University of Illinois Spring 2014

ECE 313: Conflict Final Exam

Tuesday, May 13, 2014, 7:00 p.m. — 10:00 p.m.
Room 241 Everitt Lab

1. **[18 points]** Consider an experiment in which a fair coin is repeatedly tossed every ten seconds, with the first toss happening ten seconds after time zero. Assume the tosses are independent.

   (a) (4 points) Associated to the k^{th} toss, define a random variable X_k that takes the value 1 if the outcome is heads and 0 otherwise. What is the pmf of X_k?
   Solution: The distribution of X_k is Bernoulli with $p = 0.5$, so $p_{x_k}(0) = p_{x_k}(1) = 0.5$.

   (b) (2 points) What is the name of the random process defined by the infinite sequence of random variables $X_1, X_2, \ldots$?
   Solution: The sequence $X_1, X_2, \ldots$ is a *Bernoulli process.*

   (c) (6 points) Find the pmf for the total number of heads that appear in the first 10 minutes.
   Solution: In the first ten minutes, there will be (6 tosses/min) $\times$ (10 min) = 60 tosses. Let C_{60} denote the number of heads that appear in 60 tosses; then C_{60} is binomially distributed with pmf $p_{C_{60}}(k) = \binom{60}{k}(0.5)^k(0.5)^{60-k} = \binom{60}{k}(0.5)^{60},\ 0 \le k \le 60$.

   (d) (6 points) Find the pmf for the time it takes until a heads shows for the first time.
   Solution: Since the tosses are conducted every 10 s, the time elapsed until a heads shows up for the first time is $T_1 = 10 \times L_1$ [s], where L_1 is a geometric random variable with parameter $p = 0.5$. Thus, $p_{T_1}(10k) = (0.5)^k,\ k = 1, 2, \ldots$.

2. **[24 points]** Let T_1 and T_2 denote two independent exponential random variables with the same parameter, λ.

   (a) (6 points) Find the pdf of $T_t = T_1 + T_2$.
   Solution: Since T_1 and T_2 are independent, we can use the convolution formula; thus, $f_{T_t}(t) = \int_0^t \lambda^2 e^{-\lambda u} e^{-\lambda(t-u)} du = \int_0^t \lambda^2 e^{-\lambda t} du = \lambda^2 t e^{-\lambda t}$
   ALTERNATIVELY, the sum of r independent exponentially distributed random variables with parameter r has the Erlang distribution with parameters r and λ. The Erlang distribution for $r = 2$ has pdf $\lambda^2 t e^{-\lambda t}$ for $t \ge 0$.

   (b) (6 points) Find the pdf of $T_s = \min\{T_1, T_2\}$.
   Solution: First we will find the CDF of T_s and then we will differentiate to obtain the pdf. From the definition of CDF, we have that

$$\begin{aligned} F_{T_s}(t) &= P\{T_s \le t\} = P\{\min\{T_1, T_2\} \le t\} = 1 - P\{\min\{T_1, T_2\} > t\} \\ &= 1 - P(\{T_1 > T\} \cap \{T_2 > t\}) \\ &= 1 - P\{T_1 > t\}P\{T_2 > t\} \qquad \text{(by independence of } T_1 \text{ and } T_2) \\ &= 1 - e^{-\lambda t}e^{-\lambda t} = 1 - e^{-2\lambda t} \end{aligned}$$

   Differentiating yields $f_{T_s}(t) = 2\lambda e^{-2\lambda t}$ for $t \ge 0$. That is, T_s has the exponential distribution with parameter 2λ.

