ECE 313
Final Exam

July 25, 2012

Exam Time : 120 mins

Problem 1

(7×2 pts) Cars are passing a certain point of a road according to a Poisson process with arrival rate 1 car per minute.

For each part below, circle the letter of the correct answer and fill in the blanks for your selection. Do not circle more than one answer.

a) Let X denote the number of cars passing in the first two minutes. What type of random variable is X?

i) Binomial, $n =$_______, $p =$_______

ii) Poisson, $\lambda = 2$

iii) Exponential, $\lambda =$_______

iv) Geometric, $p =$_______

v) Erlang, $r =$_______, $\lambda =$_______

b) Let T_5 denote the time of the arrival of the fifth car. What type of random variable is T_5?

i) Binomial, $n =$_______, $p =$_______

ii) Poisson, $\lambda =$_______

iii) Exponential, $\lambda =$_______

iv) Geometric, $p =$_______

v) Erlang, $r = 5$, $\lambda = 1$

c) Let S denote the sum of the number of cars passing in the first 10 seconds and the number of cars passing in the fifth minute. What type of random variable is S?

i) Binomial, $n =$_______, $p =$_______

ii) Poisson, $\lambda = 1 + \frac{1}{6}$

iii) Exponential, $\lambda =$_______

iv) Geometric, $p =$_______

v) Erlang, $r =$_______, $\lambda =$_______

d) Let M denote the number of cars passing in the first five minutes and let N be the number of cars passing in the first 15 minutes. What describes the type of the conditional pdf of M given that $N = 18$? That is, what describes $p_{M|N}(M|18)$?

i) Binomial, $n = 18$, $p = 1/3$
ii) Poisson, $\lambda =$_______
iii) Exponential, $\lambda =$_______
iv) Geometric, $p =$_______
v) Erlang, $r =$_______, $\lambda =$_______

e) A hitchhiker arrives at time $t = 60$ min. Let W be the amount of time that the hitchhiker waits until a cars passes. What type of random variable is W?

i) Binomial, $n =$_______, $p =$_______
ii) Poisson, $\lambda =$_______
iii) Exponential, $\lambda = 1$
iv) Geometric, $p =$_______
v) Erlang, $r =$_______, $\lambda =$_______

f) The hitchhiker of the previous part is superstitious and only asks for a ride if the car is red. Each passing car is red with probability 0.1 independently of its arrival time and the color of other cars. Let Q denote the number of cars passing until a red car passes (including the red car itself). What is the type of Q?

i) Binomial, $n =$_______, $p =$_______
ii) Poisson, $\lambda =$_______
iii) Exponential, $\lambda =$_______
iv) Geometric, $p = 0.1$
v) Erlang, $r =$_______, $\lambda =$_______

g) Each passing red car stops for the hitchhiker of the previous two parts with probability 0.05, independently of other cars and its arrival time. Let R be the number of cars passing until a red car stops (including the car that stops). What is the type of R?

i) Binomial, $n =$_______, $p =$_______
ii) Poisson, $\lambda =$_______
iii) Exponential, $\lambda = 0.1 \cdot 0.05 = 0.005$
iv) Geometric, $p =$_______
v) Erlang, $r =$_______, $\lambda =$_______

Problem 2

(3×5 pts) Consider random variables $X \sim N(1,1)$ and $Y \sim N(0,3^2)$. Suppose X and Y are jointly Gaussian and that the correlation coefficient of X and Y is ρ. Determine the following. (Your answers should depend on ρ.)

a) $P(2X+3Y-2 \geq 2)$.

b) The best unconstrained estimator of $X+Y$ if Y is observed.

c) The MSE for the estimator of part (b).

