Estimation Theory

Fredrik Rusek

Chapter 11

Chapter 10 – Bayesian Estimation

Section 10.8 Bayesian estimators for deterministic parameters

If no MVU estimator exists, or is very hard to find, we can apply an MMSE estimator to deterministic parameters

Recall the form of the Bayesian estimator for DC-levels in WGN

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

Compute the MSE for a given value of A

Chapter 10 – Bayesian Estimation

Section 10.8 Bayesian estimators for deterministic parameters

If no MVU estimator exists, or is very hard to find, we can apply an MMSE estimator to deterministic parameters

Recall the form of the Bayesian estimator for DC-levels in WGN

$$E(A|\mathbf{x}) = \mu_A + \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N}}(\bar{x} - \mu_A) = \alpha\bar{x} + (1-\alpha)\mu_A \qquad \alpha<1$$

Compute the MSE for a given value of A

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

Chapter 10 – Bayesian Estimation

Section 10.8 Bayesian estimators for deterministic parameters

If no MVU estimator exists, or is very hard to find, we can apply an MMSE estimator to deterministic parameters

Recall the form of the Bayesian estimator for DC-levels in WGN

$$E(A|\mathbf{x}) = \mu_A + \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N}}(\bar{x} - \mu_A) = \alpha\bar{x} + (1-\alpha)\mu_A \qquad \alpha<1$$

Compute the MSE for a given value of A

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

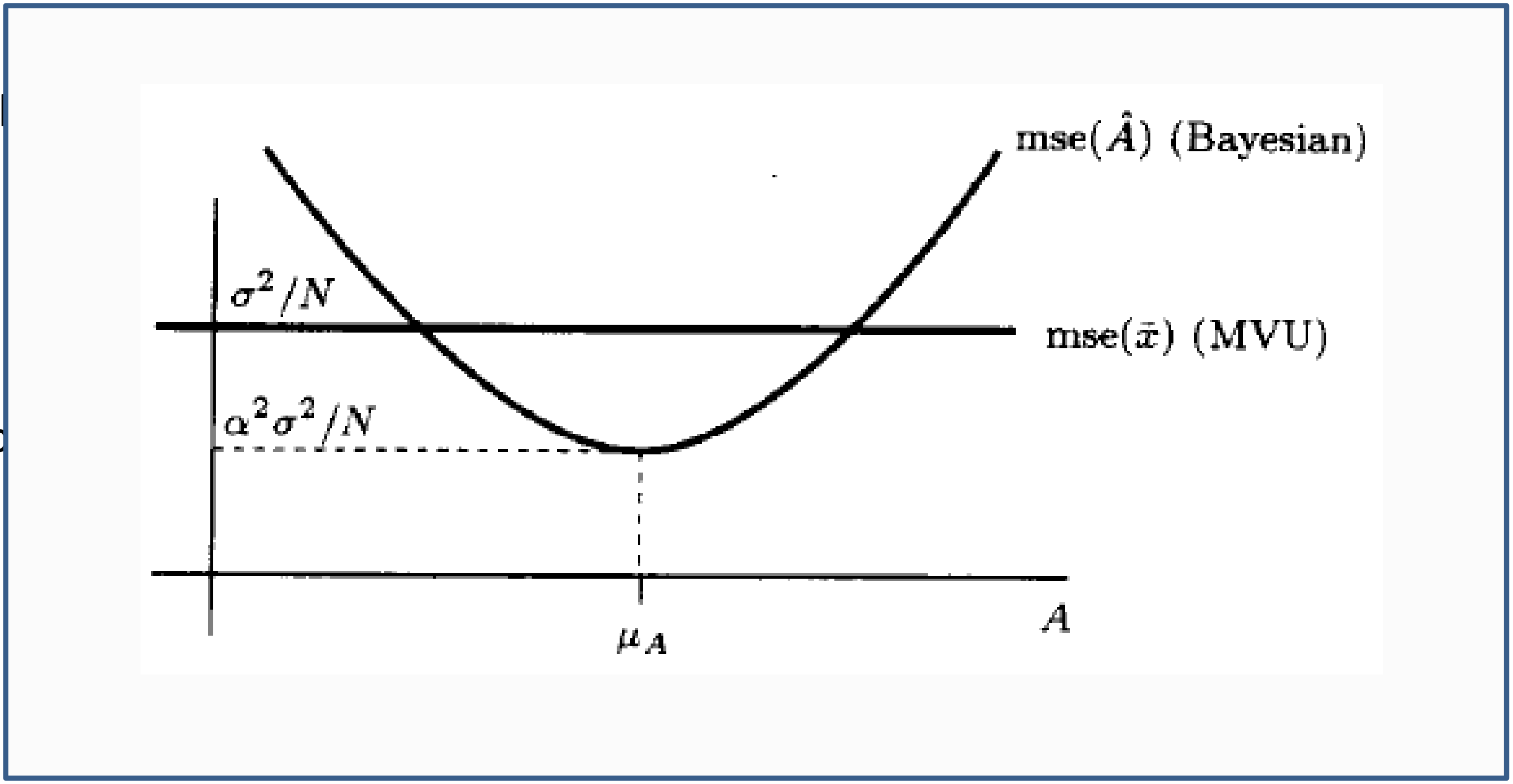
Variance smaller than classical estimator **Large bias for large A**

Chapter 10 – Bayesian Estimation

Section 10.8 Bayesian estimators for deterministic parameters

Recal

Comp



MSE for Bayesian is smaller for A close to the prior mean, but larger far away

Chapter 10 – Bayesian Estimation

Section 10.8 Bayesian estimators for deterministic parameters

However, the BMSE is smaller

$$\mathrm{Bmse}(\hat{A}) \quad = \quad E_A[\mathrm{mse}(\hat{A})]$$

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

Chapter 10 – Bayesian Estimation

Section 10.8 Bayesian estimators for deterministic parameters

However, the BMSE is smaller

$$\begin{aligned}\mathrm{Bmse}(\hat{A}) &= E_A[\mathrm{mse}(\hat{A})] \\ &= \alpha^2\frac{\sigma^2}{N} + (1-\alpha)^2 E_A[(A-\mu_A)^2]\end{aligned}$$

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

Chapter 10 – Bayesian Estimation

Section 10.8 Bayesian estimators for deterministic parameters

However, the BMSE is smaller

$$\begin{aligned}\mathrm{Bmse}(\hat{A}) &= E_A[\mathrm{mse}(\hat{A})] \\ &= \alpha^2 \frac{\sigma^2}{N} + (1-\alpha)^2 E_A[(A-\mu_A)^2] \\ &= \alpha^2 \frac{\sigma^2}{N} + (1-\alpha)^2 \sigma_A^2\end{aligned}$$

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

Chapter 10 – Bayesian Estimation

Section 10.8 Bayesian estimators for deterministic parameters

However, the BMSE is smaller

$$\begin{aligned}
\mathrm{Bmse}(\hat{A}) &= E_A[\mathrm{mse}(\hat{A})] \\
&= \alpha^2 \frac{\sigma^2}{N} + (1-\alpha)^2 E_A[(A-\mu_A)^2] \\
&= \alpha^2 \frac{\sigma^2}{N} + (1-\alpha)^2 \sigma_A^2 \\
&= \frac{\sigma^2}{N} \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N}} \\
&< \frac{\sigma^2}{N} = \mathrm{Bmse}(\bar{x}),
\end{aligned}$$

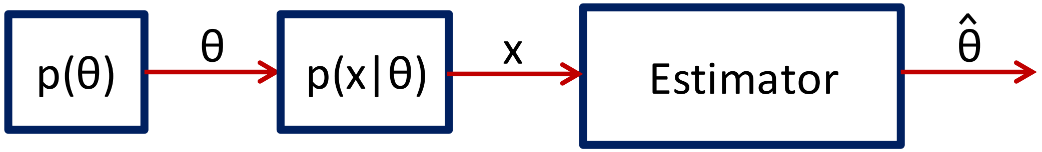
$$\alpha = \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N}}$$

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

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

Chapter 11 – General Bayesian Estimators

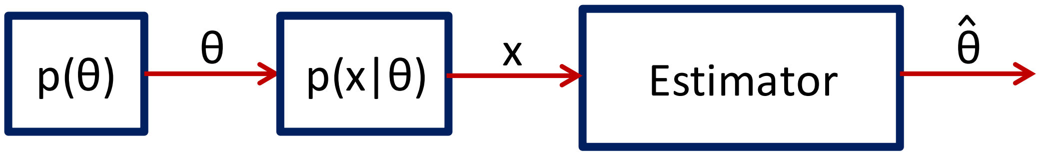
Risk Functions



Error: $\varepsilon = \theta - \hat{\theta}$

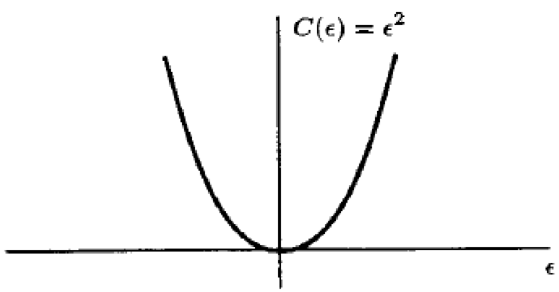
Chapter 11 – General Bayesian Estimators

Risk Functions



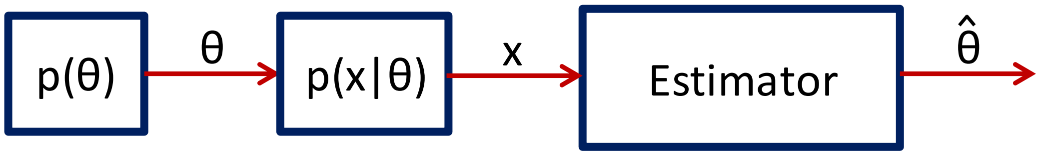
Error: $\varepsilon = \theta - \hat{\theta}$

The MMSE estimator minimizes ***Bayes Risk*** $\mathcal{R} = E[\mathcal{C}(\epsilon)]$ ***where the cost function is*** $\mathcal{C}(\epsilon) = \epsilon^2$



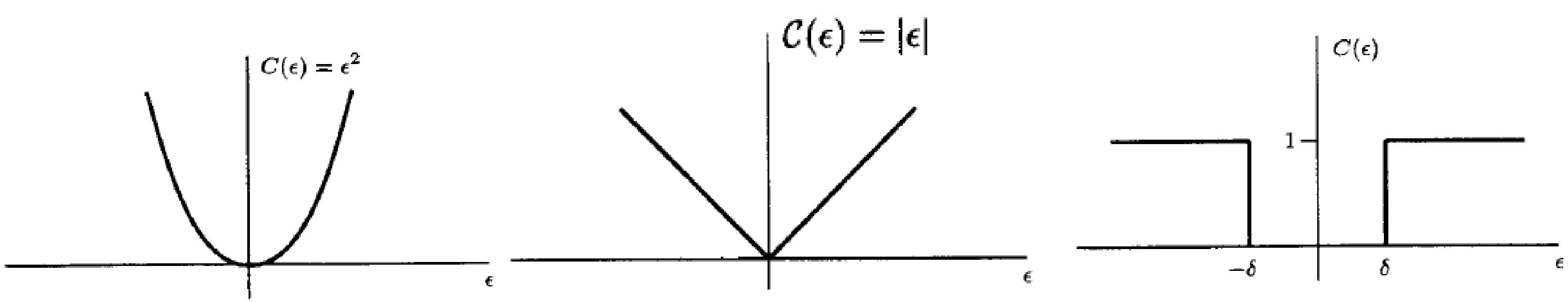
Chapter 11 – General Bayesian Estimators

