14.30 Ekonomide İstatistiksel Yöntemlere Giriş
Bahar 2009

14.30 Ekonomide İstatistiksel Yöntemlere Giriş
Ders Notları 5

Konrad Menzel

19 Şubat 2009

1. Kesikli Rasgele Değişkenler

Tanım 1. *Eğer bir rasgele değişken olan X sadece sonlu sayıda değerler (ya da sayılabilir sonsuz) alırsa (x_1, x_2, …), X'in kesikli bir dağılımı vardır.*

Tanım 2. *Eğer rasgele değişken X'in kesikli bir dağılımı varsa, X'in olasılık yoğunluk fonksiyonu (p.d.f.) aşağıdaki gibi tanımlanır:*

$$f_X(x) = P(X = x)$$

Eğer { x_1, x_2, …} X'in muhtemel bütün değerlerinin kümesi ise, o zaman herhangi bir $x \notin$ { x_1, x_2, …} için, $f_X(x) = 0$'dır. Aynı zamanda,

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

$A \subset \mathbb{R}$ için, $X \notin A$'nın olasılığı,

$$P(X \in A) = \sum_{x_i \in A} f_X(x_i)$$

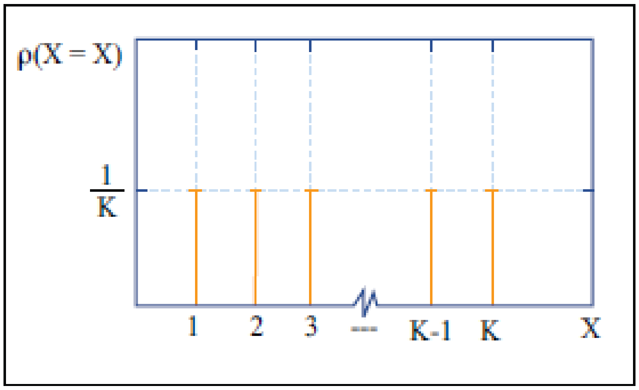
Örnek 1. *Eğer X fırlattığımız zarın yüzündeki sayı ise, bütün tam sayılar 1, 2, …, 6 eşit şanslıdır. Daha genel olarak, kesikli uniform dağılımını x_1, x_2, …, x_k sayılarının üzerinden dağılımın p.d.f'siyle tanımlayabiliriz.*

$$f_X(x) = \begin{cases} \frac{1}{k} & \text{eğer } x \in \{x_1 x_2, \dots, x_k\} \text{ ise} \\ 0 & \text{diğer bütün durumlarda} \end{cases}$$

Bu $S = \{x_1, x_2, \dots, x_k\}$ örneklem uzaylı bir deneyin basit olasılığıyla ilintilidir.

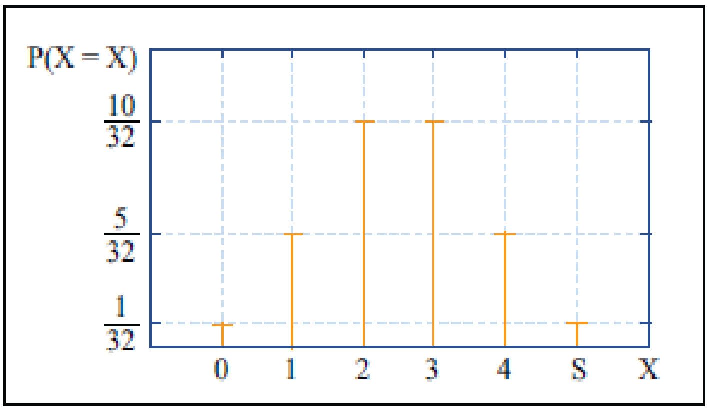
Örnek 2. *Varsayalım ki adil bir madeni parayı birbirinden bağımsız olarak 5 kere fırlattık ve X'i gözlemlenen turalar için bir rasgele değişken olarak tanımladık. O zaman, sayma kurallarımıza göre $n(S) = 2^5 = 32$ ve kombinasyon kuralının kullanarak ("k sayıda tura") = $\binom{5}{k}$ elde edilir. Dolayısıyla,*

