Lecture 8:

- Review: Expected value of a Random Variable
- Variance and Standard Deviation of a Random Variable
- Standardized Random Variables
- Cumulative Distribution Functions
- [If time] Introduction to Probability Distributions
Discrete Random Variables: Expected Value

A fundamental way of characterizing a collection of real numbers is the average or mean value of the collection:

Example: The mean/average of \( \{2, 4, 6, 9\} = 21/4 = 5.7\)

The corresponding notion for a random variable \(X\) is the Expected Value:

\[
E(X) = \sum_{a \in R_X} a \cdot f(a)
\]

Alternate notation: \(\mu_X = E(X)\)

Example: \(X = \text{“the number of dots showing on a single thrown die”}\)

\[
E(X) = \sum_{k \in R_X} \frac{k}{6} = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = \frac{21}{6} = 3.5
\]

\(R_X = \{1, 2, 3, 4, 5, 6\}\)

\(f_X = \{\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\}\)
Expected Value and Fair Games

The expected value of a random variable is a one-number summary of its behavior. It describes what you can expect as the limiting behavior over many trials.

A good example of what this means occurs with games in which you win or lose money on each round or trial. Such a game can be modeled by a random variable: $X = \text{the amount you win (+) or lose (-)}.$ A game is fair if $E(X) = 0.$

**Example:** The rules of “Chuck-a-luck” are as follows. The player makes a bet on any number 1 through 6 and then three dice are thrown. If 1, 2, or 3 dice show the same number as the player’s choice, then he or she wins back the original bet plus 1, 2, or 3 times the original bet.

- So if you bet $1 on 4 and the dice roll 2, 4, and 6, you get back $1+1=2$ for a net win of $1.
- If you bet $1 on 2 and the dice roll 2, 6, and 2, you get back $1+2=3$ for a net win of $2.$
- If you bet $1 on 5 and no 5's show and you lose $1.
Expected Value: Basic Properties

There are two important things to remember about \( E(X) \):

- It may not exist (may be infinite!)
- Linearity of Expectation

Example where \( E(X) \) is infinite (the “St.Petersburg Paradox”): Consider the following game: you roll a coin until heads appears and I give you \( 2^K \) dollars, where \( K \) is the number of rolls. Thus,

\[
X = \text{“the number of rolls until heads appears”} \quad Y = 2^X
\]

and we seek \( E(Y) \).

\[
E(Y) = 2^1 \cdot \frac{1}{2} + 2^2 \cdot \frac{1}{4} + 2^3 \cdot \frac{1}{8} + \ldots + 2^k \cdot \frac{1}{2^k} + \ldots
\]

\[
= 1 + 1 + 1 + \ldots
\]

This will happen for countably infinite RVs when the sequence does not converge.

In order to make this game fair, I would have to charge you an infinite amount of money to play, even though you only win a finite amount of money each round!
Expected Value: Basic Properties

Theorem (Linearity of Expectation)

For any random variable \( X \) and real numbers \( a \) and \( b \),

\[
E(a \cdot X + b) = a \cdot E(X) + b
\]

Proof:

\[
E(aX + b) = \sum_{k \in R_X} (a \cdot k + b) \cdot f_X(k)
\]

\[
= \sum_{k \in R_X} (a \cdot k \cdot f_X(k)) + (b \cdot f_X(k))
\]

\[
= \sum_{k \in R_X} (a \cdot k \cdot f_X(k)) + \sum_{k \in R_X} (b \cdot f_X(k))
\]

\[
= a \cdot \sum_{k \in R_X} (k \cdot f_X(k)) + b \cdot \sum_{k \in R_X} f_X(k)
\]

\[
= a \cdot E(X) + b \cdot 1.0
\]

\[
= a \cdot E(X) + b
\]

(Obvious) Corollary: For any constant \( b \), \( E(b) = b \).

This will make many calculations involving expected value MUCH easier!
Discrete Random Variables: Expected Value

But how useful is expected value in making the following decision?

Consider two games. Which one would you prefer to play?

**Game One:** For $1 per round, you can flip a coin, and I’ll give you $6 (your bet back plus $5) if heads appears, and nothing if tails appears (so you lose your bet). Call this the random variable $X_1 = \text{net payout}$.

\[
E(X_1) = -1 \cdot \frac{1}{2} + 5 \cdot \frac{1}{2} = 2.00
\]

**Game Two:** For $1 per round, you can flip a coin 20 times, and if you get 20 heads, I’ll give you $3,145,728, else nothing. Call this the random variable $X_2$.

\[
E(X_2) = -1 \cdot \frac{2^{20} - 1}{2^{20}} + 3,145,728 \cdot \frac{1}{2^{20}} = 2.00
\]

Oh, and by the way, I only have time for a single round....
Discrete Random Variables: Variance

The question is: **How much does** X **vary from E(X)? How spread out is the probability distribution around the expected value?**
The **Variance** of a random variable, $\text{Var}(X)$, is the expected deviation from $\text{E}(X)$.

But what precisely is “deviation”?

