Dawson College: **Probability and Statistics**: 201-BZS-05-S01: Fall 2016

Name: Y. Lamontagne

Test 2

This test is graded out of ~~40~~ 38 marks. No books, notes, unauthorised electronic devices are allowed. You must show all your work, the correct answer is worth 1 mark the remaining marks are given for the work. If you need more space for your answer use the back of the page.

Question 1.[1] *(5 marks)* A genetic test is used to determine if people have a predisposition for thrombosis, which is the formation of a blood clot inside a blood vessel that obstructs the flow of blood through the circulatory system. It is believed that 3% of people actually have this predisposition. The genetic test is 99% accurate if a person actually has the predisposition, meaning that the probability of a positive test result when a person actually has the predisposition is 0.99. The test is 98% accurate if a person does not have the predisposition. What is the probability that a randomly selected person who tests positive for the predisposition by the test actually has the predisposition?

Let T - test positive

A_1 - predisposition for thrombosis

A_2 - $(A_1)'$ - no predisposition for thrombosis

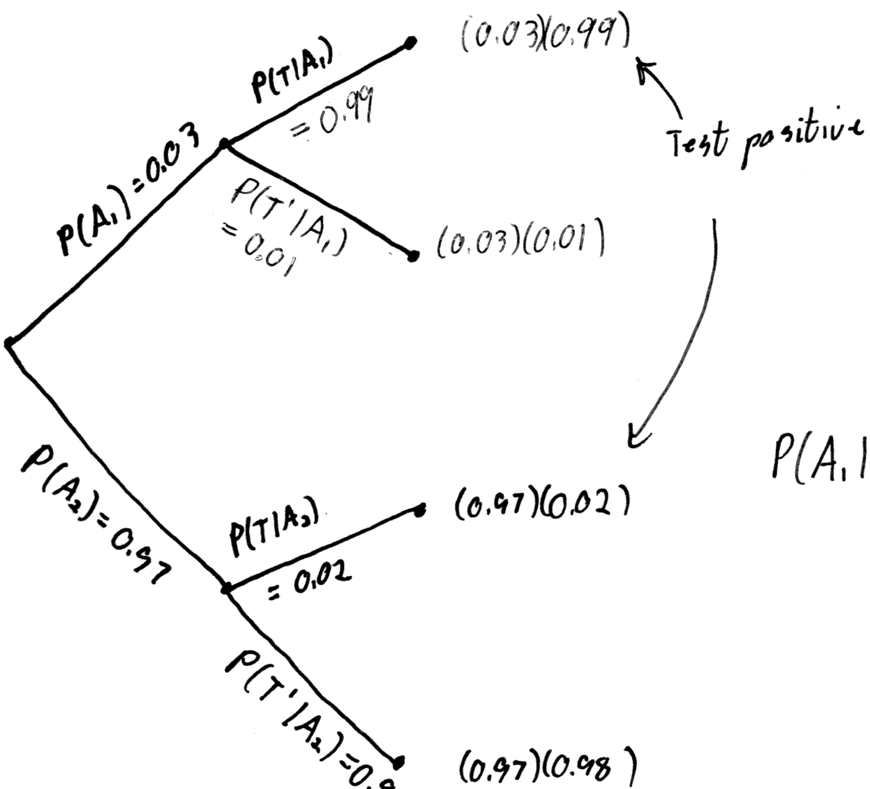
Given $p(A_1) = 0.03$ $\quad P(T \mid A_1) = 0.99$

$P(A_2) = 0.97$ $\quad P(T \mid A_2) = 0.02$

by Bayes' thm

$$P(A_1 \mid T) = \frac{P(A_1 \cap T)}{P(T)} = \frac{P(T \mid A_1)P(A_1)}{\sum_{i=1}^{2} P(T \mid A_i)P(A_i)} = \frac{0.99(0.03)}{0.99(0.03) + 0.02(0.97)}$$

$$= 60.5\%$$



$$P(A_1 \mid T) = \frac{0.99(0.03)}{0.99(0.03) + (0.02)(0.97)}$$

$$= 60.5\%$$

[1] OpenIntro Statistics by D.M. Diez, C.D. Barr and M. Çetinkaya-Rundel, OpenIntro LaTeX, code, and PDFs are released under a Creative Commons BY-SA 3.0 license.

