MATH2801/2901 Final Revision

Part I: Probability Theory

Rui Tong

UNSW Society of Statistics

29 May 2018

Table of Contents

Categorical v.s. Quantitative

Categorical

Based off some 'category'.

E.g. Sunny v.s. Cloudy, Male v.s. Female

Quantitative

Based off some 'scale'; usually involves numbers.

E.g. Weight, Precipitation, Age lived

Course Focus - Quantitative Data

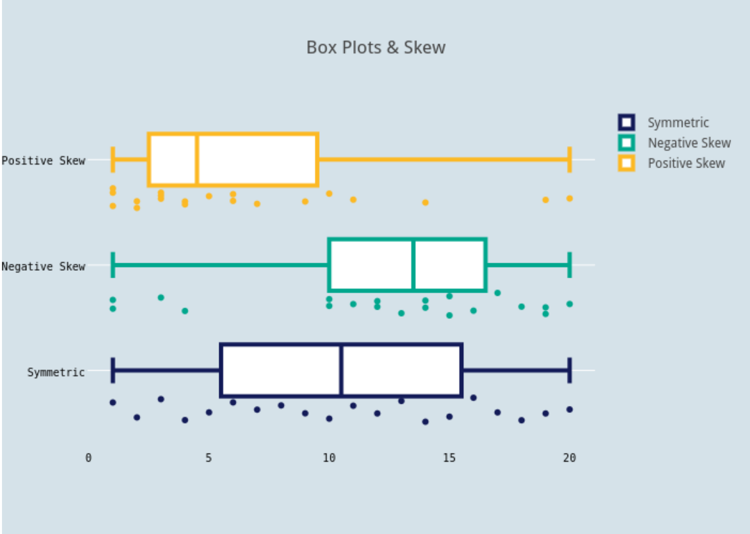
Nature of quantitative data

- Location - Where abouts is the data centered?
- Scale - To what extent is the data spread around there?
- Shape - Symmetric v.s. Skewed

Skewness of data

- *Negatively* skewed data is clustered towards the *right*.
- *Positively* skewed data is clustered towards the *left*.

Boxplots


Box Plots & Skew
Symmetric
Negative Skew
Positive Skew
Positive Skew
Negative Skew
Symmetric
0
5
10
15
20

Probability

Definition (2901)

A probability is a function $\mathbb{P}$ that assigns a value in $[0,1]$ from events in the sample space Ω, in the σ-algebra (say $\mathcal{A}$).

Definition (Probability Space) (2901)

A *probability space* is the triple $(\Omega, \mathcal{A}, \mathbb{P})$ with the axioms

$$\mathbb{P}(A) \geq 0 \qquad \forall A \in \mathcal{A}$$
$$\mathbb{P}(\Omega) = 1$$
$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

for mutually exclusive events $A_1, A_2, \cdots \in \mathcal{A}$

Probability

Definition (Probability Space) (2901)

A *probability space* is the triple $(\Omega, \mathcal{A}, \mathbb{P})$ with the axioms

$$\mathbb{P}(A) \geq 0 \qquad \forall A \in \mathcal{A}$$
$$\mathbb{P}(\Omega) = 1$$
$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

for mutually exclusive events $A_1, A_2, \cdots \in \mathcal{A}$

Don't worry too much about them.

Complementary Event

Definition (Complement)

Given an event A, the complement A^c is essentially the event representing 'not A'

Theorem (Probability of a complement)

For any event $A \in \mathcal{A}$,

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$$

Conditional Probability

Definition (Conditional Probability)

Given that the event $B \in \mathcal{A}$ has occurred, the probability of $A \in \mathcal{A}$ occuring is

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

Theorem (Multiplication Law)

If $\mathbb{P}(B) \neq 0$, then the probability of A and B occurring is

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B)$$

and similarly if $\mathbb{P}(A) \neq 0$,

$$\mathbb{P}(A \cap B) = \mathbb{P}(B \mid A)\mathbb{P}(A)$$

Conditional Probability

Theorem (Multiplication Law)

If $\mathbb{P}(B) \neq 0$, then the probability of A and B occurring is

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B)$$

Example (MATH1251)

A diagnostic test has 99% chance of correctly detecting if someone has a disease. If only 2% of the population have this disease, what is the probability that someone has the disease and was successfully tested for it?

$$\mathbb{P}(D \cap T) = \mathbb{P}(T \mid D)\mathbb{P}(D) = 0.99 \times 0.02$$

Independence

Definition (Independence)

Two events $A, B \in \mathcal{A}$ are independent if

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

Remark

If $\mathbb{P}(B) \neq 0$, then two events are independent iff

$$\mathbb{P}(A \mid B) = \mathbb{P}(A)$$

Total Probability

Theorem (Law of Total Probability)

Let the events $A_1, A_2, \ldots$ be mutually exclusive. Then

$$\begin{aligned}\mathbb{P}(B) &= \mathbb{P}(B \mid A_1)\mathbb{P}(A_1) + \mathbb{P}(B \mid A_2)\mathbb{P}(A_2) + \ldots \\ &= \sum_i \mathbb{P}(B \mid A_i)\mathbb{P}(A_i).\end{aligned}$$

We can have a finite *or* infinite number of events A_i.

