Stat 426 : Homework 1.

Moulinath Banerjee

University of Michigan

February 3, 2003

Announcement: For purposes of the homework, you can cite any results in the handouts or the text-book or any others proved in class, without proof. The homework carries a total of 60 points, but contributes 4 points towards your total grade.

• 1. Prove that for three not necessarily disjoint events $A$, $B$ and $C$,
\[ P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C). \]

Hint: You can write $A \cup B \cup C$ as $(A \cup B) \cup C$ and use the formula for the union of two events (on page 2 of the first handout) and proceed from there. (5 points)

Solution: Write,
\[
P(A \cup B \cup C) = P((A \cup B) \cup C) \\
= P(A \cup B) + P(C) - P((A \cup B) \cap C) \\
= P(A) + P(B) - P(A \cap B) + P(C) - P((A \cap C) \cup (B \cap C)) \\
= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).
\]


Solution: Let $1S$ denote the event that an event is produced by the first shift, $2S$ denote the event that it was produced by the second shift and $3S$ denote the event that it was produced by the third shift. Let $D$ denote the event that an item is defective. We have the following information:
\[ P(1S) = P(2S) = P(3S) = \frac{1}{3}, \]
and
\[ P(D \mid 1S) = \frac{1}{100}, \ P(D \mid 2S) = \frac{2}{100}, \ P(D \mid 3S) = \frac{5}{100}. \]
We need to find $P(3S \mid D)$. Using Bayes Rule we get,

$$P(3S \mid D) = \frac{P(D \mid 3S) P(3S)}{P(D \mid 3S) P(3S) + P(D \mid 2S) P(2S) + P(D \mid 1S) P(1S)}$$

$$= \frac{5/100 \times 1/3}{5/100 \times 1/3 + 2/100 \times 1/3 + 1/100 \times 1/3}$$

$$= \frac{5}{5 + 2 + 1}$$

$$= \frac{5}{8}$$

$$= 0.625.$$

3. (a) Show that if events $A_1, A_2, \ldots, A_n$ are mutually independent, then so are $A_1, A_2, \ldots, A_{n-1}, A_n$. (Hint: Use the definition of mutual independence)

Solution: What we essentially need to show here is that for any subclass $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$ of $A_1, A_2, \ldots, A_{n-1}$ it is the case that,

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k} \cap A_n) = P(A_{i_1}) \times P(A_{i_2}) \times \ldots \times P(A_{i_k}) \times P(A_n). \quad (*)$$

Now, note that as the events $A_1, A_2, \ldots, A_{n-1}$ are independent, so are the events, $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$. Let

$$B = A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}.$$

Now observe that $B$ and $A_n$ are independent and hence

$$P(B \cap A_n) = P(B) \times P(A_n).$$

Also the left side of the equation $(*)$ is precisely $P(B \cap A_n)$. Now,

$$P(B \cap A_n) = P(B) - P(B \cap A_n) = P(B) - P(B) \times P(A_n) = P(B) \times P(A_n').$$

But the expression on the extreme right of the above display, by the independence of $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$, is simply

$$P(A_{i_1}) \times P(A_{i_2}) \times \ldots \times P(A_{i_k}) \times P(A_n'),$$

which is the right side of the equation $(*)$, completing the proof.

(b) Use this result repeatedly to show that if $B_1, B_2, \ldots, B_n$ are independent events then so are $C_1, C_2, \ldots, C_n$ where each $C_i$ is either $B_i$ or $B_i'$. (Hint: Observe that it suffices to prove that if $B_1, B_2, \ldots, B_n$ are independent, then so are $B_1, B_2, \ldots, B_m, B_{m+1}^c, B_{m+2}^c, \ldots, B_n^c.$
(Why ?) \(3 + 2 = 5\) points

**Solution:** Use induction. First, note that we’ve proved that for any \(n\) events, \(D_1, D_2, \ldots, D_n\) which are independent, so are \(D_1, D_2, \ldots, D_{n-1}, D_n^c\). What this tells us is that given any sequence of events, we can retain independence by just complementing one event and keeping the others intact. Thus, we know that if \(B_1, B_2, \ldots, B_n\) are independent, so are \(B_1, B_2, \ldots, B_{n-1}, B_n^c\). Assume now that we can complement the last \(k\) events and still retain independence; in other words, \(B_1, B_2, \ldots, B_{n-k}, B_{n-k+1}^c, \ldots B_n^c\) are independent. We will show that we can complement \(B_{n-k}\) as well and retain independence. But this follows immediately, since

\[
B_1, B_2, \ldots, B_{n-k}, B_{n-k+1}^c, \ldots, B_n^c
\]

is a sequence of mutually independent events and if we complement \(B_{n-k}\), keeping the others intact, independence is retained. In other words,

\[
B_1, B_2, \ldots, B_{n-k-1}, B_{n-k}^c, B_{n-k+1}^c, \ldots, B_n^c
\]

are mutually independent.

