İstatistik

Ders notları

Hüseyin Taştan

İçindekiler

Şekil Listesi

Tablo Listesi

Önsöz

Bu kitap İstatistik I ve İstatistik II ders notlarının bir araya getirilmesiyle oluşturulmuştur. İstatistik I, betimsel istatistik, olasılık, ve dağılım teorisi konularını kapsamaktadır. Ders planına buradan ulaşabilirsiniz: https://htastan.github.io/istatistik/Istatistik_I_Syllabus_2024_Fall.pdf

İstatistik I dersinin amacı, öğrencilere temel istatistik kavramlarını ve yöntemlerini tanıtarak, veriye dayalı analiz yapabilme becerisi kazandırmaktır. Bu ders, öğrencilerin ekonomi ve sosyal bilimler alanındaki problemleri istatistiksel yöntemlerle analiz edebilme yeteneğini geliştirmeyi hedefler. Ayrıca, öğrencilerin veri toplama, düzenleme, analiz etme ve yorumlama süreçlerini etkin bir şekilde yürütebilmeleri için gerekli teorik ve uygulamalı bilgi altyapısını sunar.

Ders kapsamında, betimsel istatistik, olasılık kuramı, kesikli ve sürekli olasılık dağılımları, normal dağılım, Merkezi Limit Teoremi, ve örnekleme kavramları incelenecektir. Bu dersin ikinci kısmını oluşturan İstatistik II dersinde ise ağırlıklı olarak çıkarsama konuları ele alınacaktır (nokta ve aralık tahmini, hipotez testleri, varyans analizi). Öğrenciler bu kavramları hem teorik düzeyde hem de çeşitli uygulamalar aracılığıyla öğrenerek, özellikle iktisadi veriler üzerinde anlamlı çıkarımlar yapma yetisi kazanacaklardır. Ders, ekonominin yanı sıra sosyal bilimlerin diğer alanlarında karşılaşılan problemlerin çözümünde istatistiksel araçların nasıl kullanılacağını göstermeyi amaçlamaktadır.

İstatistik yazılımı: Derslerde ve laboratuvar oturumlarında `R` kullanacağız. `R`, istatistiksel hesaplamalar ve grafikler için kullanılan, istatistikçiler, araştırmacılar, veri bilimciler ve ekonometrisyenler ile endüstri profesyonelleri tarafından yaygın olarak tercih edilen açık kaynaklı bir yazılımdır. `R`'ın en son sürümünü şu adresten indirebilirsiniz:

https://www.r-project.org/

R için entegre bir geliştirme ortamı olarak R-studio kullanılabilir:

https://www.rstudio.com/products/RStudio/

1 Olasılık

Olasılık teorisi, günlük yaşamımızdaki pek çok belirsiz durumun analizine temel oluşturur. Ekonominin resesyona girme riski, bir spor karşılaşmasının sonucu veya bir yatırımın başarılı olma ihtimali gibi konular, olasılık teorisinin bize sunduğu araçlarla açıklanabilir. Olasılık, belirsizlikleri matematiksel bir çerçevede sistematik olarak değerlendirip öngörülerde bulunmamıza olanak tanır.

Olasılık, izleyen bölümlerde inceleyeceğimiz dağılım teorisinin de temelini oluşturur. Normal, binom ve Poisson gibi olasılık dağılımları, olayların nasıl ve ne sıklıkta gerçekleşeceğini anlamamıza olanak tanır. Bunun yanı sıra, olasılık teorisi, istatistiksel çıkarsama tekniklerinin dayandığı teorik çerçeveyi sağlar. Örneğin, güven aralıkları oluşturma, hipotez test etme ve tahmin yapma gibi süreçlerin arkasında olasılık kuramı yatar.

Bu bölümde, olasılığın temel kavramlarını inceleyeceğiz. İlk olarak, olasılığın dayandığı temel yapı taşlarından biri olan küme teorisini ve kümelerle ilgili işlemleri ele alacağız. Daha sonra, olasılığın nesnel ve öznel tanımlarını yapacağız. Geliştirdiğimiz bu araçları kullanarak olaylar üzerindeki olasılık hesaplarını ve olasılık kurallarını öğreneceğiz.

1.1 Örneklem Uzayı ve Olaylar

Olasılık teorisinde olaylar, tüm sonuçlardan oluşan örneklem uzayının belirli alt kümeleridir. Bu nedenle, olayları ve bunların arasındaki ilişkileri anlamak için küme teorisinin temel kavramlarına ihtiyaç duyarız. Küme teorisi, olasılık hesaplamalarında olayların birleşimi, kesişimi ve tamamlayıcı olayların tanımlanmasında önemli bir araçtır. Şimdi, küme kavramlarını ve olaylar arasındaki temel işlemleri inceleyerek bu ilişkileri nasıl ifade ettiğimizi göreceğiz.

1.1.1 Temel küme işlemleri

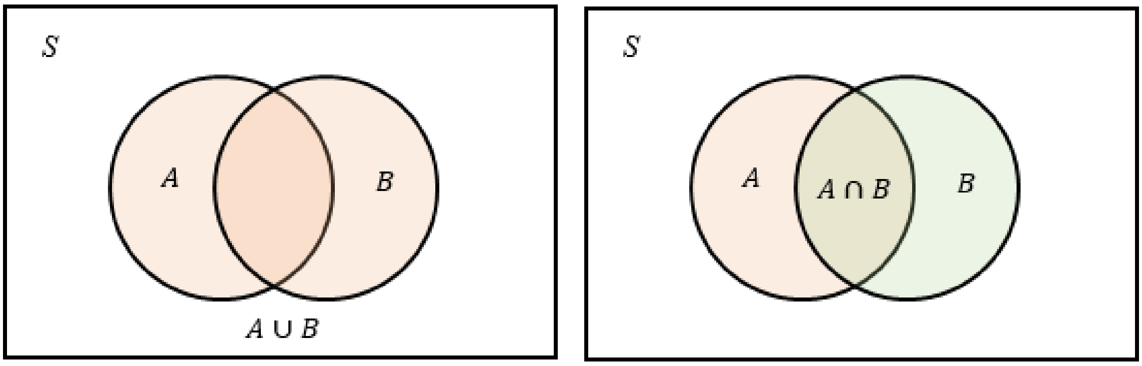
Bir küme birbirinden farklı nesnelerden oluşan bir topluluktur. Genellikle büyük harflerle gösterilir. Örneğin

$$A = \{a, b, c, d, e\},$$

ve

$$B = \{d, e, f, g, h\}$$

kümeleri beş nesneden oluşmaktadır. Bir nesne kümenin elemanıysa $\in$ ile gösterilir, örneğin $a \in A$ yazabiliriz. Tersi durumda $\notin$ notasyonunu kullanabiliriz, örneğin $a \notin B$.



(a) Birleşim: A veya B

(b) Kesişim: A ve B

Şekil 1.1: Kümelerin kesişimi ve birleşimi

Kümelerin **birleşimi** A'da ya da B de (ya da her ikisinde) yer alan nesneleri içerir (Şekil 1.1a):

$$A \cup B = \{x \mid x \in A \text{ ya da } x \in B\}$$

Kümelerin birleşimini indeks kümesi I olan bir küme dizisi için de tanımlayabiliriz:

$$\bigcup_{i \in I} A_i = \{a \mid a \in A_i \text{ her } i \in I\}$$

Örnek 1.1. Yukarıda tanımladığımız A ve B kümelerinin birleşimi:

$$A \cup B = \{a, b, c, d, e, f, g, h\}$$

olur.

Kümelerin **kesişimi** ise hem A hem de B'de yer alan nesnelerden oluşur (Şekil 1.1b):

$$A \cap B = \{x \mid x \in A \text{ ve } x \in B\}$$

Örnek 1.2. A ve B kümelerinin kesişim kümesi:

$$A \cap B = \{d, e\}$$

olur.

Küme farkı bir kümenin diğer kümeden farklı olan elemanlarının oluşturduğu yeni küme olarak tanımlanır ve $A \setminus B$ ya da $A - B$ ile gösterilir (bkz Şekil 1.2a).

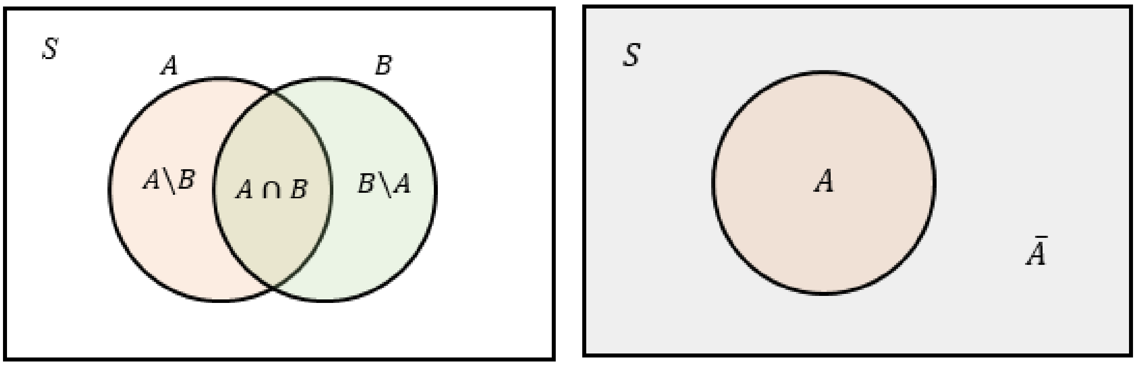
Örnek 1.3. A ve B kümelerinin farkı

$$A \setminus B = \{a, b, c\}$$

A'da olup B'de olmayan elemanları içerir. B'de olup A'da olmayanlar ise:

$$B \setminus A = \{f, g, h\}$$

olur.



(a) Küme farkı

(b) Tümleyen

Şekil 1.2: Kümelerin farkı ve tümleyeni

Küme eşitliği iki kümenin tam olarak aynı elemanlardan oluşması durumudur ve $A = B$ ile gösterilir.

Örnek 1.4.

$$A = \{a, b, c, d, e\}, \quad B = \{d, e, f, g, h\} \quad \Rightarrow \quad A \neq B.$$

Her iki küme aynı elemanlardan oluşmadığı için A ve B kümeleri eşit değildir.

Bir kümenin **tümleyeni** o kümede yer almayan sonuçlardan oluşan kümedir ve A^c ya da $\bar{A}$ ile gösterilir (bkz Şekil 1.2b).

Örnek 1.5.

$$A = \{a, b, c, d, e\}, \quad B = \{d, e, f, g, h\}$$

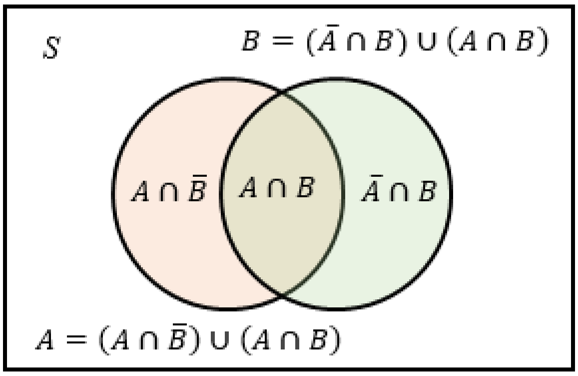
tanımlı olsun. A'nın tümleyeni

$$\bar{A} \equiv A^c = (A \cup B) \setminus A = \{a, b, c, d, e, f, g\} \setminus \{a, b, c, d, e\} = \{f, g\}$$

ve B'nin tümleyeni

$$\bar{B} \equiv B^c = (A \cup B) \setminus B = \{a, b, c, d, e, f, g\} \setminus \{d, e, f, g\} = \{a, b, c\}$$

olur.



Şekil 1.3: Kümelerin kesişimi ve tümleyeni

A ve B, S evreninde tanımlı iki olay ise

$$A = (A \cap \bar{B}) \cup (A \cap B)$$

yazılabilir. Yani, hem B'nin tümleyeninde hem de A'da yer alan elemanlar ile A da ve B de yer alan elemanların birleşimi A'yı verir. Benzer şekilde

$$B = (\bar{A} \cap B) \cup (A \cap B)$$

yazılabilir (bkz. Şekil 1.3).

Örnek 1.6. S evreni aşağıdaki gibi tanımlansın:

$$S = \{a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z\}$$

Önceki örneklerdeki gibi

$$A = \{a, b, c, d, e\}, \quad B = \{d, e, f, g, h\}$$

tanımlı olsun. A'nın tümleyeni

$$\bar{A} = \{f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z\}$$

ve B'nin tümleyeni

$$\bar{B} = \{a, b, c, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z\}$$

olduğuna göre

$$A \cap \bar{B} = \{a, b, c\}, \quad A \cap B = \{d, e\}$$

ve buradan

$$(A \cap \bar{B}) \cup (A \cap B) = \{a, b, c, d, e\} = A$$

bulunur. Benzer şekilde

$$(\bar{A} \cap B) \cup (A \cap B) = \{f, g, h\} \ \cup \ \{d, e\} = B$$

olduğu gösterilebilir.

Örnek 1.7 (R ile küme işlemleri)**.** R programını kullanarak birleşim (union), kesişim (intersect), küme farkı (setdiff) ve küme eşitliği (setequal) gibi işlemleri yapabiliriz.

```
# R'da nasıl yapılır
# A ve B kümelerini string vektör olarak tanımlayalım:
A <- c("Elma", "Armut", "Kiwi")
A
```

```
[1] "Elma"  "Armut" "Kiwi"
```

```
B <- c("Elma", "Portakal", "Erik", "Kiwi")
B
```

```
[1] "Elma"     "Portakal" "Erik"     "Kiwi"
```

```
# A ve B kümelerinin birleşimi
union(A, B)
```

```
[1] "Elma"     "Armut"    "Kiwi"     "Portakal" "Erik"
```

```
# A ve B kümelerinin kesişimi
intersect(A, B)
```

```
[1] "Elma" "Kiwi"
```

```
# A\B, A fark B = A'da olup B'de olmayan nesneler
setdiff(A, B)
```

```
[1] "Armut"
```

```
# B\A, B fark A = B'de olup A'da olmayan nesneler
setdiff(B, A)
```

```
[1] "Portakal" "Erik"
```

```
# kümeler eşit mi
setequal(A, B)
```

```
[1] FALSE
```

Bir nesnenin (ya da nesnelerin) bir kümenin elemanı olup olmadığını bulmak için `is.element()` fonksiyonu kullanılabilir:

```r
is.element("Portakal", A)
```

```
[1] FALSE
```

```r
# alternatif:
"Portakal" %in% A
```

```
[1] FALSE
```

```r
is.element(c("Portakal", "Elma"), A)
```

```
[1] FALSE  TRUE
```

1.1.2 Rassal deney, örneklem uzayı, olay

Bir küme gözlem veya ölçümlerden oluşabilir. Bu gözlemleri üreten sürece **deney** ve deneyin eğer mümkünse tekrarlanmasına ise deneme ya da **tekrar** adı verilir. Olanaklı tüm sonuçlardan oluşan kümeye **örneklem uzayı** ya da **evrensel küme** adı verilir ve S ile gösterilir. Örneğin hilesiz bir para atılırsa bu ya yazı (Y) ya da tura (T) ile sonuçlanır. Öyleyse örneklem uzayı

$$S = \{T, Y\}$$

olur.

Örneklem uzayı S'nin bir alt kümesine **olay** denir ve genellikle büyük harflerle gösterilir. Olaylar örneklem uzayında tanımlı temel olaylardan oluşur. Örneğin bir zar atıldığı zaman örneklem uzayı $S = \{1, 2, 3, 4, 5, 6\}$ kümesiyle ifade edilir. Eğer A olayı çift gelmesi olarak tanımlandıysa, yani, $A = \{2, 4, 6\}$ ise 2, 4, veya 6 olaylarından herhangi biri gerçekleşirse A olayı gerçekleşmiş olur.

İki olayın birlikte gerçekleşmisini bu olayların kesişimi ile ifade edebiliriz. Örneğin A ve B olaylarının kesişimi, $A \cap B$ hem A'nın hem de B'nin gerçekleştiğini ifade eder.

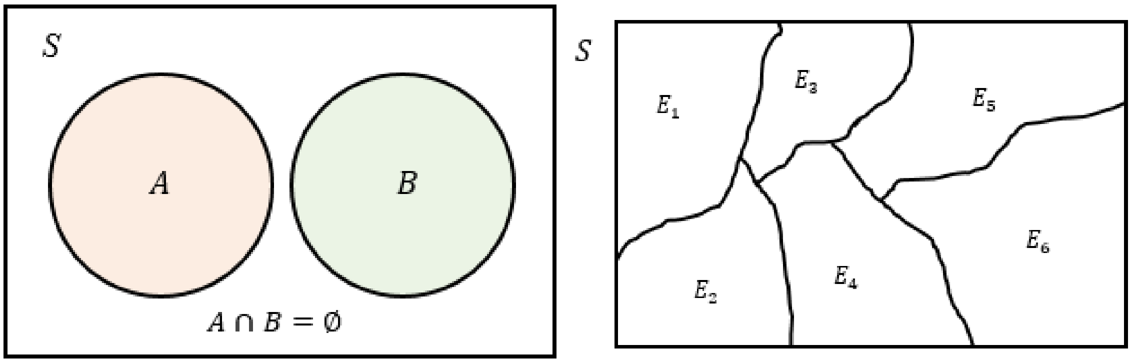
A ve B olaylarının kesişim kümesi boş kümeyse, yani hiçbir elemanı yoksa

$$A \cap B = \emptyset$$

bu olaylara **karşılıklı bağdaşmaz** ya da **ayrık** olaylar denir (bkz. Şekil 1.4). Daha genel olarak $E_1, E_2, \ldots, E_k$ örneklem uzayı S'de tanımlı olaylar ise ve her bir E_i, E_j çifti birlikte gerçekleşmiyorsa (kesişimleri boş kümeyse) bu olaylar karşılıklı bağdaşmazdır. Bu şekilde karşılıklı bağdaşmaz olayların birleşimi

$$S = E_1 \cup E_2 \cup \ldots \cup E_k$$

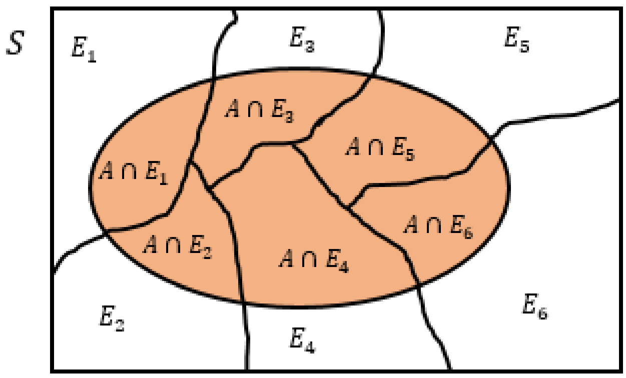
örneklem uzayını veriyorsa bu olaylara karşılıklı bağdaşmaz ve **bütünü kapsayıcı** denir (bkz. Şekil 1.4b).



(a) Karşılıklı bağdaşmaz (ayrık) olaylar

(b) Bütünü kapsayıcı ve karşılıklı bağdaşmaz

Şekil 1.4: Karşılıklı bağdaşmaz ve bütünü kapsayıcı olaylar



Şekil 1.5: Karşılıklı bağdaşmaz ve bütünü kapsayıcı temel olaylar ve A olayı

Şekil 1.5 örneklem uzayı S'de tanımlı, karşılıklı bağdaşmaz ve bütünü kapsayıcı $E_i,\ i = 1, 2, \ldots, 6$ temel olaylarını, ve A olayını göstermektedir. Her bir E_i olayı ile A'nın kesişimlerinin birleşimi A olayını verir:

$$A = (E_1 \cap A) \cup (E_2 \cap A) \cup (E_3 \cap A) \cup (E_4 \cap A) \cup (E_5 \cap A) \cup (E_6 \cap A)$$

Para atma deneyinde $A = \{T\}$ ve $B = \{Y\}$ olayları ayrık olaylardır. Her iki olay da örneklem uzayının bir alt kümesidir ve $A \subset S$ ile gösterilir. Kesişimleri ise boş kümedir, yani hiç bir elemanı yoktur. Ayrıca bu iki olay bütünü kapsayıcıdır.

Örnek 1.8. İki madeni para atılıyor ve sonuçlar kaydediliyor. Örneklem uzayını oluşturunuz. A "en az bir Yazı gelmesi" kümesinde kaç eleman vardır?

Çözüm:

İki madeni para atıldığında ortaya çıkabilecek tüm sonuçların kümesi

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

olarak yazılabilir. En az bir Yazı gelmesi şeklinde tanımlanan A olayı

$$A = \{TY, YT, YY\}$$

üç temel sonuçtan oluşur. Bu olaylardan herhangi biri gerçekleşirse A olayı gerçekleşmiş olur.

Örnek 1.9. İki zar atılıyor ve sonuçlar kaydediliyor. A olayı toplamın 4 olması B olayı ise toplamın 5'ten küçük olması şeklinde tanımlanıyor. Bu kümeleri oluşturunuz. $A \cap B$ kaç eleman içerir?

