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:

\[
E(X) = \sum_{k \in R_X} k \cdot P(X = k)
\]

Example:

\( R_Y = \{ 0, 1, 2, 3 \} \)
\( f_Y = \{ \frac{1}{6}, \frac{2}{6}, \frac{2}{6}, \frac{1}{6} \} \)

\[
E(Y) = \sum_{k \in R_Y} k \cdot f(k) = \frac{0}{6} + \frac{1}{3} + \frac{2}{3} + \frac{3}{6} = 1.5
\]
**Linearity of Expected Value**

**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:**

$$ E(aX + b) = \sum_{k \in R_X} (a \cdot k + b) \cdot P_X(k) $$

$$ = \sum_{k \in R_X} (a \cdot k \cdot P_X(k)) + \sum_{k \in R_X} (b \cdot P_X(k)) $$

$$ = a \cdot \sum_{k \in R_X} (k \cdot P_X(k)) + b \cdot \sum_{k \in R_X} P_X(k) $$

$$ = a \cdot E(X) + b \cdot 1.0 $$

$$ = a \cdot E(X) + b $$

**Corollary:** For any constant $c$, $E(c) = c$. This will make many calculations involving expected value MUCH easier!

**Linearity of Expected Value**

**Theorem (Linearity of Expectation of Sum of Random Variables)**

For any random variables $X$ and $Y$, 

$$ E(X + Y) = E(X) + E(Y) $$

**Proof:**

$$ E(X + Y) = \sum_{j \in R_X} \sum_{k \in R_Y} (j + k) \cdot P(X = j, Y = k) $$

$$ = \sum_{j \in R_X} \sum_{k \in R_Y} j \cdot P(X = j, Y = k) + k \cdot P(X = j, Y = k) $$

$$ = \sum_{j \in R_X} \sum_{k \in R_Y} j \cdot P(X = j, Y = k) + \sum_{k \in R_Y} k \cdot P(X = j, Y = k) $$

$$ = \sum_{j \in R_X} j \cdot \left( \sum_{k \in R_Y} P(X = j, Y = k) \right) + \sum_{k \in R_Y} k \cdot \left( \sum_{j \in R_X} P(X = j, Y = k) \right) $$

$$ = \sum_{j \in R_X} j \cdot P(X = j) + \sum_{k \in R_Y} k \cdot P(Y = k) $$

$$ = E(X) + E(Y) $$

where in the second-to-last step, we used the Law of Total Probability: 

$$ P(A) = \sum_{\mathclap{\gamma \in R_\gamma}} P(A \cap R_\gamma) = \sum_{\mathclap{\gamma \in R_\gamma}} P(A | R_\gamma) P(R_\gamma) $$.

Note: $X$ and $Y$ do not have to be independent! Again, this will simplify many proofs to come.
Expected Value: Basic Properties

**Theorem (Linearity of Expectation of Products of Independent Random Variables)**

For any independent random variables $X$ and $Y$,

$$E(X \cdot Y) = E(X) \cdot E(Y)$$

**Proof:**

$$E(X \cdot Y) = \sum_{j \in R_X} \sum_{k \in R_Y} j \cdot k \cdot P(X = j, Y = k)$$

$$= \sum_{j \in R_X} \sum_{k \in R_Y} j \cdot k \cdot P(X = j) \cdot P(Y = k)$$

$$= \sum_{j \in R_X} j \cdot P(X = j) \cdot \left( \sum_{k \in R_Y} k \cdot P(Y = k) \right)$$

$$= \sum_{j \in R_X} j \cdot P(X = j) \cdot E(Y)$$

$$= \left( \sum_{j \in R_X} j \cdot P(X = j) \right) \cdot E(Y)$$

$$= E(X) \cdot E(Y)$$

**Note:** $X$ and $Y$ must be independent!

---

**Corollary:** This theorem is not true if $X$ and $Y$ are dependent:

Here is a counter example.

Suppose you flip a coin, let $X =$ the number of heads showing and $Y =$ the number of tails showing. It should be obvious $X$ and $Y$ are not independent, in fact $Y = 1 - X$.

<table>
<thead>
<tr>
<th>$XY$</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>0</td>
<td>0.0</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>0.0</td>
</tr>
</tbody>
</table>

$$E(X \cdot Y) = 0 \cdot (0.0) + 0 \cdot (0.5) + 0 \cdot (0.5) + 1 \cdot (0.0) = 0.0$$

This confirms your intuition that if you flip a coin and multiply the number of heads showing by the number of tails showing (one of which is always 0), you’ll always get 0.

