu seçilir.

(a) Binom dağılımını kullanarak, varsayımın doğru olduğunu varsayarak, tam olarak 3 kuyunun kirli olma olasılığı nedir?

(b) 3'ten fazla kuyunun kirli olasılığı nedir?

Cevap 5.4:

$$a.) b(3; 10, 0.3) = \sum_{x=0}^3 b(x; 10, 0.3) - \sum_{x=0}^2 b(x; 10, 0.3) = 0.6496 - 0.3828 = 0.2668$$

$$\text{Diğer çözüm: } P(X = 3) = b(3; 10, 0.3) = \binom{10}{3} (0.3)^3 (1 - 0.3)^{10-3} = 0.2668$$

$$b.) P(X > 3) = \sum_{x=4}^{10} b(x; 10, 0.3) = 1 - P(X \leq 3) = 1 - \sum_{x=0}^3 b(x; 10, 0.3) \\ = 1 - 0.6496 = 0.3504$$

Örnek 5.5: Örnek 5.2'deki iki terimli rastgele değişkenin ortalamasını ve varyansını bulun ve ardından $\mu \pm 2\sigma$ aralığını yorumlamak için Chebyshev teoremini kullanınız.

Cevap 5.5: Örnek 5.2, $n = 15$ ve $p = 0.4$ ile iki terimli bir deney olduğundan, Teorem 5.1'e göre,

$$\mu = n * p = 15 * 0.4 = 6 \text{ ve } \sigma^2 = n * p * q = 15 * 0.4 * (1 - 0.4) = 3.6$$

3.6'nın karekökünü alarak $\sigma = 1.897$ 'yi buluruz. Dolayısıyla gerekli aralık $6 \pm (2)(1.897)$ yani 2,206 ile 9,794 arasındadır. Chebyshev'in teoremi, hastalığa yakalanan 15 hasta arasındaki iyileşme sayısının en az 3/4 olasılıkla 2.206 ile 9.794 arasında veya veriler ayrık (kesikli) olduğu için 2 ile 10 arasında (dâhil) olma olasılığının olduğunu belirtir.

Örnek 5.6: Örnek 5.4'ün durumunu ele alalım. Kuyuların %30'unun kirli olduğu fikri, bölge su kurulu tarafından ileri sürülen bir varsayımdan başka bir şey değildir. 10 kuyunun rastgele seçildiğini ve 6 kuyunun kirlilik içerdiğinin bulunduğunu varsayalım. Bu, varsayım ne anlama geliyor? Bir olasılık ifadesi kullanınız.

Cevap 5.6: Öncelikle şunu sormalıyız: "Eğer varsayım doğruysa, 6 veya daha fazla kirli kuyu bulmamız olası mıdır?"

$$P(X \geq 6) = \sum_{x=0}^{10} b(x; 10, 0.3) - \sum_{x=0}^5 b(x; 10, 0.3) = 1 - 0.9527 = 0.0473$$

Sonuç olarak, tüm kuyuların yalnızca %30'u kirliyse, 6 veya daha fazla kuyunun kirli olması pek olası değildir (%4,7 şans). Bu, varsayım üzerine önemli ölçüde şüphe uyandırır ve kirlilik sorununun çok daha şiddetli olduğunu gösterir.

E. Çok Terimli Deneyler ve Çok Terimli Dağılım

Her denemenin ikiden fazla olası sonucu olmasına izin verirse, iki terimli deney çok terimli bir deney haline gelir. Üretilen bir ürünün hafif, ağır veya kabul edilebilir olarak sınıflandırılması ve belirli bir kavşakta meydana gelen kazaların haftanın gününe göre kaydedilmesi çok terimli deneyleri oluşturur.

Genel olarak, belirli bir deneme, $p_1, p_2, \dots, p_k$ olasılıkları ile k olası $E_1, E_2, \dots, E_k$ sonuçtan herhangi biri ile sonuçlanabiliyorsa, çok terimli dağılım n bağımsız denemede, E_1 'in x_1 kez, E_2 'nin x_2 kez ve E_k 'nin x_k kez meydana gelir meydana gelme olasılığını verecektir, burada

$$x_1 + x_2 + \dots + x_k = n$$

Bu ortak olasılık dağılımını şu şekilde göstereceğiz:

$$f(x_1, x_2, \dots, x_k; p_1, p_2, \dots, p_k, n)$$

Açıkça, $p_1 + p_2 + \dots + p_k = 1$, çünkü her denemenin sonucu k olası sonuçtan biri olmalıdır.

Genel formülü türetmek için binom durumunda olduğu gibi ilerliyoruz. Denemeler bağımsız olduğundan, E_1 için x_1 , E_2 için x_2 , $\dots$, E_k için x_k $p_1^{x_1} * p_2^{x_2} * \dots * p_k^{x_k}$ olasılıkla oluşacaktır. n deneme için benzer sonuçlar veren siparişlerin toplam sayısı, birinci grupta x_1 , ikinci grupta x_2 ve k . grupta x_k olmak üzere n ögenin k gruba ayrılma sayısına eşittir. Dolayısıyla bu,

$$\binom{n}{x_1, x_2, \dots, x_k} = \frac{n!}{x_1! x_2! \dots x_k!}$$

yöntemiyle yapılabilir. Tüm bölümler birbirini dışladığından (ayrık) ve eşit olasılıkla meydana geldiğinden, çok terimli dağılımı, belirli bir düzen için olasılığı toplam bölüm sayısı ile çarparak elde ederiz.

Bir deneme $p_1, p_2, \dots, p_k$ olasılıkları ile oluşan k tane $E_1, E_2, \dots, E_k$ sonuçlarından birisi ile sonuçlanıyorsa; n tane bağımsız denemede elde edilen $E_1, E_2, \dots, E_k$ sonuçlarının oluşma sayılarını temsil eden $X_1, X_2, \dots, X_k$ rastlantı değişkenlerinin olasılık dağılımı;

$$\sum_{i=1}^k x_i = n \text{ ve } \sum_{i=1}^k p_i = 1 \text{ olmak üzere}$$

$$f(x_1, x_2, \dots, x_k; p_1, p_2, \dots, p_k, n) = \binom{n}{x_1, x_2, \dots, x_k} p_1^{x_1} * p_2^{x_2} * \dots * p_k^{x_k}$$

Örnek 5.7: Bir havaalanına uçakların geliş ve gidişlerinin karmaşıklığı o kadar fazladır ki, "ideal" koşulları modellemek için genellikle bilgisayar simülasyonu kullanılır. 3 pisti olan belirli bir havaalanı için, ideal ortamda aşağıdakilerin her bir piste gelişigüzel gelen bir ticari jet tarafından erişilme olasılıkları olduğu bilinmektedir:

Pist 1: $p_1 = 2/9$,

Pist 2: $p_2 = 1/6$,

Pist 3: $p_3 = 11/18$.

Rastgele gelen 6 uçağın aşağıdaki şekilde dağılıma olasılığı nedir?

Pist 1: 2 uçak,

Pist 2: 1 uçak,

Pist 3: 3 uçak.

Cevap 5.7: Çok terimli dağılımı kullanarak,

$$\begin{aligned} f(x_1, x_2, \dots, x_k; p_1, p_2, \dots, p_k, n) &= \binom{n}{x_1, x_2, \dots, x_k} p_1^{x_1} * p_2^{x_2} * \dots * p_k^{x_k} \\ &= f\left(2, 1, 3; \frac{2}{9}, \frac{1}{6}, \frac{11}{18}, 6\right) = \binom{6}{2, 1, 3} \left(\frac{2}{9}\right)^2 * \left(\frac{1}{6}\right)^1 * \left(\frac{11}{18}\right)^3 \\ &= \frac{6!}{2! 1! 3!} \frac{2^2}{9^2} \frac{1^1}{6^1} \frac{11^3}{18^3} = 0.1127 \end{aligned}$$

PROBLEM 8/9.1: Antidepresanlarla ilgili tutumları inceleyen ulusal bir araştırma, yanıt verenlerin yaklaşık %70'inin "antidepresanların gerçekten hiçbir şeyi iyileştirmediğini, sadece asıl sorunu örttüğünü" düşündüğünü ortaya koymuştur. Bu araştırmaya göre, rastgele seçilen sonraki 5 kişiden en az 3'ünün bu görüşe sahip olma olasılığı nedir?

PROBLEM 8/9.2: Belirli bir seçim bölgesine herhangi bir delegenin uçak, otobüs, otomobil veya trenle gelme olasılığı sırasıyla 0.4, 0.1, 0.3 ve 0.2'dir. Bu toplantıya rastgele seçilen 10 delegeden 3'ünün uçakla, 3'ünün otobüsle, 2'sinin otomobille ve 2'sinin trenle gelme olasılığı nedir?

5.3 HİPERGEOMETRİK DAĞILIM

Bölüm 5.2'deki binom dağılımı ile hipergeometrik dağılım arasındaki farkı görmenin en basit yolu, örneklemenin nasıl yapıldığını not etmektir. Hipergeometrik için uygulama türleri, binom dağılımı için olanlara çok benzer. Belirli bir kategoriye giren gözlem sayısı için hesaplama olasılıklarıyla ilgileniyoruz. Ancak iki terimli dağılım söz konusu olduğunda, denemeler arasında bağımsızlık gereklidir. Sonuç olarak, bu dağılım, örneğin, çok sayıda kaleminden (kart destesi, üretim kalemleri partisi) numune almaya uygulanırsa, numune alma, gözlemlendikten sonra her bir kalemin değiştirilmesiyle yapılmalıdır. Öte yandan, hipergeometrik dağılım bağımsızlık gerektirmez ve değiştirmeden yapılan örneklemeye dayanır.

Hipergeometrik dağılım için uygulamalar, kabul örneklemede, elektronik testlerde ve kalite güvencesinde yoğun kullanımla birlikte birçok alanda bulunur. Açıkçası, bu alanların birçoğunda, test edilen öge pahasına test yapılır. Yani, öge yok edilir ve bu nedenle numunede değiştirilemez. Bu nedenle, değiştirmeden örnekleme gereklidir. Oyun kartlarıyla ilgili basit bir örnek, ilk örneğimiz olacak.

52 iskambillik sıradan bir desteden 5 çekilişte 3 kırmızı kart görme olasılığını bulmak istiyorsak, Bölüm 5.2'deki binom dağılımı, bir sonraki çekiliş yapılmadan önce her kart değiştirilmedikçe ve deste yeniden karıştırılmadıkça geçerli değildir. Değiştirmeden örnekleme problemini çözmek için, sorunu yeniden ifade edelim. Rastgele 5 kart çekilirse, destedeki mevcut 26 karttan 3 kırmızı kart ve destedeki 26 karttan 2 siyah kart seçme olasılığıyla ilgileniyoruz. 3 kırmızı kart seçmenin $\binom{26}{3}$ yolu vardır ve bu yolların her biri için $\binom{26}{2}$ yoldan 2 siyah kart seçebiliriz. Bu nedenle, 5 çekilişte 3 kırmızı ve 2 siyah kartı seçme yollarının toplam sayısı $\binom{26}{3} \binom{26}{2}$ 'dir. Mevcut 52 karttan herhangi 5 tanesini seçmenin toplam yolu $\binom{52}{5}$ 'tir. Bu nedenle, 3'ü kırmızı ve 2'si siyah olan 5 kartı değiştirmeden seçme olasılığı şu şekilde verilir:

$$\frac{\binom{26}{3} \binom{26}{2}}{\binom{52}{5}} = \frac{\frac{26!}{3!(26-3)!} \frac{26!}{2!(26-2)!}}{\frac{52!}{5!(52-5)!}} = 0.3251$$

Genel olarak, N öğeden n büyüklüğünde rastgele bir örnek seçildiğinde, başarılar olarak etiketlenmiş k öğeden k başarı ve başarısızlık olarak etiketlenmiş $N - k$ öğeden $n - x$ başarısızlık seçme olasılığıyla ilgileniyoruz. Bu, hipergeometrik bir deney olarak bilinir, yani aşağıdaki iki özelliğe sahiptir:

1. N öğeden değiştirilmeden n büyüklüğünde rastgele bir örnek seçilir. Yani N tane öğeden n tanesi seçildikten sonra tekrar yerine konmaksızın seçim yapmaya devam edilir.
2. N öğeden k tanesi başarı olarak sınıflandırılabilir ve $N - k$ başarısızlık olarak sınıflandırılır.

Hipergeometrik bir deneyin başarısının X sayısına hipergeometrik rasgele değişken denir. Buna göre, hipergeometrik değişkenin olasılık dağılımı hipergeometrik dağılım olarak adlandırılır ve değerleri $h(x; N, n, k)$ ile gösterilir.

Hipergeometrik rasgele değişken X 'in olasılık dağılımı, k 'nin başarı olarak etiklendiği ve $N - k$ 'nin başarısızlık olarak etiklendiği N öğeden seçilen n büyüklüğündeki rastgele bir örnekteki başarıların sayısı:

$$h(x; N, n, k) = \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}}$$

x 'in aralığı, tanımdaki üç binom katsayısı tarafından belirlenebilir; burada x ve $n - x$, sırasıyla k ve $N - k$ 'dan fazla değildir ve her ikisi de 0'dan küçük olamaz. Genellikle, her ikisi de k (başarı sayısı) ve $N - k$ (başarısızlık sayısı) örnek boyutu n 'den büyükse, hipergeometrik bir rasgele değişkenin aralığı $x = 0, 1, 2, \dots, n$ olacaktır.

Örnek 5.8: Enjeksiyon cihazı olarak kullanılan belirli bir parça 10'lu partiler halinde satılır. Partide birden fazla kusur yoksa üretici partiyi kabul edilebilir bulur. Bir numune alma planı, rasgele numune almayı ve 10 parçadan 3'ünün test edilmesini içerir. 3 parçadan hiçbirisi kusurlu değilse parti kabul edilir. Bu planın faydası hakkında yorum yapınız.

Cevap 5.8: Partinin gerçekten kabul edilemez olduğunu varsayalım (yani, 10 parçadan 2'sinin kusurlu olduğunu). Numune alma planının partiyi kabul edilebilir bulma olasılığı

$$P(X = 0) = \frac{\binom{2}{0} \binom{8}{3}}{\binom{10}{3}} = 0.467$$

Bu nedenle, 2 kusurlu parça ile parti gerçekten kabul edilemez ise, bu numune alma planı kabaca zamanın %47'sinde kabule izin verecektir. Sonuç olarak, bu plan hatalı kabul edilmelidir.

Örnek 5.9: 40 parçadan oluşan lotların her biri, 3 veya daha fazla kusur içeriyorsa kabul edilemez olarak kabul edilir. Bir partiden numune alma prosedürü, rasgele 5 bileşen seçmek ve kusurlu bulunursa partiyi reddetmektir. Partinin tamamında 3 kusur varsa, numunede tam olarak 1 kusurlu bulunma olasılığı nedir?

Cevap 5.9: $n = 5$, $N = 40$, $k = 3$ ve $x = 1$ ile hipergeometrik dağılımı kullanarak, 1 kusurlu elde etme olasılığını şu şekilde buluyoruz:

$$h(x; N, n, k) = \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}} = h(1; 40, 5, 3) = \frac{\binom{3}{1} \binom{40-3}{5-1}}{\binom{40}{5}} = 0.3011$$

Bir kez daha, bu plan, zamanın yalnızca yaklaşık %30'unda kötü bir parti (3 kusurlu) tespit ettiğinden arzu edilmez.

TEOREM 5.2: Hipergeometrik dağılımı $h(x; N, n, k)$ ortalaması ve varyansı:

$$\mu = \frac{nk}{N} \text{ ve } \sigma^2 = \frac{N-n}{N-1} * n * \frac{k}{N} * \left(1 - \frac{k}{N}\right)$$

Örnek 5.10: İstatistikçiler, partileri veya çok sayıda malzemeyi kabul etmek veya reddetmek için örnekleme planlarını kullanır. Bu örnekleme planlarından birinin, 12'si kusurlu olan 100 öğelik bir partiden 10 öğeyi bağımsız olarak örneklemeyi içerdiğini varsayalım. 12'si kusurlu olan 100 öğemiz var. 10'luk bir numunede 3'ünün kusurlu olma olasılığı nedir?

Cevap 5.10: Hipergeometrik olasılık fonksiyonunu kullanarak,

$$h(x; N, n, k) = \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}} = h(3; 100, 10, 12) = \frac{\binom{12}{3} \binom{88}{7}}{\binom{100}{10}} = 0.08$$

Örnek 5.11: Örnek 5.9'daki rastgele değişkenin ortalamasını ve varyansını bulun ve ardından $\mu \pm 2\sigma$ aralığını yorumlamak için Chebyshev teoremini kullanınız.

Cevap 5.11: Örnek 5.9, Teorem 5.2'ye göre $N = 40$, $n = 5$ ve $k = 3$ ile hipergeometrik bir deney olduğundan,

$$\mu = \frac{nk}{N} = \frac{5*3}{40} = 0.375 \text{ ve } \sigma^2 = \frac{N-n}{N-1} * n * \frac{k}{N} * \left(1 - \frac{k}{N}\right) = \frac{40-5}{40-1} * 5 * \frac{3}{40} * \left(1 - \frac{3}{40}\right) = 0.3113$$

0,3113'ün karekökünü alarak $\sigma = 0,558$ 'i buluruz. Bu nedenle, gerekli aralık $0,375 \pm (2)(0,558)$ veya -0,741 ile 1,491'dir. Chebyshev'in teoremi, 3'ü kusurlu olan 40 bileşenden rastgele 5 bileşen seçildiğinde elde edilen kusur sayısının, -0.741 ile 1.491 arasında düşme olasılığının en az 3/4 olduğunu belirtir. Yani, zamanın en az dörtte üçünde, 5 bileşen 2'den az kusur içerir.

A. Binom (İki Terimli) Dağılımı ile Hipergeometrik Dağılım Arasındaki İlişki

Bu bölümde, geniş uygulanabilirliği olan birkaç önemli ayırık (kesikli) dağıtımı tartışacağız. Bu dağıtımların çoğu birbiriyle güzel bir şekilde ilişkilidir. Yeni başlayan öğrenci bu ilişkiler hakkında net bir anlayış kazanmalıdır. Hipergeometrik ve binom dağılımı arasında ilginç bir ilişki vardır. Tahmin edilebileceği gibi, eğer n, N 'ye kıyasla küçükse, N ögenin doğası her çekilişte çok az değişir. Dolayısıyla, n, N 'ye kıyasla küçük olduğunda hipergeometrik dağılımı yaklaşık olarak tahmin etmek için bir binom dağılımı kullanılabilir. Aslında, pratik bir kural olarak, $\frac{n}{N} \leq 0.05$ olduğunda yaklaşım iyidir.

Böylece, $\frac{k}{N}$ miktarı p binom parametresinin rolünü oynar. Sonuç olarak, binom dağılımı, hipergeometrik dağılımın geniş nüfuslu bir versiyonu olarak görülebilir. Ortalama ve varyans;

$$\mu = n * p = \frac{nk}{N} \text{ ve } \sigma^2 = n * p * q = n * \frac{k}{N} * (1 - \frac{k}{N})$$

Bu formülleri Teorem 5.2'dekilerle karşılaştırdığımızda, ortalamanın aynı olduğunu ancak varyansın $\frac{(N-n)}{(N-1)}$ düzeltme faktörü kadar farklı olduğunu görüyoruz; bu, n, N 'ye göre küçük olduğunda ihmal edilebilir.