Çözüm:

Bir zar atıldığında olanaklı sonuçlardan oluşan küme

$$Z_1 = \{1, 2, 3, 4, 5, 6\}$$

ve benzer şekilde ikinci zar için

$$Z_2 = \{1, 2, 3, 4, 5, 6\}$$

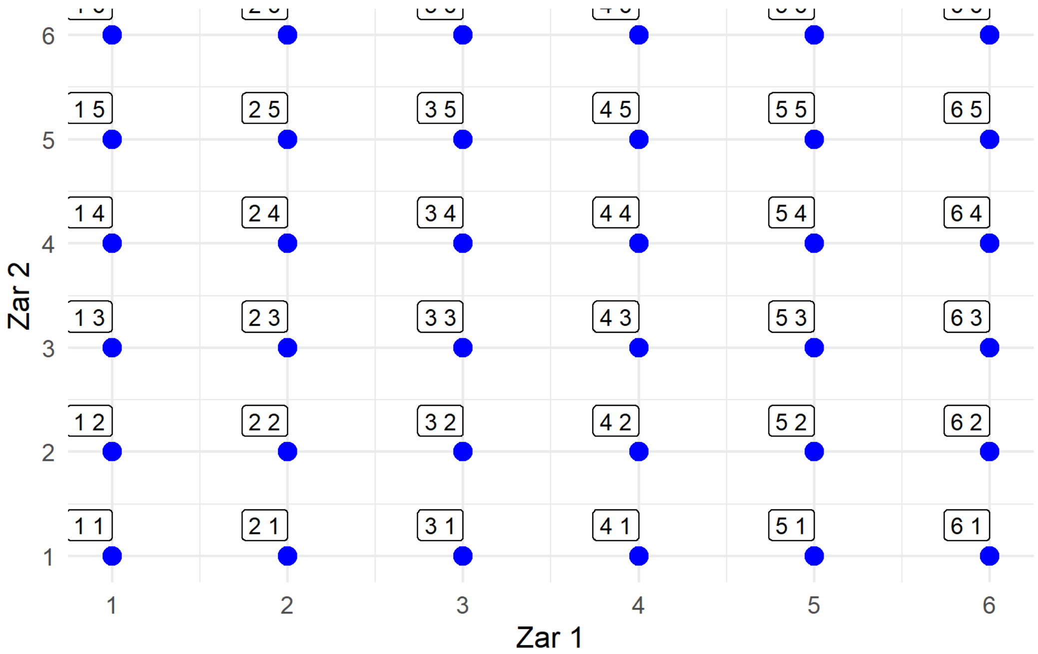
olur. İki zar atıldığında ortaya çıkabilecek sonuçlar $Z_{1i}, Z_{2j}, i, j = 1, \ldots, 6$ olacaktır. Öyleyse örneklem uzayının temel sonuçları, birinci zar ve ikinci zarın sonuçları (Z_1, Z_2) olmak üzere aşağıdaki gibi yazılabilir:

$$S = \{(1,1),\ (1,2),\ (1,3),\ (1,4), \ldots, (6,4),\ (6,5),\ (6,6)\}$$

Bu örneklem uzayı 36 temel sonuçtan oluşmaktadır (bkz Şekil 1.6).

A olayı zarların toplamının 4 olması

$$A = \{(1,3),\ (3,1),\ (2,2)\}$$



Şekil 1.6: İki zar atışı için örneklem uzayı

B ise toplamın beşten küçük olması:

$$B = \{(1,1),\ (1,2),\ (2,1),\ (1,3),\ (3,1),\ (2,2)\}$$

olarak tanımlanmıştı. Bu olayların kesişimi

$$A \cap B = \{(1,3),\ (3,1),\ (2,2)\}$$

üç eleman içermektedir.

Örnek 1.10. Örneklem uzayı S tüm pozitif tamsayılardan oluşsun (sayma sayıları). $A \subset S$ kümesi çift tamsayılardan oluşsun. A'nın tümleyenini, $\bar{A}$, bulunuz.

Çözüm: Örneklem uzayı

$$S = \{1, 2, 3, 4, ...\}$$

A olayı:

$$A = \{2, 4, 6, ...\}$$

A kümesinin tümleyeni

$$\bar{A} = \{1, 3, 5, ...\}$$

olur. Birleşimleri örneklem uzayını verir:

$$A \cup \bar{A} = S$$

Kesişimleri ise boş kümedir:

$$A \cap \bar{A} = \emptyset$$

1.2 Olasılık Tanımları

Hilesiz bir para atıldığında T ve Y olasılıklarının %50-%50 olduğunu hepimiz kabul ederiz. İki sonuçtan oluşan başka bir deney tasarlayalım: BIST100 endeksinin günlük kapanış yönünü gözlemliyoruz. Bu durumda, günlük kapanış ya artar, ya sabit kalır, ya da azalır. Bu olayların sonuçlarını

$$S = \{+, \ -\}$$

ile ifade edebiliriz. Burada +, endeksin yukarı yönlü hareketini veya sabit kalmasını, − ise aşağı yönlü hareketi belirtir. Peki, endeksin yükselme olasılığını nasıl hesaplayabiliriz? Yazı-Tura deneyinde olduğu gibi bu sonuca %50-%50 olasılık atamak uygun olur mu?

İzleyen bölümlerde A olayının olasılığını $\mathbb{P}(A)$ ile göstereceğiz:

$$0 \leq \mathbb{P}(A) \leq 1$$

$\mathbb{P}(A)$, **olasılık ölçüsünü** ifade eder ve "A olayının olasılığı" şeklinde okunur. Olasılık ölçüsü $\mathbb{P}(\cdot)$ olanaklı tüm sonuçlara 0 ile 1 arasında değerler atayan bir fonksiyon olarak düşünülebilir. $\mathbb{P}(A) = 0$, olayın gerçekleşmeyeceği, $\mathbb{P}(A) = 1$ ise kesin olarak gerçekleşeceği anlamına gelir.

1.2.1 Klasik Yaklaşım

Klasik olasılık yaklaşımı, olasılığı teorik bir çerçevede tanımlar ve bir olayın tüm olası sonuçlarının **eşit olasılıkla gerçekleşeceği** varsayımına dayanır. Bu yaklaşımda, bir olayın olasılığı, o olayın istenen sonuçlarının, tüm mümkün sonuçlara oranı olarak tanımlanır. Başka bir deyişle, belirli bir olayın gerçekleşme olasılığı, tüm olanakların eşit olduğu durumlarda, bu olayın olası sonuçlarının toplam olası sonuçlara oranıdır.

Bir deneyde gözlemlenebilecek toplam sonuç sayısını N ve ilgilenilen olayın bu sonuçlardan kaçını kapsadığını N_A ile gösterelim. Bu durumda, klasik olasılık yaklaşımına göre olayın olasılığı şöyle tanımlanır:

$$\mathbb{P}(A) = \frac{N_A}{N}$$

Örneğin bir madeni para atıldığında olanaklı sonuçlar, Tura ve Yazı, eşit olasılığa sahiptir. Olanaklı sonuç $N = 2$ ve $N_A = 1$ olduğuna göre

$$\mathbb{P}(\text{Tura}) = \mathbb{P}(\text{Yazı}) = \frac{1}{2}$$

yazabiliriz.

Bir zar atma deneyini ele alalım. Altı yüzlü hilesiz bir zarın yüzeylerinden herhangi birinin gelme olasılığı eşittir, 1/6. Bu durumda zarın her bir yüzeyi toplam altı olası sonuçtan biridir. A olayını tek sayı gelmesi olarak tanımlarsak

$$\mathbb{P}(A) = \frac{N_A}{N} = \frac{3}{6} = \frac{1}{2}$$

A olayının gerçekleşmesiyle sonuçlanan temel sonuçlar $\{1, 3, 5\}$ olduğu için $N_A = 3$ olur.

Klasik olasılık yaklaşımı, tüm olası sonuçların eşit olasılıkla gerçekleştiği durumlarda oldukça kullanışlıdır. Eşit olasılık varsayımı altında N_A ve N'nin sayma yöntemleriyle bulunması gerekebilir (bu konuları izleyen alt bölümlerde inceleyeceğiz).

Bazı durumlarda ise eşit olasılık varsayımı geçerli olmayabilir ve klasik yaklaşım yetersiz kalır. Örneğin, hisse senedi piyasasının yönü veya hava durumu gibi olayların olasılıkları, sonuçların eşit olasılıkla oluşmasını beklemek için çok fazla belirsizlik ve değişken içerir. Bu tür durumlarda, olayların gözlemler yoluyla hesaplandığı göreli sıklık yaklaşımı veya bireysel değerlendirmelere dayanan öznel yaklaşım daha uygun hale gelir.

1.2.2 Göreli Sıklık Yaklaşımı

Göreli sıklık (frekans) yaklaşımında bir deneyin çok sayıda tekrarlandığı düşünülür. Bu tanıma göre, A olayının olasılığı, deney sonsuz sayıda tekrarlandığında gerçekleşme oranına eşittir:

$$\mathbb{P}(A) = \lim_{n\to\infty} \frac{\text{A olayının gerçekleşme sayısı}}{\text{Toplam tekrar sayısı}} = \lim_{n\to\infty} \frac{n_A}{n} \tag{1.1}$$

Burada n_A, A olayının gerçekleşme sayısını, n ise tekrar sayısını göstermektedir.

Hilesiz bir para atıldığında Tura (T) sonucuna

$$\mathbb{P}(\text{Tura gelmesi}) = 0.5$$

olasılığını atamıştık. A olayını "Tura gelmesi" olarak tanımlarsak sıklık yaklaşımına göre

$$\mathbb{P}(A) = \lim_{n\to\infty} \frac{\text{Tura sayısı}}{\text{Toplam para atışı tekrar sayısı}} = 0.5$$

yazabiliriz.

Bu tanımı somutlaştırmak için `R`'da bir simülasyon yapabiliriz. Kullanıcı tarafından girilen bir vektörden rassal çekilişler yapmamızı sağlayan `sample()` fonksiyonunu kullanarak olasılıkları yaklaşık olarak hesaplayabiliriz. Olanaklı sonuçları `"T", "Y"` olarak tanımlayacağız ve yerine koyma yöntemiyle bu iki sonuçtan birini rassal olarak çekeceğiz.

```r
# sample() fonksiyonu ile madeni para deneyi
set.seed(2468) # aynı sonuçlar için gerekli başlangıç
n <- 10
yazı_tura_deneyi <- sample(c("T", "Y"),   # olanaklı sonuçlar
                           n,             # tekrar sayısı
                           replace = TRUE # yerine koyarak tekrarla
                           )
# sonuçları göster
yazı_tura_deneyi
```

```
 [1] "Y" "Y" "T" "T" "Y" "Y" "Y" "T" "Y" "T"
```

Bu deneyin 10 tekrarlık bir dizilimi yukarıda verildi. Buna göre 10 tekrardan

```
sum(yazı_tura_deneyi == "T") # Tura gelen durumların sayısı
```

```
[1] 4
```

dördünün tura geldiğini görüyoruz. Klasik olasılık tanımı n sonsuza giderken bu oranın ilgilendiğimiz olasılığa yaklaştığını söyler.

Olasılığın sıklık tanımının nasıl çalıştığını görmek için bu deneyi birden 100'e kadar tekrarlayalım, $n = 1, 2, \dots, 100$. Her tekrar için T gelme oranını hesaplayalım:

```
set.seed(123)   # replikasyon için başlangıç
N <- 100        # tekrar sayısı
P100 <- numeric(N)
for (n in 1:N){
  x <- sample(c(1, 0),      # 1=Tura, 0=Yazı
              n,
              replace = TRUE)
  P100[n] <- sum(x==1)/n   # Tura=1 gelme sayısının toplama oranı
}
P100[N]
```

```
[1] 0.44
```

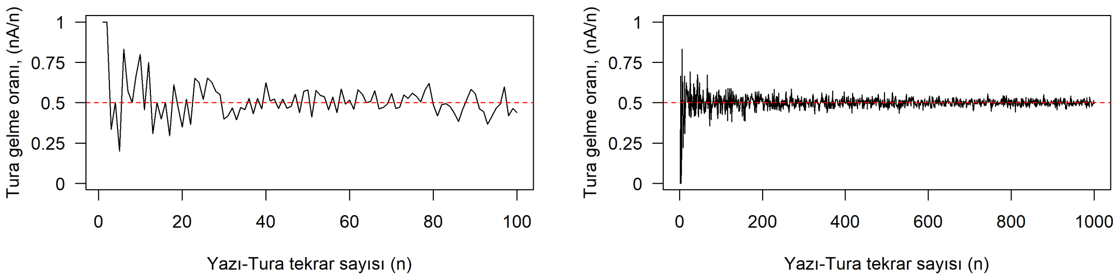
100 tekrar sonucunda Tura gelme "sıklığı" ya da "olasılığı" 0.44 oldu. Şekil 1.7a tekrar sayısı arttıkça Tura gelme oranının nasıl değiştiğini göstermektedir.

Şimdi tekrar sayısını daha da artıralım ve 1000 yapalım:

```
set.seed(155)
N <- 1000
P1000 <- numeric(N)
for (n in 1:N){
  x <- sample(c(1, 0), n, replace = TRUE)
  P1000[n] <- sum(x==1)/n
}
# 1 = Tura, 0 = Yazı
P1000[N]
```

```
[1] 0.496
```

Tura gelme sıklığı 0.496 oldu ve teorik olasılığa daha fazla yaklaştı. Her atışın sonucu bağımsız olduğundan, kısa vadede sonuçlar rastgele dağılabilir. Örneğin, ilk 10 atışta 7 kez tura gelebilir, 3 kez yazı gelebilir. Ancak uzun vadede bu sıklık, Şekil 1.7b'de görüldüğü gibi, teorik olasılık olan 0.5'e yaklaşacaktır.



(a) n=1,...,100 (b) n=1,...,1000

Şekil 1.7: Olasılığın sıklık tanımı: Tura gelme sıklığı

Eşit olasılık varsayımı altında, klasik olasılık ile göreli sıklık olasılığı aynı değere yakınsar. Eşit olasılık varsayımı, tüm sonuçların aynı olasılıkla gerçekleşeceğini öngördüğünden, bir deney çok sayıda tekrarlandığında gözlemsel olarak elde edilen göreli sıklık, teorik olarak klasik olasılık değeriyle örtüşür. Bu, özellikle bağımsız ve eşit olasılıklı olaylarda geçerlidir. Dolayısıyla, yazı-tura deneyinde paranın birçok kez atılması durumunda elde edilen yazı veya tura gelme sıklıkları, klasik olasılık değeri olan %50'ye yakınsar. Ancak eşit olasılık varsayımı geçerli değilse veya sonuçlar bağımlıysa, göreli sıklık ile klasik olasılık değerleri farklı olabilir.

Pay senedi piyasasında endeksin günlük kapanış değerinin yönünün olasılığını bulmaya çalışalım. BIST100 endeksinin 03/01/2003-29/09/2021 dönemi günlük kapanış değerleri kullanılarak yüzde değişimleri hesaplanmıştır. Günlük değişimlerin işaret + ise 1, − ise 0 değerini alan bir değişken oluşturulmuştur. `BIST100yon.rda` bu değişkeni içermektedir:

```
load("Data/BIST100yon.rda")
n <- length(BIST100yon)
n
```

```
[1] 4695
```

```
table(BIST100yon)/n
```

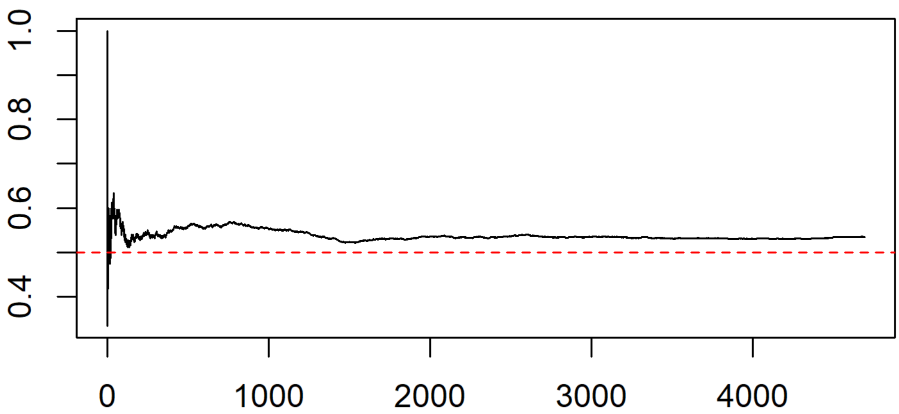
```
BIST100yon
        0         1 
0.4656017 0.5343983 
```

4695 günlük yön bilgisinden hareketle borsanın düşüşle kapandığı günlerin yaklaşık %47, artışla kapandığı günlerin ise %53 olduğunu görüyoruz.

```
artis_oran <- cumsum(BIST100yon)/(1:n)
# ilk 10 gün
options(digits=2)
artis_oran[1:10]
```

```
 [1] 1.00 0.50 0.33 0.50 0.60 0.50 0.57 0.50 0.56 0.60
```

```
# grafik
plot(artis_oran, type="l",
     xlab="",
     ylab="")
abline(h=0.5, lty="dashed", col="red")
```



Şekil 1.8: Artışla kapanan günlerin oranı, kümülatif

Şekil 1.8 dönem boyunca borsanın artışla kapandığı günlerin oranını göstermektedir. Borsanın yönünün ilk dönemler dışında pozitif olmasının olasılığı %50'nin biraz üzerindedir.

Buradan hareketle, borsanın yönünün yazı-tura atma deneyine benzediği sonucuna varamayız. Hilesiz bir para veya zar atma gibi eşit olasılıklı ve bağımsız olaylarda, sonuçları üreten süreci veya modeli biliriz. Bu tür deneylerde klasik olasılık yaklaşımı oldukça uygundur. Ancak, hisse senedi piyasasında günlük kapanış değerlerinin yönünü belirleyen bir model veya süreci bilmiyoruz; yalnızca gözlemlerimizi temel alarak bir tahmin yürütebiliyoruz. Bu nedenle, borsa gibi bağımlı ve değişken faktörlerden etkilenen durumlarda gözlemler yoluyla hesaplanan göreli sıklık yaklaşımı daha işlevsel hale gelir.

Genellikle, hisse senedi piyasasında günlük pozitif kapanışların oranı, negatif kapanışlara göre biraz daha yüksek gözlemlenir. Ancak, yazı-tura deneylerinde her bir olay birbirinden bağımsızken, borsada durum farklıdır. Bir günün kapanışı, haberler, ekonomik gelişmeler gibi birçok faktörden etkilenir ve bu etkiler birbirleriyle ilişkili olabilir. Borsanın yönünü tahmin etmek isteyen bir yatırımcı, bunu yazı-tura atar gibi yapamaz; hisse senedi ve finansal piyasalar çok sayıda faktörden etkilenir. Bu nedenle günlük değişimler geçmişteki gelişmelerden bağımsız tekrarlar gibi düşünülemez.

Bu da gösteriyor ki, yazı-tura gibi deneyler için en iyi tahmin %50 olabilirken, borsanın yönü için en iyi tahmin uzun dönem ortalaması olan %53 olmayabilir. Örneğin, son beş günlük dönemde hep artışla kapanan bir piyasada, ertesi gün artışla kapanma olasılığımız %50'den daha yüksek olabilir (veya tersi). Bu tür durumlarda bireysel bilgi, gözlem ve deneyimlerin de hesaba katıldığı öznel olasılık yaklaşımları daha uygun hale gelir; bu yaklaşımlar, yeni bilgiler ışığında olasılıkların güncellenmesine imkan tanır ve finansal piyasalardaki gibi karmaşık ve değişken sistemlerde karar vermeyi destekler.

1.2.3 Öznel Yaklaşım

Yazı-Tura ve borsanın yönü deneylerini tekrar düşündüğümüzde Tura gelme olasılığı ile borsanın günü artışla tamamlama olasılığının önemli bir farka işaret ettiğini görüyoruz. Tura gelme olasılığı, paranın hilesiz olduğu ve bağımsızlık varsayımları altında, herkes tarafından %50 olarak kabul edilir. Bu anlamda objektif/nesneldir. Borsanın yarın artışla kapanma olasılığı ise özneldir, kişiden kişiye değişir. Aslında yazı-tura deneyinde de olasılıkların öznel olarak tanımlanması mümkündür. Atılan paranın hileli olduğunu düşünüyorsak %50-%50 olasılık yerine %40-%60 olasılıkların daha uygun olduğu sonucuna varabiliriz.

Bayesçi ya da **sübjektif/öznel** yaklaşım, frekans yaklaşımına göre olasılık kavramını daha geniş bir perspektiften ele alır. Bu yaklaşım olasılığı "bir olayın gerçekleşmesine ilişkin inanç düzeyi" olarak tanımlar. Borsanın gelecekteki yönüne ya da bir futbol maçının sonucuna ilişkin olasılıklar kişiden kişiye değişebilir. Bir futbol maçı için belirlenen galibiyet olasılığı, farklı kişilerin geçmiş maç sonuçlarına ve takım performansına dair öznel inançlarına göre farklılık gösterebilir.

Bayesçi yaklaşım öznel olasılık kavramını bir adım öteye taşıyarak, gözlemler doğrultusunda bu inancın güncellenmesini sağlar. Bayes teoremi, bireyin bir olayın gerçekleşmesine ilişkin inancını yeni veri ile nasıl güncellemesi gerektiğini gösterir. Örneğin, bir yatırımcı borsanın artışla kapanma olasılığını başlangıçta %60 olarak değerlendirebilir, ancak gün içinde gelen olumlu ekonomik veriler doğrultusunda bu olasılığı %70'e çıkarabilir. Burada, Bayesçi yöntem bireyin inancını yeni bilgi doğrultusunda sistematik olarak değiştirme imkanı tanır.

Sıklık yaklaşımı çok sayıda tekrar edilen deneylerin sonuçlarına odaklanırken, öznel yaklaşım, tekrarı mümkün olmayan olaylara ilişkin bireysel inançlara dayanır. Önemli farklara sahip olsalar da hem sıklık hem de öznel olasılık yaklaşımları aynı matematiksel kurallara uyar.

1.3 Olasılık Önermeleri ve Kuralları

Bir deneyin olanaklı tüm sonuçlarını içeren örneklem uzayını S, bu kümedeki karşılıklı bağdaşmaz ve bütünü kapsayıcı temel sonuçları E_i, ve bu uzay içinde tanımlı herhangi bir olayı A ile gösterelim.

