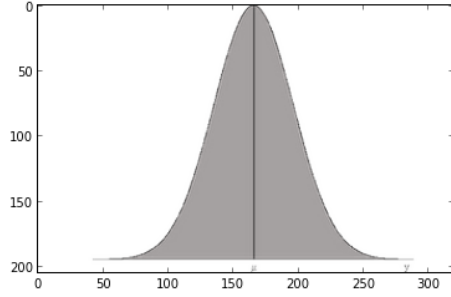


Giris

Bu notlar makine ogrenimi, veri madenciligi gibi konularda gerekli olasilik ve istatistik bilgisini paylamak icin hazirlaniyor. Notlarda olasilik ve istatistik ayni anda anlatilacak, ve uygulamalara agirlik verilecek.

Dağılımlar Hakkında

Doğadan yapılan çoğu ölçümlerin, sıklık grafiğini alınca sonucun aşağıda gibi çıkması ilginçtir.



Mesela, Türkiye'deki 2000 yetişkinin kilosunu ölçün. Grafiğini alın, kesinlikle yukarıdaki tepe şekli çıkacak. Ya da, 1000 kişinin boyunu ölçün, aynı tepe şekli. Keskin nişancının hedefe attığı kurşunların hedefe gelişini en iyi 12 en kötü 1 olmak üzere ölçün, sıklık grafiğini alın. Gene aynı tepe şekli!

Nasıl oluyor bu iş?

Açıklama için, normal dağılım eğrisinden bahsetmemiz gerekecek.

Not olarak düşelim: Sıklık grafiği, X sayısının ne kadar çıktığını sayıp, Y ekseninde bu sayıyı X'e tekabül ederek kolon olarak göstermeye denir. Mesela, 60 kilo değeri 13 kere çıktı ise, $X=60$, $Y=13$ gibi bir kolon çizilecektir.

Normal Dağılım Eğrisi

Normal dağılımın olasılık kavramı ile yakın bağları var. Bu konuda ünlü bir deney zar atma deneyidir. Mesela, elimizde tek bir zar olsun, ve bu zarı arka arkaya atalım. Sabrımız yeterse 1000 kere atalım. Sonuçta, sıklık grafiği eşit bir dağılım gösterecektir. (Zar tutmuyorsanız :))

Bunun sebeplerini anlamak zor değil. Her zar atış olayı birbirinden bağımsız, ve her sayının üstte gelme ihtimali birbirine eşit olduğu için ($1/6$), her sayıdan eşit miktarda gelecektir. Tabii bunun için deneyin birçok kere tekrarlanması gerekiyor.

Fakat, bir yerine 2 zar atalım. Hatta hatta, 4 zar atalım, ve bu sefer sıklık grafik hanesine yazmadan çıkan sayıları önce toplayalım. Bu çıkan toplamın sıklık grafiğini alalım.

İşte bu sıklık grafiği göreceğiz ki, önceden bahsettiğimiz tepe grafiğine yaklaşıyor. Ne kadar çok zar atarsanız, bu benzerlik o kadar daha fazla olacaktır.

Bunun sebebi de aslında basit: 1..6 arası sayıların tek bir zardan gelme olasılığı

aynı, evet. Fakat toplamlara gelince, mesela iki zarlı örnekte, 10 sayısının olasılığı 2 sayısından daha yüksek. Çünkü, 10 sayısını 5-5, 4-6 ya da 6-4 ile alabiliyoruz. 2 sayısı sadece 1-1 ile geliyor.

Buradan şu sonuç çıkabilir: Eğer doğada ölçtüğümüz bir kavramın oluşmasında birden fazla etken var ise, o ölçümlerin sıklığı her zaman tepe şekli ile olacaktır.

Fakat, daha da gizemli olan bir olay şudur; Sabit olan bir şeyi ölçtüğümüzde (yaptığımız hatalar sonucu) çıkan grafiğin bile tepe şekilli olması! Yani, doğru dürüst hata yapmak bile elimizde değil gibi gözüküyor... Bunun tabii ki olasılık açıklamaları olacaktır. İzlediğimiz matematikçilerden bu en son konuda net bir açıklama alamadık.

Simulasyon

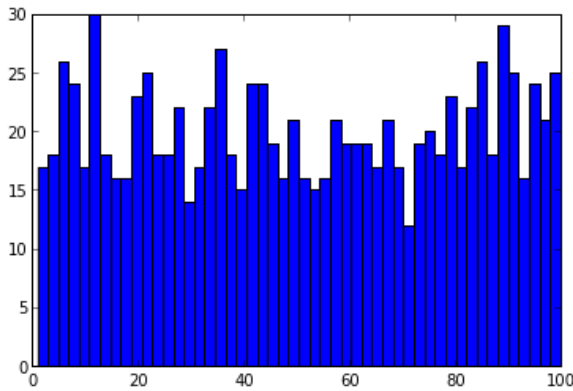
Eğer bu kavramları simulasyon ortamında göstermek istersek, Python ile bunu yapabiliriz.

İlk önce, Random.org sitesinden rasgele sayı üretip bilgiyarımıza kopyalacağız. Bahsettiğimiz site, kimsenin kullanmadığı radyo kanallarından atmosfer gürültüsü dinleyip, bu gürültüleri sayısal değere çevirerek rasgele sayı üretiyor.

Gerçek rasgele sayı üretmek pek kolay bir iş değil. Her ne kadar bilgisayarınızda rasgele sayı üreten birçok algoritma olsa bile, bu algoritmalar belli bir sayı üretiminden sonra kendini tekrar etmeye başlıyorlar. Gerçek rasgele sayılar için muhakkak dış bir kaynağa bağlanmak gerekiyor.

