MATH2901 Revision Seminar

Part I: Probability Theory

Jeffrey Yang

UNSW Society of Statistics/UNSW Mathematics Society

August 2020

Note

The theoretical content was mostly sourced from Libo's notes.

Most of the examples were taken from the slides Rui Tong (2018 StatSoc Team) made for that year's revision seminar.

All examples are presented at the end so that it isn't as obvious which techniques/methods should be used.

It is recommended that you refer to the official lecture notes when quoting definitions/results.

Table of Contents

1. Basic Probability Theory
2. Review of Random Variables
3. Computations for Common Distributions
4. Inequalities
5. Limit of Random Variables
6. Tips and Assorted Examples

1. Axioms and Basic Results
2. **Conditional Probability** and **Independence**
3. Basic Laws
4. **Bayes Formula**

Probability theory is about modelling and analysing *random experiments*. We can do this mathematically by specifying an appropriate **probability space** consisting of three components:

1. a **sample space**, Ω, which is the set of all possible outcomes.
2. an **event space**, $\mathcal{F}$, which is the set of all events "of interest". Here, we can understand $\mathcal{F}$ as being a set of subsets of Ω which satisfies certain conditions.
3. a **probability function**, $\mathbb{P} : \mathcal{F} \to [0, 1]$, which assigns each event in the event space a probability.

Of course, for the model to be meaningful, $\mathbb{P}$ should satisfy certain axioms.

Axioms

Definition (Probability Space)

A *probability space* is the triple $(\Omega, \mathcal{F}, \mathbb{P})$. Here, $\mathbb{P}$ satisfies the axioms

1. $\mathbb{P}(A) \geq 0$ for all $A \in \mathcal{F}$
2. $\mathbb{P}(\Omega) = 1$
3. $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ for mutually exclusive events $A_1, A_2, \ldots \in \mathcal{F}$

Exercise for the reader: Prove the axioms.

Elementary Results

From the axioms, we are able to derive the following fundamental results:

1. If $A_1, A_2, ..., A_k$ are mutually exclusive,
$$\mathbb{P}\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} \mathbb{P}(A_i).$$
2. $\mathbb{P}(\emptyset) = 0$.
3. For any $A \subseteq \Omega$, $0 \leq \mathbb{P}(A) \leq 1$ and $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$.
4. If $B \subset A$, then $\mathbb{P}(B) \leq \mathbb{P}(A)$. Hence, if B occurs $\rightarrow$ A occurs, then $\mathbb{P}(B) \leq \mathbb{P}(A)$.

These results can be used without proof.

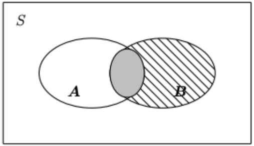
Conditional Probability

Definition (Conditional Probability)

The **conditional probability** that an event A occurs, given that an event B has occurred is

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Here, we require $\mathbb{P}(B) \neq 0$.



Given that B has occurred, the total probability for the possible results of an experiment equals $\mathbb{P}(B)$. Visually, we see that the only outcomes in A that are now possible are those in $A \cap B$.

Definition (Independent Events)

Events A and B are **independent** if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Recall that for *any* two events A and B, we have $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$. Thus, A and B are independent if and and only $\mathbb{P}(A|B) = \mathbb{P}(A)$ (or equivalently, $\mathbb{P}(B|A) = \mathbb{P}(B)$). Intuitively, this means that knowing event A has occurred does not give us any information on the probability of event B occurring (and vice versa).

Independence

Definition (Pairwise Independent Events)

For a countable sequence of events $\{A_i\}$, the events are **pairwise independent** if

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$$

for all $i \neq j$

Definition ((Mutually) Independent Events)

For a countable sequence of events $\{A_i\}$, the events are **(mutually) independent** if for any collection $A_{i_1}, A_{i_2}, ..., A_{i_n}$, we have

$$\mathbb{P}(A_{i_1} \cap ... \cap A_{i_n}) = \mathbb{P}(A_{i_1})...\mathbb{P}(A_{i_n}).$$

Independence implies pairwise independence but not vice versa. Can you think of an example of where pairwise independence does not imply independence?

The Multiplicative Law

Definition (The Multiplicative Law)

For events A_1, A_2, we have

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_2|A_1)\mathbb{P}(A_1).$$

For events A_1, A_2, A_3, we have

$$\begin{aligned}\mathbb{P}(A_1 \cap A_2 \cap A_3) &= \mathbb{P}(A_3 \cap A_2 \cap A_1) \\ &= \mathbb{P}(A_3|A_2 \cap A_1)\mathbb{P}(A_2 \cap A_1) \\ &= \mathbb{P}(A_3|A_1 \cap A_2)\mathbb{P}(A_2|A_1)\mathbb{P}(A_1).\end{aligned}$$

We trust that the reader is able to generalise this to cases involving a greater number of events.

The Multiplicative Law is highly useful when dealing with a sequence of dependent trials.

The Additive Law

Definition (The Additive Law)

For events A and B, we have

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

You might have noticed that the Additive Law resembles the Inclusion-exclusion principle. The following quote from Wikipedia sheds some light on this:

"As finite probabilities are computed as counts relative to the cardinality of the probability space, the formulas for the principle of inclusion–exclusion remain valid when the cardinalities of the sets are replaced by finite probabilities."

https://en.wikipedia.org/wiki/Inclusion-exclusion_principle

The Law of Total Probability

Definition (The Law of Total Probability)

Suppose $A_1, A_2, ..., A_k$ are **mutually exclusive** ($A_i \cap A_j = \emptyset$ for all $i \neq j$) and **exhaustive** ($\bigcup_{i=1}^{k} A_i = \Omega =$ sample space); that is, $A_1, ..., A_k$ forms a **partition** of Ω. Then, for any event B, we have

$$\mathbb{P}(B) = \sum_{i=1}^{k} \mathbb{P}(B|A_i)\mathbb{P}(A_i).$$

The Law of Total Probability relates marginal probabilities to conditional probabilities and is often used in calculations involving Bayes' Formula.

Bayes' Formula/Bayes' Theorem/Bayes' Law

Definition (Bayes' Theorem)

For a partition $A_1, A_2, ..., A_k$ and an event B,

$$\mathbb{P}(A_j|B) = \frac{\mathbb{P}(B|A_j)\mathbb{P}(A_j)}{\sum_{i=1}^{k} \mathbb{P}(B|A_i)\mathbb{P}(A_i)} = \frac{\mathbb{P}(B|A_j)\mathbb{P}(A_j)}{\mathbb{P}(B)}$$

In essence, Bayes' Theorem allows us to reverse the order of conditioning provided that we know the marginal probabilities of event A and event B. When dealing with problems involving Bayes' Theorem, it is recommended that one draws a **tree diagram**.

Here, we shall adopt the frequentist interpretation of probability where probability measures a *proportion of outcomes* – as opposed to the Bayesian interpretation where probability measures a *degree of belief*.

Table of Contents

1. Basic Probability Theory
2. Review of Random Variables
3. Computations for Common Distributions
4. Inequalities
5. Limit of Random Variables
6. Tips and Assorted Examples

Review of Random Variables

1. Discrete Random Variables
2. Continuous Random Variables
3. Cumulative Distribution Function
4. Expectation and Moments

Discrete Random Variables

Definition (Discrete Random Variable)

The random variable X is **discrete** if there are *countably many* values x for which $\mathbb{P}(X = x) > 0$.

The *probability structure* of X is typically described by its **probability (mass) function**.

Definition (Probability (Mass) Function)

The **probability function** of the discrete random variables X is given by

$$f_X(x) = \mathbb{P}(X = x).$$

The following two properties are important and apply to all discrete random variables:

1. $f_X(x) \geq 0$ for all $x \in \mathbb{R}$
2. $\sum_{\text{all } x} f_X(x) = 1$

Continuous Random Variables

A **continuous random variable** has a *continuum* of possible values.
Continuous random variables do not have a probability (mass) function, but have the analogous **(probability) density function**.

Definition ((Probability) Density Function)

The **density** function of a continuous random variable is a real-valued function f_X on $\mathbb{R}$ with the property

$$\int_A f_X(x)dx = \mathbb{P}(X \in A)$$

for any (measurable) set $A \subseteq \mathbb{R}$.

The following two properties are important and apply to all continuous random variables:

1. $f_X(x) \geq 0$ for all $x \in \mathbb{R}$.
2. $\int_{-\infty}^{\infty} f_X(x)dx = 1$.

Cumulative Distribution Function (cdf)

Definition (Cumulative Distribution Function)

The cumulative distribution function (cdf) of a random variable X is defined by

$$F_X(x) = \mathbb{P}(X \leq x).$$

Here, X can be either continuous or discrete.

In the case where X is a continuous random variable, we have the following important results:

1. $F_X = \int_{-\infty}^{\infty} f_X(t)dt.$
2. $f_X(x) = F_X'(x).$
3. $\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)dx =$ area under f_X between a and b.

Thus, if we know one of F_X or f_X, we are able to derive the other. Moreover, once we know these, we are able to derive any probability/property of X.

Important Remarks on Continuous Random Variables

Suppose X is a continuous random variable. Then we have

$$\mathbb{P}(X = a) = 0 \text{ for any } a \in \mathbb{R}.$$

Hence, it is only meaningful to talk about the probability of X lying in some subinterval(s) of R. Consequently, we have

$$\mathbb{P}(a < X < b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X \leq b).$$

Thus, we typically do not have to worry about whether an interval contains its boundary points or not.

This is NOT the case for discrete random variables.

Expectation

Definition (Expected Value of a Discrete Random Variable)

The **expected value** or **mean** of a discrete random variable X is

$$\mathbb{E}X = \mathbb{E}[X] \stackrel{\text{def}}{=} \sum_{\text{all } x} x \cdot \mathbb{P}(X = x) = \sum_{\text{all } x} x \cdot f_X(x),$$

where f_X is the probability function of X.

Definition (Expected Value of a Continuous Random Variable)

The **expected value** of **mean** of a continuous random variable X is

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) dx,$$

where f_X is the density function of X.

In either case, $\mathbb{E}[X]$ can be interpreted as the *long run average* of X.

Expectation of Transformed Random Variables

Often, we are interested in a transformation of a random variable. In particular, we often examine the *r***th moment** *of X about some constant a*, $\mathbb{E}[(X-a)^r]$. The following results provide a method of calculating the expectation of a transformation of a random variable.

Result (Transformation of a Discrete Random Variable)

Let X be a discrete random variable and let g be a function of X. Then

$$\mathbb{E}g(X) = \mathbb{E}[g(X)] = \sum_{\text{all } x} g(x) \cdot f_X(x).$$

Result (Transformation of a Continuous Random Variable)

Let X be a continuous random variable and let g be a function of X. Then

$$\mathbb{E}g(X) = \mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(X) f_X(x) dx.$$

Properties of Expectation

Result (Linearity of Expectation)

Let X, Y be random variables and a, b be constants. Then

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y].$$

Result (Expected Value of a Constant)

For a constant c,

$$\mathbb{E}[c] = c.$$

In general, for **dependent** random variables X, Y,

$$\mathbb{E}[XY] \neq \mathbb{E}[X]\mathbb{E}[Y].$$

Also, if g is a transformation of X, then typically,

$$\mathbb{E}[g(X)] \neq g(\mathbb{E}[X]).$$

Variance and Standard Deviation

Definition (Variance)

Let $\mu = \mathbb{E}[X]$. Then the **variance** of X, denoted $\text{Var}[X]$, is defined as

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

Observe that this is the second moment of X about μ.

Definition (Standard Deviation)

The **standard deviation** of X is the square root of its variance:

$$\sigma = \sqrt{\text{Var}[X]}.$$

Both variance and standard deviation are measures of the *spread* of a random variable. Standard deviations are in the same unit as X and thus, can be more readily interpreted. However, variances are easier to work with theoretically.

