Olasılık Teorisi ve İstatistik
Final Sınavı - Çözümler

İstanbul Üniversitesi - Cerrahpaşa
Bilgisayar Mühendisliği Bölümü - Güz 2019

24.Aralık.2019

S1: Bir SÜREKLİ RASGELE DEĞİŞKEN, $\boldsymbol{X}$'in "*Olasılık Yoğunluk Fonksiyonu*" aşağıdaki gibidir.

$$f(x) = \alpha\ |x|^4\ \mathrm{e}^{-|x|}\ ,\quad -\infty < x < \infty$$

Aşağıdaki soruları cevaplayınız.

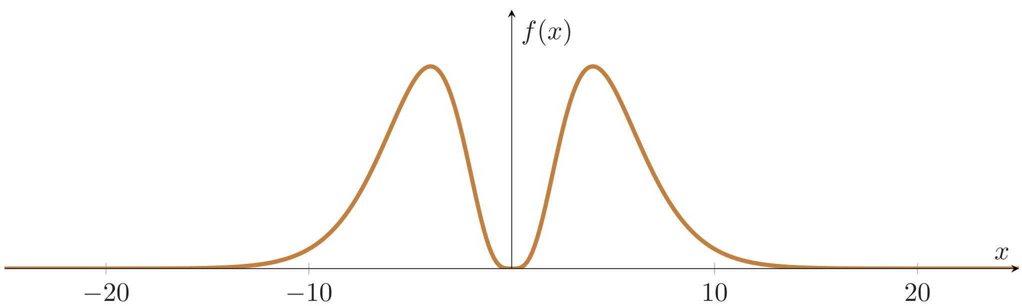
(a) `(10 puan)` $\alpha = \frac{1}{48}$ olması gerektiğini gösteriniz.

Çözüm 1a:

OYF şu şekilde de yazılabilir:

$$f(x) = \begin{cases} \alpha\ x^4\ \mathrm{e}^{x} & , \quad x \le 0 \\ \alpha\ x^4\ \mathrm{e}^{-x} & , \quad x > 0 \end{cases}$$

X'in olasılık yoğunluk fonksiyonu



$$\int_{-\infty}^{\infty} f(x)\ \mathrm{d}x = 1$$

olması gerekiyor. Ayrıca bu OYF için:

$$\int_{-\infty}^{0} f(x)\ \mathrm{d}x = \int_{0}^{\infty} f(x)\ \mathrm{d}x$$

olduğu görülebilir. Burdan da:

$$\begin{aligned}
1 &= \int_{-\infty}^{0} f(x)\ \mathrm{d}x + \int_{0}^{\infty} f(x)\ \mathrm{d}x \\
1 &= 2\int_{0}^{\infty} f(x)\ \mathrm{d}x \\
1 &= 2\alpha \int_{0}^{\infty} x^4\ \mathrm{e}^{-x}\ \mathrm{d}x \\
\frac{1}{2} &= \alpha\ \Gamma(5) \\
\alpha &= \frac{1}{2\cdot 4!} \\
\alpha &= \frac{1}{48} \qquad \blacksquare
\end{aligned}$$

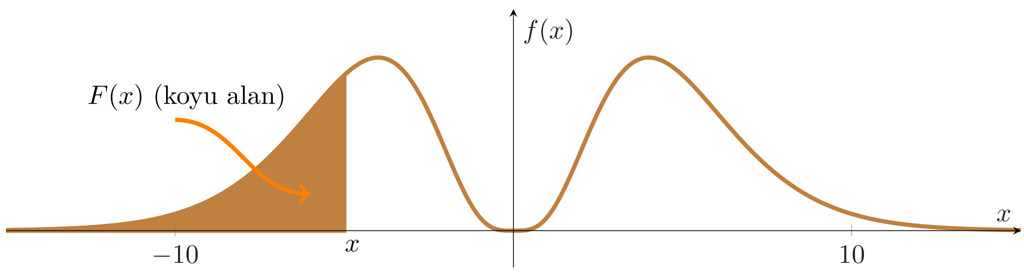
(b) `(10 puan)` X'in "*Birikimli Dağılım Fonksiyonu*"nu bulunuz.