Örnek 5.12: Bir otomobil lastiği üreticisi, yerel bir distribütöre gönderilen 5000 adetlik bir sevkiyattan 1000 tanesinin hafifçe kusurlu olduğunu bildirdi. Distribütörden bu lastiklerden rastgele 10 tane alınırsa, tam olarak 3 tanesinin kusurlu olma olasılığı nedir?

Cevap 5.12: $N = 5000$, örneklem büyüklüğü $n = 10$ 'a göre büyük olduğundan, istenen olasılığı binom dağılımını kullanarak yaklaşık olarak hesaplayacağız. (Yani $\frac{n}{N} = \frac{10}{5000} = 0.02 < 0.05$ olduğundan binom dağılımı kullanılabilir.) Kusurlu bir lastik elde etme olasılığı 0.2 ($P = \frac{k}{N} = \frac{1000}{5000}$)'dir. Bu nedenle, tam olarak 3 kusurlu lastik elde etme olasılığı;

$$h(x; N, n, k) = \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}} = h(3; 5000, 10, 1000) = \frac{\binom{1000}{3} \binom{5000-1000}{10-3}}{\binom{5000}{10}} = 0.2015$$

$$\cong b(3; 10, 0.2) = \binom{10}{3} (0.2)^3 (1 - 0.2)^{10-3} = 0.2013$$

$$\cong \sum_{x=0}^{10} b(x; 10, 0.3) - \sum_{x=0}^5 b(x; 10, 0.3) = 0.8791 - 0.6778 = 0.2013$$

N öge, sırasıyla $a_1, a_2, \dots, a_k$ öğeleri olan k hücre $A_1, A_2, \dots, A_k$ 'ya bölünebilirse, o zaman n büyüklüğündeki rastgele bir örnekleme $A_1, A_2, \dots, A_k$ 'den seçilen öğelerin sayısını temsil eden $X_1, X_2, \dots, X_k$ rasgele değişkenlerinin olasılık dağılımı şu şekildedir:

$$\sum_{i=1}^k x_i = n \text{ ve } \sum_{i=1}^k p_i = 1 \text{ olmak üzere}$$

$$f(x_1, x_2, \dots, x_k; a_1, a_2, \dots, a_k, N, n) = \frac{\binom{a_1}{x_1} \binom{a_2}{x_2} \dots \binom{a_k}{x_k}}{\binom{N}{n}}$$

Örnek 5.13: Biyolojik bir vaka çalışması için 10 kişilik bir grup kullanılmaktadır. Grupta kan grubu O olan 3, A grubu olan 4 ve B grubu olan 3 kişi vardır. 5 kişiden oluşan rastgele bir örneğin 0 kan grubuna sahip 1 kişi, A kan grubuna sahip 2 kişi ve B kan grubuna sahip 2 kişiden oluşma olasılığı nedir?

Cevap 5.13: Hipergeometrik dağılımın $x_1 = 1, x_2 = 2, x_3 = 2, a_1 = 3, a_2 = 4, a_3 = 3, N = 10$ ve $n = 5$ ile uzantısını kullanarak, istenen olasılığı buluruz:

$$\begin{aligned} f(x_1, x_2, \dots, x_k; a_1, a_2, \dots, a_k, N, n) &= \frac{\binom{a_1}{x_1} \binom{a_2}{x_2} \dots \binom{a_k}{x_k}}{\binom{N}{n}} \\ &= f(1, 2, 2; 3, 4, 3, 10, 5) = \frac{\binom{3}{1} \binom{4}{2} \binom{3}{2}}{\binom{10}{5}} = \frac{3}{14} \end{aligned}$$

EK A.1

Table A.1 Binomial Probability Sums $\sum_{x=0}^r b(x; n, p)$

n	r	P									
		0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80	0.90
1	0	0.9000	0.8000	0.7500	0.7000	0.6000	0.5000	0.4000	0.3000	0.2000	0.1000
	1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	0	0.8100	0.6400	0.5625	0.4900	0.3600	0.2500	0.1600	0.0900	0.0400	0.0100
	1	0.9900	0.9600	0.9375	0.9100	0.8400	0.7500	0.6400	0.5100	0.3600	0.1900
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	0.7290	0.5120	0.4219	0.3430	0.2160	0.1250	0.0640	0.0270	0.0080	0.0010
	1	0.9720	0.8960	0.8438	0.7840	0.6480	0.5000	0.3520	0.2160	0.1040	0.0280
	2	0.9990	0.9920	0.9844	0.9730	0.9360	0.8750	0.7840	0.6570	0.4880	0.2710
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	0.6561	0.4096	0.3164	0.2401	0.1296	0.0625	0.0256	0.0081	0.0016	0.0001
	1	0.9477	0.8192	0.7383	0.6517	0.4752	0.3125	0.1792	0.0837	0.0272	0.0037
	2	0.9963	0.9728	0.9492	0.9163	0.8208	0.6875	0.5248	0.3483	0.1808	0.0523
	3	0.9999	0.9984	0.9961	0.9919	0.9744	0.9375	0.8704	0.7599	0.5904	0.3439
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	0	0.5905	0.3277	0.2373	0.1681	0.0778	0.0313	0.0102	0.0024	0.0003	0.0000
	1	0.9185	0.7373	0.6328	0.5282	0.3370	0.1875	0.0870	0.0308	0.0067	0.0005
	2	0.9914	0.9421	0.8965	0.8369	0.6826	0.5000	0.3174	0.1631	0.0579	0.0086
	3	0.9995	0.9933	0.9844	0.9692	0.9130	0.8125	0.6630	0.4718	0.2627	0.0815
	4	1.0000	0.9997	0.9990	0.9976	0.9898	0.9688	0.9222	0.8319	0.6723	0.4095
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	0	0.5314	0.2621	0.1780	0.1176	0.0467	0.0156	0.0041	0.0007	0.0001	0.0000
	1	0.8857	0.6554	0.5339	0.4202	0.2333	0.1094	0.0410	0.0109	0.0016	0.0001
	2	0.9842	0.9011	0.8306	0.7443	0.5443	0.3438	0.1792	0.0705	0.0170	0.0013
	3	0.9987	0.9830	0.9624	0.9295	0.8208	0.6563	0.4557	0.2557	0.0989	0.0159
	4	0.9999	0.9984	0.9954	0.9891	0.9590	0.8906	0.7667	0.5798	0.3446	0.1143
	5	1.0000	0.9999	0.9998	0.9993	0.9959	0.9844	0.9533	0.8824	0.7379	0.4686
	6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
7	0	0.4783	0.2097	0.1335	0.0824	0.0280	0.0078	0.0016	0.0002	0.0000	
	1	0.8503	0.5767	0.4449	0.3294	0.1586	0.0625	0.0188	0.0038	0.0004	0.0000
	2	0.9743	0.8520	0.7564	0.6471	0.4199	0.2266	0.0963	0.0288	0.0047	0.0002
	3	0.9973	0.9667	0.9294	0.8740	0.7102	0.5000	0.2898	0.1260	0.0333	0.0027
	4	0.9998	0.9953	0.9871	0.9712	0.9037	0.7734	0.5801	0.3529	0.1480	0.0257
	5	1.0000	0.9996	0.9987	0.9962	0.9812	0.9375	0.8414	0.6706	0.4233	0.1497
	6		1.0000	0.9999	0.9998	0.9984	0.9922	0.9720	0.9176	0.7903	0.5217
	7			1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table A.1 (continued) Binomial Probability Sums $\sum_{x=0}^r b(x; n, p)$

<i>n</i>	<i>r</i>	<i>p</i>									
		0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80	0.90
8	0	0.4305	0.1678	0.1001	0.0576	0.0168	0.0039	0.0007	0.0001	0.0000	
	1	0.8131	0.5033	0.3671	0.2553	0.1064	0.0352	0.0085	0.0013	0.0001	
	2	0.9619	0.7969	0.6785	0.5518	0.3154	0.1445	0.0498	0.0113	0.0012	0.0000
	3	0.9950	0.9437	0.8862	0.8059	0.5941	0.3633	0.1737	0.0580	0.0104	0.0004
	4	0.9996	0.9896	0.9727	0.9420	0.8263	0.6367	0.4059	0.1941	0.0563	0.0050
	5	1.0000	0.9988	0.9958	0.9887	0.9502	0.8555	0.6846	0.4482	0.2031	0.0381
	6		0.9999	0.9996	0.9987	0.9915	0.9648	0.8936	0.7447	0.4967	0.1869
	7		1.0000	1.0000	0.9999	0.9993	0.9961	0.9832	0.9424	0.8322	0.5695
	8				1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	0	0.3874	0.1342	0.0751	0.0404	0.0101	0.0020	0.0003	0.0000		
	1	0.7748	0.4362	0.3003	0.1960	0.0705	0.0195	0.0038	0.0004	0.0000	
	2	0.9470	0.7382	0.6007	0.4628	0.2318	0.0898	0.0250	0.0043	0.0003	0.0000
	3	0.9917	0.9144	0.8343	0.7297	0.4826	0.2539	0.0994	0.0253	0.0031	0.0001
	4	0.9991	0.9804	0.9511	0.9012	0.7334	0.5000	0.2666	0.0988	0.0196	0.0009
	5	0.9999	0.9969	0.9900	0.9747	0.9006	0.7461	0.5174	0.2703	0.0856	0.0083
	6	1.0000	0.9997	0.9987	0.9957	0.9750	0.9102	0.7682	0.5372	0.2618	0.0530
	7		1.0000	0.9999	0.9996	0.9962	0.9805	0.9295	0.8040	0.5638	0.2252
	8			1.0000	1.0000	0.9997	0.9980	0.9899	0.9596	0.8658	0.6126
	9					1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	0	0.3487	0.1074	0.0563	0.0282	0.0060	0.0010	0.0001	0.0000		
	1	0.7361	0.3758	0.2440	0.1493	0.0464	0.0107	0.0017	0.0001	0.0000	
	2	0.9298	0.6778	0.5256	0.3828	0.1673	0.0547	0.0123	0.0016	0.0001	
	3	0.9872	0.8791	0.7759	0.6496	0.3823	0.1719	0.0548	0.0106	0.0009	0.0000
	4	0.9984	0.9672	0.9219	0.8497	0.6331	0.3770	0.1662	0.0473	0.0064	0.0001
	5	0.9999	0.9936	0.9803	0.9527	0.8338	0.6230	0.3669	0.1503	0.0328	0.0016
	6	1.0000	0.9991	0.9965	0.9894	0.9452	0.8281	0.6177	0.3504	0.1209	0.0128
	7		0.9999	0.9996	0.9984	0.9877	0.9453	0.8327	0.6172	0.3222	0.0702
	8		1.0000	1.0000	0.9999	0.9983	0.9893	0.9536	0.8507	0.6242	0.2639
	9				1.0000	0.9999	0.9990	0.9940	0.9718	0.8926	0.6513
	10					1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
11	0	0.3138	0.0859	0.0422	0.0198	0.0036	0.0005	0.0000			
	1	0.6974	0.3221	0.1971	0.1130	0.0302	0.0059	0.0007	0.0000		
	2	0.9104	0.6174	0.4552	0.3127	0.1189	0.0327	0.0059	0.0006	0.0000	
	3	0.9815	0.8389	0.7133	0.5696	0.2963	0.1133	0.0293	0.0043	0.0002	
	4	0.9972	0.9496	0.8854	0.7897	0.5328	0.2744	0.0994	0.0216	0.0020	0.0000
	5	0.9997	0.9883	0.9657	0.9218	0.7535	0.5000	0.2465	0.0782	0.0117	0.0003
	6	1.0000	0.9980	0.9924	0.9784	0.9006	0.7256	0.4672	0.2103	0.0504	0.0028
	7		0.9998	0.9988	0.9957	0.9707	0.8867	0.7037	0.4304	0.1611	0.0185
	8		1.0000	0.9999	0.9994	0.9941	0.9673	0.8811	0.6873	0.3826	0.0896
	9			1.0000	1.0000	0.9993	0.9941	0.9698	0.8870	0.6779	0.3026
	10					1.0000	0.9995	0.9964	0.9802	0.9141	0.6862
	11						1.0000	1.0000	1.0000	1.0000	1.0000

Table A.1 (continued) Binomial Probability Sums $\sum_{x=0}^r b(x; n, p)$

		p									
n	r	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80	0.90
12	0	0.2824	0.0687	0.0317	0.0138	0.0022	0.0002	0.0000			
	1	0.6590	0.2749	0.1584	0.0850	0.0196	0.0032	0.0003	0.0000		
	2	0.8891	0.5583	0.3907	0.2528	0.0834	0.0193	0.0028	0.0002	0.0000	
	3	0.9744	0.7946	0.6488	0.4925	0.2253	0.0730	0.0153	0.0017	0.0001	
	4	0.9957	0.9274	0.8424	0.7237	0.4382	0.1938	0.0573	0.0095	0.0006	0.0000
	5	0.9995	0.9806	0.9456	0.8822	0.6652	0.3872	0.1582	0.0386	0.0039	0.0001
	6	0.9999	0.9961	0.9857	0.9614	0.8418	0.6128	0.3348	0.1178	0.0194	0.0005
	7	1.0000	0.9994	0.9972	0.9905	0.9427	0.8062	0.5618	0.2763	0.0726	0.0043
	8		0.9999	0.9996	0.9983	0.9847	0.9270	0.7747	0.5075	0.2054	0.0256
	9		1.0000	1.0000	0.9998	0.9972	0.9807	0.9166	0.7472	0.4417	0.1109
	10				1.0000	0.9997	0.9968	0.9804	0.9150	0.7251	0.3410
	11					1.0000	0.9998	0.9978	0.9862	0.9313	0.7176
12						1.0000	1.0000	1.0000	1.0000	1.0000	
13	0	0.2542	0.0550	0.0238	0.0097	0.0013	0.0001	0.0000			
	1	0.6213	0.2336	0.1267	0.0637	0.0126	0.0017	0.0001	0.0000		
	2	0.8661	0.5017	0.3326	0.2025	0.0579	0.0112	0.0013	0.0001		
	3	0.9658	0.7473	0.5843	0.4206	0.1686	0.0461	0.0078	0.0007	0.0000	
	4	0.9935	0.9009	0.7940	0.6543	0.3530	0.1334	0.0321	0.0040	0.0002	
	5	0.9991	0.9700	0.9198	0.8346	0.5744	0.2905	0.0977	0.0182	0.0012	0.0000
	6	0.9999	0.9930	0.9757	0.9376	0.7712	0.5000	0.2288	0.0624	0.0070	0.0001
	7	1.0000	0.9988	0.9944	0.9818	0.9023	0.7095	0.4256	0.1654	0.0300	0.0009
	8		0.9998	0.9990	0.9960	0.9679	0.8666	0.6470	0.3457	0.0991	0.0065
	9		1.0000	0.9999	0.9993	0.9922	0.9539	0.8314	0.5794	0.2527	0.0342
	10			1.0000	0.9999	0.9987	0.9888	0.9421	0.7975	0.4983	0.1339
	11				1.0000	0.9999	0.9983	0.9874	0.9363	0.7664	0.3787
	12					1.0000	0.9999	0.9987	0.9903	0.9450	0.7458
	13						1.0000	1.0000	1.0000	1.0000	1.0000
14	0	0.2288	0.0440	0.0178	0.0068	0.0008	0.0001	0.0000			
	1	0.5846	0.1979	0.1010	0.0475	0.0081	0.0009	0.0001			
	2	0.8416	0.4481	0.2811	0.1608	0.0398	0.0065	0.0006	0.0000		
	3	0.9559	0.6982	0.5213	0.3552	0.1243	0.0287	0.0039	0.0002		
	4	0.9908	0.8702	0.7415	0.5842	0.2793	0.0898	0.0175	0.0017	0.0000	
	5	0.9985	0.9561	0.8883	0.7805	0.4859	0.2120	0.0583	0.0083	0.0004	
	6	0.9998	0.9884	0.9617	0.9067	0.6925	0.3953	0.1501	0.0315	0.0024	0.0000
	7	1.0000	0.9976	0.9897	0.9685	0.8499	0.6047	0.3075	0.0933	0.0116	0.0002
	8		0.9996	0.9978	0.9917	0.9417	0.7880	0.5141	0.2195	0.0439	0.0015
	9		1.0000	0.9997	0.9983	0.9825	0.9102	0.7207	0.4158	0.1298	0.0092
	10			1.0000	0.9998	0.9961	0.9713	0.8757	0.6448	0.3018	0.0441
	11				1.0000	0.9994	0.9935	0.9602	0.8392	0.5519	0.1584
	12					0.9999	0.9991	0.9919	0.9525	0.8021	0.4154
	13					1.0000	0.9999	0.9992	0.9932	0.9560	0.7712
	14						1.0000	1.0000	1.0000	1.0000	1.0000

Table A.1 (continued) Binomial Probability Sums $\sum_{x=0}^r b(x; n, p)$

<i>n</i>	<i>r</i>	<i>p</i>									
		0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80	0.90
15	0	0.2059	0.0352	0.0134	0.0047	0.0005	0.0000				
	1	0.5490	0.1671	0.0802	0.0353	0.0052	0.0005	0.0000			
	2	0.8159	0.3980	0.2361	0.1268	0.0271	0.0037	0.0003	0.0000		
	3	0.9444	0.6482	0.4613	0.2969	0.0905	0.0176	0.0019	0.0001		
	4	0.9873	0.8358	0.6865	0.5155	0.2173	0.0592	0.0093	0.0007	0.0000	
	5	0.9978	0.9389	0.8516	0.7216	0.4032	0.1509	0.0338	0.0037	0.0001	
	6	0.9997	0.9819	0.9434	0.8689	0.6098	0.3036	0.0950	0.0152	0.0008	
	7	1.0000	0.9958	0.9827	0.9500	0.7869	0.5000	0.2131	0.0500	0.0042	0.0000
	8		0.9992	0.9958	0.9848	0.9050	0.6964	0.3902	0.1311	0.0181	0.0003
	9		0.9999	0.9992	0.9963	0.9662	0.8491	0.5968	0.2784	0.0611	0.0022
	10		1.0000	0.9999	0.9993	0.9907	0.9408	0.7827	0.4845	0.1642	0.0127
	11			1.0000	0.9999	0.9981	0.9824	0.9095	0.7031	0.3518	0.0556
	12				1.0000	0.9997	0.9963	0.9729	0.8732	0.6020	0.1841
	13					1.0000	0.9995	0.9948	0.9647	0.8329	0.4510
	14						1.0000	0.9995	0.9953	0.9648	0.7941
	15							1.0000	1.0000	1.0000	1.0000
16	0	0.1853	0.0281	0.0100	0.0033	0.0003	0.0000				
	1	0.5147	0.1407	0.0635	0.0261	0.0033	0.0003	0.0000			
	2	0.7892	0.3518	0.1971	0.0994	0.0183	0.0021	0.0001			
	3	0.9316	0.5981	0.4050	0.2459	0.0651	0.0106	0.0009	0.0000		
	4	0.9830	0.7982	0.6302	0.4499	0.1666	0.0384	0.0049	0.0003		
	5	0.9967	0.9183	0.8103	0.6598	0.3288	0.1051	0.0191	0.0016	0.0000	
	6	0.9995	0.9733	0.9204	0.8247	0.5272	0.2272	0.0583	0.0071	0.0002	
	7	0.9999	0.9930	0.9729	0.9256	0.7161	0.4018	0.1423	0.0257	0.0015	0.0000
	8	1.0000	0.9985	0.9925	0.9743	0.8577	0.5982	0.2839	0.0744	0.0070	0.0001
	9		0.9998	0.9984	0.9929	0.9417	0.7728	0.4728	0.1753	0.0267	0.0005
	10		1.0000	0.9997	0.9984	0.9809	0.8949	0.6712	0.3402	0.0817	0.0033
	11			1.0000	0.9997	0.9951	0.9616	0.8334	0.5501	0.2018	0.0170
	12				1.0000	0.9991	0.9894	0.9349	0.7541	0.4019	0.0684
	13					0.9999	0.9979	0.9817	0.9006	0.6482	0.2108
	14					1.0000	0.9997	0.9967	0.9739	0.8593	0.4853
	15						1.0000	0.9997	0.9967	0.9719	0.8147
	16							1.0000	1.0000	1.0000	1.0000

