AAE 706: Probability

Thomas F. Rutherford

Department of Agricultural and Applied Economics
University of Wisconsin Madison

January 28, 2021

Today's Agenda

- Probability review
- Multivariate distributions and conditional probability
- Bayes theorem
- Applications

Language of Probability

- Sample Space
- Events
- Disjoint events
- Axioms of probability
- Random variable
- Distribution function
- Probability density function
- Discrete versus continuous distributions
- Multivariate models
- Independent random variables
- Marginal probability function
- Conditional distribution
- Sample information
- Prior probability
- Likelihood function
- Expected value
- Expectation operator
- Moment of a distribution
- Central moment
- Standard deviation
- Covariance
- Correlation
- Conditional expectation

Language of Advanced Probability

- Coefficient of variation
- Relative skewness
- Relative kurtosis
- Chebyshev inequality
- Moment generating function
- Conjugate distribution

Sample Space and Events

sample space S is the set of anything that can happen which is of interest to the investigator.

event A is a subset of the sample space S.

disjoint events – two events A_i and A_j are disjoint if $A_i \cap A_j = \varnothing$

Axioms of Probabilities

The Kolmogorov axioms are the foundations of probability theory. They were introduced by Russian mathematician Andrey Kolmogorov in 1933. $Pr(\cdot)$ is a probability function if it satisfies:

1. $Pr(A_i) \geq 0$ for any $A_i \in S$
2. $Pr(S) = 1$
3. $Pr(A_i \cup A_j) = Pr(A_i) + Pr(A_j)$ if A_i and A_j are disjoint events.

Partitioning the Sample Space

- A box contains three white and two red balls:

$$w_1, w_2, w_3, r_1, r_2$$

- We remove from it "at random" two balls in succession. What is the probability that the first ball is white and the second red?

First Solution

An element of our space is a pair of different balls. The total number of such pairs is 20, namely,

$$w_1 w_2, w_1 w_3, w_2 w_3, w_1 r_1, w_1 r_2, w_2 r_1, w_2 r_2, w_3 r_1, w_3 r_2, \ldots$$

All are equally probable and the probability of each elementary event is then $1/20$. The event (white first, red second) consists of the six elements

$$w_1 r_1, w_1 r_2, w_2 r_1, w_2 r_2, w_3 r_1, w_3 r_2$$

Therefore its probability equals $6/20$.

Second Solution

It is easy to see that $P(\textit{white first}) = 3/5$

After choose a white ball first, the chance of choosing a red ball second is

$$P(\textit{red second} | \textit{white first}) = 2/4$$

The probability of the two events is then:

$$P(\textit{white first}, \textit{red second}) = 3/5 \times 2/4 = 6/20$$

Random Variable

A random variable X is a real value function that has a specific value at each point of the sample space.

For any $B \subset S$, the probability that $X \in B$ is $Pr(X \in B)$.

Cumulative Distribution

A distribution function is the function $F(t) = Pr(X \leq t)$ such that:

1. $F(t)$ is non-decreasing and continuous from the right
2. $F(-\infty) = 0$
3. $F(+\infty) = 1$

Probability Functions

A probability function $= f(x)$ where x can be a discrete or a continuous variable.

- **discrete case:** when X can take a countable number of distinct values: $x^1, x^2, x^3, \ldots$ Then,

$$f(x^i) = Pr(X = x^i)$$

and

$$Pr(X \in B) = \sum_{x^i \in B} f(x^i)$$

- **continuous case** the function $f(x)$ satisfies

$$Pr(x \in B) = \int_{x \in B} f(x) \mathrm{dx}$$

where $f(x) = \frac{\mathrm{dF(x)}}{\mathrm{dx}}$

NB: $Pr(X = x) = 0$ in the continuous case.

Multivariate Distribution

In the multivariate case, $x = (x_1, x_2, \ldots, x_n)$ where n is the number of continuous random variables.

- The joint distribution function of x is

$$F_n(x) = Pr(X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n)$$

- The *marginal distribution* of the subset $(x_1, x_2, \ldots, x_k)$, $k < n$, is

$$F_k(x_1, x_2, ..., x_k) = F_n(x_1, x_2, ..., x_k, \infty, \ldots, \infty)$$

Marginal Probability

The marginal probability function is

$$f_k(x_1, \dots, x_k) = \int \dots \int f_n(x_1, \dots, x_n) dx_{k+1} \dots dx_n$$

in the continuous case, and

$$f_k(x_1, \dots, x_k) = \sum_{x_{k+1}, \dots, x_n} f_n(x_1, \dots, x_n)$$

in the discrete case, where $f_n(x_1, \dots, x_n)$ is the joint probability function.

