# Probability

Prof. F.P. Kelly

Lent 1996

These notes are maintained by Paul Metcalfe.
Comments and corrections to pdm23@cam.ac.uk.

```
Revision: 1.1
Date: 2002/09/20 14:45:43
```

The following people have maintained these notes.

| | |
|---|---|
| – June 2000 | Kate Metcalfe |
| June 2000 – July 2004 | Andrew Rogers |
| July 2004 – date | Paul Metcalfe |

# Contents

# Introduction

These notes are based on the course "Probability" given by Prof. F.P. Kelly in Cambridge in the Lent Term 1996. This typed version of the notes is totally unconnected with Prof. Kelly.

Other sets of notes are available for different courses. At the time of typing these courses were:

| | |
|---|---|
| Probability | Discrete Mathematics |
| Analysis | Further Analysis |
| Methods | Quantum Mechanics |
| Fluid Dynamics 1 | Quadratic Mathematics |
| Geometry | Dynamics of D.E.'s |
| Foundations of QM | Electrodynamics |
| Methods of Math. Phys | Fluid Dynamics 2 |
| Waves (etc.) | Statistical Physics |
| General Relativity | Dynamical Systems |
| Combinatorics | Bifurcations in Nonlinear Convection |

They may be downloaded from

`http://www.istari.ucam.org/maths/`.

# Chapter 1

# Basic Concepts

## 1.1 Sample Space

Suppose we have an experiment with a set $\Omega$ of outcomes. Then $\Omega$ is called the sample space. A potential outcome $\omega \in \Omega$ is called a sample point.

For instance, if the experiment is tossing coins, then $\Omega = \{H, T\}$, or if the experiment was tossing two dice, then $\Omega = \{(i, j) : i, j \in \{1, \dots, 6\}\}$.

A subset $A$ of $\Omega$ is called an event. An event $A$ occurs is when the experiment is performed, the outcome $\omega \in \Omega$ satisfies $\omega \in A$. For the coin-tossing experiment, then the event of a head appearing is $A = \{H\}$ and for the two dice, the event "rolling a four" would be $A = \{(1, 3), (2, 2), (3, 1)\}$.

## 1.2 Classical Probability

If $\Omega$ is finite, $\Omega = \{\omega_1, \dots, \omega_n\}$, and each of the $n$ sample points is "equally likely" then the probability of event $A$ occurring is

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}$$

**Example.** *Choose $r$ digits from a table of random numbers. Find the probability that for $0 \le k \le 9$,*

1. *no digit exceeds $k$,*

2. *$k$ is the greatest digit drawn.*

*Solution.* The event that no digit exceeds $k$ is

$$A_k = \{(a_1, \dots, a_r) : 0 \le a_i \le k, i = 1 \dots r\} .$$

Now $|A_k| = (k+1)^r$, so that $\mathbb{P}(A_k) = \left(\frac{k+1}{10}\right)^r$.

Let $B_k$ be the event that $k$ is the greatest digit drawn. Then $B_k = A_k \setminus A_{k-1}$. Also $A_{k-1} \subset A_k$, so that $|B_k| = (k+1)^r - k^r$. Thus $\mathbb{P}(B_k) = \frac{(k+1)^r - k^r}{10^r}$ $\square$

### The problem of the points

Players A and B play a series of games. The winner of a game wins a point. The two players are equally skillful and the stake will be won by the first player to reach a target. They are forced to stop when A is within 2 points and B within 3 points. How should the stake be divided?

Pascal suggested that the following continuations were equally likely

| | | | | |
|---|---|---|---|---|
| AAAA | AAAB | AABB | ABBB | BBBB |
| | AABA | ABBA | BABB | |
| | ABAA | ABAB | BBAB | |
| | BAAA | BABA | BBBA | |
| | | BAAB | | |
| | | BBAA | | |

This makes the ratio $11:5$. It was previously thought that the ratio should be $6:4$ on considering termination, but these results are not equally likely.

## 1.3 Combinatorial Analysis

The fundamental rule is:

Suppose $r$ experiments are such that the first may result in any of $n_1$ possible outcomes and such that for each of the possible outcomes of the first $i-1$ experiments there are $n_i$ possible outcomes to experiment $i$. Let $a_i$ be the outcome of experiment $i$. Then there are a total of $\prod_{i=1}^r n_i$ distinct $r$-tuples $(a_1, \dots, a_r)$ describing the possible outcomes of the $r$ experiments.

*Proof.* Induction. □

## 1.4 Stirling's Formula

For functions $g(n)$ and $h(n)$, we say that $g$ is asymptotically equivalent to $h$ and write $g(n) \sim h(n)$ if $\frac{g(n)}{h(n)} \to 1$ as $n \to \infty$.

**Theorem 1.1 (Stirling's Formula).** *As* $n \to \infty$,

$$\log \frac{n!}{\sqrt{2\pi n} n^n e^{-n}} \to 0$$

*and thus* $n! \sim \sqrt{2\pi n} n^n e^{-n}$.

We first prove the weak form of Stirling's formula, that $\log(n!) \sim n \log n$.

*Proof.* $\log n! = \sum_1^n \log k$. Now

$$\int_1^n \log x dx \leq \sum_1^n \log k \leq \int_1^{n+1} \log x dx,$$

and $\int_1^z logx\ dx = z \log z - z + 1$, and so

$$n \log n - n + 1 \leq \log n! \leq (n+1)\log(n+1) - n.$$

Divide by $n \log n$ and let $n \to \infty$ to sandwich $\frac{\log n!}{n \log n}$ between terms that tend to $1$. Therefore $\log n! \sim n \log n$. □

Now we prove the strong form.

*Proof.* For $x > 0$, we have

$$1 - x + x^2 - x^3 < \frac{1}{1+x} < 1 - x + x^2.$$

Now integrate from $0$ to $y$ to obtain

$$y - y^2/2 + y^3/3 - y^4/4 < \log(1+y) < y - y^2/2 + y^3/3.$$

Let $h_n = \log \frac{n!e^n}{n^{n+1/2}}$. Then[1] we obtain

$$\frac{1}{12n^2} - \frac{1}{12n^3} \le h_n - h_{n+1} \le \frac{1}{12n^2} + \frac{1}{6n^3}.$$

For $n \ge 2$, $0 \le h_n - h_{n+1} \le \frac{1}{n^2}$. Thus $h_n$ is a decreasing sequence, and $0 \le h_2 - h_{n+1} \le \sum_{r=2}^{n}(h_r - h_{r+1}) \le \sum_1^\infty \frac{1}{r^2}$. Therefore $h_n$ is bounded below, decreasing so is convergent. Let the limit be $A$. We have obtained

$$n! \sim e^A n^{n+1/2} e^{-n}.$$

We need a trick to find $A$. Let $I_r = \int_0^{\pi/2} \sin^r \theta \, d\theta$. We obtain the recurrence $I_r = \frac{r-1}{r} I_{r-2}$ by integrating by parts. Therefore $I_{2n} = \frac{(2n)!}{(2^n n!)^2}\pi/2$ and $I_{2n+1} = \frac{(2^n n!)^2}{(2n+1)!}$. Now $I_n$ is decreasing, so

$$1 \le \frac{I_{2n}}{I_{2n+1}} \le \frac{I_{2n-1}}{I_{2n+1}} = 1 + \frac{1}{2n} \to 1.$$

But by substituting our formula in, we get that

$$\frac{I_{2n}}{I_{2n+1}} \sim \frac{\pi}{2} \frac{2n+1}{n} \frac{2}{e^{2A}} \to \frac{2\pi}{e^{2A}}.$$

Therefore $e^{2A} = 2\pi$ as required. $\square$

[1] by playing silly buggers with $\log 1 + \frac{1}{n}$

# Chapter 2

# The Axiomatic Approach

## 2.1 The Axioms

Let $\Omega$ be a sample space. Then probability $\mathbb{P}$ is a real valued function defined on subsets of $\Omega$ satisfying :-

1. $0 \leq \mathbb{P}(A) \leq 1$ for $A \subset \Omega$,
2. $\mathbb{P}(\Omega) = 1$,
3. for a finite or infinite sequence $A_1, A_2, \dots \subset \Omega$ of disjoint events, $\mathbb{P}(\cup A_i) = \sum_i \mathbb{P}(A_u)$.

The number $\mathbb{P}(A)$ is called the probability of event $A$.

We can look at some distributions here. Consider an arbitrary finite or countable $\Omega = \{\omega_1, \omega_2, \dots\}$ and an arbitrary collection $\{p_1, p_2, \dots\}$ of non-negative numbers with sum 1. If we define

$$\mathbb{P}(A) = \sum_{i:\omega_i \in A} p_i,$$

it is easy to see that this function satisfies the axioms. The numbers $p_1, p_2, \dots$ are called a probability distribution. If $\Omega$ is finite with $n$ elements, and if $p_1 = p_2 = \dots = p_n = \frac{1}{n}$ we recover the classical definition of probability.

Another example would be to let $\Omega = \{0, 1, \dots\}$ and attach to outcome $r$ the probability $p_r = e^{-\lambda}\frac{\lambda^r}{r!}$ for some $\lambda > 0$. This is a distribution (as may be easily verified), and is called the Poisson distribution with parameter $\lambda$.

**Theorem 2.1 (Properties of $\mathbb{P}$).** *A probability $\mathbb{P}$ satisfies*

1. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$,
2. $\mathbb{P}(\emptyset) = 0$,
3. *if* $A \subset B$ *then* $\mathbb{P}(A) \leq \mathbb{P}(B)$,
4. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

*Proof.* Note that $\Omega = A \cup A^c$, and $A \cap A^c = \emptyset$. Thus $1 = \mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(A^c)$. Now we can use this to obtain $\mathbb{P}(\emptyset) = 1 - \mathbb{P}(\emptyset^c) = 0$. If $A \subset B$, write $B = A \cup (B \cap A^c)$, so that $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \cap A^c) \geq \mathbb{P}(A)$. Finally, write $A \cup B = A \cup (B \cap A^c)$ and $B = (B \cap A) \cup (B \cap A^c)$. Then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \cap A^c)$ and $\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \cap A^c)$, which gives the result. □

**Theorem 2.2 (Boole's Inequality).** *For any* $A_1, A_2, \dots \subset \Omega$,

$$\mathbb{P}\left(\bigcup_1^n A_i\right) \leq \sum_i^n \mathbb{P}(A_i)$$
$$\mathbb{P}\left(\bigcup_1^\infty A_i\right) \leq \sum_i^\infty \mathbb{P}(A_i)$$

*Proof.* Let $B_1 = A_1$ and then inductively let $B_i = A_i \setminus \bigcup_1^{i-1} B_k$. Thus the $B_i$'s are disjoint and $\bigcup_i B_i = \bigcup_i A_i$. Therefore

$$\begin{aligned}
\mathbb{P}\left(\bigcup_i A_i\right) &= \mathbb{P}\left(\bigcup_i B_i\right) \\
&= \sum_i \mathbb{P}(B_i) \\
&\leq \sum_i \mathbb{P}(A_i) \qquad \text{as } B_i \subset A_i.
\end{aligned}$$

□

**Theorem 2.3 (Inclusion-Exclusion Formula).**

$$\mathbb{P}\left(\bigcup_1^n A_i\right) = \sum_{\substack{S \subset \{1,\dots,n\} \\ S \neq \emptyset}} (-1)^{|S|-1} \mathbb{P}\left(\bigcap_{j \in S} A_j\right).$$

*Proof.* We know that $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2)$. Thus the result is true for $n = 2$. We also have that

$$\mathbb{P}(A_1 \cup \dots \cup A_n) = \mathbb{P}(A_1 \cup \dots \cup A_{n-1}) + \mathbb{P}(A_n) - \mathbb{P}((A_1 \cup \dots \cup A_{n-1}) \cap A_n).$$

But by distributivity, we have

$$\mathbb{P}\left(\bigcup_i^n A_i\right) = \mathbb{P}\left(\bigcup_1^{n-1} A_i\right) + \mathbb{P}(A_n) - \mathbb{P}\left(\bigcup_1^{n-1} (A_i \cap A_n)\right).$$

Application of the inductive hypothesis yields the result. □

**Corollary (Bonferroni Inequalities).**

$$\sum_{\substack{S \subset \{1,\dots,r\} \\ S \neq \emptyset}} (-1)^{|S|-1} \mathbb{P}\left(\bigcap_{j \in S} A_j\right) \begin{matrix} \leq \\ \textit{or} \\ \geq \end{matrix} \mathbb{P}\left(\bigcup_1^n A_i\right)$$

*according as $r$ is even or odd. Or in other words, if the inclusion-exclusion formula is truncated, the error has the sign of the omitted term and is smaller in absolute value. Note that the case $r = 1$ is Boole's inequality.*

*Proof.* The result is true for $n = 2$. If true for $n-1$, then it is true for $n$ and $1 \leq r \leq n-1$ by the inductive step above, which expresses a $n$-union in terms of two $n-1$ unions. It is true for $r = n$ by the inclusion-exclusion formula. $\square$

**Example (Derangements).** *After a dinner, the $n$ guests take coats at random from a pile. Find the probability that at least one guest has the right coat.*

*Solution.* Let $A_k$ be the event that guest $k$ has his[1] own coat.

We want $\mathbb{P}(\bigcup_{i=1}^n A_i)$. Now,

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_r}) = \frac{(n-r)!}{n!},$$

by counting the number of ways of matching guests and coats after $i_1, \ldots, i_r$ have taken theirs. Thus

$$\sum_{i_1 < \cdots < i_r} \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_r}) = \binom{n}{r} \frac{(n-r)!}{n!} = \frac{1}{r!},$$

and the required probability is

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = 1 - \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{(-1)^{n-1}}{n!},$$

which tends to $1 - e^{-1}$ as $n \to \infty$. $\square$

Furthermore, let $\mathbb{P}_m(n)$ be the probability that exactly $m$ guests take the right coat. Then $\mathbb{P}_0(n) \to e^{-1}$ and $n!\,\mathbb{P}_0(n)$ is the number of derangements of $n$ objects. Therefore

$$\begin{aligned} \mathbb{P}_m(n) &= \binom{n}{m} \frac{1 \times \mathbb{P}_0(n-m) \times (n-m)!}{n!} \\ &= \frac{\mathbb{P}_0(n-m)}{m!} \to \frac{e^{-1}}{m!} \text{ as } n \to \infty. \end{aligned}$$

## 2.2 Independence

**Definition 2.1.** *Two events $A$ and $B$ are said to be independent if*

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

*More generally, a collection of events $A_i$, $i \in I$ are independent if*

$$\mathbb{P}\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i)$$

*for all finite subsets $J \subset I$.*

**Example.** *Two fair dice are thrown. Let $A_1$ be the event that the first die shows an odd number. Let $A_2$ be the event that the second die shows an odd number and finally let $A_3$ be the event that the sum of the two numbers is odd. Are $A_1$ and $A_2$ independent? Are $A_1$ and $A_3$ independent? Are $A_1$, $A_2$ and $A_3$ independent?*

[1] I'm not being sexist, merely a lazy typist. Sex will be assigned at random...

*Solution.* We first calculate the probabilities of the events $A_1$, $A_2$, $A_3$, $A_1 \cap A_2$, $A_1 \cap A_3$ and $A_1 \cap A_2 \cap A_3$.

| Event | Probability |
|---|---|
| $A_1$ | $\frac{18}{36} = \frac{1}{2}$ |
| $A_2$ | As above, $\frac{1}{2}$ |
| $A_3$ | $\frac{6 \times 3}{36} = \frac{1}{2}$ |
| $A_1 \cap A_2$ | $\frac{3 \times 3}{36} = \frac{1}{4}$ |
| $A_1 \cap A_3$ | $\frac{3 \times 3}{36} = \frac{1}{4}$ |
| $A_1 \cap A_2 \cap A_3$ | $0$ |

Thus by a series of multiplications, we can see that $A_1$ and $A_2$ are independent, $A_1$ and $A_3$ are independent (also $A_2$ and $A_3$), but that $A_1$, $A_2$ and $A_3$ are *not* independent. □

Now we wish to state what we mean by "2 independent experiments"[2]. Consider $\Omega_1 = \{\alpha_1, \dots\}$ and $\Omega_2 = \{\beta_1, \dots\}$ with associated probability distributions $\{p_1, \dots\}$ and $\{q_1, \dots\}$. Then, by "2 independent experiments", we mean the sample space $\Omega_1 \times \Omega_2$ with probability distribution $\mathbb{P}((\alpha_i, \beta_j)) = p_i q_j$.

