ECON 3130 Notes

Jinzhou Wu

October 30, 2025

Contents

1 Probabilities and Events

Date: Aug 27, 2025

1.1 Definitions

Definition (Experiment)**.** A process that results in uncertain outcomes.

Definition (Outcome)**.** A possible result of the experiment.

Definition (Sample Space)**.** The set of all possible basic outcomes.

Definition (Event)**.** A set of basic outcomes, i.e., a subset of the sample space.

Definition (Probability)**.** A numerical measure of the chance that an event will occur.

1.2 Probability Distributions

Definition (Probability Distribution)**.** A table assigning probabilities to all basic outcomes of an experiment, where the probabilities sum to 1.

Example: Preferred Pet

Pet	Pr(Pet)
Dog	0.44
Cat	0.30
Other	0.07
None	0.19
Total	1.00

Probability Tables When analyzing multiple attributes (e.g., gender and pet preference), we summarize data in a **probability table**.

Rules:

1. Two attributes: one on rows, one on columns.
2. Events are mutually exclusive and exhaustive.
3. Joint probabilities go in cells; marginal probabilities go in row and column totals.

Example Table: Preferred Pet vs. Gender

	Dog	Cat	Other	None	Pr(Gender)
Female	0.18	0.12	0.04	0.05	0.38
Male	0.32	0.15	0.05	0.09	0.62
Refuse/Other	0.00	0.00	0.00	0.00	0.00
Pr(Pet)	0.50	0.27	0.09	0.14	1.00

1.3 Set Theory

Events can be represented as sets within a sample space.

- $A \cap B$ = Event that both A and B occur.
- $A \cup B$ = Event that A or B (or both) occur.
- A^c = Complement of A, i.e., A does not occur.

Rules for operations on sets:

- **Commutativity:**
$$A \cup B = B \cup A, \quad A \cap B = B \cap A$$

- **Associativity:**
$$A \cup (B \cup C) = (A \cup B) \cup C, \quad A \cap (B \cap C) = (A \cap B) \cap C$$

- **Distributive laws:**
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

- **DeMorgan's Laws:**
$$(A \cup B)^c = A^c \cap B^c$$
$$(A \cap B)^c = A^c \cup B^c$$

1.4 Probability Rules

Rule 1: Intersection
$$\Pr(A \cap B) = \Pr(A \mid B) \Pr(B)$$

Rule 2: Union
$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$

Rule 3: Complement
$$\Pr(A^c) = 1 - \Pr(A)$$

1.5 Conditional Probability

Definition (Conditional Probability)**.** The probability of event A given event B:

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}, \quad \Pr(B) > 0$$

Multiplication Rule
$$\Pr(A \cap B) = \Pr(A \mid B) \Pr(B)$$

2 Discrete Random Variables

Date: Sep 3, 2025

2.1 Independence

Definition (Independence)**.** Two events A and B are **independent** if and only if

$$\Pr(A \cap B) = \Pr(A)\Pr(B).$$

Theorem (Independence and Joint Events)**.** *If A and B are independent, then*

$$Prob(A|B) = Prob(A)$$

The conditional probability = the marginal probability (of the unconditioned event).

2.2 Probability Trees

A way of deriving probabilities from conditional information. They are most useful when one has information that can be arranged sequentially.

Probability Tree Rules:

- The tree is drawn on its side.
- It starts from a circle called a chance node.
- The branches of the tree at each node correspond to possible outcomes. We label the branches with conditional probabilities.
- The probabilities at the end of the tree are joint probabilities obtained by multiplication.

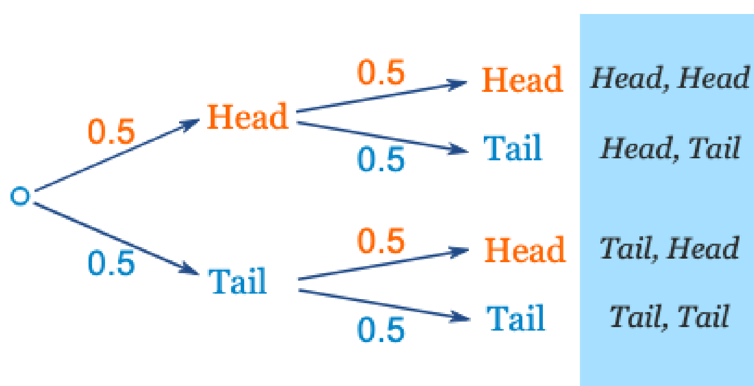


Figure 1: Example of a probability tree.

2.3 The Monty Hall Problem

A classic probability puzzle based on a game show:

- Three doors: behind one is a **car**, behind two are **goats**.

- You choose one door (say Door 1).
- The host, Monty Hall, who knows where the car is, opens another door showing a goat.
- You are given the choice to **stay** or **switch**.

Should you switch?

