Homework 1

Exercise 1. What is $P(\Omega) = P(\{1, 2, 3, 4, 5, 6\})$? ∎

$P(\Omega) = 1$

Exercise 2. Let D be the event that a person has a disease, and $\bar{D}$ be the event that the person does not have the disease. Show $P(D) + P(\bar{D}) = 1$. ∎

$\Omega = \{D, \bar{D}\}$
$P(\{D, \bar{D}\}) = P(\Omega) = 1$
$P(\{D, \bar{D}\}) = P(D) + P(\bar{D})$ since D and $\bar{D}$ are disjoint.
Therefore, $P(D) + P(\bar{D}) = 1$

Exercise 3. Let $P(D)$ be the prevalence π. Let $\theta = P(E)$, the fraction testing positive. Show

$$\theta = \pi s + (1 - \pi)(1 - f),$$

where s is the sensitivity and f is the specificity. Suppose the fraction testing positive is known from a survey to be 0.001. What is an implied bound on the specificity in this population? ∎

We can write a 2x2 table of probabilities that sum to one:

	$\bar{T}$	T
$\bar{D}$	$f(1-\pi)$	$(1-f)(1-\pi)$
D	$(1-s)\pi$	$s\pi$

θ is the fraction testing positive, so summing the righthand column of the above table, we have:
$\theta = (1 - f)(1 - \pi) + s\pi$.

Solving for the specificity f, we have $f = 1 - \frac{\theta - s\pi}{1-\pi}$. The minimum value of the sensitivity s is 0 and the maximum value is 1, corresponding to the maximum and minimum values for f.
Therefore, $f \epsilon [1 - \frac{\theta}{1-\pi}, 1 - \frac{\theta-\pi}{1-\pi}]$. If the prevalence π is equal to 0, then the minimum possible f is $1 - \theta = 0.999$. However, higher prevalences will yield lower minimums.

Exercise 4. If D is independent of E, show $\bar{D}$ is independent of E. ∎

Law of Total Probability: $P(D) + P(\bar{D}) = 1$
$P(D|E) + P(\bar{D}|E) = 1$
Because D and E are independent: $P(D|E) = P(D)$.
Multiplying by -1 and adding 1 to both sides: $1 - P(D|E) = 1 - P(D)$.
Finally, substituting from the two equations above we have:
$P(\bar{D}|E) = P(\bar{D})$. Therefore, $\bar{D}$ is independent of E.

Exercise 5. Suppose A, B, and C are binary events, that either happen or they don't. The sample space can be classified in a two by two by two table. Find a way to assign the probabilities of in this table so that A and B are independent, A and C are independent, and B and C are independent, but A, B, and C are not independent. (For instance, make $P(A \cap B) = P(A)P(B)$, $P(A \cap C) = P(A)P(C)$, and $P(B \cap C) = P(B)P(C)$, but make sure $P(A \cap B \cap C) \neq P(A)P(B)P(C)$, etc.) ∎

Answers will vary.

P(000)=0, P(010)=0.25, P(100)=0.25, P(110)=0, P(001)=0.25, P(011)=0, P(101)=0, P(111)=0.25.

Exercise 6. Given a two by two table with cell probabilities $P(AB) = P(A \cap B)$, $P(A\bar{B})$, $P(\bar{A}B)$, and $P(\bar{A}\bar{B})$, the *odds ratio* is

$$o = \frac{P(AB)P(\bar{A}\bar{B})}{P(\bar{A}B)P(A\bar{B})}.$$

Show that the odds ratio equals 1 when and only when A and B are independent. ∎

We must show two things. a) An odds ratio equal to 1 implies A and B are independent and b) A and B independent implies an odds ratio equal to 1.

a)

$o = \frac{P(AB)P(\bar{A}\bar{B})}{P(\bar{A}B)P(A\bar{B})} = 1$

$P(A|B)P(\bar{A}|\bar{B}) = P(\bar{A}|B)P(A|\bar{B})$

$P(A|B)(1 - P(A|\bar{B})) = (1 - P(A|B))P(A|\bar{B})$

$\frac{P(A|B)}{1-P(A|B)} = \frac{P(A|\bar{B})}{1-P(A|\bar{B})}$

Therefore, $P(A|B) = P(A|\bar{B})$, which implies A and B are independent.