Question 2. The *uniform probability density function* of a continuous random variable X is defined as

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a,b] \\ 0 & \text{if } x \notin [a,b] \end{cases}$$

a. *(2 marks)* Prove that the uniform probability density function is a probability density function.

b. *(2 marks)* Find μ and σ for the uniform probability density function.

c. *(2 marks)* ~~Find~~ Express $P(X > \mu \mid X < \mu + \sigma)$ in terms of integrals.

a) $f(x) \geq 0 \quad \forall x \in \mathbb{R}$

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{a} f(x)dx + \int_{a}^{b} f(x)dx + \int_{b}^{\infty} f(x)dx = \int_{-\infty}^{a} 0\,dx + \int_{a}^{b} f(x)dx + \int_{b}^{\infty} 0\,dx$$

$$= \int_{a}^{b} \frac{1}{b-a}dx = \left[\frac{x}{b-a}\right]_a^b$$

$$= \frac{b}{b-a} - \frac{a}{b-a}$$

$$= \frac{b-a}{b-a} = 1$$

$\therefore f(x)$ is a pdf.

b) $$\mu = \int_{-\infty}^{\infty} x f(x)dx = \int_{-\infty}^{a} x(0)dx + \int_{a}^{b} x \frac{1}{b-a}dx + \int_{b}^{\infty} x(0)dx$$

$$= \int_{a}^{b} \frac{x}{b-a}dx = \left[\frac{x^2}{2}\frac{1}{b-a}\right]_a^b$$

$$= \frac{b^2}{2}\frac{1}{b-a} - \frac{a^2}{2}\frac{1}{b-a}$$

$$= \frac{b^2-a^2}{2(b-a)} = \frac{\cancel{(b-a)}(b+a)}{2\cancel{(b-a)}} = \frac{b+a}{2}$$

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

$$= \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$$

$$= \int_{-\infty}^{a} x^2(0)dx + \int_{a}^{b} x^2 \frac{1}{b-a}dx + \int_{b}^{\infty} x^2(0)dx - \mu^2 = \int_{a}^{b} x^2\frac{1}{b-a}dx - \left(\frac{b+a}{2}\right)^2$$

$$= \left[\frac{x^3}{3}\frac{1}{b-a}\right]_a^b - \left(\frac{b+a}{2}\right)^2 = \frac{b^3-a^3}{3(b-a)} - \left(\frac{b+a}{2}\right)^2$$

$$= \frac{\cancel{(b-a)}(a^2+ab+b^2)}{3\cancel{(b-a)}} - \frac{(a+b)^2}{4}$$

$$= \frac{1}{12}\left[4(a^2+ab+b^2) - 3(a^2+2ba+b^2)\right]$$

$$= \frac{1}{12}\left[a^2-2ab+b^2\right] = \frac{1}{12}(a-b)^2$$

$$\therefore \sigma = \sqrt{\frac{1}{12}(a-b)^2} = \frac{1}{\sqrt{12}}|a-b|$$

c) $$P(X > \mu \mid X < \mu+\sigma)$$

$$= \frac{P((X>\mu) \cap (X<\mu+\sigma))}{P(X<\mu+\sigma)}$$

$$= \frac{\int_{\mu}^{\mu+\sigma} f(x)dx}{\int_{-\infty}^{\mu+\sigma} f(x)dx} = \frac{\int_{\mu}^{\mu+\sigma} \frac{1}{b-a}dx}{\int_{-\infty}^{\mu+\sigma} \frac{1}{b-a}dx}$$

Question 3. A small pond has 11 fishes of which 4 are tagged.

a. *(2 marks)* If a marine biologist catches 4 fish from the small pond, find a probability function for the discrete random variable X whose values x are the number of tagged fish caught by the marine biologist.

b. *(2 marks)* Prove that the function found in part a. is a probability function.

c. *(2 marks)* What is the probability that the marine biologist will catch fewer tagged fish than expected.

a) $f(x) = \dfrac{\binom{4}{x}\binom{11-4}{4-x}}{\binom{11}{4}} = \dfrac{\binom{4}{x}\binom{7}{4-x}}{\binom{11}{4}}$

b) $f(x) \geq 0 \qquad x \in \{0,1,2,3,4\}$

$$\sum_{x=0}^{4} f(x) = f(0)+f(1)+f(2)+f(3)+f(4) = \frac{\binom{4}{0}\binom{7}{4}+\binom{4}{1}\binom{7}{3}+\binom{4}{2}\binom{7}{2}+\binom{4}{3}\binom{7}{1}+\binom{4}{4}\binom{7}{0}}{\binom{11}{4}}$$

$$= \frac{35+140+126+28+1}{\binom{11}{4}} = \frac{330}{330} = 1$$

$\therefore$ the function is a probability function

c) $$\mu = \sum_{x=0}^{4} x \cdot f(x) = \frac{0\binom{4}{0}\binom{7}{4}+1\binom{4}{1}\binom{7}{3}+2\binom{4}{2}\binom{7}{2}+3\binom{4}{3}\binom{7}{1}+4\binom{4}{4}\binom{7}{0}}{\binom{11}{4}}$$

$$= \frac{0+140+2(126)+3(28)+4(1)}{\binom{11}{4}} = \frac{480}{330} = 1.45$$

or $\mu = n\frac{M}{N} = 4\,\frac{4}{11} = 1.45$

$$P(X < 1.45) = P(X=0) + P(X=1)$$

$$= f(0) + f(1)$$

$$= \frac{\binom{4}{0}\binom{7}{4}+\binom{4}{1}\binom{7}{3}}{\binom{11}{4}} = \frac{35+140}{330} = 53\ \%$$

Question 4.[2] *(5 marks)* Suppose a university announced that it admitted 2,500 students for the following year's freshman class. However, the university has dorm room spots for only 1,786 freshman students. If there is a 70% chance that an admitted student will decide to accept the offer and attend this university, what is the approximate probability that the university will not have enough dormitory room spots for the freshman class?

