Estimation Theory.

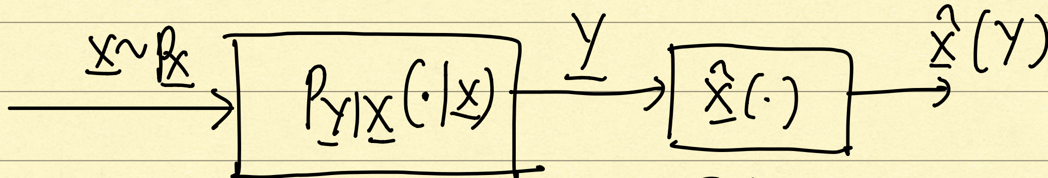
Last time "detection theory". Unknown $H \in \{H_0, H_1\}$.
Given Y, estimate H.

Modelled H: 1) H is random $\Pr(H=H_0)=p_0$
(Bayesian) $\Pr(H=H_1)=p_1$

2) No priors on H; H is non-random.
$\Pr(\text{err} \mid H_0) \le \varepsilon$ objective $\min \Pr(\text{err} \mid H_1)$.

Bayesian Parameter Estimation

Unknown continuous-valued $\underline{x} \in \mathcal{X} \subset \mathbb{R}^d$
Observations $\underline{y} \in \mathcal{Y}$



Estimator
$\hat{\underline{x}}(\cdot) : \mathcal{Y} \to \mathbb{R}^d$ or $\mathcal{X}$.

Bayesian case $\underline{x}$ has a prior dist^{n} $p_{\underline{x}}(\underline{x})$

belief about unknown $\underline{x}$.

Observation model / Likelihood $f_{\underline{y}|\underline{x}}(\cdot|\underline{x})$ or $p_{\underline{y}|\underline{x}}(\cdot|\underline{x})$
cts discrete

Ex: Y is a noise-corrupted measurement of some f^n of $\underline{X}$

$$\underline{Y} = h(X) + W. \qquad W \sim f_W(w)$$

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

$h(x) = Ax$ A matrix maps $\underline{x} \in \mathbb{R}^d$ to y.

$W \sim N(0, \Lambda)$

$$f_{Y|X}(y|x) = f_W(y - Ax)$$
$$= N(y; Ax, \Lambda)$$

$$f_W(w) = \frac{1}{(2\pi)^{d/2}|\Lambda|^{1/2}} \exp\left(-\frac{1}{2} w^T \Lambda^{-1} w\right).$$

evaluation pt

$$f_W(w) = \frac{1}{(2\pi)^{d/2}|\Lambda|^{1/2}} \exp\left(-\frac{1}{2}(y - Ax)^T \Lambda^{-1} (y - Ax)\right)$$

mean

covariance matrix

$$= N(y; Ax, \Lambda)$$

X has prior dist^n $f_X(\cdot)$

likelihood prior

Joint dist^n $f_{XY}(x,y) = f_{Y|X}(y|x) f_X(x)$

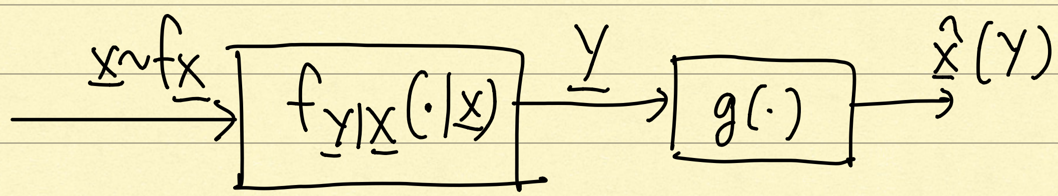
Posterior dist^n $f_{X|Y}(x|y) = \dfrac{f_{Y|X}(y|x) f_X(x)}{\ldots}$

$\int f_{Y|X}(y|x')f_X(x')dx.$

Estimator: Cost function $C: \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$

$C(x,\hat{x})$: cost of estimating x as $\hat{x}$.

$$\hat{x}(\cdot) := \underset{g(\cdot)}{\text{argmin}}\ \mathbb{E}[C(X, g(Y))] \quad \text{—} (*)$$



Rmk: $\mathbb{E}$ is over (X,Y)

$$(*)\ \mathbb{E}[C(X,g(Y))] = \int_{\mathcal{Y}} \int_{\mathcal{X}} C(x,g(y)) \overbrace{f_{XY}(x,y)}^{f_{X|Y}(x|y)f_Y(y)}\, dx\, dy$$

