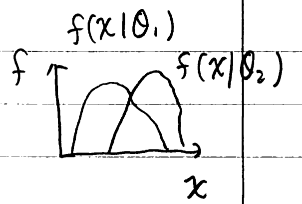
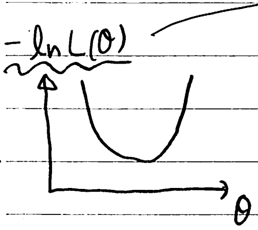


$\Updownarrow$



$L(\theta) = f(x|\theta)$ ← fixed

$f(x|\theta)$ ← x: random variable "observable", θ: parameters

• Once we wrote $\int d\theta\, g(\theta)$, this is Baysian.

good custom.

Note $\int d\theta\, L(\theta) \neq 1$.

Transformation

• What's the PDF of $y(x)$ under given PDF $f(x)$?

$y(x)$ ‖ $g(y)$

$$\Rightarrow \quad g(y) = \frac{1}{|\partial y/\partial x|} f(x(y))$$

$\Updownarrow$

• Likelihood is invariant to transf. of θ.

$$L(\phi(\theta)) = L(\theta) \qquad \text{for } \phi = \phi(\theta)$$

Sho Iwamoto

Don't forget $E[\ ]$ depends on θ. (params.)

unbiased estimator $\underset{\text{def}}{\Leftrightarrow}$ $\mathrm{bias}(\theta)=0$ for $^\forall\theta$

Parameter Estimation

$$f(x,\theta)$$

- Introduce $\hat{\theta}(x)$ "Estimator" a func. of R.V. x (i.e. R.V.!)
- Introduce $E[\circledast|\theta]$ "expectation"

$$= \int dx\,(\circledast)\,f(x|\theta)$$

$$\mathrm{bias}(\theta) := E[\hat{\theta}|\theta]-\theta$$ bias of an estimator $\hat{\theta}$ under θ.

$$\mathrm{Var}[\hat{\theta}|\theta] = E\left[(\hat{\theta}-E[\hat{\theta}])^2\,\middle|\,\theta\right]$$

$$\mathrm{MSE}[\hat{\theta},|\theta] = E\left[(\hat{\theta}-\theta)^2\,\middle|\,\theta\right]$$ Mean squared error.

$$\mathrm{Cov}_{ij}[\hat{\theta}|\theta] = \mathrm{Cov}[\hat{\theta}_i,\hat{\theta}_j|\theta] = E\left[(\hat{\theta}_i-E[\theta_i])(\hat{\theta}_j-E[\theta_j])\,\middle|\,\theta\right]$$

☆ Cramér-Rao bound

$$\mathrm{Var}[\hat{\theta}|\theta] \geq \frac{\left(1+\frac{\partial b}{\partial\theta}(\theta)\right)^2}{I(\theta)}$$ for 1-D case

$$I_{ij}(\theta) = E\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\left(-\ln L(\theta)\right)\,\middle|\,\theta\right]$$

$$= \int dx\,f(x|\theta)\,\frac{\partial^2}{\partial\theta_i\partial\theta_j}\left(-\ln L(\theta)\right)$$

Maximum Likelihood estimator (MLE)
... "asymptotically" unbiased

$$\hat{\theta}_{MLE} = \underset{\theta}{\mathrm{argmax}}\, L(\theta)$$

Example lifetime $e^{-t/\tau}$

$$f(t|\tau) = e^{-t/\tau}/\tau$$

$$-\ln L(\tau) = -\ln\left[\prod_{data} e^{-t_i/\tau}/\tau\right]$$

$$= \sum_{data}\left[\ln\tau + \frac{t_i}{\tau}\right]$$

$\xrightarrow{\frac{\partial}{\partial\tau}} \sum\left(\frac{1}{\tau} - \frac{t_i}{\tau^2}\right)$

0 for $N\tau = \sum t_i$

$$\Rightarrow \hat{\tau}_{MLE} = \frac{1}{N}\sum t_i$$

$\frac{\partial L}{\partial\tau} = 0$

$\Leftrightarrow \frac{\partial}{\partial\tau}\ln L = 0$

Example Gaussian

$$-\ln L(\mu) = -\ln\prod\left[\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right]$$

$$= N\ln\sqrt{2\pi}\sigma + \frac{1}{2\sigma^2}\sum(x_i-\mu)^2$$

$$\xrightarrow{\frac{\partial}{\partial\mu}} \frac{2}{2\sigma^2}\sum(\mu - x_i) \rightarrow \hat{\mu}_{MLE} = \frac{1}{N}\sum x_i$$

$$\mathrm{bias}(\hat{\mu}|\mu) = \cdots = 0$$

for σ^2 ?

$$\hat{\sigma}^2_{MLE} = \sum \frac{(x_i - \mu)^2}{N}$$

If we know μ, $\hat{\sigma}^2_{MLE}$ is unbiased.

If we don't know,

$$\hat{\sigma}^2_{MLE} = \sum \frac{(x_i - \hat{\mu}_{MLE})^2}{N}$$

$$\Rightarrow E[\hat{\sigma}^2 \mid \sigma^2] = \frac{N-1}{N}\sigma^2$$ "biased"

but "asymptotically unbiased".

