Probability

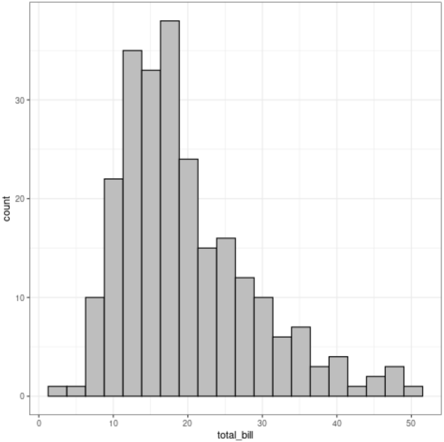
Expectation and Variance

Grinnell College

March 7, 2025

Review – Describing Distributions

When we saw *distributions* earlier in the class, they were a way to represent information from a quantitative variable.



Two of the big things we talked about were center and spread.

- Measures of center we used were median or mean
- Measures of spread we used were IQR or std. dev.

Review – Describing Distributions

Mean is the same thing as the **average** value of the variable.

$$\overline{x} = \frac{\sum x_i}{n}$$

- even if a distribution is skewed, the mean can still be useful
  - more on this in a sec

Standard Deviation is one of the measures of spread

$$s = \sqrt{\frac{1}{n-1}\sum(x_i - \overline{x})^2}$$

- interpretation: the average distance of observations from the mean
- larger value of s $\rightarrow$ more variability or spread
- sometimes variance is used instead (variance $= s^2$)

Goal for Today

We are going to apply the concept of center and spread (mean and standard deviation) to the probability distributions concept we saw on Wednesday.

Review – Probability Distributions

A **probability distribution** represents each of the *disjoint* outcomes of a random process and their associated probabilities

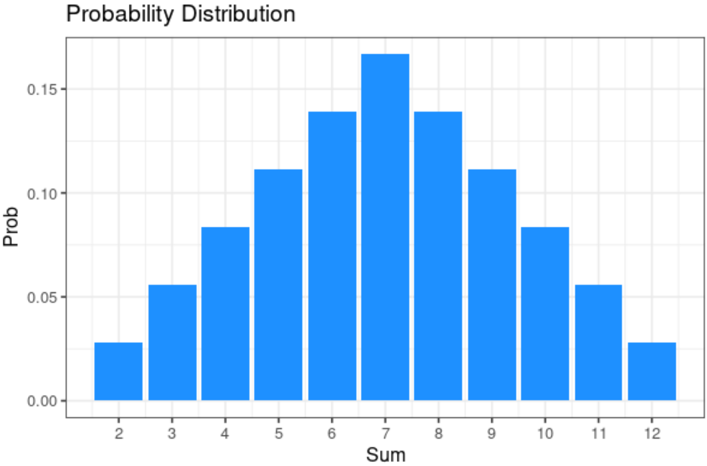
Dice Sum	2	3	4	5	6	7	8	9	10	11	12
Probability	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

For a probability distribution to be valid, the following must be true:

1. The outcomes are disjoint
2. Every probability is between 0 and 1
3. The sum of all probabilities must equal 1

Review – Probability Distributions

Dice Sum	2	3	4	5	6	7	8	9	10	11	12
Probability	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$



Random Variable

When working with a *random process* (like die rolling, or coin flipping) we can construct a quantitative variable that tells us about the outcome of that process.

Typically we will label a random variable with capital letters to distinguish it from data variables in our data sets.

Example 1: rolling a six-sided die

- X = result of die roll (can be 1,2,3,4,5 or 6)
- $P(X=6) = \frac{1}{6}$

Example 2: coin flip

- Y = 1 if heads, Y = 0 if tails
- $P(Y=1) = P(H) = P(T) = P(Y = 0) = 0.5$

Expectation

When talking about the center of a probability distribution, most often the *mean* is used (even if the distribution is skewed). We will use the term **Expected Value** to denote that this is an average for a *random variable*.