Gösterimiz için, rasgele sayıları üretip, bir dat dosyasına koyuyoruz. Python ile bu sayıları okuyup, ilk önce teker teker sayıların sıklık grafiğini, ondan sonra sayıları üçer üçer toplayıp, onların grafiğini alıp göstereceğiz. Aşağıda bu iki grafiği bulabilirsiniz.

```
A = loadtxt('rasgele.dat')
plt.hist(A, 50)
plt.savefig('dagilim_1.png')
```



```
A = loadtxt('rasgele.dat');
```

```

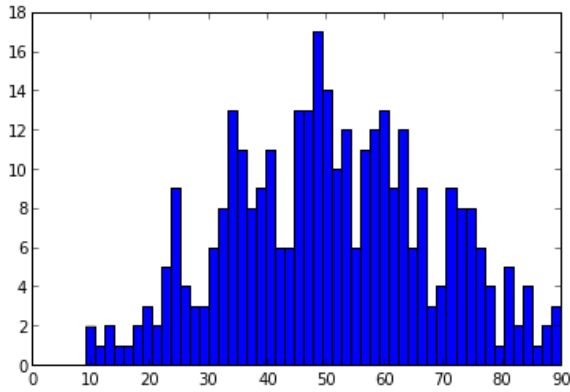
B = []

i = 1;

while (i < 998):
    toplam = 0
    s = A[i]
    toplam = toplam + s
    s = A[i+1]
    toplam = toplam + s
    s = A[i+2]
    toplam = toplam + s
    B.append(toplam/3)
    i = i + 3

plt.hist(B, 50);
plt.savefig('dagilim_2.png')

```



Olasilik

Orneklem Uzayi (Sample Space)

Orneklem uzayi Ω bir deneyin mumkun tum olasiliksal sonuclarin (outcome) kumesidir. Eger deneyimiz ardi ardina iki kere yazi (T) tura (H) atip sonucu kaydetmek ise, bu deneyin mumkun tum sonuclari soyledir

$$\Omega = \{HH, HT, TH, TT\}$$

Sonuclar ve Olaylar (Outcomes and Events)

Ω icindeki her nokta bir sonuctur (outcome). Olaylar Ω 'nin herhangi bir alt kumesidir ve sonuclardan olusurlar. Mesela ustteki yazi-tura deneyinde "iki atisin icinden ilk atisin her zaman H gelmesi olayi" boyle bir alt kumedir, bu olaya A diyelim, $A = \{HH, HT\}$.

Ya da bir deneyin sonucu ω fiziksel bir olcüm , diyelin ki sicaklik olcumu. Sicaklik \pm , reel bir sayi olduguna gore, $\Omega = (-\infty, +\infty)$, ve sicaklik olcumunun 10'dan

buyuk ama 23'ten kucuk ya da esit olma "olayi" $A = (10, 23]$. Koseli parantez kullanildi cunku sinir degerini dahil ediyoruz.

Ornek

10 kere yazi-tura at. $A =$ "en az bir tura gelme" olayi olsun. T_j ise j 'inci yazi-tura atisinda yazi gelme olayi olsun. $P(A)$ nedir?

Bunun hesabi icin en kolayi, hic tura gelmeme, yani tamamen yazi gelme olasiligini, A^c 'yi hesaplamak, ve onu 1'den cikartmaktir. c sembolu "tamamlayici (complement)" kelimesinden geliyor.

$$P(A) = 1 - P(A^c)$$

$$= 1 - P(\text{hepsi yazi})$$

$$= 1 - P(T_1)P(T_2)...P(T_{10})$$

$$= 1 - \left(\frac{1}{2}\right)^{10} \approx .999$$

Rasgele Degiskenler (Random Variables)

Bir rasgele degisken X bir eslemedir, ki bu esleme $X : \Omega \rightarrow \mathbb{R}$ her sonuc ile bir reel sayi arasindaki eslemedir.

Olasilik derslerinde bir noktadan sonra artik ornekleme uzayindan bahsedilmez, ama bu kavramin arkalarda bir yerde her zaman devrede oldugunu hic aklimizdan cikartmayalim.

Ornek

10 kere yazi-tura attik diyelim. VE yine diyelim ki $X(\omega)$ rasgele degiskeni her ω siralamasinda (sequence) olan tura sayisi. Iste bir esleme. Mesela eger $\omega = \text{HHTHHTHHTT}$ ise $X(\omega) = 6$. Tura sayisi eslemesi ω sonucunu 6 sayisina esledi.

Ornek

$\Omega = \{(x, y); x^2 + y^2 \leq 1\}$, yani kume birim cember ve icindeki reel sayilar (unit disc). Diyelim ki bu kumeden rasgele secim yapıyoruz. Tipik bir sonuc $\omega = (x, y)$ 'dir. Tipik rasgele degiskenler ise $X(\omega) = x$, $Y(\omega) = y$, $Z(\omega) = x + y$ olabilir. Goruldugu gibi bir sonuc ile reel sayi arasinda esleme var. X rasgele degiskeni bir sonucu x 'e eslemis, yani (x, y) icinden sadece x 'i cekip cikartmis. Benzer sekilde Y, Z degiskenleri var.

Toplamsal Dagilim Fonksiyonu (Cumulative Distribution Function -CDF-)

Tanim

X rasgele degiskeninin CDF'i $F_X : \mathbb{R} \rightarrow [0, 1]$ tanimi

$$F_X(x) = P(X \geq x)$$

Eger X ayriksal ise, yani sayilabilir bir kume $\{x_1, x_2, \dots\}$ icinden degerler aliyorsa olasilik fonksiyonu (probability function), ya da olasilik kutle fonksiyonu (probability mass function -PMF-)

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

Bazen f_X , ve F_X yerine sadece f ve F yazariz.

Tanim

Eger X surekli (continuous) ise, yani tum x 'ler icin $f_X(x) > 0$, $\int_{-\infty}^{+\infty} f(x)dx = 1$ olacak sekilde bir f_X mevcut ise, o zaman her $a \leq b$ icin

$$P(a < X < b) = \int_a^b f_X(x)dx$$

Bu durumda f_X olasilik yogunluk fonksiyonudur (probability density function -PDF-).

$$F_X = \int_{-\infty}^x f_X(t)dt$$

Ayrica $F_X(x)$ 'in turevi alinabildigi her x noktasinda $f_X(x) = F'_X(x)$ demektir.