James-Stein estimator $\hat{\mu}_{JS} = \hat{\mu}_{MLE}\left(1 - \frac{(d-2)}{\|\mu_{MLE}\|}\right)$

biased, but $MSE[\hat{\mu}_{JS} \mid \mu] \leq MSE[\hat{\mu}_{MLE} \mid \mu]$ $\forall \mu$

for Multi-dim. Gauss.

$MSE \simeq Var + Bias^2$

"bias/variance trade-off".

... What is good estimator?

Minimum bias?

Minimum variance?

Sho Iwamoto

No.

Date

PDF: $\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\lambda)^2}{2\sigma^2}} \prod_{i=1}^{n} \frac{e^{-t_i/\tau+\lambda}}{\tau+\lambda}$

$$-\ln L(\tau,\lambda) = -\left[C - \frac{(y-\lambda)^2}{2\sigma^2} + \sum_{i=1}^{n}\left[-\frac{t_i}{\tau+\lambda} - \ln(\tau+\lambda)\right]\right]$$

$$= \frac{(y-\lambda)^2}{2\sigma^2} + \frac{\Sigma t_i}{\tau+\lambda} + n\ln(\tau+\lambda) + C$$

(a) MLE

$$\frac{\partial}{\partial\tau}(-\ln L) = -\frac{\Sigma t_i}{(\tau+\lambda)^2} + \frac{n}{|\tau+\lambda|}$$

$$\frac{\partial}{\partial\lambda}(-\ln L) = \frac{\lambda-y}{\sigma^2} - \frac{\Sigma t_i}{(\tau+\lambda)^2} + \frac{n}{|\tau+\lambda|}$$

Both of them $= 0 \iff \begin{cases} |\tau+\lambda| = \frac{1}{n}\Sigma t_i \\ \lambda = y \end{cases}$

$$\therefore \hat{\lambda}(t_i, y) = y$$

$$\hat{\tau}(t_i; y) = \pm\frac{1}{n}\Sigma t_i - \hat{\lambda}$$

$$= \pm\frac{1}{n}\Sigma t_i - y$$

hereafter $\hat{\tau} > 0$ is assured.

(b) Variance

- $\langle y \rangle = \lambda$, $\langle t_i \rangle = \tau+\lambda$ $Var(y) = \sigma^2$, $Var(t_i) = (\lambda+\tau)^2$

$\therefore Var(\hat{\lambda}) = \sigma^2$

$$V(x) = \langle [x - \langle x \rangle]^2 \rangle$$

No.

Date . .

$$\mathrm{Var}(\hat{\tau}) = \mathrm{Var}\left(\frac{1}{n}\sum_i t_i - y\right)$$

$$= \left\langle \left(\frac{1}{n}\sum t_i - y - \left\langle \frac{1}{n}\sum_i t_i - y \right\rangle\right)^2 \right\rangle$$

$$= \left\langle \left[\frac{1}{n}\sum\left(t_i - \langle t_i \rangle\right) - \left(y - \langle y \rangle\right)\right]^2 \right\rangle$$

$$\left(= \frac{1}{n^2}\sum \mathrm{Var}(t_i) + \mathrm{Var}(y) = \frac{(\lambda+\tau)^2}{n} + \sigma^2 \right)$$

in Mathematica Notebook I did ...

$$= \left\langle \left[\frac{1}{n}\sum t_i - (\tau+\lambda) - y + \lambda\right]^2 \right\rangle$$

$$= \left\langle \frac{1}{n^2}\left(\sum t_i\right)^2 + y^2 + \tau^2 - 2\tau\frac{1}{n}\sum t_i \right\rangle$$

Formula $V(aX+bY) = a^2 V(x) + b^2 V(y) + 2ab\,\mathrm{Cov}(x,y)$

Sho Iwamoto

No.

Date

Decision Theory

$\theta \in \Theta$: parameter space $\qquad f(x|\theta)$

$x \in X$: random var space

$a \in A$: "action" $\qquad \delta : X \to A$

example: MSE

$\mathrm{Loss}(\theta, \delta(x))$ and $\mathrm{Risk}(\theta, \delta) = E_{f(x|\theta)}\left[\mathrm{Loss}(\theta, \delta(x))\right]$

- A rule δ dominates δ' if $\mathrm{Risk}(\theta, \delta) \le \mathrm{Risk}(\theta, \delta')$ $\forall \theta$
- admissible iff no other rules dominates it.

Bayes' Risk

$$r(\pi, \delta) = E_{\pi(\theta)}[R(\theta, \delta)]$$
$$= \int d\theta \int dx\, \pi(\theta)\, f(x|\theta)\, L(\theta, \delta(x))$$

Expected Loss

$$\rho(\pi, \delta) = E_{\pi(\theta|x)}\, \mathrm{Loss}(\theta, \delta)$$

- Every admissible rule is a generalized Bayes rule (for some prior $\pi(\theta)$)
- Every Bayes procedures admissible.