

Ozet Istatistikleri, Grafikleri

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 ckar.

Tanim

Surekli dagilim fonksiyonlari icin $E(X)$

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

ayriksal dagilimlar icin

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

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 hesaplayacaktır. Ortalama μ_x olarak ta gosterilebilir.

$E(X)$ 'in bir tanim olduguna dikkat, yani bu ifade tamamen bizim yarattigimiz, ortaya cikarttigimiz bir sey, matematigin baz kurallarindan gelerek turetilen bir kavram degil.

Notasyonel basitlik icin ustteki toplam / integral 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 tanımlamak gerekir, bu sart soyle olsun,

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

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

Ornek

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

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

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

2 nereden cikti? Cunku $|x|$ kullaniyoruz, o zaman sinir degerlerinde sadece sifirin sagina bakip sonucu ikiyle carpma 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.

Tanim

x_1, \dots, x_n verilerini iceren orneklemen (sample) ortalamasi

$$\bar{x} = \frac{1}{n} \sum x_i \quad (1)$$

Dikkat bu orneklemdeki verinin ortalamasi. Hicbir dagilim hakkında hicbir faraziye yapmadik. Ayrica tanim kullandik, yani bu ifadenin ne oldugu tamamen bize bagli.

Orneklem ortalamasi 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.

Tanim

Y rasgele degiskeninin varyansi (variance)

$$\text{Var}(Y) = E((Y - E(Y))^2)$$

Ifadede toplama ve bolme gibi islemler olmadigina dikkat; onun yerine kare ifadeleri üzerinde beklenti ifadesi var. Yani Y' 'nin beklentisini rasgele degiskenin kendisinden cikartip kareyi aliyoruz, ve bu islemin Y' 'den gelen tum zar atislari uzerinden beklentisi bize varyansi veriyor. Bir rasgele degisken gorunce onun yerine "dagilimdan uretilen sayi" dusunmek faydalidir, ki bu gercek dunya sartlarindan (ve buyuk miktarda olunca) veri noktalarini temsil eder.

Tanim

y_1, \dots, y_n orneklemnin varyansi (literaturde S^2 olarak gecebiliyor,

$$S^2 = \frac{1}{n} \sum (y_i - \bar{y})^2 \quad (2)$$

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

Ilginc bir cebirsel islem sudur ve bize verinin uzerinden tek bir kez gecerek hem sayisal ortalama hem de sayisal varyansi hesaplamamizi saglar. Eger \bar{y} tanimini ustteki formule sokarsak,

$$\begin{aligned} &= \frac{1}{n} \sum_i y_i^2 + \frac{1}{n} \sum_i m^2 - \frac{2}{n} \sum_i y_i \bar{y} \\ &= \frac{1}{n} \sum_i y_i^2 + \frac{\bar{y}^2 n}{n} - \frac{2\bar{y}n}{n} \bar{y} \\ &= \frac{1}{n} \sum_i y_i^2 + \bar{y}^2 - 2\bar{y}^2 \\ &= \frac{1}{n} \sum_i y_i^2 - \bar{y}^2 \end{aligned}$$

