BIOS 7731 HW8 Tim Vigers

5.3.28) Suppose $X_1, \dots, X_{n_1}$ and $Y_1, \dots, Y_{n_2}$ are independent normal samples with $\mu_1 = E[X_1]$, $\sigma_1^2 = Var[X_1]$, $\mu_2 = E[Y_1]$ and $\sigma_2^2 = var[Y_1]$. Let $\Delta = \mu_2 - \mu_1$ and if $n_1, n_2 \to \infty$ $n_1/n \to \lambda$, $0 < \lambda < 1$, Define the two-sample pivot:

$$T(\Delta) = \sqrt{\frac{n_1 n_2}{n}} \frac{(\bar{Y} - \bar{X} - \Delta)}{s} \quad \text{where} \quad n = n_1 + n_2 \quad \text{and}$$

$$s = \frac{1}{n-2}\left[\sum_{i=1}^{n_1}(x_i - \bar{x})^2 + \sum_{j=1}^{n_2}(y_j - \bar{y})^2\right]^{\frac{1}{2}} = \left[\frac{(n_1-1)s_X^2}{n-2} + \frac{(n_2-1)s_Y^2}{n-2}\right]^{1/2}$$

where s_X^2 and s_Y^2 are the sample variances of X and Y, respectively.

First, we can re-arrange the numerator of $T(\Delta)$:

$$\sqrt{\frac{n_1 n_2}{n}}(\bar{Y} - \bar{X} - \Delta) = \sqrt{\frac{n_2}{n}}\sqrt{n_1}(\bar{X} - \mu_1) - \sqrt{\frac{n_1}{n}}\sqrt{n_2}(\bar{Y} - \mu_2)$$

By Slutsky's theorem and the CLT, we know that this converges in distribution to:

$$\sqrt{1-\lambda} * N(0, \sigma_1^2) - \sqrt{\lambda} * N(0, \sigma_2^2)$$

which is distributed as $N(0, (1-\lambda)\sigma_2^2 + \lambda\sigma_2^2)$. Thus, again by Slutsky's:

$$\sqrt{\frac{n_1 n_2}{n}} \frac{(\bar{Y} - \bar{X} - \Delta)}{s} \xrightarrow{d} \frac{N(0, (1-\lambda)\sigma_1^2 + \lambda\sigma_2^2)}{\sqrt{\lambda\sigma_1^2 + (1-\lambda)\sigma_2^2}}$$

which is distributed as $N\left(0, \frac{(1-\lambda)\sigma_1^2 + \lambda\sigma_2^2}{\lambda\sigma_1^2 + (1-\lambda)\sigma_2^2}\right)$. $\square$

5.3.8) a. Let $X_1, \dots, X_{n_1}$ be iid $N(\mu_1, \sigma_1^2)$ and $Y_1, \dots, Y_{n_1}$ be iid $N(\mu_2, \sigma_2^2)$. We want to construct the likelihood ratio $(\lambda(x))$ test of $H_0: \sigma_1^2 = \sigma_2^2$ vs $H_1: \sigma_1^2 \neq \sigma_2^2$:

$$\lambda(x) = \frac{\sup\{p(x, \theta) : \theta \in \Theta_0\}}{\sup\{p(x, \theta) : \theta \in \Theta\}} < c \qquad \theta = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$$

The unrestricted likelihood is maximized a $\hat{\theta} = (\bar{x}, \bar{y}, \hat{\sigma}_1^2, \hat{\sigma}_2^2)$ where $\hat{\sigma}_1^2 = \frac{\sum^{n_1}(x_i - \bar{x})^2}{n_1}$ and $\hat{\sigma}_2^2 = \frac{\sum^{n_2}(y_i - \bar{y})}{n_2}$ (see CB example 7.2.11 for details). Under the null hypothesis, $\sigma_1^2 = \sigma_2^2$, so the likelihood is:

$$L(\theta_0 | x) = \prod_{i=1}^{n_1} \frac{1}{\sqrt{2\pi}\sigma_*} \exp\left(\frac{-1}{2} \frac{(x_i - \mu_1)^2}{\sigma_*^2}\right) \prod_{j=1}^{n_2} \frac{1}{\sqrt{2\pi}\sigma_*} \exp\left(\frac{-1}{2} \frac{(y_i = \mu_2)}{\sigma_*^2}\right)$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^{n_1+n_2} \left(\frac{1}{\sigma_*}\right)^{n_1+n_2} \exp\left(\frac{-1 \sum^{n_1}(x_i - \mu_1)^2}{2\sigma_*^2}\right) \exp\left(\frac{-1 \sum^{n_2}(y - \mu_2)^2}{2\sigma_*^2}\right)$$

