14.30 Ekonomide İstatistiksel Yöntemlere Giriş
Bahar 2009

Problem Seti 3 - Çözümleri
14.30 Ekonomide İstatistiksel Yöntemlere Giriş

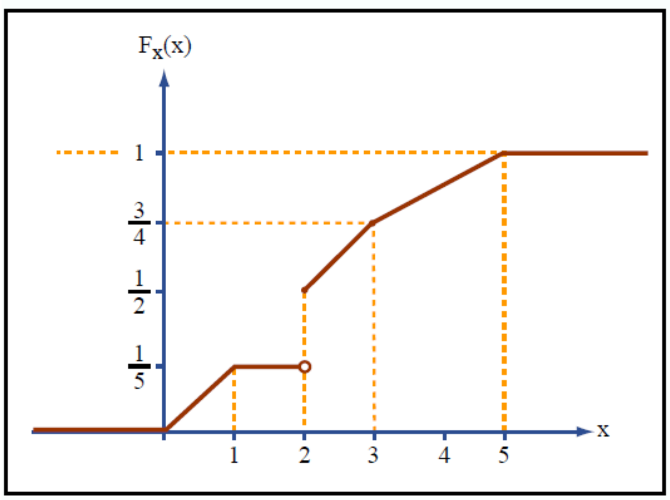
Konrad Menzel

Son Gün: 3 Mart 2009

Soru Bir

1. Kümülatif dağılım fonksiyonun(CDF) tanımını yazınız. Kelimeler ile ne anlama geldiğini açıklayınız, bir örnek veriniz.
   - 1'in Çözümü: CDF'nin bir tanımı f(.): R→ [0, 1], burada f(x) = $P_r(X \leq x)$. CDF bize rasgele değişken X'in sıralı desteğinin belli bir noktasına kadar olan birikimli olasılıkları verir. Bu, bir şeyin bir x sonucundan küçük veya eşit (ya da solunda, bu sıralamayı nasıl yorumlamak istediğinize bağlıdır) olma ihtimallerinin ne olduğunu bilebileceğimiz anlamına gelir.

2. Aşağıdaki fonksiyonun geçerli bir CDF olup olmadığını gösteriniz. Eğer geçerliyse, ilgili PDF'nin grafiğini çiziniz.

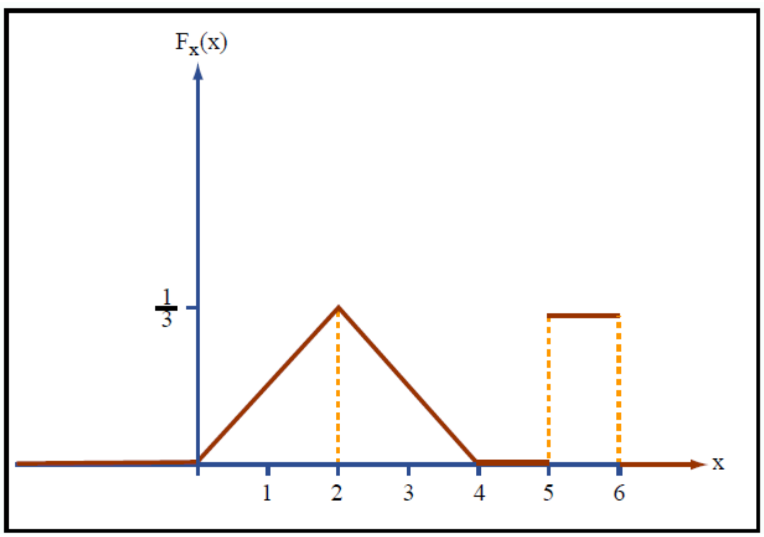


Kaynak: MIT OpenCourseWare

   - 2'nin Çözümü: Fonksiyon gerçekte geçerli bir CDF'dir. Alta sıfır ve üste bir ile sınırlandırılmıştır. Ayrıca sol, $\lim_{x \to -\infty} F_X(x) = 0$, ve sağ, $\lim_{x \to \infty} F_X(x) = 1$, limit koşullarını da sağlamaktadır. Ancak, bu karışık bir rasgele değişkendir çünkü sürekli bir dağılımı vardır ve 2 etrafında bir kütle noktası vardır. PDF ise aşağıdaki denklemdir:

$$f(x) = \begin{cases} 0 & x < 0 \cup 5 \leq x \\ \frac{1}{5} & 0 \leq x < 1 \\ 0 & 1 \leq x < 2 \\ \frac{3}{10} & x = 2 \\ \frac{1}{4} & 2 \leq x < 3 \\ \frac{1}{8} & 3 \leq x < 5 \end{cases}$$

