STAT 30100 : Mathematical Statistics-1

Winter 2020

Homework 4

Solutions by

Jinhong Du

12243476

STAT 30100, Homework 4

1. Let $X_n = \frac{1}{n}B_n$, where $B_n \sim \text{Bin}(n, p)$, with $0 < p < 1$. Let $Y_n = \max\{X_n, 1 - X_n\}$. What is the asymptotic distribution of Y_n when

   (a) $p \neq \frac{1}{2}$?

   For any n, suppose $W_1, \ldots, W_n \overset{iid}{\sim} \text{Bernoulli}(p)$, then $X_n = \frac{1}{n}\sum_{i=1}^n W_n \triangleq \overline{W}$. Also, we have $\mathbb{E}W_i = p$ and $\text{Var}(W_i) = p(1-p)$. By central limit theorem, we have $\sqrt{n}(X_n - p) \xrightarrow{D} N(0, p(1-p)) \triangleq Z$.
   Let $h(x) = \max\{x, 1-x\}\mathbb{1}_{x\in[0,1]}$, then $h'(x) = \begin{cases} -1 & , \ 0 < x < \frac{1}{2} \\ 1 & , \ \frac{1}{2} < x < 1 \\ 0 & , \ x \leq 0 \text{ or } x \geq 1 \end{cases}$, which is continous near $p \neq \frac{1}{2}$. By Cramer's Theorem, we have $\sqrt{n}[h(X_n) - h(p)] \xrightarrow{D} h'(p)Z$. So, when $p > \frac{1}{2}$, $\sqrt{n}(Y_n - p) \xrightarrow{D} N(0, p(1-p))$; when $p < \frac{1}{2}$, $\sqrt{n}(Y_n - 1 + p) \xrightarrow{D} N(0, p(1-p))$.

   (b) $p = \frac{1}{2}$?

   From (a), we have $\sqrt{n}(X_n - \frac{1}{2}) \xrightarrow{D} N(0, \frac{1}{4}) \triangleq Z$. Notice that $Y_n - \frac{1}{2} = \max\{X_n - \frac{1}{2}, \frac{1}{2} - X_n\}$. Let $g(x) = \max\{x, -x\}$. since g is continuous on $\mathbb{R}$, by Slutsky Theorem, we have $g\left(X_n - \frac{1}{2}\right) \xrightarrow{D} g(Z)$, i.e., $\sqrt{n}\left(Y_n - \frac{1}{2}\right) \xrightarrow{D} \max\{Z, -Z\} = |Z|$. Since $T = |Z| = \begin{cases} Z & , \text{ if } Z \geq 0 \\ -Z & , \text{ if } Z < 0 \end{cases}$, the density of T is given by $f_T(x) = 2\frac{2}{\sqrt{2\pi}}e^{-2x^2}\mathbb{1}_{x\geq 0}$, which is the half normal distribution with scale parameter $\frac{1}{2}$.

2. Suppose that $X_1, \ldots, X_n$ are i.i.d. random vectors in $\mathbb{R}^d$, having mean vector μ and $d \times d$ nonsingular covariance matrix Σ. Lemma 2 of Ferguson chapter 9 (pp. 56-57) says that the Hotelling T^2 statistic converges in distribution to a central χ_d^2 random variable, where

$$T^2 = (n-1)(\overline{X}_n - \mu)^\top S_n^{-1}(\overline{X}_n - \mu), \quad \overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \quad \text{and} \quad S_n = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)(X_i - \overline{X}_n)^\top.$$

   (a) On p. 57, Ferguson gives a 2-sentence proof outline for Lemma 2. Give a clearer and more detailed version of this proof.

   Proof. Since $X_1, \ldots, X_n$ are i.i.d. random variables with mean μ and covariance matrix Σ, we have $\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{D} N(0, \Sigma) \triangleq Y$. Notice that $S_n = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)(X_i - \overline{X}_n)^\top = \frac{1}{n}\sum_{i=1}^n [X_i X_i^\top - \overline{X}_n X_i^\top - X_i \overline{X}_n^\top + \overline{X}_n \overline{X}_n^\top] = \frac{1}{n}\sum_{i=1}^n X_i X_i^\top - \overline{X}_n \overline{X}_n^\top$.
   For the first term, since $\mathbb{E}(X_i X_i^\top) = \text{Var}(X_i) + \mathbb{E}(X_i)\mathbb{E}(X_i)^\top = \Sigma + \mu\mu^\top$, by Weak Law of Large Number, we have $\frac{1}{n}\sum_{i=1}^n X_i X_i^\top \xrightarrow{\mathbb{P}} \Sigma + \mu\mu^\top$. For the second term, since $\overline{X}_n \xrightarrow{\mathbb{P}} \mu$, we have $\overline{X}_n \overline{X}_n^\top \xrightarrow{\mathbb{P}} \mu\mu^\top$. So $S_n \xrightarrow{\mathbb{P}} \Sigma$, $S_n^{-\frac{1}{2}} \xrightarrow{\mathbb{P}} \Sigma^{-\frac{1}{2}}$ and therefore $S_n^{-\frac{1}{2}} \xrightarrow{D} \Sigma^{-\frac{1}{2}}$.
   By Slutsky Theorem, we have $\sqrt{n}S_n^{-\frac{1}{2}}(\overline{X}_n - \mu) \xrightarrow{D} \Sigma^{-\frac{1}{2}}Y$ and $S_n^{-\frac{1}{2}}(\overline{X}_n - \mu) \xrightarrow{D} \mathbf{0}$. Let $g : \mathbb{R}^d \mapsto \mathbb{R}$ such that $g(x) = x^\top x$ is continous on $\mathbb{R}^d$, then by Slutsky Theorem, we have $n(\overline{X}_n - \mu)^\top S_n^{-1}(\overline{X}_n - \mu) \xrightarrow{D} Y\Sigma^{-1}Y \sim \chi_d^2$ and $(\overline{X}_n - \mu)^\top S_n^{-1}(\overline{X}_n - \mu) \xrightarrow{D} 0$. Then again by Slutsky Theorem, we have $T^2 \xrightarrow{D} \chi_d^2$. $\square$

