GGY 112
İSTATİSTİK

Doç. Dr. Furkan BAŞER
Ankara Üniversitesi Uygulamalı Bilimler Fakültesi



4. HAFTA

RASGELE DEĞİŞKENLER VE BEKLENEN DEĞER



RASTGELE DEĞİŞKENLER

Rastgele bir deney yapıldığında, genelde deneysel sonucun t üm ayrıntıları ile değil yalnızca sonucun belirlediği bir sayısal niceliğin değeri ile ilgileniriz.

Deney sonucuyla belirlenen ilgilenilen nicelikler bir *rastgele değişken* olarak bilinir.



RASTGELE DEĞİŞKENLERİN ÇEŞİTLERİ

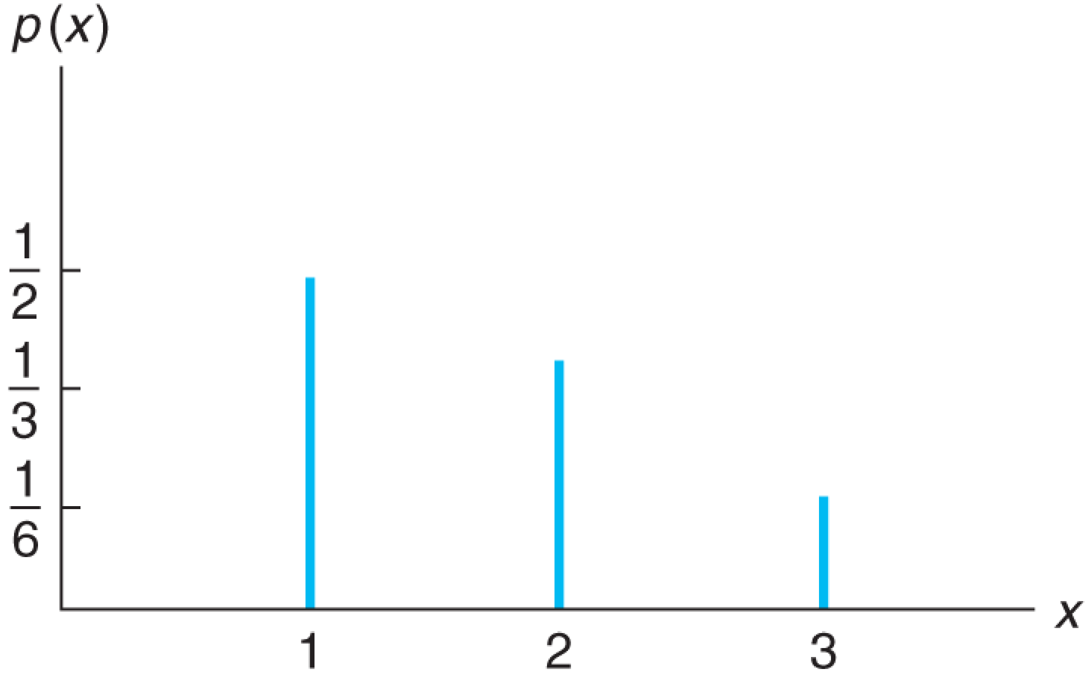
Daha önce de söz edildiği gibi, mümkün değerler kümesi bir dizi olan rastgele değişkene *kesiklidir*, denir. Kesikli bir *X* rastgele değişkeni için, *X*'in *olasılık kitle fonksiyonu* $p(a)$'yı ile tanımlarız.

