

430.714

Estimation Theory

Prof. Songhwai Oh
ECE, SNU

Bayesian Approach

What is probability?

A patient is admitted to the hospital and a potentially life-saving drug is administered. The following dialog takes place between the nurse and a concerned relative.

RELATIVE: Nurse, what is the probability that the drug will work?

NURSE: I hope it works, we'll know tomorrow.

RELATIVE: Yes, but what is the probability that it will?

NURSE: Each case is different, we have to wait.

RELATIVE: But let's see, out of a hundred patients that are treated under similar conditions, how many times would you expect it to work?

NURSE (somewhat annoyed): I told you, every person is different, for some it works, for some it doesn't.

RELATIVE (insisting): Then tell me, if you had to bet whether it will work or not, which side of the bet would you take?

NURSE (cheering up for a moment): I'd bet it will work.

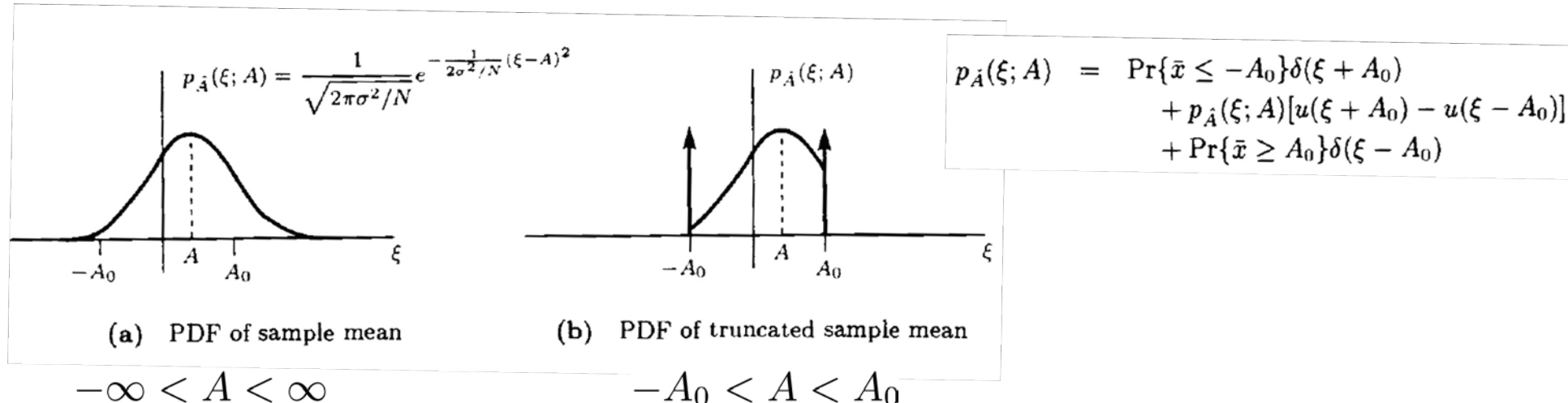
RELATIVE (somewhat relieved): OK, now, would you be willing to lose two dollars if it doesn't work, and gain one dollar if it does?

NURSE (exasperated): What a sick thought! You are wasting my time!

Two Schools of Thoughts

- Frequentist
 - Frequency of occurrences
 - Parameters are unknown constants
- Bayesian
 - Degree of belief (educated guess)
 - Parameters are random variables

Prior Knowledge: Example

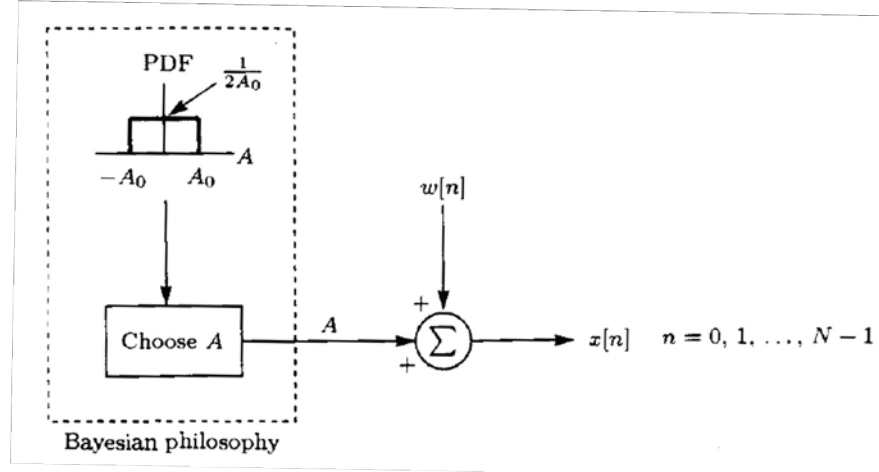


$$\hat{A} = \bar{x}$$

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

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

Bayesian approach



Bayesian MSE: $\text{Bmse}(\hat{A}) = E[(A - \hat{A})^2]$

expectation wrt the joint $p(\mathbf{x}, A)$.

