Mathematics 375 – Probability Theory
Solutions for Final Examination – December 16, 2011

I. A large number of observations of a certain continuous random variable Y were taken, and a typical subset of $n = 50$ of them were ordered to produce the data set below:

0.10	0.23	0.25	0.26	0.26	0.28	0.31	0.39	0.42	0.44
0.44	0.48	0.53	0.54	0.55	0.66	0.69	0.70	0.72	0.73
0.73	0.74	0.74	0.76	0.77	0.78	0.81	0.95	0.97	1.02
1.03	1.06	1.07	1.13	1.13	1.14	1.19	1.22	1.27	1.28
1.30	1.34	1.34	1.37	1.39	1.40	1.43	1.43	1.46	1.48

A) (10) Construct a relative frequency histogram for this data using 10 equal bins on the interval $\left[0, \frac{3}{2}\right]$.

Solution: Each bin will have width $1.5/10 = .15$, so they correspond to the intervals

$$[0, .15], [.15, .3], [.3, .45], [.45, .6], \ldots, [.9, 1.05], [1.05, 1.2], [1.2, 1.35], [1.35, 1.5]$$

There are: $1, 5, 5, 4, 8, 4, 4, 6, 6, 7$ values in these bins (respectively), so the relative frequency histogram will have boxes of heights 1/50, 1/10, 1/10, 2/25, etc.

B) (20) Two models are proposed for the distribution of Y:

1. a *uniform* distribution on $\left[0, \frac{3}{2}\right]$, or
2. a distribution described by the probability density function (pdf)

$$f(y) = \begin{cases} \frac{8}{9}\left(\frac{3}{2} - y\right) & \text{if } 0 \le y \le \frac{3}{2} \\ 0 & \text{otherwise} \end{cases}$$

Find the expected values and variances of the two proposed model distributions.

Solution: The uniform distribution has $\mu_1 = E(Y) = \frac{0+1.5}{2} = .75$ and $V(Y) = \frac{(1.5-0)^2}{12} = .1875$, so $\sigma_1 \doteq .4330$. The distribution from 2. has

$$\mu_2 = E(Y) = \int_0^{1.5} y \frac{8}{9}\left(\frac{3}{2} - y\right)\, dy = .5$$

and

$$V(Y) = \int_0^{1.5} y^2 \frac{8}{9}\left(\frac{3}{2} - y\right)\, dy - (.5)^2 = .125$$

The standard deviation is $\sigma_2 \doteq .3536$.

C) (5) The actual sample mean is $\overline{y} = .8542$ and the sample standard deviation is $s \doteq .4033$. Is either of the models in part B clearly more likely as the underlying distribution from which the values in part A were generated? Explain.

Solution: We have that μ_1 is significantly closer to $\overline{y}$ and σ_1 closer to s. So the uniform distribution seems more likely.

II. (15) A bin contains five components from supplier A, three from supplier B, and four from supplier C. If three of the components are selected randomly (without replacement) for testing, what is the probability that the components chosen will come from at most two of the suppliers?

Solution: There are 12 components in the bin, and we are selecting three of them without replacement. Hence there are $\binom{12}{3}$ different collections that can be chosen. The easiest way to do this is to compute the complementary probability of the probability that one from each supplier is included:

$$1 - \frac{\binom{5}{1}\binom{3}{1}\binom{4}{1}}{\binom{12}{3}}$$

The "brute force" method can also be used. Out of all ways of choosing three of the components, there are $\binom{5}{3}$ all from A, $\binom{3}{3}$ all from B, $\binom{4}{3}$ all from C, $\binom{5}{2}\binom{3}{1}$ consisting of one from A and two from B, etc. Hence the probability as stated is also equal to:

$$\frac{\binom{5}{3} + \binom{3}{3} + \binom{4}{3} + \binom{5}{2}\binom{3}{1} + \binom{5}{1}\binom{3}{2} + \binom{5}{2}\binom{4}{1} + \binom{5}{1}\binom{4}{2} + \binom{3}{2}\binom{4}{1} + \binom{3}{1}\binom{4}{2}}{\binom{12}{3}}.$$

III. Rutabaga seeds from grower A have an 90% germination rate, while seeds from grower B have a 80% germination rate and those from grower C have a 85% germination rate. A seed packaging company buys 25% of its seeds from grower A, 35% from grower B, and 40% from grower C, then mixes them thoroughly in making up packets for sale.