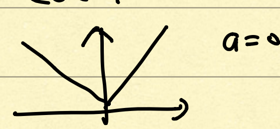
$$= \int_{\mathcal{Y}} \underbrace{\left[\int_{\mathcal{X}} C(x,g(y)) f_{X|Y}(x|y)\, dx \right]}_{= \mathbb{E}[C(X,g(Y))|Y=y]} \underbrace{f_Y(y)}_{\geq 0.}\, dy$$

Because $f_Y(y) \geq 0$, we can perform a pointwise min to find $\hat{x}(\cdot)$.

$$\hat{x}(y) = \underset{a}{\text{argmin}} \int_{\mathcal{X}} C(x,a) f_{X|Y}(x|y)\, dx.$$

Minimum Absolute-Error Estimation (scalar)

$$\boxed{C(a,\hat{a}) = |a - \hat{a}|}$$



Thm: The MAE estimate is the median of $f_{X|Y}(x|y)$.

Pf:

$$\hat{x}_{MAE}(y) = \underset{a\in\mathbb{R}}{\operatorname{argmin}} \int_{-\infty}^{\infty} |x-a|\, f_{X|Y}(x|y)\, dx$$

$$= \underset{a}{\operatorname{argmin}} \left[\int_{-\infty}^{a} |x-a|\, f_{X|Y}(x|y)\, dx + \int_{a}^{\infty} |x-a|\, f_{X|Y}(x|y)\, dx\right]$$

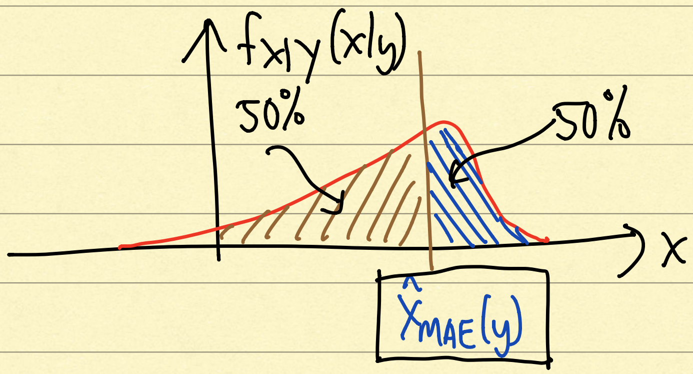
$$= \underset{a}{\operatorname{argmin}} \left[\int_{-\infty}^{a} (a-x)\, f_{X|Y}(x|y)\, dx + \int_{a}^{\infty} (x-a)\, f_{X|Y}(x|y)\, dx\right]$$

Differentiate [ ] using Leibniz's rule.

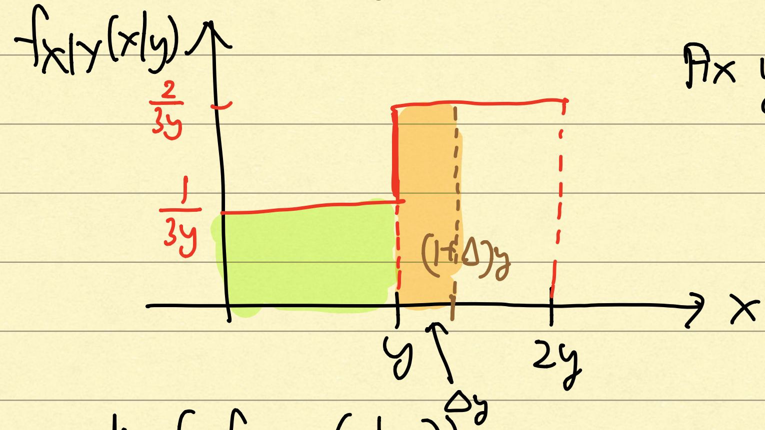
$$\frac{d}{da}\int_{b(a)}^{c(a)} g(a,x)\, dx$$

$$\frac{d}{da}[\cdots] = \int_{-\infty}^{a^*} f_{X|Y}(x|y)\, dx - \int_{a^*}^{\infty} f_{X|Y}(x|y)\, da = 0$$

$$\int_{-\infty}^{\hat{x}_{MAE}(y)} f_{X|Y}(x|y)dx = \int_{\hat{x}_{MAE}(y)}^{\infty} f_{X|Y}(x|y)\, dx.$$



