Introduction to (Bayesian) Estimation
MAE 5020

Estimation of random variables

Oklahoma State University

August 15, 2023

Overview

Objective: transition to estimation of unknown *random variables* instead of deterministic θ

- Review of Multivariate Gaussian Random Variables
- Maximum a posterior estimation (MAP)

Gaussian rv

1) reasonable approximation to observed random behavior;
2) central limit theorem indicates superposition of an arbitrarily large number of independent microscopic random phenomena is justifiably Gaussian at the macroscopic.

Univariate Gaussian

$y \sim N(y; m_y, \sigma_y^2)$ (mean m_y, variance σ_y^2) if the pdf of y is

$$p(y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left[-\frac{(y - m_y)^2}{2\sigma_y^2}\right], \quad -\infty < y < \infty. \tag{1}$$

Multivariate Gaussian

$y = (y_1, \cdots, y_m)^T$. $y \sim N(y; m_y, P_y)$, where $m_y \in \mathbb{R}^m = E(y)$ and $P_y \in \mathbb{R}^{m \times m}_{\geq 0} = E\left((y - m_y)(y - m_y)^T\right)$ if

$$p(y_1, \cdots, y_m) = p(y) = \frac{1}{\sqrt{(2\pi)^m |P_y|}} \exp\left[-\frac{(y - m_y)^T P_y^{-1} (y - m_y)}{2}\right]. \tag{2}$$

Jointly Gaussian random vectors

$x \sim N(x; m_x, P_x) \in \mathbb{R}^n$, $y \sim N(y; m_y, P_y) \in \mathbb{R}^m$. Then the joint distribution of x, y is Gaussian, given by

$$p(x, y) = \frac{1}{\sqrt{(2\pi)^{m+n}|P_z|}} \exp\left[-\frac{(z - m_z)^T P_z^{-1}(z - m_z)}{2}\right] \quad (3)$$

where $z = (x^T, y^T)^T$, $m_z = (m_x^T, m_y^T)^T$, and

$$P_z = \begin{pmatrix} P_x & P_{xy} \\ P_{yx} & P_y \end{pmatrix} \quad (4)$$

with

$$P_{xy} = E\left((x - m_x)(y - m_y)^T\right), \quad (5)$$

$$P_{yx} = E\left((y - m_y)(x - m_x)^T\right) = P_{yx}^T. \quad (6)$$

When x and y are statistically independent, $P_{xy} = 0$.

The conditional density function

This is one of the most important density functions we will be interested in.
If x, y are jointly Gaussian,

$$p(x|y) = \frac{1}{\sqrt{(2\pi)^n |D|}} \exp\left[-\frac{(x-m)^T D^{-1} (x-m)}{2} \right] \tag{7}$$

where

$$m = E(x|y) = m_x + P_{xy} P_y^{-1} (y - m_y) \tag{8}$$

$$D = P_x - P_{xy} P_y^{-1} P_{yx} > 0. \tag{9}$$

Note that m is an affine transformation of y (other terms are all deterministic!)

Linear property

When x and y are jointly Gaussian, $z = Ax + By + c$ is also Gaussian if A, B, c are deterministic.

Properties of conditional mean

Let x, y, z be n, m, r dimensional jointly Gaussian random vectors. If y, z are statistically independent

$$E(x|y,z) = E(x|y) + E(x|z) - m_x. \tag{10}$$

else,

$$E(x|y,z) = E(x|y) + E(x|\bar{z}) - m_x \tag{11}$$

where $\bar{z} = z - E(z|y)$ (the dependence on y is removed, y and $\bar{z}$ are independent).

MAP Estimation: setting

- View θ as
- Measurements $z(1), \cdots, z(k)$ assumed to depend on θ
- Priori probability model on θ: $P(\theta)$

Bayes theorem

$$p(\theta|Z(k)) = \frac{p(Z(k)|\theta)p(\theta)}{p(Z(k))}$$

$p(\theta|Z(k))$:
$p(\theta)$:
$p(Z(k)|\theta)$:

MAP Estimation

Since $p(Z(k))$ does not depend on θ:

The MAP estimate, $\hat{\theta}_{MAP}(k)$ is found to maximize $p(\theta|Z(k))$ or RHS of the above equation.

Properties

- When $p(\theta)$ is uniform
- Generally,
- The MAP estimate doesn't "carry over", i.e.,
- If there is a lot of measurements (big data),

Obtaining MAP estimates

- The prior $p(\theta)$ and $p(Z(k)|\theta)$ must be specified (or at least can be evaluated).
- Maximization of $p(\theta|Z(k))$ or $\ln p(\theta|Z(k))$ typically requires optimization.
- In the special case of linear models and Gaussian noise and prior, we can have a closed-form solution.

Example: MAP estimation of Gaussian mean

The generic linear Gaussian model

Recall $Z(k) = H(k)\theta + V(k)$, where $\theta \sim N(\theta; m_\theta, P_\theta)$ and $V(k) \sim N(0, R(k))$. Derive its MAP estimate.