Problem 1

Problem 1a

$\boldsymbol{X}_i = (X_{i1}, \cdots, X_{id})$ for $i = 1, \cdots, k$
$(\boldsymbol{X}_1, \cdots, \boldsymbol{X}_k) \sim \text{Multinomial}(n, \boldsymbol{\pi}_1, \cdots, \boldsymbol{\pi}_k)$ where $\boldsymbol{\pi}_i = (\pi_{i1}, \cdots \pi_{id})$
We know that if $\boldsymbol{Y} \sim \text{Multinomial}(n, \boldsymbol{\pi})$ then the $M_{\boldsymbol{Y}}(\boldsymbol{t}) = \mathbb{E}\left[e^{\boldsymbol{t}^T\boldsymbol{Y}}\right] = \left(\sum_{i=1}^k \pi_i e^{t_i}\right)^n$.

Let $\boldsymbol{Y}^* = (X_{1+}, \cdots, X_{k+})$ and let $\boldsymbol{t}^* = \begin{bmatrix} t_1 & t_2 & \cdots & t_k \\ \vdots & & & \\ t_1 & t_2 & \cdots & t_k \end{bmatrix}_{d \times k}$

$$\begin{aligned} M_{\boldsymbol{Y}^*}(\boldsymbol{t}^*) &= \mathbb{E}\left[\exp\{(\boldsymbol{t}^*)^T \boldsymbol{Y}^*\}\right] \\ &= \left(\sum_{i=1}^k \sum_{j=1}^d \pi_{ij} e^{t_{ij}}\right)^n \\ &= \left(\sum_{i=1}^k e^{t_i} \sum_{j=1}^d \pi_{ij}\right)^n && t_{ij} = t_i \\ &= \left(\sum_{i=1}^k e^{t_i} \pi_{i+}\right)^n && \sim \text{Multinomial}(n, \pi_{1+}, \cdots, \pi_{k+}) \end{aligned}$$

Problem 1b

Let $\boldsymbol{Y}^* = (Y_a, Y_b, Y_c)$ where $Y_a = X_1 + X_3,\ Y_b = X_2,\ Y_c = X_4 + X_5$ and

$$\begin{aligned}
M_{\boldsymbol{Y}^*}(\boldsymbol{t}) &= \mathbb{E}\left[\exp\{\boldsymbol{t}^T\boldsymbol{Y}^*\}\right] \\
&= \mathbb{E}\left[\exp\{\boldsymbol{t}^T(Y_a + Y_b + Y_c)\}\right] \\
&= \mathbb{E}\left[\exp\{\boldsymbol{t}^T X_1 + \boldsymbol{t}^T X_3 + \boldsymbol{t}^T X_2 + \boldsymbol{t}^T X_4 + \boldsymbol{t}^T X_5\}\right] \\
&= \mathbb{E}\left[\prod_{i=1}^{k} \exp\{\boldsymbol{t}^T X_i\}\right] \\
&= \prod_{i=1}^{n} \mathbb{E}\left[\exp\{\boldsymbol{t}^T X_i\}\right] \\
&= \prod_{i=1}^{n} M_{X_i}(t) \\
&= \left(\sum_{j=1}^{k} \pi_j e^{t_j}\right)^n \\
&\sim \text{Multinomial}(n, \pi_1 + \pi_3, \pi_2, \pi_4 + \pi_5, \textstyle\sum_{j=1}^5 \pi_j \leq 1)
\end{aligned}$$

Since $\sum_{j=1}^6 \pi_j = 1$ we have $\sum_{j=1}^5 \pi_j \leq 1$.

Problem 1c

Let $Y = F_X(X)$ and $U \sim \text{Uniform}(0,1)$

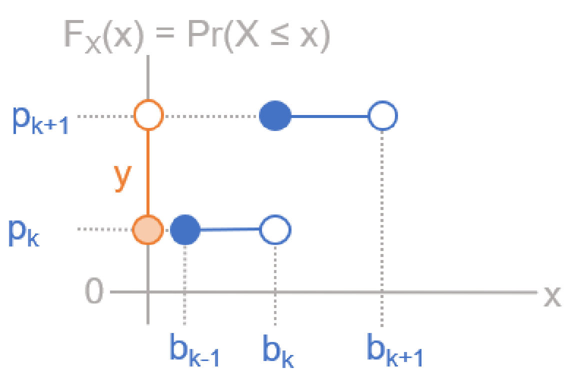
$$F_X(x) = \mathbb{P}(X \le x) = \begin{cases} p_1 & x < b_1 \\ p_2 & b_1 \le x < b_2 \\ \vdots & \\ p_k & b_{k-1} \le x < b_k \\ \vdots & \\ p_j & b_j \le x \end{cases}$$

