Rastgele Değişkenler & Beklenen Değer

Coşkunöz Vakfı - Veri Bilimci Eğitimi

Cem Yığman - 2022

Rastgele Değişkenler (Random Variables)
Tanım

- Gerçek hayat problemlerinde bir deneyin çıktılarından çok, çoğunlukla çıktı ile ilgili bir fonksiyon ile ilgileniriz.
- Bu durum, örneğin basit bir regresyon modelinde, bağımsız değişkenlerin parametrelerinden çok, bağımlı değişkenin değeri ile ilgilenmeye benzer.
- Örnekler:
    - İki hilesiz zar atıldığında, zarların toplamı kaç olacak ?
    - Hilesiz bir parayı 10 kere havaya attığımızda kaç kere yazı gelecek

Rastgele Değişkenler (Random Variables)

Tanım

X = İki hilesiz zar atıldığında toplamı gösteren rastgele değişken

$P\{X=2\} = P\{(1,1)\} = 1/36$

$P\{X=3\} = P\{(1,2), (2,1)\} = 2/36$

$P\{X=4\} = P\{(1,3), (2,2), (3,1)\} = 3/36$

$P\{X=5\} = P\{(1,4), (2,3), (3,2), (4,1)\} = 4/36$

$P\{X=6\} = P\{(1,5), (2,4), (3,3), (4,2), (5,1)\} = 5/36$

… X, 1 ile 12 arasında değeler alabiliyor.

Rastgele Değişkenler (Random Variables)

Örnek

$$1 = P\{\bigcup_{n=2}^{12}\{X = n\}\} = \sum_{n=2}^{12} P\{X = n\}$$

$$P\{X > 10\} = \sum_{n=11}^{12} P\{X = n\} = P\{X = 11\} + P\{X = 12\} = \frac{2}{36} + \frac{1}{36} = \frac{1}{12}$$

Rastgele Değişkenler (Random Variables)

Örnek

Hilesiz ve tura gelme olasılığı ***p*** olan bir parayı, tura gelene kadar attığımızı düşünelim. Deneyin sonucunda oluşacak sayıya N dersek (ilk turanın geleceği atış sayısı), N 1, 2, 3, … gibi değerler alan bir rastgele değişkendir.

$$P\{N=1\} = P\{H\} = p$$

$$P\{N=2\} = P\{(T,H)\} = (1-p)*p$$

$$P\{N=3\} = P\{(T,T,H)\} = (1-p)^2*p$$

$$P\{N=4\} = P\{(T,T,T,H)\} = (1-p)^3*p$$

$$P\{N=n\} = P\{(\underbrace{T,T,T,\ldots,T}_{\text{n-1}},H)\} = (1-p)^{n-1}*p, \qquad n \geq 1$$

Rastgele Değişkenler (Random Variables)

Ayrık Rastgele Değişkenler (Discrete Random Variables)

- Deneylerinin sayılabilir adette çıktısı olan rastgele değişkenlere “Ayrık Değişkenler” adı verilir.

- Ayrık rastgele değişkenler “Probability Mass Function” ile tanımlanır. p(a)

