## ECE 673-Random signal analysis I Midterm: solution

Q1 (1 point). An exhausted mother with a inclination for mathematics decides to pick the name of her newborn daughter by choosing 4 letters at random from the english alphabet (there are 26 possible letters). What is the probability that she will choose LISA?
Sol.: The sample space $\mathcal{S}$ contains $N_{\mathcal{S}}=26^{4}=456976$ simple events and the probability of each simple event is $1 / N_{\mathcal{S}}=2.2 \cdot 10^{-6}$. Therefore the probability of choosing "LISA" is $2.2 \cdot 10^{-6}$.
Q2 (1 point). Two fair dices are tossed. Find the probability that the same number will be observed on both dices. Next, find the probability that different numbers will be observed. Sol.: The sample space $\mathcal{S}=\left\{\left(z_{1}, z_{2}\right): z_{i} \in\{1,2,3,4,5,6\}\right\}$ contains $N_{\mathcal{S}}=6^{2}=36$ simple events and the probability of each simple event is $1 / N_{\mathcal{S}}=1 / 36$. The probability of observing two different numbers can be obtained as

$$
P[\text { different }]=1-P[\text { same }]=1-N / N_{\mathcal{S}},
$$

where $N$ is the number of simple events containing two equal numbers: $N=6$. It follows that

$$
P[\text { different }]=1-P[\text { same }]=1-6 / 36=5 / 6
$$

Q3 (1 point). NJ transit has a new bus running in Newark. The new bus runs on Sunday, Tuesday and Thursday while the old bus runs the other days. The new bus has a probability of being on time $3 / 5$ while the older bus has a probability of only $1 / 5$. If a passenger chooses an arbitrary day of the week to ride the bus, what is the probability that the bus will be on time?
Sol.: Let us define the events:

$$
\begin{aligned}
N & =\text { the passenger rides the new bus } \\
T & =\text { the bus is on time. }
\end{aligned}
$$

We have

$$
\begin{aligned}
P[T \mid N] & =3 / 5 \\
P\left[T \mid N^{c}\right] & =1 / 5
\end{aligned}
$$

and

$$
\begin{aligned}
P[N] & =3 / 7 \\
P\left[N^{c}\right] & =4 / 7
\end{aligned}
$$

Therefore, from the law of total probability we obtain

$$
P[T]=P[T \mid N] P[N]+P\left[T \mid N^{c}\right] P\left[N^{c}\right]=3 / 5 \cdot 3 / 7+1 / 5 \cdot 4 / 7=13 / 35
$$

Q4 (1 point). A drunk person wanders aimlessly on the road. At each step, he can either move forwards or backwards. He will move in one direction (either forward or backward) with probability $1 / 4$ if he has moved in the same direction at the previous step. Moreover,

the first step is certainly forward (i.e., with probability 1 ). Draw the corresponding Markov chain and find the probabilty that the drunk person moves one step forward, then one step backward, then again one step backward and finally one step forward.
Sol.: The Markov chain is depicted in the figure below ("F" represents a step forward and " B " backward). The desired probability is

$$
\begin{aligned}
P[F, B, B, F] & =P[F] P[B \mid F] P[B \mid B] P[F \mid B]= \\
& =1 \cdot 3 / 4 \cdot 1 / 4 \cdot 3 / 4=9 / 64
\end{aligned}
$$

D1 (2 points). A famous skier takes part in ten different competitions at the Olympic winter games. In each competion he has a probability equal to 0.2 to get the gold medal (and the outcomes of different competitions are assumed to be independent).
(i) If $X$ is a random variable counting the number of gold medals won by the skier, what is the probability mass function of $X$ ? Evaluate the probability that the skier will win only one gold medal.
(ii) What is the probability that the first gold medal will be won by the skier at the third competition?
(iii) Now consider a second skier, less skilled but very longevous. He gets to compete in his whole career in 1000 races, having a probability 0.005 to win in each race. Let $Y$ be a random variable counting the number of gold medals won by the second skier. Propose a reasonable approximation for the probability mass function of $Y$.
(iv) Evaluate the approximate and exact probability that $Y=2$ and compare the two probabilities.
Sol.: The problem can be modelled as independent Bernoulli trials with probability $p=0.2$. (i) The random variable $X$ is distributed as $X \sim \operatorname{bin}(10,0.2)$. The requested probability is thus