(b) Suppose that μ_n is a sequence of vectors (possibly dependent on $X_1, \ldots, X_n$) such that $\sqrt{n}(\mu_n - \mu) \to k$, where k is a fixed vector in $\mathbb{R}^d$. Formulate and prove a theorem on the limiting non-central χ_d^2 distribution for T_n^2, where

$$T_n^2 = n\left(\overline{X}_n - \mu_n\right)^T S_n^{-1}\left(\overline{X}_n - \mu_n\right).$$

Theorem. Suppose that $X_1, \ldots, X_n$ are i.i.d. random vectors in $\mathbb{R}^d$, having mean vector μ and $d \times d$ nonsingular covariance matrix Σ. Suppose that μ_n is a sequence of vectors (possibly dependent on $X_1, \ldots, X_n$) such that $\sqrt{n}(\mu_n - \mu) \to k$, where k is a fixed vector in $\mathbb{R}^d$. Then $T_n^2 \xrightarrow{D} \chi_d^2(k^2)$ where

$$T_n^2 = n\left(\overline{X}_n - \mu_n\right)^T S_n^{-1}\left(\overline{X}_n - \mu_n\right), \quad \overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \quad \text{and} \quad S_n = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)(X_i - \overline{X}_n)^\top.$$

Proof. By Slutsky Theorem, we have that $\sqrt{n}(\overline{X}_n - \mu_n) = \sqrt{n}(\overline{X}_n - \mu) - \sqrt{n}(\mu_n - \mu) \xrightarrow{D} N(-k, \Sigma) \triangleq Z$. Similar to the proof in (a), we have $\sqrt{n}S_n^{-\frac{1}{2}}(\overline{X}_n - \mu_n) \xrightarrow{D} \Sigma^{-\frac{1}{2}}Z \sim N(-k\Sigma^{-\frac{1}{2}}, I_d)$. Let $g : \mathbb{R}^d \mapsto \mathbb{R}$ such that $g(x) = x^\top x$ is continous on $\mathbb{R}^d$, then by Slutsky Theorem, we have $n\left(\overline{X}_n - \mu_n\right)^T S_n^{-1}\left(\overline{X}_n - \mu_n\right) \xrightarrow{D} \chi_d^2(k^\top \Sigma^{-1} k)$. $\square$

3. Suppose we have a bivariate Bernoulli random variable $X = (Y, Z)$, where Y and Z take value 0 or 1. Suppose we have n i.i.d. realizations of X. Let $\hat{p}_{ij}$ be the proportion of the observations in the sample such that $Y = i$ and $Z = j$, and let $\hat{p}_{i+} = \hat{p}_{i0} + \hat{p}_{i1}$ and $\hat{p}_{+j} = \hat{p}_{0j} + \hat{p}_{1j}$, where $i = 0, 1$ and $j = 0, 1$. Consider the statistic

$$S = \sum_{i=0}^1 \sum_{j=0}^1 \frac{n\left(\hat{p}_{ij} - \hat{p}_{i+}\hat{p}_{+j}\right)^2}{\hat{p}_{i+}\hat{p}_{+j}}.$$

Find the asymptotic distribution of S as $n \to \infty$ for the case when Y and Z are independent, and prove your finding. Comment on the possible uses of this asymptotic result.

Proof. Let $X_1, \ldots, X_n$ be the n realizations with $Y_1, \ldots, Y_n \overset{iid}{\sim}$ Bernoulli(p_Y) and $Z_1, \ldots, Z_n \overset{iid}{\sim}$ Bernoulli(p_Z). Notice that

$$\begin{aligned}
\hat{p}_{11} - \hat{p}_{1+}\hat{p}_{+1} &= \frac{1}{n}\sum_{k=1}^n Y_k Z_k - \left(\frac{1}{n}\sum_{k=1}^n Y_k\right)\left(\frac{1}{n}\sum_{k=1}^n Z_k\right) = \frac{1}{n}\sum_{k=1}^n (Y_k - \overline{Y}_n)(Z_k - \overline{Z}_n) \\
\hat{p}_{10} - \hat{p}_{1+}\hat{p}_{+1} &= \frac{1}{n}\sum_{k=1}^n Y_k(1 - Z_k) - \left(\frac{1}{n}\sum_{k=1}^n Y_k\right)\left(\frac{1}{n}\sum_{k=1}^n (1 - Z_k)\right) \\
&= \frac{1}{n}\sum_{k=1}^n (Y_k - \overline{Y}_n)[(1 - Z_k) - (1 - \overline{Z}_n)] = -\frac{1}{n}\sum_{k=1}^n (Y_k - \overline{Y}_n)(Z_k - \overline{Z}_n) \\
\hat{p}_{01} - \hat{p}_{0+}\hat{p}_{+1} &= \frac{1}{n}\sum_{k=1}^n (1 - Y_k)Z_k - \left(\frac{1}{n}\sum_{k=1}^n (1 - Y_k)\right)\left(\frac{1}{n}\sum_{k=1}^n Z_k\right) \\
&= \frac{1}{n}\sum_{k=1}^n [(1 - Y_k) - (1 - \overline{Y}_n)](Z_k - \overline{Z}_n) = -\frac{1}{n}\sum_{k=1}^n (Y_k - \overline{Y}_n)(Z_k - \overline{Z}_n) \\
\hat{p}_{00} - \hat{p}_{0+}\hat{p}_{+0} &= \frac{1}{n}\sum_{k=1}^n (1 - Y_k)(1 - Z_k) - \left(\frac{1}{n}\sum_{k=1}^n (1 - Y_k)\right)\left(\frac{1}{n}\sum_{k=1}^n (1 - Z_k)\right) \\
&= \frac{1}{n}\sum_{k=1}^n [(1 - Y_k) - (1 - \overline{Y}_n)][(1 - Z_k) - (1 - \overline{Z}_n)] = \frac{1}{n}\sum_{k=1}^n (Y_k - \overline{Y}_n)(Z_k - \overline{Z}_n),
\end{aligned}$$

