# Chapter 4: Statistics with R - 2nd Edition 

Robert Stinerock

Student Exercises

1. A two-step experiment consists of flipping a coin (on the first step) and casting a die (on the second). Write out the sample space. How many outcomes are possible?

Answer: Since the experiment has $N=2$ steps with $n_{1}=2$ results on the first step and $n_{2}=6$ on the second, there would be

$$
\left(n_{1}\right)\left(n_{2}\right)=(2)(6)=12
$$

outcomes for the overall experiment. The sample space is listed below.
$S=\{(H, 1),(H, 2),(H, 3),(H, 4),(H, 5),(H, 6),(T, 1),(T, 2),(T, 3),(T, 4),(T, 5),(T, 6)\}$

```
# Use the * multiplication operator.
2 * 6
## [1] 12
```

2. For a popular version of a card game, 7 cards are dealt from the deck of 52 . How many different hands of 7 cards can be drawn from 52 ?

Answer: Using the Counting Rule for Combinations, and $n=7$ from $N=52$, there are $133,784,560$ unique hands of 7 cards from a deck of 52 .

Answer:

```
\[
\binom{52}{7}=\frac{52!}{7!(52-7)!}=\frac{52!}{7!45!}=133,784,560
\]
# Use the choose(N, n) function to find number of
# combinations of N items taken n at a time.
choose(52, 7)
## [1] 133784560
```

3. The process of drawing cards from a deck can be thought of as sampling without replacement because once a card has been drawn (and added to a player's hand) it is not replaced in the deck and thus cannot be drawn again. Suppose a new computer-based card game is designed which relaxes this restriction, and allows sampling with replacement, where each card drawn is then returned to the original deck from which it may be drawn again. How many unique seven-card hands are possible?

Answer: Applying the Counting Rule for the Multiple-Step Experiment, there are $N=7$ steps with $n=52$ possible results on each step: $\left(n_{1}\right)=\left(n_{2}\right)=\ldots=\left(n_{7}\right)=52$. The number of outcomes for the overall experiment is $1,028,071,702,528$ possible unique hands. Note: R reports very large (and very small) numbers using scientific notation: $1.028072 e+12$ is actually $1.028072(10)^{12}$ (see 1 ). To learn how to express this value in non-scientific notation, see 2 .

```
# (1) Use ~ operator to raise a number to a power.
52 - 7
## [1] 1.028072e+12
# (2) To suppress an answer reported in scientific notation,
# include: options(scipen = 999)
options(scipen = 999)
52 - 7
## [1] 1028071702528
```

4. The number of outcomes when sampling with replacement exceeds the number of outcomes when sampling without replacement. Why?

Answer: The reason why there are so many more possible experimental outcomes in the game with sampling with replacement than with the game with sampling without replacement is that there are many more possible results on each of the seven steps of the multiple-step experiment. If we compare the first three steps in both games, this becomes clear. For the first game, after three steps, there are $(52)(52)(52)=140,608$ possible results whereas for the second game, there are only $(52)(51)(50)=132,600$. As more cards are dealt - the deeper into the experiment we go - the greater this divergence in the number of possible outcomes.
5. For each problem below, how many ways can $n$ objects can be selected from $N$ ?
(a) $n=2$ from $N=11$

Answer:

```
        (c}\begin{array}{c}{11}\\{2}\end{array})=\frac{11!}{2!(11-2)!}=\frac{11!}{2!9!}=5
# Use the choose(N, n) function to find number of
# combinations of N items taken n at a time.
choose(11, 2)
## [1] 55
```

(b) $n=5$ from $N=10$

Answer:

$$
\binom{10}{5}=\frac{10!}{5!(10-5)!}=\frac{10!}{5!5!}=252
$$

choose (10, 5)
\#\# [1] 252
(c) $n=19$ from $N=20$

Answer:

$$
\binom{20}{19}=\frac{20!}{19!(20-19)!}=\frac{20!}{19!1!}=20
$$

choose (20, 19)
\#\# [1] 20
(d) $n=99$ from $N=100$

Answer:

$$
\binom{100}{99}=\frac{100!}{99!(100-99)!}=\frac{100!}{99!1!}=100
$$

choose(100, 99)
\#\# [1] 100
(e) $n=0$ from $N=10$

Answer:

$$
\binom{10}{0}=\frac{10!}{0!(10-0)!}=\frac{10!}{0!10!}=1
$$

choose (10, 0)
\#\# [1] 1
6. Find the number of permutations of $n$ objects selected from $N$ objects.
(a) $n=2$ from $N=11$

