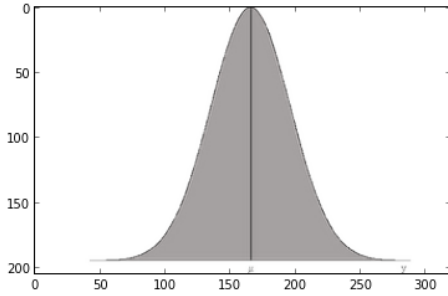


Bu notlar makine öğrenimi, veri madenciliği gibi konularda gerekli olasılık ve istatistik bilgisini paylaşmak için hazırlanıyor. Notlarda olasılık ve istatistik aynı anda anlatılacak, ve uygulamalara ağırlık verilecek.

### Dağılımlar Hakkında

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



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

Nasıl oluyor bu iş?

Açıklama için, normal dağılım eğrisinden bahsetmemiz gerekecek.

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

### Normal Dağılım Eğrisi

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

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

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

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

Bunun sebebi sezgisel olarak tahmin edilebilir, 1..6 arası sayıların tek bir zardan gelme olasılığı aynı, evet. Fakat toplamlara gelince, mesela iki zarlı örnekte, 10

sayısının olasılığı 2 sayısından daha yüksek. Çünkü, 10 sayısını 5-5, 4-6 ya da 6-4 ile alabiliyoruz. 2 sayısı sadece 1-1 ile geliyor.

Buradan şu sonuç çıkabilir: Eğer doğada ölçtüğümüz bir kavramın oluşmasında birden fazla etken var ise, o ölçümlerin sıklığı her zaman çan şekli ile olacaktır. Bir kisinin boyunu, kilosunu etkileyen pek çok diğer faktör olduğu için bu tek olcutleri dağılımlarının normal çıktığı iddia edilebilir.

Toplamların dağılımının çan eğrisine yaklaşması durumu İstatistikte Merkezi Limit Teorisi ile ispatlanmıştır.

## Simulasyon

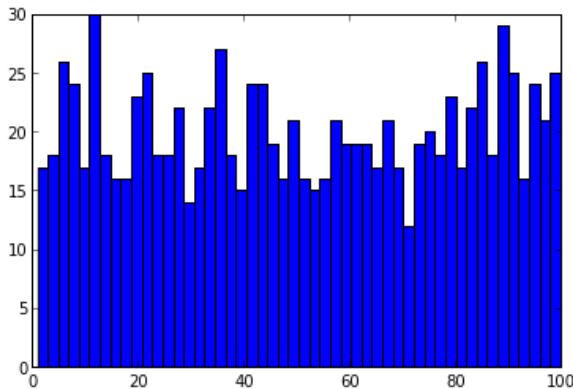
Eğer bu kavramları simulasyon ortamında göstermek istersek, Python ile bunu yapabiliriz.

İlk önce, Random.org sitesinden rasgele sayı üretip bilgisayarımıza 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 bilgisayarımızda rasgele sayı üreten birçok algoritma olsa bile, bu algoritmalar belli bir sayı üretiminden sonra kendini tekrar etmeye başlıyorlar. Gerçek rasgele sayılar için dış bir kaynağa bağlanmak bir seçenek olabilir. Ama sunu da söylemek lazım, simulasyon tekniklerinin tamamı için yarı-rasgele (pseudorandom) sayılar yeterlidir.

Siteden rasgele sayıları üretip, bir veri dosyasına koyuyoruz. Python ile bu sayıları okuyup, ilk önce teker teker sayıların sıklık grafiğini, ondan sonra sayıları üçer üçer toplayıp, onların grafiğini alıp göstereceğiz.

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



```
A = loadtxt('rasgele.dat');
B = []
```

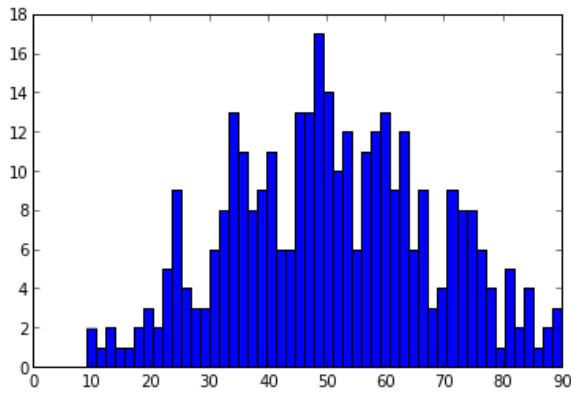
```

i = 1;

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

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