   (c) (6 points) Find the pdf of $T_p = \max\{T_1, T_2\}$.
   Solution: Similarly to part (b) above, first we will find the CDF of T_p and then will differentiate to find its pdf. Again, from the definition of CDF, we have that $F_{T_p}(t) = P(T_p \le t) = P(\max\{T_1, T_2\} \le t) = P(\{\{T_1 \le T\} \cap \{T_2 \le t\}\})$. Then, from independence, we have that $P(\{\{T_1 \le T\} \cap \{T_2 \le t\}\}) = P(T_1 \le T)P(T_2 \le t) = (1 - e^{-\lambda t})(1 - e^{-\lambda t}) = 1 - 2e^{-\lambda t} + e^{-2\lambda t}$. Finally, by differentiating, we obtain that $f_{T_p}(t) = 2\lambda(e^{-\lambda t} - e^{-2\lambda t})$.

(d) (6 points) Without doing any calculation, determine which is larger, $E[T_s]$ or $E[T_p]$. Explain your reasoning.
Solution: Let $D = T_p - T_s$, so that $T_p = T_s + D$. Then $P\{D > 0\} = 1$, because the minimum of two numbers is less than the maximum. So $E[D] > 0$. Therefore, $E[T_p] = E[T_s] + E[D] > E[T_s]$.

3. **[24 points]** At a tennis tournament, Serena is up against an opponent who she never met before. The match is best of five sets, where the one who first wins three sets is the winner. There are three equally likely scenarios for Serena:

- She is a stronger player and she wins each set with probability 3/4.
- She is a weaker player and she wins each set with probability 1/4.
- She is equally good as her opponent and she wins each set with probability 1/2.

Given which scenario holds, the outcomes of the sets are mutually independent.

(a) (6 points) What is the probability the match ends in three straight sets?
Solution: $P\{WWW\} = P\{LLL\} = 1/3 \times ((1/4)^3 + (3/4)^3 + (1/2)^3) = 3/16$. So the probability the match ends in three sets is $P\{WWW, LLL\} = 3/8$.

(b) (6 points) What is the probability Serena loses the match?
Solution: 1/2 by symmetry.

(c) (6 points) It turns out that Serena lost the match by 1-3. What was the probability for that to happen?
Solution:

$$\begin{aligned} P(\text{lost 1-3}) &= P\{WLLL, LWLL, LLWL\} = 3P\{WLLL\} \\ &= 3 \cdot \frac{1}{3} \cdot \left((1/4)^3(3/4) + (3/4)^3(1/4) + (1/2)^4\right) \\ &= \frac{3}{256} + \frac{27}{256} + \frac{16}{256} = \frac{46}{256} = \frac{23}{128} \end{aligned}$$

(d) (6 points) After the match Serena decides to reevaluate herself. Given that she lost by 1-3, what is the conditional probability Serena is a stronger player than her opponent?
Solution: $\frac{3}{3+27+16} = \frac{3}{46}$.

4. **[24 points]** Suppose (X, Y) is uniformly distributed within the triangular region $\{(u, v) : 0 \le v \le u \le 1\}$, i.e.,

$$f_{XY}(u, v) = \begin{cases} 2 & 0 \le v \le u \le 1 \\ 0 & \text{else} \end{cases}$$

(a) (6 points) Are X and Y independent? Explain your answer.
Solution: No, the support of $f_{X,Y}$ is not a product set.

(b) (6 points) Suppose we want to estimate Y from X in the way that minimizes the mean-square error (MSE). Find the best unconstrained estimator $g^*(X)$ and the corresponding MSE.
Solution: Conditioned on $X = u$, Y is uniformly distributed on $[0, u]$. Therefore $g^*(u) = E[Y|X = u] = u/2$. The MMSE is $E[(Y - g^*(X))^2] = E[Y^2] - E[g^*(X)^2]$, where $E[Y^2] = \int_0^1 2(1-v)v^2 dv = \frac{1}{6}$, $E[X^2] = \int_0^1 2uu^2 du = \frac{1}{2}$. So

$$\text{MMSE} = E[Y^2] - E[X^2]/4 = \frac{1}{24}.$$

(c) (6 points) Find the best linear estimator $L^*(X)$ and the corresponding MSE.
Solution: Since g^* is linear, it is also the best linear estimator. Therefore L^* coincides with g^* and the MSE is also $\frac{1}{24}$. Alternatively, we could use the formulas for L^* and for the associated MSE.

