2028A: Basic Statistical Methods
Handout - January, 30 - February, 3

STATISTICAL ESTIMATION

1. Main Concepts

- Statistical inference is almost always directed to drawing conclusions about characteristics of a population. In statistical inference, a characteristic of a population is called *parameter*. Examples of parameters are the *mean* of the population or *variability* of the population, the *proportion* of a category and so on.
- Since an investigator obtains a sample of the population, he/she needs to "guess" the parameters of interest based on the sample data. Evaluating the population parameters based on the sample data is called *estimation*.
- Since the observed sample data are observed with some uncertainty and variability, the observed values or observations, conventionally denoted with small letters $x_1, \ldots, x_n$, are assumed to come from a sample of random variables, conventionally denoted with $X_1, \ldots, X_n$.

  In order to evaluate or estimate the parameters using the sample data, we need to assume a probability distribution for the population. So we assume that the sample of r.v.'s $X_1, \ldots, X_n$ comes from a population with density function f. If $X_1, \ldots, X_n$ are independent random variables from a distribution of density or mass function f, from probability theory, we have that their joint density or mass function is:

$$f(x_1, x_2, \ldots, x_n) = f(x_1)f(x_2)\ldots f(x_n).$$

- We can evaluate the population parameters using a function of the sample of random variables. A *statistic* is a function of the observable random variables $X_1, \ldots, X_n$, which does not depend on any unknown parameter. A statistic is a random variable itself.
- Statistics are used to evaluate characteristics/parameters of a population. A *point estimate* of an unknown parameter θ is a single number that can be regarded as a sensible value for θ. A point estimate is obtained by selecting a suitable statistic and computing its value from the recorded/observed data. The selected statistic is called the *point estimator* for θ.

  For example, if the statistic $S(X_1, \ldots, X_n)$ is suitable for evaluating a parameter θ, then $S(X_1, \ldots, X_n)$ is the point estimator for θ and $S(x_1, \ldots, x_n)$ for observations $x_1, \ldots, x_n$ is the point estimate of θ. Note that $S(x_1, \ldots, x_n)$ is a value/number and $S(X_1, \ldots, X_n)$ is a random variable.

Notation: Conventionally, we denote the population parameters with Greek letters: μ denotes the population mean, σ^2 denotes the variability parameter and θ denotes a general parameter. Also we denote their point estimates or estimators by $\widehat{\mu}$ for the mean, $\widehat{\sigma}^2$ for the variability parameter and $\widehat{\theta}$ for a general parameter.

2. Examples of Statistics

   - *Sample mean* is a point estimator for the mean parameter μ:

     $$\bar{X} = \frac{X_1 + \ldots + X_n}{n}$$

     for $X_1, \ldots, X_n$ a random sample.
     The corresponding point estimate is:

     $$\bar{x} = \frac{x_1 + \ldots + x_n}{n}$$

     for $x_1, \ldots, x_n$ observed values.
     Note that $\bar{x}$ is a value/number and $\bar{X}$ is a random variable.

   - *Sample variance* is a point estimator for the variability parameter σ^2:

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

     for $X_1, \ldots, X_n$ a random sample and $\bar{X}$ its sample mean.
     The corresponding point estimate is:

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

     for $x_1, \ldots, x_n$ observed values and $\bar{x}$ its observed sample mean.

   - *Sample proportion* is a point estimator for the proportion parameter p for a category in a population. For a sample of size n, the observed frequency of a given category (e.g. female students among all students) is denoted by x and the random frequency of a given category is denoted by X. Note that the distribution of X is Binomial, $X \sim Bin(n, p)$, where p is the proportion parameter, which is unknown.
     The sample proportion is:

     $$\widehat{p} = \frac{X}{n}$$

     and the observed sample proportion is:

     $$\widehat{p} = \frac{x}{n}.$$

3. Example: Proportion Parameter

   An automobile manufacturer has developed a new type of bumper, which is supposed to absorb impacts with less damage than previous bumpers. The manufacturer has used this bumper in a sequence of 25 controlled crashes against wall, each at 10 mph, using one of its compact models. Let X be the number of crashes that result in no visible damage to the automobile. He observes that only 15 out of 25 cars have no damage after crash.

