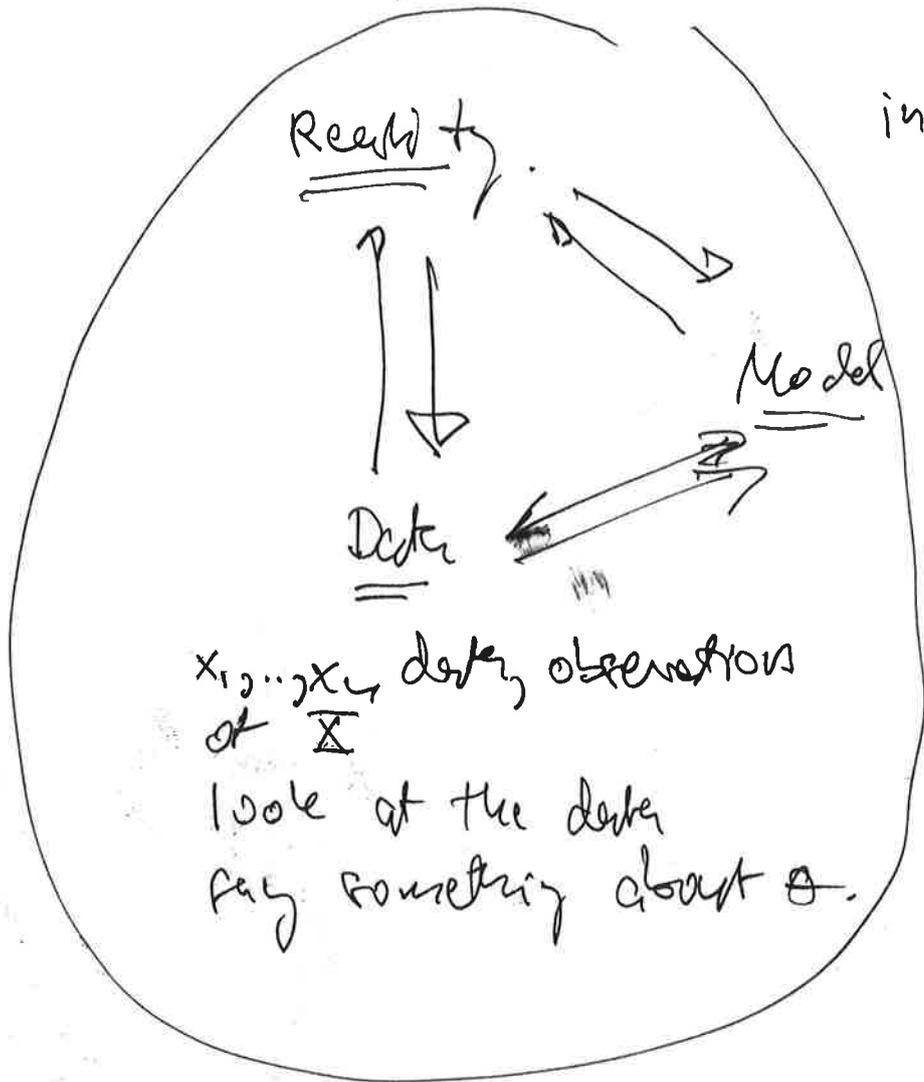


Inference theory

Lecture 1

inference theory



mathematical randomness.

X r.v.

$$F_X(x) = P(X \leq x)$$

Typically F_X is not fully known.

unknown θ parameter

(probability theory)

EX: $X \in \text{Bin}(10, p)$

unknown

#

Data come from a distribution F_θ ,
 θ unknown.

Ex: (count)

$X \in \text{Bin}(10, \theta)$, x obs of X

$$F_\theta \text{ d.f. } F_\theta(x) = \sum_{i \leq x} f_\theta(i) = \sum_{i \leq x} \binom{10}{i} \theta^i (1-\theta)^{10-i}$$

θ unknown, know that it lies in a
parameter space $\Omega = \{\text{set of allowed } \theta\}$.

Estimate θ , say something about θ ,
or more generally, about $g(\theta)$

We call the thing we want to estimate,
 $g(\theta)$, an estimand.

