

## 6.041/6.431 Spring 2010 Final Exam Solutions

### Question 1

- (a) Let  $X$  be a Poisson random variable with parameter  $\lambda$ . Which **one** of the following statements is implied by applying the Markov inequality to  $X$ ?

- i.  $\mathbf{P}(X \geq 2) \leq 2/\lambda$
- ii.  $\mathbf{P}(X \geq 2) \leq 2/\lambda^2$
- iii.  $\mathbf{P}(X \geq 2) \leq \lambda/2$
- iv.  $\mathbf{P}(|X - \lambda| \geq 2) \leq \lambda^2/4$
- v.  $\mathbf{P}(|X - \lambda| \geq 2) \leq \lambda/4$

The Markov inequality is  $\mathbf{P}(X \geq a) \leq \mathbf{E}[X]/a$  for any nonnegative-valued random variable  $X$  and positive number  $a$ . A Poisson random variable takes only nonnegative values, so the Markov inequality applies. Since  $\mathbf{E}[X] = \lambda$ , applying the Markov inequality with  $a = 2$  gives statement iii.

- (b) Let  $X$  be a standard normal random variable, and let  $\Phi$  denote its CDF as usual. Which **one** of the following statements can be derived with the help of the Chebyshev inequality applied to  $X$ ?

- i.  $\Phi(-t) \leq 1/(2t^2)$  for every positive  $t$
- ii.  $\Phi(-t) \leq 1/(4t^2)$  for every positive  $t$
- iii.  $\Phi(-t) \geq 1/(2t)$  for every positive  $t$
- iv.  $\Phi(-t) \geq 1/(2t^2)$  for every positive  $t$
- v.  $\Phi(-t) \geq 1/(4t^2)$  for every positive  $t$

The Chebyshev inequality

$$\mathbf{P}(|X - \mathbf{E}[X]| \geq c) \leq \text{var}(X)/c^2 \quad \text{for all } c > 0$$

specialized to mean 0 and variance 1 is

$$\mathbf{P}(|X| \geq c) \leq 1/c^2 \quad \text{for all } c > 0.$$

The symmetry of the standard normal PDF implies  $\mathbf{P}(X \leq -c) = \mathbf{P}(X \geq c)$ . Thus the inequality above simplifies to

$$\mathbf{P}(X \leq -c) \leq 1/(2c^2) \quad \text{for all } c > 0.$$

This is statement i. above.

- (c) Let  $X_1, X_2, \dots$  be independent random variables with the uniform distribution over  $[-1, 1]$ . Which **one** of the following statements is true?

- i. With  $Y_n = \sum_{k=1}^n X_k$ , the sequence  $Y_1, Y_2, \dots$  converges in probability to 0.

- ii. With  $Y_n = \sum_{k=1}^n X_k^2$ , the sequence  $Y_1, Y_2, \dots$  converges in probability to 0.
- iii. With  $Y_n = \sum_{k=1}^n X_k^k$ , the sequence  $Y_1, Y_2, \dots$  converges in probability to 0.
- iv. With  $Y_n = \sum_{k=1}^n X_k/n$ , the sequence  $Y_1, Y_2, \dots$  converges in probability to 0.
- v. With  $Y_n = \sum_{k=1}^n X_k/\sqrt{n}$ , the sequence  $Y_1, Y_2, \dots$  converges in probability to 0.

**Statement iv. follows from the weak law of large numbers. All the other statements are false.**

- (d) Let  $X_1, X_2, \dots, X_{100}$  be independent exponential random variables with parameter 0.25, and let  $Y = X_1 + X_2 + \dots + X_{100}$ . Also, as usual let  $\Phi$  denote the CDF of a standard normal random variable. Which of the following is the most accurate approximation?

