1. Basic Concepts

1.1 Sample Space

A coin toss 3) A roulette wheel spin 5) Spin a pointer
2) A die throw 4) Pick for theoretical lottery

\( \Omega \), a set of all possible outcomes.

1) \( \Omega = \{ H, T \} \)
2) \( \Omega = \{ 1, 2, 3, 4, 5, 6 \} \)
3) \( \Omega = \{ 1, 2, \ldots, 36, 0 \ldots \} \)
4) \( \Omega = \{ \text{all 6-subsets of } [1, 2, 3, \ldots, 49] \} \)
5) \( \Omega = \{ 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}, 2\pi \} \)

The sample space is the set of all possible outcomes.

A subset of \( \Omega \) is called an event.

Examples of events:

1) Heads \( A = \{ H \} \)
2) Prime \( A = \{ 2, 3, 5 \} \)
3) Even \( A = \{ 2, 4, \ldots \} \)
4) Run \( A = \{ k, k+4, k+8, \ldots, k+5 \} : 1 \leq k \leq 44 \}
5) \leq 6 o'clock \( A = [0, \frac{\pi}{2}] \)

Outcome: \( \omega \in \Omega \) elementary event.

If \( \omega \) occurs in the experiment, we say "A occurs" iff \( \omega \in A \)

\[ A \cup B \] either A or B occurs
\[ A \cap B \] both A and B
\[ A \setminus B = A \cap \overline{B} \] A, but not B
\[ A \subset B \] If A, then B
\[ A = B \] equivalence
\[ A \wedge B = \emptyset \] Cannot have both A and B, mutually exclusive.

1.2 Combinatorial Probability

If \( \Omega \) is finite, \( \Omega = \{ \omega_1, \omega_2, \ldots, \omega_n \} \)

Assume each \( \omega_i \) is equally likely.

Let \( P \) be the probability

\[ P(A) = \frac{|A|}{n} \] for \( A \subseteq \Omega \)

Example: A hand of 13 cards is dealt from 52. What is the probability

that it contains:

i) Exactly one ace
ii) Exactly one ace and two kings

i) \( |A| = \binom{4}{1}\binom{48}{12} \) total # hands = \( \binom{52}{13} \), answer \( \frac{\binom{4}{1}\binom{48}{12}}{\binom{52}{13}} \)

ii) \( |A| = \binom{4}{1}\binom{4}{2}\binom{48}{9} \) answer \( \frac{\binom{4}{1}\binom{4}{2}\binom{48}{9}}{\binom{52}{13}} \)
Example: Table of random integers, of which we pick $r$.

Assume each $\omega \in \Omega$ is equiprobable.

What is the probability that:

1) No digit exceeds $k$? 
   $k \in \{0, 1, ..., 9\}$

Solution

$a)$ $P(\text{No digit } \leq k) = \frac{k^r}{10^r}$

$b)$ $P(\text{Greater digit } = k) = \frac{1}{10^r}$

1.3 Permutations and Combinations

Permutation: n objects, choose r to form an ordered subset.

$nPr = \frac{n!}{(n-r)!}$

Combination: An unordered subset of objects, a set of r.

$nCr = \frac{n!}{r!(n-r)!}$

Questions: Urn contains b blue balls and r red. Remove then at random without replacement. Find the probability that the first red ball that it is the $(k+1)^{th}$ ball overall.

Solution

Let $R_k$ be the index of the first red ball. What is the probability that $R_k = k+1$?

$p(R = k+1) = P(B_k R_k) = \frac{\text{# such sequences, length } (b+r)}{\text{total # of sequences}}$

$\frac{\binom{b+r-1}{k+1} \binom{b+r}{r-1}}{\binom{b+r}{r}}$

Menage Problem: M/W couples seated randomly at a circular table, alternating MWMWMW... . Find $p(\text{wife is seated beside her partner})$.

$= \frac{1}{n^r} \sum_{k=0}^{n} (-1)^k \binom{n}{k} \frac{2n}{n-k} \binom{2n-k}{n-k} \frac{n-k}{(n-k)!}$

for large n and r.

Example: n keys in pocket, 1 lock, pick keys at random until success.

$a)$ With replacement.

$p(\text{success on } r^{th} \text{ attempt}) = (n-1)^{r-1} \frac{1}{n} = \frac{1}{n} \left(1 - \frac{1}{n}\right)^r \approx \frac{1}{e}$

$b)$ Without replacement.

$p(\text{success on } r^{th} \text{ attempt}) = \frac{(n-1) \cdots (n-r+1) \frac{1}{n^r}}{n(n-1) \cdots \left(n-r+1\right)} = \frac{1}{n^r}$

Two Facts:

Stirling's Formula

$\frac{n!}{e^n \sqrt{2\pi n}} \to 1 \quad \text{as } n \to \infty$

Weak Version

$\frac{\log n!}{n \log n} \to 1 \quad \text{as } n \to \infty$

$log n! = \sum_{k=1}^{n} log k \leq \int_{1}^{n} \log x \, dx \leq \sum_{k=1}^{n} \log k \leq \int_{1}^{n+1} \log x \, dx$

$[x \log x - x]^n \leq \log n! \leq [x \log x]^{n+1}$
\[
\frac{n \log n - n + 1}{n \log n} \leq \log n! \leq \frac{(n+1) \log (n+1) - (n+1)+1}{n \log n}
\]
\[
\text{tends to } 1 \leq \frac{\log n!}{n \log n} \leq 1
\]

**Binomial Expansion**

\[
(1+x)^n = 1 + nx + \frac{n(n-1)}{2!} x^2 + \ldots + x^n
\]

\[
= \sum_{k=0}^{\infty} x^k \binom{n}{k} \quad \text{assuming } x \in \mathbb{R}, \ n \in \{1, 2, 3, \ldots\}
\]

\[
(1+x)^n = \sum_{k=0}^{\infty} \binom{n}{k} x^k
\]

\[
(1-x)^n = 1 - nx + \frac{n(n-1)}{2!} x^2 - \ldots + (-1)^n x^n
\]

\[
\text{If the Taylor Expansion converges correctly then}
\]

\[
f(x) = f(0) + f'(0) x + f''(0) \frac{x^2}{2!} + \ldots
\]

\[
f'(0) = a(1+0)^{n-1} = a^n
\]

\[
f''(0) = a(1-1) - (n-1) a^{n-2} = -a^{n-2} (n-1)
\]

\[
f(x) = \sum_{k=0}^{\infty} \binom{n}{k} x^k
\]

\[
f(x) = \sum_{k=0}^{\infty} \frac{n!}{k!(n-k)!} x^k
\]

\[
\text{Fine providing } |x| < 1
\]

**Example.**

Top a fair coin \(2^n\) times. \(H = \# heads\) \(Q = \{H, T\}^n\)

\[
f(H=n) = \binom{2^n}{n}
\]

\[
\text{(Stirling)} \quad \frac{n^n e^{-n} \sqrt{2n\pi}}{2^{2n}} \sim \frac{1}{\sqrt{n\pi}}
\]

\[
\frac{2^{2n} e^{n+\frac{1}{12n}}}{\pi^n 2^{n+\frac{1}{12n}}}
\]

**Exercise**

\[
p(H=n) \text{ when } 3n \text{ coins are tossed}
\]

\[
p(H=n) = \binom{3n}{n} \frac{n!}{(2n)!} 2^{-3n}
\]

\[
= \frac{(3n)!}{n! (2n)!} 2^{-3n}
\]

\[
\sim \frac{(3n)^{3n} e^{-3n} \sqrt{6\pi n}}{n^n 2^{3n} (2n)^{2n} \sqrt{2\pi n}}
\]

\[
\sim \frac{(3n)^{3n} e^{-3n} \sqrt{6\pi n}}{n^n 2^{3n} (2n)^{2n} \sqrt{2\pi n}}
\]

\[
\sim \frac{1}{n^n 2^{3n} \sqrt{2\pi n}}
\]

\[
\sim \frac{3^{3n}}{2^{5n} \sqrt{3\pi n}}
\]

\[
\sqrt{\frac{g}{f}} = \frac{\sqrt{13}}{2} \times \frac{\sqrt{13}}{2}
\]

\[
= \frac{3^{3n}}{2^{5n}} \times \sqrt{\frac{6}{8\pi}} \times \frac{1}{\sqrt{2\pi n}} \times \frac{1}{2^{3n}}
\]

\[
= \frac{3^{3n}}{2^{5n}} \times \sqrt{\frac{6}{8\pi}} \times \frac{1}{\pi n} = \frac{3^{3n}}{2^{5n}} \sqrt{\frac{3}{4\pi n}}
\]
2. Probability Space

2.1 (i) Sample Space \( \Omega \)

(ii) Collection of events \( \mathcal{F} \)

(iii) Probability function

The power set of \( \Omega \) is the set of all subsets of \( \Omega \). Denoted \( 2^\Omega \) or \( \{0, 1\}^\Omega \).

In general, the event space \( \mathcal{F} \) is not equal to \( 2^\Omega \) but not equal, in general.

Reason: If \( \Omega \) is uncountable, \( 2^\Omega \) is too big.

Reasonable conditions on events:

If \( A, B \) are events, then so are \( A \cup B, A \cap B, \Omega \setminus A \).

Definition: An event space (or \( \sigma \)-field or \( \sigma \)-algebra) is a collection \( \mathcal{F} \) of subsets of the sample space \( \Omega \) such that:

a) \( \emptyset \in \mathcal{F} \)

b) If \( A_1, A_2, \ldots \in \mathcal{F} \) then \( \bigcup_{i=1}^{\infty} A_i \in \mathcal{F} \)

c) If \( A \in \mathcal{F} \) then \( \overline{A} = \Omega \setminus A \in \mathcal{F} \)

Notes:

(i) \( \Omega = \Omega \setminus \emptyset \in \mathcal{F} \), by (a), (c) \( \mathcal{F} \) is an event space.

(ii) Finite unions of events \( \{A_i \} \) in \( \mathcal{F} \) \( (A_i = \emptyset \text{ for } i \geq n+1) \)

(iii) \[ \bigcap_{i} A_i = \bigcap_{i} \overline{A_i} \] \( \mathcal{F} \) is closed under countable intersections and also finite intersections \( (A_i = \Omega \text{ for } i \geq n+1) \)

(iv) \( A \cap B = A \cap \overline{B} \in \mathcal{F} \) if \( A, B \in \mathcal{F} \)

Similarly, \( A \Delta B = (A \cap B) \cup (B \cap \overline{A}) \)

(iv) (a) is equivalent to requiring \( \mathcal{F} \neq \emptyset \)

(Since if \( \mathcal{F} = \emptyset \), then \( A \in \mathcal{F} \) if \( A \cap \overline{A} = \emptyset \in \mathcal{F} \)).
Definition Let \( S \) be a set and \( \mathcal{F} \) an event space of \( S \). The pair 
\((S, \mathcal{F})\) is a 'measurable pair'. A probability measure is a function, \( P : \mathcal{F} \to \mathbb{R} \) such that 

a) \( 0 \leq P(A) \leq 1 \) for \( A \in \mathcal{F} \) 

b) \( P(S) = 1, \quad P(\emptyset) = 0 \) 

c) If \( A_1, A_2, \ldots \subseteq \mathcal{F} \) are pairwise disjoint.

Then \( P(\bigcup A_i) = \sum P(A_i) \) 'countable additivity' 

(Non-examinable) The problem with event spaces.

Theorem Assuming the Continuum Hypothesis, there is no measure \( \mu \) on the set of all subsets of \( I = [0, 1] \) with \( \mu(I) = 1 \), and \( \mu([x, y]) = 0 \) for \( 0 < x < y < 1 \).

