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

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 olcum , 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.

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

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

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

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

Rasgele Degiskenler (Random Variables)

Bir rasgele degisken X bir eslemedir, ki bu esleme $X : \Omega \to \mathbb{R}$ 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 $\omega = HHTHHTHHTT$ ise $X(\omega) = 6$. Tura sayisi eslemesi ω sonucunu 6 sayisina esledi.

Ornek

$\Omega = \{(x,y); x^2 + y^2 \leqslant 1\}$, yani kume birim cember ve icindeki reel sayilar (unit disc). Diyelim ki bu kumeden rasgele secim yapiyoruz. 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 $F_X : \mathbb{R} \to [0,1]$ tanimi

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

Eger X ayriksal ise, yani sayilabilir bir kume $\{x_1, x_2, ...\}$ 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 $a \leqslant b$ 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 entegral hesabi yapmamiz gereklidir. Ek olarak PDF 1'den buyuk olabilir, ama PMF olamaz. PDF'in 1'den buyuk olabilmesi entegrali bozmaz mi? Unutmayalim, entegral 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)

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

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.

Iki rasgele degisken X ve Y dagilimsal olarak birbirine esitligi, yani $X \overset{d}{=} Y$ eger $F_X(x) = F_Y(x)$, $\forall x$. Bu X, Y birbirine esit, birbirinin aynisi demek degildir. Bu degiskenler hakkindaki 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 $P(X = 0) = 1 - p$ olacaktir, ki $p \in [0,1]$ olmak uzere. O zaman X'in dagilimi Bernoulli deriz, ve $X \sim \mathrm{Bernoulli}(p)$ 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 vardir, ve onlardan ayri olarak parametreler vardir. Bu iki kavrami birbiriyle karistirmayalim.

Birbicim (Uniform) Dagilim

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

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

ise ve $a < b$ olacak sekilde. CDF hesabi olasilik egrisinin entegralini temel alir, birbicim 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.

Ileride gorecegiz ki μ bu dagilimin "ortasi", ve σ onun etrafa ne kadar "yayildigi" (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. Gelenege 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 bulunamamaktadir, bunun sebebi ise Normal PDF'in entegralinin analitik olarak alinamiyor olusudur.

Bazi faydali puf noktalari

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 $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2, ...$ ve her X_i digerlerinden bagimsiz ise, o zaman

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

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

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

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

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

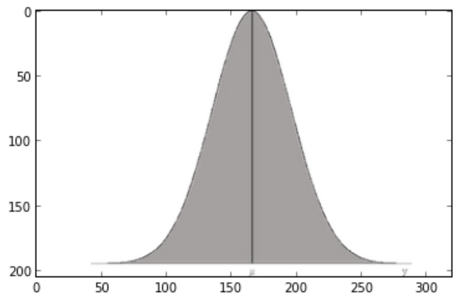
Ilk gecisi nasil elde ettik? Bir olasilik ifadesi $P(\cdot)$ icinde esitligin iki tarafina ayni anda ayni toplama, cikarma operasyonlarini yapabiliriz.

Son ifadenin anlami sudur. Eger standart Normal'in CDF'ini hesaplayabiliyorsak, istedigimiz Normal olasilik hesabini yapabiliriz demektir, cunku artik X iceren bir hesabin Z'ye nasil tercume edildigini goruyoruz.

Tum istatistik yazilimlari $\Phi(z)$ ve $\Phi(z)^{-1}$ hesabi icin gerekli rutinlere sahiptir. Tum istatistik kitaplarinda $\Phi(z)$'nin belli degerlerini tasiyan bir tablo vardir. Ders notlarimizin sonunda da benzer bir tabloyu bulabilirsiniz.

Dağılımlar Hakkında

Doğadan yapılan çoğu ölçümlerin, sıklık grafiğini alınca sonucun aşağıda gibi çıkmasi ilginctir.



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 ekseni üzerinde 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 cıkabilir: 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

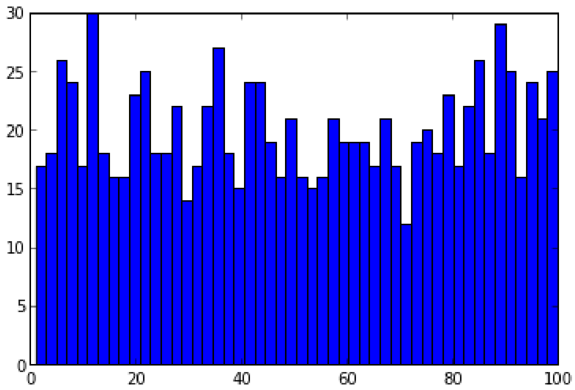
Eğer bu kavramları simulasyon ortamında göstermek istersek, Python ile bunu yapabiliriz.

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.

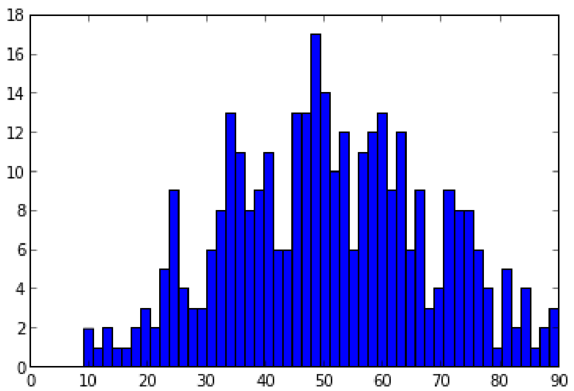
Gerçek rasgele sayı üretmek pek kolay bir iş değil. Her ne kadar bilgisayarınızda rasgele sayı üreten birçok algoritma olsa bile, bu algoritmalar belli bir sayı üretiminden sonra kendini tekrar etmeye başlıyorlar. Gerçek rasgele sayılar için muhakkak dış bir kaynağa bağlanmak gerekiyor.

