ASEN 5044, Fall 2018

Statistical Estimation for Dynamical Systems

Lecture 17 [Special Topic #3]: Multivariate Gaussian Conditional PDFs and Bayes' Rule for Gaussian Random Vectors

Prof. Nisar Ahmed (Nisar.Ahmed@Colorado.edu)

Friday 10/12/2018

University of Colorado Boulder

Today:

- Conditional expectations and expected values
  - Conditional mean and covariance
- Conditional multivariate Gaussian PDFs
  - Given Gaussian $p(x_1,x_2)$, what is $p(x_1|x_2)$ or $p(x_2|x_1)$?
- Bayes' rule for multivariate Gaussian PDFs
  - Given $p(x,y)=p(x)p(y|x)$ in factored form, what is $p(x|y=c)$ for Gaussians?

Conditional Expectations and Expected Values

- Recall definition of conditional pdf for random variables x_1 and x_2

$$p(x_1|x_2=c) \triangleq \frac{p(x_1, x_2=c)}{p(x_2=c)} = \frac{p(x_1, x_2=c)}{\int_{-\infty}^{\infty} p(x_1, x_2=c)\,dx_1}$$

$$p(x_2|x_1=d) \triangleq \frac{p(x_1=d, x_2)}{p(x_1=d)} = \frac{p(x_1=d, x_2)}{\int_{-\infty}^{\infty} p(x_1=d, x_2)\,dx_2}$$

- Concept of expected values naturally extends to **conditional expected values**

given some fxns $f_1(x_1)$ & $f_2(x_2)$:

$$E[f_1(x_1)|x_2=c] \triangleq \int_{-\infty}^{\infty} f_1(x_1)\cdot p(x_1|x_2=c)\cdot dx_1 \quad \left(= E[f_1(x_1)]_{p(x_1|x_2=c)}\right)$$

$$E[f_2(x_2)|x_1=d] \triangleq \int_{-\infty}^{\infty} f_2(x_2)\cdot p(x_2|x_1=d)\cdot dx_2 \quad \left(= E[f_2(x_2)]_{p(x_2|x_1=d)}\right)$$

Conditional Mean and Covariance

- For random vectors with some defined joint distribution, can thus also compute **conditional mean vector** and **conditional covariance matrix**

for $\underbrace{x_1 \in \mathbb{R}^{n_1} \;\&\; x_2 \in \mathbb{R}^{n_2}}_{\text{random vectors}}$ & $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^{n}$ $(n = n_1 + n_2)$

- Conditional mean vector of x_1 given $x_2 = c$: $m_{1|2} \triangleq E[x_1 | x_2 = c] = \int_{-\infty}^{\infty} x_1 \cdot p(x_1 | x_2 = c)\, dx_1 \in \mathbb{R}^{n_1}$

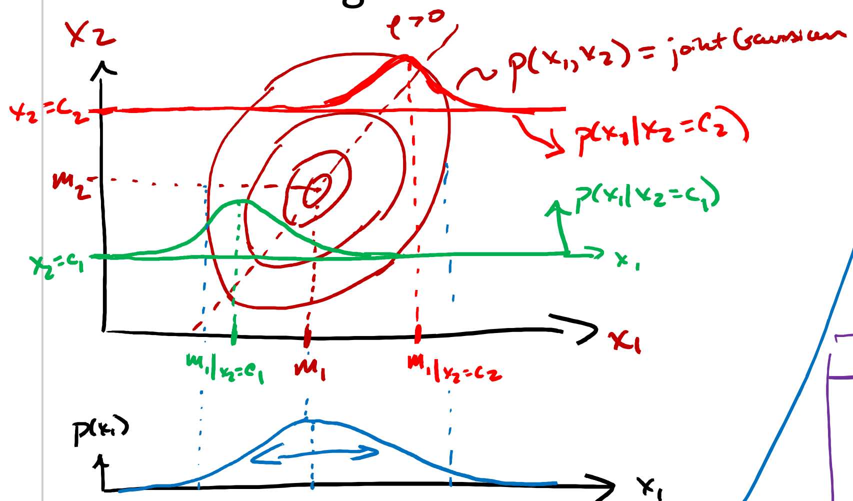
- conditional covariance matrix of x_1 given $x_2 = c$: $C_{1|2} \triangleq E[(x_1 - m_{1|2})(x_1 - m_{1|2})^T | x_2 = c]$

$$= \int_{-\infty}^{\infty} (x_1 - m_{1|2})(\ldots)^T \cdot p(x_1 | x_2 = c)\, dx_1 \in \mathbb{R}^{n_1 \times n_1}$$