Now, suppose $A \subset \Omega_1$ and $B \subset \Omega_2$. The event $A$ can be interpreted as an event in $\Omega_1 \times \Omega_2$, namely $A \times \Omega_2$, and similarly for $B$. Then

$$\mathbb{P}(A \cap B) = \sum_{\substack{\alpha_i \in A \\ \beta_j \in B}} p_i q_j = \sum_{\alpha_i \in A} p_i \sum_{\beta_j \in B} q_j = \mathbb{P}(A)\,\mathbb{P}(B)\,,$$

which is why they are called "independent" experiments. The obvious generalisation to $n$ experiments can be made, but for an infinite sequence of experiments we mean a sample space $\Omega_1 \times \Omega_2 \times \dots$ satisfying the appropriate formula $\forall n \in \mathbb{N}$.

You might like to find the probability that $n$ independent tosses of a biased coin with the probability of heads $p$ results in a total of $r$ heads.

## 2.3 Distributions

The binomial distribution with parameters $n$ and $p$, $0 \le p \le 1$ has $\Omega = \{0, \dots, n\}$ and probabilities $p_i = \binom{n}{i} p^i (1-p)^{n-i}$.

**Theorem 2.4 (Poisson approximation to binomial).** *If $n \to \infty$, $p \to 0$ with $np = \lambda$ held fixed, then*

$$\binom{n}{r} p^r (1-p)^{n-r} \to e^{-\lambda} \frac{\lambda^r}{r!}.$$

[2] or more generally, $n$.

*Proof.*

$$\begin{aligned}\binom{n}{r}p^r(1-p)^{n-r} &= \frac{n(n-1)\dots(n-r+1)}{r!}p^r(1-p)^{n-r}\\ &= \frac{n}{n}\frac{n-1}{n}\dots\frac{n-r+1}{n}\frac{(np)^r}{r!}(1-p)^{n-r}\\ &= \prod_{i=1}^{r}\left(\frac{n-i+1}{n}\right)\frac{\lambda^r}{r!}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-r}\\ &\to 1\times\frac{\lambda^r}{r!}\times e^{-\lambda}\times 1\\ &= e^{-\lambda}\frac{\lambda^r}{r!}.\end{aligned}$$

□

Suppose an infinite sequence of independent trials is to be performed. Each trial results in a success with probability $p \in (0,1)$ or a failure with probability $1-p$. Such a sequence is called a sequence of Bernoulli trials. The probability that the first success occurs after exactly $r$ failures is $p_r = p(1-p)^r$. This is the *geometric distribution* with parameter $p$. Since $\sum_0^\infty p_r = 1$, the probability that all trials result in failure is zero.

## 2.4 Conditional Probability

**Definition 2.2.** *Provided $\mathbb{P}(B) > 0$, we define the conditional probability of $A|B$*[3] *to be*

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

*Whenever we write $\mathbb{P}(A|B)$, we assume that $\mathbb{P}(B) > 0$.*

Note that if $A$ and $B$ are independent then $\mathbb{P}(A|B) = \mathbb{P}(A)$.

**Theorem 2.5.**

1. $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\,\mathbb{P}(B)$,
2. $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A|B \cap C)\,\mathbb{P}(B|C)\,\mathbb{P}(C)$,
3. $\mathbb{P}(A|B \cap C) = \frac{\mathbb{P}(A \cap B|C)}{\mathbb{P}(B|C)}$,
4. *the function $\mathbb{P}(\circ|B)$ restricted to subsets of $B$ is a probability function on $B$.*

*Proof.* Results 1 to 3 are immediate from the definition of conditional probability. For result 4, note that $A \cap B \subset B$, so $\mathbb{P}(A \cap B) \leq \mathbb{P}(B)$ and thus $\mathbb{P}(A|B) \leq 1$. $\mathbb{P}(B|B) = 1$ (obviously), so it just remains to show the last axiom. For disjoint $A_i$'s,

$$\begin{aligned}\mathbb{P}\left(\bigcup_i A_i \middle| B\right) &= \frac{\mathbb{P}(\bigcup_i (A_i \cap B))}{\mathbb{P}(B)}\\ &= \frac{\sum_i \mathbb{P}(A_i \cap B)}{\mathbb{P}(B)}\\ &= \sum_i \mathbb{P}(A_i|B)\,, \text{ as required.}\end{aligned}$$

□

[3] read "$A$ given $B$".

**Theorem 2.6 (Law of total probability).** *Let $B_1, B_2, \ldots$ be a partition of $\Omega$. Then*

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A|B_i)\,\mathbb{P}(B_i)\,.$$

*Proof.*

$$\begin{aligned}\sum \mathbb{P}(A|B_i)\,\mathbb{P}(B_i) &= \sum \mathbb{P}(A \cap B_i) \\ &= \mathbb{P}\left(\bigcup_i A \cap B_i\right) \\ &= \mathbb{P}(A)\,,\ \text{as required.}\end{aligned}$$

□

**Example (Gambler's Ruin).** *A fair coin is tossed repeatedly. At each toss a gambler wins £1 if a head shows and loses £1 if tails. He continues playing until his capital reaches $m$ or he goes broke. Find $p_x$, the probability that he goes broke if his initial capital is £$x$.*

*Solution.* Let $A$ be the event that he goes broke before reaching £$m$, and let $H$ or $T$ be the outcome of the first toss. We condition on the first toss to get $\mathbb{P}(A) = \mathbb{P}(A|H)\,\mathbb{P}(H) + \mathbb{P}(A|T)\,\mathbb{P}(T)$. But $\mathbb{P}(A|H) = p_{x+1}$ and $\mathbb{P}(A|T) = p_{x-1}$. Thus we obtain the recurrence

$$p_{x+1} - p_x = p_x - p_{x-1}.$$

Note that $p_x$ is linear in $x$, with $p_0 = 1$, $p_m = 0$. Thus $p_x = 1 - \frac{x}{m}$. □

**Theorem 2.7 (Bayes' Formula).** *Let $B_1, B_2, \ldots$ be a partition of $\Omega$. Then*

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\,\mathbb{P}(B_i)}{\sum_j \mathbb{P}(A|B_j)\,\mathbb{P}(B_j)}.$$

*Proof.*

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A \cap B_i)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_i)\,\mathbb{P}(B_i)}{\sum_j \mathbb{P}(A|B_j)\,\mathbb{P}(B_j)},$$

by the law of total probability. □

# Chapter 3

# Random Variables

Let $\Omega$ be finite or countable, and let $p_\omega = \mathbb{P}(\{\omega\})$ for $\omega \in \Omega$.

**Definition 3.1.** *A random variable $X$ is a function $X : \Omega \mapsto \mathbb{R}$.*

Note that "random variable" is a somewhat inaccurate term, a random variable is neither random nor a variable.

**Example.** *If $\Omega = \{(i, j), 1 \le i, j \le t\}$, then we can define random variables $X$ and $Y$ by $X(i, j) = i + j$ and $Y(i, j) = \max\{i, j\}$*

Let $R_X$ be the image of $\Omega$ under $X$. When the range is finite or countable then the random variable is said to be discrete.

We write $\mathbb{P}(X = x_i)$ for $\sum_{\omega : X(\omega) = x_i} p_\omega$, and for $B \subset \mathbb{R}$

$$\mathbb{P}(X \in B) = \sum_{x \in B} \mathbb{P}(X = x) .$$

Then

$$(\mathbb{P}(X = x) , x \in R_X)$$

is the distribution of the random variable $X$. Note that it is a probability distribution over $R_X$.

## 3.1 Expectation

**Definition 3.2.** *The expectation of a random variable $X$ is the number*

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} p_w X(\omega)$$

*provided that this sum converges absolutely.*

Note that

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{\omega\in\Omega} p_w X(\omega) \\
&= \sum_{x\in R_X} \sum_{\omega : X(\omega)=x} p_\omega X(\omega) \\
&= \sum_{x\in R_X} x \sum_{\omega : X(\omega)=x} p_\omega \\
&= \sum_{x\in R_X} x\mathbb{P}(X = x)\,.
\end{aligned}$$

Absolute convergence allows the sum to be taken in any order.

If $X$ is a positive random variable and if $\sum_{\omega\in\Omega} p_\omega X(\omega) = \infty$ we write $\mathbb{E}[X] = +\infty$. If

$$\begin{aligned}
\sum_{\substack{x\in R_X\\ x\geq 0}} x\mathbb{P}(X = x) = \infty \text{ and} \\
\sum_{\substack{x\in R_X\\ x< 0}} x\mathbb{P}(X = x) = -\infty
\end{aligned}$$

then $\mathbb{E}[X]$ is undefined.

**Example.** *If* $\mathbb{P}(X = r) = e^{-\lambda}\frac{\lambda^r}{r!}$, *then* $\mathbb{E}[X] = \lambda$.

*Solution.*

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{r=0}^{\infty} r e^{-\lambda}\tfrac{\lambda^r}{r!} \\
&= \lambda e^{-\lambda} \sum_{r=1}^{\infty} \frac{\lambda^{r-1}}{(r-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda
\end{aligned}$$

$\square$

**Example.** *If* $\mathbb{P}(X = r) = \binom{n}{r} p^r (1-p)^{n-r}$ *then* $\mathbb{E}[X] = np$.

*Solution.*

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{r=0}^{n} r p^r (1-p)^{n-r} \binom{n}{r} \\
&= \sum_{r=0}^{n} r \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r} \\
&= n \sum_{r=1}^{n} \frac{(n-1)!}{(r-1)!(n-r)!} p^r (1-p)^{n-r} \\
&= np \sum_{r=1}^{n} \frac{(n-1)!}{(r-1)!(n-r)!} p^{r-1} (1-p)^{n-r} \\
&= np \sum_{r=1}^{n-1} \frac{(n-1)!}{(r)!(n-r)!} p^r (1-p)^{n-1-r} \\
&= np \sum_{r=1}^{n-1} \binom{n-1}{r} p^r (1-p)^{n-1-r} \\
&= np
\end{aligned}$$

□

For any function $f \colon \mathbb{R} \mapsto \mathbb{R}$ the composition of $f$ and $X$ defines a new random variable $f$ and $X$ defines the new random variable $f(X)$ given by

$$f(X)(w) = f(X(w)).$$

**Example.** *If $a$, $b$ and $c$ are constants, then $a + bX$ and $(X - c)^2$ are random variables defined by*

$$\begin{aligned}
(a + bX)(w) &= a + bX(w) \qquad \textit{and} \\
(X - c)^2(w) &= (X(w) - c)^2.
\end{aligned}$$

Note that $\mathbb{E}[X]$ is a constant.

**Theorem 3.1.**

1. *If $X \geq 0$ then $\mathbb{E}[X] \geq 0$.*
2. *If $X \geq 0$ and $\mathbb{E}[X] = 0$ then $\mathbb{P}(X = 0) = 1$.*
3. *If $a$ and $b$ are constants then $\mathbb{E}[a + bX] = a + b\mathbb{E}[X]$.*
4. *For any random variables $X$, $Y$ then $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.*
5. *$\mathbb{E}[X]$ is the constant which minimises $\mathbb{E}\left[(X - c)^2\right]$.*

*Proof.* 1. $X \geq 0$ means $X_w \geq 0 \; \forall \; w \in \Omega$

$$\text{So } \mathbb{E}[X] = \sum_{\omega \in \Omega} p_\omega X(\omega) \geq 0$$

2. If $\exists \omega \in \Omega$ with $p_\omega > 0$ and $X(\omega) > 0$ then $\mathbb{E}[X] > 0$, therefore $\mathbb{P}(X = 0) = 1$.

3.

$$\begin{aligned}\mathbb{E}[a+bX] &= \sum_{\omega\in\Omega}(a+bX(\omega))\,p_\omega \\ &= a\sum_{\omega\in\Omega}p_\omega + b\sum_{\omega\in\Omega}p_\omega X(\omega) \\ &= a+\mathbb{E}[X]\,.\end{aligned}$$

4. Trivial.

5. Now

$$\begin{aligned}\mathbb{E}\big[(X-c)^2\big] &= \mathbb{E}\big[(X-\mathbb{E}[X]+\mathbb{E}[X]-c)^2\big] \\ &= \mathbb{E}\big[[(X-\mathbb{E}[X])^2\big] + 2(X-\mathbb{E}[X])(\mathbb{E}[X]-c) + [(\mathbb{E}[X]-c)]^2\big] \\ &= \mathbb{E}\big[(X-\mathbb{E}[X])^2\big] + 2(\mathbb{E}[X]-c)\mathbb{E}[(X-\mathbb{E}[X])] + (\mathbb{E}[X]-c)^2 \\ &= \mathbb{E}\big[(X-\mathbb{E}[X])^2\big] + (\mathbb{E}[X]-c)^2.\end{aligned}$$

This is clearly minimised when $c=\mathbb{E}[X]$.

□

**Theorem 3.2.** *For any random variables* $X_1, X_2, ...., X_n$

$$\mathbb{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{E}[X_i]$$

*Proof.*

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^n X_i\right] &= \mathbb{E}\left[\sum_{i=1}^{n-1} X_i + X_n\right] \\ &= \mathbb{E}\left[\sum_{i=1}^{n-1} X_i\right] + \mathbb{E}[X]\end{aligned}$$

Result follows by induction. □

## 3.2 Variance

$$\begin{aligned}\operatorname{Var}X &= \mathbb{E}\big[X^2\big]-\mathbb{E}[X]^2 \qquad \text{for Random Variable } X \\ &= \mathbb{E}[X-\mathbb{E}[X]]^2 = \sigma^2 \\ \text{Standard Deviation} &= \sqrt{\operatorname{Var}X}\end{aligned}$$

**Theorem 3.3.** *Properties of Variance*

(i) $\operatorname{Var}X \geq 0$ if $\operatorname{Var}X=0$, then $\mathbb{P}(X=\mathbb{E}[X])=1$
Proof - from property 1 of expectation
(ii) If $a, b$ constants, $\operatorname{Var}(a+bX)=b^2\operatorname{Var}X$

*Proof.*

$$
\begin{aligned}
\operatorname{Var} a + bX &= \mathbb{E}[a + bX - a - b\mathbb{E}[X]] \\
&= b^2 \mathbb{E}[X - \mathbb{E}[X]] \\
&= b^2 \operatorname{Var} X
\end{aligned}
$$

□

(iii) $\operatorname{Var} X = \mathbb{E}\big[X^2\big] - \mathbb{E}[X]^2$

*Proof.*

$$
\begin{aligned}
\mathbb{E}[X - \mathbb{E}[X]]^2 &= \mathbb{E}\big[X^2 - 2X\mathbb{E}[X] + (\mathbb{E}[X])^2\big] \\
&= \mathbb{E}\big[X^2\big] - 2\mathbb{E}[X]\,\mathbb{E}[X] + \mathbb{E}[X]^2 \\
&= \mathbb{E}\big[X^2\big] - (\mathbb{E}[X])^2
\end{aligned}
$$

□

**Example.** *Let X have the geometric distribution* $\mathbb{P}(X = r) = pq^r$ *with* $r = 0, 1, 2...$ *and* $p + q = 1$*. Then* $\mathbb{E}[X] = \frac{q}{p}$ *and* $\operatorname{Var} X = \frac{q}{p^2}$.

