

Tutorial 2: Maximum-Likelihood and Bayesian Parameter Estimation

Rui Zhao

rzhao@ee.cuhk.edu.hk

Jan. 23, 2014

- 1 ML estimate: $\hat{\Sigma}$ of multivariate Gaussian
- 2 ML estimate bias: $\hat{\sigma}^2$ of univariate Gaussian
- 3 Bayesian estimate: Univariate Gaussian
 - brief review
 - posteriori $p(\mu|\mathcal{D})$
 - conditional probability density $p(x|w_i, \mathcal{D}_i)$

Outline

- 1 ML estimate: $\hat{\Sigma}$ of multivariate Gaussian
- 2 ML estimate bias: $\hat{\sigma}^2$ of univariate Gaussian
- 3 Bayesian estimate: Univariate Gaussian
 - brief review
 - posteriori $p(\mu|\mathcal{D})$
 - conditional probability density $p(x|w_i, \mathcal{D}_i)$

- 1 ML estimate: $\hat{\Sigma}$ of multivariate Gaussian
- 2 ML estimate bias: $\hat{\sigma}^2$ of univariate Gaussian
- 3 Bayesian estimate: Univariate Gaussian
 - brief review
 - posteriori $p(\mu|\mathcal{D})$
 - conditional probability density $p(x|w_i, \mathcal{D}_i)$

1. ML estimate: $\hat{\Sigma}$ of multivariate Gaussian

Multivariate Gaussian Case: unknown μ and Σ

$$\hat{\mu} = \frac{1}{n} \sum_{k=1}^n \mathbf{x}_k \quad (1)$$

$$\hat{\Sigma} = \frac{1}{n} \sum_{k=1}^n (\mathbf{x}_k - \hat{\mu})(\mathbf{x}_k - \hat{\mu})^t \quad (2)$$

- $\hat{\mu}$ is the sample mean.
- $\hat{\Sigma}$ is the arithmetic average of the n matrices $(\mathbf{x}_k - \hat{\mu})(\mathbf{x}_k - \hat{\mu})^t$.

1. ML estimate: $\hat{\Sigma}$ of multivariate Gaussian

Multivariate normal density

$$p(\mathbf{x} \mid \mu, \Sigma) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \mu)^t \Sigma^{-1} (\mathbf{x} - \mu) \right] \quad (3)$$

Draw $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ independently from $p(\mathbf{x} \mid \mu, \Sigma)$, and the joint density (likelihood) is:

$$p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \mid \mu, \Sigma) = \quad (4)$$

$$\frac{1}{(2\pi)^{nd/2} |\Sigma|^{n/2}} \exp \left[-\frac{1}{2} \sum_{k=1}^n (\mathbf{x}_k - \mu)^t \Sigma^{-1} (\mathbf{x}_k - \mu) \right] \quad (5)$$

Log-likelihood $l(\mu, \Sigma)$ is

$$l(\mu, \Sigma) = -\frac{nd}{2} \ln(2\pi) - \frac{n}{2} \ln |\Sigma| - \frac{1}{2} \sum_{k=1}^n (\mathbf{x}_k - \mu)^t \Sigma^{-1} (\mathbf{x}_k - \mu) \quad (6)$$

1. ML estimate: $\hat{\Sigma}$ of multivariate Gaussian

Multivariate normal density

Let $\mathbf{A} = \Sigma^{-1}$

$$l(\mu, \Sigma) = -\frac{nd}{2} \ln(2\pi) - \frac{n}{2} \ln |\Sigma| - \frac{1}{2} \sum_{k=1}^n (\mathbf{x}_k - \mu)^t \Sigma^{-1} (\mathbf{x}_k - \mu) \quad (7)$$

$$l(\mu, \Sigma) = -\frac{nd}{2} \ln(2\pi) + \frac{n}{2} \ln \mathbf{A} - \frac{1}{2} \sum_{k=1}^n (\mathbf{x}_k - \mu)^t \mathbf{A} (\mathbf{x}_k - \mu) \quad (8)$$

$$\frac{\partial l(\mu, \Sigma)}{\partial \mathbf{A}} = \frac{n}{2} \mathbf{A}^{-1} - \frac{1}{2} \sum_{k=1}^n (\mathbf{x}_k - \mu)(\mathbf{x}_k - \mu)^t = 0 \quad (9)$$

Replace \mathbf{A} by Σ^{-1}

$$\hat{\Sigma} = \frac{1}{n} \sum_{k=1}^n (\mathbf{x}_k - \hat{\mu})(\mathbf{x}_k - \hat{\mu})^t, \quad (\hat{\mu} = \frac{1}{n} \sum_{k=1}^n \mathbf{x}_k) \quad (10)$$

