STAT 210
Applied Statistics and Data Analysis
Review of Inference I: Pointwise Estimation

Joaquín Ortega
KAUST

Fall 2020

Pointwise estimation

Introduction

One of the fundamental problems in Statistics is the study of a population using a sample drawn at random.

We assume that we have defined the characteristic we want to study and that the model consists of a family of distribution function $F(t;\theta)$ that depends on a parameter θ.

The model gives a framework for statistical analysis and is usually a simple approximation to a complicated reality.

A model that determines the distribution up to the value of one or several parameters is known as a **parametric model**.

The parameter θ may be a number or a vector.

The problem we want to consider here is the estimation of this parameter from a random sample drawn from the population.

Estimation

By **random sample**, we mean a subset $x_1, x_2, \ldots, x_n$ of elements in the population that are selected according to a random mechanism in which all elements of the population have the same probability of being selected.

We can think of the values $x_1, x_2, \ldots, x_n$ as the *realization* of a colection of random variables $X_1, X_2, \ldots, X_n$ where $X_i, i = 1, \ldots, n$ represent independent variables with distribution function $F(t; \theta)$.

We use the notation **iid** for independent, identically distributed.

As a rule, we will use capital letters X, Y, Z for random variables and small case letter x, y, z for their values once they have been observed. Also, rv stands for random variable.

Estimation

To estimate the unknown parameter, we use a **statistic**, which is a function of the random sample.

The sample mean is an example of a statistic, as are the sample variance, the standard deviation, the median, the quantiles, the correlation coefficient, or the slope of a regression line.

A statistic is any function of the random sample.

When we use a statistic to estimate a specific parameter of the population, we speak of an **estimator**.

Estimation

Thus, the sample mean is an estimator for the population mean. The sample mean is given by the formula

$$\bar{X} = \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{1}$$

When we replace the variables $X_1, X_2, \ldots, X_n$ by their observed values $x_1, x_2, \ldots, x_n$, we have an **estimate** of the unknown population mean:

$$\bar{x} = \bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i. \tag{2}$$

Example 1.

Suppose that we have a parametric model given by the family of Gaussian distributions $N(\mu, \sigma^2)$, with parameters μ and σ^2, where μ is the population mean and σ^2 is the variance.

These parameters are unknown, and we draw a random sample from the population. Suppose the sample has size $n = 10$ and we observe the following values:

```
round(smpl1,2)
```

```
##  [1] 4.56 5.46 5.01 4.64 3.73 1.17 7.08 5.73 4.79
## [10] 0.94
```

We can use the sample mean $\bar{x}_{10}$ to estimate μ. In `R`, we use the function `mean`:

```
(mean.smpl1 <- mean(smpl1))
```

```
## [1] 4.311172
```

Example 1.

Our estimate for the mean of the population, based on the sample contained in the vector `smpl1` is 4.31. We also use the notation $\hat{\mu}$ or $\hat{\mu}_n$ to denote this estimate

$$\hat{\mu} = \bar{x} = 4.31.$$

Similarly, we can use the sample variance s^2 to estimate σ^2. The formula for the sample variance is

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

In `R` we use the function `var` to obtain the sample variance:

```
(var.smpl1 <- var(smpl1))
```

```
## [1] 3.714826
```

We also use the notation $\hat{\sigma}^2$ to denote the estimate of the variance:

$$\hat{\sigma}^2 = 3.715$$

Sampling distribution

Sampling distribution

Since a statistic is a function of a random sample, and a random sample is a collection of iid random variables, a statistic is also a random variable.

If we repeat the sampling procedure to obtain a new sample, we will get different values for the sample and, therefore, a different value for the statistic.

The statistic has a probability distribution characterized by its distribution function. This distribution is known as the **sampling distribution** for the statistic.

It is not always easy to determine the sampling distribution, but it is always possible to calculate the mean and variance for the sampling distribution (assuming that they exist).

Sampling distribution

Suppose we have a random sample $X_1, \ldots, X_n$ from a distribution function $F(t; \theta)$ that has finite mean and variance, denoted by μ and σ^2, respectively:

$$E(X_i) = \mu, \qquad Var(X_i) = \sigma^2, \qquad i = 1, 2, \ldots, n.$$

The sample mean is given by the average of the values in the sample, and using the linearity of the expected value we get

$$E(\hat{\mu}_n) = E(\bar{X}) = E\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = E(X_i) = \mu \quad (3)$$

When the expected value of the estimator is the parameter we want to estimate, we say that the estimator is **unbiased**. The sample mean is an unbiased estimator of the population mean.

