Math 425
Introduction to Probability
Lecture 9

Kenneth Harris
kaharri@umich.edu

Department of Mathematics
University of Michigan

January 28, 2009

Conditional probability

Definition (Conditioning Rule)

Let E and F be events.
The conditional probability of E given F is defined by:

$$\mathbf{P}(E \mid F) = \frac{\mathbf{P}(E \cap F)}{\mathbf{P}(F)}.$$

provided $\mathbf{P}(F) > 0$.

Note. From now on I will assume $\mathbf{P}(F) > 0$ when I write $\mathbf{P}(E \mid F)$.

Conjunctions

☞ We can often compute $\mathbf{P}(F)$ and $\mathbf{P}(E \mid F)$ easily.
If so, we can compute the conjunction $\mathbf{P}(E \cap F)$ as well.

Lemma (Multiplication Rule)

For any events E and F (where $\boldsymbol{P}(F) > 0$),

$$\boldsymbol{P}(E \cap F) = \boldsymbol{P}(E) \cdot \boldsymbol{P}(F \mid E)$$

Example

Example

A drawer contains 5 red socks and 3 blue socks.
If you remove the socks at random, what is the probability of holding a blue pair?

☞ You could do this by enumerating all the possible combinations in the sample space for this experiment, but conditioning provides a simpler method.

Example – continued

☞ 5 red and 3 blue socks.
Consider the two events

- B_i: the ith sock picked is blue ($i = 1, 2$).

We want to determine $\mathbf{P}(B_1 \cap B_2)$.
Use the multiplication rule,

$$\begin{aligned} \mathbf{P}(B_1 \cap B_2) &= \mathbf{P}(B_1) \cdot \mathbf{P}(B_2 \mid B_1) \\ &= \frac{3}{8} \cdot \frac{2}{7} = \frac{3}{28}. \end{aligned}$$

Reason: If you choose 1 blue sock then there are 2 blue socks and 7 socks remaining. So,

$$\mathbf{P}(B_2 \mid B_1) = \frac{2}{7}.$$

Partition Rule

☞ It is often easier to compute the probability of an event by dividing the sample space into two disjoint groups. (Compare to Ross, equation 3.3.1, p. 72).

Theorem (Partition Rule)

$$\mathbf{P}(E) = \mathbf{P}(E \mid F) \cdot \mathbf{P}(F) + \mathbf{P}(E \mid F^c) \cdot \mathbf{P}(F^c)$$

Proof. Since $E \cap F$ and $E \cap F^c$ are mutually disjoint

$$\begin{aligned} \mathbf{P}(E) &= \mathbf{P}(E \cap F) + \mathbf{P}(E \cap F^c) \\ &= \mathbf{P}(E \mid F) \cdot \mathbf{P}(F) + \mathbf{P}(E \mid F^c) \cdot \mathbf{P}(F^c) \end{aligned}$$

The last line is by the Conditioning Rule.

Example 1: tests

Example

In a population of individuals a proportion p are subject to a disease (such as AIDS). A test is available which indicates whether an individual has the disease (a positive result). No test is perfect though. Suppose the following:

- The probability of positive test when an individual has the disease is 95%, (So 5% of the time the test fails to indicate the disease – a false negative.)
- The probability of positive test when an individual does not have the disease is 5% – false positive.

What is the probability that a randomly selected individual is positive?

Example 1 – continued

Solution. Let P be the event that the result is positive and D the event the person has the disease. Then,

$$\mathbf{P}(D) = p \quad \mathbf{P}(P \mid D) = 0.95 \quad \mathbf{P}(P \mid D^c) = 0.05$$

Use the Partition Rule:

$$\begin{aligned} \mathbf{P}(P) &= \mathbf{P}(P \mid D) \cdot \mathbf{P}(D) + \mathbf{P}(P \mid D^c) \cdot \mathbf{P}(D^c) \\ &= 0.95p + 0.05(1 - p) \\ &= 0.9p + 0.05. \end{aligned}$$

☞ You can expect alot of positive results, most of which will be false positives, if the disease is rare (i.e., p is small).

