Mathematics 375 – Probability and Statistics I
Final Examination Solutions – December 14, 2009

Directions

Do all work in the blue exam booklet. There are 200 possible regular points and 10 possible Extra Credit points. If an answer can be expressed in terms of binomial or multinomial coefficents, factorials, the gamma function, etc. please leave it in that form rather than evaluating. Even if you cannot see how to completely solve a problem, do not leave it blank. Partial credit will be given for relevant definitions, ideas that could lead to solutions, and so forth.

I. Twenty students in a probability and statistics class were asked to report the number of pets in their families, giving the following data:

$$3, 3, 2, 4, 1, 0, 1, 2, 3, 5, 1, 2, 2, 1, 1, 0, 2, 4, 3, 1$$

A) (10) Construct a relative frequency histogram for this data using one "bin" for each integer value.

Solution: There are two 0's, six 1's, five 2's, four 3's, two 4's, and one 5 in the data. So the relative frequency histogram will show a rectangle of height $2/20 = 1/10$ for the value 0, a rectangles of height $6/20 = 3/10$ for the value 1, $5/20 = 1/4$ for the value 2, $4/20 = 1/5$ for the value 3, $2/20 = 1/10$ for the value 4, and $1/20$ for the value 5.

B) (10) How many of the data values are within two sample standard deviations of the sample mean? How does this compare with the "empirical rule?"

Solution: The sample mean is $\overline{x} = (\sum_{i=1}^{20} x_i)/20 = 2.05$. The sample standard deviation is

$$s = \sqrt{\frac{1}{19}\sum_{i=1}^{20}(x_i - \overline{x})^2} \doteq 1.36$$

So the interval $[\overline{x} - 2s, \overline{x} + 2s]$ is approximately $[-.67, 4.77]$. 19 out of the 20 data points are in this range. This is exactly 95% of the data points, so there is close agreement with the "empirical rule."

II. Suppose you have a key ring with N keys, exactly one of which is your car key. You are parked in a country lane on a moonless night and can't see the keys or the ring. So you try the keys in your car door lock one by one until you find the right one.

A) (10) Assuming you are not careful and you mix up all the keys after each try, what is the probability that you take at least y tries to find the right key?

Solution: Since you mix the keys back up each time, you have a probability of finding the correct key of $1/N$ on each try, and the tries are independent. The number of the try on which you first find the key is a *geometric* random variable with $p = 1/N$ and $q = (N-1)/N$. The probability that you take at least y tries to find the right key is the sum of the geometric series:

$$\sum_{n=y}^{\infty} pq^{n-1} = \frac{pq^{y-1}}{1-q} = q^{y-1} = ((N-1)/N)^{y-1}.$$

B) (5) Let Y be the discrete random variable giving the number of the trial on which you find the right key. What is the expected value of Y?

Solution: From the formula for geometric random variables, $E(Y) = 1/p = N$.

C) (10) Now suppose you are more careful and you take each wrong key off the key ring and put it aside after you try it so that it does not get mixed back in with the keys you have not tried. Let Z be the discrete random variable giving the number of the trial on which you find the right key by this method. What is the expected value of Z?

Solution: Now, you have a probability of $1/N$ of finding the correct key on the first trial and a probability of $(N-1)/N$ of not finding the correct one. If you don't find the correct one, then the first choice is removed from the "pool" and you have a $1/(N-1)$ chance of finding the correct key on the second try (given that you didn't find it on the first), and an $(N-2)/(N-1)$ probability of not finding the correct one. Similarly for the later trials. The unconditional probability of finding the correct key on the zth trial is

$$\frac{N-1}{N} \cdot \frac{N-2}{N-1} \cdot \frac{N-3}{N-2} \cdots \frac{N-z+1}{N-z+2} \cdot \frac{1}{N-z+1} = \frac{1}{N}$$

(the same for all trials $z = 1, \ldots, N$), and zero for $z > N$.

$$E(Z) = 1 \cdot \frac{1}{N} + 2 \cdot \frac{1}{N} + \cdots + N \cdot \frac{1}{N} = \frac{N(N+1)}{2} \cdot \frac{1}{N} = \frac{N+1}{2}$$

As should be intuitively clear, this method is superior to the one from part A. You'll find the correct key faster on average this way!

