STAT 234 Lecture 15B Population & Sample (Section 1.1)
Lecture 16A Point Estimates, Mean Squared Error, Unbiased Estimates (Section 7.1)

Yibi Huang
Department of Statistics
University of Chicago

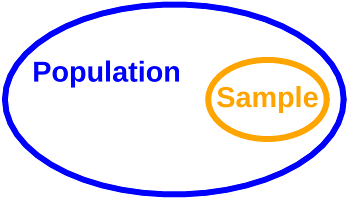
Four Keywords

Population: The entire group of individuals in which we are interested but can't usually assess directly.

Example: All humans, all working-age people in California, all crickets

Sample: The part of the population we actually examine and for which we do have data.

How well the sample represents the population depends on the sample design.



A **parameter** is a number describing a characteristic of the population

A **statistic** is a number describing a characteristic of a sample.

Good and Bad Sampling Methods

Whether information collected from the sample tells us truth about the population depends on how subjects is selected from the population to form a sample.

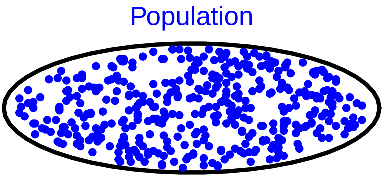
Simple Random Sampling

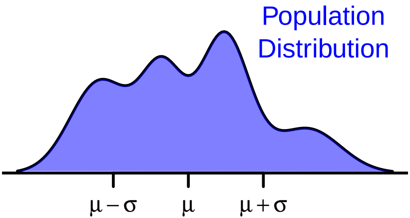
- Basic idea: put the names of individuals in the population in a box, shake well, and make draws from the box at random **without replacement**
- produces nearly **i.i.d.** observations if the sample size is less than 10% of the population size

Convenience sampling: just sampling from those who are easily accessible

- E.g. “Man on the street” survey (cheap, convenient, popular with TV “journalism”)

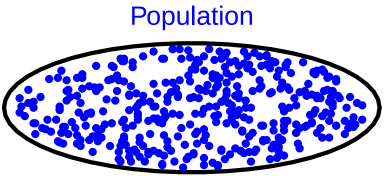
Population & Parameter

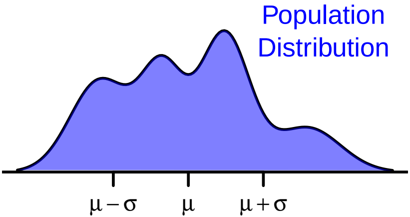




- Suppose we are interested in some numerical characteristic X about individuals in a certain population.

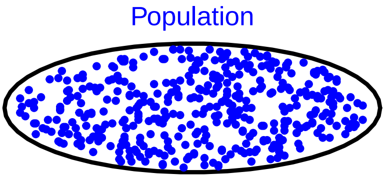
Population & Parameter

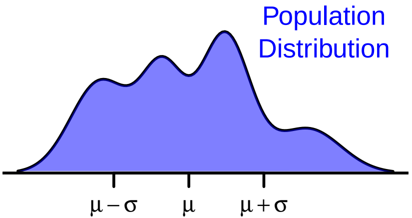




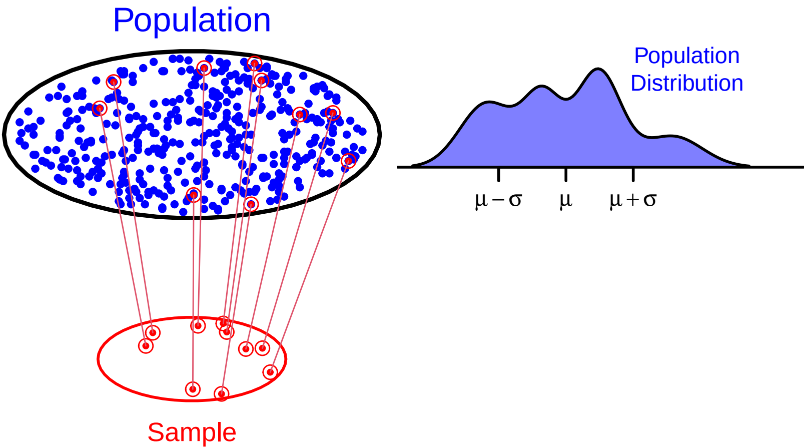
- Suppose we are interested in some numerical characteristic X about individuals in a certain population.
- If it's possible to interview each individual in the population and record his/her X value, we can then make a histogram for the recorded X-values and that's the *population distribution*.