Solution (cont.)

$$\sum_{i=0}^{1}\sum_{j=0}^{1}\frac{1}{\hat{p}_{i+}\hat{p}_{+j}}=\left(\frac{1}{\overline{Y}_n\overline{Z}_n}+\frac{1}{(1-\overline{Y}_n)\overline{Z}_n}+\frac{1}{\overline{Y}_n(1-\overline{Z}_n)}+\frac{1}{(1-\overline{Y}_n)(1-\overline{Z}_n)}\right)$$
$$=\frac{1}{\overline{Y}_n\overline{Z}_n(1-\overline{Y}_n)(1-\overline{Z}_n)},$$

let $T_n=\frac{1}{n}\sum_{k=1}^{n}(Y_k-\overline{Y}_n)(Z_k-\overline{Z}_n)$, we have $S=\frac{nT_n^2}{\overline{Y}_n\overline{Z}_n(1-\overline{Y}_n)(1-\overline{Z}_n)}$. Since Y and Z are independent, we have

$$\mathbb{E}(T_n)=\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}(Y_k-\overline{Y}_n)\mathbb{E}(Z_k-\overline{Z}_n)=0$$
$$\begin{aligned}\mathrm{Var}(T_n)=\mathbb{E}(T_n^2)&=\frac{1}{n^2}\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}(Y_i-\overline{Y}_n)(Z_i-\overline{Z}_n)(Y_j-\overline{Y}_n)(Z_j-\overline{Z}_n)\right]\\&=\frac{1}{n^2}\sum_{i\neq j}\mathbb{E}[(Y_i-\overline{Y}_n)(Y_j-\overline{Y}_n)]\mathbb{E}[(Z_i-\overline{Z}_n)(Z_j-\overline{Z}_n)]\\&\quad+\frac{1}{n^2}\sum_{k=1}^{n}\mathbb{E}[(Y_k-\overline{Y}_n)^2]\mathbb{E}[(Z_k-\overline{Z}_n)^2].\end{aligned}$$

While

$$\begin{aligned}\mathbb{E}[(Y_i-\overline{Y}_n)(Y_j-\overline{Y}_n)]&=\mathbb{E}(Y_i)\mathbb{E}(Y_j)-\mathbb{E}(Y_i\overline{Y}_n)-\mathbb{E}(Y_j\overline{Y}_n)+\mathbb{E}(\overline{Y}_n^2)\\&=p_Y^2-\frac{2(n-1)}{n}p_Y^2-\frac{2}{n}p_Y+\frac{1}{n}p_Y(1-p_Y)+p_Y^2\\&=-\frac{1}{n}p_Y(1-p_Y)\\\mathbb{E}[(Y_k-\overline{Y}_n)^2]&=\mathbb{E}[(Y_k-p_Y)^2]-2\mathbb{E}[(Y_k-p_Y)(\overline{Y}_n-p_Y)]+\mathbb{E}[(p_Y-\overline{Y}_n)^2]\\&=p_Y(1-p_Y)-\frac{2}{n}\sum_{j=1}^{n}\mathbb{E}[(Y_k-p_Y)(Y_j-p_Y)]+\frac{1}{n}p_Y(1-p_Y)\\&=\frac{n+1}{n}p_Y(1-p_Y)-\frac{2}{n}p_Y(1-p_Y)-\frac{2}{n}\sum_{j\neq k}\mathbb{E}(Y_k-p_Y)\mathbb{E}(Y_j-p_Y)]\\&=\frac{n-1}{n}p_Y(1-p_Y)\end{aligned}$$

and analogously $\mathbb{E}[(Z_i-\overline{Z}_n)(Z_j-\overline{Z}_n)]=-\frac{1}{n}p_Z(1-p_Z)$, $\mathbb{E}[(Z_k-\overline{Z}_n)^2]=\frac{n-1}{n}p_Z(1-p_Z)$, so $\mathrm{Var}(T_n)=\frac{n-1}{n^2}p_Y(1-p_Y)p_Z(1-p_Z)$. By central limit theorem, we have $\sqrt{n}\sqrt{\frac{n}{n-1}}T_n\xrightarrow{D}N\left(0,p_Y(1-p_Y)p_Z(1-p_Z)\right)\triangleq W$. Since $\lim_{n\to\infty}\sqrt{\frac{n}{n-1}}=1$, by Slutsky Theorem, we have

$$\sqrt{n}T_n\xrightarrow{D}W.$$

Since by law of large numbers, we have $\overline{Y}_n\xrightarrow{\mathbb{P}}p_Y$ and $\overline{Z}_n\xrightarrow{\mathbb{P}}p_Z$. By Slutsky Theorem, we have $\left(\sqrt{n}T_n,\overline{Y}_n,\overline{Z}_n\right)^\top\xrightarrow{D}(W,p_Y,p_Z)^\top$.

