40.001: Probability

Week 8, Class 1

ALL LECTURES: WEEKS 8 TO 13

Term 4, 2017

SINGAPORE UNIVERSITY OF
TECHNOLOGY AND DESIGN

What will we do in these six weeks?

- Chapter 5: Continuous random variables.
- Chapter 6: Joint distribution of random variables.
- Chapter 7: Properties of expectations.
- Chapter 8: Limit theorems.

Continuous random variables

Suppose we have a random quantity which takes values in an uncountable set.

- Height of an individual (in cms),
- temperature of an object (in $^\circ$C).

A random variable X is continuous if there exists (an absolutely continuous) function $f : \mathbb{R} \to [0, \infty)$ such that for any set $B \subset \mathbb{R}$ (measurable)

$$\mathbb{P}(X \in B) = \int_B f(x)\,\mathrm{d}x.$$

The function f is called the probability density function (p.d.f.) of X. [1]

[1] Ignore the words in grey for now; but they are the more technically correct versions of what we will use in this class.

Properties of p.d.f.

$$1.\ \mathbb{P}(X \in \mathbb{R}) = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1,$$

$$2.\ \mathbb{P}(X \in [a,b]) = \mathbb{P}(a \leq X \leq b) \ = \int_a^b f(x)\,\mathrm{d}x$$

$$\approx \sum_{h=0}^{n-1} f\left(a + \frac{h(b-a)}{n}\right) \times \frac{b-a}{n}.$$

$$3.\ \mathbb{P}(X = y) = \int_y^y f(x)\,\mathrm{d}x = 0.$$

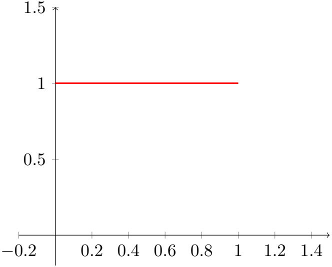
The (cumulative) distribution function (c.d.f.) takes the form

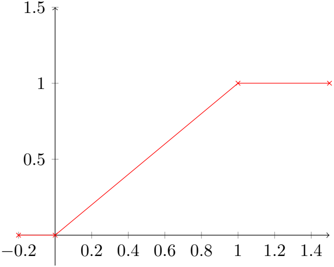
$$F(a) = \mathbb{P}(X \leq a) = \int_{-\infty}^{a} f(x)\,\mathrm{d}x.$$

So $F'(a) = f(a)$.

Example

Example 1:





$$f(x) = \begin{cases} 1 & 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

$$F(x) = \begin{cases} 0 & x < 0, \\ x & 0 \le x \le 1, \\ 1 & x > 1. \end{cases}$$

Example

Example 2:

$$f(x) = \begin{cases} 3x^2 & 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \qquad F(x) = \begin{cases} 0 & x < 0, \\ x^3 & 0 \le x \le 1, \\ 1 & x > 1. \end{cases}$$

Example 3:

$$f(x) = \begin{cases} 6x(1-x) & 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \qquad F(x) = \begin{cases} 0 & x < 0, \\ 3x^2 - 2x^3 & 0 \le x \le 1, \\ 1 & x > 1. \end{cases}$$

Example 4:

$$f(x) = \begin{cases} 48x(1-4x) & 0 \le x \le \frac{1}{4}, \\ 48(4x-3)(1-x) & \frac{3}{4} \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Expectation and variance

Suppose $X \sim f(x)$. For small $\mathrm{d}x$,

$$f(x)\,\mathrm{d}x \approx \mathbb{P}(x \leq X \leq x + \mathrm{d}x).$$

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x\, f(x)\,\mathrm{d}x.$$

For a function $g : \mathbb{R} \to \mathbb{R}$, $\mathbb{E}(g(X))$ is computed analogously:

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x)\, f(x)\,\mathrm{d}x.$$

We define the variance as before:

$$\mathsf{Var}(X) = \mathbb{E}[X - \mathbb{E}(X)]^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

Expectation

- For $a, b \in \mathbb{R}$, we have $\mathbb{E}(aX + b) = a\,\mathbb{E}(X) + b$.
- For a non-negative random variable X,

$$\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X > x)\,\mathrm{d}x.$$

$$\begin{aligned}
\int_0^\infty \mathbb{P}(X > x)\,\mathrm{d}x &= \int_0^\infty \int_x^\infty f(y)\,\mathrm{d}y\,\mathrm{d}x \\
&= \int_0^\infty \int_0^y f(y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_0^\infty \left[\int_0^y \mathrm{d}x\right] f(y)\,\mathrm{d}y = \int_0^\infty y\,f(y)\,\mathrm{d}y = \mathbb{E}(X).
\end{aligned}$$

- For $a, b \in \mathbb{R}$, we have $\mathsf{Var}(aX + b) = a^2\mathsf{Var}(X)$.

Exercise

- Find the Expectation and Variance for:

$$1.\ \ f(x) = \begin{cases} 1 & 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

$$2.\ \ f(x) = \begin{cases} 6x(1-x) & 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

1. $\mathbb{E}(X) = 1/2$, $\text{Var}(X) = 1/12$, 2. $\mathbb{E}(X) = 1/2$, $\text{Var}(X) = 1/20$.

- Find $\mathbb{E}(e^X)$ for the first p.d.f. : Uniform density between $(0, 1)$.

  Ans. $e - 1$

- For a pdf f: Is $f(x) \ge 0$? (yes)

  Is $f(x) \le 1$? (not necessarily)

Uniform distribution

- A random variable X is Uniformly distributed over (a, b) for $a < b$, if it's p.d.f. is given by

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b, \\ 0 & \text{otherwise.} \end{cases}$$

- We write $X \sim \text{Unif}(a, b)$.

Uniform distribution

- Find the c.d.f. $F(x)$. Also find $\mathbb{E}(X)$, Var(X) and $\mathbb{E}(X^n)$ for any positive integer n.
- $\mathbb{E}(X) = \frac{a+b}{2}$, Var$(X) = \frac{(b-a)^2}{12}$, $\mathbb{E}(X^n) = \frac{b^{n+1}-a^{n+1}}{(n+1)(b-a)}$.

$$F(x) = \begin{cases} 0 & x \leq a, \\ \frac{x-a}{b-a} & a < x < b, \\ 1 & x \geq b. \end{cases}$$

Exercise

1. Suppose $X \sim$ Unif$(-1, 1)$. Find $\mathbb{P}(|X| < \frac{1}{3} \,||X| < \frac{1}{2})$.
   Ans. 2/3.
2. A machine makes circular plates which has diameter Uniformly distributed between $(0, 2)$ (in cms); (you may assume that the radius is Uniformly distributed between (0,1)).
   Unfortunately plates of area $3.1\,\text{cm}^2$ or more cannot be used.
   If 200 plates are made in 1 hour, what is the expected number of discarded plates?
   Ans. The radius of a plate $R \sim$ Unif$(0, 1)$. Hence probability that a plate is discarded is:
   $$p = \mathbb{P}(\text{Area} \geq 3.1 \text{ cm}^2) = \mathbb{P}(\pi R^2 \geq 3.1) = \mathbb{P}(R \geq \sqrt{3.1/\pi}) = 0.0066.$$
   Let $X = \#$ of discarded plates. Then $X \sim$ Bin(n, p) where $n = 200$ and $p = 0.0066$.
   Hence $\mathbb{E}(X) = np = 1.32$.

Exercise

- A stick of length one is split at a point U having Uniform density on $(0,1)$. Find the expected length of the part of the stick containing a fixed point p.

  Answer. For $0 < p < 1$ and the break point $u \in (0,1)$, define $L_p(u)$ to be the length of the the stick containing the breaking point. Then

  $$L_p(u) = \begin{cases} u, & u > p \\ 1-u, & u \le p. \end{cases}$$

  Hence for $U \sim$ Unif$[0,1]$, we compute $\mathbb{E}\, L_p(U) = 0.5 + p(1-p)$.

- For what value of p is the expected length of the stick containing p maximized? Ans. 0.5.

Exponential distribution

- A random variable X follows an exponential distribution with parameter $\lambda > 0$, if it's p.d.f. is given by

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

- We write $X \sim \text{Exp}(\lambda)$.
- The c.d.f. of X:

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

- $\mathbb{E}(X) = 1/\lambda$, $\text{Var}(X) = 1/\lambda^2$. PLOT

Often useful in modelling the amount of time until a specific event occurs (earthquake, machine failure, queue time, next phone call, waiting time in Poisson processes).

Example

Suppose the duration of a phone call you get is $X \sim \text{Exp}(1/10)$. Find

1. $\mathbb{P}$(a phone call lasts more than 10 minutes). e^{-1}
2. $\mathbb{P}$(a phone call lasts between 10 and 20 minutes). $e^{-1} - e^{-2}$
3. $\mathbb{P}$(a phone call lasts at least 20 minutes given that it has already lasted 10 minutes). e^{-1}

Memoryless property: For any $s, t \geq 0$,

$$\mathbb{P}(X > s + t \,|\, X > t) = \mathbb{P}(X > s).$$

Exponential is the unique continuous distribution with this property.

How about the discrete case? Answer: Geometric(p) (check!!).

Hazard rate function

Let X denote the lifetime of an item, with density f and distribution F. Hence X is a non-negative random variable. The probability that an item aged t will fail within time $\mathrm{d}t$ is

$$\mathbb{P}(X \in (t, t + \mathrm{d}t) | X > t) \approx \frac{f(t)\,\mathrm{d}t}{1 - F(t)}.$$

We define the hazard rate function to be

$$\lambda(t) = \frac{f(t)}{1 - F(t)}.$$

- For $X \sim \text{Exp}(\lambda)$, we have $\lambda(t) = \lambda$.

Hazard rate function

For a continuous random variable X, the hazard rate $\lambda(t)$ uniquely determines the c.d.f F.

$$\int_0^t \lambda(s)\,\mathrm{d}s = \int_0^t \frac{f(s)}{1-F(s)}\mathrm{d}s = -\ln(1-F(t)),$$

$$F(t) = 1 - \exp\left\{-\int_0^t \lambda(s)\,\mathrm{d}s\right\}$$

Hazard rate example

The death (failure) rate of a smoker is twice that of a non-smoker.

Does this mean a non-smoker has twice the probability of surviving a certain number of years than a smoker of same age?

Answer.

Note that we have $\lambda_s(t) = 2\lambda_n(t)$, where λ_s, λ_n are death rate/ hazard rate/ failure rate of smokers and non-smokers respectively.