   - The manufacturer is interested in the proportion of all such crashes that result in no damage. We define this feature through the parameter p, which is the probability of no damage in a single crash.

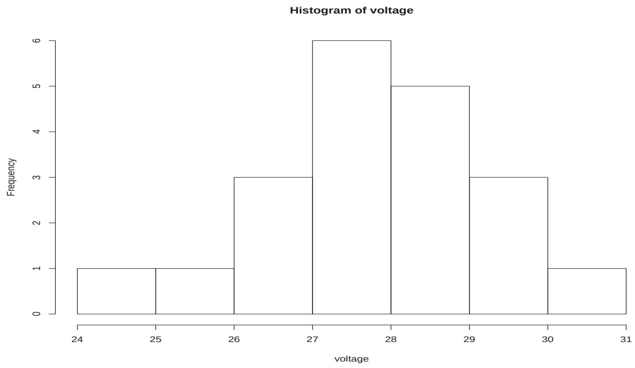

Figure 1: Histogram of the dilectic breakdown voltage.

- A *point estimator* of p is:

$$\widehat{p} = \frac{X}{n}, \ n \text{ is the sample size}$$

- X is observed to be $x = 15$.
- The corresponding *point estimate* is:

$$\widehat{p} = \frac{x}{n} = \frac{15}{25} = .60.$$

4. Example: Mean parameter

   The accompanying sample consists of $n = 20$ observations on dilectic breakdown voltage of a piece of epoxy resin.

   ```
   24.46 25.61 26.25 26.42 26.66 27.15 27.31 27.54 27.74 27.94 27.98
   28.04 28.28 28.49 28.50 28.87 29.11 29.13 29.50 20.88
   ```

   The histogram of the observations is approximately symmetric (see figure 1).

   In this example, we are interested in obtaining the average dilectic breakdown voltage of the piece of epoxy resin.

   - We denote or define the characteristic of interest μ, which is the mean of the population.
   - The sample of r.v.'s of the observed values is $X_1, X_2, \ldots, X_{20}$.
   - Examples of point estimators are:

     (a)
     $$\bar{X} = \frac{X_1 + X_2 + \ldots + X_{20}}{20}$$

     (b)
     $$\widetilde{X} = \text{median}(X_1, \ldots, X_{20})$$

(c)

$$\widehat{X} = \frac{\min_{i=1}^{20}\{X_i\} + \max_{i=1}^{20}\{X_i\}}{2}$$

- The corresponding point estimates are:

  (a)

  $$\bar{x} = \frac{x_1 + x_2 + \ldots + x_{20}}{20} = \frac{555.86}{20} = 27.793$$

  (b)

  $$\widetilde{x} = \text{median}(x_1, \ldots, x_{20}) = \frac{27.94 + 27.98}{2} = 27.960$$

  (c)

  $$\widehat{x} = \frac{\min_{i=1}^{20}\{x_i\} + \max_{i=1}^{20}\{x_i\}}{2} = \frac{24.46 + 30.88}{2} = 27.670$$

  Note: Each of the estimators uses a different measure of the center of the sample to estimate μ. Which of the estimates is closest to the true value? We cannot answer this without knowing the true value.

5. Example: Variability parameter

   For the dilectic breakdown voltage data, we would like to estimate the variability parameter.

   - We denote or define the characteristic of interest σ^2, which is the variance of the population.
   - The sample of r.v.'s of the observed values is $X_1, X_2, \ldots, X_{20}$.
   - Examples of point estimators are:

     (a)

     $$S^2 = \frac{\sum_{i=1}^{20}(X_i - \bar{X})^2}{19}$$

     (b)

     $$IQR = Q2(X_1, \ldots, X_{20}) - Q1(X_1, \ldots, X_{20})$$

     where $Q2(X_1, \ldots, X_{20})$ and $Q1(X_1, \ldots, X_{20})$ are the upper and, respectively, the lower quartiles of the sample.