$$
P[X=1]=\binom{10}{1} 0.2 \cdot(0.8)^{9}=0.27
$$

(ii) Defining the random variable $Z=\{$ index of the first competition where the skier wins $\}$, we have that $Z \sim \operatorname{geom}(0.2)$. Therefore, the requested probability reads

$$
P[Z=3]=0.2 \cdot(0.8)^{2}=0.128
$$

(iii) A reasonable approximation is $Y \sim \operatorname{Pois}(\lambda)$ with $\lambda=1000 \cdot 0.005=5$.
(iv) Since the actual distribution of $Y$ is $\operatorname{bin}(1000,0.005)$, the real probability at hand is

$$
\begin{aligned}
P[Y & =2]=\binom{1000}{2}(0.005)^{2}(0.995)^{998}= \\
& =8.39 \times 10^{-2}
\end{aligned}
$$

On the other hand, the approximate probability (assuming that $Y \sim \operatorname{Pois}(5)$ ) reads

$$
P[Y=2]=\frac{5^{2}}{2!} \exp (-5)=8.42 \times 10^{-2}
$$

The approximation is reasonably good since the number of trials is large (1000) and the probability of success in each trial is small (0.005).
D2 (2 points). Write a MATLAB program that estimates mean and variance of a random variable $Y=5 X$, where $X \sim \exp (2)$, and compares the estimates to the true values $E[Y]$ and $\operatorname{var}(Y)$. (Hint: generate $X$ using the Inverse Probability Integral method)
Sol.: In order to use the Inverse Probability Integral, we need to calculate the inverse CDF of the exponential random variable $X$. Since the CDF reads

$$
F_{X}(x)=\left\{\begin{array}{cc}
0 & x \leq 0 \\
1-\exp (-2 x) & x>0
\end{array}\right.
$$

1. we have

$$
F_{X}^{-1}(u)=\left\{\begin{array}{cc}
-\frac{1}{2} \log (1-u) & 0 \leq u \leq 1 \\
0 & \text { otherwise }
\end{array} .\right.
$$

Now, the following code can be used in order to estimate mean and variance of $Y$ :
$\mathrm{N}=1000$; \%number of Monte Carlo iterations
$\mathrm{x}=\mathrm{zeros}(\mathrm{N}, 1)$;
for $\mathrm{i}=1: \mathrm{N} \%$ for each iteration
$u=\operatorname{rand}(1)$;
$\mathrm{x}(\mathrm{i})=-1 / 2^{*} \log (1-\mathrm{u})$;
$y(\mathrm{i})=5^{*} \mathrm{x}(\mathrm{i})$;
end
meanestimate $=\operatorname{sum}(y) / \mathrm{N}$
truemean $=5^{*} 1 / 2$
varestimate $=\operatorname{sum}\left(\mathrm{y} .{ }^{\wedge} 2\right) / \mathrm{N}-(\text { meanestimate })^{\wedge} 2$
truevar $=5^{\wedge} 2^{*} 1 / 2^{\wedge} 2$
D3 (2 points). Two discrete random variables $X$ and $Y$ are defined as follows. The PMF of $X$ is

$$
p_{X}[k]=\left\{\begin{array}{cc}
1 / 4 & k=1 \\
3 / 8 & k=2 \\
3 / 8 & k=3
\end{array}\right.
$$

and the conditional of $Y$ given $X$ is $p_{Y \mid X}[k \mid j]$ defined as

| $Y \backslash X$ | 1 | 2 | 3 |
| :---: | :---: | :---: | :---: |
| 2 | 1 | 0 | 0 |
| 3 | 0 | $2 / 3$ | $1 / 3$ |
| 4 | 0 | $1 / 3$ | $2 / 3$ |.