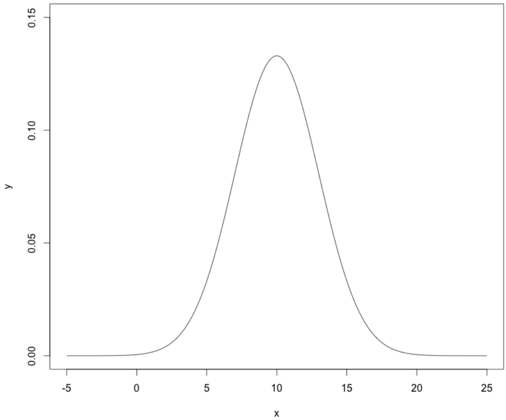
Now, $\mathbb{P}(A$-year old non-smoker reaches age $B)$

$$= \frac{1 - F(B)}{1 - F(A)} = \exp\left(-\int_A^B \lambda_n(t)\,\mathrm{d}t\right).$$

The probability for a smoker is the square of this.

Normal distribution

The normal (or *Gaussian*) distribution, or bell curve, is one of the most well-known distributions.



Normal distribution

- A random variable X follows a Normal distribution with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$, if it's p.d.f. is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$

- We denote $X \sim N(\mu, \sigma^2)$.
- Symmetry around μ: $f(\mu + x) = f(\mu - x)$.
- f is a valid p.d.f. (check)
- errors in measurement, height/weight of individuals, etc. are modelled with normal distribution.

Proof that f is a valid pdf:

Use polar co-ordinate transformation with $x = r\cos\theta, y = r\sin\theta$

$$\begin{aligned}
\left(\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x\right)^2 &= \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x \int_{-\infty}^{\infty} f(y)\,\mathrm{d}y \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\,\mathrm{d}x\,\mathrm{d}y \\
&= \frac{1}{2\pi}\int_{0}^{\infty}\int_{0}^{2\pi} e^{-r^2/2}\cdot r\,\mathrm{d}\theta\,\mathrm{d}r \\
&= \int_{0}^{\infty} re^{-r^2/2}\,\mathrm{d}r = \left[-e^{-r^2/2}\right]_0^{\infty} = 1.
\end{aligned}$$

Normal distribution

- For $a, b \in \mathbb{R}$, If $X \sim N(\mu, \sigma^2)$ and $Y = aX + b$ then $Y \sim N(a\mu + b, a^2\sigma^2)$.
- Now, if $X \sim N(\mu, \sigma^2)$ and $Z = \frac{X-\mu}{\sigma}$ then $Z \sim N(0, 1)$.

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad -\infty < z < \infty.$$

This is called the Standard Normal Distribution.

Normal distribution: mean and variance

Suppose $Z \sim N(0,1)$.

- Expectation:

$$\mathbb{E}\, Z = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-z^2/2}\, \mathrm{d}x = \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}\Big|_{-\infty}^{\infty} = 0,$$

$$\begin{aligned} \mathbb{E}\, Z^2 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-z^2/2}\, \mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}} \left(-z e^{-z^2/2}\right)\Big|_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} = 1. \end{aligned}$$

- Variance: Hence Var$(Z) = 1$.
- If $X = \sigma\, Z + \mu$, or $Z = (X - \mu)/\sigma$ then $X \sim N(\mu, \sigma^2)$.
- $\mathbb{E}\, X = \mu$, Var$(X) = \sigma^2$.

Normal distribution: Cumulative distribution function

- The c.d.f. for the standard normal $Z \sim N(0,1)$ is denoted

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2}\, \mathrm{d}x.$$

- Φ cannot be expressed in terms of elementary functions.
- $\Phi(-z) = 1 - \Phi(z)$.
- $\Phi(1) - \Phi(-1) \approx 0.683,\ \Phi(2) - \Phi(-2) \approx 0.955,\ \Phi(3) - \Phi(-3) \approx 0.997.$
- If $X \sim N(\mu, \sigma^2)$,

$$F_X(a) = \mathbb{P}(X \leq a) = \mathbb{P}\left(\frac{X-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}\right) = \mathbb{P}\left(Z \leq \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{a-\mu}{\sigma}\right).$$

NORMAL TABLE

PLOT

Examples

1. Suppose your test scores follow a normal distribution, and you get an A if you score more than 1 standard deviation above the mean, find the percentage of students who gets an A.

   $1 - \Phi(1) = 0.1587$

2. The age of NY times subscribers is $N(\mu = 35, \sigma^2 = 25)$. Find
   (a) $\mathbb{P}$ (A subscriber is between ages 30 and 40). $\Phi(1) - \Phi(-1) = 0.6827$
   (b) $\mathbb{P}$ (A subscriber is less than 45 years old). $1 - \Phi(1) = 0.9772$

3. Suppose $Z \sim N(0, 1)$. Find
   (a) $\mathbb{P}(|Z - 1| < 2)$. $\Phi(3) - \Phi(-1) = 0.84$
   (b) $\mathbb{P}(|Z| < 3)$. $\Phi(3) - \Phi(-3) = 0.9973$
   (c) $\mathbb{P}(\frac{Z}{1+Z^2} < \frac{1}{2})$. 1

NORMAL TABLE

Example

Value-at-Risk: The Value-at-Risk or VaR$_p$ of an investment is the amount v such that there is only $100p\%$ probability of a loss greater than v from the investment.

- Suppose X denotes (in million dollars) the gains from the investment. (So $-X$ is the loss)
- Let $X \sim N(\mu = 10, \sigma^2 = 100)$ and $p = 0.01$. Calculate VaR$_p(X)$.

  Answer.

  $$p = 0.01 = \mathbb{P}(-X > v) = \mathbb{P}(X < -v) = \Phi\left(\frac{-v-\mu}{\sigma}\right) = 1 - \Phi\left(\frac{v+\mu}{\sigma}\right)$$

  Hence

  $$\frac{v+\mu}{\sigma} = \frac{v+10}{10} = 2.326.$$

  So, v= 13.326.

Median and Quantiles

- Median: The median of a distribution is the minimum value under which 1/2 of the population lie. So if F is the distribution function then

$$F(\text{median}) = 0.5.$$

- Quantile: The p^{th} quantile x_p for $X \sim F$ is

$$x_p = \inf\{x : F(x) \geq p\} =: F^{\leftarrow}(p).$$

Example:

1. For the pdf $f(x) = 2x, 0 \leq x \leq 1$ find the p^{th} quantile of $X \sim f$.
2. Find the p^{th} quantile for $X \sim \text{Exp}(1)$.

Normal approximation to binomial

- For large n the binomial distribution can be approximated quite accurately by a Normal distribution with the same mean and variance.
- Digression: NY TIMES ARTICLE OPINION POLLS QUINCUNX
- Suppose $S_n \sim \text{Bin}(n, p)$.
- So $\mu = \mathbb{E}\, S_n = np$, $\text{Var}(S_n) = \sigma^2 = np(1-p)$.
- For $a < b$ (and large n)

$$\mathbb{P}\left\{a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right\} \approx \Phi(b) - \Phi(a).$$

- This is essentially De-Moivre Laplace Central Limit Theorem.
- Since a discrete distribution is approximated by a continuous distribution, we think

$$\mathbb{P}(X = i) = \mathbb{P}(i - 0.5 < X < i + 0.5)$$

and proceed to calculate. This is called *continuity correction*. PLOT

Normal approximation to binomial: example

Suppose the ideal class size in SUTD is 350. From experience, we know any accepted student has 25% chance of enrolling.

- Suppose 1200 students are accepted. What is the probability that more than 350 students enrol?
- Approximately how many students should be accepted to guarantee with 99% probability that at least 300 students are enrolled.

SOLUTION: If X denoted the number of enrolled students $X \sim \text{Bin}(n = 1200, p = 0.25)$.
(i) We want $\mathbb{P}(X > 350)$ Now $\mu = \mathbb{E}\, X = 300, \sigma^2 = \text{Var}(X) = 225$. Hence

$$\mathbb{P}(X > 350) = \mathbb{P}(X > 350.5) = \mathbb{P}\left(\frac{X - 300}{15} > \frac{350.5 - 300}{15}\right)$$
$$\approx 1 - \Phi(3.3) = 0.0005.$$

(ii) Find $n = M$ s.t, $\mathbb{P}(X > 300) = 0.99$, or, $1 - \Phi\left(\frac{300 - M/4}{\sqrt{3M}/4}\right) = 0.99$. We want $\left(\frac{300 - M/4}{\sqrt{3M}/4}\right) = \Phi^{-1}(0.01) = -2.33$.
We get $M \approx 1344$.

Gamma distribution

- A random variable X follows a Gamma distribution with parameters $\alpha > 0$ and $\lambda > 0$, if it's p.d.f. is given by

$$f(x) = \frac{\lambda^{\alpha} x^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda x}, \quad x \geq 0,$$

  where

$$\Gamma(\alpha) = \int_0^{\infty} e^{-y} y^{\alpha-1} \mathrm{d}y.$$

- We denote $X \sim \text{Gamma}(\alpha, \lambda)$.

Gamma distribution: properties

- $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$.
- $\Gamma(n) = (n - 1)!$
- $\Gamma(1/2) = \sqrt{\pi}$ (check!!)
- $\mathbb{E}(X) = \frac{\alpha}{\lambda}$.
- $\mathbb{E}(X^2) = \frac{\alpha(\alpha+1)}{\lambda^2}$.
- Var$(X) = \frac{\alpha}{\lambda^2}$.
- If $\alpha = 1$, we have the Exp(λ) distribution.
- If $\alpha = n$, some positive integer then a Gamma(n, λ) distribution is the (independent) sum of n exponential distributions with parameter λ.

PLOT

Beta distribution

- A random variable X follows a Beta distribution with parameters $a > 0, b > 0$, if it's p.d.f. is given by

$$f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}, \quad 0 < x < 1.$$

where

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1} \mathrm{d}x = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

- We denote $X \sim$ Beta(a, b).
- $\mathbb{E}\, X = \frac{a}{a+b}$, Var$(X) = \frac{ab}{(a+b)^2(a+b+1)}$.
- If $a = b = 1$, then $X \sim$ Unif$(0, 1)$.

PLOT

Exercise

1. Suppose a standard test of one hour follows a beta distribution with mean 45 minutes and standard deviation 10 minutes. What is the probability that the test finishes in 40 minutes?

2. Tires of a certain kind have lifetime which follows an exponential distribution with mean 10000 kms. How many spare tires should one keep on a 15000 km trip to make at least 60% sure that they don't need to buy more tires during the trip ? Tires work/ fail independently (and we assume a tire once changed doesn't fail during the trip anymore)

Exercise: solutions

1. If $X =$ time to take the test (in hours) then $X \sim$ Beta(a, b), with

$$\mathbb{E}\, X = \frac{a}{a+b} = \frac{3}{4}, \quad \text{and} \quad \text{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)} = \left(\frac{1}{6}\right)^2.$$

We can solve for a and b to get $a = \frac{69}{16}$ and $b = \frac{23}{16}$.

