UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Problem Set 14 (Optional)
Fall 2021

1. **Balls in Bins Estimation**

We throw $n \geq 1$ balls into $m \geq 2$ bins. Let X and Y represent the number of balls that land in bin 1 and 2 respectively.

(a) Calculate $\mathbb{E}[Y \mid X]$.

(b) What are $L[Y \mid X]$ and $Q[Y \mid X]$ (where $Q[Y \mid X]$ is the best quadratic estimator of Y given X)?
Hint: Your justification should be no more than two or three sentences, no calculations necessary! Think carefully about the meaning of the MMSE.

(c) Unfortunately, your friend is not convinced by your answer to the previous part. Compute $\mathbb{E}[X]$ and $\mathbb{E}[Y]$.

(d) Compute $\text{var}(X)$.

(e) Compute $\text{cov}(X, Y)$.

(f) Compute $L[Y \mid X]$ using the formula. Ensure that your answer is the same as your answer to part (b).

Solution:

(a) $\mathbb{E}[Y \mid X = x] = (n - x)/(m - 1)$, because once we condition on x balls landing in bin 1, the remaining $n - x$ balls are distributed uniformly among the other $m - 1$ bins. Therefore,

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

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

(c) Let X_i be the indicator that the ith ball falls in bin 1. Then, $X = \sum_{i=1}^{n} X_i$, and by linearity of expectation, $\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = n/m$, since there are n indicators and each ball has a probability $1/m$ of landing in bin 1. By symmetry, $\mathbb{E}[Y] = n/m$ as well.

(d) The number of balls that falls into the first bin is binomially distributed with parameters n and $1/m$. Hence the variance is $n(1/m)(1 - 1/m)$.

(e) Let X_i be as before, and let Y_i be the indicator that the ith ball falls into bin 2.

$$\text{cov}(X, Y) = \sum_{i=1}^{n} \sum_{j=1}^{n} \text{cov}(X_i, Y_j)$$

We can compute $\text{cov}(X_i, Y_i) = \mathbb{E}[X_i Y_i] - \mathbb{E}[X_i]\mathbb{E}[Y_i] = 0 - (1/m)(1/m) = -1/m^2$ (note that $\mathbb{E}[X_i Y_i] = 0$ because it is impossible for a ball to land in both bins 1 and 2). Also, we have $\text{cov}(X_i, Y_j) = 0$ because the indicator for the ith ball is independent of the indicator for the jth ball when $i \neq j$. Hence, $\text{cov}(X, Y) = n(-1/m^2) = -n/m^2$.

(f)

$$\begin{aligned}
L[Y \mid X] &= \mathbb{E}[Y] + \frac{\operatorname{cov}(X,Y)}{\operatorname{var}(X)}(X - \mathbb{E}[X]) \\
&= \frac{n}{m} + \frac{-n/m^2}{n(1/m)(1-1/m)}\Big(X - \frac{n}{m}\Big) \\
&= \frac{n}{m} - \frac{1}{m-1}\Big(X - \frac{n}{m}\Big) \\
&= \frac{mn - n - mX + n}{m(m-1)} = \frac{n-X}{m-1}
\end{aligned}$$

2. MMSE and Conditional Expectation

Let $X, Y_1, \ldots, Y_n$ be square integrable random variables. The MMSE of X given $(Y_1, \ldots, Y_n)$ is defined as the function $\phi(Y_1, \ldots, Y_n)$ which minimizes the mean square error

$$\mathbb{E}[(X - \phi(Y_1, \ldots, Y_n))^2].$$

(a) For this part, assume $n = 1$. Show that the MMSE is precisely the conditional expectation $\mathbb{E}[X|Y]$. *Hint:* expand the difference as $(X - \mathbb{E}[X|Y] + \mathbb{E}[X|Y] - \phi(Y))$.

(b) Argue that

$$\mathbb{E}\big[(X - \mathbb{E}[X \mid Y_1, \ldots, Y_n])^2\big] \le \mathbb{E}\Bigg[\bigg(X - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X \mid Y_i]\bigg)^2\Bigg].$$

That is, the MMSE does better than the average of the individual estimates given each Y_i.

