3 Building Blocks of Probability

1. **Sample space**: Denoted by $S$. Collection of all possible outcomes of random experiment, and each outcome corresponds to one and only one element in the sample space (i.e., sample point).

2. **Event**: Any subset of $S$.

3. **Probability**: Denoted by $P$. Measure of likelihood of event.

**Examples:**

1. Toss coin twice. Event: One Heads and one Tails.
3. Choose a point from interval or a square, outcome.
4. Stock price after one month from today, outcome.
Operations on events

- $A \cap B$, $A \cup B$, and $A'$ (complement). Two events $A$ and $B$ are mutually exclusive (or disjoint) if $A \cap B = \emptyset$.

- $A \cap B$: both $A$ and $B$ occur.
- $A \cup B$: $A$ or $B$ occur.
- $A'$: $A$ does not occur.
PROBABILITY AXIOMS

Let \( P(A) \) denote the probability of event \( A \). Then

1. \( 0 \leq P(A) \leq 1 \) for every event \( A \subset S \).
2. \( P(S) = 1 \).
3. If \( A_1, A_2, \ldots \), is a sequence of mutually exclusive events, then

\[
P(A_1 \cup A_2 \cup \cdots) = \sum_{i=1}^{\infty} P(A_i).
\]

A FEW ELEMENTARY CONSEQUENCES FROM AXIOMS:

- \( P(A) = P(A \cap B) + P(A \cap B') \).
- \( P(A \cup B) = P(A) + P(B) - P(A \cap B) \).
- \( P(\bar{A}) = 1 - P(A) \). In particular, \( P(\emptyset) = 0 \).
Conditional Probability and Independence

Conditional probability: Given that event \( B \) happens, what is the probability that \( A \) happens?

\[
P(A|B) = \frac{P(A \cap B)}{P(B)}
\]

1. Rewrite the definition:

\[
P(A \cap B) = P(B)P(A|B) = P(A)P(B|A).
\]
Law of total probability and Bayes’ rule

Let \( \{B_1, B_2, \ldots, B_n\} \) be a partition of the sample space \( S \), i.e.

\[
S = B_1 \cup \ldots \cup B_n
\]

and

\[
B_i \cap B_j = \emptyset, \ i \neq j
\]

1. Law of total probability:

\[
P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)
\]

2. Bayes’ rule:

\[
P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}
\]
Independence:

1. Two events $A$ and $B$ are independent if
   \[ P(A \cap B) = P(A)P(B). \]
2. If $A$ and $B$ are independent then observe that $P(A|B) = P(A)$
   and $P(B|A) = P(B)$. 
Random Variables

Random variable is a variable whose value is the numerical outcome of a random experiment.

**Random Variable:** A random variable $X$ is a function $X : S \rightarrow \mathbb{R}$. In other words, for every sample point (i.e., possible outcome) $s \in S$, its associated numerical value is $X(s)$.

**Notation:** For any subset $I \subset \mathbb{R}$, the event
\[ \{X \in I\} = \{s : X(s) \in I\}. \]
For example $\{X \leq 2\}$, $\{X = 1\}$, $\{0 < X < 3\}$ are all events, and we can use ideas from probability on events to study.
Discrete random variables

Discrete random variable: A random variable that can only take finitely many or countably many possible values.

1. Toss a fair coin. Win $1 if heads, and lose $1 if tails. $X$ is the total winning.
2. Toss a fair coin. $X$ is the first time a heads appears.
3. Randomly select a college student. His/her IQ and SAT score.
Characterizing Discrete Random Variable

1. Distribution: Let \( \{x_1, x_2, \ldots \} \) be the possible values of \( X \). Let

\[
P(X = x_i) = p_i,
\]

where \( p_i \geq 0 \) and

\[
\sum_i p_i = 1.
\]

2. Cumulative Distribution Function or cdf

\[
F(x) = P(X \leq x) = \sum_{x_i \leq x} p_i.
\]

3. Graphic description: (1) probability function \( \{p_i\} \), (2) cdf \( F \).
Expected Value

Discrete random variable $X$. Possible values $\{x_1, x_2, \ldots \}$, and $P(X = x_i) = p_i$.