Hence

$$\begin{aligned}
\mathbb{P}(\text{test finishes in 40 minutes}) &= \mathbb{P}\left(X \leq \frac{2}{3}\right) \\
&= \frac{1}{B\left(\frac{69}{16}, \frac{23}{16}\right)} \int_0^{2/3} x^{53/16}(1-x)^{7/16}\,\mathrm{d}x \\
&= 0.28
\end{aligned}$$

Exercise

1. Tyres of a certain kind have lifetime which follows an exponential distribution with mean 10000 kms. How many spare tyres (think of the minimum number) should one keep on a 15000 km trip to make at least 60% sure that they don't need to buy more tyres during the trip?
   Assume that tyres work/ fail independently (and we assume a tyre once changed doesn't fail during the trip anymore)

This is not a "normal approximation of binomial" problem.

Exercise: solution

2. Let $Y =$ lifetime of a tyre, then $Y \sim$ Exp$(1/10000)$. The probability that a tyre expires (flattens) before 15000 kms is

$$p = \mathbb{P}(Y \leq 15000) = 1 - e^{-15000/10000} = 0.7769.$$

Let $X =$ Number of tyres that flatten/expire before 15000 kms. Then $X \sim$ Bin$(4, p)$. We want the minimum number of tyres s such that, $\mathbb{P}(X \leq s) \geq 0.6$.
Observe that

$$\mathbb{P}(X \leq 3) = 1 - p^4 = 0.6358, \mathbb{P}(X \leq 2) = 1 - p^4 - 4p^3(1-p) = 0.2173.$$

Hence one should keep at least 3 spare tyres.

Functions of Random Variables

1. If $X \sim \text{Unif}(0, 1)$, find the p.d.f. of $Y = X^n$ for an integer $n > 0$.
2. If X has p.d.f. f_X, what is the p.d.f. of $Y = |X|$?

Given the distribution F_X of a random variable X, what is the distribution of $Y = g(X)$?

Theorem: If X is a continuous random variable with p.d.f. f_X and $g : \mathbb{R} \to \mathbb{R}$ is strictly monotone and differentiable function of x, then $Y = g(X)$ has p.d.f.

$$f_Y(y) = f_X\big(g^{-1}(y)\big) \left| \frac{d}{dy} g^{-1}(y) \right|,$$

if $y = g(x)$ for some x and $f_Y(y) = 0$ otherwise.

Example

1. If $X \sim \text{Exp}(1)$, find the distribution of $Y = e^{-X}$.
   $Y \sim \text{Unif}(0, 1)$
2. If X has p.d.f. f_X, find the distribution of $Y = X^2$. Compute f_Y when $X \sim N(0, 1)$.
   $Y \sim \text{Gamma}(\alpha = 1/2, \gamma = 1/2)$
3. If $X \sim N(\mu, \sigma^2)$ find the distribution of $Y = e^X$. This is called the Lognormal Distribution.
   Income, daily returns from investments, etc. are often modelled using lognormal distribution.

Joint distribution of Random Variables: Introduction

Roll a fair dice twice.

$$X = \text{maximum of the two rolls,}$$

$$Y = \text{minimum of the two rolls.}$$

- Find $\mathbb{P}(X = 1)$ and $\mathbb{P}(Y = 2)$.
- Can they happen together? $\mathbb{P}(X = 1, Y = 2) = 0.$
- Let $p(x, y) = \mathbb{P}(X = x, Y = y)$ for $1 \leq x, y \leq 6$.
- $p(x, y)$ is the joint distribution or joint mass function of the random variables X and Y.
- $p_X(x) = \mathbb{P}(X = x)$ and $p_Y(y) = \mathbb{P}(Y = y)$ gives the marginal distributions or marginal probability mass functions of X and Y respectively.

Joint distribution of Random Variables: Introduction

	1	2	3	4	5	6	p_X
1	$\frac{1}{36}$	0	0	0	0	0	$\frac{1}{36}$
2	$\frac{1}{18}$	$\frac{1}{36}$	0	0	0	0	$\frac{3}{36}$
3	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{36}$	0	0	0	$\frac{5}{36}$
4	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{36}$	0	0	$\frac{7}{36}$
5	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{36}$	0	$\frac{9}{36}$
6	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{36}$	$\frac{11}{36}$
p_Y	$\frac{11}{36}$	$\frac{9}{36}$	$\frac{7}{36}$	$\frac{5}{36}$	$\frac{3}{36}$	$\frac{1}{36}$	1

What is the distribution of $X - Y$?

X	0	1	2	3	4	5
Probability	$\frac{1}{6}$	$\frac{5}{18}$	$\frac{2}{9}$	$\frac{1}{6}$	$\frac{1}{9}$	$\frac{1}{18}$

Definitions

For random variables X and Y,

- Joint (cumulative) distribution function of X, Y:

$$F(a, b) = \mathbb{P}(X \leq a, Y \leq b).$$

- (Marginal) distribution function of X and Y:

$$F_X(a) = \mathbb{P}(X \leq a) = \lim_{b \to \infty} F(a, b) = F(a, \infty),$$

$$F_Y(b) = \mathbb{P}(Y \leq b) = \lim_{a \to \infty} F(a, b) = F(\infty, b).$$

Definitions

- Also note that for $a_1 < a_2$ and $b_1 < b_2$,

$$\begin{aligned}\mathbb{P}(a_1 < X \leq a_2,\, b_1 < Y \leq b_2)& \\ &= F(a_2, b_2) - F(a_1, b_2) - F(a_2, b_1) + F(a_1, b_1).\end{aligned}$$

- Exercise: Find $\mathbb{P}(X > a, Y > b)$ in terms of F.

  Ans.

$$1 - F(a, \infty) - F(\infty, b) + F(a, b)$$

Properties

For random variables X and Y with joint (cumulative) distribution function

$$F(a, b) = \mathbb{P}(X \leq a, Y \leq b),$$

- $\lim_{x \to -\infty} \lim_{y \to -\infty} F(x, y) = 0$, $\lim_{x \to \infty} \lim_{y \to \infty} F(x, y) = 1$,
- $0 \leq F(x, y) \leq 1$,
- $F(x_1, y_1) \leq F(x_2, y_2)$ if $x_1 \leq x_2, y_1 \leq y_2$.

Discrete case

Suppose X, Y are both discrete. The Joint probability mass function:

$$p(x, y) = \mathbb{P}(X = x, Y = y),$$

$$p_X(x) = \mathbb{P}(X = x) = \sum_{y:\, p(x,y)>0} p(x, y),$$

$$p_Y(y) = \mathbb{P}(Y = y) = \sum_{x:\, p(x,y)>0} p(x, y).$$

$$\sum_{x,y:\, p(x,y)>0} p(x, y) = 1.$$

Theorem: If (X, Y) had joint p.m.f. $p(x, y)$ and $g : \mathbb{R}^2 \to \mathbb{R}$ then

$$\mathbb{E}[g(X, Y)] = \sum_x \sum_y g(x, y) p(x, y).$$

Exercise

Suppose (X, Y) has joint pmf given by

$$p(x, y) = c(x + y), \quad 1 \leq x, y \leq n.$$

Here x, y are integers.

1. Find c. $\quad c = \frac{1}{n^2(n+1)}$
2. Find the marginal distributions of X and Y.

   $p_X(x) = p_Y(x) = \frac{1}{2n} + \frac{x}{n(n+1)}$
3. Find $\mathbb{P}(X = Y)$ and $\mathbb{P}(X > Y)$.

   $\mathbb{P}(X = Y) = \frac{1}{n}, \mathbb{P}(X > Y) = \frac{1}{2}\left(1 - \frac{1}{n}\right)$
4. Find $\mathbb{E}\,X, \mathbb{E}\,Y, \mathbb{E}\,XY$.

   $\mathbb{E}\,X = \mathbb{E}\,Y = \frac{2n+1}{6} + \frac{n+1}{4}, \quad \mathbb{E}\,XY = \frac{(n+1)(2n+1)}{6}$

Definitions: joint distribution of random variables

For random variables X and Y,

- Joint (cumulative) distribution function of X, Y:

$$F(a, b) = \mathbb{P}(X \leq a, Y \leq b).$$

- (Marginal) distribution function of X and Y:

$$F_X(a) = \mathbb{P}(X \leq a) = \lim_{b \to \infty} F(a, b) = F(a, \infty),$$

$$F_Y(b) = \mathbb{P}(Y \leq b) = \lim_{a \to \infty} F(a, b) = F(\infty, b).$$

Continuous case

X and Y are jointly continuous if there exists a function $f(x,y) : \mathbb{R}^2 \to \mathbb{R}$ such that for any $C \subset \mathbb{R}^2$

$$\mathbb{P}((X,Y) \in C) = \int_{(x,y)\in C} f(x,y)\,\mathrm{d}x\mathrm{d}y.$$

We call f the joint probability density function of X and Y.

- The joint cdf is calculated

$$F(a,b) = \int_{-\infty}^{b} \int_{-\infty}^{a} f(x,y)\,\mathrm{d}x\mathrm{d}y.$$

- Again, we have

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x\mathrm{d}y = 1.$$

Continuous case: properties

- If $A \subset \mathbb{R}, B \subset \mathbb{R}$, write $C = \{(x, y) : x \in A, y \in B\}$. Then
$$\mathbb{P}(X \in A, Y \in B) = \int_B \int_A f(x, y)\,\mathrm{d}x\mathrm{d}y.$$

- Marginal pdf:
$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\,\mathrm{d}y, \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\,\mathrm{d}x.$$

- Note that
$$\mathbb{P}(a < X < a + \mathrm{d}a, b < Y < b + \mathrm{d}b) \approx f(a, b)\,\mathrm{d}a\mathrm{d}b.$$

  Hence we have $$f(a, b) = \frac{\partial^2 F(a, b)}{\partial a \partial b}.$$

Exercise

1. Suppose (X, Y) has joint pdf given by

$$f(x,y) = \begin{cases} 2e^{-x}e^{-2y}, & 0 < x < \infty, 0 < y < \infty, \\ 0 & \text{otherwise.} \end{cases}$$

   (a) Find marginal distributions of X and Y.

   $f_X(x) = e^{-x}, x > 0$ and $f_Y(y) = 2e^{-2y}, y > 0$.

   (b) Find $\mathbb{P}(X > 1, Y < 1)$ and $\mathbb{P}(X < Y)$.

   $\mathbb{P}(X > 1, Y < 1) = e^{-1}(1 - e^{-2})$ and $\mathbb{P}(X < Y) = \frac{1}{3}$.

   (c) Find the distribution of $Z = \frac{X}{Y}$.

   $f_Z(z) = \frac{2}{(2+z)^2}, z \geq 0$.

2. Suppose (X, Y) has joint pdf given by

$$f(x,y) = \begin{cases} c, & 0 \leq x + y \leq 1, x > 0, y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

   Find c.

   Answer. $c = 2$.