$$P(X=0)=\binom{5}{0}\frac{1}{32}=\frac{1}{32}, \qquad P(X=1)=\binom{5}{1}\frac{1}{32}=\frac{5}{32},$$
$$P(X=2)=\binom{5}{2}\frac{1}{32}=\frac{10}{32}, \qquad P(X=3)=\binom{5}{3}\frac{1}{32}=\frac{10}{32},$$
$$P(X=4)=\binom{5}{4}\frac{1}{32}=\frac{5}{32}, \quad \text{ve} \quad P(X=5)=\binom{5}{5}\frac{1}{32}=\frac{1}{32}.$$



Kaynak: MIT OpenCourseWare

Bu olasılıkların toplamının 1'e eşit olduğuna not ediniz.



Zar atma örneğinde, her tek sonuç tam olarak rasgele değişkenin bir değeri ile ilintiliydi. Diğer yandan, beş madeni para atma durumunda, diyelim ki, X = 2, X = 0 ile karşılaştırıldığında, sonuçların sayıları arasında çok büyük farklar vardı. Sonuçların rasgele bir değişkenin gerçekleşen değerleri ile eşleştirilmesi, rasgele deneyin sonuçları eşit olasılığa sahip olsa bile, çok çarpık bir dağılıma yol açabilir.

1.1. Binom Dağılımı

Önceki örneği genelleştirmek için, her birisi “başarılı” ve “başarısız” olarak sonuçlanacak (olasılıkları eşit olmak zorunda olmayan) ardışık (dizi) n tane bağımsız ve benzer denemeyi gözlemlediğimizi ve toplam başarılı X sayısıyla ilgilendiğimiz varsayalım.

Örnek 3. *Kalite kontrolü için, bir üretim fabrikasında araba parçalar yığınından 100 parçalık bir örneklem seçtiğimizi varsayalım. Kalite kontrolünü geçen parça “başarılı” olarak belirlenirken, bir veya iki kritere uymayan parça ise “başarısız” olacaktır. Örneklemden hareketle parçaların %1’nden fazlasının standartlara uymadığına inanmak için iyi bir nedenimizin olmadığı sonucuna varmak istiyoruz. Bunun için de toplam içindeki bozuk parçaların toplam payının bir fonksiyonu olarak başarısızlığın dağılımıyla ilgileniyoruz.*

Varsayalım ki başarının olasılığı *p*’ye eşit olsun, dolaysıyla başarısızlığın olasılığı $q = 1 - p$’dir. Denemeler varsayımsal olarak bağımsız olduğu için, herhangi bir verili x başarı ve n – x başarısız dizisinin *sabit sıralı* olasılığı aşağıdaki gibidir:

$$p^x(1 - p)^{n - x}$$

Ancak, sadece *başarılı* X *sayısı* ile ilgilendiğimiz için, $\binom{n}{x}$ sayıda başarı x’li farklı dizi olduğunu hesaba katmak zorundayız.

$$P(X = x) = \binom{n}{x} p^x(1 - p)^{n - x}$$

Tanım 3. *Aşağıdaki gibi bir olasılık yoğunluk fonksiyonu olan X rasgele değişkenin X ∽ B(n, p) şeklinde yazılan p ve n parametreli bir binom dağılımlı olduğu söylenebilir.*

$$f_X(x) = P(X = x) = \begin{cases} \binom{n}{x} p^x(1-p)^{n-x} & \text{Eğer} \quad x \in \{1, 2, \ldots, n\} \text{ ise} \\ 0 & \text{Diğer bütün durumlarda} \end{cases}$$

Daha önce her bir rasgele deney için ayrı ayrı olarak olasılık dağılımı türettiğimize, olası sonuçlarını, her olayın sonuçlarını vs. yazdığımıza dikkat etmelisiniz. Binom dağılımı rasgele deneylerin tümüne hizmet eden bir modeldir. Bu kategoriye düşen herhangi bir örnek için, verilen bir (*n, p*) parametre setinin sadece olasılıklara bakarız.

