

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 Introduction to Statistical Methods in Economics

Lecture Notes 8

Konrad Menzel

March 3, 2009

1 Conditional p.d.f.s

Definition 1 *The conditional p.d.f. of Y given X is*

$$f_{Y|X}(y|x) = \frac{f_{XY}(x, y)}{f_X(x)}$$

Note that if X and Y are discrete,

$$f_{Y|X}(y|x) = \frac{P(Y = y|X = x)}{P(X = x)}$$

which just corresponds to the conditional probability of the event corresponding to $X = x$ given $Y = y$ as defined two weeks ago.

Note that

- for a particular value of the conditioning variable, the conditional p.d.f. has all the properties of a usual p.d.f. (i.e. positive, integrates to 1)
- the definition generalizes to any number of random variables on either side

Example 1 *Let's go back to the data on extra-marital affairs, and look at the variables we are actually most interested in: the number of affairs during the last year, Z , and self-reported "quality" of the marriage, X . The joint p.d.f. is given by* Since three quarters of respondents reported not having had

		Z			f_X
		0	1	2	
X	1	17.80%	4.49%	6.82%	29.12%
	2	24.29%	3.83%	4.16%	32.28%
	3	32.95%	3.33%	2.33%	38.60%
	f_Z	75.04%	11.65%	13.31%	100.00%

an affair, it might be more instructive to look at the p.d.f. of the number of affairs Z conditional on the rating of marriage quality. Conditional on the low rating, $X = 1$, we have

$$f_{Z|X}(0|1) = \frac{f_{XZ}(1, 0)}{f_X(1)} = \frac{17.80\%}{29.12\%} = 61.13\%$$

		Z		
		0	1	2
X	$f_{Z X}$			
	1	61.13%	15.42%	23.42%
	2	75.25%	11.86%	12.88%
	3	85.36%	8.63%	6.04%

Putting the conditional c.d.f.s for the values of $X = 1, 2, 3$ together in a table, we get

Why is this exercise interesting? - while in the table with the joint p.d.f., the overall picture was not very clear, we can see that for lower values of marriage quality X , the conditional p.d.f. puts higher probability mass on higher numbers of affairs.

Does this mean that dissatisfaction with marriage causes extra-marital affairs? Certainly not: we could just do the reverse exercise, and look at the conditional p.d.f. of reported satisfaction with marriage, X , given the number of affairs, Z . E.g.

$$f_{X|Z}(1,0) = \frac{f_{XZ}(1,0)}{f_Z(0)} = \frac{17.80\%}{75.04\%} = 23.72\%$$

or, summarizing the conditional p.d.f.s in a table:

We see that the conditional p.d.f. of X given a larger number of affairs, Z , puts more probability on lower

		Z		
		0	1	2
X	$f_{X Z}$			
	1	23.72%	38.54%	51.24%
	2	32.37%	32.88%	31.25%
	3	43.91%	28.58%	17.51%

satisfaction with the marriage. So we could as well read the numbers as extra-marital affairs ruining the relationship. This is often referred to as "reverse causality": even though we may believe that A causes B , B may at the same time cause A .

Therefore, even though the conditional distributions shift in a way which is compatible with either story, we can't interpret the relationship as "causal" in either direction, because both stories are equally plausible, and presumably in reality, there is some truth to either of them.

2 Review

I don't expect you to memorize any of the examples we did in class, however, especially for "text" problems they can be extremely helpful as "models" for particular situations/problems. Often you can find a solution strategy to a given question by seeing analogies to examples we discussed in the lecture.

1. Probability

Sample Space, Set Theory and Basic Operations

won't discuss this

Definition of probability

(P1) $P(A) \geq 0$ for all $A \subset S$

(P2) $P(S) = 1$

(P3) if A_1, A_2, \dots is a sequence of *disjoint* sets,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Special Case: Simple Probabilities

- S finite
- $P(A) = \frac{n(A)}{n(S)}$ where $n(B)$ denotes the number of outcomes in set B

Properties of Probability Functions

- $P(A^C) = 1 - P(A)$
- $P(\emptyset) = 0$
- if $A \subset B$, then $P(A) \leq P(B)$
- $0 \leq P(A) \leq 1$ for any event $A \subset S$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Calculating Probabilities

Try to attack problems in this order

- define sample space and the event of interest in terms of outcomes
- for simple probabilities, make sure that you defined the sample space in a way that makes each outcome equally likely
- if you are stuck, start writing out outcomes in sample space explicitly

Counting Rules

- basic set-up: have set X_1, \dots, X_N of N objects
- multiplication rule: have to be able to factor an experiment into k parts such that the *number of outcomes* m_i in each of them does not depend on the outcomes in the other parts. Sometimes tricky (e.g. chess example).
- several different ways of drawing k objects from the set (should remember those for the exam):
 1. k draws with replacement, order matters: N^k possibilities
 2. k draws without replacement, order matters (special case: permutation, $k = N$): $\frac{N!}{(N-k)!}$ possibilities

- 3. k draws without replacement, order doesn't matter (combination): $\binom{N}{k} = \frac{N!}{(N-k)!k!}$ possibilities (e.g. in binomial distribution, count the number of all different sequences of "successes" which give the same overall *number* of successes).
- *partitions*: number of ways of allocating N objects across k groups, identities of objects don't matter (e.g. number of different allocations a five identical blue balls to four urns): in general $\binom{N+k-1}{k-1}$ possibilities, will discuss this below
- we saw that in one way or another, all these counting rules derive from the multiplication rule, where sometimes we had to divide by the number of different possibilities of obtaining the same event (e.g. different orders of drawing the same combination).

