Econ 300/QAC 201: Quantitative Methods in Economics/Applied Data Analysis

4th Class **6/10/10**

"There are two times in a man's life when he should not speculate: when he can't afford it, and when he can." -- Mark Twain

summarizing our important results so far:

1. addition law of probability:

$$Pr(A_1 \cup A_2) = Pr(A_1) + Pr(A_2) - Pr(A_1 \cap A_2)$$

2. finding probability of an event by finding the probability of its complement:

$$Pr(E) = 1 - Pr(\overline{E})$$

3. definition of conditional probability:

$$Pr(A_2 | A_1) = Pr(A_1 \cap A_2) / Pr(A_1)$$

4. multiplication law of probability:

$$Pr(A_1 \cap A_2) = Pr(A_2)Pr(A_1 | A_2) = Pr(A_1)Pr(A_2 | A_1)$$

5. importance of the assumption of independence for simplifying multiplication and addition laws:

if A_1 and A_2 are independent, so:

$Pr(A_1 | A_2) = Pr(A_1)$ and $Pr(A_2 | A_1) = Pr(A_2)$, then:

$$Pr(A_1 \cap A_2) = Pr(A_1) * Pr(A_2)$$

$$Pr(A_1 \cup A_2) = Pr(A_1) + Pr(A_2) - Pr(A_1) * Pr(A_2)$$

clarify the relationship between mutual exclusiveness, where
($\Pr(A_1 \cap A_2) = 0$, and independence

define statistical dependence as opposite of independence:
$\Pr(A_1 | A_2) \neq \Pr(A_1)$ and $\Pr(A_2 | A_1) \neq \Pr(A_2)$

if two events are mutually exclusive they must be dependent:

$$\Pr(A_2 | A_1) = (\Pr(A_1 \cap A_2) \;/\; \Pr(A_1) = \frac{0}{\Pr(A_1)} \;=\; 0$$

but $0 \neq \Pr(A_2)$ in general

so $\Pr(A_2 | A_1) \neq \Pr(A_2)$

so if two events are independent they must not be mutually exclusive

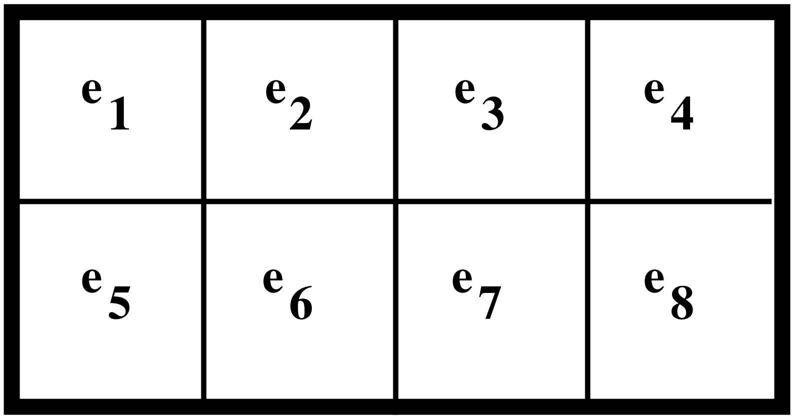
but if two events are dependent they need not be mutually exclusive

Let me do a problem to illustrate the three cases that can occur:

consider 3 tosses of a fair coin. Eight outcomes are possible:

e_1	HHH
e_2	HHT
e_3	HTH
e_4	HTT
e_5	THH
e_6	THT
e_7	TTH
e_8	TTT

we can show the sample space graphically:

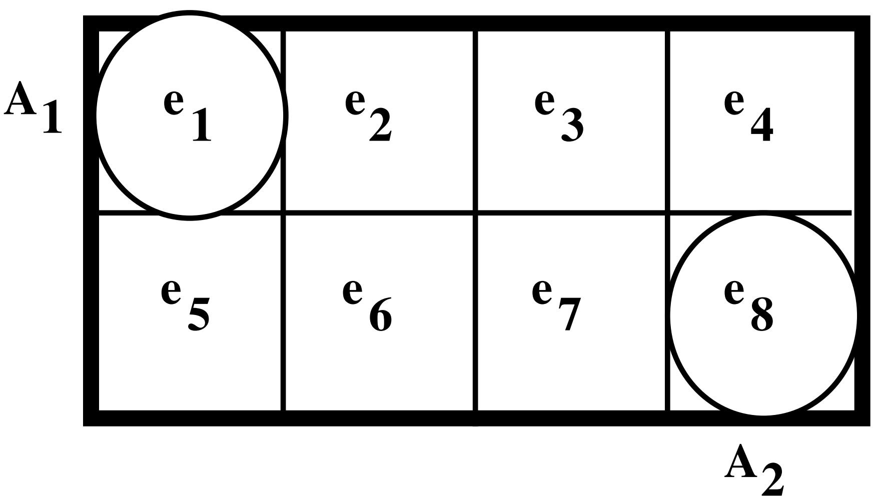


consider three cases:

