14.30 Introduction to Statistical Methods in Economics
Spring 2009

14.30 Introduction to Statistical Methods in Economics
Lecture Notes 3

Konrad Menzel

February 10, 2009

1 Counting Rules and Probabilities

Recall that with simple probabilities, each outcome is equally likely, and for a finite sample space, we can give the probability of an event A as

$$P(A) = \frac{n(A)}{n(S)}$$

We'll now see how to make good use of counting rules to calculate these probabilities.

Example 1 *Draw two cards from a deck of 52 cards with replacement, assuming that each card is picked with equal probability. What is the possibility of drawing two different cards?*

$$S = \{(A\clubsuit, A\clubsuit), (A\clubsuit, A\spadesuit), \ldots\} \implies n(S) = 52^2$$

The event "two different cards" consists of

$$A = \{(A\clubsuit, A\spadesuit), (A\clubsuit, A\heartsuit), \ldots\} \implies n(A) = \frac{52!}{(52-2)!} = 52 \cdot 51$$

so that

$$P(A) = \frac{n(A)}{n(S)} = \frac{52!}{(52-2)!(52)^2} = \frac{51}{52} \approx 0.98$$

Alternatively, we could have used proposition 1 on probabilities:

$$P(A) = 1 - P(A^C) = 1 - P("\text{two cards are same}") = 1 - P("\text{2nd card same as 1st}") = 1 - \frac{1}{52} = \frac{51}{52}$$

In some other examples, computing the probability of an event through its complement may simplify the calculation a lot.

Example 2 *Suppose Oceania attacks the capital of Eurasia[1] with 16 missiles, 8 of which carry a nuclear warhead. Suppose the Eurasian army can track all 16 projectiles and has 12 missiles each of which can intercept one incoming missile with absolute certainty, but can't tell which of the missiles carry a conventional load. What is the combinatorial probability that Eurasia cannot avert disaster and at least one of the nuclear warheads reaches its target? What would be your intuitive guess?*

---

[1]The names are taken from Orwell's novel "1984", so this is not supposed to be a real-world example.

Since in any event, exactly 4 projectiles reach their target, the sample space S consists of all combinations of 4 missiles out of 16. Therefore the number of elements of S is given by the binomial coefficient

$$n(S) = \binom{16}{4} = \frac{16!}{12!4!}$$

In order to evaluate the probability, one approach is to use the complement rule. The event complementary to $A = $ "At least one nuclear warhead hits target" is $A^C = $ "All missiles hitting the target are conventional", and the outcomes in A^C are given by all combinations of 4 missiles out of 8 (the conventional ones), so that

$$n(A^C) = \binom{8}{4} = \frac{8!}{4!4!}$$

Therefore,

$$P(A) = 1 - P(A^C) = 1 - \frac{12!4!8!}{16!4!4!} = 1 - \frac{12!\,8!}{16!\,4!} = 1 - \frac{1}{16 \cdot 15 \cdot 14 \cdot 13} \frac{8 \cdot 7 \cdot 6 \cdot 5}{1} = \frac{25}{26}$$

So it turns out that this probability is extremely close to one - I'm not sure whether you would have expected this, but despite being politically incorrect, this example shows that our intuitions may fail easily in combinatoric problems, last but not least because of the high numbers of possibilities.

Example 3 *The famous birthday "paradox" is about a (once) popular party game: given you have a group of n friends, what is the probability that at least a pair of them has the same birthday (assuming that all birthdays are equally likely, which is actually only approximately true empirically)? Again, let's look at the complementary event A^C that each of your n friends has a different birthday: since this corresponds to drawing n out of 365 without replacement, we can use the corresponding formula*

$$n(A^C) = \frac{365!}{(365 - n)!}$$

so that we can calculate the probability $P(A)$ of at least two of your friends having the same birthday:

$$P(A) = 1 - P(A^C) = 1 - \frac{n(A^C)}{n(S)} = 1 - \frac{365!}{(365 - n)!365^n}$$

This formula is not yet particularly easy to read, so let's just write down the probabilities in decimals for a few values of n:

n	$P(A)$
5	0.027
10	0.117
15	0.253
20	0.411
25	0.569
30	0.706
50	0.97
366	1

Many people find these probabilities very high, but it's usually because one is tempted to start thinking about the problem by calculating the probability that any of your n friends has the same birthday as you. *You can convince yourself that this probability is*

$$P(\tilde{A}) = 1 - \left(1 - \frac{1}{365}\right)^n = 1 - \left(\frac{364}{365}\right)^n$$

for which our list looks different: The reason for this discrepancy is that in the previous situation, A also

n	$P(A)$
5	0.014
10	0.027
15	0.040
20	0.053
25	0.066
30	0.079
50	0.128
366	0.634

contained all pairs among *your n friends, and that number went up very quickly.*