Extended Partition Rule

☞ It is often more convenient to divide a sample space into several groups. (Compare to Ross, equation 3.3.4, p. 81).

Theorem (Extended Partition Rule)

Let E be some event and suppose $F_1, F_2, \ldots, F_n$ is a collection of mutually exclusive events, one of which must occur:

$$\bigcup_{k=1}^{n} F_i = S.$$

Then,

$$\boldsymbol{P}(E) = \sum_{k=1}^{n} \boldsymbol{P}(E \mid F_k) \cdot \boldsymbol{P}(F_k).$$

Extended Partition Rule

Proof.
We can break the event E into cases (since some one of the events F_k must occur):

$$E = \bigcup_{k=1}^{n} (E \cap F_k).$$

The events $E \cap F_k$ are mutually exclusive (since the events F_k are).
By the Sum Rule

$$\begin{aligned} \mathbf{P}(E) &= \sum_{k=1}^{n} \mathbf{P}(E \cap F_k) \\ &= \sum_{k=1}^{n} \mathbf{P}(E \mid F_k) \cdot \mathbf{P}(F_k) \end{aligned}$$

The last line is by the Multiplication Rule.

Example 2: coins

Example

You have 3 double-headed coins, 1 double-tailed coin and 5 normal coins. You select one coin at random and flip it.
What is the probability of heads?

Example 2 – continued

Solution. Let D, T, N be the event of choosing a double-headed, double-tailed and normal coin. Let H be the event that the coin shows heads.
We are given the following

$$\mathbf{P}(D) = \frac{1}{3} \quad \mathbf{P}(T) = \frac{1}{9} \quad \mathbf{P}(N) = \frac{5}{9}$$
$$\mathbf{P}(H \mid D) = 1 \quad \mathbf{P}(H \mid T) = 0 \quad \mathbf{P}(H \mid N) = \frac{1}{2}.$$

So,

$$\begin{aligned} \mathbf{P}(H) &= \mathbf{P}(H \mid D) \cdot \mathbf{P}(D) + \mathbf{P}(H \mid T) \cdot \mathbf{P}(T) + \mathbf{P}(H \mid N) \cdot \mathbf{P}(N) \\ &= 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{9} + \frac{1}{2} \cdot \frac{5}{9} \\ &= \frac{11}{18}. \end{aligned}$$

Converting conditional probabilities

☞ The following rule is the ♡ of Bayes' Rule.

Theorem (Rule for Converting Conditional Probabilities)

Let E and H be events.

Then

$$\mathbf{P}(H \mid E) = \frac{\mathbf{P}(E \mid H) \cdot \mathbf{P}(H)}{\mathbf{P}(E)}$$

Proof. Apply the Conditioning Rule twice

$$\mathbf{P}(H \mid E) = \frac{\mathbf{P}(E \cap H)}{\mathbf{P}(E)} = \frac{\mathbf{P}(E \mid H) \cdot \mathbf{P}(H)}{\mathbf{P}(E)}.$$

Bayes' Rule

☞ Bayes' Rule is one of the most important in all probability.

Theorem (Bayes' Rule)

$$\mathbf{P}(H \mid E) = \frac{\mathbf{P}(E \mid H) \cdot \mathbf{P}(H)}{\mathbf{P}(E \mid H) \cdot \mathbf{P}(H) + \mathbf{P}(E \mid H^c) \cdot \mathbf{P}(H^c)}$$

Proof. By the Partitioning Rule

$$\mathbf{P}(E) = \mathbf{P}(E \mid H) \cdot \mathbf{P}(H) + \mathbf{P}(E \mid H^c) \cdot \mathbf{P}(H^c)$$

Plug this into the Rule for Converting Conditional Probabilities

$$\mathbf{P}(H \mid E) = \frac{\mathbf{P}(E \mid H) \cdot \mathbf{P}(H)}{\mathbf{P}(E)}$$

Example 1 – false positives

Example

In a population of individuals a proportion p are subject to a disease:

- The probability of positive test when an individual has the disease is 95%, (The cases where the test fails to indicate the disease is called a false negative.)
- The probability of positive test when an individual does not have the disease is 5%. (This is called a false positive).