Risk Functions



Error: $\varepsilon = \theta - \hat{\theta}$

The MMSE estimator minimizes ***Bayes Risk*** $\mathcal{R} = E[\mathcal{C}(\epsilon)]$ ***where the cost function is*** $\mathcal{C}(\epsilon) = \epsilon^2$



$$\mathcal{C}(\epsilon) = \begin{cases} 0 & |\epsilon| < \delta \\ 1 & |\epsilon| > \delta \end{cases}$$

Chapter 11 – General Bayesian Estimators

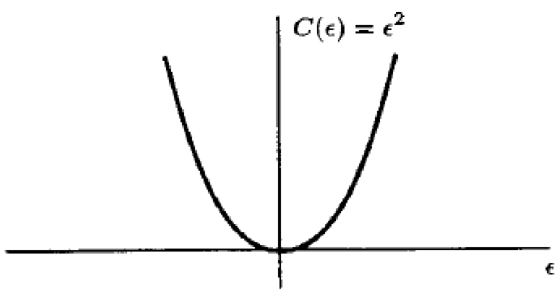
Risk Functions

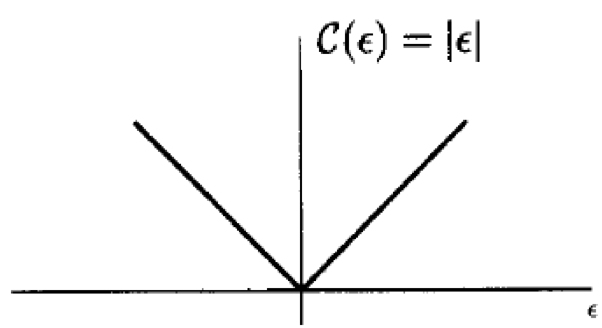
An estimator that minimizez Bayes risk, for some cost, is termed a Bayes estimator

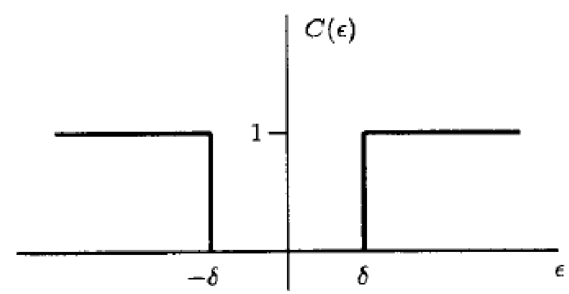
$$p(\theta) \xrightarrow{\theta} p(x|\theta) \xrightarrow{x} \boxed{\text{Estimator}} \xrightarrow{\hat{\theta}}$$

Error: $\varepsilon = \theta - \hat{\theta}$

The MMSE estimator minimizes ***Bayes Risk*** $\mathcal{R} = E[\mathcal{C}(\epsilon)]$ ***where the cost function is*** $\mathcal{C}(\epsilon) = \epsilon^2$



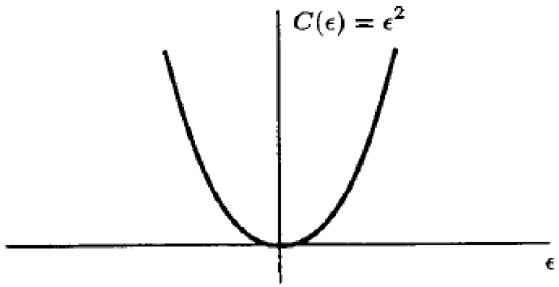




$$\mathcal{C}(\epsilon) = \begin{cases} 0 & |\epsilon| < \delta \\ 1 & |\epsilon| > \delta \end{cases}$$

Chapter 11 – General Bayesian Estimators

Let us now optimize $\mathcal{R} = E[\mathcal{C}(\epsilon)]$ **for different** $\mathcal{C}(\epsilon)$



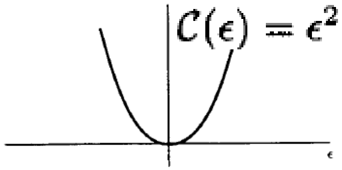
For a quadratic cost, we already know that $\hat{\theta} = E(\theta|\mathbf{x})$

Chapter 11 – General Bayesian Estimators

Let us now optimize $\mathcal{R} = E[\mathcal{C}(\epsilon)]$ **for different** $\mathcal{C}(\epsilon)$

Bayes risk equals

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



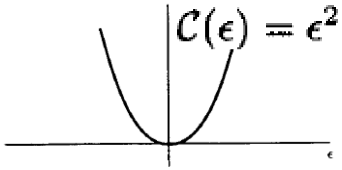
$$\hat{\theta} = E(\theta|\mathbf{x})$$

Chapter 11 – General Bayesian Estimators

Let us now optimize $\mathcal{R} = E[\mathcal{C}(\epsilon)]$ **for different** $\mathcal{C}(\epsilon)$

Bayes risk equals

$$\begin{aligned}\mathcal{R} &= E[\mathcal{C}(\epsilon)] \\ &= \int\int \mathcal{C}(\theta - \hat{\theta}) p(\mathbf{x}, \theta)\, d\mathbf{x}\, d\theta \\ &= \int \underbrace{\left[\int \mathcal{C}(\theta - \hat{\theta}) p(\theta|\mathbf{x})\, d\theta\right]}_{\text{Minimize this to minimize Bayes risk}} p(\mathbf{x})\, d\mathbf{x}.\end{aligned}$$



$$\hat{\theta} = E(\theta|\mathbf{x})$$

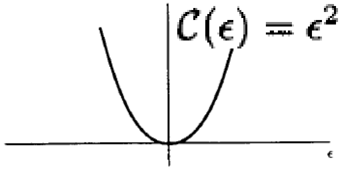
Chapter 11 – General Bayesian Estimators

Let us now optimize $\mathcal{R} = E[\mathcal{C}(\epsilon)]$ **for different** $\mathcal{C}(\epsilon)$

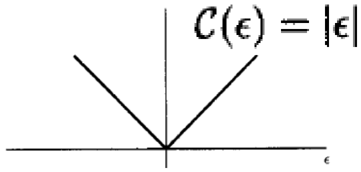
Bayes risk equals

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

$$g(\hat{\theta}) = \int |\theta - \hat{\theta}|p(\theta|\mathbf{x})\,d\theta$$



$$\hat{\theta} = E(\theta|\mathbf{x})$$



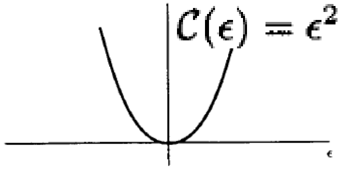
Chapter 11 – General Bayesian Estimators

Let us now optimize $\mathcal{R} = E[\mathcal{C}(\epsilon)]$ **for different** $\mathcal{C}(\epsilon)$

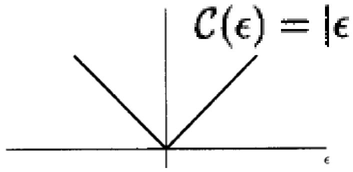
Bayes risk equals

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

$$\begin{aligned}
g(\hat{\theta}) &= \int |\theta - \hat{\theta}| p(\theta|\mathbf{x})\, d\theta \\
&= \int_{-\infty}^{\hat{\theta}} (\hat{\theta} - \theta) p(\theta|\mathbf{x})\, d\theta + \int_{\hat{\theta}}^{\infty} (\theta - \hat{\theta}) p(\theta|\mathbf{x})\, d\theta
\end{aligned}$$



$\hat{\theta} = E(\theta|\mathbf{x})$



Chapter 11 – General Bayesian Estimators

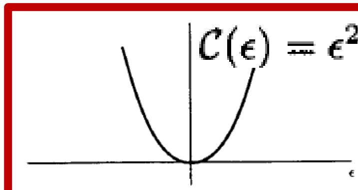
Let us now optimize $\mathcal{R} = E[\mathcal{C}(\epsilon)]$ **for different** $\mathcal{C}(\epsilon)$

Bayes risk equals

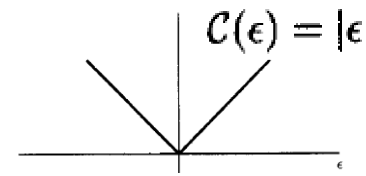
$$\begin{aligned}\mathcal{R} &= E[\mathcal{C}(\epsilon)] \\ &= \int\int \mathcal{C}(\theta-\hat{\theta})p(\mathbf{x},\theta)\,d\mathbf{x}\,d\theta \\ &= \int\left[\int \mathcal{C}(\theta-\hat{\theta})p(\theta|\mathbf{x})\,d\theta\right]p(\mathbf{x})\,d\mathbf{x}.\end{aligned}$$

$$\begin{aligned}g(\hat{\theta}) &= \int |\theta-\hat{\theta}|p(\theta|\mathbf{x})\,d\theta \\ &= \int_{-\infty}^{\hat{\theta}}(\hat{\theta}-\theta)p(\theta|\mathbf{x})\,d\theta + \int_{\hat{\theta}}^{\infty}(\theta-\hat{\theta})p(\theta|\mathbf{x})\,d\theta\end{aligned}$$

We need $\frac{dg(\hat{\theta})}{d\hat{\theta}}$ **, but the limits of the integral depends on** $\hat{\theta}$ **Not standard differential**



$\hat{\theta} = E(\theta|\mathbf{x})$



Chapter 11 – General Bayesian Estimators

Interlude: Leibnitz's rule (very useful)

$$\boxed{\begin{aligned}\frac{\partial}{\partial u}\int_{\phi_1(u)}^{\phi_2(u)} h(u,v)\,dv &= \int_{\phi_1(u)}^{\phi_2(u)} \frac{\partial h(u,v)}{\partial u}\,dv + \frac{d\phi_2(u)}{du}h(u,\phi_2(u)) \\ &\quad - \frac{d\phi_1(u)}{du}h(u,\phi_1(u)).\end{aligned}}$$

Chapter 11 – General Bayesian Estimators

Leibnitz's rule (very useful)

$$\boxed{\begin{aligned}\frac{\partial}{\partial u}\int_{\phi_1(u)}^{\phi_2(u)} h(u,v)\,dv \quad &= \quad \int_{\phi_1(u)}^{\phi_2(u)} \frac{\partial h(u,v)}{\partial u}\,dv + \frac{d\phi_2(u)}{du}h(u,\phi_2(u)) \\ &\qquad - \frac{d\phi_1(u)}{du}h(u,\phi_1(u)).\end{aligned}}$$