- i.  $\mathbf{P}\left(\frac{Y - 25}{5/2} \leq z\right) \approx \Phi(z)$
- ii.  $\mathbf{P}\left(\frac{Y - 25}{25/4} \leq z\right) \approx \Phi(z)$
- iii.  $\mathbf{P}\left(\frac{Y - 25}{40} \leq z\right) \approx \Phi(z)$
- iv.  $\mathbf{P}\left(\frac{Y - 400}{40} \leq z\right) \approx \Phi(z)$
- v.  $\mathbf{P}\left(\frac{Y - 400}{1600} \leq z\right) \approx \Phi(z)$

**Since  $Y$  is a sum of i.i.d. random variables, it is reasonable to approximate its CDF with the CDF of a normal random variable with the same mean and variance.  $\mathbf{E}[Y] = 100 \mathbf{E}[X_1] = 400$  and  $\sigma_Y = \sqrt{\text{var}(Y)} = \sqrt{100 \text{var}(X_1)} = \sqrt{100 \cdot 16} = 40$ . Thus  $(Y - 400)/40$  is standardized, and iv. is the best approximation.**

- (e) Let  $Y$  be the number of heads in 100 independent tosses of a fair coin. Also, as usual let  $\Phi$  denote the CDF of a standard normal random variable. Which of the following is the most accurate approximation of  $\mathbf{P}(Y \geq 60)$ ?

- i.  $\Phi(-2)$
- ii.  $\Phi(-1.9)$
- iii.  $\Phi(1.9)$
- iv.  $\Phi(2)$
- v.  $\Phi(2.1)$

**From the list of possible expressions, we see that we must use a normal approximation. The De Moivre–Laplace approximation dictates that the normal approximation to  $\mathbf{P}(Y \geq 59.5)$  will be more accurate than the normal approximation to  $\mathbf{P}(Y \geq 60)$ . Using  $\mathbf{E}[Y] = 100 \cdot 0.5 = 50$  and  $\sigma_Y = \sqrt{100 \cdot 0.5 \cdot 0.5} = 5$ , the desired normal approximation is**

$$\mathbf{P}(Y \geq 59.5) = \mathbf{P}\left(\frac{Y - 50}{5} \geq \frac{59.5 - 50}{5}\right) \approx 1 - \Phi(1.9) = \Phi(-1.9),$$

**choice ii.**

(f) Suppose  $X$  has the exponential distribution with some unknown parameter  $\theta$ . Which of the following is a 50% confidence interval for  $\theta$ ?

- i.  $[(\ln(4/3))X, (\ln 4)X]$
- ii.  $[(\ln(4/3))X, (\ln 2)X]$
- iii.  $[(\ln(4/3))/X, (\ln 4)/X]$
- iv.  $[(\ln(4/3))/X, (\ln 2)/X]$
- v.  $[(\ln(4/3))/X, 1/X]$

**Note that**

$$\mathbf{P}_{\theta}(\theta \in [aX, bX]) = \mathbf{P}_{\theta}(X \in [\theta/b, \theta/a]) = \int_{\theta/b}^{\theta/a} \theta e^{-\theta x} dx = -e^{-\theta x} \Big|_{x=\theta/b}^{x=\theta/a} = e^{-\theta^2/b} - e^{-\theta^2/a},$$

**which can be arbitrarily small (by taking  $\theta \rightarrow \infty$ ). Thus, i. and ii. do not provide confidence intervals at all. Contrarily, the calculation**

$$\mathbf{P}_{\theta}(\theta \in [a/X, b/X]) = \mathbf{P}_{\theta}(X \in [a/\theta, b/\theta]) = \int_{a/\theta}^{b/\theta} \theta e^{-\theta x} dx = -e^{-\theta x} \Big|_{x=a/\theta}^{x=b/\theta} = e^{-a} - e^{-b}$$

**shows that iii., iv., and v. are all confidence intervals. Of these, iii. has the correct length; iv. and v. are shorter and hence *not* 50% confidence intervals.**

### Question 2

An atom of the radioactive element Vestium decays to an atom of Hockfieldium after a time that is an exponential random variable with parameter  $\lambda$ . Hockfieldium is a stable element; i.e., it is not radioactive. Each radioactive atom decays independently of any other atoms.

(a) Suppose a box has  $n$  atoms of Vestium at time 0, where  $n$  is a positive integer. Let  $V$  be the remaining atoms of Vestium in the box at time  $t$ , where  $t$  is a positive real number. Find the PMF of  $V$ .