   - The corresponding point estimates are:

     (a)

     $$s^2 = \frac{\sum_{i=1}^{20}(x_i - \bar{x})^2}{19} = \frac{\sum_{i=1}^{20} x_i^2 - 20\bar{x}^2}{19} \frac{14972.02 - 14898.16}{19} = 3.887$$

     (b)

     $$IQR = Q2(x_1, \ldots, x_{20}) - Q1(x_1, \ldots, x_{20}) = 28.4925 - 26.6 = 1.89.$$

6. Evaluation of Estimators

In the best of all possible worlds, we could find an estimator $\widehat{\theta}$ such that $\widehat{\theta} = \theta$. However, $\widehat{\theta}$ is a random variable so for some samples $\widehat{\theta}$ will yield a value larger than θ, whereas for other samples $\widehat{\theta}$ will yield a value smaller than θ. Therefore, we write:

$$\widehat{\theta} = \theta + \text{ error of estimation.}$$

We want to find/construct $\widehat{\theta}$ such that the error of estimation (due to sample variability and uncertainty) is small. We define "small" throughout two different concepts:

(a) We define $\widehat{\theta}$ an *unbiased* estimator of θ if:

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

for every possible value of θ. That is, the distribution of $\widehat{\theta}$ is centered at θ. The difference

$$\mathbb{E}(\widehat{\theta}) - \theta$$

is called *bias* of $\widehat{\theta}$.

If two different point estimators are being compared, we will prefer the one with smaller absolute bias. So if $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are two estimators of θ then:

$$\begin{array}{l} \text{Prefer } \widehat{\theta}_1 \text{ if } |\mathbb{E}(\widehat{\theta}_1) - \theta| < |\mathbb{E}(\widehat{\theta}_2) - \theta| \\ \text{Prefer } \widehat{\theta}_2 \text{ if } |\mathbb{E}(\widehat{\theta}_1) - \theta| > |\mathbb{E}(\widehat{\theta}_2) - \theta| \end{array}$$

<u>Example:</u>

Take two instruments measuring blood pressure. One of them has been calibrated, where the second one gives values larger than the true ones. For one individual, we measure the blood pressure for 10 times: $x_1, \ldots, x_{10}$ with the first instrument and $y_1, \ldots, y_{10}$ with the second instrument. Say θ is the blood pressure of the same individual. If we estimate θ as:

$$\bar{x} = \frac{x_1 + \ldots + x_{10}}{10}$$

based on the measurements of the first instrument and

$$\bar{y} = \frac{y_1 + \ldots + y_{10}}{10}$$

based on the measurements of the second instrument then $\bar{x} \approx \theta$ but $\bar{y} > \theta$. So the second instrument yields an estimate of θ that have systematic error component or bias given by $\bar{y} - \theta > 0$.

(b) Suppose $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are two unbiased estimators of θ. Which one to choose? Even though the distribution of each estimator is centered at the true value of θ, the spreads of the distributions about the true value θ may be different.

Among all unbiased estimators of θ, we then choose the one that has minimum variance. The resulting estimator is called *minimum variance unbiased estimator* (MVUE) of θ.

(c) What if we compare two estimators with positive bias? Say we compare two estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ of θ such that:

$$\mathbb{E}(\widehat{\theta}_1) = \mu_1,\ \mathbb{E}(\widehat{\theta}_2) = \mu_2$$
$$\mathbb{V}(\widehat{\theta}_1) = \sigma_1^2,\ \mathbb{V}(\widehat{\theta}_2) = \sigma_2^2.$$

We want to compare both the bias and the variance simultaneously. For that we define the *mean square error*:

$$MSE(\widehat{\theta}) = \mathbb{E}(\widehat{\theta} - \theta)^2 = \mathbb{V}(\widehat{\theta}) + (\mathbb{E}(\widehat{\theta}) - \theta)^2.$$

Therefore, we compute $MSE(\widehat{\theta}_1)$ and $MSE(\widehat{\theta}_2)$, and choose the estimator with the smallest mean square error.

7. Example: Unbiased estimators

(a) *Sample proportion*

For $X \sim Bin(n, p)$, we estimate the success probability p as:

$$\widehat{p} = \frac{X}{n}, \text{ given } n.$$

Is this estimator unbiased for p?

$$\mathbb{E}[\widehat{p}] = \mathbb{E}[\frac{X}{n}] = \frac{1}{n}\mathbb{E}[X] = \frac{1}{n}(np) = p$$

Therefore, no matter what the true value of p is, the distribution of the estimator $\widehat{v}p$ will be centered at the true value.

