ST111/112
Probability A & B

Revision Guide

Written by Joy Tolia

WMS the Warwick Mathematics Society

Contents

Introduction

This revision guide for ST111/112 Probability A & B has been designed as an aid to revision, not a substitute for it. It contains a lot of theory and **NO EXAMPLES**, but the main point of Probability is not theory, **it's practice**. So, the best way to revise is to use this revision guide as a **quick reference** and just keep trying example done in lectures, example sheets and past exam questions. The recommended**textbook** by the lecturer is **"Probability (Springer Texts in Statistics)" by Pitman** so make sure you do as many examples as you can from this book. Ensure that you:

- **practise, practise, PRACTISE!!!**

Finally, good luck on the exam!

Disclaimer: Use at your own risk. No guarantee is made that this revision guide is accurate or complete, or that it will improve your exam performance, or make you 20% cooler. Use of this guide *will* increase entropy, contributing to the heat death of the universe.

Authors

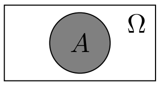
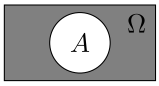
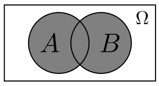
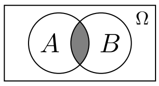
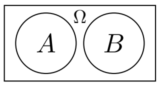
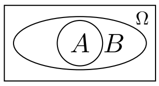
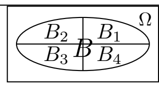
Originally written in 2015 by J. Tolia (j.tolia@warwick.ac.uk). Based upon lectures given by Sigurd Assing 2014-2015. Any corrections or improvements should be entered into our feedback form at http://tinyurl.com/WMSGuides (alternatively email revision.guides@warwickmaths.org).

Probability A

1 Distributions

To formally define probabilities, we take a sample space Ω, and assign probabilities to *events*, i.e. subsets of Ω:

Let us start with looking at the following terminology:

Event Language	Set Language	Set Notation	Venn Diagram
Outcome space or sample space	Whole set	Ω	
Event	Subset of Ω	A, B, C	
Impossible event	Empty set	$\varnothing$	
Not A, opposite of A	Complement of A	A^c or $\Omega \backslash A$	
A or B	Union of A and B	$A \cup B$	
A and B	Intersection of A and B	AB or $A \cap B$	
A and B are mutually exclusive	A and B are disjoint	$A \cap B = \phi$	
If A then B	A is a subset of B	$A \subseteq B$	
B partitioned into $B_1, \dots, B_n$	$B = B_1 \cup \dots \cup B_n, \quad B_i \cap B_j = \phi, \quad i \neq j$		

Axioms 1.1 (Axioms of Probability $\mathbb{P}$ for finite Ω)**.** The following can be thought of as rules for the map $\mathbb{P}$[1] $: \mathcal{P}(\Omega) \to [0, 1]$ where Ω is a finite set and $\mathcal{P}(\Omega)$ is the power set of Ω i.e. the collection of all subsets of Ω:

1. $\mathbb{P}[B] \geq 0, \quad \forall B \in \Omega.$
2. $\mathbb{P}[B] = \mathbb{P}[B_1] + \dots + \mathbb{P}[B_n]$ if B is partitioned into $B_1, \dots, B_n$.
3. $\mathbb{P}[\Omega] = 1$

Proposition 1.2. Let $A, B \subseteq \Omega$ then:

1. Law of Complements: $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
2. Difference Rule: Suppose $A \subseteq B$ then $\mathbb{P}[A] \leq \mathbb{P}[B]$ and $\mathbb{P}[B \setminus A] = \mathbb{P}[B] - \mathbb{P}[A]$.
3. Inclusion - Exclusion: $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$.
4. $\mathbb{P}[\varnothing] = 0$.

Proof. We prove these using the axioms of probability we stated earlier.

1. As $\Omega = A + A^c$, we have $1 = \mathbb{P}[\Omega] = \mathbb{P}[A] + \mathbb{P}[A^c]$ as A and A^c form a partition of Ω, finally we subtract $\mathbb{P}[A]$.

[1] We call $\mathbb{P}$ the probability distribution over Ω.

2. Let $A \subseteq B$, we have $B = A \cup (B \setminus A)$. Hence $\mathbb{P}[B] = \mathbb{P}[A] + \mathbb{P}[B \setminus A]$ as A and $B \setminus A$ form a partition of B and rearranging gives us $\mathbb{P}[B \setminus A] = \mathbb{P}[B] - \mathbb{P}[A]$. We have that $\mathbb{P}[B \setminus A] \geq 0$ which implies $\mathbb{P}[B] - \mathbb{P}[A] \geq 0$.
3. We can write $A \cup B = A \cup (B \setminus (A \cap B))$ which forms a partition, draw a Venn diagram if this is not clear. We have $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B \setminus (A \cap B)] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$ by the difference rule.
4. Take $A = \Omega$ then $A^c = \varnothing$ and use the law of complements.

□

Definition 1.3 (Conditional Probability)**.** Let $\mathbb{P}$ be an arbitrary probability distribution over Ω and $A, B \subseteq \Omega$ such that $\mathbb{P}[B] > 0$. We define the *conditional probability* of A given B by:

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}$$

Remark 1.4. When working on examples to do with conditional probabilities, it might be useful to use tree diagrams.

Lemma 1.5 (Multiplication Law)**.** From the definition of conditional probability we get the multiplication law:

$$\mathbb{P}[A \cap B] = \mathbb{P}[A|B] \cdot \mathbb{P}[B]$$

Theorem 1.6 (Law of Total Probability)**.** Let $B_1, \ldots, B_n$ form a partition of Ω with $\mathbb{P}[B_i] > 0$ for all $i = 1, \ldots, n$. Then for $A \subset \Omega$:

$$\mathbb{P}[A] = \sum_{i=1}^{n} \mathbb{P}[A|B_i] \cdot \mathbb{P}[B_i]$$

Proof. First, we see that the sets $(A \cap B_1), \ldots, (A \cap B_n)$ form a partition of A. So by the second statement in the axiom, we get $\mathbb{P}[A] = \sum_{i=1}^{n} \mathbb{P}[A \cap B_i]$. Finally, using the multiplication law to get $\mathbb{P}[A \cap B_i] = \mathbb{P}[A|B_i] \cdot \mathbb{P}[B_i]$ for all $i = 1, \ldots, n$, we have $\mathbb{P}[A \cap B_i] = \sum_{i=1}^{n} \mathbb{P}[A|B_i] \cdot \mathbb{P}[B_i]$.