Örnek 4. *Bir sınıf arkadaşınızdan biraz para koparmak için, tura gelme olasılığı* $p_L = \frac{4}{5}$ *olan yamuk bir 1 madeni parayı temin ettiğinizi düşünün. Maalesef, o para sizin diğer normal paralarınızla karıştı ve ancak 9 liranın 8 lirasını kola makinesine attığınızda fark ettiniz. Aceleyle hemen parayı 10 kere fırlatınız ve toplam olarak 8 tanesi tura geldi. Bilinen eski bir madeni para hilesiyle arkadaşınızı soymaya devam etmeye çalışmak iyi* *bir fikir olur muydu? Yoksa $p_F = \frac{1}{2}$ olasılıklı sıradan (adil) oyuna bağlı mı kalırdınız? A = "kalan madeni para yamuktur" ve B = "10'da 8 Tura" için P(A|B) kaçtır? Eğer madeni para adil ise,*

$$P(B|A^C) = \binom{10}{8} p_F^8 (1-p_F)^{10-8} = \binom{10}{8} \frac{1}{2^{10}}$$

Eğer yamuk ise

$$P(B|A) = \binom{10}{8} p_L^8 (1-p_L)^{10-8} = \binom{10}{8} \frac{4^8 \cdot 1^2}{5^{10}}$$

Şimdi Bayes Teoremin ne söyleyeceğine bakalım:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)} = \frac{\binom{10}{8} \frac{4^8 \cdot 1^2}{5^{10}} \frac{1}{9}}{\binom{10}{8} \frac{4^8 \cdot 1^2}{5^{10}} \frac{1}{9} + \binom{10}{8} \frac{1}{2^{10}} \frac{8}{9}} = \frac{\frac{4^8}{5^{10}}}{\frac{4^8}{5^{10}} + \frac{8}{2^{10}}} \approx 46.21\%$$

Dolayısıyla, geri kalan madeni paranın gerçekte daha çok sıradan bir madeni para olma olasılığı yüksektir - ki toplam öyle demiyor, çünkü tura olasılığı hala

$$P(H|B) = \frac{4}{5} \cdot 46.21\% + \frac{1}{2} \cdot 53.79\% \approx 63.86\%$$

Ancak, elbette parayı birkaç kere daha atmaktır daha iyi bir fikir olurdu. Eğer deneyi keyfi olarak sık sık tekrarlarsanız, o zaman en sonunda iki para arasındaki farkı ihtiyari bir doğruluk derecesiyle tespit edebilecek duruma gelirsiniz. Bir parantez açmak gerekirse, bu uygulamayı basit hipotez testi için bir örnek olarak görebilirsiniz. Diyelim ki, turanın oranını daha önce ki gibi tutmak şartıyla, diğer bir 10 denemede yine 8 tura elde ettiniz (buna olay C diyelim). O zaman, önceki adımların aynısın takip ederek, bu sefer evvelki *P(B)* yerine sonraki *P(H|B)*'ye dayanarak koşulu olasılık şöyle olur:

$$P(H|B \cap C) = \frac{\binom{10}{8} \frac{4^8 \cdot 1^2}{5^{10}} \cdot 46.21\%}{\binom{10}{8} \frac{4^8 \cdot 1^2}{5^{10}} \cdot 46.21\% + \binom{10}{8} \frac{1}{2^{10}} \cdot 53.79\%} = \frac{46.21\%}{46.21\% + 53.79\% \cdot \frac{5^{10}}{4^8 2^{10}}} \approx 85.51\%$$

Diğer bir seçenek olarak, eğer iki seriyi 16 tura ve 20 deneme olarak toplulaştırırsak,

$$P(A|B') = \frac{\binom{20}{16} \frac{4^{16}}{5^{20}}}{\binom{20}{16} \frac{4^{16}}{5^{20}} + \binom{20}{16} \frac{8}{2^{20}}} \approx 85.51\%$$

elde deriz.