$$p(a) = P\{X = a\}$$



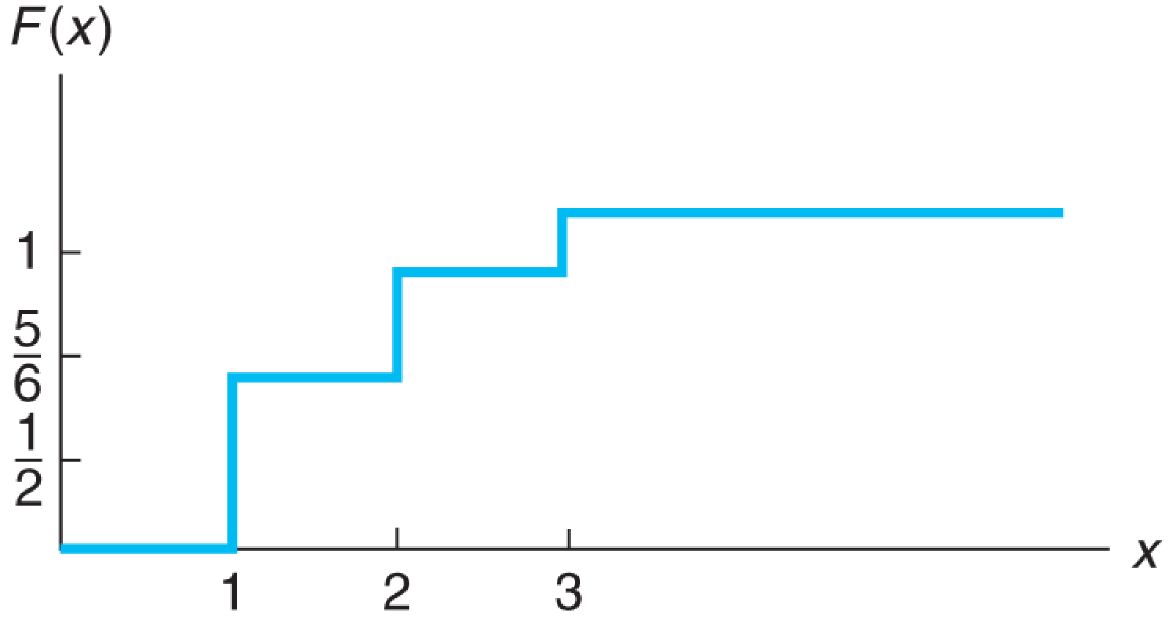
RASTGELE DEĞİŞKENLERİN ÇEŞİTLERİ

p(x)'in grafiği, Örnek 4.2a.

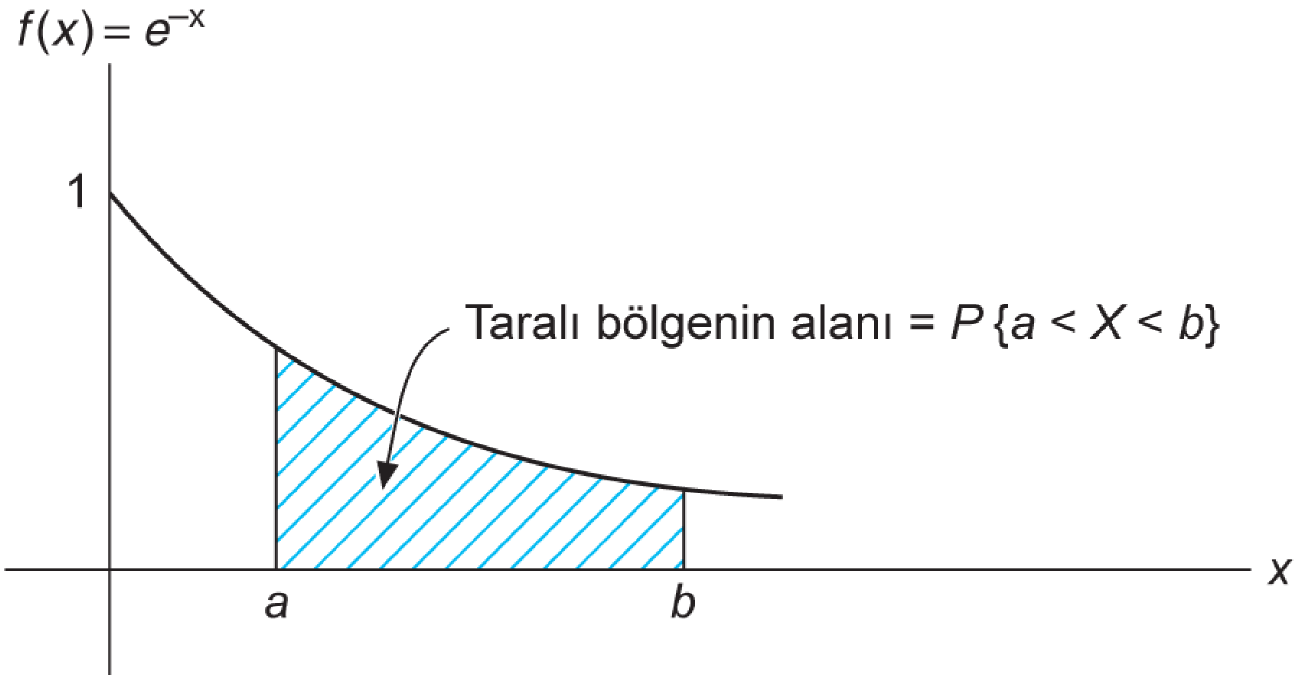


RASTGELE DEĞİŞKENLERİN ÇEŞİTLERİ

F(x)’in grafiği



RASTGELE DEĞİŞKENLERİN ÇEŞİTLERİ



$$f(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$ *olasılık yoğunluk fonksiyonu.*

