CS 237: Probability in Computing

Wayne Snyder
Computer Science Department
Boston University

Lecture 13:

- Review: Expectation of Random Variables, Games
- Properties of Expectation
  - Linearity of Expectation
  - Expectation of Sum of Independent Random Variables
- Variance and Standard Deviation of Random Variables
  - Properties of Variance

Discrete RandomVariables: 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:

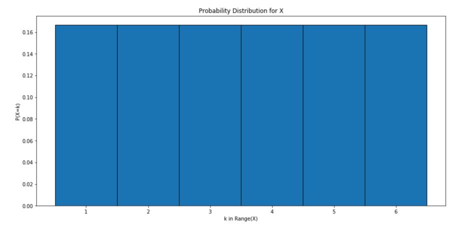
$$E(X) = \sum_{k \in R_X} k \cdot P(X = k)$$

Example: X = “the number of dots showing on a single thrown die”

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

$$f_X = \{ \tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6} \}$$

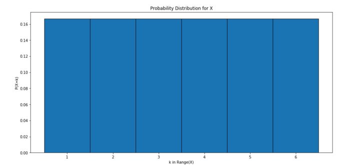
$$E(X) = \sum_{k \in R_X} \frac{k}{6} = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5$$



Discrete RandomVariables: Expected Value

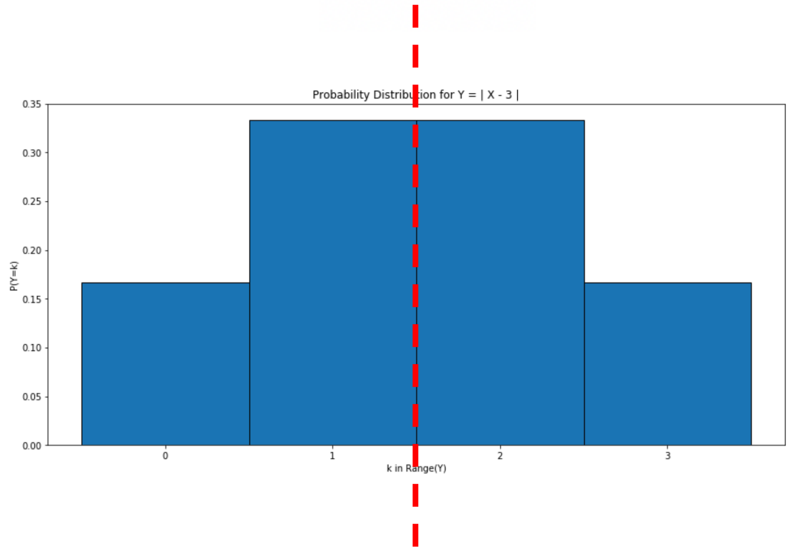
$$R_X = \{1, 2, 3, 4, 5, 6\}$$

$$f_X = \{\ \tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6}\ \}$$



Example: Y = | X – 3 |

$E(\,Y\,)$



$$R_Y = \{\ 0, 1, 2, 3\ \}$$

$$f_Y = \{\ \tfrac{1}{6}, \tfrac{2}{6}, \tfrac{2}{6}, \tfrac{1}{6}\ \}$$

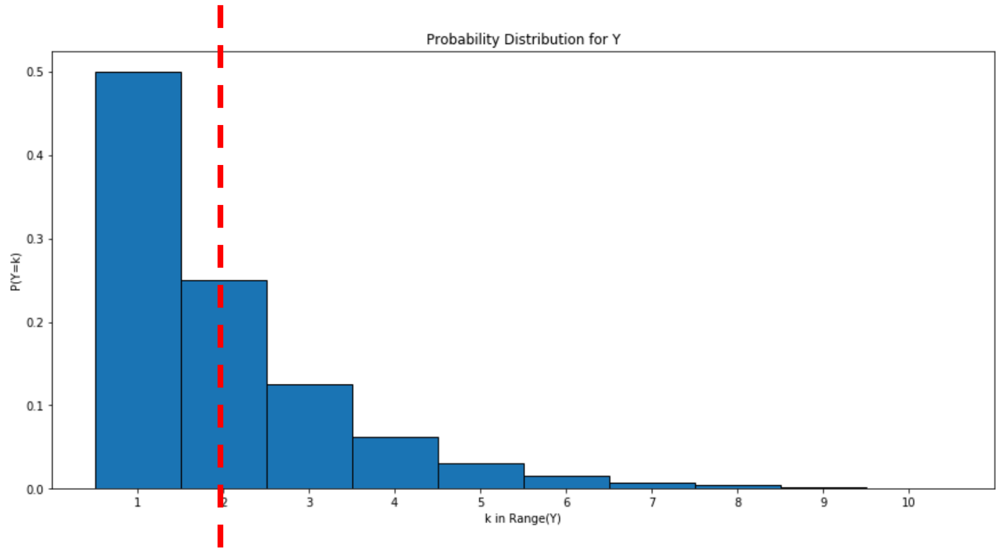
$$E(\,Y\,) \;=\; \sum_{k\,\in\,R_Y} k * f(k) \;=\; \frac{0}{6} + \frac{1}{3} + \frac{2}{3} = \frac{3}{6} \;=\; 1.5$$