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. Aşâğıda 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')
```



Ornek

$X \sim N(3,5)$ 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 oldugu icin yukaridaki form ortaya cikti.

Ornek

Simdi oyle 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 degerine tekabul eden esik degerini bulup, ustteki formul uzerinden geriye tercume 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 Dagilimi

X, ν derece bagimsizlikta t dagilimina sahiptir, ki bu $X \sim t_\nu$ diye yazilir eger

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

t dagilimi Normal dagilima benzer ama daha kuyrugu daha kalindir. Aslinda Normal dagilimi t dagiliminin $\nu = \infty$ oldugu hale tekabul eder. Cauchy dagilimi da t'nin ozel bir halidir, $\nu = 1$ halidir. Bu durumda yogunluk fonksiyonu

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

Bu formul hakikaten bir yogunluk mudur? Kontrol icin entegralini alalim,

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

Cogunlukla entegre edilen yerde "1 arti ya da eksi bir seyin karesi" turunde bir ifade gorulurse, yerine gecirme (subsitution) 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

$$= \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{dx}{1+x^2} = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{1}{\sec^2\theta}\sec^2\theta d\theta = \frac{1}{\pi}\int_{-\infty}^{\infty} 1\ d\theta =$$

$$= \frac{1}{\pi}\theta|_{-\infty}^{\infty} = \frac{1}{\pi}[\arctan(\infty) - \arctan(-\infty)]$$

$= \frac{1}{\pi}[\frac{\pi}{2} - (-\frac{\pi}{2})] = 1$

χ^2 Dagilimi

X'in p derece serbestlige sahip bir χ^2 dagilima sahip ise $X \sim \chi^2_p$ olarak gosterilir, yogunluk

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

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

Iki Degiskenli Dagilimlar

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 $F_{X,Y}(x, y) = P(X \leqslant x, Y \leqslant y)$ diye gosterilir.

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

Ornek

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

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

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

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

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

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

$$1 = \int \int f(x,y)dydx = c \int_{-1}^{1} \int_{x^2}^{1} x^2 y$$

$$= c \int_{-1}^{1} x^2 \int_{x^2}^{1} y dy dx = \int_{-1}^{1} x^2 (\frac{1}{2} - \frac{x^4}{2}) dx = 1$$

$$= c \int_{-1}^{1} x^2 (\frac{1-x^4}{2}) dx = 1$$

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

Devam edersek $c = 21/4$ buluruz.

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

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

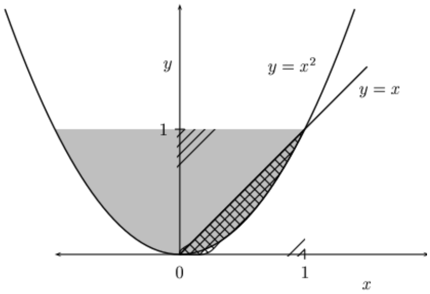
Simdi bunlara bir de $y \leqslant x$ eklememiz lazim. Yani ortadaki esitsizlige bir oge daha eklenir.

$-1 \leqslant x \leqslant 1$

$x^2 \leqslant y \leqslant x$

$y \leqslant 1$

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



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

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

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

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

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

"Hafizasiz" Dagilim, Ustel (Exponential) Dagilim

Ustel dagilimin hafizasiz oldugu soylenir. Bunun ne anlama geldigini anlatmaya ugrasalim. Diyelim ki rasgele degisken X bir aletin omrunu temsil ediyor, yani bir $p(x)$ fonksiyonuna bir zaman "sordugumuz" zaman bize dondurulen olasilik, o aletin x zamani kadar daha islemesinin olasiligi. Eger $p(2) = 0.2$ ise, aletin 2 yil daha yasamasinin olasiligi 0.2.

Bu hafizasizligi, olasilik matematigi ile nasil temsil ederiz?

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

Yani oyle bir dagilim var ki elimizde, $X > t$ bilgisi veriliyor, ama (kalan) zamani hala $P(X > s)$ olasiligi veriyor. Yani t kadar zaman gectigi bilgisi hicbir seyi degistirmiyor. Ne kadar zaman gecmis olursa olsun, direk s ile gidip ayni olasilik hesabini yapiyoruz.

Sartsal (conditional) formulunu uygularsak ustteki soyle 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 olmasi icin X ne sekilde dagilmis olmalidir? Ustteki denklem sadece X dagilim fonksiyonu ustel (exponential) olursa mumkundur, cunku sadece o zaman

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

gibi bir iliski kurulabilir.

Ornek

Diyelim ki bir bankadaki bekleme zamani ortalam 10 dakika ve ustel olarak dagilmis. Bir musterinin i) bu bankada 15 dakika beklemesinin ihtimali nedir? ii) Bu musterinin 10 dakika bekledikten sonra toplam olarak 15 dakika (ya da daha fazla) beklemesinin olasiligi nedir?

Cevap

i) Eger X musterinin bankada bekledigi zamani temsil ediyorsa

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

ii) Sorunun bu kismi musteri 10 dakika gecirdikten sonra 5 dakika daha gecirmesinin olasiligini soruyor. Fakat ustel dagilim "hafizasiz" oldugu icin kalan zamani alip yine direk ayni fonksiyona geciyoruz,

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

Bilesen (Marginal) Dagilimlar

Surekli rasgele degiskenler icin bilesen yogunluk

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

ve

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

Ustteki integraller gercek bir dagilim fonksiyonu $f(x,y)$ verilince alt ve ust limit te tanimlamak zorundadir. Cunku bilesen yogunluk icin bir veya daha fazla degiskeni "integralle disari atmak (integrate out)" ettigimiz soylenir, eger ayriksal (discrete) ortamda olsaydik bu atilan degiskenin tum degerlerini goze alarak toplama yapan bir formul yazardik. Surekli ortamda integral kullaniyoruz, ama tum degerlerin uzerinden yine bir sekilde gecmemiz gerekiyor. Iste alt ve ust limitler bunu gerceklestiriyor. Bu alt ve ust limitler, atilan degiskenin "tum degerlerine" bakmasi gerektigi icin $-\infty, +\infty$ olmalidir. Eger problem icinde degiskenin belli degerler arasinda oldugu belirtilmis ise (mesela alttaki ornekte $x > 0$) o zaman entegral limitleri alt ve ust sinirini buna gore degistirebilir.

Ornek

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

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

Ornek

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

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

Tanim

Iki 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 $X \amalg Y$ 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

$X \amalg Y$ dogrudur.

Ornek

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

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

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

Cevap

Bagimsizligi kullanarak birlesik dagilimi hesaplayabiliriz

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

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

$$P(X + Y \leqslant 1) = \int\int_{x+y \leqslant 1} f(x,y) dy dx$$

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

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

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

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

Kosullu Dagilimlar (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,

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

$P(X < 1/4 | Y = 1/3)$ 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) Dagilimlar ve IID Orneklemler (Samples)

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

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

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

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

$$= \frac{1}{(2\pi)^{k/2}} \exp\left\{ -\frac{1}{2} z^T z \right\}$$

Bu durumda Z'nin *standart* cok 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 cok degiskenli Normal dagilimina sahip oldugunu soyleriz, ve bunu $X \sim N(\mu, \Sigma)$ olarak gosteririz, eger dagilimin yogunlugu

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

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

