Stat 426 : Homework 1.

Moulinath Banerjee

University of Michigan

October 11, 2004

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,

$$P(A \cup B \cup C) = P((A \cup B) \cup C)$$

$$= P(A \cup B) + P(C) - P((A \cup B) \cap C)$$

$$= P(A) + P(B) - P(A \cap B) + P(C) - P((A \cap C) \cup (B \cap C))$$

$$= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

• 2. 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,

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

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,

\[
\begin{align*}
P(X > l) &= \sum_{i=1}^{\infty} P(X = i) \\
&= \sum_{i=1}^{\infty} q^{i-1} p \\
&= p \left( q + q^2 + q^3 + \ldots \right) \\
&= p q \left( 1 + q + q^2 + \ldots \right) \\
&= p q \left( 1/1 - q \right) \\
&= q^l,
\end{align*}
\]

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 $m \geq 2$,

$$P(X = m) = pq^{m-1}.$$

**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

$$P(X = m) = P(X > m - 1) - P(X > m) = q^{m-1} - q^m = pq^{m-1}.$$

But for $m = 1$,

$$P(X = m) = P(X = 1) = p = pq^{m-1},$$

trivially and the proof is complete. (5 + 5 = 10 points)
• 3. If $X$ is a 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 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.

• 4. (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 \( x > 0 \),
\[
f(x) = \frac{1}{2} \lambda \exp(-\lambda |x|).
\]

5 (i) The p.m.f of \( X_1 \) is as follows:
\[
P(X_1 = 1) = p \quad \text{and} \quad P(X_1 = 0) = 1 - p.
\]
The joint p.m.f of \((X_1, X_2)\) is:
\[
P(X_1 = 1, X_2 = 1) = P(X_1 = 1) P(X_2 = 1 | X_1 = 1) = p \frac{Np - 1}{N - 1},
\]
and by using similar conditioning arguments,
\[
P(X_1 = 1, X_2 = 0) = p \frac{N(1 - p)}{N - 1}, \quad P(X_1 = 0, X_2 = 1) = (1 - p) \frac{Np}{N - 1}
\]
and
\[
P(X_1 = 0, X_2 = 0) = (1 - p) \frac{N(1 - p) - 1}{N - 1}.
\]
To compute the p.m.f. of \( X_2 \), it suffices to find the probability that \( X_2 \) equals 1. This is computed as:
\[
P(X_2 = 1) = P(X_1 = 1, X_2 = 1) + P(X_1 = 0, X_2 = 1)
= p (Np - 1) + (1 - p) Np
= p.
\]
This shows that \( X_1 \) and \( X_2 \) have the same distribution. Also, \( X_1 \) and \( X_2 \) are not independent.
To see this, note, for example, that
\[
P(X_2 = 1 | X_1 = 1) = \frac{Np - 1}{N - 1} \neq p = P(X_2 = 1).
\]
(ii) When \( n = N \), the population size, the sample \((X_1, X_2, \ldots, X_N)\) collected without replacement is the entire population, so that the sum of the \(X_i\)'s which we denote by \(S\) gives the total number of democrats. Thus \( S \) equals \( Np \) for any sample. In other words \( P(S = Np) = 1 \). This implies that \( E(S) = Np \) and \( \text{Var}(S) = 0 \).

(iii) We have \( \hat{p} = n^{-1} (X_1 + X_2 + \ldots + X_n) \). Thus,

\[
E(\hat{p}) = n^{-1}(E X_1 + E X_2 + \ldots + E X_n) = p,
\]

since all the \( X_i\)'s have the same distribution and \( E X_1 = 1 P(X_1 = 1) + 0 P(X_1 = 0) = p \).

We have:

\[
\text{Var}(\hat{p}) = n^{-2} \text{Var} \left( \sum_{i=1}^{n} X_i \right)
\]

\[
= n^{-2} \left( \sum_{i=1}^{n} \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j) \right)
\]

\[
= n^{-2} \left( n \text{Var}(X_1) + n(n-1) \text{Cov}(X_1, X_2) \right),
\]

on using the fact that all the \( X_i\)'s have the same marginal distribution and all pairs \((X_i, X_j)\) for \( i \neq j \) have the same distribution. Now,

\[
\text{Var}(X_1) = p(1-p) \text{ and } \text{Cov}(X_1, X_2) = -p(1-p)/(N-1)
\]

by direct computation (using the marginal pmf of \( X_1 \) and the joint p.m.f. of \((X_1, X_2)\)). This leads to,

\[
\text{Var}(\hat{p}) = n^{-2} (n p(1-p) - n(n-1) p(1-p)/(N-1)) = \frac{p(1-p)}{n} \left[ 1 - \frac{n-1}{N-1} \right].
\]