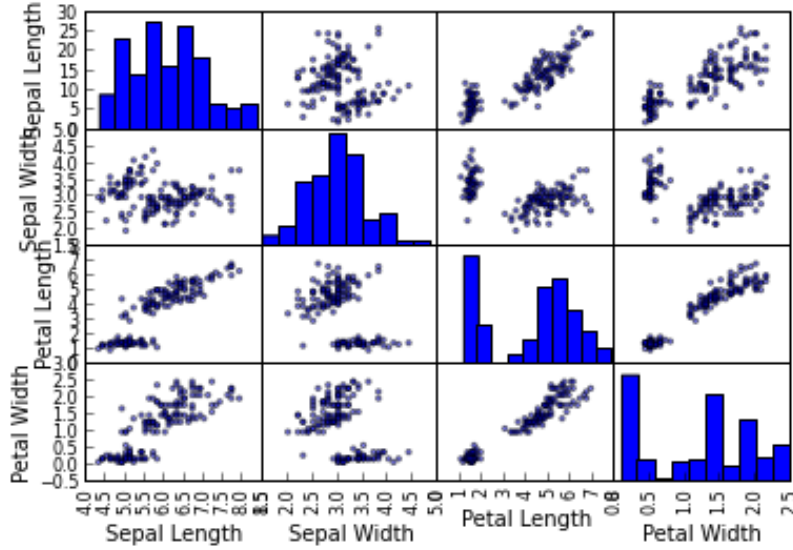
Bu arada standard sapm varyansin karesidir, ve biz karekok olan versiyon ile calismayi tercih ediyoruz. Niye? Cunku o zaman veri noktalarinin ve yayilma olcusunun birimleri birbiri ile ayni olacak. Eger veri setimiz bir alisveris sepetindeki malzemelerin lira cinsinden degerleri olsaydi, varyans bize sonucu "kare lira" olarak verecekti ve bunun pek anlami olmayacakti.

Kovaryans ve Korelasyon (Covariance and Correlation)

iki veya daha fazla boyutun arasindaki iliskileri gormek veri analizinde onemli bir yer tutar. Yontemlerden birisi verideki mumkun her ikili iliskiyi grafiksel olarak gosterme. Pandas `scatter_matrix` bunu yapabilir. Iris veri seti uzerinde

gorelim, her boyut hem y-ekseni hem x-ekseninde verilmiş, ilişkiyi gormek için ekseninde o boyutu bulup kesime noktalarındaki grafiğe bakmak lazım.

```
import pandas as pd
df = pd.read_csv('iris.csv')
df = df.ix[:,0:4]
pd.scatter_matrix(df)
plt.savefig('stat_summary_01.png')
```



İlişki olduğu zaman o ilişkiye tekabül eden grafikte “düz çizgiye benzer” bir görüntü olur, demek ki değişkenlerden biri artınca diğeri de artıyor, azalınca diğeri de azalıyor demektir. Eğer ilinti yok ise bol gürültülü, ya da yuvarlak kureye benzer bir şekil çıkar. Üstteki grafiğe göre yaprak genişliği (petal width) ile yaprak boyu (petal length) arasında bir ilişki var.

Tanım

X, Y rasgele değişkenlerin arasındaki kovaryans,

$$\text{Cov}(X, Y) = E(X - E(X))(Y - E(Y))$$

Yani hem X hem Y’nin beklentilerinden ne kadar saptıklarını her veri ikilisi için, çıkartarak tespit ediyoruz, daha sonra bu farkları birbiriyle çarpıyoruz, ve beklentisini alıyoruz (yani tüm olasılık üzerinden ne olacağını hesaplıyoruz). Çarpımların bildiğimiz özelliğine göre, artı değer artı değerle çarpılınca artı, eksi ile eksi artı, eksi ile artı eksi verir, ve bu bilgi bize ilinti bulma hakkında güzel bir ipucu sunar. Pozitif sonucun pozitif korelasyon, negatif ise tersi şekilde ilinti olduğu sonucuna böylece kolayca erisebiliriz.

Ayrı ayrı X, Y değişkenleri yerine çok boyutlu X kullanırsak, ki boyutları m, n olsun yani m veri noktası ve n boyut (özellik, öge) var, tanımı şöyle ifade edebiliriz,

$$\Sigma = \text{Cov}(X) = E((X - E(X))^T(X - E(X)))$$

Verisel Kovaryans (Empirical Covariance)

Tanim

$$S = \frac{1}{n}(X - E(X))^T(X - E(X))$$

Pandas ile `cov` cagrisi bu hesabi hizli bir sekilde yapar,

```
print df.cov()
```