Properties of Variance

Result (Alternative Formula for Variance)

Let X be a random variable and let $\mu = \mathbb{E}[X]$. Then

$$\text{Var}[X] = \mathbb{E}[X^2] - \mu^2.$$

The variance will often be calculated using this formula.

Result (Nonlinearity of Variance)

Let X be a random variable and let a be a constant. Then

$$\text{Var}[X + a] = \text{Var}[X]$$
$$\text{Var}[aX] = a^2 \, \text{Var}[X].$$

Pay attention to the nonlinearity of variance. A common mistake is treating variance as being linear.

Moment Generating Functions

Definition (Moment Generating Function)

The **moment generating function** (mgf) of a random variable X is

$$\mu_X(u) = \mathbb{E}[e^{uX}].$$

We say that the moment generating function of X exists if $m_X(u)$ is finite in some interval containing zero.

The following result regarding the rth moment of X, $\mathbb{E}[X^r]$, shows why the moment generating function is called as such.

Result (rth Moment of a random variable)

Let X be a random variable. Then in general, we have

$$\mathbb{E}[X^r] = m_X^{(r)}(0)$$

for $r = 0, 1, 2, \ldots$.

Properties of Moment Generating Functions

Result (Uniqueness)

Let X and Y be two random variables all of whose moments exist. If

$$m_X(u) = m_Y(u)$$

for all u in a neighbourhood of 0 (i.e., for all $|\mu| < \epsilon$ for some $\epsilon < 0$), then

$$F_X(x) = F_Y(y) \text{ for all } x \in \mathbb{R}.$$

That is, the moment generating function of a random variable is **unique**.

Properties of Moment Generating Functions

Result (Convergence)

Let $\{X_n : n = 1, 2, ...\}$ be a sequence of random variables, each with moment generating function $m_{X_n}(u)$. Furthermore, suppose that

$$\lim_{n\to\infty} m_{X_n}(u) \text{ for all } u \text{ in a neighbourhood of } 0$$

and $m_X(u)$ is a moment generating function of a random variable X. Then

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \text{ for all } x \in \mathbb{R}.$$

That is, convergence of moment generating functions implies convergence of cumulative distribution functions.

Location and Scale Families of Densities

Result (Location Family of Densities)

A **location family** of densities based on the random variable U is the family of densities $f_X(x)$ where $X = U + c$ for all possible c. Here, $f_X(x)$ is given by:

$$f_X(x) = f_U(x - c).$$

Result (Scale Family of Densities)

A **scale family** of densities based on the random variable U is the family of densities $f_X(x)$ where $X = cU$ for all possible c. $f_X(x)$ is given by:

$$f_X(x) = c^{-1} f_U\left(\frac{x}{c}\right).$$

In essence, the density function of a continuous random variable may belong to a *family* of density functions that all have a similar form. This relates to the concept of **parameters** in statistics.

Proving the above results is not difficult and is a good exercise.

Table of Contents

1. Basic Probability Theory
2. Review of Random Variables
3. Computations for Common Distributions
4. Inequalities
5. Limit of Random Variables
6. Tips and Assorted Examples

Computations for Common Distributions

1. Common distributions summary
2. Computing the MGF of random variables
3. Computing the moments directly
4. Computing the distribution of simple transformed random variables
5. **Bivariate density computations**
6. Computing the sum of random variables using the **convolution formula**, **bivariate transformation** and **MGFs**
7. Identifying distributions from the MGF

Bernoulli Distribution

Definition (Bernoulli Distribution)

For a Bernoulli trial, define the random variable

$$X = \begin{cases} 1 & \text{if the trial results in success} \\ 0 & \text{otherwise} \end{cases}$$

Then X is said to have a **Bernoulli distribution**.

Result (Probability Function of X)

If X is a Bernoulli random variable defined according to a Bernoulli trial with success probability $0 < p < 1$ then the probability function of X is

$$f_X(x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0. \end{cases}$$

An equivalent way of writing this is $f_X(x) = p^x(1-p)^x$, $x = 0, 1$.

Binomial Distribution

Definition (Binomial Random Variable)

Consider a sequence of n independent Bernoulli trials, each with success probability p. If

$$X = \text{ total number of successes}$$

then X is a **Binomial** random variable with parameters n and p. A common shorthand is

$$X \sim \text{Bin}(n, p).$$

Here, "$\sim$" has the meaning "is distributed as" or "has distribution".

Results

If $X \sim \text{Bin}(n, p)$ then

1. $f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, ..., n$,
2. $\mathbb{E}[X] = np$,
3. $\text{Var}(X) = np(1-p)$.

Geometric Distribution

Definition (Geometric Distribution)

If

$$X = \text{ number of trials until first success},$$

then X is said to have a geometric distribution with parameter p (the probably of success on each trial).

Results

If $X \sim \text{Geom}(p)$, then

1. $f_X(x; p) = p(1-p)^{x-1}$, $x = 1, 2, \ldots$
2. $\mathbb{E}[X] = \frac{1}{p}$,
3. $\text{Var}(X) = \frac{1-p}{p^2}$.

Poisson Distribution

Definition (Poisson Distribution)

The random variable X has a **Poisson distribution** with parameter $\lambda > 0$ if its probability function is

$$f_X(x;\lambda) = \mathbb{P}(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, \ldots$$

A common abbreviation is $X \sim \text{Poisson}(\lambda)$.

The Poisson distribution is a model of the occurrence of point events in a continuum. The number of points occurring in a time interval t is a random variable with a Poisson(λt) distribution

Results

If $X \sim \text{Poisson}(\lambda)$

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

Exponential Distribution

The exponential distribution is useful for describing the probability structure of *positive* random variables. The exponential distribution is closely related to the Poisson distribution; the time until the next event has an exponential distribution with parameter $\beta = \frac{1}{\lambda}$.

Definition

A random variable X is said to have an **exponential distribution** with parameter $\beta > 0$ if X has density function:

$$f_X(x; \beta) = \frac{1}{\beta} e^{-x/\beta}, x > 0.$$

Results

1. $\mathbb{E}(X) = \beta$,
2. $\text{Var}(X) = \beta^2$,
3. **Memoryless property**: $\mathbb{P}(X > s + t | X > s) = \mathbb{P}(X > t)$.

Uniform Distribution

Definition (Uniform Distribution)

A continuous random variable X that can take values in the interval (a, b) with equal likelihood is said to have a **uniform distribution** on (a,b). A common shorthand is

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

Results

If $X \sim \text{Unif}(a, b)$, then

1. $f_X(x; a, b) = \frac{1}{b-a}$, $a < x < b$,
2. $\mathbb{E}(X) = \frac{a+b}{2}$,
3. $\text{Var}(X) = \frac{(b-a)^2}{12}$.

Gamma Function

The Gamma Function extends the factorial function to the real numbers.

Definition (Gamma Function)

The **Gamma function** at $x \in \mathbb{R}$ is given by

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt.$$

Results

1. $\Gamma(x) = (x-1)\Gamma(x-1)$,
2. $\Gamma(n) = (n-1)!,$, $n = 1, 2, 3...$
3. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$,
4. $\int_0^\infty x^m e^{-x} dx = m!$ for $m = 0, 1, 2, \ldots$.

Beta Function

Definition (Beta Function)

The **Beta function** at $x, y \in \mathbb{R}$ is given by

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt.$$

Result

For all $x, y \in \mathbb{R}$,

$$B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

Phi Function

Definition (Φ)

For all $x \in R$,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt.$$

We cannot simplify the above expression for $\Phi(x)$ as there is no closed-form anti-derivative.

Φ gives the cumulative distribution function of the standard normal distribution.

Results

1. $\lim_{x \to -\infty} \Phi(x) = 0$,
2. $\lim_{x \to \infty} \Phi(x) = 1$,
3. $\Phi(0) = \frac{1}{2}$,
4. Φ is monotonically increasing over $\mathbb{R}$.

Normal Distribution

Definition (Normal Distribution)

The random variable X is said to have a **normal distribution** with parameters μ and σ^2 (where $-\infty < \mu < \infty$ and $\sigma^2 > 0$) if X has density function

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

A common shorthand is

$$X \sim N(\mu, \sigma^2).$$

Results

If $X \sim N(\mu, \sigma^2)$,

1. $\mathbb{E}(X) = \mu$,
2. $\mathrm{Var}(X) = \sigma^2$,

Computing Normal Distribution Probabilities

Result

If $Z \sim N(0,1)$ then

$$\mathbb{P}(Z \leq x) = F_Z(x) = \Phi(x).$$

In other words, the Φ function is the cumulative distribution function of the $N(0,1)$ random variable.

Result

If $X \sim N(\mu, \sigma^2)$ then

$$Z = \frac{X - \mu}{\sigma} \sim N(0,1).$$

Gamma Distribution

Definition (Gamma Distribution)

A random variable X is said to have a **Gamma distribution** with parameters α and β (where $\alpha, \beta > 0$) if X has density function:

$$f_X(x; \alpha, \beta) = \frac{e^{-\alpha/\beta} x^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha}, x > 0$$

A common shorthand is:

$$X \sim \text{Gamma}(\alpha, \beta).$$

Results

If $X \sim \text{Gamma}(\alpha, \beta)$ then

1. $\mathbb{E}(X) = \alpha\beta$,
2. $\text{Var}(X) = \alpha\beta^2$.

Moreover, Y has an Exponential distribution iff $Y \sim \text{Gamma}(1, \beta)$.

Beta Distribution

Definition (Beta Distribution)

A random variable X is said to have a **Beta distribution** with parameters $a\alpha, \beta > 0$ if its density function is

$$f_X(x; \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}, 0 < x < 1.$$

The Beta distribution generalises the $\mathrm{Unif}(0, 1)$ distribution which can be thought of as a Beta distribution with $a = b = 1$.

Results

If X has a distribution with parameters α and β, then

1. $\mathbb{E}(X) = \frac{\alpha}{\alpha+\beta}$,
2. $\mathrm{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$.

Joint Probability Function and Density Function

Definition (Joint Probability Function)

If X and Y are discrete random variables, then the **joint probability function** of X and Y is

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

the probability that $X = x$ **and** $Y = y$.

Definition (Joint Density Function)

The **joint density function** of continuous random variables X and Y is a bivariate function $f_{X,Y}$ with the property

$$\int\int_A f_{X,Y}(x, y)dxdy = \mathbb{P}((X, Y) \in A)$$

for any (measurable) subset A of $\mathbb{R}^2$.

Joint Cumulative Distribution Function

Definition (Joint Cumulative Distribution Function)

The **joint cdf** of X and Y is

$$\begin{aligned} F_{X,Y}(x,y) &= \mathbb{P}(X \leq x, Y \leq y) \\ &= \begin{cases} \sum_{u \leq x} \sum_{v \leq y} \mathbb{P}(X = u, Y = v) & (X \text{ discrete}) \\ \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u,v)\,dudv & (X \text{ continuous}). \end{cases} \end{aligned}$$

Result

1. If X and Y are discrete radom variables, then $\sum_{\text{all } x} \sum_{\text{all } y} f_{X,Y}(x,y) = 1$.
2. If X and Y are continuous random variables then $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dxdy = 1$.

Expectation of Joint Functions

Result (Expectation of Joint Functions)

If g is any function of $g(X, Y)$,

$$\mathbb{E}[g(X, Y)] = \begin{cases} = \sum_{\text{all } x} \sum_{\text{all } y} g(x, y)\mathbb{P}(X = x, Y = y) & \text{discrete} \\ = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y)dxdy & \text{continuous} \end{cases}$$

Marginal Probability Function

Result (Marginal Probability Function)

If X and Y are discrete, then $f_X(x)$ and $f_Y(y)$ can be calculated from $f_{X,Y}(x,y)$ as follows:

$$f_X(x) = \sum_{\text{all } y} f_{X,Y}(x,y)$$
$$f_Y(y) = \sum_{\text{all } x} f_{X,Y}(x,y).$$

$f_X(x)$ is sometimes referred to as the **marginal probability function** of X.