□

Theorem 1.7 (Baye's Rule)**.** Let $A \subseteq \Omega$ be an event and $B_1, \ldots, B_n$ form a partition of Ω. Then for all $i = 1, \ldots n$:

$$\mathbb{P}[B_i|A] = \frac{\mathbb{P}[B_i] \cdot \mathbb{P}[A|B_i]}{\mathbb{P}[B_1] \cdot \mathbb{P}[A|B_1] + \cdots + \mathbb{P}[B_n] \cdot \mathbb{P}[A|B_n]} = \frac{\mathbb{P}[B_i] \cdot \mathbb{P}[A|B_i]}{\sum_{j=1}^{n} \mathbb{P}[B_j] \cdot \mathbb{P}[A|B_j]}$$

Definition 1.8 (Independence)**.** Let $A, B \subseteq \Omega$ be two events. We say that A and B are *independent events* if:

$$\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B]$$

2 Repeated Trials and Sampling

Definition 2.1 (Independence of multiple events)**.** Suppose we have events $A_1, \ldots, A_n \subseteq \Omega$, we say these events are *independent* if:

$$\mathbb{P}[A_{i_1} \cap \cdots \cap A_{i_k}] = \mathbb{P}[A_{i_1}] \times \cdots \times \mathbb{P}[A_{i_k}]$$

for all $1 \leq i_1 < \cdots < i_k \leq n$ and $k = 1, \ldots, n$.

Definition 2.2. Suppose Ω is partitioned into 2 sets, i.e. $\Omega = \{0, 1\}$ or $\Omega = \{A, A^c\}$, where the event A can be thought of as success and A^c can be thought of as failure. Then the *Bernoulli*(p) distribution over Ω where $p \in [0, 1]$ is given by:

$$\mathbb{P}[A] = p, \quad \mathbb{P}[A^c] = 1 - \mathbb{P}[A] = 1 - p$$

Proposition 2.3. Suppose we are looking at Bernoulli(p) distribution then we have the following:

$$\mathbb{P}\left[k \text{ successes in } n \text{ trials}\right] = \binom{n}{k} p^k (1-p)^{n-k}$$

where $\binom{n}{k} = \frac{n!}{(n-k)!k!}$.

Definition 2.4. The distribution $\mathbb{P}$ over $\Omega = \{0, 1, \ldots, n\}$ satisfying

$$\mathbb{P}\left[\{k\}\right] = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, \ldots, n$$

is called the *binomial*(n, p) distribution

Definition 2.5. The distribution $\mathbb{P}$ over $\Omega = \{0, 1, \ldots, n\}$ satisfying

$$\mathbb{P}\left[\{g\}\right] = \frac{\binom{G}{g}\binom{N-G}{n-g}}{\binom{N}{n}}, \quad g = 0, \ldots, n$$

for some $n, G, N \in \mathbb{N}$ such that $n \leq N, G \leq N$, is called *hypergeometric*(n, N, G) distribution.

3 Random Variables

We continue assuming that Ω is a finite outcome space throughout this section.

Definition 3.1 (Random Variable)**.** A function $X : \Omega \to \mathbb{R}$ is called a *random variable.* The *range* of X is the set of all possible values X can take so $\text{range}(X) := \{X(\omega) : \omega \in \Omega\}$.

Proposition 3.2. Let $X : \Omega \to \mathbb{R}$ be a random variable and $\mathbb{P}$ be a distribution over Ω. Then following map defines a distribution over range(X), let $B \subseteq \text{range}(X)$:

$$B \mapsto \mathbb{P}\left[X \in B\right] := \mathbb{P}\left[\{\omega \in \Omega : X(\omega) \in B\}\right]$$

This distribution is called the distribution of the random variable X.

Proof. We have to show that the distribution of the random variable X follows the three statement in the axiom of probability distribution.

1. $\mathbb{P}\left[X \in B\right] = \mathbb{P}\left[\{\omega \in \Omega : X(\omega) \in B\}\right] \geq 0$ for all $B \subseteq \text{range}(X)$, as $\mathbb{P}$ itself is a distribution on Ω.
2. Let $B_1, \ldots, B_n$ be a partition of $B \in \text{range}(X)$ then $\{\omega \in \Omega : X(\omega) \in B_1\}, \ldots, \{\omega \in \Omega : X(\omega) \in B_n\}$ is a partition of $\{\omega \in \Omega : X(\omega) \in B\}$, therefore:

$$\begin{aligned} \mathbb{P}\left[X \in B\right] &= \mathbb{P}\left[\{\omega \in \Omega : X(\omega) \in B\}\right] \\ &= \mathbb{P}\left[\{\omega \in \Omega : X(\omega) \in B_1\}\right] + \cdots + \mathbb{P}\left[\{\omega \in \Omega : X(\omega) \in B_n\}\right] \\ &= \mathbb{P}\left[X \in B_1\right] + \cdots + \mathbb{P}\left[X \in B_n\right] \end{aligned}$$

3. $\mathbb{P}\left[X \in \text{ range}(X)\right] = \mathbb{P}\left[\{\omega \in \Omega : X(\omega) \in \text{ range}(X)\}\right] = 1$ as X is a function from Ω.

□

Remark 3.3. The distribution of the random variable X is defined by $\mathbb{P}\left[X = x\right]$ where $x \in \text{range}(X)$ because

$$\mathbb{P}\left[X \in B\right] = \sum_{x \in B} \mathbb{P}\left[X = x\right], \quad \forall B \subseteq \text{range}(X)$$

Example 3.4. Let us look at the random variable on the outcome space $\Omega = \{\text{heads,tails}\}$ then define X by $X(\text{heads}) = 1$, $X(\text{tails}) = 0$, hence $\text{range}(X) = \{0, 1\}$. Suppose we are looking at a fair coin then $\mathbb{P}\left[\{heads\}\right] = 0.5 = \mathbb{P}\left[\{tails\}\right]$. Then $\mathbb{P}\left[X = 1\right] = \mathbb{P}\left[\{\omega \in \Omega : X(\omega) = 1\}\right] = \mathbb{P}\left[\{heads\}\right] = 0.5$. Similarly $\mathbb{P}\left[X = 0\right] = \mathbb{P}\left[\{\omega \in \Omega : X(\omega) = 0\}\right] = \mathbb{P}\left[\{tails\}\right] = 0.5$.

