Outline

0. Review & course evaluation
1. Maximum likelihood estimation
2. Cramer Rao inequality
3. Bayes' estimation

0. Review

What we have done so far?

Find the best parameter estimation based on measurements & a priori information with different principles

Supposed mathematical model: $\vec{y} = H\vec{x}$; $\tilde{\vec{y}} = H\vec{x} + \vec{v}$; $\hat{\vec{y}} = H\hat{\vec{x}}$

a) Least squares approximation

$J = \min \frac{1}{2} (\tilde{\vec{y}} - \hat{\vec{y}})^T (\tilde{\vec{y}} - \hat{\vec{y}})$

$\hat{\vec{x}} = (H^T H)^{-1} H^T \tilde{\vec{y}}$

weighted LSA

$\hat{\vec{x}} = (H^T W H)^{-1} H^T W \tilde{\vec{y}}$ No mathematical criteria for W

b) Minimum variance estimation (Probability)

$\hat{\vec{x}} = M\tilde{\vec{y}} + \vec{n} \Rightarrow \vec{n} = 0 \quad MH = I$

$J = \min \frac{1}{2} Tr[E\{(\hat{\vec{x}} - \vec{x})(\hat{\vec{x}} - \vec{x})^T\}]$

$\hat{\vec{x}} = (H^T R^{-1} H)^{-1} H^T R^{-1} \tilde{\vec{y}}$

is identical to LSA, provided that $W = R^{-1}$

Minimum variance estimation with a priori

$\hat{\vec{x}} = M\tilde{\vec{y}} + N\hat{\vec{x}}_a + \vec{n} \Rightarrow \vec{n} = 0 \quad MH + N = I$

Solution $\hat{\vec{x}} = (H^T R^{-1} H + Q^{-1})^{-1} (H^T R^{-1} \tilde{\vec{y}} + Q^{-1} \hat{\vec{x}}_a)$

1. Maximum likelihood estimation

Not always biased

Recall: a Gaussian distribution $Y \sim N(\mu, \sigma^2)$

the probability density function

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

$$E[\tilde{y}_i] = \mu$$

$$\Rightarrow \quad Var[\tilde{y}_i] = \sigma^2$$

Question: Given $\{\tilde{y}_1, \tilde{y}_2 \cdots \tilde{y}_m\}$

how to estimate μ and σ^2 from observed data?

$$\hat{\mu} = \frac{1}{m}\sum_{i=1}^{m} \tilde{y}_i ; \qquad \text{Let } \vec{x} = [\mu, \sigma^2]$$

Basic idea:

A good estimate of $\vec{x}$ would be the value of $\hat{\vec{x}}$ that maximize the probability (likelihood) of getting the data we observed.

Define likelihood function

$$L(\tilde{\vec{y}}; \vec{x}) = f(\tilde{y}_1; \vec{x}) f(\tilde{y}_2; \vec{x}) \cdots f(\tilde{y}_m; \vec{x})$$

$$= \prod_{i=1}^{m} f(\tilde{y}_i; \vec{x})$$

In textbook $f(\tilde{y}_i; \vec{x})$ is written as $p(\tilde{y}_i | \vec{x})$

For the Gaussian example:

$$L(\tilde{\vec{y}}; \vec{x}) = \prod_{i=1}^{m} f(\tilde{y}_i; \vec{x}) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{m}{2}} \exp\left(-\frac{\sum_{i=1}^{m}(\tilde{y}_i - \mu)}{2\sigma^2}\right)$$

The Gaussian distribution is a monotonic exponential function for μ and σ^2.

$$\ln[L(\tilde{\vec{y}}; \vec{x})] = -\frac{m}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}(\tilde{y}_i - \mu)$$

Max $L(\tilde{\vec{y}}; \vec{x})$ is equivalent to Max $\ln[L(\tilde{\vec{y}}; \vec{x})]$

$$J = \max \ln[L(\tilde{\vec{y}}; \hat{\vec{x}})]$$

Necessary condition: $\frac{\partial}{\partial x}\ln[L(\tilde{\vec{y}}; \hat{\vec{x}})] = 0$