Bayes' Law

Theorem (Bayes' Law)

Let the events $A_1, A_2, \ldots$ be mutually exclusive. Then

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

Often used in conjunction with the law of total probability to obtain

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\sum_i \mathbb{P}(B \mid A_i)\mathbb{P}(A_i)}$$

Bayes' Law

Theorem (Bayes' Law)

Let the events $A_1, A_2, \dots$ be mutually exclusive. Then

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

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.

Bayes' Law

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.

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

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

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

Bayes' Law

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

Discrete Random Variables

Essentially, a r.v. X assigns a value to an event.

Definition (Discrete Random Variable)

X is a discrete random variable if it can only take countably many values.

The probability function is denoted

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

In 2801, this is also denoted $f_X(x)$ for the discrete case.

Validity of the discrete random variable

Properties of the discrete random variable

A discrete random variable must satisfy

- $\mathbb{P}(X = x) \geq 0$ for all x
- $\sum_{\text{all } x} \mathbb{P}(X = x) = 1$

Example

A discrete random satisfies $\mathbb{P}(X = 1) = \frac{1}{3}$ and $\mathbb{P}(X \neq -1, X \neq 1) = 0$. What must $\mathbb{P}(X = -1)$ equal to?

Validity of the discrete random variable

Properties of the discrete random variable

A discrete random variable must satisfy

- $\mathbb{P}(X = x) \geq 0$ for all x
- $\sum_{\text{all } x} \mathbb{P}(X = x) = 1$

Example

A discrete random satisfies $\mathbb{P}(X = 1) = \frac{1}{3}$ and $\mathbb{P}(X \neq -1, X \neq 1) = 0$. What must $\mathbb{P}(X = -1)$ equal to?

From the second property, $\mathbb{P}(X = -1) = 1 - \frac{1}{3} = \frac{2}{3}$.

Continuous Random Variables

Definition (Continuous Random Variable)

X is a continuous random variable if it takes uncountably many values.

The density function is denoted

$$f_X(x)$$

Validity of the continuous random variable

Properties of the continuous random variable

A continuous random variable must satisfy

- $f_X(x) \geq 0$ for all x
- $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$

Example

Can $f_X(x) = 2e^{-x}$ for $x \geq 0$ be a continuous random variable?

No, because $\int_{-\infty}^{\infty} f_X(x) = \int_0^{\infty} 2e^{-x}\, dx = 2$.

Remark

If X is a continuous random variable, then $\mathbb{P}(X = x) = 0$ for any x. We *must* consider the probability that it lies in some interval.

If X is a continuous random variable, it's always defined on some interval (can be $\mathbb{R}$). As a convention, wherever it's not defined we just assume that the density is 0.

Cumulative Distribution Function

Definition (Cumulative Distribution Function)

The CDF $F_X(x)$ is the function given by $\boxed{F_X(x) = \mathbb{P}(X \leq x)}$

Properties of the CDF (2901)

The CDF must satisfy the following properties

- $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to +\infty} F_X(x) = 1$
- $F_X(x)$ is non-decreasing
- Right-continuous

Important property of the CDF

Assuming $a < b$,

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

Where people lose marks

The CDF isn't just defined over some small interval. It's defined over all of $\mathbb{R}$.

Finding a Cumulative Distribution Function

Discrete case

Add up all the probabilities you require.

Continuous case

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

Lemma (Continuous case):

$$\mathbb{P}(a < X \leq b) = \int_{a}^{b} f_X(t)\,dt$$

Finding a Cumulative Distribution Function

Example

Derive the CDF of X if $X \sim \text{Unif}(0, 1)$. That is to say,

$$f_X(x) = \begin{cases} 1 & x \in (0, 1) \\ 0 & \text{otherwise} \end{cases}$$

$$F_X(x) = \int_0^x 1 \, dt = x.$$

Trap! We need to consider the cases for *every* real number x!

Finding a Cumulative Distribution Function

Example

Derive the CDF of X if $X \sim \text{Unif}(0, 1)$. That is to say,

$$f_X(x) = \begin{cases} 1 & x \in (0, 1) \\ 0 & \text{otherwise} \end{cases}$$

For $x \leq 0$, we have

$$F_X(x) = \int_{-\infty}^{x} 0 \, dt = 0$$

Finding a Cumulative Distribution Function

Example

Derive the CDF of X if $X \sim \text{Unif}(0,1)$. That is to say,

$$f_X(x) = \begin{cases} 1 & x \in (0,1) \\ 0 & \text{otherwise} \end{cases}$$

For $x \leq 0$, we have

$$F_X(x) = \int_{-\infty}^{x} 0\,dt = 0$$

For $0 < x < 1$, we have

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f_X(t)\,dt \\ &= \int_{-\infty}^{0} 0\,dt + \int_{0}^{x} 1\,dt \\ &= x \end{aligned}$$

Finding a Cumulative Distribution Function

Example

Derive the CDF of X if $X \sim \text{Unif}(0,1)$. That is to say,

$$f_X(x) = \begin{cases} 1 & x \in (0,1) \\ 0 & \text{otherwise} \end{cases}$$

For $x \geq 1$, we have

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f_X(t)\,dt \\ &= \int_{-\infty}^{0} 0\,dt + \int_0^1 1\,dt + \int_1^x 0\,dt \\ &= 1 \end{aligned}$$