Table A.1 (continued) Binomial Probability Sums $\sum_{x=0}^r b(x; n, p)$

<i>n</i>	<i>r</i>	<i>p</i>									
		0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80	0.90
17	0	0.1668	0.0225	0.0075	0.0023	0.0002	0.0000				
	1	0.4818	0.1182	0.0501	0.0193	0.0021	0.0001	0.0000			
	2	0.7618	0.3096	0.1637	0.0774	0.0123	0.0012	0.0001			
	3	0.9174	0.5489	0.3530	0.2019	0.0464	0.0064	0.0005	0.0000		
	4	0.9779	0.7582	0.5739	0.3887	0.1260	0.0245	0.0025	0.0001		
	5	0.9953	0.8943	0.7653	0.5968	0.2639	0.0717	0.0106	0.0007	0.0000	
	6	0.9992	0.9623	0.8929	0.7752	0.4478	0.1662	0.0348	0.0032	0.0001	
	7	0.9999	0.9891	0.9598	0.8954	0.6405	0.3145	0.0919	0.0127	0.0005	
	8	1.0000	0.9974	0.9876	0.9597	0.8011	0.5000	0.1989	0.0403	0.0026	0.0000
	9		0.9995	0.9969	0.9873	0.9081	0.6855	0.3595	0.1046	0.0109	0.0001
	10		0.9999	0.9994	0.9968	0.9652	0.8338	0.5522	0.2248	0.0377	0.0008
	11		1.0000	0.9999	0.9993	0.9894	0.9283	0.7361	0.4032	0.1057	0.0047
	12			1.0000	0.9999	0.9975	0.9755	0.8740	0.6113	0.2418	0.0221
	13				1.0000	0.9995	0.9936	0.9536	0.7981	0.4511	0.0826
	14					0.9999	0.9988	0.9877	0.9226	0.6904	0.2382
	15					1.0000	0.9999	0.9979	0.9807	0.8818	0.5182
	16						1.0000	0.9998	0.9977	0.9775	0.8332
	17							1.0000	1.0000	1.0000	1.0000
18	0	0.1501	0.0180	0.0056	0.0016	0.0001	0.0000				
	1	0.4503	0.0991	0.0395	0.0142	0.0013	0.0001				
	2	0.7338	0.2713	0.1353	0.0600	0.0082	0.0007	0.0000			
	3	0.9018	0.5010	0.3057	0.1646	0.0328	0.0038	0.0002			
	4	0.9718	0.7164	0.5187	0.3327	0.0942	0.0154	0.0013	0.0000		
	5	0.9936	0.8671	0.7175	0.5344	0.2088	0.0481	0.0058	0.0003		
	6	0.9988	0.9487	0.8610	0.7217	0.3743	0.1189	0.0203	0.0014	0.0000	
	7	0.9998	0.9837	0.9431	0.8593	0.5634	0.2403	0.0576	0.0061	0.0002	
	8	1.0000	0.9957	0.9807	0.9404	0.7368	0.4073	0.1347	0.0210	0.0009	
	9		0.9991	0.9946	0.9790	0.8653	0.5927	0.2632	0.0596	0.0043	0.0000
	10		0.9998	0.9988	0.9939	0.9424	0.7597	0.4366	0.1407	0.0163	0.0002
	11		1.0000	0.9998	0.9986	0.9797	0.8811	0.6257	0.2783	0.0513	0.0012
	12			1.0000	0.9997	0.9942	0.9519	0.7912	0.4656	0.1329	0.0064
	13				1.0000	0.9987	0.9846	0.9058	0.6673	0.2836	0.0282
	14					0.9998	0.9962	0.9672	0.8354	0.4990	0.0982
	15					1.0000	0.9993	0.9918	0.9400	0.7287	0.2662
	16						0.9999	0.9987	0.9858	0.9009	0.5497
	17						1.0000	0.9999	0.9984	0.9820	0.8499
	18							1.0000	1.0000	1.0000	1.0000

Table A.1 (continued) Binomial Probability Sums $\sum_{x=0}^r b(x; n, p)$

<i>n</i>	<i>r</i>	<i>p</i>									
		0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80	0.90
19	0	0.1351	0.0144	0.0042	0.0011	0.0001					
	1	0.4203	0.0829	0.0310	0.0104	0.0008	0.0000				
	2	0.7054	0.2369	0.1113	0.0462	0.0055	0.0004	0.0000			
	3	0.8850	0.4551	0.2631	0.1332	0.0230	0.0022	0.0001			
	4	0.9648	0.6733	0.4654	0.2822	0.0696	0.0096	0.0006	0.0000		
	5	0.9914	0.8369	0.6678	0.4739	0.1629	0.0318	0.0031	0.0001		
	6	0.9983	0.9324	0.8251	0.6655	0.3081	0.0835	0.0116	0.0006		
	7	0.9997	0.9767	0.9225	0.8180	0.4878	0.1796	0.0352	0.0028	0.0000	
	8	1.0000	0.9933	0.9713	0.9161	0.6675	0.3238	0.0885	0.0105	0.0003	
	9		0.9984	0.9911	0.9674	0.8139	0.5000	0.1861	0.0326	0.0016	
	10		0.9997	0.9977	0.9895	0.9115	0.6762	0.3325	0.0839	0.0067	0.0000
	11		1.0000	0.9995	0.9972	0.9648	0.8204	0.5122	0.1820	0.0233	0.0003
	12			0.9999	0.9994	0.9884	0.9165	0.6919	0.3345	0.0676	0.0017
	13			1.0000	0.9999	0.9969	0.9682	0.8371	0.5261	0.1631	0.0086
	14				1.0000	0.9994	0.9904	0.9304	0.7178	0.3267	0.0352
	15					0.9999	0.9978	0.9770	0.8668	0.5449	0.1150
	16					1.0000	0.9996	0.9945	0.9538	0.7631	0.2946
	17						1.0000	0.9992	0.9896	0.9171	0.5797
	18							0.9999	0.9989	0.9856	0.8649
	19							1.0000	1.0000	1.0000	1.0000
20	0	0.1216	0.0115	0.0032	0.0008	0.0000					
	1	0.3917	0.0692	0.0243	0.0076	0.0005	0.0000				
	2	0.6769	0.2061	0.0913	0.0355	0.0036	0.0002				
	3	0.8670	0.4114	0.2252	0.1071	0.0160	0.0013	0.0000			
	4	0.9568	0.6296	0.4148	0.2375	0.0510	0.0059	0.0003			
	5	0.9887	0.8042	0.6172	0.4164	0.1256	0.0207	0.0016	0.0000		
	6	0.9976	0.9133	0.7858	0.6080	0.2500	0.0577	0.0065	0.0003		
	7	0.9996	0.9679	0.8982	0.7723	0.4159	0.1316	0.0210	0.0013	0.0000	
	8	0.9999	0.9900	0.9591	0.8867	0.5956	0.2517	0.0565	0.0051	0.0001	
	9	1.0000	0.9974	0.9861	0.9520	0.7553	0.4119	0.1275	0.0171	0.0006	
	10		0.9994	0.9961	0.9829	0.8725	0.5881	0.2447	0.0480	0.0026	0.0000
	11		0.9999	0.9991	0.9949	0.9435	0.7483	0.4044	0.1133	0.0100	0.0001
	12		1.0000	0.9998	0.9987	0.9790	0.8684	0.5841	0.2277	0.0321	0.0004
	13			1.0000	0.9997	0.9935	0.9423	0.7500	0.3920	0.0867	0.0024
	14				1.0000	0.9984	0.9793	0.8744	0.5836	0.1958	0.0113
	15					0.9997	0.9941	0.9490	0.7625	0.3704	0.0432
	16					1.0000	0.9987	0.9840	0.8929	0.5886	0.1330
	17						0.9998	0.9964	0.9645	0.7939	0.3231
	18						1.0000	0.9995	0.9924	0.9308	0.6083
	19							1.0000	0.9992	0.9885	0.8784
	20								1.0000	1.0000	1.0000

OLASILIK VE İSTATİSTİK

10. HAFTA

BAZI KESİKLİ (AYRIK) OLASILIK DAĞILIMLARI (DEVAM)

5.4 NEGATİF BİNOM (İKİ TERİMLİ) DAĞILIM

Özelliklerin iki terimli bir deney için listelenenlerle aynı olduğu bir deneyi düşünelim, tek istisna, denemelerin sabit sayıda başarı elde edilene kadar tekrarlanmasıdır. Bu nedenle, n 'nin sabit olduğu n denemede x 'in başarı olasılığı yerine, artık k 'ıncı başarının x 'inci denemede meydana gelme olasılığıyla ilgileniyoruz. **Bu tür deneylere negatif binom deneyleri denir.**

Örnek olarak, kullanıldığı vakaların %60'ında etkili olduğu bilinen bir ilacın kullanımını düşünün. İlaç, hastaya bir dereceye kadar rahatlama sağlamada etkiliyse, başarılı olarak kabul edilecektir. Rahatlama yaşayan beşinci hastanın, belirli bir hafta içinde ilacı alan yedinci hasta olma olasılığını bulmakla ilgileniyoruz. Başarıyı S ile ve başarısızlığı F ile belirterek, istenen sonuca ulaşmanın olası bir sırası,

$$(0.6)(0.4)(0.6)(0.6)(0.6)(0.4)(0.6) = (0.6)^5(0.4)^2.$$

olasılıkla meydana gelen SFSSSFS'dir.

Beşinci başarı olması gereken son sonuç dışında, F'leri ve S'leri yeniden düzenleyerek olası tüm sıralamaları listeleyebiliriz. Olası siparişlerin toplam sayısı, bir gruba 2 başarısızlık ve diğer gruba atanan 4 başarı ile ilk altı denemenin iki gruba ayrılma sayısına eşittir. Bu, birbirini dışlayan $\binom{6}{4} = 15$ yolla yapılabilir. Dolayısıyla X, beşinci başarının gerçekleştiği sonucu temsil ediyorsa, o zaman

$$P(X = 7) = \binom{6}{4} (0.6)^5 (0.4)^2 = 0.1866.$$

A. Negatif Binom Rastgele Değişken Nedir?

Bir negatif iki terimli deneyde k başarı elde etmek için gereken X deneme sayısına negatif iki terimli rasgele değişken denir ve bunun olasılık dağılımına negatif iki terimli dağılım denir. Olasılıkları istenen başarı sayısına ve belirli bir denemede başarı olasılığına bağlı olduğundan, bunları $b^*(x; k, p)$ ile göstereceğiz. $b^*(x; k, p)$ genel formülünü elde etmek için, belirli bir sırayla $k - 1$ başarı ve $x - k$ başarısızlıktan önceki x 'inci denemede başarı olasılığını göz önünde bulundurun. Denemeler bağımsız olduğu için, istenen her sonuca karşılık gelen tüm

olasılıkları çarpabiliriz. Her başarı p olasılıkla ve her başarısızlık $q = 1 - p$ olasılıkla gerçekleşir. Bu nedenle, belirtilen siparişin başarıyla sonuçlanma olasılığı

$$p^{k-1} q^{x-k} p = p^k q^{x-k}.$$

Herhangi bir sırayla $k-1$ başarı ve $x-k$ başarısızlık meydana geldikten sonra, başarı ile biten deneydeki toplam örnek noktası sayısı, $x-1$ denemelerin $k-1$ ile iki gruba ayrılma sayısına eşittir. bir gruba karşılık gelen başarı ve diğer gruba karşılık gelen $x-k$ başarısızlık. Bu sayı $\binom{x-1}{k-1}$ terimiyle belirtilir, her biri birbirini dışlar ve eşit olasılıkla $p^k q^{x-k}$ oluşur. $p^k q^{x-k}$ 'yi $\binom{x-1}{k-1}$ ile çarparak genel formülü elde ederiz.

Tekrarlanan bağımsız denemeler, p olasılıkla başarılı ve $q = 1 - p$ olasılıkla başarısızlıkla sonuçlanabiliyorsa, o zaman k . başarının gerçekleştiği deneme sayısı X rasgele değişkeninin olasılık dağılımı, ,

$$b * (x; k, p) = \binom{x-1}{k-1} p^k q^{x-k}, x = k, k + 1, k + 2, \dots$$

Örnek 5.14: Bir NBA (Ulusal Basketbol Birliği) şampiyonluk serisinde, yedi maçın dördünü kazanan takım galip gelir. A ve B takımlarının şampiyonluk maçlarında karşı karşıya geldiğini ve A takımının B takımına karşı bir oyunu kazanma olasılığının 0.55 olduğunu varsayalım.

- a.) A takımının seriyi 6 maçta kazanma olasılığı nedir?
- b.) A takımının seriyi kazanma olasılığı nedir?
- c.) A ve B takımları, 5 maçın 3'ünü kazanarak belirlenen bölgesel bir eleme serisinde karşı karşıya gelseydi, A takımının seriyi kazanma olasılığı nedir?

Cevap 5.14:

$$a.) b * (x; k, p) = \binom{x-1}{k-1} p^k q^{x-k} = b * (6; 4, 0.55) = \binom{6-1}{4-1} (0.55)^4 (1 - 0.55)^{6-4} = 0.1853$$

b.) P(A takımını şampiyonluk serisini kazanır)

$$b * (4; 4, 0.55) + b * (5; 4, 0.55) + b * (6; 4, 0.55) + b * (7; 4, 0.55) \\ 0.0915 + 0.1647 + 0.1853 + 0.1668 = 0.6083.$$

c.) P(A takımını elemeyi kazanır)

$$b * (3; 3, 0.55) + b * (4; 3, 0.55) + b * (5; 3, 0.55)$$

Negatif binom dağılımı, adını $p^k q^{-k}$ açılımındaki her terimin $x = k, k + 1, k + 2, \dots$ için $b * (x; k, p)$ değerlerine karşılık gelmesinden alır. Negatif binom dağılımının $k = 1$ olduğu özel durumunu ele alırsak, tek bir başarı için gereken deneme sayısı için bir olasılık dağılımımız olur. Bir örnek, bir tura gelene kadar bir madeni paranın atılması olabilir. İlk turanın dördüncü atışta gelme olasılığı ilgimizi çekebilir. Negatif binom dağılımı

$$b * (x; 1, p) = p q^{x-1}, \quad x = 1, 2, 3, \dots$$

formuna indirgenir.

Ardışık terimler geometrik bir ilerleme oluşturduğundan, bu özel durumu geometrik dağılım olarak adlandırmak ve değerlerini $g(x; p)$ ile gösterilmektedir.

Tekrarlanan bağımsız denemeler, p olasılıkla başarılı ve $q = 1 - p$ olasılıkla başarısızlıkla sonuçlanabiliyorsa, ilk başarının meydana geldiği denemenin sayısı X rasgele değişkeninin olasılık dağılımı,

$$g(x; p) = p q^{x-1}, \quad x = 1, 2, 3, \dots$$

Örnek 5.15: Belirli bir üretim süreci için ortalama olarak her 100 parçadan 1'inin kusurlu olduğu bilinmektedir. İncelenen beşinci parçanın bulunan ilk kusurlu parça olma olasılığı nedir?

Cevap 5.15: $x = 5$ ve $p = 0.01$ ile geometrik dağılımı kullanarak,

$$g(x; p) = p q^{x-1} = g(5; 0.01) = (0.01)(1 - 0.01)^{5-1} = 0.0096.$$

Örnek 5.16: "Yoğun bir zamanda" bir telefon santrali kapasitesine çok yakındır, bu nedenle arayanlar aramalarını yapmakta güçlük çekerler. Bir bağlantı kurmak için gereken deneme sayısını bilmek ilgi çekici olabilir. Meşgul bir süre boyunca bağlantı olasılığını $p = 0.05$ olarak kabul ettiğimizi varsayalım. Başarılı bir çağrı için 5 denemenin gerekli olma olasılığını bulunuz.

Cevap 5.16: $x = 5$ ve $p = 0.05$ verimleri ile geometrik dağılımı kullanarak;

$$g(x; p) = p q^{x-1} = g(5; 0.05) = (0.05)(1 - 0.05)^{5-1} = 0.041.$$

Çoğu zaman, geometrik dağılımla ilgili uygulamalarda ortalama ve varyans önemlidir. Örneğin, Örnek 5.16'da, bağlantı kurmak için gerekli olan beklenen arama sayısı oldukça önemlidir.

Aşağıdaki teorem, ispat olmaksızın geometrik dağılımın ortalamasını ve varyansını belirtmektedir.

TEOREM 5.3: Geometrik dağılımı izleyen rastgele bir değişkenin ortalaması ve varyansı;

$$\mu = \frac{1}{p} \text{ ve } \sigma^2 = \frac{1-p}{p^2}$$

5.5 POISSON DAĞILIMI VE POISSON SÜRECİ

X rasgele değişkeninin sayısal değerlerini, belirli bir zaman aralığında veya belirli bir bölgede meydana gelen sonuçların sayısını veren deneylere Poisson deneyleri denir. Belirli bir zaman aralığı, bir dakika, bir gün, bir hafta, bir ay ve hatta bir yıl gibi herhangi bir uzunlukta olabilir. Örneğin, bir Poisson deneyi, bir ofis tarafından saat başına alınan telefon aramalarının sayısını, kışın kar nedeniyle okulun kapalı olduğu gün sayısını veya yağmur nedeniyle ertelenen oyunların sayısını temsil eden X rasgele değişkeni için gözlemler oluşturabilir. bir beyzbol sezonunda. Belirtilen bölge bir çizgi parçası, bir alan, bir hacim veya belki de bir malzeme parçası olabilir. Bu gibi durumlarda X , dönüm başına tarla faresi sayısını, belirli bir kültürdeki bakteri sayısını veya sayfa başına yazım hatası sayısını temsil edebilir. Bir Poisson deneyi, Poisson sürecinden türetilir ve aşağıdaki özelliklere sahiptir.

A. Poisson Sürecinin Özellikleri

1.) Uzayın belirli bir zaman aralığında veya belirli bir bölgesinde meydana gelen sonuçların sayısı, başka herhangi bir ayrık zaman aralığında veya bölgede meydana gelen sayıdan bağımsızdır. Bu anlamda Poisson sürecinin hafızası olmadığını söylüyoruz.

2.) Tek bir sonucun çok kısa bir zaman aralığında veya küçük bir bölgede meydana gelme olasılığı, zaman aralığının uzunluğu veya bölgenin büyüklüğü ile orantılıdır ve bu zaman aralığı veya bölge dışında meydana gelen sonuçların sayısına bağlı değildir.

3.) Bu kadar kısa bir zaman aralığında birden fazla sonucun ortaya çıkma veya bu kadar küçük bir bölgede düşme olasılığı yok denecek kadar azdır.

