2402 Assignment 2 刘恒徽.

Q1: $X \sim N(4, 11)$ $Y \sim N(6, 14)$

then $Z = X - Y \sim N(-2, 25)$

$\therefore P(4 < Z^2 < 16) = P(-4 < Z < -2) + P(2 < Z < 4)$

$P(-4 < Z < -2) = P(\frac{-4+2}{5} < S < \frac{-2+2}{5}) = P(-\frac{2}{5} < S < 0) = 0.1554$

$P(2 < Z < 4) = P(\frac{4}{5} < S < \frac{4+2}{5}) = P(\frac{4}{5} < S < \frac{6}{5}) = 0.0968$

$\therefore P(4 < Z^2 < 16) = 0.0968 + 0.1554 = 0.2522$

So the probability is 0.2522, approximately 0.25

Q2. $X \sim B(n, P)$ Easy to see that a success have occured in the $a+b-1$ events, so total number is C^a_{a+b-1}

Then $P = C^a_{a+b-1} P^a (1-P)^b$

3(a). suppose we picked $X = i$, then the other dice must be smaller than or equal to i, so the probability is $\frac{1}{6}$ and $\frac{i}{6}$, but we have to subtract the possibility of both getting i,

then $P = C^1_2 \frac{1}{6} \frac{i}{6} - \frac{1}{6} \times \frac{1}{6} = \frac{2i-1}{36}$

then $p(X) = \frac{2X-1}{36}$, $X = 1, 2, \ldots 6$

(b). $P(Y) = p(1) + \cdots + P(k) = C^1_k (\frac{1}{6})^1 (\frac{Y-1}{6})^{k-1} + \cdots + C^k_k (\frac{1}{6})^k (\frac{Y-1}{6})^0$

$= (\frac{1}{6} + \frac{Y-1}{6})^k - C^0_k (\frac{1}{6})^0 (\frac{Y-1}{6})^k$

$= (\frac{Y}{6})^k - (\frac{Y-1}{6})^k$

So $p(Y) = (\frac{Y}{6})^k - (\frac{Y-1}{6})^k$, $Y = 1, 2, 3 \ldots, 6$

4.(a). $P(\text{family 2 girls}) = P(\text{2 children}) \times P(\text{2 girls}) = (1-r) \times \frac{1}{4}$

(b). $P(\text{elder boy \& younger girl}) = (1-r) \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{4} \times (1-r)$

(c). $P(\text{at lease 1 boy}) = P(1\text{ child}) \times P(\text{boy}) + P(2\text{ child}) \times (1 - P(\text{no boy}))$
$= r \times \frac{1}{2} + (1-r) \times (1 - \frac{1}{4}) = \frac{3}{4} - \frac{1}{4}r$

5(a). $TP = P(D^+) \times P(T^+|D^+) = 1\% \times 99\% = 0.99\%$

$TN = P(D^-) \times P(T^-|D^-) = 99\% \times 95\% = 94.05\%$

$\therefore FN = P(D^+) - TP = 1\% - 0.99\% = 0.01\%$

$FP = P(D^-) - TN = 99\% - 94.05\% = 4.95\%$

(b). $P(D^+|T^+) = \frac{P(TP)}{P(T^+)} = \frac{0.99\%}{0.99\% + 4.95\%} = \frac{0.99\%}{5.94\%} = 16.67\%$

$\therefore$ the probability that the patient has the disease when positive is 16.67%

6. $P(Y=0) = \frac{1}{4} + \frac{1}{4} \times \frac{12}{13} \times \frac{12}{13} + \frac{1}{2} \times \frac{12}{13} = \frac{625}{676}$

$P(Y=1) = C_2^1 (\frac{1}{13})(\frac{12}{13}) \times \frac{1}{4} + \frac{1}{2} \times \frac{1}{13} = \frac{24}{676} + \frac{26}{676} = \frac{50}{676}$

$P(Y=2) = \frac{1}{4} \times \frac{1}{13} \times \frac{1}{13} = \frac{1}{676}$

(a). $P(Y=1|X=2) = \frac{P(Y=1, X=2)}{P(X=2)} = \frac{C_2^1 (\frac{1}{13})(\frac{12}{13}) \times \frac{1}{4}}{\frac{1}{4}} = \frac{24}{169}$