**First (doomed) attempt:** deviation = distance from expected value

\[
\text{deviation} = X - \text{E}(X) \quad \text{Var}(X) = \text{E} \left[ X - \text{E}(X) \right]
\]

**Example:** $X_1$ = “Flip a coin and return the number of heads showing”  
$X_2$ = “Flip a coin and return $100 \times$ the number of heads showing”

\[
\begin{align*}
R_{X_1} &= \{ 0, 1 \} \quad f_{X_1} = \{ \frac{1}{2}, \frac{1}{2} \} \quad \text{E}(X) = 0.5 \\
R_{X_2} &= \{ 0, 100 \} \quad f_{X_2} = \{ \frac{1}{2}, \frac{1}{2} \} \quad \text{E}(X) = 50 \\
R_{X_1-0.5} &= \{ -0.5, 0.5 \} \quad f_{X_1-0.5} = \{ \frac{1}{2}, \frac{1}{2} \} \quad \text{E}(X_1 - 0.5) = \text{E}(X_1) - 0.5 = 0.0 \\
R_{X_2-50} &= \{ -50, 50 \} \quad f_{X_2-50} = \{ \frac{1}{2}, \frac{1}{2} \} \quad \text{E}(X_2 - 50) = \text{E}(X_2) - 50 = 0.0
\end{align*}
\]

Note that $\text{E}(X)$ is a constant:

\[
\text{E}(X - \text{E}(X)) = \text{E}(X) - \text{E}(X) = 0.0
\]
Discrete Random Variables: Variance

Example: \( X_1 = \) “Flip a coin and return the number of heads showing” 
\( X_2 = \) “Flip a coin and return 100 \(*\) the number of heads showing”

Second attempt: deviation = absolute value of distance from \( E(X) \)

\[
\text{deviation} = | X - E(X) | \\
\text{Var}(X) = E[ | X - E(X) | ]
\]

\[
R_{X_1} = \{ 0, 1 \} \\
f_{X_1} = \{ \frac{1}{2}, \frac{1}{2} \} \\
E(X) = 0.5
\]

\[
R_{X_2} = \{ 0, 100 \} \\
f_{X_2} = \{ \frac{1}{2}, \frac{1}{2} \} \\
E(X) = 50
\]

\[
R_{|X_1 - 0.5|} = \{ 0.5 \} \\
f_{|X_1 - 0.5|} = \{ 1.0 \} \\
E(|X_1 - 0.5|) = 0.5
\]

\[
R_{|X_2 - 50|} = \{ 50 \} \\
f_{|X_2 - 50|} = \{ 1.0 \} \\
E(|X_2 - 50|) = 50
\]
Discrete Random Variables: Variance

Example: \( X_1 = \) “Flip a coin and return the number of heads showing”
\( X_2 = \) “Flip a coin and return \( 100 \times \) the number of heads showing”

Second attempt: deviation = absolute value of distance from \( E(X) \)

\[ \text{deviation} = |X - E(X)| \quad \text{Var}(X) = E[|X - E(X)|] \]

Looks promising! What’s wrong with that?

\textbf{Ugh!} Requires case analysis and blows up with an exponential number of cases, and resulting in functions that are not continuous; such "piece-wise" functions are very hard to work with!

\[ f(x) = |x - 30| + |x + 50| + |x/2 + 10| \]
Discrete Random Variables: Variance

Ok, finally, here is the best definition:

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

This is the standard definition and has several advantages:

- It is much easier to work with mathematically;
- Like the absolute value, it gives only positive values.

But it gives results which are not very intuitive!

Alternate notation for expected value:

\[ \mu_X = E(X) \]

or just \( \mu \) if \( X \) is obvious.

\[ R_{X_1} = \{ 0, 1 \} \quad f_{X_1} = \{ \frac{1}{2}, \frac{1}{2} \} \quad E(X) = 0.5 \]

\[ R_{X_2} = \{ 0, 100 \} \quad f_{X_2} = \{ \frac{1}{2}, \frac{1}{2} \} \quad E(X) = 50 \]

\[ R_{(X_1-0.5)^2} = \{ 0.25 \} \quad f_{(X_1-0.5)^2} = \{ 1.0 \} \quad E[(X_1 - 0.5)^2] = 0.25 \]

\[ R_{(X_2-50)^2} = \{ 2500 \} \quad f_{(X_2-50)^2} = \{ 1.0 \} \quad E[(X_2 - 50)^2] = 2500 \]

And what about the units? If these are dollars, then this is 2500 squared dollars...
Discrete Random Variables: Standard Deviation

Therefore a more common measure of spread around the mean is the Standard Deviation:

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

\[
\begin{align*}
R_{X_1} &= \{0, 1\} & f_{X_1} &= \{\frac{1}{2}, \frac{1}{2}\} & E(X) &= 0.5 \\
R_{X_2} &= \{0, 100\} & f_{X_2} &= \{\frac{1}{2}, \frac{1}{2}\} & E(X) &= 50 \\
R_{(X_1-0.5)^2} &= \{0.25\} & f_{(X_1-0.5)^2} &= \{1.0\} & \text{Var}(X_1) &= 0.25 & \sigma_{X_1} &= 0.5 \\
R_{(X_2-50)^2} &= \{2500\} & f_{(X_2-50)^2} &= \{1.0\} & \text{Var}(X_2) &= 2500 & \sigma_{X_2} &= 50
\end{align*}
\]

This has all the advantages of the variance, plus two more:

- The units are correct; and
- It corresponds to a well-known geometric notion, the Euclidean Distance.
Distance Metrics measure how much two vectors differ from one another.

