3

(a)

$$f(x_{11}, x_{12}, x_{21}, x_{22}) = \frac{n!}{\prod_{i,j} x_{ij}!} \prod_{(i,j)} P_{ij}^{x_{ij}}$$

$$= \frac{n!}{\prod_{(i,j)} x_{ij}!} \left[\exp\left\{ x_{11}\log P_{11} + x_{12}\log P_{12} + x_{21}\log P_{21} + x_{22}\log P_{22} \right\} \right.$$

$$= \left(\frac{n!}{\prod_{(i,j)} x_{ij}!}\right) \left[\exp\left\{ x_{11}\log P_{11} + x_{12}\log P_{12} + x_{21}\log P_{21} + (n - x_{11} - x_{12} - x_{21})\log P_{22} \right\} \right.$$

$$= \left(\frac{n!}{\prod_{(i,j)} x_{ij}!}\right) P_{22}^{n} \left[\exp\left\{ x_{11}\log\left(\frac{P_{11}}{P_{22}}\right) + x_{12}\log\left(\frac{P_{12}}{P_{22}}\right) + x_{21}\log\left(\frac{P_{21}}{P_{22}}\right) \right\} \right]$$

This is in the exponential family of distributions with dimension (rank) of 3.

(b) $\Rightarrow$

Show that if A and B are independent, then

$$\log\left(\frac{P_{11}}{P_{22}}\right) = \log\left(\frac{P_{12}}{P_{22}}\right) + \log\left(\frac{P_{21}}{P_{22}}\right). \qquad (1)$$

If $A \perp\!\!\!\perp B$, then

$$P(A \cap B) = P(A)\,P(B) \qquad (2)$$

Now $P(A \cap B) = P_{11}$, $P(A) = P_{11} + P_{21}$, $P(B) = P_{11} + P_{12}$

Thus (2) implies

$$P_{11} = (P_{11} + P_{21})(P_{11} + P_{12})$$

$$\Rightarrow \quad P_{11} = P_{11}^2 + P_{21}P_{11} + P_{11}P_{12} + P_{21}P_{12}$$

$$\Rightarrow \quad P_{11}(P_{11} - 1 + P_{21} + P_{12}) + P_{21}P_{12} = 0$$

$$\Rightarrow \quad P_{11}(-P_{22}) + P_{21}P_{12} = 0$$

$$\Rightarrow \quad P_{21}P_{12} = P_{11}P_{22} \qquad (3.)$$

Now by exponentiating both sides of (1), we get

$$\frac{P_{11}}{P_{12}} = \frac{P_{12}}{P_{22}} \frac{P_{21}}{P_{22}}$$

$\Rightarrow P_{12} P_{21} = P_{11} P_{22}$

Thus (3) and (1) are equivalent.

$\Leftarrow$

Show that if $P_{21} P_{12} = P_{11} P_{22}$, then $A \perp\!\!\!\perp B$.

We need to show $P(A \cap B) = P(A)\,P(B)$

and thus we need to show

$$P_{11} = (P_{11} + P_{21})(P_{11} + P_{12}) \qquad (1)$$

Let us show that the right side equals the left using the fact that $P_{12} P_{21} = P_{11} P_{22}$. The right side of (1) equals:

$$P_{11}^2 + P_{21} P_{11} + P_{11} P_{12} + P_{21} P_{12}$$

$$= P_{11}^2 + P_{21} P_{11} + P_{11} P_{12} + P_{11} P_{22}$$

$$= P_{11}^2 + P_{11}(P_{21} + P_{12} + P_{22})$$

$= P_{11}^2 + P_{11}(P_{21} + P_{12} + 1 - P_{11} - P_{12} - P_{21})$

$= P_{11}^2 + P_{11}(1 - P_{11})$

$= P_{11}$.

