MIT OpenCourseWare
http://ocw.mit.edu

14.30 Ekonomide İstatistiksel Yöntemlere Giriş
Bahar 2009

Bu materyale atıfta bulunmak ve kullanım koşulları için http://ocw.mit.edu/terms sayfasını ziyaret ediniz.

Problem Seti 5 - Çözümler
14.30 Ekonomide İstatistiksel Yöntemlere Giriş

Konrad Menzel

Son Gün: 31 Mart 2009

Soru Bir

Bağımsız rasgele değişkenlerin toplamı veya ortalaması ile ilgilendiğimizde Büklüm teoremi çok yararlı bir hiledir. Son problem setinde, aşağıdaki X rasgele değişkeni ile ilgilendik.

$$f(x) = \begin{cases} e^{-x} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}.$$

Şimdi, varsayalım ki $X = X_1 = X_2 = \ldots = X_k$ bağımsız ve aynı dağılımlı (i.i.d.) rasgele değişkenlerdir.

1. Büklüm formülünü kullanarak, $Y_2 = \frac{1}{2}(X_1 + X_2)$ için PDF tanımlayınız. (İpucu: $Z_1 = X_1$ ile $Z_2 = X_1 + X_2$'ü tanımla ve sonra dönüştürme yöntemini kullanarak Z_2'den Y_2'yi elde ediniz.)

   - 1'in Çözümü: Büklüm formülünü kullanmak için, X_1 ve X_2'nin bileşik PDF'si ile y_1 ve x_1'in fonksiyonu olarak x_2'ye ihtiyaç vardır. Fonksiyon şöyledir: $x_2 = 2y_2 - x_1$. Aynı zamanda, bağımsız oldukları için, sadece iki marjinali, $f_{x1}(x_1)$ ile $f_{x2}(x_2)$, birbiriyle çarparak bileşik PDF'yi oluşturabiliriz: Bu bize bileşik PDF'yi verir, yani $x_1 > 0$ ve $x_2 > 0$ için $f(x_1, x_2) = e^{-(x1+x2)}$'dir, diğer bütün durumlarda ise sıfırdır. Bu probleme uyarlanmış büklüm fonksiyonu (limitleri hesaba katarak) (burada Jacobian'ı kullanarak y_1 ile x_1'in fonksiyonu olarak x_2 için değişken değiştirme formülünü hesaba katarız) şöyledir:

$$
\begin{aligned}
f_{Y_2}(y_2) &= \int_0^{y_2} f(x_1, 2y_2 - x_1)\,|J|\,dx_1 \\
&= \int_0^{y_2} e^{-(x_1+2y_2-x_1)}\,|2|\,dx_1 \\
&= 2\int_0^{y_2} e^{-2y_2}dx_1 \\
&= 2\left(x_1 e^{-2y_2}\right)_{x_1=0}^{x_1=2y_2} \\
f_{Y_2}(y_2) &= 4y_2 e^{-2y_2}.
\end{aligned}
$$

2. Beklenen değer $\mathbb{E}[Y_2]$'yi hesaplayınız.
    - 2'nin Çözümü: Beklenen değeri hesaplamak için, sürekli rasgele değişkenin formülünü kullanırız:

$$
\mathbb{E}[Y_2] = \int_{-\infty}^{\infty} y_2 f_{Y_2}(y_2)dy_2
$$

İki yinelenen parçalı integrali kullanarak aşağıdaki integrali hesaplamak zorundayız:

$$
\begin{aligned}
\mathbb{E}[Y_2] &= \int_0^{\infty} y_2\left(4y_2 e^{-2y_2}\right)dy_2 \\
&= \int_0^{\infty} (y_2)^2\, 4e^{-2y_2}dy_2 \\
&= \left[(y_2)^2\left(-2e^{-2y_2}\right)\right]_0^{\infty} - \int_0^{\infty}(2y_2)\left(-2e^{-2y_2}\right)dy_2 \\
&= \left[(y_2)^2\left(-2e^{-2y_2}\right)\right]_0^{\infty} - \left[(2y_2)\left(-e^{-2y_2}\right)\right]_0^{\infty} + \int_0^{\infty} 2\left(-e^{-2y_2}\right)dy_2 \\
&= \left[(y_2)^2\left(-2e^{-2y_2}\right)\right]_0^{\infty} - \left[(2y_2)\left(-e^{-2y_2}\right)\right]_0^{\infty} + \left[e^{-2y_2}\right]_0^{\infty} \\
&= (0-0) + (0-0) + (0+1) \\
\mathbb{E}[Y_2] &= 1
\end{aligned}
$$

Ayrı bir seçenek olarak, beklenen değerin özeliklerini de kullanabiliriz (büklüm zorunlu değil):

$$\begin{aligned}
\mathbb{E}[Y_2] &= \mathbb{E}[\frac{1}{2}(X_1 + X_2)] \\
&= \frac{1}{2}\mathbb{E}[X_1 + X_2] \\
&= \frac{1}{2}\left(\mathbb{E}[X_1] + \mathbb{E}[X_2]\right) \\
&= \frac{1}{2}\left(2\mathbb{E}[X]\right) \\
&= \mathbb{E}[X] \\
&= \int_{-\infty}^{\infty} x f_X(x) dx \\
&= \int_{0}^{\infty} x e^{-x} dx \\
&= (xe^{-x})_0^{\infty} - \int_0^{\infty} -e^{-x} dx \\
&= 0 + (0 + 1) \\
\mathbb{E}[Y_2] &= 1
\end{aligned}$$