(b) *Sample mean*

Take $X_1, \ldots, X_n$ from some distribution of mean μ and variance σ^2. That is, $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2$.

A point estimator of μ is:

$$\widehat{\mu} = \bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

Is this estimator unbiased for μ?

$$\mathbb{E}[\widehat{\mu}] = \mathbb{E}[\tfrac{X_1+X_2+\ldots+X_n}{n}] =$$
$$\tfrac{\mathbb{E}[X_1]+\mathbb{E}[X_2]+\ldots+\mathbb{E}[X_n]}{n} = \tfrac{n\mu}{n} = \mu$$

(c) *Sample variance*

Take $X_1, \ldots, X_n$ from some distribution of mean μ and variance σ^2. That is, $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2$.

A point estimator of σ^2 is:

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

A point estimator of σ is:

$$\widehat{\sigma} = S = \sqrt{\frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1}}$$

Is $\widehat{\sigma}^2$ estimator unbiased for σ^2?

$$\mathbb{E}[\widehat{\sigma}^2] = \mathbb{E}\left[\tfrac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1}\right] =$$
$$\tfrac{1}{n-1}\left(\sum_{i=1}^n \mathbb{E}[X_i^2] - n\mathbb{E}\left[\tfrac{(\sum_{i=1}^n X_i)^2}{n^2}\right]\right) =$$
$$\tfrac{1}{n-1}\left(\sum_{i=1}^n (\mathbb{V}[X_i] + \mathbb{E}[X_i]^2) - n\mathbb{E}\left[\tfrac{\mathbb{V}(\sum_{i=1}^n X_i) + (\mathbb{E}(\sum_{i=1}^n X_i))^2}{n^2}\right]\right) =$$
$$\tfrac{1}{n-1}\left(\sum_{i=1}^n (\sigma^2 + \mu^2) - \mathbb{E}\left[\tfrac{n\sigma^2 + (n\mu)^2}{n}\right]\right) = \tfrac{1}{n-1}(n\sigma^2 - \sigma^2) = \sigma^2$$

8. Example: Mean Square Error

Suppose $\mathbb{E}[X_1] = \mathbb{E}[X_2] = \mathbb{E}[X_3] = \mu$ and $\mathbb{V}[X_1] = 7$, $\mathbb{V}[X_2] = 13$, $\mathbb{V}[X_1] = 20$. Consider the point estimates of μ:

- $$\widehat{\mu}_1 = \frac{X_1 + X_2 + X_3}{3}$$
- $$\widehat{\mu}_2 = \frac{X_1}{4} + \frac{X_2}{3} + \frac{X_3}{5}$$
- $$\widehat{\mu}_3 = \frac{X_1}{6} + \frac{X_2}{3} + \frac{X_3}{4} + 2$$

(a) Calculate the bias of each point estimate. Is any one of them unbiased?

- $$\mathbb{E}[\widehat{\mu}_1] = \frac{\mathbb{E}[X_1] + \mathbb{E}[X_2] + \mathbb{E}[X_3]}{3} = \mu$$

  Therefore, the bias is 0.
- $$\mathbb{E}[\widehat{\mu}_2] = \frac{\mathbb{E}[X_1]}{4} + \frac{\mathbb{E}[X_2]}{3} + \frac{\mathbb{E}[X_3]}{5} = \frac{\mu}{4} + \frac{\mu}{3} + \frac{\mu}{5} = 1.217\mu$$

  Therefore, the absolute bias is $.217\mu$.

- $$\mathbb{E}[\widehat{\mu}_3] = \frac{\mathbb{E}[X_1]}{6} + \frac{\mathbb{E}[X_2]}{3} + \frac{\mathbb{E}[X_3]}{4} + 2 = \frac{\mu}{6} + \frac{\mu}{3} + \frac{\mu}{4} + 2$$

  Therefore, the bias is $2 - \frac{\mu}{4}$.

(b) Calculate the variance of each point estimate. Which one has the smallest variance?