**Since every atom either remains as Vestium or not, with the same probability and independent of every other atom, the number of remaining Vestium atoms is a binomial random variable. “Success” in the Bernoulli trial that yields this binomial random variable is for the atom to not decay by time  $t$ ; using the CDF of the exponential random variable with parameter  $\lambda$ , the success probability is  $e^{-\lambda t}$ . The desired PMF is thus**

$$p_V(k) = \begin{cases} \binom{n}{k} (e^{-\lambda t})^k (1 - e^{-\lambda t})^{n-k}, & \text{for } k = 0, 1, \dots, n; \\ 0, & \text{otherwise.} \end{cases}$$

(b) An atom of Vestium can itself be the product of the radioactive decay of an atom of Grayon. The decay of any one atom of Grayon to an atom of Vestium occurs after a time that is an exponential random variable with parameter  $\mu$ .

Suppose a box initially contains two atoms of Grayon and nothing else. Find the expected time until the box is no longer radioactive, i.e., it contains neither Grayon nor Vestium—only Hockfieldium.

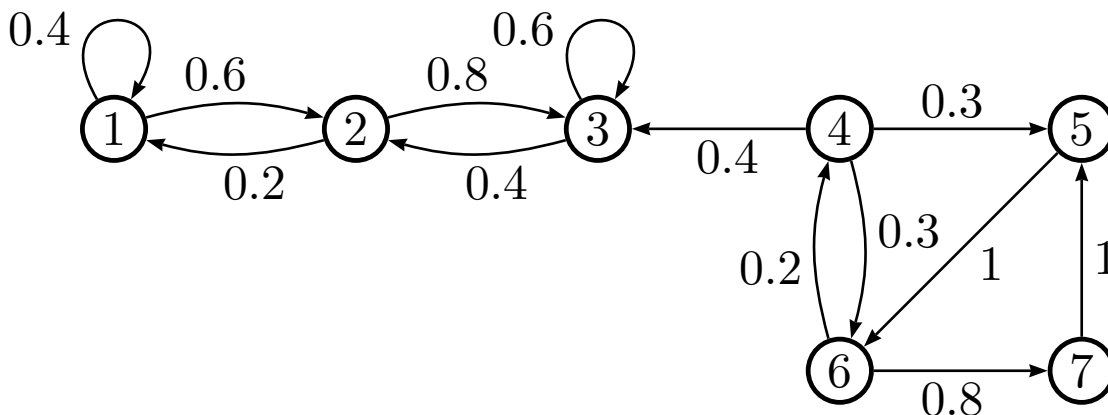
**This part can be solved with an understanding of merged Poisson processes. After a time  $T_1$ , an atom of Grayon decays to an atom of Vestium;  $T_1$  is exponentially distributed**

with parameter  $2\mu$ . After an additional increment of  $T_2$ , another decay occurs; since the box contains one atom of Grayon and one atom of Vestium during this increment,  $T_2$  is exponentially distributed with parameter  $\mu + \lambda$ . Now it makes a difference whether the second decay results in one atom of Grayon and one atom of Hockfieldium or two atoms of Vestium. In the first case, the two remaining increments are exponentially distributed with parameters  $2\lambda$  and  $\lambda$ . In the second case, the two remaining increments are exponentially distributed with parameters  $\mu$  and  $\lambda$ . Using the probability  $\mu/(\mu + \lambda)$  of the first case, the desired expected time is

$$\frac{1}{2\mu} + \frac{1}{\mu + \lambda} + \left[ \frac{\mu}{\mu + \lambda} \left( \frac{1}{2\lambda} + \frac{1}{\lambda} \right) + \frac{\lambda}{\mu + \lambda} \left( \frac{1}{\mu} + \frac{1}{\lambda} \right) \right].$$

### Question 3

Consider a Markov chain  $X_0, X_1, X_2, \dots$  described by the following transition diagram.



- (a) Find the conditional probability of  $X_{103} = 3$  given  $X_{100} = 1$ .

Precisely two transition sequences are consistent with  $X_{100} = 1$  and  $X_{103}$ :

$$1 \rightarrow 1 \rightarrow 2 \rightarrow 3 \quad \text{and} \quad 1 \rightarrow 2 \rightarrow 3 \rightarrow 3.$$

Adding the probabilities of these sequences gives

$$\mathbf{P}(X_{103} = 3 \mid X_{100} = 1) = (0.4)(0.6)(0.8) + (0.6)(0.8)(0.6) = (0.6)(0.8) = 0.48.$$

- (b) Identify the transient and recurrent states.

**States 4, 5, 6, and 7 are transient; states 1, 2, and 3 are recurrent.**

- (c) Find the steady-state distribution of the Markov chain.

