Math 425
Introduction to Probability
Lecture 11

Kenneth Harris
kaharri@umich.edu

Department of Mathematics
University of Michigan

February 2, 2009

Example: HIV

Example

The genome of the HIV virus consists of a chain of 10^4 nucleotides (one of four at each position: adenine, thymine cytosine, guanine). The probability of one link in the chain mutating is 3×10^{-5}. Each mutation in a position is independent of a mutation in each other position.

- What is the probability that the HIV genome is replicated without mutation?
- What is the probability that the replication of the HIV genome results in a chain that differs in exactly one position?

Example continued

☞ Let $u = 3 \times 10^{-5}$ (probability of mutation) and $L = 10^4$ (length of genome).

By independence, the probability that the HIV genome is replicated without mutation is

$$(1-u)^L \approx 0.74$$

☞ Again, by independence, the probability that the replication of the HIV genome occurs with exactly one mutation is

$$\binom{L}{1} \cdot u(1-u)^{L-1} \approx 0.22$$

Example continued

☞ Some mutations confer drug resistance, and some of these are as little as one step away.

If 10^9 newly infected cells are being produced every day, then any particular one-error mutant arises

$$u(1-u)^{L-1} \cdot 10^9 = 22,000 \text{ times}.$$

Question. What is the probability that every one-error mutation arises at least once each day?

This is the coupon problem later in the lecture.

Conditional probabilities are probabilities

☞ It is useful to know that conditional probabilities obey the probability axioms. (See Lecture 9 and 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*

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

We will write $\mathbf{P}_E(\cdot) = \mathbf{P}(\cdot \mid E)$ for the new probability function updated on the basis that E occurs.

Conditional probabilities are probabilities

☞ We can conditionalize on the new probability $\mathbf{P}_E(\cdot)$.
In terms of the old probability $\mathbf{P}(\cdot)$:

$$\begin{aligned}
\mathbf{P}_E(F \mid G) &= \frac{\mathbf{P}_E(F \cap G)}{\mathbf{P}_E(G)} \\
&= \frac{\frac{\mathbf{P}(E \cap F \cap G)}{\mathbf{P}(E)}}{\frac{\mathbf{P}(E \cap G)}{\mathbf{P}(E)}} \\
&= \frac{\mathbf{P}(E \cap F \cap G)}{\mathbf{P}(E \cap G)} \\
&= \mathbf{P}(F \mid E \cap G)
\end{aligned}$$

Conditional independence revisited

☞ Recall, that E and F are conditionally independent given G if

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

☞ If E and F are conditionally independent given G, then

$$\mathbf{P}_E(F \mid G) = \mathbf{P}(F \mid G)$$

since

$$\mathbf{P}_E(F \mid G) = \mathbf{P}(F \mid E \cap G) = \mathbf{P}(F \mid G)$$

Compare to Example 5f of Ross, p. 109.

Example: Urns

Example

An urn has red and blue balls, but I do not know the proportions. However, the proportion is known to be one of the following (where each is equally likely)

- A Three-quarters red,
- B One half red,
- C One-quarter red

A ball is selected and it is red. The ball is returned to the urn, and the urn mixed well.

What is the probability the next ball is red?

Example: Urns

☞ I have three hypotheses about the contents of the urn:

$$\mathbf{P}(A) = \mathbf{P}(B) = \mathbf{P}(C) = \frac{1}{3},$$

Let R be the event the ball drawn is red, then

$$\mathbf{P}(R \mid A) = \frac{3}{4} \quad \mathbf{P}(R \mid B) = \frac{1}{2} \quad \mathbf{P}(R \mid C) = \frac{1}{4}$$

☞ By the partition rule

$$\begin{aligned} \mathbf{P}(R_1) &= \mathbf{P}(R_1 \mid A) \cdot \mathbf{P}(A) + \mathbf{P}(R_1 \mid B) \cdot \mathbf{P}(B) + \mathbf{P}(R_1 \mid C) \cdot \mathbf{P}(C) \\ &= \frac{3}{4} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{4} \cdot \frac{1}{3} = \frac{1}{2}. \end{aligned}$$

Example: Urns

☞ Let R_1 be the event that the first ball drawn is red. Then, we need to update our three hypotheses:

$$\begin{aligned} \mathbf{P}(A \mid R_1) &= \mathbf{P}(R_1 \mid A) \cdot \frac{\mathbf{P}(A)}{\mathbf{P}(R_1)} \\ &= \frac{3}{4} \cdot \frac{2}{3} = \frac{1}{2} \\ \mathbf{P}(B \mid R_1) &= \mathbf{P}(R_1 \mid B) \cdot \frac{\mathbf{P}(B)}{\mathbf{P}(R_1)} \\ &= \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3} \\ \mathbf{P}(C \mid R_1) &= \mathbf{P}(R_1 \mid C) \cdot \frac{\mathbf{P}(C)}{\mathbf{P}(R_1)} \\ &= \frac{1}{4} \cdot \frac{2}{3} = \frac{1}{6} \end{aligned}$$

