

Several Random Variables

Dr. Edmund Lam

Department of Electrical and Electronic Engineering
The University of Hong Kong

ELEC2844: Probabilistic Systems Analysis
(Second Semester, 2020–21)

<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 multiple 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
- Bayes' rule

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

- 1 Derived distributions
- 2 Sum of random variables
- 3 Covariance and correlation
 - Covariance
 - A detailed example
 - Correlation
- 4 Gaussian random variables
 - Central limit theorem
 - Higher dimensional Gaussian
 - Generalization of Gaussian random variable
- 5 (Advanced Topic) Expectation and variance
 - Iterated expectation
 - Total variance
- 6 (Advanced Topic) Random number of independent random variables

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 .

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:

$$F_Z(z) = \mathbb{P}(Z \leq z) = \mathbb{P}(g(X, Y) \leq z) = \iint_{\{x, y \mid g(x, y) \leq z\}} f_{X,Y}(x, y) \, dx dy$$

$$f_Z(z) = \frac{d}{dz} F_Z(z)$$

(1)

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 ?

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 $\mathbb{P}(X \leq z) = \mathbb{P}(Y \leq z) = z$.

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

Differentiating,

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

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 ?

Example

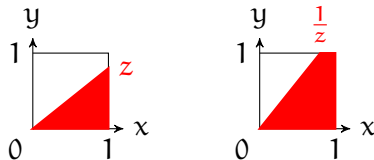
$X \sim U(0, 1)$ and $Y \sim U(0, 1)$, X and Y are independent, and $Z = Y/X$. What is the PDF of Z ?

ANS: Case 1: $0 \leq z \leq 1$. We need to find $\mathbb{P}(\frac{Y}{X} \leq z) = \mathbb{P}(Y \leq zX)$. Given $X = x$, we have $\mathbb{P}(Y \leq zX) = zx$. But we need to integrate over all possible values of X . Therefore,

$$\mathbb{P}(Y \leq zX) = \int_0^1 (zx) dx = \left[\frac{1}{2}zx^2 \right]_0^1 = \frac{1}{2}z.$$

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

$$\mathbb{P}(Y/X \leq z) = \mathbb{P}(X/Y \geq z') = 1 - \mathbb{P}(X/Y \leq z') = 1 - \frac{1}{2}z' = 1 - \frac{1}{2z}$$



Example

Combining,

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

Differentiating,

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

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 ?

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: $z \geq 0$.

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

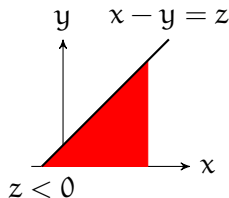
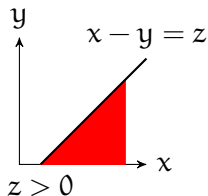
Example

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

$$F_Z(z) = \mathbb{P}(Z \leq z) = \mathbb{P}(-Z \geq -z) = \mathbb{P}(Z \geq -z) = 1 - F_Z(-z)$$

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

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



Example

Combining,

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

Differentiating,

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

We can express in a single formula

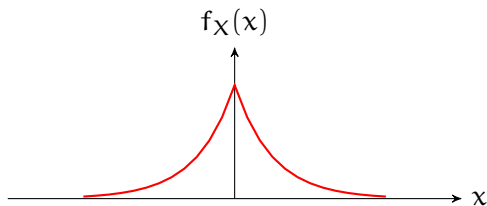
$$f_Z(z) = \frac{\lambda}{2}e^{-\lambda|z|} \quad (3)$$

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

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

$$\text{var}(X) = 2 \left(\int_0^{\infty} x^2 \frac{\lambda}{2} e^{-\lambda x} dx \right) = \lambda \left[\left(-\frac{x^2}{\lambda} - \frac{2x}{\lambda^2} - \frac{2}{\lambda^3} \right) e^{-\lambda x} \right]_0^{\infty} = \frac{2}{\lambda^2}$$

$$\begin{aligned} M_X(s) &= \int_{-\infty}^{\infty} e^{sx} \frac{\lambda}{2} e^{-\lambda|x|} dx \\ &= \frac{\lambda}{2} \left[\int_{-\infty}^0 e^{sx} e^{\lambda x} dx + \int_0^{\infty} e^{sx} e^{-\lambda x} dx \right] \\ &= \frac{\lambda}{2} \left\{ \left[\frac{1}{s+\lambda} e^{(s+\lambda)x} \right]_{-\infty}^0 + \left[\frac{1}{s-\lambda} e^{(s-\lambda)x} \right]_0^{\infty} \right\} \\ &= \frac{\lambda}{2} \left(\frac{1}{s+\lambda} - \frac{1}{s-\lambda} \right) = \frac{\lambda^2}{\lambda^2 - s^2} \quad \text{where } |s| < \lambda \end{aligned}$$

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 $\angle Z$?

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 $R \geq 0$ and $\Theta \in [0, 2\pi]$, where

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

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 $0 \leq s \leq r$ and $0 \leq \phi \leq \theta$, which is *a sector of a circle with radius r and angle θ* , denoted A .

