Estimation Theory

Prototype problem: Estimate $\vec{x}$ based on $\vec{y}$.
the value of — on observations of a related vector

Ex: $\vec{X} = \{$position, velocity$\}$ of an aircraft. $\vec{y}$: {radar return measurements from several sensors}.

Two different views.

① $\vec{x}$ is random. $p_{\vec{y}|\vec{x}}(\vec{y}|\vec{x})$

② $\vec{x}$ is a parameter. $p_{\vec{y}}(\vec{y};\vec{x})$.

\# Estimation of random vectors: a Bayesian formulation.

"Bayesian estimation theory"

$p_{\vec{x}}(\vec{x})$: prior density of $\vec{x} \in \mathbb{R}^n$.

$p_{y|x}(\vec{y}|\vec{x})$: measurement model.

$$p_{x|y}(\vec{x}|\vec{y}) = \frac{p_{y|x}(\vec{y}|\vec{x})\, p_{\vec{x}}(\vec{x})}{p_y(\vec{y})} \quad : \text{posterior density}$$

Estimator $\hat{x}(y)$: ① choosing a deterministic scalar valued $C(\vec{a}, \hat{a})$ that specifies the cost of estimating $\vec{a}$ as $\hat{a}$.

Bayesian framework: ② choose the estimator the minimizes the average cost.

$$\hat{x}(\cdot) = \arg\min_{f(\cdot)} \underbrace{E\left[C(\vec{x}, f(\vec{y}))\right]}_{\text{over } \vec{x}, \vec{y} \text{ jointly}}$$

$$E\left[C(\vec{x}, f(\vec{y}))\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} C(\vec{x}, f(\vec{y}))\, p_{x,y}(\vec{x},\vec{y})\, d\vec{x}\, d\vec{y}.$$

$$= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} C(\vec{x}, f(\vec{y}))\, p_{x|y}(\vec{x}|\vec{y})\, d\vec{x}\right] \underbrace{p_y(\vec{y})}_{>0}\, d\vec{y}$$

$$\Rightarrow \quad \hat{X}(y) = \arg\min_{a} \int_{-\infty}^{\infty} C(\vec{x}, \vec{a})\, p_{x|y}(\vec{x}|\vec{y})\, d\vec{x}.$$

$$= \arg\min_{a} \int_{-\infty}^{\infty} C(\vec{x}, \vec{a})\, p_{y|x}(\vec{y}|\vec{x})\, p_x(\vec{x})\, d\vec{x}.$$

① Minimum Absolute-Error Estimation (MAE)

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

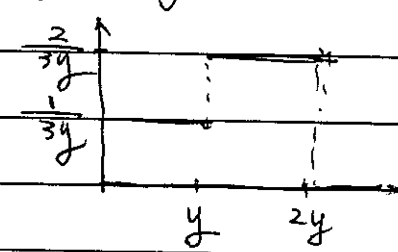
(scalar case, vector can be handled in a component-wise manner).

$\hat{X}_{MAE}(\vec{y})$: $\displaystyle\int_{-\infty}^{\hat{x}_{MAE}(y)} p_{x|y}(x|y)\, dx = \int_{\hat{x}_{MAE}(y)}^{+\infty} p(x|y)\, dx = \frac{1}{2}$

Median of a probability density.

Example

$$p_{x|y}(x|y) = \begin{cases} \frac{1}{3y}, & 0 < x < y \\ \frac{2}{3y}, & y < x < 2y \\ 0. & \text{o.w.} \end{cases}$$



Try $\hat{x}_{MAE}(y) = (1+\Delta)y$ for $\Delta > 0$. then: $\frac{1}{3y}\cdot y + \frac{2}{3y}\cdot\Delta y = \frac{1}{2} \Rightarrow \Delta = \frac{1}{4}$.

Note that: the median of a density is not necessarily unique.