Lemma 3.5. Let X be a random variable and $f : \mathbb{R} \to \mathbb{R}$ be a function then $Z = f(X)$ is also a random variable.

Remark 3.6. The distribution of $Z = f(X)$ is determined by:

$$\mathbb{P}\left[Z = z\right] = \mathbb{P}\left[f(X) = z\right] = \sum_{x \in \text{range}(X), z = f(x)} \mathbb{P}\left[X = x\right], \quad z \in \text{range}(Z)$$

3.1 Joint Distribution

Definition 3.7. Let $X, Y : \Omega \to \mathbb{R}$ be random variables and $\mathbb{P}$ be a distribution over Ω. We define the range$(X, Y) := \{(x, y) : x \in \text{range}(X), y \in \text{range}(Y)\}$. The *joint distribution* of (X, Y) is determined by $p : \text{range}\,(X, Y) \to [0, 1]$ with:

$$p\,(x, y) := \mathbb{P}\,[X = x, Y = y]\,, \quad (x, y) \in \text{range}\,(X, Y)$$

where $\mathbb{P}\,[X = x, Y = y] := \mathbb{P}\,[\{\omega \in \Omega : X\,(\omega) = x \text{ and } Y\,(\omega) = y\}]$.

Lemma 3.8. Let (X, Y) be a pair of random variables whose distribution is given by $p\,(x, y)$ for $x \in \text{range}(X), y \in \text{range}(Y)$. Then the *marginal distributions* of X and Y are given respectively by:

$$\mathbb{P}\,[X = x] = \sum_{y \in \text{range}(Y)} p\,(x, y)$$

$$\mathbb{P}\,[Y = y] = \sum_{x \in \text{range}(X)} p\,(x, y)$$

Proof. We have that $\cup_{y \in \text{range}(Y)} \{X = x, Y = y\}$ forms a partition of $\{X = x\}$, it may be useful for you to write these sets in terms of $\omega \in \Omega$. Now using the second statement in the axiom of probability, we get:

$$\mathbb{P}\,[X = x] = \mathbb{P}\left[\bigcup_{y \in \text{range}(Y)} (X = x, Y = y)\right] = \sum_{y \in \text{range}(Y)} \mathbb{P}\,[X = x, Y = y] = \sum_{y \in \text{range}(Y)} p\,(x, y)$$

□

Definition 3.9. The following two definitions give us intuition on the different ways two random variables can be similar.

1. Let $X : \Omega_1 \to \mathbb{R}$ be a random variable with $\mathbb{P}_1$ be a distribution over Ω_1 and let $Y : \Omega_2 \to \mathbb{R}$ be a random variable with $\mathbb{P}_2$ be a distribution over Ω_2. If range$(X) =$ range(Y) and $\mathbb{P}_1\,[X = x] = \mathbb{P}_2\,[Y = x]$ for all $x \in \text{range}(X)$ then we say that X and Y have the *same distribution*.
2. Let $X, Y : \Omega \to \mathbb{R}$ be random variables on the same outcome space Ω and $\mathbb{P}$ is a probability distribution over Ω. If we have the following:

$$1 = \mathbb{P}\,[X = Y] := \mathbb{P}\,[\{\omega \in \Omega : X\,(\Omega) = Y\,(\omega)\}]$$

then we say that X and Y are *equal* and write $X = Y$.

Definition 3.10. The following are probability of events determined by random variables $X, Y : \Omega \to \mathbb{R}$ where $\mathbb{P}$ is the probability distribution on Ω. We omit writing $x \in \text{range}(X)$ and $y \in \text{range}(Y)$ and just write x and y instead in the following statements.

1. $\mathbb{P}\,[X < Y] := \sum_{x<y} p\,(x, y) = \sum_x \sum_{y:y>x} p\,(x, y)$.
2. $\mathbb{P}\,[X = Y] := \sum_{x=y} p\,(x, y) = \sum_x p\,(x, x)$.
3. $\mathbb{P}\,[X + Y = z] := \sum_{x+y=z} p\,(x, y) = \sum_x p\,(x, z - x)$.
4. $\mathbb{P}\,[g\,(X, Y) = z] := \sum_{g(x,y)=z} p\,(x, y) = \sum_x \sum_{y:g(x,y)=z} p\,(x, y)$ for any $g : \mathbb{R}^2 \to \mathbb{R}$.

Definition 3.11 (Conditional Distribution of Random Variables)**.** Let (X, Y) be a pair of random variables and $\mathbb{P}$ be a distribution on Ω. Given $A \subseteq \Omega$, the following defines the *conditional distribution of Y given A*:

$$\mathbb{P}\,[Y \in B|A]\,, \quad B \subseteq \text{range}(Y)$$

The conditional distribution is determined by $\mathbb{P}\,[Y = y|A]$ for $y \in \text{range}(Y)$.

Remark 3.12. We can write the event $\{X = x, Y = y\}$ as $\{X = x\} \cap \{Y = y\}$ so using the definition of conditional probability from the first section we have:

$$p\,(x, y) = \mathbb{P}\,[X = x, Y = y] = \mathbb{P}\,[X = x] \cdot \mathbb{P}\,[Y = y|X = x]$$

Definition 3.13 (Independence of Random Variables)**.** Let $X, Y \Omega \to \mathbb{R}$ be random variables and $\mathbb{P}$ be the probability distribution on Ω. We say X and Y are *independent* if and only if for all $x \in \text{range}(X), y \in \text{range}(Y)$ we have:

$$\mathbb{P}[X = x, Y = y] = \mathbb{P}[X = x] \cdot \mathbb{P}[Y = y]$$

Remark 3.14. If X and Y are independent then we have for all $A \subseteq \text{range}(X), B \subseteq \text{range}(Y)$, we have $\mathbb{P}[X \in A, Y \in B] = \mathbb{P}[X \in A] \cdot \mathbb{P}[Y \in B]$.

