STAT 305 D Homework 7

Due March 28, 2013 at 12:40 PM in class

1. Vardeman and Jobe chapter 5 section 4 problem 1 (page 300)

   Explain in qualitative terms what it means for two random variables X and Y to be independent. What advantage is there when X and Y can be described as independent?

   If X and Y are independent, then they are uncorrelated—they do not influence each other in any way. Put differently, if the probability distribution of X and Y are both known completely, observing the actual value of X does not in any way change the probability distribution of the yet-to-be-observed Y, and vice-versa.

   One practical advantage of X and Y being independent is that the variance of a linear combination of the two can be easily computed using equation (5-59). Another advantage is that it is easy to describe the joint probability distribution of X and Y—it is just the product of the marginal distributions. In general, independence allows the probability of two events happening together to be computed as the product of the probabilities of each event.

2. Vardeman and Jobe chapter 5 section 4 problem 2 (page 300)

   Quality audit records are kept on numbers of major and minor failures of circuit packs during burn-in of large electronic switching devices. They indicate that for a device of this type, the random variables

   $$X = \text{the number of major failures}$$

   and

   $$Y = \text{the number of minor failures}$$

   can be described at least approximately by the accompanying joint distribution.

$y \backslash x$	0	1	2
0	.15	.05	.01
1	.10	.08	.01
2	.10	.14	.02
3	.10	.08	.03
4	.05	.05	.03

   (a) Find the marginal probability functions for both X and Y—$f_X(x)$ and $f_Y(y)$.
   (b) Are X and Y independent? Explain.
   (c) Find the mean and variance of X—EX and $\text{Var}\,X$.
   (d) Find the mean and variance of Y—EY and $\text{Var}\,Y$.
   (e) Find the conditional probability function for Y, given that $X = 0$—i.e., that there are no major circuit pack failures. (That is, find $f_{Y|X}(y \mid 0)$.)

(a) Compute $f_X(x)$ by summing down the columns and $f_Y(y)$ by summing across the rows.

x	0	1	2
$f_X(x)$	.5	.4	.1

y	0	1	2	3	4
$f_Y(y)$	.21	.19	.26	.21	.13

(b) No, since $f(x,y) \neq f_X(x)f_Y(y)$. For example, $f(0,0) = .15$ and $f_X(0)f_Y(0) = .5(.21) = .105$.

(c) Use equations (5-1) and (5-2), and the marginal distribution of X from part (a).

$$EX = (0)(.5) + (1)(.4) + (2)(.1) = .6$$

and

$$\text{Var}X = (0)^2(.5) + (1)^2(.4) + (2)^2(.1) - (.6)^2 = .44.$$

(d) Use equations (5-1) and (5-2), and the marginal distribution of Y from part (a).

$$EY = (0)(.21) + (1)(.19) + (2)(.26) + (3)(.21) + (4)(.13) = 1.86$$

and

$$\text{Var}Y = (0)^2(.21) + (1)^2(.19) + (2)^2(.26) + (3)^2(.21) + (4)^2(.13) - (1.86)^2 = 1.7404.$$

(e) Use equation (5-42). For example

$$f_{Y|X}(0|0) = \frac{f(0,0)}{f_X(0)} = \frac{.15}{.5} = .3$$

The rest are given in the table below.

y	0	1	2	3	4
$f_{Y\|X}(y\|0)$	.3	.2	.2	.2	.1

Using equation (5-1),

$$E(Y|X=0) = (0)(.3) + (1)(.2) + (2)(.2) + (3)(.2) + (4)(.1) = 1.6.$$

3.

a. $E(X+2Y)$, where $X \sim$ Binomial$(n = 3, p_1 = 0.7)$ and is independent of $Y \sim$Geometric $(p_2 = 0.3)$

$$E(X) = np_1 = 3 \cdot 0.7 = 2.1 \quad \text{since } X \text{ is Binomial}(3,.7)$$
$$E(Y) = 1/p_2 = 1/0.3 = 3.33 \quad \text{since } Y \text{ is Geometric}(0.3)$$