Marginal Density Function

Result (Marginal Density Function)

If X and Y are continuous, then $f_X(x)$ and $f_Y(y)$ can be calculated from $f_{X,Y}(x,y)$ as follows:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$
$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx.$$

$f_X(x)$ is sometimes referred to as the **marginal density function** of X.

Conditional Probability Function

Definition (Conditional Probability Function)

If X and Y are discrete, the **conditional probability function** of X given $Y = y$ is

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

Similarly,

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

Result

$$\mathbb{P}(Y \in A|X = x) = \sum_{y \in A} f_{Y|X}(y|X = x).$$

Conditional Density Function

Definition (Conditional Density Function)

If X and Y are continuous, the **conditional density function** of X given $Y = y$ is

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

Similarly,

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

Shorthand: $f_{Y|X}(y|x)$.

Result

$$\mathbb{P}(a \leq Y \leq b|X = x) = \int_a^b f_{Y|X}(y|x)dy.$$

Conditional Expected Value and Variance

Result (Conditional Expected Value)

The **conditional expected value** of X given $Y = y$ is

$$\mathbb{E}(X|Y=y) = \begin{cases} \sum_{\text{all } x} x\mathbb{P}(X=x|Y=y) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx & \text{if } X \text{ is continuous} \end{cases}$$

Result (Conditional Variance)

The **conditional variance** of X given $Y = y$ is

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

where

$$\mathbb{E}(X^2|Y=y) = \begin{cases} \sum_{\text{all } x} x^2\mathbb{P}(X=x|Y=y) \\ \int_{-\infty}^{\infty} x^2 f_{X|Y}(x|y)dx. \end{cases}$$

Independent Random Variables

Definition (Independent)

Random variables X and Y are **independent** if and only if for all x, y

$$f_{X,Y}(x,y) = f_X(x)f_Y(y).$$

Result

Random variables X and Y are independent if and only if for all x, y

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

.

Result

If X and Y are independent

$$F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y).$$

Covariance

Definition (Covariance

The **covariance** of X and Y is

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)].$$

$\mathrm{Cov}(X, Y)$ measures how much X and Y vary about their means and also how much they vary together linearly.

Results

1. $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$
2. $\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mu_X \mu_Y$
3. If X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$.

Correlation

Definition (Correlation)

The **correlation** between X and Y is

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}}.$$

$\mathrm{Corr}(X, Y)$ measures the strength of the linear association between X and Y. Independent random variables are uncorrelated, but uncorrelated variables are not necessarily independent (Consider the case where $\mathbb{E}(X) = 0$ and $Y = X^2$).

Results

1. $|\mathrm{Corr}(X, Y)| \leq 1$
2. $|\mathrm{Corr}(X, Y)| = 1$ if and only if $\mathbb{P}(Y = a + bX) = 1$ for some constants a, b.

Transformations

Result

For discrete X,

$$f_Y(y) = \mathbb{P}(Y = y) = \mathbb{P}[h(X) = y] = \sum_{x:h(x)=y} \mathbb{P}(X = x).$$

Result

For continuous X, if h is *monotonic* over the set $\{x : f_X(x) > 0\}$ then

$$\begin{aligned} f_Y(y) &= f_X(x) \left| \frac{dx}{dy} \right| \\ &= f_X(h^{-1}(y)) \left| \frac{dx}{dy} \right| \end{aligned}$$

for y such that $f_X(h^{-1}(y)) > 0$.

Linear Transformation

Result

For a continuous random variable X, if $Y = aX + b$ is a linear transformation of X with $a \neq 0$, then

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

for all y such that $f_X(\frac{y-b}{a}) > 0$.

Leading into Bivariate Transformations...

If X and Y have joint density $f_{X,Y}(x, y)$ and U is a function of X and Y, we can find the density of U by calculating $F_U(u) = \mathbb{P}(U \leq u)$ and differentiating.

Result

If U and V are functions of continuous random variables X and Y, then

$$f_{U,V}(u,v) = f_{X,Y}(x,y) \cdot |J|$$

where

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$$

is a determinant called the **Jacobian** of the transformation.
The full specification of $f_{U,V}(u,v)$ requires that the range of (u,v) values corresponding to those (x,y) for which $f_{X,Y}(x,y) > 0$ is determined.

Bivariate Transformations

To find $f_U(u)$ by bivariate transformation:

1. Define some bivariation transformation (U, V).
2. Find $f_{U,V}(u, v)$.
3. We want the marginal distribution of U. So now find $\int_{-\infty}^{\infty} f_{U,V}(u, v)$.

Using a bivariate transformation to find the distribution of U is often more convenient that deriving it via the cumulative distribution function. Using the cdf requires double integration, which we can avoid when we use a bivariate transformation.

Sum of Independent Random Variables - Probability Function/Density Function Approach

Result (Discrete Convolution Formula)

Suppose that X and Y are independent random variables raking only non-negative integer values, and let $Z = X + Y$. Then

$$f_X(z) = \sum_{y=0}^{z} f_X(z-y)f_Y(y), z = 0, 1, ...$$

Result (Continuous Convolution Formula)

Suppose X and Y are independent continuous random variables with $X \sim f_X(x)$ and $Y \sim f_Y(y)$. Then $Z = X + Y$ has density

$$f_Z(z) = \int_{\text{all possible } y} f_X(z-y)f_Y(y)dy.$$

Sum of Gammmas

Result

If $X_1, X_2, ..., X_n$ are independent with $X_i \sim \text{Gamma}(\alpha_i, \beta)$, then

$$\sum_{i=1}^{n} X_i \sim \text{Gamma}(\sum_{i=1}^{n} \alpha_i, \beta).$$

Sum of Independent Random Variables - Moment Generating Function Approach

Result

Suppose that X and Y are independent random variables with moment generating functions m_X and m_Y. Then

$$m_{X+Y}(u) = m_X(u)m_Y(u).$$

This generalises to n independent random variables.

Result

If $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\sigma_Y, \sigma_Y^2)$ are independent then

$$X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2).$$

This also generalises.

Table of Contents

1. Basic Probability Theory
2. Review of Random Variables
3. Computations for Common Distributions
4. Inequalities
5. Limit of Random Variables
6. Tips and Assorted Examples

Inequalities

1. Jensen's Inequality
2. Markov's inequality
3. Chebyshev's inequality

Result (Jensen's Inequality)

If h is a *convex* function (i.e., concave up) and X is a random variable, then

$$\mathbb{E}[h(X)] \geq h(\mathbb{E}[X]).$$

There are several formulations of Jensen's inequality and the above formulation is the one most relevant for probability theory. Students studying MATH2701 next term will be delighted to re-encounter Jensen's inequality albeit in a different formulation.

Result (Markov's Inequality)

Let X be a nonnegative random variable and $a > 0$. Then

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

That is, the probability that X is *at least* a is *at most* the expectation of X divided by a.

Chebychev's Inequality

Result (Chebychev's Inequality)

If X is any random variable with $\mathbb{E}[X] = \mu$, $\mathrm{Var}(X) = \sigma^2$ then

$$\mathbb{P}(|X - \mu| > k\sigma) \leq \frac{1}{k^2}.$$

That is, the probability that X is more than k standard deviations from its mean is less than $\frac{1}{k^2}$.

The significance of Chebychev's Inequality is that we can make specific probabilistic statements about a random variable given only its mean and standard deviation – observe that we have not made any assumptions on the distribution of X.

Interestingly, Chebychev's Inequality can be easily derived as a corollary of Markov's Inequality (Hint: consider the random variable $(X - \mathbb{E}[X])^2$)

Table of Contents

1. Basic Probability Theory
2. Review of Random Variables
3. Computations for Common Distributions
4. Inequalities
5. Limit of Random Variables
6. Tips and Assorted Examples

Limit of Random Variables

1. Almost Sure Convergence
2. Convergence in Probability
3. Convergence in Distribution
4. Central Limit Theorem
5. Law of Large Numbers
6. The Statement and Application of the Delta Method

Almost Sure Convergence

Definition (Almost Sure Convergence)

The sequence of numerical random variables $X_1, X_2, \ldots$ is said to converge **almost surely** to a numerical random variable X, denoted $X_n \xrightarrow{\text{a.s.}} X$ if

$$\mathbb{P}\left(\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right) = 1.$$

Within the context of probability theory, an event is said to happen **almost surely** if it happens with probability 1. Note that it is possible for the set of exceptions to be non-empty as long as it has probability 0.

The previous definition of almost sure convergence can be difficult to work with so we will make use of an alternative definition.

Result (Alternative Definition of Almost Sure Convergence)

$$X_n \stackrel{\text{a.s.}}{\to} X$$

if and only if for every $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{k\geq n} |X_k - X| > \epsilon\right) = 0.$$

Almost Sure Convergence is the mode of convergence used in the **Strong Law of Large Numbers**.

Convergence in Probability

The main idea behind Convergence in Probability is that the probability of an "unusual" event becomes smaller as the sequence progresses.

Definition (Convergence in Probability)

The sequence of random variables $X_1, X_2, \ldots$ **converges in probability** to a random variable X if, for all $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0.$$

This is usually written as

$$X_n \xrightarrow{\mathbb{P}} X.$$

For context, an estimator is called **consistent** if it converges in probability to the quantity being estimated. Furthermore, convergence in probability is the type of convergence established by the **Weak Law of Large Numbers**.

Relationship between Almost Sure Convergence and Convergence in Probability

Result (Almost Sure Convergence implies Convergence in Probability)

$$X_n \stackrel{\text{a.s.}}{\to} X \implies X_n \stackrel{\mathbb{P}}{\to} X$$

and

$$X_n \stackrel{\text{a.s.}}{\to} 0 \iff \sup_{k \geq n} |X_k| \stackrel{\mathbb{P}}{\to} 0.$$

To understand why almost sure convergence is stronger than convergence in probability, almost sure convergence depends on a joint distribution whereas convergence in probability depends only on a marginal distribution.
Wise words of wisdom: Almost sure convergence means no noodle leaves the strip (for large enough n), convergence in probability means the proportion of noodles leaving the strip goes to 0 (as $n \to \infty$).

Convergence in Distribution

Convergence in Distribution is concerned with whether the distributions of X_i converges to the distribution of some random variable X. In other words, we increasingly expect to see the next outcome in a sequence of random experiments become better modelled by a given probability distribution.

Definition (Convergence in Distribution)

Let $X_1, X_2, \ldots$ be a sequence of random variables. We say that X_n **converges in distribution** to X if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

for all x where F_X is continuous. A common shorthand is $X_n \xrightarrow{\mathrm{d}} X$. We say that F_X is the **limiting distribution** of X_n.

Convergence in distribution often arises in applications of the **central limit theorem.**

More on Convergence in Distribution

Convergence in distribution allows us to make approximate probability statements about X_n, for large n, if we can derive the limiting distribution $F_X(n)$.

Result (Establishing Convergence in Distribution using Moments)

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables, each with mgf $M_{X_i}(t)$. Furthermore, suppose that

$$\lim_{n\to\infty} M_{X_n}(t) = M_X(t).$$

If $M_X(t)$ is a moment generating function then there is a unique F_X (which gives a random variable X) whose moments are determined by $M_X(t)$ and for all points of continuity $F_X(x)$ we have

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x).$$

Relationship between Convergence in Probability and Convergence in Distribution

Result (Convergence in Probability implies Convergence in Distribution)

$$X_n \xrightarrow{\mathbb{P}} X \implies X_n \xrightarrow{\text{d}} X.$$

Convergence in probability is concerned with the convergence of the actual values (the x_i's) whereas convergence in distribution is concerned with the convergence of the distributions (the $F_{X_i}(x)$'s).

