

MIT OpenCourseWare  
<http://ocw.mit.edu>

14.30 Introduction to Statistical Methods in Economics  
Spring 2009

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

### 14.30 Exam I

Spring 2009

**Instructions:** This exam is closed-book and closed-notes. You may use a calculator. Please read through the exam first in order to ask clarifying questions and to allocate your time appropriately. In order to receive partial credit in the case of computational errors, please show all work. You have approximately 85 minutes to complete the exam. Good luck!

**1.(20 points) Short Questions** True, false, or uncertain? Always give a brief explanation or a simple counter-example.

- (a) If  $X$  and  $Y$  are independent, then the support of the joint distribution  $f_{XY}(x, y)$  of the random variables  $X$  and  $Y$  is rectangular.
- (b) A continuous p.d.f. can never take a value greater than 1.
- (c) The following function is a c.d.f. of a random variable:

$$G(x) = \begin{cases} 0 & \text{if } x < 1 \\ 0.2 & \text{if } 1 \leq x < 2 \\ 0.2x & \text{if } 2 \leq x < 4 \\ 1 & \text{if } x \geq 4 \end{cases}$$

- (d) The function in part (c) is a c.d.f. of a discrete random variable.

**2. (15 points)** Two airlines, Sun Express, and Commuter Jet, are serving four major US airports. We are interested in the reliability of each of these airlines, and we have the following data on the probability of a significant delay (20 minutes or more) for departing flights from each of these cities.

	Sun Express		Commuter Jet	
	% delayed	# flights/day	% delayed	# flights/day
Atlanta	12%	5	11%	10
Chicago	30%	3	25%	12
Phoenix	5%	8	5%	2
Seattle	25%	4	20%	6

- (a) Given these numbers, and all else equal, which of the two airlines appears to have a better record?

A website reports the overall on-time rate for each airline which is defined as one minus the probability of a significant delay for a randomly chosen flight by that airline, where each flight departing from any of these four cities may be chosen with equal probability.

- (b) Use the law of total probability to calculate the on-time rate for each of the two airlines. Based on your result, which airline would you prefer if you made your choice based only on the information from the website?
- (c) How do you reconcile your result from (b) with your answer in (a)? Discuss the relationship in terms of conditional and unconditional probabilities, and feel free to draw parallels with examples discussed in class.

**3. (15 points)**

(a) Suppose

$$f_{XY}(x, y) = \begin{cases} cx^2y^2 & \text{if } 0 \leq x \leq 3 \text{ and } 0 \leq y \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

What does  $c$  have to be so that  $f_{XY}(x, y)$  is a proper joint p.d.f. of two random variables  $X$  and  $Y$ ? Given the correct value of  $c$ , are  $X$  and  $Y$  independent?

(b) Now let's modify the problem a little bit: suppose

$$g_{XY}(x, y) = \begin{cases} bx^2y^2 & \text{if } 0 \leq x \leq y \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

Draw the support of  $g_{XY}(x, y)$  in the  $(x, y)$  plane (i.e. the area in which the function is strictly positive). What does  $b$  now have to be for  $g_{XY}(x, y)$  to be a p.d.f.? Are  $X$  and  $Y$  still independent? Why or why not?

**4. (30 points)**

You are a miner during the Alaska Gold Rush of the 1890s, and you are prospecting a tract of land along a small creek in the forests and want to determine whether you are likely to find gold there. It is costly to rush back to town to register a formal claim on the lot which would allow you to exploit it, so you want to be at least 60% certain that there is gold in the ground. Let  $G$  be the event that there is gold on the specific tract you are surveying, and suppose that without additional information, the probability of finding gold is  $P(G) = 0.1$ .

After years out in the wild, you have learned that there are some important signs to look for: if there is gold, the rocks below the ground are darker than usual, the bushes have shorter roots, the cranberries you find in the place have a sweeter taste, and the first moose you'll see will wag its tail to the right. If there is no gold, you may still observe any of these indicators with some probability, to be specific:

B	$P(B G)$	$P(B G^C)$
Moose wags tail to the right	1.0	0.7
Dark rocks	1.0	0.4
Short roots	1.0	0.5
Sweet cranberries	1.0	0.3

Regardless of whether there is gold or not, the pieces of evidence are independent of each other. Clearly, if you observe that either of the above is not true, the tract contains no gold for sure, and you should move on.

However, there is fierce competition: other diggers are also surveying the same area, and the tract will be awarded to the first person to show up at the land registry and pay the official fee. Therefore, you want to base your conclusions on as few pieces of evidence as possible and rush to town as soon as you are certain enough that there is gold, i.e. as soon as  $P(G|B) \geq 0.6$ .

- State Bayes' Law. Derive a formula for the posterior  $P(G|B)$  that depends only on  $P(G)$  and the ratio  $L = \frac{P(B|G)}{P(B|G^C)}$ . Does  $P(G|B)$  increase or decrease in  $L$ ?
- Compute  $L$  for each of the different pieces of evidence listed in the table. Given independence, how do you calculate the ratio  $L$  for combinations of pieces of evidence, e.g.  $B =$  "dark rocks and short roots"?
- Given the values of  $L$ , in which order should you check for the different indicators? If there is gold in the ground, what is the minimum number of pieces of evidence that you have to check before  $P(G|B) \geq 0.6$ ?

Now fast forward to the year 2009, and suppose that you are the chief prospector of a multinational mining company looking for gold in northern Canada. You can drill the ground at  $N$  different spots and have the samples analyzed in a lab. For a given sample, a chemical test gives a positive result with

probability  $p$  if there is gold, and if there is no gold, the probability of a positive result is  $1 - p$ . The test is independent across the  $N$  samples. Define the event  $B =$  "  $x$  out of  $N$  samples produced a positive result."

- (d) Depending on  $N$  and  $x$ , what is the probability  $P(B|G)$ ? What is the ratio  $L(N, x) := \frac{P(B|G)}{P(B|G^c)}$ ?
- (e) Suppose  $p = 0.8$ , and we obtained a positive test result for 4 out of 12 samples. What is  $L(12, 4)$ ? Given your result from (a), what is the posterior probability  $P(G|B)$ ?
- (f) Suppose instead that we obtained a positive test result for  $x$  samples, and a negative result for  $x + 4$  samples (so the total number of samples is  $2x + 4$ ). How does your answer differ from (e)?

**5. (10 points)** Is the following statement true, false, or uncertain: "If  $P(A|B) > P(A)$  and  $P(A|C) > P(A)$ , then  $P(A|BC) > P(A)$ ?" Construct a simple *numerical* counterexample with a state space  $S = \{s_1, s_2, s_3, s_4\}$ , specifying the probability for each of the 4 outcomes, and defining events  $A$ ,  $B$ , and  $C$ .