Answer:

$$
P_{n}^{N}=\frac{N!}{(N-n)!}=\frac{11!}{(11-2)!}=\frac{11!}{9!}=110
$$

```
# Use the ratio of factorials to find the number of
# permutations of N items taken n at a time.
factorial(11) / factorial(11 - 2)
## [1] 110
```

(b) $n=5$ from $N=10$

Answer:

$$
P_{n}^{N}=\frac{N!}{(N-n)!}=\frac{10!}{(10-5)!}=\frac{10!}{5!}=30,240
$$

factorial(10) / factorial(10 - 5)
\#\# [1] 30240
(c) $n=19$ from $N=20$

Answer:

$$
P_{n}^{N}=\frac{N!}{(N-n)!}=\frac{20!}{(20-19)!}=\frac{20!}{1!} \approx(2.433)(10)^{18}
$$

factorial(20) / factorial(20 - 19)
\#\# [1] 2432902008176640000
(d) $n=1$ from $N=20$

Answer:

$$
P_{n}^{N}=\frac{N!}{(N-n)!}=\frac{20!}{(20-1)!}=\frac{20!}{19!}=20
$$

factorial(20) / factorial(20 - 1)
\#\# [1] 20
(e) $n=0$ from $N=10$

Answer:

$$
P_{n}^{N}=\frac{N!}{(N-n)!}=\frac{10!}{(10-0)!}=\frac{10!}{10!}=1
$$

```
factorial(10) / factorial(10 - 0)
## [1] 1
```

7. A simple random sample (SRS) of size $n$ drawn from a finite population of size $N$ is a sample that has been chosen such that all possible samples of size $n$ are equallylikely. If one uses an SRS procedure to draw a sample of size $n$ from $N$, what is the probability of each sample?

Answer: If one uses an SRS procedure to select a sample of size $n$ from a finite population of size $N$, then the probability of any given sample is

$$
\frac{1}{\binom{N}{n}}
$$

8. Suppose we select a SRS of $n=2$ balls from an urn containing a population of $N=5$ balls (painted with the letters A, B, C, D, and E). List the sample space of all possible samples or experimental outcomes.

Answer: The sample space $S$ of all possible samples of size $n=2$ is

$$
S=\{(A B),(A C),(A D),(A E),(B C),(B D),(B E),(C D),(C E),(D E)\}
$$

9. Referring to the previous exercise, what is the probability of each possible sample?

Since there are 10 possible samples, and since SRS requires that each be equallylikely, the probability of any single SRS must be $1 / 10$.

$$
\begin{aligned}
& p(A B)=p(A C)=p(A D)=p(A E)=p(B C)=p(B D) \\
& \quad=p(B E)=p(C D)=p(C E)=p(D E)=0.10
\end{aligned}
$$

Second, applying the rule from above

$$
=\frac{1}{\binom{N}{n}}=\frac{1}{\binom{5}{2}}=\frac{1}{10}=0.10
$$

```
1 / choose(5, 2)
## [1] 0.1
```

10. Although this rule (used in the preceding two exercises) provides the probability for any SRS of size $n$ selected from a finite population of size $N$, does it apply when drawing samples from an infinite population?

Answer: The rule does not apply because it is not possible to specify the size of the population $N$ in advance. In fact, the expression could not even be calculated since there is no value for $N$. This happens when, for example, a sample of items is being selected as they come off an assembly line, or when a sample of customers is being chosen as they depart from a big-box retailer.
11. Suppose we select a SRS of $n=3$ balls from an urn containing a population of $N=6$ balls (painted with the numbers $1,2,3,4,5$, and 6 ). List the sample space $S$ of all possible outcomes.