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

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

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

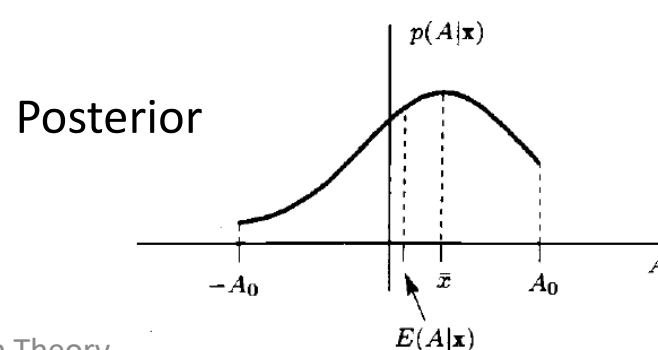
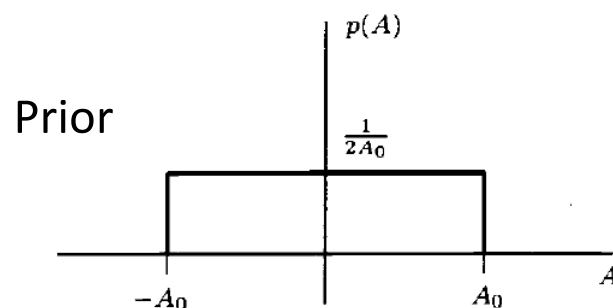
$$\frac{\partial}{\partial \hat{A}} \int (A - \hat{A})^2 p(A|\mathbf{x}) dA = \int \frac{\partial}{\partial \hat{A}} (A - \hat{A})^2 p(A|\mathbf{x}) dA$$

$$= \int -2(A - \hat{A}) p(A|\mathbf{x}) dA$$

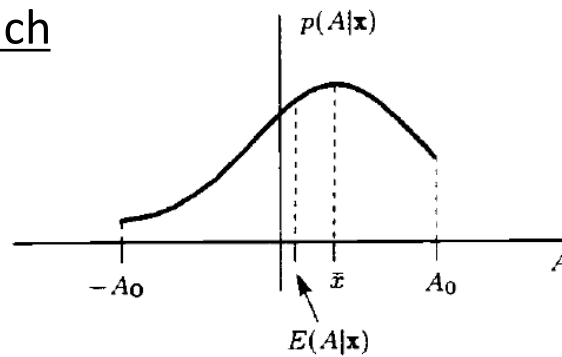
$$= -2 \int A p(A|\mathbf{x}) dA + 2\hat{A} \int p(A|\mathbf{x}) dA$$

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

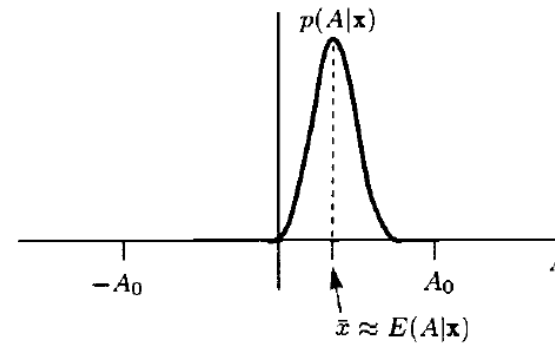
$$= E(A|\mathbf{x})$$



Bayesian approach



(a) Short data record



(b) Large data record

Minimum Mean Square Error (MMSE) estimator: $\hat{\theta} = E(\theta|\mathbf{x}) = \int \theta p(\theta|\mathbf{x}) d\theta$

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

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

Bayesian approach: Gaussian Case (Summary)

$$x_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2) \quad X = \{x_1, \dots, x_N\}$$

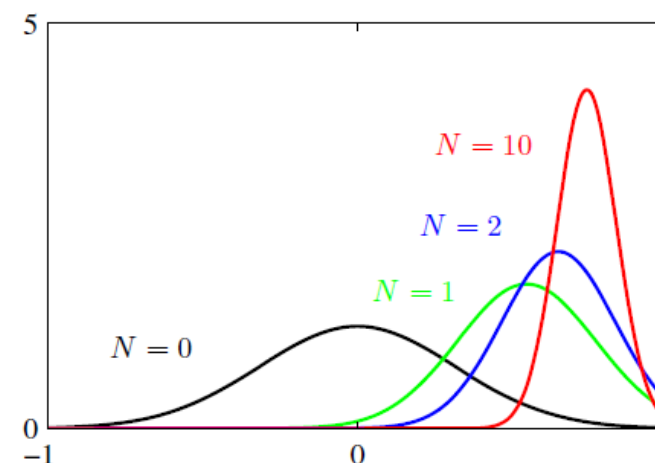
Prior: $p(\mu) = \mathcal{N}(\mu | \mu_0, \sigma_0^2)$

