Lec. 28

Announcement.

Make-up meeting on this Friday.

Review

MLE

$\{f_{\underline{Y};\underline{\theta}}(\underline{y}) \,;\, \underline{\theta} \in A\}$

$\{\underline{Y} = \underline{y}\}$

$$\hat{\underline{\theta}}_{ML}(\underline{y}) \triangleq \arg\max_{\underline{\theta}} f_{\underline{Y};\underline{\theta}}(\underline{y})$$

likelihood function

$$\triangleq \arg\max_{\underline{\theta}} \ln f_{\underline{Y};\underline{\theta}}(\underline{y})$$

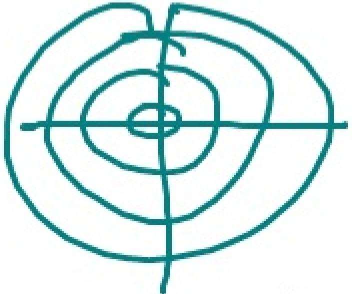
log-likelihood function

$$\frac{\partial}{\partial \underline{\theta}} \ln f_{\underline{Y};\underline{\theta}}(\underline{y})$$

score function

P.R

$$\hat{\underline{\theta}}_{ML}(\underline{y}) \triangleq \arg\max_{\underline{\theta}} f_{\underline{Y};\underline{\theta}}(\underline{y})$$



(M) EM algorithm $\hat{\underline{\theta}}^{(i+1)}$ is no worse than $\hat{\underline{\theta}}^{(i)}$, $\forall i$

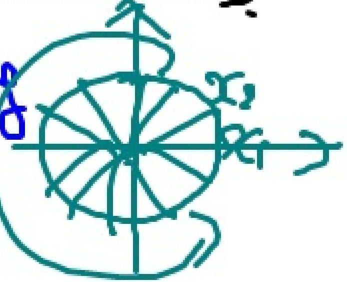
in the sense that $f_{\underline{Y};\hat{\underline{\theta}}^{(i+1)}}(\underline{y}) \geq f_{\underline{Y};\hat{\underline{\theta}}^{(i)}}(\underline{y})$, $\forall i$

Real Analysis

Let $(x_n)_{n\geq 1}$ be a sequence of real numbers.

(i) $x_{n+1} \geq x_n, \forall n$ non-decreasing

(ii) $x_n \leq M < \infty, \forall n$. bounded



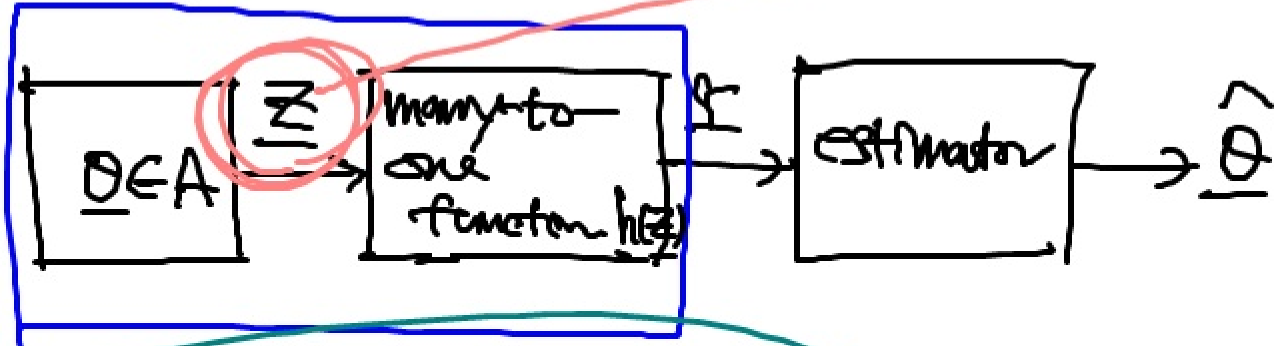
Then, $(x_n)_{n\geq 1}$ converges.

p.3

$\{f_{\underline{Y};\underline{\theta}}(\underline{y}) : \underline{\theta}\in A\}$ Given