*Solution.*

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_{r=0}^{\infty} rpq^r = pq \sum_{r=0}^{\infty} rq^{r-1} \\
&= \frac{1}{pq} \sum_{r=0}^{\infty} \frac{d}{dq}(q^r) = pq \frac{d}{dq}\Big(\frac{1}{1-q}\Big) \\
&= pq(1-q)^{-2} = \frac{q}{p} \\
\mathbb{E}\big[X^2\big] &= \sum_{r=0}^{\infty} r^2 p^2 q^{2r} \\
&= pq\Bigg(\sum_{r=1}^{\infty} r(r+1)q^{r-1} - \sum_{r=1}^{\infty} rq^{r-1}\Bigg) \\
&= pq(\frac{2}{(1-q)^3} - \frac{1}{(1-q)^2} = \frac{2q}{p^2} - \frac{q}{p} \\
\operatorname{Var} X &= \mathbb{E}\big[X^2\big] - \mathbb{E}[X]^2 \\
&= \frac{2q}{p^2} - \frac{q}{p} - \frac{q^2}{p} \\
&= \frac{q}{p^2}
\end{aligned}
$$

□

**Definition 3.3.** *The co-variance of random variables $X$ and $Y$ is:*

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

*The correlation of $X$ and $Y$ is:*

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

Linear Regression

**Theorem 3.4.** $\text{Var}\,(X+Y) = \text{Var}\,X + \text{Var}\,Y + 2Cov(X,Y)$

*Proof.*

$$\begin{aligned}\text{Var}\,(X+Y) &= \mathbb{E}\big[(X+Y)^2 - \mathbb{E}[X] - \mathbb{E}[Y]\big]^2\\ &= \mathbb{E}\big[(X-\mathbb{E}[X])^2 + (Y-\mathbb{E}[Y])^2 + 2(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\big]\\ &= \text{Var}\,X + \text{Var}\,Y + 2Cov(X,Y)\end{aligned}$$

□

## 3.3 Indicator Function

**Definition 3.4.** *The Indicator Function $I[A]$ of an event $A \subset \Omega$ is the function*

$$I[A](w) = \begin{cases} 1, & \textit{if}\ \omega \in A;\\ 0, & \textit{if}\ \omega \notin A.\end{cases} \tag{3.1}$$

NB that $I[A]$ is a random variable

1.

$$\begin{aligned}\mathbb{E}[I[A]] &= \mathbb{P}(A)\\ \mathbb{E}[I[A]] &= \sum_{\omega\in\Omega} p_\omega I[A](w)\\ &= \mathbb{P}(A)\end{aligned}$$

2. $I[A^c] = 1 - I[A]$

3. $I[A \cap B] = I[A]I[B]$

4.

$$\begin{aligned}I[A\cup B] &= I[A] + I[B] - I[A]I[B]\\ I[A\cup B](\omega) &= 1 \text{ if } \omega\in A \text{ or } \omega \in B\\ I[A\cup B](\omega) &= I[A](\omega) + I[B](\omega) - I[A]I[B](\omega) \text{ WORKS!}\end{aligned}$$

**Example.** $n \geq$ *couples are arranged randomly around a table such that males and females alternate. Let $N$ = The number of husbands sitting next to their wives. Calculate* *the* $\mathbb{E}[N]$ *and the* $\operatorname{Var} N$.

$$
\begin{aligned}
N &= \sum_{i=1}^{n} I[A_i] \qquad A_i = \textit{ event couple i are together} \\
\mathbb{E}[N] &= \mathbb{E}\left[\sum_{i=1}^{n} I[A_i]\right] \\
&= \sum_{i=1}^{n} \mathbb{E}[I[A_i]] \\
&= \sum_{i=1}^{n} \frac{2}{n} \\
\textit{Thus } \mathbb{E}[N] &= n\frac{2}{n} = 2 \\
\mathbb{E}\big[N^2\big] &= \mathbb{E}\left[\left(\sum_{i=1}^{n} I[A_i]\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_{i=1}^{n} I[A_i]^2 + 2\sum_{i\leq j} I[A_i]I[A_j]\right)\right] \\
&= n\mathbb{E}\big[I[A_i]^2\big] + n(n-1)\mathbb{E}[(I[A_1]I[A_2])] \\
\mathbb{E}\big[I[A_i]^2\big] &= \mathbb{E}[I[A_i]] = \frac{2}{n} \\
\mathbb{E}[(I[A_1]I[A_2])] &= I\mathbb{E}[[A_1 \cap B_2]] = \mathbb{P}(A_1 \cap A_2) \\
&= \mathbb{P}(A_1)\,\mathbb{P}(A_2|A_1) \\
&= \frac{2}{n}\left(\frac{1}{n-1}\frac{1}{n-1} - \frac{n-2}{n-1}\frac{2}{n-1}\right) \\
\operatorname{Var} N &= \mathbb{E}\big[N^2\big] - \mathbb{E}[N]^2 \\
&= \frac{2}{n-1}\left(1 + 2(n-2)\right) - 2 \\
&= \frac{2(n-2)}{n-1}
\end{aligned}
$$

## 3.4 Inclusion - Exclusion Formula

$$\begin{aligned}
\bigcup_1^N A_i &= \left(\bigcap_1^N A_i^c\right)^c \\
I\left[\bigcup_1^N A_i\right] &= I\left[\left(\bigcap_1^N A_i^c\right)^c\right] \\
&= 1 - I\left[\bigcap_1^N A_i^c\right] \\
&= 1 - \prod_1^N I[A_i^c] \\
&= 1 - \prod_1^N (1 - I[A_i]) \\
&= \sum_1^N I[A_i] - \sum i_1 \leq i_2 I[A_1]I[A_2] \\
&+ ... + (-1)^{j+1} \sum_{i_1 \leq i_2 ... \leq i_j} I[A_1]I[A_2]...I[A_j] + ...
\end{aligned}$$

Take Expectation

$$\begin{aligned}
\mathbb{E}\left[\bigcup_1^N A_i\right] &= \mathbb{P}\left(\bigcup_1^N A_i\right) \\
&= \sum_1^N \mathbb{P}(A_i) - \sum i_1 \leq i_2 \mathbb{P}(A_1 \cap A_2) \\
&+ ... + (-1)^{j+1} \sum_{i_1 \leq i_2 ... \leq i_j} \mathbb{P}\big(A_{i_1} \cap A_{i_2} \cap .... \cap A_{i_j}\big) + ...
\end{aligned}$$

## 3.5 Independence

**Definition 3.5.** *Discrete random variables $X_1, ..., X_n$ are independent if and only if for any $x_1...x_n$ :*

$$\mathbb{P}(X_1 = x_1, X_2 = x_2.......X_n = x_n) = \prod_1^N \mathbb{P}(X_i = x_i)$$

**Theorem 3.5 (Preservation of Independence).** *If $X_1, ..., X_n$ are independent random variables and $f_1, f_2...f_n$ are functions $\mathbb{R} \to \mathbb{R}$ then $f_1(X_1)...f_n(X_n)$ are independent random variables*

*Proof.*

$$\begin{aligned}
\mathbb{P}(f_1(X_1) = y_1, \ldots, f_n(X_n) = y_n) &= \sum_{\substack{x_1: f_1(X_1)=y_1, \ldots \\ x_n: f_n(X_n)=y_n}} \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) \\
&= \prod_1^N \sum_{x_i: f_i(X_i)=y_i} \mathbb{P}(X_i = x_i) \\
&= \prod_1^N \mathbb{P}(f_i(X_i) = y_i)
\end{aligned}$$

□

**Theorem 3.6.** *If $X_1 \ldots\ldots X_n$ are independent random variables then:*

$$\mathbb{E}\left[\prod_1^N X_i\right] = \prod_1^N \mathbb{E}[X_i]$$

*NOTE* that $\mathbb{E}[\sum X_i] = \sum \mathbb{E}[X_i]$ without requiring independence.

*Proof.* Write $R_i$ for $R_{X_i}$ the range of $X_i$

$$\begin{aligned}
\mathbb{E}\left[\prod_1^N X_i\right] &= \sum_{x_1 \in R_1} \ldots\ldots \sum_{x_n \in R_n} x_1 .. x_n \mathbb{P}(X_1 = x_1, X_2 = x_2 \ldots\ldots, X_n = x_n) \\
&= \prod_1^N \left( \sum_{x_i \in R_i} \mathbb{P}(X_i = x_i) \right) \\
&= \prod_1^N \mathbb{E}[X_i]
\end{aligned}$$

□

**Theorem 3.7.** *If $X_1, \ldots, X_n$ are independent random variables and $f_1 \ldots. f_n$ are function $\mathbb{R} \to \mathbb{R}$ then:*

$$\mathbb{E}\left[\prod_1^N f_i(X_i)\right] = \prod_1^N \mathbb{E}[f_i(X_i)]$$

*Proof.* Obvious from last two theorems! □

**Theorem 3.8.** *If $X_1, \ldots, X_n$ are independent random variables then:*

$$\operatorname{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \operatorname{Var} X_i$$

*Proof.*

$$\begin{aligned}
\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) &= \mathbb{E}\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] - \mathbb{E}\left[\sum_{i=1}^{n} X_i\right]^2 \\
&= \mathbb{E}\left[\sum_i X_i^2 + \sum_{i\neq j} X_i X_j\right] - \mathbb{E}\left[\sum_{i=1}^{n} X_i\right]^2 \\
&= \sum_i \mathbb{E}\left[X_i^2\right] + \sum_{i\neq j} \mathbb{E}[X_i X_j] - \sum_i \mathbb{E}[X_i]^2 - \sum_{i\neq j} \mathbb{E}[X_i]\,\mathbb{E}[X_j] \\
&= \sum_i \left(\mathbb{E}\left[X_i^2\right] - \mathbb{E}[X_i]^2\right) \\
&= \sum_i \operatorname{Var} X_i
\end{aligned}$$

□

**Theorem 3.9.** *If $X_1, \ldots, X_n$ are independent identically distributed random variables then*

$$\operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\operatorname{Var} X_i$$

*Proof.*

$$\begin{aligned}
\operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) &= \frac{1}{n^2}\operatorname{Var} X_i \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{Var} X_i \\
&= \frac{1}{n}\operatorname{Var} X_i
\end{aligned}$$

□

**Example.** *Experimental Design.* *Two rods of unknown lengths $a, b$. A rule can measure the length but with but with error having 0 mean (unbiased) and variance $\sigma^2$. Errors independent from measurement to measurement. To estimate $a, b$ we could take separate measurements $A, B$ of each rod.*

$$\begin{aligned}
\mathbb{E}[A] &= a \qquad \operatorname{Var} A = \sigma^2 \\
\mathbb{E}[B] &= b \qquad \operatorname{Var} B = \sigma^2
\end{aligned}$$

*Can we do better? YEP! Measure* $a+b$ *as* $X$ *and* $a-b$ *as* $Y$

$$\begin{aligned} \mathbb{E}[X] &= a+b \qquad \operatorname{Var} X = \sigma^2 \\ \mathbb{E}[Y] &= a-b \qquad \operatorname{Var} Y = \sigma^2 \\ \mathbb{E}\left[\frac{X+Y}{2}\right] &= a \\ \operatorname{Var}\frac{X+Y}{2} &= \frac{1}{2}\sigma^2 \\ \mathbb{E}\left[\frac{X-Y}{2}\right] &= b \\ \operatorname{Var}\frac{X-Y}{2} &= \frac{1}{2}\sigma^2 \end{aligned}$$

*So this is better.*

**Example.** *Non standard dice. You choose 1 then I choose one. Around this cycle*

$a \to B$ $\mathbb{P}(A \geq B) = \frac{2}{3}$. *So the relation 'A better that B' is not transitive.*

# Chapter 4

# Inequalities

## 4.1 Jensen's Inequality

A function $f;(a,b) \to \mathbb{R}$ is convex if

$$f(px+qy) \le pf(x) + (1-p)f(y) \text{ - } \forall x, y \in (a,b) \text{ - } \forall p \in (0,1)$$

Strictly convex if strict inequality holds when $x \ne y$

f is concave if $-f$ is convex. f is strictly concave if $-f$ is strictly convex

Concave

neither concave or convex.

We know that if f is twice differentiable and $f^{''}(x) \geq 0$ for $x \in (a, b)$ the if f is convex and strictly convex if $f^{''}(x) \geq 0$ for$x \in (a, b)$.

**Example.**

$$\begin{aligned} f(x) &= -\log x \\ f^{'}(x) &= \frac{-1}{x} \\ f^{''}(x) &= \frac{1}{x^2} \geq 0 \end{aligned}$$

$f(x)$ *is strictly convex on* $(0, \infty)$

**Example.**

$$\begin{aligned} f(x) &= -x \log x \\ f^{'}(x) &= -(1 + logx) \\ f^{''}(x) &= \frac{-1}{x} \leq 0 \end{aligned}$$

*Strictly concave.*

**Example.** $f(x = x^3$ *is strictly convex on* $(0, \infty)$ *but not on* $(-\infty, \infty)$

**Theorem 4.1.** *Let* $f : (a, b) \to \mathbb{R}$ *be a convex function. Then:*

$$\sum_{i=1}^{n} p_i f(x_i) \geq f\left(\sum_{i=1}^{n} p_i x_i\right)$$

$x_1, \ldots, X_n \in (a, b)$, $p_1, \ldots, p_n \in (0, 1)$ *and* $\sum p_i = 1$. *Further more if f is strictly convex then equality holds if and only if all x's are equal.*

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

*Proof.* By induction on n $n = 1$ nothing to prove $n = 2$ definition of convexity. Assume results holds up to n-1. Consider $x_1, ..., x_n \in (a, b)$, $p_1, ..., p_n \in (0, 1)$ and $\sum p_i = 1$

$$\text{For } i = 2...n, \text{ set } p_i^{'} = \frac{p_i}{1 - p_i}, \text{ such that } \sum_{i=2}^{n} p_i^{'} = 1$$

Then by the inductive hypothesis twice, first for n-1, then for 2

$$\begin{aligned}\sum_{1}^{n} p_i f_i(x_i) &= p_1 f(x_1) + (1 - p_1) \sum_{i=2}^{n} p_i^{'} f(x_i) \\ &\geq p_1 f(x_1) + (1 - p_1) f\left(\sum_{i=2}^{n} p_i^{'} x_i\right) \\ &\geq f\left(p_1 x_1 + (1 - p_1) \sum_{i=2}^{n} p_i^{'} x_i\right) \\ &= f\left(\sum_{i=1}^{n} p_i x_i\right)\end{aligned}$$

f is strictly convex $n \geq 3$ and not all the $x_i' s$ equal then we assume not all of $x_2 ... x_n$ are equal. But then

$$(1 - p_j) \sum_{i=2}^{n} p_i^{'} f(x_i) \geq (1 - p_j) f\left(\sum_{i=2}^{n} p_i^{'} x_i\right)$$

So the inequality is strict. □

**Corollary (AM/GM Inequality).** *Positive real numbers* $x_1, \ldots, x_n$

$$\left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} \leq \frac{1}{n} \sum_{i=1}^{n} x_i$$

Equality holds if and only if $x_1 = x_2 = \cdots = x_n$

*Proof.* Let

$$\mathbb{P}(X = x_i) = \frac{1}{n}$$

then $f(x) = -\log x$ is a convex function on $(0, \infty)$.

So

$$\begin{aligned}\mathbb{E}[f(x)] &\geq f\left(\mathbb{E}[x]\right) \text{ (Jensen's Inequality)} \\ -\mathbb{E}[\log x] &\geq \log \mathbb{E}[x] \qquad [1] \\ \text{Therefore } -\frac{1}{n} \sum_{1}^{n} \log x_i &\leq -\log \frac{1}{n} \sum_{1}^{n} x \\ \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} &\leq \frac{1}{n} \sum_{i=1}^{n} x_i \qquad [2]\end{aligned}$$

For strictness since f strictly convex equation holds in [1] and hence [2] if and only if $x_1 = x_2 = \cdots = x_n$ □

If $f : (a, b) \to \mathbb{R}$ is a convex function then it can be shown that at each point $y \in (a, b) \exists$ a linear function $\alpha_y + \beta_y x$ such that

$$\begin{aligned} f(x) &\leq \alpha_y + \beta_y x \qquad x \in (a, b) \\ f(y) &= \alpha_y + \beta_y y \end{aligned}$$

If f is differentiable at y then the linear function is the tangent $f(y) + (x - y)f^{'}(y)$

Let $y = \mathbb{E}[X]$, $\alpha = \alpha_y$ and $\beta = \beta_y$

$$f(\mathbb{E}[X]) = \alpha + \beta\mathbb{E}[X]$$

So for any random variable X taking values in $(a, b)$

$$\begin{aligned} \mathbb{E}[f(X)] &\geq \mathbb{E}[\alpha + \beta X] \\ &= \alpha + \beta\mathbb{E}[X] \\ &= f(\mathbb{E}[X]) \end{aligned}$$

## 4.2 Cauchy-Schwarz Inequality

**Theorem 4.2.** *For any random variables $X, Y$,*

$$\mathbb{E}[XY]^2 \leq \mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right]$$

*Proof.* For $a, b \in \mathbb{R}$ Let

$$\begin{aligned} Let Z &= aX - bY \\ \text{Then} 0 \leq \mathbb{E}\left[Z^2\right] &= \mathbb{E}\left[(aX - bY)^2\right] \\ &= a^2\mathbb{E}\left[X^2\right] - 2ab\mathbb{E}[XY] + b^2\mathbb{E}\left[Y^2\right] \end{aligned}$$

quadratic in a with at most one real root and therefore has discriminant $\leq 0$.

