KONU / SUBJECT: OLASILIK VE İSTATİSTİK

TARİH / DATE: SAAT / TIME:

GÖKHAN ÇETİNEL → 6402

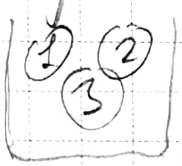
— Deterministik Deneyler → Çıkışı önceden bilinen deneyler
— İstatiksel Deneyler → Çıkışı önceden bilinmediği (F=no) rastlantısal sonuçlar veren deneylerdir.

S: Örnek Uzay → Bir deneyin tüm çıkışlarının oluşturduğu kümedir. S ile gösterilir.

Çıkış → Rastlantı deneyinin ayrıştırılamaz mümkün tüm sonuçlarından her birine çıkış denir.

Rastlantı Deneyi → Gözlemlenen çıkışların rastgele değiştiği yani çıkışların önceden kestirilemediği deneylerdir.

*Deney Tanımlayalım;



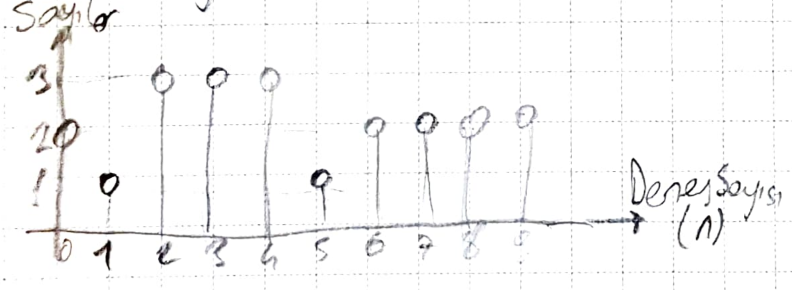
$S = \{1, 2, 3\}$ Çıkış

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

Çıkış değil (Alt küme)

İstatiksel Düzenlilik:

Mühendislikte çoğu olasılık modeli rastlantı deneylerinin çok sayıda tekrarlanmasından elde edilen ortalamaların yaklaşık olarak aynı değeri verdiği gerçeğine dayanır. Bu özelliğe istatiksel düzenlilik denir.



(Frekanslar)

$\dfrac{\text{İstenilen çıkışlar}}{\text{tüm çıkışlar}}$

2 → 0,5
3 → 0,3
1 → 0,2

1

Varsa deneyinin aynı koşullar altında "n" kez tekrarlandığını varsayalım. Çıkışın 0, 1 ve 2 olma sayısı sırası ile $N_0(n)$, $N_1(n)$ ve $N_2(n)$ ile gösterilsin. k çıkışının bağıl frekansı $f_k(n)$

$f_0(n) = \dfrac{N_0(n)}{n}$

$f_1(n) = \dfrac{N_1(n)}{n}$

$f_2(n)$

$$\boxed{f_k(n) = \frac{N_k(n)}{n}}$$

$N_0(n) \rightarrow$ sıfır olma sayısı

www.teksanjenerator.com.tr

İstatistiksel düzenlilik n çok büyük değerler aldığında bağıl frekansın sabit bir değer etrafında çok az değişeceği $\boxed{\lim_{n \to \infty} f_k(n) = \underbrace{P_k}_{\text{olasılık}} \text{ (sabit)}}$ [P_k sabitine k çıkışının olasılığı denir.]

Bağıl Frekansın Özellikleri

1- Bir rastlantı deneyinin k çıkışı olduğunu varsayalım. $S=\{1,2,\dots,K\}$ n denemede herhangi bir çıkışın gözükme sayısı $\boxed{0 \leq N_k(n) \leq n}$ $k=1,2,3,\dots,K$

Eşitsizliğin her tarafını "n" e bölersek.

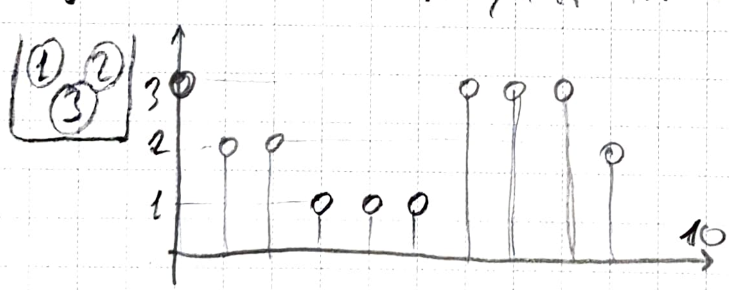
$$0 \leq \frac{N_k(n)}{n} \leq 1 \qquad \boxed{0 \leq f_k(n) \leq 1} \quad k=1,2,3,\dots,K$$

2- Tüm çıkışların gözükme sayısının toplamı $\sum_{k=1}^{K} N_k(n) = n$

Eşitliğin iki tarafını "n"e bölelim $\sum_{k=1}^{K} \frac{N_k(n)}{n} = 1$

$$\boxed{\sum_{k=1}^{K} f_k(n) = 1}$$

3- Vazo deneyinde "üzerinde çift sayı yazan top çekilir" olayını ele alalım. $A=\{\text{çift numaralı top çek}\}$



$\frac{\text{favoriler}}{\text{Tüm çıkış}}$

$P_1 + P_3 = \frac{3}{10} + \frac{4}{10}$

$$\boxed{f_A(n) = \frac{N_2(n)}{n} + \frac{N_4(n)}{n} + \dots + \frac{N_{2n}(n)}{n}}$$ ile hesaplanır.

[Seçilen top tekrar yerine konuyor]

Olasılık Modelleri

Modern teoriye göre olasılık teorisi aşağıdaki modelleri yapar.

1- Başlangıç deneyi tanımlanmış ve tüm çıkışları içeren örnek uzayı S belirlenmiştir.

2- Olaylar yani S'in alt kümeleri belirlenmiştir.

3- Her bir A olayına aşağıdaki aksiyomları sağlayacak bir sayı değeri P[A] atanmıştır.

$$\boxed{\begin{array}{l} 0 \le P[A] \le 1 \\ P[S] = 1 \\ A \cap B = \emptyset \text{ ise } P[A \text{ veya } B] = P[A] + P[B] \end{array}}$$

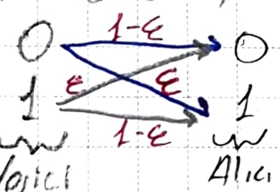
Ayrık Olay

modern olasılık aksiyomları

veya = birleşim
ve = kesişim

⇒ Güvensiz kanallar üzerinden haberleşme

İkili haberleşme sistemi (ε = Hata olasılığı) $\varepsilon = 10^{-3}$ olsun



Gönderilen Bit {0,1} → Hata Düzelten Kodlayıcı → Haberleşme Kanalı → Hata Düzelten Kod Çözücü → Alınan Bit

0

[Kodlayıcı Hata oranına bağlı olarak 3 kere gönderiyor]

000	→	0 ✓	
001	→	0 ✓	
010	→	0 ✓	
011	→	1 ✗	→ $(1-\varepsilon).\varepsilon.\varepsilon$
100	→	0 ✓	
101	→	1 ✗	→ $\varepsilon.(1-\varepsilon).\varepsilon$
110	→	1 ✗	→ $\varepsilon.\varepsilon.(1-\varepsilon)$
111	→	1 ✗	→ $\varepsilon.\varepsilon.\varepsilon$

[Çoğunluğa bakılır. Sıfır gönderiliyor]

$$P(\text{Hata}) = 3\varepsilon^2(1-\varepsilon) + \varepsilon^3 = 3\varepsilon^2 - 2\varepsilon^3$$

$$P(\text{Hata}) = 3 \cdot 10^{-6} - 2 \cdot 10^{-9}$$

↓

[Hata payı 10^{-3} ten 10^{-6} ya kadar düştü.]

KONU / SUBJECT: Cevapların yanına açıklama yaz

TARİH / DATE:

SAAT / TIME:

TEKSAN

NOT= Hata tespiti ve düzeltmesi, kablosuz ve diğer gürültülü kanallar üzerinden güvenli haberleşmeyi mümkün kılar. Olasılık teorisi oluşabilecek hataları belirlemede kullanılır.

Küme Gösterimleri

Tanım: (Tümleyen) A olayının tümleyeni, A olayında bulunmayan çıkışların kümesidir. Ve A' ile gösterilir. A olayının meydana gelme olasılığı p ise A' olayının meydana gelme olasılığı 1-p dir.

$$\boxed{P[A'] = 1 - P[A]}$$

⇒ Bir zar atıldığında 3 veya 5 gelmeme olasılığını hesaplayınız?

A={3 ve ya 5 gelme} = $P[3 \cup 5] = P[3] + P[5] = \frac{2}{6}$

$S = \{1, 2, 3, \ldots, 6\}$

$P[A'] = 1 - \frac{2}{6} = \frac{4}{6}$

Tanım: Kümelerin Kesişimi $A \cap B$ olayı A ve B kümelerinin her ikisinde de bulunan çıkışlardan oluşur. $P[A \cap B] = P[AB]$

$A \cap B = \emptyset$ ise olaylara ayrık olaylar denir.

De Morgan Kuralları

$(A \cup B)' = A'B'$

$(AB)' = A' \cup B'$

⇒ Bir fabrikada kontrol edilen nesnelerden %40'ı mekanik arızalı, %50 si elektrik arızalı, %25'i iki arızaya sahip olsun aşağıdaki olasılıkları bulunuz?

A={mekanik arızalı}

B={elektrik arızalı}

a) $P[A] = ?$ 0,4

b) $P[AB] = ?$ 0,25

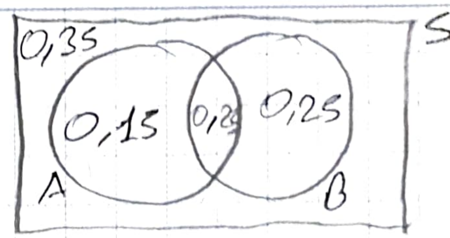
c) $P[A'B] = ?$ 0,25

d) $P[A'B'] = ?$ 0,35 $= P[(A \cup B)']$

e) $P[A' \cup B'] = ?$ $= P[(AB)'] = 1 - P[AB] = 0{,}75$

$A \to \%40 = 0,4$

$B \to \%50 = 0,5$



⇒

	B	B'	
A	0,33	? 0,15	? 0,48
A'	? 0,31	? 0,21	0,52
	0,64	? 0,36	1

A ve B olaylarına ilişkin iki yönlü tablo şekilde verilmiştir.

a) Tüm soru işaretlerini bulun.

b) $A \cup B$ olasılığını hesaplayın

1. sorun için; $(A \cap B) \cup (A' \cap B)$

$(A \cup A') \cap B$

B

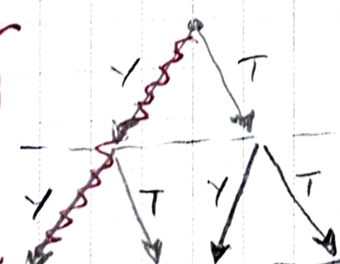
b) $P(A \cup B) = P(A) + P(B) - P(AB)$

$0,48 + 0,64 - 0,33 = 0,79$

Ağaç Diyagramı

Ağaç diyagramı ard arda meydana gelen olayların çıkışlarının tanımlanmasında kullanılır. Ayrıca koşullu olasılıkların hesaplanmasında büyük kolaylık sağlar.

Olasılıklar Çarpılır



$P(Y_1 Y_2)$ $P(Y_1 T_2)$ $P(T_1 Y_2)$ $P(T_1 T_2)$

(+) toplanır

İkinci atışta tura gelme olasılığı.

⇒ Yerine Koyma $\boxed{\begin{matrix}3S\\5K\end{matrix}}$

$P[S] = 3/8$ S   K $P[K] = 5/8$

$P[S] = \frac{2}{7}$ $P[K] = \frac{5}{7}$ $\frac{3}{7}$ $\frac{4}{7}$

$P[K_1K_2] = \frac{5}{8}\,\frac{4}{7}$

$P[S_1S_2] = \frac{3}{8}\cdot\frac{2}{7}$ $P[S_1K_2] = \frac{3}{8}\,\frac{5}{7}$ $P[K_1S_2] = \frac{5}{8}\,\frac{3}{7}$

$P[\text{2 topta kırmızı}] = P[S_1K_2] + P[K_1K_2] = \frac{3}{8}\cdot\frac{5}{7} + \frac{5}{8}\frac{4}{7} = \frac{35}{56}$

Olasılığın Türleri

1- Teorik (klasik) Olasılık

Teorik olasılık örnek uzaydaki çıkışların olasılıkları eşit olduğunda kullanılır. Buna göre A olayının olasılığı

$$P[A] = \frac{\text{A olayının eleman sayısı}}{\text{örnek uzayın eleman sayısı}}$$

2- Deneysel (istatiksel) Olasılık

Deneysel olasılık gözlemlenen veriye bağlıdır. A olayının olasılığı bağıl frekansına eşittir.

$$P[A] = \frac{\text{A olayının frekansı}}{\text{Toplam gözlem sayısı}}$$

3- Öznel Olasılık

Deneyi yapan kişinin önceki tecrübelerine bağlı olarak yaptığı fikir.

⇒ İki zar aynı anda atıldığında toplamın 7 ve ya 11 gelme olasılığını hesaplayınız?

$P(7 \vee 11) = ?$

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)				(2,5)	
3	(3,1)			(3,4)		
4	(4,1)		(4,3)			
5	(5,1)	(5,2)				
6	(6,1)					(6,6)
7						

$P(7) = \frac{6}{36}$

$P(11) = \frac{2}{36}$

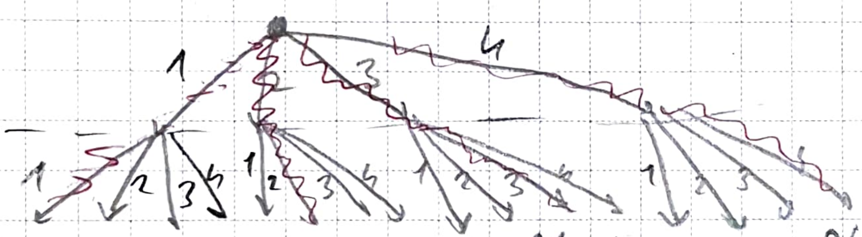
$P(7 \cup 11) = P(7) + P(11) - \cancel{P(7 \cap 11)} = \frac{8}{36}$

⇒ 4 yüzeyli simetrik bir zar peş peşe iki kez atılıyor.

a) İki zarda da aynı sayı gelme olasılığı

b) Her bir zarda 3ten küçük sayı gelme olasılığı

c) İki zar arasındaki farkın en çok 1 olma olasılığı



a) $P(1,1) \cdot P(2,2) \cdot P(3,3) \cdot P(4,4)$

b) $P(1,2)$

c) $P(\text{Fark}) = P(\text{Fark}=0) + P(\text{Fark}=1)$

12 123 234 34

ÖDEV: İki zar aynı anda atılsaydı cevaplar nasıl değişir

Olasılıkta Sayma Kuralları

Bazı deneylerde ağaç diyagramı yoluyla örnek uzayın elemanlarını sistematik olarak listelemek yararlıdır. Örnek uzaydaki nokta sayısı çok fazla olduğunda da sayma kurallarını kullanarak tek tek listelemeden soruları çözebiliriz.

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

TEKSAN

Teorem (Çarpım İlkesi) Bir işlem n_1 farklı şekilde gerçekleşiyorsa ve bu n_1 işlemin her biri n_2 farklı şekilde gerçekleşiyorsa iki işlem $n_1 \cdot n_2$ farklı şekilde gerçekleşebilir.

$n_1 \cdot n_2 \cdot n_3 \cdot \ldots \cdot n_k$ = örnek uzayın eleman sayısı

Teorem (Permütasyon) Permütasyon terimi sıra önemli olduğunda nesnelerin dizilişi ile ilgilenir. n farklı nesneden r adet seçilme sayısı

$$_rP_n = \frac{n!}{(n-r)!}$$ ile hesaplanır.

Teorem (Kombinasyon) Kombinasyon terimi sıra önemli olmadığında nesnelerin dizilişleriyle ilgilenir. n farklı nesneden r boyutlu alt küme oluşturma sayısı

$$\binom{n}{r} = \frac{n!}{(n-r)! \, r!}$$ ile hesaplanır.

r! → elemanların kendi arasında ~~farklı~~ dizilişleri

		II. ZAR		
		1	2	3
I. ZAR	1	(1,1)	(1,2)	(1,3)
	2	(2,1)	(2,2)	(2,3)
	3	(3,1)	(3,2)	(3,3)

⇒ Bir pakette ikisi arızalı 6 ampul vardır. Kullanmak için 3 ampul seçildiğinde hiçbirinin arızalı olmama olasılığı = ?

2A
4S

$$P[A] = \frac{\binom{4}{3}}{\binom{6}{3}}$$

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

⇒ Bir poker oyununda bir el 5 karttan oluşuyor. Bir elde iki As üç joker olma olasılığı nedir?

$$52\begin{pmatrix}4\text{AS}\\4\text{ Joker}\end{pmatrix} \xrightarrow{5\text{ seç}} \begin{matrix}2\text{AS}\\3\text{ Joker}\end{matrix}$$

$$P[A] = \frac{\binom{4}{2}\binom{4}{3}}{\binom{52}{5}} = \frac{24}{2{,}598{,}960}$$

⇒ Bir üniversite deposunda 10'u lazer 15'i kartuşlu olmak üzere 25 yazıcı bulunmaktadır. Bir teknisyen tarafından rastgele 6 yazıcı seçiliyor.

a) 3 tanesinin lazer yazıcı olma olasılığı

b) En az üçünün kartuşlu olma olasılığı nedir?

a) $$\frac{\binom{10}{3}\binom{15}{3}}{\binom{25}{6}}$$

b) $P(K=3)+P(K=4)+P(K=5)+P(K=6)$

$= P(K \geq 3)$

$$P(K=3) = \frac{\binom{15}{3}\binom{10}{3}}{\binom{25}{6}}$$

$$P(K=4) = \frac{\binom{15}{4}\binom{10}{2}}{\binom{25}{6}}$$

$$P(K=5) = \frac{\binom{15}{5}\binom{10}{1}}{\binom{25}{6}}$$

$$P(K=6) = \frac{\binom{15}{6}\binom{10}{0}}{\binom{25}{6}}$$

Tümleyen mantığı ile yapılabilir.

$$\Rightarrow \sum_{k=3}^{6} \frac{\binom{15}{k}\binom{10}{6-k}}{\binom{25}{6}}$$

$= 0{,}853$

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

Teorem: n adet farklı nesneyi sırasıyla $n_1, n_2, \ldots n_k$ eleman içeren k alt gruba bölme sayısı

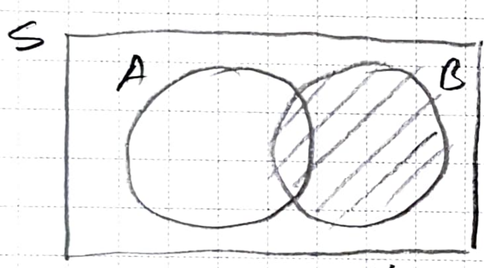
$$\frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_k!} \qquad \boxed{\sum_{i=1}^{k} n_i = n}$$

Grub elemanları toplamı

⇒ 10 kişi 3 tarlada çalışmak üzere gruplara bölünüyor 5 kişi A tarlasına, 3 kişi B tarlasına 2 kişi C tarlasına gönderiliyor. Kaç farklı durum vardır.

$$\frac{10!}{5! \cdot 3! \cdot 2!} = 2520$$

Koşullu Olasılık ve Bağımsızlık



B bilindiğine göre A olayının gerçekleşme olasılığı

I. durum $P(A) = \frac{\text{A'nın çıkış sayısı}}{\text{Örnek uzay}}$

II. durum $P(A/B) = \frac{P(AB)}{P(B)}$

Tanım: Koşullu olasılık B olayının meydana geldiğinin bilinmesi durumunda A olayının meydana gelme olasılığı $P(A \backslash B)$ ile gösterilir. Ve B bilindiğinde A diye okunur.

$$P(A \backslash B) = \frac{P(AB)}{P(B)}, \quad P(B) \neq \emptyset$$

NOT: Tüm çıktılar eşit olasılıklı olduğunda formülü kullanmadan koşullu olasılığı bulmak mümkündür. Eğer B'nin gerçekleştiğini önceden biliyorsak sadece B'deki çıkışları dikkate alarak olasılıkları hesaplayabiliriz.

Örneğin; zar atma deneyinde tek sayı geldiğini biliyorsa 3 gelme olasılığı

$$P(3/\text{tek sayı}) = \frac{1}{3} \quad S = \{1, 3, 5\}$$

$$\boxed{P(AB) = P(A \backslash B) \cdot P(B)}$$ Çarpım kuralı

Tanım: Bağımsızlık A ve B olaylarının istatiksel olarak bağımsız olmaları için; $\boxed{P(AB) = P(A) \cdot P(B)}$ şartı sağlanmalıdır.

$$P(A \backslash B) = \frac{P(AB)}{P(B)} = \frac{P(A) \cdot \cancel{P(B)}}{\cancel{P(B)}}$$

$$\boxed{P(A \backslash B) = P(A)}$$ B bilindiğinde A'nın olasılığı değişmiyor.