Sampling distribution

To calculate the variance of $\hat{\mu}_n$ observe that

$$\hat{\mu}_n - \mu = \Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) - \mu = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)$$

In consequence

$$\begin{aligned}
Var(\hat{\mu}_n) &= E\big[(\hat{\mu}_n - \mu)^2\big] = E\Big[(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu))^2\Big] \\
&= \frac{1}{n^2}\Big[\sum_{i=1}^{n} E(X_i - \mu)^2 + \sum_{i \neq j} E(X_i - \mu)(X_j - \mu)\Big] \\
&= \frac{1}{n^2}\Big[\sum_{i=1}^{n} Var(X_i) + \sum_{i \neq j} Cov(X_i, X_j)\Big] \qquad (4)
\end{aligned}$$

where $Cov(X, Y) = E\big((X - \mu_X)(Y - \mu_Y)\big)$ is the covariance between X and Y with $\mu_X = E(X)$ and similarly for μ_Y.

Sampling distribution

Formula (4) always holds. In the case we are considering, the variables in the sum are independent, and this implies that their covariance is zero.

Using this in (4) we get

$$Var(\hat{\mu}_n) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n}\sigma^2. \tag{5}$$

Therefore, the variance of the sample mean decreases with n.

Since the variance measures how much the distribution of a random variable is concentrated around its mean, this says that, as the sample size n grows, the distribution of the estimator $\hat{\mu}_n$ will concentrate around the true value μ.

Example 1 revisited

The data in the vector `smpl1` that we used in example 1 were simulated from a normal distribution with parameters $\mu = 4.5$ and $\sigma^2 = 4$. The sample average we obtained was $\hat{\mu} = 4.136$.

If we draw another sample form the same distribution, we obtain a different estimate:

```r
smpl2 <- rnorm(10,4.5,2)
round(smpl2,2)
```

```
##  [1] 5.89 6.84 3.75 4.36 5.67 5.49 1.30 3.84 7.03
## [10] 6.51
```

```r
mean(smpl2)
```

```
## [1] 5.066089
```

Example 1 revisited

If we do this again, we will get yet another different value, since the sample average, as we have seen, is a random variable.

What we just proved says that the expected value of these sample means is the true value (4.5) and that their variance is equal to the population variance (4) divided by the sample size (10), i.e., 0.4.

To show that this is the case, we will carry out a simulation to obtain an empirical approximation to the distribution of $\hat{\mu}$.

In the `R` code below, we will generate 2500 samples of size ten from the population, which has $N(4.5, 4)$ distribution and calculate the sample mean for each of these samples. This gives a sample of 2500 empirical averages of size 10.

Example 1 revisited

```
smpl.mat <- matrix(rnorm(25000,4.5,2), ncol = 10)
mean.vec <- apply(smpl.mat,1,mean)
hist(mean.vec, breaks = 20, freq = FALSE,
     xlab = 'Sample mean',
     main = 'Histogram of 25000 sample means',
     xlim = c(0,10))
lines(density(mean.vec, adjust = 1.5), col = 'red',
      lwd = 2)
curve(dnorm(x,4.5,2), 0, 10, add = TRUE, col = 'blue',
      lwd = 2)
```

Example 1 revisited

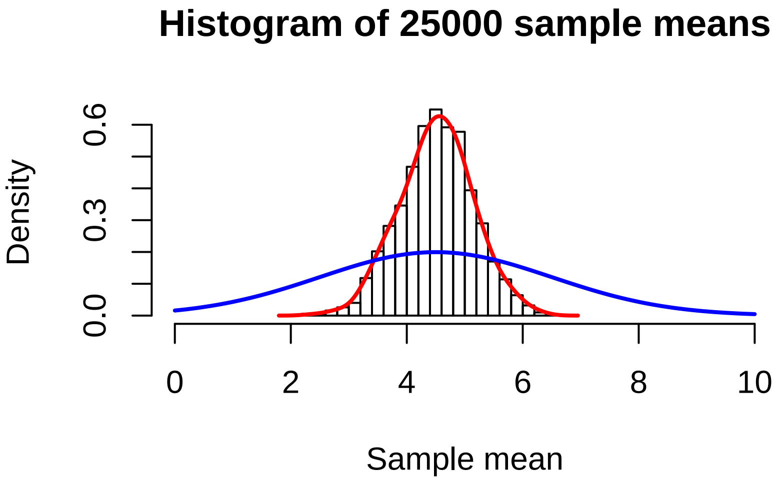

Figure 1: Histogram of 25000 empirical means for samples of size 10. In blue, population distribution, in red, estimated sampling density.