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

Example

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

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

Step 4: Integrate joint PDF to find marginal PDF

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

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

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*, $\Theta \sim \mathcal{U}(0, 2\pi)$
- ② The magnitude follows a distribution known as *Rayleigh distribution*, $R \sim \text{Ray}(\sigma)$ where

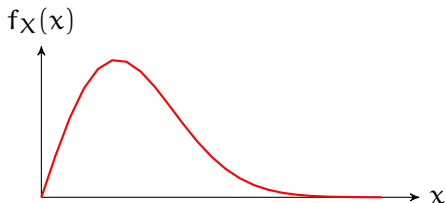
$$f_R(r) = \frac{r}{\sigma^2} e^{-r^2/(2\sigma^2)} \quad (4)$$

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: $\text{var}(X)$
Rayleigh: $X \sim \text{Ray}(\sigma)$	$\sigma\sqrt{\frac{\pi}{2}}$	$\frac{4 - \pi}{2}\sigma^2$



- 1 Derived distributions
- 2 Sum of random variables**
- 3 Covariance and correlation
 - Covariance
 - A detailed example
 - Correlation
- 4 Gaussian random variables
 - Central limit theorem
 - Higher dimensional Gaussian
 - Generalization of Gaussian random variable
- 5 (Advanced Topic) Expectation and variance
 - Iterated expectation
 - Total variance
- 6 (Advanced Topic) Random number of independent random variables

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:

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

Sum of two independent random variables

Continuous random variables:

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

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

$$\text{Discrete:} \quad p_Z(z) = \sum_x p_X(x) p_Y(z-x) \quad (5)$$

$$\text{Continuous:} \quad f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx \quad (6)$$

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

Recall we have also looked at the moment generating functions:

$$M_Z(s) = \mathbb{E}(e^{sZ}) = \mathbb{E}(e^{s(X+Y)}) = \mathbb{E}(e^{sX}e^{sY}) = \mathbb{E}(e^{sX}) \mathbb{E}(e^{sY})$$

$$M_Z(s) = M_X(s)M_Y(s) \quad (7)$$

The moment generating function (MGF) of Z is the **product** of the MGF of X and Y .

Example

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

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 $0 \leq x \leq 1$, we have $f_X(x) = 1$.

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

Example

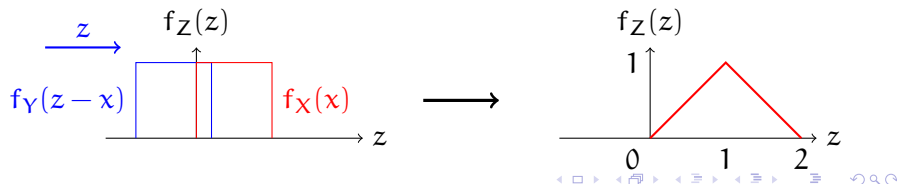
For $0 \leq z \leq 1$, we need to enforce $0 \leq z - x$, which means the upper limit of x can only be z :

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

For $1 \leq z \leq 2$, we need to enforce $z - x \leq 1$, 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 ?

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

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

The case for $z < 0$ is similar.

- 1 Derived distributions
- 2 Sum of random variables
- 3 Covariance and correlation
 - Covariance
 - A detailed example
 - Correlation
- 4 Gaussian random variables
 - Central limit theorem
 - Higher dimensional Gaussian
 - Generalization of Gaussian random variable
- 5 (Advanced Topic) Expectation and variance
 - Iterated expectation
 - Total variance
- 6 (Advanced Topic) Random number of independent random variables

Definition

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

$$\text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \quad (8)$$

- $\text{cov}(X, Y) = 0 \implies X$ 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

Alternative form:

$$\text{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X] \mathbb{E}[Y] \quad (9)$$

Proof:

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

Properties

Some properties (a and b are scalars):

$$\text{cov}(X, X) = \text{var}(X) \quad (10)$$

$$\text{cov}(X, aY + b) = a \cdot \text{cov}(X, Y) \quad (11)$$

$$\text{cov}(X, Y + Z) = \text{cov}(X, Y) + \text{cov}(X, Z) \quad (12)$$

Properties

Some properties (a and b are scalars):

$$\text{cov}(X, X) = \text{var}(X) \quad (10)$$

$$\text{cov}(X, aY + b) = a \cdot \text{cov}(X, Y) \quad (11)$$

$$\text{cov}(X, Y + Z) = \text{cov}(X, Y) + \text{cov}(X, Z) \quad (12)$$

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

X and Y independent $\implies X$ and Y uncorrelated

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

Example

Consider four points, at $(1, 0)$, $(0, 1)$, $(-1, 0)$, $(0, -1)$, each with probability $1/4$.

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

However,

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

Covariance

If X and Y are independent (uncorrelated),
 $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$. But more generally,

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y) \quad (13)$$

Covariance

If X and Y are independent (uncorrelated),
 $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$. But more generally,

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y) \quad (13)$$

Even more generally,

$$\text{var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{var}(X_i) + \sum_{\{(i,j) \mid i \neq j\}} \text{cov}(X_i, X_j) \quad (14)$$

Covariance

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

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

- 1 Derived distributions
- 2 Sum of random variables
- 3 **Covariance and correlation**
 - Covariance
 - **A detailed example**
 - Correlation
- 4 Gaussian random variables
 - Central limit theorem
 - Higher dimensional Gaussian
 - Generalization of Gaussian random variable
- 5 (Advanced Topic) Expectation and variance
 - Iterated expectation
 - Total variance
- 6 (Advanced Topic) Random number of independent random variables

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.

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

$$\mathbb{E}(X) = \frac{1}{2} \cdot (2) + \frac{1}{2} \cdot (0) = 1$$

$$\mathbb{E}(X^2) = \frac{1}{2} \cdot (2)^2 + \frac{1}{2} \cdot (0)^2 = 2$$

$$\text{var}(X) = 2 - 1^2 = 1$$

- 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.

$$\mathbb{E}(X) = \frac{1}{6} \cdot (3) + \frac{3}{6} \cdot (1) + \frac{2}{6} \cdot (0) = 1$$

$$\mathbb{E}(X^2) = \frac{1}{6} \cdot (3)^2 + \frac{3}{6} \cdot (1)^2 + \frac{2}{6} \cdot (0)^2 = 2$$

$$\text{var}(X) = 2 - 1^2 = 1$$

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

Example

Attempt #2: Analytical derivation.