3. Aşağıdaki fonksiyonun geçerli bir PDF olduğunu gösteriniz ve ilgili CDF'yi çiziniz.

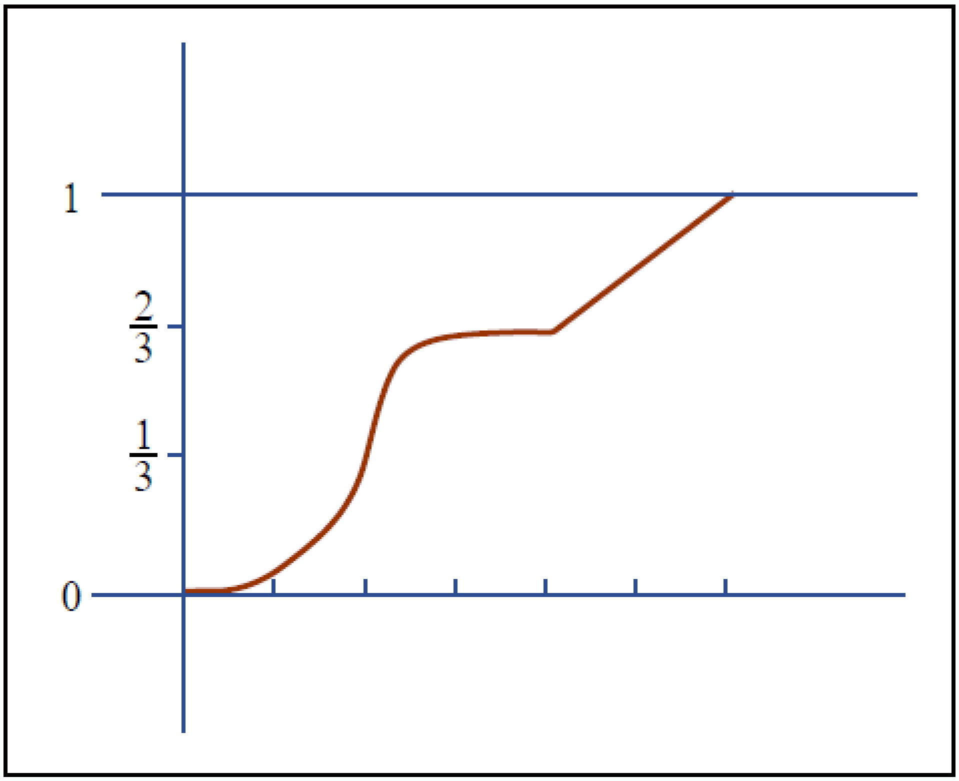


Kaynak: MIT OpenCourseWare

- 3'ün Çözümü: Bu fonksiyon gerçekte geçerli bir PDF'dir. Her noktada pozitiftir ve integrali birdir (üçgenin alanı 2/3'tür ve 5 ile 6 arasındaki aralığın alanı 1/3'tür, bu da toplamı 1 yapar).CDF doğrudan türetilir. Kolay integral için analitik olarak önce PDF'yi ve sonra CDF'yi aşağıya yazacağım:

$$f_X(x) = \begin{cases} 0 & x < 0 \cup 6 \leq x \\ \frac{1}{6}x & 0 \leq x < 2 \\ \frac{1}{6}(4-x) & 2 \leq x < 4 \\ 0 & 4 \leq x < 5 \\ \frac{1}{3} & 5 \leq x < 6 \end{cases} \qquad F_X(x) = \begin{cases} 0 & x < 0 \cup 6 \leq x \\ \frac{1}{12}x^2 & 0 \leq x < 2 \\ \frac{2}{3} - \frac{1}{12}(4-x)^2 & 2 \leq x < 4 \\ \frac{2}{3} & 4 \leq x < 5 \\ \frac{2}{3} + \frac{1}{3}(x-5) & 5 \leq x < 6 \end{cases}$$

