MATH 371
Midterm 2

Name : SOLUTIONS

November 8, 2017

Instructions: Give numerical answers unless instructed that a formula suffices. Round answer to two decimal places when appropriate. Show work for partial credit. Good luck.

1. (20 pts. – 5 pts. each)

(a) The lifetime of an oil-drilling bit is approximately normally distributed with mean 90 hours and standard deviation 15 hours. What is the probability that the drill bit will fail before 70 hours of use?

$\mu = 90$ hrs $\sigma = 15$ hrs $Z = \frac{Y-90}{15} \sim N(0,1)$.

Symmetry

$$P(Y \le 70) = P\left(\frac{Y-90}{15} \le \frac{70-90}{15}\right) = P\left(Z \le \frac{-20}{15}\right) = P(Z \le -1.\bar{33}) = P(Z \ge 1.\bar{33}) = \boxed{.0918}$$

Table

(b) The amount of snowfall over a one week period in November is uniformly distributed with a minimum value of 0 inches and a maximum value of 7 inches. What is the probability that 5 or more inches falls during one week in November?

$P(S \ge 5)$ $S \sim Unif(0,7)$

$= \boxed{2/7}$

(c) The magnitude of earthquakes in the Aleutians has a mean of 4 on the Richter scale and is often modeled with an exponential distribution. Find the probability that an earthquake in the Aleutians has magnitude 5 or greater on the Richter scale.

$M \sim Exp(4)$

$$P(M \ge 5) = \int_5^\infty \frac{1}{4} e^{-m/4}\, dm = \lim_{R\to\infty} -e^{-m/4}\Big|_5^R = 0 - (-e^{-5/4}) = \boxed{\frac{1}{e^{5/4}} \approx .29}$$

(d) According to the 2010 US census data, 71.2% of the population live in urbanized areas ($\geq 50,000$ people), 9.5% of the population live in urban clusters ($2,500 - 50,000$ inhabitants), and the rest live in rural areas.

i. In a sample of size $n = 10$, give a formula that computes the probability that in this sample, 6 live in urbanized areas, 3 in urban clusters, and 1 in a rural area. (Do not evaluate.)

Multinomial $(10;\ .712, .095, .193)$

$$\text{Prob} = \binom{10}{6\ 3\ 1} (.712)^{10} (.095)^{3} (.193)^{1}$$

ii. In this sample, how many people do you expect to live in a rural area?

$\mathbb{E}(\text{Rural Area}) = n p_3 = 10(.193) = 1.93$

2. (10 pts. – 5 pts. each) Let X be a continuous random variable with density function given by

$$f(x) = \begin{cases} \frac{1}{4}(x+1), & \text{if } 0 \le x \le 2 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Give the cumulative distribution function $F_X(x)$ for X. (Be sure to specify values of $F_X(x)$ for all real numbers x.)

If $0 < x < 2$, then $F_X(x) = \int_0^x \frac{1}{4}(t+1)dt = \frac{1}{8}t^2 + \frac{1}{4}t \Big|_0^x = \frac{1}{8}x^2 + \frac{1}{4}x$

Thus, $\boxed{F_X(x) = \begin{cases} 0, & x \le 0 \\ 1/8\, x^2 + 1/4\, x, & 0 < x < 2 \\ 1, & x > 2 \end{cases}}$

(b) Set up, **but do not integrate**, an integral to compute the expected value of $U = 3X - 1$.

$\mathbb{E}(u) = \mathbb{E}(3X-1) = \int_0^2 (3x-1)\frac{1}{4}(x+1)dx = \boxed{\frac{1}{4}\int_0^2 3x^2 + 2x - 1 \, dx}$

3. (12 pts.) Nutritionists recommend that people get between 20% and 35% of their daily calories from healthy fats. In a particular US state, data is collected and in this population it is found that that the mean is 40% daily calories from fats with variance $\sigma^2 = .04$.