Population & Parameter



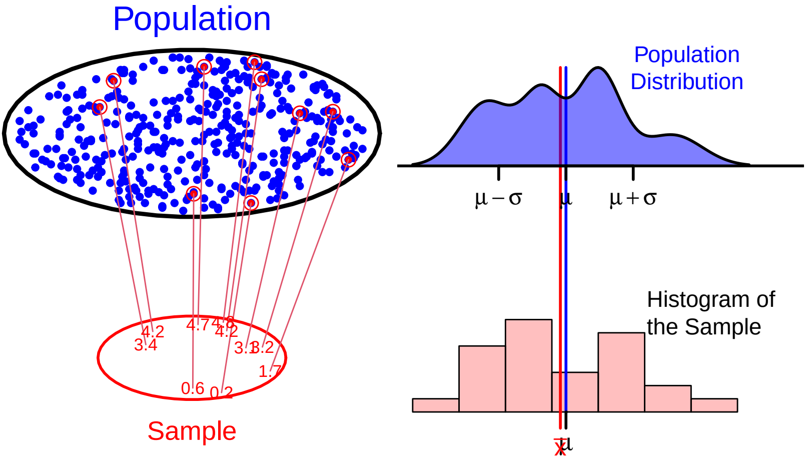


- Suppose we are interested in some numerical characteristic X about individuals in a certain population.
- If it's possible to interview each individual in the population and record his/her X value, we can then make a histogram for the recorded X-values and that's the *population distribution*.
- The population distribution is arbitrary (not necessarily normal). We are interested in the *parameters* of the population distribution, e.g. the *population mean* μ and the *population variance* σ^2, or other parameters that describe the population distribution.



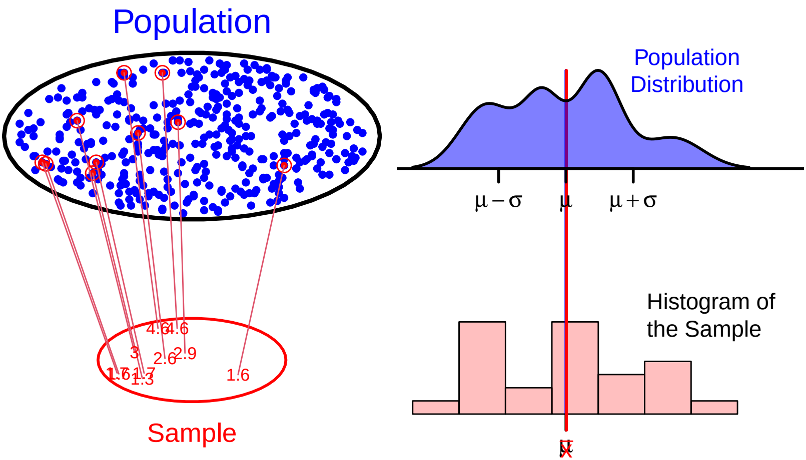
A (simple) random sample is taken from the population.

