UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Homework 13
Spring 2023

1. **Balls in Bins Estimation**

   You throw n balls into m bins, where $n \geq 1$ and $m \geq 2$. Each ball lands in each bin with the same probability, independently of all other events. Let X and Y be the number of balls in bin 1 and 2 respectively.

   a. What is $\mathbb{E}(Y \mid X)$?

   b. Define $\mathbb{Q}(Y \mid X)$ to be the best quadratic function in X that minimizes mean squared error when used to estimate Y. Without doing any mathematical work, what are $\mathbb{L}(Y \mid X)$ and $\mathbb{Q}(Y \mid X)$? Justify your answer.

   c. Your friend from UCLA who hasn't learned about the Hilbert space of random variables isn't convinced by your explanation. Use the formula

   $$\mathbb{L}(Y \mid X) = \mathbb{E}(Y) + \frac{\operatorname{cov}(X, Y)}{\operatorname{var}(X)}(X - \mathbb{E}(X))$$

   to calculate the LLSE and verify your claim.

Solution:

a. $\mathbb{E}(Y \mid X = x) = \frac{n-x}{m-1}$, because conditioned on x balls landing in bin 1, the remaining $n - x$ balls are distributed uniformly among the other $m - 1$ bins. Thus

$$\mathbb{E}(Y \mid X) = \frac{n - X}{m - 1}.$$

b. By part a, $\mathbb{E}(Y \mid X)$ is a linear function of X. Since the best estimator of Y given X is linear, it must also be the best *linear* and *quadratic* estimator of Y given X, i.e. $\mathbb{E}(Y \mid X)$, $\mathbb{L}(Y \mid X)$, and $\mathbb{Q}(Y \mid X)$ all coincide.

c. $X, Y \sim \text{Binomial}(n, \frac{1}{m})$, so $\mathbb{E}(Y) = \frac{n}{m}$ and $\operatorname{var}(X) = n(\frac{1}{m})(1 - \frac{1}{m})$. Now let X_i be the indicator that ball i falls in bin 1 and Y_j the indicator that ball j falls in bin 2. Then, by the bilinearity of covariance,

$$\begin{aligned}
\operatorname{cov}(X, Y) &= \sum_{i=1}^{n} \operatorname{cov}(X_i, Y_i) + \sum_{i \neq j} \cancel{\operatorname{cov}(X_i, Y_j)} \\
&= \sum_{i=1}^{n} \cancel{\mathbb{E}(X_i Y_i)} - \mathbb{E}(X_i)\,\mathbb{E}(Y_i) \\
&= -\frac{n}{m^2}.
\end{aligned}$$

Plugging into the formula,

$$\begin{aligned}\mathbb{L}(Y \mid X) &= \frac{n}{m} + \frac{-\frac{n}{m^2}}{n(\frac{1}{m})(1-\frac{1}{m})}\left(X - \frac{n}{m}\right)\\ &= \frac{n}{m} - \frac{1}{m-1}\left(X - \frac{n}{m}\right)\\ &= \frac{mn - n - mX + n}{m(m-1)} = \frac{n-X}{m-1},\end{aligned}$$

which indeed equals the MMSE.

2. **Basic Properties of Jointly Gaussian Random Variables**

Let $(X_1, \ldots, X_n)$ be a collection of jointly Gaussian random variables with mean vector μ and covariance matrix Σ. Their joint density is given by, for $x \in \mathbb{R}^n$,

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\left\{-\frac{1}{2}(x-\mu)^\mathsf{T}\Sigma^{-1}(x-\mu)\right\}.$$

a. Show that $X_1, \ldots, X_n$ are independent if and only if they are pairwise uncorrelated.

b. Show that any linear combination of $X_1, \ldots, X_n$ will also be a Gaussian random variable. *Hint*: Consider using moment-generating functions.

Solution:

a. Independence implies uncorrelatedness in general, so suppose $X_1, \ldots, X_n$ are pairwise uncorrelated, in which case Σ is a diagonal matrix with diagonal entries $\sigma_1^2, \ldots, \sigma_n^2$. Then

$$\begin{aligned} f(x) &= \frac{1}{\sqrt{(2\pi)^n \prod_{i=1}^n \sigma_i^2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^n \frac{1}{\sigma_i^2}(x_i-\mu_i)^2\right\} \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{-\frac{1}{2\sigma_i^2}(x_i-\mu_i)^2\right\} \\ &= \prod_{i=1}^n f_{X_i}(x_i), \end{aligned}$$

so pairwise uncorrelatedness implies the independence of $X_1, \ldots, X_n$ jointly Gaussian.

b. The moment-generating function of a linear combination $Y = a^\mathsf{T} X = \sum_{i=1}^n a_i X_i$ is

$$\begin{aligned} \phi_Y(t) &= \mathbb{E}(\exp\{ta^\mathsf{T} X\}) = \phi_X(ta) \\ &= \exp\left\{ta^\mathsf{T}\mu - \frac{1}{2}t^2 a^\mathsf{T}\Sigma a\right\}. \end{aligned}$$

Therefore Y is Gaussian with distribution $\mathcal{N}(a^\mathsf{T}\mu, a^\mathsf{T}\Sigma a)$.

