

Supplementary Notes

Estimator properties

In this section we recap some important statistical properties of our estimators (for further details please refer to the statistical inference course).

A population parameter θ represents a quantitative measurement of population values. Since θ is typically unknown, we need to estimate it by drawing samples from the population of interest. Our estimator, $\hat{\theta}$, is then computed based on the samples we have obtained. The question is then, how well does $\hat{\theta}$ do its job in estimating θ ? To evaluate this, we examine key statistical properties of estimators.

Consistency

As we see in Figure 1, there is some discrepancy between the sample means and the true population mean. Thus, it is expected that, as the sample size increases, our estimator $\hat{\theta}$ will get closer to the true value θ . Thus an estimator $\hat{\theta}$ is said to be *consistent* if it converges to the true parameter θ as the sample size n increases, i.e., $\hat{\theta} \rightarrow \theta$ as $n \rightarrow \infty$. E.g., let $\hat{\theta} = \bar{x} + \frac{1}{n}$ be a biased but consistent estimator of $\theta = 5$ (\bar{x} represents the sample mean). Figure 2 shows how as sample size increases, the estimator converges to the true parameter.

Expected value

The expected value of an estimator is a weighted average of all possible estimates. Here, the weights are given by the probability of selecting a particular sample s , i.e., $p(s)$. Mathematically, this can be written as:

$$\mathbb{E}(\hat{\theta}) = \sum_{s \in \Omega} p(s) \hat{\theta}(s)$$

Where Ω is the sample space, i.e. the number of possible samples. Furthermore, if all possible samples are equally likely then $p(s) = \frac{1}{\Omega}$ and the expected values becomes:

$$\mathbb{E}(\hat{\theta}) = \frac{1}{\Omega} \sum_{s \in \Omega} \hat{\theta}(s).$$

Sampling from Normal Distribution

Black curve: True population ($N(0,1)$)

Colored points: Sample observations

Dashed lines: Sample means

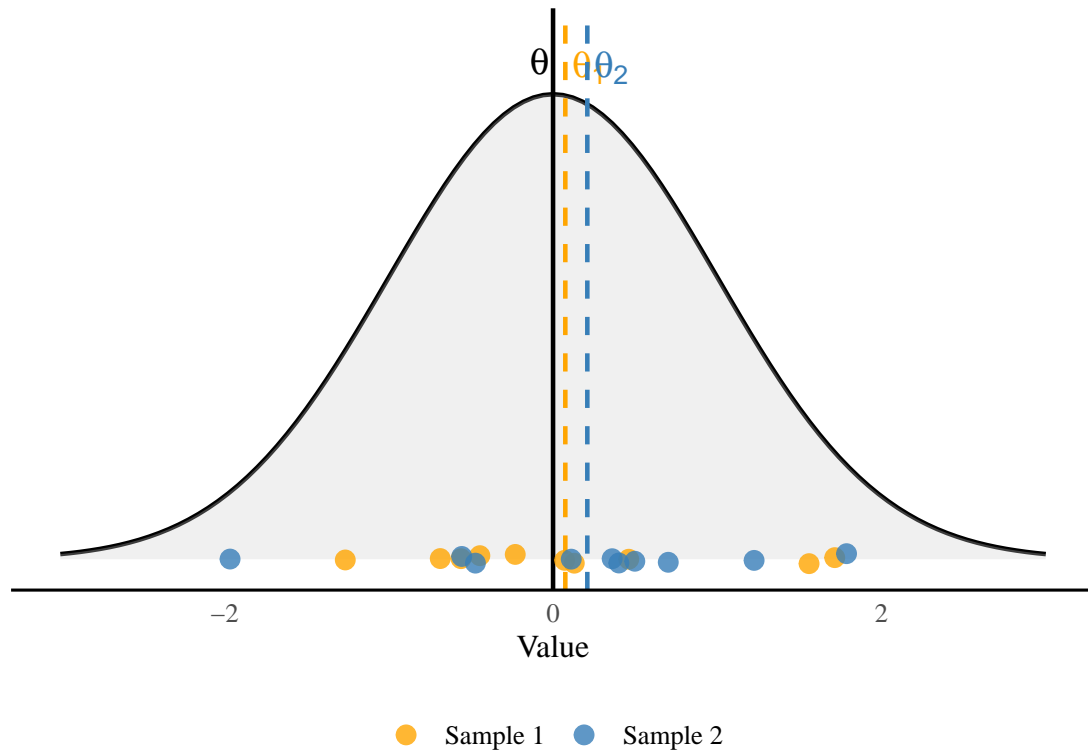


Figure 1: Sampling from Normal Distribution. The black curve represent the true population distribution with mean represented by the black solid line. The colored point are two different samples with their corresponding sample mean shown in colored dashed lines.