Should we implement universal (or random) testing?

That is, should we be confident a positive test result indicates the disease when we test everyone?

Example 1 – continued

Solution. Let D be the event of having a disease and R be the event of a positive result.

We are given

$$\mathbf{P}(D) = p \quad \mathbf{P}(R \mid D) = 0.95 \quad \mathbf{P}(R \mid D^c) = 0.05 \quad \mathbf{P}(R) = 0.9p + 0.05$$

(We computed the last in Example 1 of the previous section.)

☞ You want to know $\mathbf{P}(D \mid R)$. Use Bayes' Rule

$$\mathbf{P}(D \mid R) = \frac{\mathbf{P}(R \mid D)\mathbf{P}(D)}{\mathbf{P}(R)} = \frac{0.95p}{0.9p + 0.05}$$

Example 1 – continued

☞ D: disease, R: positive test result

$$\mathbf{P}(D \mid R) = \frac{0.95p}{0.9p + 0.05}$$

☞ If the disease is rare: $p = 0.006$ (about the rate of AIDS in the US). Then

$$\mathbf{P}(D \mid R) \approx 0.1$$

90% of positive cases will be false positives.

☞ If the disease is common: $p = 0.1$. Then

$$\mathbf{P}(D \mid R) \approx 0.69$$

Two-thirds of positives are now people with the disease.

Extended Bayes Rule

☞ The extended version of Bayes' Theorem just applies the Extended Partition Rule to the Rule for Converting Conditional Probabilities. (See Ross Proposition 3.1, p. 81.)

Theorem (Extended Bayes' Rule)

Let E be some event and suppose $H_1, H_2, \ldots, H_n$ is a collection of mutually exclusive events, one of which must occur:

$$\bigcup_{k=1}^{n} H_k = S.$$

Then,

$$\mathbf{P}(H_i \mid E) = \frac{\mathbf{P}(E \mid H_i) \cdot \mathbf{P}(H_i)}{\sum_{k=1}^{n} \mathbf{P}(E \mid H_k) \cdot \mathbf{P}(H_k)}.$$

Example 3: An urn problem

Example

An urn contains two balls, which have been randomly chosen to be either red or blue. We perform the following experiment to determine the composition of the urn.

1. Select a ball and record its color.
2. Replace the ball in the urn.
3. Mix the contents of the urn well.

Suppose we perform this experiment twice, and each time get a red ball.

What is the most likely composition of the urn?

Example 3 – continued

Solution. Prior to performing the experiment we have three hypotheses about the urn:

- D: the balls in the urn have different colors,
- R: both balls in the urn are red,
- B: both balls in the urn are blue.

Since the balls were randomly chosen to be placed in the urn, we have the following probabilities (prior to conducting the experiment):

$$\mathbf{P}(D) = \frac{1}{2} \qquad \mathbf{P}(R) = \mathbf{P}(B) = \frac{1}{4}.$$

☞ Our experiment produced the following outcome:

- R_2: Two red balls are drawn.

The likelihood of this event given each hypothesis

$$\mathbf{P}(R_2 \mid D) = \frac{1}{4} \qquad \mathbf{P}(R_2 \mid R) = 1 \qquad \mathbf{P}(R_2 \mid B) = 0$$

Example 3 – continued

☞ Use the Partition Rule:

$$\begin{aligned}\mathbf{P}(R_2) &= \mathbf{P}(R_2 \mid D)\mathbf{P}(D) + \mathbf{P}(R_2 \mid R)\mathbf{P}(R) + \mathbf{P}(R_2 \mid B)\mathbf{P}(B) \\ &= \frac{1}{4}\cdot\frac{1}{2} + 1\cdot\frac{1}{4} + 0\cdot\frac{1}{4} = \frac{3}{8}\end{aligned}$$