Take $b \neq 0$

$$\mathbb{E}[XY]^2 \leq \mathbb{E}\left[X^2\right] \mathbb{E}\left[Y^2\right]$$

□

**Corollary.**

$$|Corr(X, Y)| \leq 1$$

*Proof.* Apply Cauchy-Schwarz to the random variables $X - \mathbb{E}[X]$ and $Y - \mathbb{E}[Y]$ □

## 4.3 Markov's Inequality

**Theorem 4.3.** *If X is any random variable with finite mean then,*

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[|X|]}{a} \textit{ for any a} \geq 0$$

*Proof.* Let

$$A = |X| \geq a$$
$$Then\, |X| \geq aI[A]$$

Take expectation

$$\mathbb{E}[|X|] \geq a\mathbb{P}(A)$$
$$\mathbb{E}[|X|] \geq a\mathbb{P}(|X| \geq a)$$

□

## 4.4 Chebyshev's Inequality

**Theorem 4.4.** *Let X be a random variable with* $\mathbb{E}\left[X^2\right] \leq \infty$. *Then* $\forall \epsilon \geq 0$

$$\mathbb{P}(|X| \geq \epsilon) \leq \frac{\mathbb{E}\left[X^2\right]}{\epsilon^2}$$

*Proof.*

$$I[|X| \geq \epsilon] \leq \frac{x^2}{\epsilon^2}\ \forall x$$

Then

$$I[|X| \geq \epsilon] \leq \frac{x^2}{\epsilon^2}$$

Take Expectation

$$\mathbb{P}(|X| \geq \epsilon) \leq \mathbb{E}\left[\frac{x^2}{\epsilon^2}\right] = \frac{\mathbb{E}\left[X^2\right]}{\epsilon^2}$$

□

*Note*

1. The result is "distribution free" - no assumption about the distribution of X (other than $\mathbb{E}\left[X^2\right] \leq \infty$).

2. It is the "best possible" inequality, in the following sense

$$\begin{aligned} X &= +\epsilon \text{ with probability } \frac{c}{2\epsilon^2} \\ &= -\epsilon \text{ with probability } \frac{c}{2\epsilon^2} \\ &= 0 \text{ with probability } 1 - \frac{c}{\epsilon^2} \\ \text{Then } \mathbb{P}(|X| \geq \epsilon) &= \frac{c}{\epsilon^2} \\ \mathbb{E}\left[X^2\right] &= c \\ \mathbb{P}(|X| \geq \epsilon) &= \frac{c}{\epsilon^2} = \frac{\mathbb{E}\left[X^2\right]}{\epsilon^2} \end{aligned}$$

3. If $\mu = \mathbb{E}[X]$ then applying the inequality to $X - \mu$ gives

$$\mathbb{P}(X - \mu \geq \epsilon) \leq \frac{\operatorname{Var} X}{\epsilon^2}$$

Often the most useful form.

## 4.5 Law of Large Numbers

**Theorem 4.5 (Weak law of large numbers).** *Let $X_1, X_2.....$ be a sequences of independent identically distributed random variables with Variance $\sigma^2 \leq \infty$ Let*

$$S_n = \sum_{i=1}^{n} X_i$$

*Then*

$$\forall \epsilon \geq 0, \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) \to 0 \textit{ as } n \to \infty$$

*Proof.* By Chebyshev's Inequality

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{S_n}{n}-\mu\right|\geq\epsilon\right) &\leq \frac{\mathbb{E}\left[\left(\frac{S_n}{n}-\mu\right)^2\right]}{\epsilon^2} \\
&= \frac{\mathbb{E}\left[(S_n-n\mu)^2\right]}{n^2\epsilon^2} \text{ properties of expectation} \\
&= \frac{\operatorname{Var} S_n}{n^2\epsilon^2} \text{ Since } \mathbb{E}[S_n]=n\mu \\
\text{But } \operatorname{Var} S_n &= n\sigma^2 \\
\text{Thus } \mathbb{P}\left(\left|\frac{S_n}{n}-\mu\right|\geq\epsilon\right) &\leq \frac{n\sigma^2}{n^2\epsilon^2}=\frac{\sigma^2}{n\epsilon^2}\to 0
\end{aligned}$$

□

**Example.** *$A_1, A_2$... are independent events, each with probability p. Let $X_i = I[A_i]$. Then*

$$\frac{S_n}{n}=\frac{nA}{n}=\frac{\textit{number of times A occurs}}{\textit{number of trials}}$$

$$\mu=\mathbb{E}[I[A_i]]=\mathbb{P}(A_i)=p$$

*Theorem states that*

$$\mathbb{P}\left(\left|\frac{S_n}{n}-p\right|\geq\epsilon\right)\to 0 \textit{ as } n\to\infty$$

*Which recovers the intuitive definition of probability.*

**Example.** *A Random Sample of size n is a sequence $X_1, X_2, \dots, X_n$ of independent identically distributed random variables ('n observations')*

$$\bar{X}=\frac{\sum_{i=1}^n X_i}{n} \textit{ is called the SAMPLE MEAN}$$

*Theorem states that provided the variance of $X_i$ is finite, the probability that the sample mean differs from the mean of the distribution by more than $\epsilon$ approaches 0 as $n \to \infty$.*

We have shown the weak law of large numbers. Why weak? $\exists$ a strong form of larger numbers.

$$\mathbb{P}\left(\frac{S_n}{n}\to\mu \text{ as } n\to\infty\right)=1$$

This is NOT the same as the weak form. What does this mean?

$\omega \in \Omega$ determines

$$\frac{S_n}{n}, \qquad n=1,2,\dots$$

as a sequence of real numbers. Hence it either tends to $\mu$ or it doesn't.

$$\mathbb{P}\left(\omega : \frac{S_n(\omega)}{n}\to\mu \text{ as } n\to\infty\right)=1$$

# Chapter 5

# Generating Functions

In this chapter, assume that X is a random variable taking values in the range $0, 1, 2, \ldots$. Let $p_r = \mathbb{P}(X = r)\; r = 0, 1, 2, \ldots$

**Definition 5.1.** *The* Probability Generating Function *(p.g.f) of the random variable X,or of the distribution* $p_r = 0, 1, 2, \ldots$*, is*

$$p(z) = \mathbb{E}\left[z^X\right] = \sum_{r=0}^{\infty} z^r \mathbb{P}(X = r) = \sum_{r=0}^{\infty} p_r z^r$$

*This* $p(z)$ *is a polynomial or a power series. If a power series then it is convergent for* $|z| \leq 1$ *by comparison with a geometric series.*

$$|p(z)| \leq \sum_r p_r \left|z\right|^r \leq \sum_r p_r = 1$$

**Example.**

$$\begin{aligned} p_r &= \frac{1}{6}\; r = 1, \ldots, 6 \\ p(z) = \mathbb{E}\left[z^X\right] &= \frac{1}{6}\left(1 + z + \ldots z^6\right) \\ &= \frac{z}{6}\frac{1 - z^6}{1 - z} \end{aligned}$$

**Theorem 5.1.** *The distribution of X is uniquely determined by the p.g.f* $p(z)$.

*Proof.* We know that we can differential p(z) term by term for $|z| \leq 1$

$$p^{'}(z) = p_1 + 2p_2 z + \ldots$$
$$\text{and so } p^{'}(0) = p_1 \qquad (p(0) = p_0)$$

Repeated differentiation gives

$$p^{(i)}(z) = \sum_{r=i}^{\infty} \frac{r!}{(r-i)!} p_r z^{r-i}$$

and has $p^{(i)} = 0 = i! p_i$ Thus we can recover $p_0, p_1, \ldots$ from p(z) $\square$

**Theorem 5.2 (Abel's Lemma).**

$$\mathbb{E}[X] = \lim_{z \to 1} p'(z)$$

*Proof.*

$$p'(z) = \sum_{r=i}^{\infty} r p_r z^{r-1} \qquad |z| \leq 1$$

For $z \in (0,1)$, $p'(z)$ is a non decreasing function of z and is bounded above by

$$\mathbb{E}[X] = \sum_{r=i}^{\infty} r p_r$$

Choose $\epsilon \geq 0$, N large enough that

$$\sum_{r=i}^{N} r p_r \geq \mathbb{E}[X] - \epsilon$$

Then

$$\lim_{z \to 1} \sum_{r=i}^{\infty} r p_r z^{r-1} \geq \lim_{z \to 1} \sum_{r=i}^{N} r p_r z^{r-1} = \sum_{r=i}^{N} r p_r$$

True $\forall \epsilon \geq 0$ and so

$$\mathbb{E}[X] = \lim_{z \to 1} p'(z)$$

□

Usually $p'(z)$ is continuous at z=1, then $\mathbb{E}[X] = p'(1)$.

$$\left( \text{Recall } p(z) = \frac{z}{6} \frac{1 - z^6}{1 - z} \right)$$

**Theorem 5.3.**

$$\mathbb{E}[X(X-1)] = \lim_{z \to 1} p''(z)$$

*Proof.*

$$p''(z) = \sum_{r=2}^{\infty} r(r-1) p z^{r-2}$$

Proof now the same as Abel's Lemma □

**Theorem 5.4.** *Suppose that $X_1, X_2, \ldots, X_n$ are independent random variables with p.g.f's $p_1(z), p_2(z), \ldots, p_n(z)$. Then the p.g.f of*

$$X_1 + X_2 + \ldots X_n$$

*is*

$$p_1(z) p_2(z) \ldots p_n(z)$$

*Proof.*

$$\begin{aligned} \mathbb{E}\left[z^{X_1 + X_2 + \ldots X_n}\right] &= \mathbb{E}\left[z^{X_1} . z^{X_2} \ldots . z^{X_n}\right] \\ &= \mathbb{E}\left[z^{X_1}\right] \mathbb{E}\left[z^{X_2}\right] \ldots \mathbb{E}\left[z^{X_n}\right] \\ &= p_1(z) p_2(z) \ldots p_n(z) \end{aligned}$$

□

**Example.** *Suppose X has Poisson Distribution*

$$\mathbb{P}(X = r) = e^{-\lambda}\frac{\lambda^r}{r!} \qquad r = 0, 1, \ldots$$

*Then*

$$\begin{aligned}\mathbb{E}\left[z^X\right] &= \sum_{r=0}^{\infty} z^r e^{-\lambda}\frac{\lambda^r}{r!} \\ &= e^{-\lambda}e^{-\lambda z} \\ &= e^{-\lambda(1-z)}\end{aligned}$$

*Let's calculate the variance of X*

$$p^{'} = \lambda e^{-\lambda(1-z)} \qquad p^{''} = \lambda^2 e^{-\lambda(1-z)}$$

*Then*

$$\mathbb{E}[X] = \lim_{z\to 1} p^{'}(z) = p^{'}(1)(\text{ Since } p^{'}(z) \text{ continuous at } z = 1\text{ })\mathbb{E}[X] = \lambda$$

$$\begin{aligned}\mathbb{E}[X(X-1)] &= p^{''}(1) = \lambda^2 \\ \operatorname{Var} X &= \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 \\ &= \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2 \\ &= \lambda^2 + \lambda - \lambda^2 \\ &= \lambda\end{aligned}$$

**Example.** *Suppose that Y has a Poisson Distribution with parameter $\mu$. If X and Y are independent then:*

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

*But this is the p.g.f of a Poisson random variable with parameter $\lambda + \mu$. By uniqueness (first theorem of the p.g.f) this must be the distribution for $X + Y$*

**Example.** *X has a binomial Distribution,*

$$\begin{aligned}\mathbb{P}(X = r) &= \binom{n}{r}p^r(1-p)^{n-r} \qquad r = 0, 1, \ldots \\ \mathbb{E}\left[z^X\right] &= \sum_{r=0}^{n}\binom{n}{r}p^r(1-p)^{n-r}z^r \\ &= (pz + 1 - p)^n\end{aligned}$$

*This shows that $X = Y_1 + Y_2 + \cdots + Y_n$. Where $Y_1 + Y_2 + \cdots + Y_n$ are independent random variables each with*

$$\mathbb{P}(Y_i = 1) = p \qquad \mathbb{P}(Y_i = 0) = 1 - p$$

Note *if the p.g.f factorizes look to see if the random variable can be written as a sum.*

## 5.1 Combinatorial Applications

Tile a $(2 \times n)$ bathroom with $(2 \times 1)$ tiles. How many ways can this be done? Say $f_n$

$$f_n = f_{n-1} + f_{n-2} \qquad f_0 = f_1 = 1$$

Let

$$F(z) = \sum_{n=0}^{\infty} f_n z^n$$

$$\begin{aligned}
f_n z^n &= f_{n-1} z^n + f_{n-2} z^n \\
\sum_{n=2}^{\infty} f_n z^n &= \sum_{n=2}^{\infty} f_{n-1} z^n + \sum_{n=0}^{\infty} f_{n-2} z^n \\
F(z) - f_0 - z f_1 &= z(F(z) - f_0) + z^2 F(z) \\
F(z)(1 - z - z^2) &= f_0(1 - z) + z f_1 \\
&= 1 - z + z = 1.
\end{aligned}$$

Since $f_0 = f_1 = 1$, then $F(z) = \frac{1}{1-z-z^2}$
Let

$$\alpha_1 = \frac{1+\sqrt{5}}{2} \qquad \alpha_2 = \frac{1-\sqrt{5}}{2}$$

$$\begin{aligned}
F(z) &= \frac{1}{(1-\alpha_1 z)(1-\alpha_2 z)} \\
&= \frac{\alpha_1}{(1-\alpha_1 z)} - \frac{\alpha_2}{(1-\alpha_2 z)} \\
&= \frac{1}{\alpha_1 - \alpha_2} \left( \alpha_1 \sum_{n=0}^{\infty} \alpha_1^n z^n - \alpha_2 \sum_{n=0}^{\infty} \alpha_2^n z^n \right)
\end{aligned}$$

The coefficient of $z_1^n$, that is $f_n$, is

$$f_n = \frac{1}{\alpha_1 - \alpha_2} (\alpha_1^{n+1} - \alpha_2^{n+1})$$

## 5.2 Conditional Expectation

Let $X$ and $Y$ be random variables with joint distribution

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

Then the distribution of X is

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

This is often called the Marginal distribution for $X$. The conditional distribution for $X$ given by $Y = y$ is

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

**Definition 5.2.** *The conditional expectation of X given $Y = y$ is,*

$$\mathbb{E}[X = x|Y = y] = \sum_{x \in R_x} x\mathbb{P}(X = x|Y = y)$$

*The conditional Expectation of X given Y is the random variable $\mathbb{E}[X|Y]$ defined by*

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

*Thus $\mathbb{E}[X|Y] : \Omega \to \mathbb{R}$*

**Example.** *Let $X_1, X_2, \ldots, X_n$ be independent identically distributed random variables with $\mathbb{P}(X_1 = 1) = p$ and $\mathbb{P}(X_1 = 0) = 1 - p$. Let*

$$Y = X_1 + X_2 + \cdots + X_n$$

*Then*

$$\begin{aligned}
\mathbb{P}(X_1 = 1|Y = r) &= \frac{\mathbb{P}(X_1 = 1, Y = r)}{\mathbb{P}(Y = r)} \\
&= \frac{\mathbb{P}(X_1 = 1, X_2 + \cdots + X_n = r - 1)}{\mathbb{P}(Y = r)} \\
&= \frac{\mathbb{P}(X_1)\,\mathbb{P}(X_2 + \cdots + X_n = r - 1)}{\mathbb{P}(Y = r)} \\
&= \frac{p\binom{n-1}{r-1}p^{r-1}(1-p)^{n-r}}{\binom{n}{r}p^r(1-p)^{n-r}} \\
&= \frac{\binom{n-1}{r-1}}{\binom{n}{r}} \\
&= \frac{r}{n}
\end{aligned}$$

*Then*

$$\begin{aligned}
\mathbb{E}[X_1|Y = r] &= 0 \times \mathbb{P}(X_1 = 0|Y = r) + 1 \times \mathbb{P}(X_1 = 1|Y = r) \\
&= \frac{r}{n}
\end{aligned}$$

$$\mathbb{E}[X_1|Y = Y(\omega)] = \frac{1}{n}Y(\omega)$$

*Therefore* $\mathbb{E}[X_1|Y] = \frac{1}{n}Y$

Note *a random variable - a function of $Y$.*

## 5.3 Properties of Conditional Expectation

**Theorem 5.5.**

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

*Proof.*

$$
\begin{aligned}
\mathbb{E}[\mathbb{E}[X|Y]] &= \sum_{y\in R_y} \mathbb{P}(Y=y)\,\mathbb{E}[X|Y=y] \\
&= \sum_y \mathbb{P}(Y=y) \sum_{x\in R_x} \mathbb{P}(X=x|Y=y) \\
&= \sum_y \sum_x x\mathbb{P}(X=x|Y=y) \\
&= \mathbb{E}[X]
\end{aligned}
$$

□

**Theorem 5.6.** *If $X$ and $Y$ are independent then*

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

*Proof.* If $X$ and $Y$ are independent then for any $y \in R_y$

$$\mathbb{E}[X|Y=y] = \sum_{x\in R_x} x\mathbb{P}(X=x|Y=y) = \sum_x x\mathbb{P}(X=x) = \mathbb{E}[X]$$

□

**Example.** *Let $X_1, X_2, \dots$ be i.i.d.r.v's with p.g.f $p(z)$. Let N be a random variable independent of $X_1, X_2, \dots$ with p.g.f $h(z)$. What is the p.g.f of:*

$$X_1 + X_2 + \cdots + X_N$$

$$
\begin{aligned}
\mathbb{E}\left[z^{X_1+,\dots,X_n}\right] &= \mathbb{E}\left[\mathbb{E}\left[z^{X_1+,\dots,X_n}|N\right]\right] \\
&= \sum_{n=0}^{\infty} \mathbb{P}(N=n)\,\mathbb{E}\left[z^{X_1+,\dots,X_n}|N=n\right] \\
&= \sum_{n=0}^{\infty} \mathbb{P}(N=n)\,(p(z))^n \\
&= h(p(z))
\end{aligned}
$$

*Then for example*

$$\mathbb{E}[X_1+,\dots,X_n] = \left.\frac{d}{dz}h(p(z))\right|_{z=1}$$

$$= h^{'}(1)p^{'}(1) = \mathbb{E}[N]\,\mathbb{E}[X_1]$$

*Exercise* Calculate $\frac{d^2}{dz^2}h(p(z))$ and hence

$$\mathrm{Var}\, X_1+,\dots,X_n$$

In terms of $\mathrm{Var}\, N$ and $\mathrm{Var}\, X_1$

## 5.4 Branching Processes

$X_0, X_1 \dots$ sequence of random variables. $X_n$ number of individuals in the $n^{th}$ generation of population. Assume.