Notes (i) \( P \) is finitely additive. 

\[ P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum P(A_i) \]

(ii) \( P(\emptyset) = 0 \) follows by \( \emptyset \cup \emptyset = \emptyset \Rightarrow P(\emptyset) = 0 \)

\[ \emptyset \cup \emptyset \cup \cdots = S ; \quad P(S) = P(\emptyset) + \cdots = 1 \]

Definition A probability space is \( (\Omega, \mathcal{F}, P) \) with 

a) \( \Omega \) is a set \( \rightarrow \) Experiment 

b) \( \mathcal{F} \) is an event space in \( \Omega \)

c) \( P \) is a probability measure on \( (\Omega, \mathcal{F}) \)

Example (i) Bernoulli Distribution \( \Omega = \{0, 1\} \), \( \mathcal{F} = 2^\Omega \)

\[ 0 \leq p \leq 1 \]

\[ P(A) \begin{cases} 1 - p & \text{if } A = 0 \\ p & \text{if } A = 1 \\ \emptyset & \text{if } A = \Omega \\ \emptyset & \text{if } A = \emptyset \end{cases} \]

"Two of a possible biased coin"
(ii) Combinatorial Probability

\[ \Omega = \{ \omega_1, \omega_2, \ldots, \omega_n \}, \quad n = 2^a \]

\[ A \in \mathcal{F}, \quad P(A) = \frac{|A|}{n} \]

(iii) Poisson Distribution

\[ \Omega = \{ \omega_1, \omega_2, \ldots \} \]

\[ (p_i : i \geq 1) \text{ a real sequence with } p_i \geq 0 \quad \sum p_i = 1 \]

\[ P(A) = \sum_{i=0}^{\infty} p_i \]

\[ E[\lambda] = \frac{\lambda^k}{k!} \quad c = e^{-\lambda} \quad \text{Poisson Probability, parameter } \lambda \]

**Theorem:** Let \( (\Omega, \mathcal{F}, P) \) be a probability space.

If \( A, B \in \mathcal{F} \) then

a) \( P(A) + P(A^c) = 1 \)

b) If \( B \subseteq A, \quad P(A) = P(B) + P(A \setminus B) \geq P(B) \)

c) \( P(A \cup B) = P(A) + P(B) - P(A \cap B) \)

**Proof:**

\[ A \cup B = A \cup (B \setminus A) \]

\[ B \setminus A = B \setminus (A \cap B) \quad \text{since } A \cap B \subseteq B \]

\[ P(A \cup B) = P(A) + P(B \setminus A \cap B) = P(A) + (P(B) - P(A \cap B)) \]
\[ i = A, \quad \lambda = k \]
\[ + \frac{1}{r} = (A + k) \quad k = A \]
\[ X = \beta \]

\[ \alpha = \beta \]
\[ \gamma = (A) \]

\[ \sigma = \delta \]
\[ \gamma = (A) \]

\[ i = (A + \lambda) \quad (\sigma) \]
\[ A = \beta \]

\[ (A + \lambda) \quad (A + \lambda) = (A + \lambda) \quad A = \beta \]

\[ A = (A + \lambda) \]
\[ \lambda + \alpha \]

\[ (A + \lambda) \quad (\sigma) \]
\[ \gamma = (A) \]

\[ \lambda + \alpha \]

\[ (A + \lambda) \quad (\sigma) \]
\[ \gamma = (A) \]
**Venn**

**Theorem: Inclusion-Exclusion Principle**

\[
p(A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n) = \sum_{i=1}^{n} p(A_i) - \sum_{i<j} p(A_i \cap A_j) + \sum_{i<j<k} p(A_i \cap A_j \cap A_k) - \ldots + (-1)^{n} p(A_1 \cap \ldots \cap A_n)
\]

**Proof - By Induction on \( n \)**

True for \( n = 2 \)

We assume true for \( n = k \)

\[
p(A_1 \cup \ldots \cup A_k \cup A_{k+1}) = p(A_1 \cup \ldots \cup A_k) + p(A_{k+1}) - p[(A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_k) \cap A_{k+1}]
\]

Expand using the induction hypothesis, and collect term. \( \Box \)

**Boole's Inequality** (Sub-additivity of probability)

\[
p(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} p(A_i)
\]

**Proof**

Trivial for \( n = 1 \) and use a proof by induction. \( \Box \)

(Note! Also true for a countable union if \( n = \infty \), but can't use induction)

**Bonferroni Inequality**

\[
p(\bigcap_{i=1}^{n} A_i) \geq 1 - \sum_{i=1}^{n} p(A_i) = 1 - n + \sum_{i=1}^{n} p(A_i)
\]

**Proof**

\[
p(\bigcap_{i=1}^{n} A_i) = 1 - p(U \backslash \bigcup_{i=1}^{n} A_i) \geq 1 - \sum_{i=1}^{n} p(A_i) \text{ by Boole}
\]

**Neo-Bonferroni Inequality**

\[
p(A_1 \cup A_2 \cup \ldots \cup A_n) \geq \sum_{i} p(A_i) - \sum_{i<j} p(A_i \cap A_j) - \ldots + (-1)^{n} \sum_{i_1 < \ldots < i_r} p(A_{i_1} \cap \ldots \cap A_{i_r})
\]

with \( \geq \) if \( r \) is even

\( \leq \) if \( r \) is odd.
Example (Derangement)

After dinner, the porter hands the hats to guests at random. There are $n$ hats and $n$ guests. What is the probability that nobody receives the correct hat?

Solution 

$\Omega = \{\text{permutations of } 1, 2, \ldots, n\}$

where the permutation $(\omega_1, \omega_2, \ldots, \omega_n)$ means that guest $i$ receives the hat of guest $\omega_i$.

Let $A_i = \{\omega \in \Omega : \omega_i = i\}$ = the $i^{th}$ person receives the correct hat.

We want $P(\bigcap A_i) = 1 - P(\bigcup A_i)$

$P(A_i \cap \ldots \cap A_r) = \frac{(n-r)!}{n!}$

$\sum_{i<r} P(A_i) = \frac{(n-r)!}{n!} \cdot \binom{n}{r} = \frac{1}{r!}$

$P(\bigcup A_i) = 1 - \frac{1}{2} + \frac{1}{3!} - \ldots + (-1)^{n+1} \frac{1}{n!} = 1 - P(\bigcap \bar{A}_i)$

$\lim_{n \to \infty} \frac{1}{n!} \approx e^{-1}$

Let $p_m(n) = P(\text{exactly } m \text{ people receive the correct hat})$

$= \binom{n}{m} \frac{(n-m)!}{n!} p_n^{(n-m)} = p_n^{m/m} \approx \frac{e^{-1}}{m!} \quad \text{as } n \to \infty$

The Poisson distribution with parameter $\lambda = 1$.

2.2 Conditional Probability

$[\Omega, \mathcal{F}, P]$ is a probability space.

Event $A$, probability $P(A)$

New information: A certain event $B$ has occurred.

What now is the probability of $A$?

The definition is $c P(A \cap B)$ for some $c$.
The probability of $A$ given $B$ must be 1.

$cP(A \cap B) = 1$, but $p(A \cap B) = p(B) \Rightarrow c = \frac{1}{p(B)}$

**Definition** The "conditional probability of $A$ given $B"$ denoted $p(A \mid B)$ is

\[
p(A \mid B) = \frac{p(A \cap B)}{p(B)}
\]

**Note:** Well defined if and only if $p(B) \neq 0$

**Theorem** Let $B$ satisfy $0 < p(B) < 1$

Then $p(A) = p(A \mid B) p(B) + p(A \mid B^c) p(B^c)$

**Proof**

\[
A = (A \cap B) \cup (A \cap B^c) \text{ a disjoint union,}
\]

\[
p(A) = p(A \cap B) + p(A \cap B^c)
= p(A \mid B) p(B) + p(A \mid B^c) p(B^c)
\]
\[ p(B_i | A) = \frac{p(A \cap B_i)}{p(A)} \]

\[ = \frac{\sum_{i} p(A \cap B_i) p(B_i)}{p(A)} \times \frac{p(B_i)}{p(B)} \]

\[ = \frac{p(A \cap B_i) p(B_i)}{\sum_{i} p(A \cap B_i) p(B_i)} \]
Probability

\[ p(A \mid B) = \frac{p(A \cap B)}{p(B)} \quad (\text{for } p(B) > 0) \]

More generally:

**Theorem** If \( B_1, B_2, \ldots \) is a partition of \( \Omega \) with \( p(B_i) > 0 \) \( \forall i \) then

\[ p(A) = \sum_{i} p(A \mid B_i) p(B_i) \quad (\text{the Law of Total Probability}) \]

"2 stage experiment"

**Example** A fair coin is tossed once. If heads, 1 die is tossed; if tails, 2 dice are tossed. What is the probability that the sum of any die value is four?

**Solution** \[ \Omega = \{0,1\} \times \{1,2,\ldots,6\}^2 \quad \text{for example} \]

Let \( A = \{ \text{total is 4} \} \), \( B = \{ \text{coin shows heads} \} \)

\[ p(A) = p(A \mid B) p(B) + p(A \mid \bar{B}) p(\bar{B}) \]

\[ = \frac{1}{2} \times \frac{1}{2} + \frac{2}{36} \times \frac{1}{2} = \frac{1}{8} \]

**Properties of Conditional Probability**

a) \( p(A \cap B) = p(A \mid B) p(B) \)

b) \( p(A \mid B) = \frac{p(B \mid A) p(A)}{p(B)} \)

c) \( p(A \cap B \cap C) = p(A \mid B \cap C) p(B \cap C) p(C) \)

d) \( p(A \mid B \cap C) = \frac{p(A \mid B \cap C \cap D)}{p(B \cap C)} \)

**Theorem (Bayes' Formula)**

Let \( B_1, B_2, \ldots \) be a partition of \( \Omega \), \( p(B_i) > 0 \) \( \forall i \).

Then

\[ p(B_i \mid A) = \frac{p(A \mid B_i) p(B_i)}{\sum_{i} p(A \mid B_i) p(B_i)} \]

Map: \( p(B_i) \rightarrow p(B_i \mid A) \)
Example: False Positives

There is a rare disease with incidence in the population 1 in 100,000. The test is fairly reliable.

If you have the disease, the test is positive with probability 0.95.
If not, the test is positive with probability 0.005.

If the test is positive, what is the probability the patient has the disease?

Solution

\[ D = \{ \text{has disease} \} \quad \Gamma = \{ \text{Test is positive} \} \]

\[
p(D | \Gamma) = \frac{\rho(D \cap \Gamma)}{p(D | \Gamma) p(D)}
\]

\[
= \frac{1}{100000 \times 0.95} \times 0.005
\]

\[
= 0.002
\]

Test is essentially useless.

Principle of Uniformity:

In the absence of information, take the prior to be uniform.

E.g. If there are 2 possibilities, take \( \frac{1}{2} \) and \( \frac{1}{2} \). ⇒ "Bayes Postulate"

1939 Harold Jeffreys

Simpson's Paradox (British Medical Journal 1986, kidney Stone removal)

Before 1980, Open Surgery was performed and afterwards, a new operation PN. In 1972-1980, \( \frac{273}{350} \approx 78\% \) were successful.

In 1980-1985, \( \frac{289}{350} \approx 83\% \) were successful.

We deduce that PN is better than OS.

Small (<2cm) | Large (>2cm) | We then see that OS is better in BOTH

| OS | 93% \( \frac{81}{87} \) | 73% \( \frac{192}{263} \) |
| PN | 87% \( \frac{234}{270} \) | 69% \( \frac{55}{80} \) |
02/04/11

Probability

The following are not inconsistent:

\[ p(R_{1A}) > p(R_{1B}) \]
\[ p(R_{1A} \cap S) < p(R_{1B} \cap S) \]
\[ p(R_{1A} \cap L) < p(R_{1B} \cap L) \]

\[ R = \{ \text{success} \} \]

A: PN
B: OS
S: small
L: large
2.3 Independence

Intuition: if \( p(A) = p(A \mid B) \) then \( B \) is of little relevance to \( A \).

Definition

Events are independent if \( p(A \cap B) = p(A) \cdot p(B) \)

More generally, a family \((A_i : i \in I)\) is independent if

\[
p\left( \bigwedge_{i \in S} A_i \right) = \prod_{i \in S} p(A_i) \quad \text{for all finite subsets } S \subseteq I,
\]

and pairwise independent if \( p(A_i \cap A_j) = p(A_i) \cdot p(A_j) \quad \forall i, j \in I \)

Independence \( \Rightarrow \) Pairwise Independence. The converse is false.

Example

\( \Omega = \{1, 2, 3, 4\} \quad p(A) = \frac{1}{4} \)

\( A_1 = \{1, 2\}, A_2 = \{2, 3\}, A_3 = \{1, 3\} \)

'Independence' \( \Rightarrow \) 'repeated trials'

Example Two dice are thrown, equiprobable outcomes

\( A \): attribute of the number on the first die.

\( B \): attribute of the number on the second die.

\[
p(A \cap B) = \frac{\# \text{outcomes: 1 on 1st, 2 on 2nd}}{36} = \frac{\# \text{outcomes, with A} \times \# \text{outcomes, with B}}{6 \times 6}
\]

More general: Product probability space

\( \Omega_1 = \{ \alpha_1, \alpha_2, \ldots \} \quad p_1(\alpha_i) = p_i \)

\( \Omega_2 = \{ \beta_1, \beta_2, \ldots \} \quad p_2(\beta_i) = q_i \)
Let \( S = S_1 \times S_2 = \{ (\alpha_i, \beta_i) : i, j \geq 1 \} \)

\[ Y = \text{something suitable} \]

\[ p[(\alpha_i, \beta_i)] = p_i \cdot q_i, \quad i, j \geq 1 \]

Then \( A_1 \subseteq S_1, \ A_2 \subseteq S_2 \)

\[ p(A_1 \times A_2) = \sum_{i \in A_1} \sum_{j \in A_2} p_i \cdot q_j = \sum_{i \in A_1} p_i \cdot \sum_{j \in A_2} q_j = p_i(A_1) \cdot p_j(A_2) \]

**Language**

Flips of a coin → Interpreted to imply independence

Throws of a die → Between outcomes

**Example**

\( n \) flips of a coin that shows heads with probability \( p \) each time.