Üç olayın Bağımsızlığı (A, B, C)

İkili bağımsız
$$\begin{cases} P(AB) = P(A) \cdot P(B) \\ P(AC) = P(A) \cdot P(C) \\ P(BC) = P(B) \cdot P(C) \end{cases}$$

Hepsi bağımsız çıktığında üçlü ye bakılır.

Üçlü bağımsız $\{ P(ABC) = P(A) \cdot P(B) \cdot P(C)$

4 şartın dördü de sağlanıyorsa bu 3 olay bağımsızdır. Her hangi birinin sağlanmaması durumunda olaylar bağımsız değildir denir.

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

TEKSAN

⇒ "0" ve "1" den oluşan 3 bit bir haberleşme kanalından iletiliyor. İletilirken 1 bitin değişme olasılığı 0,1 olarak veriliyor.

a) En az 1 bitin değişme olasılığı

b) Sadece 1 bitin değişme olasılığı nedir?

B_1 B_1' P(değişme) = 0,1

B_2 B_2' P(Değişmeme) = 0,9

B_3 B_3'

değişmiyor değişiyor

a) {en az 1 bit değişme} = A

$P(A) = 1 - P(\text{hiçbirinin değişmediği}) = 1 - (0,9)^3 = \underline{0,271}$

$\rightarrow P(B_1 B_2 B_3) = P(B_1)P(B_2)P(B_3) = (0,9)^3$

b) B = {sadece 1 bit değişsin}

$P(B) = P(B_1' B_2 B_3) + P(B_1 B_2' B_3) + P(B_1 B_2 B_3')$

$= (0,1)(0,9)(0,9) + (0,9)(0,1)(0,9) + (0,9)(0,9)(0,1)$

$= 3.(0,1)(0,9)^2 = \underline{0,243}$

⇒ İki zar peş peşe atılsın

a) Birinci zardan 2 gelme olasılığı

b) İkinci zardan 2 gelme olasılığı

c) İki zar toplamının 4 olma olasılığı

d) İlk zardan 2 geldiği bilindiğine göre toplamın 4 olma olasılığı

e) İlk zardan 2 gelme ve toplamın 4 olma olasılığı

a) $P(B_2) = 1/6$

b) $P(İ_2) = 1/6$

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

c) $A = \{\text{İki zar toplamı } 4\}$

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

d) $P(A/B_2) = \dfrac{P(AB_2)}{P(B_2)} = \dfrac{1/36}{1/6} = 1/6$

e) $P(B_2A) = P(A/B_2)\,P(B_2) = 1/36$

⇒ $[-1,2]$ aralığında rastgele bir X sayısı seçiliyor

$A = \{X < 0\}$

$B = \{X > 0{,}75\}$

$C = \{|X - 0{,}5| < 0{,}5\}$ olsun

a) A, B, C, AB, AC olasılıklarını bulunuz.

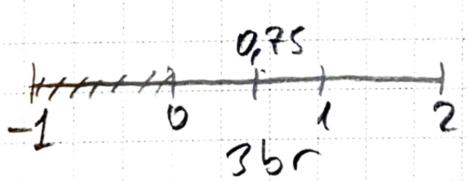
b) A ve B bağımsız mıdır?

c) A ve C bağımsız mıdır?

$S = \{[-1,2]\}$ → sürekli örnek uzay (sayılamaz)

→ ayrık örnek uzay (sayılabilir)

$P(S) = 1$



3br — 1

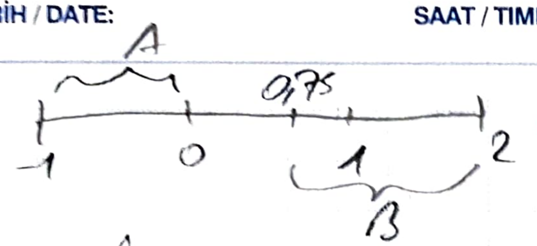
1,25 — x

- $P(A) = \{x < 0\} = 1/3$
- $P(B) = \dfrac{5}{12}$
- $P(C) = 1/3$

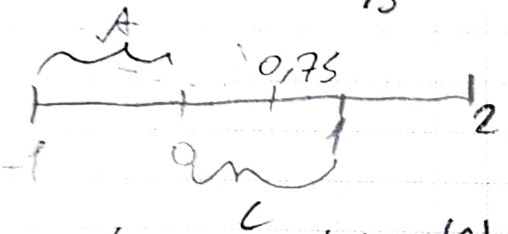
$-0{,}5 < x - 0{,}5 < 0{,}5$

$0 < x < 1$

(Eğer örnek uzay sürekliyse nokta olasılığı sıfırdır.)



• $P(AB)=0$



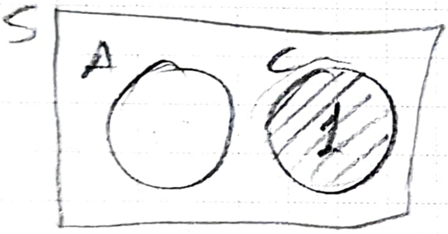
• $P(AC)=0$

b) $P(AB) = P(A) . P(B)$ ise bağımsızdır.

$0 \neq \frac{1}{3} \cdot \frac{5}{12}$ bağımsız değildir.

c) $P(AC) = P(A) . P(C)$

$0 \neq \frac{1}{3} \cdot \frac{1}{3}$ bağımsız değildir.



⇛ 7 kırmızı, top içeren bir çantadan yerine koymaksızın rastgele bir top seçiliyor.

a) 1. topun Yeşil 2. topun Kırmızı
b) 2 topun yeşil
c) Birinci topun yeşil olduğu bilindiğine göre 2. topun kırmızı olma

$\boxed{\begin{matrix}7K\\3Y\end{matrix}}$ Yerine koyma

$P[K_1] = \frac{7}{10}$ $P[Y_1] = \frac{3}{10}$

$P[K_2/K_1] = \frac{6}{9}$ $P[Y_2/K_1] = \frac{3}{9}$ $P[K_2/Y_1] = \frac{7}{9}$ $P[Y_2/Y_1] = \frac{2}{9}$

$P[K_1K_2] = P[K_1]P[K_2/K_1]$

$= \frac{7}{10} \cdot \frac{6}{9}$

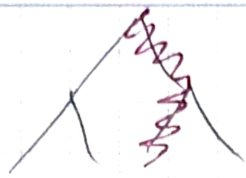
$P[K_1Y_2] = P[Y_2/K_1]P[K_1]$

$= \frac{7}{10} \cdot \frac{3}{9}$

KONU / SUBJECT: 2.4

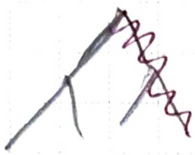
TARİH / DATE: SAAT / TIME:

TEKSAN

a) 

$P[Y_1 K_2] = P[K_2 / Y_1] P[Y_1] = \frac{3}{10} \cdot \frac{7}{9}$

$= P[K_2 Y_1]$

b) 

$P[Y_1 Y_2] = P[Y_2 / Y_1] P[Y_1] = \frac{3}{10} \cdot \frac{2}{9}$

$= P[Y_2 Y_1]$

c) $P[K_2 / Y_1] = \frac{7}{9}$

ÖDEV ⇒ Yapılan bir anket sonucu Birleşmiş Milletlerde aşağıdaki tablo elde edilmiştir.

Etnik grub	Yetişkin sayısı (milyon)	Lise derecesi %	Üniversite %	Yüksek lisans ve üzeri %
Amerikan	1,1	6,4	6,1	3,3
Siyahi	16,8	5,3	7,5	3,8
Asyalı	4,3	7,7	22,7	13,9
İspanyol	11,2	4,8	5,8	3,3
Beyaz	132,0	6,3	13,9	7,7

Sağ üç sütun koşullu yüzdeleri veriyor

a) Kaç asyalı yüksek lisans ve üzeri eğitim seviyesine sahiptir.

b) Yetişkin amerikanların yüzde kaçı üniversite mezunudur.

c) 1 kişinin lise mezunu olduğu bilindiğine göre ispanyol olma olasılığını hesaplayınız.

BAYES KURALI

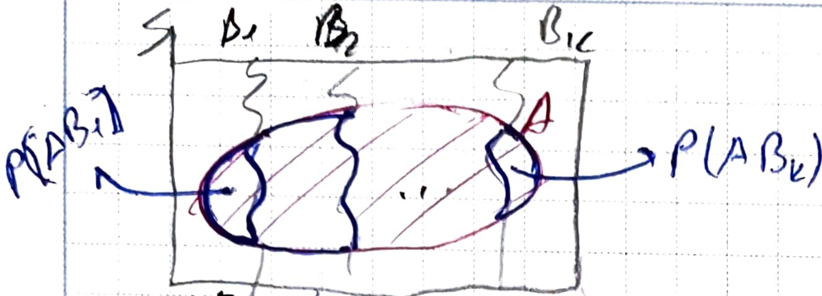
$B_1, B_2, \ldots, B_K$ olayları S örnek uzayında aşağıdaki koşulları sağlayan olaylar olsun.

a) $B_i B_j = \emptyset$ tüm i ve j'ler için

b) $B_1 \cup B_2 \cup \ldots B_K = S$

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:



⇒ $P[A/B_i]$ olasılıklarını biliyorsak $P[A]$ olasılığını nasıl hesaplarız?

$$P[A] = P[AB_1] + P[AB_2] + \dots + P[AB_k]$$
$$= P[A/B_1].P[B_1] + P[A/B_2].P[B_2] + \dots$$

çarpım kuralı

$$\boxed{P[A] = \sum_{i=1}^{k} P[A/B_i]\,P[B_i]}$$

Toplam olasılık

⇒ $P[B_i/A]$ nasıl hesaplanır?

$$P[B_i/A] = \frac{P[AB_i]}{P[A]}$$

$$\boxed{P[B_i/A] = \frac{P[A/B_i]\,P[B_i]}{\sum_{i=1}^{k} P[A/B_i]\,P[B_i]}}$$

Bayes Kuralı (olaya ters gitmek istendiğinde)

⇒ Bir fabrikada üretilen ürünlerin sırasıyla %30, %45 ve %25'i A, B ve C makinalarından üretilmektedir. Bu üç makinanın arızalı üretim yapma olasılıkları sırasıyla %2, %3 ve %2 olsun. Rastgele bir ürün seçilsin.

a- Ürünün arızalı olma olasılığı

b- Ürünün arızalı olduğu bilindiğine göre C makinada üretilme olasılığı

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

TEKSAN

		Arıza	
P(A)	%30	%2	→ P(Arıza\A)
P(B)	%45	%3	→ P(Arıza\B)
P(C)	%25	%2	→ P(Arıza\C)

a) $P(\text{Arıza}) = \underbrace{P(A, \text{Arıza})} + P(B, \text{Arıza}) + P(C, \text{Arıza})$

$$= P(A).P(\text{Arıza}\backslash A) + P(B)\,P(\text{Arıza}\backslash B) + P(C)\ldots$$

$$= 0{,}3 . 0{,}02 + 0{,}45 . 0{,}03 + 0{,}25 . 0{,}02$$

$$= 0{,}0245$$

$$\boxed{\begin{array}{l} P(A\backslash B) = \dfrac{P(AB)}{P(B)} \\ P(AB) = P(B)P(A\backslash B) \end{array}}$$

b)

$$P(C\backslash \text{Arızalı}) = \frac{P(C, \text{Arızalı})}{P(\text{Arızalı})}$$

$$= \frac{P(C)\,P(\text{Arızalı}\backslash C)}{P(\text{Arızalı})}$$

$$= \frac{0{,}25 . 0{,}02}{0{,}0245}$$

ÖDEV ⇒ Genetik bir hastalık DNA görüntüleme testi kullanıldığında teşhis edilmektedir. Hastalığın 1000 kişiden 1'inde görüldüğü biliniyor. Yapılan test %0,5 hatalı pozitif olmaya sahiptir. Seçilen bir kişinin test sonucunun pozitif olduğu bilindiğine göre kişinin gerçekten hasta olma olasılığı nedir?

Bayes 1/6

→ Sonuç +

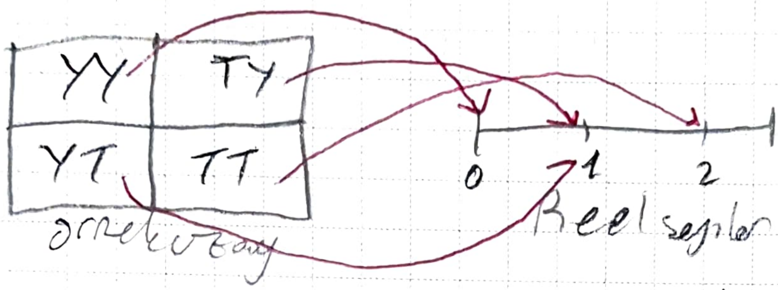
Hatalı Pozitif → P(Hasta \ Hasta değil)

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

AYRIK RASTLANTI DEĞİŞKENLERİ

Tanım: Rastlantı değişkeni Bir rastlantı değişkeni rastlantı deneyinin her bir çıktısına bir sayı atayan fonksiyondur. Tanım kümesi örnek uzay değer kümesi ise reel sayılardır.



Tura gelme sayısı (X)	YY	YT	TY	TT
	0	1	1	2

$$\left.\begin{aligned} P(0) &= 1/4 \\ P(1) &= 1/2 \\ P(2) &= 1/4 \end{aligned}\right\} \text{PMF}$$

1) Örnek uzayda boş eleman kalmaz
2) Tanım kümesindeki bir eleman değer kümesinde sadece 1 noktaya gidebilir
3) Tanım kümesinde birden fazla eleman değer kümesinde aynı noktaya gidebilir

$(X, Y, Z, \dots)$ rastlantı değişkeninin adı büyük harf.
(x_1, x_2, y_1, y_2) olduğu değer küçük harf

NOT: Rastlantı değişkenleri ayrık ve sürekli olmaları üzerine ikiye ayrılır. Sayılabilir bir kümeden değerler alan rastlantı değişkenlerine ayrık rastlantı değişkeni denir.

$X = \{0, 1, 2\}$ → Ayrık

$X = \{X : [0, 2]\}$ → Sürekli

Tanım: Olasılık Kütle Fonksiyonu (Probability Mass Function, PMF)

Bir X rastlantı değişkeni sonlu ve sayılabilir x değerleri alıyorsa ayrıktır denir. Ve olasılık kütle fonksiyonu (pmf) ile ifade edilir.

1- $P(X=x) = P_X(x) \geq 0$

2- $\sum_{x} P(X=x) = 1$

$P(X=0) = [0,1]$

negatif
pozitif
1'den büyük

⇒ 3'ü arızalı 8 bilgisayardan iki tane seçelim arızalı bilgisayar sayısının olasılık kütle fonksiyonunu (pmf) sını bulunuz ve çiziniz?

$\boxed{\begin{array}{c} 3A \\ 5S \end{array}}$

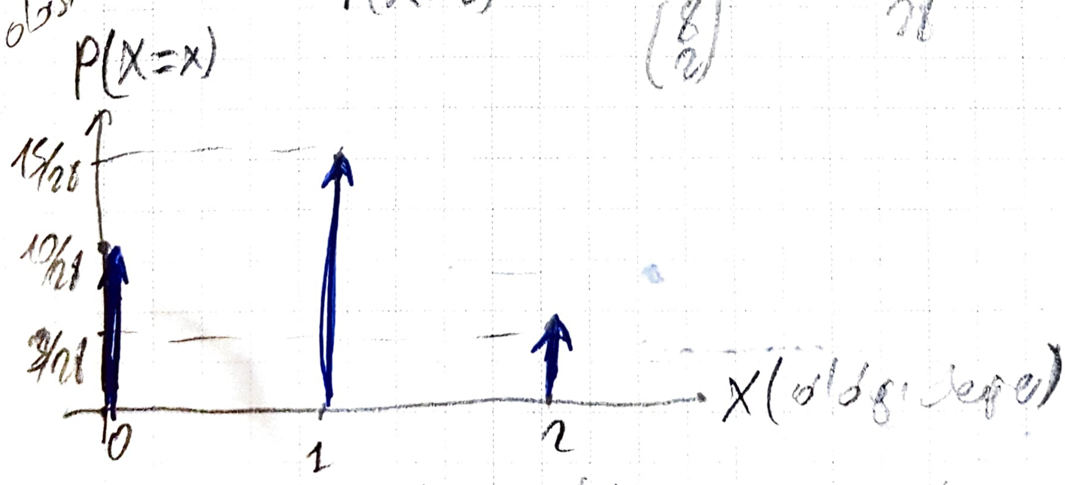
$X =$ arızalı bil. sayısı $= \{0, 1, 2\}$

$$P(X=0) = \frac{\binom{3}{0}\binom{5}{2}}{\binom{8}{2}} = \frac{10}{28}$$

Rastlantı değerl kombinasyonsa PMF de olabilir

$$P(X=1) = \frac{\binom{3}{1}\binom{5}{1}}{\binom{8}{2}} = \frac{15}{28}$$

$$P(X=2) = \frac{\binom{3}{2}\binom{5}{0}}{\binom{8}{2}} = \frac{3}{28}$$



Arada boşluk olduğundan ayrıktır

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

TEKSAN

Tanım: Toplam Dağılım Fonksiyonu (Yığılımlı dağılım fonk)
(Cumulative Distribution Function, CDF)

Ayrık bir X rastlantı değişkeninin toplam dağılım fonksiyonu $F(x)$ ile gösterilir ve şu şekilde tanımlanır.

$\boxed{F(x) = P(X \leq x)}$ ve ya $F(x) = \sum_{y \leq x} P(y)$

Özellikler

1- $\lim_{x \to -\infty} F(x) = 0$

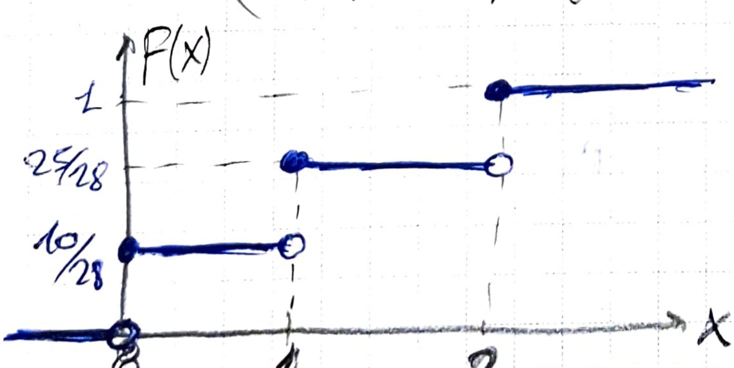
2- $\lim_{x \to \infty} F(x) = 1$

3- $F(x)$ kesinlikle azalmayan bir fonksiyondur

4- $P(x) = F(x) - F(x-)$ ⟶ bundan önceki değer

⟹ Bir önceki örnek için X rastlantı değişkeninin toplam dağılım fonk (CDF) bulunuz ve çiziniz
[Ayrıkca küçük eşit veya değile dikkat]

$$F(x) = \begin{cases} x < 0, & F(x) = 0 \\ 0 \leq x < 1, & F(x) = P(x=0) = \frac{10}{28} \\ 1 \leq x < 2, & F(x) = P(x=0) + P(x=1) = \frac{25}{28} \\ x \geq 2, & F(x) = P(x=0) + P(x=1) + P(x=2) = 1 \end{cases}$$



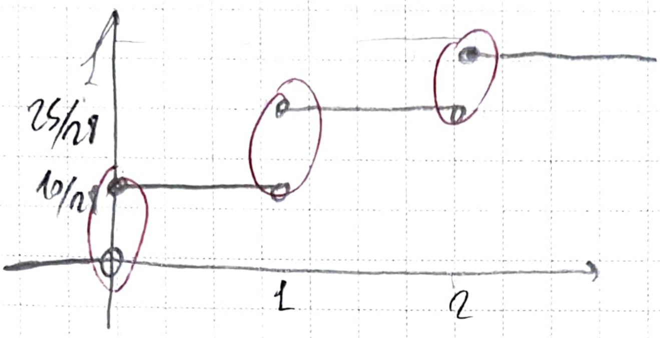
KONU / SUBJECT: RD → Rastlantı değişkeni

TARİH / DATE:

SAAT / TIME:

TEKSAN

NOT: Ayrık bir rastlantı değişkeni için CDF basamak şeklinde bir fonksiyondur. Bu fonksiyondaki sıçramalar, RD aldığı değerlerde gerçekleşir.