Böylece, güncellemeyi aynı anda veya farklı zamanda yapmış olmamız fark etmeyecektir. Bu genel olarak arzulanan Bayesyen güncellemenin özeliğidir: nihai sonuç kullandığımız genel bilgiye bağlıdır, güncelleme sırasına değil.

2. Sürekli Rasgele Değişken

Verinin birçok türü bir çeşit ölçümün sonucudur: en azından kavramsal olarak, bir reel sayı aralığındaki herhangi bir değeri (bazen tümünü) alabilen ağırlık, uzunluk, gelir vs. gibi. Bu durumda, bir kesikli değişkenin olasılık yoğunluk fonksiyonun tanımı kullanışlı değildir, çünkü (a) olası sonuçların miktarı sayılabilir değildir, bu nedenle de sadece tek tek sayıların olasılıklarının toplamını alamayız ve (b) bölünmez bir bütünün *belirli* bir değerinin olasılığı sıfırdır. Bu yüzden, kesikli durumdan ayrı olarak, bu tür rasgele değişkenler ile ilgilenmek zorundayız.

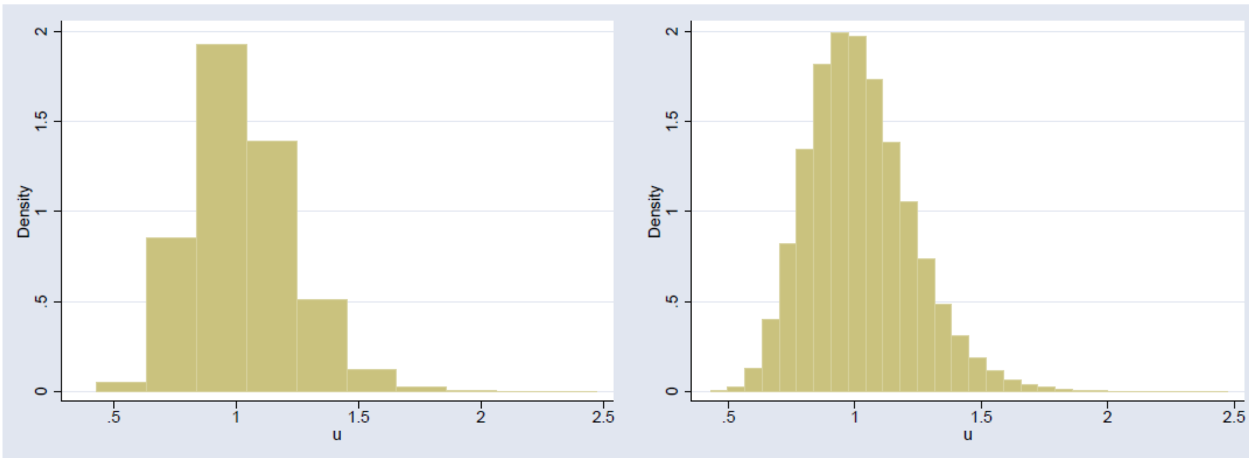
Tanım 4. *Eğer X bir sayı doğrusundaki bir aralığın (sınırlandırılmış veya sınırlandırılmamış) herhangi bir değerini alırsa, rasgele değişken X'in sürekli bir dağılımı vardır.*