② Maximum A posteriori Estimation (MAP).

$$C(a,\hat{a}) = \begin{cases} 1 & |a-\hat{a}| > \varepsilon \\ 0 & \text{o.w.} \end{cases}$$, which uniformly penalizes all est. errors with magnitude bigger than ε.

$$\hat{X}_{\underline{MUC}} = \arg\min_a \left[1 - \int_{a-\varepsilon}^{a+\varepsilon} p(x|y)\,dx \right] = \arg\max_a \int_{a-\varepsilon}^{a+\varepsilon} p(x|y)\,dx$$

~~Max:~~ Minimum uniform cost (MUC)

that: find the interval of length 2ε where the posterior density $p(x|y)$ is most ~~concentrated~~

Let $\varepsilon \to 0$. $\hat{X}_{MAP} = \arg\max_a p(a|y) = \lim_{\varepsilon\to 0} \hat{X}_{MUC}(y)$.

the peak value of a density is referred to as its <u>mode</u>

Generalize to vector form. $\hat{X}_{MAP} = \arg\max_{\vec{a}} p_{\vec{x}|\vec{y}}(\vec{a}|\vec{y})$

Cond.

① $\frac{\partial}{\partial x} P_{x|y}(x|y) = 0$ ② $\frac{\partial^2}{\partial x^2} p(x|y) < 0$.

<u>Bias and Variance</u>:

$$e(\vec{x},\vec{y}) = \hat{X}(\vec{y}) - \vec{X}$$

(I) Bias. $\vec{b} = \langle \hat{X}(\vec{y}) - \vec{X} \rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (\hat{X}(y) - \vec{X})\, p(\vec{x},\vec{y})\, d\vec{x}\, d\vec{y}$.

(II) Error Variance: $\Lambda_e = \langle (e(\vec{x},\vec{y}) - b)(e(\vec{x},\vec{y}) - b)^T \rangle$

$$\langle ee^T \rangle = \Lambda_e + bb^T.$$

③ Bayes' Least-Square Estimation (BLS).

$$\cancel{C(a,\hat{a}) = \|a-\hat{a}\|^2 = (a-\hat{a})^T(a-\hat{a})}.$$

$$\hat{X}_{BLS}(\vec{y}) = \arg\min_a \int_{-\infty}^{\infty} (X-a)^T(X-a)\, p(x|y)\, d\vec{x}.$$

$$\frac{\partial}{\partial a}: \quad \cancel{\int_{-\infty}^{\infty} (X-\hat{a})\, p(x|y)\, d\vec{x} = 0} \Rightarrow$$

$$a = \cancel{\int_{-\infty}^{\infty} X\, p(x|y)\, d\vec{x} = E[\vec{X}|y=y]}$$

$$\hat{X}_{BLS}(y) = a = E[X|Y=y] \quad \text{: the mean of the posterior density } P_{X|y}(x/y).$$

when $a = \hat{X}_{BLS}(y)$. $\quad e(x,y) = \hat{X}(y) - x.$

(I) $\hat{X}_{BLS}$ is always unbiased: $E[\hat{X}_{BLS}(y)] = E[E[X|Y]] = E[X]$!

(II) $\Lambda_{BLS} = E[ee^T] = E[(X-E[X|Y])(X-E[X|Y])^T]$

$$= E\left[E[(X-E[X|Y])(X-E[X|Y])\,|\,Y]\right]$$

$$= E[\Lambda_{X|Y}(Y)] \quad \text{Covariance of the posterior prob}$$

(III) Minimum value of the expected cost objective func.

$$E[C(X,\hat{X}_{BLS}(y))] = E[(E(X|Y)-X)^T(E(X|y)-X)]$$

$$= E[\mathrm{tr}\,(E[X|y]-X)(E[X|y]-X)^T] = \mathrm{tr}\, E[(E(X|y)-X)(E(X|y)-X)^T]$$

$$= \mathrm{tr}\,\Lambda_{BLS}$$

$\hat{X}_{BLS}(y)$:

Example. $y = \operatorname{sgn} x + w$, where $x \sim \text{uni}[-1,1]$

$w \sim \text{uni}[-1,1]$

x, w indep. w each other.

