Chapter 4 sections

- 4.1 Expectation
- 4.2 Properties of Expectations
- 4.3 Variance
- 4.4 Moments
- 4.5 The Mean and the Median
- 4.6 Covariance and Correlation
- 4.7 Conditional Expectation
- **SKIP:** 4.8 Utility
Summarizing distributions

- The distribution of $X$ contains everything there is to know about the probabilistic properties of $X$.
- However, sometimes we want to summarize the distribution of $X$ in one or a few numbers
  - e.g. to more easily compare two or more distributions.
- Examples of descriptive quantities:
  - Mean ( = Expectation )
    - Center of mass - weighted average
  - Median, Moments
  - Variance, Interquartile Range (IQR), Covariance, Correlation
Definition of Expectation $\mu = E(X)$

**Def: Mean aka. Expected value**

Let $X$ be a random variable with p.d.f $f(x)$. The *mean*, or *expected value* of $X$, denoted $E(X)$, is defined as follows:

- **$X$ discrete:**
  $E(X) = \sum_{All \ x} xf(x)$
  assuming the sum exists.

- **$X$ continuous:**
  $E(X) = \int_{-\infty}^{\infty} xf(x) \, dx$
  assuming the integral exists.

If the sum or integral does not exist we say that the expected value does not exist.

The mean is often denoted with $\mu$. 

Recall the distribution of $Y =$ the number of heads in 3 tosses (coin toss example from Lecture 4)

<table>
<thead>
<tr>
<th>$y$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_Y(y)$</td>
<td>$\frac{1}{8}$</td>
<td>$\frac{3}{8}$</td>
<td>$\frac{3}{8}$</td>
<td>$\frac{1}{8}$</td>
</tr>
</tbody>
</table>

then

$$E(Y) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{12}{8} = \frac{3}{2} = 1.5$$
Examples

- Recall the distribution of $Y$ = the number of heads in 3 tosses (coin toss example from Lecture 4)

<table>
<thead>
<tr>
<th>$y$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_Y(y)$</td>
<td>$\frac{1}{8}$</td>
<td>$\frac{3}{8}$</td>
<td>$\frac{3}{8}$</td>
<td>$\frac{1}{8}$</td>
</tr>
</tbody>
</table>

then

$$E(Y) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{12}{8} = \frac{3}{2} = 1.5$$

- Find $E(X)$ where $X \sim \text{Binom}(n, p)$. The pf of $X$ is

$$f(x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad \text{for} \ x = 0, 1, \ldots, n$$

- Find $E(X)$ where $X \sim \text{Uniform}(a, b)$. The pdf of $X$ is

$$f(x) = \frac{1}{b - a} \quad \text{for} \ a \leq x \leq b$$
**Theorem 4.1.1**

Let $X$ be a random variable with probability density function $f(x)$ and $g(x)$ be a real-valued function. Then

- $X$ discrete:
  \[
  E(g(X)) = \sum_{\text{All } x} g(x)f(x)
  \]

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

Example: Find $E(X^2)$ where $X \sim \text{Uniform}(a, b)$. 

Expectation of $g(X, Y)$

**Theorem 4.1.2**

Let $X$ and $Y$ be random variables with joint probability density function $f(x, y)$ and let $g(x, y)$ be a real-valued function. Then

- $X$ and $Y$ discrete:
  
  $$E(g(X, Y)) = \sum_{\text{All } x, y} g(x, y)f(x, y)$$

- $X$ and $Y$ continuous:
  
  $$E(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f(x, y) \, dx \, dy$$

**Example:** Find $E\left(\frac{X+Y}{2}\right)$ where $X$ and $Y$ are independent and $X \sim \text{Uniform}(a, b)$ and $Y \sim \text{Uniform}(c, d)$. 
Properties of Expectation

Theorems 4.2.1, 4.2.4 and 4.2.6:

- $E(aX + b) = aE(X) + b$ for constants $a$ and $b$.
- Let $X_1, \ldots, X_n$ be $n$ random variables, all with finite expectations $E(X_i)$, then
  
  $$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$$

- Corollary: $E(a_1 X_1 + \cdots + a_n X_n + b) = a_1 E(X_1) + \cdots + a_n E(X_n) + b$ for constants $b, a_1, \ldots, a_n$.

- Let $X_1, \ldots, X_n$ be $n$ independent random variables, all with finite expectations $E(X_i)$, then
  
  $$E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} E(X_i)$$

CAREFUL !!! In general $E(g(X)) \neq g(E(X))$.
For example: $E(X^2) \neq [E(X)]^2$
Examples

If $X_1, X_2, \ldots, X_n$ are i.i.d. $\text{Bernoulli}(p)$ random variables then

$Y = \sum_{i=1}^{n} X_i \sim \text{Binomial}(n, p)$.

$$E(X_i) = 0 \times (1 - p) + 1 \times p = p \quad \text{for } i = 1, \ldots, n$$

$$\Rightarrow E(Y) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} p = np$$

Note: i.i.d. stands for independent and identically distributed
Definition of Variance \( \sigma^2 = \text{Var}(X) \)

Def: Variance

Let \( X \) be a random variable (discrete or continuous) with a finite mean \( \mu = E(X) \). The **Variance of \( X \)** is defined as

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

The **standard deviation of \( X \) is defined as** \( \sqrt{\text{Var}(X)} \)

We often use \( \sigma^2 \) for variance and \( \sigma \) for standard deviation.