Example: Urns

☞ Since $\mathbf{P}(\cdot \mid R_1)$ is a probability, we write $\mathbf{P}_{R_1}(\cdot) = \mathbf{P}(\cdot \mid R_1)$.
The new probabilities are:

$$\mathbf{P}_{R_1}(A) = \frac{1}{2} \qquad \mathbf{P}_{R_1}(B) = \frac{1}{3} \qquad \mathbf{P}_{R_1}(C) = \frac{1}{6}.$$

☞ Let R_2 be the event that the second ball drawn is red.
R_2 is conditionally independent of R_1 given any of A, B, C.

$$\mathbf{P}_{R_1}(R_2 \mid A) = \mathbf{P}(R_2 \mid A) = \frac{3}{4} \qquad \mathbf{P}_{R_1}(R_2 \mid B) = \frac{1}{2} \qquad \mathbf{P}_{R_1}(R_2 \mid C) = \frac{1}{4}.$$

☞ By the Partition Rule

$$\begin{aligned} \mathbf{P}_{R_1}(R_2) &= \mathbf{P}_{R_1}(R_2 \mid A) \cdot \mathbf{P}_{R_1}(A) + \mathbf{P}_{R_1}(R_2 \mid B) \cdot \mathbf{P}_{R_1}(B) + \mathbf{P}_{R_1}(R_2 \mid C) \cdot \mathbf{P}_{R_1}(C) \\ &= \frac{3}{4} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{4} \cdot \frac{1}{6} = \frac{7}{12}. \end{aligned}$$

Example: genetics

Example

A gene responsible for a rare condition R has two alleles, A and a. All people have a gene pair, one of which they receive from their mother and the other from their father. Only people receiving an a from each parent have the condition R. There are four possible gene pairs

$$AA \quad Aa \quad aA \quad aa$$

A child receives independently one gene from each parent, and either gene pair in the parent is equally likely to be transmitted to the child.

☞ Neither parent has the condition R, but both carry the a gene. They have a child without the condition R.

- What is the probability that their child carries the a gene?

Example – continued

☞ We have the following information from the problem:

- Both parents are *Aa* or *aA*.
- Their male child could be any of *AA*, *Aa* or *aA*.

☞ Let $H_{AA}, H_{Aa}, H_{aA}, H_{aa}$ be the hypotheses about the genetic make-up of the child. Then, before we have any information about whether the child has condition R,

$$\mathbf{P}(H_{AA}) = \mathbf{P}(H_{aa}) = \mathbf{P}(H_{Aa}) = \mathbf{P}(H_{aA}) = \frac{1}{4},$$

and

$$\mathbf{P}(R^c \mid H_{AA}) = \mathbf{P}(R^c \mid H_{aA}) = \mathbf{P}(R^c \mid H_{Aa}) = 1 \qquad \mathbf{P}(R^c \mid H_{aa}) = 0.$$

We also know

$$\mathbf{P}(H_{aa} \mid R^c) = 0 \qquad \mathbf{P}(H_{aA} \mid R^c) = \mathbf{P}(H_{Aa} \mid R^c)$$

Example – continued

☞ What is the probability that the child carries an *a* gene, H_a, given it does not have condition R:

$$\begin{aligned}\mathbf{P}(H_a \mid R^c) &= 2 \cdot \mathbf{P}(H_{Aa} \mid R^c) \\ &= \frac{2 \cdot \mathbf{P}(R^c \mid H_{Aa}) \cdot \mathbf{P}(H_{Aa})}{\mathbf{P}(R^c)}\end{aligned}$$

We can compute the right-side except for $\mathbf{P}(R^c)$:

$$\begin{aligned}\mathbf{P}(R^c) &= \mathbf{P}(R^c \mid H_{AA}) \cdot \mathbf{P}(H_{AA}) + 2 \cdot \mathbf{P}(R^c \mid H_{Aa}) \cdot \mathbf{P}(H_{Aa}) \\ &\quad + \mathbf{P}(R^c \mid H_{aa}) \cdot \mathbf{P}(H_{aa}) \\ &= 1 \cdot \frac{1}{4} + 2 \cdot 1 \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} = \frac{3}{4}\end{aligned}$$

So

$$\mathbf{P}(H_a \mid R^c) = \frac{\frac{1}{2}}{\frac{3}{4}} = \frac{2}{3}$$

Example – continued

☞ Their child (1) mates with another carrier (2) of the *a* gene.
What is the probability that their first child has the condition R?

☞ Let C_1 be the event that the child has condition R.

$$\begin{aligned}\mathbf{P}(C_1) &= \mathbf{P}(1 \text{ gives } a) \cdot \mathbf{P}(2 \text{ gives } a) \\ &= \mathbf{P}(1 \text{ gives } a \mid H_a) \cdot \mathbf{P}(H_a) \cdot \mathbf{P}(2 \text{ gives } a) \\ &= (\frac{1}{2} \cdot \frac{2}{3}) \cdot \frac{1}{2} = \frac{1}{6}\end{aligned}$$

☞ So, the probability that the child has the condition R is $\frac{1}{6}$.