Çözüm **1b** **:**

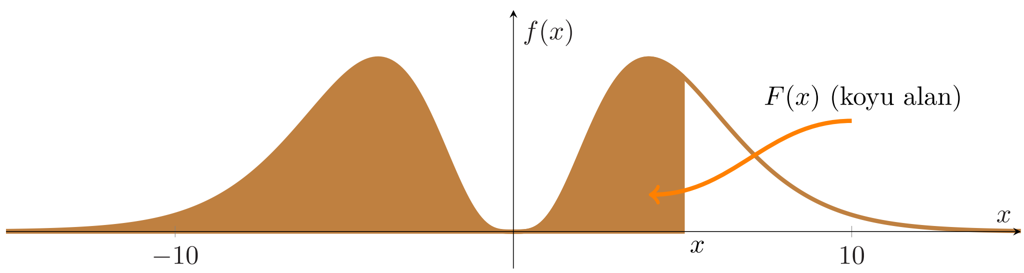
X'in Birikimli Dağılım fonksiyonu aşağıdaki şekilde veriliyor.

$$F(x) = \int_{\infty}^{x} f(\tau)\ \mathrm{d}\tau$$

OYF'nin formülü $x < 0$ ve $x > 0$ için farklı, ona dikkat etmek lazım. BDF, x'e kadar OYF eğrisinin altında kalan alan olarak tanımlanıyor. $x < 0$ için bu alan şekildeki gibidir:



$x > 0$ için:



Önce $x < 0$ için $F(x)$'i bulalım.

$$F(x) = \int_{-\infty}^{x} \alpha\ \tau^4\ \mathrm{e}^{\tau}\ \mathrm{d}\tau$$

Bunu yapmak için parçalı integral metodunu kullanacağız. Fakat öncesinde şöyle yapalım. $L_n(x)$ diye bir fonksiyon tanımlayalım, öyle ki:

$$L_n(x) = \int_{-\infty}^{x} \tau^n \ \mathrm{e}^{\tau} \ \mathrm{d}\tau$$

olsun. Bunu çözmeye çalışalım:

$$\begin{aligned} u &= \tau^n \\ \mathrm{d}v &= \mathrm{e}^{\tau} \ \mathrm{d}\tau \\ \mathrm{d}u &= n \ \tau^{n-1} \\ v &= \mathrm{e}^{\tau} \end{aligned}$$

yazarsak ve

$$\int u \ \mathrm{d}v = uv - \int v \ \mathrm{d}u$$

kullanarak

$$\begin{aligned} \int_{-\infty}^{x} \tau^n \ \mathrm{e}^{\tau} \ \mathrm{d}\tau &= \tau^n \ \mathrm{e}^{\tau}\Big|_{-\infty}^{x} - n \int_{-\infty}^{x} \tau^{n-1} \ \mathrm{e}^{\tau} \ \mathrm{d}\tau \\ L_n(x) &= x^n \ \mathrm{e}^x - n \ L_{n-1}(x) \end{aligned}$$