X = # of students accept the university offer.

p - success - accepting university offer = 70%

$$P(X=x) = \binom{2500}{x}(0.70)^x(0.30)^{2500-x}$$

P(not enough dorm room for accepted students)

$$= P(1787 \le X \le 2500)$$

$$= \sum_{x=1787}^{2500} P(X=x) \approx P_{normal}(1786.5 \le X \le 2500.5)$$ continuity correction

normal dist

can use normal dist. to approx since $np \ge 10$ and $n(1-p) \ge 10$

$$\mu = np = 2500(0.7) = 1750$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{2500(0.7)(0.3)} = 22.91$$

$$= P\left(\frac{1786.5-1750}{22.91} \le Z \le \frac{2500.5-1750}{22.91}\right)$$

$$= P(1.59 \le Z \le 32.76)$$

$$= P(Z \le 32.76) - P(Z \le 1.59)$$

$$= 1 - 0.9441$$

$$= 0.0559$$

$\therefore$ about 6% probability of not enough dorm rooms.

---

[2]OpenIntro Statistics by D.M. Diez, C.D. Barr and M. Çetinkaya-Rundel, OpenIntro LaTeX, code, and PDFs are released under a Creative Commons BY-SA 3.0 license.

Question 5.[3] If $P(A) = 0.3$, $P(B) = 0.7$ and assuming that events A and B arise from independent random processes,

i. *(1 mark)* Compute $P(A \text{ and } B)$.

$$P(A \wedge B) = P(A)P(B) \quad \text{since A and B are independent}$$
$$= 0.3(0.7)$$
$$= 0.21$$

ii. *(1 mark)* Compute $P(A \text{ or } B)$.

$$P(A \cup B) = P(A) + P(B) - P(A \wedge B)$$
$$= 0.3 + 0.7 - 0.21$$
$$= 0.79$$

iii. *(1 mark)* Compute $P(A \mid B)$.

$$P(A \mid B) = P(A) \quad \text{by definition of independence}$$
$$= 0.3$$

Question 6. (2 marks) Prove: If A_1, A_2, A_3 are events then $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2)$.

$$P(A_1 \wedge A_2 \wedge A_3) = P((A_1 \wedge A_2) \wedge A_3)$$
$$= P(A_3 \mid A_1 \wedge A_2)\, P(A_1 \wedge A_2)$$
$$= P(A_3 \mid A_1 \wedge A_2)\, P(A_2 \mid A_1)\, P(A_1)$$

Question 7.[4] Below are final exam scores of 20 Introductory Statistics students.

57, 66, 69, 71, 72, 73, 74, 77, 78, 78, 79, 79, 81, 81, 82, 83, 83, 88, 89, 94

a. *(3 marks)* The mean score is 77.7 with a standard deviation of 8.44. Use this information to determine if the scores approximately follow the 68 – 95 – 99.7% Rule.

$77.7 \pm 8.44 \to (69.26, 86.14)$ within 1 SD of $\bar{x}$ 14 observation/20 = 70%

$77.7 \pm 2(8.44) \to (60.82, 94.58)$ within 2 SD of $\bar{x}$ 19 observation/20 = 95%

$77.7 \pm 3(8.44) \to (52.38, 103.02)$ within 3 SD of $\bar{x}$ 20 observation/20 = 100%

The scores approx. follow the 68-95-99.7% rule.

b. *(2 marks)* What does the Chebyshev Theorem tell us about the data and 2 standard deviation.

$\left(1 - \frac{1}{2^2}\right)\% = 75\%$. at least 75% of the data lie within 2 standard deviation of the mean.

[3]OpenIntro Statistics by D.M. Diez, C.D. Barr and M. Çetinkaya-Rundel, OpenIntro LaTeX, code, and PDFs are released under a Creative Commons BY-SA 3.0 license.

[4]OpenIntro Statistics by D.M. Diez, C.D. Barr and M. Çetinkaya-Rundel, OpenIntro LaTeX, code, and PDFs are released under a Creative Commons BY-SA 3.0 license.

