Agenda

0. Download Rmd.
   Sections - Josh - week 4
1. Bootstrap Review and Confidence Intervals
2. Data Visualization
3. Sufficiency and Rao-Blackwell.

Bootstrap

We have some estimator and we want to construct a CI

What if it's hard to get mathematical result?

$\Rightarrow$ Bootstrap.

Use existing data to generate "new" datasets by resampling our existing dataset.

CI's

Consider $\hat{\theta}$ as our point estimate from our dataset.

Normal Interval

$$\hat{\theta} \pm z_{1-\frac{\alpha}{2}} \sqrt{V_{Boot}}$$

← sample variance of bootstrapped estimates.

standard normal quantile

Percentile Interval

$$\left(\underbrace{\hat{\theta}^*_{\frac{\alpha}{2}}}, \underbrace{\hat{\theta}^*_{1-\frac{\alpha}{2}}} \right)$$

Bootstrapped sample quantiles.

Pivotal Interval

$$\left(2\hat{\theta} - \hat{\theta}^*_{1-\frac{\alpha}{2}},\ 2\hat{\theta} - \hat{\theta}^*_{\frac{\alpha}{2}}\right)$$

Boot. Sample Quantiles.

Assumptions

1. Normal interval only works if estimator normal.

2. Percentile interval
   Assume there exists function m such that $m(\hat{\theta})$ is normal $\operatorname{var}(m(\hat{\theta}))$ is independent ~~of~~ θ.

3. Pivotal interval none of these assumptions.

Is exact if $\hat{\theta} - \theta$ is a pivot.

Estimate vs. estimator.

$T(X_1, \dots, X_n)$ ex. $= \overline{X}$

Estimate is a realization of estimator.

Def. Pivot.

Function of estimator and parameter whose dist. does not depend on parameter e.g. $X_1, \dots, X_n \sim N(\mu, 1)$

use estimator

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

note that

$$\frac{1}{n}\sum_{i=1}^{n} X_i - \mu \sim N(0,1)$$

$$\frac{\hat{\sigma}^2 \leftarrow \text{sample variance}}{\sigma \leftarrow \text{true var}} \sim \chi^2_{df.}$$

w/ MLE

$$\frac{\sqrt{n}(\hat{\theta}_{MLE} - \theta)}{\sqrt{I(\theta)^{-1}}} \dot{\sim} N(0,1)$$

Data Visualization

1. Histograms.
2. Box plots.
3. QQ plots.

Quantile Review

Theoretical:

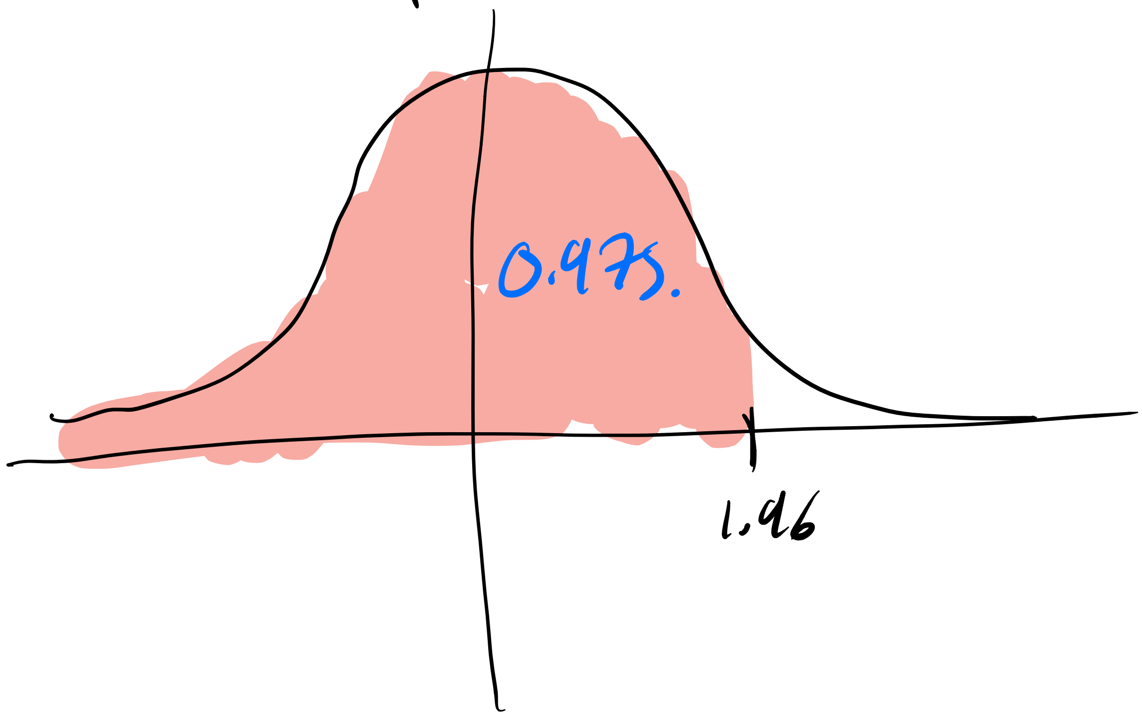
The quantile function is the inverse CDF.

