X. Point Estimation

Instructor: **Yanlin Qi**

Institute of Transportation Studies
Department of Statistics
University of California, Davis

Point Estimation

Scenario change

- We have completed the probability portion of the course:
  - distributions of discrete random variables
  - distributions of continuous random variables
  - joint distributions of two (discrete) random variables
  - sampling distributions of statistics

Point Estimation

Previous settings

we assumed that we had full knowledge of the distribution, including both:

- **The distribution type**
- **The values of the associated parameters**
  - Examples
    - $Bernoulli(0.5)$
    - $Pois(2.2)$
    - $N(65, 22)$
    - $Exp(\frac{1}{15})$

Point Estimation

Practical Settings

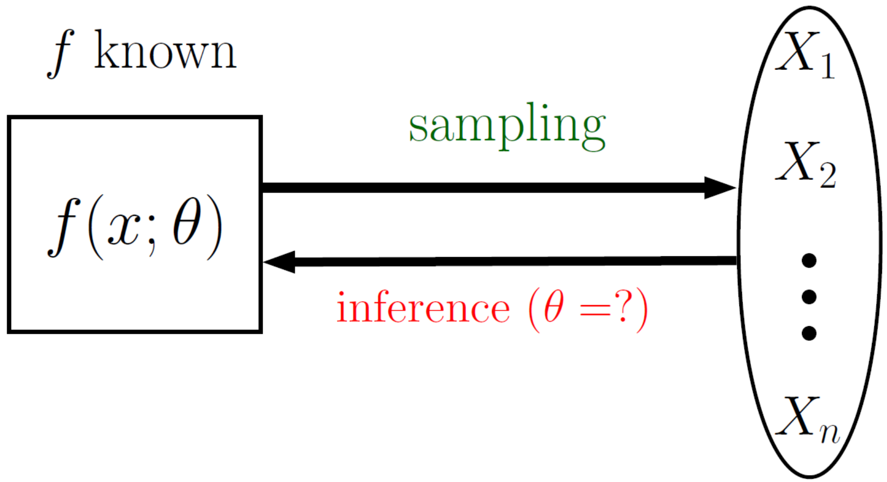
- Typically, we only know the type of the distribution for the population or can make a reasonable assumption about it
- However, we do not know the values of its parameters.

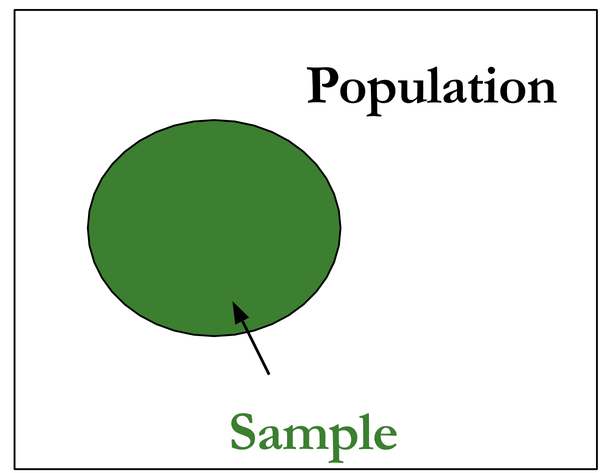
Challenges

- In most cases, it is often too difficult or expensive to study the whole population to find the exact value of a distribution parameter.

Point Estimation

- A more efficient way is to use a sample from the population to infer about the population parameters. This is called statistical inference.





Point Estimation

For example, in the brown egg problem, we **only know (or can assume)** that the weights of all the brown eggs produced at the farm (population) **follow a normal distribution** (this is our model).

We will **<u>need to determine</u>** the values of its parameters μ (mean weight) and σ^2 (variance).

