

Ayrık Olasılık

CSC-2259 Ayrık Yapılar

Konstantin Busch - LSU

1

Ayrık Olasılığa Giriş

Hilesiz zar



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

Olası tüm sonuçlar

2

E olayının olasılığı:

$$p(E) = \frac{\text{olay kümesinin büyüklüğü}}{\text{örnek uzayın büyüklüğü}} = \frac{|E|}{|S|}$$

Not: $0 \leq p(E) \leq 1$

için $0 \leq |E| \leq |S|$

Konstantin Busch - LSU

3



Zarın 3 gelme olasılığı nedir?

Olay uzayı: $E = \{3\}$

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

Olasılık: $p(E) = \frac{|E|}{|S|} = \frac{1}{6}$

4



Zarın 2 veya 5 gelme olasılığı nedir?

Olay uzayı: $E = \{2,5\}$

Örnek uzayı: $S = \{1,2,3,4,5,6\}$

Olasılık:
$$p(E) = \frac{|E|}{|S|} = \frac{2}{6}$$

5

İki adet hilesiz zar



Örnek uzayı: 36 olası sonuç vardır

$$S = \{(1,1), (1,2), (1,3), \dots, (6,6)\}$$

Birinci zar

İkinci zar

Beklenen çift

6



Zarın (1,1) ikilisi gelme olasılığı nedir?

Olay uzayı: $E = \{(1,1)\}$

Örnek uzay: $S = \{(1,1), (1,2), (1,3), \dots, (6,6)\}$

Olasılık:
$$p(E) = \frac{|E|}{|S|} = \frac{1}{36}$$

7



Zarın çift gelme olasılığı nedir?

Olay uzayı: $E = \{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)\}$

Örnek uzay: $S = \{(1,1), (1,2), (1,3), \dots, (6,6)\}$

Olasılık:
$$p(E) = \frac{|E|}{|S|} = \frac{6}{36}$$

8

Düzensiz sayılar oyunu

Oyun yöneticisi 40 sayıdan kazanan 6 sayı kümesini seçer.

Seçim sayıları: 1,2,3,...,40

Örnek kazanan sayılar: 4,7,16,25,33,39

Oyuncunun seçtiği 6 sayının kümesi
(Sıradan bağımsız)

Örnek kazanan sayılar: 8,13,16,23,33,40

Oyuncunun kazanma olasılığı nedir?

Konstantin Busch - LSU

9

Kazanan olay:

$$E = \{\{4,7,16,25,33,39\}\} \quad |E| = 1$$

Kazanan 6 sayının bir tek kümesi

Örnek uzay:

$$S = \{40 \text{ sayıdan kazanan 6 sayı kümesinin tüm alt kümeleri}\} \\ = \{\{1,2,3,4,5,6\}, \{1,2,3,4,5,7\}, \{1,2,3,4,5,8\}, \dots\}$$

$$|S| = \binom{40}{6} = 3,838,380$$

Konstantin Busch - LSU

10

Oyuncunun kazanma olasılığı:

$$P(E) = \frac{|E|}{|S|} = \frac{1}{\binom{40}{6}} = \frac{1}{3,838,380}$$

Kart Oyunları

Destede 52 kart vardır.

13 çeşit kartlar (2,3,4,5,6,7,8,9,10,a,k,q,j),
Her bir türde 4 takım vardır (h,d,c,s)

Oyuncu 4 kart verir.

Oyuncunun kartlarının hepsinin aynı
tür olma olasılığı nedir?

Olay: $E = \{\{2_h, 2_d, 2_c, 2_s\}, \{3_h, 3_d, 3_c, 3_s\}, \dots, \{j_h, j_d, j_c, j_s\}\}$

$|E| = 13$ 4 kartın her bir kümesi
aynı türdedir

Örnek uzay:

$$S = \{52 \text{ kartın olası tüm 4 kartların kümesi}\} \\ = \{\{2_h, 2_d, 2_c, 2_s\}, \{2_h, 2_d, 2_c, 3_h\}, \{2_h, 2_d, 2_c, 3_d\}, \dots\}$$

$$|S| = \binom{52}{4} = \frac{52!}{4!48!} = \frac{52 \cdot 51 \cdot 50 \cdot 49}{4 \cdot 3 \cdot 2} = 270,725$$