3. Büklüm formülünü kullanarak, $Y_3 = \frac{1}{3}$ (X_1 + X_2 + X_3) için PDF tanımlayınız. (İpucu: Bölüm 1'deki ipucunu kullanarak Z_3 = X_1 + X_2 + X_3'ü tanımlayınız, X_3 ve Z_2 ile büklüm yaparak problemi Z_2 ve Z_3'e dönüştürünüz.)
   - 3'ün Çözümü: $F_{Y2}(y_2)$ elde ettiğimiz yerdeki 1'in sonuçlarına aynı yöntemi uygulayabiliriz. X_1, X_2 ve X_3'ün bileşik PDF'sini şöyle yazarız:

$$\begin{aligned}
f_{X_1,X_2,X_3}(x_1, x_2, x_3) &= f_{X_1}(x_1) f_{X_2}(x_2) f_{X_3}(x_3) \\
&= e^{-(x_1+x_2+x_3)}
\end{aligned}$$

Daha sonra 5'i çözmemize yardımcı olması için, bu problem için birazcık farklı bir yaklaşım sergileyeceğiz. Z_2 = X_1 + X_2 ile Z_3 = X_1 + X_2 + X_3'ü tanımlayınız. Şimdi herhangi bir Jacobian dönüşümü ihmal edebiliriz ve doğrudan büklüm formülünü uygulayabiliriz. ilk önce, X_1 ve X_2'nin bileşik dağılımına sadece büklüm formülünü uygulayarak Z_2'nin dağılımı için çözebiliriz.

$$\begin{aligned}
f_{Z_2}(z_2) &= \int_0^{z_2} f(x_1, z_2 - x_1)dx_1 \\
&= \int_0^{y_2} e^{-(x_1+z_2-x_1)}dx_1 \\
&= \int_0^{z_2} e^{-z_2}dx_1 \\
&= \left(x_1 e^{-z_2}\right)_{x_1=0}^{x_1=z_2} \\
f_{Z_2}(z_2) &= z_2 e^{-z_2}.
\end{aligned}$$

Şimdi, Büklümü Z_2 ile X_3'ün bileşik dağılımına uygulayarak Z_3'ün dağılımını elde edebiliriz:

$$\begin{aligned}
f_{Z_3}(z_3) &= \int_{-\infty}^{\infty} f_{X_3,Z_2}(x_3, z_2)dz_2 \\
&= \int_0^{z_3} f_{X_3}(z_3 - z_2) f_{Z_2}(z_2)dz_2 \\
&= \int_0^{z_3} z_2 e^{-z_3}dz_2 \\
&= \int_0^{z_3} z_2 e^{-z_3}dz_2 \\
f_{Z_3}(z_3) &= \frac{1}{2}(z_3)^2 e^{-z_3}
\end{aligned}$$

Ancak biz Y_3 ile ilgileniyoruz, Z_3 ile değil. Bu nedenle, $Z_3 = 3Y_3$'ü kullanarak, rasgele değişkenin basit bir sürekli ters dönüşümünü uygulayabiliriz:

$$\begin{aligned}
f_{Y_3}(y_3) &= f_{Z_3}(3y_3)|3| \\
&= \frac{3}{2}(3y_3)^2 e^{-3y_3} \\
f_{Y_3}(y_3) &= \frac{3}{2}(3y_3)^2 e^{-3y_3}
\end{aligned}$$

4. Beklenen değer $\mathbb{E}[Y_3]$'ü hesaplayınız.
   - 4'ün Çözümü: Beklenen değerin özeliklerini çok kolay bir şekilde uygulayarak tekrar $\mathbb{E}[Y_3]$ = 1'i elde edebiliriz. Ancak, bunun büklüm formülünü kullanarak ta yapabiliriz. 2'den öğrenmemiz gereken şey, herhangi bir k ∈ N ve y = ∞ ile y = 0 için $y^k e^{-y} = 0$ olduğundan parçalı integralin önemli olmadığıdır. Bu nedenle sadece parçalı integralin son kısmındaki e^{-3y_3}olan terim ile ilgileniyoruz (isterseniz doğrulayabilirsiniz):

$$\begin{aligned}
\mathbb{E}[Y_3] &= \int_0^\infty y_3 \left(\frac{3}{2} (3y_3)^2 e^{-3y_3} \right) dy_3 \\
&= \int_0^\infty y_3 \left(\frac{3}{2} (3y_3)^2 e^{-3y_3} \right) dy_3 \\
&= \left[e^{-3y_3} \right]_0^\infty \\
&= (0+1) \\
\mathbb{E}[Y_3] &= 1
\end{aligned}$$

5. Büklüm formülünü kullanarak, $Y_k = \frac{1}{k}(X_1 + X_2 + \ldots + X_k) = \frac{1}{k}\sum_{i=1}^{k} X_k$ için PDF tanımlayınız.(İpucu: bölüm 1 ve bölüm 2'deki ipuçlarının yöntemlerini kullanarak bir örnek süreç belirleyiniz.)
   - 5'in Çözümü: 3'te tanımlanan diziyi kullanarak büklüm formülünü kullanabiliriz. Daha fazla kısmi toplam tanımlarız: $Z_1 = X_1$, $Z_2 = X_1 + X_2$, $Z_3 = X_1 + X_2 + X_3$, …, $Z_k = \sum_{i=1}^{k} X_i$. Bu bize ardışık Z'lerin farkı olan Z'nin fonksiyonu olan X için bir ters fonksiyon oluşturmamıza olanak verir. Bu bize şu Jacobian'ı verir:

$$J = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

   Bunun determinantı 1'dir. Böylece, Jacobian'ı görmezlikten gelebiliriz. Şimdi bileşik dağılımı yazabiliriz:

$$
\begin{aligned}
f(z_1, z_2 - z_1, \ldots, z_k - z_{k-1}) &= f(x_1(z_1, ..., z_k), \ldots, x_k(z_1, ..., z_k)) \\
&= e^{-z_k}.
\end{aligned}
$$

x'lerin bütün bileşik dağılımları z_k'nın bir fonksiyonu olduğu için, k kadar x bağımsız x'in toplamlarının dağılımı için z_k'nın yeterli olduğunu söyleyebiliriz. Şimdi, yapmamız gereken tek şey z_k'nin marjinalini bulmak için z_1 den z_{k-1}'e kadar integralleri almaktır. 3'ten bildiğimiz kadarıyla, büklüm aşağıdaki sonuçları doğurur:

$$
f_{Z_2}(z_2) = z_2 e^{-z_2}
$$

ve

$$
f_{Z_3}(z_3) = \frac{1}{2}(z_3)^2 e^{-z_3}
$$

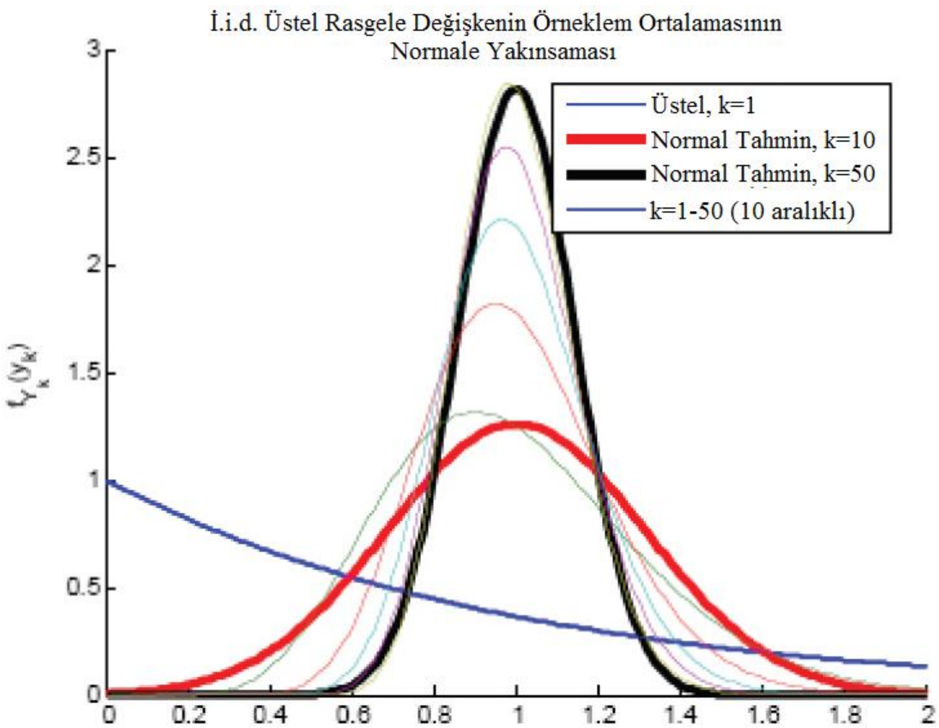
Bunu tekrarlarsak Z_k için genel bir formül elde ederiz:

$$
f_{Z_k}(z_k) = \left(\prod_{i=1}^{k-1} \frac{1}{i}\right)(z_k)^{k-1} e^{-z_k}
$$

$Z_k = kY_k$ için ters dönüşüm formülünü uygulayarak:

$$
f_{Y_k}(y_k) = \left(\prod_{i=1}^{k-1} \frac{1}{i}\right)(y_k)^{k-1} k^k e^{-k y_k}
$$

Bu aradığımız sonuçtur. İlginç olan ise, k sonsuza giderken bunun Normal'e veya Gaussian'a yakınsamasıdır. Bu daha genel bir teoremin özel bir durumudur: Merkezi Limit Teoremi. Aşağıda üstel dağılımdan Normal dağılıma kadar dağılımlardan k kadar rasgele çekilişin ortalamasının yakınsamasını görebilirsiniz.



6. Beklenen değer $\mathbb{E}[Y_k]$'yı hesaplayınız.
   - 6'nın Çözümü: 4'e bakınız. Beklenen değer 1'dir. :)

7. Bu bize k büyüklüğündeki bir örneklemin ortalaması hakkında ne söylüyor? Bu özellik üstel dağılıma mı özgü? Açıklayınız.
   - 7'nin Çözümü: k büyüklüğündeki bir i.i.d. (bağımsız ve aynı dağılımlı) örneklemin ortalaması tek çekilişin beklenen değerinin aynısıdır. Bunun anlamı, k büyüklüğündeki bir örneklemin ortalaması, dağılımın ortalamasının sapmasız bir tahminidir. Bütün integral boyunca buna dikkat etmemiş olabilirsiniz, fakat eğer beklenen değeri bütün beklenen değer özeliklerini kullanarak hesaplamış olsaydınız, üstel dağılımın herhangi bir özeliğini kullanarak bunu keşf etmeyeceğinizi anlardınız.

Soru 2

(Bain/Engelhardt, s.228)

Varsayalım ki X_1, X_2, …, X_k bağımsız rasgele değişkenlerdir ve bütün i = 1, 2, …, k için $Y_i = u_i(X_i)$ olsun. Y_1, Y_2, …, Y_k'nın bağımsız olduğunu gösteriniz. Sadece X_i'nin bağımsız ve $X_i = wi(Y_i)$'nin bire-bir olduğu durumu ele alınız. *İpucu*: Eğer $x_i = w_i(y_i)$ ters dönüşüm ise, o zaman Jacobian aşağıdaki, gibidir:

$$J = \prod_{i=1}^{k} \frac{d}{dy_i} w_i(y_i).$$

Fazladan puan için Jacobian ile ilgili *ipucunu* ispatlayınız.