Consider the function $g:\mathbb{R}^3\mapsto\mathbb{R}$ such that $g((a,b,c)^\top)=a^2\left(\frac{1}{bc}+\frac{1}{(1-b)c}+\frac{1}{b(1-c)}+\frac{1}{(1-b)(1-c)}\right)$, which is continous except when $b\in\{0,1\}$ or $c\in\{0,1\}$, i.e., $\mathbb{P}(\{x\in\mathbb{R}^3: g \text{ is continous at } x\})=1$. So by Slutsky Theorem, we have $g\left(\left(\sqrt{n}T_n,\overline{Y}_n,\overline{Z}_n\right)^\top\right)\xrightarrow{D}g\left((W,p_Y,p_Z)^\top\right)=\left(\frac{W}{p_Yp_Z(1-p_Y)(1-p_Z)}\right)^2\overset{D}{=\!=}\chi_1^2$, i.e.,

$$S\xrightarrow{D}\chi_1^2.$$

The result is useful for testing whetehr Y and Z are independent or not.

□

4. (a) One measure of the homogeneity of a multinomial population with k cells and probabilities $\boldsymbol{p} = (p_1, \ldots, p_k)^\top$, is the sum of the squares of the probabilities, $S(\boldsymbol{p}) = \sum_{i=1}^k p_i^2$. Note that $\frac{1}{k} \le S(\boldsymbol{p}) \le 1$, with higher values indicating greater heterogeneity. Given a sample of size n from this population (with replacement), we can estimate $S(\boldsymbol{p})$ by $S(\hat{\boldsymbol{p}})$, where $\hat{\boldsymbol{p}} = (\hat{p}_1, \ldots, \hat{p}_k)^\top$ and $\hat{p}_i$ is the proportion of the observations that fall in cell i. Find the asymptotic distribution of $S(\hat{\boldsymbol{p}})$. Remember to consider separately the case when $\boldsymbol{p}$ is uniform, i.e. $(\frac{1}{k}, \ldots, \frac{1}{k})^\top$.

Let $X_i \in \mathbb{R}^k$ to be $\mathbf{e}_j$ if the ith trial resulted in outcome j for $i = 1, \ldots, n$. Then $\hat{\boldsymbol{p}} = \frac{1}{n}\sum_{i=1}^n X_i$. We have $\sqrt{n}(\hat{\boldsymbol{p}} - \boldsymbol{p}) \xrightarrow{D} N(0, \Sigma) \triangleq Y$ where $\Sigma = P - \boldsymbol{p}\boldsymbol{p}^\top$ and $P = \mathrm{diag}(\boldsymbol{p})$. Also, $S'(\boldsymbol{p}) = 2\boldsymbol{p}$.
(1) If $\boldsymbol{p} \ne (\frac{1}{k}, \ldots, \frac{1}{k})^\top$, by Cramer's Theorem, we have

$$\sqrt{n}[S(\hat{\boldsymbol{p}}) - S(\boldsymbol{p})] \xrightarrow{D} N(0, 4\boldsymbol{p}^\top P\boldsymbol{p} - 4\boldsymbol{p}^\top\boldsymbol{p}\boldsymbol{p}^\top\boldsymbol{p})$$

(2) If $\boldsymbol{p} = (\frac{1}{k}, \ldots, \frac{1}{k})^\top = \frac{1}{k}\mathbf{1}$, then $4\boldsymbol{p}^\top P\boldsymbol{p} - 4\boldsymbol{p}^\top\boldsymbol{p}\boldsymbol{p}^\top\boldsymbol{p} = 4\frac{1}{k^2}(\boldsymbol{p}^\top\mathbf{1} - \boldsymbol{p}^\top\mathbf{1}\boldsymbol{p}^\top\mathbf{1}) = 0$ and thus the above method cannot be applied directly to obtain the asymptotic distribution. Since

$$\sqrt{n}(\hat{\boldsymbol{p}} - \boldsymbol{p})^\top P^{-1}\sqrt{n}(\hat{\boldsymbol{p}} - \boldsymbol{p}) = nk(\hat{\boldsymbol{p}}^\top\hat{\boldsymbol{p}} - 2\frac{1}{k}\hat{\boldsymbol{p}}\mathbf{1} + \boldsymbol{p}^\top\boldsymbol{p}) = nk(\hat{\boldsymbol{p}}^\top\hat{\boldsymbol{p}} - \frac{1}{k}),$$

by Slutsky Theorem, we have $nk(\hat{\boldsymbol{p}}^\top\hat{\boldsymbol{p}} - \frac{1}{k}) \xrightarrow{D} Y^\top P^{-1}Y$. Since $(P^{-1}\Sigma)^2 = k^2(P^2 - \boldsymbol{p}\boldsymbol{p}^\top P - P\boldsymbol{p}\boldsymbol{p}^\top + \boldsymbol{p}\boldsymbol{p}^\top\boldsymbol{p}\boldsymbol{p}^\top) = I - 2k\boldsymbol{p}\boldsymbol{p}^\top + k\boldsymbol{p}\boldsymbol{p}^\top = k(P - \boldsymbol{p}\boldsymbol{p}^\top) = P^{-1}\Sigma$, we have $P^{-1}\Sigma$ is idempotent. Also, $\mathrm{rank}(P^{-1}\Sigma) = \mathrm{tr}(P^{-1}\Sigma) = \mathrm{tr}(I_k - \mathbf{1}\boldsymbol{p}^\top) = k - 1$. So by Theorem 2 in the handout of quadratic forms, $Y^\top P^{-1}Y \sim \chi^2_{k-1}$. Therefore, $nk(\hat{\boldsymbol{p}}^\top\hat{\boldsymbol{p}} - \frac{1}{k}) \xrightarrow{D} \chi^2_{k-1}$

(b) Another measure of homogeneity often used is Shannon entropy, defined as $H(\boldsymbol{p}) = -\sum_{i=1}^k p_i \log(p_i)$, with $0 \le H(\boldsymbol{p}) \le \log k$, and with higher values indicating greater homogeneity. What is the asymptotic distribution of $H(\hat{\boldsymbol{p}})$? Remember to consider separately the case when p is uniform.