Solution:

(a) We write

$$\begin{aligned}
&\mathbb{E}[(X - \phi(Y))^2] \\
&\quad = \mathbb{E}[(X - \mathbb{E}[X|Y] + \mathbb{E}[X|Y] - \phi(Y))^2] \\
&\quad = \mathbb{E}[(X - \mathbb{E}[X|Y])^2] + 2\mathbb{E}[(X - \mathbb{E}[X|Y])(\mathbb{E}[X|Y] - \phi(Y))] + \mathbb{E}[(\mathbb{E}[X|Y] - \phi(Y))^2] \\
&\quad \ge \mathbb{E}[(X - \mathbb{E}[X|Y])^2] + 2\mathbb{E}[(X - \mathbb{E}[X|Y])(\mathbb{E}[X|Y] - \phi(Y))].
\end{aligned}$$

Now, if we let $f(Y) = \mathbb{E}[X|Y] - \phi(Y)$, notice that the cross term vanishes:

$$\begin{aligned}
\mathbb{E}[(X - \mathbb{E}[X|Y])f(Y)] &= \mathbb{E}[Xf(Y)] - \mathbb{E}[\mathbb{E}[X|Y]f(Y)] \\
&= \mathbb{E}[Xf(Y)] - \mathbb{E}[\mathbb{E}[Xf(Y)|Y]] \\
&= \mathbb{E}[Xf(Y)] - \mathbb{E}[Xf(Y)] = 0.
\end{aligned}$$

Therefore we've shown

$$\mathbb{E}[(X - \phi(Y))^2] \ge \mathbb{E}[(X - \mathbb{E}[X|Y])^2]$$

for any ϕ. Hence $\mathbb{E}[X|Y]$ is the MMSE.

(b) Notice that $\sum_{i=1}^n \mathbb{E}[X \mid Y_i]$ is a function of $Y_1, \dots, Y_n$. The argument is true, since $\mathbb{E}[X \mid Y_1, \dots, Y_n]$ is the MMSE estimate of X among all functions of $Y_1, \dots, Y_n$. Written out explicitly,

$$\begin{aligned}
&\mathbb{E}\left[\left(X - \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X \mid Y_i]\right)^2\right] \\
&\quad = \mathbb{E}\left[\left(X - \mathbb{E}[X \mid Y_1, \dots, Y_n] + \mathbb{E}[X \mid Y_1, \dots, Y_n] - \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X \mid Y_i]\right)^2\right] \\
&\quad = \mathbb{E}[(X - \mathbb{E}[X \mid Y_1, \dots, Y_n])^2] + \mathbb{E}\left[\left(\mathbb{E}[X \mid Y_1, \dots, Y_n] - \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X \mid Y_i]\right)^2\right] \\
&\quad \ge \mathbb{E}[(X - \mathbb{E}[X \mid Y_1, \dots, Y_n])^2],
\end{aligned}$$

where in the second equality we expanded the square, and used the orthogonality property to vanish the cross-term.

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 $P(N = n \mid T = t)$, for $t > 0$ and $n \in \mathbb{Z}_+$.

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

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

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

Hence, the MMSE is $\mathbb{E}[N \mid T] = 1 + \lambda p T$. 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 $\lambda p T$, so the conditional expectation of the number of points at time T, N, is $1+\lambda p T$.

4. **Gaussian Random Vector MMSE**

Let
$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\right)$$
be a Gaussian random vector.

Let
$$W = \begin{cases} 1, & \text{if } Y > 0 \\ 0, & \text{if } Y = 0 \\ -1, & \text{if } Y < 0 \end{cases}$$
be the sign of Y. Find $\mathbb{E}[WX \mid Y]$. Is the LLSE the same as the MMSE?

Solution:

First of all notice that since W is a function of Y we have that
$$\mathbb{E}[WX \mid Y] = W\mathbb{E}[X \mid Y].$$

Now since X, Y are jointly Gaussian we have that
$$\mathbb{E}[X \mid Y] = L[X \mid Y] = 1 + \frac{Y}{2}.$$

All in all
$$\mathbb{E}[WX \mid Y] = \begin{cases} 1 + \frac{Y}{2}, & \text{if } Y > 0 \\ 0, & \text{if } Y = 0 \\ -1 - \frac{Y}{2}, & \text{if } Y < 0. \end{cases}$$

No, the LLSE and MMSE are different in this case. Recall that the LLSE is a *linear* function of Y, in particular:
$$L[WX|Y] = \frac{\text{Cov}(WX, Y)}{\text{Var}(Y)} Y + \mathbb{E}[WX].$$
Since the coefficient of Y is just some constant, this cannot be the same as the MMSE, which varies with $|Y|$ (note that we do not have to explicitly compute any of the moments here).

5. **Gaussian Sine**

Let X, Y, Z be jointly Gaussian random variables with covariance matrix
$$\begin{bmatrix} 4 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 4 \end{bmatrix}$$
and mean vector $[0, 2, 0]$. Compute $\mathbb{E}[(\sin X) Y (\sin Z)]$. *Hint:* Condition on (X, Z).