- Çözüm: $X_i = w_i(Y_1, Y_2, \ldots, Y_k)$'yı yazabiliriz, burada $w_i(.)$ $u_i(.)$'nin ters fonksiyonudur yani $w_i(.) = u_i^{-1}(.)$. u_i sadece X_i'nin bir fonksiyonu olduğu için, ters dönüşüm sadece Y_i veya $X_i = w_i(Y_i)$'nin bir fonksiyonu olabilir. Bu sezgisel yaklaşımla, pdf'lerin faktörlere ayrılabileceğini göstererek bağımsızlığı gösterebiliriz:

$$f_{X_1,\ldots,X_k}(x_1, \ldots, x_k) \quad = \quad f_{X_1}(x_1) \cdots f_{X_k}(x_k).$$

Şimdi dönüşüm yöntemini uygulayabiliriz:

$$f_{Y_1,\ldots,Y_k}(y_1, \ldots, y_k) = f_{X_1,\ldots,X_k}(x_1, \ldots, x_k)\,|J|$$

burada

$$J \quad = \quad \begin{vmatrix} \frac{d}{dy_1} w_1(y_1) & \cdots & \frac{d}{dy_k} w_1(y_1) \\ \vdots & \ddots & \vdots \\ \frac{d}{dy_1} w_k(y_k) & \cdots & \frac{d}{dy_k} w_k(y_k) \end{vmatrix}$$

$$J \quad = \quad \begin{vmatrix} \frac{d}{dy_1} w_1(y_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{d}{dy_k} w_k(y_k) \end{vmatrix}$$

Burada her bir w_i y_i'nin bir fonksiyonudur, y_{-i}'nin değil. Dolaysıyla, diyagonaldeki bütün türevler sıfırdır. Böylece, diyagonal bir matrisimiz olduğu için Jacobian çok kolay hesaplanabilir:

Diaygonal Matris:

$J = \prod_{i=1}^{k} \frac{d}{dy_i} w_i(y_i).$

Bu durumda, artık bütün parçaları yerine koyabiliriz:

$$
\begin{aligned}
f_{Y_1,\ldots,Y_k}(y_1,\ldots,y_k) &= f_{X_1}(w_1(x_1))\cdots f_{X_k}(w_k(y_k))\left|\prod_{i=1}^{k}\frac{d}{dy_i}w_i(y_i)\right| \\
&= f_{X_1}(w_1(x_1))\cdots f_{X_k}(w_k(y_k))\prod_{i=1}^{k}\left|\frac{d}{dy_i}w_i(y_i)\right| \\
&= \left(f_{X_1}(w_1(y_1))\left|\frac{d}{dy_1}w_1(y_1)\right|\right)\cdots\left(f_{X_k}(w_k(y_k))\left|\frac{d}{dy_k}w_k(y_k)\right|\right) \\
&= f_{Y_1}(y_1)\cdots f_{Y_k}(y_k)
\end{aligned}
$$

burada X'lerin bağımsızlığını kullandık. İlk satırda Jacobian, ikinci satırda bir çarpımın mutlak değerinin mutlak değerlerin çarpımına eşitliği, üçüncü satırda çarpımın değişme ve birleşme özelikleri ile son satırda tek bir rasgele değişkenin dönüşümünün tanımı için ipuçları verdik. Dolayısıyla, $Y_1, \ldots, Y_k$ bağımsızdır. Bitti!

Soru Üç

Sonraki bir problem setine aktarılmıştır.

Soru Dört

Örneklemin özeliklerini analiz etmek için sıra istatistiği çok yararlıdır.

1. CDF $F_X(x)$'li bir rasgele değişken X'in *n* büyüklüğündeki örnekleminin *k*.ncı sıra istatistiği için pdf ve cdf genel formüllerini yazınız.

- 1'in Çözümü: *Pdf f(x)* (eğer $a < x < b$ ise $f(x) > 0$'dır, diğer bütün durumlarda ise sıfırdır) ve *cdf F(x)* ile k'ncı sıra istatistiğinin formülü, Y_k, eğer $a < y_k < b$ ise şöyledir:

$$
g_k(y_k) = \frac{n!}{(k-1)!(n-k)!}\left[F(y_k)\right]^{k-1}\left[1-F(y_k)\right]^{n-k} f(y_k)
$$

diğer bütün durumlarda bu sıfıra eşittir. Marjinal cdf aşağıdaki gibidir:

$$
G_k(y_k) = \sum_{j=k}^{n}\binom{n}{j}\left[F(y_k)\right]^{j}\left[1-F(y_k)\right]^{n-j}
$$

Soru Beş

(Bain/Engelhardt, s.229)

X_1 ve X_2 sürekli bir dağılımdan elde edilen n = 2 büyüklüğünde rasgele bir örneklem olsun. Dağılımın pdf'si, eğer *0 < x < 1* ise, f(x) = 2x'tir, diğer durumlarda ise sıfırdır.