Let \( S_n \) be the number of heads. Find \( p(S_n = k) \).

**Solution 1**

\[ p(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \]

The Binomial Distribution.

**Solution 2**

Let \( X \) be the outcome of the first flip

\[ p(S_n = k) = p(S_n = k \mid X = H) \cdot p(X = H) + p(S_n = k \mid X = T) \cdot p(X = T) \]

\[ p(S_n = k) = p(S_{n-1} = k-1) \cdot p + p(S_{n-1} = k) \cdot (1-p) \]

Valid for \( k \geq 0 \)

Let \( p_n(k) = p(S_n = k) \)

\[ p_n(k) = p \cdot p_{n-1}(k-1) + (1-p) \cdot p_{n-1}(k) \]

A discrete recurrence relation.
\[ p_n(k) = \mathbb{P}[p_{n-2}(k-2) + (1-p) p_{n-2}(k-1)] \\
+ (1-p) [p_{n-2}(k-1) + (1-p) p_{n-2}(k)] \\
= \sum_{r=0}^{\infty} p^r (1-p)^{n-r} \binom{n}{r} p_{n-r}(k-r) \quad s \geq 0 \]

(Prove by induction)

\[ = \sum_{r=0}^{\infty} p^r (1-p)^{n-r} \binom{n}{r} p_0(k-r) \quad p_0(k-r) = \delta_{kr} \]

\[ = p^k (1-p)^{n-k} \binom{n}{k} \]

**Geometric Distribution**

Same coin is tossed repeatedly until the first head appears. Let \( R \) be the number of flips required.

\[ p(R = r) = (1-p)^{r-1} p, \quad r = 1, 2, 3, \ldots \]

Take care: Can see the geometric distribution as

\[ p_r = (1-p)^{r-1} p, \quad r = 0, 1, 2, \ldots \]

**Example 2**

Random walks

\[ \begin{array}{cccc}
0 & 1 & 2 & \vdots \\
\downarrow & \searrow & \nearrow & \uparrow \\
1 & 2 & 3 & \ddots \\
\end{array} \]

Start at \( k \). At each step, move one step right with probability \( p \), or left with probability \( q \). Different steps are independent.

Assume there are absorbing barriers at 0 and \( N \). This can be compared to a gambler who plays a game until either he reaches 0—he leaves bankrupt, or \( N \)—leaves and buys a large car.
What is the probability of ultimate bankruptcy?

\[ p(\text{we reach } 0 \text{ before we reach } N) \]

Let \( A = \{ \text{absorbed at } 0 \} \)

\[ B = \{ \text{1st step is to the right} \} \]

\[ p(A) = p(A | B) p(B) + p(A | \overline{B}) p(\overline{B}) \]

Let \( p_k = p(A \mid \text{start at } k) \)

\[ p_k = p_{k+1} p + p_{k-1} q, \quad \text{for } 0 < k < N \]
Random walk:

\[ P_{k+1} = P_k \alpha_{k+1} + q \alpha_{k-1} \]
\[ \Rightarrow P_k (P_{k+1} - P_k) = q (P_k - P_{k-1}) \]

Boundary conditions \( P_0 = 1, P_N = 0 \). Try \( P_k = e^{\theta k} \)

\[ \frac{\theta e^2 - \theta + q = 0}{(\theta e - q)(\theta - 1) = 0} \]
\[ \theta = \frac{q}{\beta}, 1 \]

If \( q \neq \beta \), roots are distinct

\( (P \neq \frac{1}{2}) \) General solution \( P_k = A \left( \frac{\alpha}{\beta} \right)^k + B \cdot 1^k \)

\[ P_k = \frac{\left( \frac{\alpha}{\beta} \right)^k - \left( \frac{\alpha}{\beta} \right)^N}{1 - \left( \frac{\alpha}{\beta} \right)^N} \]

If \( q = \beta = \frac{1}{2} \) (symmetric) \( P_k = 1 - \frac{k}{N} \)

3. Discrete Random Variables \((\sigma, \Omega, P)\)

Concept definition: A random variable is a function \( X: \Omega \rightarrow \mathbb{R} \)

Example 1: \( X = \# \) heads after two coin tosses

\( \Omega = \{0, 1\}^2 \)

\( \Omega = \{w_1, w_2\} \in \Omega \)

\( X(w) = w_1 + w_2 \)

Example 2: Throw 3 dice. \( X \) is the largest number shown.

Definition: The distribution function of a random variable \( X \) is defined by \( F: \mathbb{R} \rightarrow [0, 1] \), \( F(x) = P \{ \omega \in \Omega : X(\omega) \leq x \} \)

\[ F(x) = \lim_{x \to \infty} \frac{\omega \in \Omega : x(\omega) \leq \omega} \]
Example 1. Fair coin, two tosses, \( X = \# \text{heads} \)

- \( X \) takes values in \([0, 1, 2]\)

**Definition:** The **mass function** of the random variable \( X \) is the function \( f : \mathbb{R} \to [0, 1] \) given by \( f(x) = p(X = x) \)

**Definition:** The random variable \( X \) is discrete if there exists a countable set \( S = \{x_1, x_2, \ldots\} \) such that \( p(X \in S) = 1 \).

If \( X \) is discrete, we usually work with the mass function.

1. **Binomial Distribution** (coin toss)
   
   \[ p_0 = 1 - p, \quad p_1 = p, \quad \text{where } p \in [0, 1] \]
   
   \[ f(0) = 1 - p, \quad f(1) = p \]

2. **Binomial Distribution** \( n, p, \text{ bin}(n, p) \)
   
   \[ f(k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \ldots, n \]

3. **Poisson Distribution** \( \text{Po}(\lambda) \)
   
   \[ f(k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, \ldots \]

**Relationship between Binomial and Poisson distributions**

A book: One page has \( n = 10^5 \) characters, each of which is misprinted with probability \( p = 10^{-5} \) independently of the others.

The distribution of \( N \), the total number of misprints, is \( \text{bin}(n, p) \).

As \( n \to \infty, p \to 0 \) such that \( np \to \lambda \) as \( n \to \infty \):

\[ p(n = k) = \binom{n}{k} p^k (1 - p)^{n-k} = \frac{n(n-1)\ldots(n-k+1)}{n^k} \frac{\lambda^k}{(1 - \frac{\lambda}{n})^n} \frac{1}{k!} \]

As \( n \to \infty \)

\[ p(n = k) \to e^{-\lambda} \frac{\lambda^k}{k!} \]

So, \( \text{bin}(n, \frac{\lambda}{n}) \to \text{Po}(\lambda) \) as \( n \to \infty \)
Example 4  Geometric distribution

$$f(k) = A \beta^k, \quad k \geq 1, 0 < \beta < 1$$

$$\sum_k f(k) = \sum_k A \beta^k = \frac{A \beta}{1 - \beta} = 1, \quad A = \frac{1 - \beta}{\beta}$$

$$f(k) = (1 - \beta) \beta^{k-1}$$

Example 5  Negative binomial

Toss a coin, $p(\text{heads}) = p$, until exactly $r$ heads are tossed.

What is the probability that this requires $k$ tosses ($= P_k$).

$$P_k = P\left(\sum_{i=0}^{k-1} \text{heads}, \text{heads} \quad r \quad k \quad -1\right) = \binom{k-1}{r-1} (1-p)^{k-r} p^r$$

$$= p^r (1-p)^{k-r} \binom{k-1}{r-1}, \quad k = r, r+1, \ldots$$

$$\sum_k P_k = p^r \sum_{k=r}^{\infty} (1-p)^{k-r} \binom{k-1}{r-1}$$

$$= p^r \sum_{k=0}^{\infty} (1-p)^{k} \binom{k+1}{r}$$

$$= p^r \frac{(1-p)^{r}}{(1-p)^{r}} = 1 \quad \text{for} \quad r \geq 0$$

3.2 Expectation (of Discrete Random Variables)

$$(\Omega, \mathcal{F}, \mathbb{P})$, discrete random variable $X$

Definition  The expectation (or mean value) of $X$ is

$$E(X) = \sum_{\mathbb{P}(X = x) > 0} x \mathbb{P}(X = x) = \sum_{x} x P_X(x)$$

whenever this sum converges absolutely.

Composition of Functions

Composition creates a new random variable

$$\begin{align*}
\Omega & \xrightarrow{X} \mathbb{R} \\
\mathbb{R} & \xrightarrow{g} \mathbb{R} \\
Y = g(X) \\
[Y(w) = g(X(w)), \quad w \in \Omega]
\end{align*}$$
Theorem (Law of the unconscious statistician)

\[ E(g(x)) = \sum_{x} g(x) f_X(x) \]

Proof

\[ Y = g(X), \quad E(Y) = \sum_{y} y \cdot P(Y = y) \]

\[ = \sum_{y} y \left[ \sum_{x: g(x) = y} P(X = x) \right] = \sum_{x} \frac{\# g(x) \cdot P(X = x)}{\# g(x)} \]

Properties of Expectation

1) If \( X \geq 0 \), then \( E(X) \geq 0 \)

2) If \( X \geq 0 \), \( E(X) = 0 \), then \( P(X = 0) = 1 \)

Proof: If \( X \geq 0 \), \( E(X) = \sum_{x \geq 0} x \cdot P(X = x) = 0 \)

\[ \Rightarrow x \cdot P(x = x) = 0 \quad \text{for all } x > 0. \]

3) \( E(\alpha X + \beta) = \sum_{x} (\alpha x + \beta) \cdot P(X = x) = \alpha \sum_{x} x \cdot P(X = x) + \beta \sum_{x} P(x) = \alpha E(X) + \beta \]

4) \( E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y) \)

\[ E(\alpha X + \beta Y) = \sum_{x,y} (\alpha x + \beta y) \cdot P(X = x, Y = y) \]

\[ = \alpha \sum_{x,y} x \cdot P(X = x, Y = y) + \beta \sum_{x,y} P(x, y) \]

\[ = \alpha \sum_{x} x \cdot P(X = x) + \beta \sum_{y} y \cdot P(Y = y) \]

\[ = \alpha E(X) + \beta E(Y) \]

"Expectation is a linear operator" \( E(X) \); measure of distribution center.

Variance: a measure of dispersion.

Definition The Variance of \( X \) is:

\[ \text{Var}(X) = \text{E}[X - \text{E}(X)]^2 \]

and the standard deviation \( \sigma(X) = \sqrt{\text{Var}(X)} \)
a) Variance is non-linear
\[ \text{Var}(aX + b) = a^2 \text{Var}(X) \]
and hence \( \sigma(aX + b) = |a| \sigma(X) \)

b) \[ \text{Var}(X) = E[(X - E(X))^2] = E(X^2 - 2X(E(X)) + (E(X))^2) \]
\[ = E(X^2) - 2E(X)E(X) + [E(X)]^2 \]
\[ = E(X^2) - (E(X))^2 \]

Warning: Be careful with parentheses, e.g., what does \( E(X^2) \) mean?

(Ex)^2 or E(X^2)?

