Revised: 2/7/2021

Probability III

Michael C. Hackett

Assistant Professor, Computer Science

Community College of Philadelphia

Random Variables

- A **random variable** is a random process or variable with a numerical outcome, with a probability for each of these possible outcomes.
  - The number selected on a Roulette wheel is a ***discrete random variable***: The result is 0 through 36, each with the same probability
  - The sum of rolling a pair of dice is also a discrete random variable: The result is 2 through 12, each sum having (mostly) different probabilities.
  - Discrete random variables are *countable*.

- Random variables are usually represented with a capital letter.
  - For example, we'll use X to represent the revenue per download for a hypothetical dating app for statisticians called "Statr"

Random Variables

- This hypothetical app, Statr, comes in three versions:
  - A free version with limited functionality
    - The free version earns $0 for the developers per download
  - A free, ad-supported version with full functionality
    - The ad-supported version earns $0.50 for the developers per download
  - A paid version with no ads and full functionality
    - The paid version earns $1 for the developers per download

Random Variables

- We'll say we anticipate the following number of downloads next month:
  - The free version will have 500 downloads

    $x_1 = \$0.00$

    $500 \times \$0.00 = \$0.00$
  - The ad-supported version will have 750 downloads

    $x_2 = \$0.50$

    $750 \times \$0.50 = \$375.00$
  - The paid version will have 250 downloads

    $x_3 = \$1.00$

    $250 \times \$1.00 = \$250.00$

i	1	2	3
x_i	\$0	\$0.50	\$1.00

Random Variables

- We can also calculate the probability for each type of download based on the previous slide's data:

$$\frac{Number\ of\ downloads\ for\ that\ type}{Total\ number\ of\ downloads}$$

i	1	2	3
x_i	\$0	\$0.50	\$1.00
$P(X = x_i)$	0.33	0.50	0.17

- The average revenue per download (*the **expected value** of X*) is:

$$E(X) = \frac{\$0 + \$375 + \$250}{1500} = \$0.42$$

Random Variables

i	1	2	3
x_i	\$0.00	\$0.50	\$1.00
$P(X = x_i)$	0.33	0.50	0.17
$x_i \times P(X = x_i)$	\$0.00	\$.25	\$0.17

- The expected value of X could also be computed by adding each outcome weighted by its probability:

$$E(X) = \$0.00 \times P(X = \$0.00) + \$0.50 \times P(X = \$0.50) + \$1.00 \times P(X = \$1.00)$$

$$E(X) = \$0.00 \times 0.33 + \$0.50 \times 0.50 + \$1.00 \times 0.17$$

$$E(X) = \$0.00 + \$0.25 + \$0.17 = \$0.42$$

Random Variables

- The equation for expected value on the previous slide can be generalized to:

$$E(X) = x_1 \times P(X = x_1) \; + \cdots + x_n \times P(X = x_n)$$

or

$$E(X) = \sum_{i=1}^{n} x_i \times P(X = x_i)$$

Random Variables

- The expected value of a random variable represents the average outcome.

$$E(X) = 0.42$$

could also be written as

$$\mu = \$0.42$$

Variance of Random Variables

- In a previous lecture, we saw that variance was computed by taking the average of the sums of squared deviations.

$$\sigma^2 = \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \ldots + (x_n - \mu)^2}{n - 1}$$

Variance of Random Variables

- For random variables, we again compute the sums of squared deviations from the mean/expected value.
  - We take their sums, weighted by their corresponding probabilities, just as was done for the expectation.

$$Var(X) = \sigma^2 = (x_1 - \mu)^2 \times P(X = x_1) + \cdots + (x_n - \mu)^2 \times P(X = x_n)$$

or

$$Var(X) = \sigma^2 = \sum_{i=1}^{n} (x_i - \mu)^2 \times P(X = x_i)$$

Variance of Random Variables

i	1	2	3
x_i	\$0.00	\$0.50	\$1.00
$P(X = x_i)$	0.33	0.50	0.17
$x_i \times P(X = x_i)$	\$0.00	\$.25	\$0.17
$(x_i - \mu)^2$	\$0.18	\$0.006	\$0.34
$(x_i - \mu)^2 \times P(X = x_1)$	\$0.06	\$0.003	\$0.06

- The expected value of X could also be computed by adding each outcome weighted by its probability:

$$Var(X) = \sigma^2 = (\$0 - 0.42)^2 \times P(X = \$0) + (\$0.50 - 0.42)^2 \times P(X = \$0.50) + (\$1.00 - 0.42)^2 \times P(X = \$1.00)$$

$$\sigma^2 = 0.18 \times 0.33 + 0.006 \times 0.50 + 0.34 \times 0.17$$

$$\sigma^2 = 0.06 + 0.003 + 0.06 = 0.12$$

Variance of Random Variables

- The standard deviation for a random variable is still calculated by taking the square root of the variance:

$$\sigma^2 = 0.12$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{0.12} = \$0.35$$

Linear Combinations

- The Statr example focused on one random variable, X, which was the revenue per download of the app.

- It is also possible to combine random variables.

- Let’s say another hypothetical app, “Stuber”, a ride-sharing app for statisticians, averages 150 downloads per day.