Burdan

$$\begin{aligned} L_0(x) &= \mathrm{e}^x \\ L_1(x) &= x \ \mathrm{e}^x - \mathrm{e}^x \\ L_2(x) &= x^2 \ \mathrm{e}^x - 2(x \ \mathrm{e}^x - \mathrm{e}^x) = x^2 \ \mathrm{e}^x - 2 \ x \ \mathrm{e}^x + 2 \ \mathrm{e}^x \\ L_3(x) &= x^3 \ \mathrm{e}^x - 3(x^2 \ \mathrm{e}^x - 2 \ x \ \mathrm{e}^x + 2 \ \mathrm{e}^x) \\ &= x^3 \ \mathrm{e}^x - 3x^2 \ \mathrm{e}^x + 6 \ x \ \mathrm{e}^x - 6 \ \mathrm{e}^x \\ L_4(x) &= x^4 \ \mathrm{e}^x - 4(x^3 \ \mathrm{e}^x - 3x^2 \ \mathrm{e}^x + 6 \ x \ \mathrm{e}^x - 6 \ \mathrm{e}^x) \\ &= x^4 \ \mathrm{e}^x - 4x^3 \ \mathrm{e}^x + 12x^2 \ \mathrm{e}^x - 24 \ x \ \mathrm{e}^x + 24 \ \mathrm{e}^x \end{aligned}$$

Burdan

$$\begin{aligned} F(x) &= \frac{1}{48} \ L_4(x) \\ &= \frac{1}{48} \ (x^4 \ \mathrm{e}^x - 4x^3 \ \mathrm{e}^x + 12x^2 \ \mathrm{e}^x - 24 \ x \ \mathrm{e}^x + 24 \ \mathrm{e}^x) \end{aligned}$$

Bulunur.

$x > 0$ için ise şunu dikkate alalım. OYF y eksenine göre simetrik olduğu için, yani $f(x) = f(-x)$ olduğu için, $F(x) = 1 - F(-x)$ olur, (**-'lere dikkat edelim...**). O zaman

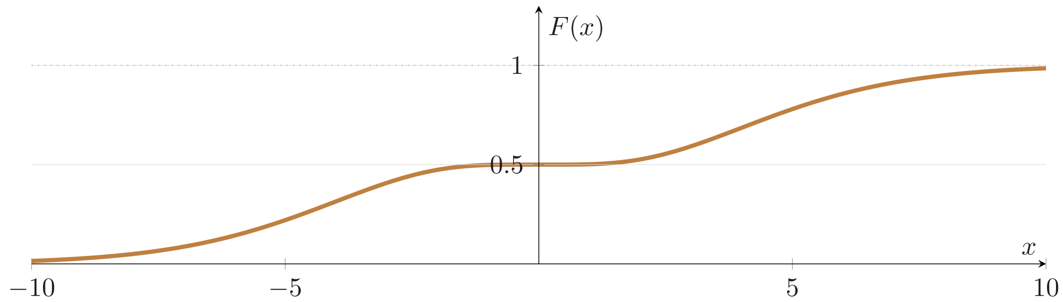
$$F(x) = 1 - \left[\frac{1}{48} \ (x^4 \ \mathrm{e}^{-x} + 4x^3 \ \mathrm{e}^{-x} + 12x^2 \ \mathrm{e}^{-x} + 24 \ x \ \mathrm{e}^{-x} + 24 \ \mathrm{e}^{-x})\right] \qquad , \ x > 0 \quad \text{için}$$

bulunur. Böylece son olarak:

$$F(x) = \begin{cases} \frac{1}{48}\left(x^4\ e^{x} - 4x^3\ e^{x} + 12x^2\ e^{x} - 24\ x\ e^{x} + 24\ e^{x}\right) & , \quad x \le 0 \\ 1 - \left[\frac{1}{48}\left(x^4\ e^{-x} + 4x^3\ e^{-x} + 12x^2\ e^{-x} + 24\ x\ e^{-x} + 24\ e^{-x}\right)\right] & , \quad x > 0 \end{cases}$$

bulunur. Bunun (-10,10) aralığındaki grafiğine bakarsak:

X'in birikimli dağılım fonksiyonu



olduğu görülür.

(c) `(10 puan)` X'in bir fonksiyonu $Y = g(X) = e^{-2|X|}$ verilmiş ise Y'nin beklenen değerini bulunuz.