$\rightarrow$ analogously can define $m_{2|1}$ & $C_{2|1}$ for cond. mean & covar. of x_2 given $x_1 = d$

Conditional Gaussian PDFs

- For a Gaussian random vector with multivariate Gaussian joint pdf over its elements, conditioning on some subset of vector elements leads to a **conditional Gaussian pdf**



i.e. let $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ for $x \in \mathbb{R}^n$, $x_1 \in \mathbb{R}^{n_1}$, $x_2 \in \mathbb{R}^{n_2}$

$$p(x) = p\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \mathcal{N}_x(m_x, C_x)$$

$$m_x = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}, \quad C_x = \left[\begin{array}{c:c} C_{11} & C_{12} \\ \hdashline C_{21} & C_{22} \end{array}\right], \quad \begin{array}{l} C_{11} \in \mathbb{R}^{n_1 \times n_1} \\ C_{12} \in \mathbb{R}^{n_1 \times n_2} \\ C_{21} = C_{12}^T \\ C_{22} \in \mathbb{R}^{n_2 \times n_2} \end{array}$$

Then: $p(x_1 | x_2 = a) = \mathcal{N}_{x_1|x_2}(m_{1|2}, C_{1|2})$ [cond. Gaussian pdf]

where cond. mean: $m_{1|2} = m_1 + C_{12} \cdot C_{22}^{-1}[a - m_2] \in \mathbb{R}^{n_1}$

(*) cond. covar: $C_{1|2} = C_{11} - C_{12} \cdot C_{22}^{-1} C_{21} \in \mathbb{R}^{n_1 \times n_1}$ (pos def)

aka Fundamental eqs. of linear est.

$\Rightarrow$ $\boxed{\text{(*) Conditioning a joint Gaussian pdf} \Rightarrow \text{another Gaussian!}}$

Derivation Sketch: Conditional Gaussian PDFs

- <u>Key idea:</u> start with definition of conditional pdf for vector random variables, then simplify and massage exponential term to look like multivariate Gaussian pdf over variables to left of conditioning sign (assuming conditioning vars are known/constant)

$$p(x_1|x_2=c) = \frac{p(x_1, x_2=c)}{p(x_2=c)} = \frac{\mathcal{N}_x(m_x, C_x)\big|_{x_2=c}}{\mathcal{N}_{x_2}(m_2, C_{22})\big|_{x_2=c}}$$

$$= \frac{\frac{1}{(2\pi)^{n/2}|C_x|^{1/2}} \cdot \exp\left\{-\frac{1}{2}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}\right)^T \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}^{-1} (\cdots)\right\}\Big|_{x_2=c}}{\frac{1}{(2\pi)^{n/2}|C_{22}|^{1/2}} \cdot \exp\left\{-\frac{1}{2}(x_2 - m_2)^T C_{22}^{-1} (\cdots)\right\}\Big|_{x_2=c}} = \frac{|C_{22}|^{1/2}}{|C_x|^{1/2}} \cdot \exp\left\{-\frac{1}{2}(*)\right\}$$

$$(*) = \left(\begin{bmatrix} x_1 \\ c \end{bmatrix} - \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}\right)^T \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} (\cdots) + (c - m_2)^T C_{22}^{-1} (\cdots)$$

$$\left(\text{where } \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}^{-1}\right)$$

$\Rightarrow$ expand as fxn of x_1 only & "completing the square"

