Several Random Variables

Prof. Edmund Lam

Department of Electrical and Electronic Engineering
The University of Hong Kong

ELEC2844: Probabilistic Systems Analysis
(Second Semester, 2023–24)

https://www.eee.hku.hk/~elec2844

Multiple random variables

We have mostly looked at one random variable X, including whether it is discrete or continuous.

We have also looked at two random variables briefly, discussing:

- Joint PMF / PDF and marginal PMF / PDF
- Conditional PMF / PDF and independence
- Expectation and variance of the sum of independent random variables (extending from two to many random variables)
- Bayes' rule

In this lecture, we will further investigate topics relating to multiple random variables.

1 Independent, identically distributed random variables
2 Derived distributions
- Laplacian distribution
- Rayleigh distribution
3 Sum of random variables
4 Covariance and correlation
- Covariance
- Correlation
- A detailed example solved with covariance
5 Gaussian random variables
- Central limit theorem
- Higher dimensional Gaussian
- Generalization of Gaussian random variable

Joint PDF

Recall that we have defined the joint PDF of two random variables:

$$\mathbb{P}((X,Y) \in B) = \iint_{(x,y)\in B} f_{X,Y}(x,y)\,dx\,dy$$

for every subset B of the 2D plane.

We can extend to multiple random variables $(X_1, \ldots, X_n)$:

$$\boxed{\mathbb{P}((X_1,\ldots,X_n) \in B) = \int\cdots\int_{(x_1,\ldots,x_n)\in B} f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)\,dx_1\ldots dx_n} \tag{1}$$

for every subset B of the n-dimensional plane.

Marginal PDF

We can also have marginal PDF through integrating some of its variables. e.g. for $f_{X,Y,Z}(x,y,z)$, we can have

$$f_X(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y,Z}(x,y,z)\,dy\,dz \tag{2}$$

$$f_{X,Y}(x,y) = \int_{-\infty}^{\infty} f_{X,Y,Z}(x,y,z)\,dz \tag{3}$$

Example

A joint Gaussian random variable has a PDF given by

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}[(x-\mu_x)^2/\sigma_x^2+(y-\mu_y)^2/\sigma_y^2]}$$

Find the marginal PDF $f_X(x)$.
(The marginal PDF $f_Y(y)$ will be similar.)

Example

ANS: We can compute the marginal PDF of a joint Gaussian as

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy \\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}[(x-\mu_x)^2/\sigma_x^2+(y-\mu_y)^2/\sigma_y^2]}\,dy \\ &= \frac{1}{\sqrt{2\pi}\sigma_x} e^{-(x-\mu_x)^2/2\sigma_x^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_y} e^{-(y-\mu_y)^2/2\sigma_y^2}\,dy \\ &= \frac{1}{\sqrt{2\pi}\sigma_x} e^{-(x-\mu_x)^2/2\sigma_x^2} \end{aligned}$$

This is also a Gaussian!

You may notice that in this example, X and Y are independent, because $f_{X,Y}(x,y) = f_X(x)\,f_Y(y)$. Later on, we will see that for Gaussian, even if the random variables are not independent, the marginal PDF will still be Gaussian.

i.i.d. random variables

Generally, dealing with multidimensional joint PDFs can be painful:

- In continuous, you do multidimensional integrations
- In discrete (e.g. in a computer), you do summation with multidimensional arrays (also known as tensors)

The process is greatly simplified if the random variables are independent:

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = f_{X_1}(x_1)\cdot f_{X_2}(x_2)\ldots f_{X_n}(x_n) \tag{4}$$

We can do integration (summation) over each random variable when needed.

i.i.d. random variables

Often, we further simplify to assume that the random variables are also "identically distributed", i.e., $f_{X_n}(x_n)$ is the same for any n. We use the shorthand "i.i.d." to denote "independent, identically distributed":

$$\boxed{f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)} \tag{5}$$

Example

Suppose we have $X_i \sim \mathcal{N}(0,1)$ for $i = 1,\ldots,n$. Find the joint PDF.

ANS: We first note that $f_{X_i}(x_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}}$. Therefore,

$$\begin{aligned} f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} \\ &= (2\pi)^{-N/2} e^{-\sum_{i=1}^{n} \frac{x_i^2}{2}} \end{aligned}$$

The simplicity of this formula justifies the frequent use of i.i.d. Gaussian in many engineering applications.

1. Independent, identically distributed random variables
2. Derived distributions
   - Laplacian distribution
   - Rayleigh distribution
3. Sum of random variables
4. Covariance and correlation
   - Covariance
   - Correlation
   - A detailed example solved with covariance
5. Gaussian random variables
   - Central limit theorem
   - Higher dimensional Gaussian
   - Generalization of Gaussian random variable

Procedure

The procedure for transforming one random variable to another is applicable to several random variables.

Given: PDF of X and Y, and $Z = g(X,Y)$, find the PDF of Z.

A two-step procedure:

$$\boxed{\begin{aligned} F_Z(z) &= \mathbb{P}(Z \leqslant z) = \mathbb{P}(g(X,Y) \leqslant z) = \iint_{\{x,y \mid g(x,y) \leqslant z\}} f_{X,Y}(x,y)\,dx\,dy \\ f_Z(z) &= \frac{d}{dz} F_Z(z) \end{aligned}} \tag{6}$$

Example

$X \sim U(0,1)$ and $Y \sim U(0,1)$, X and Y are independent, and $Z = \max\{X, Y\}$. What is the PDF of Z?

ANS: We know $\mathbb{P}(X \leqslant z) = \mathbb{P}(Y \leqslant z) = z$.

$$\begin{aligned} F_Z(z) = \mathbb{P}(\max\{X,Y\} \leqslant z) &= \mathbb{P}(X \leqslant z, Y \leqslant z) \\ &= \mathbb{P}(X \leqslant z)\,\mathbb{P}(Y \leqslant z) = z^2 \end{aligned}$$

Differentiating,

$$f_Z(z) = \begin{cases} 2z & 0 \leqslant z \leqslant 1 \\ 0 & \text{otherwise.} \end{cases}$$

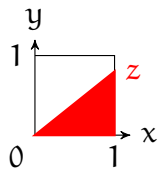
Example

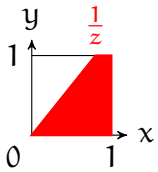
$X \sim U(0,1)$ and $Y \sim U(0,1)$, X and Y are independent, and $Z = Y/X$. What is the PDF of Z?

ANS: Case 1: $0 \leqslant z \leqslant 1$. We need to find $\mathbb{P}\left(\frac{Y}{X} \leqslant z\right) = \mathbb{P}(Y \leqslant zX)$. Given $X = x$, we have $\mathbb{P}(Y \leqslant zX) = zx$. But we need to integrate over all possible values of X. Therefore,
$\mathbb{P}(Y \leqslant zX) = \int_0^1 (zx)\,dx = \left[\frac{1}{2}zx^2\right]_0^1 = \frac{1}{2}z$.