Suppose $X \sim F$

$$F(z) = P(X \le z) = p$$

$F^{-1}(p) = z.$

e.g. 0.975 quantile $N(0,1)$
approx 1.96.



Sample:

Suppose $X \sim F$

$$P(F(X) \le z) = P(X \le F^{-1}(z))$$
$$= F(F^{-1}(z))$$

$= z.$

cdf of $F(X)$ is z

$\Rightarrow F(x) \sim unif(0,1).$

$\Rightarrow P(F(X) \leq z) = z$

$X_{(k)}$ is kth order Statistic

$$F(X_{(k)}) \approx \frac{k}{n}.$$

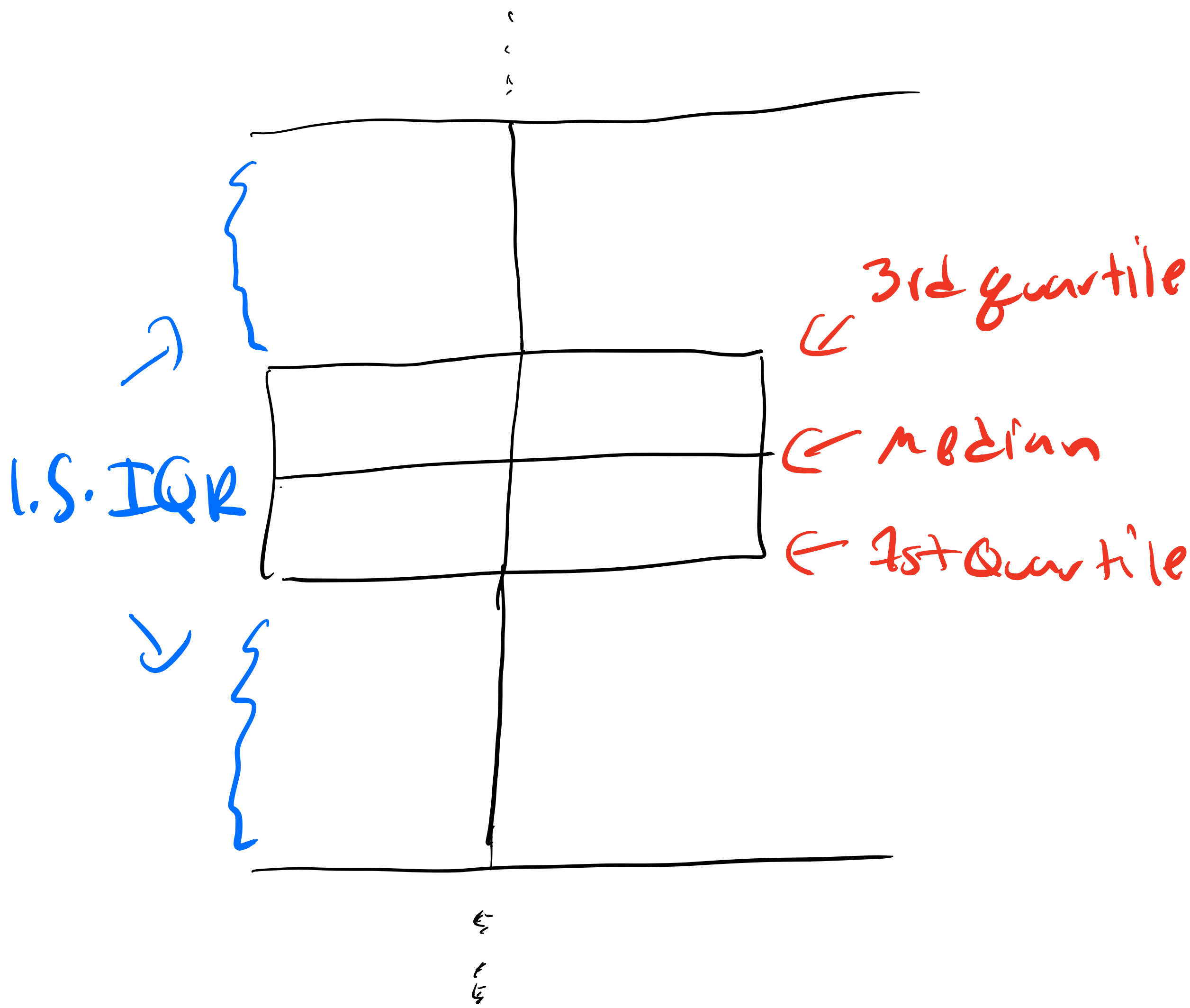
or

$$X_{(k)} = F^{-1}\left(\frac{k}{n}\right)$$

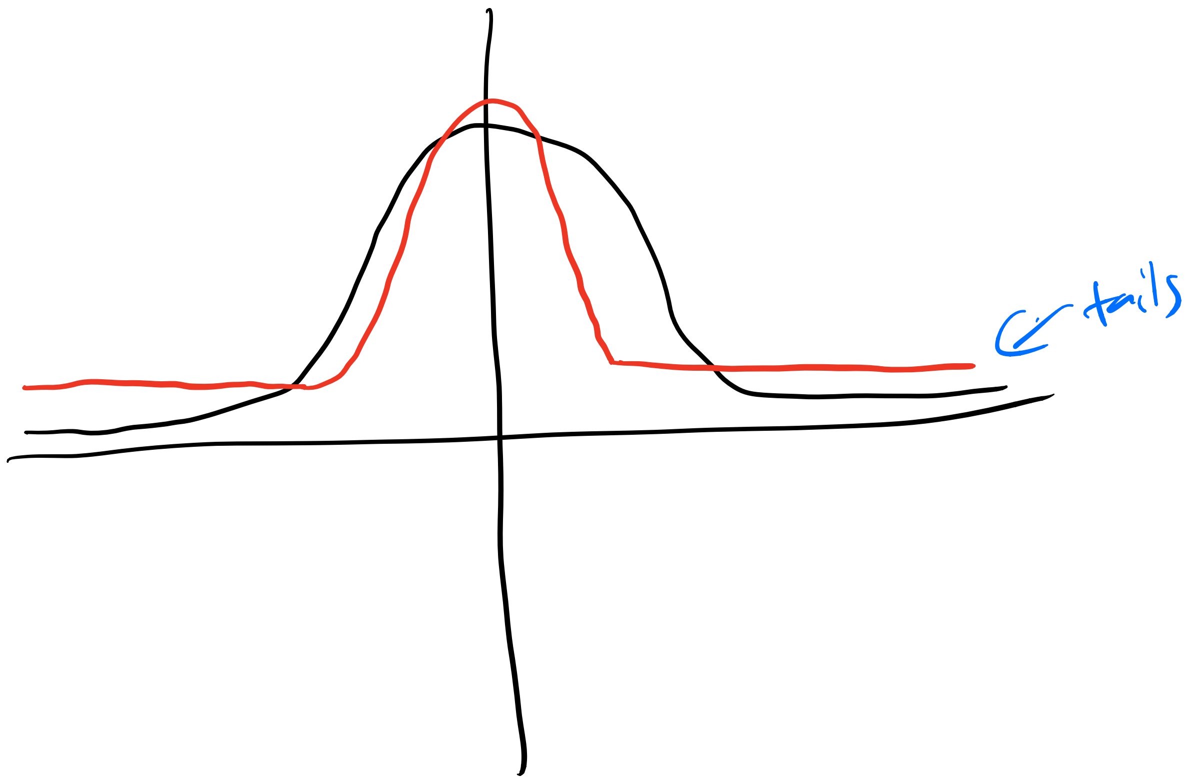
in practice we

$$X_{(k)} \approx F^{-1}\left(\frac{k}{n+1}\right)$$

Boxplots



$$IQR = \underbrace{Q_3}_{\text{3rd quartile}} - \underbrace{Q_1}_{\text{1st quartile}} = q_n(0.75) - q_n(0.25)$$