This completes the proof

(c) We want to test the hypothesis

$H_0: \theta = 0$
$H_1: \theta \neq 0$

where $\boxed{\theta = \log\left(\frac{P_{11}}{P_{22}}\right) - \log\left(\frac{P_{12}}{P_{22}}\right) - \log\left(\frac{P_{21}}{P_{22}}\right)}$ (1)

Thus, we want to test for independence of A and B.

We use Theorem 2.7 on pages 328-331 of the Bios 761 notes.

We can write $f(x_{11}, x_{12}, x_{21}, x_{22})$ in the form

$f(x_{11}, x_{12}, x_{21}, x_{22})$

$$= \left(\frac{n!}{\prod_{(i,j)} x_{ij}!} \right) p_{22}^{n} \exp\left\{ x_{11}\theta + (x_{11}+x_{12}) \log\left(\frac{p_{12}}{p_{22}}\right) + (x_{11}+x_{21}) \log\left(\frac{p_{21}}{p_{22}}\right) \right\}$$

where θ is defined by (1). Thus, according to Theorem 2.7, we can derive the rejection region of the UMPU test by considering the conditional distribution of $[X_{11} \mid X_{11}+X_{12}=n_1,\ X_{11}+X_{21}=m_1]$

Instead of computing this distribution directly, we consider first the conditional distribution subject only to the condition $X_{11}+X_{12}=n_1$, and hence $X_{21}+X_{22}=n-n_1=n_2$. Thus

$P(X_{11}=k,\ X_{21}=l \mid X_{11}+X_{12}=n_1)$

$$= \binom{n_1}{k}\binom{n_2}{l}\left(\frac{p_{11}}{p_{11}+p_{12}}\right)^{k}\left(\frac{p_{12}}{p_{11}+p_{12}}\right)^{n_1-k} \cdot$$

$$\left(\frac{p_{21}}{p_{21}+p_{22}}\right)^{l}\left(\frac{p_{22}}{p_{21}+p_{22}}\right)^{n_2-l},$$

which is the distribution of two independent binomial variables, the number of successes in in n_1 and n_2 trials with probability $p_1 = \frac{p_{11}}{p_{11}+p_{12}}$ and $p_2 = \frac{p_{21}}{p_{21}+p_{22}}$.

Actually this is clear without computation since we are now dealing with samples of fixed size n_1 and n_2 from the subpopulations B and B^c and the probability of A in these subpopulations is p_1 and p_2. If we now add the additional restriction $X_{11}+X_{21}=m_1$, then the conditional distribution of X_{11} subject to the two conditions $X_{11}+X_{12}=n_1$, $X_{11}+X_{21}=m_1$

is the Same as that of $X_{11} \mid X_{11}+X_{21}=m_1$

This Probability is thus calculated as

$$P(X_{11}=k \mid X_{11}+X_{21}=m_1)$$

$$\frac{P(X_{11}=k,\ X_{11}+X_{21}=m_1)}{P(X_{11}+X_{21}=m_1)} = \frac{P(X_{11}=k,\ X_{21}=m_1-k)}{P(X_{11}+X_{21}=m_1)}$$

$$= \frac{\binom{n_1}{k}\binom{n_2}{m_1-k} P_1^{k}(1-P_1)^{n_1-k} P_2^{m_1-k}(1-P_2)^{n_2-m_1+k}}{\sum_{j=0}^{S}\binom{n_1}{j}\binom{n_2}{m_1-j} P_1^{j}(1-P_1)^{n_1-j} P_2^{m_1-j}(1-P_2)^{n_2-m_1+j}}$$

$S = \min(m_1, n_1)$

$$= \boxed{G_{m_1}(\rho)\binom{n_1}{k}\binom{n_2}{m_1-k}\rho^{m_1-k} \qquad k=0,\ldots S}$$

where

$$\boxed{\rho = \frac{\left(\frac{P_2}{1-P_2}\right)}{\left(\frac{P_1}{1-P_1}\right)}}$$

and

$$G_{m_1}(\rho) = \left(\sum_{j=0}^{s} \binom{n_1}{j} \binom{n_2}{m_1 - j} \rho^j \right)^{-1}.$$