	Sepal Length	Sepal Width	Petal Length	Petal Width
Sepal Length	0.685694	-0.039268	1.273682	0.516904
Sepal Width	-0.039268	0.188004	-0.321713	-0.117981
Petal Length	1.273682	-0.321713	3.113179	1.296387
Petal Width	0.516904	-0.117981	1.296387	0.582414

Eger kendimiz bu hesabi yapmak istersek,

```
means = df.mean()
n = df.shape[0]
df2 = df.apply(lambda x: x - means, axis=1)
print np.dot(df2.T, df2) / n
```

```
[[ 0.68112222 -0.03900667  1.26519111  0.51345778]
 [-0.03900667  0.18675067 -0.319568   -0.11719467]
 [ 1.26519111 -0.319568    3.09242489  1.28774489]
 [ 0.51345778 -0.11719467  1.28774489  0.57853156]]
```

Medyan ve Yuzdelikler (Percentile)

Ustteki hesaplarin cogu 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 hucrelerine / 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,

```
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 edilmes 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 noktalar (outliers) acik sekilde gosterirler. Isim nereden geliyor? Box yani kutu, dagilimin agirliklerinin 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 bunlari 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 farklilikini 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 farklilikini hemen gosterir.

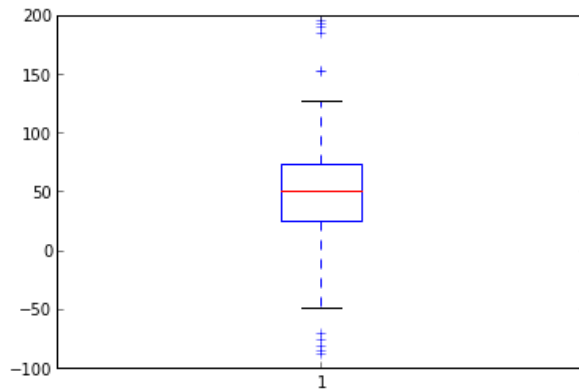
Python uzerinde basit bir BW grafigi

```
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])
```

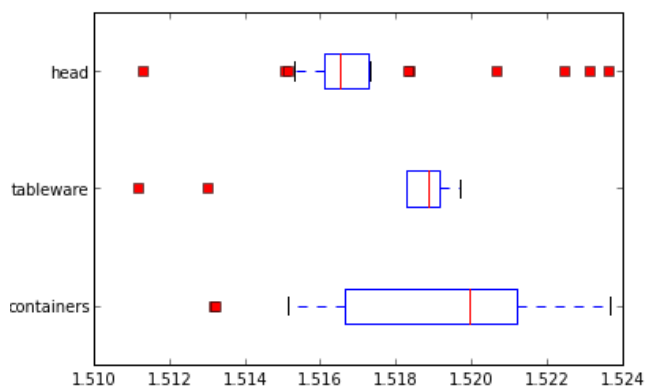
```
plt.yticks([1, 2, 3], ['containers', 'tableware', 'head'])
```