(1) If $\boldsymbol{p} \ne (\frac{1}{k}, \ldots, \frac{1}{k})^\top$, since $H'(\boldsymbol{p}) = (-\log(p_1) - 1, \ldots, -\log(p_k) - 1)^\top$, by Cramer's Theorem, we have

$$\sqrt{n}[H(\hat{\boldsymbol{p}}) - H(\boldsymbol{p})] \xrightarrow{D} N(0, H'(\boldsymbol{p})^\top(P - \boldsymbol{p}\boldsymbol{p}^\top)H'(\boldsymbol{p}))$$

(2) If $\boldsymbol{p} = (\frac{1}{k}, \ldots, \frac{1}{k})^\top = \frac{1}{k}\mathbf{1}$, then $H'(\boldsymbol{p})^\top(\hat{\boldsymbol{p}} - \boldsymbol{p}) = [\log(k) - 1]\mathbf{1}^\top(\hat{\boldsymbol{p}} - \boldsymbol{p}) = 0$, $H'(\boldsymbol{p})^\top(P - \boldsymbol{p}\boldsymbol{p}^\top)H'(\boldsymbol{p}) = [\log(k) - 1]^2 - [\log(k) - 1]^2\mathbf{1}^\top\boldsymbol{p}\boldsymbol{p}^\top\mathbf{1} = 0$ and thus the above method cannot be applied directly to obtain the asymptotic distribution. Consider the Taylor expansion

$$\begin{aligned}
&H(\hat{\boldsymbol{p}}) - H(\boldsymbol{p}) \\
=&H'(\boldsymbol{p})^\top(\hat{\boldsymbol{p}} - \boldsymbol{p}) + \frac{1}{2}(\hat{\boldsymbol{p}} - \boldsymbol{p})^\top H''(\boldsymbol{p})(\hat{\boldsymbol{p}} - \boldsymbol{p}) + \frac{1}{6}\sum_{i_1=1}^k\sum_{i_2=1}^k\sum_{i_3=1}^k \left.\frac{\partial^3 H(\boldsymbol{t})}{\partial t_{i_1}\partial t_{i_2}\partial t_{i_3}}\right|_{\boldsymbol{t}=\boldsymbol{w}} \prod_{l=1}^3(\hat{p}_{i_l} - p_{i_l}) \\
=&-\frac{1}{2k}(\hat{\boldsymbol{p}} - \boldsymbol{p})^\top(\hat{\boldsymbol{p}} - \boldsymbol{p}) + R_3(\hat{\boldsymbol{p}}, \boldsymbol{p})
\end{aligned}$$

where $\boldsymbol{w}$ is in the line segment between $\hat{\boldsymbol{p}}$ and $\boldsymbol{p}$. So

$$n[H(\hat{\boldsymbol{p}}) - H(\boldsymbol{p})] = -n\frac{1}{2k}(\hat{\boldsymbol{p}} - \boldsymbol{p})^\top(\hat{\boldsymbol{p}} - \boldsymbol{p}) + nR_3(\hat{\boldsymbol{p}}, \boldsymbol{p}).$$

Since by law of large number, $\hat{\boldsymbol{p}} - \boldsymbol{p} \xrightarrow{\mathbb{P}} 0$ and by central limit theorem $\sqrt{n}(\hat{\boldsymbol{p}} - \boldsymbol{p}) \xrightarrow{D} N(0, \Sigma)$ while $\hat{p}_i - p_i$ has order 3 in $R_3(\hat{\boldsymbol{p}}, \boldsymbol{p})$, we have $nR_3(\hat{\boldsymbol{p}}, \boldsymbol{p}) \xrightarrow{\mathbb{P}} 0$. Form (a) we also have $-n\frac{1}{2k}(\hat{\boldsymbol{p}} - \boldsymbol{p})^\top(\hat{\boldsymbol{p}} - \boldsymbol{p}) \xrightarrow{D} -\frac{1}{2}\chi^2_{k-1}$. Therefore, $2n[H(\hat{\boldsymbol{p}}) - \log(k)] \xrightarrow{D} -\chi^2_{k-1}$.

5. Suppose a sample of size 60 is taken from the hatchlings of a litter of lady bird beetles, and the offspring are divided into the two-by-two contingency table using the dichotomies: male/female and spotted/plain. The data are: 15 spotted-male, 21 spotted-female, 17 plain-male, and 7 plain-female.

   (a) Write the formula for Pearson's χ^2 statistic for testing the hypothesis that all 4 cells have equal probability, $\frac{1}{4}$. Calculate the value of the statistic for this specific data set. How many degrees of freedom does the χ^2 have?

   Chi-squared test for given probabilities,

   $$X = n\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(\hat{p}_{ij}-p_{ij})^2}{p_{ij}} = 60\cdot\frac{\left(\frac{15}{60}-\frac{1}{4}\right)^2+\left(\frac{21}{60}-\frac{1}{4}\right)^2+\left(\frac{17}{60}-\frac{1}{4}\right)^2+\left(\frac{7}{60}-\frac{1}{4}\right)^2}{\frac{1}{4}} \approx 6.933333$$

   Since we are testing for 4 categories, the degree of freedom is $4-1=3$. Under H_0, $X\sim\chi^2_3$.