ORTAK DAĞILIMLI RASTGELE DEĞİŞKENLER

Verilen bir deney için, yalnızca bireysel rastgele değişkenlerin olasılık dağılım fonksiyonları ile değil, ancak aynı zamanda iki veya daha çok sayıda rastgele değişken arasındaki ilişkilerle de ilgileniriz. Söz gelimi, kanserin olası nedenleri ile ilgili bir deneyde günlük ortalama içilen sigara sayısı ile bireyin kanser hastalığına yakalandığı yaş arasındaki bağıntı ile ilgilenebiliriz.

ORTAK DAĞILIMLI RASTGELE DEĞİŞKENLER

$i \backslash j$	0	1	2	3	Satır Toplamı $= P\{X = i\}$
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
3	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
Sütun Toplamı $= P\{Y = j\}$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	

$$P\{X = i, Y = j\}$$

ORTAK DAĞILIMLI RASTGELE DEĞİŞKENLER

i \ j	0	1	2	3	Satır Toplamı $= P\{B = i\}$
0	.15	.10	.0875	.0375	.3750
1	.10	.175	.1125	0	.3875
2	.0875	.1125	0	0	.2000
3	.0375	0	0	0	.0375
Sütun Toplamı $= P\{G = j\}$	.3750	.3875	.2000	.0375	

$$P\{B = i, G = j\}$$

BAĞIMSIZ RASTGELE DEĞİŞKENLER

Herhangi iki *A* ve *B* gerçek sayılar kümesi için

$$P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\}$$

olduğunda, *X* ve *Y* rastgele değişkenlerine *bağımsızdır*, denir.

KOŞULLU DAĞILIMLAR

İki rastgele değişken arasındaki ilişki, birinin verilen bir değerine karşılık diğerinin koşullu dağılımı ele alınarak açıklanabilir. Herhangi iki *E* ve *F* olayı için, *P(F)* > 0 olmak üzere, *F* verildiğinde *E*'nin koşullu olasılığının

$$P(E|F) = \frac{P(EF)}{P(F)}$$

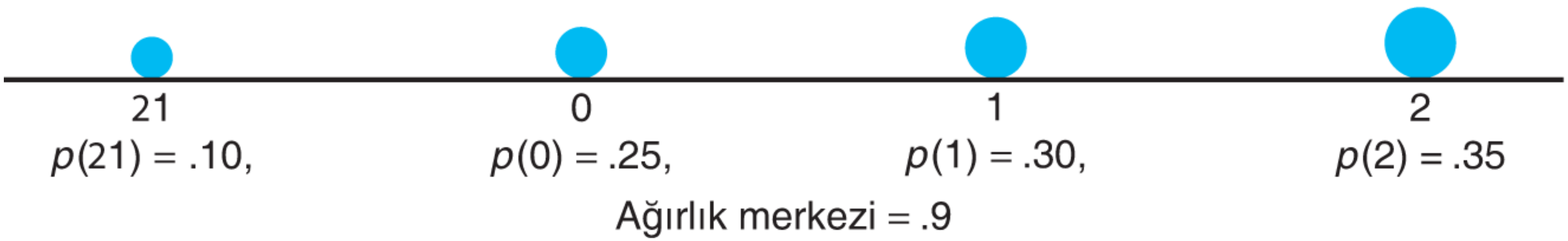
BEKLENTİ

Olasılık teorisinde çok önemli kavramlardan biri bir rastgele değişkenin beklentisidir. *X*;*x*1, *x*2, . . . mümkün değerlerini alan kesikli bir rastgele değişkense, *X’* in *E*[*X*] ile gösterilen *beklentisi* ya da *beklenen değeri* ile tanımlanır.

$$E[X] = \sum_i x_i P\{X = x_i\}$$



BEKLENTİ



BEKLENEN DEĞERİN ÖZELLİKLERİ

Bir *X* rastgele değişkeninin ve olasılık dağılımının (yani, kesikli durumda olasılık kitle fonksiyonunun, sürekli durumda olasılık yoğunluk fonksiyonunun) verildiğini varsayalım. Aynı zamanda *X'* in beklenen değerini değil ancak *X'* in bir fonksiyonunun, diyelim *g* (*X*)'in beklenen değerini hesaplamayla ilgilendiğimizi varsayalım.