Çözüm 1c:

$$\begin{aligned} E[g(X)] &= \int_{-\infty}^{\infty} g(x) f(x)\ dx \\ &= \int_{-\infty}^{\infty} \alpha |x|^4\ e^{-3|x|}\ dx \\ &= 2\int_{0}^{\infty} \alpha x^4\ e^{-3x}\ dx \quad \text{(simetriden faydalanarak)} \end{aligned}$$

$$y = 3x \ \text{ dersek } \ dx = \frac{dy}{3} \ \text{ olur.}$$

$$\begin{aligned} E[g(X)] &= 2\int_{0}^{\infty} \alpha\ (y/3)^4\ e^{-y}\ (1/3)\ dx \\ &= 2\ \alpha\ \frac{1}{3^5}\ \Gamma(5) \\ &= 2\ \frac{1}{3^5 \cdot 48}\ 4! \\ E[g(X)] &= \frac{1}{243} \qquad \blacksquare \end{aligned}$$

S2: X ve Y **AYRIK** rastgele değişkenleri için "*Birleşik Olasılık Kütle Fonksiyonu*" aşağıdaki gibidir. Verilen soruları cevaplayınız.

$$p_{XY}(x,y) = \begin{cases} \alpha\ (x+y) & , \quad 1 \le x \le y \le 3 \qquad (x, y \text{ tam sayı}) \\ 0 & , \quad \text{diğer} \end{cases}$$

(a) `(10 puan)` $\alpha = \frac{1}{24}$ olması gerektiğini gösteriniz.

Çözüm (2a):

OKF'nin aşağıdaki şartı sağlaması gerekir:

$$1 = \sum_{x=1}^{3} \sum_{y=x}^{3} \mathrm{p}_{XY}(x, y)$$
$$1 = \alpha \sum_{x=1}^{3} \sum_{y=x}^{3} (x + y)$$

Bunu tablo ile yapmakta fayda var. Değer uzaylarını gözönüne alarak aşağıdaki tabloya olasılıkları yazalım:

	x		
y	**1**	**2**	**3**
1	2α	0	0
2	3α	4α	0
3	4α	5α	6α

Toplamın 1 olması lazım. Burdan

$$24\alpha = 1$$
$$\alpha = \frac{1}{24}$$

(b) `(10 puan)` $Y = 2$ olma şartı altında X'in şartlı olasılık kütle fonksiyonunu bulunuz. (formül olarak bulmak zorunda değilsiniz, zaten değer uzayı çok küçük, bütün noktalardaki ihtimalleri listeleyebilirsiniz)

Çözüm (2b):

$Y = y$ olma şartı altında X'in OKF'si:

$$\mathrm{p}_{\mathrm{X}}(x|y) = \frac{\mathrm{p}_{\mathrm{XY}}(x, y)}{\mathrm{p}_{\mathrm{Y}(y)}}$$

Y'nin marjinal dağılımını bir önceki tablodan bulabiliriz. AŞağıya tekrar yazıp, sutünları ve satırları toplayarak marjinal dağılımları bulabiliriz.

	x			
y	**1**	**2**	**3**	$f_Y(y)$
1	2α	0	0	2α
2	3α	4α	0	7α
3	4α	5α	6α	15α
$f_X(x)$	9α	9α	6α	

$Y = 2$ olma şartı altında X, sadece 1 ve 2 değerlerini alabilir. Burdan:

$$p_X(1|2) = P(X = 1|Y = 2) = \frac{p_{XY}(1,2)}{p_{Y(2)}} = \frac{3\alpha}{7\alpha} = \frac{3}{7}$$

$$p_X(2|2) = P(X = 2|Y = 2) = \frac{p_{XY}(2,2)}{p_{Y(2)}} = \frac{4\alpha}{7\alpha} = \frac{4}{7}$$

bulunur.

(c) `(10 puan)` X ve Y'nin korelasyon katsayısını bulunuz.