1. En büyük ve en küçük sıra istatistiklerin, Y_1 ve Y_2, marjinal pdf'lerini bulunuz.

- 1'in Çözümü: n = 2 büyüklüğündeki bir örneklemden edinilen en küçük ve en büyük sıra istatistiğinin marjinal pdf'leri marjinal pdf formülü kullanılarak türetilebilir. Ancak önce X'in CDF'sine ihtiyacımız var:

$$F(x) = \int_0^x 2x'dx' = x^2$$

Şimdi marjinal pdf formülünü kullanabiliriz:

$$\begin{aligned} g_1(y_1) &= \frac{2!}{(1-1)!(2-1)!}[F(y_1)]^{1-1}[1-F(y_1)]^{2-1}f(y_1) \\ &= 2\left[1-(y_1)^2\right]2y_1 \\ g_1(y_1) &= 4y_1\left[1-(y_1)^2\right] \end{aligned}$$

ve

$$\begin{aligned} g_2(y_2) &= \frac{2!}{(2-1)!(2-2)!}[F(y_2)]^{2-1}[1-F(y_2)]^{2-2}f(y_2) \\ &= 2\left[(y_2)^2\right]2y_2 \\ g_2(y_2) &= 4(y_2)^3 \end{aligned}$$

2. Beklenen değerlerini, $\mathbb{E}[Y_1]$ ve $\mathbb{E}[Y_2]$, hesaplayınız.

- 2'nin Çözümü: Bir rasgele değişkenin beklenen değer formülünü uygularız:

$$
\begin{aligned}
\mathbb{E}[Y_1] &= \int_0^1 y_1 g_1(y_1) dy_1 \\
&= \int_0^1 y_1 4y_1 \left[1-(y_1)^2\right] dy_1 \\
&= \int_0^1 \left[4(y_1)^2 - 4(y_1)^4\right] dy_1 \\
&= \left[\frac{4}{3}(y_1)^3 - \frac{4}{5}(y_1)^5\right]_0^1 \\
&= 4\left(\frac{1}{3} - \frac{1}{5}\right) \\
\mathbb{E}[Y_1] &= \frac{8}{15}.
\end{aligned}
$$

Böylece, birinci sıra istatistiğinin beklenen değeri $\frac{8}{15}$'tir. İkinci (maksimum) sıra istatistiği için aşağıdakini elde ederiz:

$$
\begin{aligned}
\mathbb{E}[Y_2] &= \int_0^1 y_2 g_2(y_2) dy_2 \\
&= \int_0^1 y_2 4(y_2)^3 dy_2 \\
&= \int_0^1 \left[4(y_2)^4\right] dy_2 \\
&= \left[\frac{4}{5}(y_2)^5\right]_0^1 \\
&= 4\left(\frac{1}{5}\right) \\
\mathbb{E}[Y_2] &= \frac{4}{5}.
\end{aligned}
$$

Şansımıza, maksimumun beklenen değerini minimumun (ilk) sıra istatistiğinden büyük bulduk.

3. Y_1 ve Y_2'nin bileşik pdf'sini bulunuz.

3'ün Çözümü: Y_1 ve Y_2'nin bileşik pdf'si y_1 ve y_2'nin permütasyonları hesaba katılarak pdf f(x)'ten elde edilebilir (f(x) ile tanımlanmış bir dağılımdan bağımsız olarak örneklenmiş n sayıda gözlem için bileşik pdf şöyle yazılabilir: $g(y_1, \ldots, y_n) = n!f(y_1) \ldots f(y_n)$):

$$\begin{aligned} f_{Y_1,Y_2}(y_1,y_2) &= 2!f(y_1)f(y_2) \\ &= 2!\,(2y_1)\,(2y_2) \\ f_{Y_1,Y_2}(y_1,y_2) &= 8y_1y_2 \end{aligned}$$

4. $R = Y_2 - Y_1$ örneklem aralığının pdf'sini bulunuz.

- 4'ün Çözümü: örneklem genişliğinin pdf'si sadece bileşik PDF'nin dönüşümüdür. Bu şimdiye kadar yaptığımız bütün büklümlere benzer. Şunu düşünün: $Y_1 = Y_1$ ve $Y_2 = R + Y_1$. O zaman dönüşüm yöntemlerini kullanabiliriz:

$$f_{Y_1,R}(y_1,r) = f_{Y_1,Y_2}(y_1, r+y_1)\,|J|$$

burada $J = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1$. Böylece, 3'teki bileşik pdf formülünü uygulayarak şunu elde ederiz:

$$f_{Y_1,R}(y_1,r) = 8y_1(r+y_1)\,|1|$$

Burada y_1'e göre integralini alıp marginal R'yi buluruz. Sınırlardaki değişime dikkat ederek devam edelim (emin olmak için önce Y_1 ve Y_2 ile sonra R ve Y_1 ile bir resim çiziniz):

$$\begin{aligned}
f_R(r) &= \int_0^{1-r} 8y_1(r+y_1)dy_1 \\
&= \left[8(\frac{1}{2}y_1^2 r + \frac{1}{3}y_1^3)\right]_0^{1-r} \\
&= 8\left[\frac{1}{2}(1-r)^2 r + \frac{1}{3}(1-r)^3\right] \\
f_R(r) &= \frac{4}{3}(1-r)^2(2+r)
\end{aligned}$$

5. Örneklem aralığının beklenen değerini hesaplayınız, $\mathbb{E}$[R].