Definition: The k\textsuperscript{th} moment of \( X \) is \( M_k = E(X^k) \), \( k \in \mathbb{N}_0 \)

Note:

a) \( \text{Var}(X) = M_2 - (M_1)^2 \)
b) \( \text{Var}(X) > 0 \)
c) \( \text{Var}(X) = 0 \) if and only if \( p(X = c) = 1 \) for some \( c \in \mathbb{R} \)
\[ M_k = E(X^k) \]\text{ moments} \\

**Example 1: Bernoulli's Distribution** 

\[ p(X=0) = q, \quad p(X=1) = p, \quad p+q = 1 \]

\[ E(X) = 0 \cdot q + 1 \cdot p = p \]

\[ \text{Var}(X) = E(X^2) - (E(X))^2 = p - p^2 = pq \]

**Probability Space** \((\Omega, \mathcal{F}, P)\) event \(A \subseteq \Omega\) 

**Indicator function** of \(A\) is the random variable \(1_A: \Omega \rightarrow \{0, 1\}\) 

\[ 1_A(\omega) = \begin{cases} 0, & \omega \notin A \\ 1, & \omega \in A \end{cases} \]

\(1_A\) is a Bernoulli random variable. 

\[ p(1_A = 0) = p(A), \quad E(1_A) = p(A) \]

**Example 2: Binomial Distribution** 

\[ X \sim \text{bin}(n, p) \]

\[ E(X) = \sum_{k=0}^{n} k \cdot p(X=k) = \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k} = np \]

or \(X\) is the sum of \(n\) Bernoulli variables each with mean \(p\), so 

\[ E(X) = np, \] and in fact \(\text{Var}(X) = npq\)

**Example 3: Poisson Distribution** \(Po(\lambda)\) 

\[ E(X) = \sum_{k=0}^{\infty} k \cdot p(X=k) = \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} = \lambda \]

**Example 4: Geometric, parameter \(p\)** 

\[ p(X=r) = (1-p)^{r-1} p, \quad r \geq 1 \]

\[ E(X) = \sum_{r=1}^{\infty} r \cdot (1-p)^{r-1} p \]

\[ \sum_{r=0}^{\infty} x^r = \frac{1}{1-x}, \quad |x| < 1, \quad \sum_{r=0}^{\infty} r x^{r-1} = \frac{1}{(1-x)^2}, \quad |x| < 1 \]

\[ E(X) = p \cdot \frac{1}{p^2} = \frac{1}{p} \]
and \( \text{Var}(X) = \frac{q\nu}{\mu^2} \) \( \nu = 1 - p \)

### 3.3 Probability Generating Functions

**Definition** Random variable \( X \) taking values in \( \{0, 1, 2, \ldots \} \)

The probability generating function of \( X \) is the function

\[
G : S \to \mathbb{R} \quad G(s) = \sum_{k=0}^{\infty} s^k p(X=k) = E(s^X)
\]

wherever this sum converges absolutely, and as big an \( S \) as possible. Note: this sum converges absolutely whenever \( |s| < 1 \) \(-1 < s < 1\). Sometimes we write \( G_X \) for \( G \).

\[
G_X(0) = p(X=0) \quad G_X(1) = 1
\]

**Theorem**

The distribution of \( X \) is uniquely determined by its PGF \( G \).

**Proof**

\[
p_k = p(X=k)
\]

\[
G(s) = p_0 + sp_1 + s^2 p_2 + \ldots
\]

Converges on \(-1, 1\] \( S=0 \Rightarrow G(0) = p_0 \)

\[
G'(0) = p_1 \quad \ldots \quad G^{(k)}(0) = k! \cdot p_k
\]

**Why?**

1) An elegant method for handling sums of random variables.
2) A good method for calculating moments

\[
G(s) = \sum_{k=0}^{\infty} s^k p(X=k) \quad s \in (-1, 1]
\]

\[
G'(s) = \sum_{k=0}^{\infty} k s^{k-1} p(X=k)
\]
Theorem \( E(X) = G'(1) \)

Non rigorous, \( G(x) = E(x^k) \), \( G'(x) = E(xS^{x-1}) \) \( \Rightarrow G'(1) = E(x) \)

Problem \( s = 1 \) might be on the edge of the domain of convergence of \( G \). We need Abel's Lemma.

Further such results

\[
E(X) = G'(1) \quad G^{(k)}(1) = E\left[\frac{X(X-1)(X-2)\ldots(X-k+1)}{k!}\right]
\]

\[
G''(1) = E\left[\frac{X(X-1)}{2}\right] = E\left[X^2 - X\right] = E(X^2) - E(X)
\]

\[
\therefore \quad \text{Var}(X) = G''(1) + G'(1) - [G'(1)]^2
\]

Example: Bern(p)

\( G(x) = qS^0 + px^1 = q + px \)

Bin(n, p) \( G(x) = (q + px)^n \)

Poi(\lambda) \( G(x) = e^{\lambda(1-x)} \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda s)^k}{k!} \)

Geom(p) \( G(x) = \frac{ps}{1-qs} \)

Application of generating functions to tiling a bathroom

Let \( f_n \) be the number of possible ways to tile this with tiles of size \( 2 \times 1 \). \( f_n = f_{n-1} + f_{n-2} \)

\( f_0 = 1 \), \( f_1 = 1 \)

\[
\sum_{n=0}^{\infty} f_n s^n = \sum_{n=0}^{\infty} f_{n-1} s^n + \sum_{n=0}^{\infty} f_{n-2} s^n
\]

\[
F(s) - f_0 - f_1 s = s[F(s) - f_0] + s^2 F(s)
\]

\[
F(s) = \frac{f_0 (1-s) + f_1 s}{1-s-s^2} = \frac{1}{1-s-s^2}
\]
Number of $2 \times 1$ tilings of a $2 \times n$ area, $f_n = \# \text{ tilings}$

$n \geq 2 
\begin{align*}
f_n &= f_{n-1} + f_{n-2} , \\
f_0 &= f_1 = 1 \\
\alpha &= \frac{1 + \sqrt{5}}{2} \\
F(s) &= \sum_{n=0}^{\infty} f_n s^n = \frac{1}{1-s-s^2} = \left(1 - \frac{1}{\alpha_1} \right) \left(1 - \frac{1}{\alpha_2} \right) \\
\alpha_1 &= \frac{1 + \sqrt{5}}{2} \\
\alpha_2 &= \frac{1 - \sqrt{5}}{2} \\
\frac{1}{\alpha_1 - \alpha_2} &= \frac{\alpha_1}{1-\frac{1}{\alpha_1} s} + \frac{\alpha_2}{1-\frac{1}{\alpha_2} s} \\
\alpha &= \frac{1}{\alpha_1 - \alpha_2} \\
f_n &= \text{coefficient of } s^n = \frac{1}{\alpha_1 - \alpha_2} \left(\alpha_1^{n+1} - \alpha_2^{n+1} \right)
\end{align*}

The method of generating functions is robust:

\[ f_n = n f_{n-1} + f_{n-2} , \quad s^n f_n = s^n f_{n-1} + s^n f_{n-2} \]

\[ F(s) = f_0 - f_1 s = 3 ... \]

3.4 Independent Random Variables

**Definition**: Discrete random variables $X, Y$ are independent if

\[ p(x=x, y=y) = p(x=x)p(y=y) \quad \forall x, y \in \mathbb{R} \]

This can be extended to families of random variables.

\[ \{X_i : i \in I\} \text{ is independent if} \]

\[ p(x_i = x_i \quad \forall i \in S) = \prod_{i \in S} p(X_i = x_i) \quad \text{finite } S \subseteq I \]

The function $f_{x,y}(x, y) = p(x=x, y=y)$ is called the joint (probability) mass function of the pair $x, y$.

**Definition**

The covariance of $X$ and $Y$:

\[ \text{cov}(X, Y) = E\left[ (X-E_X)(Y-E_Y) \right] \]

and the correlation coefficient is

\[ \rho_{x,y} = \frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X) \text{ Var}(Y)}} \]
$X, Y$ are called **uncorrelated** if \( \rho(X, Y) = 0 \)

Note \( \text{cov}(X, Y) = E(\{X - E(X)\} \{Y - E(Y)\}) = E(XY) - E(X)E(Y) \)

**Theorem**

a) If $X, Y$ are independent then $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$ for $g, h : \mathbb{R} \rightarrow \mathbb{R}$

b) If $X, Y$ are independent, $\text{cov}(X, Y) = 0$, hence
\[ \text{Var}(X) \text{Var}(Y) \neq 0 \]

c) There exist random variables $X, Y$ that are dependent and uncorrelated.

**Proof**

a) \[ E[g(X)h(Y)] = \sum_{x,y} g(x)h(y) \rho(X=x, Y=y) = \sum_{x,y} g(x)h(y) \rho(X=x) \rho(Y=y) \text{ by independence} \]
\[ = \sum_{x} g(x) \rho(X=x) \sum_{y} h(y) \rho(Y=y) = E[g(X)]E[h(Y)] \]

b) If independent, \( E(XY) = E(X)E(Y) \Rightarrow \text{Cov}(X, Y) = 0 \)

c) $U, V$ are $\text{Bern}(1/2)$, independent.

\[ X = U + V, \quad Y = |U - V| \]

**Exercise**

Show
\[ \rho(X=2, Y=1) \neq \rho(X=2) \rho(Y=1) \]

and $X, Y$ are uncorrelated.
Correlation as a measure of dependence

a) It is a single number

b) $-1 \leq \rho(X, Y) \leq 1$ (assume variances $\neq 0$)

Theorem (Schwarz's or Cauchy-Schwarz inequality)

$$\left[ E(XY) \right]^2 \leq E(X^2)E(Y^2) \quad \text{True for all random variables}$$

Proof: Let $Z = X + tY$ where $t \in \mathbb{R}$

$$0 \leq E(Z^2) = E(X^2 + 2tXY + t^2Y^2)$$

$$= E(X^2) + 2tE(XY) + t^2E(Y^2) \geq 0 \forall t \in \mathbb{R}$$

$\Rightarrow$ Quadratic in $t^2$, discriminant $\leq 0$ so there is at most 1 real root.

$$4[E(XY)^2] - 4E(X^2)E(Y^2) \leq 0$$

$$\rho(X, Y)^2 = \frac{\text{cov}(X, Y)^2}{\text{Var}(X)\text{Var}(Y)} \leq 1$$

Use inequality on $E[XY-EX-EY]^2$

c) $\rho^2 = 1$ iff $X - EX + t(Y - EY) = 0$ for some $t \in \mathbb{R}$

iff $X + tY = C$ for some $t, C \in \mathbb{R} \ast$

and $\rho = 1$ iff $t$ in $(\ast)$ satisfies $t < 0, \rho = -1$ iff $t > 0$

d) $\rho(X, Y) = 0$ if $X, Y$ are independent

e) $\rho(aX+b, cY+d) = \rho(X, Y)$ if $ac > 0$
Theorem: a) \( \text{Var}(X+Y) = \text{Var}(X) + 2\text{Cov}(X,Y) + \text{Var}(Y) \)

b) If \( X \) and \( Y \) are independent, \( \text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) \)

Proof:
\[
\text{Var}(X+Y) = E[(X+Y - E(X+Y))^2] \\
= E[(X-E(X))^2 + 2(X-E(X))(Y-E(Y)) + (Y-E(Y))^2] \\
= \text{Var}(X) + 2\text{Cov}(X,Y) + \text{Var}(Y)
\]

Examples:

a) Variance of binomial \((n, p)\) is \( np(1-p) \)

b) Negative Binomial, parameters \( k, p \), has the same distribution as the sum of \( k \) independent \( \text{Geom}(p) \) random variables, \( k \sim \frac{(1-p)}{p} \)

Sums of random variables

Theorem:
\[
p(X+Y = z) = \sum_x p(X=x, Y=z-x)
\]

Proof:
\[
[X+Y = z] = \bigcup_x \{X=x, Y=z-x\}, \text{ a disjoint union}
\]

Since \( X, Y \) are independent, \( p(X+Y = z) = \sum_x p(X=x, Y=z-x) \)

Corollary:
If \( X, Y \) are independent then for \( Z = X+Y \)
\[
f_Z(z) = \sum_x f_X(x)f_Y(z-x) \]

Convolution \( f_Z = f_X \ast f_Y \)

Theorem: If \( X \) and \( Y \) are independent,
\[
G_{X+Y}(s) = G_X(s)G_Y(s)
\]
Proof: \( G_{X+Y}(s) = E(s^{X+Y}) = E(s^X s^Y) = E(s^X) E(s^Y) = G_X G_Y \)

Example

1. Let \( X \) be \( P_0(A) \), \( Y \) be \( P_0(B) \) which are independent

\[
G_X(s) = e^{A(s-1)}
\]

\[
G_X(s) = e^{A(s-1)} e^{B(s-1)} = e^{A+B(s-1)} \Rightarrow X+Y \sim P_0(A+B)
\]

2. What is the pgf of the negative binomial distribution with parameters \( k, p \)?

It is the \( k \text{th} \) power of the pgf of \( \text{Geom}(p) \)

i.e. \( (1-q s)^{-k} \)

3. Indicator Functions

\[
I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \\ \end{cases}
\]

Note: \( E(I_A) = P(A) \), \( \text{Var}(I_A) = P(A)P(A) \)

Basic Facts

i) \( I_{A \cap B} = I_A I_B \)

ii) \( I_A = 1 - I_A \)

\[
I_{A \cup B} = 1 - I_{A \cap B} = 1 - I_{A \cap B} = 1 - \frac{1}{1} - I_{A \cap B} = 1 - \frac{1}{1} - I_{A \cap B} = 1 - (1 - I_A)(1 - I_B)
\]

\[
= I_A + I_B - I_{A \cap B}
\]

Take expectations: \( P(A \cup B) = P(A) + P(B) - P(A \cap B) \)

Example. Inclusion-Exclusion Formula

\[
I_{\bigcup_{i} A_i} = 1 - \sum_{i} (1 - I_{A_i}) = 1 - \left[ 1 - \sum_{i} I_{A_i} + \sum_{i} \sum_{j<i} I_{A_i} I_{A_j} \right]
\]

Take Expectations: \( P(\bigcup_{i} A_i) = \sum_{i} P(A_i) - \sum_{i < j} P(A_i A_j) \) etc.