Finding a Cumulative Distribution Function

Example

Derive the CDF of X if $X \sim \text{Unif}(0,1)$. That is to say,

$$f_X(x) = \begin{cases} 1 & x \in (0,1) \\ 0 & \text{otherwise} \end{cases}$$

Therefore:

$$F_X(x) = \begin{cases} 0 & \text{if } x \le 0 \\ x & \text{if } 0 < x < 1 \\ 1 & \text{if } x \ge 1 \end{cases}$$

E.g. $F_X(\frac{1}{2}) = \mathbb{P}(X \le \frac{1}{2}) = \frac{1}{2}$

Remark

That was not necessarily the most efficient way of doing the problem.

We could've recycled some earlier computations along the way.

CDF of a continuous random variable

Lemma

$$\frac{d}{dx}F_X(x) = f_X(x)$$

Quantiles

Definition (Quantiles)

The k-th quantile of X is the solution to the equation

$$F_X(x) = k.$$

Example: The median is just the value of x such that $F_X(x) = \frac{1}{2}$.

Useful remark (2901)

The function Q_X is just the inverse function of F_X.

Example

Find the lower quartile (25% quantile) of the $\text{Exp}\left(\frac{1}{2}\right)$ distribution.

Quantiles

Example

Find the lower quartile (25% quantile) of the $\mathrm{Exp}\left(\frac{1}{2}\right)$ distribution.

The density function is $f_X(x) = \frac{1}{2}e^{-x/2}$ for $x \geq 0$. We're only interested in the CDF for $x \geq 0$.

$$F_X(x) = \int_0^x \frac{1}{2}e^{-t/2}\,dt = 1 - e^{-x/2}$$

(for $x \geq 0$).

Quantiles

Example

Find the lower quartile (25% quantile) of the $\text{Exp}\left(\frac{1}{2}\right)$ distribution.

Setting $F_X(x) = \frac{1}{4}$ gives

$$\begin{aligned} \frac{1}{4} &= 1 - e^{-x/2} \\ e^{-x/2} &= \frac{3}{4} \\ \frac{x}{2} &= -\ln\frac{3}{4} \\ x &= 2\ln\frac{4}{3} \end{aligned}$$

Expectation

Definition (Expected Value)

For a discrete random variable X, its expectation is

$$\mathbb{E}[X] = \sum_{\text{all } x} x\,\mathbb{P}(X = x).$$

For a continuous random variable X, its expectation is

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

Expectation

Definition (Expected Value after Transform)

For a discrete random variable X.

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

For a continuous random variable X,

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

Properties of the Expectation

Theorem (Properties of taking expectation)

- $\mathbb{E}[aX] = a\mathbb{E}[X]$
- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$
- $\mathbb{E}[1] = 1$

Critical misassumption

In general, for any function f,

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

Variance and Standard Deviation

Let $\mathbb{E}[X] = \mu$

Definition (Variance)

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

Theorem (Variance Formula)

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

Definition (Standard Deviation)

$$\text{SD}(X) = \sigma_X = \sqrt{\text{Var}(X)}$$

Variance and Standard Deviation

Example (Trivial for 2901)

Prove the variance formula from the definition

$$\begin{aligned}\mathbb{E}\left[(X-\mu)^2\right] &= \mathbb{E}\left[X^2 - 2\mu X + \mu^2\right] \\ &= \mathbb{E}[X^2] - 2\mu\mathbb{E}[X] + \mu^2\mathbb{E}[1] \\ &= \mathbb{E}[X^2] - 2\mu\mu + \mu^2 \\ &= \mathbb{E}[X^2] - \mu^2\end{aligned}$$

Properties of the Variance

Theorem (Properties of taking variances)

- $\text{Var}(X + b) = \text{Var}(X)$
- $\text{Var}(aX) = a^2 \text{Var}(X)$
- $\text{Var}(aX + b) = a^2 \text{Var}(X)$
- $\text{Var}(1) = 0$

Critical misassumption

In general, for any two random variables X and Y,

$$\text{Var}(X + Y) \neq \text{Var}(X) + \text{Var}(Y)$$

Expectation Computations

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

Expectation Computations

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

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

Expectation Computations

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

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

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

Expectation Computations

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

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

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

Expectation Computations

Example (2901 oriented)

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

Expectation Computations

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 $x = 1, 2, \dots$

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

Expectation Computations

Example (2901 oriented)

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

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

Expectation Computations

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

Expectation Computations

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

Expectation Computations

Example (2901 oriented)

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

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

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

Expectation Computations (2901)

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)

Expectation Inequalities

Theorem (Chebychev's (Second) Inequality)

Let $\mathbb{E}[X] = \mu$ and $SD(X) = \sigma$. Then, *regardless of the distribution* of X,

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

Note that this is an *upper* bound.

Expectation Inequalities

Example - Bounding problem (MATH2801 notes)

A factory produces 500 machines a day on average. It is subject to a variance of 100. Let X be the amount of machines produced tomorrow. Find a *lower* bound for the probability that between 400 to 600 machines are produced tomorrow.