III. (Hypothetical) In the city of Worcester, 50 percent of registered voters are Democrats, 30 percent are Republicans, and 20 percent are Independents. In the recent Senate primary, 15 percent of Democrats voted, 10 percent of Republicans voted, and 35 percent of Independents voted.

A) (10) If a single voter is selected at random, what is the probability that he or she voted in the Senate primary?

Solution: Let V be the event that the voter voted in the primary, and D, R, I be the three party affiliations. By the Law of Total Probability,

$$\begin{aligned} P(V) &= P(V|D)P(D) + P(V|R)P(R) + P(V|I)P(I) \\ &= (.15)(.5) + (.10)(.30) + (.35)(.20) \\ &= .175. \end{aligned}$$

B) (10) Given that a registered voter did vote, what is the probability that the voter was an Independent?

Solution: We want $P(I|V)$, so we apply Bayes' Rule:

$$P(I|V) = \frac{P(V|I)P(I)}{P(V|D)P(D) + P(V|R)P(R) + P(V|I)P(I)} = \frac{(.35)(.20)}{.175} = .4.$$

IV. Let X be a random variable for which $E(X) = \mu$ and $V(X) = \sigma^2$. Let c be an arbitrary constant.

A) (10) Show that $E((X-c)^2) = (\mu - c)^2 + \sigma^2$.

Solution: By the linearity of the expected value,

(1) $E((X-c)^2 = E(X^2 - 2cX + c^2) = E(X^2) - 2cE(X) + c^2 = E(X^2) - 2c\mu + c^2.$

Now by one of our standard results $V(X) = \sigma^2 = E(X^2) - \mu^2$. So $E(X^2) = \sigma^2 + \mu^2$. Therefore, substituting into (1), we have

$$E((X-c)^2) = \sigma^2 + \mu^2 - 2c\mu + c^2 = \sigma^2 + (\mu - c)^2,$$

which is what we wanted to show.

B) (5) For which c is $E((X-c)^2)$ a minimum?

Solution: Note that $(\mu - c)^2 \geq 0$ for all c and $= 0$ only if $c = \mu$. Therefore, $E((X-c)^2)$ is minimized when $c = \mu$.

V. The number of customers arriving at a fast food restaurant in a typical hour has a Poisson distribution with mean 20.

A) (5) What is the probability that 17 customers arrive in a given hour?

Solution: Let Y be the number of customers. From the Poisson pmf formula:

$$P(Y = 17) = \frac{20^{17}e^{-20}}{17!} \doteq .076.$$

B) (10) If the store is open 12 hours a day, what is the probability that exactly 17 customers will arrive in exactly 4 out of the 12 hours?

Solution: Assuming the numbers of customers in each hour are independent of each other, then the number N of hours in which 17 customers arrive has a binomial distribution with $n = 12$ and $p = .076$ from part A. Hence

$$P(N = 4) = \binom{12}{4}(.076)^4(.924)^8 \doteq .0088.$$

VI. A continuous random variable Y has p.d.f

$$f(y) = \begin{cases} c(y^2 + y) & \text{if } 1 \le y \le 2 \\ 0 & \text{otherwise} \end{cases}$$

A) (5) What is the value of c?

Solution: In order for $f(y)$ to be a legal p.d.f., we must have

$$1 = \int_1^2 c(y^2 + y)\ dy = c\left(\frac{y^3}{3} + \frac{y^2}{2}\Big|_1^2\right) = \frac{23c}{6}.$$

So $c = \frac{6}{23}$.

B) (5) What is $P(1 \le y \le 3/2)$?