Answer:

$$
\begin{aligned}
& S=\{(1,2,3),(1,2,4),(1,2,5),(1,2,6),(1,3,4),(1,3,5),(1,3,6),(1,4,5),(1,4,6),(1,5,6), \\
& (2,3,4),(2,3,5),(2,3,6),(2,4,5),(2,4,6),(2,5,6),(3,4,5),(3,4,6),(3,5,6),(4,5,6)\}
\end{aligned}
$$

12. Referring to the above exercise, how many sample points are there? Use combinations.

Answer:

$$
C_{3}^{6}=\binom{6}{3}=\frac{6!}{3!(6-3)!}=\frac{6!}{(3!)(3!)}=\frac{720}{(6)(6)}=20
$$

The counting rule confirms that there are 20 possible simple random samples.

```
choose(6, 3)
## [1] 20
```

13. Again referring to the above exercise, how many permutations of 3 items can be taken from 6 ?

Answer: The number of permutations of 3 items taken from 6 is

$$
P_{n}^{N}=\frac{N!}{(N-n)!}=\frac{6!}{(6-3)!}=\frac{6!}{3!}=120
$$

```
factorial(6) / factorial(6 - 3)
```

\#\# [1] 120
14. Referring once more to the above exercise, it is clearly impractical (and unnecessary) to list all 120 permutations. It is, however, possible to gain a bit more insight into this issue. List each permutation of the first triple in the answer to above question: $(1,2,3)$.

Answer: There are 6 possible ways that $(1,2,3)$ can be ordered. Here they are

$$
(1,2,3)(1,3,2)(2,1,3)(2,3,1)(3,1,2)(3,2,1)
$$

If $(1,2,3)$ can be arranged in 6 ways, it should be clear that this is true of all 20 combinations. Since each of the 20 combinations can be arranged in 6 different ways, there are $(20)(6)=120$ permutations.
15. Consider the experiment of rolling a pair of dice. Suppose we are interested in the sum of the face values showing on the dice.
(a) What is the probability of obtaining a value of 3 or less?

Answer: There are 3 different ways the experiment can result in two values summing to 3 or less. Let $E_{\leq 3}$ be the event that the two dice sum to 3 or less.

$$
E_{\leq 3}=\{(1,1),(1,2),(2,1)\}
$$

and

$$
p\left(E_{\leq 3}\right)=p(1,1)+p(1,2)+p(2,1)=\frac{1}{36}+\frac{1}{36}+\frac{1}{36}=\frac{3}{36}=\frac{1}{12}
$$

Therefore, there is a $1 / 12$ probability of obtaining a value of 3 or less.
(b) What is the probability of obtaining a value of 10 or more?

Answer: There are 6 different ways the experiment can result in values summing to 10 or more. Let $E_{\geq 10}$ be the event that the two dice sum to 10 or more.

$$
E_{\geq 10}=\{(6,4),(5,5),(4,6),(6,5),(5,6),(6,6)\}
$$

and

$$
\begin{gathered}
p\left(E_{\geq 10}\right)=p(6,4)+p(5,5)+p(4,6)+p(6,5)+p(5,6)+p(6,6)= \\
\frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{1}{36}=\frac{6}{36}=\frac{1}{6}
\end{gathered}
$$

Therefore, there is a $1 / 6$ probability of obtaining a value of 10 or more.
(c) Which of the three methods of assigning probability values has been used?

Answer: the classical method of assigning probabilities has been used because the assumption of equally-likely outcomes seems valid.
16. A class has 20 enrolled students. What is the probability that two or more students have the same birthday? As a simplifying assumption, ignore leap years.

Hint: Recall that the probability of an event $A$ equals one minus the probabilty of the complement of that event, $A^{c}$. That is, $p(A)=1-p\left(A^{c}\right)$.

Answer: The probability that at least two people (out of 20) share the same birthday is one minus the probability that no one shares the same birthday. Let $A$ be the event at least two people share the same birthday, and $A^{c}$ be the event that no two people (out of 20) share the same birthday. Then

$$
\begin{gathered}
p(A)=1-p\left(A^{c}\right)=1-\left(\frac{364}{365}\right)\left(\frac{363}{365}\right)\left(\frac{362}{365}\right) \ldots\left(\frac{347}{365}\right)\left(\frac{346}{365}\right) \\
p(A)=1-0.5886=0.4114
\end{gathered}
$$

```
# (1) Use *, ^, and / operators to find products and ratios.
```



```
# (2) The probability of any event equals one minus the
# probability of the complement of that event.
probA = 1 - probAcomp
# (3) Examine contents of probA.
probA
## [1] 0.4114384
# (4) Alternatively, use function prod(N:n) to find product
# N*(N-1)*(N-2)*(N-3)*...*(n+2)*(n+1)*(n).
NewprobAcomp = prod(365 : 346) / (365) ~ 20
# (5) The probability of any event equals one minus the
# probability of the complement of that event.
NewprobA = 1 - NewprobAcomp
    # (6) Examine contents of NewprobA. The results are the same.
NewprobA
    ## [1] 0.4114384
```