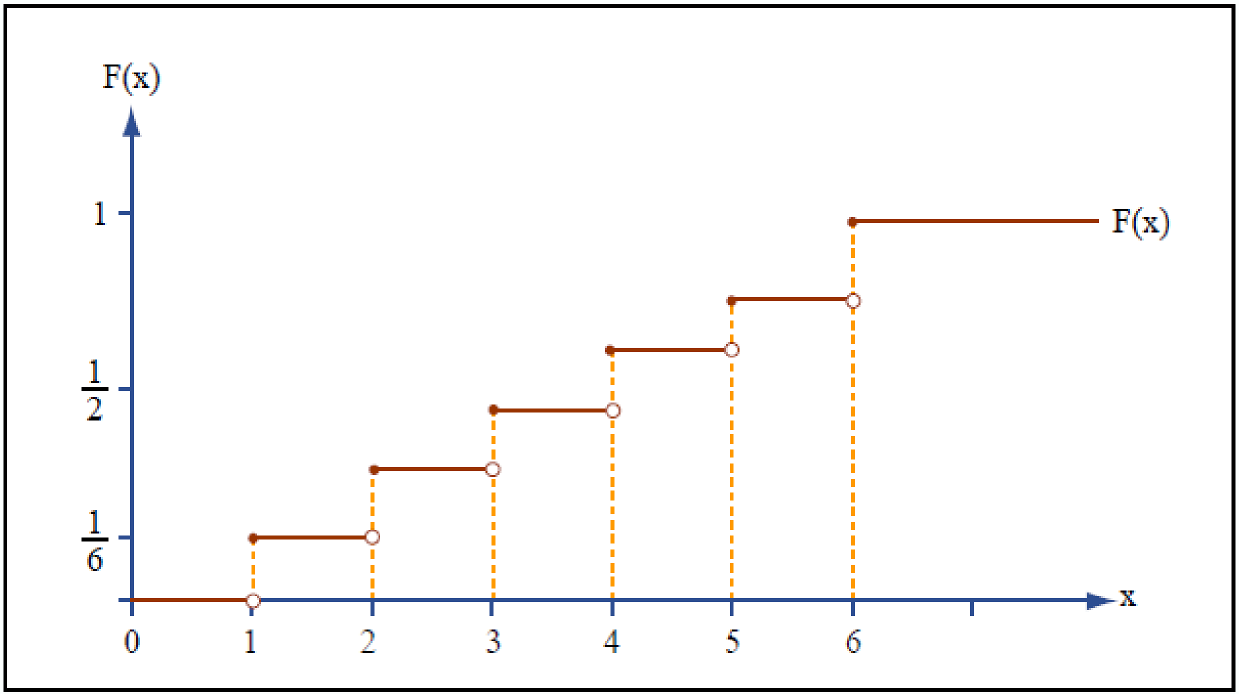
Ben bir grafik tasarımcısı değilim ancak en azından detaylara biraz dikkat ederseniz bu grafiği çizmek nispetten kolay olsa gerek.



Kaynak: MIT OpenCourseWare

Soru İki

1. C.d.f.si sürekli olmayan fakat sağdan-sürekli olan bir p.d.f gösteriniz.
   - 1'in Çözümü: Bu en az bir kütle noktası(mass point) olan bir dağılımdan (ya da tamamen kesikli olan bir dağılımdan) gelmek zorundadır. Konrad'ın ders notları CDF'nin bir örneğini içermektedir:



Kaynak: MIT OpenCourseWare

Bu grafik ile ilintili PDF, basit haliyle bir zarın bir yüzündeki değer olarak tanımlanan X rasgele değişkeni tarafından türetilen bir PDF'dir. PDF şöyledir:

$$f(x) = \begin{cases} \frac{1}{6} & \text{eğer } x = 1,2,3,4,5,6 \text{ ise} \\ 0 & \text{diğer bütün durumlarda} \end{cases}$$

Doğru/Yanlış/Belirsiz: Eğer ifade doğru ise her zaman kısa bir açıklama yapınız, ya da eğer ifade yanlış veya belirsiz ise her zaman karşı örnek geliştiriniz ve karşı örneği kısaca açıklayınız.