```



## Olasilik

### Orneklem Uzayi (Sample Space)

Orneklem uzayi  $\Omega$  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)

$\Omega$  icindeki her nokta bir sonuctur (outcome). Olaylar  $\Omega$ '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  $\omega$  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.

$$\begin{aligned} P(A) &= 1 - P(A^c) \\ &= 1 - P(\text{hepsi yazi}) \\ &= 1 - P(T_1)P(T_2)\dots P(T_{10}) \\ &= 1 - \left(\frac{1}{2}\right)^{10} \approx .999 \end{aligned}$$

### Rasgele Degiskenler (Random Variables)

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

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

### Ornek

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

### Ornek

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

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

Eger  $X$  ayrık ise, yani sayılabilir bir küme  $\{x_1, x_2, \dots\}$  icinden değerler alıyorsa olasılık fonksiyonu (probability function), ya da olasılık kütle fonksiyonu (probability mass function -PMF-)

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

Bazen  $f_X$ , ve  $F_X$  yerine sadece  $f$  ve  $F$  yazarız.

Tanım

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

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

Bu durumda  $f_X$  olasılık yoğunluk fonksiyonudur (probability density function -PDF-).

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

Ayrıca  $F_X(x)$ 'in türevi alınabildiği her  $x$  noktasında  $f_X(x) = F'_X(x)$  demektir.

Dikkat! Eger  $X$  sürekli ise o zaman  $P(X = x) = 0$  değerindedir.  $f(x)$  fonksiyonunu  $P(X = x)$  olarak görmek hatalıdır. Bu sadece ayrık rasgele değişkenler için işler. Sürekli durumda olasılık hesabı için belli iki nokta arasında integral hesabı yapmamız gereklidir. Ek olarak PDF 1'den büyük olabilir, ama PMF olamaz. PDF'in 1'den büyük olabilmesi integrali bozmaz mı? Unutmayalım, integral hesabı yapıyoruz, noktasal değerlerin 1 olması tüm 1'lerin toplandığı anlamına gelmez. Bakınız *Entegralleri Nasıl Düşünelim* yazımız.

Tanım

$X$  rasgele değişkeninin CDF'i  $F$  olsun. Ters CDF (inverse cdf), ya da çeyrek fonksiyonu (quantile function)

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

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

Eger inf kavramını bilmiyorsak simdilik onu minimum olarak düşünebiliriz.

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

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

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

olarak bilinir.

İki rasgele değişken  $X$  ve  $Y$  dağılımsal olarak birbirine eşitliği, yani  $X \stackrel{d}{=} Y$  eğer  $F_X(x) = F_Y(x)$ ,  $\forall x$ . Bu  $X, Y$  birbirine eşit, birbirinin aynısı demek değildir. Bu değişkenler hakkındaki tüm olasılıksal işlemler, sonuçlar aynı olacak demektir.

Uyari! “ $X$ ’in dağılımı  $F$ ’tir” beyanını  $X \sim F$  şeklinde yazmak bir gelenek. Bu biraz kötü bir gelenek aslında çünkü  $\sim$  sembolü aynı zamanda yaklaşıksallık kavramını belirtmek için de kullanılıyor.

### Bernoulli Dağılımı

$X$ ’in bir yazı-tura atısını temsil ettiğini düşünelim. O zaman  $P(X = 1) = p$ , ve  $P(X = 0) = 1 - p$  olacaktır, ki  $p \in [0, 1]$  olmak üzere. O zaman  $X$ ’in dağılımı Bernoulli deriz, ve  $X \sim \text{Bernoulli}(p)$  diye gösteririz. Olasılık 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 arasındaki herhangi bir reel sayı.

Uyari!

$X$  bir rasgele değişken;  $x$  bu değişkenin alabileceği spesifik bir değer;  $p$  değeri ise bir **parametre**, yani sabit, önceden belirlenmiş reel sayı. Tabii istatistik problemlerde (olasılık problemlerinin tersi olarak düşünürsek) cogenlukla o sabit parametre bilinmez, onun veriden hesaplanması, kestirilmesi gerekir. Her halukarda, coğu istatistik modelde rasgele değişkenler vardır, ve onlardan ayrı olarak parametreler vardır. Bu iki kavramı birbiriyle karıştırmayalım.