(i) Find the joint PMF of $X$ and $Y, p_{X, Y}[k, j]$.
(ii) Evaluate mean and covariance matrix of the random vector $\mathbf{Z}=[X Y]^{T}$.
(iii) Say that we are interested in estimating $Y$. If $X$ is not known (only the PMF of $Y$ is known), then what is the best predictor for $Y$ ? What is the corresponding error?
(iv) Now, if $X$ is known, what is the best linear predictor of $Y(\hat{Y}=a X+b)$ ? What is the corresponding error? Provide an explanation in terms of regression lines.
Sol.: (i) The joint PMF of $X$ and $Y$ reads $p_{X, Y}[k, j]=p_{Y \mid X}[k \mid j] \cdot p_{X}[j]$

| $Y \backslash X$ | 1 | 2 | 3 |
| :---: | :---: | :---: | :---: |
| 2 | $1 / 4$ | 0 | 0 |
| 3 | 0 | $1 / 4$ | $1 / 8$ |
| 4 | 0 | $1 / 8$ | $1 / 4$ |.

(ii) The mean vector is

$$
E[\mathbf{Z}]=\left[\begin{array}{c}
E[X] \\
E[Y]
\end{array}\right]=\left[\begin{array}{l}
17 / 8 \\
25 / 8
\end{array}\right]
$$

since

$$
E[X]=1 \cdot 1 / 4+2 \cdot 3 / 8+3 \cdot 3 / 8=17 / 8
$$

and

$$
\begin{aligned}
p_{Y}[k] & = \begin{cases}1 / 4 & k=2 \\
3 / 8 & k=3 \\
3 / 8 & k=4\end{cases} \\
E[Y] & =2 \cdot 1 / 4+3 \cdot 3 / 8+4 \cdot 3 / 8=25 / 8 .
\end{aligned}
$$

On the other hand, the covariance matrix reads

$$
\mathbf{C}_{\mathbf{Z}}=\left[\begin{array}{cc}
\operatorname{var}(X) & \operatorname{cov}(X, Y) \\
\operatorname{cov}(X, Y) & \operatorname{var}(Y)
\end{array}\right]=\left[\begin{array}{cc}
39 / 64 & 31 / 64 \\
31 / 64 & 39 / 64
\end{array}\right]
$$

since

$$
\begin{aligned}
\operatorname{var}(X) & =E\left[X^{2}\right]-E[X]^{2}=\left(1^{2} \cdot 1 / 4+2^{2} \cdot 3 / 8+3^{2} \cdot 3 / 8\right)-(17 / 8)^{2}= \\
& =39 / 64 \\
\operatorname{var}(Y) & =E\left[Y^{2}\right]-E[Y]^{2}=\left(2^{2} \cdot 1 / 4+3^{2} \cdot 3 / 8+4^{2} \cdot 3 / 8\right)-(25 / 8)^{2}= \\
& =39 / 64 \\
\operatorname{cov}(X, Y) & =E[X Y]-E[X] E[Y]= \\
& =1 \cdot 2 \cdot 1 / 4+2 \cdot 3 \cdot 1 / 4+3 \cdot 3 \cdot 1 / 8+4 \cdot 2 \cdot 1 / 8+4 \cdot 3 \cdot 1 / 4-(17 / 8) \cdot(25 / 8)= \\
& =31 / 64 . .
\end{aligned}
$$


(iii) If only the PMF of $Y$ is known, the best predictor $\hat{Y}$ is the average $E[Y]=25 / 8$ and the corresponding mean square error is the variance $\operatorname{var}(Y)=39 / 64$.
(iv) The best linear predictor of $Y$ given $X$ is

$$
\begin{aligned}
\hat{Y} & =E[Y]+\frac{\operatorname{cov}(X, Y)}{\operatorname{var}(X)}(X-E[X])= \\
& =25 / 8+\frac{31 / 64}{39 / 64}(X-17 / 8)= \\
& =31 / 39 \cdot X+56 / 39=0.79 \cdot X+1.43 .
\end{aligned}
$$

The mean square error the best linear predictor reads

$$
\begin{aligned}
m s e & =\operatorname{var}(Y)-\frac{\operatorname{cov}(X, Y)^{2}}{\operatorname{var}(X)}= \\
& =39 / 64-\frac{(31 / 64)^{2}}{39 / 64}=35 / 156=0.22<\operatorname{var}(Y)=0.6
\end{aligned}
$$

The regression line is shown in the figure below.