A'nın olasılığı $\mathbb{P}(A)$ olmak üzere olasılık önermeleri şunlardır:

Önerme 1.1 (Olasılıklar 0 ile 1 arasında değerler alır). *S içinde tanımlı bir A olayı*

$$0 \leq \mathbb{P}(A) \leq 1$$

koşulunu sağlamak zorundadır. Olasılıklar negatif ya da 1'den büyük olamaz. $\mathbb{P}(A) = 0$ ise bu olayın kesin olarak gerçekleşmeyeceği, $\mathbb{P}(A) = 1$ ise kesin olarak gerçekleşeceği anlamına gelir.

Önerme 1.2 (Bir olayın olasılığı). *S içinde tanımlı bir A olayının olasılığı*

$$\mathbb{P}(A) = \sum_{i \in A} \mathbb{P}(E_i) \tag{1.2}$$

ile bulunabilir. Bu toplam A olayının gerçekleşmesi ile sonuçlanan temel olaylar üzerinedir.

Önerme 1.3 (Olasılıkların toplamı). *S örneklem uzayının olasılığı*

$$\mathbb{P}(S) = \sum_{i \in S} \mathbb{P}(E_i) = 1$$

olur. Örneklem uzayında yer alan temel sonuçları olasılıklarının toplamı birdir.

Bu önermelerden hareketle olasılık hesaplamalarına ilişkin tüm kuralları türetebiliriz.

Kural 1: Örneklem uzayındaki her bir temel sonucun gerçekleşme olasılığı aynı ise ve örneklem uzayında toplam n temel sonuç varsa

$$\mathbb{P}(E_i) = \frac{1}{\text{Temel sonuç sayısı}} = \frac{1}{n}$$

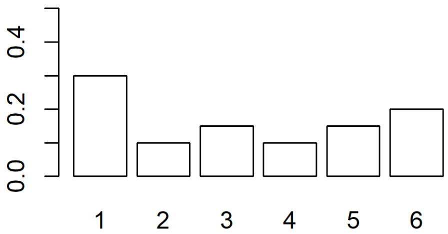
olur.

Örnek 1.11. Altı yüzü olan hilesiz bir zar atıldığında temel sonuçların her biri için olasılık 1/6 olur.

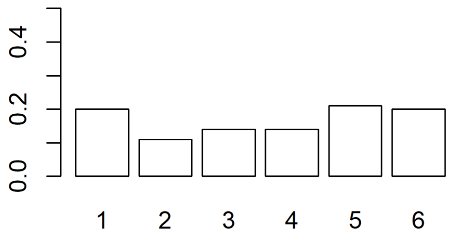
$$S = \{1, 2, 3, 4, 5, 6\}$$

$$\mathbb{P}(E_1) = ... = \mathbb{P}(E_6) = \frac{1}{6}$$

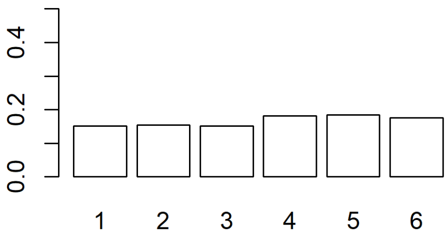
Şekil 1.9 zar atımı deneyinde temel olayların sıklıklarını göstermektedir. Tekrar sayısı n arttıkça temel olayların sıklıkları teorik olasılık değerine (1/6) yaklaşmaktadır.



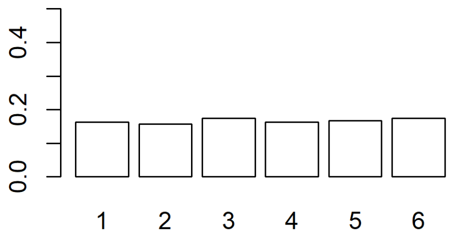
(a) n=20



(b) n=100



(c) n=1000



(d) n=10000

Şekil 1.9: Zar atımı deneyinde gerçekleşen sıklıklar

Kural 2: Eğer bir A olayı örneklem uzayında tanımlı n temel olaydan n_A kadarını içeriyorsa

$$\mathbb{P}(A) = \frac{n_A}{n}$$

olur.

Örnek 1.12. Hilesiz zar deneyinde A olayını zarın çift gelmesi olarak tanımlarsak bu olayın gerçekleşmesiyle sonuçlanan temel olaylar

$$A = \{2, 4, 6\}$$

olur. Buradan

$$\mathbb{P}(A) = \frac{3}{6} = \frac{1}{2}$$

bulunur. `R` programında bu deneyin 1000 tekrarını yapalım:

```
set.seed(101)
options(digits = 4)
x <- sample(c(1,2,3,4,5,6), 1000, replace=TRUE)
xcift <- x[x==2|x==4|x==6]
# A olayına yol açan temel olaylar
table(xcift)/1000
```

```
xcift
    2     4     6
0.159 0.177 0.166
```

```
sum(table(xcift)/1000)
```

```
[1] 0.502
```

```
#
A <- as.numeric((x==2) | (x==4) | (x==6))
table(factor(A, labels=c("Tek","Çift")))/1000
```

```
  Tek  Çift
0.498 0.502
```

```
#barplot(table(A)/1000, names.arg = c("Çift","Tek"), col="white")
```

Örnek 1.13. Hilesiz bir madeni para 3 kere atılıyor ve gelen sonuçlar kaydediliyor. En az 2 tura gelme olasılığını bulun ve `R` programında simülasyonunu yapın.

Çözüm:

Örneklem uzayını oluşturalım:

$$S = \{TTT,\ TTY,\ TYT,\ YTT,\ YYT,\ TYY,\ YTY,\ YYY\}$$

Bu temel olayların gerçekleşme şansı aynıdır: 1/8. Bunların arasında A olayına yol açan temel olaylar

$$A = \{TTT,\ TTY,\ TYT,\ YTT\}$$

olur. Öyleyse A olayının olasılığı

$$\mathbb{P}(A) = \frac{n_A}{n} = \frac{4}{8} = 0.5$$

olur.

```
set.seed(444)
N <- 1000
E <- matrix(NaN, nrow = N, ncol = 8)
colnames(E) <- c("TTT", "TTY", "TYT", "YTT", "YYT", "TYY", "YTY", "YYY")
para1 <- sample(c(1, 0), N, replace = TRUE)  # 1.para, T=1, Y=0
para2 <- sample(c(1, 0), N, replace = TRUE)  # 2.para, T=1, Y=0
para3 <- sample(c(1, 0), N, replace = TRUE)  # 2.para, T=1, Y=0
# temel olaylar
E[,1] <- (para1==1 & para2==1 & para3==1)  #
E[,2] <- (para1==1 & para2==1 & para3==0)
E[,3] <- (para1==1 & para2==0 & para3==1)
E[,4] <- (para1==0 & para2==1 & para3==1)
E[,5] <- (para1==0 & para2==0 & para3==1)
E[,6] <- (para1==1 & para2==0 & para3==0)
E[,7] <- (para1==0 & para2==1 & para3==0)
E[,8] <- (para1==0 & para2==0 & para3==0)
```

Burada $N \times 8$ boyutlu E matrisi temel olaylardan hangisinin gerçekleştiğini gösteren bir indikatör matrisidir. Örneğin ilk 10 denemenin sonuçları aşağıdaki gibidir:

```
E[1:10,]
```

```
     TTT TTY TYT YTT YYT TYY YTY YYY
[1,]   0   0   0   0   0   0   1   0
[2,]   0   0   0   0   0   1   0   0
[3,]   0   0   1   0   0   0   0   0
[4,]   0   0   1   0   0   0   0   0
[5,]   1   0   0   0   0   0   0   0
```

```
 [6,]    0    0    0    0    0    0    1    0
 [7,]    0    0    0    0    0    0    0    1
 [8,]    0    0    0    0    0    0    0    1
 [9,]    1    0    0    0    0    0    0    0
[10,]    0    0    0    0    0    1    0    0
```

Satırlar denemeleri, sütunlar ise temel olayları göstermektedir. Hücrenin 0 olması o denemede sütunda belirtilen temel olayın gerçekleşmediği, 1 olması ise gerçekleştiği anlamına gelmektedir. Örneğin ilk denemede (tekrarda) YTY olayının gerçekleştiğini görüyoruz. Her satırda sütunlardan sadece birinin 1 değerini alması gerekir, yani temel olaylar aynı anda gerçekleşemez (karşılıklı bağdaşmaz). Bunun doğrudan sonucu E indikatör matrisinin satır toplamlarının her bir satır için 1 olması ve bunların toplamının N olmasıdır. Örneğimizde

```
sum(rowSums(E))
```

```
[1] 1000
```

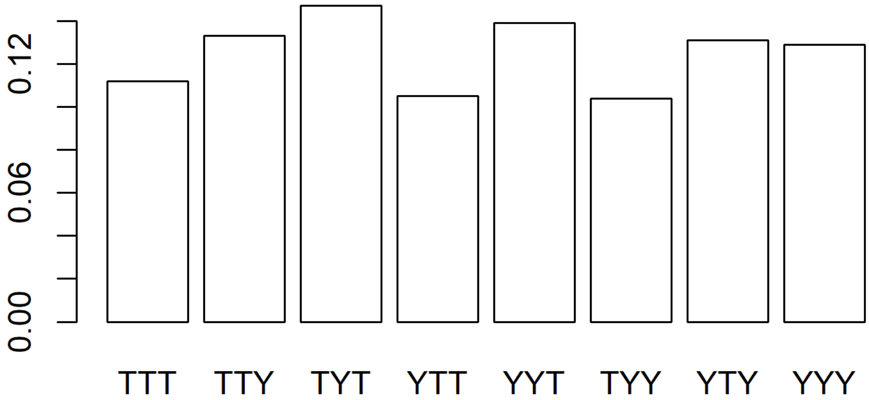
olduğu kolayca görülür. Bu matrisin sütun ortalamaları temel olayların sıklıklarını gösterir:

```
# sütun ortalamaları sıklıkları verir
colMeans(E)
```

```
  TTT   TTY   TYT   YTT   YYT   TYY   YTY   YYY 
0.112 0.133 0.147 0.105 0.139 0.104 0.131 0.129 
```

Şekil 1.10 temel olayların sıklıklarının çubuk grafiğini göstermektedir.

```
barplot(colMeans(E), col="white")
```



Şekil 1.10: İki para deneyinde temel olayların sıklıkları

A olayının olasılığı A kümesinde yer alan temel olayların olasılıklarının toplamıdır. Yukarıdaki E matrisinden de görüleceği gibi birinci ve ikinci tekrarlarda A olayı gerçekleşmemişken, üçüncü, dördüncü ve beşinci tekrarlarda gerçekleşmiştir. Bunu ilgili sütunların toplamından hareketle bir indikatör değişkeni yardımıyla hesaplayabiliriz.

```
# A = en az 2 Tura olasılığı
# A = 1, olay gerçekleşti,
#   = 0, olay gerçekleşmedi
A <- E[,1]+E[,2]+E[,3]+E[,4]
sum(A)/N
```

```
[1] 0.497
```

1000 tekrarlı bu simülasyonda 3 para atımı deneyinde en az iki Tura gelme sıklığı 0.497 olmuştur. Tekrar sayısı arttırılarak teorik olasılığa daha fazla yaklaşılır.

Kural 3 (Tümleyen kuralı): A olayının gerçekleşmemesini A kümesinin tümleyeni olarak tanımlarsak:

$$\mathbb{P}(A) + \mathbb{P}(\bar{A}) = \mathbb{P}(S) = 1 \tag{1.3}$$

Buradan

$$\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A) \tag{1.4}$$

olur. Boş kümenin olasılığı her zaman sıfırdır:

$$\mathbb{P}(\emptyset) = \mathbb{P}(\bar{S}) = 0$$

Kural 4 (Toplama Kuralı): A ve B ayrık olaylar değilse, yani **birlikte gerçekleşmeleri mümkünse**, olayların birleşiminin olasılığı

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \tag{1.5}$$

olur. Kesişim olasılığı, $\mathbb{P}(A \cap B)$, bu olayların birlikte gerçekleşme olasılıklarıdır.

Örnek 1.14. Bir sınıfta 30 öğrenci vardır. Öğrencilerden 12'si matematik, 18'i ise fizik dersine kayıtlıdır. 6 öğrenci her iki derse de kayıtlıdır. Rasgele seçilen bir öğrencinin matematik veya fizik dersine kayıtlı olma olasılığı nedir? Sadece matematik veya sadece fizik dersine kayıtlı olma olasılığı nedir?

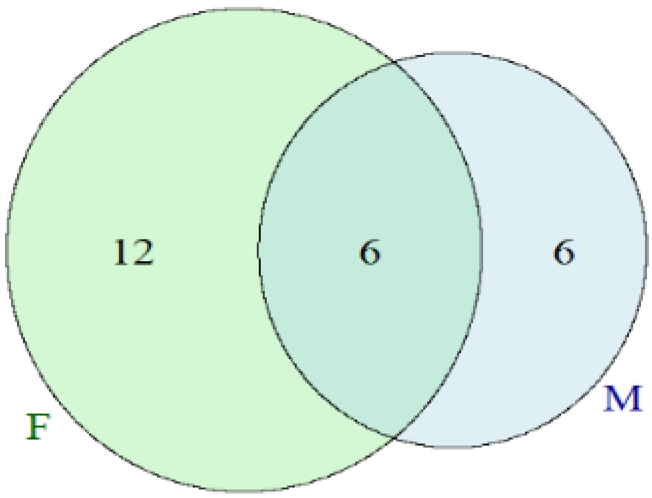
Çözüm:

- M: Matematik dersini alması, $\mathbb{P}(M) = 12/30 = 0.4$
- F: Fizik dersini alması, $\mathbb{P}(F) = 18/30 = 0.6$
- $M \cap F$ İkisin birden alması, $\mathbb{P}(M \cap F) = 6/30 = 0.2$

Toplama kuralından hareketle rasgele seçilen bir öğrencinin matematik veya fizik dersine kayıtlı olma olasılığı

$$\mathbb{P}(M \cup F) = \mathbb{P}(M) + \mathbb{P}(F) - \mathbb{P}(M \cap F) = 0.4 + 0.6 - 0.2 = 0.8$$

%80 olur (bkz. Şekil 1.11).



Şekil 1.11: Matematik (M) ve Fizik (F) ders tercihlerinin Venn diyagramı

Şekil 1.11 ders tercihlerinin Venn diyagramını göstermektedir. Sadece matematik dersini alanların sayısı 12-6=6'dır. Öyleyse sadece matematik dersini alma olasılığı $\mathbb{P}(\text{sadece } M) =$

$6/30 = 0.2$. Sadece fizik dersini alanların sayısı ise 18-6=12'dir, yani $\mathbb{P}(\text{sadece } F) = 12/30 = 0.4$. Buradan

$$\mathbb{P}(\text{sadece } M \cup \text{sadece } F) = \mathbb{P}(\text{sadece } M) + \mathbb{P}(\text{sadece } F) = 0.2 + 0.4 = 0.6$$

bulunur. $\mathbb{P}(\text{sadece } M \cap \text{sadece } F) = \mathbb{P}(\emptyset) = 0$ olduğuna dikkat ediniz.

A ve B karşılıklı bağdaşmaz (ayrık) iki olay ise bu olayların birleşiminin olasılığı

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) \tag{1.6}$$

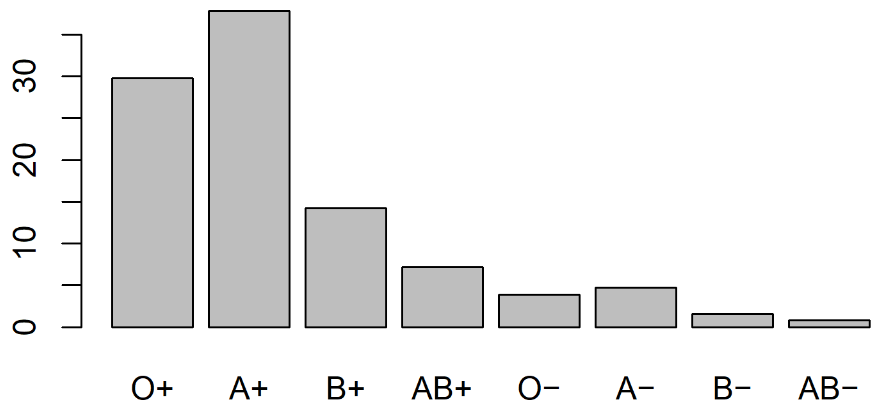
olur. Bu, toplama kuralının $\mathbb{P}(A \cap B) = 0$ olduğu özel bir durumudur.

Örnek 1.15. Türkiye'e kan grubu sıklıkları yüzde olarak aşağıdaki tabloda verilmiştir. Şekil 1.12 bu oranların (%) çubuk çizimini göstermektedir.

```
# Türkiye'de gözlemlenen kan grubu sıklıkları
# kaynak: https://en.wikipedia.org/wiki/Blood_type_distribution_by_country
kan_grubu <- c(29.8, 37.8, 14.2, 7.2, 3.9, 4.7, 1.6,0.8)
names(kan_grubu) <- c("O+", "A+", "B+", "AB+", "O-", "A-", "B-", "AB-")
print(kan_grubu)
```

```
  O+   A+   B+  AB+   O-   A-   B-  AB- 
29.8 37.8 14.2  7.2  3.9  4.7  1.6  0.8 
```

```
barplot(kan_grubu)
```



Şekil 1.12: Türkiye'de gözlemlenen kan grubu sıklıkları, %

Buna göre rassal seçilmiş bir bireyin kan grubunun O+ ya da A+ olma olasılığı kaçtır?

Çözüm:

Bu olaylar karşılıklı bağdaşmaz olduğu için (bir kişinin sadece bir kan grubu olabilir) istenen olasılık

$$\mathbb{P}(\mathrm{O}^+ \cup \ \mathrm{A}^+) = \%29.8 + \%37.8 = \%67.6$$

olur.

Toplama kuralı genelleştirilebilir. A, B ve C üç olay ise bu olayların birleşiminin olasılığı aşağıdaki gibi yazılabilir:

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C) \tag{1.7}$$

Tanım 1.1 (Boole eşitsizliği)**.** $A_1, \dots, A_k$ bir olaylar dizisi ise bu olayların birleşiminin, $\bigcup_{i=1}^{k} A_i$ olasılığı

$$\mathbb{P}\left(\bigcup_{i=1}^{k} A_i\right) \leq \sum_{i=1}^{k} \mathbb{P}(A_i) \tag{1.8}$$

olasılıklar toplamından küçük ya da eşit olur. $A_1, \dots, A_k$ olayları karşılıklı bağdaşmaz ise

$$\mathbb{P}\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} \mathbb{P}(A_i) \tag{1.9}$$

eşitlik sağlanır. Aksi durumda

$$\mathbb{P}\left(\bigcup_{i=1}^{k} A_i\right) < \sum_{i=1}^{k} \mathbb{P}(A_i) \tag{1.10}$$

olur.

Örnek 1.16. Bir anket çalışmasında 100 kişi, favori içecekleri olarak çay, kahve ve meyve suyu tercihlerini belirtmişlerdir. Anket sonuçları aşağıdaki gibidir:

- 55 kişi çayı tercih ediyor.
- 45 kişi kahveyi tercih ediyor.
- 32 kişi meyve suyunu tercih ediyor.
- 18 kişi hem çayı hem de kahveyi tercih ediyor.
- 12 kişi hem çayı hem de meyve suyunu tercih ediyor.
- 8 kişi hem kahveyi hem de meyve suyunu tercih ediyor.
- 6 kişi her üç içeceği de tercih ediyor.

Buna göre aşağıdaki soruları cevaplayınız:

a) Bir kişinin çay ya da kahveyi tercih etme olasılığı nedir?
b) Bir kişinin kahve ya da meyve suyu tercih etme olasılığı nedir?

c) Bir kişinin çay, kahve, ya da meyve suyundan en az birini tercih etme olasılığı nedir?

Çözüm:

Olayları tanımlayalım:

- : “çay tercih ediyor.”
- K: “Kahve tercih ediyor.”
- M: “Meyve suyu tercih ediyor”

Olasılıklar:

$$\mathbb{P}() = 0.55, \quad \mathbb{P}(K) = 0.45, \quad \mathbb{P}(M) = 0.32$$

$$\mathbb{P}(\cap K) = 0.18, \quad \mathbb{P}(\cap M) = 0.12, \quad \mathbb{P}(K \cap M) = 0.08$$

$$\mathbb{P}(\cap K \cap M) = 0.06$$

Tercih yapan 100 kişi vardı. Bunlardan

- Sadece çay tercih edenler: $55 - 18 - 12 + 6 = 31$
- Sadece kahve tercih edenler: $45 - 18 - 8 + 6 = 25$
- Sadece meyve suyu tercih edenler: $32 - 12 - 8 + 6 = 18$
- Çay ve kahve (meyve suyu hariç): $18 - 6 = 12$
- Çay ve meyve suyu (kahve hariç): $12 - 6 = 6$
- Kahve ve meyve suyu (çay hariç): $8 - 6 = 2$
- Üçünü birden tercih eden: 6

kişidir (bkz. Şekil 1.13).

Bir kişinin çay ya da kahveyi tercih etme olasılığı

$$\mathbb{P}(\cup K) = \mathbb{P}() + \mathbb{P}(K) - \mathbb{P}(\cap K) = 0.55 + 0.45 - 0.18 = 0.82$$

%82’dir.

Bir kişinin kahve ya da meyve suyu tercih etme olasılığı

$$\mathbb{P}(K \cup M) = \mathbb{P}(K) + \mathbb{P}(M) - \mathbb{P}(K \cap M) = 0.45 + 0.32 - 0.08 = 0.69$$

%69’dur.