### Düz (Uniform) Dağılım

$X$  düz,  $\text{Uniform}(a, b)$  olarak dağılmış deriz, ve bu  $X \sim \text{Uniform}(a, b)$  olarak yazılır eğer

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

ise ve  $a < b$  olacak şekilde. CDF hesabı olasılık eğrisinin integralini temel alır, düz dağılım bir  $a, b$  arasında  $1/b - a$  yüksekliğinde bir dikdörtgen şeklinde olacağı için, bu dikdörtgendeki herhangi bir  $x$  noktasında CDF dağılımı, yani o  $x$ ’in başlayıp sol tarafın alanının hesabı basit bir dikdörtgensel alan hesabıdır, yani  $x - a$  ile  $1/b - a$ ’nin çarpımıdır, 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. Bazilari bu dagilimi

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

olarak gosterebiliyor, cunku bu sekilde (birazdan gorecegimiz) cok boyutlu Gaussian formulu ile alaka daha rahat gozukuyor.

Ileride gorecegiz ki  $\mu$  bu dagilimin “ortasi”, ve  $\sigma$  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. Gele nege gore standart Normal dagilim rasgele degiskeni  $Z$  ile gosterilmelidir, PDF ve CDF  $\phi(z)$  ve  $\Phi(z)$  olarak gosterilir.

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

Bazi faydali puf noktaları

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

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

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

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

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

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

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

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

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

Örnek

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

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

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

Örnek

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

Cevap

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

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

O zaman

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

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

$t$  (Student's  $t$ ) ve Cauchy Dağılımı

$X$ ,  $v$  derece bağımsızlıkta  $t$  dağılımına sahiptir, ki bu  $X \sim t_v$  diye yazılır eğer

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



t dagilimi Normal dagilima benzer ama daha kuyrugü daha kalindir. Aslında Normal dagilimi t dagiliminin  $v = \infty$  olduđu hale tekabül eder. Cauchy dagilimi da t'nin özel bir halidir,  $v = 1$  halidir. Bu durumda yoğunluk fonksiyonu

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

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

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

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

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

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

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

O zaman

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

$\chi^2$  Dagilimi

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

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

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

## İki Degiskenli Dagilimler

### Tanim

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

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

### Ornek

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

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

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

### Cevap

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

### Ornek

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

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

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

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

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

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

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

Devam edersek  $c = 21/4$  buluruz.

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

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

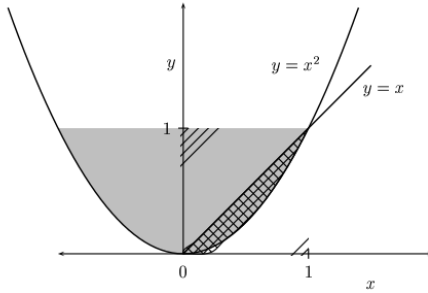
Simdi bunlara bir de  $y \leq x$  eklememiz lazim. Yani ortadaki esitsizlige bir oge daha eklenir.

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

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

$$y \leq 1$$

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



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

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

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

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

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

“Hafizasız” Dagilim, Ustel (Exponential) Dagilim

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

Bu hafizasızlığı, olasılık matematigi ile nasıl temsil ederiz?

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

Yani öyle bir dagilim var ki elimizde,  $X > t$  bilgisi veriliyor, ama (kalan) zamani hala  $P(X > s)$  olasılığı veriyor. Yani  $t$  kadar zaman gectigi bilgisi hiçbir şeyi degistirmiyor. Ne kadar zaman gecmis olursa olsun, direk  $s$  ile gidip ayni olasılık hesabini yapıyoruz.

Sartsal (conditional) formülünü uygularsak üstteki şöyle olur

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

ya da

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

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

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

gibi bir ilişki kurulabilir.

Örnek

Diyelim ki bir bankadaki bekleme zamani ortalama 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 olasılığı nedir?

Cevap

i) Eger  $X$  musterinin bankada beklediği zamani temsil ediyorsa

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

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

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

Bilesen (Marginal) Dağılımlar

Surekli rasgele degiskenler için bilesen yogunluk

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

ve

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

Üstteki integraller gerçek bir dağılım fonksiyonu  $f(x, y)$  verilince alt ve üst limit te tanımlamak zorundadır. Çünkü bilesen yogunluk için bir veya daha fazla degiskeni “integrallerle dışarı atmak (integrate out)” ettigimiz söylenir, eğer ayrık-sal (discrete) ortamda olsaydık bu atılan degiskenin tüm degerlerini goze alarak toplama yapan bir formül yazardık. Surekli ortamda integral kullanıyoruz, ama tüm degerlerin üzerinden yine bir sekilde gecmemiz gerekiyor. İşte alt ve üst limitler bunu gerçekleştiriyor. Bu alt ve üst limitler, atılan degiskenin “tüm degerlerine” bakması gerektiği için  $-\infty, +\infty$  olmalıdır. Eğer problem içinde degiskenin belli degerler arasında olduğu belirtilmiş ise (mesela alttaki örnekte  $x > 0$ ) o zaman entegral limitleri alt ve üst sinirini buna göre degistirebilir.

