EVALUATION OF AN ESTIMATOR

MEAN SQUARE ERROR.

$\hat{\theta}$ OR W "DISTANCE MEASURE" BETWEEN W AND θ

$E_\theta(W-\theta)^2$, FOR θ FIX, WE CAN COMPUTE: MSE

WHY $(W-\theta)^2$ AND NOT $|W-\theta|$. MATHEMATICALLY, IT IS EASIER TO DEAL WITH A SQUARE RATHER THAN AN ABSOLUTE VALUE.

ALSO

$$E_\theta(W-\theta)^2 = \underbrace{\mathrm{VAR}_\theta\, W}_{\text{MEASURE OF PRECISION}} + \underbrace{(\theta - E_\theta(W))^2}_{\text{BIAS}^2}$$

PROOF:

$$E_\theta(W-\theta)^2 = E_\theta\left((W-E_\theta(W)) - (\theta - E_\theta(W))\right)^2$$

$$= E_\theta(W-E_\theta(W))^2 - 2E_\theta(W-E_\theta(W))(\theta - E_\theta(W)) + (\theta - E_\theta(W))^2.$$

(MID-TERM IS EQUAL TO ZERO).

IDEALLY, WE WOULD LIKE TO GET W^* SUCH THAT $MSE(W^*) \leq MSE(W)$ FOR ANY ESTIMATOR W. AND ANY VALUE θ.

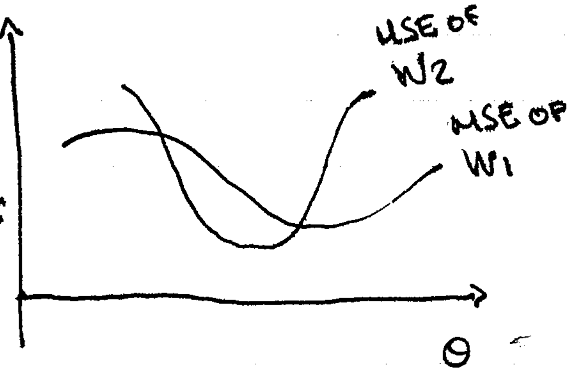
USUALLY, FOR ANY TWO ESTIMATOR W_1, W_2 THE MSEs CROSS WITH EACH OTHER

FOR CERTAIN VALUES OF θ:

$MSE_{W_1} < MSE_{W_2}$

FOR OTHER VALUES OF θ:

$MSE_{W_2} < MSE_{W_1}$



EXAMPLE: $X_1, X_2, \ldots X_n$ iid BERNOULLI (θ)

WE HAD TWO ESTIMATORS:

$\hat{\theta}_{MLE} = \bar{X}$ AND $\hat{\theta}_B = \dfrac{\alpha + \sum_{i=1}^{n} X_i}{(\alpha+\beta+n)}$ (α, β PARAMETERS OF THE PRIOR).

FIND MSE OF BOTH ESTIMATORS

$E(\hat{\theta}_{MLE}) = E(\bar{X}) = \theta$ BIAS $\hat{\theta}_{MLE} = 0$

$MSE(\hat{\theta}_{MLE}) = VAR(\bar{X}) = \dfrac{\theta(1-\theta)}{n}$ (1)

FOR BAYES ESTIMATOR:

$E(\hat{\theta}_B) = \dfrac{\alpha + \sum_{i=1}^{n} E(X_i)}{(\alpha+\beta+n)} = \dfrac{\alpha + n\theta}{(\alpha+\beta+n)} \neq \theta$ (NOT UNBIASED)

$BIAS^2(\hat{\theta}_B) = \left(\theta - \dfrac{\alpha+n\theta}{(\alpha+\beta+n)}\right)^2 = \left(\dfrac{(\alpha+\beta)\theta - \alpha}{(\alpha+\beta+n)}\right)^2$

$VAR(\hat{\theta}_B) = \dfrac{1}{(\alpha+\beta+n)^2} VAR(\alpha + \sum_{i=1}^{n} X_i) = \dfrac{n\,VAR(X)}{(\alpha+\beta+n)^2} = \dfrac{n\theta(1-\theta)}{(\alpha+\beta+n)^2}$

$\Rightarrow MSE(\hat{\theta}_B) = \dfrac{n\theta(1-\theta)}{(\alpha+\beta+n)^2} + \dfrac{((\alpha+\beta)\theta-\alpha)^2}{(\alpha+\beta+n)^2}$ (2)

DIFFICULT TO ESTABLISH COMP. BETWEEN (1) AND (2)

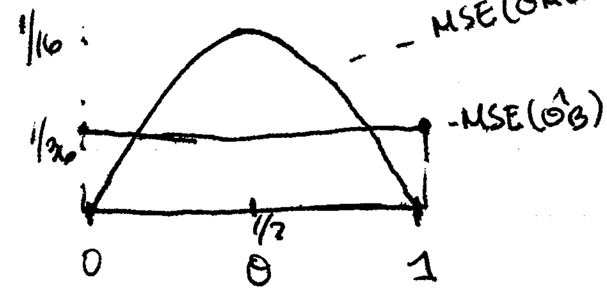
JUST TO GIVES US AN IDEA:

$n=4$, $\alpha=\beta=1$ (UNIFORM PRIOR)

$\Rightarrow MSE(\hat{\theta}_{MLE}) = \dfrac{\theta(1-\theta)}{4}$

$MSE(\hat{\theta}_B) = \dfrac{4\theta(1-\theta)}{36} + \dfrac{(2\theta-1)^2}{36} = \dfrac{4\theta - 4\theta^2 + 4\theta^2 - 4\theta + 1}{36} = \dfrac{1}{36}$

GRAPH



IT SEEMS THAT FOR THE MOST OF THE VALUES OF Θ, $\hat{\Theta}_B$ IS A BETTER CHOICE TO $\hat{\Theta}_{MLE}$, EXCEPT IF Θ CLOSE TO ZERO OR CLOSE TO ONE.