Example 1 revisited

The figure shows that there is much less variability in the means that in the sample, as the variance has considerably reduced, even though the size of the sample is only 10.

The red curve looks like a normal distribution, and we will see in the next section that this is indeed the case.

Sampling distribution for the mean

Sampling distribution for the mean of a normal sample

When data come from a normal distribution, it is possible to obtain the sampling distribution for the average of a random sample.

To do this, we need the following property of independent normal random variables: Let X and Y be independent random variables with means μ_X, μ_Y, and variances σ_X^2, σ_Y^2, respectively. Then their sum $X + Y$ also has a normal distribution with mean $\mu_X + \mu_Y$, and variance $\sigma_X^2 + \sigma_Y^2$.

We can use this property to derive the sampling distribution for the mean. Equation (1) shows that $\hat{\mu}$ is a sum of independent normal random variables.

Therefore, the distribution of the sum will also be normal, and we calculated the mean and variance for $\hat{\mu}$. We conclude that

$$\hat{\mu} \sim N(\mu, \sigma^2/n).$$

Asymptotic sampling distribution for the mean

What happens if the original sample does not come from a normal distribution?

Let $X_1, X_2, \ldots, X_n$ be independent random variables, and suppose that all the variables have the same distribution with mean μ and variance σ^2, but we do not assume that they have a normal distribution. As before, the sample average is given by

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

This variable has, approximately, a normal distribution. This is a consequence of the Central Limit Theorem, one of the fundamental results of Probability Theory.

Central Limit Theorem

Central Limit Theorem

Let $X_n, n \geq 1$ be a sequence of independent random variables having the same distribution with mean μ and variance σ^2. Then,

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

converges to a standard normal distribution, as $n \to \infty$. We denote this by

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{w} N(0,1),$$

where the w stands for weak convergence, which is convergence in distribution. What this means is that for any number x, and n large,

$$P\Big(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq x\Big) \approx \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx.$$

Central Limit Theorem

We can also say that for a sample $X_1, \ldots, X_n$ of size n, all variables identically distributed with mean μ and variance σ^2, the (sampling) distribution for the average $\bar{X}_n$ is approximately normal with mean μ and variance σ^2/n, if the size of the sample n is large.

We denote this by

$$\bar{X}_n \approx N(\mu, \sigma^2/n)$$

The standard deviation of the sampling distribution is called the **standard error** and plays an important role in Statistics.

Simulation Example

We can use stochastic simulation to give an example of how the Central Limit Theorem works.

We simulate 10000 samples of size 20 from an exponential distribution with parameter one and estimate the sample means.

Then we plot a histogram, and the estimated density for the sample means and compare them with a normal distribution.

Recall that the mean and variance for an exponential distribution of parameter 1 are both 1.

```
exp.mean <- numeric(10000)
for (i in 1:10000) {
  exp.mean[i] <- mean(rexp(20))}
```

Simulation Example

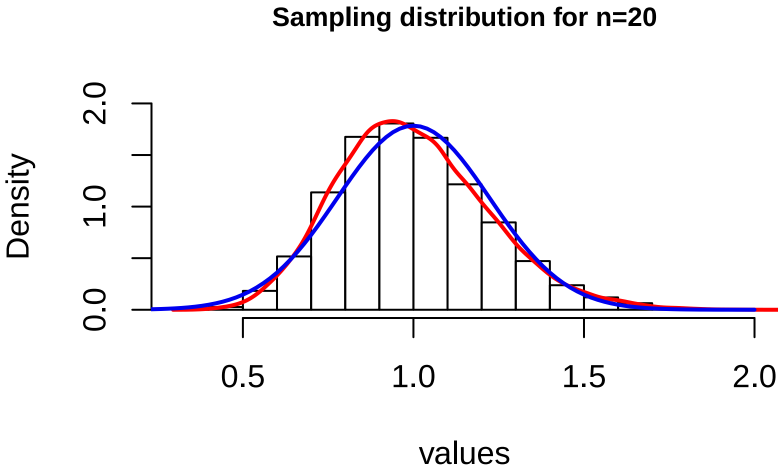

Figure 2: Sampling distribution for the empirical mean of exponential samples. The parameter for the exponential distribution is 1 and the sample size is 20. The blue curve represents the normal density.