The MLEs for μ_1 and μ_2 will be the same as the unrestricted likelihood, because we are only interested in restricting σ_1^2 and σ_2^2. So to find our MLE for σ_*^2, we plug in our MLEs for μ_1 and μ_2, and take the log of the likelihood:

$$\ell(\theta_0 | x) = -\frac{m}{2}\log(2\pi) - \frac{m}{2}\log(\sigma_*^2) - \frac{1}{2\sigma_*^2}\left(\sum^{n_1}(x_i - \bar{x})^2 + \sum^{n_2}(y_i - \bar{y})^2\right)$$

where $m = n_1 + n_2$. Next we take the derivative of $\ell(\theta_0 | x)$ and set it to 0, then solve for σ_*^2:

$$\frac{\partial}{\partial \sigma_*^2} \ell(\theta_0|x) = \frac{-m}{2\sigma_*^2} + \frac{1}{2\sigma_*^4}\left(\sum^{n_1}(x_i-\bar{x})^2 + \sum^{n_2}(y_i-\bar{y})^2\right)$$

so $$\frac{\left(\sum^{n_1}(x_i-\bar{x})^2 + \sum^{n_2}(y_i-\bar{y})^2\right)}{\sigma_*^4} = \frac{m}{\sigma_*^2}$$

and $$\hat{\sigma}_*^2 = \frac{\left(\sum^{n_1}(x_i-\bar{x})^2 + \sum^{n_2}(y_i-\bar{y})^2\right)}{n_1+n_2}.$$

Now we have the whole test statistic:

$$\lambda(x) = \frac{\left(\frac{1}{\sqrt{2\pi}}\right)^m \left(\frac{1}{\hat{\sigma}_*^2}\right)^{\frac{m}{2}} \exp\left(\frac{-1}{2\hat{\sigma}_*^2}\left(\sum^{n_1}(x_i-\bar{x})^2 + \sum^{n_2}(y_i-\bar{y})^2\right)\right)}{\left(\frac{1}{\sqrt{2\pi}}\right)^m \left(\frac{1}{\hat{\sigma}_1^2}\right)^{\frac{n_1}{2}} \left(\frac{1}{\hat{\sigma}_2^2}\right)^{\frac{n_2}{2}} \exp\left(\frac{-1\sum^{n_1}(x_i-\bar{x})^2}{2\hat{\sigma}_1^2} - \frac{\sum^{n_2}(y_i-\bar{y})^2}{2\hat{\sigma}_2^2}\right)}$$

Because the MLEs for μ_1 and μ_2 are the same under both hypotheses, this simplifies to (BD pg. 262 for details):

$$\left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_*^2}\right)^{\frac{n_1}{2}} \left(\frac{\hat{\sigma}_2^2}{\hat{\sigma}_*^2}\right)^{\frac{n_2}{2}} = \left(\frac{(n_1-1)s_1^2/n_1}{((n_1-1)s_1^2+(n_2-1)s_2^2)/m}\right)^{n_1/2} \left(\frac{(n_2-1)s_2^2/n_2}{((n_1-1)s_1^2+(n_2-1)s_2^2)/m}\right)^{\frac{n_2}{2}}$$

With some simple algebra, the above reduces to a function of n_1, n_2 (known constants) and $\frac{s_1^2}{s_2^2}$. Thus,

n_1 and n_2 are absorbed into the constant c, and the test is based solely on $\frac{s_1^2}{s_2^2}$.

b. By CB definition 5.3.6, the random variable $F = \dfrac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2}$ has Snedecor's F distribution with $n-1$ and $m-1$ degrees of freedom. This assumes that $X_1, \ldots X_n \sim N(\mu_X, \sigma_X^2)$ and $Y_1, \ldots, Y_m \sim N(\mu_Y, \sigma_Y^2)$. This problem fulfills all of the assumptions above, therefore $\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1, n_2-2}$.