Kesikli rasgele değişkenler için olasılık yoğunluk fonksiyonu tanımlamak nispetten daha basitti çünkü sınırlı sayıda değeri vardı. Sürekli bir rasgele değişken sayılabilir rakamlardan daha fazla değer alır, bu yüzden de elde etmek fazlaca çaba gerektirir. Şöyle ki: rasgele değişkenin alabileceği muhtemel değerleri bir "grup"a koyarak dağılımı "kesikleştiriyoruz", yani $P(X = x)$ olasılığına bakmak yerine $P(x_1 \leq X \leq x_2)$ gibi bir aralığın olasılığına bakıyoruz. Bunun grafiksel gösterimi *histogram*dır: sayı doğrusu üzerinde bir grup sayıyı, $x_0 < x_1 < \ldots < x_n$, sabitleştiriyoruz ve X değerinin "grup"lara düşme olasılığını hesaplıyoruz. Grup birbirinin devamı iki sayı aralığındır, yani $P(x_{i-1} \leq X \leq x_i)$. Ondan sonra $[x_0, x_1]$ aralığındaki değerler için aşağıdaki fonksiyonu tanımlıyoruz:

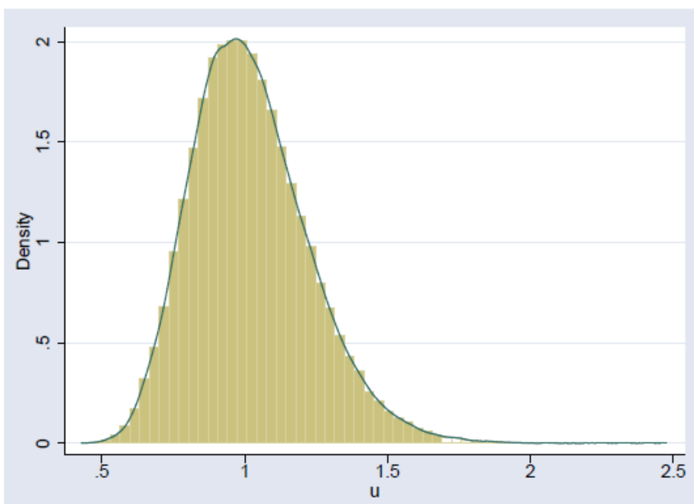
$$h_n(x) = \begin{cases} \frac{P(x_0 \leq X \leq x_1)}{x_1 - x_0} & \text{Eğer} \quad x \in [x_0, x_1) \\ \vdots & \\ \frac{P(x_{i-1} \leq X \leq x_i)}{x_i - x_{i-1}} & \text{Eğer} \quad x \in [x_{i-1}, x_i) \\ \vdots & \\ \frac{P(x_{n-1} \leq X \leq x_n)}{x_n - x_{n-1}} & \text{Eğer} \quad x \in [x_{n-1}, x_n) \end{cases}$$

$[x_{i-1}, x_i)$ aralığının uzunluğuyla bölmek, verili bir aralıkta grafiğin altında kalan alanın rasgele değişken X'in aralıktaki bir değeri alma olasılığına eşit olmasını garantiler. Yani şunu hesaplayabiliriz:

$$P(x_j \le X \le x_k) = \sum_{i=j+1}^{k} P(x_{i-1} \le X \le x_i) = \sum_{i=j+1}^{k} \left[\int_{x_{i-1}}^{x_i} h_n(t)dt \right] = \int_{x_j}^{x_k} h_n(t)dt \quad \text{for } x_j < x_k$$