Want to show: $F_Y(y) \le \mathbb{P}(U \le y) = y \ \forall \ 0 < y < 1$

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \le y) \\ &= \mathbb{P}(F_X(x) \le y) \qquad \text{Let } p_k \le y < p_{k+1} \\ &= \mathbb{P}(x_k \le x < x_{k+1}) = p_k \le y \end{aligned}$$

Want to show: $F_Y(y) < \mathbb{P}(U \le y) = y$ for some $0 < y < 1$

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \le y) \\ &= \mathbb{P}(F_X(x) \le y) \qquad \text{Let } p_k < y < p_{k+1} \\ &= \mathbb{P}(x_k \le x < x_{k+1}) = p_k < y \end{aligned}$$



Problem #1.7

Problem #1.7.a

$$
\begin{aligned}
L(\pi) &= \prod_{i=1}^{n} \pi^{y_i} \cdot (1-\pi)^{1-y_i} \\
\ell(\pi) &= \left(\textstyle\sum_{i=1}^{n} y_i\right) \log \pi + \left(\textstyle\sum_{i=1}^{n} (1-y_i)\right) \log(1-\pi) \\
\frac{\partial \ell(\pi)}{\partial \pi} &= \frac{\sum_{i=1}^{n} y_i}{\pi} + \frac{\sum_{i=1}^{n}(1-y_i)}{1-pi} \overset{set}{=} 0 \\
\implies \textstyle\sum_{i=1}^{n} y_i - \pi \sum_{i=1}^{n} y_i &= \pi \textstyle\sum_{i=1}^{n}(1-y_i) \\
\textstyle\sum_{i=1}^{n} y_i &= \pi \underbrace{\left(\textstyle\sum_{i=1}^{n}(1-y_i) + \sum_{i=1}^{n} y_i\right)}_{=n} \\
\hat{\pi} &= \frac{1}{n} \textstyle\sum_{i=1}^{n} y_i \\
\hat{\pi} &= \frac{1}{20} \cdot 20 = 1
\end{aligned}
$$

Problem #1.7.b

Wald Statistics:

$$
\begin{aligned}
W^2 &= \frac{n\left(\hat{\pi} - \pi_0\right)^2}{\hat{\pi}(1-\hat{\pi})} \\
&= \frac{20 \cdot (1-0.5)^2}{1 \cdot (1-1)} = \frac{5}{0} \quad \text{DNE, } \infty
\end{aligned}
$$

Wald CI:

$$
\begin{aligned}
&\hat{\pi} \pm z_{.025} \sqrt{\hat{\pi}(1-\hat{\pi})/n} \\
&1 \pm 1.96 \cdot 0 \\
\implies &\text{Wald CI: } [1, 1]
\end{aligned}
$$

This is not sensible to use since we are given values that we cannot quantify (i.e. in order to find $\hat{\pi}$, dividing by zero).

Problem #1.7.c

S^2	S	apval	CI
20	4.4721	<0.0001	[0.8389, 1]

Problem #1.7.d

L^2	L	CI
27.7259	5.2655	[0.9084, 1]

Problem #1.7.e

CP Confidence Interval: [0.8316, 1] and pval < 0.0001

Problem #1.7.f

$$z_{\alpha/2} = \frac{\hat{\pi} - \pi_T}{\sqrt{\pi_T(1-\pi_T)/n}}$$

$$n = \frac{z_a^2 \pi_T(1-\pi_T)}{(\hat{\pi} - \pi_T)^2}$$

$$n = 138.2925$$

A sample size of 138 is needed.

Problem #1.8

$H_0 : \pi_G = 0.75$ vs $H_1 : \pi_G \neq 0.75$

$\hat{\pi}_G = 0.7743$

$Z = 1.8601$

apval $= 0.0629$

Wald CI: [0.7496, 0.7989]

Using the traditional threshold of 0.05, we would not reject the null that $\pi_G = 0.75$ (i.e. that the ratio of Green to Yellow is $3 : 1$).