6.436/15.085
Midterm Exam

Date: October 23, 2006

Problem 1 *Which of the following functions is a distribution function? For those which are compute the density function. For those which are not explain what fails.*

A.

$$F(x) = \begin{cases} 1 - e^{-x^2}, & x \geq 0; \\ 0, & \textit{otherwise.} \end{cases}$$

B.

$$F(x) = \begin{cases} e^{-\frac{1}{x}}, & x > 0; \\ 0, & \textit{otherwise.} \end{cases}$$

C.

$$F(x) = \begin{cases} 0, & x \leq 0; \\ \frac{1}{3}, & 0 < x \leq \frac{1}{2} \\ 1 & x > \frac{1}{2}. \end{cases}$$

Solution:

A. F is a non-decreasing continuous function satisfying $\lim_{x\to-\infty} F(x) = 0, \lim_{x\to\infty} F(x) = 1$. Thus it is a distribution function. Its density is $f(x) = 0, x < 0$ and $f(x) = 2x\exp(-x^2), x \geq 0$.

B. Same answer. Note that F is continuous at zero as well (although only right-continuity is needed) as $\lim_{x\downarrow 0} \exp(-1/x) = 0$. Its density is $f(x) = 0, x < 0$ and $f(x) = \exp(-1/x)/x^2, x \geq 0$.

C. F is not right-continuous, so it is not a distribution function.

Problem 2 *Buses arrive at twenty minutes intervals starting at noon. A man arrives at a random time X after noon, where X is distributed as*

$$F(x) = \begin{cases} 0, & x < 0; \\ \frac{x}{60}, & 0 \leq x < 60 \\ 1 & x \geq 60. \end{cases}$$

Finding the expected time that the man has to wait for the bus.

Solution: Observe that F is the uniform distribution on $[0, 60]$. We use the following fact: let $U \stackrel{d}{=} [A, B]$ and let $A \le a < b \le B$. Then conditioned on $U \in [a, b]$, $U \stackrel{d}{=} U(a, b)$. One line proof $\mathbb{P}(U \in [a, x] | U \in [a, b]) = \frac{(x-a)/(B-A)}{(b-a)/(B-A)} = (x - a)/(b - a)$.

For each $t = 1, 2, 3$, conditioned on the event $X \in [20(t - 1), 20t]$, the waiting time is $20t - X$, which by our observation is uniformly distributed over $[0, 20]$. Thus the expected waiting time is the half-length, that is $20/2 = 10$. Since the events $X \in [20(t - 1), 20t]$ are equally likely for $t = 1, 2, 3$, the expected waiting time is 10 minutes.

Problem 3 *Let X, Y be independent geometrically distributed r.v. with parameter p. Let $Z = \mathbb{E}[X|X + Y]$. Find the expected value and the variance of Z.*

Solution: We use the following trick: observe that $\mathbb{E}[X|X + Y] + \mathbb{E}[Y|X + Y] = \mathbb{E}[X + Y|X+Y] = X+Y$. By symmetry $\mathbb{E}[X|X+Y] = \mathbb{E}[Y|X+Y]$ implying that $Z = \mathbb{E}[X|X+Y] = (X + Y)/2$. Thus the expected value of Z is $\mathbb{E}[Z] = (1/2)(\mathbb{E}[X] + \mathbb{E}[Y]) = 1/p$. Since X and Y are independent, then $\mathrm{Var}(Z) = (1/4)(\mathrm{Var}(X) + \mathrm{Var}(Y)) = (1 - p)/(2p^2)$, as the variance of a geometric r.v. is $(1 - p)/p^2$.

Problem 4 *A sequence of events $A_1, \ldots, A_n, n \ge 3$ is given. It is known that at least one of these events happens, but also at most two of these events can happen at the same time. Also it is known that $\mathbb{P}(A_r) = p$ for all $r = 1, \ldots, n$ and $\mathbb{P}(A_r \cap A_s) = q$ for all $1 \le r < s \le n$. Show that*

A. $p \ge \frac{1}{n}$.

B. $q \le \frac{2}{n(n-1)}$.

Solution:

A. We are given that $\cup_{1 \le i \le n} A_n = \Omega$. Therefore $1 = \mathbb{P}(\cup_{1 \le i \le n} A_n) \le \sum_i \mathbb{P}(A_i) = np$, implying $p \ge 1/n$.

B. Consider the events $B_{i,j} \triangleq A_i \cap A_j, i \ne j$. Observe that these events are disjoint as $B_{i,j} \cap B_{i,k} = A_i \cap A_j \cap A_k = \emptyset$ and $B_{i,j} \cap B_{k,l} = A_i \cap A_j \cap A_k \cap A_l = \emptyset$. Therefore

$$1 = \mathbb{P}(\Omega) \ge \sum_{i<j} \mathbb{P}(B_{i,j}) = \frac{n(n-1)}{2} q$$

and the required bound for q is obtained.