Not: Karekok kavrami tekil sayilardan 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 $\Sigma^{-1/2} = (\Sigma^{1/2})^{-1}$.

Cok Boyutlu Gaussian'i Parcalamak (Partitioning)

Diyelim ki Normal bir vektor X'i $X = (X_1, X_2)$ olarak parcaladik. Bunu Gaussian'a etkileri ne olur? Ayni sekilde $\mu = (\mu_1, \mu_2)$ olarak parcalayabiliriz. Σ ise

$$\Sigma = \left[\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array} \right]$$

olarak parcalanabilir. a, b'nin parcalarinin boyutlari p, q olsun, $n = p + q$.

Simdi birlesik Gaussian'i

$$f(x; \mu, \Sigma) = \frac{1}{(2\pi)^{(p+q)/2} \det(\Sigma)^{1/2}} \exp \bigg\{ -\frac{1}{2} \left[\begin{array}{c} x_1 - \mu_1 \\ x_2 - \mu_2 \end{array} \right]^T \left[\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array} \right]^{-1} \left[\begin{array}{c} x_1 - \mu_1 \\ x_2 - \mu_2 \end{array} \right] \bigg\}$$

Birlesik yogunlugu parcalar uzerinden belirtirsek, bu yogunlugu X_2 icin bilesen yogunluga ve X_1 icin bir kosullu yogunluga ayirabiliriz. Yani

$$f(x_1, x_2) = f(x_1 | x_2) f(x_2)$$

tanimindaki parcalari elde etmeye calisacagiz. Ama bundan once boluntulenmis matrislere yakindan bakalim.

Bir boluntulenmis (partitioned) matrisin tersini almak icin, o matrisin parcalarinin tersini almak dogru degildir, yani

$$\left[\begin{array}{cc} E & F \\ G & H \end{array} \right]^{-1} \neq \left[\begin{array}{cc} E^{-1} & F^{-1} \\ G^{-1} & H^{-1} \end{array} \right]$$

Tersini alma islemi icin bazi numaralar lazim. Ana numara boluntulenmis matrisi kosegen bir matris haline getirmek, cunku kosegen matrislerin tersi, kosegendeki elemanlarin tersidir, yani ters alma operasyonu bu tur matrislerin "icine isler", o yuzden bir sekilde bir kosegen matris elde etmeye ugrasacagiz. Bunun icin boluntulenmis matrisimizi sagdan ve soldan bazi matrislerle carpacagiz. Ayrica sunu da bilelim,

$XYZ = W$

durumunda Y'nin tersini almak istersek, sag ve soldaki X, Z matrislerinin tersini almak gerekmez, niye?

$X^{-1}XYZ = X^{-1}W$

$YZZ^{-1} = X^{-1}WZ^{-1}$

$Y = X^{-1}WZ^{-1}$

Simdi iki tarafin da tersini alalim,

$Y^{-1} = ZW^{-1}X$

Tamam, baslayalim.

$$M = \begin{bmatrix} E & F \\ G & H \end{bmatrix}$$

matrisini kosegen yapacagiz. Eger sadece alt sol koseyi sifirlayasaydik, bunu yapacak ozel bir matrisle soldan carpardik,

$$\begin{bmatrix} I & -FH^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} E & F \\ G & H \end{bmatrix} = \begin{bmatrix} E & F \\ 0 & H \end{bmatrix}$$

Sadece ust sag koseyi sifirlamak isteseydik, sagdan carpardik

$$\begin{bmatrix} E & F \\ G & H \end{bmatrix} \begin{bmatrix} I & 0 \\ -H^{-1}G & I \end{bmatrix} = \begin{bmatrix} E & 0 \\ G & H \end{bmatrix}$$

Hepsini biraraya koyalim,

$$\begin{bmatrix} I & -FH^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} E & F \\ G & H \end{bmatrix} \begin{bmatrix} I & 0 \\ -H^{-1}G & I \end{bmatrix} = \begin{bmatrix} E - FH^{-1}G & 0 \\ 0 & H \end{bmatrix}$$

Bu carpimin dogrulugu carpim elle yapilarak kontrol edilebilir.

Ustte gordugumuz gibi

$XYZ = W$

ifadesindeki Y'nin tersi

$Y^{-1} = ZW^{-1}X$

ile olur.

$$\underbrace{\begin{bmatrix} I & -FH^{-1} \\ 0 & I \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} E & F \\ G & H \end{bmatrix}}_{Y} \underbrace{\begin{bmatrix} I & 0 \\ -H^{-1}G & I \end{bmatrix}}_{Z} = \underbrace{\begin{bmatrix} E - FH^{-1}G & 0 \\ 0 & H \end{bmatrix}}_{W}$$

O zaman

$$\left[\begin{array}{cc} E & F \\ G & H \end{array}\right]^{-1} = \left[\begin{array}{cc} I & 0 \\ -H^{-1}G & I \end{array}\right] \left[\begin{array}{cc} E-FH^{-1}G & 0 \\ 0 & H \end{array}\right]^{-1} \left[\begin{array}{cc} I & -FH^{-1} \\ 0 & I \end{array}\right]$$

Daha kisa olmasi esitligin sag tarafinda, ortadaki matris icin $E - FH^{-1}G$ yerine M/H kullanalim (bu arada M/H lineer cebirde "M'in H'e gore Schur tamamlayicisi (complement)" olarak bilinir),

$$\left[\begin{array}{cc} E & F \\ G & H \end{array}\right]^{-1} = \left[\begin{array}{cc} I & 0 \\ -H^{-1}G & I \end{array}\right] \left[\begin{array}{cc} (M/H)^{-1} & 0 \\ 0 & H^{-1} \end{array}\right] \left[\begin{array}{cc} I & -FH^{-1} \\ 0 & I \end{array}\right]$$

Esitligin sag tarafindaki carpimi gerceklestirirsek,

$$= \left[\begin{array}{cc} (M/H)^{-1} & -(M/H)^{-1}FH^{-1} \\ -H^{-1}G(M/H)^{-1} & H^{-1} + H^{-1}G(M/H)^{-1}FH^{-1} \end{array}\right]$$

Bu final ifade boluntulenmis bir matrisin tersini o matrisin icindeki parcalar uzerinden temsil eden bir ifadedir.

Icinde bir kosesi sifir olan boluntulenmis matrislerde determinantlar soyle isler,

$$\det\left(\left[\begin{array}{cc} E & 0 \\ G & H \end{array}\right]\right) = \det\left(\left[\begin{array}{cc} E & F \\ 0 & H \end{array}\right]\right) = \det(E)\det(H)$$

Ayrica

$$\det(AB) = \det(A)\det(B)$$

O zaman (2)'nin determinantini alirsak, det yerine || kullandik,

$$|M| = |M/H||H|$$

Bu ifade gayet dogal duruyor (bir raslanti herhalde, ya da Schur tamamlayicisi isareti ozellikle boyle secilmis),

Boluntulenmis bir matrisin devrigini almak icin her blogunun ayri ayri devrigi alinir, ve tum bloklarin yani boluntulenmis tamaminin bir daha devrigi alinir, yani

$$\left[\begin{array}{cc} A & B \\ C & D \end{array}\right]^{T} = \left[\begin{array}{cc} A^T & C^T \\ B^T & D^T \end{array}\right]$$

Simdi cok degiskenli Normal icin bilesen ve kosullu yogunluk hesaplarina gelelim. Gaussian formulunun exp kismini alirsak,

$$\exp\left\{-\frac{1}{2}\left[\begin{array}{c} x_1-\mu_1 \\ x_2-\mu_2 \end{array}\right]^T \left[\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array}\right]^{-1} \left[\begin{array}{c} x_1-\mu_1 \\ x_2-\mu_2 \end{array}\right]\right\}$$

(3)'teki acilimi kullanirsak, ve $E = \Sigma_{11}, F = \Sigma_{12}, ..$ olacak sekilde,

$$\exp\left\{-\frac{1}{2}\left[\begin{array}{c} x_1-\mu_1 \\ x_2-\mu_2 \end{array}\right]^T \left[\begin{array}{cc} I & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & I \end{array}\right] \left[\begin{array}{cc} (\Sigma/\Sigma_{22}) & 0 \\ 0 & \Sigma_{22}^{-1} \end{array}\right] \left[\begin{array}{cc} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{array}\right] \left[\begin{array}{c} x_1-\mu_1 \\ x_2-\mu_2 \end{array}\right]\right\}$$

Acilimi tamamen yaparsak,

$$
\begin{aligned}
= \quad & \exp\left\{ -\tfrac{1}{2}(x_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2))^T(\Sigma/\Sigma_{22})^{-1}(x_1-\mu_1-\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2)) \right\} \cdot \\
& \exp\left\{ 1\tfrac{1}{2}(x_2-\mu_2)^T\Sigma_{22}^{-1}(x_2-\mu_2) \right\}
\end{aligned}
$$

Not: $\Sigma_{12}^T = \Sigma_{21}$. Ustte birinci exp icinde sol bolumde devrigin icindeki ifadelerden, mesela x_1^T, μ_1^T'den ve Σ_{21}'li ifadeden devrik islemini cekip, buyuk paranteze alininca bu degisim oldu.

Simdi mesela 1. exp'ye dikkat edersek, ortada $(\Sigma/\Sigma_{22})^{-1}$ var, ve bu ifadenin solunda ve saginda birbirinin devrigi olan ayni terimler duruyor. Ifadenin tamami bir Normal dagilim. Ayni sey 2. exp icin gecerli.

Isin exp tarafini halletik. Simdi exp oncesindeki kesiri (4) kullanarak parcalayalim,

$$
\frac{1}{(2\pi)^{(p+q)/2}\det(\Sigma)^{1/2}} = \frac{1}{(2\pi)^{(p+q)/2}\bigg(\det(\Sigma/\Sigma_{22})\det(\Sigma_{22})\bigg)^{1/2}}
$$

$$
= \left(\frac{1}{(2\pi)^{p/2}\det(\Sigma/\Sigma_{22})^{1/2}} \right) \left(\frac{1}{(2\pi)^{q/2}\det(\Sigma_{22})^{1/2}} \right)
$$

Bu parcalarin her birini ayri bir exp onunde kullanabiliriz, ve ikinci exp ifadesinin

$$
\frac{1}{(2\pi)^{q/2}\det(\Sigma_{22})^{1/2}} \exp\left\{ \frac{1}{2}(x_2-\mu_2)^T\Sigma_{22}^{-1}(x_2-\mu_2) \right\}
$$

oldugunu goruyoruz. Bu ifade $f(x_2)$ bilesen yogunlugudur! O zaman geri kalanlar, yani diger kesir ve birinci exp hep beraber $f(x_1|x_2)$ yogunlugu olmalidir. Yani,

$$
\frac{1}{(2\pi)^{p/2}\det(\Sigma/\Sigma_{22})^{1/2}} \cdot
$$

$$
\exp\left\{ -\frac{1}{2}(x_1-\mu_1-\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2))^T(\Sigma/\Sigma_{22})^{-1}(x_1-\mu_1-\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2)) \right\}
$$

Buradan genel bir kural cikartabiliriz,

1) X_2'nin bilesen yogunlugu $X_2 \sim N(\mu_2, \Sigma_{22})$

2) $X_2 = x_2$ olmak uzere X_1'in kosullu dagilimi

$$
X_1|X_2 = x_2 \sim N\bigg(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2) \ , \ \Sigma/\Sigma_{22} \bigg)