(a) (7 pts.) You want to model the distribution of percentage of daily calories from fat using one of the following: Uniform, Normal, Exponential, Gamma, or Beta distributions. Using your knowledge of the qualitative features of these distributions, which one is the most plausible candidate? Explain briefly, giving some rationale for rejecting or retaining each of the distributions as a model.

Beta! The random variable is X: proportion of daily calories from fats.

Reject: Normal since range is all of $\mathbb{R}$

Uniform: wrong shape

Exponential / Gamma since these have range all $x > 0$.

(b) (5 pts.) Using the values of the mean and variance that are given, for the distribution you specified in part (a) decide what parameter values you would use to model the distribution of daily fat intake percentage and find $P(Y \ge .5)$. (*Hint:* Try positive integers less than 10.)

$\mu = \frac{\alpha}{\alpha+\beta} = .4 \qquad \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = .04$

Definitely want $\alpha, \beta > 1$ so the density's shape is correct.

By trial and error, $\alpha = 4, \beta = 6$ has wrong variance, but $\alpha = 2, \beta = 3$ has

$\mu = \frac{2}{2+3} = .4 \qquad \sigma^2 = \frac{2(3)}{(2+3)^2(2+3+1)} = \frac{1}{5^2} = \frac{1}{25} = .04!$ $\qquad \boxed{\alpha = 2, \beta = 3}$

3) continued ...

$$P(Y \geq .5) = \int_0^{.5} \frac{1}{n} y^1 (1-y)^2 \, dy \qquad \text{where} \quad n = \int_0^1 y(1-y)^2 \, dy$$

Alternatively, $\frac{1}{n} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} = \frac{\Gamma(5)}{\Gamma(2)\Gamma(3)} = \frac{4!}{1 \cdot 2!} = 12$

The full calculation is $n = \int_0^1 y - 2y^2 + y^3 \, dy = \frac{1}{2}y^2 - \frac{2}{3}y^3 + \frac{1}{4}y^4 \Big|_0^1 = \frac{3}{4} - \frac{2}{3} = \frac{1}{12}$

Either way,

$$P(Y \geq .5) = \int_{1/2}^1 12y(1-y)^2 \, dy = 12\left[\frac{1}{2}y^2 - \frac{2}{3}y^3 + \frac{1}{4}y^4\right]\Bigg|_{1/2}^1$$

$$= 12\left[\frac{1}{12} - \left(\frac{1}{8} - \frac{2}{3} \cdot \frac{1}{8} + \frac{1}{64}\right)\right] = 12\left[\frac{1}{12} - \frac{1}{8} + \frac{1}{12} - \frac{1}{64}\right] = 2 - 12\left[\frac{1}{8} + \frac{1}{64}\right]$$

$$= 2 - \frac{12 \cdot 9}{64} = 2 - \frac{27}{16} = \boxed{\frac{5}{16}} \approx .3125$$

4. (20 pts. – 5 pts. each) Let Y_1 denote gender of student and Y_2 the number of classes taken during the fall semester. Consider the following table which gives the joint probability function $p(y_1, y_2)$:

		Female ($Y_1 = 0$)	Male ($Y_1 = 1$)	$p_2(y_2)$
Y_2	1	.08	.10	.18
	2	.09	.04	.13
	3	.11	.10	.21
	4	.16	.09	.25
	5	.12	.04	.16
	6	.05	.02	.07
$p_1(y_1)$		.61	.39	

(a) Give $P(Y_1 \geq 0, Y_2 \geq 4)$.