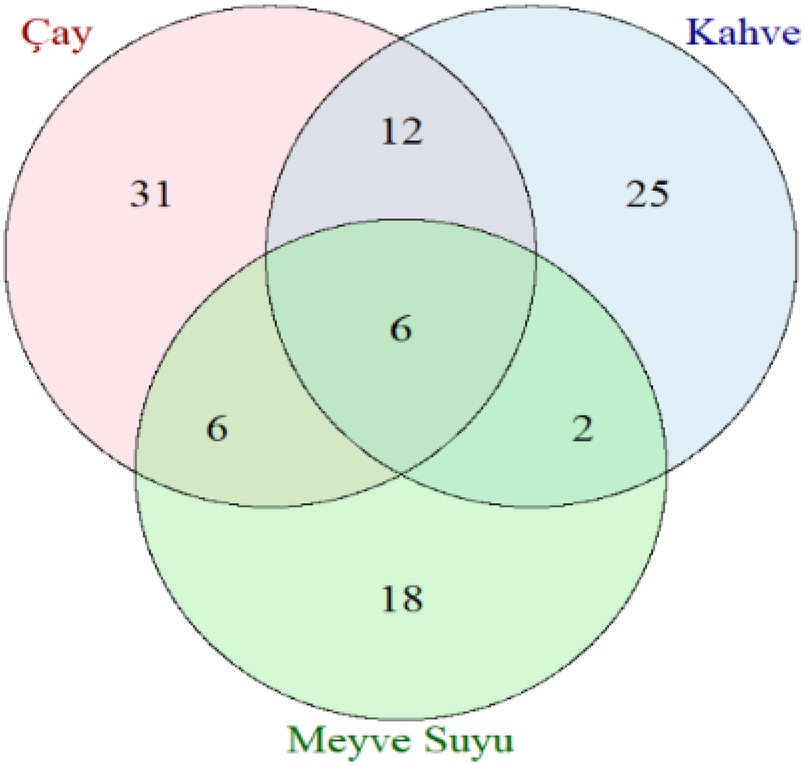
Bir kişinin çay, kahve, ya da meyve suyundan en az birini tercih etme olasılığı

$$\mathbb{P}(\cup K \cup M) = \mathbb{P}() + \mathbb{P}(K) + \mathbb{P}(M) - \mathbb{P}(\cap K) - \mathbb{P}(\cap M) - \mathbb{P}(K \cap M) + \mathbb{P}(\cap K \cap M)$$

ve buradan

$$\mathbb{P}(\cup K \cup M) = 0.55 + 0.45 + 0.32 - 0.18 - 0.12 - 0.08 + 0.06 = 1$$

olur. Bu herkesin en az bir içecek tercihinde bulunduğunu gösterir.



Şekil 1.13: İçecek tercihlerinin Venn diyagramı (kişi sayıları gösterilmiştir)

1.4 Sayma Kuralları ve Olasılık

Klasik olasılıkta örneklem uzayındaki nesnelerin ve bunların arasında ilgilendiğimiz olayın gerçekleşmesiyle sonuçlananların sayısının olasılık hesaplamaları için önemli olduğunu gördük. Çoğu durumda bu nesnelerin açıkça listelenmesi zordur ya da mümkün değildir. Listelemek yerine belirli kuralları takip ederek bu nesneleri sayabiliriz.

Sayma işlemlerinde dikkat etmemiz gereken ilk nokta nesnelerin yerine konarak mı yoksa yerine konmadan mı seçildiğidir. Yerine koymadan örneklemede nesneler teker teker seçilir ve tekrar yerine konmaz. Diğer durumda ise seçilen nesneler yerine konur ve tekrar seçilme şansı olur. Örneğin `sample()` fonksiyonu ile 1-10 arası rakamdan 5'ini yerine koyarak ve koymadan seçelim:

```r
set.seed(3)
sample(1:10, 5, replace = TRUE)
```

```
[1]  5 10  7  4 10
```

```r
sample(1:10, 5, replace = FALSE)
```

```
[1]  8 10  4  2  5
```

`replace = TRUE` ile yerine koyarak rassal seçim yaptık. Bu durumda bazı nesnelerin tekrar seçildiğini görüyoruz.

Tanım 1.2 (Çarpım kuralı)**.** Eğer bir işlemi ya da eylemi yapmanın n_1 yolu varsa, başka bir işlemi yapmanın da n_2 yolu varsa bu işlemleri birlikte yapmanın $n_1 \times n_2$ yolu vardır. Bu kural ikiden fazla eylemler için genelleştirilebilir.

Eğer her tekrarda olası sonuçların sayısı n ise deney r kere tekrar edildiğinde ortaya çıkan tüm sonuçların sayısı n^r olur. Bu kuralı 3 para atımı deneyinde örtük olarak kullanmıştık. Her para atımı tekrarının 2 sonucu vardır ve 3 kez tekrar edilirse olanaklı tüm sonuçların sayısı $2^3 = 8$ olur.

Tanım 1.3 (Permütasyon)**.** Yerine koymadan seçim yaptığımızda ve sıranın önemli olduğu durumlarda n nesnenin kaç farklı şekilde düzenlenebileceğini bulmak için faktöriyel işlemcisini

$$n! = n \times (n-1) \times (n-2) \times ... \times 2 \times 1$$

kullanabiliriz. Faktöriyel (!) işlemcisi ayırdedilebilir nesnelerin kaç farklı şekilde sıralanabileceğini gösterir. Örneğin $\{a, b, c, d, e\}$ nesnelerini sıralamanın $5! = 120$

```r
factorial(5)
```

```
[1] 120
```

yolu vardır. Hiç bir nesne içermeyen bir kümeyi sıralamanın ise sadece 1 yolu vardır. Bu nedenle $0! = 1$ olarak tanımlıdır.

Bu işlemi boşluk doldurmanın farklı yolları olarak düşünebiliriz. Aşağıdaki gibi 5 tane boşluk ya da kutu varsa, elimizdeki 5 nesneyi bu kutulara kaç farklı şekilde yerleştirebiliriz?

$$- - - - -$$

Birinci kutu ya da boşluğu doldurmanın 5 yolu vardır:

$$5 - - - -$$

İlk seçim yapıldıktan sonra geriye 4 nesne kalır, öyleyse ikinci kutuyu doldurmanın 4 yolu vardır:

$$5\ 4 - - -$$

Benzer şekilde ilerlenirse faktöriyel tanımına ulaşılır ve çarpım kuralı uygulanarak ortaya çıkacak sonuç sayısı

$$5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 \equiv 5! = 120$$

olur. Eğer yerine koyarak seçim yapsaydık bu durumda toplam nesne sayısı her boşluk doldurma eyleminde aynı olduğu için çarpım kuralından hareketle

$$5 \cdot 5 \cdot 5 \cdot 5 \cdot 5 = 5^5 = 3125$$

olur.

Eğer n nesne arasından yerine koymadan r tanesini seçmek istiyorsak ve sıra önemliyse **permütasyon** tanımına ulaşırız:

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

Eğer n nesne arasından n nesneyi yerine koymadan seçiyorsak ve bu nesnelerin sıralaması önemliyse, permütasyonun özel bir durumu olan faktoriyel tanımını elde ederiz:

$$P_n^n = \frac{n!}{(n-n)!} = \frac{n!}{0!} = n!$$

Örnek 1.17. $\{a, b, c, d, e\}$ harfleri arasından ikisini sıralı olarak seçmenin kaç farklı yolu vardır?

Çözüm:

Sıra önemli olduğuna göre (a, b) ve (b, a) nesneleri ayrı ayrı sayılmalıdır. Bu şekilde oluşturulabilecek 5'in 2'li permütasyonları

$$P_2^5 = \frac{5!}{(5-2)!} = \frac{5 \cdot 4 \cdot 3!}{3!} = 20$$

olur.

Örnek 1.18. 5 farklı renkten (kırmızı, mavi, yeşil, sarı, beyaz) üç renk seçip, kırmızı ve mavi renklerin birlikte olmadığı sıralı kaç nesne oluşturulabilir?

Çözüm: 5 renk arasından 3'ünü sıralı seçmenin

$$P_3^5 = \frac{5!}{(5-3)!} = 60$$

farklı yolu vardır. Bunlardan kırmızı ve mavinin olduğu nesnelerin çıkarılması gerekir. Kırmızı ve mavi içeren permütasyonlarda üçüncü rengi seçmenin

$$P_1^3 = \frac{3!}{(3-1)!} = 3$$

yolu vardır. Bu 3 rengin kendi içinde sıralanma sayısı

$$P_3^3 = 3! = 6$$

olur. Öyleyse kırmızı ve mavi içeren permütasyonların toplam sayısı

$$3 \times 6 = 18$$

olur. Buradan kırmızı ve mavi içermeyen permütasyonların sayısı

$$60 - 18 = 42$$

olarak bulunur.

Tanım 1.4 (Kombinasyon)**.** Permütasyonda sıra önemliydi. Eğer yerine koymadan seçim yapıyorsak ve sıralama önemli değilse n nesnenin r kombinasyonu

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

ile bulunur.

Örneğin $\{a, b, c, d, e\}$ arasından $r = 2$ harfin sırası önemsiz olarak seçilme sayısı

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

olur.

`R`'da bu hesaplamayı yapmak için

```r
choose(5,2)
```

```
[1] 10
```

kullanılabilir. Ayrıca `combn()` fonksiyonu ile bu nesneleri listeleyebiliriz:

```r
combn(c("a","b","c","d","e"), 2)
```

```
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
[1,] "a"  "a"  "a"  "a"  "b"  "b"  "b"  "c"  "c"  "d"
[2,] "b"  "c"  "d"  "e"  "c"  "d"  "e"  "d"  "e"  "e"
```

Permütasyonda (ab) ile (ba) farklı sonuç olarak kabul ediliyordu. Kombinasyonda ise sıra önemli olmadığından aynı olarak kabul edilir. Bu nedenle 5 nesnenin 2'li permütasyonu 20 iken, kombinasyonu 10 olur.

Örnek 1.19. Bir şirket, yönetim kuruluna 3 yeni üye seçecektir. Başvuranlar arasında 6 erkek ve 4 kadın aday bulunmaktadır.

a) Hiç kadın seçilmeme olasılığını bulunuz.
b) En az bir kadın seçilme olasılığını bulunuz.
c) Tam olarak 2 kadın seçilme olasılığını bulunuz.

Çözüm:

Öncelikle toplam kaç farklı şekilde 3 kişi seçilebileceğini bulalım:

$$\text{Toplam seçim sayısı} = \binom{10}{3} = \frac{10!}{3!(10-3)!} = \frac{10 \cdot 9 \cdot 8}{3 \cdot 2 \cdot 1} = 120$$

Bu 120 farklı kombinasyonun herbirinin seçilme şansının eşit olduğunu varsayıyoruz.

a) Hiç kadın seçilmemesi için 3 kişinin tamamı erkekler arasından seçilmelidir. 6 erkek arasından 3'ünü seçmenin farklı yolları

$$\binom{6}{3} = \frac{6!}{3!(6-3)!} = 20$$

Böylece

$$\mathbb{P}(\text{hiç kadın yok}) = \frac{\binom{6}{3}}{\binom{10}{3}} = \frac{20}{120} = \frac{1}{6}$$

bulunur.

b) En az bir kadın seçilme olasılığı, hiç kadın seçilmeme olasılığının tümleyenidir:

$$\mathbb{P}(\text{en az bir kadın}) = 1 - \mathbb{P}(\text{hiç kadın yok}) = 1 - \frac{1}{6} = \frac{5}{6}$$

c) Tam olarak 2 kadın seçilme olasılığı:

4 kadın arasından 2 kişiyi seçmenin

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

ve 6 erkek arasından 1 kişiyi seçmenin

$$\binom{6}{1} = \frac{6!}{1!(6-1)!} = 6$$

farklı yolu vardır. Öyleyse, en az iki kadının olduğu seçimlerin sayısı $6 \cdot 6 = 36$ olur. Buradan

$$\mathbb{P}(2 \text{ kadın}) = \frac{\binom{4}{2} \cdot \binom{6}{1}}{\binom{10}{3}} = \frac{6 \cdot 6}{120} = \frac{36}{120} = \frac{3}{10}$$

bulunur.

Doğrulama yapmak için olanaklı tüm kombinasyonların sayısını bulabiliriz:

- Hiç kadın yok: $\binom{6}{3} = 20$ durum
- 1 kadın: $\binom{4}{1} \cdot \binom{6}{2} = 4 \cdot 15 = 60$ durum
- 2 kadın: $\binom{4}{2} \cdot \binom{6}{1} = 6 \cdot 6 = 36$ durum
- 3 kadın: $\binom{4}{3} \cdot \binom{6}{0} = 4 \cdot 1 = 4$ durum

Toplam: $20 + 60 + 36 + 4 = 120$ durum ($\binom{10}{3}$'e eşit)

Örnek 1.20. Hilesiz bir madeni para 10 kere atılıyor. 4 Tura gelme olasılığı kaçtır?

Çözüm:

Bir madeni para atıldığında her birinin olasılığı $p = 0.5$ olan iki sonuçtan biri gerçekleşir, yazı (Y) ya da tura (T). Bu şekilde p ve $1-p$ olasılıklarla iki sonucu olan denemelere **Bernoulli denemeleri** denir. Bu örneğimizde Bernoulli denemesini bağımsız şekilde $n = 10$ kere tekrarlıyoruz ve $k = 4$ sonucun Tura olma olasılığını hesaplamak istiyoruz.

Öncelikle 4 Tura gelmesinin kaç farklı biçimde oluşabileceğini düşünelim:

$$\frac{}{1}, \frac{}{2}, \frac{}{3}, \frac{}{4}, \frac{}{5}, \frac{}{6}, \frac{}{7}, \frac{}{8}, \frac{}{9}, \frac{}{10}$$

Örneğin bunlardan biri aşağıdaki gibi olabilir:

$$\frac{Y}{1}, \frac{T}{2}, \frac{T}{3}, \frac{Y}{4}, \frac{Y}{5}, \frac{T}{6}, \frac{Y}{7}, \frac{T}{8}, \frac{Y}{9}, \frac{Y}{10}$$

Bu dizilimin gerçekleşme olasılığı kaçtır?

Sonuçlar birbirinden bağımsız olduğu için

$$\frac{1}{2} \times \frac{1}{2} \times ... \times \frac{1}{2} = \left(\frac{1}{2}\right)^{10} = 0.0009765625$$

buluruz. Sıra önemsiz olduğu için 4 Tura ve 6 Yazı ile sonuçlanan başka bir dizilimin olasılığı da aynı olur. Bizim istediğimiz olasılık tam olarak 4 Tura gelme olasılığıdır. Bunun için 4 Tura ve 6 Yazı gelmesinin **olanaklı kaç dizilimi** olduğunu bulmamız gerekir. Karşılıklı bağdaşmaz bu dizilimlerin sayısı 10 denemenin 4'lü kombinasyonu ile hesaplanabilir:

$$\binom{10}{4} = \frac{10!}{4!(10-4)!} = 210$$

```
choose(10, 4)
```

```
[1] 210
```

Öyleyse her birinin gerçekleşme şansı 0.0009765625 olan 210 farklı Yazı-Tura dizilimi gözlemlenebilir. Tura sayısının tam olarak 4 olma olasılığı tüm bu olasılıkların toplamıdır. Ya da dizilim sayısı ile tekil bir dizilimin olasılığını çarparsak

$$\frac{10!}{4!(10-4)!} \times \left(\frac{1}{2}\right)^{10} = 0.2051$$

buluruz.

Örnek 1.21. Önceki örnekte para hilesiz kabul edildiği için T ve Y gelme olasılıkları 0.5 olarak alınmıştı. Şimdi paranın yanlı olduğunu ve Tura gelme olasılığının $p = 0.6$, Yazı gelme olasılığının ise $1 - p = 0.4$ olduğunu düşünelim. Bu durumda 10 denemede 4 Tura gelme olasılığı kaçtır?

Çözüm:

Parayı bir kez atarsak T gelme şansı 0.6 olur. İkinci paranın da T gelme olasılığı ise çarpım kuralıyla $0.6 \times 0.6 = 0.36$ olur. Birinci paranın T ikinci paranın Y olma olasılığı ise $0.6 \times 0.4 = 0.24$ olur.

Benzer şekilde önceki örnekte verdiğimiz $Y, T, T, Y, Y, T, Y, T, Y, Y$ dizilimin gerçekleşme olasılığı:

$$0.4 \times 0.6 \times 0.6 \times 0.4 \times 0.4 \times 0.6 \times 0.4 \times 0.6 \times 0.4 \times 0.4 = 0.6^4 \times 0.4^6 = 0.0005308416$$

olur. 4 T ve 6 Y içeren olanaklı tüm dizilimlerin olasılığı aynıdır. Öyleyse toplam olasılık

$$\frac{10!}{4!(10-4)!} \times 0.6^4 \times 0.4^6 = 0.111$$

yaklaşık 0.111 olur. 10 denemede gelen Tura sayısını tüm olanaklı $0, 1, \ldots, 10$ değerler için hesaplayıp grafiğini çizebiliriz:

```
x <- 0:10
p <- numeric(11)
for (i in 0:10){
  p[i+1] <-  choose(10, i)*(0.6)^i*(0.4)^(10-i)
}
names(p) <- x
p
```

```
        0         1         2         3         4         5         6         7 
0.0001049 0.0015729 0.0106168 0.0424673 0.1114767 0.2006581 0.2508227 0.2149908 
        8         9        10 
0.1209324 0.0403108 0.0060466 
```

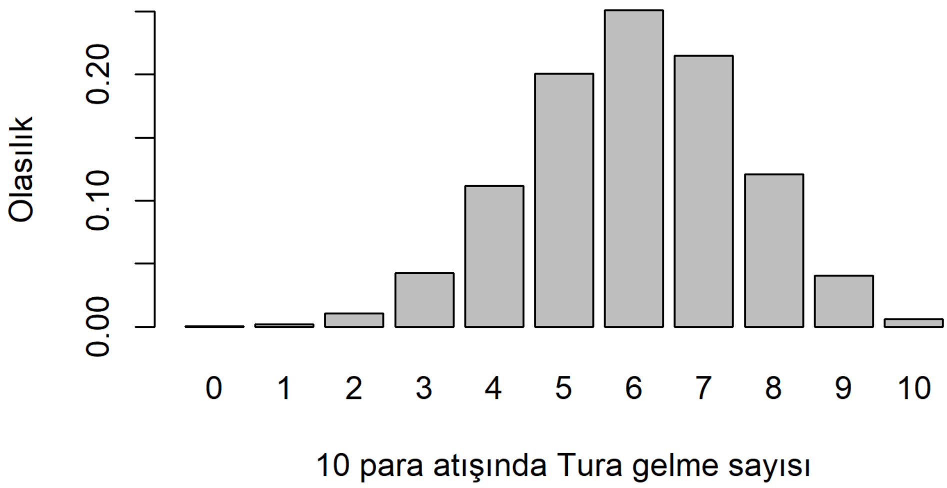
Olasılıklar toplamının

```
sum(p)
```

```
[1] 1
```

olduğuna dikkat ediniz. Şekil 1.14 bu olasılıkların dağılımını göstermektedir.

```
barplot(p,
        xlab = "10 para atışında Tura gelme sayısı",
        ylab = "Olasılık")
```



Şekil 1.14: p=0.6 için 10 denemede x Tura gelme sayısının olasılıkları

Önceki örnekte n bağımsız Bernoulli denemesinde x Tura olasılığı **Binom dağılımı**

$$\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$$

ile betimlenebilir. Bu dağılımın özelliklerini daha sonra inceleyeceğiz.

	Yerine koyarak	**Yerine koymadan**
Sıralı	$P(n,k) = n^k$	$P(n,k) = \frac{n!}{(n-k)!}$
	Yerine koyarak permütasyon	Yerine koymadan permütasyon
Sırasız	$C(n+k-1,k) = \binom{n+k-1}{k}$	$C(n,k) = \binom{n}{k}$
	Yerine koyarak kombinasyon	Yerine koymadan kombinasyon

Örnek 1.22 (Doğumgünü problemi)**.** Bir odada n kişi vardır. Bu kişilerin doğumgünlerinin gün/ay olarak açıklandığını düşünelim. Bu durumda en az bir eşleşme olasılığı kaçtır? Bir yılda 365 gün olduğunu varsayınız (artık yılları göz ardı ediyoruz). Ayrıca, yine basitlik amacıyla, doğumların tüm yıla eşit dağıldığını varsayalım.

Çözüm

Olayları tanımlayalım:

- A: "en az bir kişinin aynı doğumgününe sahip olması",
- $\bar{A}$: "hiçbir çiftin aynı doğumgününe sahip olmaması".

Bu iki olay birbirinin tümleyeni olduğuna göre istenen olasılık

$$\mathbb{P}(A) = 1 - \mathbb{P}(\bar{A})$$

ile bulunabilir.

Odadaki her birey için 365 olanaklı doğumgünü vardır. Öyleyse tüm bireyler için örneklem uzayındaki nesne sayısı 365^n olur. Buradan

$$\mathbb{P}(A) = 1 - \mathbb{P}(\bar{A}) = 1 - \frac{\#(\bar{A})}{\#(S)}$$

yazılabilir. Burada $\#(\bar{A})$, $\bar{A}$ olayının gerçekleşme sayısıdır.

Hiç eşleşmenin olmadığı durumda, birinci birey için 365 olanaklı gün, ikinci birey için 364 olanaklı gün, üçüncü birey için 363 olanaklı gün, vb. olur. Birey n için ise $365 - n + 1$ olanaklı doğumgünü tanımlanabilir. Öyleyse, toplam gerçekleşme sayısı

$$\#(\bar{A}) = 365 \times 364 \times 363 \times ... \times (365 - n + 1)$$

olur. Buradan

$$\mathbb{P}(A) = 1 - \frac{\#(\bar{A})}{\#(S)} = 1 - \frac{365 \times 364 \times 363 \times ... \times (365 - n + 1)}{365^n}$$

Bu olasılık odadaki birey sayısına göre değişecektir.