$$p(y|x) = \begin{cases} 1/2, & y \in [0,2] \\ 0 & \text{o.w.} \end{cases} \quad \text{when } x > 0.$$

$$= \begin{cases} 1/2 & y \in [-2,0] \\ 0. & \text{o.w.} \end{cases} \quad \text{when } x < 0.$$

$$\Rightarrow p_{x,y}(x,y) = p(y|x)\,p(x) = \begin{cases} 1/4 & x \in [0,1],\ y \in [0,2] \\ 1/4 & x \in [-1,0],\ y \in [-2,0] \\ 0 & \text{o.w.} \end{cases}$$

$$\Rightarrow p(x|y) = \begin{cases} 1 & x \in [0,1] \\ 0. & \text{o.w.} \end{cases}, \text{ when } y > 0.$$

$$\begin{cases} 1 & x \in [-1,0] \\ 0, & \text{o.w} \end{cases} \quad \text{when } y < 0$$

$$\Rightarrow \hat{x}_{BLS}(y) = E[x|Y=y] = \frac{1}{2}\operatorname{sgn}(y) = \begin{cases} 1/2, & y > 0 \\ -1/2, & y < 0 \end{cases}.$$

Error variance $\lambda_{x|y}(y) = \frac{1}{12} \Rightarrow \lambda_{BLS} = E[\lambda_{x|y}(y)] = 1/12$.

Chebyshev's Inequality:

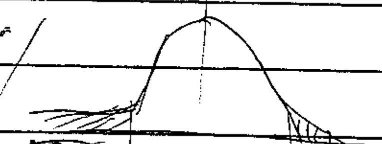
Let X be a random variable with expectation μ and finite variance σ^2.
~~Then for any real number $k>0$,~~

$$\Pr\left[\,|X-\mu| \geq k\sigma\,\right] \leq \frac{1}{k^2}$$

Let $Y = X-\mu$. then $\langle y \rangle = 0$.

Proof: $$\Pr\left[\,|X-\mu| \geq k\sigma\,\right] = \int_{-\infty}^{-k\sigma} p(y)\,dy + \int_{k\sigma}^{\infty} p(y)\,dy$$

~~$\Pr[|y| \geq k\sigma]$ $|y|^2 \geq k\sigma$~~



$|X-\mu| \geq$ $|y| \geq k\sigma$

$\Rightarrow y^2 \geq k^2\sigma^2$.

$\Rightarrow P[|y| \geq k\sigma] \leq P[y^2 \geq k^2\sigma^2] = \int$

$\int_{k\sigma}^{\infty} y^2$

$\leq \int_{k\sigma} \frac{y^2}{k^2\sigma^2} p(y)\,dy$

$$\leq \int_{-\infty}^{-k\sigma} \frac{y^2}{k^2\sigma^2} p(y)\,dy + \int_{k\sigma}^{\infty} \frac{y^2}{k^2\sigma^2} p(y)\,dy$$

$$\leq \frac{1}{k^2\sigma^2} \int_{-\infty}^{\infty} y^2 p(y)\,dy = \frac{1}{k^2}.$$

① Markov's Inequality:

$$\Pr\left[\,|X| \geq a\,\right] \leq \frac{E[|X|]}{a} \quad a>0.$$

② ~~Chebyshev's Inequality:~~

$$\Pr\left[\,|X-\mu| \geq a\,\right] \leq \frac{Var[X]}{a^2}.$$

$\int_{-\infty}^{a} p(y)\,dy + \int$ $\int_{\Omega:|y|\geq a} p(y)\,dy = \frac{Var[Y]}{a^2}$

$\leq \int_{\Omega(|y|\geq a)} \frac{y^2}{a^2} p(y)\,dy$

Multi-dim Chebyshev's Inequality:

Thm. Let X be an N-dim random vector with $\mu = E[X]$ and cov. matrix

$$V = E[(X-\mu)(X-\mu)^T].$$ If $V \succ 0$, then for any $t > 0$,

$$\Pr\left[\sqrt{(X-\mu)^T V^{-1}(X-\mu)} > t\right) \le \frac{N}{t^2}, \quad \text{with } N = \mathrm{tr}(V^{-1}V)$$

Proof: Define $Y = (X-\mu)^T V^{-1}(X-\mu)$. y is positive.