Case 2: $z > 1$. Let $z' = 1/z$.

$$\mathbb{P}(Y/X \leqslant z) = \mathbb{P}(X/Y \geqslant z') = 1 - \mathbb{P}(X/Y \leqslant z') = 1 - \tfrac{1}{2}z' = 1 - \tfrac{1}{2z}$$





Example

Combining,

$$F_Z(z) = \mathbb{P}\left(\frac{Y}{X} \leqslant z\right) = \begin{cases} \frac{z}{2} & 0 \leqslant z \leqslant 1 \\ 1 - \frac{1}{2z} & z > 1 \\ 0 & \text{otherwise.} \end{cases}$$

Differentiating,

$$f_Z(z) = \begin{cases} \frac{1}{2} & 0 \leqslant z \leqslant 1 \\ \frac{1}{2z^2} & z > 1 \\ 0 & \text{otherwise.} \end{cases}$$

1 Independent, identically distributed random variables

2 Derived distributions
- Laplacian distribution
- Rayleigh distribution

3 Sum of random variables

4 Covariance and correlation
- Covariance
- Correlation
- A detailed example solved with covariance

5 Gaussian random variables
- Central limit theorem
- Higher dimensional Gaussian
- Generalization of Gaussian random variable

Example

$X \sim \text{Exp}(\lambda)$ and $Y \sim \text{Exp}(\lambda)$, X and Y are independent, and $Z = X - Y$. What is the PDF of Z?

ANS: Case 1: $z \geqslant 0$.

$$\begin{aligned}
F_Z(z) = \mathbb{P}(X - Y \leqslant z) &= 1 - \mathbb{P}(X - Y > z) \\
&= 1 - \int_0^\infty \left(\int_{z+y}^\infty f_{X,Y}(x,y)\, dx \right) dy \\
&= 1 - \int_0^\infty \lambda e^{-\lambda y} \left(\int_{z+y}^\infty \lambda e^{-\lambda x} dx \right) dy \\
&= 1 - \int_0^\infty \lambda e^{-\lambda y} \left(e^{-\lambda(z+y)} \right) dy \\
&= 1 - e^{-\lambda z} \int_0^\infty \lambda e^{-2\lambda y} dy \\
&= 1 - \tfrac{1}{2} e^{-\lambda z}
\end{aligned}$$

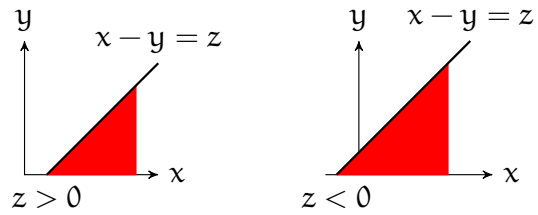
Example

Case 2: $z < 0$. Then, $-Z = Y - X$ which has the same distribution as Z.

$$F_Z(z) = \mathbb{P}(Z \leqslant z) = \mathbb{P}(-Z \geqslant -z) = \mathbb{P}(Z \geqslant -z) = 1 - F_Z(-z)$$

Since $-z > 0$, we can make use of case 1,

$$F_Z(z) = 1 - \left(1 - \tfrac{1}{2} e^{-\lambda(-z)}\right) = \tfrac{1}{2} e^{\lambda z}$$



Example

Combining,

$$F_Z(z) = \begin{cases} 1 - \frac{1}{2} e^{-\lambda z} & z \geqslant 0 \\ \frac{1}{2} e^{\lambda z} & z < 0 \end{cases}$$

Differentiating,

$$f_Z(z) = \begin{cases} \frac{\lambda}{2} e^{-\lambda z} & z \geqslant 0 \\ \frac{\lambda}{2} e^{\lambda z} & z < 0 \end{cases}$$

We can express in a single formula

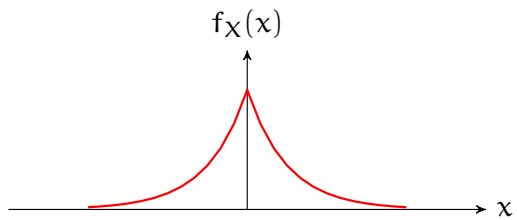
$$\boxed{f_Z(z) = \frac{\lambda}{2} e^{-\lambda |z|}} \tag{8}$$

called Laplacian random variable, and denote $Z \sim \text{Lap}(\lambda)$.

(4) Laplacian random variable

We can add Laplacian to our list of continuous random variables.

	mean: $\mathbb{E}(X)$	variance: $\text{var}(X)$
Laplacian: $X \sim \text{Lap}(\lambda)$	0	$\frac{2}{\lambda^2}$



(4) Laplacian random variable

$$\mathbb{E}(X) = 0 \qquad \text{(by symmetry)}$$

$$\begin{aligned}
\mathrm{var}(X) &= 2\left(\int_0^\infty x^2 \frac{\lambda}{2} e^{-\lambda x} dx\right) \\
&= \lambda \left[\left(-\frac{x^2}{\lambda} - \frac{2x}{\lambda^2} - \frac{2}{\lambda^3}\right) e^{-\lambda x}\right]_0^\infty \qquad \text{(from integration table)} \\
&= \frac{2}{\lambda^2}
\end{aligned}$$

1. Independent, identically distributed random variables
2. Derived distributions
   - Laplacian distribution
   - Rayleigh distribution
3. Sum of random variables
4. Covariance and correlation
   - Covariance
   - Correlation
   - A detailed example solved with covariance
5. Gaussian random variables
   - Central limit theorem
   - Higher dimensional Gaussian
   - Generalization of Gaussian random variable

Example

$X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(0,1)$, X and Y are independent, and $Z = X + iY$, where $i = \sqrt{-1}$. What are the PDFs of $|Z|$ and $\angle Z$?

ANS: We work out the solution in a few steps.

Step 1: Representing X and Y in a complex plane, we can convert to polar coordinates with random variables $R \geqslant 0$ and $\Theta \in [0, 2\pi]$, where

$$X = R\cos\Theta \qquad Y = R\sin\Theta$$

We also note that the joint PDF of X and Y is

$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}$$

Example

Step 2: From $f_{X,Y}(x,y)$ we find $F_{R,\Theta}(r,\theta)$.