We require some bound for $\mathbb{P}(400 \leq X \leq 600)$. Observe that:

$$\begin{aligned}\mathbb{P}(400 \leq X \leq 600) &= \mathbb{P}(-100 \leq X - 500 \leq 100) \\ &= \mathbb{P}(|X - 500| \leq 100) \\ &= \mathbb{P}(|X - \mu| \leq k\sigma^2)\end{aligned}$$

where $\mu = 500$, $\sigma^2 = 100$ and therefore $\boxed{\sigma = 10}$ and $\boxed{k = 10}$.

Expectation Inequalities

Example - Bounding problem (MATH2801 notes)

A factory produces 500 machines a day on average. It is subject to a variance of 100. Let X be the amount of machines produced tomorrow. Find a *lower* bound for the probability that between 400 to 600 machines are produced tomorrow.

From Chebychev's (second) inequality,

$$\mathbb{P}(|X-\mu| > 10\sigma) < \frac{1}{10^2}$$
$$\therefore 1 - \mathbb{P}(|X-\mu| \le 10\sigma) < \frac{1}{100}$$
$$\mathbb{P}(400 \le X \le 600) > \frac{99}{100}$$

Expectation Inequalities

Theorem (Markov's inequality) (2901)

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

Theorem (Jensen's inequality) (2901)

If h is a convex function (aka. concave up function), then

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

Moment Generating Functions

Definition (Moments)

The r-th moment of a random variable X is $\mathbb{E}[X^r]$.

Definition (MGF)

The moment generating function of a random variable X is

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

Properties of the MGF

Theorem (MGF uniquely characterises distributions)

$$m_X(u) = m_Y(u) \iff F_X(x) = F_Y(x)$$

Theorem (MGF of a sum of independent r.v.s)

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

Lemma (Computing moments)

The r-th moment, is the limit as $u \to 0$, *of* the r-th derivative:

$$\mathbb{E}[X^r] = \lim_{u \to 0} \frac{d^r}{dx} m_X(u)$$

Properties of the MGF

Definition (Existence of MGF) (2901)

The MGF must be finite for some interval $[-h, h]$ containing 0.

(However it need not be defined at 0...)

What??

Example

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

What??

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

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

What??

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

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

What??

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

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

What??

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

What??

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.

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

Using the MGF

Example

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

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

Using the MGF

Example

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

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

Bernoulli distribution

Definition (Bernoulli Distribution)

A random variable X follows a $\text{Ber}(p)$ distribution if

$$\mathbb{P}(X = x) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \end{cases}$$

Significance of each parameter

p is the probability of success.

Usage

Used to model (the likelihood of) something that either does or does not happen.

Binomial distribution

Definition (Binomial Distribution)

A random variable X follows a $\text{Bin}(n, p)$ distribution if

$$\mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, \ldots, n$$

Significance of each parameter

- n is the number of trials.
- p is the probability of success.

Usage

Used to model how many successes in a total of n Bernoulli trials.

Hypergeometric distribution (ignored in 2901)

Definition (Hypergeometric Distribution)

A random variable X follows a $\text{Hyp}(N, m, n)$ distribution if

$$\mathbb{P}(X = x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}} \qquad 0 \leq x \leq \min(m, n)$$

Significance of each parameter

- n is the number of times we select the items.
- N is the size of the population.
- m is number of items in the pop. satisfying some criteria.

Usage

Used to model how likely we choose x out of the m desirable items.

Hypergeometric V.S. Binomial

Hypergeometric assumes no replacement changes things. Binomial is typically for situations with 'replacement'.

Geometric Distribution

Definition (Geometric Distribution)

A random variable X follows a $\mathrm{Geom}(p)$ distribution if

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

Significance of each parameter

p is the probability of success.

Usage

Used to model how many Bernoulli trials we need before we reach the *first* success outcome.

Poisson Distribution

Definition (Geometric Distribution)

A random variable X follows a Poisson(λ) distribution if

$$\mathbb{P}(X = x) = e^{-\lambda}\frac{\lambda^x}{x!} \qquad x = 0.1, 2, \dots$$

Significance of each parameter

λ is the average number of occurrences of an event

Usage

Used to model events that are rare. Recommended when an occurrence of an event is independent from another occurrence.

Example - Computing probabilities

Example

5 cards without replacement from an ordinary deck of playing cards. What is the probability of getting exactly 2 red cards (i.e., hearts or diamonds)?

Example - Computing probabilities

Example

5 cards without replacement from an ordinary deck of playing cards. What is the probability of getting exactly 2 red cards (i.e., hearts or diamonds)?

- No replacement - Hypergeometric
- $N = 52$ (number of cards)
- $m = 26$ (number of favourable cards, i.e. red cards)
- $n = 5$ (number of draws)

Example - Computing probabilities

Example

5 cards without replacement from an ordinary deck of playing cards. What is the probability of getting exactly 2 red cards (i.e., hearts or diamonds)?