   (b) Find the noncentrality parameter for the alternative that specifies P(spottedmale)=.20, P(spotted-female)=.35, P(plain-male)=.30, and P(plain-female)=.15.

   For the null hypothesis, $p = p_0 = (\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4})^\top$. For the alternative hypothesis, $p = p_1 = (0.2, 0.35, 0.30, 0.15)^\top$. Set $\delta = \sqrt{n}(p_1 - p_0) = 2\sqrt{15}(-0.05, 0.1, 0.05, -0.1)^\top$. Under the alternative p_1, the chi square statistic is approximately noncentral χ^2_3 with noncentrality parameter $\lambda = \delta^\top \operatorname{diag}(p_0)^{-1}\delta = 60\times 4\times(0.05^2+0.1^2+0.05^2+0.1^2) = 6$.

   (c) Find the sample size needed to get power .9 at this alternative when testing at the 5% level of significance.

   For $\alpha = 0.05$, let $x_{1-\alpha}$ be the $(1-\alpha)$-quantile of χ^2_3. For the alternative, the noncentrality parameter is given by $\lambda = n\times 0.1$ for given sample size n. The power is approximated by $\mathbb{P}(\chi^2_3(0.1n) > x_{1-\alpha})$. Let $\mathbb{P}(\chi^2_3(0.1n) > x_{1-\alpha}) = 0.9$, we have $n\geq 142$. So the sample size should be at least 142.

```
library(pwr)
pwr.chisq.test(w=sqrt(6/60),df=3,sig.level=0.05,power=0.9)
```

```
##
##      Chi squared power calculation
##
##               w = 0.3162278
##               N = 141.7149
##              df = 3
##       sig.level = 0.05
##           power = 0.9
##
## NOTE: N is the number of observations
```

```
1-pchisq(qchisq(.95,3), 3, ncp=t(delta) %*% solve(diag(p0)) %*% delta/60*142)
```

```
## [1] 0.9006331
```

6. (Casella and Berger Problem 5.19)

(a) Prove that the χ^2 distribution is *stochastically increasing* in its degrees of freedom; that is, if $p > q$, then for any a, $\mathbb{P}(\chi_p^2 > a) \geq \mathbb{P}(\chi_q^2 > a)$, with strict inequality for some a.

Proof. For any $k \in \mathbb{Z}^+$, $\chi_k^2 = \sum_{i=1}^k Z_i^2$ for some $Z_1, \dots, Z_k \overset{iid}{\sim} N(0,1)$. For $p > q$, we can construct a sample space such that $\chi_p^2 = \sum_{i=1}^p Z_i^2$ and $\chi_q^2 = \sum_{i=1}^q Z_i^2$ for $Z_1, \dots, Z_p \overset{iid}{\sim} N(0,1)$. Then $X = \sum_{i=q+1}^p Z_i^2 \sim \chi_{p-q}^2$. For $a > 0$,

$$\begin{aligned}
\mathbb{P}(\chi_p^2 \leq a) &= \mathbb{P}(\chi_q^2 + X \leq a) \\
&= \mathbb{P}(\chi_q^2 + X \leq a | X > 0)\mathbb{P}(X > 0) + \mathbb{P}(\chi_q^2 + X \leq a | X \leq 0)\mathbb{P}(X \leq 0) \\
&= \mathbb{P}(\chi_q^2 + X \leq a | X > 0) \\
&< \mathbb{P}(\chi_q^2 \leq a)
\end{aligned}$$

since $\mathbb{P}(X > 0) = 1$. For $a \leq 0$,

$$\mathbb{P}(\chi_p^2 \leq a) = \mathbb{P}(\chi_q^2 \leq a) = 0.$$

Therefore, $\mathbb{P}(\chi_p^2 > a) \geq \mathbb{P}(\chi_q^2 > a)$, with strict inequality for $a > 0$. □

(b) Use the results of part (a) to prove that for any ν, $kF_{k,\nu}$ is stochastically increasing in k.

Proof. Notice that $F_{k,\nu} = \frac{\chi_k^2/k}{\chi_\nu^2/\nu}$ where χ_k^2 and χ_ν^2 are independent. From (a) χ_k^2 is stochastically increasing in k, so $kF_{k,\nu} = \frac{\chi_k^2}{\chi_\nu^2/\nu}$ is stochastically increasing in k. □

(c) Show that for any k, ν, and α, $kF_{\alpha,k,\nu} > (k-1)F_{a,k-1,\nu}$ (The notation $F_{\alpha,k-1,\nu}$ denotes a level-α *cutoff point*; see Section 8.3.1. Also see Miscellanea 8.5.1 and Exercise 11.15.)