- 4. We call \(X\) a geometric random variable if \(X\) takes values \(\{1, 2, 3, \ldots\}\) and 
\[
P(X = m) = pq^{m-1},
\]
where \(0 < p, q < 1\) and also \(p + q = 1\). Refer to the handout for a random experiment that produces a geometric random variable.

(a) Prove that for any two positive integers \(m, n\), it is the case that,

\[
P(X > m + n \mid X > m) = P(X > n).
\]

This is the **memoryless property** and has been discussed in class. To show this, first prove that the memoryless property is equivalent to the assertion that

\[
P(X > m + n) = P(X > m) P(X > n).
\]

To show this, note that

\[
P(X > m + n \mid X > m) = \frac{P(\{X > m + n\} \cap \{X > m\})}{P(X > m)} = \frac{P(X > m + n)}{P(X > m)}.
\]

Thus,

\[
P(X > m + n \mid X > m) = P(X > n),
\]

is equivalent to

\[
\frac{P(X > m + n)}{P(X > m)} = P(X > n),
\]

which is precisely the same as saying,

\[
P(X > m + n) = P(X > m) P(X > n) \quad (*).
\]

3
Next, show that for the geometric distribution, for any positive integer \( l \),

\[
P(X > l) = q^l,
\]

and proceed.

We show this as follows. Now,

\[
P(X > l) = \sum_{i=1}^{\infty} P(X = i) = \sum_{i=1}^{\infty} q^{i-1} p = p \left( q^1 + q^{l+1} + q^{l+2} + \ldots \right) = p q^l \left( 1 + q + q^2 + \ldots \right) = p q^l \left( 1 / 1 - q \right) = q^l,
\]

since \( q = 1 - p \). Now, (\( \star \)) is readily seen to hold, since

\[
P(X > m + n) = q^{m+n} = q^m q^n = P(X > m) P(X > n).
\]

(b) We will prove the converse of (a). We will show that if \( X \) is a discrete random variable taking values \( \{1, 2, 3, \ldots\} \) with probabilities \( \{p_1, p_2, p_3, \ldots\} \) and satisfies the memoryless property, then \( X \) must follow a geometric distribution.

Follow these steps to establish the fact that \( X \) is geometric. Using the fact that \( X \) has the memoryless property, show that

\[
P(X > m) = (P(X > 1))^m,
\]

for any \( m \geq 2 \). As a first step towards proving this show that

\[
P(X > 2) = (P(X > 1))^2.
\]

The above can be proved through induction. Assume that

\[
P(X > k) = (P(X > 1))^k.
\]
We will show that,
\[ P(X > k + 1) = (P(X > 1))^{k+1}. \]
Now, the memoryless property, combined with what we’ve assumed, tells us that
\[ P(X > k + 1) = P(X > k)P(X > 1) = (P(X > 1))^k P(X > 1) = (P(X > 1))^{k+1}. \]
But we know that our assumption is true for \( k = 2 \), since
\[ P(X > 2) = P(X > 1 + 1) = (P(X > 1))^2. \]
It follows that
\[ P(X > m) = (P(X > 1))^m, \]
for all \( m \geq 2 \). Define \( p = P(X = 1) \) and \( q = P(X > 1) \). You now have,
\[ P(X > m) = q^m, \]
for any \( m \geq 2 \). Use this to show that for any \( m \geq 2 \),
\[ P(X = m) = p q^{m-1}. \]
**Hint:** Note that the event \( \{ X > m - 1 \} \) is the disjoint union of the events \( \{ X > m \} \) and \( \{ X = m \} \).

From the hint, it follows that
\[ P(X = m) = P(X > m - 1) - P(X > m) = q^{m-1} - q^m = p q^{m-1}. \]
But for \( m = 1 \),
\[ P(X = m) = P(X = 1) = p = p q^{m-1}, \]
trivially and the proof is complete. \((5 + 5 = 10 \text{ points})\)

- **5.** If \( X \) is random variable with distribution function \( F \), with continuous non-vanishing density \( f \), obtain the density function of the random variable \( Y = X^2 \), from first principles; i.e. **without** using the extended change of variable theorem on Page 14 of the first handout.