A) (10) What is the probability that a randomly selected seed will germinate?

Solution: By the Law of Total Probability:

$$\begin{aligned} P(G) &= P(G \mid A)P(A) + P(G \mid B)P(B) + P(G \mid C)P(C) \\ &= (.9)(.25) + (.8)(.35) + (.85)(.4) \\ &= .845. \end{aligned}$$

B) (10) Given that a seed did not germinate, what is the probability that it came from grower B?

Solution: By the definition of conditional probabilities (or Bayes' Rule),

$$P(B \mid \overline{G}) = \frac{P(\overline{G} \mid B)P(B)}{P(\overline{G})} = \frac{(.2)(.35)}{(.155)} \doteq .4516.$$

IV. A permutation of $A = \{1, 2, 3, 4, 5\}$ is a one-to-one, onto mapping from this set to itself. For instance, the f defined by

$$f(1) = 2, f(2) = 4, f(3) = 3, f(4) = 1, f(5) = 5$$

is one such permutation. Consider a permutation selected at random, and let Y be the discrete random variable representing *number of elements of the set A that are mapped to themselves*. For instance, if the f above was selected, then the value of Y is 2 (since $f(3) = 3$ and $f(5) = 5$). For this problem, you are given the information that the moment-generating function of Y is

$$m(t) = \frac{44}{120} + \frac{45}{120}e^{t} + \frac{20}{120}e^{2t} + \frac{10}{120}e^{3t} + \frac{1}{120}e^{5t}$$

A) (10) Find the mean and variance of Y.

Solution: The mean of Y is given by

$$E(Y) = m'(0) = \frac{45+40+30+5}{120} = 1.$$

The variance is

$$E(Y^2) - (E(Y))^2 = m''(0) - 1 = \frac{45+80+90+25}{120} - 1 = 1.$$

B) (10) What is the probability that a randomly selected permutation maps *no element of A to itself*?

Solution: Since Y is a discrete random variable, the moment-generating function is

$$m(t) = \sum_y e^{ty} p(y)$$

Therefore, the coefficients in the formula given for $m(t)$ must be the probabilities that Y takes the corresponding values:

$$\begin{aligned} P(Y=0) &= \frac{44}{120} \\ P(Y=1) &= \frac{45}{120} \\ P(Y=2) &= \frac{20}{120} \\ P(Y=3) &= \frac{10}{120} \\ P(Y=4) &= 0 \\ P(Y=5) &= \frac{1}{120} \end{aligned}$$

The probability we want is $P(Y=0) = \frac{44}{120} = \frac{11}{30}$.

V. A candy maker produces candy bars that have a stated label weight of 57.4 grams. The manufacturing process is subject to some randomness, though, and the actual weights of the bars are normally distributed with mean $\mu = 58$ grams and $\sigma = 2$ grams.

A) (5) What is the probability that a single candy bar has weight < 57 grams?

Solution: Say the weight of a single candy bar is $Y \sim Normal(58, 2^2)$. Then

$$P(Y < 57) = P\left(\frac{Y-58}{2} < \frac{57-58}{2}\right) = P(Z < -1/2) = P(Z > 1/2) = .3085.$$

B) (5) If candy bars are selected independently and at random from the production line, what is the probability that at least 4 trials will be necessary to find one that has weight < 57 grams?