$X = \{0, 1, 2\}$

$P(X=0) = \frac{10}{28} - 0$

$P(X=1) = \frac{25}{28} - \frac{10}{28} = \frac{15}{28}$

$P(X=2) = 1 - \frac{25}{28} = \frac{3}{28}$

$\Rightarrow P(x) = \frac{k}{2^x} \quad x = -1, 0, 1, 2$

X rastlantı değişkeninin olasılık dağılımı verilmektedir.

a) Bu dağılımın geçerli olması için $k = ?$

b) Olasılık kütle fonk. bul ve çiz? PMF

c) Toplam dağılım fonk. bul ve çiz? CDF

a) $P(X=-1) = \frac{k}{2^{-1}} = 2k$

$P(X=0) = \frac{k}{2^0} = k$

$P(X=1) = \frac{k}{2^1} = k/2$

$P(X=2) = \frac{k}{2^2} = \frac{k}{4}$

$3k + \frac{k}{2} + \frac{k}{4} = 1, \quad \frac{15k}{4} = 1, \quad \boxed{k = \frac{4}{15}}$

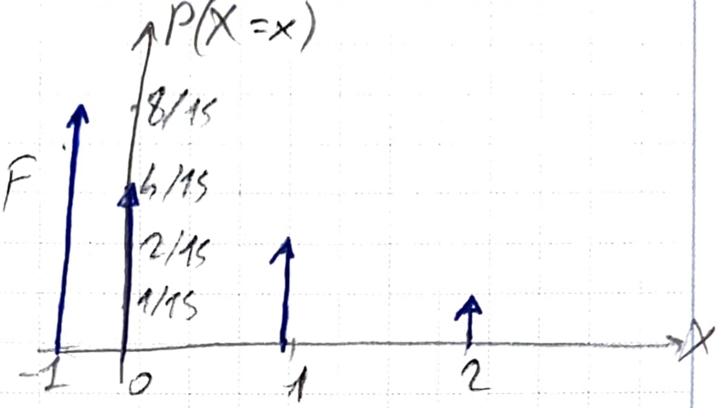
b) $P(x=-1) = \frac{8}{15}$

$P(x=0) = \frac{4}{15}$

$P(x=1) = \frac{2}{15}$

$P(x=2) = \frac{1}{15}$

PMF



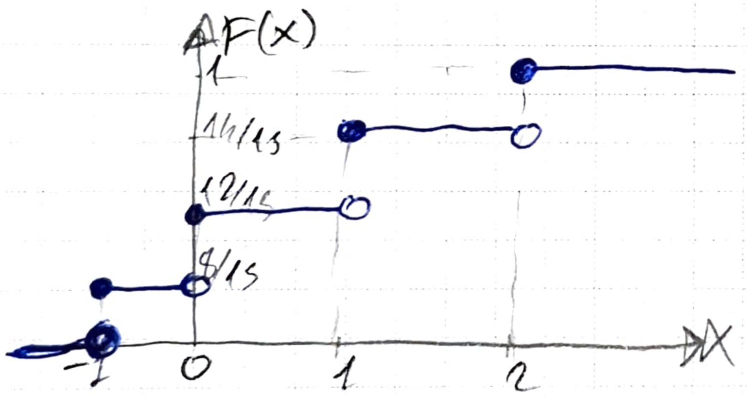
x'i yerine yaz

c) $F(x) = \begin{cases} x < -1, & F(x)=0 \\ -1 \leq x < 0, & F(x) = P(x=-1) = \frac{8}{15} \\ 0 \leq x < 1, & F(x) = P(x=-1) + P(x=0) = \frac{12}{15} \\ 1 \leq x < 2, & F(x) = P(x=-1) + P(x=0) + P(x=1) = \frac{14}{15} \\ x \geq 2, & F(x) = 1 \end{cases}$

Artacak

0'dan başla

1'de biter



RASTLANTI DEĞİŞKENİN ORTALAMA DEĞERİ

(BEKLENEN DEĞER)

Bir rastlantı değişkeni hakkında bilmek istediğimiz en önemli şeylerden biri rastlantı değişkeninin ortalama olarak hangi değerleri alacağıdır. Bu değer olasılıklarla ağırlıklandırılmış beklenen değerdir.

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

Tanım: Beklenen Değer (Mean) Ayrık bir X rastlantı değişkeninin ortalama değeri $E[X]$ ile gösterilir ve şu şekilde hesaplanır.

$$E[X] = \sum_x x \cdot P(X=x) = \mu$$

$$\boxed{\begin{matrix} 0 & 1 \\ & 2 \end{matrix}} = \{0,1,1,1,0,2,2,0,0,0\}$$

$$P(X=0) = \frac{5}{10} \quad P(X=1) = 3/10 \quad P(X=2) = 2/10$$

$$E[X] = \sum_x x P(X=x)$$

$$= 0 \cdot \left(\frac{5}{10}\right) + 1 \cdot \left(\frac{3}{10}\right) + 2 \cdot \left(\frac{2}{10}\right)$$

$$= 0{,}7$$

Tanım: Fonksiyonun Ortalama Değeri

X, $P(x)$ olasılıklarına sahip ayrık bir rastlantı değişkeni $g(x)$ ise gerçel değerli bir fonksiyon olsun. Bu durumda;

$$E[g(X)] = \sum_X g(x) P(x) \text{ ile hesaplanır}$$

Tanım: Varyans (Değişinti)

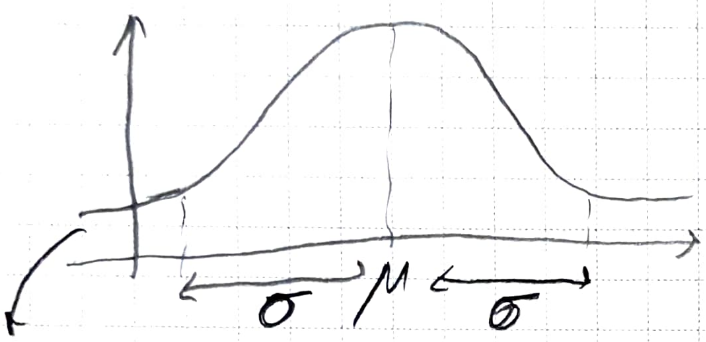
X rastlantı değişkeninin varyansı $V(x) = \sigma^2$ ile gösterilir ve şu şekilde hesaplanır.

$$V(x) = E[(X-\mu)^2] \quad , \mu = E[X]$$

[kesinlikle pozitif çıkar]

Tanımı: Standart Sapma (σ)

Varyansın kareköküne standart sapma denir. Ortalama değer olasılık dağılımının merkezini verirken standart sapma yayılımını verir.



Gauss dağılımı

Teorem: Doğrusal Fonksiyonlar

X bir rastlantı değişkeni a ve b sabitler olmak üzere

i) $E[aX+b] = aE[X]+b$ [Sabitin ortalaması kendisidir]

ii) $V(aX+b) = V(aX) + \underbrace{V(b)}_{0} = a^2 V(X)$ 0 → DC voltaj

iii) $\sigma_{aX+b} = |a|\sigma$

iv) $E[X_1+X_2+X_3+\ldots] = E[X_1]+E[X_2]+\ldots$

- $V(x) = E[(x-\mu)^2] \quad , \mu = E[x]$

$$= E[x^2 - 2x\mu + \mu^2]$$

$$= E[x^2] - E[2x\mu] + E[\mu^2]$$

$$= E[x^2] - 2\underbrace{\mu}_{E[x]} E[x] + \underbrace{\mu^2}_{E[x]^2}$$

$$= E[x^2] - (E[x])^2$$

$$\boxed{Var(x) = E[x^2] - \mu^2}$$

$$E[x^2] = \sum_x x^2 . P(x)$$

$\Rightarrow$

x	P(x)
0	10/28
1	15/28
2	3/28

a) $E[x] = ?$
b) $Var(x) = ?$
c) $E[3x+7] = ?$

a) $E[x] = \sum_x x P(x)$

$$= 0 . \left(\frac{10}{28}\right) + 1 . \left(\frac{15}{28}\right) + 2\left(\frac{3}{28}\right)$$

$$= \frac{21}{28}$$

b) $Var(x) = E[x^2] - (E[x])^2$

$$E[x^2] = 0^2 . \left(\frac{10}{28}\right) + 1^2\left(\frac{15}{28}\right) + 2^2\left(\frac{3}{28}\right)$$

$$= \frac{27}{28}$$

$$Var(x) = \frac{27}{28} - \left(\frac{21}{28}\right)^2$$ "pozitif çıkar"

c) $E[3x+7] = \sum_x \underbrace{(3x+7)}_{g(x)} . P(x)$

$$= 7 . \frac{10}{28} + 10\, \frac{15}{28} + 13\, \frac{3}{28}$$

I. Yol

$$E[3x+7] = 3 . E[x] + 7 = 3 . \frac{21}{28} + 7$$

II. Yol

Chebyshev Eşitsizliği

X rastlantı değişkeni μ ortalamalı ve σ^2 varyanslı bir rastlantı değişkeni olsun. Herhangi bir pozitif bir k değeri için

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2} \quad \text{(üst sınır)}$$

veya

$$P(\mu - k\sigma < X < \mu + k\sigma) > 1 - \frac{1}{k^2} \quad \text{(Alt sınır)}$$

NOT: Chebyshev eşitsizliği rastlantı değişkeninin ortalama değeri ve varyansı bilindiğinde bazı olasılıkları tahmin etmeye çalışırsak kullanılır. Bu eşitsizlik bize kabaca bir tahmin verir.

⇒ Bir araba aküsünün ortalama çalışma ömrü 30 ay, standart sapması 5 ay dır.

a) Akünün en az 18 ay çalışma olasılığını tahmin ediniz?

b) Tüm akülerin en az %80 inin çalıştığı bir aralık belirleyiniz.

a) $\mu = 30 \quad \sigma = 5$

$$P(\underbrace{\mu - k\sigma}_{18 \text{ ay (en az)}} < X < \mu + k\sigma)$$

$$30 - k.5 = 18$$

$$k = \frac{12}{5} = 2{,}4$$

$$P(18 < X < 42) > 1 - \frac{1}{(2{,}4)^2} = 0{,}826$$

b) $1-\frac{1}{k^2}=0,9$ $k=\sqrt{10}=3,16$

$$P(\underbrace{30-3,16\times5}_{14,2}<X<\underbrace{30+3,16\times5}_{45,8})>0,9$$

⇒ Bir alışveriş merkezine gelen günlük müşteri sayısı ortalaması 20, varyansı 4'dür. Müşteri sayısı X'in [16,24] aralığında olma olasılığı hakkında ne söylene bilir?

$\mu=20$, $\sigma=\sqrt{\sigma^2}=\sqrt{4}=2$

$P(16\leq X\leq 24)$

$P(15<X<25)>1-\frac{1}{k^2}$

$15=20-k.2$
$k=2,5$

$25=20+k.2$
$k=2,5$

$1-\frac{1}{(2,5)^2}=0,84$

$P(15<X<25)>0,84$

Önemli Ayrık Dağılımlar

1- Bernoulli Dağılımı (kardeş var yok)

X rastantı değişkeni nesnelerin durumunu gösteren bir değişken olsun. X sadece $X\in\{0,1\}$ değerlerini alıyorsa bu değişkene bernoulli değişkeni denir.

$P(X=1)$ = başarı olasılığı $=P$ → (istenilen şey)

$P(X=0)$ = başarısızlık olasılığı $=1-p=q$

(if - Else)

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

TEKSAN

⇒ Bernoulli değişkeninin ortalama değeri ve varyansını hesaplayınız?

$$E[X] = \sum_x x.P(X=x)$$
$$= 1.p + 0.q$$
$$= p$$

$$Var(x) = E[x^2] - (E[x])^2$$

$$E[x^2] = \sum_x x^2.P(X=x)$$
$$= 1^2 p + 0^2 q$$
$$= p$$

$$Var(x) = p - p^2 = p(1-p) = pq$$

$$\boxed{\begin{array}{l} E[x]=p \\ Var(x)=pq \end{array}}$$

2- Binom Dağılımı (iki terimli dağılım) (n kez bernoulli)

n nesne incelediğimizi ve arızalı nesne sayısını merak ettiğimizi var sayalım. Bu işlem deneyin n kez tekrarlanmasıdır. Ve bernoulli denemeleri olarak adlandırılır. Binom dağılımı aşağıdaki özelliklere sahiptir;

1- Her denemenin sonucu başarı ya da başarısızlıktır.

2- Denemeler birbirinden bağımsızdır.

3- X rastlantı değişkeni n denemedeki başarı sayısı olarak tanımlanır.

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

TEKSAN

Kütle dağılımı

n nesnede k başarı

$$P(X=k) = \binom{n}{k} p^k \cdot q^{n-k} \quad , k=0,\dots,n$$

$$\left.\begin{aligned} E[X] &= n.p \\ Var(X) &= npq \end{aligned}\right\} n \text{ kez}$$

⇒ Bir saklama işlemi sonucu elemanın sağlam kalma olasılığı 0,75 dir.

a) Test edilen 8 nesneden 2 tanesinin sağlam kalması

b) En az iki tanesinin sağlam kalması

c) En çok 6 tanesinin sağlam kalması

$p = 0{,}75 \quad , \quad n = 0{,}25$

a) $P(X=2) = \binom{8}{2}(0{,}75)^2 \cdot (0{,}25)^6$

b) $P(X \geq 2) = \sum_{k=2}^{8} \binom{8}{k}(0{,}75)^k (0{,}25)^{8-k}$

$P(x \geq 2) = 1 - P(x < 2)$ // Diğer yol

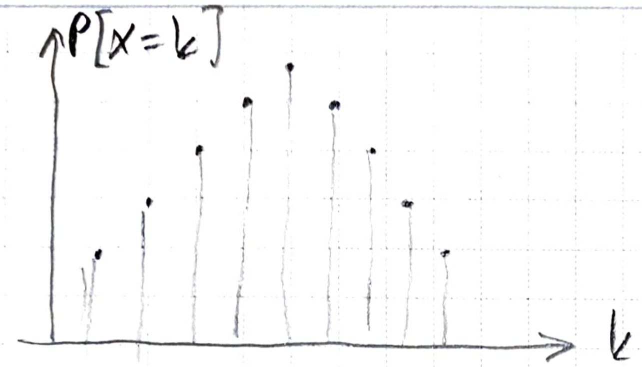
$$= 1 - \underbrace{\binom{8}{0}(0{,}75)^0 (0{,}25)^8}_{P(X=0)} - \underbrace{\binom{8}{1}(0{,}75)^1 (0{,}25)^7}_{P(X=1)}$$

c) $P(X \leq 6) = 1 - P(X=7) - P(X=8)$

$$= 1 - \binom{8}{7}(0{,}75)^7 (0{,}25)^1 - \binom{8}{8}(0{,}75)^8 (0{,}25)^0$$

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:



3- Geometrik Dağılım (kapı anahtar)

İki terimli dağılımda deneme sayısı n sabittir. Ayrıca ilgilenilen değişken başarı sayısıdır. Bazı durumlarda istenilen sayıda başarı elde edilmesi için yapılması gereken deneme sayısı sorulur.
İlk başarı elde edilinceye kadar gereken deneme sayısı Y, p parametrede geometrik rastlantı değişkenidir.

Geometrik BD nin PMF'si;

$$P(Y=y)=(1-p)^{y-1}\cdot p \qquad y=1,2,\ldots$$

$$E[Y]=\frac{1}{p}$$

$$var(Y)=\frac{1-p}{p^2}$$

⇒ Bir üretim süreci için her 100 nesneden 1 inin bozuk olduğu biliniyor.

a) İlk bozuk nesnenin denenen 5'inci nesne olma olasılığı

b) İlk bozuk nesne bulunana kadar yapılması gereken ortalama deney sayısı

d) $p = \frac{1}{100} = 0{,}01$

$P(Y=5) = (0{,}99)^4\, 0{,}01$ // Bağımsız deney

b) $E[Y] = 100$

$E[Y] = \sum_y y\,P(Y=y) = \sum_{y=1}^{\infty} y(1-p)^{y-1} p = 100$

4- Negatif İki Terimli Dağılım

p başarı olasılığı olmak üzere bağımsız bernoulli denemeleri dizisinde r'inci başarının elde edilmesi için gereken deneme sayısı Y ile gösterilsin. Y rastlantı değişkeninin PMF si;

$P(Y=y) = \binom{y-1}{r-1} p^r q^{y-r}, \quad y = r, r+1, \ldots$

r. başarı gelene kadar yapılan deney sayısı

$\underbrace{\ldots\ldots}_{\text{öncekiler bilinmiyor } (y-1)}$ $\overset{y}{r}$. başarı

⇒ Bir eğitmen, ilk çocuklarını bekleyen 5 çifti eğitme olmak istiyor. Sorduğu çiftlerin %20 si eğitime katılmak istiyor. Bu 5 çifti bulmadan önce 15 çifte eğitime katılıp katılmayacağı sorma olasılığı nedir.

Deneme sayısı = 15

$P(Y=15) = \binom{14}{4} (0{,}2)^5 (0{,}8)^{10}$

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

TEKSAN

⇒ NBA ligi serisinde 7 maçtan 4 ününü kazanan takım şampiyon olmaktadır. A takımının, B takımına göre kazanma olasılığı 0,55 olsun. A ve B takımları eşleştiğinde A takımının 6ıncı maçta şampiyon olma olasılığı nedir?
A takımının şampiyon olma olasılığı nedir?

. . . . . . 6.maç
4.galibiyet

$$P(Y=6)=\binom{5}{3}(0{,}55)^4(0{,}45)^2=0{,}1853$$

$$P(A)=P(Y=4)+P(Y=5)+P(Y=6)+P(Y=7)$$
$$=\binom{3}{3}(0{,}55)^4\cdot(0{,}45)^0+\binom{4}{3}(0{,}55)^4(0{,}45)^1$$
$$+\binom{5}{3}(0{,}55)^4(0{,}45)^2+\binom{6}{3}(0{,}55)^4(0{,}45)^3$$

5- Poisson Dağılımı

(Birim zamandaki kazalar veya elektron)

İstenilen sınırlar içerisinde istenilen sayıda olayın gerçekleşme olasılığını gösteren rastlantı değişkenidir. Çoğunlukla büyük ölçekli olaylarda kullanılır.

Poisson dağılımının olasılık kütle fonk.;

$$pois(x,\mu)=P(X=x)=\frac{e^{-\mu}\mu^x}{x!}\quad,\ x=0,1,\ldots$$

μ: Birim zamanda meydana gelen olay sayısı

$$E[X]=var[X]=\mu$$

⇒ Bir laboratuvar deneyinde bir milisaniyede sayıcıdan geçen ortalama radyoaktif parçacık sayısı 4'dür. a) Verilen bir milisaniyede sayıcıya giren parçacık sayısının 6 olma olasılığı nedir. b) Sayıcıya en az 6 parçacık girme olasılığı nedir.

$\mu = 4$

a) $P(X=6) = \dfrac{e^{-4} \cdot 4^6}{6!} = 0{,}1042$

b) $P(X \geq 6) = 1 - P(X<6)$

$= 1 - \sum_{x=0}^{5} \dfrac{e^{-4} \cdot 4^x}{x!}$

⇒ Endüstriyel bir fabrikada nadiren iş kazası oluşmaktadır. Bir günde iş kazası olma olasılığı binde 5'dir. 400 günlük bir periyod için

a) Sadece 1 günde kaza olma olasılığı

b) En çok 2 günde kaza olma olasılığı

$\mu = \dfrac{5}{1000} \times 400 = 2$

a) $P(x=1) = \dfrac{e^{-2} 2^1}{1!}$

b) $P(X \leq 2) = P(x=0) + P(x=1) + P(x=2)$

$= \sum_{x=0}^{2} \dfrac{e^{-2} 2^x}{x!}$

Moment Üreten Fonksiyonlar (MGF)
(Moment Generating Function)

Y rastlantı değişkeninin bir fonksiyonu g(Y) olsun. g(Y)'nin ortalama değeri;

$$E[g(Y)] = \sum_{y} g(Y) \cdot P(Y=y)$$ ile hesaplanır

$Y = e^{ty}$

durumunda bu ifadeye moment üreten fonksiyon denir.

$$\boxed{m(t) = E[e^{ty}]}$$

$E[Y^n]$ = n. moment

$E[Y]$ = 1. moment = ortalama değer

$E[Y^2]$ = 2. moment

$$\frac{d}{dt}[m(t)]\Big|_{t=0} = E[Y]$$

$$E[Y \cdot e^{ty}]\Big|_{t=0} = E[Y]$$

$$E[Y^2 e^{ty}]\Big|_{t=0} = E[Y^2]$$

$$\boxed{\frac{d^n}{dt^n} m(t)\Big|_{t=0} = E[Y^n]}$$

NOT = Rastlantı değişkenlerinin kuvvetlerinin ortalama değerine moment denir. Moment üreten fonk aşağıdaki özelliklere sahiptir.