1. $X_0 = 1$
2. Each individual lives for unit time then on death produces $k$ offspring, probability $f_k$. $\sum f_k = 1$
3. All offspring behave independently.

$$X_{n+1} = Y_1^n + Y_2^n + \cdots + Y_n^n$$

Where $Y_i^n$ are i.i.d.r.v's. $Y_i^n$ number of offspring of individual $i$ in generation $n$.

Assume

1. $f_0 \geq 0$
2. $f_0 + f_1 \leq 1$

Let F(z) be the probability generating function of$Y_i^n$.

$$F(z) = \sum_{n=0}^{\infty} f_k z^k = \mathbb{E}\big[z^{X_i}\big] = \mathbb{E}\Big[z^{Y_i^n}\Big]$$

Let

$$F_n(z) = \mathbb{E}\big[z^{X_n}\big]$$

Then $F_1(z) = F(z)$ the probability generating function of the offspring distribution.

**Theorem 5.7.**

$$F_{n+1}(z) = F_n(F(z)) = F(F(\dots(F(z))\dots))$$

*$F_n(z)$ is an n-fold iterative formula.*

*Proof.*

$$\begin{aligned}
F_{n+1}(z) &= \mathbb{E}\big[z^{X_{n+1}}\big] \\
&= \mathbb{E}\big[\mathbb{E}\big[z^{X_{n+1}}|X_n\big]\big] \\
&= \sum_{n=0}^{\infty} \mathbb{P}(X_n = k)\,\mathbb{E}\big[z^{X_{n+1}}|X_n = k\big] \\
&= \sum_{n=0}^{\infty} \mathbb{P}(X_n = k)\,\mathbb{E}\Big[z^{Y_1^n + Y_2^n + \cdots + Y_n^n}\Big] \\
&= \sum_{n=0}^{\infty} \mathbb{P}(X_n = k)\,\mathbb{E}\Big[z^{Y_1^n}\Big]\ldots\mathbb{E}\Big[z^{Y_n^n}\Big] \\
&= \sum_{n=0}^{\infty} \mathbb{P}(X_n = k)\,(F(z))^k \\
&= F_n(F(z))
\end{aligned}$$

□

**Theorem 5.8.** *Mean and Variance of population size*

$$\begin{aligned}
\textit{If } m &= \sum_{k=0}^{\infty} k f_k \leq \infty \\
\textit{and } \sigma^2 &= \sum_{k=0}^{\infty} (k-m)^2 f_k \leq \infty
\end{aligned}$$

*Mean and Variance of offspring distribution.*

*Then* $\mathbb{E}[X_n] = m^n$

$$\operatorname{Var} X_n = \begin{cases} \frac{\sigma^2 m^{n-1}(m^n-1)}{m-1}, & m \neq 1 \\ n\sigma^2, & m = 1 \end{cases} \tag{5.1}$$

*Proof.* Prove by calculating $F^{'}(z)$, $F^{''}(z)$ *Alternatively*

$$\begin{aligned}
\mathbb{E}[X_n] &= \mathbb{E}[\mathbb{E}[X_n|X_{n-1}]] \\
&= \mathbb{E}[m|X_{n-1}] \\
&= m\mathbb{E}[X_{n-1}] \\
&= m^n \text{ by induction} \\
\mathbb{E}\big[(X_n - mX_{n-1})^2\big] &= \mathbb{E}\big[\mathbb{E}\big[(X_n - mX_{n-1})^2|X_n\big]\big] \\
&= \mathbb{E}[\operatorname{Var}(X_n|X_{n-1})] \\
&= \mathbb{E}\big[\sigma^2 X_{n-1}\big] \\
&= \sigma^2 m^{n-1}
\end{aligned}$$

Thus

$$\mathbb{E}\big[X_n^2\big] - 2m\mathbb{E}[X_n X_{n-1}] + m^2\mathbb{E}\big[X_{n-1}^2\big]^2 = \sigma^2 m^{n-1}$$

Now calculate

$$
\begin{aligned}
\mathbb{E}[X_n X_{n-1}] &= \mathbb{E}[\mathbb{E}[X_n X_{n-1} | X_{n-1}]] \\
&= \mathbb{E}[X_{n-1}\mathbb{E}[X_n | X_{n-1}]] \\
&= \mathbb{E}[X_{n-1} m X_{n-1}] \\
&= m\mathbb{E}\left[X_{n-1}^2\right] \\
\text{Then } \mathbb{E}\left[X_n^2\right] &= \sigma^2 m^{n-1} + m^2 \mathbb{E}[X_{n-1}]^2 \\
\operatorname{Var} X_n &= \mathbb{E}\left[X_n^2\right] - \mathbb{E}[X_n]^2 \\
&= m^2\mathbb{E}\left[X_{n-1}^2\right] + \sigma^2 m^{n-1} - m^2\mathbb{E}[X_{n-1}]^2 \\
&= m^2 \operatorname{Var} X_{n-1} + \sigma^2 m^{n-1} \\
&= m^4 \operatorname{Var} X_{n-2} + \sigma^2 (m^{n-1} + m^n) \\
&= m^{2(n-1)} \operatorname{Var} X_1 + \sigma^2 (m^{n-1} + m^n + \cdots + m^{2n-3}) \\
&= \sigma^2 m^{n-1} (1 + m + \cdots + m^n)
\end{aligned}
$$

□

To deal with extinction we need to be careful with limits as $n \to \infty$. Let

$$
\begin{aligned}
A_n &= X_n = 0 \\
&= \text{Extinction occurs by generation } n \\
\text{and let } A &= \bigcup_1^\infty A_n \\
&= \text{the event that extinction ever occurs}
\end{aligned}
$$

Can we calculate $\mathbb{P}(A)$ from $\mathbb{P}(A_n)$?

More generally let $A_n$ be an increasing sequence

$$A_1 \subset A_2 \subset \dots$$

and define

$$A = \lim_{n\to\infty} A_n = \bigcup_1^\infty A_n$$

Define $B_n$ for $n \geq 1$

$$
\begin{aligned}
B_1 &= A_1 \\
B_n &= A_n \cap \left( \bigcup_{i=1}^{n-1} A_i \right)^c \\
&= A_n \cap A_{n-1}^c
\end{aligned}
$$

$B_n$ for $n \geq 1$ are disjoint events and

$$\begin{aligned}
\bigcup_{i=1}^{\infty} A_i &= \bigcup_{i=1}^{\infty} B_i \\
\bigcup_{i=1}^{n} A_i &= \bigcup_{i=1}^{n} B_i \\
\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) &= \mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i\right) \\
&= \sum_{1}^{\infty} \mathbb{P}(B_i) \\
&= \lim_{n\to\infty} \sum_{1}^{n} \mathbb{P}(B_i) \\
&= \lim_{n\to\infty} \bigcup_{i=1}^{n} B_i \\
&= \lim_{n\to\infty} \bigcup_{i=1}^{n} A_i \\
&= \lim_{n\to\infty} \mathbb{P}(A_n)
\end{aligned}$$

Thus

$$\mathbb{P}\Big(\lim_{n\to\infty} A_n\Big) = \lim_{n\to\infty} \mathbb{P}(A_n)$$

Probability is a continuous set function. Thus

$$\begin{aligned}
\mathbb{P}(\text{extinction ever occurs}) &= \lim_{n\to\infty} \mathbb{P}(A_n) \\
&= \lim_{n\to\infty} \mathbb{P}(X_n = 0) \\
&= q, \qquad \text{Say}
\end{aligned}$$

*Note* $\mathbb{P}(X_n = 0)$, $n = 1, 2, 3, \ldots$ is an increasing sequence so limit exists. But

$$\mathbb{P}(X_n = 0) = F_n(0) \qquad F_n \text{ is the p.g.f of } X_n$$

So

$$q = \lim_{n\to\infty} F_n(0)$$

Also

$$\begin{aligned}
F(q) &= F\left(\lim_{n\to\infty} F_n(0)\right) \\
&= \lim_{n\to\infty} F\left(F_n(0)\right) \qquad \text{Since F is continuous} \\
&= \lim_{n\to\infty} F_{n+1}(0)
\end{aligned}$$

Thus $F(q) = q$

"q" is called the *Extinction Probability*.

*Alternative Derivation*

$$\begin{aligned} q &= \sum_k \mathbb{P}(X_1 = k)\,\mathbb{P}(extinction|X_1 = k) \\ &= \sum \mathbb{P}(X_1 = k)\,q^k \\ &= F(q) \end{aligned}$$

**Theorem 5.9.** *The probability of extinction, $q$, is the smallest positive root of the equation $F(q) = q$. $m$ is the mean of the offspring distribution.*

$$\textit{If } m \leq 1 \textit{ then } q = 1, \textit{ while if } m \geq 1 then q \leq 1$$

*Proof.*

$$F(1) = 1 \qquad m = \sum_0^\infty k f_k^{'} = \lim_{z \to 1} F^{'}(z)$$

$$F^{''}(z) = \sum_{j=z}^{\infty} j(j-1)z^{j-2} \text{ in } 0 \leq z \leq 1 \text{ Since } f_0 + f_1 \leq 1 \text{ Also } F(0) = f_0 \geq 0$$

Thus if $m \leq 1$, there does not exists a $q \in (0,1)$ with $F(q) = q$. If $m \geq 1$ then let $\alpha$

be the smallest positive root of $F(z) = z$ then $\alpha \leq 1$. Further,

$$\begin{aligned} F(0) &\leq F(\alpha) = \alpha \\ F(F(0)) &\leq F(\alpha) = \alpha \\ F_n(0) &\leq \alpha \qquad \forall n \geq 1 \\ q &= \lim_{n \to \infty} F_n(0) \leq 0 \\ q &= \alpha \qquad \text{Since } q \text{ is a root of } F(z) = z \end{aligned}$$

□

## 5.5 Random Walks

Let $X_1, X_2, \ldots$ be i.i.d.r.vs. Let

$$S_n = S_0 + X_1 + X_2 + \cdots + X_n \qquad \text{Where, usually } S_0 = 0$$

Then $S_n$ $(n = 0, 1, 2, \ldots$ is a 1 dimensional *Random Walk*.

We shall assume

$$X_n = \begin{cases} 1, & \text{with probability } p \\ -1, & \text{with probability } q \end{cases} \tag{5.2}$$

This is a simple random walk. If $p = q = \frac{1}{2}$ then the random walk is called *symmetric*

**Example (Gambler's Ruin).** *You have an initial fortune of $A$ and I have an initial fortune of $B$. We toss coins repeatedly I win with probability $p$ and you win with probability $q$. What is the probability that I bankrupt you before you bankrupt me?*

*Set* $a = A + B$ *and* $z = B$ *Stop a random walk starting at* $z$ *when it hits* $0$ *or* $a$.

*Let* $p_z$ *be the probability that the random walk hits* $a$ *before it hits* $0$, *starting from* $z$. *Let* $q_z$ *be the probability that the random walk hits* $0$ *before it hits* $a$, *starting from* $z$. *After the first step the gambler's fortune is either* $z - 1$ *or* $z + 1$ *with prob* $p$ *and* $q$ *respectively. From the law of total probability.*

$$p_z = qp_{z-1} + pp_{z+1} \qquad 0 \leq z \leq a$$

*Also* $p_0 = 0$ *and* $p_a = 1$. *Must solve* $pt^2 - t + q = 0$.

$$t = \frac{1 \pm \sqrt{1 - 4pq}}{2p} = \frac{1 \pm \sqrt{1 - 2p}}{2p} = 1 \text{ or } \frac{q}{p}$$

*General Solution for* $p \neq q$ *is*

$$p_z = A + B\left(\frac{q}{p}\right)^z \qquad A + B = 0 A = \frac{1}{1 - \left(\frac{q}{p}\right)^a}$$

*and so*

$$p_z = \frac{1 - \left(\frac{q}{p}\right)^z}{1 - \left(\frac{q}{p}\right)^a}$$

*If* $p = q$, *the general solution is* $A + Bz$

$$p_z = \frac{z}{a}$$

*To calculate* $q_z$, *observe that this is the same problem with* $p, q, z$ *replaced by* $p, q, a - z$ *respectively. Thus*

$$q_z = \frac{\left(\frac{q}{p}\right)^a - \left(\frac{q}{p}\right)^z}{\left(\frac{q}{p}\right)^a - 1} \text{ if } p \neq q$$

*or*

$$q_z = \frac{a-z}{z} \text{ if } p = q$$

*Thus* $q_z + p_z = 1$ *and so on, as we expected, the game ends with probability one.*

$$\begin{aligned}
\mathbb{P}(\textit{hits } 0 \textit{ before } a) &= q_z \\
q_z &= \frac{\left(\frac{q}{p}\right)^a - (\frac{q}{p})^z}{\left(\frac{q}{p}\right)^a - 1} \text{ if } p \neq q \\
\textit{Or } &= \frac{a-z}{z} \text{ if } p = q
\end{aligned}$$

*What happens as* $a \to \infty$*?*

$$\begin{aligned}
\mathbb{P}(\textit{ paths hit } 0 \textit{ ever}) &= \bigcup_{a=z+1}^{\infty} \textit{path hits } 0 \textit{ before it hits } a \\
\mathbb{P}(\textit{hits } 0 \textit{ ever}) &= \lim_{a\to\infty} \mathbb{P}(\textit{hits } 0 \textit{ before } a) \\
&= \lim_{a\to\infty} q_z \\
&= \left(\frac{q}{p}\right) \qquad p \geq q \\
&= 1 \qquad p = q
\end{aligned}$$

*Let* $G$ *be the ultimate gain or loss.*

$$G = \begin{cases} a - z, & \textit{with probability } p_z \\ -z, & \textit{with probability } q_z \end{cases} \tag{5.3}$$

$$\mathbb{E}[G] = \begin{cases} ap_z - z, & \textit{if } p \neq q \\ 0, & \textit{if } p = q \end{cases} \tag{5.4}$$

Fair game remains fair *if the coin is fair then then games based on it have expected reward* $0$*.*

*Duration of a Game* Let $D_z$ be the expected time until the random walk hits $0$ or $a$, starting from $z$. Is $D_z$ finite? $D_z$ is bounded above by $x$ the mean of geometric random variables (number of window's of size a before a window with all $+1's$ or $-1's$). Hence $D_z$ is finite. Consider the first step. Then

$$\begin{aligned}
D_z &= 1 + pD_{z+1} + qD_{z-1} \\
\mathbb{E}[\text{duration}] &= \mathbb{E}[\mathbb{E}[\text{duration} \mid \text{first step}]] \\
&= p\left(\mathbb{E}[\text{duration} \mid \text{first step up}]\right) + q\left(\mathbb{E}[\text{duration} \mid \text{first step down}]\right) \\
&= p(1 + D_{z+1}) + q(1 + D_{z-1})
\end{aligned}$$