**Hint:** Express the probability of the event \( X^2 \leq y \) in terms of the distribution function \( F \) of \( X \) and proceed from there. \((5 \text{ points})\)

**Solution:** Let \( G(y) \) denote the distribution function of \( Y \). Then
\[ G(y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = F(\sqrt{y}) - F(-\sqrt{y}). \]
To get the density function of \( y \), say \( g(y) \), we just differentiate the distribution function \( G(y) \). Thus,
\[ g(y) = G'(y) = \frac{1}{2 \sqrt{y}} f(\sqrt{y}) + \frac{1}{2 \sqrt{y}} f(-\sqrt{y}), \]
by a direct application of the chain rule for derivatives.
• 6. (i) If $X$ and $Y$ are independent standard normal variables find the probability of the event 
\( \{X^2 + Y^2 \leq 1\} \).

**Solution:** From what we derived in class, we know that $R^2 = X^2 + Y^2$ follows an
\( \exp(1/2) \) distribution. The distribution function is given by
\[
F(y) = 1 - \exp(-y/2).
\]
Thus the required probability is
\[
F(1) = 1 - \exp(-1/2).
\]
(ii) Let $T$ be an exponential random variable with parameter $\lambda$ and let $W$ be a random
variable independent of $T$ which assumes the value 1 with probability 1/2 and the value $-1$ with probability 1/2. Show that the density of $X = WT$ is,
\[
f_X(x) = \frac{\lambda}{2} e^{-\lambda|x|},
\]
using first principles. This distribution is called the **double exponential** distribution.

**Hint:** It would help to split up the event \( \{X \leq x\} \) as the union of \( \{X \leq x, W = 1\} \) and \( \{X \leq x, W = -1\} \). (5 + 5 = 10 points)

**Solution:** We proceed as suggested by the hint, to obtain the distribution function
of the random variable $X$. Denote the distribution of $X$ by $F$. Then,
\[
F(x) &= P(X \leq x, W = 1) + P(X \leq x, W = -1) \\
&= P(WT \leq x, W = 1) + P(WT \leq x, W = -1) \\
&= P(T \leq x, W = 1) + P(-T \leq x, W = -1) \\
&= \frac{1}{2} P(T \leq x) + \frac{1}{2} P(-T \leq x).
\]
Now, for $x < 0$, $P(T \leq x) = 0$ and
\[
P(-T \leq x) = P(T \geq -x) = \exp(\lambda x).
\]
Thus, for $x < 0$,
\[
F(x) = \frac{1}{2} \exp(\lambda x),
\]
and consequently, the density, $f(x)$ is
\[
f(x) = F'(x) = \frac{1}{2} \lambda \exp(\lambda x) = \frac{1}{2} \lambda \exp(-\lambda | x |).
\]
It is shown similarly, that for $x > 0$,
\[
f(x) = \frac{1}{2} \lambda \exp(-\lambda | x |).
\]
7. (i) If $X$ and $Y$ are independent Poisson random variables with parameters $\lambda_1$ and $\lambda_2$, then show that $X + Y$ is also Poisson with parameter $\lambda_1 + \lambda_2$. Recall that if $W$ follows Poisson($\theta$), then the p.m.f. of $W$ is,

$$P(W = m) = \frac{e^{-\theta} \theta^m}{m!}.$$  

**Hint:** Write $P(X + Y = m)$ as $\sum_{i=0}^{m} P(X = i, X + Y = m)$ and proceed.

**Solution:**

$$P(X + Y = m) = \sum_{i=0}^{m} P(X = i, X + Y = m)$$

$$= \sum_{i=0}^{m} P(X = i, Y = m - i)$$

$$= \sum_{i=0}^{m} P(X = i)P(Y = m - i)$$

$$= \sum_{i=0}^{m} e^{-\lambda_1} \frac{\lambda_1^i}{i!} e^{-\lambda_2} \frac{\lambda_2^{m-i}}{(m-i)!}$$

$$= e^{-(\lambda_1 + \lambda_2)} \frac{(\lambda_1 + \lambda_2)^m}{m!} \sum_{i=0}^{m} \binom{m}{i} \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^i \left( \frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{m-i}$$

$$= e^{-(\lambda_1 + \lambda_2)} \frac{(\lambda_1 + \lambda_2)^m}{m!} \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} + \frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^m$$

$$= e^{-(\lambda_1 + \lambda_2)} \frac{(\lambda_1 + \lambda_2)^m}{m!}.$$

(ii) Use this result repeatedly to show that if $X_1, X_2, \ldots, X_n$ are independent Poisson random variables with parameters $\lambda_1, \lambda_2, \ldots, \lambda_n$ respectively, then $X_1 + X_2 + \ldots + X_n$ follows Poisson with parameter $\lambda_1 + \lambda_2 + \ldots + \lambda_n$.

