Point Estimation and Sampling Distribution

Miaoyan Wang

Department of Statistics
UW Madison

Topics

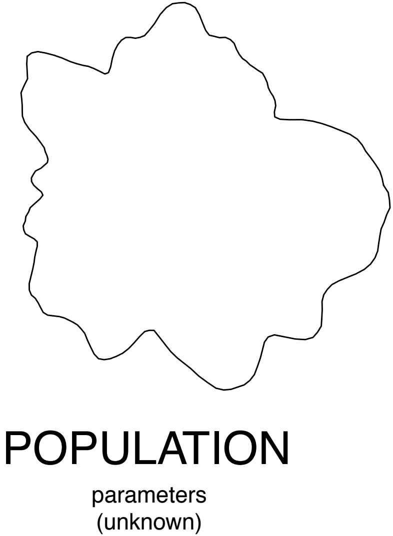
- Sample and Population
- Point estimation
  - Methods of Moments
  - Maximum Likelihood Estimation (next lecture)
- Sampling Distribution

Introduction

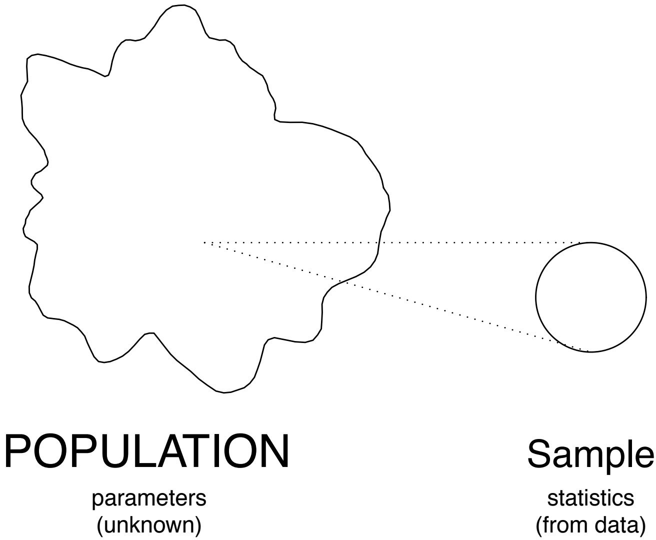
- A demanding course. Requires a general ability to do rigorous mathematical proofs and hands-on data analyses.
- Most students already have:
  - strong preparation in probability
  - linear algebra and analysis
  - basic statistics theory
  - hands on experience modeling data in R

  With hard work, you can make up one or one and a half deficits. More than that, you will feel lost.
- The course assumes a significant level of mathematical maturity. The minimal levels of math preparation can be found in the book "Plane Answers to Complex Questions" (on canvas).
  - Linear Algebra: Appendix B1-B4 and B7
  - Probability: Appendix C, D