Independence

The random variables $(x_1, x_2, ..., x_n)$ are *independent* if

$$F_n(x_1, x_2, ..., x_n) = F_1(x_1)F_2(x_2)...F_n(x_n)$$

or

$$f_n(x_1, x_2, ..., x_n) = f_1(x_1)f_2(x_2)...f_n(x_n).$$

Conditional Distribution

Let $f(x, y)$ be the joint probability function for (x, y). Then,

$g_1(x) = \int f(x, y)dy$ is the marginal probability function for x, and

$g_2 y) = \int f(x, y)dx$ is the marginal probability function for y.

- The conditional probability function of x given y is

$$h_1(x|y) = \frac{f(x, y)}{g_2(y)}$$

- and the conditional probability function of y given x is

$$h_2(y|x) = \frac{f(x, y)}{g_1(x)}$$

Bayes Theorem

$$h_2(y|x) = \frac{h_1(x|y)g_2(y)}{\int h_1(x|y)g_2(y)dy}$$

in the continuous case, and

$$h_2(y|x) = \frac{h_1(x|y)g_2(y)}{\sum_y h_1(x|y)g_2(y)}$$

in the discrete case.

In which x corresponds to *sample information*, $g_2(y)$ is the *prior probability*, $h_1(x|y)$ is the *likelihood function of the sample* and $h_2(y|x)$ is the *posterior probability*.

Proof of Bayes Theorem

In the continuous case:

$$h_2(y|x) = \frac{f(x,y)}{g_1(x)} = \frac{f(x,y)}{\int f(x,y)dy} = \frac{h_1(x|y)g_2(y)}{\int h_1(x|y)g_2(y)dy}$$

Example

A box contains 2,000 components, of which 5% are defective. A second box contains 500 components, of which 40% are defective. Two other boxes contain 1,000 components each, with 10% defective components. We select at random one of the above boxes and remove from it (at random) a single component. What is the probability that this component is defective?

Example (cont.)

We have 4,000 good (g) components and 500 defective (d) components arranged as follows:

Box 1	1,900g	100d
Box 2	300g	200d
Box 3	900g	100d
Box 4	900g	100d

Thus our probability space has 4,500 elements. We denote by the event consisting of all elements in the ith box and by $\mathcal{D}$ the event consisting all 500 defective elements. We want to determine $P(\mathcal{D})$. Clearly,

$$\mathcal{B}_1 + \mathcal{B}_2 + \mathcal{B}_3 + \mathcal{B}_4 = \mathcal{S} \qquad \mathcal{B}_i\mathcal{B}_j = 0 \quad i \neq j$$

By random selection of a box one means that

$$P(\mathcal{B}_1) = P(\mathcal{B}_2) = P(\mathcal{B}_3) = P(\mathcal{B}_4) = 0.25$$

Example (cont.)

Once a box is selected, the probability that the removed element is defective equals the ratio of the number of defective to the total number of elements in that box. In our language this means that

$$P(\mathcal{D}|\mathcal{B}_1) = \frac{100}{2,000} = 0.05$$

$$P(\mathcal{D}|\mathcal{B}_2) = \frac{100}{1,000} = 0.1$$

$$P(\mathcal{D}|\mathcal{B}_3) = \frac{200}{500} = 0.4$$

$$P(\mathcal{D}|\mathcal{B}_4) = \frac{100}{1,000} = 0.1$$

Example (cont.)

Based on the definition of conditional probability given earlier we have:

$$P(\mathcal{D}) = P(\mathcal{D}|\mathcal{B}_1)P(\mathcal{B}_1) + P(\mathcal{D}|\mathcal{B}_2)P(\mathcal{B}_2) + P(\mathcal{D}|\mathcal{B}_3)P(\mathcal{B}_3) + P(\mathcal{D}|\mathcal{B}_4)P(\mathcal{B}_4)$$

Inserting values for conditional probabilities, we obtain

$$P(\mathcal{D}) = 0.05 \times 14 + 0.4 \times +0.1 \times 44 + 0.1 \times Y = 0.1625$$

Bayesian Example

We examine the selected component in the previous example and we find it defective. What is the probability that it was taken from box 2?