1. Expectation (expected value, mean, “$\mu$”):

$$EX = \sum_i x_i P(X = x_i) = \sum_i x_i p_i.$$
Properties of Expected Value

1. **Theorem**: Consider a function $h$ and random variable $X$. Then
   
   $$E[h(X)] = \sum_i h(x_i)P(X = x_i)$$

2. **Corollary**: Let $a, b$ be real numbers. Then
   
   $$E[aX + b] = aEX + b.$$
Variance and Standard Deviation

1. Variance ("\(\sigma^2\)") and Standard deviation ("\(\sigma\)"):

\[
\text{Var}[X] \triangleq \mathbb{E} \left[ (X - EX)^2 \right], \quad \text{Std}[X] \triangleq \sqrt{\text{Var}X}.
\]

2. Example (interpretation of variance): A random variable \(X\),

\[
P(X = 0) = 1/2, \quad P(X = a) = 1/4, \quad P(X = -a) = 1/4.
\]
Properties of Variance

1. Proposition: \( \text{Var}[X] = E[X^2] - (EX)^2. \)

2. Proposition: Let \( a, b \) be real numbers. Then \( \text{Var}[aX + b] = a^2\text{Var}[X]. \)
Examples of Discrete Probability Distributions

1. Bernoulli random variable: A random variable $X$ takes values in $\{0, 1\}$ such that

$$P(X = 1) = p, \quad P(X = 0) = 1 - p.$$ 

Used when there are only two possible outcomes for an experiment.

2. Binomial random variable: A random variable $X$ that takes values in $\{0, 1, \ldots, n\}$ such that

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$ 

This binominal distribution is denoted by $Bin(n, p)$, and we write $X \sim Bin(n, p)$. Number of successes in $n$ independent Bernoulli($p$) trials.
Examples of Discrete Probability Distributions

1. Geometric random variable: A random variable $X$ that takes values in \{1, \ldots, \} such that

$$P(X = k) = p \times (1 - p)^{k-1}, 0 < p < 1$$

we say that $X$ has geometric distribution with parameter $p$. The number of independent Bernoulli($p$) trials needed for a single success.

2. Negative binomial random variable: A random variable $X$ taking values in $r, r+1, \ldots$ such that

$$P(X = k) = \binom{k - 1}{r - 1}(1 - p)^{k-r} p^r,$$

we say that $X$ has negative binomial distribution with parameters $r$ and $p$. The number of independent Bernoulli($p$) trials needed to get $r$ successes.
Examples of Discrete Probability Distributions

1. Poisson random variable: A random variable $X$ that takes values in $\{0, 1, 2, \ldots\}$ and

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

we say $X$ has Poisson distribution with parameter $\lambda$.

2. If $Y \sim Bin(n, p)$ for $n$ large and $p$ small, then

$$P(Y = k) \approx e^{-\lambda} \frac{\lambda^k}{k!}, \text{ where } \lambda = n \times p.$$ 

3. Also talked about hypergeometric and multinomial distributions.
Example of Discrete Random Variables

1. Flip a fair coin \( n \) times, let \( X \) be the number of tails. What is the distribution of \( X \)?

2. Flip a fair coin, let \( X \) be the number of flips until a tail occurs. What is the distribution of \( X \)?
Questions

1. What is the mean and variance of a Bernoulli random variable with parameter $p$?
2. What is the mean and variance of a geometric random variable with parameter $\rho$?
3. What is the mean and variance of a Poisson($\lambda$) random variable?
4. A jumbo jet has 4 engines that operate independently. Each engine has probability 0.002 of failure. At least 2 operating engines needed for a successful flight. Probability of an unsuccessful flight? Use binomial distribution.
5. Two fair dice are thrown $n$ times. Let $X$ be the number of throws in which the number on the first die is smaller than that on the second die. What is the distribution of $X$?
**Continuous Random Variable**

- **Continuous random variable**: A random variable that can take any value on an interval of $\mathbb{R}$.
- **Distribution**: A density function $f : \mathbb{R} \rightarrow \mathbb{R}^+$ such that
  1. non-negative, i.e., $f(x) \geq 0$ for all $x$.
  2. for every subset $I \subset \mathbb{R}$,

\[
P(X \in I) = \int_I f(x) \, dx
\]

3. \[
\int_{\mathbb{R}} f(x) \, dx = 1.
\]

4. Relation between density $f$ and cdf $F$:

\[
F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y) \, dy
\]

\[
f(x) = F'(x).
\]
Continuous Random Variables

Given a continuous random variable $X$,

$$P(X = a) = 0$$

for any real number $a$. Therefore,

$$P(X < a) = P(X \leq a), \quad P(X > a) = P(X \geq a)$$

$$P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b).$$

Example: Suppose $X$ has density $f$ that takes the following form:

$$f(x) = \begin{cases} 
  cx(2 - x), & 0 \leq x \leq 2, \\
  0, & \text{otherwise}.
\end{cases}$$