b)

If A and B are independent:

$o = \frac{P(AB)P(\bar{A}\bar{B})}{P(\bar{A}B)P(A\bar{B})} = \frac{P(A)P(B)P(\bar{A})P(\bar{B})}{P(\bar{A})P(B)P(A)P(\bar{B})} = 1$

Exercise 7. Given a two by two by two table representing events A, B, and C, show

$$P(A|BC) = \frac{P(AB|C)P(C)}{P(BC)}.$$

$P(ABC) = P(AB|C)P(C)$

$P(ABC) = P(A|BC)P(BC)$

Setting these expressions equal to each other, we have:

$P(AB|C)P(C) = P(A|BC)P(BC)$

Rearranging, we have: $P(A|BC) = \frac{P(AB|C)P(C)}{P(BC)}$.

Exercise 8. Two events B and C are conditionally independent given A if $P(BC|A) = P(B|A)P(C|A)$. Assume that B is conditionally independent of C given A. First show given this assumption, $P(B|AC) = P(B|A)$. Then show that

$$P(A|BC) = \frac{P(B|A)P(A|C)}{P(B|C)}.$$

1)

We are given: $P(BC|A) = P(B|A)P(C|A)$

We can also write: $P(BC|A) = P(B|AC)P(C|A)$

$P(B|A)P(C|A) = P(B|AC)P(C|A)$

Therefore, $P(B|AC) = P(B|A)$.

2)

$P(AB|C) = P(A|BC)P(B|C)$

$P(AB|C) = P(B|AC)P(A|C)$

Setting these expressions equal to each other, we have:

$P(A|BC)P(B|C) = P(B|AC)P(A|C)$

$P(A|BC) = \frac{P(B|AC)P(A|C)}{P(B|C)}$

Using 1), we have:

$P(A|BC) = \frac{P(B|A)P(A|C)}{P(B|C)}$

Exercise 9. Suppose a person has a sexual partner and that the probability that the partner is infected with HIV is π. Suppose that the probability of infection given that the partner is infected is β. What is the probability of infection in this partnership? ∎

$\beta\pi$

Exercise 10. Show that $P(A \cup B \cup C) \leq P(A) + P(B) + P(C)$. Suppose a person has a needlestick injury with blood contaminated with Hepatitis B, Hepatitis C, and HIV; suppose the probability of HBV transmission is 0.3 from any needlestick, HCV transmission is 0.03 from any needlestick, and HIV transmission is 0.003 from any needlestick. Show the total probability of any infection is less than or equal to 0.333, no matter what the independence or nonindependence between the infections is. ∎

$P(A \cup B \cup C) \leq P(A) + P(B) + P(C)$

Therefore, the total probability of any infection is less than $0.3 + 0.03 + 0.003 = 0.333$.

Exercise 11. Suppose two parents are carriers of the sickle gene and that the chance of passing the gene to the next generation is 1/2. What is the probability distribution of the number of children out of 4 who would be born homozygous (received the gene from both parents)? Assume independent and identical risk. ∎

$P(X = x) = \binom{4}{x}(0.25)^x(0.75)^{4-x}$

Exercise 12. Suppose a person has n sexual partners chosen at random from a pool of partners in which the prevalence is π and the risk of transmission given a positive partner is β. Show the probability of infection is $1 - (1 - \beta p)^n$. ∎