IF NOW WE TAKE $n = 100$

$$MSE(\hat{\Theta}_{MLE}) = \frac{\Theta(1-\Theta)}{100} \quad \text{AND} \quad MSE(\hat{\Theta}_B) = =$$

$$\frac{100\,\Theta(1-\Theta)}{(102)^2} + \frac{(2\Theta-1)^2}{(102)^2} = \frac{100\Theta - 100\Theta^2 + 4\Theta^2 - 4\Theta + 1}{(102)^2} = \frac{96(\Theta-\Theta^2)+1}{(102)^2}$$

IN GENERAL, COMPARISONS OF MSE DEPEND ON n AND THE PRIOR, IF WE ARE USING BAYES ESTIMATOR.

A RELATED CONCEPT:

W IS AN UNBIASED ESTIMATOR OF $\Theta \Leftrightarrow E_\Theta(W) = \Theta$

$(\text{BIAS}_\Theta(W) = 0)$

$\hat{\Theta}_{MLE}$ IS UNBIASED BUT $\hat{\Theta}_B$ IS NOT (BERNOULLI EX).

$X_1, X_2, \ldots X_n$, $X \sim U(0,\Theta)$ TWO POSSIBLE ESTIMATORS FOR Θ ARE $\hat{\Theta}_1 = 2\bar{X}$ AND $\hat{\Theta}_2 = \max\{X_1, X_2, \ldots X_n\} = X_{(n)}$

ARE THEY UNBIASED?

$E(\hat{\Theta}_1) = 2E(\bar{X}) = 2E(X) = 2\frac{\Theta}{2} = \Theta$, $\hat{\Theta}_1$ UNBIASED

FOR $\hat{\Theta}_2$

$$P(X_{(n)} \le x) = \prod_{i=1}^{n} F_X(x) = \left[\frac{x}{\Theta}\right]^n = x^n/\Theta^n \quad , 0 \le x \le \Theta$$

THE pdf OF $X_{(n)}$ IS $f_{X_{(n)}}(x) = \dfrac{n x^{n-1}}{\Theta^n}$; $0 \le x \le \Theta$

$$\Rightarrow E(\hat{\theta}_2) = \int_0^\theta x \left[\frac{n x^{n-1}}{\theta^n}\right] dx = \frac{n}{\theta^n} \, x^{n+1} \, \frac{1}{(n+1)} = \frac{n}{(n+1)} \theta$$

NOT UNBIASED. BUT IF MAKE

$$\hat{\theta}_3 = \frac{n+1}{n} \hat{\theta}_2 \quad \Rightarrow \quad E(\hat{\theta}_3) = \theta$$

LOOKING FOR AN ESTIMATOR SUCH THAT ~~HAS~~ "BEST MSE", IS PRACTICALLY IMPOSSIBLE.

IDEA: CONSIDER ONLY ESTIMATORS W THAT ARE UNBIASED

$$\Rightarrow MSE_\theta(W) = Var_\theta W$$

THEN, AMONG THE CLASS OF UNBIASED ESTIMATORS, FIND W^* SUCH THAT

$$Var_\theta W^* \leq Var_\theta W \quad \text{FOR ALL } \theta.$$

W^* IS THE UNIFORM ~~MINIMUM~~ MINIMUM VARIANCE UNBIASED ESTIMATOR OF θ. (UMVUE). HOW TO FIND A UMVUE?

SUPPOSE WE SPECIFY A LOWER BOUND $B(\theta)$ FOR THE VARIANCE OF ANY UNBIASED ESTIMATOR OF θ, i.e.

$$B(\theta) \leq Var_\theta W.$$

~~NOW~~ UNDER CERTAIN CONDITIONS, WE WILL SHOW THAT

$$B(\theta) = \frac{1}{n E_\theta\left(\frac{\partial}{\partial\theta} \log f(x|\theta)\right)^2} \qquad \text{CRÁMER-RAO LOWER BOUND}$$

SUPPOSE $X_1, X_2, \ldots X_n$ ARE iid POISSON (λ) RVs.