- No replacement - Hypergeometric
- $N = 52$ (number of cards)
- $m = 26$ (number of favourable cards, i.e. red cards)
- $n = 5$ (number of draws)

We are considering $x = 2$.

$$\mathbb{P}(X = 2) = \frac{\binom{26}{2}\binom{52-26}{5-2}}{\binom{52}{5}} \approx 0.3251$$

Remark

If we had replacement, we would have a probability $p = \frac{26}{52} = \frac{1}{2}$, so we would consider $\text{Bin}\left(5, \frac{1}{2}\right)$

Example - Computing probabilities

Example

A busy switchboard receives 150 calls an hour on average. Assume that every call is indep and can be modelled with a Poisson distribution. from each other. 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 - Computing probabilities

Example

A busy switchboard receives 150 calls an hour on average. Assume that every call is indep and can be modelled with a Poisson distribution. from each other. 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 - Computing probabilities

Example

A busy switchboard receives 150 calls an hour on average. Assume that every call is indep and can be modelled with a Poisson distribution. from each other. 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 - Computing probabilities

Example

A busy switchboard receives 150 calls an hour on average. Assume that every call is indep and can be modelled with a Poisson distribution. from each other. 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,

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

Exponential Distribution

Definition (Exponential Distribution)

A random variable T follows an $\text{Exp}(\beta)$ distribution if

$$f_T(t) = \frac{1}{\beta}e^{-t/\beta} \qquad t > 0$$

Significance of each parameter

$\beta = \frac{1}{\lambda}$. It is the average time taken until the next occurrence of the event

Usage

Based off the memory-less property (see next slide).

Exponential Distribution - Lack of Memory

Theorem (Memory-less property)
A continuous distribution satisfies the memoryless property

$$\mathbb{P}(T > s + t \mid T > s) = \mathbb{P}(T > t)$$

if and only if it is an exponential distribution.

Usage

The exponential distribution is used to measure the time taken between consecutive independent events.

Example

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

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

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

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**.

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

Uniform Distribution

Definition (Uniform Distribution)

A random variable X follows a $\text{Unif}(a, b)$ distribution if

$$f_X(x) = \frac{1}{b-a} \qquad a < x < b.$$

Significance of the parameters

a and b are the two endpoints.

Gamma Distribution (2901)

Definition (Gamma Distribution)

A random variable X follows a Gamma(α, β) distribution if

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

Significance of the parameters

- β is the same as in the exponential distribution
- α - not too obvious, don't worry about it.

Relationships between Random Variables (2901)

Acronym - 'iid.' stands for independent, identically distributed

Theorem (Bernoulli sums to Binomial)

If $X_1, \ldots, X_n$ is a sequence of Ber(p) random variables, then

$$Y := \sum_{i=1}^{n} X_i \sim \text{Bin}(n, p)$$

Theorem (Exponential sums to Gamma)

If $X_1, \ldots, X_n$ is a sequence of Exp(β) random variables, then

$$Y := \sum_{i=1}^{n} X_i \sim \text{Gamma}(\alpha, \beta)$$

(We'll come back to this later.)

Normal Distribution

Definition (Normal Distribution)

A random variable X follows a $\mathcal{N}(\mu, \sigma^2)$ distribution if

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

Significance of the parameters

- μ is its mean
- σ^2 is its variance

Definition (Standard Normal Distribution)

If $Z \sim \mathcal{N}(0, 1)$, then Z follows the standard normal distribution.

Transforms

Loose definition (Transform)

The transformation of a random variable X under some function h, is just $h(X)$.

Comparing Distributions - QQ Plots

Definition (Quantile-Quantile Plot)

For two data sets, the plot of their quantiles against each other is called a Quantile-Quantile Plot.

Using QQ plots

We seek if the QQ plot between our data and that from a *known* distribution is linear. If this is the case, then they are *linear* transforms of each other.

Sketch of execution

Given some data, we plot its quantiles against that of $\mathcal{N}(0,1)$. If the graph is linear, then the unknown data is also from a normal distribution.

Transforms on a Discrete Random Variable

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?

Transforms on a Discrete Random Variable

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 $Z = X^2$.

Transforms on a Discrete Random Variable

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

Transforms on a Discrete Random Variable

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.

Transforms on a Discrete Random Variable

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.

Transforms on a Discrete Random Variable

Just to think about... (2901 oriented)

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

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

Transforms on a Continuous Random Variable

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 $Y = X^2$?

Transforms on a Continuous Random Variable

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 $\frac{dx}{dy} = \frac{1}{2\sqrt{y}}$

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

Transforms on a Continuous Random Variable

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 $\frac{dx}{dy} = \frac{1}{2\sqrt{y}}$

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

Transforms on a Continuous Random Variable

Method 2

Brute force via the CDF. (Used when h is not invertible over our region.)

Example

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

Transforms on a Continuous Random Variable

Example

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

$f_X(x) = \frac{1}{20}$ for $x \in (-10, 10)$. But clearly $h(x) = x^2$ is not invertible over this interval!

Transforms on a Continuous Random Variable

Example

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

$$\begin{aligned} F_Y(y) = \mathbb{P}(Y \leq y) &= \mathbb{P}(X^2 \leq y) \\ &= \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

Transforms on a Continuous Random Variable

Example

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

$$\begin{aligned} F_Y(y) = \mathbb{P}(Y \leq y) &= \mathbb{P}(X^2 \leq y) \\ &= \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

Taking derivatives w.r.t y with the chain rule:

$$\begin{aligned} f_Y(y) &= \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y}) \\ &= \frac{1}{2\sqrt{y}} \times \frac{1}{20} + \frac{1}{2\sqrt{y}} \times \frac{1}{20} \\ &= \frac{1}{20\sqrt{y}} \end{aligned}$$

Where everybody loses marks

For what values of x is the transformed random variable defined for???