Solution

a) $2X+3Y-2$ is Gaussian with mean

$$E[2X+3Y-2] = 2E[X] - 2 = 0$$

and variance

$$\begin{aligned}\mathrm{Var}[2X+3Y-2] &= \mathrm{Cov}(2X+3Y-2, 2X+3Y-2)\\ &= 4\mathrm{Var}[X] + 9\mathrm{Var}[Y] + 2\cdot 6\mathrm{Cov}(X,Y)\\ &= 4 + 9\times 9 + 12\rho\sigma_X\sigma_Y\\ &= 85 + 36\rho.\end{aligned}$$

So

$$\begin{aligned}P(2X+3Y-2 \geq 2) &= P\left(\frac{2X+3Y-2}{\sqrt{85+36\rho}} \geq \frac{2}{\sqrt{85+36\rho}}\right)\\ &= Q\left(\frac{2}{\sqrt{85+36\rho}}\right).\end{aligned}$$

b) $X+Y$ and Y are jointly Gaussian with covariance

$$\begin{aligned}\mathrm{Cov}(X+Y,Y) &= \mathrm{Cov}(X,Y) + \mathrm{Var}[Y]\\ &= \rho\sigma_X\sigma_Y + \sigma_Y^2 = 9 + 3\rho.\end{aligned}$$

Hence,

$$\begin{aligned}E[X+Y|Y] &= \hat{E}[X+Y|Y]\\ &= E[X+Y] + \frac{\mathrm{Cov}(X+Y,Y)}{\mathrm{Var}[Y]}(Y - E[Y])\\ &= 1 + \frac{9+3\rho}{9}Y = 1 + Y + \rho Y/3.\end{aligned}$$

Alternatively, we could write

$$\begin{aligned}E[X+Y|Y] &= Y + E[X|Y]\\ &= Y + E[X] + \sigma_X\rho\frac{Y-E[Y]}{\sigma_Y}\\ &= Y + 1 + \rho Y/3.\end{aligned}$$

c) The MSE is $\mathrm{Var}[X+Y](1-\rho_{X+Y,X})^2$. To find $\mathrm{Var}[X+Y]$ we write

$$\begin{aligned}\mathrm{Var}[X+Y] &= \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}(X,Y)\\ &= 1+9+2\rho 3 = 10+6\rho,\end{aligned}$$

and ρ_{X+Y} equals

$$\rho_{X+Y,Y} = \frac{\mathrm{Cov}(X+Y,Y)}{\sigma_Y\sqrt{\mathrm{Var}[X+Y]}} = \frac{9+3\rho}{3\sqrt{10+6\rho}} = \frac{3+\rho}{\sqrt{10+6\rho}}.$$

So

$$\mathrm{MSE} = (10+6\rho)\left(1-\frac{9+6\rho+\rho^2}{10+6\rho}\right) = 1-\rho^2.$$

Alternatively, we could write

$$\begin{aligned}\mathrm{MSE} &= E\left[((X+Y)-E[X+Y|Y])^2\right]\\ &= E\left[(X-E[X|Y])^2\right]\\ &= \mathrm{Var}[X]\left(1-\rho_{X,Y}^2\right) = 1-\rho^2.\end{aligned}$$

Problem 3

(3+5+5 pts) The accuracy of medical tests can be stated in terms of:

- Sensitivity = conditional probability that test is positive given that disease is present.
- Specificity = conditional probability that test is negative given that disease is not present.
- Probability of false positive = conditional probability that test is positive given that disease is not present.
- Probability of false negative = conditional probability that test is negative given disease is present.

a) True ☐ False ☐ A high sensitivity test may have a large probability of false positive.

For the following parts of the problem consider a certain HIV test for which sensitivity is equal to 99.7% and specificity is 98.5%. Furthermore, assume that 0.4% of the general population in the US is HIV positive.

b) Find the probability of error for this test if it is administered to a random member of the general population.

c) Suppose the test result is positive for a random person in the general population. What is the conditional probability that he or she is actually HIV positive?