1- X ve Y rastlantı değişkenlerinin moment üreten fonksiyonları $M_x(t)$ ve $M_y(t)$ olsun. $M_x(t) = M_y(t)$ ise X ve Y aynı olasılık dağılımına sahiptir. (teklik özelliği)

2- $M_{x+b}(t) = e^{bt} M_x(t)$

b = sabit

$$E[e^{(x+b)t}]$$
$$E[e^{xt} e^{bt}]$$
$$e^{bt} E[e^{xt}]$$

3- $M_{ax}(t) = M_x(at)$

(a, sabit)

$$E[e^{axt}]$$
$$E[e^{atx}]$$

4- $Y = X_1 + X_2 + X_3 + \dots + X_n$ ise

$M_y(t) = M_{x_1}(t) \cdot M_{x_2}(t) \cdot \dots \cdot M_{x_n}(t)$

$\Rightarrow$ $P(x) = \frac{1}{3}$ $x = -1, 0, 1$

a) Moment üreten fonksiyonu kullanarak ortalama değeri ve varyansı hesaplayınız?

b) $M_{x+5}(t) = ?$

a) $M(t) = E[e^{tx}] = \sum_{x=-1}^{1} e^{tx} p(x)$

$= e^{-t} p(x=-1) + e^{t.0} p(x=0) + e^{t.1} p(x=1)$

$$\boxed{M_x(t) = \frac{1}{3}(e^{-t} + 1 + e^{t})}$$

$$E[x] = m'(t=0)$$

$$m'(t) = \frac{1}{3}(-e^{-t}+e^{t})\Big|_{t=0} = 0 = E[x]$$

$$var(x) = E[x^2] - E[x]^2$$

$$E[x^2] = m''(t=0)$$

$$m''(t) = \frac{1}{3}(e^{-t}+e^{t})\Big|_{t=0} = \frac{2}{3}$$

$$Var(x) = 2/3$$

b) $M_{x+5}(t) = e^{5t} M_x(t)$

$$M_{x+5}(t) = e^{5t}\,\frac{1}{3}(e^{-t}+1+e^{t})$$

$$= \frac{1}{3}(e^{4t}+e^{5t}+e^{6t})$$

Toplam Dağılım Fonksiyonundan (CDF) Yararlanarak Olasılık Hesaplama

$$F(x) = P(X \leq x)$$

$$P(a<x<b) = [F(x=b) - F(x=a)] - P(x=b)$$

$$P(a \leq x < b) = [F(x=b) - F(x=a)] + P(x=a) - P(x=b)$$

$$P(a<x \leq b) = [F(x=b) - F(x=a)]$$

$$P(a \leq x \leq b) = [F(x=b) - F(x=a)] + P(x=a)$$

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

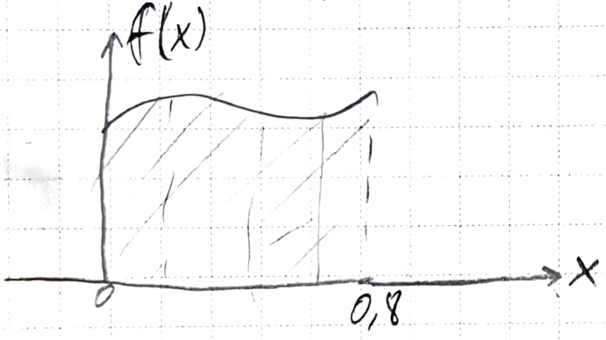
TEKSAN

Sürekli Olasılık Dağılımları

Şimdiye kadar tartışılan tüm rastlantı değişkenleri ayrıktı yani sonlu sayıda ve sayılabilir değerler alıyorlardı. Ancak pratikte karşılaşılan rastlantı değişkenlerinin çoğu belirli bir aralıkta sonsuz değer alan sürekli rastlantı değişkenleridir.

Sürekli rastlantı değişkenlerinin tanımlanmak için kullandığımız araçlar ayrık durumdan farklı olsada aralarında bir ilişki vardır.

Tanım: Olasılık Yoğunluk Fonksiyonu (Probability Density Function, PDF)



Sürekli bir X rastlantı değişkeninin Olasılık yoğunluk fonk (PDF) reel sayılar üzerinde tanımlıdır. ve aşağıdaki özellikleri sağlar.

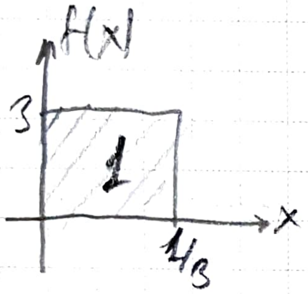
i) $f(x) \geq 0$, tüm x'ler için

ii) $\int_{-\infty}^{\infty} f(x)\,dx = 1$

iii) $P(a \leq x \leq b) = \int_{a}^{b} f(x)\,dx$

*NOT = Sürekli olasılık yoğunlukları sürekli bir aralık üzerinde sonsuz nokta için tanımlı olduğundan bir noktanın olasılığı sıfırdır.

Ayrı iki nokta arasındakidan olan o aralık tam olasılık değeridir. Olasılık yoğunluk fonksiyonunu yüksekliği 1'den büyük olabilir. Ancak altındaki kalan alan kesinlikle 1 olmak zorundadır.



⇒ Bir laboratuvar deneyinde °C olarak reaksiyon sıcaklığı dağılımı aşağıda verilen X rastlantı değişkeni ile modelleniyor.

a) Verilen dağılım geçerli midir?

b) X in sıfır ile 1 arasında olma olasılığı

$$f(x) = \begin{cases} \dfrac{x^2}{3} & , -1 \leq x \leq 2 \\ 0 & , \text{diğer} \end{cases}$$

a) $\int_{-\infty}^{\infty} f(x)\,dx = 1$

$$\int_{-1}^{2} \frac{x^2}{3}\,dx = \frac{x^3}{9}\Big|_{-1}^{2} = \frac{8}{9} - \left(\frac{-1}{9}\right)$$

$= 1$ olduğundan geçerlidir.

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

Tanım: Toplam Dağılım Fonksiyonu (CDF)

Sürekli bir X rastlantı değişkeni için toplam dağılım fonksiyonu şu şekilde tanımlanır.

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

Ayrık durumdaki tüm özellikler sürekli durumda da aynen geçerlidir.

Bu eşitliğin iki sonucu vardır;

1- $P(a<x<b)=F(x=b)-F(x=a)=P(a\leq x\leq b)$
$=P(a\leq x<b)$
$=P(a<x\leq b)$

2- $f(x)=\frac{d}{dx}F(x)$

$F(x)\rightarrow CDF$
$f(x)\rightarrow PDF$

$\Rightarrow$ Bir önceki örneğin toplam dağılım fonksiyonunu bulunuz. Toplam dağılım fonksiyonundan yararlanarak $P(0<x<1)$ olasılığını hesaplayınız?

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

$$F(x)=\begin{cases} x<-1, & F(x)=0 \\ -1\le x\le 2, & F(x)=\int_{-\infty}^{\infty} f(t)\,dt = \dfrac{x^3+1}{9} \\ x>2, & F(x)=1 \end{cases}$$

$$f(t)=\frac{t^2}{3}$$

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

$$\boxed{F(x)=\frac{x^3+1}{9}}$$

b)

$$P(0<x<1)=F(x=1)-F(x=0)$$
$$=\left(\frac{x^3+1}{9}\right)_{x=1}-\left(\frac{x^3+1}{9}\right)_{x=0}$$
$$=\frac{2}{9}-\frac{1}{9}$$
$$=\frac{1}{9}$$

$\int_0^1 f(x)\,dx \rightarrow$ Bu integralle de sonuç bulunabilir

$\Rightarrow$ Bir aracın bozulma süresi X olsun, X'in olasılık yoğunluk fonksiyonu aşağıda veriliyor. Toplam dağılım fonk. bul.

$$f(x)=\begin{cases} \dfrac{3x^2}{64}e^{-\left(\frac{x^3}{64}\right)}, & x>0 \\ 0, & \text{diğer} \end{cases}$$

$$F(x)=\int_{-\infty}^{x} f(t)\,dt=\int_0^x \frac{3t^2}{64}e^{-\left(\frac{t^3}{64}\right)}dt$$

$$\left[\begin{array}{l} -\dfrac{t^3}{64}=u \\ -\dfrac{3t^2}{64}dt=du \end{array}\right]$$

$$-\int_0^x e^u\,du = -e^{-\left(\frac{t^3}{64}\right)}\Big|_0^x$$

$$\boxed{F(x) = -e^{\left(-\frac{x^3}{64}\right)} + 1}$$

Cihazın 2 ay çalışması sıfır
2 aya kadar çalışması;
$P(0<x<2) = F(x=2) - F(\ldots$

Yüzlük saatler cinsinden akülerin çalışma ömrü aşağıdaki yoğunluğa sahip X rastlantı değişkeniyle modelleniyor.

$$f(x) = \begin{cases} \frac{1}{2} e^{-\frac{x}{2}} & , x > 0 \\ 0 & , \text{diğer} \end{cases}$$

a) Bir akünün çalışma ömrünün 200 saatten az ve ya 400 saatten fazla olma olasılığı

b) Bir akünün 200 saatten fazla kullanıldığı bilindiğine göre 300 saatten fazla kullanılma olasılığı nedir?

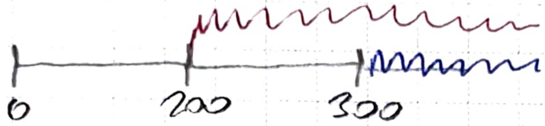
a) $P(A) = \int_0^2 \frac{1}{2} e^{-x/2}\,dx + \int_4^\infty \frac{1}{2} e^{-x/2}\,dx - 0$ kesişim

$= \frac{1}{2}(-2)e^{-x/2}\Big|_0^2 + \frac{1}{2}(-2)e^{-x/2}\Big|_4^\infty$

$= -(e^{-1}-1) - (0 - e^{-2})$

$= 1 - e^{-1} + e^{-2}$

b) $$P(x>3 \mid x>2) = \frac{P(x>3, x>2)}{P(x>2)}$$



$$= \frac{P(x>3)}{P(x>2)}$$

$$= \frac{\int_3^\infty \frac{1}{2} e^{-x/2} dx}{\int_2^\infty \frac{1}{2} e^{-x/2} dx} = \frac{-e^{-x/2} \Big|_3^\infty}{-e^{-x/2} \Big|_2^\infty} = \frac{e^{-3/2}}{e^{-1}}$$

Tanım: Ortalama Değer

Sürekli bir X rastlantı değişkeninin olasılık yoğunluk fonksiyonu (PDF) si f(x) olsun. X 'in ortalama değeri şu şekilde hesaplanır.

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

g(x), X 'e bağlı gerçel değerli bir fonksiyon ise g(x) 'in ortalama değeri şu şekilde hesaplanır.

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

Tanım: Varyans

Sürekli X rastlantı değişkeninin varyansı

$$\sigma^2 = var(x) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$$

$$= E[x^2] - (E[x])^2$$

Varyansın karekökü standart sapmayı verir.

NOT: Varyans ve ortalama değere ait ayrık durumda verilen özellikler sürekli durumda da aynen geçerlidir.

⟹

$$f(x) = \begin{cases} 3x^2 & , 0 \leq x \leq 1 \\ 0 & , \text{diğer} \end{cases}$$

Yoğunluk fonksiyonu verilen X rastlantı değişkeni için

a) Ortalama değeri ve varyansını hesaplayın

b) $y(x) = 4x+3$ fonksiyonunun ortalama değerini ve varyansını bulunuz.

a) $E[x] = \int_{-\infty}^{\infty} x.f(x)\,dx = \int_0^1 \underbrace{x\,3x^2}_{3x^3}\,dx = \frac{3x^4}{4}\Big|_0^1 = 3/4 = 0{,}75$

$var(x) = E[x^2] - (E[x])^2$

$E[x^2] = \int_0^1 \underbrace{x^2\,3x^2}_{3x^4}\,dx = \frac{3x^5}{5}\Big|_0^1 = 3/5$

$Var(x) = \frac{3}{5} - \left(\frac{3}{4}\right)^2 = 0{,}0375$

b) $u(x) = 4x+3$

$$E[4x+3] = E[4x] + E[3]$$
$$= 4E[X] + 3$$
$$= 4 \cdot \frac{3}{4} + 3$$
$$= 6$$

$$Var(4x+3) = 16\,var(x) + 0$$
$$= 16(0{,}0375)$$
$$= 0{,}6$$

[Sabitin ort değer kendisi varyansı sıfır]

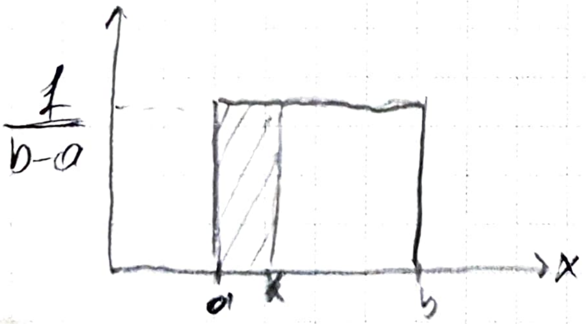
Önemli Sürekli Dağılımlar

1) Düzgün Dağılım

En basit sürekli dağılımlardan biri sürekli düzgün dağılımdır. [AB] aralığında olasılık sabittir. Bu aralıkta X rastlantı değişkeninin PDF si şu şekilde verilir.

$$f(x) = \begin{cases} \frac{1}{b-a} & , a \leq x \leq b \\ 0 & , \text{diğer} \end{cases}$$

Düzgün dağılımın CDF si ve ortalama değerini bulalım.



CDF

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

$$F(x) = \int_{a}^{x} \frac{1}{b-a}\,dt$$

$$= \frac{t}{b-a}\Big|_a^x = \frac{x-a}{b-a}$$

$$\boxed{F(x) = \frac{x-a}{b-a}}$$

Ortalama Değer

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

$$= \int_a^b x \frac{1}{b-a}$$

$$= \frac{1}{b-a} \frac{x^2}{2}\Big|_a^b$$

$$= \frac{b^2-a^2}{2(b-a)}$$

$$\boxed{E[x] = \frac{a+b}{2}}$$

Varyans

$$\sigma^2 = \frac{(b-a)^2}{12}$$

$$\sigma^2 = E[x^2] - (E[x])^2$$

$$E[x^2] = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_a^b x^2 \frac{1}{b-a}\,dx$$

$$= \frac{1}{b-a} \frac{x^3}{3}\Big|_a^b = \frac{b^3-a^3}{3(b-a)}$$

$$\sigma^2 = \frac{b^3-a^3}{3(b-a)} - \left(\frac{b+a}{2}\right)^2$$

$$\boxed{\sigma^2 = \frac{(b-a)^2}{12}}$$

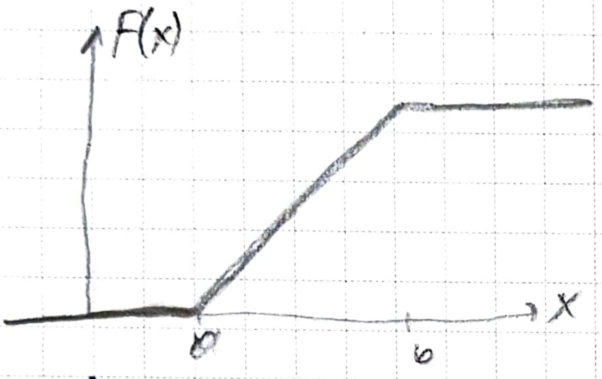
KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

TEKSAN

CDF'nin Grafiği

$$F(x) = \frac{x-a}{b-a} \qquad a \leq x \leq b$$

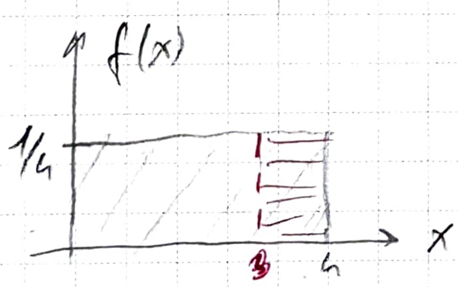


Bir şirketin konferans salonu 4 saatten fazla kullanılmıyor. Ancak salonda uzun yada kısa çok sayıda konferans veriliyor. X konferans uzunluğunun düzgün dağılımlı olduğu varsayılarak

a- X in PDF sini bulun

b- Her hangi bir konferansın en az 3 saat sürme olasılığı=?

a- $f(x) = \begin{cases} \frac{1}{4}, & 0 \leq x \leq 4 \\ 0, & \text{diğer} \end{cases}$



b- $P(X \geq 3) = \int_3^4 \underbrace{\frac{1}{4}}_{f(x)} dx = \frac{1}{4}$

⇒ Bir devre kartının bozulması sistemin çalışmasını yeni bir kart örnekilene kadar engelliyor. Üretim süresi X en az 1 en çok 4 gün arasında düzgün dağılımda dağılıyor. Bozulma maliyeti C ve sabit maliyet C_0 şu şekilde ilişkilendiriliyor.

$$C = C_0 + C_1 X^2$$

a- Üretim süresinin 2 gün ve ya daha fazla olması

b- Tek bir hata için toplam maliyeti bulunuz.

$$f(x) = \frac{1}{b-a} = \frac{1}{3}, \quad 1 \le x \le 4$$

a) $$P(X \geq 2) = \int_2^4 \frac{1}{3}\,dx = \frac{x}{3}\Big|_2^4 = \frac{2}{3}$$

b) $$E[C] = E[C_0 + C_1 x^2] = E[C_0] + C_1 E[X^2]$$

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

$$= \int_1^4 x^2 \frac{1}{3}\,dx = \frac{x^3}{9}\Big|_1^4 = \frac{64-1}{9} = \frac{63}{9} = 7$$

$$\boxed{E[C] = C_0 + C_1 . 7}$$

2) Üstel Dağılım

β parametreli üstel dağılımlı sürekli X rastlantı değişkeninin olasılık yoğunluk fonksiyonu şu şekilde verilir

$$f(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} & , x>0 \\ 0 & , \text{diğer} \end{cases}$$

$$E[X] = \beta = \int_0^\infty x \frac{1}{\beta} e^{-x/\beta} dx$$

$$var[X] = \beta^2$$

Üstel dağılımın CDF sını bulalım.

$$F(x) = \int_{-\infty}^{x} f(t) dt$$

$$F(x) = \int_0^x \frac{1}{\beta} e^{-t/\beta} dt$$

$$= \frac{1}{\beta}(-\beta) e^{-t/\beta} \Big|_0^x$$

$$= -e^{-x/\beta} + 1$$

$$\boxed{F(x) = 1 - e^{-x/\beta}}$$

NOT: Bir elemanın çalışma ömrü f yoğunluklu X rastlantı değişkeni ile modellensin. hata fonksiyonu r(t) şu şekilde verilir.

$$r(t) = \frac{f(t)}{1-F(t)} = \frac{\frac{1}{\beta} e^{-t/\beta}}{1-(1-e^{-t/\beta})} = \frac{1}{\beta}$$

Hata oranının ($\mu = \lambda = \frac{1}{\beta}$) elemanın yaşına bağlı olmadığına dikkat ediniz. Buna üstel dağılımın hafızasızlık özelliği denir.

NOT Bir olayın λ hızında gerçekleştiğini varsayalım. Bu durumda $[0,t]$ aralığında ortalama olay sayısı $\mu = \lambda . t$ olur ve poisson dağılımı ile modellenir. Üstel dağılım ise sabit hata oranına sahip ard arda gerçekleşen olaylar arasındaki sürenin modellenmesi için uygundur.

⇒ Arızadan dolayı cihazın çalışmadığı sürenin $\beta = 6$ saat parametreli üstel dağılım ile modellendiğini varsayalım. Bir sonraki arıza aralığının 5 ile 10 saat arasında olma olasılığı nedir.

$\beta = 6$ saat

$$f(x) = \frac{1}{\beta} e^{-x/\beta} = \frac{1}{6} e^{-x/\beta}$$

$$P(5 < x < 10) = \int_5^{10} \frac{1}{6} e^{-x/6} dx$$

$$P(5 < x < 10) = F(10) - F(5)$$

$$F(x) = 1 - e^{-x/\beta}$$

$$F(10) = 1 - e^{-10/6}$$

$$F(5) = 1 - e^{-5/6}$$

0,2457

$P(5<x<10) = \left(1 - e^{-10/6}\right) -$

3- Normal Dağılım

Sürekli dağılımlar arasında en yaygın olarak kullanılan dağılım, Gauss dağılımı olarak da bilinen normal dağılımdır.

Bu dağılım ölçüm hataları, market stok hareketlilikleri, yapay zeka, otomatik kontrol gibi bir çok alanda kullanılan popüler bir modeldir.

Normal dağılım çan eğrisi şeklinde yapıya sahiptir. Çan eğrisinin merkezi ortalama değer, yayılımı ise standart sapma ile oluşur.