3. **Gaussian Random Vector MMSE**

Consider the Gaussian random vector

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\right),$$

and define the sign of Y to be the random variable

$$W = \begin{cases} 1 & \text{if } Y > 0 \\ 0 & \text{if } Y = 0 \\ -1 & \text{if } Y < 0 \end{cases}.$$

a. Find $\mathbb{E}(WX \mid Y)$.

b. Is the LLSE $\mathbb{L}(WX \mid Y)$ the same as the MMSE you found in part a?

c. Are WX and Y jointly Gaussian?

Solution:

a. As W is a function of Y, $\mathbb{E}(WX \mid Y) = W\,\mathbb{E}(X \mid Y)$. As X, Y are jointly Gaussian,

$$\mathbb{E}(X \mid Y) = \mathbb{L}(X \mid Y) = 1 + \frac{1}{2}Y.$$

Putting these two equations together,

$$\mathbb{E}(WX \mid Y) = \begin{cases} 1 + \frac{1}{2}Y & \text{if } Y > 0 \\ 0 & \text{if } Y = 0 \\ -1 - \frac{1}{2}Y & \text{if } Y < 0. \end{cases}$$

b. No, the LLSE and MMSE differ. The LLSE is a *linear* function of Y, whose coefficient of Y is constant, whereas the coefficient of Y in the MMSE varies with its sign.

c. By part b, WX and Y are not jointly Gaussian, because the LLSE and MMSE coincide for jointly Gaussian random variables.

4. **MMSE for Jointly Gaussian**

Find $\mathbb{E}(X \mid Y, Z)$ for the jointly Gaussian random vector

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \sim \mathcal{N}\left(\mu = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},\ \Sigma = \begin{bmatrix} 5 & 3 & 1 \\ 3 & 9 & 3 \\ 1 & 3 & 2 \end{bmatrix}\right).$$

Solution: Since everything is zero-mean, we have that

$$\begin{aligned}
\mathbb{E}(X \mid Y, Z) &= \Sigma_{X,(Y,Z)} \Sigma_{(Y,Z)}^{-1} \begin{bmatrix} Y \\ Z \end{bmatrix} \\
&= \begin{bmatrix} \mathbb{E}(XY) & \mathbb{E}(XZ) \end{bmatrix} \begin{bmatrix} \mathbb{E}(Y^2) & \mathbb{E}(YZ) \\ \mathbb{E}(YZ) & \mathbb{E}(Z^2) \end{bmatrix}^{-1} \begin{bmatrix} Y \\ Z \end{bmatrix} \\
&= \begin{bmatrix} 3 & 1 \end{bmatrix} \begin{bmatrix} 9 & 3 \\ 3 & 2 \end{bmatrix}^{-1} \begin{bmatrix} Y \\ Z \end{bmatrix} \\
&= \frac{1}{3} Y.
\end{aligned}$$

5. **Even-Times Kalman Filter**

Consider a random process $(X_n)_{n\in\mathbb{N}}$ with state space model

$$\begin{aligned} X_{n+1} &= aX_n + V_n, \quad V_n \sim_{\text{i.i.d.}} \mathcal{N}(0, \sigma_V^2) \\ Y_n &= X_n + W_n, \quad W_n \sim_{\text{i.i.d.}} \mathcal{N}(0, \sigma_W^2) \end{aligned}$$

where $(V_n)_{n\in\mathbb{N}}$ and $(W_n)_{n\in\mathbb{N}}$ are independent. We can only observe the process at even times, i.e. we observe the random variables $Y_0, Y_2, Y_4, \ldots$.

a. Derive a recurrence relation for the estimator $\hat{X}_{2n|2n} := \mathbb{L}(X_{2n} \mid Y_0, Y_2, \ldots, Y_{2n})$ in terms of $\hat{X}_{2n-2|2n-2}$.

b. Derive a recurrence relation for $\hat{X}_{2n+1|2n}$ in terms of $\hat{X}_{2n|2n}$.

Solution:

a. The even-times state transition model is given by

$$\begin{aligned} X_{2n+2} &= aX_{2n+1} + V_{2n+1} \\ &= a^2 X_{2n} + (aV_{2n} + V_{2n+1}), \end{aligned}$$

where the new noise terms $aV_{2n} + V_{2n+1} \sim \mathcal{N}(0, (a^2+1)\sigma_V^2)$ are also independent. Thus, we can rewrite the Kalman filter equations for the updated model:

$$\begin{aligned} \hat{X}_{2n+2|2n+2} &= a^2 \hat{X}_{2n|2n} + K_{2n+2}\tilde{Y}_{2n+2} \\ \tilde{Y}_{2n+2} &= Y_{2n+2} - a^2 \hat{X}_{2n|2n}, \end{aligned}$$