$$p(x_i) > 0 \qquad i = 1,2,3,...$$
$$p(x_i) = 0 \qquad diger$$

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

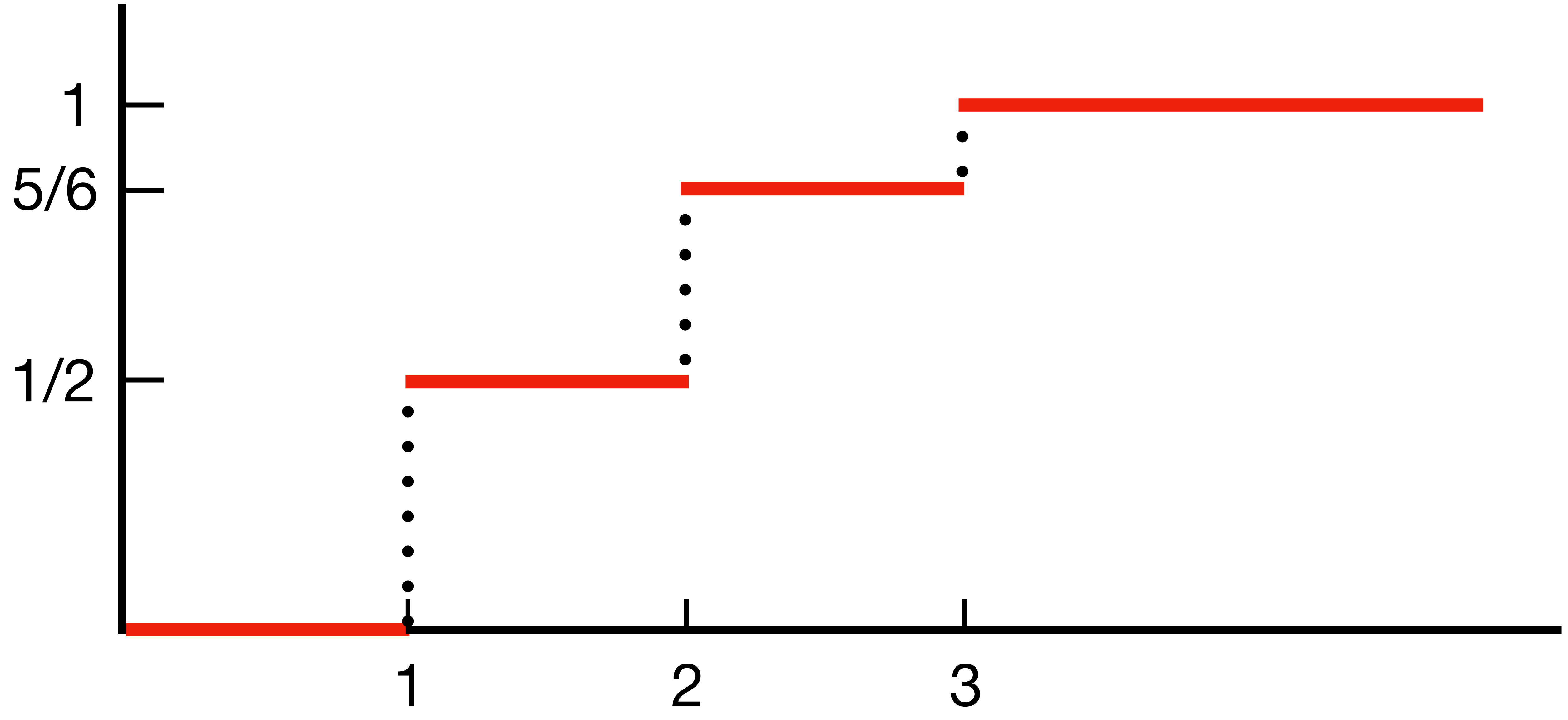
Rastgele Değişkenler (Random Variables)

Ayrık Rastgele Değişkenler (Discrete Random Variables)

Örnek

$$p(1) = \frac{1}{2} \qquad p(2) = \frac{1}{3} \qquad p(3) = \frac{1}{6}$$

$$F(a) = \begin{cases} 0, & a < 1 \\ \frac{1}{2}, & 1 \leq a < 2 \\ \frac{5}{6}, & 2 \leq a < 3 \\ 1, & 3 \leq a \end{cases}$$



Rastgele Değişkenler (Random Variables)

Ayrık Rastgele Değişkenler / 1 - Bernoulli Değişkeni

Deneyin başarı ya da başarısızlıkla sonuçlandığı değişkenlerdir.

$$p(0) = P\{X = 0\} = 1 - p$$

$$p(1) = P\{X = 1\} = p$$

p(1) başarılı deney sonuçlarını, p(0) ise başarısız deney sonuçlarını gösterir.

$$0 \leq p \leq 1$$

Rastgele Değişkenler (Random Variables)