Proof. For any k, ν, and $\alpha \in (0,1)$, since $F_{\alpha,k,\nu} > 0, F_{\alpha,k-1,\nu} > 0$, from (a) we have

$$\begin{aligned}
\mathbb{P}(kF_{k,\nu} > (k-1)F_{\alpha,k-1,\nu}) &< \mathbb{P}((k-1)F_{k-1,\nu} > (k-1)F_{\alpha,k-1,\nu}) \\
&= \alpha \\
&= \mathbb{P}(kF_{k,\nu} > kF_{\alpha,k,\nu})
\end{aligned}$$

Since the cumulative distribution function is nondecreasing, we have $kF_{\alpha,k,\nu} > (k-1)F_{a,k-1,\nu}$. □

7. (Casella and Berger Problem 5.20)

(a) We can see that the t distribution is a mixture of normals using the following argument:

$$\mathbb{P}\left(T_\nu \leq t\right) = \mathbb{P}\left(\frac{Z}{\sqrt{\chi_\nu^2/\nu}} \leq t\right) = \int_0^\infty \mathbb{P}(Z \leq t\sqrt{x}/\sqrt{\nu})\mathbb{P}\left(\chi_\nu^2 = x\right) \mathrm{d}x,$$

where T_ν is a t random variable with ν degrees of freedom. Using the Fundamental Theorem of Calculus and interpreting $\mathbb{P}(\chi_\nu^2 = \nu x)$ as a pdf, we obtain

$$f_{T_\nu}(t) = \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2 x}{2\nu}} \frac{\sqrt{x}}{\sqrt{\nu}} \frac{1}{\Gamma(\frac{\nu}{2})2^{\nu/2}} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}} \mathrm{d}x,$$

a scale mixture of normals. Verify this formula by direct integration.

Proof. For $t > 0$,

$$\begin{aligned}
f_{T_\nu}(t) &= \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2 x}{2\nu}} \frac{\sqrt{x}}{\sqrt{\nu}} \frac{1}{\Gamma(\frac{\nu}{2})2^{\nu/2}} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}} \mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}\Gamma(\frac{\nu}{2})2^{\frac{\nu}{2}}\sqrt{\nu}} \int_0^\infty x^{\frac{\nu-1}{2}} e^{-\frac{1}{2}\left(\frac{t^2}{\nu}+1\right)x} \mathrm{d}x \\
&\overset{y=\frac{1}{2}\left(\frac{t^2}{\nu}+1\right)x}{=\!=\!=\!=} \frac{1}{\sqrt{2\pi}\Gamma(\frac{\nu}{2})2^{\frac{\nu}{2}}\sqrt{\nu}} 2^{\frac{\nu+1}{2}} \left(\frac{t^2}{\nu}+1\right)^{-\frac{\nu+1}{2}} \int_0^\infty x^{\frac{\nu-1}{2}} e^{-y} \mathrm{d}y \\
&= \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})} \left(\frac{t^2}{\nu}+1\right)^{-\frac{\nu+1}{2}},
\end{aligned}$$

which is the density function for t random variable with ν degree of freedom. □

(b) A similar formula holds for the F distribution; that is, it can be written as a mixture of chi squareds. If $F_{1,\nu}$ is an F random variable with 1 and ν degrees of freedom, then we can write

$$\mathbb{P}\left(F_{1,\nu} \le \nu t\right) = \int_0^\infty \mathbb{P}\left(\chi_1^2 \le ty\right) f_\nu(y) \mathrm{d}y,$$

where $f_\nu(y)$ is a χ_ν^2 pdf. Use the Fundamental Theorem of Calculus to obtain an integral expression for the pdf of $F_{1,\nu}$ and show that the integral equals the pdf.

Proof. Since for $t > 0$,

$$\begin{aligned}
\mathbb{P}\left(F_{1,\nu} \le \nu t\right) &= \int_0^\infty P\left(Z^2 \le ty\right) f_\nu(y) \mathrm{d}y \\
&= \int_0^\infty \int_0^{ty} \frac{1}{\sqrt{2\pi}} z^{-\frac{1}{2}} e^{-\frac{z}{2}} \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} y^{\frac{\nu}{2}-1} e^{-\frac{y}{2}} \mathrm{d}z\mathrm{d}y \\
&\overset{\nu z = yx}{=\!=\!=} \int_0^\infty \int_0^{\nu t} \frac{1}{2^{\frac{\nu+1}{2}}\Gamma(\frac{1}{2})\Gamma(\frac{\nu}{2})} \nu^{-\frac{1}{2}} x^{-\frac{1}{2}} y^{\frac{\nu-1}{2}} e^{-\frac{(\frac{x}{\nu}+1)y}{2}} \mathrm{d}x\mathrm{d}y \\
&= \int_0^{\nu t} \int_0^\infty \frac{1}{2^{\frac{\nu+1}{2}}\Gamma(\frac{1}{2})\Gamma(\frac{\nu}{2})} \nu^{-\frac{1}{2}} x^{-\frac{1}{2}} y^{\frac{\nu-1}{2}} e^{-\frac{(\frac{x}{\nu}+1)y}{2}} \mathrm{d}y\mathrm{d}x \qquad \text{(Fubini's Theorem)}
\end{aligned}$$

the integral expression for the density of $F_{1,\nu}$ is

$$f_{F_{1,\nu}}(x) = \int_0^\infty \frac{1}{2^{\frac{\nu+1}{2}}\Gamma(\frac{1}{2})\Gamma(\frac{\nu}{2})} \nu^{-\frac{1}{2}} x^{-\frac{1}{2}} y^{\frac{\nu-1}{2}} e^{-\frac{(\frac{x}{\nu}+1)y}{2}} \mathrm{d}y.$$

Next we show this is euqal to the pdf,

$$\begin{aligned}
&\int_0^\infty \frac{1}{2^{\frac{\nu+1}{2}}\Gamma(\frac{1}{2})\Gamma(\frac{\nu}{2})} \nu^{-\frac{1}{2}} x^{-\frac{1}{2}} y^{\frac{\nu-1}{2}} e^{-\frac{(\frac{x}{\nu}+1)y}{2}} \mathrm{d}y \\
&\overset{w=\frac{(\frac{x}{\nu}+1)}{2}y}{=\!=\!=\!=} \int_0^\infty \frac{1}{\Gamma(\frac{1}{2})\Gamma(\frac{\nu}{2})} \nu^{-\frac{1}{2}} x^{-\frac{1}{2}} \left(\frac{x}{\nu}+1\right)^{-\frac{\nu+1}{2}} w^{\frac{\nu-1}{2}} e^{-w} \mathrm{d}w \\
&= \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{\nu}{2})} \nu^{-\frac{1}{2}} x^{-\frac{1}{2}} \left(\frac{x}{\nu}+1\right)^{-\frac{\nu+1}{2}}.
\end{aligned}$$

□

(c) Verify that the generalization of part (b),

$$\mathbb{P}\left(F_{m,\nu} \leq \frac{\nu}{m}t\right) = \int_0^\infty \mathbb{P}\left(\chi_m^2 \leq ty\right) f_\nu(y)\mathrm{d}y,$$

is valid for all integers $m > 1$.

