E4.10
C2.1
SC4

IMPERIAL COLLEGE LONDON

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
EXAMINATIONS 2007

MSc and EEE PART IV: MEng and ACGI

Corrected Copy

PROBABILITY AND STOCHASTIC PROCESSES

Thursday, 26 April 2:30 pm

Time allowed: 3:00 hours

There are SIX questions on this paper.

Answer FOUR questions.

All questions carry equal marks

Any special instructions for invigilators and information for candidates are on page 1.

Examiners responsible	First Marker(s) :	K.K. Leung
	Second Marker(s) :	R.B. Vinter

Special Instructions for Invigilator: **None**

Information for Students: **Complementary Normal Distribution**

$$Q(x) = 1 - \Phi(x) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dx$$

If needed, for any y different from all the x values given below, $Q(y)$ can be approximated by linear interpolation of the values of $Q(x)$ at the two x values closest to y.

x	$Q(x)$	x	$Q(x)$
0	5.00e-01	2.7	3.47e-03
0.1	4.60e-01	2.8	2.56e-03
0.2	4.21e-01	2.9	1.87e-03
0.3	3.82e-01	3.0	1.35e-03
0.4	3.45e-01	3.1	9.68e-04
0.5	3.09e-01	3.2	6.87e-04
0.6	2.74e-01	3.3	4.83e-04
0.7	2.42e-01	3.4	3.37e-04
0.8	2.12e-01	3.5	2.33e-04
0.9	1.84e-01	3.6	1.59e-04
1.0	1.59e-01	3.7	1.08e-04
1.1	1.36e-01	3.8	7.24e-05
1.2	1.15e-01	3.9	4.81e-05
1.3	9.68e-02	4.0	3.17e-05
1.3	8.08e-02	4.5	3.40e-06
1.5	6.68e-02	5.0	2.87e-07
1.6	5.48e-02	5.5	1.90e-08
1.7	4.46e-02	6.0	9.87e-10
1.8	3.59e-02	6.5	4.02e-11
1.9	2.87e-02	7.0	1.28e-12
2.0	2.28e-02	7.5	3.19e-14
2.1	1.79e-02	8.0	6.22e-16
2.2	1.39e-02	8.5	9.48e-19
2.3	1.07e-02	9.0	1.13e-19
2.4	8.20e-03	9.5	1.05e-21
2.5	6.21e-03	10.0	7.62e-24
2.6	4.66e-03		

1. a. A communication link transmits digits of 0 and 1. With probability 0.9, the receiver can correctly detect a digit 1 when it is in fact sent. However, with probability of 0.05, the receiver incorrectly detects a digit 1 when a digit 0 is in fact sent. If digits 0 and 1 are actually sent with probability 0.6 and 0.4, respectively, what is the probability the sent digit is 1, given that the receiver does detect a digit 1?

(Hint: Use Bayes' rule.) [9]

b. X and Y are two independent identically distributed (i.i.d.) scalar random variables with the probability density function (pdf)

$$f(t) = \frac{1}{W} \qquad 0 \le t \le W \text{ where } W \text{ is a constant.}$$

Define a new random variable $Z \equiv X - Y$. Find the pdf for Z. [7]

c. Let $\{X_i\}$ be a sequence of identically distributed mutually independent Bernoulli random variables with

$$P(X_i = 1) = p \quad \text{and} \quad P(X_i = 0) = 1 - p$$

where $0 < p < 1$. Let $S_N \equiv X_1 + X_2 + \dots + X_N$ where N is an integer-valued random variable with

$$P(N = m) = \frac{a^m e^{-a}}{m!} \quad \text{for } m = 0, 1, 2, \dots \text{ and } a > 0 \text{ is a constant.}$$

Show that the probability generating function (i.e., the z-transform) for S_N is

$$S^*(z) = e^{ap(z-1)}.$$

[9]

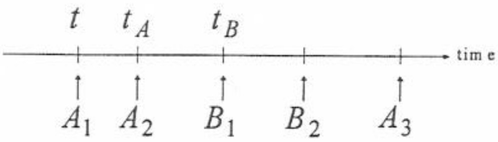
2. a. A scalar random variable X has the following probability density function (pdf)

$$f_X(t) = \frac{a}{2} e^{-a|t|} \qquad -\infty < t < \infty$$

where $a > 0$ is a constant. Derive the Laplace transform (L.T.) of $f_X(t)$. Hence, or otherwise, obtain the mean and variance of X.

