Lab 1

Jordan Garrett

1/7/2022

About me

Purpose of labs

- Facilitate learning theoretical concepts covered in lecture
- Answer questions about homework
- Review for midterms and the final
- Implement statistical analyses from lecture in RStudio

Lab 1 Outline

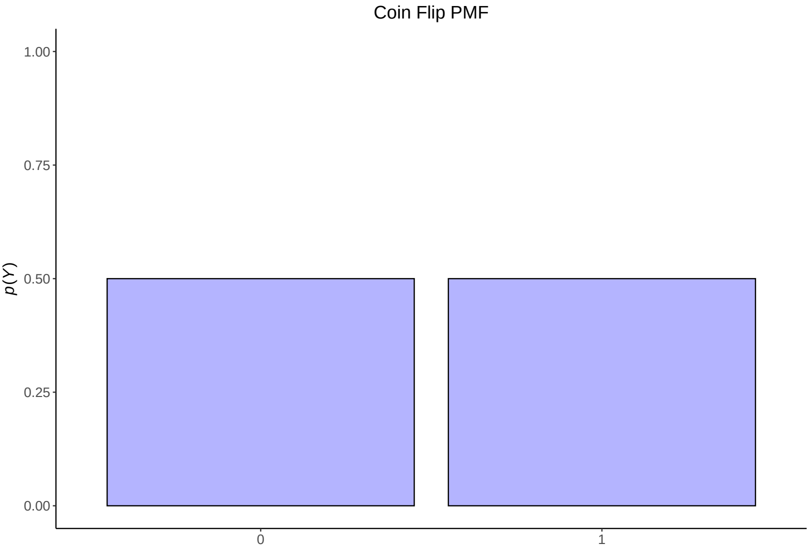
- Cover preliminary statistics and probability concepts
- Intro to R and setup

Random Variable

- A variable that takes on specific values with specific probabilities.
- Measurable functions that map outcomes of a stochastic process to a measurable space.
- Typically denoted by a capital letter (e.g., X, Y, Z).
- Denote the outcome of a coin flip as Y
  - $Y = 1$ if Heads, else $Y = 0$
  - $p(Y{=}1) = 0.5$

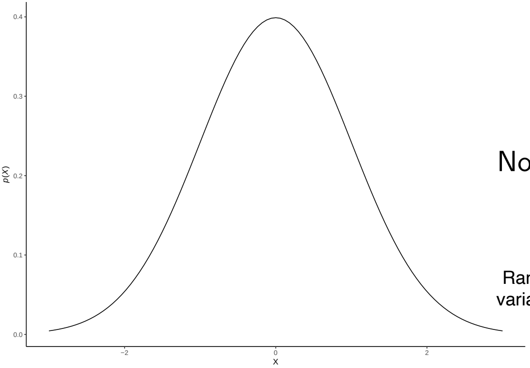
Probability distribution

- Probability **mass** function (PMF)
  - Assigns probabilities to individual values of a *discrete* random variable



Probablity distribution

- Probability **density** function (PDF)
  - Similar to a PMF, but instead specifies the probability that a *continuous* variable takes on a range of values.



Normal Distribution Notation

$$X \sim N(\mu, \sigma^2)$$

Random variable X

has a probability distribution

that is Normal

with a mean and variance

Expected value and Variance

- The expected value of a random variable Y is denoted as $E(Y)$.
  - Probability weighted average of all possible values.

```
y <- c(70,80,85,90,100)
p.y <- c(0.18,0.34,0.35,0.11,0.02)

E.y <- sum(y*p.y)
E.y
```

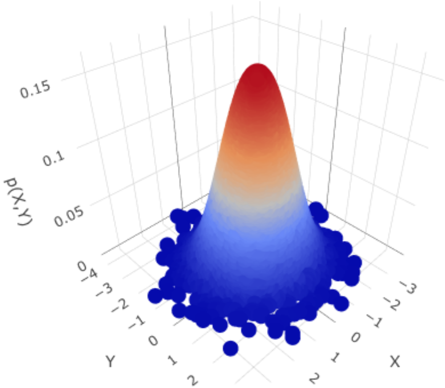
```
## [1] 81.45
```

- Variance
  - A measure of how disperse all possible values of a random variable are from the expected value (i.e. population or sample mean)

Multivariate Distributions

PDF → **Joint Probability Density**

- For the random variables X and Y, the joint pdf characterizes the probability that each X and Y takes on a set of values.



Multivariate Distributions

Variance $\rightarrow$ **Covariance**

- Measure of how much two random variables vary together.
- Formally, is the expected value of each random variable's deviation from its respective expected value.

$$\text{cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$

- Typically represented as a matrix.

Marginal Distribution

- Probability distribution of an outcome for one random variable in the presence of all other outcomes for another random variable

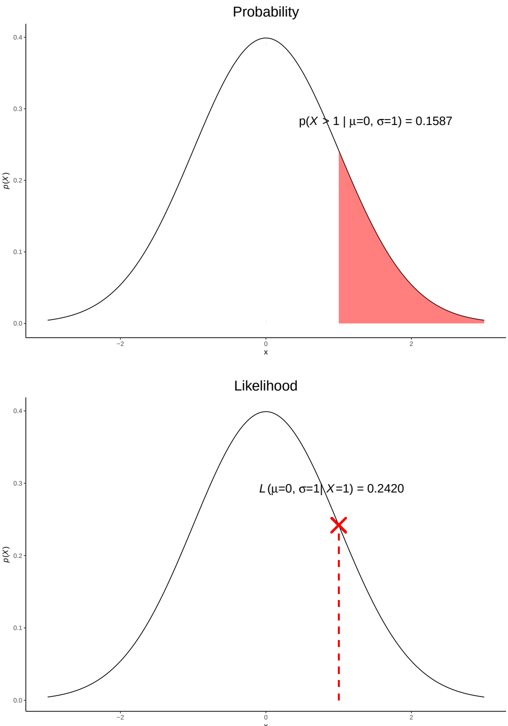
	x_1	x_2	x_3	x_4	$p_Y(y_i)$
y_1	0.125	0.0625	0.03125	0.03125	**0.25**
y_2	0.09375	0.1875	0.09375	0.09375	**0.46875**
y_3	0.28125	0	0	0	**0.28125**
$p_X(x_i)$	**0.5**	**0.25**	**0.125**	**0.125**	1

Pearson correlation & Statistical independence

- Pearson correlation coefficient ρ measures the linear relationship between two random variables
- Two random variables are statistically independent if the realization of one does not affect the outcome of the other.

Likelihood

- Throughout this course we are going to use statistical models (e.g., regression, ANOVA) to describe patterns of variability in random variables.
- These models have *parameters* (e.g., mean of a sampling distribution)
- A **likelihood** function is the joint probability of observed data as a function of parameters in a statistical model.


Probability
p(X > 1 | μ=0, σ=1) = 0.1587
p(X)
x
Likelihood
L (μ=0, σ=1| X=1) = 0.2420
p(X)
x

R

Programming Tips

- Google is your best friend!
- More than a single way to skin a cat (code).
- Learn more than one language.
- Have fun!