Weak Law of Large Numbers

Result (Weak Law of Large Numbers)

Suppose $X_1, X_2, \ldots$ are independent, each with mean μ and variance $0 < \sigma^2 < \infty$. If

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \text{ then } \bar{X}_n \xrightarrow{\mathbb{P}} \mu.$$

The Weak Law of Large Numbers describes how the sample average converges to the distributional average as the sample size increases.

Slutsky's Theorem

Result (Slutsky's Theorem)

Let $X_1, X_2, \ldots$ be a sequence of random variables such that $X_n \xrightarrow{\text{d}} X$ for some distribution X and let $Y_1, Y_2, \ldots$ be another sequence of random variables such that $Y_n \xrightarrow{\mathbb{P}} c$ for some constant c. Then

1. $X_n + Y_n \xrightarrow{\text{d}} X + c$
2. $X_n Y_n \xrightarrow{\text{d}} cX$.

Slutsky's Theorem extends some properties of algebraic operations on convergent sequences of real numbers to sequences of random variables and is useful for establishing convergence in distribution results.

Strong Law of Large Numbers

The Weak Law corresponds to convergence in probability while the Strong Law corresponds to almost sure convergence.

Result (Strong Law of Large Numbers)

Let $X_1, X_2, \ldots$ be independent with common mean $\mathbb{E}[X] = \mu$ and variance $\text{Var}(X) = \sigma^2 < \infty$, then

$$\bar{X}_n \overset{\text{a.s.}}{\to} \mu.$$

Central Limit Theorem

Result (Central Limit Theorem)

Suppose $X_1, X_2, \ldots$ are independent and identically distributed random variables with common mean $\mu = \mathbb{E}(X_i)$ and common variance $\sigma^2 = \mathrm{Var}(X_i) < \infty$. For each $n \geq 1$ let $\bar{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$. Then

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{\text{d}} Z$$

where $Z \sim N(0,1)$. It is common to write

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{\text{d}} N(0,1).$$

Note that $\mathbb{E}(\bar{X}_n = \mu$ and $\mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$ so that Central Limit Theorem states that the limiting distribution of any **standardised** average of independent random variables is the standard Normal distribution.

Alternative Forms of the Central Limit Theorem

Sometimes, probabilities involving related quantities such as the sum $\sum_{i=1}^{n} X_i$ are required. Since $\sum_{i=1}^{n} X_i = n\bar{X}$, the Central Limit Theorem also applies to the sum of a sequence of random variables.

Results

1. $\sqrt{n}(\bar{X} - \mu) \xrightarrow{\text{d}} N(0, \sigma^2)$
2. $\frac{\sum_i X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{\text{d}} N(0, 1)$
3. $\frac{\sum_i X_i - n\mu}{\sqrt{n}} \xrightarrow{\text{d}} N(0, \sigma^2)$

Applications of the Central Limit Theorem

Central Limit Theorem for Binomial Distribution

Suppose $X \sim \text{Bin}(n, p)$. Then

$$\frac{X - np}{\sqrt{np(1-p)}} \xrightarrow{\text{d}} N(0, 1)$$

Normal Approximation to the Poisson Distribution

Suppose $X \sim \text{Poisson}(\lambda)$. Then

$$\lim_{\lambda \to \infty} \mathbb{P}\left(\frac{X - \lambda}{\sqrt{\lambda}} \leq x \right) = \mathbb{P}(Z \leq x)$$

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

Continuity correction: add $\frac{1}{2}$ to the numerator.

The Delta Method

The Delta Method

Let $Y_1, Y_2, \ldots$ be a sequence of random variables such that

$$\frac{\sqrt{n}(Y_n - \theta)}{\sigma} \xrightarrow{\text{d}} N(0, 1).$$

Suppose the function g is differentiable in the neighbourhood of θ and $g'(\theta) \neq 0$. Then

$$\sqrt{n}(g(Y_n) - g(\theta)) \xrightarrow{\text{d}} N(0, \sigma^2[g'(\theta)]^2).$$

Alternatively,

$$\frac{g(Y_n) - g(\theta)}{g'(\theta)/\sqrt{n}} \xrightarrow{\text{d}} N(0, 1).$$

There are several ways of stating the Delta method.

Table of Contents

1. Basic Probability Theory
2. Review of Random Variables
3. Computations for Common Distributions
4. Inequalities
5. Limit of Random Variables
6. Tips and Assorted Examples

General Tips

1. When finding a pdf/cdf, always specify the domain over which it is defined.
2. Make sure that the relevant conditions are satisfied before applying some result/theorem (e.g., are the random variables i.i.d? Does a sequence of random variables have the right mode of convergence?)
3. Does the pdf in question belongs to a known family? If so, you can typically make use of some known results.
4. $\xrightarrow{\text{a.s.}} \implies \xrightarrow{\mathbb{P}} \implies \xrightarrow{\text{d}}$
5. Know how to do calculations involving bivariate distributions
6. Know how to apply (bivariate) transformations
7. Know how to sum independent random variables using the convolution formula approach and the moment generating function approach.
8. Know the Large Numbers Laws, the CLT and the Delta method.

Example 1

Example (MATH1251) (contd.)

99% of the people with the disease receive a positive test. 98% of those without receive a negative test. If 2% of the population have the disease, determine the probability of someone having the disease *given* they received a positive test.

Example 1

We require $\mathbb{P}(D \mid T) = \dfrac{\mathbb{P}(T \mid D)\mathbb{P}(D)}{\mathbb{P}(T)}$.

$$\begin{aligned}\mathbb{P}(T) &= \mathbb{P}(T \mid D)\mathbb{P}(D) + \mathbb{P}(T \mid D^c)\mathbb{P}(D^c)\\ &= \mathbb{P}(T \mid D)\mathbb{P}(D) + \big(1 - \mathbb{P}(T^c \mid D^c)\big)\mathbb{P}(D^c)\\ &= 0.99 \times 0.02 + (1 - 0.98) \times 0.98 = 0.0394\end{aligned}$$

$$\therefore \mathbb{P}(D \mid T) = \frac{0.99 \times 0.02}{0.0394} \approx 0.5025$$

Example 1

A lot of people get stuck with Bayes' law, especially when used with other results. Use a tree diagram!

Example 2

Example

Given the distribution of X below, compute its expectation and standard deviation.

x	0	3	9	27
$\mathbb{P}(X = x)$	0.3	0.1	0.5	0.1

Example 2

Example

Given the distribution of X below, compute its expectation and standard deviation.

x	0	3	9	27
$\mathbb{P}(X = x)$	0.3	0.1	0.5	0.1

$$\begin{aligned}\mathbb{E}[X] &= \sum_{\text{all } x} x\,\mathbb{P}(X = x) \\ &= 0 \times 0.3 + 3 \times 0.1 + 9 \times 0.5 + 27 \times 0.1 \\ &= 7.5\end{aligned}$$

Example 2

Example

Given the distribution of X below, compute its expectation and standard deviation.

x	0	3	9	27
$\mathbb{P}(X = x)$	0.3	0.1	0.5	0.1

$$\mathbb{E}[X] = 7.5$$

$$\begin{aligned}\mathbb{E}[X^2] &= 0^2 \times 0.3 + 3^2 \times 0.1 + 9^2 \times 0.5 + 27^2 \times 0.1 \\ &= 114.3\end{aligned}$$

Example 2

Example

Given the distribution of X below, compute its expectation and standard deviation.

x	0	3	9	27
$\mathbb{P}(X = x)$	0.3	0.1	0.5	0.1

$$\mathbb{E}[X] = 7.5$$
$$\mathbb{E}[X^2] = 114.3$$

$$\sigma_X = \sqrt{\mathbb{E}[X^2] - (\mathbb{E}[X])^2} = \sqrt{114.3 - 7.5^2} = \sqrt{58.05} \approx 7.619$$

Example 2

Example (2901 oriented)

Let $X \sim \text{Geom}(p)$. Prove that $\mathbb{E}[X] = \frac{1}{p}$.

Example 2

Example (2901 oriented)

Let $X \sim \text{Geom}(p)$. Prove that $\mathbb{E}[X] = \frac{1}{p}$.

Recall: $\mathbb{P}(X = x) = p(1-p)^{x-1}$ for $x = 1, 2, \ldots$

$$\mathbb{E}[X] = \sum_{\text{all } x} x\,\mathbb{P}(X = x) = \sum_{x=1}^{\infty} xp(1-p)^{x-1}$$

Example 2

Example (2901 oriented)

Let $X \sim \text{Geom}(p)$. Prove that $\mathbb{E}[X] = \frac{1}{p}$.

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x=1}^{\infty} xp(1-p)^{x-1} \\
&= \sum_{y=0}^{\infty} (y+1)p(1-p)^{y} && (y = x-1) \\
&= (1-p)\left[\sum_{y=0}^{\infty} (y+1)p(1-p)^{y-1}\right]
\end{aligned}$$

Example 2

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_{x=1}^{\infty} xp(1-p)^{x-1} \\
&= \sum_{y=0}^{\infty} (y+1)p(1-p)^{y} && (y = x-1) \\
&= (1-p)\left[\sum_{y=0}^{\infty} (y+1)p(1-p)^{y-1}\right] \\
&= (1-p)\sum_{y=0}^{\infty} yp(1-p)^{y-1} + (1-p)\sum_{y=0}^{\infty} p(1-p)^{y-1}
\end{aligned}
$$

Example 2

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_{x=1}^{\infty} xp(1-p)^{x-1} \\
&= (1-p)\sum_{y=0}^{\infty} yp(1-p)^{y-1} + (1-p)\sum_{y=0}^{\infty} p(1-p)^{y-1} \\
&= (1-p)\sum_{y=1}^{\infty} yp(1-p)^{y-1} + (1-p)\sum_{y=1}^{\infty} p(1-p)^{y-1} \\
&\quad + p(1-p)^{-1} \qquad\qquad \text{(evaluating at } y = 0) \\
&= (1-p)\mathbb{E}[X] + (1-p)\left(1 + p(1-p)^{-1}\right)
\end{aligned}
$$

Example 3

Example (2901 oriented)

Let $X \sim \text{Geom}(p)$. Prove that $\mathbb{E}[X] = \frac{1}{p}$.

$$\therefore p\mathbb{E}[X] = \Big((1-p) + p\Big)$$

$$\mathbb{E}[X] = \frac{1}{p}$$

Example 3

In general, can be done with the aid of Taylor series or binomial theorem. But preferably just do this:

Method (Deriving Expected Value from definition) (2901)

Keep rearranging the expression until you make the entire density, or $\mathbb{E}[X]$, appear again.

- Discrete case - Use a change of summation index at some point
- Continuous case - Use integration by parts (or occasionally integration by substitution)

Example

Let $f_X(x) = \frac{2}{\theta^2}x$ for $0 < x < \theta$. Compute the MGF and (2901) assert its existence.

Example 4

Example

Let $f_X(x) = \frac{2}{\theta^2}x$ for $0 < x < \theta$. Compute the MGF and (2901) assert its existence.

Integrate by parts

$$
\begin{aligned}
m_X(u) = \mathbb{E}[e^{uX}] &= \frac{2}{\theta^2}\int_0^\theta xe^{ux}\,dx \\
&= \frac{2}{\theta^2}\left(\frac{xe^{ux}}{u}\bigg|_0^\theta - \int_0^\theta \frac{e^{ux}}{u}\,dx \right)
\end{aligned}
$$

Example 4

Example

Let $f_X(x) = \frac{2}{\theta^2}x$ for $0 < x < \theta$. Compute the MGF and (2901) assert its existence.