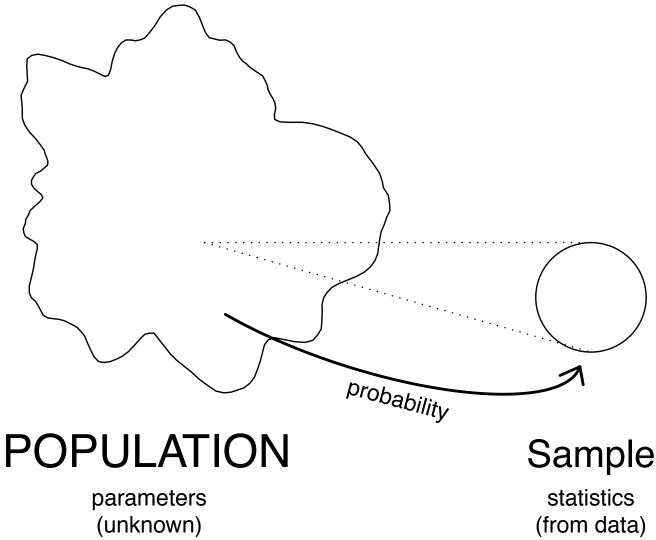
Probability vs. Statistics



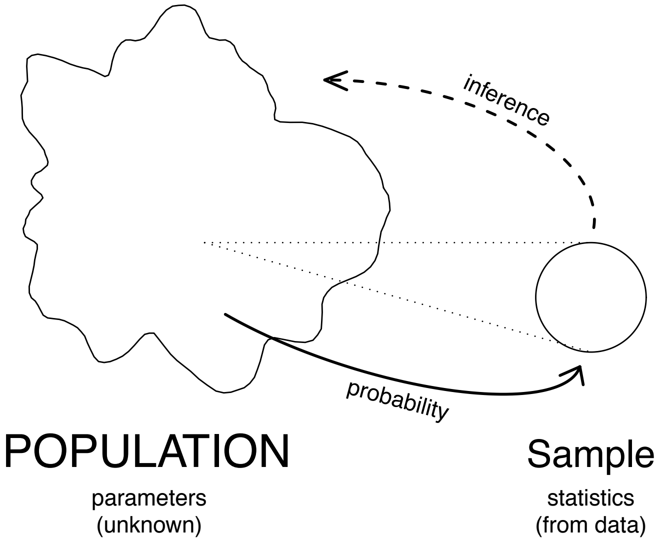
Probability vs. Statistics



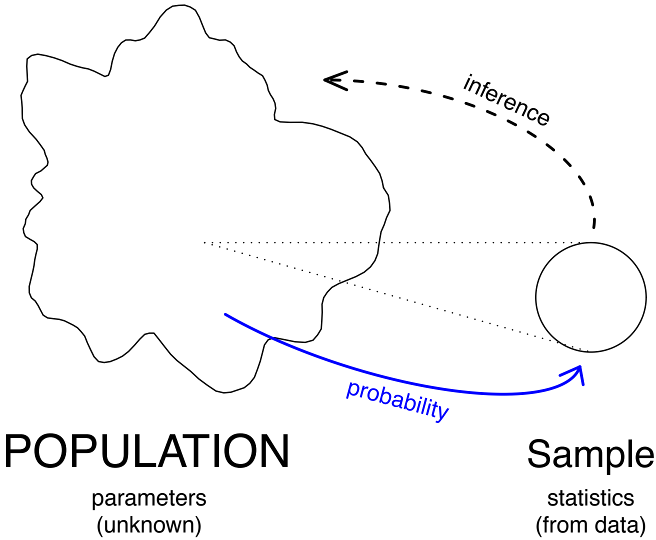
Probability vs. Statistics



Probability vs. Statistics



Probability vs. Statistics



Population vs. Sample

- Population attributes
  - $X, Y, \ldots$ (capital letters): **random variable** following some probability model or data generating process
  - $\theta, \mu, \sigma, \ldots$(Greek letters): intrinsic **population parameters** in some probability model
- Sample attributes
  - $x_1, x_2, \bar{x}, s, \ldots$ (small letters): (a function of) the **observed** values/outcome of r.v.'s in a particular data set.
  - $\hat{\theta}, \hat{\mu}, \hat{\sigma}, \ldots$("hat"): **estimated parameter/estimate** from a particular data set.
- Example: A survey conducted by a research in art education found that 17% of those surveyed had taken one course in dance in their life.
  Q: Is the number 17% a sample attribute or a population attribute? What is its standard error?

Sample and Statistics

- Let $(X_1, \ldots, X_n)$ be a random sample of size n. Any random variable $T = f(X_1, \ldots, X_n)$ as a function of $(X_1, \ldots, X_n)$ is called a **statistic**.
  - If we treat each X_i as a random variable, T is called an **estimator**.
  - If we plug X_i by the observed value from a particular sample, T is called an **estimate**.
- Dance Survey Problem: Is the 17% an estimate, estimator, or parameter? What is the statistics in this setting?
- Example:
  - The sample mean, defined by $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$, is a **statistics**.
  - The sample variance, defined by $S^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1}$, is a **statistics**.
  - Why capital letter?
- Note: A statistic/estimator/estimate cannot involve any unknown parameter in its expression. For example, $\bar{X} - \mu$ is not a statistic if the population mean μ is unknown.