Dikkat! Eger X surekli ise o zaman $P(X = x) = 0$ degerindedir. $f(x)$ fonksiyonunu $P(X = x)$ olarak gormek hatalidir. Bu sadece ayriksal rasgele degiskenler icin isler. Surekli durumda olasilik hesabi icin belli iki nokta arasinda integral hesabi yapmamiz gereklidir. Ek olarak PDF 1'den buyuk olabilir, ama PMF olamaz. PDF'in 1'den buyuk olabilmesi integrali bozmaz mi? Unutmayalim, integral hesabi yapiyoruz, noktasal degerlerin 1 olmasi tum 1'lerin toplandigi anlamina gelmez. Bakiniz *Entegralleri Nasil Dusunelim* yazimiz.

Tanim

X rasgele degiskeninin CDF'i F olsun. Ters CDF (inverse cdf), ya da ceyrek fonksiyonu (quantile function)

$$F^{-1}(q) = \inf \left\{ x : F(x) \leq q \right\}$$

ki $q \in [0, 1]$. Eger F kesinlikle artan ve surekli bir fonksiyon ise $F^{-1}(q)$ tekil bir x sayisi ortaya cikarir, ki $F(x) = q$.

Eger inf kavramini bilmiyorsak simdilik onu minimum olarak dusunebiliriz.

$F^{-1}(1/4)$ birinci ceyrek

$F^{-1}(1/2)$ medyan (median, ya da ikinci ceyrek),

$F^{-1}(3/4)$ ucuncu ceyrek

olarak bilinir.

İki rasgele degisken X ve Y dagilimsal olarak birbirine esitligi, yani $X \stackrel{d}{=} Y$ eger $F_X(x) = F_Y(x)$, $\forall x$. Bu X, Y birbirine esit, birbirinin aynisi demek degildir. Bu degiskenler hakkındaki tum olasiliksal islemler, sonuclar ayni olacak demektir.

Uyari! “ X ’in dagilimi F ’tir” beyanini $X \sim F$ seklinde yazmak bir gelenek. Bu biraz kotu bir gelenek aslinda cunku \sim sembolu ayni zamanda yaklasiksallik kavramini belirtmek icin de kullaniliyor.

Bernoulli Dagilimi

X ’in bir yazi-tura atisini temsil ettigini dusunelim. O zaman $P(X = 1) = p$, ve $P(X = 0) = 1 - p$ olacaktır, ki $p \in [0, 1]$ olmak uzere. O zaman X ’in dagilimi Bernoulli deriz, ve $X \sim \text{Bernoulli}(p)$ diye gosteririz. Olasilik fonksiyonu $f(x) = p^x(1 - p)^{(1-x)}$, $x \in \{0, 1\}$.

Yani x ya 0, ya da 1. Parametre p , 0 ile 1 arasindaki herhangi bir reel sayi.

Uyari!

X bir rasgele degisken; x bu degiskenin alabilecegi spesifik bir deger; p degeri ise bir **parametre**, yani sabit, onceden belirlenmis reel sayi. Tabii istatistiki problemlerde (olasilik problemlerinin tersi olarak dusunursek) cogunlukla o sabit parametre bilinmez, onun veriden hesaplanmasi, kestirilmesi gerekir. Her halukarda, cogu istatistiki modelde rasgele degiskenler vardır, ve onlardan ayri olarak parametreler vardır. Bu iki kavrami birbiriyle karistirmayalim.

Duz (Uniform) Dagilim

X duz, $\text{Uniform}(a, b)$ olarak dagilmis deriz, ve bu $X \sim \text{Uniform}(a, b)$ olarak yazilir eger

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \text{ icin} \\ 0 & \text{digerleri} \end{cases}$$

ise ve $a < b$ olacak sekilde. CDF hesabi olasilik egrisinin integralini temel alir, duz dagilim bir a, b arasinda $1/b - a$ yuksekliginde bir dikdortgen seklinde olacagi icin, bu dikdortgendeki herhangi bir x noktasinda CDF dagilimi, yani o x ’in baslayip sol tarafin alaninin hesabi basit bir dikdortgensel alan hesabidir, yani $x - a$ ile $1/b - a$ ’nin carpimidir, o zaman

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & x \in [a, b] \\ 1 & x > b \end{cases}$$

Normal (Gaussian) Dagilim

$X \sim N(\mu, \sigma^2)$ ve PDF

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma^2}(x - \mu)^2 \right\}, x \in \mathbb{R}$$

ki $\mu \in \mathbb{R}$ ve $\sigma > 0$ olacak sekilde.

İleride gorecegiz ki μ bu dagilimin “ortasi”, ve σ onun etrafa ne kadar “yay-ildigi” (spread). Normal dagilim olasilik ve istatistikte cok onemli bir rol oynar. Dogadaki pek cok olay yaklasiksal olarak Normal dagilima sahiptir. Sonra gorecegimiz uzere, mesela bir rasgele degiskenin degerlerinin toplami her zaman Normal dagilima yaklasir (Merkezi Limit Teorisi -Central Limit Theorem-).

Eger $\mu = 0$ ve $\sigma = 1$ ise X 'in standart Normal dagilim oldugunu soyleriz. Gele- nege gore standart Normal dagilim rasgele degiskeni Z ile gosterilmelidir, PDF ve CDF $\phi(z)$ ve $\Phi(z)$ olarak gosterilir.

$\Phi(z)$ 'nin kapali form (closed-form) tanimi yoktur. Bu, matematikte “analitik bir forma sahip degil” demektir, formulu bulunamamaktadır, bunun sebebi ise Normal PDF'in integralinin analitik olarak alinamiyor olusudur.

Bazi faydali puf noktaları

1. Eger $X \sim N(\mu, \sigma^2)$ ise, o zaman $Z = (X - \mu)/\sigma \sim N(0, 1)$.
2. Eger $Z \sim N(0, 1)$ ise, o zaman $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$
3. Eger $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2, \dots$ ve her X_i digerlerinden bagimsiz ise, o zaman

$$\sum_{i=1}^n X_i \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$$

Tekrar $X \sim N(\mu, \sigma^2)$ alirsak ve 1. kuraldan devam edersek / temel alirsak su da dogru olacaktır.

$$P(a < X < b) = ?$$

$$= P\left(\frac{a - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{b - \mu}{\sigma}\right)$$

$$= P\left(\frac{a - \mu}{\sigma} < Z < \frac{b - \mu}{\sigma}\right) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

İlk gecisi nasıl elde ettik? Bir olasılık ifadesi $P(\cdot)$ içinde eşitliğin iki tarafına aynı anda aynı toplama, çıkarma operasyonlarını yapabiliriz.