Bir Poisson deneyi sırasında ortaya çıkan sonuçların X sayısına Poisson rasgele değişkeni denir ve olasılık dağılımına Poisson dağılımı denir. Ortalama sonuç sayısı $\mu = \lambda t$ 'den hesaplanır; burada t , ilgili "zaman", "mesafe", "alan" veya "hacimdir". Olasılıklar, sonuçların meydana gelme oranı olan λ 'ya bağlı olduğundan, onları $p(x; \lambda t)$ ile göstereceğiz. Yukarıda

listelenen bir Poisson işleminin üç özelliğine dayalı olarak $p(x; \lambda t)$ formülünün türetilmesi bu kitabın kapsamı dışındadır. Poisson olasılıklarını hesaplamak için aşağıdaki formül kullanılır.

Belirli bir zaman aralığında veya t ile belirtilen belirli bir bölgede meydana gelen sonuçların sayısını temsil eden Poisson rasgele değişkeni X 'in olasılık dağılımı

$$p(x; \lambda t) = \frac{e^{-\lambda t} (\lambda t)^x}{x!}, x = 0, 1, 2, \dots$$

burada λ , birim zaman, mesafe, alan veya hacim başına ortalama sonuç sayısıdır ve $e = 2.71828$.

Ek'te Tablo A.2'de, 0.1 ile 18.0 arasında değişen λt 'nin seçilmiş değerleri için aşağıdaki Poisson olasılık toplamaları verilmiştir.

$$P(r; \lambda t) = \sum_{x=0}^r p(x; \lambda t)$$

Bu tablonun kullanımını aşağıdaki iki örnek ile açıklıyoruz.

Örnek 5.17: Bir laboratuvar deneyinde, bir sayaçtan 1 milisaniyede geçen ortalama radyoaktif parçacık sayısı 4'tür. Bir milisaniyede 6 parçacığın sayaca girme olasılığı nedir?

Cevap 5.17: $x = 6$ ve $\lambda t = 4$ ile Poisson dağılımını kullanarak ve Tablo A.2'ye bakarak,

$$p(x; \lambda t) = \frac{e^{-\lambda t} (\lambda t)^x}{x!} = p(6; 4) = \frac{e^{-4} (4)^6}{6!} = 0.1041$$

Ya da

$$\begin{aligned} P(r; \lambda t) &= \sum_{x=0}^r p(x; \lambda t) = P(6; 4) \\ &= \sum_{x=0}^6 p(x; 4) - \sum_{x=0}^5 p(x; 4) = 0.8893 - 0.7851 = 0.1042 \end{aligned}$$

Örnek 5.18: 10, her gün belirli bir limana gelen ortalama petrol tankeri sayısıdır. Limandaki tesisler günde en fazla 15 tanker kabul edebilmektedir. Belirli bir günde tankerlerin geri çevrilme olasılığı nedir?

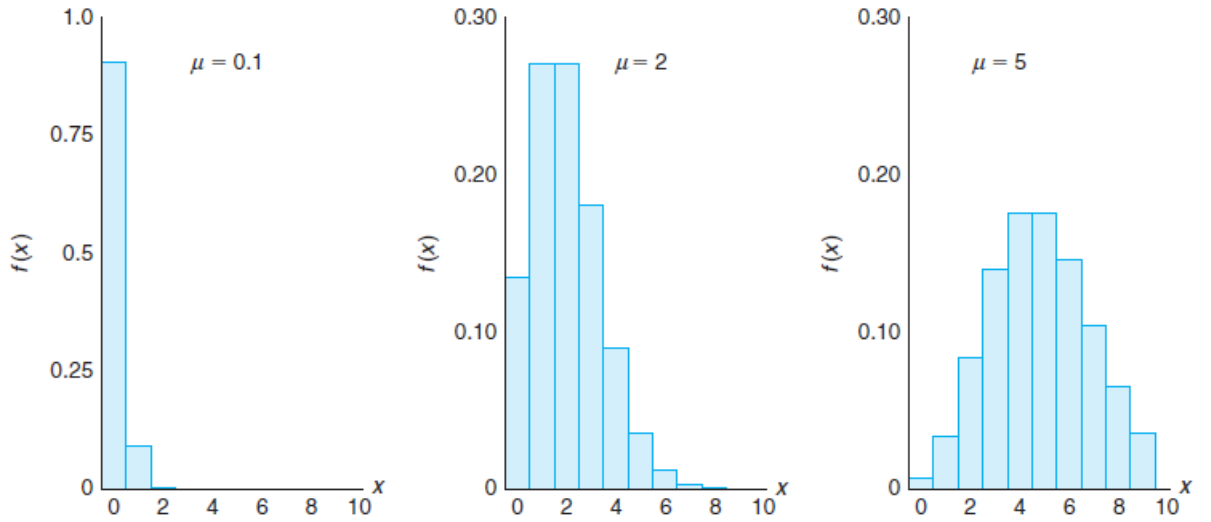
Cevap 5.18: Her gün gelen tanker sayısı X olsun. Sonra, Tablo A.2'yi kullanarak,

$$P(X > 15) = 1 - P(X \leq 15) = 1 - \sum_{x=0}^{15} p(x; 10) = 1 - 0.9513 = 0.0487.$$

TEOREM 5.4: Poisson dağılımı $p(x; \lambda t)$ 'nin hem ortalaması hem de varyansı λt 'dir.

B. Poisson Olasılık Fonksiyonunun Doğası

Pek çok ayırık (kesikli) ve sürekli dağılım gibi, Poisson dağılımının biçimi, ortalama büyüdükçe giderek daha simetrik, hatta çan şekli haline gelir. Şekil 5.1, $\mu = 0.1$, $\mu=2$ ve $\mu=5$ için olasılık fonksiyonunun grafiklerini göstererek bunu göstermektedir. $\mu=5$ 'e kadar büyüdüğünde simetriye yakınlığa dikkat edin.



Şekil 5.1: Farklı araçlar için Poisson yoğunluk fonksiyonları.

C. Bir Poisson Dağılımı ile Binom Dağılımının Yaklaşımı

Poisson dağılımının iki terimli dağılımla ilişkili olduğu, Poisson sürecinin üç ilkesinden anlaşılmalıdır. Poisson, Örnek 5.17 ve 5.18'de gösterildiği gibi genellikle uzay ve zaman problemlerinde uygulama bulsa da, binom dağılımının sınırlayıcı bir biçimi olarak görülebilir. Binom durumunda, n oldukça büyük ve p küçükse, koşullar Poisson sürecinin sürekli uzay veya zaman sonuçlarını simüle etmeye başlar. Binom durumundaki Bernoulli denemeleri arasındaki bağımsızlık, Poisson sürecinin 2. ilkesi ile tutarlıdır. p parametresinin 0'a yakın olmasına izin vermek, Poisson sürecinin 3. prensibi ile ilgilidir. Aslında, n büyükse ve p 0'a yakınsa, binom olasılıklarını tahmin etmek için $\mu = n * p$ ile Poisson dağılımı kullanılabilir. p

1'e yakınsa, başarı ve başarısızlık olarak tanımladığımız şeyi değiştirerek, böylece p 'yi 0'a yakın bir değere değiştirerek, binom olasılıklarına yaklaşmak için Poisson dağılımını kullanmaya devam edebiliriz.

Bunu genellikle $n * p \leq 10$ şartının sağlandığı binom dağılımına poisson dağılımı yaklaşımı uygulanabilir.

TEOREM 5.5: X , olasılık dağılımı $b(x; n, p)$ olan bir iki terimli rasgele değişken olsun.

$n \rightarrow \infty, p \rightarrow 0$ ve $np \xrightarrow{n \rightarrow \infty} \mu$ sabit kaldığında,

$b(x; n, p) \xrightarrow{n \rightarrow \infty} p(x; \mu)$.

Örnek 5.19: Belirli bir endüstriyel tesiste kazalar nadiren meydana gelir. Herhangi bir günde kaza olma olasılığının 0.005 olduğu ve kazaların birbirinden bağımsız olduğu bilinmektedir.

- a.) Herhangi bir 400 günlük süre içinde herhangi bir günde kaza olma olasılığı nedir?
- b.) Kazalı en fazla üç günün olma olasılığı nedir

Cevap 5.19: X , $n = 400$ ve $p = 0.005$ olan bir binom rasgele değişkeni olsun. Böylece, $np = 2$. Poisson yaklaşımı kullanılarak, (a) ve

a.) $P(X = 1) = p(x; \lambda t) = \frac{e^{-\lambda t} (\lambda t)^x}{x!} = p(1; 2) = \frac{e^{-2} (2)^1}{1!} = 0.271$. Ya da

$$b(x; n, p) = b(1; 400, 0.005) = \binom{400}{1} (0.005)^1 (1 - 0.005)^{400-1} = 0.27066.$$

b.) $P(X \leq 3) = \sum_{x=0}^3 p(x; \lambda t) = \sum_{x=0}^3 p(x; 2) = 0.8571$.

Örnek 5.20: Cam ürünlerin yapıldığı bir üretim sürecinde kusurlar veya kabarcıklar meydana gelir ve bu da parçayı zaman zaman pazarlama için istenmeyen hale getirir. Üretilen bu ürünlerin ortalama her 1000'inden 1'inde bir veya daha fazla baloncuk olduğu bilinmektedir. 8000 üründen rastgele bir örneklemin baloncuk içeren 7'den az öge verme olasılığı nedir?

Cevap 5.20: Bu aslında $n = 8000$ ve $p = 0.001$ olan bir binom deneyidir. p , 0'a çok yakın ve n oldukça büyük olduğundan, Poisson dağılımına yaklaşacağız;

$$\mu = np = (8000)(0.001) = 8.$$

Dolayısıyla, eğer X baloncukların sayısını temsil ediyorsa,

OLASILIK VE İSTATİSTİK

11.HAFTA

BAZI SÜREKLİ OLASILIK DAĞILIMLARI

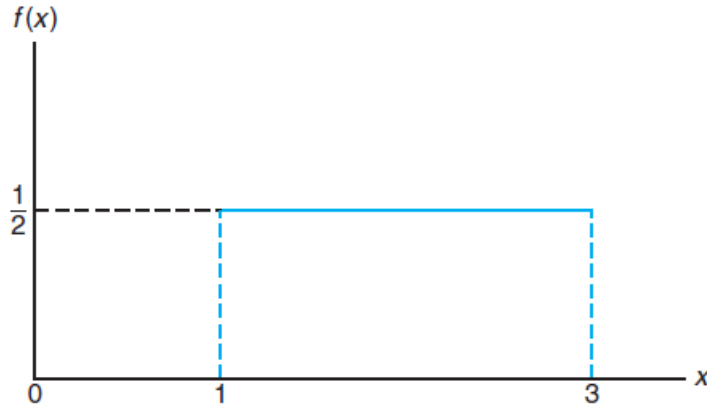
6.1 SÜREKLİ DÜZGÜN (BİR BİÇİMLİ/UNIFORM) DAĞILIM

Tüm istatistiklerdeki en basit sürekli dağılımlardan biri, sürekli düzgün (bir biçimli/tekdüze) dağılımdır. Bu dağılım, "düz" bir yoğunluk fonksiyonu ile karakterize edilir ve bu nedenle olasılık, $[A, B]$ gibi kapalı bir aralıkta düzgündür.

Sürekli düzgün rasgele değişken X 'in $[A, B]$ aralığındaki yoğunluk fonksiyonu

$$f(x; A, B) = \begin{cases} \frac{1}{B - A}, & A \leq x \leq B \\ 0, & \text{Diğer durumlar} \end{cases}$$

Yoğunluk fonksiyonu, tabanı $B - A$ ve sabit yüksekliği $\frac{1}{B-A}$ olan bir dikdörtgen oluşturur. Sonuç olarak, düzgün dağılım genellikle dikdörtgen dağılım olarak adlandırılır. Ancak aralığın her zaman kapalı olmayabileceğini unutmayın: $[A, B]$. (A, B) de olabilir. $[1, 3]$ aralığındaki düzgün bir rasgele değişken için yoğunluk fonksiyonu Şekil 6.1'de gösterilmektedir.



Şekil 6.1: $[1, 3]$ aralığında rastgele bir değişken için yoğunluk fonksiyonu.

Yoğunluk fonksiyonunun basit doğası nedeniyle, tekdüze dağılım için olasılıkların hesaplanması kolaydır. Bununla birlikte, bu dağılımın uygulanmasının, $[A, B]$ içinde sabit uzunlukta bir aralıkta düşme olasılığının sabit olduğu varsayımına dayandığına dikkat edin.

Örnek 6.1: Belirli bir şirketteki büyük bir konferans odasının en fazla 4 saat için rezerve edilebileceğini varsayalım. Hem uzun hem de kısa konferanslar oldukça sık gerçekleşir. Aslında, bir konferansın X uzunluğunun $[0, 4]$ aralığında düzgün bir dağılıma sahip olduğu varsayılabilir.

a.) Olasılık yoğunluk fonksiyonu nedir?

b.) Herhangi bir konferansın en az 3 saat sürme olasılığı nedir?

Cevap 6.1:

a.) Bu durumda düzgün dağılmış rasgele değişken X için uygun yoğunluk fonksiyonu şu şekildedir

$$f(x) = \begin{cases} \frac{1}{4-0}, & 0 \leq x \leq 4 \\ 0, & \text{Diğer durumlar} \end{cases}$$

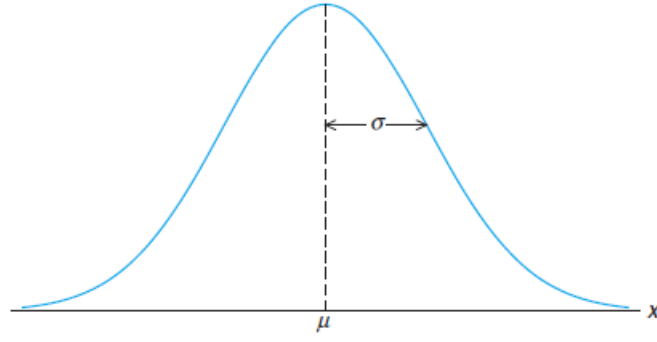
$$b.) P(X \geq 3) = \int_3^4 \frac{1}{4} dx = \frac{1}{4}.$$

TEOREM 6.1: Düzgün dağılımın ortalaması ve varyansı:

$$\mu = \frac{A+B}{2}$$
$$\sigma^2 = \frac{(B-A)^2}{12}$$

6.2 NORMAL DAĞILIM

Tüm istatistik alanındaki en önemli sürekli olasılık dağılımı normal dağılımdır. Normal eğri olarak adlandırılan grafiği, doğada, endüstride ve araştırmada meydana gelen birçok olguyu yaklaşık olarak tanımlayan Şekil 6.2'deki çan şeklindeki eğridir. Örneğin, meteorolojik deneyler, yağış çalışmaları ve üretilen parçaların ölçümleri gibi alanlardaki fiziksel ölçümler, genellikle normal bir dağılımla fazlasıyla yeterince açıklanır. Ek olarak, bilimsel ölçümlerdeki hatalar, normal bir dağılımla son derece iyi bir şekilde tahmin edilir. 1733'te Abraham DeMoivre, normal eğrinin matematiksel denklemini geliştirdi. Endüktif istatistik teorisinin çoğunun kurulduğu bir temel sağladı. Normal dağılım, denklemini aynı niceliğin tekrarlanan ölçümlerindeki hataların incelenmesinden de türeten Karl Friedrich Gauss (1777-1855) onuruna, genellikle Gauss dağılımı olarak anılır.



Şekil 6.2: Normal eğri.

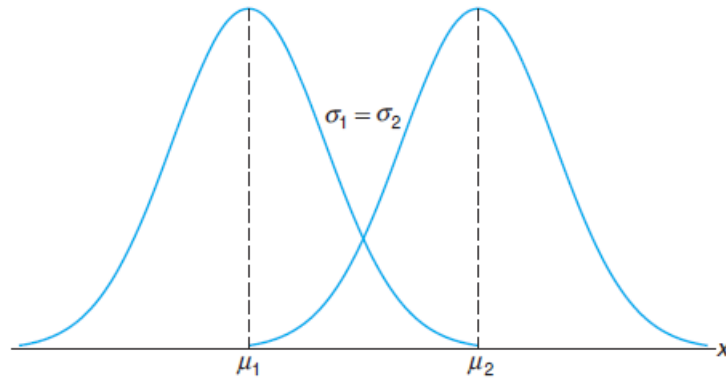
Şekil 6.2'deki çan şeklindeki dağılıma sahip sürekli bir rasgele değişken X , normal bir rasgele (rastlantı) değişken olarak adlandırılır. Normal değişkenin olasılık dağılımı için matematiksel denklem, sırasıyla μ ve σ olmak üzere iki parametreye, ortalamasına ve standart sapmasına bağlıdır. Dolayısıyla, X 'in yoğunluk değerlerini $n(x; \mu, \sigma)$ ile gösteriyoruz.

Ortalama μ ve varyans σ^2 ile normal rasgele değişken X 'in yoğunluğu

$$n(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi} * \sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, -\infty < x < \infty,$$

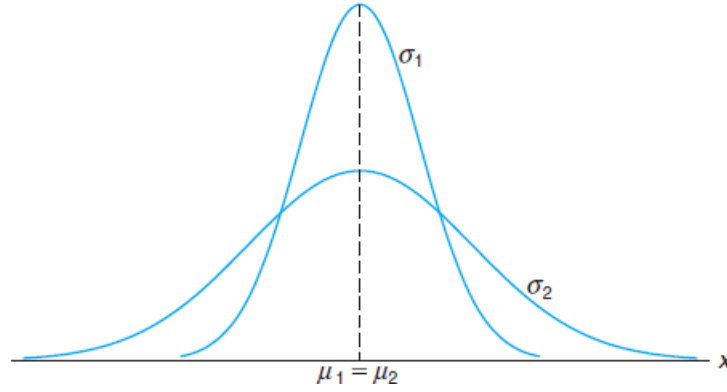
burada $\pi = 3,14159 \dots$ ve $e = 2.71828 \dots$

μ ve σ belirtildiğinde, normal eğri tamamen belirlenir. Örneğin, $\mu = 50$ ve $\sigma = 5$ ise, x 'in çeşitli değerleri ve çizilen eğri için $n(x; 50, 5)$ koordinatları hesaplanabilir. Şekil 6.3'te, aynı standart sapmaya ancak farklı ortalamalara sahip iki normal eğri çizdik. İki eğri form olarak aynıdır ancak yatay eksen boyunca farklı konumlarda ortalanmıştır.



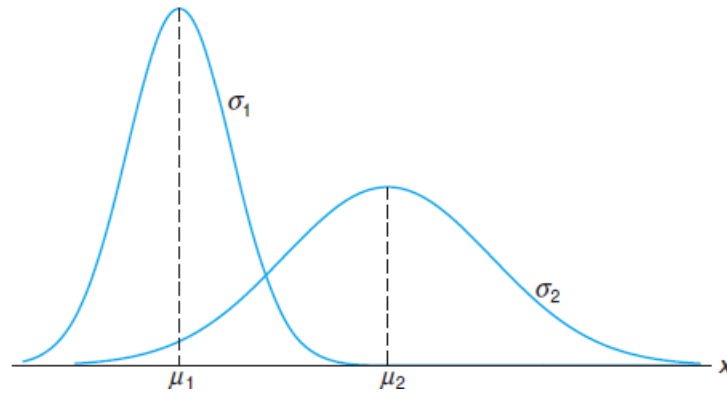
Şekil 6.3: $\mu_1 < \mu_2$ ve $\sigma_1 = \sigma_2$ ile normal eğriler.