Equation holds for $0 \leq z \leq a$ with $D_0 = D_a = 0$. Let's try for a particular solution $D_z = C_z$

$$C_z = C_p(z+1) + C_q(z-1) + 1$$

$$C = \frac{1}{q-p} \qquad \text{for } p \neq q$$

Consider the homogeneous relation

$$pt^2 - t + q = 0 \qquad t_1 = 1 \qquad t_2 = \frac{q}{p}$$

General Solution for $p \neq q$ is

$$D_z = A + B\left(\frac{q}{p}\right)^z + \frac{z}{q = p}$$

Substitute $z = 0, a$ to get $A$ and $B$

$$D_z = \frac{z}{q-p} - \frac{a}{q-p}\frac{1-\left(\frac{q}{p}\right)^z}{1-\left(\frac{q}{p}\right)^a} \qquad p \neq q$$

If $p = q$ then a particular solution is $-z^2$. General solution

$$D_z - z^2 + A + Bz$$

Substituting the boundary conditions given.,

$$D_z = z(a-z) \qquad p = q$$

**Example.** *Initial Capital.*

| *p* | *q* | *z* | *a* | $\mathbb{P}(ruin)$ | $\mathbb{E}$[*gain*] | $\mathbb{E}$[*duration*] |
|---|---|---|---|---|---|---|
| *0.5* | *0.5* | *90* | *100* | *0.1* | *0* | *900* |
| *0.45* | *0.55* | *9* | *10* | *0.21* | *-1.1* | *11* |
| *0.45* | *0.55* | *90* | *100* | *0.87* | *-77* | *766* |

*Stop the random walk when it hits* $0$ *or* $a$.
*We have* absorption *at* $0$ *or* $a$. *Let*

$$U_{z,n} = \mathbb{P}(\textit{r.w. hits 0 at time n—starts at z})$$

$$U_{z,n+1} = pU_{z+1,n} + qU_{z-1,n} \qquad 0 \leq z \leq a \qquad n \geq 0$$
$$U_{0,n} = U_{a,n} = 0 \qquad n \geq 0$$
$$U_{a,0} = 1U_{z,0} = 0 \qquad 0 \leq z \leq a$$
$$\textit{Let } U_z = \sum_{n=0}^{\infty} U_{z,n}s^n.$$

*Now multiply by* $s^{n+1}$ *and add for* $n = 0, 1, 2\ldots$

$$U_z(s) = psU_{z+1}(s) + qsU_{z-1}(s)$$

*Where* $U_0(s) = 1$ *and* $U_a(s) = 0$

*Look for a solution*

$$U_x(s) = (\lambda(s))^z\,\lambda(s)$$

*Must satisfy*

$$\lambda(s) = ps\left((\lambda(s))^2 + qs\right.$$

*Two Roots,*

$$\lambda_1(s),\ \lambda_2(s) = \frac{1 \pm \sqrt{1 - 4pqs^2}}{2ps}$$

*Every Solution of the form*

$$U_z(s) = A(s)\left(\lambda_1(s)\right)^z + B(s)\left(\lambda_2(s)\right)^z$$

*Substitute* $U_0(s) = 1$ *and* $U_a(s) = 0.A(s) + B(s) = 1$ *and*

$$A(s)\left(\lambda_1(s)\right)^a + B(s)\left(\lambda_2(s)\right)^a = 0$$

$$U_z(s) = \frac{\left(\lambda_1(s)\right)^a\left(\lambda_2(s)\right)^z - \left(\lambda_1(s)\right)^z\left(\lambda_2(s)\right)^a}{\left(\lambda_1(s)\right)^a - \left(\lambda_2(s)\right)^a}$$

*But* $\lambda_1(s)\lambda_2(s) = \dfrac{q}{p}$ *recall quadratic*

$$U_z(s) = \left(\frac{q}{p}\right)\frac{\left(\lambda_1(s)\right)^{a-z} - \left(\lambda_2(s)\right)^{a-z}}{\left(\lambda_1(s)\right)^a - \left(\lambda_2(s)\right)^a}$$

*Same method give generating function for absorption probabilities at the other barrier. Generating function for the duration of the game is the sum of these two generating functions.*

# Chapter 6

# Continuous Random Variables

In this chapter we drop the assumption that $\Omega$ id finite or countable. Assume we are given a probability $p$ on some subset of $\Omega$.

For example, spin a pointer, and let $\omega \in \Omega$ give the position at which it stops, with $\Omega = \omega : 0 \leq \omega \leq 2\pi$. Let

$$\mathbb{P}(\omega \in [0, \theta]) = \frac{\theta}{2\pi} \qquad (0 \leq \theta \leq 2\pi)$$

**Definition 6.1.** *A continuous random variable $X$ is a function $X : \Omega \to \mathbb{R}$ for which*

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

*Where $f(x)$ is a function satisfying*

1. $f(x) \geq 0$
2. $\int_{-\infty}^{+\infty} f(x)dx = 1$

*The function f is called the* Probability Density Function.

For example, if $X(\omega) = \omega$ given position of the pointer then x is a continuous random variable with p.d.f

$$f(x) = \begin{cases} \frac{1}{2\pi}, & (0 \leq x \leq 2\pi) \\ 0, & \text{otherwise} \end{cases} \tag{6.1}$$

This is an example of a uniformly distributed random variable. On the interval $[0, 2\pi]$

in this case. Intuition about probability density functions is based on the approximate relation.

$$\mathbb{P}(X \in [x, x + x\delta x]) = \int_x^{x+x\delta x} f(z)dz$$

Proofs however more often use the distribution function

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

$F(x)$ is increasing in $x$.

If $X$ is a continuous random variable then

$$F(x) = \int_{-\infty}^{x} f(z)dz$$

and so $F$ is continuous and differentiable.

$$F^{'}(x) = f(x)$$

(At any point x where then fundamental theorem of calculus applies).

The distribution function is also defined for a discrete random variable,

$$F(x) = \sum_{\omega : X(\omega) \leq x} p_\omega$$

and so F is a step function.

In either case

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

**Example.** *The exponential distribution. Let*

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & 0 \leq x \leq \infty \\ 0, & x \leq 0 \end{cases} \tag{6.2}$$

*The corresponding pdf is*

$$f(x) = \lambda e^{-\lambda x} \qquad 0 \leq x \leq \infty$$

*this is known as the exponential distribution with parameter $\lambda$. If $X$ has this distribution then*

$$\begin{aligned} \mathbb{P}(X \leq x + z | X \leq z) &= \frac{\mathbb{P}(X \leq x + z)}{\mathbb{P}(X \leq z)} \\ &= \frac{e^{-\lambda(x+z)}}{e^{-\lambda z}} \\ &= e^{-\lambda x} \\ &= \mathbb{P}(X \leq x) \end{aligned}$$

*This is known as the memoryless property of the exponential distribution.*

**Theorem 6.1.** *If $X$ is a continuous random variable with pdf $f(x)$ and $h(x)$ is a continuous strictly increasing function with $h^{-1}(x)$ differentiable then $h(x)$ is a continuous random variable with pdf*

$$f_h(x) = f\left(h^{-1}(x)\right) \frac{d}{dx} h^{-1}(x)$$

*Proof.* The distribution function of $h(X)$ is

$$\mathbb{P}(h(X) \leq x) = \mathbb{P}\big(X \leq h^{-1}(x)\big) = F\left(h^{-1}(x)\right)$$

Since $h$ is strictly increasing and $F$ is the distribution function of X Then.

$$\frac{d}{dx}\mathbb{P}(h(X) \leq x)$$

is a continuous random variable with pdf as claimed $f_h$. *Note* usually need to repeat proof than remember the result. $\square$

**Example.** *Suppose* $X \sim U[0,1]$ *that is it is uniformly distributed on* $[0,1]$ *Consider* $Y = -\log x$

$$\begin{aligned}\mathbb{P}(Y \leq y) &= \mathbb{P}(-\log X \leq y)\\ &= \mathbb{P}\big(X \geq e^{-Y}\big)\\ &= \int_{e^{-Y}}^{1} 1 dx\\ &= 1 - e^{-Y}\end{aligned}$$

*Thus Y is exponentially distributed.*

More generally

**Theorem 6.2.** *Let* $U \sim U[0,1]$. *For any continuous distribution function F, the random variable* $X$ *defined by* $X = F^{-1}(u)$ *has distribution function* $F$.

*Proof.*

$$\begin{aligned}\mathbb{P}(X \leq x) &= \mathbb{P}\big(F^{-1}(u) \leq x\big)\\ &= \mathbb{P}(U \leq F(x))\\ &= F(x) \sim U[0,1]\end{aligned}$$

□

*Remark*

1. a bit more messy for discrete random variables

$$\mathbb{P}(X = X_i) = p_i \qquad i = 0, 1, \dots$$

   Let

$$X = x_j \text{ if } \sum_{i=0}^{j-1} p_i \leq U \leq \sum_{i=0}^{j} p_i \qquad U \sim U[0,1]$$

2. useful for simulations

## 6.1 Jointly Distributed Random Variables

For two random variables $X$ and $Y$ the joint distribution function is

$$F(x,y) = \mathbb{P}(X \leq x, Y \leq y) \qquad F : \mathbb{R}^2 \to [0,1]$$

Let

$$\begin{aligned}F_X(x) &= \mathbb{P}(Xz \leq x)\\ &= \mathbb{P}(X \leq x, Y \leq \infty)\\ &= F(x, \infty)\\ &= \lim_{y \to \infty} F(x,y)\end{aligned}$$

This is called the marginal distribution of X. Similarly

$$F_Y(x) = F(\infty, y)$$

$X_1, X_2, \ldots, X_n$ are jointly distributed continuous random variables if for a set $c \in \mathbb{R}^b$

$$\mathbb{P}((X_1, X_2, \ldots, X_n) \in c) = \iint \ldots \int_{(x_1,\ldots,x_n)\in c} f(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

For some function f called the joint probability density function satisfying the obvious conditions.

1. $$f(x_1, \ldots, x_n) dx_1 \geq 0$$

2. $$\iint \ldots \int_{\mathbb{R}^n} f(x_1, \ldots, x_n) dx_1 \ldots dx_n = 1$$

**Example.** $(n = 2)$

$$\begin{aligned} F(x, y) &= \mathbb{P}(X \leq x, Y \leq y) \\ &= \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) du dv \\ \textit{and so } f(x, y) &= \frac{\partial^2 F(x, y)}{\partial x \partial y} \end{aligned}$$

**Theorem 6.3.** *provided defined at $(x, y)$. If $X$ and $y$ are jointly continuous random variables then they are individually continuous.*

*Proof.* Since X and Y are jointly continuous random variables

$$\begin{aligned} \mathbb{P}(X \in A) = \mathbb{P}(X \in A, Y \in (-\infty, +\infty)) \\ &= \int_A \int_{-\infty}^{\infty} f(x, y) dx dy \\ &= f_A f_X(x) dx \\ \text{where } f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy \end{aligned}$$

is the pdf of $X$. □

Jointly continuous random variables $X$ and $Y$ are *Independent* if

$$\begin{aligned} f(x, y) = f_X(x) f_Y(y) \\ \text{Then } \mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\, \mathbb{P}(Y \in B) \end{aligned}$$

Similarly jointly continuous random variables $X_1, \ldots, X_n$ are independent if

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

Where $f_{X_i}(x_i)$ are the pdf's of the individual random variables.

**Example.** *Two points $X$ and $Y$ are tossed at random and independently onto a line segment of length L. What is the probability that:*

$$|X - Y| \leq l?$$

*Suppose that "at random" means uniformly so that*

$$f(x, y) = \frac{1}{L^2} \qquad x, y \in [0, L]^2$$

*Desired probability*

$$
\begin{aligned}
&= \iint_A f(x,y)dxdy \\
&= \frac{\textit{area of } A}{L^2} \\
&= \frac{L^2 - 2\frac{1}{2}(L-l)^2}{L^2} \\
&= \frac{2Ll - l^2}{L^2}
\end{aligned}
$$

**Example (Buffon's Needle Problem).** *A needle of length l is tossed at random onto a floor marked with parallel lines a distance L apart* $l \leq L$*. What is the probability that the needle intersects one of the parallel lines.*

*Let* $\theta \in [0, 2\pi]$ *be the angle between the needle and the parallel lines and let* $x$ *be the distance from the bottom of the needle to the line closest to it. It is reasonable to suppose that X is distributed* Uniformly.

$$X \sim U[0, L] \qquad \Theta \sim U[0, \pi)$$

*and* $X$ *and* $\Theta$ *are independent. Thus*

$$f(x, \theta) = \frac{1}{l\pi} 0 \leq x \leq L \textit{ and } 0 \leq \theta \leq \pi$$

*The needle intersects the line if and only if* $X \leq \sin\theta$ *The event A*

$$
\begin{aligned}
&= \iint_A f(x,\theta)dxd\theta \\
&= l\int_0^\pi \frac{\sin\theta}{\pi L}d\theta \\
&= \frac{2l}{\pi L}
\end{aligned}
$$

**Definition 6.2.** *The* expectation *or mean of a continuous random variable* $X$ *is*

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

*provided not both of* $\int_{-\infty}^{\infty} xf(x)dx$ *and* $\int_{-\infty}^{0} xf(x)dx$ *are infinite*

**Example (Normal Distribution).** *Let*

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad -\infty \leq x \leq \infty$$

*This is non-negative for it to be a pdf we also need to check that*

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

*Make the substitution* $z = \frac{x-\mu}{\sigma}$. *Then*

$$
\begin{aligned}
I &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{\frac{-(x-\mu)^2}{2\sigma^2}}dx \\
&= \frac{1}{\sqrt{2\pi}}\int\int_{-\infty}^{\infty} e^{\frac{-z^2}{2}}dz
\end{aligned}
$$

*Thus* $I^2 = \frac{1}{2\pi}\left[\int_{-\infty}^{\infty} e^{\frac{-x^2}{2}}dx\right]\left[\int_{-\infty}^{\infty} e^{\frac{-y^2}{2}}dy\right]$

$$
\begin{aligned}
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{\frac{-(y^2+x^2)}{2}}dxdy \\
&= \frac{1}{2\pi}\int_0^{2\pi}\int_0^{\infty} re^{\frac{-\theta^2}{2}}drd\theta \\
&= \int_0^{2\pi} d\theta = 1
\end{aligned}
$$

*Therefore* $I = 1$. *A random variable with the pdf f(x) given above has a* Normal distribution *with parameters* $\mu$ *and* $\sigma^2$ *we write this as*

$$X \sim N[\mu, \sigma^2]$$

*The Expectation is*

$$
\begin{aligned}
\mathbb{E}[X] &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} xe^{\frac{-(x-\mu)^2}{2\sigma^2}}dx \\
&= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} (x-\mu)e^{\frac{-(x-\mu)^2}{2\sigma^2}}dx + \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} \mu e^{\frac{-(x-\mu)^2}{2\sigma^2}}dx.
\end{aligned}
$$

*The first term is convergent and equals zero by symmetry, so that*

$$\begin{aligned}\mathbb{E}[X] &= 0 + \mu \\ &= \mu\end{aligned}$$

**Theorem 6.4.** *If $X$ is a continuous random variable then,*

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X \geq x)\,dx - \int_0^\infty \mathbb{P}(X \leq -x)\,dx$$

*Proof.*

$$\begin{aligned}\int_0^\infty \mathbb{P}(X \geq x)\,dx &= \int_0^\infty \left[\int_x^\infty f(y)dy\right] dx \\ &= \int_0^\infty \int_0^\infty I[y \geq x] f(y) dy dx \\ &= \int_0^\infty \int_0^y dx f(y) dy \\ &= \int_0^\infty y f(y) dy\end{aligned}$$

$$\text{Similarly} \int_0^\infty \mathbb{P}(X \leq -x)\,dx = \int_{-\infty}^0 y f(y) dy$$

result follows. □

*Note* This holds for discrete random variables and is useful as a general way of finding the expectation whether the random variable is discrete or continuous.