Ayrık Rastgele Değişkenler / 2 - Binom Değişkeni

n adet deneme yapılan bir deneyde başarılı olma olasılığı *p* olan deneyin kaç kere başarılı olacağını gösterir.

$$p(i) = \binom{n}{i} p^i (1-p)^{n-i}$$

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

Örnek: 4 hilesiz para havaya atılıyor. Olayların bağımsız olduğu kabul edilirse, 2 yazı ve 2 tura gelme olasılığı nedir?

$$p(X=2) = \binom{4}{2} * \frac{1}{2}^2 * (1 - \frac{1}{2})^2 \quad = \frac{4!}{2! * 2!} * \frac{1}{4} * \frac{1}{4}$$

$$p(X=2) = \frac{3}{8}$$

Rastgele Değişkenler (Random Variables)

Ayrık Rastgele Değişkenler / 2 - Binom Değişkeni

Örnek: Bir uçağın motorunun uçuş sırasında arızalanması olasılığı (1-p) ve uçaklar sahip oldukları motorun %50'sinin çalışması durumunda bile uçabildiğine göre, hangi durumda 4 motorlu bir uçak iki motorlu bir uçağa tercih edilir?

Rastgele Değişkenler (Random Variables)

Ayrık Rastgele Değişkenler / 2 - Binom Değişkeni

$$p(i) = \binom{n}{i} p^i (1-p)^{n-i}$$

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

Çalışır durumdaki motorların sayısı = Binom

4 motorlu bir uçağın başarılı bir uçuş yapması;

$$\binom{4}{2} p^2(1-p)^2 + \binom{4}{3} p^3(1-p) + \binom{4}{4} p^4(1-p)^0 = 6p^2(1-p)^2 + 4p^3(1-p) + p^4$$

2 motorlu bir uçağın başarılı bir uçuş yapması;

$$\binom{2}{1} p(1-p) + \binom{2}{2} p^2(1-p)^0 = 2p(1-p) + p^2$$

Birinci formülün değeri ikinciden büyükse 4 motorlu uçak daha güvenlidir

$$6p^2(1-p)^2 + 4p^3(1-p) + p^4 \geq 2p(1-p) + p^2 = p \geq \frac{2}{3}$$

Motorun arızalanmaması olasılığı %66’dan büyükse 4 motorlu uçak daha güvenlidir.

Rastgele Değişkenler (Random Variables)

Ayrık Rastgele Değişkenler / 3 - Geometrik Değişken

p olasılığına sahip bir olayın kaçıncı denemede gerçekleşeceği ile ilgilenir.

$$p(n) = P\{X = n\} = (1 - p)^{n-1}p, \qquad\qquad i = 1,2,3,...$$

Rastgele Değişkenler (Random Variables)

Ayrık Rastgele Değişkenler / 4 - Poisson Değişken

λ parametresi ile tanımlanan ayrık rastgele değişkene Poisson değişkeni denir. Poisson değişkenleri belirli bir zaman dilimi içinde gerçekleşen olayları saymak için kullanılır.

Olayların belirli zaman dilimi içinde sabit bir λ sayısı ile gerçekleştiğini varsayar.

İlk uygulamalarından birinde bir yıl içinde at tepmesi sonucu ölen Prusya'lı askerlerin sayısını belirlemek için kullanıldı.

Bir dükkanı bir gün boyunca ziyaret eden müşterilerin sayısı …

$$p(i) = P\{X = i\} = e^{-\lambda}\frac{\lambda^i}{i!}, \qquad i = 1,2,3,...$$

Rastgele Değişkenler (Random Variables)

Devamlı Rastgele Değişkenler (Continuous Random Variables)