Solution Let C_i = event that the car is behind door i. Let H = event that Monty opens **Door 3**. Initially:

$$\Pr(C_1) = \Pr(C_2) = \Pr(C_3) = \tfrac{1}{3}.$$

Assume you pick Door 1 and Monty opens Door 3 revealing a goat. By Bayes' theorem:

$$\Pr(C_i \mid H) = \frac{\Pr(H \mid C_i)\Pr(C_i)}{\Pr(H)}.$$

Monty's behavior:

$$\Pr(H \mid C_1) = \tfrac{1}{2}, \qquad \Pr(H \mid C_2) = 1, \qquad \Pr(H \mid C_3) = 0.$$

Thus, the total probability of H:

$$\Pr(H) = \Pr(H \mid C_1)\Pr(C_1) + \Pr(H \mid C_2)\Pr(C_2) + \Pr(H \mid C_3)\Pr(C_3),$$

$$\Pr(H) = \tfrac{1}{2} \cdot \tfrac{1}{3} + 1 \cdot \tfrac{1}{3} + 0 = \tfrac{1}{2}.$$

Finally, the posterior probabilities:

$$\Pr(C_1 \mid H) = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \tfrac{1}{3}, \qquad \Pr(C_2 \mid H) = \frac{1 \cdot \frac{1}{3}}{\frac{1}{2}} = \tfrac{2}{3}.$$

The probability that your original choice is correct is $\frac{1}{3}$, while switching gives $\frac{2}{3}$. **Always switch.**

2.4 Random Variables

Definition (Random Variable)**.** An experiment with numerical outcomes (or outcomes that can be mapped to numbers).

Notation We denote a random variable with a capital letter, like X or Y.

Probability Mass Function (PMF)

$$f(x) = Pr(X = x)$$

Every discrete random variable has a PMF that describes the probability of each possible value.

2.5 Measures of Central Tendency

Expected Value

$$E[X] = \sum_{x \in S} x \cdot f(x)$$

Mode

$$Mode[X] = \text{value (or values) that maximize } f(x)$$

Median

$$Median[X] = \text{value } m \text{ such that } Pr(X \leq m) = Pr(X \geq m) = 0.5$$

2.6 Measures of Dispersion

Variance

$$Var[X] = E[(X - E[X])^2]$$

$$Var[X] = \sum_{x \in S} (x - E[X])^2 \cdot f(x)$$

A Variance Identity

$$\begin{aligned}
Var[X] &= E[(X - E[X])^2] \\
&= E[X^2 - 2XE[X] + (E[X])^2] \\
&= E[X^2] - E[2XE[X]] + E[(E[X])^2] \\
&= E[X^2] - 2E[X]E[X] + (E[X])^2 \\
&= \boxed{E[X^2] - (E[X])^2}
\end{aligned}$$

Variance of a Linear Function of X

$$\begin{aligned}
Var[aX + b] &= E[(aX + b - E[aX + b])^2] \\
&= E[(aX + b - aE[X] - b)^2] \\
&= E[(aX - aE[X])^2] \\
&= E[a^2(X - E[X])^2] \\
&= a^2 E[(X - E[X])^2] \\
&= \boxed{a^2 Var[X]}
\end{aligned}$$

Standard Deviation

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

2.7 Conditional Expectations

Conditional Expectations. If X and Y are random variables, the *conditional expectation* of X given that Y takes on a certain value y is:

$$E[X \mid Y = y] = \sum_{x} x \cdot \Pr(X = x \mid Y = y)$$

- This is the same formula as the one for expected value, except that we have replaced the marginal probability with a conditional probability.
- The conditional expectation of X given $Y = y$ is the probability-weighted average of X given $Y = y$.
- $E[X \mid Y = y]$ is a function that operates on values of Y.

2.8 Law of Iterated Expectations

Theorem (Law of Iterated Expectations)**.**

$$E[E[X \mid Y]] = E[X]$$

The expected value of the conditional expectation of X given Y is equal to the expected value of X.

- This is also called the **law of total expectation**.
- It is useful when you know the distribution of X conditional on Y, but not the marginal distribution of X.
- You can compute $E[X]$ by first computing $E[X \mid Y]$ for each value of Y, then taking the expectation over Y.

$$E[X] = \sum_y E[X \mid Y = y] \cdot \Pr(Y = y) = \sum_y \sum_x x \cdot \Pr(X = x \mid Y = y) \cdot \Pr(Y = y)$$

3 More Discrete Random Variables

Date: Sep 8, 2025

3.1 Moments

Definition (Moment)**.** Numbers that represent qualities of a particular distribution. Expected values of a random variables raised to a different powers. rth moment centered around b of a random variable X is defined as

$$\mu_r = E[(X - b)^r]$$

The mean is the first moment centered around zero. The variance is the second moment centered around E[X].

3.2 Moment-generating function (MGF)

Definition (Moment-generating function)**.** The moment-generating function (MGF) of a random variable X is defined as

$$M_X(t) = E[e^{tX}] = \sum_x e^{tx} f(x) \quad \text{(pmf)}$$

for all t in an open interval containing 0 such that the expectation exists.

The mgf is not always defined for a random variable. If two random variables have the same mgf, then they also have the same pmf.

Finding moments using MGF The rth moment of a random variable X can be found by taking the rth derivative of the mgf and evaluating it at $t = 0$:

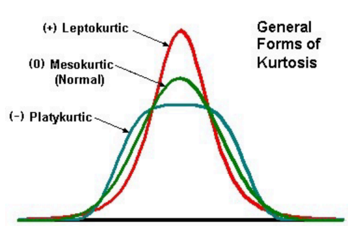
$$\mu_r = E[X^r] = M_X^{(r)}(0) = \sum_x x^r f(x)$$

3.3 Standardized moments

$$\hat{\mu}_k = \frac{E[(X - \mu)^k]}{E[(X - \mu)^2]^{k/2}}$$

kth centered moment divided by the standard deviation raised to the k. Standardized moments are unit-invariant.