Simulation Example

We repeat this simulation with samples of size 50.

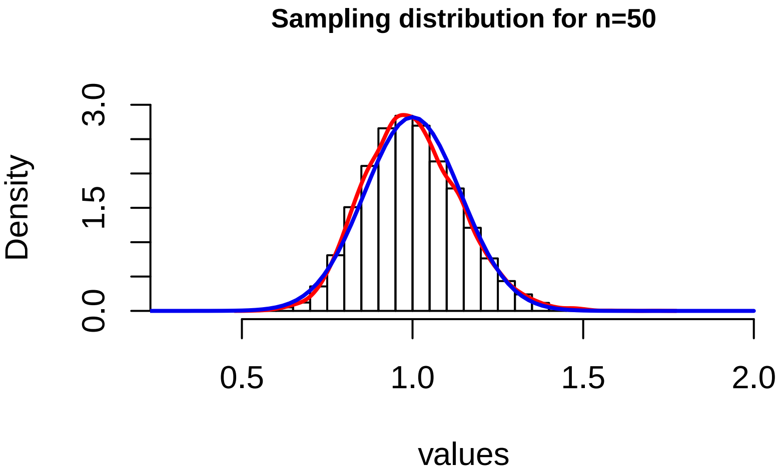

Figure 3: Sampling distribution for the empirical mean of exponential samples. The parameter for the exponential distribution is 1 and the sample size is 50. The blue curve represents the normal density.

Simulation Example

Now the fit is better. Observe also how the standard deviation of the sampling distribution decreases as the sample size increases.

If we sample from a standard normal distribution, then, as we saw before, the average value for the sample also follows a normal distribution with mean zero and variance $1/n$.

In this case, the normal distribution is not an approximation to the sampling distribution: it is precisely the sampling distribution. Therefore, we can plot the changes in the distribution as the sample size increases.

Simulation Example

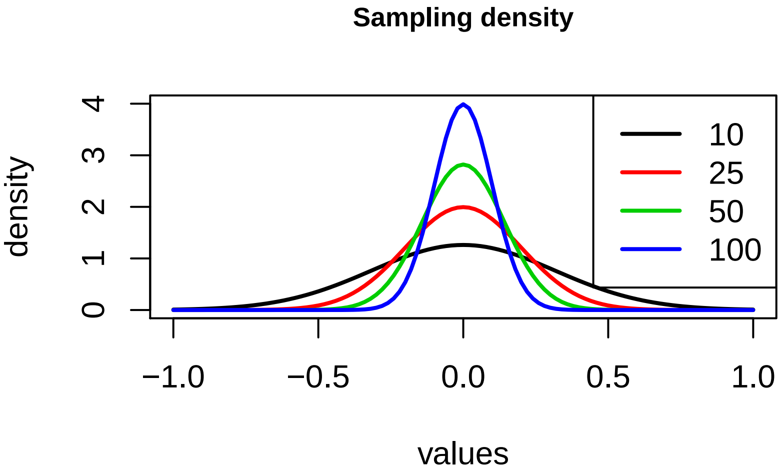

Figure 4: Sampling density for the empirical mean of standard normal samples for different sample sizes.

Summary

Summary

Let $X_1, \ldots, X_n$ be iid rv's with mean μ and variance σ^2.

1. The sampling density of $\bar{X}_n$ has mean μ .
2. The standard deviation of the sampling density, known as the standard error, is $\sigma/\sqrt{n}$.
3. When n is large, the sampling density of $\bar{X}_n$ approaches the normal distribution, regardless of the distribution of the population.
4. When the population distribution is normal, so is the sampling density for any value of n.
5. If the variance is not known, and we estimate it from the sample, the sampling distribution for the normalized sample mean is t_{n-1}.
6. When n is small, and the population distribution is not normal, we cannot assume that the sampling distribution is normal.

Summary

A Rule of Thumb

For a sample size $n > 30$, the sampling distribution for the mean for any population with mean μ and variance σ^2 can be approximated by a normal distribution with mean μ and variance σ^2/n.

When n is small, and the population is not normal, all we can say is that

1. The mean of the sampling density of the mean equals μ, the mean of the population.
2. The standard deviation of the sampling distribution of the mean is σ^2/n.

One way to proceed in this case is to use the bootstrap, a technique that will study later in this course.