- 1 ML estimate: $\hat{\Sigma}$ of multivariate Gaussian
- 2 ML estimate bias: $\hat{\sigma}^2$ of univariate Gaussian
- 3 Bayesian estimate: Univariate Gaussian
 - brief review
 - posteriori $p(\mu|\mathcal{D})$
 - conditional probability density $p(x|w_i, \mathcal{D}_i)$

- 1 ML estimate: $\hat{\Sigma}$ of multivariate Gaussian
- 2 ML estimate bias: $\hat{\sigma}^2$ of univariate Gaussian
- 3 Bayesian estimate: Univariate Gaussian
 - brief review
 - posteriori $p(\mu|\mathcal{D})$
 - conditional probability density $p(x|w_i, \mathcal{D}_i)$

2. ML estimate bias: $\hat{\sigma}^2$ of univariate Gaussian

Univariate Gaussian Case

ML estimator $\hat{\sigma}^2$ is biased.

$$E[\hat{\sigma}^2] = E\left[\frac{1}{n} \sum_{k=1}^n (x_k - \hat{\mu})^2\right] = \frac{n-1}{n} \sigma^2 \quad (11)$$

An elementary unbiased estimator for σ^2 is given by $\frac{1}{n-1} \sum_{k=1}^n (x_k - \hat{\mu})^2$.

2. ML estimate bias: $\hat{\sigma}^2$ of univariate Gaussian

Univariate Gaussian Case

$$E\left[\frac{1}{n} \sum_{k=1}^n (x_k - \hat{\mu})^2\right] = E\left[\frac{1}{n} \sum_{k=1}^n \left(x_k - \frac{1}{n} \sum_{j=1}^n x_j\right)^2\right] \quad (12)$$

$$= E\left[\frac{1}{n} \sum_{k=1}^n \left(x_k^2 - \frac{2}{n} x_k \sum_{j=1}^n x_j + \frac{1}{n^2} \left(\sum_{j=1}^n x_j\right)^2\right)\right] \quad (13)$$

$$= E\left[\frac{1}{n} \left(\sum_{k=1}^n x_k^2 - \frac{2}{n} \left(\sum_{k=1}^n x_k\right)^2 + \frac{n}{n^2} \left(\sum_{k=1}^n x_k\right)^2\right)\right] \quad (14)$$

$$= E\left[\frac{1}{n} \left(\sum_{k=1}^n x_k^2 - \frac{1}{n} \left(\sum_{k=1}^n x_k\right)^2\right)\right] \quad (15)$$

$$= \frac{1}{n} E\left[\sum_{k=1}^n x_k^2\right] - \frac{1}{n^2} E\left[\left(\sum_{k=1}^n x_k\right)^2\right] \quad (16)$$

$$= E[x^2] - \frac{1}{n^2} E\left[\sum_{k=1}^n x_k^2 + \sum_{i \neq j} x_i x_j\right] \quad (17)$$

$$= E[x^2] - \frac{1}{n^2} E\left[\sum_{k=1}^n x_k^2\right] - \frac{1}{n^2} E\left[\sum_{i \neq j} x_i x_j\right] \quad (18)$$

$$= E[x^2] - \frac{1}{n} E[x^2] - \frac{n^2 - n}{n^2} E[x_i x_j] = \frac{n-1}{n} \left(E[x^2] - (E[x])^2\right) \quad (19)$$

$$= \frac{n-1}{n} \sigma^2 \quad (20)$$

- 1 ML estimate: $\hat{\Sigma}$ of multivariate Gaussian
- 2 ML estimate bias: $\hat{\sigma}^2$ of univariate Gaussian
- 3 Bayesian estimate: Univariate Gaussian
 - brief review
 - posteriori $p(\mu|\mathcal{D})$
 - conditional probability density $p(x|w_i, \mathcal{D}_i)$

- 1 ML estimate: $\hat{\Sigma}$ of multivariate Gaussian
- 2 ML estimate bias: $\hat{\sigma}^2$ of univariate Gaussian
- 3 Bayesian estimate: Univariate Gaussian
 - brief review
 - posteriori $p(\mu|\mathcal{D})$
 - conditional probability density $p(x|w_i, \mathcal{D}_i)$

- 1 ML estimate: $\hat{\Sigma}$ of multivariate Gaussian
- 2 ML estimate bias: $\hat{\sigma}^2$ of univariate Gaussian
- 3 Bayesian estimate: Univariate Gaussian
 - brief review
 - posteriori $p(\mu|\mathcal{D})$
 - conditional probability density $p(x|w_i, \mathcal{D}_i)$