Sample and Statistics

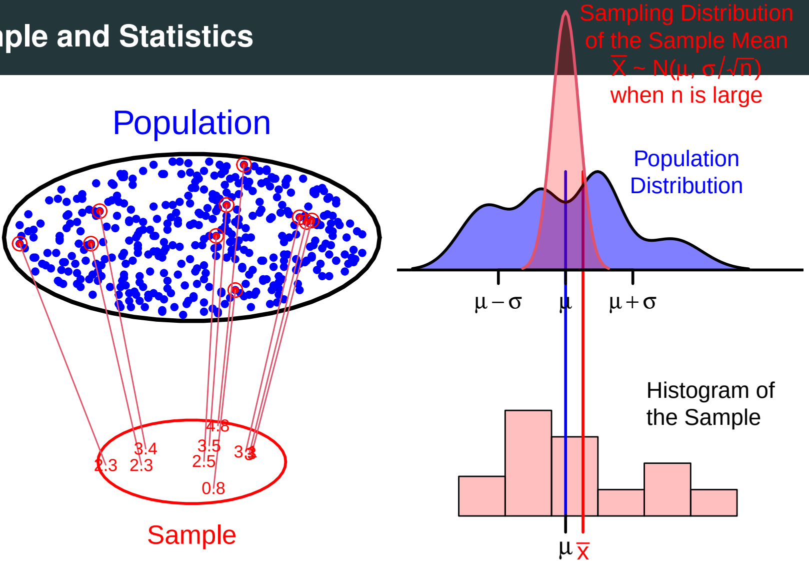


The X-values $(X_1, X_2, \ldots, X_n)$ for individuals in the sample are recorded. One can use a histogram for the sample to estimate the population distribution.

Sample and Statistics



A *statistic* T is a function of the sample $T = g(X_1, \ldots, X_n)$ that is used to estimate the parameter of a population, e.g., the sample mean $\overline{X} = \frac{1}{n}\sum_i X_i$ is a statistic to estimate the population mean μ.



A *statistic* is a random variable that it's value will vary from sample to sample, e.g., the sample mean $\overline{X}$ is random and it's distribution is approx $N(\mu, \sigma/\sqrt{n})$ when n is large by CLT.

Sample and Statistic

A **statistic** $T = T(X_1, X_2, \ldots, X_n)$ is a function of the random sample $X_1, X_2, \ldots, X_n$.

- A statistic cannot involve any *unknown* parameter, for example, $\overline{X} - \mu$ is not a statistic if the population mean μ is unknown.
- A statistic T itself is a random variable, which its own probability. This distribution of T allows us to determine the accuracy and reliability of our estimate.

Examples

- the sample mean $\overline{X}$ is a statistic to estimate the population mean μ
- the sample variance $s^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}$ is a statistic to estimate the population variance σ^2

Statistical Inference

Suppose we want to estimate a parameter θ of a population distribution $f(x|\theta)$, e.g., θ can be

- population mean $\mu = \int x f(x|\theta) dx$
- population variance $\sigma^2 = \int (x-\mu)^2 f(x|\theta) dx$
- success probability p for a Binomial distribution, and so on

based on a random sample $X_1, X_2, \ldots, X_n$ of size n.

Statistical Inference

Suppose we want to estimate a parameter θ of a population distribution $f(x|\theta)$, e.g., θ can be

- population mean $\mu = \int x f(x|\theta) dx$
- population variance $\sigma^2 = \int (x-\mu)^2 f(x|\theta) dx$
- success probability p for a Binomial distribution, and so on

based on a random sample $X_1, X_2, \ldots, X_n$ of size n.

In Chapter 3-6, we assume the parameters μ, σ^2, p's are *known* and can be used to calculate probabilities for the random variables $X_1, X_2, \ldots, X_n$.

Suppose we want to estimate a parameter θ of a population distribution $f(x|\theta)$, e.g., θ can be

- population mean $\mu = \int x f(x|\theta) dx$
- population variance $\sigma^2 = \int (x-\mu)^2 f(x|\theta) dx$
- success probability p for a Binomial distribution, and so on

based on a random sample $X_1, X_2, \ldots, X_n$ of size n.

In Chapter 3-6, we assume the parameters μ, σ^2, p's are *known* and can be used to calculate probabilities for the random variables $X_1, X_2, \ldots, X_n$.

In Statistics problems, the parameters of the population distribution are *unknown* as we cannot observe the entire population. *Statistical inference* use the *observed random sample* $X_1, \ldots, X_n$ to *infer the UNKNOWN parameter(s):* μ, σ^2, p*, etc*.

Statistical Inference

Statistical Inference include

- *point estimate*: finding an estimated value for the unknown parameter .... (Chapter 7)
- *confidence interval*: giving a range for the plausible values of the unknown parameter .... (Chapter 8)
- *hypothesis testing*: testing whether the unknown parameter is a given value or not .... (Chapter 9)

Point Estimate

The *point estimate* of a parameter θ, is a statistic $T = T(X_1, \ldots, X_n)$ computed from the sample $\{X_1, \ldots, X_n\}$ that is a sensible guess for the unknown θ.