Under the null hypotheses of $\theta = 0$, we have $P_1 = P_2$, so that $\rho = 1$, and the conditional distribution of

$[X_{11} \mid X_{11} + X_{12} = n_1, \; X_{11} + X_{21} = m_1, \; H_0]$

$$= \frac{\binom{n_1}{k}\binom{n_2}{m_1 - k}}{\binom{n_1 + n_2}{m_1}} = HG(n_1, n_2, m_1)$$

hypergeometric dist

so that the test becomes Fisher's exact test

Now the rejection region is of the form

$$\phi(X_{11}) = \begin{cases} 1 & \text{if } X_{11} < c_1 \text{ or } X_{11} > c_2 \\ \gamma_i & \text{if } X_{11} = c_i, \; i = 1, 2 \\ 0 & \text{otherwise} \end{cases}$$

where

$n_1 = X_{11} + X_{12}$, $\quad c_i \equiv c_i(n_1, m_1)$, $i = 1, 2$.

$m_1 = X_{11} + X_{21}$

The c_i's are determined from the

equations

$$E_{H_0}\left[\phi(X_{11}) \,\middle|\, X_{11}+X_{12}=n_1, X_{11}+X_{21}=m_1\right] = \alpha$$

and
$$E_{H_0}\left[X_{11}\,\phi(X_{11}) \,\middle|\, X_{11}+X_{12}=n_1, X_{11}+X_{21}=m_1\right]$$

$$= \alpha\, E_{H_0}\left[X_{11} \,\middle|\, X_{11}+X_{12}=n_1, X_{11}+X_{21}=m_1\right]$$

Where E_{H_0} denotes expectation with respect to H_0.

Under H_0

$$\left[X_{11} \,\middle|\, X_{11}+X_{12}=n_1, X_{11}+X_{21}=m_1\right] \sim HG(n_1, n_2, m_1).$$

The power function of the test is

$$E_{H_1} \phi(X_{11})$$

$$= P\left(X_{11} < c_1 \text{ or } X_{11} > c_2 \mid X_{11}+X_{12}=n_1, X_{11}+X_{21}=m_1\right]$$

and this conditional probability is calculated from the noncentral hypergeometric distribution

$$\left[X_{11} \mid X_{11}+X_{12}=n_1, X_{11}+X_{21}=m_1, H_1\right]$$

$$= G_{m_1}(\rho) \binom{n_1}{k} \binom{n_2}{m_1-k} \rho^{m_1-k}, \quad k=0,\dots,s$$

$$= \text{Non-central } HG(n_1, n_2, m_1, \rho)$$

When $\rho=1$, it is a central HG distribution

d) $H_0: P(A) \geq P(B)$

$H_1: P(A) < P(B)$

We first express the hypothesis in terms of θ.

Note that $P(A) = P(B)$ implies that

$$P_{11} + P_{21} = P_{11} + P_{12}$$

$$\Leftrightarrow \quad P_{21} = P_{12}.$$

In terms of θ, this corresponds to selecting $a_0 = 0$, $a_1 = 1$, $a_2 = -1$ from the general form of θ given in part c). Thus $\theta = \log\left(\frac{P_{12}}{P_{22}}\right) - \log\left(\frac{P_{21}}{P_{22}}\right)$.

Thus $$P_{21} = P_{12} \Leftrightarrow \underbrace{\log\left(\frac{P_{12}}{P_{22}}\right) - \log\left(\frac{P_{21}}{P_{22}}\right)}_{\theta} = 0$$