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

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

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

$$\mathbb{E}(X_i) = \frac{1}{n} \cdot 1 + \frac{n-1}{n} \cdot 0 = \frac{1}{n}$$

$$\mathbb{E}(X) = \mathbb{E}(X_1) + \dots + \mathbb{E}(X_n) = \frac{1}{n} + \dots + \frac{1}{n} = 1$$

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,

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

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

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

Hence,

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

Example

Overall,

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

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 Derived distributions
- 2 Sum of random variables
- 3 Covariance and correlation**
 - Covariance
 - A detailed example
 - **Correlation**
- 4 Gaussian random variables
 - Central limit theorem
 - Higher dimensional Gaussian
 - Generalization of Gaussian random variable
- 5 (Advanced Topic) Expectation and variance
 - Iterated expectation
 - Total variance
- 6 (Advanced Topic) Random number of independent random variables

Definition

Often, we work with a “normalized” version of covariance, known as **correlation coefficient**:

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \text{var}(Y)}} \quad (15)$$

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

$$-1 \leq \rho(X, Y) \leq 1 \quad (16)$$

- $|\rho|$ 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

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

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:

$$(\mathbb{E}[XY])^2 \leq \mathbb{E}[X^2] \mathbb{E}[Y^2] \quad (17)$$

for any random variable X and Y .

Proof of correlation coefficient magnitude (1)

We start with a lemma known as Schwarz inequality:

$$(\mathbb{E}[XY])^2 \leq \mathbb{E}[X^2] \mathbb{E}[Y^2] \quad (17)$$

for any random variable X and Y .

Proof: We assume $\mathbb{E}[Y^2] \neq 0$, 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 ≥ 0 .

Proof of correlation coefficient magnitude (2)

Proof:

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

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

Proof of correlation coefficient magnitude (3)

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

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

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

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

Then, we make use of Schwarz inequality

$$(\rho(X, Y))^2 = \frac{(\text{cov}(X, Y))^2}{\text{var}(X) \text{var}(Y)} = \frac{(\mathbb{E}[\tilde{X}\tilde{Y}])^2}{\mathbb{E}[\tilde{X}^2] \mathbb{E}[\tilde{Y}^2]} \leq 1$$

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

Proof of correlation coefficient magnitude (4)

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

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

Therefore,

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

Proof of correlation coefficient magnitude (5)

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

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

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,

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

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 $\rho(X, Y)$?

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

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

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

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

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

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

Alternatively, we can apply the formula

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

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

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

- 1 Derived distributions
- 2 Sum of random variables
- 3 Covariance and correlation
 - Covariance
 - A detailed example
 - Correlation
- 4 Gaussian random variables**
 - Central limit theorem**
 - Higher dimensional Gaussian
 - Generalization of Gaussian random variable
- 5 (Advanced Topic) Expectation and variance
 - Iterated expectation
 - Total variance
- 6 (Advanced Topic) Random number of independent random variables

More about Gaussian

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

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \quad (18)$$

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

$$f_X(x) = \sqrt{\frac{\beta}{2\pi}} e^{-\frac{1}{2}\beta(x-\mu)^2} \quad (19)$$

Ubiquity of Gaussian random variables

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

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

$$Z_n = \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \quad (20)$$

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

Central limit theorem

We can easily show that

$$\begin{aligned}\mathbb{E}(Z_n) &= \frac{\mathbb{E}(X_1 + \cdots + X_n) - n\mu}{\sigma\sqrt{n}} = 0 \\ \text{var}(Z_n) &= \frac{\text{var}(X_1 + \cdots + X_n)}{\sigma^2 n} = \frac{n\sigma^2}{n\sigma^2} = 1\end{aligned}$$

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

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

as n approaches infinity. We will not be showing the proof here (which normally uses 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 $\mathbb{P}(S_n \leq c)$ 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

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

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?

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 $S_{100} = X_1 + \dots + X_{100}$.

$$\textcircled{1} \quad n\mu = 100 \times (5 + 50)/2 = 2750 ; \quad n\sigma^2 = 100 \times (50 - 5)^2/12 = 16875$$

$$\textcircled{2} \quad z = \frac{3000 - 2750}{\sqrt{16875}} = \frac{250}{129.9} = 1.92$$

$$\textcircled{3} \quad \mathbb{P}(S_{100} > 3000) = 1 - \mathbb{P}(S_{100} \leq 3000) \approx 1 - \Phi(1.92) = 0.0274$$

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.

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 $S_{100} = X_1 + \dots + X_{100}$. Note that the events $\{N_{320} \geq 100\}$ and $\{S_{100} \leq 320\}$ are equivalent.

$$\textcircled{1} \quad n\mu = 100 \times (1 + 5)/2 = 300 ; \quad n\sigma^2 = 100 \times (5 - 1)^2/12 = 400/3$$

$$\textcircled{2} \quad z = \frac{320 - 300}{\sqrt{400/3}} = 1.73$$

$$\textcircled{3} \quad \mathbb{P}(S_{100} \leq 320) \approx \Phi(1.73) = 0.9582$$

- 1 Derived distributions
- 2 Sum of random variables
- 3 Covariance and correlation
 - Covariance
 - A detailed example
 - Correlation
- 4 Gaussian random variables**
 - Central limit theorem
 - Higher dimensional Gaussian**
 - Generalization of Gaussian random variable
- 5 (Advanced Topic) Expectation and variance
 - Iterated expectation
 - Total variance
- 6 (Advanced Topic) Random number of independent random variables