Intervals that random variables are defined on

In general, once you transform a random variable, the new interval it's defined on *may not be the same as the old one*.

Finishing off the earlier problems

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.

Finishing off the earlier problems

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 $0 < y < 100$.

Probabilities in the Normal Distribution

Theorem (Standardisation of a Normal r.v.)

Let X be a $\mathcal{N}(\mu, \sigma^2)$ random variable. Then,

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

Definition (Phi function)

$\Phi(x)$ is the CDF of the $\mathcal{N}(0, 1)$ distribution. It has properties

- $\lim_{x \to -\infty} \Phi(x) = 0$ and $\lim_{x \to +\infty} \Phi(x) = 1$
- $\Phi(-x) = 1 - \Phi(x)$
- $\Phi(0) = 0.5$
- Monotonic increasing (just like every CDF)
- Accessible on R via `pnorm(x, lower.tail = TRUE)`

Probabilities in the Normal Distribution

Example (2801 notes)

The distribution of young men's heights is approximately normally distributed with mean 174 cm and variance 40.96 cm. What is the probability that a randomly selected young man's height is one-hundred-and-seventy-something cm tall?

Let X be the height of a young man. Then $X \sim \mathcal{N}(174, 40.96)$. We require:

Probabilities in the Normal Distribution

Example (2801 notes)

The distribution of young men's heights is approximately normally distributed with mean 174 cm and variance 40.96 cm. What is the probability that a randomly selected young man's height is one-hundred-and-seventy-something cm tall?

Let X be the height of a young man. Then $X \sim \mathcal{N}(174, 40.96)$. We require:

$$
\begin{aligned}
\mathbb{P}(170 \leq X < 180) &= \mathbb{P}\left(\frac{170-174}{6.4} \leq \frac{1X-174}{6.4} < \frac{180-174}{6.4}\right) \\
&= \mathbb{P}(-0.625 \leq Z < 0.9375) \\
&= \Phi(0.9375) - \Phi(-0.625) \\
&= \texttt{pnorm(0.9375) - pnorm(-0.625)} \\
&\approx 0.5597638
\end{aligned}
$$

Probabilities in the Normal Distribution

Example (2801 notes)

The distribution of young men's heights is approximately normally distributed with mean 174 cm and variance 40.96 cm. What is the probability that a randomly selected young man's height is one-hundred-and-seventy-something cm tall?

Remark: We could have also done this with

```
pnorm(180,mean=174,sd=6.4) -pnorm(170,mean=174,sd=6.4)
```

Normal Distribution

Corollary (Reversing the standardisation) (2901)

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

$$X = \mu + \sigma Z \sim \mathcal{N}(\mu, \sigma)$$

Probability Theory - Random variables context

The notation $\mathbb{P}(X = x, Y = y)$ means $\mathbb{P}\big((X = x) \cap (Y = y)\big)$.

Lemma (common sense put to mathematical terms - 2901)

$$\mathbb{P}(X > a, X > b) = \mathbb{P}(X > \max\{a, b\})$$

$$\mathbb{P}(X < a, X < b) = \mathbb{P}(X < \min\{a, b\})$$

Another one (2901)

$$\mathbb{P}(X + Y = a) = \mathbb{P}(X = a - Y)$$

Definition (Conditional Probability)

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

Joint Discrete Distribution

Definition (Joint Probability Function)

If X and Y are both discrete random variables, then their joint probability function is denoted

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

In 2801, this is also denoted $f_{X,Y}(x, y)$

Properties of the joint probability function

- $\mathbb{P}(X = x, Y = y) \geq 0$ for all x, y
- $\sum_{\text{all } x} \sum_{\text{all } y} = 1$

Joint Continuous Distribution

Definition (Joint Density Function)

If X and Y are both continuous random variables, then their joint density function is denoted

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

Properties of the continuous random variable

- $f_{X,Y}(x,y) \geq 0$ for all x, y
- $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1$

Computing Probabilities - Bivariate Discrete

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

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

Computing Probabilities - Bivariate Discrete

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

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

Computing Probabilities - Bivariate Discrete

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

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

Computing Probabilities - Bivariate Continuous

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 $\mathbb{P}(X \leq Y^2)$.

Computing Probabilities - Bivariate Continuous

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 $\mathbb{P}(X \leq Y^2)$.

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

Computing Probabilities - Bivariate Continuous

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 $\mathbb{P}(X \leq Y^2)$.

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

Expectation

Note that $\mathbb{E}[X, Y]$ is not well defined.