[10]

b. Consider the merging of two independent Poisson processes A and B with their respective arrival rates of λ_A and λ_B. That is, the probability density functions (pdf's) for their interarrival times are given by $\lambda_A e^{-\lambda_A t}$ and $\lambda_B e^{-\lambda_B t}$, respectively. A realization of the merged arrival process is shown in the following figure.



A_i : The i^{th} arrival of Poisson process A and

B_i : The i^{th} arrival of Poisson process B.

Let us prove that the merged process is also a Poisson process as follows. Without loss of generality, consider that an arrival from either process A or B has just occurred at time t (e.g., A_1 in the above diagram). Further, let t_A and t_B be the time when the *first* arrival from process A and B arrive after time t, respectively.

i. Are the time durations $(t_A - t)$ and $(t_B - t)$ independent of the evolution of the processes A and B prior to time t? Why? [2]

ii. Determine the respective pdf's for the time durations $(t_A - t)$ and $(t_B - t)$. [3]

iii. Show that the pdf for the time duration from time t to the next arrival time, regardless of whether it arrives from process A or B, is given by

$$(\lambda_A + \lambda_B) e^{-(\lambda_A + \lambda_B)t}.$$

[7]

iv. Comment on the result for merging more than two Poisson arrival processes. [3]

3. a. Consider four scalar random variables, X, Y, U and V, which are related by the following linear transformation

$$U = aX + bY$$
$$V = cX + dY$$

where a, b, c and d are constants.

i. Show that $\sigma_U^2 = a^2\sigma_X^2 + 2ab\text{cov}[XY] + b^2\sigma_Y^2$ where σ_Z^2 denotes the variance for any given scalar random variable Z and $\text{cov}[XY]$ is the covariance of X and Y. [4]

ii. Show that $\text{cov}[UV] = ac\sigma_X^2 + (bc + ad)\text{cov}[XY] + bd\sigma_Y^2$. [4]

b. Suppose that two random processes, $X(t)$ and $Y(t)$, are jointly wide-sense stationary. Their cross-correlation functions are given by

$$R_{XY}(\tau) = E[X(t)Y(t+\tau)]$$
$$R_{YX}(\tau) = E[Y(t)X(t+\tau)]$$

where t and τ are arbitrary.

i. Prove that $R_{XY}(-\tau) = R_{YX}(\tau)$. [3]

ii. Prove that $|R_{XY}(\tau)| \le \frac{1}{2}[R_X(0) + R_Y(0)]$ where $R_X(0) = E[X^2(t)]$ and $R_Y(0) = E[Y^2(t)]$. [5]

(Hint: Consider $E[\{X(t) \pm kY(t+\tau)\}^2] \ge 0$ and then set k appropriately.)

iii. Prove that $|R_{XY}(\tau)| \le \sqrt{R_X(0)R_Y(0)}$. [7]

(Hint: Again consider $E[\{X(t) + kY(t+\tau)\}^2] \ge 0$ with appropriate k.)

iv. Comment on which one of the upper bounds for $|R_{XY}(\tau)|$ in parts ii and iii is tighter and why. [2]

4. a. A production line is designed to manufacture 8,200 Ω resistors. Collectively, the resistance (denoted by X) of the produced resistors can be modelled as a Gaussian random variable with a mean $\mu = 8{,}200\,\Omega$ and a standard deviation $\sigma = 615\,\Omega$. Following manufacture, all resistors are measured exactly. Resistors having values within 5% of 8,200 Ω (i.e., within $8{,}200 \pm 410\,\Omega$) are separated from the others to be marked and sold. The conditioning event A is defined as

$$A = \{\, 7{,}790\Omega < X \leq 8{,}610\Omega \,\}.$$

i. Using the complementary normal distribution given at the beginning of this examination paper, determine the probability of event A. [6]

ii. Give an expression for the standard deviation of the resistance given the condition A. (No need to carry out the calculation.) [3]

b. Consider that X is a scalar random variable with mean 0 and finite variance σ^2.