FIND CRÁMER-RAO LOWER BOUND FOR λ

$$f(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots \quad \lambda > 0$$

$$\log f(x|\lambda) = x \log\lambda - \lambda - \log x!$$

NOTES:

NEED TO CHECK: THAT FOR iid OBSERVATIONS (COROLLARY 7.3.10)

$$E\left(\frac{d\log f(\underline{x}|\theta)}{d\theta}\right)^2 = n\,E\left(\frac{d\log f(x|\theta)}{d\theta}\right)^2$$

NOTES SKIP.

$$\frac{d\log f(\underline{x}|\theta)}{d\theta} = \frac{d}{d\theta}\log\prod_{i=1}^{n} f(x_i|\theta) = \frac{d}{d\theta}\sum_{i=1}^{n}\log(f(x_i|\theta))$$

$$= \sum_{i=1}^{n}\frac{d}{d\theta}\log f(x_i|\theta)$$

$$\left(\sum_{i=1}^{n}\frac{d\log f(x_i|\theta)}{d\theta}\right)^2 = \sum_{i=1}^{n}\left(\frac{d\log f(x_i|\theta)}{d\theta}\right)^2 + 2\sum_{i}\sum_{\substack{j\\ i<j}}\left(\frac{d\log f(x_i|\theta)}{d\theta}\,\frac{d\log f(x_j|\theta)}{d\theta}\right)$$

IF WE TAKE EXPECTED VALUES: $\Rightarrow$

$$E\left(\frac{d\log f(\underline{x}|\theta)}{d\theta}\right)^2 = \sum_{i=1}^{n} E\left(\frac{d\log f(x_i|\theta)}{d\theta}\right)^2 + 2\sum_{i}\sum_{j} E\left(\frac{d\log f(x_i|\theta)}{d\theta}\right)E\left(\frac{d\log f(x_j|\theta)}{d\theta}\right) = n\,E\left(\frac{d\log f(x|\theta)}{d\theta}\right)^2$$

TO FACILITATE CALCULATIONS: EX 7.39

$$E\left(\frac{d\log f(x|\theta)}{d\theta}\right)^2 = -E\left(\frac{d^2\log f(x|\theta)}{d\theta^2}\right)$$

PROOF:

$$E\left(\frac{d\log f(x|\theta)}{d\theta}\right)^2 = \int\left[\frac{d f(x|\theta)}{d\theta}\,\frac{1}{f(x|\theta)}\right]^2 f(x|\theta)\,dx$$

$$= \int\left[\frac{d f(x|\theta)}{d\theta}\right]^2\frac{1}{f(x|\theta)}\,dx \;:=\; -\int\frac{d^2\log f(x|\theta)}{d\theta^2}\,f(x|\theta)\,dx$$

HOWEVER, $\frac{d}{d\theta}\left[\frac{f'(x|\theta)}{f(x|\theta)}\right] = \frac{d^2 f(x|\theta)}{d\theta^2}\,\frac{f(x|\theta)}{f(x|\theta)^2} - \left[\frac{d f(x|\theta)}{d\theta}\right]^2\frac{1}{f(x|\theta)^2}$

AND $\int\frac{d^2 f(x|\theta)}{d\theta^2}\,\frac{f(x|\theta)}{f(x|\theta)^2}\,dx = \frac{d^2}{d\theta^2}\int f(x|\theta)\,dx = 0$

$$\frac{\partial \log f(x|\lambda)}{\partial \lambda} = \frac{x}{\lambda} - 1 = \frac{x-\lambda}{\lambda}$$

$$\Rightarrow E\left(\frac{\partial \log f(x|\lambda)}{\partial \lambda}\right)^2 = E\left(\frac{x-\lambda}{\lambda}\right)^2 = \frac{1}{\lambda^2} E(x-\lambda)^2 = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

$$\Rightarrow B(\lambda) = \frac{1}{n(1/\lambda)} = \lambda/n.$$

SO IF W IS ANY ESTIMATOR OF λ SUCH THAT $E(W)=\lambda$

$$\Rightarrow \quad \lambda/n \le Var_\lambda(W)$$

ALSO, WE KNOW THAT $\hat{\theta} = \bar{X}$ IS UNBIASED BECAUSE $E(\bar{X}) = \lambda$. ALSO $Var(\bar{X}) = \frac{\lambda}{n} = B(\lambda)$.

NECESSARILY $\bar{X}$ IS UMVUE.

IN GENERAL, WE WILL CONSIDER UNBIASED ESTIMATORS FOR ANY DIFFERENTIABLE FUNCTION $\tau(\theta)$, i.e. $E(W) = \tau(\theta)$

$$\Rightarrow \quad Var_\theta(W) \geq \frac{\left[\frac{d\tau(\theta)}{d\theta}\right]^2}{n E_\theta\left(\frac{d \log f(x|\theta)}{d\theta}\right)^2} = \left[\frac{d\tau(\theta)}{d\theta}\right]^2 B(\theta)$$

FOR THE POISSON EXAMPLE, SUPPOSE WE WANT TO ESTIMATE $P[X=0] = e^{-\lambda} \Rightarrow \tau(\lambda) = e^{-\lambda} \Rightarrow \frac{d\tau(\lambda)}{d\lambda} = -e^{-\lambda}$

THE LOWER BOUND IN VARIANCE FOR THE CLASS OF UNBIASED ESTIMATORS OF λ IS: $(e^{-2\lambda})(\lambda/n)$

CAN WE FIND AN UNBIASED ESTIMATOR THAT REACHES THE BOUND? NO. THERE IS ONE AND ONLY FUNCTION OF λ FOR WHICH SUCH THING HAPPENS.

$$W = \frac{1}{n}\sum_{i=1}^{n} I(x_i)_{\{0\}}; \quad E(W) = \frac{1}{n} n e^{-\lambda} = e^{-\lambda}, \quad Var(W) = \frac{1}{n} e^{-\lambda}(1-e^{-\lambda})$$