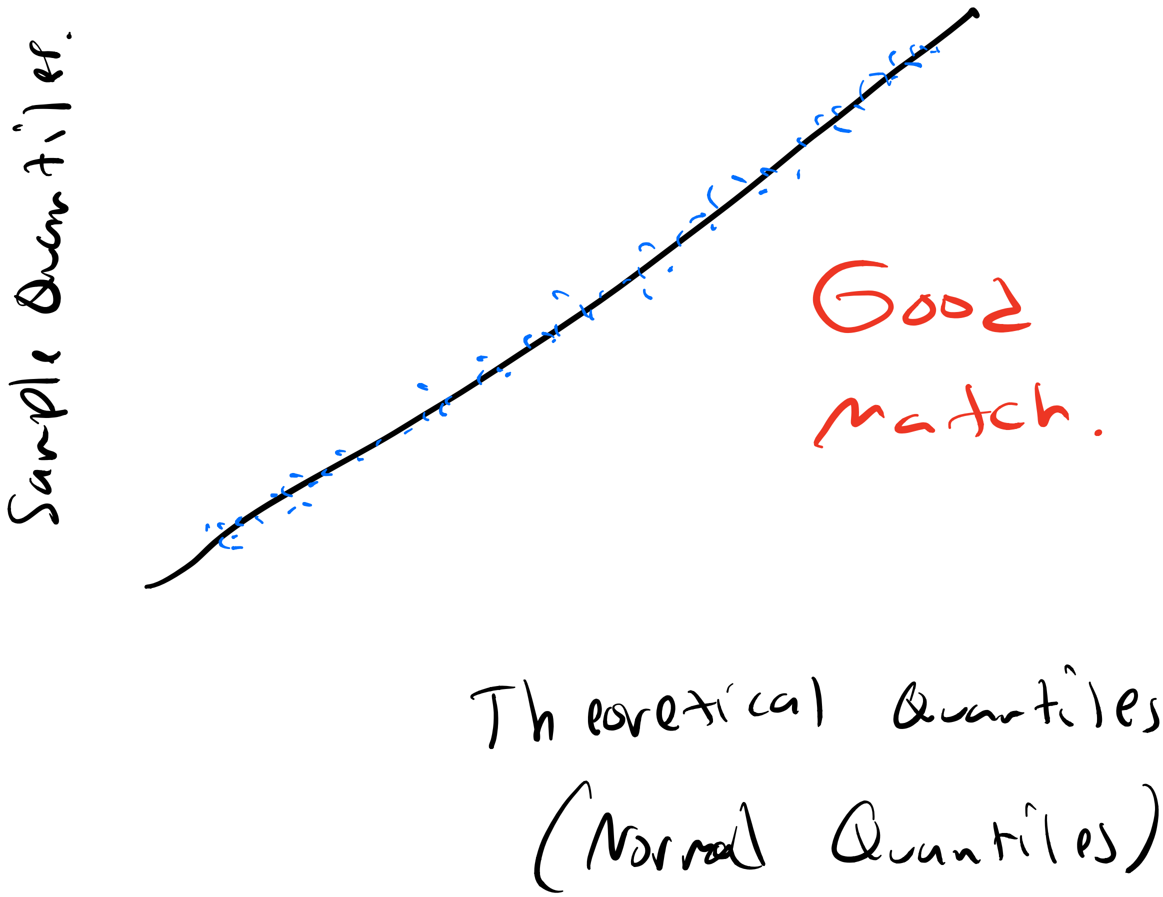


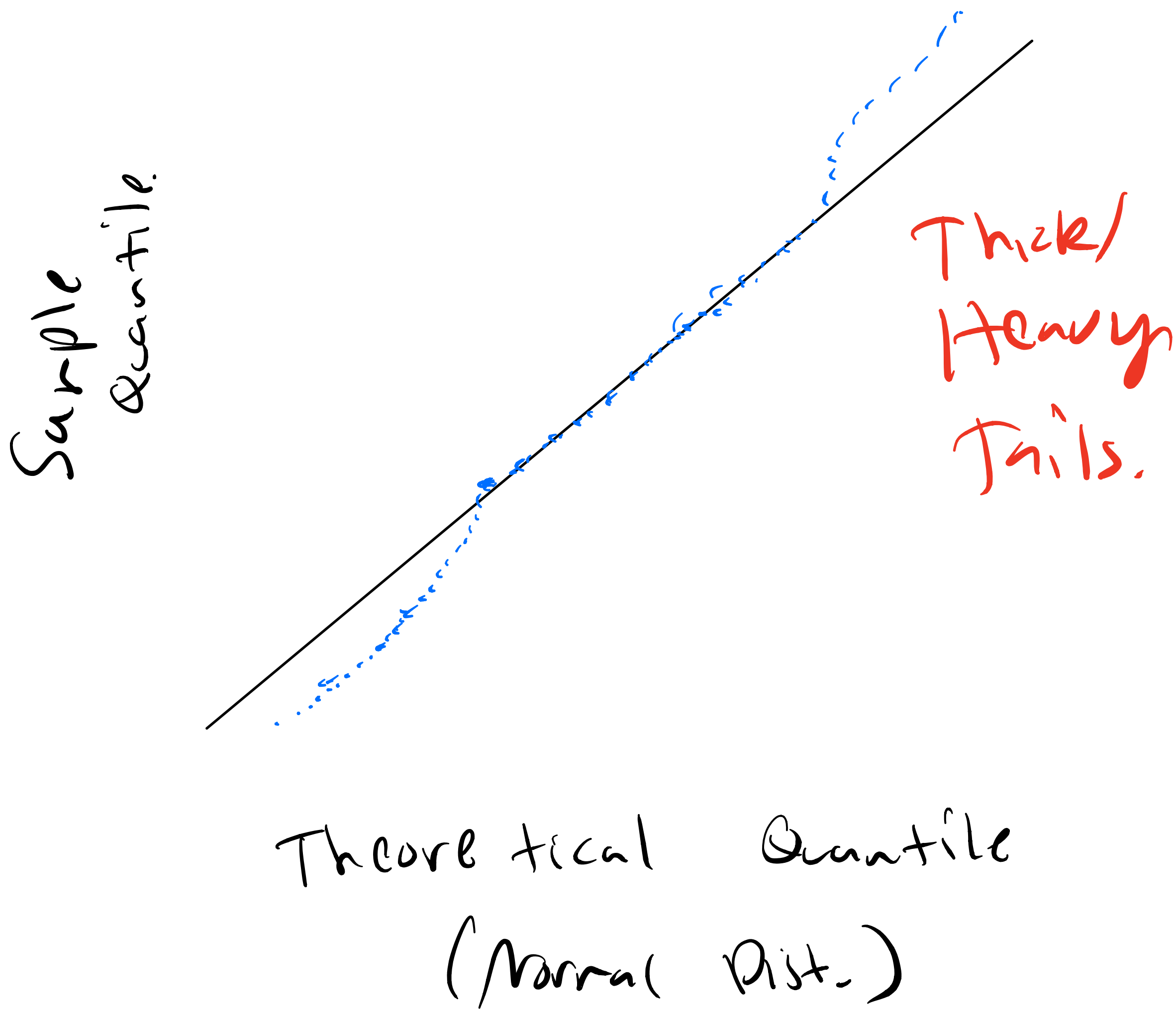
Red dist is thicker/ heavier tailed than black distribution