- Deneylerinin sayılamayacak adette çıktısı olan rastgele değişkenlere “Devamlı Değişkenler” adı verilir.
- Devamlı rastgele değişkenler “Probability Density Function” ile tanımlanır.f(x)
- (-∞, ∞) aralığındaki bir B setindeki bütün x değerleri için tanımlanmış bir f(x) fonksiyonu

$$P(x \in B) = \int_B f(x)dx$$

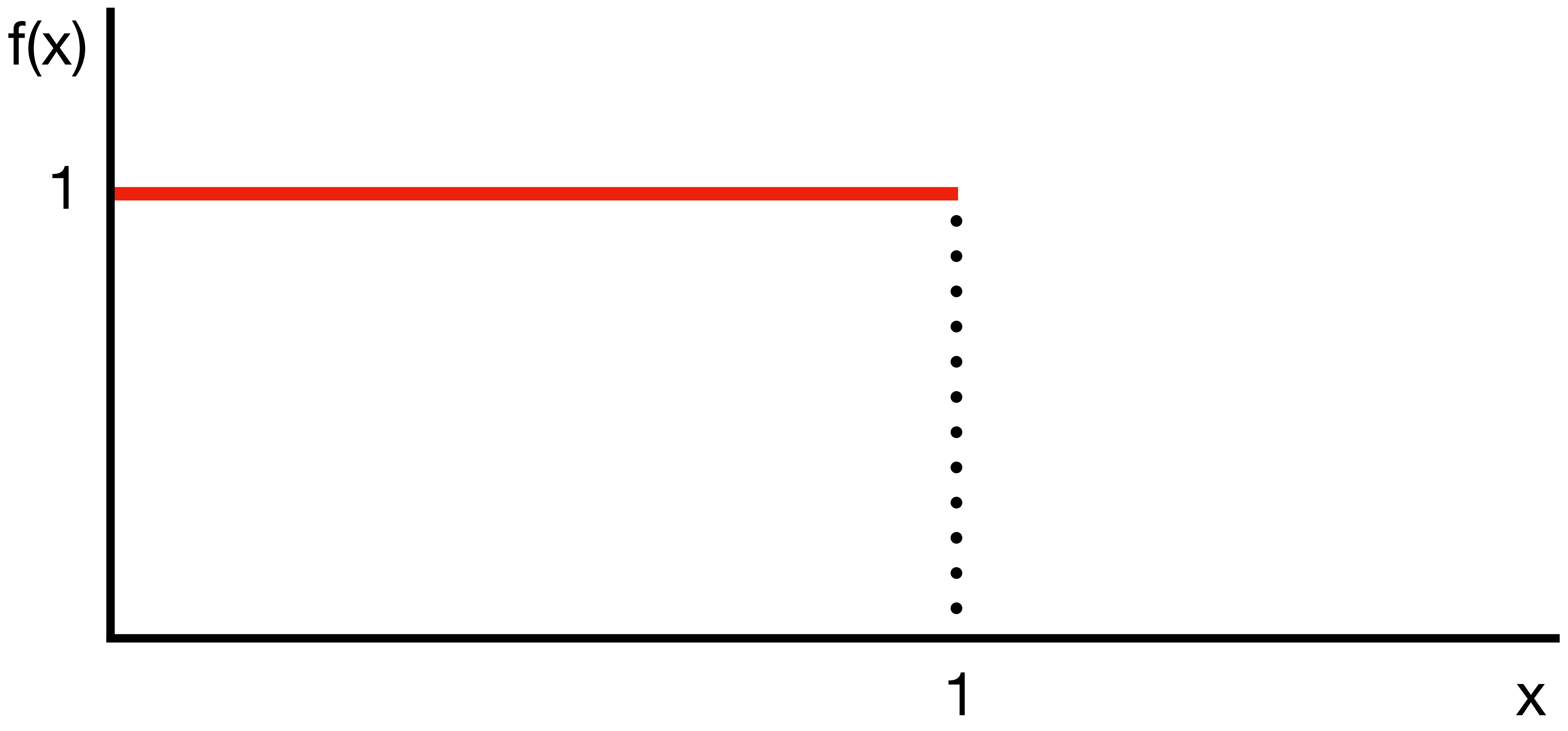
$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

$$P(X = a) = \int_a^a f(x)dx = 0$$

Rastgele Değişkenler (Random Variables)

