

[POINTEST]

(a) Consider the statistical functional  $T(\nu)$  which returns the second moment of  $\nu$  (in other words,  $T(\nu) = E[X^2]$  where  $X$  is  $\nu$ -distributed), and let  $\theta = T(\nu)$ . Is the plug-in estimator of  $\theta$  biased? Is it consistent?

(b) Now consider the estimator  $\hat{\theta}$  of  $\theta$  which is defined to be the sum of (i) the square of the plug-in estimator of the mean of  $\nu$  and (ii) the plug-in estimator of the variance of  $\nu$ . Is  $\hat{\theta}$  biased? Is it consistent?

(a) The plug-in estimator of  $\theta$  is  $\frac{1}{n} \sum_{i=1}^n X_i^2$ , which is unbiased by linearity of expectation and consistent by the law of large numbers.

(b) We have

$$\begin{aligned}\hat{\theta} &= \bar{X}^2 + \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \\ &= \bar{X}^2 + \frac{1}{n} \sum_{i=1}^n X_i^2 - 2\bar{X} \frac{1}{n} \sum_{i=1}^n X_i + \bar{X}^2 \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2.\end{aligned}$$

Therefore, this estimator is actually the same as the estimator in (a), and it is therefore also unbiased and consistent.

(a)  $X_1, X_2, \dots$  are i.i.d. Bernoulli random variables with unknown  $p$  and estimator

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i$$

(b)  $X_1, X_2, \dots$  are i.i.d.  $\mathcal{N}(\mu, \sigma^2)$ , with unknown  $\mu$  and  $\sigma^2$  and estimator

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

(c)  $X_1, X_2, \dots$  are i.i.d. uniform random variables on an unknown bounded interval. For  $n \geq 100$  we estimate the mean using

$$\hat{\mu} = \frac{\sum_{i=1}^{100} X_i}{100}$$

(d)  $X_1, X_2, \dots$  are i.i.d.  $\mathcal{N}(\mu, \sigma^2)$ , with unknown  $\mu$  and  $\sigma^2$ . For  $n \geq 100$  we estimate the standard deviation using

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{100} (X_i - \bar{X})^2}{99}}$$

(a) **Unbiased and consistent.** The expectation of  $\hat{p}$  is  $(1/n)(np) = p$ , and the variance converges to 0 since  $\hat{p}$  is an average of i.i.d., finite-variance random variables. Therefore, the mean squared error converges to 0 as  $n \rightarrow \infty$ .

(b) **Biased and consistent.** The estimator is biased because its value is always slightly smaller than the unbiased estimator (which has  $n - 1$  instead of  $n$  in the denominator). The estimator is nevertheless consistent, since the bias and the variance both converge to 0 as  $n \rightarrow \infty$ .

(c) **Unbiased and inconsistent.** The mean of  $\hat{\mu}$  is  $(1/100)(100\mu) = \mu$ , so the estimator is unbiased. The variance isn't zero and doesn't depend on  $n$ , so it cannot converge to 0 as  $n \rightarrow \infty$ . Therefore, the estimator is inconsistent.

(d) **Biased and inconsistent.** This estimator is inconsistent for the same reason as (c). The bias is trickier. Since the variance of  $\hat{\sigma}$  is positive, then we have

$\mathbb{E}[\hat{\sigma}^2] - \mathbb{E}[\hat{\sigma}]^2 > 0$ , which implies that

$$\mathbb{E}[\hat{\sigma}]^2 < \mathbb{E}[\hat{\sigma}^2] = \mathbb{E} \left[ \frac{1}{99} \sum_{i=1}^{100} (X_i - \bar{X})^2 \right] = \sigma^2$$