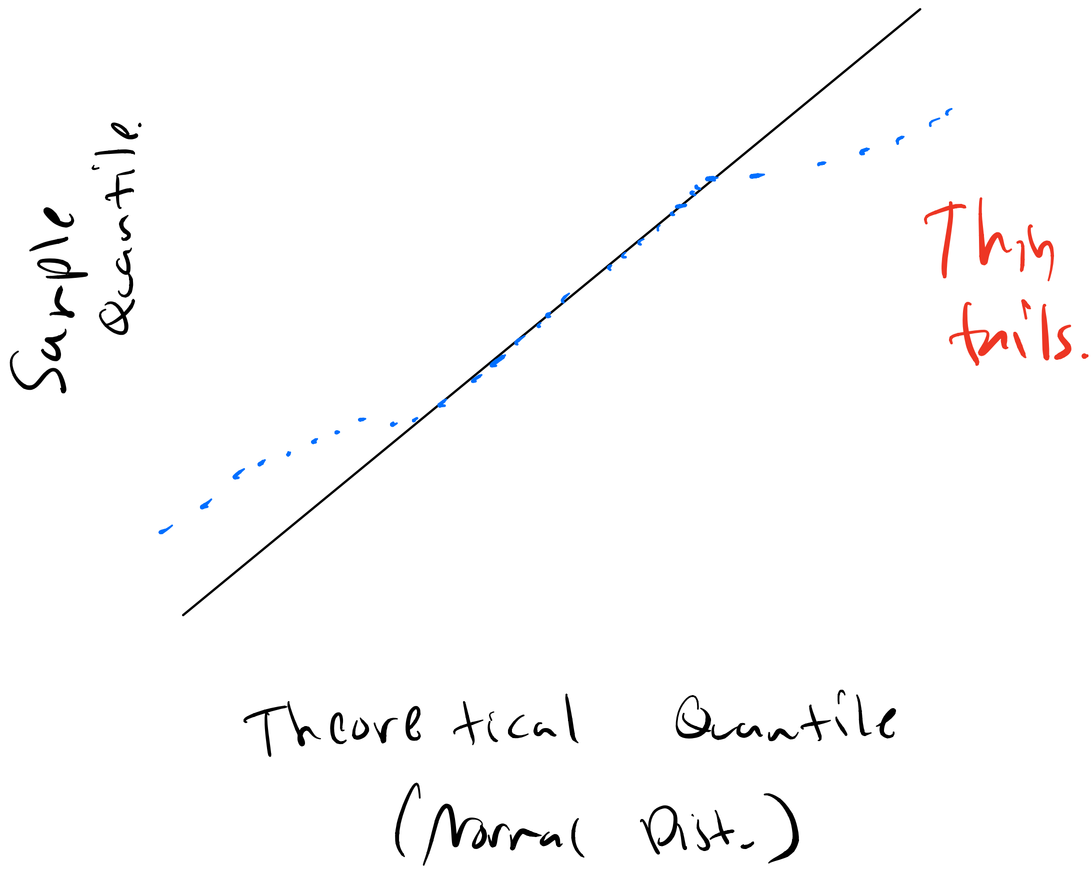
Ex. t-distribution has thick/ heavy tails.

QQ-plots

Suppose I believe data is normally distributed.




Sample Quantile
Thick/Heavy Tails.
Theoretical Quantile
(Normal Dist.)