5.3.18) The aim of a variance stabilizing transformation is to find a function $h()$ such that $h(\bar{X})$ has asymptotically constant variance. In other words, we want to find $h()$ such that $\sigma^2[h'(\mu)]^2 = c > 0$ where σ^2 is the variance of the family of interest. For the case of $X_i \sim \text{Bernoulli}(\theta)$, $\text{var}(X) = \theta(1-\theta)$ so we set:

$$\theta(1-\theta)[h'(\theta)]^2 = c \quad , \text{ or } \quad h'(\theta) = \sqrt{\frac{c}{\theta(1-\theta)}}$$

In order to find $h()$, we find the indefinite integral:

$$\int \frac{\sqrt{c}}{\sqrt{\theta(1-\theta)}} d\theta = \sqrt{c} \int \frac{1}{\sqrt{\theta}\sqrt{1-\theta}} d\theta$$

substituting $u = \sqrt{\theta}$ $\quad d\theta = 2\sqrt{\theta}\, du$ gives us:

$$2\sqrt{c} \int \frac{u}{u\sqrt{1-u^2}} du = 2\sqrt{c} \int \frac{1}{\sqrt{1-u^2}} = 2\sqrt{c}\, \sin^{-1}(u) + d$$

$$= 2\sqrt{c}\, \sin^{-1}(\sqrt{\theta}) + d$$

where d is arbitrary.

Thus, because the rest of the terms are constant, we say $h(t) = \sin^{-1}(\sqrt{t})$. If we additionally want $h(0) = 0$ and $h(1) = 1$, we need a normaling constant:

$$h'(t) = \sin^{-1}(\sqrt{t}), \quad h(0) = 0, \quad \text{but } h(1) = \pi/2$$

Because $\sin^{-1}(0) = 0$, we only need to worry about $t=1$. So we use $h(t) = 2/\pi \sin^{-1}(\sqrt{t})$ and thus

$$h(t) = 2/\pi \sin^{-1}(\sqrt{t}) \qquad h(0) = 0 \qquad \text{and} \qquad h(1) = \pi/2 \; 2/\pi = 1$$

$\square$.

5.3.20) Let $B_{m,n}$ have a Beta distribution with parameters m, n s.t. as $m, n \to \infty$ $m/m+n \to \alpha$ with $0 < \alpha < 1$. Using the hint we can rewrite $B_{m,n}$ as:

$$\frac{m\bar{X}/n\bar{Y}}{1 + m\bar{X}/n\bar{Y}} = \frac{m\bar{X}}{n\bar{Y} + m\bar{X}}$$ where $X_1, \ldots, X_m$ and $Y_1, \ldots Y_n$ are iid exponential (1).

Then, by the CLT and Slutsky's we have:

$$B_{m,n} = \frac{m\bar{X}}{n\bar{Y} + m\bar{X}} - \frac{m}{m+n} \xrightarrow{d} N(0, \Sigma)$$

where Σ is the variance-covariance matrix of $m\bar{X}$ and $n\bar{Y}$ (BD 5.3.22). Then, by the multivariate delta method we can set:

$$g(u_n) = g(m, n, \bar{x}, \bar{y}) = \frac{m\bar{X}}{n\bar{Y} + m\bar{X}} \qquad \text{and}$$

$$g(m, n) = m/n+m$$

Thus, we can estimate σ_0^2 where $(g(U_n) - g(u)) \to N(0, \sigma_0^2)$ [Notice that this σ_0^2 contains a multivariate version of $\sqrt{n}$]:

$$\sigma_0^2 \approx g'(u) \Sigma [g'(u)]^T$$

First we find $g'(u) = \left(\frac{n}{(m+n)^2} , \frac{-m}{(m+n)^2} \right)$

and $\Sigma = \begin{pmatrix} var(m\bar{X}) & 0 \\ 0 & var(n\bar{Y}) \end{pmatrix} = \begin{pmatrix} m^2/m & 0 \\ 0 & n^2/n \end{pmatrix}$

Then we have $\sigma_0^2 = \frac{n m2}{(m+n)^4} + \frac{m^2 n}{(m+n)^4} = \frac{nm(m+n)}{(m+n)^4} = \frac{nm}{(m+n)^3}$

Because $\frac{m}{m+n} \to \alpha$ as $m, n \to \infty$, $\sigma_0^2 \to \frac{\alpha(1-\alpha)}{m+n}$.

