UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Homework 12
Fall 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. **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.

3. **Geometric MMSE**

   Let N be a geometric random variable with parameter $1-p$, and $(X_i)_{i\in\mathbb{N}}$ be i.i.d. exponential random variables with parameter λ. Let $T = X_1 + \cdots + X_N$. Compute the LLSE and MMSE of N given T.

   Hint: Compute the MMSE first.

   Solution: First, we calculate $\Pr(N = n \mid T = t)$, for $t > 0$ and $n \in \mathbb{Z}_+$.

$$\begin{aligned}
\Pr(N = n \mid T = t) &= \frac{\Pr(N = n) f_{T|N}(t \mid n)}{\sum_{k=1}^{\infty} \Pr(N = k) f_{T|N}(t \mid k)} \\
&= \frac{(1-p)p^{n-1}\lambda^n t^{n-1} e^{-\lambda t}/(n-1)!}{\sum_{k=1}^{\infty}(1-p)p^{k-1}\lambda^k t^{k-1} e^{-\lambda t}/(k-1)!} \\
&= \frac{\lambda(\lambda pt)^{n-1}/(n-1)!}{\lambda\sum_{k=1}^{\infty}(\lambda pt)^{k-1}/(k-1)!} = \frac{(\lambda pt)^{n-1}}{e^{\lambda pt}(n-1)!}, \qquad n \in \mathbb{Z}_+.
\end{aligned}$$

   Next, we calculate $\mathbb{E}[N \mid T = t]$.

$$\begin{aligned}
\mathbb{E}[N \mid T = t] &= \sum_{n=1}^{\infty} n \frac{(\lambda pt)^{n-1}}{e^{\lambda pt}(n-1)!} \\
&= \sum_{n=1}^{\infty} \frac{(\lambda pt)^{n-1}}{e^{\lambda pt}(n-1)!} + \sum_{n=1}^{\infty}(n-1)\frac{(\lambda pt)^{n-1}}{e^{\lambda pt}(n-1)!} \\
&= 1 + \frac{\lambda pt}{e^{\lambda pt}} \sum_{n=2}^{\infty} \frac{(\lambda pt)^{n-2}}{(n-2)!} = 1 + \frac{\lambda pt}{e^{\lambda pt}} e^{\lambda pt} = 1 + \lambda pt.
\end{aligned}$$

   Hence, the MMSE is $\mathbb{E}[N \mid T] = 1 + \lambda pT$. The MMSE is linear, so it is also the LLSE.

   In terms of a Poisson process, T represents the first arrival of a marked Poisson process with rate λ, where arrivals are marked independently with probability $1-p$. The marked Poisson process has rate $\lambda(1-p)$. The unmarked points form a Poisson process of rate λp. In time T, the expected number of unmarked points is λpT, so the conditional expectation of the number of points at time T, N, is $1 + \lambda pT$.

4. **Exam Difficulty**

The difficulty of an EECS 126 exam, Θ, is uniformly distributed on $[0, 100]$ (continuously). Alice gets a score X that is uniformly distributed on $[0, \Theta]$, and she wants to estimate the difficulty of the exam given her score.

a. What is the MLE of Θ? What is the MAP of Θ?

b. What is the LLSE for Θ?

Solution:

a. Since the prior on Θ is uniform, the MLE and MAP estimates will be the same. Both are equal to $\hat{\Theta} = X$, as

$$\operatorname*{argmax}_{\theta} f_{X|\Theta}(x \mid \theta) = \operatorname*{argmax}_{\theta} \frac{1}{\theta} \cdot \mathbb{1}_{x \le \theta \le 100} = x.$$

b. Recall that the LLSE of Θ given X can be found as

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

First, $\mathbb{E}(\Theta) = 50$ and $\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X \mid \Theta)) = \mathbb{E}(\frac{\Theta}{2}) = 25$. Let us find $\operatorname{var}(X)$ using the law of total variance and $\operatorname{cov}(\Theta, X) = \mathbb{E}(\Theta X) - \mathbb{E}(\Theta)\,\mathbb{E}(X)$:

$$\mathbb{E}(\operatorname{var}(X \mid \Theta)) = \mathbb{E}\Big(\frac{\Theta^2}{12}\Big) = \int_0^{100} \frac{\theta^2}{12} \cdot \frac{1}{100}\,\mathrm{d}\theta = \frac{10000}{36}.$$

$$\operatorname{var}(\mathbb{E}(X \mid \Theta)) = \operatorname{var}\Big(\frac{\Theta}{2}\Big) = \frac{1}{4}\frac{10000}{12} = \frac{10000}{48}.$$

$$\operatorname{var}(X) = \mathbb{E}(\operatorname{var}(X \mid \Theta)) + \operatorname{var}(\mathbb{E}(X \mid \Theta)) = \frac{70000}{144}.$$

$$\mathbb{E}(\Theta X) = \mathbb{E}(\mathbb{E}(\Theta X \mid \Theta)) = \mathbb{E}\Big(\frac{\Theta^2}{2}\Big) = \frac{10000}{6}.$$

$$\operatorname{cov}(\Theta, X) = \mathbb{E}(\Theta X) - \mathbb{E}(\Theta)\,\mathbb{E}(X) = \frac{1250}{3}.$$

Putting everything together, the LLSE is

$$\mathbb{L}(\Theta \mid X) = 50 + \frac{6}{7}(X - 25).$$

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:

$$\hat{X}_{2n+2|2n+2} = a^2 \hat{X}_{2n|2n} + K_{2n+2}\tilde{Y}_{2n+2}$$

$$\tilde{Y}_{2n+2} = Y_{2n+2} - a^2 \hat{X}_{2n|2n},$$

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 $\hat{X}_{n|n} := \mathbb{L}(X_n \mid Y_0, \dots, Y_n)$.

   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 $\tilde{Y}_n \perp \text{span}\{1, Y_0, \dots, Y_{n-1}\}$,

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

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

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

   Now, to update the estimator covariance, we find

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

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

In short, we have the new update equations

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