- $$\mathbb{V}[\widehat{\mu}_1] = \frac{\mathbb{V}[X_1] + \mathbb{V}[X_2] + \mathbb{V}[X_3]}{3^2} = \frac{7 + 13 + 20}{3^2} = 4.444$$
- $$\mathbb{V}[\widehat{\mu}_2] = \frac{\mathbb{V}[X_1]}{4^2} + \frac{\mathbb{V}[X_2]}{3^2} + \frac{\mathbb{V}[X_3]}{5^2} = \frac{7}{4^2} + \frac{13}{3^2} + \frac{20}{5^2} = 2.682$$
- $$\mathbb{V}[\widehat{\mu}_3] = \frac{\mathbb{V}[X_1]}{6^2} + \frac{\mathbb{V}[X_2]}{3^2} + \frac{\mathbb{V}[X_3]}{4^2} = \frac{7}{6^2} + \frac{13}{3^2} + \frac{20}{4^2} = 2.889$$

Therefore, μ_2 has the smallest variance.

(c) Which point estimate has the smallest mean square error when $\mu = 3$?

- $$\text{MSE}[\widehat{\mu}_1] = \mathbb{V}[\widehat{\mu}_1] + (\mathbb{E}[\widehat{\mu}_1] - \mu)^2 = 4.444 + (3 - 3)^2 = 4.444$$
- $$\text{MSE}[\widehat{\mu}_2] = \mathbb{V}[\widehat{\mu}_2] + (\mathbb{E}[\widehat{\mu}_2] - \mu)^2 = 2.682 + (0.217 \times 3)^2 = 3.104$$
- $$\text{MSE}[\widehat{\mu}_3] = \mathbb{V}[\widehat{\mu}_3] + (\mathbb{E}[\widehat{\mu}_3] - \mu)^2 = 2.889 + (2 - \frac{3}{4})^2 = 4.452.$$

9. Sampling Distribution of Point Estimators

Besides reporting the value of a point estimate, some indication of its *precision* should be given. The usual measure of precision is the *standard error of the estimator* used to obtain the point estimate for a parameter θ. Denote this error:

$$\sigma_\theta = \sqrt{\mathbb{V}[\widehat{\theta}]}$$

where $\widehat{\theta}$ is a point estimator of θ.

If the standard error itself involves unknown parameters whose values can be estimated, substitution of these estimates into σ_θ yields the *estimated standard error* of the estimator. We denote the estimated standard error by $\widehat{\sigma}_\theta$.

But how can we obtain $\widehat{\theta}$? There are three alternative:

(a) Obtain the exact probability distribution of the estimator based on the probability assumption on the population.

(b) Obtain the approximate probability distribution of the estimator for large sample sizes.

(c) Approximate the (estimated) standard error using simulation techniques.

We will discuss only the first two approaches for the sample proportion, sample mean and sample variance.

10. Sampling Distribution of the Sample Proportion

For fixed n, X has a $Bin(n,p)$ of unknown parameter p, the probability of a success.

(a) Point estimator for p:

$$\widehat{p} = \frac{X}{n}.$$

(b) Since X is a sum of n Bernoulli trials, for large n we can approximate the distribution of $\widehat{p}$ with a normal distribution (Central Limit Theorem). For $X_1, \ldots, X_n$ Bernoulli trials:

$$\widehat{p} = \frac{X_1 + \ldots + X_n}{n} \sim\approx N(\mu_{\widehat{p}}, \sigma_{\widehat{p}})$$

where

$$\begin{aligned} \mu_{\widehat{p}} &= \mathbb{E}[\widehat{p}] = \mathbb{E}\left[\tfrac{X}{n}\right] = \tfrac{np}{n} = p \\ \sigma_{\widehat{p}} &= \mathbb{V}[\widehat{p}] = \mathbb{V}\left[\tfrac{X}{n}\right] = \tfrac{np(1-p)}{n^2} = \tfrac{p(1-p)}{n} \end{aligned}$$