Def An ^{stochastic} estimator is a function
of the data x_1, \dots, x_n ,

$$T(x_1, \dots, x_n)$$

or, if x obs of X

$$T(x).$$

(2)

so

$$\delta : (x_1, \dots, x_n) \longrightarrow \{g(\theta) : \theta \in \mathcal{R}\}$$

Note that δ is supposed to be possible to calculate, in particular should ~~not~~ not depend on the unknown θ .
#

Note that F_θ is completely specified apart from knowledge about the value of θ . Therefore

if we estimate θ , we have an estimator of F_θ . Thus

$$\hat{F}_n = \delta(x_1, \dots, x_n)$$

then

$$\hat{F} = F_{\hat{\theta}_n}$$

plug-in estimator.

Plug-in approach does not always work.

EX:

x_1, \dots, x_n indep obs of $X \sim F$

Estimator of F is the empirical d.f.

$$\cancel{F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(x_i \leq x)}$$

$$F_n(u) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(x_i \leq u)$$

where

$$\mathbb{1}(x_i \leq u) = \begin{cases} 1 & \text{if } x_i \leq u \\ 0 & \text{if not} \end{cases}$$

Estimator of F . Want to estimate

$$g(\theta) = g(F) = \frac{d}{dt} (F) \Big|_{t=t_0}$$

Here we assume that X cont. r.v.
so that \exists f. density function

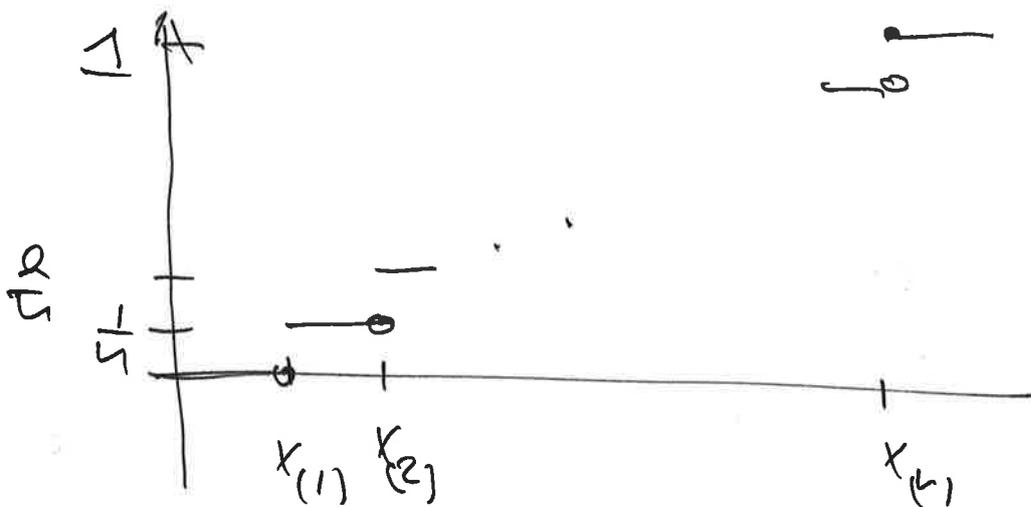
$$f = F'$$

(4)

For the plug-in approach would give us:

$$\begin{aligned} \hat{f} &= g(\hat{\theta}_n) = g(F_n) \\ &= \frac{d}{dt} F_n(t) \end{aligned}$$

Does not exist?



$$x_{(1)} \leq \dots \leq x_{(4)}$$

Note that F_n is not differentiable,
so plug-in approach does not work
#

5

We want to estimate $g(\theta)$, the
 expected value, with an estimator, $\hat{g}(\theta)$,
 $\delta(x)$ or $\delta(x_1, \dots, x_n)$.

How good is the estimator?

To measure that we introduce a
 loss function

$L(\theta, d) =$ loss when estimating
 $g(\theta)$ with the
 value $d (= \delta(x))$.

Ex:

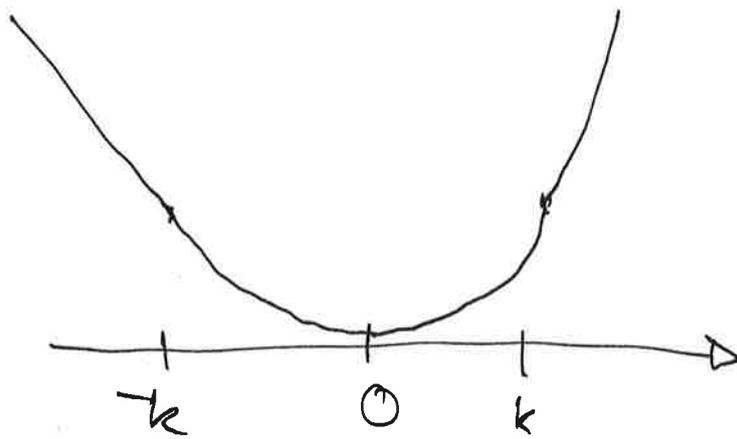
$$L_1(\theta, d) = |g(\theta) - d|$$

$$L_2(\theta, d) = (g(\theta) - d)^2$$

$$L_3(\theta, d) = (g(\theta) - d)^2 \cdot w(\theta)$$

$$L_4(\theta, d) = \begin{cases} (g(\theta) - d)^2 & \text{if } |g(\theta) - d| \leq k \\ |g(\theta) - d| + k^2 & \text{if } |g(\theta) - d| > k \end{cases}$$