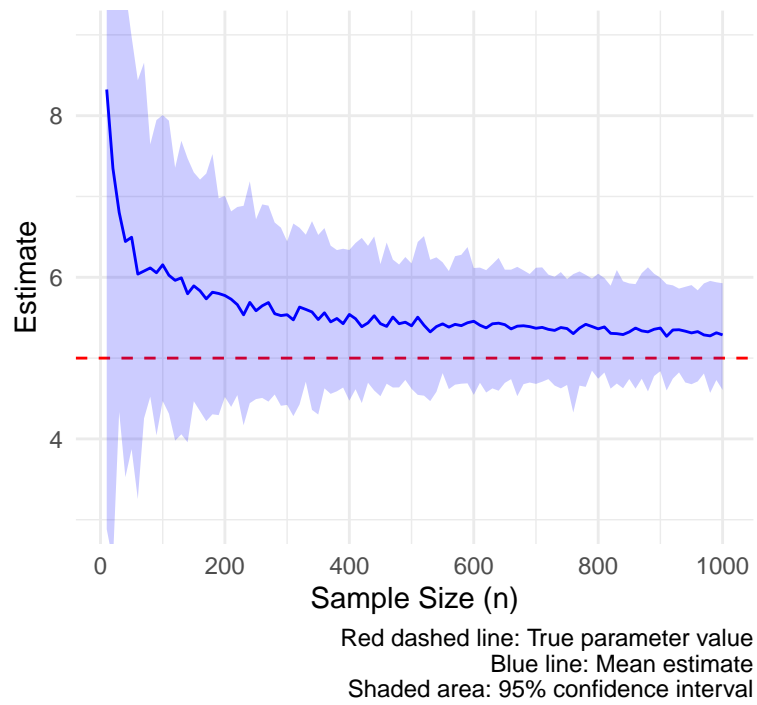


Figure 2: Demonstration of estimator consistency. The dashed red line indicates the true parameter value. The blue line tracks the mean estimate across increasing sample sizes, while the shaded region represents the 95% confidence interval.

Note that the expected value is a function of both, the sampling design (due to $p(s)$) and the population being sampled (through the sample estimate $\hat{\theta}(s)$).

Task 1

Suppose we have a tiny population of three trees with the following diameters (cm) :

Tree	diameter (cm)
1	9.1
2	10.2
3	17.2

Now, we sample only **2 trees** and measure the diameter of each tree in the sample. It is clear from here that there are only $\Omega = 3$ possible samples we could've taken:

1. $s_1 = \{\text{tree 1, tree 2}\}$
2. $s_2 = \{\text{tree 1, tree 3}\}$
3. $s_3 = \{\text{tree 2, tree 3}\}$

Imagine that samples are selected with an equal probability so that $p(s_1) = p(s_2) = p(s_3) = 1/3$.

The true population mean diameter is $\mu_y = (9.1 + 10.2 + 17.2)/3 = 12.17$.

Suppose we chose the sample average \bar{y} as our estimate of μ_y . Then, the expected value of \bar{y} is given by

$$\begin{aligned}\mathbb{E}(\bar{y}) &= p(s_1) \left(\frac{9.1 + 10.2}{2} \right) + p(s_2) \left(\frac{9.1 + 17.2}{2} \right) + p(s_3) \left(\frac{10.2 + 17.2}{2} \right) \\ &= \frac{1}{3} \left(\frac{19.3}{2} + \frac{26.3}{2} + \frac{27.4}{2} \right) \\ &= 12.17 = \mu_y\end{aligned}$$

What would have happened if our sampling design had unequal sampling probabilities given by $p(s_1) = 1/2$, $p(s_2) = 1/3$ and $p(s_3) = 1/6$? Calculate $\mathbb{E}(\hat{y})$ under this sampling scenario.

