دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته
دانشکده مهندسی برق و کامپیوتر

تئوری تخمین
ESTIMATION THEORY

نیم سال اول سال تحصیلی ۰۴-۰۵

تخمین بیز

- مقدمه
- کاربرد اطلاعات پیشین در تخمین
- خواص PDF گوسی
- تخمین بیز خطی
- تخمین بیز پارامترهای ثابت
- تخمین بیشینه احتمال پسین
- تخمین کمینه خطای میانگین مربعات خطی

مقدمه

- در تخمین کلاسیک، پارامتر مورد نظر ثابت و مجهول فرض می شد، اما در تخمین بیز، پارامتر مورد نظر یک متغیر تصادفی است که می خواهیم مقدار یک تحقق آن را تخمین بزنیم.
- در این تخمین، ما یک دانش اولیه از پارامتر مجهول (تابع PDF) داریم که می خواهیم آن را در تخمین وارد کنیم.
- تخمین بیز در مواردی که MVU به ازای همه مقادیر θ وجود ندارد نیز کاربرد دارد، زیرا با در نظر گرفتن یک PDF برای متغیر تصادفی می توان تخمینی به دست آورد که به طور متوسط کارا باشد.

کاربرد اطلاعات پیشین در تخمین

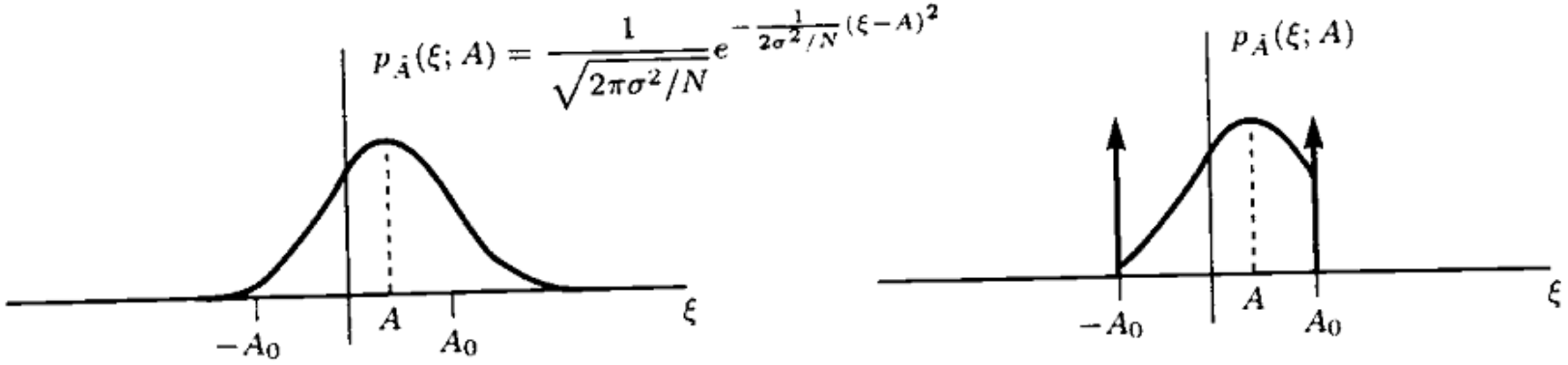
- مثال ۱. کاربرد اطلاعات پیشین در تخمین
  - در مثال معروف تخمین ولتاژ DC، فرض کنید می دانیم که مقدار ولتاژ محدود در بازه $-A_0 \leq A \leq A$ است.
  - دقت کنید که در تخمین کلاسیک $-\infty < A < +\infty$ بود.
  - حال فرض کنید که تخمینگر متوسط گیر ($\hat{A} = \bar{x}$) را به صورت زیر تغییر دهیم:

$$\check{A} = \begin{cases} -A_0 & \bar{x} < -A_0 \\ \bar{x} & -A_0 \leq \bar{x} \leq A_0 \\ A_0 & \bar{x} > A_0 \end{cases}$$

  - PDF تخمینگر جدید:

$$\begin{aligned} p_{\check{A}}(\xi; A) = {} & \Pr\{\bar{x} \leq -A_0\}\delta(\xi + A_0) \\ & + p_{\hat{A}}(\xi; A)[u(\xi + A_0) - u(\xi - A_0)] \\ & + \Pr\{\bar{x} \geq A_0\}\delta(\xi - A_0) \end{aligned}$$

کاربرد اطلاعات پیشین در تخمین



(a) PDF of sample mean

(b) PDF of truncated sample mean

$$
\begin{aligned}
\text{mse}(\hat{A}) &= \int_{-\infty}^{\infty} (\xi - A)^2 p_{\hat{A}}(\xi; A)\, d\xi \\
&= \int_{-\infty}^{-A_0} (\xi - A)^2 p_{\hat{A}}(\xi; A)\, d\xi + \int_{-A_0}^{A_0} (\xi - A)^2 p_{\hat{A}}(\xi; A)\, d\xi \\
&\quad + \int_{A_0}^{\infty} (\xi - A)^2 p_{\hat{A}}(\xi; A)\, d\xi \\
&> \int_{-\infty}^{-A_0} (-A_0 - A)^2 p_{\hat{A}}(\xi; A)\, d\xi + \int_{-A_0}^{A_0} (\xi - A)^2 p_{\hat{A}}(\xi; A)\, d\xi \\
&\quad + \int_{A_0}^{\infty} (A_0 - A)^2 p_{\hat{A}}(\xi; A)\, d\xi \\
&= \text{mse}(\check{A}).
\end{aligned}
$$

کاربرد اطلاعات پیشین در تخمین

- تخمینگر جدید بایاس دارد، ولی خطای میانگین مربعات کمتری دارد.
- سوال: با وجود اطلاعات پیشین، بهترین تخمینگر کدام است؟
- می توان خطای میانگین مربعات را معیار سنجش قرار داد.
- خطای mse در تخمین کلاسیک:

$$\text{mse}(\hat{A}) = \int (\hat{A} - A)^2 p(\mathbf{x}; A)\, d\mathbf{x}$$

- خطای mse در تخمین بیز:

$$\text{Bmse}(\hat{A}) = \int\int (A - \hat{A})^2 p(\mathbf{x}, A)\, d\mathbf{x}\, dA.$$

- ضمنا داریم:

$$p(\mathbf{x}, A) = p(A|\mathbf{x})p(\mathbf{x}) \quad \Rightarrow \quad \text{Bmse}(\hat{A}) = \int \left[\int (A - \hat{A})^2 p(A|\mathbf{x})\, dA \right] p(\mathbf{x})\, d\mathbf{x}.$$

کاربرد اطلاعات پیشین در تخمین

- اگر مقدار داخل کروشه به ازای هر x کمینه شود، خطای mse بیز کمینه می شود. با مشتق گیری داریم:

$$\begin{aligned}\frac{\partial}{\partial \hat{A}} \int (A-\hat{A})^2 p(A|\mathbf{x})\, dA &= \int \frac{\partial}{\partial \hat{A}} (A-\hat{A})^2 p(A|\mathbf{x})\, dA \\ &= \int -2(A-\hat{A}) p(A|\mathbf{x})\, dA \\ &= -2\int A p(A|\mathbf{x})\, dA + 2\hat{A}\int p(A|\mathbf{x})\, dA\end{aligned}$$