Konstantin Busch - LSU

13

Elde aynı türde 4 kart olma olasılığı:

$$P(E) = \frac{|E|}{|S|} = \frac{13}{\binom{52}{4}} = \frac{13}{270,725}$$

Konstantin Busch - LSU

14

Sıralı sayılar oyunu

Oyun yöneticisi bir kutuda 1....50 ile etiketlenmiş toplardan 5 tane seçer

Seçim numaraları: 1,2,3,...,50

Örnek olarak kazanan numaralar: 37,4,16,33,9

Oyuncunun seçtiği 5 sayı kümesi
(Sıralama önemli)

Örnek olarak oyuncunun sayıları: 40,16,13,25,33

Oyuncunun kazanma olasılığı nedir?

Konstantin Busch - LSU

15

Tekrar yerine koyma olmayan örneklem:

Bütün toplar seildikten sonra, top geri kutuya atılmaz

Örnek uzay büyüklüğü:
50 topun 5li permutasyonları

$$|S| = P(50,5) = \frac{50!}{(50-5)!} = \frac{50!}{45!} = 50 \cdot 49 \cdot 48 \cdot 47 \cdot 46 = 245,251,200$$

$$\text{Kazanma olasılığı: } P(E) = \frac{|E|}{|S|} = \frac{1}{245,251,200}$$

Konstantin Busch - LSU

16

Yerine koyarak örnekleme:

Bir top seçildikten sonra,
top tekrar kutuya konulur

Örneklem uzayının büyüklüğü:

Yinelemeyle 50 topun 5 li permutasyonu

$$|S| = 50^5 = 312,500,000$$

Kazanma olasılığı: $P(E) = \frac{|E|}{|S|} = \frac{1}{312,500,000}$

Konstantin Busch - LSU

17

Ters olasılık:

$$p(\bar{E}) = 1 - p(E)$$

İspat: $\bar{E} = S - E$

$$p(\bar{E}) = \frac{|S - E|}{|S|} = \frac{|S| - |E|}{|S|} = 1 - \frac{|E|}{|S|} = 1 - p(E)$$

İspatın sonu

Konstantin Busch - LSU

18

Örnek: Sekiz bitlik ikili dizide en az bir tanesinin 0 olma olasılığı nedir?

$$E = \{01111111, 10111111, \dots, 00111111, \dots, 00000000\}$$

$$\overline{E} = \{11111111\}$$

$$p(E) = 1 - p(\overline{E}) = 1 - \frac{|\overline{E}|}{|S|} = 1 - \frac{1}{2^8}$$

Olasılığın birleşimi: $E_1, E_2 \subseteq S$

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$$

İspat: $|E_1 \cup E_2| = |E_1| + |E_2| - |E_1 \cap E_2|$

$$\begin{aligned} p(E_1 \cup E_2) &= \frac{|E_1 \cup E_2|}{|S|} = \frac{|E_1| + |E_2| - |E_1 \cap E_2|}{|S|} \\ &= \frac{|E_1|}{|S|} + \frac{|E_2|}{|S|} - \frac{|E_1 \cap E_2|}{|S|} \\ &= p(E_1) + p(E_2) - p(E_1 \cap E_2) \end{aligned}$$

Örnek: Sekiz bitlik binary dizinin 0 ile başlaması **veya** 11 ile bitmesi olasılığı nedir?

0 ile başlayan diziler:

$$E_1 = \{00000000, 00000001, \dots, 01111111\}$$

$$|E_1| = 2^7 \quad (\text{all binary strings with 7 bits } 0xxxxxxx)$$

11 ile biten diziler:

$$E_2 = \{00000011, 00000111, \dots, 11111111\}$$

$$|E_2| = 2^6 \quad (\text{all binary strings with 6 bits } xxxxxx11)$$

Konstantin Busch - LSU

21

0 ile başlayan **ve** 11 ile biten diziler:

$$E_1 \cap E_2 = \{000000011, 00000111, \dots, 01111111\}$$

$$|E_1 \cap E_2| = 2^5 \quad (\text{all binary strings with 5 bits } 0xxxxx11)$$

0 ile başlayan **veya** 11 ile biten diziler:

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$$

$$= \frac{|E_1|}{|S|} + \frac{|E_2|}{|S|} - \frac{|E_1 \cap E_2|}{|S|}$$

$$= \frac{2^7}{2^8} + \frac{2^6}{2^8} - \frac{2^5}{2^8} = \frac{1}{2} + \frac{1}{4} - \frac{1}{8} = \frac{5}{8}$$

Konstantin Busch - LSU

22

Olasılık Teorisi

Örnek uzay: $S = \{x_1, x_2, \dots, x_n\}$

Olasılık dağılım fonksiyonu: p

$$0 \leq p(x_i) \leq 1$$

$$\sum_{x=1}^n p(x_i) = 1$$

Olası durumlar: $p(x_i) \neq p(x_j)$

Örnek: Hileli para

2/3 olasılıkla tura (T)

1/3 olasılıkla yazı (Y)



Örnekleme uzayı: $S = \{Y, T\}$

$$p(T) = \frac{2}{3} \quad p(Y) = \frac{1}{3} \quad p(T) + p(Y) = \frac{2}{3} + \frac{1}{3} = 1$$

Düsgün olasılık dağılımı:

$$p(x_i) = \frac{1}{n}$$

Örnek uzay: $S = \{x_1, x_2, \dots, x_n\}$

Örnek: Hilesiz para

1/2 olasılıkla Tura (T) veya Yazı (Y)

$$S = \{H, T\} \quad p(H) = \frac{1}{2} \quad p(T) = \frac{1}{2}$$

Konstantin Busch - LSU

25

Olayın olasılığı: E

$$E = \{x_1, x_2, \dots, x_k\} \subseteq S$$

$$p(E) = \sum_{i=1}^k p(x_i)$$

Düsgün olasılık dağılımı için: $p(E) = \frac{|E|}{|S|}$

Konstantin Busch - LSU

26

Örnek: Hileli zar $S = \{1,2,3,4,5,6,6\}$

$$p(1) = p(2) = p(3) = p(4) = p(5) = \frac{1}{7} \quad p(6) = \frac{2}{7}$$

Zarın 2 veya 6 gelme olasılığı nedir?

$$E = \{2,6\}$$

$$p(E) = p(2) + p(6) = \frac{1}{7} + \frac{2}{7} = \frac{3}{7}$$

Konstantin Busch - LSU

27

Olayların kombinasyonları:

Tümleyin: $p(\overline{E}) = 1 - p(E)$

Birleşim:

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$$

Ayrı olayların birleşimi: $p\left(\bigcup_i E_i\right) = \sum_i p(E_i)$

Konstantin Busch - LSU

28

Şartlı Olasılık

Hilesiz üç paranın atılışı



Yazı



Tura



Yazı

Şart: İlk para yazı olsun

Soru: İlk paranın yazı olduğunu düşünerek, tek sayı atışlarının yazı olma olasılığı nedir?

Konstantin Busch - LSU

29

Örneklem uzayı:

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

Verilen koşuldaki sınırlı örneklem:

$$F = \{YYY, YYT, YTY, YTT\}$$

Birinci para Yazı

Konstantin Busch - LSU

30

Şart olmayan olaylar:

$$E = \{YYY, YTT, TYT, TTY\}$$

Yazıların tek sayıları için

Şartlı olaylar:

$$E_F = E \cap F = \{YYY, YTT\}$$

Birinci para yazı

$$E = \{YYY, YTT, TYT, TTY\}$$

$$E_F = E \cap F = \{YYY, YTT\}$$

Verilen şartlarda,
Örnek uzaya F 'nin değişimi

$$p(E_F) = \frac{|E \cap F|}{|F|} = \frac{|E \cap F| / |S|}{|F| / |S|} = \frac{p(E \cap F)}{p(F)} = \frac{2/8}{4/8} = 0.5$$

(para hilesiz

Şart ile olayın belirtimi:

$$E_F = E | F$$

E olayı F 'yi belirtir.

$$p(E_F) = p(E | F) = \frac{p(E \cap F)}{p(F)}$$

Şartlı olasılığın tanımı:

(rasgele olasılık dağılımı için)

E ve F olaylarıyla S örneklem uzayı verilsin
(burada $p(F) > 0$);

F ile verilen E 'nin şartlı olasılığı:

$$p(E | F) = \frac{p(E \cap F)}{p(F)}$$

Örnek: İki çocuğun olan bir ailenin bir çocuğu erkek ise iki tanesinin erkek olma ihtimali nedir?