Question 8.[5] The expression used in evaluating the *R score* is:

$$\text{R score} = (Z + ISG + 5) \times 5$$

where Z and ISG are the numerical expressions for the Z score and the indicator of the strength of the group, respectively. The ISG is defined as

$$ISG = \frac{\text{average grade of the group at the secondary level} - 75\%}{14}$$

a. *(2 marks)* Suppose that the average grade of the group at the secondary now taking Statistics and Probability at CEGEP is 86%. The Statistics and Probability class average is 75% and standard deviation is 12. Compute the R score of a student who has an average of 80%.

Z-score

$= \frac{80-75}{12}$

$= 0.4167$

$ISG = \frac{86-75}{14} = 0.7857$

R score $= (0.4167 + 0.7857 + 5)5$

$= 31.01$

b. *(2 marks)* For a different course a student gets an R score of 34 what is the student's average in the course. assuming that the students taking the class are the same as in part a. and the grades are normally distributed.

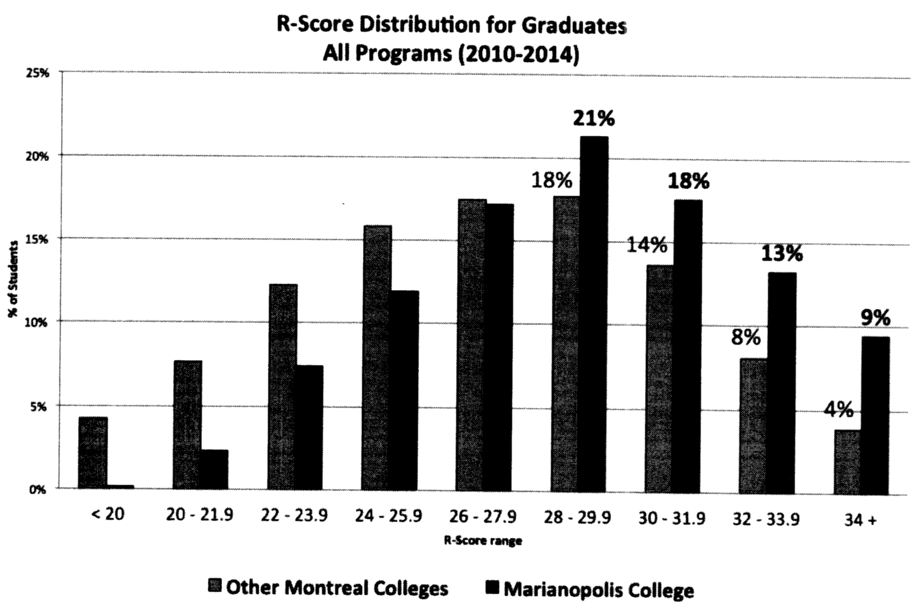
$34 = (Z\text{ score} + 0.7857 + 5) \times 5$

Z score $= \frac{34}{5} - 0.7857 - 5 = 1.01$

$1.01 = \frac{X - \bar{X}}{S}$

$X = \bar{X} + 1.01S$

c. *(2 marks)* The following graph is presented on the Marionapolis College website.[6]



Assuming that Z-score follow a standard normal distribution and the above claim is true, give a possible explanation of the above results.

The skew is due to the ISG being higher.

---

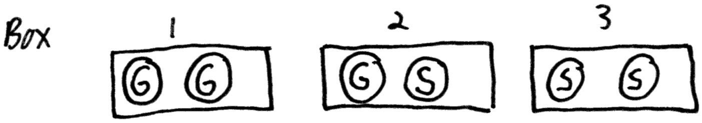
[5]CREPUQ. The R score : what it is, and what it does. September 3, 2004

[6]http://bemarianopolis.ca/choose-us/r-score/

Bonus Question.[7] *(3 marks)*
Bertrand's box paradox is a paradox of elementary probability theory, first posed by Joseph Bertrand in his 1889 work Calcul des probabilitiés.

There are three boxes: a box containing two gold coins, a box containing two silver coins, a box containing one gold coin and a silver coin.

After choosing a box at random and withdrawing one coin at random, if that happens to be a gold coin, what is probability of the next coin also being a gold coin?



$$P(\text{Next coin gold} \mid \text{first coin is gold})$$

$$= \frac{P(\text{Next coin gold} \cap \text{first coin gold})}{P(\text{first coin gold})}$$

$$= \frac{1/3}{\underbrace{\tfrac{1}{3}0}_{\text{box 3 is picked}} + \underbrace{\tfrac{1}{3}\left(\tfrac{1}{2}\right)}_{\text{if box 2 is picked}} + \underbrace{\tfrac{1}{3}(1)}_{\text{if box 1 is picked}}} = \frac{1/3}{1/6 + 1/3} = \frac{1/3}{1/2} = \frac{2}{3}$$

---

[7] Wikipedia contributors. "Bertrand's box paradox." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 9 Aug. 2016. Web. 9 Aug. 2016.