Recap: Non-Bayesian Parameter Estimation

Given Y, estimate x (non-random)

$p_Y(y;x)$ or $f_Y(y;x)$

space of Y

Def: An estimator $\hat{x}: \mathcal{Y} \to \mathbb{R}$ is valid if it doesn't depend on x.

Def: An estimator $\hat{x}: \mathcal{Y} \to \mathbb{R}$. is unbiased if

$$\forall x \qquad \mathbb{E}[\hat{x}(Y)] = \int_{\mathcal{Y}} f_Y(y;x)\, y\, dy = x.$$

Def: Admissible estimator $\equiv$ Valid & Unbiased.

Def: Fisher information in $\underline{Y}$ about x

$$J_{\underline{Y}}(x) = \mathbb{E}\left[\left(\frac{\partial}{\partial x} \log f_{\underline{Y}}(\underline{Y};x)\right)^2\right]$$

Prop: $J_{\underline{Y}}(x) = \mathbb{E}\left[-\frac{\partial^2}{\partial x^2} \log f_Y(Y;x)\right]$.

Thm: If $f_Y(\cdot\,;x)$ satisfies (CRLB)

$$\mathbb{E}\left[\frac{\partial}{\partial x} \log f_{\underline{Y}}(\underline{Y};x)\right] = 0 \quad \forall x,$$

then for any admissible estimator $\hat{x}(\cdot)$,

$$\lambda_{\hat{x}}(x) = \mathrm{Var}_x(\hat{x}(\underline{Y})) \geq \frac{1}{J_{\underline{Y}}(x)}.$$

Lem: [Cauchy-Schwarz inequality]

Let Z, W be 2 arb. zero-mean rv

$$\mathbb{E}[ZW] \leq \sqrt{\mathbb{E}[Z^2]\,\mathbb{E}[W^2]}.$$

Pf: Recall that since $\hat{x}(\cdot)$ is unbiased

$$\mathbb{E}_x[e(Y)] = \mathbb{E}_x[\hat{x}(Y) - x] = 0.$$

$\log = \log_e$

$$\mathrm{Var}_x(e(Y)) = \mathbb{E}_x[e^2(Y)] = \lambda_{\hat{x}}(x)$$

NB: $\mathbb{E}_x$ & Var_x are taken wrt $f_Y(\cdot\,; \underline{x})$

Def: Score function $h(y) = \frac{\partial}{\partial x} \log f_Y(y;x)$

$$h(y) = \frac{1}{f_Y(y;x)} \frac{\partial}{\partial x} f_Y(y;x).$$

$$\mathbb{E}_x[h(Y)] = \mathbb{E}_x\left[\frac{1}{f_Y(Y;x)} \frac{\partial}{\partial x} f_Y(Y;x)\right]$$

$$= \int_y \cancel{f_Y(y;x)} \left[\frac{1}{\cancel{f_Y(y;x)}} \frac{\partial}{\partial x} f_Y(y;x)\right] dy$$

$$= \frac{\partial}{\partial x}\int_{-\infty}^{\infty} f_Y(y;x)\,dy = 0.$$

$$\mathrm{Var}(h(Y)) = \mathbb{E}[h(Y)^2] = \mathbb{E}\left[\left(\frac{\partial}{\partial x}\log f_Y(Y;x)\right)^2\right]$$

$$=: J_{\underline{Y}}(x)$$

We have $e(Y), h(Y)$ both zero mean.

$$\mathbb{E}[e(Y)h(Y)] \le \sqrt{\mathbb{E}[e(Y)^2]\,\mathbb{E}[h(Y)^2]}$$

$$\overset{\shortparallel}{1} \qquad \le \sqrt{\lambda_{\hat{x}}(x)\,J_{\underline{Y}}(x)}$$

$$\mathbb{E}[e(Y)\,h(Y)]$$

$$= \int \cancel{f_Y(y;x)}\,(\hat{x}(y)-x)\,\underbrace{\frac{\partial}{\partial x}\log f_Y(y;x)}_{= \frac{1}{\cancel{f_Y(y;x)}}\frac{\partial}{\partial x}f_Y(y;x)}\,dy$$