Independence, Conditional Probability, Bayes' Theorem

- A and B are independent if $P(AB) = P(A)P(B)$
- conditional probability $P(A|B) = \frac{P(AB)}{P(B)}$ if $P(B) > 0$
- $P(A|B) = P(A)$ if, and only if, A and B are independent
- law of total probability: if B_1, \dots, B_n partition of S ,

$$P(A) = P(A|B_1)P(B_1) + \dots + P(A|B_n)P(B_n)$$

The law of total probability links conditional to marginal probabilities, i.e. how to relate $P(A)$ to $P(A|B_1), \dots, P(A|B_n)$. Classical application: aggregating over subpopulations/subcases, e.g. death rates over different types of bypass surgery.

- Bayes' Theorem (simple formulation): if $P(B) > 0$, then

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)}$$

Bayes' Theorem tells us how to switch the order of conditioning, i.e. how to go from $P(B|A)$ to $P(A|B)$. Classical application: update beliefs about event A given data B , e.g. medical test example.

- "base rates" matter a lot

For the exam, you should know these relationships by heart.

2. Random Variables and Distribution Functions

- random variables give numerical characterization of random events.
- random variable X is a function from the sample space S to the real numbers \mathbb{R} .
- probability function on S induces a probability distribution of X in \mathbb{R} ,

$$P(X \in A) = P\{s \in S : X(s) \in A\}$$

- probability density function (PDF) $f_X(x)$ is defined by

$$P(X = x) = f_X(x) \quad \text{if } X \text{ is discrete}$$

$$P(X \in A) = \int_A f_X(t) dt$$

- the cumulative distribution function (CDF) $F_X(x)$ is defined by

$$F_X(x) = P(X \leq x)$$

As an important example for a discrete random variable, we spent some time looking at the Binomial distribution, which describes the number X of "successes" in a sequence of N independent trials, with a success probability equal to p for each trial. The p.d.f for the binomial distribution was (you should know this for the exam)

$$f_X(x) = P(X = x) = \binom{N}{x} p^x (1-p)^{N-x}$$

Relationship between CDF and PDF

Getting from the PDF to the CDF

- if X is discrete, add up

$$F_X(x) = \sum_{x_i \leq x} f_X(x_i)$$

- if X is continuous, integrate

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(t) dt$$

Getting from the CDF to the PDF

- if X is discrete,

$$f_X(x) = F_X(x^+) - F_X(x^-)$$

- if X is continuous

$$f_X(x) = F'_X(x) = \frac{d}{dx} F_X(x)$$

Also, recall main properties of CDF

- $0 \leq F_X(x) \leq 1$ for all $x \in \mathbb{R}$
- $F_X(x)$ nondecreasing in x
- $F_X(x)$ continuous from the right
- $F_X(x)$ is continuous everywhere if and only if X is continuous

Joint Distributions

Looked at

- joint PDF for X and Y (discrete or continuous)
- marginal distribution of X with PDF

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy$$

- independence of random variables, most importantly

$$f_{XY}(x, y) = f_X(x)f_Y(y) \quad \text{for all } (x, y) \in \mathbb{R}^2$$

if and only if X and Y are independent

- conditional distribution of Y given X ,

$$f_{Y|X}(y|x) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

3 Random Problems

Example 2 (Spring 2003 Exam) A Monet expert is given a painting purported to be a lost Monet. He is asked to assess the chances that it is genuine and has the following information:

- In general, only 1% of the "found" paintings he receives turn out to be genuine, an event we'll call G
- "Found" paintings have a different frequency of use of certain pigments than genuine Monets do:
 - (a) cadmium yellow Y appears in 20% "found" paintings, but only 10% genuine ones
 - (b) raw umber U appears in 80% of "found" paintings, but only 40% of genuine ones
 - (c) burnt sienna S appears in 40% of "found", but 60% of genuine paintings
- This particular painting uses burnt sienna, but not cadmium yellow or raw umber.

What is the probability that this particular painting is genuine? Do we have to make any additional assumptions to answer the question?

This problem has the following structure: the problem seems to tell us what colors ("data" SY^CUC) are how likely to appear *given* that the painting is genuine ("state of the world" G), i.e. $P(B|A)$. But we actually want to know how likely the painting is genuine *given* the colors that were used in it, i.e. $P(A|B)$. So we are trying to switch the order of conditioning, so we'll try to use Bayes' Theorem.