Solution:

Conditioning on (X, Z), we have by tower property that

$$\mathbb{E}[\sin(X)Y\sin(Z)] = \mathbb{E}[\mathbb{E}[\sin(X)Y\sin(Z)|X,Z]] = \mathbb{E}[\sin(X)\sin(Z)\mathbb{E}[Y|X,Z]]. \tag{1}$$

The inner conditional expectation can be computed as

$$\begin{aligned}\mathbb{E}[Y|X,Z] &= \mu_Y + \Sigma_{Y,(X,Z)}\Sigma_{(X,Z)}^{-1}\begin{bmatrix}X\\Z\end{bmatrix}\\ &= 2 + [1\ 1]\begin{bmatrix}1/4 & 0\\ 0 & 1/4\end{bmatrix}\begin{bmatrix}X\\Z\end{bmatrix}\\ &= 2 + \frac{1}{4}X + \frac{1}{4}Z.\end{aligned}$$

Then (1) becomes (we use the fact that X and Z are uncorrelated and therefore independent, and since sin is odd and X and Z are symmetric about 0, $\mathbb{E}[\sin(X)] = \mathbb{E}[\sin(Z)] = 0$):

$$\begin{aligned}&\mathbb{E}[\sin(X)\sin(Z)(2 + \frac{1}{4}X + \frac{1}{4}Z)]\\ &= 2\mathbb{E}[\sin(X)]\mathbb{E}[\sin(Z)] + \mathbb{E}[\sin(X)X/4]\mathbb{E}[\sin(Z)] + \mathbb{E}[\sin(X)]\mathbb{E}[\sin(Z)Z/4]\\ &= 0.\end{aligned}$$

6. Error of the Kalman Filter for a Linear Stochastic System

The linear stochastic system

$$\begin{bmatrix}X_{1,k+1}\\X_{2,k+1}\end{bmatrix} = \begin{bmatrix}2 & 1\\1 & 2\end{bmatrix}\begin{bmatrix}X_{1,k}\\X_{2,k}\end{bmatrix} + \begin{bmatrix}1\\-1\end{bmatrix}w_k, \qquad k \geq 0,$$

starts from an arbitrary (known) initial condition $\begin{bmatrix}x_{1,0}\\x_{2,0}\end{bmatrix}$ and the system noise variables $(w_k,\ k \geq 0)$ are i.i.d. normal with mean 0 and variance 1.

The state variables are not directly observable. However, we can observe

$$Y_k = X_{1,k} + X_{2,k}, \qquad k \geq 0.$$

Let $\hat{X}_{k|k}$ denote the minimum mean square error estimator of $X_k = \begin{bmatrix}X_{1,k}\\X_{2,k}\end{bmatrix}$ given $(Y_0, \ldots, Y_k)$. Determine the asymptotic behavior of the covariance matrix of the estimation error.

Note: This problem needs thought. Note that there is no observation noise, so the assumption used in the derivation of the Kalman filter equations, that the covariance matrix of the observation noise is positive definite, is no longer valid.

Solution:

First, we observe that

$$\begin{bmatrix}2 & 1\\1 & 2\end{bmatrix}\begin{bmatrix}1 & -1\\-1 & 1\end{bmatrix}\begin{bmatrix}2 & 1\\1 & 2\end{bmatrix} = \begin{bmatrix}2 & 1\\1 & 2\end{bmatrix}\begin{bmatrix}1 & -1\\-1 & 1\end{bmatrix} = \begin{bmatrix}1 & -1\\-1 & 1\end{bmatrix}.$$

We claim by induction that for each $k \in \mathbb{N}$,

$$\operatorname{var} X_k = k \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

The base case is true since X_0 is deterministic, and supposing that the claim holds true for $k \in \mathbb{N}$, then by independence of X_k and w_k,

$$\begin{aligned} \operatorname{var} X_{k+1} &= \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} (\operatorname{var} X_k) \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} + \operatorname{var} w_k \\ &= k \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} + \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \\ &= k \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} = (k+1) \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}. \end{aligned}$$

With the claim now established, we see that $\operatorname{var} Y_k = (k+1)^2(1+1-2) = 0$, so Y_k is a constant for all $k \in \mathbb{N}$; hence, $\hat{X}_{k|k} = \mathbb{E}[X_k]$. Thus, the covariance matrix of the estimation error at time k is simply $\operatorname{var} X_k$, and as $k \to \infty$ the covariance matrix blows up.