$$= \int (\hat{x}(y)-x)\frac{\partial}{\partial x}f_Y(y;x)\,dy$$

$$= \frac{\partial}{\partial x}\underbrace{\int \hat{x}(y)f_Y(y;x)\,dy}_{\substack{\mathbb{E}[\hat{x}(Y)]=x \\ \text{unbiasedness}}} - x\,\underbrace{\frac{\partial}{\partial x}\int \overbrace{f_Y(y;x)\,dy}^{=1}}_{0}$$

$$= 1-0 = 1.$$

$$1 = \mathbb{E}[e(Y)h(Y)] \leq \sqrt{\lambda_{\hat{x}}(x) J_{\underline{Y}}(x)}$$

$$Var(\hat{x}(Y)) = \lambda_{\hat{x}}(x) \geq \frac{1}{J_{\underline{Y}}(x)} \qquad ////.$$

<u>Ex:</u> $Y = x + W \qquad W \sim N(0, \sigma^2)$

$$\log f_Y(y; x) = -\frac{(y-x)^2}{2\sigma^2} + \text{const.}$$

$$f_Y(y; x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-x)^2}{2\sigma^2}\right)$$

$$\frac{\partial}{\partial x} \log f_Y(y; x) = -\frac{1}{\sigma^2}(x - Y) = \frac{1}{\sigma^2} W.$$

$$J_Y(x) = \mathbb{E}\left[\left(\frac{\partial}{\partial x} \log f_Y(Y; x)\right)^2\right]$$

$$= \frac{1}{\sigma^4} \underset{\substack{\parallel \\ Var(W) = \sigma^2.}}{\mathbb{E}(W^2)} = \frac{1}{\sigma^2}.$$

CRLB states that any estimator $\hat{x} : \mathbb{R} \to \mathbb{R}$ satisfies

$$\underline{\lambda_{\hat{x}}(x) \geq \sigma^2.}$$

$$\frac{\partial}{\partial x} \log f_Y(y; x) = -\frac{1}{\sigma^2}(x - Y)$$

$$\frac{\partial^2}{\partial x^2} \log f_Y(y;x) = -\frac{1}{\sigma^2}.$$

$$J_Y(x) = -\mathbb{E}\left[\frac{\partial^2}{\partial x^2} \log f_Y(Y;x)\right] = \frac{1}{\sigma^2}.$$

$Y = x + W$. We observe Y.

$\hat{x}(Y) = Y$ $\qquad$ $\mathbb{E}[\hat{x}(Y)] = \mathbb{E}\, Y = x \Rightarrow$ Unbiased.

$$\mathrm{Var}(\hat{x}(Y)) = \mathrm{Var}(Y) = \mathrm{Var}(x+W) = \mathrm{Var}(W) = \sigma^2$$

The CRLB is attained with equality.
Hence $\hat{x}(\cdot)$ is efficient.

Prop: $J_{\underline{Y}}(x) = \mathbb{E}\left[\left(\frac{\partial}{\partial x} \log f_{\underline{Y}}(\underline{Y};x)\right)^2\right]$

$$= \mathbb{E}\left[-\frac{\partial^2}{\partial x^2} \log f_Y(Y;x)\right].$$

Pf: Observe that

$$\int f_Y(y;x)\, dy = 1$$

Differentiating this wrt x

$$\int \frac{\partial}{\partial x} f_Y(y;x)\, dy = 0 \Rightarrow \int f_Y(y;x) \frac{\partial}{\partial x} \log f_Y(y;x)\, dy = 0. \quad (*)$$

$$f_Y(y;x)\,\frac{\partial}{\partial x}\log f_Y(y;x)$$

$$\|$$

Diff. (*) wrt x

$$\int f_Y(y;x)\frac{\partial^2}{\partial x^2}\log f_Y(y;x) + \left(\frac{\partial}{\partial x}\log f_Y(y;x)\right)\boxed{\frac{\partial}{\partial x} f_Y(y;x)}\, dy = 0.$$

$$\mathbb{E}\left[\frac{\partial^2}{\partial x^2}\log f_Y(Y;x)\right] + \underbrace{\mathbb{E}\left[\left(\frac{\partial}{\partial x}\log f_Y(Y;x)\right)^2\right]}_{J_Y(x)} = 0$$

Maximum Likelihood Estimation

Def: The <u>MLE</u> $\hat{x}_{ML}(y)$ is defined as

$$\hat{x}_{ML}(y) = \underset{x}{\operatorname{argmax}}\ \underbrace{f_Y(y;x)}_{\text{likelihood}}$$

Def: An unbiased estimator is <u>efficient</u> if it satisfies the CRLB with equality.