i. Prove that for any $a, b > 0$

$$P[X \geq a] \leq \frac{E[(X+b)^2]}{(a+b)^2} = \frac{\sigma^2 + b^2}{(a+b)^2}.$$ [12]

(Hint: Apply Markov's inequality.)

ii. By optimizing the value for b in the result of part i, prove that

$$P[X \geq a] \leq \frac{\sigma^2}{\sigma^2 + a^2}.$$ [4]

5. a. A random variable X is observed and used to predict the value of a second random variable Y. Let us choose the predictor for Y, denoted by $g(X)$, based on observations of X so that the mean squared error $E[(Y-g(X))^2]$ is minimized. Under this criterion, show that the best possible predictor of Y is $g(X)=E[Y \mid X]$. That is, prove the following

$$E[(Y-g(X))^2] \geq E[(Y-E[Y \mid X])^2]. \qquad [12]$$

(Hint: Consider $E[(Y-g(X))^2 \mid X]$ first.)

b. In a digital signal-processing system, raw continuous analog data X is characterized by a probability distribution and density functions, $F_X(x)$ and $f_X(x)$, respectively. X must be quantized to obtain a digital representation. To quantize the raw data X, an increasing sequence of numbers a_i for $i = 0, \pm 1, \pm 2, \ldots$, such that $\lim_{i \to \infty} a_i = \infty$ and $\lim_{i \to -\infty} a_i = -\infty$, is fixed and the raw data are then quantized according to the interval $(a_i, a_{i+1}]$ in which X lies. Let y_i be the discretized value when $X \in (a_i, a_{i+1}]$. Let Y denote the observed discretized value as

$$Y = y_i \quad \text{if} \quad X \in (a_i, a_{i+1}].$$

The distribution of Y is given by

$$P(Y = y_i) = F_X(a_{i+1}) - F_X(a_i).$$

Suppose now that we want to choose the values y_i, $i = 0, \pm 1, \pm 2, \ldots$ so as to minimize $E[(X-Y)^2]$, the expected mean square difference between the raw data and their quantized version.

i. Find the optimal values y_i, $i = 0, \pm 1, \pm 2, \ldots$ [8]

(Hint: Apply the best predictor in part a.)

ii. For the optimal quantizer Y, show that $E[Y] = E[X]$ so the mean square error quantizer preserves the input mean. [5]

6. Consider a communication channel with memory buffer to store and transmit data packets. The channel time is divided into slots and exactly one packet can be transmitted in each time slot. Packets arrive at the channel according to an independent renewal process $\{a_i, i = 0, 1, 2, ...\}$ where a_i denotes the probability that i packets arrive in a time slot. (Obviously, $\sum_{i=0}^{\infty} a_i = 1$.) Completion of a packet transmission and packet arrivals, if any, are assumed to occur immediately before and right at the end of a time slot, respectively. The buffer is limited in size and can store up to K packets, including the one being transmitted. New arriving packets fill the buffer to up its size limit and "overflow" packets finding the buffer fully occupied are lost from the system.

a. Draw a state transition diagram for the Markov chain representing the channel. [3]

b. Provide the state-transition probability matrix P for the Markov chain. [9]

(Hint: Be careful about the cases where more packets arriving than the buffer can hold.)

c. Denote the state probabilities at steady state by $\underline{\pi} = (\pi_0, \pi_1, \pi_2, ..., \pi_K)$ where π_i is the probability of having i packets in the system at the end of a time slot. Provide the equations from which $\underline{\pi}$ can be solved. (No need to solve them.) [3]

d. Define the link utilization ρ as the fraction of time during which the channel is transmitting packets at steady state. Express ρ in terms of $\{\pi_0, \pi_1, \pi_2, ..., \pi_K\}$. [3]

e. Obtain the probability P_o that packet overflow (loss) occurs in an arbitrary time slot in terms of $\{\pi_0, \pi_1, \pi_2, ..., \pi_K\}$ and $\{a_i, i = 0, 1, 2, ...\}$. [4]

f. Now, assume that the number of packets arriving in a time slot depends on the number of packets currently existing in the system. For example, the more packets existing in the system, the smaller the probability that new packets will arrive in the next time slot. Can the channel still be modelled as a Markov chain and why? [3]