Discrete Random Variables: Expected Value

Example: Y = "tosses of a fair coin until a heads appears"

$$R_Y = \{\ 1, 2, 3,\ \dots\ \}$$

$$f_Y = \{\ \tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{8},\ \dots \}$$



$$E(Y) = \sum_{k \in R_Y} k * f(k) = \frac{1}{2} + \frac{2}{4} + \frac{3}{8} + \frac{4}{16} \dots = 2.0$$

Expected Value: Basic Properties

Theorem (Linearity of Expectation)

For any random variable **X** and real numbers **a** and **b**,

$$E(\, a * X + b \,) \;=\; a * E(X) + b$$

Proof:

$$\begin{aligned}
E(\, aX + b \,) &= \sum_{k \in R_X} (a * k + b) * P_X(k) \\
&= \sum_{k \in R_X} (a * k * P_X(k)) + (b * P_X(k)) \\
&= \sum_{k \in R_X} (a * k * P_X(k)) \;+\; \sum_{k \in R_X} (b * P_X(k)) \\
&= a * \sum_{k \in R_X} (k * P_X(k)) \;+\; b * \sum_{k \in R_X} P_X(k) \\
&= a * E(X) \;+\; b * 1.0 \\
&= a * E(X) \;+\; b
\end{aligned}$$

This will make many calculations involving expected value MUCH easier!

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

Expected Value: Basic Properties

Theorem (Expectation of Sums of Random Variables):

If X and Y are two discrete random variables (not necessarily independent), then:

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

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

If $S_1, \ldots, S_n$ is a partition of the sample space S, and A is an event, then $A \cap S_1, S \cap S_2, \ldots, S \cap S_n$ is a partition of the even A, and

$$P(A) \; = \sum_{1 \le i \le n} P(A, S_i)$$

(This is essentially case analysis, breaking A up into n disjoint cases.)

Expected Value: Basic Properties

Theorem (Expectation of Product of Independent Random Variables):

If X and Y are two *independent* discrete random variables, then:

$$
\begin{aligned}
E(X \cdot Y) &= \sum_{j \in R_X} \sum_{k \in R_Y} j \cdot k \cdot P(X = j, Y = k) \\
&= \sum_{j \in R_X} \sum_{k \in R_Y} j \cdot k \cdot P(X = j) \cdot P(Y = k) \\
&= \sum_{j \in R_X} j \cdot P(X = j) \cdot \left(\sum_{k \in R_Y} k \cdot P(Y = k) \right) \\
&= \sum_{j \in R_X} j \cdot P(X = j) \cdot E(Y) \\
&= E(Y) \cdot \sum_{j \in R_X} j \cdot P(X = j) \\
&= E(Y) \cdot E(X) \\
&= E(X) \cdot E(Y)
\end{aligned}
$$

where in the second step, we used the independence of X and Y.