Determine $c$, cdf $F$, and find $P(X < 1)$. 
**Expected Value and Variance**

1. **Expectation** (expected value, mean, “$\mu$”) of random variable $X$ with density $f$:

   $$EX = \int_{-\infty}^{\infty} xf(x) \, dx$$

2. **Theorem**: Consider a function $h$ and random variable $h(X)$. Then

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

3. **Variance and Standard Deviation**

   $$\text{Var}[X] = E[(X - EX)^2] = \int_{-\infty}^{\infty} (x - EX)^2 f(x) \, dx$$

   and

   $$\text{Std}(X) = \sqrt{\text{Var}[X]}.$$
**Uniform distribution**

1. **Uniform distribution on \([0, 1]\).** A random variable \(X\) with density

\[
 f(x) = \begin{cases} 
 1 & , \quad 0 \leq x \leq 1 \\
 0 & , \quad \text{otherwise.}
\end{cases}
\]

2. Expected value and variance?

3. **Uniform distribution on \([a, b]\).** A random variable \(X\) with density

\[
 f(x) = \begin{cases} 
 \frac{1}{b - a} & , \quad a \leq x \leq b \\
 0 & , \quad \text{otherwise.}
\end{cases}
\]

4. **Remark:** If \(X\) is uniformly distributed on \([0, 1]\), then \(Y \equiv (b - a)X + a\) is uniformly distributed on \([a, b]\).

5. Expected value and variance?


**Exponential Distribution**

1. **Exponential distribution with rate** $\lambda$. A random variable $X$ with density

\[ f(x) = \begin{cases} 
\lambda e^{-\lambda x}, & 0 \leq x < \infty \\
0, & \text{otherwise}.
\end{cases} \]

We say $X \sim \text{exponential}(\lambda)$.

2. The only continuous distribution with the memoryless property: If $X \sim \text{exponential}(\lambda)$ then

\[ P(X > t + h|X > t) = P(X > h). \]
Gamma Distribution

1. Gamma distribution with shape parameter $k$, and scale parameter $\lambda$ is a random $X$ variable with density

$$f(x) = \begin{cases} \lambda^k x^{k-1} e^{-\lambda x} / \Gamma(k) , & 0 \leq x < \infty \\ 0 , & \text{otherwise} \end{cases}$$

where

$$\Gamma(k) = \int_0^\infty x^{k-1} e^{-x} dx,$$

we write $X \sim \text{Gamma}(k, \lambda)$.

2. If $X_1, \ldots X_n$ are i.i.d and $X_i \sim \text{exponential}(\lambda)$ for $1 \leq i \leq n$ then

$$X = X_1 + \ldots + X_n \sim \text{Gamma}(n, \lambda).$$
Other continuous distributions

1. **Weibull Distribution** with parameters $\lambda$ and $a$ has density

\[ f(x) = \begin{cases} a\lambda x^{a-1}e^{-(\lambda x)^a}, & 0 \leq x < \infty \\ 0, & \text{otherwise,} \end{cases} \]

used to model failure times.

2. **Beta Distribution** with parameters $a$ and $b$ has density

\[ f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}, \]

for $0 \leq x \leq 1$, otherwise $f(x) = 0$.

3. **Standard Cauchy Distribution** has density

\[ f(x) = \frac{1}{\pi(1 + x^2)}, -\infty < x < \infty \]

this distribution doesn’t have mean or variance.
**Joint Distributions**

1. For 2 discrete random variables $X \in \{x_1, \ldots, \}$ and $Y \in \{y_1, \ldots, \}$, describe their joint distribution via their ‘joint pmf’: $p_{i,j} = P(X = x_i, Y = y_j)$. This satisfies
   1.1 $p_{i,j} \geq 0$ for all $i, j$.
   1.2 $\sum_i \sum_j p_{i,j} = 1$.

2. For 2 continuous random variables $X$ and $Y$ describe their joint distribution via the ‘joint pdf’: $f(x, y)$, which satisfies
   2.1 $f(x, y) \geq 0$ for all $x, y$,
   2.2 $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dx \, dy = 1$. 
Marginal Distributions, Independent Random Variables

For two cont. random variables \( X \) and \( Y \) with joint pdf \( f(x, y) \), calculate marginal pdf of \( X \), \( f_X \) as

\[
f_X(x) = \int_{-\infty}^{\infty} f(x, y) \, dy.
\]

Similar calculation for discrete random variables.