3.1 Bayesian estimate: brief review

Bayesian estimate

Given sample set \mathcal{D} , then posteriori for estimation is

$$P(w_i|\mathbf{x}, \mathcal{D}) = \frac{p(\mathbf{x}|w_i, \mathcal{D})P(w_i|\mathcal{D})}{\sum_{j=1}^c p(\mathbf{x}|w_j, \mathcal{D})P(w_j|\mathcal{D})} \quad (21)$$

- consider each class individually: $p(\mathbf{x}|w_i, \mathcal{D}) \rightarrow p(\mathbf{x}|\mathcal{D})$
- prior is known $P(w_i|\mathcal{D})$

Target: estimate posteriori $p(\mathbf{x}|\mathcal{D}) \rightarrow p(\mathbf{x})$

$$p(\mathbf{x}|\mathcal{D}) = \int p(\mathbf{x}, \theta|\mathcal{D})d\theta \quad (22)$$

$$= \int p(\mathbf{x}|\theta, \mathcal{D})p(\theta|\mathcal{D})d\theta = \int p(\mathbf{x}|\theta)p(\theta|\mathcal{D})d\theta \quad (23)$$

- 1 ML estimate: $\hat{\Sigma}$ of multivariate Gaussian
- 2 ML estimate bias: $\hat{\sigma}^2$ of univariate Gaussian
- 3 Bayesian estimate: Univariate Gaussian
 - brief review
 - posteriori $p(\mu|\mathcal{D})$
 - conditional probability density $p(x|w_i, \mathcal{D}_i)$

Outline

- 1 ML estimate: $\hat{\Sigma}$ of multivariate Gaussian
- 2 ML estimate bias: $\hat{\sigma}^2$ of univariate Gaussian
- 3 Bayesian estimate: Univariate Gaussian
 - brief review
 - posteriori $p(\mu|\mathcal{D})$
 - conditional probability density $p(x|w_i, \mathcal{D}_i)$

3.2 Bayesian estimate of univariate Gaussian: posteriori $p(\mu|\mathcal{D})$

Bayesian estimate

$$P(w_i|\mathbf{x}, \mathcal{D}) \Rightarrow p(\mathbf{x}|w_i, \mathcal{D}) \Rightarrow p(\mathbf{x}|\mathcal{D}) \Rightarrow p(\mathbf{x}|\theta) \text{ \& } p(\theta|\mathcal{D})$$

- $p(\mathbf{x}|\theta)$ is pre-assumed in form: $p(x|\mu) \sim \mathcal{N}(\mu, \sigma^2)$
- $p(\theta|\mathcal{D})$ is the posteriori: $p(\mu|\mathcal{D})$

Posteriori $p(\mu|\mathcal{D})$ of univariate Gaussian

$$p(\mu|\mathcal{D}) = \frac{p(\mathcal{D}|\mu)p(\mu)}{\int p(\mathcal{D}|\mu)p(\mu)d\mu} \quad (24)$$

$$= \alpha \prod_{k=1}^n p(x_k|\mu)p(\mu) \quad (25)$$

where $p(x_k|\mu) \sim \mathcal{N}(\mu, \sigma^2)$, and $p(\mu) \sim \mathcal{N}(\mu_0, \sigma_0^2)$.

3.2 Bayesian estimate of univariate Gaussian: posteriori $p(\mu|\mathcal{D})$

Posteriori $p(\mu|\mathcal{D})$ of univariate Gaussian

$$p(\mu|\mathcal{D}) = \alpha \prod_{k=1}^n p(x_k|\mu)p(\mu) \quad (26)$$

$$= \alpha \prod_{k=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x_k - \mu}{\sigma}\right)^2\right] \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left[-\frac{1}{2}\left(\frac{\mu - \mu_0}{\sigma_0}\right)^2\right] \quad (27)$$

$$= \alpha' \exp\left[-\frac{1}{2}\left(\sum_{k=1}^n \left(\frac{\mu - x_k}{\sigma}\right)^2 + \left(\frac{\mu - \mu_0}{\sigma_0}\right)^2\right)\right] \quad (28)$$

$$= \alpha'' \exp\left[-\frac{1}{2}\left[\left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right)\mu^2 - 2\left(\frac{1}{\sigma^2} \sum_{k=1}^n x_k + \frac{\mu_0}{\sigma_0^2}\right)\mu\right]\right] \quad (29)$$

