Axioms of Probability

1. $0 \leq \Pr(A) \leq 1$
2. $\Pr(S) = 1$
3. If $A_1, A_2, ...$ are mutually exclusive (disjoint),
   i.e. $A_i \cap A_j = \emptyset$ when $i \neq j$, then $\Pr(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \Pr(A_i)$
   In particular, if events A and B are mutually exclusive, then $\Pr(A \cup B) = \Pr(A) + \Pr(B)$

Properties of Probability

1. $\Pr(\emptyset) = 0$
2. If $A_1, A_2, ..., A_n$ are mutually exclusive events, then $\Pr(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \Pr(A_i)$
3. $\Pr(A') = 1 - \Pr(A)$
4. $\Pr(A) = \Pr(A \cap B) + \Pr(A \cap B')$
5. $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$
6. $Pr(A \cup B \cup C) = \Pr(A) + \Pr(B) + \Pr(C) - \Pr(A \cap B) - \Pr(B \cap C) - \Pr(A \cap C) + \Pr(A \cap B \cap C)$

Conditional Probability, $P(A \mid B)$

- $\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$, if $\Pr(A) \neq 0$
- For fixed A, $\Pr(B \mid A)$ satisfies the postulates of probability.
- False positive: $\Pr(+ \mid \text{condition})$

Multiplication rule

- $\Pr(A \cap B) = \Pr(A) \Pr(B \mid A) = \Pr(B) \Pr(A \mid B)$, providing $\Pr(A) > 0, \Pr(B) > 0$
- $\Pr(A \cap B \cap C) = \Pr(A) \Pr(B \mid A) \Pr(C \mid A \cap B)$
- $\Pr(A_1 \cap ... \cap A_n) = \Pr(A_1) \Pr(A_2 \mid A_1) \Pr(A_3 \mid A_1 \cap A_2) ... \Pr(A_n \mid A_1 \cap ... \cap A_{n-1})$

The Law of Total Probability

- Let $A_1, A_2, ..., A_n$ be a partition of sample space S (mutually exclusive and exhaustive events s.t. $A_i \cap A_j = \emptyset$ for $i \neq j$ and $\cup_{i=1}^{n} A_i = S$).
- Then $\Pr(B) = \sum_{i=1}^{n} \Pr(B \cap A_i) = \sum_{i=1}^{n} \Pr(A_i) \Pr(B \mid A_i)$

Bayes' Theorem

- Let $A_1, A_2, ..., A_n$ be a partition of S
- $\Pr(A_k \mid B) = \frac{\Pr(A_k) \Pr(B|A_k)}{\sum_{i=1}^{n} \Pr(A_i) \Pr(B|A_i)} = \frac{\Pr(A_k) \Pr(B|A_k)}{\Pr(B)}$, $k \in [1, n]$

Independent Events

- Definition: iff $\Pr(A \cap B) = \Pr(A) \Pr(B)$

Properties

- Suppose $\Pr(A) > 0, \Pr(B) > 0$, A and B are independent:
  - $\Pr(B \mid A) = \Pr(B)$ and $\Pr(A \mid B) = \Pr(A)$
  - A and B cannot be mutually exclusive (and vice versa)
- The sample space S and $\emptyset$ are independent of any event
- If $A \subset B$, then A and B are dependent unless $B = S$

Theorem

If A, B are indep, then so are A and B', A' and B, A' and B'.

n Independent Events

- **Pairwise Independent Events**:
  Events $A_1, A_2, ..., A_n$ are pairwise indep
  iff $\Pr(A_i \cap A_j) = \Pr(A_i) \Pr(A_j)$
- **Mutually Independent**:
  Events $A_1, A_2, ..., A_n$ are (mutually) independent iff for any subset $\{A_{i_1}, A_{i_2}, ..., A_{i_k}\}$ of $A_1, A_2, ..., A_n$,
  $\Pr(A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_k}) = \Pr(A_{i_1}) \Pr(A_{i_2}) ... \Pr(A_{i_k})$

Remarks

- $A_1, A_2, ..., A_n$ are mutually independent $\Leftrightarrow$ for any pair of events A_j, A_k where $j \neq k$, the multiplication rule holds, for any 3 distinct events, the multiplication rule holds, and so on $\Pr(A_1 \cap A_2 \cap ... \cap A_n) = \Pr(A_1)\Pr(A_2)...\Pr(A_n)$. In total there are $2^n - n - 1$ different cases.
- Mutually indep $\Rightarrow$ pairwise indep (not the converse)
- Suppose $A_1, A_2, ..., A_n$ are mutually indep events, let $B_i = A_i$ or A_i', $i \in [1, n]$. Then $B_1, B_2, ..., B_n$ are also mutually indep events.

Discrete Probability Distributions

Discrete R.V.

Let X be an R.V. If R_X is **finite or countable infinite**, X is discrete R.V.