Solution: This probability is given by the integral

$$\begin{aligned} P(1 \le y \le 3/2) &= \int_1^{3/2} \frac{6}{23}(y^2 + y)\ dy \\ &= \frac{6}{23}\left(\frac{y^3}{3} + \frac{y^2}{2}\Big|_1^{3/2}\right) \\ &= \frac{17}{46} \doteq .3696. \end{aligned}$$

C) (10) Find $E(Y)$ and $V(Y)$.

Solution: Proceeding as usual,

$$E(Y) = \int_1^2 y \cdot \frac{6}{23}(y^2 + y)\ dy = \frac{73}{46}.$$

Then

$$E(Y^2) = \int_1^2 y^2 \cdot \frac{6}{23}(y^2 + y)\ dy = \frac{597}{230}.$$

So

$$V(Y) = \frac{597}{230} - \left(\frac{73}{46}\right)^2 = \frac{817}{10580}.$$

VII. (15) The diameters of the bolts in a large box follow a normal distribution with mean 2 cm and a standard deviation of 0.03 cm. The diameters of the holes in the nuts in another large box are also normally distributed, but with mean 2.02 cm and standard deviation 0.04 cm. A bolt and a nut will fit together if the hole in the nut is larger than the diameter of the bolt, but the difference is no larger than .05 cm. If a nut and a bolt are selected independently and randomly from the two boxes, what is the probability that they will fit together?

Solution: Let Y_1 be the diameter of the bolt and Y_2 be the diameter of the hole in the nut. We want to know the probability

$$P(Y_1 \le Y_2 \le Y_1 + .05) = P(0 \le Y_2 - Y_1 \le .05).$$

Now by our standard results, $U = Y_2 - Y_1$ has a normal distribution with mean $2.02 - 2 = .02$ and standard deviation $\sqrt{(.03)^2 + (.04)^2} = .05$. Hence we standardize and use the normal table:

$$\begin{aligned} P(0 \le U \le .05) &= P\left(\frac{0 - .02}{.05} \le Z \le \frac{.05 - .02}{.05}\right) \\ &= P(-.4 \le Z \le .6) = 1 - .3446 - .2743 = .3811. \end{aligned}$$

VIII. Let X and Y be independent random variables with moment-generating functions

$$m_X(t) = m_Y(t) = e^{t^2+3t}$$

A) (10) What is the moment generating function of $Z = 3X + 2Y - 4$?

Solution: The moment generating function of Z is

$$m_Z(t) = m_X(3t)m_Y(2t)e^{-4t} = e^{9t^2+9t} \cdot e^{4t^2+6t} \cdot e^{-4t} = e^{13t^2+11t}.$$

B) (10) What is the distribution of Z?

Solution: By the Uniqueness Theorem, this is the moment-generating function of a normally-distributed random variable with mean $\mu = 11$ and variance $\sigma^2 = 26$.

IX. Twenty students in a probability and statistics class take a final examination independently of one another. The number of minutes each student requires to complete the exam is a random variable with an exponential distribution with mean 120 minutes. The students all start work at 8:30am.

A) (10) What is the probability that at least one student out of the 20 will complete the exam by 10:00 am?

Solution: Let Y be the time that one student takes to finish. The probability we want is the complementary probability to the probability of the event that all the students take 90 minutes or more. Since

$$P(Y \geq 90) = \int_{90}^{\infty} e^{-y/120}/120 \; dy = e^{-90/120} = e^{-3/4} \doteq .4724,$$

the desired probability is

$$1 - \left(e^{-3/4}\right)^{20}.$$

(This is very close to 1.) An alternate method is to compute $P(Y \leq 90) = .5276$, and then use the binomial probability:

$$\sum_{y=1}^{20} \binom{20}{y} (.5276)^y (.4724)^{20-y}.$$

B) (10) What is the distribution of the total time taken by all of the students to complete the exam?

Solution: Using moment-generating functions, by the independence, if $T = Y_1 + \cdots + Y_{20}$ is the total time,

$$m_T(t) = \left(\frac{1}{(1-120t)}\right)^{20} = \frac{1}{(1-120t)^{20}}.$$

By the Uniqueness Theorem, we see that T has a gamma distribution with $\alpha = 20$ and $\beta = 120$.