$\Rightarrow$ cond. Gaussian pdf for $x_1 | x_2 = c$

- Will post supplementary material to Canvas for complete derivation…

Induction of Multivariate Gaussian PDFs

- Suppose we have random vectors x and y, where

$$\boxed{p(x) = \mathcal{N}_x(m_x, C_x)} \quad \& \quad y = Hx + w, \quad w \sim \mathcal{N}_w(0, R) \quad \& \quad w \perp\!\!\!\perp x$$

$$[x \in \mathbb{R}^n, \; y \in \mathbb{R}^p, \; w \in \mathbb{R}^p]$$

- This is actually equivalent to knowing p(x,y) = p(x)p(y|x) (i.e. factorization of joint pdf), since we can "induce" p(y|x) from the given information

Note: Since $y = Hx + w \xrightarrow{\text{for some known } x} y - Hx = w \longrightarrow p(w) = p(y - Hx)$

$$\longrightarrow p(w) = \mathcal{N}_w(0, R) \longrightarrow p_{y-Hx}(y - Hx) = \mathcal{N}_{y-Hx}(0, R)$$

$$= \text{const.} \cdot \exp\left\{-\tfrac{1}{2}(y - Hx)^T R^{-1} (\cdots)\right\}$$

$\longrightarrow$ if we are given x, then clearly $p(y|x) = \mathcal{N}_y(Hx, R)$

Induction of Multivariate Gaussian PDFs

- From knowing p(x) and p(y|x), we can further “induce” all required statistics for joint pdf:

$$p(x,y) = p\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \mathcal{N}(m, C), \quad m = \begin{bmatrix} m_x \checkmark \\ \underline{m_y} \end{bmatrix}, \quad C = \begin{bmatrix} C_{xx} \checkmark & C_{xy} \checkmark \\ \underline{C_{yx}} & \underline{C_{yy}} \end{bmatrix}$$

→ Already know m_x & C_{xx} are from $p(x) = \mathcal{N}_x(m_x, C_{xx})$

$$\rightarrow m_y = E[y] = E[Hx + w] = E[Hx] + E[\cancel{w}]_{0} = H \cdot E[x] = \boxed{Hm_x = m_y}$$

$$\rightarrow C_{xy} = E[(x-m_x)(y-m_y)^T] = E[(x-m_x)(Hx + w - Hm_x)^T]$$

$$= E[(x-m_x)\{(Hx - Hm_x) + w\}^T]$$

$$= E[(x-m_x)(H[x-m_x])^T] + \underbrace{\cancel{E[(x-m_x)w^T]}}_{0}$$

$$= E[(x-m_x)(x-m_x)^T H^T] \qquad = E[(x-m_x)] \cdot \cancel{E[w^T]} \quad b/c \; x \perp\!\!\!\perp w$$

$$= \underbrace{E[(x-m_x)(x-m_x)^T]}_{C_{xx}} \cdot H^T = \boxed{C_{xx} H^T = C_{xy}} \quad (= C_{yx}^T)$$

$$\rightarrow \text{Similarly: can show: } C_{yy} = E[(y-m_y)(\ldots)^T] = E[(Hx + w - Hm_y)(\ldots)^T] = (\ldots) = \boxed{\begin{gathered} = C_{yy} \\ HC_x H^T + R \end{gathered}}$$

Upshot: Bayes' Rule for Multivariate Gaussian PDFs

- Thus, starting from p(x) and p(y|x), <u>can immediately write down Bayes' posterior pdf</u> p(x|y) <u>using formulas for conditional Gaussian mean and covariance of x given y</u>:

$$p(x|y) = \mathcal{N}_{x|y}(m_{x|y}, C_{x|y}) \quad \left(= \frac{p(x)\cdot p(y|x)}{\int_{-\infty}^{\infty} p(x)\,p(y|x)\,dx} \propto p(x)\cdot p(y|x) \right)$$

$$\boxed{m_{x|y} = m_x + C_{xx} H^T \left[H C_x H^T + R \right]^{-1} \cdot (y - H m_x)}$$

$$\boxed{C_{x|y} = C_{xx} - C_{xy} C_{yy}^{-1} C_{yx} = C_{xx} - C_{xx} H^T \left[H C_x H^T + R \right]^{-1} H \cdot C_{xx}}$$