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 :

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

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:

$$f_{\mathbf{X}}(\mathbf{x}) = \sqrt{\frac{1}{(2\pi)^n \det(\boldsymbol{\Sigma})}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})} \quad (21)$$

- $\mathbf{x}$: vector of observations
- $\boldsymbol{\mu}$: vector of the means
- $\boldsymbol{\Sigma}$: 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:

$$f_{\mathbf{X}}(\mathbf{x}) = \sqrt{\frac{\det(\boldsymbol{\beta})}{(2\pi)^n}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\beta} (\mathbf{x}-\boldsymbol{\mu})} \quad (22)$$

Once again, the multivariate Gaussian is completely determined by $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ 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

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

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

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

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

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

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 Derived distributions
- 2 Sum of random variables
- 3 Covariance and correlation
 - Covariance
 - A detailed example
 - Correlation
- 4 Gaussian random variables**
 - Central limit theorem
 - Higher dimensional Gaussian
 - **Generalization of Gaussian random variable**
- 5 (Advanced Topic) Expectation and variance
 - Iterated expectation
 - Total variance
- 6 (Advanced Topic) Random number of independent random variables

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:

$$f_X(x) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|x-\mu|/\alpha)^\beta} \quad (23)$$

- 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

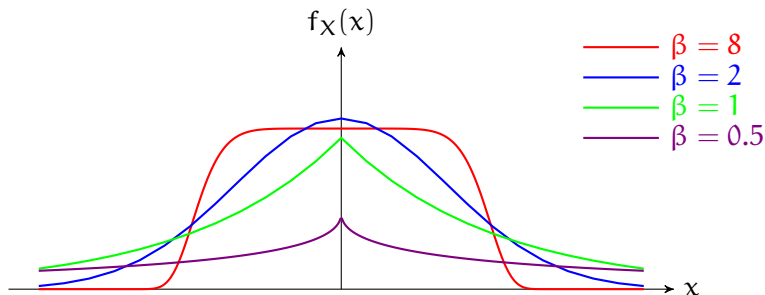
Case II: $\beta = 1$

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

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 $\beta \rightarrow \infty$, 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

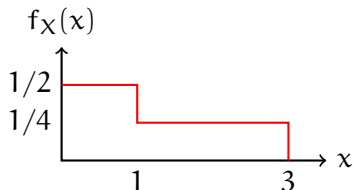
- 1 Derived distributions
- 2 Sum of random variables
- 3 Covariance and correlation
 - Covariance
 - A detailed example
 - Correlation
- 4 Gaussian random variables
 - Central limit theorem
 - Higher dimensional Gaussian
 - Generalization of Gaussian random variable
- 5 (Advanced Topic) Expectation and variance
 - Iterated expectation
 - Total variance
- 6 (Advanced Topic) Random number of independent random variables

Motivating example

A continuous random variable X has the PDF

$$f_X(x) = \begin{cases} \frac{1}{2} & 0 \leq x \leq 1 \\ \frac{1}{4} & 1 < x \leq 3 \\ 0 & \text{otherwise.} \end{cases}$$

as depicted below. What is $\mathbb{E}(X)$ and $\text{var}(X)$?



Motivating example

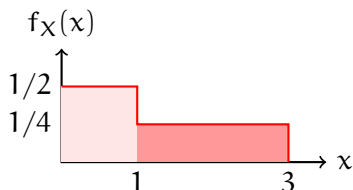
We can solve it directly:

$$\begin{aligned}\mathbb{E}(X) &= \int_0^1 x \frac{1}{2} dx + \int_1^3 x \frac{1}{4} dx \\ &= \frac{1}{4} [x^2]_0^1 + \frac{1}{8} [x^2]_1^3 = \frac{1}{4} + \frac{9}{8} - \frac{1}{8} = \frac{5}{4}\end{aligned}$$

$$\begin{aligned}\mathbb{E}(X^2) &= \int_0^1 x^2 \frac{1}{2} dx + \int_1^3 x^2 \frac{1}{4} dx \\ &= \frac{1}{6} [x^3]_0^1 + \frac{1}{12} [x^3]_1^3 = \frac{1}{6} + \frac{27}{12} - \frac{1}{12} = \frac{7}{3}\end{aligned}$$

$$\text{var}(X) = \frac{7}{3} - \left(\frac{5}{4}\right)^2 = \frac{37}{48}$$

Motivating example



It seems we can “divide” X in two ranges: $[0, 1]$ and $[1, 3]$.
 Let us define an auxiliary random variable Y where

$$Y = \begin{cases} 1 & x < 1 \\ 2 & x \geq 1 \end{cases}$$

Can we compute $\mathbb{E}(X)$ and $\text{var}(X)$ in terms of $\mathbb{E}(X|Y)$ and $\text{var}(X|Y)$?

Definition

Let us first consider $\mathbb{E}[X|Y]$. Remember that *unconditional average* = *averaging the conditional averages*:

$$\mathbb{E}(X) = \sum_y p_Y(y) \mathbb{E}(X|Y=y) \quad (24)$$

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} \mathbb{E}(X|Y=y) f_Y(y) dy \quad (25)$$

Definition

Let us first consider $\mathbb{E}[X|Y]$. Remember that *unconditional average* = *averaging the conditional averages*:

$$\mathbb{E}(X) = \sum_y p_Y(y) \mathbb{E}(X|Y=y) \quad (24)$$

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} \mathbb{E}(X|Y=y) f_Y(y) dy \quad (25)$$

- ① $\mathbb{E}[X|Y=y]$ is a constant (for a fixed value of y). So more generally, it is a function of y .
- ② $\mathbb{E}[X|Y]$ is therefore a function of Y , i.e., with PMF $p_Y(y)$ or PDF $f_Y(y)$.
- ③ So, the above are formulas for *expectation of* $\mathbb{E}[X|Y]$.