But:

$$E(X) \cdot E(Y) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$
**Expected Value of the Standard Distributions**

**Expected Value of Bernoulli**

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

\[ R_X = \{ 0, 1 \} \]

\[ P_X = \{ 1 - p, p \} \]

\[ E(X) = 1 \cdot p + 0 \cdot (1 - p) = p \]

**Expected Value of Binomial**

\[ X \sim B(N, p) \]

\[ R_X = \{ 0, \ldots, N \} \]

\[ P_X(k) = \binom{N}{k} p^k (1 - p)^{N-k} \]

Formally, if \( Y \sim \text{Bernoulli}(p) \), and

\[ X = \text{"The number of successes in } N \text{ trials of } Y = \frac{Y + Y + \ldots + Y}{N} \]

By the linearity of expectation of sums of RVs we immediately have:

\[ E(X) = E(Y_1) + E(Y_2) + \ldots + E(Y_N) = N \cdot E(Y) = N \cdot p \]
Geometric Distribution: Expected Value

To derive the Expected Value, we can use the fact that $X \sim \text{Geometric}(p)$ has the memoryless property and break into two cases, depending on the result of the first Bernoulli trial.

Let

- $X_S$ = “result of $X$ when there is a success on the first trial”
- $X_F$ = “result of $X$ when there is a failure on the first trial”

Clearly,

- $E(X_S) = 1$
- $E(X_F) = 1 + E(X)$ for the remaining trials

By the memoryless property!

Thus we have:

$$
\begin{align*}
\mu_X &= 1 + (1 - p)(1 + \mu_X) \\
&= p + 1 - p + \mu_X - p \mu_X \\
&= 1 + \mu_X - p \mu_X \\
0 &= 1 - p \mu_X \\
p \mu_X &= 1 \\
\mu_X &= \frac{1}{p}
\end{align*}
$$

E$(X) = \frac{1}{p}$

Pascal Distribution: Expected Value

Since the Pascal is simply an “iterated” version of the Geometric, we can use the theorem again!

Formally, if $Y \sim \text{Bernoulli}(p)$ and

- $X = “The number of trials of $Y$ until $m$ successes occur”$

= $Y_1 + \ldots + Y_m$

$m$ times

Then

$X \sim \text{Pascal}(m, p)$

and by the linearity of expectation for sums of RVs we have

$$
E(X) = E(Y_1) + \cdots + E(Y_m) = m \cdot E(Y) = \frac{m}{p}
$$
**Problems with Expected Value**

Expected value is a one-number summary of a possibly very complex distribution!

How useful is this? Are there any issues or problems to keep in mind??

**Issue #1:** The expected value may not exist (may be $\infty$)!

For countably infinite discrete RVs, we calculate the expected value using an infinite sum, for example the expected number of trials until the first success in Geometric(1/2):

$$E(X) = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + \cdots + k \cdot \frac{1}{2^k} + \cdots = 2.0$$

The problem is that such sequences may not converge, and the expected value may be infinite (you’ll see what I mean when you do the current homework).

---

**Problems with Expected Value**

Expected value is a one-number summary of a possibly very complex distribution!

How useful is this? Are there any issues or problems to keep in mind??

**Issue #2:** *(The Inspection Paradox)*

For some problems, the expected value is very subtle, and depends, essentially on your point of view, or perhaps on naively confusing "average value" with "expected value."

Example: Suppose Wayne is teaching two classes at BU, LX 496 with 20 students, and CS 237 with 120 students.

What is the "average class size" for Wayne?

$$\frac{20 + 120}{2} = 70$$

What is the "average class size" for a student in one of these classes?

$$\frac{20 \cdot 20 + 120 \cdot 120}{140} = 105.7143$$
Discrete Random Variables: Expected Value

Issue #3: Expectation is only one point of comparison, there are many others!

Example: 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 $11 (net: $10) if heads appears, and nothing if tails appears (net: -$1). Call this the random variable $X_1$:

$$E(X_1) = 10 \cdot \frac{1}{2} - 1 \cdot (1 - \frac{1}{2}) = 4.50$$

Game Two: For $1 per round, you can flip a coin 20 times, and if you get 20 heads, I'll give you $5,767,168, else you lose the $1. Call this the random variable $X_2$:

$$E(X_2) = 5,767,167 \cdot \frac{1}{2^{20}} - 1 \cdot (1 - \frac{1}{2^{20}}) = 4.50$$

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

Any takers for Game TWO??

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 \( E(X) \).

But how to define "deviation"? Whatever we pick it should work on simple examples.