Son ifadenin anlamı sudur. Eğer standart Normal'in CDF'ini hesaplayabiliyorsak, istediğimiz Normal olasılık hesabını yapabiliriz demektir, çünkü artık X içeren bir hesabın Z 'ye nasıl tercüme edildiğini görüyoruz.

Tüm istatistik yazılımları $\Phi(z)$ ve $\Phi(z)^{-1}$ hesabı için gerekli rutinlere sahiptir. Tüm istatistik kitaplarında $\Phi(z)$ 'nin belli değerlerini taşıyan bir tablo vardır. Ders notlarımızın sonunda da benzer bir tabloyu bulabilirsiniz.

Örnek

$X \sim N(3, 5)$ ise $P(X > 1)$ nedir? Cevap:

$$P(X > 1) = 1 - P(X < 1) = 1 - P(Z < \frac{1-3}{\sqrt{5}}) = 1 - \Phi(-0.8944) = .81$$

Soru tam $P(a < X < b)$, sadece b olduğu için yukarıdaki form ortaya çıktı.

Örnek

Şimdi öyle bir q bul ki $P(X < q) = .2$ olsun. Yani $\Phi^{-1}(.2)$ 'yi bul. Yine $X \sim N(3, 5)$.

Cevap

Demek ki tablodan .2 değerine tekabül eden esik değerini bulup, üstteki formül üzerinden geriye tercüme etmemiz gerekiyor. Normal tablosunda $\Phi(-0.8416) = .2$,

$$.2 = P(X < q) = P(Z < \frac{q - \mu}{\sigma}) = \Phi(\frac{q - \mu}{\sigma})$$

O zaman

$$-0.8416 = \frac{q - \mu}{\sigma} = \frac{q - 3}{\sqrt{5}}$$

$$q = 3 - 0.8416\sqrt{5} = 1.1181$$

t (Student's t) ve Cauchy Dağılımı

X , ν derece bağımsızlıkta t dağılımına sahiptir, ki bu $X \sim t_\nu$ diye yazılır eğer

$$f(x) = \frac{\Gamma(\nu + 1)/2}{\sqrt{\nu\pi}\Gamma(\nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}$$

t dağılımı Normal dağılıma benzer ama daha kuyruğu daha kalındır. Aslında Normal dağılımı t dağılımının $\nu = \infty$ olduğu hale tekabül eder. Cauchy dağılımı

da t 'nin özel bir halidir, $v = 1$ halidir. Bu durumda yoğunluk fonksiyonu

$$f(x) = \frac{1}{\pi(1+x^2)}$$

Bu formül hakikaten bir yoğunluk mudur? Kontrol için entegralini alalım,

$$\int_{-\infty}^{\infty} f(x) dx = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{dx}{1+x^2}$$

Cogunlukla entegre edilen yerde "1 arti ya da eksi bir seyin karesi" turunde bir ifade gorulurse, yerine gecirme (substitution) islemi trigonometrik olarak yapilir.

$$x = \tan \theta, \theta = \arctan x$$

$$1+x^2 = 1+\tan^2 \theta = \sec^2 \theta$$

$$dx/d\theta = \sec^2 \theta$$

O zaman

$$\begin{aligned} &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{dx}{1+x^2} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1}{\sec^2 \theta} \sec^2 \theta d\theta = \frac{1}{\pi} \int_{-\infty}^{\infty} 1 d\theta = \\ &= \frac{1}{\pi} \theta|_{-\infty}^{\infty} = \frac{1}{\pi} [\arctan(\infty) - \arctan(-\infty)] \\ &= \frac{1}{\pi} \left[\frac{\pi}{2} - \left(-\frac{\pi}{2} \right) \right] = 1 \end{aligned}$$

χ^2 Dagilimi

X 'in p derece serbestlige sahip bir χ^2 dagilima sahip ise $X \sim \chi_p^2$ olarak gosterilir, yoğunluk

$$f(x) = \frac{1}{\Gamma(p/2)2^{p/2}} x^{(p/2)-1} e^{-x/2}, x > 0$$

Eger Z_1, \dots, Z_p bagimsiz standart Normal rasgele degiskenler ise, $\sum_{i=1}^p Z_i^2 \sim \chi_p^2$ esitligi dogrudur.

İki Degiskenli Dagilimler

Tanim

Surekli ortamda (X, Y) rasgele degiskenleri icin yogunluk fonksiyonu $f(x, y)$ tanimlanabilir eger i) $f(x, y) > 0, \forall (x, y)$ ise, ve ii) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$ ise ve her kume $A \subset \mathbb{R} \times \mathbb{R}$ icin $P((X, Y) \in A) = \int \int_A f(x, y) dx dy$. Hem ayriksal hem surekli durumda birlesik (joint) CDF $F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$ diye gosterilir.

Bu tanimda A kumesi olarak tanimlanan kavram uygulamalarda bir olaya (event) tekabul eder. Mesela

Ornek

(X, Y) 'in birim kare uzerinde duz (uniform) olsun. O zaman

$$f(x, y) = \begin{cases} 1 & \text{eger } 0 \leq x \leq 1, 0 \leq y \leq 1 \text{ ise} \\ 0 & \text{diger durumlarda} \end{cases}$$

$P(X < 1/2, Y < 1/2)$ 'yi bul.

Cevap

Burada verilen $A = \{X < 1/2, Y < 1/2\}$ bir altkumedir ve bir olaydir. Olaylari boyle tanimlamamis miydik? Orneklem uzayinin bir altkumesi olay degil midir? O zaman f 'i verilen altkume uzerinden entegre edersek, sonuca ulasmis oluruz.