- 5'in Çözümü: Şöyle hesaplanır:

$$\begin{aligned}
\mathbb{E}[R] &= \int_0^1 r\frac{4}{3}(1-r)^2(2+r)\,dr \\
&= \frac{4}{3}\int_0^1 \left(1-2r+r^2\right)\left(2r+r^2\right)dr \\
&= \frac{4}{3}\int_0^1 \left(2r-4r^2+2r^3+r^2-2r^3+r^4\right)dr \\
&= \frac{4}{3}\int_0^1 \left(2r-3r^2+r^4\right)dr \\
&= \frac{4}{3}\left(r^2-r^3+\frac{1}{5}r^5\right)_0^1 \\
&= \frac{4}{3}\left(\frac{1}{5}\right) \\
\mathbb{E}[R] &= \frac{4}{15}.
\end{aligned}$$

Örneklem genişliğinin beklenen değerinin birinci ve sonuncu sıra istatistiği (maksimum ve minimum sıra istatistikleri) arasındaki farka eşit olduğunu belirtmek önemlidir.

Soru Altı

(Bain/Engelhardt, s.229)

Pdf'si *0 ≤ x < ∞* durumunda *f(x)* = $\frac{1}{x2}$ ve diğer durumlarda sıfır olan bir dağılımdan elde edilen *n* büyüklüğündeki bir rasgele örneklem düşününüz.

1. Sıralı istatistiğin bileşik pdf'sini bulunuz.

- 1'in Çözümü: Bileşik pdf X_1'den X_n'e olan rasgele değişkenleri temsil eden örneklemin bileşik pdf'lerinin yeniden etiketlenmesiyle temsil edilebilir.

Özelikle şunu yazabiliriz:

$$\begin{aligned} g(y_1, ..., y_n) &= n! f(y_1) \cdots f(y_n) \\ &= n! \frac{1}{y_1^2} \cdots \frac{1}{y_n^2} \end{aligned}$$

2. En küçük sıralı istatistiğin pdf'si olan Y_1'i bulunuz.

- 2'nin Çözümü: En küçük sıralı istatistiğin pdf'si bütün diğer sıra istatistiklerine göre integral aldığımız zamanki PDF'dir. Bunun için halihazırda bir formülümüz var, dolaysıyla her şeyi yeni baştan türetmektense, onu kullanacağız. k =1 ve n'i kullanarak aşağıdakini elde ederiz:

$$
\begin{aligned}
g_k(y_k) &= \frac{n!}{(k-1)!(n-k)!}\,[F(y_k)]^{k-1}\,[1-F(y_k)]^{n-k}\,f(y_k) \\
g_1(y_1) &= n\,[1-F(y_1)]^{n-1}\,f(y_1) \\
&= n\left[1-\int_1^{y_1} f(\tilde{y}_1)d\tilde{y}_1\right]^{n-1} f(y_1) \\
&= n\left[1-\left(-\frac{1}{y_1}+1\right)\right]^{n-1}\frac{1}{y_1^2} \\
&= n\left[\frac{1}{y_1}\right]^{n-1}\frac{1}{y_1^2} \\
g_1(y_1) &= \frac{n}{y_1^{n+1}}
\end{aligned}
$$

3. Beklenen değer $\mathbb{E}[Y_1]$'i hesaplayınız. Eğer yoksa, nedenini açıklayınız.

- 3'ün Çözümü: 2'deki cevabımızda var olan rasgele değişkenin beklenen değerinin formülünü uygulayarak yine bir integral alacağız:

$$
\begin{aligned}
\mathbb{E}[Y_1] &= \int_1^{\infty} y_1\frac{n}{y_1^{n+1}}dy_1 \\
&= n\int_1^{\infty}\frac{1}{y_1^{n}}dy_1 \\
&= \frac{n}{n-1}\left[-\frac{1}{y_1^{n-1}}\right]_1^{\infty} \\
&= \frac{n}{n-1}[0+1] \\
\mathbb{E}[Y_1] &= 1+\frac{1}{n-1}.
\end{aligned}
$$

4. En büyük sıralı istatistiğin pdf'si Y_n'i bulunuz.

- 4'üm Çözümü: En büyük sıra istatistiğinin hesaplanmasında kullanılan formül en küçük sıra istatistiğinin hesaplanmasında kullanılan formülün aynısıdır:

$$
\begin{aligned}
g_k(y_k) &= \frac{n!}{(k-1)!(n-k)!}\left[F(y_k)\right]^{k-1}\left[1-F(y_k)\right]^{n-k} f(y_k) \\
g_n(y_n) &= n\left[F(y_n)\right]^{n-1} f(y_n) \\
&= n\left[\int_1^{y_n} f(\tilde{y}_n)d\tilde{y}_n\right]^{n-1} f(y_n) \\
&= n\left[\left(-\frac{1}{y_n}+1\right)\right]^{n-1}\frac{1}{y_n^2} \\
g_n(y_n) &= n\left[1-\frac{1}{y_n}\right]^{n-1}\frac{1}{y_n^2}.
\end{aligned}
$$

5. f(x)'ten yapılan tek çekiliş X'in beklenen değeri $\mathbb{E}[X]$'i hesaplayınız. İntegral sapıyor mu? O $\mathbb{E}[Y_n]$'nin varlığı hakkında ne söyler? Açıklayınız.

- 5'in Çözümü:

$$
\begin{aligned}
\mathbb{E}[X] &= \int_1^{\infty} x\frac{1}{x^2}dx \\
&= \int_1^{\infty}\frac{1}{x}dx \\
&= \left[\log x\right]_1^{\infty} \\
&= \infty - 0 \\
\mathbb{E}[X] &= \infty,
\end{aligned}
$$

ya da kelimeler ile ifade edecek olursak, integral sapar. Bu, en azından bir tane sıra istatistiğinin beklenen değerinin olamayacağını önerir. Bu durum, X'in beklenen değeri sıra istatistiğinden ya da her bir sıra istatistiğinin hesaplanıp sonra bütün n sıra istatistiklere göre ortalaması alınarak hesaplanabildiği için en küçük sıra istatistiğinin beklenen değeri olmasına rağmen geçerlidir.