For a *fixed* set of (r,θ), the CDF integrates all the points (s,ϕ) with $0 \leqslant s \leqslant r$ and $0 \leqslant \phi \leqslant \theta$, which is *a sector of a circle with radius* r *and angle* θ, denoted A.

$$\begin{aligned}
F_{R,\Theta}(r,\theta) &= \mathbb{P}(R \leqslant r, \Theta \leqslant \theta) \\
&= \iint_A \frac{1}{2\pi} e^{-(x^2+y^2)/2} dx dy \\
&= \frac{1}{2\pi} \int_0^\theta \int_0^r e^{-s^2/2} s\, ds\, d\phi
\end{aligned}$$

Example

Step 3: Differentiate $F_{R,\Theta}(r,\theta)$ to obtain $f_{R,\Theta}(r,\theta)$.

$$f_{R,\Theta}(r,\theta) = \frac{\partial^2}{\partial r \partial \theta} F_{R,\Theta}(r,\theta) = \frac{r}{2\pi} e^{-r^2/2} \qquad r \geqslant 0,\ \theta \in [0, 2\pi]$$

Step 4: Integrate joint PDF to find marginal PDF

$$f_R(r) = \int_0^{2\pi} f_{R,\Theta}(r,\theta)\, d\theta = r e^{-r^2/2} \qquad r \geqslant 0$$

$$f_\Theta(\theta) = \int_0^{\infty} \frac{r}{2\pi} e^{-r^2/2} dr = \frac{1}{2\pi}\left[-e^{-r^2/2}\right]_0^{\infty} = \frac{1}{2\pi} \qquad \theta \in [0, 2\pi]$$

In our question, $|Z| = R$ and $\angle Z = \Theta$

Example

$X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(0,1)$, X and Y are independent, and $Z = X + iY$, then

1. The angle follows a *uniform distribution*, $\Theta \sim U(0, 2\pi)$
2. The magnitude follows a distribution known as *Rayleigh distribution*, $R \sim \mathrm{Ray}(\sigma)$ where

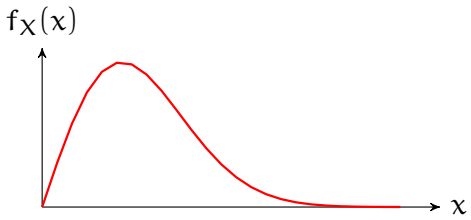
$$\boxed{f_R(r) = \frac{r}{\sigma^2} e^{-r^2/(2\sigma^2)}} \tag{9}$$

with σ^2 being the variance of the normal distributions X and Y.

(5) Rayleigh random variable

We can add Rayleigh to our list of continuous random variables.

	mean: $\mathbb{E}(X)$	variance: $\mathrm{var}(X)$
Rayleigh: $X \sim \mathrm{Ray}(\sigma)$	$\sigma\sqrt{\frac{\pi}{2}}$	$\frac{4-\pi}{2}\sigma^2$



1 Independent, identically distributed random variables
2 Derived distributions
- Laplacian distribution
- Rayleigh distribution

3 Sum of random variables
4 Covariance and correlation
- Covariance
- Correlation
- A detailed example solved with covariance

5 Gaussian random variables
- Central limit theorem
- Higher dimensional Gaussian
- Generalization of Gaussian random variable

Sum of two independent random variables

X and Y are two random variables, possibly of different distributions, but independent of each other. We are interested to know the distribution of $Z = X + Y$.

Discrete random variables:

$$\begin{aligned} p_Z(z) = \mathbb{P}(X+Y=z) &= \sum_{\{(x,y)\,|\,x+y=z\}} \mathbb{P}(X=x, Y=y) \\ &= \sum_x \mathbb{P}(X=x, Y=z-x) \\ &= \sum_x \mathbb{P}(X=x)\,\mathbb{P}(Y=z-x) \\ &= \sum_x p_X(x)\,p_Y(z-x) \end{aligned}$$

Sum of two independent random variables

Continuous random variables:

$$\begin{aligned} \mathbb{P}(Z \leqslant z \,|\, X=x) &= \mathbb{P}(X+Y \leqslant z \,|\, X=x) \\ &= \mathbb{P}(x+Y \leqslant z \,|\, X=x) \\ &= \mathbb{P}(x+Y \leqslant z) \\ &= \mathbb{P}(Y \leqslant z-x) \end{aligned}$$

Differentiating, we get $f_{Z\,|\,X}(z\,|\,x) = f_Y(z-x)$.

$$\begin{aligned} f_{X,Z}(x,z) &= f_X(x)\,f_{Z\,|\,X}(z\,|\,x) = f_X(x)\,f_Y(z-x) \\ f_Z(z) &= \int_{-\infty}^{\infty} f_{X,Z}(x,z)\,dx = \int_{-\infty}^{\infty} f_X(x)\,f_Y(z-x)\,dx \end{aligned}$$

Sum of two independent random variables

$$\text{Discrete:} \quad p_Z(z) = \sum_x p_X(x)\,p_Y(z-x) \tag{10}$$

$$\text{Continuous:} \quad f_Z(z) = \int_{-\infty}^{\infty} f_X(x)\,f_Y(z-x)\,dx \tag{11}$$

The PMF (PDF) of Z is the convolution of the PMF (PDF) of X and Y.

Example

Two random variables X and Y are independent and uniformly distributed between 0 and 1. Find the PDF of $Z = X + Y$.

We make use of the formula

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_X(x)\,f_Y(z-x)\,dx \\ &= \int_0^1 f_Y(z-x)\,dx \end{aligned}$$

since we know that only for $0 \leqslant x \leqslant 1$, we have $f_X(x) = 1$.

We also require $0 \leqslant z - x \leqslant 1$ for $f_Y(z-x) = 1$. To see the resulting limits of the integration, we divide into two cases: $0 \leqslant z \leqslant 1$ and $1 \leqslant z \leqslant 2$.

Example

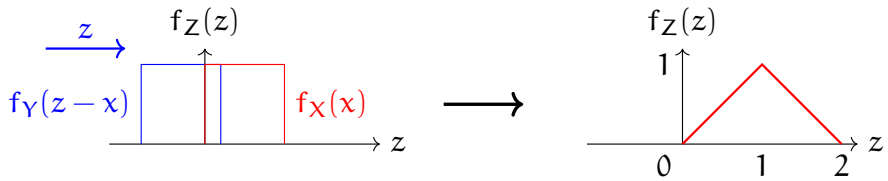
For $0 \leqslant z \leqslant 1$, we need to enforce $0 \leqslant z - x$, which means the upper limit of x can only be z:

$$f_Z(z) = \int_0^1 f_Y(z-x)\,dx = \int_0^z (1)dx = z$$

For $1 \leqslant z \leqslant 2$, we need to enforce $z - x \leqslant 1$, which means the lower limit of x can only be $z-1$:

$$f_Z(z) = \int_0^1 f_Y(z-x)\,dx = \int_{z-1}^1 (1)dx = 1-(z-1) = 2-z$$

We could have done the convolution graphically:



Example

$X \sim \text{Exp}(\lambda)$ and $Y \sim \text{Exp}(\lambda)$, X and Y are independent, and $Z = X - Y$. What is the PDF of Z?