Use Markov's Inequality:

$$\Pr[Y \geq t] = \Pr[\sqrt{y} \geq t] \Rightarrow \Pr[y > t^2] \leq \frac{E[y]}{t^2}$$

$$E[Y] = E[(X-\mu)^T V^{-1}(X-\mu)] = \mathrm{tr}(V^{-1}V] = N.$$

Additional Properties of BLS Estimators.

Thm. [orthogonality] An estimator $\hat{x}(\cdot)$ is the Bayes' Least square estimator, i.e., $\hat{x}(\cdot) = \hat{x}_{BLS}(\cdot)$ if and only if the associated estimation error $e(x,y) = \hat{x}(y) - x$ is orthogonal to any (vector-valued) function $g(\cdot)$ of the data, i.e.,
$$E\left[(\hat{x}(y) - x)\, g^T(y)\right] = 0.$$

Prof: rewrite cond. as: $\underbrace{E\left[x\, g^T(y)\right]}_{\text{w.r.t. both } x, y} = \underbrace{E\left[\hat{x}(y)\, g^T(y)\right]}_{\text{w.r.t. } x}$

$$E\left[x g^T(y)\right] = E\left[E\left[x g^T(y) \mid y\right]\right] = E\left[E[x|y] \cdot g^T(y)\right]$$

① Suppose $\hat{x}(\cdot) = \hat{x}_{BLS}(\cdot)$. Then, $\hat{x}(y) = E[x|y] \Rightarrow$

② Suppose: $E\left[\left(E[x|y] - \hat{x}(y)\right) g^T(y)\right] = 0$, hold for any $g^T(y)$.

Choose $g(y) = E[x|y] - \hat{x}(y)$, then: $\underbrace{E\left[\left(E[x|y] - \hat{x}(y)\right)\left(E(x|y) - \hat{x}(y)\right)^T\right]}_{\text{w.r.t. } y} = 0.$

$E[z z^T] = 0 \Rightarrow z = 0$, $Pr[z=0] = 1$.

Convergence in covariance.

$\Rightarrow E[x|y] = \hat{x}(y)$.

////

Thm. Let Λ_e be the error cov. of any estimator $\hat{X}(\cdot)$. Then the error cov. of the BLS estimator, i.e. Λ_{BLS} satisfies:

$$\Lambda_{BLS} \preccurlyeq \Lambda_e. \quad \text{with equality iff} \quad \hat{X}(y) - E[\hat{X}(y) - X] = \hat{X}_{BLS}(x|y) = E[X|Y].$$

Prop. BLS is guaranteed to yield less uncertainty in the value of X than any other estimator — biased or unbiased.

Proof: Let $b = E[\hat{X}(y) - X]$ be the bias of $\hat{X}(\cdot)$.

- Let $g(y) = \hat{X}(y) - \hat{X}_{BLS}(y) - b$.

$$\Lambda_e = E\left[(\hat{X}(y) - X - b)(\hat{X}(y) - X - b)^T\right]$$

$$= E\left[[g(y) + (\hat{X}_{BLS}(y) - X)][g(y) + (\hat{X}_{BLS}(y) - X)]^T\right]$$

$$= E[g(y)g(y)^T] + E\left[(\hat{X}_{BLS}(y) - X)(\hat{X}_{BLS}(y) - X)^T\right]$$

$$+ \underbrace{E\left[(\hat{X}_{BLS}(y) - X)g^T(y)\right]}_{=0} + \underbrace{E\left[(\hat{X}_{BLS}(y) - X)g^T(y)\right]^T}_{=0}$$

$$\Rightarrow \Lambda_e - \Lambda_{BLS} = E[g(y)g^T(y)] \succcurlyeq 0. \quad \text{with equality iff } g(y) = 0.$$

Nonrandom Parameter Estimation:

$p(\vec{y};\vec{x})$. treat $\vec{x}$ as parameters, not a random vector.