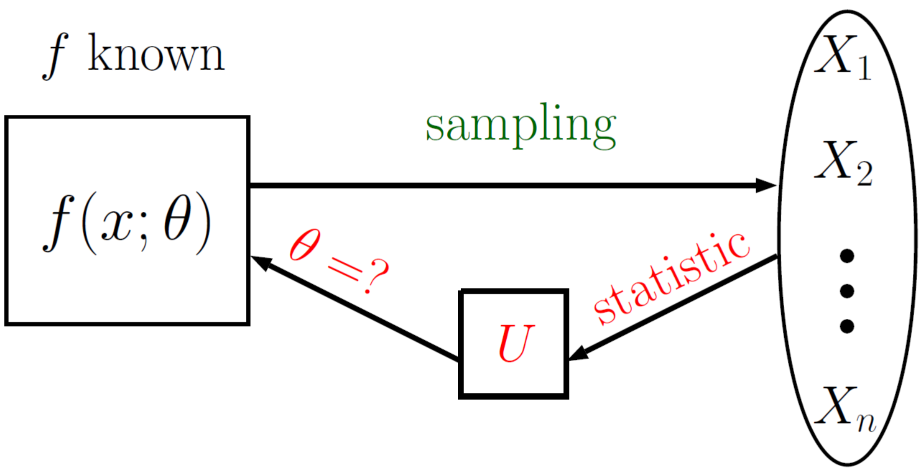
<u>Inference</u> about <u>the population mean μ</u> and <u>the variance σ^2</u> can be made based on a random sample $X_1, \ldots, X_n$ from the distribution (e.g., weights of a carton of eggs selected from the population).

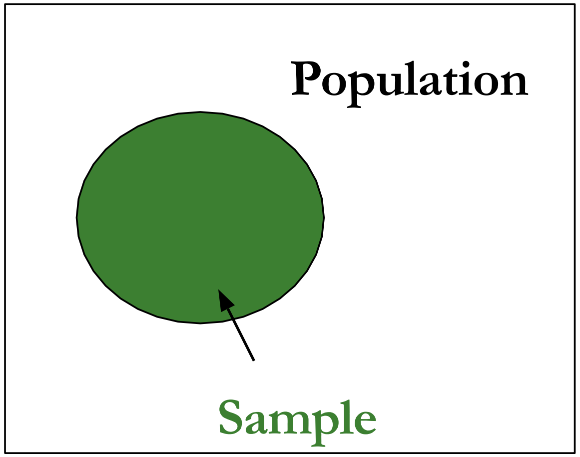
Point Estimation

- We may consider **<u>three</u>** different kinds of inference tasks:
  - **Point estimation:** What is the single (best) guess of the population mean μ?
  - **Interval estimation:** In what interval (range) does μ lie "with high probability"?
  - **Hypothesis testing:** The label says $\mu = 65$ g, but the average weight of the eggs in a randomly selected carton is only 63.9 g. Is this a contradiction?

Point Estimation

- For each task, inference will be performed through a statistic:





Point Estimation

- Consider the brown egg example again.

- **Example 0.1.** Suppose the weights of the 12 eggs in a selected carton are

$$x_1 = 63.3,\ x_2 = 63.4,\ x_3 = 64.0,\ x_4 = 63.0,\ x_5 = 70.4,\ x_6 = 65.7,$$

$$x_7 = 63.7,\ x_8 = 65.8,\ x_9 = 67.5,\ x_{10} = 66.4,\ x_{11} = 66.8,\ x_{12} = 66.0$$

Obviously, one can use the sample mean $\bar{x} = 65.5$ g as a reasonable guess of the population mean μ.

- We say that $\bar{x} = 65.5$ g is a **point estimate** of μ.
- However, point estimates will likely vary from sample to sample.