Therefore, the (approximate) sampling distribution is:

$$\widehat{p} \sim\approx N(p, \frac{p(1-p)}{n})$$

(c) The approximate (large n) standard error is:

$$\sigma_p = \sqrt{\mathbb{V}[\widehat{p}]} \approx \sqrt{\frac{p(1-p)}{n}},$$

which is a measure of accuracy of $\widehat{p}$, a point estimator for p. But the standard error depends on the unknown parameter p. Therefore, we estimate the standard error by plugging-in $\widehat{p}$:

$$\widehat{\sigma}_p \approx \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}.$$

For the observation x from a binomial experiment, the standard error based on this observation is:

$$\widehat{\sigma}_p \approx \sqrt{\frac{\frac{x}{n}(1-\frac{x}{n})}{n}} = \sqrt{\frac{x(x-n)}{n^2}}$$

11. Example: Sampling Distribution of the Sample Proportion

When tossing a die, a success is when we score '6'. A first experiment is to roll the die for 100 times and only 30 success are observed. In a second experiment, we roll the die for 1000 times and we observe 300 successes. For the first experiment, the point estimate is:

$$\widehat{p}_1 = \frac{30}{100} = .3.$$

For the second experiment, the point estimate is:

$$\widehat{p}_2 = \frac{300}{1000} = .3.$$

For the first experiment, the standard error is:

$$\sigma_{p_1} = \frac{1}{100}\frac{30(100-30)}{100} = .046.$$

For the second experiment, the standard error is:

$$\sigma_{p_2} = \frac{1}{1000}\frac{300(1000-300)}{1000} = .014.$$

So even though the point estimates coincides for the two experiments, the standard error for the second experiment is smaller and the second experiment is more accurate.

12. Sampling Distribution of the Sample Mean

$X_1, \dots, X_n$ a random sample from a distribution of mean μ and variance σ ($\mathbb{E}[X_i] = \mu$,$\mathbb{V}[X_i] = \sigma$). We are interested in the mean of the population, μ.

(a) Point estimator for μ:

$$\widehat{\mu} = \frac{X_1 + \ldots + X_n}{n} = \bar{X}.$$

(b) For large n, apply CLT:

$$\widehat{\mu} \sim\approx N(\mu, \frac{\sigma^2}{n}),$$

which is the (approximate) sampling distribution of $\widehat{\mu}$.
If σ is known, then the measure of precision for μ is:

$$\sigma_\mu = \frac{\sigma}{\sqrt{n}}.$$

If σ is unknown, then the measure of precision is estimated by substituting σ^2 with its point estimate:

$$\widehat{\sigma}_\mu = \sqrt{\frac{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}{n}}$$

where $x_1, \dots, x_n$ are observed values.

Similar to the sample proportion, the approximation is more accurate as n large.

13. Sampling Distribution of the Sample Variance

$X_1, \dots, X_n$ a random sample from normal distribution of mean μ and variance σ ($X_i \sim N(\mu, \sigma^2)$). We are interested in the variability of the population, σ^2.

(a) *The Chi-Square Distribution*
If a random variable X has a standard normal distribution, then $Y = X^2$ is said to have a *chi-square distribution with one degree of freedom* and denote χ_1^2.
If $X_1, \ldots, X_n$ is a random sample from $N(0,1)$, then

$$Y = X_1^2 + \ldots + X_n^2 \sim \chi_n^2$$

and it is said that Y has a chi-square distribution with n degrees of freedom.
The expectation and the variance of $X \sim \chi_n^2$ are:

$$\mathbb{E}[X] = n$$
$$\mathbb{V}[X] = 2n$$

(b) Point estimator for σ^2:

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

(c) The exact sampling distribution of $\widehat{\sigma}^2$ is:

$$\widehat{\sigma}^2 \sim \frac{\sigma^2}{n-1}\chi_{n-1}^2.$$

(d) The standard deviation of $\widehat{\sigma}$ is:

$$\sqrt{\mathbb{V}[\widehat{\sigma}]} = \sqrt{\frac{\sigma^2}{n-1}(2(n-1))} = \sqrt{2\sigma^2}.$$

Since the standard error depends on the unknown σ^2, we substitute it with its estimator S^2:

$$\widehat{\sigma}_\sigma = \sqrt{2S^2} = \sqrt{2\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}.$$

14. Sampling Distribution of the Sample Mean under Normal Distribution

$X_1, \ldots, X_n$ a random sample from normal distribution of mean μ and variance σ ($X_i \sim N(\mu, \sigma^2)$).

