Exercise 11

February 6, 2021

1. Consider the time-series model

   $x_{t+1} = Ax_t + w, w \sim \mathcal{N}(0, Q)$
   $y_t = Cx_t + v, v \sim \mathcal{N}(0, R)$

   Where w, v are i.i.d. Gaussian noise variables. Further, assume that $p(x_0) = \mathcal{N}(\mu_0, \Sigma_0)$

   (a) What is the form of $p(x_0, x_1, .., x_T))$? Justify your answer (you do not have to explicitly compute the joint distribution).

   (b) Assume that $p(x_t|y_1, ..., y_t) = \mathcal{N}(\mu_t, \Sigma_t)$

      i. Compute $p(x_{t+1}|y_1, ..., y_t)$

      ii. Compute $p(x_{t+1}, y_{t+1}|y_1, ..., y_t)$

      iii. At time $t + 1$, we observe the value $y_{t+1} = \hat{y}$. Compute the conditional distribution $p(x_{t+1}|y_1, ..., y_{t+1})$

   Solution:

   (a) $\mathbf{x}_{t+1}$ is obtained from $\mathbf{x}_t$ by a linear transformation, $\mathbf{Ax}_t$ and adding a Gaussian random variabme $\mathbf{w}$. Initial distribution for $\mathbf{x}_0$ is a Gaussian distribution, a linear transformation of a Gaussian random variable is also a Gaussian random variable, whareas a sum of Gaussian random variables is a Gaussian random variable. Thus, the joint distribution $p(\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_T)$ is also a Gaussian distribution.

(b) i. Let $\mathbf{z} = \mathbf{A}\mathbf{x}_{t+1}$. Since this is a linear transformation of a Gaussian random variable, $\mathbf{x}_t \sim \mathcal{N}(\mu_t, \mathbf{\Sigma})$, then $\mathbf{z}$ is distributed as:

$$\mathbf{z} \sim \mathcal{N}(\mathbf{A}\mu_t, \mathbf{A}\mathbf{\Sigma}\mathbf{A}^T),$$

whereas the mean and the covariance of a sum of two Gaussian random variables are given by the sum of the means and the covariances of these variables, i.e.,

$$\mathbf{x}_{t+1} = \mathbf{z} + \mathbf{w} \sim \mathcal{N}(\mathbf{A}\mu_t, \mathbf{A}\mathbf{\Sigma}\mathbf{A}^T + \mathbf{Q}),$$

That is:

$$p(\mathbf{x}_{t+1}|\mathbf{y}_1, ..., \mathbf{y}_t) = \mathcal{N}(\mathbf{x}_{t+1}|\mathbf{A}\mu_t, \mathbf{A}\mathbf{\Sigma}\mathbf{A}^T + \mathbf{Q}).$$

ii. If we assume that $\mathbf{x}_{t+1}$ is fixed, then $\mathbf{y}_{t+1} = \mathbf{C}\mathbf{x}_{t+1} + \mathbf{v}$ follows the same distribution as $\mathbf{v}$, but with the mean shifted by $\mathbf{C}\mathbf{x}_{t+1}$, i.e.

$$p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \mathbf{y}_1, ..., \mathbf{y}_t) = \mathcal{N}(\mathbf{y}_{t+1}|\mathbf{C}\mathbf{x}_{t+1}, \mathbf{R}).$$

The the joint probability is obtained as

$$p(\mathbf{y}_{t+1}, \mathbf{x}_{t+1}|\mathbf{y}_1, ..., \mathbf{y}_t) = p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \mathbf{y}_1, ..., \mathbf{y}_t)p(\mathbf{x}_{t+1}|\mathbf{y}_1, ..., \mathbf{y}_t) = \mathcal{N}(\mathbf{y}_{t+1}|\mathbf{C}\mathbf{x}_{t+1}, \mathbf{R})\mathcal{N}(\mathbf{x}_{t+1}|\mathbf{A}\mu_t, \mathbf{A}\mathbf{\Sigma}\mathbf{A}^T + \mathbf{Q}).$$

iii. Let us introduce temporary notation

$$\mu_{t+1} = \mathbf{A}\mu_t,$$
$$\mathbf{\Sigma}_{t+1} = \mathbf{A}\mathbf{\Sigma}\mathbf{A}^T + \mathbf{Q},$$
$$p(\mathbf{x}_{t+1}|\mathbf{y}_1, ..., \mathbf{y}_t) = \mathcal{N}(\mu_{t+1}, \mathbf{\Sigma}_{t+1})$$

Then $\mathbf{y}_{t+1}$ is obtained in terms of the parameters of distribution $p(\mathbf{x}_{t+1}|\mathbf{y}_1, ..., \mathbf{y}_t)$ following the same steps as question 1), with the result

$$p(\mathbf{y}_{t+1}|\mathbf{y}_1, ..., \mathbf{y}_t) = \mathcal{N}(\mathbf{y}_{t+1}|\mathbf{C}\mu_{t+1}, \mathbf{C}\mathbf{\Sigma}_{t+1}\mathbf{C}^T + \mathbf{R}) = \mathcal{N}\left(\mathbf{y}_{t+1}|\mathbf{C}\mathbf{A}\mu_t, \mathbf{C}(\mathbf{A}\mathbf{\Sigma}\mathbf{A}^T +\right.$$