**Transient states have zero steady state probability, so  $\pi_4 = \pi_5 = \pi_6 = \pi_7 = 0$ . We can find  $\pi_1$ ,  $\pi_2$ , and  $\pi_3$  through two balance equations and normalization:**

$$\begin{aligned} \pi_2 &= 3\pi_1 \\ \pi_3 &= 2\pi_2 \\ \pi_1 + \pi_2 + \pi_3 &= 1 \end{aligned}$$

**These have solution  $\pi_1 = 1/10$ ,  $\pi_2 = 3/10$ ,  $\pi_3 = 3/5$ .**

- (d) Find  $\lim_{n \rightarrow \infty} \mathbf{P}(X_n = 1 | X_n = X_{n-1})$ .

$X_n = X_{n-1}$  implies  $X_n \in \{1, 3\}$  because only states 1 and 3 have self-transitions. Thus we can write

$$\begin{aligned} \mathbf{P}(X_n = 1 | X_n = X_{n-1}) &= \mathbf{P}(X_n = 1 | \{X_n = X_{n-1} = 1\} \cup \{X_n = X_{n-1} = 3\}) \\ &= \frac{\mathbf{P}(X_n = X_{n-1} = 1)}{\mathbf{P}(\{X_n = X_{n-1} = 1\} \cup \{X_n = X_{n-1} = 3\})} \end{aligned}$$

Using the steady-state distribution computed in the previous part,

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbf{P}(X_n = X_{n-1} = 1) &= \lim_{n \rightarrow \infty} \mathbf{P}(X_n = 1 | X_{n-1} = 1) \mathbf{P}(X_{n-1} = 1) \\ &= p_{11} \lim_{n \rightarrow \infty} \mathbf{P}(X_{n-1} = 1) = p_{11}\pi_1 = \frac{1}{25}, \end{aligned}$$

and similarly

$$\lim_{n \rightarrow \infty} \mathbf{P}(X_n = X_{n-1} = 3) = p_{33}\pi_3 = \frac{9}{25}.$$

Combining the calculations above,

$$\lim_{n \rightarrow \infty} \mathbf{P}(X_n = 1 | X_n = X_{n-1}) = \frac{1/25}{1/25 + 9/25} = \frac{1}{10}.$$

- (e) Suppose  $\mathbf{P}(X_0 = 6) > 0$ . Let  $T$  be the smallest number such that  $X_T = 4$ . Find  $\mathbf{E}[T | X_0 = 6]$ .

The transitions from state 6 are to state 4 and to state 7. Moreover,  $X_1 = 7$  implies  $X_2 = 5$ , which implies  $X_3 = 6$ . From this we get a very simple recursion for  $\mathbf{E}[T | X_0 = 6]$  by using the total expectation theorem:

$$\begin{aligned} \mathbf{E}[T | X_0 = 6] &= \mathbf{P}(X_1 = 4 | X_0 = 6)\mathbf{E}[T | X_1 = 4, X_0 = 6] + \mathbf{P}(X_1 = 7 | X_0 = 6)\mathbf{E}[T | X_1 = 7, X_0 = 6] \\ &= 0.2 \cdot 1 + 0.8(3 + \mathbf{E}[T | X_0 = 6]) \end{aligned}$$

Solving for  $\mathbf{E}[T | X_0 = 6]$  yields  $\mathbf{E}[T | X_0 = 6] = 13$ .

- (f) Again suppose  $\mathbf{P}(X_0 = 6) > 0$ . What is the conditional probability that state 5 is entered at least once given that  $X_0 = 6$ ?

The Markov chain eventually ends up in the recurrent class  $\{1, 2, 3\}$ . At issue is whether it ever passes through state 5 before reaching the recurrent class, when it starts at  $X_0 = 6$ . Let  $q_k$  denote the probability that the Markov chain enters state 5 at least once starting from state  $k$ ; the quantity of interest is  $q_6$ . The following equations follow from the total probability theorem and the definition of  $q_k$ :

$$\begin{aligned} q_4 &= 0.4 \cdot \underbrace{0}_{q_3} + 0.3 \cdot \underbrace{1}_{q_5} + 0.3q_6 \\ q_6 &= 0.2q_4 + 0.8 \cdot \underbrace{1}_{q_7} \end{aligned}$$

Solving this system of equations gives  $q_6 = 43/47$ .