Thus there is a 0.4114 probability that at least two people share the same birthday.
17. An experiment consists of casting a single die; the sample space is $S=\{1,2,3,4,5,6\}$. Define three events: $A=\{1,2\}, B=\{3,4\}$, and $C=\{2,3,5,6\}$.
(a) What are $p(A), p(B)$, and $p(C)$ ?

$$
\begin{aligned}
& p(A)=\frac{2}{6}=\frac{1}{3} \\
& p(B)=\frac{2}{6}=\frac{1}{3}
\end{aligned}
$$

$$
p(C)=\frac{4}{6}=\frac{2}{3}
$$

(b) What is $p(A \cup B)$ ?

$$
p(A \cup B)=p(A)+p(B)-p(A \cap B)=\frac{1}{3}+\frac{1}{3}-0=\frac{2}{3}
$$

Note that since $A$ and $B$ are mutually exclusive, $p(A \cap B)=0$
(c) What is $A^{c}$ and $B^{c}$ ?

$$
\begin{aligned}
& A^{c}=\{3,4,5,6\} \\
& B^{c}=\{1,2,5,6\}
\end{aligned}
$$

(d) What is $p\left(A^{c}\right)$ ? What is $p\left(B^{c}\right)$

$$
\begin{aligned}
& p\left(A^{c}\right)=\frac{4}{6}=\frac{2}{3} \\
& p\left(B^{c}\right)=\frac{4}{6}=\frac{2}{3}
\end{aligned}
$$

18. Suppose that we have an experiment with the sample space consisting of seven equally-likely sample points, $S=\left\{S_{1}, S_{2}, S_{3}, S_{4}, S_{5}, S_{6}, S_{7}\right\}$. Define three events in the following way: $A=\left\{S_{1}, S_{4}, S_{6}\right\}, B=\left\{S_{2}, S_{4}, S_{7}\right\}$, and $C=\left\{S_{2}, S_{3}, S_{5}, S_{7}\right\}$
(a) What is $p(A), p(B)$, and $p(C)$ ?

$$
\begin{aligned}
& p(A)=\frac{3}{7} \\
& p(B)=\frac{3}{7}
\end{aligned}
$$

$$
p(C)=\frac{4}{7}
$$

(b) What is $p(A \cap B)$ ?

Since $(A \cap B)=\left\{S_{4}\right\}$, we know

$$
p(A \cap B)=p\left(S_{4}\right)=\frac{1}{7}
$$

(c) What is $p(A \cup B)$ ?

$$
p(A \cup B)=p(A)+p(B)-p(A \cap B)=\frac{3}{7}+\frac{3}{7}-\frac{1}{7}=\frac{5}{7}
$$

(d) What is $p\left(C^{c}\right)$ ?

Since $C^{c}=\left\{S_{1}, S_{4}, S_{6}\right\}$, we know

$$
p\left(C^{c}\right)=p\left(S_{1}\right)+p\left(S_{4}\right)+p\left(S_{6}\right)=\frac{3}{7}
$$

19. An experiment consists of the roll of a single die; the sample space is $S=\{1,2,3,4,5,6\}$. Define two events: $A=\{1,3,5\}$ and $B=\{3,4,5\}$.
(a) What is $(A \cup B)$ ?

Answer: $(A \cup B)=\{1,3,4,5\}$
(b) What is $(A \cap B)$ ?

Answer: $(A \cap B)=\{3,5\}$
(c) What is $p(A \cap B)$ ?

Answer:

$$
p(A \cap B)=p(3)+p(5)=\frac{1}{6}+\frac{1}{6}=\frac{2}{6}=\frac{1}{3}
$$

(d) What is $(A \cup B)$ ?