(a) *t-distribution*
A *t-distribution* with n degrees of freedom is defined by:

$$t \sim \frac{N(0,1)}{\sqrt{\frac{\chi_n^2}{n}}}$$

where $N(0,1)$ and χ_n^2 are independent. Therefore, if we divide a normal standard random variable with the square root of an independent χ_n^2/n random variable, the resulting random variable has t distribution with n degrees of freedom. We denote it with t_n.

The expectation and the variance of a $X \sim t_n$ are:

$$\mathbb{E}[X] = 0$$
$$\mathbb{V}[X] = \frac{n}{n-2}$$

(b) According to the definition of the t-distribution and the sampling distribution of $\widehat{\sigma}^2 = S^2$,we have

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t_{n-1}.$$

15. Example: Proportion parameter

Of n_1 randomly selected male smokers, X_1 smoked filter cigarettes, whereas of n_2 randomly selected female, X_2 smoked filter cigarettes. Let p_1 and p_2 denote the probabilities that a randomly selected male and female, respectively, smoke filter cigarettes.

(a) Show that $(X_1/n_1) - (X_2/n_2)$ is an unbiased estimator for $p_1 - p_2$.

(b) What is the standard error of the estimator in part (a)?

(c) How would you use the observed values x_1 and x_2 to estimate the standard error of your estimator?

(d) If $n_1 = n_2 = 200$ and $x_1 = 127$, $x_2 = 176$, use the estimator of part (a) to obtain an estimate of $p_1 - p_2$.

(e) Use the result of part (c) and the data of part (d) to estimate the standard error of the estimator.

(a)

$$\mathbb{E}\left((X_1/n_1) - (X_2/n_2)\right) = \tfrac{1}{n_1}\mathbb{E}(X_1) - \tfrac{1}{n_2}\mathbb{E}(X_2) =$$
$$\tfrac{1}{n_1} n_1 p_1 - \tfrac{1}{n_2} n_2 p_2 = p_1 - p_2$$

(b)

$$\mathbb{V}\left((X_1/n_1) - (X_2/n_2)\right) = \tfrac{1}{n_1^2}\mathbb{V}(X_1) + \tfrac{1}{n_2^2}\mathbb{V}(X_2) =$$
$$\tfrac{1}{n_1^2} n_1 p_1(1-p_1) + \tfrac{1}{n_2^2} n_2 p_2(1-p_2) = \tfrac{p_1(1-p_1)}{n_1} + \tfrac{p_2(1-p_2)}{n_2}$$

(c) With $\widehat{p}_1 = \frac{X_1}{n_1}$ and $\widehat{p}_2 = \frac{X_2}{n_2}$ the estimated standard error is:

$$\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}$$

(d)

$$\widehat{p}_1 - \widehat{p}_2 = \frac{127}{200} - \frac{176}{200} = -.245$$

(e)

$$\sqrt{\frac{(.635)(.365)}{200} + \frac{(.880)(.120)}{200}} = .041$$

16. Example: Comparing point estimators

Let X_1, X_2, X_3 be a random sample from the following distribution:

$$f_X(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

(This is an $\exp\left(\frac{1}{\theta}\right)$ distribution.)

Consider the following estimators for θ.

- $\widehat{\theta}_1 = X_1$
- $\widehat{\theta}_2 = \frac{X_1+X_2}{2}$
- $\widehat{\theta}_3 = \frac{X_1+2X_2}{3}$
- $\widehat{\theta}_4 = \frac{X_1+X_2+X_3}{3}$

(a) Which of these estimators is unbiased?

(b) Among the unbiased estimators, which one has the smallest variance?