Devamlı Rastgele Değişkenler / 1 - Tekdüze Değişken (Uniform)

$$f(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & otherwise \end{cases}$$



Rastgele Değişkenler (Random Variables)

Devamlı Rastgele Değişkenler / 1 - Tekdüze Değişken (Uniform)

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha}, & \alpha < x < \beta \\ 0, & otherwise \end{cases}$$

Örnek: X (0,10) aralığında tekdüze olarak dağılmışsa aşağıdaki olasılıkları hesaplayın.

a) X<3 b)X>7 c)1<X<6

$$a)P\{X < 3\} = \int_0^3 \frac{1}{10}dx = \frac{3}{10}$$

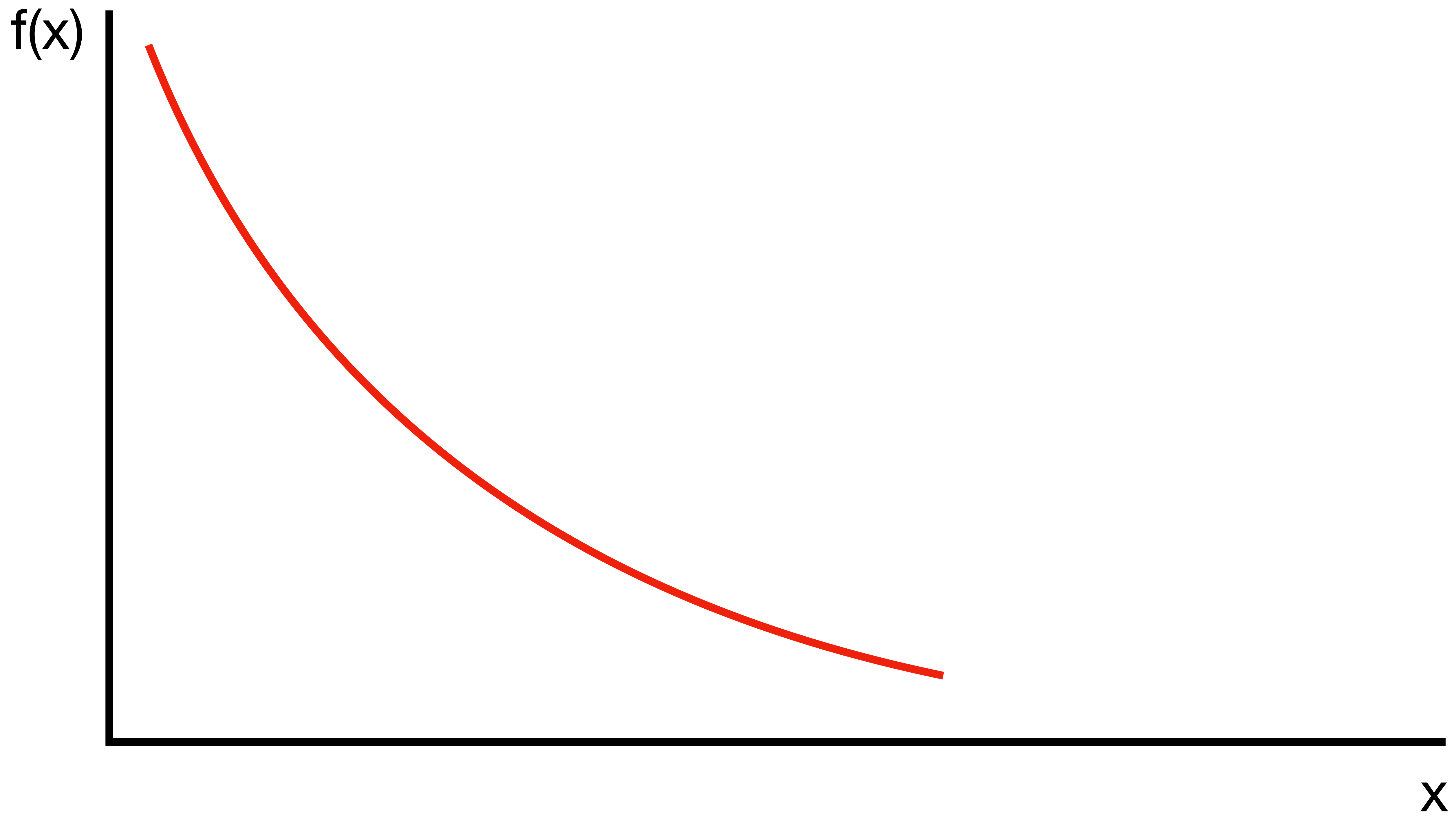
$$b)P\{X > 7\} = \int_7^{10} \frac{1}{10}dx = \frac{3}{10}$$