Linear Combinations

- Each day of the week will be represented by random variables D_1 through D_7
  - D_1 is Sunday and D_7 is Saturday

- The total weekly downloads are the sum of those seven variables
  - $W = D_1 + D_2 + D_3 + D_4 + D_5 + D_6 + D_7$

Linear Combinations

- The average (expected value) number of downloads per day was stated to be 150
- The average (expected value) of weekly downloads is the sum of the averages (expected values) of the seven variables:

$$E(W) = E(D_1) + E(D_2) + E(D_3) + E(D_4) + E(D_5) + E(D_6) + E(D_7)$$

$$E(W) = 150 + 150 + 150 + 150 + 150 + 150 + 150 = 1050$$

Linear Combinations

- This example has described a linear combination of random variables.

- More formally, the linear combination of random variables is given by

$$aX + bY$$

  - Where X and Y are random variables, and a and b are fixed numbers.

Linear Combinations

- In our example, we had more than just two (X and Y) random variables
  - We had seven: D_1 through D_7
  - $aD_1 + bD_2 + cD_3 + dD_4 + eD_5 + fD_6 + gD_7$

- There were no explicit coefficients, but we could use a fixed coefficient of 1 for a, b, c…
  - $W = 1D_1 + 1D_2 + 1D_3 + 1D_4 + 1D_5 + 1D_6 + 1D_7$

Linear Combinations

- To compute the expected value of this linear combination of random variables:

$$E(W) = 1 \times E(D_1) + 1 \times E(D_2) + 1 \times E(D_3) + 1 \times E(D_4) + 1 \times E(D_5) + 1 \times E(D_6) + 1 \times E(D_7)$$

$$E(W) = 150 + 150 + 150 + 150 + 150 + 150 + 150 = 1050$$

- Same as we previously had calculated but done more formally.

Linear Combinations

- We'll demonstrate another example that uses values other than 1 for coefficients.

- Let's imagine we purchased one Bitcoin (\$38138.39) and one Litecoin (\$155.67), two cryptocurrencies.
  - Prices on 2/7/21

- We wish to calculate the expected gain or loss of our cryptocurrency, $E(C)$, at the end of the month
  - Based on the expected change in value of Bitcoin and Litecoin at the end of the month.

Linear Combinations

$$aX + bY$$

- Our random variables are the change in price of the two cryptocurrencies.
  - X is the change in Bitcoin price
  - Y is the change in Litecoin price
- Our coefficients are our fixed values
  - a is the purchase price of our Bitcoin: \$38138.39
  - b is the purchase price of our Litecoin: \$155.67
- C is the value of our cryptocurrency

$$C = \$38138.39 \times X + \$155.67 \times Y$$

Linear Combinations

- We'll say we expect Bitcoin's price to drop 5% (-0.05) and Litecoin's price to increase 3% (0.03)

$$E(C) = \$38138.39 \times E(X) + \$155.67 \times E(Y)$$

$$E(C) = \$38138.39 \times -0.05 + \$155.67 \times 0.03$$

$$E(C) = -\$1906.92 + \$4.67 = -\$1902.25$$

- We expect our cryptocurrency to lose \$1902.25 of its USD value over the next month

Variance in Linear Combinations

- Bitcoin and Litecoin (like cryptocurrency in general) experience a lot of volatility with dramatic price increases and decreases, often within the same 24 hour period.

- We can calculate the variability of a linear combination to describe such uncertainties.
  - Cryptocurrency, though, has notoriously high variability
  - For this example, we'll even assume that each has low 30-day variance
    - Bitcoin with a variance of 0.08 and Litecoin with a variance of 0.04

Variance in Linear Combinations

$$Var(aX + bY) = a^2 \times Var(X) + b^2 \times Var(Y)$$

- The equation above is for calculating the variance of linear combinations.

$$Var(aX + bY) = 38138.39^2 \times 0.08 + 155.67^2 \times 0.04$$
$$Var(aX + bY) = 1454536791.79 \times 0.08 + 24233.15 \times 0.04$$
$$Var(aX + bY) = 116{,}362{,}943.34 + 969.33 = 116{,}363{,}912.67$$

Variance in Linear Combinations

- We still calculate the standard deviation by taking the square root of the variance.

$$\sigma = \sqrt{116{,}363{,}912.67} = \$10{,}787.21$$

- Basically, this would indicate that in ~68% of future months, our cryptocurrency is expected to be worth $\pm\$10{,}787.21$ our cryptocurrency's average worth.
    - This is assuming a lot of factors and that is still has a LOT of variability

Linear Combinations

- We can also determine the variability of Stuber downloads
  - We expected each day to have 150 downloads
  - We'll say each day has a standard deviation of 25 downloads
    - This makes the daily variance $25^2 = 625$
  - The coefficient for each day was 1

$$\sigma^2 = 1^2 \times 625 + 1^2 \times 625 + 1^2 \times 625 + 1^2 \times 625 + 1^2 \times 625 + 1^2 \times 625 + 1^2 \times 625$$

$$\sigma^2 = 625 + 625 + 625 + 625 + 625 + 625 + 625 = 4375$$

$$\sigma = \sqrt{4375} = 66.14$$

- This would indicate ~68% of the time, the number of daily app downloads is 150 ± 66.14