Skew is standardized third moment. Negative skew = left tail. Positive skew = right tail. Kurtosis is standardized fourth moment.



Kurtosis measures how fat the tails are. Positive means more extreme values. Negative means fewer extreme values. Note that all three of these random variables have the same variance.

3.4 Bernoulli distribution

Definition (Beroulli distribution)**.** A Bernoulli random variable X takes the value 1 with probability p and the value 0 with probability $1 - p$.

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

The mean of a Bernoulli random variable is $E[X] = p$. The variance of a Bernoulli random variable is $Var(X) = p(1 - p)$.

3.5 Binomial distribution

Definition (Binomial distribution)**.** A Binomial random variable X with parameters n and p counts the number of successes in n independent Bernoulli trials, each with success probability p.

The mean of a Binomial random variable is $E[X] = np$. The variance of a Binomial random variable is $Var(X) = np(1 - p)$ (by the linearity of expectation).

- There are a fixed number of independent trials.
- Each trial has two basic outcomes (e.g., "success"/"failure").
- The probability of a success, p, is constant across trials.

Normal approximation to the Binomial When the number of trials n is large, the Binomial distribution can be approximated by a Normal distribution with the same mean and variance:

$$X \sim \text{Binomial}(n, p) \approx N(np, np(1 - p))$$

This approximation is most accurate when $np \geq 10$ and $n(1 - p) \geq 10$.

Binomial probability mass function (pmf) The probability mass function (pmf) for a Binomial random variable X is:

$$\Pr(X = k \mid n, p) = \frac{n!}{k!(n-k)!} p^k (1 - p)^{n-k} = \binom{n}{k} p^k (1 - p)^{n-k}$$

where:

- n = number of trials (maximum possible number of successes)
- k = number of successes
- $n - k$ = number of failures
- p = probability of a success on each trial
- $1 - p$ = probability of a failure on each trial

X can also be defined as the sum of n independent Bernoulli random variables, each with parameter p:

$$X = \sum_{i=1}^{n} X_i, \quad X_i \sim \text{Bernoulli}(p)$$

4 Continuous Random Variables

Date: Sep 10, 2025

Definition (Continuous Random Variable)**.** A random variable X is said to be continuous if there exists a function $f(x)$ such that for any set A,

$$P(X \in A) = \int_A f(x)\,dx$$

where $f(x)$ is called the probability density function (pdf) of X. Key properties of a legal pdf: $f(x) \geq 0\ \forall x \in S$ and $\int_S f(x)dx = 1$.

4.1 Cumulative Distribution Function (CDF)

$$F(a) = \Pr(X \leq a) = \int_{-\infty}^{a} f(x)dx$$

4.2 Uniform Random Variable

If $X \sim \text{Uniform}(a, b)$, then every value between a and b is equally likely. The probability density function (pdf) is:

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

The cumulative distribution function (CDF) is:

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

The probability that X falls between a and b is:

$$\Pr(a < X < b) = \int_a^b \frac{1}{b-a}dx = 1$$

The expected value and variance of a uniform random variable are:

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

4.3 Expected Value

The expected value (mean) of a continuous random variable X with pdf $f(x)$ is defined as:

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

4.4 Variance

The variance of a continuous random variable X with pdf $f(x)$ is defined as:

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

4.5 Moment-Generating Function (MGF)

The moment-generating function (MGF) of a continuous random variable X with pdf $f(x)$ is defined as:

$$M(t) = \int_{-\infty}^{+\infty} e^{tx} f(x)\,dx, \qquad -h < t < h$$

where h is a constant such that $M(t)$ is finite for t in $(-h, h)$.

For discrete random variables, the MGF is:

$$M(t) = \sum_{x \in S} e^{tx} f(x)$$

The MGF characterizes the distribution if it is finite in some neighborhood around t.

The r-th moment of X can be obtained by differentiating $M(t)$ r times and evaluating at $t = 0$:

$$M^{(r)}(0) = E[X^r]$$

4.6 Quantiles

The p'th quantile π_p is defined as the value such that

$$p = \int_{-\infty}^{\pi_p} f(x)\,dx$$

5 Normal Random Variables

Date: Sep 15, 2025

5.1 Definition and Properties

Definition (Normal Random Variable)**.** A normal random variable is a continuous random variable whose probability density function (pdf) is given by:

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

where μ is the mean and σ^2 is the variance (σ is the standard deviation).

- The normal distribution is symmetric about its mean μ.
- The total area under the pdf is 1.
- The standard normal distribution has $\mu = 0$ and $\sigma^2 = 1$.
- Probabilities are computed as areas under the curve:

$$\Pr(a < X < b) = \int_a^b f(x)dx$$

- $\mathbb{E}(X) = \mu$
- $\mathrm{Var}(X) = \sigma^2$ (and thus $\mathrm{SD}(X) = \sigma$)
- 68%, 95%, 99.7% of outcomes are within 1, 2, 3 standard deviations of the mean, respectively.

A normal random variable X is denoted by:

$$X \sim N(\mu, \sigma^2)$$

which we read as "X is normally distributed with mean μ and variance σ^2."

If $\mu = 0$ and $\sigma = 1$, then the random variable is a **standard normal**, denoted Z.