Point Estimation

- To study such randomness, we need to consider a random sample $X_1, \ldots, X_{12}$ from the population and examine the associated statistic:

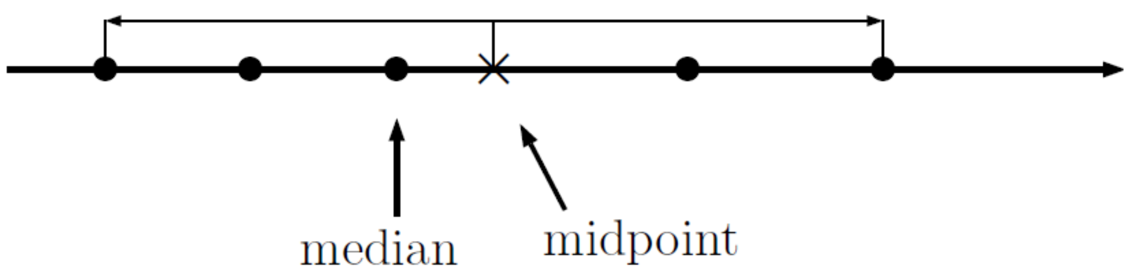
$$\bar{X} = \frac{1}{12}\sum_{i=1}^{12} X_i.$$

- The statistic $\bar{X}$ is called a **point estimator** of μ.

  - Note that a point estimator is a random variable (also a statistic) while a point estimate is an observed value of the point estimator (obtained through a realization of the sampling process).

Point Estimation

- **Question**. Are there other estimators for μ in the brown egg example and what are the corresponding point estimates (based on the same sample)?
- Sample median $\tilde{X}$. Point estimate is $\tilde{x} = \frac{65.7+65.8}{2} = 65.75$
- Midpoint of the range M . Point estimate is $m = \frac{63.0+70.4}{2} = 66.7$



- **Conclusion**: <u>Point estimators of μ are not unique.</u>

  → Follow-up question: Which one is the best?

Point Estimation

- **General definition**

- More generally, consider a distribution $f\,(x;\theta)$ with known type f but unknown parameter value θ. For example,

  - f is the normal PDF and θ represents μ (assuming σ^2 known);

  - f is the Poisson PMF and θ is the associated parameter λ;

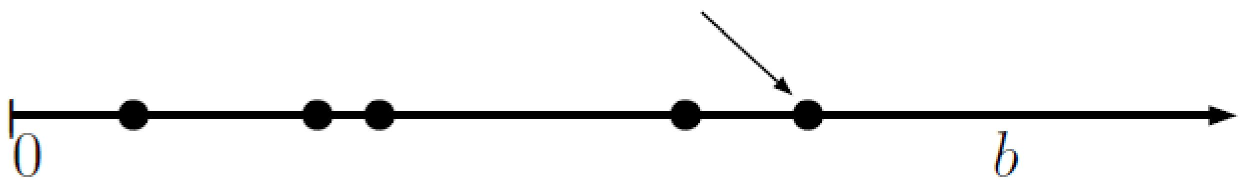
- **Def 0.1.** A **point estimator** $\hat{\theta}$ of θ is any (reasonable) statistic that is used to estimate θ.

- For any specific realization of the random sample, the corresponding value of $\hat{\theta}$ is called a **point estimate** of θ.

Point Estimation

- Example 0.2. Suppose we draw a random sample $X_1, \ldots, X_n$ from the uniform distribution Uniform(0, *b*). Then the sample maximum

$$X_{\max} = \max_{1 \le i \le n} X_i$$



can be used as a point estimator for *b*.

Point Estimation

- **Follow-up question.** Is there another statistic that may be used to estimate the unknown parameter b in the preceding example?

Point Estimation

❖ **New question.** Given a random sample $X_1, ..., X_n$ from a population with unknown variance σ^2, what estimators can we use for σ^2?

