UNIVERSITY OF MASSACHUSETTS
Department of Biostatistics and Epidemiology
BioEpi 540W - Introduction to Biostatistics
Fall 2004

Exercises with Solutions– Topic 2
Introduction to Probability

Due: Monday October 18, 2004

READINGS

1. Study Guide (Rosner B. Study Guide for Fundamentals of Biostatistics, 5th Edition) Chapter 3, pp 13-18.
2. **Especially pp 17-18. This is a short introduction to ROC Curves.**
3. Text (Rosner B. Fundamentals of Biostatistics, 5th Edition) Chapter 3, pp 45-66.

EXERCISES:

1. Let A and B be two events such that P(A)=0.6 and P(B)=0.6. Which of the following MUST be true ?

A) A and B are independent
B) A and B are mutually exclusive
C) A and B are not mutually exclusive
D) P(A and B)=0.2
E) P(A or B) = 1.2

2. Let A and B be two events such that P(A)=0.6 and P(A|B)=0.6. Which of the following is TRUE ?

A) A and B are independent
B) A and B are mutually exclusive
C) P(A or B) = 0.36
D) P(A and B) = 0.36
E) P(A or B) = 1.2

3. We choose 367 people at random. What is the probability that at least two of them share the same birthday (month and day) ? (Note : Assume that all birthdays are equally likely and all years have 365 days).

A) 0.03
B) 0.11
C) 0.24
D) 0.57
E) 0.97
F) 1.0

4. We choose 5 people at random. What is the probability that at least two of them share the same birthday (month and day)? (Note : Assume that all birthdays are equally likely and all years have 365 days).

A) 0.03
B) 0.11
C) 0.24
D) 0.57
E) 0.97
F) 1.0

5. (Source: http://www.math.uah.edu/stat/prob/prob6.html#Hemophilia
note 9-23-04: I'm not sure this link is still working – cb.)

The common form of hemophilia is due to a defect on the X chromosome (one of the two chromosomes that determine gender). We will let h denote the defective gene, linked to hemophilia, and H the corresponding normal gene. Women have two X chromosomes, and h is recessive. Thus, a woman with gene type HH is normal; a woman with gene type hH or Hh is free of the disease, but is a carrier; and a woman with gene type hh has the disease. A man has only one X chromosome (his other sex chromosome, the Y chromosome, plays no role in the disease). A man with gene type h has hemophilia, and a man with gene type H is healthy. The following exercises explore the transmission of the disease.

Suppose that a woman initially has a 50% chance of being a carrier.

5A) Given that she has 2 healthy sons, compute the conditional probability that that she is a carrier.

5B) Compute the conditional probability that the third son will be healthy.

Note – Exercise 5 is hard. Try it, play with it, talk about it, but don't anxt over it. If you are not able to solve it, that's okay!

SOLUTIONS:

1. Let A and B be two events such that P(A)=0.6 and P(B)=0.6. Which of the following MUST be true?

   Answer: C – A and B are not mutually exclusive.

   The definition of a probability distribution says that the sum of the elementary sample points must be 1. If the events "A" and "B" have a total probability that is greater than 1, then they must contain some elementary sample points in common. Therefore, they cannot be mutually exclusive.

---

2. Let A and B be two events such that P(A)=0.6 and P(A|B)=0.6. Which of the following is TRUE ?

   Answer: A – A and B are independent.

   That both P(A) = 0.6 and P(A|B) = 0.6 says that the likelihood of the event A occuring is not altered in any way by the occurrence of event B. So A and B must be independent.

---

3. We choose 367 people at random. What is the probability that at least two of them share the same birthday (month and day) ? (Note : Assume that all birthdays are equally likely and all years have 365 days).

   Answer: F – 1.0

   Since it is not possible that all 367 persons can have a unique birthday, for the reason that a year contains 365 days, it must be that at least one person shares a birthday with someone else.

---

4. We choose 5 people at random. What is the probability that at least two of them share the same birthday (month and day) ? (Note : Assume that all birthdays are equally likely and all years have 365 days).

Answer: A – 0.03

P[at least 2 share the same birthday]

= 1 - P[all 5 people have unique birthdays]

=
1 - { P[2nd is unique|1st] P[3rd is unique|1st, 2nd] P[4th is unique|1st, 2nd, 3rd] P[5th is unique|1st, 2nd, 3

$$= 1 - \left\{ \left[\frac{364}{365}\right]\left[\frac{363}{365}\right]\left[\frac{362}{365}\right]\left[\frac{361}{365}\right]\right\}$$

= 0.03

5A. Given that she has 2 healthy sons, compute the conditional probability that that she is a carrier.

Answer: 0.33

Terminology: genotype= the gene type of a person
Terminology: phenotype= the expressed characteristic (such as disease) resulting from a genotype

Let us define the following events:

Event	Description
W^+	Woman is a carrier (hH or Hh)
W^{HH}	Woman is free of the carrier gene (HH)
W^{hh}	Woman is Diseased (hh)

Event	Description
A^+	1st Son is healthy
$A^{-\,-}$	1st Son is not healthy

Similarly, we define events for the second son as B^+ and B^-, and for the third son as the events C^+ and C^-.