$$

Σ/Σ_{22} nedir? Hatirlarsak, $M/H = E - FH^{-1}G$, ve $E = \Sigma_{11}, F = \Sigma_{12}, ..$ o zaman

$\Sigma/\Sigma_{22} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$

Yani

$$X_1|X_2 = x_2 \sim N\bigg(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2)\ ,\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\bigg)$$

Hatirlama Numarasi

Normal Dagilimin formulunu bazen hatirlayamayabiliriz. Peki daha basit bir formulden 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 bolersek, 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 noktasi (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 $\sqrt{2}$'e bolmek lazim, tabii ayni anda onun etkisini sifirlamak icin normalize eden sabiti dengelemek amaciyla $\sqrt{2}$'ye bolmek lazim,

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

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

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

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

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

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.

Beklenti (Expectation)

Bu deger, dagilim $f(x)$'in tek sayilik bir ozetidir. Yani beklenti hesabina bir taraftan bir dagilim fonksiyonu girer, diger taraftan tek bir sayi disari cikar. Surekli dagilim fonksiyonlari icin $E(X)$

$$E(X) = \int xf(x)dx$$

ayriksal durumda

$$E(X) = \sum_x xf(x)$$

olarak hesaplanir. Hesabin, her x degerini onun olasiligi ile carpip topladigina dikkat. Bu tur bir hesap dogal olarak tum x'lerin ortalamasini verecektir, ve dolayli olarak dagilimin ortalamasini hesaplayacaktir. Ortalama μ_x olarak ta gosterilebilir.

Notasyonel basitlik icin ustteki toplam / entegral yerine

$$= \int x \ dF(x)$$

diyecegiz, bu notasyonel bir kullanim sadece, unutmayalim, reel analizde $\int x \ dF(x)$'in ozel bir anlami var (hoca tam diferansiyel dF'den bahsediyor).

Beklentinin taniminin kapsamli / eksiksiz olmasi icin $E(X)$'in "mevcudiyeti" icin de bir sart tanimlamak gerekir, bu sart soyle olsun,

$$\int_x |x| dF_X(x) < \infty$$

ise beklenti mevcut demektir. Tersi sozkonusu ise beklenti mevcut degildir.

Ornek

$X \sim Unif(-1, 3)$ olsun. $E(X) = \int xdF(x) = \int xf_X(x)dx = \frac{1}{4}\int_{-1}^{3} xdx = 1$.

Ornek

Cauchy dagiliminin $f_X(x) = \{\pi(1+x^2)\}^{-1}$ oldugunu soylemistik. Simdi beklentiyi hesaplayalim. Parcali entegral teknigi lazim, $u = x$, $dv = 1/1 + x^2$ deriz, ve o zaman $v = \tan^{-1}(x)$ olur, bkz *Ters Trigonometrik Formuller* yazimiz. Demek ki

$$\int |x| dF(x) = \frac{2}{\pi}\int_0^{\infty} \frac{xdx}{1+x^2}$$

2 nereden cikti? Cunku $|x|$ kullaniyoruz, o zaman sinir degerlerinde sadece sifirin sagina bakip sonucu ikiyle carpmak yeterli. Bir sabit oldugu icin π ile beraber disari cikiyor. Simdi

$$\int u dv = uv - \int v du$$

uzerinden

$$= [x \tan^{-1}(x)]_0^\infty - \int_0^\infty \tan^{-1}(x) dx = \infty$$

Yani ustteki hesap sonsuzluga gider. O zaman ustteki tanimimiza gore Cauchy dagiliminin beklentisi yoktur.

Toplamların Moment'leri

Olasılık matematiğinde "moment üreten işlevler" olarak adlandırılan, başlangıçta pek yararlı gibi gözükmesede bir takım matematiksel özellikleri olduğu için, ispatlarda oldukça işe yarayan bir kavram vardır.

Her rasgele değişkenin bir dağılımı olduğunu biliyoruz. Her rasgele değişkenin de ayrıca bir moment üreten fonksiyonu da vardır. Ayrıca, moment üreten fonksiyon ile rasgele değişken arasında bire-bir olarak bir ilişki mevcuttur. "Bu neye yarar?" diye sorulabilir; Cevap olarak, mesela cebirsel olarak türete türete bir moment'e geldiğimiz düşünelim, ve tekrar başka bir taraftan, başka bir formülden gene türete türete tekrar aynı moment işlevine geliyorsak, bu demektir ki, iki taraftan gelen rasgele değişkenler (ve tekabül eden dağılımları) birbirine eşittir. Bazı şartlarda moment üreten işlevler ile cebir yapmak, dağılım fonksiyonlarından daha rahat olmaktadır.

Her rasgele değişken için, moment üreten işlev şöyle bulunur.

X rasgele degiskenin moment ureten operasyonu

$M(t) = E(e^{tX})$ olarak gosterilir

Ayriksal operasyonlar icin

$$M(t) = \sum_x e^{tx} p(x)$$

Surekli islevler icin

$$M(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

Kuram

Gelelim yazımızın esas konusu olan kuramımıza.

Eğer $X_1, X_2 ... X_n$ bağımsız rasgele değişken ise, ve her değişkenin $M_i(t)$ $i = 1, 2, 3, ...n$

olarak, öz olarak aynı olan birer moment üreten işlevi var ise, o zaman,

$$Y = \sum_{i=1}^{n} aX_i$$

acilimi

$$M_y(t) = \prod_{i=1}^{n} M(a_i t)$$

olacaktir.

Ispat

$M_y(t) = E(e^{tY} = E(e^{t(a_1X_1+a_2X_2+..+a_nX_n)}$

$= E[\exp(ta_1X_1ta_2X_2... + ta_nX_n)]$

$= E[\exp(ta_1X_1) + \exp(ta_2X_2) + ... + \exp(ta_nX_n)]$

$= E[\exp(ta_1X_1)] + E[\exp(ta_2X_2)] + ... + E[\exp(ta_nX_n)]$

Daha once belirttigimiz gibi

$M_i(t) = E[\exp(tX_i)]$

olduguna gore ve t yerine ta_i koyuldugunu dusunelim

$$M_y(t) = \prod_{i=1}^{n} M_y(a_i t)$$

olacaktir.

Bunu $M_y(t) = (M_i(a_i t))^n$ seklinde de gosterebiliriz. Ortalama (Mean) ve Medyan (Median)

Ozet Istatistikleri

Genellikle istatistik kitaplari hemen ortalama (mean), medyan (median) ve baglantili ozet istatistiklerinden (summary statistics) bahsederek ise girerler. Bu istatistikleri dikkatle kullanmak gerekir, cunku her turlu veri, her yerde gecerli degildirler. Mesela ortalama sadece tek merkezi bir tepesi olan (unimodal) dagilimlar icin gecerlidir. Eger bu temel varsayim gecerli degilse, ortalama kullanarak yapilan hesaplar bizi yanlis yollara goturur. Ayrica bir dagilimi simetrik olup olmadigi da ortalama ya da medyan kullanilip kullanilmamasi kararinda onemlidir. Eger simetrik, tek tepeli bir dagilim var ise, ortalama ve medyan birbirine yakin olacaktir. Fakat veri baska turde bir dagilim ise, o zaman bu iki olcut birbirinden cok farkli olabilir.

Once ortalama ve standart sapmayi (standart deviation) gorelim.

$$m = \frac{1}{n} \sum_i x_i$$

Standart sapma veri noktalarin "ortalamadan farkinin ortalamasini" verir. Tabii bazen noktalar ortalamanin altinda, bazen ustunde olacaktir, bizi bu negatiflik, pozitiflik ilgilendirmez, biz sadece farkla alakaliyiz. O yuzden her sapmanin karesini aliriz, bunlari toplayip nokta sayisina boleriz .

$$s^2 = \frac{1}{n} \sum_i (x_i - m)^2$$

Eger m tanimini ustte yerine koyarsak,

$$= \frac{1}{n} \sum_i x_i^2 + \frac{1}{n} \sum_i m^2 - \frac{2}{n} \sum_i x_i m$$

$$= \frac{1}{n} \sum_i x_i^2 + \frac{m^2 n}{n} - \frac{2mn}{n} m$$

$$= \frac{1}{n} \sum_i x_i^2 + m^2 - 2m^2$$

$$= \frac{1}{n} \sum_i x_i^2 - m^2$$

Bu olcuye varyans (variance) denir ve teorik olarak ortalamadan daha onemli oldugu soylenebilir. Fakat dagilimin yayilma olcusu olarak biz bu olcuyu oldugu gibi degil, onun karesini kullanacagiz (ki standart sapma buna deniyor aslinda). Niye? Cunku o zaman veri noktalarinin ve yayilma olcusunun birimleri birbiri ile ayni olacak. Eger veri setimiz bir alisveris sepetindeki malzemelerin lira cinsinden degerleri olsaydi, varyans bize sonucu "karekok lira" olarak verecekti ve bunun pek anlami olmayacakti.

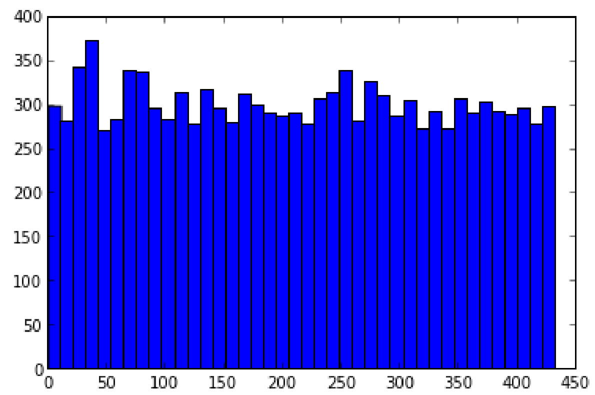
Medyan ve Yuzdelikler (Percentile)

Ustteki hesaplar sayilari toplayip, bolmek uzerinden yapildi. Medyan ve diger yuzdeliklerin hesabi (ki medyan 50. yuzdelige tekabul eder) icin eldeki tum degerleri "siraya dizmemiz" ve sonra 50. yuzdelik icin ortadakine bakmamiz gerekiyor. Mesela eger ilk 5. yuzdeligi ariyorsak ve elimizde 80 tane deger var ise, bastan 4. sayiya / vektor hucresine / ogeye bakmamiz gerekiyor. Eger 100 eleman var ise, 5. sayiya bakmamiz gerekiyor, vs.

Bu siraya dizme islemi kritik. Kiyasla ortalama hesabi hangi sirada olursa olsun, sayilari birbirine topluyor ve sonra boluyor. Zaten ortalama ve sapmanin istatistikte daha cok kullanilmasinin tarihi sebebi de aslinda bu; bilgisayar oncesi cagda sayilari siralamak (sorting) zor bir isti. Bu sebeple hangi sirada olursa olsun, toplayip, bolerek hesaplanabilecek ozetler daha makbuldu. Fakat artik siralama islemi kolay, ve veri setleri her zaman tek tepeli, simetrik olmayabiliyor.

Ornek veri seti olarak unlu `dellstore2` tabanindaki satis miktarlari kullanirsak,