Örnek

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

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

Örnek

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

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

Tanim

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

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

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

Teori

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

Ornek

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

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

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

Cevap

Bagimsizligi kullanarak birlesik dagilimi hesaplayabiliriz

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

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

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

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

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

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

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

### Kosullu 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,

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

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

Cevap

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

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

Ana formulumuz neydi?

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

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

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

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

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

## Cok Degiskenli Normal

Tek degiskenli Normal dagilimin iki parametresi vardi,  $\mu, \sigma$ . Cok degiskenli formda  $\mu$  bir vektor,  $\sigma$  yerine ise  $\Sigma$  matrisi var. Once rasgele degiskeni tanımlayalım,

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

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

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

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

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

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

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

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

Hatirlama Numarasi



Normal Dağılımın formülünü bazen hatırlayamayabiliriz. Peki daha basit bir formülden başlayarak onu türetebilir miyiz? Bu mümkün. Daha önce  $e^{-x^2}$  Nasıl Entegre Edilir? yazısında

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

olduğunu görmüştük. Dikkat edersek bu integral bir formülün olasılıksal dağılım olup olmadığını kontrol etmek için kullandığımız integrale benziyor. Eğer integral 1 çıkarsa onun olasılıksal dağılım olduğunu biliyoruz. Üstteki sonuç  $\sqrt{\pi}$ , fakat iki tarafı  $\sqrt{\pi}$ 'ye bölersek, sağ taraf 1 olur ve böylece solda bir dağılım elde ederiz. Yani

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

formülünde integralin sağındaki kısım bir dağılımdır. Bu formülü dönüştürerek Gaussian'a erisebiliriz. Üstteki formülün orta noktası (mean) sıfır, varyansı (variance), yani  $\sigma^2 = 1/2$  (bunu da ezberlemek lazım ama o kadar dert değil). O zaman  $\sigma = 1/\sqrt{2}$ .

İlk amacımız  $\sigma = 1$ 'e erismek olsun (çünkü oradan herhangi bir  $\sigma$ 'ya atlayabiliriz), bunun için  $x$ 'i  $\sqrt{2}$ 'ye bölmek lazım, tabii aynı anda onun etkisini sıfırlamak için normalize eden sabiti dengelemek amacıyla  $\sqrt{2}$ 'ye bölmek lazım,

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

$\sigma = 1$ 'e erisince oradan herhangi bir  $\sigma$  için,  $\sigma$  değişkenine bölelim, yine hem e üstüne hem sabite bu eki yapalım,

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

Şimdi herhangi bir ortalama  $\mu$  için bu değişkeni formüle sokalım, bunun için  $\mu$ 'yu  $x$ 'den çıkarmak yeterli

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

e üstündeki kare alma işlemini 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.

### Rasgele Degiskenler, Yogunluklar

Simdi konularin uzerinden bir daha gecelim; rasgele degisken,  $X, Y$  gibi buyuk harflerle gosterilen buyuklukler “bir zar atis sonucu icleri doldurulan” degiskenlerdir. Bu zar atisi her zaman  $X$ ’in,  $Y$ ’nin bagli oldugu dagilima gore olacaktır. Eger  $X \sim N(10, 2)$  ise, bir formül, hesabin icinde  $X$  gordugumuz zaman cogunlukla o noktaya 10’a yakin degerler olacagini biliriz. Tabii ki “kesin” her zaman ne olacagini bilmeyiz, zaten bir modelde noktasal deger (tipik cebirsel degiskenler) yerine rasgele degisken kullanmanin sebeplerinden biri budur.

Rasgele degiskenler matematiksel formuller icinde kullanilabilir dogal olarak, mesela  $Z = X + Y =$  gibi. O zaman elde edilen yeni degisken de bir rasgele degisken olur.

### Kaynaklar

[1] [http://en.wikipedia.org/wiki/Confidence\\_interval](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