1) A_1 and A_2 are mutually exclusive and therefore dependent

A_1 = all heads $\quad\quad$ $Pr(A_1) = \frac{1}{8}$

A_2 = all tails $\quad\quad$ $Pr(A_2) = \frac{1}{8}$

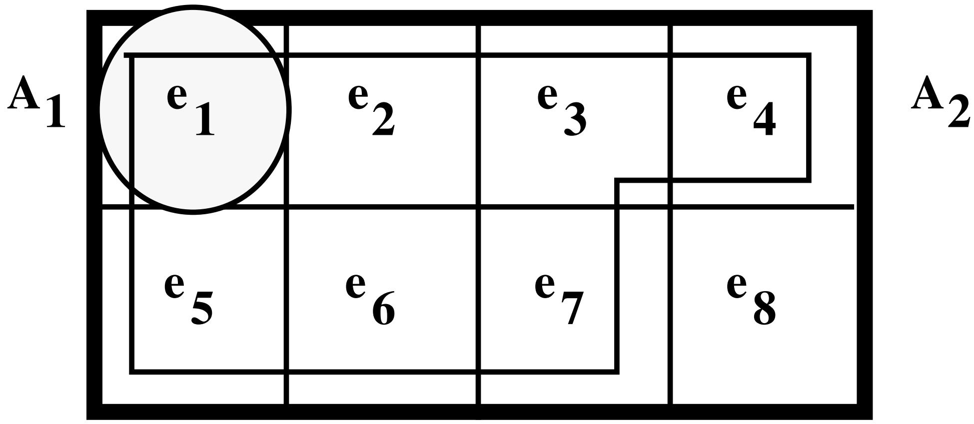


$Pr(A_1 \mid A_2) = 0$
$Pr(A_2 \mid A_1) = 0$

2) A_1 and A_2 are dependent, but not mutually exclusive

A_1 = all heads $\quad\quad$ $Pr(A_1) = \frac{1}{8}$

A_2 = at least one head $\qquad Pr(A_2) = \frac{7}{8}$



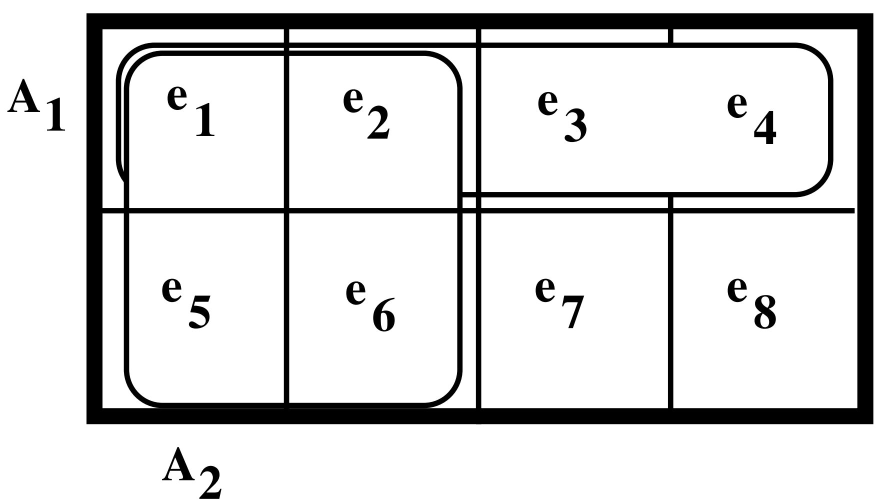
$Pr(A_1 | A_2) = \frac{1}{7}$

$Pr(A_2 | A_1) = 1$

3) A_1 and A_2 are independent

A_1 = first coin is heads $\qquad Pr(A_1) = \frac{1}{2}$

A_2 = second coin is heads $\qquad Pr(A_2) = \frac{1}{2}$



$Pr(A_1 | A_2) = \frac{1}{2} = Pr(A_1)$

$Pr(A_2 | A_1) = \frac{1}{2} = Pr(A_2)$

Now do a few problems to illustrate use of probability:

for fans of probability questions, “Ask Marilyn” has tons of them. for example, from her column of Nov. 14, 1993:

Question: There are five cars on display as prizes, and their five ignition keys are in a box. You get to pick one key out of the box and try it in the ignition of one car. If it fits, you win the car. What are your chances of winning a car?

answer: 1/5. You may have been tempted to say 1/25, which is the probability that you win the red Ferrari secretly preferred by a member of the audience, but if you prefer the red Ferrari, the probability switches back to 1/5.