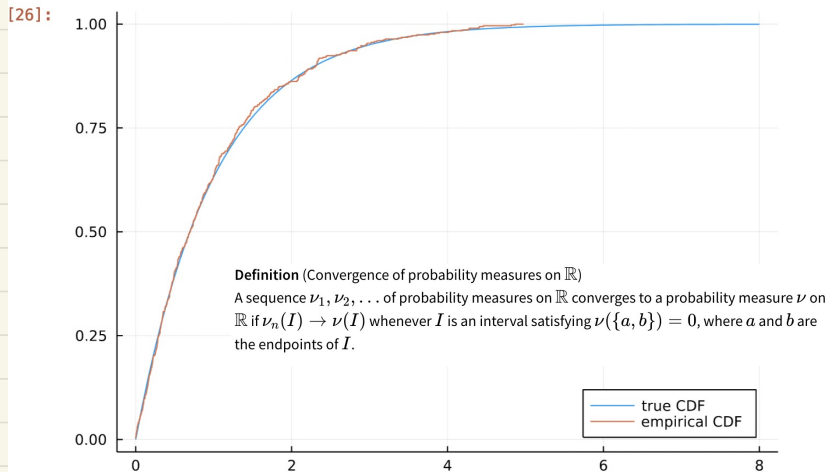
Thus the bias of  $\hat{\sigma}$  is negative.

### Example

Draw 500 independent observations from an exponential distribution with parameter 1. Plot the function  $\hat{F}$  which maps  $x$  to the proportion of observations at or to the left of  $x$  on the number line. We call  $\hat{F}$  the **empirical CDF**. Compare the graph of the empirical CDF to the graph of the CDF of the exponential distribution with parameter 1.

```
[26]: using Plots, Distributions
```

```
n = 500
xs = range(0, 8, length=100)
plot(xs, x-> 1-exp(-x), label = "true CDF", legend = :bottomright)
plot!(sort(rand(Exponential(1),n)), (1:n)/n,
       seriestype = :steppre, label = "empirical CDF")
```



This example suggests an idea for estimating  $\hat{\theta}$ : since the unknown distribution  $\nu$  is typically close to the measure  $\hat{\nu}$  which places mass  $\frac{1}{n}$  at each of the observed observations, we can build an estimator of  $T(\nu)$  by plugging  $\hat{\nu}$  into  $T$ .

#### Definition (Plug-in estimator)

The **plug-in estimator** of  $\theta = T(\nu)$  is  $\hat{\theta} = T(\hat{\nu})$ .

The estimator  $\hat{\theta}_n$  is said to be a *consistent estimator* of  $\theta$  if, for any positive number  $\varepsilon$ ,

$$\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| \leq \varepsilon) = 1$$

or, equivalently,

$$\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| > \varepsilon) = 0.$$

The notation  $\hat{\theta}_n$  expresses that the estimator for  $\theta$  is calculated by using a sample of size  $n$ . For example,  $\bar{Y}_2$  is the average of two observations whereas  $\bar{Y}_{100}$  is the average of the 100 observations contained in a sample of size  $n = 100$ . If  $\hat{\theta}_n$  is an unbiased estimator, the following theorem can often be used to prove that the estimator is consistent.

An unbiased estimator  $\hat{\theta}_n$  for  $\theta$  is a consistent estimator of  $\theta$  if

$$\lim_{n \rightarrow \infty} V(\hat{\theta}_n) = 0.$$

Suppose that  $\hat{\theta}_n$  converges in probability to  $\theta$  and that  $\hat{\theta}'_n$  converges in probability to  $\theta'$ .

- a**  $\hat{\theta}_n + \hat{\theta}'_n$  converges in probability to  $\theta + \theta'$ .
- b**  $\hat{\theta}_n \times \hat{\theta}'_n$  converges in probability to  $\theta \times \theta'$ .
- c** If  $\theta' \neq 0$ ,  $\hat{\theta}_n/\hat{\theta}'_n$  converges in probability to  $\theta/\theta'$ .
- d** If  $g(\cdot)$  is a real-valued function that is continuous at  $\theta$ , then  $g(\hat{\theta}_n)$  converges in probability to  $g(\theta)$ .

Suppose that  $U_n$  has a distribution function that converges to a standard normal distribution function as  $n \rightarrow \infty$ . If  $W_n$  converges in probability to 1, then the distribution function of  $U_n/W_n$  converges to a standard normal distribution function.