Example

\( n \geq 2 \) Male/Female Couples. The \( n \) men are seated randomly in odd positions at a round table, and the women randomly in the even positions.
Let \( N_0 := \text{# Men seated beside the right woman} \).

What are the mean and variance of \( N \)?

**Solution**

Let \( A_i \) be the event that the \( i \)th couple are next to each other.

\[
N = \sum_{c} I_{A_c}, \quad E(N) = \sum_{c} E(I_{A_c}) = \sum_{c} P(A_i)
\]

\[
E(N) = n \cdot P(A_i) \quad \text{by symmetry}
\]

\[
E(N) = n \cdot \frac{2}{n} = 2 \quad \Rightarrow \quad \frac{1}{A} = 1
\]

\[
E(N^2) = E(\sum_{c} I_{A_c}^2 + 2 \sum_{c < c'} I_{A_c} I_{A_{c'}})
\]

\[
= E(N) + 2 \sum_{c < c'} P(A_c \cap A_{c'})
\]

\[
= E(N) + 2 \cdot \frac{n}{n(n-1)} P(A_1 \cap A_2) \quad \text{by symmetry}
\]

\[
= E(N) + n(n-1) P(A_1) P(A_2 | A_1) \quad \text{no for uniform}
\]

\[
= E(N) + n(n-1) P(A_1) \cdot \left[ \frac{1}{n-1} \cdot \frac{1}{n-1} + \frac{n}{n-1} \cdot \frac{1}{n-1} \right]
\]

\[
= 2 + 2 \frac{(2n-3)}{n-1}
\]

\[
\text{Var}(X) = E(N^2) - (E(N))^2 = \frac{2(n-2)}{n-1}
\]

**In fact** \( P(N = k) = f_n(k) \)

\[
P(N = k) \xrightarrow{k \to \infty} \frac{2^k e^{-2}}{k!} = \text{Po}(2)
\]
3.6 Joint Distributions, Conditional Distributions

For $X, Y$ discrete, the joint mass function $f(x, y) = P(X=x, Y=y)$.

The marginal mass functions are:

$$f_X(x) = \sum_y f(x, y)$$

$$f_Y(y) = \sum_x f(x, y)$$

The conditional mass function of $X$ given $Y$

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

It is well defined if and only if $f_Y(y) \neq 0$.

The conditional expectation of $X$ given $Y = y$

$$E(X|Y = y) = \sum_x x f_{X|Y}(x|y)$$

Writing $E(Y) = E(X|Y = y)$ we normally define the conditional expectation of $X$ given $Y$ as $E(Y)$. This is a random variable. $E(Y) = E(X|Y)$.

Example $X_1, X_2, \ldots, X_n$ are independent $\text{Bern}(p)$

Find $E(X_1 | Y = y)$,

$Y = X_1 + X_2 + \ldots + X_n$ independent

Solution $E(X_1 | Y = y) = \frac{P(X_1 = 1 | Y = y)}{P(Y = y)}$

$$E(X_1 | Y = y) = \frac{P(X_1 = 1, Y = y)}{P(Y = y)} = \frac{P(X_1 = 1) P(X_2 + X_3 + \ldots + X_n)}{P(Y = y)}$$

$$= \frac{1 \cdot (p-1)(p-1) \cdots (1-p)^{n-y}}{\binom{n}{y} p^n (1-p)^{n-y}} = \frac{y}{n}$$

$E(X_1 | Y) = \frac{y}{n}$

A more clever method $X_1, X_2, \ldots, X_n$, iid (independent, identically distributed)

$Y = X_1 + X_2 + \ldots + X_n$
\[ E(Y|Y) = Y \]
\[ E(x_1 + \ldots + x_n | Y) = E(x_1 | Y) + \ldots + E(x_n | Y) = \frac{1}{n} E(x_1 | Y) \text{ by symmetry} \]
\[ \therefore E(x_1 | Y) = \frac{Y}{n} \]

**Theorem** (Properties of Conditional Expectation)

a) \( E(E(X|Y)) = E(X) \) \[[Very Useful!!]\]

b) If \( X \) and \( Y \) are independent, \( E(X|Y) = E(X) \), a constant

**Proof**

a) \[
E(E(X|Y)) = \sum_y E(X|Y=y) p(Y=y) \\
= \sum_y \left[ \sum_x x p(x|Y=y) \right] p(Y=y) \\
= \sum_x x p(x, Y=y) = \sum_x x p(x) = E(X)
\]

b) Obvious from the definition of conditional expectation.

Reminder: \( X_1, X_2, \ldots \) iid, taking values in \([0, 1, 2, \ldots] \)

\[ S = X_1 + X_2 + \ldots + X_n \]

\[ G_S(s) = G_{X_1}(s) \cdots G_{X_n}(s) = G_X(s)^n \]

**Theorem** The random sum formula

Let \( N, X_1, X_2, \ldots \) be independent, taking values in \([0, 1, 2, \ldots] \)

Suppose the \( X_i \) are iid with pgf \( G \). Then

\[ T = X_1 + X_2 + \ldots + X_N \] has pgf \( G_T(s) = G_N(G(s)) \)

**Example** \[ p(N = k) = 1 \quad G_N(s) = S^k \]

**Proof** \[
G_T(s) = E(s^T) = E \left( E(s^{T|N}) \right) \]

\[
G_T(s) = E \left[ G(s)^N \right] = G_N(G(s))
\]
\[ E(S^r) = E(E(S^r | N)) = \sum_n E(S^r | N=n) \rho(N=n) \]
\[ = \sum_n G_n(s) \rho(N=n) = G_N(G(s)) \]

**Example** \( \rho(N=k) = 1, \ G_N(s) = s^k \)

**Theorem** \( E(T) = E(N)E(X_i) \)

**Proof** \( E(T) = G_T'(1) = G_N'(G(0)) G_1'(1) = E(N)E(X_i) \)

**Exercise** Find \( \text{Var}(T) \)

**In general** \( \text{Var}(T) \neq E(N)\text{Var}(X_i) \)
3.7 Branching Process

A model for population growth (bacterial, spread of a family name).
Sometimes known as the Goulton-Watson (-Bienaymé) process.
It deals with growth in generations.

Let $X_n$ be the number of individuals in the $n$th generation.

Assumptions

a) $X_0 = 1$, a progenitor
b) $X_i$ is the number of offspring of the progenitor with mass function $f(k) = \rho(X_i = k)$.
c) Each member of the process has a family whose size has mass function $f$.
d) All offspring have family sizes which are independent of one another.

We can draw a family tree of the branching process, a random tree.

$X_{n+1} = Y_1 + Y_2 + Y_3 + \ldots + Y_{X_n}$, where the $Y_i$ are iid, mass function $f$, and independent of $X_n$. $X_{n+1}$ is a sum of a random number, $X_n$, of independent family sizes.

Let $G_n(s) = E(S^{X_n})$, the pgf of $X_n$.

Theorem

$G_{n+1}(s) = G_n(G(s))$, where $G(s)$ is the pgf of a family size.
i.e. $G(s) = E(S^{X_i}) = \sum_k s^k f(k)$

$X_{n+1} = A_i + \ldots + A_{X_n}$, and the $A_i$ are iid with distribution of $X_n$. 
\[ C_{n+1}(S) = G_n(\xi) \]

**Proof**

By decomposition of the tree, and the random sum formula.

Hence \( C_n(S) = G(G_{n-1}(S)) = G(G(\ldots(G(S))\ldots)), n \text{ times}. \)

**Corollary**

Let \( \mu = E(X_i) < \infty, \sigma^2 = Var(X_i) < \infty, \) then:

\[ E(X_n) = \mu^n, \quad Var(X_n) = \begin{cases} \mu \sigma^2 & \text{if } \mu = 1 \\ \frac{\sigma^2 \mu^{n-1} (\mu^n - 1)}{\mu - 1} & \text{if } \mu \neq 1 \end{cases} \]

**Proof**

\[ G_n(1) = G_{n-1}'(G(1)) G'(1) \quad \quad G_n(S) = G_{n-1}(G(S)) \]

\[ = G_{n-1}'(1) G'(1) \]

\[ E(X_n) = E(X_{n-1}) \mu = E(X_0) \mu^n = \mu^n \]

The calculation of variance is left as an exercise.

**Example**

Let \( X_i \) have the geometric distribution. \( p(X_i = k) = pq^k \)

\[ k = 0, 1, 2, \ldots, p + q = 1, \quad p \neq q, \quad pq \neq 0 \]

\[ G(S) = \sum_{k=0}^{\infty} pq^k = \frac{p}{1-q}S \quad \text{if } |q| < 1 \]

\[ G_n(S) = p \left( \frac{q^n - p^n}{q - p} \right) - qS(q^n - p^n) \quad |S| \leq 1 \]

**Proof**

By induction.
Hence $p(X_n = k)$ for general $k$, $n$.

\[ \mu = E(X_1) = \frac{a}{\varphi} \neq 1 \]

\[ \therefore E(X_n) = \mu^n = \left(\frac{a}{\varphi}\right)^n \]

**Extinction**

\[ p(X_n = 0) = G_n(0) = \frac{\mu^n - 1}{\mu^{n+1} - 1} \xrightarrow{n \to \infty} \begin{cases} 1 & \text{if } \mu < 1 \\ \frac{1}{\mu} & \text{if } \mu > 1 \end{cases} \]
Problem of extinction

\[ G(s) = E(s^{X_1}) \]

Let \( A_n = \{ X_n = 0 \} \subseteq A_{n+1} \quad \text{and} \quad A_n \subseteq A_{n+1} \subseteq \ldots \subseteq \lim_{n \to \infty} A_n \]

\[ \lim_{n \to \infty} A_n := \bigcap_{n=1}^{\infty} A_n = \{ \text{ultimate extinction} \} \]

Theorem

If \( B_1, B_2, \ldots \) is an increasing sequence of events, then

\[ p(\bigcup_{i=1}^{\infty} B_i) = \lim_{i \to \infty} p(B_i) \quad \text{i.e.} \quad p(\lim_{i \to \infty} B_i) = \lim_{i \to \infty} p(B_i) \]

"Probability measures are continuous set functions"

A similar statement for intersections of decreasing sequences exists.

Proofs

If \( B_n \setminus B_{n-1} =: C_n, \quad C_1 = B_1 \), then the \( C_n \) are disjoint

\[ p(\bigcup_{n=1}^{\infty} B_n) = \lim_{n \to \infty} p(\bigcup_{n=1}^{\infty} C_n) = \lim_{n \to \infty} \sum_{i=1}^{n} p(C_i) = \sum_{i=1}^{\infty} p(C_i) = p(\bigcup_{i=1}^{\infty} C_i) = p(\bigcup_{i=1}^{\infty} B_i) \]

Branching Processes

Let \( \eta = p(\text{ultimate extinction}) = \lim_{n \to \infty} p(X_n = 0) \) by the last theorem.

Theorem

\( \eta \) is the smallest non-negative root of the equation \( x = G(x) \)

Proofs

Let \( \eta_n = p(X_n = 0), \quad \text{so} \quad \eta_n \nearrow \eta \)

\[ \eta = G(\eta) = G(G(\eta)) = \ldots = G(\eta_n) \]

As \( n \to \infty \), \( \eta_n \nearrow \eta \), \( G(\eta_{n-1}) \to G(\eta) \) by continuity of \( G \)

\( \eta = G(\eta) \)
Let $\eta$ be any non-negative root of $x = G(x)$

$\eta_1 = G(0) \leq G(\eta) = \eta$

$\eta_2 = G(\eta_1) \leq G(\eta) = \eta$

By induction $\eta_n \leq \eta \forall \eta$, hence $\eta \leq \eta$

Theorem

a) If $\mu < 1$, then $\eta = 1$

b) If $\mu > 1$, then $\eta < 1$

c) If $\mu = 1$, and $\text{Var}(X) > 0$, then $\eta = 1$

Proof

$G(x)$

(see picture)

$G'(1) = \mu$

a) When $\mu < 1$, the only solution to $x = G(x)$ is $x = 1$, so $\eta = 1$

b) $\mu > 1$ Then there exists another root in $[0, 1)$ and this is $\eta$. 
3.8 Random Walk

Consider a random walk on \([0, 1, \ldots, N]\), with absorbing barriers at 0 and N.

Let \(M = \#\) steps up to the moment of absorption at either 0 or N.