We want the conditional probability $P(\mathcal{B}_2|\mathcal{D})$ of the event $\mathcal{B}_2$ (second box), assuming $\mathcal{D}$ (defective component). Since

$$P(\mathcal{D}) = 0.1625, \quad P(\mathcal{D}|\mathcal{B}_2) = 0.4, \quad P(\mathcal{B}_2) = 1/4$$

we obtain:

$$P(\mathcal{B}_2|\mathcal{D}) = \frac{0.4 \times 1/4}{0.1625} \approx 0.615$$

Thus the a posteriori (conditional) probability of having selected box 2, assuming that the selected part is defective, equals 0.615.

Frequency Interpretation

The frequency interpretation of the a posteriori probability $P(\mathcal{B}_2|\mathcal{D})$ is the following:

Repeat the experiment (selection of a box and removal of an article) n times. In $n_{\mathcal{D}}$ of these trials the selected part is defective. In the above $n_{\mathcal{D}}$ trials we count the number of times the selected part was taken from box 2. If this number is $n_{\mathcal{B}_2\mathcal{D}}$, then

$$P(\mathcal{B}_2|\mathcal{D}) \approx \frac{n_{\mathcal{B}_2\mathcal{D}}}{n_{\mathcal{D}}}$$

Expectations

The expected value of some function $r(x)$ is given by

$$E[r(x)] = \int r(x)f(x)dx,$$

in the continuous case, or

$$E[r(x)] = \sum_{x} r(x)f(x),$$

in the discrete case, where E is the "expectation operator."

The kth moment

Choose $r(x) = x^k, k = 1, 2, \ldots$ Then,

$$m_k = E(x^k)$$

is the k-th moment of x.

- If $k = 1$, then $m_1 = E(x)$, the mean (or average) of x, a common measure of the "central location" of x.
- If $k = 2$, then $m_2 = E(x^2)$, the second moment of x.
- If $k = 3$, then $m_3 = E(x^3)$, the third moment of x, ...

Central Moment

The k-th central moment: Choose $r(x) = (x - m_1)^k, \quad k = 2, 3, \ldots$ Then,

$$M_k = E[(x - m_1)^k]$$

is the k-th central moment of x.

- If $k = 2$, then $M_2 = E[(x - m_1)^2])$, the *variance* of x, a common measure of the "spread" or "dispersion" of x.
- If $k = 3$, then $M_3 = E[(x - m_1)^3]$, the third central moment of x
- ...

Other Statistics

variance

$$V(x) = E[(x - m_1)^2 = E[x^2 + m_1^2 - 2xm_1] = m_2 - m_1^2$$

standard deviation

$$\sqrt{M_2}$$

coefficient of variation

$$\frac{\sqrt{M_2}}{m_1}$$

relative kurtosis

$$\frac{M_4}{{M_2}^2}$$

relative skewness

$$\frac{M_3}{{M_2}^{3/2}}$$

Covariance

$$
\begin{aligned}
Cov(x,y) &\equiv E[(x-E(x))(y-E(y))] \\
&= E[xy - xE(y) - yE(x) + E(x)E(y)] \\
&= E(xy) - E(x)E(y)
\end{aligned}
$$

is related to the correlation:

$$
\rho(x,y) \equiv \frac{Cov(x,y)}{\sqrt{(M_2(x)M_2(y)}} \quad \Rightarrow \quad -1 \leq \rho \leq 1
$$

More Bayesian Inference

You need to estimate the probability of event z which depeonds on the state of nature. Assume there are three states of nature, and your prior probabilities of these states are $Pr(a_1) = .15$, $Pr(a_2) = .30$ and $Pr(a_3) = .55$. To gain further information, you consult an expert who gives you a forecast ($Pr(z)$) with conditional probabilities:

$$Pr(z|a_1) = .3 \quad Pr(z|a_2) = .50 \quad Pr(z|a_3) = .10$$

If you are a Bayesian learner, what probabilities do you want to use in your decision?

More Bayesian Inference (Solution)

The first step here is to calculate the probability of the forecast. This can be calculated from the priors for a_i and the conditionals:

$$Pr(z) = \sum_i Pr(z|a_i)Pr(a_i) = 0.25$$

Once we know $Pr(z)$ we can calculate posterior probabilities using Bayes' theorem:

$$Pr(a_i|z) = \frac{Pr(z|a_i)Pr(a_i)}{Pr(z)} = (0.18, 0.6, 0.22)$$

The revised estimate of the three states of nature would be impossible without Bayesian inference.