1. Eğer P(A|B) > P(A) ve P(A|C) > P(A), o zaman P(A|B, C) > P(A).
   - 1'in Çözümü: Yanlış. Çünkü iki koşullu olasılığın büyük olması, onların bileşik olasılığının büyük olmayacağı anlamına gelmez. Kış mevsimi veri iken (olay B), hastalanma olasılığı (olay A) bir örnek olabilir. Kış mevsiminde hastalanma olasılığınız yıl ortalamasına göre daha yüksektir. Aynı zamanda, almanız gerekmeyen bir ilacı aldığınız (rasgele olay C) zaman da büyük ihtimalle hastalanırsınız. Ancak, eğer sadece herkesin en çok hastalandığı kış döneminde koruyucu ilaç alırsanız, hastalanma ihtimaliniz daha düşük olur. Aynı şey grip aşısı için de geçerlidir(aşı olursanız hastalanma olasılığı

düşüktür, kış mevsimindeyseniz hastalanma olasılığı yüksektir, interaktif olasılık potansiyel olarak düşüktür).

2. Sürekli bir p.d.f hiçbir zaman 1'den büyük değer almaz.
   - 2'nin Çözümü: Yanlış. Sürekli bir PDF örneği uniform PDF olabilir, Yani eğer $0 \leq x \leq 1/6$ ise $f(x) = 6$'dır ve diğer bütün durumlarda da sıfırdır. Bu PDF açıkça iyi tanımlanmıştır ve sayılmayan sonuçlar kümesinde birden büyük bir değerdir. Dahası, bu ifade en uygun şekilde CDF için geçerlidir.
3. $P(A) = P(A|B)P(B)$ A ile B'nin bağımsız olduğu anlamına gelir.
   - 3'ün Çözümü: Yanlış. Bunun anlamı, toplam olasılık kanununun ifade ettiği gibi A ile B gerçekte ilişkilidir yani

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$

Bu nedenle, bunun anlamı ya $P(A|B^C) = 0$ (bunun anlamı A ancak B gerçekleşince gerçekleşir, bu da ikisinin bağımsız olmadığını ima eder) ya da $P(B^C) = 0$ (bunun anlamı B = S (örneklem uzayı)'dir ve o zaman A ile B bağımsızdır). A ile B'nin basit bir örneği şöyle olabilir: bir tek zar atış uzayında, A = {4 atma} ve B = {Çift sayı atma}. Bu örnekte, $P(A) = 1/6$, $P(B) = 1/2$, ve $P(A|B) = 1/3$. Ancak, $P(A \cap B) = 1/6 \neq P(A)P(B) = (1/6)(1/2)$. Ancak, eğer B = S ise, o zaman $P(S) = 1$ iken gerçekte elimizde $P(A|S)P(S) = P(A)P(S) = P(A).1 = P(A)$ olduğu anlaşılıyor. Bu, ikisinin de bağımsız olduğu ve sorudaki koşulların aynı anda sağlandığı anlamına gelir. Bu hala ilginç bir sonuç değildir. :)

Soru üç

(kaynak: Bain/Engelhardt, Ch. 2, ex. 8)

Negatif olmayan tam-sayı alan bir rasgele X değişkeni x = 0, 1, 2 için $F(x) = 1 - (1/2)^{x+1}$ CDF formuna sahiptir. Bu CDF $x < 0$ için sıfır değerini alır.

1. X'in PDF'sini bulunuz.
   - 1'in Çözümü: PDF CDF'nin her noktasındaki farktır. Bu rasgele bir tam sayı olduğu için, bitişik kesikli sonuçların farklarını olasılık olarak hesaplamamız gerekiyor.

$$
\begin{aligned}
F(x) - F(x-1) &= \left(1 - (1/2)^{x+1}\right) - \left(1 - (1/2)^{x-1+1}\right) \\
&= (1/2)^x (1 - (1/2)) \\
f(x) &= \frac{1}{2^{x+1}}
\end{aligned}
$$

Aynı zamanda negatif herhangi bir tam sayı için PDF f(x) = 0 gibi bir şeydir.