weight
function



$$\text{L}_k(\theta, d) =$$

Used in robust statistics.

#

A loss function L is a function such that

$$\text{L}(\theta, d) \geq 0 \quad \text{all } \theta, d$$

$$L(\theta, g(\theta)) = 0 \quad \text{all } \theta.$$

Recall that we have an estimator, statistic,
 $\delta(X)$, or $\delta(x_1, \dots, x_n)$, and treated as
 a random element

$$\delta(X)$$

$$\delta(X_1, \dots, X_n)$$

Then the loss is also a random element

$$L(\theta, \delta(X))$$

Then we can measure how well δ estimates $f(\theta)$ by the expected loss

$$R(\theta, \delta) = E_{\theta}(L(\theta, \delta(X)))$$

is called the risk, or risk function.

Now, let us ~~try to~~ try to find a δ that minimizes

$$R(\theta, \delta)$$

for all θ . We want to define, if possible

$$\delta^* = \underset{\delta}{\operatorname{argmin}} R(\theta, \delta)$$

and such that

$$\delta^*$$

has

$$R(\theta, \delta^*) \leq R(\theta, \delta)$$

for every other estimator, and all θ .

(8)

This is not possible!

Ex:

$x = 7$ obs of $X \in \text{Bin}(10, \theta)$. Want to find an estimator $\hat{\delta}(x)$ f.t.c. $R(\theta, \hat{\delta}) \leq R(\theta, \delta)$, $\forall \delta$, $\forall \theta$. of $\theta = f(\theta)$

Suppose that loss function

$$\begin{aligned} \cancel{L(\theta, d)} &\in \theta - d \\ L(\theta, d) &= (\theta - d)^2 \end{aligned}$$

and

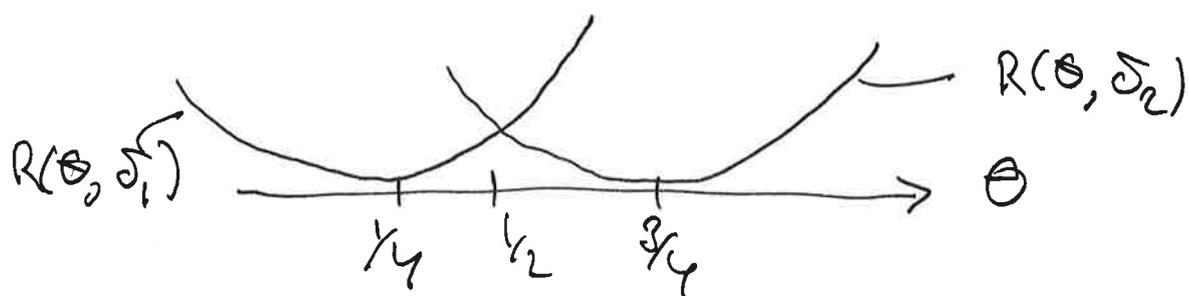
$$R(\theta, \hat{\delta}(X)) = E_{\theta}((\theta - \hat{\delta}(X))^2)$$

$$\hat{\delta}_1(x) = \frac{1}{4}$$

$$\hat{\delta}_2(x) = \frac{3}{4}$$

$$\bullet R(\theta, \hat{\delta}_1) = E((\theta - \frac{1}{4})^2) = (\theta - \frac{1}{4})^2$$

$$R(\theta, \hat{\delta}_2) = E((\theta - \frac{3}{4})^2) = (\theta - \frac{3}{4})^2$$



Several approaches to coming to terms with this problem.