- اگر مشتق مساوی صفر قرار داده شود:

$$\hat{A} = \int A p(A|\mathbf{x})\, dA$$

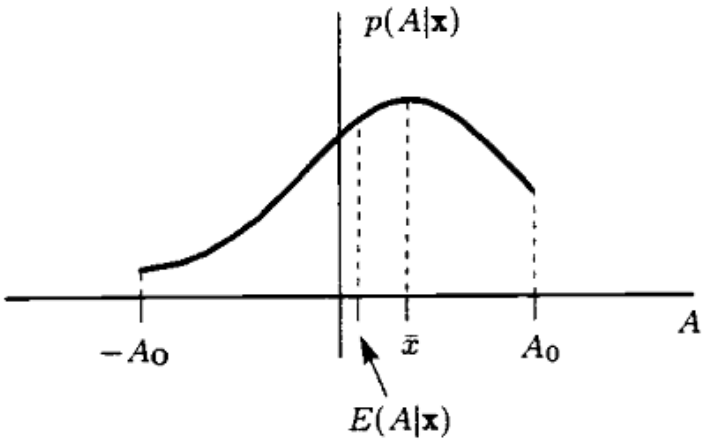
$$\hat{A} = E(A|\mathbf{x}) \quad \Rightarrow \quad \textit{mean of the posterior PDF } p(A|\mathbf{x})$$

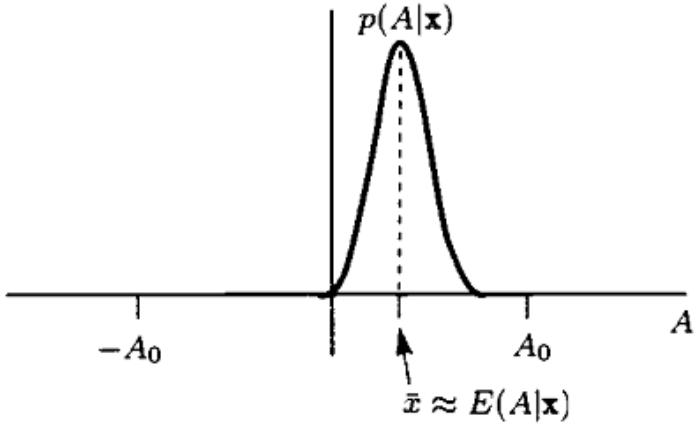
کاربرد اطلاعات پیشین در تخمین

- استفاده از قضیه بیز برای تبدیل PDF پسین:

$$p(A|\mathbf{x}) = \frac{p(\mathbf{x}|A)p(A)}{p(\mathbf{x})} = \frac{p(\mathbf{x}|A)p(A)}{\int p(\mathbf{x}|A)p(A)\,dA}.$$

- در داده وسیع، اثر اطلاعات پیشین کمتر می شود:





کاربرد اطلاعات پیشین در تخمین

- مثال ۲. ولتاژ DC در نویز گوسی
  - در اینجا فرض می کنیم اطلاعات پیشین ما از سطح ولتاژ، به صورت یک PDF گوسی است:

$$p(A) = \frac{1}{\sqrt{2\pi\sigma_A^2}} \exp\left[-\frac{1}{2\sigma_A^2}(A-\mu_A)^2\right].$$

  - ضمنا داریم:

$$\begin{aligned} p(\mathbf{x}|A) &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right] \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]\right] \exp\left[-\frac{1}{2\sigma^2}(NA^2-2NA\bar{x})\right] \end{aligned}$$

کاربرد اطلاعات پیشین در تخمین

- حال برای به دست آوردن تخمین MMSE داریم:

$$p(A|\mathbf{x}) = \frac{p(\mathbf{x}|A)p(A)}{\int p(\mathbf{x}|A)p(A)\,dA}$$

$$= \frac{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}\sqrt{2\pi\sigma_A^2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]\right]\exp\left[-\frac{1}{2\sigma^2}(NA^2-2NA\bar{x})\right]}{\int_{-\infty}^{\infty}\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}\sqrt{2\pi\sigma_A^2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]\right]\exp\left[-\frac{1}{2\sigma^2}(NA^2-2NA\bar{x})\right]} \cdot \frac{\exp\left[-\frac{1}{2\sigma_A^2}(A-\mu_A)^2\right]}{\exp\left[-\frac{1}{2\sigma_A^2}(A-\mu_A)^2\right]\,dA}$$

$$= \frac{\exp\left[-\frac{1}{2}\left(\frac{1}{\sigma^2}(NA^2-2NA\bar{x})+\frac{1}{\sigma_A^2}(A-\mu_A)^2\right)\right]}{\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}\left(\frac{1}{\sigma^2}(NA^2-2NA\bar{x})+\frac{1}{\sigma_A^2}(A-\mu_A)^2\right)\right]} = \frac{\exp\left[-\frac{1}{2}Q(A)\right]}{\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}Q(A)\right]\,dA}$$

کاربرد اطلاعات پیشین در تخمین

- داریم:

$$\begin{aligned} Q(A) &= \frac{N}{\sigma^2}A^2 - \frac{2NA\bar{x}}{\sigma^2} + \frac{A^2}{\sigma_A^2} - \frac{2\mu_A A}{\sigma_A^2} + \frac{\mu_A^2}{\sigma_A^2} \\ &= \left(\frac{N}{\sigma^2} + \frac{1}{\sigma_A^2}\right)A^2 - 2\left(\frac{N}{\sigma^2}\bar{x} + \frac{\mu_A}{\sigma_A^2}\right)A + \frac{\mu_A^2}{\sigma_A^2}. \end{aligned}$$

- با قرار دادن:

$$\begin{aligned} \sigma_{A|x}^2 &= \frac{1}{\frac{N}{\sigma^2} + \frac{1}{\sigma_A^2}} \\ \mu_{A|x} &= \left(\frac{N}{\sigma^2}\bar{x} + \frac{\mu_A}{\sigma_A^2}\right)\sigma_{A|x}^2. \end{aligned}$$

- خواهیم داشت:

$$Q(A) = \frac{1}{\sigma_{A|x}^2}\left(A^2 - 2\mu_{A|x}A + \mu_{A|x}^2\right) - \frac{\mu_{A|x}^2}{\sigma_{A|x}^2} + \frac{\mu_A^2}{\sigma_A^2} = \frac{1}{\sigma_{A|x}^2}(A - \mu_{A|x})^2 - \frac{\mu_{A|x}^2}{\sigma_{A|x}^2} + \frac{\mu_A^2}{\sigma_A^2}$$

کاربرد اطلاعات پیشین در تخمین

- بنابراین:

$$p(A|\mathbf{x}) = \frac{\exp\left[-\frac{1}{2\sigma_{A|x}^2}(A-\mu_{A|x})^2\right]\exp\left[-\frac{1}{2}\left(\frac{\mu_A^2}{\sigma_A^2}-\frac{\mu_{A|x}^2}{\sigma_{A|x}^2}\right)\right]}{\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2\sigma_{A|x}^2}(A-\mu_{A|x})^2\right]\exp\left[-\frac{1}{2}\left(\frac{\mu_A^2}{\sigma_A^2}-\frac{\mu_{A|x}^2}{\sigma_{A|x}^2}\right)\right]dA}$$

$$= \frac{1}{\sqrt{2\pi\sigma_{A|x}^2}}\exp\left[-\frac{1}{2\sigma_{A|x}^2}(A-\mu_{A|x})^2\right]$$

$$\hat{A} = E(A|\mathbf{x}) = \mu_{A|x} = \frac{\frac{N}{\sigma^2}\bar{x}+\frac{\mu_A}{\sigma_A^2}}{\frac{N}{\sigma^2}+\frac{1}{\sigma_A^2}}$$

کاربرد اطلاعات پیشین در تخمین

- در نهایت تخمینگر MMSE به صورت زیر به دست می آید:

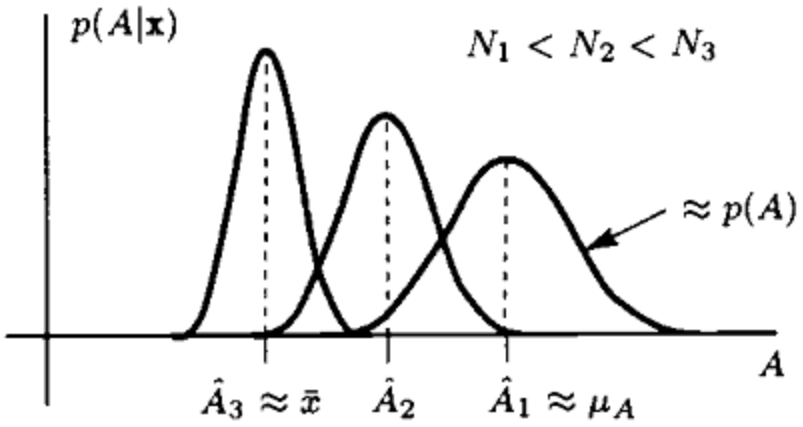
$$\hat{A} = \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N}}\bar{x} + \frac{\frac{\sigma^2}{N}}{\sigma_A^2 + \frac{\sigma^2}{N}}\mu_A = \alpha\bar{x} + (1-\alpha)\mu_A$$

$$\text{var}(A|\mathbf{x}) = \sigma_{A|x}^2 = \frac{1}{\frac{N}{\sigma^2} + \frac{1}{\sigma_A^2}}$$

- که:

$$\alpha = \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N}}.$$

- اثر افزایش داده:



کاربرد اطلاعات پیشین در تخمین

- در مورد خطای تخمین MMSE داریم:

$$\begin{aligned}\text{Bmse}(\hat{A}) &= \int\int (A-\hat{A})^2 p(\mathbf{x},A)\,d\mathbf{x}\,dA \\ &= \int\int (A-\hat{A})^2 p(A|\mathbf{x})\,dA p(\mathbf{x})\,d\mathbf{x}.\end{aligned}$$

$$\begin{aligned}\text{Bmse}(\hat{A}) &= \int\int [A-E(A|\mathbf{x})]^2 p(A|\mathbf{x})\,dA p(\mathbf{x})\,d\mathbf{x} \\ &= \int \text{var}(A|\mathbf{x}) p(\mathbf{x})\,d\mathbf{x}.\end{aligned}$$

- در مثال اخیر داشتیم:

$$\begin{aligned}\text{Bmse}(\hat{A}) &= \int \sigma^2_{A|x} p(\mathbf{x})\,d\mathbf{x} \\ &= \frac{1}{\frac{N}{\sigma^2}+\frac{1}{\sigma_A^2}}\end{aligned}$$

$$\Rightarrow \quad \text{Bmse}(\hat{A}) = \frac{\sigma^2}{N}\left(\frac{\sigma_A^2}{\sigma_A^2+\frac{\sigma^2}{N}}\right) \quad \Rightarrow \quad \text{Bmse}(\hat{A}) < \frac{\sigma^2}{N}$$

خواص PDF گوسی

Theorem 10.1 (Conditional PDF of Bivariate Gaussian) *If x and y are distributed according to a bivariate Gaussian PDF with mean vector $[E(x)\,E(y)]^T$ and covariance matrix*

$$\mathbf{C} = \begin{bmatrix} \text{var}(x) & \text{cov}(x,y) \\ \text{cov}(y,x) & \text{var}(y) \end{bmatrix}$$

so that

$$p(x,y) = \frac{1}{2\pi \det^{\frac{1}{2}}(\mathbf{C})} \exp\left[-\frac{1}{2}\begin{bmatrix} x-E(x) \\ y-E(y) \end{bmatrix}^T \mathbf{C}^{-1} \begin{bmatrix} x-E(x) \\ y-E(y) \end{bmatrix}\right],$$

then the conditional PDF $p(y|x)$ is also Gaussian and

$$E(y|x) = E(y) + \frac{\text{cov}(x,y)}{\text{var}(x)}(x - E(x)) \tag{10.16}$$

$$\text{var}(y|x) = \text{var}(y) - \frac{\text{cov}^2(x,y)}{\text{var}(x)}. \tag{10.17}$$

$$\text{var}(y|x) = \text{var}(y)\left[1 - \frac{\text{cov}^2(x,y)}{\text{var}(x)\text{var}(y)}\right] = \text{var}(y)(1-\rho^2)$$

خواص PDF گوسی

Theorem 10.2 (Conditional PDF of Multivariate Gaussian) *If* $\mathbf{x}$ *and* $\mathbf{y}$ *are jointly Gaussian, where* $\mathbf{x}$ *is* $k \times 1$ *and* $\mathbf{y}$ *is* $l \times 1$*, with mean vector* $[E(\mathbf{x})^T \; E(\mathbf{y})^T]^T$ *and partitioned covariance matrix*

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{xx} & \mathbf{C}_{xy} \\ \mathbf{C}_{yx} & \mathbf{C}_{yy} \end{bmatrix} = \begin{bmatrix} k \times k & k \times l \\ l \times k & l \times l \end{bmatrix} \tag{10.23}$$

so that

$$p(\mathbf{x}, \mathbf{y}) = \frac{1}{(2\pi)^{\frac{k+l}{2}} \det^{\frac{1}{2}}(\mathbf{C})} \exp\left[-\frac{1}{2} \left(\begin{bmatrix} \mathbf{x} - E(\mathbf{x}) \\ \mathbf{y} - E(\mathbf{y}) \end{bmatrix} \right)^T \mathbf{C}^{-1} \left(\begin{bmatrix} \mathbf{x} - E(\mathbf{x}) \\ \mathbf{y} - E(\mathbf{y}) \end{bmatrix} \right) \right],$$

then the conditional PDF $p(\mathbf{y}|\mathbf{x})$ *is also Gaussian and*

$$\begin{aligned} E(\mathbf{y}|\mathbf{x}) &= E(\mathbf{y}) + \mathbf{C}_{yx}\mathbf{C}_{xx}^{-1}(\mathbf{x} - E(\mathbf{x})) \\ \mathbf{C}_{y|x} &= \mathbf{C}_{yy} - \mathbf{C}_{yx}\mathbf{C}_{xx}^{-1}\mathbf{C}_{xy}. \end{aligned}$$

مدل خطی بیز

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

where $\mathbf{x}$ is an $N \times 1$ data vector, $\mathbf{H}$ is a known $N \times p$ matrix, $\boldsymbol{\theta}$ is a $p \times 1$ random vector with prior PDF $\mathcal{N}(\boldsymbol{\mu}_\theta, \mathbf{C}_\theta)$, and $\mathbf{w}$ is an $N \times 1$ noise vector with PDF $\mathcal{N}(\mathbf{0}, \mathbf{C}_w)$ and independent of $\boldsymbol{\theta}$. This data model is termed the *Bayesian general linear model.*