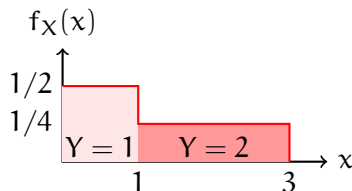
Definition

We therefore have this *law of iterated expectation*:

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}_Y[\mathbb{E}_X[X|Y]] \quad (26)$$

Note that we put in the subscript X and Y to emphasize what the random variable is with each expectation operation. We do not always make that designation.

Motivating example



Note that $\mathbb{P}(Y = 1) = \mathbb{P}(Y = 2) = 1/2$. Also, conditioning on $Y = 1$ or $Y = 2$, the r.v. X is uniform, such that

$$\mathbb{E}(X|Y = 1) = \frac{1}{2} \quad \text{and} \quad \mathbb{E}(X|Y = 2) = 2$$

Therefore,

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}(\mathbb{E}[X|Y]) = \mathbb{P}(Y = 1) \mathbb{E}[X|Y = 1] + \mathbb{P}(Y = 2) \mathbb{E}[X|Y = 2] \\ &= \frac{1}{2} \left(\frac{1}{2} \right) + \frac{1}{2} (2) = \frac{5}{4} \end{aligned}$$

- 1 Derived distributions
- 2 Sum of random variables
- 3 Covariance and correlation
 - Covariance
 - A detailed example
 - Correlation
- 4 Gaussian random variables
 - Central limit theorem
 - Higher dimensional Gaussian
 - Generalization of Gaussian random variable
- 5 (Advanced Topic) Expectation and variance**
 - Iterated expectation
 - **Total variance**
- 6 (Advanced Topic) Random number of independent random variables

Total variance

There's also a law of total variance:

$$\text{var}(X) = \mathbb{E}[\text{var}(X|Y)] + \text{var}(\mathbb{E}[X|Y]) \quad (27)$$

Both the law of iterated expectation and law of total variance allow us to start with expressions of $\mathbb{E}(X|Y)$ and $\text{var}(X|Y)$ to arrive at $\mathbb{E}(X)$ and $\text{var}(X)$.

Total variance

To show the law of total variance, we first define two quantities:

$$\hat{X} = \mathbb{E}(X|Y) \quad (28)$$

$$\tilde{X} = \hat{X} - X \quad (29)$$

$\hat{X}$ is an **estimator** of X given Y , whereas $\tilde{X}$ is the estimation **error**.

Total variance

To show the law of total variance, we first define two quantities:

$$\hat{X} = \mathbb{E}(X|Y) \quad (28)$$

$$\tilde{X} = \hat{X} - X \quad (29)$$

$\hat{X}$ is an **estimator** of X given Y , whereas $\tilde{X}$ is the estimation **error**.

In our example:

$$\hat{X} = \begin{cases} \frac{1}{2} & y = 1 (0 \leq X < 1) \\ 2 & y = 2 (1 \leq X < 3) \end{cases}$$

$$\tilde{X} = \begin{cases} \frac{1}{2} - X & y = 1 (0 \leq X < 1) \\ 2 - X & y = 2 (1 \leq X < 3) \end{cases}$$

Total variance

We aim to show:

- 1 We can divide X into two parts, such that

$$\text{var}(X) = \text{var}(\tilde{X}) + \text{var}(\hat{X})$$

The second term $\text{var}(\hat{X})$ is $\text{var}(\mathbb{E}(X|Y))$.

- 2 The variance of the estimator is

$$\text{var}(\tilde{X}) = \mathbb{E}(\text{var}(X|Y))$$

To show $\text{var}(X) = \text{var}(\tilde{X}) + \text{var}(\hat{X})$, we need to demonstrate that the *estimator is uncorrelated with the estimation error*. We will do that later.

Total variance

Instead, we first look at the estimator. We want to show in two steps:

- 1 Its expected value is zero:

$$\mathbb{E}(\tilde{X}) = 0$$

- 2 Its variance is then:

$$\text{var}(\tilde{X}) = \mathbb{E}(\tilde{X}^2) = \mathbb{E}(\text{var}(X|Y))$$

Total variance

The estimator $\hat{X} = \mathbb{E}(X|Y)$ is *unbiased*, because

$$\mathbb{E}(X) = \mathbb{E}[\mathbb{E}(X|Y)] = \mathbb{E}(\hat{X}),$$

and therefore $\mathbb{E}(\tilde{X}) = \mathbb{E}(\hat{X} - X) = 0$.

Total variance

The error has an expected value of 0, but what about its variance, $\text{var}(\tilde{X})$?

Total variance

The error has an expected value of 0, but what about its variance, $\text{var}(\tilde{X})$?

By definition,

$$\text{var}(X|Y) = \mathbb{E}\left[(X - \mathbb{E}[X|Y])^2 | Y\right] = \mathbb{E}\left[(X - \hat{X})^2 | Y\right] = \mathbb{E}[\tilde{X}^2 | Y]$$

Total variance

The error has an expected value of 0, but what about its variance, $\text{var}(\tilde{X})$?