Definition (Expectation)

Suppose that g is a function from $\mathbb{R}^2$ to $\mathbb{R}$.
For discrete random variables X and Y,

$$\mathbb{E}[g(X, Y)] = \sum_{\text{all } x} \sum_{\text{all } y} g(x, y)\mathbb{P}(X = x, Y = y)$$

For continuous random variables X and Y,

$$\mathbb{E}[g(X, Y)] = \iint_{\mathbb{R}^2} g(x, y) f_{X,Y}(x, y)\, dx\, dy$$

Expectation Computations

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

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

Expectation Computations

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

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

Expectation Computations

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

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

Mostly 2901-oriented interlude

Problem

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

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

Cumulative Distribution Function (Bivariate)

Definition (Cumulative Distribution Function)

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

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

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:

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

Marginal Functions

Definition (Marginal Probability Function)

For discrete r.v.s X and Y with mass function $\mathbb{P}(X = x, Y = y)$,

$$\mathbb{P}(X = x) = \sum_{\text{all } y} \mathbb{P}(X = x, Y = y)$$
$$\mathbb{P}(Y = y) = \sum_{\text{all } x} \mathbb{P}(X = x, Y = y)$$

Definition (Marginal Density Function)

For continuous r.v.s X and Y with density function $f_{X,Y}(x, y)$,

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

Independence

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

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

Independence

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.

Independence (Alternate method 1)

Lemma (Independence of random variables)

Two random variables are independent if and only if

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

i.e. you can replace the density with the CDF.

Conditional Functions

Definition (Conditional Probability Function)

The conditional probability function of X, **given** $Y = y$, is

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

Definition (Conditional Density Function)

The conditional density function of X, **given** $Y = y$, is

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

Conditional Functions

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

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

Conditional Functions

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

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

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

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

Independence (Alternate method 2)

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.

Conditional Expectation and Variance

Definition (Conditional Expectation)

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

Definition (Conditional Variance)

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

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

Conditional Expectation and Variance

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

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

Conditional Expectation and Variance

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

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

Conditional Expectation and Variance

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

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

Covariance

Let $\mathbb{E}[X] = \mu_X$ and $\mathbb{E}[Y] = \mu_y$.

Definition (Covariance)

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

Theorem (Covariance Formula)

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

Definition (Correlation)

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\text{SD}(X)\,\text{SD}(Y)} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\,\text{Var}(Y)}}$$

Covariance results

Theorem (Further properties of taking variances)

- $\mathrm{Var}(aX + bY) = a^2 \mathrm{Var}(X) + b^2 \mathrm{Var}(Y) + 2ab \,\mathrm{Cov}(X, Y)$
- $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2 \,\mathrm{Cov}(X, Y)$

Theorem (Properties of taking covariances)

- $\mathrm{Cov}(aX + bY, Z) = a^2 \mathrm{Cov}(X, Z) + b^2 \mathrm{Cov}(Y, Z)$
- $\mathrm{Cov}(X, aY + bZ) = a^2 \mathrm{Cov}(X, Y) + b^2 \mathrm{Cov}(X, Z)$
- $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$

Theorem (Consequence of zero covariance)

$$\mathrm{Cov}(X, Y) = 0 \iff \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

Working with the covariance - Definition

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

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

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

Working with the covariance - Definition

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 $\mathbb{E}[Y]$

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

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

Working with the covariance - Definition

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:

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

Working with the covariance - Definition

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 $\mathbb{E}[X]\mathbb{E}[Y]$

Working with the covariance - Formulae

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

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

Working with the covariance - Formulae

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

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

Working with the covariance - Formulae

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

Uncorrelatedness $\not\Rightarrow$ Independence

In general, the implication is one-sided.

Exception: X and Y are bivariate normal.

Having a hard time with formulas?

1. Know all the formulae for the single variable case
2. Know that $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$
3. All of the bivariate formulae stem from these

The Bivariate Transform (2901)

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

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

The Bivariate Transform (2901)

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

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

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

The Bivariate Transform (2901)

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

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

The Bivariate Transform (2901)

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 $v > 0$.

The Bivariate Transform (2901)

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:

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

Bivariate Transform in Sums (Continuous case) (2901)

Method:

1. Set $U = X + Y$ and $V = Y$
2. Apply the bivariate transform to find $f_{U,V}$
3. Compute the marginal density f_U

Convolutions

For random variables X and Y, let $Z = X + Y$.

Lemma (Discrete Convolution)

$$\mathbb{P}(Z = z) = \sum_y \mathbb{P}(X = z - y)\mathbb{P}(Y = y)$$

Lemma (Continuous Convolution)

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z - y) f_Y(y)\, dy$$

Convolutions

The hard part is (again) figuring what to sum/integrate over.

Working with convolutions

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

Working with convolutions

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.

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

Working with convolutions

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

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

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

Working with convolutions

Example

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

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

Working with convolutions

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, $z = x + y = 2, 3, 4, \ldots$

Working with convolutions

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.

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

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

Working with convolutions

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.

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.

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

Working with convolutions

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.

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

Working with convolutions

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.

Via Moment Generating Functions

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.

Via Moment Generating Functions

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.