Slowly tidy everything up

$$
\begin{aligned}
m_X(u) = \mathbb{E}[e^{uX}] &= \frac{2}{\theta^2}\int_0^\theta xe^{ux}\,dx \\
&= \frac{2}{\theta^2}\left(\left.\frac{xe^{ux}}{u}\right|_0^\theta - \int_0^\theta \frac{e^{ux}}{u}\,dx \right) \\
&= \frac{2\theta e^{u\theta}}{u\theta^2} - \frac{2}{\theta^2}\left(\left.\frac{e^{ux}}{u^2}\right|_0^\theta \right)
\end{aligned}
$$

Example 4

Example

Let $f_X(x) = \frac{2}{\theta^2}x$ for $0 < x < \theta$. Compute the MGF and (2901) assert its existence.

Slowly tidy everything up

$$\begin{aligned} m_X(u) = \mathbb{E}[e^{uX}] &= \frac{2}{\theta^2}\int_0^\theta x e^{ux}\, dx \\ &= \frac{2}{\theta^2}\left(\frac{xe^{ux}}{u}\bigg|_0^\theta - \int_0^\theta \frac{e^{ux}}{u}\, dx \right) \\ &= \frac{2\theta e^{u\theta}}{u\theta^2} - \frac{2}{\theta^2}\left(\frac{e^{ux}}{u^2}\bigg|_0^\theta \right) \\ &= \frac{2(u\theta e^{u\theta} - e^{u\theta} + 1)}{u^2\theta^2} \end{aligned}$$

Example 4

Example

Let $f_X(x) = \frac{2}{\theta^2}x$ for $0 < x < \theta$. Compute the MGF and (2901) assert its existence.

$$m_X(u) = \frac{2(u\theta e^{u\theta} - e^{u\theta} + 1)}{u^2\theta^2}$$

GeoGebra simulation

Example 4

Example

Let $f_X(x) = \frac{2}{\theta^2}x$ for $0 < x < \theta$. Compute the MGF and (2901) assert its existence.

Idea: Can check that the limit as $u \to 0$ is finite. The finiteness of the limit implies the required result.

$$\begin{aligned}
\lim_{u\to 0} \frac{2(u\theta e^{u\theta} - e^{u\theta} + 1)}{u^2\theta^2} &\overset{LH}{=} \lim_{u\to 0} \frac{2\left(\theta e^{u\theta} + u\theta^2 e^{u\theta} - \theta e^{u\theta}\right)}{2u\theta^2} \\
&= \lim_{u\to 0} e^{u\theta} \\
&= 1
\end{aligned}$$

Example 5

Example

Use the MGF of $X \sim \text{Bin}(n, p)$ to prove that $\mathbb{E}[X] = np$.

$$\mathbb{E}[X] = \lim_{u \to 0} \frac{d}{du}(1 - p + pe^u)^n$$

Example 5

Example

Use the MGF of $X \sim \text{Bin}(n, p)$ to prove that $\mathbb{E}[X] = np$.

$$
\begin{aligned}
\mathbb{E}[X] &= \lim_{u \to 0} \frac{d}{du}(1 - p + pe^u)^n \\
&= \lim_{u \to 0} n(1 - p + pe^u)^{n-1} \cdot pe^u \\
&= n(1 - p + p)^{n-1} \cdot p \\
&= np
\end{aligned}
$$

Example 6

Example

A busy switchboard receives 150 calls an hour on average. Assume that calls are independent from each other and can be modelled with a Poisson distribution. Find the probability of

1. Exactly 3 calls in a given *minute*
2. At least 10 calls in a given *5 minute period*.

Naive:

$$X \sim \text{Poisson}(150).$$

Example 6

Example

A busy switchboard receives 150 calls an hour on average. Assume that calls are independent from each other and can be modelled with a Poisson distribution. Find the probability of

1. Exactly 3 calls in a given *minute*
2. At least 10 calls in a given *5 minute period*.

In Q1, take $X \sim \text{Poisson}(150/60) = \text{Poisson}(2.5)$. Then,

$$\mathbb{P}(X = 3) = e^{-2.5}\frac{2.5^3}{3!} \approx 0.2138$$

Example 6

Example

A busy switchboard receives 150 calls an hour on average. Assume that calls are independent from each other and can be modelled with a Poisson distribution. Find the probability of

1. Exactly 3 calls in a given *minute*
2. At least 10 calls in a given *5 minute period*.

In Q2, take $Y \sim \text{Poisson}(2.5 \times 5) = \text{Poisson}(12.5)$. Then,

$$\begin{aligned}\mathbb{P}(Y \geq 10) &= 1 - \mathbb{P}(Y \leq 9)\\ &= 1 - e^{-12.5}\left(\frac{12.5^0}{0!} + \cdots + \frac{12.5^9}{9!}\right)\end{aligned}$$

Example 6

Example

A busy switchboard receives 150 calls an hour on average. Assume that calls are independent from each other and can be modelled with a Poisson distribution. Find the probability of

1. Exactly 3 calls in a given *minute*
2. At least 10 calls in a given *5 minute period*.

In Q2, take $Y \sim \text{Poisson}(2.5 \times 5) = \text{Poisson}(12.5)$. Then,

$$\begin{aligned}\mathbb{P}(Y \geq 10) &= 1 - \mathbb{P}(Y \leq 9) \\ &= 1 \texttt{ - ppois(9,lambda=12.5,lower=TRUE)} \\ &\approx 0.7985689\end{aligned}$$

Example 7

Example (2901 course pack)

If, on average, 5 servers go offline during the day, what is the chance that no servers will go offline in the next hour?

Example 7

Example (2901 course pack)

If, on average, 5 servers go offline during the day, what is the chance that no servers will go offline in the next hour?

The number of servers going offline in a day is $X \sim \text{Poisson}(5)$.

So the time taken for the next server to go offline is $T \sim \text{Exp}(0.2)$, measured in **days**.

Example 7

Example (2901 course pack)

If, on average, 5 servers go offline during the day, what is the chance that no servers will go offline in the next hour?

The number of servers going offline in a day is $X \sim \text{Poisson}(5)$.

So the time taken for the next server to go offline is $T \sim \text{Exp}(0.2)$, measured in **days**.

$$\therefore \text{We require } \mathbb{P}\left(T > \frac{1}{24}\right)$$

Example 7

Example (2901 course pack)

If, on average, 5 servers go offline during the day, what is the chance that no servers will go offline in the next hour?

The number of servers going offline in a day is $X \sim \text{Poisson}(5)$.

So the time taken for the next server to go offline is $T \sim \text{Exp}(0.2)$, measured in **days**.

$$\mathbb{P}\left(T > \frac{1}{24}\right) = \int_{1/24}^{\infty} 5e^{-5t}\,dt$$
$$= e^{-5/24}$$

Example 8

Formula (Transforming a Discrete r.v.)

$$\mathbb{P}\big(h(X) = y\big) = \sum_{x:h(x)=y} \mathbb{P}(X = x)$$

Um, ye wat?

Example 8

Example

A random variable has the following distribution:

x	-1	0	1	2
$\mathbb{P}(X = x)$	0.38	0.21	0.14	0.27

Determine the distribution of $Y = X^3$ and $Z = X^2$.

Example 8

Example

A random variable has the following distribution:

x	-1	0	1	2
$\mathbb{P}(X = x)$	0.38	0.21	0.14	0.27

Determine the distribution of $Y = X^3$ and $Z = X^2$.

If X can take the values $-1, 0, 1, 2$,
then $Y = X^3$ takes the values $-1, 0, 1, 8$.

$$\mathbb{P}(Y = -1) = \mathbb{P}(X^3 = -1) = \mathbb{P}(X = -1) = 0.38$$

Similarly, $\mathbb{P}(Y = 0) = 0.21$, $\mathbb{P}(Y = 1) = 0.14$, $\mathbb{P}(Y = 8) = 0.27$.

Example 8

Example

A random variable has the following distribution:

x	-1	0	1	2
$\mathbb{P}(X = x)$	0.38	0.21	0.14	0.27

Determine the distribution of $Y = X^3$ and $Z = X^2$.

On the other hand, X^2 can only take the values of $0, 1, 4$.

$$\mathbb{P}(Z = 0) = \mathbb{P}(X^2 = 0) = \mathbb{P}(X = 0) = 0.21$$

...and $\mathbb{P}(Z = 4)$ is still equal to 0.27.

Example 8

Example

A random variable has the following distribution:

x	-1	0	1	2
$\mathbb{P}(X = x)$	0.38	0.21	0.14	0.27

Determine the distribution of $Y = X^3$ and $Z = X^2$.

On the other hand, X^2 can only take the values of $0, 1, 4$.

$$\mathbb{P}(Z = 0) = \mathbb{P}(X^2 = 0) = \mathbb{P}(X = 0) = 0.21$$

$$\mathbb{P}(Z = 1) = \mathbb{P}(X^2 = 1) = \mathbb{P}(X = \pm 1) = 0.38 + 0.14 = 0.62$$

...and $\mathbb{P}(Z = 4)$ is still equal to 0.27.

Example 8

Just to think about... (2901 oriented)

If $X \sim \text{Poisson}(\lambda)$, what must be the distribution of $Y = X^2$

$$\mathbb{P}(Y = y) = \begin{cases} e^{-\lambda} \frac{\lambda^{\sqrt{y}}}{(\sqrt{y})!} & \text{if } y = 0, 1, 4, 9, \ldots \\ 0 & \text{otherwise} \end{cases}$$

Example 9

Method 1 (Continuous random variable transform theorem)

Consider the transform $y = h(x)$. If h is monotonic wherever $f_X(x)$ is non-zero, then the density of $Y = h(X)$ is

$$f_Y(y) = f_X\left(h^{-1}(y)\right)\left|\frac{dx}{dy}\right|$$

Example

Let $X \sim \text{Exp}(\lambda)$. What is the density of $Y = X^2$?

Example 9

Example

Let $X \sim \text{Exp}(\lambda)$. What is the density of $Y = X^2$?

- $f_X(x) = \frac{1}{\lambda} e^{-x/\lambda}$ for all $x > 0$.
- $h(x) = x^2$ is invertible for all $x > 0$, with $h^{-1}(y) = \sqrt{y}$.
- $x = \sqrt{y}$, so $\frac{dx}{dy} = \frac{1}{2\sqrt{y}}$

$$\therefore f_Y(y) = f_X(\sqrt{y}) \left| \frac{1}{2\sqrt{y}} \right|$$

Example 9

Example

Let $X \sim \text{Exp}(\lambda)$. What is the density of $Y = X^2$?

- $f_X(x) = \frac{1}{\lambda} e^{-x/\lambda}$ for all $x > 0$.
- $h(x) = x^2$ is invertible for all $x > 0$, with $h^{-1}(y) = \sqrt{y}$.
- $x = \sqrt{y}$, so $\frac{dx}{dy} = \frac{1}{2\sqrt{y}}$

$$\begin{aligned} \therefore f_Y(y) &= f_X(\sqrt{y}) \left| \frac{1}{2\sqrt{y}} \right| \\ &= \frac{1}{\lambda} e^{-\sqrt{y}/\lambda} \left| \frac{1}{2\sqrt{y}} \right| \\ &= \frac{1}{2\lambda\sqrt{y}} e^{-\sqrt{y}/\lambda} \end{aligned}$$

Example 10

Example

Let $X \sim \text{Exp}(\lambda)$. What is the density of $Y = X^2$?

$$f_Y(y) = \frac{1}{2\lambda\sqrt{y}} e^{-\sqrt{y}/\lambda}$$

Since $x > 0$ and $y = x^2$, $y > 0$ as well.

Example 10

Example

Let $X \sim \text{Unif}(-10, 10)$. What is the density of $Y = X^2$?

$$f_Y(y) = \frac{1}{20\sqrt{y}}$$

Since $-10 < x < 10$ and $y = x^2$, we must have $0 < y < 100$.

