HW1

Brody Kendall

10/7/2021

1. Let $X_i = (X_{i1}, ..., X_{id})$ be a d-dimensional random vector, $i = 1, ..., K$. Let $(X_1, ..., X_K)\tilde{}multinomial(n, \pi_1, , \pi_K)$, where $\pi_i = (\pi_{i1}, ..., \pi_{id})$ is a d-dimensional vector, $i = 1, ..., K$. Show that $(X_{1+}, ..., X_{K+})\tilde{}multinomial(n, \pi_{1+}, ..., \pi_{K+})$, where $X_{j+} = \Sigma_{i=1}^d X_{ji}$ and $\pi_{j+} = \Sigma_{i=1}^d \pi_{ji}$, for $j = 1, ..., K$.

The moment generating function of $(X_{1+}, ..., X_{K+})$ is given by $E[exp(\Sigma_{j=1}^K t_j X_{j+})] = E[exp(\Sigma_{j=1}^K t_j \Sigma_{i=1}^d X_{ji})] = E[exp(\Sigma_{j=1}^K \Sigma_{i=1}^d t_j X_{ji})] = (\Sigma_{j=1}^K \Sigma_{i=1}^d \pi_{ji} e^{t_j})^n = (\Sigma_{j=1}^K \pi_{j+} e^{t_j})^n$ which is of the form of a moment generating function of a multinomial distribution with parameters $(n, \pi_{1+}, ..., \pi_{K+})$. That is, $(X_{1+}, ..., X_{K+})\tilde{}multinomial(n, \pi_{1+}, ..., \pi_{K+})$

2. Let $(X_1, ..., X_6)\tilde{}multinomial(n, \pi_1, ..., \pi_6)$. Show that $(X_1+X_3, X_2, X_4+X_5)\tilde{}multinomial(n, \pi_1 + \pi_3, \pi_2, \pi_4 + \pi_5; \Sigma_{i=1}^5 \pi_i \leq 1)$

The moment generating function of $(X_1 + X_3, X_2, X_4 + X_5)$ is given by $E[exp(t_1(X_1 + X_3) + t_2 X_2 + t_3(X_4 + X_5))]$
$= E[exp(t_1 X_1 + t_1 X_3 + t_2 X_2 + t_3 X_4 + t_3 X_5)]$
$= (\pi_1 e^{t_1} + \pi_3 e^{t_1} + \pi_2 e^{t_2} + \pi_4 e^{t_3} + \pi_5 e^{t_3})^n = ((\pi_1 + \pi_3) e^{t_1} + \pi_2 e^{t_2} + (\pi_4 + \pi_5) e^{t_3})^n$
which is of the form of a moment generating function of a multinomial distribution with parameters $(n, \pi_1 + \pi_3, \pi_2, \pi_4 + \pi_5)$. That is, $(X_1 + X_3, X_2, X_4 + X_5)\tilde{}multinomial(n, \pi_1 + \pi_3, \pi_2, \pi_4 + \pi_5; \Sigma_{i=1}^5 \pi_i \leq 1)$

3. The probability integral transform theorem shows that if X is continuous with cdf F_X, then $Y = F_X(X)$ is uniformly distributed on (0,1). In this problem, we investigate the relationship between discrete random variables and uniform random variables. Let X be a discrete random variable with cdf F_X and define the random variable Y as $Y = F_X(X)$. Let U be a uniform random variable on (0,1). Show that the cdf of Y satisfies $F_Y(y) \leq P(U \leq y) = y$, for all $0 < y < 1$ and $F_Y(y) < P(U \leq y) = y$, for some $0 < y < 1$. Note that in this case, Y is said to be stochastically greater than a Uniform(0,1) random variable.

Consider the set $A_y = \{x : F_X(x) \leq y\}$. Since F_X is non-decreasing,
(1) $A_y = (-\infty, x_y]$ or
(2) $A_y = (-\infty, x_y)$

If (1) is true then $F_Y(y) = P(Y \leq y) = P(F_X(X) \leq y) = P(X \in A_y) = F_X(x_y) \leq y$.