Let \(e_k = E(M \mid \text{start at } k)\)

\[
e_k = E(E(M \mid 1^{st\ step}) = p(e_{k+1} + 1) + q(e_{k-1} + 1)
\]

\[e_0 = 0 = e_N\]

\[\text{(*) } pe_{k+1} - e_k + q e_{k-1} = -1\]

General solution \(e_k = \begin{cases} A\left(\frac{q}{p}\right)^k + B & q \neq p \\ A + Bk & q = p \end{cases}\)

Particular Solution \(e_k = \begin{cases} -\frac{k}{p-q} & p \neq q \\ \frac{N}{p-q} & p = q \end{cases}\)

\(p \neq q, \quad e_k = -\frac{k}{p-q} + c, S., \quad A = \frac{N}{(p-q)(p+q)-1}, \quad B = -A\)

\(p = q, \quad e_k = k(N-k)\)

4. Continuous Random Variables

4.1 Density Functions \((\xi, \gamma, \rho), X : \Omega \rightarrow \mathbb{R}\), Distribution function

Distribution Function \(F_X(x) = \rho(x \leq x)\)

Definition \(X\) is called "continuous" if there exists \(f : \mathbb{R} \rightarrow \mathbb{R}\) such that

a) \(\rho(x \leq x) = \int_{-\infty}^x f(u) \, du, \quad x \in \mathbb{R}\)

b) \(f(u) > 0 \quad \forall u\)

If this holds, \(f\) is called the probability density function of \(X\).
If $F_x(x) = \int_{-\infty}^{x} f(u)\,du$, $f \geq 0$

Note

i) If $F_x$ is differentiable, we take the pdf to be $f_x = F'_x$

ii) Assume henceforth that $X$ has pdf $f_x$

$p(x = x) = 0$ $\forall x \in \mathbb{R}$

Proofs

$$\{X = x\} = \bigcap \{X \in (x - \frac{1}{n}, x]\}$$

$$p(X = x) = \lim_{n \to \infty} p(x - \frac{1}{n} < X \leq x)$$

$$= \lim_{n \to \infty} \left[ \int_{x - \frac{1}{n}}^{x} f(u)\,du - \int_{-\infty}^{x - \frac{1}{n}} f(u)\,du \right]$$

$$= \lim_{n \to \infty} \int_{x - \frac{1}{n}}^{x} f(u)\,du = 0$$

iii) $p(a \leq X \leq b) = \int_{a}^{b} f(u)\,du$

Proof: $p(a \leq X \leq b) = p(X = a) + p(a < X \leq b)$

$$= 0 + \int_{a}^{b} f(u)\,du - \int_{-\infty}^{a} f(u)\,du = \int_{a}^{b} f(u)\,du$$

iv) The pdf $f$ is characterized by: $f(u) > 0$ $\forall u$

$f$ is integrable with $\int_{-\infty}^{\infty} f(u)\,du = 1$

v) For a mass function, the element of probability is $f(X)$; for a density function, it is $f(x)\,dx$

Note: Proofs for discrete distributions are often valid also for continuous distributions with $P(X \in B) \to \int_{B} f(u)\,du$
Examples

1. Uniform Distribution  $\text{Unif}([a, b])$

\[ f(u) = \begin{cases} 0 & \text{if } u \notin [a, b] \\ \frac{1}{b-a} & \text{if } u \in [a, b] \end{cases} \text{ for some } c \]

\[ \int_{-\infty}^{\infty} f(u) \, du = c(b-a) = 1 \Rightarrow c = \frac{1}{b-a} \]

---

2. Exponential Distribution  $\text{Exp}(\lambda)$

\[ f(u) = \begin{cases} 0 & u \leq 0 \\ \lambda e^{-\lambda u} & u > 0 \end{cases} \]

\[ F(x) = \int_{-\infty}^{x} f(u) \, du = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-\lambda x} & x > 0 \end{cases} \]

Important Property: ‘lack of memory’ or ‘memoryless’ property

Let $X$ be $\text{Exp}(\lambda)$. We need $p(X > y + z \mid X > y)$

\[ p(X > y + z \mid X > y) = \frac{p(X > y + z)}{p(X > y)} = e^{-\lambda(y+z)} \]

Conversely, if $F$ is a distribution function with pdf $f$, and

\[ \frac{1 - F(y+z)}{1 - F(y)} = 1 - F(z) \quad y, z > 0 \]

then $F$ is the distribution function for an exponential distribution.

This is a key property in the theory of Markov processes and more generally, to random/stochastic processes.
3. Normal / Gaussian Distribution

\[ N(0, 1) : f(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}, u \in \mathbb{R} \]

More generally: \[ N(\mu, \sigma^2) : g(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], x \in \mathbb{R} \]

i.e. \( N(0, 1) \) changed by location \( \mu \) and scaled by \( \sigma \)

4.2 If \( X \) has pdf \( f \), and \( h : \mathbb{R} \rightarrow \mathbb{R} \), what is the pdf of \( h(X) \)?

More generally: What is the distribution of \( h(X) \) in terms of the distribution function of \( X \)?

\[ p[h(x) \leq y] = p[X \in h^{-1}(-\infty, y)] + \text{calculation} \]

\[ = F_{h(x)}(y) = p(X \leq h^{-1}(y)) \text{ if } h \text{ strictly increasing} \]

\[ = F_X[h^{-1}(y)] \]

assuming \( h \) is sufficiently smooth

Take \( X \) to have pdf \( f \). \[ f_{h(x)}(y) = f[h^{-1}(y)] \frac{dy}{h^{-1}(y)} \]
\[ Y = h(X) \] If all functions are sufficiently smooth
\[ f_Y(y) = f_X[h^{-1}(y)] \left| \frac{dy}{dh^{-1}(y)} \right| \]

**Example**

If \( X \sim \text{Unif}[0,1] \), \( h(x) = -\log x \), \( Y = h(X) \)
\[ p(Y \leq y) = p(-\log x \leq y) = p(\log x \geq -y) = p(x \geq e^{-y}) = 1 - e^{-y}, \ y > 0 \]
\[ f_Y(y) = e^{-y}, \ y > 0 \]

**Important Method**

\( X \sim \text{Unif}[0,1] \), let \( F \) be a continuous distribution function
Let \( Y = F^{-1}(X) \)

[For the sake of rigor,
\( F^{-1}(y) \) is the infimum of \( \{ x : F(x) = y \} \]
\[ p(x \leq y) = p(F^{-1}(x) \leq y) = p(x \leq F(y)) = F(y) \]
\( Y \) has distribution \( F \).

See Monte Carlo Methods

**Example** If \( X \sim \text{N}(0,1) \)
Let \( Y = \sigma X + \mu \) \( \sigma, \mu \in \mathbb{R} \)
\( h(x) = \sigma x + \mu = y \), \( x = \frac{y - \mu}{\sigma} = h^{-1}(\sigma) \)
\[ f_Y(y) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left[ -\frac{1}{2} \left( \frac{y - \mu}{\sigma} \right)^2 \right] \frac{1}{101} \]
3. Expectation

If $X$ is discrete, $E(X) = \sum x P(X = x)$

If $X$ is continuous, $E(X) = \int_{-\infty}^{\infty} xf(x) \, dx$ whenever the integral is absolutely convergent.

**Theorem**

If $X$ has pdf $f$, and $\int_{-\infty}^{\infty} |g(x)| f(x) \, dx < \infty$ then

$E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) \, dx$

**Proposition** If $X$ is a continuous random variable,

$E(X) = \int_{-\infty}^{\infty} P(X > x) \, dx - \int_{-\infty}^{\infty} P(X < -x) \, dx$

If $E(X)$ exists, this may be used as a definition of $E(X)$ for any $X$, regardless of type.

**Proof**

$\int_{-\infty}^{\infty} P(X > x) \, dx = \int_{-\infty}^{\infty} \left[ \int_{x}^{\infty} f(u) \, du \right] \, dx$

$= \int_{-\infty}^{\infty} du \int_{x}^{\infty} f(u) \, dx = \int_{-\infty}^{\infty} u \, f(u) \, du$

and similarly $\int_{-\infty}^{\infty} P(X < -x) \, dx = -\int_{-\infty}^{\infty} u \, f(u) \, du$

**Proof**

$\int_{-\infty}^{\infty} P(g(X) > y) \, dy = \int_{-\infty}^{\infty} dy \int_{x : g(x) > y} f(x) \, dx$

$= \int_{x : g(x) > 0} f(x) \int_{x : g(x) > y} dy = \int_{x : g(x) > 0} g(x) f(x) \, dx$

The 2nd integral is $-\int_{x : g(x) < 0} g(x) f(x) \, dx$

hence the claim is proved.

Interchanging orders of integration is validated by a result called Fubini's Theorem.

Note: Using discrete theory, one now defines mean, variance, moments, covariance...
5.1 Three Inequalities

5.1 Jensen's Inequality

Definition. A function \( u : (a, b) \rightarrow \mathbb{R} \) is called convex if

\[
u(px + (1-p)y) \leq p u(x) + (1-p) u(y)\]

\( \forall x, y \in (a, b), \ p \in [0, 1] \)

\( u \) is concave if \(-u\) is convex. Note: Convexity \( \Rightarrow \) continuity

Examples

\( u(x) = -\log(x) \), \( u(x) = \frac{1}{x} \) on \((0, \infty)\)

Fact. If \( u'' \) exists and satisfies \( u''(x) \geq 0 \) for \( x \in (a, b) \), then \( u \) is convex.

Theorem

Let \( X \) be a random variable taking values in some open interval \((a, b)\) and let \( u \) be convex on \((a, b)\). Then

\[
E(u(X)) \leq u(E(X))
\]

Example. AM - GM inequality

Let \( x_1, x_2, \ldots, x_n > 0 \)

Let \( p(X = x_i) = \frac{1}{m}, \ i = 1, 2, \ldots, m \)

\( u(x) = -\log(x) \)

By Jensen's Inequality

\[
u \left( \frac{1}{m} \sum_{i=1}^{m} x_i \right) \leq \sum_{i=1}^{m} \frac{1}{m} u(x_i)
\]

\[-\log(AM) \leq -\log(GM)\]

\( AM \geq GM \)
Example: \( p(X > 0) = 1, \ u(x) = \frac{1}{x} \)

\( \frac{1}{EX} \leq E\left(\frac{1}{X}\right) \) \( \) In general \( E\left(\frac{1}{x}\right) \neq \frac{1}{EX} \)

unless \( \text{Var}(X) = 0 \).

**Proof of Jensen's Inequality**

**Theorem (Supporting hyperplane theorem)**

\( u \) is convex on \((a, b)\) if and only if

\( \forall x \in (a, b), \ \exists \lambda \in \mathbb{R} \) such that \( u(y) \geq \lambda (y - x) + u(x) \) \( \forall y \in (a, b) \)

The proof is to be discussed later.

Let \( x = EX \). By the supporting hyperplane theorem,

\( \exists \lambda \) such that \( u(y) \geq \lambda (y - EX) + u(EX) \)

\( \therefore u(x) \geq \lambda (x - EX)) + u(EX) \)

\( E[u(x)] \geq 0 + u(EX) \)

**More on Jensen**

**Lemma:** If \( u \) is convex then

\( u\left(\sum \pi_i x_i\right) \leq \sum \pi_i u(x_i) \) for \( x_i \in (a, b) \)

\( \pi_i > 0 \) with \( \sum \pi_i = 1 \).

This is equivalent to Jensen's Inequality for discrete random variables taking finitely many values.
Proof: (by induction on \( m \))

\( m = 2 \) holds by the definition of continuity. Assume this is true for \( m = k \geq 2 \):

\[
U\left( \sum_{i} p_i x_i \right) = U\left( 1 - \prod_{i=1}^{k} \frac{p_i}{1-p_i} + \prod_{i=1}^{k} x_i \right) \\
\leq (1 - p_{k+1}) U(\cdots) + p_{k+1} U(x_{k+1}) \quad \text{by convexity} \\
\leq (1 - p_{k+1}) \sum_{i=1}^{k} \frac{p_i}{1-p_i} U(x_i) + p_{k+1} x_{k+1} \quad \text{by the induction hypothesis}
\]

Supporting hyperplane theorem

\( p \) is the Euclidean distance \( p \) to \((a, b)\). Find \( \hat{p} \) such that \( d(\hat{p}) \leq d(p) \forall p \) on the curve. Since convex functions are continuous, \( \inf_{p} d(p) \) is attained at some \( \hat{p} \). Find the perpendicular bisector of the line from \((a, b)\) to \( \hat{p} \). This bisector does not intersect the curve (by convexity) as this would cause a contradiction.

This line is a "separating hyperplane".

From separating to supporting hyperplanes

Find \((a_i, b_i)\) such that \( b_i < U(a_i) \) and \( a_i \to c, b_i \to U(c) \)

For each \( i \), there exists a separating line \( y = a_i x + b_i \)

\( \{ (a_i, b_i) : i \geq 1 \} \) possesses some limit point \((a, b)\). The line \( y = ax + b \) is the required supporting hyperplane.
S.2 Chebyshev's Inequality

If \( \text{Var}(X) \) is small, in what sense is \( X \) near to a constant?