- The sample variance is the most common point estimator:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

- Another possibility is to use population variance estimator**:**

$$S'^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{n-1}{n}S^2$$

Point Estimation

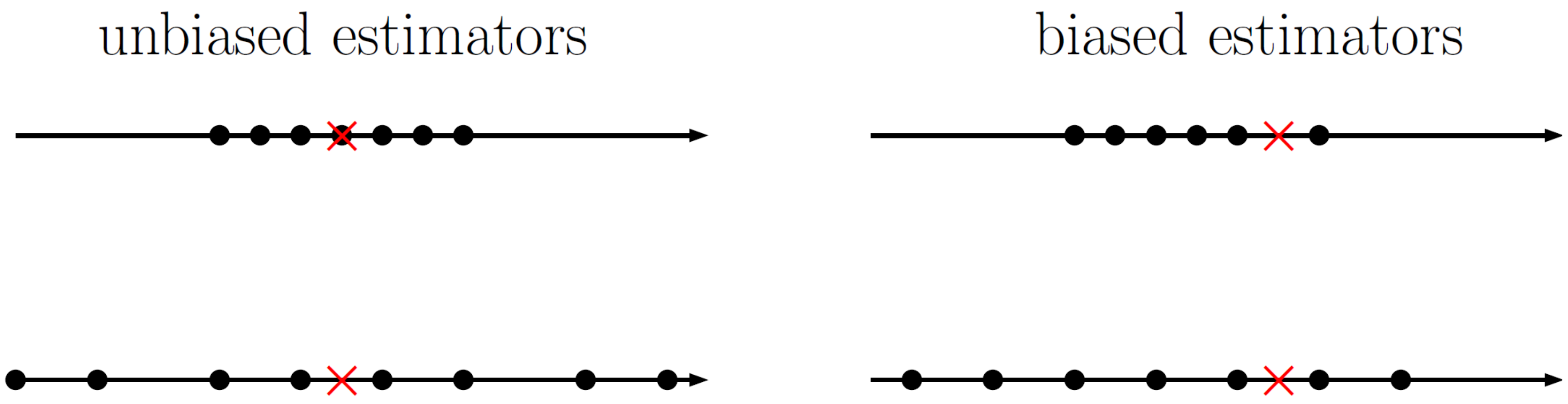
- **Example 0.3.** Given the same sample from before,

$$x_1 = 63.3, x_2 = 63.4, x_3 = 64.0, x_4 = 63.0, x_5 = 70.4, x_6 = 65.7,$$
$$x_7 = 63.7, x_8 = 65.8, x_9 = 67.5, x_{10} = 66.4, x_{11} = 66.8, x_{12} = 66.0$$

- A point estimate of σ^2 based on S^2 is $s^2 = 4.72$. In contrast, $s'^2 = 4.32$.

Evaluation of estimators

- The best estimators are unbiased and have least possible variance.



Point Estimation

- **Def 0.2.** A point estimator $\hat{\theta}$ of θ is said to be unbiased if

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

- It does not systematically overestimate or underestimate the true value

- Otherwise, it is biased and the bias of θ is defined as

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta.$$

Point Estimation

- ***Theorem* 0.1.**
  - Let $X_1, \ldots, X_n \overset{iid}{\sim} f(x)$ with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$ for all $1 \leq i \leq n$. The statistics $\bar{X}, S^2$ are always unbiased estimators of μ, σ^2 respectively.

- *Proof*. The $\bar{X}$ part directly follows from a previous sampling result:

$$E(\bar{X}) = \mu.$$

Point Estimation

- The variance part can be proved based on the following identity

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right]$$

- That is,

$$\begin{aligned}
\mathrm{E}(S^2) &= \frac{1}{n-1}\left[\sum_{i=1}^{n}\mathrm{E}(X_i^2) - n\mathrm{E}(\bar{X}^2)\right] \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n}(\mu^2+\sigma^2) - n(\mu^2 + \frac{\sigma^2}{n})\right] \\
&= \frac{1}{n-1}\left[n(\mu^2+\sigma^2) - (n\mu^2+\sigma^2)\right] \\
&= \sigma^2
\end{aligned}$$

- (In the above we have used the formula $E(Y^2) = E(Y)^2 + Var(Y)$ for any random variable Y).