2. P[10<X≤20]'nı bulunuz.
   - 2'nin Çözümü: Aranızda geometrik serileri daha önce görenler, bunun sonlu farkların toplamı olduğunu hemen anlamışlardır.

$$
P[10 < X \leq 20] = \sum_{x=11}^{20} \frac{1}{2^{x+1}} = \frac{\frac{1}{2^{11+1}} - \frac{1}{2^{20+1+1}}}{1 - \frac{1}{2}} = \frac{1}{2^{11}} - \frac{1}{2^{21}}
$$

Ya da halı hazırda CDF'miz var olduğu için, sadece aşağıdakini kullanabilirsiniz:

$$
\begin{aligned}
P[10 < X \leq 20] &= F(20) - F(10) \\
&= \left(1 - \frac{1}{2^{21}}\right) - \left(1 - \frac{1}{2^{11}}\right) \\
&= \frac{1}{2^{11}} - \frac{1}{2^{21}}
\end{aligned}
$$

3. P[X çift sayıdır]'nı bulunuz.
   - 3'ün Çözümü: sadece çift sayıları istediğimiz için, diğer sayılar için hesaplama yapmaya ihtiyacımız var. Bunu her bir x'i 2 ile çarparak kolayca yapabilriz. Bu bize PDFdeki bütün çift sayıları atlamamıza olanak verecektir. Ardından sonsuz toplamı almak yeterlidir:

$$
\begin{aligned}
P[X \text{ çifttir }] &= \sum_{x=0}^{\infty} \frac{1}{2^{2x+1}} \\
&= \frac{1}{2} \sum_{x=0}^{\infty} \frac{1}{4^x} \\
&= \frac{1}{2} \frac{1}{1 - \frac{1}{4}} \\
P[X \text{ çifttir }] &= \frac{2}{3}
\end{aligned}
$$

Soru Dört

1. Varsayalım ki bir rasgele değişken x'in [0, 1] aralığıyla orantılı bir PDF'si vardır. Bu PDF için bir formül yazınız. İlgili CDF nedir?
   - 1'in Çözümü: PDF x ile orantılıdır:

     eğer $x \in [0, 1]$ ise $f(x) = kx$'tir, bütün diğer durumlarda ise sıfırdır

     PDF'nin integral özeliğiyle, $k = 2$ olduğunu biliyoruz. CDF sadece PDF'nin integralidir:

     eğer $x \in [0, 1]$ ise $F(x) = X^2$

     Tamamlamak için CDF'yi şöyle tanımlıyoruz: $F(x) = 1$ eğer $x > 1$ ise ve $F(X) = 0$ eğer $X < 0$ ise.

2. Şimdi ise, o rasgele değişken x'in [0, 1] aralığıyla orantılı bir CDF'si vardır. Bu CDF için bir formül yazınız. İlgili PDF nedir?
   - 2'nin Çözümü: CDF x ile orantılıdır:

     eğer $x \in [0, 1]$ ise $F(x) = kx$

CDF'nin özelliklerine göre $k = 1$ CDF'nin 0 ve 1 ile sınırlandırıldığını kesinleştirir. Aynı zamanda bütünlük için CDF'yi şöyle tanımlıyoruz: $F(x) = 1$ eğer $x > 1$ ise ve $F(X) = 0$ eğer $X < 0$ ise (bunu sınavda ve problem setinde yapmayı unutmayınız – reel sayıların desteği için PDF ve CDF'yi tanımlamaya gereksiniminiz var). İlgili PDF türevdir:

eğer $x \in [0, 1]$ ise $f(x) = 1$'dir, bütün diğer durumlarda ise sıfırdır

Soru Beş

Varsayalım ki X ile Y'nin bileşik PDF'si şöyledir:

$$f_{X,Y} = \begin{cases} kx^3y & \text{eğer } 0 < x < y < 1 \text{ ise} \\ 0 & \text{diğer bütün durumlarda} \end{cases}$$

1. k'nin değeri nedir?
   - 1'in Çözümü: f_{XY}'nin bir PDF olabilmesi için, k'nin değeri yoğunluğun integralinin 1 olmasını sağlamak zorundadır:

$$
\begin{aligned}
\int_{X\times Y} f_{X,Y}\,dydx &= 1\\
\int_0^1\int_x^1 kx^3y\,dydx &=\\
\int_0^1 kx^3\left[\frac{1}{2}y^2\right]_x^1 dx &=\\
\frac{1}{2}k\int_0^1 x^3\left(1-x^2\right)dx &=\\
\frac{1}{2}k\int_0^1 \left(x^3-x^5\right)dx &=\\
\frac{1}{2}k\left(\frac{1}{4}x^4-\frac{1}{6}x^6\right)_0^1 &=\\
k &= 24
\end{aligned}
$$

Böylece yoğunluğun uygun bir olasılık yoğunluk fonksiyonu olabilmesi için integralin 1 olması ve bunun olabilmesi içinde k'nın 24'e eşit olması gerekir.