Joint distribution (two or more random variables)

- The joint cdf of random variables $X_1, X_2, \ldots, X_n$ is given by

$$F(x_1, x_2, \ldots, x_n) = \mathbb{P}(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n).$$

- We have $(X_1, X_2, \ldots, X_n)$ jointly continuous if there exists a function $f(x_1, \ldots, x_n)$ such that

$$\mathbb{P}((X_1, \ldots, X_n) \in C) = \int_{(x_1, \ldots, x_n) \in C} f(x_1, \ldots, x_n) \mathrm{d}x_1 \ldots \mathrm{d}x_n.$$

Independence

- Random variables X and Y are independent if for any two sets $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$ we have

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\,\mathbb{P}(Y \in B).$$

- Independence of X and Y is also equivalent to

$$F(a,b) = \mathbb{P}(X \leq a, Y \leq b) = \mathbb{P}(X \leq a)\,\mathbb{P}(Y \leq b) = F_X(a)F_Y(b).$$

  for all $a, b \in \mathbb{R}$.

- In the discrete case independence is also equivalent to: $p(x,y) = p_X(x)p_Y(y), \quad \forall x, y$.
- In the continuous case independence is also equivalent to: $f(x,y) = f_X(x)f_Y(y), \quad \forall x, y$.

Exercise

- The number of customers joining a queue at a bank is $X \sim \text{Poisson}(\lambda)$. If each individual has probability p of being a male (independent of the others), show that the numbers of males and females joining the queue are independent random variables.

$$\mathbb{P}(M = i, F = j) = \mathbb{P}(M = i, F = j | M + F = i + j)\, \mathbb{P}(M + F = i + j).$$

$$\mathbb{P}(M + F = i + j) = e^{-\lambda} \frac{\lambda^{i+j}}{(i + j)!}.$$

$$\mathbb{P}(M = i, F = j | M + F = i + j) = \binom{i + j}{i} p^i (1 - p)^j.$$

$$\mathbb{P}(M = i, F = j) = e^{-\lambda} \lambda^{i+j} \frac{p^i}{i!} \frac{(1 - p)^j}{j!}.$$

Sum over j to find $\mathbb{P}(M = i)$.

Exercises

1. Suppose you have an iPhone X and a Samsung Galaxy Note8 having lifetime distributions Exp(μ) and Exp(λ) respectively in years. What is the probability that the iPhone fails before the Galaxy Note? Assume the two lifetimes are independent.
   (You can think mean life of the iPhone to be $1/\mu = 3$ and that of the Galaxy Note to be $1/\lambda = 5$ years.)
2. Michael and Ash agree to meet at a cafe between 5 pm and 6 pm. They arrive independently following a Unif$[0, 1]$ distribution (in hours) starting at 5pm and wait at the cafe for 15 minutes (each). What is the probability that they meet?

Independence

- Random variables X and Y are independent if for any two sets $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$ we have

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\,\mathbb{P}(Y \in B).$$

- Proposition: Continuous (discrete) random variables X and Y are independent if and only if their joint pdf (pmf) can be written as

$$f(x,y) = h(x)g(y), \quad -\infty < x < \infty, -\infty < y < \infty.$$

Independence

Are X and Y independent in the following?

$$(1) f(x,y) = \begin{cases} 6e^{-2x-3y}, & 0 < x < \infty, 0 < y < \infty \\ 0 & \text{otherwise.} \end{cases} \quad \text{(Yes)}$$

$$= 2e^{-2x}\mathbf{1}_{\{0<x<\infty\}} \times 3e^{-3y}\mathbf{1}_{\{0<y<\infty\}} = h(x)g(y).$$

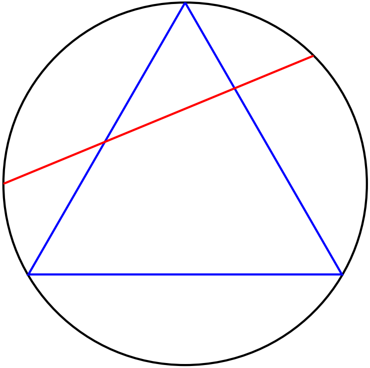
$$(2) f(x,y) = \begin{cases} 24xy, & 0 < x < 1, 0 < y < 1, 0 < x+y < 1 \\ 0 & \text{otherwise.} \end{cases} \quad \text{(No)}$$

Symbol shock

- $\prod$: product of numbers.
  $a_1 \times a_2 \times \ldots \times a_n = \prod_{i=1}^{n} a_i$
- $\perp$: independence.
  X and Y are independent: $X \perp Y$.
- iid: independent and identically distributed
- $\mathbf{1}_A(x)$: indicator of the set A.

$$\mathbf{1}_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{otherwise.} \end{cases}$$

A puzzle: random chord



Find the probability that a random chord in the circle has length greater than the length of the side of the inscribed equilateral triangle.

Independence (two or more random variables)

- $X_1, X_2, \ldots, X_n$ are independent if for all $x_1, x_2, \ldots, x_n \in \mathbb{R}$,

$$\mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \leq x_i).$$

- If X_i's are continuous random variables then $\forall x_1, x_2, \ldots, x_n \in \mathbb{R}$,

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i).$$

- If X_i's are discrete random variables then $\forall x_1, x_2, \ldots, x_n \in \mathbb{R}$,

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p_{X_i}(x_i).$$

In statistical analysis of sampled data, this is often very useful. A sequence of independent and identically distributed random variables are often called iid samples.

Exercises: sum of independent random variables

1. X and Y are independent and $X \sim \text{Bin}(n, p)$ and $Y \sim \text{Bin}(m, p)$. Find the distribution of $Z = X + Y$.

   Answer: $Z \sim \text{Bin}(m + n, p)$.

2. X and Y are independent and $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$. Find the distribution of $Z = X + Y$.

   Answer: $Z \sim \text{Poisson}(\lambda + \mu)$.

In both cases, Z is called the convolution of X and Y.

Sum of independent random variables

X and Y are independent and continuous random variables with pdf's f_X and f_Y respectively. Find the distribution of $Z = X + Y$.

Solution:

$$\begin{aligned}
F_Z(z) &= \mathbb{P}(X + Y \leq z) \\
&= \iint\limits_{\{x+y\leq z\}} f_X(x) f_Y(y)\, \mathrm{d}x\mathrm{d}y \\
&= \int\limits_{y=-\infty}^{\infty} \int\limits_{x=-\infty}^{z-y} f_X(x) f_Y(y)\, \mathrm{d}x\mathrm{d}y = \int\limits_{y=-\infty}^{\infty} F_X(z-y) f_Y(y)\, \mathrm{d}y.
\end{aligned}$$

We call Z the convolution of X and Y. Differentiating w.r.t. z we get :

$$f_Z(z) = \int\limits_{y=-\infty}^{\infty} f_X(z-y) f_Y(y)\, \mathrm{d}y.$$

Exercises: sum of independent random variables

1. X and Y are independent and identically distributed (iid) Unif$[0, 1]$. Find the distribution of $Z = X + Y$.
2. X and Y are iid Exp(λ). Find the distribution of $Z = X + Y$.
3. If $X \sim$ Gamma(α, λ) and $Y \sim$ Gamma(β, λ) and $X \perp Y$ (independent). Find the distribution of $Z = X + Y$.

   Recall
   $$f_X(x) = \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, \quad 0 < x < \infty.$$

These become easier when we learn moment generating functions.

Examples: sum of independent random variables

Try these at your convenience.

1. If $X_1, \ldots, X_n$ are iid Exp(λ) then $\sum_{i=1}^n X_i \sim$ Gamma(n, λ).
2. The sum of squares of n independent $N(0, 1)$ random variables is Gamma distributed with parameters $(n/2, 1/2)$, known as a chi-squared distribution.

   If $n = 2$, we have an Exp$(1/2)$ r.v.

These also become much easier when we learn moment generating functions.

Normal

Let X and Y be independent normal random variables with parameters (μ_1, σ_1^2) and (μ_2, σ_2^2) respectively. Then $X + Y$ is also normal, with parameters $(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Proof: later. Generalizes to $X_1 + \cdots + X_n$. What about $X - Y$?

In fact, any for any $a, b \in \mathbb{R}$ we have

$$aX + bY \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2).$$

Example: You play 100 games against a superior opponent (probability of winning each game $= 0.4$), and 100 games against an inferior opponent (probability of winning $= 0.8$). What is the probability that you win more than 110 games?

Conditional distribution: discrete case

1. Discrete case: X and Y have joint pmf $p(x, y)$ and marginals p_X, p_Y. Then

$$p_{X|Y}(x|y) = \mathbb{P}(X = x|Y = y) = \frac{p(x, y)}{p_Y(y)}$$

   if $p_Y(y) > 0$ and is 0 otherwise.

   We can define the conditional CDF here as:

$$F_{X|Y}(a|y) = \mathbb{P}(X \leq a|Y = y) = \sum_{x \leq a} p_{X|Y}(x|y).$$

2. If X and Y are independent then for all x, y with $p_Y(y) > 0$

$$p_{X|Y}(x|y) = \frac{p(x, y)}{p_Y(y)} = \frac{p_X(x)p_Y(y)}{p_Y(y)} = p_X(x).$$

Example

1. X and Y are independent and $X \sim$ Poisson(λ) and $Y \sim$ Poisson(μ). Find the conditional distribution of $X|X+Y=n$.
   Answer: $Z \sim$ Bin$(n, \frac{\lambda}{\lambda+\mu})$.
2. X and Y are independent and $X \sim$ Bin(n_1, p) and $Y \sim$ Bin(n_2, p). Find the conditional distribution of $X|X+Y=k$.
   Answer: $Z \sim$ Hypergeometric(n_1+n_2, n_1, k).

Conditional distribution: continuous case

1. X and Y have joint pdf $f(x, y)$ and marginal pdfs f_X, f_Y. Then

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}, \quad x \in \mathbb{R}, y \in \mathbb{R}$$

   if $f_Y(y) > 0$ and is 0 otherwise.

   We can define the conditional CDF of X given $Y = y$ as:

$$F_{X|Y}(x|y) = \int_{-\infty}^{x} f_{X|Y}(a|y)\,\mathrm{d}a.$$

2. If X and Y are independent then for all x, y with $f_Y(y) > 0$

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

Example

1. Suppose (X, Y) has uniform distribution on the circle of radius 1. Find the conditional pdf of $X|Y = y$.
2. X and Y are independent and identically distributed (i.i.d.) Exp(1). Let $T = X + Y$. Find the conditional pdf of $X|T = t$.

- Keep in mind: BAYES' RULE:

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

Exercise

X and Y are independent and identically distributed (i.i.d.) Exp(1). Let $T = X + Y$. Find the conditional pdf of $X|T = t$.