We have:

$$\frac{d}{d\hat{\theta}} \int_{-\infty}^{\hat{\theta}} (\hat{\theta}-\theta)p(\theta|\mathbf{x})\,d\theta + \int_{\hat{\theta}}^{\infty} (\theta-\hat{\theta})p(\theta|\mathbf{x})\,d\theta$$

$$h(\hat{\theta},\theta) = (\hat{\theta}-\theta)p(\theta|\mathbf{x})$$

Chapter 11 – General Bayesian Estimators

Leibnitz's rule (very useful)

$$\frac{\partial}{\partial u}\int_{\phi_1(u)}^{\phi_2(u)} h(u,v)\,dv = \int_{\phi_1(u)}^{\phi_2(u)} \frac{\partial h(u,v)}{\partial u}\,dv + \boxed{\frac{d\phi_2(u)}{du}h(u,\phi_2(u))} - \frac{d\phi_1(u)}{du}h(u,\phi_1(u)).$$

$$\frac{d}{d\hat{\theta}}\boxed{\int_{-\infty}^{\hat{\theta}}(\hat{\theta}-\theta)p(\theta|\mathbf{x})\,d\theta} + \int_{\hat{\theta}}^{\infty}(\theta-\hat{\theta})p(\theta|\mathbf{x})\,d\theta$$

$$h(u,\phi_2(u)) =$$

$$h(\hat{\theta},\theta) = (\hat{\theta}-\theta)p(\theta|\mathbf{x})$$

Chapter 11 – General Bayesian Estimators

Leibnitz's rule (very useful)

$$\frac{\partial}{\partial u}\int_{\phi_1(u)}^{\phi_2(u)} h(u,v)\,dv = \int_{\phi_1(u)}^{\phi_2(u)} \frac{\partial h(u,v)}{\partial u}\,dv + \boxed{\frac{d\phi_2(u)}{du}h(u,\phi_2(u))} - \frac{d\phi_1(u)}{du}h(u,\phi_1(u)).$$

$$\frac{d}{d\hat{\theta}}\boxed{\int_{-\infty}^{\hat{\theta}}(\hat{\theta}-\theta)p(\theta|\mathbf{x})\,d\theta} + \int_{\hat{\theta}}^{\infty}(\theta-\hat{\theta})p(\theta|\mathbf{x})\,d\theta$$

$$h(u,\phi_2(u)) = \begin{cases} u = \hat{\theta} \\ \phi_2(u) = \hat{\theta} \end{cases}$$

$$h(\hat{\theta},\theta) = (\hat{\theta}-\theta)p(\theta|\mathbf{x})$$

Chapter 11 – General Bayesian Estimators

Leibnitz's rule (very useful)

$$\frac{\partial}{\partial u}\int_{\phi_1(u)}^{\phi_2(u)} h(u,v)\,dv = \int_{\phi_1(u)}^{\phi_2(u)} \frac{\partial h(u,v)}{\partial u}\,dv + \boxed{\frac{d\phi_2(u)}{du}h(u,\phi_2(u))}$$
$$- \frac{d\phi_1(u)}{du}h(u,\phi_1(u)).$$

$$\frac{d}{d\hat{\theta}}\boxed{\int_{-\infty}^{\hat{\theta}}(\hat{\theta}-\theta)p(\theta|\mathbf{x})\,d\theta} + \int_{\hat{\theta}}^{\infty}(\theta-\hat{\theta})p(\theta|\mathbf{x})\,d\theta$$

$$h(u,\phi_2(u)) = h(\hat{\theta},\hat{\theta}) =$$

$$h(\hat{\theta},\theta) = (\hat{\theta}-\theta)p(\theta|\mathbf{x})$$

Chapter 11 – General Bayesian Estimators

Leibnitz's rule (very useful)

$$\frac{\partial}{\partial u}\int_{\phi_1(u)}^{\phi_2(u)} h(u,v)\,dv = \int_{\phi_1(u)}^{\phi_2(u)} \frac{\partial h(u,v)}{\partial u}\,dv + \boxed{\frac{d\phi_2(u)}{du}h(u,\phi_2(u))} - \frac{d\phi_1(u)}{du}h(u,\phi_1(u)).$$

$$\frac{d}{d\hat{\theta}}\boxed{\int_{-\infty}^{\hat{\theta}}(\hat{\theta}-\theta)p(\theta|\mathbf{x})\,d\theta} + \int_{\hat{\theta}}^{\infty}(\theta-\hat{\theta})p(\theta|\mathbf{x})\,d\theta$$

$$h(u,\phi_2(u)) = h(\hat{\theta},\hat{\theta}) = (\hat{\theta}-\hat{\theta})p(\hat{\theta}|\mathbf{x}) = 0$$

$$h(\hat{\theta},\theta) = (\hat{\theta}-\theta)p(\theta|\mathbf{x})$$

Chapter 11 – General Bayesian Estimators

Leibnitz's rule (very useful)

$$\frac{\partial}{\partial u}\int_{\phi_1(u)}^{\phi_2(u)} h(u,v)\,dv = \int_{\phi_1(u)}^{\phi_2(u)} \frac{\partial h(u,v)}{\partial u}\,dv + \cancel{\frac{d\phi_2(u)}{du}h(u,\phi_2(u))}$$
$$-\boxed{\frac{d\phi_1(u)}{du}h(u,\phi_1(u)).}$$

$$\frac{d}{d\hat{\theta}}\boxed{\int_{-\infty}^{\hat{\theta}}(\hat{\theta}-\theta)p(\theta|\mathbf{x})\,d\theta} + \int_{\hat{\theta}}^{\infty}(\theta-\hat{\theta})p(\theta|\mathbf{x})\,d\theta$$

$$h(u,\phi_2(u)) = h(\hat{\theta},\hat{\theta}) = (\hat{\theta}-\hat{\theta})p(\hat{\theta}|\mathbf{x}) = 0 \qquad \frac{d\phi_1(u)}{du} =$$

$$h(\hat{\theta},\theta) = (\hat{\theta}-\theta)p(\theta|\mathbf{x})$$

Chapter 11 – General Bayesian Estimators

Leibnitz's rule (very useful)

$$\frac{\partial}{\partial u}\int_{\phi_1(u)}^{\phi_2(u)} h(u,v)\,dv = \int_{\phi_1(u)}^{\phi_2(u)} \frac{\partial h(u,v)}{\partial u}\,dv + \cancel{\frac{d\phi_2(u)}{du}h(u,\phi_2(u))} - \boxed{\frac{d\phi_1(u)}{du}h(u,\phi_1(u)).}$$

$$\frac{d}{d\hat{\theta}}\boxed{\int_{-\infty}^{\hat{\theta}}(\hat{\theta}-\theta)p(\theta|\mathbf{x})\,d\theta} + \int_{\hat{\theta}}^{\infty}(\theta-\hat{\theta})p(\theta|\mathbf{x})\,d\theta$$

Lower limit does not depend on u:

$u = \hat{\theta}$

$$h(u,\phi_2(u)) = h(\hat{\theta},\hat{\theta}) = (\hat{\theta}-\hat{\theta})p(\hat{\theta}|\mathbf{x}) = 0 \qquad \frac{d\phi_1(u)}{du} = 0$$

$$h(\hat{\theta},\theta) = (\hat{\theta}-\theta)p(\theta|\mathbf{x})$$

Chapter 11 – General Bayesian Estimators

Leibnitz's rule (very useful)

$$\frac{\partial}{\partial u}\int_{\phi_1(u)}^{\phi_2(u)} h(u,v)\,dv = \boxed{\int_{\phi_1(u)}^{\phi_2(u)} \frac{\partial h(u,v)}{\partial u}\,dv} + \cancel{\frac{d\phi_2(u)}{du} h(u,\phi_2(u))} - \cancel{\frac{d\phi_1(u)}{du} h(u,\phi_1(u))}.$$

$$\frac{d}{d\hat{\theta}} \boxed{\int_{-\infty}^{\hat{\theta}} (\hat{\theta}-\theta)p(\theta|\mathbf{x})\,d\theta} + \int_{\hat{\theta}}^{\infty} (\theta-\hat{\theta})p(\theta|\mathbf{x})\,d\theta$$

$$\boxed{\frac{\partial h(u,v)}{\partial u}}$$

$$h(\hat{\theta},\theta) = (\hat{\theta}-\theta)p(\theta|\mathbf{x})$$

Chapter 11 – General Bayesian Estimators

Leibnitz's rule (very useful)

$$\frac{\partial}{\partial u}\int_{\phi_1(u)}^{\phi_2(u)} h(u,v)\,dv = \boxed{\int_{\phi_1(u)}^{\phi_2(u)} \frac{\partial h(u,v)}{\partial u}\,dv} + \cancel{\frac{d\phi_2(u)}{du} h(u,\phi_2(u))}$$

$$- \cancel{\frac{d\phi_1(u)}{du} h(u,\phi_1(u))}.$$

$$\frac{d}{d\hat{\theta}} \boxed{\int_{-\infty}^{\hat{\theta}} (\hat{\theta}-\theta)p(\theta|\mathbf{x})\,d\theta} + \int_{\hat{\theta}}^{\infty} (\theta-\hat{\theta})p(\theta|\mathbf{x})\,d\theta$$

$$\boxed{\frac{\partial h(u,v)}{\partial u}} \qquad \frac{d}{d\hat{\theta}}h(\hat{\theta},\theta) = \frac{d}{d\hat{\theta}}(\hat{\theta}-\theta)p(\theta|\mathbf{x}) = p(\theta|\mathbf{x})$$

$$h(\hat{\theta},\theta) = (\hat{\theta}-\theta)p(\theta|\mathbf{x})$$

Chapter 11 – General Bayesian Estimators

Leibnitz's rule (very useful)

$$\frac{\partial}{\partial u}\int_{\phi_1(u)}^{\phi_2(u)} h(u,v)\,dv = \boxed{\int_{\phi_1(u)}^{\phi_2(u)} \frac{\partial h(u,v)}{\partial u}\,dv} + \cancel{\frac{d\phi_2(u)}{du} h(u,\phi_2(u))}$$
$$- \cancel{\frac{d\phi_1(u)}{du} h(u,\phi_1(u))}.$$

$$\frac{d}{d\hat{\theta}} \int_{-\infty}^{\hat{\theta}} (\hat{\theta}-\theta)p(\theta|\mathbf{x})\,d\theta + \int_{\hat{\theta}}^{\infty} (\theta-\hat{\theta})p(\theta|\mathbf{x})\,d\theta$$
$$= \int_{-\infty}^{\hat{\theta}} p(\theta|\mathbf{x})\,d\theta - \int_{\hat{\theta}}^{\infty} p(\theta|\mathbf{x})\,d\theta$$

Chapter 11 – General Bayesian Estimators

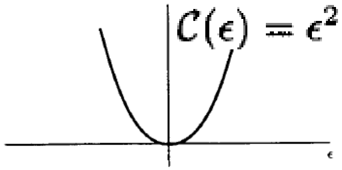
Let us now optimize $\mathcal{R} = E[\mathcal{C}(\epsilon)]$ **for different** $\mathcal{C}(\epsilon)$

Bayes risk equals

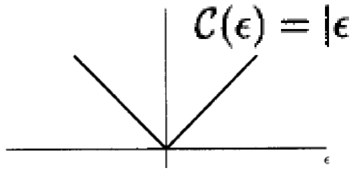
$$
\begin{aligned}
\mathcal{R} &= E[\mathcal{C}(\epsilon)] \\
&= \int\int \mathcal{C}(\theta - \hat{\theta}) p(\mathbf{x}, \theta)\, d\mathbf{x}\, d\theta \\
&= \int \left[\int \mathcal{C}(\theta - \hat{\theta}) p(\theta|\mathbf{x})\, d\theta \right] p(\mathbf{x})\, d\mathbf{x}.
\end{aligned}
$$

$$
\begin{aligned}
g(\hat{\theta}) &= \int |\theta - \hat{\theta}| p(\theta|\mathbf{x})\, d\theta \\
&= \int_{-\infty}^{\hat{\theta}} (\hat{\theta} - \theta) p(\theta|\mathbf{x})\, d\theta + \int_{\hat{\theta}}^{\infty} (\theta - \hat{\theta}) p(\theta|\mathbf{x})\, d\theta
\end{aligned}
$$