**Theorem (Markov's Inequality)**

If \( E(X) \) exists, then \( P(|X| \geq a) \leq \frac{E(|X|)}{a} \) for \( a > 0 \)

**Proof**

Let \( A = \{ |X| \geq a \} \). Then \( |X| \geq a \) \( A \)

\[ E(|X|) \geq E(a \cdot 1_A) = a \cdot P(A) \]

**Theorem (Chebyshev's Inequality)**

\[ P(|X - E(X)| \geq a) = \frac{\text{Var}(X)}{a^2}, \quad a > 0 \]

**Proof**

\[ P(|X - E(X)| \geq a) = P\left( \frac{|X - E(X)|^2}{a^2} \geq 1 \right) \leq \frac{1}{a^2} E\left[ (X - E(X))^2 \right] \text{ by Markov} \]

\[ P\left( |X - E(X)| \geq a \right) \leq \frac{\text{Var}(X)}{a^2} \]

\[ P(X \geq a) = P(e^{\theta X} \geq e^{\theta a}) \quad \theta > 0 \]

\[ \leq \frac{E(e^{\theta X})}{e^{\theta a}} \]

\[ \therefore P(X \geq a) \leq \inf \left\{ e^{-\theta a} E(e^{\theta X}) : \theta > 0 \right\} \]

This leads to the theory of large deviations.
5.3 Law of large numbers (relating to repeated experimentation)

Theorem: Let $X_1, X_2, \ldots$ be iid random variables with finite variance and mean $\mu$. Let $S_n = \sum_{i=1}^{n} X_i$

a) $E\left[\left(\frac{S_n}{n} - \mu\right)^2\right] \to 0$ as $n \to \infty$

b) $P\left(\left|\frac{S_n}{n} - \mu\right| > \epsilon\right) \to 0$ as $n \to \infty$ for $\epsilon > 0$

c) "mean square convergence" "convergence in $L^2$"

d) The weak law of large numbers. There is also a strong law.

Language

$X_n \Rightarrow X$ in mean square or $L^2$, if $E\left[(X_n - X)^2\right] \to 0$

$X_n \Rightarrow X$ in probability if $\forall \epsilon > 0$

$P\left(\left|X_n - X\right| > \epsilon\right) \to 0$ as $n \to \infty$

There are also other forms of convergence

Repeated Experimentation

Repeat an Experiment. Each time, we observe whether $A$ occurs or not, $A_i = \{A \text{ occurs on the } i^{th} \text{ experiment}\}$

$\frac{1}{n} \sum_{i=1}^{n} I_{A_i}$ should converge to something which we can interpret as $P(A)$.

Proof:

a) $E\left(\frac{S_n}{n}\right) = \frac{1}{n} E(S_n) = \frac{1}{n} n \mu = \mu$

$E\left[\left(\frac{S_n}{n} - \mu\right)^2\right] = \text{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2} \text{Var}(S_n) = \frac{1}{n} \text{Var}(X) \to 0$ as $n \to \infty$
1) \( P \left( \left| \frac{S_n}{n} - \mu \right| > \varepsilon \right) \leq \frac{\text{Var} \left( \frac{S_n}{n} \right)}{\varepsilon^2} \) by Chebyshev

\[
= \frac{\text{Var}(x)}{n\varepsilon^2} \to 0 \quad n \to \infty
\]
4.4 Functions of Random Variables

We may have a collection \( X = (X_1, X_2, \ldots, X_n) \) on \((\Omega, \mathcal{F}, \mathbb{P})\).

We use a joint distribution function:

\[
F_X(x) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n)
\]

If \( F_X(x) = \int_{\mathbb{R}^n} f_X(u) \, du \quad \text{for } x \in \mathbb{R}^n \)

and \( f_X(x) > 0 \), \( F_X \) is the joint distribution function of \( X \).

Normally \( f_X(x) = \frac{1}{\prod_{i=1}^{n} \Delta x_i} F_X(x) \)

Note: If \( F_{x,y}(x, y) = \mathbb{P}(X \leq x, Y \leq y) \)

a) The marginal distribution function of \( X \) is \( F_X(x) = F_{x,y}(x, y) \)

b) The marginal pdf of \( X \) is \( f_X(x) = \frac{d}{dx} F_{x,y}(x, \infty) \)

\[
f_X(x) = \frac{d}{dx} \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(u, v) \, du \, dv = \int_{-\infty}^{\infty} f(x, v) \, dv
\]

c) The "basic element of probability" is \( \Delta \)

\[
p(x < X < x + dx, y < Y < y + dy) = f_{x,y}(x, y) \, dx \, dy
\]

\[
p[(X, Y) \in A] = \int_A f_{x,y}(x, y) \, dx \, dy
\]

d) \( X, Y \) are independent if the joint distribution function factorizes

as \( F_{x,y}(x, y) = F_X(x) F_Y(y) \), \( x, y \in \mathbb{R} \)

i.e., in the continuous case \( f_{x,y}(x, y) = f_x(x) f_y(y) \), \( x, y \in \mathbb{R} \)

Reminder: \( A_1, A_2, \ldots, A_n \)

are independent events if and only if

\[
p(A_1 \cap A_2 \cap \cdots \cap A_n) = \prod_{i=1}^{n} p(A_i)
\]

We can discuss the independence of a family \( \{X_1, X_2, \ldots, X_n\} \) of random variables

in a similar fashion.
Application
If X, Y are independent, \( E[g(X)h(Y)] = E[g(X)]E[h(Y)] \)
and hence \( \text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) \)

e) The conditional distribution function of Y given X is
\[
\lim_{dx \to 0} \frac{1}{dx} \int_{-\infty}^{\infty} f(x, y) dy = \frac{f(x, y)}{f_x(x)}
\]

Definition
The conditional density function of Y given X is
\[
f_{Y|X}(y|X) = f_{X,Y}(x,y) / f_X(x)
\]

f) The conditional expectation of Y given X is
\[
\Psi(x) = E(Y|X=x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy
\]

Theorem \( E[E(Y|X)] = E(Y) \)

4.5 Change of variable

General Question
If X, Y have pdf f
\[ U = u(X,Y), V = v(X,Y) \]
What is the joint pdf of (U, V) ?
\[ T: \mathbb{R}^2 \to \mathbb{R}^2, T: (x,y) \mapsto (u(x,y), v(x,y)) \]
\[ (U, V) = T(X, Y) \]
03/03/11

\[ p[(u, v) \in B] = p[(x, y) \in T^{-1}(B)] = \iiint_{(B)} f(u, v) \, du \, dv \]
\[ D = \{(x, y) : f(x, y) > 0\} \]

Let \( S \) be \( T(D) \). \( T : D \to S \)

Assume that \( T \) is invertible on \( S \), i.e. \( T \) is bijective on \( D \).

\[ p[(u, v) \in B] = \iiint_{(B)} f(x, y) \, dx \, dy \]
\[ = \iiint_{(B)} f[x(u, v), y(u, v)] |J| \, du \, dv \]

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

\[ \therefore f_{u, v}(u, v) = \begin{cases} f[x(u, v), y(u, v)] |J| & , \text{ } u, v \in S \\ 0 & , \text{ } u, v \notin S \end{cases} \]

Example

\( X, Y \) are independent, \( \operatorname{Exp}(1) \).

Let \( U = X + Y \), \( V = \frac{X}{X + Y} \)

\[ f(x, y) = e^{-x-y} \quad \text{for } x, y > 0 \]

\( u = x + y \), \( v = \frac{x}{x + y} \), \( \therefore x = uv \), \( y = u(1 - v) \)

\( T : (0, \infty)^2 \to (0, \infty) \times (0, 1) \), a bijection

Jacobian = \[ \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} v & u \\ -v & -u \end{vmatrix} = -u \]

\[ \therefore f_{u, v}(u, v) = e^{-uv-u(v)} |J| = 0 \quad u > 0, \quad 0 < v < 1 \]

\[ \begin{align*}
\text{Here: } f_u(u) &= u e^{-u}, u > 0 \quad f_v(v) = 1, \quad 0 < v < 1
\end{align*} \]
and \( f_{u,v}(u, v) = f_u(u) f_v(v) \)

so \( u \) and \( v \) are independent
Example \(X, Y\) are independent, \(N(0, 1)\)
\[
R = \sqrt{x^2 + y^2} \quad \theta = \arctan \left( \frac{y}{x} \right)
\]
Use \(x = r \cos \theta, \ y = r \sin \theta\)
\[
f_{R, \theta}(r, \theta) = f_{x, y}(r \cos \theta, r \sin \theta) = \frac{r}{2\pi} e^{-\frac{1}{2} r^2} \quad r > 0, \theta \in [0, 2\pi]
\]
Therefore \(R, \theta\) are independent, \(\theta\) is uniform \([0, 2\pi]\)
\(R\) has pdf \(r e^{-\frac{1}{2} r^2}, r > 0\)

4.6 Bivariate (multivariate) normal distribution

\[
N(\mu, \sigma^2) \quad f(x) = \frac{1}{\sqrt{2\pi \sigma^2}} \exp \left[ -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right]
\]
Exercise. The mean is \(\int_{-\infty}^{\infty} x f(x) \, dx = \mu\)
The variance is \(\int_{-\infty}^{\infty} (x - \mu)^2 f(x) \, dx = \sigma^2\)
If \(X \sim N(\mu, \sigma^2)\), \(Y = \frac{x - \mu}{\sigma} \) is \(N(0, 1)\).

Bivariate Case

\[
f(x, y) = c_1 \exp \left[ -c_2 Q(x, y) \right]
\]
where \(Q\) is a quadratic form in \(x\) and \(y\). We take
\[
Q(x, y) = \left( \frac{x - \mu_1}{\sigma_1} \right)^2 + \left( \frac{y - \mu_2}{\sigma_2} \right)^2 - 2\rho \left( \frac{x - \mu_1}{\sigma_1} \right) \left( \frac{y - \mu_2}{\sigma_2} \right)
\]
\[
f(x, y) = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1 - \rho^2}} \exp \left[ -\frac{1}{2} Q(x, y) \right], \quad x, y \in \mathbb{R}
\]
Parameter: \(\sigma_1, \sigma_2 > 0, \mu_1, \mu_2 \in \mathbb{R}, \rho < 1\)
Note \(Q(x, y) = (x - \mu)^T \Sigma^{-1} (x - \mu)\)
\(X = (x, y), \ M = (\mu_1, \mu_2), \ \Sigma = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}\)
\[
\Sigma = \begin{pmatrix} \text{Var}(x) & \text{Cov}(x, y) \\ \text{Cov}(x, y) & \text{Var}(y) \end{pmatrix}
\]
Let \( U = \frac{X - \mu_1}{\sigma_1} \), \( V = \frac{Y - \mu_2}{\sigma_2} \)

\[
    f_{u,v}(u,v) = \frac{1}{2\pi(1-\rho^2)} \exp \left[ -\frac{u^2 - 2\rho uv + v^2}{2(1-\rho^2)} \right]
\]

\[
    f_u(u) = \int_{-\infty}^{\infty} f_{u,v}(u,v) \, dv
\]

\[
    = \int_{-\infty}^{\infty} \frac{1}{2\pi(1-\rho^2)} \exp \left[ -\frac{(u - \rho v)^2}{2(1-\rho^2)} \right] \, dv
    = \frac{1}{\sqrt{2\pi(1-\rho^2)}} N(\rho u, 1-\rho^2)
\]

\( U \) is normal with parameters \( 0 \) and \( 1 \), \( V \) is also \( N(0,1) \).

\( f_{u,v}(u,v) = f_u(u) f_{v|u}(v|u) \). Given \( U = u \), \( V \) is \( N(\rho u, 1-\rho^2) \)

\[
    A(u,v) = (u,v) A(u)
\]

\[
    A = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} = \begin{pmatrix} \rho & \rho \\ \rho & \rho \end{pmatrix}^{-1} = \left( \frac{\text{Var}(V)}{\text{cov}(u,v)} \right)^{-1}
\]

Correlation and Covariance

\[
    E(uv) = E[E(uv|u)] = E[uE(v|u)]
\]

\[
    E(uv) = E(U \cdot \rho V) = \rho E(U^2) = \rho \text{Var}(U) = 1
\]

\[
    \therefore \rho \text{ is } \rho \text{ Cov}(U,V) = \text{corr}(U,V)
\]

**Very Important Properties**

1. \( U, V \) are independent if and only if they are uncorrelated (i.e., \( \rho = 0 \))

2. If \( U \) and \( V \) have a bivariate normal distribution then

   \( \alpha U + \beta V \) has a normal distribution for any given \( \alpha, \beta \in \mathbb{R} \).