Model answer ~~2006~~ 2007 Prob. & Stochastic Process Exam
Autumn

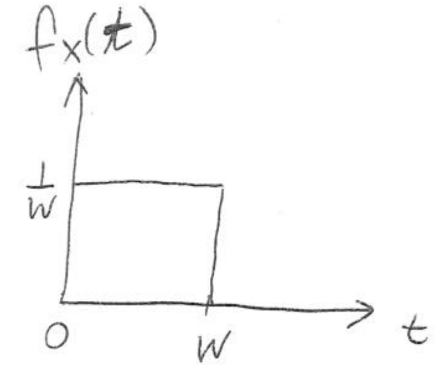
master - 18/4/07

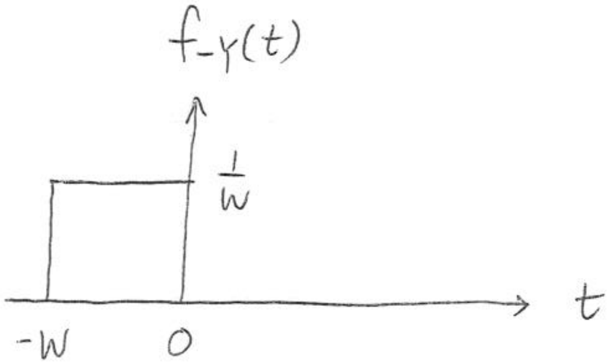
1.a. Define $H \triangleq$ Event that a 1 is sent
$H^c \triangleq$ Event that a 0 is sent
$D \triangleq$ Event that a 1 is detected.

$$P(H/D) = \frac{P(HD)}{P(D)}$$
$$= \frac{P(D/H)P(H)}{P(HD) + P(H^cD)}$$
$$= \frac{P(D/H)P(H)}{P(D/H)P(H) + P(D/H^c)P(H^c)}$$
$$= \frac{0.9 \times 0.4}{0.9 \times 0.4 + 0.05 \times 0.6}$$
$$= \frac{0.36}{0.36 + 0.03} \quad \left(= \frac{0.36}{0.39} \right)$$

$$\Rightarrow P(H/D) = \frac{12}{13}$$

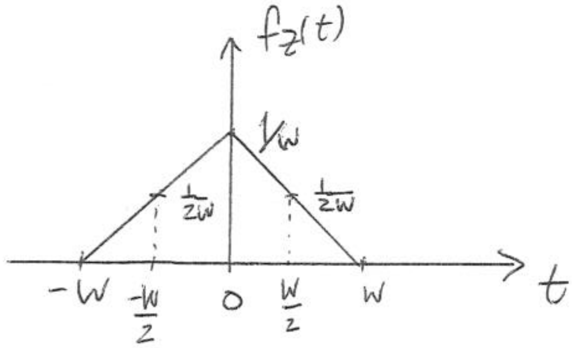
1. b.





Convolution

Since $Z = X + (-Y)$, $f_Z(t) = f_X(t) \circledast f_{-Y}(t)$.



1c.

$$S^*(z) = E\left[z^{X_1 + X_2 + \cdots + X_N} \right]$$

$$= \sum_{m=0}^{\infty} E\left[z^{X_1 + X_2 + \cdots + X_m} \middle| N = m \right] P(N=m)$$

$$= \sum_{m=0}^{\infty} \prod_{i=1}^{m} E\left[z^{X_i} \right] \cdot \frac{a^m e^{-a}}{m!}$$

$$= \sum_{m=0}^{\infty} \prod_{i=1}^{m} \left[z\, P(X_i = 1) + P(X_i = 0) \right] \frac{a^m e^{-a}}{m!}$$

$$= \sum_{m=0}^{\infty} \prod_{i=1}^{m} \left(zp + 1 - p \right) \frac{a^m e^{-a}}{m!}$$