2. x'in marjinal PDF'si olan $f_X(x)$ nedir?
    - 2'nin Çözümü: Marjinal PDF'nin tanımı şöyledir:

$$f_X(x) = \int_Y f(x,y)dy$$

Kelimeler ile ifade edecek olursak, x'in marjinalini elde etmek için, yapmamız gereken tek şey bileşik yoğunluktaki diğer rasgele değişkenlere göre integral almaktır, yani bu durumda y'e göre integral almalıyız. Bu integrali halı hazırda bir önceki problemde aldık, bu nedenle sadece yukarıdaki çalışmadan elde edilen çözümü aşağıya yazacağım.

$$
\begin{aligned}
f_X(x) &= \int_Y f(x,y)dy\\
&= 12x^3\left(1-x^2\right)
\end{aligned}
$$

y'nin marjinal PDF'si için de çözebilirdik. Bu bizi farklı bir integral almaya zorlayacaktı. Ancak bu pratiği burada yapmayacağım

3. x=1/2'de, x'in marginal CDF değeri, $F_x(x)$, nedir?
   - 3'ün Çözümü: x'in marjinal CDF'si marjinal PDF'nin CDF'si olarak tanımlanır ya da

$$F_X(x) = \int_{-\infty}^{x} f_X(\tilde{x})d\tilde{x}.$$

   Bu nedenle, bu PDF'nin integralini alarak, ki halı hazırda 1'de bunu yaptık, şunu elde ederiz:

$$\begin{aligned} F_X(x) &= 12\left(\frac{1}{4}\tilde{x}^4 - \frac{1}{6}\tilde{x}^6\right)_0^x \\ &= 12\left(\frac{1}{4}x^4 - \frac{1}{6}x^6\right) \\ F_X(x) &= 6x^4\left(\frac{1}{2} - \frac{1}{3}x^2\right). \end{aligned}$$

Bunu 1/2 için çözdüğümüzde şunu elde ederiz:

$$F_X\left(\tfrac{1}{2}\right) = 6\left(\tfrac{1}{2}\right)^4\left(\tfrac{1}{2} - \tfrac{1}{3}\left(\tfrac{1}{2}\right)^2\right) = \tfrac{3}{16}\cdot\tfrac{5}{6} = \tfrac{5}{32}$$

4. y'nin koşullu PDF'si nedir (x'e koşullanmış, yani f(y|x))? X ile Y bağımsız mı? Açıklayınız.
   - 4'ün Çözümü: Yine koşullu dağılımın tanımını aşağıya yazmalıyız (bu sadece koşullu olasılığın tanımının basitçe bir uzantısıdır):

$$f(y|x) = \frac{f(x,y)}{f_X(x)}.$$

Açıkçası, elimizde bu denklemin bütün parçaları var. Aslında, sabitin integralini alarak koşullu dağılımı elde etmek zorunda olmayışımız yeterince ilginçtir. Bu birçok Bayezyen uygulama ile Metropolis- Hastings Markov Chain Monte Carlo gibi ileri istatistiki yöntemlere uygundur. Bunun hakkında daha çok bilgiyi daha sonra alacaksınız, fakat burada sadece gösterim amacıyla k'lı yoğunlukları kullanacağız:

$$\begin{aligned} f(y|x) &= \frac{f(x,y)}{f_X(x)} \\ &= \frac{kx^3y}{\frac{1}{2}kx^3(1-x^2)} \\ f(y|x) &= \frac{2y}{1-x^2}. \end{aligned}$$

Yinelemek gerekirse, ilgilenen okuyucular için, y'nin marjinali fY(y)'den f(x|y)'yi hesaplamak da mümkündür. Bu durumda, ya doğrudan koşullu dağılımı hesaplayacaktık ya da Bayes Kuralı kullanacaktık:

$$f(x|y) = \frac{f(y|x)f_X(x)}{f_Y(y)}$$

Sorunun ikinci bölümünü de cevaplandırmalıyız: X ve Y bağımsız mı? y'nin koşullu dağılımının x'in bir fonksiyonu olmasından ötürü ikisi de bağımsız değildir. İlk bakışta "faktör kuralı" na göre bileşik marjinallerin faktörünü alabileceğimiz gibi görünüyor, ancak unutmayalım ki bunu ancak bileşik dağılımın desteği dikdörtgen olduğu zaman yapabiliriz. Ancak, marjinalleri hesaplamak onun gerçekten önemli olduğunu anlamanın en kolay yoludur. Dikkatli bir şekilde dikdörtgen olmayan bir destek oluşturmak mümkün iken ve buna rağmen bileşik dağılımı oluşturan iki bağımsız rasgele değişkene sahip olabilecekken, bu durum bizim örneğimizde işe yaramıyor.