$$\therefore p(\mu|\mathcal{D}) \sim \mathcal{N}(\mu_n, \sigma_n^2), \text{ i.e. } p(\mu|\mathcal{D}) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left[-\frac{1}{2}\left(\frac{\mu - \mu_n}{\sigma_n}\right)^2\right] \quad (30)$$

$$\Rightarrow \frac{1}{\sigma_n^2} = \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}, \quad \frac{\mu_n}{\sigma_n^2} = \frac{1}{\sigma^2} \sum_{k=1}^n x_k + \frac{\mu_0}{\sigma_0^2} = \frac{n}{\sigma^2} \hat{\mu}_n + \frac{\mu_0}{\sigma_0^2} \quad (31)$$

$$\Rightarrow \mu_n = \left(\frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2}\right)\hat{\mu}_n + \frac{\sigma^2}{\sigma_0^2 + \sigma^2}\mu_0, \quad \sigma_n^2 = \frac{\sigma_0^2\sigma^2}{n\sigma_0^2 + \sigma^2} \quad (32)$$

Outline

- 1 ML estimate: $\hat{\Sigma}$ of multivariate Gaussian
- 2 ML estimate bias: $\hat{\sigma}^2$ of univariate Gaussian
- 3 Bayesian estimate: Univariate Gaussian
 - brief review
 - posteriori $p(\mu|\mathcal{D})$
 - conditional probability density $p(x|w_i, \mathcal{D}_i)$

- 1 ML estimate: $\hat{\Sigma}$ of multivariate Gaussian
- 2 ML estimate bias: $\hat{\sigma}^2$ of univariate Gaussian
- 3 Bayesian estimate: Univariate Gaussian
 - brief review
 - posteriori $p(\mu|\mathcal{D})$
 - conditional probability density $p(x|w_i, \mathcal{D}_i)$

3.3 Bayesian estimate of univariate Gaussian: conditional probability density $p(x|w_i, \mathcal{D}_i)$

Conditional probability density $p(x|w_i, \mathcal{D}_i)$

$$p(x|w_i, \mathcal{D}_i) \quad (33)$$

$$\sim p(x|\mathcal{D}) \quad (34)$$

$$= \int p(x|\mu)p(\mu|\mathcal{D})d\mu \quad (35)$$

$$\begin{aligned} &= \int \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 \right] \frac{1}{\sqrt{2\pi}\sigma_n} \exp \left[-\frac{1}{2} \left(\frac{\mu-\mu_n}{\sigma_n} \right)^2 \right] d\mu \\ &= \frac{1}{2\pi\sigma\sigma_n} \exp \left[-\frac{1}{2} \frac{(x-\mu_n)^2}{\sigma^2 + \sigma_n^2} \right] f(\sigma, \sigma_n) \end{aligned} \quad (36)$$

$$\therefore p(x|\mathcal{D}) \sim \mathcal{N}(\mu_n, \sigma^2 + \sigma_n^2) \quad (37)$$

$$f(\sigma, \sigma_n) = \int \exp \left[-\frac{1}{2} \frac{\sigma^2 + \sigma_n^2}{\sigma^2 \sigma_n^2} \left(\mu - \frac{\sigma_n^2 x + \sigma^2 \mu_n}{\sigma^2 + \sigma_n^2} \right)^2 \right] d\mu \quad (38)$$

3.3 Bayesian estimate of univariate Gaussian: conditional probability density $p(x|w_i, \mathcal{D}_i)$

Conditional probability density $p(x|w_i, \mathcal{D}_i)$

$$p(x|w_i, \mathcal{D}_i) \sim p(x|\mathcal{D}) \quad (39)$$

$$= \int p(x|\mu)p(\mu|\mathcal{D})d\mu \quad (40)$$

$$= \int \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left[-\frac{1}{2}\left(\frac{\mu-\mu_n}{\sigma_n}\right)^2\right] d\mu \quad (41)$$

$$= \frac{1}{2\pi\sigma\sigma_n} \int \exp\left[-\frac{1}{2}\left(\frac{x^2-2x\mu+\mu^2}{\sigma^2} + \frac{\mu^2-2\mu\mu_n+\mu_n^2}{\sigma_n^2}\right)\right] d\mu \quad (42)$$

$$= \frac{1}{2\pi\sigma\sigma_n} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma^2} + \frac{\mu_n^2}{\sigma_n^2}\right)\right] \int \exp\left[-\frac{1}{2}\left(\frac{\mu^2-2x\mu}{\sigma^2} + \frac{\mu^2-2\mu\mu_n}{\sigma_n^2}\right)\right] d\mu \quad (43)$$