Let's Make a Deal

On the TV game show there are three curtains. Behind one of the curtains is a car, and behind the other two are goats. The game show host knows which curtain the car is behind. You are asked to pick a curtain, and will be given the prize behind it. Right after you pick, however, the host reveals one of the curtains you did not pick, which has a goat behind it (because he knows which curtain the car is behind, he won't accidentally reveal the car). The host then asks you whether you would like to consider switching to the remaining unopened curtain. The question is: Do you stay, do you switch, or does it even matter?

Let's Make a Deal (solution)

The probability of choosing the car is $1/3$. This probability is unchanged when a curtain is drawn to reveal a goat. A switch will always select the car provided that the car is not behind the curtain initial selected. For this reason, the probability of choosing the car after switching is $2/3$.

Let's Make a Deal (Excel)

Create an Excel spreadsheet with 100 simulated cases which confirms your analysis.

Let's Make a Deal (Excel)

Create an Excel spreadsheet with 100 simulated cases which confirms your analysis.
Let column G report the chosen curtain (randomly assigned from {1,2,3}).
Column H represents the curtain with the car (also randomly assigned).
We enter a 1 in column I if our choice is the curtain with the car and a zero otherwise (`=if(g2=h2,1,0)`). We enter a 1 in column J if we get the car by switching (`=if(g2<>h2,1,0)`. Generate 100 rows and compare `=sum(i2..i101)` with `=sum(j2..j101)`.

	F	G	H	I	J	K	L	M
1		Chosen	Car	V(noswitch)	V(switch)		Wins:	
2	1	1	1	1	0		Switch:	61
3	2	3	3	1	0		NoSwitch:	38
4	3	3	1	0	0			
5	4	3	1	0	1			
6	5	2	2	1	1			
7	6	3	1	0	0			
8	7	2	2	1	1			
9	8	2	1	0	0			
10	9	2	2	1	1			
11	10	2	2	1	0			
12	11	2	2	1	0			
13	12	1	2	0	0			

Pseudo Random Number Generation

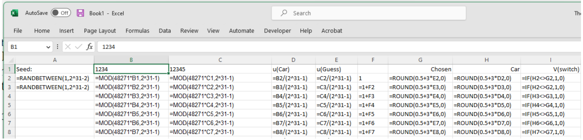
The Lehmer random number generator is a type of *linear congruential generator* (LCG) that operates in multiplicative group of integers modulo n. The general formula is

$$X_{k+1} = a \cdot X_k \mod m$$

in which m is a *prime number* or a *power of a prime number*, and multiplier a is an element of high *multiplicative order* modulo m, a *primitive root modulo n*, and the seed X_0 is *coprime* to m

Useful specification for use in Excel:

$$a = 48271. \quad \text{and} \quad m = 2^{31} - 1$$



	A	B	C	D	E	F	G	H	I
1	Seed:	1234	12345	u(Car)	u(Guess)		Chosen	Car	V(switch)
2	=RANDBETWEEN(1,2^31-2)	=MOD(48271*B1,2^31-1)	=MOD(48271*C1,2^31-1)	=B2/(2^31-1)	=C2/(2^31-1)	1	=ROUND(0.5+3*E2,0)	=ROUND(0.5+3*D2,0)	=IF(H2<>G2,1,0)
3	=RANDBETWEEN(1,2^31-2)	=MOD(48271*B2,2^31-1)	=MOD(48271*C2,2^31-1)	=B3/(2^31-1)	=C3/(2^31-1)	=1+F2	=ROUND(0.5+3*E3,0)	=ROUND(0.5+3*D3,0)	=IF(H2<>G2,1,0)
4		=MOD(48271*B3,2^31-1)	=MOD(48271*C3,2^31-1)	=B4/(2^31-1)	=C4/(2^31-1)	=1+F3	=ROUND(0.5+3*E4,0)	=ROUND(0.5+3*D4,0)	=IF(H3<>G3,1,0)
5		=MOD(48271*B4,2^31-1)	=MOD(48271*C4,2^31-1)	=B5/(2^31-1)	=C5/(2^31-1)	=1+F4	=ROUND(0.5+3*E5,0)	=ROUND(0.5+3*D5,0)	=IF(H4<>G4,1,0)
6		=MOD(48271*B5,2^31-1)	=MOD(48271*C5,2^31-1)	=B6/(2^31-1)	=C6/(2^31-1)	=1+F5	=ROUND(0.5+3*E6,0)	=ROUND(0.5+3*D6,0)	=IF(H5<>G5,1,0)
7		=MOD(48271*B6,2^31-1)	=MOD(48271*C6,2^31-1)	=B7/(2^31-1)	=C7/(2^31-1)	=1+F6	=ROUND(0.5+3*E7,0)	=ROUND(0.5+3*D7,0)	=IF(H6<>G6,1,0)
8		=MOD(48271*B7,2^31-1)	=MOD(48271*C7,2^31-1)	=B8/(2^31-1)	=C8/(2^31-1)	=1+F7	=ROUND(0.5+3*E8,0)	=ROUND(0.5+3*D8,0)	=IF(H7<>G7,1,0)