Objective: If an <u>eff. est.</u> exists, it must be the MLE.

Consider the CS Inequality. $\mathbb{E}[ZW] \le \sqrt{\mathbb{E}(Z^2)\,\mathbb{E}(W^2)}$

It is an equality iff $Z = cW$ for const $c \in \mathbb{R}$.

$$\mathbb{E}_x[e(Y)h(Y)] \le \sqrt{\mathbb{E}_x[e(Y)^2]\,\mathbb{E}_x[h(Y)^2]}$$

Equality in CRLB holds iff $\exists$ const. $\underline{k(x)}$

$$e(y) = k(x)\,h(y) \quad \text{for all } y.$$

$$\hat{x}(y) = x + k(x)\frac{\partial}{\partial x}\log f_Y(y;x)$$

But we need the RHS to be independent of x.

$$\mathbb{E}[e^2(Y)] = \mathbb{E}[e(Y)k(x)h(Y)] = k(x)\,\overbrace{\mathbb{E}[e(Y)h(Y)]}^{=1}$$

$$\| \qquad = k(x)$$

$$\lambda_{\hat{x}}(x) = J_Y(x)^{-1}$$

$$\boxed{\hat{x}(y) = x + \frac{1}{J_Y(x)}\underbrace{\left(\frac{\partial}{\partial x}\log f_Y(y;x)\right)}_{h(y)}}$$

An efficient estimator takes this form.
It is auto. unbiased.

$$\mathbb{E}[\hat{x}(Y)] = \mathbb{E}\left[x + \frac{1}{J_Y(x)}h(Y)\right]$$

$$= x + \frac{1}{J_Y(x)}\mathbb{E}[h(Y)] = x.$$

Thm: When an eff. est. exists, it is the MLE, i.e.,

$$\hat{x}_{eff}(\cdot) = \hat{x}_{ML}(\cdot).$$

Pf: $$\hat{x}_{eff}(y) = \underset{\hat{x}_{ML}(y).}{\underset{\shortparallel}{x}} + \underbrace{\frac{1}{J_Y(x)} \frac{\partial}{\partial x} \log f_Y(y;x)}_{\to 0}.$$

Since the est. $\hat{x}_{eff}$ is efficient, the RHS doesn't dep. on x.

So, let's substitute $x = \hat{x}_{ML}(y)$.

$$\left.\frac{\partial}{\partial x} \log f_Y(y;x)\right|_{x=\hat{x}_{ML}(y)} = 0.$$

$$\hat{x}_{ML}(y) = \underset{x}{\text{argmax}} \ \log f_Y(y;x)$$

Eg: $$f_Y(y;x) = \begin{cases} \frac{1}{x} e^{-y/x}, & y \geq 0 \\ 0 & y < 0 \end{cases}$$

$\mathbb{E}_x[Y] = x.$

$f_X(x) = \lambda e^{-\lambda x} \mathbb{1}\{x \geq 0\}$

$\mathbb{E}X = \frac{1}{\lambda}$, $\text{Var} X = \frac{1}{\lambda^2}.$

Observe Y. Compute the MLE

$$\log f_Y(y;x) = -\log x - y/x.$$

$$\frac{\partial}{\partial x} \log f_Y(y;x) = -\frac{1}{x} + \frac{y}{x^2} = 0 \qquad = \frac{y-x}{x^2}$$

$$\hat{x}_{ML}(y) = y.$$

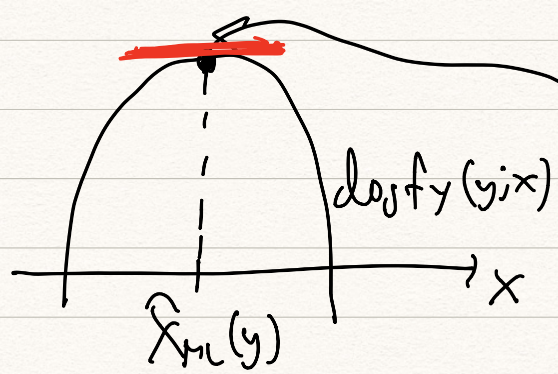
$$\lambda_{ML}(x) = \mathrm{Var}_x(\hat{x}_{ML}(Y)) = \mathrm{Var}(Y) = \underline{\underline{x^2}}$$