Toparlayacak olursak, en küçük sıra istatistiği beklenen değere sahip olduğu için ve X için toplam beklenen değer mevcut olmadığından, en büyük sıra istatistiğinin sonlu bir beklenen değerinin olmaması gerektiğini söyleyebiliriz. Çoklu sıra istatistiği için var olmayan beklenen değer mümkün iken, hangi sıra istatistiğinin işi bozduğunu söyleyemem. Formüller ile oynayarak kendi başınıza bulmakta serbestsiniz.

6. n = 2 için örneklem aralığı, $R = Y_n - Y_1$'in pdf'sini elde ediniz. *İpucu*: kısmı fraksiyonları kullanınız, Yahoo'da "QuickMath"'i arayarak kısmı farksiyon hakkındaki yardımı bilgisayardan alabilirsiniz: http://www.quickmath.com/webMathematica3/quickmath/page.jsp?s1=algebra&s2=partial fractions&s3=basic

- 6'nın Çözümü: Yine iki rasgele değişken tanımlayacağız. Onlar, $S = Y_1$ ve $R = Y_n - Y_1$'dir ve tersleri $S = Y_1$ ve $Y_2 = R + S$'dir. 1'deki bileşik pdf şimdi kullanılabilir:

$$
\begin{aligned}
g(y_1, \ldots, y_n) &= n! \frac{1}{y_1^2} \cdots \frac{1}{y_n^2} \\
g(y_1, y_2) &= \frac{2}{y_1^2 y_2^2}
\end{aligned}
$$

Jacobian'nın determinantının 1 olduğunu not ederek, $J = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1$, dönüşümün formülünü kullanırız:

$$
\begin{aligned}
f_{R,S}(r, s) &= g(s, r+s)\,|1| \\
&= \frac{2}{s^2(r+s)^2} \\
f_R(r) &= \int_1^{\infty} \frac{2}{s^2(r+s)^2} ds
\end{aligned}
$$

burada, bilgisayar kullanmıyorsanız, kısmi fraksiyonlar ile çözmek çok acı verici olur[1].

Bu, kısmı fraksiyon ayrıştırmadır:

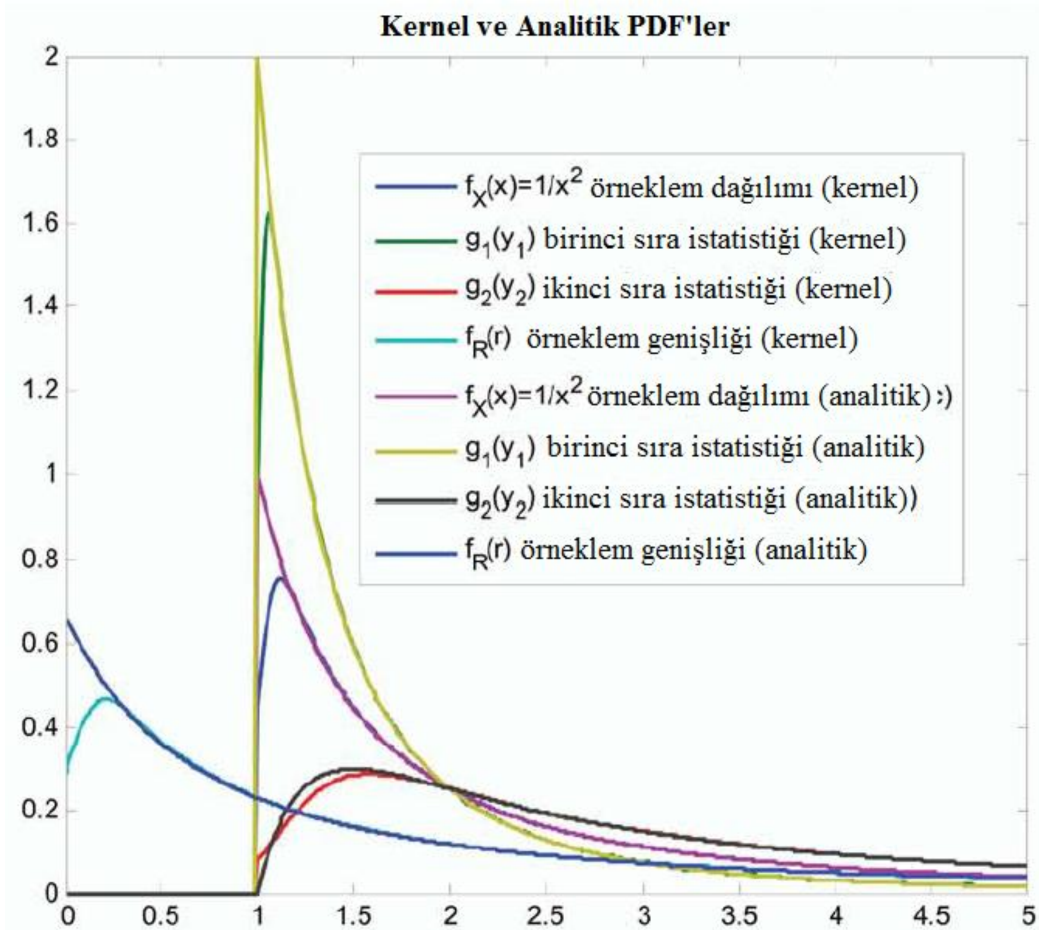
[1] Bunu yapan bir web sitesi bulundu: http://www.quickmath.com/webMathematica3/quickmath/page.jsp?s1=algebra&s2=partialfractions&s3=basic