Sufficient condition: $\frac{\partial^2}{\partial x^2}\ln[L(\tilde{\vec{y}}; \hat{\vec{x}})] < 0$

$$\frac{\partial}{\partial \hat{\mu}}\ln[L(\tilde{\vec{y}}; \hat{\vec{x}})] = \frac{1}{\hat{\sigma}^2}\sum_{i=1}^{m}(\tilde{y}_i - \hat{\mu}) = 0$$

$$\Rightarrow \hat{\mu} = \frac{1}{m}\sum_{i=1}^{m}\tilde{y}_i$$

$$\frac{\partial}{\partial \hat{\sigma}^2}\ln[L(\tilde{\vec{y}}; \hat{\vec{x}})] = -\frac{m}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4}\sum_{i=1}^{m}(\tilde{y}_i - \hat{\mu})^2 = 0$$

$$\Rightarrow \hat{\sigma}^2 = \frac{1}{m}\sum_{i=1}^{m}(\tilde{y}_i - \hat{\mu})^2$$

$$E[\hat{u}] = E\left[\frac{1}{m}\sum_{i=1}^{m}\tilde{y}_i\right] = \frac{1}{m}E\left[\sum_{i=1}^{m}\tilde{y}_i\right] = \frac{1}{m}\cdot m\mu = \mu$$

$$\mathrm{Var}[\hat{u}] = \mathrm{Var}\left[\frac{1}{m}\sum_{i=1}^{m}\tilde{y}_i\right]$$

$$= \frac{1}{m^2}\,\mathrm{Var}\left[\sum_{i=1}^{m}\tilde{y}_i\right]$$

$$= \frac{1}{m^2}(m\sigma^2) = \frac{\sigma^2}{m}$$

$$\mathrm{Var}[\hat{u}] = E[\hat{u}^2] - (E[u])^2$$

$$E[\hat{u}^2] = \frac{\sigma^2}{m} + \mu$$

$$\hat{\sigma}^2 = \frac{1}{m}\sum_{i=1}^{m}(\tilde{y}_i - \hat{u})^2$$ ✡ sample variance

$$E[\hat{\sigma}^2] = E\left[\frac{1}{m}\sum_{i=1}^{m}(\tilde{y}_i - \hat{u})^2\right]$$

$$= \frac{1}{m}E\left[\sum_{i=1}^{m}(\tilde{y}_i^2 - 2\hat{u}\tilde{y}_i + \hat{u}^2)\right]$$

$$= \frac{1}{m}E\left[\sum_{i=1}^{m}\tilde{y}_i^2 - 2\hat{u}\sum_{i=1}^{m}\tilde{y}_i + m\hat{u}^2\right]$$

$$= \frac{1}{m}E\left[\sum_{i=1}^{m}\tilde{y}_i^2 - 2m\hat{u}^2 + m\hat{u}^2\right]$$

$$= \frac{1}{m}E\left[\sum_{i=1}^{m}\tilde{y}_i^2 - m\hat{u}^2\right]$$

$$= \frac{1}{m}(m\cdot E[\tilde{y}_i^2]) - \frac{1}{m}(m\cdot E[\hat{u}^2])$$

$$= \mu^2 + \sigma^2 - \left(\frac{\sigma^2}{m} + \mu^2\right)$$

$$= \frac{m-1}{m}\sigma^2 \Rightarrow \hat{\sigma}^2 \text{ is biased}$$

$$\hat{\sigma} = \frac{1}{m-1}\sum_{i=1}^{m}(\tilde{y}_i - \hat{u})^2 \quad \text{is unbiased}$$

Example:

Suppose we have data from $Y \sim B(1,p)$ (discrete random variable)

$$f(\tilde{y}_i;\vec{x}) = x^{\tilde{y}_i}(1-x)^{1-\tilde{y}_i} \quad \text{where } x=p,\ \tilde{y}_i = 0,1$$

$$= \begin{cases} x & \text{if } \tilde{y}_i = 1 \\ 1-x & \text{if } \tilde{y}_i = 0 \end{cases}$$

e.g. toss a coin n times.

$$L(\tilde{\vec{y}};\vec{x}) = \prod_{i=1}^{m} x^{\tilde{y}_i}(1-x)^{(1-\tilde{y}_i)}$$

$$\ln L(\tilde{\vec{y}};x) = \sum_{i=1}^{m} \tilde{y}_i \ln x + \sum_{i=1}^{m} (1-\tilde{y}_i)\ln(1-x)$$

$$J = \text{Max } \ln L(\tilde{\vec{y}};\hat{x})$$

Necessary condition: $\frac{\partial}{\partial x} L(\tilde{\vec{y}};\hat{x}) = 0$

$$\Rightarrow \sum_{i=1}^{m} \tilde{y}_i \frac{1}{\hat{x}} + \sum_{i=1}^{m} (1-\tilde{y}_i)\frac{1}{1-\hat{x}} = 0$$