Probability Function (p.f.) or Probability Mass Function (p.m.f.)

- For a discrete R.V., each value X has a certain probability $f(x)$. Such a function $f(x)$ is called the p.f.
- The collection of pairs $(x_i, f(x_i))$ is prob distribution of X
- The probability of $X = x_i$ denoted by $f(x_i)$ must satisfy:
  1. $f(x_i) \geq 0 \forall x_i$
  2. $\sum_{i=1}^{\infty} f(x_i) = 1$

Continuous Probability Distributions

Continuous R.V.

Suppose that R_X is an **interval or a collection of intervals**, then X is a continuous R.V.

Probability Density Function (p.d.f.)

- Let X be a continuous R.V.
- p.d.f. $f(x)$ is a function satisfying:
  1. $f(x) \geq 0 \; \forall x \in R_X$
  2. $\int_{R_X} f(x)dx = 1$ or $\int_{-\infty}^{\infty} f(x)dx = 1$ as $f(x) = 0 \; \forall x \notin R_X$
  3. $\forall c, d : c < d$ (i.e. $(c, d) \subset R_X$), $\Pr(c \leq X \leq d) = \int_c^d f(x)dx$

Remarks

- $\Pr(c \leq X \leq d) = \int_c^d f(x)dx$ represents area under the graph of the p.d.f. $f(x)$ between $x = c$ and $x = d$
- Let x_0 be a fixed value, $\Pr(X = x_0) = 0$
- $\leq$ and $<$ can be used interchangeably in a prob statement.
- $\Pr(A) = 0$ does not necessarily imply $A = \emptyset$
- $R_X \in [a, b] \Rightarrow f(x) = 0 \; \forall x \notin [a, b]$

Cumulative Distribution Function (c.d.f.)

- Let X be an R.V., discrete or continuous.
- $F(x)$ is a c.d.f. of X where $F(x) = \Pr(X \leq x)$

c.d.f. for Discrete R.V.

- $F(x) = \sum_{t \leq x} f(t) = \sum_{t \leq x} \Pr(X = t)$
- c.d.f. of a discrete R.V. is a step function
- $\forall a, b$ s.t. $a \leq b$, $\Pr(a \leq X \leq b) = \Pr(X \leq b) - \Pr(X < a) = F(b) - F(a^-)$ where a^- is the largest possible value of X that is strictly less than a
- $R_X \subset \mathbb{Z}, a, b \in \mathbb{Z} \Rightarrow$
  - $\Pr(a \leq X \leq b) = \Pr(X = a$ or $a + 1$ or ... or $b) = F(b) - F(a - 1)$
  - Taking $a = b$, $\Pr(X = a) = F(a) - F(a - 1)$

c.d.f. for Continuous R.V.

- $F(x) = \int_{-\infty}^{\infty} f(t)dt$
- $f(x) = \frac{dF(x)}{dx}$ if the derivative exists
- $\Pr(a \leq X \leq b) = \Pr(a < X \leq b) = F(b) - F(a)$
- $F(x)$ is a non-decreasing function: $x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$
- $0 \leq F(x) \leq 1$

Mean and Variance of an R.V.

Expected Value / Mean / Mathematical Expectation

- **Discrete:** $E(X) = \mu_X = \sum_i x_i f(x_i) = \sum_x x f(x)$
- **Continuous:** $E(X) = \mu_X = \int_{-\infty}^{\infty} x f(x) dx$
- **Remark**: The expected value exists provided the sum/integral exists

Expectation of a function of an R.V.

$\forall g(X)$ with p.f. $f_X(x)$

- **Discrete:** $E[g(X)] = \sum_x g(x) f_X(x)$
- **Continuous:** $E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$
- Provided the sum/integral exists.

Variance ($\sigma_X^2 = V(X)$)

- $g(x) = (x - \mu_X)^2$, Let X be an R.V. with p.f. $f(x)$
- $\sigma_X^2 = V(X) = E[(X - \mu_X)^2]$
- $E[(X - \mu_X)^2] = \begin{cases} \sum_x (x - \mu_X)^2 f_X(x) \text{ if X is discrete} \\ \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(X) dx \text{ if X is continuous} \end{cases}$
- $V(X) \geq 0$, $V(X) = E(X^2) - [E(X)]^2$
- **Standard deviation** $= \sigma_X = \sqrt{V(X)}$

Properties of Expectation

1. $E(aX + b) = aE(X) + b$
2. $V(X) = E(X^2) - [E(X)]^2$
3. $V(aX + b) = a^2 V(X)$

Joint Distributions

1. X and Y are independent if $f_{x,y}(x, y) = f_x(x) f_y(y)$ for all x and y.
2. Marginal distribution: $f_x(x) = \int_{-\infty}^{\infty} f_{x,y}(x, y) dy$.
3. Conditional distribution: $f_{x|y}(x|y) = f_{x,y}(x, y)/f_y(y)$.
4. Covariance: $cov(x, y) = E(XY) - E(X)E(Y)$. If X and Y are independent, $cov(x, y) = 0$. Also, $var(aX + bY) = a^2 var(X) + b^2 var(Y) + 2ab \cdot cov(X, Y)$.