ANS: We have already seen the answer is Laplacian. But we can also proceed by noting that $Z = X + (-Y)$, and $f_{-Y}(y) = f_Y(-y)$ by symmetry, so

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)\, f_{-Y}(z-x)\,dx = \int_{-\infty}^{\infty} f_X(x)\, f_Y(x-z)\,dx$$

Now consider $z \geqslant 0$, so $f_Y(x-z)$ is nonzero only when $x \geqslant z$.

$$f_Z(z) = \int_z^{\infty} \lambda e^{-\lambda x} \lambda e^{-\lambda(x-z)} dx = \lambda^2 e^{\lambda z} \int_z^{\infty} e^{-2\lambda x} dx$$
$$= \lambda^2 e^{\lambda z} \frac{1}{2\lambda} e^{-2\lambda z} = \frac{\lambda}{2} e^{-\lambda z}$$

The case for $z < 0$ is similar.

1. Independent, identically distributed random variables
2. Derived distributions
   - Laplacian distribution
   - Rayleigh distribution
3. Sum of random variables
4. Covariance and correlation
   - Covariance
   - Correlation
   - A detailed example solved with covariance
5. Gaussian random variables
   - Central limit theorem
   - Higher dimensional Gaussian
   - Generalization of Gaussian random variable

Definition of covariance

The covariance of two random variables X and Y is denoted by $\text{cov}(X, Y)$, and is defined by

$$\boxed{\text{cov}(X,Y) = \mathbb{E}\big[\big(X - \mathbb{E}[X]\big)\big(Y - \mathbb{E}[Y]\big)\big]} \tag{12}$$

- $\text{cov}(X,Y) = 0 \implies X$ and Y are uncorrelated.
- $\text{cov}(X,Y)$ is positive $\implies X - \mathbb{E}(X)$ and $Y - \mathbb{E}(Y)$ tend to have the same sign.
- $\text{cov}(X,Y)$ is negative $\implies X - \mathbb{E}(X)$ and $Y - \mathbb{E}(Y)$ tend to have the opposite sign.

Variance measures the "spread" of a random variable. Covariance measures the "spread" across two random variables.

Definition of covariance

A very useful alternative form for calculations:

$$\boxed{\mathrm{cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]} \tag{13}$$

Proof:

$$\begin{aligned}\mathrm{cov}(X,Y) &= \mathbb{E}\big[\big(X-\mathbb{E}[X]\big)\big(Y-\mathbb{E}[Y]\big)\big]\\ &= \mathbb{E}(XY - X\mathbb{E}[Y] - Y\mathbb{E}[X] + \mathbb{E}[X]\,\mathbb{E}[Y])\\ &= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] - \mathbb{E}[X]\,\mathbb{E}[Y] + \mathbb{E}[X]\,\mathbb{E}[Y]\\ &= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]\end{aligned}$$

Properties

Some properties of covariance (a and b are scalars):

$$\boxed{\begin{aligned}\mathrm{cov}(X,X) &= \mathrm{var}(X) &(14)\\ \mathrm{cov}(X,aY+b) &= a\cdot\mathrm{cov}(X,Y) &(15)\\ \mathrm{cov}(X,Y+Z) &= \mathrm{cov}(X,Y)+\mathrm{cov}(X,Z) &(16)\end{aligned}}$$

Also, since if X and Y are independent, $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$, so

$$\boxed{X \text{ and } Y \text{ independent} \Longrightarrow X \text{ and } Y \text{ uncorrelated}}$$

Converse is NOT true! (illustrated in the next example)

Uncorrelated does not imply independent

Consider four points, at $(1,0)$, $(0,1)$, $(-1,0)$, $(0,-1)$, each with probability $1/4$. Are the two coordinates uncorrelated? Independent?

ANS: We note that

- $\mathbb{E}(X) = \mathbb{E}(Y) = 0$
- $\mathbb{E}(XY) = 0$
- Hence, $\mathrm{cov}(X,Y) = 0$

Therefore, they are uncorrelated.

However, they are not independent because fixing Y (e.g. $Y=1$), it determines X ($X=0$).

Example

Let $X \sim U(0,2\pi)$.

- If $Y = \cos X$ and $Z = \sin X$, are Y and Z uncorrelated?
- If $Y = \cos(X/4)$ and $Z = \sin(X/4)$, are Y and Z uncorrelated?

ANS: We should check $\mathbb{E}(Y)$, $\mathbb{E}(Z)$, and $\mathbb{E}(YZ)$, to see if $\mathrm{cov}(Y,Z) = 0$. For $Y = \cos X$ and $Z = \sin X$:

$$\begin{aligned}\mathbb{E}(Y) &= \int_0^{2\pi} \frac{1}{2\pi}\cos x\,dx = 0\\ \mathbb{E}(Z) &= \int_0^{2\pi} \frac{1}{2\pi}\sin x\,dx = 0\\ \mathbb{E}(YZ) &= \int_0^{2\pi} \frac{1}{2\pi}\cos x\sin x\,dx = \int_0^{2\pi}\frac{1}{4\pi}\sin(2x)\,dx = 0\end{aligned}$$

Therefore, $\mathrm{cov}(Y,Z) = 0$, so Y and Z are uncorrelated.

Example

For $Y = \cos(X/4)$ and $Z = \sin(X/4)$:

$$\mathbb{E}(Y) = \int_0^{2\pi} \frac{1}{2\pi} \cos\frac{x}{4} dx = \frac{1}{2\pi}\left[4\sin\frac{x}{4}\right]_0^{2\pi} = \frac{2}{\pi}$$
$$\mathbb{E}(Z) = \int_0^{2\pi} \frac{1}{2\pi} \sin\frac{x}{4} dx = \frac{1}{2\pi}\left[-4\cos\frac{x}{4}\right]_0^{2\pi} = \frac{2}{\pi}$$
$$\mathbb{E}(YZ) = \int_0^{2\pi} \frac{1}{2\pi} \cos\frac{x}{4}\sin\frac{x}{4} dx = \frac{1}{4\pi}\left[-2\cos\frac{x}{2}\right]_0^{2\pi} = \frac{1}{\pi}$$

Therefore, $\mathrm{cov}(Y, Z) = \frac{1}{\pi} - \frac{4}{\pi^2}$, so Y and Z are negatively correlated.

Use of covariance

We know if X and Y are independent, $\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y)$.