X. Suppose Y_1, Y_2 have joint density

$$f(y_1, y_2) = \begin{cases} 24y_1y_2 & \text{if } y_1 \geq 0,\ y_1 + y_2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

A) (10) Are Y_1, Y_2 independent?

Solution: We need the marginal densities to determine this. Computing,

$$f_1(y_1) = \int_0^{1-y_1} 24y_1y_2 \; dy_2 = 12y_1(1-y_1)^2$$

if $0 \leq y_1 \leq 1$ and zero otherwise. For future reference, note that this says Y_1 has a beta distribution with $\alpha = 2$ and $\beta = 3$. By symmetry,

$$f_2(y_2) = \int_0^{1-y_2} 24y_1y_2 \; dy_1 = 12y_2(1-y_2)^2$$

if $0 \le y_2 \le 1$ and zero otherwise. So Y_1 and Y_2 are actually identically distributed(!) Since $f_1(y_1)f_2(y_2) \neq f(y_1, y_2)$, Y_1 and Y_2 are *not independent.*

B) (10) What is $V(8Y_1 - 2Y_2)$?

Solution: There are many ways to do this, some harder and some easier. Perhaps the most "painless" way is to use the observation made above in part A that Y_1 and Y_2 both have beta distributions. This means that we can immediately write down:

$$E(Y_1) = E(Y_2) = \frac{\alpha}{\alpha+\beta} = \frac{2}{5}$$

and

$$V(Y_1) = V(Y_2) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{1}{25}.$$

The only integral we need to compute is

$$E(Y_1Y_2) = \int_0^1 \int_0^{1-y_1} 24y_1^2y_2^2 \, dy_2 \, dy_1 = \frac{2}{15}.$$

Then

$$\mathrm{Cov}(Y_1, Y_2) = \frac{2}{15} - \left(\frac{2}{5}\right)^2 = -\frac{2}{75}.$$

Finally,

$$\begin{aligned} V(8Y1 - 2Y_2) &= 64V(Y_1) + 4V(Y_2) - 32\mathrm{Cov}(Y_1, Y_2) \\ &= \frac{64}{25} + \frac{4}{25} + \frac{64}{75} \\ &= \frac{268}{75} \doteq 3.5733. \end{aligned}$$

C) (10) Use the method of distribution functions to find the density function for $U = Y_1 + Y_2$.

Solution: We have

$$\begin{aligned} F_U(u) &= P(U \le u) \\ &= P(Y_1 + Y_2 \le u) \\ &= \int_0^u \int_0^{u-y_1} 24y_1y_2 \, dy_2 \, dy_1 \\ &= u^4 \end{aligned}$$

if $0 \le u \le 1$, 1 if $u \ge 1$, and zero otherwise. Then $f_U(u) = 4u^3$ if $0 \le u \le 1$ and zero otherwise. (That is, U has a beta distribution with $\alpha = 4$ and $\beta = 1$.)

Extra Credit (10) One form of the Law of Large Numbers states that if $X_1, \ldots, X_n$ are independent and identically distributed random variables for which $E(X_i) = \mu$ and $V(X_i) = \sigma^2$ exist, then for any $\epsilon > 0$,

$$\lim_{n\to\infty} P(|\overline{X} - \mu| < \epsilon) = 1.$$

Prove this statement without assuming anything else about the distribution of the X_i.

Solution: The random variable $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ has mean μ and variance $\sigma^2 = \frac{1}{n^2}V(X_1) + \cdots + \frac{1}{n^2}V(X_n) = \frac{\sigma^2}{n}$. For a given ϵ, we apply Tchebysheff's Theorem with $k = \frac{\epsilon n}{\sigma^2}$:

$$\begin{aligned} P(|\overline{X} - \mu| < \epsilon) &= P(|\overline{X} - \mu| < k\frac{\sigma^2}{n}) \\ &\geq 1 - \frac{\sigma^4}{\epsilon^2 n^2}. \end{aligned}$$

As $n \to \infty$ the last line here approaches 1, which implies what we wanted to show.