5. X + Y < 1'in olasılığı nedir?

- 5’in Çözümü: X + Y < 1 bölgesi üzerinden integral almak istiyoruz. Bunu birçok yoldan yapabiliriz, fakat ben sadece onu doğrudan hesaplayacağım. İntegralin sınırları konusunda çok dikkatli olmamız gerektiğini unutmamak gerekiyor. y’ye göre integral alındığında PDF’nin ilk tanımından gelen x < y < 1-x eşitsizliğini hatırlamak zorundasınız. İkinci eşitsizlik X + Y < 1 ⇒ Y < 1 – X koşulundan geliyor. İki eşitsizliğin birleştirilmesi, x < 1/2'nin aynı zamanda integral alınırken x üzerine koymak zorunda olduğumuz koşul olduğunu da hatırlatmaktadır. Şimdi integrali alabiliriz

$$
\begin{aligned}
P(X+Y<1) &= 24\int_0^{\frac{1}{2}}\int_x^{1-x} x^3 y\,dy\,dx \\
&= 24\int_0^{\frac{1}{2}} x^3\left[\frac{1}{2}y^2\right]_x^{1-x} dx \\
&= 12\int_0^{\frac{1}{2}} x^3\left(1-2x+x^2-x^2\right)dx \\
&= 12\int_0^{\frac{1}{2}} dx \\
&= 12\left[\frac{1}{4}x^4-\frac{2}{5}x^5\right]_0^{\frac{1}{2}} \\
&= \frac{12}{2^4}\left[\frac{1}{4}-\frac{1}{5}\right] \\
&= \frac{3}{80}
\end{aligned}
$$

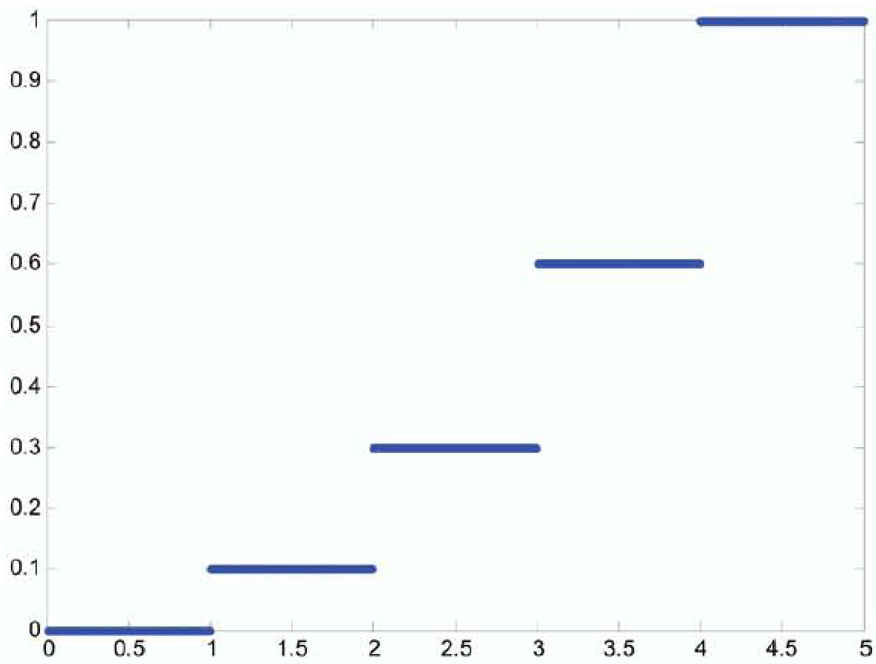
Ne mutlu ki bu 1’den küçüktür. :) Aynı zamanda, yoğunluk şeklinin (1,1) etrafına daha fazla yığılma yapabileceğini de ilaveten kontrol etmek gerekiyor. Biz sadece üçgen bölgesinin 0 < x < y < 1 ile sınırlandırılmış daha küçük bir bölümünün integralini aldık.