Erkek veya kızın eşit olasılıkta olacağını varsayalım

Örnek Uzay: $S = \{EE, EK, KE, KK\}$

Şart: $F = \{EE, EK, KE\}$
Bir çocuk erkektir

Konstantin Busch - LSU

35

Durum: $E = \{EE\}$
Her iki çocuk da erkek

Durumun koşullu olasılığı:

$$p(E | F) = \frac{p(E \cap F)}{p(F)} = \frac{p(\{EE\})}{p(\{EE, EK, KE\})} = \frac{1/4}{3/4} = \frac{1}{3}$$

Konstantin Busch - LSU

36

Bağımsız Durumlar

Durumlar E_1 ve E_2 ancak ve ancak bağımsız ise:

$$p(E_1 \cap E_2) = p(E_1)p(E_2)$$

Eş değer tanım (Eğer $p(E_2) > 0$):

$$p(E_1 | E_2) = p(E_1)$$

Örnek: 4 bitlik düzgün rasgele katarlar:

E_1 : 1 ile başlayan katar

E_2 : çift 1 içeren katar

$$E_1 = \{1000, 1001, 1010, 1011, 1100, 1101, 1110, 1111\}$$

$$E_2 = \{0000, 0011, 0101, 0110, 1001, 1010, 1100, 1111\}$$

$$E_1 \cap E_2 = \{1111, 1100, 1010, 1001\}$$

$$|E_1| = |E_2| = 8 \quad p(E_1) = p(E_2) = \frac{8}{16} = \frac{1}{2}$$

$$|E_1 \cap E_2| = 4 \quad p(E_1 \cap E_2) = \frac{4}{16} = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = p(E_1)p(E_2)$$

Durum E_1 ve E_2 bağımsızdır

Bernoulli denemesi:

İki sonuçlu deneyim:

Başarı veya başarısızlık

Başarı olasılığı: p

Başarısızlık olasılığı: $q = 1 - p$

Örnek: Hileli para

Başarı = Tura

$$p = p(T) = \frac{2}{3}$$



Başarısızlık = Yazı

$$q = p(Y) = \frac{1}{3}$$

Bağımsız Bernoulli Denemeler:

Ardışık Bernoulli denemelerinin sonuçları birbirlerine bağlı değildir

Örnek: Ardışık madeni paralar

Hileli parayı 5 kere atın



3 tura yapma ihtimali nedir?

Tura olasılığı: $p = \frac{2}{3}$ (başarı)

Yazı olasılığı: $q = \frac{1}{3}$ (başarısız)

Konstantin Busch - LSU

41



⋮

3'ü tura olan 5 bozuk para
silsilesindeki sıralamak için
yolların toplam sayıları:

TTTY

TYTTY

TYTYT

YTTYT

⋮

$$\binom{5}{3}$$

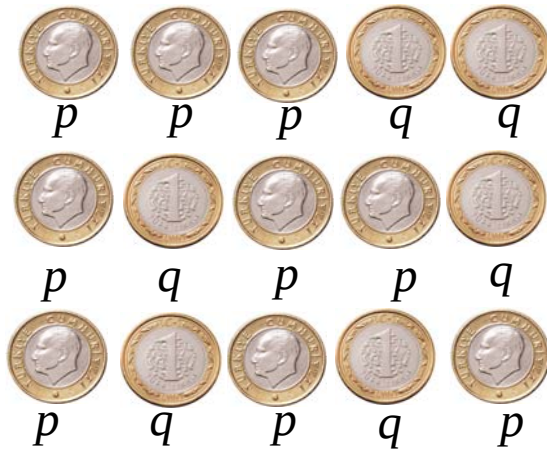
Konstantin Busch - LSU

42

Her habgi bir silsiledeki 3 tura ve 2 yazı durumundaki olasılıklar:

$$p^3 q^2$$

Örnek olarak:



TTTTYY

$$pppqq = p^3 q^2$$

TYTTTY

$$pqppq = p^3 q^2$$

TYTYT

$$pqrpq = p^3 q^2$$

Konstantin Busch - LSU

43

3 tura olma olasılığı:

$$p^3 q^2 + p^3 q^2 + \cdots + p^3 q^2 = \binom{5}{3} p^3 q^2$$

1^{nci}

Dizi başarısı
(3 Tura)