where the Kalman gain is given by

$$\begin{aligned} K_{2n+2} &= \frac{\sigma^2_{2n+2|2n}}{\sigma^2_{2n+2|2n} + \sigma_W^2} \\ \sigma^2_{2n+2|2n} &= (a^2\sigma^2_{2n|2n})^2 + (a^2+1)\sigma_V^2 \\ \sigma^2_{2n+2|2n+2} &= (1 - K_{2n+2})\sigma^2_{2n+2|2n}. \end{aligned}$$

b. By the linearity of the LLSE,

$$\hat{X}_{2n+1|2n} = a\hat{X}_{2n|2n}.$$

6. **Kalman Filter with Correlated Noise**

Consider the state space model

$$\begin{aligned} X_n &= aX_{n-1} + V_n \\ Y_n &= X_n + V_n, \end{aligned}$$

with $X_0 = 0$ and $(V_n)_{n\geq 0} \sim_{\text{i.i.d.}} \mathcal{N}(0,1)$. Note that the observation noise is the same as the process noise V_n, not independent of it, so this is different from the usual Kalman filter model. Derive recursive update equations for $\hat{X}_{n|n} := \mathbb{L}(X_n \mid Y_0, \ldots, Y_n)$.

Hint: You may use the fact that the equations will be of the form

$$\begin{aligned} \hat{X}_{n|n} &= a\hat{X}_{n-1|n-1} + K_n\tilde{Y}_n \\ \tilde{Y}_n &= Y_n - a\hat{X}_{n-1|n-1}, \end{aligned}$$

where you should find the Kalman gain and the estimator covariance recurrence relation

$$\begin{aligned} K_n &= \,? \\ \sigma^2_{n|n-1} &= a^2\sigma^2_{n-1|n-1} + 1 \\ \sigma^2_{n|n} &= \,?(\sigma^2_{n|n-1}). \end{aligned}$$

Solution: By the linearity of the LLSE and the innovation $\tilde{Y}_n \perp \operatorname{span}\{1, Y_0, \ldots, Y_{n-1}\}$,

$$\begin{aligned} \hat{X}_{n|n} &= \mathbb{L}(X_n \mid Y_0, \ldots, Y_n) \\ &= \mathbb{L}(aX_{n-1} \mid Y_0, \ldots, Y_{n-1}) + \mathbb{L}(X_n \mid \tilde{Y}_n) \\ &= a\hat{X}_{n-1|n-1} + \frac{\operatorname{cov}(X_n, \tilde{Y}_n)}{\operatorname{var}(\tilde{Y}_n)}\tilde{Y}_n. \end{aligned}$$

To find K_n, the coefficient of $\tilde{Y}_n$, we compute

$$\begin{aligned} \operatorname{cov}(X_n, \tilde{Y}_n) &= \operatorname{cov}(X_n, X_n + V_n - a\hat{X}_{n-1|n-1}) \\ &= \sigma^2_{n|n-1} + \operatorname{cov}(X_n, V_n) \\ &= \sigma^2_{n|n-1} + 1 \\ \operatorname{var}(\tilde{Y}_n) &= \operatorname{cov}(X_n, \tilde{Y}_n) + \operatorname{cov}(V_n, \tilde{Y}_n) \\ &= \sigma^2_{n|n-1} + 1 + \operatorname{cov}(V_n, X_n + V_n - a\hat{X}_{n-1|n-1}) \\ &= \sigma^2_{n|n-1} + 3. \end{aligned}$$

Now, to update the estimator covariance, we find

$$\begin{aligned} \sigma^2_{n|n} &= \operatorname{var}(X_n - \hat{X}_{n|n}) \\ &= \operatorname{var}(X_n - a\hat{X}_{n-1|n-1}) - 2\operatorname{cov}(X_n - a\hat{X}_{n-1|n-1}, K_n\tilde{Y}_n) + \operatorname{var}(K_n\tilde{Y}_n) \\ &= \operatorname{var}(X_n - a\hat{X}_{n-1|n-1}) - 2\operatorname{cov}(X_n, K_n\tilde{Y}_n) + K_n^2\operatorname{var}(\tilde{Y}_n) \\ &= \sigma^2_{n|n-1} - 2K_n\operatorname{cov}(X_n, \tilde{Y}_n) + K_n\operatorname{cov}(X_n, \tilde{Y}_n) \end{aligned}$$

$$
\begin{aligned}
&= \sigma^2_{n|n-1} - K_n(\sigma^2_{n|n-1} + 1) \\
&= (1 - K_n)\sigma^2_{n|n-1} - K_n.
\end{aligned}
$$

In short, we have the new update equations

$$
\begin{aligned}
K_n &= \frac{\sigma^2_{n|n-1} + 1}{\sigma^2_{n|n-1} + 3}, \\
\sigma^2_{n|n} &= (1 - K_n)\sigma^2_{n|n-1} - K_n.
\end{aligned}
$$