Example – continued

☞ The probability now that (1) has the *a* gene given that C_1^c, its first child does not have condition R:

$$\begin{aligned}\mathbf{P}(H_a \mid C_1^c) &= \frac{\mathbf{P}(C_1^c \mid H_a) \cdot \mathbf{P}(H_a)}{\mathbf{P}(C_1^c)} \\ &= \frac{\frac{3}{4} \cdot \frac{2}{3}}{\frac{5}{6}} \\ &= \frac{3}{5}.\end{aligned}$$

Reason: The child has condition R if it is *aa*, which requires receiving one *a* from each parent (which are both *Aa* or *aA*):

$$\mathbf{P}(C_1^c \mid H_a) = 1 - \mathbf{P}(C_1 \mid H_a) = 1 - \frac{1}{4} = \frac{3}{4}$$

Example – continued

☞ Let $\mathbf{P}_{C_1^c}(\cdot) = \mathbf{P}(\cdot \mid C_1^c)$, the new probability function given the new information that the first child does not have condition R.

☞ We have also updated the likelihood that mate (1) carries the a gene in the light of this new information:

$$\mathbf{P}_{C_1^c}(H_a) = \frac{3}{5} < \mathbf{P}(H_a) = \frac{2}{3}$$

Example – continued

☞ The couple have a second child. What is the probability that it has condition R given its sibling did not have the condition?

☞ Let C_2 be the event that the second child has condition R.

$$\begin{aligned}
\mathbf{P}_{C_1^c}(C_2) &= \mathbf{P}_{C_1^c}(1 \text{ gives } C_2\ a) \cdot \mathbf{P}_{C_1^c}(2 \text{ gives } C_2\ a) \\
&= \mathbf{P}_{C_1^c}(1 \text{ gives } C_2\ a \mid H_a) \cdot \mathbf{P}_{C_1^c}(H_a) \cdot \mathbf{P}_{C_1^c}(2 \text{ gives } C_2\ a) \\
&= \left(\frac{1}{2} \cdot \frac{3}{5}\right) \cdot \frac{1}{2} = \frac{3}{20}.
\end{aligned}$$

Reason: The events that the first and second child receive an a gene from (1) are independent given that (1) has the a gene. So,

$$\mathbf{P}_{C_1^c}(1 \text{ gives } C_2\ a \mid H_a) = \mathbf{P}(1 \text{ gives } C_2\ a \mid H_a) = \frac{1}{2}$$

Example: Coupon problem

Example

What is the probability that each of the values of a fair die turn-up in n throws of the die?

What is the average number of throws required for all six values to appear on at least one throw? Sorry, you'll have to wait for next chapter for this question.

Generalization

Example

An urn contains c balls, each of a different color. A ball is drawn, its color noted, returned to the urn, and the urn is thoroughly mixed.

What is the probability p that every color turns up at least once when the experiment is repeated n times?

☞ Compare to Example 4i of Ross, p. 93. The probabilities here are uniform – each "coupon" (i.e. ball, here) is equally likely to be drawn.

Example continued

☞ Let A_i $(1 \leq i \leq c)$ be the event that the ith color has not been picked.
By the Inclusion-Exclusion Rule

$$\begin{aligned} p &= 1 - \mathbf{P}(A_1 \cup A_2 \cup \ldots \cup A_c) \\ &= 1 - \sum_{j=1}^{c} \mathbf{P}(A_j) + \sum_{j<k} \mathbf{P}(A_j \cap A_k) - \\ &\quad \ldots + (-1)^c \mathbf{P}(A_1 \cap A_2 \cap \ldots \cap A_c) \end{aligned}$$

☞ Since each color is equally likely to be picked,

$$\mathbf{P}(A_i) = \mathbf{P}(A_j) \qquad \mathbf{P}(A_i \cap A_j) = \mathbf{P}(A_k \cap A_\ell) \qquad \text{etc.}$$

So,

$$p = 1 - \binom{c}{1} \cdot \mathbf{P}(A_1) + \binom{c}{2} \cdot \mathbf{P}(A_1 \cap A_2) - \ldots + (-1)^c \binom{c}{c} \cdot \mathbf{P}(\bigcap_i A_i).$$

Example continued

☞ By the independence of the picks,

$$\mathbf{P}(A_1 \cap A_2 \cap \ldots \cap A_k) = (1 - \frac{k}{c})^n$$

This is the probability that NONE of these k colors are picked in n draws.

☞ So,

$$\begin{aligned} p &= 1 - \binom{c}{1}(1 - \frac{1}{c})^n + \binom{c}{2}(1 - \frac{2}{c})^n + \ldots + (-1)^c \binom{c}{c}(1 - \frac{c-1}{c})^n \\ &= \sum_{k=0}^{c-1} (-1)^k \binom{c}{k} (1 - \frac{k}{c})^n \end{aligned}$$

Die example revisited

☞ The probability p of throwing at least each value on a fair die in n throws is

$$p = \sum_{k=0}^{5} (-1)^k \binom{6}{k} (1 - \frac{k}{6})^n$$

tosses	p
6	0.0154
12	0.4378
13	0.5139
15	0.6442
18	0.7847
24	0.9254