Ex 1: $$f_{X|Y}(x|y) = \begin{cases} \frac{1}{3y} & 0<x<y \\ \frac{2}{3y} & y<x<2y \end{cases}$$



Fix y

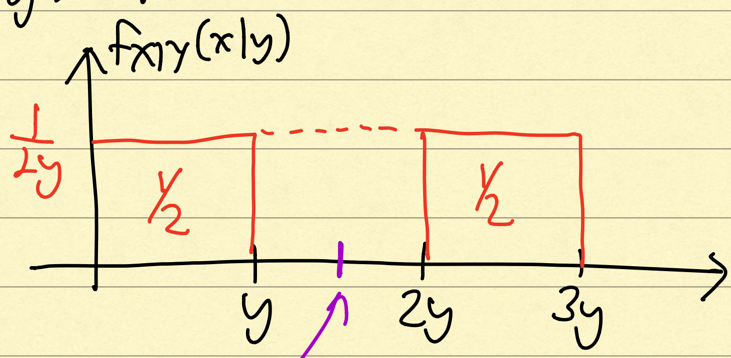
$\hat{X}_{MAE}(y) = \text{median}(f_{X|Y}(\cdot|y))$.

Let Δ be s.t. $\hat{X}_{MAE}(y) = (1+\Delta)\,y$. $\Delta > 0$.

$$\frac{1}{3y}\cdot y + \Delta y \cdot \frac{2}{3y} = \frac{1}{2}. \Rightarrow \Delta = \frac{1}{4}.$$

$\hat{X}_{MAE}(y) = \frac{5}{4}y$. median is unique.

Ex: $y > 0$.



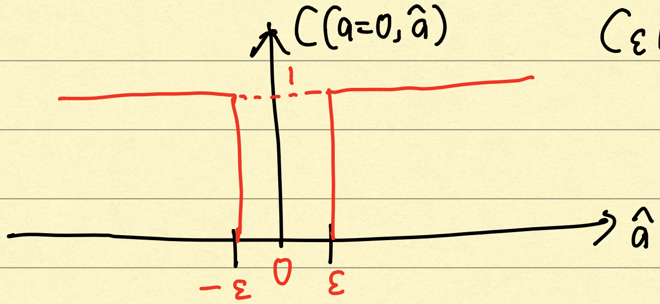
Any point in $[y, 2y]$ can be taken as $\hat{X}_{MAE}(y)$.

Maximum A Posterior Estimation

Minimum Uniform Cost (MUC)

$$C_\varepsilon(a,\hat{a}) = \begin{cases} 1 & |a-\hat{a}| > \varepsilon \\ 0 & \text{else.} \end{cases}$$

$C_\varepsilon(a,\hat{a}) = 1\{|a-\hat{a}|>\varepsilon\}$



Rmk: Uniformly penalizes all errors $> \varepsilon$.

Thm: In the limit $\varepsilon \to 0$, MUC estimate returns the mode of the belief/posterior $f_{X|Y}(\cdot|y)$.

$$\hat{X}_{MAP}(y) = \underset{a}{\operatorname{argmax}}\ f_{X|Y}(a|y).$$

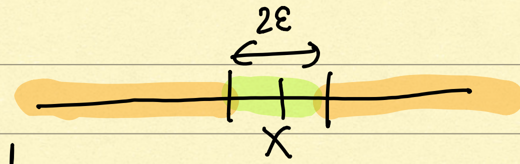
Pf. Substitute this cost f^n into

$$\hat{x}(y) = \underset{a}{\operatorname{argmin}} \int_{\mathcal{X}} 1\{|a-x|>\varepsilon\} f_{X|Y}(x|y)\,dx.$$

$$= \underset{a}{\operatorname{argmin}} \int_{\mathcal{X}} (\cancel{1} - 1\{|x-a|\le\varepsilon\}) f_{X|Y}(x|y)\,dx.$$

