Homework 5: Due 7/3/14

1. Let X and Y be continuous random variables with joint/marginal p.d.f.'s

$$f(x,y) = 2, \qquad 0 \le x \le y \le 1,$$
$$f_1(x) = 2(1-x), \qquad 0 \le x \le 1,$$
$$f_2(y) = 2y, \qquad 0 \le y \le 1.$$

Find the conditional p.d.f. $h(y|x)$ and the conditional probability $P\left(\frac{1}{2} \le Y \le \frac{3}{4} \mid X = \frac{1}{4}\right)$. What is the expected value of Y when $X = \frac{1}{4}$?

Solution: The conditional p.d.f $h(y|x) = f(x,y)/f_1(x)$ is immediately seen to be

$$h(y|x) = \frac{2}{2(1-x)} = \frac{1}{1-x}.$$

To find $P(\frac{1}{2} \le Y \le \frac{3}{4} \mid X = \frac{1}{4})$ we integrate the conditional p.d.f. $h(y|\frac{1}{4})$ on the interval $1/2 \le y \le 3/4$, and we obtain

$$P\left(\frac{1}{2} \le Y \le \frac{3}{4} \,\middle|\, X = \frac{1}{4}\right) = \int_{1/2}^{3/4} \frac{1}{1-\frac{1}{4}}\, dy = \frac{1}{4}\left(\frac{1}{\frac{3}{4}}\right) = \boxed{\frac{1}{3}}.$$

Since expectation is linear, we have $E(Y|X = \frac{1}{4}) = E(4/3) = \boxed{4/3}$.

□

2. Let X and Y be discrete random variables with joint p.m.f.

$$f(x,y) = \frac{x+y}{32}, \qquad x = 1, 2, \quad y = 1, 2, 3, 4.$$

Find the marginal p.m.f.'s of X and Y and the conditional p.m.f.'s $g(x|y)$ and $h(y|x)$. Find $P(1 \le Y \le 3 \mid X = 1)$ and $P(Y \le 2 \mid X = 2)$. Finally, find $E(Y \mid X = 1)$ and find $Var(Y \mid X = 1)$.

Solution: The marginal p.m.f.'s of X and Y are immediately seen to be

$$f_1(x) = \sum_{y=1,2,3,4} f(x,y) = \frac{4x+10}{32},$$
$$f_2(y) = \sum_{x=1,2} f(x,y) = \frac{3+2y}{32}.$$

The conditional p.m.f.'s are thus seen to be

$$h(y|x) = \frac{f(x,y)}{f_1(x)} = \frac{(x+y)/32}{(4x+10)/32} = \frac{x+y}{4x+10},$$
$$g(x|y) = \frac{f(x,y)}{f_2(y)} = \frac{(x+y)/32}{(3+2y)/32} = \frac{x+y}{3+2y}.$$

Also, we have

$$P(1 \leq Y \leq 3 \mid X = 1) = \sum_{y=1,2,3} h(y|1) = \frac{2}{14} + \frac{3}{14} + \frac{4}{14} = \boxed{\frac{9}{14}}.$$

Note, this can also be computed as $1 - h(4|1) = 1 - \frac{5}{14} = \frac{9}{14}$. Next, we compute

$$P(Y \leq 2 \mid X = 2) = \sum_{y=1,2} h(y|2) = \frac{2+1}{18} + \frac{2+2}{18} = \boxed{\frac{7}{18}}.$$

Finally, we have that

$$E(Y \mid X = 1) = \sum_{y=1,2,3,4} y \cdot h(y|1) = (1)\frac{2}{14} + (2)\frac{3}{14} + (3)\frac{4}{14} + (4)\frac{4}{14} = \boxed{\frac{18}{7}},$$

and

$$Var(Y \mid X = 1) = E(Y^2 \mid X = 1) - E(Y \mid X = 1) = \frac{57}{7} - \left(\frac{18}{7}\right)^2 = \boxed{1.503}.$$

□

3. Let W equal the weight of a box of oranges which is supposed to weight 1-kg. Suppose that $P(W < 1) = .05$ and $P(W > 1.05) = .1$. Call a box of oranges light, good, or heavy depending on if $W < 1$, $1 \leq W \leq 1.05$, or $W > 1.05$, respectively. In $n = 50$ independent observations of these boxes, let X equal the number of light boxes and Y the number of good boxes.

Find the joint p.m.f. of X and Y. How is Y distributed? Name the distribution and state the values of the parameters associated to this distribution. Given $X = 3$, how is Y distributed? Determine $E(Y \mid X = 3)$ and find the correlation coefficient ρ of X and Y.