Herhangi bir n için bu olasılığı hesaplayan bir `R` fonksiyonu yazabiliriz:

```r
dogumgunu_olasiligi <- function(n) {
  # P(A^c): hiç eşleşmenin olmaması
  P_Ac <- prod((365 - (0:(n - 1))) / 365)
  # P(A): en az iki kişinin aynı doğumgününe sahip olması
  P_A <- 1 - P_Ac
  return(P_A)
}
```

```r
# 10 kişi içinde aynı doğumgünü olasılığı kaçtır?
dogumgunu_olasiligi(10)
```

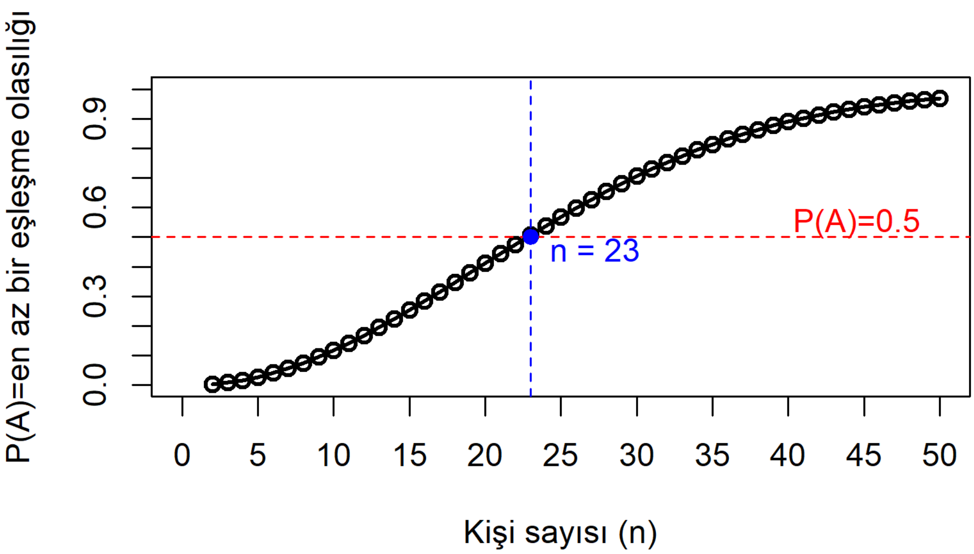
```
[1] 0.1169
```

10 kişilik bir grupta en az bir doğumgünü eşleşmesinin olma olasılığı yaklaşık olarak %12'dir. Seçilmiş bazı değerler için bu olasılıkları hesaplayalım:

```r
dgunu <- rbind(kisi_sayisi = c(10,20,23,40,50,60),
               olasilik = c(dogumgunu_olasiligi(10), dogumgunu_olasiligi(20),
                            dogumgunu_olasiligi(23), dogumgunu_olasiligi(40),
                            dogumgunu_olasiligi(50), dogumgunu_olasiligi(60))
               )
print(dgunu, digits=3)
```

```
              [,1]   [,2]   [,3]   [,4]  [,5]   [,6]
kisi_sayisi 10.000 20.000 23.000 40.000 50.00 60.000
olasilik     0.117  0.411  0.507  0.891  0.97  0.994
```

Gruptaki kişi sayısı arttıkça eşleşme olasılığı artmaktadır. Şekil 1.15 kişi sayısına göre olasılıkların nasıl değiştiğini göstermektedir. 23 kişilik bir grupta en az iki kişinin aynı doğumgününe sahip olma olasılığı %50.7'dir. 50 kişilik bir grupta ise yaklaşık %97, 60 kişilik bir grupta ise %99'dur.



Şekil 1.15: Doğumgünü problemi: kişi sayısına göre en az bir doğumgünü eşleşme olasılığı

1.5 Koşullu Olasılık

A ve B iki olay olsun. Bu olayların birlikte gerçekleşme ($A \cap B$) olasılığı nedir? Bu durumda bir olay verilmişken (gerçekleşmişken) diğer olayın olasılığını, yani **koşullu olasılıkları** hesaplamamız gerekir.

B verilmişken A olayının koşullu olasılığı aşağıdaki gibi tanımlıdır:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \tag{1.11}$$

Benzer şekilde A verilmişken B olayının koşullu olasılığı

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \tag{1.12}$$

olarak tanımlanır. $\mathbb{P}(B|A)$ ve $\mathbb{P}(A|B)$ notasyonundaki | sembolünün sağındaki olay verilmiş/gerçekleşmiş olaydır ve belirsizlik barındırmaz. | sembolünü `R`'daki "ya da" işlemiyle karıştırmayınız.

$\mathbb{P}(A|B)$ ve $\mathbb{P}(B|A)$ koşullu olasılıklarının birbirine eşit olması gerekmez. Paydada yer alan tekil olasılıklar 0 ise koşullu olasılık tanımsız olur. Koşullu olasılık tanımında *payın aynı olduğuna* dikkat ediniz.

Kural 5 (Olasılığın Çarpım Kuralı) A ve B olaylarının birlikte gerçekleşme olasılıkları, yani birleşik olasılıklar, koşullu olasılıktan hareketle aşağıdaki gibi yazılabilir:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) \tag{1.13}$$

ve

$$\mathbb{P}(A \cap B) = \mathbb{P}(B|A)\mathbb{P}(A) \tag{1.14}$$

Örnek 1.23. Hilesiz bir zar atılıyor ve çift sayı geldiği görülüyor. Zarın 6 olma olasılığı kaçtır?

Çözüm:

$$\mathbb{P}(\{6\} \mid \{2,4,6\}) = \frac{\mathbb{P}(\{6\} \ \cap \ \{2,4,6\})}{\mathbb{P}(\{2,4,6\})} = \frac{1/6}{3/6} = \frac{1}{3}$$

`R` ile simülasyonunu yapalım:

```
set.seed(899)
# payda: A ve B olasılığı
A_ve_B <- replicate(10000, {
  zar <- sample(1:6, 1, replace = TRUE)
  (zar==6) & (zar==2 | zar==4 | zar==6)
})
P_A_ve_B <- mean(A_ve_B)
# pay: B olasılığı
B <- replicate(10000, {
  zar <- sample(1:6, 1, replace = TRUE)
  (zar==2 | zar==4 | zar==6)
})
P_B <- mean(B)
# koşullu olasılık
P_A_ve_B/P_B
```

```
[1] 0.3347
```

Örnek 1.24. Örnek 1.16 verilerini kullanacağız. Bir anket çalışmasında 100 kişi, favori içecekleri olarak çay, kahve ve meyve suyu tercihlerini belirtmişlerdir. Anket sonuçları aşağıdaki gibidir:

- 55 kişi çayı tercih ediyor.
- 45 kişi kahveyi tercih ediyor.
- 32 kişi meyve suyunu tercih ediyor.
- 18 kişi hem çayı hem de kahveyi tercih ediyor.
- 12 kişi hem çayı hem de meyve suyunu tercih ediyor.
- 8 kişi hem kahveyi hem de meyve suyunu tercih ediyor.
- 6 kişi her üç içeceği de tercih ediyor.

Buna göre aşağıdaki soruları cevaplayınız:

a) Meyve suyu tercih etmediği bilinen bir kişinin kahve tercih etme olasılığı nedir?
b) Meyve suyu tercih etmediği bilinen bir kişinin çay tercih etme olasılığı nedir?

Çözüm:

Olayları tanımlayalım:

- : "çay tercih ediyor."
- K: "Kahve tercih ediyor."
- M: "Meyve suyu tercih ediyor"

Olasılıklar:

$$\mathbb{P}() = 0.55, \quad \mathbb{P}(K) = 0.45, \quad \mathbb{P}(M) = 0.32$$
$$\mathbb{P}(\cap K) = 0.18, \quad \mathbb{P}(\cap M) = 0.12, \quad \mathbb{P}(K \cap M) = 0.08$$
$$\mathbb{P}(\cap K \cap M) = 0.06$$

Meyve suyu tercih edenlerin sayısı 32, etmeyenler 68 kişidir. Rassal seçilen birinin meyve suyu içmeme olasılığı

$$\mathbb{P}(\bar{M}) = 0.68$$

olur. Meyve suyu tercih etmeyen 68 kişiden 25'i sadece kahve, 12 kişi çay ve kahve (meyve suyu hariç) tercih ediyor. Öyleyse

$$\mathbb{P}(K \cap \bar{M}) = 0.37$$

Meyve suyu tercih etmediği bilinen bir kişinin kahve tercih etme olasılığı:

$$\mathbb{P}(K|\bar{M}) = \frac{\mathbb{P}(K \cap \bar{M})}{\mathbb{P}(\bar{M})} = \frac{0.37}{0.68} \approx 0.54$$

yaklaşık %54 olur.

Meyve suyu tercih etmeyen 68 kişiden 31 kişi sadece çay ve 12 kişi çay ve kahve (meyve suyu hariç) tercih ediyor. Meyve suyu tercih etmediği bilinen bir kişinin çay tercih etme olasılığı:

$$\mathbb{P}(|\bar{M}) = \frac{\mathbb{P}(\cap \bar{M})}{\mathbb{P}(\bar{M})} = \frac{0.43}{0.68} \approx 0.63$$

yaklaşık %63 olur.

Tanım 1.5 (Bağımsızlık)**.** Eğer A olayının gerçekleşmesi B olayını etkilemiyorsa (ya da tersi) bu iki olay birbirinden **bağımsızdır**. Bu durumda olayların birlikte gerçekleşme olasılığı

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \tag{1.15}$$

olur. Örneğin hilesiz bir para peşpeşe iki kere atıldığında birinci sonuç ikinci sonuç hakkında bilgi içermez, yani bu olaylar bağımsızdır. Öyleyse ikisinin de tura olma olasılığı $0.5 \times 0.5 = 0.25$ olur.

Eğer A ve B *birbirinden bağımsız* olaylarsa bir olayın gerçekleşmesi diğeri hakkında bilgi vermez. Bu durumda

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

ve

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

olur.

Örnek 1.25. Borsanın günü artışla kapatma olasılığı %50'dir. Havanın güneşli olma olasılığı %60'dır. Ayrıca, borsanın artması ve havanın güneşli olma olasılığı %30'dur. "Borsanın artması" ve "Havanın güneşli olması" olayları istatistik bakımından birbirinden bağımsız mıdır?

Çözüm:

Olayları tanımlayalım:

- A: "Borsanın artması", $\mathbb{P}(A) = 0.5$
- B: "Güneşli hava", $\mathbb{P}(B) = 0.6$

Bu iki olayın birlikte gerçekleşme olasılığı ise

$$\mathbb{P}(A \cap B) = 0.30$$

olarak verilmiştir. Buradan

$$\mathbb{P}(A)\mathbb{P}(B) = (0.5) \cdot (0.6) = 0.30 = \mathbb{P}(A \cap B)$$

olduğu için bu iki olay birbirinden bağımsızdır.

Örnek 1.26. Bir ülkede hanelerin %26'sı düşük gelir düzeyine sahiptir. Hanelerin %25'i tasarruf yaptığını, %75'i ise tasarruf yapmadığını beyan etmiştir. Ayrıca, hanelerin %3 kadarı düşük gelir grubundadır ve tasarruf yapmaktadır. Buna göre, düşük gelir ve tasarruf birbirinden bağımsız olaylar mıdır?

Çözüm: Verilenlerden

$$\mathbb{P}(\text{Tasarruf}) = 0.25, \quad \mathbb{P}(\text{Düşük Gelir}) = 0.26$$

$$\mathbb{P}(\text{Tasarruf } \cap \text{Düşük Gelir}) = 0.03$$

Buradan

$\mathbb{P}(\text{Tasarruf}) \cdot \mathbb{P}(\text{Düşük Gelir}) = (0.25) \cdot (0.26) = 0.065 \neq 0.03 = \mathbb{P}(\text{Tasarruf } \cap \text{Düşük Gelir})$

bulunur. Öyleyse tasarruf ve gelir düzeyi birbirinden istatistik bakımından bağımsız değildir.

Önerme 1.4. *B olayı verilmişken A ve B olaylarının birlikte gerçekleşme olasılıkları B verilmişken A olayının koşullu olasılığıdır:*

$$\mathbb{P}(A \cap B|B) = \mathbb{P}(A|B)$$

***İspat**: Doğrudan koşullu olasılık tanımını kullanarak*

$$\begin{aligned}\mathbb{P}(A \cap B|B) &= \tfrac{\mathbb{P}((A\cap B)\cap B)}{\mathbb{P}(B)} \\ &= \tfrac{\mathbb{P}(A\cap(B\cap B)}{\mathbb{P}(B)} \\ &= \tfrac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = \mathbb{P}(A|B)\end{aligned}$$

Alıştırma: $\mathbb{P}(A \cup B|B) = ?$

1.6 Bayes Kuralı

Teorem 1.1 (Toplam olasılık kuralı)**.** *A ve B örneklem uzayı S içinde tanımlı iki olay olsun. B^c B olayının tümleyeni olmak üzere*

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c) \tag{1.16}$$

olarak yazılabilir.

***İspat**:*

Herhangi bir olay ile bu olayın tümleyeninin birleşimi örneklem uzayını verir, yani

$$S = B \cup B^c \tag{1.17}$$

olur. Yine herhangi bir olayın örneklem uzayı ile kesişimi o olayı verir:

$$A = A \cap S \tag{1.18}$$

Bu ikisini birleştirirsek

$$A = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c) \tag{1.19}$$

yazabiliriz. $(A \cap B)$ ve $(A \cap B^c)$ karşılıklı bağdaşmaz olaylar olduğu için

$$\mathbb{P}(A) = \mathbb{P}((A \cap B) \cup (A \cap B^c)) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) \tag{1.20}$$

bulunur. Koşullu olasılık tanımından hareketle teoremin son eşitliğine kolayca ulaşılır.

Bu kural genelleştirilebilir.

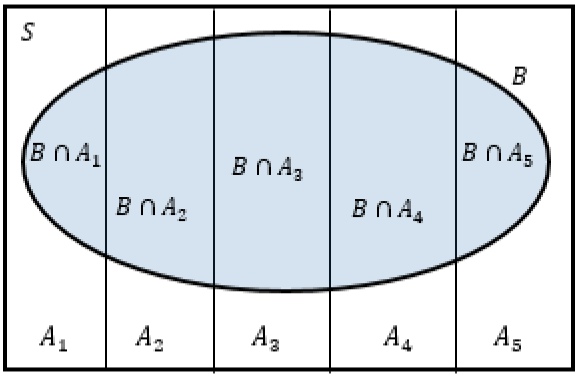
Teorem 1.2 (Genelleştirilmiş toplam olasılık kuralı)**.** *$A_1, A_2, \ldots, A_k$ olayları karşılıklı bağdaşmaz ve birleşimleri örneklem uzayı S olsun, yani*

$$A_i \cap A_j = \emptyset, \quad i \neq j \tag{1.21}$$

ve

$$\bigcup_{i=1}^{k} A_i = S \tag{1.22}$$

Bu iki özelliği sağlayan $A_1, A_2, \ldots, A_k$ olaylarına S'nin bir partisyonu denir (Şekil 1.16).



Şekil 1.16: Örneklem uzayının bir partisyonu

Bu durumda toplam olasılık kuralı

$$\mathbb{P}(B) = \sum_{i=1}^{k} \mathbb{P}(B \cap A_i) = \sum_{i=1}^{k} \mathbb{P}(B|A_i)\mathbb{P}(A_i) \tag{1.23}$$

olur.

Teorem 1.3 (Bayes kuralı)**.** *A ve B örneklem uzayı S'de tanımlı iki olay olsun. B verilmişken A'nın koşullu olasılığı*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c)} \tag{1.24}$$

olarak yazılabilir. Burada

- *$\mathbb{P}(A)$ önsel olasılık (prior)*
- *$\mathbb{P}(A|B)$ ardıl olasılık (posterior)*

olarak isimlendirilir. önsel olasılık B olayına ilişkin bilginin olmadığı durumda kişisel inanç düzeyini yansıtır. Bayes Teoremi B olayı gözlemlendikten sonra önsel olasılık $\mathbb{P}(A)$'nın nasıl güncelleştirildiğini ve ardıl olasılığa ulaşıldığını gösterir.

İspat:

Koşullu olasılık tanımından hareketle

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A)$$

yazılabileceğini biliyoruz. Paydaya bunu yazıp, paya toplam olasılık kuralını uygularsak sonuca ulaşırız.

Bayesçi istatistik bu kural üzerinde inşa edilmiştir. Bayes kuralı $\mathbb{P}(A)$ ile temsil edilen önsel (prior) olasılıkların, B olayı gerçekleştiğinde nasıl güncelleneceğine ilişkin bir yol sunar. Böylece B bilgisinden hareketle ardıl (posterior) olasılık $\mathbb{P}(A|B)$:

$$\mathbb{P}(A|B) = q \cdot \mathbb{P}(A), \quad q = \frac{\mathbb{P}(B|A)}{\mathbb{P}(B)}$$

ile bulunur.

Teorem 1.4 (Toplam olasılıklar ile Bayes kuralı)**.** $A_1, A_2, \ldots, A_k$ *olayları* S*'ni bir partisyonu ise Bayes kuralı*

$$\mathbb{P}(A_j|B) = \frac{\mathbb{P}(B|A_j)\mathbb{P}(A_j)}{\sum_{i=1}^{k} \mathbb{P}(B|A_i)\mathbb{P}(A_i)} \tag{1.25}$$

Paydada yer alan $\mathbb{P}(B)$*'nin toplam olasılık formunda yazıldığına dikkat ediniz.*

Örnek 1.27. Bir hisse senedi piyasası uzmanı piyasanın yönüne ilişkin tahminlerde bulunmaktadır. Geçmiş verilerden hareketle uzmanın doğru tahmin yapma, yani hisse gerçekten yükseldiğinde yükselecek tahmini yapmış olması ve hisse düştüğünde uzmanın düşecek tahmini yapmış olma olasılığı $P(D) = 0.8$ olarak bulunmuştur. Hisse senedinin gerçekten yükselme olasılığı $P(\text{Gerçek+}) = 0.6$ düşme olasılığı $P(\text{Gerçek}-) = 0.4$ olsun. Bu borsa uzmanı bir hissenin yükseleceğini söylediğinde gerçekten yükselme (doğru olma) olasılığı kaçtır?

Çözüm

Verilerden analistin doğru tahmin olasılığı

$$\mathbb{P}(D) = P(\text{Tahmin } + | \text{ Gerçek } +) = \mathbb{P}(\text{Tahmin } - | \text{ Gerçek } -) = 0.8$$

ve yanlış tahmin yapma olasılığı

$$\mathbb{P}(Y) = \mathbb{P}(\text{Tahmin } + | \text{ Gerçek } -) = \mathbb{P}(\text{Tahmin } - | \text{ Gerçek } +) = 0.2$$

olarak yazılabilir.

Bayes teoreminden hareketle

$$\mathbb{P}(\text{Gerçek} + |\text{Tahmin } +) = \frac{\mathbb{P}(\text{Tahmin } + | \text{ Gerçek } +)\mathbb{P}(\text{Gerçek } +)}{\mathbb{P}(\text{Tahmin } +)}$$

Paydada yer alan uzmanın yükseleceğini tahmin etme olasılığını bulmak için toplam kuralını uygulayabiliriz:

$$\begin{aligned}\mathbb{P}(\text{Tahmin }+) &= \mathbb{P}(\text{Tahmin }+|\text{Gerçek}+)\mathbb{P}(\text{Gerçek}+) + \mathbb{P}(\text{Tahmin }+|\text{Gerçek}-)\mathbb{P}(\text{Gerçek}-)\\ &= 0.8 \times 0.6 + 0.2 \times 0.4\\ &= 0.48 + 0.08\\ &= 0.56\end{aligned}$$

Buradan

$$\begin{aligned}\mathbb{P}(\text{Gerçek} + |\text{Tahmin }+) &= \frac{\mathbb{P}(\text{Tahmin }+|\text{Gerçek}+)\mathbb{P}(\text{Gerçek}+)}{\mathbb{P}(\text{Tahmin }+)}\\ &= \frac{0.8 \times 0.6}{0.56}\\ &= 0.857\end{aligned}$$

bulunur.

1.7 Birleşik Olasılıklar ve Çapraz Tablolar

280 kişiden oluşan bir tüketici kümesine elektronik ticaret için geliştirilmiş yeni bir uygulamayı beğenip beğenmedikleri sorulmuştur. Bu tüketici kümesinin tüm tüketicilerden oluşan kümeyi iyi temsil ettiğini varsayalım. Anketi cevaplayan tüketicilerin cinsiyetleri (Kadın veya Erkek) ve ürünü beğenip beğenmedikleri (Evet veya Hayır) bir tablo halinde özetlenebilir. Tablo 1.2 bu verilerin özetini vermektedir. Bu tabloda yer alan sayılar ilgili satır ve sütunlara karşılık gelen özellikleri ya da cevapları sağlayan gözlem sayılarıdır. Örneğin ürünü beğenen kadınların sayısı 75 (Evet ve Kadın), beğenmeyen kadınların sayısı 63 (Hayır ve Kadın) olarak bulunmuştur. Bu kümedeki kadınların sayısı 138, erkeklerin sayısı 142, ürünü beğenenlerin sayısı 158 ve beğenmeyenlerin sayısı 122'dir.

Tablo 1.2: 2×2 Çapraz Tablo

	Kadın	Erkek	Toplam
Evet	75	83	158
Hayır	63	59	122
Toplam	138	142	280

Bu tabloyu sıklıklar yerine olasılıklarla (oranlar) da ifade edebiliriz (bkz. Tablo 1.3). Bu tablodaki hücreler ilgili satır ve sütunun gerçekleşmesine ilişkin arakesit/kesişim olasılıklarıdır. Örneğin rassal olarak bir tüketici seçersek bu tüketicinin erkek olması ve uygulamayı beğenme olasılığı $\mathbb{P}(Evet \cap \; Erkek) = 0.30$'dur. Benzer şekilde diğer olasılıklar da tanımlanabilir.