Şekil 6.4'te, aynı ortalamaya ancak farklı standart sapmalara sahip iki normal eğri çizdik. Bu sefer iki eğrinin de yatay ekseninde tam olarak aynı konumda merkezlendiğini ancak standart sapması büyük olan eğrinin daha düşük olduğunu ve daha uzağa yayıldığını görüyoruz. Bir olasılık eğrisi altındaki alanın 1'e eşit olması gerektiğini ve bu nedenle gözlem seti ne kadar değişken olursa, karşılık gelen eğrinin o kadar düşük ve geniş olacağını unutmayın.



Şekil 6.4: $\mu_1 = \mu_2$ ve $\sigma_1 < \sigma_2$ ile normal eğriler.

Şekil 6.5, farklı ortalamalara ve farklı standart sapmalara sahip iki normal eğriyi göstermektedir. Açıkçası, yatay ekseninde farklı konumlarda merkezlenirler ve şekilleri iki farklı σ değerini yansıtır.



Şekil 6.5: $\mu_1 < \mu_2$ ve $\sigma_1 < \sigma_2$ ile normal eğriler.

Şekil. 6.2 ile Şekil. 6.5'in incelenmesine ve $n(x; \mu, \sigma)$ 'nin birinci ve ikinci türevlerinin incelenmesine dayanarak, normal eğrinin aşağıdaki özelliklerini listeliyoruz:

1. Normal eğrinin maksimum olduğu nokta olan, $x = \mu$ 'de gerçekleşir.
2. Eğri, ortalama $x = \mu$ boyunca dikey bir eksen etrafında simetriktir.
3. Normal eğri; $\mu - \sigma < X < \mu + \sigma$ aralığında ise içbükeydir (konkav) ve diğer yerlerde ise dışbükeydir(konveks).
4. Ortalamadan her iki yönde de uzaklaştıkça normal eğri yatay eksene asimptotik olarak yaklaşır.
5. Eğrinin altındaki ve yatay eksenin üzerindeki toplam alan 1'e eşittir.

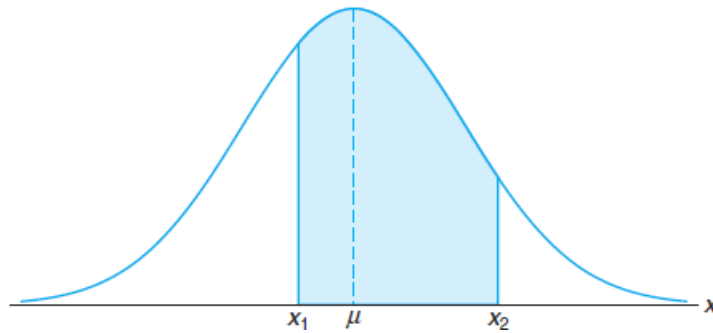
TEOREM 6.2: $n(x; \mu, \sigma)$ 'nin ortalaması ve varyansı sırasıyla μ ve σ^2 'dir. Bu nedenle, standart sapma σ 'dir.

6.3 NORMAL EĞRİNİN ALTINDA KALAN ALANLAR

Herhangi bir sürekli olasılık dağılımı veya yoğunluk fonksiyonunun eğrisi, $x = x_1$ ve $x = x_2$ ordinatıyla sınırlanan eğrinin altındaki alan, X rasgele değişkeninin $x = x_1$ ve $x = x_2$ arasında bir değer alma olasılığına eşit olacak şekilde oluşturulur. Böylece, Şekil 6.6'daki normal eğri için,

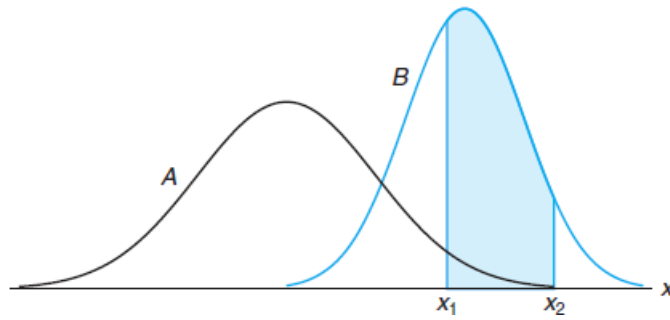
$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} n(x; \mu, \sigma) dx = \frac{1}{\sqrt{2\pi} * \sigma} \int_{x_1}^{x_2} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx$$

gölge bölgenin alanıyla temsil edilir.



Şekil 6.6: $P(x_1 < X < x_2)$ = gölge bölgenin alanı.

Şekil 6.3, 6.4 ve 6.5'te normal eğrinin incelenen dağılımın ortalamasına ve standart sapmasına nasıl bağlı olduğunu gördük. Herhangi iki ordinat arasındaki eğrinin altındaki alan da μ ve σ değerlerine bağlı olmalıdır. Bu, farklı ortalamalara ve varyanslara sahip iki eğri için $P(x_1 < X < x_2)$ 'ye karşılık gelen gölgeli bölgelere sahip olduğumuz Şekil 6.7'de açıkça görülmektedir. X 'in A dağılımını tanımlayan rasgele değişken olduğu $P(x_1 < X < x_2)$, A eğrisinin altındaki gölgeli alanla gösterilir. X , B dağılımını tanımlayan rasgele değişken ise, o zaman $P(x_1 < X < x_2)$ tüm gölgeli bölge tarafından verilir. Açıkçası, iki gölgeli bölgenin boyutu farklıdır; bu nedenle, her dağılımla ilişkili olasılık, X 'in verilen iki değeri için farklı olacaktır.



Şekil 6.7: Farklı normal eğriler için $P(x_1 < X < x_2)$.

Normal eğri altındaki alanların hesaplanmasında kullanılabilecek birçok istatistiksel yazılım türü vardır. Normal yoğunluk fonksiyonlarının integrallerini çözmede karşılaşılan zorluk, hızlı başvuru için normal eğri alanlarının tablo haline getirilmesini gerektirir. Bununla birlikte, akla gelebilecek her μ ve σ değeri için ayrı tablolar oluşturmaya çalışmak umutsuz bir iş olacaktır. Neyse ki, herhangi bir X normal rasgele değişkeninin tüm gözlemlerini, ortalama 0 ve varyans 1 olan normal bir rasgele değişken Z 'nin yeni bir gözlem grubuna dönüştürebiliriz. Bu,

$$Z = \frac{X - \mu}{\sigma}$$

dönüşümü aracılığıyla yapılabilir.

X , x değerini aldığında, karşılık gelen Z değeri $z = (x - \mu)/\sigma$ ile verilir. Bu nedenle, X , $x = x_1$ ve $x = x_2$ değerleri arasında kalırsa, Z rasgele değişkeni karşılık gelen $z_1 = (x_1 - \mu)/\sigma$ ve $z_2 = (x_2 - \mu)/\sigma$ değerleri arasında olacaktır. Sonuç olarak, Z 'nin ortalaması 0 ve varyansı 1 olan normal bir rasgele değişken olarak görüldüğü

$$P(x_1 < X < x_2) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \int_{x_1}^{x_2} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{1}{2}(z)^2} dz$$

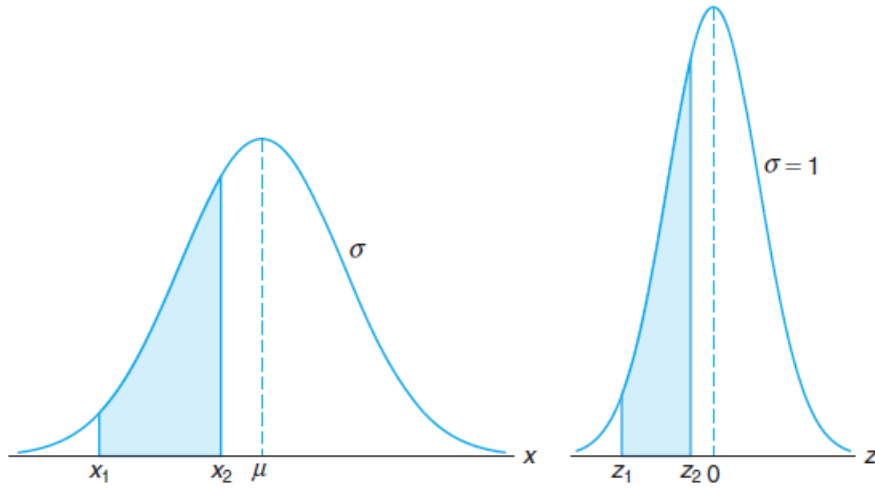
$$= \int_{z_1}^{z_2} n(z; 0,1) dz = P(z_1 < Z < z_2)$$

şeklinde yazabiliriz.

Tanım 6.1: Ortalaması 0 ve varyansı 1 olan normal bir rastgele değişkenin dağılımına standart normal dağılım denir.

Orijinal ve dönüştürülmüş dağılımlar Şekil 6.8'de gösterilmektedir. x_1 ve x_2 arasında düşen tüm X değerleri, z_1 ve z_2 arasında karşılık gelen z değerlerine sahip olduğundan, Şekil 6.8'deki $x = x_1$ ve $x = x_2$ ordinatları arasındaki X eğrisinin altındaki alan, $z = z_1$ ve $z = z_2$ 'nin dönüştürülmüş ordinatları arasındaki Z eğrisinin altındaki alana eşittir.

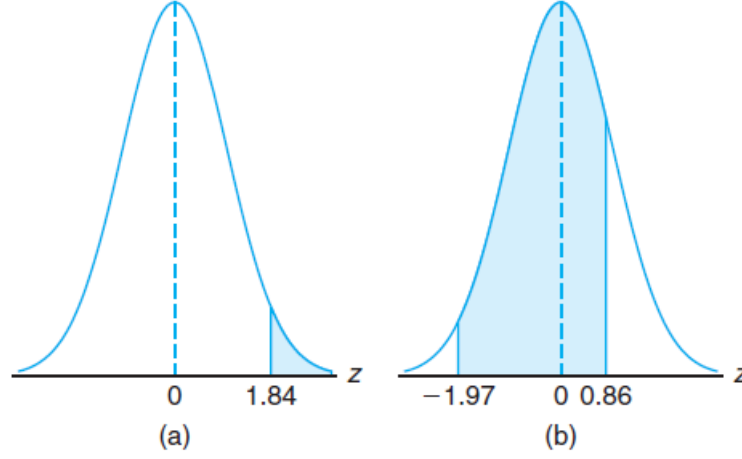
Z dönüşüm tablosu Ek'te Tablo A.3 ile gösterilmiştir.



Şekil 6.8: Orijinal ve dönüştürülmüş normal dağılımlar.

Örnek 6.2: Standart bir normal dağılım verildiğinde,

- a.) $z = 1,84$ 'ün sağında ve
b.) $z = -1,97$ ile $z = 0,86$ arasında kalan eğrinin altındaki alanı bulunuz.



Şekil 6.9: Örnek 6.2 için alanlar.

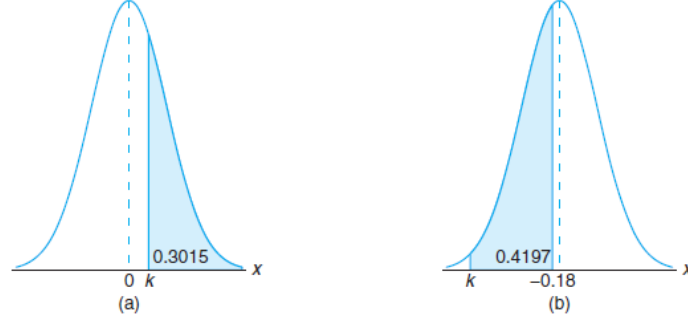
Cevap 6.2:

- a.) Şekil 6.9(a)'daki $z = 1.84$ 'ün sağındaki alan, 1 eksi Tablo A.3'teki $z = 1.84$ 'ün solundaki alana eşittir, yani $P(Z > 1.84) = 1 - P(Z < 1.84) = 1 - 0.9671 = 0.0329$.
- b.) Şekil 6.9(b)'deki $z = -1.97$ ile $z = 0.86$ arasındaki alan, $z = 0.86$ 'nın solundaki alan eksi $z = -1.97$ 'nin solundaki alana eşittir. Yani $P(-1.97 < Z < 0.86) = P(Z < 0.86) - P(Z < -1.97)$ Tablo A.3'ten istenen alanı $0.8051 - 0.0244 = 0.7807$ olarak buluyoruz.

Örnek 6.3: Aşağıda verilen standart bir normal dağılım için,

a.) $P(Z > k) = 0.3015$

b.) $P(k < Z < -0.18) = 0.4197$ olduğu durumlar için k 'nın değerleri nedir?

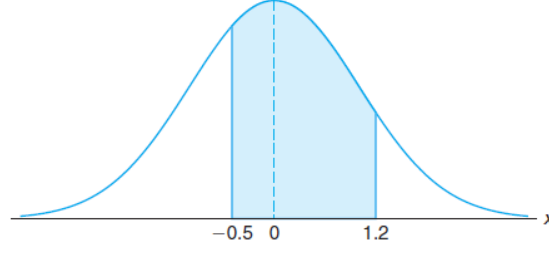


Şekil 6.10: Örnek 6.3 için alanlar.

Cevap 6.3: Dağılımlar ve istenen alanlar Şekil 6.10'da gösterilmiştir.

- a.) Şekil 6.10(a)'da, sağda 0.3015'lik bir alan bırakan k değerinin, solda 0.6985'lik bir alan bırakması gerektiğini görüyoruz. Yani, $P(Z > k) = 1 - P(Z < k) = 0.3015$ olmaktadır. Bu durumda $P(Z < k) = 0.6985$ Tablo A.3'ten $k = 0.52$ olduğu sonucu çıkar.
- b.) Tablo A.3'ten -0.18'in solundaki toplam alanın 0.4286'ya eşit olduğunu not ediyoruz. Şekil 6.10(b)'de, k ile -0.18 arasındaki alanın 0.4197 olduğunu görüyoruz, dolayısıyla k 'nin solundaki alan $0.4286 - 0.4197 = 0.0089$ olmalıdır. Dolayısıyla, Tablo A.3'ten $k = -2.37$ elde ederiz. Yani, $P(k < Z < -0.18) = P(Z < -0.18) - P(Z < k) = 0.4197$. Tablo A.3'ten $P(Z < -0.18) = 0.4286$ olarak görülmektedir. Böylece $P(Z < k) = 0.4286 - 0.4197 = 0.0089$ değeri Tablo A3'ten $k = -2.37$ 'de sağlanmaktadır.

Örnek 6.4: $\mu = 50$ ve $\sigma = 10$ ile normal dağılıma sahip bir X rastgele değişkeni verildiğinde, X 'in 45 ile 62 arasında bir değer alma olasılığını bulunuz.



Şekil 6.11: Örnek 6.4 için alan.

Cevap 6.4: $x_1 = 45$ ve $x_2 = 62$ 'ye karşılık gelen z değerleri

$$Z = \frac{x - \mu}{\sigma} = z_1 = \frac{45 - 50}{10} = -0.5 \text{ ve } z_2 = \frac{62 - 50}{10} = 1.2$$

Bu nedenle,

$$P(45 < X < 62) = P(-0.5 < Z < 1.2)$$

Şekil 6.11'deki taralı bölgenin alanı ile gösterilir. Bu alan, $z = 1.2$ 'nin solundaki tüm alandan $z = -0.5$ ordinatının solundaki alanı çıkararak bulunabilir. Yani

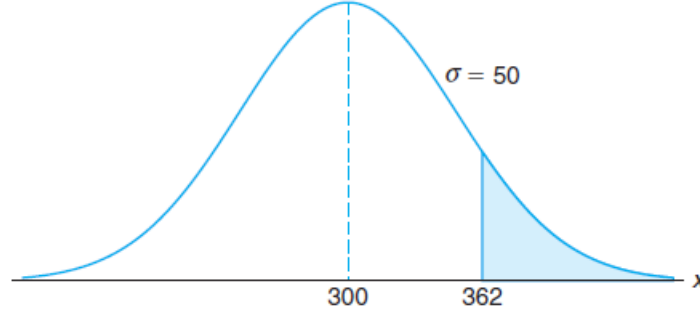
$$P(-0.5 < Z < 1.2) = P(Z < 1.2) - P(Z < -0.5)$$

Tablo A.3'ü kullanarak,

$$P(-0.5 < Z < 1.2) = P(Z < 1.2) - P(Z < -0.5) = 0.8849 - 0.3085 = 0.5764.$$

Örnek 6.5: X 'in $\mu = 300$ ve $\sigma = 50$ ile normal bir dağılıma sahip olduğuna göre, X 'in 362'den büyük bir değer alma olasılığını bulunuz

Cevap 6.5: İstenen alan gölgeli olarak normal olasılık dağılımı Şekil 6.12'de gösterilmiştir. $P(X > 362)$ 'yi bulmak için, $x = 362$ 'nin sağındaki normal eğrinin altındaki alanı değerlendirmemiz gerekir. Bu, $x = 362$ 'yi karşılık gelen z değerine dönüştürerek, Tablo A.3'ten z 'nin solundaki alanı elde ederek ve ardından bu alanı 1'den çıkararak yapılabilir.



Şekil 6.12: Örnek 6.5 için Alan.

$$z = \frac{X - \mu}{\sigma} = \frac{362 - 300}{50} = 1.24$$

$$P(X > 362) = P(Z > 1.24) = 1 - P(Z < 1.24) = 1 - 0.8925 = 0.1075.$$

Chebyshev teoremine göre, rastgele bir değişkenin ortalamasının 2 standart sapması içinde bir değer alma olasılığı en az $3/4$ 'tür. Rastgele değişkenin normal bir dağılımı varsa, $x_1 = \mu - 2\sigma$ ve $x_2 = \mu + 2\sigma$ 'ya karşılık gelen z değerleri kolayca şu şekilde hesaplanır:

$$z_1 = \frac{(\mu - 2\sigma) - \mu}{\sigma} = -2 \text{ ve } z_2 = \frac{(\mu + 2\sigma) - \mu}{\sigma} = 2$$

Böylece

$$\begin{aligned} P(\mu - 2\sigma < X < \mu + 2\sigma) &= P(-2 < Z < 2) \\ &= P(Z < 2) - P(Z < -2) = 0.9772 - 0.0228 = 0.9544. \end{aligned}$$

A. Normal Eğriyi Tersten Kullanma

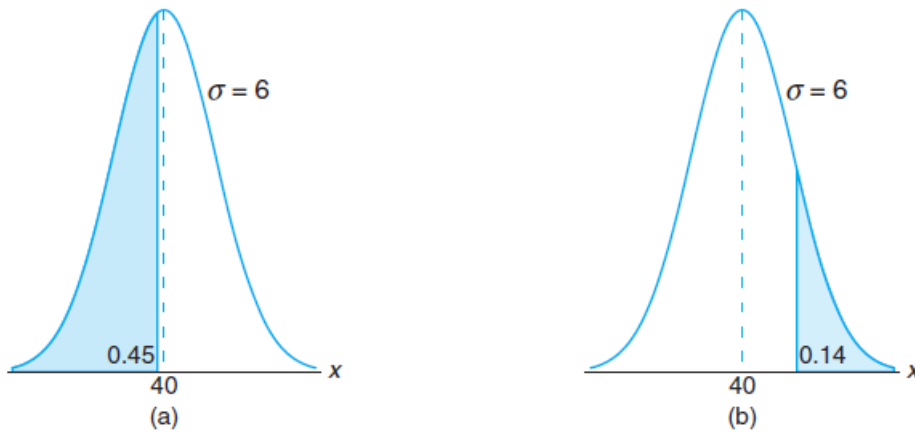
Bazen Tablo A.3'te listelenen değerler arasında kalan belirli bir olasılığa karşılık gelen z değerini bulmamız gerekir (bkz. Örnek 6.6). Kolaylık sağlamak için, her zaman belirtilen olasılığa en yakın tablosal olasılığa karşılık gelen z değerini seçeceğiz.