Normal dağılımlı X rastlantı değişkeninin PDF'si

ortalama değer

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right], \quad -\infty<x<\infty$$

standart sapma — varyans

$$X \sim N[\mu, \sigma^2]$$

$$\Rightarrow X \sim N[0, 25] \qquad f(x) = \frac{1}{5\sqrt{2\pi}} e^{-\frac{(x-0)^2}{2(25)}} = \frac{1}{\sqrt{2\pi}\,5} e^{-x^2/50}$$

$\mu = 0$, $\sigma^2 = 25$

$$P(a<x<b) = \int_a^b f(x)\,dx \quad ?$$

(Integral bulunamaz buradan Standart Normal dağılım)

Standart Normal Dağılım

Ortalama değeri sıfır, varyansı 1 olan normal dağılıma standart normal dağılım denir ve z değişkeni ile gösterilir.

$\mu = 0, \ \sigma^2 = 1$

$$\boxed{f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}}$$

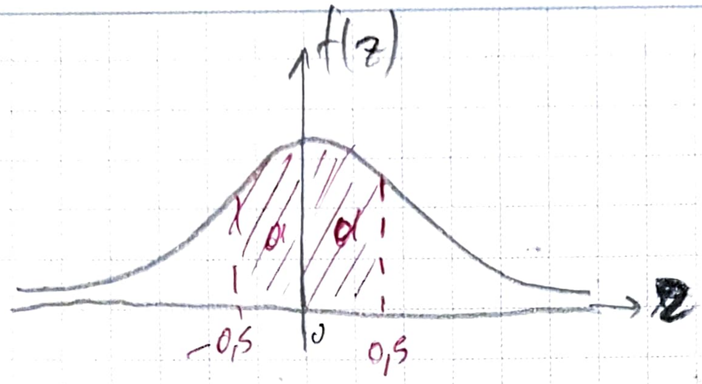
*(Tablolar)

$z - [0,1]$

Gausu → standart gaus → Tablolar

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:



Tablo okuma

$P(0<z<a)$

0,005
0,22
0,35

0,3 0,05

z	0,00	0,01	0,02	...	0,05
0,0					X
0,1					
0,2			X		
0,3					X

⇒ Normal dağılımlı X rastlantı değişkeni Standart normal dağılıma nasıl dönüştürülür?

Herhangi bir X normal dağılımını gözlemi standart normal dağılım z ye dönüştürmek için aşağıdaki ifade kullanılır.

$$z = \frac{X-\mu}{\sigma}$$

⇒ $X \sim N[50,100]$, $P(45<X<62)=?$

$$z = \frac{X-\mu}{\sigma}$$

$$z_1 = \frac{45-50}{10} = -0,5$$

$$z_2 = \frac{62-50}{10} = 1,2$$

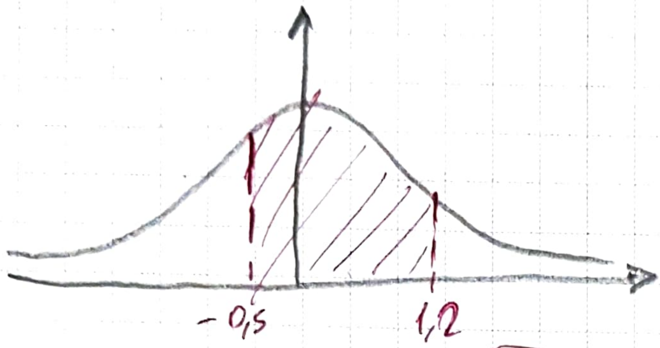
$P(-0,5<z<1,2)=?$

Tabloda pozitifler var.

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

TEKSAN



$$P(-0{,}5 < z < 1{,}2) = \underbrace{\boxed{P(-0{,}5 < z < 0)}}_{\text{Tablo } z=0{,}5} + \underbrace{\boxed{P(0 < z < 1{,}2)}}_{\text{Tablo } z=1{,}2}$$

$$= P(z=0{,}5) + P(z=1{,}2)$$

$$= 0{,}1915 + 0{,}3849$$

$$= 0{,}5764$$

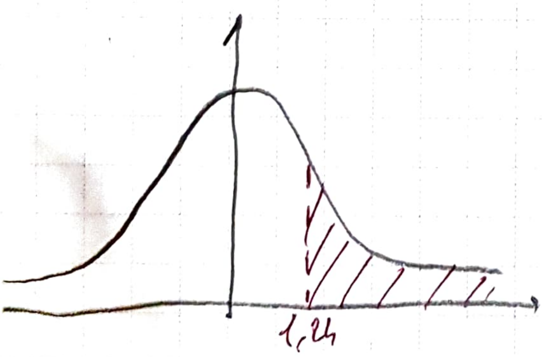
⟹ X rastlantı değişkeni 300 ortalamalı 50 standart sapmalı gauss dağılımlı olsun. X'in 362'den büyük olma olasılığını hesaplayınız.

$$X \sim N[300, 2500]$$

$$z = \frac{X - 300}{50}$$

$$P(x > 362) \longrightarrow P(z > 1{,}24)$$

$$z = \frac{362 - 300}{50} = 1{,}24$$



$$P(z > 1{,}24) = 0{,}5 - \underbrace{P(0 < z < 1{,}24)}_{z=1{,}24}$$

$$= 0{,}1075$$

www.teksanjenerator.com.tr

KONU / SUBJECT: Z yi veriyor X i buluyor

TARİH / DATE:

SAAT / TIME: (tersi)

TEKSAN

Tanım: Percentile

X rastlantı değişkeninin P ıncı persentili % P/100 kadarlık alanı geride bırakan q noktasıdır. Şu şekilde tanımlanır.

$P(X \leq q) = \% \frac{P}{100}$ örneğin 50 inci persantil olasılık dağılımının medyan değerine karşılık gelir.

$X = \{-100, -25, -10, 0, 10, 20, 30\}$

medyan

$P(X \leq q) = 0{,}25$

z	0,00	0,06
0,0		
0,1		
0,0		0,25

ters okuma

$P(0 \leq z \leq z_0) = 0{,}49$ ise $z_0 = ?$

z	0,03
2,3	0,49

$z_0 = 2{,}33$

$P(0 \leq z \leq 2{,}33)$

$$2{,}33 = \frac{x - \mu}{\sigma}$$

$$x = \sigma(2{,}33) + \mu$$

Tanım: %68 - %95 kuralı

Normal bir dağılım sahip rastlantı değişkenin aldığı değerlerin %68'i $[\mu-\sigma, \mu+\sigma]$ aralığında yer alır.

%95 ise $[\mu-2\sigma, \mu+2\sigma]$ aralığında yer alır.

%99,7'si $[\mu-3\sigma, \mu+3\sigma]$ aralığındadır.

Normal Dağılım ile Binom Dağılımın İlişkisi

X rastlantı değişkeni ortalama değeri $\mu = np$ ve varyansı $\sigma^2 = npq$ olan binom rastlantı değişkeni olsun

$$\boxed{Z_n = \frac{X-np}{\sqrt{npq}}}$$

ile verilen rastlantı değişkeni n yeterince büyük olduğunda standart normal dağılıma yakınsar.

Standart gauss $Z = \dfrac{\overset{\text{binom}}{X} - \mu}{\sigma}$

⇒ X rastlantı değişkeni $n = 15$, $p = 0{,}4$ olan binom rastlantı değişkeni olsun. $P(7 \le X \le 9)$ olasılığını binom dağılımı ile ve standart normal dağılım ile hesaplayınız?

Binom dağılımı

$$P(X=k) = \binom{n}{k} p^k \cdot q^{n-k}$$

$$P(7 \le X \le 9) = \sum_{k=7}^{9} \binom{15}{k} (0{,}4)^k \cdot (0{,}6)^{15-k}$$

$$= \boxed{0{,}3564}$$

$$\mu = n.p = 15.(0,4)$$

$$\sigma^2 = npq = 15(0,4)(0,6)$$

$$z_n = \frac{x - 15(0,4)}{\sqrt{15(0,4)(0,6)}}$$

$$z_n = \frac{x-6}{1,897}$$

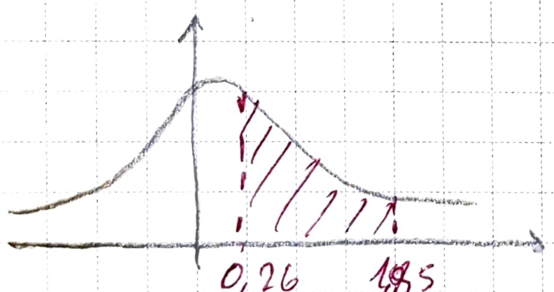
$$\left.\begin{array}{l} x_1 = 6,5 \\ x_2 = 9,5 \end{array}\right\}$$

0,5 üste ve 0,5 alta alınır 7 ve 9'u kapsasın diye

$$z_1 = \frac{6,5-6}{1,897} = 0,26$$

$$z_2 = \frac{9,5-6}{1,897} = 1,85$$

$$\boxed{\overset{\text{binom}}{P(7 \le x \le 9)} \longrightarrow \overset{\text{normal}}{P(0,26 \le z \le 1,85)}}$$



$$= P(z = 1,85) - P(z = 0,26) = \boxed{0,3652}$$

NOT: Örneklerde görüldüğü gibi standart normal dağılım ile binom dağılımı arasında yakın bir ilişki vardır. n değeri yeterince büyük seçildiğinde binom dağılımı normal dağılıma yakınsar.

⇒ Bir bireyin nadir görülen bir kan hastalığına yakalanma olasılığı 0,4 dır. 100 kişinin hastalığa yakalandığı biliniyorsa en çok 30 kişinin hasta olma olasılığı

$n = 100 \ , \ p = 0{,}4$

- $P(X \leq 30) = \sum_{k=0}^{30} \binom{100}{k} (0{,}4)^k (0{,}6)^{100-k}$ (Binom)

- $\mu = np = 100\,(0{,}4) = 40$

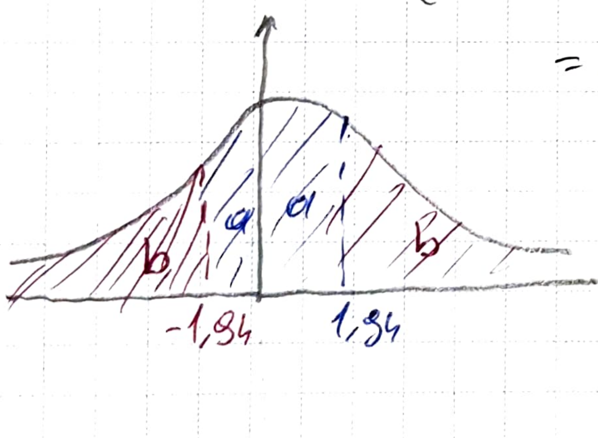
$\sigma^2 = npq = 100 \cdot 0{,}4 \cdot 0{,}6 = 24$

$$z = \frac{X - 40}{\sqrt{24}}$$

$$z = \frac{30{,}5 - 40}{\sqrt{24}} = -1{,}94$$

$P(X \leq 30) \longrightarrow P(z < -1{,}94)$

$= 0{,}5 - P(z = 1{,}94)$



KONU / SUBJECT: Vize Örnekleri

TARİH / DATE:

SAAT / TIME:

TEKSAN

⇒ X rastlantı değişkeni üniversite kütüphanesinde 1 saatlik periyotta kitap alan öğrenci sayısı olsun X değişkeninin 15 dk da ortalama 4,2 kitap alımı ile modellenen bir dağılıma sahip olduğu var sayılarak

a) $P(2<X<6)$

b) $P(X>4)$

Poisson

15 dakikada 4,2 → ortalama değer tam sayı olmayabilir.

60 dk x

$\mu = 4,2 \times 4 = 16,8$

$$P[X=k] = \frac{\mu^k}{k!} e^{-\mu} \quad \text{(Poisson)}$$

$$P[X=k] = \frac{(16,8)^k}{k!} e^{-16,8}$$

a) $P(2<X<6) = \sum_{k=3}^{5} \frac{(16,8)^k}{k!} e^{-16,8}$ böyle bırak! AMK

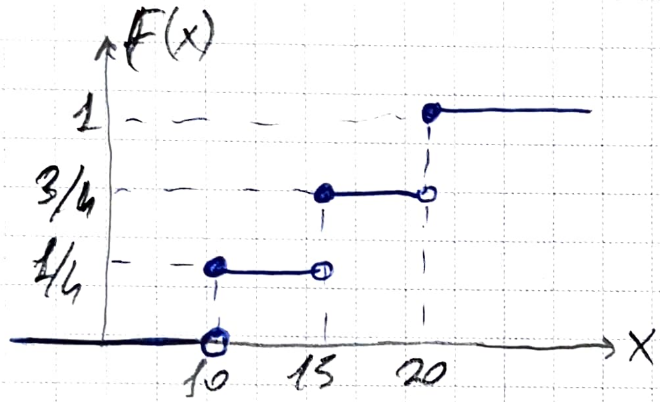
b) $\underbrace{P(X>4)}_{\sum_{k=5}^{\infty}} = 1 - P(X \leq 4) = 1 - \sum_{k=0}^{4} \frac{(16,8)^k}{k!} e^{-16,8}$

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

⇒ Toplam dağılım fonksiyonu (CDF) aşağıda verilen X rastlantı değişkeninin ortalama değerini ve varyansını bulunuz?

$$F(x) = \begin{cases} 0, & x < 10 \\ 1/4, & 10 \leq x < 15 \\ 3/4, & 15 \leq x < 20 \\ 1, & 20 \leq x \end{cases}$$



$$\left.\begin{aligned} P(x=10) &= 1/4 \\ P(x=15) &= 3/4 - 1/4 = 1/2 \\ P(x=20) &= 1 - 3/4 = 1/4 \end{aligned}\right\} \text{PMF}$$

$$E[x] = \sum_x x.P(x)$$

$$= 10.\frac{1}{4} + 15.\frac{1}{2} + 20.\frac{1}{4} = 15$$

$$Var(x) = E[x^2] - E[x]^2$$

$$E[x^2] = \sum_x x^2 P(x) = 10^2\frac{1}{4} + 15^2\frac{1}{2} + 20^2\frac{1}{4} = 237,5$$

$$Var(x) = (237,5) - (15)^2 = 12,5$$

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

TEKSAN

⇒ X rastlantı değişkeninin yoğunluk fonksiyonu aşağıda veriliyor.

$$f(x) = \begin{cases} cx(1-x^2) & , 0 \leq x \leq 1 \\ 0 & , \text{diğer} \end{cases}$$

a) Fonksiyonun geçerli bir yoğunluk fonksiyonu olmasılığını $c = ?$

b) X in toplam dağılım fonksiyonunu bulunuz.

c) $P[0 < x < 0,5]$, $P[x = 1]$, $P[0,25 < x < 1,5]$ olasılıklarını hesaplayınız?

a) $\int_{-\infty}^{\infty} f(x)dx = 1$

$\int_{0}^{1} cx(1-x^2)\,dx = 1$

$\frac{cx^2}{2} - \frac{cx^4}{4}\Big|_0^1 = 1$

$c = 4$

b) $F(x) = \int_{-\infty}^{x} f(t)\,dt$

$F(x) = \int_{0}^{x} 4t(1-t^2)\,dt$

$= \left(\frac{4t^2}{2} - \frac{4t^4}{4}\right)\Big|_0^x$

$\boxed{F(x) = 2x^2 - x^4}$

Çizmek istersek $0 \leq x \leq 1$ aralığında çizeriz

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

c) $P[0<x<0{,}5] = \int_0^{0{,}5} 4x(1-x^2)\,dx$

$= F(0{,}5) - F(0)$

$P[x=1]=0$

$P[0{,}25<x<1{,}5] = \int_{0{,}25}^{1} 4x(1-x^2)\,dx$

$= \underset{1}{F(1{,}5)} - F(0{,}25)$

sınırlar $0 \le x \le 1$ arasında olduğundan (–) değerler ve 1'den büyükler 1 ve 0 kabul

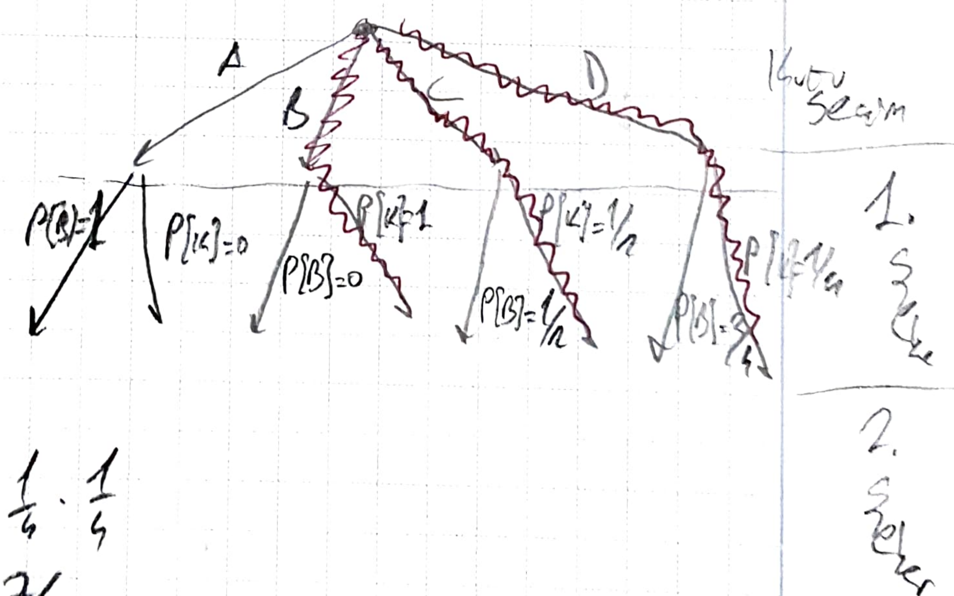
$\Rightarrow$ A kutusunda iki beyaz B kutusunda iki kırmızı C kutusunda iki beyaz iki kırmızı D kutusunda 3 beyaz 1 kırmızı şeker vardır. Kutuların eşit olasılıklı olduğunu varsayarak

a) Seçilen şekerin kırmızı olma olasılığı

b) Seçilen şekerin kırmızı olduğu bilindiğine göre D kutusundan gelme olasılığı

A → 3B
B → 2K
C → 2B 2K
D → 3B 1K

$P[A]=P[B]=P[C]=P[D]=1/4$



a) $P[K] = \frac{1}{4}\cdot 1 + \frac{1}{4}\cdot\frac{1}{2} + \frac{1}{4}\cdot\frac{1}{4}$

$= \frac{1}{4} + \frac{1}{8} + \frac{1}{16} = 7/16$

KONU / SUBJECT:

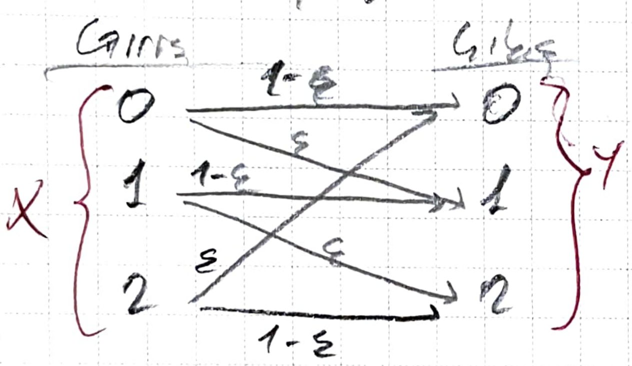
TARİH / DATE: SAAT / TIME:

b) $P[D/K] = \dfrac{P[D,K]}{P[K]} = \dfrac{P[K/D]\,P[D]}{P[K]} = \dfrac{1/4 \cdot 1/4}{7/16} = 1/7$

⇒ Şekilde gibi bir haberleşme sistemi verilsin. sıfır 1 ve 2 sembollerinin eşit olasılıkla ortaya çıktığını varsayalım.

a) Çıkış sembollerinin olasılıklarını hesaplayınız.

b) Çıkışta 1 gözlemlendiğini varsayarak girişin 0, 1, 2 olma olasılıklarını hesaplayınız?