2   Independent Events

Intuitively, we want to define a notion that for two different events A and B the occurrence of A does not "affect" the likelihood of the occurrence of B, e.g. if we toss a coin two times, the outcome of the second toss should not be influenced in any way by the outcome of the first. In order to keep notation simple, we will from now on denote

$$P(A \cap B) = P(AB)$$

Definition 1 *The events A and B are said to be* independent *if*

$$P(A \cap B) = P(A)P(B)$$

From this you can see that independence is merely a property of the probability distribution, not necessarily of the physical nature of the events. So while in some examples (e.g. the series of coin tosses) we have a good intuition for independence, in most cases we'll have no choice but need to check the formal condition.

Example 4 *Say we roll a fair die once, what are the probabilities for the events*

$$\begin{aligned} A &= \{2,4,6\} \\ B &= \{1,2,3,4\} \end{aligned}$$

and their intersection? Counting outcomes, $P(A) = \frac{n(A)}{n(S)} = \frac{3}{6} = \frac{1}{2}$, and similarly, $P(B) = \frac{4}{6} = \frac{2}{3}$. The probability of the intersection of events is

$$P(AB) = P(\{2,4\}) = \frac{2}{6} = \frac{1}{3} = P(A)P(B)$$

so the events are independent even though they resulted from the same *roll.*
In order to see how independence depends crucially on the underlying probability distribution, now suppose

that the die had been manipulated so that $P(6) = \frac{3}{8}$, whereas for all other numbers $n = 1, \ldots, 5$, $P(n) = \frac{1}{8}$.
Then, by axiom (P3) on adding probabilities of disjoint events,

$$P(A) = \frac{1}{8} + \frac{1}{8} + \frac{3}{8} = \frac{5}{8}, \qquad P(B) = \frac{4}{8} = \frac{1}{2}$$

and

$$P(AB) = \frac{2}{8} = \frac{1}{4} < P(A)P(B) = \frac{5}{16}$$

One interpretation of independence is the following: suppose we know that B has occurred, does that knowledge change our beliefs about the likelihood of A (and vice versa)? We'll formalize this in the next section, and it will turn out that if A and B are independent, there is nothing about event A to be learnt from the knowledge that B occurred.

Proposition 1 *If A and B are independent, then A and B^C are also independent.*

PROOF: Since we can partition A into the disjoint events AB and AB^C, we can write

$$P(AB^C) \overset{(P3)}{=} P(A) - P(AB) \overset{\text{Indep.}}{=} P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^C)$$

proving independence $\square$

We can now extend the definition of independence to more than two events:

Definition 2 *A collection of events $A_1, A_2, \ldots$ are independent if for any subset $A_{i_1}, A_{i_2}, \ldots$ of these events (all indices being different)*

$$P(A_{i_1} \cap A_{i_2} \cap \ldots) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot \ldots$$

E.g. for three events A, B, C,

$$P(AB) = P(A)P(B), \qquad P(AC) = P(A)P(C), \qquad P(BC) = P(B)P(C)$$

and

$$P(ABC) = P(A)P(B)P(C)$$

Example 5 *Let the sample space be $S = \{s_1, s_2, s_3, s_4\}$, and $P(s_i) = \frac{1}{4}$ for all outcomes. Then each of the events*

$$A = \{s_1, s_2\}, \qquad B = \{s_1, s_3\}, \qquad C = \{s_1, s_4\}$$

occurs with probability $\frac{1}{2}$. The probability for the event $A \cap B$ is

$$P(AB) = P(\{s_1\}) = \frac{1}{4} = P(A)P(B)$$

and likewise for any other pair of events, so the events are pairwise independent. However, taken together, the full collection is not independent since

$$P(ABC) = P(\{s_1\}) = \frac{1}{4} \neq \frac{1}{8} = P(A)P(B)P(C)$$

Intuitively, once we know that both A and B occurred, we know for sure that C occurred.

4

3 Conditional Probability

Suppose the occurrence of A affects the occurrence (or non-occurrence) of B and vice versa. How do we describe the probability of B given knowledge about A? - We already argued heuristically that if the events are independent, then A does not reveal any information about B. But what if it does? How do we change probabilities as a result?

