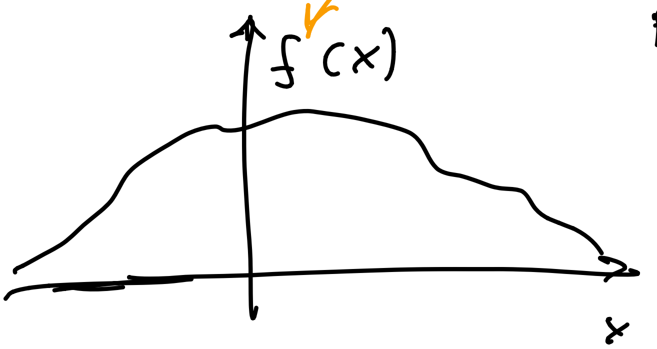
Frequentist Inference

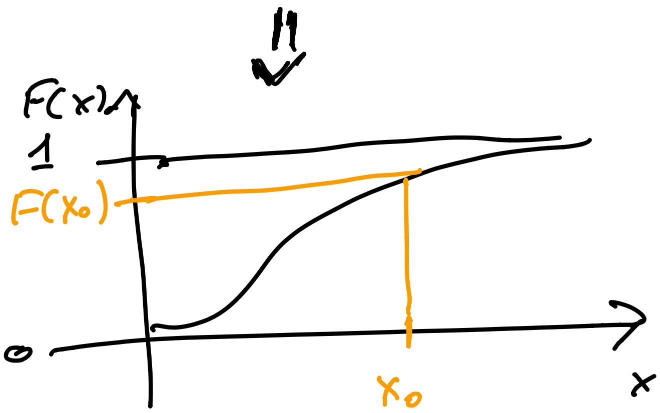
- Usually, for classical inference we assume that the observed sample comes from a probability distribution F.

(could be a named distribution or not).



Probability Density Function (pdf) → Continuous R.V.'s

Probability Mass Function (pmf) → Discrete R.V.'s



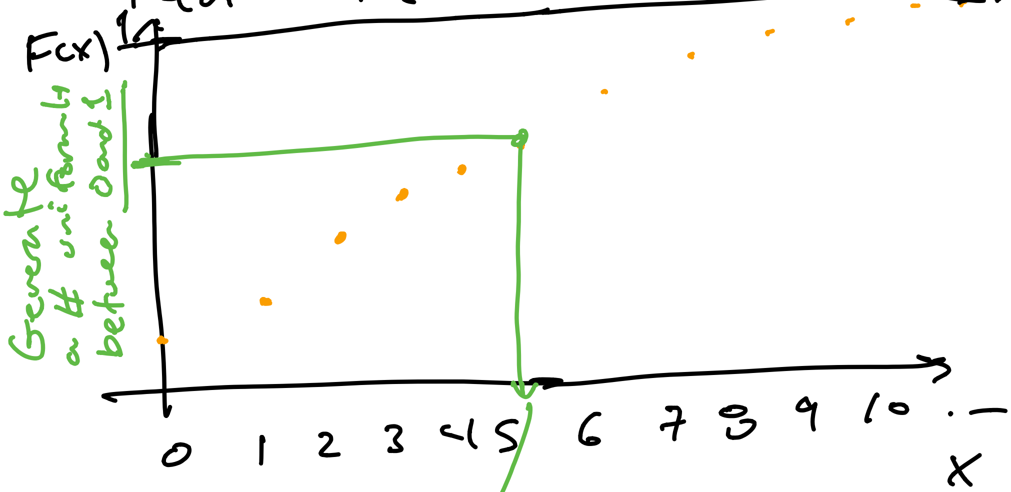
cdf (cumulative Density Function).

How much probability have we accumulated by a particular value x_0

Further, we call our data vector $\underline{x} = (x_1, x_2, \dots, x_n)$

The data vector $\underline{x}$ is a realization of a vector of random variables $\underline{X} = (X_1, X_2, \dots, X_n)$, which are distributed according to F.

Example 1: Let X be a Poisson(5) random variable, then the cdf looks like . . .

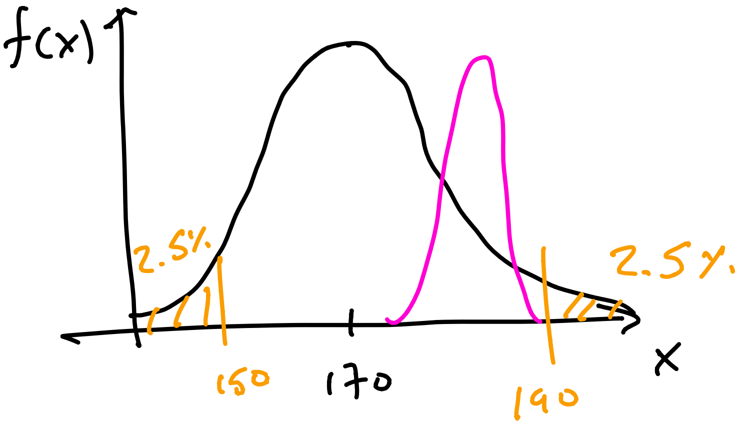


To create this plot, we needed to know that $X \sim Poi(5)$

Hey! That corresponds to $x = 5$ (realization of X).