Remark 3.15 (Joint distributions of more than two random variables)**.** Let $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ be random variables. The joint distribution of these random variables is determined by $p : \text{range}(X_1) \times \cdots \times \text{range}(X_n) \to [0, 1]$ satisfying

$$\sum_{x_1 \in \text{range}(X_1)} \cdots \sum_{x_n \in \text{range}(X_n)} p(x_1, \ldots, x_n) = 1$$

where $p(x_1, \ldots, x_n) := \mathbb{P}[X_1 = x_1, \ldots, X_n = x_n]$ for all $x_1 \in \text{range}(X_1), \ldots, x_n \in \text{range}(X_n)$. Finally we say $X_1, \ldots, X_n$ are independent if and only if for all $x_1 \in \text{range}(X_1), \ldots, x_n \in \text{range}(X_n)$, we have:

$$p(x_1, \ldots, x_n) = \mathbb{P}[X_1 = x_1] \times \cdots \times \mathbb{P}[X_n = x_n]$$

3.2 Expectation

Definition 3.16. Let $X : \Omega \to \mathbb{R}$ be a random variable and $\mathbb{P}$ be a distribution over Ω, we define the real number $\mathbb{E}[X]$, known as the *expectation* of X by:

$$\mathbb{E}[X] = \sum_{x \in \text{range}(X)} x \cdot \mathbb{P}[X = x]$$

Example 3.17. Let us look at the random variable X which has $\text{range}(X) = \{0, 1\}$ and $\mathbb{P}[X = 1] = 0.5$ and $\mathbb{P}[X = 0] = 0.5$. We constructed this random variable from a outcome space of a fair coin. The expectation of this random variable is:

$$\mathbb{E}[X] = \sum_{x \in \text{range}(X)} x \cdot \mathbb{P}[X = x] = 1 \cdot \mathbb{P}[X = 1] + 0 \cdot \mathbb{P}[X = 0] = 0.5$$

Example 3.18 (Indicators)**.** Let Ω be an outcome space and $\mathbb{P}$ be the probability distribution on Ω. Fix $A \subseteq \Omega$. We define the *indicator random variable* $\mathbb{1}_A : \Omega \to \mathbb{R}$ by:

$$\mathbb{1}_A(\omega) = \begin{cases} 1 \text{ if } \omega \in A \\ 0 \text{ otherwise} \end{cases}$$

Note that $\text{range}(\mathbb{1}_A) = \{0, 1\}$ so:

$$\mathbb{E}[\mathbb{1}_A] = 0 \cdot \mathbb{P}[\mathbb{1}_A = 0] + 1 \cdot \mathbb{P}[\mathbb{1}_A = 1] = \mathbb{P}[\{\omega \in \Omega : \mathbb{1}_A(\omega) = 1\}] = \mathbb{P}[\{\omega \in A\}] = \mathbb{P}[A]$$

Theorem 3.19 (Functions of Random Variables)**.** Let $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ be random variables and $\mathbb{P}$ be a distribution on Ω. For any $g : \mathbb{R}^n \to \mathbb{R}$, we have:

$$\mathbb{E}[g(X_1, \ldots, X_n)] = \sum_{x_1 \in \text{range}(X_1)} \cdots \sum_{x_n \in \text{range}(X_n)} g(x_1, \ldots, x_n) \cdot \mathbb{P}[X_1 = x_1, \ldots, X_n = x_n]$$

Proof. We prove this for the case $n = 1$. Let X be a random variable, by definition:

$$\mathbb{E}[g(X)] = \sum_{y \in \text{range}(g(X))} y \cdot \mathbb{P}[g(X) = y]$$

where

$$y \cdot \mathbb{P}[g(X) = y] = y \cdot \sum_{x : g(x) = y} \mathbb{P}[X = x] = \sum_{x : g(x) = y} g(x) \cdot \mathbb{P}[X = x]$$

finally:

$$\mathbb{E}\left[g\left(X\right)\right] = \underbrace{\sum_{y \in \text{range}(g(X))} \sum_{x : g(x) = y}}_{\sum_{x \in \text{range}(X)}} g(x) \cdot \mathbb{P}\left[X = x\right]$$

□

Theorem 3.20 (Addition Rule)**.** Let $X_1, \ldots, X_n$ be random variables then:

$$\mathbb{E}\left[X_1 + \cdots + X_n\right] = \mathbb{E}\left[X_1\right] + \ldots \mathbb{E}\left[X_n\right]$$

Proof. We prove this for case $n =$. Let (X, Y) be a pair of random variables, with joint distribution function $p\left(x, y\right)$, $x \in \text{range}(X)$ and $y \in \text{range}(Y)$. We have $\mathbb{P}\left[X = x\right] = \sum_{y \in \text{range}(Y)} p(x, y)$ and $\mathbb{P}\left[Y = y\right] = \sum_{x \in \text{range}(X)} p(x, y)$. Therefore we have:

$$\mathbb{E}\left[X\right] = \sum_{x \in \text{range}(X)} x \cdot \mathbb{P}\left[X = x\right] = \sum_{x \in \text{range}(X)} \sum_{y \in \text{range}(Y)} x \cdot p(x, y)$$

Similarly $\mathbb{E}\left[Y\right] = \sum_{y \in \text{range}(Y)} \sum_{x \in \text{range}(X)} y \cdot p(x, y)$. Now applying the formula from previous theorem with $g(x, y) = x + y$, and since we can interchange the order of finite sums, we get:

$$\begin{aligned}
\mathbb{E}\left[X + Y\right] &= \sum_{x \in \text{range}(X)} \sum_{y \in \text{range}(Y)} (x + y) \cdot p(x, y) \\
&= \sum_{x \in \text{range}(X)} \sum_{y \in \text{range}(Y)} x \cdot p(x, y) + \sum_{y \in \text{range}(Y)} \sum_{x \in \text{range}(X)} y \cdot p(x, y) \\
&= \mathbb{E}\left[X\right] + \mathbb{E}\left[Y\right]
\end{aligned}$$

□

Theorem 3.21 (Expectation of Product of Independent Random Variables)**.** Let $X_1, \ldots, X_n$ be independent random variables then:

$$\mathbb{E}\left[X_1 \times \cdots \times X_n\right] = \mathbb{E}\left[X_1\right] \times \cdots \times \mathbb{E}\left[X_n\right]$$