We need $\dfrac{dg(\hat{\theta})}{d\hat{\theta}}$ **, but the limits of the integral depends on** $\hat{\theta}$ **Not standard differential**



$\hat{\theta} = E(\theta|\mathbf{x})$



Chapter 11 – General Bayesian Estimators

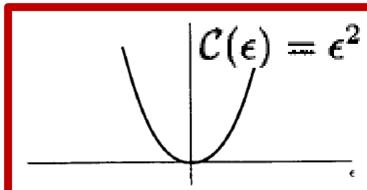
Let us now optimize $\mathcal{R} = E[\mathcal{C}(\epsilon)]$ **for different** $\mathcal{C}(\epsilon)$

Bayes risk equals

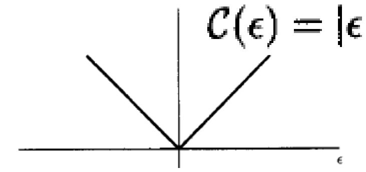
$$\begin{aligned}\mathcal{R} &= E[\mathcal{C}(\epsilon)] \\ &= \int\int \mathcal{C}(\theta - \hat{\theta})p(\mathbf{x}, \theta)\, d\mathbf{x}\, d\theta \\ &= \int \left[\int \mathcal{C}(\theta - \hat{\theta})p(\theta|\mathbf{x})\, d\theta\right] p(\mathbf{x})\, d\mathbf{x}.\end{aligned}$$

$$\begin{aligned}g(\hat{\theta}) &= \int |\theta - \hat{\theta}| p(\theta|\mathbf{x})\, d\theta \\ &= \int_{-\infty}^{\hat{\theta}} (\hat{\theta} - \theta)p(\theta|\mathbf{x})\, d\theta + \int_{\hat{\theta}}^{\infty} (\theta - \hat{\theta})p(\theta|\mathbf{x})\, d\theta\end{aligned}$$

$$\frac{dg(\hat{\theta})}{d\hat{\theta}} = \int_{-\infty}^{\hat{\theta}} p(\theta|\mathbf{x})\, d\theta - \int_{\hat{\theta}}^{\infty} p(\theta|\mathbf{x})\, d\theta = 0$$



$\hat{\theta} = E(\theta|\mathbf{x})$



Chapter 11 – General Bayesian Estimators

Let us now optimize $\mathcal{R} = E[\mathcal{C}(\epsilon)]$ **for different** $\mathcal{C}(\epsilon)$

Bayes risk equals

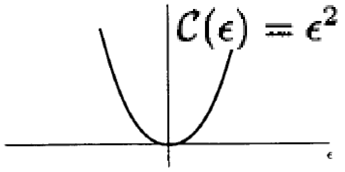
$$\begin{aligned}\mathcal{R} &= E[\mathcal{C}(\epsilon)] \\ &= \int\int \mathcal{C}(\theta - \hat{\theta})p(\mathbf{x}, \theta)\, d\mathbf{x}\, d\theta \\ &= \int \left[\int \mathcal{C}(\theta - \hat{\theta})p(\theta|\mathbf{x})\, d\theta\right] p(\mathbf{x})\, d\mathbf{x}.\end{aligned}$$

$$\begin{aligned}g(\hat{\theta}) &= \int |\theta - \hat{\theta}|p(\theta|\mathbf{x})\, d\theta \\ &= \int_{-\infty}^{\hat{\theta}} (\hat{\theta} - \theta)p(\theta|\mathbf{x})\, d\theta + \int_{\hat{\theta}}^{\infty} (\theta - \hat{\theta})p(\theta|\mathbf{x})\, d\theta\end{aligned}$$

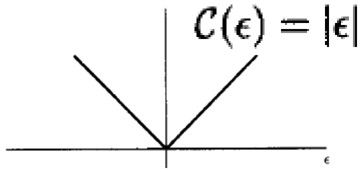
$$\frac{dg(\hat{\theta})}{d\hat{\theta}} = \int_{-\infty}^{\hat{\theta}} p(\theta|\mathbf{x})\, d\theta - \int_{\hat{\theta}}^{\infty} p(\theta|\mathbf{x})\, d\theta = 0$$

$$\int_{-\infty}^{\hat{\theta}} p(\theta|\mathbf{x})\, d\theta = \int_{\hat{\theta}}^{\infty} p(\theta|\mathbf{x})\, d\theta$$

$\hat{\theta}$ is the median of the posterior



$\hat{\theta} = E(\theta|\mathbf{x})$



Chapter 11 – General Bayesian Estimators

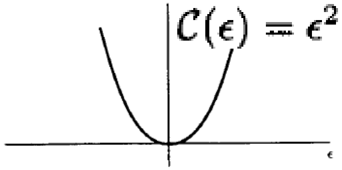
Let us now optimize $\mathcal{R} = E[\mathcal{C}(\epsilon)]$ **for different** $\mathcal{C}(\epsilon)$

Bayes risk equals

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

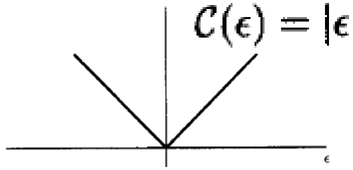
$$
g(\hat{\theta}) = \int_{-\infty}^{\hat{\theta}-\delta} 1 \cdot p(\theta|\mathbf{x})\, d\theta + \int_{\hat{\theta}+\delta}^{\infty} 1 \cdot p(\theta|\mathbf{x})\, d\theta
$$

$\mathcal{C}(\epsilon) = \epsilon^2$



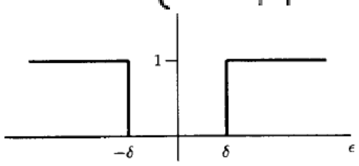
$\hat{\theta} = E(\theta|\mathbf{x})$

$\mathcal{C}(\epsilon) = |\epsilon|$



$\hat{\theta}$ = median

$$
\mathcal{C}(\epsilon) = \begin{cases} 0 & |\epsilon| < \delta \\ 1 & |\epsilon| > \delta \end{cases}
$$



Chapter 11 – General Bayesian Estimators

Let us now optimize $\mathcal{R} = E[\mathcal{C}(\epsilon)]$ **for different** $\mathcal{C}(\epsilon)$

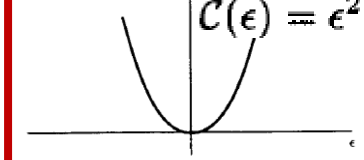
Bayes risk equals

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

$$g(\hat{\theta}) = \int_{-\infty}^{\hat{\theta}-\delta} 1 \cdot p(\theta|\mathbf{x})\, d\theta + \int_{\hat{\theta}+\delta}^{\infty} 1 \cdot p(\theta|\mathbf{x})\, d\theta$$

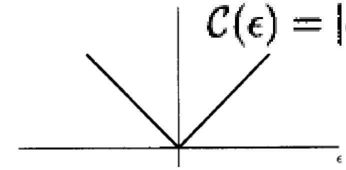
$$= 1 - \int_{\hat{\theta}-\delta}^{\hat{\theta}+\delta} p(\theta|\mathbf{x})\, d\theta.$$

$\mathcal{C}(\epsilon) = \epsilon^2$



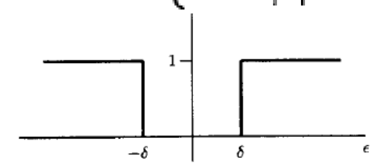
$\hat{\theta} = E(\theta|\mathbf{x})$

$\mathcal{C}(\epsilon) = |\epsilon|$



$\hat{\theta}$ = median

$$\mathcal{C}(\epsilon) = \begin{cases} 0 & |\epsilon| < \delta \\ 1 & |\epsilon| > \delta \end{cases}$$



Chapter 11 – General Bayesian Estimators

Let us now optimize $\mathcal{R} = E[\mathcal{C}(\epsilon)]$ **for different** $\mathcal{C}(\epsilon)$

Bayes risk equals

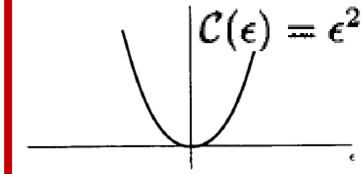
$$\begin{aligned}\mathcal{R} &= E[\mathcal{C}(\epsilon)] \\ &= \int\int \mathcal{C}(\theta-\hat{\theta})p(\mathbf{x},\theta)\,d\mathbf{x}\,d\theta \\ &= \int\left[\int \mathcal{C}(\theta-\hat{\theta})p(\theta|\mathbf{x})\,d\theta\right]p(\mathbf{x})\,d\mathbf{x}.\end{aligned}$$

$$g(\hat{\theta}) = \int_{-\infty}^{\hat{\theta}-\delta} 1\cdot p(\theta|\mathbf{x})\,d\theta + \int_{\hat{\theta}+\delta}^{\infty} 1\cdot p(\theta|\mathbf{x})\,d\theta$$

$$= 1 - \int_{\hat{\theta}-\delta}^{\hat{\theta}+\delta} p(\theta|\mathbf{x})\,d\theta.$$

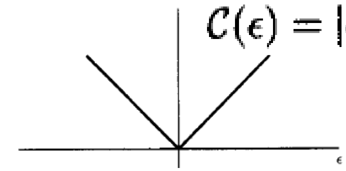
$\hat{\theta}$ **= arg max** $\int_{\hat{\theta}-\delta}^{\hat{\theta}+\delta} p(\theta|\mathbf{x})\,d\theta.$

$\mathcal{C}(\epsilon) = \epsilon^2$



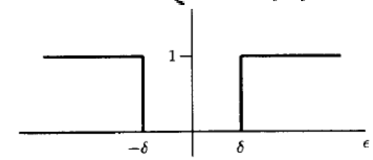
$\hat{\theta} = E(\theta|\mathbf{x})$

$\mathcal{C}(\epsilon) = |\epsilon|$



$\hat{\theta}$ **= median**

$$\mathcal{C}(\epsilon) = \begin{cases} 0 & |\epsilon| < \delta \\ 1 & |\epsilon| > \delta \end{cases}$$



Chapter 11 – General Bayesian Estimators

Let us now optimize $\mathcal{R} = E[\mathcal{C}(\epsilon)]$ **for different** $\mathcal{C}(\epsilon)$

Bayes risk equals

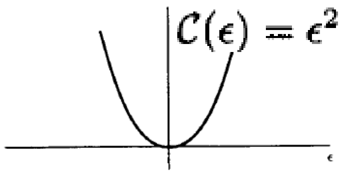
$$\begin{aligned}\mathcal{R} &= E[\mathcal{C}(\epsilon)] \\ &= \int\int \mathcal{C}(\theta-\hat{\theta})p(\mathbf{x},\theta)\,d\mathbf{x}\,d\theta \\ &= \int\left[\int \mathcal{C}(\theta-\hat{\theta})p(\theta|\mathbf{x})\,d\theta\right]p(\mathbf{x})\,d\mathbf{x}.\end{aligned}$$

$$g(\hat{\theta}) = \int_{-\infty}^{\hat{\theta}-\delta} 1\cdot p(\theta|\mathbf{x})\,d\theta + \int_{\hat{\theta}+\delta}^{\infty} 1\cdot p(\theta|\mathbf{x})\,d\theta$$