5.2 Moment Generating Function (mgf) of the Normal

In general, the moment generating function (mgf) of a random variable X is:

$$M(t) = \mathbb{E}[e^{tX}] = \int_{-\infty}^{+\infty} e^{tx} f(x)dx$$

For the normal distribution:

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

This simplifies to:

$$M(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$$

The mgf can be used to compute all moments of the normal distribution by differentiating $M(t)$ with respect to t and evaluating at $t = 0$.

5.3 Linear Combinations of Independent Normals

1. **Sums.** The sum of independent normally distributed random variables is also normally distributed.
   Example: If X and Y are independent and $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, then $W = X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

2. **Constants.** If $Z \sim N(0, 1)$, then for constants a and b, $a + bZ \sim N(a, b^2)$. In general, multiplying a normal random variable by a constant and/or adding a constant yields another normal random variable.

3. **Standardizing.** If $X \sim N(\mu, \sigma^2)$, then the standardized variable $Z = \frac{X-\mu}{\sigma}$ follows a standard normal distribution: $Z \sim N(0, 1)$.

6 More Random Variables

Date: Sep 17, 2025

6.1 Negative Binomial Random Variables

Definition (Negative Binomial Random Variable)**.** A random variable X is said to have a **negative binomial distribution** with parameters r and p, denoted $X \sim \text{NegBin}(r, p)$, if it represents the number of independent Bernoulli trials needed to achieve **exactly r successes**, where each trial has two possible outcomes (success or failure) and the probability of success on each trial is constant and equal to p.

The probability mass function is given by:

$$f(x; r, p) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$$

for $x = r, r+1, r+2, \ldots$

- $f(x; r, p)$: The probability that the r-th success occurs on the x-th trial.
- $\binom{x-1}{r-1}$: The number of ways to arrange $r-1$ successes in the first $x-1$ trials.
- p^r: Probability of r successes.
- $(1-p)^{x-r}$: Probability of $x-r$ failures.

The mean and variance of a negative binomial random variable $X \sim \text{NegBin}(r, p)$ are:

$$E[X] = \frac{r}{p}$$

$$\text{Var}[X] = \frac{r(1-p)}{p^2}$$

Derivation:

Let X be the number of trials needed to get r successes. X can be written as the sum of r independent geometric random variables Y_i (each representing the number of trials needed to get the i-th success after the $(i-1)$-th success):

$$X = Y_1 + Y_2 + \cdots + Y_r$$

Each Y_i is Geom(p), so $E[Y_i] = 1/p$ and $\text{Var}[Y_i] = (1-p)/p^2$. By linearity of expectation and independence:

$$E[X] = r \cdot E[Y_1] = \frac{r}{p}$$

$$\text{Var}[X] = r \cdot \text{Var}[Y_1] = \frac{r(1-p)}{p^2}$$

6.2 Geometric Random Variables

A **geometric random variable** is a special case of the negative binomial distribution where $r = 1$. That is, it counts the number of trials needed to get the first success.

The probability mass function is:

$$f(x; p) = \binom{x-1}{1-1} p^1 (1-p)^{x-1} = p(1-p)^{x-1}$$

for $x = 1, 2, 3, \ldots$

The mean and variance are:

$$E[X] = \frac{1}{p}$$
$$\text{Var}[X] = \frac{1-p}{p^2}$$

Derivation:

For $X \sim \text{Geom}(p)$,

$$E[X] = \sum_{x=1}^{\infty} xp(1-p)^{x-1}$$

This is a standard result and can be shown using the formula for the expectation of a geometric series:

$$E[X] = \frac{1}{p}$$

Similarly, for the variance:

$$\text{Var}[X] = \frac{1-p}{p^2}$$

6.3 Poisson Random Variables

A **Poisson random variable** is defined by a single parameter λ that represents the rate of occurrences.

Definition (Poisson Random Variable)**.** A random variable X is said to have a **Poisson distribution** with parameter $\lambda > 0$, denoted $X \sim \text{Poisson}(\lambda)$, if it represents the number of occurrences of an event in a fixed interval of time or space, where occurrences happen independently and at a constant average rate λ.

The probability mass function is:

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

for $k = 0, 1, 2, \ldots$

Derivation from Binomial:

Suppose we divide a unit interval into n subintervals, each of length $1/n$. In each subinterval, the probability of an occurrence is λ/n, so the total expected number of occurrences in the unit interval is λ.

Let $Y \sim \text{Bin}(n, \lambda/n)$ be the number of occurrences in n subintervals.

The probability mass function is:

$$P(Y = x) = \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

As $n \to \infty$, the binomial distribution converges to the Poisson distribution:

$$P(X = x) = \lim_{n\to\infty} \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} = \frac{\lambda^x e^{-\lambda}}{x!}$$

This shows that the Poisson distribution can be viewed as the limiting case of the binomial distribution when the number of trials is large and the probability of success is small, but the expected number of successes remains fixed.

Conditions for a Poisson process:

- Numbers of occurrences in nonoverlapping subintervals are independent.
- Probability of exactly one occurrence in a sufficiently short subinterval of length h is approximately λh.
- Probability of two or more occurrences in a sufficiently short subinterval is essentially zero.

The mean and variance of a Poisson random variable $X \sim \text{Poisson}(\lambda)$ are:

$$E[X] = \lambda$$

$$\text{Var}[X] = \lambda$$

Derivation:

The mean is

$$E[X] = \sum_{k=0}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!}$$

This can be shown to equal λ (using properties of the exponential and the fact that $\sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} = \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{\lambda}$).