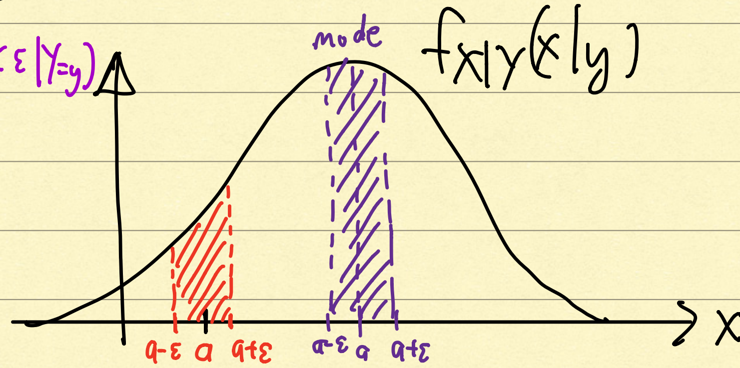
||

||



$$= \underset{a}{\operatorname{argmax}} \int_{a-\varepsilon}^{a+\varepsilon} f_{X|Y}(x|y)\, dx.$$

$$= \underset{a}{\operatorname{argmax}} \Pr(|X-a|<\varepsilon \mid Y=y)$$



$$\lim_{\varepsilon \to 0^+} \hat{x}_{MUC,\varepsilon}(y) = \underset{a \in \mathbb{R}}{\operatorname{argmax}} f_{X|Y}(a|y)$$

mode of dist^n.

Bayes Least Squares Estimation

$$C(\underline{a}, \hat{\underline{a}}) = \|a - \hat{a}\|^2 = (a-\hat{a})^T(a-\hat{a}) = \sum_{i=1}^{N} (a_i - \hat{a}_i)^2$$

Mean-squared error (MSE)

Thm: The BLS estimate is the conditional mean of the belief/ posterior $f_{X|Y}(\cdot|y)$.

$$C(x,a) = (x-a)^2$$

i.e, $\hat{x}_{BLS}(y) = E[X|Y=y]$.

Pf: $\hat{x}_{BLS}(y) = \underset{a}{\operatorname{argmin}} \int_{-\infty}^{\infty} (x-a)^2 f_{X|Y}(x|y)\, dx.$

Differentiating wrt a

$$\frac{\partial}{\partial a}\int_{-\infty}^{\infty}(x-a)^2 f_{X|Y}(x|y)\,dx.$$

$$= \int_{-\infty}^{\infty} \cancel{-2}(x-a)\, f_{X|Y}(x|y)\,dx = 0$$

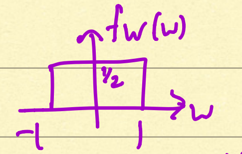
$$\int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\,dx = \underbrace{a^*}_{\hat{x}_{BLS}(y)} \cancelto{1}{\int_{-\infty}^{\infty} f_{X|Y}(x|y)\,dx}$$

$$\Rightarrow \quad \hat{x}_{BLS}(y) = \int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\,dx = E[X|Y=y].$$

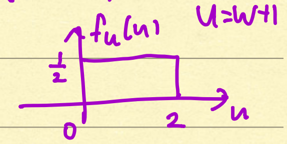
Ex: $X \perp\!\!\!\perp W$ $\qquad X, W \sim \underline{\text{Unif}[-1,1]}$.

$$Y = \underbrace{\overbrace{\text{sgn}(X)}^{\in\{-1,+1\}}}_{+1,\ x\geq 0} + W.$$

$$\text{sgn}(t) = \begin{cases} 1 & t \geq 0 \\ -1 & t < 0 \end{cases}$$

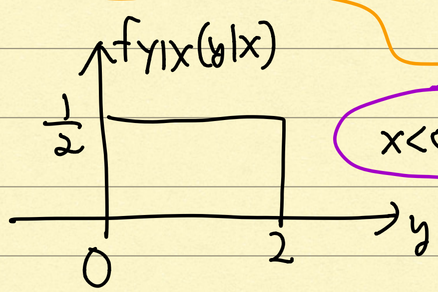


$f_{Y|X}(\cdot|x) \sim \text{Unif}[0,2] \qquad x \geq 0$



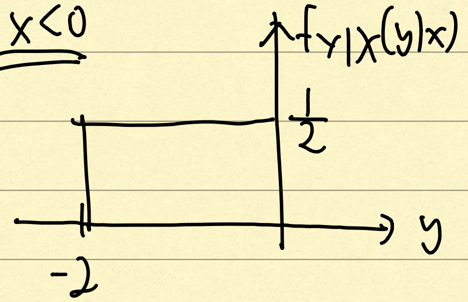
$f_{Y|X}(\cdot|x) \sim \text{Unif}[-2,0] \qquad x < 0$

