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

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

(a) The plug-in estimator of θ 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

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

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

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

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

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

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

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

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

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

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

(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 $n \to \infty$.

(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 $n \to \infty$.

(c) **Unbiased and inconsistent**. The mean of $\widehat{\mu}$ 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 $n \to \infty$. 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 $\widehat{\sigma}$ is positive, then we have
$\mathbb{E}[\hat{\sigma}^2] - \mathbb{E}[\hat{\sigma}]^2 > 0$, which implies that

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

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

Example

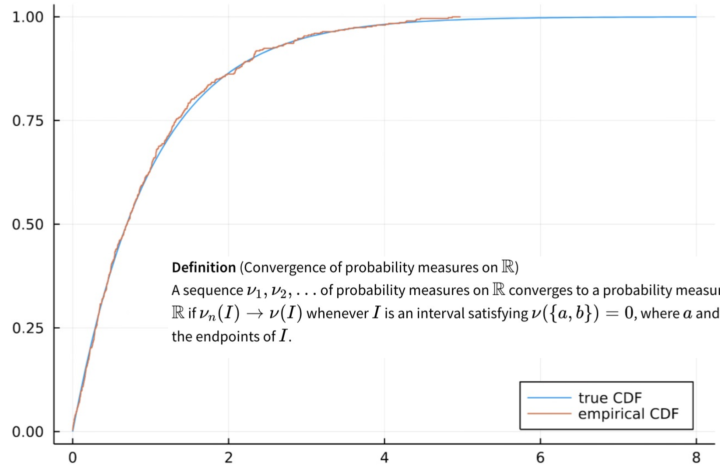
Draw 500 independent observations from an exponential distribution with parameter 1. Plot the function $\widehat{F}$ which maps x to the proportion of observations at or to the left of x on the number line. We call $\widehat{F}$ the **empirical CDF**. Compare the graph of the empirical CDF to the graph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

[26]:

```julia
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")
```

[26]:



Definition (Convergence of probability measures on $\mathbb{R}$)
A sequence $\nu_1, \nu_2, \ldots$ of probability measures on $\mathbb{R}$ converges to a probability measure ν on $\mathbb{R}$ if $\nu_n(I) \to \nu(I)$ whenever I is an interval satisfying $\nu(\{a, b\}) = 0$, where a and b are the endpoints of I.

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

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

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

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

or, equivalently,

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

The notation $\hat{\theta}_n$ expresses that the estimator for θ is calculated by using a sample of size n. For example, $\overline{Y}_2$ is the average of two observations whereas $\overline{Y}_{100}$ 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 θ is a consistent estimator of θ if

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

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

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 θ/θ'.
d If $g(\cdot)$ is a real-valued function that is continuous at θ, 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 $n \to \infty$. If W_n converges in probability to 1, then the distribution function of U_n/W_n converges to a standard normal distribution function.

Consider the problem of flipping a coin n times and estimating the probability of heads, i.e. $X_i \sim \text{Bernoulli}(p)$. Consider the following estimators:

(a) $\hat{p_1} = 0.5$

(b) $\hat{p_2} = \frac{1}{n}\sum_{i=1}^{n} X_i$

(c) $\hat{p_3} = \frac{1}{2}\hat{p_1} + \frac{1}{2}\hat{p_2}$

Calculate the bias, variance, are consistent?

(a)

Bias = $E[\hat{p_1}] - p = E[0.5] - p = 0.5 - p$

Var = $Var[\hat{p_1}] = Var[0.5] = 0$

MSE = $Bias^2 + Var = (0.5 - p)^2$

(b)

Bias = $E[\hat{p_2}] - p = E[\frac{1}{n}\sum_{i=1}^{n} X_i] - p = \frac{1}{n}\sum_{i=1}^{n} E[X_i] - p = p - p = 0$

Var = $Var[\hat{p_2}] = Var[\frac{1}{n}\sum_{i=1}^{n} X_i]$
$= \frac{1}{n^2}Var[\sum_{i=1}^{n} X_i] = \frac{1}{n^2}\sum_{i=1}^{n} p(1-p) = \frac{p(1-p)}{n}$

MSE = $Bias^2 + Var = \frac{p(1-p)}{n}$

(c)

Bias = $E[\frac{1}{2}\hat{p_1} + \frac{1}{2}\hat{p_2}] - p$
$= \frac{1}{2}(E[\hat{p_1}] + E[\hat{p_2}]) - p = \frac{1}{2}(0.5 - p) - p = 0.25 - 1.5p$

Var = $Var[\frac{1}{2}\hat{p_1} + \frac{1}{2}\hat{p_2}] = \frac{1}{4}Var[\hat{p_1} + \hat{p_2}] = \frac{1}{4}Var[0.5 + \frac{1}{n}\sum_{i=1}^{n} X_i]$
$= \frac{1}{4}Var[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{p(1-p)}{4n}$

MSE = $Bias^2 + Var = (0.25 - 1.5p)^2 + \frac{p(1-p)}{4n}$

We can see that $\hat{p_2}$ is consistent by the law of large numbers.
$\hat{p_1} \to 0.5$ and $\hat{p_3} \to 0.25 + 0.5p$ so the other estimators are not consistent.