Probability: Random Variables and Large Samples

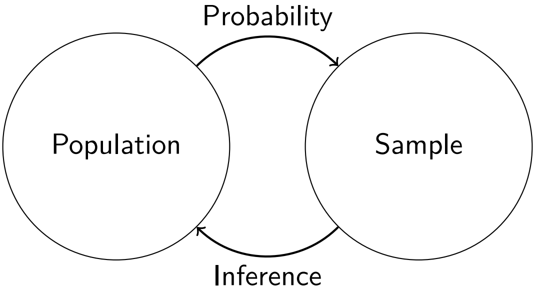
PSC4375: Week 10

Prof. Weldzius

Villanova University

Slides Updated: 2025-03-24

Learning about populations



- We want to learn about the chance process that generated our data.
  - What's the true support for Trump in the population?
  - We only get to see a sample from the population.
  - Stare at the results of 1000 coin flips and determine if the coin was fair.
- We have probability to help us, but. . .

What are random variables?

$$\{\text{draw a Trump supporter}\} \xleftrightarrow{???} \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Randomly selecting senators, redux

	Democrats	*Republicans*	*Independents*	*Total*
Men	29	43	2	74
Women	16	10	0	26
Total	45	53	2	100

- Draw a Senator's name from a hat and define the random variable:
- A **random variable** is a mapping from the outcomes to numbers.
  - Example: $X = 1$ if selected Senator is a woman, $X = 0$ otherwise
- **Random**: before we draw, there is uncertainty about the value of X!
- Straightforward probability connection:

$$\mathbb{P}(X = 1) = \mathbb{P}(\text{draw a woman senator}) = \frac{26}{100}$$

Bernoulli r.v.

- An r.v. is X is said to follow a Bernoulli distribution with probability p if:
    - X takes on only two values, 0 and 1, and
    - $\mathbb{P}(X = 1) = p$ and $\mathbb{P}(X = 0) = 1 - p$
- Simplest possible random variable: indicator/binary variable.
- Distribution of a Bernoulli r.v. entirely determined by p.
    - Infinite number of possible Bernoulli r.v.s: one for each value of p.

Why random variables?

- Why go through the trouble of defining random variables?
  - Allows us to think about the uncertainty of our estimates.
  - Before analyzing sample means useful to detour into sample sums.
- Extremely small data example: sample two senators with replacement.
  - $X_1 = 1$ if senator 1 is a woman, $X_1 = 0$ otherwise
  - $X_2 = 1$ if senator 2 is a woman, $X_2 = 0$ otherwise
- Define the sum of these: $S = X_1 + X_2$ (also an r.v.)
- What sums should we expect to see?
  - How surprised should we be if $S = 1$?
  - What is the probability of each possible value, $\mathbb{P}(S = k)$

Binomial distribution

- **Binomial r.v.**: X takes on any integer between 0 and n.
  - Number of heads in n independent coin flips with probability p of heads.
  - "Binomial with n trials and probability of success p''
- Example: random draws of two senators, Y is how many are women?
  - Binomial with $n = 2$ and $p = 0.26$.
- **Probability mass function** gives the probability of any possible value:

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

Binomial distribution (n = 10, p = 0.5)

More calls to senators

You work as a lobbyist and you've been asked to check to see the gender balance of the calls placed to Senate offices from your firm. The firm has placed 1000 calls over the last year. If the firm was randomly choosing senators (with replacement) each call, what numbers of women senators contacted would be more or less plausible?

- That math formula for $\mathbb{P}(X = k)$ looked not very fun...
- We can **simulate** data from this distribution using `rbinom()`.

```
rbinom(n = 5, size = 1000, prob = 0.26)
```

```
## [1] 269 247 259 268 266
```

Simulations

```
sims <- 10000

draws <- rbinom(sims, size = 1000, prob = 0.26)
```

```
length(draws)
```

```
## [1] 10000
```

```
mean(draws)
```

```
## [1] 260.0883
```

Simulations