**Theorem 4.3.1 – Another way of calculating variance**

For any random variable \( X \)

\[
\text{Var}(X) = E(X^2) - [E(X)]^2
\]
Examples - calculating the variance

- Recall the distribution of $Y =$ the number of heads in 3 tosses (coin toss example from Lecture 4)

<table>
<thead>
<tr>
<th>$y$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_Y(y)$</td>
<td>$\frac{1}{8}$</td>
<td>$\frac{3}{8}$</td>
<td>$\frac{3}{8}$</td>
<td>$\frac{1}{8}$</td>
</tr>
</tbody>
</table>

We already found that $\mu = E(Y) = 1.5$. Then

$$Var(Y) = (0 - 1.5)^2 \frac{1}{8} + (1 - 1.5)^2 \frac{3}{8} + (2 - 1.5)^2 \frac{3}{8} + (3 - 1.5)^2 \frac{1}{8}$$

$$= 0.75$$

- Find $Var(X)$ where $X \sim Uniform(a, b)$
Properties of the Variance

Theorems 4.3.2, 4.3.3, 4.3.4 and 4.3.5

- \( \text{Var}(X) \geq 0 \) for any random variable \( X \).
- \( \text{Var}(X) = 0 \) if and only if \( X \) is a constant, i.e. \( P(X = c) = 1 \) for some constant \( c \).
- \( \text{Var}(aX + b) = a^2 \text{Var}(X) \)
- If \( X_1, \ldots, X_n \) are independent we have
  \[
  \text{Var} \left( \sum_{i=1}^{n} X_i \right) = \sum_{i=1}^{n} \text{Var}(X_i)
  \]
Examples

If $X_1, X_2, \ldots, X_n$ are i.i.d. Bernoulli($p$) random variables then $Y = \sum_{i=1}^{n} X_i \sim \text{Binomial}(n, p)$.

$$E(X_i) = p \quad \text{for } i = 1, \ldots, n$$

$$E(X_i^2) = 0^2 \times (1 - p) + 1^2 \times p = p \quad \text{for } i = 1, \ldots, n$$

$$\Rightarrow \text{Var}(X_i) = E(X_i^2) - [E(X_i)]^2 = p - p^2 = p(1 - p)$$

$$\Rightarrow \text{Var}(Y) = \text{Var} \left( \sum_{i=1}^{n} X_i \right) = \sum_{i=1}^{n} \text{Var}(X_i) = \sum_{i=1}^{n} p(1 - p)$$

$$= np(1 - p)$$
Measures of location and scales

The mean is a measure of location, the variance is a measure of scale.

Different mean, same variance

Same means, different variance
Moments and Central moments

Def: Moments

Let $X$ be a random variable and $k$ be a positive integer.

- The expectation $E(X^k)$ is called the $k^{th}$ moment of $X$.
- Let $E(X) = \mu$. The expectation $E((X - \mu)^k)$ is called the $k^{th}$ central moment of $X$.

- The first moment is the mean: $\mu = E(X^1)$.
- The first central moment is zero: $E(X - \mu) = E(X) - E(X) = 0$.
- The second central moment is the variance: $\sigma^2 = E((X - \mu)^2)$. 

Moments and Central moments

- **Symmetric distribution**: If the p.d.f \( f(x) \) is symmetric with respect to a point \( x_0 \), i.e. \( f(x_0 + \delta) = f(x_0 - \delta) \) for all \( \delta \)
- If the mean of a symmetric distribution exists, then it is the point of symmetry.
- If the distribution of \( X \) is symmetric w.r.t. its mean \( \mu \) then \( E ((X - \mu)^k) = 0 \) for \( k \) odd (if the central moment exists)
- **Skewness**: \( E ((X - \mu)^3) / \sigma^3 \)
Moment generating function

Def: Moment Generating Function
Let $X$ be a random variable. The function

$$
\psi(t) = E\left( e^{tX} \right) \quad t \in \mathbb{R}
$$

is called the moment generating function (m.g.f.) of $X$

Theorem 4.4.2
Let $X$ be a random variables whose m.g.f. $\psi(t)$ is finite for $t$ in an open interval around zero. Then the $n$th moment of $X$ is finite, for $n = 1, 2, \ldots$, and

$$
E(X^n) = \left. \frac{d^n}{dt^n} \psi(t) \right|_{t=0}
$$
Let $X \sim \text{Gamma}(n, \beta)$. Then $X$ has the pdf

$$f(x) = \frac{1}{(n-1)! \beta^n} x^{n-1} e^{-x/\beta} \quad \text{for } x > 0$$

Find the m.g.f. of $X$ and use it to find the mean and the variance of $X$. 
Properties of m.g.f.

Theorems 4.4.3 and 4.4.4:

- $\psi_{aX+b}(t) = e^{bt} \psi_X(at)$
- Let $Y = \sum_{i=1}^{n} X_i$ where $X_1, \ldots, X_n$ are independent random variables with m.g.f. $\psi_i(t)$ for $i = 1, \ldots, n$ Then
  \[ \psi_Y(t) = \prod_{i=1}^{n} \psi_i(t) \]

Theorem 4.4.5: Uniqueness of the m.g.f.

Let $X$ and $Y$ be two random variables with m.g.f.'s $\psi_X(t)$ and $\psi_Y(t)$.
If the m.g.f.'s are finite and $\psi_X(t) = \psi_Y(t)$ for all values of $t$ in an open interval around zero, then $X$ and $Y$ have the same distribution.
Example

- Let $X \sim N(\mu, \sigma^2)$. $X$ has the pdf

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

and the m.g.f. for the normal distribution is

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

Homework (not to turn in): Show that $\psi(t)$ is the m.g.f. of $X$.

- Let $X_1, \ldots, X_2$ be independent Gaussian random variables with means $\mu_i$ and variances $\sigma_i^2$. What is the distribution of $Y = \sum_{i=1}^{n} X_i$?