Çözüm 2c:

Korelasyon katsayısı ρ_{XY} önceki sorudaki tabloyu da kullanarak aşağıdaki adımlardan bulunabilir.

$$\begin{aligned}
\rho_{XY} &= \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} \\
\sigma_X^2 &= E(X^2) - \mu_X^2 \\
\mu_X &= 1 \cdot 9\alpha + 2 \cdot 9\alpha + 3 \cdot 6\alpha = 45\alpha = 1.875 \\
E(X^2) &= 1^2 \cdot 9\alpha + 2^2 \cdot 9\alpha + 3^2 \cdot 6\alpha = 99\alpha = 4.125 \\
\sigma_X^2 &= 0.6049 \\
\sigma_X &= 0.7806 \\
\mu_Y &= 1 \cdot 2\alpha + 2 \cdot 7\alpha + 3 \cdot 15\alpha = 61\alpha = 2.5417 \\
E(Y^2) &= 1^2 \cdot 2\alpha + 2^2 \cdot 7\alpha + 3^2 \cdot 15\alpha = 165\alpha = 6.875 \\
\sigma_Y &= \sqrt{E(Y^2) - \mu_Y^2} = 0.6442 \\
E(XY) &= \sum_{x=1}^{3} \sum_{y=x}^{3} x \cdot y \cdot p_{XY}(x,y) \\
&= (1 \cdot 1)(2\alpha) + (1 \cdot 2)(3\alpha) + (1 \cdot 3)(4\alpha) \\
&\quad + (2 \cdot 2)(4\alpha) + (2 \cdot 3)(5\alpha) + (3 \cdot 3)(6\alpha) = 120\alpha = 5 \\
\sigma_{XY} &= E(XY) - \mu_X \, \mu_Y \\
&= 5 - 1.875 \cdot 2.5417 = 0.2343 \\
\rho_{XY} &= \frac{0.2343}{0.7808 \times 0.6442} \\
\rho_{XY} &= 0.4658 \quad \blacksquare
\end{aligned}$$

S3: Bir sunucu odasında kullanılacak hard disklerdeki bozuk sektör sayısı zamana bağlı olarak Poisson sürecini takip etmekte ve **1 yıl** içinde bir hard diskte ortalamada 0.5 bozuk sektör görülmektedir. Aşağıdaki soruları cevaplayınız.

(a) `(10 puan)` Bir hard diskte 4üncü bozuk sektörün görülmesi ile 5inci bozuk sektörün görülmesi arasında geçen sürenin 3 aydan küçük olma ihtimali nedir?

Çözüm 3a:

Poisson süreçleri hafızasızdır yani 4üncü hatadan sonra 5inci hatanın görülmesi arasında geçen süre baştan itibaren ilk hata görülünceye kadar geçen süre ile aynı dağılıma sahiptir. Hata görülünceye kadar geçen süre de üstel dağılımla modellenebilir. T, bu süreçte hata görülünceye kadar geçen süre ise, T,parametresi $\lambda = 0.5$ hata/yıl olan bir üstel dağılmlı rastgele değişkendir. OYFsi:

$$f(t) = 0.5\ \mathrm{e}^{-0.5t}$$

olur. Bize sorulan T'nin $\frac{3}{12} = 0.25$'den küçük olma ihtimalidir (Bir yılda 12 ay var)

$$\begin{aligned} P(T < 0.25) &= \int_0^{0.25} 0.5\ \mathrm{e}^{-0.5t}\ \mathrm{d}t \\ &= -\mathrm{e}^{-0.5t}\Big|_0^{0.25} \\ &= 0.1175 \quad \blacksquare \end{aligned}$$

(b) `(10 puan)` Bu sunucu odasına aynı hard diskten 250 adet alındığını varsayalım. 2 yıl sonunda üzerinde üzerindeki bozuk sektör sayısı en fazla 1 olan disk sayısının 175'den fazla olma ihtimali yaklaşık olarak nedir?