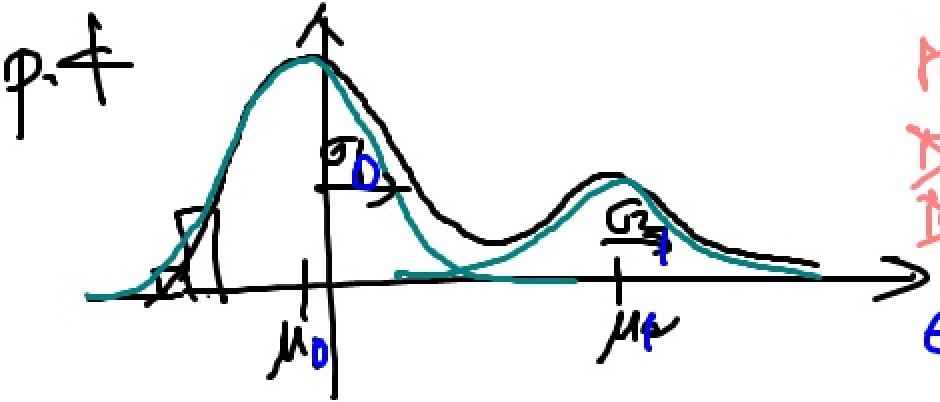
Incomplete observation

Complete observation



$\underline{y} = \underline{h}(\underline{z}),\ \{f_{\underline{Z};\underline{\theta}}(\underline{z})\ ;\ \underline{\theta}\in A\}$

Virtual decomposition &



raw data → histogram.

~~raw~~

Gaussian mixture.

$$\boxed{\begin{matrix}\alpha = p \\ \mu_0, \mu_1 \\ \cancel{\sigma_1}, \cancel{\sigma_2}\end{matrix}} \xrightarrow{\underline{Y}} \boxed{\text{MLE}} \to \hat{\theta}_{MLE}(\underline{Y})$$

$$f_{Y_1, Y_2}(y_1, y_2) = f_{Y_1}(y_1) f_{Y_2}(y_2)$$

Simplified problem

$\sigma_1 = 1, \sigma_2 = 1$

$\mu_0 = 0, \mu_1 = 1$

p ← unknown

$$\hat{p}_{ML}(\underline{y}) = \arg\max_{0<p<1} \prod_{i=1}^{N} \left(\frac{p}{\sqrt{2\pi}} e^{-\frac{(y_i - 1)^2}{2}} + \frac{1-p}{\sqrt{2\pi}} e^{-\frac{y_i^2}{2}} \right)$$

$y=1$ 러

$$\boxed{f_{\underline{Z};\underline{\theta}}(\underline{z}) = f_{\underline{Z},\underline{Y};\underline{\theta}}(\underline{z},\underline{y}), \quad \forall \underline{y}=h(\underline{z}).}$$

$A \subset B$

$\{\underline{Z}=\underline{z}\}$ $\{Z=1\}$

$P(A) = P(A \cap B)$

$= P(A)P(B|A)$

$\{\underline{Y}=\underline{y}\}$ $\{Y=1\}$

$= \{Z=1, \text{ or } Z=-1\}$

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

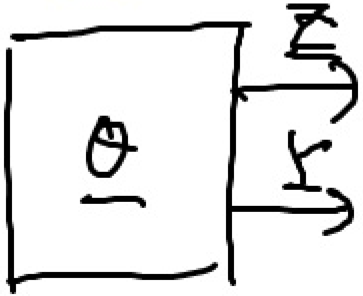
$$= f_{\underline{Y};\underline{\theta}}(\underline{y}) \cdot f_{\underline{Z}|\underline{Y};\underline{\theta}}(\underline{z}|\underline{y}), \quad \forall \underline{y}=h(\underline{z})$$

p6

$$\ln f_{\underline{Y};\underline{\theta}}(\underline{y}) = \ln f_{\underline{Z};\underline{\theta}}(\underline{z}) - \ln f_{\underline{Z}|\underline{Y};\underline{\theta}}(\underline{z}|\underline{y})$$

$$\forall\, \underline{y} = h(\underline{z})$$

$$E_{\hat{\underline{\theta}}}\left[\ln f_{\underline{Y};\underline{\theta}}(\underline{Y}) \,\middle|\, \underline{Y}=\underline{y}\right] = E_{\hat{\underline{\theta}}}\left[\ln f_{\underline{Z};\underline{\theta}}(\underline{Z}) \,\middle|\, \underline{Y}=\underline{y}\right] - E_{\hat{\underline{\theta}}}\left[\ln f_{\underline{Z}|\underline{Y};\underline{\theta}}(\underline{Z}|\underline{Y}) \,\middle|\, \underline{Y}=\underline{y}\right]$$