Solution: The random variables are said to come from a *trinomial distribution* in this case since there are three exhaustive and mutually exclusive outcomes light, good, or heavy, having probabilities $p_1 = .05$, $p_2 = .85$, and $p_3 = .1$, respectively. It is easy to see that the trinomial p.m.f. in this case is

$$f(x, y) = \frac{50!}{x!y!(50 - x - y)!} p_1^x p_2^y p_3^{50-x-y}.$$

That is, $f(x, y)$ is exactly the joint p.m.f. of X and Y, the number of light boxes and good boxes, where $p_1 = .05$, $p_2 = .85$, and $p_3 = .1$ are the various probabilities of each of the three events: light, good, and heavy.

The random variable is binomially distributed, but the parameters of the distribution depend on the value of the random variable X. We have that the random variable Y is (conditionally) binomially distributed $\mathrm{b}(n - x, \frac{p_2}{1-p_1})$ since the marginal distributions of X, Y are $\mathrm{b}(n, p_1)$, $\mathrm{b}(n, p_2)$ and the conditional p.m.f. of Y is thus, with $n = 50$,

$$\begin{aligned} h(y|x) = f(x, y)\frac{1}{f_1(x)} &= \frac{n!}{x!y!(n - x - y)!} p_1^x p_2^y p_3^{n-x-y} \cdot \frac{x!(n - x)!}{n!} \frac{1}{p_1^x (1 - p_1)^{n-x}} \\ &= \frac{n!x!(n - x)!}{x!y!(n - x - y)!n!} \cdot \frac{p_1^x}{p_1^x} \cdot \frac{p_2^y}{(1 - p_1)^y} \cdot \frac{p_3^{n-x-y}}{(1 - p_1)^{n-x}(1 - p_1)^{-y}} \\ &= \frac{(n - x)!}{y!(n - x - y)!} \left(\frac{p_2}{1 - p_1}\right)^y \left(\frac{p_3}{1 - p_1}\right)^{n-x-y}. \end{aligned}$$

In this case, with the specified values of n, p_1, p_2, and p_3, we have for $X = 3$

$$h(y|3) = \frac{47!}{y!(47 - y)!} (.8947)^y (.1053)^{47-y},$$

so that Y is conditionally $\mathrm{b}(47, .8947)$ when $X = 3$. Since $\mu = np$ for binomial distribution, we have that $E(Y|X = 3) = (47)(.8947) = \boxed{42.05}$.

It is not hard to see that in fact $E(Y|x) = (n - x)\frac{p_2}{1-p_1}$ in general, and that a similar formula holds for $E(X|y)$. The correlation coefficient is now found using the fact that since each of the conditional expectations $E(Y|x) = (n-x)\frac{p_2}{1-p_1}$ and $E(X|y) = (n-y)\frac{p_1}{1-p_2}$ is *linear*, then the square of the correlation coefficient ρ^2 is equal to the product of the respective coefficients of x and y in the conditional expectations.

$$\rho^2 = \left(\frac{-p_2}{1-p_1}\right)\left(\frac{-p_1}{1-p_2}\right) = \frac{p_1 p_2}{(1-p_1)(1-p_2)},$$

from which it follows that

$$\rho = \boxed{-\sqrt{\frac{p_1 p_2}{(1-p_1)(1-p_2)}} = -.0819}.$$

The fact that the correlation coefficient is negative follows from the fact that, for example,

$$E(Y|x) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X),$$

and noting that the coefficient of x in $E(Y|x)$ is seen to be negative (and also $\sigma_Y, \sigma_X > 0$).

$\square$

4. Let X have the uniform distribution $U(0,2)$ and let the conditional distribution of Y, given that $X = x$, be $U(0,x)$. Find the joint p.d.f. $f(x,y)$ of X and Y, and be sure to state the domain of $f(x,y)$. Find $E(Y|x)$.

Solution: We have that

$$f(x,y) = \frac{y}{x}, \qquad 0 < x \leq 2,\ 0 \leq y \leq x.$$

Now,

$$E(Y|x) = \int_0^x y \cdot \frac{y}{x}\, dx = \left. \frac{1}{3x}\, y^3 \right|_0^x = \frac{x^2}{3}.$$

$\square$

5. The *support* of a random variable X is the set of x-values such that $f(x) \neq 0$. Given that X has p.d.f. $f(x) = x^2/3$, $-1 < x < 2$, what is the support of X^2? Find the p.m.f. of the random variable $Y = X^2$.

Solution: The p.d.f. $g(y)$ of $Y = X^2$ is obtained as follows. We note that the possible y-values that can be obtained are in the range $0 \leq y \leq 4$, so the support of $g(y)$ needs to be the interval $[0, 4]$. Now, on the interval $[1, 4]$, there is a one-to-one transformation represented by $x = \sqrt{y}$. We first find $G(y) = P(Y \leq y) = P(X^2 \leq y) = P(X \leq \sqrt{y})$ for $X = x$ in $[1, 2]$, corresponding to $Y = y$ in $[1, 4]$. We have

$$G(y) = \int_1^{\sqrt{y}} f(x)\, dx = \int_1^{\sqrt{y}} \frac{x^2}{3}\, dx,$$

and in particular $g(y) = G'(y) = \frac{(\sqrt{y})^2}{3} \cdot (\sqrt{y})'$, from the chain rule and the Fundamental Theorem of Calculus. Simplifying, we have $g(y) = \frac{\sqrt{y}}{6}$, $1 < y < 4$.