① $\hat{x}(\cdot) = \arg\min_f E[(\vec{x}-f(y))^2]$ is not valid.

$\Rightarrow \hat{x}(\cdot) = x$.

② Bias & Variance: $e(\vec{y}) = \hat{x}(y) - \vec{x}$. only a random w.r.t. $\vec{y}$.

$b_{\hat{x}}(x) = E[e(\vec{y})] = \int_{-\infty}^{+\infty} \hat{x}(y) p_y(y;x)\,dy - x$.

Error variance: $\Lambda_e(x) = E[[e(y)-b_{\hat{x}}(x)][e(y)-b_{\hat{x}}(x)]^T]$. w.r.t. $\vec{y}$.

$E[e(y)e^T(y)] = \Lambda_e(x) + b_{\hat{x}}(x) b_{\hat{x}}(x)^T$

Note that. $\Lambda_e(x) = \Lambda_{\hat{x}}(x) = \mathrm{Var}[\hat{x}(y)-\vec{x}] = \mathrm{Var}[\hat{x}(y)-\vec{x}-b_{\hat{x}}(x)] = \Lambda_e(x)$.

Scalar case: $\lambda_e(x) = \lambda_{\hat{x}}(x)$.

③ Minimum-Variance Unbiased Estimators. (MVU)

$A = \{\hat{x}(\cdot) \mid \hat{x}(\cdot) \text{ is valid and } b_{\hat{x}}(x) = 0\}$. valid and unbiased (for all x).

$\hat{x}_{MVU}(\cdot) = \arg\min_{\hat{x}\in A} \lambda_{\hat{x}}(x)$. for all x.

④ Cramer-Rao bound.

Thm. Consider the est. of an unknown scalar parameter x given measurement vector $\vec{y}$ with density $p_y(\vec{y};x)$. When it exists, the Cramer-Rao Bd gives a lower bd on the variance of any valid unbiased estimator $\hat{x}(\cdot)$ for x.

$\lambda_{\hat{x}}(x) \geq \frac{1}{I_y(x)}$, where $I_y(x) = E\left[\left(\frac{\partial}{\partial x}\ln p_y(y;x)\right)^2\right]$ Fisher info.

$\rightarrow$ unbiased.

Proof. $e(\vec{y}) = \hat{x}(\vec{y}) - x$. $\quad Var[e(\vec{y})] = E[e^2(\vec{y})] = \lambda_{\hat{x}}(x)$.

Define: $f(\vec{y}) = \frac{\partial}{\partial x} \ln p_y(\vec{y}; x)$.

① $E[f(\vec{y})] = \int p(\vec{y};x) \frac{\partial}{\partial x} \ln p(\vec{y};x)\, d\vec{y} = \frac{\partial}{\partial x} \int p(\vec{y};x)\, d\vec{y} = 0$.

② $Var[f(\vec{y})] = E\left[\left(\frac{\partial}{\partial x} \ln p(\vec{y};x)\right)^2\right] =$

Note that: $E\left[-\frac{\partial^2}{\partial x^2} \ln p(\vec{y};x)\right]$

$\int_{-\infty}^{\infty} p(\vec{y};x)\, d\vec{y} = 1$

$\cancel{\frac{\partial}{\partial x} \ln p(\vec{y};x)}$ $\int \frac{\partial}{\partial x} \ln p(\vec{y};x) \cdot p(\vec{y};x)\, d\vec{y} =$

take $\frac{\partial}{\partial x}$ again!

③ $E[e(\vec{y}) \cdot f(\vec{y})] = Cov(e(\vec{y}), f(\vec{y})) \le [Var(e(\vec{y})) \cdot Var(f(\vec{y}))]^{1/2}$.