$P(Y_1 \geq 0, Y_2 \geq 4) = p(1,4) + p(2,4) + p(1,5) + p(2,5) + p(1,6) + p(2,6)$

$= .16 + .09 + .12 + .04 + .05 + .02$

$= \boxed{.48}$

(b) Give $P(Y_2 \geq 3 \mid Y_1 = 0)$. Explain briefly that meaning of this probability.

$$P(Y_2 \geq 3 \mid Y_1 = 0) = \frac{P(Y_1 = 0, Y_2 \geq 3)}{P(Y_1 = 0)} = 1 - \frac{.08 + .09}{p_1(0)} = 1 - \frac{.17}{.61} \approx 1 - .2767 \approx \boxed{.72}$$

"Probability of taking at least 3 classes given that the student is female"

(c) Give the probability $P(Y_2 < 2)$. See $p_2(y_2)$ above

$P(Y_2 < 2) = p_2(1) = \boxed{.18}$

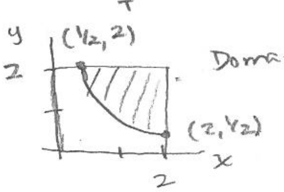
(d) Are Y_1, and Y_2 independent? *Prove* your answer.

No. For instance, $p(0,1) = .08 \neq p_1(0)\,p_2(1) = .61(.18) \approx .11$

Dependent.

5. (10 pts.) The random variables X, Y are independent and uniformly distributed over the interval from 0 to 2. What is the probability that $XY \geq 1$? (Evaluate any integrals in your answer.)

Since X and Y are independent and Unif(0,2), there joint density on the square is $f(x,y) = f_x(x)\, f_y(y) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$ $\quad 0 \leq x \leq 2,\ 0 \leq y \leq 2$



Domain of $f(x,y)$

To find $P(XY \geq 1)$, consider those points (x, y) with $y \geq \frac{1}{x}$ which is shown in the shaded area.

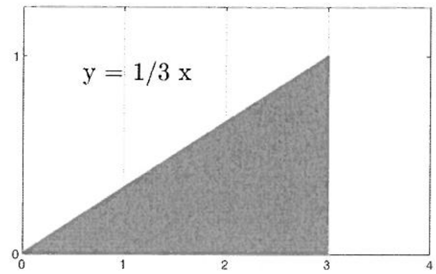
$$P(XY \geq 1) = \int_{1/2}^{2} \int_{1/x}^{2} \frac{1}{4}\, dy\, dx = \int_{1/2}^{2} \frac{1}{4}(2 - 1/x)\, dx = \frac{1}{2}x - \frac{1}{4}\ln|x| \Big|_{1/2}^{2} = (1 - \tfrac{1}{4}\ln 2) - (\tfrac{1}{4} - \tfrac{1}{4}\ln(1/2))$$

$y = 2$, $y = 1/x$

$$= \frac{3}{4} - \frac{1}{4}(\ln 2 - \ln \tfrac{1}{2}) = \frac{3}{4} - \frac{1}{4}(\ln 2 + \ln 2) = \boxed{\frac{3}{4} - \frac{1}{2}\ln 2 \approx .40}$$

6. (20 pts. – 5 pts. each) Suppose X and Y are jointly continuous random variables with joint density function

$$f(x,y) = \begin{cases} 2y, & \text{if } 0 \le y \le \frac{1}{3}x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$



(a) Find the marginal density function for Y. (Include the domain in your answer.)

$$f_Y(y) = \int_{3y}^{3} 2y\,dt = 2yt\Big|_{3y}^{3} = 2y(3-3y) = 6y - 6y^2$$

Therefore, $f_Y(y) = \begin{cases} 6y - 6y^2, & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$

(b) Find the conditional density function of X, given $Y = y$. (Include the domain in your answer.)

$$f(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{2y}{2y(3-3y)} = \frac{1}{3-3y} \quad \text{for } 0 < y < 1, \ 0 < x < 3$$

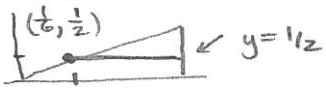
Notice strict inequalities

$f(x|y) = 0$ for $y \in (-\infty, 0) \cup (1, \infty)$

(c) Find $P(X \ge 2 \mid Y = \frac{1}{2})$.