Question: Consider the following five events in Reno-style Roulette (36 numbers divided evenly between red and black, plus 2 green numbers, 0 and 00):

A is the event that the wheel comes up green on the first spin
B is the event that the wheel comes up black on the first spin
C is the event that the wheel comes up green or black on the first spin
D is the event that the wheel comes up green or red on the first spin
E is the event that the wheel comes up green on the second spin.

a) which two of these events are complements? answer: B and D
b) which two of these events are mutually exclusive but not complements?
answer: A and B
c) which two of these events are independent? answer: E and any of the others
d) which one of these events is the union of two of the other four? which two?
answer: C is the union of A and B
e) which one of these events is the intersection of two of the other four? which two?
answer: A is the intersection of C and D

Question: Suppose that 33% of all pets wear collars, that 60% of all pets are cats, and that 81% of all pets are cats or wear collars.

a) What percentage of pets are cats who wear collars?

answer:

$\Pr(A_1 \cup A_2) = \Pr(A_1) + \Pr(A_2) - \Pr(A_1 \cap A_2)$

rearranging, $\Pr(A_1 \cap A_2) = \Pr(A_1) + \Pr(A_2) - \Pr(A_1 \cup A_2)$

If A_1 is the event of collar-wearing and A_2 is the event of cat-being, then

Pr(collar-wearing cat) = .33 + .60 - .81 = .12 or 12%

b) what percentage of cats wear collars?

answer: $\frac{.12}{.60}$ = .20 or 20%

rework cats and collars problem showing it in a 2x2 table:

		collar		
		y	n	
cat	y	.12	.48	.60
	n	.21	.19	.40
		.33	.67	

you only know the .33 and .60 and .81, from these you get the marginal distributions and knowing .81 gets you 1 - .81 = .19

Question: You have a sample containing 40% men and 60% women. Suppose that 6% of the men are colorblind and 1% of the women are colorblind.

a) what is the probability that a person selected at random is colorblind?

answer: can set up a 2x2 table to show how the sample breaks into 4 groups:

	sex	
	male	female
not colorblind	37.6%	59.4%
colorblind	2.4%	0.6%

so Pr(colorblind) = 3%

b) if a person selected at random is colorblind, what is the probability that this person is a man?

answer: $\Pr(\text{man} \mid \text{colorblind}) = \frac{2.4}{3} = 80\%$

Question: In an episode of the TV show "Star Trek" entitled "The Empath," some aliens are holding a few members of the crew of the Starship Enterprise captive. One of these aliens states that the probability is 87% that Dr. McCoy will die, while the probability is

93% that Mr. Spock will suffer severe brain damage. What is the probability that neither event will occur? Guess what actually does happen.

answer:

Pr(neither occurs) = .13*.07 = .0091 or .9%

Question: On the TV game show Jeopardy on April 3, 1985, the final question had four possible answers. All three contestants had absolutely no idea, so each had to guess at random and independently of the other contestants. What is the probability that at least one contestant would get the right answer?

answer:

Pr(no one gets it) = $(\frac{3}{4})3$ = .42, so Pr at least one gets it = 1 - .42 = .58 or 58%

(the question was, "In a deck of cards, which one of the four suits does not have a one-eyed face card? The answer is clubs, and none of the contestants got it right.)

Important application of probability theory to consideration of rationality, which is after all an underpinning of economic theory: Consider relationship of probability theory to people's intuitive understanding of likelihood

Question: Linda is 31, single, outspoken, and very bright. She majored in philosophy in college. As a student, she was deeply concerned with discrimination and other social issues, and participated in anti-nuclear demonstrations. Which statement is more likely:

a. Linda is a bank teller

b. Linda is a bank teller and active in the feminist movement

answer: as a principle of probability, the likelihood of any two uncertain events happening together is always less than the odds of either happening alone, so a. is more likely.

Principle: the problem-solving method of representativeness, in which the mind compares an event to a mental model to decide how likely the event is, can lead you astray.