$= E\left[(\hat{x}(\vec{y}) - x) \cdot \frac{\partial}{\partial x} \ln p(\vec{y};x)\right] = \int_{-\infty}^{+\infty} (\hat{x}(\vec{y}) - x) \cdot \frac{\partial}{\partial x} p(\vec{y};x)\, d\vec{y} =$

$= \frac{\partial}{\partial x} \int_{-\infty}^{\infty} \hat{x}(\vec{y})\, p(\vec{y};x)\, d\vec{y} - \underbrace{\int_{-\infty}^{\infty} x \frac{\partial}{\partial x} p(\vec{y};x)\, d\vec{y}}_{=0}$

$= 1 - 0 = 1$

$\Rightarrow Var[e(\vec{y})] \ge \frac{1}{Var[f(\vec{y})]} = \frac{1}{I_y(x)}$. ▨

Comments:

① For biased estimator: $\hat{x}(\vec{y})$: $E[e(\vec{y})] = b(x)$. Then:

$Cov(e(\vec{y}), f(\vec{y})) = b'(x) + 1$. $\cancel{Var[e(\vec{y})]}$

Then: $Var[e(\vec{y})] \ge \frac{(1 + b'(x))^2}{I_y(x)} \Rightarrow \boxed{E[e^2(\vec{y})] \ge \frac{(1 + b'(x))^2}{I_y(x)} + b(x)^2}$

② Vector case: $e(\vec{y}) = \hat{x}(\vec{y}) - \vec{x}$ Suppose unbiased.

$I \le Var[e(\vec{y})] \cdot I_y(x) \Rightarrow Var[e(\vec{y})] \ge I_y(x)^{-1}$.

where $I_y(x) = E\left[\left(\frac{\partial \ln p(\vec{y};\vec{x})}{\partial x}\right)^T \left(\frac{\partial \ln p(\vec{y};\vec{x})}{\partial x}\right)\right] = -E\left[\frac{\partial^2 \ln p(\vec{y};x)}{\partial x^2}\right]$

Comments on Cramer-Rao Bound: / Fisher info.

①. may not exist. $$p_y(y;x) = \begin{cases} 1, & x<y<x+1 \\ 0 & \text{o.w.} \end{cases}$$

not strictly positive for all x and y.

②. $I_y(x)$ is both nonnegative and additive.

$\vec{y} = [y_1, y_2, \cdots, y_M]^T$, y_i i.i.d. then

$$I_y(x) = \sum_{i=1}^{M} I_{y_i}(x).$$

~~③. Fisher info. measures on average, how "peaky" $\ln p(y;x)$ is as a func of $\hat{x}$.~~

④ Example: $y = x + w$, $w \sim N(0,\sigma^2)$. $\ln p(y;x) = -\frac{1}{2\sigma^2}(y-x)^2 - \frac{1}{2}\ln(2\pi\sigma^2)$.

$$I_y(x) = \frac{1}{\sigma^2}.$$

Efficiency and Consistency:

1. Cramer-Rao Bd is satisfied iff $e(\vec{y}) = k(x)\cdot f(\vec{y})$. $\forall \vec{y}$. (pf.)

Def. estimators satisfying CR bd: efficient estimator.

$$\hat{x}(\vec{y}) = x + k(x)\frac{\partial}{\partial x}\ln p(y;x).$$

① Unbiased $\Rightarrow \int p(\vec{y})\hat{x}(\vec{y})\,d\vec{y} = x \Rightarrow \quad k(x)\cdot\int \frac{\partial}{\partial x}\ln p(y;x)\cdot p(y;x)\,dy = 0$

② $\cancel{E[e^2(y)] = \frac{1}{I_y(x)} \Rightarrow \quad k^2(x)\, I_y(x) = \frac{1}{I_y(x)} \Rightarrow \quad k(x) = \frac{1}{I_y(x)}}$.

Efficient estimator: $\hat{x}(\vec{y}) = x + \frac{1}{I_y(x)}\frac{\partial}{\partial x}\ln p(y;x)$.

2. $\hat{x}_M = \hat{x}(y_1, y_2, \cdots, y_M)$. based on $(y_1, y_2, \cdots, y_M)$ observations.

We say $\hat{x}_M$ is a consistent estimator for x if $\hat{x}_M \xrightarrow{m.s} x$ ~~as $M\to\infty$~~.