$x \geq 0 \qquad f_{Y|X}(y|x) = \begin{cases} 1/2 & 0 < y < 2 \\ 0 & \text{else} \end{cases}$



$x < 0 \qquad f_{Y|X}(y|x) = \begin{cases} \frac{1}{2}, & -2 < y < 0 \\ 0, & \text{else} \end{cases}$

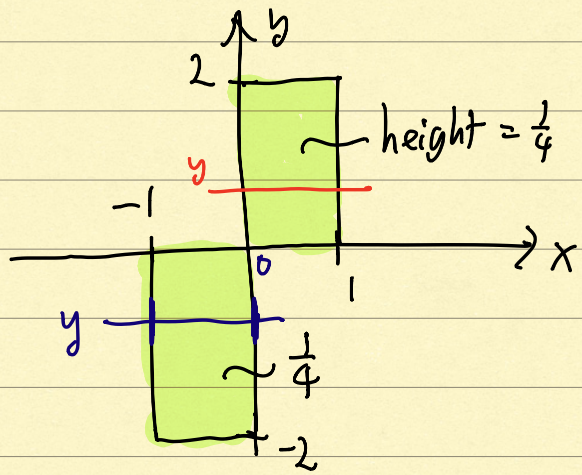
$\underline{x<0}$

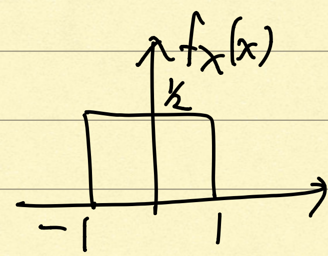


$f_{XY}(x,y) = f_{Y|X}(y|x)\,\underline{f_X(x)}$
$= \frac{1}{2} \cdot \frac{1}{2}$

Joint density $\underline{f_{XY} = f_{Y|X} \cdot f_X}$

$$f_{XY}(x,y) = \begin{cases} \frac{1}{4} & 0 \le x \le 1,\ 0<y<2 \\ \frac{1}{4} & -1 \le x < 0,\ -2<y<0 \\ 0 & \text{else} \end{cases}$$

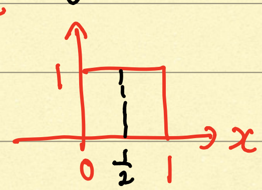




$$\hat{X}_{BLS}(y) = \mathbb{E}[X \mid Y=y]$$

Suppose $\underline{y \in [0,2]}$, $\underline{f_{X|Y}}(x|y) \sim$ Unif $[0,1]$

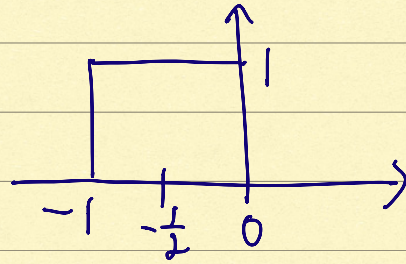
$$\hat{x}_{BLS}(y) = \int_0^1 x f_{X|Y}(x|y)\,dx$$



$$= \int_0^1 x\,dx = \frac{1}{2}$$

Suppose $y \in [-2, 0)$, $f_{X|Y}(x|y)$

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



$$\hat{x}_{BLS}(y) = \begin{cases} \frac{1}{2} & y \in [0, 2] \\ -\frac{1}{2} & y \in [-2, 0) \end{cases}$$

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

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

$$\mathrm{Var}(X|Y=y) = \frac{1}{12}$$

$y \in [-2, 2]$

NonBayesian Parameter Estimation (x is non-random)

cf. Bayesian Parameter Est. $X \sim f_X$

Observation model $f_Y(y; x)$

Eg: 1) $p_Y(y; x) = \frac{e^{-x} x^y}{y!}$, $y = 0, 1, \cdots$

Poisson parameterized by $x = \mathbb{E}[Y]$.

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

Gaussian parameterized by $x = \mathbb{E}[Y]$.

3) $f_Y(y;x) = \begin{cases} 1/x & y \in [0,x] \\ 0 & \text{else} \end{cases}$

Unif. dist. parameterized by right support point x.