Now $P(A) \geq P(B)$ iff

$$P_{21} \geq P_{12}, \text{ iff } P_{12} - P_{21} \leq 0$$

iff $\theta \leq 0$, and therefore the hypotheses can be restated as

$$H_0: \theta \leq 0$$

$$H_1: \theta > 0$$

Now we can write

$$f(x_{11}, x_{12}, x_{21}, x_{22}) = \left(\frac{n!}{\prod\limits_{(i,j)} x_{ij}!}\right) \cdot p_{22}^{\,n}$$

$$\cdot \exp\left\{ x_{11} \log\left(\frac{p_{11}}{p_{22}}\right) + x_{12}\left(\log\left(\frac{p_{12}}{p_{22}}\right) - \log\left(\frac{p_{21}}{p_{22}}\right)\right) + x_{21} \log\left(\frac{p_{21}}{p_{22}}\right) + x_{12} \log\left(\frac{p_{21}}{p_{22}}\right) \right\}$$

$$= \left(\frac{n!}{\pi x_{ij}!}\right) p_{22}^{\,n} \exp\left\{ \theta\, x_{12} + x_{11} \log\left(\frac{p_{11}}{p_{2}}\right) + (x_{21} + x_{12}) \log\left(\frac{p_{21}}{p_{22}}\right) \right\}$$

thus, according to theorem 2-7 of the notes, we can construct a UMPU size α test by considering $[X_{12} | X_{11}, X_{21} + X_{12}]$

Thus

$$P\{X_{12} = k \mid X_{11} = j,\ X_{21} + X_{12} = \ell\} = \frac{P\{X_{12} = k,\ X_{11} = j,\ X_{21} = \ell - k,\ X_{22} = n - \ell - j\}}{P\{X_{11} = j,\ X_{21} + X_{12} = \ell,\ X_{22} = n - \ell - j\}}$$

$$= \frac{\left[\frac{n!}{k!\,j!\,(\ell-k)!\,(n-\ell-j)!}\right] p_{11}^{j}\, p_{12}^{k}\, p_{21}^{\ell-k}\, p_{22}^{n-\ell-j}}{\sum_{r=0}^{\ell}\left[\frac{n!}{j!\,r!\,(\ell-r)!\,(n-\ell-j)!}\right] p_{11}^{j}\, p_{12}^{r}\, p_{21}^{\ell-r}\, p_{22}^{n-\ell-j}}$$

$$= \frac{\left[\frac{1}{k!\,(\ell-k)!}\right] p_{12}^{k}\, p_{21}^{\ell-k}}{\sum_{r=0}^{\ell}\frac{1}{r!\,(\ell-r)!}\, p_{12}^{r}\, p_{21}^{\ell-r}}$$

$$= \frac{\left[\frac{\ell!}{k!\,(\ell-k)!}\right] p_{12}^{k}\, p_{21}^{\ell-k}}{\sum_{r=0}^{\ell}\left[\frac{\ell!}{r!\,(\ell-r)!}\right] p_{12}^{r}\, p_{21}^{\ell-r}}$$

$$= \frac{\binom{\ell}{k} p_{12}^{k}\, p_{21}^{\ell-k}}{(p_{12} + p_{21})^{\ell}}$$

$$= \binom{\ell}{k}\left(\frac{p_{12}}{p_{12}+p_{21}}\right)^{k}\left(\frac{p_{21}}{p_{12}+p_{21}}\right)^{\ell-k}$$

Thus $X_{12} | X_{11}, X_{12} + X_{21} = \ell \sim Bin\left(p = \frac{p_{12}}{p_{12}+p_{21}},\ n = \ell\right)$

Following Theorem 2.7, the UMPU size α test is of the form

$$\phi(X_{12}) = \begin{cases} 1 & \text{if } X_{12} > c \\ \gamma & \text{if } X_{12} = c \\ 0 & \text{otherwise} \end{cases}$$

where

$$c \equiv c(X_{11}, X_{21} + X_{12}) \equiv c(j, l)$$

where c is chosen to make the test size α by solving

$$E_{H_0}\left[\phi(X_{12}) \mid X_{11} = j, X_{21} + X_{12} = l\right] = \alpha$$

Thus c is determined by solving

$$\boxed{\begin{aligned} & P(X_{12} > c \mid X_{11} = j, X_{21} + X_{12} = l] \\ & + \gamma P(X_{12} = c \mid X_{11} = j, X_{21} + X_{12} = l] = \alpha \end{aligned}}$$