$$= \sum_{m=0}^{\infty} (zp + 1 - p)^m \cdot \frac{a^m e^{-a}}{m!}$$

$$= e^{-a} \sum_{m=0}^{\infty} \frac{[a(zp + 1 - p)]^m}{m!}$$

$$= e^{-a}\, e^{ap(z-1)+a}$$

$$\Rightarrow \quad S^*(z) = e^{ap(z-1)}$$

2a. $$X^*(s) = \int_{-\infty}^{\infty} e^{-st}\,\frac{a}{2}\,e^{-a|t|}\,dt$$

$$= \frac{a}{2}\left[\int_{-\infty}^{0} e^{-st} e^{+at}\,dt + \int_{0}^{\infty} e^{-st} e^{-at}\,dt\right]$$

$$= \frac{a}{2}\left[\int_{-\infty}^{0} e^{(-s+a)t}\,dt + \int_{0}^{\infty} e^{-(s+a)t}\,dt\right]$$

$$= \frac{a}{2}\left[\frac{1}{-s+a}\cdot e^{(-s+a)t}\Big|_{-\infty}^{0} + \frac{-1}{s+a}\,e^{-(s+a)t}\Big|_{0}^{\infty}\right]$$

$$= \frac{a}{2}\left[\frac{1}{-s+a} + \frac{1}{s+a}\right]$$

$$= \frac{a}{2}\cdot\frac{s+a-s+a}{a^2-s^2}$$

$$\Rightarrow \quad X^*(s) = \frac{a^2}{a^2-s^2}$$

$$\bar{X} = -\frac{dX^*(s)}{ds}\Big|_{s=0} = -\left[a^2(-1)(a^2-s^2)^{-2}(-2s)\Big|_{s=0}\right]$$

$$\Rightarrow \quad \bar{X} = 0$$

$$\overline{X^2} = \frac{d^2X^*(s)}{ds^2}\Big|_{s=0} = \frac{d}{ds}\left[2a^2 s(a^2-s^2)^{-2}\right]\Big|_{s=0}$$

$$= \left[2a^2(a^2-s^2)^{-2} + 2a^2 s\,(-2)(a^2-s^2)^{-3}(-2s)\right]\Big|_{s=0}$$

$$\Rightarrow \quad \overline{X^2} = \frac{2a^2}{a^4} = \frac{2}{a^2}$$

$$\sigma_X^2 = \overline{X^2} - \bar{X}^2 = \frac{2}{a^2}$$

2b. i) Poisson process has the memoryless property.
Thus, the time intervals $(t_A - t)$ and $(t_B - t)$ after time t are independent of the evolution of the processes prior to time t.

ii) The time intervals $(t_A - t)$ and $(t_B - t)$ have their pdf's

$$\lambda_A e^{-\lambda_A t} \quad \text{and} \quad \lambda_B e^{-\lambda_B t},$$

respectively.

iii) Let T be the time interval from time t until the next arrival, regardless of the source.

Then, $\quad T = \min\left[(t_A - t), (t_B - t)\right]$

$$P(T > x) = P\{\min[(t_A - t)(t_B - t)] > x\}$$

$$= P\{(t_A - t) > x\} \cdot P\{(t_B - t) > x\}$$

$$= e^{-\lambda_A x} \cdot e^{-\lambda_B x}$$

$$\Rightarrow P(T > x) = e^{-(\lambda_A + \lambda_B)x}$$

$$\Rightarrow F_T(x) = 1 - P(T \leq x)$$

$$= 1 - e^{-(\lambda_A + \lambda_B)x}$$

$$\Rightarrow f_T(x) = (\lambda_A + \lambda_B) e^{-(\lambda_A + \lambda_B)x}$$

2b. iv) For merging N Poisson processes, the pdf of the interarrival time for the merged process is

$$(\lambda_1 + \dots + \lambda_N) e^{-(\lambda_1 + \dots + \lambda_N)t}$$

where λ_i is the arrival rate of the i^{th} Poisson process.