$f_{\underline{Y}}(y;\underline{x})$

$$\mu_{\underline{Y}}(\underline{x}) = \mathbb{E}_{\underline{x}}[\underline{Y}] = \int_{-\infty}^{\infty} y f_Y(y;x)\, dy$$

$$\Lambda_{\underline{Y}}(\underline{x}) = \mathrm{Cov}_{\underline{x}}(\underline{Y}) = \int_{-\infty}^{\infty} (y-\mu_Y(x))(y-\mu_Y(x))^T f_Y(y;x)\, dy$$

Scalar case $\quad \mathrm{Cov}_x(Y) = \mathrm{Var}_x(Y)$
$$= \mathbb{E}_x[(Y - \mathbb{E}_x[Y])^2].$$

x: scalar.

$$\hat{x}(\cdot) = \underset{g(\cdot)}{\mathrm{argmin}}\ \mathbb{E}_Y[(x - \underset{\overset{\|}{\hat{x}(Y)=x}}{g(Y)})^2].$$

$$\| $$

$$\int_{-\infty}^{\infty} f_Y(y;x)(x-g(y))^2\, dy$$

random

If $\hat{x}(Y) = x$, then the above cost $= 0$

But this is unreasonable $\because$ x is precisely what we're trying to estimate.

Def: An estimator $\hat{x}: \mathcal{Y} \to \mathcal{X}$ is valid if it does not depend on the parameters estimated.

Bias, Error Covariance

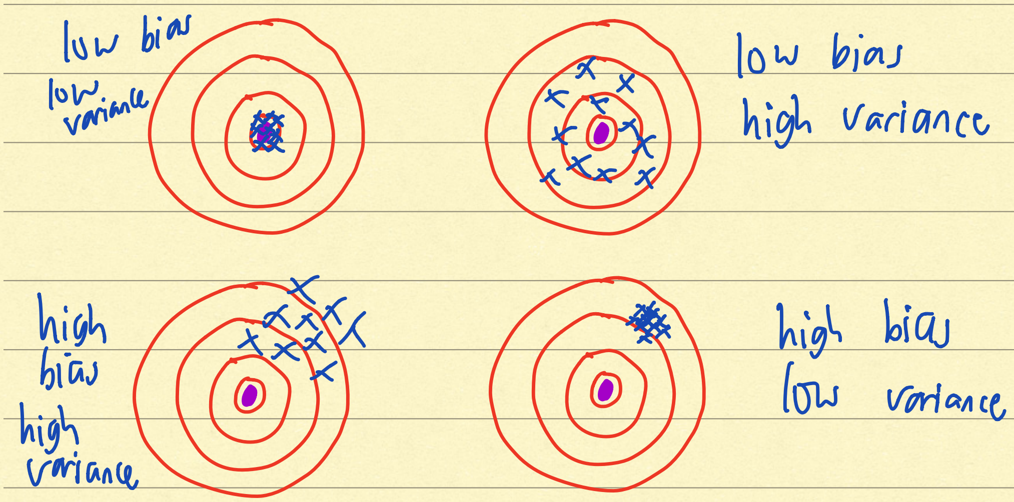
$\mathrm{Var}(X) = \mathbb{E}[(X-\mathbb{E}X)^2]$

$\hat{x}(\cdot) : \mathcal{Y} \to \mathcal{X}$

Error $E = e(Y) = \hat{x}(Y) - x = \hat{X} - x$

Bias: $b_{\hat{X}}(x) = \mathbb{E}[e(Y)] = \mathbb{E}[\hat{x}(Y)] - x.$

Error Covariance: $\Lambda_E(x) = \mathbb{E}[(e(Y) - b_{\hat{X}}(x))(e(Y) - b_{\hat{X}}(x))^T].$



Def: An estimator $\hat{x}(\cdot) : \mathcal{Y} \to \mathcal{X}$ for a non-random parameter x is <u>unbiased</u> if $b_{\hat{X}}(x) = 0 \quad \forall x \in \mathcal{X}$.

Rmk: $\Lambda_E(x) = \Lambda_{\hat{X}}(x)$

Error covariance "covariance of estimator"

$$\text{LHS} = \mathbb{E}\left[(e(Y) - b_{\hat{x}}(x))(e(Y) - b_{\hat{x}}(x))^T\right].$$

$$= \text{Cov}(e(Y))$$

$$= \text{Cov}(\hat{x}(Y) - \underbrace{\cancel{x}}_{\text{deterministic}}) = \text{Cov}(\hat{x}(Y)) = \Lambda_{\hat{x}}(x).$$

Prop: $\text{MSE} = \mathbb{E}[e(Y)e(Y)^T]$.