Solution

a) True. For example consider a test whose result is always positive. This test has sensitivity one since it never misses. If the probability of disease being present is zero, then the probability of false positive is also one. High sensitivity tests are common in medicine because it is important not to miss the presence of dangerous diseases even at the cost of some false positive cases.

b) The probability of false positive is $1 - 0.985 = 0.015$ and the probability of false negative is $1 - .997 = 0.003$. The fraction of population that is not HIV positive is $1 - 0.004 = 0.996$. So

$$p_e = 0.996\,(0.015) + 0.004\,(0.003) = 0.014952 = 1.4\%.$$

Approximation: note that $1 - 0.004 \simeq 1$ and that $0.004\,(0.003)$ is much smaller than 0.015. So

$$p_e = (1 - 0.004)\,(0.015) + 0.004\,(0.003) \simeq 1\,(0.015) + 0 = 0.015 = 1.5\%.$$

This approximation shows that the main component of the error comes from false positives.

c) Let H^+ denote the event that the person is HIV positive and H^- denote the event that the person is HIV negative. Also, let T^+ be the event that the test result is positive.

$$\begin{aligned} P\left(H^+|T^+\right) &= \frac{P\left(T^+|H^+\right)P\left(H^+\right)}{P\left(T^+|H^+\right)P\left(H^+\right) + P\left(T^+|H^-\right)P\left(H^-\right)} \\ &= \frac{0.997(0.004)}{0.997(0.004) + 0.015\,(0.996)} = 0.211 = 21.1\%. \end{aligned}$$

Note that this probability is rather small. The main reason is that a priori probability of HIV positive is very small, that is, 0.004.
Approximation: note that $0.997 \simeq 0.996 \simeq 1$. So

$$P\left(H^+|T^+\right) = \frac{0.997(0.004)}{0.997(0.004) + 0.015\,(0.996)} \simeq \frac{0.004}{0.004 + 0.015} = \frac{4}{19} = 0.210 = 21.0\%.$$

Problem 4

(23 pts) Let R be the triangle with vertices $(-1, 0), (0, 1)$, and $(1, 0)$. Also, let X and Y be jointly continuous with joint pdf

$$f_{X,Y}(x, y) = \begin{cases} 1, & (x, y) \in R \\ 0, & (x, y) \notin R \end{cases}$$

a) (4 pts) Are X and Y independent? Justify your answer: if they are independent find the marginals, and if they are not independent provide a reason.

b) (4 pts) Find $P(Y \le |X|)$.

c) (3 pts) Determine the set of values for x such that $f_{Y|X}(y|x)$ is defined.

d) (4 pts) For the set of values of x in part (c), find $f_{Y|X}(y|x)$ for $-\infty < y < \infty$.

e) (4 pts) Find the best linear estimator of Y if X is observed.

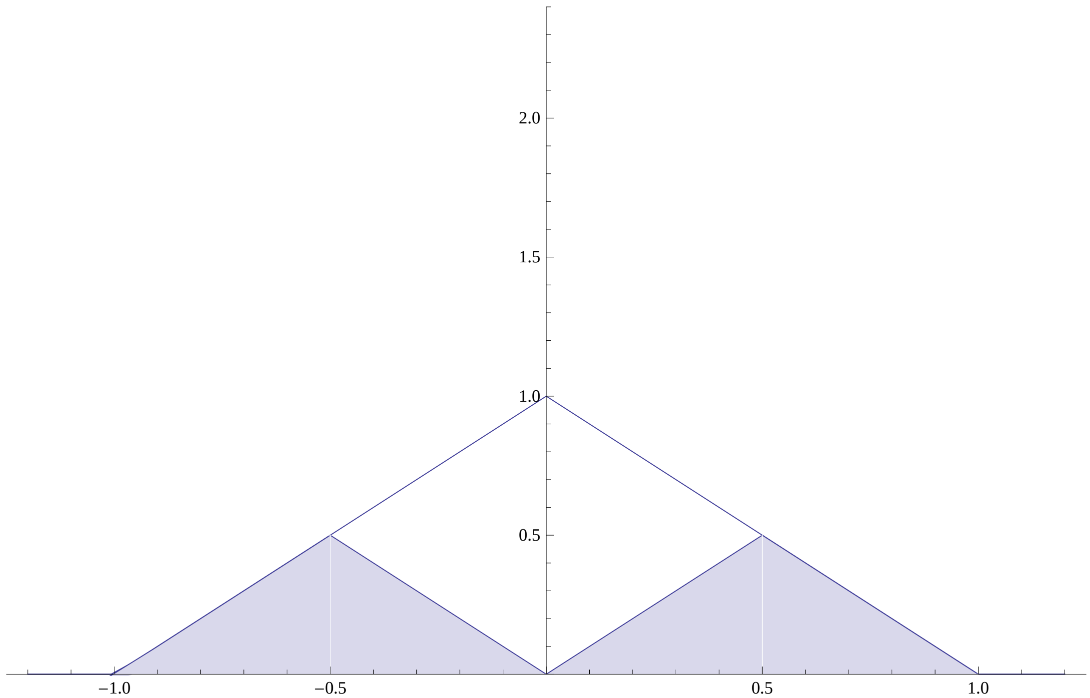
f) (4 pts) Find the MSE for the estimator in part (e).