2^{nci}

Dizi başarısı
(3 Tura)

$\binom{5}{3}$ nci

Dizi başarısı
(3 Tura)

Konstantin Busch - LSU

44

Hileli bozuk paranın 5 kez atılması:



Tam olarak 3 tura olma ihtimali:

$$\binom{5}{3} p^3 q^2 = \frac{5!}{3!2!} \left(\frac{2}{3}\right)^3 \left(\frac{1}{3}\right)^2 \approx 0.0086$$

Belirtilen silsileyerleşiminde 3 tura ve 2 yazı elde etme olasılığı

Silsilenin 5 paranın 3'nün tura olarak düzenlemek için mümkün olan tüm yollar

Konstantin Busch - LSU

45

Teorem: n bağımsız Bernoulli denemesinde k başarı elde etme olasılığı:

$$\binom{n}{k} p^k q^{n-k}$$

Ayrıca şöyle bilinir
binomial olasılık dağılımı:

$$b(k; n, p) = \binom{n}{k} p^k q^{n-k}$$

Konstantin Busch - LSU

46

İspat:

$$\binom{n}{k} p^k q^{n-k}$$

k başarıları ve $n-k$
başarısızlığıyla
toplam dizilim sayısı

Bir dizilimin,
belirtilen noktalarda
 k başarı ve $n-k$
başarısızlığı ihtimali

Örnek:

BNBNNB...BBN

İspat sonu

Konstantin Busch - LSU

47

Örnek:

Rasgele aynı ikili katarlar

0 bit için olasılık 0.9

1 bitlik olasılık 0.1

10 bitten 8 bit 0'ın olasılığı nedir?

0100001000 gibi

$$p = 0.9 \quad q = 0.1 \quad k = 8 \quad n = 10$$

$$b(k; n, p) = \binom{n}{k} p^k q^{n-k} = \binom{10}{8} (0.9)^8 (0.1)^2 = 0.1937102445$$

Konstantin Busch - LSU

48

Doğum günü problemi

Doğum günü çatışması: Aynı gün içinde iki kişinin doğum günü var

Problem:

Doğum günü çatışması ihtimalinin en az $\frac{1}{2}$ olması için kaç kişi odada olmalıdır?

Varsayım: her hengi bir günde doğma olasılığı eşittir

Bir yılda 366 gün var

Eğer insanların sayısı 367 veya daha fazla ise doğum günü çatışması güvercin yuvası prensibiyle onaylanır

İnsanları $n \leq 366$ olduğunu varsayalım

Hesaplamalarda;

p_n : n insanın hepsinin farklı doğum gününe sahip olma olasılığı

Yardımcı olası için;

$1 - p_n$: n insan arasında doğum günü çartışması olasılığı

Konstantin Busch - LSU

51

Örnek uzay: Kartezyen çarpım

$$S = \{1,2,\dots,366\} \times \{1,2,\dots,366\} \times \dots \times \{1,2,\dots,366\}$$

1^{inci} kişinin doğum
günü seçimleri

2^{inci} kişinin doğum
günü seçimleri

n ^{inci} kişinin
doğum günü
seçimleri

$$S = \{(1,1,\dots,1), (2,1,\dots,1), \dots, (366,366,\dots,366)\}$$

Örnek uzay büyüklüğü: $|S| = 366 \cdot 366 \cdot \dots \cdot 366 = 366^n$

Konstantin Busch - LSU

52

Olay kümesi: Her birinin doğum günü farklı

$$E = \{(1,2,\dots,366), (366,1,\dots,365), \dots, (366,365,\dots,1)\}$$

1inci kişinin
doğum
günü

2inci kişinin
doğum günü

ninci kişinin
doğum günü

#seçim

#seçim

#seçim

Örnek büyüklüğü:

$$|E| = P(366, n) = \frac{366!}{(366-n)!} = 366 \cdot 365 \cdot 364 \cdots (366-n+1)$$

Doğum günü çatışmasının olmama olasılığı

$$p_n = \frac{|E|}{|S|} = \frac{366 \cdot 365 \cdot 364 \cdots (366-n+1)}{366^n}$$

