8 Model Inference

Goal: Approximate the sampling distribution of statistics without straightforward closed forms.

Ex For iid samples $X_1, \ldots, X_n$ from some dist w/ mean μ & variance σ^2, might use

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$ to estimate μ.

95% approx CI for μ? $\bar{X} \pm 2 \, \hat{\sigma}/\sqrt{n}$

how do we know? $\mathrm{Var}\,\bar{X} = \sigma^2/n$

But what about median$\{X_1, \ldots, X_n\}$?

What is $\mathrm{Var}(\text{median}\{X_1, \ldots, X_n\})$?

<u>Basic idea</u> : Approximate sampling dist. of a statistic using resampling methods.

$\boxed{\text{Setup}}$

If $X_1, \dots, X_n$ are iid samples from dist F [think of F as cdf w/ pdf f], seek to estimate

$\Theta = T(F)$.

$\boxed{\text{Ex}}$ Population mean $\mu = \Theta$ or variance $\Theta = \sigma^2$

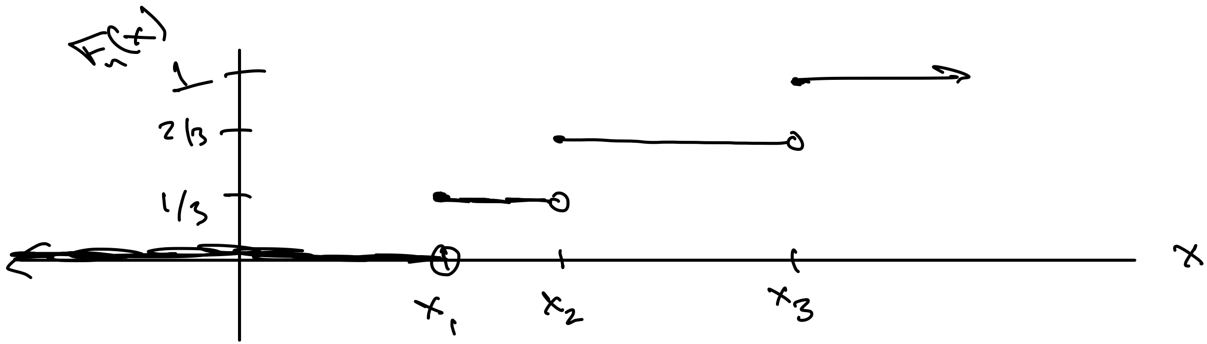
$$\mu = \Theta = \int x f(x)\,dx = \int x\,dF(x) = T(F)$$

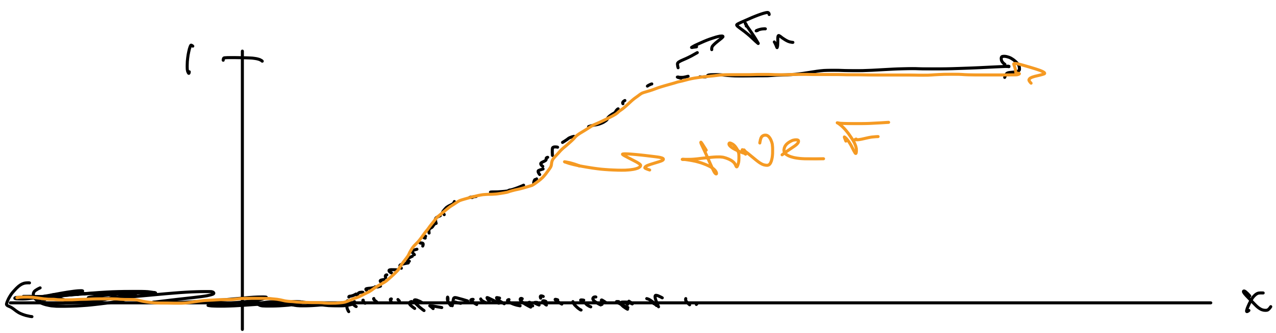
$$\sigma^2 = \Theta = \int (x-\mu)^2 f(x)\,dx = \int (x-\mu)^2\,dF(x) = T(F)$$

note: as we change F we get different μ/σ^2s

DEF Given data $X_1 = x_1, \dots, X_n = x_n$, the empirical cdf is $F_n(x)$:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}[x \geq x_i]$$





Remark To estimate $\theta = T(F)$ we often use the "plug-in" estimator $\hat{\theta}_n = T(F_n)$.

Ex

$$\theta = \int x \, dF(x) = T(F)$$

$$\hat{\theta} = \int x \, dF_n(x) = \bar{x} = T(F_n)$$

8.1 Jackknife

If $\underline{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ are iid samples from F, then

to estimate θ we use $\hat{\theta} = s(\underline{x})$ ($=$ statistic based on $\underline{x}$).