Examples of Point Estimates

Example 1: If $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \sigma^2)$, the point estimate for the population mean μ can be

- the sample mean $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$
- the median of $X_1, \ldots, X_n$
- the average of $X_1, \ldots, X_n$ after discarding the minimum & maximum

The point estimate for the population variance σ^2 can be

- the sample variance $S^2 = \dfrac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{n-1}$
- an alternative estimator would result from using divisor n instead of $n-1$

$$\widehat{\sigma}^2 = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{n}$$

Examples of Point Estimates

Example 2: If $X \sim Bin(n, p)$ is Binomial, the point estimate for the success probability p can be

- the sample proportion $\widehat{p} = \dfrac{X}{n}$
- Wilson's plus-four estimate $\tilde{p} = \dfrac{X+2}{n+4}$
  - adding successes and two failures to the sample and then calculate the sample proportion of successes

Mean Squared Error

With many possible point estimates $\widehat{\theta}$'s for a parameter θ, how to choose a good one among them?

A population criterion is to compare their *Mean Squared Error (MSE)*, defined as

$$\text{Mean Squared Error (MSE)} = \mathrm{E}[(\widehat{\theta} - \theta)^2]$$

MSE = (Bias)^2+ Variance

Recall the shortcut formula for the variance of any variable Y

$$\mathrm{Var}(Y) = \mathrm{E}(Y^2) - (\mathrm{E}(Y))^2,$$

Rearranging the terms, we get

$$\mathrm{E}(Y^2) = (\mathrm{E}(Y))^2 + \mathrm{Var}(Y).$$

Plugging in $Y = \widehat{\theta} - \theta$, then $\mathrm{E}(\widehat{\theta} - \theta) = \mathrm{E}(\widehat{\theta}) - \theta$, we get

$$\begin{array}{ccccc} \mathrm{E}[(\widehat{\theta} - \theta)^2] & = & [\mathrm{E}(\widehat{\theta}) - \theta]^2 & + & \mathrm{Var}(\widehat{\theta} - \theta) \\ \| & & \| & & \| \\ \text{MSE} & = & (\text{Bias})^2 & + & \mathrm{Var}(\widehat{\theta}) \end{array}$$

where the *bias* of an point estimate $\widehat{\theta}$ for θ is defined to be the difference between the expected value of the estimate and the true value of the parameter

$$\text{Bias} = \mathrm{E}(\widehat{\theta}) - \theta$$

Examples of MSE

If $X_1, \ldots, X_n$ are i.i.d. with population mean μ and population variance σ^2, using the sample mean $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ the point estimate for the population mean μ

- the bias is $\mathrm{E}(\overline{X}) - \mu = \mu - \mu = 0$
- the variance is $\mathrm{Var}(\overline{X}) = \sigma^2/n$

The MSE for $\overline{X}$ is hence

$$\text{MSE} = (\text{Bias})^2 + \text{Variance} = 0^2 + \frac{\sigma^2}{n} = \frac{\sigma^2}{n}$$

Unbiased Estimators

A point estimator $\widehat{\theta}$ is said to be an *unbiased estimator* of θ if

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

for every possible value of θ.

Unbiased Estimators

A point estimator $\widehat{\theta}$ is said to be an *unbiased estimator* of θ if

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

for every possible value of θ.

For unbiased estimators, $\boxed{\text{MSE = Variance}}$.

Proof of Expected Value of Sample Variance (1)

Recall X_i's are i.i.d. with $\mathrm{E}(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$. In L13, we showed that the mean and variance of the sample mean $\overline{X}$ are

$$\mathrm{E}(\overline{X}) = \mu \quad \text{and} \quad \mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$$

Applying the general result that $\mathrm{E}(Y^2) = \mathrm{Var}(Y) + (\mathrm{E}(Y))^2$ to X_i and $\overline{X}$, we get

$$\mathrm{E}(X_i^2) = \mathrm{Var}(X_i) + (\mathrm{E}(X_i))^2 = \sigma^2 + \mu^2$$

$$\mathrm{E}(\overline{X}^2) = \mathrm{Var}(\overline{X}) + \left(\mathrm{E}(\overline{X})\right)^2 = \frac{\sigma^2}{n} + \mu^2$$

Proof of Expected Value of Sample Variance (2)

Taking expectation on both sides of the shortcut formula for sample variance,

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

and plugging in $\mathrm{E}(X_i^2) = \sigma^2 + \mu^2$ and $\mathrm{E}(\overline{X}^2) = \frac{\sigma^2}{n} + \mu^2$, we get

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

MSE of Sample Variance (May Skip)

To obtain the MSE, we also need to calculate the variance for the sample variance.