Expected Value: Basic Properties

Note: This theorem is not true if X and Y are dependent:

It is easy to see that this result is *not* true for dependent variables: Consider the following. Flip a coin and let X count the number of heads and Y count the number of tails. Clearly X and Y are *not* indepedent, and in fact $Y = 1 - X$. Clearly

$$E(X) \cdot E(Y) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

But we have (showing the product and then the probability in parentheses):

XY	0	1
0	0 (0)	0 (1/2)
1	0 (1/2)	1 (0)

So

$$E(X \cdot Y) = 0 \cdot 0 + 0 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} + 1 \cdot 0 = 0$$

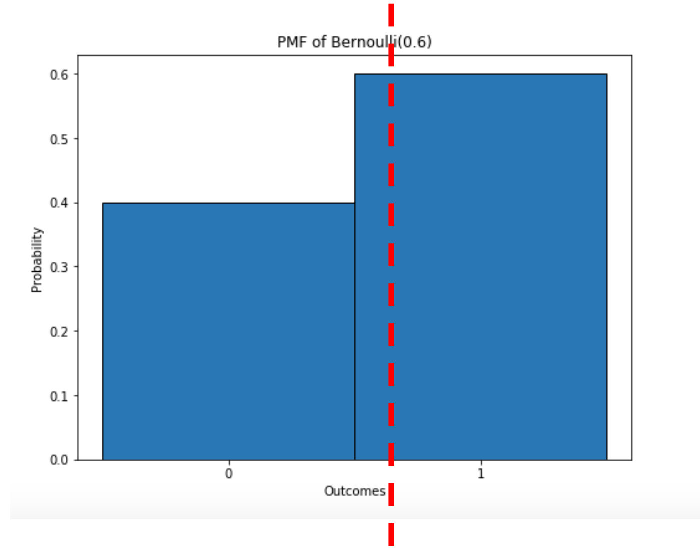
Expected Value of the Standard Distributions

Expected Value of Bernoulli

$$X \sim \text{Bernoulli}(p)$$

$$R_X = \{0, 1\}$$

$$P_X = \{1 - p, p\}$$

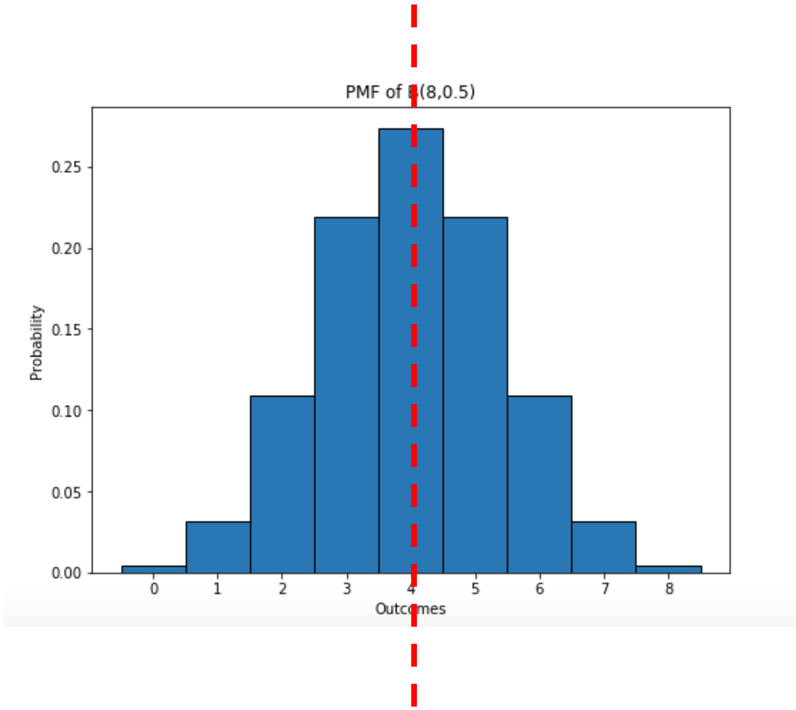


$$E(X) = 1 \cdot p + 0 \cdot (1 - p) = p$$

Expected Value of the Standard Distributions

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 ~ Bernoulli(p), and

$$X = \text{“The number of successes in N trials of Y”} = \overbrace{Y + Y + \ldots + Y}^{\text{N times}}$$

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

E(X) = N * p

Geometric Distribution: Expected Value

To derive the expected value, we can use the fact that X ~ G(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,

- $E(X_S) = 1$
- $E(X_F) = 1 + E(\text{ X for the remaining trials })$

  $= 1 + E(\text{ X })$

  By the memoryless property!