Sekil 1: Aynı Dağılımın, sırasıyla, 10 ve 30 Gruplu Histogramı

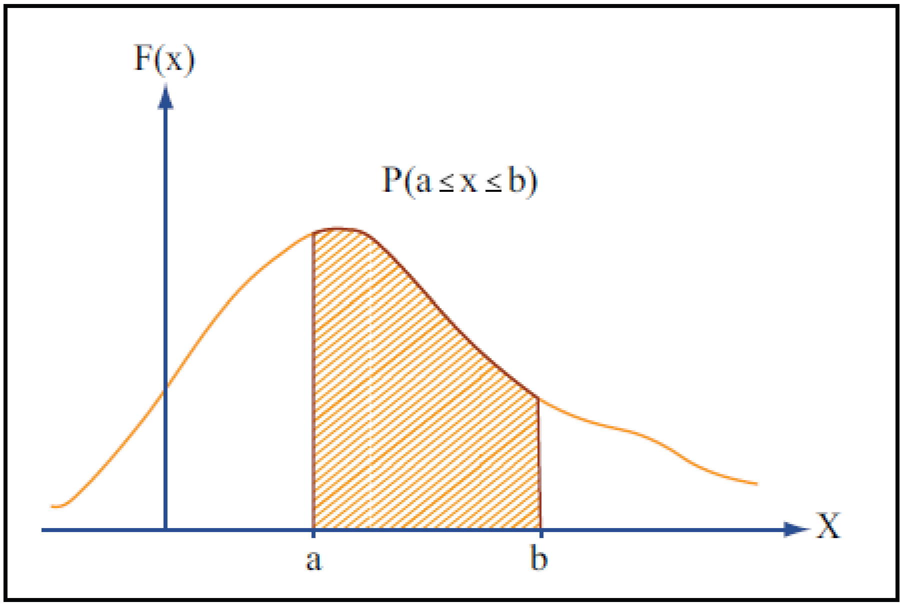


Şekil 2: 60 Gruplu Histogram ve Sürekli Yoğunluk

Bu henüz tam tatmin edici değil, çünkü bu sadece bize $x_0 < x_1 < \ldots < x_n$ sayılarının iki noktası arasına düşen X'in olasılığını hesaplamamıza yarar, yoksa [x_j, x_k] gibi bir alt aralığın içine düşeni değil. Bu sorunu, x_1, x_2, … silsilesini küçülterek ve böylece aralığı daraltarak çözebiliriz. Bir birine komşu iki nokta, x_{i-1}, x_i, arasını ihtiyari bir küçük "dx" değeri kadar daraltırsak, X'in a ve b gibi iki nokta arasına düşmesini a'dan b'ye $h_\infty(x)$ integrali olarak veren $h_\infty(x)$ fonksiyonunu elde ederiz. Bu limit bir sürekli rasgele fonksiyonun olasılık yoğunluk fonksiyonu olarak adlandırılır.

Tanım 5. Eğer rasgele değişken X sürekli bir dağılıma sahip ise, X'in olasılık yoğunluk fonksiyonu (p.d.f.) pozitif bir $f_x(x)$ fonksiyonu olarak tanımlanır yani $A \subset \mathbb{R}$ gibi herhangi bir aralık için

$$P(X \in A) = \int_A f_X(t)dt$$



Kaynak: MIT OpenCourseWare

Şekil 3. Sürekli Rasgele bir Değişkenin P.D.F.'si

Olasılık fonksiyonun aksiyomlarından, herhangi bir sürekli p.d.f.'nin aşağıdaki ilişkiyi sağlamak zorunda olduğunu görebiliriz.

$$f_X(x) \geq 0 \qquad \forall x \in \mathbb{R}$$

ve

$$\int_{-\infty}^{\infty} f_X(x) = 1$$

Böylece, Eğer *A = [a, b]* $\subset \mathbb{R}$ için *P(X ∈ A)*'yi bilmek istersek, aşağıdakini hesaplayabiliriz:

$$P(X \in [a,b]) = P(a \leq X \leq b) = \int_a^b f(t)dt$$

Açıklama 1. *Eğer X sürekli bir dağılıma sahip ise, herhangi bir x ∈ ℝ için*

$$P(X = x) = 0$$

Bu, kısmi olarak sezgilere aykırı gibi görünebilir çünkü biz gerçekte kesikli olan şeyleri (gelir, işsizlik süresi gibi) tahmin etmek için sürekli dağılımı kullanıyoruz. Şimdiye kadar, hesapladığımız herhangi bir olasılık için herhangi bir sürekli rasgele değişken örneği görmedik.