3a. i) $\sigma_U^2 = E[(U - E(U))^2]$

$= E[(aX + bY - a\bar{X} - b\bar{Y})^2]$

$= E\{[a(X-\bar{X}) + b(Y-\bar{Y})]^2\}$

$= E\{a^2(X-\bar{X})^2 + 2ab(X-\bar{X})(Y-\bar{Y})$

$+ b^2(Y-\bar{Y})^2\}$

$\Rightarrow \sigma_U^2 = a^2\sigma_X^2 + 2ab\, E[(X-\bar{X})(Y-\bar{Y})]$

$+ b^2\sigma_Y^2$

$\Rightarrow \sigma_U^2 = a^2\sigma_X^2 + 2ab\, cov(XY) + b^2\sigma_Y^2$

ii) $cov(UV) = E[(U-\bar{U})(V-\bar{V})]$

$= E[(aX + bY - a\bar{X} - b\bar{Y})(cX + dY - c\bar{X} - d\bar{Y})]$

$= E\{[a(X-\bar{X}) + b(Y-\bar{Y})][c(X-\bar{X}) + d(Y-\bar{Y})]\}$

$= E\{ac(X-\bar{X})^2 + ad(X-\bar{X})(Y-\bar{Y})$

$+ bc(Y-\bar{Y})(X-\bar{X}) + bd(Y-\bar{Y})^2\}$

$\Rightarrow cov(UV) = ac\,\sigma_X^2 + (ad + bc)\, cov(XY) + bd\,\sigma_Y^2$

3b. i) $R_{XY}(-\tau) = E[X(t)\, Y(t-\tau)]$

$$= E[X(t'+\tau)\, Y(t')] \qquad t' = t-\tau$$

$$= E[Y(t')\, X(t'+\tau)]$$

$$= R_{YX}(\tau)$$

$$\Rightarrow \quad R_{XY}(-\tau) = R_{YX}(\tau)$$

ii) For any arbitrary real k,

$$E\{[X(t) \pm k Y(t+\tau)]^2\} \geq 0$$

$$\Rightarrow \quad E\{X^2(t) \pm 2k X(t) Y(t+\tau) + k^2 Y^2(t+\tau)\} \geq 0$$

$$\Rightarrow \quad R_X(0) \pm 2k R_{XY}(\tau) + k^2 R_Y(0) \geq 0$$

Set $k=1$ in the above,

$$\Rightarrow \quad R_X(0) \pm 2 R_{XY}(\tau) + R_Y(0) \geq 0$$

$$\Rightarrow \quad \pm 2 R_{XY}(\tau) \leq R_X(0) + R_Y(0)$$

$$\Rightarrow \quad |R_{XY}(\tau)| \leq \frac{1}{2}\left(R_X(0) + R_Y(0)\right)$$

iii) Consider $E\{[X(t) + kY(t+\tau)]^2\} \geq 0$

where k is a positive, real constant.

Thus, we have $k^2 R_Y(0) + 2k R_{XY}(\tau) + R_X(0) \geq 0$

The quadratic (i.e., the left-hand side of the above inequality) will never be negative if k does not have real roots. That is, if the discriminant is such that

$$\left[2 R_{XY}(\tau)\right]^2 - 4 R_Y(0) R_X(0) \leq 0$$

$$\Rightarrow \quad |R_{XY}(\tau)| \leq \sqrt{R_X(0) R_Y(0)}$$

iv) The square of any real number is greater than or equal to zero. Thus,

$$\left(\sqrt{R_X(0)} - \sqrt{R_Y(0)}\right)^2 \geq 0$$

$$\Rightarrow \quad R_X(0) - 2\sqrt{R_X(0) R_Y(0)} + R_Y(0) \geq 0$$

$$\Rightarrow \quad \frac{1}{2}\left(R_X(0) + R_Y(0)\right) \geq \sqrt{R_X(0) R_Y(0)}$$

Therefore, $|R_{XY}(\tau)| \leq \sqrt{R_X(0) R_Y(0)}$ is a stronger bound than

$$|R_{XY}(\tau)| \leq \frac{1}{2}\left(R_X(0) + R_Y(0)\right)$$

4a. i) $P(A) = P(X \le 8{,}610) - P(X \le 7{,}790)$

$$= \Phi\left(\frac{8610-8200}{615}\right) - \Phi\left(\frac{7790-8200}{615}\right)$$

$$= \Phi(0.667) - \Phi(-0.667)$$

From the table (at the beginning of exam paper), $Q(0.6) = 0.274$, $Q(0.7) = 0.242$

$$Q(0.667) = 0.274 + \frac{2}{3}(0.242 - 0.274)$$

$$= 0.252667$$

$$\Rightarrow \quad P(A) = (1 - 0.252667) - 0.252667$$

$$= 0.495$$

ii)