```
data = np.loadtxt("dell.csv")
plt.hist(data,40)
plt.savefig('05_02.png')
```




```
print np.mean(data)
```

```
213.948899167
```

```
print np.median(data)
```

```
214.06
```

```
print np.std(data)
```

```
125.118481954
```

```
print np.mean(data)+2*np.std(data)
```

```
464.185863074
```

```
print np.percentile(data, 95)
```

```
410.4115
```

Goruldugu gibi uc nokta hesabi icin ortalamadan iki sapma otesini kullanirsak, 464.18, fakat 95. yuzdeligi kullanirsak 410.41 elde ediyoruz. Niye? Sebep ortalamanin kendisi hesaplanirken cok uc degerlerin toplama dahil edilmis olmasi ve bu durum, ortalamanin kendisini daha buyuk seviyeye dogru itiyor. Yuzdelik hesabi ise sadece sayilari siralayip belli bazi elemanlari otomatik olarak uc nokta olarak addediyor.

Box Whisker Grafikleri

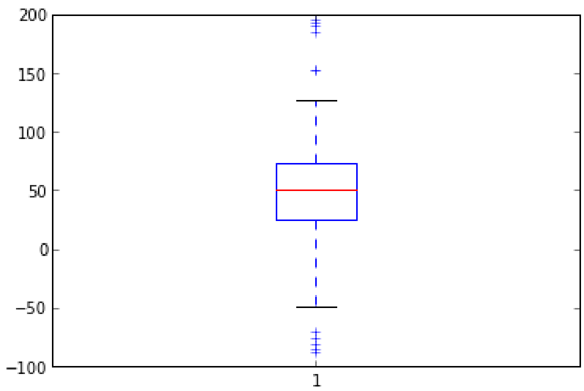
Tek boyutlu bir verinin dagilimini gormek icin Box ve Whisker grafikleri faydali araclardir; medyan (median), dagilimin genisligini ve siradisi noktalari (outliers) acik sekilde gosterirler. Isim nereden geliyor? Box yani kutu, dagilimin agirliginin nerede oldugunu gosterir, medyanin sagindada ve solunda olmak uzere iki ceyregin arasindaki kisimdir, kutu olarak resmedilir. Whiskers kedilerin biyiklarina verilen isimdir, zaten grafikte birazcik biyik gibi duruyorlar. Bu uzantilar medyan noktasindan her iki yana kutunun iki kati kadar uzatilir sonra verideki "ondan az olan en buyuk" noktaya kadar geri cekilir. Tum bunlarin disinda kalan veri ise teker teker nokta olarak grafikte basilir. Bunlar siradisi (outlier) olduklari icin daha az olacaklari tahmin edilir.

BW grafikleri iki veriyi dagilimsal olarak karsilastirmak icin birebirdir. Mesela Larsen and Marx adli arastirmacilar cok az veri iceren Quintus Curtius Snodgrass veri setinin degisik oldugunu ispatlamak icin bir suru hesap yapmislardir, bir suru matematiksel isleme girmislerdir, fakat basit bir BW grafigi iki setin farkliligini hemen gosterir.

BW grafikleri iki veriyi dagilimsal olarak karsilastirmak icin birebirdir. Mesela Larsen and Marx adli arastirmacilar cok az veri iceren Quintus Curtius Snodgrass veri setinin degisik oldugunu ispatlamak icin bir suru hesap yapmislardir, bir suru matematiksel isleme girmislerdir, fakat basit bir BW grafigi iki setin farkliligini hemen gosterir.

Python uzerinde basit bir BW grafigi

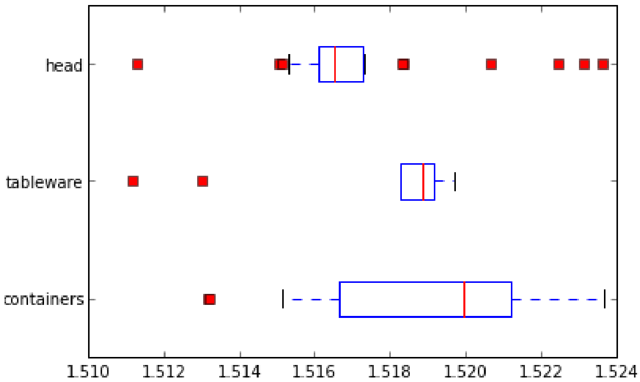
```
spread= rand(50) * 100
center = ones(25) * 50
flier_high = rand(10) * 100 + 100
flier_low = rand(10) * -100
data =concatenate((spread, center, flier_high, flier_low), 0)
plt.boxplot(data)
plt.savefig('05_03.png')
```



Bir diger ornek Glass veri seti uzerinde

```
data = loadtxt("glass.data",delimiter=",")
head = data[data[:,10]==7]
tableware = data[data[:,10]==6]
containers = data[data[:,10]==5]

print head[:,1]

data =(containers[:,1], tableware[:,1], head[:,1])

plt.yticks([1, 2, 3], ['containers', 'tableware', 'head'])