Önceki iki örnek, önce bir x değerinden bir z değerine gidilerek ve ardından istenen alan hesaplanarak çözüldü. Örnek 6.6'da, işlemi tersine çeviriyoruz ve bilinen bir alan veya olasılıkla başlıyoruz, z değerini buluyoruz ve ardından z formülünü yeniden düzenleyerek x 'i buluyoruz:

$$z = \frac{x - \mu}{\sigma}, x = \sigma z + \mu$$

Örnek 6.6: $\mu = 40$ ve $\sigma = 6$ ile bir normal dağılım verildiğinde,

- soldaki alanın %45'i ve
- sağdaki alanın %14'ü olan x 'in değerini bulunuz.



Şekil 6.13: Örnek 6.6 için alanlar.

Cevap 6.6:

- Şekil 6.13(a)'da istenen x değerinin solundaki 0.45'lik bir alan gölgelendirilmiştir. Sola 0.45'lik bir alan bırakan bir z değerine ihtiyacımız var. Tablo A.3'ten $P(Z < -0.13) = 0.45$ buluruz, dolayısıyla istenen z değeri -0.13'tür. Dolayısıyla,

$$x = (6)(-0.13) + 40 = 39.22$$

Yani; $P(Z < z) = 0.45$; $P(X < x) = P\left(Z < \frac{x-\mu}{\sigma}\right) = 0.45$ Tabloda 0.45 sonucu doğrudan yoksa buna en yakın sonuç esas alınır. Dolayısıyla bu rakam 0.4487 olarak güncellenebilir. Böylece tabloda 0.4487'yi sağlayan değer -0.13'tür. Böylece

$$\frac{x - \mu}{\sigma} = -0.13 \Rightarrow \frac{x - 40}{6} = -0.13 \Rightarrow x = 39.22$$

b.) Şekil 6.13(b)'de istenen x değerinin sağında 0.14'e eşit bir alanı gölgelendiriyoruz. Bu sefer, alanın 0,14'ünü sağa ve dolayısıyla 0,86'lık bir alanı sola bırakan bir z değerine ihtiyacımız var. Yine Tablo A.3'ten $P(Z < 1.08) = 0.86$ buluyoruz, yani istenen z değeri 1.08 ve $x = (6)(1.08) + 40 = 46.48$.

$$\text{Yani; } P(X > x) = 1 - P(X < x) = 0.14$$

$$\Rightarrow P(X < x) = 0.86 = \left(Z < \frac{x-40}{6}\right) = 0.86 \text{ Tablodan}$$

$$\frac{x-40}{6} = 1.08 \text{ olarak bulunmaktadır. Böylece } x = 46.48.$$

EK A3:

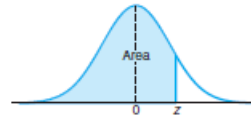


Table A.3 Areas under the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

OLASILIK VE İSTATİSTİK

12.HAFTA

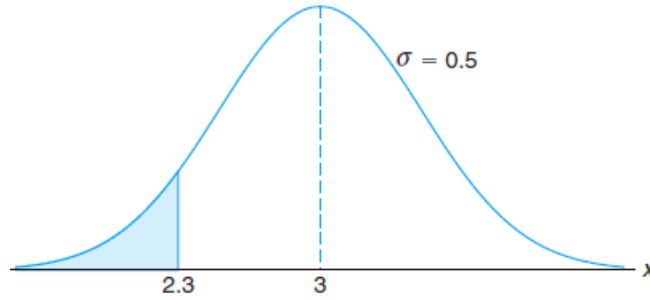
BAZI SÜREKLİ OLASILIK DAĞILIMLARI (DEVAM)

6.4 NORMAL DAĞILIM UYGULAMALARI

Normal dağılımın geçerli olduğu pek çok problemde bazıları aşağıdaki örneklerde ele alınmaktadır.

Örnek 6.7: Belirli bir depolama pili tipi, 0.5 yıllık standart sapma ile ortalama 3.0 yıl dayanır. Pil ömrünün normal dağıldığını varsayarak, belirli bir pilin 2.3 yıldan daha az dayanma olasılığını bulunuz.

Cevap 6.7: Önce, verilen pil ömrü dağılımını ve istenen alanı gösteren Şekil 6.14 gibi bir diyagram oluşturulmaktadır.



Şekil 6.14: Örnek 6.7 için Alan.

$P(X < 2.3)$ 'ü bulmak için, 2.3'ün solundaki normal eğrinin altındaki alanı değerlendirmemiz gerekir. Bu, karşılık gelen z değerinin solundaki alanı bularak gerçekleştirilir. Dolayısıyla,

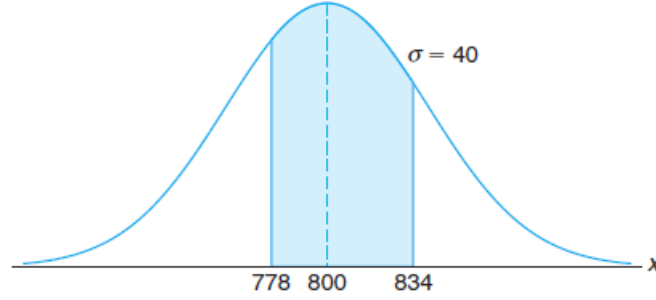
$$z = \frac{2.3 - 3}{0.5} = -1.4$$

Tablo A.3'ten

$$P(X < 2.3) = P(Z < -1.4) = 0.0808.$$

Örnek 6.8: Bir elektrik firması, normal dağılan, ortalama 800 saat ve standart sapması 40 saat olan, yanmadan önce bir ömre sahip ampuller üretmektedir. Bir ampulün 778 ile 834 saat arasında yanma olasılığını bulunuz.

Cevap 6.8: Ampul ömrünün dağılımı Şekil 6.15'te gösterilmektedir.



Şekil 6.15: Örnek 6.8 için Alan.

$x_1 = 778$ ve $x_2 = 834$ 'e karşılık gelen z değerleri;

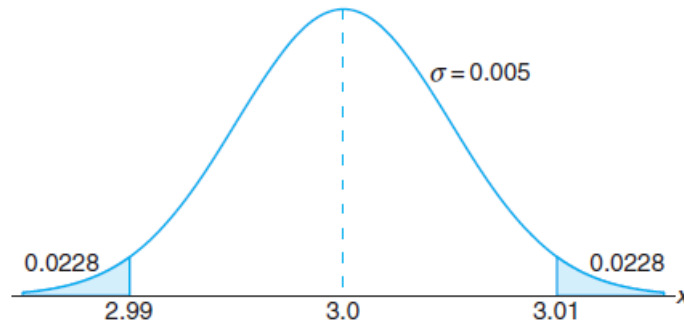
$$z_1 = \frac{778-800}{40} = -0.55 \text{ ve } z_2 = \frac{834-800}{40} = 0.85$$

Dolayısıyla,

$$\begin{aligned} P(778 < X < 834) &= P(-0.55 < Z < 0.85) = P(Z < 0.85) - P(Z < -0.55) \\ &= 0.8023 - 0.2912 = 0.5111. \end{aligned}$$

Örnek 6.9: Firma ürettiği bilyelerin çap uzunluğunu 3.0 ± 0.01 cm olarak belirlemektedir. Bu aralığın dışındaki çap uzunluğuna sahip olan bilyeler kusurlu kabul edilmektedir. Bilye çaplarının ortalama değeri $\mu = 3.0$ ve standart sapması $\sigma = 0.005$ ile normal bir dağılıma sahip olduğu bilinmektedir. Üretilen bilyelerin ortalama olarak ne kadarı kusurludur?

Cevap 6.9: Çapların dağılımı Şekil 6.16'da gösterilmektedir.



Şekil 6.16: Örnek 6.9 için Alan.

Çap limitlerine karşılık gelen değerler $x_1 = 2.99$ ve $x_2 = 3.01$ 'dir. Karşılık gelen z değerleri;

$$z_1 = \frac{2.99-3}{0.005} = -2.0 \text{ ve } z_2 = \frac{3.01-3}{0.005} = 2.0$$

Dolayısıyla, $P(2.99 < X < 3.01) = P(-2.0 < Z < 2.0)$.

Tablo A.3'ten $P(Z < -2.0) = 0.0228$. Normal dağılımın simetrisi nedeniyle, $P(Z < -2.0) + P(Z > 2.0) = 2(0.0228) = 0.0456$ olduğunu buluruz. Sonuç olarak, üretilen bilyalı rulmanların ortalama %4.56'sının kusurlu çıkacağı tahmin edilmektedir.

Örnek 6.10: Ürünlerin belirli bir boyutunun $1.50 \pm d$ içinde olup olmadığı test edilmektedir. Bu ölçümün ortalaması 1.50 ve standart sapması 0.2 ile normal dağıldığı bilinmektedir. Test edilen ürünlerin %95'inin sağlam kabul edilebilmesi için “d” değerini belirleyiniz.

Cevap 6.10: $x_1 = 1.5 - d$ ve $x_2 = 1.5 + d$ 'e karşılık gelen z değerleri;

$$z_1 = \frac{(1.5-d)-(1.5)}{0.2} = -\frac{d}{0.2} \text{ ve } z_2 = \frac{(1.5+d)-(1.5)}{0.2} = \frac{d}{0.2}$$

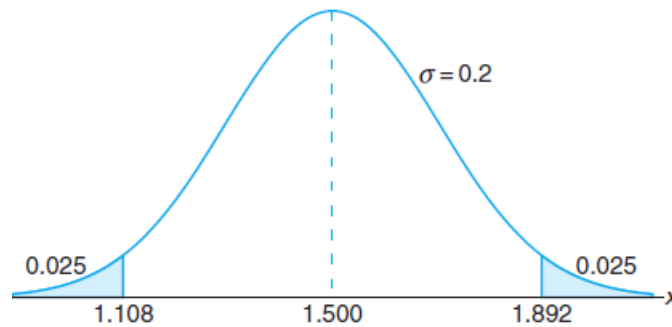
Dolayısıyla,

$$\begin{aligned} P(1.5 - d < X < 1.5 + d) &= P\left(-\frac{d}{0.2} < Z < \frac{d}{0.2}\right) \\ &= P\left(Z < \frac{d}{0.2}\right) - P\left(Z < -\frac{d}{0.2}\right) = 1 - 2 * P\left(Z < -\frac{d}{0.2}\right) = 0.95 \\ &\Rightarrow P\left(Z < -\frac{d}{0.2}\right) = 0.025 \end{aligned}$$

Dolayısıyla,

$$d = 0.392.$$

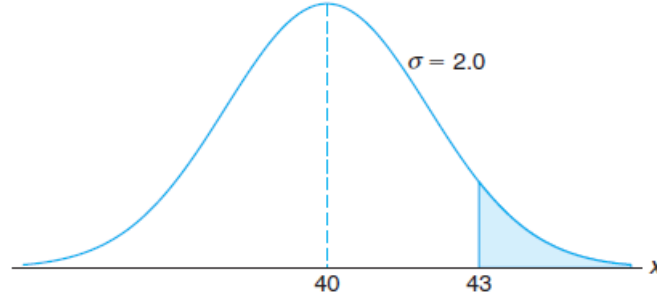
Çözümün gösterimi Şekil 6.17'de gösterilmektedir.



Şekil 6.17: Örnek 6.10 için Alan.

Örnek 6.11: Belirli bir makine, ortalama direnci 40 ohm ve standart sapması 2 ohm olan elektrik dirençleri yapıyor. Direncin normal bir dağılım izlediğini ve herhangi bir doğruluk derecesinde ölçülebileceğini varsayarsak, dirençlerin yüzde kaçının direnci 43 ohm'dan büyük direnç değerine sahiptir?

Cevap 6.11: Şekil 6.18'de $x = 43$ 'ün sağındaki alanı bulmalıyız.



Şekil 6.18: Örnek 6.11 için Alan.

Bu, $x = 43$ 'ü karşılık gelen z değerine dönüştürerek, Tablo A.3'ten z 'nin solundaki alanı elde ederek ve ardından bu alanı 1'den çıkararak yapılabilir:

$$z = \frac{43 - 40}{2} = 1.5$$

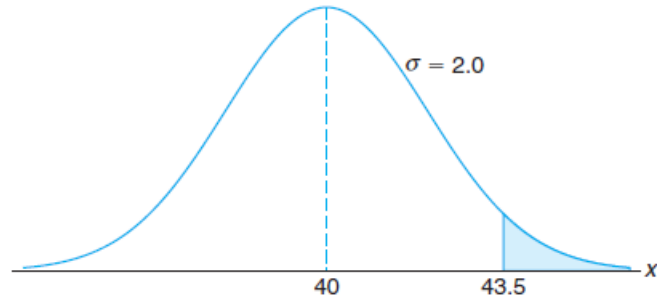
Böylece,

$$P(X > 43) = P(Z > 1.5) = 1 - P(Z < 1.5) = 1 - 0.9332 = 0.0668.$$

Dolayısıyla, dirençlerin %6,68'i 43 ohm'u aşan bir dirence sahip olacaktır.

Örnek 6.12: Örnek 6.11'teki direnç değerleri içi en yakın tamsayıya yuvarlanarak elde ediliyorsa 43 ohm'u aşan dirençlerin yüzdesini bulunuz.

Cevap 6.12: Gerekli alan, Şekil 6.19'da 43.5'in sağında taralı bölgedir.



Şekil 6.19: Örnek 6.12 için Alan.

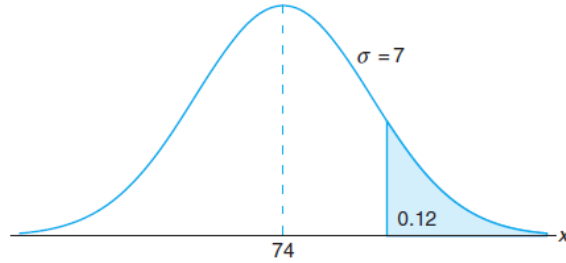
$$z = \frac{43.5 - 40}{2} = 1.75$$

Böylece,

$$P(X > 43.5) = P(Z > 1.75) = 1 - P(Z < 1.75) = 1 - 0.9599 = 0.0401.$$

Örnek 6.13: Bir sınavın ortalama notu 74 ve standart sapması 7'dir. Sınıfın %12'si A harf notunu aldıysa ve notlar normal bir dağılım izleyecek şekilde eğri ise, mümkün olan en düşük A ve mümkün olan en yüksek B nedir? (NOT: 0-100 arası tamsayı değerleridir).

Cevap 6.13: Bu örnekte, bilinen bir olasılık alanıyla başlıyoruz, z değerini buluyoruz ve sonra $x = \sigma z + \mu$ formülünden x'i belirliyoruz. A alan öğrencilerin oranına karşılık gelen 0.12'lik bir alan Şekil 6.20'de gölgelendirilmiştir.



Şekil 6.20: Örnek 6.13 için Alan.

Alanın 0.12'sini sağa ve dolayısıyla 0.88'lik bir alanı sola bırakan bir z değerine ihtiyacımız var. Tablo A.3'ten, $P(Z < 1.18)$ 0.88'e en yakın değere sahiptir, dolayısıyla istenen z değeri 1.18'dir. Yani;

$$P(X > A) = 0.12.$$

$$z = \frac{A - 74}{7}$$

$$P\left(Z > \frac{A - 74}{7}\right) = 1 - P\left(Z < \frac{A - 74}{7}\right) = 0.12$$

$$P\left(Z < \frac{A - 74}{7}\right) = 0.88$$

$$\frac{A - 74}{7} = 1.175$$

$$A = 82.225$$

Bu nedenle, en düşük A 83 ve en yüksek B 82'dir.

Örnek 6.14: Örnek 6.13'te sınıfın %60'ı en az hangi nottan daha düşük bir not almıştır?

Cevap 6.14:

$$P(X < x) = 0.6.$$

$$z = \frac{x - 74}{7}$$

$$P\left(Z < \frac{x - 74}{7}\right) = 0.6$$

$$\frac{x - 74}{7} = 0.25$$

$$x = 75.75$$

Yani notların %60'ı 75 ve altındadır.

6.5 İKİ TERİMLİ (BİNOM) DAĞILIMIN NORMAL DAĞILIMA YAKLAŞIMI

İki terimli deneylerle ilişkili olasılıklar, iki terimli dağılımın $b(x; n, p)$ formülünden veya n küçük olduğunda Tablo A.1'den kolaylıkla elde edilebilir. Ek olarak, binom olasılıkları birçok bilgisayar yazılım paketinde kolayca bulunur. Bununla birlikte, iki terimli ve normal dağılım arasındaki ilişkiyi öğrenmek öğreticidir. Bölüm 5.5'te, n oldukça büyük ve p 0 veya 1'e çok yakın olduğunda, iki terimli olasılıkları tahmin etmek için Poisson dağılımının nasıl kullanılabileceğini gösterdik. Hem iki terimli hem de Poisson dağılımları ayrıktır (kesiklidir). Kesintisiz bir numune alanı üzerindeki yaklaşık olasılıklara sürekli bir olasılık dağılımının ilk uygulaması, normal eğrinin kullanıldığı Örnek 6.12'de gösterildi. Normal dağılım, ikincisi simetrik bir çan şekli aldığı anda, ayrı bir dağılıma genellikle iyi bir yaklaşımdır. Teorik açıdan bakıldığında, bazı dağılımlar, parametreleri belirli sınırlara yaklaştıkça normale yakınsar. Normal dağılım, uygun bir yaklaşık dağılımdır, çünkü kümülatif dağılım fonksiyonu çok kolay bir şekilde tablolaştırılabilir. Biri kümülatif dağılım işleviyle çalışırken, pratik problemlerde binom dağılımına normal tarafından güzel bir şekilde yaklaşılr. Şimdi, n yeterince büyük olduğunda, iki terimli özellikleri yaklaşık olarak tahmin etmek için normal eğri altındaki alanları kullanmamıza izin veren bir teorem ifade ediyoruz.

TEOREM 6.3: X , ortalama $\mu = np$ ve varyans $\sigma^2 = npq$ olan bir binom rasgele değişkeniyse, o zaman Z dağılımının sınırlayıcı biçimi,

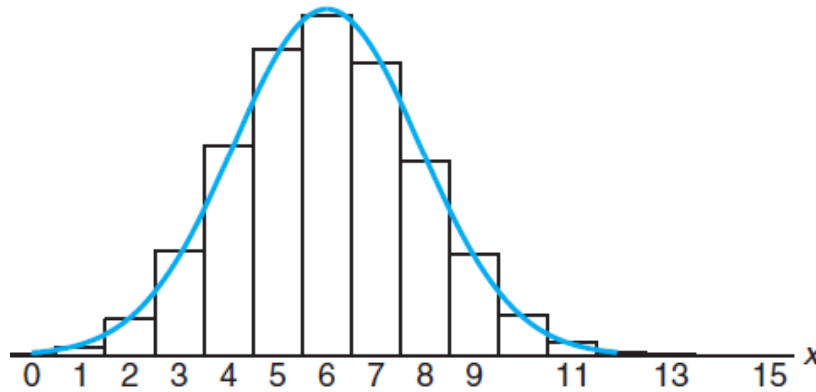
$$Z = \frac{X - np}{\sqrt{npq}}$$

$n \rightarrow \infty$ olarak, standart normal dağılım $n(z; 0, 1)$ olur.