$\therefore P(Y=1|X=2) = \frac{24}{169}$

(b). $P(Y=0)=\frac{625}{676}$ $P(Y=1)=\frac{50}{676}$ $P(Y=2)=\frac{1}{676}$

(c) $P(X=2|Y=1)=\frac{P(X=2,Y=1)}{P(Y=1)}=\frac{C_2^1(\frac{1}{13})(\frac{11}{13})\times\frac{1}{4}}{\frac{50}{676}}=\frac{24}{50}=\frac{12}{25}$

$\therefore P(X=2|Y=1)$ is $\frac{12}{25}$

7. $P(\text{none defective})=0.6\times\frac{C_{20}^2}{C_{40}^2}+0.3\times\frac{C_{19}^2}{C_{20}^2}+0.1\times\frac{C_{18}^2}{C_{20}^2}$

$=0.6+0.3\times\frac{19\times18}{20\times19}+0.1\times\frac{18\times17}{20\times19}=\frac{1806}{1900}$

(a). $P(D_0|E)=\frac{P(D_0E)}{P(E)}=\frac{0.6}{\frac{1806}{1900}}\approx 63.12\%$

(b). $P(D_1|E)=\frac{P(D_1E)}{P(E)}=\frac{0.3\times\frac{18}{20}}{\frac{1806}{1900}}\approx 28.41\%$

(c). $P(D_2|E)=\frac{P(D_2E)}{P(E)}=\frac{0.1\times\frac{18\times17}{20\times19}}{\frac{1806}{1900}}\approx 8.47\%$

8. $r=\frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{\Sigma(X-\bar{X})^2\cdot\Sigma(Y-\bar{Y})^2}}=\frac{(-14.2)\times(-386)+\cdots+14.8\times384}{\sqrt{518.8\times354022}}$

≈ 0.9758 $\therefore r=0.9758$

$\because \Sigma(X-\bar{X})^2=(-14.2)^2+\cdots+(14.8)^2=518.8$

$\therefore SS(X)=\sqrt{518.8}\approx 22.7772$

$\Sigma(y-\bar{y})^2=(-386)^2+\cdots+(384)^2=354022$

$\therefore SS(y)=\sqrt{354022}\approx 594.9975$

(a). $b = r\frac{S_y}{S_x} = 0.9758 \times \frac{594.9975}{22.7772} = 25.4903$

$a = \bar{y} - b\bar{x} = 1036 - 36.2 \times 36.2 = 113.2$

$\therefore \hat{y} = 25.49x + 113.2$

(b). $r = 0.9758$

$\therefore$ corelation coefficient is 0.9758

(c). if $X=0$, $\hat{y} = 113.2$ $\therefore$ Expected sales is 113.2 million

if $X = 58$, then $\hat{y} = 25.49 \times 58 + 113.2 = 1591.62$

$\therefore$ if advertising is 58 million, then predicted sales would be 1591.62 million

9. $L(X) = f(X_1) + \cdots + f(X_n) = \theta^n e^{-\theta(X_1+\cdots+X_n)}$

$\ln L(X) = \ln \theta^n e^{-\theta(X_1+\cdots+X_n)} = \ln\theta^n + \ln e^{-\theta(X_1+\cdots+X_n)}$

$= n\ln\theta - (X_1+\cdots+X_n)\theta$

let $a = n$, $b = (X_1+\cdots+X_n)$, then when $\theta = \frac{a}{b}$,

$\ln L(X)$ achieves maximum. So when $\theta = \frac{a}{b} = \frac{n}{X_1+\cdots+X_n}$

So maximum likelihood estimator for θ is $\frac{n}{X_1+\cdots+X_n}$

10 (1). Take word2 for example: word2 = word2.lower()

n_word2 = email.count(word2) p_spam = p_spam * 0.66

p_nospam = p_nospam * 0.10

posterior_odds_spam = prior_odds_spam * p_spam

posterior_odds_no_spam = (1 - prior_odds_spam) * p_no_spame

p_isSpam = posterior_odds_spam / (posterior_odds_spam + posterior_odds_no_spame)

(2.)