SOLUTION: The joint pdf of X, Y is given by

$$f_{X,Y}(u, v) = e^{-u}e^{-v}, \quad 0 < u < \infty, 0 < v < \infty.$$

For $x < t$,

$$\begin{aligned}
\mathbb{P}(X \le x, T \le t) = \mathbb{P}(X \le x, X + Y \le t) &= \int_0^x \int_0^{t-u} e^{-u}e^{-v}\mathrm{d}v\,\mathrm{d}u \\
&= \int_0^x e^{-u} \int_0^{t-u} e^{-v}\mathrm{d}v\,\mathrm{d}u \\
&= \int_0^x e^{-u}\left[1 - e^{-(t-u)}\right]\mathrm{d}u = 1 - e^{-x} - xe^{-t}.
\end{aligned}$$

Hence, $$f_{X,T}(x, t) = \frac{\partial^2}{\partial x \partial t}(F_{X,T}(x, t)) = e^{-t}, \quad 0 < x < t < \infty,$$

and, $$f_T(t) = \int_0^\infty f_{X,T}(x, t)\mathrm{d}x = \int_0^t e^{-t}\mathrm{d}x = te^{-t}, \quad 0 < x < t < \infty.$$

Therefore, $$f_{X|T}(x|t) = \frac{f_{X,T}(x, t)}{f_T(t)} = \frac{1}{t}, \quad 0 < x < t.$$

Example

For the bold at heart: tread carefully.

(X, Y) follows Bivariate Normal with parameters $(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$ where $-1 < \rho < 1$ if its pdf for $-\infty < x < \infty, -\infty < y < \infty$ is given by:

$$f(x, y) = \frac{\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]\right\}}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}.$$

Then

$$X|Y = y \sim N\left(\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y - \mu_Y), \sigma_X^2(1-\rho^2)\right),$$

$$Y|X = x \sim N\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X), \sigma_Y^2(1-\rho^2)\right).$$

Conditional distribution: discrete and continuous

- If X is continuous with pdf f and N is discrete then we can find the conditional distribution of $X|N = n$.

$$\mathbb{P}(x < X < x + dx|N = n) = \frac{\mathbb{P}(N = n|x < X < x + dx)}{\mathbb{P}(N = n)} \mathbb{P}(x < X < x + dx).$$

  Now we can think of

$$f_{X|N}(x|n) = \frac{\mathbb{P}(N = n|X = x)}{\mathbb{P}(N = n)} f(x).$$

- Exercise: Say n Bernoulli trials have a common probability of success $U \sim$ Unif$(0, 1)$. So $X|U = u \sim$ Bin(n, u). What is the conditional density of the success probability given that there were k successes?
- What is the distribution of X?

Change of variables: functions of random variables

- (X_1, X_2) are jointly continuous with pdf f_{X_1,X_2}.
- g_1, g_2 are two functions $y_1 = g_1(x_1, x_2)$ and $y_2 = g_2(x_1, x_2)$ such that they can be uniquely solved for x_1 and x_2 with $x_1 = h_1(y_1, y_2)$ and $x_2 = h_2(y_1, y_2)$ (one-one functions).
- We want joint pdf of $Y_1 = g_1(X_1, X_2),\ Y_2 = g_2(X_1, X_2)$ from that of $X_1,\ X_2$.
- Assume g_1, g_2 has continuous partial derivatives at all points (x_1, x_2).
- Calculate the determinant of the Jacobian:

$$J(x_1, x_2) = \left\| \begin{matrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{matrix} \right\|$$

- Check that $|J(x_1, x_2)| \neq 0, \quad \forall (x_1, x_2)$.
- Then

$$\boxed{f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(x_1, x_2)\ |J(x_1, x_2)|^{-1}}$$

where $x_1 = h_1(y_1, y_2),\ x_2 = h_2(y_1, y_2)$.

Examples

Suppose $(X, Y) \sim f_{X,Y}$.

1. If $Z_1 = X + Y, Z_2 = X - Y$, find f_{Z_1,Z_2}.

   – Here, $J(x, y) = -2$, hence

$$f_{Z_1,Z_2}(z_1, z_2) = \frac{1}{2} f_{X,Y}\left(\frac{z_1 + z_2}{2}, \frac{z_1 - z_2}{2}\right).$$

   – In this case, now assume that X, Y are iid Unif$(0, 1)$. Find f_{Z_1,Z_2}.

   – Here $f_{Z_1,Z_2}(y_1, y_2) = \frac{1}{2}$ on a 'diamond'.

Examples

Suppose $(X,Y) \sim f_{X,Y}$.

1. If $R = \sqrt{X^2+Y^2}, \Theta = \arctan\left(\frac{Y}{X}\right)$, find $f_{R,\Theta}$.

   – Here, $J(x,y) = (\sqrt{x^2+y^2})^{-1}$, hence

$$f_{R,\Theta}(r,\theta) = f_{X,Y}(x,y)|J(x,y)|^{-1} = r f_{X,Y}(r\cos\theta, r\sin\theta).$$

   – In this case, now assume that X and Y are iid $N(0,1)$. Find the joint pdf of (R,Θ).

   – Here

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

   R and Θ are independent.

Examples

1. If $X \sim$ Gamma(α, λ) and $Y \sim$ Gamma(β, λ), $X \perp Y$ find the pdf of $U := \frac{X}{X+Y}$. (Try the easier case $\alpha = \beta = \lambda = 1$.)

   METHOD 1: Clearly $0 < U < 1$. Hence for any $0 < u < 1$,

$$\mathbb{P}(U \leq u) = \mathbb{P}(X \leq u(X+Y)) = \mathbb{P}((1/u - 1)X \leq Y)$$
$$= \int_{x=0}^{\infty} \int_{y=(1/u-1)x}^{\infty} e^{-x} e^{-y} \, \mathrm{d}y \mathrm{d}x = \int_{0}^{\infty} e^{-x/u} \mathrm{d}x = u.$$

   METHOD 2: change of variable $U = \frac{X}{X+Y}$, $V = X + Y$.

2. If X_1, X_2, X_3 are iid Unif$(0,1)$ find the distribution of $Z = \max(X_1, X_2, X_3)$.

Expectation

- Discrete case: $\mathbb{E}\,X = \sum_{x:p(x)>0} xp(x)$.
- Continuous case: $\mathbb{E}\,X = \int_{-\infty}^{\infty} xf(x)\mathrm{d}x.$
- If X, Y have joint pdf (or pmf) given by $f(x,y)$ (or $p(x,y)$), then

$$\mathbb{E}(g(X,Y)) = \begin{cases} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\,f(x,y)\,\mathrm{d}x\mathrm{d}y & \text{(continuous case)} \\ \sum_{x,y:p(x,y)>0} g(x,y)\,p(x,y) & \text{(discrete case)} \end{cases}$$

Expectation: properties

- If $\mathbb{P}(a \leq X \leq b) = 1$, then $a \leq \mathbb{E}(X) \leq b$.
- If $g(X, Y) = X + Y$, we get $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$.
- (If $\mathbb{E}\,|X_i| < \infty$ then)
  $\mathbb{E}(X_1 + \ldots + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \ldots \mathbb{E}(X_n)$.
- If $X \geq Y$ then $\mathbb{E}(X) \geq \mathbb{E}(Y)$. [2]

Examples

1. $X \sim$ Bin(n, p). Hence $X = X_1 + X_2 + \ldots X_n$ where $X_i \sim$ Bernoulli(p) (independent). So $\mathbb{E}\,X = np$.
2. If $X_1, \ldots X_n$ are iid with expected value μ. Then $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is called the sample mean. Find $\mathbb{E}(\overline{X})$. Clearly $\mathbb{E}(\overline{X}) = \mu$.

[2] What does $X \geq Y$ mean? For every realization of the random experiment $\omega \in \Omega$, we have $X(\omega) \geq Y(\omega)$.

Examples

1. An accident occurs at a point uniformly distributed on a 1km stretch of road. An ambulance is independently located at a point also uniformly distributed on the road. Find the expected distance the ambulance has to travel to the site of the accident.

   SOLUTION: $\mathbb{E}(|X-Y|) = \int_0^1 \int_0^1 |x-y|\, \mathrm{d}x \mathrm{d}y = 1/3.$

2. Coupon Collector's problem: Suppose there are N types of coupons. Each type of coupon is equally likely to appear in a draw; find the expected number of coupons one needs to collect to get a complete set.

   SOLUTION:

   $X_i = \#$ of additional coupons needed to get a new type, after i types have been collected.
   $\mathbb{P}(X_i = k) = (i/N)^{k-1}(1 - i/N)$. $\quad X_i \sim$ Geometric$(\frac{N-i}{N})$
   $\mathbb{E}(X) = \mathbb{E}(X_0 + \cdots + X_{N-1}) = \sum_{i=0}^{N-1} N/(N-i) = N \sum_{j=1}^{N} 1/j.$
   For large N: $\mathbb{E}\, X \approx N \log N$

More examples

1. N people throw their (indistinguishable) hats into the middle of a room; later each person randomly selects a hat. Find the expected number of people who pick their own hat.

   SOLUTION: Define the indicator random variables

$$X_i = \begin{cases} 1 & i^{\text{th}} \text{ person gets his hat} \\ 0 & \text{otherwise.} \end{cases}$$

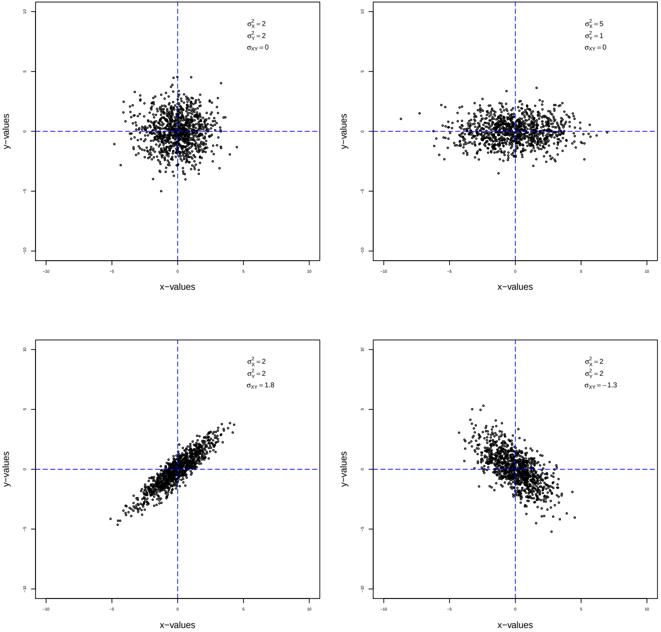
$$\mathbb{E}(X) = \sum_{i=1}^{N} \mathbb{E}\, X_i = N \cdot \tfrac{1}{N} = 1.$$

2. Hypergeometric distribution: Suppose n balls are randomly selected from N balls, m of which are white. Find the expected number of white balls selected.

   SOLUTION: Let $X = X_1 + \cdots + X_m$ denote the number of white balls selected. Here

$$X_i = \begin{cases} 1 & \text{if the } i^{\text{th}} \text{ white ball is selected,} \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\mathbb{E}(X_i) = \mathbb{P}(X_i = 1) = \binom{N-1}{n-1} / \binom{N}{n} = \frac{n}{N}$. Hence $\mathbb{E}(X) = \sum_{i=1}^{m} \mathbb{E}\, X_i = \frac{nm}{N}$.