Solution

a) X and Y are not independent as the support is not a product set.

b) The probability is the ratio of the shaded area to the area of R.

$$P\left(Y \leq |X|\right) = \frac{2\left(\frac{1}{2}\cdot 1\cdot \frac{1}{2}\right)}{\frac{1}{2}\cdot 2\cdot 1} = \frac{1}{2}.$$



c) $f_{Y|X}\left(y|x\right)$ is defined whenever $f_X\left(x\right) > 0$. So $f_{Y|X}\left(y|x\right)$ is defined for $\{x : -1 < x < 1\}$.

d) For $-1 < x < 0$,

$$f_X\left(x\right) = \int_0^{x+1} dy = 1 + x = 1 - |x|$$

and for $0 \leq x < 1$,

$$f_X\left(x\right) = \int_0^{1-x} dy = 1 - x = 1 - |x|\,.$$

So, for $-1 < x < 1$,

$$f_{Y|X}\left(y|x\right) = \begin{cases} \frac{1}{1-|x|}, & 0 \leq y \leq 1 - |x| \\ 0, & y < 0 \text{ or } y > 1 - |x|\,. \end{cases}$$

e) First, we find the covariance of X and Y. Since $E[X]=0$, we have $\mathrm{Cov}(X,Y)=E[XY]$.

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \int_{-1}^{0}\int_{0}^{1+x} xydydx + \int_{0}^{1}\int_{0}^{1-x} xydydx \\ &= \int_{-1}^{0}\left(\frac{xy^2}{2}\right)_0^{1+x} dx + \int_{0}^{1}\left(\frac{xy^2}{2}\right)_0^{1-x} dx \\ &= \int_{-1}^{0}\frac{x(1+x)^2}{2}dx + \int_{0}^{1}\frac{x(1-x)^2}{2}dx \\ &= \int_{0}^{1}\frac{-x(1-x)^2}{2}dx + \int_{0}^{1}\frac{x(1-x)^2}{2}dx \\ &= 0.\end{aligned}$$

The fact that covariance equals zero could also be deduced from the fact that the support is symmetric with respect to the y axis.
Since $\mathrm{Cov}(X,Y)=0$, the best linear estimator is the constant estimator. So

$$\hat{E}[Y|X]=E[Y].$$

For $0\le y\le 1$,

$$f_Y(y)=\int_{y-1}^{1-y} dx = 2(1-y).$$

Thus

$$E[Y\hat{|}X] = E[Y] = \int_0^1 2y(1-y)\,dy = \left(y^2-\frac{2y^3}{3}\right)_0^1 = 1-\frac{2}{3}=\frac{1}{3}.$$

f) The MSE equals $\mathrm{Var}[Y]$. For $0\le y\le 1$,

$$E\left[Y^2\right]=\int_0^1 y^2 2(1-y)\,dy = \left(\frac{2y^3}{3}-\frac{2y^4}{4}\right)_0^1=\frac{1}{6}.$$

$$\text{MSE } = \mathrm{Var}[Y] = \frac{1}{6}-\frac{1}{9}=\frac{1}{18}.$$

Problem 5

(3×5 pts) This problem is concerned with counting and conditional probabilities.

a) The letters of the word “bookkeeper” are rearranged randomly.

   i) What the probability that the resulting word is again “bookkeeper”?

   ii) What is the conditional probability that the resulting word is “bookkeeper” given that the two k’s appear consecutively and the two o’s appear consecutively?

b) An integer is randomly chosen from the set $\{i : 1\le i\le 999\}$. What is the probability that all digits of the chosen number are even? (Hint: A possible approach is to first solve the problem for the case that the number is randomly chosen from the set $\{i : 0\le i\le 999\}$ instead).