**Solution:** Use induction. From part (i) we know that the result is true for $n = 2$. Assume the result to be true for $n = l$. We will show that it holds for $n = l + 1$. So, let $X_1, X_2, \ldots, X_{l+1}$ be independent Poisson random variables with parameters $\lambda_1, \lambda_2, \ldots, \lambda_{l+1}$. It is required to show that the sum $X_1 + X_2 + \ldots + X_{l+1}$ follows a Poisson distribution with parameter $\lambda_1 + \lambda_2 + \ldots + \lambda_{l+1}$. Now, by our assumption, $S_l = X_1 + X_2 + \ldots + X_l$ is Poisson with parameter $A_l = \lambda_1 + \lambda_2 + \ldots + \lambda_l$. Also, $X_{l+1}$ is independent of $S_l$ and has a Poisson distribution with parameter $\lambda_{l+1}$. From part (i) it follows that $S_{l+1} = S_l + X_{l+1}$ is a Poisson random variable with parameter $A_l + \lambda_{l+1}$, which is precisely what we set out to prove.

(iii) Show from first principles that the conditional distribution of the random vector $(X_1, X_2, \ldots, X_n)$ given that the sum $X_1 + X_2 + \ldots + X_n = K$ for some integer $K$ follows the multinomial distribution with parameters $(K, p_1, p_2, \ldots, p_n)$ where each $p_i$ is given by
\[ \lambda_i / (\sum_{j=1}^{n} \lambda_j) \].

The above result has an interesting “Poisson process interpretation” which we will discuss if we have time. (3 + 2 + 5 = 10 points)

**Solution:** We need to find

\[ P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \mid X_1 + X_2 + \ldots + X_n = K) \].

Clearly, the above conditional probability is 0 if \( \sum_{i=1}^{k} x_i \) does not equal \( K \) and so we only need to compute the conditional probability for all \( n \)-tuples \( (x_1, x_2, \ldots, x_n) \) that sum to \( K \). Consider such an \( n \)-tuple. Then,

\[
P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \mid \sum_{i=1}^{n} X_i = K) = \frac{P(X_1 = x_1, \ldots, X_n = x_n, \sum_{i=1}^{n} X_i = K)}{P(\sum_{i=1}^{n} X_i = K)}
= \frac{P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)}{P(\sum_{i=1}^{n} X_i = K)}
= \frac{\Pi_{i=1}^{n} e^{-\lambda_i} \lambda_i^{x_i}/x_i!}{e^{-\sum_{i=1}^{n} \lambda_i} (\sum_{i=1}^{n} \lambda_i)^K/K!}
= \frac{K!}{\Pi_{i=1}^{n} x_i!} \Pi_{i=1}^{n} \left( \frac{\lambda_i}{\lambda_1 + \lambda_2 + \ldots + \lambda_n} \right)^{x_i},
\]

which is all that we need to show.

- 8. Problem 54 on Page 109 of the text book. (10 points)

**Solution:** The joint density of \((X, Y)\) is,

\[ f(x, y) = \lambda^2 e^{-\lambda(x+y)}, \quad x > 0, y > 0. \]

Set

\[ U = X + Y, \quad V = X/Y. \]

The forward transformation is then,

\[ (u, v) = g(x, y) = (g_1(x, y), g_2(x, y)) = (x + y, x/y) \]

and this is a 1-1 continuously differentiable transformation from \((0, \infty) \times (0, \infty)\) to \((0, \infty) \times (0, \infty)\). The inverse transformation involves expressing \((x, y)\) in terms of \((u, v)\) and is seen to be,

\[ (x, y) = h(u, v) = (h_1(u, v), h_2(u, v)) = (uv/v + 1, u/v + 1). \]

Using the Jacobian theorem, we find that the joint density of \((U, V)\) is

\[ g(u, v) = \lambda^2 e^{-\lambda(h_1(u, v) + h_2(u, v))} \left| \frac{\partial h_1}{\partial u} - \frac{\partial h_2}{\partial v} \right| . \]
Straightforward differentiation yields,

$$
\left| \frac{\partial h_1}{\partial u} \frac{\partial h_2}{\partial v} - \frac{\partial h_2}{\partial u} \frac{\partial h_1}{\partial v} \right| = u/(v + 1)^2.
$$

Also, \( h_1(u, v) + h_2(u, v) = u \). Thus, the joint density of \((U, V)\) is,

$$
g(u, v) = \lambda^2 e^{-\lambda u} \frac{u}{(v + 1)^2},
$$

which splits as

$$
g(u, v) = \lambda^2 e^{-\lambda u} \times \frac{1}{(v + 1)^2},
$$

downing the independence of \( U \) and \( V \).