ε = hata olasılığı

$P[0] = P[1] = P[2] = 1/3$

a) $P[Y=0] = P(X=0, Y=0) + P(X=2, Y=0)$

$= P(Y=0 / X=0) \cdot P(X=0) + P(Y=0 / X=2)\, P(X=2)$

$= (1-\varepsilon)\dfrac{1}{3} + \varepsilon \cdot \dfrac{1}{3} = \dfrac{1}{3}$

$P[Y=1] = \underbrace{P(Y=1/X=0) \cdot P(X=0)}_{P(X=0,\,Y=1)} + \underbrace{P(Y=1/X=1)\, P(X=1)}_{P(X=1,\,Y=1)}$

$= \varepsilon \dfrac{1}{3} + (1-\varepsilon)\dfrac{1}{3} = 1/3$

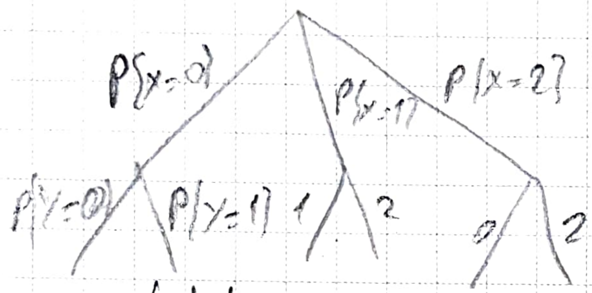
$P[Y=2] = P[Y=2/X=1]\, P(X=1) + P(Y=2/X=2)\, P(X=2)$

$= \varepsilon \dfrac{1}{3} + (1-\varepsilon)\dfrac{1}{3} = 1/3$

b) $P[X=0/Y=1] = \dfrac{P(X=0, Y=1)}{P(Y=1)} = \dfrac{P[Y=1/X=0]\,P(X=0)}{P(Y=1)} = \varepsilon$

$P[X=1/Y=1] = \dfrac{P[Y=1/X=1] \cdot P[X=1]}{P[Y=1]} = \dfrac{(1-\varepsilon)1/3}{1/3} = 1-\varepsilon$

$P[X=2/Y=1] = \dfrac{P[Y=1/X=2]\,P[X=2]}{P[Y=1]} = 0$



⇒ Bir toplulukta sigara içme oranı %60 dır. Topluluktan rastgele 10 kişi seçelim

n denemede k başarı Binom

a) 3 kişinin sigara içme
b) En az iki kişinin sigara içme

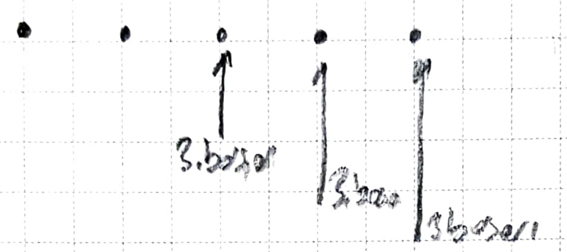
$p=0,6 \quad q=0,4 \quad n=10$

a) $P[k=3] = \binom{10}{3}(0,6)^3(0,4)^7$

b) $P[k \geq 2] = \sum_{k=2}^{10} \binom{10}{k}(0,6)^k(0,4)^{10-k}$

$= 1 - \sum_{k=0}^{1} \binom{10}{k}(0,6)^k(0,4)^{10-k}$

⇒ 5 maçlık bir seride en çok maçı kazanan şampiyon olmaktadır. A takımının oynadığı bir oyunu kazanma olasılığı %60 dır. A takımının şampiyon olma olasılığını bulunuz?



y denemede r başarı

$\binom{y-1}{r-1} p^r q^{y-r}$ $y=r, r+1, \ldots$ (sınav kağıdı olacak)

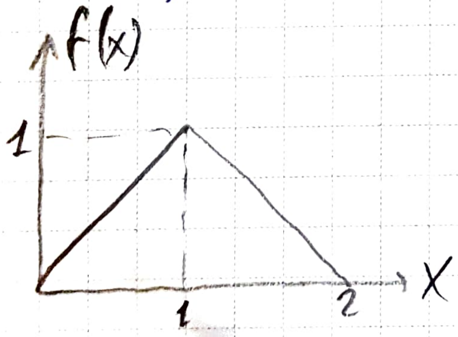
$$P\{A\ \text{ş}\} = \underbrace{\binom{2}{2}(0,6)^3(0,4)^0}_{\substack{y=3\\ r=3}} + \underbrace{\binom{3}{2}(0,6)^3(0,4)^1}_{\substack{y=4\\ r=3}} + \underbrace{\binom{4}{2}(0,6)^3(0,4)^2}_{\substack{y=5\\ r=3}}$$

⇒ X rastlantı değişkeninin yoğunluk fonksiyonu şekilde gösterilmektedir.

a) X'in CDF sini bulunuz?

b) X'in moment çıkaran fonksiyonunu bul?

c) X'in ortalama değerini hesaplayınız.



$$f(x) = \begin{cases} x, & 0 \le x < 1 \\ 2-x, & 1 \le x < 2 \\ 0, & \text{diğer} \end{cases}$$

$$m = \frac{y-y_1}{x-x_1}$$

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

a)

$$F(x)=\begin{cases} 0, & x<0 \\ x^2/2, & 0\le x\le 1 \\ 2x-x^2/2-1, & 1\le x<2 \\ 1, & x\geq 2 \end{cases}$$

(eşitsizlik ade önemli değil)

$$F(x)=\int_{-\infty}^{x} f(t)\,dt=\underbrace{\int_0^1 t\,dt}_{1/2}+\int_1^x (2-t)\,dt \quad , 1\le x\le 2$$

$$*F(x)=\int_0^x t\,dt \quad , 0\le x\le 1 \qquad F(x)=\frac{x^2}{2}$$

$$F(x)=\frac{1}{2}+\left(2t-\frac{t^2}{2}\right)\Big|_1^x$$

$$=\frac{1}{2}+\left(2x-\frac{x^2}{2}\right)-\frac{3}{2}$$

$$=2x-\frac{x^2}{2}-1$$

b)

$$m(t)=E[e^{tx}]=\int_{-\infty}^{\infty} e^{tx} f(x)\,dx$$

$$m(t)=\int_0^1 e^{tx}\cdot x\,dx+\int_1^2 e^{tx}(2-x)\,dx$$

$$\frac{d}{dt}m(t)\Big|_{t=0}=E[X]$$

c)

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

$$=\int_0^1 x(x)\,dx+\int_1^2 x(2-x)\,dx$$

modülasyon → sinyali uzak yere
demodülasyon

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

TEKSAN

⇒ Bir modem kanala +2 Voltluk işaret veriyor. Kanal bu işarete olasılıkları sırasıyla 4/10, 3/10, 2/10, 1/10 olan $\{0, -1, -2, -3\}$ kümesinden rastgele bir gürültü ekliyor. Kanal çıkışı Y olsun.

a) Y nin PMF sini bulunuz.

b) Kanal girişinin çıkışına eşit olma olasılığı nedir.

c) Kanal çıkışının pozitif olma olasılığı

d) Y nin ortalama değerini ve varyansını bulunuz?

	Gürültü	Y	
+2V	0	2V	4/10
	-1	1V	3/10
	-2	0V	2/10
	-3	-1V	1/10

$P[Y=2] = 4/10$

$P[Y=1] = 3/10$

$P[Y=0] = 2/10$

$P[Y=-1] = 1/10$

KONU / SUBJECT: FINAL
TARİH / DATE: SAAT / TIME:

TEKSAN

ORTAK OLASILIK DAĞILIMLARI

Daha önce tartışılan rastlantı değişkenlerinin tamamı tek boyutlu idi. Yani rastlantısal nicelikleri bir zaman anında tek başına inceledik. Bazı durumlarda birden fazla rastlantı değişkenini aynı anda ölçmek isteyebiliriz. Örneğin kimyasal bir deneyde açığa çıkan gazın hacmini ve yoğunluğunu ölçmek isteyebiliriz. Bu deney bize iki boyutlu bir rastlantı değişkeni verir.

Hacim: V
Yoğunluk: D olsun

$P(V,D)$ ayrıksa PMF
$f(V,D)$ sürekli ise PDF

Tanım: Ortak PMF

Eğer X ve Y ayrık rastlantı değişkeni ise $X=x$ ve $Y=y$ iken ortak olasılık kütle fonksiyonu $P(X,Y)=P(X=x,Y=y)$ ile gösterilir.
X ve Y değişkenlerinin ortak davranışı hakkında bilgi verir ve aşağıdaki özellikleri sağlar.

1- $P(x,y)\geqslant 0$ tüm (x,y) çifti için

2- $\sum\limits_{x}\sum\limits_{y} P(x,y)=1$

Tanım: Marjinal PMF

X ve Y nin ayrı ayrı marjinal olasılık fonksiyonları şu şekilde hesaplanır.

$$P(x)=\sum_{y} P(x,y)$$

$$P(y)=\sum_{x} P(x,y)$$

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

$P(x,y)$

X \ Y	0	1	3
0	0,1	0,3	0,2
1	0,1	0,2	0,1

$P(X=0)$

$P(x=0, y=3)$

$P(X=0, Y=0)$

X	

Y	

Marjinal [illegible]

$$P(X=0) = P(X=0, Y=0) + P(X=0, Y=1) + P(X=0, Y=3) = 0,6$$

$$P(X=1) = P(X=1, Y=0) + P(X=1, Y=1) + P(X=1, Y=3) = 0,4$$

$$P(Y=0) = P(X=0, Y=0) + P(X=1, Y=0) = 0,2$$

$$P(Y=1) = 0,5$$

$$P(Y=3) = 0,3$$

Tanım: Ortak Yoğunluk ve Marjinal Yoğunluk

X ve Y sürekli rastlantı değişkenlerinin ortak olasılık yoğunluk fonksiyonları (PDF) $f(x,y)$ ile gösterilir. Ve şu özellikleri sağlar.

1- $f(x,y) \geq 0$ tüm x ve y çiftleri için

2- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$

Marjinal PDF

$$f(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$ y'nin sınırları

$$f(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$ x'in sınırları

⇒ Endüstriyel bir kimya fabrikası temel iki katkı maddesi içeren bir ürün üretmektedir. X ve Y katkı maddelerinin konsantrasyonlarının (mg/lt) ortak dağılımları şu şekilde veriliyor.

$$f(x,y)=\begin{cases} 2(1-x) & ,0<x<1 \ ,0<y<1 \\ 0 & ,\text{diğer} \end{cases}$$

a- Bu geçerli bir dağılım mıdır?
b- Marjinal yoğunlukları bulunuz
c- $P(0<x<0{,}5 \ ,0{,}4<y<0{,}7)$ olasılığını hesapla

a- $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy=1$ olmalı

$$\int_0^1\int_0^1 2(1-x)\,dx\,dy \Rightarrow 2x-x^2\Big|_0^1=1 \ , \ \int_0^1 1\,dy=y\Big|_0^1=1$$

b- $f(x)=\int f(x,y)\,dy$

$$=\int_0^1 2(1-x)\,dy$$

$$=2y-2xy\Big|_0^1$$

$f(x)=2-2x$

$f(y)=\int_0^1 2(1-x)\,dx$

$$=2x-x^2\Big|_0^1=1$$

$f(y)=1$

c-

$$P(0<x<0{,}5 \ , \ 0{,}4<y<0{,}7)=\int_{0{,}4}^{0{,}7}\int_0^{0{,}5} 2(1-x)\,dx\,dy$$

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

TEKSAN

Koşullu Olasılık Dağılımları

X ve Y rastlantı değişkenlerinin sürekli ve ayrık durum için koşullu olasılıkları şu şekilde hesaplanır.

$$P(X=x, Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)}$$ ayrık

$$f(x/y) = \frac{f(x,y)}{f(y)}, \quad f(y) \neq 0$$

$$f(y/x) = \frac{f(x,y)}{f(x)}, \quad f(x) \neq 0$$

sürekli

X ⇒ X birim sıcaklık değişimi ve Y spektrum kayma oranı olmak üzere X,Y rastlantı değişkeni çiftinin ortak yoğunluk fonksiyonu veriliyor

$$f(x,y) = \begin{cases} 10xy^2, & 0<x<y<1 \\ 0, & \text{diğer} \end{cases}$$

a- Marjinal yoğunlukları bulunuz

b- Koşullu yoğunlukları bulunuz

a- $f(x) = \int f(x,y)\,dy = \int_x^1 10xy^2\,dy = \frac{10xy^3}{3}\Big|_x^1 = \frac{10x}{3} - \frac{10x^4}{3}$

$= \frac{10}{3}(x - x^4)$

$f(y) = \int f(x,y)\,dx = \int_0^y 10xy^2\,dx = 5x^2y^2\Big|_0^y = 5y^4$

b- $f(x/y) = \dfrac{f(x,y)}{f(y)}$

$= \dfrac{10xy^2}{5y^4} = 2xy^{-2}$

$f(y/x) = \dfrac{f(x,y)}{f(x)}$

$= \dfrac{10xy^2}{\frac{10}{3}(x-x^4)} = \dfrac{3y^2}{1-x^3}$

NOT:

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

$= \int_0^1 x \frac{10}{3}(x-x^4) dx$

$= \int_0^1 \frac{10}{3}(x^2-x^5) dx$

$= \frac{10}{3}\left(\frac{x^3}{3} - \frac{x^6}{6}\right)\Big|_0^1$

$= \frac{10}{3}\left(\frac{1}{3} - \frac{1}{6}\right)$

$= \frac{10}{18}$

[Integral sınırları ortadan değerde genel alınır]

$E[Y] = \int_0^1 y 5y^4 dy$

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

TEKSAN

Tanım: Bağımsızlık

X ve Y rastlantı değişkenleri aşağıdaki koşulu sağlıyorsa istatiksel olarak bağımsızdır denir.

$P(x,y) = P(x).P(y)$ Ayrık durum için

$f(x,y) = f(x).f(y)$ sürekli durum için

⇒ Ortak yoğunluk fonksiyonları verilen X ve Y rastlantı değişkenleri bağımsız mıdır

$f(x,y) = 2(1-x) \quad 0<x<1 \; , \; 0<y<1$

$f(x) = \int_0^1 2(1-x)\,dy = 2y - 2xy \Big|_0^1 = 2-2x$

$f(y) = \int_0^1 2(1-x)\,dx = 2x - x^2 \Big|_0^1 = 1$

Bağımsızlık için $f(x,y) = f(x).f(y)$ olmalı

$2(1-x) \stackrel{?}{=} (2-2x).1$

eşit olduğundan X ve Y bağımsızdır

⇒ X_1 toplam satış sayısı ve X_2 satılan elektronik malzeme sayısı olsun X_1 ve X_2 nin ortak davranışı için aşağıdaki tablo veriliyor

X_1 \ X_2	0	1	2
0	0,1	0	0
1	0,1	0,2	0
2	0,1	?	0,15

a- Soru işaretinin değeri = ?

b- Marjinal fonksiyonları bulunuz

c- $X_1 \leq 1$ ve $X_2 \leq 1$ olasılığını hesaplayınız

d- $X_1 = 2$ koşulu altında X_2'nin olasılıklarını hesapla

e- X_1 ve X_2 bağımsız mıdır

a,b) $P(x) = \sum_{y} P(x,y)$

$P(y) = \sum_{x} P(x,y)$

$P(x_1=0) = 0{,}1+0+0 = 0{,}1$

$P(x_1=1) = 0{,}1+0{,}2+0 = 0{,}3$

$P(x_1=2) = 0{,}1+?+0{,}15 = 0{,}6$

$\left.\right\} 1$

$? = 0{,}35$

$P(x_2=0) = 0{,}3$

$P(x_2=1) = 0{,}55$

$P(x_2=2) = 0{,}15$

$\left.\right\} 1$ // sağlama

c- $P(x_1 \leq 1, x_2 \leq 1) = P(x_1=0, x_2=1) + P(x_1=1, x_2=0) + P(x_1=1, x_2=1) + P(x_1=0, x_2=0)$

$= 0 + 0{,}1 + 0{,}2 + 0{,}1 = 0{,}4$

d- $X_1 = 2$ koşulu

$P(x_2=0 / x_1=2) = \dfrac{P(x_1=2, x_2=0)}{P(x_1=2)} = \dfrac{0{,}1}{0{,}6} = \dfrac{1}{6}$

$P(x_2=1 / x_1=2) = \dfrac{P(x_1=2, x_2=1)}{P(x_1=2)} = \dfrac{0{,}35}{0{,}6}$

$P(x_2=2 / x_1=2) = \dfrac{P(x_2=2, x_1=2)}{P(x_1=2)} = \dfrac{0{,}15}{0{,}6}$

e- $P(x_1, x_2) \stackrel{?}{=} P(x_1) P(x_2)$

$P(x_1=0, x_2=0) \stackrel{?}{=} P(x_1=0) \cdot P(x_2=0)$

$0{,}1 \stackrel{?}{=} (0{,}1)(0{,}3)$

bağımsız değil

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

TEKSAN

ÖDEV

$$f(x,y) = \begin{cases} k, & 0 \leq x \leq y \leq 1 \\ 0, & \text{diğer} \end{cases}$$

X ve Y rastlantı değişkenlerinin ortak davranışı yukarıda verilmiştir.

a) Bu ifadenin geçerli olabilmesi için k değeri ne olmalıdır.

b) X ve Y'nin marjinal yoğunluklarını bulunuz.

c) Ortalama değerleri hesaplayınız.

ÖDEV $[0,1] \times [0,1]$ boyutlarında bir kareden birbirinden bağımsız iki koordinat seçiliyor. Bu koordinatlar X ve Y olsun

a) Seçilen noktanın orjine uzaklığının 1'den küçük olma olasılığı $(P(X^2+Y^2<1))$

b) $Y=0{,}5$ olduğunda göre X'in koşullu yoğunluğunu bulunuz?

Tanım: Fonksiyonun Ortalama Değeri

X ve Y birer rastlantı değişkeni $g(x,y)$ gerçel değerli bir fonksiyon olsun. Fonksiyonun ortalama değeri şu şekilde hesaplanır.

$$E[g(x,y)] = \sum_x \sum_y g(x,y) P(x,y) \longrightarrow \text{X ve Y Ayrık}$$

$$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) . f(x,y) \, dx \, dy \longrightarrow \text{X ve Y Sürekli}$$

Tanım: Kovaryans

X ve Y rastlantı değişkenlerinin kovaryansı şu şekilde hesaplanır

$$cov(X,Y) = E[(x-\mu_x)(y-\mu_y)], \quad \mu_x = E[X], \quad \mu_y = E[Y]$$

Kovaryans bize iki rastlantı değişkeni arasındaki ilişki hakkında bilgi verir. Kovaryans pozitifse X ile Y arasında pozitif bir ilişki vardır. Yani X artarken Y de artar.

$$cov(X,X) = E[(x-\mu_x)^2] = var(X)$$

Tanım: Korelasyon

X ve Y rastlantı değişkenleri arasındaki korelasyon katsayısı

$$\rho = \frac{cov(X,Y)}{\sqrt{var(X)\,var(Y)}}$$

Özellikleri:

1- Korelasyon katsayısı (ρ) -1 ve $+1$ arası değerler alır $[-1,+1]$

2- Korelasyon katsayısı boyutsuzdur. (Skalerlik)

3- $\rho = -1$ ve ya $\rho = +1$ ise Y, X'in doğrusal bir fonksiyonudur.

4- X ya da Y doğrusal bir işleme tabi tutulduğunda korelasyon katsayısı değişmez.