(d) (6 points) Find the correlation coefficient $\rho_{X,Y}$.
Solution: By symmetry, $\sigma_X = \sigma_Y$. And by the general formula,
$L^*(X) = \mu_Y + \frac{\rho_{X,Y}\sigma_Y}{\sigma_X}(X - \mu_X)$, whereas $L^*(X) = \frac{X}{2}$. Matching the slopes of these two expressions for L^* yields $\rho_{X,Y} = 1/2$.
ALTERNATIVELY, we can calculate $\text{Cov}(X,Y)$, $\text{Var}(X)$, and $\text{Var}(Y)$ (which is equal to $\text{Var}(X)$) and use the definition of $\rho_{X,Y}$.

5. **[18 points]** Consider a vehicle traveling along a straight road in the direction of increasing mile markers. Suppose that at time zero, the location of the vehicle U is uniformly distributed between zero and one (i.e. between mile marker zero and mile marker one), and suppose the speed of the vehicle V is a constant over time, which is exponentially distributed with parameter $\lambda > 0$, and is independent of U. Thus, the location of the vehicle as a function of time t (measured in hours) is given by $X_t = U + Vt$.

(a) (6 points) Identify the joint pdf of (U, V). Be sure to specify the support (i.e. where the joint pdf is not zero.)
Solution:

$$f_{U,V}(u,v) = f_U(u)f_V(v) = \begin{cases} \lambda e^{-\lambda v} & 0 \le u \le 1, \quad v \ge 0 \\ 0 & else \end{cases}$$

(b) (6 points) Find the mean and variance of X_t for a fixed $t > 0$. Make your answer as explicit as possible.
Solution: $E[X_t] = E[U] + tE[V] = \frac{1}{2} + \frac{t}{\lambda}$, and, since U and V are independent, $\text{Var}(X_t) = \text{Var}(U) + t^2\text{Var}(V) = \frac{1}{12} + \frac{t^2}{\lambda^2}$.

(c) (6 points) Let T be the time the vehicle reaches the one mile marker. Find the CDF of T.
Solution: Using $U + VT = 1$ yields $T = \frac{1-U}{V}$. Clearly $F_T(t) = 0$ for $t < 0$. For $t \ge 0$,

$$\begin{aligned} F_T(t) &= P\left\{\frac{1-U}{V} \le t\right\} = P\left\{V \ge \frac{1-U}{t}\right\} \\ &= P\{(U,V) \in \text{shaded region}\} = \int_0^1 \int_{\frac{1-u}{t}}^{\infty} \lambda e^{-\lambda v} dv du \\ &= \int_0^1 e^{-\lambda\left(\frac{1-u}{t}\right)} du = \frac{t}{\lambda} e^{-\lambda\left(\frac{1-u}{t}\right)}\Big|_{u=0}^{1} = \frac{t}{\lambda}(1 - e^{-\frac{\lambda}{t}}) \end{aligned}$$

6. **[14 points]** Suppose 192 tickets are sold for an airplane flight with 150 seats. Suppose each passenger arrives for the flight with probability $p = 0.75$.

(a) (7 points) Assuming each passenger arrives independently with probability p, use the central limit theorem (CLT) to express the approximate probability the flight is oversold (i.e. strictly more than 150 passengers arrive for the flight) in terms of the Q function. You do *not* need to use the continuity correction.
Solution: Let X denote the number of passengers that arrive for the flight. By the assumptions, X has the binomial distribution with parameters $n = 192$ and $p = 0.75$. Thus $E[X] = (192)(0.75) = 144$ and $\text{Var}(X) = 192(0.75)(0.25) = \frac{3 \cdot 192}{16} = 36$. So using the Gaussian approximation without a continuity correction,

$$P\{X > 150\} = P\left\{\frac{X - 144}{6} > \frac{150 - 144}{6}\right\} \approx Q\left(\frac{150 - 144}{6}\right) = Q(1).$$

For the continuity correction we would start with $P\{X \geq 150.5\}$ yielding $Q(6.5/6)$. We could also start with $P\{X \geq 151\}$, giving $Q(7/6)$.

