上海科技大学
ShanghaiTech University

SI140 Probability & Mathematical Statistics

Homework 10

陈昱聪

Chen Yucong ><E<>N

Student ID: 2019533079

Email: chenyc@shanghaitech.edu.cn

School of Information Science and Technology

◎ Group#2 (TA: 曾理)

Problem 7.66

Solution:

(a)

$$p_1 = \frac{a+b}{a+b+c+d} \qquad p_2 = \frac{c}{a+b+c+d} \qquad p_3 = \frac{d}{a+b+c+d}$$

$$\begin{aligned}
&P(X = n_1, Y = n_2, Z = n_3) \\
&= \frac{n!}{n_1!n_2!n_3!} p_1^{n_1} p_2^{n_2} p_3^{n_3} \\
&= \frac{n!}{n_1!n_2!n_3!} \left(\frac{a+b}{a+b+c+d}\right)^{n_1} \left(\frac{c}{a+b+c+d}\right)^{n_2} \left(\frac{d}{a+b+c+d}\right)^{n_3}
\end{aligned}$$

For all non-negative integers such that $n_1 + n_2 + n_3 = n$
$P(X = n_1, Y = n_2, Z = n_3) = 0$ otherwise.

(b) This could be seen as a "HGeom" distribution with 3 categories, thus

$$P(X = n_1, Y = n_2, Z = n_3) = \frac{\binom{a+b}{n_1}\binom{c}{n_2}\binom{d}{n_3}}{\binom{a+b+c+d}{n}}$$

For all non-negative integers such that $n_1 + n_2 + n_3 = n$
$P(X = n_1, Y = n_2, Z = n_3) = 0$ otherwise.

Problem 7.70

Solution:

(a)

$$
\begin{aligned}
P(X_1 = n_1, X_2 = n_2, X_3 = n_3) &= \frac{n!}{n_1!n_2!n_3!}(p^2)^{n_1}\left[2p(1-p)\right]^{n_2}\left[(1-p)^2\right]^{n_3} \\
&= \frac{n!}{n_1!n_2!n_3!}2^{n_2}p^{2n_1+n_2}(1-p)^{n_2+2n_3}
\end{aligned}
$$

For all non-negative integers such that $n_1 + n_2 + n_3 = n$
$P(X_1 = n_1, X_2 = n_2, X_3 = n_3) = 0$ otherwise.

(b) Let X be #have an A and with $p(2-p)$ in success and $q_B = (1-p)^2$ in failure. Since $p_c + q_c = 1$, this is binomial given $X \sim Bin(n, p(2-p))$.

(c) Let Y be #A in the $2n$ genes. Since the frequency of A in population is p, we get Y is binomial given $Y \sim Bin(2n, p)$

(d) Find the estimator of p by caculating the sample proportion. We know that #A in the sample of size n is $2X_1 + X_2$, and the total number of genes is $2(X_1 + X_2 + X_3) = 2n$. We get

$$
\hat{p} = \frac{2X_1 + X_2}{2(X_1 + X_2 + X_3)} = \frac{2X_1 + X_2}{2n}
$$

(e) Since we can't get any information within X_1 and X_2, now we consider X_3. Find the estimator of p by caculating the sample proportion. We get

$$
(1-\hat{p})^2 = \frac{X_3}{(X_1 + X_2 + X_3)} = \frac{X_3}{n}
$$

$$
\Rightarrow \hat{p} = 1 - \sqrt{\frac{X_3}{n}}
$$

Problem 7.71

Solution:

(a) For any arbitrary linear combination of $X+Y$ and $X-Y$:

$$t(X+Y)+s(X-Y)$$

can also be written as a linear combination of X and Y:

$$(t+s)X+(t-s)Y$$

which is Normal since (X,Y) is Bivariate Normal. So $(X+Y,X-Y)$ is also Bivariate Normal.