The variance is

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

It can be shown that $E[X^2] = \lambda^2 + \lambda$, so

$$\text{Var}[X] = (\lambda^2 + \lambda) - \lambda^2 = \lambda$$

7 Even More Random Variables

Date: Sep 22, 2025

7.1 Exponential Random Variables

An **exponential random variable** W with rate parameter $\lambda > 0$ is a continuous random variable with cumulative distribution function (CDF)

$$F(w) = \begin{cases} 0 & \text{if } w < 0 \\ 1 - e^{-\lambda w} & \text{if } w \geq 0 \end{cases}$$

and probability density function (PDF)

$$f(w) = \begin{cases} \lambda e^{-\lambda w} & \text{if } w \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

The expected value and variance of W are given by

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

The exponential distribution is commonly used to model the time between independent events that occur at a constant average rate. It is the time until the first event in a Poisson process.

7.2 Gamma Random Variables

A **gamma random variable** X with shape parameter $\alpha > 0$ and scale parameter $\theta > 0$ has pdf

$$f(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta}, \quad 0 \leq x < \infty$$

The gamma distribution generalizes the exponential distribution. If $\alpha = 1$, the gamma distribution reduces to the exponential distribution. The gamma distribution models the waiting time until the α-th event in a Poisson process with rate $\lambda = 1/\theta$.

The mean and variance are:

$$\mathbb{E}[X] = \frac{t}{\lambda} = \alpha\theta, \qquad \mathrm{Var}[X] = \frac{t}{\lambda^2} = \alpha\theta^2$$

where $\Gamma(t)$ is the **gamma function**:

$$\Gamma(t) = \int_0^\infty y^{t-1} e^{-y} dy, \quad t > 0$$

Derivation of the gamma function:

We use integration by parts:

$$\int u\, dv = uv - \int v\, du$$

Let $u = y^{t-1}$, $dv = e^{-y}dy$, so $du = (t-1)y^{t-2}dy$, $v = -e^{-y}$.

Applying integration by parts:

$$\Gamma(t) = \int_0^\infty y^{t-1} e^{-y} dy = \left[-y^{t-1}e^{-y}\right]_0^\infty + \int_0^\infty (t-1) y^{t-2} e^{-y} dy$$

The boundary term vanishes for $t > 1$, so:

$$\Gamma(t) = (t-1)\int_0^\infty y^{t-2}e^{-y}dy = (t-1)\Gamma(t-1)$$

This gives the recurrence relation:

$$\Gamma(t) = (t-1)\Gamma(t-1)$$

For $t = 1$:

$$\Gamma(1) = \int_0^\infty e^{-y}dy = 1$$

Gamma Random Variables When α is an Integer If α is a positive integer, the gamma distribution describes the waiting time until the α-th event in a Poisson process with rate $\lambda = 1/\theta$.

The cumulative distribution function (cdf) is:

$$F(x) = \Pr(X \leq x) = 1 - \Pr(X > x)$$

This is equivalent to the probability that at least α events have occurred by time x:

$$F(x) = 1 - \Pr(\text{fewer than } \alpha \text{ occurrences in } [0, x])$$

For a Poisson process, the probability of k events in $[0, x]$ is:

$$\Pr(k \text{ events in } [0, x]) = \frac{(\lambda x)^k e^{-\lambda x}}{k!}$$

So,

$$F(x) = 1 - \sum_{k=0}^{\alpha-1} \frac{(\lambda x)^k e^{-\lambda x}}{k!}$$

Or, equivalently, using $\theta = 1/\lambda$:

$$F(x) = 1 - \sum_{k=0}^{\alpha-1} \frac{(x/\theta)^k e^{-x/\theta}}{k!}$$

This expresses the cdf of the gamma distribution when α is an integer in terms of the Poisson distribution.

8 Bivariate Distributions

Date: Oct 1, 2025

8.1 Linearity of Expectation:

For random variables A and B, $\mathbb{E}[A+B] = \mathbb{E}[A] + \mathbb{E}[B]$.

$$\begin{aligned}
\mathbb{E}[A+B] &= \sum_a \sum_b (a+b) f(a,b) \\
&= \sum_a \sum_b a f(a,b) + \sum_a \sum_b b f(a,b) \\
&= \left(\sum_a \sum_b a f(a,b) \right) + \left(\sum_a \sum_b b f(a,b) \right) \\
&= \sum_a a \left(\sum_b f(a,b) \right) + \sum_b b \left(\sum_a f(a,b) \right) \\
&= \sum_a a f_A(a) + \sum_b b f_B(b) \\
&= \mathbb{E}[A] + \mathbb{E}[B]
\end{aligned}$$

where $f(a,b)$ is the joint probability mass function, and $f_A(a)$ and $f_B(b)$ are the marginal distributions of A and B, respectively.

8.2 Covariance

Definition. For two random variables X and Y, the covariance is defined as:

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

For discrete random variables, this can be written as:

$$\mathrm{Cov}[X,Y] = \sum_x \sum_y (x - \mathbb{E}[X])(y - \mathbb{E}[Y]) \Pr(X = x, Y = y)$$

- $\mathrm{Cov}(X,Y) > 0$: When X increases, Y tends to increase (positive association).
- $\mathrm{Cov}(X,Y) < 0$: When X increases, Y tends to decrease (negative association).
- $\mathrm{Cov}(X,Y) = 0$: No (linear) association between X and Y.