In order to find $g(y)$ on $0 < y < 1$, we need to work a little harder. For x in the interval $-1 < x < 1$ there is a two-to-one transformation given by $x = -\sqrt{y}$ for $-1 < x < 0$, and $x = \sqrt{y}$ for $0 < x < 1$. We then calculate $G(y)$ for $0 < y < 1$ (*i.e.*, $-1 < x < 1$) as before, but now using two integrals $G(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} < X < 0) + P(0 < X < \sqrt{y})$, so for y in the interval $0 < y < 1$ we have

$$G(y) = \int_{-\sqrt{y}}^{0} f(x)\, dx + \int_0^{\sqrt{y}} f(x)\, dx.$$

Again, from the chain rule and the Fundamental Theorem of Calculus we have

$$g(y) = G'(y) = -f(-\sqrt{y}) \cdot (-\sqrt{y})' + f(\sqrt{y}) \cdot (\sqrt{y})' = \frac{-y}{3} \cdot \frac{-1}{2\sqrt{y}} + \frac{y}{3} \cdot \frac{1}{2\sqrt{y}} = \frac{\sqrt{y}}{3}.$$

So,

$$\boxed{g(y) = \begin{cases} \sqrt{y}/3 & \text{if } 0 < y < 1, \\ \sqrt{y}/6 & \text{if } 1 < y < 4. \end{cases}}$$

There is no problem defining $g(0) = g(1) = 0$, or even just leaving the p.d.f. undefined at the points $y = 0$ and $y = 1$.

□

6. Let X_1, X_2 denote two independent random variables each with the $\chi^2(2)$ distribution. Find the joint p.d.f. of $Y_1 = X_1$ and $Y_2 = X_1 + X_2$. What is the support of Y_1, Y_2 (i.e., what is the domain of the joint p.d.f., where $f(y_1, y_2) \neq 0$)? Are Y_1 and Y_2 independent?

Solution: We have that X_1, X_2 have the same p.d.f.

$$h(x) = \frac{1}{2}\, e^{-x/2}, \qquad 0 \leq x < \infty,$$

corresponding to the $\chi^2(r)$ distribution with $r = 2$ degrees of freedom. By the way, this is the same as saying that X_1, X_2 follow exponential distributions with $\theta = 2$. Since X_1, X_2 are independent, the joint p.d.f. of X_1 and X_2 is

$$f(x_1, x_2) = h(x_1)h(x_2) = \frac{1}{4}\, e^{-(x_1+x_2)/2}.$$

The change of variables formula is $g(y_1, y_2) = |J| f(v_1(y_1, y_2), v_2(y_1, y_2))$ using the determinant of the Jacobian

$$J = \begin{vmatrix} \dfrac{\partial v_1(x_1, x_2)}{\partial y_1} & \dfrac{\partial v_1(x_1, x_2)}{\partial y_2} \\ \\ \dfrac{\partial v_2(x_1, x_2)}{\partial y_1} & \dfrac{\partial v_2(x_1, x_2)}{\partial y_2} \end{vmatrix}$$

where $v_i(y_1, y_2)$ is the inverse of u_i, $Y_i = u_i(X_1, X_2)$, $i = 1, 2$. In this case, $Y_1 = u_1(X_1, X_2) = X_1$, so $X_1 = v_1(Y_1, Y_2) = Y_1$, and $Y_2 = u_2(X_1, X_2) = X_1 + X_2$, so $X_2 = v_2(Y_1, Y_2) = Y_2 - Y_1$. Hence,

$$|J| = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1,$$

so the joint p.d.f. of Y_1 and Y_2 is

$$g(y_1, y_2) = |1| f(y_1, y_2 - y_1) = \frac{1}{4} e^{-y_2/2}, \quad 0 \leq y_1 \leq y_2 < \infty.$$

To determine if Y_1 and Y_2 are independent we compute the marginal p.d.f.'s. We have,

$$g_1(y_1) = \int_0^{y_2} \frac{1}{4} e^{-y_2/2}\, dy_1 = \frac{y_2}{4} e^{-y_2/2},$$

and

$$g_2(y_2) = \int_{y_1}^{\infty} \frac{1}{4} e^{-y_2/2}\, dy_2 = \frac{-1}{2} e^{-y_2/2}\Big|_{y_1}^{\infty} = \frac{1}{2} e^{-y_1/2}.$$

Since, $g(y_1, y_2) \neq g_1(y_1) g_2(y_2)$, Y_1 and Y_2 are dependent.

□