Lecture 11:

- Variance and Limit Theorems
- Continuous Distributions
  - Basic Definitions
  - Importance of the CDF
  - Uniform Continuous Distribution
Review: Variance and Standard Deviation

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

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

\[ \sigma_X = \text{def } \sqrt{\text{Var}(X)} \]
Variance of the Special Distributions

\[ X \sim \text{Bernoulli}(p) \]

\[
E(X) = p \\
Var(X) = E(X^2) - E(X)^2 \\
= E(X) - p^2 \\
= p \cdot (1 - p)
\]

\[ Y \sim \text{Binomial}(N, p) \]

\[
Y = X_1 + X_2 + \cdots + X_N \\
E(Y) = N \cdot p \\
Var(Y) = N \cdot p \cdot (1 - p)
\]

\[ Z \sim \text{Geometric}(p) \]

\[
E(Z) = \frac{1}{p} \\
Var(Z) = \frac{1 - p}{p^2}
\]

\[ W \sim \text{Pascal}(m, p) \]

\[
W = Z_1 + Z_2 + \cdots + Z_m \\
E(W) = \frac{m}{p} \\
Var(W) = \frac{m(1 - p)}{p^2}
\]
Variance and StdDev: Important Properties

Useful formula for the Variance and Standard Deviation, showing that variance and the standard deviation are NOT linear functions:

**Theorem:** \( Var(aX + b) = a^2 \cdot Var(X) \)

**Proof:**

\[
\begin{align*}
Var(aX + b) &= E\left[ (aX + b) - \mu_{aX+b} \right]^2 \\
&= E\left[ (aX + b) - (a\mu_X + b) \right]^2 \\
&= E\left[ a(X - \mu_X)^2 \right] \\
&= a^2 \cdot E\left[ (X - \mu_X)^2 \right] \\
&= a^2 \cdot Var(X)
\end{align*}
\]

**Corollary:**

\[ \sigma_{aX+b} = |a| \cdot \sigma_X \]
Variance and StdDev: Important Properties

However, independence, as usual, makes things simpler:

**Theorem:** (Variance of Sum of **Independent** Random Variables)

Let X and Y be independent random variables, then

\[ \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) \]

**Proof:**

\[
\begin{align*}
\text{Var}(X + Y) &= \mathbb{E}[(X + Y)^2] - (\mathbb{E}(X + Y))^2 \\
&= \mathbb{E}[X^2 + 2XY + Y^2] - (\mathbb{E}(X) + \mathbb{E}(Y))^2 \\
&= \mathbb{E}(X^2) + 2\mathbb{E}(XY) + \mathbb{E}(Y^2) - [\mathbb{E}(X)^2 - 2\mathbb{E}(Y)^2(\mathbb{E}(Y)^2) - \mathbb{E}(Y)^2] \\
&= \mathbb{E}(X^2) - \mathbb{E}(X)^2 + \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 + 2[\mathbb{E}(XY) - \mathbb{E}(Y)^2(\mathbb{E}(Y))]
\end{align*}
\]

This term is called the Covariance of X and Y, \( \text{Cov}(X,Y) \), and measures how much they “vary together”. For **independent RV**, \( \text{Cov}(X,Y) = 0 \). This will be back in a few weeks....
Limit Theorems for Random Variables

Some very useful and important theorems give us bounds for how spread out a distribution can be. These help understand what data is useful and what might be an “outlier.”

Markov Inequality:

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

Proof:

\[
E(X) = \sum_{k \in R_X} k \cdot P(X = k) \\
\geq \sum_{k \in R_X, k \geq a} k \cdot P(X = k) \\
\geq \sum_{k \in R_X, k \geq a} a \cdot P(X = k) \\
\geq a \cdot \sum_{k \in R_X, k \geq a} P(X = k) = a \cdot P(X \geq a)
\]
Limit Theorems for Random Variables

Markov Inequality:

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

Examples

In 2019, the average salary in the US = $56,516.

What is the upper bound on the percentage of millionaires in the US?

\[ P(X \geq 10^6) \leq \frac{56,516}{10^6} = 0.056516 \]

What is lower bound on how much you have to earn to guarantee you are in the top 1%?

\[ P(X \geq 5,651,600) \leq \frac{56,516}{5,651,600} = 0.01 \]

Note: This is a theoretical upper bound!
Limit Theorems for Random Variables

If we know the variance, we can use Markov’s result to get a more precise bound, but involving the absolute deviation from the mean.