$$(1-\hat{x})\sum_{i=1}^{m} \tilde{y}_i + \hat{x}\sum_{i=1}^{m}(1-\tilde{y}_i) = 0$$

$$\sum_{i=1}^{m} \tilde{y}_i - \hat{x}\sum_{i=1}^{m} \tilde{y}_i + m\hat{x} - \hat{x}\sum_{i=1}^{m} \tilde{y}_i = 0$$

$$\Rightarrow \hat{x} = \frac{\sum_{i=1}^{m} \tilde{y}_i}{m}$$ This is the estimation for p

$$E[\tilde{y}_i] = p$$

$$E[\hat{x}] = E\left[\frac{1}{m}\sum_{i=1}^{m} \tilde{y}_i\right] = p$$ This is an unbiased estimation

Information matrix (scalar in this case)

$$L(\tilde{\vec{y}};x) = \prod_{i=1}^{m} f(\tilde{y}_i;x)$$

$$\ln L(\tilde{\vec{y}};x) = \sum_{i=1}^{m} \ln f(\tilde{y}_i;x)$$

$$\frac{\partial}{\partial x}\ln L(\tilde{\vec{y}};x) = \sum_{i=1}^{m} \frac{\partial}{\partial x}\ln f(\tilde{y}_i;x)$$

$$\frac{\partial^2}{\partial x^2}\ln L(\tilde{\vec{y}};x) = \sum_{i=1}^{m} \frac{\partial^2}{\partial x}\ln f(\tilde{y}_i;x)$$

$$F = \frac{m}{x(1-x)}$$

$$f(\tilde{y}_i;\vec{x}) = x^{\tilde{y}_i}(1-x)^{1-\tilde{y}_i}$$

$$\ln f(\tilde{y}_i;x) = \tilde{y}_i \ln x + (1-\tilde{y}_i)\ln(1-x)$$

$$\frac{\partial}{\partial x}\ln f(\tilde{y}_i;x) = \frac{\tilde{y}_i}{x} - \frac{1-\tilde{y}_i}{1-x}$$

$$\frac{\partial^2}{\partial x^2}\ln f(\tilde{y}_i;x) = -\frac{\tilde{y}_i}{x^2} - \frac{1-\tilde{y}_i}{(1-x)^2}$$

$$E\left[\frac{\partial^2}{\partial x^2}\ln f(\tilde{y}_i;x)\right] = -\frac{1}{x} - \frac{1}{1-x}$$

$$F = -E\left[\frac{\partial^2}{\partial x^2}\ln f(\tilde{y}_i;x)\right] = \frac{1}{x} + \frac{1}{1-x} = \frac{1}{x(1-x)}$$

$$p = \text{Var}[\hat{x}] = \text{Var}\left[\frac{1}{m}\sum_{i=1}^{m} \tilde{y}_i\right] = \frac{1}{m^2}\cdot m\,\text{Var}(\tilde{y}_i) = \frac{1}{m}x(1-x)$$

$$\Rightarrow p = F^{-1}$$

2. Cramer - Rao inequality

$\hat{u} = \frac{1}{m}\sum_{i=1}^{m} \tilde{y}_i$

Minimum variance estimation

$\tilde{y}_i = u + v_i$ where $v \sim N(0, \sigma^2)$

$\vec{\tilde{y}} = Hx + \vec{v}$ where $x = u$, $H = [1, \cdots 1]^T \in \mathbb{R}^{m \times 1}$

$R = \text{cov}[\vec{v}\vec{v}^T] = \begin{bmatrix} \sigma^2 & & \\ & \ddots & \\ & & \sigma^2 \end{bmatrix}$

$\hat{x} = \hat{u} = (H^T R^{-1} H)^{-1} H^T R^{-1} \vec{\tilde{y}}$

$= \frac{1}{m}\sum_{i=1}^{m} \tilde{y}_i$

$f(\vec{\tilde{y}}; u) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{m}{2}} \exp\left(-\frac{\sum_{i=1}^{m}(\tilde{y}_i - u)}{2\sigma^2}\right)$

$\ln f(\vec{\tilde{y}}; u) = -\frac{m}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}(\tilde{y}_i - u)^2$

$\frac{\partial}{\partial u}\ln f(\vec{\tilde{y}}; u) = \frac{1}{\sigma^2}\sum_{i=1}^{m}(\tilde{y}_i - u)$