In general, the variance of the sample variance of i.i.d. $X_1, \dots, X_n$ can be shown to be

$$\mathrm{Var}(S^2) = \frac{\mathrm{E}[(X_i - \mu)^4]}{n} - \frac{(n-3)\sigma^4}{n(n-1)}.$$

If $X_1, X_2, \dots, X_n$ are i.i.d. normal, one can show $\mathrm{E}[(X_i - \mu)^4] = 3\sigma^4$ and hence the variance of S^2 simplifies to

$$\mathrm{Var}(S^2) = \frac{3(n-1)\sigma^4}{n(n-1)} - \frac{(n-3)\sigma^4}{n(n-1)} = \frac{3(n-1)-(n-3)}{n(n-1)}\sigma^4 = \frac{2\sigma^4}{(n-1)}.$$

The MSE of S^2 is hence

$$\text{MSE} = (\text{Bias})^2 + \text{Variance} = 0^2 + \frac{2\sigma^4}{n-1} = \frac{2\sigma^4}{n-1}$$

A Biased Estimator for σ^2 w/ Smaller MSE (May Skip)

Consider an alternative estimator for σ^2 that using divisor *n+1* instead of $n - 1$

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

The expected value and variance of $\widehat{\sigma}^2$ are respectively

$$\mathrm{E}(\widehat{\sigma}^2) = \frac{(n-1)\,\mathrm{E}(S^2)}{n+1} = \frac{(n-1)\sigma^2}{n+1},$$

$$\mathrm{Var}(\widehat{\sigma}^2) = \left(\frac{n-1}{n+1}\right)^2 \mathrm{Var}(S^2) = \left(\frac{n-1}{n+1}\right)^2 \frac{2\sigma^4}{(n-1)} = \frac{2(n-1)\sigma^4}{(n+1)^2}$$

if $X_1, \ldots, X_n$'s are i.i.d. normal. Hence, $\widehat{\sigma}^2$ is a **biased** estimator for σ^2 with

$$\text{Bias} = \mathrm{E}(\widehat{\sigma}^2) - \sigma^2 = \frac{(n-1)\sigma^2}{n+1} - \sigma^2 = \frac{-2\sigma^2}{n+1}.$$

The MSE of $\widehat{\sigma}^2$ is

$$\text{MSE} = (\text{Bias})^2 + \text{Variance} = \left(\frac{-2\sigma^2}{n+1}\right)^2 + \frac{2(n-1)\sigma^4}{(n+1)^2} = \frac{2n\sigma^4}{(n+1)^2}$$

which is lower than the MSE of $\dfrac{2\sigma^4}{n-1}$ for the sample variance S^2.

The MSE of $\widehat{\sigma^2}$ is

$$\text{MSE} = (\text{Bias})^2 + \text{Variance} = \left(\frac{-2\sigma^2}{n+1}\right)^2 + \frac{2(n-1)\sigma^4}{(n+1)^2} = \frac{2n\sigma^4}{(n+1)^2}$$

which is lower than the MSE of $\dfrac{2\sigma^4}{n-1}$ for the sample variance S^2.

A biased estimator might have smaller MSE if it has a smaller variance.

MSE of the Sample Proportion $\widehat{p} = \dfrac{X}{n}$

If $X \sim Bin(n, p)$ is Binomial, a point estimate for the success probability p is the sample proportion $\widehat{p} = \dfrac{X}{n}$. As X is Binomial,

$$\begin{aligned} \mathrm{E}(X) &= np & \Rightarrow \quad \mathrm{E}(\widehat{p}) &= \frac{\mathrm{E}(X)}{n} = \frac{np}{n} = p \\ \mathrm{Var}(X) &= np(1-p) & \Rightarrow \quad \mathrm{Var}(\widehat{p}) &= \frac{\mathrm{Var}(X)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n} \end{aligned}$$