To compute the **Expected Value**, we need to use the outcomes and account for how likely they are to come up. The *expected value* of a random variable X is denoted E(X)

Formula:

$$E(X) = \sum x_i p_i$$

- x_i represent the value of outcome i
- p_i is the probability associated with outcome i

Expectation – Example

Let's look at the probability distribution for rolling one 6-sided die.

Die Result	1	2	3	4	5	6
Probability	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Let X = result of roll, then

$$\begin{aligned} E(X) &= \sum x_i p_i \\ &= 1(\frac{1}{6}) + 2(\frac{1}{6}) + 3(\frac{1}{6}) + 4(\frac{1}{6}) + 5(\frac{1}{6}) + 6(\frac{1}{6}) \\ &= \frac{21}{6} = 3.5 \end{aligned}$$

- Interpretation: the *expected* outcome is 3.5
- Interpretation: If you roll many 6-sided dice and compute the average, you can expect a value close to 3.5

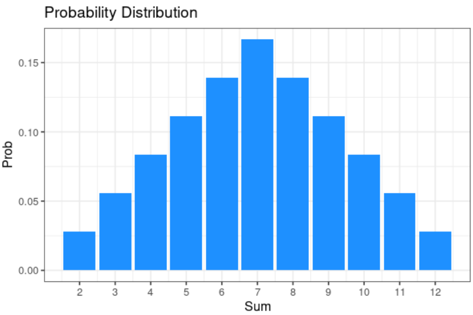
Expectation for Multiple RV's

Let's return to the 2 die rolling example. When computing the expected value of the sum, it is time consuming because of how many outcomes there are. There is a simpler rule.

Expectation of Sums
Let X and Y be two random variables. Their sum $X + Y$ has expected value $\mathrm{E}(X + Y) = \mathrm{E}(X) + \mathrm{E}(Y)$

This works for more than 2 RV's too. The rule makes finding an average of lots of processes very easy

Sum of Dice Example



For rolling 2 dice and adding them up, visually we can see the expected value (mean) should be 7. Let's use the result from the last slide

- Let X = result of die 1
- Let Y = result of die 2
- Let Z = sum of two dice = X + Y
- Then E(Z) = E(X) + E(Y) = 3.5 + 3.5 = 7

Variance

We may also want to talk about the *variability* of a process. It's nice to know the average of a d6 is 3.5, but how much variability can I expect when I roll it?

Working with **variance** ($=s^2$) is usually easier than std.dev. directly, although interpretations with std.dev. are easier

- variance of a random variable is denoted Var(X)

Formula: Let $\mu = E(X)$.

$$\begin{aligned} Var(X) &= E[(X-\mu)^2] = E(X^2) - \mu^2 \\ &= \left(\sum x_i^2 p_i\right) - \mu^2 \end{aligned}$$

- variance = expected squared deviation from the mean
- we won't do much calculation of Variance directly
- convert to std.dev. to do interpretations

Variance Rules

There are rules for variances similar to expected values when working with multiple random variables (you won't be tested on these, but *simple* HW problems may show up)

- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ if X and Y are independent
- $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y)$ if X and Y are independent
- $\text{Var}(cX) = c^2\text{Var}(X)$

Variance Example

Let's go back to the 6-sided die rolling example.

Die Result	1	2	3	4	5	6
Probability	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Let X = result of roll, then

$$\begin{aligned} Var(X) &= (\sum x_i^2 p_i) - \mu^2 \\ &= [1^2(\frac{1}{6}) + 2^2(\frac{1}{6}) + 3^2(\frac{1}{6}) + 4^2(\frac{1}{6}) + 5^2(\frac{1}{6}) + 6^2(\frac{1}{6})] - 3.5^2 \\ &= \frac{91}{6} - 3.5^2 = \frac{105}{36} \approx 2.92 \end{aligned}$$

- variance $= 2.92 \rightarrow$ s.d. $= \sqrt{2.92} = 1.71$
- interpretation: a d6 result is 1.71 away from the mean, on average