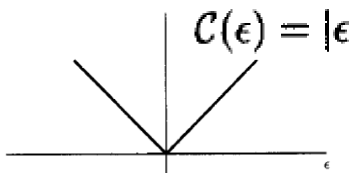
$$= 1 - \int_{\hat{\theta}-\delta}^{\hat{\theta}+\delta} p(\theta|\mathbf{x})\,d\theta.$$

$\hat{\theta}$ **= arg max** $\int_{\hat{\theta}-\delta}^{\hat{\theta}+\delta} p(\theta|\mathbf{x})\,d\theta.$

Let δ->0: $\hat{\theta}$ **= arg max** $p(\theta|\mathbf{x})$ **(maximum a posterori (MAP))**

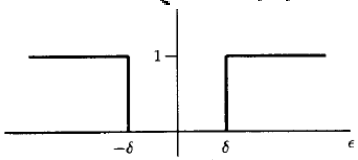


$\hat{\theta} = E(\theta|\mathbf{x})$



$\hat{\theta}$ **= median**

$$\mathcal{C}(\epsilon) = \begin{cases} 0 & |\epsilon| < \delta \\ 1 & |\epsilon| > \delta \end{cases}$$

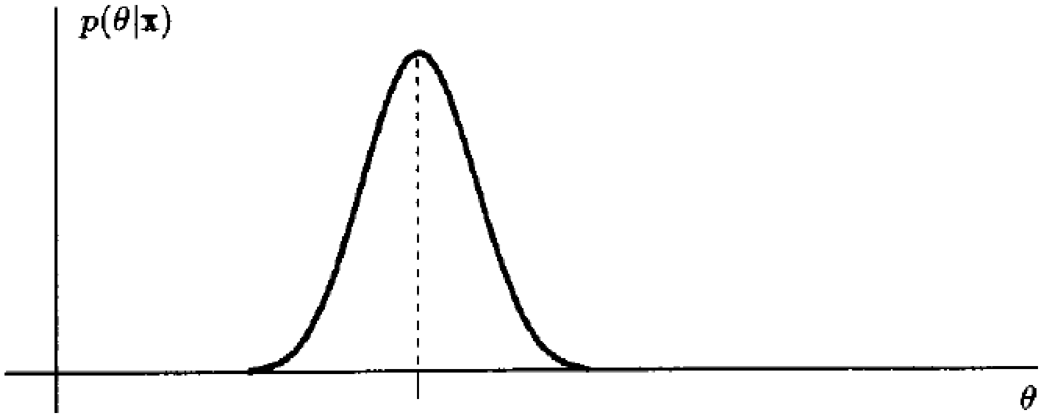


$\hat{\theta}$ **= arg max** $\int_{\hat{\theta}-\delta}^{\hat{\theta}+\delta} p(\theta|\mathbf{x})\,d\theta.$

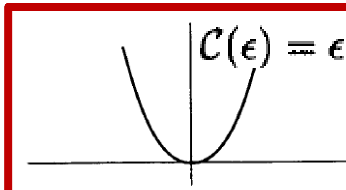
Chapter 11 – General Bayesian Estimators

Gausian posterior

What is relation between mean, median and max ?

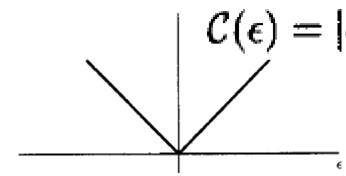


$\mathcal{C}(\epsilon) = \epsilon^2$



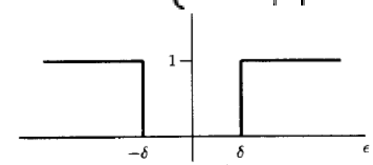
$\hat{\theta} = E(\theta|\mathbf{x})$

$\mathcal{C}(\epsilon) = |\epsilon|$



$\hat{\theta}$ = median

$$\mathcal{C}(\epsilon) = \begin{cases} 0 & |\epsilon| < \delta \\ 1 & |\epsilon| > \delta \end{cases}$$

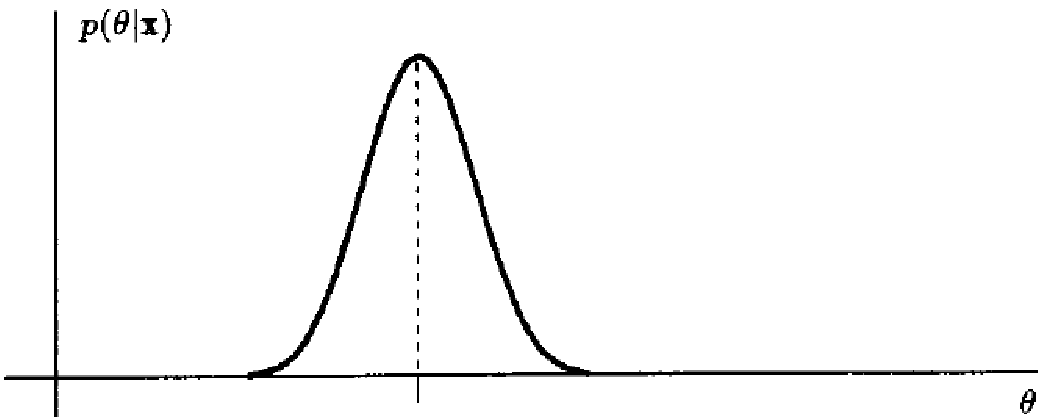


$\hat{\theta}$ = arg max $\int_{\hat{\theta}-\delta}^{\hat{\theta}+\delta} p(\theta|\mathbf{x})\,d\theta$

Chapter 11 – General Bayesian Estimators

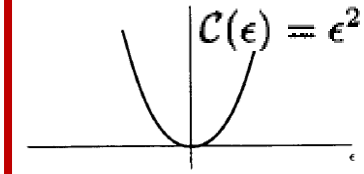
Gausian posterior

What is relation between mean, median and max ?



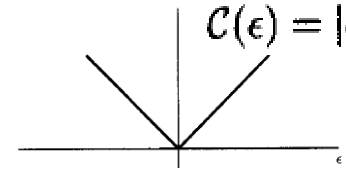
Gaussian posterior makes the three risk functions identical

$\mathcal{C}(\epsilon) = \epsilon^2$



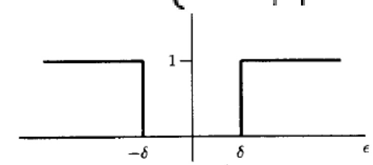
$\hat{\theta} = E(\theta|\mathbf{x})$

$\mathcal{C}(\epsilon) = |\epsilon|$



$\hat{\theta}$ = median

$$\mathcal{C}(\epsilon) = \begin{cases} 0 & |\epsilon| < \delta \\ 1 & |\epsilon| > \delta \end{cases}$$



$\hat{\theta}$ = arg max $\int_{\hat{\theta}-\delta}^{\hat{\theta}+\delta} p(\theta|\mathbf{x})\,d\theta$

Chapter 11 – General Bayesian Estimators

Extension to vector parameter

Suppose we have a vector parameter of unknowns $\boldsymbol{\theta}$

Consider estimation of θ_1. It still holds that the MAP estimator uses $p(\theta_1|\mathbf{x})$

Chapter 11 – General Bayesian Estimators

Extension to vector parameter

Suppose we have a vector parameter of unknowns $\boldsymbol{\theta}$

Consider estimation of θ_1. It still holds that the MAP estimator uses $p(\theta_1|\mathbf{x})$

The parameters θ_2 …. θ_N are nuisance parameters, but we can integrate them away

$$p(\theta_1|\mathbf{x}) = \int \cdots \int p(\boldsymbol{\theta}|\mathbf{x})\, d\theta_2 \cdots d\theta_p$$

Chapter 11 – General Bayesian Estimators

Extension to vector parameter

Suppose we have a vector parameter of unknowns **θ**

Consider estimation of θ_1. It still holds that the MAP estimator uses $p(\theta_1|\mathbf{x})$

The parameters θ_2 …. θ_N are nuisance parameters, but we can integrate them away

$$p(\theta_1|\mathbf{x}) = \int \cdots \int p(\boldsymbol{\theta}|\mathbf{x})\, d\theta_2 \cdots d\theta_p$$

The estimator is

$$\boxed{\hat{\theta}_1 = \int \theta_1 p(\theta_1|\mathbf{x})\, d\theta_1}$$

Chapter 11 – General Bayesian Estimators

Extension to vector parameter

Suppose we have a vector parameter of unknowns $\boldsymbol{\theta}$

Consider estimation of θ_1. It still holds that the MAP estimator uses $p(\theta_1|\mathbf{x})$

The parameters θ_2 …. θ_N are nuisance parameters, but we can integrate them away

$$p(\theta_1|\mathbf{x}) = \int \cdots \int p(\boldsymbol{\theta}|\mathbf{x})\, d\theta_2 \cdots d\theta_p$$

The estimator is

$$\boxed{\begin{aligned} \hat{\theta}_1 &= \int \theta_1 p(\theta_1|\mathbf{x})\, d\theta_1 \\ &= \int \theta_1 \left[\int \cdots \int p(\boldsymbol{\theta}|\mathbf{x})\, d\theta_2 \ldots d\theta_p\right] d\theta_1 \end{aligned}}$$

Chapter 11 – General Bayesian Estimators

Extension to vector parameter

Suppose we have a vector parameter of unknowns **θ**

Consider estimation of θ_1. It still holds that the MAP estimator uses $p(\theta_1|\mathbf{x})$

The parameters θ_2 …. θ_N are nuisance parameters, but we can integrate them away

$$p(\theta_1|\mathbf{x}) = \int \cdots \int p(\boldsymbol{\theta}|\mathbf{x})\, d\theta_2 \cdots d\theta_p$$

The estimator is

$$\boxed{\begin{aligned} \hat{\theta}_1 &= \int \theta_1 p(\theta_1|\mathbf{x})\, d\theta_1 \\ &= \int \theta_1 \left[\int \cdots \int p(\boldsymbol{\theta}|\mathbf{x})\, d\theta_2 \ldots d\theta_p\right] d\theta_1 \\ &= \int \theta_1 p(\boldsymbol{\theta}|\mathbf{x})\, d\boldsymbol{\theta} \end{aligned}}$$

Chapter 11 – General Bayesian Estimators

Extension to vector parameter

In vector form

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} \int \theta_1 p(\boldsymbol{\theta}|\mathbf{x})\, d\boldsymbol{\theta} \\ \int \theta_2 p(\boldsymbol{\theta}|\mathbf{x})\, d\boldsymbol{\theta} \\ \vdots \\ \int \theta_p p(\boldsymbol{\theta}|\mathbf{x})\, d\boldsymbol{\theta} \end{bmatrix} = E(\boldsymbol{\theta}|\mathbf{x})$$

$$\begin{aligned} \hat{\theta}_1 &= \int \theta_1 p(\theta_1|\mathbf{x})\, d\theta_1 \\ &= \int \theta_1 \left[\int \cdots \int p(\boldsymbol{\theta}|\mathbf{x})\, d\theta_2 \ldots d\theta_p \right] d\theta_1 \\ &= \int \theta_1 p(\boldsymbol{\theta}|\mathbf{x})\, d\boldsymbol{\theta} \end{aligned}$$

Chapter 11 – General Bayesian Estimators

Extension to vector parameter

Observations

Classical approach (non-Bayesian):
We must estimate all unknown paramters jointly, except if.....what holds????