By definition,

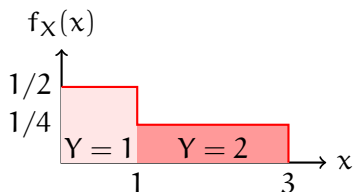
$$\text{var}(X|Y) = \mathbb{E}[(X - \mathbb{E}[X|Y])^2 | Y] = \mathbb{E}[(X - \hat{X})^2 | Y] = \mathbb{E}[\tilde{X}^2 | Y]$$

Therefore,

$$\text{var}(\tilde{X}) = \mathbb{E}(\tilde{X}^2) - (\mathbb{E}(\tilde{X}))^2 = \mathbb{E}(\mathbb{E}[\tilde{X}^2 | Y]) = \mathbb{E}(\text{var}(X|Y))$$

Total variance

In our example:



Conditioning on $Y = 1$ or $Y = 2$, the r.v. X is uniform, such that

$$\text{var}(X|Y = 1) = \frac{1^2}{12} \quad \text{and} \quad \text{var}(X|Y = 2) = \frac{2^2}{12}$$

Therefore,

$$\begin{aligned} \mathbb{E}(\text{var}(X|Y)) &= \mathbb{P}(Y = 1) \text{var}(X|Y = 1) + \mathbb{P}(Y = 2) \text{var}(X|Y = 2) \\ &= \frac{1}{2} \left(\frac{1}{12} \right) + \frac{1}{2} \left(\frac{4}{12} \right) = \frac{5}{24} \end{aligned}$$

Total variance

Now, we go back to show that *the estimator is uncorrelated with the estimation error*:

$$\text{cov}(\hat{X}, \tilde{X}) = 0 \quad (30)$$

First, we have

$$\mathbb{E}(\tilde{X} | Y) = \mathbb{E}((\hat{X} - X) | Y) = \mathbb{E}(\hat{X} | Y) - \mathbb{E}(X | Y) = 0$$

because given Y , then $\hat{X}$ is a fixed value, and therefore $\mathbb{E}(\hat{X} | Y) = \hat{X}$.

The second term, $\mathbb{E}(X | Y)$ is $\hat{X}$ by definition.

Total variance

Second, note that for any function $g(\cdot)$, we have

$$\mathbb{E}(Xg(Y) | Y) = g(Y)\mathbb{E}(X | Y),$$

because given the value of Y , the function $g(Y)$ is a constant and therefore can be pulled outside the expectation.

Total variance

Second, note that for any function $g(\cdot)$, we have

$$\mathbb{E}(Xg(Y) | Y) = g(Y)\mathbb{E}(X | Y),$$

because given the value of Y , the function $g(Y)$ is a constant and therefore can be pulled outside the expectation.

As a special case, we have

$$\mathbb{E}(\hat{X}\tilde{X}) = \mathbb{E}(\mathbb{E}[\hat{X}\tilde{X} | Y]) = \mathbb{E}(\hat{X}\mathbb{E}[\tilde{X} | Y]) = 0$$

because $\hat{X}$ is a function of Y only, and $\mathbb{E}[\tilde{X} | Y] = 0$ as calculated earlier.

Total variance

Third,

$$\text{cov}(\hat{X}, \tilde{X}) = \mathbb{E}(\hat{X}\tilde{X}) - \mathbb{E}(\hat{X})\mathbb{E}(\tilde{X}) = 0 - \mathbb{E}(X) \cdot 0 = 0$$

Total variance

Third,

$$\text{cov}(\hat{X}, \tilde{X}) = \mathbb{E}(\hat{X}\tilde{X}) - \mathbb{E}(\hat{X})\mathbb{E}(\tilde{X}) = 0 - \mathbb{E}(X) \cdot 0 = 0$$

Because $\text{cov}(\hat{X}, \tilde{X}) = 0$, we can conclude

$$\text{var}(X) = \text{var}(\tilde{X}) + \text{var}(\hat{X})$$

Total variance

Third,

$$\text{cov}(\hat{X}, \tilde{X}) = \mathbb{E}(\hat{X}\tilde{X}) - \mathbb{E}(\hat{X})\mathbb{E}(\tilde{X}) = 0 - \mathbb{E}(X) \cdot 0 = 0$$

Because $\text{cov}(\hat{X}, \tilde{X}) = 0$, we can conclude

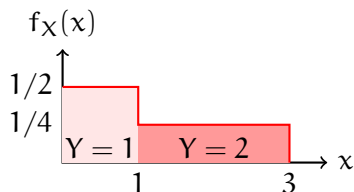
$$\text{var}(X) = \text{var}(\tilde{X}) + \text{var}(\hat{X})$$

The law of total variance is precisely the same equation:

$$\text{var}(X) = \underbrace{\mathbb{E}[\text{var}(X|Y)]}_{\text{var}(\tilde{X})} + \underbrace{\text{var}(\mathbb{E}[X|Y])}_{\text{var}(\hat{X})}$$

Total variance

In our example:



We already note that $\mathbb{E}(X) = \frac{5}{4}$, which we call μ here.

$$\begin{aligned}\text{var}(\mathbb{E}[X|Y]) &= \mathbb{P}(Y = 1) (\mathbb{E}[X|Y = 1] - \mu)^2 + \mathbb{P}(Y = 2) (\mathbb{E}[X|Y = 2] - \mu)^2 \\ &= \frac{1}{2} \left(\frac{1}{2} - \frac{5}{4} \right)^2 + \frac{1}{2} \left(2 - \frac{5}{4} \right)^2 = \frac{9}{16} \\ \text{var}(X) &= \mathbb{E}[\text{var}(X|Y)] + \text{var}(\mathbb{E}[X|Y]) = \frac{5}{24} + \frac{9}{16} = \frac{37}{48}\end{aligned}$$

Example

We have a biased coin where the probability of heads, denoted by Y , is a continuous uniform random variable in the range of $[0, 1]$. We toss the coin n times, and let X be the number of heads obtained. Find $\mathbb{E}(X)$ and $\text{var}(X)$.