EX: 5.40 LET $X_1, X_2 \ldots X_n$ BE iid BERNOULLI (p)
SHOW THAT THE VARIANCE OF $\bar{X}$ ATTAINS THE CRAMÉR-RAO LOWER BOUND

$$\mathrm{VAR}(\bar{X}) = \mathrm{VAR}\left(\sum_{i=1}^{n} X_i\right) = \frac{n}{n^2}\mathrm{VAR}(X) = \frac{p(1-p)}{n}.$$

$$f(x|p) = p^x(1-p)^{1-x}, \quad 0<p<1, \quad x=0,1.$$

$$\log f(x|p) = x\log p + (1-x)\log(1-p)$$

$$\frac{d\log f(x|p)}{dp} = \frac{x}{p} - \frac{(1-x)}{(1-p)}$$

$$E_{x|p}\left(\frac{d\log f(x|p)}{dp}\right)^2 = E_{x|p}\left(\frac{x}{p} - \frac{(1-x)}{(1-p)}\right)^2$$

$$= E_{x|p}\left(\frac{x-p}{p(1-p)}\right)^2 = \frac{1}{p^2(1-p)^2}\mathrm{VAR}(x) = \frac{1}{p(1-p)}$$

$$B(p) = 1/n\left(\frac{1}{p(1-p)}\right) = \frac{p(1-p)}{n}$$

$$\frac{d^2\log f(x|p)}{dp^2} = -\frac{x}{p^2} - \frac{(1-x)}{(1-p)^2}; \quad E_{x|p}\left(-\frac{d^2\log f(x|p)}{dp^2}\right)$$

$$= E_{x|p}\left(\frac{x}{p^2}\right) + \frac{E_{x|p}(1-x)}{(1-p)^2} = \frac{1}{p} + \frac{1}{(1-p)} = \frac{1}{p(1-p)}$$

COROLLARY 7.3.15 IF $W(\underline{X})$ IS AN UNBIASED ESTIMATOR OF $\tau(\theta)$, THEN $W(\underline{X})$ ATTAINS THE CRAMÉR-RAO LOWER BOUND IF AND ONLY IF

$$a(\theta)[W(\underline{x}) - \tau(\theta)] = \frac{d}{d\theta}\log L(\theta|\underline{x})$$

UNIQUENESS OF A FUNCTION $\tau(\theta)$ FOR WHICH THE BOUND IS REACHED

HOW TO USE THIS ?

$X_1, X_2 \ldots X_n$, $X \sim Exp(\theta = 1/\lambda)$ iid OBS.

$$L(\lambda|\underline{x}) = \prod_{i=1}^{n} f(x_i|\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$$

$$\log L(\lambda|\underline{x}) = n \log \lambda - \lambda \sum_{i=1}^{n} x_i$$

$$\frac{d \log L(\lambda|\underline{x})}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = \underbrace{-n}_{a(\lambda)} \Big[\underbrace{\bar{X}}_{W(\underline{x})} - \underbrace{\frac{1}{\lambda}}_{\tau(\lambda)} \Big]$$

SO THE ONLY FUNCTION OF λ FOR WHICH THE BOUND IS ATTAINED IS $\tau(\lambda) = 1/\lambda$

ANOTHER EXAMPLE, LET $X_1, X_2 \ldots X_n$ BE iid OBSERVATIONS WHERE $X \sim$ POISSON(λ). WE SHOW LAST TIME THAT $\bar{X}$ IS UMVUE. ALTERNATIVE PROOF: LET.

$$L(\lambda|\underline{x}) = \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}}{\prod_{i=1}^{n} x_i!} \; ; \; \log(L(\lambda|\underline{x})) = \sum_{i=1}^{n} x_i \log \lambda - n\lambda - \log \prod_{i=1}^{n} x_i!$$

$$\frac{d \log L(\lambda|\underline{x})}{d\lambda} = \frac{\sum x_i}{\lambda} - n + 0 = \frac{\sum_{i=1}^{n} x_i - n\lambda}{\lambda} =$$

$$= \underbrace{\frac{n}{\lambda}}_{a(\lambda)} \big(\underbrace{\bar{X}}_{W(\underline{x})} - \underbrace{\lambda}_{\tau(\lambda)} \big)$$

FINAL COMMENTS

- THERE IS ONLY ONE FUNCTION $\tau(\lambda)$ FOR WHICH THERE EXISTS A $W(\underline{x})$ SUCH THAT $VAR(W(\underline{x}))$ REACHES THE LOWER BOUND. (REMEMBER $e^{-\lambda}$ IN POISSON EX).

PROOF OF CRAMÉR-RAO'S INEQUALITY:

TAKE ANY ESTIMATOR $W = W(\underline{X})$ SUCH THAT $E_\theta(W) = \tau(\theta)$

CONSIDER

$$\frac{d\,\tau(\theta)}{d\theta} = \frac{d\,E_\theta(W(\underline{x}))}{d\theta} = \frac{d}{d\theta}\int_{\mathcal{X}} W(\underline{x})\, f(\underline{x}|\theta)\,d\underline{x}$$

$$= \int_{\mathcal{X}} W(\underline{x})\,\frac{d f(\underline{x}|\theta)}{d\theta}\,d\underline{x} = \int_{\mathcal{X}} W(\underline{x})\,\underbrace{\frac{d f(\underline{x}|\theta)}{d\theta}\,\frac{1}{f(\underline{x}|\theta)}}_{\frac{d\log f(\underline{x}|\theta)}{d\theta}}\,f(\underline{x}|\theta)\,d\underline{x}$$

$$= \int_{\mathcal{X}} W(\underline{x})\,\frac{d\log f(\underline{x}|\theta)}{d\theta}\,f(\underline{x}|\theta)\,d\underline{x} = \underbrace{E\left(W(\underline{x})\,\frac{d\log f(\underline{x}|\theta)}{d\theta}\right)}_{\text{LOOKS LIKE A COVARIANCE}}$$