Let's first compile the information contained in the problem:

$$\begin{aligned} P(Y) &= 0.2 \\ P(Y|G) &= 0.1 \\ P(U) &= 0.8 \\ P(U|G) &= 0.4 \\ P(S) &= 0.4 \\ P(S|G) &= 0.6 \end{aligned}$$

and

$$P(G) = 0.01$$

But what do we need to apply Bayes' theorem? - the theorem tells us that

$$P(G|SY^C U^C) = \frac{P(SY^C U^C|G)P(G)}{P(SY^C U^C)}$$

However, know only *marginal* probability of each color, but would need *joint* probabilities (both conditional on G and unconditional).

Therefore we have to make an additional assumption at this point, and the simplest way to attack this is to assume that the use of pigments is independent across the three colors, both unconditionally and conditional on G , i.e.

$$P(SY^C U^C|G) = P(S|G)P(Y^C|G)P(U^C|G) = 0.6 \cdot 0.9 \cdot 0.6$$

and

$$P(SY^C U^C) = P(S)P(Y^C)P(U^C) = 0.4 \cdot 0.8 \cdot 0.2$$

Using Bayes' theorem we get therefore that under the independence assumption

$$P(G|SY^C U^C) = \frac{0.6 \cdot 0.9 \cdot 0.6 \cdot 0.01}{0.4 \cdot 0.8 \cdot 0.2} = \frac{81}{1600}$$

To see how much this assumption mattered, can invent a different dependence structure among the different types of pigments: suppose that for genuine Monets, every painting using sienna S also uses umbra U for sure. Then, by the definition of conditional probabilities

$$P(SY^C U^C|G) \leq P(SU^C|G) = P(U^C|SG)P(S|G) = 0 \cdot 0.6 = 0$$

so that for a true Monet, it is *impossible* to find sienna S , but not umbra, therefore we'd know *for sure* that the painting in questions can't be a Monet (note that since our painting had this combination, it has to be possible for "found" paintings in general).

So, summing up, this problem did *not* give us enough information to answer the question.

Example 3 (Exam Fall 1998) Recycling is collected at my house sometime between 10am and noon, and any particular minute is as likely as any other. Garbage is collected sometime between 8:30am and 11:00am, again with any particular instant as likely as any other. The two collection times are independent.

- (a) What is the joint p.d.f. of the two collection times, R and G ?
- (b) What is the probability that the recycling is collected before garbage?

The *marginal* distribution of R is (continuous) uniform with density

$$f_R(r) = \begin{cases} \frac{1}{2} & \text{if } r \in [10, 12] \\ 0 & \text{otherwise} \end{cases}$$

The *marginal* distribution of G is discrete with p.d.f.

$$f_G(g) = \begin{cases} \frac{2}{5} & \text{if } g \in [8.5, 11] \\ 0 & \text{otherwise} \end{cases}$$

By independence, the joint p.d.f. is

$$f_{GR}(g, r) = f_G(g)f_R(r) = \begin{cases} \frac{1}{5} & \text{if } r \in [10, 12] \text{ and } g \in [8.5, 11] \\ 0 & \text{otherwise} \end{cases}$$

The probability of the event $R \leq G$ can be calculated as

$$P(R \leq G) = \int_{8.5}^{11} \int_{10}^{\max\{g, 10\}} \frac{1}{5} dr dg = \int_{8.5}^{11} \frac{\max\{g - 10, 0\}}{5} dg = \int_{10}^{11} \frac{g - 10}{5} dg = \left[\frac{g^2}{10} - 2g \right]_{10}^{11} = \frac{21}{10} - 2 = \frac{1}{10}$$

Example 4 *One of your classmates asked how one can solve the following problem: how many different ways are there to allocate N indistinguishable blackboards to k different classrooms? This corresponds to choosing a partition of blackboards over k classes. We can do the calculation as follows:*

- *introduce $k - 1$ "separators" Z_1, Z_2, \dots, Z_{k-1} , which we mix with the blackboards B_1, B_2, \dots, B_N*
- *represent each allocation of blackboards to rooms as a reordering of $B_1, B_2, \dots, B_N, Z_1, Z_2, \dots, Z_{k-1}$. The blackboards before the first Z to appear in the sequence are those which we are going to put up in the first classroom, the boards up to the second "separator" go into room 2, etc. If the sequence is e.g. $Z_5, B_7, B_2, B_5, Z_4, B_9, \dots$, then there is going to be no blackboards in room 1, boards 7, 2, and 5 go to room 2 etc.*
- *the number of different orderings of the sequence is $(N + (k - 1))!$*
- *since blackboards and separators are equivalent (classrooms aren't), we have to divide by the number of permutations of each the blackboards ($N!$ permutations), and the separators ($(k - 1)! permutations$).*
- *putting all pieces together, we have*

$$p = \frac{(N + k - 1)!}{N!(k - 1)!} = \binom{N + k - 1}{k - 1}$$

possible allocations.