Proof. We prove this for case $n =$. Let (X, Y) be a pair of independent random variables. Then:

$$\begin{aligned}
\mathbb{E}\left[X \times Y\right] &= \sum_{x \in \text{range}(X)} \sum_{y \in \text{range}(Y)} xy \cdot \underbrace{\mathbb{P}\left[X = x, Y = y\right]}_{= \mathbb{P}[X = x] \times \mathbb{P}[Y = y] \text{ by independence}} \\
&= \sum_{x \in \text{range}(X)} \sum_{y \in \text{range}(Y)} \left(x\mathbb{P}\left[X = x\right]\right)\left(y\mathbb{P}\left[Y = y\right]\right) \\
&= \left(\sum_{x \in \text{range}(X)} x\mathbb{P}\left[X = x\right]\right) \times \left(\sum_{y \in \text{range}(Y)} y\mathbb{P}\left[Y = y\right]\right) \\
&= \mathbb{E}\left[X\right] \times \mathbb{E}\left[Y\right]
\end{aligned}$$

□

Theorem 3.22 (Markov Inequality)**.** Let $X : \Omega \to [0, \infty)$ be a random variable and $\mathbb{P}$ be a distribution on Ω. Then for any $a > 0$, we have:

$$\mathbb{P}\left[X \geq a\right] = \frac{\mathbb{E}\left[X\right]}{a}$$

Proof. As $\text{range}(X) = \{x \in \mathbb{R} : x \geq 0\}$:

$$\mathbb{E}\left[X\right] = \sum_{x \geq 0} x \cdot \mathbb{P}\left[X = x\right] \geq \sum_{x \geq a} x \cdot \mathbb{P}\left[X = x\right] = \sum_{x \geq a} a \cdot \mathbb{P}\left[X = x\right] = a \cdot \mathbb{P}\left[X \geq a\right]$$

□

Probability B

4 Weak Law of Large Numbers

Theorem 4.1 (Weak Law of Large Numbers)**.** let $\mathbb{P}$ be probability a distribution on Ω, where Ω is finite. Let $X_i : \Omega \to \mathbb{R}$, $i \in \mathbb{N}$ be a sequence of random variables such that:

1. $X_1, \ldots, X_n$ are independent for all $n \in \mathbb{N}$
2. $\mathbb{E}[X_i] = \mathbb{E}[X_1]$ for all $i \in \mathbb{N}$
3. $\sup_{i \geq i} \mathbb{E}\left[X_i^2\right] < \infty$

Then, for all $\epsilon > 0$, we have:

$$\lim_{n \to \infty} = \mathbb{P}\left[\left|\frac{X_1 + \cdots + X_n}{n} - \mathbb{E}[X_1]\right| \geq \epsilon\right] = 0$$

5 Standard Deviation and Variance

In this section, let $X : \Omega \to \mathbb{R}$ be a random variable and let $\mathbb{P}$ be a distribution on Ω, we still hold the assumption that Ω is finite from Probability A. We use μ to denote the expected value of X:

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

Definition 5.1. 1. The *variance* of X, denoted by $\operatorname{Var}[X]$, is defined as the following:

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

2. The *Standard deviation* of X, denotes by $\operatorname{SD}[X]$ or σ, is defined as the following:

$$\operatorname{SD}[X] = \sqrt{\operatorname{Var}[X]}$$

Remark 5.2. We have that $\operatorname{Var}[X] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$ by the following:

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

Theorem 5.3 (Chebychev's Inequality)**.** Let X be a random variable and $k > 0$, then:

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq k \times \operatorname{SD}[X]] \leq \frac{1}{k^2}$$

Proof.

$$\begin{aligned}
\mathbb{P}[|X - \mathbb{E}[X]| \geq k \times \operatorname{SD}[X]] &= \mathbb{P}\left[|X - \mathbb{E}[X]|^2 \geq k^2 \times \operatorname{Var}[X]\right] \\
&\leq \frac{\mathbb{E}\left[(X - \mathbb{E}[X])^2\right]}{k^2 \times \operatorname{Var}[X]} \\
&= \frac{1}{k^2}
\end{aligned}$$

□

Proposition 5.4. Let $X_1, \ldots, X_n$ be independent random variables. Then:

$$\operatorname{Var}[X_1, \ldots, X_n] = \sum_{i=1}^{n} \operatorname{Var}[X_i]$$

Proof. We prove the proposition for $n = 2$, let $\mu_1 = \mathbb{E}[X_1], \mu_2 = \mathbb{E}[X_2]$. Then:

$$\begin{aligned}
\operatorname{Var}[X_1 + X_2] &= \mathbb{E}\left[(X_1 + X_2 - (\mu_1 - \mu_2))^2\right] \\
&= \mathbb{E}\left[((X_1 - \mu_1) + (X_2 - \mu_2))^2\right] \\
&= \mathbb{E}\left[(X_1 - \mu_1)^2 - 2(X_1 - \mu_1)(X_2 - \mu_2) + (X_2 - \mu_2)^2\right] \\
&= \mathbb{E}\left[(X_1 - \mu_1)^2\right] - 2\mathbb{E}[X_1 - \mu_1] \cdot \mathbb{E}[X_2 - \mu_2] + \mathbb{E}\left[(X_2 - \mu_2)^2\right] \\
&= \operatorname{Var}[X_1] + 0 + \operatorname{Var}[X_2]
\end{aligned}$$

□

Corollary 5.5. If $X_1, \ldots, X_n$ are independent random variables such that $\operatorname{Var}[X_i] = \operatorname{Var}[X_1]$ for all $i = 1, \ldots, n$, then:

$$\operatorname{Var}[X_1 + \cdots + X_n] = n \cdot \operatorname{Var}[X_1]$$

Proposition 5.6. Let X be a random variable and $c \in \mathbb{R}$, then:

$$\operatorname{Var}[cX] = c^2 \operatorname{Var}[X]$$

Proof. As we have $\mathbb{E}[cX] = c \cdot \mathbb{E}[X] = c\mu$, so:

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

□

6 Normal Approximation

Definition 6.1.