We wish to find the probability that a woman is a carrier, given she has had two healthy sons. This probability is represented by $P\left(W^+ \mid A^+B^+\right)$. To find this probability, we use Bayes Rule. Thus,

$$P\left(W^+ \mid A^+B^+\right) = \frac{P\left(A^+B^+ \mid W^+\right)P\left(W^+\right)}{P\left(A^+B^+ \mid W^+\right)P\left(W^+\right) + P\left(A^+B^+ \mid W^{HH}\right)P\left(W^{HH}\right) + P\left(A^+B^+ \mid W^{hh}\right)P\left(W^{hh}\right)}.$$

We know that $P\left(W^+\right) = 0.5$, $P\left(W^{hh}\right) = 0.25$, and . $P\left(W^{HH}\right) = 0.25$. Notice that when a woman is not carrier, she could be either diseased (hh) or free of the carrier gene (HH).

We assume that the probability that a son is healthy is independent of the probability any other son is healthy from the same mother. Independence implies that the probability that a woman has a healthy son remains the same, regardless of the health of her previous sons. As a result, we can write:

$$P\left(A^+B^+ \mid W^+\right) = P\left(A^+ \mid W^+\right)\ P\left(B^+ \mid W^+\right)$$

and

$$P\left(A^+B^+ \mid W^{hh}\right) = P\left(A^+ \mid W^{hh}\right)\ P\left(B^+ \mid W^{hh}\right)$$

and

$$P\left(A^+B^+ \mid W^{HH}\right) = P\left(A^+ \mid W^{HH}\right)\ P\left(B^+ \mid W^{HH}\right).$$

Now $P(A^+|W^+)=0.5$, $P(A^+|W^{hh})=0$, and $P(A^+|W^{HH})=1.0$.

As a result, $P\left(A^+B^+ \mid W^+\right) = 0.25$, $P\left(A^+B^+ \mid W^{hh}\right) = 0$, and $P\left(A^+B^+ \mid W^{HH}\right) = 1$.

Using these expressions,

$$P\left(W^{+} \mid A^{+}B^{+}\right)=\frac{P\left(A^{+}B^{+} \mid W^{+}\right)P\left(W^{+}\right)}{P\left(A^{+}B^{+} \mid W^{+}\right)P\left(W^{+}\right)+P\left(A^{+}B^{+} \mid W^{HH}\right)P\left(W^{HH}\right)+P\left(A^{+}B^{+} \mid W^{hh}\right)P\left(W^{hh}\right)}$$

$$=\frac{(0.25)(0.5)}{(0.25)(0.5)+(1)(0.25)+(0)(0.25)}$$

$$=\frac{1}{3}$$

As a result, the probability that a woman is a carrier, given she has had two healthy sons is 0.333, or 33.3%.

5B) Compute the conditional probability that the third son will be healthy.

Answer: 5/6 = 0.83

Now we want to know the probability that a third son is healthy, given the previous two are healthy. This probability is equal to the following:

$$P\left(A^{+}B^{+}C^{+} \mid A^{+}B^{+}\right)=\frac{P\left(A^{+}B^{+}C^{+}\right)}{P\left(A^{+}B^{+}\right)}.$$

We evaluate the bottom term, and then the top term. To evaluate the bottom term, we use the law of total probability such that

$$P\left(A^{+}B^{+}\right)=P\left(A^{+}B^{+} \mid W^{+}\right)\ P\left(W^{+}\right)+P\left(A^{+}B^{+} \mid W^{hh}\right)\ P\left(W^{hh}\right)+P\left(A^{+}B^{+} \mid W^{HH}\right)\ P\left(W^{HH}\right).$$

Previously, we determined expressions for each of these terms. As a result,

$$P\left(A^{+}B^{+}\right)=(0.25)\ (0.5)+0\ (0.25)\ +1\ (0.25)=0.375$$

Similarly,

$$\begin{aligned}P\left(A^{+}B^{+}C^{+}\right)&=P\left(A^{+}B^{+}C^{+} \mid W^{+}\right)\ P\left(W^{+}\right)\\&+P\left(A^{+}B^{+}C^{+} \mid W^{hh}\right)\ P\left(W^{hh}\right)\\&+P\left(A^{+}B^{+}C^{+} \mid W^{HH}\right)\ P\left(W^{HH}\right)\end{aligned}$$

Now $P\left(A^{+}B^{+}C^{+} \mid W^{+}\right)=P\left(A^{+} \mid W^{+}\right)P\left(B^{+} \mid W^{+}\right)P\left(C^{+} \mid W^{+}\right)=(0.5)^{3}=0.125$ since sons are independent. Similarly, $P\left(A^{+}B^{+}C^{+} \mid W^{hh}\right)=0$ and $P\left(A^{+}B^{+}C^{+} \mid W^{HH}\right)=1$.

Thus, $P\left(A^{+}B^{+}C^{+}\right)=(0.125)\ (0.5)+0\ (0.25)\ +1\ (0.25)=0.3125$.

As a result,

$$
\begin{aligned}
P\left(A^{+}B^{+}C^{+} \mid A^{+}B^{+}\right) &= \frac{P\left(A^{+}B^{+}C^{+}\right)}{P\left(A^{+}B^{+}\right)} \\
&= \frac{0.3125}{0.375} \\
&= \frac{5}{6}
\end{aligned}
$$