$$= \underbrace{\Lambda_{\hat{x}}(x)}_{} + \underbrace{b_{\hat{x}}(x)\, b_{\hat{x}}(x)^T}_{}$$

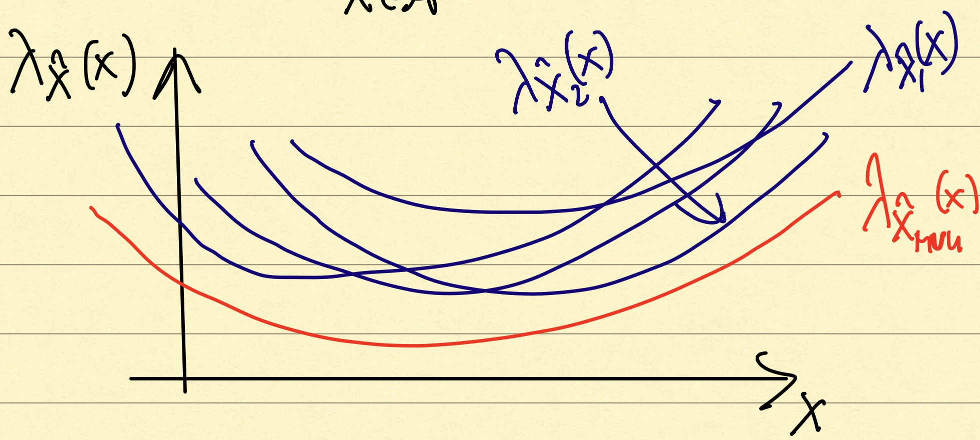
MSE = variance + squared bias.

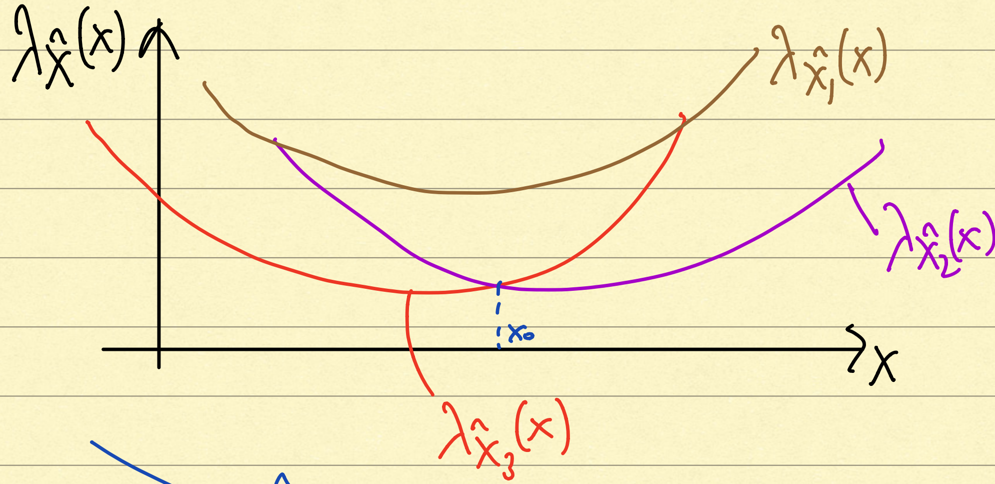
Def: An <u>admissible</u> estimator is one that is valid & unbiased.

$$\mathcal{A} = \{\hat{x}(\cdot) ; \hat{x}(\cdot) \text{ valid } \& \ b_{\hat{x}}(x) = 0 \ \forall x\}$$

Def: Minimum variance unbiased estimators (MVUE)

$$\hat{x}_{MVU}(\cdot) = \underset{\hat{x} \in \mathcal{A}}{\text{argmin}} \ \lambda_{\hat{x}}(x) \quad \forall x.$$





$$\hat{x}_{ZSS} = \xcancel{\begin{cases} \hat{x}_2 & \text{when } x < x_0 \\ \hat{x}_3 & \text{when } x \geq x_0 \end{cases}}$$ not a valid estimator.

Rmk: An MVU estimator may not exist.

Restrict ourselves to admissible estimators (unbiased)

MSE $\downarrow$ min variance $(\hat{x}(\cdot))$.

<u>Thm</u>: If $f_Y(y;x)$ satisfies the reg. condition

$$\mathbb{E}\left[\frac{\partial}{\partial x} \log f_Y(y;x)\right] = 0 \quad \text{—} (*)$$

then for any $\hat{x} \in \mathcal{A}$,

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

where

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

is the Fishe info. in Y about X.