Example 11

Example

The joint probability distribution of X and Y is

		y		
		0	1	2
	0	1/16	1/8	1/8
x	1	1/8	1/16	0
	2	3/16	1/4	1/16

Determine $\mathbb{P}(X = 0, Y = 1)$, $\mathbb{P}(X \geq 1, Y < 1)$ and $\mathbb{P}(X - Y = 1)$

$$\mathbb{P}(X = 0, Y = 1) = \frac{1}{8}$$

Example 11

Example

The joint probability distribution of X and Y is

		y		
		0	1	2
x	0	1/16	1/8	1/8
	1	1/8	1/16	0
	2	3/16	1/4	1/16

Determine $\mathbb{P}(X = 0, Y = 1)$, $\mathbb{P}(X \geq 1, Y < 1)$ and $\mathbb{P}(X - Y = 1)$

$$\mathbb{P}(X \geq 1, Y < 1) = \mathbb{P}(X = 1, Y = 0) + \mathbb{P}(X = 2, Y = 0)$$
$$= \frac{1}{8} + \frac{3}{16} = \frac{5}{16}$$

Example 11

Example

The joint probability distribution of X and Y is

		y		
		0	1	2
x	0	1/16	1/8	1/8
	1	1/8	1/16	0
	2	3/16	1/4	1/16

Determine $\mathbb{P}(X = 0, Y = 1)$, $\mathbb{P}(X \geq 1, Y < 1)$ and $\mathbb{P}(X - Y = 1)$

$$\begin{aligned}\mathbb{P}(X - Y = 1) &= \mathbb{P}(X = 2, Y = 1) + \mathbb{P}(X = 1, Y = 0) \\ &= \frac{1}{4} + \frac{1}{8} = \frac{3}{8}\end{aligned}$$

Example 12

Joint continuous distributions

Unless you know how to use indicator functions really well (2901), sketch the region!

Example

$$f_{X,Y}(x,y) = \frac{1}{x^2y^2} \qquad x \geq 1, y \geq 1$$

is the joint density of the continuous r.v.s X and Y. Find $\mathbb{P}(X < 2, Y \geq 4)$ and $\mathbb{P}(X \leq Y^2)$.

Example 12

Example

$$f_{X,Y}(x,y) = \frac{1}{x^2y^2} \qquad x \geq 1, y \geq 1$$

is the joint density of the continuous r.v.s X and Y. Find $\mathbb{P}(X < 2, Y \geq 4)$ and $\mathbb{P}(X \leq Y^2)$.

$$\begin{aligned}\mathbb{P}(X < 2, Y \geq 4) &= \int_1^2 \int_4^\infty \frac{1}{x^2y^2}\,dy\,dx \\ &= \int_1^2 \frac{1}{4x^2}\,dx \\ &= \frac{1}{8}\end{aligned}$$

Example 12

Example

$$f_{X,Y}(x,y) = \frac{1}{x^2y^2} \qquad x \geq 1, y \geq 1$$

is the joint density of the continuous r.v.s X and Y. Find $\mathbb{P}(X < 2, Y \geq 4)$ and $\mathbb{P}(X \leq Y^2)$.

$$\begin{aligned}\mathbb{P}(X \leq Y^2) &= \int_1^\infty \int_1^{x^2} \frac{1}{x^2y^2}\,dy\,dx \\ &= \int_1^\infty \left(\frac{1}{x^2} - \frac{1}{x^4}\right)dx \\ &= \frac{2}{3}\end{aligned}$$

Example 13

Example

Find $\mathbb{E}[Y^2 \ln X]$ for the following distribution

		y	
		1	2
x	1	1/10	1/5
	2	3/10	2/5

$$\mathbb{E}[Y^2 \ln X] = 1^2 \ln 1 \mathbb{P}(X = 1, Y = 1) + 2^2 \ln 1 \mathbb{P}(X = 1, Y = 2)$$

Example 13

Example

Find $\mathbb{E}[Y^2 \ln X]$ for the following distribution

		y	
		1	2
x	1	1/10	1/5
	2	3/10	2/5

$$\begin{aligned}\mathbb{E}[Y^2 \ln X] = 1^2 \ln 1 \mathbb{P}(X = 1, Y = 1) + 2^2 \ln 1 \mathbb{P}(X = 1, Y = 2) \\ + 1^2 \ln 2 \mathbb{P}(X = 2, Y = 1) + 2^2 \ln 2 \mathbb{P}(X = 2, Y = 2)\end{aligned}$$

Example 13

Example

Find $\mathbb{E}[Y^2 \ln X]$ for the following distribution

		y	
		1	2
x	1	1/10	1/5
	2	3/10	2/5

$$\begin{aligned}
\mathbb{E}[Y^2 \ln X] &= 1^2 \ln 1 \mathbb{P}(X = 1, Y = 1) + 2^2 \ln 1 \mathbb{P}(X = 1, Y = 2) \\
&\quad + 1^2 \ln 2 \mathbb{P}(X = 2, Y = 1) + 2^2 \ln 2 \mathbb{P}(X = 2, Y = 2) \\
&= \left(\frac{3}{10} + 2 \times \frac{2}{5} \right) \ln 2 = \frac{11 \ln 2}{10}
\end{aligned}$$

Example 14

Problem

Examine the existence of $\mathbb{E}[XY]$ for the earlier example:

$$f_{X,Y}(x,y) = \frac{1}{x^2y^2} \text{ for } x, y \geq 1.$$

Example 15

Definition (Cumulative Distribution Function)

The CDF $F_{X,Y}(x, y)$ is the function given by

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

Finding a CDF (Continuous case)

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v)\, du\, dv$$

Example

For the earlier example, $F_{X,Y}(x, y) = 0$ if $x < 1$ or $y < 1$. Else:

$$F_{X,Y}(x, y) = \int_{1}^{x} \int_{1}^{y} \frac{1}{u^2 v^2}\, du\, dv = \left(1 - \frac{1}{x}\right)\left(1 - \frac{1}{y}\right)$$

Example 16

Recall that $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Definition (Independence of random variables)

Two random variables are independent when:

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y) \qquad \text{(discrete case)}$$
$$f_{X,Y}(x, y) = f_X(x)f_Y(y) \qquad \text{(continuous case)}$$

Example

Test if X and Y are independent, for

$$f_{X,Y}(x, y) = \frac{1}{x^2y^2} \qquad x, y \geq 1.$$

Example 16

Example

Test if X and Y are independent, for

$$f_{X,Y}(x,y) = \frac{1}{x^2y^2} \qquad x, y \geq 1.$$

$$\begin{aligned} f_X(x) &= \int_1^\infty \frac{1}{x^2y^2}\, dy \\ &= \frac{1}{x^2} \qquad x \geq 1 \end{aligned}$$

$$\text{Similarly } f_Y(y) = \frac{1}{y^2} \qquad y \geq 1.$$

Therefore since $f_{X,Y}(x,y) = f_X(x)f_Y(y)$, X and Y are independent.

Example 17

Example

Determine $\mathbb{P}(X = x \mid Y = 2)$, i.e. $f_{X|Y}(x \mid 2)$, for

		y	
		1	2
x	1	1/10	1/5
	2	3/10	2/5

$$\begin{aligned}\mathbb{P}(Y = 2) &= \mathbb{P}(X = 1, Y = 2) + \mathbb{P}(X = 2, Y = 2)\\ &= \frac{1}{5} + \frac{2}{5}\\ &= \frac{3}{5}.\end{aligned}$$

Example 17

Example

Determine $\mathbb{P}(X = x \mid Y = 2)$, i.e. $f_{X|Y}(x \mid 2)$, for

		y	
		1	2
x	1	1/10	1/5
	2	3/10	2/5

$$\mathbb{P}(Y = 2) = \frac{3}{5}$$

$$\mathbb{P}(X = 1 \mid Y = 2) = \frac{\mathbb{P}(X = 1, Y = 2)}{\mathbb{P}(Y = 2)} = \frac{1}{3}$$

$$\mathbb{P}(X = 2 \mid Y = 2) = \frac{\mathbb{P}(X = 2, Y = 2)}{\mathbb{P}(Y = 2)} = \frac{2}{3}$$

Example 18

Lemma (Independence of random variables)

Two random variables are independent if and only if

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

or

$$f_{X|Y}(x \mid y) = f_X(x)$$

Investigation

For the earlier example with $f_{X,Y}(x, y) = x^{-2}y^{-2}$ for $x \geq 1$, $y \geq 1$, prove the independence of X and Y using this lemma instead.

Example 19

Definition (Conditional Expectation)

$$\mathbb{E}[X \mid Y = y] = \begin{cases} \displaystyle\sum_{\text{all } x} x\mathbb{P}(X = x \mid Y = y) & \text{discrete case} \\ \displaystyle\int_{-\infty}^{\infty} x\, f_{X|Y}(x \mid y)\, dx & \text{continuous case} \end{cases}$$

Definition (Conditional Variance)

$$\mathrm{Var}(X \mid Y = y) = \mathbb{E}[X^2 \mid Y = y] - \left(\mathbb{E}[X \mid Y = y]\right)^2$$

(And similarly for Y. Basically, just add the condition to the original formula.)

Example 19

Example

Find $\mathbb{E}[X \mid Y = 2]$ and $\text{Var}(X \mid Y = 2)$ for

		y	
		1	2
x	1	1/10	1/5
	2	3/10	2/5

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

Example 19

Example

Find $\mathbb{E}[X \mid Y = 2]$ and $\mathrm{Var}(X \mid Y = 2)$ for

		y	
		1	2
x	1	1/10	1/5
	2	3/10	2/5

$$\begin{aligned}\mathbb{E}[X^2 \mid Y = 2] &= 1^2 \cdot \mathbb{P}(X = 1 \mid Y = 2) + 2^2 \cdot \mathbb{P}(X = 2 \mid Y = 2)\\ &= 1^2 \times \frac{1}{3} + 2^2 \times \frac{2}{3}\\ &= 3.\end{aligned}$$

Example 19

Example

Find $\mathbb{E}[X \mid Y = 2]$ and $\mathrm{Var}(X \mid Y = 2)$ for

		y	
		1	2
x	1	1/10	1/5
	2	3/10	2/5

$$\mathrm{Var}(X^2 \mid Y = 2) = 3 - \left(\frac{5}{3}\right)^2 = \frac{2}{9}$$

Example 20

Example

Let $f_{X,Y}(x,y) = xy$ for $x \in [0,1]$, $y \in [0,2]$. Determine their covariance in the old fashioned way.

Step 1: Determine the marginal densities

$$f_X(x) = \int_0^2 xy\, dy = 2x \qquad (0 \leq x \leq 1)$$

$$f_Y(y) = \int_0^1 xy\, dx = \frac{y}{2} \qquad (0 \leq y \leq 2)$$

Example 20

Example

Let $f_{X,Y}(x, y) = xy$ for $x \in [0, 1]$, $y \in [0, 2]$. Determine their covariance in the old fashioned way.

Step 2: Find the marginal expectations $\mathbb{E}[X]$ and $\mathbb{E}[Y]$

$$\mathbb{E}[X] = \int_0^1 2x^2 \, dx = \frac{2}{3}$$

$$\mathbb{E}[Y] = \int_0^2 \frac{y^2}{2} \, dy = \frac{4}{3}$$

Example 20

Example

Let $f_{X,Y}(x, y) = xy$ for $x \in [0, 1]$, $y \in [0, 2]$. Determine their covariance in the old fashioned way.

Step 3: Find $\mathbb{E}[XY]$

$$\mathbb{E}[XY] = \int_0^1 \int_0^2 xy \, dy \, dx = \cdots = \frac{8}{9}$$

Step 4: Plug in:

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \frac{8}{9} - \frac{2}{3} \times \frac{4}{3} = 0.$$

Example 20

Example

Let $f_{X,Y}(x,y) = xy$ for $x \in [0,1]$, $y \in [0,2]$. Determine their covariance in the old fashioned way.