Çözüm 3b:

X: Bir hard diskte 2 yıl sonunda görülen hata sayısı dersek, X bir Poisson rastgele değişkeni olur. Parametresi $\alpha = 2$ yıl $\times$ 0.5 hata/yıl $= 1$ hata olur. Olasılık kütle fonksiyonu: $\mathrm{p}(x) = \dfrac{\mathrm{e}^{-1}1^x}{x!} = \dfrac{\mathrm{e}^{-1}}{x!}$ olur. 2 yıl sonunda hata sayısının 1 veya daha az olma ihtimaline p diyelim:

$$\begin{aligned} p = P(X \leq 1) &= P(X = 0) + P(X = 1) \\ &= \mathrm{e}^{-1}(1 + 1) \\ &= 0.7358 \end{aligned}$$

Y: 250 hard disk içinden 2 yılın sonunda hata sayısı 1 veya daha az olan disk sayısı ise, Y binom dağılımlıdır ve parametreleri $p = 0.7358$ ve $n = 250$'dir. Bize sorulan $P(Y > 175)$. Bunu merkezi limit teoreminden normal dağılıma yaklaşımla çözebiliriz çünkü $np = 183.9397$ ve $n(1 - p) = 66.05$ değerleri büyük. Ayrıca $P(Y > 175) = P(Y > 175.5)$ diyerek süreklilik düzeltmesi de uygulayabiliriz. Y'nin ortalaması $\mu_Y = np = 183.9397$ ve standart sapması $\sigma_Y = \sqrt{np(1-p)} = 6.9713$ Öyle ki:

$$\begin{aligned} P(Y > 175.5) &= P\left(\frac{Y - \mu_Y}{\sigma_Y} > \frac{175.5 - 183.9397}{6.9713}\right) \\ &= P(Z > -1.2106) \\ &= 1 - \phi(-1.2106) \\ &= 1 - 0.1131 \\ P(Y > 175.5) &= 0.8869 \quad \blacksquare \end{aligned}$$

S4: Bir matbaada basılan her davetiye A ve B kodlu 2 makinadan biri ile yapılmaktadır. A makinasından çıkan bir davetiyenin yırtık çıkma ihtimali `0.15` ve B makinasından çıkan bir davetiyenin yırtık çıkma ihtimali `0.45`'tir. Herhangi bi davetiyenin A makinası ile yapılmış olma ihtimali `p`, B makinası ile yapılmış olma ihtimali `q`'dur.

(a) `(10 puan)` Bu matbaada basılan herhangi bir davetiyenin yırtık çıkma ihtimalinin `0.39` olması için `p` ve `q` ne olmalıdır?

Çözüm 4a **:**

Olaylarımızı tanımlayalım:

A: A makinası ile yapılma olayı, $P(A) = p$

B: B makinası ile yapılma olayı, $P(B) = q$.

A ve B olayları birbirini tamamlayan ve birbirini dışlayan olaylar, o yüzden $q = 1 - p$

Y: Davetiyenin yırtık çıkma olayı, $P(Y|A) = 0.15$, $P(Y|B) = 0.45$. $P(Y) = 0.39$ verilmiş.

Toplu olasılık kuralından:

$$\begin{aligned} P(Y) &= P(Y|A)P(A) + P(Y|B)P(B) \\ 0.39 &= 0.15p + 0.45(1-p) \\ 0.30p &= 0.06 \\ p &= 0.2 \\ q &= 0.8 \qquad \blacksquare \end{aligned}$$

(b) `(10 puan) p = 0.15` ise rastgele seçilen yırtık bi davetiyeyi B makinasının üretmiş olma ihtimali nedir?

Çözüm 4b **:**

$$\begin{aligned} P(Y) &= P(Y|A)P(A) + P(Y|B)P(B) \\ P(Y) &= 0.15 \times 0.15 + 0.45 \times 0.85 \\ P(Y) &= 0.4050 \\ P(B|Y) &= \frac{P(Y|B)P(B)}{P(Y)} \\ P(B|Y) &= 0.9444 \qquad \blacksquare \end{aligned}$$