$$c)P\{1 < X < 6\} = \int_1^6 \frac{1}{10}dx = \frac{1}{2}$$

Rastgele Değişkenler (Random Variables)

Devamlı Rastgele Değişkenler / 2 - Üstel Değişken (Exponential)

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases}$$



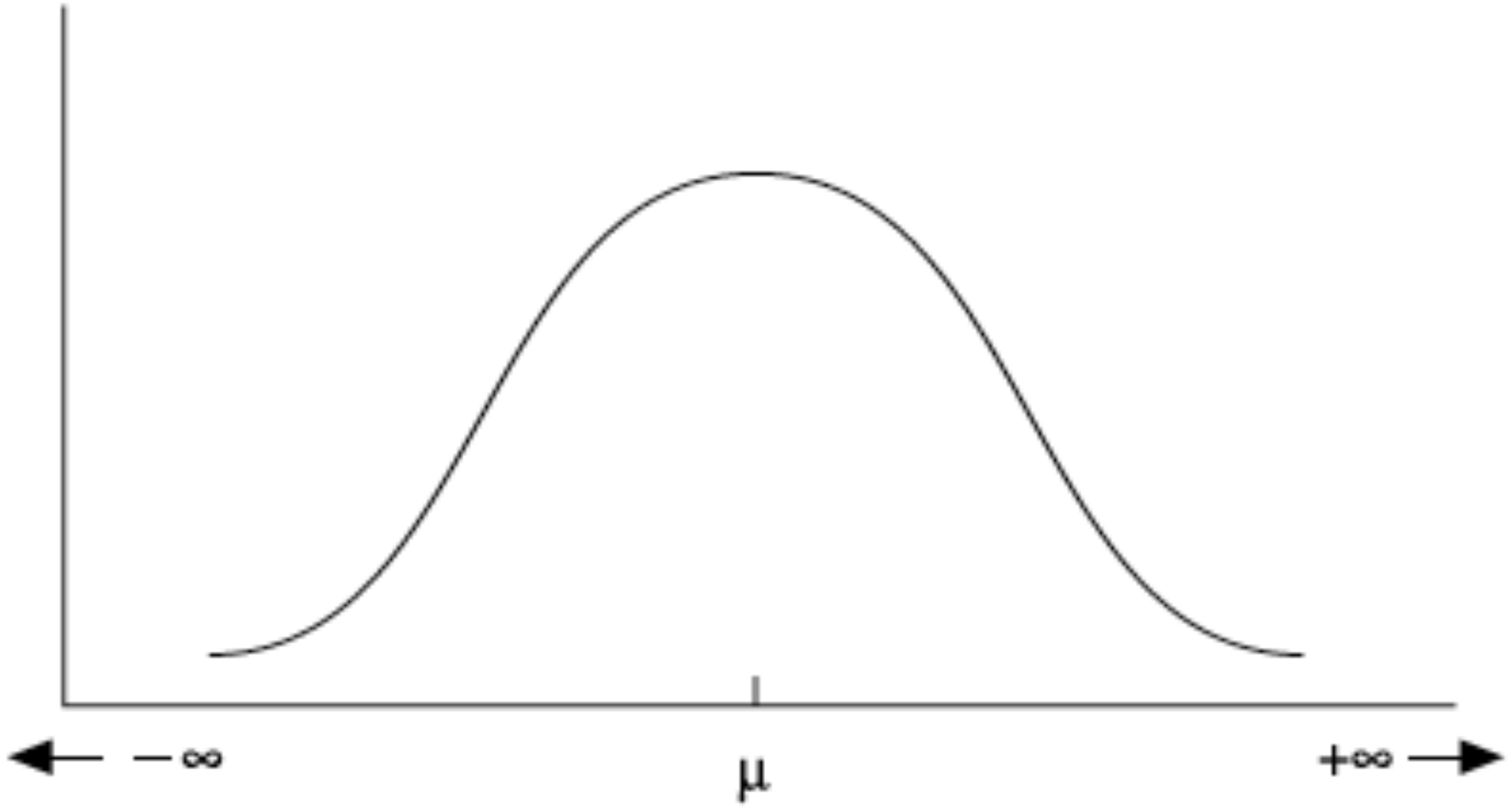
Rastgele Değişkenler (Random Variables)