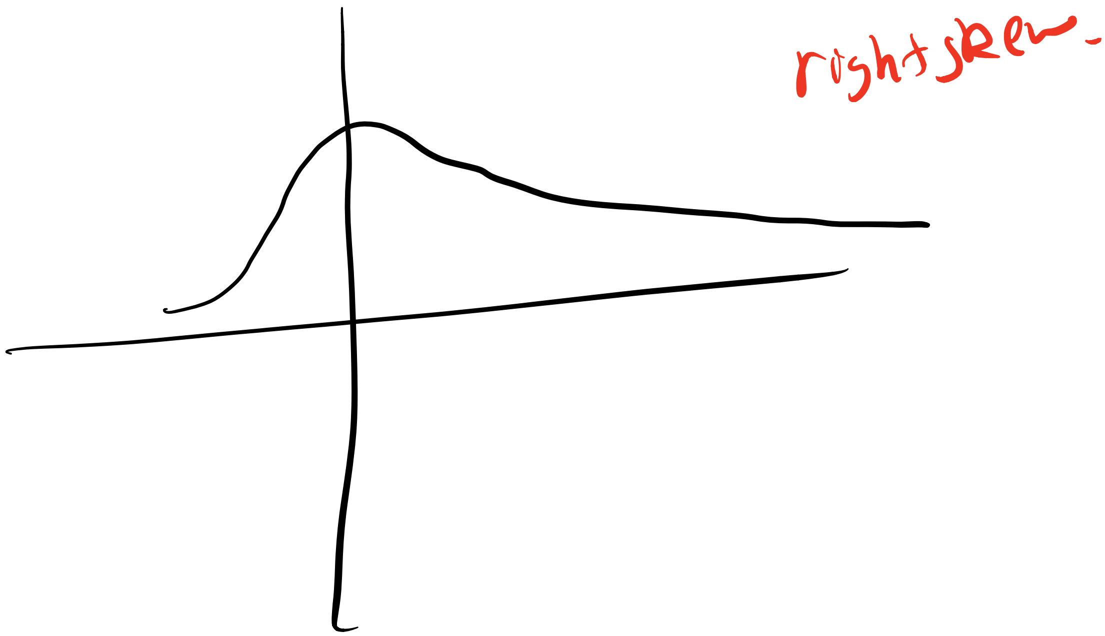
Ex. of point here

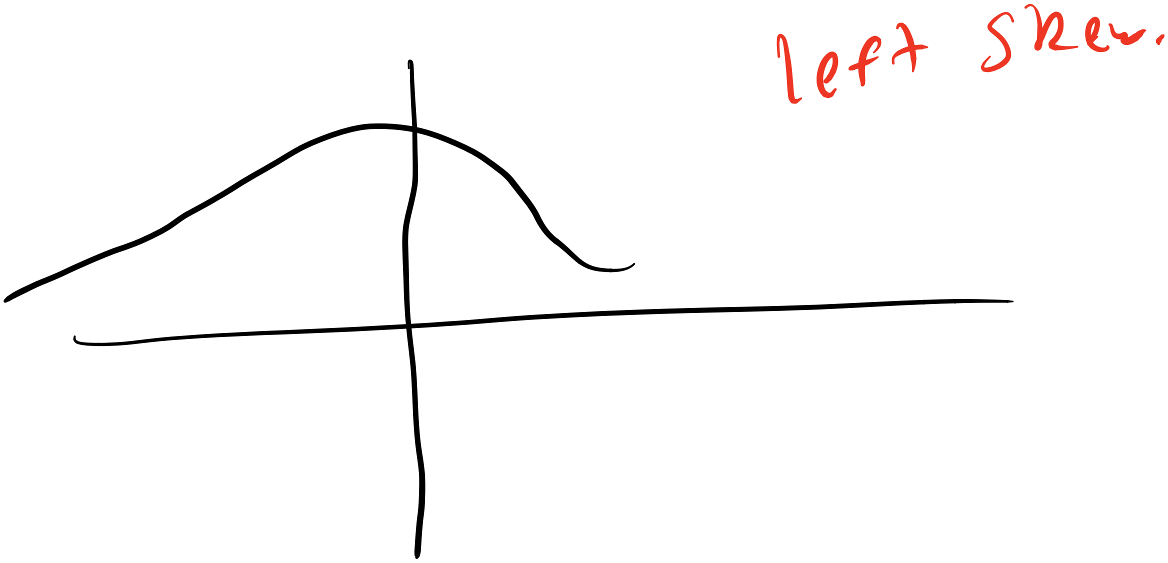
Suppose $X_{(1)}, \dots, X_{(1001)}$

$$X_{(501)} \approx F^{-1}\left(\frac{501}{1001}\right)$$

$$\left(F^{-1}\left(\frac{501}{1001}\right), X_{(501)}\right)$$

Skewed Distributions





Sufficiency

Review 4 nice things.

1. Unbiased
2. Consistent.
3. Efficient
4. Sufficient.

Data is costly to use and store. We only want what we need for estimation procedure.

Consider estimating θ from $X_1, \dots, X_n$.

Can we find some function of $t(X_1, \ldots, X_n)$ that contains all info on θ?

Def. Sufficiency

We say $T(X_1, \ldots, X_n)$ is sufficient for θ if

$$X_1, \ldots, X_n \mid \left(T(X_1, \ldots, X_n) = t\right)$$

does not depend on θ for any t.

T is called a sufficient statistic.

You can prove sufficiency with the def inition.