$$= \frac{1}{2\pi\sigma\sigma_n} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma^2} + \frac{\mu_n^2}{\sigma_n^2}\right)\right] \int \exp\left[-\frac{1}{2}\left(\frac{\sigma_n^2+\sigma^2}{\sigma_n^2\sigma^2}\mu^2 - 2\left(\frac{x}{\sigma^2} + \frac{\mu_n}{\sigma_n^2}\right)\mu\right)\right] d\mu \quad (44)$$

$$= \frac{1}{2\pi\sigma\sigma_n} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma^2} + \frac{\mu_n^2}{\sigma_n^2}\right)\right] \int \exp\left[-\frac{1}{2}\frac{\sigma_n^2+\sigma^2}{\sigma_n^2\sigma^2}\left(\mu^2 - 2\frac{\sigma_n^2\sigma^2}{\sigma_n^2+\sigma^2}\left(\frac{x}{\sigma^2} + \frac{\mu_n}{\sigma_n^2}\right)\mu\right)\right] d\mu \quad (45)$$

$$= \frac{1}{2\pi\sigma\sigma_n} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma^2} + \frac{\mu_n^2}{\sigma_n^2} - \frac{(\sigma_n^2x + \sigma^2\mu_n)^2}{\sigma_n^2\sigma^2(\sigma_n^2 + \sigma^2)}\right)\right] f(\sigma, \sigma_n) \quad (46)$$

$$f(\sigma, \sigma_n) = \int \exp\left[-\frac{1}{2}\frac{\sigma_n^2+\sigma^2}{\sigma_n^2\sigma^2}\left(\mu - \frac{\sigma_n^2x + \sigma^2\mu_n}{\sigma^2 + \sigma_n^2}\right)^2\right] d\mu$$

3.3 Bayesian estimate of univariate Gaussian: conditional probability density $p(x|w_i, \mathcal{D}_i)$

Conditional probability density $p(x|w_i, \mathcal{D}_i)$

$$p(x|\mathcal{D}) \quad (47)$$

$$= \frac{1}{2\pi\sigma\sigma_n} \exp \left[-\frac{1}{2} \left(\frac{x^2}{\sigma^2} + \frac{\mu_n^2}{\sigma_n^2} \right) \right] \int \exp \left[-\frac{1}{2} \frac{\sigma_n^2 + \sigma^2}{\sigma_n^2 \sigma^2} \left(\mu^2 - 2 \frac{\sigma_n^2 \sigma^2}{\sigma_n^2 + \sigma^2} \left(\frac{x}{\sigma^2} + \frac{\mu_n}{\sigma_n^2} \right) \mu \right) \right] d\mu \quad (48)$$

$$= \frac{1}{2\pi\sigma\sigma_n} \exp \left[-\frac{1}{2} \left(\frac{x^2}{\sigma^2} + \frac{\mu_n^2}{\sigma_n^2} - \frac{(\sigma_n^2 x + \sigma^2 \mu_n)^2}{\sigma_n^2 \sigma^2 (\sigma_n^2 + \sigma^2)} \right) \right] f(\sigma, \sigma_n) \quad (49)$$

$$= \frac{1}{2\pi\sigma\sigma_n} \exp \left[-\frac{1}{2} \left(\frac{x^2}{\sigma^2} + \frac{\mu_n^2}{\sigma_n^2} - \frac{\sigma_n^2 x^2}{\sigma^2 (\sigma_n^2 + \sigma^2)} - \frac{2x\mu_n}{\sigma_n^2 + \sigma^2} - \frac{\sigma^2 \mu_n^2}{\sigma_n^2 (\sigma_n^2 + \sigma^2)} \right) \right] f(\sigma, \sigma_n) \quad (50)$$

$$= \frac{1}{2\pi\sigma\sigma_n} \exp \left[-\frac{1}{2} \left(\frac{x^2}{\sigma_n^2 + \sigma^2} - \frac{2x\mu_n}{\sigma_n^2 + \sigma^2} + \frac{\mu_n^2}{\sigma_n^2 + \sigma^2} \right) \right] f(\sigma, \sigma_n) \quad (51)$$

$$= \frac{1}{2\pi\sigma\sigma_n} \exp \left[-\frac{1}{2} \frac{(x - \mu_n)^2}{\sigma^2 + \sigma_n^2} \right] f(\sigma, \sigma_n) \quad (52)$$

$$f(\sigma, \sigma_n) = \int \exp \left[-\frac{1}{2} \frac{\sigma_n^2 + \sigma^2}{\sigma_n^2 \sigma^2} \left(\mu - \frac{\sigma_n^2 x + \sigma^2 \mu_n}{\sigma^2 + \sigma_n^2} \right)^2 \right] d\mu$$