Thus we have:

$$\begin{aligned} \mu_X &= 1p + (1-p)(1+\mu_X) \\ &= p + 1 - p + \mu_X - p\,\mu_X \\ &= 1 + \mu_X - p\,\mu_X \end{aligned}$$

$$0 = 1 - p\,\mu_X$$

$$p\,\mu_X = 1$$

$$\mu_X = 1/p$$

Geometric Distribution

Example

Suppose you draw cards WITH replacement until you get an Ace. How many draws would you expect it to take?

Solution:

On average, how many independent games of poker are required until a particular player is dealt a **Royal Flush**?

Solution: This is $G(0.00000154)$. $E(X) = 1/\ 0.00000154 = 649{,}350.6493$

Geometric Distribution

Example

Suppose you draw cards WITH replacement until you get an Ace. How many draws would you expect it to take?

Solution: This is G(1/13). E(X) = 13

On average, how many independent games of poker are required until a particular player is dealt a **Royal Flush**?

Solution: This is G(0.00000154). E(X) = 1/ 0.00000154 = 649,350.6493

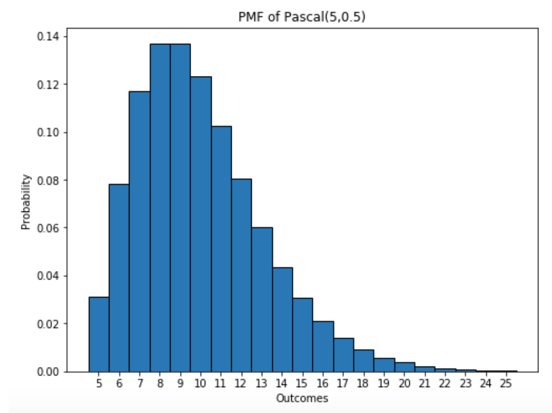
Pascal Distribution: Expected Value

Since the Pascal is simply an "iterated" version of the Geometric, we can use the linearity of expectation again!