ONE THING . .

$$E\left(\frac{d\log f(\underline{x}|\theta)}{d\theta}\right) = \int_{\mathcal{X}} \frac{d f(\underline{x}|\theta)}{d\theta}\,\frac{1}{f(\underline{x}|\theta)}\,f(\underline{x}|\theta)\,d\theta$$

$$= \frac{d}{d\theta}\int_{\mathcal{X}} f(\underline{x}|\theta)\,d\underline{x} = \frac{d\,(1)}{d\theta} = 0$$

$\mathrm{COV}(X,Y) = E(XY)$ AS LONG AS $E(X) = 0$ OR $E(Y) = 0$

$$\Rightarrow E\left(W(\underline{x})\,\frac{d\log f(\underline{x}|\theta)}{d\theta}\right) = \mathrm{cov}\left(W, \frac{d\log f(\underline{x}|\theta)}{d\theta}\right) = \frac{d\,\tau(\theta)}{d\theta}$$

CAUCHY-SCHWARTZ INEQUAL: $(\mathrm{COV}(X,Y))^2 \le \mathrm{VAR}(X)\,\mathrm{VAR}(Y)$

$$\Rightarrow \left(\frac{d\,\tau(\theta)}{d\theta}\right)^2 \le \mathrm{VAR}(W(\underline{x}))\,\mathrm{VAR}\left(\underbrace{\frac{d\log f(\underline{x}|\theta)}{d\theta}}_{\text{HAS A ZERO EXPECTED VALUE}}\right)$$

$$\Rightarrow \frac{(d\,\tau(\theta)/d\theta)^2}{\underbrace{E\left(\frac{d\log f(\underline{x}|\theta)}{d\theta}\right)^2}_{\text{FISHER'S INFORMATION NUMBER}}} \le \mathrm{VAR}(W(\underline{x}))$$

NOTICE THAT IF THE NUMBER INCREASES, THEN THE LOWER BOUND DECREASES

$\Rightarrow$ MORE PRECISION.

- THE THEORY APPLIES ONLY WHEN $\frac{\partial}{\partial\theta}\int = \int\frac{\partial}{\partial\theta}$

VALID FOR EXPONENTIAL FAMILY MODELS

HOWEVER IF THE SUPPORT OF THE pdf DEPENDS ON θ in $\left(f(x|\theta) = \frac{1}{\theta} I_{(0,\theta)}(x)\right)$ THIS LAST CONDITION IS NOT MET.

EX: LET $X_1, X_2 . X_n$ BE iid OBSERVATIONS SUCH THAT

$$f(x|\mu) = e^{-(x-\mu)} I_{(\mu,\infty)}(x)$$

IF WE ATTEMPT TO FIND CRAMÉR-RAO LOWER BOUND FOR μ

$$\log f(x|\mu) = -(x-\mu)$$

$$\frac{d\log f(x|\mu)}{d\mu} = 1 \qquad E_{x|\mu}\left(\frac{d\log f(x|\mu)}{d\mu}\right)^2 = \frac{1}{\text{CONSTANT}}$$

$$f(x_1, x_2 . x_n|\mu) = e^{-\sum_{i=1}^{n}(x_i-\mu)} I_{(x_{(1)},\infty)}(\mu)$$

$T = X_{(1)}$ IS SUFFICIENT, UNBIASED? WHAT IS $E(X_{(1)})$?

$$P(X_{(1)} \geq x) = \prod_{i=1}^{n} P(X_i > x) = [P(X>x)]^n = [1-P[X\leq x]]^n$$

$$P[X\leq x] = \int_{\mu}^{x} e^{-(u-\mu)}du = -e^{-(u-\mu)}\Big|_{\mu}^{x} = 1-e^{-(x-\mu)}$$

$$\Rightarrow P(X_{(1)} \geq x) = e^{-n(x-\mu)} \Rightarrow P(X_{(1)} \leq x) = 1-e^{-n(x-\mu)}$$

$$f_{X_{(1)}}(x) = n e^{-n(x-\mu)}$$

$z = x-\mu$

$$E(X_{(1)}) = n\int_{\mu}^{\infty} x e^{-n(x-\mu)}dx = n\int_{0}^{\infty}(z+\mu)e^{-nz}dz$$

$$= n\int_{0}^{\infty} z e^{-nz}dz + n\mu\int_{0}^{\infty} e^{-nz}dz = \frac{1}{n}+\mu$$

THEN $T^* = X_{(1)} - \frac{1}{n}$ IS UNBIASED FOR μ.. IS THIS UMVUE?

$$\mathrm{VAR}\left(X_{(1)}-\frac{1}{n}\right)=\mathrm{VAR}(X_{(1)})$$

$$E(X_{(1)}^2)=n\int_{\mu}^{\infty}x^2e^{-(n)(x-\mu)}dx=n\int_{0}^{\infty}(z+\mu)^2e^{-nz}dz$$

$$=n\int_0^{\infty}z^2e^{-nz}dz+2n\mu\int_0^{\infty}ze^{-nz}dz+\mu^2n\int_0^{\infty}e^{-nz}dz$$

$$=2/n^2+2\mu(1/n)+\mu^2$$

$$\mathrm{VAR}(X_{(1)})=2/n^2+\cancel{2\mu/n}+\cancel{\mu^2}-\frac{1}{n^2}-\cancel{\frac{2\mu}{n}}-\cancel{\mu^2}=\frac{1}{n^2}<\underline{1}$$

THE BOUND DOES NOT APPLY ! (SEE ALSO EXAMPLE 7.3.13)

WHAT TO DO ?