If $X$ takes values in the set $[0, 1, \ldots, ]$ Theorem states

$$\mathbb{E}[X] = \sum_{n=0}^\infty \mathbb{P}(X \geq n)$$

and a direct proof follows

$$\begin{aligned}\sum_{n=0}^\infty \mathbb{P}(X \geq n) &= \sum_{n=0}^\infty \sum_{m=0}^\infty I[m \geq n] \mathbb{P}(X = m) \\ &= \sum_{m=0}^\infty \left(\sum_{n=0}^\infty I[m \geq n]\right) \mathbb{P}(X = m) \\ &= \sum_{m=0}^\infty m \mathbb{P}(X = m)\end{aligned}$$

**Theorem 6.5.** *Let $X$ be a continuous random variable with pdf $f(x)$ and let $h(x)$ be a continuous real-valued function. Then provided*

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

$$\mathbb{E}[h(x)] = \int_{-\infty}^\infty h(x) f(x) dx$$

*Proof.*

$$
\begin{aligned}
\int_0^\infty \mathbb{P}(h(X) \geq y)\, dy \\
&= \int_0^\infty \left[ \int_{x:h(x)\geq 0} f(x)dx \right] dy \\
&= \int_0^\infty \int_{x:h(x)\geq 0} I[h(x) \geq y] f(x) dx dy \\
&= \int_{x:h(x)} \left[ \int_0^{h(x)\geq 0} dy \right] f(x)dx \\
&= \int_{x:h(x)\geq 0} h(x) f(x) dy
\end{aligned}
$$

$$
\text{Similarly } \int_0^\infty \mathbb{P}(h(X) \leq -y) = -\int_{x:h(x)\leq 0} h(x) f(x) dy
$$

So the result follows from the last theorem. □

**Definition 6.3.** *The variance of a continuous random variable $X$ is*

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

Note *The properties of expectation and variance are the same for discrete and continuous random variables just replace $\sum$ with $\int$ in the proofs.*

**Example.**

$$
\begin{aligned}
\operatorname{Var} X &= \mathbb{E}\big[X^2\big] - \mathbb{E}[X]^2 \\
&= \int_{-\infty}^\infty x^2 f(x)dx - \left( \int_{-\infty}^\infty x f(x) dx \right)^2
\end{aligned}
$$

**Example.** *Suppose $X \sim N[\mu, \sigma^2]$ Let $z = \frac{X-\mu}{\sigma}$ then*

$$
\begin{aligned}
\mathbb{P}(Z \leq z) &= \mathbb{P}\left( \frac{X-\mu}{\sigma} \leq z \right) \\
&= \mathbb{P}(X \leq \mu + \sigma z) \\
&= \int_{-\infty}^{\mu+\sigma z} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx \\
\textit{Let } \left( u = \frac{x-\mu}{\sigma} \right) &= \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{\frac{-u^2}{2}} du \\
&= \Phi(z) \text{ The distribution function of a } N(0,1) \text{ random variable} \\
Z &\sim N(0,1)
\end{aligned}
$$

*What is the variance of $Z$?*

$$\begin{aligned}
\operatorname{Var} X &= \mathbb{E}\left[Z^2\right] - \mathbb{E}[Z]^2 \text{ \textit{Last term is zero}} \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{\frac{-z^2}{2}} \, dz \\
&= \left[-\frac{1}{\sqrt{2\pi}} z e^{\frac{-z^2}{2}}\right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{\frac{-z^2}{2}} \, dz \\
&= 0 + 1 = 1 \\
\operatorname{Var} X &= 1
\end{aligned}$$

Variance of $X$?

$$\begin{aligned}
X &= \mu + \sigma z \\
\text{Thus } \mathbb{E}[X] &= \mu \text{ we know that already} \\
\operatorname{Var} X &= \sigma^2 \operatorname{Var} Z \\
\operatorname{Var} X &= \sigma^2 \\
X &\sim (\mu, \sigma^2)
\end{aligned}$$

## 6.2 Transformation of Random Variables

Suppose $X_1, X_2, \ldots, X_n$ have joint pdf $f(x_1, \ldots, x_n)$ let

$$\begin{aligned}
Y_1 &= r_1(X_1, X_2, \ldots, X_n) \\
Y_2 &= r_2(X_1, X_2, \ldots, X_n) \\
&\vdots \\
Y_n &= r_n(X_1, X_2, \ldots, X_n)
\end{aligned}$$

Let $R \in \mathbb{R}^n$ be such that

$$\mathbb{P}((X_1, X_2, \ldots, X_n) \in R) = 1$$

Let $S$ be the image of $R$ under the above transformation suppose the transformation from $R$ to $S$ is 1-1 (bijective).

Then $\exists$ inverse functions

$$x_1 = s_1(y_1, y_2, \ldots, y_n)$$
$$x_2 = s_2(y_1, y_2, \ldots, y_n) \ldots$$
$$x_n = s_n(y_1, y_2, \ldots, y_n)$$

Assume that $\frac{\partial s_i}{\partial y_j}$ exists and is continuous at every point $(y_1, y_2, \ldots, y_n)$ in $S$

$$J = \begin{vmatrix} \frac{\partial s_1}{\partial y_1} & \cdots & \frac{\partial s_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial s_n}{\partial y_1} & \cdots & \frac{\partial s_n}{\partial y_n} \end{vmatrix} \tag{6.3}$$

If $A \subset R$

$$\begin{aligned} \mathbb{P}((X_1, \ldots, X_n) \in A)\,[1] &= \int_A \cdots \int f(x_1, \ldots, x_n) dx_1 \ldots dx_n \\ &= \int_B \cdots \int f\left(s_1, \ldots, s_n\right) |J|\, dy_1 \ldots dy_n \end{aligned}$$

Where B is the image of A

$$= \mathbb{P}((Y_1, \ldots, Y_n) \in B)\,[2]$$

Since transformation is 1-1 then [1],[2] are the same

Thus the density for $Y_1, \ldots, Y_n$ is

$$g((y_1, y_2, \ldots, y_n) = f\left(s_1(y_1, y_2, \ldots, y_n), \ldots, s_n(y_1, y_2, \ldots, y_n)\right) |J|$$
$$y_1, y_2, \ldots, y_n \in S$$
$$= 0 \text{ otherwise.}$$

**Example (density of products and quotients).** *Suppose that $(X, Y)$ has density*

$$f(x, y) = \begin{cases} 4xy, & \textit{for } 0 \le x \le 1, 0 \le y \le 1 \\ 0, & \textit{Otherwise.} \end{cases} \tag{6.4}$$

*Let $U = \frac{X}{Y}$ and $V = XY$*

$$
\begin{aligned}
X &= \sqrt{UV} & Y &= \sqrt{\frac{V}{U}} \\
x &= \sqrt{uv} & y &= \sqrt{\frac{v}{u}} \\
\frac{\partial x}{\partial u} &= \frac{1}{2}\sqrt{\frac{v}{u}} & \frac{\partial x}{\partial v} &= \frac{1}{2}\sqrt{\frac{u}{v}} \\
\frac{\partial y}{\partial u} &= \frac{-1}{2}\frac{v^{\frac{1}{2}}}{u^{\frac{3}{2}}} & \frac{\partial y}{\partial v} &= \frac{1}{2\sqrt{uv}}.
\end{aligned}
$$

*Therefore* $|J| = \frac{1}{2u}$ *and so*

$$
\begin{aligned}
g(u, v) &= \frac{1}{2u}(4xy) \\
&= \frac{1}{2u} \times 4\sqrt{uv}\sqrt{\frac{v}{u}} \\
&= 2\frac{u}{v} \qquad \textit{if } (u, v) \in D \\
&= 0 \qquad \textit{Otherwise}.
\end{aligned}
$$

*Note* $U$ and $V$ are NOT independent

$$g(u, v) = 2\frac{u}{v} I[(u, v) \in D]$$

not product of the two identities.

When the transformations are linear things are simpler still. Let $A$ be the $n \times n$ invertible matrix.

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = A \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}.$$

$$|J| = \det A^{-1} = \det A^{-1}$$

Then the pdf of $(Y_1, \ldots, Y_n)$ is

$$g(y_1, \ldots,_n) = \frac{1}{\det A} f(A^{-1}g)$$

**Example.** *Suppose* $X_1, X_2$ *have the pdf* $f(x_1, x_2)$. *Calculate the pdf of* $X_1 + X_2$.

*Let* $Y = X_1 + X_2$ *and* $Z = X_2$. *Then* $X_1 = Y - Z$ *and* $X_2 = Z$.

$$A^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \tag{6.5}$$

$$\det A^{-1} = 1\frac{1}{\det A}$$

*Then*

$$g(y, z) = f(x_1, x_2) = f(y - z, z)$$

*joint distributions of $Y$ and $X$.*

*Marginal density of Y is*

$$g(y) = \int_{-\infty}^{\infty} f(y-z, z)dz \qquad -\infty \leq y \leq \infty$$

$$\text{or } g(y) = \int_{-\infty}^{\infty} f(z, y-z)dz \text{ By change of variable}$$

*If $X_1$ and $X_2$ are independent, with pgf's $f_1$ and $f_2$ then*

$$f(x_1, x_2) = f(x_1)f(x_2)$$

$$\text{and then } g(y) = \int_{-\infty}^{\infty} f(y-z)f(z)dz$$

*- the convolution of $f_1$ and $f_2$*

*For the pdf f(x) $\hat{x}$ is a mode if $f(\hat{x}) \geq f(x) \forall x$*

*$\hat{x}$ is a median if*

$$\int_{-\infty}^{\hat{x}} f(x)dx - \int_{\hat{x}}^{\infty} f(x)dx = \frac{1}{2}$$

*For a discrete random variable, $\hat{x}$ is a median if*

$$\mathbb{P}(X \leq \hat{x}) \geq \frac{1}{2} \text{ or } \mathbb{P}(X \geq \hat{x}) \geq \frac{1}{2}$$

*If $X_1, \ldots, X_n$ is a sample from the distribution then recall that the sample mean is*

$$\frac{1}{n}\sum_{1}^{n} X_i$$

*Let $Y_1, \ldots, Y_n$ (the* statistics*) be the values of $X_1, \ldots, X_n$ arranged in increasing order. Then the sample median is $Y_{\frac{n+1}{2}}$ if $n$ is odd or any value in*

$$\left[Y_{\frac{n}{2}}, Y_{\frac{n+1}{2}}\right] \text{ if n is even}$$

*If $Y_n = max X_1, \ldots, X_n$ and $X_1, \ldots, X_n$ are iidrv's with distribution $F$ and density $f$ then,*

$$\begin{aligned}\mathbb{P}(Y_n \leq y) &= \mathbb{P}(X_1 \leq y, \ldots, X_n \leq y) \\ &= (F(y))^n\end{aligned}$$

*Thus the density of $Y_n$ is*

$$\begin{aligned} g(y) &= \frac{d}{dy}\left(F(y)\right)^n \\ &= n\left(F(y)\right)^{n-1} f(y) \end{aligned}$$

*Similarly $Y_1 = minX_1, \dots, X_n$ and is*

$$\begin{aligned} \mathbb{P}(Y_1 \le y) &= 1 - \mathbb{P}(X_1 \ge y, \dots, X_n \ge y) \\ &= 1 - \left(1 - F(y)\right)^n \end{aligned}$$

*Then the density of $Y_1$ is*

$$= n\left(1 - F(y)\right)^{n-1} f(y)$$

*What about the joint density of $Y_1, Y_n$?*

$$\begin{aligned} G(y, y_n) &= \mathbb{P}(Y_1 \le y_1, Y_n \le y_n) \\ &= \mathbb{P}(Y_n \le y_n) - \mathbb{P}(Y_n \le y_n, Y_1 \ge_1) \\ &= \mathbb{P}(Y_n \le y_n) - \mathbb{P}(y_1 \le X_1 \le y_n, y_1 \le X_2 \le y_n, \dots, y_1 \le X_n \le y_n) \\ &= \left(F(y_n)\right)^n - \left(F(y_n) - F(y_1)\right)^n \end{aligned}$$

*Thus the pdf of $Y_1, Y_n$ is*

$$\begin{aligned} g(y_1, y_n) &= \frac{\partial^2}{\partial y_1 \partial y_n} G(y_1, y_n) \\ &= n(n-1)\left(F(y_n) - F(y_1)\right)^{n-2} f(y_1)f(y_n) \qquad -\infty \le y_1 \le y_n \le \infty \\ &= 0 \qquad \textit{otherwise} \end{aligned}$$

*What happens if the mapping is not 1-1? $X = f(x)$ and $|X| = g(x)$?*

$$\mathbb{P}(|X| \in (a, b)) = \int_a^b \left(f(x) + f(-x)\right) dx \quad g(x) = f(x) + f(-x)$$

*Suppose $X_1, \dots, X_n$ are iidrv's. What is the pdf of $Y_1, \dots, Y_n$ the order statistics?*

$$g(y_1, \dots, y_n) = \begin{cases} n! f(y_1) \dots f(y_n), & y_1 \le y_2 \le \dots \le y_n \\ 0, & \textit{Otherwise} \end{cases} \tag{6.6}$$

**Example.** *Suppose $X_1, \dots, X_n$ are iidrv's exponentially distributed with parameter $\lambda$. Let*

$$\begin{aligned} z_1 &= Y_1 \\ z_2 &= Y_2 - Y_1 \\ &\vdots \\ z_n &= Y_n - Y_{n-1} \end{aligned}$$

*Where $Y_1, \dots, Y_n$ are the order statistics of $X_1, \dots, X_n$. What is the distribution of the $z's$?*

$Z = AY$
*Where*

$$A = \begin{pmatrix} 1 & 0 & 0 & \dots 0 & 0 \\ -1 & 1 & 0 & \dots 0 & 0 \\ 0 & -1 & 1 & \dots 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & -1 & 1 \end{pmatrix} \tag{6.7}$$

$det(A) = 1$

$$\begin{aligned} h(z_1, \dots, z_n) &= g(y_1, \dots, y_n) \\ &= n!f(y_1) \dots f(y_n) \\ &= n!\lambda^n e^{-\lambda y_1} \dots e^{-\lambda y_n} \\ &= n!\lambda^n e^{-\lambda(y_1 + \dots + y_n)} \\ &= n!\lambda^n e^{-\lambda(z_1 2z_2 + \dots + nz_n)} \\ &= \prod_{i=1}^{n} \lambda i e^{-\lambda i z_{n+1-i}} \end{aligned}$$

*Thus $h(z_1, \dots, z_n)$ is expressed as the product of n density functions and*

$$Z_{n+1-i} \sim exp(\lambda i)$$

*exponentially distributed with parameter $\lambda i$, with $z_1, \dots, z_n$ independent.*

**Example.** *Let $X$ and $Y$ be independent $N(0.1)$ random variables. Let*

$$D = R^2 = X^2 + Y_2$$

*then* $\tan \Theta = \frac{Y}{X}$ *then*

$$d = x^2 + y^2 \textit{ and } \theta = \arctan\left(\frac{y}{x}\right)$$

$$|J| = \begin{vmatrix} 2x & 2y \\ \frac{\frac{-y}{x^2}}{1+\left(\frac{y}{x}\right)^2} & \frac{\frac{1}{x}}{1+\left(\frac{y}{x}\right)^2} \end{vmatrix} = 2 \tag{6.8}$$

$$\begin{aligned} f(x,y) &= \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \frac{1}{\sqrt{2\pi}} e^{\frac{-y^2}{2}} \\ &= \frac{1}{2\pi} e^{\frac{-(x^2+y^2)}{2}} \end{aligned}$$

*Thus*

$$g(d,\theta) = \frac{1}{4\pi} e^{\frac{-d}{2}} \quad 0 \le d \le \infty \quad 0 \le \theta \le 2\pi$$

*But this is just the product of the densities*

$$\begin{aligned} g_D(d) &= \frac{1}{2} e^{\frac{-d}{2}} \qquad 0 \le d \le \infty \\ g_\Theta(\theta) &= \frac{1}{2\pi} \qquad 0 \le \theta \le 2\pi \end{aligned}$$

*Then* $D$ *and* $\Theta$ *are independent.* $d$ $\sim$*exponentially mean 2.* $\Theta \sim U[0, 2\pi]$.

Note *this is useful for the simulations of the normal random variable.*

*We know we can simulate* $N[0,1]$ *random variable by* $X = f'(U)$ *when* $U \sim U[0,1]$ *but this is difficult for* $N[0,1]$ *random variable since*

$$F(x) = \Theta(x) = \int_{-\infty}^{+x} \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$$

*is difficult.*

*Let* $U_1$ *and* $U_2$ *be independent* $\sim$ $U[0,1]$. *Let* $R^2 = -2\log U$, *so that* $R^2$ *is exponential with mean 2.* $\Theta = 2\pi U_2$. *Then* $\Theta \sim U[0, 2\pi]$. *Now let*

$$\begin{aligned} X &= R\cos\Theta = \sqrt{-2\log U_1}\cos(2\pi U_2) \\ Y &= R\sin\Theta = \sqrt{-2\log U_2}\sin(2\pi U_1) \end{aligned}$$

*Then* $X$ *and* $Y$ *are independent* $N[0,1]$ *random variables.*

**Example (Bertrand's Paradox).** *Calculate the probability that a "random chord" of a circle of radius 1 has length greater that* $\sqrt{3}$. *The length of the side of an inscribed equilateral triangle.*

*There are at least 3 interpretations of a random chord.*

*(1) The ends are independently and uniformly distributed over the circumference.*

*answer* $= \frac{1}{3}$

*(2)The chord is perpendicular to a given diameter and the point of intersection is uniformly distributed over the diameter.*

$$a^2 + \left(\frac{\sqrt{3}}{2}\right)^2 = \left(\sqrt{3}\right)^2$$

*answer* $= \frac{1}{2}$

*(3) The foot of the perpendicular to the chord from the centre of the circle is uniformly distributed over the diameter of the interior circle.*

*interior circle has radius* $\frac{1}{2}$.

$$\textit{answer } = \frac{\pi\left(\frac{1}{2^2}\right)}{\pi 1^2} = \frac{1}{4}$$

## 6.3 Moment Generating Functions

If X is a continuous random variable then the analogue of the pgf is the moment generating function defined by

$$m(\theta) = \mathbb{E}\big[e^{\theta x}\big]$$

for those $\theta$ such that $m(\theta)$ is finite

$$m(\theta) = \int_{-\infty}^{\infty} e^{\theta x} f(x) dx$$

where $f(x)$ is the pdf of $X$.