$$J_Y(x) = \mathbb{E}\left[\left(\frac{\partial}{\partial x}\log f_Y(Y;x)\right)^2\right]$$

$$= \mathbb{E}\left[\frac{(Y-x)^2}{x^4}\right] = \frac{1}{x^4}\mathrm{Var}(Y) = \frac{1}{x^2}.$$

$$\lambda_{ML}(x) = \frac{1}{J_Y(x)} \Rightarrow \text{CRLB is tight!}$$

Rmk: (96) $\because$ $\hat{x}_{ML}(y) = \underset{x}{\operatorname{argmax}} \log f_Y(y;x)$



$$\left.\frac{\partial}{\partial x}\log f_Y(y;x)\right|_{x=\hat{x}_{ML}(y)} = 0.$$

Eg: $f_Y(y;x) = \begin{cases} xe^{-xy}, & y \geq 0 \\ 0 & y < 0 \end{cases}$

$\mathbb{E}_x Y = \frac{1}{x}$ Is the MLE efficient? No!

$$\log f_Y(y;x) = \log x - xy$$

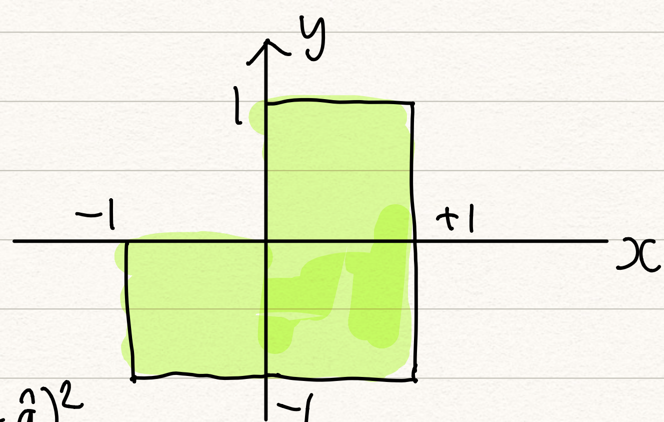
$$\frac{\partial}{\partial x} \log f_Y(y;x) = \frac{1}{x} - y = 0.$$

$$\hat{x}_{ML}(y) = \frac{1}{y}.$$

$$\underline{\mathbb{E}[\hat{x}_{ML}(Y)]} = \underline{\mathbb{E}[\tfrac{1}{Y}]} \neq \frac{1}{\mathbb{E}Y} = \frac{1}{\cancel{1/x}} = x.$$

If unbiased, $\mathbb{E}[\hat{x}_{ML}(Y)] = x$.

1.

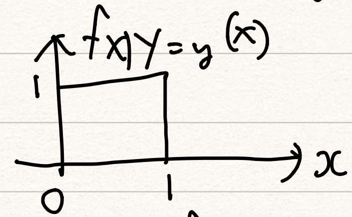


$C(a,\hat{a}) = (a-\hat{a})^2$

a) $\hat{x}_{BLS}(Y) = \mathbb{E}[X|Y]$

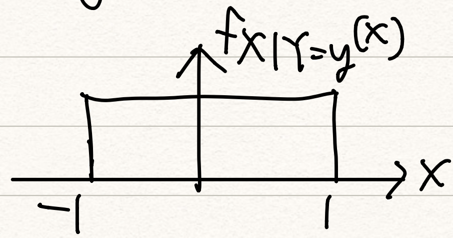
Attempt find $\mathbb{E}[X|Y=y]$

$y>0$



$\mathbb{E}[X|Y=y] = \frac{1}{2}.$

$y \leq 0$



$\mathbb{E}[X|Y=y] = 0.$

$f_{X|Y=y}(x) = f_{X|Y}(x|y)$

$$\hat{x}_{BLS}(y) = \begin{cases} \frac{1}{2} & y>0 \\ 0 & y \leq 0 \end{cases}$$

b) $\lambda_{BLS} = \underset{Y}{\mathbb{E}}[\underbrace{\lambda_{X|Y}}_{\text{Var}(X|Y=y)}].$