- برای استفاده از قضیه قبل داریم:

$$\begin{aligned} E(\mathbf{x}) &= E(\mathbf{H}\boldsymbol{\theta} + \mathbf{w}) = \mathbf{H}E(\boldsymbol{\theta}) = \mathbf{H}\boldsymbol{\mu}_\theta \\ E(\mathbf{y}) &= E(\boldsymbol{\theta}) = \boldsymbol{\mu}_\theta \end{aligned}$$

$$\begin{aligned} \mathbf{C}_{xx} &= E\left[(\mathbf{x} - E(\mathbf{x}))(\mathbf{x} - E(\mathbf{x}))^T\right] \\ &= E\left[(\mathbf{H}\boldsymbol{\theta} + \mathbf{w} - \mathbf{H}\boldsymbol{\mu}_\theta)(\mathbf{H}\boldsymbol{\theta} + \mathbf{w} - \mathbf{H}\boldsymbol{\mu}_\theta)^T\right] \\ &= E\left[(\mathbf{H}(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta) + \mathbf{w})(\mathbf{H}(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta) + \mathbf{w})^T\right] \\ &= \mathbf{H}E\left[(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta)(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta)^T\right]\mathbf{H}^T + E(\mathbf{w}\mathbf{w}^T) \\ &= \mathbf{H}\mathbf{C}_\theta\mathbf{H}^T + \mathbf{C}_w \end{aligned}$$

$$\begin{aligned} \mathbf{C}_{yx} &= E\left[(\mathbf{y} - E(\mathbf{y}))(\mathbf{x} - E(\mathbf{x}))^T\right] \\ &= E\left[(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta)(\mathbf{H}(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta) + \mathbf{w})^T\right] \\ &= E\left[(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta)(\mathbf{H}(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta))^T\right] \\ &= \mathbf{C}_\theta\mathbf{H}^T. \end{aligned}$$

مدل خطی بیز

Theorem 10.3 (Posterior PDF for the Bayesian General Linear Model) *If the observed data* $\mathbf{x}$ *can be modeled as*

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w} \tag{10.27}$$

where $\mathbf{x}$ *is an* $N \times 1$ *data vector,* $\mathbf{H}$ *is a known* $N \times p$ *matrix,* $\boldsymbol{\theta}$ *is a* $p \times 1$ *random vector with prior PDF* $\mathcal{N}(\boldsymbol{\mu}_\theta, \mathbf{C}_\theta)$*, and* $\mathbf{w}$ *is an* $N \times 1$ *noise vector with PDF* $\mathcal{N}(\mathbf{0}, \mathbf{C}_w)$ *and independent of* $\boldsymbol{\theta}$*, then the posterior PDF* $p(\boldsymbol{\theta}|\mathbf{x})$ *is Gaussian with mean*

$$E(\boldsymbol{\theta}|\mathbf{x}) = \boldsymbol{\mu}_\theta + \mathbf{C}_\theta\mathbf{H}^T(\mathbf{H}\mathbf{C}_\theta\mathbf{H}^T + \mathbf{C}_w)^{-1}(\mathbf{x} - \mathbf{H}\boldsymbol{\mu}_\theta) \tag{10.28}$$

and covariance

$$\mathbf{C}_{\theta|x} = \mathbf{C}_\theta - \mathbf{C}_\theta\mathbf{H}^T(\mathbf{H}\mathbf{C}_\theta\mathbf{H}^T + \mathbf{C}_w)^{-1}\mathbf{H}\mathbf{C}_\theta. \tag{10.29}$$

In contrast to the classical general linear model, $\mathbf{H}$ need not be full rank to ensure the invertibility of $\mathbf{H}\mathbf{C}_\theta\mathbf{H}^T + \mathbf{C}_w$.

مدل خطی بیز

- مثال ۳.

$x[n] = A + w[n]$ for $n = 0, 1, \ldots, N-1$ with $A \sim \mathcal{N}(\mu_A, \sigma_A^2)$ and $w[n]$ is WGN with variance σ^2

- هدف: تخمین A
- مدل خطی مساله:

$$\mathbf{x} = \mathbf{1}A + \mathbf{w}.$$

- طبق روابط پیش گفته داریم:

$$E(A|\mathbf{x}) = \mu_A + \sigma_A^2 \mathbf{1}^T (\mathbf{1}\sigma_A^2 \mathbf{1}^T + \sigma^2 \mathbf{I})^{-1} (\mathbf{x} - \mathbf{1}\mu_A).$$

- با توجه به تساوی زیر:

$$\left(\mathbf{I} + \frac{\sigma_A^2}{\sigma^2}\mathbf{1}\mathbf{1}^T\right)^{-1} = \mathbf{I} - \frac{\frac{\sigma_A^2}{\sigma^2}\mathbf{1}\mathbf{1}^T}{1 + N\frac{\sigma_A^2}{\sigma^2}}$$

مدل خطی بیز

- مثال ۳. ادامه
  - خواهیم داشت:

$$
\begin{aligned}
E(A|\mathbf{x}) &= \mu_A + \frac{\sigma_A^2}{\sigma^2}\mathbf{1}^T\left(\mathbf{I} - \frac{\mathbf{1}\mathbf{1}^T}{N + \frac{\sigma^2}{\sigma_A^2}}\right)(\mathbf{x} - \mathbf{1}\mu_A) \\
&= \mu_A + \frac{\sigma_A^2}{\sigma^2}\left(\mathbf{1}^T - \frac{N}{N + \frac{\sigma^2}{\sigma_A^2}}\mathbf{1}^T\right)(\mathbf{x} - \mathbf{1}\mu_A) \\
&= \mu_A + \frac{\sigma_A^2}{\sigma^2}\left(1 - \frac{N}{N + \frac{\sigma^2}{\sigma_A^2}}\right)(N\bar{x} - N\mu_A) \\
&= \mu_A + \frac{N}{N + \frac{\sigma^2}{\sigma_A^2}}(\bar{x} - \mu_A) \\
&= \mu_A + \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N}}(\bar{x} - \mu_A).
\end{aligned}
$$

مدل خطی بیز

- مثال۳. ادامه

$$\mathrm{var}(A|\mathbf{x}) = \sigma_A^2 - \sigma_A^2 \mathbf{1}^T (\mathbf{1}\sigma_A^2 \mathbf{1}^T + \sigma^2 \mathbf{I})^{-1} \mathbf{1}\sigma_A^2.$$

$$\begin{aligned}
\mathrm{var}(A|\mathbf{x}) &= \sigma_A^2 - \frac{\sigma_A^2}{\sigma^2} \mathbf{1}^T \left(\mathbf{I} - \frac{\mathbf{1}\mathbf{1}^T}{N + \frac{\sigma^2}{\sigma_A^2}} \right) \mathbf{1}\sigma_A^2 \\
&= \sigma_A^2 - \frac{\sigma_A^4}{\sigma^2} \left(N - \frac{N^2}{N + \frac{\sigma^2}{\sigma_A^2}} \right) \\
&= \frac{\frac{\sigma^2}{N}\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N}}
\end{aligned}$$

مدل خطی بیز

- روابط مهم این بخش:

$$E(\boldsymbol{\theta}|\mathbf{x}) = \boldsymbol{\mu}_\theta + \left(\mathbf{C}_\theta^{-1} + \mathbf{H}^T\mathbf{C}_w^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{C}_w^{-1}(\mathbf{x} - \mathbf{H}\boldsymbol{\mu}_\theta)$$

$$\mathbf{C}_{\theta|x} = \left(\mathbf{C}_\theta^{-1} + \mathbf{H}^T\mathbf{C}_w^{-1}\mathbf{H}\right)^{-1}$$

$$\mathbf{C}_{\theta|x}^{-1} = \mathbf{C}_\theta^{-1} + \mathbf{H}^T\mathbf{C}_w^{-1}\mathbf{H}.$$