Example 6 *If we throw a fair die, and I tell you that in fact the outcome was an even number, i.e that the event $B = \{2, 4, 6\}$ occurred. What's the probability of having rolled a 6? Since there are only three equally likely possibilities in B, 6 being one of them, we'd intuitively expect the answer to be $\frac{1}{3}$. Here we basically simplified the sample space, to $\tilde{S} = B = \{2, 4, 6\}$, and calculated the simple probability for the redefined problem.*

Definition 3 *Suppose A and B are events defined on S such that $P(B) > 0$.* The conditional probability of A given *that B occurred is given by*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Intuitively, the numerator redefines which outcomes in A are possible once B is known to have occurred, the denominator does the same thing for the whole sample space.
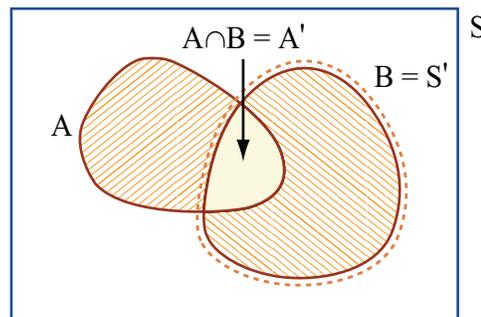
Figure 1: The Event A Conditional on B

Remark 1 *Conditional probability and independence: if A and B are independent,*

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$$

so B occurring tells us nothing about A, so the conditional probability is the same as the unconditional probability.

Example 7 *This example is adapted from Greg Mankiw's blog:[2] On platforms like Intrade, you can trade assets which pay 1\$ if a given event (e.g. Yankees win the World Series) occurs. If the market works as it should, the prices of this type of assets at a given time t can be interpreted as probabilities given the information traders have at that point in time. On the political market on Intrade, you can trade assets for the events*

---

[2]http://gregmankiw.blogspot.com/2006/11/bayes-likes-obama

- A_i that candidate i wins the presidential elections (without conditioning on nomination)

- B_i that candidate i wins the nomination of her/his party

- C_k that the nominee of party k wins the election

We can now use the probabilities which are implied by the asset for the corresponding event to answer the question what the market "thinks" which candidate of either party has the highest probability of winning the presidential elections if nominated $P(A_i|B_i)$ - i.e. nominating which candidate would give each party the highest chances of winning the presidency.
We can (relatively) safely assume that a candidate who is not nominated by the party has no chances of winning the presidency, so that

$$A_i \subset B_i \implies A_i \cap B_i = A_i$$

so that

$$P(A_i|B_i) = \frac{P(A_i \cap B_i)}{P(B_i)} = \frac{P(A_i)}{P(B_i)}$$

so that we only have to plug in the prices for the corresponding assets. Based on asset prices on February 6 on the Intrade political markets, we get the following numbers (in the last column I report the values from Mankiw's original blog entry, as of November 2006).

candidate	$P(A_i)$	$P(B_i)$	$P(A_i|B_i)$	$\tilde{P}_{Nov'06}(A_i|B_i)$
Clinton	28.7%	45.2%	63.5%	51%
Huckabee	0.5%	2.0%	25.0%	NA
McCain	34.4%	93.0%	37.0%	63%
Obama	35.0%	53.0%	66.0%	88%
Paul	0.4%	1.2%	33%	NA
Romney	1.2%	2.6%	46.2%	50%

In order to distinguish it from the *conditional* probabilities $P(A|B_i)$, $P(A)$ is also called the *marginal* probability of A. The relationship between marginal and conditional probabilities is given by the Law of Total Probability:

Theorem 1 *(Law of Total Probability)* *Suppose that $B_1, \ldots, B_n$ is a partition of the sample space S such that $P(B_i) > 0$ for every $i = 1, \ldots, n$. Then*

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

for any event B.

PROOF: From the definition of a conditional probability, $P(A|B_i)P(B_i) = P(A \cap B_i)$ for any event B_i. Since $B_1, \ldots, B_n$ is a partition of the sample space, $(A \cap B_1), \ldots, (A \cap B_n)$ are disjoint and exhaustive for A - i.e. constitute a partition for A. Therefore, by the axiom (P3) on probabilities of unions of disjoint sets, $\sum_{i=1}^{n} P(A \cap B_i) = P(A)$ $\square$

Example 8 *In medical data we can often find that patients treated by older, and more experienced, heart surgeons have in fact* higher *post-operative death rates than those operated by younger surgeons - say we observe a death rate of 6.0% for experienced surgeons, and onlyl 5.5% for unexperienced surgeons.*

Does this mean that the surgeons' skill decreases with age? Probably not - let's suppose there are four different types of procedures a surgeon may have to perform - single, double, triple, and quadruple bypass (the terminology refers to the number of coronary arteries that have to be bypassed artificially). The complexity of the procedure and the risk to the patient increase in the number of bypasses, and it might also be that the patients who are generally "sicker" may tend to require a more complicated procedure. Suppose we are told that for each procedure, patients of the experienced surgeon face significantly lower death rates, but that the overall patient mortality is lower for unexperienced surgeons. In light of the law of total probability, how does that fit together? Let's look at an example (these numbers are of course made up):