First, $f(x \mid y = 1/2) = \dfrac{1}{3 - 3(\frac{1}{2})} = \dfrac{2}{3}$. Thus, $P(X \ge 2 \mid Y = 1/2) = \displaystyle\int_2^3 \frac{2}{3}\,dx = \boxed{2/3}$

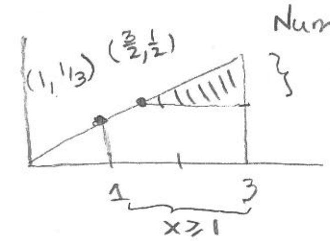
Checking the domain. All okay.



(d) Set up, **but do not integrate**, a formula to compute $P(X \ge 1 \mid Y \ge \frac{1}{2})$. A complete answer has formulas for integrands, not names of functions like $f(x,y)$, and limits of integration.

Note this.

$$P(X \ge 1 \mid Y \ge 1/2) = \frac{P(X \ge 1, Y \ge 1/2)}{P(Y \ge 1/2)} = \frac{P(X \ge \frac{3}{2}, Y \ge \frac{1}{2})}{P(Y \ge 1/2)}$$



Numerator: $\} \ Y \ge 1/2$

$$= \boxed{\frac{\displaystyle\int_{3/2}^{3} \int_{1/2}^{1/3 x} 2y\,dy\,dx}{\displaystyle\int_{1/2}^{1} 6y - 6y^2\,dy}} = 1 \text{ from geometry}$$

7. (13 pts.) Suppose Y_1, Y_2, Y_3 are random variables with

$$E(Y_1) = -1, \quad E(Y_2) = 2, \quad E(Y_3) = 4,$$

$$V(Y_1) = 3, \quad V(Y_2) = 5, \quad V(Y_3) = 4,$$

$$\text{Cov}(Y_1, Y_2) = -1, \quad \text{Cov}(Y_1, Y_3) = 1, \quad \text{Cov}(Y_2, Y_3) = 0.$$

(a) (3 pts.) Find $E(3Y_1 - 2Y_2 + Y_3)$.

$$\mathbb{E}(3Y_1 - 2Y_2 + Y_3) = 3\,\mathbb{E}(Y_1) - 2\,\mathbb{E}(Y_2) + \mathbb{E}(Y_3) = 3(-1) - 2(2) + 4 = \boxed{-3}$$

(b) (3 pts.) Find $V(3Y_1 - 2Y_2 + Y_3)$.

$$V(3Y_1 - 2Y_2 + Y_3) = (3)^2 V(Y_1) + (-2)^2 V(Y_2) + (1)^2 V(Y_3) + 2(3)(-2)\,\text{Cov}(Y_1, Y_2) + 2(3)(1)\,\text{Cov}(Y_1, Y_3)$$
$$+ 2(-2)(1)\,\text{Cov}(Y_1, Y_3)$$

$$= 9(3) + 4(5) + 4 - 12(-1) + 6(1) - 4(0)$$

$$= 27 + 20 + 4 + 12 + 6$$

$$= \boxed{69}$$

(c) (3 pts.) Suppose someone observes (Y_1, Y_2, Y_3) and tells you that the value for Y_2 is above the mean of that variable μ_2. What, if anything, does this suggest about the values of Y_1 and Y_3 that were probably observed.

It suggests that y_1 is less than μ_1 since $\text{Cov}(Y_1, Y_2) < 0$, but tells you nothing about y_3 since $\text{Cov}(Y_2, Y_3) = 0$.

(d) (4 pts.) Give the smallest interval you can about the mean of Y_2 in which observations of Y_2 must lie with probability .99. Round to two decimal places. (Do *not* assume that Y_2 is normal.)

Y_2 has $\mu_2 = 2$ and $\sigma_2 = \sqrt{5} \approx 2.2361$

By Tchebychev's Theorem, $P(|Y_2 - \mu_2| < k\sigma) \geq 1 - \frac{1}{k^2} = .99 \Rightarrow \frac{1}{k^2} = \frac{1}{100} \Rightarrow k = 10$

Thus, $P(|Y_2 - 2| < 10(2.2361)) \geq .99$

The interval is $2 \pm 10(2.2361) \approx 2 \pm 22.36 = (-20.36, 24.36)$