Answer:

$$
p(A \cup B)=p(1)+p(3)+p(4)+p(5)=\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}=\frac{4}{6}=\frac{2}{3}
$$

20. Referring to the previous exercise, could we use $p(A \cup B)=p(A)+p(B)-p(A \cap B)$ ?

Answer: Yes, we could.

$$
p(A \cup B)=p(A)+p(B)-p(A \cap B)=\frac{3}{6}+\frac{3}{6}-\frac{2}{6}=\frac{4}{6}=\frac{2}{3}
$$

21. Suppose we have two events $A$ and $B$ with $p(A)=0.40$ and $p(B)=0.20$; the probability of the intersection of these two events is 0.10 .
(a) What is $p(A \cup B)$ ?

$$
p(A \cup B)=p(A)+P(B)-p(A \cap B)=0.40+0.20-0.10=0.50
$$

(b) What is $p(B \mid A)$ ?

$$
p(B \mid A)=\frac{p(A \cap B)}{P(A)}=\frac{0.10}{0.40}=0.25
$$

(c) What is $p(A \mid B)$ ?

$$
p(A \mid B)=\frac{p(A \cap B)}{P(B)}=\frac{0.10}{0.20}=0.50
$$

(d) Are events $A$ and $B$ independent?

Answer: Since $p(A \mid B) \neq p(A)$ and $p(B \mid A) \neq p(B)$, it is clear that events $A$ and $B$ are not independent.
22. Suppose we have two events $A$ and $B$ with $p(A)=0.40$ and $p(B)=0.20$; the conditional probability of event $A$ given $B$ is 0.60 .
(a) What is the probability of the intersection of events $A$ and $B$ ?

$$
p(A \cap B)=p(A \mid B) p(B)=(0.60)(0.20)=0.12
$$

(b) What is the conditional probability of event $B$ given $A$ ?

$$
p(B \mid A)=\frac{p(B \cap A)}{p(A)}=\frac{0.12}{0.40}=0.30
$$

(c) What is the probability of the union of events $A$ and $B$ ?

$$
p(A \cup B)=p(A)+p(B)-p(A \cap B)=0.40+0.20-0.12=0.48
$$

23. Suppose an experiment has three mutually-exclusive events, $A, B$, and $C$, with $p(A)=0.34, p(B)=0.33$, and $p(C)=0.33$.
(a) What is the probability of the union of events $A$ and $B$ ?

$$
p(A \cup B)=p(A)+p(B)-p(A \cap B)=0.34+0.33-0=0.67
$$

since $(A \cap B)$ contains no elements; that is, it is the empty set.
(b) What is the probability of the intersection of events $A$ and $B$ ?

$$
p(A \cap B)=0
$$

since the event $(A \cap B)$ is the empty set, there is zero probability of finding any element that belongs to both events $A$ and $B$.
(c) What is the conditional probability of event $B$ given $A$ ?

$$
p(B \mid A)=\frac{p(A \cap B)}{p(A)}=\frac{0}{0.34}=0.00
$$

24. Suppose we have two events $A$ and $B$ with $p(A)=0.40$ and $p(B)=0.20$. Moreover, $p(A \mid B)=p(A)$ and $p(B \mid A)=p(B)$.
(a) What is the probability of the intersection of events $A$ and $B$ ?

$$
p(A \cap B)=p(A) p(B)=(0.40)(0.20)=0.08
$$

(b) What is the conditional probability of event $A$ given $B$ ?

$$
p(A \mid B)=\frac{p(A \cap B)}{p(B)}=\frac{0.08}{0.20}=0.40
$$

(c) What is the probability of the union of events $A$ and $B$ ?

$$
p(A \cup B)=p(A)+p(B)-p(A \cap B)=(0.40)+(0.20)-(0.08)=0.52
$$

25. The following data from a sample of 100 families show the record of college attendance by fathers and their oldest sons: in 22 families, both father and the son attended college; in 31 families, neither father nor son attended college; in 12 families, the father attended college while the son did not; and in 35 families, the son attended college but the father did not.
(a) What is the probability a son attended college given that his father attended college?

|  | Son Attended | Son Did Not | Totals |
| :---: | :---: | :---: | :---: |
| Father Attended | 22 | 12 | 34 |
| Father Did Not | 35 | 31 | 66 |
| Totals | 57 | 43 | 100 |

Organizing this information in a simple table
Define the following events
$S_{c}$ is event that the son attended college
$S_{n}$ is event that the son did not attend college
$F_{c}$ is event that the father attended college
$F_{n}$ is event that the father did not attend college

$$
p\left(S_{c} \mid F_{c}\right)=\frac{p\left(S_{c} \cap F_{c}\right)}{p\left(F_{c}\right)}=\frac{22}{34}=0.6471
$$

Answer: 0.6471 . There is almost a $65 \%$ chance that a son attended college given that his father did.
(b) What is the probability a son attended college given that his father did not attend college?

$$
p\left(S_{c} \mid F_{n}\right)=\frac{p\left(S_{c} \cap F_{n}\right)}{p\left(F_{n}\right)}=\frac{35}{66}=0.5303
$$

Answer: 0.5303 . There is a $53 \%$ chance that a son attended college if his father did not.
(c) Is attending college by the son independent of whether his father attended college?