Probability Distributions

Binomial

Number of successes in n trials, trials must be independent and have the same probability.
$X \sim B(n, p), P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$
$E(X) = np, V(X) = np(1 - p)$

Negative Binomial

Number of trials before k success. $X \sim NB(k, p)$, X = number of trials till k-th success occurs
PF = $fx(X) = P(X = x) = \binom{x-1}{k-1} p^k (1 - p)^{x-k}$
E(X) = $\frac{k}{p}, V(X) = \frac{(1-p)k}{p^2}$
Geometric distribution is special case where k = 1

Poisson

Number of events occuring in a fixed period of time/space.
$X \sim Poisson(\lambda), \lambda > 0$ is the expected number of occurences in the given time.
$fx(k) = P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$
$E(X) = V(X) = \lambda$

Uniform

$x \sim U(a, b)$
Discrete
Expectation: $\frac{1}{k}\sum_{i=1}^{k} x_i$ Variance: $\frac{1}{k}\sum_{i=1}^{k} x_i^2 - \mu_x^2$
Continuous
Expectation: $\frac{a+b}{2}$ Variance: $\frac{(b-a)^2}{12}$

Exponential

$x \sim exp(\lambda)$
PDF: $f_x(x) = \begin{cases} \lambda e^{-\lambda x} & if x \geq 0, \\ 0 & if x < 0 \end{cases}$ CDF: $f_x(x) = \begin{cases} 1 - e^{-\lambda x} & if x \geq 0, \\ 0 & if x < 0 \end{cases}$
$E(X) = \frac{1}{\lambda}$ $V(X) = \frac{1}{\lambda^2}$
No memory property, a to b = 0 to b - a
$P(X > x) = e^{-\lambda x}$

Normal Distribution

$X \sim N(\mu, \sigma^2)$
PDF: $fx(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$
$E(X) = \mu$ $V(X) = \sigma^2$

Normalisation

$Z = \frac{X-\mu}{\sigma}, Z \sim N(0, 1)$, PDF $= \phi(\cdot)$ CDF $= \Phi(\cdot)$

Sampling and Sampling Distributions

For random samples of size n taken from population with mean μ_x and variance σ_x^2
Sampling distribution of sample mean $\overline{X}$ has mean μ_x and variance $\frac{\sigma_x^2}{n}$

Sample Variance

$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$

Central Limit Theorem

CLT states that samples of large n, sums and means of random samples follow normal distribution.
$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0, 1)$, $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$

χ^2 Distribution

Let Z be standard normal variable. Z^2 follows a χ^2 distribution with one degree of freedom.
$Z_1^2 + Z_2^2 + \ldots + Z_n^2$ follows a χ^2 distribution with n degrees of freedom.
Let $Y \sim \chi^2(n), E(Y) = n, var(Y) = 2n$
For large n, $\chi^2(n)$ follows $N(n, 2n)$.

t-Distribution

$Z \sim N(0, 1)$ and $U \sim \chi^2(n)$ then $T = \frac{Z}{\sqrt{U/n}}$ follows t-distribution with n degrees of freedom.
$E(T) = 0, var(T) = n/(n-2)$
$\frac{X-\mu}{S/\sqrt{n}}$ follows t distribution

F-Distribution

$X \sim F(m,n), F = \frac{U/m}{V/n}$ where $U \sim \chi^2(m)$ and $V \sim \chi^2(n)$
$E(X) = \frac{n}{n-2}$

Estimation

CI For Mean

Case	Population	σ	n	Confidence Interval
I	Normal	known	any	$\overline{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$
II	any	known	large	$\overline{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$
III	Normal	unknown	small	$\overline{x} \pm t_{n-1;\alpha/2} \cdot \frac{s}{\sqrt{n}}$
IV	any	unknown	large	$\overline{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$

Paired Data

if n is small and population is normally distributed
$\overline{d} \pm t_{n-1;\alpha/2} \cdot \frac{s_D}{\sqrt{n}}$
if n is large,
$\overline{d} \pm z_{\alpha/2} \cdot \frac{s_D}{\sqrt{n}}$

Hypothesis Tests

1. Set up null vs alternative hypotheses.
2. Determine the level of significance.
3. Identify statistic, distribution, small/large sample & rejection criteria.
4. Compute based on data and compare by one/double-side.
5. Make conclusion on whether to reject null hypothesis

With variance known and population normal, use $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$

Comparing means: Paired Data

$D_i = X_i - Y_i$
Test Statistic $= T = \frac{\overline{D}-\mu_D}{S_D/\sqrt{n}}$