Chapter 11 – General Bayesian Estimators

Extension to vector parameter

Observations

Classical approach (non-Bayesian):
We must estimate all unknown paramters jointly, except if Fisher information is diagonal

Vector MMSE estimator minimizes the MSE for each component of the unknown vector parameter **θ**, i.e.,

$$[\hat{\boldsymbol{\theta}}]_i = [E(\boldsymbol{\theta}|\mathbf{x})]_i \text{ minimizes } E[(\theta_i - \hat{\theta}_i)^2]$$

Chapter 11 – General Bayesian Estimators

Performance of MMSE estimator

$$\mathrm{Bmse}(\hat{\theta}_1) = E[(\theta_1 - \hat{\theta}_1)^2]$$

Chapter 11 – General Bayesian Estimators

Performance of MMSE estimator

$$\mathrm{Bmse}(\hat{\theta}_1) = E[(\theta_1 - \hat{\theta}_1)^2]$$

$$= \int (\theta_1 - \hat{\theta}_1)^2 p(\mathbf{x}, \theta_1)\, d\theta_1 d\mathbf{x}$$

Function of x

Chapter 11 – General Bayesian Estimators

Performance of MMSE estimator

$$\begin{aligned}\mathrm{Bmse}(\hat{\theta}_1) &= E[(\theta_1 - \hat{\theta}_1)^2] \\ &= \int (\theta_1 - \hat{\theta}_1)^2 p(\mathbf{x}, \theta_1)\, d\theta_1 d\mathbf{x} \\ &= \int \left[\int (\theta_1 - E(\theta_1|\mathbf{x}))^2\, p(\theta_1|\mathbf{x})\, d\theta_1\right] p(\mathbf{x})\, d\mathbf{x}\end{aligned}$$

Bayes rule

MMSE estimator

Chapter 11 – General Bayesian Estimators

Performance of MMSE estimator

$$\begin{aligned}
\mathrm{Bmse}(\hat{\theta}_1) &= E[(\theta_1 - \hat{\theta}_1)^2] \\
&= \int (\theta_1 - \hat{\theta}_1)^2 p(\mathbf{x}, \theta_1)\, d\theta_1 d\mathbf{x} \\
&= \int \left[\int (\theta_1 - E(\theta_1|\mathbf{x}))^2 \, p(\theta_1|\mathbf{x})\, d\theta_1 \right] p(\mathbf{x})\, d\mathbf{x} \\
&= \int \mathrm{var}(\theta_1|\mathbf{x}) p(\mathbf{x})\, d\mathbf{x}.
\end{aligned}$$

By definition

Chapter 11 – General Bayesian Estimators

Performance of MMSE estimator

$$
\begin{aligned}
\mathrm{Bmse}(\hat{\theta}_1) &= E[(\theta_1 - \hat{\theta}_1)^2] \\
&= \int (\theta_1 - \hat{\theta}_1)^2 p(\mathbf{x}, \theta_1)\, d\theta_1 d\mathbf{x} \\
&= \int \left[\int (\theta_1 - E(\theta_1|\mathbf{x}))^2 p(\theta_1|\mathbf{x})\, d\theta_1 \right] p(\mathbf{x})\, d\mathbf{x} \\
&= \int \mathrm{var}(\theta_1|\mathbf{x}) p(\mathbf{x})\, d\mathbf{x}. \\
&= \int \left[\int (\theta_1 - E(\theta_1|\mathbf{x}))^2 p(\boldsymbol{\theta}|\mathbf{x})\, d\boldsymbol{\theta} \right] p(\mathbf{x})\, d\mathbf{x}.
\end{aligned}
$$

$$
p(\theta_1|\mathbf{x}) = \int \cdots \int p(\boldsymbol{\theta}|\mathbf{x})\, d\theta_2 \ldots d\theta_p
$$

Chapter 11 – General Bayesian Estimators

Performance of MMSE estimator

$$\mathrm{Bmse}(\hat{\theta}_1) = E[(\theta_1 - \hat{\theta}_1)^2]$$

$$= \int (\theta_1 - \hat{\theta}_1)^2 p(\mathbf{x}, \theta_1)\, d\theta_1 d\mathbf{x}$$

$$= \int \left[\int (\theta_1 - E(\theta_1|\mathbf{x}))^2 p(\theta_1|\mathbf{x})\, d\theta_1\right] p(\mathbf{x})\, d\mathbf{x}$$

$$= \int \mathrm{var}(\theta_1|\mathbf{x}) p(\mathbf{x})\, d\mathbf{x}.$$

$$= \int \underbrace{\left[\int (\theta_1 - E(\theta_1|\mathbf{x}))^2 p(\boldsymbol{\theta}|\mathbf{x})\, d\boldsymbol{\theta}\right]} p(\mathbf{x})\, d\mathbf{x}.$$

Element [1,1] of

$$\boxed{\mathrm{Bmse}(\hat{\theta}_i) = \int [\mathbf{C}_{\theta|x}]_{ii} p(\mathbf{x})\, d\mathbf{x}}$$

$$\mathbf{C}_{\theta|x} = E_{\theta|x}\left[(\boldsymbol{\theta} - E(\boldsymbol{\theta}|\mathbf{x}))(\boldsymbol{\theta} - E(\boldsymbol{\theta}|\mathbf{x}))^T\right]$$

Chapter 11 – General Bayesian Estimators

Additive property

Independent observations $\mathbf{x}_1, \mathbf{x}_2$
Estimate $\boldsymbol{\theta}$
Assume that $\mathbf{x}_1, \mathbf{x}_2$, $\boldsymbol{\theta}$ are jointly Gaussian

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

Theorem 10.2

Chapter 11 – General Bayesian Estimators

Additive property

Independent observations $\mathbf{x}_1,\mathbf{x}_2$
Estimate $\boldsymbol{\theta}$
Assume that $\mathbf{x}_1,\mathbf{x}_2$, $\boldsymbol{\theta}$ are jointly Gaussian

Typo in book, should include means as well

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

$$\begin{aligned}\mathbf{C}_{xx}^{-1} &= \begin{bmatrix} \mathbf{C}_{x_1x_1} & \mathbf{C}_{x_1x_2} \\ \mathbf{C}_{x_2x_1} & \mathbf{C}_{x_2x_2} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \mathbf{C}_{x_1x_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{x_2x_2} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \mathbf{C}_{x_1x_1}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{x_2x_2}^{-1} \end{bmatrix}\end{aligned}$$

$$\begin{aligned}\mathbf{C}_{\theta x} &= E\left[\boldsymbol{\theta}\begin{bmatrix}\mathbf{x}_1 \\ \mathbf{x}_2\end{bmatrix}^T\right] = \\ &= \begin{bmatrix}\mathbf{C}_{\theta x_1} & \mathbf{C}_{\theta x_2}\end{bmatrix}\end{aligned}$$

Independent observations

Chapter 11 – General Bayesian Estimators

Additive property

Independent observations $\mathbf{x}_1,\mathbf{x}_2$
Estimate $\boldsymbol{\theta}$
Assume that $\mathbf{x}_1,\mathbf{x}_2$, $\boldsymbol{\theta}$ are jointly Gaussian

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

$$\boxed{\begin{aligned}\hat{\boldsymbol{\theta}} &= E(\boldsymbol{\theta}) + \left[\begin{array}{cc}\mathbf{C}_{\theta x_1} & \mathbf{C}_{\theta x_2}\end{array}\right]\left[\begin{array}{cc}\mathbf{C}_{x_1x_1}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{x_2x_2}^{-1}\end{array}\right]\left[\begin{array}{c}\mathbf{x}_1 - E(\mathbf{x}_1) \\ \mathbf{x}_2 - E(\mathbf{x}_2)\end{array}\right] \\ &= E(\boldsymbol{\theta}) + \mathbf{C}_{\theta x_1}\mathbf{C}_{x_1x_1}^{-1}(\mathbf{x}_1 - E(\mathbf{x}_1)) + \mathbf{C}_{\theta x_2}\mathbf{C}_{x_2x_2}^{-1}(\mathbf{x}_2 - E(\mathbf{x}_2)).\end{aligned}}$$

MMSE estimate can be updated sequentially !!!

Chapter 11 – General Bayesian Estimators

MAP estimator

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

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

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

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

Chapter 11 – General Bayesian Estimators

MAP estimator

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

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

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

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

Benefits compared with MMSE

Not needed (typically hard to find)

Optimization generally easier than finding the conditional expectation

Chapter 11 – General Bayesian Estimators

MAP vs ML estimator

Alexander Aljechin (1882-1946) became world chess champion 1927 (by defeating Capablanca)

Aljechin defended his title twice, and regained it once

Chapter 11 – General Bayesian Estimators

MAP vs ML estimator

Alexander Aljechin (1882-1946) became world chess champion 1927 (by defeating Capablanca)

Aljechin defended his title twice, and regained it once

Magnus Calrsen became world champion 2013, and defended the title Once in 2014

Chapter 11 – General Bayesian Estimators

MAP vs ML estimator

Alexander Aljechin (1882-1946) became world chess champion 1927 (by defeating Capablanca)

Aljechin defended his title twice, and regained it once

Magnus Calrsen became world champion 2013, and defended the title Once in 2014

Now consider a title game in 2015. Observe Y=y1, where y1=win
Two hypotheses:

- H1: Aljechin defends title
- H2: Carlsen defends title

Chapter 11 – General Bayesian Estimators

MAP vs ML estimator

Alexander Aljechin (1882-1946) became world chess champion 1927 (by defeating Capablanca)

Aljechin defended his title twice, and regained it once

Magnus Calrsen became world champion 2013, and defended the title Once in 2014

Now consider a title game in 2015. Observe Y=y1, where y1=win
Two hypotheses:

- H1: Aljechin defends title
- H2: Carlsen defends title

Given the above statistics

$$f(y1|H1)>f(y1|H2)$$

Chapter 11 – General Bayesian Estimators

MAP vs ML estimator

Alexander Aljechin (1882-1946) became world chess champion 1927 (by defeating Capablanca)

Aljechin defended his title twice, and regained it once

Magnus Calrsen became world champion 2013, and defended the title Once in 2014

Now consider a title game in 2015. Observe Y=y1, where y1=win
Two hypotheses:

- **H1: Aljechin defends title**
- **H2: Carlsen defends title**

Given the above statistics

$$f(y1|H1)>f(y1|H2)$$

ML rule: Aljechin takes title (although he died in 1946)

Chapter 11 – General Bayesian Estimators

MAP vs ML estimator

Alexander Aljechin (1882-1946) became world chess champion 1927 (by defeating Capablanca)

Aljechin defended his title twice, and regained it once

Magnus Calrsen became world champion 2013, and defended the title Once in 2014

Now consider a title game in 2015. Observe $Y=y1$, where $y1$=win
Two hypotheses:

- H1: Aljechin defends title
- H2: Carlsen defends title

Given the above statistics

$$f(y1|H1)>f(y1|H2)$$

MAP rule: $f(H1)=0$, -> Carlsen defends title

Chapter 11 – General Bayesian Estimators

Example DC-level in white noise, uniform prior U[-A_0,A_0]