Solution

a)

i) The total number of possible cases is

$$\frac{10!}{2!2!3!}$$

so the desired probability is

$$\frac{2!2!3!}{10!} = \frac{24}{10!}.$$

ii) Consider the two k's to be a single symbol and the two o's to be a single symbol. Then the total number of cases is $8!/3!$ so the desired probability is $3!/8!$.

b) There are $5 \times 5 \times 5 = 125$ numbers in the set $\{i : 0 \leq i \leq 999\}$ whose digits are all even. Since one of these numbers is 0, in the set $\{i : 1 \leq i \leq 999\}$, there are $125 - 1$ numbers whose digits are all even. So the desired probability is

$$\frac{124}{999}.$$

Problem 6

(6 + 4 pts) Suppose X is a uniform random variable over the interval $\left[0, \frac{3}{2}\right]$. Let Y be defined as $Y = \lfloor X \rfloor$, where $\lfloor x \rfloor$ is defined as the largest integer that is less than or equal to x.

a) Find the pmf of Y.

b) What is $E\left[Y|X = 1.25\right]$?

Solution

a) The support of Y is the set $\{0, 1\}$.

$$p_Y(0) = P(0 \leq X < 1) = \frac{1}{3/2} = \frac{2}{3},$$
$$p_Y(1) = P(1 \leq X) = \frac{1/2}{3/2} = \frac{1}{3}.$$

So

$$p_Y(j) = \begin{cases} \frac{2}{3}, & j = 0, \\ \frac{1}{3}, & j = 1, \\ 0, & \text{else.} \end{cases}$$

b) $E\left[Y|X = 1.25\right] = E\left[1|X = 1.25\right] = 1$.

Problem 7

(4+6 pts) Consider a 10-bit code designed as follows. Each codeword is of the form $x_1x_2x_3x_4p_1x_5x_6x_7x_8p_2$. For $1 \leq i \leq 8$, x_i are data bits. p_1 and p_2 are parity bits defined as

$$p_1 = x_1 \oplus x_2 \oplus x_3 \oplus x_4,$$
$$p_2 = x_5 \oplus x_6 \oplus x_7 \oplus x_8$$

where $\oplus$ denotes the XOR operator: $0 \oplus 0 = 1 \oplus 1 = 0$ and $0 \oplus 1 = 1 \oplus 0 = 1$.

a) What is the minimum number of bit errors in an undetected error pattern? Describe an error pattern with that many bit errors.

b) Suppose each bit is in error with probability $p = 0.001$, independently of the other bits. Using a union bound based on your answer to part (a), find an upper bound on the probability of undetected errors.

Solution

a) Two. Any single bit error is detected. If there are two bit errors but one occurs in the first five bits and the second occurs in the second five bits, both errors are detected. If there are two bit errors and both occur in the first five bits or both occur in the second five bits, the error is undetected. Examples of error patterns with two bit errors are: 100010000, 0000011000.

b) There must be at least two errors in the first five bits or at least two errors in the second five bits. Let $E_{i,j}$ denote the event that i'th and the j'th bit are erroneous. Let $P(E)$ denote the probability of undetected errors. So

$$\begin{aligned}
P(E) &= P\left((E_{1,2} \cup E_{1,3} \cup \cdots \cup E_{4,5}) \bigcup (E_{6,7} \cup E_{6,8} \cup \cdots \cup E_{9,10})\right) \\
&\leq (P(E_{1,2}) + P(E_{1,3}) + \cdots + P(E_{4,5})) + (P(E_{6,7}) + P(E_{6,8}) + \cdots + P(E_{9,10})) \\
&= \binom{5}{2} P(E_{1,2}) + \binom{5}{2} P(E_{6,7}) \\
&= 2\binom{5}{2} p^2 = 20 \times 10^{-6} = 2 \times 10^{-5}.
\end{aligned}$$