With covariance, we can have a more general expression:

$$\boxed{\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\mathrm{cov}(X, Y)} \tag{17}$$

(So, $\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y)$ is also true as long as they are uncorrelated.)
Even more generally,

$$\boxed{\mathrm{var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{var}(X_i) + \sum_{\{(i,j)\,|\,i\neq j\}} \mathrm{cov}(X_i, X_j)} \tag{18}$$

Covariance

Proof: Let $\tilde{X}_i = X_i - \mathbb{E}(X_i)$.

$$\begin{aligned}
\mathrm{var}\left(\sum_{i=1}^n X_i\right) = \mathrm{var}\left(\sum_{i=1}^n \tilde{X}_i\right) &= \mathbb{E}\left[\left(\sum_{i=1}^n \tilde{X}_i\right)^2\right] \\
&= \mathbb{E}\left[\sum_{i=1}^n\sum_{j=1}^n \tilde{X}_i\tilde{X}_j\right] \\
&= \sum_{i=1}^n\sum_{j=1}^n \mathbb{E}\left[\tilde{X}_i\tilde{X}_j\right] \\
&= \sum_{i=1}^n \mathbb{E}\left[\tilde{X}_i^2\right] + \sum_{\{(i,j)\,|\,i\neq j\}} \mathbb{E}\left[\tilde{X}_i\tilde{X}_j\right] \\
&= \sum_{i=1}^n \mathrm{var}(X_i) + \sum_{\{(i,j)\,|\,i\neq j\}} \mathrm{cov}(X_i, X_j)
\end{aligned}$$

1. Independent, identically distributed random variables
2. Derived distributions
   - Laplacian distribution
   - Rayleigh distribution
3. Sum of random variables
4. Covariance and correlation
   - Covariance
   - Correlation
   - A detailed example solved with covariance
5. Gaussian random variables
   - Central limit theorem
   - Higher dimensional Gaussian
   - Generalization of Gaussian random variable

Definition

Often, we work with a "normalized" version of covariance, known as correlation coefficient:

$$\boxed{\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)\,\mathrm{var}(Y)}}} \tag{19}$$

Assuming X and Y both have nonzero variance, the numerator determines similar properties as covariance:

- $\rho(X,Y) = 0 \implies X$ and Y are uncorrelated.
- $\rho(X,Y)$ is positive $\implies X - \mathbb{E}(X)$ and $Y - \mathbb{E}(Y)$ tend to have the same sign.
- $\rho(X,Y)$ is negative $\implies X - \mathbb{E}(X)$ and $Y - \mathbb{E}(Y)$ tend to have the opposite sign.

Definition

$\rho(X,Y)$ is normalized in the sense that

$$\boxed{-1 \leqslant \rho(X,Y) \leqslant 1} \tag{20}$$

- $|\rho|$ is a measure of the extent to which $X - \mathbb{E}(X)$ and $Y - \mathbb{E}(Y)$ are "correlated" (i.e., cluster together).
- $|\rho| = 1$ if and only if

$$Y - \mathbb{E}[Y] = c\big(X - \mathbb{E}[X]\big)$$

where c is a constant of the same sign as ρ

Proof of correlation coefficient magnitude (1)

We start with a lemma known as Schwarz inequality:

$$\boxed{\big(\mathbb{E}[XY]\big)^2 \leqslant \mathbb{E}\big[X^2\big]\,\mathbb{E}\big[Y^2\big]} \tag{21}$$

for any random variable X and Y.

Proof: We assume $\mathbb{E}\big[Y^2\big] \neq 0$, because otherwise we have $Y = 0$ with probability 1, and therefore $\mathbb{E}[XY] = 0$, so equality holds. With this assumption, we start with an expression

$$\mathbb{E}\left[\left(X - \frac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]}Y\right)^2\right]$$

which must be $\geqslant 0$.

Proof of correlation coefficient magnitude (2)

Proof:

$$\begin{aligned}
0 &\leqslant \mathbb{E}\left[\left(X - \frac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]}Y\right)^2\right] \\
&= \mathbb{E}\left[X^2 - 2\frac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]}XY + \frac{(\mathbb{E}[XY])^2}{(\mathbb{E}[Y^2])^2}Y^2\right] \\
&= \mathbb{E}\big[X^2\big] - 2\frac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]}\mathbb{E}[XY] + \frac{(\mathbb{E}[XY])^2}{(\mathbb{E}[Y^2])^2}\mathbb{E}\big[Y^2\big] \\
&= \mathbb{E}\big[X^2\big] - \frac{(\mathbb{E}[XY])^2}{\mathbb{E}[Y^2]}
\end{aligned}$$

Therefore, $\big(\mathbb{E}[XY]\big)^2 \leqslant \mathbb{E}\big[X^2\big]\,\mathbb{E}\big[Y^2\big]$, and thus $\dfrac{\big(\mathbb{E}[XY]\big)^2}{\mathbb{E}[X^2]\,\mathbb{E}[Y^2]} \leqslant 1$.

Proof of correlation coefficient magnitude (3)

For general random variables X and Y, we first "center" them to form

$$\tilde{X} = X - \mathbb{E}[X]$$
$$\tilde{Y} = Y - \mathbb{E}[Y]$$

- $\mathrm{var}\left(\tilde{X}\right) = \mathrm{var}(X) = \mathbb{E}\left[\tilde{X}^2\right]$
- $\mathrm{cov}(X,Y) = \mathbb{E}\left[\left(X - \mathbb{E}[X]\right)\left(Y - \mathbb{E}[Y]\right)\right] = \mathbb{E}\left[\tilde{X}\tilde{Y}\right]$

Then, we make use of Schwarz inequality

$$\left(\rho(X,Y)\right)^2 = \frac{\left(\mathrm{cov}(X,Y)\right)^2}{\mathrm{var}(X)\,\mathrm{var}(Y)} = \frac{\left(\mathbb{E}\left[\tilde{X}\tilde{Y}\right]\right)^2}{\mathbb{E}\left[\tilde{X}^2\right]\mathbb{E}\left[\tilde{Y}^2\right]} \leqslant 1$$

So $|\rho(X,Y)| \leqslant 1$.

Proof of correlation coefficient magnitude (4)

Next, we show what happens when $Y - \mathbb{E}[Y] = c\left(X - \mathbb{E}[X]\right)$, or $\tilde{Y} = c\tilde{X}$:

- $\mathbb{E}\left(\tilde{X}\tilde{Y}\right) = c\mathbb{E}\left(\tilde{X}^2\right)$
- $\mathbb{E}\left(\tilde{Y}^2\right) = c^2\mathbb{E}\left(\tilde{X}^2\right)$

Therefore,

$$\rho(X,Y) = \frac{c\mathbb{E}\left[\tilde{X}^2\right]}{\sqrt{c^2\mathbb{E}\left[\tilde{X}^2\right]\mathbb{E}\left[\tilde{X}^2\right]}} = \frac{c}{|c|} = \begin{cases} 1 & c > 0 \\ -1 & c < 0 \end{cases}$$

Proof of correlation coefficient magnitude (5)