$\sigma_X^2 = 2$
$\sigma_Y^2 = 2$
$\sigma_{XY} = 0$
$\sigma_X^2 = 5$
$\sigma_Y^2 = 1$
$\sigma_{XY} = 0$
$\sigma_X^2 = 2$
$\sigma_Y^2 = 2$
$\sigma_{XY} = 1.8$
$\sigma_X^2 = 2$
$\sigma_Y^2 = 2$
$\sigma_{XY} = -1.3$
x−values
y−values

Independence, covariance

- If X and Y are independent then for any functions g and h,,

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

  in particular, $\mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y)$.
- Recall Var$(X) = \mathbb{E}[X - \mathbb{E}(X)]^2$.
- The covariance of X and Y is defined as

  Cov$(X, Y) = \mathbb{E}\left[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right].$
- Equivalently,

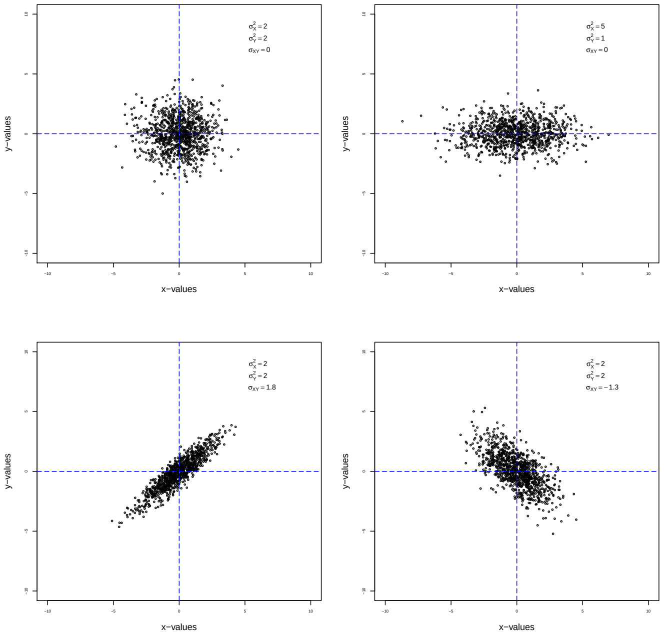
  Cov$(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\,\mathbb{E}(Y).$
- Hence, if X and Y are independent Cov$(X, Y) = 0$.

Properties of covariance

- $\text{Cov}(X, X) = \text{Var}(X)$.
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.
- $\text{Cov}(aX + b, cY + d) = ac\,\text{Cov}(X, Y)$.
- Does $\text{Cov}(X, Y) = 0$ imply independence? No

  Counterexample: Consider

  $\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \mathbb{P}(X = -1) = \frac{1}{3}$ and $Y = 0$ iff $X \neq 0$.



$\sigma_X^2 = 2$
$\sigma_Y^2 = 2$
$\sigma_{XY} = 0$
$\sigma_X^2 = 5$
$\sigma_Y^2 = 1$
$\sigma_{XY} = 0$
$\sigma_X^2 = 2$
$\sigma_Y^2 = 2$
$\sigma_{XY} = 1.8$
$\sigma_X^2 = 2$
$\sigma_Y^2 = 2$
$\sigma_{XY} = -1.3$
x−values
y−values

Properties of covariance

- Suppose $\mathbb{E}\, X_i = \mu_i$ and $\mathbb{E}\, Y_j = \nu_j$ where $i = 1, \ldots, m$ and $j = 1, \ldots, n$. Then

$$\text{Cov}\Big(\sum_{i=1}^{m} X_i, \sum_{j=1}^{n} Y_j\Big) = \sum_{i=1}^{m}\sum_{j=1}^{n} \text{Cov}(X_i, Y_j)$$

- $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\,\text{Cov}(X, Y)$.
- In general,

$$\text{Var}\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} \text{Var}(X_i) + 2\sum_{i<j}^{n} \text{Cov}(X_i, X_j)$$

Examples

Suppose $X_i, \dots, X_n$ are iid $\sim F$. Let $\mathbb{E}(X_1) = \mu$ and Var$(X_1) = \sigma^2$.

- Define
$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$
- Show that
$$\mathbb{E}\,\overline{X} = \mu, \quad \text{Var}(\overline{X}) = \frac{\sigma^2}{n}, \quad \mathbb{E}\,S^2 = \sigma^2.$$
- If $X \sim \text{Bin}(n, p)$ find Var(X).

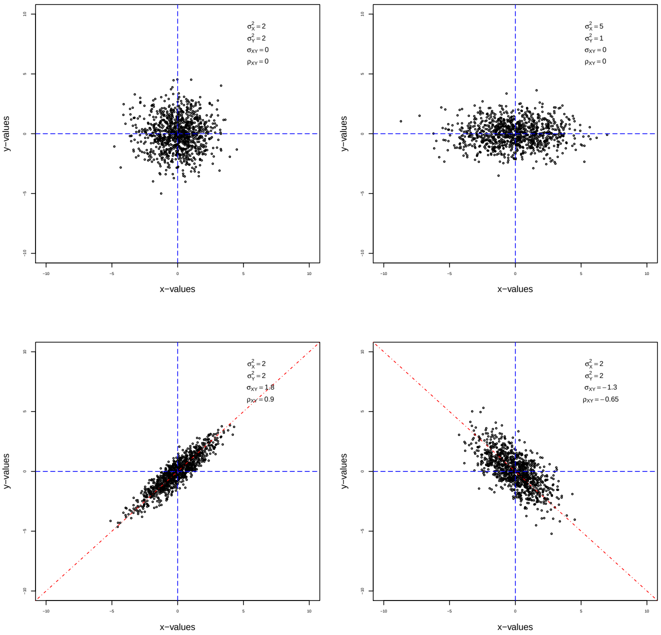
Correlation

The correlation between random variables X and Y is defined as

$$\rho(X,Y) = \text{corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\,\text{Var}(Y)}}.$$

This ρ is called Pearson's correlation coefficient and measures linear dependence between X and Y.

- Check that $-1 \leq \rho(X,Y) \leq 1$.
  Let $\sigma_X^2 = \text{Var}(X)$ and $\sigma_Y^2 = \text{Var}(Y)$. Then Var $\left(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y}\right) \geq 0.$
  Thus $2(1 \pm \rho(X,Y)) \geq 0$. Hence the result.



$\sigma_X^2 = 2$
$\sigma_Y^2 = 2$
$\sigma_{XY} = 0$
$\rho_{XY} = 0$
$\sigma_X^2 = 5$
$\sigma_Y^2 = 1$
$\sigma_{XY} = 0$
$\rho_{XY} = 0$
$\sigma_X^2 = 2$
$\sigma_Y^2 = 2$
$\sigma_{XY} = 1.8$
$\rho_{XY} = 0.9$
$\sigma_X^2 = 2$
$\sigma_Y^2 = 2$
$\sigma_{XY} = -1.3$
$\rho_{XY} = -0.65$
x−values
y−values

Example

- Random walk in a plane: Start from the origin O, take n unit steps, each along a uniformly chosen direction. Find the expected square of the distance from O.
  - $\Theta_1, \Theta_2, \ldots, \Theta_n$ are iid Unif$(0, 2\pi)$.
  - $X_i = \cos\Theta_i,\ Y_i = \sin\Theta_i$.
  - $D_n^2 = \left(\sum_{i=1}^n X_i\right)^2 + \left(\sum_{i=1}^n Y_i\right)^2$
    $= n + \sum_{i\neq j}(\cos\Theta_i \cos\Theta_j + \sin\Theta_i \sin\Theta_j).$
  - Take the expectation and use independence. $\mathbb{E}(D_n^2) = n$.

What about $\mathbb{E}(D_n)$?

Conditional expectation

Discrete case: We know that $p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)}$.

Hence

$$\begin{aligned}\mathbb{E}(X|Y=y) &= \sum_x x\ \mathbb{P}(X=x|Y=y) \\ &= \sum_x x\, p_{X|Y}(x|y) = g(y)\end{aligned}$$

where g is a function of y.

Tower property (Law of Total Expectation): $\boxed{\mathbb{E}(X) = \mathbb{E}\left[\mathbb{E}(X|Y)\right]}$

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}[X|Y]\right] = \mathbb{E}[g(Y)] &= \sum_y g(y)\ \mathbb{P}(Y=y) \\ &= \sum_y \left[\sum_x x\ \mathbb{P}(X=x|Y=y)\right] \mathbb{P}(Y=y) \\ &= \sum_x x \left[\sum_y \mathbb{P}(X=x|Y=y)\,\mathbb{P}(Y=y)\right] = \sum_x x\ \mathbb{P}(X=x) = \mathbb{E}(X).\end{aligned}$$

Example

1. X, Y independent $X \sim$ Poisson$(\lambda), Y \sim$ Poisson(μ). Find $\mathbb{E}(X|X+Y=n)$.

   $X|X+Y=n \sim$ Bin$(n, \frac{\lambda}{\lambda+\mu})$.

   $\mathbb{E}(X|X+Y=n) = n\frac{\lambda}{\lambda+\mu}$.

2. X, Y iid Bin(n,p). Find $\mathbb{E}(X|X+Y=m)$.

   $X|X+Y=m \sim$ Hypergeometric(n,n,m). Why? See below.

   $\mathbb{E}(X|X+Y=m) = \frac{m}{2}$.

For $0 \le k \le \min(m,n)$ and $0 \le m \le 2n$,

$$\begin{aligned}\Pr(X=k|X+Y=m) &= \frac{\Pr(X=k, X+Y=m)}{\Pr(X+Y=m)} = \frac{\Pr(X=k, Y=m-k)}{\Pr(X+Y=m)} \\ &= \frac{\Pr(X=k)\Pr(Y=m-k)}{\Pr(X+Y=m)} \quad \text{(using independence of } X \text{ and } Y\text{)} \\ &= \binom{n}{k}\binom{n}{m-k}\Big/\binom{2n}{m}.\end{aligned}$$

Example

1. A miner is trapped in a mine with three doors. Door 1 leads to safety in 3 hours. Door 2 leads back to the mine in 5 hours and door 3 leads back to the mine in 7 hours. Unfortunately the doors are not marked and he picks a door at random. Find the miner's expected time to freedom in the following cases:
   - he choses a door randomly in front of him at all times (not so smart). 15
   - he choses randomly from a door not already chosen (a little smarter). 9

Miner problem: solution

1. Let X = number of hours till freedom for the miner.

   Let Y = door number chosen by the miner whenever he is in the cave. Clearly, $\Pr(Y = 1) = \Pr(Y = 2) = \Pr(Y = 3) = 1/3$.

   Moreover, we can see that

   - $\mathbb{E}(X|Y = 1) = 3$.
   - $\mathbb{E}(X|Y = 2) = 5 + \mathbb{E}(X)$.
   - $\mathbb{E}(X|Y = 3) = 7 + \mathbb{E}(X)$.