IDEA: CONSIDER W ANY UNBIASED ESTIMATOR $\tau(\theta)$..

AND LET T BE A SUFFICIENT STATISTIC FOR θ.

LETS DEFINE A NEW ESTIMATOR $\phi(T)=E(W|T)$

IN THE CONTINUOUS CASE, $E(W|T)=\int w\,f(w|t)\,dw$

SINCE T IS SUFFICIENT $f(w|t)$ DOES NOT DEPEND ON θ (DEF. OF SUFFICIENT STATISTIC) THEN $\phi(T)$ DOES NOT DEPEND ON θ AND CAN BE CONSIDERED AN ESTIMATOR.

ON THE OTHER PART,

$$E(\phi(T))=E\,E(W|T)=E(W)=\tau(\theta)$$

(REMEMBER THAT $E(E(X|Y))=E(X)$). ALSO,

$$\mathrm{VAR}(W)=\mathrm{VAR}(E(W|T))+E(\mathrm{VAR}(W|T))$$

$$=\mathrm{VAR}(\phi(T))+E(\mathrm{VAR}(W|T))$$

$$\geq\mathrm{VAR}(\phi(T))$$

IN TERMS OF VARIANCE, $\phi(T)$ IS UNIFORMLY BETTER THAN W (RAO-BLACKWELL) (WE CAN ALWAYS IMPROVE AN UNBIASED ESTIMATOR BY CONDITIONING ON A SUFFICIENT STATISTIC).

NOW, LETS THINK THAT T IS BOTH SUFFICIENT AND COMPLETE TWO ESTIMATORS $\phi(T)$ AND $\phi^*(T)$ SUCH THAT

$$E(\phi(T)) = \tau(\theta) \quad ; \quad E(\phi^*(T)) = \tau(\theta)$$

THIS IMPLIES,

$$E(\phi(T) - \phi^*(T)) = E(\phi(T)) - E(\phi^*(T)) = \tau(\theta) - \tau(\theta) = 0$$

SINCE T IS COMPLETE $\Rightarrow \phi(T) - \phi^*(T) = 0$

$\Rightarrow \phi(T) = \phi^*(T)$

THEN, IF T IS SUFFICIENT AND COMPLETE, THERE IS ONE AND ONLY ONE FUNCTION OF T THAT ~~ESTIMATES~~ RESULTS IN AN UNBIASED ESTIMATOR OF $\tau(\theta)$

BY RAO-BLACKWELL:

$$\phi(T) = E(W \mid T) \; ; \quad \text{WHERE } E(W) = \tau(\theta)$$

AND $\quad VAR(\phi(T)) \leq VAR(W)$

SINCE $\phi(T)$ IS UNIQUE $\Rightarrow \phi(T)$ IS A UNIFORMLY MINIMUM VARIANCE UNBIASED ESTIMATOR.

LEHMANN-SCHEFFÉ THEOREM: LET $\phi(T)$ BE AN UNBIASED ESTIMATOR OF $\tau(\theta)$ THAT IS A FUNCTION OF A SUFFICIENT AND COMPLETE STATISTIC. THEN, $\phi(T)$ IS UMVUE.

$$E(g(x_{(1)})) = n\int_{\mu}^{\infty} g(x)e^{-n(x-\mu)}dx = 0 \quad -n\int_{-\infty}^{-\mu} g(-z)e^{-n(-z-\mu)}dz = 0$$

$z = -x$ TAKING DERIVATIVE

$-ng(\mu)e^{-n(\mu-\mu)} = 0$

$g(\mu) = 0$

63

IN THE EXAMPLE, $f(x|\mu) = e^{-(x-\mu)} I_{(\mu,\infty)}(x)$

$X_{(1)} = \min\{X_1, X_2, \ldots X_n\}$ IS SUFFICIENT AND COMPLETE

$T^* = X_{(1)} - 1/n$ IS UNBIASED FOR μ. T^* IS UMVUE.

EX: LET $X_1, X_2, \ldots, X_n$ IID OBSERVATIONS $X \sim$ POISSON (λ)

$f(x|\lambda) = \dfrac{\lambda^x e^{-\lambda}}{x!}, \; x = 0, 1, 2, \ldots$

WE KNOW THAT $T = \sum_{i=1}^{n} X_i$ IS SUFFICIENT AND COMPLETE

$\Rightarrow \bar{X} = \sum_{i=1}^{n} X_i/n = T/n$ AND $E(\bar{X}) = \lambda \Rightarrow \bar{X}$ IS UMVUE.