Chebyshev’s Inequality:

Proof:

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

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

Used Markov’s Inequality here.
Limit Theorems for Random Variables

Alternate form of Chebyshev’s Inequality:

\[ P\left( |X - \mu| \geq k\sigma \right) \leq \frac{1}{k^2} \]

Proof:

\[
P\left( |X - \mu| \geq k\sigma \right) = P\left( (X - \mu)^2 \geq (k\sigma)^2 \right)
= \frac{Var(X)}{k^2 \sigma^2}
= \frac{Var(X)}{k^2 \cdot Var(X)}
= \frac{1}{k^2}
\]
Limit Theorems for Random Variables

Alternate form of Chebyshev’s Inequality:

\[ P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \]

Example. Suppose the expected US salary is 56,516 and the standard deviation is 20,000. Then the highest percentage of people who earn 2 standard deviations away from the mean, less than $16,516 or greater than $96,516 is \( \frac{1}{4} \).

Thus, if we speak of “how many standard deviations away from the mean” as our measure of “how unusual the outcome is,” we have an absolute mathematical upper bound on this probability:

- 2 standard deviations \( \leq \frac{1}{4} \)
- 3 standard deviations \( \leq \frac{1}{9} \)
- 4 standard deviations \( \leq \frac{1}{16} \) etc.
Discrete vs Continuous Distributions

Recall: A Random Variable $X$ is a function from a sample space $S$ into the reals:

$$X : S \rightarrow \mathcal{R}$$

A random variable is called continuous if $R_x$ is uncountable.

What needs to change when working with continuous as opposed to discrete distributions?

Recall: The probability of a random experiment such as a spinner outputting any particular, exact real number is 0:

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

This result extends to any countable collection of real numbers!

So we can only think about intervals:

$$P(0.5 < X < 0.75) = 0.25$$
Probability Functions: Equiprobable vs Not Equiprobable

When the sample space is uncountable, say with the spinner, it is possible for the probability function to be equiprobable or non-equiprobable.

**Uncountable and Equiprobable:**

*Example:* Spin the spinner and report the real number showing.

\[ S = [0..1) \] Any point is equally likely

**Uncountable and NOT Equiprobable:**

*Example:* Heights of Human Beings:

People are more likely to be close to the average height than at the extremes!
Review: Cumulative Distribution Functions

The Cumulative Distribution Function (CDF) for a random variable $X$ shows what happens when we keep track of the sum of the probability distribution from left to right over its range:

$$F_X(k) = P(X \leq k) = \sum_{a \leq k} P_X(a)$$

Example: $X = “The number of dots showing on a thrown die”$

Probability Distribution Function $P_X$  Cumulative Distribution Function $F_X$
Discrete vs Continuous Distributions: PDF vs PMF

Because of the anomalies having to do with continuous probability, we need to keep the following important points in mind:

(A) We will no longer be able to use a discrete Probability Mass Function, but instead a Probability Density Function (PDF), $f_X(a)$.

(A) The probability function $f_X$ does NOT represent the probability of a point in the domain, since as we know:

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

therefore we can ONLY work with intervals $P(X \leq a)$, $P(X > a)$, $P(a \leq X \leq b)$, etc. and $f_X$ is not as important as the CDF $F_X$.

(B) In calculating $F_X$ and working with intervals, we can not use discrete sums as we did in the discrete case, but will have to use integrals:

$$\sum_{x=a}^{b}$$

(C) The range $R_X$ will be all the reals $(-\infty, \infty)$ and so we don’t specify it each time.
Discrete vs Continuous Distributions

**Discrete Random Variables**

The Probability Mass Function (PMF) of a discrete random variable $X$ is a function from the range of $X$ into $\mathbb{R}$:

$$P_X : R_X \mapsto \mathbb{R}$$

such that

(i) $\forall y \in R_X \ P_X(y) \geq 0.0$

(ii) $\sum_{y \in R_X} P_X(y) = 1.0$

**Continuous Random Variables**

The Probability Density Function (PDF) of a continuous random variable $X$ is a function from $\mathbb{R}$ to $\mathbb{R}$:

$$f_X : \mathbb{R} \mapsto \mathbb{R}$$

such that

(i) $\forall y \ f_X(y) \geq 0.0$

(ii) $\int_{-\infty}^{\infty} f_X(y) \ dy = 1.0$
Continuous Distributions

Let’s clarify these ideas with an example....