In two dimensions, our last two definitions of variance correspond to

1-norm = SAD = Sum of Absolute Differences (or "Manhattan Distance")

2-norm distance = ED = Euclidean Distance = Square root of sum of squared distances

\[ 1\text{-norm distance} = \sum_{i=1}^{n} |x_i - y_i| \]

\[ 2\text{-norm distance} = \left( \sum_{i=1}^{n} |x_i - y_i|^2 \right)^{1/2} \]

Pythagorean Theorem

Standard Deviation is consistent with the 2-norm distance metric.
Discrete Random Variables: Standard Deviation

Useful formulae for the Variance and Standard Deviation:

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

\[ \begin{align*}
\text{Var}(X) &= E[(X - \mu_X)^2] \\
&= E[X^2 - 2\mu X + \mu^2] \\
&= E(X^2) - E(2\mu X) + E(\mu^2) \\
&= E(X^2) - 2\mu \times E(X) + \mu^2 \\
&= E(X^2) - 2\mu^2 + \mu^2 \\
&= E(X^2) - \mu^2
\end{align*} \]
Discrete Random Variables: Standard Deviation

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

**Theorem:** \( \text{Var}(aX + b) = a^2 \times \text{Var}(X) \)

**Proof:**

\[
\text{Var}(aX + b) = E \left[ \left( (aX + b) - \mu_{aX+b} \right)^2 \right] \\
= E \left[ \left( (aX + b) - (a\mu_X + b) \right)^2 \right] \\
= E \left[ \left( a(X - \mu_X) \right)^2 \right] \\
= E \left[ a^2 \times (X - \mu_X)^2 \right] \\
= a^2 \times E \left[ (X - \mu_X)^2 \right] \\
= a^2 \times \text{Var}(X)
\]

**Corollary:**

\[
\sigma_{aX+b} = |a| \times \sigma_X
\]
Let's apply this idea to our game:

**Game One:** For $1 per round, you can flip a coin, and I'll give you $6 (your bet back plus $5) if heads appears, and nothing if tails appears (so you lose your bet). Call this the random variable $X_1 = \text{net payout}$.

$$E(X_1) = -1 \cdot \frac{1}{2} + 5 \cdot \frac{1}{2} = 2.00$$

**Game Two:** For $1 per round, you can flip a coin 20 times, and if you get 20 heads, I'll give you $3,145,728, else nothing. Call this the random variable $X_2$.

$$E(X_2) = -1 \cdot \frac{2^{20} - 1}{2^{20}} + 3,145,727 \cdot \frac{1}{2^{20}} = 2.00$$

Oh, and by the way, I only have time for a single round....

$$E(X_1^2) = \frac{(-1)^2}{2} + \frac{5^2}{2} = \frac{26}{2} = 13$$

$$Var(X_1) = E(X_1^2) - E(X_1)^2 = 13 - 2^2 = 9$$

$$\sigma_{X_1} = \sqrt{9} = 3$$

$$E(X_2^2) = (-1)^2 \cdot \frac{2^{20} - 1}{2^{20}} + 3,145,727^2 \cdot \frac{1}{2^{20}} = \frac{3,145,727^2}{2^{20}} + \frac{2^{20} - 1}{2^{20}} = 9,437,179$$

$$Var(X_2) = E(X_2^2) - E(X_2)^2 = 9,437,179 - 2^2 = 9,437,175$$

$$\sigma_{X_2} = \sqrt{9437175} = 3071.9985 = 3,072.00$$
Discrete Random Variables: Standardized RVs

In order to compare random variables and avoid “comparing apples and oranges,” we use the notion of a Standardized Random Variable which has

- $E(X) = 0.0$
- $\text{Var}(X) = \sigma_x = 1.0$

In order to create a standard version of a random variable, we simply subtract the expected value (so now the expected value is 0.0) and divide by the standard deviation (so now it is 1.0):

$$X^* \overset{\text{def}}{=} \frac{X - \mu_x}{\sigma_x}$$
Cumulative Distribution Functions

One more topic before considering the standard distributions....

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:

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

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

<table>
<thead>
<tr>
<th>$k$</th>
<th>$f_X(k)$</th>
<th>$F_X(k)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.16</td>
<td>0.16</td>
</tr>
<tr>
<td>2</td>
<td>0.16</td>
<td>0.32</td>
</tr>
<tr>
<td>3</td>
<td>0.16</td>
<td>0.48</td>
</tr>
<tr>
<td>4</td>
<td>0.16</td>
<td>0.64</td>
</tr>
<tr>
<td>5</td>
<td>0.16</td>
<td>0.80</td>
</tr>
<tr>
<td>6</td>
<td>0.16</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Probability Distribution Function $f_X$          Cumulative Distribution Function $F_X$