$$f_{X|A}(x) = \begin{cases} \dfrac{1}{0.495} \cdot \dfrac{1}{\sqrt{2\pi}\,\sigma_x} \, e^{-(x-\mu_x)^2/2\sigma_x^2} & \text{if } 7{,}790 < x \le 8{,}610 \\ 0 & \text{otherwise} \end{cases}$$

$$\Rightarrow \quad \sigma_{X|A}^2 = \frac{1}{0.495} \cdot \frac{1}{\sqrt{2\pi}\,\sigma_x} \int_{7{,}790}^{8{,}610} (x-\mu_x)^2 \, e^{-(x-\mu_x)^2/2\sigma_x^2} \cdot dx$$

where $\mu_x = 8200$, $\sigma_x = 615$.

4b. i) Consider $P(X \geq a)$

For $b > 0$,

$X \geq a \iff (X+b) \geq (a+b) \Rightarrow (X+b)^2 \geq (a+b)^2$

Thus, $P(X \geq a) = P((X+b) \geq (a+b))$

$\leq P((X+b)^2 \geq (a+b)^2)$

Let $Y = (X+b)^2$.

$P((X+b)^2 \geq (a+b)^2) = P(Y \geq (a+b)^2)$

$\leq \dfrac{E(Y)}{(a+b)^2}$ by Markov inequality.

Hence, $P(X \geq a) \leq \dfrac{E[(X+b)^2]}{(a+b)^2}$

$= \dfrac{E[X^2 + 2bX + b^2]}{(a+b)^2}$

$\Rightarrow \quad P(X \geq a) \leq \dfrac{\sigma^2 + b^2}{(a+b)^2} \quad \because E(X) = 0.$

ii) Let $f(b) = \dfrac{\sigma^2 + b^2}{(a+b)^2}$

$\dfrac{df(b)}{db} = (\sigma^2 + b^2)(-2)(a+b)^{-3} + (a+b)^{-2}(2b)$

$= \dfrac{-2\sigma^2 + 2ab}{(a+b)^3}$

Set $\dfrac{df(b)}{db} = 0$. We obtain $b = \sigma^2/a$ which minimizes $f(b)$.

Therefore, $P(X \geq a) \leq \dfrac{\sigma^2 + \sigma^2/a}{(a + \sigma^2/a)^2} = \dfrac{\sigma^2}{a^2 + \sigma^2}.$

5a. $E[(Y-g(X))^2 | X]$

$= E[(Y - E[Y|X] + E[Y|X] - g(X))^2 | X]$

$= E[(Y - E[Y|X])^2 | X]$

$\quad + E[(E[Y|X] - g(X))^2 | X]$

$\quad + 2E[(Y - E[Y|X])(E[Y|X] - g(X)) | X]$ (*)

Given fixed X, $E[Y|X]$ and $g(X)$ can be viewed as constant. Thus,

$E[(Y - E[Y|X])(E[Y|X] - g(X)) | X]$

$= \{E[Y|X] - g(X)\} E[(Y - E[Y|X]) | X]$

$= \{E[Y|X] - g(X)\}\{\underbrace{E[Y|X] - E[Y|X]}_{0}\}$

$= 0$

Hence, from (*) we get

$$E[(Y-g(X))^2 | X] = E[(Y - E[Y|X])^2 | X] + E[(E[Y|X] - g(X))^2 | X]$$

$$\Rightarrow E[(Y-g(X))^2 | X] \geq E[(Y - E[Y|X])^2 | X]$$

as the 2nd term above is always positive or zero.

Unconditioning X in the above yields

$$E[(Y-g(X))^2] \geq E[(Y - E[Y|X])^2].$$

That is, the best predictor for Y is $E[Y|X]$.