Devamlı Rastgele Değişkenler / 3 - Normal Değişken

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \qquad -\infty < x < \infty$$

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (x_i - \mu)^2}$$



Rastgele Değişkenin Beklenen Değeri

Ayrık Değişkenler

$$E[X] = \sum_{x:p(x)>0} xp(x)$$

X'in beklenen değeri, alabileceği sayısal değerler ile o değerin olasılığının ağırlıklı ortalamasıdır.

$$P(1) = \frac{1}{4} \quad P(2) = \frac{1}{2} \quad P(3) = \frac{1}{4}$$

$$E[X] = \sum_{x:p(x)>0} xp(x) = 1 * \frac{1}{4} + 2 * \frac{1}{2} + 3 * \frac{1}{9} = \frac{19}{12}$$

Rastgele Değişkenin Beklenen Değeri

Ayrık Değişkenler

$$E[X] = \sum_{x:p(x)>0} xp(x)$$

Hilesiz bir zarı attığımızda deneyin toplam beklenen değeri kaçtır?

Rastgele Değişkenin Beklenen Değeri

Ayrık Değişkenler

$$E[X] = \sum_{x:p(x)>0} xp(x)$$

Hilesiz bir zarı attığımızda deneyin toplam beklenen değeri kaçtır?

$$P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6}$$

$$E[X] = \sum_{x:p(x)>0} xp(x) = 1 * \frac{1}{6} + 2 * \frac{1}{6} + 3 * \frac{1}{6} + 4 * \frac{1}{6} + 5 * \frac{1}{6} + 6 * \frac{1}{6} = \frac{7}{2}$$

Rastgele Değişkenin Beklenen Değeri

Ayrık Değişkenler

- Bernoulli

$$E[X] = p$$

- Binom

$$E[X] = np$$

- Poisson

$$E[X] = \lambda$$

Rastgele Değişkenin Beklenen Değeri

Devamlı Değişkenler

- Tekdüze

$$E[X] = \frac{\beta + \alpha}{2}$$

- Üstel

$$E[X] = \frac{1}{\lambda}$$

- Normal

$$E[X] =$$

Rastgele Değişkenin Beklenen Değeri

Devamlı Değişkenler

- Tekdüze

$$E[X] = \frac{\beta + \alpha}{2}$$

- Üstel

$$E[X] = \frac{1}{\lambda}$$

- Normal

$$E[X] = \mu$$

Strong Law of Large Numbers

Büyük Sayılar Kanunu

X_1, X_2, X_3, … bağımsız ancak aynı dağılıma sahip rastgele değişkenler olsun.

Eğer;

$$E[X_i] = \mu$$

$$\frac{X_1 + X_2 + X_3 + \ldots + X_n}{n} \to \mu \qquad as \quad n \to \infty$$

Rastgele Değişkenler & Beklenen Değer

Coşkunöz Vakfı - Veri Bilimci Eğitimi

Cem Yığman - 2022