تخمین بیز پارامتر ثابت

- همانطور که گفته شد، تخمینگر MVU همیشه وجود ندارد، ولی تخمینگر MMSE با رویکرد تخمین بیز همیشه وجود دارد و تخمینگری را ارئه می دهد که «به طور متوسط» بهینه است؛ ولی ممکن است برای یک مقدار خاص θ بهتر از MVU نباشد.
- به عنوان نمونه، مثال قبل را در نظر بگیرید:

$$\begin{aligned}\hat{A} &= \frac{\sigma_A^2}{\sigma_A^2+\sigma^2/N}\bar{x}+\frac{\sigma^2/N}{\sigma_A^2+\sigma^2/N}\mu_A \\ &= \alpha\bar{x}+(1-\alpha)\mu_A\end{aligned}$$

- اگر A یک پارامتر ثابت باشد، می توان مقدار mse را به صورت زیر محاسبه نمود:

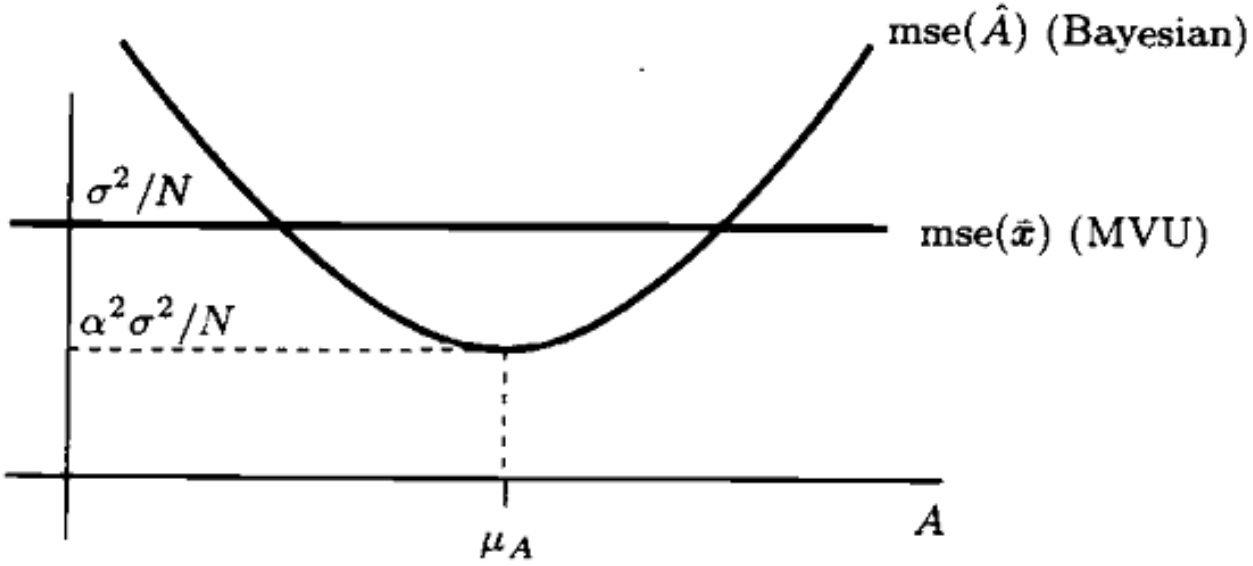
$$\text{mse}(\hat{A}) = \text{var}(\hat{A}) + b^2(\hat{A})$$

$$b(\hat{A}) = E(\hat{A}) - A$$

تخمین بیز پارامتر ثابت

- بنابراین:

$$\begin{aligned}\text{mse}(\hat{A}) &= \alpha^2 \text{var}(\bar{x}) + [\alpha A + (1-\alpha)\mu_A - A]^2 \\ &= \alpha^2 \frac{\sigma^2}{N} + (1-\alpha)^2 (A - \mu_A)^2.\end{aligned}$$



تخمین بیز پارامتر ثابت

- مشخص است که اگر میانگین پیشین به A نزدیک باشد، خطای کمتری خواهیم داشت.
- این خطای کمتر به بهای بایاس به دست آمده است:

$$E(\hat{A}) = \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2/N} A + \frac{\sigma^2/N}{\sigma_A^2 + \sigma^2/N} \mu_A.$$

- ضمنا در صورتی که هیچ اطلاعات پیشین در مورد پارامتر مورد نظر نداشته باشیم، در واقع مانند این است که: $\sigma_A^2 \to \infty$
- در این صورت: $\alpha \to 1$ و بنابراین:

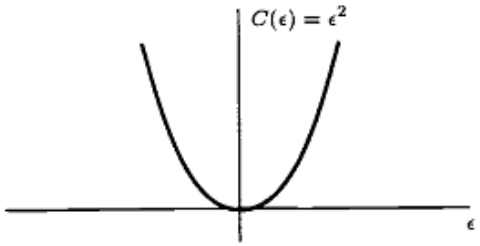
$mse(\hat{A}) \to \frac{\sigma^2}{N}$ (نمودار منحنی، به خط صاف میل می کند)

$$E(\hat{A}) \to A$$

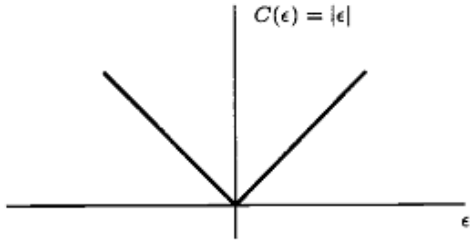
انواع تابع خطا

The Bayes risk $\mathcal{R}$ is

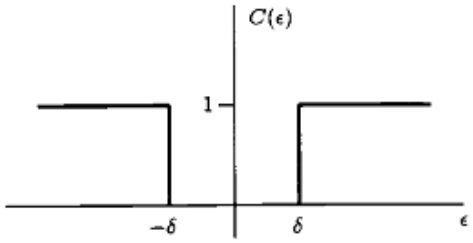
$$\begin{aligned}\mathcal{R} &= E[\mathcal{C}(\epsilon)] \\ &= \int\int \mathcal{C}(\theta-\hat{\theta})p(\mathbf{x},\theta)\,d\mathbf{x}\,d\theta \\ &= \int\left[\int \mathcal{C}(\theta-\hat{\theta})p(\theta|\mathbf{x})\,d\theta\right]p(\mathbf{x})\,d\mathbf{x}.\end{aligned}$$



(a) Quadratic error

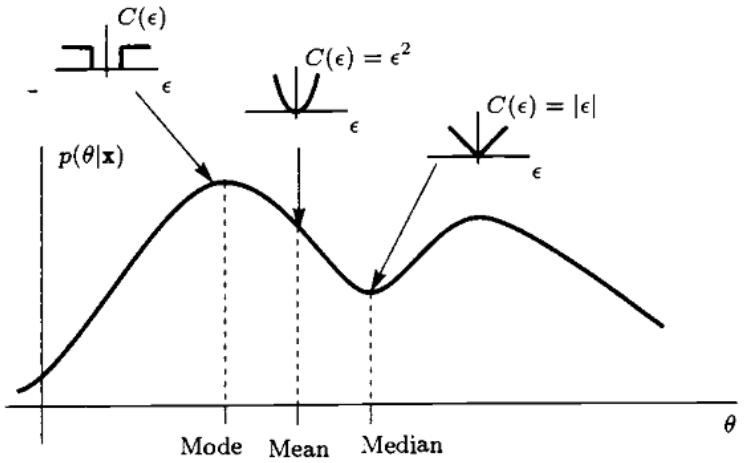


(b) Absolute error

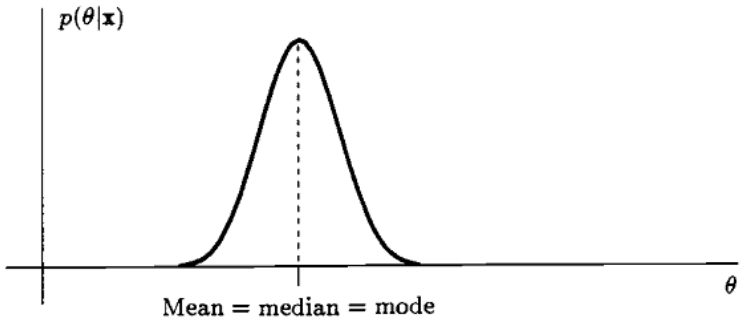


(c) Hit-or-miss error

انواع تابع خطا



(a) General posterior PDF



(b) Gaussian posterior PDF

Figure 11.2 Estimators for different cost functions

تخمینگر بیشینه احتمال پسین (MAP)

