Point Estimation

– Math 161a, Spring 2019, San Jose State University

Prof. Guangliang Chen

April 25, 2019

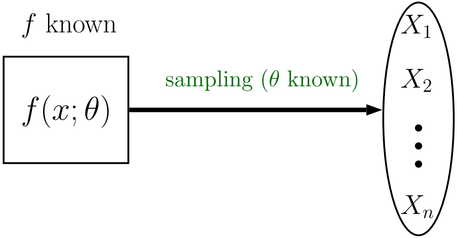
Outline

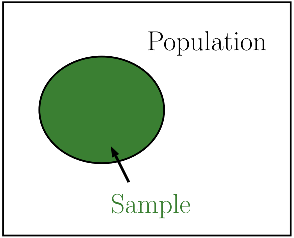
Scenario change

We have just completed the probability chapters of the course, which concern the distributions of

- Ch. 3,4 **a single random variable** (discrete or continuous),
- Sec. 5.1 **two discrete random samples jointly**
- 5.3, 5.4 **a statistic** (function of a random sample from some distribution)

In the above (theoretical) settings, we study those distributions with the knowledge of both the distribution type and the values of the associated parameters.





In the probability chapters:

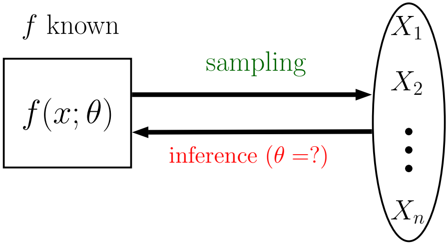
- Both the distribution type f and the value of the associated parameter θ are known (e.g., Bernoulli(p), Exp(λ), N($\mu, 2^2$))
- $n = 1$ (single random variable), $n > 1$ (random sample)

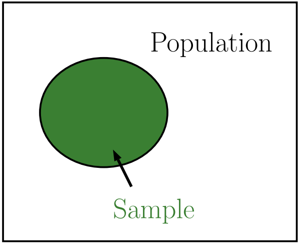
Point Estimation

In practical settings we usually only know, or can only assume, the type of the distribution for the population, but not the values of its parameters.

It is often impossible/too expensive to access the whole population.

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





For example, in the egg weight 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).

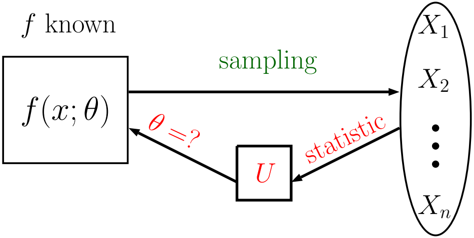
We will need to infer the values of its parameters μ (mean weight) and σ^2 (variance).

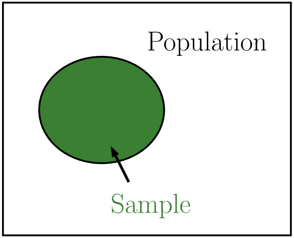
Inference about the population mean μ and the variance σ^2 can be made based on a random sample $X_1, \ldots, X_{12}$ from the distribution (e.g., weights of a carton of eggs selected from the population).

We may consider three 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 is performed through a statistic:





Point estimation

Consider the 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 single 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:

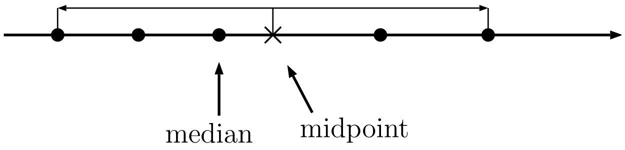
$$\bar{X} = \frac{1}{12} \sum_{i=1}^{12} X_i.$$

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

Note. Point estimator is a random variable (statistic) while point estimate is a specific number (obtained through a realization of the sampling process).

Question. Are there other estimators for μ in the 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 $\longrightarrow$ 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,

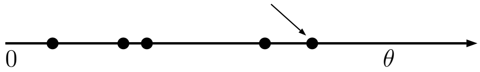
- f is the normal pdf and θ represents μ (assuming σ^2 known);
- f is Poisson pmf and θ is the parameter λ;

Definition 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$(0, \theta)$. Then the sample maximum

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



can be used as a point estimator for θ.

Question. Any other statistic may be used to estimate θ?

What estimators can we use for the population variance σ^2?

- The sample variance

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

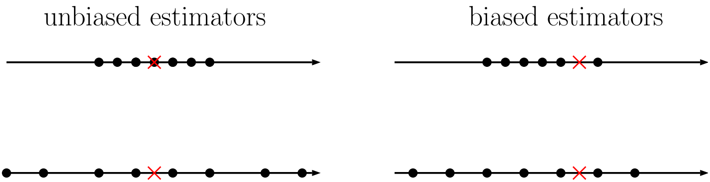
- Another possibility is to use

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

Example 0.3. In the egg example, 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**.



Definition 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. $\bar{X}, S^2$ 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).

Example 0.4. 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

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

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$(0, \theta)$: $\bar{X}$ is an unbiased estimator of $\theta/2$, which implies that $2\bar{X}$ is an unbiased estimator of θ.

Example 0.5. For a random sample of size n from the Unif$(0, \theta)$ distribution (where θ is unknown), it can be shown that the sample maximum is a biased estimator of θ:

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

with negative bias

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

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

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

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

Between two unbiased estimators (of some parameter), the one with smaller variance is better.

Definition 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.