#### Question 4

**Breaking a stick more than twice.** We start with a stick of length  $\ell$ . We break at a point which is chosen randomly and uniformly over its length, and keep the piece that contains the left end of the stick. We then repeat the same process several times on the piece that we were left with. Denote by  $X_n$  the length of the piece we are left with after breaking  $n$  times.

- (a) Find  $\mathbf{E}[X_n]$ .

$X_1$  is uniformly distributed over  $[0, \ell]$  so  $\mathbf{E}[X_1] = \ell/2$ . For any  $n > 1$ , the conditional distribution of  $X_n$  conditioned on  $\{X_{n-1} = t\}$  is uniform over  $[0, t]$  so  $\mathbf{E}[X_n | X_{n-1}] = X_{n-1}/2$ . By the law of iterated expectations,

$$\mathbf{E}[X_n] = \mathbf{E}[\mathbf{E}[X_n | X_{n-1}]] = \mathbf{E}[X_{n-1}/2] = \mathbf{E}[X_{n-1}]/2.$$

A simple induction then gives  $\mathbf{E}[X_n] = \ell/2^n$ .

- (b) After breaking the stick  $n$  times, we randomly pick one of the  $n + 1$  pieces, each of the pieces being equally likely to be picked. Calculate the expected length of the chosen piece.

Let  $L_0, L_1, \dots, L_n$  denote the (random) lengths of the  $n + 1$  pieces and let  $M$  denote the length of the chosen piece. Then

$$\begin{aligned} \mathbf{E}[M] &= \sum_{i=0}^n \mathbf{P}(\text{piece } i \text{ chosen}) \mathbf{E}[M | \text{piece } i \text{ chosen}] && \text{(total expectation theorem)} \\ &= \sum_{i=0}^n \frac{1}{n+1} \mathbf{E}[L_i] && \text{(pieces equally-likely to be chosen)} \\ &= \frac{1}{n+1} \mathbf{E} \left[ \sum_{i=0}^n L_i \right] \\ &= \frac{1}{n+1} \mathbf{E}[\ell] && \text{(sum of lengths of pieces not random!)} \\ &= \frac{\ell}{n+1}. \end{aligned}$$

- (c) Does the sequence  $X_1, X_2, \dots$  converge in probability to a number? If so, to what value? Prove.

The sequence converges in probability to 0. To prove this, let  $\epsilon$  be any positive number and note the following:

$$\begin{aligned} \mathbf{P}(|X_n - 0| \geq \epsilon) &= \mathbf{P}(X_n \geq \epsilon) && \text{(since } X_n \text{ is nonnegative)} \\ &\leq \mathbf{E}[X_n]/\epsilon && \text{(by the Markov inequality)} \\ &= \frac{\ell}{2^n \epsilon} && \text{(using the result of part (a)).} \end{aligned}$$

Thus

$$\lim_{n \rightarrow \infty} \mathbf{P}(|X_n - 0| \geq \epsilon) \leq \lim_{n \rightarrow \infty} \frac{\ell}{2^n \epsilon} = 0.$$

Since probabilities are nonnegative, this shows

$$\lim_{n \rightarrow \infty} \mathbf{P}(|X_n - 0| \geq \epsilon) = 0,$$

completing the proof.

### Question 5

Let  $W_1, W_2$ , and  $W_3$  be independent, continuous random variables each uniformly distributed over  $[0, 1]$ . Let  $X = W_1 + W_2$  and  $Y = X + W_3$ .

- (a) Find the linear least mean squares (LLMS) estimator of  $X$  from  $Y$ .

The linear least mean squares estimator of  $X$  from  $Y$  is given by

$$\hat{X}_{\text{LLMS}} = \frac{\text{cov}(X, Y)}{\text{var}(Y)}(Y - \mathbf{E}[Y]) + \mathbf{E}[X],$$

and the required quantities are straightforward to compute. With independence required in some steps,

$$\mathbf{E}[X] = \mathbf{E}[W_1 + W_2] = \mathbf{E}[W_1] + \mathbf{E}[W_2] = \frac{1}{2} + \frac{1}{2} = 1$$

$$\mathbf{E}[Y] = \mathbf{E}[X + W_2] = \mathbf{E}[X] + \mathbf{E}[W_2] = 1 + \frac{1}{2} = \frac{3}{2}$$