Procedure	Unexperienced		Experienced	
	Death Rate	Percentage of Cases	Death Rate	Percentage of Cases
Single Bypass	4.0%	50.0%	2.0%	25.0%
Double Bypass	6.0%	40.0%	4.0%	25.0%
Triple Bypass	10.0%	9.0%	6.0%	25.0%
Quadruple Bypass	20.0%	1.0%	12.0%	25.0%
Total	5.5%	100.0%	6%	100.0%

In the notation of the Law of Total Probability, the overall death rate $P(A)$ for, say, experienced surgeons can be computed from the death rate conditional on procedure B_i, $P(A|B_i)$, and the base rates/proportions $P(B_i)$ of cases corresponding to each procedure.

We can see that since experienced surgeons are assigned a disproportionately large share of risky cases (presumably because you need more experience for these), their average *(better:* marginal*) death rate is higher than that of unexperienced surgeons, even though they perform better conditional on each treatment category. This phenomenon is often referred to as a* composition effect*.*

So what is the practical importance of each type of probabilities? If you were to choose among surgeons for a bypass operation, the type of procedure should only depend on your health status, not whether the surgeon is experienced or not, so in that situation you should only care about the conditional probabilities. It's harder to come up with a good use for the marginal death rates.

In most statistical analysis, you'd in fact be interested in conditional death rates (e.g. if you are interested in the effect of experience on mortality), and the variable "type of procedure" would be treated as what statisticians call a confounding factor. A classical problem in statistics and econometrics is that often many of the relevant confounding factors are not observed, and you'll learn about "tricks" of dealing with that problem.

Remark 2 *Another closely related concept is that of* conditional independence*, which is going to be very important in Econometrics. Two events A and B are said to be independent* conditional *on event C if*

$$P(AB|C) = P(A|C)P(B|C)$$

It is important to note that

- *unconditional independence does* not *imply conditional independence*

- *conditional independence does* not *imply unconditional independence*

i.e. whether A and B are independent depends crucially on what we are conditioning on. There'll be an exercise with a counterexample on the next problem set.

4 Conditional Independence (not covered in lecture)

We can extend the definition of independence to conditional probabilities:

Definition 4 *Two events A and B are said to be independent* conditional *on event C if the conditional probabilities satisfy*

$$P(AB|C) = P(A|C)P(B|C)$$

This definition corresponds exactly to that of unconditional independence which we looked at before, only that we restrict ourselves to the new sample space $S' = C$. The notion of conditional independence is going to play an important role later on in econometrics, so it deserves a special mention. As a technical point, it is important to notice that conditional independence does not imply unconditional independence, or vice versa. In other words, whether two events are independent depends crucially on what else we are conditioning on. I mentioned this in passing in the last lecture, and I'm now going to provide the following example as an illustration:

Example 9 *Let's look again at a roll of a fair die, i.e. $S = \{1,2,3,4,5,6\}$, where each outcome occurs with probability $\frac{1}{6}$.*

(1) Making two independent events dependent: Consider the events $A = \{1,2,3,4\}$ and $B = \{2,4,6\}$. We already saw in a previous example that these events are independent:

$$P(A)P(B) = P(\{1,2,3,4\})P(\{2,4,6\}) = \frac{4}{6} \cdot \frac{3}{6} = \frac{1}{3} = P(\{2,4\}) = P(AB)$$

Now let event $C = \{3,6\}$. Then

$$P(A|C) = \frac{P(AC)}{P(C)} = \frac{P(\{3\})}{P(\{3,6\})} = \frac{1}{2} = \frac{P(\{6\})}{P(C)} = \frac{P(BC)}{P(C)} = P(B|C)$$

However,

$$P(AB|C) = \frac{P(\emptyset)}{P(C)} = 0 \neq P(A|C)P(B|C)$$

i.e. A and B are not independent conditional on C since their intersections with C are disjoint.

(2) Making two dependent events independent: Let $D = \{2,3,4\}$ and $E = \{2,4,6\}$. We can check that D and E are dependent: we can see that $P(D) = P(E) = \frac{1}{2}$. However,

$$P(DE) = P(\{2,4\}) = \frac{1}{3} \neq \frac{1}{4} = P(D)P(E)$$

But if we condition on $F = \{3,4,5,6\}$,

$$P(DE|F) = \frac{P(\{4\})}{P(\{3,4,5,6\})} = \frac{1}{4}$$

whereas

$$P(D|F)P(E|F) = \frac{P(\{3,4\})}{P(\{3,4,5,6\})} \cdot \frac{P(\{4,6\})}{P(\{3,4,5,6\})} = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

so that by conditioning on F, D and E became independent.