$= \frac{(b-a)^2}{12}$

$$\text{Var}(X|Y=y) = \begin{cases} 1/12 & y>0 \\ 1/3. & y \leq 0 \end{cases}$$

$\lambda_{BLS} = \Pr(Y>0)\frac{1}{12} + \Pr(Y \leq 0)\frac{1}{3}.$

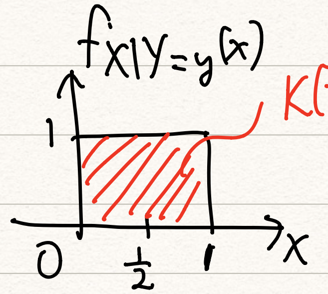
$= \quad 1/3 \cdot 1/12 + 2/3 \cdot 1/3 = 1/4.$

c) $C(x,\hat{x}) = \begin{cases} (x-\hat{x})^2 & x<0 \\ K(x-\hat{x})^2 & x \geq 0. \end{cases}$

$K \to 0^+$

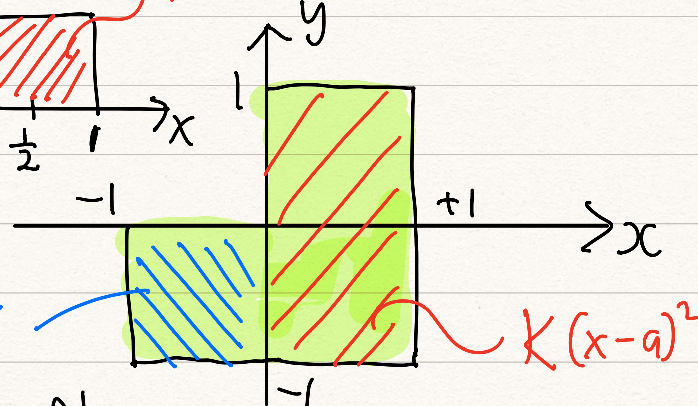
$K \to \infty.$

$\hat{x}_{MLS}(y)$



$\underline{y>0,\ x>0}$

$\hat{x}_{MLS}(y)$
$= \hat{x}_{BLS}(y) = \frac{1}{2}.$



$$\hat{x}_{MLS}(y) = \underset{a}{\operatorname{argmin}} \underbrace{\int_{-1}^{1} C(x,a) f_{X|Y}(x|y)\,dx}_{J(a)}.$$

$\in [-1,0]$

$\underline{y \leq 0}$

$$J(a) = \int_{-1}^{0} C(x,a) f_{X|Y}(x|y)\,dx + \int_{0}^{1} C(x,a) f_{X|Y}(x|y)\,dx$$

$$= \int_{-1}^{0} (x-a)^2 \frac{1}{2}\,dx + \int_{0}^{1} K(x-a)^2 \frac{1}{2}\,dx.$$

$$\frac{dJ}{da} = \frac{1-K}{2} + a(1+K) = 0$$

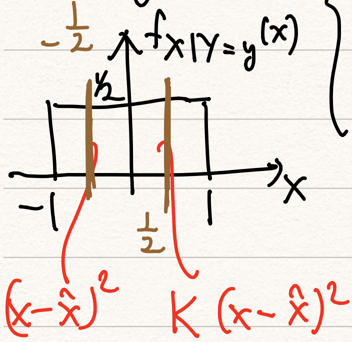
$\Downarrow$

$$a = \frac{1}{2}\,\frac{K-1}{K+1}$$

$$\boxed{\hat{x}_{MLS}(y) = \begin{cases} 1/2 & y > 0 \\ \frac{1}{2}\cdot\frac{K-1}{K+1} & y \le 0 \end{cases}}$$

d) $y<0$, cost f^{Δ} is a scaled version of original BLS cost f^{Δ} so estimator stays the same.