Ornek

Eger dagilim kare olmayan bir bolge uzerinden tanimliysa hesaplar biraz daha zorlasabilir. (X, Y) yogunlugu

$$f(x, y) = \begin{cases} cx^2y & \text{eger } x^2 \leq y \leq 1 \\ 0 & \text{digerleri} \end{cases}$$

Niye c bilinmiyor? Belki problemin modellemesi sirasinda bu bilinmez olarak ortaya cikmistir. Olabilir. Bu degeri hesaplayabiliriz, cunku $f(x, y)$ yogunluk olmalı, ve yogunluk olmanin sarti $f(x, y)$ entegre edilince sonucun 1 olmasi.

Once bir ek bilgi uretelim, eger $x^2 \leq 1$ ise, o zaman $-1 \leq x \leq 1$ demektir. Bu lazim cunku entegrale sinir degeri olarak verilecek.

$$\begin{aligned} 1 &= \int \int f(x, y) dy dx = c \int_{-1}^1 \int_{x^2}^1 x^2 y dy dx \\ &= c \int_{-1}^1 x^2 \int_{x^2}^1 y dy dx = \int_{-1}^1 x^2 \left(\frac{1}{2} - \frac{x^4}{2} \right) dx = 1 \\ &= c \int_{-1}^1 x^2 \left(\frac{1 - x^4}{2} \right) dx = 1 \end{aligned}$$

$$= \frac{c}{2} \int_{-1}^1 x^2 - x^6 dx = 1$$

Devam edersek $c = 21/4$ buluruz.

Simdi, diyelim ki bizden $P(X \geq Y)$ 'yi hesaplamamiz isteniyor. Bu hangi A bolgesine tekabul eder? Elimizdekiler

$$-1 \leq x \leq 1, x^2 \leq y, y \leq 1$$

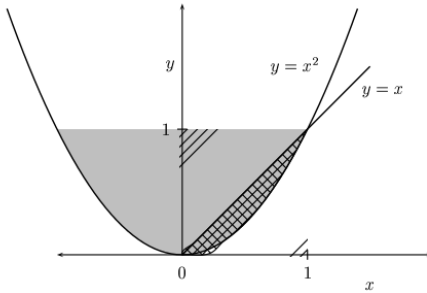
Simdi bunlara bir de $y \leq x$ eklememiz lazim. Yani ortadaki esitsizlige bir oge daha eklenir.

$$-1 \leq x \leq 1$$

$$x^2 \leq y \leq x$$

$$y \leq 1$$

$x^2 \leq y$ 'yi hayal etmek icin $x^2 = y$ 'yi dusunelim, bu bir parabol olarak cizilebilir, ve parabolun ustunde kalanlar otomatik olarak $x^2 \leq y$ olur, bu temel irdelemelerden biri.



Ayni sekilde $y \leq x$ icin $y = x$ 'i dusunelim, ki bu 45 derece aciyla cizilmis duz bir cizgi. Cizginin alti $y \leq x$ olur. Bu iki bolgenin kesisimi yukaridaki resimdeki golgeli kisim.

Ek bir bolge sarti $0 \leq x \leq 1$. Bu sart resimde bariz goruluyor, ama cebirsel olarak bakarsak $y \geq x^2$ oldugunu biliyoruz, o zaman $y \geq 0$ cunku x^2 muhakkak bir pozitif sayi olmalidir. Diger yandan $x \geq y$ verilmiş, tum bunlari yanyana koyarsak $x \geq 0$ sarti ortaya cikar.

Artik $P(X \geq Y)$ hesabi icin haziriz,

$$P(X \geq Y) = \frac{21}{4} \int_0^1 \int_{x^2}^x x^2 y dy dx = \frac{21}{4} \int_0^1 x^2 \left[\int_{x^2}^x y dy \right] dx$$

$$= \frac{21}{4} \int_0^1 x^2 \frac{x^2 - x^4}{2} dx = \frac{3}{20}$$

“Hafizasız” Dagilim, Ustel (Exponential) Dagilim

Ustel dagilimin hafizasız olduğu söylenir. Bunun ne anlama geldiğini anlatmaya uğrasalım. Diyelim ki rasgele değişken X bir aletin omrunu temsil ediyor, yani bir $p(x)$ fonksiyonuna bir zaman “sordüğümüz” zaman bize dondurulan olasılık, o aletin x zamanı kadar daha işlemesinin olasılığı. Eğer $p(2) = 0.2$ ise, aletin 2 yıl daha yaşamasının olasılığı 0.2.

Bu hafizasızlığı, olasılık matematiği ile nasıl temsil ederiz?

$$P(X > s + t | X > t) = P(X > s), \forall s, t \geq 0$$

Yani öyle bir dağılım var ki elimizde, $X > t$ bilgisi veriliyor, ama (kalan) zamanı hala $P(X > s)$ olasılığı veriyor. Yani t kadar zaman geçtiği bilgisi hiçbir şeyi değiştirmiyor. Ne kadar zaman geçmiş olursa olsun, direkt s ile gidip aynı olasılık hesabını yapıyoruz.

Sartsal (conditional) formülünü uygularsak üstteki şöyle olur

$$\frac{P(X > s + t, X > t)}{P(X > t)} = P(X > s)$$

ya da

$$P(X > s + t, X > t) = P(X > s)P(X > t)$$

Bu son denklemin tatmin olması için X ne şekilde dağılmış olmalıdır? Üstteki denklem sadece X dağılım fonksiyonu ustel (exponential) olursa mümkündür, çünkü sadece o zaman

$$e^{-\lambda(s+t)} = e^{-\lambda s} e^{-\lambda t}$$

gibi bir ilişki kurulabilir.

Örnek

Diyelim ki bir bankadaki bekleme zamanı ortalama 10 dakika ve ustel olarak dağılmış. Bir müşterinin i) bu bankada 15 dakika beklemesinin ihtimali nedir? ii) Bu müşterinin 10 dakika bekledikten sonra toplam olarak 15 dakika (ya da daha fazla) beklemesinin olasılığı nedir?