Hence:

$$\begin{aligned} E(X+2Y) &= E(X) + 2E(Y) \\ &= 2.1 + 2 \cdot 3.33 \\ &= 8.76 \end{aligned}$$

b. $Var(X+2Y)$, where $X \sim$ Binomial$(n=3, p_1=.7)$ and is independent of $Y \sim$Geometric $(p_2 = 0.3)$

$$Var(X) = np_1(1-p_1) = 3 \cdot 0.7 \cdot (1-0.7) = 0.63 \quad \text{since } X \text{ is Binomial(3,.7)}$$
$$E(Y) = (1-p_2)/p_2^2 = (1-0.3)/0.3^2 = 7.78 \quad \text{since } Y \text{ is Geometric(0.3)}$$

Hence:

$$\begin{aligned} Var(X+2Y) &= Var(X) + 2^2 E(Y) \\ &= 0.63 + 4 \cdot 7.78 \\ &= 31.75 \end{aligned}$$

c. $E(3X+5Y-8Z)$, where X, Y, and Z are independent, $E(X) = 4$, $E(Y) = 3$, $E(Z) = 10$

$$\begin{aligned} E(3X+5Y-8Z) &= 3E(X) + 5E(Y) - 8E(Z) \\ &= 3 \cdot 4 + 5 \cdot 3 - 8 \cdot 10 \\ &= -53 \end{aligned}$$

d. $Var(3X+5Y-8Z)$, where X, Y, and Z are independent, $Var(X) = 9$, $Var(Y) = 16$, $Var(Z) = 2$

$$\begin{aligned} Var(3X+5Y-8Z) &= 3^2Var(X) + 5^2Var(Y) + 8^2Var(Z) \\ &= 9 \cdot 9 + 25 \cdot 16 + 64 \cdot 2 \\ &= 609 \end{aligned}$$

e. $E(\overline{X})$, where $X_1, X_2, \ldots, X_n$ are iid with mean 0 and variance 2.

$$\begin{aligned} E(\overline{X}) &= E(\frac{1}{n}X_1 + \frac{1}{n}X_2 + \cdots + \frac{1}{n}X_n) \\ &= \frac{1}{n}E(X_1) + \frac{1}{n}E(X_2) + \cdots + \frac{1}{n}E(X_n) \\ &= \frac{1}{n}0 + \frac{1}{n}0 + \cdots + \frac{1}{n}0 \\ &= 0 \end{aligned}$$

f. $P(\overline{X} \leq 6)$, where $X_1, X_2, \ldots, X_{40} \sim$ iid $N(5, 2)$.

Let n = 40.

$$\begin{aligned}
E(\overline{X}) &= E(\frac{1}{n}X_1 + \frac{1}{n}X_2 + \cdots + \frac{1}{n}X_n) \\
&= \frac{1}{n}E(X_1) + \frac{1}{n}E(X_2) + \cdots + \frac{1}{n}E(X_n) \\
&= \frac{1}{n}5 + \frac{1}{n}5 + \cdots + \frac{1}{n}5 \\
&= n\frac{1}{n}5 \\
&= 5
\end{aligned}$$

$$\begin{aligned}
Var(\overline{X}) &= Var(\frac{1}{n}X_1 + \frac{1}{n}X_2 + \cdots + \frac{1}{n}X_n) \\
&= \frac{1}{n^2}Var(X_1) + \frac{1}{n^2}Var(X_2) + \cdots + \frac{1}{n^2}VaX_n) \\
&= \frac{1}{n^2}2 + \frac{1}{n^2}2 + \cdots + \frac{1}{n^2}2 \\
&= n\frac{1}{n^2}2 \\
&= 2/n \\
&= 2/40 \\
&= 1/20
\end{aligned}$$