Tablo 1.3: 2×2 Çapraz Tablo

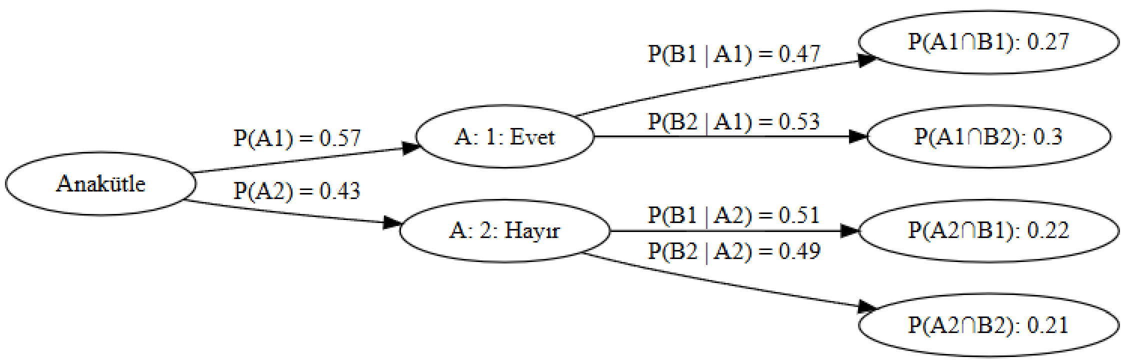
	Kadın	Erkek	Toplam
Evet	0.27	0.30	0.57
Hayır	0.22	0.21	0.43
Toplam	0.49	0.51	1

Bu tabloda satır ve sütunlarda temsil edilen kategorik özellikleri örneklem uzayındaki olay olarak tanımlayabiliriz. Bu iki olay kümesi (cinsiyet ve uygulamayı beğenme) A ve B grubu olaylar olarak tanımlanabilir. Örneğimizde A olayları ürünü beğenip beğenmeme durumunu, B olayları ise cinsiyeti temsil etmektedir. Her iki olay grubunun karşılıklı bağdaşmaz ve bütünü kapsayıcı olduğuna dikkat ediniz. Tablo 1.4 her biri iki gruptan oluşan bu çapraz tabloyu göstermektedir.

Tablo 1.4: 2×2 Çapraz Tablo (birleşik olasılıklar)

	B_1	B_2	Toplam
A_1	$\mathbb{P}(A_1 \cap B_1)$	$\mathbb{P}(A_1 \cap B_2)$	$\mathbb{P}(A_1)$
A_2	$\mathbb{P}(A_2 \cap B_1)$	$\mathbb{P}(A_2 \cap B_2)$	$\mathbb{P}(A_2)$
Toplam	$\mathbb{P}(B_1)$	$\mathbb{P}(B_2)$	1

Birleşik olasılıkları (arakesit ya da kesişim olasılıkları) gösteren bu çapraz tablodan hareketle koşullu olasılıkları ve marjinal olasılıkları kolayca hesaplayabiliriz. Bu olasılıkları bir ağaç diyagramı ile de temsil edebiliriz.



Şekil 1.17: Ağaç diyagramı ile olasılıkların gösterilmesi

Şekil 1.17, Tablo 1.3'e verilen çapraz tablonun ağaç diyagramını göstermektedir. Bu diyagramın sol tarafında tüm anakütle temsil edilir. Sağ tarafta A_1 ve A_2 olaylarının marjinal olasılıkları gösterilmiştir (sırasıyla, evet ve hayır). Çapraz tablodan da görüleceği gibi, ürünü beğenen birinin kadın olma olasılığı

$$\mathbb{P}(B_1|A_1) = \frac{\mathbb{P}(A_1 \cap B_1)}{\mathbb{P}(A_1)} = \frac{0.27}{0.57} \approx 0.47$$

olur. Diğer koşullu olasılıklar benzer şekilde hesaplanabilir. Ağaç diyagramının en sonunda birleşik olasılıklar yer almaktadır. Marjinal olasılıklarla koşullu olasılıkların çarpımına eşittir.

Tablo 1.5: İki grup olay için birleşik olasılık tablosu, $r \times c$

	B_1	B_2	...	B_c	
A_1	$\mathbb{P}(A_1 \cap B_1)$	$\mathbb{P}(A_1 \cap B_2)$	...	$\mathbb{P}(A_1 \cap B_c)$	$\mathbb{P}(A_1)$
A_2	$\mathbb{P}(A_2 \cap B_1)$	$\mathbb{P}(A_2 \cap B_2)$	...	$\mathbb{P}(A_2 \cap B_c)$	$\mathbb{P}(A_2)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_r	$\mathbb{P}(A_r \cap B_1)$	$\mathbb{P}(A_r \cap B_2)$	...	$\mathbb{P}(A_r \cap B_c)$	$\mathbb{P}(A_r)$
	$\mathbb{P}(B_1)$	$\mathbb{P}(B_2)$	...	$\mathbb{P}(B_c)$	1

Buradaki analiz r ve c kategoriyi içeren iki grup olay için genelleştirilebilir. Tablo 1.5 karşılıklı bağdaşmaz ve bütünü kapsayıcı $A_1, A_2, \dots, A_r$ ve $B_1, B_2, \dots, B_c$ olay kümelerinin birleşik olasılıklarını göstermektedir. A grubu olaylar r kategoriden oluşurken, B grubu olaylar c kategori içermektedir.

Örnek 1.28. Bir anket çalışmasında 450 öğrenci üzerinde sosyal medya kullanımı ve akademik başarı (GPA) durumu incelenmiştir. Günlük sosyal medya kullanım süresi ile akademik başarı arasında aşağıdaki tablo oluşturulmuştur.

Sosyal Medya Kullanımı (Günlük)	Düşük (GPA < 2.0)	Orta (2.0 GPA < 3.0)	Yüksek (GPA 3.0)	Toplam
Düşük (0-1 saat)	10	30	80	120
Orta (1-3 saat)	20	60	70	150
Yüksek (3 saatten fazla)	90	50	40	180
Toplam	120	140	190	450

A_i sosyal medya kullanım düzeyini, B_j GPA düzeyini göstersin:

- A_1: Düşük düzey sosyal medya kullanımı
- A_2: Orta düzey sosyal medya kullanımı
- A_3: Yüksek düzey sosyal medya kullanıma

- B_1: Düşük GPA
- B_2: Orta GPA
- B_3: Yüksek GPA

Birleşik olasılıkları, $\mathbb{P}(A_i \cap B_j)$, gösteren çapraz tablo:

Sosyal Medya / GPA	B_1: Düşük GPA	B_2: Orta GPA	B_3: Yüksek GPA	Toplam
A_1: **Düşük**	$\frac{10}{450} = 0.022$	$\frac{30}{450} = 0.067$	$\frac{80}{450} = 0.178$	$\frac{120}{450} = 0.267$
A_2: **Orta**	$\frac{20}{450} = 0.044$	$\frac{60}{450} = 0.133$	$\frac{70}{450} = 0.156$	$\frac{150}{450} = 0.333$
A_3: **Yüksek**	$\frac{90}{450} = 0.200$	$\frac{50}{450} = 0.111$	$\frac{40}{450} = 0.089$	$\frac{180}{450} = 0.400$
Toplam	$\frac{120}{450} = 0.267$	$\frac{140}{450} = 0.311$	$\frac{190}{450} = 0.422$	1.000

Buna göre aşağıdaki soruları yanıtlayalım.

- **Rassal seçilmiş bir öğrencinin sosyal medya kullanımının yüksek olma olasılığı kaçtır?**

Tablodan yüksek sosyal medya kullanıcısı olan toplamda 180 öğrenci olduğunu görüyoruz. Öyleyse bir öğrencinin sosyal medya kullanımının yüksek olma olasılığı:

$$\mathbb{P}(\text{Yüksek sosyal medya kullanımı}) = \mathbb{P}(A_3) = \frac{180}{450} = 0.4$$

Bu da %40'lık bir olasılığa karşılık gelir. Alternatif olarak bu olasılığı kesişim olasılıkların toplamı olarak da bulabiliriz:

Yüksek sosyal medya kullanımı ve GPA   3.0 olasılığı:

$$\mathbb{P}(\text{Yüksek sosyal medya ve GPA} \quad 3.0) = \mathbb{P}(A_3 \cap B_3) = \frac{40}{450} = 0.0889$$

Yüksek sosyal medya kullanımı ve 2.0   GPA < 3.0 olasılığı:

$$\mathbb{P}(\text{Yüksek sosyal medya ve } 2.0 \quad \text{GPA} < 3.0) = \mathbb{P}(A_3 \cap B_2) = \frac{50}{450} = 0.111$$

Yüksek sosyal medya kullanımı ve GPA < 2.0 olasılığı:

$$\mathbb{P}(\text{Yüksek sosyal medya ve GPA} < 2.0) = \mathbb{P}(A_3 \cap B_1) = \frac{90}{450} = 0.2$$

Toplam yüksek sosyal medya kullanımı olasılığı:

$$\mathbb{P}(A_3) = \mathbb{P}(A_3 \cap B_1) + \mathbb{P}(A_3 \cap B_2) + \mathbb{P}(A_3 \cap B_3) = 0.2 + 0.111 + 0.089 = 0.4$$

arakesit (birleşik) olasılıkların toplamı olarak bulunabilir.

- **Rassal seçilmiş bir öğrencinin not ortalamasının (GPA) düşük olma olasılığı kaçtır?**

Bir öğrencinin düşük GPA'ye sahip olma olasılığı

$$\mathbb{P}(B_1) = \mathbb{P}(\text{Düşük GPA}) = \frac{120}{450} \approx 0.2667$$

yaklaşık %27'dir.

- **Bir öğrencinin yüksek sosyal medya kullanımı ve düşük GPA düzeyine sahip olma olasılığı nedir?**

Tablodan yüksek sosyal medya kullanımı ve düşük GPA'ye sahip öğrenci sayısının 90 olduğunu görüyoruz. Öyleyse bir öğrencinin yüksek sosyal medya kullanımı ve düşük GPA'ye sahip olma olasılığı:

$$\mathbb{P}(\text{Yüksek sosyal medya ve düşük GPA}) = \mathbb{P}(A_3 \cap B_1) = \frac{90}{450} = 0.2$$

%20'lik bir olasılığa karşılık gelir.

- **Bir öğrencinin düşük sosyal medya kullanımı ve yüksek GPA'ye sahip olma olasılığı nedir?**

$$\mathbb{P}(A_1 \cap B_3) = \mathbb{P}(\text{Düşük sosyal medya ve yüksek GPA}) = \frac{80}{450} \approx 0.178$$

Yaklaşık %17.8 olur.

- **Sosyal medya kullanımının yüksek olduğu bir öğrencinin düşük GPA düzeyine sahip olma olasılığı nedir?**

Bu soruda öğrencinin yüksek sosyal medya kullanıcısı olduğu bilinmektedir. İstenen koşullu olasılık

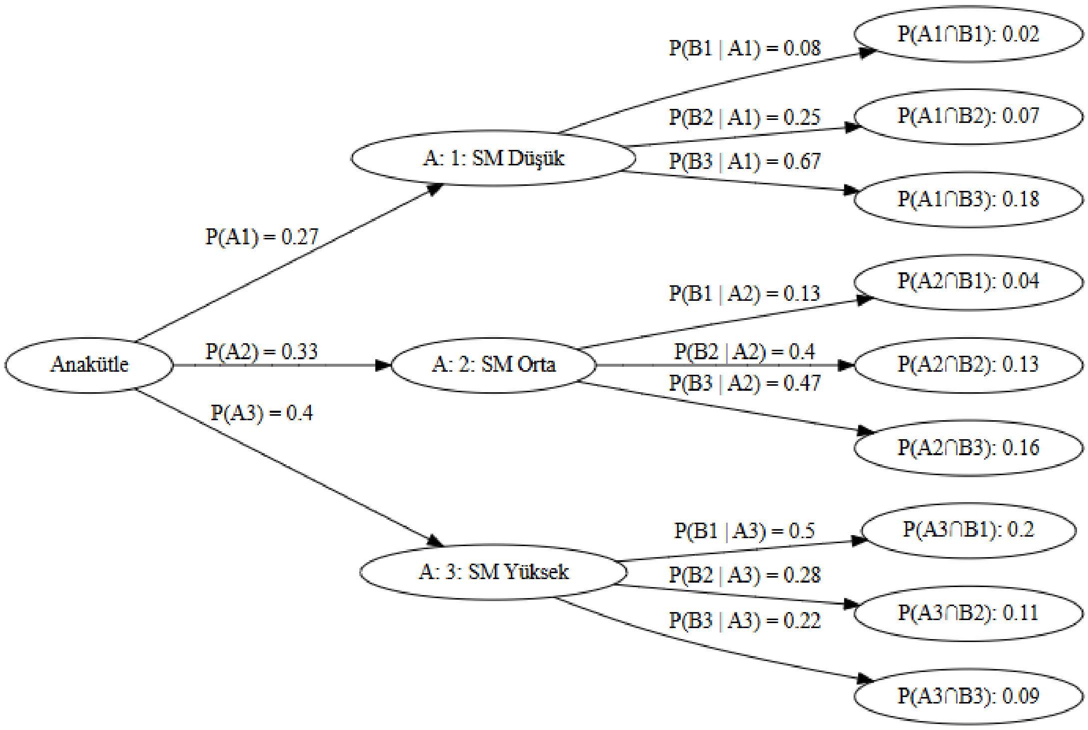
$$\mathbb{P}(\text{Düşük GPA} \mid \text{Yüksek sosyal medya}) = \frac{P(\text{Yüksek sosyal medya ve Düşük GPA})}{P(\text{Yüksek sosyal medya})}$$

ya da

$$\mathbb{P}(B_1|A_3) = \frac{\mathbb{P}(A_3 \cap B_1)}{\mathbb{P}(A_3)} = \frac{0.2}{0.4} = 0.5$$

bulunur.

Şekil 1.18 bu problemin olasılık ağaç diyagramını göstermektedir. Olayların tanımı şöyledir:



Şekil 1.18: Ağaç diyagramı: Sosyal medya ve akademik başarı

- A_1: Düşük sosyal medya kullanımı
- A_2: Orta düzey sosyal medya kullanımı
- A_3: Yüksek düzey sosyal medya kullanımı
- B_1: Düşük GPA
- B_2: Orta GPA
- B_3: Yüksek GPA

Düşük düzeyde sosyal medya kullanıcısı olan bir öğrencinin yüksek GPA olma olasılığını bulalım:

$$\mathbb{P}(B_3|A_1) = \frac{\mathbb{P}(B_3 \cap A_1)}{\mathbb{P}(A_1)} = \frac{0.18}{0.27} \approx 0.67$$

Diğer koşullu olasılıklar benzer şekilde bulunabilir.

1.8 Çözümlü Alıştırmalar

Alıştırma 1.1. İki hilesiz zar atılıyor. Zarların toplamının 7 olma olasılığı nedir?

Çözüm:

A olayı "İki zarın toplamının 7 olması" olsun. Örneklem uzayı 36 temel sonuçtan oluşur ve herbirinin olasılığı eşittir (bkz. Örnek 1.9).

Toplam 6 farklı durumda A olayı gerçekleşir: (1,6), (2,5), (3,4), (4,3), (5,2), (6,1). Öyleyse istenen olasılık

$$P(A) = \frac{6}{36} = \frac{1}{6}.$$

olur.

Alıştırma 1.2. Önceki soruda iki zarın toplamının 8 olma olasılığını bulunuz.

Çözüm:

Toplamı 8 yapan ikililer: (2,6), (3,5), (4,4), (5,3), (6,2)

Her bir zar atışı $\frac{1}{6}$ olasılıkla gerçekleşir.

İstenen olasılık

$$\mathbb{P}(\text{toplam } 8) = \frac{5}{36}$$

olur.

Alıştırma 1.3. Bir torbada 8 kırmızı ve 4 mavi top bulunmaktadır. İki top yerine koymadan çekiliyor. İki topun da mavi olma olasılığı nedir?

Çözüm:

Olayları tanımlayalım:

- M_1: "İlk topun mavi olması"
- M_2: "İkinci topun mavi olması"

İstenen olasılık her ikisinin de mavi olmasıdır, yani:

$$\mathbb{P}(M_1 \cap M_2)$$

Koşullu olasılık tanımından hareketle

$$\mathbb{P}(M_2|M_1) = \frac{\mathbb{P}(M_1 \cap M_2)}{\mathbb{P}(M_1)}$$

Buradan da

$$\mathbb{P}(M_1 \cap M_2) = \mathbb{P}(M_1) \cdot \mathbb{P}(M_2|M_1) = \frac{4}{12} \cdot \frac{3}{11} = \frac{1}{11}.$$

bulunur.

Bir torbada 5 kırmızı, 3 yeşil ve 2 mavi bilye vardır. Torbadan yerine koymadan art arda iki bilye çekiliyor. İlk bilyenin kırmızı, ikinci bilyenin yeşil olma olasılığını bulun.

Çözüm:

Olayları tanımlayalım:

- K: "İlk bilyenin kırmızı olması"
- Y: "İkinci bilyenin yeşil olması"

Olasılık çarpım kuralına göre

$$\mathbb{P}(K \cap Y) = \mathbb{P}(K) \cdot \mathbb{P}(Y|K)$$

yazılabilir. İlgili olasılıklar $\mathbb{P}(K) = \frac{5}{10} = \frac{1}{2}$ ve $\mathbb{P}(Y|K) = \frac{3}{9}$ olduğuna göre

$$\mathbb{P}(K \cap Y) = \frac{1}{2} \cdot \frac{3}{9} = \frac{1}{6}$$

bulunur.

Alıştırma 1.4. Bir sınıftaki 30 öğrenciden 18'i matematik dersini, 15'i fizik dersini seçmiştir. Her iki dersi de seçen öğrenci sayısı 10'dur. Rastgele seçilen bir öğrencinin sadece matematik dersini seçmiş olma olasılığını bulunuz.

Çözüm:

Venn diyagramı kullanarak:

- Sadece matematik: 18 - 10 = 8 öğrenci
- Sadece fizik: 15 - 10 = 5 öğrenci
- Her iki ders: 10 öğrenci
- Hiçbir ders: 30 - 8 - 5 - 10 = 7 öğrenci

olduğu gösterilebilir. Öyleyse

$$P(\text{sadece matematik}) = \frac{8}{30} = \frac{4}{15}$$

bulunur.

Alıştırma 1.5. A ve B şehirleri arasında uçuş yapan bir havayolu şirketinin uçuşlarının %20'si gecikmeli olmaktadır. Gecikmeli uçuşların %15'inde bagajlar kaybolmakta, zamanında olan uçuşların ise %5'inde bagajlar kaybolmaktadır. Rastgele seçilen bir yolcunun bagajının kaybolmuş olma olasılığını bulunuz.

Çözüm:

Olayları tanımlayalım:

- G: Uçuşun gecikmeli olması
- K: Bagajın kaybolması

Verilenler:

- $\mathbb{P}(G) = 0.20$
- $\mathbb{P}(K|G) = 0.15$
- $\mathbb{P}(G^c) = 0.8$, uçuşun gecikmesiz olması (G tümleyen)

- $\mathbb{P}(K|G^c) = 0.05$

Toplam olasılık kuralına göre:

$$\mathbb{P}(K) = \mathbb{P}(K|G) \cdot \mathbb{P}(G) + \mathbb{P}(K|G^c) \cdot \mathbb{P}(G^c)$$

$$\mathbb{P}(K) = 0.15 \cdot 0.20 + 0.05 \cdot 0.80 = 0.03 + 0.04 = 0.07$$

bulunur.

Alıştırma 1.6. Bir torbada 3 kırmızı, 4 mavi top vardır. Torbadan yerine koyarak 3 top çekiliyor. Çekilen topların hepsinin aynı renkte olma olasılığını bulunuz.

Çözüm:

Olay tanımları:

- K: "Hepsi kırmızı"
- M: "Hepsi mavi"

Yerine koyarak çekim yapıldığından olaylar bağımsızdır. Öyleyse hepsinin aynı olma olasılığı

$$\mathbb{P}(\text{hepsi aynı}) = \mathbb{P}(K) + \mathbb{P}(M)$$

yazılabilir.

$$\mathbb{P}(K) = \left(\frac{3}{7}\right)^3 = \frac{27}{343}$$

$$\mathbb{P}(M) = \left(\frac{4}{7}\right)^3 = \frac{64}{343}$$

$$\mathbb{P}(\text{hepsi aynı}) = \frac{27}{343} + \frac{64}{343} = \frac{91}{343} \approx 0.265$$

Alıştırma 1.7. Bir araştırma projesinde görev alacak 4 kişilik bir ekip oluşturulacaktır. Başvuranlar arasında 7 deneyimli (5 yıldan fazla tecrübeli) ve 5 yeni başlayan (2 yıldan az tecrübeli) araştırmacı bulunmaktadır.

a) Seçilen ekipte en az 2 deneyimli araştırmacı bulunması olasılığını bulunuz.
b) Ekipte tam olarak 3 deneyimli araştırmacı bulunması olasılığını bulunuz.
c) Ekibin tamamen yeni başlayan araştırmacılardan oluşması olasılığını bulunuz.

Çözüm:

Önce toplam kaç farklı şekilde 4 kişilik ekip oluşturulabileceğini bulalım:

$$\text{Toplam seçim sayısı} = \binom{12}{4} = \frac{12!}{4!(12-4)!} = \frac{12 \cdot 11 \cdot 10 \cdot 9}{4 \cdot 3 \cdot 2 \cdot 1} = 495$$

a) En az 2 deneyimli araştırmacı olması olasılığını bulmak için bu durumu sağlayan ekiplerin sayısını bulalım:

- 2 deneyimli + 2 yeni başlayan içeren ekip sayısı

$$\binom{7}{2}\binom{5}{2} = 210$$

- 3 deneyimli + 1 yeni başlayan ekip sayısı

$$\binom{7}{3}\binom{5}{1} = 175$$

- 4 deneyimli + 0 yeni başlayan ekip sayısı

$$\binom{7}{4}\binom{5}{0} = 35$$

En az 2 deneyimli araştırmacı durumunu sağlayın ekiplerin sayısı

$$210 + 175 + 35 = 420$$

olur. Öyleyse

$$\mathbb{P}(\text{en az 2 deneyimli}) = \frac{\binom{7}{2}\binom{5}{2} + \binom{7}{3}\binom{5}{1} + \binom{7}{4}\binom{5}{0}}{\binom{12}{4}} = \frac{420}{495} \approx 0.85$$

bulunur.

b) Tam olarak 3 deneyimli olması olasılığı:

$$P(3 \text{ deneyimli}) = \frac{\binom{7}{3}\binom{5}{1}}{\binom{12}{4}} = \frac{35 \cdot 5}{495} = \frac{175}{495} \approx 0.353$$

c) Tamamen yeni başlayanlardan oluşma olasılığı:

$$P(\text{hepsi yeni}) = \frac{\binom{5}{4}\binom{7}{0}}{\binom{12}{4}} = \frac{5}{495} = \frac{1}{99} \approx 0.01$$

Alıştırma 1.8. Bir öğrenci 10 sorudan oluşan bir testte her soruya rastgele cevap veriyor. Her soruda 4 seçenek vardır. Öğrencinin en az 6 soruyu doğru cevaplaması olasılığını bulunuz.