Cevap

i) Eğer X müşterinin bankada beklediği zamanı temsil ediyorsa

$$P(X > 15) = e^{-15 \cdot 1/10} = e^{-3/2} \approx 0.223$$

ii) Sorunun bu kısmi müşteri 10 dakika gecirdikten sonra 5 dakika daha gecirmesinin olasılığını soruyor. Fakat üstel dağılım “hafızasız” olduğu için kalan zamanı alıp yine direkt aynı fonksiyona geçiyoruz,

$$P(X > 5) = e^{-5 \cdot 1/10} = e^{-1/2} \approx 0.60$$

Bilesen (Marginal) Dağılımlar

Süreklili rasgele değişkenler için bileşen yoğunluk

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

ve

$$f_Y(y) = \int f(x, y) dx$$

Üstteki integraller gerçek bir dağılım fonksiyonu $f(x, y)$ verilince alt ve üst limit te tanımlamak zorundadır. Çünkü bileşen yoğunluk için bir veya daha fazla değişkeni “integralle dışarı atmak (integrate out)” ettiğimiz söylenir, eğer ayrık-sal (discrete) ortamda olsaydık bu atılan değişkenin tüm değerlerini goze alarak toplama yapan bir formül yazardık. Sürekli ortamda integral kullanıyoruz, ama tüm değerlerin üzerinden yine bir şekilde geçmemiz gerekiyor. İşte alt ve üst limitler bunu gerçekleştiriyor. Bu alt ve üst limitler, atılan değişkenin “tüm değerlerine” bakması gerektiği için $-\infty, +\infty$ olmalıdır. Eğer problem içinde değişkenin belli değerler arasında olduğu belirtilmiş ise (mesela alttaki örnekte $x > 0$) o zaman entegral limitleri alt ve üst sınırını buna göre değiştirebilir.

Örnek

$f_{X,Y}(x, y) = e^{-(x+y)}$, olsun ki $x, y \geq 0$. O zaman $f_X(x)$

$$f_X(x) = e^{-x} \int_0^{\infty} e^{-y} dy = e^{-x} \cdot 1 = e^{-x}$$

Örnek

$$f(x, y) = \begin{cases} x + y & \text{eğer } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{diğer} \end{cases}$$

$$f_Y(y) = \int_0^1 (x + y) dx = \int_0^1 x dx + \int_0^1 y dx = \frac{1}{2} + y \quad (1)$$

Tanim

İki rasgele degisken A, B bagimsizdir eger tum A, B degerleri icin

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

esitligi dogru ise. Bu durumda $X \perp Y$ yazilir.

Teori

X, Y 'nin birlesik PDF'i $f_{X,Y}$ olsun. O zaman ve sadece $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ ise $X \perp Y$ dogrudur.

Ornek

Diyelim ki X, Y bagimsiz, ve ikisinin de ayni yogunlugu var.

$$f(x) = \begin{cases} 2x & \text{eger } 0 \leq x \leq 1 \\ 0 & \text{digerleri} \end{cases}$$

$P(X + Y < 1)$ 'i hesaplayin.

Cevap

Bagimsizligi kullanarak birlesik dagilimi hesaplayabiliriz

$$f(x, y) = f_X(x)f_Y(y) = \begin{cases} 4xy & \text{eger } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{digerleri} \end{cases}$$

Simdi bu birlesik yogunluk uzerinden istedigimiz bolgeyi hesaplariz, bolgeyi tanımlayan $X + Y \leq 1$ ifadesi.

$$P(X + Y \leq 1) = \iint_{x+y \leq 1} f(x, y) dy dx$$

Entegralin limiti ustteki hali sembolik, hesap icin bu yeterli degil, eger $x + y \leq 1$ ise, $y \leq 1 - x$ demektir, ve bolge $y = 1 - x$ cizgisinin alti olarak kabul edilebilir. x, y zaten sifirdan buyuk olmalı, yani sola dogru yatık cizginin alti ve y, x eksenlerinin ustu kismini olusturan bir ucgen,

$$= \int_0^1 \int_0^{1-x} 4yx \, dy dx = 4 \int_0^1 x \left[\int_0^{1-x} y \, dy \right] dx$$

Numaraya dikkat, hangi degisken uzerinden integral aldigimiza bakarak, onun haricindekileri sabit kabul ederek bu “sabitleri” integral disina atiyoruz, boylece isimizi kolaylastiriyoruz. Hesabi tamamlarsak,

$$4 \int_0^1 x \frac{(1-x)^2}{2} dx = \frac{1}{6}$$

Kosullu Dagilimler (Conditional Distributions)

Surekli rasgele degiskenler icin kosullu olasilik yogunluk fonksiyonlari

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Devam edelim, eger kosullu yogunluk uzerinden olay hesabi yapmak istersek, ve $f_Y(y) > 0$ oldugunu farzederek,

$$P(X \in A|Y = y) = \int_A f_{X|Y}(x|y) dx$$

Ornek

(1) sonucunu aldigimiz ornege donelim,

$$f(x,y) = \begin{cases} x+y & \text{eger } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{diger} \end{cases}$$

$P(X < 1/4|Y = 1/3)$ nedir?

Cevap

Ustteki olasilik hesabi icin $f_{X|Y}$ fonksiyonuna ihtiyacimiz var. (1)'de gordugumu uzere,

$$f_Y(y) = \frac{1}{2} + y$$

Ana formulumuz neydi?

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

$$= \frac{x+y}{\frac{1}{2} + y}$$

$$P(X < 1/4|Y = 1/3) = \int_0^{1/4} \frac{x + \frac{1}{3}}{\frac{1}{2} + \frac{1}{3}} dx = \frac{\frac{1}{32} + \frac{1}{3}}{\frac{1}{2} + \frac{1}{3}} = \frac{14}{32}$$

Cok Degiskenli (Multivariate) Dagilimler ve IID Orneklemler (Samples)

$X = (X_1, \dots, X_n)$ olsun, ki (X_1, \dots, X_n) 'lerin herbiri bir rasgele degisken, o zaman X 'e rasgele vektor (random vector) ismi verilir. $f(x_1, \dots, x_n)$ 'in PDF'i temsil ettigini dusunelim. Bu PDF'i baz alarak aynen iki degiskenli (bivariate) orneklerde oldugu gibi, benzer tekniklerle bilesenleri, kosullu dagilimleri, vs. hesaplamak mumkundur.