$\frac{\partial^2}{\partial u^2}\ln f(\vec{\tilde{y}}; u) = -\frac{m}{\sigma^2}$

Let $F = \frac{m}{\sigma^2}$ Information matrix (Fisher)

$P = E\{(\hat{\vec{x}} - \vec{x})(\hat{\vec{x}} - \vec{x})^T\}$ Covariance of $\hat{\vec{x}}$

$= E\{(\hat{u} - u)^2\}$

$= E\{\hat{u}^2 - 2u\hat{u} + u^2\}$

$= E[\hat{u}^2] - 2uE[\hat{u}] + u^2$

$= \frac{\sigma^2}{m} + u^2 - 2u^2 + u^2$

$= \frac{\sigma^2}{m}$

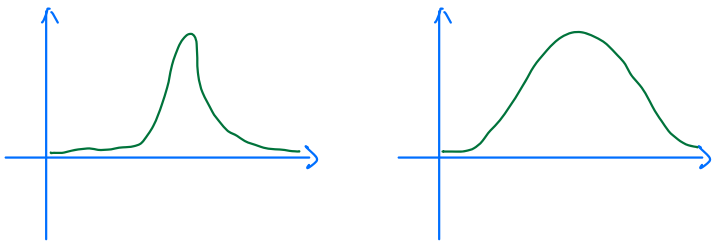
$P = F^{-1}$ $\qquad P \geq F^{-1}$

Define Fisher matrix as

$$F = E\left\{ \left[\frac{\partial}{\partial x} \ln f(\tilde{y}; \vec{x}) \right] \left[\frac{\partial}{\partial x} \ln f(\tilde{y}; \vec{x}) \right]^T \right\}$$

can be computed by

$$F = -E\left\{ \frac{\partial^2}{\partial \vec{x} \, \partial \vec{x}^T} \ln f(\tilde{y}; \vec{x}) \right\}$$



$$F = \frac{m}{\sigma^2} \qquad F^{-1} = \frac{\sigma^2}{m}$$

$$\frac{\partial}{\partial x} \ln f(\tilde{y}; \vec{x}) = \frac{1}{f} \cdot \frac{\partial}{\partial x} f$$

$$F = E\left\{ \left[\frac{\partial}{\partial x} \ln f(\tilde{y}; \vec{x}) \right] \left[\frac{\partial}{\partial x} \ln f(\tilde{y}; \vec{x}) \right]^T \right\}$$

$$= \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial x} \ln f(\tilde{y}; \vec{x}) \right] \left[\frac{\partial}{\partial x} \ln f(\tilde{y}; \vec{x}) \right]^T f(\tilde{y}; \vec{x}) \, d\tilde{y}$$

$$= \int_{-\infty}^{\infty} \frac{1}{f} \frac{\partial f}{\partial x} \left(\frac{\partial f}{\partial x} \right)^T \frac{1}{f} f \, d\tilde{y}$$

$$= \int_{-\infty}^{\infty} \frac{1}{f} \frac{\partial f}{\partial x} \left(\frac{\partial f}{\partial x} \right)^T d\tilde{y}$$

$F^{-1} \downarrow \rightarrow F \uparrow \rightarrow f \downarrow$ & $\frac{\partial f}{\partial x} \uparrow$.

CR inequality.

$$P \equiv E\{(\hat{\vec{x}} - \vec{x})(\hat{\vec{x}} - \vec{x})^T\}$$

$$F \equiv E\{[\frac{\partial}{\partial \vec{x}} \ln[f(\tilde{\vec{y}}; \vec{x})]][\frac{\partial}{\partial \vec{x}} \ln[f(\tilde{\vec{y}}; \vec{x})]]^T\}$$

The CR inequality is $P \geq F^{-1}$

Start the proof by

① $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(\tilde{\vec{y}}; \vec{x})\, d\tilde{y}_1\, d\tilde{y}_2 \cdots d\tilde{y}_m = 1$

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

② Take partial derivative of ① w.r.t. $\vec{x}$

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

③ $\hat{\vec{x}}$ is assumed to be unbiased

$$E[(\hat{\vec{x}} - \vec{x})] = \int_{-\infty}^{\infty} \underbrace{(\hat{\vec{x}} - \vec{x})}_{\text{vector}} \underbrace{f(\tilde{\vec{y}}; \vec{x})}_{\text{scalar}}\, d\tilde{\vec{y}} = \vec{0}$$