- در این تخمینگر هدف یافتن پارامتر مجهول به نحوی است که احتمال پسین بیشینه شود.
- این تخمینگر تابع خطا از نوع hit-or-miss را کمینه می کند.

$$\hat{\theta} = \arg\max_{\theta} p(\theta|\mathbf{x}).$$

$$p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)p(\theta)}{p(\mathbf{x})} \quad \Rightarrow \quad \hat{\theta} = \arg\max_{\theta} p(\mathbf{x}|\theta)p(\theta)$$

- می توان نوشت:

$$\hat{\theta} = \arg\max_{\theta} \left[\ln p(\mathbf{x}|\theta) + \ln p(\theta)\right].$$

تخمینگر بیشینه احتمال پسین (MAP)

- مثال۴. فرض کنید داریم:

$$p(x[n]|\theta) = \begin{cases} \theta \exp(-\theta x[n]) & x[n] > 0 \\ 0 & x[n] < 0 \end{cases}$$

- به طوری که x[n] ها IID مشروط هستند:

$$p(\mathbf{x}|\theta) = \prod_{n=0}^{N-1} p(x[n]|\theta)$$

- و ضمنا داریم:

$$p(\theta) = \begin{cases} \lambda \exp(-\lambda\theta) & \theta > 0 \\ 0 & \theta < 0. \end{cases}$$

- تخمین MAP از θ را پیدا کنید.

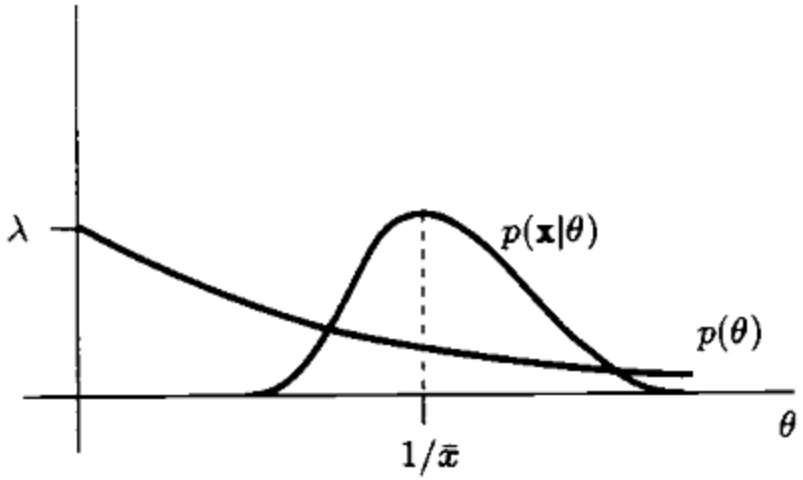
$$\begin{aligned} g(\theta) &= \ln p(\mathbf{x}|\theta) + \ln p(\theta) \\ &= \ln \left[\theta^N \exp\left(-\theta \sum_{n=0}^{N-1} x[n] \right) \right] + \ln \left[\lambda \exp(-\lambda\theta) \right] \\ &= N \ln \theta - N\theta\bar{x} + \ln \lambda - \lambda\theta \end{aligned}$$

تخمینگر بیشینه احتمال پسین (MAP)

- مثال۴. ادامه
  - بیشینه تابع فوق را با مشتق گیری پیدا می کنیم:

$$\frac{dg(\theta)}{d\theta} = \frac{N}{\theta} - N\bar{x} - \lambda$$

$$\hat{\theta} = \frac{1}{\bar{x} + \frac{\lambda}{N}}.$$



تخمینگر کمینه خطای میانگین مربعات خطی LMMSE

- در این تخمینگر به دنبال تخمینگر خطی زیر هستیم:

$$\hat{\theta} = \sum_{n=0}^{N-1} a_n x[n] + a_N$$

- که خطای میانگین مربعات زیر را کمینه می کند:

$$\text{Bmse}(\hat{\theta}) = E\left[(\theta - \hat{\theta})^2\right]$$

- ضمنا ما pdf مشترک داده و پارامتر را نداریم، و فقط ممان های اول و دوم را در اختیار داریم.
- در اینجا تخمینگر LMMSE به صورت زیر حاصل خواهد شد:

$$\hat{\theta} = E(\theta) + \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}(\mathbf{x} - E(\mathbf{x})).$$

- و ضمنا:

$$\text{Bmse}(\hat{\theta}) = C_{\theta\theta} - \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{C}_{x\theta}.$$

تخمینگر کمینه خطای میانگین مربعات خطی LMMSE

- مثال۵. ولتاژ DC در نویز WGN با احتمال پیشین یکنواخت

$$x[n] = A + w[n] \qquad n = 0, 1, \ldots, N-1$$

where $A \sim \mathcal{U}[-A_0, A_0]$, $w[n]$ is WGN with variance σ^2, and A and $w[n]$ are independent.

- به دلیل نیاز به انتگرال گیری، تخمینگر MMSE را به صورت بسته نمی توان به دست آورد.
- چون E(x)=0، داریم:

$$\begin{aligned}
\mathbf{C}_{xx} &= E(\mathbf{x}\mathbf{x}^T) \\
&= E\left[(A\mathbf{1} + \mathbf{w})(A\mathbf{1} + \mathbf{w})^T\right] \\
&= E(A^2)\mathbf{1}\mathbf{1}^T + \sigma^2\mathbf{I} \\
\mathbf{C}_{\theta x} &= E(A\mathbf{x}^T) \\
&= E\left[A(A\mathbf{1} + \mathbf{w})^T\right] \\
&= E(A^2)\mathbf{1}^T
\end{aligned}$$

تخمینگر کمینه خطای میانگین مربعات خطی LMMSE

- مثال۵. ادامه
  - بنابراین:

$$\begin{aligned}\hat{A} &= \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{x} \\ &= \sigma_A^2\mathbf{1}^T(\sigma_A^2\mathbf{1}\mathbf{1}^T+\sigma^2\mathbf{I})^{-1}\mathbf{x}\end{aligned}$$

  - این تخمینگر همان فرم تخمینگر مثال۳ را دارد. یعنی با توجه به صفر بودن میانگین A خواهیم داشت:

$$\hat{A}=\frac{\sigma_A^2}{\sigma_A^2+\frac{\sigma^2}{N}}\bar{x}.$$

  - و ضمنا چون $\sigma_A^2 = E(A^2) = (2A_0)^2/12 = A_0^2/3$ ، در نهایت:

$$\hat{A}=\frac{\frac{A_0^2}{3}}{\frac{A_0^2}{3}+\frac{\sigma^2}{N}}\bar{x}.$$

LMMSE برداری

- در حالتی که چند پارامتر مجهول را به روش LMMSE تخمین می زنیم، هدف تخمین هر کدام از پارامترها به شکل زیر است:

$$\hat{\theta}_i = \sum_{n=0}^{N-1} a_{in} x[n] + a_{iN}$$

- به طوری که تابع زیر کمینه شود:

$$\text{Bmse}(\hat{\theta}_i) = E\left[(\theta_i - \hat{\theta}_i)^2\right] \qquad i = 1, 2, \ldots, p$$