Point Estimation

❖ *Remark*. The theorem implies that S'^2 is a biased estimator of σ^2:

$$E(S'^2) = E\left(\frac{n-1}{n}S^2\right) = \frac{n-1}{n}\sigma^2 \neq \sigma^2$$

and the bias is

$$B(S'^2) = E(S'^2) - \sigma^2 = -\frac{1}{n}\sigma 2$$

That is, S'^2 tends to underestimate σ^2.

Point Estimation

❖ *Remark*. Note that μ may represent <u>different parameters</u> for <u>different populations</u>:

- Normal: $\bar{X}$ is an unbiased estimator of μ;
- Bernoulli: $\bar{X}$ is an unbiased estimator of p;
- Poisson: $\bar{X}$ is an unbiased estimator of λ;
- Uniform$(0, b)$: $\bar{X}$ is an unbiased estimator of $\frac{b}{2}$, which implies that $2\bar{X}$ unbiased estimator of b.

Point Estimation

- **Example 0.4.** For a random sample of size *n* from the $\underline{\boldsymbol{Uniform(0,b)}}$ distribution (**<u>where b is unknown</u>**), it can be shown that the sample maximum is a biased estimator of b:

$$\mathrm{E}(X_{\max}) = \frac{n}{n+1}b$$

with negative bias

$$B(X_{\max}) = \mathrm{E}(X_{\max}) - b = \frac{n}{n+1}b - b = -\frac{1}{n+1}b$$

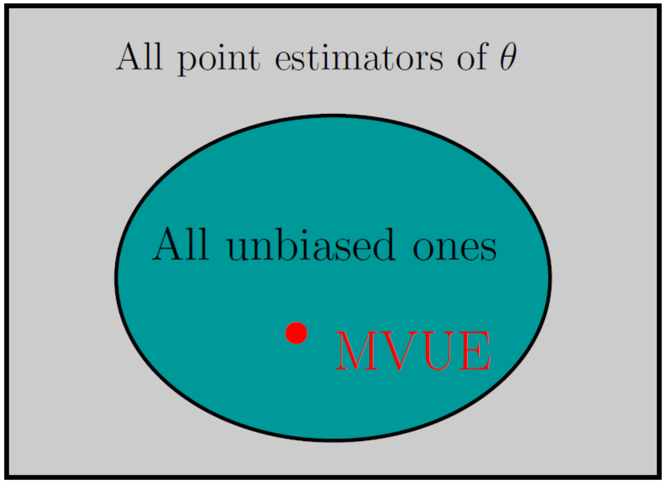
However, $\frac{n+1}{n}X_{\max}$ is an unbiased estimator of

$$\mathrm{E}\left(\frac{n+1}{n}X_{\max}\right) = \frac{n+1}{n}\mathrm{E}\left(X_{\max}\right) = \frac{n+1}{n}\cdot\frac{n}{n+1}b = b$$

(Recall that $2\bar{X}$ is another unbiased estimator of b).

Point Estimation

- Between two unbiased estimators (of some parameter), the one with smaller variance is better.
- **Def 0.3.** The unbiased estimator $\hat{\theta}^*$ of θ that has the smallest variance is called a minimum variance unbiased estimator (MVUE).



- *Theorem* 0.2. For normal populations, $\bar{X}$ is a MVUE for μ.

Methods of Point Estimation

Method of Maximum Likelihood

1. One of the best methods to obtain a point estimator
2. Suppose that X is a RV with PDF $f(x;\theta)$, where θ is a single unknown parameter
3. Let $x_1, x_2, \ldots, x_n$ be the observed values in a RS of size n
4. The likelihood function (LF) of the RS is

$$L(\theta) = f(x_1;\theta) \cdot f(x_2;\theta) \cdot \cdots \cdot f(x_n;\theta)$$

   - Note that the LF is now a function of only θ

5. The maximum likelihood estimator (MLE) of θ is:
   - The value that **maximizes** the LF $L(\theta)$

Normal Distribution MLE for μ

1. Let X be normally distributed with unknown mean μ and known variance σ^2
2. The likelihood function for a RS of size n is

$$L(\mu) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i-\mu)^2/(2\sigma^2)} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\sum_{i=1}^{n}(x_i-\mu)^2/(2\sigma^2)}$$