   Actually, this characterizes the normal distribution.
Hence, for example: \((X_1, X_2, \ldots, X_n)\) is said to have a multivariate normal distribution, MUN, if:

\[
\sum a_i X_i \text{ is univariate normal, for any } a_1, a_2, \ldots, a_n \in \mathbb{R}.
\]

In general:

For \(X = (X_1, X_2, \ldots, X_n)\)

- The mean vector is \(\mu = (EX_1, EX_2, \ldots, EX_n)\)
- The covariance matrix is \(V = (V_{ij})_{n \times n} = \text{Cov}(X_i, X_j)\)

\[
E\left[ (X - \mu)(X - \mu)^T \right]
\]
6. Geometrical Probability

6.1 Bertrand's Paradox

A chord of the unit circle is picked at random. What is the probability that an equilateral triangle with the chord as its base, fits within the circle?

a) Assume $D$ is Unit $[0, 1]$. $D \geq \frac{1}{2}$ gives the largest allowed triangle, so the triangle lies in the circle if and only if $D > \frac{1}{2}$. \( p(D > \frac{1}{2}) = \frac{1}{2} \)

b) Assume the acute angle $\theta$ between the chord and tangent at an endpoint is Unit $[0, \frac{\pi}{2}]$.

\[
\text{Probability} = p(\theta > \frac{\pi}{3}) = \frac{\frac{\pi}{3}}{\frac{\pi}{2}} = \frac{2}{3}
\]

c) Pick a point uniformly on the disc. Draw a chord with this point on centre.

\[ p(D \leq d) = \frac{\pi d^2}{\pi} = d^2 \quad \text{for} \quad d \in (0, 1) \]

Answer = \( p(D > \frac{1}{2}) = 1 - p(D \leq \frac{1}{2}) = 1 - \frac{1}{4} = \frac{3}{4} \)

d) Choose $p$ and $q$ as independent points on the circumference, each having the uniform distribution. Answer = $\frac{2}{3}$

6.2 Buffon's Needle

A unit needle is dropped at random onto a plane, ruled by parallel straight lines, each parallel lines of which is a unit distance apart.

What is the probability the needle intersects some line?
Let \((X, Y)\) be the coordinates of the centre of the needle, \(\theta\) be the inclination to the \(x\)-axis. Assume:

\[ Z = Y - L Y \quad \text{in} \quad \text{Unif}[0, 1] \quad \text{.} \quad Y, \theta \text{ are independent} \]

\[ \theta \quad \text{in} \quad \text{Unif}[0, \pi] \]

\[ f_{Z, \theta}(z, \theta) = \frac{1}{\pi} \quad \text{for} \quad 0 \leq z \leq 1, \quad 0 \leq \theta \leq \pi \]

For what pairs \((z, \theta)\) is there an intersection?

This intersection occurs if \(Z < \frac{1}{2} \sin \theta\) or \(Z \geq 1 - \frac{1}{2} \sin \theta\) \((*)\)

\[ p(\text{intersection}) = \int_0^\pi f_{Z, \theta}(z, \theta) \, dz \, d\theta \]

\[ B = \{ (z, \theta) \in [0, 1] \times [0, \pi] \mid (*) \text{ holds} \} \]

\[ p(\text{intersection}) = \frac{1}{\pi} \int_0^\pi \sin \theta \, d\theta = \frac{1}{\pi} \left[ -\cos \theta \right]_0^\pi = \frac{2}{\pi} \]

Therefore, Buffon's Needle can be used to estimate \(\pi\).

By repeated experimentation one may obtain a numerical estimate for \(\pi\).

The rate of convergence depends on the variance of the number of intersections is \(n\) throws. Mathematically, let \(I = \{ \text{intersection} \}\)

Let \(I_i\) be the indicator function of \(I\).

\[ E(I_i) = \frac{2}{\pi} \quad , \quad \text{Var}(I_i) = \frac{2}{\pi} \left( 1 - \frac{2}{\pi} \right) \]

Buffon's Cross

Throw \(n\) times. \(Z^* = \# \text{ intersections overall}\)

\[ E(\frac{Z^*}{n}) = \frac{n}{\pi}, \quad \frac{1}{n} \text{Var}(\frac{Z^*}{n}) = \frac{3 - \frac{12}{\pi}}{\pi} - \frac{\frac{4}{\pi^2}}{\pi^2} \]
Buffon's cross provides a better estimate as this estimate converges faster.

What happens if we use a needle of length \( L < 1 \)?

Intersections occur if \( z \leq \frac{L}{2} \sin \theta \) or \( z \geq 1 - \frac{L}{2} \sin \theta \)

The probability turns out to be \( \frac{2L}{\pi} \)

**Buffon's Noodle**

What can we say about dropping a flexible needle?
Baffles Noodle

Take a ruled plane

Drop a noodle of length L onto the grid.

I := \# intersections with the lines

\[
E(I) = \sum_{\text{segments}} \frac{1}{\text{segment length}} \cdot \frac{2L}{\pi} \approx \frac{2L}{\pi}
\]

6.3 Broken Sticks

Take a stick of unit length and break it in two places, X, Y, chosen uniformly on [0, 1], independently of each other.

What is the probability that the three small sticks can form a triangle?

\[ U = \min \{X, Y\} \quad V = |X - Y| \quad W = 1 - U - V \]

Condition to be able to construct a Triangle:
\[ U < V + W \quad V < U + W \quad W < U + V \]

\[ \Rightarrow U, V, W < \frac{1}{2} \quad [\text{what about equality?}] \]

\[ \begin{align*}
& \text{Shaded area } \mathcal{B} \\
& \text{either } X < Y \quad \text{or } X > Y \\
& X < \frac{1}{2} \\
& X - Y < \frac{1}{2} \\
& 1 - Y < \frac{1}{2} \\
& X > Y \Rightarrow Y < \frac{1}{2} \\
& X - Y < \frac{1}{2} \\
& 1 - X < \frac{1}{2} \\
\end{align*} \]

\[ \mathbb{P} [(X, Y) \in \mathcal{B}] = |\mathcal{B}| = \frac{1}{4} \]

Note: Generalize to n breaks and the answer is \( 1 - \frac{n+1}{2^n} \)
7. Central Limit Theorem

Consider $X_1, X_2, \ldots$ iid. $E(X_i) = \mu$, $\text{Var}(X_i) = \sigma^2 \in (0, \infty)$

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

Law of Large Numbers $S_n \xrightarrow{P} n\mu$

Central Limit Theorem $S_n \xrightarrow{D} \mu + N(0, \sigma^\infty)$, $N$ is Normal $(0, 1)$

"Normalise" $\frac{S_n - n\mu}{\sqrt{n\sigma^2}}$

Central Limit Theorem:

Under the above assumptions, $P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq x\right) \to \Phi(x)$

where $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$

"$\frac{S_n - n\mu}{\sqrt{n\sigma^2}}$" is asymptotically $N(0, 1)$.

Definition

The Moment Generating Function (MGF) is defined as, for a random variable $X$: $M_X(t) = E(e^{tX})$ for any $t$ for which this is finite.

Note: If $X$ takes values in $\{0, 1, 2, \ldots\}$,

$M_X(t) = E[(e^t)^X] = G_X(e^t)$

Examples

a) $\text{Exp}(\lambda)$

$M_X(t) = E(e^{tx}) = \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx$

$M_X(t) = \lambda \int_0^\infty e^{-\lambda(x-t)} dx = \left\{ \begin{array}{ll} \lambda & 0 \leq t < \lambda \\ \infty & t \geq \lambda \end{array} \right.$

b) $N(0, 1)$

$M_X(t) = \int_{-\infty}^{\infty} e^{tx} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx = \int_{-\infty}^{\infty} \exp \left[-\frac{1}{2}(x-t)^2\right] \frac{e^{tx^2}}{\sqrt{2\pi}} dx$

$= e^{\frac{1}{2}t^2}, \ t \in \mathbb{R}$
Cauchy Distribution: \( f(x) = \frac{1}{\pi(1+x^2)} \)

\[ M(t) = \int_{-\infty}^{\infty} \frac{e^{tx}}{\pi(1+x^2)} \, dx \]

The Cauchy distribution has infinite mean and variance.

**Properties of the Moment-Generating Function**

**A) Uniqueness Theorem**

If \( M(t) < \infty \) on some neighborhood of the origin 0, then there is a unique distribution with MGF \( M(t) \).

**B) Continuity Theorem**
For a random variable $X$, the mgf is $M(t) = E(e^{tx})$

A. Uniqueness

If $M$ is the mgf of some distribution, and $M(t) < \infty$ for $|t| < \epsilon$ and some $\epsilon > 0$, then this is the unique distribution with mgf $M$.

B. Continuity Theorem

If $Y_1, Y_2, \ldots$ are random variables, such that for all $t$:

$M_{Y_n}(t) \rightarrow e^{t^2}$ as $n \rightarrow \infty$

then $P(Y_n \leq x) \rightarrow \Phi(x)$ as $n \rightarrow \infty$

where $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$

$C. \quad M_{X+Y}(t) = E(e^{t(x+y)}) = e^{tb} E(e^{tX}) = e^{tb} M_X(at)$

$D. \quad M_{X+Y}(t) = E[e^{t(x+y)}] = E[e^{tx} \cdot e^{ty}] = M_X(t) M_Y(t)$

$E. \quad M_X(t) = E(e^{tx}) = E(1 + tx + \frac{t^2 x^2}{2!} + \ldots) = 1 + tE(x) + \frac{t^2 E(x^2)}{2!} + \ldots$

Generating function, $G_x(t) = \sum \frac{t^n}{n!} q_n$

Exponential generating function of $e^x$

The above is ok if $M < \infty$ on some neighborhood of 0.

Central Limit Theorem

If $X_1, X_2, \ldots$ are i.i.d., mean $\mu$, variance $\sigma^2 \neq 0$

$S_n = \sum X_i$

$P(\frac{S_n - n\mu}{n\sigma^2} \leq x) \rightarrow \Phi(x)$
Proof. WLOG, take $\mu = 0$, $\sigma^2 = 1$ (let $X_i = \frac{X_i - \mu}{\sigma}$)

$M_{\frac{S_n}{\sigma}}(t) = M_{\frac{S_n}{\sigma}}(\frac{t}{\sqrt{n}})$ by C

$= M_{\frac{t}{\sqrt{n}}} (\frac{t}{\sqrt{n}})^n$ by D

$= (1 + \frac{t^2}{2n} + o(\frac{t^2}{n}))^n$ by E

$\to e^{\frac{t^2}{2}}$ as $n \to \infty$

Hence the claim holds, by the continuity theorem.

Example

An unknown fraction $p$ of the population vote for unlimited Higher Education fees. It is desired to estimate $p$ by asking a sample of size $n$. Allow an error in estimate $\leq 0.05$.

What $n$ should be used?

Assume each individual votes yes with probability $p$ independently of all others. Let $X_i = 1$ if the $i^{th}$ says yes, $0$ if no.

$S_n = \sum X_i$, the $\bar{X} = \frac{S_n}{n}$ to estimate $p$.

$p\left(\left|\frac{S_n}{n} - p\right| < 0.005\right)$

$= p\left(\left|\frac{S_n - np}{np(1-p)}\right| < 0.005\sqrt{n}\right) = p(1-p) \leq \frac{1}{4}$

$\geq p\left(\left|\frac{S_n - np}{np(1-p)}\right| < 0.005\sqrt{n}\right)$

We agree to tolerate mistakes that have probability $\leq 5\%$, say.

as $n \to \infty$, this is approximately

$\approx \int_{-0.005\sqrt{n}}^{0.005\sqrt{n}} e^{-\frac{u^2}{2}} du = 2 \Phi (0.005\sqrt{n}) - 1 \times 0.95$

if $n \approx 40,000$
An application of the Central Limit Theorem

1) \( X_i \sim \text{Bern}(\rho) \)

\[ S_n = \sum_{i=1}^{n} X_i \]

\[ p(\frac{S_n - np}{\sqrt{np(1-p)}} \leq \alpha) = \Phi(\alpha) \]

\[ \sum_{k: \frac{k}{n} \leq \alpha} \binom{n}{k} \rho^k (1-\rho)^{n-k} \]

\[ \Rightarrow \int_{-\infty}^{\frac{\alpha}{\sqrt{\rho}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \]

2) \( \rho = \frac{1}{2} \)

\[ \sum_{k=\frac{k}{n} \leq \frac{\alpha}{\sqrt{\rho}}} \binom{n}{k} \approx 2^n \int_{-\alpha}^{\alpha} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \]

3) \( X_1, X_2, \ldots \sim \text{iid, Poisson}(1) \)

\[ S_n = \sum_{i=1}^{n} X_i \sim \text{Poisson}(n), \quad \mu = \sigma^2 = n \]

\[ p(\frac{S_n - n}{\sqrt{n}} \leq \alpha) \approx \Phi(\alpha) \]

\[ \sum_{k: \frac{k}{n} \leq \alpha} \frac{e^n}{k!} \approx e^n \int_{-\alpha}^{\alpha} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \]

Also

\[ e^{-n} \left( 1 + n + \frac{n^2}{2!} + \ldots + \frac{n^n}{n!} \right) \xrightarrow{n \to \infty} \frac{1}{2} \]