The posterior is

$$p(A|\mathbf{x}) = \begin{cases} \dfrac{\dfrac{1}{\sqrt{2\pi\sigma^2/N}} \exp\left[-\dfrac{1}{2\frac{\sigma^2}{N}}(A-\bar{x})^2\right]}{\displaystyle\int_{-A_0}^{A_0} \dfrac{1}{\sqrt{2\pi\sigma^2/N}} \exp\left[-\dfrac{1}{2\frac{\sigma^2}{N}}(A-\bar{x})^2\right] dA} & |A| \leq A_0 \\ 0 & |A| > A_0 \end{cases}$$

We got stuck here:
Cannot put the denominator in closed form
Cannot integrate the nominator

Lets try with the MAP estimator

Chapter 11 – General Bayesian Estimators

Example DC-level in white noise, uniform prior U[-A_0,A_0]

The posterior is

$$p(A|\mathbf{x}) = \begin{cases} \dfrac{\dfrac{1}{\sqrt{2\pi\sigma^2/N}} \exp\left[-\dfrac{1}{2\frac{\sigma^2}{N}}(A-\bar{x})^2\right]}{\displaystyle\int_{-A_0}^{A_0} \dfrac{1}{\sqrt{2\pi\sigma^2/N}} \exp\left[-\dfrac{1}{2\frac{\sigma^2}{N}}(A-\bar{x})^2\right] dA} & |A| \leq A_0 \\ 0 & |A| > A_0 \end{cases}$$

Denominator: Does not depend on A -> irrelevant

Chapter 11 – General Bayesian Estimators

Example DC-level in white noise, uniform prior U[-A_0,A_0]

The posterior is

$$p(A|\mathbf{x}) = \begin{cases} \dfrac{\dfrac{1}{\sqrt{2\pi\sigma^2/N}}\exp\left[-\dfrac{1}{2\frac{\sigma^2}{N}}(A-\bar{x})^2\right]}{\displaystyle\int_{-A_0}^{A_0}\dfrac{1}{\sqrt{2\pi\sigma^2/N}}\exp\left[-\dfrac{1}{2\frac{\sigma^2}{N}}(A-\bar{x})^2\right]dA} & |A| \le A_0 \\ 0 & |A| > A_0 \end{cases}$$

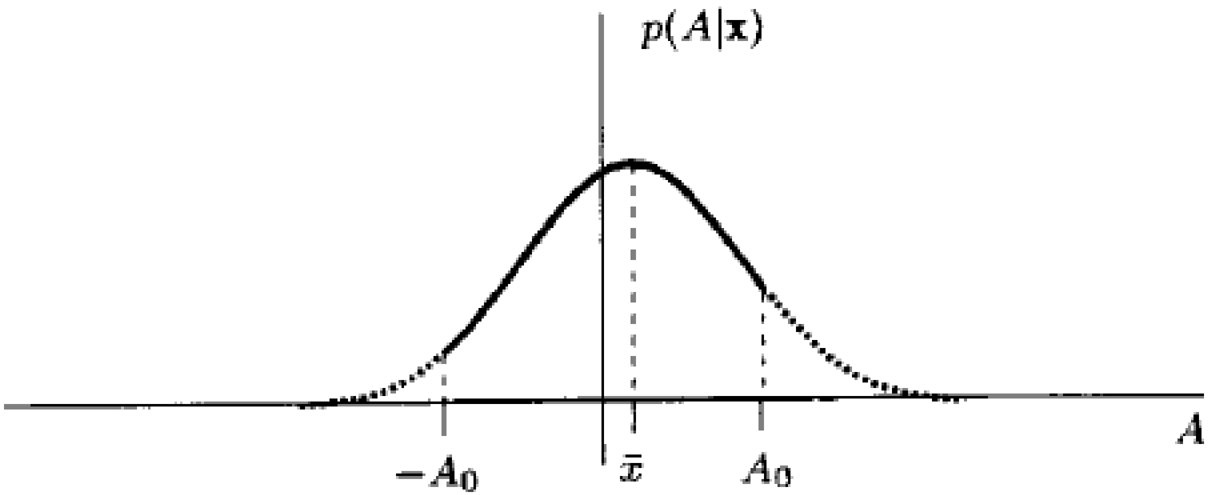
Denominator: Does not depend on A -> irrelevant
We need to maximize the nominator

Chapter 11 – General Bayesian Estimators

Example DC-level in white noise, uniform prior U[-A_0,A_0]

$$p(A|\mathbf{x}) = \begin{cases} \dfrac{\dfrac{1}{\sqrt{2\pi\sigma^2/N}} \exp\left[-\dfrac{1}{2\frac{\sigma^2}{N}}(A-\bar{x})^2\right]}{\displaystyle\int_{-A_0}^{A_0} \dfrac{1}{\sqrt{2\pi\sigma^2/N}} \exp\left[-\dfrac{1}{2\frac{\sigma^2}{N}}(A-\bar{x})^2\right] dA} & |A| \leq A_0 \\ 0 & |A| > A_0 \end{cases}$$

$$\hat{A} = \begin{cases} \bar{x} & -A_0 \leq \bar{x} \leq A_0 \end{cases}$$



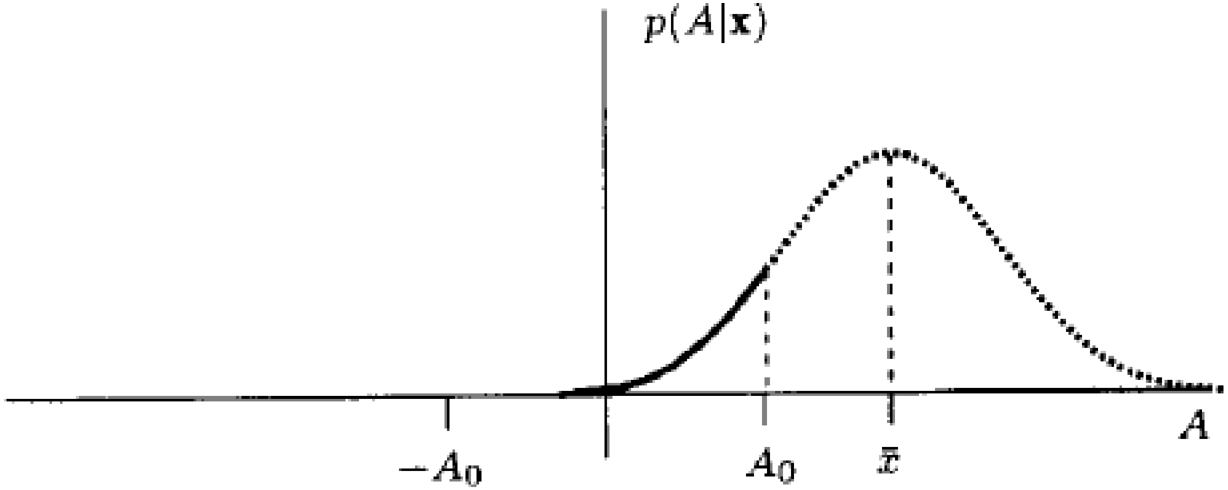
(a) $-A_0 \leq \bar{x} \leq A_0$

Chapter 11 – General Bayesian Estimators

Example DC-level in white noise, uniform prior U[-A_0,A_0]

$$p(A|\mathbf{x}) = \begin{cases} \dfrac{\dfrac{1}{\sqrt{2\pi\sigma^2/N}} \exp\left[-\dfrac{1}{2\frac{\sigma^2}{N}}(A-\bar{x})^2\right]}{\displaystyle\int_{-A_0}^{A_0} \dfrac{1}{\sqrt{2\pi\sigma^2/N}} \exp\left[-\dfrac{1}{2\frac{\sigma^2}{N}}(A-\bar{x})^2\right] dA} & |A| \leq A_0 \\ 0 & |A| > A_0 \end{cases}$$

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



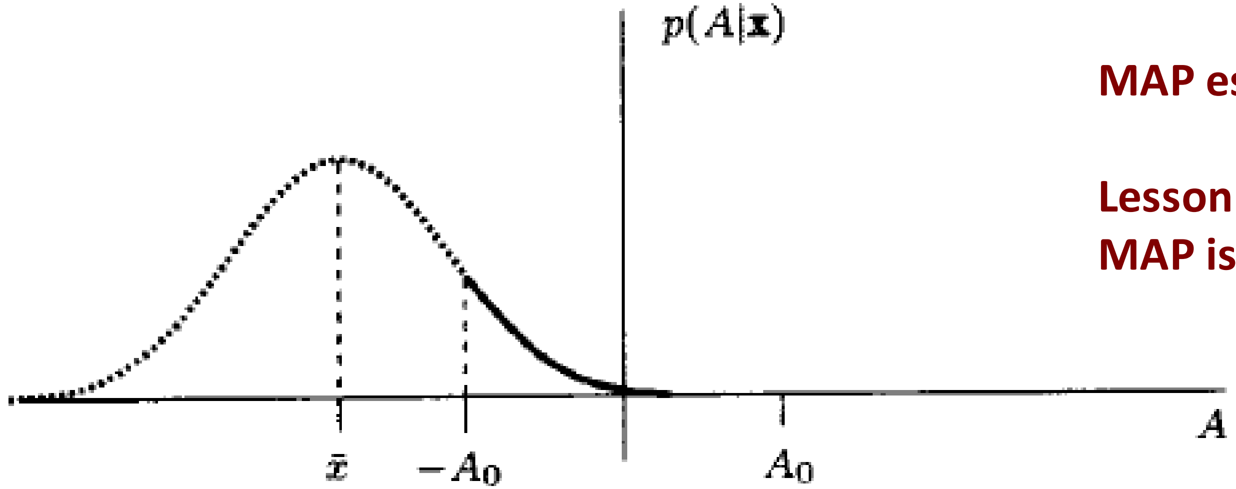
(b) $\bar{x} > A_0$

Chapter 11 – General Bayesian Estimators

Example DC-level in white noise, uniform prior U[-A_0,A_0]

$$p(A|\mathbf{x}) = \begin{cases} \dfrac{\dfrac{1}{\sqrt{2\pi\sigma^2/N}}\exp\left[-\dfrac{1}{2\frac{\sigma^2}{N}}(A-\bar{x})^2\right]}{\displaystyle\int_{-A_0}^{A_0}\dfrac{1}{\sqrt{2\pi\sigma^2/N}}\exp\left[-\dfrac{1}{2\frac{\sigma^2}{N}}(A-\bar{x})^2\right]dA} & |A| \le A_0 \\ 0 & |A| > A_0 \end{cases}$$

$$\hat{A} = \begin{cases} -A_0 & \bar{x} < -A_0 \\ \bar{x} & -A_0 \le \bar{x} \le A_0 \\ A_0 & \bar{x} > A_0. \end{cases}$$



(c) $\bar{x} < -A_0$

MAP estimator can be found!