Example

We have a biased coin where the probability of heads, denoted by Y , is a continuous uniform random variable in the range of $[0, 1]$. We toss the coin n times, and let X be the number of heads obtained. Find $\mathbb{E}(X)$ and $\text{var}(X)$.

ANS: X is dependent on Y , so Eq. (26) and (27) would be useful.

Example

We have a biased coin where the probability of heads, denoted by Y , is a continuous uniform random variable in the range of $[0, 1]$. We toss the coin n times, and let X be the number of heads obtained. Find $\mathbb{E}(X)$ and $\text{var}(X)$.

ANS: X is dependent on Y , so Eq. (26) and (27) would be useful.

Since $\mathbb{E}(X|Y = y) = ny$, we have $\mathbb{E}(X|Y) = nY$.

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}[X|Y]) = \mathbb{E}(nY) = n\mathbb{E}(Y) = \frac{n}{2}$$

Example

Similarly, since $\text{var}(X|Y=y) = ny(1-y)$, so $\text{var}(X|Y) = nY(1-Y)$.

$$\mathbb{E}(\text{var}(X|Y)) = \mathbb{E}(nY(1-Y)) = n\mathbb{E}(Y) - n\mathbb{E}(Y^2) = \frac{n}{2} - \frac{n}{3} = \frac{n}{6}$$

because $\mathbb{E}(Y^2) = \text{var}(Y) + (\mathbb{E}(Y))^2 = \frac{1}{12} + (\frac{1}{2})^2 = \frac{1}{3}$. Also,

$$\text{var}(\mathbb{E}[X|Y]) = \text{var}(nY) = n^2 \cdot \frac{1}{12}$$

Combining,

$$\text{var}(X) = \mathbb{E}[\text{var}(X|Y)] + \text{var}(\mathbb{E}[X|Y]) = \frac{n}{6} + \frac{n^2}{12}$$

- 1 Derived distributions
- 2 Sum of random variables
- 3 Covariance and correlation
 - Covariance
 - A detailed example
 - Correlation
- 4 Gaussian random variables
 - Central limit theorem
 - Higher dimensional Gaussian
 - Generalization of Gaussian random variable
- 5 (Advanced Topic) Expectation and variance
 - Iterated expectation
 - Total variance
- 6 (Advanced Topic) Random number of independent random variables

Example

You visit a number of bookstores in search of a particular textbook on probability. Any given bookstore carries the book with probability p , independent of the others. In a typical bookstore, the amount of time you spend is exponentially distributed with parameter λ , and independent of the time you spend in other bookstores. You will keep on visiting bookstores until you find the book (because the lectures are too boring, you'd rather learn from a book). What are the mean, variance, and PDF of the total time spent in search of the book?

Example

You visit a number of bookstores in search of a particular textbook on probability. Any given bookstore carries the book with probability p , independent of the others. In a typical bookstore, the amount of time you spend is exponentially distributed with parameter λ , and independent of the time you spend in other bookstores. You will keep on visiting bookstores until you find the book (because the lectures are too boring, you'd rather learn from a book). What are the mean, variance, and PDF of the total time spent in search of the book?

It is a sum of a *geometric number* of independent exponential random variables.

Example

You visit a number of bookstores in search of a particular textbook on probability. Any given bookstore carries the book with probability p , independent of the others. In a typical bookstore, the amount of time you spend is exponentially distributed with parameter λ , and independent of the time you spend in other bookstores. You will keep on visiting bookstores until you find the book (because the lectures are too boring, you'd rather learn from a book). What are the mean, variance, and PDF of the total time spent in search of the book?

It is a sum of a *geometric number* of independent exponential random variables.

We have sufficient tools now to approach such type of problems involving summing of a random number of independent random variables.

Setting

We consider

$$Y = X_1 + \dots + X_N$$

where

- N is a random variable that takes nonnegative integer values.
- $X_1, X_2, \dots$ are identically distributed random variables.
- $N, X_1, X_2, \dots$ are independent, meaning that any finite subcollection of these random variables are independent.
- $\mathbb{E}(X)$ and $\text{var}(X)$ are the common mean and variance of each X_i .

Expectation

We first calculate $\mathbb{E}(Y)$:

$$\begin{aligned}\mathbb{E}(Y | N = n) &= \mathbb{E}(X_1 + \dots + X_N | N = n) \\ &= \mathbb{E}(X_1 + \dots + X_n | N = n) \\ &= \mathbb{E}(X_1 + \dots + X_n) \\ &= n\mathbb{E}(X)\end{aligned}$$

This is true for every nonnegative integer n and, therefore,

$$\mathbb{E}(Y | N) = N\mathbb{E}(X)$$

Using the law of iterated expectations, we obtain

$$\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}[Y | N]) = \mathbb{E}(N\mathbb{E}[X]) = \mathbb{E}(N) \mathbb{E}(X)$$

Variance

Similarly, to compute $\text{var}(Y)$:

$$\begin{aligned}\text{var}(Y | N = n) &= \text{var}(X_1 + \dots + X_N | N = n) \\ &= \text{var}(X_1 + \dots + X_n | N = n) \\ &= \text{var}(X_1 + \dots + X_n) \\ &= n\text{var}(X)\end{aligned}$$

This is true for every nonnegative integer n and, therefore,

$$\text{var}(Y | N) = N\text{var}(X)$$

Using the law of total variance, we obtain

$$\begin{aligned}\text{var}(Y) &= \mathbb{E}[\text{var}(Y | N)] + \text{var}(\mathbb{E}[Y | N]) \\ &= \mathbb{E}[N\text{var}(X)] + \text{var}(N\mathbb{E}[X]) \\ &= \mathbb{E}(N) \text{var}(X) + (\mathbb{E}[X])^2 \text{var}(N)\end{aligned}$$

Putting together

Summary:

$$\mathbb{E}(Y) = \mathbb{E}(N) \mathbb{E}(X) \quad (31)$$

$$\text{var}(Y) = \mathbb{E}(N) \text{var}(X) + (\mathbb{E}[X])^2 \text{var}(N) \quad (32)$$

Furthermore, through the transform method, we can derive the overall distribution.