The <u>jackknife</u> yields estimates of <u>bias</u> + <u>standard error</u>

of $\hat{\theta}$, and works best for smooth functions of F (e.g. moments)

$\boxed{\text{Recall}}$

Bias of $\hat{\theta}$ is $\quad \text{bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$

$\boxed{\text{EX}}$ $x_1, \ldots, x_n$ iid mean μ var σ^2 [not necessarily normal!]

$\bar{X}$ has 0 bias b/c $E\bar{X} = \mu$

But $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ has $E\,\hat{\sigma}^2 = \frac{n-1}{n}\sigma^2$

so $\quad \text{bias}(\hat{\sigma}^2) = E(\hat{\sigma}^2) - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2$

$$= \left(\frac{n-1}{n} - 1\right)\sigma^2 = \left(\frac{n-1-n}{n}\right)\sigma^2$$

$$= -\frac{1}{n}\sigma^2$$

$\Rightarrow \hat{\sigma}^2$ is biased

DEF Define the leave-one-out obs as

$$\underline{x}_{(i)} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{i-1} \\ x_{i+1} \\ \vdots \\ x_n \end{pmatrix} \rightarrow x_i \text{ removed, vector length } n-1$$

and $\hat{\theta}_{(i)} = s(\underline{x}_{(i)})$ to be the ith jackknife replicate of $\hat{\theta}$.

The jackknife estimate of bias is

$$\widehat{bias} = (n-1)\left[\hat{\theta}_{(\cdot)} - \hat{\theta}\right]$$

where $\hat{\theta}_{(\cdot)} = \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_{(i)}$

+ the jackknife estimate of standard error is

$$\widehat{SE} = \left(\underbrace{\frac{n-1}{n}} \sum_{i=1}^{n} \left(\hat{\Theta}_{(i)} - \hat{\Theta}_{(\cdot)} \right)^2 \right)^{1/2}$$

not standardizing by $\frac{1}{n}$ b/c every $\hat{\Theta}_{(i)}$ deletes only a single data point, so $\hat{\Theta}_{(i)}$ + $\hat{\Theta}$ are very similar

$\boxed{\text{Aside}}$ Now that we have $\widehat{\text{bias}}$, why not unbias $\hat{\theta}$

via

$$\hat{\theta}_{new} = \hat{\theta} - \widehat{\text{bias}} \, ?$$

B/c $\widehat{\text{bias}}$ is a statistic & it has some variance, so the MSE may get worse

8.2 Nonparametric bootstrap

We resample the vector $\underline{x}$ with replacement, call

$$\underline{x}^* = \begin{pmatrix} x_1^* \\ x_2^* \\ \vdots \\ x_n^* \end{pmatrix} = n \text{ samples from } \underline{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

$$= \text{"bootstrap dataset"}$$

So x_1 may reappear 3x & x_2 not at all in $\underline{x}^*$

$\boxed{\text{Ex}}$ $\underline{x} = \begin{pmatrix} 0 \\ 12 \\ 0.5 \end{pmatrix}$ One possible bootstrap dataset is

$$\underline{x}^* = \begin{pmatrix} 12 \\ 12 \\ 12 \end{pmatrix} \quad \text{or} \quad \underline{x}^* = \begin{pmatrix} 0.5 \\ 0.5 \\ 0 \end{pmatrix}$$

there are $3^3 = 27$ datasets possible.

$\boxed{\text{Note}}$ $\underline{x}$ containing n samples has n^n bootstrap datasets.

Way to think about this procedure

$\underline{x}$ is sample from F

$\underline{x}^*$ " " " F_n

DEF Corresponding to each bootstrap dataset

$\underline{x}^{*b}$, $b = 1, 2, \ldots, n^n$, the bth bootstrap replicate of $\hat{\theta}$ is

$$\hat{\theta}^*(b) = s(\underline{x}^{*b})$$

Remark These bootstrap replicates can be thought of as (not independent) samples from the sampling dist of $\hat{\theta}$.

$$\{\hat{\theta}^*(b)\}_{b=1}^{n^n} \text{ samples from } f(\hat{\theta})$$

$\boxed{\text{Problem}}$ n^n is too big. In practice we resample $\underline{X}$ giving $\underline{X}^{*1}, \underline{X}^{*2}, \dots, \underline{X}^{*B}$ B times and treat

$\hat{\Theta}^*(1), \dots, \hat{\Theta}^*(B)$ as an estimate of sampling dist of $\hat{\Theta}$.