Bu işi nasıl yürüteceğiz? Bir yolu şöyledir: *g (X)*'in kendisi rastgele değişken olduğundan, *X'* in dağılım bilgisinden hesaplanabilmesi gereken bir olasılık dağılımına sahip olmalıdır.

RASTGELE DEĞİŞKENLERİN TOPLAMLARININ BEKLENEN DEĞERİ

İki boyutlu versiyonu, *X* ve *Y* rastgele değişkenler ve *g* de bu iki rastgele değişkenin bir fonksiyonuysa, o zaman

$$\begin{aligned} E[g(X,Y)] &= \sum_y \sum_x g(x,y)p(x,y) \quad \text{kesikli durumda} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)\,dx\,dy \quad \text{sürekli durumda} \end{aligned}$$

olur.



VARYANS

Olasılık dağılım fonksiyonu ile birlikte bir *X* rastgele değişkeni verildiğinde, uygun tanımlanmış belirli ölçülerle kitle fonksiyonunun temel özellikleri özetlenebilseydi, son derece yararlı olurdu.

Böyle bir ölçü *X’* in beklenen değeri, *E*[*X*] olurdu.
Ancak *E*[*X*], *X’* in mümkün değerlerinin ağırlıklı ortalaması olmakla birlikte, bu değerlerin değişimi ya da yayılımı hakkında herhangi bir şey anlatmıyor.



VARYANS

$$W = 0,\ 1 \text{ olasılıkla}$$

$$Y = \begin{cases} -1, & \frac{1}{2} \text{ olasılıkla} \\ 1, & \frac{1}{2} \text{ olasılıkla} \end{cases}$$

$$Z = \begin{cases} -100, & \frac{1}{2} \text{ olasılıkla} \\ 100, & \frac{1}{2} \text{ olasılıkla} \end{cases}$$

RASTGELE DEĞİŞKENLERİN TOPLAMLARININ KOVARYANS VE VARYANSI

Rastgele değişkenlerin toplamının beklentisinin beklentilerinin toplamına eşit olduğunu gösterdik. Ancak varyanslar için ilişkili sonuç genelde geçerli değildir.

$$\begin{aligned}
\operatorname{Var}(X+X) &= \operatorname{Var}(2X) \\
&= 2^2 \operatorname{Var}(X) \\
&= 4 \operatorname{Var}(X) \\
&\neq \operatorname{Var}(X) + \operatorname{Var}(X)
\end{aligned}$$



MOMENT ÜRETEN FONKSİYONLAR

X rastgele değişkeninin moment üreten fonksiyonu *φ(t)* tüm *t* değerleri için

$$\phi(t) = E[e^{tX}] = \begin{cases} \displaystyle\sum_x e^{tx} p(x) & \text{X kesikli ise} \\ \displaystyle\int_{-\infty}^{\infty} e^{tx} f(x)\, dx & \text{X sürekli ise} \end{cases}$$

ile tanımlanır.



CHEBYSHEV EŞİTSİZLİĞİ VE ZAYIF BÜYÜK SAYILAR KANUNU

MARKOV EŞİTSİZLİĞİ

X yalnızca negatif olmayan değerleri alan bir rastgele değişkense herhangi bir *a* > 0 değeri için

$$P\{X \geq a\} \leq \frac{E[X]}{a}$$

sağlanır.

CHEBYSHEV EŞİTSİZLİĞİ VE ZAYIF BÜYÜK SAYILAR KANUNU

CHEBYSHEV EŞİTSİZLİĞİ

X, μ ortalamalı ve σ2 varyanslı bir rastgele değişkense, herhangi bir $k > 0$ değeri için

$$P\{|X-\mu| \geq k\} \leq \frac{\sigma^2}{k^2}$$