Rmk: $(*)$

$$\mathbb{E}\left[\frac{\partial}{\partial x} \log f_Y(y;x)\right]$$

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

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

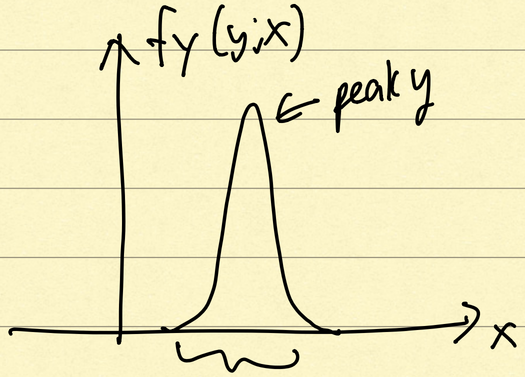
$$= \frac{\partial}{\partial x} \int f_Y(y;x)\, dy = 0.$$

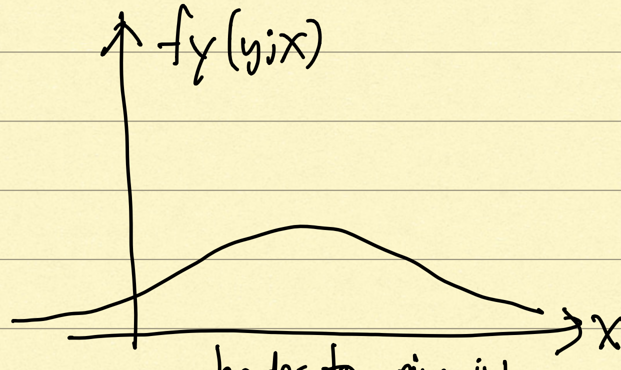
$$\boxed{\mathrm{Var}_x(\hat{x}(Y)) \geq \frac{1}{J_Y(x)}}$$ Why?

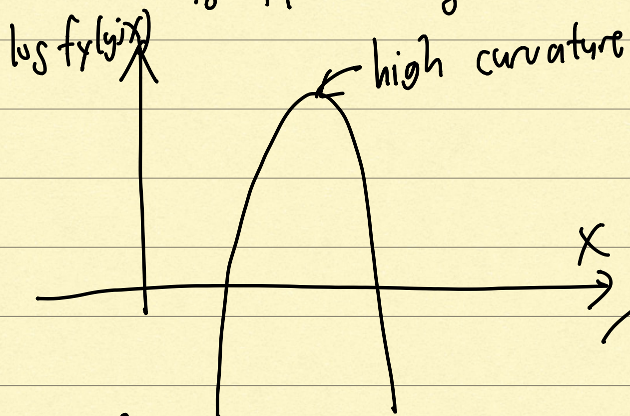
Prop: $$J_Y(x) = \mathbb{E}\left[\left(\frac{\partial}{\partial x} \log f_Y(y;x)\right)^2\right]$$

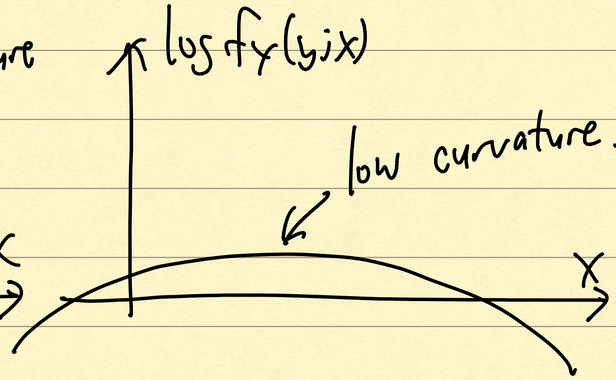
$$= \mathbb{E}\left[-\frac{\partial^2}{\partial x^2} \log f_Y(y;x)\right]$$

<u>2 different scenarios</u>









$-\frac{\partial^2}{\partial x^2} \log f_Y(y;x)$ measure curvature

Fisher information large

Fishe inf. Small.

CRLB $\quad Var(\hat{x}(Y)) \geq \underbrace{\frac{1}{\text{large}}}_{\text{small.}}$

$Var(\hat{x}(Y)) \geq \underbrace{\frac{1}{\text{small}}}_{\text{large}}$