Cok Degiskenli Normal

Tek degiskenli Normal dagilimin iki parametresi vardi, μ, σ . Cok degiskenli formda μ bir vektor, σ yerine ise Σ matrisi var. Once rasgele degiskeni tanımlayalım,

$$Z = \begin{bmatrix} Z_1 \\ \vdots \\ Z_k \end{bmatrix}$$

ki $Z_1, \dots, Z_k \sim N(0, 1)$. Z 'nin yogunlugu

$$\begin{aligned} f(z) &= \prod_{i=1}^k f(z_i) = \frac{1}{(2\pi)^{k/2}} \exp \left\{ -\frac{1}{2} \sum_{j=1}^k z_j^2 \right\} \\ &= \frac{1}{(2\pi)^{k/2}} \exp \left\{ -\frac{1}{2} z^T z \right\} \end{aligned}$$

Bu durumda Z 'nin *standart* çok degiskenli Normal dagilima sahip oldugu soylenir, ve $Z \sim N(0, I)$ olarak gosterilir. Buradaki 0 degeri icinde k tane sifir olan bir vektor olarak, I ise $k \times k$ birim (identity) matrisi olarak anlasilmalidir.

Daha genel olarak bir vektor X 'in çok degiskenli Normal dagilimina sahip oldugunu soyleriz, ve bunu $X \sim N(\mu, \Sigma)$ olarak gosteririz, eger dagilimin yogunlugu

$$f(x; \mu, \Sigma) = \frac{1}{(2\pi)^{k/2} \det(\Sigma)^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\}$$

Σ pozitif kesin (positive definite) bir matristir. Hatırlayalım, bir matris pozitif kesindir eger tum sifir olmayan x vektorleri icin $x^T \Sigma x > 0$ ise.

Not: Karekok kavrami tekil sayılardan matrislere de aktarilabilir. Bir matris B 'nin A 'nin karekoku oldugu soylenir, eger $B \cdot B = A$ ise.

Devam edersek, eger Σ pozitif kesin ise bir $\Sigma^{1/2}$ matrisini oldugu gosterilebilir, ki bu matrise Σ 'nin karekoku ismi verilir, ve bu karekokun su ozellikleri vardir, (i) $\Sigma^{1/2}$ simetriktir, (ii) $\Sigma = \Sigma^{1/2} \Sigma^{1/2} = I$ ve $\Sigma^{-1/2} = (\Sigma^{1/2})^{-1}$.

Hatirlama Numarasi

Normal Dagilimin formulu bazen hatırlayamayabiliriz. Peki daha basit bir formolden baslayarak onu turetebilir miyiz? Bu mumkun. Daha once e^{-x^2} Nasil

Entegre Edilir? yazisinda

$$\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$$

oldugunu gormustuk. Dikkat edersek bu integral bir formulun olasiliksal dagilim olup olmadigini kontrol etmek icin kullandigimiz integrale benziyor. Eger integral 1 cikarsa onun olasiliksal dagilim oldugunu biliyoruz. Ustteki sonuc $\sqrt{\pi}$, fakat iki tarafi $\sqrt{\pi}$ 'ye bolerseniz, sag taraf 1 olur ve boylece solda bir dagilim elde ederiz. Yani

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{\pi}} e^{-x^2} dx = 1$$

formulunde integralin sagindaki kisim bir dagilimdir. Bu formulu donusturerek Gaussian'a erisebiliriz. Ustteki formulun orta noktası (mean) sifir, varyansi (variance), yani $\sigma^2 = 1/2$ (bunu da ezberlemek lazim ama o kadar dert degil). O zaman $\sigma = 1/\sqrt{2}$.

Ilk amacimiz $\sigma = 1$ 'e erismek olsun (cunku oradan herhangi bir σ 'ya atlayabiliriz), bunun icin x 'i $\sqrt{2}$ 'ye bolmek lazim, tabii ayni anda onun etkisini sifirlamak icin normalize eden sabiti dengelemek amaciyla $\sqrt{2}$ 'ye bolmek lazim,

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{x}{\sqrt{2}}\right)^2} dx$$

$\sigma = 1$ 'e erisince oradan herhangi bir σ icin, σ degiskenine bolelim, yine hem e ustune hem sabite bu eki yapalim,

$$= \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{x}{\sqrt{2}\sigma}\right)^2} dx$$

Simdi herhangi bir ortalama μ icin bu degiskeni formule sokalim, bunun icin μ 'yu x 'den cikarmak yeterli

$$= \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)^2} dx$$

e ustundeki kare alma islemini acarsak,

$$= \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Boylece integral icindeki kisim tek boyutlu Gaussian formuna erismis oluyor.

Çebişev Eşitsizliği

Olasılık matematiğinde, büyük sayılar kuramı adında anılan ve olasılık matematiğinin belkemiğini oluşturan kuramı ispatlamak için, diğer bir kuram olan Çebişev eşitsizliğini de anlamamız gerekiyor. Çebişev eşitsizliği bir rasgele değişken, onun ortalaması (beklentisi) ve herhangi bir sabit sayı arasındaki üçlü arasında bir 'eşitsizlik' bağlantısı kurar, ve bu bağlantı diğer olasılık işlemlerimizde ispat verisi olarak işimize yarar.

Teori: Herhangi bir t değeri için,

$$P(|X - \mu| > t) \leq \frac{\sigma^2}{t^2}$$

İspata başlayalım. Entegral ile olasılık hesabı yapmak için bize bir x uzayı lazım.

$$R = x : |x - \mu| > t$$

Yani R uzayı, x ile ortalamasının farkının, t 'den büyük olduğu bütün sayıların kümesidir.

O zaman,

$$P(|X - \mu| > t) = \int_R f(x) dx$$

Dikkat edelim $P(\cdot)$ içindeki formül, küme tanımı ile aynı. O yüzden $P(\cdot)$ hesabı ortada daha olmayan, ama varolduğu kesin bir dağılım fonksiyonu tanımlamış da oluyor. Buna $f(x)$ deriz. $P(\cdot)$ 'in, $f(x)$ fonksiyonunun R üzerinden entegral olduğunu olasılığa giriş dersinden bilmemiz lazım.

