S520 Problem Set 3 Solutions

Arturo Valdivia

Due on 1/31/2022

Problem 1. ISI Section 3.7 Problem 8

Let
$D = \{$ person suffering form disease$\}$

$+ = \{$ tested positive$\}$
$- = \{$tested negative$\}$

Observe that "$-$" is the complement of "$+$".

Sensitivity is the probability of testing positive given that the person suffers from disease. We also call the probability of true positives. Similarly, specificity is the probability of testing negative given that the person doesn't suffer from disease, or the probability for true negatives. In the question we have

Sensitivity $= P(+|D) = 0.71$
Specificity $= P(-|D^c) = 0.88$

and $P(D) = 0.03$

a. The probability of false positives are given by $P(+|D^c)$ which is the complement of $P(-|D^c)$ so

$$P(-|D^c) = 1 - P(+|D^c) = 1 - 0.88 = 0.12$$

b. Similarly, for the probability of false negatives, we get

$$P(-|D) = 1 - P(+|D) = 1 - 0.71 = 0.29$$

c. The tree diagram looks like the one presented in our class notes `S520_012423_notes_tree_diagrams_annotated.pdf` with:

- $P(D) = P(A) = 0.03$,
- $P(D^c) = P(A^c) = 0.97$,
- $P(+|D) = P(B|A) = 0.71$
- $P(-|D) = P(B^c|A) = 0.29$
- $P(+|D^c) = P(B|A^c) = 0.12$
- $P(-|D^c) = P(B^c|A^c) = 0.88$

d. We are looking for $P(+)$

$$P(+) = P(D) \cdot P(+|D) + P(D^c) \cdot P(+|D^C) = 0.03 * 0.71 + 0.97 * 0.12 = 0.138$$

```
0.03 * 0.71 + 0.97 * 0.12
```

```
## [1] 0.1377
```

e) We need to find $P(D|+)$. Using Bayes' rule we get:

$$P(D|+) = \frac{P(D \cap +)}{P(+)} = \frac{P(D) \cdot P(+|D)}{P(+)} = \frac{0.03 * 0.71}{0.033} = 0.15$$

```
(0.03 * 0.71)/(0.03 * 0.71 + 0.97 * 0.12)
```

```
## [1] 0.1546841
```

There is about 15% chance that the woman will have pre-eclampsia given that she tested positive.

Problem 2. ISI Section 4.5 Problem 3

The PMF is : a.

$$f(x) = \begin{cases} 0.5 & x = 1 \\ 0.3 & x = 3 \\ 0.2 & x = 7 \\ 0 & otherwise \end{cases}$$

b.

$$P(X \le y) = F(y)$$

$$F(y) = \begin{cases} 0 & -\infty < y < 1 \\ 0.5 & 1 \le y < 3 \\ 0.8 & 3 \le y < 7 \\ 1 & 7 \le y < \infty \end{cases}$$

c.

$$EX = \sum_{x \in \{1,3,7\}} x \cdot f(x) = 1 \cdot 0.5 + 3 \cdot 0.3 + 7 \cdot 0.2 = 2.8$$

d.

$$VarX = \sum_{x \in \{1,3,7\}} (x - EX)^2 \cdot f(x) = (1 - 2.8)^2 \cdot 0.5 + (3 - 2.8)^2 \cdot 0.3 + (7 - 2.8)^2 \cdot 0.2 = 5.16$$

e.

$$\sqrt{VarX} = \sqrt{5.16} = 2.27$$

Here are the calculations in R:

```
x = c(1, 3, 7)
fx = c(0.5, 0.3, 0.2)
mu = sum(x*fx)
sigma.sq = sum((x-mu)^2*fx)
sigma = sqrt(sigma.sq)
c(mu, sigma.sq, sigma)
```

```
## [1] 2.800000 5.160000 2.271563
```

Problem 3

a.

$$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$$

b.

$$X(S) = \{-15, 0, 15, 30\}$$

c.

$$F(y) = \begin{cases} 0 & -\infty < y < -15 \\ 1/8 & -15 \le y < 0 \\ 1/2 & 0 \le y < 15 \\ 7/8 & 15 \le y < 30 \\ 1 & 30 \le y < \infty \end{cases}$$

d.

$$f(x) = \begin{cases} 1/8 & y = -15 \\ 3/8 & y = 0 \\ 3/8 & y = 15 \\ 1/8 & y = 30 \\ 0 & otherwise \end{cases}$$

e.

$$EX = \sum_{x \in \{-15,0,15,30\}} xf(x) = (-15) \cdot (1/8) + 0 \cdot (3/8) + 15(3/8) + 30(1/8) = 7.5$$

f.

$$VarX = \sum (x - E(X)^2) \cdot f(x) = (-15 - 7.5)^2(1/8) + (0 - 7.5)^2(3/8) + (15 - 7.5)^2(3/8) + (30 - 7.5)^2(1/8) = 168.75$$

$$sd = \sqrt{VarX} = \sqrt{168.75} = 12.99$$

Here is the R code for parts e - f.

```
x = c(-15,0,15,30)
fx = c(1/8, 3/8, 3/8, 1/8)
EX = sum(x*fx)
VarX = sum((x - EX)^2*fx)
c(EX, VarX, sqrt(VarX))
```

```
## [1]   7.50000 168.75000  12.99038
```

Problem 4

a. S is the set of all possible outcomes. An outcome contains as many cards drawn needed until we get an Ace. A couple of outcomes would be, for example, $(3\heartsuit, A\clubsuit)$ or $(K\diamondsuit, 10\clubsuit, 3\heartsuit, 2\clubsuit, A\spadesuit)$. Moreover, because draws happen with replacement, there is no limit about how many draws may be needed, so

b.
$$Y(S) = \{1, 2, 3, ...\} = \mathcal{N}$$

c. There are exactly 4 aces in the deck, so the chances of selecting an ace is $4/52 = 1/13$.

$$f(-4) = 0, f(\pi) = 0, f(4) = \left(\frac{12}{13}\right)^3 \cdot \frac{1}{13} = 0.060502$$

$$F(-2) = 0, F(2) = f(1) + f(2) = \frac{1}{13} + \frac{12}{13} \cdot \frac{1}{13} = 0.148$$

Using R:

```
# f(4)
p=1/13
p*(1-p)^3
```

```
## [1] 0.06050208
```

```
# F(2)
(p) + (1-p)*(p)
```

```
## [1] 0.147929
```

d.
$$f(y) = \left(\frac{12}{13}\right)^{y-1} \frac{1}{13},\ y = 1, 2, 3, ...$$

and $f(y) = 0$ for any $y \notin \mathbb{N}$

Problem 5 :

a. Since $X \sim Bernoulli(p)$:

$$f(x) = \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \end{cases}$$

$$EX = \sum_{x \in \{0,1\}} x \cdot f(x) = 0 \cdot f(0) + 1 \cdot f(1) = 0(1-p) + 1p = p$$

$$VarX == (0-\mu)^2 \cdot f(0) + (1-mu)^2 \cdot f(1) = (0-p)^2(1-p) + (1-p)^2 p = p(1-p)[p + (1-p)] = p(1-p)$$

b. Since $Y \sim Binomial(n, p)$ is, by definition, the sum of n independent Bernoulli trials, all with the same parameter p:

$$Y = \sum_{i=1}^{n} X_i$$

where $X_i \sim Bernoulli(p)$

We can then use the properties of expected value and variance, to get:

$$EY = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} EX_i = \sum_{i=1}^{n} p = np$$

$$VarY = Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} VarX_i = \sum_{i=1}^{n} p(1-p) = np(1-p)$$