$y \ge 0$, $\begin{cases} K=1 & \hat{x}_{MLS}(y) = 0 \\ K\to\infty & \hat{x}_{MLS}(y) = \frac{1}{2} \\ K\to 0^{+} & \underline{\hat{x}_{MLS}(y) = -\frac{1}{2}.} \end{cases}$



---

4. $Y_1 = x + W_1$ $\quad$ $\mathrm{Var}\, W_1 = 1$

$x \ne 0.$ $\quad$ $Y_2 = x + W_2.$ $\quad$ $\mathrm{Var}\, W_2 = \begin{cases} 1 & x>0 \\ 2 & x<0. \end{cases}$

W_i : zm. Gaussian rvs. $W_1 \perp\!\!\!\perp W_2$.

a) $f_{\underline{Y}}(y; x) = \begin{cases} N(y_1; x, 1)\, N(y_2; x, 1) & x>0. \\ N(y_1; x, 1)\, N(y_2; x, 2) & x<0 \end{cases}$

$\underline{Y} = (Y_1, Y_2)$, $y = (y_1, y_2)$ $\qquad \exp\left(-\frac{(y-x)^2}{2\sigma^2}\right)$

$$\log f_{\underline{Y}}(y;x) \overset{c}{=} \begin{cases} -\frac{1}{2}(y_1-x)^2 - \frac{1}{2}(y_2-x)^2 & x>0 \\ -\frac{1}{2}(y_1-x)^2 - \frac{1}{4}(y_2-x)^2 & x<0 \end{cases}$$

$$\frac{\partial^2}{\partial x^2} \log f_{\underline{Y}}(y;x) = \begin{cases} -2 & x>0 \\ -\frac{3}{2} & x<0 \end{cases}$$

$$J_Y(x) = -\mathbb{E}\left[\frac{\partial^2}{\partial x^2} \log f_Y(y;x)\right]$$

$$= \begin{cases} 2 & x>0 \\ 3/2 & x<0. \end{cases}$$

Cramér-Rao LB. $\quad \lambda_{\hat{x}}(x) \geq \begin{cases} \frac{1}{2} & x>0 \\ \frac{2}{3}. & x<0 \end{cases}$

$\forall\ \hat{x}(\cdot)$

b) $\hat{X}_1 = \frac{Y_1+Y_2}{2}$, $\qquad \hat{X}_2 = \frac{2Y_1}{3} + \frac{Y_2}{3}$.

$$\mathbb{E}[\hat{X}_1(Y_1, Y_2)] = \mathbb{E}\hat{X}_1 = \frac{1}{2}\left(\mathbb{E}Y_1 + \mathbb{E}Y_2\right) = \frac{1}{2}(x+x)$$

$$= x.$$

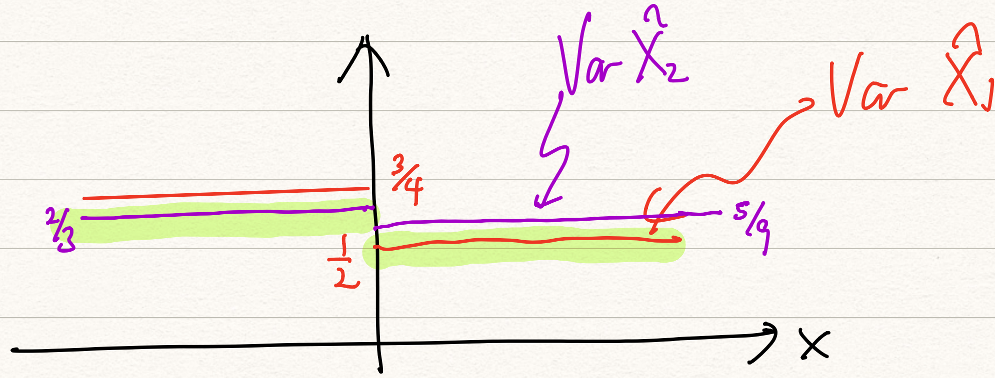
$\mathbb{E}[\hat{X}_2] = x.$

$$\mathrm{Var}(\hat{X}_1) = \frac{1}{4}\mathrm{Var}\, Y_1 + \frac{1}{4}\mathrm{Var}\, Y_2$$

$$= \begin{cases} \frac{1}{2}, & x > 0 \\ \frac{3}{4}, & x < 0 \end{cases}$$

$$\mathrm{Var}(\hat{X}_2) = \frac{4}{9}\mathrm{Var}\, Y_1 + \frac{1}{9}\mathrm{Var}\, Y_2$$

$$= \begin{cases} 5/9, & x > 0 \\ 2/3, & x < 0 \end{cases}$$



$$\lambda_{\hat{X}}(x) \geq \begin{cases} \frac{1}{2} & x > 0 \\ \frac{2}{3}. & x < 0 \end{cases}$$