Solution: If N is the number of the trial on which the first bar with weight < 57 grams is found, then N has a geometric distribution with $p = .3085$ and $q = .6915$. So

$$P(N \geq 4) = 1 - P(N \leq 3) = 1 - .3085 - (.6915)(.3085) - (.6915)^2(.3085) = .3307$$

This is the same as the sum of the geometric series

$$\sum_{n=4}^{\infty} q^{n-1}p = \frac{q^3 p}{1-q} = q^3.$$

C) (5) If 20 candy bars are selected independently and at random from the production line, what is the probability that 3 or fewer will have weight < 57 grams?

Solution: (Assuming the selection is done without replacement), the number M of bars with weight < 57 out of the 20 has a $Binomial(20, .3085)$ distribution, so the probability is

$$P(M \leq 3) = \sum_{m=0}^{3} \binom{20}{m} (.3085)^m (.6915)^{20-m} \doteq .0931$$

D) (10) The production process can be changed to alter the average weight μ of the bars, but $\sigma = 2$ cannot be changed. What value for μ will make the probability that a candy bar weighs < 57.4 grams equal to .01?

Solution: We want

$$.01 = P(Y < 57.4) = P\left(\frac{Y-\mu}{2} < \frac{57.4-\mu}{2}\right) = P\left(Z < \frac{57.4-\mu}{2}\right)$$

Hence $\frac{57.4-\mu}{2} = -2.33$ approximately. Solving for μ, $\mu = 57.4 + 2(2.33) = 62.06$.

VI.

A) (10) State the Central Limit Theorem.

Solution: Let $Y_1, \ldots, Y_n$ be independent, identically distributed random variables, all with finite expected value μ and variance σ^2. Let $\overline{Y} = \frac{1}{n}(Y_1 + \cdots + Y_n)$ as usual. Then

$$U_n = \frac{\overline{Y} - \mu}{\sigma/\sqrt{n}}$$

converges to a standard normal Z in distribution as $n \to \infty$.

B) (10) From experience, a charity knows that the contributions it receives have mean $\mu = 3125$ and $\sigma = 250$. If $n = 2000$ of the contributions are selected at random, estimate the probability that the average of those 2000 contributions is > 3200.

Solution: There are several ways to attack this. I gave full credit for either of the following approaches.

Method 1: The number $n = 2000$ is large enough that we can think of using the Central Limit Theorem to estimate this probability. The distribution of $\overline{Y}$ will be approximately normal with mean 3125 and SD $250/\sqrt{2000}$. So

$$P(\overline{Y} > 3200) = P\left(\frac{\overline{Y} - 3125}{250/\sqrt{2000}} > \frac{3200 - 3125}{250/\sqrt{2000}}\right) \doteq P(Z > 13.42)$$

This is essentially $= 0$.

Method 2: Since we don't know the actual distribution of the contribution amounts, it would also make sense to estimate this probability using Tchebysheff's theorem. For any $k > 0$ we have (writing $\overline{\mu}$ and $\overline{\sigma}$ for the mean and SD of $\overline{Y}$),

$$P(|\overline{Y} - \overline{\mu}| > k\overline{\sigma}) < \frac{1}{k^2}$$

So since $3200 = 3125 + (13.42)\overline{\sigma}$,

$$P(\overline{Y} > 3200) + P(\overline{Y} < 3050) < \frac{1}{13.42^2} \doteq .0056$$

(This is a much weaker bound, of course.)

VII. Let Y be a random variable with a pdf of the form

$$f(y) = \begin{cases} ky^4 e^{-2y} & \text{if } y > 0 \\ 0 & \text{otherwise} \end{cases}$$

A) (15) Determine the value of k and find $E(Y)$ and $V(Y)$.

Solution: We recognize this as a Gamma density with $\alpha = 5$ and $\beta = 1/2$. Hence

$$k = \frac{1}{(1/2)^5 \Gamma(5)} = \frac{32}{24} = \frac{4}{3}.$$

By the formulas for Gamma distributions, $E(Y) = \alpha\beta = 5/2$ and $V(Y) = \alpha\beta^2 = 5/4$. (These can also be computed directly, of course.)