(e) The likelihood ratio test for testing $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$ is given by

(i)

$$\Lambda_n = \frac{\dfrac{n!}{\prod X_{ij}!} \exp\left\{ (X_{11}+X_{12}) \log\left(\dfrac{\hat{P}_{12}}{\hat{P}_{22}}\right) + (X_{11}+X_{21}) \log\left(\dfrac{\hat{P}_{21}}{\hat{P}_{22}}\right) \right\}}{\left(\dfrac{n!}{\prod X_{ij}!}\right) \exp\left\{ X_{11} \log\left(\dfrac{X_{11}}{X_{22}}\right) + X_{12} \log\left(\dfrac{X_{12}}{X_{22}}\right) + X_{21} \log\left(\dfrac{X_{21}}{X_{22}}\right) \right\}}$$

the MLE's under H_0 are

$$\hat{P}_{11} = \left(\frac{n_1 m_1}{n^2}\right) = \frac{(X_{11}+X_{12})(X_{11}+X_{21})}{n^2}$$

$$\hat{P}_{12} = \left(\frac{n_1 m_2}{n^2}\right) = \frac{(X_{11}+X_{12})(X_{12}+X_{22})}{n^2}$$

$$\hat{P}_{21} = \left(\frac{n_2 m_1}{n^2}\right) = \frac{(X_{21}+X_{22})(X_{11}+X_{21})}{n^2}$$

$$\hat{P}_{22} = \left(\frac{n_2 m_2}{n^2}\right) = \frac{(X_{21}+X_{22})(X_{12}+X_{22})}{n^2}$$

the MLE's of P_{ij} under H_1 are

$$\hat{P}_{ij}^{(H_1)} = \frac{X_{ij}}{n}, \quad i=1,2, \quad j=1,2$$

after a little algebra, we can write Λ_n as

$$\Lambda_n = \frac{\prod_{j=1}^{2}\prod_{i=1}^{2} \hat{P}_{ij}^{X_{ij}}}{\left(\frac{X_{ij}}{n}\right)^{X_{ij}}}, \quad \text{where } \hat{P}_{ij} \text{ is given on the previous page}$$

using the fact that $\frac{\hat{P}_{ij}}{\left(\frac{X_{ij}}{n}\right)} \xrightarrow{P} 1$ under H_0,

and

$$\log(1+x) = x - x^2/2 + o(|x|^2) \quad \text{as } |x| \to 0$$

we obtain that

$$-2\log\Lambda_n = -2\sum_{j=1}^{2}\sum_{i=1}^{2} X_{ij}\log\left(1+\frac{\hat{P}_{ij}}{X_{ij}/n}-1\right)$$

$$= -2\sum_{i=1}^{2}\sum_{j=1}^{2} X_{ij}\left(\frac{\hat{P}_{ij}}{X_{ij}/n}-1\right) + \sum_{j=1}^{2}\sum_{i=1}^{2} X_{ij}\left(\frac{\hat{P}_{ij}}{X_{ij}/n}-1\right)^2$$

$$+ o_p(1)$$

$$= \sum_{i=1}^{2}\sum_{j=1}^{2} \left[\frac{(X_{ij} - n\hat{P}_{ij})^2}{X_{ij}}\right] + o_p(1)$$

$$= \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(X_{ij} - n\hat{P}_{ij})^2}{n\hat{P}_{ij}} + o_p(1)$$

Where this last equality follows from the fact that

$$\sum_{j=1}^{2}\sum_{i=1}^{2} \hat{P}_{ij} = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{X_{ij}}{n} = 1$$

(ii) asymptotically

$$-2\log\Lambda_n \overset{H_0}{\sim} \chi_1^2$$

and under H_1,

$$-2\log\Lambda_n \overset{H}{\sim} \chi_1^2(\delta)$$

$$\delta = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(nP_{ij} - nP_{i+}P_{+j})^2}{nP_{ij}}$$