Çözüm:

Her soru için doğru cevap olasılığı $p = \frac{1}{4}$ olmak üzere Binom formülünü (bkz. Örnek 1.20 ve Örnek 1.21) kullanabiliriz:

$$P(X \geq 6) = \sum_{k=6}^{10} \binom{10}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{10-k}$$

$$= \binom{10}{6}\left(\frac{1}{4}\right)^6\left(\frac{3}{4}\right)^4 + \binom{10}{7}\left(\frac{1}{4}\right)^7\left(\frac{3}{4}\right)^3 + \binom{10}{8}\left(\frac{1}{4}\right)^8\left(\frac{3}{4}\right)^2 + \binom{10}{9}\left(\frac{1}{4}\right)^9\left(\frac{3}{4}\right)^1 + \binom{10}{10}\left(\frac{1}{4}\right)^{10}$$
$$\approx 0.0197$$

yaklaşık %2'dir.

Alıştırma 1.9. Bir ailedeki üç çocuğun en az ikisinin kız olma olasılığını bulunuz. Her doğumda kız ve erkek olma olasılıkları eşittir.

Çözüm:

K: Kız, E: Erkek

3 çocuk için olası tüm kombinasyonlar: KKK, KKE, KEK, EKK, EEK, EKE, KEE, EEE (toplam $2^3 = 8$ olası durum)

En az iki kız: KKK, KKE, KEK, EKK

Her bir doğum bağımsız ve $P(K) = P(E) = \frac{1}{2}$

$$
\begin{aligned}
P(\text{en az iki kız}) &= P(KKK) + P(KKE) + P(KEK) + P(EKK) \\
&= \left(\frac{1}{2}\right)^3 + 3 \cdot \left(\frac{1}{2}\right)^3 \\
&= \frac{1}{8} + \frac{3}{8} = \frac{1}{2}
\end{aligned}
$$

Alıştırma 1.10. Bir torbada 1'den 100'e kadar tamsayılarla numaralandarılmış toplar bulunuyor. Çekilen bir sayının 9 ile tam bölünebilme olasılığı nedir? Çekilen sayının 9 ya da 3 ile tam bölünebilme olasılığı nedir?

Çözüm: Örneklem uzayı

$$S = \{1, 2, 3, \dots, 99, 100\}$$

A olayı 9 ile tam bölünme, B olayı 3 ile tam bölünme olsun:

$$A = \{9, 18, 27, 36, 45, 54, 63, 72, 81, 90, 99\}$$

$$B = \{3, 6, 9, \dots, 96, 99\}$$

A olayı 11 temel sonuçtan, B olayı ise 33 temel sonuçtan oluşmaktadır. Öyleyse

$$\mathbb{P}(A) = \frac{11}{100} = 0.11$$

9 ile tam bölünme olasılığı %11'dir. Benzer şekilde

$$\mathbb{P}(B) = \frac{33}{100} = 0.33$$

bulunur. Bu iki olayın kesişim kümesi, yani hem 3 hem de 9 ile tam bölünen sayılardan oluşan küme

$$A \cap B = \{9, 18, 27, 36, 45, 54, 63, 72, 81, 90, 99\} = A$$

Bunun olasılığı da $\mathbb{P}(A \cap B) = 0.11$ olur.

Çekilen bir sayının 9 ile ya da 3 ile tam bölünme olasılığı:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = 0.11 + 0.33 - 0.11 = 0.33$$

%33 olur.

Alıştırma 1.11. Bir torbada 1'den 100'e kadar tamsayılarla numaralandarılmış toplar bulunuyor. Çekilen bir sayının 9, 3, veya 5 ile tam bölünme olasılığı kaçtır? Çekilen bir sayının hem 9, hem 3 hem de 5 ile tam bölünme olasılığı kaçtır?

Çözüm: A 9 ile tam bölünme, B 3 ile tam bölünme, ve C 5 ile tam bölünme olayları olsun. Örneklem uzayı önceki soruda tanımlandığı gibi 1-100 arası tamsayılardan oluşmaktadır. A olayı 11 temel sonuç, B olayı 33 temel sonuç, C olayı 20 temel sonuç içermektedir.

Çekilen bir sayının 9, 3, veya 5 ile tam bölünme olasılığı

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

olarak yazılabilir. $\mathbb{P}(A) = 0.11$, $\mathbb{P}(B) = 0.33$ ve $P(C) = 0.2$ olduğunu biliyoruz. Kesişimleri bulalım:

$$A \cap B = \{9, 18, 27, 36, 45, 54, 63, 72, 81, 90, 99\} = A$$

$$A \cap C = \{45, 90\}$$
$$B \cap C = \{15, 30, 45, 60, 75, 90\}$$
$$A \cap B \cap C = \{45, 90\}$$

Öyleyse

$$\mathbb{P}(A \cup B \cup C) = 0.11 + 0.33 + 0.20 - 0.11 - 0.02 - 0.06 + 0.02 = 0.47$$

%47 bulunur.

Bu olayların kesişim kümesi

$$A \cap B \cap C = \{45, 90\}$$

olduğu gösterilebilir. Öyleyse

$$\mathbb{P}(A \cap B \cap C) = \frac{2}{100} = 0.02$$

olur.

Alıştırma 1.12. Bir torbada 1'den 100'e kadar tamsayılarla numaralandarılmış toplar bulunuyor. Bir top rassal olarak çekiliyor ve 5 ile tam bölünebildiği görülüyor. Bu sayının 3 ile tam bölünebilme olasılığı kaçtır?

Çözüm:

Olayları tanımlayalım:

- A: "Çekilen sayı 5 ile tam bölünebilir."
- B: "Çekilen sayı 3 ile tam bölünebilir."

İstenen olasılık

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

Hem 5 hem de 3 ile tam bölünen toplam 6 sayı vardır:

$$A \cap B = \{15, 30, 45, 60, 75, 90\}$$

Öyleyse

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{6/100}{20/100} = 0.3$$

olur.

Alıştırma 1.13. Bir şehirde hanelerin %52'sinde sigara içilmektedir. Özel sağık sigortası olanların oranı ise 0.20'dir. Ayrıca, özel sigortası olan ve sigara içilen hanelerin oranı %15'dir. "Özel sağlık sigortası var" ve "Sigara içiliyor" olayları istatistik bakımından birbirinden bağımsız mıdır?

Çözüm:

Olayları tanımlayalım:

- A: "Özel sağlık sigortası var", $\mathbb{P}(A) = 0.2$
- B: "Sigara içiliyor", $\mathbb{P}(B) = 0.52$

Bu iki olayın birlikte gerçekleşme olasılığı ise

$$\mathbb{P}(A \cap B) = 0.15$$

olarak verilmiştir. Buradan

$$\mathbb{P}(A)\mathbb{P}(B) = (0.2) \cdot (0.52) \neq 0.15 = \mathbb{P}(A \cap B)$$

olduğu için bu iki olay birbirinden bağımsız değildir.

Alıştırma 1.14. İki madeni para atılıyor. A olayı "iki paranın tura olması" ve B olayı "en az bir paranın yazı gelmesi" olarak tanımlanıyor. Örneklem uzayını oluşturunuz ve bu deneyin `R`'da simülasyonunu yapınız.

Çözüm:

T ve Y, Tura ve Yazı sonuçlarını temsil etmek üzere örneklem uzayını aşağıdaki gibi yazabiliriz:

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

İlgilendiğimiz olaylar:

$$A = \{TT\}$$

$$B = \{TY, YT, YY\}$$

Örneklem uzayı S'de 4 temel sonuç vardır ve bunların gerçekleşme olasılıkları eşittir. Öyleyse $P(A) = 0.25$ ve $P(B) = 0.75$ yazabiliriz. Burada paraların hilesiz olduğu ve yinelemelerin birbirlerini etkilemediği varsayımlarını yaptık.

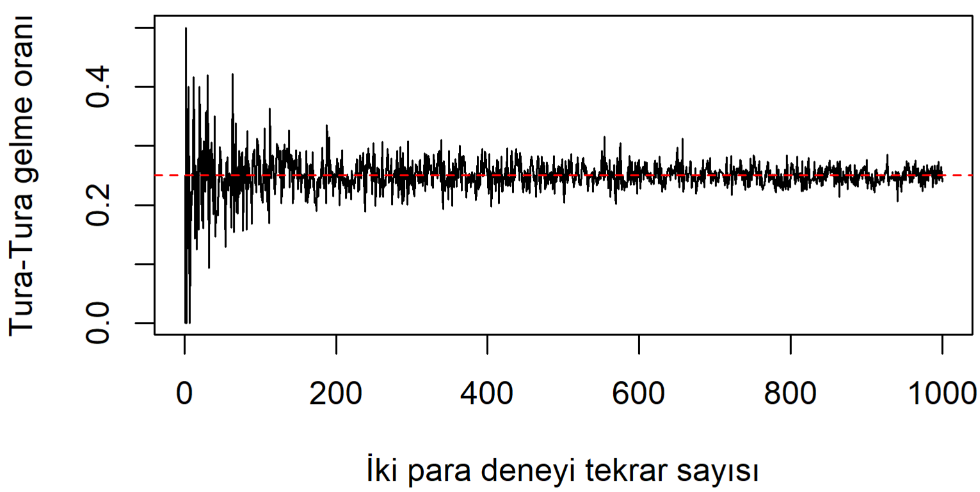
Şimdi R ile bu deneyin simülasyonunu yapalım:

```
set.seed(155)
N <- 1000
P_A <- numeric(N)

for (n in 1:N){
  para1 <- sample(c(1, 0), n, replace = TRUE) # 1.para, T=1, Y=0
  para2 <- sample(c(1, 0), n, replace = TRUE) # 2.para, T=1, Y=0
  P_A[n] <- sum(para1==1 & para2==1)/n        # (1,1) olayının sıklığı
}
P_A[N]
```

```
[1] 0.24
```

```
# grafik
plot(P_A, type="l",
     xlab="İki para deneyi tekrar sayısı",
     ylab="Tura-Tura gelme oranı")
abline(h=0.25, lty="dashed", col="red")
```



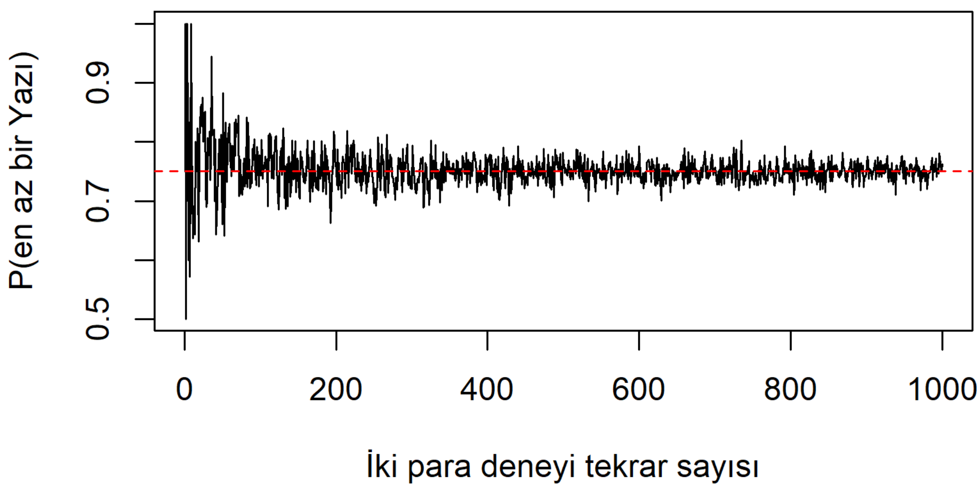
Şekil 1.19: İki para deneyinde Tura-Tura gelme sıklığı

```r
set.seed(156)
N <- 1000
P_B <- numeric(N)

for (n in 1:N){
  para1 <- sample(c(1, 0), n, replace = TRUE)  # 1.para, T=1, Y=0
  para2 <- sample(c(1, 0), n, replace = TRUE)  # 2.para, T=1, Y=0
  P_B[n] <- sum(para1==0 | para2==0)/n  # B="en az biri 0" olayının sıklığı
}
P_B[N]
```

```
[1] 0.763
```

```r
# grafik
plot(P_B, type="l",
     xlab="İki para deneyi tekrar sayısı",
     ylab="P(en az bir Yazı)")
abline(h=0.75, lty="dashed", col="red")
```



Şekil 1.20: İki para deneyinde en az birinin Yazı gelme sıklığı

Sıklık yaklaşımının simülasyonu teorik olasılıklarla yaklaşık aynı sonucu verdi.

R'da olasılıkların simülasyonunu yapmanın başka bir yolu `replicate()` fonksiyonudur. İki para atışında TT gelme olasılığını yaklaştırmak istersek aşağıdaki komutları kullanabiliriz:

```r
set.seed(567)
TT <- replicate(10000,
                {
                  iki_para <- sample(c("T", "Y"), 2, replace = TRUE)
                  iki_para[1]=="T" & iki_para[2]=="T"
                  }
                )
mean(TT)
```

```
[1] 0.25
```

Burada `replicate()` fonksiyonunun ilk girdisi tekrar sayısıdır. Küme parantezi `{}` içinde ise replikasyonu yapacağımız `sample()` fonksiyonu yer almaktadır. Sonuçlar ise `iki_para` vektöründe saklanacaktır. `iki_para[1]=="T" & iki_para[2]=="T"` komutu ise çıktıyı belirleyecektir. Her iki koşul doğru olduğunda TRUE (1) ya da FALSE (0) değerini alacaktır. Bunun ortalaması ise gerçekleşme oranını verecektir. Alıştırma olarak `replicate()` komutunu kullanarak "en az birinin Yazı gelmesi" olasılığını yaklaştırınız.

Alıştırma 1.15 (Monty Hall problemi)**.** 1970'lerde ABD'de popüler olan bir yarışma programındasınız (Let's Make a Deal). Karşınızda üç kapı var: Kapı 1, Kapı 2 ve Kapı 3. Bu kapılardan birinin arkasında bir araba ödülü var, diğer iki kapının arkasında ise keçi bulunuyor (yani ödül yok, keçiyi istemediğinizi varsayıyoruz). Sunucu (Monty Hall) size kapılardan birini seçmenizi istiyor. Seçiminizi yaptıktan sonra, Monty Hall sizin seçmediğiniz kapılardan birini açıyor ve arkasında bir keçi olduğunu gösteriyor. Sonra size şu soruyu soruyor: Seçiminizi değiştirmek ister misiniz yoksa aynı kapıda mı kalmak istersiniz?

Seçiminizi değiştirirseniz arabayı kazanma olasılığınız nedir? Aynı kapıda kalırsanız arabayı kazanma olasılığınız nedir?

Çözüm:

Sadece 1 kapının ardında araba ödülü olduğuna göre kazanma olasılığı $\frac{1}{3}$, keçi olma olasılığı ise $\frac{2}{3}$'tür:

$$\mathbb{P}(\text{Araba}) = \frac{1}{3}, \quad \mathbb{P}(\text{Keçi}) = \frac{2}{3}$$

Yarışmacının 1. kapıyı seçtiğini düşünelim. A_1, A_2, ve A_3, sırasıyla, arabanın 1., 2., ve 3. kapıda olduğunu göstersin. Sunucu, her zaman sizin seçmediğiniz kapılardan birini açacaktır (keçi olacak şekilde). Bu bilgi ışığında seçtiğiniz kapıyı değiştirirseniz kazanma olasılığınız kaç olur?

Ödül (araba) kazanma olasılığı genel olarak aşağıdaki gibi yazılabilir:

$$\mathbb{P}(\text{Ödül kazan}) = \mathbb{P}(\text{Ödül kazan} \mid A_1)\mathbb{P}(Araba) + \mathbb{P}(\text{Ödül kazan} \mid A_2)\mathbb{P}(Araba) + \mathbb{P}(\text{Ödül kazan} \mid A_3)\mathbb{P}(Araba)$$

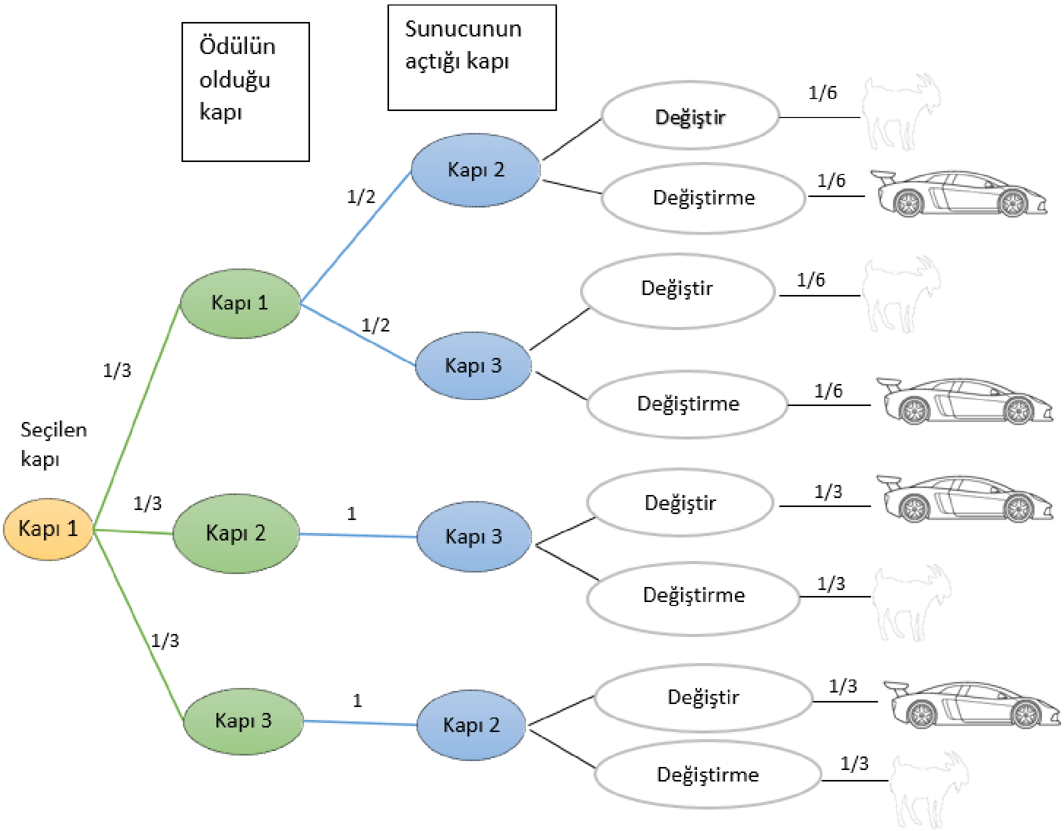
$$\mathbb{P}(\text{Ödül kazan}) = \mathbb{P}(\text{Ödül kazan} \mid A_1) \cdot \frac{1}{3} + \mathbb{P}(\text{Ödül kazan} \mid A_2) \cdot \frac{1}{3} + \mathbb{P}(\text{Ödül kazan} \mid A_3) \cdot \frac{1}{3}$$

Yarışmacının "değiştir" stratejisini takip ettiğini düşünelim. Yarışmacı Kapı 1'i seçmişti. Eğer gerçekten ödül Kapı 1'de ise araba ödülü kesin olarak kaybedilir, yani

$$\mathbb{P}(\text{Ödül kazan} \mid A_1) = 0$$

olur. Eğer araba Kapı 2 ya da Kapı 3'te ise "değiştir" stratejisi kazanmayla sonuçlanır:

$$\mathbb{P}(\text{Ödül kazan}) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = \frac{2}{3}$$



Şekil 1.21: Ödülün 1. kapıda olduğu durumda Monty Hall problemi

Sonuç olarak seçimi değiştirdiğimizde kazanma olasılığımız daha yüksektir. Eğer seçiminizi değiştirirseniz, ilk başta keçiyi seçmişseniz (olasılık 2/3), araba olan kapıya geçeceksiniz ve kazanacaksınız. Eğer arabanın olduğu kapıyı seçmişseniz (olasılık 1/3), seçim değiştirerek kaybedeceksiniz. Bu durumda, seçiminizi değiştirirseniz arabayı kazanma olasılığınız 2/3 olur. Seçiminizi değiştirmezseniz, arabayı kazanma olasılığınız sadece 1/3 olacaktır (bkz. Şekil 1.21).