5b. i) Let the (random) event

$$I = i \quad \text{if} \quad a_i < X \leq a_{i+1}.$$

$$\Rightarrow E[(X-Y)^2] = \sum_i E[(X-y_i)^2 \mid I=i]\, P(I=i)$$

By the result in part a, the best predictor for y_i for all i is given by

$$y_i = E[X \mid I=i]$$

$$\Rightarrow \quad y_i = E[X \mid a_i < X \leq a_{i+1}]$$

$$\Rightarrow \quad y_i = \int_{a_i}^{a_{i+1}} \frac{x f_X(x)\,dx}{F_X(a_{i+1}) - F_X(a_i)}$$

ii) Since $Y = E[X \mid I]$, we have

$$E[Y] = E_I\{E[X \mid I]\}$$

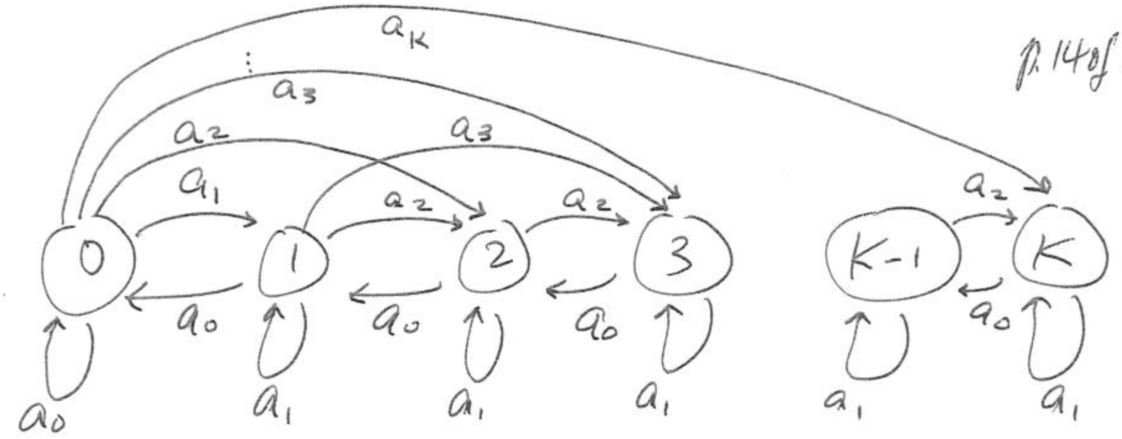
$$= \sum_i E[X \mid I=i] \cdot P(I=i)$$

$$= \sum_i \int_{a_i}^{a_{i+1}} \frac{x f_X(x)\,dx}{\cancel{P(I=i)}} \cdot \cancel{P(I=i)}$$

$$= \sum_i \int_{a_i}^{a_{i+1}} x f_X(x)\,dx$$

$$\Rightarrow \quad E[Y] = E[X].$$

6a.



b. P : state transition prob. matrix

$$
\begin{array}{c c}
 & \begin{array}{cccccc} 0 & 1 & 2 & 3 & \cdots & K-1 & K \end{array} \\
\begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ \vdots \\ K-1 \\ K \end{array} &
\begin{pmatrix}
a_0 & a_1 & a_2 & a_3 & & a_{K-1} & 1-\sum_{i=0}^{K-1} a_i \\
a_0 & a_1 & a_2 & a_3 & \cdots & a_{K-1} & 1-\sum_{i=0}^{K-1} a_i \\
0 & a_0 & a_1 & a_2 & & a_{K-2} & 1-\sum_{i=0}^{K-2} a_i \\
0 & 0 & a_0 & a_1 & & a_{K-3} & 1-\sum_{i=0}^{K-3} a_i \\
 & & & \vdots & & & \\
0 & 0 & 0 & 0 & \cdots & a_1 & 1-\sum_{i=0}^{1} a_i \\
0 & 0 & 0 & 0 & \cdots & a_0 & 1-a_0
\end{pmatrix}
\end{array}
$$

c. The linear equations for the Markov chain

$$\underline{\pi} = \underline{\pi} P$$

and $\sum_{i=0}^{K} \pi_i = 1$

d. $\rho = \sum_{i=1}^{K} \pi_i = 1 - \pi_0$

6e. $$P_o = \sum_{i=0}^{K} P(\text{overflow} \mid N=i)\, P(N=i)$$

where $N \triangleq$ the number of packets existing in the system

$$\Rightarrow P_o = \sum_{i=0}^{K} \sum_{k=K-i+1}^{\infty} a_k \pi_i$$

6f. The channel can still be modelled as a Markov chain ~~depend~~ despite the dependence. This is so because the future evolution of the system is fully characterized by the current state of the system (i.e., the number of existing packets in the system).