☞ Use Bayes' Rule to recompute the probability of each hypothesis:

$$\begin{aligned}\mathbf{P}(D \mid R_2) &= \frac{\mathbf{P}(R_2 \mid D)\cdot\mathbf{P}(D)}{\mathbf{P}(R_2)} = \frac{\frac{1}{8}}{\frac{3}{8}} = \frac{1}{3} \\ \mathbf{P}(R \mid R_2) = &= \frac{\mathbf{P}(R_2 \mid R)\cdot\mathbf{P}(R)}{\mathbf{P}(R_2)} = \frac{\frac{1}{4}}{\frac{3}{8}} = \frac{2}{3} \\ \mathbf{P}(B \mid R_2) &= \frac{\mathbf{P}(R_2 \mid B)\cdot\mathbf{P}(B)}{\mathbf{P}(R_2)} = \frac{0}{\frac{3}{8}} = 0\end{aligned}$$

☞ The most likely explanation is that the balls in the urn are both red.

Conditional probabilities are probabilities

☞ It is useful to know that conditional probabilities obey the probability axioms. (See Ross, Proposition 3.5.1, p. 102)

Theorem

Let F be any event with $\mathbf{P}(F) > 0$. Then, the function $\mathbf{P}(\cdot \mid F)$ on the event space S is a probability function.
That is, $\mathbf{P}(\cdot \mid F)$ satisfies the probability axioms.

1. $0 \leq \mathbf{P}(E \mid F) \leq 1$ *for all events E,*
2. $\mathbf{P}(S \mid F) = 1$,
3. *If E_1 and E_2 are mutually exclusive events, then*

$$\mathbf{P}(E_1 \cup E_2 \mid F) = \mathbf{P}(E_1 \mid F) + \mathbf{P}(E_2 \mid F)$$

Proof

For any event F with $\mathbf{P}(F) > 0$:
① Since $E \cap F \subseteq F$,

$$0 \leq \frac{\mathbf{P}(E \cap F)}{\mathbf{P}(F)} = \mathbf{P}(E \mid F) \leq 1.$$

②

$$\mathbf{P}(S \mid F) = \frac{\mathbf{P}(S \cap F)}{\mathbf{P}(F)} = \frac{\mathbf{P}(F)}{\mathbf{P}(F)} = 1$$

③ If E_1, E_2 are mutually exclusive, then so is $E_1 \cap F$ and $E_2 \cap F$.
By the distributive law

$$(E_1 \cup E_2) \cap F = (E_1 \cap F) \cup (E_2 \cap F)$$

So,

$$\begin{aligned}\mathbf{P}(E_1 \cup E_2 \mid F) &= \frac{\mathbf{P}(E_1 \cap F) + \mathbf{P}(E_2 \cap F)}{\mathbf{P}(F)} \\ &= \mathbf{P}(E_1 \mid F) + \mathbf{P}(E_1 \mid F)\end{aligned}$$

Example: Cesarian sections

Example

98% of all babies survive delivery. However, 15% of all deliveries involve Cesarian (C) sections, and then the survival rate drops to 96%.
If a randomly chosen pregnant woman does not have a C section, then what is the probability that the baby survives?

☞ Let B be the event that the baby survives and C the event that a C section is performed.
We want to compute $\mathbf{P}(B \mid C^c)$

Example – continued

☞ We are given the following data

$$\mathbf{P}(B) = 0.98 \quad \mathbf{P}(C) = 0.15 \quad \mathbf{P}(B \mid C) = 0.96.$$

☞ By converting conditional probabilities

$$\mathbf{P}(B \mid C^c) = \frac{\mathbf{P}(C^c \mid B) \cdot \mathbf{P}(B)}{\mathbf{P}(C^c)}$$

The right-side can be computed by converting again

$$\begin{aligned} \mathbf{P}(C \mid B) &= \frac{\mathbf{P}(B \mid C) \cdot \mathbf{P}(C)}{\mathbf{P}(B)} = \frac{0.96 \cdot 0.15}{0.98} = 0.1469 \\ \mathbf{P}(C^c \mid B) &= 1 - \mathbf{P}(C \mid B) = 0.8531 \end{aligned}$$

☞ The baby's probability of surviving when no C section is performed:

$$\mathbf{P}(B \mid C^c) = \frac{0.8531 \cdot 0.98}{0.85} = 0.9835.$$

Example: prostate cancer

Example

Prostate cancer is a common type of cancer found in men. One test for prostate cancer measures the level of a protein PSA (prostate specific antigen) produced only by the prostate gland. The test is notoriously unreliable though.