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

TEKSAN

$$\mathrm{cov}(X,Y) = E[(x-\mu_x)(y-\mu_y)]$$

$$= E[XY - X\mu_y - y\mu_x + \mu_x\mu_y]$$

$$= E[XY] - E[X\mu_y] - E[Y\mu_x] + E[\mu_x\mu_y]$$

$$= E[XY] - E[Y].E[X] - \cancel{E[X]E[Y]} + \cancel{E[X]E[Y]}$$

$\mu_y = E[Y]$
$\mu_x = E[X]$

$$\boxed{\mathrm{cov}(X,Y) = E[X,Y] - E[X]E[Y]}$$

$\Rightarrow$

$$f(x,y) = \begin{cases} 8xy & , 0 \le x \le 1 \quad 0 \le y \le 1 \\ 0 & , \text{diğer} \end{cases}$$

X maratonu tamamlayan erkek koşucu oranı Y ise bayan koşucu oranı olmak üzere X ve Y nin ortak davranış yukarıda verilmiştir. X ve Y nin korelasyonunu ve kovaryansını hesaplayınız.

$$\mathrm{cov}(X,Y) = E[XY] - E[X]E[Y]$$

$$E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x,y)\, dx\, dy$$

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

$$E[Y] = \int_{-\infty}^{\infty} y f(y)\, dy$$

$$f(x) = \int_{-\infty}^{\infty} f(x,y)\, dy$$

$$f(x) = \int_{0}^{1} 8xy\, dy$$

$$= 4xy^2 \Big|_0^1 = 4x$$

$$f(y) = \int_{-\infty}^{\infty} f(x,y)\, dx$$

$$f(y) = \int_{0}^{1} 8xy\, dx$$

$$= 4x^2y \Big|_0^1 = 4y$$

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

TEKSAN

$$E[X] = \int_0^1 x\,4x\,dx = \frac{4x^3}{3}\Big|_0^1 = \frac{4}{3}$$

$$E[Y] = \int_0^1 y\,4y\,dy = \frac{4y^3}{3}\Big|_0^1 = \frac{4}{3}$$

$$E[XY] = \int_0^1 \underbrace{\int_0^1 xy\,8xy\,dx}_{8x^2y^2}\,dy$$

$$\frac{8x^3y^2}{3}\Big|_0^1 = \frac{8y^2}{3}$$

$$\frac{8y^3}{9}\Big|_0^1 = \frac{8}{9}$$

$$\mathrm{cov}(X,Y) = \frac{8}{9} - \frac{4}{3}\cdot\frac{4}{3} = \frac{8}{9} - \frac{16}{9} = \frac{8}{9}$$

$$\rho = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)\,\mathrm{var}(Y)}}$$

$$\mathrm{var}(X) = E[X^2] - (E[X])^2$$

$$E[X^2] = \int_0^1 x^2\,4x\,dx = x^4\Big|_0^1 = 1$$

$$\mathrm{var}(X) = 1 - \left(\frac{4}{3}\right)^2 = 1 - \frac{16}{9} = -\frac{7}{9} \quad \times$$

× Varyans negatif çıktığı için ortak davranış hatalı

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

TEKSAN

$$f(x,y) = \begin{cases} 2, & 0 \leq x \leq y \leq 1 \\ 0, & \text{diğer} \end{cases}$$

X ve Y rastlantı değişkenlerinin ortak yoğunlukları veriliyor kovaryanslarını hesaplayınız?

$$f(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_x^1 2\,dy = 2y \Big|_x^1 = 2 - 2x$$

$$f(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^y 2\,dx = 2x \Big|_0^y = 2y$$

$$\mathrm{cov}(X,Y) = E[XY] - E[X]E[Y]$$

$$E[X] = \int x f(x)\,dx = \int_0^1 x(2-2x)\,dx = x^2 - \frac{2}{3}x^3 \Big|_0^1 = \frac{1}{3}$$

$$E[Y] = \int y \cdot f(y)\,dy = \int_0^1 y(2y)\,dy = \frac{2y^3}{3} \Big|_0^1 = \frac{2}{3}$$

$$E[XY] = \int_0^1 \int_0^y xy\,(2)\,dx\,dy =$$

$$\int_0^1 \int_x^1 xy(2)\,dy\,dx$$

ortak davranış olduğunda her iki integral 0 dan 1'e olamaz. Biri ortak diğeri sayı olmak zorunda

$$= x^2 y \Big|_0^y = y^3$$

$$\int_0^1 y^3\,dy = \frac{y^4}{4} \Big|_0^1 = \frac{1}{4}$$

$$\mathrm{cov}(X,Y) = \frac{1}{4} - \frac{1}{3} \cdot \frac{2}{3} = \frac{1}{4} - \frac{2}{9} = \frac{1}{36}$$

ÖDEV

$$f(X,Y)=\begin{cases} 8xy, & 0\leq x\leq 1,\ 0\leq y\leq x \\ 0, & \text{diğer} \end{cases}$$

X ve Y nin kovaryansını ve korelasyonunu hesaplayınız.

Tanım: Kovaryans ve Bağımsızlık

X ve Y rastlantı değişkenleri bağımsız ise kovaryansları sıfırdır.

$$cov(X,Y)=0$$

$$E[XY]-E[X]E[Y]=0$$

$$E[XY]=E[X]E[Y]$$

$$\rho=\frac{cov(X,Y)}{\sqrt{var(X)var(Y)}}=0$$

cov sıfırsa ilişkisiz

NOT: Her rastlantı değişkeni bağımsız ise kesinlikle ilişkisizdir ($\rho=0$). Ancak tersi her zaman doğru değildir.

Tanım: Toplamın Varyansı

X ve Y birer rastlantı değişkeni, $U=aX+bY+c$ ise;

$$Var(U)=a^2var(X)+b^2var(Y)+2ab\,cov(X,Y)$$

X ve Y bağımsız ise;

$$var(U)=a^2var(X)+b^2var(Y)$$

olur.

$X \backslash Y$	0	1	2
0	0,1	0	0
1	0,1	0,2	0
2	0,1	0,35	0,15

X ve Y nin ortak PMF si tabloda veriliyor

a- $cov(X,Y)$

b- $\rho = ?$

c- $3X-4Y+8$ fonksiyonun varyansını bulunuz.

a- $P(X=x) = \sum_y P(x,y)$

$P(Y=y) = \sum_x P(x,y)$

X	0	1	2
P(x)	0,1	0,3	0,6

Y	0	1	2
P(y)	0,3	0,55	0,15

$E[XY] = \sum_x \sum_y xy\,P(x,y)$

$= 1.1.(0,2) + 2.1(0,35) + 2.2(0,15)$

$= 0,2 + 0,7 + 0,6$

$= 1,5$

$E[X] = \sum_x x\,P(x) = 0,3 + 1,2 = 1,5$

$E[Y] = \sum_y y\,P(y) = 0,55 + 0,3 = 0,85$

$cov(X,Y) = 1,5 - (1,5)0,85 = 9/40$

b- $Var(x) = E[X^2] - (E[X])^2$

$E[X^2] = \sum_x x^2 P(x) = 1^2 0,3 + 2^2 0,6 = 2,7$

$E[Y^2] = \sum_y y^2 P(y) = 1^2 0,55 + 2^2 0,15 = 1,15$

$var(X) = 2,7 - (1,5)^2 = 9/20$

$var(Y) = 1,15 - (0,85)^2 = 0,43$

$\rho = \dfrac{9/40}{\sqrt{0,45 \times 0,43}} \cong 0,51$

c- $var(3X - 4Y + 8)$ aralarında ilişki var

$= 9\,var(X) + 16 \cdot var(Y) + 2 \cdot (3) \cdot (-4) \cdot cov(X,Y)$

$= 9\,\dfrac{9}{20} + 16 \cdot 0,43 - 24 \cdot (9/40)$

Tanım: Koşullu Ortalama Değer

X ve Y her hangi iki rastlantı değişken olsun Y=y koşulu altında X'in ortalama değeri şu şekilde hesaplanır.

$E[X/Y=y] = \int_{-\infty}^{\infty} x f(x/y)\,dx$ → sürekli

$E[X/Y=y] = \sum_{x} x\,P(x/y)$ → Ayrık

$E[y/X=x] = \int_{-\infty}^{\infty} y f(y/x)\,dy$

$E[y/X=x] = \sum_{y} y\,P(y/x)$

⟹ X rastlantı değişkeni, bir zar atıldığında gözüken rakamlar olsun. Y rastlantı değişkeni ise, $X \leq 3$ ise sıfır aksi durumda bir değerini alan bir değişken olsun. Buna göre

a- $E[X/Y=0] = ?$

b- $E[X/Y]$ rastlantı değişkenini tanımla

$X \leq 3$

X	1	2	3	4	5	6
Y	0	0	0	1	1	1

a- $E[X/Y=0] = \sum_x x P(X/Y=0)$

$$P(X=1,2,3/Y=0) = \frac{P(X=1,2,3,Y=0)}{P(Y=0)} = \frac{1/6}{3/6} = \frac{1}{3}$$

$P(X=1/Y=0) = 1/3$
$P(X=2/Y=0) = 1/3$
$P(X=3/Y=0) = 1/3$

X \ Y	0	1
1		
2		
3		
4		
5		
6		

$P(X=4,5,6/Y=0)$
$P(X=4/Y=0) = 0$
$P(X=5/Y=0) = 0$
$P(X=6/Y=0) = 0$

$= 1P(X=1/Y=0) + 2P(X=2/Y=0) + 3P(X=3/Y=0) + \dots + P(X=6/Y=0)$

$= 1\frac{1}{3} + 2\frac{1}{3} + 3 \cdot \frac{1}{3}$

$= 2$

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

TEKSAN

b- Z rastlantı değişkeni $\rightarrow E[X/Y] \rightarrow E[X/Y=0]$, $E[X/Y=1]$

$$E[X/Y=1] = \sum_{x} x\,P(X/Y=1)$$

$$P(X=4,5,6/Y=1) = \frac{P(X=4,5,6,\ Y=1)}{P(Y=1)}$$

$$P(X=4/Y=1) = \frac{1/6}{3/6} = 1/3$$

$$P(X=5/Y=1) = 1/3$$

$$P(X=6/Y=1) = 1/3$$

$$E[X/Y=1] = 4\cdot\frac{1}{3} + 5\cdot\frac{1}{3} + 6\cdot\frac{1}{3} = 5$$

$E[X/Y]$	2	5
$P(X/Y)$	1/2	1/2

NOT: $E[X] = E[E[X/Y]]$

ÖDEV Notta verilen ifadenin doğruluğunu ispatlayınız

RASTLANTI DEĞİŞKENLERİNİN FONKSİYONLARI

Karşılaştığımız bazı durumlarda olasılık dağılımı bilinen rastlantı değişkeninin yerine, bu rastlantı değişkeninin fonksiyonunun dağılımı önemlidir.

Bu bölümde X'in olasılık dağılımı bilinirken $Y=g(x)$ ile verilen Y rastlantı değişkeninin olasılık dağılımını bulmak amaçlanmaktadır. Bunun için iki yöntem açıklanacaktır.

$$Y = g(x) \to \text{PMF}, \ \text{PDF}$$

1- Toplam Dağılım Fonksiyonu (CDF) Metodu

CDF metodu her yerde kullanılan bir metoddur. Amaç $Y=g(x)$'in toplam dağılım fonksiyonu X'in CDF'si cinsinden bulmaktır. Burada aralıkların nasıl değiştiği mutlaka incelenmelidir. İşleme $P(Y<y)$ hesaplanarak ve $F(x)$ cinsinden ifade edilerek başlanır.

$\Rightarrow$ X rastlantı değişkeni $1/\lambda$ ortalama değerli üstel rastlantı değişkeni olsun. $Y=b.X$, $b>0$ ise Y'nin CDF'sini ve PDF'sini bulunuz.

$$F(x) = 1 - e^{-\lambda x \to y/b} \quad [0,\infty] \qquad \text{CDF}$$

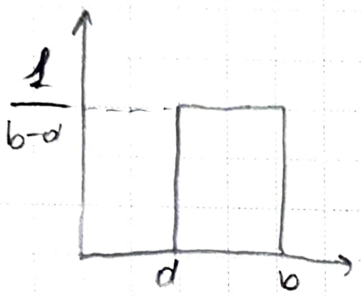
$$P(Y<y) = P(bx<y) = P\left(x < \frac{y}{b}\right) \Rightarrow F_x\left(\frac{y}{b}\right)$$

$$F_x\left(\frac{y}{b}\right) = 1 - e^{-\lambda \frac{y}{b}} \quad (0,\infty) \qquad \text{Y'nin CDF'si}$$

$$\boxed{F(y) = 1 - e^{-\lambda \frac{y}{b}}} \ \text{CDF} \qquad f(y) = \frac{d}{dy}F(y)$$

$$\boxed{f(y) = \frac{b}{\lambda} e^{-\lambda y/b}} \ \text{PDF} \qquad \frac{b}{\lambda} \text{ ortalamalı üstel rastlantı değişkeni}$$

⇒ X rastlantı değişkeni $[a,b]$ aralığında düzgün dağılımlı bir rastlantı değişkeni olsun. $Y = cX + d$, $c > 0$ ise Y'nin CDF'sini bulunuz?



$$f(x) = \frac{1}{b-a} \quad , [a,b]$$

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{a}^{x} \frac{1}{b-a}\,dt = \frac{x-a}{b-a}$$

x yerine $\frac{y-d}{c}$

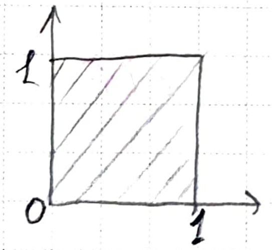
$$P(Y<y) = P(cX+d<y) = P\left(x < \frac{y-d}{c}\right) = F_x\left(\frac{y-d}{c}\right)$$

$$F(y) = \frac{\frac{y-d}{c} - a}{b-a} = \frac{y-d-ac}{bc-ac} \quad , [ac+d, bc+d]$$

düzgün dağılımlı

- $Y = cX + d$ eşitliğinde alt sınır (a) ve üst sınır (b) yi koyarak aralığı bulabiliriz. Yoğunluk fonksiyonu sorulsaydı Y'ye göre türev alınacaktı.

⇒ X rastlantı değişkeni $[0,1]$ aralığında düzgün dağılımlı rastlantı değişkeni olsun. $Y = -\ln(1-X)$ ise Y'nin CDF'si ve PDF'sini bulunuz.



$$\left.\begin{array}{l} -\ln(1-0) \to 0 \\ -\ln(1-1) \to \infty \end{array}\right\rangle [0,\infty)$$

$$f(x) = 1 \quad , [0,1]$$

$$F(x) = \int_0^x 1\,dt = X$$

$$P(Y<y) = P(-\ln(1-x)<y) = P(x < -e^{-y}+1) = F_x(1-e^{-y})$$

$$F(y) = F_x(1-e^{-y}) = 1-e^{y} \quad [0,\infty) \quad \text{CDF}$$

$$f(y) = \frac{dF(y)}{dy} = e^{-y} \quad \text{PDF}$$

- Öncekilerde doğrusal fonksiyon değildi o yüzden dağılım türü değişmedi

NOT: $Y=g(X)$ ile ifade edilen dönüşüm doğrusal ise rastlantı değişkeninin türü değişmez. Ancak aralığı değişe bilir.

Dönüşüm doğrusal değilse rastlantı değişkenine ait dağılımın hem türü hem de aralığı değişir.

⇒ X rastlantı değişkeninin PDF'si aşağıda veriliyor

$$f(x)=\begin{cases} 3x^2 & , 0\le x\le 1 \\ 0 & , \text{diğer} \end{cases}$$

$U=40(1-X)$ ise U'nun PDF'sini bulun?

Önce CDF'yi buluyoruz.

$$F(x)=\int_{-\infty}^{x} f(t)\,dt=\int_0^x 3t^2\,dt=x^3$$

$$P(U<u)=P(40(1-X)<u)=P\left(X>1-\frac{u}{40}\right)=1-P\left(X<1-\frac{u}{40}\right)$$

$$F(u)=1-\left(1-\frac{u}{40}\right)^3 \quad 0\le u\le 40 \quad \text{CDF}$$

$$f(u)=\frac{dF(u)}{du}=-3\left(1-\frac{u}{40}\right)^2\cdot\frac{-1}{40}=\frac{3}{40}\left(1-\frac{u}{40}\right)^2$$

$$0\le u\le 40 \quad \text{PDF}$$

2- Dönüşüm Metodu

Teorem: Ayrık Dönüşümler

X rastlantı değişkeni $P(x)$ olasılık kütle fonksiyonu ayrık bir rastlantı değişkeni olsun. X ve U değerleri arasında $U=h(X)$ şeklinde $X=w(u)$ çözümüne sahip bire bir dönüşüm tanımlansın. U'nun PMF'si $g(u)=P[w[u]]$ 'dur

$\Rightarrow$ X aşağıda verilen kütle fonksiyonuna sahip geometrik rastlantı değişkeni olsun

$$P(x) = \frac{3}{4}\left(\frac{1}{4}\right)^{x-1} \quad x = 1, 2, 3, \ldots$$

$U = x^2$'nin PMF sini bulun?

$$U = x^2 \qquad x = \sqrt{U}$$

$$P(u) = \frac{3}{4}\left(\frac{1}{4}\right)^{\sqrt{u}-1} \qquad u = 1, 4, 9, \ldots$$

Teorem: Sürekli Dönüşümler

X, $f(x)$ yoğunluk fonksiyonuna sahip sürekli bir rastlantı değişkeni olsun. $y = h(x)$, $x = w(y)$ şeklinde bir çözüme sahip birebir dönüşüm olsun. Y'nin yoğunluk fonksiyonu şu şekilde bulunur.

$$f_Y(y) = f(x)\left|\frac{dx}{dy}\right| = f[w(y)] \cdot x\, |J|$$

$\longrightarrow$ Jacobien

$\Rightarrow$ X sürekli rastlantı değişkeninin yoğunluk fonksiyonu aşağıda veriliyor.

$$f(x) = \begin{cases} \frac{x+1}{2} & , -1 \leq x \leq 1 \\ 0 & , \text{diğer} \end{cases}$$

$Y = X^2$ ise Y'nin yoğunluk fonksiyonunu bulun?

$$Y = X^2 \qquad x_1 = \sqrt{Y} \qquad x_2 = -\sqrt{Y}$$

$$J_1 = \frac{1}{2\sqrt{Y}} \qquad J_2 = -\frac{1}{2\sqrt{Y}}$$

$$f_Y(y) = f_X(\sqrt{y})\left|\frac{1}{2\sqrt{y}}\right| + f_X(-\sqrt{y})\left|\frac{-1}{2\sqrt{y}}\right|$$

$$= \frac{\sqrt{y}+1}{2} \cdot \frac{1}{2\sqrt{y}} + \frac{-\sqrt{y}+1}{2} \cdot \frac{1}{2\sqrt{y}}$$

$$= \frac{1}{2\sqrt{y}}(1) = \boxed{\frac{1}{2\sqrt{y}}} \qquad \boxed{0 \leq y \leq 1}$$

NOT: Öteleme ve Ölçekleme Parametresi

X herhangi bir dağılıma sahip bir rastlantı değişkeni $Y = a + bX$ olsun. $X = \frac{y-a}{b}$ dir

a → öteleme parametresi

b → ölçekleme parametresi

Bu durumda Y'nin yoğunluk fonksiyonu

$$F_y(y) = \frac{1}{|b|} f\left(\frac{y-a}{b}\right)$$ ile hesaplanır.

Merkezi Limit Teoremi (CLT)

Tanım: Örnek Ortalama

Herhangi bir dağılıma sahip bağımsız rastlantı değişkenlerinden oluşan gruba örnek denir ve şu şekilde gösterilir.

$$\underline{X} = \{X_1, X_2, \ldots, X_n\}$$

Örnek ortalaması yani tüm gözlemlerin ortalama değeri İstatistikte önemli bir rol oynar. Örnek ortalaması $\bar{X}$ ile gösterilir.

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$ şeklinde hesaplanır.

Eğer $E[X_i] = \mu$ ve $var(X_i) = \sigma^2$ ise

$$E[\bar{X}] = E\left[\frac{X_1 + X_2 + \ldots + X_n}{n}\right] = \frac{1}{n}(n.\mu) = \mu$$

$$var(\bar{X}) = var\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \frac{\sigma^2}{n}$$

CLT'ye göre normal dağılımlı rastlantı değişkenlerinin toplamı yine normal dağılımlıdır. Diğer türden rastlantı değişkenlerinin toplamı ise gözlem sayısı n arttıkça normal dağılıma yakınsar.

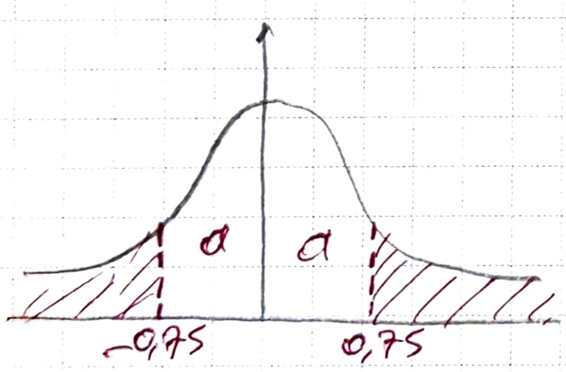
$\bar{X}$, μ ortalamalı ve σ^2 varyanslı herhangi bir dağılımdan gelen bir örneğin ortalaması olsun. n'nin büyük değerleri için örnek μ ortalamalı ve $\frac{\sigma^2}{n}$ varyanslı normal dağılımla modellenebilir.

⇒ Bir firma ortalama ömrü 800 saat standart sapması 400 saat olan ampuller üretmektedir. 16 ampulün rastgele seçildiğini varsayarsak ortalama ömürlerinin 725 saatten az olma olasılığı nedir.

$\mu = 800$ saat $\quad \sigma = 400 \quad P(\bar{X} < 725) = ? \quad n = 16$

$Var(\bar{X}) = \frac{\sigma^2}{n}$, $std(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{400}{4} = 100$

$z = \frac{x - \mu}{\sigma} \Rightarrow \frac{725 - 800}{100} = -0{,}75$



$P(z < -0{,}75) = 0{,}5 - P(z = 0{,}75) = 0{,}2266$

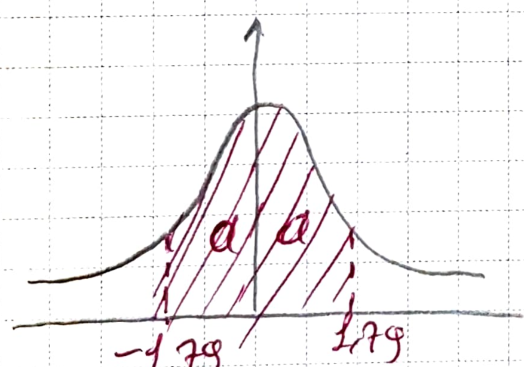
NOT: Merkezi limit teoremi (CLT) ilk olarak binom dağılımında keşfedilmiştir. Binom dağılımlı Y rastlantı değişkeni n adet bernoulli denemesinin toplamıdır. CLT'ye göre binom dağılımı p ortalamalı, $\frac{p(1-p)}{n}$ varyanslı normal dağılıma yakınsar.