Probability of infection per partner: βp

Probability didn't get infected in a partnership: $1-\beta p$

Probability didn't get infected in n partnerships: $(1-\beta p)^n$

Probability got infected in n partnerships: $1-(1-\beta p)^n$

Exercise 13. Show the expectation of a Bernoulli random variable is p. ∎

$E[X] = \sum_{n=0}^{1} np(n) = 0(1-p) + 1p = p$

Exercise 14. It has been estimated that the fraction of wild-type tubercle bacilli which harbor a resistance mutation to isoniazid, a first-line drug, is 10^{-6}, and the fraction that harbor mutations to rifampicin is 10^{-8}. Assuming these are independent, what is the probability of harboring mutations conferring resistance to both (and by definition being multidrug resistant)? Assume that a cavitary lesion contains 10^9 organisms. What is the expected number with isoniazid mutations? Rifampicin mutations? MDR (both isoniazid and rifampicin mutations)? ∎

INH mutations: $10^{-6}10^{9} = 10^3$

RIF mutations: $10^{-8}10^{9} = 10$

Both mutations: $10^{-8}10^{-6}10^{9} = 10^{-5}$

Exercise 15. Let X and Y be jointly distributed with $P(X = 0, Y = 0) = 0$, $P(X = 0, Y = 1) = 1/2$, $P(X = 1, Y = 0) = 1/2$, $P(X = 1, Y = 1) = 0$. Here, $P(X = 1) = 1/2$ and $P(Y = 0) = 1/2$, but $P(X = 1, Y = 0) \neq 1/2 \times 1/2$, so X and Y are NOT independent. Compute $E[XY]$ from the definition. What is $E[X]E[Y]$? ∎

$E[XY] = \sum_x \sum_y xyP(X = x, Y = y) = 0 \times 0 \times 0 + 0 \times 1 \times 0.5 + 1 \times 0 \times 0.5 + 1 \times 1 \times 0 = 0$

$E[X]E[Y] = 0.5^2 = 0.25$

Exercise 16. Since the number of successes in N independent identical Bernoulli trials is the sum of their values, every 1 being a success, show that if X is a Binomial with N trials and p successes per trial, $EX = Np$. ∎

$E[X] = E[\sum_{i=1}^{N} Y_i]$, where Y_i is a Bernoulli RV with expectation p.

$= \sum_{i=1}^{N} E[Y_i] = Np$.

Exercise 17. Suppose 51 people receive a needlestick injury with hepatitis C contaminated blood. Assume independent and identical risk, and that 3 individuals were infected as a result. Compute the probability of this event for $p = 0$, $p = 0.001$, $p = 0.01$, $p = 3/51$, $p = 0.1$, $p = 0.5$. ∎

```
get.probability <- function(p){
        dbinom(3,51,p)
```

```
}
get.probability(c(0,0.001,0.01,3/51,0.1,0.5))
```

	p	probability
1	0.000E+00	0.000E+00
2	1.000E-03	1.985E-05
3	1.000E-02	1.286E-02
4	5.882E-02	2.309E-01
5	1.000E-01	1.325E-01
6	5.000E-01	9.248E-12

Exercise 18. Show the variance of a Bernoulli random variable is $p(1-p)$. Use this and the rule for summing variances of independent variables to show the variance of a Binomial with N trials and p per trial is $varX = Np(1-p)$. ∎

Let Y be a Bernouilli random variable with expectation p.
$var(Y) = E[(Y - E[Y])^2]$
$= E[Y^2] - E[2YE[Y]] + E[E[Y]^2]$
$= \sum_y y^2 P(Y = y) - E[Y]^2$
$= 0^2(1-p) + 1p - p^2 = p - p^2 = p(1-p)$
Let X be the sum of N independent trials.
$var(X) = var[\sum_{i=1}^N Y_i] = Nvar(Y) = Np(1-p)$

Exercise 19. Suppose you believe that a certain drug will work around 30 percent of the time, and you want to estimate this to within a standard deviation of plus or minus five percent in absolute terms. About how large does N have to be? ∎

From lecture: $SE(\hat{p}) = \sqrt{\frac{p(1-p)}{N}}$
$SE(\hat{p}) = \sqrt{\frac{p(1-p)}{N}} \leq 0.05$
$p(1-p) \leq 0.05^2 N$
$N \geq (0.3)(0.7)/0.05^2 = 84$

Exercise 20. How long did this assignment take you? We want future assignments to take about six hours. ∎