   Now, using the Tower property, we have

$$
\begin{aligned}
\mathbb{E}(X) &= \mathbb{E}[\mathbb{E}[X|Y]] \\
&= \mathbb{E}(X|Y = 1)\Pr(Y = 1) + \mathbb{E}(X|Y = 2)\Pr(Y = 2) + \mathbb{E}(X|Y = 3)\Pr(Y = 3) \\
&= 5 + \frac{2}{3}\mathbb{E}(X).
\end{aligned}
$$

   Hence $\mathbb{E}(X) = 15$.

2. Let W = number of hours till freedom for the miner and Z be the random variable providing the sequence of doors, i.e., 1, 21, 31, 231, 321, which will be chosen with probabilities 1/3, 16, 1/6, 1/6, 1/6 respectively. Then a straightforward calculation gives $\mathbb{E}(W) = 9$.

Conditional expectation

Continuous case: We know that $f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$.

Hence

$$\mathbb{E}(X|Y=y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\mathrm{d}x = g(y)$$

where g is a function of y.

We can prove as before that the Tower property (law of total expectation) holds

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

Calculating probabilities via conditioning

Suppose E is an event whose probability we want to calculate.

Conditioning may help.

$$X = \begin{cases} 1 & \text{if } E \text{ occurs} \\ 0 & \text{if } E \text{ does not occur} \end{cases}$$

Then $\mathbb{E}(X) = \mathbb{P}(E)$. Clearly

$$\mathbb{E}[X|Y = y] = \mathbb{P}(X = 1|Y = y) = \mathbb{P}(E|Y = y).$$

Hence

$$\mathbb{P}(E) = \begin{cases} \int\limits_{-\infty}^{\infty} \mathbb{P}(E|Y = y)\, f_Y(y)\mathrm{d}y & \text{(continuous case)} \\ \sum\limits_{y} \mathbb{P}(E|Y = y)\, \mathbb{P}(Y = y) & \text{(discrete case)} \end{cases}$$

Example

Suppose X and Y are independent with $X \sim$ Exp(λ) and $Y \sim$ Exp(μ).

1. Find $\mathbb{P}(X < Y)$. $\frac{\lambda}{\lambda+\mu}$
2. Find $\mathbb{P}(X + Y \leq t)$ if $\lambda = \mu$. $X + Y \sim$ Gamma$(2, \lambda)$

Example: solution

Suppose X and Y are independent with $X \sim \text{Exp}(\lambda)$ and $Y \sim \text{Exp}(\mu)$.

1.

$$
\begin{aligned}
\Pr(X < Y) &= \int_0^\infty \Pr(X < Y | Y = y) f_Y(y)\,\mathrm{d}y \\
&= \int_0^\infty \Pr(X < y | Y = y) f_Y(y)\,\mathrm{d}y = \int_0^\infty \Pr(X < y) \mu e^{-\mu y}\,\mathrm{d}y \quad (X \perp Y) \\
&= \int_0^\infty (1 - e^{-\lambda y}) \mu e^{-\mu y}\,\mathrm{d}y = 1 - \frac{\mu}{\lambda + \mu} = \frac{\lambda}{\lambda + \mu}.
\end{aligned}
$$

2. Similarly as above since $X \perp Y$ and $\lambda = \mu$,

$$
\begin{aligned}
\Pr(X + Y \leq t) &= \int_0^\infty \Pr(X + Y \leq t | Y = y) f_Y(y)\,\mathrm{d}y \\
&= \int_0^\infty \Pr(X \leq t - y | Y = y) f_Y(y)\,\mathrm{d}y = \int_0^t \Pr(X \leq t - y) \lambda e^{-\lambda y}\,\mathrm{d}y \\
&= \int_0^t (1 - e^{-\lambda(t-y)}) \lambda e^{-\lambda y}\,\mathrm{d}y = 1 - e^{-\lambda t} - \lambda t e^{-\lambda t}.
\end{aligned}
$$

Summary

1. Modeling random phenomenon: discrete or continuous.
2. Doing so for more than one random variable, random vectors, ... may be random processes.
3. Calculating probabilities (likelihood/ chance), expectations (likely/ average outcome), variances (volatility/ dispersion).

   Next:
4. One technique that makes calculation easier: MGFs.
5. Behaviour of a sample of homogeneous items.

Moment generating function

Fact: (most) pdfs and pmfs are uniquely determined by their moments: $\mathbb{E}\,X, \mathbb{E}\,X^2, \mathbb{E}\,X^3, \ldots$

Definition: For any random variable X, the moment generating function or m.g.f. is given by

$$M_X(t) = \mathbb{E}(e^{tX}), \quad t \in \mathbb{R} \text{ such that } \mathbb{E}(e^{tX}) < \infty.$$

Hence depending on whether X is discrete or continuous, we have,

$$M_X(t) = \sum_x e^{tx}\, p(x) \qquad \text{or} \qquad \int_{\mathbb{R}} e^{tx} f(x)\, \mathrm{d}x.$$

We can show that

$$M_X(t) = \mathbb{E}(e^{tX}) = \mathbb{E}\Big(\sum_{n=0}^{\infty} \frac{X^n}{n!} t^n\Big) = \sum_{n=0}^{\infty} \frac{t^n}{n!}\, \mathbb{E}(X^n).$$

MGF properties

Note that $M_X(t) = \mathbb{E}(e^{tX})$. Hence $M_X(0) = 1$.

$$
\begin{aligned}
M_X'(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\,\mathbb{E}(e^{tX}) \\
&= \mathbb{E}\left(\frac{\mathrm{d}}{\mathrm{d}t} e^{tX}\right) \quad \text{(this needs justification)} \\
&= \mathbb{E}\left(X e^{tX}\right).
\end{aligned}
$$

- $t = 0 : M_X'(0) = \mathbb{E}(X)$.
- Similarly we have $M_X{}''(0) = \mathbb{E}(X^2), M_X{}'''(0) = \mathbb{E}(X^3)$.
- In general, $\mathbb{E}(X^n) = M_X^{(n)}(0)$.

MGF examples

We can easily find $\mathbb{E}(X)$ and Var(X) using mgf.

1. $X \sim \text{Bin}(n, p), \quad M_X(t) = (pe^t + 1 - p)^n, t \in \mathbb{R}.$
2. $X \sim \text{Poisson}(\lambda), \quad M_X(t) = e^{\lambda(e^t - 1)}, \quad t \in \mathbb{R}.$
3. $X \sim \text{Geometric}(p), \quad M_X(t) = \frac{e^t p}{(1 - e^t(1-p))}, \quad t > -\ln p.$
4. $X \sim \text{Exp}(\lambda), \quad M_X(t) = \frac{\lambda}{\lambda - t}, \quad t < \lambda.$
5. $Z \sim \mathcal{N}(0, 1), \quad M_Z(t) = e^{t^2/2}, \ t \in \mathbb{R}.$
6. $X \sim \mathcal{N}(\mu, \sigma^2), \quad M_X(t) = M_{\mu + \sigma Z}(t) = \exp\{\frac{\sigma^2 t^2}{2} + \mu t\}, t \in \mathbb{R}.$
7. $X \sim \text{Gamma}(\alpha, \lambda), \quad M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^\alpha, \quad t < \lambda.$

MGF: properties

- Theorem (without proof) A moment generating function (when it exists) uniquely determines the distribution of a random variable.
- If $X \sim F_X$ and $Y \sim F_Y$. Then $M_X(t) = M_Y(t)$ for all t if and only if $F_X(z) = F_Y(z)$ for all $z \in \mathbb{R}$ (means X and Y have the same distribution).
- If X and Y are independent, then for all t where $\mathbb{E}(e^{t(X+Y)}) < \infty$,

$$M_{X+Y}(t) = \mathbb{E}(e^{t(X+Y)}) = \mathbb{E}(e^{tX})\,\mathbb{E}(e^{tY}) = M_X(t)\,M_Y(t).$$

Exercise

Show the following:

1. If $X \perp Y$ with $X \sim$ Poisson(λ) and $Y \sim$ Poisson(μ), then $X + Y \sim$ Poisson$(\lambda + \mu)$.
2. If $X \perp Y$ with $X \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$, then $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.
3. If $X_1, X_2, \ldots, X_n$ are iid Exp(λ) then $X_1 + \ldots + X_n \sim$ Gamma(n, λ).

Introduction: Limit Theorems

- Why do we care about mean, variance, moments of a distribution?
- Can we say something meaningful with this knowledge?
- How does the average of a random variable behave?
- Can we say anything about the distribution of the average?

Markov's Inequality

If X is a non-negative random variable, then for $a > 0$,

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

Proof: Let I be 1 if $X \geq a$ and 0 otherwise. Then $\mathbb{E}(I) = \mathbb{P}(A)$.

Now since $I \leq \frac{X}{a}$, we have

$$\mathbb{P}(X \geq a) = \mathbb{E}(I) \leq \frac{\mathbb{E}(X)}{a}.$$

Example: Suppose $X \sim \text{Bin}(100, 0.2)$. How big is $\mathbb{P}(X \geq 50)$?

- Markov: $\mathbb{P}(X \geq 50) \leq \mathbb{E}(X)/50 = 20/50 = 0.4$.
- Normal Approximation: $\mathbb{P}(X \geq 50) \approx 1 - \Phi(29.5/4) = 0$.

  Caveat: Markov's inequality does not always provide a great

Chebyshev's inequality

Let X be a random variable with mean μ and variance σ^2.Then for any $k > 0$,

$$\mathbb{P}(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}.$$

Proof: A direct consequence of Markov's inequality.

Back to previous example:

Suppose $X \sim$ Bin$(100, 0.2)$. How big is $\mathbb{P}(X \geq 50)$?

- Markov: $\mathbb{P}(X \geq 50) \leq \mathbb{E}(X)/50 = 20/50 = 0.4$.
- Normal Approximation: $\mathbb{P}(X \geq 50) \approx 1 - \Phi(29.5/4) = 0$.
- Chebyshev:

  $\mathbb{P}(X \geq 50) = \mathbb{P}(|X - 20| \geq 30) \leq \text{Var}(X)/30^2 = 16/900 = 0.0178$.

Examples

1. Give an upper limit for $\mathbb{P}(|X-\mu| \geq 2\sigma) \leq 1/4$.
   The bounds of Chebyshev are also not very tight. If $X \sim N(\mu, \sigma^2)$ then $\mathbb{P}(|X-\mu| \geq 2\sigma) \approx 0.0456.$
2. A factory produces items with mean 50 and variance 25. What can be said about $\mathbb{P}$(the production is between 40 and 60)? What about greater than 75?
3. X has mean 5 and standard deviation 2. Is it possible that $\mathbb{P}(-3 \leq X \leq 13) = 0.9$?
4. If $\sigma^2 = 0$, then $\mathbb{P}(X = \mu) = 1$.