Since $p\left(S_{c} \mid F_{c}\right)=0.6471>p\left(S_{c}\right)=0.5700$, the events are not independent. In other words, knowing whether the father attended college tells us something about whether the son is likely to have attended.
26. We have the following probabilities and events: the prior probabilities for two events $E_{1}$ and $E_{2}$ are $p\left(E_{1}\right)=0.50$ and $p\left(E_{2}\right)=0.50$ and the conditional probilities are
$p\left(E_{3} \mid E_{1}\right)=0.15$ and $p\left(E_{3} \mid E_{2}\right)=0.70$. Summarize this information in a table, and answer the 3 questions below.

| Event | Prior | Conditional | Joint | Posterior |
| :---: | :---: | :---: | :---: | :---: |
| $E_{1}$ | $p\left(E_{1}\right)=0.50$ | $p\left(E_{3} \mid E_{1}\right)=0.15$ | $p\left(E_{1} \cap E_{3}\right)=0.0750$ | $p\left(E_{1} \mid E_{3}\right)=0.1765$ |
| $E_{2}$ | $p\left(E_{2}\right)=0.50$ | $p\left(E_{3} \mid E_{2}\right)=0.70$ | $p\left(E_{2} \cap E_{3}\right)=0.3500$ | $p\left(E_{2} \mid E_{3}\right)=0.8235$ |
|  | 1.000 |  | $p\left(E_{3}\right)=0.4250$ | 1.000 |

(a) What are $p\left(E_{1} \cap E_{3}\right)$ and $p\left(E_{2} \cap E_{3}\right)$ ?

Answer: Multiplying the prior probability times the conditional probability

$$
\begin{aligned}
& p\left(E_{1} \cap E_{3}\right)=p\left(E_{3} \mid E_{1}\right) p\left(E_{1}\right)=(0.15)(0.50)=0.075 \\
& p\left(E_{2} \cap E_{3}\right)=p\left(E_{3} \mid E_{2}\right) p\left(E_{2}\right)=(0.70)(0.50)=0.3500
\end{aligned}
$$

(b) What is $p\left(E_{3}\right)$ ?

Answer: Summing the two joint probabilities
$p\left(E_{3}\right)=p\left(E_{3} \mid E_{1}\right) p\left(E_{1}\right)+p\left(E_{3} \mid E_{2}\right) p\left(E_{2}\right)=(0.15)(0.50)+(0.70)(0.50)=0.4250$
(c) What are $p\left(E_{1} \mid E_{3}\right)$ and $p\left(E_{2} \mid E_{3}\right)$ ?

Answer: The divide the joint probabilities by $p\left(E_{3}\right)$

$$
\begin{aligned}
& p\left(E_{1} \mid E_{3}\right)=\frac{p\left(E_{1} \cap E_{3}\right)}{p\left(E_{3}\right)}=\frac{0.0750}{0.4250}=0.1765 \\
& p\left(E_{2} \mid E_{3}\right)=\frac{p\left(E_{2} \cap E_{3}\right)}{p\left(E_{3}\right)}=\frac{0.3500}{0.4250}=0.8235
\end{aligned}
$$

27. Consider the following facts: (a) in random testing, you test positive for a disease; (b) all people who have the disease test positive for it; (c) in $5 \%$ of the cases, this test shows positive even when a person does not have the disease (that is, there is a $5 \%$ chance of a false positive); and (d) in the population-at-large, one person in a thousand has the disease. What is the probability that you have the disease? Hint: define the following events: $D$ is event person has the disease; $D^{c}$ is event person does not have the disease; $P$ is event person tests positive for the disease.

| Event | Prior | Conditional | Joint | Posterior |
| :---: | :---: | :---: | :---: | :---: |
| $D$ | $p(D)=0.001$ | $p(P \mid D)=1.000$ | $p(P \cap D)=0.001$ | $p(D \mid P)=0.0196$ |
| $D^{c}$ | $p\left(D^{c}\right)=0.999$ | $p\left(P \mid D^{c}\right)=0.050$ | $p\left(P \cap D^{c}\right)=0.050$ | $p\left(D^{c} \mid P\right)=0.9804$ |
|  | 1.000 |  | $p(P)=0.0510$ | 1.000 |