(b) (7 points) Suppose instead that the 192 passengers consist of 96 pairs of passengers, such that for each pair, both passengers in the pair arrive for the flight with probability $p = 0.75$; otherwise neither passenger in the pair arrives for the flight. Assume pairs arrive independently. The number of seats on the plane is still 150. Find the Gaussian approximation for the probability the flight is oversold.

Solution: If we let Y_i denote the number of passengers that arrive from the i^{th} pair, then the total number of passengers that arrive can be written as $S = Y_1 + \cdots + Y_{96}$, where the Y's are independent and $P\{Y_i = 2\} = p$ and $P\{Y_i = 0\} = 1 - p$. Thus, S has mean $96(2)p = 144$ and variance $96(4)p(1-p) = 72$. (So the effect of pairing gives a larger variance than in part (a).) Thus,

$$P\{S > 150\} = P\left\{\frac{S - 144}{\sqrt{72}} > \frac{150 - 144}{\sqrt{72}}\right\} \approx Q\left(\frac{150 - 144}{\sqrt{72}}\right) = Q(1/\sqrt{2}).$$

(Slightly different correct answers are possible as for part (a). Another way to solve this part is to focus on the number of pairs of passengers that arrive, with the flight being oversold if more than 75 pairs of passengers arrive.)

7. **[24 points]** Suppose under hypothesis H_0, the observation X has density f_0, and under hypothesis H_1, the observation X has density f_1, where the densities are given by

$$f_0(u) = \begin{cases} \frac{1}{2} & |u| \leq 1 \\ 0 & |u| \geq 1 \end{cases} \qquad f_1(u) = \begin{cases} 1 - |u| & |u| \leq 1 \\ 0 & |u| \geq 1 \end{cases}$$

(a) (6 points) Describe the ML decision rule for deciding which hypothesis is true for observation X.

Solution: ML decision rule: declare H_1 if $|X| < \frac{1}{2}$, and declare H_0 otherwise.

(b) (6 points) Find $p_{\text{false alarm}}$ and p_{miss} for the ML rule.

Solution:

$$p_{\text{false alarm}} = P(\text{declare } H_1 | H_0 \text{ is true}) = P\left(-\frac{1}{2} \leq X \leq \frac{1}{2} \middle| H_0\right) = \frac{1}{2}.$$

$$p_{\text{miss}} = P(\text{declare } H_0 | H_1 \text{ is true}) = P\left(|X| \geq \frac{1}{2} \middle| H_1\right) = \frac{1}{4}.$$

(c) (6 points) Suppose it is assumed a priori that H_0 is true with probability π_0 and H_1 is true with probability $\pi_1 = 1 - \pi_0$. For what values of π_0 does the MAP decision rule declare H_0 with probability one, no matter which hypothesis is really true?

Solution: In order for H_0 to be declared with probability one, we need $\frac{f_1(u)}{f_0(u)} \leq \frac{\pi_0}{\pi_1}$ for all u. The maximum of $\frac{f_1(u)}{f_0(u)}$ is 2 (occurs at $u = 0$–draw a sketch) so we need $2 \leq \frac{\pi_0}{\pi_1}$, or $\pi_0 \geq \frac{2}{3}$.

(d) (6 points) For the prior distribution $\pi_0 = 0.2$ and $\pi_1 = 0.8$, find $P\left(H_0 \;\middle|\; |X| < 0.5\right)$.