Question: In a typical English text, does the letter k appear more often as the first letter in a word or the third letter?

answer: k appears about twice as often in the third position. Most people judge that k is commoner at the beginning of words because it is easier to recall words that begin with a letter than those that contain it somewhere in the middle
Principle: the problem-solving method/mental short cut of availability, which occurs when people judge the likelihood of something happening by how easily they can call other examples of the same thing to mind, can lead you astray.

Question: A cab was involved in a hit-and-run accident. Two cab companies serve the city: the Green, which operates 85 percent of the cabs, and the Blue, which operates the remaining 15 percent. A witness identifies the hit-and-run cab as Blue. When the court tests the reliability of the witness under circumstances similar to those on the night of the accident, he correctly identifies the color of a cab 80 percent of the time and misidentifies it the other 20 percent. What's the probability that the cab involved in the accident was Blue, as the witness stated?

answer:

can show this application of Bayes' Theorem using a tree form:

	.80 "green"	= .68	
.85 green			so 71% "green"
	.20 "blue"	= .17	
	.20 "green"	= .03	
.15 blue			and 29% "blue"
	.80 "blue"	= .12	
		100% of sample	

So $Pr(\text{blue} \mid \text{"blue"}) = \frac{.12}{.29} = .41$ or 41%

Principle: it is important to consider the base rate, or background data, against which the probability of an event must be judged (i.e. need to apply Bayes' Theorem and take prior probability into account).

Bayes' Theorem

idea is to take into account prior knowledge of the distribution of outcomes along with data regarding outcomes in arriving at a posterior probability for an event

Let's do another problem before proceeding to a formal statement of it:

Question: Suppose that 80% of Econ 300 students study for their first quiz, that 90% of those who study pass the quiz, and that 40% of those who do not study pass the quiz. A student chosen at random is found to have failed: what is the probability that this student studied?

answer:

	.90 pass	= .72	
.80 study			so 80% pass
	.10 fail	= .08	
	.40 pass	= .08	
.20 don't study			and 20% fail
	.60 fail	= .12	
		100% of sample	

So $\Pr(\text{studied} \mid \text{failed}) = \frac{.08}{.20} = .40$ or 40%

this is a very old result, Rev. Thomas Bayes (1702-61) having had this result published posthumously in 1763.

formal statement of Bayes' Theorem:

Let the events $A_1, \ldots, A_k$ form a partition of the space S such that $\Pr(A_j) > 0$ for $j = 1,\ldots,k$, and let B be any event such that $\Pr(B) > 0$. Then for $i = 1,\ldots,k$,

$$\Pr(A_i \mid B) = \frac{\Pr(A_i)\Pr(B \mid A_i)}{\sum_{j=1}^{k} \Pr(A_j)\Pr(B \mid A_j)}$$

Proof: by the definition of conditional probability,

$$\Pr(A_i \mid B) = \Pr(A_i \cap B) \;/\; \Pr(B)$$

of which the numerator $\Pr(A_i \cap B) = \Pr(A_i)\Pr(B \mid A_i)$ and the denominator $\Pr(B) = \sum_{j=1}^{k} \Pr(A_j)\Pr(B \mid A_j)$. Q.E.D.

Problem: three different machines, M_1, M_2, and M_3, are used to produce a large batch of manufactured items. Suppose that 20% were produced by M_1, 30% by M_2, and 50% by M_3. Suppose further that 1% of the items produced by M_1 are defective, 2% of those produced by M_2 are defective, and 3% of those produced by M_3 are defective. Finally, suppose that one item is selected at random from the entire batch and is found to be defective. Determine the probability that this item was produced by machine M_2.

answer: Let A_i be the event that the selected item was produced by machine M_i $(i = 1,2,3)$ and let B be the event that the selected item is defective. We are attempting to calculate the conditional probability $\Pr(A_2 \mid B)$.

From the problem's conditions, $\Pr(A_1) = .2$, $\Pr(A_2) = .3$, $\Pr(A_3) = .5$

Also, the probabilities $\Pr(B \mid A_i)$ that an item produced by machine M_i is defective are: $\Pr(B \mid A_1) = .01$, $\Pr(B \mid A_2) = .02$, $\Pr(B \mid A_3) = .03$

Then it follows from Bayes' Theorem that:

$$\Pr(A_2 \mid B) = \frac{\Pr(A_2)\Pr(B \mid A_2)}{\sum_{j=1}^{3} \Pr(A_j)\Pr(B \mid A_j)}$$

$$= \frac{(.3)(.02)}{(.2)(.01) + (.3)(.02) + (.5)(.03)} = \frac{.006}{.002 + .006 + .015} = \frac{6}{23} = .26$$

Note that $\Pr(A_2 \mid B)$ is lower than $\Pr(A_2)$.