(a)
- $$\mathbb{E}(\widehat{\theta}_1) = \mathbb{E}(X_1) = (X_1 \sim exp(\frac{1}{\theta})) = \theta \;\Rightarrow\; \text{unbiased}$$
- $$\mathbb{E}(\widehat{\theta}_2) = \mathbb{E}\left(\tfrac{X_1+X_2}{2}\right) = \tfrac{1}{2}\mathbb{E}(X_1 + X_2) = \\ \tfrac{1}{2}(\mathbb{E}(X_1) + \mathbb{E}(X_2)) = \tfrac{1}{2}(\theta + \theta) = \theta \;\Rightarrow\; \text{unbiased}$$
- $$\mathbb{E}(\widehat{\theta}_3) = \mathbb{E}\left(\tfrac{X_1+2X_2}{3}\right) = \tfrac{1}{3}\mathbb{E}(X_1 + 2X_2) = \\ \tfrac{1}{3}(\mathbb{E}(X_1) + 2\mathbb{E}(X_2)) = \tfrac{1}{3}(\theta + 2\theta) = \theta \;\Rightarrow\; \text{unbiased}$$
- $$\mathbb{E}(\widehat{\theta}_4) = \mathbb{E}(\bar{X}) = \mathbb{E}(X) = \theta \;\Rightarrow\; \text{unbiased}$$

Therefore, $\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\theta}_3, \widehat{\theta}_4$ are unbiased.

(b)
- $$\mathbb{V}(\widehat{\theta}_1) = \mathbb{V}(X_1) = (X_1 \sim exp(\frac{1}{\theta})) = \theta^2$$
- $$\mathbb{V}(\widehat{\theta}_2) = \mathbb{V}\left(\tfrac{X_1+X_2}{2}\right) = \tfrac{1}{4}\mathbb{V}(X_1 + X_2) = (\text{since } X_1 \perp X_2) = \\ \tfrac{1}{4}(\mathbb{V}(X_1) + \mathbb{V}(X_2)) = \tfrac{1}{4}(\theta^2 + \theta^2) = \tfrac{\theta^2}{2}$$

- $$\mathbb{V}(\widehat{\theta}_3) = \mathbb{V}\left(\tfrac{X_1+2X_2}{3}\right) = \tfrac{1}{9}\mathbb{V}(X_1+2X_2) = (\text{since } X_1 \perp X_2) = \tfrac{1}{9}(\mathbb{V}(X_1)+4\mathbb{V}(X_2)) = \tfrac{1}{9}(\theta^2+4\theta^2) = \tfrac{5\theta^2}{9}$$

- $$\mathbb{V}(\widehat{\theta}_4) = \mathbb{V}(\bar{X}) = \left(\mathbb{V}(\bar{X}) = \frac{\mathbb{V}(X)}{n}\right) = \frac{\theta^2}{3}$$

$$\Rightarrow \; \mathbb{V}(\widehat{\theta}_4) < \mathbb{V}(\widehat{\theta}_2) < \mathbb{V}(\widehat{\theta}_3) < \mathbb{V}(\widehat{\theta}_1)$$

Therefore, $\widehat{\theta}_4$ has the smallest variance.

17. Example: Mean square error (MSE)

The reading on a voltage meter connected to a test circuit is uniformly distributed over the interval $(\theta, \theta+1)$, where θ is the true but unknown voltage of the circuit. Suppose that $X_1, X_2, \ldots, X_n$ denotes a random sample of such readings.

(a) Show that $\bar{X}$ is a biased estimator of θ, and compute the bias.

(b) Find $MSE(\bar{X})$. Based on the MSE, is $\bar{X}$ a consistent estimator of θ?

(a)

$$\mathbb{E}(\bar{X}) = \mathbb{E}(X) = \theta + \tfrac{1}{2} \neq \theta \;\Rightarrow\; \bar{X} \text{ is not unbiased}$$
$$Bias(\bar{X}) = \mathbb{E}(\bar{X}) - \theta = \theta + \tfrac{1}{2} - \theta = \tfrac{1}{2}$$

(b)

$$MSE(\bar{X}) = \mathbb{V}(\bar{X}) + [Bias(\bar{X})]^2 = \frac{\mathbb{V}(X)}{n} + \left(\frac{1}{2}\right)^2 = \frac{1}{12n} + \frac{1}{4}$$

Since $MSE(\bar{X}) = \frac{1}{12n} + \frac{1}{4} \longrightarrow_{n\to\infty} \frac{1}{4} \neq 0$, $\bar{X}$ is not consistent for θ.