plt.boxplot(data,0,'rs',0,0.75)
plt.savefig('05_04.png')
```

```
[ 1.51131  1.51838  1.52315  1.52247  1.52365  1.51613  1.51602  1.51623
  1.51719  1.51683  1.51545  1.51556  1.51727  1.51531  1.51609  1.51508
  1.51653  1.51514  1.51658  1.51617  1.51732  1.51645  1.51831  1.5164
  1.51623  1.51685  1.52065  1.51651  1.51711]
```



Guven Araligi (Confidence Intervals)

Bu kavram istatistikte tartisilan konulardan biri. Bayes ve Frenkansçı (Frequentist) istatistik arasindaki felsefi farklardan biri burada ortaya cikiyor. Frekansci tanim soyledir:

"Bir parametre θ icin $1-\alpha$ seviyesinde bir $C_n = (a, b)$ guven araligi tanimlanabilir – bu aralik $a = a(X_1, .., X_n)$ ve $b = b(X_1, .., X_n)$ adli iki fonksiyon uzerinden tanimlanabilir. Bu fonksiyonlar veri uzerinde isleyen, *verinin* fonksiyonlaridir, ve sonucta

$$\mathbb{P}_\theta(\theta \in C_n) \geqslant 1 - \alpha, \ \ \forall \theta \in \Theta$$

Yani (a, b) araligi $1 - \alpha$ olasiliginda θ 'yi icine alir / hapseder.

Daha detayli olarak deney arka arkaya pek cok kez tekrarlandiginda parametrenin tahmininin $1 - \alpha$ oraninda tanimlanan araliga dusecegi soylenir. $1 - \alpha$ sayisina guven araliginin kapsami (coverage) ismi de verilir. Genellikle insanlar yuzde 95 guven araligini kullanirlar, ve bu yuzdeye tekabul eden $\alpha = 0.05$

rakami kullanilir.

Uzerine basarak belirtmek gerekir ki C_n rasgele (random) bir degerdir, ama θ sabittir, cunku C_n verinin bir fonksiyonudur, ve veriden, yani bir orneklemden gelecegi icin o da rasgele olmalidir.

Eger θ bir vektor ise o zaman bir aralik yerine bir guven kumesi kullanilir (mesela bir kure, ya da elips)."

Fakat frekansci yaklasimda aralik fonksiyonlari a, b ile guven araligi arasindaki baglanti net degildir. Hangi fonksiyon secimi hangi α'ya sebebiye vermektedir? Bu durum net oldugu durumlarda bile teorik olarak saglamligi suphelidir, ayrica hesabin sozel olarak ortaya konmasinda bazi eksikler vardir. "Deney arka arkaya pek cok kez tekrarlandiginda parametrenin tahmini, $1-\alpha$ guven araliga dusecektir" ibaresi mesela; "deney tekrari" her durumda gecerli olmayabilir. Meteoroloji "yarin yuzde 80 ihtimali ile yagmur yagacak" diyorsa, o hesap sartlarinin bir daha ortaya cikmasinin olasiligi cok dusuktur, Kaos Teorisi bize en azindan bunu soyluyor.

Wiki sayfasinda [1] tartismanin boyutlari gorulebilir.

Son onyillarda ortaya cikan yaklasim ise Bayes Teorisini devreye sokmak. Bir guven araligi tanimlamanin en saglam yolu bu hesabi bir dagilimi baz alarak yapmak. Eger sonuc olarak bir tekil sayi degil, bir dagilim elde edersek bu dagilim uzerinde guvenlik hesaplarini yapmak cok kolay hale gelir. Mesela sonuc (sonsal dagilim) bir Gaussian dagilim ise, bu dagilimin yuzde 95 agirliginin nerede oldugu, ve nasil hesaplandigi bellidir.

Bayes Teorisi

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Veri analizi baglaminda diyelim ki deneyler yaparak tahmini olarak hesaplamak (estimate) istedigimiz bir parametre var, bu bir protonun kutlesi ya da bir ameliyat sonrasi hayatta kalma orani olabilir. Bu durumlarda iki ayri "olaydan" bahsetmemiz gerekir, B olayi spesifik bazi olcumlerin elde edilmesi "olayidir", mesela olcum uc sayidan olusuyorsa, biz bir olcumde spesifik olarak $\{0.2, 4, 5.4\}$ degerlerini elde etmisiz. Ikinci olay bilmedigimiz parametrenin belli bir degere sahip olmasi olacak. O zaman Bayes Teorisinin su sekilde tekrar yazabiliriz,

$$P(parametre|veri) \propto P(data|parametre)P(parametre)$$

$\propto$ isareti orantili olmak (proportional to) anlamina geliyor. Boleni attik cunku o bir sabit (tamamen veriye bagli, tahmini hesaplamak istedigimiz parametreye bagli degil). Tabii bu durumda sol ve sag taraf birbirine esit olmaz, o yuzden esitlik yerine orantili olmak isaretini kullandik. Bu cercevede "belli bir numerik sabit cercevesinde birbirine esit (equal within a numeric constant)" gibi cumleler de gorulebilir.

Ornek

Diyelim ki bir bozuk para ile 10 kere yazi-tura attik, ve sonuc altta

T H H H H T T H H H

Bu veriye bakarak paranin hileli olup olmadigini anlamaya calisacagiz. Bayes ifadesini bu veriye gore yazalim,

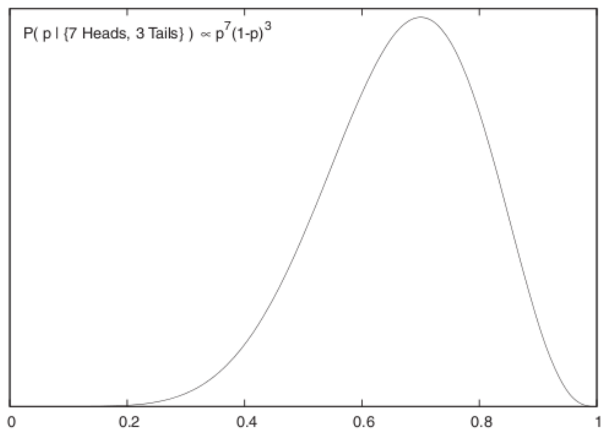
$$P(p|\{\text{T H H H H T T H H H}\} \propto P(\{\text{T H H H H T T H H H}|p)P(p)\}$$

$P(p)$ ifadesi ne anlama gelir? Aslinda bu ifadeyi $P([Dagilim] = p)$ olarak gormek daha iyi, artik p parametresini bir dagilimdan gelen bir tekil deger olarak gordugumuze gore, o dagilimin belli bir p'ye esit oldugu zamani modelliyoruz burada. Her halukarda $P(p)$ dagilimini, yani onsel (prior) olasiligi bilmiyoruz, hesaptan once her degerin mumkun oldugunu biliyoruz, o zaman bu onsel dagilimi duz (flat) olarak aliriz, yani $P(p) = 1$.

$P(\{\text{T H H H H T T H H H}|p)$ ifadesi goz korkutucu olabilir, ama buradaki her ogenin bagimsiz ozdesce dagilmis (independent identically distributed) oldugunu gorursek, ama bu ifadeyi ayri ayri $P(\{T|p)$ ve $P(\{H|p)$ carpimlari olarak gorebiliriz. $P(\{T|p) = p$ ve $P(\{H|p) = 1 - p$ oldugunu biliyoruz. O zaman

$$P(p|\{7 \text{ Tura}, 3 \text{ Yazi}\} \propto p^7(1-p)^3$$

Grafiklersek,



Boylece p icin bir sonsal (posterior) dagilim elde ettik. Artik bu dagilimin yuzde 95 agirliginin nerede oldugunu rahatca gorebiliriz / hesaplayabiliriz. Dagilimin tepe noktasinin $p = 0.7$ civarinda oldugu goruluyor. Bir dagilimla daha fazlasini yapmak ta mumkun, mesela bu fonksiyonu p'ye bagli baska bir fonksiyona karsi entegre etmek mumkun, mesela beklentiyi bu sekilde hesaplayabiliriz.

Onsel dagilimin her noktaya esit agirlik veren birornek (uniform) secilmis olmasi, yani problemi cozmeye sifir bilgiden baslamis olmamiz, yontemin bir zayifligi olarak gorulmemeli. Yontemin kuvveti elimizdeki bilgiyle baslayip onu net bir sekilde veri ve olurluk uzerinden sonsal tek dagilima goturebilmesi. Baslangic ve sonuc arasindaki baglanti gayet net. Fazlasi da var; ilgilendigimiz alani (domain) ogrendikce, basta hic bilmedigimiz onsel dagilimi daha net, bilgili bir sekilde secebiliriz ve bu sonsal dagilimi da daha olmasi gereken modele daha yaklastirabilir.

Gaussian Kontrolu

Diyelim ki Gaussian dagilimina sahip oldugunu dusundugumuz $\{x_i\}$ verilerimiz var. Bu verilerin Gaussian dagilimina uyup uymadigini nasil kontrol edecegiz? Normal bir dagilimin her veri noktasi icin soyle temsil edebiliriz,

$$y_i = \Phi\left(\frac{x_i - \mu}{\sigma}\right)$$

Burada Φ standart Gaussian'i temsil ediyor (detaylar icin *Istatistik Ders 1*) ve CDF fonksiyonuna tekabul ediyor. CDF fonksiyonunun ayni zamanda ceyregi (quantile) hesapladigi soylenir, aslinda CDF son derece detayli bir olasilik degeri verir fakat evet, dolayli yoldan noktanin hangi ceyrek icine dustugu de gorulecektir.

Simdi bir numara yapalim, iki tarafa ters Gaussian formulunu uygulayalim, yani Φ^{-1}.

$$\Phi^{-1}(y_i) = \Phi^{-1}\left(\Phi\left(\frac{x_i - \mu}{\sigma}\right)\right)$$

$$\Phi^{-1}(y_i) = \frac{x_i - \mu}{\sigma}$$

$$x_i = \Phi^{-1}(y_i)\sigma + \mu$$

Bu demektir ki elimizdeki verileri $\Phi^{-1}(y_i)$ bazinda grafiklersek, bu noktalar egimi σ, baslangici (intercept) μ olan bir duz cizgi olmalidir. Eger kabaca noktalar duz cizgi olusturmuyorsa, verimizin Gaussian dagilima sahip olmadigina karar verebiliriz.

Ustte tarif edilen grafik, olasilik grafigi (probability plot) olarak bilinir.

Ters Gaussian teorik fonksiyonunu burada vermeyecegiz, Scipy `scipy.stats.invgauss` hesaplar icin kullanilabilir. Fakat y_i'nin kendisi nereden geliyor? Eger y_i, CDF'in bir sonucu ise, pur veriye bakarak bir CDF degeri de hesaplayabilmemiz gerekir. Bunu yapmak icin bir baska numara lazim.

1. Eldeki sayilari artan sekilde siralayin

2. Her veri noktasina bir derece (rank) atayin (siralama sonrasi hangi seviyede oldugu yeterli, 1'den baslayarak).

3. Ceyrek degeri y_i bu sira / $n+1$, n eldeki verinin buyuklugu.

Bu teknik niye isliyor? x'in CDF'i $x_i < x$ sartina uyan x_i'lerin orani degil midir? Yani bir siralama soz konusu ve ustteki teknik te bu siralamayi biz elle yapmis olduk, ve bu siralamadan gereken bilgiyi aldik.

Moment Fonksiyonlari

Olasılık matematiğinde "moment üreten işlevler" olarak adlandırılan, başlangıçta pek yararlı gibi gözükmesede bir takım matematiksel özellikleri olduğu için, ispatlarda oldukça işe yarayan bir kavram vardır.

Her rasgele değişkenin bir dağılımı olduğunu biliyoruz. Her rasgele değişkenin de ayrıca bir moment üreten fonksiyonu da vardır. Ayrıca, moment üreten fonksiyon ile rasgele değişken arasında bire-bir olarak bir ilişki mevcuttur. "Bu neye yarar?" diye sorulabilir; Cevap olarak, mesela cebirsel olarak türete türete bir moment'e geldiğimiz düşünelim, ve tekrar başka bir taraftan, başka bir formülden gene türete türete tekrar aynı moment işlevine geliyorsak, bu demektir ki, iki taraftan gelen rasgele değişkenler (ve tekabül eden dağılımları) birbirine eşittir. Bazı şartlarda moment üreten işlevler ile cebir yapmak, dağılım fonksiyonlarından daha rahat olmaktadır.

Her rasgele değişken için, moment üreten işlev şöyle bulunur.

X rasgele degiskenin moment ureten operasyonu

$M(t) = E(e^{tX})$ olarak gosterilir

Ayriksal operasyonlar icin

$$M(t) = \sum_x e^{tx} p(x)$$

Surekli islevler icin

$$M(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

Kuram

Gelelim yazımızın esas konusu olan kuramımıza.

Eğer $X_1, X_2 ... X_n$ bağımsız rasgele değişken ise, ve her değişkenin $M_i(t)$ $i = 1, 2, 3, ...n$ olarak, öz olarak aynı olan birer moment üreten işlevi var ise, o zaman,

$$Y = \sum_{i=1}^{n} aX_i$$

acilimi

$$M_y(t) = \prod_{i=1}^{n} M(a_i t)$$

olacaktir.

Ispat

$M_y(t) = E(e^{tY} = E(e^{t(a_1 X_1 + a_2 X_2 + .. + a_n X_n)}$

$= E[\exp(ta_1 X_1 ta_2 X_2 ... + ta_n X_n)]$

$= E[\exp(ta_1 X_1) + \exp(ta_2 X_2) + ... + \exp(ta_n X_n)]$

$= E[\exp(ta_1 X_1)] + E[\exp(ta_2 X_2)] + ... + E[\exp(ta_n X_n)]$

Daha once belirttigimiz gibi

$M_i(t) = E[\exp(tX_i)]$

olduguna gore ve t yerine ta_i koyuldugunu dusunelim

$$M_y(t) = \prod_{i=1}^{n} M_y(a_i t)$$

olacaktir.

Bunu $M_y(t) = (M_i(a_i t))^n$ seklinde de gosterebiliriz.

Orneklem Dagilimlari (Sampling Distributions)

Diyelim ki elimizde (hakkinda) ogrenmek istedigimiz bir sayisal obek (population) var. Bu obekteki her elemani ayri ayri incelemek istemiyoruz, problem degil, nufustan bir orneklem (sample) aliriz. Eger bu orneklem nufusu yeterince iyi temsil ediyorsa, problem cikmaz. Bu temsiliyeti garantilemenin iyi bir yolu orneklemi rasgele yapmaktir.

Simdi, diyelim ki, bu orneklemi bir sekilde ozetlemek istiyoruz yani orneklem verisi kullanilarak hesaplanmis temsili bir istatistik (descriptive statistic) elde edecegiz.

Fakat orneklemimiz rasgele idi. Bu istatistigimiz (ki o da sonucta bir rasgele degiskendir ve onun da bir dagilimi vardir), nasil bir dagilima sahiptir? Yani nufus dagilimi (population distribution), ve orneklem dagiliminin (sampling distribution) birbiriyle baglantisiyla ilgileniyoruz.

Teori

Eger $X_1, .., X_n$ bir $N(\mu, \sigma)$ dagiliminda alinmis orneklem olsun. O zaman orneklem ortalamasinin dagilimi $N(\mu, \sigma/\sqrt{n})$.

[TBD - Ispat]

Teori

Eger $X_1, .., X_n$ bir $N(\mu, \sigma)$ dagiliminda alinmis orneklem olsun. O zaman su buyukluk

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

t_{n-1} dagilimina, yani $n - 1$ serbestlik derecesindeki (degree of freedom) bir Student's t Dagilimidir.

[TBD - Ispat]

Hipotez Testleri (Hypothesis Testing)

Hipotez testi (bir veriye dayanarak) farzedilen bir parametreyi bir sabit degerle karsilastirmak, ya da iki parametreyi birbiriyle karsilastirmak icin kullanilir.

Bir hipotez testi, sonucta sadece iki cevap verebilecek bir sorudur; bu sonuclar "reddetmek" ya da "reddetmemek" olabilir. Dikkat: bu sonuclardan biri "kabul etmek" degil, bir istatistiki hipotezi kabul etmek mumkun degildir. Tek soyleyebildigimiz "bir hipotezi reddetmek icin elimizde yeterli veri olmadigini" soylemektir. Ama reddedebiliyorsak, bu sonucta daha bir kesinlik vardir.

Tek Orneklem t Testi (One-sample t test)

Bu test verinin Normal dagilimdan geldigini farzeder, tek orneklem durumunda elde $x_1, ..., x_n$ verisi vardir, ve bu veri $N(\mu, \Sigma)$ dagilimindan gelmistir ve test etmek istedigimiz hipotez / karsilastirma $\mu = \mu_0$.