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

X = "The number of trials of Y until m successes occur" /

$$= \underbrace{Y_1 + \dots + Y_m}_{\text{m times}}$$



Then

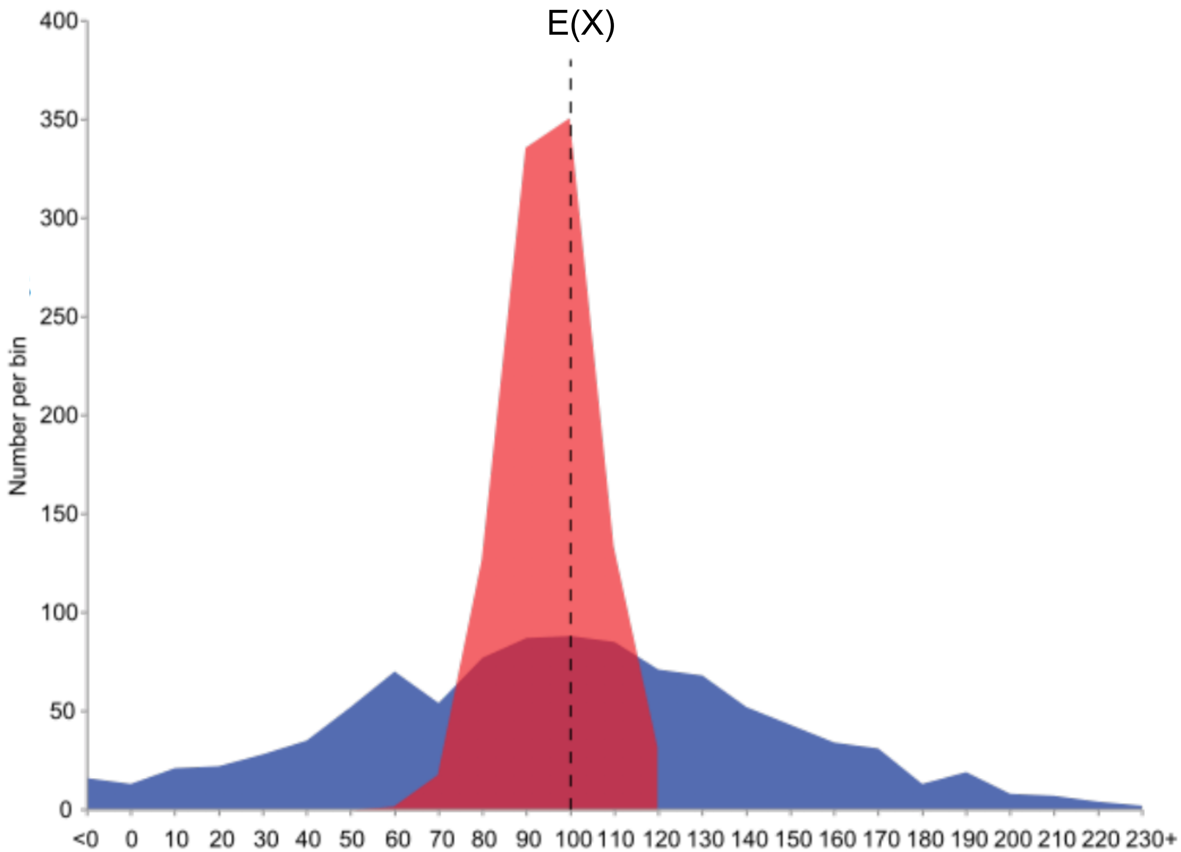
$$X \sim Pascal(m, p)$$

and by the linearity of expectation we have

$$E(X) = E(Y_1) + \cdots + E(Y_m) = m \cdot E(Y) = m \,/\, p$$

Discrete RandomVariables: Variance

The question is: How much does X vary from E(X)? How spread out is the probability distribution around the expected value?



Discrete RandomVariables: Variance

The **Variance** of a random variable, 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**

deviation = X – E(X) Var(X) = E[X – E(X)]

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"

$R_{X_1} = \{ 0, 1 \}$ $P_{X_1} = \{ \frac{1}{2}, \frac{1}{2} \}$ $E(X) = 0.5$

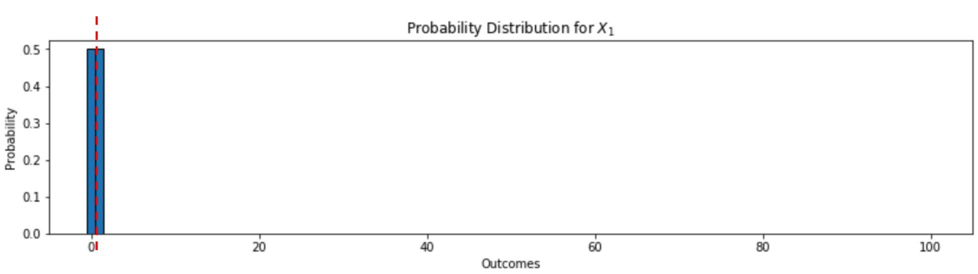
$R_{X_2} = \{ 0, 100 \}$ $P_{X_2} = \{ \frac{1}{2}, \frac{1}{2} \}$ $E(X) = 50$

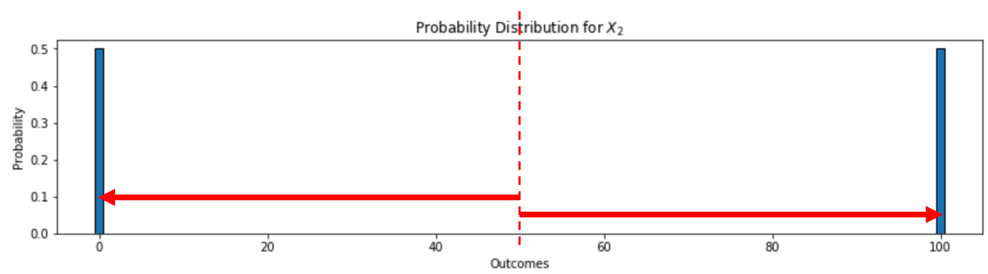
$R_{X_1-0.5} = \{ -0.5, 0.5 \}$ $P_{X_1-0.5} = \{ \frac{1}{2}, \frac{1}{2} \}$ $E(X_1 - 0.5) = E(X_1) - 0.5 = 0.0$

$R_{X_2-50} = \{ -50, 50 \}$ $P_{X_2-0.5} = \{ \frac{1}{2}, \frac{1}{2} \}$ $E(X_2 - 50) = E(X_2) - 50 = 0.0$

Note that E(X) is a constant:

$E(X - E(X)) = E(X) - E(X) = 0.0$





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)