1. The *probability density function*, denoted by ϕ, of the *standard normal distribution* is given by:

$$\phi(y) := \frac{1}{2\pi} \exp\left(-\frac{y^2}{2}\right)$$

2. The *cumulative distribution function*, denoted by Φ, of the standard normal distribution is given by:

$$\Phi(x) := \int_{-\infty}^{x} \phi(y)\, dy$$

We can use the normal distribution to approximate different distributions, such as the binomial distribution. The normal distribution is characterised by its mean and variance, hence if we can find the mean and the variance or standard deviation of a given distribution then we can approximate it with the normal distribution.

Suppose we are looking at a random variable S_n which is binomially distributed with binomial(n, p), then $\mu = \mathbb{E}[S_n] = np$ and $\sigma = \operatorname{SD}[S_n] = \sqrt{np(1-p)}$. Then for $0 \leq k_a \leq k_b \leq n$:

$$\mathbb{P}[k_a \leq S_n \leq k_b] \approx \Phi\left(\frac{k_b + \frac{1}{2} - \mu}{\sigma}\right) - \Phi\left(\frac{k_a - \frac{1}{2} - \mu}{\sigma}\right)$$

Remark 6.2 (Symmetry for Φ)**.** Given $z > 0$ we have that $\Phi(-z) = 1 - \Phi(z)$. This can be seen by looking at a graph of the probability density function of the standard normal distribution $\phi(x)$ as it is symmetric around $x = 0$.

7 Poisson Distribution

Definition 7.1. Let $X : \Omega \to \{0, 1, 2, \dots\}$ be a random variable and $\mathbb{P}$ a distribution over Ω such that, given $mu > 0$:

$$\mathbb{P}[X = k] = \frac{e^{-\mu}\mu^k}{k!}, \quad k = 0, 1, 2, \dots$$

Then we say that X has a *Poisson(μ) distribution.*

Remark 7.2. For X from the above definition to be a random variable, Ω has to be at least countably infinite.

From now on we can assume that if Ω is countably infinite i.e. $\Omega = \{\omega_1, \omega_2, \dots\}$ and $p\Omega \to [0, 1]$ satisfies $\sum_{k=1}^{\infty} p(\omega_k) = 1$ then $\mathbb{P} : \mathcal{P} \to [0, 1]$ satisfies Axioms 1.1, where $\mathbb{P}$ is given by:

$$\mathbb{P}[B] = \sum_{\omega \in B} p(\omega), \quad B \subseteq \Omega$$

For a random variable X defined in Definition 7.1, we calculate the expected value using the following:

$$\mathbb{E}[X] = \sum_{x \in \text{range}(X)} x \times \mathbb{P}[X = x] = \sum_{k=0}^{\infty} k \times \mathbb{P}[X = k]$$

Proposition 7.3. Let X be a random variable that is distribution Poisson(μ), then the $\mathbb{E}[X] = \mu$ and $\text{Var}[X] = \mu$.

Proof.

$$\begin{aligned}
\mathbb{E}[X] &= \lim_{N\to\infty} \sum_{k=0}^{N} k \times \mathbb{P}[X = k] \\
&= \lim_{N\to\infty} \sum_{k=1}^{N} k \cdot \frac{e^{-\mu}\mu^k}{k!} \\
&= \mu e^{-\mu} \lim_{N\to\infty} \sum_{k=1}^{N} \frac{\mu^{k-1}}{(k-1)!} \\
&= \mu e^{-\mu} \lim_{N\to\infty} \sum_{k=0}^{N-1} \frac{\mu^k}{k!} \\
&= \mu e^{-\mu} e^{\mu} = \mu
\end{aligned}$$

$$\begin{aligned}
\text{Var}[X] &= \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 \\
&= \lim_{N\to\infty} \sum_{k=1}^{N} k^2 \cdot \frac{e^{-\mu}\mu^k}{k!} - \mu^2 \\
&= \lim_{N\to\infty} \left[e^{-\mu} \sum_{k=2}^{N} k(k-1) \frac{\mu^k}{k!} \underbrace{+ e^{-\mu} \sum_{k=1}^{N} k \cdot \frac{\mu^k}{k!}}_{=\mu} \right] - \mu^2 \\
&= \mu^2 e^{-\mu} \lim_{N\to\infty} \left[\sum_{k=2}^{N} \frac{\mu^{k-2}}{(k-2)!} \right] + \mu - \mu^2 \\
&= \mu^2 e^{-\mu} \lim_{N\to\infty} \left[\sum_{k=0}^{N-2} \frac{\mu^k}{k!} \right] + \mu - \mu^2 \\
&= \mu^2 e^{-\mu} e^{\mu} + \mu - \mu^2 = \mu
\end{aligned}$$

□

8 Continuous Random Variables

8.1 Probability Densities

Definition 8.1. Let $\mathbb{P}$ be a distribution over Ω, where Ω is uncountably infinite. Let $X : \Omega \to \mathbb{R}$, where range (X) includes whole intervals. The *probability density function (pdf)*, $f : \mathbb{R} \to \mathbb{R}$, of X is given by:

$$\mathbb{P}[a \leq X \leq b] = \int_a^b f(x)dx$$

with the properties:

1. f is piecewise continuous
2. $$f(x) \geq 0, \quad \forall x \in \mathbb{R}$$
3. $$\int_{-\infty}^{\infty} f(x)dx = 1$$

Proposition 8.2. Let f be a probability density function of $X : \Omega \to \mathbb{R}$, $\mathbb{P}$ be a distribution over Ω. Suppose $\tilde{f} : \mathbb{R} \to \mathbb{R}$ is piecewise continuous such that $\tilde{f}(x) \geq 0, \forall x \in \mathbb{R}$, $\int_{-\infty}^{\infty} \tilde{f}(x)\, dx = 1$ and:

$$\int_a^b \tilde{f}(x)\, dx = \int_a^b f(x)\, dx, \quad \forall a \leq b$$

Then $\tilde{f}(x) = f(x)$ for all $x \in \mathbb{R}$ such that f and $\tilde{f}$ are continuous at x.