$\hat{X}_1$ achieves the CRB if $x > 0$
$\hat{X}_2$ achieves the CRB if $x < 0$.

No single estimator that doesn't depend on x that achieves the CRB $\forall x$.

---

2. $$P_Y(y;q) = q(1-q)^y, \quad y \in \mathbb{N} \cup \{0\}.$$

MLE: $$\frac{\partial}{\partial q} \log P_Y(y;q) = \frac{\partial}{\partial q}\left[\log q + y \log(1-q)\right]$$

$$= \frac{1}{q} - \frac{y}{1-q} = 0$$

$$1 - q - qy = 0$$

$$\hat{q}_{ML}(y) = \frac{1}{1+y}.$$

$$\mathbb{E}[\hat{q}_{ML}(Y)] = \mathbb{E}\left[\frac{1}{1+Y}\right] = \sum_{y=0}^{\infty} \frac{q(1-q)^y}{y+1}$$

$$= q + \sum_{y \geq 1} \frac{q(1-q)^y}{y+1} > q.$$

Biased estimated.

b) $$\hat{x}_{BLS}(y) = \mathbb{E}[X \mid Y = y]$$

$$= \int x f_{X|Y}(x|y)\, dx$$

$$= \frac{\int_0^1 x f_{Y|X}(y|x) \underbrace{f_X(x)}_{=1} dx}{}$$

$$\int_0^1 f_{Y|X}(y|x') \underbrace{f_X(x')}_{=1} dx'$$

$$= \frac{\int_0^1 x \cdot x(1-x)^y dx}{\int_0^1 x(1-x)^y dx}$$

$$= \frac{2}{y+3}.$$

c) $R_{\hat{X}} = \mathbb{E}\left[\frac{X}{\hat{x}(Y)}\right] = 1$

i) $R_{ML} = \mathbb{E}\left[\frac{q}{\frac{1}{Y+1}}\right] = \mathbb{E}[(Y+1)q]$

$$= q\,\mathbb{E}[1+Y]$$

$$= q\left(1+\frac{1-q}{q}\right) = 1$$

$$R_{BLS} = \mathbb{E}\left[\frac{X}{2/Y+3}\right] = \frac{1}{2}\Big(\underbrace{\mathbb{E}[XY]}_{=\frac{1}{2}} + 3\underbrace{\mathbb{E}X}_{=1/2}\Big)$$

$$\mathbb{E}[XY] = \mathbb{E}[\mathbb{E}[XY|X]]$$

$$= \mathbb{E}[X\,\mathbb{E}[Y|X]]$$

$$= \mathbb{E}\left[\cancel{X}\frac{1-X}{\cancel{X}}\right] = \mathbb{E}[1-X] = \frac{1}{2}.$$

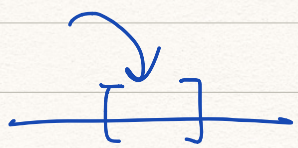
$$\mathbb{E}[Y|X=x] = \frac{1-x}{x}$$

$$\text{RBLS} = \frac{1}{2}\left(\frac{1}{2}+\frac{3}{2}\right) = 1.$$

Exam Format

Q1: Probability
→ Conditional Expectation
→ Moment Generating Function.
→ Convergence

Q2: Poisson processes
→ Properties
→ SIP, IIP



Q3: Markov chains
→ Classif. of states
→ Period
→ Stationary dist.
→ Conv. rate $|\lambda_2|$
→ Absorption.

Q4: Detection & Estimation

Hypothesis Test.
Sufficient statistic
Prob. of error q_0, q_1

Bias
Cramér - Rao Bound
Efficient
ML.

Randomize?