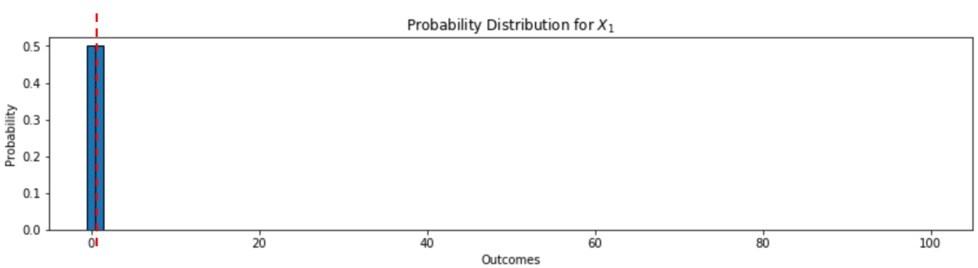
deviation = | X – E(X) | Var(X) = E[| X – E(X) |]

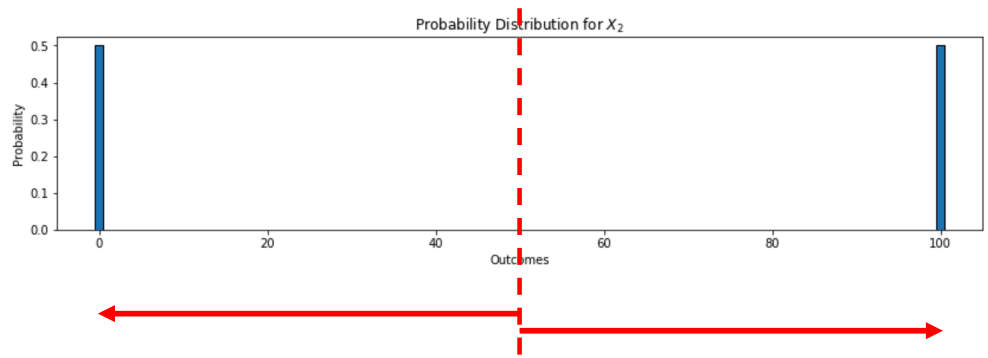
$R_{X_1} = \{\, 0,\ 1 \,\} \qquad P_{X_1} = \{\, \frac{1}{2}, \frac{1}{2} \,\} \qquad E(X) = 0.5$

$R_{X_2} = \{\, 0,\ 100 \,\} \qquad P_{X_2} = \{\, \frac{1}{2}, \frac{1}{2} \,\} \qquad E(X) = 50$

$R_{|X_1 - 0.5|} = \{\, 0.5 \,\} \qquad P_{|X_1 - 0.5|} = \{\, 1.0 \,\} \qquad E(\,|X_1 - 0.5|\,) = 0.5$

$R_{|X_2 - 50|} = \{\, 50 \,\} \qquad P_{|X_2 - 50|} = \{\, 1.0 \,\} \qquad E(\,|X_2 - 50|\,) = 50$





Discrete RandomVariables: 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"

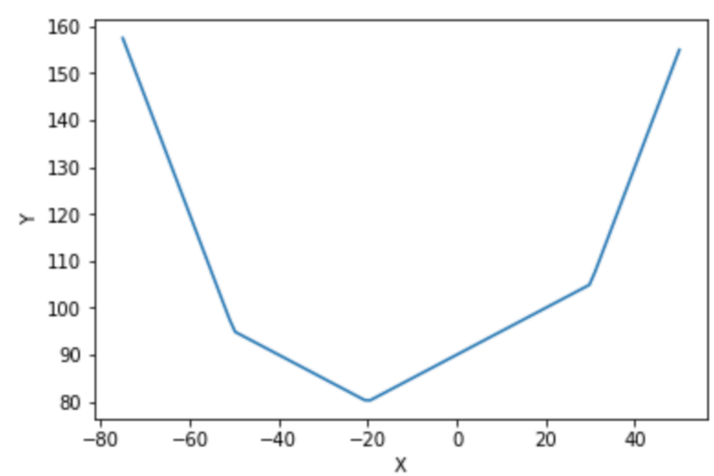
Second attempt: deviation = absolute value of distance from E(X)

deviation = $| X - E(X) |$ 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:

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

Alternate notation for expected value:

$$\mu_X = E(X)$$

or just μ if **X** is obvious.