Thus, perhaps surprisingly, even if you test positive for this disease, there is less than a $2 \%$ chance that you actually have it.
28. We have the following probabilities and events: the prior probabilities for 3 events $E_{1}, E_{2}$, and $E_{3}$ are $p\left(E_{1}\right)=0.50, p\left(E_{2}\right)=0.10$, and $p\left(E_{3}\right)=0.40$ and the conditional probabilities are $p\left(E_{4} \mid E_{1}\right)=0.15, p\left(E_{4} \mid E_{2}\right)=0.45$ and $p\left(E_{4} \mid E_{3}\right)=0.70$. Summarize this information in a table, and answer the 3 questions below.

| Event | Prior | Conditional | Joint | Posterior |
| :---: | :---: | :---: | :---: | :---: |
| $E_{1}$ | $p\left(E_{1}\right)=0.50$ | $p\left(E_{4} \mid E_{1}\right)=0.15$ | $p\left(E_{1} \cap E_{4}\right)=0.0750$ | $p\left(E_{1} \mid E_{4}\right)=0.1875$ |
| $E_{2}$ | $p\left(E_{2}\right)=0.10$ | $p\left(E_{4} \mid E_{2}\right)=0.45$ | $p\left(E_{2} \cap E_{4}\right)=0.0450$ | $p\left(E_{2} \mid E_{4}\right)=0.1125$ |
| $E_{3}$ | $p\left(E_{3}\right)=0.40$ | $p\left(E_{4} \mid E_{3}\right)=0.70$ | $p\left(E_{3} \cap E_{4}\right)=0.2800$ | $p\left(E_{3} \mid E_{4}\right)=0.7000$ |
|  | 1.000 |  | $p\left(E_{4}\right)=0.4000$ | 1.000 |

(a) What are $p\left(E_{1} \cap E_{4}\right), p\left(E_{2} \cap E_{4}\right)$, and $p\left(E_{3} \cap E_{4}\right)$ ?

Answer: Multiplying the prior probability times the conditional probability

$$
\begin{aligned}
& p\left(E_{1} \cap E_{4}\right)=p\left(E_{4} \mid E_{1}\right) p\left(E_{1}\right)=(0.15)(0.50)=0.0750 \\
& p\left(E_{2} \cap E_{4}\right)=p\left(E_{4} \mid E_{2}\right) p\left(E_{2}\right)=(0.45)(0.10)=0.0450 \\
& p\left(E_{3} \cap E_{4}\right)=p\left(E_{4} \mid E_{3}\right) p\left(E_{3}\right)=(0.70)(0.40)=0.2800
\end{aligned}
$$

(b) What is $p\left(E_{4}\right)$ ?

Answer: Summing the three joint probabilities

$$
\begin{aligned}
& p\left(E_{4}\right)=p\left(E_{4} \mid E_{1}\right) p\left(E_{1}\right)+p\left(E_{4} \mid E_{2}\right) p\left(E_{2}\right)+p\left(E_{4} \mid E_{3}\right) p\left(E_{3}\right) \\
& \quad=(0.15)(0.50)+(0.45)(0.10)+(0.70)(0.40)=0.4000
\end{aligned}
$$

(c) What are $p\left(E_{1} \mid E_{4}\right), p\left(E_{2} \mid E_{4}\right)$, and $p\left(E_{3} \mid E_{4}\right)$ ?

Answer: The divide the joint probabilities by $p\left(E_{4}\right)$

$$
p\left(E_{1} \mid E_{4}\right)=\frac{p\left(E_{1} \cap E_{4}\right)}{p\left(E_{4}\right)}=\frac{0.0750}{0.4000}=0.1875
$$

$$
\begin{aligned}
& p\left(E_{2} \mid E_{4}\right)=\frac{p\left(E_{2} \cap E_{4}\right)}{p\left(E_{4}\right)}=\frac{0.0450}{0.4000}=0.1125 \\
& p\left(E_{3} \mid E_{4}\right)=\frac{p\left(E_{3} \cap E_{4}\right)}{p\left(E_{4}\right)}=\frac{0.2800}{0.4000}=0.7000
\end{aligned}
$$