Weak law of large numbers

Suppose X_i's are iid with mean $\mathbb{E}(X_1) = \mu$. Then for any fixed $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left| \frac{X_1 + \cdots + X_n}{n} - \mu \right| > \epsilon \right) = 0.$$

Proof: A direct consequence of Chebyshev's inequality if we assume $\sigma^2 = \text{Var}(X_1) < \infty$.

So the sample mean converges in probability towards the population mean. We write $\overline{X} \to \mu$ in probability or $\overline{X} \xrightarrow{P} \mu$, where

$$\overline{X} = \frac{X_1 + \ldots + X_n}{n}.$$

Central limit theorem

Suppose X_i's are iid with mean $\mathbb{E}(X_1) = \mu$ and Var$(X_1) = \sigma^2$. Then

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{X_1 + \cdots + X_n - n\,\mu}{\sqrt{n}\,\sigma} \leq a\right) = \lim_{n\to\infty} \mathbb{P}\left(\frac{n(\overline{X} - \mu)}{\sqrt{n}\,\sigma} \leq a\right) = \Phi(a),$$

where Φ is the distribution of the standard normal.

This means for large n, $\overline{X} \sim N(\mu, \sigma^2/n)$ approximately.

- Note relationship between CLT and WLLN.
- Hence Normal approximation to Binomial is justified.

Proof of CLT (sketch) [not in syllabus]

Lemma: Suppose Z_n if a sequence of random variables with mfg M_n. Then $M_{Z_n}(t) \to M_Z(t)$ for all t if and only if $\mathbb{P}(Z_n \le z) \to \mathbb{P}(Z \le z)$ for all $z \in \mathbb{R}$.

Proof of CLT: Define
$$W_n = \frac{n(\overline{X} - \mu)}{\sqrt{n}\,\sigma} = \frac{X_1 + \cdots + X_n - n\,\mu}{\sqrt{n}\,\sigma}.$$

Assume without loss of generality $\mu = 0, \sigma^2 = 1$. Then

$$M_{W_n}(t) = \mathbb{E}\left(\exp\left(\frac{1}{\sqrt{n}}(X_1 + \ldots + X_n)t\right)\right) = \left(M_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n.$$

Taking logarithms,

$$\begin{aligned}
\log M_{W_n}(t) &= n \log M_{X_1}\left(\frac{t}{\sqrt{n}}\right) \\
&= n \log\left[M_{X_1}(0) + M'_{X_1}(0)\frac{t}{\sqrt{n}} + M''_{X_1}(0)\frac{t^2}{2n} + o(n^{-1})\right] \\
&= n \log\left[1 + \frac{t^2}{2n} + o(n^{-1})\right] \\
&\approx n\left(\frac{t^2}{2n} + o(n^{-1})\right) \to \frac{t^2}{2} \qquad (n \to \infty).
\end{aligned}$$

Examples

1. 100 fair dice are rolled; find the probability that the sum ≤ 330.

   SOLUTION: $\mathbb{E}(X_i) = 7/2,$ $\text{Var}(X_i) = 35/12$.

   CLT:

$$\begin{aligned}\mathbb{P}(S_n \leq 330) &= \mathbb{P}(S_n \leq 330.5)\\ &\approx \mathbb{P}\big(Z \leq \sqrt{100}\left(\frac{330.5}{100} - \frac{7}{2}\right) / \sqrt{35/12}\big) \approx 0.13.\end{aligned}$$

Examples

2. Measurements X_i are iid with mean μ and $\sigma = 2$. How many of them are needed to be 95% sure that $\overline{X}$ is within 0.5 of μ?

   SOLUTION:

   Chebyshev:

   $0.05 = \mathbb{P}(|\overline{X} - \mu| \geq 0.5) \leq \frac{\sigma^2/n}{0.5^2} = \frac{16}{n}$. Hence $n = 320$.

   CLT:

$$0.95 = \mathbb{P}(|\overline{X} - \mu| \leq 0.5) = \mathbb{P}\left(\frac{|\overline{X} - \mu|}{\sigma/\sqrt{n}} \leq \frac{0.5}{\sigma/\sqrt{n}}\right)$$

$$\approx \mathbb{P}\left(|Z| \leq \frac{\sqrt{n}}{4}\right).$$

   Hence $n \approx 62$.

Important inequalites

MARKOV'S INEQUALITY: If X is a non-negative random variable, then for $a > 0$,

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

CHEBYSHEV'S INEQUALITY: Let X be a random variable with mean μ and variance σ^2. Then for any $k > 0$,

$$\mathbb{P}(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}.$$

Proof: A direct consequence of Markov's inequality.

Weak law of large numbers (WLLN)

Suppose X_i's are iid with mean $\mathbb{E}(X_1) = \mu$. Then for any fixed $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| > \epsilon\right) = 0.$$

Proof: A direct consequence of Chebyshev's inequality if we assume $\sigma^2 = \text{Var}(X_1) < \infty$.

So the sample mean converges in probability towards the population mean. We write $\overline{X}_n \to \mu$ in probability or $\overline{X}_n \xrightarrow{P} \mu$, where

$$\overline{X}_n = \frac{X_1 + \ldots + X_n}{n}.$$

STRONG LAW OF LARGE NUMBERS (SLLN): [Not tested]

Notions of Convergence [not tested]

Suppose we have a probability space: $(\Omega, \mathcal{A}, \mathbb{P})$

(Sample space, collection of sets in Ω , probability).

$X_n, n \geq 1$ and X are functions from $\Omega \to \mathbb{R}$ (in other words, random variables).

1. SURE CONVERGENCE:

$$\lim_{n\to\infty} X_n(\omega) = X(\omega), \quad \forall \omega \in \Omega. \qquad \texttt{convergence of functions (\`a la Math I)}$$

2. ALMOST SURE CONVERGENCE:

$$\mathbb{P}(\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)) = 1. \qquad \texttt{SLLN}$$

3. CONVERGENCE IN PROBABILITY: For any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(\omega \in \Omega : |X_n(\omega) - X(\omega)| > \epsilon) = 0. \qquad \texttt{WLLN}$$

4. CONVERGENCE IN DISTRIBUTION: If $X_n \sim F_n$ and $X \sim F$, then

$$\lim_{n\to\infty} F_n(x) = F(x), \qquad \texttt{CLT}$$

for all points $x \in \mathbb{R}$ where F is continuous.

Central limit theorem

Suppose X_i's are iid with mean $\mathbb{E}(X_1) = \mu$ and Var$(X_1) = \sigma^2$. Let

$$S_n = \sum_{i=1}^{n} X_i \,, \quad \overline{X}_n = \frac{S_n}{n}.$$

Then for any $a \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - n\,\mu}{\sqrt{n}\,\sigma} \le a\right) = \lim_{n\to\infty} \mathbb{P}\left(\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \le a\right) = \Phi(a),$$

where Φ is the distribution of the standard normal.

- We write $\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{\mathcal{D}} Z$ where $Z \sim N(0,1)$.
- This means for large n, $\overline{X}_n \sim N(\mu, \sigma^2/n)$ approximately.
- Hence Normal approximation to Binomial is justified.
- Note relationship between CLT and WLLN.

Exercise

1. Peter has $\$16$ and Mary has $\$32$. They bet $\$1$ every time on 900 independent tosses of a fair coin. What is the probability that neither are in debt at the end?
2. If you predict in advance the outcomes of 7 tosses of a fair coin, what is the probability of success? If 1000 people do the same prediction independently, what is the probability that at least one of them will be correct?
3. If X follows Poisson(3) what is the third moment of X?

Exercise

4. Liam's bowl of spaghetti contains n strands. He selects two ends at random and joins them together. He does this until there are no ends left. What is the expected number of spaghetti hoops in the bowl?
5. You are randomly dealt 5 cards from a standard deck of cards. What is the expected number of aces you see? What is the variance of the number of aces you see?

Exercise

6. An expedition is sent to the Himalayas with the objective of catching a pair of wild yaks for breeding. Assume yaks are loners and roam about the Himalayas at random. The probability $p \in (0, 1)$ that a given trapped yak is male is independent of prior outcomes. Find the expected number of yaks that must be caught until a pair is obtained.
Solution: $1 + \frac{p}{1-p} + \frac{1-p}{p}$.

Exercise

7. Suppose a chicken lays a Poisson number of eggs, N, with mean λ and each egg fertilizes, independently of others, with probability p. If X is the number of eggs actually fertilized, find the correlation between the number of eggs laid and the number fertlized.

Solution: Note that $\Pr(N = n) = e^{-\lambda}\frac{\lambda^n}{n!}$, for $n = 0, 1, \ldots$ and $\mathbb{E}(N) =$ Var$(N) = \lambda$. For the random variable X, we have $X|N = n \sim$ Bin(n, p). So

$$\begin{aligned}
\Pr(X = k) &= \sum_{n=0}^{\infty} \Pr(X = k, N = n) = \sum_{n=k}^{\infty} \Pr(X = k|N = n) \Pr(N = n) \\
&= \sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} e^{-\lambda} \frac{\lambda^n}{n!} \\
&= \frac{(p\lambda)^k}{k!} e^{-\lambda} \sum_{n-k=0}^{\infty} \frac{((1-p)\lambda)^{n-k}}{(n-k)!} = \frac{(p\lambda)^k}{k!} e^{-\lambda} \cdot e^{(1-p)\lambda} = e^{-p\lambda} \frac{(p\lambda)^k}{k!}.
\end{aligned}$$

Hence $X \sim$ Poisson$(p\lambda)$ and $\mathbb{E}(X) =$ Var$(X) = p\lambda$.

Exercise

7. Suppose a chicken lays a Poisson number of eggs, N, with mean λ and each egg fertilizes, independently of others, with probability p. If X is the number of eggs actually fertilized, find the correlation between the number of eggs laid and the number fertlized.

Solution (continued): Now we find Cov(X, N). Clearly, using tower property (law of total expectation),

$$\begin{aligned}\mathbb{E}(XN) &= \mathbb{E}[\mathbb{E}[(XN)|N]] = \mathbb{E}[N\,\mathbb{E}[X|N]] = \mathbb{E}[N \times Np] \quad (\text{since } X|N = n \sim \text{Bin}(n, p)) \\ &= p(\mathbb{E}(N^2)) = p(\text{Var}(N) + [\mathbb{E}(N)]^2) = p(\lambda + \lambda^2).\end{aligned}$$

Hence

$$\text{Cov}(X, N) = \mathbb{E}[XN] - \mathbb{E}[X]\,\mathbb{E}[N] = p\lambda + p\lambda^2 - (p\lambda)(\lambda) = p\lambda.$$

And we have

$$\text{corr}(X, N) = \frac{\text{Cov}(X, N)}{\sqrt{\text{Var}(X)\text{Var}(N)}} = \frac{p\lambda}{\sqrt{p\lambda \times \lambda}} = \sqrt{p}.$$