Thus the sample proportion $\widehat{p}$ is **unbiased** with the MSE

$$\text{MSE} = (\text{Bias})^2 + \text{Variance} = 0^2 + \frac{p(1-p)}{n} = \frac{p(1-p)}{n}.$$

MSE for Wilson's "Plus-Four" Estimate for Proportions

Recall Wilson's plus-four estimate is

$$\tilde{p} = \frac{X+2}{n+4}.$$

It's expected value and variance are respectively,

$$\mathrm{E}(\tilde{p}) = \frac{\mathrm{E}(X)+2}{n+4} = \frac{np+2}{n+4}, \text{ and } \mathrm{Var}(\tilde{p}) = \frac{\mathrm{Var}(X)}{(n+4)^2} = \frac{np(1-p)}{(n+4)^2}.$$

Its bias and MSE are respectively

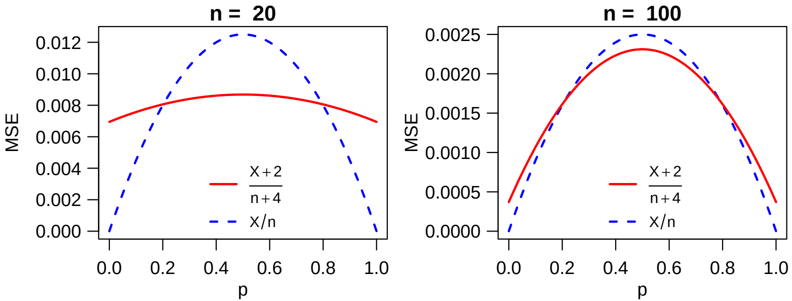
$$\text{Bias} = \mathrm{E}(\tilde{p}) - p = \frac{np+2}{n+4} - p = \frac{2-4p}{n+4}$$

$$\text{MSE} = (\text{Bias})^2 + \text{Variance} = \left(\frac{2-4p}{n+4}\right)^2 + \frac{np(1-p)}{(n+4)^2}$$

MSE's for Sample Proportion & Wilson's "Plus-Four"

Below are the graphs of the MSE for $\widehat{p} = X/n$ and $\tilde{p} = \dfrac{X+2}{n+4}$

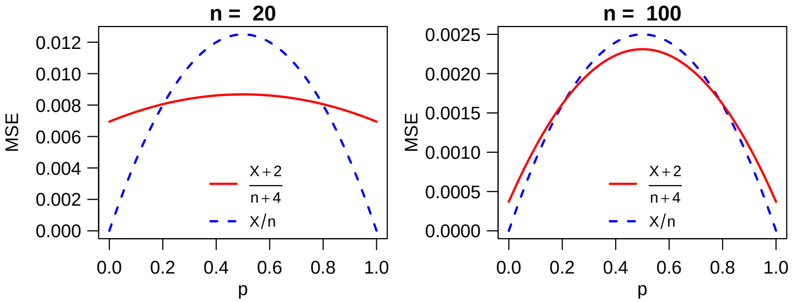
$$\text{MSE}(\widehat{p}) = \frac{p(1-p)}{n}, \quad \text{MSE}(\tilde{p}) = \left(\frac{2-4p}{n+4}\right)^2 + \frac{np(1-p)}{(n+4)^2}$$



MSE's for Sample Proportion & Wilson's "Plus-Four"

Below are the graphs of the MSE for $\widehat{p} = X/n$ and $\tilde{p} = \dfrac{X+2}{n+4}$

$$\mathrm{MSE}(\widehat{p}) = \frac{p(1-p)}{n}, \quad \mathrm{MSE}(\tilde{p}) = \left(\frac{2-4p}{n+4}\right)^2 + \frac{np(1-p)}{(n+4)^2}$$



- $\widehat{p} = X/n$ has a smaller MSE only when p is close to 0 or 1
- $\tilde{p} = \frac{X+2}{n+4}$ has a smaller MSE when p is NOT close to 0 or 1
- The two MSE's are close when n is large

Other Criteria for Choosing Point Estimates

Chapter 7 of MMSA introduced several other criteria for choosing points estimates including

- maximum likelihood estimate (MLE) in Section 7.2
- sufficient statistic in Section 7.3
- information and efficiency in Section 7.4

All of them are important in Statistics but we don't have enough time to cover them in STAT 234. If interested, you can take STAT 244 to learn them.