```
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]

```



Parametre Tahmin Ediciler (Estimators)

Maksimum Olurluk (maximum likelihood) kavramini kullanarak ilginç bazı sonuclara erismek mumkun; bu sayede dagilim fonksiyonlari ve veri arasinda bazi sonuclar elde edebiliriz. Maksimum olurluk nedir? MO ile verinin her noktası teker teker olasilik fonksiyonuna gecilir, ve elde edilen olasilik sonuclari birbiri ile carpilir. Cogunlukla formül icinde bilinmeyen bir(kac) parametre vardır, ve bu carpim sonrasI, icinde bu parametre(ler) olan yeni bir formül ortaya cıkar. Bu nihai formülün kısmi türevi alınıp sifira esitlenince cebirsel bazı teknikler ile bilinmeyen parametre bulunabilir. Bu sonuc eldeki veri bağlamında en mumkun (olur) parametre degeridir. Oyle ya, mesela Gaussian $N(10, 2)$ dagilimi var ise, 60,90 gibi degerlerin “olurlugu” dusuktur. Gaussin uzerinde örnek,

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

Carpim sonrasI

$$\begin{aligned} f(x_1, \dots, x_n; \mu, \sigma) &= \prod \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma^2}(x_i - \mu)^2 \right\} \\ &= \frac{(2\pi)^{-n/2}}{\sigma^n} \exp \left\{ -\frac{\sum (x_i - \mu)^2}{2\sigma^2} \right\} \end{aligned}$$

Ustel kısım $-n/2$ nereden geldi? Cunku bolen olan karekoku uste cikardik, boylece $-1/2$ oldu, n cunku n tane veri noktası yuzunden formül n kere carpiliyor. Veri noktaları x_i icinde. Eger log, yani \ln alırsak \exp 'den kurtuluruz, ve biliyoruz ki log olurlugu maksimize etmek normal olurlugu maksimize etmek ile aynı şeydir, cunku \ln transformasyonu monoton bir transformasyondur. Ayrıca olurluk icbukeydir (concave) yani kesin tek bir maksimumu vardır.

$$\ln f = -\frac{1}{2}n \ln(2\pi) - n \ln \sigma - \frac{\sum (x_i - \mu)^2}{2\sigma^2}$$

Türevi alıp sifira esitleyelim

$$\frac{\partial(\ln f)}{\partial \mu} = \frac{\sum (x_i - \mu)^2}{2\sigma^2} = 0$$

$$\hat{\mu} = \frac{\sum x_i}{n}$$

Bu sonuc (1)'deki formül, yani örneklem ortalaması ile aynı! Fakat buradan hemen bir bağlantıya zıplamadan önce sunu hatırlayalım - örneklem ortalaması formülünü *biz* tanımladık. “Tanım” diyerek bir ifade yazdık, ve budur dedik.

Simdi sonradan, verinin dagiliminin Gaussian oldugunu farzederek, bu verinin mumkun kilabilecegi en optimal parametre degeri nedir diye hesap ederek ayni formule eristik, fakat bu bir anlamda bir guzel raslanti oldu.. Daha dogrusu bu aynilik Gaussian / Normal dagilimlarinin “normalligi” ile alakali muhakkak, fakat ornekleme ortalamasi hicbir dagilim faraziyesi yapmiyor, herhangi bir dagilimdan geldiği bilinen ya da bilinmeyen bir veri uzerinde kullanilabiliyor (ise yaramayabilir de tabii). Bunu unutmayalim. Istatistikte matematigin lakaytlasmasi (sloppy) kolaydir, o sebeple neyin tanim, neyin hangi faraziye gore optimal, neyin nufus (population) neyin ornekleme (sample) oldugunu hep hatirlamamiz lazim.

Devam edelim, maksimum olurluk ile $\hat{\sigma}$ hesaplayalim,

$$\frac{\partial(\ln f)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{\sum (x_i - \mu)^2}{2\sigma^3} = 0$$

Cebirsel birkac duzenleme sonrasi ve μ yerine yeni hesapladigimiz $\hat{\mu}$ kullanarak,

$$\hat{\sigma}^2 = \frac{\sum (x_i - \hat{\mu})^2}{n}$$

Bu da ornekleme varyansi ile ayni!

Büyük Sayılar Kanunu (Law of Large Numbers)

Olasılık kuramında önemli matematiksel bir denklem, büyük sayılar kanunudur. Bu kanun, ornekleme (sample) ile rasgele degiskenler, yani matematiksel olasilik dagilimleri olan dunya arasinda bir baglanti gorevi gorur.

