<u>PS 699</u>: Problem Set 2 – Probability

1) The variance of a Bernoulli random variable is $\pi(1-\pi)$ where π is the probability of a successful Bernoulli trial (*i.e.*, probability that X=1). π is also the mean of a Bernoulli random variable; i.e., $E(X)=\pi$.

a) first, using the formula for the p.f. of a Bernoulli random variable, and the formula for a variance, prove that $V(X) = \pi(1-\pi)$.

b) now, at what π is the variance of X highest? Even if you know the answer already, prove it by showing your maximizing V(X) with respect to π, *i.e.* $\text{Max}_\pi\ \pi(1-\pi)$.

2) Dice have six sides, all equally likely to "come up" (*i.e.*, appear on the top face when the dice are rolled). If you roll two dice and call their sum X,

a) draw (using excel, or by hand--carefully and as clearly as possible if the latter) the p.f. of X, *i.e.* draw f(X).

b) from what distribution would you say the random variable X is drawn?

c) what is the mean of X?

d) what is the variance of X?

e) based on your answer to these questions, give a general formula for the mean and variance of this type of probability distribution. (General here means that if I met another random variable like this one, and if I know the parameters of its probability function, I can use the formulae you give to find mean and variance.)

3) You may use the statistical tables from Greene or the probability functions in Stata or Excel or any other program as you wish to solve the following:

a) X is a normal random variable with mean 3 and variance 16. What is the probability that any draw, x, on this variable is greater than 7? Less than 1? Between 7 and 1? What is the 95% confidence interval for draws x? (That is, what symmetric range around the mean contains 95% of the probability. Drawing pictures of the distribution and shading the area you want may help if you're having trouble with this problem.)

b) Redo 3a only now X is a t with 10 degrees of freedom (n.b., the statements of mean and variance of X from 3a no longer apply).

c) Redo 3a only now X is a Chi-Squared with 3 degrees of freedom.

(4) *Jensen's Inequality*: Suppose X is a random variable equal to 1 with probability 1/3, 2 with probability 1/3, and 3 with probability 1/3. What is

(a) $E(X)$? (b) $[E(X)]^2$? (c) $E(X^2)$? (d) $ln[E(X)]$? (e) $E[ln(X)]$?

(f) Compare (b) to (c) (*i.e.*, is b larger or smaller than a?) and (d) to (e). What it is about the functions $f(x)=x^2$ and $f(x)=ln(x)$ that makes these comparisons differ?

(5) The following table gives probabilities for a discrete bivariate distribution: f(x,y).

X \ Y	y = 1	y = 2	y = 3	y = 4
x = 1	.1	0	.1	.15
x = 2	.1	.1	0	0
x = 3	0	0	0	.1
x = 4	.1	.25	0	0

What is the probability...

a) x is 4 given that y is 2? b) y is 1 given that x is 4?

What is the conditional probability function...

c) $f(x|y=1)$? d) $f(y|x=1)$? e) What outcome for y makes x=1 most likely?

Joint and marginal probabilities:

f) What is the probability (x<3 and y<3)? g) What is Pr(x=4)? h) What is Pr(y=2)?

(6) Suppose we have two independent, standard-normal-distribution random-variables Y and X.

a) do I need to tell you what kind of independence Y and X exhibit before you can do the rest of this problem, (see questions *b) - e)* below) or will any of the three in the class notes suffice? Is this generally the case?

What is the probability that...

b) x>1 and y>1? c) x>1 or y>1?

What is the probability that...

d) x>1 given that y>1? e) y>1 given that x>1?

(7) You estimate the following regression model: $\mathbf{y} = b_0 + b_1\mathbf{x} + b_2\mathbf{z} + b_3\mathbf{x}\cdot\mathbf{z} + \mathbf{e}$

[· here is element-by-element multiplication (a.k.a., *Hadamard product*).] Stata reports b_0=3.5, b_1=2, b_2=1 b_3=-3, assuming that **x** and **z** both vary from 0-1, graph the following in a spreadsheet (label appropriately and print):

a) $\partial y/\partial x$ on the "y" axis against z on the "x" axis;

b) $\partial y/\partial z$ on the "y" axis against x on the "x" axis.

c) In terms of the regression model, explain (in words) what these two lines indicate.

d) Generically (i.e., give the formula), what is the variance of the $\partial y/\partial x$ you just estimated? (Important here to note that, in the context of (classical) statistical estimation, the **b**'s here are the random variables and the **x**'s are constant ("across repeated samples"). Remembering that, just use the formula for V(a+bX+cY) given in the lecture notes; again, in that presentation, the random variables were X and Y, and a, b, and c were constants. Take care to align notation correctly with the substance. This, too, is useful practice.)

e) Generically, what is the variance of $\partial y/\partial z$? I.e., give the formula for V(E($\partial y/\partial z$)), that being another way to denote the question (the E(·) here refers to *expectation*, i.e., your *estimates*).

Suppose, finally, that Stata tells you V[(b_1,b_2,b_3)'] was given by the matrix:

$$V\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \mathbf{\Sigma_b} = \begin{bmatrix} +0.256 & -0.027 & -0.309 \\ -0.027 & +1.296 & -1.624 \\ -0.309 & -1.624 & +2.652 \end{bmatrix}$$

f) Graph $\partial y/\partial x \pm 1$ standard-deviation (technically, *standard-error*) (using spreadsheet formula may help...);

g) Graph $\partial y/\partial z \pm 1$ standard-error.

h) Try to explain why $\partial y/\partial x$±1s.e. and $\partial y/\partial z$±1s.e. are shaped like they are.

Spreadsheet help: create a column of numbers counting by, say, 0.05, from 0 to 1 (which I told you is the range of x and z). Use the formulae you found in questions *a)* and *b)*, entered as spreadsheet formula referencing each of these columns of x and z to create two more columns, $\partial y/\partial x$ and $\partial y/\partial z$. (Spreadsheets formulae have the form, e.g., =4.7+3*(A3) where (A3) is the cell where the value that you want to multiply by 3 is located.) Write the appropriate equation in this format in the cell of a next column. Now copy that formula down to the last row in which you have the x and z counters. (You can look at how the spreadsheet automatically updated the formulae as it copied down just by selecting any one of the cells in this new column.) Use the formulae for V(E($\partial y/\partial x$)) and V(E($\partial y/\partial z$)) from question *e)* to create columns of data equal $\partial y/\partial x$±1s.e. and $\partial y/\partial z$±1s.e. Have the spreadsheet plot a graph with a counter column as the x-axis, the $\partial y/\partial x$ (or $\partial y/\partial z$) as the first y-variable, and the $\partial y/\partial x$±1s.e. (or...) as the second and third y-variable. Format the lines and label the graph appealingly. Can you explain what you just graphed?