Covariance measures the linear association between two random variables.

8.3 Variance of a Sum

For any two random variables X and Y,

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + 2\,\mathrm{Cov}(X,Y) + \mathrm{Var}(Y)$$

Derivation

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

If X and Y are independent, then $\mathrm{Cov}(X,Y)=0$, so

$$\mathrm{Var}(X+Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)$$

Intuition If two random variables negatively covary, they will tend to "cancel each other out" and reduce the variance of the sum; if they positively covary, they will tend to magnify the variance of the sum.

Additional Useful Facts about Covariance

- For constants a, b, c, d,
$$\mathrm{Cov}(aX+b,\ cY+d)=ac\,\mathrm{Cov}(X,Y)$$
- Covariance is linear in each argument:
$$\mathrm{Cov}(X,\ Y+Z)=\mathrm{Cov}(X,Y)+\mathrm{Cov}(X,Z)$$
- If X and Y are independent, then $\mathrm{Cov}(X,Y)=0$.
- The converse is not true: $\mathrm{Cov}(X,Y)=0$ does **not** imply X and Y are independent.

Independence Implies Zero Covariance If X and Y are independent, then their joint pmf factors $f(x,y)=f_X(x)f_Y(y)$. Hence,

$$\begin{aligned}
\mathrm{Cov}[X,Y] &= \sum_x\sum_y (x-\mu_X)(y-\mu_Y)f(x,y) \\
&= \sum_x\sum_y (x-\mu_X)(y-\mu_Y)f_X(x)f_Y(y) \\
&= \sum_x (x-\mu_X)f_X(x)\sum_y (y-\mu_Y)f_Y(y) \\
&= \sum_x (x-\mu_X)f_X(x)\cdot 0 \\
&= 0
\end{aligned}$$

where $\mu_X=\mathbb{E}[X]$ and $\mu_Y=\mathbb{E}[Y]$. Thus, independence implies zero covariance.

9 Bivariate Distributions Continued

Date: Oct 6, 2025

9.1 Correlation Coefficient

- Covariance varies with the measurement scale of the variables.
- The correlation coefficient ρ is scale-invariant.
- ρ ranges from -1 (perfect negative linear association) to 1 (perfect positive linear association).
- $\rho = 0$ indicates no linear relationship.

$$\rho = \frac{\mathrm{Cov}(X, Y)}{\mathrm{SD}(X)\,\mathrm{SD}(Y)}$$

9.2 Conditional Distributions

- For two discrete random variables X and Y, the conditional probability mass function of X given Y is:
$$g(x|y) = \frac{f(x, y)}{f_Y(y)}, \quad \text{provided that } f_Y(y) > 0$$
- $f(x, y)$ is the joint probability mass function of X and Y.
- $f_Y(y)$ is the marginal probability mass function of Y.
- $g(x|y)$ is not defined for values of y where $f_Y(y) = 0$.

If X and Y are independent, then $g(x|y) = f_X(x)$ for all y such that $f_Y(y) > 0$.

9.3 Continuous Bivariate Distributions

- A joint probability density function (pdf) $f(x, y)$ for continuous random variables X and Y must satisfy:
  1. $f(x, y) \geq 0$, and $f(x, y) = 0$ when $(x, y) \notin S$, where S is the sample space.
  2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$
  3. For any region A, $\Pr[(X, Y) \in A] = \iint_A f(x, y)\, dx\, dy$
- The marginal pdfs are obtained by integrating out the other variable:
$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy, \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$
- Independence: X and Y are independent if and only if $f(x, y) = f_X(x) f_Y(y)$ for all (x, y).

10 Functions of Random Variables

Date: Oct 8, 2025

10.1 Expectations

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

10.2 The Bivariate Normal Distribution

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left[-\frac{q(x, y)}{2}\right]$$

where

$$q(x, y) = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]$$

- μ_X, μ_Y are the means of X and Y
- σ_X, σ_Y are the standard deviations
- ρ is the correlation coefficient between X and Y

10.3 Conditional Expectation of the Bivariate Normal

If X and Y are bivariate normal:

- The marginal distributions of X and Y are also normal.
- $X + Y$ and X are jointly normal (as are $X + Y$ and Y).
- The conditional distribution $Y|X$ is also normal with:

$$\begin{aligned} E[Y|x] &= \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X) \\ &= \mu_Y - \rho\frac{\sigma_Y}{\sigma_X}\mu_X + \rho\frac{\sigma_Y}{\sigma_X}x \end{aligned}$$

  Since $\rho = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$, we can substitute to get:

$$E[Y|x] = \mu_Y - \frac{\sigma_{XY}}{\sigma_X^2}\mu_X + \frac{\sigma_{XY}}{\sigma_X^2}x$$

  where σ_{XY} is the covariance between X and Y.
  And $Var(Y|x) = \sigma_Y^2(1 - \rho^2)$.

- X and Y being Normal does not imply the joint distribution is Bivariate Normal.