Common Sums

For independent random variables:

- Sum of normal is normal - add means and variances
- Sum of n exponentials with the same parameter β is Gamma(n, β)
- Sum of Gamma with same second component is still Gamma - just add the first component
- Sum of Poisson is Poisson - add the parameter
- Sum of n Bernoullis with the same parameter p is Bin(n, p)
- Sum of Binomial with the same probability parameter p is still binomial - just add the first component

Modes of Convergence (2901)

Definition(Convergence Almost Surely)

$$X_n \stackrel{\text{a.s.}}{\to} X \iff \mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1$$

Definition (Convergence in Probability)

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

Definition (Convergence in Distribution)

$$X_n \stackrel{d}{\to} X \iff \lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

Convergence in Distribution Proof (2901)

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 $Y \sim \text{Exp}(1)$.

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

Convergence in Distribution Proof (2901)

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 $Y \sim \text{Exp}(1)$.

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

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

Convergence in Distribution Proof (2901)

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 $Y \sim \text{Exp}(1)$.

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

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

Convergence in Distribution Proof (2901)

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 $Y \sim \text{Exp}(1)$.

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

Convergence in Distribution Proof (2901)

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 $Y \sim \text{Exp}(1)$.

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

Convergence in Distribution Proof (2901)

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 $Y_n \xrightarrow{d} Y$.

Stronger forms of convergence

Lemma ('Strength' of convergence)

Almost sure convergence $\implies$ Convergence in $\mathbb{P}$ $\implies$ Convergence in d

Takeout for 2801

When using a theorem that says $\xrightarrow{\mathcal{D}}$, you can replace it with $\xrightarrow{P}$.

Law of Large Numbers

Lemma (Weak Law of Large Numbers)

For a sequence of i.i.d. r.v.s $X_1, \ldots, X_n$, with mean μ and finite variance σ^2,

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}} \mu$$

Lemma (Strong Law of Large Numbers)

For a sequence of i.i.d. r.v.s $X_1, \ldots, X_n$, with mean μ and finite variance σ^2,

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\text{a.s.}} \mu$$

Law of Large Numbers

For the interested reader: The strong law fails usually when your random variable is badly behaved.

Slutsky's Theorem

Theorem (Slutsky's Theorem)

Let $X_1, \ldots, X_n$ be a sequence of random variables with $X_n \xrightarrow{d} X$.

Let $Y_1, \ldots, Y_n$ be a sequence of random variables with $Y_n \xrightarrow{P} c$, where c is some constant. Then:

$$X_n + Y_n \xrightarrow{d} X + c$$

$$X_n Y_n \xrightarrow{d} cX$$

2801 note: Can replace $X_n \xrightarrow{\mathcal{D}} X$ with $X_n \xrightarrow{P} X$!

★ Central Limit Theorem ★

★ Theorem (CLT) ★

For a sequence of i.i.d. r.v.s $X_1, \dots, X_n$ with mean μ and finite variance σ^2

$$\frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1)$$

where $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$

(In the special case that the X_i's are normally distributed, the LHS **is** standard-normal distributed.)

Key property of the CLT

The actual distribution of $X_1, \dots, X_n$ does not matter.

Working with the CLT

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.

Working with the CLT

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 $X_1, \dots, X_{40}$ 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 $Z \sim \mathcal{N}(0, 1)$.

Working with the CLT

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?

$$\therefore \mathbb{P}(\overline{X_{40}} > 80) = \mathbb{P}\left(\frac{\overline{X_{40}} - 68}{4/\sqrt{40}} > \frac{80-68}{4/\sqrt{40}}\right)$$

$$\approx \mathbb{P}\left(Z > \frac{80-68}{4/\sqrt{40}}\right)$$

Working with the CLT

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?

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

Remark: Averages v.s. Sums

Earlier: CLT for averages.

If we consider $S = \sum_{i=1}^{n} X_i$, we have

$$\frac{S - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

We call this the CLT for sums.

Working with the CLT

Quick remark: Continuity correction for discrete random variables

- Not examinable for 2801
- Most likely not examinable either for 2901

Approximating a Binomial with a Normal

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

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

Approximating a Binomial with a Normal

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 $X \sim \text{Bin}\left(200, \frac{1}{5}\right)$.

Approximating a Binomial with a Normal

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,

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

Approximating a Binomial with a Normal

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,

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

Approximating a Binomial with a Normal

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,

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

Ending note for today

Whenever you find the probability/density function, always specify what range it's defined over!!!

Appendix: `R`

Some examples with $\text{Bin}(n, p)$:

- `dbinom(x, size=n, prob=p)` $= \mathbb{P}(X = x)$
- `pbinom(x, size=n, prob=p, lower.tail=TRUE)` $= \mathbb{P}(X \leq x)$
- `pbinom(x, size=n, prob=p, lower.tail=FALSE)` $= \mathbb{P}(X > x)$
- `qbinom(k, size=n, prob=p, lower.tail=TRUE)` $=$
  k-th quantile $=$ Solution to $\mathbb{P}(X \leq x) \leq k$

Some examples with $\mathcal{N}(\mu, \sigma^2)$

- `pnorm(x, mean=mu, sd=sigma, lower.tail=TRUE)` $= \mathbb{P}(X \leq x)$
- `qnorm(k, mean=mu, sd=sigma, lower.tail=TRUE)` $=$
  k-th quantile $=$ Solution to $\mathbb{P}(X \leq x) \leq k$

`rnorm(n, mean=mu, sd=sigma)` just randomly generates a bunch of values from $\mathcal{N}(\mu, \sigma^2)$ for you.