We now show the reverse: when $\rho(X,Y) = \pm 1$,

$$\mathbb{E}\left[\left(\tilde{X} - \frac{\mathbb{E}\left[\tilde{X}\tilde{Y}\right]}{\mathbb{E}\left[\tilde{Y}^2\right]}\tilde{Y}\right)^2\right] = \mathbb{E}\left[\tilde{X}^2\right] - \frac{\left(\mathbb{E}\left[\tilde{X}\tilde{Y}\right]\right)^2}{\mathbb{E}\left[\tilde{Y}^2\right]}$$
$$= \mathbb{E}\left[\tilde{X}^2\right]\left(1 - [\rho(X,Y)]^2\right) = 0$$

This means, with probability 1, $\tilde{X} - \frac{\mathbb{E}[\tilde{X}\tilde{Y}]}{\mathbb{E}[\tilde{Y}^2]}\tilde{Y}$ is equal to zero. It follows that, with probability 1,

$$\tilde{X} = \frac{\mathbb{E}\left[\tilde{X}\tilde{Y}\right]}{\mathbb{E}\left[\tilde{Y}^2\right]}\tilde{Y} = \sqrt{\frac{\mathbb{E}\left[\tilde{X}^2\right]}{\mathbb{E}\left[\tilde{Y}^2\right]}}\rho(X,Y)\tilde{Y}$$

i.e., the ratio of $\tilde{X}$ and $\tilde{Y}$ is determined by the sign of $\rho(X,Y)$.

Example

For n independent coin toss, let X be the number of heads and Y be the number of tails. What is the correlation coefficient $\rho(X,Y)$?

ANS: Since $X + Y = n$, they are "perfectly correlated." We expect $\rho(X,Y) = \pm 1$. Moreover, since $X + Y = \mathbb{E}(X) + \mathbb{E}(Y) = n$, we have

$$X - \mathbb{E}(X) = -\left(Y - \mathbb{E}(Y)\right)$$

so the sign of $\rho(X,Y)$ should be negative, i.e., it is equal to -1.

Alternatively, we can apply the formula

$$\mathrm{cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = -\mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = -\mathrm{var}(X)$$

and note that $\mathrm{var}(X) = \mathrm{var}(Y)$, by symmetry, and therefore

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)\,\mathrm{var}(Y)}} = \frac{-\mathrm{var}(X)}{\sqrt{\mathrm{var}(X)\,\mathrm{var}(X)}} = -1$$

Computing correlation from data

(From supplementary text) Suppose we have a set of observation data $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$ where x_i comes from a random variable X, and y_i comes from a random variable Y. We want to know if X and Y are correlated, or more precisely, we want to compute the correlation coefficient $\rho(X, Y)$ from the empirical data.

To do so, we compute

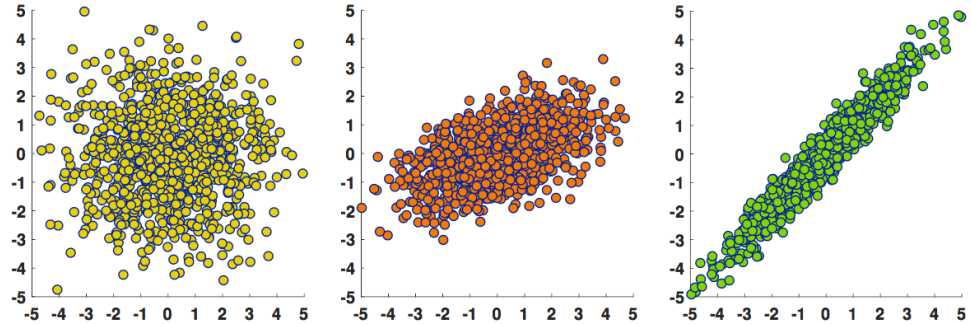
$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

$$\widehat{\rho} = \frac{\frac{1}{N}\sum_{i=1}^{N} x_i y_i - \bar{x}\bar{y}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2}}$$

Computing correlation from data

Examples:



(a) $\widehat{\rho} = -0.0038$ (b) $\widehat{\rho} = 0.5321$ (c) $\widehat{\rho} = 0.9656$

(a) is uncorrelated; (b) is mildly correlated; (c) is highly correlated.

1. Independent, identically distributed random variables
2. Derived distributions
   - Laplacian distribution
   - Rayleigh distribution
3. Sum of random variables
4. Covariance and correlation
   - Covariance
   - Correlation
   - A detailed example solved with covariance
5. Gaussian random variables
   - Central limit theorem
   - Higher dimensional Gaussian
   - Generalization of Gaussian random variable

Example

In a class with n students, after the final exam, they were ranked from 1 to n (no two students shared the same rank). The names and the marks were put in a spreadsheet, but a careless teacher sorted the names in some random way without linking them to the marks. Consequently, the matching between the student and his or her actual rank became random. For the student originally with rank k, the new rank now takes a discrete random variable which is uniform between 1 and n. Note that in the new rank, again no two students share the same rank.

What is the expected number of correct ranking (i.e., new rank is the same as original rank), and its variance?

This question is a work of fiction. Any resemblance to actual persons, living or dead, or actual events is purely coincidental.

Example

Attempt #1: Test some small cases. Let X = correct ranking.

1. Two students AB; after the randomization, half the time the order remains AB; half the time the order becomes BA.

$$\begin{aligned}\mathbb{E}(X) &= \tfrac{1}{2}\cdot(2)+\tfrac{1}{2}\cdot(0)=1\\ \mathbb{E}(X^2) &= \tfrac{1}{2}\cdot(2)^2+\tfrac{1}{2}\cdot(0)^2=2\\ \mathrm{var}(X) &= 2-1^2=1\end{aligned}$$

2. Three students ABC; There are now 3! permutations, with 1/6 being all correct, 3/6 having one correct, and 2/6 being all wrong.

$$\begin{aligned}\mathbb{E}(X) &= \tfrac{1}{6}\cdot(3)+\tfrac{3}{6}\cdot(1)+\tfrac{2}{6}\cdot(0)=1\\ \mathbb{E}(X^2) &= \tfrac{1}{6}\cdot(3)^2+\tfrac{3}{6}\cdot(1)^2+\tfrac{2}{6}\cdot(0)^2=2\\ \mathrm{var}(X) &= 2-1^2=1\end{aligned}$$

Guess: $\mathbb{E}(X)=\mathrm{var}(X)=1$ *for all* n*?*

Example

Attempt #2: Analytical derivation.

Let $X_i = 1$ if the ith student has the correct rank, and zero otherwise. So,

$$X = X_1 + X_2 + \ldots + X_n$$

For each X_i, we have $\mathbb{P}(X_i = 1) = 1/n$; therefore,

$$\begin{aligned}\mathbb{E}(X_i) &= \tfrac{1}{n}\cdot 1+\tfrac{n-1}{n}\cdot 0=\tfrac{1}{n}\\ \mathbb{E}(X) &= \mathbb{E}(X_1)+\ldots+\mathbb{E}(X_n)=\tfrac{1}{n}+\ldots+\tfrac{1}{n}=1\end{aligned}$$

Hence, although each one having a correct rank decreases with n, there are more students, and the net result is that the expected number of correctness remains 1.