10.4 Functions of a Random Variable

10.4.1 Distribution Function Technique

If $Y = u(X)$, then

$$G(y) = \Pr(Y \leq y) = \Pr(u(X) \leq y)$$

and the density is

$$g(y) = G'(y)$$

Suppose $X \sim N(0,1)$ and $Y = \exp(X)$ (log-normal distribution). Let $\Phi(x)$ be the cdf and $\phi(x)$ the pdf of the standard normal. Then $\Phi'(x) = \phi(x)$.

$$G(y) = \Pr(\exp(X) \leq y) = \Pr(X \leq \ln y) = \Phi(\ln y)$$

$$g(y) = G'(y) = \Phi'(\ln y) \cdot \frac{1}{y} = \phi(\ln y) \cdot \frac{1}{y}$$

for $y > 0$; otherwise, $g(y) = 0$.

10.4.2 Change-of-Variable Technique

If $Y = u(X)$, define $v()$ such that $X = v(Y)$.

Suppose $c_1 < x < c_2$ and $d_1 = u(c_1) < y < d_2 = u(c_2)$ ($v()$ is increasing),

$$G(y) = \int_{c_1}^{v(y)} f(x)dx, \qquad d_1 < y < d_2$$

$$G'(y) = g(y) = f(v(y)) \cdot v'(y), \qquad d_1 < y < d_2$$

Log-normal example:
$u(x) = \exp(x)$, $v(y) = \ln(y)$, $v'(y) = 1/y$

$$g(y) = \phi(\ln y) \cdot \frac{1}{y} \qquad \text{when } y > 0$$

Suppose $c_1 < x < c_2$ and $d_1 = u(c_1) > y > d_2 = u(c_2)$ ($v()$ is decreasing),

$$G(y) = \int_{v(y)}^{c_2} f(x)dx, \qquad d_2 < y < d_1$$

$$G'(y) = g(y) = f(v(y)) \cdot (-v'(y)), \qquad d_2 < y < d_1$$

Note: The negative sign appears because $v'(y) < 0$ when $v()$ is decreasing.
In general, we have:

$$g(y) = f(v(y)) \cdot |v'(y)|, \qquad y \in S_Y$$

11 More Functions of Random Variables

Date: Oct 15, 2025

11.1 Cauchy Distribution

Definition (Cauchy Distribution)**.** Let $W \sim U(-\pi/2, \pi/2)$, i.e., W is uniformly distributed between $-\pi/2$ and $\pi/2$. Define $X = \tan W$. Then X has the standard Cauchy distribution.

The probability density function (pdf) of the standard Cauchy distribution is:

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}$$

Properties:

- The Cauchy distribution is symmetric about $x = 0$.
- It has much heavier tails than the normal distribution.
- Neither the mean nor the variance exists.
- The median and mode are both 0.

Derivation of the PDF:
Let $W \sim U(-\pi/2, \pi/2)$, so its pdf is

$$f_W(w) = \frac{1}{\pi}, \quad w \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$$

Define $X = \tan W$. To find the pdf of X, use the change of variables formula:

$$g_X(x) = f_W(v(x)) \left|v'(x)\right|$$

where $v(x) = \arctan x$ and $v'(x) = \frac{1}{1+x^2}$.
Substituting, we get:

$$g_X(x) = f_W(\arctan x) \cdot \frac{1}{1+x^2} = \frac{1}{\pi} \cdot \frac{1}{1+x^2} = \frac{1}{\pi(1+x^2)}$$

Notes:

1. The graph of the pdf is called the witch of Agnesi.
2. X is the ratio of two independent standard normals.
3. X is a t distribution with one degree of freedom.

11.2 Change-of-Variable Technique (2 Variables)

Suppose $Y_1 = u_1(X_1, X_2)$ and $Y_2 = u_2(X_1, X_2)$, and we can invert these to get $X_1 = v_1(Y_1, Y_2)$ and $X_2 = v_2(Y_1, Y_2)$. Then, the joint pdf of (Y_1, Y_2) is given by:

$$g(y_1, y_2) = |J| \, f\big(v_1(y_1, y_2), v_2(y_1, y_2)\big)$$

where f is the joint pdf of (X_1, X_2), and J is the Jacobian determinant:

$$J = \begin{vmatrix} \frac{\partial v_1}{\partial y_1} & \frac{\partial v_1}{\partial y_2} \\ \frac{\partial v_2}{\partial y_1} & \frac{\partial v_2}{\partial y_2} \end{vmatrix}$$

12 The Central Limit Theorem

Date: Oct 20, 2025

12.1 Sample Mean

Suppose we have n random variables X_i that are independent and identically distributed (iid) with mean μ and standard deviation σ.

Define the sample mean as:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

The expected value and variance of the sample mean are:

$$E[\overline{X}] = \mu \qquad Var[\overline{X}] = \frac{\sigma^2}{n}$$

As $n \to \infty$, $Var[\overline{X}] \to 0$.

Law of Large Numbers (LLN): As n increases, the sample mean $\overline{X}$ converges to the population mean μ.

Central Limit Theorem (CLT): As n increases, the distribution of $\overline{X}$ approaches a normal distribution, regardless of the underlying distribution of X_i.