Example 2 : Height of all students at Tec.

$X \sim Norm(170, 10)$



→ In our class a realization $x = 164$

If $\underline{x}$ is a vector . . .

$\underline{x} = (164, 173, 170, 181, 152, 190, \dots)$

In this example (as in practically all we'll be dealing with) $X_1, X_2, \dots, X_n \overset{iid}{\sim} Norm(170, 10)$

iid = Independent and identically distributed.

Simply stated:

F : A prob. distribution

X_i's : "Individual sample from F"

x_i's : The very particular values for each of the X_i's we got in our sample.

- Our objective is to learn properties of F based on $\underline{x}$, the vector of realizations.

→ A usual property we'd like to learn is the expectation of a single random draw X_i from F.

$$\theta = E_F(X)$$

. Obvious estimate is $\underbrace{\hat{\theta}}_{\text{The estimator of theta.}} = \bar{x}$ (sample mean)

- As mentioned previously $\hat{\theta}$ is calculated from $\underline{x}$ following an algorithm $t(\underline{x}) = \frac{\sum_{i=1}^{n} x_i}{n}$.

Then, our sample-specific estimator $\hat{\theta} = t(\underline{x})$ is a realization of $\underbrace{\hat{\Theta}}_{\text{"Vampire"}} = t(\underline{X})$, the estimator of a sample.

⇒ $\hat{\theta}$ is a number, while $\hat{\Theta}$ in uncertain (which we want to quantify) and accurate to a degree

$\hat{\theta} =$

- We have chosen $t(\underline{X})$ such that it is a "good" estimator of θ (the true value)

Define $\mu = E_F(\hat{\theta})$ ← $t(\underline{x})$. Then

- Bias is defined as: $\underbrace{\mu}_{\text{"The expected value of the estimator"}} - \underbrace{\theta}_{\text{True value of the parameter we are estimating.}}$

Biased estimator

- Variance is defined as: $E_F\{(\hat{\theta} - \mu)^2\}$

Bias-Variance tradeoff

- want to minimize: $\text{Bias}^2 + \text{Var}$.

- Frequentism usually makes inferences w.r.t an infinite sequence of trials.

Hypothetical datasets $\underline{X}^{(1)}, \underline{X}^{(2)}, \underline{X}^{(3)} \ldots$

generate infinite $\hat{\theta}^{(1)}, \hat{\theta}^{(2)}, \hat{\theta}^{(3)}, \ldots$

- In other words, you need F to get $\hat{\theta}$.

But $\hat{\theta}$ depends on θ and θ is a property of F.

Circular Dependency alert!

To circumvent this, we'll follow the usual

Frequentist Assumptions

① Plug-in principle. Substitute quantities that we do not know with an estimate.

Example $\sigma_{\bar{X}} = \left[\frac{Var_F(X)}{n}\right]^{1/2}$ The variance of random variable $X \sim F$ is not known, because F is not known,

$\bar{X} = \sum x_i / n$

-> To solve this, we estimate $Var_F(X)$ with our sample, using sample variance $\hat{Var}_F = \frac{\sum (x_i - \bar{X})^2}{n-1}$

$$\Rightarrow \hat{\sigma}_{\bar{X}} = \left[\frac{(x_i - \bar{X})^2}{n(n-1)}\right]^{1/2}$$ as usual :).

② Use of Taylor series approximations.

Can relate $f(X)$ to ① by local linear approximations

Example: If $\hat{\theta} = \bar{X}^2 \Rightarrow \frac{d\hat{\theta}}{d\bar{X}} = 2\bar{X}$.

Since $\bar{X}$ is a constant, we get:

$$se(\bar{X}^2) \approx 2\underbrace{|\bar{X}|}_{\text{Absolute value used}} \cdot \hat{\sigma}_{\bar{X}}$$

Absolute value used bc. $se \geq 0$.

$N(\cdot)$, $Poi(\cdot)$, $F(\cdot)$, χ^2, ...