Kanun kabaca bildigimiz günlük bir gerceğin matematiksel ispatıdır da denebilir. Yazı-tura atarken yazı çıkma ihtimalinin 1/2 olduğunu biliyoruz (cunku zar atisini bir dagilim gibi goruyoruz). Herhalde çoğumuz da bu yazı-tura işleminin “bir çok kere” tekrarlandığı durumda, toplam sonucun aşağı yukarı yarısının yazı olacağını tahmin biliyoruz.

Matematiksel olarak, farzedelim ki her yazı-tura atışı bir deney olsun. Her ayrı deneyin sonucu X_1, X_2, \dots, X_n olarak rasgele değişkenlerle tanımlanmış olsun, bu degiskenlerin dagilimi ayni (cunku ayni zar), ama birbirlerinden bagimsizlar (cunku her deney digerinden alakasiz). Değişkenlerin sonucu 1 ya da 0 değeri taşıyacak, Yazı=1, Tura=0.

Buyuk Sayılar Kanunu tum bu deney sonuclarinin, yani rasgele degiskenlerin av-erajı alinirsa, yani $\bar{X} = X_1 + \dots + X_n$ ile, elde edilen sonucun X_i ’lerin (ayni olan) beklentisine yaklasacaginin soyler, yani n büyüdükçe \bar{X}_n ’in 1/2’ye yaklaştığını ispatlar, yani $E[X_i] = 1/2$ degerine. Notasyonel olarak $E(X_i) = \mu$ olarak da gosterilebilir.

Formüsel olarak, herhangi bir $\epsilon > 0$ icin,

$$\lim_{n \rightarrow \infty} P(|\bar{X} - \mu| \leq \epsilon) = 1$$

ya da

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

ya da

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

Burada ne soylandigine dikkat edelim, X_i dagilimi *ne olursa olsun*, yani ister Binom, ister Gaussian olsun, *orneklem* üzerinden hesaplanan sayisal ortalamanin (empirical mean) formülsele olasilik beklentisine yaklastigini soyluyoruz! X_i 'ler en absurt dagilimler olabilirler, bu dagilimlerin fonksiyonu son derece cetrefil, tek tepeli (unimodal) bile olmayabilir, o formüller üzerinden beklenti için gereken integralin belki analitik cozumu bile mevcut olmayabilir! Ama yine de ortalama, o dagilimlerin beklentisine yaklasacaktır. Istatistik ile olasilik teorisi arasindaki cok onemli bir baglanti bu.

Sonuc sasirtici, fakat bir ek daha yapalim, sezgisel (intuite) olarak bakarsak aslinda sonuc cok sasirtici olmayabilir. Niye? Diyelim ki genel veri $N(\mu, \sigma^2)$ seklinde bir Normal dagilimdan geliyor ve orneklem de bu sebeple ayni dagilima sahip. Bu durumda orneklemdeki veri noktalarinin μ 'ya yakin degerler olmasini beklemek mantikli olmaz mi? Cunku bu dagilim "zar atinca" ya da bir genel nufustan bir "ornek toplayinca" (ki bunu bir anlamda istatistiksel bir zar atisi olarak gorebiliriz) onu μ, σ^2 'e gore atacak. Ornekleme zar atisi sonuclari olarak gordugumuze gore elde edilen verilerin bu sekilde olacagi sasirtici olmamali. Ve bu zar atislarinin ortalamasinin, son derece basit bir aritmetik bir islemlle hesaplaniyor olsa bile, μ 'ye yaklasmasi normal olmalı.

Bu arada, bu argumana tersten bakarsak Monte Carlo entegralinin niye isledigine gorebiliriz (bkz *Monte Carlo, Entegraller, MCMC* yazisi).

Ozellikle orneklem ile genel nufus (population) arasinda kurulan baglantiya dikkat edelim. Istatigin onemli bir bolumunun bu baglanti oldugu soylenebilir. Her orneklem, bilmedigimiz ama genel nufusu temsil eden bir dagilimla ayni dagilima sahip olan X_i 'dir dedik, ve bu ayniliktan ve bagimsizlikten yola cikarak bize genel nufus hakkında bir ipucu saglayan bir kanun gelistirdik (ve birazdan ispatlayacagiz).

Ispata baslayalim.

X_1, X_2, \dots, X_n bagimsiz degiskenler olsun.

$$E(X_i) = \mu$$

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

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

\bar{X}_n de bir rasgele deęiřkendir, ünkü \bar{X}_n deęiřkeni her X_i daęılımıyla alakalı.