~~Mean square convergence: $\lim_{M\to\infty} E[(\hat{x}_M - x)^2] = 0$~~

Example.

① $y = x + w, \quad w \sim N(0, \sigma^2)$.

$I_y(x) = \frac{1}{\sigma^2}, \qquad \lambda_{\hat{x}}(x) \geq \sigma^2$.

Let $\hat{x}(y) = y$, efficient estimator. not a func. of x, therefore valid.

$\lambda_{\hat{x}}(x) = \sigma^2$

② $y_i = x + w_i, \quad w_i \sim N(0, \sigma^2), \quad i = 1, \cdots, M$.

$I_y(x) = \frac{M}{\sigma^2}. \qquad \hat{x}(y) = \frac{1}{M}\sum_{i=1}^{M} y_i$, efficient estimator

$\lambda_{\hat{x}} = 1/I_y(x) = \sigma^2/M$.

Consistent because $\lambda_{\hat{x}} = \frac{\sigma^2}{M} \to 0$, as $M \to \infty$.

Maximum Likelihood Estimation:

1. $\hat{x}_{eff}(\vec{y}) = x + \frac{1}{I_y(x)}\frac{\partial}{\partial x}\ln p_y(y; x)$. because it is valid, indep. of x.

So we can choose any x in the RHS. Let $\hat{x}_{ML}(y) = \arg\max_x \ln p_y(y; x)$

then: $\hat{x}_{eff}(\vec{y}) = \hat{x}_{ML}(y)$.

✓ Conclude: when it exists the (unique) efficient estimator is equivalent to the ML estimator of the problem!

2. Note: However, it doesn't mean $\hat{x}_{ML}$ is always efficient!

① $\frac{\partial}{\partial x} p(y; x) = 0$ ② $\frac{\partial^2}{\partial x^2} p_y(y; x) < 0$. Search ~~[illegible]~~ over these local maxima & any boundary values

Asymptotic properties:

① $\hat{X}_M(y_1, y_2, \cdots, y_M)$ is asymptotically unbiased if

$$E[\hat{X}_M] \to X \quad \text{as } M \to \infty.$$

②. Weakly asymptotically efficient if:

$$\Lambda_{\hat{X}_M} - \frac{1}{I_{y_1 \ldots y_M}(x)} \to 0 \quad \text{as } M \to \infty.$$

③ Strongly asymptotically efficient if.

$$\lambda_{\hat{X}_M} \cdot I_{y_1 \ldots y_M}(x) \to 1, \quad \text{as } M \to \infty.$$

Example: ① $y = hx + w$, $w \sim N(0, \sigma_w^2)$. $\hat{X}_{ML}(y) = \frac{y}{h}$.

$E[\hat{X}_{ML}(y) - x] = 0$. $\lambda_{ML}(x) = \sigma_w^2 / h^2 = 1/I_y(x)$. so ML is efficient.

② $\vec{y} = H\vec{x} + \vec{w}$, $\vec{w} \sim N(0, \Lambda_w)$

$$p(\vec{y}; \vec{x}) = N(y; H\vec{x}, \Lambda_w) \sim \exp\left\{-\frac{1}{2}(\vec{y} - H\vec{x})^T \Lambda_w^{-1} (\vec{y} - H\vec{x})\right\}$$

$$\hat{X}_{ML}(\vec{y}) = \arg\min_x J(x) = \frac{1}{2}(\vec{y} - H\vec{x})^T \Lambda_w^{-1} (\vec{y} - H\vec{x}).$$

$$\hat{X}_{ML}(\vec{y}) = \underline{(H^T \Lambda_w^{-1} H)^{-1}} \, \underline{H^T \Lambda_w^{-1}} \vec{y}$$

$E[\hat{X}_{ML}(\vec{y})] = \vec{x}$, unbiased.

$\Lambda_{ML} = \underline{(H^T \Lambda_w^{-1} H)^{-1}}$. $\quad I_y(x) = -\frac{d^2}{dx^2} J(x) = H^T \Lambda_w^{-1} H$.

↳ $\hat{X}_{ML}$ efficient.