**Theorem 6.6.** *The moment generating function determines the distribution of $X$, provided $m(\theta)$ is finite for some interval containing the origin.*

*Proof.* Not proved. □

**Theorem 6.7.** *If $X$ and $Y$ are independent random variables with moment generating function $m_x(\theta)$ and $m_y(\theta)$ then $X + Y$ has the moment generating function*

$$m_{x+y}(\theta) = m_x(\theta) \times m_y(\theta)$$

*Proof.*

$$\begin{aligned} \mathbb{E}\left[e^{\theta(x+y)}\right] &= \mathbb{E}\left[e^{\theta x} e^{\theta y}\right] \\ &= \mathbb{E}\left[e^{\theta x}\right] \mathbb{E}\left[e^{\theta y}\right] \\ &= m_x(\theta) m_y(\theta) \end{aligned}$$

□

**Theorem 6.8.** *The $r^{th}$ moment of $X$ ie the expected value of $X^r$, $\mathbb{E}[X^r]$, is the coefficient of $\frac{\theta^r}{r!}$ of the series expansion of $n(\theta)$.*

*Proof.* Sketch of....

$$\begin{aligned} e^{\theta X} &= 1 + \theta X + \frac{\theta^2}{2!} X^2 + \dots \\ \mathbb{E}\left[e^{\theta X}\right] &= 1 + \theta \mathbb{E}[X] + \frac{\theta^2}{2!} \mathbb{E}\left[X^2\right] + \dots \end{aligned}$$

□

**Example.** *Recall $X$ has an exponential distribution, parameter $\lambda$ if it has a density $\lambda e^{\lambda x}$ for $0 \le x \le \infty$.*

$$\begin{aligned} \mathbb{E}\left[e^{\theta X}\right] &= \int_0^\infty e^{\theta x} \lambda e^{\lambda x} dx \\ &= \lambda \int_0^\infty e^{-(\lambda - \theta) x} dx \\ &= \frac{\lambda}{\lambda - \theta} = m(\theta) \text{ for } \theta \le \lambda \\ \mathbb{E}[X] = m^{'}(0) &= \left[\frac{\lambda}{(\lambda - \theta)^2}\right]_{\theta=0} = \frac{1}{\lambda} \\ \mathbb{E}\left[X^2\right] &= \left[\frac{2\lambda}{(\lambda - \theta)^2}\right]_{\theta=0} = \frac{2}{\lambda^2} \end{aligned}$$

*Thus*

$$\begin{aligned} \operatorname{Var} X &= \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 \\ &= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} \end{aligned}$$

**Example.** *Suppose $X_1, \dots, X_n$ are iidrvs each exponentially distributed with parameter $\lambda$.*

*Claim : $X_1, \dots, X_n$ has a gamma distribution, $\Gamma(n, \lambda)$ with parameters $n, \lambda$. With density*

$$\frac{\lambda^n e^{-\lambda x} x^{n-1}}{(n-1)!} \qquad 0 \leq x \leq \infty$$

*we can check that this is a density by integrating it by parts and show that it equals 1.*

$$\begin{aligned} \mathbb{E}\Big[e^{\theta(X_1+\dots+X_n)}\Big] &= \mathbb{E}\big[e^{\theta X_1}\big] \dots \mathbb{E}\big[e^{\theta X_n}\big] \\ &= \big[\mathbb{E}\big[e^{\theta X_1}\big]\big]^n \\ &= \left(\frac{\lambda}{\lambda-\theta}\right)^n \end{aligned}$$

*Suppose that $Y \sim \Gamma(n, \lambda)$.*

$$\begin{aligned} \mathbb{E}\big[e^{\theta Y}\big] &= \int_0^\infty e^{\theta x} \frac{\lambda^n e^{-\lambda x} x^{n-1}}{(n-1)!} dx \\ &= \left(\frac{\lambda}{\lambda-\theta}\right)^n \int_0^\infty \frac{(\lambda-\theta)^n e^{-(\lambda-\theta)x} x^{n-1}}{(n-1)!} dx \end{aligned}$$

*Hence claim, since moment generating function characterizes distribution.*

**Example (Normal Distribution).** $X \sim N[0, 1]$

$$\begin{aligned} \mathbb{E}\big[e^{\theta X}\big] &= \int_{-\infty}^{\infty} e^{\theta x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{x-\mu}{2\sigma^2}\right)^2} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-1}{2\sigma^2}(x^2 - 2x\mu + \mu^2 - 2\theta\sigma^2 x)\right] dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-1}{2\sigma^2}\left((x-\mu-\theta\sigma^2)^2 - 2\mu\sigma^2\theta - \theta^2\sigma^4\right)\right] dx \\ &= e^{\mu\theta+\theta^2\frac{\sigma^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-1}{2\sigma^2}(x-\mu-\theta\sigma^2)^2\right] dx \end{aligned}$$

*The integral equals 1 are it is the density of $N[\mu + \theta\sigma^2, \sigma^2]$*

$$= e^{\mu\theta+\theta^2\frac{\sigma^2}{2}}$$

*Which is the moment generating function of $N[\mu, \sigma^2]$ random variable.*

**Theorem 6.9.** *Suppose $X, Y$ are independent $X \sim N[\mu_1, \sigma_1^2]$ and $Y \sim N[\mu_2, \sigma_2^2]$ then*

1. $$X + Y \sim N[\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2]$$

2. $$aX \sim N[a\mu_1 + a^2\sigma^2]$$

*Proof.* 1.

$$\begin{aligned}\mathbb{E}\left[e^{\theta(X+Y)}\right] &= \mathbb{E}\left[e^{\theta X}\right]\mathbb{E}\left[e^{\theta Y}\right] \\ &= e^{(\mu_1\theta+\frac{1}{2}\sigma_1^2\theta^2)}e^{(\mu_2\theta+\frac{1}{2}\sigma_2^2\theta^2)} \\ &= e^{(\mu_1+\mu_2)\theta+\frac{1}{2}(\sigma_1^2+\sigma_2^2)\theta^2}\end{aligned}$$

which is the moment generating function for

$$N[\mu_1+\mu_2, \sigma_1^2+\sigma_2^2]$$

2.

$$\begin{aligned}\mathbb{E}\left[e^{\theta(aX)}\right] &= \mathbb{E}\left[e^{(\theta a)X}\right] \\ &= e^{\mu_1(\theta a)+\frac{1}{2}\sigma_1^2(\theta a)^2} \\ &= e^{(a\mu_1)\theta+\frac{1}{2}a^2\sigma_1^2\theta^2}\end{aligned}$$

which is the moment generating function of

$$N[a\mu_1, a^2\sigma_1^2]$$

□

## 6.4 Central Limit Theorem

$X_1, \dots, X_n$ iidrv's, mean $0$ and variance $\sigma^2$. $X_i$ has density

$$\operatorname{Var} X_i = \sigma^2$$

$X_1 + \cdots + X_n$ has Variance

$$\operatorname{Var} X_1 + \cdots + X_n = n\sigma^2$$

$\frac{X_1+\cdots+X_n}{n}$ has Variance

$$\operatorname{Var} \frac{X_1 + \cdots + X_n}{n} = \frac{\sigma^2}{n}$$

$\frac{X_1+\cdots+X_n}{\sqrt{n}}$ has Variance

$$\operatorname{Var} \frac{X_1 + \cdots + X_n}{\sqrt{n}} = \sigma^2$$

**Theorem 6.10.** *Let $X_1, \ldots, X_n$ be iidrv's with $\mathbb{E}[X_i] = \mu$ and $\operatorname{Var} X_i = \sigma^2 \leq \infty$.*

$$S_n = \sum_1^n X_i$$

*Then $\forall (a, b)$ such that $-\infty \leq a \leq b \leq \infty$*

$$\lim_{n \to \infty} \mathbb{P}\left(a \leq \frac{S_n - n\sigma}{\sigma\sqrt{n}} \leq b\right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} dz$$

*Which is the pdf of a $N[0, 1]$ random variable.*

*Proof.* Sketch of proof.....

WLOG take $\mu = 0$ and $\sigma^2 = 1$. we can replace $X_i$ by $\frac{X_i - \mu}{\sigma}$. mgf of $X_i$ is

$$\begin{aligned} m_{X_i}(\theta) &= \mathbb{E}\left[e^{\theta X_i}\right] \\ &= 1 + \theta \mathbb{E}[X_i] + \frac{\theta^2}{2}\mathbb{E}\left[X_i^2\right] + \frac{\theta^3}{3!}\mathbb{E}\left[X_i^3\right] + \dots \\ &= 1 + \frac{\theta^2}{2} + \frac{\theta^3}{3!}\mathbb{E}\left[X_i^3\right] + \dots \end{aligned}$$

The mgf of $\frac{S_n}{\sqrt{n}}$

$$\begin{aligned} \mathbb{E}\left[e^{\theta \frac{S_n}{\sqrt{n}}}\right] &= \mathbb{E}\left[e^{\frac{\theta}{\sqrt{n}}(X_1 + \dots + X_n)}\right] \\ &= \mathbb{E}\left[e^{\frac{\theta}{\sqrt{n}}X_1}\right] \dots \mathbb{E}\left[e^{\frac{\theta}{\sqrt{n}}X_n}\right] \\ &= \mathbb{E}\left[e^{\frac{\theta}{\sqrt{n}}X_1}\right]^n \\ &= \left(m_{X_1}\left(\frac{\theta}{\sqrt{n}}\right)\right)^n \\ &= \left(1 + \frac{\theta^2}{2n} + \frac{\theta^3 \mathbb{E}\left[X^3\right]}{3! n^{\frac{3}{2}}}\right)^n \qquad \Longrightarrow e^{\frac{\theta^2}{2}} \text{ as } n \to \infty \end{aligned}$$

Which is the mgf of $N[0,1]$ random variable. $\square$

*Note* if $S_n \sim Bin[n,p]$ $X_i = 1$ with probability p and $= 0$ with probability $(1-p)$. Then

$$\frac{S_n - np}{\sqrt{npq}} \simeq N[0,1]$$

This is called the normal approximation the the binomial distribution. Applies as $n \to \infty$ with $p$ constant. Earlier we discussed the Poisson approximation to the binomial. which applies when $n \to \infty$ and $np$ is constant.

**Example.** *There are two competing airlines. n passengers each select 1 of the 2 plans at random. Number of passengers in plane one*

$$S \sim Bin[n, \frac{1}{2}]$$

*Suppose each plane has s seats and let*

$$\begin{aligned} f(s) &= \mathbb{P}(S \le s) \\ \frac{S - np}{\sqrt{npq}} &\simeq n[0,1] \\ f(s) &= \mathbb{P}\left(\frac{S - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}} \le \frac{s - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}}\right) \\ &= 1 - \Phi\left(\frac{2s - n}{\sqrt{n}}\right) \end{aligned}$$

*therefore if* $n = 1000$ *and* $s = 537$ *then* $f(s) = 0.01$. *Planes hold 1074 seats only 74 in excess.*

**Example.** *An unknown fraction of the electorate, p, vote labour. It is desired to find p within an error no exceeding 0.005. How large should the sample be.*

*Let the fraction of labour votes in the sample be* $p^{'}$. *We can never be certain (without complete enumeration), that* $\left|p - p^{'}\right| \leq 0.005$. *Instead choose n so that the event* $\left|p - p^{'}\right| \leq 0.005$ *have probability* $\geq 0.95$.

$$\begin{aligned}\mathbb{P}\left(\left|p - p^{'}\right| \leq 0.005\right) &= \mathbb{P}(|S_n - np| \leq 0.005n) \\ &= \mathbb{P}\left(\frac{|S_n - np|}{\sqrt{npq}} \leq \frac{0.005\sqrt{n}}{\sqrt{n}}\right)\end{aligned}$$

*Choose n such that the probability is* $\geq 0.95$.

$$\int_{-1.96}^{1.96} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx = 2\Phi(1.96) - 1$$

*We must choose n so that*

$$\frac{0.005\sqrt{n}}{\sqrt{n}} \geq 1.96$$

*But we don't know* $p$. *But* $pq \leq \frac{1}{4}$ *with the worst case* $p = q = \frac{1}{2}$

$$n \geq \frac{1.96^2}{0.005^2}\frac{1}{4} \simeq 40,000$$

*If we replace 0.005 by 0.01 the* $n \geq 10,000$ *will be sufficient. And is we replace 0.005 by 0.045 then* $n \geq 475$ *will suffice.*

Note *Answer does not depend upon the total population.*

## 6.5 Multivariate normal distribution

Let $x_1, \ldots, X_n$ be iid $N[0,1]$ random variables with joint density $g(x_1, \ldots x_n)$

$$\begin{aligned} g(x_1, \ldots x_n) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{\frac{-x_i^2}{2}} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} e^{\frac{-1}{2} \sum_{i=1}^{n} x_i^2} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} e^{\frac{-1}{2} \vec{x} \wedge \vec{x}} \end{aligned}$$

Write

$$\vec{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$$

and let $\vec{z} = \vec{\mu} + A\vec{X}$ where $A$ is an invertible matrix $\left(\vec{x} = A^{-1}(\vec{x} - \vec{\mu})\right)$. Density of $\vec{z}$

$$\begin{aligned} f(z_1, \ldots, z_n) &= \frac{1}{(2\pi)^{\frac{n}{2}}} \frac{1}{\det A} e^{\frac{-1}{2}\left(A^{-1}(\vec{z}-\vec{\mu})\right)^T\left(A^{-1}(\vec{z}-\vec{\mu})\right)} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{\frac{-1}{2}(\vec{z}-\vec{\mu})^T \Sigma^{-1} (\vec{z}-\vec{\mu})} \end{aligned}$$

where $\Sigma AA^T$. This is the multivariate normal density

$$\vec{z} \sim MVN[\vec{\mu}, \Sigma]$$

$$Cov(z_i, z_j) = \mathbb{E}[(z_i - \mu_i)(z_j - \mu_j)]$$

But this is the $(i,j)$ entry of

$$\begin{aligned} \mathbb{E}\left[(\vec{z} - \vec{\mu})(\vec{z} - \vec{\mu})^T\right] &= \mathbb{E}\Big[(A\vec{X})(A\vec{X})^T\Big] \\ &= A\mathbb{E}\left[XX^T\right] A^T \\ &= AIA^T \qquad\qquad = AA^T = \Sigma \text{ Covariance matrix} \end{aligned}$$

If the covariance matrix of the MVN distribution is diagonal, then the components of the random vector $\vec{z}$ are independent since

$$f(z_1, \ldots, z_n) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{\frac{1}{2}} \sigma_i} e^{\frac{-1}{2}\left(\frac{z_i - \mu_i}{\sigma_i}\right)^2}$$

Where

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_n^2 \end{pmatrix}$$

Not necessarily true if the distribution is no MVN recall sheet 2 question 9.

**Example (bivariate normal).**

$$f(x_1, x_2) = \frac{1}{2\pi(1-p^2)^{\frac{1}{2}}\sigma_1\sigma_2} \times \\ \exp\left[-\frac{1}{2(1-p^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - \\ 2p\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right]\right]$$

$\sigma_1, \sigma_2 \le 0$ *and* $-1 \le p \le +1$. *Joint distribution of a* bivariate normal random variable.

**Example.** *An example with*

$$\Sigma^{-1} = \frac{1}{1-p^2}\begin{pmatrix} \sigma_1^2 & p\sigma_1^{-1}\sigma_2^{-1} \\ p\sigma_1^{-1}\sigma_2^{-1} & \sigma_2^2 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \sigma_1^2 & p\sigma_1\sigma_2 \\ p\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

$\mathbb{E}[X_i] = \mu_i$ *and* $\operatorname{Var} X_i = \sigma_i^2$. $Cov(X_1, X_2) = \sigma_1\sigma_2 p$.

$$Correlation(X_1, X_2) = \frac{Cov(X_1, X_2)}{\sigma_1\sigma_2} = p$$