The required conditional distribution is then obtained as

$$p(\mathbf{x}_{t+1}|\mathbf{y}_1, ..., \mathbf{y}_t, \mathbf{y}_{t+1}) = \frac{p(\mathbf{y}_{t+1},\mathbf{x}_{t+1}|\mathbf{y}_1,...,\mathbf{y}_t)}{p(\mathbf{y}_{t+1}|\mathbf{y}_1,...,\mathbf{y}_t)} = \frac{\mathcal{N}(\mathbf{y}_{t+1}|\mathbf{C}\mathbf{x}_{t+1},\mathbf{R})\mathcal{N}(\mathbf{x}_{t+1}|\mathbf{A}\mu_t,\mathbf{A}\mathbf{\Sigma}\mathbf{A}^T+\mathbf{Q})}{\mathcal{N}(\mathbf{y}_{t+1}|\mathbf{C}\mathbf{A}\mu_t,\mathbf{C}(\mathbf{A}\mathbf{\Sigma}\mathbf{A}^T+\mathbf{Q})\mathbf{C}^T+\mathbf{R})}$$

2. Consider a Gaussian random variable $x \sim \mathcal{N}(x|\mu_x, \Sigma_x)$ where $x \in R^D$ Furthermore, we have
$y = Ax + b + w$
where $y \in R^E, A \in R^{E\times D}, b \in R^E$, and $w \sim \mathcal{N}(w|0, Q)$ is independent Gaussian noise. "Independent" implies that x and w are independent random variables and that Q is diagonal.

   (a) Write down the likelihood $p(y|x)$

   (b) The distribution $p(y) = \int p(y|x)p(x)dx$ is Gaussian. Compute the mean μ_y and the covariance Σ_y. Derive your result in detail.

   (c) The random variable y is being transformed according to measurement mapping $z = Cy + v$, where $z \in R^F, C \in R^{F\times E}, v \sim \mathcal{N}(v|0, R)$ is independent Gaussian noise.

      i. Write down $p(z|y)$

      ii. Compute the mean μ_z and the covariance Σ_z. Derive your result in detail.

   (d) Now, a value $\hat{y}$ is measured. Compute the posterior distribution $p(x|\hat{y})$

Solution

(a) If $\mathbf{x}$ is fixed, then $\mathbf{y}$ has the same distribution as $\mathbf{w}$, but with the mean shifter by $\mathbf{Ax} + \mathbf{b}$, that is

$$p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}|\mathbf{Ax} + \mathbf{b}, \mathbf{Q})$$

(b) Let us consider random variable $\mathbf{u} = \mathbf{Ax}$, it is distributed according to

$$p(\mathbf{u}) = \mathcal{N}(\mathbf{u}|\mathbf{A}\mu_x, \mathbf{A}\mathbf{\Sigma}_x\mathbf{A}^T).$$

Then $\mathbf{y}$ is a sum of two Gaussian random variables $\mathbf{u}$ and $\mathbf{w}$ with its mean additionally shifted by $\mathbf{b}$, that is

$$p(\mathbf{y}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\mu_x + \mathbf{b}, \mathbf{A}\mathbf{\Sigma}_x\mathbf{A}^T + \mathbf{Q}),$$

that is

$\mu_y = \mathbf{A}\mu_x + \mathbf{b},$
$\mathbf{\Sigma}_y = \mathbf{A}\mathbf{\Sigma}_x\mathbf{A}^T + \mathbf{Q}.$

(c) Like in b), assuming that $\mathbf{y}$ is fixed we obtain the conditional distribution

$$p(\mathbf{z}|\mathbf{y}) = \mathcal{N}(\mathbf{z}|\mathbf{C}\mathbf{y}, \mathbf{R})$$

Since $\mathbf{C}\mathbf{y}$ is a Gausssian random variable with distribution $\mathcal{N}(\mathbf{C}\mu_y, \mathbf{C}\mathbf{\Sigma}_y\mathbf{C}^T)$ we obtain the distribution of $\mathbf{z}$ as that of a sum of two Gaussian random variables:

$$p(\mathbf{z}) = \mathcal{N}(\mathbf{z}|\mathbf{C}\mu_y, \mathbf{C}\mathbf{\Sigma}_y\mathbf{C}^T + \mathbf{R}) = \mathcal{N}(\mathbf{z}|\mathbf{C}(\mathbf{A}\mu_x + \mathbf{b}), \mathbf{C}(\mathbf{A}\mathbf{\Sigma}_x\mathbf{A}^T + \mathbf{Q})\mathbf{C}^T + \mathbf{R})$$

(d) The posterior distribution $p(\mathbf{x}|\mathbf{y})$ can be obtained by applying the Bayes' theorem:

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} = \frac{\mathcal{N}(\mathbf{y}|\mathbf{A}\mathbf{x}+\mathbf{b},\mathbf{Q})\mathcal{N}(\mathbf{x}|\mu_x,\mathbf{\Sigma}_x)}{\mathcal{N}(\mathbf{y}|\mathbf{A}\mu_x+\mathbf{b},\mathbf{A}\mathbf{\Sigma}_x\mathbf{A}^T+\mathbf{Q})}$$

3. Given a continuous random variable X, with cdf $F_X(x)$. Show that the random variable $Y := F_X(X)$ is uniformly distributed.

Solution:

Cdf is related to pdf as

$F_x(x) = \int_{-\infty}^{x} dx' f_x(x'),$
$\frac{d}{dx}F_x(x) = f_x(x)$

and changes in the interval $[0, 1]$.
The pdf of variable $y = F_x(x)$ then can be defined as $f_y(y) = f_x(x)\left|\frac{dx}{dy}\right| = \frac{f_x(x)}{\left|\frac{dy}{dx}\right|} = \frac{f_x(x)}{\left|\frac{dF_x(x)}{dx}\right|} = \frac{f_x(x)}{f_x(x)} = 1,$

i.e. y is uniformly distributed in interval $[0, 1]$.