④ Take partial derivative of ③ w.r.t. $\vec{x}$

$$\frac{\partial}{\partial \vec{x}} \int_{-\infty}^{\infty} (\hat{\vec{x}} - \vec{x}) f(\tilde{\vec{y}}; \vec{x})\, d\tilde{\vec{y}}$$

$$= \int_{-\infty}^{\infty} (\hat{\vec{x}} - \vec{x}) \left[\frac{\partial}{\partial \vec{x}} f(\tilde{\vec{y}}; \vec{x})\right]^T d\tilde{\vec{y}} - I\int_{-\infty}^{\infty} f(\hat{\vec{y}}; \vec{x})\, d\hat{\vec{y}} = 0$$

$$\int_{-\infty}^{\infty} (\hat{\vec{x}} - \vec{x}) \left[\frac{\partial}{\partial \vec{x}} f(\hat{\vec{y}}; \vec{x})\right]^T d\tilde{\vec{y}} = I$$

⑤ $\frac{\partial}{\partial \vec{x}} \ln f(\hat{\vec{y}}; \vec{x}) = \frac{1}{f} \cdot \frac{\partial}{\partial \vec{x}} f$

$$\frac{\partial}{\partial \vec{x}} f(\hat{\vec{y}}; \vec{x}) = \left[\frac{\partial}{\partial \vec{x}} \ln f(\hat{\vec{y}}; \vec{x})\right] \cdot f(\tilde{\vec{y}}; \vec{x})$$

⑥ plug ⑤ into ④

$$\int_{-\infty}^{\infty} (\hat{\vec{x}} - \vec{x}) \left[\frac{\partial}{\partial \vec{x}} \ln f(\tilde{\vec{y}}; \vec{x})\right]^T \underbrace{f(\hat{\vec{y}}; \vec{x})}_{= f^{1/2} \cdot f^{1/2}} d\tilde{\vec{y}} = I$$

scalar

Let $\vec{a} \equiv f(\tilde{\vec{y}}; \vec{x})^{1/2} (\hat{\vec{x}} - \vec{x})$

$$\vec{b} \equiv f(\tilde{\vec{y}}; \vec{x})^{1/2} \left[\frac{\partial}{\partial \vec{x}} \ln f(\tilde{\vec{y}}; \vec{x})\right]$$

$$\int_{-\infty}^{\infty} (\vec{a}\,\vec{b}^T)\, d\tilde{\vec{y}} = I$$

$$P \equiv E\{(\hat{\vec{x}} - \vec{x})(\hat{\vec{x}} - \vec{x})^T\}$$

$$= \int_{-\infty}^{\infty} (\hat{\vec{x}} - \vec{x})(\hat{\vec{x}} - \vec{x})^T f(\tilde{\vec{y}}; \vec{x})\, d\tilde{\vec{y}}$$

$$= \int_{-\infty}^{\infty} \vec{a}\,\vec{a}^T\, d\tilde{\vec{y}}$$

$$F \equiv E\left\{\left[\frac{\partial}{\partial \vec{x}} \ln[f(\tilde{\vec{y}}; \vec{x})]\right]\left[\frac{\partial}{\partial \vec{x}} \ln[f(\tilde{\vec{y}}; \vec{x})]\right]^T\right\}$$

$$= \int_{-\infty}^{\infty} \vec{b}\,\vec{b}^T\, d\tilde{\vec{y}}$$

From ⑥, we have

$$I = \int_{-\infty}^{\infty} \vec{a}\,\vec{b}^T\, d\tilde{\vec{y}}$$

For any random vectors $\vec{\alpha}, \vec{\beta}$

$$\vec{\alpha}^T I \vec{\beta} = \int_{-\infty}^{\infty} \vec{\alpha}^T \vec{a}\,\vec{b}^T \vec{\beta}\, d\tilde{\vec{y}}$$

Schwartz inequality

$$\left[\int_{-\infty}^{\infty} g(\tilde{\vec{y}}; \vec{x})\, h(\tilde{\vec{y}}; \vec{x})\, d\tilde{\vec{y}}\right]^2 \le \int_{-\infty}^{\infty} g^2(\tilde{\vec{y}}; \vec{x})\, d\tilde{\vec{y}} \int_{-\infty}^{\infty} h^2(\tilde{\vec{y}}; \vec{x})\, d\tilde{\vec{y}}$$