Eğer $x \in R$ dersek o zaman

$$\frac{|x - \mu|^2}{t^2} \geq 1$$

t 'nin denkleme bu şekilde nereden geldiği şaşkınlık yaratabilir. Daha önce tanımlanan şu ibareye dikkat edelim, $x : |x - \mu| > t$ diye belirtmiştik. Bu ifadeyi değiştirerek, yukarıdaki denkleme gelebiliriz.

Devam edersek, elimizdeki 1'den büyük bir değer var. Bu değeri kullanarak, aşağıdaki tanımı yapmamız doğru olacaktır.

$$\int_R f(x) dx \leq \int_R \frac{(x - \mu)^2}{t^2} f(x) dx \leq \int_{-\infty}^{\infty} \frac{(x - \mu)^2}{t^2} f(x) dx$$

Ortadaki integral niye birinci integralden büyük? Çünkü ortadaki integraldeki $F(x)dx$ ibaresinden önce gelen kısmın, her zaman 1'den büyük olacağını belirttiğimize göre, ikinci integralin birinciden büyük olması normaldir.

Evet...Üçüncü integral ispata oldukça yaklaştı aslında. Standart sapma işaretini hala ortada göremiyoruz, fakat son integraldeki ibare standart sapma değerini zaten içeriyor. Önce daha önceki olasılık matematiği bilgimize dayanarak, standart sapmanın tanımını yazıyoruz. Dikkat edelim, bu ibare şu anki ispatımız dahilinden değil, haricinden önceki bilgimize dayanarak geldi. Standart sapmanın tanımı şöyledir.

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

O zaman

$$\frac{\sigma^2}{t^2} = \int_{-\infty}^{\infty} \frac{(x - \mu)^2}{t^2} f(x) dx$$

yani

$$\int_{\mathbb{R}} f(x) dx \leq \frac{\sigma^2}{t^2} = \int_{-\infty}^{\infty} \frac{(x - \mu)^2}{t^2} f(x) dx$$

ki $\int_{\mathbb{R}} f(x) dx$ zaten $P(|X - \mu| > t)$ olarak tanımlanmıştı.

Büyük Sayılar Kanunu

Olasılık kuramında önemli matematiksel bir denklem, büyük sayılar kanunudur. Bu kanun, tahmini olarak bildiğimiz günlük bir gerçeğin matematiksel ispatıdır da denebilir.

Yazı-tura atarken yazı çıkma ihtimalinin 1/2 olduğunu biliyoruz. Herhalde çoğumuz da bu yazı-tura işleminin "bir çok kere" tekrarlandığı durumda, toplam sonucun aşağı yukarı "yarısının" yazı olacağını tahmin biliyoruz. Bu tahminin matematiksel olarak söylemi, büyük sayılar kanunudur. Yıllarca önce Öklid'in geometriyi ispat ederek yaptığı gibi, matematiğe eklediğimizi her yeni bilgi dağarcığını önce matematiksel olarak ispatlamamız gerekiyor.

Farzedelim ki her yazı-tura atışı bir deney olsun. Her ayrı deneyin sonucu X_1, X_2, \dots, X_n olarak rasgelen değişkenlerle tanımlanmış olsun. Bu değişkenler ya 1 ya da 0 değeri taşıyacak, Yazı=1, Tura=0 olmak üzere.

Buna göre, n tane deneyden sonra elimize gelmesi gereken yazı oranı şudur.

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

Büyük sayılar kanunu, n büyüdükçe X_n 'in $1/2$ 'ye yaklaştığını ispatlar.

Başlayalım.

X_1, X_2, \dots, X_n bağımsız değişkenler olsun.

$$E(X_i) = \mu$$

$$\text{Var}(X_i) = \sigma$$

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

O zaman her $\epsilon > 0$ için ve $n \rightarrow \infty$, $p(|\bar{X}_n - \mu|) \rightarrow 0$.

Bu tanımlara göre, her rasgele değişkenin (deneyin) ortalaması aynı değerdir diyoruz. Bu zaten beklenir bir tanımdı, çünkü her rasgele değişkenin dağılımının aynı olduğunu kabul etmiştik. Her yazı tura aynı şartlar altında atılmazlar mı?

\bar{X}_n de bir rasgele değişkendir, çünkü Büyük sayılar kanununu, matematiksel olarak, \bar{X}_n değişkeninin ortalamasını tekil olarak her X_i dağılımının (aynı olan) ortalaması arasında birkü onun da formülü başka rasgele değişkenlere dayanıyor.

İspat devam etmek için, şapkalı X_n dağılımının beklentisini bulmamız gerekiyor.

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

$$E(\bar{X}_n) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right)$$

E doğrusal bir işlec (linear operator) olduğu için dışarıdan içeri doğru nüfuz eder.

$$= \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} n\mu$$

$$= \mu$$

\bar{X}_n dağılımının standart sapmasını da bulalım.

Diğer bir olasılık kuramına göre

$$Y = a + bX$$

$$\text{Var}(Y) = b^2 \text{Var}(X)$$

oldugunu biliyoruz. O zaman,

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

$$\text{Var}(\bar{X}_n) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i)$$

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n}$$

Artık Çebişev kuramını kullanmaya hazırız.

$n \rightarrow \infty$,

$$P(|\bar{X}_n - \mu| > \epsilon) \rightarrow 0$$

Cebisev'den

$$P(|\bar{X}_n - \mu| > \epsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\epsilon^2}$$

$$P(|\bar{X}_n - \mu| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \rightarrow 0$$

$\sigma^2/n\epsilon^2$ 'in sifira gitmesi normal cunku n sonsuza gidiyor.

Peki $P(|\bar{X}_n - \mu| > \epsilon)$ 'nin sifira gittigini gosterdik mi?

$\sigma^2/n\epsilon^2$ 'nin sifira gittigini gosterdik. $\sigma^2/n\epsilon^2$ de $P(|\bar{X}_n - \mu| > \epsilon)$ 'den buyuk olduguna gore, demek ki o da sifira iner.

Kaynaklar

[1] http://en.wikipedia.org/wiki/Confidence_interval

[2] Janert, P., Data Analysis with Open Source Tools

[3] Introduction to Probability Models, Sheldon Ross, 8th Edition, sf. 273