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

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

Example: \( X_1 = "\text{Flip a coin and return the number of heads showing}" \)
\( X_2 = "\text{Flip a coin and return 100 * the number of heads showing}" \)

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

Note that \( E(X) \) is a constant:

\[
E(X - E(X)) = E(X) - E(X) = 0.0
\]

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

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

Example: \( X_1 = "\text{Flip a coin and return the number of heads showing}" \)
\( X_2 = "\text{Flip a coin and return 100 * the number of heads showing}" \)

\[
\begin{align*}
R_{X_1} &= \{ 0, 1 \} \quad P_{X_1} = \{ \frac{1}{2}, \frac{1}{2} \} \quad E(X) = 0.5 \\
R_{X_2} &= \{ 0, 100 \} \quad P_{X_2} = \{ \frac{1}{2}, \frac{1}{2} \} \quad E(X) = 50 \\
R_{X_1,0.51} &= \{ 0.51 \} \quad P_{X_1,0.51} = \{ 1.0 \} \quad E(1X_1 - 0.51) = 0.5 \\
R_{X_2,501} &= \{ 501 \} \quad P_{X_2,501} = \{ 1.0 \} \quad E(1X_2 - 501) = 50
\end{align*}
\]
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 \( E(X) \)

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

Looks promising! What’s wrong with that?

**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! Anyone want to take the derivative of the following function?

\[
f(x) = |x - 30| + |x + 50| + |x/2 + 10|
\]

---

Discrete Random Variables: Variance

Ok, finally, here is the best definition:

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

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!

- \( R_{X_1} = \{0, 1\} \quad P_{X_1} = \{\frac{1}{3}, \frac{2}{3}\} \quad E(X) = 0.5 \)
- \( R_{X_2} = \{0, 100\} \quad P_{X_2} = \{\frac{1}{2}, \frac{1}{2}\} \quad E(X) = 50 \)
- \( R_{X_3, -0.5^2} = \{0.25\} \quad P_{X_3, -0.5^2} = \{1.0\} \quad E[(X_3 - 0.5)^2] = 0.25 \)
- \( R_{X_4, -50^2} = \{2500\} \quad P_{X_4, -50^2} = \{1.0\} \quad E[(X_4 - 50)^2] = 2500 \)

Alternate notation for expected value:

\[
\mu_X = E(X)
\]

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

And what about the units? If these are dollars, then this is 2500 squared dollars...
Therefore a more common measure of spread around the mean is the Standard Deviation:

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

\[ R_{X_1} = \{ 0, 1 \} \quad P_{X_1} = \{ \frac{1}{2}, \frac{1}{2} \} \quad E(X) = 0.5 \]
\[ R_{X_2} = \{ 0, 100 \} \quad P_{X_2} = \{ \frac{1}{2}, \frac{1}{2} \} \quad E(X) = 50 \]
\[ R_{X_{0.5}^2} = \{ 0.25 \} \quad P_{X_{0.5}^2} = \{ 1.0 \} \quad \text{Var}(X_1) = 0.25 \quad \sigma_{X_1} = 0.5 \]
\[ R_{X_{50}^2} = \{ 2500 \} \quad P_{X_{50}^2} = \{ 1.0 \} \quad \text{Var}(X_2) = 2500 \quad \sigma_{X_2} = 50 \]

This has all the advantages of the variance, plus three more:
- It explains simple examples;
- The units are correct; and
- It corresponds to a well-known geometric notion, the Euclidean Distance.

---

**Discrete Random Variables: Variance and StdDev**

Let's apply this idea to our games:

**Game One:** For $1 per round, you can flip a coin, and I'll give you $11 (net: $10) if heads appears, and nothing if tails appears (net: -$1). Call this the random variable \( X_1 \):

\[
E(X_1) = 10 \cdot \frac{1}{2} - 1 \cdot (1 - \frac{1}{2}) = \$4.50
\]

**Game Two:** For $1 per round, you can flip a coin 20 times, and if you get 20 heads, I'll give you $5,767,168; else you lose the $1. Call this the random variable \( X_2 \):

\[
E(X_2) = 5,767,167 \cdot \frac{1}{2^{20}} - 1 \cdot (1 - \frac{1}{2^{20}}) = \$4.50
\]

\[
\text{Var}(X_1) = E[(X_1 - \mu_X)^2]
\]
\[
= \frac{(10 - 4.5)^2}{2} + \frac{(-1 - 4.5)^2}{2}
\]
\[
= \frac{5.5^2 + (-5.5)^2}{2}
\]
\[
= 5.5^2
\]
\[
= 30.25
\]
\[
\sigma_{X_1} = \$5.50
\]

\[
\text{Var}(X_2) = E[(X_2 - \mu_X)^2]
\]
\[
= \frac{(5,767,167 - 4.5)^2}{2^{20}} + \frac{(-5.5)^2}{2^{20}} \cdot \frac{2^{20}}{2^{20}} - 1
\]
\[
= 31,719,393.75
\]

\[
\sigma_{X_2} = \$5,631
\]
**Discrete Random Variables: Variance and StdDev**

Useful formulae for the Variance and Standard Deviation:

**Theorem:**

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

**Proof:**

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

Recall that \( E(X) \) is a constant!

\[
\text{Var}(X_2) = E[(X_2 - \mu_X)^2] \\
= \frac{(5,767,167 - 4.5)^2}{2^{20}} + (-5.5)^2 \cdot \frac{2^{20} - 1}{2^{20}} \\
= 31,719,393.75
\]

\( \sigma_{X_1} = \$5,631 \)

\[
\text{Var}(X_1) = E(X_1^2) - E(X_1)^2 \\
= \frac{(5,767,167)^2}{2^{20}} + (-1) \cdot \frac{2^{20} - 1}{2^{20}} \\
= 31,719,414 + 1 = 31,719,414 \\
\text{Var}(X_1) = 31,719,414 - 4.5^2 \\
= 31,719,393.75 \\
\sigma_{X_1} = \$5,631
\]

**Discrete Random Variables: Variance and StdDev**

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

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

**Proof:**

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

**Corollary:**

\[ \sigma_{aX+b} = |a| \cdot \sigma_X \]
Discrete Random Variables: Variance and StdDev

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:**

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

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…

---

Variance of Standard Distributions

<table>
<thead>
<tr>
<th>( X \sim \text{Bernoulli}(p) )</th>
<th>( Y \sim \text{Binomial}(N, p) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E(X) = p )</td>
<td>( E(Y) = N \cdot p )</td>
</tr>
<tr>
<td>( \text{Var}(X) = E(X^2) - E(X)^2 )</td>
<td>( \text{Var}(Y) = N \cdot p \cdot (1 - p) )</td>
</tr>
<tr>
<td>( = E(X) - p^2 )</td>
<td></td>
</tr>
<tr>
<td>( = p \cdot (1 - p) )</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( Z \sim \text{Geometric}(p) )</th>
<th>( W \sim \text{Pascal}(m, p) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E(Z) = 1/p )</td>
<td>( E(W) = \frac{m}{p} )</td>
</tr>
<tr>
<td>( \text{Var}(Z) = \frac{1 - p}{p^2} )</td>
<td>( \text{Var}(W) = \frac{m(1 - p)}{p^2} )</td>
</tr>
</tbody>
</table>

| \( W = Z_1 + Z_2 + \ldots + Z_m \) | |
Expectation and Variance: Summary of Theorems

However, independence, as usual, makes things simpler:

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

Let $X$ and $Y$ be independent random variables, then

$$Var(X + Y) = Var(X) + Var(Y)$$

**Proof:**

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

$$= E[X^2 + 2XY + Y^2] - (E(X) + E(Y))^2$$

$$= E(X^2) + 2E(XY) + E(Y^2) - [E(X)^2 - 2E(Y)E(Y) - E(Y)^2]$$

$$= E(X^2) - E(X)^2 + E(Y)^2 - E(Y)^2 + 2[E(XY) - E(Y)E(Y)]$$

$$= Var(X) + Var(Y)$$

---

Variance of Standard Distributions

<table>
<thead>
<tr>
<th>$X \sim Bernoulli(p)$</th>
<th>$Y \sim Binomial(N, p)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E(X) = p$</td>
<td>$Y = X_1 + X_2 + \cdots + X_N$</td>
</tr>
<tr>
<td>$Var(X) = E(X^2) - E(X)^2$</td>
<td>$E(Y) = N \cdot p$</td>
</tr>
<tr>
<td>$= E(X) - p^2$</td>
<td>$Var(Y) = N \cdot p \cdot (1 - p)$</td>
</tr>
<tr>
<td>$= p \cdot (1 - p)$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$Z \sim Geometric(p)$</th>
<th>$W \sim Pascal(m, p)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E(Z) = 1/p$</td>
<td>$W = Z_1 + Z_2 + \cdots + Z_m$</td>
</tr>
<tr>
<td>$Var(Z) = \frac{1 - p}{p^2}$</td>
<td>$E(W) = \frac{m}{p}$</td>
</tr>
<tr>
<td></td>
<td>$Var(W) = \frac{m(1 - p)}{p^2}$</td>
</tr>
</tbody>
</table>