Let $g(\tilde{\vec{y}}; \vec{x}) = \vec{\alpha}^T \vec{a}^T \quad h(\tilde{\vec{y}}; \vec{x}) = \vec{b}^T \beta$

$$\left[\int_{-\infty}^{\infty} \vec{\alpha}^T (\vec{a}\,\vec{b}^T) \beta\, d\tilde{\vec{y}}\right]^2 \le \int_{-\infty}^{\infty} \vec{\alpha}^T \vec{a}\,\vec{a}\,\vec{\alpha}\, d\tilde{\vec{y}} \int_{-\infty}^{\infty} \vec{\beta}^T \vec{b}\,\vec{b}^T \beta\, d\tilde{\vec{y}}$$

$$[\alpha^T \beta]^2 \le (\vec{\alpha}^T P \alpha)(\beta^T F \beta)$$

Let $\beta = F^{-1}\alpha; \quad \beta^T = \alpha^T (F^{-1})^T$

$$(\alpha^T P \alpha)(\alpha^T (F^{-1})^T F F^{-1} \alpha) \ge [\alpha^T F^{-1} \alpha]^2$$

$$\alpha^T P \alpha\, \alpha^T (F^{-1})^T \alpha - \alpha^T F^{-1} \alpha\, \alpha^T F^{-1} \alpha \ge 0$$

$$[\alpha^T P \alpha - \alpha^T F^{-1} \alpha]\, \alpha^T F^{-1} \alpha \ge 0$$

$$\alpha^T P \alpha - \alpha^T F^{-1} \alpha \ge 0$$

$$\alpha^T (P - F^{-1}) \alpha \ge 0$$

$$P - F^{-1} \ge 0$$

$$P \ge F^{-1}$$

When are we gonna use this

1. Evaluate the estimators
2. feasibility studies.

Example (CRLB cannot be attained)

$\{\tilde{y}_i\} \sim N(\mu, \sigma^2)$, given μ, estimate σ^2

$$F = -E\left\{\frac{\partial^2}{\partial \vec{x}\,\partial \vec{x}^T} \ln[f(\tilde{y};\vec{x})]\right\}$$

$$f(\tilde{y}_i;\vec{x}) = \frac{1}{2\pi\sigma^2}\exp\left[-\frac{(\tilde{y}_i-\mu)^2}{2\sigma^2}\right]$$

$$\ln f(\tilde{y}_i;\vec{x}) = -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{(\tilde{y}_i-\mu)^2}{2\sigma^2}$$

$$\frac{\partial}{\partial\sigma^2}\ln f(\tilde{y}_i;\vec{x}) = -\frac{1}{2}\frac{1}{\sigma^2} + \frac{(\tilde{y}_i-\mu)^2}{2(\sigma^2)^2}$$

$$\frac{\partial^2}{\partial(\sigma^2)^2}\ln f = \frac{1}{2(\sigma^2)^2} - \frac{2(\tilde{y}_i-\mu)^2}{2(\sigma^2)^3}$$

$$F = -E\left[\frac{1}{2\sigma^4} - \frac{(\tilde{y}_i-\mu)^2}{\sigma^6}\right]$$

$$= \frac{1}{2\sigma^4}$$

The lower bound is $F^{-1} = \frac{2\sigma^4}{m}$

The unbiased estimation is $\hat{\sigma} = \frac{1}{m-1}\sum_{i=1}^{m}(\tilde{y}_i - \hat{\mu})^2$

$$\mathrm{Var}(\hat{\sigma}) = \mathrm{Var}\left(\frac{1}{m-1}\sum_{i=1}^{m}(\tilde{y}_i-\hat{\mu})^2\right)$$

$$= \frac{2\sigma^4}{m-1} = P > F^{-1}$$

3. Bayesian Estimation

In textbook 2.5 & 2.6, $P(\tilde{\vec{y}}|\vec{x})$ means likelihood function like $f(\tilde{\vec{y}};\vec{x})$ in the notes

Basic idea: consider the parameters (to be estimated) are random variables with a priori distribution

Bayesian estimation will combine the a priori information with measurements

Bayes' theorem

$$\boxed{P(\vec{x}|\tilde{\vec{y}})} = \frac{P(\tilde{\vec{y}}|\vec{x})P(\vec{x})}{P(\tilde{\vec{y}})}$$

a posteriori distribution of $\vec{x}$

3.1 Maximum a posteriori estimation (MAP)

$$J = \max \ln P(\vec{x}|\tilde{\vec{y}})$$

Note $P(\tilde{\vec{y}})$ does not depend on x

Maximize J is equalent to

$$J_{MAP} = \max \ln P(\tilde{\vec{y}}|\vec{x})P(\vec{x})$$

$$= \max \ln P(\tilde{\vec{y}}|\vec{x}) + \ln P(\vec{x})$$