3. Now, taking the logarithm and then the derivative

$$\ln L(\mu) = -\frac{n}{2}\ln(2\pi\sigma^2) - (2\sigma^2)^{-1}\sum_{i=1}^{n}(x_i-\mu)^2 \qquad \frac{d\ln L(\mu)}{d\mu} = (\sigma^2)^{-1}\sum_{i=1}^{n}(x_i-\mu)$$

4. Equating this last result to zero and solving for μ yields

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{X}$$

Normal Distribution MLEs for μ and σ^2

- Let X be normally distributed with unknown mean μ and unknown variance σ^2
- The likelihood function for a RS of size n and its logarithms are the same as in the previous case
- Now, by taking the derivatives and equating them to zero

$$\frac{\partial \ln L(\mu,\sigma^2)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)=0 \qquad \frac{\partial \ln L(\mu\sigma^2)}{\partial(\sigma^2)} = -\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2=0$$

- Solving for μ and σ^2 yields

$$\hat{\mu}=\overline{X} \qquad \hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$$

Notice this is not unbiased

Properties of a MLE

1. **General Conditions:** Under very general and not restrictive conditions, when the sample size n is large and if $\hat{\theta}$ is the MLE of θ
2. **Approximately Unbiased:** $\hat{\theta}$ is an approximately unbiased estimator for θ
3. **Minimum Variance:** The variance of $\hat{\theta}$ is nearly as small as the variance that could be obtained with any other estimator
4. **Normal Distribution:** $\hat{\theta}$ has an approximate normal distribution
5. **PDF Requirement:** To use the ML estimation, the PDF of the population must be either known or assumed

Summary

Summary

Assume a distribution $f(x)$ with an unknown parameter θ and a random sample $X_1, \ldots, X_n$ from this population.

- **Basic concepts**
  - **Point estimator:** a *statistic* used to estimate the parameter θ, denoted as $\hat{\theta}$. The observed value of $\hat{\theta}$ corresponding to a specific sample is called a **point estimate**.

Summary

- **Unbiasedness**:
- $\hat{\theta}$ is unbiased if $E(\hat{\theta}) = \theta$. Otherwise, the bias is $B(\hat{\theta}) = E(\hat{\theta}) - \theta$.
- When two estimators $\widehat{\theta}_1$, $\widehat{\theta}_2$ are both unbiased, we prefer the one with smaller variance.
- The unbiased estimator $\hat{\theta}^*$ with the smallest variance is called a minimum variance unbiased estimator (MVUE) for θ.

Summary

Important results

- **Sample mean $\bar{X} = \frac{1}{n}\sum X_i$ is always unbiased** (as an estimator for population mean μ). For example,
  - For Normal populations $\mathrm{N}(\mu, \delta^2), \overline{\mathrm{X}}$ is unbiased for μ;

    For $\mathrm{N}(\mu, \delta^2)$, also has the smallest variance (among all unbiased estimators) and thus is a MVUE for μ

  - For Poisson populations $\mathrm{Pois}(\lambda)$, $\overline{\mathrm{X}}$ is unbiased for λ;
  - For Uniform distributions $\mathrm{Unif}(0, \theta)$, $\overline{\mathrm{X}}$ is unbiased for $\frac{\theta}{2}$

Summary

- **Sample variance** $S^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$ **is always unbiased** (as an estimator for population variance σ^2). For example:
  - For Normal populations $\mathrm{N}(\mu, \delta^2)$, S^2 is unbiased for δ^2;
  - For Poisson populations $\mathrm{Pois}(\lambda)$, S^2 is unbiased for λ;

Note that $S'^2 = \frac{1}{n}\sum(X_i - \bar{X})^2$ is always a biased estimator for δ^2, the bias is $B(S'^2) = -\frac{1}{n}\delta^2$