Lesson learned (generally true)
MAP is easier to find than MMSE

Chapter 11 – General Bayesian Estimators

Element-wise MAP for vector-valued parameter

$$p(\theta_1|\mathbf{x}) = \int \cdots \int p(\boldsymbol{\theta}|\mathbf{x})\, d\theta_2 \ldots d\theta_p.$$

$$\hat{\theta}_1 = \arg\max_{\theta_1} p(\theta_1|\mathbf{x})$$

"No-integration-needed" benefit gone

Chapter 11 – General Bayesian Estimators

Element-wise MAP for vector-valued parameter

$$p(\theta_1|\mathbf{x}) = \int \cdots \int p(\boldsymbol{\theta}|\mathbf{x})\, d\theta_2 \ldots d\theta_p$$

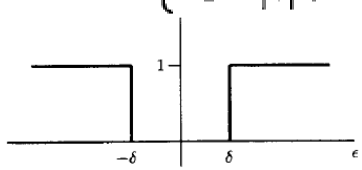
$$\hat{\theta}_1 = \arg\max_{\theta_1} p(\theta_1|\mathbf{x})$$

"No-integration-needed" benefit gone

The estimator $\hat{\theta}_i = \arg\max_{\theta_i} p(\theta_i|\mathbf{x})$

Minimizes the "hit-or-miss" risk $\mathcal{R}_i = E[\mathcal{C}(\theta_i - \hat{\theta}_i)]$ for each I, where δ->0

$$\mathcal{C}(\epsilon) = \begin{cases} 0 & |\epsilon| < \delta \\ 1 & |\epsilon| > \delta \end{cases}$$



Chapter 11 – General Bayesian Estimators

Element-wise MAP for vector-valued parameter

$$\boxed{\begin{aligned} p(\theta_1|\mathbf{x}) &= \int \cdots \int p(\boldsymbol{\theta}|\mathbf{x})\, d\theta_2 \ldots d\theta_p \\ \hat{\theta}_1 &= \arg\max_{\theta_1} p(\theta_1|\mathbf{x}) \end{aligned}}$$

"No-integration-needed" benefit gone

Let us now define another risk function

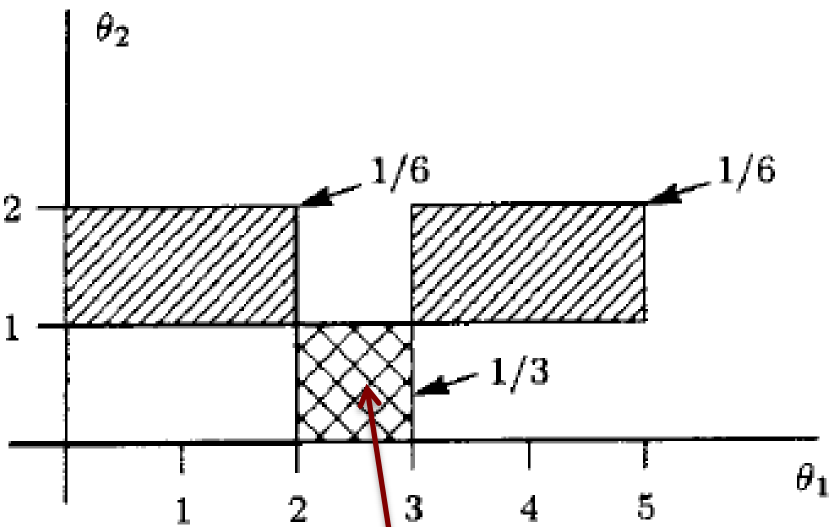
$$\mathcal{C}(\epsilon) = \begin{cases} 1 & ||\epsilon|| > \delta \\ 0 & ||\epsilon|| < \delta \end{cases} \qquad \boldsymbol{\epsilon} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$$

Easy to prove that as δ->0, Bayes risk is minimized by the vector-MAP-estimator

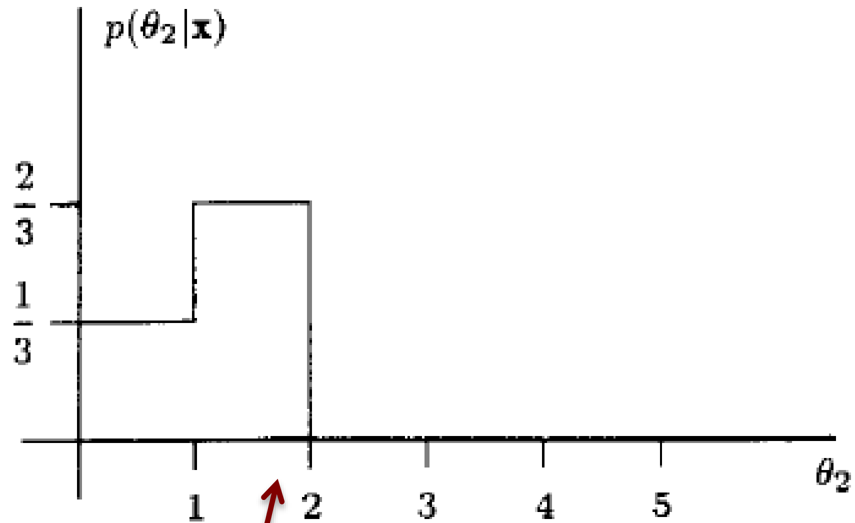
$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\mathbf{x})$$

Chapter 11 – General Bayesian Estimators

Element-wise MAP and vector valued MAP are not the same



(a) Posterior PDF $p(\theta_1, \theta_2|\mathbf{x})$



(b) Posterior PDF $p(\theta_2|\mathbf{x})$

Vector-valued MAP solution

Element-wise MAP solution

Chapter 11 – General Bayesian Estimators

Two properties of vector-MAP

- For jointly Gaussian $\mathbf{x}$ and $\boldsymbol{\theta}$, the conditional mean $E(\boldsymbol{\theta}|\mathbf{x})$ coincides with the peak of $p(\boldsymbol{\theta}|\mathbf{x})$. **Hence, the vector-MAP and the MMSE coincide.**

Chapter 11 – General Bayesian Estimators

Two properties of vector-MAP

- For jointly Gaussian $\mathbf{x}$ and $\boldsymbol{\theta}$, the conditional mean $E(\boldsymbol{\theta}|\mathbf{x})$ coincides with the peak of $p(\boldsymbol{\theta}|\mathbf{x})$. **Hence, the vector-MAP and the MMSE coincide.**

- **Invariance does *not* hold for MAP (as opposed to MLE)**

$$\hat{\alpha} = g(\hat{\theta})$$

Chapter 11 – General Bayesian Estimators

Invariance

Why does invariance hold for MLE?

With $\alpha = g(\theta)$, it holds that $p(x|\alpha) = p_\theta(x|g^{-1}(\alpha))$

Chapter 11 – General Bayesian Estimators

Invariance

Why does invariance hold for MLE?

With $\alpha=g(\theta)$, it holds that $p(x|\alpha) = p_\theta(x|g^{-1}(\alpha))$

However, MAP involves the prior, and it doesn't hold that $p_\alpha(\alpha)=p_\theta(g^{-1}(\alpha))$, since the two distributions are related through the Jacobian

Chapter 11 – General Bayesian Estimators

Example

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

Exponential

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

Inverse gamma

Chapter 11 – General Bayesian Estimators

Example

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

Exponential

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

Inverse gamma

MAP

$$g(\theta) = \ln p(\mathbf{x}|\theta) + \ln p(\theta)$$

Chapter 11 – General Bayesian Estimators

Example

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

Exponential

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

Inverse gamma

MAP

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

Chapter 11 – General Bayesian Estimators

Example

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

Exponential

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

Inverse gamma

MAP

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

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

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

Chapter 11 – General Bayesian Estimators

Example

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

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

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

Consider estimation of

$$\alpha = 1/\theta.$$

$$\hat{\alpha} = \frac{1}{\hat{\theta}} = \bar{x} + \frac{\lambda}{N}$$ **? (holds for MLE)**

Chapter 11 – General Bayesian Estimators

Example

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

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

$$p(x[n]|\alpha) = \begin{cases} \frac{1}{\alpha} \exp\left(-\frac{x[n]}{\alpha}\right) & x[n] > 0 \\ 0 & x[n] < 0 \end{cases}$$

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

Consider estimation of

$$\alpha = 1/\theta.$$

$$\hat{\alpha} = \frac{1}{\hat{\theta}} = \bar{x} + \frac{\lambda}{N}$$ **? (holds for MLE)**

Chapter 11 – General Bayesian Estimators

Example

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

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

$$p(x[n]|\alpha) = \begin{cases} \frac{1}{\alpha} \exp\left(-\frac{x[n]}{\alpha}\right) & x[n] > 0 \\ 0 & x[n] < 0 \end{cases}$$

$$p_\alpha(\alpha) \quad =$$

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

Consider estimation of

$$\alpha = 1/\theta.$$

$$\hat{\alpha} = \frac{1}{\hat{\theta}} = \bar{x} + \frac{\lambda}{N}$$ **? (holds for MLE)**

Chapter 11 – General Bayesian Estimators

Example

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

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

$$p(x[n]|\alpha) = \begin{cases} \frac{1}{\alpha} \exp\left(-\frac{x[n]}{\alpha}\right) & x[n] > 0 \\ 0 & x[n] < 0 \end{cases}$$

$$p_\alpha(\alpha) = \frac{p_\theta(\theta(\alpha))}{\left|\frac{d\alpha}{d\theta}\right|}$$

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

Consider estimation of

$$\alpha = 1/\theta.$$

$$\hat{\alpha} = \frac{1}{\hat{\theta}} = \bar{x} + \frac{\lambda}{N}$$ **? (holds for MLE)**

Chapter 11 – General Bayesian Estimators

Example

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

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

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

Consider estimation of

$$\alpha = 1/\theta.$$

$$\hat{\alpha} = \frac{1}{\hat{\theta}} = \bar{x} + \frac{\lambda}{N}$$ **? (holds for MLE)**

$$p(x[n]|\alpha) = \begin{cases} \frac{1}{\alpha}\exp\left(-\frac{x[n]}{\alpha}\right) & x[n] > 0 \\ 0 & x[n] < 0 \end{cases}$$

$$\begin{aligned} p_\alpha(\alpha) &= \frac{p_\theta(\theta(\alpha))}{\left|\frac{d\alpha}{d\theta}\right|} \\ &= \begin{cases} \frac{\lambda \exp(-\lambda/\alpha)}{\alpha^2} & \alpha > 0 \\ 0 & \alpha < 0. \end{cases} \end{aligned}$$

Chapter 11 – General Bayesian Estimators

Example

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

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

$$p(x[n]|\alpha) = \begin{cases} \frac{1}{\alpha} \exp\left(-\frac{x[n]}{\alpha}\right) & x[n] > 0 \\ 0 & x[n] < 0 \end{cases}$$

$$\begin{aligned} p_\alpha(\alpha) &= \frac{p_\theta(\theta(\alpha))}{\left|\frac{d\alpha}{d\theta}\right|} \\ &= \begin{cases} \frac{\lambda \exp(-\lambda/\alpha)}{\alpha^2} & \alpha > 0 \\ 0 & \alpha < 0. \end{cases} \end{aligned}$$

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

Consider estimation of

$$\alpha = 1/\theta.$$

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

$$g(\alpha) = \ln p(\mathbf{x}|\alpha) + \ln p(\alpha) = \quad ....$$

$$\frac{dg}{d\alpha} = -\frac{N+2}{\alpha} + \frac{N\bar{x} + \lambda}{\alpha^2}$$

$$\boxed{\hat{\alpha} = \frac{N\bar{x} + \lambda}{N+2}}$$