- (i) allow only certain estimators.
 - a) allow only unbiased estimators, i.e. estimators δ s.t.

$$E_{\theta}(\delta(X)) = g(\theta) \quad \text{all } \theta.$$

Then we can let

$$\mathcal{U} = \{ \text{all unbiased estimators } \delta \text{ of } g(\theta) \}$$

and try to do

$$\hat{\delta} = \underset{\delta \in \mathcal{U}}{\text{argmin}} R(\theta, \delta)$$

and hope that the resulting estimator satisfies

$$R(\theta, \hat{\delta}) \leq R(\theta, \delta) \quad \forall \delta, \theta.$$

In this course we will show that this works for a large class of distributions, exponential family of distributions.

$\hat{\delta}$ is called UMVU (uniform minimum variance unbiased)

b) allow only estimators that preserve some scaling properties

ex: estimator of length, given in km. change scale of r.v. \bar{X} into meter. We want then the resulting estimator to change the scale also.

#

Equivariant estimator. Try to do

$$\hat{\delta} = \underset{\delta \in \{\text{equivariant}\}}{\text{argmin}} R(\theta, \delta)$$

goal hope

$$R(\theta, \delta^1) \leq R(\theta, \delta) \text{ for all } \delta, \theta$$

We will prove that that works.

This works for ~~family~~ families of distributions called

group families

Then δ^1 is called MRE

minimum risk equivariant

estimator.

(ii) a) change the measure of error, risk function

$$\int R(\theta, \delta) \lambda(\theta) d\theta$$

weight function density

$$\lambda \geq 0$$

$$\int \lambda(u) du = 1$$

this means that

we treat θ as ~~if~~ a r.v.

with density λ .

Note

$$R(\theta, \delta) = E_{\theta}(L(\theta, \delta(X)))$$

so that

$$\int R(\theta, \delta) \lambda(\theta) d\theta = E_{\lambda}(R(\theta, \delta)) =$$

(Recall $E(h(X)) = \int h(y) f_X(y) dy$)

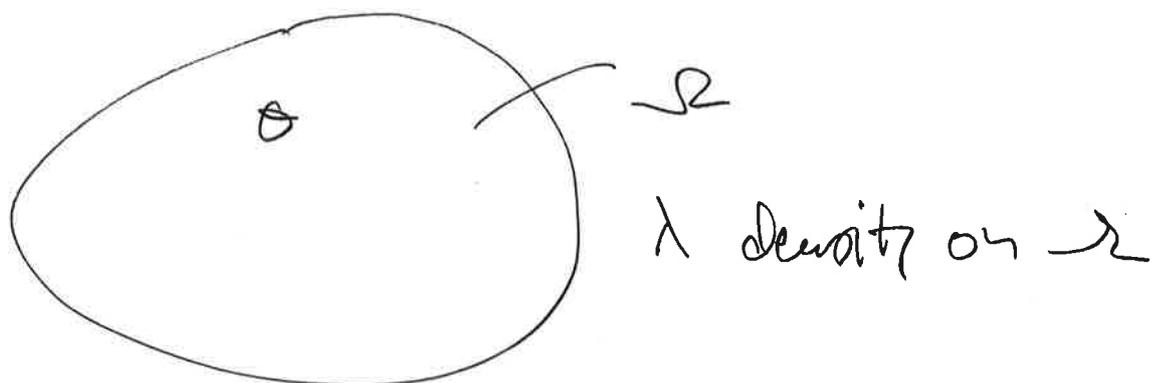
$$= E_{\lambda}(E_{\theta}(R(\theta, \delta(X))))$$

Try to find

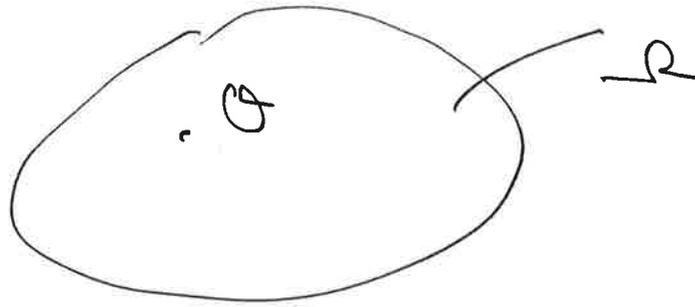
$$\delta_{\lambda} = \arg \min_{\delta} \int R(\theta, \delta) \lambda(\theta) d\theta$$

δ_{λ} is called a Bayes estimator.

5)



b)



$$\sup_{\theta \in \Omega} R(\theta, \delta)$$

is the maximum risk T_{θ} to have

$$\delta = \arg \min_{\delta} \sup_{\theta \in \Omega} R(\theta, \delta)$$

δ is called a minimax estimator.

Relation

Bayes estimator \longleftrightarrow minimax estimator.

(least favorable distributions...)