The whole dataset is a sufficient statistic.

$$T(X_1, \dots, X_n) = (X_1, \dots, X_n)$$

$$\theta = (\mu, \sigma^2)$$

$$\hat{\theta}_{MLE} = \begin{pmatrix} \bar{X} \\ \frac{1}{n} \sum_{i=1} (X_i - \bar{X})^2 \end{pmatrix}$$

Factorization Theorem

T is sufficient for

θ if and only if

$$f_\theta(x_1, \dots, x_n) = \underbrace{g_\theta(T)}_{\text{solely a function of } T \text{ and } \theta} \cdot \underbrace{h(x_1, \dots, x_n)}_{\text{no dep on } \theta!}$$

"θ depends on the data only through a sufficient statistic"

Ex. $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, 1)$

want to find suff. stat. for μ.

$$f_\mu(x_1, \dots, x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-1}{2}(x_i - \mu)^2\right)$$

$$= \frac{1}{(\sqrt{2\pi})^n} \cdot \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2\right)$$

$$= \frac{1}{(\sqrt{2\pi})^n} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - 2x_i\mu + \mu^2)\right)$$

$$= \frac{1}{(\sqrt{2\pi})^n} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right) \cdot \exp\left(\sum_{i=1}^{n} x_i \mu\right)$$

$$\cdot \exp\left(-\frac{1}{2}\sum_{i=1}^{n} \mu^2\right)$$

$$\sum_{i=1}^{n} x_i \mu = \mu \sum_{i=1}^{n} x_i = \mu \cdot n \cdot \bar{x}$$

$$= \overbrace{\frac{1}{(\sqrt{2\pi})^n} \cdot \exp\left(-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right)}^{h(\underline{x})}$$

$$\underbrace{\cdot \exp\left(\mu n \bar{x}\right) \cdot \exp\left(-\frac{1}{2}\sum_{i=1}^{n} \mu^2\right)}_{g_\mu(\bar{x})}$$

$$f_\theta(x_1, \ldots, x_n) = \underbrace{g_\theta(T)}_{\text{solely a function of } T \text{ and } \theta} \cdot \underbrace{h(x_1, \ldots, x_n)}_{\text{no dep on } \theta\,!}$$

Fact: Any 1-1 of a sufficient statistic is also sufficient

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

Fact: If T is sufficient for θ, $\hat{\theta}_{MLE}$ is a function of T.

why?

T suff. iff

$$f_\theta(x_1, \dots, x_n) = g_\theta(T) \cdot h(x_1, \dots, x_n)$$

$$\log f_\theta(x_1, \dots, x_n) = \log g_\theta(T) + \log h(x_1, \dots, x_n)$$

maximizing w/ respect to θ is equivalent to maximizing $\log g_\theta(T)$.

How can we measure goodness of estimator?

One way is $MSE(\hat{\theta})$.

the mean squared error

$$MSE(\hat{\theta}) = E_\theta\left((\hat{\theta}-\theta)^2\right)$$

Note we can we decompose

$$MSE(\hat{\theta}) = \text{bias}^2 + \text{var}(\hat{\theta})$$
$$= \left(E(\hat{\theta})-\theta\right)^2 + \text{var}(\hat{\theta}).$$

"bias-variance trade off"

Rao-Blackwell Theorem

Let $\hat{\theta}$ be an estimator for θ. Assume $E(\hat{\theta}^2) < \infty$

Suppose T sufficient for θ.

$$\tilde{\theta} = E(\hat{\theta} \mid T)$$

"Rao-Blackwellized Estimator"

then $MSE(\tilde{\theta}) \leq MSE(\hat{\theta})$

strict if $\tilde{\theta} \neq \hat{\theta}$

Note $\tilde{\theta} = E(\hat{\theta} \mid T)$

has no dependence on θ

because $\hat{\theta} \mid T$ has no

dependence on θ, by

sufficiency of T.

Ch 8. Q. 68.c)

Show $T = \sum_{i=1}^{n} x_i$ is sufficient for λ, $x_i \overset{iid}{\sim} \text{pois}(\lambda)$.

$$f_\lambda(x_1, \dots, x_n) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

$$= \underbrace{\lambda^{\sum x_i} \cdot e^{-n\lambda}}_{g_\lambda(\sum x_i)} \cdot \underbrace{\frac{1}{\prod_{i=1}^{n} x_i!}}_{h(x_1, \dots, x_n)}$$

$$g_\lambda\left(\sum x_i\right) \cdot h(x_1, \dots, x_n)$$