Doğum günü çatışması olasılığı: $1 - p_n$

$$n = 22 \quad 1 - p_n \approx 0.475$$

$$n = 23 \quad 1 - p_n \approx 0.506$$

Doğum günü çatışması olasılığı: $1 - p_n$

$$n = 23 \quad 1 - p_n \approx 0.506$$

Bu yüzden: $n \geq 23$ kişinin doğum çatışmasının en az $\frac{1}{2}$ olma ihtimali

Doğum günü problemi analizlerinde çatışmaları en aza indirmek için uygun Hash tablo boyutu tanımlamaları kullanılır

Hash fonksiyon çatışması: $h(k_1) = h(k_2)$

Randomize algoritması:

seçimlerin rasgele olduğu algoritmalar
(Örnek: quicksort)

Las Vegas algoritması:

Çıktıları her zaman doğru olan randomize algoritmalar. (quicksort gibi)

Monte Carlo algoritması:

Çıktıları bazen doğru olan randomize algoritmalar (yanlış çıktı üretebilir)

Konstantin Busch - LSU

57

A Monte Carlo algoritması

```
Primality_Test( $n, k$ ) {  
  for( $i = 1$  to  $k$ ) {  
     $b \leftarrow \text{random\_number}(1, \dots, n)$   
    if (Miller_Test( $n, b$ ) == failure)  
      return(false) //  $n$  is not prime  
  }  
  return(true) // most likely  $n$  is prime  
}
```

Konstantin Busch - LSU

58

```

Miller_Test( n,b) {
     $n-1 = 2^s t$ 
     $s, t \geq 0, \quad s \leq \log n, \quad t$  is odd
    for (  $j = 0$  to  $s-1$  ) {
        if (  $b^t \equiv 1 \pmod{n}$  or  $b^{2^j t} \equiv -1 \pmod{n}$  )
            return(success)
        }
    return(failure)
}

```

Konstantin Busch - LSU

59

n asal sayında her $1 \leq b \leq n$ için Miller testini geçer

n birleşin sayıda $\frac{n}{4}$ sayıdan daha az için

$1 \leq b \leq n$ aralığında Miller testini geçer

Olasılık olarak yanlış pozitif: $\frac{1}{4}$

Konstantin Busch - LSU

60

Eğer asallık test algoritması **yanlış** ise sayı kesin olarak asal değildir.

Eğer algoritma çalıştırılması **doğru** ise cevap yüksek olasılıkla doğrudur (sayı asal):

$$1 - \left(\frac{1}{4}\right)^k = 1 - \frac{1}{n} \approx 1$$

$$k = \log_4 n \text{ için}$$

Konstantin Busch - LSU

61

Bayes Teorem

$$p(E) \neq 0 \quad p(F) \neq 0$$

$$p(F | E) = \frac{p(E | F)p(F)}{p(E | F)p(F) + p(E | \overline{F})p(\overline{F})}$$

Uygulamalar: Makine öğrenmesi
Spam Filtreleri

Konstantin Busch - LSU

62

Bayes Teorem İspatı:

$$\begin{array}{l}
 \left. \begin{array}{l} p(F | E) = \frac{p(E \cap F)}{p(E)} \\ p(E | F) = \frac{p(E \cap F)}{p(F)} \end{array} \right\} \xrightarrow{\text{yellow arrow}} \begin{array}{l} p(E \cap F) = p(F | E)p(E) \\ p(E \cap F) = p(E | F)p(F) \end{array} \\
 \downarrow \text{yellow arrow} \\
 p(F | E)p(E) = p(E | F)p(F) \\
 \downarrow \text{yellow arrow} \\
 p(F | E) = \frac{p(E | F)p(F)}{p(E)}
 \end{array}$$

Konstantin Busch - LSU

63

$$\begin{array}{l}
 \left. \begin{array}{l} E = (E \cap F) \cup (E \cap \bar{F}) \\ (E \cap F) \cap (E \cap \bar{F}) = \emptyset \end{array} \right\} \xrightarrow{\text{yellow arrow}} p(E) = p(E \cap F) + p(E \cap \bar{F}) \\
 \hline
 \begin{array}{l} p(E \cap F) = p(E | F)p(F) \\ p(E \cap \bar{F}) = p(E | \bar{F})p(\bar{F}) \end{array} \\
 \downarrow \text{yellow arrow} \\
 p(E) = p(E | F)p(F) + p(E | \bar{F})p(\bar{F})
 \end{array}$$