(b) $\text{Cov}(X+Y,X-Y)=\text{Var}(X)-\text{Var}(Y)=0$, so $X+Y$ and $X-Y$ are independent.

$$\text{Var}(X+Y)=\text{Var}(X)+\text{Var}(Y)+2\rho=2+2\rho$$

$$\text{Var}(X-Y)=\text{Var}(X)+\text{Var}(Y)-2\rho=2-2\rho$$

$$\Rightarrow (X+Y)\sim\mathcal{N}(0,2+2\rho)\qquad (X-Y)\sim\mathcal{N}(0,2-2\rho)$$

$$\text{So the joint PDF is: } f(a,b)=\frac{1}{4\pi\sqrt{1-\rho^2}}e^{-\frac{1}{4}\left(a^2/(1+\rho)+b^2/(1-\rho)\right)}$$

Problem 7.75

Solution:

(a) For any arbitrary linear combination of X, Y and $X+Y$:

$$t_1X + t_2Y + t_3(X+Y)$$

can also be written as a linear combination of X and Y:

$$(t_1+t_3)X + (t_2+t_3)Y$$

which is Normal since $(t_1+t_3)X+(t_2+t_3)Y$ is a linear combination of two i.i.d. r.v.s. of Normal. So $(X, Y, X+Y)$ is Multivariate Normal.

(b) Let $t_1 = t_2 = -1, t_3 = 1$, since $P(-X-Y+SX+SY=0) = P(S=1) = \frac{1}{2}$, this combination is not a continuous r.v. so that $(X, Y, SX+SY)$ is **not** Multivariate Normal.

(c) Find the PDF of $Z = t_1SX + t_2SY$, $W = t_1X + t_2Y$

$$\begin{aligned}
f_Z(a) =& f_Z(a|S=1)P(S=1) + f_Z(a|S=-1)P(S=-1) \\
=& \frac{1}{2}\, f_Z(a|S=1) + \frac{1}{2}\, f_Z(-a|S=-1) \\
=& \frac{1}{2}\, f_W(a) + \frac{1}{2}\, f_W(-a) \\
=& f_W(a) \qquad \text{(By considering } W \sim \mathcal{N}(0, t_1^2\, t_2^2)\text{, it is symmetric about 0)}
\end{aligned}$$

So for any t_1 and t_2, $t_1SX + t_2SY$ is just the same as $t_1X + t_2Y$, since the latter one is Normal, we get $t_1SX + t_2SY$ is Normal, so (SX, SY) is Multivariate Normal.

Problem 7.76

Solution:

First to find c such that $\text{Cov}(Y - cX, X) = 0$:

$$\text{Cov}(Y - cX, X) = \text{Cov}(Y, X) - c\text{Var}(X) = \rho\text{Std}(X)\text{Std}(Y) - c\text{Var}(X) = \rho\sigma_1\sigma_2 - c\sigma_1^2 = 0$$

$$\Rightarrow c = \rho\,\frac{\sigma_2}{\sigma_1}$$

Since X and Y are independent r.v.s. with Normal, their linear combinations are still Normal, thus we know $(Y - cX, X)$ is Bivariate Normal. From Theorem 7.5.7. we know that when $c = \rho\,\frac{\sigma_2}{\sigma_1}$, $Y - cX$ and X are independent.

Problem 7.79

Solution:

(Let D, R, N be r.v.s. of #registered Democrats, #registered Repubicans, #people showed up at the polls)

(a) We know that for a registered voter, the probability of showing up to the polls and being a Democrat is ps. By using the conclusion of Chicken-Egg story, X is Poisson and

$$X \sim \text{Pois}(ps\lambda)$$

(b) By using the conclusion of Chicken-Egg story, $X|V$ is Binomial and

$$X|V \sim \text{Bin}(v, ps)$$

(c) Since the data of Repubicans does not affect X due to the independence, we know that this is a Binomial with the probability of success is s, that is

$$X|D, R \sim \text{Bin}(d, s)$$

(d) After knowing the true number of Democrats in those v registered voters, that is d, we can update the proportion of Democrats to be $\frac{d}{v}$ instead of p. So we have the possibility of $\frac{d}{v}$ Democrats and selected n people (i.e. who showed up) without replacement. So this is a Hypergeometric distribution that is

$$X|D, R, N \sim \text{HGeom}(d, r, n)$$