Example

The calculation of the variance is more complicated because X_i and X_j are correlated, for $i \neq j$. First,

$$\mathrm{var}(X_i) = \frac{1}{n}\left(1-\frac{1}{n}\right) \qquad \text{(Bernoulli)}$$

Then, let us calculate $\mathbb{E}(X_iX_j)$ for $i \neq j$.

$$\begin{aligned}\mathbb{E}(X_iX_j) &= \mathbb{P}(X_i = 1 \text{ and } X_j = 1)\\ &= \mathbb{P}(X_i = 1)\,\mathbb{P}(X_j = 1 \mid X_i = 1)\\ &= \frac{1}{n}\cdot\frac{1}{n-1} = \frac{1}{n(n-1)}\end{aligned}$$

Hence,

$$\mathrm{cov}(X_i, X_j) = \mathbb{E}(X_iX_j) - \mathbb{E}(X_i)\,\mathbb{E}(X_j) = \frac{1}{n(n-1)} - \frac{1}{n}\cdot\frac{1}{n} = \frac{1}{n^2(n-1)}$$

Example

Overall,

$$\begin{aligned}\mathrm{var}(X) = \mathrm{var}\left(\sum_{i=1}^{n} X_i\right) &= \sum_{i=1}^{n}\mathrm{var}(X_i) + \sum_{\{(i,j)\,|\,i\neq j\}}\mathrm{cov}(X_i, X_j)\\ &= n\left[\frac{1}{n}\left(1-\frac{1}{n}\right)\right] + n(n-1)\left[\frac{1}{n^2(n-1)}\right]\\ &= 1\end{aligned}$$

Hence, the variance also remains 1 irrespective of the number of students.

Quite surprising that irrespective of n*, you always expect to get* 1 *correct and all the rest wrong, and even with the same variance!*

1 Independent, identically distributed random variables
2 Derived distributions
- Laplacian distribution
- Rayleigh distribution
3 Sum of random variables
4 Covariance and correlation
- Covariance
- Correlation
- A detailed example solved with covariance
5 Gaussian random variables
- Central limit theorem
- Higher dimensional Gaussian
- Generalization of Gaussian random variable

More about Gaussian

We have already come across Gaussian r.v. with parameters $(\mu, \sigma > 0)$ has PDF

$$\boxed{f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}} \tag{22}$$

Sometimes, we use a parameter $\beta > 0$, which is the inverse variance such that $\beta = 1/\sigma^2$, and call it *precision*:

$$\boxed{f_X(x) = \sqrt{\frac{\beta}{2\pi}} e^{-\frac{1}{2}\beta(x-\mu)^2}} \tag{23}$$

Ubiquity of Gaussian random variables

A very important mathematical result supporting the general use of Gaussian random variables is the central limit theorem.

Let $X_1, X_2, \ldots$ be a sequence of n independent, identically distributed random variables with mean μ and variance σ^2. We define

$$\boxed{Z_n = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}} \tag{24}$$

The central limit theorem states that Z_n "converges" to $\mathcal{N}(0, 1)$.

Central limit theorem

We can easily show that

$$\mathbb{E}(Z_n) = \frac{\mathbb{E}(X_1 + \cdots + X_n) - n\mu}{\sigma\sqrt{n}} = 0$$

$$\mathrm{var}(Z_n) = \frac{\mathrm{var}(X_1 + \cdots + X_n)}{\sigma^2 n} = \frac{n\sigma^2}{n\sigma^2} = 1$$

Also, the "convergence" is in the technical sense that the CDF of Z_n converges to the standard normal CDF,

$$\lim_{n\to\infty} \mathbb{P}(Z_n \leqslant z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2} dx$$

as n approaches infinity. We will not be showing the proof here (which normally uses something called the moment generating function).

Central limit theorem

Normal Approximation Based on the Central Limit Theorem

Let $S_n = X_1 + \cdots + X_n$, where the X_i are independent identically distributed random variables with mean μ and variance σ^2. If n is large, the probability $\mathbb{P}(S_n \leqslant c)$ can be approximated by treating S_n as if it were normal, according to the following procedure:

1. Calculate the mean $n\mu$ and the variance $n\sigma^2$ of S_n.
2. Calculate the normalized value $z = (c - n\mu)/\sigma\sqrt{n}$.
3. Use the approximation

$$\mathbb{P}(S_n \leqslant c) \approx \Phi(z)$$

where $\Phi(z)$ is available from standard normal CDF.

Example

We have 100 packages with independent weights uniformly distributed between 5 and 50 kilograms. What is the (approximate) probability that the total weight exceeds 3000 kilograms?

ANS: Each package has weight X_i, and $S_{100} = X_1 + \ldots + X_{100}$.

1. $n\mu = 100 \times (5+50)/2 = 2750$; $n\sigma^2 = 100 \times (50-5)^2/12 = 16875$
2. $z = \dfrac{3000 - 2750}{\sqrt{16875}} = \dfrac{250}{129.9} = 1.92$
3. $\mathbb{P}(S_{100} > 3000) = 1 - \mathbb{P}(S_{100} \leqslant 3000) \approx 1 - \Phi(1.92) = 0.0274$

Example

A 3D printer prints out different designs in an amount of time that is uniformly distributed between 1 and 5 hours, independent of each other. Find the (approximate) probability that the number of parts processed within 320 hours, denoted by N_{320}, is at least 100.

ANS: Each design takes X_i, and $S_{100} = X_1 + \ldots + X_{100}$. Note that the events $\{N_{320} \geqslant 100\}$ and $\{S_{100} \leqslant 320\}$ are equivalent.

1. $n\mu = 100 \times (1+5)/2 = 300$; $n\sigma^2 = 100 \times (5-1)^2/12 = 400/3$
2. $z = \dfrac{320 - 300}{\sqrt{400/3}} = 1.73$
3. $\mathbb{P}(S_{100} \leqslant 320) \approx \Phi(1.73) = 0.9582$

1. Independent, identically distributed random variables
2. Derived distributions
   - Laplacian distribution
   - Rayleigh distribution
3. Sum of random variables
4. Covariance and correlation
   - Covariance
   - Correlation
   - A detailed example solved with covariance
5. Gaussian random variables
   - Central limit theorem
   - Higher dimensional Gaussian
   - Generalization of Gaussian random variable

Higher dimensional Gaussian

Often, we can deal with random variables that are jointly Gaussian. Suppose we have two independent Gaussian random variables X and Y:

$$f_{X,Y}(x,y) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-(x-\mu_x)^2/2\sigma_x^2} \frac{1}{\sqrt{2\pi}\sigma_y} e^{-(y-\mu_y)^2/2\sigma_y^2}$$
$$= \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}[(x-\mu_x)^2/\sigma_x^2+(y-\mu_y)^2/\sigma_y^2]}$$

The *contour lines* of this two-dimensional plots are concentric circles if $\sigma_x = \sigma_y$. Otherwise, they are ellipses.