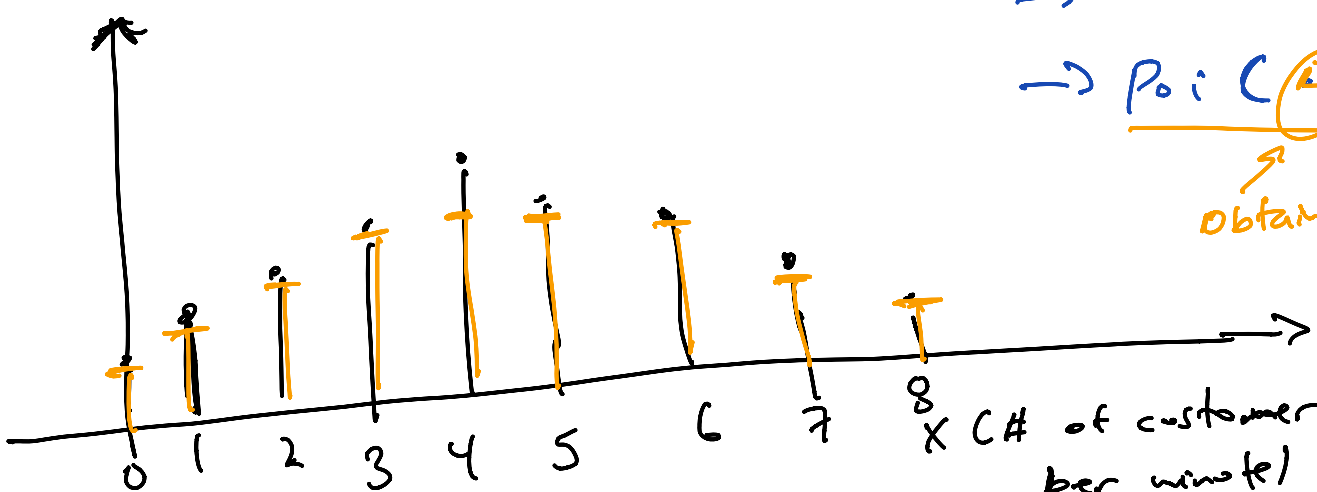
(3) Parametric Families and Maximum Likelihood Estimation (MLE)

→ Assume a parametric distribution and use the data find the most likely values for the parameters

→ Bin(·)

→ Poi(4) ~F

obtained by MLE.

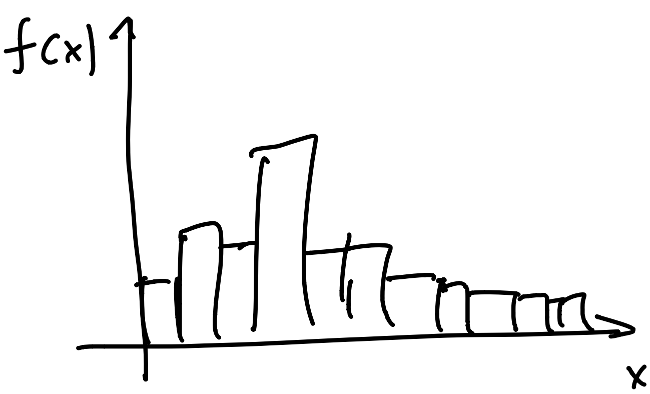


(4) Simulation and Bootstrap.

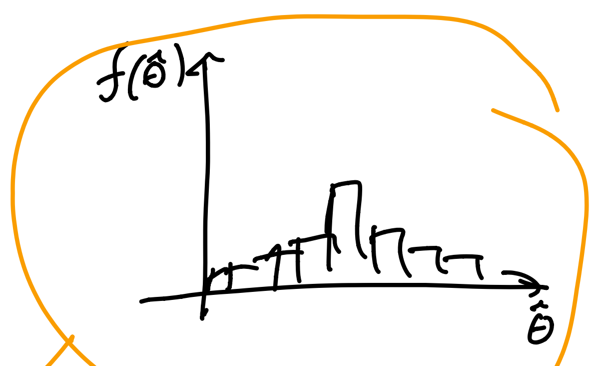
Use your sample $\underline{x}$ as an estimate $\hat{F}$ of F.

. Implement a repeated sequence of trials to get

$$\hat{\theta}^{(b)} = t(\underbrace{\underline{x}^{(b)}}_{\text{Bootstrap replicates}}) \quad \text{for } b=1,\dots,B$$



$$\Rightarrow \hat{\theta}^{(1)}, \hat{\theta}^{(2)}, \dots, \hat{\theta}^{(B)}$$



This is our estimate for the distribution of $\hat{\Theta}$

For example, $V(\hat{\Theta})$ would be estimated with the sample std dev of the $\hat{\theta}^{(b)}$ ____ .