Proof. Since for $t > 0$,

$$\begin{aligned}
\mathbb{P}\left(F_{m,\nu} \leq \frac{\nu}{m}t\right) &= \int_0^\infty \mathbb{P}\left(\chi_m^2 \leq ty\right) f_\nu(y)\mathrm{d}y \\
&= \int_0^\infty \int_0^{ty} \frac{1}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})} z^{\frac{m}{2}-1} e^{-\frac{z}{2}} \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} y^{\frac{\nu}{2}-1} e^{-\frac{y}{2}} \mathrm{d}z\mathrm{d}y \\
&\overset{\nu z = myx}{=\!=\!=} \int_0^\infty \int_0^{\frac{\nu}{m}t} \frac{1}{2^{\frac{m+\nu}{2}}\Gamma(\frac{m}{2})\Gamma(\frac{\nu}{2})} \left(\frac{m}{\nu}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1} y^{\frac{\nu+m}{2}-1} e^{-\frac{1}{2}\left(\frac{m}{\nu}x+1\right)y} \mathrm{d}x\mathrm{d}y \\
&= \int_0^{\frac{\nu}{m}t} \int_0^\infty \frac{1}{2^{\frac{m+\nu}{2}}\Gamma(\frac{m}{2})\Gamma(\frac{\nu}{2})} \left(\frac{m}{\nu}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1} y^{\frac{\nu+m}{2}-1} e^{-\frac{1}{2}\left(\frac{m}{\nu}x+1\right)y} \mathrm{d}y\mathrm{d}x \\
& \qquad \text{(Fubini's Theorem)}
\end{aligned}$$

the integral expression for the density of $F_{m,\nu}$ is

$$f_{F_{m,\nu}}(x) = \int_0^\infty \frac{1}{2^{\frac{m+\nu}{2}}\Gamma(\frac{m}{2})\Gamma(\frac{\nu}{2})} \left(\frac{m}{\nu}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1} y^{\frac{\nu+m}{2}-1} e^{-\frac{1}{2}\left(\frac{m}{\nu}x+1\right)y} \mathrm{d}y.$$

Next we show this is equal to the pdf,

$$\begin{aligned}
& \int_0^\infty \frac{1}{2^{\frac{m+\nu}{2}}\Gamma(\frac{m}{2})\Gamma(\frac{\nu}{2})} \left(\frac{m}{\nu}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1} y^{\frac{\nu+m}{2}-1} e^{-\frac{1}{2}\left(\frac{m}{\nu}x+1\right)y} \mathrm{d}y \\
\overset{w=\frac{1}{2}\left(\frac{m}{\nu}x+1\right)y}{=\!=\!=\!=} & \int_0^\infty \frac{1}{\Gamma(\frac{m}{2})\Gamma(\frac{\nu}{2})} \left(\frac{m}{\nu}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1} \left(\frac{m}{\nu}x+1\right)^{-\frac{\nu+m}{2}} w^{\frac{\nu+m}{2}-1} e^{-w} \mathrm{d}w \\
= & \frac{\Gamma(\frac{m+\nu}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{\nu}{2})} \left(\frac{m}{\nu}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1} \left(\frac{m}{\nu}x+1\right)^{-\frac{\nu+m}{2}}.
\end{aligned}$$

□

8. (Casella and Berger Problem 5.22) Let X and Y be iid $N(0,1)$ random variables, and define $Z = \min(X, Y)$. Prove that $Z^2 \sim \chi_1^2$.

Proof. Since X and Y are independent, $X - Y \sim N(0,1)$,

$$\mathbb{P}(X \leq Y) = \mathbb{P}(X - Y \geq 0) = \frac{1}{2}, \qquad \mathbb{P}(X > Y) = 1 - \mathbb{P}(X \leq Y) = \frac{1}{2},$$

we have

$$\begin{aligned}
\mathbb{P}(Z^2 \leq z) &= \mathbb{P}(Z^2 \leq z, X \leq Y) + \mathbb{P}(Z^2 \leq z, X > Y) \\
&= \mathbb{P}(Z^2 \leq z | X \leq Y)\mathbb{P}(X \leq Y) + \mathbb{P}(Z^2 \leq z | X > Y)\mathbb{P}(X > Y) \\
&= \mathbb{P}(X^2 \leq z | X \leq Y)\mathbb{P}(X \leq Y) + \mathbb{P}(Y^2 \leq z | X > Y)\mathbb{P}(X > Y) \\
&= \frac{1}{2}\mathbb{P}(X^2 \leq z) + \frac{1}{2}\mathbb{P}(Y^2 \leq z) \\
&= \mathbb{P}(X^2 \leq z).
\end{aligned}$$

Since $X^2 \sim \chi_1^2$ and the distribution function of Z^2 is the same as X^2's, we have $Z^2 \sim \chi_1^2$. □