A probability like $\Pr(A_2)$ is often called the <u>prior probability</u> that the selected item has been produced by machine M_2 and a probability like $\Pr(A_2 \mid B)$ is called the <u>posterior probability</u> that the selected item was produced by M_2 because it is the probability of this event after it is known that the selected item is defective.

<u>Probability distributions</u>

change our terminology slightly now

a discrete random variable X takes on various values x with probabilities specified by its probability distribution p(x), so $\Pr(X=1) = p(1)$

reduction of sample space: care about numbers, not order of events, for most problems

can add up relevant probabilities p(x) to answer new questions:

e.g., $Pr(X < 2) = p(0) + p(1)$

can graph x vs. p(x) —x on horizontal axis, p(x) on vertical axis [draw graph], where the heights of the lines at each discrete value of x sum up to 1 (so p(x) is bounded between 0 and 1 inclusive)

We now distinguish between the mean and variance of a random variable from its probability distribution and the mean and variance of a sample of observations. The mean is known more generally as the first moment and the variance is known as the second moment:

	sample moments	population moments
mean:	$\overline{X} = \sum x\frac{f}{n}$	$\mu = \sum xp(x)$
variance:	$s^2 \approx MSD = \sum (x - \overline{X})^2\frac{f}{n}$	$\sigma^2 = \sum (x - \mu)^2 p(x)$

Note. occasionally one also uses the third moment, also known as skewness, which tells whether there is more weight in the left or right tail (if positive, the distribution is skewed to the right, i.e., has a longer right tail; if negative, it is skewed to the left, i.e., has a longer left tail): $\sum (x - \mu)^3 p(x)$; and the fourth moment, also known as kurtosis, which tells how much weight is in the tail as opposed to the center of the distribution (gives an idea of flatness vs. spikeness of distribution—the closer to 0, the more "light-tailed" it is): $\sum (x - \mu)^4 p(x)$. These are shown in the population moment form; they have sample moment analogies as well.

platykurtic—thinner tails (flatter); mesokurtic—like a normal distribution; leptokurtic—thicker tails (more spiked)

Important simplification of the calculation of population variance:

$$\sigma^2 = \sum x^2 p(x) - \mu^2$$

We will also start using the expected value symbol now:

for the random variable X,

μ_X = mean of X
= average of X
= expected value of X, or E(X)

so $E(X) = \sum xp(x)$

This carries over into functions of X:

if $G = g(X)$, then $E(G) = E[g(X)] = \sum g(x)p(x)$

note one possible function $G = g(X) = (X - \mu)^2$

so $E[(X - \mu)^2] = \sum (x - \mu)^2 p(x)$

$= \sigma^2_X$

and σ^2 can be calculated as $\sigma^2 = E(X^2) - \mu^2$ [see handout for full proof]

If G is a linear function of X:

$G = a + bX$

then the formulas for μ_G and σ_G are simple:

$\mu_G = a + b\mu_X$

$\sigma_G = |b| \sigma_X$

If G is nonlinear, have to use the general formulas to calculate μ_G and σ_G:

$\mu_G = \sum g(x)p(x)$

$$\sigma^2_G = \sum (g(x) - \mu_G)^2 p(x)$$

For tomorrow, finish reading Chapter 4.

Elementary Rules of Expectations

1. $E(k) = k$ (where k is any number; i.e., the expectation of a constant is the constant)

2. $E(k*X) = k*E(X)$ (where k is any number)

3. $E(X + Y) = E(X) + E(Y)$

Using these rules one can show that:

$$\sigma^2 = E[(X - \mu)^2] = E(X^2) - \mu^2$$

Proof:

$$E[(X - \mu)^2] = E[(X - \mu)*X - (X - \mu)*\mu]$$

$$= E[X^2 - \mu*X - X*\mu + \mu^2]$$

$$= E(X^2) - E(\mu*X) - E(X*\mu) + E(\mu^2) \qquad \text{[Rule 3]}$$

$$= E(X^2) - \mu*E(X) - \mu*E(X) + \mu^2$$

[by Rules 1 and 2 and the fact that μ is a constant]

$$= E(X^2) - \mu^2 - \mu^2 + \mu^2 \text{[since } E(X) = \mu\text{]}$$

$$= E(X^2) - \mu^2$$