Hence, $\overline{X} \sim N(5, 1/20)$

$$\begin{aligned}
P(\overline{X} \leq 6) &= P(\frac{\overline{X} - 5}{\sqrt{1/20}} \leq \frac{6 - 5}{\sqrt{1/20}}) \\
&= P(Z \leq 4.472) \\
&= 0.9999999
\end{aligned}$$

g. $P(\overline{X} > 11)$, where $X_1, X_2, \ldots, X_{20} \sim$ iid $N(3, 150)$

$$\begin{aligned}
E(\overline{X}) &= E(X_1) = 3 \\
Var(\overline{X}) &= Var(X_1)/20 = 150/20 = 15/2
\end{aligned}$$

Hence, $\overline{X} \sim N(3, 15/2)$

$$\begin{aligned} P(\overline{X} > 11) &= P(\frac{\overline{X} - 3}{\sqrt{15/2}} > \frac{11 - 3}{\sqrt{15/2}}) \\ &= P(Z > 2.92) \\ &= 1 - P(Z \leq 2.92) \\ &= 1 - \Phi(2.92) \\ &= 0.00175 \end{aligned}$$

h. $P(|\overline{X} - 3| > 1)$, where $X_1, X_2, \ldots, X_{10} \sim$ iid $N(3, 18)$

$$\begin{aligned} E(\overline{X}) &= E(X_1) = 3 \\ Var(\overline{X}) &= Var(X_1)/10 = 18/10 = 1.8 \end{aligned}$$

Hence, $\overline{X} \sim N(3, 1.8)$

$$\begin{aligned} P(|\overline{X} - 3| > 1) &= P(\overline{X} - 3 > 1 \text{ or } P(\overline{X} - 3 < -1) \\ &= P(\overline{X} - 3 > 1) + P(\overline{X} - 3 < -1) \\ &= P(\frac{\overline{X} - 3}{\sqrt{1.8}} > \frac{1}{\sqrt{1.8}}) + P(\frac{\overline{X} - 3}{\sqrt{1.8}} < \frac{-1}{\sqrt{1.8}}) \\ &= P(Z > 0.745) + P(Z < -0.745) \\ &= P(Z < -0.745) + P(Z < 0.745) \quad \text{by symmetry} \\ &= 2P(Z < -0.745) \\ &= 2P(Z \leq -0.745) \\ &= 2\Phi(-0.745) \\ &= 0.456 \end{aligned}$$

4.

a. P($\overline{X} \leq 23$), where $X_1, X_2, \ldots X_{40}$ ~iid Binomial with mean 20 and variance 100.

$$\begin{aligned} E(\overline{X}) &= E(X_1) = 20 \\ Var(\overline{X}) &= Var(X_1)/40 = 100/40 = 5/2 \end{aligned}$$

And since we have at least 25 X_i's, $\overline{X}$ is approximately N(20, 5/2) by the Central Limit Theorem. Hence, if $Z \sim N(0,1)$, then:

$$\begin{aligned} P(\overline{X} \leq 23) &= P\left(\frac{\overline{X}-20}{\sqrt{5/2}} \leq \frac{23-20}{\sqrt{5/2}}\right) \\ &\approx P(Z \leq 1.90) \\ &= \Phi(1.90) \\ &= 0.9712 \end{aligned}$$

b. P($\overline{X} > 1$), where $X_1, X_2, \ldots, X_{86} \sim$ iid Scaled Exponential with mean 0 and variance 28.

$$\begin{aligned} E(\overline{X}) &= E(X_1) = 0 \\ Var(\overline{X}) &= Var(X_1)/86 = 28/86 = 14/43 \end{aligned}$$

And since we have at least 25 X_i's, $\overline{X}$ is approximately N(0, 14/43) by the Central Limit Theorem. Hence, if $Z \sim N(0,1)$, then:

$$\begin{aligned} P(\overline{X} > 1) &= P\left(\frac{\overline{X}-0}{\sqrt{14/43}} > \frac{1-0}{\sqrt{14/43}}\right) \\ &\approx P(Z > 1.752) \\ &= 1 - P(Z \leq 1.752) \\ &= 1 - 0.960 \\ &= 0.04 \end{aligned}$$

c. $P(|\overline{X}+8| > 12$, where $X_1, X_2, \ldots, X_{70} \sim$ iid Weibull with mean 3 and variance 37.