```
from scipy.stats import ttest_1samp, wilcoxon, ttest_ind
import pandas as pd
daily_intake = np.array([5260,5470,5640,6180,6390,6515, 6805,7515,7515,8230,8770])
df = pd.DataFrame(daily_intake)
print df.describe()
```

```
                 0
count    11.000000
mean   6753.636364
std    1142.123222
min    5260.000000
25%    5910.000000
50%    6515.000000
75%    7515.000000
max    8770.000000
```

```
t_statistic, p_value = ttest_1samp(daily_intake, 7725)
print "one-sample t-test", p_value
```

```
one-sample t-test 0.0181372351761
```

Sonuc `p_value 0.05`'ten kucuk cikti yani yuzde 5 onemliligini (significance) baz aldik bu durumda veri hipotezden onemli derecede (significantly) uzakta. Demek ki ortalamanin 7725 oldugu hipotezini reddetmemiz gerekiyor.

Testi iki orneklemli kullanalim, gruplar 0/1 degerleri ile isaretlendi, ve test etmek istedigimiz iki grubun ortalamasinin (mean) ayni oldugu hipotezini test etmek. t-test bu arada varyansin ayni oldugunu farzeder.

```
energ = np.array([
[9.21, 0],
[7.53, 1],
[7.48, 1],
[8.08, 1],
[8.09, 1],
[10.15, 1],
[8.40, 1],
[10.88, 1],
[6.13, 1],
[7.90, 1],
[11.51, 0],
[12.79, 0],
[7.05, 1],
[11.85, 0],
[9.97, 0],
[7.48, 1],
[8.79, 0],
[9.69, 0],
[9.68, 0],
[7.58, 1],
[9.19, 0],
[8.11, 1]])
group1 = energ[energ[:, 1] == 0][:, 0]
group2 = energ[energ[:, 1] == 1][:, 0]
t_statistic, p_value = ttest_ind(group1, group2)
print "two-sample t-test", p_value
```

```
two-sample t-test 0.00079899821117
```

$p - value < 0.05$ yani iki grubun ortalamasi ayni degildir. Ayni oldugu hipotezi reddedildi.

Eslemeli t-Test (Paired t-test)

Eslemeli testler ayni deneysel birimin olcumu alindigi zaman kullanilabilir, yani olcum alinan ayni grupta, deney sonrasi deneyin etki edip etmedigi test edilebilir. Bunun icin ayni olcum deney sonrasi bir daha alinir ve "farklarin ortalamasinin sifir oldugu" hipotezi test edilebilir. Altta bir grup hastanin deney oncesi ve sonrasi ne kadar yiyecek tukettigi listelenmis.