$$\text{var}(X) = \text{var}(W_1 + W_2) = \text{var}(W_1) + \text{var}(W_2) = \frac{1}{12} + \frac{1}{12} = \frac{1}{6}$$

$$\text{var}(Y) = \text{var}(W_1 + W_2 + W_3) = \text{var}(W_1) + \text{var}(W_2) + \text{var}(W_3) = \frac{1}{12} + \frac{1}{12} + \frac{1}{12} = \frac{1}{4}$$

$$\begin{aligned} \mathbf{E}[XY] &= \mathbf{E}[X(X + W_3)] = \mathbf{E}[X^2] + \mathbf{E}[XW_3] = \text{var}(X) + (\mathbf{E}[X])^2 + \mathbf{E}[X]\mathbf{E}[W_3] \\ &= \frac{1}{6} + 1^2 + 1 \cdot \frac{1}{2} = \frac{5}{3} \end{aligned}$$

$$\text{cov}(X, Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y] = \frac{5}{3} - 1 \cdot \frac{3}{2} = \frac{1}{6}$$

Thus, the desired estimator is

$$\hat{X}_{\text{LLMS}} = \frac{2}{3} \left( Y - \frac{3}{2} \right) + 1 = \frac{2}{3}Y.$$

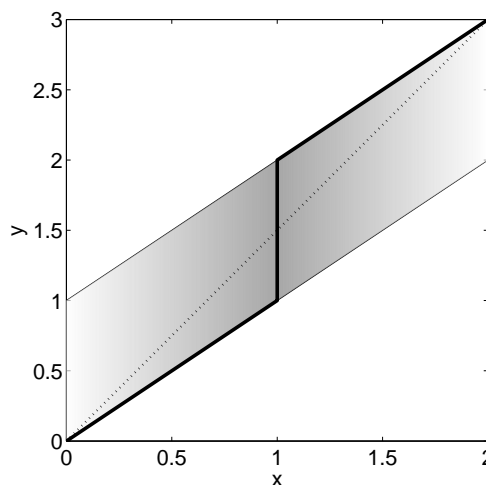
It is also possible to find the LMS estimator (without presuming it to be linear) and then notice that it is linear.

- (b) Find the maximum a posteriori probability (MAP) estimator of  $X$  from  $Y$ .

As the sum of two uniform random variables,  $X$  has the triangular PDF

$$f_X(x) = \begin{cases} 1 - |x - 1|, & \text{for } x \in [0, 2]; \\ 0, & \text{otherwise.} \end{cases}$$

The conditional PDF  $f_{Y|X}(y|x)$  is uniform over  $[x, x+1]$ . Thus the joint PDF  $f_{X,Y}(x,y)$  is nonzero on a parallelogram as marked below, with constant value on vertical slices within the parallelogram but non-constant value on horizontal slices. This is depicted with shading below.



The MAP estimate of  $X$  from  $Y = y$  is obtained by finding the maximum of  $f_{X,Y}(x, y)$  along the horizontal slice determined by  $Y = y$ . This maximum is obtained on the bold curve above. Thus,

$$\hat{X}_{\text{MAP}} = \begin{cases} Y, & \text{for } Y \in [0, 1]; \\ 1, & \text{for } Y \in [1, 2]; \\ Y - 1, & \text{for } Y \in (2, 3]. \end{cases}$$

(The LLMS estimator is shown with a dotted line.)

### Question 6

Consider the problem of choosing between hypotheses  $H_0$  and  $H_1$  upon observing  $X$ . The observation  $X$  has the following PDFs under the two hypotheses:

$$\begin{aligned} f_X(x; H_0) &= \begin{cases} e^{-x}, & \text{for } x \geq 0; \\ 0, & \text{otherwise,} \end{cases} \\ f_X(x; H_1) &= \begin{cases} 2e^{-2x}, & \text{for } x \geq 0; \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

- (a) Find the probabilities of Type I and Type II errors under the maximum likelihood (ML) rule. Recall: Type I error is rejecting  $H_0$  when  $H_0$  is true; Type II error is accepting  $H_0$  when  $H_1$  is true.