İspat devam etmek iin, \bar{X}_n daęılımının beklentisini bulmamız gerekiyor.

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

E dogrusal bir islec (linear operator) olduęu iin disaridan iceri dogru nufuz eder.

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

Dikkat edelim, bu *ortalamanın* beklentisi, ortalamanın kendisinin hangi degere yaklasacagini hala gostermiyor. Eger oyle olsaydi isimiz bitmis olurdu :) Daha yapacak cok is var.

Simdi \bar{X}_n daęılımının standart sapmasını da bulalım. Diger bir olasilik kuramina gore

$$Y = a + bX$$

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

oldugunu biliyoruz. O zaman,

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

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

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n} \quad (3)$$

Artık Çebişev kuramını kullanmaya hazırız. İspatlamaya çalıştığımız neydi? $n \rightarrow \infty$ iken,

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

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.

Buyuk Sayilar Kanunu orneklem ortalamasinin ve varyansinin X_i 'in beklentisi ve varyansi ile baglanti kurar. Merkezi Limit Teorisi bir adim daha atar, ve der ki " \bar{X} 'in dagilimi Gaussian dagilim olmalidir yani normal egrisi seklinde cikmalidir!". Teorinin detaylari bu bolumde bulunabilir.

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.

$$\mathbb{R} = x : |x - \mu| > t$$

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

O zaman,

$$P(|X - \mu| > t) = \int_{\mathbb{R}} f(x) dx$$

Dikkat edelim $P(\cdot)$ 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 integral 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, çünkü birinci integral $f(x)$ olasılık dağılımına bağlı, integral ise bir alan hesabıdır ve olasılık dağılımlarının sonsuzlar arasındaki integrali her zaman 1 çıkar, kaldı ki üstteki x 'in uzayını daha da daralttık.

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

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

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

Merkezi Limit Teorisi (Central Limit Theorem -CLT-)

Buyuk Sayilar Kanunu orneklem ortalamasinin gercek nufus beklentisine yaklasacagini ispatladi. Orneklem herhangi bir dagilimdan gelebiliyordu. CLT bu

teoriyi bir adim ilerletiyor ve diyor ki kendisi de bir rasgele degisken olan orneklem ortalamasi \bar{X} Normal dagilima sahiptir! Daha detaylandirmal gerekirse,

Diyelim ki X_1, \dots, X_i orneklemleri birbirinden bagimsiz, ayni dagilimli ve ortalamasi μ , standart sapmasi σ olan (ki o da ayni dagilima sahip) bir nufustan geliyorlar. Orneklem ortalamasi \bar{X} , ki bu rasgele degiskenin beklentisinin μ , ve (3)'e gore standart sapmasinin σ/\sqrt{n} oldugunu biliyoruz. Dikkat: \bar{X} 'in kendisinden degil, *beklentisinden* bahsediyoruz, yani ortalama dagiliminin ortalamasi (kafalar yeterince karisti mi?). Teori der ki n buyudukce \bar{X} dagilimi bir $N(\mu, \sigma/\sqrt{n})$ dagilimina yaklasir.

Biraz daha temizlemek gerekirse, herhangi bir normal dagilimi standart normal'e donusturmeyi daha once gormustuk, beklentiyi cikartip standart sapmaya boluyoruz, o zaman orneklem dagilimi \bar{X} ,

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

dagilimina yaklasir diyoruz, ki $Z = N(0, 1)$ dagilimidir, beklentisi sifir, standart sapmasi 1 degerindedir.

Bu teoremin ispatini simdilik vermeyecegiz.

Kaynaklar

[1] <http://mathworld.wolfram.com/MaximumLikelihood.html>

[2] Introduction to Probability and Statistics Using R