```
intake = np.array([
[5260, 3910],
[5470, 4220],
[5640, 3885],
[6180, 5160],
[6390, 5645],
[6515, 4680],
[6805, 5265],
[7515, 5975],
```

```
[7515, 6790],
[8230, 6900],
[8770, 7335],
])
pre = intake[:, 0]
post = intake[:, 1]
t_statistic, p_value = ttest_1samp(post - pre, 0)
print "paired t-test", p_value
```

```
paired t-test 3.05902094293e-07
```

Wilcoxon isaretli-sirali testi (Wilcoxon signed-rank test)

t Testleri Normal dagilima gore sapmalari yakalamak acisindan, ozellikle buyuk orneklemler var ise, oldukca saglamdir. Fakat bazen verinin Normal dagilimdan geldigi faraziyesini yapmak istemeyebiliriz. Bu durumda *dagilimdan bagimsiz metotlar* daha uygundur, bu tur metotlar icin verinin yerine cogunlukla onun sira istatistiklerini (order statistics) kullanir.

Tek orneklemli Wilcoxon testi icin prosedur μ_0'i tum veriden cikartmak ve geri kalan (farklari) isaretine bakmadan numerik degerine gore siralamak, ve bu sira degerini bir kenara yazmak. Daha sonra geri donup bu sefer cikartma islemi sonucunun isaretine bakmak, ve eksi isareti tasiyan sira degerlerini toplamak, ayni islemi arti isareti icin yapmak, ve eksi toplami arti toplamindan cikartmak. Sonucta elimize bir istatistik W gelecek. Bu test istatistigi aslinda $1..n$ tane sayi icinden herhangi birini $1/2$ olasiligiyla secmek, ve sonuclari toplamaya tekabul etmektedir. Ve bu sonuc yine `0.05` ile karsilastirilir.

```
z_statistic, p_value = wilcoxon(daily_intake - 7725)
print "one-sample wilcoxon-test", p_value
```

```
one-sample wilcoxon-test 0.0279991628713
```

Hipotezi yine reddettik.

Ustte yaptigimiz eslemeli t-testi simdi Wilcoxon testi ile yapalim,

```
z_statistic, p_value = wilcoxon(post - pre)
print "paired wilcoxon-test", p_value
```

```
paired wilcoxon-test 0.00463608893545
```

Binom Testi

Binom dagilimi belli sayida "deney" icinde her seferinde p olasiligi tasiyan iki kategorili bir olaydan *kac tane* olabilecegini modeller. Dagilim

$$P(K = k) = \binom{n}{k} p^k q^{n-k}$$

olarak belirtilir, ki $q = 1 - p$ degeridir. Bu dagilimin $n > 20$, yani "yeterince buyuk" degerleri icin Normal (Gaussian) Dagilima yaklastigi / onun gibi oldugu bilinmektedir. Bu yaklasilan dagilim ortalamasi np ve standart sapmasi $\sqrt{npq}$ olan bir Normal dagilim olacaktir, yani $N(np, \sqrt{npq})$.

Devam edelim, madem elimizde bir normal dagilim var, bu normal dagilimi diger her normal dagilim gibi standardize edebiliriz,

$$Z = \frac{K - np}{\sqrt{npq}} \sim N_{0,1}$$

Burada rasgele degisken K, yani her binom deneyi ardindan ele gececek basari sayisi burada. Binom testi icin test etmek istedigimiz sey budur. O zaman p yerine test ettigimiz ana binom dagilimindan gelen ana orani $\hat{p}$ kullaniriz, ki bu basari / tum deney sayisi olarak hesaplanir, K yeni deneydeki ele gecen basari sayisidir, n ornekleminin buyuklugudur, bu sayilari yerine koyarak Z dagilimindan bir guven rakami (confidence) elde edebiliriz.

Bir ornek uzerinde gorelim: diyelim ki elimizde bir Web sitesinin gunluk ziyaret, tiklama sayilarini gosteren bir veri seti var (CVR ziyaretcilerin, sitedeki tiklayan musteriye "cevirme' orani, -conversion-)

```
import pandas as pd
from scipy import stats
a = pd.DataFrame({'tiklama': [20.,2.,40.,5.,10.,100.],
                  'ziyaret': [100.,10.,300.,400.,30.,800.]})
a['cvr'] = a['tiklama'] / a['ziyaret']
print a
```

```
   tiklama  ziyaret       cvr
0       20      100  0.200000
1        2       10  0.200000
2       40      300  0.133333
3        5      400  0.012500
4       10       30  0.333333
5      100      800  0.125000
```

Diyelim ki bu veri seti icin cvr'in 0.16, yani yuzde 16 oldugunu onceden biliyoruz. Ustteki basari orani binom dagili ile modellenebilir, ziyaretler "deneylerdir", yani orneklem buyuklugunu gosterirler. Tiklama ise basaridir,

```
p_hat = 0.16
btest = lambda x: (x['cvr']-p_hat) / np.sqrt( p_hat*(1-p_hat)/x['ziyaret'])
a['guven'] = a.apply(btest, axis=1)
a['guven'] = np.round(stats.zprob(a['guven'])*100,2)
print a
```

```
   tiklama  ziyaret       cvr  guven
0       20      100  0.200000  86.24
1        2       10  0.200000  63.50
2       40      300  0.133333  10.39
3        5      400  0.012500   0.00
4       10       30  0.333333  99.52
5      100      800  0.125000   0.35
```

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 $X_1, X_2 ... X_n$ 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.

$X_1, X_2, .., X_n$ bagimsiz degiskenler olsun.

$E(X_i) = \mu$

$Var(X_i) = \sigma$

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

O zaman her $\epsilon > 0$ icin ve $n \to \infty$, $p(|\bar{X_n} - \mu|) \to 0$.

Bu tanımlara göre, her rasgele değişkenin (deneyin) ortalaması ayni 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, çünku Büyük sayılar kanununu, matematiksel olarak,$\bar{X_n}$ değişkeninin ortalamasını tekil olarak her Xi dağılımının (aynı olan) ortalaması arasında birkü onun da formülü başka rasgelen değişkenlere dayanıyor.

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

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

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

E dogrusal bir islec (linear operator) oldugu icin disaridan iceri dogru nufuz 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.

Diger bir olasilik kuramina gore

$$Y = a + bX$$

$$Var(Y) = b^2 Var(X)$$

oldugunu biliyoruz. O zaman,

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

$$Var(\bar{X_n}) = Var(\frac{1}{n} \sum_{i=1}^{n} X_i) = \frac{1}{n^2} \sum_{i=1}^{n} Var(X_i)$$

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

Artık Çebişev kuramını kullanmaya hazırız.

$n \to \infty$,

$$P(|\bar{X_n} - \mu| > \epsilon) \to 0$$

Cebisev'den

$$P(|\bar{X_n} - \mu| > \epsilon) \leqslant \frac{Var(\bar{X_n})}{\epsilon^2}$$

$$P(|\bar{X_n} - \mu| > \epsilon) \leqslant \frac{\sigma^2}{n\epsilon^2} \to 0$$

$\sigma^2/n\epsilon^2$'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.

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,

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

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

Eger $x \in R$ dersek o zaman

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

t'nin denkleme bu şekilde nereden geldiği şaşkınlık yaratabilir. Daha önce tanımlanan şu ibareye dikkat edelim, $x : |x-u| > t$ 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.

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

Ortadaki entegral niye birinci entegralden büyük? Çünkü ortadaki entegraldeki $F(x)dx$ ibaresinden önce gelen kısmın, her zaman 1'den büyük olacağını belirttiğimize göre, ikinci entegralin birinciden büyük olması normaldir.

Evet...Üçüncü entegral ispata oldukça yaklaştı aslında. Standart sapma işaretini hala ortada göremiyoruz, fakat son entegraldeki ibare standart sapma değerini zaten içeriyor. Önce daha önceki olasılık natematiği 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

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

ki $\int_R f(x)dx$ zaten $P(|X-\mu| > t)$ olarak tanimlanmisti.

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

https://gist.github.com/mblondel/1761714

Introductory Statistics with R

Introduction to Probability and Statistics Using R