$\mu = np$ ve $\sigma^2 = np(1 - p)$ ile normal dağılımın, n büyük olduğunda ve p 0 veya 1'e çok yakın olmadığında binom dağılımına yalnızca çok doğru bir yaklaşım sağlamakla kalmaz, aynı zamanda oldukça n küçük ve p 1/2'ye oldukça yakın olduğunda bile iyi bir yaklaşım göstermektedir.

Binom dağılımına normal yaklaşımı göstermek için, önce $b(x; 15, 0.4)$ için histogramı çizeriz ve sonra X binom değişkeni ile aynı ortalama ve varyansa sahip özel normal eğriyi üst üste bindiririz. Dolayısıyla, bir normal eğri çizeriz $\mu = np = (15)(0.4) = 6$ ve $\sigma^2 = npq = (15)(0.4)(0.6) = 3.6$

$b(x; 15, 0.4)$ 'ün histogramı ve tamamıyla ortalaması ve varyansı ile belirlenen karşılık gelen üst üste bindirilmiş normal eğri Şekil 6.22'de gösterilmektedir.

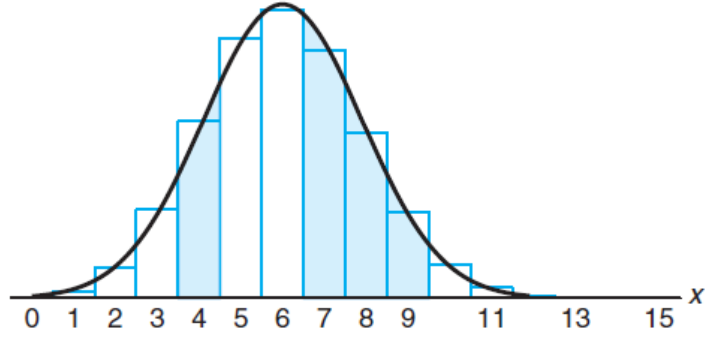


Şekil 6.22: $b(x; 15, 0.4)$ 'ün normal yaklaşımı.

Binom rasgele değişkeni X 'in belirli bir x değerini alma olasılığı, tabanı x merkezli olan çubuğun alanına eşittir. Örneğin, X 'in 4 değerini üstlenme olasılığı, tabanı $x = 4$ merkezli dikdörtgenin alanına eşittir. Tablo A.1'i kullanarak, bu alanın

$$P(X = 4) = b(4; 15, 0.4) = 0.1268$$

olduğunu ve yaklaşık olarak Şekil 6.23'teki iki koordinat $x_1 = 3.5$ ve $x_2 = 4.5$ arasındaki normal eğri altındaki taralı bölgedir.



Şekil 6.23: $b(x; 15, 0.4)$ ve $\sum_{x=7}^9 b(x; 15, 0.4)$ için normal yaklaşım.

Z değerlerine çevirirsek

$$z_1 = \frac{3.5-6}{1.897} = -1.32 \text{ ve } z_2 = \frac{4.5-6}{1.897} = -0.79$$

elde ederiz.

X bir binom rasgele değişkeni ve Z bir standart normal değişken ise, o zaman

$$\begin{aligned} P(X = 4) &= b(4; 15, 0.4) \approx P(-1.32 < Z < -0.79) \\ &= P(Z < -0.79) - P(Z < -1.32) = 0.2148 - 0.0934 = 0.1214. \end{aligned}$$

Bu, 0.1268 tam değeri ile çok yakından uyuşmaktadır.

Normal yaklaşım, n 'nin büyük değerleri için binom toplamalarının hesaplanmasında en kullanışlıdır. Şekil 6.23'e atıfta bulunarak, X 'in 7'den 9'a kadar (dâhil) bir değer alma olasılığıyla ilgilenebiliriz. Kesin olasılık, tabanları $x = 7, 8$ ve 9 merkezli dikdörtgenlerin alanlarının toplamına eşit olan

$$\begin{aligned} P(7 \leq X \leq 9) &= \sum_{x=7}^9 b(x; 15, 0.4) - \sum_{x=0}^6 b(x; 15, 0.4) \\ &= 0.9662 - 0.6098 = 0.3564, \end{aligned}$$

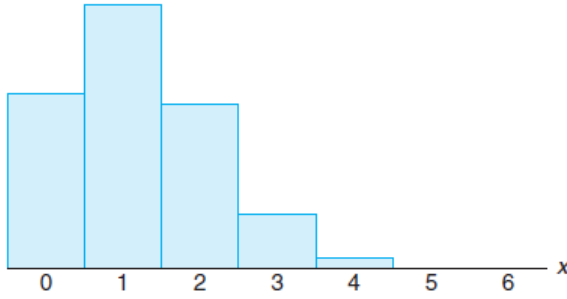
ile verilir.

Normal yaklaşım için, Şekil 6.23'te $x_1 = 6.5$ ve $x_2 = 9.5$ ordinatları arasındaki eğrinin altındaki gölgeli bölgenin alanını buluyoruz. Karşılık gelen z değerleri;

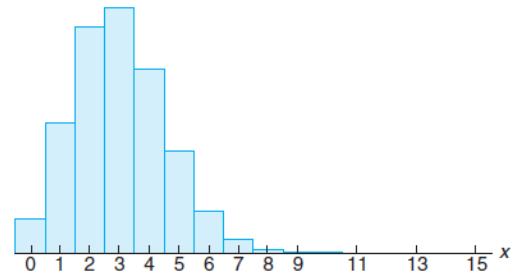
$$z_1 = \frac{6.5-6}{1.897} = 0.26 \text{ ve } z_2 = \frac{9.5-6}{1.897} = 1.85. \text{ Şimdi,}$$

$$\begin{aligned} P(7 \leq X \leq 9) &\approx P(0.26 < Z < 1.85) = P(Z < 1.85) - P(Z < 0.26) \\ &= 0.9678 - 0.6026 = 0.3652. \end{aligned}$$

Bir kez daha, normal eğri yaklaşımı, 0.3564 tam değeriyle çok yakından uyuşan bir değer sağlar. Eğrinin histograma ne kadar iyi uyduğuna bağlı olan doğruluk derecesi, n arttıkça artacaktır. Bu, p 1/2'ye çok yakın olmadığında ve histogram artık simetrik olmadığında özellikle doğrudur. Şekil 6.24 ve 6.25, sırasıyla $b(x; 6, 0.2)$ ve $b(x; 15, 0.2)$ için histogramları göstermektedir. Normal bir eğrinin histograma $n = 15$ olduğunda, $n = 6$ olduğu duruma göre çok daha iyi uyacağı açıktır.



Şekil 6.24: $b(x; 6, 0.2)$ için histogram.



Şekil 6.25: $b(x; 15, 0.2)$ için histogram.

Binom'a normal yaklaşımla ilgili çizimlerimizde, x 'in solundaki normal eğri altındaki alanı ararsak, $x + 0.5$ kullanmanın daha doğru olduğu ortaya çıktı. Bu, kesikli bir dağılımın sürekli bir dağılımla yaklaşık olarak tahmin edildiği gerçeğini karşılamak için yapılan bir düzeltmedir. +0.5 düzeltmesine süreklilik düzeltmesi denir. Yukarıdaki tartışma, iki terimli için aşağıdaki resmi normal yaklaşıma yol açar:

X , n ve p parametreleri ile bir binom rasgele değişkeni olsun. Büyük n için X , $\mu = np$ ve $\sigma^2 = npq = np(1 - p)$ ve

$$P(X \leq x) = \sum_{k=0}^x b(k; n, p)$$

$\approx x + 0.5$ 'in solundaki normal eğri altındaki alan

$$= P\left(Z \leq \frac{x + 0.5 - np}{\sqrt{npq}}\right),$$

ile yaklaşık olarak normal bir dağılıma sahiptir ve np ve $n(1 - p)$, **5'ten büyük veya ona eşitse yaklaşım iyi olacaktır.**

Daha önce belirttiğimiz gibi, yaklaşıklık kalitesi büyük n için oldukça iyidir. $p = 1/2$ 'ye yakınsa, makul bir yaklaşım için orta veya küçük bir örneklem boyutu yeterli olacaktır. Yaklaştırma kalitesinin bir göstergesi olarak Tablo 6.1'i sunuyoruz. Hem normal yaklaşım hem de gerçek binom kümülatif olasılıkları verilmiştir. $p = 0.05$ ve $p = 0.10$ 'da, $n = 10$ için yaklaşıklığın oldukça kaba olduğuna dikkat edin. Ancak, $n = 10$ için bile, $p = 0.50$ için iyileşmeye dikkat edin. Öte yandan, $p, p = 0.05$ 'te sabitlendiğinde, $n = 20$ 'den $n = 100$ 'e gidildikçe yaklaşımın iyileşmesine dikkat edin.

Tablo 6.1: Normal Yaklaşım ve Gerçek Kümülatif Binom Olasılıkları

r	$p = 0.05, n = 10$		$p = 0.10, n = 10$		$p = 0.50, n = 10$	
	Binomial	Normal	Binomial	Normal	Binomial	Normal
0	0.5987	0.5000	0.3487	0.2981	0.0010	0.0022
1	0.9139	0.9265	0.7361	0.7019	0.0107	0.0136
2	0.9885	0.9981	0.9298	0.9429	0.0547	0.0571
3	0.9990	1.0000	0.9872	0.9959	0.1719	0.1711
4	1.0000	1.0000	0.9984	0.9999	0.3770	0.3745
5			1.0000	1.0000	0.6230	0.6255
6					0.8281	0.8289
7					0.9453	0.9429
8					0.9893	0.9864
9					0.9990	0.9978
10					1.0000	0.9997

r	$p = 0.05$					
	$n = 20$		$n = 50$		$n = 100$	
	Binomial	Normal	Binomial	Normal	Binomial	Normal
0	0.3585	0.3015	0.0769	0.0968	0.0059	0.0197
1	0.7358	0.6985	0.2794	0.2578	0.0371	0.0537
2	0.9245	0.9382	0.5405	0.5000	0.1183	0.1251
3	0.9841	0.9948	0.7604	0.7422	0.2578	0.2451
4	0.9974	0.9998	0.8964	0.9032	0.4360	0.4090
5	0.9997	1.0000	0.9622	0.9744	0.6160	0.5910
6	1.0000	1.0000	0.9882	0.9953	0.7660	0.7549
7			0.9968	0.9994	0.8720	0.8749
8			0.9992	0.9999	0.9369	0.9463
9			0.9998	1.0000	0.9718	0.9803
10			1.0000	1.0000	0.9885	0.9941

Örnek 6.15: Bir hastanın nadir görülen bir kan hastalığından kurtulma olasılığı 0.4'tür. 100 kişinin bu hastalığa yakalandığı biliniyorsa, 30'dan azının hayatta kalma olasılığı nedir?

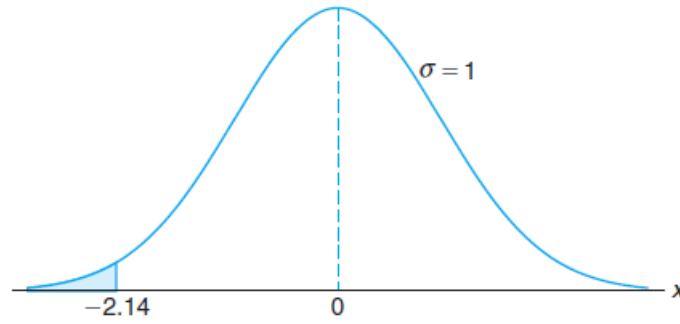
Cevap 6.15: Binom değişkeni X 'in hayatta kalan hasta sayısını temsil etmesine izin verin. $n = 100$ olduğundan, μ ile normal eğri yaklaşımını kullanarak oldukça doğru sonuçlar elde etmeliyiz.

İstenen olasılığı elde etmek için $x = 29.5$ 'in solundaki alanı bulmalıyız. (30'dan az olması)

29.5'e karşılık gelen z değeri

$$z = \frac{29.5 - 40}{4.899} = -2.14, \text{ 'dir}$$

ve 100 hastanın 30'undan azının hayatta kalma olasılığı Şekil 6.26'daki gölgeli bölge tarafından verilmiştir.



Şekil 6.26: Örnek 6.15 için Alan.

Dolayısıyla,

$$P(X < 30) \approx P(Z < -2.14) = 0.0162.$$

Örnek 6.16: 80 soruluk çoktan seçmeli bir sınav yapılmaktadır. Her bir sorunun 4 cevap seçeneği olup 1 tanesi doğrudur. Cevaplar rastgele seçilirse 25 ile 30 arası sorunun doğru cevaplanma olasılığı nedir?

Cevap 6.16: 80 sorunun her biri için doğru cevabı tahmin etme olasılığı $p = 1/4$ 'tür. X , tahminden kaynaklanan doğru cevapların sayısını temsil ediyorsa, o zaman

$$P(25 \leq X \leq 30) = \sum_{x=25}^{30} b(x; 80, 1/4).$$

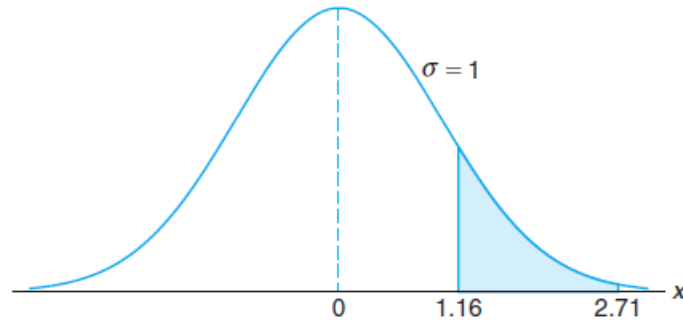
Normal eğri yaklaşımını kullanarak,

$$\mu = np = (80) \left(\frac{1}{4} \right) = 20$$
$$\sigma = \sqrt{npq} = \sqrt{(80)(1/4)(3/4)} = 3.873,$$

$x_1 = 24.5$ ile $x_2 = 30.5$ arasındaki alana ihtiyacımız var. Karşılık gelen z değerleri

$$z_1 = \frac{24.5 - 20}{3.873} = 1.16 \quad z_2 = \frac{30.5 - 20}{3.873} = 2.71.$$

25 ile 30 soruyu doğru tahmin etme olasılığı, Şekil 6.27'deki gölgeli bölge tarafından verilmektedir.



Şekil 6.27: Örnek 6.16 için Alan.

Tablo A.3'ten şunu buluyoruz:

$$P(25 \leq X \leq 30) = \sum_{x=25}^{30} b(x; 80, 0.25) \approx P(1.16 < Z < 2.71)$$
$$= P(Z < 2.71) - P(Z < 1.16) = 0.9966 - 0.8770 = 0.1196.$$

EK A3:

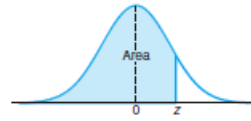


Table A.3 Areas under the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

OLASILIK VE İSTATİSTİK

13.HAFTA

BAZI SÜREKLİ OLASILIK DAĞILIMLARI (DEVAM)

6.6 GAMA VE ÜSTEL DAĞILIMLAR

Normal dağılım, mühendislik ve bilimdeki birçok problemi çözmek için kullanılabilse de, farklı türde yoğunluk fonksiyonları gerektiren çok sayıda durum vardır. Bu tür iki yoğunluk fonksiyonu, gama ve üstel dağılımlar bu bölümde tartışılmaktadır.

Üstel dağılımın, gama dağılımının özel bir durumu olduğu ortaya konmuştur. Her ikisi de çok sayıda uygulama içermektedir. Üstel ve gama dağılımları hem kuyruk teorisi hem de güvenilirlik problemlerinde önemli bir rol oynar. Servis tesislerine varışlar arasındaki süre ile bileşen parçalarının ve elektrik sistemlerinin arızalanma süresi genellikle üstel dağılımla güzel bir şekilde modellenir. Gama ve üstel arasındaki ilişki, gama'nın benzer problemlerde kullanılmasına izin verir. Daha fazla detay ve resim bu bölümde daha sonra verilecektir.

Gama dağılımı, adını matematiğin birçok alanında incelenen iyi bilinen gama fonksiyonundan alır. Gama dağılımına geçmeden önce, bu işlevi ve bazı önemli özelliklerini gözden geçirelim.

Tanım 6.2: Gama fonksiyonu

$$\Gamma(a) = \int_{-\infty}^{\infty} x^{a-1} e^{-x} dx, a > 0 \text{ için}$$

ile tanımlanır.

Aşağıda gama işlevinin birkaç basit özelliği verilmiştir.

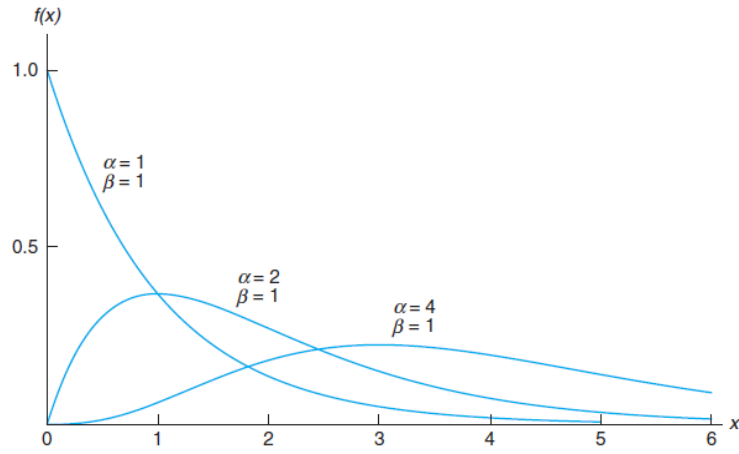
- a.) $\Gamma(n) = (n - 1)(n - 2) \cdots (1)\Gamma(1)$, n pozitif tam sayısı için. Yani, $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$.
- b.) $\Gamma(n) = (n - 1)!$, Pozitif bir tamsayı için.
- c.) $\Gamma(1) = 1$.
- d.) $\Gamma(1/2) = \sqrt{\pi}$

Sürekli rasgele değişken X , yoğunluk fonksiyonu

$$f(x; \alpha, \beta) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, & x > 0, \\ 0, & \text{elsewhere,} \end{cases}$$

tarafından verilirse, α ve β parametreleriyle bir gama dağılımına sahiptir, burada $\alpha > 0$ ve $\beta > 0$.

α ve β parametrelerinin belirli belirli değerleri için çeşitli gama dağılımlarının grafikleri Şekil 6.28'de gösterilmektedir.



Şekil 6.28: Gama dağılımları.

$\alpha = 1$ olan özel gama dağılımına üstel dağılım denir.

Sürekli rasgele değişken X , yoğunluk fonksiyonu

$$f(x; \beta) = \begin{cases} \frac{1}{\beta} e^{-x/\beta}, & x > 0, \\ 0, & \text{elsewhere,} \end{cases}$$

tarafından verilirse, β parametresiyle üstel bir dağılıma sahiptir, burada $\beta > 0$.

Aşağıdaki teorem ve sonuç, gama ve üstel dağılımların ortalamasını ve varyansını verir.

TEOREM 6.4: Gama dağılımının ortalaması ve varyansı

$\mu = \alpha\beta$ ve $\sigma^2 = \alpha\beta^2$ 'dir.

Sonuç 6.1: Üstel dağılımın ortalaması ve varyansı

$\mu = \beta$ ve $\sigma^2 = \beta^2$ 'dir.