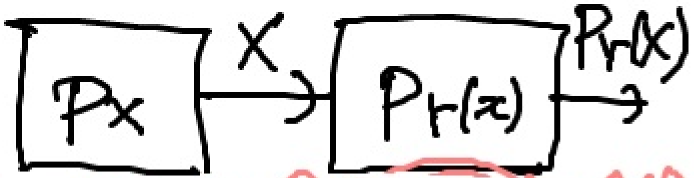


$$\ln f_{\underline{Y};\underline{\theta}}(\underline{y}) = E_{\hat{\underline{\theta}}}\left[\ln f_{\underline{Z};\underline{\theta}}(\underline{Z}) \,\middle|\, \underline{Y}=\underline{y}\right] - E_{\hat{\underline{\theta}}}\left[\ln f_{\underline{Z}|\underline{Y};\underline{\theta}}(\underline{Z}|\underline{y}) \,\middle|\, \underline{Y}=\underline{y}\right]$$

p.17

$$D(p_X \| p_Y) = E\left[\log \frac{p_X(X)}{p_Y(X)}\right] \ge 0.$$

Kullback-Leibler Distance



$$\ln f_{\underline{Y};\theta}(\underline{y}) = E_{\hat\theta}\left[\ln f_{\underline{Z};\theta}(\underline{Z}) \mid \underline{Y}=\underline{y}\right] - E_{\hat\theta}\left[\ln f_{\underline{Z}|\underline{Y};\theta}(\underline{Z}|\underline{y}) \mid \underline{Y}=\underline{y}\right] \quad \cdots (*) \quad \theta = \hat\theta^{(r+1)} \quad \hat\theta = \hat\theta^{(r)}$$

$$\ge E_{\hat\theta}\left[\ln f_{\underline{Z};\theta}(\underline{Z}) \mid \underline{Y}=\underline{y}\right] - E_{\hat\theta}\left[\ln f_{\underline{Z}|\underline{Y};\hat\theta}(\underline{Z}|\underline{y}) \mid \underline{Y}=\underline{y}\right] \quad \cdots (**)$$

To show

$$\ln f_{\underline{Y};\hat\theta^{(r+1)}}(\underline{y}) \ge \ln f_{\underline{Y};\hat\theta^{(r)}}(\underline{y})$$

$$\ln f_{\underline{Y};\hat\theta^{(r+1)}}(\underline{y}) = \hat{E}_{\hat\theta^{(r)}}(r+1) - \tilde{E}_{\hat\theta^{(r)}}(r+1) \ge \hat{E}_{\hat\theta^{(r)}}[r+1] - \tilde{E}_{\hat\theta^{(r)}}[r]$$

p.8

$$\ln f_{\underline{Y};\underline{\hat{\theta}}^{(t+1)}}(\underline{y}) \geq E_{\underline{\hat{\theta}}^{(t)}}\left[\ln f_{\underline{Z};\underline{\hat{\theta}}^{(t+1)}}(\underline{Z}) \,\middle|\, \underline{Y}=\underline{y}\right]$$

$$- E_{\underline{\hat{\theta}}^{(t)}}\left[\ln f_{\underline{Z}|\underline{Y};\underline{\hat{\theta}}^{(t)}}(\underline{Z}|\underline{y}) \,\middle|\, \underline{Y}=\underline{y}\right], \ \forall \underline{\hat{\theta}}^{(t+1)}$$

We choose $\underline{\hat{\theta}}^{(t+1)}$ s.t.

$$E_{\underline{\hat{\theta}}^{(t)}}\left[\ln f_{\underline{Z};\underline{\theta}}(\underline{Z}) \,\middle|\, \underline{Y}=\underline{y}\right]\Big|_{\underline{\theta}=\underline{\hat{\theta}}^{(t+1)}}$$

$$\geq \underbrace{E_{\underline{\hat{\theta}}^{(t)}}\left[\ln f_{\underline{Z};\underline{\hat{\theta}}^{(t)}}(\underline{Z}) \,\middle|\, \underline{Y}=\underline{y}\right]}_{\text{a function of } \underline{\theta}.}$$

① Expectation step to form a function of $\underline{\theta}$.

② Maximization step.

$\ln f_{\underline{Y};\underline{\hat{\theta}}^{(t)}}(\underline{y})$

$$\geq E_{\underline{\hat{\theta}}^{(t)}}\left[\ln f_{\underline{Z};\underline{\hat{\theta}}^{(t)}}(\underline{Z}) \,\middle|\, \underline{Y}=\underline{y}\right] - E_{\underline{\hat{\theta}}^{(t)}}\left[\cdots \underline{\hat{\theta}}^{(t)}\right]$$