12.2 Markov's Inequality

For a **nonnegative** random variable X and any $a > 0$,

$$\Pr(X \geq a) \leq \frac{E[X]}{a}$$

Proof:

Let $f(x)$ be the probability mass function of X.

$$\begin{aligned}
E[X] &= \sum_{x} x f(x) \\
&= \sum_{x \geq a} x f(x) + \sum_{x < a} x f(x) \\
&\geq \sum_{x \geq a} x f(x) \\
&\geq \sum_{x \geq a} a f(x) \\
&= a \sum_{x \geq a} f(x) \\
&= a \Pr(X \geq a)
\end{aligned}$$

Therefore,

$$\Pr(X \geq a) \leq \frac{E[X]}{a}$$

12.3 Chebyshev's Inequality

For any random variable X and any $r > 0$,

$$\Pr(|X - E[X]| \geq r) \leq \frac{Var[X]}{r^2}$$

Proof:
Define $A = \{x : |x - E[X]| \geq r\}$.

$$\begin{aligned}
Var[X] &= \sum_x (x - E[X])^2 f(x) \\
&= \sum_{x \in A} (x - E[X])^2 f(x) + \sum_{x \notin A} (x - E[X])^2 f(x) \\
&\geq \sum_{x \in A} (x - E[X])^2 f(x) \\
&\geq \sum_{x \in A} r^2 f(x) \\
&= r^2 \sum_{x \in A} f(x) = r^2 \Pr(A) = r^2 \Pr(|X - E[X]| \geq r)
\end{aligned}$$

Therefore,

$$\frac{Var[X]}{r^2} \geq \Pr(|X - E[X]| \geq r)$$

13 Proving the Central Limit Theorem

Date: Oct 22, 2025

- The Central Limit Theorem (CLT) states that for a random sample $X_1, X_2, \ldots, X_n$ from a distribution with finite mean μ and variance σ^2, the standardized sample mean

  $$W = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

  converges in distribution to the standard normal:

  $$W \xrightarrow{d} N(0, 1)$$

  as $n \to \infty$.

- This means that as the sample size n increases, the distribution of the sample mean $\overline{X}$ approaches a normal distribution with mean μ and variance $\frac{\sigma^2}{n}$:

  $$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

- The CLT is fundamental in statistics because it justifies the use of normal approximations for inference, even when the underlying distribution is not normal.

- Key conditions: finite mean and variance, and sufficiently large n (typically $n \geq 30$).

13.1 The mgf Method

- If the sequence of mgfs $M_n(t)$ of random variables X_n converges to an mgf $M(t)$ for t in an open interval around 0, then the distributions of X_n converge to the distribution with mgf $M(t)$.

- In the context of the Central Limit Theorem, we show that the mgf of the standardized sum converges to the mgf of the standard normal distribution.

- This establishes that the limiting distribution is normal.

13.2 Proof

Big Picture

- Let $W = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$.

- We want to show that $W \xrightarrow{d} N(0, 1)$.

- The mgf of the standard normal is $\exp(t^2/2)$.

- Steps:

  1. Derive the mgf of W.
  2. Take the limit of that mgf as $n \to \infty$.

3. Show that's the mgf of the standard normal.

Let $W = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\overline{X}-\mu)}{\sigma} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}$.
The moment generating function (mgf) of W is:

$$M_W(t) = E[\exp(tW)] = E\left[\exp\left(\frac{t}{\sqrt{n}\sigma}\left(\sum_{i=1}^{n} X_i - n\mu\right)\right)\right]$$

$$= E\left[\prod_{i=1}^{n} \exp\left(\frac{t}{\sqrt{n}\sigma}(X_i - \mu)\right)\right]$$

Since X_i are independent,

$$= \prod_{i=1}^{n} E\left[\exp\left(\frac{t}{\sqrt{n}\sigma}(X_i - \mu)\right)\right]$$

Let $Y_i = \frac{X_i - \mu}{\sigma}$ (standardized), so

$$= \left(E\left[\exp\left(\frac{t}{\sqrt{n}} Y_i\right)\right]\right)^n$$

Let $m(t)$ be the mgf of Y_i:

$$m(t) = E[\exp(tY_i)]$$

So,

$$M_W(t) = \left(m\left(\frac{t}{\sqrt{n}}\right)\right)^n$$

Now, expand $m(t)$ using Taylor's theorem around $t = 0$:

$$m(t) = 1 + m'(0)t + \frac{m''(t_1)t^2}{2}$$

where t_1 is between 0 and t.
Since $E[Y_i] = 0$ and $E[Y_i^2] = 1$, we have $m(0) = 1$, $m'(0) = 0$, $m''(0) = 1$.
So,

$$m(t) = 1 + \frac{m''(t_1)t^2}{2}$$

$$= 1 + \frac{t^2}{2} + \frac{(m''(t_1) - 1)t^2}{2}$$

Plug in $t/\sqrt{n}$:

$$m\left(\frac{t}{\sqrt{n}}\right) = 1 + \frac{t^2}{2n} + \frac{(m''(t_1) - 1)t^2}{2n}$$

So,

$$M_W(t) = \left[1 + \frac{t^2}{2n} + \frac{(m''(t_1) - 1)t^2}{2n}\right]^n$$

As $n \to \infty$, $t_1 \to 0$ and $m''(t_1) \to m''(0) = 1$:

$$\lim_{n\to\infty} M_W(t) = \lim_{n\to\infty} \left[1 + \frac{t^2}{2n}\right]^n$$

Recall $\lim_{n\to\infty}(1+x/n)^n = \exp(x)$:

$$\lim_{n\to\infty} M_W(t) = \exp\left(\frac{t^2}{2}\right)$$

This is the mgf of the standard normal distribution. Therefore, by the mgf method, $W \xrightarrow{d} N(0,1)$ as $n \to \infty$.