Sample and Statistics

- Key: A statistic (a function from sample) can be viewed as a **random variable** varying from sample to sample.
- How to infer the **population attributes (parameter)** from the sample statistic?
- Point estimate
  - Objective: obtain an "good" guess of a population parameter.
  - Methods: methods of moment, least sum of squares, MLE, etc.
- Interval estimate.
  - Objective: obtain an "good" interval in which the population parameter will most likely lie on.
  - Methods: Sampling distribution of statistics.

Point estimation (Method of moments)

- Use the data you have to calculate **sample moments** or **centered sample moments**.
- To fit a certain distribution, use relation to moments formula:
  - Option 1:

$$\mathbb{E}(X^k) \text{ “=” } \hat{\mu}_k \equiv \frac{1}{n}\sum_{i=1}^{n} x_i^k$$

    where $\mathbb{E}(X^k)$ is k-th population moments and $\hat{\mu}_k$ is k-th sample moment (**from data**);
  - Option 2:

$$\mathbb{E}\left[(X - \mathbb{E}X)^k\right] \text{ “=” } \hat{\mu}'_k \equiv \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k$$

    where $\mathbb{E}\left[(X - \mathbb{E}X)^k\right]$ is k-th centered population moments and $\hat{\mu}'_k$ is k-th centered sample moment.

Example: Method of Moments

- Suppose Michale recorded the temperatures (^oF) at noon for recent 10 days

50	60	45	52	67	76	80	68	75	82

- Sample Mean: $$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 65.5$$

  Sample Variance: $$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 173.83$$

  So 2nd centered sample moment:

$$\hat{\mu}'_2 = \frac{n-1}{n}s^2 = 173.83 \times 9/10 = 156.45.$$

- Note: 2nd centered sample moment μ'_2 is different from sample variance s^2.
- Question: why $n-1$ nor n in sample variance?

Example: Method of Moments

Temperature:	50	60	45	52	67	76	80	68	75	82

- Model 1: Suppose we want to fit an i.i.d. uniform $U(a, b)$ model

$$f_X(x) = \frac{1}{b-a} \quad a \le x \le b.$$

What is the estimate of a and b?

Recall that $E(X) = \frac{(a+b)}{2}$, and $Var(X) = \frac{(b-a)^2}{12}$. Now use "relation to moment" formula

$$\frac{(a+b)}{2} = E(X) \text{“=”} \ \bar{x} = 65.6,$$

$$\frac{(b-a)^2}{12} = Var(X) \text{“=”} \ \hat{\mu}_2 = 156.45.$$

Therefore we have $\hat{a} = 43.93, \hat{b} = 87.26$.

Example

Temperature:	50	60	45	52	67	76	80	68	75	82

- Model 2: Suppose we want to fit with an i.i.d. $N(\mu, \sigma)$ model

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \left[-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 \right].$$

i.e. what is the estimate of μ and σ?

Remember $E(X) = \mu$, and $Var(X) = \sigma^2$. Now use "relation to moment" formula

$$\mu = E(X) \text{"="}\ \bar{x} = 65.6,$$

$$\sigma^2 = Var(X) \text{"="}\ \hat{\mu}_2 = 156.5.$$

Solving the above gives $\hat{\mu} = 65.6$ and $\hat{\sigma} = 12.6$.

Generalization: method of moments

In general, estimate m parameters, need m sample moments

Exponential-(λ) Distribution

- Sample Mean

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Population Mean

$$E(X) = 1/\lambda$$

- Parameter Estimate:

$$\hat{\lambda} = 1/\bar{x}$$

Possion(λ) Distribution

- Sample Mean

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Population Mean

$$E(X) = \lambda$$

- Parameter Estimate:

$$\hat{\lambda} = \bar{x}$$

Generalization: method of moments

Aren't there other estimators?