	spam	0.8	0.000163	0.000130	0.101751
	¬spam	0.2	0.005753	0.001151	

spam	0.8	0.296208	0.236966	0.980170
¬spam	0.2	0.023970	0.004794	

```
# Input the email with your keyboard #
email = input('Please enter the test Email: ')     # gets the test email
email = email.lower()                              # convert string to lowercase
print(email)

# Input the spam detection keyword #
word1 = input('Please enter the first key word: ') # gets the first keyword
word1 = word1.lower()                              # convert the first key word to lowercase

word2 = input('Please enter the second key word: ') # gets the second keyword
word2 = word2.lower()                              # convert the second keyword to lowercase

word3 = input('Please enter the third key word: ') # gets the third keyword
word3 = word3.lower()                              # convert the third keyword to lowercase

word4 = input('Please enter the fourth key word: ') # gets the fourth keyword
word4 = word4.lower()                              # convert the fourth keyword to lowercase

word5 = input('Please enter the fifth key word: ') # gets the fifth keyword
word5 = word5.lower()                              # convert the fifth keyword to lowercase

# Detect whether the key word occured in your email #
n_word1 = email.count(word1)  # count the number of the first keyword in the email
n_word2 = email.count(word2)  # count the number of the second keyword in the email
n_word3 = email.count(word3)  # count the number of the third keyword in the email
n_word4 = email.count(word4)  # count the number of the fourth keyword in the email
n_word5 = email.count(word5)  # count the number of the fifth keyword in the email

# The prior odds of spam is 80% #
p_spam = 1.0                                       # initialize the likelihood ratio of the spam email
p_no_spam = 1.0                                    # initialize the likelihood ratio of the no spam email
prior_odds_spam = 0.8                      # prior odds of spam

# Calculate the Likelihood ratio #
if n_word1 != 0:                                   # the first keyword occurred in the email
    p_spam = p_spam * 0.88
    p_no_spam = p_no_spam * 0.47
if n_word2 != 0:                                   # the second keyword occurred in the email
    p_spam = p_spam * 0.66
    p_no_spam = p_no_spam * 0.10
if n_word3 != 0:                                   # the third keyword occurred in the email
    p_spam = p_spam * 0.11
    p_no_spam = p_no_spam * 0.60
if n_word4 != 0:                                   # the fourth keyword occurred in the email
```

```
    p_spam = p_spam * 0.51
    p_no_spam = p_no_spam * 0.51
if n_word5 != 0:                                    # the fifth keyword occurred in the email
    p_spam = p_spam * 0.005
    p_no_spam = p_no_spam * 0.40

# Calculate the Posterior odds #
posterior_odds_spam = prior_odds_spam * p_spam                  # posterior odds of spam
posterior_odds_no_spam = (1 - prior_odds_spam) * p_no_spam       # posterior odds of no spam

p_isSpam = posterior_odds_spam / (posterior_odds_spam + posterior_odds_no_spam)                                  #
probability of spam P(spam | email)

print("p_spam: %.6f, p_no_spam: %.6f, posterior_odds_spam: %.6f, posterior_odds_no_spam: %.6f,
p_isSpam: %.6f"%(p_spam, p_no_spam, posterior_odds_spam, posterior_odds_no_spam, p_isSpam))
```

$
donate
research
contact
CS2402

Dear research student, we have just uploaded the competition questions for CS2402. Please sign up and contact me before this Friday (29, March, 2024) if you have any questions. The award for the winner is $10,000,000! Also, you can donate your award to the whole class.

I have decided to donate what I have to you. I was diagnosed with cancer of the lungs few years ago. I have been inspired by God to donate my inheritance to you for the good work of God and charity purpose. I am doing this because my family are unbelievers and I will not allow them inherit this money for their own selfishness. I decided to bequeath the sum of $10,000,000.00 to you. If you are much more interested, Contact Thomas with this specified email: thmasbfd@gmail.com).

Result1:

```
dear research student, we have just uploaded the competition questions for cs2402. please sign up and contact
me before this friday (29, march, 2024) if you have any questions. the award for the winner is $10,000,000! also,
you can donate your award to the whole class.
p_spam: 0.000163, p_no_spam: 0.005753, posterior_odds_spam: 0.000130, posterior_odds_no_spam: 0.001151, p_isSpam:
0.101751
```

Result2:

i have decided to donate what i have to you. i was diagnosed with cancer of the lungs few years ago. i have been inspired by god to donate my inheritance to you for the good work of god and charity purpose. i am doing this because my family are unbelievers and i will not allow them inherit this money for their own selfishness. i decided to bequeath the sum of $10,000,000.00 to you. if you are much more interested, contact thomas with this specified email: thmasbfd@gmail.com).

p_spam: 0.296208, p_no_spam: 0.023970, posterior_odds_spam: 0.236966, posterior_odds_no_spam: 0.004794, p_isSpam: 0.980170