Konstantin Busch - LSU

64

$$p(F | E) = \frac{p(E | F)p(F)}{p(E)}$$

$$p(E) = p(E | F)p(F) + p(E | \bar{F})p(\bar{F})$$



$$p(F | E) = \frac{p(E | F)p(F)}{p(E | F)p(F) + p(E | \bar{F})p(\bar{F})}$$

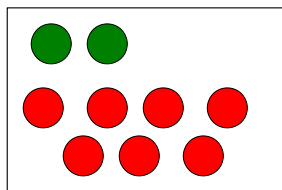
İspatın sonu

Konstantin Busch - LSU

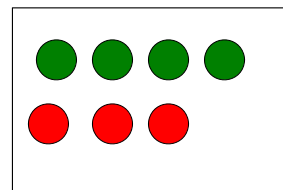
65

Örnek: Rasgele kutuyu seçtikten sonra kutudan rasgele topu seçin.

Kutu 1



Kutu 2



Soru:

Eğer bir kırmızı top seçilmişse, bunun Kutu 1'den olma olasılığı nedir?

Konstantin Busch - LSU

66

E : Kırmızı top seçimi

F : Kutu 1 seçimi

$\bar{E}$: Yeşil top seçimi

$\bar{F}$: Kutu 2 seçimi

Soru olasılığı: $P(F | E) = ?$

Soru:

Eğer bir kırmızı top seçilmişse, bunun Kutu 1'den olma olasılığı nedir?

Konstantin Busch - LSU

67

Bayes Teoremi:

$$p(F | E) = \frac{p(E | F)p(F)}{p(E | F)p(F) + p(E | \bar{F})p(\bar{F})}$$

Hesaplamak için tek ihtiyacımız:

$$p(F) \quad p(\bar{F}) \quad p(E | F) \quad p(E | \bar{F})$$

Konstantin Busch - LSU

68

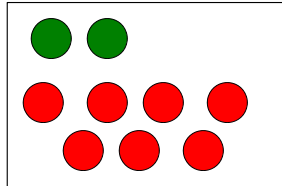
E : Kırmızı top seçimi

F : Kutu 1 seçimi

$\bar{E}$: Yeşil top seçimi

$\bar{F}$: Kutu 2 seçimi

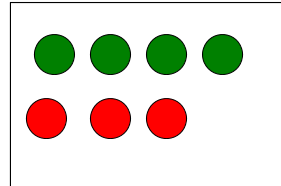
Kutu 1



$$p(F) = 1/2 = 0.5$$

Kutu 1 seçme olasılığı

Kutu 2



$$p(\bar{F}) = 1/2 = 0.5$$

Kutu 2 seçme olasılığı

Konstantin Busch - LSU

69

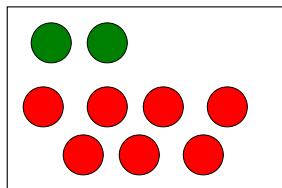
E : Kırmızı top seçimi

F : Kutu 1 seçimi

$\bar{E}$: Yeşil top seçimi

$\bar{F}$: Kutu 2 seçimi

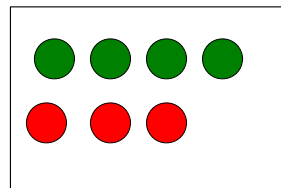
Kutu 1



$$p(E | F) = 7/9 = 0.777...$$

Kutu 1'den kırmızı top
seçme olasılığı

Kutu 2



$$p(E | \bar{F}) = 3/7 = 0.428....$$

Kutu 2'den kırmızı top
seçme olasılığı

Konstantin Busch - LSU

70

$$p(F | E) = \frac{p(E | F)p(F)}{p(E | F)p(F) + p(E | \bar{F})p(\bar{F})}$$

$$p(F) = 1/2 = 0.5 \quad p(\bar{F}) = 1/2 = 0.5$$

$$p(E | F) = 7/9 = 0.777... \quad p(E | \bar{F}) = 3/7 = 0.428....$$

$$p(F | E) = \frac{0.777 \times 0.5}{0.777 \times 0.5 + 0.428 \times 0.5} = \frac{0.777}{0.777 + 0.428} = 0.644$$

Son sonuç

Ya daha fazla kutuya sahip olsaydık?