This is the standard definition and has several advantages:

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

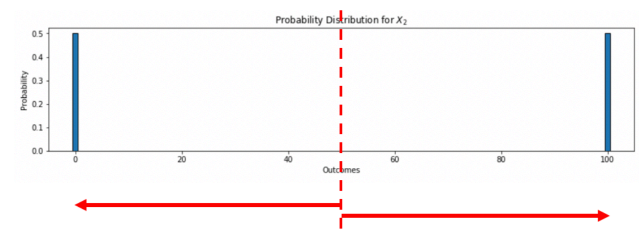
But it gives results which are not very intuitive!

$R_{X_1} = \{\,0,\ 1\,\}$ $\quad P_{X_1} = \{\,\frac{1}{2}, \frac{1}{2}\,\}$ $\quad E(X) = 0.5$

$R_{X_2} = \{\,0,\ 100\,\}$ $\quad P_{X_2} = \{\,\frac{1}{2}, \frac{1}{2}\,\}$ $\quad E(X) = 50$

$R_{(X_1-0.5)^2} = \{\,0.25\,\}$ $\quad P_{(X_1-0.5)^2} = \{\,1.0\,\}$ $\quad E[\,(X_1 - 0.5)^2] = \underline{0.25}$

$R_{(X_2-50)^2} = \{\,2500\,\}$ $\quad P_{(X_2-50)^2} = \{\,1.0\,\}$ $\quad E[\,(X_2 - 50)^2\,] = \underline{2500}$



And what about the units?
If these are dollars, then this is 2500 squared dollars...

Discrete RandomVariables: Standard Deviation

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

$$\sigma_X =_{def} \sqrt{Var(X)}$$

$$R_{X_1} = \{\, 0,\ 1 \,\} \qquad P_{X_1} = \{\, \tfrac{1}{2}, \tfrac{1}{2} \,\} \qquad E(X) = 0.5$$

$$R_{X_2} = \{\, 0,\ 100 \,\} \quad P_{X_2} = \{\, \tfrac{1}{2}, \tfrac{1}{2} \,\} \qquad E(X) = 50$$

$$R_{(X_1-0.5)^2} = \{\, 0.25 \,\} \qquad P_{(X_1-0.5)^2} = \{\, 1.0 \,\} \qquad Var(X_1) = 0.25 \qquad \sigma_{X_1} = 0.5$$

$$R_{(X_2-50)^2} = \{\, 2500 \,\} \qquad P_{(X_2-50)^2} = \{\, 1.0 \,\} \qquad Var(X_2) = 2500 \qquad \sigma_{X_2} = 50$$

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 RandomVariables: 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 :

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

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 :

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

$$\begin{aligned} Var(X_1) &= E[(X_1 - \mu_X)^2] \\ &= \frac{(10-4.5)^2}{2} + \frac{(-1-4.5)^2}{2} \\ &= \frac{5.5^2 + (-5.5)^2}{2} \\ &= 5.5^2 \\ &= 30.25 \end{aligned}$$

$$\sigma_{X_1} = \$5.50$$

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

$$\sigma_{X_2} = \$5,631$$

Discrete RandomVariables: Variance and StdDev

Useful formulae for the Variance and Standard Deviation:

Theorem:

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

Proof:

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

Recall that E(X) is a constant!

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

$$\sigma_{X_2} = \$5,631$$

$$Var(X_2) = E(X_2^2) - E(X_2)^2$$

$$E(X_2^2) = \frac{(5,767,167)^2}{2^{20}} + (-1) \cdot \frac{2^{20} - 1}{2^{20}}$$
$$= 31,719,413 + 1 = 31,719,414$$

$$Var(X_2) = 31,719,414 - 4.5^2$$
$$= 31,719,393.75$$

$$\sigma_{X_2} = \$5,631$$

Discrete RandomVariables: Variance and StdDev

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

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

Proof:

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

Corollary:

$$\sigma_{aX+b} = |a| * \sigma_X$$

Discrete RandomVariables: 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

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

Proof:

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

This term is called the Covariance of X and Y, **Cov(X,Y)**, and measures how much they "vary together". **For independent RV, Cov(X,Y) = 0.** This will be back in a few weeks….