CS 237: Probability in Computing

Wayne Snyder
Computer Science Department
Boston University

Lecture 14:

- A Story about Expectation: "Regression to the Mean"
- Variance and Standard Deviation of Discrete Random Variables
  - Properties of Variance and Standard Deviation
  - Variance and Standard Deviation of Standard Distributions
- Other Single-Value Measures of Random Variables
  - Mode
  - Skew
- Limit Theorems [if time]

Review: Expectation:

Expected Value of a Random Variable:

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

Alternate notation for expected value:

$$\mu_X = E(X)$$

or just μ if **X** is obvious.

Properties of Expectation:

Linearity of Expection for (Arbitrary) Random Variables

$E(\, a * X + b \,) = a * E(X) + b$ for constants a and b

Linearity of Sums of (Arbitrary) Random Variables:

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

Linearity of Products of <u>Independent</u> Random Variables:

$E(\, X * Y \,) = E(X) * E(Y)$

Obvious fact:

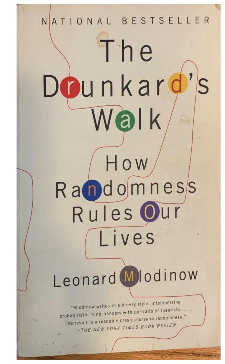
For any constant a,

$E(a) = a$

Less obvious but also true:

$E(\, E(X) \,) = E(X)$

A Story about Expectation: "Regression to the Mean"



The greatest challenge in understanding the role of randomness in life is that although the basic principles of randomness arise from everyday logic, many of the consequences that follow from those principles prove counterintuitive. Kahneman and Tversky's studies were themselves spurred by a random event. In the mid-1960s, Kahneman, then a junior psychology professor at Hebrew University, agreed to perform a rather unexciting chore: lecturing to a group of Israeli air force flight instructors on the conventional wisdom of behavior modification and its application to the psychology of flight training. Kahneman drove home the point that rewarding positive behavior works but punishing mistakes does not. One of his students interrupted, voicing an opinion that would lead Kahneman to an epiphany and guide his research for decades.[6]

"I've often praised people warmly for beautifully executed maneuvers, and the next time they always do worse," the flight instructor said. "And I've screamed at people for badly executed maneuvers, and by and large the next time they improve. Don't tell

me that reward works and punishment doesn't work. My experience contradicts it." The other flight instructors agreed. To Kahneman the flight instructors' experiences rang true. On the other hand, Kahneman believed in the animal experiments that demonstrated that reward works better than punishment. He ruminated on this apparent paradox. And then it struck him: the screaming preceded the improvement, but contrary to appearances it did not cause it.

How can that be? The answer lies in a phenomenon called regression toward the mean. That is, in any series of random events an extraordinary event is most likely to be followed, due purely to chance, by a more ordinary one. Here is how it works: The student pilots all had a certain personal ability to fly fighter planes. Raising their skill level involved many factors and required extensive practice, so although their skill was slowly improving through flight training, the change wouldn't be noticeable from one maneuver to the next. Any especially good or especially poor performance was thus mostly a matter of luck. So if a pilot made an exceptionally good landing—one far above his normal level of performance—then the odds would be good that he would perform closer to his norm—that is, worse—the next day. And if his instructor had praised him, it would appear that the praise had done no good. But if a pilot made an exceptionally bad landing—running the plane off the end of the runway and into the vat of corn chowder in the base cafeteria—then the odds would be good that the next day he would perform closer to his norm—that is, better. And if his instructor had a habit of screaming "you clumsy ape" when a student performed poorly, it would appear that his criticism did some good. In this way an *apparent* pattern would emerge: student performs well, praise does no good; student performs poorly, instructor compares student to lower primate at high volume, student improves. The instructors in Kahneman's class had concluded from such experiences that their screaming was a powerful educational tool. In reality it made no difference at all.

Discrete RandomVariables: Variance

Variance of a Random Variable:

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

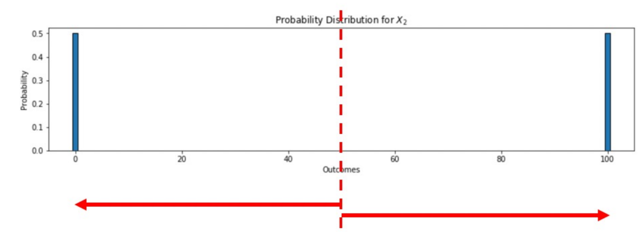
Example: X_1 = “Flip a coin and return the number of heads showing”
X_2 = “Flip a coin and return 100 * the number of heads showing”

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

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

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

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



But this is not very intuitive! 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!

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

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

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

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

$$\begin{aligned} Var(X_2) &= 31,719,414 - 4.5^2 \\ &= 31,719,393.75 \end{aligned}$$

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

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:

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

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….

Variance of Standard Distributions

$$X \sim Bernoulli(p)$$

$$E(X) = p$$

$$\begin{aligned} Var(X) &= E(X^2) - E(X)^2 \\ &= E(X) - p^2 \\ &= p \cdot (1 - p) \end{aligned}$$

$$Y \sim Binomial(N, p)$$

$$Y = X_1 + X_2 + \cdots + X_N$$

$$E(Y) = N \cdot p$$

$$Var(Y) = N \cdot p \cdot (1 - p)$$

$$Z \sim Geometric(p)$$

$$E(Z) = 1/p$$

$$Var(Z) = \frac{1-p}{p^2}$$

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

$$W = Z_1 + Z_2 + \cdots + Z_m$$

$$E(W) = \frac{m}{p}$$

$$Var(W) = \frac{m(1-p)}{p^2}$$