BACK TO OLD PROBLEM $\tau(\lambda) = e^{-\lambda} = P[X=0]$

FIRST, WE NEED ANY UNBIASED ESTIMATOR FOR $e^{-\lambda}$

$W_1 = I_{\{0\}}(X_1) \qquad E(W_1) = 1 \cdot P(X_1 = 0) + 0 \cdot P(X_1 > 0) = e^{-\lambda}$

BY OUR THEORY, $\phi(t) = E\left(I_{\{0\}}(X_1) \mid \sum_{i=1}^{n} X_i = t\right)$

$= 1 \cdot P(X_1 = 0 \mid \sum_{i=1}^{n} X_i = t) + 0 \cdot P(X_1 > 0 \mid \sum_{i=1}^{n} X_i = t)$

$$P(X_1 = 0 \mid \sum_{i=1}^{n} X_i = t) = \frac{P(X_1 = 0 \text{ AND } \sum_{i=1}^{n} X_i = t)}{P(\sum_{i=1}^{n} X_i = t)}$$

$$= P(X_1 = 0)\, P\left(\sum_{i=2}^{n} X_i = t\right) \Big/ P\left(\sum_{i=1}^{n} X_i = t\right) = \frac{e^{-\lambda}\dfrac{((n-1)\lambda)^t e^{-(n-1)\lambda}}{\cancel{t!}}}{\dfrac{(n\lambda)^t \cancel{e^{-n\lambda}}}{\cancel{t!}}}$$

$= \left(\dfrac{n-1}{n}\right)^{\sum_{i=1}^{n} X_i}$ THIS IS UMVUE!

LOSS-FUNCTION OPTIMALITY. SECTION 7.3.4.

GENERAL APPROACH TO ESTIMATION (INFERENCE)

IDEA IS TO FORMULATE AN ESTIMATION PROBLEM AS A "DECISION THEORY" PROBLEM.

AS USUAL, WE HAVE $X_1, X_2 \ldots X_n$ iid OBS. SUCH THAT $X \sim f(x|\theta)$; $\theta \in \Theta \equiv$ PARAMETER SPACE

$\mathcal{A}$ IS THE SET OF ALLOWABLE ACTIONS OR DECISIONS (USUALLY Θ IN ESTIMATION)

WE NEED A MEASURE OF DISCREPANCY BETWEEN OUR DECISION (POINT ESTIMATE) AND θ.

LOSS FUNCTION: $L(\theta, a)$ $a \equiv$ "ACTION" (LOSS $\approx$ ERROR)

PROPERTIES:

- $L(\theta, a) \geq 0$ FOR a AND θ
- $L(\theta, a) = 0$ WHEN $a = \theta$

EXAMPLE: SEVERAL POSSIBLE LOSS FUNCTIONS ARE:

i) $L(\theta, a) = (a-\theta)^2$ SQUARE-ERROR LOSS FUNCTION

ii) $L(\theta, a) = |a-\theta|$ ABSOLUTE-ERROR LOSS FUNCTION

iii) $L(\theta, a) = \begin{cases} A & \text{IF } |a-\theta| > \epsilon \\ 0 & \text{IF } |a-\theta| \leq \epsilon \end{cases}$ WHERE $A > 0$.

iv) $L(\theta, a) = c(\theta)|a-\theta|^r$ FOR $c(\theta) \geq 0$ AND $r > 0$

GENERAL LOSS FUNCTION THAT CONSIDERS i) AND ii) AS PARTICULAR CASES: THE SELECTION OF THE LOSS FUNCTION DEPENDS ON THE EXPERIMENTER AND HOW HE OR SHE QUANTIFIES THE ERRORS AT ESTIMATION. THIS IS USUALLY DIFFICULT.

THE INTEGRAL IN BRACKETS IS THE "POSTERIOR EXPECTED LOSS"
THE BAYES RULE SHOULD MINIMIZE THE POSTERIOR EXPECTED LOSS.
TO FIND THE BAYES RULE MAY NOT BE EASY FOR AN ARBITRARY LOSS FUNCTION.

EXAMPLE:

SQUARE-ERROR LOSS: $L(\theta,a) = (\theta-a)^2$

POSTERIOR EXPECTED LOSS.

$$\int_{\Theta} (\theta-a)^2 \pi(\theta|\underline{X})\, d\theta = E\left((\theta-a)^2 \,|\, \underline{X}=\underline{x}\right)$$

WE CAN RE-WRITE AS: $\int_{\Theta} \theta^2 \pi(\theta|\underline{X}) d\theta - 2a \int_{\Theta} \theta \pi(\theta|\underline{X}) + a^2$

$$= E(\theta^2|\underline{X}) - 2a\, E(\theta|\underline{X}) + a^2 \equiv BR(a)$$

MINIMIZE AS A FUNCTION OF a

FIRST DERIVATIVE WITH RESPECT TO a: $-2\,E(\theta|\underline{X}) + 2a = 0 \Rightarrow \hat{a} = E(\theta|\underline{X})$

2ND DERIVATIVE: $2 > 0 \Rightarrow \hat{a} = \delta^{\pi}$ THE BAYES RULE.