Soru Altı

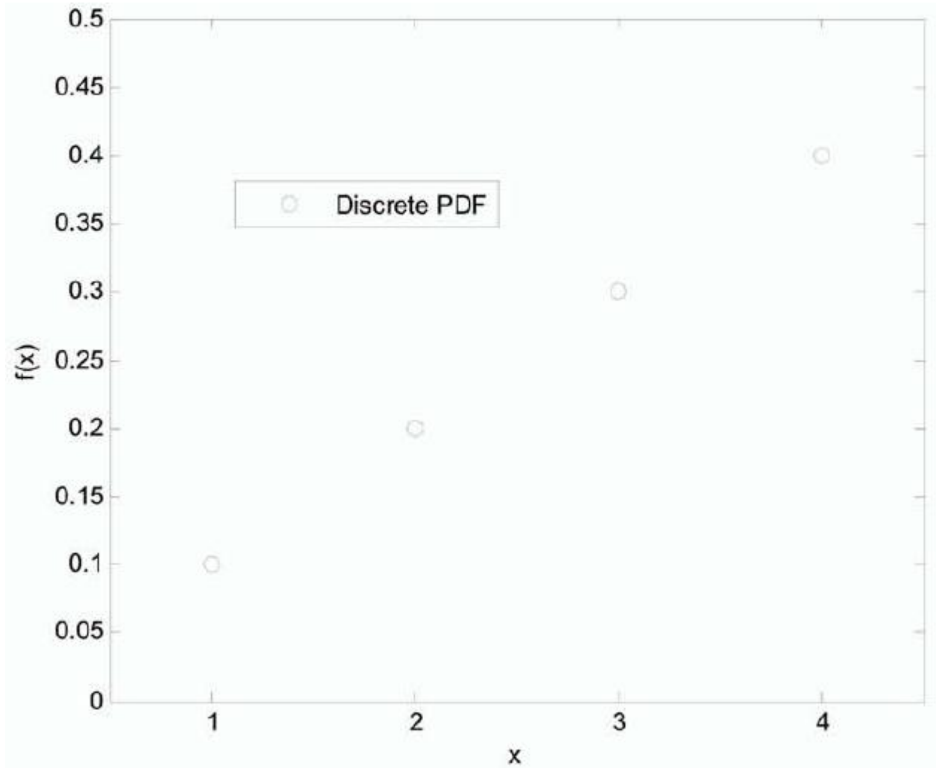
(Bain/Engelhardt, Ch. 2, ex. 10)

X bir kesikli rasgele değişken ve P[X = x] = 0 olsun. Varsayalım ki x = 1, 2, 3, 4 değerlerinde CDF *F(x) = 0.05x(1+x)*'dir.

1. CDF'nin grafiğini çiziniz.
   - 1'in Çözümü: Bu kesikli CDF'yi Matlab gibi bir programda floor $\lfloor x \rfloor$ koduyla çizdirebiliriz: F(x) = 0.05 $\lfloor x \rfloor$(1 + $\lfloor x \rfloor$). [0, 5] aralığında kesikli CDF şöyle olur:



2. Kesikli pdf, f(x)'in grafiğini çiziniz
   - 2'nin Çözümü: PDF'yi de çizmek kolaydır. Sadece 1, 2, 3, ve 4 yığılma noktalarındaki farkı kullanıp, bunun kesikli PDF olduğunu bilmemiz yeterlidir:



3. E[X]'in tanımını yazınız ve E[X]'i bulunuz.
   - 3'ün Çözümü: Her ne kadar kesikli ve sürekli rasgele değişkenler limit argümanı bakımından ilişkili olsa da, $\mathbb{E}$[X]'in tanımı sürekli ve kesikli rasgele değişkene göre farklılık gösterir. Bu problemde sadece kesikli olanı kullanacağımız halde, bütünlük açısından aşağıya ikisini de yazacağım.

$$\text{Sürekli}: \; E[X] = \int_{-\infty}^{\infty} x f(x) dx$$
$$\text{Kesikli}: \; E[X] = \sum_{x \in X} x f(x)$$

burada f(x) rasgele değişken X'in PDF'sidir. Bu problem de aşağıdaki gibi basit bir toplama vardır:

$$\begin{aligned} E[X] &= \sum_{x \in X} x f(x) \\ &= 1 \cdot .1 + 2 \cdot .2 + 3 \cdot .3 + 4 \cdot .4 \\ E[X] &= 3 \end{aligned}$$