- این تخمینگر به شکل زیر حاصل می شود:

$$E(\boldsymbol{\theta}) + \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}(\mathbf{x} - E(\mathbf{x}))$$

- و ضمنا:

$$\begin{aligned} \mathbf{M}_{\hat{\theta}} &= E\left[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T\right] \\ &= \mathbf{C}_{\theta\theta} - \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{C}_{x\theta} \end{aligned} \quad \Rightarrow \quad \text{Bmse}(\hat{\theta}_i) = [\mathbf{M}_{\hat{\theta}}]_{ii}.$$

LMMSE برداری برای مدل داده خطی

- در اینجا مدل داده خطی است:

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

- هدف این است که LMMSE را برای این مدل داده به دست آوریم.
در اینجا داریم:

$$\begin{aligned} E(\mathbf{x}) &= \mathbf{H}E(\boldsymbol{\theta}) \\ \mathbf{C}_{xx} &= \mathbf{H}\mathbf{C}_{\theta\theta}\mathbf{H}^T + \mathbf{C}_w \\ \mathbf{C}_{\theta x} &= \mathbf{C}_{\theta\theta}\mathbf{H}^T. \end{aligned}$$

- بنابراین کافی است مقادیر فوق را در رابطه تخمینگر LMMSE جایگزین کنیم.

LMMSE برداری برای مدل داده خطی

Theorem 12.1 (Bayesian Gauss-Markov Theorem) *If the data are described by the Bayesian linear model form*

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w} \tag{12.25}$$

where $\mathbf{x}$ *is an* $N \times 1$ *data vector,* $\mathbf{H}$ *is a known* $N \times p$ *observation matrix,* $\boldsymbol{\theta}$ *is a* $p \times 1$ *random vector of parameters whose realization is to be estimated and has mean* $E(\boldsymbol{\theta})$ *and covariance matrix* $\mathbf{C}_{\theta\theta}$*, and* $\mathbf{w}$ *is an* $N \times 1$ *random vector with zero mean and covariance matrix* $\mathbf{C}_w$ *and is uncorrelated with* $\boldsymbol{\theta}$ *(the joint PDF* $p(\mathbf{w}, \boldsymbol{\theta})$ *is otherwise arbitrary), then the LMMSE estimator of* $\boldsymbol{\theta}$ *is*

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= E(\boldsymbol{\theta}) + \mathbf{C}_{\theta\theta}\mathbf{H}^T(\mathbf{H}\mathbf{C}_{\theta\theta}\mathbf{H}^T + \mathbf{C}_w)^{-1}(\mathbf{x} - \mathbf{H}E(\boldsymbol{\theta})) && (12.26) \\ &= E(\boldsymbol{\theta}) + (\mathbf{C}_{\theta\theta}^{-1} + \mathbf{H}^T\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}_w^{-1}(\mathbf{x} - \mathbf{H}E(\boldsymbol{\theta})). && (12.27) \end{aligned}$$

The performance of the estimator is measured by the error $\boldsymbol{\epsilon} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$ *whose mean is zero and whose covariance matrix is*

$$\begin{aligned} \mathbf{C}_\epsilon &= E_{x,\theta}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T) \\ &= \mathbf{C}_{\theta\theta} - \mathbf{C}_{\theta\theta}\mathbf{H}^T(\mathbf{H}\mathbf{C}_{\theta\theta}\mathbf{H}^T + \mathbf{C}_w)^{-1}\mathbf{H}\mathbf{C}_{\theta\theta} && (12.28) \\ &= (\mathbf{C}_{\theta\theta}^{-1} + \mathbf{H}^T\mathbf{C}_w^{-1}\mathbf{H})^{-1}. && (12.29) \end{aligned}$$

The error covariance matrix is also the minimum MSE matrix $\mathbf{M}_{\hat{\theta}}$ *whose diagonal elements yield the minimum Bayesian MSE*

$$\begin{aligned} [\mathbf{M}_{\hat{\theta}}]_{ii} &= [\mathbf{C}_\epsilon]_{ii} \\ &= \mathrm{Bmse}(\hat{\theta}_i). && (12.30) \end{aligned}$$

LMMSE متوالی

- در این روش، می توانیم به ازای هر داده جدید، تخمین قبلی را بروزرسانی کنیم.
- بنابراین در صورت در اختیار داشتن تخمین قبلی، نیازی به محاسبات ماتریسی برای تخمین جدید نیست.
- در مثال تخمین ولتاژ DC در نویز گوسی با اطلاعات قبلی گوسی، فرض کنید میانگین PDF پیشین صفر باشد. در این صورت تخمین A با استفاده از x[0] تا x[N-1] را به صورت زیر به دست آوردیم:

$$\hat{A}[N-1] = \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N}} \bar{x}$$

- و خطای آن:

$$\text{Bmse}(\hat{A}[N-1]) = \frac{\sigma_A^2 \sigma^2}{N\sigma_A^2 + \sigma^2}.$$

LMMSE متوالی

- حال فرض کنید داده x[N] به داده های قبلی اضافه شده است.

چگونه می توانیم بدون اینکه تخمین را برای N+1 داده تکرار کنیم، با استفاده از تخمین قبلی و داده جدید، تخمین را بروز کنیم؟ داریم:

$$
\begin{aligned}
\hat{A}[N] &= \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N+1}} \frac{1}{N+1} \sum_{n=0}^{N} x[n] \\
&= \frac{N\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} \frac{1}{N} \left(\sum_{n=0}^{N-1} x[n] + x[N] \right) \\
&= \frac{N\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} \frac{\sigma_A^2 + \frac{\sigma^2}{N}}{\sigma_A^2} \hat{A}[N-1] + \frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} x[N] \\
&= \frac{N\sigma_A^2 + \sigma^2}{(N+1)\sigma_A^2 + \sigma^2} \hat{A}[N-1] + \frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} x[N] \\
&= \hat{A}[N-1] + \left(\frac{N\sigma_A^2 + \sigma^2}{(N+1)\sigma_A^2 + \sigma^2} - 1 \right) \hat{A}[N-1] + \frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} x[N] \\
&= \hat{A}[N-1] + \frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} (x[N] - \hat{A}[N-1]).
\end{aligned}
$$

LMMSE متوالی

- همانطور که دیده شد، تخمین جدید جمع تخمین قبلی است به علاوه ضریبی از خطای $x[N] - \hat{A}[N-1]$
- این ضریب بهره یا scaling را می توان به صورت زیر بیان کرد:

$$K[N] = \frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} = \frac{\text{Bmse}(\hat{A}[N-1])}{\text{Bmse}(\hat{A}[N-1]) + \sigma^2}$$

- ضریب بهره با افزایش N به سمت صفر میل می کند، یعنی با افزایش داده، داده های جدید اطلاعات زیادی در مورد پارامتر مورد نظر ارائه نمی کنند.