FOR OTHER LOSS FUNCTIONS, THE BAYES RULE MAY BE DIFFERENT THAN $E(\theta|\underline{X})$

EXAMPLE: SUPPOSE NOW THAT $L(\theta,a) = \frac{(\theta-a)^2}{\theta^2}$

$$\Rightarrow L(\theta,a) = (\theta^2 - 2a\theta + a^2)/\theta^2 = 1 - 2a/\theta + (a/\theta)^2$$

$$\Rightarrow E(L(\theta,a)) = 1 - 2a\, E(1/\theta \,|\, \underline{X}=\underline{x}) + a^2\, E(1/\theta^2 \,|\, \underline{X}=\underline{x})$$

$$\Rightarrow \delta^{\pi} = \frac{E(1/\theta \,|\, \underline{X}=\underline{x})}{E(1/\theta^2 \,|\, \underline{X}=\underline{x})}$$ IF BOTH EXPECTATIONS EXISTS.

SUPPOSE THAT AN APPROPRIATE LOSS-FUNCTION HAS BEEN DEFINED. ALSO, OUR ACTIONS a ARE FUNCTIONS OF THE SAMPLE $\underline{X} = (X_1, X_2, \ldots X_n)$ "ESTIMATORS" $\delta(\underline{X})$

AVERAGE LOSS OF AN ESTIMATOR "RISK FUNCTION"

$R(\theta, \delta) = E_{\underline{X}|\theta}(L(\theta, \delta(\underline{X})))$ * IN THE EXAMPLE.

THEN TWO ESTIMATORS δ_1, δ_2 ARE TO BE COMPARED BY THEIR RISKS: $R(\theta, \delta_1)$ AND $R(\theta, \delta_2)$.

δ_1 IS PREFERRED OVER δ_2 IF ONLY IF

$$R(\theta, \delta_1) < R(\theta, \delta_2) \quad \text{FOR ALL VALUES OF } \theta$$

EXAMPLE: CONSIDER THE SAME LOSS FUNCTIONS AS BEFORE, THE CORRESPONDING RISKS ARE:

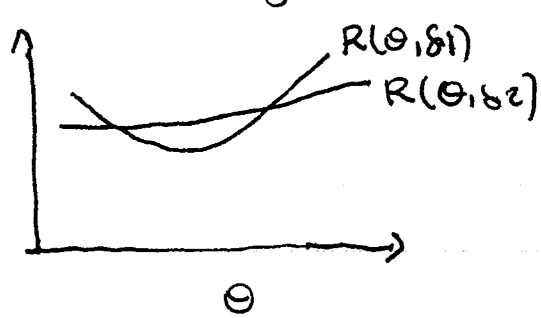
i) $E_{\underline{X}|\theta}(\delta(\underline{x}) - \theta)^2$ OUR FAMILIAR MEAN SQUARE ERROR

ii) $E_{\underline{X}|\theta} |\delta(\underline{X}) - \theta|$ THE MEAN ABSOLUTE ERROR

iii) $A \cdot P_r\{|\delta(\underline{X}) - \theta| > \epsilon\}$

iv) $C(\theta)\ E_{\underline{X}|\theta} |\delta(\underline{X}) - \theta|^r$

AS WITH THE MSE, USUALLY FOR TWO ESTIMATORS δ_1 AND δ_2



IN GENERAL, THERE WILL NOT EXIST AN ESTIMATOR WITH UNIFORMLY SMALLEST RISK.

THE PROBLEM IS THE DEPENDENCE OF THE RISK FUNCTION ON θ.

UNDER A BAYESIAN APPROACH, WE HAVE A PRIOR $\pi(\theta)$ FOR θ. A WAY OF REMOVING THE DEPENDENCY OF θ, IN RELATION TO $R(\theta, \delta)$, IS TO AVERAGE THE RISK FUNCTION FUNCTION WITH RESPECT TO $\pi(\theta)$.

BAYES RISK:

$$\int_{\Theta} R(\theta,\delta)\,\pi(\theta)\,d\theta$$

WE CAN ATTEMPT TO FIND AN ESTIMATOR THAT YIELDS THE SMALLEST VALUE OF THE BAYES RISK. SUCH AN ESTIMATOR IS CALLED THE BAYES RULE WITH RESPECT TO π AND DENOTED BY δ^{π}.

THE BAYES RISK CAN BE WRITTEN AS:

$$\int_{\Theta} R(\theta,\delta)\,\pi(\theta)\,d\theta = \int_{\Theta}\left[\int_{\mathcal{X}} R(\theta,\delta(\underline{x}))\,f(\underline{x}\mid\theta)\,d\underline{x}\right]\pi(\theta)\,d\theta$$

BY BAYES THEOREM: $f(\underline{x}\mid\theta)\,\pi(\theta) = \pi(\theta\mid\underline{x})\,m(\underline{x})$

($\pi(\theta\mid\underline{x})$ THE POSTERIOR AND $m(\underline{x})$ THE MARGINAL DENSITY OF $\underline{x}$). WE CAN WRITE BAYES RISK AS.

$$\int_{\mathcal{X}}\left[\int_{\Theta} L(\theta,\delta(\underline{x}))\,\pi(\theta\mid\underline{x})\,d\theta\right] m(\underline{x})\,d\underline{x}$$

IF FOR EVERY $\underline{x}$, WE FIND $\delta^{*}(\underline{x})$ SUCH THAT

$$\int_{\Theta} L(\theta,\delta^{*}(\underline{x}))\,\pi(\theta\mid\underline{x})\,d\theta \le \int_{\Theta} L(\theta,\delta(\underline{x}))\,\pi(\theta\mid\underline{x})\,d\underline{x}$$

$\Rightarrow \quad \delta^{*} = \delta^{\pi}$ THE BAYES RULE