$$\begin{aligned} E(\overline{X}) &= E(X_1) = 3 \\ Var(\overline{X}) &= Var(X_1)/70 = 37/70 \end{aligned}$$

And since we have at least 25 X_i's, $\overline{X}$ is approximately N(3, 37/70) by the Central Limit Theorem. Hence, if $Z \sim N(0,1)$, then:

$$
\begin{aligned}
P(|\overline{X}+8| > 12) &= P(\overline{X}+8 > 12 \text{ or } \overline{X}+8 < -12) \\
&= P(\overline{X}+8 > 12) + P(\overline{X}+8 < -12) \\
&= P(\overline{X}-3 > 1) + P(\overline{X}-3 < -23) \\
&= P\left(\frac{\overline{X}-3}{\sqrt{37/70}} > \frac{1}{\sqrt{37/70}}\right) + P\left(\frac{\overline{X}-3}{\sqrt{37/70}} < \frac{-23}{\sqrt{37/70}}\right) \\
&\approx P(Z > 1.375) + P(Z < -31.635) \\
&= 1 - P(Z \leq 1.375) + P(Z \leq -31.635) \\
&= 1 - 0.915 + 0 \\
&= 0.0846
\end{aligned}
$$

5.

Students are going to measure Young's Modulus for copper by measuring the elongation of a piece of copper wire under a tensile force. For a cylindrical wire of diameter D subjected to a tensile force F, if the initial length (length before applying the force) is L_0 and final length is L_1, Young's Modulus for the material in question is

$$Y = \frac{4FL_0}{\pi D^2(L_1 - L_0)}$$

The test and measuring equipment used in a particular lab are characterized by the standard deviations

$$\sigma_F \approx 10 \text{ lb} \qquad \sigma_D \approx .001 \text{ in.}$$
$$\sigma_{L_0} = \sigma_{L_1} = .01 \text{ in.}$$

and in the setup employed, $F \approx 300$ lb, $D \approx .050$ in., $L_0 \approx 10.00$ in., and $L_1 \approx 10.10$ in.

(a) Treating the measured force, diameter, and lengths as independent variables with the preceding means and standard deviations, find an approximate standard deviation to attach to an experimentally derived value of Y. (Partial derivatives of Y at the nominal values of F, D, L_0, and L_1 are approximately $\frac{\partial Y}{\partial F} \approx 5.09 \times 10^4$, $\frac{\partial Y}{\partial D} \approx -6.11 \times 10^8$, $\frac{\partial Y}{\partial L_0} \approx 1.54 \times 10^8$, and $\frac{\partial Y}{\partial L_1} \approx -1.53 \times 10^8$ in the appropriate units.)

(b) Uncertainty in which of the variables is the biggest contributor to uncertainty in Y?

(a) Using equation (5-59),

$$\begin{aligned}\mathrm{Var}Y &\approx (5.09 \times 10^4)^2(10)^2 + (-6.11 \times 10^8)^2(.001)^2 + (1.54 \times 10^8)^2(.01)^2 \\ &\quad +(-1.53 \times 10^8)^2(.01)^2 \\ &= 2.59081 \times 10^{11} + 3.733211 \times 10^{11} + 2.3716 \times 10^{12} + 2.3409 \times 10^{12} \\ &= 5.3449 \times 10^{12},\end{aligned}$$

so the approximate standard deviation of Y is

$$\sqrt{5.3449 \times 10^{12}} = 2,311,904.41.$$

(b) L_0, followed closely by L_1. These two variables' terms (variance $\times$ squared partial derivative) contribute the most to the propagation of error formula.

6. Weekly feedback. You get full credit as long as you write something.

   a. Is there any aspect of the subject matter that you currently struggle with? If so, what specifically do you find difficult or confusing? The more detailed you are, the better I can help you.

      You got full credit as long as you wrote something.

   b. Do you have any questions or concerns about the material, class logistics, or anything else? If so, fire away. You got full credit as long as you wrote something.