Exponential-(λ) Distribution

- 2^{nd} centered sample moment

$$\mu_2' \equiv \frac{n-1}{n}s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - x)^2$$

- Population Variance

$$Var(X) = 1/\lambda^2$$

- Parameter Estimate:

$$\hat{\lambda} = \sqrt{1/\mu_2'} = \sqrt{ns^2/(n-1)}$$

Poisson(λ) Distribution

- 2^{nd} centered sample moment

$$\mu_2' \equiv \frac{n-1}{n}s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - x)^2$$

- Population Variance

$$Var(X) = \lambda$$

- Parameter Estimate:

$$\hat{\lambda} = \mu_2' = \frac{n-1}{n}s^2$$

Method of Moments

- Advantages
  - Simple to calculate
  - Asymptotically normal (convergence to normal in distribution as the sample size $n \to \infty$)
- Disadvantages:
  - Inconsistent results (more than one estimator equation)
  - Could be biased

Sampling Distribution

Sampling Distribution: the probability distribution of a sample statistic under an assumed model.
Let $(X_1, X_2, \ldots, X_n)$ be an **i.i.d.** sample drawn from $N(\mu, \sigma^2)$.

Parameter (Population)	Estimator (Sample)	Distribution (do we need $n \to \infty$?)	Property
mean μ	$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$	$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \to N(0,1)$	Unbiased
variance σ^2	$\hat{\sigma}^2 (= S^2) = \frac{1}{(n-1)}\sum_{i=1}^{n}(X_i - \bar{X})^2$	$\frac{(n-1)\hat{\sigma}^2}{\sigma^2} \to \chi^2(n-1)$	Unbiased

- An estimator $\hat{\theta}$ for a parameter θ is called **unbiased estimator** if

$$\mathbb{E}(\hat{\theta}) = \theta$$

- Overestimate: $\mathbb{E}(\hat{\theta}) > \theta$; Underestiamte: $\mathbb{E}(\hat{\theta}) < \theta$.

Bias of Variance Estimator in MOM

Bias of variance estimator in MOM

Suppose $X_1, \ldots, X_n$ are i.i.d. r.v.'s sampled from $N(\mu, \sigma^2)$. Let $\hat{\sigma}_{\text{MOM}}$ be the MOM estimator of σ. Show that the $\hat{\sigma}_{\text{MOM}}$ underestimates σ^2.

Proof.

Recall that $\hat{\sigma}^2_{\text{MOM}} = \frac{1}{n}\sum_i (X_i - \bar{X})^2$, where $\bar{X} = \frac{1}{n}\sum_i X_i$. Then under the model $X_i \sim_{\text{i.i.d}} N(\mu, \sigma^2)$, we have

$$\mathbb{E}(n\hat{\sigma}^2_{\text{MOM}}) = \sum_i X_i^2 - n(\bar{X})^2 = n(\mu^2 + \sigma^2) - n(\mu^2 + \frac{\sigma^2}{n}) = (n-1)\sigma^2.$$

Therefore, $\mathbb{E}(\hat{\sigma}^2_{\text{MOM}}) = \frac{n-1}{n}\sigma^2 \leq \sigma^2$. □

Exercise: Prove that $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$.
What is the statistical implication? How is it different from $\text{Var}(X_i) = \sigma^2$, for all $i = 1, \ldots, n$.

Sampling Distribution of Estimators/Statistics

In general, let $\hat{\theta}$ be an estimator. How to find its bias?

- Express the estimator $\hat{\theta}$ as a function of sample $(X_1, ..., X_n)$. (hint: don't plug in the numerical value associated with a particular sample.)
- Treat each component $X_1, ..., X_n$ as **a random variable** with the **population** distribution.
- Use the properties of expectation and variance to calculate the expectation of $\hat{\theta}$.
- Compare $\mathbb{E}(\hat{\theta})$ with the real population parameter θ.