That was a horrible idea.

- Can prove that X and Y are independent
- Can use the Fubini-Tonelli theorem to just check that $\mathbb{E}[XY]$ equals $\mathbb{E}[X]\mathbb{E}[Y]$

Example 21

Example (2901)

Let $Z \sim \mathcal{N}(0,1)$ and W satisfy $\mathbb{P}(X=1)=\mathbb{P}(X=-1)=\frac{1}{2}$. Suppose that W and Z are independent and define $X := WZ$.

Show that $\mathrm{Cov}(X,Z)=0$.

Noting that $\mathbb{E}[Z]=0$,

$$\mathrm{Cov}(X,Z)=\mathbb{E}[XZ]-\mathbb{E}[X]\mathbb{E}[Z]=\mathbb{E}[XZ]$$

Example 21

Example (2901)

Let $Z \sim \mathcal{N}(0,1)$ and W satisfy $\mathbb{P}(X=1) = \mathbb{P}(X=-1) = \frac{1}{2}$. Suppose that W and Z are independent and define $X := WZ$.

Show that $\text{Cov}(X, Z) = 0$.

Noting that $\mathbb{E}[Z] = 0$,

$$\text{Cov}(X, Z) = \mathbb{E}[XZ] - \mathbb{E}[X]\mathbb{E}[Z] = \mathbb{E}[XZ]$$

Subbing in $X = WZ$ and using independence gives

$$\text{Cov}(X, Z) = \mathbb{E}[WZ^2] = \mathbb{E}[W]\mathbb{E}[Z^2]$$

Example 21

Example (2901)

Let $Z \sim \mathcal{N}(0,1)$ and W satisfy $\mathbb{P}(X=1) = \mathbb{P}(X=-1) = \frac{1}{2}$. Suppose that W and Z are independent and define $X := WZ$.

Show that $\text{Cov}(X, Z) = 0$.

Observe that

$$\mathbb{E}[W] = 1\mathbb{P}(X=1) - 1\mathbb{P}(X=-1) = 0.$$

Hence $\text{Cov}(X, Z) = \mathbb{E}[W]\mathbb{E}[Z^2] = 0$.

Example 22

Theorem (Bivariate Transform Formula)

Suppose X and Y have joint density function $f_{X,Y}$ and let U and V be transforms on these random variables. Then the joint density of U, V is

$$f_{U,V}(u,v) = f_{X,Y}(x,y)\,|\det(J)|$$

where J is the Jacobian matrix

$$J = \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix}$$

Remember: x above y and u left of v

Example (Course pack)

Let X and Y be i.i.d. Exp(4) r.v.s. Find the joint density of U and V if

$$U = \frac{1}{2}(X - Y) \text{ and } V = Y.$$

Example 22

Example (Course pack)

Let X and Y be i.i.d. Exp(4) r.v.s. Find the joint density of U and V if

$$U = \frac{1}{2}(X - Y) \text{ and } V = Y.$$

We have $y = v$ and

$$u = \frac{1}{2}(x - v) \implies x = 2u + v.$$

$$\therefore J = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix} \text{ and } \det(J) = 2.$$

Example 22

Example (Course pack)

Let X and Y be i.i.d. Exp(4) r.v.s. Find the joint density of U and V if

$$U = \frac{1}{2}(X - Y) \text{ and } V = Y.$$

$$f_{X,Y}(x, y) = \frac{1}{16} e^{-(x+y)/4}$$

Since $y = v$ and $x = 2u + v$, we get $x + y = 2u + 2v$. Therefore

$$f_{U,V}(u, v) = \frac{1}{8} e^{-(u+v)/2}.$$

Example 22

Example (Course pack)

Let X and Y be i.i.d. Exp(4) r.v.s. Find the joint density of U and V if

$$U = \frac{1}{2}(X - Y) \text{ and } V = Y.$$

We know that $y > 0$. Since $v = y$, it immediately follows that $v > 0$.

Example 22

Example (Course pack)

Let X and Y be i.i.d. Exp(4) r.v.s. Find the joint density of U and V if

$$U = \frac{1}{2}(X - Y) \text{ and } V = Y.$$

We know that $y > 0$. Since $v = y$, it immediately follows that $v > 0$. However, $x > 0$ and $x = 2u + v$. Therefore:

$$2u + v > 0$$
$$u > -\frac{v}{2}$$

Example 23

Example

Let X and Y be i.i.d. Geom(p). Use convolutions to find the probability function of $Z := X + Y$.

The probability functions are $\mathbb{P}(X = x) = p(1-p)^x$ for $x = 1, 2, 3, \ldots$, and $\mathbb{P}(Y = y) = p(1-p)^y$ for $y = 1, 2, 3, \ldots$. Therefore:

$$\mathbb{P}(X = z - y) = p(1-p)^{z-y}$$

for $z - y = 1, 2, 3, \ldots$,

Example 23

Example

Let X and Y be i.i.d. Geom(p). Use convolutions to find the probability function of $Z := X + Y$.

The probability functions are $\mathbb{P}(X = x) = p(1-p)^x$ for $x = 1, 2, 3, \ldots$, and $\mathbb{P}(Y = y) = p(1-p)^y$ for $y = 1, 2, 3, \ldots$. Therefore:

$$\mathbb{P}(X = z - y) = p(1-p)^{z-y}$$

for $z - y = 1, 2, 3, \ldots,$ i.e.

$$\boxed{y - z = \ldots, -3, -2, -1} \iff \boxed{y = \ldots, z-3, z-2, z-1}$$

Example

Let X and Y be i.i.d. Geom(p). Use convolutions to find the probability function of $Z := X + Y$.

Hence $\mathbb{P}(X = z - y)\mathbb{P}(Y = y) = p(1-p)^{z-y}p(1-p)^y = p^2(1-p)^z$, when

$$y = 0, 1, 2, \ldots$$
$$\textbf{and } y = \ldots, z-3, z-2, z-1.$$

Therefore, $y = 0, 1, 2, \ldots, z-3, z-2, z-1$.

Example 23

Example

Let X and Y be i.i.d. Geom(p). Use convolutions to find the probability function of $Z := X + Y$.

$$\begin{aligned} \therefore \mathbb{P}(Z = z) &= \sum_{y=0}^{z-1} p^2(1-p)^z \\ &= zp^2(1-p)^z \qquad \text{(sum only depends on } y\text{!)} \end{aligned}$$

Example 23

Example

Let X and Y be i.i.d. Geom(p). Use convolutions to find the probability function of $Z := X + Y$.

$$\therefore \mathbb{P}(Z = z) = \sum_{y=0}^{z-1} p^2(1-p)^z$$

$$= zp^2(1-p)^z \qquad \text{(sum only depends on } y\text{!)}$$

Since $x = 1, 2, \ldots$ and $y = 1, 2, \ldots$, i.e. x and y are natural numbers greater than or equal to 1, $z = x + y = 2, 3, 4, \ldots$

Example 24

Example

Let X and Y be i.i.d. $\mathrm{Exp}(1)$. Prove that $Z := X + Y$ follows a $\mathrm{Gamma}(2, 1)$ distribution using a convolution.

The densities are $f_X(x) = e^{-x}$ for $x > 0$, and $f_Y(y) = e^{-y}$ for $y > 0$. Therefore:

$$f_X(z - y) = e^{-z+y}, \text{ for } \boxed{z - y > 0}, \text{ i.e. } \boxed{y < z}$$

Example 24

Example

Let X and Y be i.i.d. $\text{Exp}(1)$. Prove that $Z := X + Y$ follows a $\text{Gamma}(2, 1)$ distribution using a convolution.

The densities are $f_X(x) = e^{-x}$ for $x > 0$, and $f_Y(y) = e^{-y}$ for $y > 0$. Therefore:

$$f_X(z - y) = e^{-z+y}, \text{ for } \boxed{z - y > 0}, \text{ i.e. } \boxed{y < z}$$

Hence $f_X(z - y)f_Y(y) = e^{-z}$ when $y < z$ **and** $y > 0$. i.e.

$$f_X(z - y)f_Y(y) = e^{-z} \text{ for } 0 < y < z$$

Example 24

Example

Let X and Y be i.i.d. Exp(1). Prove that $Z := X + Y$ follows a Gamma$(2, 1)$ distribution using a convolution.

$$\begin{aligned}\therefore f_Z(z) &= \int_0^z e^{-z}\, dy \\ &= e^{-z} z\end{aligned}$$

Example 24

Example

Let X and Y be i.i.d. Exp(1). Prove that $Z := X + Y$ follows a Gamma$(2, 1)$ distribution using a convolution.

$$\begin{aligned} \therefore f_Z(z) &= \int_0^z e^{-z}\,dy \\ &= e^{-z}z \\ &= \frac{e^{-z/1}z^{2-1}}{\Gamma(2)1^2} \end{aligned}$$

Since $x > 0$ and $y > 0$, $z = x + y > 0$. Thus Z has the density of a Gamma(2,1) random variable.

Example 25

Theorem (MGF of a sum)

If X and Y are independent random variables, then

$$m_{X+Y}(u) = m_X(u)m_Y(u)$$

Example

Let X and Y be i.i.d. $\text{Exp}(1)$. Prove that $Z := X + Y$ follows a $\text{Gamma}(2, 1)$ distribution from quoting MGFs.

Example 25

Example

Let X and Y be i.i.d. Exp(1). Prove that $Z := X + Y$ follows a Gamma(2, 1) distribution from quoting MGFs.

$m_X(u) = \frac{1}{1-u}$ and $m_Y(u) = \frac{1}{1-u}$. So clearly

$$m_Z(u) = m_X(u)m_Y(u) = \left(\frac{1}{1-u}\right)^2,$$

which is the MGF of a Gamma(2,1) distribution. Hence Z follows this distribution as well.

Example 26

Example

Let $X_1, \ldots, X_n$ be a sequence of i.i.d. Unif$(0,1)$ random variables. Define $Y_n = n \min\{U_1, \ldots, U_n\}$. Prove that $Y_n \xrightarrow{d} Y$, where $Y \sim \text{Exp}(1)$.

$$\begin{aligned}
F_{Y_n}(y) = \mathbb{P}(Y_n \leq y) &= \mathbb{P}(n \min\{U_1, \ldots, U_n\} \leq y) \\
&= \mathbb{P}\left(\min\{U_1, \ldots, U_n\} \leq \frac{y}{n}\right)
\end{aligned}$$

Example 26

Example

Let $X_1, \ldots, X_n$ be a sequence of i.i.d. $\text{Unif}(0,1)$ random variables. Define $Y_n = n \min\{U_1, \ldots, U_n\}$. Prove that $Y_n \xrightarrow{d} Y$, where $Y \sim \text{Exp}(1)$.

$$\begin{aligned}
F_{Y_n}(y) = \mathbb{P}(Y_n \le y) &= \mathbb{P}(n \min\{U_1, \ldots, U_n\} \le y) \\
&= \mathbb{P}\left(\min\{U_1, \ldots, U_n\} \le \frac{y}{n}\right) \\
&= 1 - \mathbb{P}\left(\min\{U_1, \ldots, U_n\} \ge \frac{y}{n}\right)
\end{aligned}$$

In general, if $\min\{x_1, \ldots, x_n\} \le x$, then **not every** $x_i \le x$.