Genelleştirilmiş Bayes Theorem:

$$p(F_j | E) = \frac{p(E | F_j)p(F_j)}{\sum_{i=1}^n p(E | F_i)p(F_i)}$$

Örnek uzay: $S = F_1 \cup F_2 \cup \dots \cup F_n$

Karşılıklı özel olaylar

Spam Filtereleri

Eğitim kümesi: Spam (kötü) epostalar B

İyi epostalar G

Bir kullanıcı eğitim kümesindeki her bir epostaları iyi veya kötü olarak sınıflandırır.

Konstantin Busch - LSU

73

G ve B 'den oluşan kelimeleri bulun.

$$n_B(w)$$

w sözcüğü içeren spam epostaların sayısı

$$n_G(w)$$

w sözcüğü içeren iyi epostaların sayısı

$$p(w) = \frac{n_B(w)}{|B|}$$

w sözcüğü içeren spam epostaların olasılığı

$$q(w) = \frac{n_G(w)}{|G|}$$

w sözcüğü içeren iyi epostaların olasılığı

Konstantin Busch - LSU

74

Yeni bir X eposta geldiğinde

S : X spam olması durum

E : X 'in w içermesi durumu

w sözcüğü içeren X 'in spam olma olasılığı nedir?

$$P(S | E) = ?$$

Eğer bu olasılık **0.9** den az ise ret edilir.

Konstantin Busch - LSU

75

$$p(S | E) = \frac{p(E | S)p(S)}{p(E | S)p(S) + p(E | \bar{S})p(\bar{S})}$$

Hesaplamak için ihtiyacımız olan:

$p(S)$	$p(\bar{S})$	$p(E S)$	$p(E \bar{S})$
0.5	0.5	$p(w) = \frac{n_B(w)}{ B }$	$q(w) = \frac{n_G(w)}{ G }$
Basitleştirilmiş varsayım		Eğitim kümesinden hesaplanan	

Konstantin Busch - LSU

76

Örnek:

"Rolex" kelimesi için eğitim kümesi:

"Rolex" 2000 spam e-postalardan
250'sinde gerçekleşir.

"Rolex" 1000 iyi e-postanın 5'inde görülür.

Eğer yeni e-posta "Rolex" kelimesini
içeriyorsa, bunun spam olma olasılığı nedir?

Konstantin Busch - LSU

77

"Rolex" 2000 spam e-postalardan
250'sinde gerçekleşir.

$$n_B(\text{Rolex}) = 250$$

$$p(\text{Rolex}) = \frac{n_B(\text{Rolex})}{|B|} = \frac{250}{2000} = 0.125$$

"Rolex" 1000 iyi e-postanın 5'inde görülür.

$$n_G(\text{Rolex}) = 5$$

$$q(\text{Rolex}) = \frac{n_G(\text{Rolex})}{|G|} = \frac{5}{1000} = 0.005$$

Konstantin Busch - LSU

78

Eğer yeni e-posta "Rolex" kelimesini içeriyorsa, bunun spam olma olasılığı nedir?

S : X 'in spam durum

E : "Rolex" kelimesi içeren X durumu

$$P(S | E) = ?$$

$$p(S | E) = \frac{p(E | S)p(S)}{p(E | S)p(S) + p(E | \bar{S})p(\bar{S})}$$

Hesaplamak için olan ihtiyacımız

$p(S)$	$p(\bar{S})$	$p(E S)$	$p(E \bar{S})$
0.5	0.5	$p(\text{Rolex}) = 0.125$	$q(\text{Rolex}) = 0.005$
Basitleştirilmiş varsayım		Eğitim kümesinde hesaplanan	

$$p(S | E) = \frac{0.125 \cdot 0.5}{0.125 \cdot 0.5 + 0.005 \cdot 0.5} = \frac{0.125}{0.13} = 0.961...$$

Yeni e-postalar spam olarak kabul edilir.
Çünkü:

$$p(S | E) = 0.961 > 0.9 \text{ spam eşiği}$$

Daha iyi spam filtreleri için iki kelime kullanır:

$$p(S | E_1 \cap E_2) = \frac{p(E_1 | S)p(E_2 | S)}{p(E_1 | S)p(E_2 | S) + p(E_1 | \bar{S})p(E_2 | \bar{S})}$$

Varsayım: E_1 ve E_2 bağımsızdır

İki kelime bir birinden
bağımsız görünür