Proof. Let $x \in \mathbb{R}$ such that f and $\tilde{f}$ are continuous at x, then we have for all small enough $h > 0$:

$$\int_x^{x+h} \tilde{f}(u)\, du = \int_x^{x+h} f(u)\, du$$

This implies:

$$\tilde{f}(x) = \lim_{h \downarrow 0} \int_x^{x+h} \tilde{f}(u)\, du = \lim_{h \downarrow 0} \int_x^{x+h} f(u)\, du = f(x)$$

□

Definition 8.3. Let $X : \Omega \to \mathbb{R}$ be a random variable and $\mathbb{P}$ a distribution over Ω such that $f : \mathbb{R} \to \mathbb{R}$ is the probability density function of X. Assume that

$$\int_{-\infty}^{\infty} |x|\, f(x)\, dx < \infty$$

Then, the *expectation* and *variance* of X are defined by:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} xf(x)\, dx = \lim_{N \to \infty} \int_{-N}^{N} xf(x)\, dx$$

$$\mathrm{Var}[X] = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 f(x)\, dx = \lim_{N \to \infty} \int_{-N}^{N} (x - \mathbb{E}[X])^2 f(x)\, dx$$

Remark 8.4.

1. All properties we know about expectation and variance still hold.
2. If $g : \mathbb{R} \to \mathbb{R}$ is a continuous function and $\int_{-\infty}^{\infty} |g(x)|\, f(x)dx < \infty$ where f is a probability density function. Then:
$$\mathbb{E}[g(x)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

8.2 Uniform Distribution

Definition 8.5. Let $X : \Omega \to \mathbb{R}$ be a random variable and $\mathbb{P}$ a distribution over Ω such that:

$$\mathbb{P}\left[a \leq X \leq b\right] = \int_a^b f(x)dx, \quad \forall a \leq b$$

where

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

for some real numbers $a \leq b$. Then we say that X has a *uniform*(a, b) *distribution.*

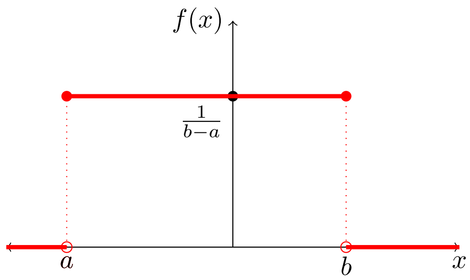


Figure 1: Probability Density Function of Uniform(a,b) Distribution

Proposition 8.6. Let $X : \Omega \to \mathbb{R}$ be a random variable and $\mathbb{P}$ a distribution over Ω such that X has uniform(a,b) distribution. Then:

$$\mathbb{E}\left[X\right] = \frac{a+b}{2}, \quad \mathrm{Var}\left[X\right] = \frac{(b-a)^2}{12}$$

Proof.

$$\mathbb{E}\left[X\right] = \int_a^b \frac{x}{b-a}dx = \frac{b^2-a^2}{2(b-a)} = \frac{(b+a)(b-a)}{2(b-a)} = \frac{a+b}{2}$$

$$\begin{aligned} \mathrm{Var}\left[X\right] &= \mathbb{E}\left[X^2\right] - \mathbb{E}\left[X\right]^2 \\ &= \int_a^b \frac{x^2}{b-a}dx - \frac{a^2+2ab+b^2}{4} \\ &= \frac{b^3-a^3}{3(b-a)} - \frac{a^2+2ab+b^2}{4} \\ &= \frac{(b-a)\left(b^2+ab+a^2\right)}{3(b-a)} - \frac{a^2+2ab+b^2}{4} \\ &= \frac{4b^2+4ab+4a^2-3a^2-6ab-3b^2}{12} \\ &= \frac{(b-a)^2}{12} \end{aligned}$$

□

8.3 Normal Distribution

Definition 8.7. Let $Z : \Omega \to \mathbb{R}$ be a random variable and $\mathbb{P}$ a distribution over Ω such that:

$$\mathbb{P}\left[a \leq Z \leq b\right] = \int_a^b \phi(y)dy, \quad \forall a \leq b$$

where ϕ is a probability density function defined in Definition 6.1. Then we call Z a standard normal random variable and Z has a *normal*$(0, 1)$ *distribution.*

Definition 8.8. Let $X : \Omega \to \mathbb{R}$ be a random variable and $\mathbb{P}$ a distribution over Ω such that:

$$\mathbb{P}\left[a \leq X \leq b\right] = \int_a^b f(x)dx, \quad \forall a \leq b$$

where

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}$$

for some $\mu \in \mathbb{R}$ and $\sigma > 0$. Then X is said to have a *normal*$\left(\mu, \sigma^2\right)$ *distribution.*

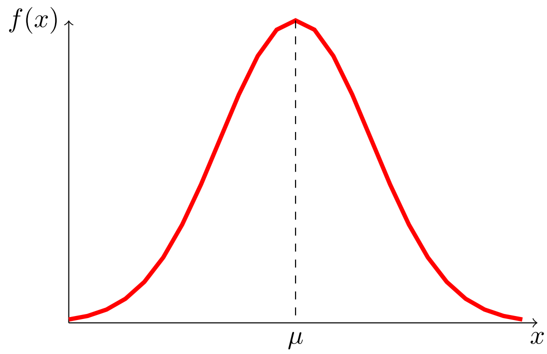


Figure 2: Probability Density Function of Normal$\left(\mu, \sigma^2\right)$ Distribution

Proposition 8.9. Let Z and X be random variables with normal(0,1) and normal$\left(\mu, \sigma^2\right)$ distributions, respectively. Then $X = \mu + \sigma Z$.