Problem 5 *Suppose X and Y have joint density function $f(x, y) = 2e^{-x-y}, 0 < x < y < \infty$. Are X and Y independent? Find the covariance of X and Y.*

Solution: Let us compute the marginal densities of X and Y:

$$f_X(x) = \int_{y \geq x} 2\exp(-x-y)dy = 2\exp(-x)(-\exp(-y))\Big|_x^\infty = 2\exp(-2x).$$
$$f_Y(y) = \int_{x \leq y} 2\exp(-x-y)dy = 2\exp(-y)(-\exp(-x))\Big|_0^y = 2\exp(-y)(1-\exp(-y)).$$

We see that $f_X(x)f_Y(y) \neq f(x, y)$. Thus X and Y are not independent. An even simpler way to see this is to observe that $f_{X|Y}(x|Y = y) = f(x, y)/f_Y(y) > 0$ when $y \geq x$ and $f_{X|Y}(x|Y = y) = 0$ when $y < x$. Thus it is not true that $f_{X|Y}(x|Y = y) = f_X(x)$ and X and Y are not independent. We have

$$\mathbb{E}[X] = \int_{x \geq 0} 2x\exp(-2x)dx$$

which we recognize as the expected value of Exp(2), thus equal to 1/2. Now

$$\mathbb{E}[Y] = \int_{y \geq 0} 2y\exp(-y)dy - \int_{y \geq 0} 2y\exp(-2y)dy.$$

The first term is twice the expected value of Exp(1), the second is the expected value of Exp(2). Thus the difference is $2 - (1/2) = 3/2$.

Now let us compute the covariance. We have

$$\begin{aligned}\mathbb{E}[XY] &= \int_{0 \leq x \leq y} 2xy\exp(-x-y)dxdy \\ &= \int_{x \geq 0} 2x\exp(-x)dx \int_{y \geq x} y\exp(-y)dy.\end{aligned}$$

We have

$$\int_{y \geq x} y\exp(-y)dy = x\exp(-x) - \int_{y \geq x} (-\exp(-y))dy = (x+1)\exp(-x),$$

giving

$$\begin{aligned}\mathbb{E}[XY] &= \int_{x \geq 0} 2x(x+1)\exp(-2x)dx \\ &= \int_{x \geq 0} 2x^2\exp(-2x)dx + \int_{x \geq 0} 2x\exp(-2x)dx\end{aligned}$$

We recognize the second term as the expected value of Exp(2) distribution, namely 1/2. The first term is the second moment of the same distribution. We find it as

$$\int_{x\geq 0} 2x^2 \exp(-2x)dx = -\int_{x\geq 0} 4x(-(1/2)\exp(-2x))dx = \int_{x\geq 0} 2x\exp(-2x)dx = 1/2.$$

Putting everything together we find

$$Cov(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 1 - (3/4) = 1/4.$$

Problem 6 *Given events $A_1, \ldots, A_n$ show that*

$$\mathbb{P}(\cup_{1\leq r\leq n} A_r) \leq \min_{1\leq k\leq n} \Big(\sum_{1\leq r\leq n} \mathbb{P}(A_r) - \sum_{r:r\neq k} \mathbb{P}(A_r \cap A_k)\Big).$$

Solution: Fix an arbitrary $k, 1 \leq k \leq n$. Observe

$$\cup_{1\leq r\leq n} A_r = \Big(A_k \cap \cup_{1\leq r\leq n} A_r\Big) \cup \Big(A_k^c \cap \cup_{1\leq r\leq n} A_r\Big)$$

Observe also that these events are disjoint. Thus

$$\begin{aligned}
\mathbb{P}(\cup_{1\leq r\leq n} A_r) &= \mathbb{P}(A_k \cap \cup_{1\leq r\leq n} A_r) + \mathbb{P}(A_k^c \cap \cup_{1\leq r\leq n} A_r) \\
&= \mathbb{P}(A_k) + \mathbb{P}(\cup_{1\leq r\leq n}(A_r \cap A_k^c)) \\
&= \mathbb{P}(A_k) + \mathbb{P}(\cup_{r\neq k}(A_r \cap A_k^c)) \\
&\leq \mathbb{P}(A_k) + \sum_{r\neq k} \mathbb{P}(A_r \cap A_k^c) \\
&= \mathbb{P}(A_k) + \sum_{r\neq k} (\mathbb{P}(A_r) - \mathbb{P}(A_r \cap A_k)) \\
&= \sum_r \mathbb{P}(A_r) - \sum_{r\neq k} \mathbb{P}(A_r \cap A_k))
\end{aligned}$$

Since we established the inequality for all k, the proof is complete.