Consider the spinner example from way back when:

X = “the real number in [0..1) that the spinner lands on”

The probability density function is:

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

Note that the area is 1.0 and for any \( 0 \leq a \leq 1 \), we have
\( f(a) = 1.0 \), so it is uniform across \([0..1)\). But clearly \( P(X = a) = 0.0 \).
Continuous Distributions

Now recall that the ONLY way to deal with continuous probability is to use intervals and to use area (or extent) for the probability. Hence we will calculate probabilities of intervals using the CDF:

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

\[ F(a) = \int_{0}^{a} 1 \, dx = x \bigg|_{0}^{a} = a \]

\[ F(a) = \begin{cases} 
0 & \text{if } a < 0 \\
a & \text{if } 0 \leq a \leq 1 \\
1 & \text{if } a > 1 
\end{cases} \]

\[ P(X < 0.75) = F(0.75) = 0.75 \]
Continuous Distributions

\[ P(0.5 < X < 0.75) = P(X < 0.75) - P(X < 0.5) \]
\[ = F(0.75) - F(0.5) \]
\[ = 0.75 - 0.5 \]
\[ = 0.25 \]

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

\[ F(a) = \int_{0}^{a} 1 \, dx = x \bigg|_{0}^{a} = a \]

\[ F(a) = \begin{cases} 
0 & \text{if } a < 0 \\
a & \text{if } 0 \leq a \leq 1 \\
1 & \text{if } a > 1
\end{cases} \]
Continuous Distributions

**Bottom Line:** In order to deal with continuous distributions, you have to do integrals....

**Example:** Suppose our PDF looked like this:

\[
f(x) = \begin{cases} 
    \frac{x}{2} & \text{if } 0 \leq x \leq 2 \\
    0 & \text{otherwise}
\end{cases}
\]

To calculate the probability of intervals, we need to determine the CDF, which means doing the following integral:

\[
F(a) = \int_{-\infty}^{a} f(x) \, dx = \int_{0}^{a} \frac{x}{2} \, dx = \left[ \frac{x^2}{4} \right]_{0}^{a} = \frac{a^2}{4}
\]

So for example,

\[
P(0.5 < X < 1.0) = \frac{1^2}{4} - \frac{0.5^2}{4} = \frac{4 - 1}{16} = \frac{3}{16} = 0.1875
\]
Continuous Distributions

Discrete Random Variables

\[ F_X(b) = P(X \leq b) = \sum_{y \leq b} P_X(y) \]

\[ P(a \leq X \leq b) = \sum_{a \leq y \leq b} P_X(y) \]

\[ E(X) = \sum_{y \in R_X} y \cdot P_X(y) \]

Continuous Random Variables

\[ F_X(b) = P(X < b) = \int_{-\infty}^{b} f(x) \, dx \]

\[ P(a < X < b) = \int_{a}^{b} f(x) \, dx \]

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

Same for both Discrete and Continuous Random Variables

\[ \text{Var}(X) = \text{Var}[\mu_X] \]

\[ \sigma_X = \sqrt{\text{Var}(X)} \]

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

All previous theorems about \( E(X) \) and \( \text{Var}(X) \) still hold, it does not matter whether \( X \) is continuous or discrete!
Uniform Distribution

The simplest continuous distribution is similar to the spinner, but with arbitrary endpoints:

If \( X = \) “a random real number uniformly chosen from the interval \([a..b]\)”

then \( X \) is a uniform random variable from \( a \) to \( b \), denoted

\[ X \sim U(a, b) \]

and where

\[
f_X(x) = \begin{cases} 
\frac{1}{b-a} & \text{if } a \leq x \leq b \\
0 & \text{otherwise}
\end{cases}
\]

\[
F_X(x) = \begin{cases} 
0 & \text{if } x < a \\
\frac{x-a}{b-a} & \text{if } a \leq x \leq b \\
1 & \text{if } x > b
\end{cases}
\]
Uniform Distribution

\[ X \sim U(a, b) \]

\[
f_X(x) = \begin{cases} 
\frac{1}{b-a} & \text{if } a \leq x \leq b \\
0 & \text{otherwise} 
\end{cases}
\]

\[
F_X(x) = \begin{cases} 
0 & \text{if } x < a \\
\frac{x-a}{b-a} & \text{if } a \leq x \leq b \\
1 & \text{if } x > b 
\end{cases}
\]

\[
E(X) = \int_a^b x \cdot \frac{1}{b-a} \, dx = \frac{1}{b-a} \cdot \frac{x^2}{2} \bigg|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{b + a}{2}
\]

\[
E(X^2) = \int_a^b x^2 \cdot \frac{1}{b-a} \, dx = \frac{1}{b-a} \cdot \frac{x^3}{3} \bigg|_a^b = \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + ab + b^2}{3}
\]

\[
Var(X) = E(X^2) - E(X)^2 = \frac{a^2 + ab + b^2}{3} - \frac{a^2 - 2ab + b^2}{4} = \frac{a^2 + 2ab + b^2}{12} = \frac{(b-a)^2}{12}
\]