Using the ML rule is to reject  $H_0$  upon observing  $X = x$  if and only if  $f_X(x; H_1) > f_X(x; H_0)$ . The rejection region is therefore  $x \geq 0$  satisfying  $2e^{-2x} > e^{-x}$ , which is  $x \in [0, \ln 2]$ . The probability of Type I error is

$$\alpha = \mathbf{P}(X \in [0, \ln 2]; H_0) = \int_0^{\ln 2} e^{-x} dx = -e^{-x} \Big|_{x=0}^{x=\ln 2} = 1 - e^{-\ln 2} = \frac{1}{2}.$$

The probability of Type II error is

$$\beta = \mathbf{P}(X \notin [0, \ln 2]; H_1) = \int_{\ln 2}^{\infty} 2e^{-2x} dx = -e^{-2x} \Big|_{x=\ln 2}^{x=\infty} = e^{-2 \ln 2} = \frac{1}{4}.$$

- (b) Among all hypothesis tests with probability of Type I error at most  $1/4$ , what is the minimum possible probability of Type II error?

The set of optimal trade-offs is given by the set of all likelihood ratio tests. For nonnegative  $x$ , the likelihood ratio

$$L(x) = \frac{f_X(x; H_1)}{f_X(x; H_0)} = \frac{2e^{-2x}}{e^{-x}} = 2e^{-x}$$

is monotonically decreasing, so the rejection regions for likelihood ratio tests are simply all the intervals of the form  $[0, b)$ . The probability of Type I error is

$$\alpha(b) = \mathbf{P}(X \in [0, b); H_0) = \int_0^b e^{-x} dx = -e^{-x} \Big|_{x=0}^{x=b} = 1 - e^{-b}.$$

The probability of Type II error is

$$\beta(b) = \mathbf{P}(X \notin [0, b); H_1) = \int_b^{\infty} 2e^{-2x} dx = -e^{-2x} \Big|_{x=b}^{x=\infty} = e^{-2b}.$$

Thus, the optimal trade-off between Type I and Type II errors satisfies  $\beta = (1 - \alpha)^2$ . The problem specifies  $\alpha \leq 1/4$ , which implies  $\beta \geq 9/16$ .



**Question 7**

Let  $X$  and  $Y$  be continuous random variables.

- (a) If  $X$  and  $Y$  are independent, does it follow that  $\mathbf{E}[X | Y] = \mathbf{E}[X]$ ? Justify your answer carefully. (If answering “no,” give a counterexample with a clear explanation of why it is a counterexample.)

**Yes, it does follow. Let  $y$  be any value such that  $f_Y(y) > 0$ . Then since independence of  $X$  and  $Y$  implies  $f_{X|Y}(x|y) = f_X(x)$  for all  $x$ ,**

$$\mathbf{E}[X | Y = y] = \int x f_{X|Y}(x|y) dx = \int x f_X(x) dx = \mathbf{E}[X].$$

**Thus, the random variable  $\mathbf{E}[X | Y]$  is the constant  $\mathbf{E}[X]$ .**

- (b) If  $X$  and  $Y$  satisfy  $\mathbf{E}[X | Y] = \mathbf{E}[X]$ , does it follow that  $X$  and  $Y$  are independent? Justify your answer carefully. (If answering “no,” give a counterexample with a clear explanation of why it is a counterexample.)

**No, it does not follow. One specific counterexample is as follows. Let  $X = Z \cdot Y$  where  $Y$  is standard normal,  $Z$  equals  $-1$  or  $1$  with equal probability, and  $Y$  and  $Z$  are independent. Conditioned on  $\{Y = y\}$ ,  $X$  is  $-y$  or  $y$  with equal probability. Thus,**

$$\mathbf{E}[X | Y = y] = \frac{1}{2}(-y) + \frac{1}{2}y = 0,$$

**from which it follows that  $\mathbf{E}[X | Y] = 0$ . Also,  $\mathbf{E}[X] = \mathbf{E}[\mathbf{E}[X | Y]] = \mathbf{E}[0] = 0$ . However,  $X$  and  $Y$  are certainly *not* independent; the unconditional distribution of  $X$  is standard normal but its conditional distribution conditioned on  $\{Y = y\}$  is discrete.**

**Independence is a very restrictive condition. While we do not have *independence*, it does follow from  $\mathbf{E}[X | Y] = \mathbf{E}[X]$  that  $X$  and  $Y$  are *uncorrelated*:**

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