Solution:

$$P\left(H_0 \;\middle|\; |X| < 0.5\right) = \frac{P(H_0, |X| < 0.5)}{P(|X| < 0.5)} = \frac{0.2 \times 0.5}{0.2 \times 0.5 + 0.8 \times 0.75} = \frac{1}{7}$$

8. **[24 points]** Suppose X has the pdf:

$$f_X(u) = \begin{cases} 0.5u & 0 \le u \le 2 \\ 0 & \text{else} \end{cases}$$

(a) (6 points) Find $E[X^2]$.
Solution:

$$E[X^2] = \int_0^2 0.5u^3 du = \left.\frac{u^4}{8}\right|_0^2 = 2.$$

(b) (6 points) Find $P(\lfloor X^2 \rfloor = 1)$, where $\lfloor u \rfloor$ is the greatest integer less than or equal to u.
Solution:

$$P(\lfloor X^2 \rfloor) = 1) = P(1 \le X^2 < 2) = P(1 \le X < \sqrt{2}) = \int_1^{\sqrt{2}} 0.5u \; du = \left.\frac{u^2}{4}\right|_1^{\sqrt{2}} = \frac{1}{4}.$$

(c) (6 points) Find the cumulative distribution function (CDF) of $Y = \ln X$.
Solution: The support of Y is $(-\infty, \ln 2]$.

$$F_Y(c) = P\{Y \le c\} = P\{\ln X \le c\} = P\{X \le e^c\} = \int_0^{e^c} 0.5u \; du = \left.\frac{u^2}{4}\right|_0^{e^c} = \frac{e^{2c}}{4}$$

for $c \le \ln 2$, and $F_Y(c) = 1$ for $c > \ln 2$

(d) (6 points) If U is uniformly distributed over the interval $[0, 1]$, find a function g such that $g(U)$ has the same distribution as X.
Solution:

$$F_X(c) = P\{X \le c\} = \int_0^c 0.5u \; du = \left.\frac{u^2}{4}\right|_0^c = \frac{c^2}{4}.$$

Let $\frac{c^2}{4} = u$, then $c = 2\sqrt{u}$, hence $g(u) = F^{-1}(u) = 2\sqrt{u}$, for $0 \le u \le 1$.

9. **[30 points]** (3 points per answer)
In order to discourage guessing, 3 points will be deducted for each incorrect answer (no penalty or gain for blank answers). A net negative score will reduce your total exam score.

(a) This part concerns different choices for the joint pdf of random variables X and Y. The constant c stands for a normalizing constant in each part.

TRUE	FALSE	
☐	☐	X and Y are independent if $f_{X,Y}(u,v) = \begin{cases} ce^{uv} & 0 \le u \le 1, 0 \le v \le 1 \\ 0 & \text{else} \end{cases}$
☐	☐	X and Y are independent if $f_{X,Y}(u,v) = \begin{cases} ce^{uv} & u^2 + v^2 \le 1 \\ 0 & \text{else} \end{cases}$

Solution: False,False

(b) Suppose X is a *continuous* type random variable with CDF $F_X(u)$.

TRUE	FALSE	
☐	☐	If $b > a$, then $F_X(b) > F_X(a)$.
☐	☐	If $F_X(b) > F_X(a)$, then $b > a$.
☐	☐	$F_X(\alpha) = \frac{1}{4}$ for some α with $-\infty < \alpha < \infty$.

Solution: False, True, True.

(c) Let (X, Y) be uniformly distributed over $[0,1] \times [0,1]$.

TRUE	FALSE	
☐	☐	The support of $Z = X - Y$ is $[-1, 1]$.
☐	☐	$Z = X + Y$ is uniformly distributed over $[0, 2]$.

Solution: True, False

(d) Let $X_1, X_2, \ldots$ be an infinite sequence of independent Bernoulli random variables all with parameter $p \in (0, 1)$.

TRUE	FALSE	
☐	☐	$C_{10} = \sum_{k=1}^{10} X_k$ has a geometric distribution with parameter p.
☐	☐	Let $C_m = \sum_{k=1}^{m} X_k$, then the mean of $C_m - C_{m-1}$ is p.
☐	☐	$X = C_{10} - C_8$, and $Y = C_9 - C_7$ are independent random variables.

Solution: False, True, False