B) (10) Let $Y_1, \ldots, Y_n$ be independent random variables all with density as above. What is the distribution of $\overline{Y} = \frac{1}{n}(Y_1 + \cdots + Y_n)$?

Solution: We know that

$$m_{Y_i}(t) = \frac{1}{\left(1 - \frac{t}{2}\right)^5}$$

(for $1 - t/2 > 0$, or $t < 2$) for all i. Since $Y_1, \ldots, Y_n$ are assumed independent, so are $\frac{1}{n}Y_i$. The moment-generating function

$$m_{\frac{1}{n}Y_i} = \frac{1}{\left(1 - \frac{t}{2n}\right)^5},$$

so multiplying n of these together, we get:

$$m_{\overline{Y}} = \frac{1}{\left(1 - \frac{t}{2n}\right)^{5n}}$$

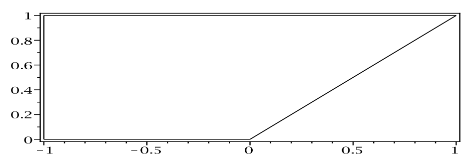
By the Uniqueness Theorem, $\overline{Y}$ has a Gamma distribution with $\alpha = 5n$, and $\beta = \frac{1}{2n}$.

VIII. Let Y_1 and Y_2 be random variables with joint pdf

$$f(y_1, y_2) = \begin{cases} 3y_1^2 y_2 & \text{if } -1 \le y_1 \le 1, 0 \le y_2 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

A) (10) What is $P(Y_2 > Y_1)$? (*Hint:* Draw a picture!)

Solution: We want to integrate the joint density over the region above the line $y_2 = y_1$, and inside the rectangle where $f(y_1, y_2)$ is nonzero, shown in this figure:



The easiest way to set this up is to integrate with respect to y_1 first (since the left- and right-hand boundary curves are the same for all y_2):

$$P(Y_2 > Y_1) = \int_0^1 \int_{-1}^{y_2} 3y_1^2 y_2 \; dy_1 dy_2 = 7/10.$$

B) (10) Are Y_1 and Y_2 independent? Why or why not?

Solution: Yes, since the joint density has the form $g(y_1)h(y_2)$ and is nonzero only on a rectangular region. (See Theorem 5.5 in our text).

C) (10) Determine $V(3Y_1 + 2Y_2)$.

Solution: By the general formulas for linear combinations,

$$V(3Y_1 + 2Y_2) = 9V(Y_1) + 4V(Y_2) + 12\text{Cov}(Y_1, Y_2).$$

The covariance term is zero because of the result of part B. We can compute the variances of Y_1 and Y_2 in several ways. Here is the most direct approach.

$$\begin{aligned}
E(Y_1) &= \int_{-1}^1 \int_0^1 y_1 \cdot 3y_1^2 y_2 \; dy_2 dy_1 = 0 \\
E(Y_1^2) &= \int_{-1}^1 \int_0^1 y_1^2 \cdot 3y_1^2 y_2 \; dy_2 dy_1 = 3/5 \\
E(Y_2) &= \int_{-1}^1 \int_0^1 y_2 \cdot 3y_1^2 y_2 \; dy_2 dy_1 = 2/3 \\
E(Y_1^2) &= \int_{-1}^1 \int_0^1 y_1^2 \cdot 3y_1^2 y_2 \; dy_2 dy_1 = 1/2 \\
V(Y_1) &= 3/5 \\
V(Y_2) &= 1/18 \\
V(3Y_1 + 2Y_2) &= \frac{27}{5} + \frac{4}{18} = \frac{253}{45}.
\end{aligned}$$

D) (10) Let $U = Y_1^2$. Determine the pdf for U.