3. Örnekler

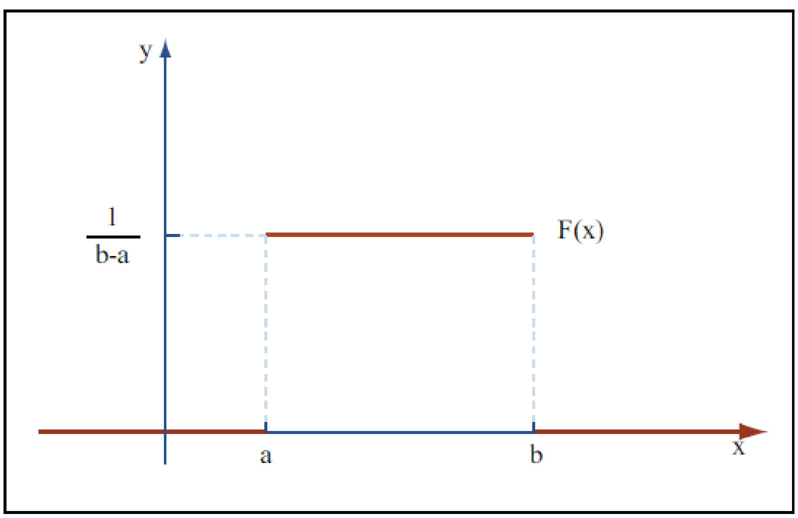
Bir rasgele değişkenin, reel eksen üzerindeki bazı [a, b] aralıklarında yer aldığını varsayalım, X'in bazı [a',b'] (burada a ≤ a' ≤ b'≤ b) alt gruplarına ait olma olasılığı, alt aralığın uzunluğu ile orantılıdır. .

Tanım 6. *Eğer aşağıdaki gibi bir olasılık yoğunluk fonksiyonuna sahip ise, X rasgele değişkeni [a, b], a < b, aralığında uniform dağılır.*

$$f_X(x) = \begin{cases} \frac{1}{a-b} & eğer\ a \leq x \leq b\ ise \\ 0 & diğer\ bütün\ durumlarda \end{cases}$$

Sembolik olarak şöyle yazarız:

$$X \sim U[a, b]$$



Kaynak: MIT OpenCourseWare

Şekil 4. Uniform Rasgele Değişken için p.d.f., *X ~ [a, b]*

Örneğin, eğer *X ~ U[0, 10]* ise, o zaman

$$P(3 < X < 4) = \int_3^4 f(t)dt = \int_3^4 \frac{1}{10}dt = \frac{1}{10}$$

P(3 ≤ X ≤ 4) nedir? Olasılık P(X = 3) = 0 = P(X = 4) olduğu için, bu P(3 < X < 4)'ün aynısıdır.

Örnek 5. *Varsayalım ki X'in p.d.f.'si şöyledir:*

$$f_X(x) = \begin{cases} ax^2 & eğer\ 0 < x < 3\ ise \\ 0 & diğer\ bütün\ durumlarda \end{cases}$$

a ne olmak zorundadır? P(X ∈ ℝ) = 1 olduğu için, yoğunluk 1'e entegre olmalı, böylece a

$$1 = \int_{-\infty}^{\infty} f_X(t)dt = \int_0^3 at^2 dt = \left[\frac{ax^3}{3}\right]_0^3 = \frac{27}{3}a - 0 = 9a$$

olmak zorundadır. Böylece a = $\frac{1}{9}$.

P(1 < X < 2) nedir? Aşağıdaki integrali hesaplayalım

$$P(1 < X < 2) = \int_1^2 \frac{t^2}{9}dt = \frac{2^3}{9 \cdot 3} - \frac{1^3}{9 \cdot 3} = \frac{7}{27}$$

P(1 < X) nedir?

$$P(1 < X) = \int_1^{\infty} f_X(t)dt = \int_1^3 \frac{t^2}{9}dt = \frac{27 - 1}{27} = \frac{26}{27}$$