Subtleties in Conditioning

Consider the following two statements:

i. The Jones family has two children at least one of whom is a boy. What is the probability that the Jones family has two boys?

ii. You meet a boy at the park who is from the Smith family in which there are two children. What is the probability that the Smith family has two boys?

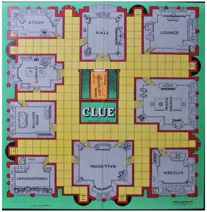
Explain why the answer to the first question is $1/3$ and the answer to the second question is $1/2$.

- The Jones family has two children at least one of whom is a boy. What is the probability that the Jones family has two boys?
  The set of equi-probable mutually exclusive outcomes is (BB, BG, GB, GG). "At least one boy" rules out (GG). The posterior probability of (BB) is then $1/3$.
  Here is another way to think about the problem. In the population of two-child families, 75% have at least one boy. In the population of children of two child families, 25% have two boys. Letting x denote the fraction of two-child families with at least one boy who have two boys, it follows that:
  $$x \times 75 = 25$$
  or $x = 1/3$.

ii. You meet a boy at the park who is from the Smith family in which there are two children. What is the probability that the Smith family has two boys?
We have the same set of equi-probable mutually exclusive outcomes is (BB, BG, GB, GG). "Meeting a boy" rules out (GB, GG). The posterior probability of (BB) is then $1/2$.

This word problem illustrates the subtlties involved in probabalistic statements. Language is important and needs to be precise. I first saw this problem in 1974, and I've remained suspicious about probability ever since. For more examples of the challenges presented by statistical reasoning, see *How to Lie with Statistics* by Darrell Huff and Irving Geis, W.W.Norton and Company, 1954.

Clue: The Board Game



Mr. Boddy is apparently the victim of foul play and is found in one of the rooms. The object of the game is to discover the answer to these three questions:

i. **Who**? Which one of the six suspects did it?

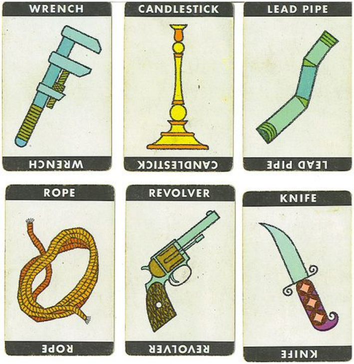
ii. **Where**? (i.e., in which of the nine rooms.)

iii. **What weapon?**

The answer lies in the little envelope resting on the stairway marked X in the center of the board. The envelope contains 3 cards. One card tells who did it- another card reveals the room in which it all happened, and the third card discloses the weapon used.¡p¿

Details of the Game

- Nine rooms, six weapons and six suspects = 21 cards.
- One room, one weapon, and one suspect are place in the envelope.
- Each of six players is randomly dealt six cards.



Rules of the Game

The player who, by the process of deduction and good plain common sense, first identifies the 3 solution cards hidden in the little envelope, wins the game. This is accomplished by players moving into the rooms and making "suggestions" of what they believe is the room, the person and the weapon for the purpose of gaining information. This information may reveal which cards are in other players' hands and which cards are missing and must, therefore, be hidden in the little envelope.

"Accusing" a suspect and naming the weapon and the room under suspicion is one of the most exciting features of this game.

You can gain more insight into the game by watching the movie on Netflix.

Simplified Game

Forget the board. Instead, deal the cards and take turns making suggestions. After one player makes a suggestion, each of the remaining players must either pass or reveal (to the player making the suggestion) one of the suggested cards (room, weapon or suspect).

Applying Bayes Theorem

1. Assuming that the cards are randomly dealt, what is your prior about the locations of the cards?
2. What is the size of the sample space defined in terms of elementary equi-probable events?
3. After examining the cards you have been dealt, how do you update your prior? What do you regard as the most likely contents of the envelope?
4. You observe a suggestion and refutation. How do you update your priors?