Bu oyunun `R`'da simülasyonunu yapabiliriz. Önce tesadüfi şekilde araba ve keçileri kapılara atayalım:

```
set.seed(1234)
kapilar <- c("keçi", "keçi", "araba")
kapilar <- sample(kapilar) # rassal olarak ödülü atama
kapilar
```

```
[1] "keçi"  "araba" "keçi"
```

Oyuncunun tercihini rassal olarak belirleyelim:

```
# oyuncunun seçimi
tercih <- sample(1:3, 1)
tercih
```

```
[1] 3
```

Sunucu Monty Hall kalan kapılardan keçili olanı açacaktır:

```
# Monty keçili bir kapıyı açıyor (yarışmacının seçtiği kapıyı açamaz)
monty_nin_secenekleri <- setdiff(1:3, tercih)
monty_nin_secenekleri
```

```
[1] 1 2
```

```
# bu iki kapıdan birini rassal olarak seçer.
# Eğer iki kapıdan birinde araba diğerinde keçi varsa her zaman keçiyi seçer.
# Eğer ikisinde de keçi varsa rassal olarak birini seçer (1/2 olasılıkla)
if (sum(kapilar[monty_nin_secenekleri]=="keçi")==1) {
  monty_nin_actigi_kapi <-
    monty_nin_secenekleri[kapilar[monty_nin_secenekleri]=="keçi"]
} else {
  monty_nin_actigi_kapi <- sample(monty_nin_secenekleri, 1)
}
# ekrana yaz
monty_nin_actigi_kapi
```

```
[1] 1
```

Şimdi yarışmacı "Değiştir" stratejisini uygulayıp uygulamamaya karar verecek. Önündeki seçeneği belirleyelim:

```r
# Yarışmacı ya kapıda kalacak ya da diğer kapıya geçecek
# Seçim değiştirmek için açılmayan kapıyı belirle
degistir_tercih <- setdiff(1:3, c(tercih, monty_nin_actigi_kapi))
degistir_tercih
```

```
[1] 2
```

Yarışmacı ya ilk tercihinde kalacak (Değiştirme) ya da bu kapıya geçecek (Değiştir). Eğer değiştirirse ve bu kapıda araba varsa kazanacak, aksi takdirde kaybedecek:

```r
# Kazanma durumu:
if (kapilar[degistir_tercih] == "araba") {
  degistir_kazanır <- 1       # Seçimi değiştirdiğinde kazandı
}
if (kapilar[tercih] == "araba") {
  degistirmemek_kazanır <- 1  # Seçimini değiştirmediğinde kazandı
}
```

Şimdi olasılığın sıklık tanımından hareketle bu oyununun çok sayıda tekrarını yapan bir fonksiyon yazalım ve "Değiştir" stratejisinin olasılığını yaklaştırmaya çalışalım.

```r
# Monty Hall probleminin simülasyonu için bir fonksiyon
sim_monty_hall <- function(nsim = 10000){
  degistir_kazanır <- 0
  degistirmemek_kazanır <- 0

  for (i in 1:nsim) {
    kapilar <- c("keçi", "keçi", "araba")
    kapilar <- sample(kapilar) # rassal olarak ödülü atama
    # oyuncunun seçimi
    tercih <- sample(1:3, 1)
    # Monty keçili bir kapıyı açıyor
    monty_nin_secenekleri <- setdiff(1:3, tercih)
    if (sum(kapilar[monty_nin_secenekleri]=="keçi")==1) {
      monty_nin_actigi_kapi <-
        monty_nin_secenekleri[kapilar[monty_nin_secenekleri]=="keçi"]
      } else {
        monty_nin_actigi_kapi <- sample(monty_nin_secenekleri, 1)
      }
    degistir_tercih <- setdiff(1:3, c(tercih, monty_nin_actigi_kapi))
    # Kazanma durumu:
    # Değiştirirse:
    if (kapilar[degistir_tercih] == "araba") {
      degistir_kazanır <- degistir_kazanır + 1
```

```
    }
    # Değiştirmezse:
    if (kapilar[tercih] == "araba") {
      degistirmemek_kazanır <- degistirmemek_kazanır + 1
    }
  }


  # Sonuçların oranlarını hesapla
  degistir_kazanma_orani <- degistir_kazanır / nsim
  degistirmeme_kazanma_orani <- degistirmemek_kazanır / nsim

  # Sonuçları yazdır
  cat("Seçim değiştirildiğinde kazanma oranı: ", degistir_kazanma_orani * 100, "%\n")
  cat("Seçimde kaldığında kazanma oranı: ", degistirmeme_kazanma_orani * 100, "%\n")

  # Sonuçları geri döndür (grafik için)
  return(c(degistir_kazanma_orani, degistirmeme_kazanma_orani))
}
```

Oyunu 10000 kez tekrarlayalım ve değiştirme stratejisinin kazanma olasılığını yaklaştıralım:

```
# Simülasyonu 10,000 kez çalıştır
set.seed(123)
sonuclar <- sim_monty_hall(nsim=10000)
```

```
Seçim değiştirildiğinde kazanma oranı:  66.47 %
Seçimde kaldığında kazanma oranı:  33.53 %
```

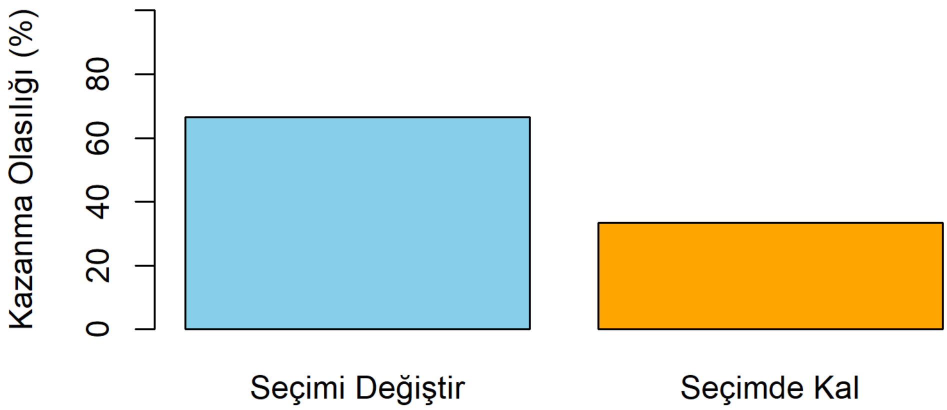
Değiştirme stratejisinin olasılığı %66.5 olarak bulundu (bkz. Şekil 1.22). Bu teorik olasılığa $2/3 \approx 0.6667$ oldukça yakındır.

Bayes teoremi veya olasılık kuralları kullanılarak da bu sonuç gösterilebilir. A_1, A_2, ve A_3, sırasıyla, arabanın 1., 2., ve 3. kapıda olması şeklinde tanımlamıştık. M_1, M_2, ve M_3, sırasıyla, sunucunun 1., 2., ve 3. kapıları açması olsun.

Oyuncunun 1. kapıyı seçtiğini, sunucunun 2. kapıyı açtığını düşünelim. Bayes teoreminden hareketle

$$\mathbb{P}(A_1 \mid M_2) = \frac{\mathbb{P}(M_2 \mid A_1)\mathbb{P}(A_1)}{\mathbb{P}(M_2)}$$

Araba 1. kapının ardındaysa Monty'nin 2. kapıyı açma olasılığı 1/2'dir. Öyleyse



Şekil 1.22: Monty Hall probleminin simülasyon sonuçları

$$\mathbb{P}(A_1 \mid M_2) = \frac{\mathbb{P}(M_2 \mid A_1)\mathbb{P}(A_1)}{\mathbb{P}(M_2)} = \frac{(1/2)\cdot(1/3)}{1/2} = \frac{1}{3}$$

Yarışmacının ilk tercihi 1. kapıydı. Monty Hall 2. kapıyı açtıktan sonra bu tercihin (değiştirme yok) kazanma şansı üçte birdir. Öyleyse değiştirdiği durumda kazanma şansı 2/3 olur.

Alıştırma 1.16. Bir şehirde %2 oranında yaygın bir hastalık olduğunu varsayalım. Bu hastalığın test edilmesi için kullanılan bir testin:

- Hastalığı olan kişiyi doğru tespit etme olasılığı (duyarlılık) %95'tir, yani $\mathbb{P}(T + |H+) = 0.95$.

- Hastalığı olmayan kişiye yanlışlıkla pozitif vermesi olasılığı (yanlış pozitif oranı) %10'dur, yani $P(T + |H-) = 0.10$.

Buna göre, test sonucu pozitif çıktığında, test edilen kişinin **gerçekten** hasta olma olasılığı nedir?

Çözüm:

Bayes teoremini kullanarak hesaplayacağız. Verilenler:

$$\mathbb{P}(H+) = 0.02 \quad \text{(hastalığın toplumdaki prevalansı)}$$

$$\mathbb{P}(H-) = 0.98 \quad \text{(hastalığın olmama olasılığı)}$$

$$\mathbb{P}(T+|H+) = 0.95 \quad \text{(testin hasta birine pozitif vermesi olasılığı)}$$

$$\mathbb{P}(T+|H-) = 0.10 \quad \text{(testin sağlıklı birine pozitif vermesi olasılığı)}$$

Aradığımız olasılık $P(H+|T+)$, yani test sonucu pozitif çıktığında kişinin hasta olma olasılığıdır.

Bayes teoremi:

$$\mathbb{P}(H+|T+) = \frac{\mathbb{P}(T+|H+) \cdot \mathbb{P}(H+)}{\mathbb{P}(T+)}$$

Burada $\mathbb{P}(T+)$, yani testin pozitif olma olasılığı, iki durumun toplamından oluşur:

$$\mathbb{P}(T+) = \mathbb{P}(T+|H+) \cdot \mathbb{P}(H+) + \mathbb{P}(T+|H-) \cdot \mathbb{P}(H-)$$

$$\mathbb{P}(T+) = (0.95 \cdot 0.02) + (0.10 \cdot 0.98)$$

Buradan

$$\mathbb{P}(T+) = 0.019 + 0.098 = 0.117$$

buluruz. Şimdi $P(H+|T+)$'yi hesaplayalım:

$$\mathbb{P}(H+|T+) = \frac{0.95 \cdot 0.02}{0.117} = \frac{0.019}{0.117} \approx 0.162$$

Test sonucu pozitif çıktığında, kişinin gerçekten hasta olma olasılığı yaklaşık **0.162** veya **%16.2**'dir.

1.9 Problemler

Problem 1.1. *Bir sınıfta 40 öğrenci vardır. Öğrencilerden 16'sı Ekonometri dersine, 28'i ise Makine Öğrenmesi dersine kayıtlıdır. 10 öğrenci her iki derse de kayıtlıdır. Rasgele seçilen bir öğrencinin:*

a) Sadece Ekonometri veya sadece Makine Öğrenmesi dersine kayıtlı olma olasılığı nedir?
b) Ekonometri veya Makine Öğrenmesi dersine (veya her ikisine) kayıtlı olma olasılığı nedir?

Problem 1.2. *Hilesiz bir para iki kez atılıyor.*

a) örneklem uzayını oluşturun.
b) En az bir Tura gelme olasılığı kaçtır?
c) İkisinin de Yazı olma olasılığı kaçtır?

Problem 1.3. *İİBF öğrencileri İktisat, İşletme ve Siyaset Bilimi bölümlerine, sınıflarına (1,2,3,4) ve cinsiyetlerine göre (kadın, erkek) sınıflandırılacaktır. Bu şekilde kaç sınıflandırma mümkündür?*

Problem 1.4. *1,2,5,6,9 rakamları kullanılarak*

a) 3 basamaklı kaç sayı oluşturulabilir (rakamlar sadece bir kez kullanılmalı)?
b) 3 basamaklı kaç sayı oluşturulabilir (rakamlar tekrar edebilir)?
c) 3 basamaklı kaç çift sayı oluşturulabilir?

Problem 1.5. *4 iktisatçı ve 3 işletmeci arasından 2 iktisatçı ve 1 işletmecinin olduğu 3 kişilik bir komite kurulacaktır. Kaç farklı komite kurulabilir?*

Problem 1.6. *İki hilesiz zar atılıyor.*

a) Toplamın 7 olma olasılığı kaçtır?
b) Toplamın 11 olma olasılığı kaçtır?
c) Toplamın 7 ya da 11 olma olasılığı kaçtır?
d) Toplamın en fazla 6 olma olasılığı kaçtır?

Problem 1.7. *Standart altı yüzlü 3 zar atılıyor. Örneklem uzayı kaç elemandan oluşur? Bu üç zarın toplamının 10 olma olasılığı kaçtır?*

Problem 1.8. *Bir zar, çift sayıların gelme olasılığı tek sayıların gelme olasılığının 2 katı olacak şekilde hileli hale getiriliyor. Bu durumda zarın*

a) 5 veya 6 gelme olasılığı kaçtır?
b) 4'ten küçük olma olasılığı kaçtır?
c) Zarın çift geldiği biliniyorsa 6 olma olasılığı kaçtır?

Problem 1.9. *Bir kutuda 4 kırmızı, 5 mavi ve 3 yeşil top bulunmaktadır. Rasgele 2 top çekiliyor. İlk top kırmızı çıktıktan sonra, ikinci topun mavi çıkma olasılığı nedir?*

Problem 1.10. *Bir bölümde 84 öğrenci vardır. Öğrencilerden 52'si Ekonometri, 32'si Makine Öğrenmesi dersine kayıtlıdır. 20 öğrenci her iki derse de kayıtlıdır. Tesadüfen seçilen bir öğrencinin Ekonometri veya Makine Öğrenmesi dersine kayıtlı olma olasılığı nedir?*

Problem 1.11. *Bir öğrencinin istatistik dersinden geçme olasılığı 0.8, mikroiktisattan geçme olasılığı 0.7, her ikisinden de geçme olasılığı 0.6 olsun. Bir öğrencinin en azından bir dersten geçme olasılığı nedir?*

Problem 1.12. *Bir torbada 4 Kırmızı ve 2 Sarı top bulunmaktadır. Yerine koyarak 3 top rassal seçiliyor. En az 2 Sarı top gelme olasılığı nedir?*

Problem 1.13. *6 erkek ve 6 kadın arasından 6 kişilik bir komite kurulacaktır.*

a) Yarısının kadınlardan oluşma olasılığı kaçtır?
b) Tamamının erkeklerden oluşma olasılığı kaçtır?

Problem 1.14. *Çoktan seçmeli bir sınavda 20 soru vardır. Her soru 5 şık içermektedir. Bir öğrenci sınava çalışmadığı için soruları rassal olarak cevaplamaya karar vermiştir.*

a) Doğru cevap sayısının 10 olma olasılığı kaçtır?
b) Doğru cevap sayısının 5'ten büyük olma olasılığı kaçtır?

Problem 1.15. *Bir uçağın zamanında kalkma olasılığı 0.85, zamanında varma olasılığı 0.82, zamanında kalkma ve varma olasılığı ise 0.8'dir. Buna göre*

a) Zamanında kalkan bir uçağın, zamanında varma olasılığı nedir?
b) Zamanında varan bir uçağın, zamanında kalkmış olma olasılığı nedir?

Problem 1.16. *8 üründen oluşan bir parti yedek parça içinde 2'si hatalıdır. Bir fabrika bu yedek parçalardan 3'ünü rassal biçimde satın alıyor.*

a) Seçilen ürünlerin üçünün de hatasız olma olasılığı kaçtır?
b) Birinin hatalı olma olasılığı kaçtır?
c) Üçünün de hatalı olma olasılığı kaçtır?

Problem 1.17. *200 üniversite öğrencisiyle yapılan bir ankette aşağıdaki sonuçlara ulaşılmıştır:*

- *150 kişi Instagram kullanıyor*
- *100 kişi Twitter/X kullanıyor*
- *80 kişi TikTok kullanıyor*
- *60 kişi Instagram ve Twitter kullanıyor*
- *40 kişi Instagram ve TikTok kullanıyor*
- *30 kişi Twitter ve TikTok kullanıyor*
- *20 kişi üçünü birden kullanıyor*

Buna göre

a) Bir öğrencinin sadece Instagram kullanma olasılığı kaçtır?
b) Twitter/X kullandığı bilinen birinin Instagram kullanma olasılığı kaçtır?
c) TikTok kullandığı bilinen bir öğrencinin Twitter kullanma olasılığı kaçtır?
d) Hiç sosyal medya kullanmama olasılığı kaçtır?

Problem 1.18. *Bir fabrikada kullanılan makinelerin %5'inde belirli bir arıza meydana gelmektedir. Bu arızayı tespit etmek için kullanılan bir sensörün*

- *Arızalı makinelerde arızayı doğru tespit etme olasılığı %90'dır.*
- *Arızası olmayan makinelerde yanlış alarm verme olasılığı %8'dir.*

Bir makinede sensör arıza tespit ettiğinde, bu makinenin gerçekten arızalı olma olasılığı nedir?

Problem 1.19. *Bir marketteki müşterilerin %60'ı kadın, %40'ı erkektir. Kadın müşterilerin %30'u, erkek müşterilerin %20'si organik ürünler satın almaktadır. Marketten rasgele bir müşteri seçildiğinde, organik ürün satın alma olasılığı nedir? Seçilen müşteri organik ürün satın aldıysa, bu müşterinin kadın olma olasılığı nedir?*

Problem 1.20. *Elinizde bir hilesiz ve bir bir hileli madeni para var. Hileli para, yazı-tura atıldığında %80 olasılıkla tura (T) gelmektedir. Rassal olarak bu paralardan birini seçiyorsunuz ve 3 kez atıyorsunuz. Sonuçların üçü de tura geliyor. Bu bilgilere dayanarak, seçtiğiniz paranın hilesiz olma olasılığı nedir?*

Problem 1.21. *Bir hapishanede, idama mahkum edilmiş 3 kişi vardır: A, B ve C. Hapishane müdürü, bu mahkumlardan yalnızca birinin affedileceğini ve diğer ikisinin idam edileceğini söylemiştir. Mahkumlardan biri olan A, müdüre gidip şu istekte bulunur:*

"Benim B ve C ile eşit şansa sahip olduğumu biliyorum, ama bu iki kişiden en az biri idam edilecek. Beni bilgilendirmek zorunda değilsiniz, ancak B ya da C'den birinin idam edileceğini bana söyleyebilir misiniz?"

Müdür, mahkumlardan hangisinin affedileceğini bilmektedir ve mahkum A'nın isteği üzerine, gerçekten idam edilecek olanlardan birini (örneğin B'yi) açıklar ve onun idam edileceğini söyler. Şimdi, A şunu düşünmeye başlar: B'nin idam edileceğini öğrendiğime göre, benim affedilme olasılığım arttı mı? C'nin idam edilmeyeceği anlamına mı geliyor? Yoksa hala C'nin affedilme olasılığı daha mı yüksek?

Soru: A'nın affedilme olasılığı, müdürün B'nin idam edileceğini açıklamasından sonra nasıl değişir? Müdürün bu bilgiyi vermesinden sonra A'nın affedilme olasılığı hala $1/3$ *mü yoksa artar mı? C'nin affedilme olasılığı nedir?*

Problem 1.22. *Bir yatırım şirketi, iki farklı fonun performansını karşılaştırmak için bir araştırma yapmıştır. Aşağıdaki çapraz tablo, bu fonların yüksek ve düşük getiri kategorilerine göre sınıflandırılmış yatırımcı sayılarını göstermektedir:*

Fon / Getiri	*Yüksek Getiri*	*Düşük Getiri*	*Toplam*
Fon A	*120*	*80*	*200*
Fon B	*90*	*60*	*150*
Toplam	*210*	*140*	*350*

a) A fonuna yatırım yapan birinin yüksek getiri elde etme olasılığı nedir?
b) B fonuna yatırım yapan birinin düşük getiri elde etme olasılığı kaçtır?

Problem 1.23. *Bir yatırım şirketi, üç farklı sektörün performansını karşılaştırmak için bir araştırma yapmıştır. Aşağıdaki çapraz tablo, bu sektörlerin yüksek ve düşük performans kategorilerine göre sınıflandırılmış yatırımcı sayılarını göstermektedir:*

Sektör / Performans	*Yüksek Performans (A)*	*Düşük Performans (B)*	*Toplam*
Teknoloji	*180*	*70*	*250*
Enerji	*120*	*80*	*200*
Sağlık	*100*	*50*	*150*
Toplam	*400*	*200*	*600*

Buna göre

a) Teknoloji ve düşük performanslı yatırım olasılığı kaçtır?
b) Teknoloji sektörüne yatırım yapan birinin yüksek performans elde etme olasılığı kaçtır?

Problem 1.24. *Aşağıdaki tablo, dört farklı sektörün firmalarının yaş kategorilerine göre olasılık oranlarını göstermektedir.*

Sektör	*Yaş <= 3*	*3 < Yaş <= 10*	*Yaş > 10*
İmalat	*0.05*	*0.10*	*0.14*
İnşaat	*0.03*	*0.08*	*0.06*
Ticaret	*0.10*	*0.12*	*0.09*
Hizmet	*0.04*	*0.08*	*0.11*

Bu çapraz tabloya göre

a) Rassal seçilmiş bir firmanın inşaat sektöründe olma olasılı nedir?
b) Rassal seçilmiş bir firmanın yaşının 3 veya daha küçük olma olasılığı kaçtır?
c) Bir firmanın hizmet sektöründe ve 10 yaşından büyük olma olasılığı kaçtır?
d) Bir firmanın ticaret sektöründe olduğu biliniyorsa bu firmanın yaşının 10'dan büyük olma olasılığı nedir?
e) Bir firmanın yaşının 3 ile 10 arasında olduğu biliniyorsa bu firmanın imalat sanayinde olma olasılığı nedir?

Problem 1.25. *Aşağıdaki tablo, hanehalklarının gelir düzeyi ve internet aboneliklerinin hızına ilişkin dağılımını göstermektedir.*

İnternet	*Gelir Düzeyi*			
	Düşük	*Orta*	*Yüksek*	*Toplam*
Hızlı	*180*	*360*	*240*	*780*
Yavaş	*300*	*200*	*100*	*600*
Toplam	*480*	*560*	*340*	*1380*

Buna göre

a) Rassal seçilen bir hanenin yüksek gelirli olma olasılığı kaçtır?

b) Rassal seçilen bir hanenin yavaş internete sahip olma olasılığı kaçtır?
c) Bir hanenin orta gelir düzeyinde ve hızlı internet abonesi olma olasılığı kaçtır?
d) Yüksek gelir düzeyine sahip olduğu bilinen bir hanenin yavaş internetinin olma olasılığı kaçtır?