LMMSE متوالی

- همچنین می توان خطای تخمین را نیز بروز کرد:

$$\begin{aligned}\text{Bmse}(\hat{A}[N]) &= \frac{\sigma_A^2\sigma^2}{(N+1)\sigma_A^2+\sigma^2} \\ &= \frac{N\sigma_A^2+\sigma^2}{(N+1)\sigma_A^2+\sigma^2}\frac{\sigma_A^2\sigma^2}{N\sigma_A^2+\sigma^2} \\ &= (1-K[N])\text{Bmse}(\hat{A}[N-1]).\end{aligned}$$

- خلاصه:

Estimator Update:

$$\hat{A}[N] = \hat{A}[N-1] + K[N](x[N] - \hat{A}[N-1])$$

where

$$K[N] = \frac{\text{Bmse}(\hat{A}[N-1])}{\text{Bmse}(\hat{A}[N-1]) + \sigma^2}.$$

Minimum MSE Update:

$$\text{Bmse}(\hat{A}[N]) = (1-K[N])\text{Bmse}(\hat{A}[N-1]).$$

LMMSE متوالی

- تعمیم به مدل خطی و حالت برداری: **x=Hθ+w**

Then, the sequential LMMSE estimator becomes (see Appendix 12A)

Estimator Update:

$$\hat{\boldsymbol{\theta}}[n] = \hat{\boldsymbol{\theta}}[n-1] + \mathbf{K}[n](x[n] - \mathbf{h}^T[n]\hat{\boldsymbol{\theta}}[n-1])$$

where

$$\mathbf{K}[n] = \frac{\mathbf{M}[n-1]\mathbf{h}[n]}{\sigma_n^2 + \mathbf{h}^T[n]\mathbf{M}[n-1]\mathbf{h}[n]}.$$

Minimum MSE Matrix Update:

$$\mathbf{M}[n] = (\mathbf{I} - \mathbf{K}[n]\mathbf{h}^T[n])\mathbf{M}[n-1].$$

- ضمناً:

$$\mathbf{M}[n] = E\left[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[n])(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[n])^T\right]. \qquad \mathbf{H}[n] = \begin{bmatrix} \mathbf{H}[n-1] \\ \mathbf{h}^T[n] \end{bmatrix} = \begin{bmatrix} n \times p \\ 1 \times p \end{bmatrix}.$$

LMMSE متوالی

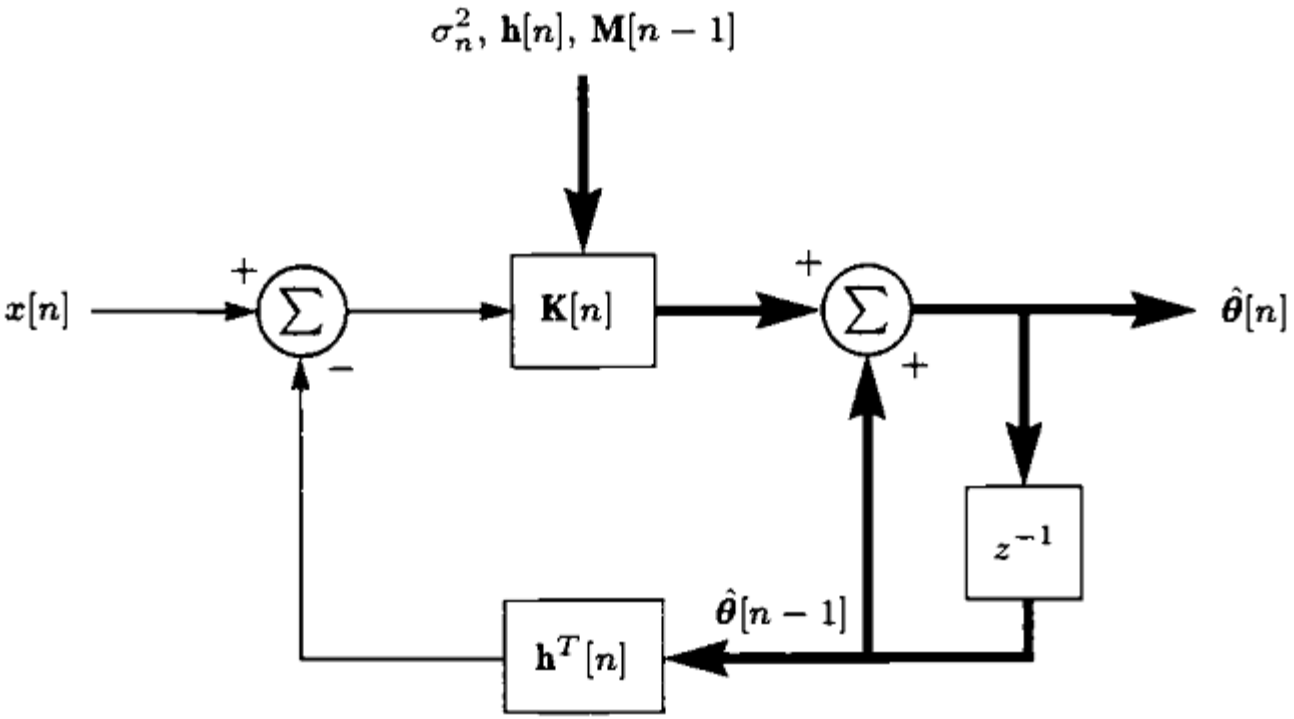


Figure 12.6 Sequential linear minimum mean square estimator

LMMSE متوالی

- مثال۶. تحلیل فوریه بیز متوالی

$$x[n] = a\cos 2\pi f_0 n + b\sin 2\pi f_0 n + w[n] \qquad n = 0, 1, \ldots, N-1$$

- هدف این است که با دریافت متوالی هر کدام از x[0]، x[1] و ...، مقدار تخمین از بردار پارامتر $\boldsymbol{\theta} = [a\, b]^T$ بروز شود.
- فرض می کنیم اطلاعات پیشین ما از بردار پارامتر شامل آمارگان مرتبه اول $E(\boldsymbol{\theta})$ و ماتریس کوواریانس $\sigma_\theta^2\mathbf{I}$ باشد. در این صورت داریم:

$$\begin{aligned} \hat{\boldsymbol{\theta}}[-1] &= E(\boldsymbol{\theta}) \\ \mathbf{M}[-1] &= \mathbf{C}_{\theta\theta} = \sigma_\theta^2\mathbf{I}. \end{aligned}$$

- بنابراین اولین تخمین با استفاده از داده x[0] به دست می آید:

$$\hat{\boldsymbol{\theta}}[0] = \hat{\boldsymbol{\theta}}[-1] + \mathbf{K}[0](x[0] - \mathbf{h}^T[0]\hat{\boldsymbol{\theta}}[-1]).$$

LMMSE متوالی

- مثال۶. ادامه
  - که $\mathbf{h}^T[0]$ اولین سطر از ماتریس مشاهده H است:

$$\mathbf{h}^T[0] = \begin{bmatrix} 1 & 0 \end{bmatrix}.$$

  - سایر سطرهای این ماتریس:

$$\mathbf{h}^T[n] = \begin{bmatrix} \cos 2\pi f_0 n & \sin 2\pi f_0 n \end{bmatrix}.$$

  - ضمنا بردار بهره 2*1 به صورت زیر به دست می آید:

$$\mathbf{K}[0] = \frac{\mathbf{M}[-1]\mathbf{h}[0]}{\sigma^2 + \mathbf{h}^T[0]\mathbf{M}[-1]\mathbf{h}[0]}$$

  - همچنین ماتریس mse جدید:

$$\mathbf{M}[0] = (\mathbf{I} - \mathbf{K}[0]\mathbf{h}^T[0])\mathbf{M}[-1]$$

  - تخمین های مرتبه بالاتر به همین صورت به دست می آیند.