☞ The probability that a noncancerous man will have elevated PSA is 13.5%, with this probability increasing to 26.8% if the man does have cancer.

☞ Suppose a doctor believes that a certain patient has probability p of having prostrate cancer, before testing PSA levels.

(a) What is the probability of his having cancer if he has elevated PSA levels?

(b) What is the probability of his having cancer if he does not have elevated PSA levels?

Example – continued

☞ Let E be the event of having an elevated PSA and H be the hypothesis the patient has cancer.
We are given the following

$$\mathbf{P}(H) = p \quad \mathbf{P}(E \mid H) = 0.268 \quad \mathbf{P}(E \mid H^c) = 0.135.$$

We want to compute (a) $\mathbf{P}(H \mid E)$ and (b) $\mathbf{P}(H \mid E^c)$.

☞ Use Bayes' Rule to compute:

$$\begin{aligned} (a): \quad \mathbf{P}(H \mid E) &= \frac{\mathbf{P}(E \mid H) \cdot \mathbf{P}(H)}{\mathbf{P}(E)} \\ (b): \quad \mathbf{P}(H \mid E^c) &= \frac{\mathbf{P}(E^c \mid H) \cdot \mathbf{P}(H)}{\mathbf{P}(E^c)} \end{aligned}$$

We need only compute $\mathbf{P}(E)$. Here we use the Partition Rule:

$$\begin{aligned} \mathbf{P}(E) &= \mathbf{P}(E \mid H) \cdot \mathbf{P}(H) + \mathbf{P}(E \mid H^c) \cdot \mathbf{P}(H^c) \\ &= 0.268 \cdot p + 0.135 \cdot (1 - p) = 0.135 + 0.133p \\ \mathbf{P}(E^c) &= 1 - \mathbf{P}(E) = 0.865 - 0.133p \end{aligned}$$

Example – continued

$$\mathbf{P}(E) = 0.135 + 0.133p \qquad \mathbf{P}(E^c) = 0.865 - 0.133p$$

$$(a): \quad \mathbf{P}(H \mid E) = \frac{\mathbf{P}(E \mid H) \cdot \mathbf{P}(H)}{\mathbf{P}(E)} = \frac{0.268p}{0.135 + 0.133p}$$

$$\begin{aligned} (b): \quad \mathbf{P}(H \mid E^c) &= \frac{\mathbf{P}(E^c \mid H) \cdot \mathbf{P}(H)}{\mathbf{P}(E^c)} = \frac{(1 - \mathbf{P}(E \mid H)) \cdot \mathbf{P}(H)}{\mathbf{P}(E^c)} \\ &= \frac{0.732p}{0.865 - 0.133p} \end{aligned}$$

Example – completed

$$\mathbf{P}(H \mid E) = \frac{0.268p}{0.135 + 0.133p} \qquad \mathbf{P}(H \mid E^c) = \frac{0.732p}{0.865 - 0.133p}$$

- ☞ If the doctor is confident the patient has cancer: $p = 0.75$.

$$\begin{aligned} \mathbf{P}(H \mid E) &= 0.8562 \\ \mathbf{P}(H \mid E^c) &= 0.7174 \end{aligned}$$

- ☞ If the doctor is riding the fence: $p = 0.5$.

$$\begin{aligned} \mathbf{P}(H \mid E) &= 0.6650 \\ \mathbf{P}(H \mid E^c) &= 0.4584 \end{aligned}$$

- ☞ Testing PSA levels is not definitive on its own.