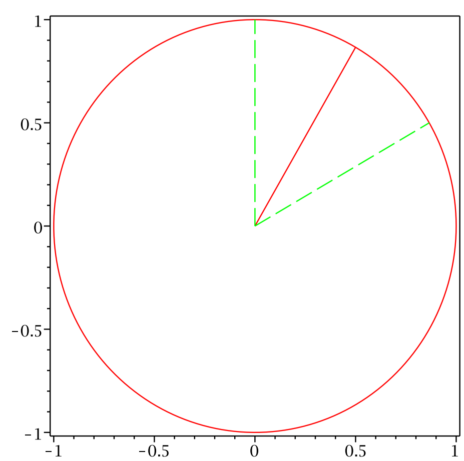
Solution: If $u = 0$, we have $P(U \le u) = 0$. For $u \ge 0$ we have

$$\begin{aligned} F_U(u) &= P(U \le u) \\ &= P(Y_1^2 \le u) \\ &= P(-\sqrt{u} \le Y_1 \le \sqrt{u}) \\ &= \begin{cases} \int_{-\sqrt{u}}^{\sqrt{u}} \int_0^1 3y_1^2 y_2 \; dy_2 dy_1 & \text{if } 0 \le u \le 1 \\ 1 & \text{if } u > 1 \end{cases} \\ &= \begin{cases} u^{3/2} & \text{if } 0 \le u \le 1 \\ 1 & \text{if } u > 1 \end{cases} \\ \text{Hence } f_U(u) &= F_u'(u) \\ &= \begin{cases} \frac{3}{2}u^{1/2} & \text{if } 0 \le u \le 1 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Extra Credit. (20) A volcano is located at the center $(x, y) = (0, 0)$ of a circular island 1 mile in radius. When the volcano erupts, it emits lava that spreads over a random sector of the island. The diagram below shows one such sector between two radii drawn with dashes. The middle radius of the sector (solid line in diagram) makes an random angle with the positive x-axis, Θ, which is uniformly distributed on the interval $[-\pi, \pi]$. The total central angle of the sector (the angle between the two dashed radii in the diagram) is a second random variable Φ that has pdf

$$f(\varphi) = \begin{cases} c(\pi/2 - \varphi) & \text{if } 0 \le \varphi \le \pi/2 \\ 0 & \text{otherwise} \end{cases}$$

for some constant c. The angles Θ and Φ are independent.



You own a villa at $(x, y) = (1, 0)$ on the coast of the island. What is the probability that any single eruption buries your villa in lava?

Solution: The density for Θ is $g(\theta) = \frac{1}{2\pi}$ for $-\pi \le \theta \le \pi$ and 0 otherwise. The constant c in the density for Φ can be determined as usual: $1 = \int_0^{\pi/2} c(\pi/2 - \varphi)\ d\varphi = c\pi^2/8$, so $c = 8/\pi^2$. Since Θ and Φ are independent, the joint density is the product

$$\frac{1}{2\pi} \cdot \frac{8}{\pi^2}(\pi/2 - \varphi)$$

on the rectangle $-\pi \le \theta \le \pi$ and $0 \le \varphi \le \pi/2$, and zero outside that rectangle. Now, the lava from an eruption will bury your villa at $(1, 0)$ on the x-axis if the angle 0 is in the interval of polar angles corresponding to the sector covered by the lava:

$$\Theta - \Phi/2 \le 0 \le \Theta + \Phi/2,$$

or equivalently,

$$-\Phi/2 \le \Theta \le \Phi/2.$$

The probability this happens is

$$\int_0^{\pi/2} \int_{-\varphi/2}^{\varphi/2} \frac{1}{2\pi} \cdot \frac{8}{\pi^2}(\pi/2 - \varphi)\ d\theta d\varphi = \frac{1}{12}.$$

(Note: This can be rearranged to show that the probability we want is equal to

$$\frac{E(\Phi)}{2\pi} = \frac{\pi}{6} \cdot \frac{1}{2\pi}$$

Better make sure your homeowner's insurance is up to date!)