Higher dimensional Gaussian

They are jointly Gaussian even if they are not independent! Suppose they have a correlation denoted by ρ. The joint PDF is

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho\frac{(x-\mu_x)}{\sigma_x}\frac{(y-\mu_y)}{\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right]}$$

The *contour lines* of this two-dimensional plots are in general ellipses that may be tilted at an angle with respect to the axes.

The *marginal distribution* is once again Gaussian!

Higher dimensional Gaussian

The beauty (if you agree) of Gaussian random variables is that this can be generalized to even higher dimensions, called multivariate Gaussian. We can write in this compact form:

$$\boxed{f_X(\mathbf{x}) = \sqrt{\frac{1}{(2\pi)^n \det(\mathbf{\Sigma})}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}} \quad (25)$$

- $\mathbf{x}$: vector of observations
- $\boldsymbol{\mu}$: vector of the means
- $\mathbf{\Sigma}$: covariance matrix, which is symmetric positive definite

The marginal distributions are also multivariate Gaussian!

Higher dimensional Gaussian

Note the need to invert the covariance matrix above. An alternative is to use a precision matrix $\boldsymbol{\beta}$ as in the 1D case:

$$\boxed{f_X(\mathbf{x}) = \sqrt{\frac{\det(\boldsymbol{\beta})}{(2\pi)^n}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\beta}(\mathbf{x}-\boldsymbol{\mu})}} \quad (26)$$

Once again, the multivariate Gaussian is completely determined by $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ or $\boldsymbol{\beta}$, and we can efficiently find the marginals from these quantities. This computational efficiency, together with the central limit theorem, justifies the very frequent use of multivariate Gaussian when dealing with high dimensional data.

Higher dimensional random variables

We can have relationships on the expectation and variance for higher dimension random variables. The following is true not just for multivariate Gaussian.

Remember that when X and Y are scalar random variable, and X has mean μ and variance σ^2, then if $Y = aX + b$, we have

$$\mathbb{E}(Y) = a\mu + b$$
$$\text{var}(Y) = a^2\sigma^2$$

When X is a vector, and $Y = \mathbf{a}^\mathsf{T} X + b$ (where $\mathbf{a}$ is a vector of coefficients), and X has mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, we have

$$\mathbb{E}(Y) = \mathbf{a}^\mathsf{T}\boldsymbol{\mu} + b$$
$$\text{var}(Y) = \mathbf{a}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{a}$$

Higher dimensional random variables

When X is a vector, and $Y = \mathbf{A}X + \mathbf{b}$ (where $\mathbf{A}$ is a matrix and $\mathbf{b}$ is a vector), and X has mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, we have

$$\mathbb{E}(Y) = \mathbf{A}\boldsymbol{\mu} + \mathbf{b}$$
$$\text{cov}(Y) = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\mathsf{T}$$

Since (multivariate) Gaussian is *completely determined* by its mean and (co-)variance, these expressions are useful for any linear transformation of (multivariate) Gaussian random variables.

1. Independent, identically distributed random variables
2. Derived distributions
   - Laplacian distribution
   - Rayleigh distribution
3. Sum of random variables
4. Covariance and correlation
   - Covariance
   - Correlation
   - A detailed example solved with covariance
5. Gaussian random variables
   - Central limit theorem
   - Higher dimensional Gaussian
   - Generalization of Gaussian random variable

Generalized Gaussian

While Gaussian distribution is extremely useful, sometimes we deal with situations where there are "heavier tails" (more likely to be far from the center) or "lighter tails". In such cases, we can use the generalized Gaussian distribution:

$$\boxed{f_X(x) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|x-\mu|/\alpha)^\beta}} \tag{27}$$

- μ: mean
- α: scale (positive, real)
- β: shape (positive, real)
- $\Gamma(\cdot)$: Gamma function. (Remember $\Gamma(n) = (n-1)!$ for integer values of n, but it "interpolates" the non-integer values.)

It may look complicated, but let's first consider some special cases.

Generalized Gaussian

Case I: $\beta = 2$

It is known that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. Then,

$$\begin{aligned} f_X(x) &= \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|x-\mu|/\alpha)^\beta} \\ &= \frac{2}{2\alpha\sqrt{\pi}} e^{-(|x-\mu|/\alpha)^2} \\ &= \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \end{aligned}$$

by letting $\sigma = \alpha/\sqrt{2}$. Therefore, when $\beta = 2$, the generalized Gaussian distribution becomes a Gaussian distribution with mean μ and variance $\alpha^2/2$.

Generalized Gaussian

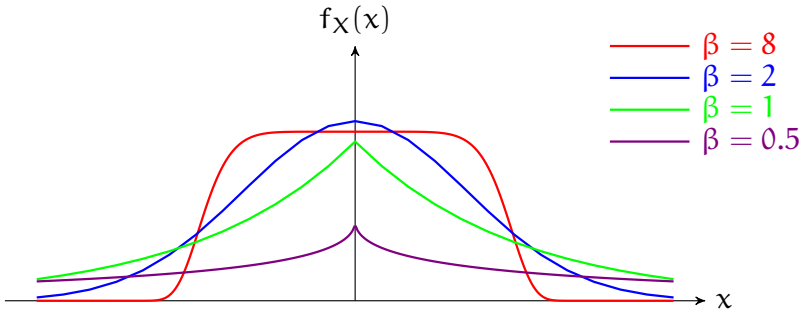
Case II: $\beta = 1$

$$\begin{aligned} f_X(x) &= \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|x-\mu|/\alpha)^\beta} \\ &= \frac{1}{2\alpha(0!)} e^{-(|x-\mu|/\alpha)} \\ &= \frac{\lambda}{2} e^{-\lambda|x-\mu|} \end{aligned}$$

by letting $\lambda = 1/\alpha$. Therefore, when $\beta = 1$, the generalized Gaussian distribution becomes a Laplacian distribution with mean μ and variance $2\alpha^2$.

Generalized Gaussian

Let's plot for different values of β: (assuming $\mu = 0$)



- As $\beta \to \infty$, the distribution approaches $U(\mu - \alpha, \mu + \alpha)$
- It has increasingly heavier tail when β gets small
- It has increasingly lighter tail when β gets large
- The measure of "tailedness" is called kurtosis

Conclusions

By now, we have covered both basic and several advanced topics dealing with discrete and continuous random variables, including cases involving multiple random variables and their interactions.