⇒ Bir para 500 kez atılıyor. Buna göre örnek ortalamasının 0,46 ile 0,54 arasında olma olasılığı nedir?

$n=500 \quad p=1/2 \quad P(0,46<\bar{x}<0,54)$

$\mu=p=0,5 \quad \sigma^2=\frac{(0,5)^2}{500}=0,0224$

$z_1=\frac{0,46-0,5}{0,0224} \qquad z_2=\frac{0,54-0,5}{0,0224}$



$P(-1,79<Z<1,79)=2P(Z=1,79)=0,027$

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

TEKSAN

İstatiksel Çıkarım

Bu bölümde popülasyon parametrelerinin tahmini problemi tartışılıcaktır. Bir popülasyon parametresi olan θ'nın tahmini tek bir $\hat{\theta}$ değeri verir ve nokta kestirimidir.

İstatiksel çıkarımın temel amaçları güven aralığı ve hipotez testidir. Güven aralığı bir nokta kestiriminin hata aralığını tanımlar. Hipotez testi ise parametreler hakkında bazı iddiaları ispatlamaya çalışır.

Güven Aralığı

Bu aralık parametrelerin doğru değerlerinin bulunduğu aralıktır. $\bar{X}$, varyansı bilinen normal bir popülasyondan alınan n adet rastgele örneğin ortalaması olsun. μ için %(1-α).100 güven aralığı

$$\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

popülasyon

örnekleme

$z_{\alpha/2}$, $\alpha/2$ kadarlık alanı sağ tarafta bırakan z değeridir.

⟹ Bir kurutma fabrikasında olasılan ortalama kurutma süreleri aşağıda veriliyor.

(3,4) (3,7) (2,5) (2,8) (4,8) (4,4) (2,9) (4,0)
(3,6) (5,2) (2,8) (3,0) (3,3) (4,8) (5,6)

$\sigma = 1$ alarak ortalama kurutma süresi için %95 güven aralığını hesaplayınız.

$n = 15$, $\sigma = 1$

%(1-α).100 = %95 $\alpha = 0{,}05$ $\alpha/2 = 0{,}025$

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

tablodan $z_{0,025} \rightarrow 1,96$

$$\bar{X} = \sum_{i=1}^{15} X_i/15 = 3,79$$

$$\left(3,79 - 1,96\frac{1}{\sqrt{15}} < \mu < 3,79 + 1,96\frac{1}{\sqrt{15}}\right)$$

$$(3,79 \mp 0,51) = (3,28 < \mu < 4,3)$$

⇒ Bir nehrin 36 farklı konumundan alınan çinko konsantrasyonunun ortalama değeri litre başına 2,6 miligramdır. Ortalama çinko konsantrasyonu için standart sapmayı (σ) 0,3 olarak %95 ve %99 güven aralığını hesaplayınız?

$\bar{X} = 2,6$ mg/lt

$n = 36$

$\sigma = 0,3$

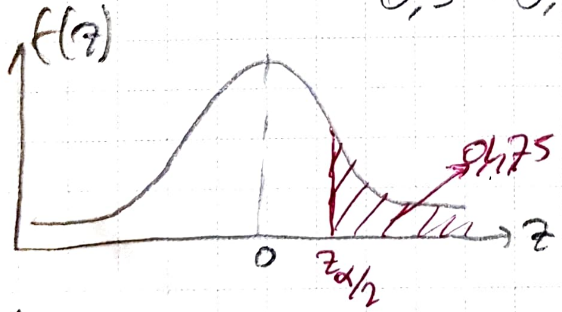
a) %95

$(1-\alpha) = 0,95$

$\alpha = 0,05$

$z_{0,025} = 1,96$

$0,5 - 0,025 = 0,475$

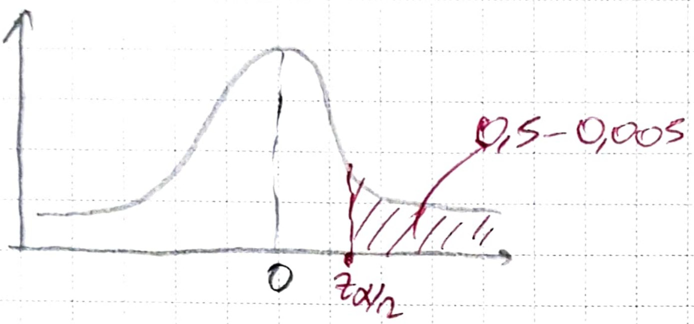


$$\left(2,6 - 1,96\frac{0,3}{\sqrt{36}} < \mu < 2,6 + 1,96\frac{0,3}{\sqrt{36}}\right)$$

$(2,50 < \mu < 2,70)$

b) %99

$$(1-\alpha)=0,99$$
$$\alpha=0,01$$
$$\alpha/2=0,005$$



$$z_{\alpha/2}=2,575$$

$$\left(2,6-2,575\cdot\frac{0,3}{\sqrt{36}}<\mu<2,6+2,575\frac{0,3}{\sqrt{36}}\right)$$

$$(2,47<\mu<2,73)$$

NOT: Güven derecesi arttıkça, güven aralığı genişler.

Örnek Büyüklüğünün Hesaplanması

Verilen bir doğrulukla bilinmeyen bir parametreyi belirlemek için toplanması gereken gözlem sayısı örnek büyüklüğüdür. m örnek büyüklüğü olmak üzere %(1-α)100 güven aralığı için şu şekilde hesaplanır;

$$m=\mp z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$n=\left(\frac{z_{\alpha/2}\cdot\sigma}{m}\right)^2$$

$$\bar{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}}<\mu<\bar{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Bir maddenin pH değerini %99 güven aralığında 0,1 ara ile tahmin etmek istiyoruz. Önceki araştırmalara göre pH değerinin 5 ile 7 arasında değiştiği biliniyor. Buna göre örnek büyüklüğü ne olmalıdır?

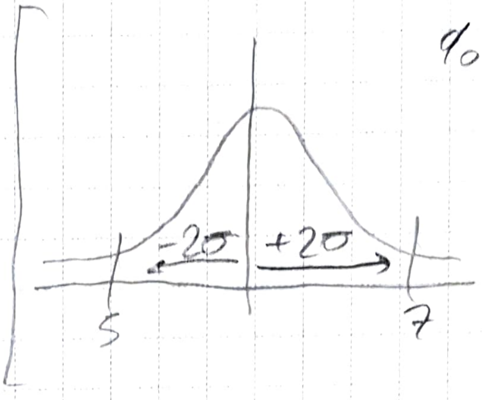
$m = 0,1$

$(1-\alpha) = 0,99$

$\alpha = 0,01$

$\alpha/2 = 0,005$

$z_{\alpha/2} = 2,575$



%68, %99 kuralı

$\sigma = \frac{7-5}{4}$

$\sigma = 0,5$

$$n = \left(\frac{(2,575)(0,5)}{0,1} \right)^2 \cong 166$$

Hipotez Testi

Hipotez karşılaşılan özel duruma ilişkin bir önermedir. İstatiksel hipotez testleri bir olasılık dağılımının bir parametre değerine ait hipotezdir. Bir istatiksel hipotez doğru ya da yanlış olabilir. Çünkü bu bir önermedir. Gerçeği öğrenebilmek için popülasyon parametresinin gerçek değeri, θ hesaplanmalıdır. $\hat{\theta}$ örneklem istatiğinin değeri θ_0 ileri sürülen θ değeri olmak üzere

$\hat{\theta} - \theta_0 > 0$

$\hat{\theta} - \theta_0 = 0$

$\hat{\theta} - \theta_0 < 0$

gibi durumlar ortaya çıkabilir.

Hipotez testleri parametrik testler ve parametrik olmayan testler olmak üzere ikiye ayrılır. Parametrik testler eşit aralıklı, yada oranlı ölçeklerin kullanıldığı testlerdir. Parametrik olmayan testler örneklem dağılımı ne durursa olsun uygulanılan testlerdir.

Hipotez Testinin Adımları

1- Hipotezin ifade edilmesi. H_0 = sıfır hipotezi
H_1 = karşı hipotez

H_0 hipotezi $\hat{\theta}$ değeri ile θ_0 değeri arasında anlamlı bir farklılık yoktur. Tezini iddia eder. H_1 hipotezi ise bu iki değer arasında anlamlı bir farklılık olduğunu iddia eder.

$H_0: \theta = \theta_0$

anlamlı farklılık var $\left\{ \begin{array}{l} H_1: \theta \neq \theta_0 \\ H_1: \theta > \theta_0 \\ H_1: \theta < \theta_0 \end{array} \right.$

a- İki yönlü testlerde hipotezler
$H_0: \theta = \theta_0$
$H_1: \theta \neq \theta_0$

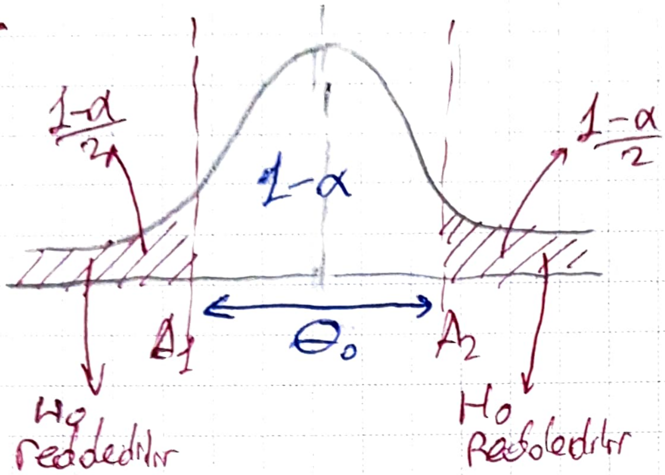
b- Tek yönlü üst kuyruk testlerinde hipotezler
$H_0: \theta = \theta_0$
$H_1: \theta > \theta_0$

c- Tek yönlü alt kuyruk testlerinde hipotezler
$H_0: \theta = \theta_0$
$H_1: \theta < \theta_0$

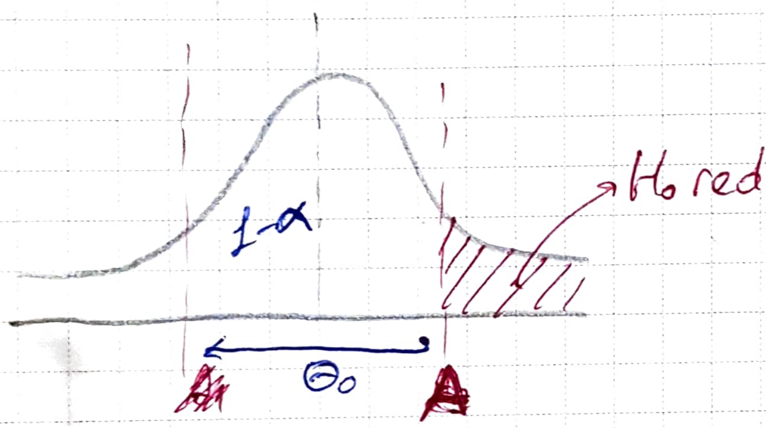
KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

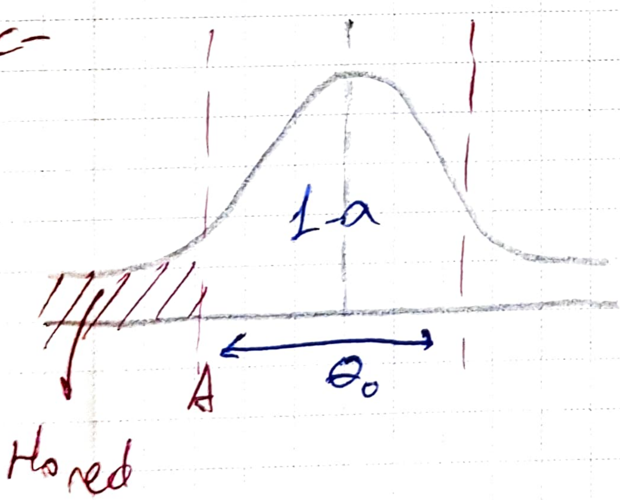
a-



b-



c-



KONU / SUBJECT:

TARİH / DATE: SAAT / TIME: TEKSAN

2- Anlamlılık düzeyinin belirlenmesi.
Hipotez testlerinde sıfır hipotezinin yanlışlıkla kabul edilmesi ya da reddedilmesi sonucu işlenen hataya yorumlama hatası denir. Gerçekte doğru olan sıfır hipotezinin reddedilmesi durumunda işlenen hataya birinci tip hata ya da α hatası denir. α tipi hata yapmanın max olasılığına testin anlam düzeyi denir. Test için α değeri önceden belirlenmelidir. Hipotez yanlış olduğu halde kabul edilirse ikinci tip hata oluşur. Buna β tipi hata denir. Çoğunlukla α tipi hata kontrol edilir.

3- Örneklerin seçilmesi, Test istatistiğinin belirlenmesi
4- İstatistiksel kararın verilmesi
5- Probleme ilişkin kararın verilmesi

⇒ Margarin üreten bir fabrikada 250 grlık paketler halinde üretim yapılması ön görülmektedir. Margarin paketlerinin kontrol amacıyla rastgele 100 paket seçilmiş ve ortalama ağırlığın 244 gr olduğu standart sapmanın 18 gr olduğu bulunmuştur. $\alpha = 0{,}05$ anlam düzeyinde paketlerin ağırlığının belirlenen standartta olup olmadığı söylenebilir mi

1- Hipotez testinin ifade edilmesi

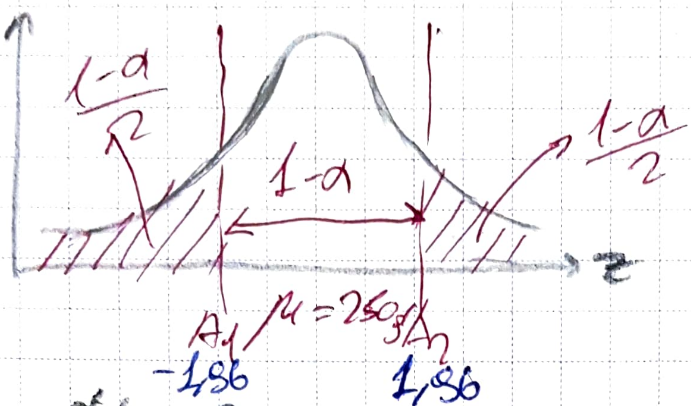
$\mu_0 = 250gr$

$H_0: \mu = 250 gr$

$H_1: \mu \neq 250 gr$ (iki yönlü)

2- Anlamlılık düzeyi

$\alpha = 0,05$



$\alpha/2 = 0,025$

3- Örneklerin seçilmesi; Test istatistiğinin Hesaplanması

$n = 100$ paket

$\bar{X} = 244gr \rightarrow \bar{X} = 244gr$

$S = 18gr \rightarrow S_{\bar{X}} = \frac{S}{\sqrt{n}} = \frac{18}{\sqrt{100}} = 1,8$ (CLT)

$\mu_0 = 250gr$

$\alpha = 0,05$

$$z = \frac{X-\mu}{\sigma} = \frac{\bar{X}-\mu_0}{S_{\bar{X}}}$$

$$z = \frac{244-250}{1,8} = -3,33$$

4- İstatistiksel Karar

$\alpha/2 = 0,025$

$z_{\alpha/2} = 1,96$

$$\left(\underbrace{\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{A_1} < \mu < \underbrace{\bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{A_2}\right)$$

$$A_1 = 244 - 1,96\frac{18}{\sqrt{100}} \rightarrow 246,47$$

$$A_2 = 244 + 1,96\frac{18}{\sqrt{100}} \rightarrow 253,53$$

Gerçek değeri 250 olduğu için 244 yerine 250 yazılabilir

5- Kararın verilmesi

$Z_{hesap} = -3,33$

$Z_{Tablo} = 1,96$

$|Z_{hes}| > Z_{tab}$ ise H_0 reddedilir.

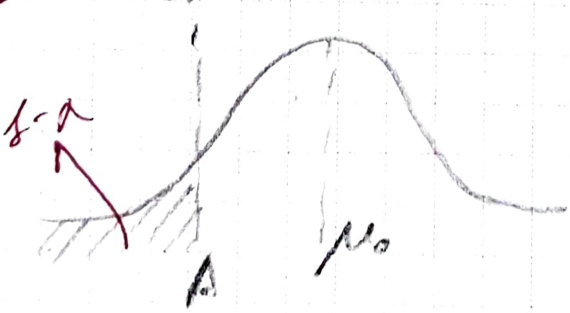
Üretilen margarin paketlerinin 250 gr olmadığı ve 250 gr dan anlamlı bir farklılık gösterdiği paketleme üretim sisteminin planlı şekilde üretim yapmadığı sonucuna ulaşılmıştır.

Bir firmanın geliştirdiği yeni sistemin ortalama paketleme süresini ürün başına 15 dakikanın altına indirdiği iddia edilmektedir. Bu iddiayı araştırmak amacıyla paketleme esnasında rastgele seçilen 225 ürünün yeni sistemde ortalama paketleme süresi 13 dk ve standart sapması 4,2 dk olarak belirlenmiştir. Yeni sistemle ilgili iddia haklı mı 0,01 anlam düzeyinde karar veriniz?

1- $H_0 : \mu_0 = 15$ dk

$H_1 : \mu_0 < 15$ dk

2- $\alpha = 0{,}01$ $1-\alpha = 0{,}99$



3- $\mu_0 = 15dk$

$\bar{X} = 13dk$

$n = 225$ ürün

$s = 4,2$ dk $\longrightarrow S_{\bar{x}} = \frac{4,2}{\sqrt{225}} = 0,28$

$\alpha = 0,01$

$$Z = \frac{x-\mu}{\sigma} = \frac{\bar{X}-\mu_0}{S_{\bar{x}}} = \frac{13-15}{0,28} = -7,142$$

4- $Z_{hes} = -7,142$

$Z_{tab} = -2,33$

$|Z_{hes}| > Z_{tab}$ ise H_0 red
H_1 kabul

5- Firmanın geliştirdiği yeni sistem paketleme süresini 15dk'nın altına düşürmüştür.

- $\alpha = 0,01$

$Z_\alpha = 0,5 - 0,01 = 0,49$

	0,00	0,03
0,0		
2,3		0,49

Gauss'a göre
Tersten bak

KONU / SUBJECT:

TARİH / DATE: SAAT / TIME:

TEKSAN

⇒ (Küçük Örneklem Testi) $n<30$

NOT: Bu durumda

$$t = \frac{\bar{X} - \mu_0}{S_{\bar{x}}}$$

$$S_{\bar{x}} = \frac{s}{\sqrt{n-1}}$$

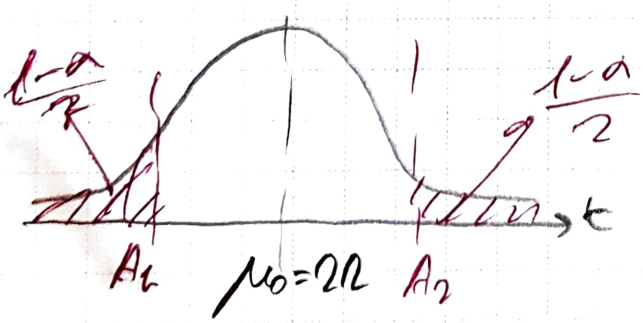
Kullanılır. Buna $n-1$ serbestlik dereceli t-testi denir.

Bir su dolum sisteminde doldurulan damacanalardaki su miktarının 22 lt olması planlanmıştır. Damacanalardaki su miktarının belirlenen standartlarda olup olmadığını kontrol etmek amacıyla rastgele 17 damacana seçilerek bunlardaki ortalama su miktarının 21,7 lt standart sapmasının 0,8 lt olduğu görülmüştür. Dolum sisteminin doğru çalışıp çalışmadığı hakkında 0,01 anlamlılık düzeyinde karar veriniz.

1- $H_0: \mu_0 = 22$ lt

$H_1: \mu_0 \neq 22$ lt

2- $\alpha = 0{,}01$



$$\alpha = \frac{0{,}01}{2} = 0{,}005$$

$$n-1 = 16$$

3- $\mu_0 = 22$

$\bar{X} = 21,7$

$n = 17$

$s = 0,8 \longrightarrow S_{\bar{x}} = \frac{0,8}{\sqrt{17-1}} = 0,2$

$\alpha = 0,01$

$$t = \frac{21,7 - 22}{0,2} = -1,5$$

4- $t_{hes} = -1,5$

$t_{tab}\left(\frac{1-\alpha}{2}, 16\right) = 2,921$

$|t_{hes}| > t_{tab}$

$1,5 < 2,921$ H_0 kabul
H_1 red

5- Bu su dolum sistemi doğru çalışmaktadır.