Likelihood: $p(\mathbf{X} | \mu) = \prod_{n=1}^N p(x_n | \mu) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{n=1}^N (x_n - \mu)^2 \right\}$

Posterior: $p(\mu | \mathbf{X}) \propto p(\mathbf{X} | \mu) p(\mu)$

$$p(\mu | \mathbf{X}) = \mathcal{N}(\mu | \mu_N, \sigma_N^2)$$

$$\mu_N = \frac{\sigma^2}{N\sigma_0^2 + \sigma^2} \mu_0 + \frac{N\sigma_0^2}{N\sigma_0^2 + \sigma^2} \mu_{\text{ML}} \quad \mu_{\text{ML}} = \frac{1}{N} \sum_{n=1}^N x_n$$
$$\frac{1}{\sigma_N^2} = \frac{1}{\sigma_0^2} + \frac{N}{\sigma^2}$$



$$N=0, \mu_N = \mu_0$$

$$N \rightarrow \infty, \mu_N \rightarrow \mu_{\text{ML}}$$

For fixed N , take the limit $\sigma_0^2 \rightarrow \infty$
Then we get the MLE.

Bayesian approach: Gaussian Case

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

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

$$\begin{aligned} \text{Posterior: } p(A|\mathbf{x}) &= \frac{p(\mathbf{x}|A)p(A)}{\int p(\mathbf{x}|A)p(A) dA} \\ &= \frac{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}} \sqrt{2\pi\sigma_A^2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n] \right] \exp \left[-\frac{1}{2\sigma^2} (NA^2 - 2NA\bar{x}) \right]}{\int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}} \sqrt{2\pi\sigma_A^2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n] \right] \exp \left[-\frac{1}{2\sigma^2} (NA^2 - 2NA\bar{x}) \right]} \cdot \frac{\exp \left[-\frac{1}{2\sigma_A^2} (A - \mu_A)^2 \right]}{\exp \left[-\frac{1}{2\sigma_A^2} (A - \mu_A)^2 \right]} dA \\ &= \frac{\exp \left[-\frac{1}{2} \left(\frac{1}{\sigma^2} (NA^2 - 2NA\bar{x}) + \frac{1}{\sigma_A^2} (A - \mu_A)^2 \right) \right]}{\int_{-\infty}^{\infty} \exp \left[-\frac{1}{2} \left(\frac{1}{\sigma^2} (NA^2 - 2NA\bar{x}) + \frac{1}{\sigma_A^2} (A - \mu_A)^2 \right) \right] dA} \\ &= \frac{\exp \left[-\frac{1}{2} Q(A) \right]}{\int_{-\infty}^{\infty} \exp \left[-\frac{1}{2} Q(A) \right] dA} \end{aligned}$$

Bayesian approach: Gaussian Case

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

$$\sigma_{A|x}^2 = \frac{1}{\frac{N}{\sigma^2} + \frac{1}{\sigma_A^2}}$$

$$\mu_{A|x} = \left(\frac{N}{\sigma^2} \bar{x} + \frac{\mu_A}{\sigma_A^2} \right) \sigma_{A|x}^2$$

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

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

Gaussian distribution !!!

MMSE

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

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

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

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

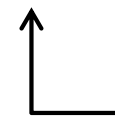
Bayesian approach: Gaussian Case

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

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

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

$$\text{Bmse}(\hat{A}) < \frac{\sigma^2}{N}$$



minimum MSE without prior information

With prior knowledge, we can improve the estimator.

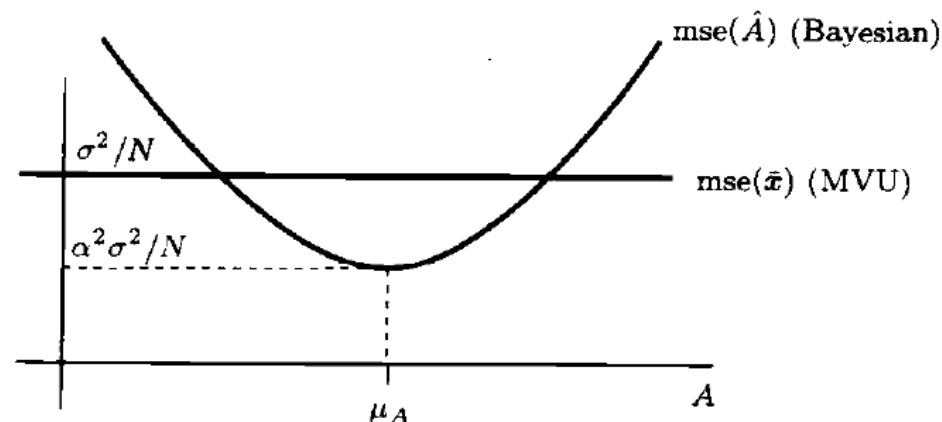
Bayesian Estimation for Deterministic Parameters

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

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

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

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



If the prior model is incorrect, the Bayesian estimator can perform worse than MVU.