- Two cont. random variables \( X \) and \( Y \) with joint pdf \( f(x, y) \) and marginals \( f_X(x) \) and \( f_Y(y) \) are independent if and only if

\[
f(x, y) = f_X(x)f_Y(y) \quad \text{for all } x, y.
\]

Similar result for discrete.

- For two independent random variables \( X \) and \( Y \), and any two functions \( h \) and \( g \),

\[
E[h(X)g(Y)] = E[h(X)]E[g(Y)].
\]
**Covariance and Correlation**

1. For two random variables $X$ and $Y$ we define


2. A normalized version of covariance is correlation

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

3. If $X$ and $Y$ are independent what is $\text{Cov}(X, Y)$?
**Sums of Random Variables**

1. If \( Y = a_1X_1 + \ldots + a_nX_n \), where \( a_i \) are constants and \( X_i \) are random variables then

\[
E[Y] = a_1E[X_1] + \ldots + a_nE[X_n],
\]

and

\[
\text{Var}[Y] = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \text{Cov}(X_i, X_j).
\]

e.g. If \( n = 2 \), \( Y = a_1X_1 + a_2X_2 \) and

\[
\text{Var}[Y] = 2a_1a_2 \text{Cov}(X_1, X_2) + a_1^2 \text{Var}[X_1] + a_2^2 \text{Var}[X_2].
\]

2. If the \( X_i \), \( 1 \leq i \leq n \) are independent then

\[
\text{Var}[Y] = a_1^2 \text{Var}[X_1] + \ldots + a_n^2 \text{Var}[X_n].
\]
Normal Distribution

Normal distribution $N(\mu, \sigma^2)$. A random variable $X$ is normally distributed with mean $\mu$ and variance $\sigma^2$ if it has density

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

We say that $X \sim N(\mu, \sigma^2)$.

1. $f$ defines a probability density function.
2. $EX = \mu$.
3. $\text{Var}[X] = \sigma^2$. 
The standard normal means $N(0, 1)$. Its cdf is often denoted by $\Phi$, that is,

$$
\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{y^2}{2} \right\} \, dy
$$

Suppose $Z$ is $N(0, 1)$, then

$$
P(Z > x) = P(Z < -x) = \Phi(-x).
$$

In particular,

$$
\Phi(x) + \Phi(-x) = 1.
$$
Standardization and Sums of Gaussians

1. Suppose $X$ has normal distribution $N(\mu, \sigma^2)$. Then its linear transform $Y = aX + b$ is also normally distributed with distribution $N(a\mu + b, a^2 \sigma^2)$.

2. Suppose $X$ has distribution $N(\mu, \sigma^2)$. Then its standardization

$$Z = \frac{X - \mu}{\sigma}$$

is $N(0, 1)$.

3. If $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$, and $\text{corr}(X_1, X_2) = \rho$ then

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2).$$
Central Limit Theorem

Why do we call the distribution ‘normal’?

- **Theorem:** If $X_1, \ldots, X_n$ is a collection of independent samples from the same cdf $F$, with $EX_1 = \mu$ and $\text{Var}(X_1) = \sigma^2$. Then the central limit theorem says that for $n$ sufficiently large

\[
\frac{1}{n}(X_1 + \ldots + X_n) \approx N(\mu, \sigma^2/n)
\]

and also

\[
X_1 + \ldots + X_n \approx N(n\mu, n\sigma^2).
\]

- Remember this result only applies if the random variables $X_i$ have a variance $\sigma^2$. For example this result doesn’t hold for the Cauchy distribution because it doesn’t have a variance.
Distributions Related to Gaussian

- If $X_i$, $1 \leq i \leq n$ are i.i.d $N(0, 1)$ random variables, then

\[ Y = X_1^2 + \ldots + X_n^2, \]

has a chi-square distribution with $n$ degrees of freedom. This is written $Y \sim \chi^2_n$. Note for integer $\nu$, $\chi^2_{\nu}$ is a special case of the gamma distribution ($\lambda = 1/2$, $k = \nu/2$).

- If $X \sim N(0, 1)$, $Y \sim \chi^2_{\nu}$, and $X$ and $Y$ are independent then

\[ \frac{X}{\sqrt{Y/\nu}} \sim t_{\nu}, \]

i.e. the ratio has a $t$-distribution with $\nu$ degrees of freedom. The Cauchy distribution is a special case of this distribution ($\nu = 1$).