Moment generating function

To find $M_Y(s)$:

$$\begin{aligned}
 \mathbb{E}(e^{sY} | N = n) &= \mathbb{E}(e^{s(X_1 + \dots + X_N)} | N = n) \\
 &= \mathbb{E}(e^{sX_1} \dots e^{sX_n} | N = n) \\
 &= \mathbb{E}(e^{sX_1}) \dots \mathbb{E}(e^{sX_n}) \\
 &= (M_X(s))^n
 \end{aligned}$$

where $M_X(s)$ is the transform associated with (identically distributed) X_i for each i . Later on, we will also make use of the representation

$$(M_X(s))^n = e^{\log(M_X(s))^n} = e^{n \log M_X(s)}$$

Moment generating function

Now, we consider two formulas:

$$\begin{aligned}
 M_Y(s) &= \mathbb{E}(e^{sY}) = \mathbb{E}(\mathbb{E}[e^{sY} | N = n]) = \mathbb{E}\left((M_X(s))^N\right) \\
 &= \sum_{n=0}^{\infty} (M_X(s))^n f_N(n) \\
 &= \sum_{n=0}^{\infty} e^{n \log M_X(s)} f_N(n) \\
 M_N(n) &= \mathbb{E}(e^{sN}) = \sum_{n=0}^{\infty} e^{ns} f_N(n)
 \end{aligned}$$

So we can conclude

$$M_Y(s) = M_N(\log M_X(s)) \quad (33)$$

Example

You visit a number of bookstores in search of a particular textbook on probability. Any given bookstore carries the book with probability p , independent of the others. In a typical bookstore, the amount of time you spend is exponentially distributed with parameter λ , and independent of the time you spend in other bookstores. You will keep on visiting bookstores until you find the book (because the lectures are too boring, you'd rather learn from a book). What are the mean, variance, and PDF of the total time spent in search of the book? $\implies$ *A sum of a geometric number of independent exponential random variables*

$N = \text{number of bookstores} \sim \text{Geo}(p)$

$Y = \text{total time spent}$

Example

Now we make use of the results derived above, and that

$$M_X(s) = \frac{\lambda}{\lambda - s} \quad M_N(s) = \frac{pe^s}{1 - (1-p)e^s},$$

we can derive

Example

Now we make use of the results derived above, and that

$$M_X(s) = \frac{\lambda}{\lambda - s} \quad M_N(s) = \frac{pe^s}{1 - (1-p)e^s},$$

we can derive

$$\mathbb{E}(Y) = \mathbb{E}(N) \mathbb{E}(X) = \frac{1}{p} \cdot \frac{1}{\lambda}$$

$$\text{var}(Y) = \mathbb{E}(N) \text{var}(X) + (\mathbb{E}[X])^2 \text{var}(N) = \frac{1}{p} \cdot \frac{1}{\lambda^2} + \frac{1}{\lambda^2} \cdot \frac{1-p}{p^2} = \frac{1}{\lambda^2 p^2}$$

$$M_Y(s) = M_N(\log M_X(s)) = \frac{p \cdot \frac{\lambda}{\lambda-s}}{1 - (1-p) \frac{\lambda}{\lambda-s}} = \frac{p\lambda}{p\lambda - s}$$

Example

Now we make use of the results derived above, and that

$$M_X(s) = \frac{\lambda}{\lambda - s} \quad M_N(s) = \frac{pe^s}{1 - (1-p)e^s},$$

we can derive

$$\mathbb{E}(Y) = \mathbb{E}(N) \mathbb{E}(X) = \frac{1}{p} \cdot \frac{1}{\lambda}$$

$$\text{var}(Y) = \mathbb{E}(N) \text{var}(X) + (\mathbb{E}[X])^2 \text{var}(N) = \frac{1}{p} \cdot \frac{1}{\lambda^2} + \frac{1}{\lambda^2} \cdot \frac{1-p}{p^2} = \frac{1}{\lambda^2 p^2}$$

$$M_Y(s) = M_N(\log M_X(s)) = \frac{p \cdot \frac{\lambda}{\lambda-s}}{1 - (1-p) \frac{\lambda}{\lambda-s}} = \frac{p\lambda}{p\lambda - s}$$

which is the transform of an exponentially distributed r.v. with parameter $p\lambda$,

$$f_Y(y) = p\lambda e^{-p\lambda y} \quad y \geq 0$$

Example

How about *a sum of a geometric number of independent geometric random variables*?

Example

How about *a sum of a geometric number of independent geometric random variables*?

$$N \sim \text{Geo}(p)$$

$$X_i \sim \text{Geo}(q)$$

$$Y = X_1 + \dots + X_N$$

We have

$$\begin{aligned} M_Y(s) &= M_N(\log M_X(s)) = \frac{p \frac{qe^s}{1-(1-q)e^s}}{1 - (1-p) \frac{qe^s}{1-(1-q)e^s}} \\ &= \frac{pqe^s}{1 - (1-pq)e^s} \end{aligned}$$

which is the transform of a geometrically distributed r.v. with parameter pq .

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.