Example 26

Example

Let $X_1, \dots, X_n$ be a sequence of i.i.d. $\mathrm{Unif}(0,1)$ random variables. Define $Y_n = n \min\{U_1, \dots, U_n\}$. Prove that $Y_n \xrightarrow{d} Y$, where $Y \sim \mathrm{Exp}(1)$.

$$\begin{aligned} F_{Y_n}(y) = \mathbb{P}(Y_n \leq y) &= \mathbb{P}(n \min\{U_1, \dots, U_n\} \leq y) \\ &= \mathbb{P}\left(\min\{U_1, \dots, U_n\} \leq \frac{y}{n}\right) \\ &= 1 - \mathbb{P}\left(\min\{U_1, \dots, U_n\} \geq \frac{y}{n}\right) \\ &= 1 - \mathbb{P}\left(U_1 > \frac{y}{n}, \dots, U_n > \frac{y}{n}\right) \end{aligned}$$

But it **is** true that if $\min\{U_1, \dots, U_n\} \geq x$, then every $x_i \geq x$.

Example 26

Example

Let $X_1, \ldots, X_n$ be a sequence of i.i.d. Unif$(0,1)$ random variables. Define $Y_n = n\min\{U_1, \ldots, U_n\}$. Prove that $Y_n \xrightarrow{d} Y$, where $Y \sim \text{Exp}(1)$.

$$\begin{aligned} F_{Y_n}(y) &= 1 - \mathbb{P}\left(U_1 > \frac{y}{n}, \ldots, U_n > \frac{y}{n}\right) \\ &= 1 - \mathbb{P}\left(U_1 > \frac{y}{n}\right) \ldots \mathbb{P}\left(U_n > \frac{y}{n}\right) && \text{(independence)} \\ &= 1 - \left[\mathbb{P}\left(U_1 > \frac{y}{n}\right)\right]^n && \text{(id. distributed)} \end{aligned}$$

Example 26

Example

Let $X_1, \dots, X_n$ be a sequence of i.i.d. Unif$(0,1)$ random variables. Define $Y_n = n\min\{U_1, \dots, U_n\}$. Prove that $Y_n \xrightarrow{d} Y$, where $Y \sim \text{Exp}(1)$.

$$
\begin{aligned}
F_{Y_n}(y) &= 1 - \mathbb{P}\left(U_1 > \frac{y}{n}, \dots, U_n > \frac{y}{n}\right) \\
&= 1 - \mathbb{P}\left(U_1 > \frac{y}{n}\right) \dots \mathbb{P}\left(U_n > \frac{y}{n}\right) && \text{(independence)} \\
&= 1 - \left[\mathbb{P}\left(U_1 > \frac{y}{n}\right)\right]^n && \text{(id. distributed)} \\
&= 1 - \left[\int_{y/n}^{1} 1\, dt\right]^n = 1 - \left(1 - \frac{y}{n}\right)^n
\end{aligned}
$$

Example 26

Example

Let $X_1, \ldots, X_n$ be a sequence of i.i.d. Unif$(0,1)$ random variables. Define $Y_n = n \min\{U_1, \ldots, U_n\}$. Prove that $Y_n \xrightarrow{d} Y$, where $Y \sim \text{Exp}(1)$.

$$\therefore \lim_{n \to \infty} F_{Y_n}(y) = 1 - e^{-y} = F_Y(y)$$

Hence $Y_n \xrightarrow{d} Y$.

Example 27

Example (Libo's notes)

Australians have average weight about 68 kg and variance about 16 kg^2. Suppose 40 random Australians are chosen. What is the (approximate) probability that the average weight of these Australians is over 80?

Let $X_1, \ldots, X_{40}$ be the weights of the Australians.

Example 27

Example (Libo's notes)

Australians have average weight about 68 kg and variance about 16 kg^2. Suppose 40 random Australians are chosen. What is the (approximate) probability that the average weight of these Australians is over 80?

Let $X_1, \ldots, X_{40}$ be the weights of the Australians. Then $n = 40$, $\mu = 68$ and $\sigma = 4$, so by the CLT:

$$\frac{\overline{X} - 68}{4/\sqrt{40}} \xrightarrow{d} Z$$

where $Z \sim \mathcal{N}(0, 1)$.

Example 27

Example (Libo's notes)

Australians have average weight about 68 kg and variance about 16 kg^2. Suppose 40 random Australians are chosen. What is the (approximate) probability that the average weight of these Australians is over 80?

$$\begin{aligned}\therefore \mathbb{P}(\overline{X_{40}} > 80) &= \mathbb{P}\left(\frac{\overline{X_{40}} - 68}{4/\sqrt{40}} > \frac{80-68}{4/\sqrt{40}}\right) \\ &\approx \mathbb{P}\left(Z > \frac{80-68}{4/\sqrt{40}}\right)\end{aligned}$$

Example 27

Example (Libo's notes)

Australians have average weight about 68 kg and variance about 16 kg^2. Suppose 40 random Australians are chosen. What is the (approximate) probability that the average weight of these Australians is over 80?

$$
\begin{aligned}
\therefore \mathbb{P}(\overline{X_{40}} > 80) &= \mathbb{P}\left(\frac{\overline{X_{40}} - 68}{4/\sqrt{40}} > \frac{80-68}{4/\sqrt{40}}\right) \\
&\approx \mathbb{P}\left(Z > \frac{80-68}{4/\sqrt{40}}\right) \\
&= \mathbb{P}(Z > 3\sqrt{40}) \\
&= \texttt{1-pnorm(3*sqrt(40))} \\
&\text{or } \texttt{pnorm(3*sqrt(40), lower.tail=FALSE)}
\end{aligned}
$$

Example 28

Lemma (Normal Approximation to Binomial)

Let $X \sim \text{Bin}(n, p)$, which is a sum of n independent $\text{Ber}(p)$ r.v.s. Then

$$\frac{X - np}{\sqrt{np(1-p)}} \xrightarrow{d} \mathcal{N}(0, 1)$$

Example 28

Example

An unfortunate soul decided to sit his exam despite having a migraine and the flu. Fortunately, it was not a university exam, and the paper involved only 200 multiple choice questions with 5 options. Therefore, he randomly guesses every answer. What is the (approximate) probability he fails?

Let X be how many he gets correct. Then $X \sim \text{Bin}\left(200, \frac{1}{5}\right)$.

Example 28

Example

An unfortunate soul decided to sit his exam despite having a migraine and the flu. Fortunately, it was not a university exam, and the paper involved only 200 multiple choice questions with 5 options. Therefore, he randomly guesses every answer. What is the (approximate) probability he fails?

Let X be how many he gets correct. Then $X \sim \text{Bin}\left(200, \frac{1}{5}\right)$.
We may approximate X with $Y \sim \mathcal{N}(40, 32)$. Then,

$$\mathbb{P}(X < 100) \approx \mathbb{P}(Y < 100)$$

Example 28

Example

An unfortunate soul decided to sit his exam despite having a migraine and the flu. Fortunately, it was not a university exam, and the paper involved only 200 multiple choice questions with 5 options. Therefore, he randomly guesses every answer. What is the (approximate) probability he fails?

Let X be how many he gets correct. Then $X \sim \text{Bin}\left(200, \frac{1}{5}\right)$.
We may approximate X with $Y \sim \mathcal{N}(40, 32)$. Then,

$$\begin{aligned}
\mathbb{P}(X < 100) &\approx \mathbb{P}(Y < 100) \\
&= \mathbb{P}\left(\frac{Y - 40}{\sqrt{32}} < \frac{100 - 40}{\sqrt{32}}\right) \\
&= \mathbb{P}\left(Z < \frac{60}{\sqrt{32}}\right) \\
&= \mathbb{P}(Z < 10.6066)
\end{aligned}$$

Example 28

Example

An unfortunate soul decided to sit his exam despite having a migraine and the flu. Fortunately, it was not a university exam, and the paper involved only 200 multiple choice questions with 5 options. Therefore, he randomly guesses every answer. What is the (approximate) probability he fails?

Let X be how many he gets correct. Then $X \sim \text{Bin}\left(200, \frac{1}{5}\right)$.
We may approximate X with $Y \sim \mathcal{N}(40, 32)$. Then,

$$\begin{aligned}
\mathbb{P}(X < 100) &\approx \mathbb{P}(Y < 100) \\
&= \mathbb{P}\left(\frac{Y - 40}{\sqrt{32}} < \frac{100 - 40}{\sqrt{32}}\right) \\
&= \mathbb{P}\left(Z < \frac{60}{\sqrt{32}}\right) \\
&= \mathbb{P}(Z < 10.6066) \qquad \text{Oh my...}
\end{aligned}$$

Example 29

Example (Libo's notes)

Let $X_1, X_2, \ldots$ be a sequence of i.i.d random variables with mean 2 and variance 7. Obtain a large sample approximation for the distribution of $(X_n)^3$.

Example 29

The CLT gives

$$\sqrt{n}(\bar{X}_n - 2) \xrightarrow{\text{d}} N(0, 7)$$

Applying the Delta Method with $g(x) = x^3$ leads to $g'(x) = 3x^2$ and then

$$\sqrt{n}[(\bar{X}_n^3) - 2^3] \xrightarrow{\text{d}} N(0, 7 \cdot (3 \cdot 2^2)^2).$$

Simplifying, we have

$$\sqrt{n}[(\bar(X)_n^3 - 8] \xrightarrow{\text{d}} N(0, 1008).$$

Thus, for large n, the approximate distribution of $(\bar{X}_n)^3$ is $N(8, \frac{1008}{n})$.

Example 30

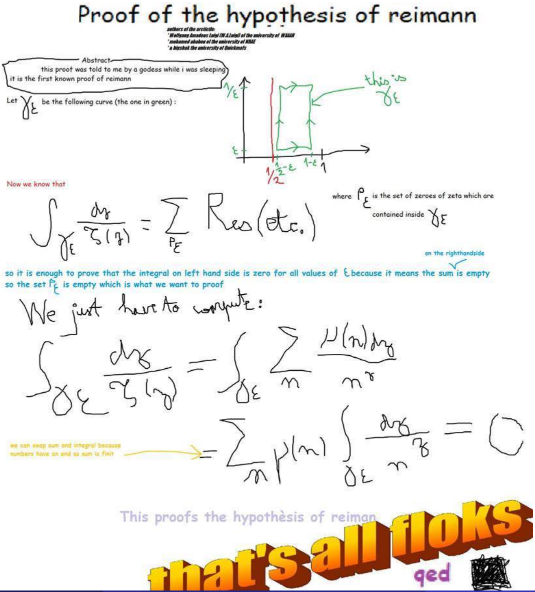
Example (More 2801 focused)

The Riemann zeta function is defined for complex s with real part greater than 1 by the absolutely convergent infinite series

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

Prove that the real part of every non-trivial zero of the Riemann zeta function is $\frac{1}{2}$.

Example 30


Proof of the hypothesis of reimann
authors of the arcticle:
'Wolfgang Amadeus Luigi (W.A.Luigi) of the university of WAAAH
'mohamed ababou of the university of NHAE
'a bigshak the university of Quickmafs
Abstract
this proof was told to me by a godess while i was sleeping
it is the first known proof of reimann
Let γ_ε be the following curve (the one in green) :
this is γ_ε
$1/\varepsilon$
ε
$1/2$
$\frac{1}{2}-\varepsilon$
$1-\varepsilon$
1
Now we know that
$\int_{\gamma_\varepsilon} \frac{dz}{\zeta(z)} = \sum_{P_\varepsilon} \mathrm{Res}(\text{etc.})$
where P_ε is the set of zeroes of zeta which are contained inside γ_ε
so it is enough to prove that the integral on left hand side is zero for all values of ε because it means the sum on the righthandside is empty
so the set P_ε is empty which is what we want to proof
We just have to compute:
$\int_{\gamma_\varepsilon} \frac{dz}{\zeta(z)} = \int_{\gamma_\varepsilon} \sum_n \frac{\mu(n)\,dz}{n^z}$
we can swap sum and integral because numbers have an end so sum is finit
$= \sum_n \mu(n) \int_{\gamma_\varepsilon} \frac{dz}{n^z} = 0$
This proofs the hypothèsis of reiman
that's all floks
qed