Thus, given the above we can see that:

$$\frac{\sqrt{m+n}\left(B_{m,n} - m/m+n\right)}{\sqrt{\alpha(1-\alpha)}} \longrightarrow N(0,1) \qquad \square.$$

5.4.1) To be completely honest, I have no idea what to do with this problem (which is why I tried the extra credit instead). It's pretty clear that we need to use Taylor Series or the Delta method at the end, but I have no idea how to get there. One approach I tried was a pretty simple rearangement of the first formula:

$$\frac{1}{\sqrt{n}\,\tau(\theta)} \sum_{i=1}^{n} \left[\psi(X_i - \theta_n) - \lambda(\theta_n) \right] \xrightarrow{\mathcal{L}} N(0,1) \Rightarrow$$

$$\frac{\sqrt{n}}{n} \sum_{i=1}^{n} \left[\psi(X_i - \theta_n) - \lambda(\theta_n) \right] \xrightarrow{\mathcal{L}} N(0, \tau^2(\theta))$$

$$\frac{\left[\sum^{n} (X_i - \theta_n) - \lambda(\theta_n) \right]}{n} \xrightarrow{\mathcal{L}} N(0, \tau^2(\theta)/n)$$

The left side of this looks a little bit like $\bar{X}$, so I want to use the CLT but can't figure out how.

Another vague idea is that $\lambda'(\theta)$ exists and is in the variance of the final distribution. So, this suggests that $\lambda(\theta)$ is a version of $g()$ in this form of the Delta method:

$$\sqrt{n}\,(g(Y_n) - g(\theta)) \xrightarrow{\mathcal{L}} N(0, \sigma^2 (g'(\theta))^2)$$

But unfortunately I don't get much further than this (I've tried several different methods but they're not worth writing out here).

I look forward to seeing the solution for this one!

5) $k_n[\delta_n - g(\theta)]$ converges in distribution to H.

a. If $\frac{k'_n}{k_n} \to d \neq 0$ then by Slutsky's theorem

$$\frac{k'_n}{k_n} k_n[\delta_n - g(\theta)] \xrightarrow{\mathcal{L}} dH.$$

b. If $\frac{k'_n}{k_n} \to 0$, then (again) by Slutsky's :

$$\frac{k'_n}{k_n} k_n[\delta_n - g(\theta)] \to (0)H = 0$$

By the same token, if $\frac{k'_n}{k_n} \to \infty$ then :

$$\frac{k'_n}{k_n} k_n[\delta_n - g(\theta)] \to \infty * H = \infty.$$

c. If $k_n \to \infty$, then $\frac{1}{k_n}[k_n(\delta_n - g(\theta))] \xrightarrow{\mathcal{L}} 0 * H$

by BD A.14.19. Then, by A.14.4, $|\delta_n - g(\theta)| \xrightarrow{P} 0$ and therefore $\delta_n \xrightarrow{P} g(\theta)$. $\square$

6) Start with the proposition in the hint, which is Chebychev's inequality. Because g is non-negative and non-decreasing on the range of a random variable X:

$$g(a)1(Z \geq a) \leq g(Z)1(Z \geq a) \leq g(Z) \qquad \text{(BD A.15.5)}$$

Therefore $\quad g(a)\,P(Z \geq a) \leq E[g(Z)] \quad$ by A.10.8

and $\quad p(Z \geq a) \leq \dfrac{E[g(Z)]}{g(a)}.$

In this case we have a function p that meets the assumptions of g above, so, we know that

$$P(|\delta_n - g(\theta)| \geq \varepsilon) \leq \frac{E[p(\delta_n - g(\theta))]}{g(\varepsilon)}$$

If $E[p(\delta_n - g(\theta))] \to 0$, then

$$\lim_{n\to\infty} P(|\delta_n - g(\theta)| \geq \varepsilon) = 0 \qquad \text{because} \qquad (*)$$

$P(|\delta_n - g(\theta)| \geq \varepsilon)$ cannot be < 0, and therefore we have strict equality for Chebychevs. $(*)$ is the definition of a consistent estimator, and therefore δ_n is consistent for $g(\theta)$.