$$\frac{2}{s^2(r+s)^2} = \frac{4}{r^3(r+s)} + \frac{2}{r^2(r+s)^2} - \frac{4}{r^3 s} + \frac{2}{r^2 s^2}$$

bunun şimdi integralini alabiliriz. Pdf için bulduğumuz sonuç şöyledir:

$$\begin{aligned}
f_R(r) &= \int_1^\infty \left[\frac{4}{r^3(r+s)} + \frac{2}{r^2(r+s)^2} - \frac{4}{r^3 s} + \frac{2}{r^2 s^2}\right] ds \\
&= \left[\frac{4}{r^3}\log(\frac{r+s}{s}) - \frac{2}{r^2(r+s)} - \frac{2}{r^2 s}\right]_1^\infty \\
&= \left[\frac{4}{r^3}\log(1+\frac{r}{s}) - \frac{2}{r^2(r+s)} - \frac{2}{r^2 s}\right]_1^\infty \\
&= \left[0 - 0 - 0 - \frac{4}{r^3}\log(1+\frac{r}{s}) + \frac{2}{r^2(r+1)} + \frac{2}{r^2}\right] \\
f_R(r) &= \frac{2}{r^2}\left[1 + \frac{1}{1+r} - \frac{2}{r} log(1+r)\right]
\end{aligned}$$

burada $r > 0$'dır. Örneklem genişliğinin PDF'si ve $n = 2$ büyüklüğündeki bir örneklemin birinci ve ikinci sıra istatistikleri aşağıdaki grafikte gösterilmiştir ve $n = 2$ büyüklüğündeki bir milyon örneklemden teorinin göstergesine kadar olan çekirdek tahmin ile karşılaştırılmıştır.



7. Beklenen değer $\mathbb{E}[R]$'yi hesaplayınız. Eğer yoksa, nedenini açıklayınız.

7'nin Çözümü: Maksimum sıra istatistiğinin beklenen değeri gibi genişliğin beklenen değeri de mevcut değildir ve $\mathbb{E}[R] = \mathbb{E}[Y_1] - \mathbb{E}[Y_2]$'dir. $\mathbb{E}[Y_1]$ mevcut olsa bile, hesaplama için her ikisinin de mevcut olması gerekir.

8. n'nin tekli sayı olduğu ve $r = (n + 1)/2 \in \mathbb{N}$ durumunda, örneklemin medyanın pdf'sini, Y_r, bulunuz. Pdf'yi sadece r ve y_r cinsinden ifade ediniz (bütün *n*'leri ve *k*'leri elimine ediniz).

- 8'in Çözümü: Örneklemin medyanı r = k = (n+1)/2'li sıra istatistiğidir. Daha önce yaptığımız gibi hesaplayabiliriz. Formülü kullanarak:

$$
\begin{aligned}
g_r(y_r) &= \frac{n!}{(r-1)!(n-r)!}\left[F(y_r)\right]^{r-1}\left[1-F(y_r)\right]^{n-r} f(y_r) \\
&= \frac{(2r-1)!}{(r-1)!(r-1)!}\left[1-\frac{1}{y_r}\right]^{r-1}\left[\frac{1}{y_r}\right]^{r-1}\frac{1}{y_r^2} \\
g_r(y_r) &= \frac{(2r-1)!}{[(r-1)!]^2}\frac{(y_r-1)^{r-1}}{y_r^{2r}}
\end{aligned}
$$

9. Beklenen değer $\mathbb{E}[Y_r]$'yi hesaplayınız. Neden mevcut olmadığını açıklayınız.

- 9'un Çözümü: Beklenen değeri hesaplayalım:

$$
\begin{aligned}
\mathbb{E}[Y_r] &= \int_1^\infty y_r g_r(y_r) dy_r \\
&= \int_1^\infty y_r \frac{(2r-1)!}{[(r-1)!]^2}\frac{(y_r-1)^{r-1}}{y_r^{2r}} dy_r \\
&= \frac{(2r-1)!}{[(r-1)!]^2}\int_1^\infty \frac{(y_r-1)^{r-1}}{y_r^{2r-1}} dy_r \\
&= \frac{(2r-1)!}{[(r-1)!]^2}\left[\left(-\frac{1}{r}\frac{(y_r-1)^r}{y_r^{2r-1}}\right)_1^\infty - \int_1^\infty -\frac{2r-1}{r}\frac{(y_r-1)^r}{y_r^{2r}} dy_r\right] \\
&= \left[\frac{(2r-1)!}{[(r-1)!]^2}\left(-\frac{1}{r}\frac{(y_r-1)^r}{y_r^{2r}}\right)_1^\infty + \frac{2r-1}{r}\int_1^\infty (y_r-1)\left(\frac{(2r-1)!}{[(r-1)!]^2}\frac{(y_r-1)^{r-1}}{y_r^{2r}}\right) dy_r\right] \\
&= \left[0 + \frac{2r-1}{r}\mathbb{E}\left[Y_r - 1\right]\right] \\
\mathbb{E}[Y_r] &= \frac{2r-1}{r}\left[\mathbb{E}\left[Y_r\right] - 1\right]
\end{aligned}
$$

Şimdi $\mathbb{E}[Y_r]$ için çözebiliriz:

$$
\mathbb{E}[Y_r] = \frac{2r-1}{r-1}.
$$

Birkaç sayısal uygulama bunun gerçekte doğru formül olduğunu gösterir: r =2 için, medyanı 3 olarak elde ederiz ve r = 3 için medyan 2.5'tir. r = 5 için medyan 2.25'tir.