A. Poisson Süreci ile İlişkisi

Üstel dağılımın uygulamalarını takip edeceğiz ve ardından gama dağılımına geri döneceğiz. Üstel dağılımın en önemli uygulamaları, Poisson sürecinin uygulandığı durumlardır. Poisson işleminin Poisson dağılımı adı verilen ayrık dağılımın kullanımına izin verdiğini hatırlamalıyız. Poisson dağılımının, belirli bir zaman diliminde veya uzayda belirli sayıda "olay" olasılığını hesaplamak için kullanıldığını hatırlayın. Birçok uygulamada, zaman periyodu veya uzay aralığı rastgele değişkendir. Örneğin, bir endüstri mühendisi, büyük bir şehirde trafiğin yoğun olduğu saatlerde sıkışık bir kavşağa varışlar arasındaki T süresini modellemekle ilgilenebilir. Bir varış, Poisson olayını temsil eder.

Üstel dağılım (genellikle negatif üstel olarak adlandırılır) ile Poisson süreci arasındaki ilişki oldukça basittir. Bölüm 5'te, Poisson dağılımı, λ parametresi ile tek parametrelili bir dağılım olarak geliştirilmiştir; burada λ , birim "zaman" başına ortalama olay sayısı olarak yorumlanabilir. Şimdi, ilk olayın meydana gelmesi için gereken süre ile açıklanan rasgele değişkeni ele alalım. Poisson dağılımını kullanarak, t zamanına kadar olan sürede herhangi bir olay meydana gelmeme olasılığının şu şekilde verildiğini buluruz:

$$p(0; \lambda t) = \frac{e^{-\lambda t} (\lambda t)^0}{0!} = e^{-\lambda t}.$$

Şimdi yukarıdakilerden faydalanabiliriz ve X 'in ilk Poisson olayının zamanı olmasına izin verebiliriz. İlk olaya kadar geçen sürenin x 'i geçme olasılığı, x 'te hiçbir Poisson olayının olmama olasılığı ile aynıdır. İkincisi, elbette, $e^{-\lambda x}$ ile verilir.

Sonuç olarak, $P(X > x) = e^{-\lambda x}$.

Böylece, X için kümülatif dağılım fonksiyonu

$$P(0 \leq X \leq x) = 1 - e^{-\lambda x}.$$

Şimdi, üstel dağılımın varlığını tanıyabilmemiz için, $\lambda = 1/\beta$ ile üstel dağılımın yoğunluk fonksiyonu olan

$$f(x) = \lambda e^{-\lambda x},$$

yoğunluk fonksiyonunu elde etmek için yukarıdaki kümülatif dağılım fonksiyonunu farklılaştırıyoruz.

B. Üstel ve Gama Dağılımlarının Uygulamaları

Yukarıda, üstel dağılımın "varışa kadar geçen süre" veya Poisson olayı problemlerine kadar geçen süre problemlerine uygulanması için temel sağladık. Burada bazı uygulamaları göstereceğiz ve ardından bu modelleme uygulamalarında gama dağılımının rolünü tartışmaya devam edeceğiz. Üstel dağılımın ortalamasının, Poisson dağılımındaki parametrenin tersi olan β parametresi olduğuna dikkat edin. Okuyucu, sık sık Poisson dağılımının hafızası olmadığına söylendiğini hatırlamalıdır, bu da ardışık zaman dilimlerindeki oluşumların bağımsız olduğunu ima eder. Önemli parametre β , olaylar arasındaki ortalama süredir. Ekipman arızasının genellikle bu Poisson sürecine uyduğu güvenilirlik teorisinde, β arızalar arasındaki ortalama süre olarak adlandırılır. Birçok ekipman arızası Poisson sürecini izler ve bu nedenle üstel dağılım geçerlidir. Diğer uygulamalar, biyomedikal deneylerde hayatta kalma sürelerini ve bilgisayar yanıt süresini içerir.

Aşağıdaki örnekte, üstel dağılımın bir güvenilirlik sorununa basit bir uygulamasını gösteriyoruz. Binom dağılımı da çözümde rol oynar.

Örnek 6.17: Bir sistemin, yıl cinsinden arızalanma süresi T tarafından verilen belirli bir tür bileşen içerdiğini varsayalım. Rastgele değişken T , ortalama arıza süresi $\beta = 5$ olan üstel dağılımla güzel bir şekilde modellenmiştir. Bu bileşenlerden 5 tanesi farklı sistemlere kurulursa, 8 yılın sonunda en az 2 tanesinin hala çalışır durumda olma olasılığı nedir?

Cevap 6.17: Belirli bir bileşenin 8 yıl sonra hala çalışıyor olma olasılığı

$$P(T > 8) = \frac{1}{5} \int_8^{\infty} e^{-t/5} dt = e^{-8/5} \approx 0.2.$$

ile verilir.

X , 8 yıl sonra çalışan bileşenlerin sayısını temsil etsin. Sonra binom dağılımını kullanarak,

$$P(X \geq 2) = \sum_{x=2}^5 b(x; 5, 0.2) = 1 - \sum_{x=0}^1 b(x; 5, 0.2) = 1 - 0.7373 = 0.2627.$$

Örnek 6.18: Belirli bir santrale gelen telefon aramalarının, dakikada ortalama 5 arama gelen bir Poisson sürecini izlediğini varsayalım. Santrale 2 arama gelene kadar 1 dakikanın geçmesi olasılığı nedir?

Cevap 6.18: Poisson süreci, $\beta = 1/5$ ve $\alpha = 2$ ile bir gama dağılımını takip eden 2 Poisson olayına kadar geçen süre için geçerlidir. 2 çağrı gelmeden önce geçen süreyi dakika cinsinden X ile belirtin. Gerekli olasılık

$$P(X \leq 1) = \int_0^1 \frac{1}{\beta^2} x e^{-x/\beta} dx = 25 \int_0^1 x e^{-5x} dx = 1 - e^{-5}(1 + 5) = 0.96.$$

ile verilir.

Gama dağılımının kökeni, α Poisson olaylarının meydana gelmesine kadar zaman (veya uzay) ile ilgili olsa da, net bir Poisson yapısı olmamasına rağmen bir gama dağılımının çok iyi çalıştığı birçok örnek vardır. Bu, hem mühendislik hem de biyomedikal uygulamalardaki sağlam kalma süresi problemleri için özellikle doğrudur.

Örnek 6.19: Farelerle yapılan bir biyomedikal çalışmada, bir toksik maddenin dozunun hayatta kalma süreleri üzerindeki etkisini belirlemek için bir doz-yanıt araştırması kullanılır. Zehirli madde, sıklıkla jet yakıtından atmosfere salınan maddedir. Belirli bir toksik madde dozu için çalışma, hafta cinsinden sağlam kalma süresinin $\alpha = 5$ ve $\beta = 10$ ile bir gama dağılımına sahip olduğunu belirler. Bir farenin 60 haftadan fazla hayatta kalmama olasılığı nedir?

Cevap 6.19: Rastgele değişken X , hayatta kalma süresi (ölüm süresi) olsun. Gerekli olasılık;

$$P(X \leq 60) = \frac{1}{\beta^5} \int_0^{60} \frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(5)} dx.$$

Yukarıdaki integral, gama dağılımı için kümülatif dağılım işlevi haline gelen eksik gama işlevinin kullanılmasıyla çözülebilir. Bu fonksiyon

$$F(x; \alpha) = \int_0^x \frac{y^{\alpha-1} e^{-y}}{\Gamma(\alpha)} dy.$$

olarak yazılır.

$y = x/\beta$ yani $x = \beta y$ kabul edersek, Ek A.23'teki tamamlanmamış gama fonksiyonu tablosunda $F(6; 5)$ olarak gösterilen

$$P(X \leq 60) = \int_0^6 \frac{y^4 e^{-y}}{\Gamma(5)} dy, \text{ 'yi}$$

elde ederiz. Bunun, gama dağılımı için olasılıkların hızlı bir şekilde hesaplanmasına izin verdiğini unutmayın. Nitekim bu problem için farenin 60 günden fazla hayatta kalmama olasılığı

$$P(X \leq 60) = F(6; 5) = 0.715.$$

tarafından verilmektedir.

Örnek 6.20: Önceki verilerden, belirli bir ürünle ilgili müşteri şikâyetleri arasında ay cinsinden geçen sürenin $\alpha = 2$ ve $\beta = 4$ olan bir gama dağılımı olduğu bilinmektedir. Kalite kontrol gerekliliklerini sıkılaştırmak için değişiklikler yapılmıştır. Bu değişikliklerin ardından ilk şikâyetin üzerinden 20 ay geçmiştir. Kalite kontrol sıkılaştırması etkiliymiş gibi mi görünüyor?

Cevap 6.20: Değişikliklerden önceki koşullar altında $\alpha = 2$ ve $\beta = 4$ ile bir gama dağılımı izleyen ilk şikâyetin zamanı X olsun. Soru, α ve β 'nin sırasıyla 2 ve 4 değerlerinde kaldığı göz önüne alındığında, $X \geq 20$ 'nin ne kadar nadir olduğu etrafında toplanıyor. Yani önceki koşullarda 20 ay gibi bir “şikâyet süresi” makul müdür? Böylece, Örnek 6.19'daki çözüm izlenerek,

$$P(X \geq 20) = 1 - \frac{1}{\beta^\alpha} \int_0^{20} \frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)} dx.$$

Yine, $y = x/\beta$ kullanarak,

$$P(X \geq 20) = 1 - \int_0^5 \frac{ye^{-y}}{\Gamma(2)} dy = 1 - F(5; 2) = 1 - 0.96 = 0.04,$$

burada $F(5; 2) = 0.96$, Tablo A.23'ten bulunur.

Sonuç olarak, gama dağılımının $\alpha = 2$ ve $\beta = 4$ koşullarının, gözlemlenen şikâyet süresinin 20 ay kadar büyük olduğu verileriyle desteklenmediği sonucuna varabiliriz. Bu nedenle, kalite kontrol çalışmasının etkili olduğu sonucuna varmak mantıklıdır.

6.7 Kİ-KARE DAĞILIMI

Gama dağılımının bir başka çok önemli özel durumu, v 'nin pozitif bir tam sayı olduğu $\alpha = v/2$ ve $\beta = 2$ kabul edilerek elde edilir. Sonuç, ki-kare dağılımı olarak adlandırılır. Dağılımın, serbestlik derecesi adı verilen v tek bir parametresi vardır.

Sürekli rasgele değişken X , yoğunluk fonksiyonu

$$f(x; v) = \begin{cases} \frac{1}{2^{v/2} \Gamma(v/2)} x^{v/2-1} e^{-x/2}, & x > 0, \\ 0, & \text{elsewhere,} \end{cases}$$

ile veriliyorsa, v pozitif bir tam sayı olmak üzere, v serbestlik dereceli bir ki-kare dağılımına sahiptir.

Ki-kare dağılımı, istatistiksel çıkarımda hayati bir rol oynar. Hem metodoloji hem de teoride önemli uygulamaları vardır. Ki-kare dağılımı, istatistiksel hipotez testi ve tahmininin önemli bir bileşenidir.

Örnekleme dağılımları, varyans analizi ve parametrik olmayan istatistiklerle ilgili konular, ki-kare dağılımının kapsamlı kullanımını içerir.

TEOREM 6.5: Ki-kare dağılımının ortalaması ve varyansı

$\mu = v$ ve $\sigma^2 = 2v$ 'dir.

6.8 BETA DAĞILIMI

Tek tip dağılımın (bir biçimli/uniform) bir uzantısı, bir beta dağılımıdır. Bir beta fonksiyonu tanımlayarak başlayalım.

Tanım 6.3: Bir beta fonksiyonu

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}, \alpha, \beta > 0$$

ile tanımlanır, burada $\Gamma(\alpha)$ gama fonksiyonudur.

Sürekli rasgele değişken X , yoğunluk fonksiyonu aşağıdaki gibi verilirse, $\alpha > 0$ ve $\beta > 0$ parametreleriyle bir beta dağılımına sahiptir:

$$f(x) = \begin{cases} \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, & 0 < x < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

$(0, 1)$ üzerindeki düzgün dağılımın, $\alpha = 1$ ve $\beta = 1$ parametreleriyle bir beta dağılımı olduğuna dikkat edin.

TEOREM 6.6: α ve β parametreleriyle bir beta dağılımının ortalaması ve varyansı sırasıyla,

$$\mu = \frac{\alpha}{\alpha + \beta} \text{ ve } \sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

$(0, 1)$ üzerindeki düzgün dağılım için, ortalama ve varyans sırasıyla;

$$\mu = \frac{1}{1+1} = \frac{1}{2} \text{ ve } \sigma^2 = \frac{(1)(1)}{(1+1)^2(1+1+1)} = \frac{1}{12}$$

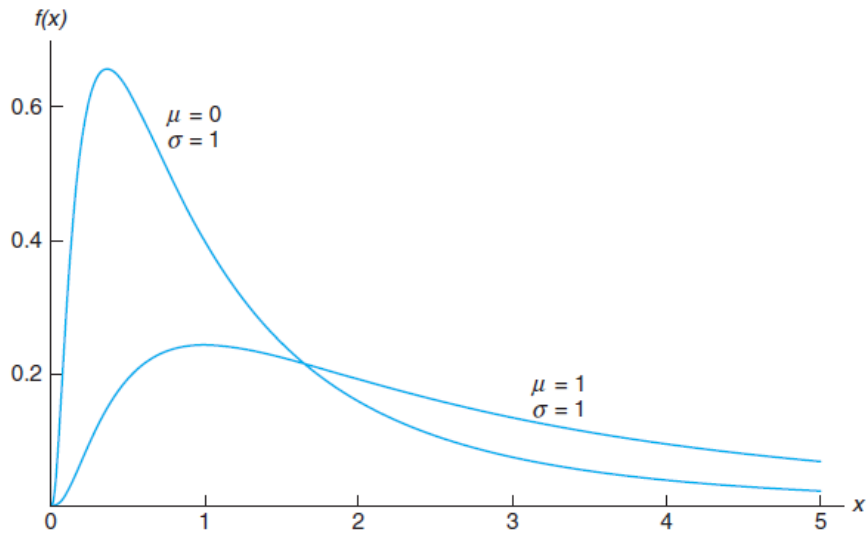
6.9 LOGNORMAL DAĞILIM

Lognormal dağılım, çok çeşitli uygulamalar için kullanılır. Dağılım, doğal log dönüşümünün normal dağılımla sonuçlandığı durumlarda geçerlidir.

Sürekli rasgele değişken X , rasgele değişken $Y = \ln(X)$ ortalama μ ve standart sapma σ ile normal bir dağılıma sahipse, bir lognormal dağılıma sahiptir. X 'in elde edilen yoğunluk fonksiyonu şu şekildedir:

$$f(x; \mu, \sigma) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2\sigma^2} [\ln(x) - \mu]^2}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Lognormal dağılımların grafikleri Şekil 6.29'da gösterilmektedir.



Şekil 6.29: Lognormal dağılımlar.

TEOREM 6.7: Lognormal dağılımın ortalaması ve varyansı

$\mu = e^{\mu + \sigma^2/2}$ ve $\sigma^2 = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$ şeklindedir.

Birikimli (kümülatif) dağılım fonksiyonu (cdf), normal dağılımla ilişkisi nedeniyle oldukça basittir. Dağılım fonksiyonunun kullanımı aşağıdaki örnekte gösterilmektedir.

Örnek 6.21: Tarihsel olarak kimyasal fabrikalar tarafından üretilen kirletici konsantrasyonlarının, lognormal dağılıma benzeyen davranışlar sergilediği bilinmektedir. Belirli bir kirleticinin milyonda parça cinsinden konsantrasyonunun $\mu = 3.2$ ve $\sigma = 1$ parametreleriyle lognormal bir dağılıma sahip olduğunu varsayalım. Konsantrasyonun milyonda 8 parçayı aşma olasılığı nedir?

Cevap 6.21: Rastgele değişken X 'in kirletici konsantrasyonu olduğunu varsayalım. Daha sonra;

$$P(X > 8) = 1 - P(X \leq 8).$$

$\ln(X)$, ortalama $\mu = 3,2$ ve standart sapma $\sigma = 1$ olan normal bir dağılıma sahip olduğundan,

$$P(X \leq 8) = \Phi \left[\frac{\ln(8) - 3.2}{1} \right] = \Phi(-1.12) = 0.1314.$$

Burada, standart normal dağılımın kümülatif dağılım fonksiyonunu belirtmek için Φ kullanıyoruz. Sonuç olarak, kirletici konsantrasyonunun milyonda 8 parçayı aşma olasılığı 0,1314'tür.

Örnek 6.22: Lokomotifler için belirli bir elektronik kontrol tipinin binlerce mil cinsinden ömrü, $\mu = 5.149$ ve $\sigma = 0.737$ ile yaklaşık olarak lognormal bir dağılıma sahiptir. Böyle bir elektronik kontrolün ömrünün yüzde 5'ini bulunuz.

Cevap 6.22: Tablo A.3'ten $P(Z < -1.645) = 0.05$ olduğunu biliyoruz. Böyle bir elektronik kontrolün ömrünü X ile belirtin. $\ln(X)$, ortalama $\mu = 5.149$ ve $\sigma = 0.737$ ile normal bir dağılıma sahip olduğundan, X 'in yüzde 5'lik değeri

$$\ln(x) = 5.149 + (0.737)(-1.645) = 3.937$$

olarak hesaplanabilir.

Dolayısıyla, $x = 51.265$. Bu, kontrollerin yalnızca %5'inin 51.265 milden daha az kullanım ömrüne sahip olacağı anlamına gelir.

EK A3:

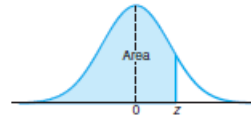


Table A.3 Areas under the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

EK A.23

Table A.23 The Incomplete Gamma Function: $F(x; \alpha) = \int_0^x \frac{1}{\Gamma(\alpha)} y^{\alpha-1} e^{-y} dy$

x	α									
	1	2	3	4	5	6	7	8	9	10
1	0.6320	0.2640	0.0800	0.0190	0.0040	0.0010	0.0000	0.0000	0.0000	0.0000
2	0.8650	0.5940	0.3230	0.1430	0.0530	0.0170	0.0050	0.0010	0.0000	0.0000
3	0.9500	0.8010	0.5770	0.3530	0.1850	0.0840	0.0340	0.0120	0.0040	0.0010
4	0.9820	0.9080	0.7620	0.5670	0.3710	0.2150	0.1110	0.0510	0.0210	0.0080
5	0.9930	0.9600	0.8750	0.7350	0.5600	0.3840	0.2380	0.1330	0.0680	0.0320
6	0.9980	0.9830	0.9380	0.8490	0.7150	0.5540	0.3940	0.2560	0.1530	0.0840
7	0.9990	0.9930	0.9700	0.9180	0.8270	0.6990	0.5500	0.4010	0.2710	0.1700
8	1.0000	0.9970	0.9860	0.9580	0.9000	0.8090	0.6870	0.5470	0.4070	0.2830
9		0.9990	0.9940	0.9790	0.9450	0.8840	0.7930	0.6760	0.5440	0.4130
10		1.0000	0.9970	0.9900	0.9710	0.9330	0.8700	0.7800	0.6670	0.5420
11			0.9990	0.9950	0.9850	0.9620	0.9210	0.8570	0.7680	0.6590
12			1.0000	0.9980	0.9920	0.9800	0.9540	0.9110	0.8450	0.7580
13				0.9990	0.9960	0.9890	0.9740	0.9460	0.9000	0.8340
14				1.0000	0.9980	0.9940	0.9860	0.9680	0.9380	0.8910
15					0.9990	0.9970	0.9920	0.9820	0.9630	0.9300