(2) must be true for at least one $y \in (0,1)$ as X is discrete, so there must be at least one point where F_X is not left-continuous (i.e. a "jump point"). When this is the case, $F_Y(y) = lim_{x \uparrow y} F_X(x) < y$.

Therefore, the cdf of Y satisfies $F_Y(y) \leq P(U \leq y) = y$, for all $0 < y < 1$ and $F_Y(y) < P(U \leq y) = y$, for some $0 < y < 1$

4. Problem 1.7

a. Since the new drug is better every time, y = n = 20, so $\hat{\pi} = y/n = 20/20 = 1$

b. Wald statistic $W^2 = \frac{20(1-0.5)^2}{1(0)} = \infty$. A 95% Wald CI is $1 \pm 1.96\sqrt{1(0)/20} = 1 \pm 0$ or $(1,1)$. These are clearly not sensible.

c. $S^2 = \frac{20(1-0.5)^2}{.5(.5)} = 20$, $apval(S^2) = P_{H_0}(S^2 \geq 20) < .0001$. A 95% Score CI is $\{\beta : \frac{n(\hat{\pi}-\beta)^2}{\beta(1-\beta)} \leq \chi^2_{.05}(1)\} = \{\beta : \frac{20(1-\beta)^2}{\beta(1-\beta)} \leq 3.841\} = (0.8389, 1)$. Therefore, we reject the null hypothesis and conclude that there is sufficient evidence that $\pi \neq 0.5$

d. $L^2 = 2[Y log(\frac{Y}{n\pi_0}) + (n-Y) log(\frac{n-Y}{n-n\pi_0})] = 2[20 log(\frac{20}{20(.5)}) + 0] = 27.7$, $apval(L^2) = P_{H_0}(L^2 \geq 27.7) < .0001$. A likelihood-based 95% CI is $(e^{(-1.96^2/40)}, 1) = (0.908, 1)$. Again, we reject the null hypothesis and conclude that there is sufficient evidence that $\pi \neq 0.5$

e. Exact binomial p-value $= 2(.5)^{20} = 0.00000191$. Using SAS, we find that the 95% CI is (0.8316, 1). Again, we reject the null hypothesis and conclude that there is sufficient evidence that $\pi \neq 0.5$

f. Using SAS, we find that the sample size necessary for a one-sided test with $\alpha = 0.05, \pi_0 = 0.5, \pi = 0.9$, and power=0.95 is $n = 11$.

5. Problem 1.8:

$y = (854, 249) \leftarrow Y\tilde{}binomial(n = 1103, \pi)$.

Test $H_0 : \pi_g = \frac{3}{4}$ vs H_1 : not H_0

$dim(H_0) = 0$

$dim(H_0 \cup H_1) = 1$ so $\nu = 1 - 0 = 1$

$S^2(y) = X^2(y) = \frac{n(\hat{\pi}-\pi_0)^2}{\pi_0(1-\pi_0)} = \frac{1103(\frac{854}{1103} - \frac{3}{4})^2}{\frac{3}{4}(\frac{1}{4})} \approx 3.46$

$apval(X^2(y)) = P_{H_0}(X^2(y) \geq 3.46) \approx 0.0629$

$pval(X^2(y)) = P_{H_0}(X^2(Y) \geq 3.46) = P_{H_0}(Y \in \{x : X^2(x) \geq 3.46\})$ where $Y\tilde{}multinomial(1103, \frac{3}{4}, \frac{1}{4})$ under H_0

$pval(X^2(y)) \approx 0.0658$

Since the approximate and exact p-values are both greater than $\alpha = 0.05$, we (narrowly) fail to reject the null hypothesis and conclude that there is insufficient evidence that 3:1 is not the true ratio.