Proof. Fix $a \leq b$, we want to show $\mathbb{P}\left[a \leq X \leq b\right] = \mathbb{P}\left[a \leq \mu + \sigma Z \leq b\right]$:

$$\begin{aligned}
\mathbb{P}\left[a \leq \mu + \sigma Z \leq b\right] &= \mathbb{P}\left[\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right] \\
&= \int_{(a-\mu)/\sigma}^{(b-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \\
\text{Substitute } x = \mu + \sigma y &= \int_a^b \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \frac{dx}{\sigma} \\
&= \mathbb{P}\left[a \leq X \leq b\right]
\end{aligned}$$

□

Proposition 8.10. Let $Z : \Omega \to \mathbb{R}$ be a random variable and $\mathbb{P}$ a distribution over Ω such that Z has normal(0,1) distribution. Then:

$$\mathbb{E}\left[Z\right] = 0, \quad \text{Var}\left[Z\right] = 1$$

Proof.

$$\mathbb{E}\left[Z\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \left[-e^{-x^2/2}\right]_{-\infty}^{\infty} = 0$$

$$\text{Var}\left[Z\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}\left[X\right]^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx - 0$$

We need to integrate $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx$. We use integration by parts:

$$\int_{-\infty}^{\infty} u(x) \cdot \frac{dv}{dx}(x) dx = \left[u(x) \cdot v(x)\right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{du}{dx}(x) \cdot v(x) dx$$

Where:

$$u(x) = x, \quad \frac{du}{dx}(x) = 1, \quad v(x) = -e^{x^2/2}, \quad \frac{dv}{dx}(x) = xe^{-x^2/2}$$

Therefore we get:

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx &= \frac{1}{\sqrt{2\pi}} \left[x \cdot -e^{-x^2/2} \right]_{-\infty}^{\infty} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} -e^{-x^2/2} dx \\ &= 0 + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx \end{aligned}$$

Now we need to integrate $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx$. We know this is equal to 1 as $\frac{1}{\sqrt{2\pi}} e^{x^2/2}$ is a probability density function. However we will show how to calculate this integral using polar coordinates:

$$\begin{aligned} \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx \right)^2 &= \frac{1}{2\pi} \left(\int_{-\infty}^{\infty} e^{-x^2/2} dx \right) \cdot \left(\int_{-\infty}^{\infty} e^{-y^2/2} dy \right) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2+y^2}{2} \right) dxdy \\ \text{Change to polar coordinates} &= \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} \exp\left(-\frac{r^2}{2} \right) r dr d\theta \\ &= \frac{1}{2\pi} \int_0^{2\pi} d\theta \cdot \int_0^{\infty} r e^{-r^2/2} dr \\ &= \left[-e^{-r^2/2} \right]_0^{\infty} \\ &= 0 - (-1) \\ &= 1 \end{aligned}$$

This means that $\text{Var}[Z] = 1$. □

Corollary 8.11. Let $X : \Omega \to \mathbb{R}$ be a random variable and $\mathbb{P}$ a distribution over Ω such that X has normal$\left(\mu, \sigma^2\right)$ distribution. Then:

$$\mathbb{E}[X] = \mu, \quad \text{Var}[X] = \sigma^2$$

Proof. Let Z be a random variable with normal(0,1) distribution then from Proposition 8.9, we can write $X = \mu + \sigma Z$. Then:

$$\mathbb{E}[X] = \mathbb{E}[\mu + \sigma Z] = \mu + \sigma \mathbb{E}[Z] = \mu + 0$$
$$\text{Var}[X] = \text{Var}[\mu + \sigma Z] = \text{Var}[\mu] + \text{Var}[\sigma Z] = 0 + \sigma^2 \text{Var}[Z] = \sigma^2$$

□

8.4 Gamma and Exponential Distribution

Definition 8.12. Let $X : \Omega \to \mathbb{R}$ be a random variable and $\mathbb{P}$ a distribution over Ω such that:

$$\mathbb{P}[a \leq X \leq b] = \int_a^b f(x) dx, \quad \forall a \leq b$$

where

$$f(x) \begin{cases} \left(\lambda e^{-\lambda x} (\lambda x)^{r-1} \right) / (r-1)! & \text{if } u > 0 \\ 0 & \text{if } u \leq 0 \end{cases}$$

for some $\lambda > 0$ and integer $r \geq 1$. Then we say that X has a *gamma*(r, λ) *distribution.*

Proposition 8.13. Let $X : \Omega \to \mathbb{R}$ be a random variable and $\mathbb{P}$ a distribution over Ω such that X has gamma(r, λ) distribution. Then:

$$\mathbb{E}[X] = \frac{r}{\lambda}, \quad \text{Var}[X] = \frac{r}{\lambda^2}$$

Definition 8.14. Let $X : \Omega \to \mathbb{R}$ be a random variable and $\mathbb{P}$ a distribution over Ω such that:

$$\mathbb{P}[a \le X \le b] = \int_a^b f(x)\,dx, \quad \forall a \le b$$

where

$$f(x) \begin{cases} \lambda e^{-\lambda x} & \text{if } u > 0 \\ 0 & \text{if } u \le 0 \end{cases}$$

for some $\lambda > 0$. Then we say that X has an *exponential*(λ) *distribution.*

Remark 8.15. Exponential distribution is just a special case of the gamma distribution where $r = 1$.

Proposition 8.16. Let $X : \Omega \to \mathbb{R}$ be a random variable and $\mathbb{P}$ a distribution over Ω such that X has exponential(λ) distribution. Then:

$$\mathbb{E}[X] = \frac{1}{\lambda}, \quad \mathrm{Var}[X] = \frac{1}{\lambda^2}$$

Theorem 8.17 (Memoryless Property)**.** Let T be a positive random variable, i.e. $T : \Omega \to (0, \infty)$ and $\mathbb{P}$ is a distribution over Ω.

$$T \text{ has an exponential}(\lambda) \text{ distribution for some } \lambda > 0$$

$$\Updownarrow$$

$$\mathbb{P}[T > t + s | T > t] = \mathbb{P}[T > s], \quad \forall s, t \ge 0$$

Proof. We only show the $\Downarrow$ part. Fix $s, t \ge 0$, suppose $T : \Omega \to (0, \infty)$ is a random variable and $\mathbb{P}$ is a distribution over Ω such that T has an exponential(λ) distribution for some $\lambda > 0$. Then:

$$\begin{aligned} \mathbb{P}[T > t + s | T > t] &= \frac{\mathbb{P}[\{T > t + s\} \cap \{T > t\}]}{\mathbb{P}[T > t]} \\ \text{As } \{T > t + s\} \cap \{T > t\} = \{T > t + s\} &= \frac{\mathbb{P}[T > t + s]}{\mathbb{P}[T > t]} \\ &= \frac{\exp(-\lambda(t + s))}{\exp(-\lambda t)} \\ &= \exp(-\lambda s) \\ &= \mathbb{P}[T > s] \end{aligned}$$

□