San José State University

Math 161A: Applied Probability & Statistics I

Point Estimation

Prof. Guangliang Chen

Section 6.1 Some general concepts of point estimation

Scenario change

We have completed the probability portion of the course:

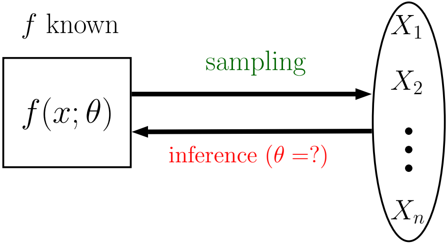
- distributions of discrete random variables (Chapter 3)
- distributions of continuous random variables (Chapter 4)
- joint distributions of two (discrete) random variables (Section 5.1)
- sampling distributions of statistics (Sections 5.3, 5.4)

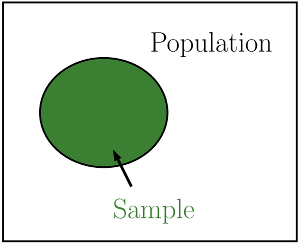
In the previous settings (which were very mathematical), we assumed that we had full knowledge of the distribution in terms of both the distribution type and the values of the associated parameters (e.g., Bernoulli(0.5), Pois(2.2), N($65, 2^2$), Exp($\frac{1}{45}$)).

In practical settings we usually only know the type of the distribution for the population (or can make a reasonable assumption about the distribution type), but not the values of its parameters.

In most cases, it is too difficult or expensive to access the whole population to determine the exact value of a distribution parameter.

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





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).

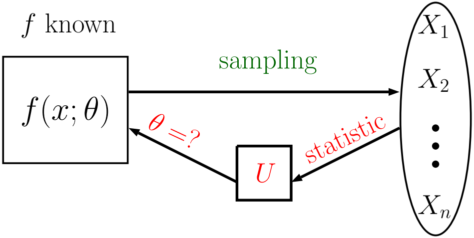
We will need to determine the values of its parameters μ (mean weight) and σ^2 (variance).

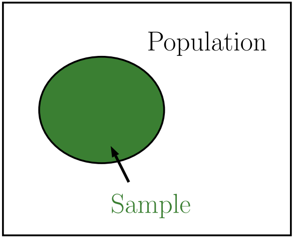
Inference about the population mean μ and the variance σ^2 can be made based on a random sample $X_1, \ldots, X_n$ from the distribution (e.g., weights of a carton of eggs selected from the population).

We may consider three 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?

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.

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

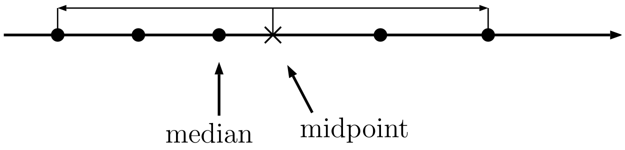
$$\bar{X} = \frac{1}{12}\sum_{i=1}^{12} X_i.$$

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).

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: Point estimators of μ are not unique.
$\longrightarrow$ Follow-up question: Which one is the best?

General definition

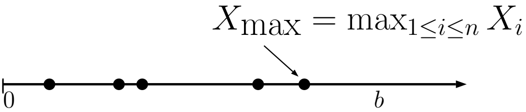
More generally, consider a distribution $f(x;\theta)$ with known type f but unknown parameter value θ. For example,

- 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.

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



can be used as a point estimator for b.

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

New question. Given a random sample $X_1, \ldots, X_n$ 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

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

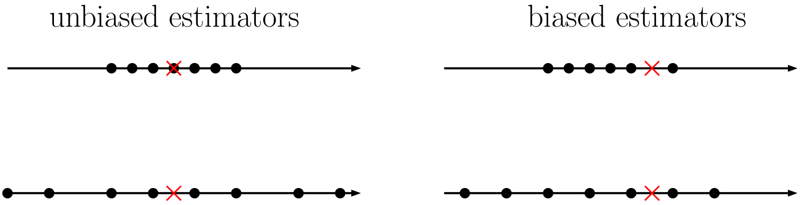
Example 0.3. Given the same sample from before,

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

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**.



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

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

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

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

Theorem 0.1. Let $X_1, \ldots, X_n \overset{iid}{\sim} f(x)$ with $\mathrm{E}(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$ for all $1 \le i \le n$. 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:

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

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 $\mathrm{E}(Y^2) = \mathrm{E}(Y)^2 + \mathrm{Var}(Y)$ for any random variable Y).

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

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

and the bias is

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

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

Remark. Note that μ may represent different parameters for different populations:

- 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 $b/2$, which implies that $2\bar{X}$ is an unbiased estimator of b.

Example 0.4. For a random sample of size n from the Unif$(0, b)$ distribution (where b is unknown), 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

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

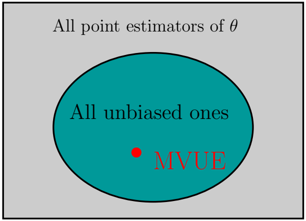
However, $\frac{n+1}{n} X_{\max}$ is an unbiased estimator of b:

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

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

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.

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.
  - Unbiasedness: $\hat{\theta}$ is unbiased if $\mathrm{E}(\hat{\theta}) = \theta$. Otherwise, the bias is $\mathrm{B}(\hat{\theta}) = \mathrm{E}(\hat{\theta}) - \theta$. When two estimators $\hat{\theta}_1, \hat{\theta}_2$ 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.

- **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 $N(\mu, \sigma^2)$, $\bar{X}$ is unbiased for μ;
    * For Poisson populations Pois(λ), $\bar{X}$ is unbiased for λ;
    * For Uniform distributions Unif$(0, \theta)$, $\bar{X}$ is unbiased for $\frac{\theta}{2}$;

In the case of a normal population $N(\mu, \sigma^2)$, $\bar{X}$ also has the smallest variance (among all unbiased estimators) and thus is a MVUE for μ.

- **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 $N(\mu, \sigma^2)$, S^2 is unbiased for σ^2;
  * For Poisson populations Pois(λ), 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 σ^2; the bias is $\mathrm{B}(S'^2) = -\frac{1}{n}\sigma^2$.