See Solution

$$\begin{aligned}\mathbb{E}(\bar{y}) &= p(s_1) \left(\frac{9.1 + 10.2}{2} \right) + p(s_2) \left(\frac{9.1 + 17.2}{2} \right) + p(s_3) \left(\frac{10.2 + 17.2}{2} \right) \\ &= \frac{1}{2} \left(\frac{19.3}{2} \right) + \frac{1}{3} \left(\frac{26.3}{2} \right) + \frac{1}{6} \left(\frac{27.4}{2} \right) \\ &= \frac{19.3}{4} + \frac{26.3}{6} + \frac{27.4}{12} \\ &= 11.5 \neq \mu_y\end{aligned}$$

In practice, $\mathbb{E}(\hat{\theta})$ cannot be evaluated because we can not measure all the N elements of our population of interest! (and if we could then θ could be evaluated directly on our population rather than obtaining a samples from it). However, understanding this concept is important for two other relevant quantities.

Bias

The bias of an estimator is the difference in magnitude between its expected value and the population parameter for which an estimated is desired:

$$\mathbf{Bias}(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta$$

When $\mathbb{E}(\hat{\theta}) = \theta$, $\hat{\theta}$ is said to be an *unbiased estimator* of θ .

Note that the bias is not a property of an individual estimate (e.g., $\hat{\theta}(s)$). E.g., we can see in Figure 1 there are some discrepancies between the true population parameter θ and the sample estimates $\hat{\theta}(1)$ and $\hat{\theta}(2)$. Such discrepancies $\hat{\theta}(s) - \theta$ are known as *sampling errors*. This doesn't mean that an error has been made during the sampling, it just indicates that the value being estimated will differ from the true value because is being estimated from just a fraction of the elements of the population. Thus, one may wonder how much an estimate $\hat{\theta}$ from one sample will differ from that calculated from a different sample. In principle we would like this difference to be small because that will ensure that no matter which sample we take, the estimated value will be similar across samples.

Variance

The **variance** of an estimator or *sampling variance* is the average squared distance between individual estimates $\hat{\theta}(s)$ and their expected value $\mathbb{E}(\hat{\theta})$, i.e.,

$$\text{Var}(\hat{\theta}) = \sum_{s \in \Omega} p(s) \left(\hat{\theta}(s) - \mathbb{E}(\hat{\theta}) \right)^2$$

Notice how the variance of an estimator does not depend (unlike the bias) on the true parameter θ

Task 2

Using the data from the previous task and assuming equal sampling probabilities, calculate $\text{Var}(\hat{y})$

See Solution

$$\begin{aligned} \text{Var}(\bar{y}) &= p(s_1) \left(\frac{19.3}{2} - 12.17 \right)^2 + p(s_2) \left(\frac{26.3}{2} - 12.17 \right)^2 + p(s_3) \left(\frac{27.4}{2} - 12.17 \right)^2 \\ &= \frac{1}{3} (9.65) = 3.22 \end{aligned}$$

Precision

The **precision** of an estimator is a qualitative measurement that assesses how small or large the variability of an estimator is and does not relate to the true value. To illustrate this, suppose we draw 20 random samples with equal probability from our population of interest. **Fig-bias_prec** shows 100 replicates/data sets of this experiment with 4 different estimators:

- $\hat{\theta}_1(s) = \frac{1}{20} \sum_{i=1}^{20} x_i$ - the mean of the s th sample which is **unbiased** and **precise**
- $\hat{\theta}_2(s) = \arg \min (x_1, \dots, x_{20})$ - the s th sample minimum value which is **biased** and **imprecise**.
- $\hat{\theta}_3(s) = x_1$ - the first observation of the unordered sample which is **unbiased** and **imprecise**.
- $\hat{\theta}_4(s) = 2 + \frac{1}{20} \sum_{i=1}^{20} x_i$ - the mean + constant which is **biased** but **precise**.

Imagine a dartboard where the bullseye represents the true population parameter (θ). Each throw corresponds to a sample estimate ($\hat{\theta}(s)$). Precision refers to how tightly clustered the darts are - a precise estimator produces estimates that land close together, regardless of their position relative to the bullseye. Bias, on the other hand, is the systematic offset from the bullseye - even with perfect precision, a biased estimator would consistently miss the center in the same direction.

