Statistics 581, Problem Set 6 Solutions

Wellner; 11/5/2014

1. Suppose that $X_1, \ldots, X_n$ are i.i.d. with the Weibull distribution F_θ given by

$$1 - F_\theta(x) = \exp(-(x/\alpha)^\beta), \qquad x \geq 0$$

where $\theta = (\alpha, \beta) \in (0, \infty) \times (0, \infty)$.
(a) Find the inverse (or quantile function) $F_\theta^{-1}(u)$ corresponding to F_θ in terms of α, β, and $u \in (0, 1)$, and show that

$$\log F_\theta^{-1}(u) = \log \alpha + \frac{1}{\beta} \log \log \left(\frac{1}{1-u} \right) .$$

(b) Fix $r \in (0, 1/2)$ and $s \in (1/2, 1)$ Use the $r-$th and $s-$th quantiles of the X_i's, namely $\mathbb{F}_n^{-1}(r)$ and $\mathbb{F}_n^{-1}(s)$, to obtain simple consistent estimators $\hat{\alpha}_n$ and $\hat{\beta}_n$ of α and β. Prove that your estimators are consistent.
(c) Prove that your estimators $\hat{\alpha}_n$ and $\hat{\beta}_n$ satisfy

$$\sqrt{n} \begin{pmatrix} \hat{\alpha}_n - \alpha \\ \hat{\beta}_n - \beta \end{pmatrix} \to_d N_2(0, \Sigma)$$

and identify Σ as a function of α, β, r, and s.
(d) How would you choose r and s to minimize the asymptotic variance of $\hat{\beta}_n$?

Solution: (a) Since $1 - F_\theta(x) = \exp(-(x/\alpha)^\beta)$, it follows we can solve $F_\theta(x) = u$ for $x = F_\theta^{-1}(u)$. This yields

$$F_\theta^{-1}(u) = \alpha(-\log(1-u))^{1/\beta} ,$$

or

$$\log F_\theta^{-1}(u) = \log \alpha + \frac{1}{\beta} \log \log \left(\frac{1}{1-u} \right) . \tag{0.1}$$

(b) Since we can estimate $F_\theta^{-1}(r)$ and $F_\theta^{-1}(s)$ respectively by $\mathbb{F}_n^{-1}(r)$ and $\mathbb{F}_n^{-1}(s)$ respectively, the relationship in (0.1) suggests that we estimate α and β as the solutions $\hat{\alpha}$ and $\hat{\beta}$ of the pair of equations

$$\log \mathbb{F}_n^{-1}(r) = \log \hat{\alpha} + \frac{1}{\hat{\beta}} \log \log 1/(1-r) , \tag{0.2}$$

$$\log \mathbb{F}_n^{-1}(s) = \log \hat{\alpha} + \frac{1}{\hat{\beta}} \log \log 1/(1-s) . \tag{0.3}$$

Letting $A_r \equiv \log \log 1/(1-r)$, and $B_s \equiv \log \log 1/(1-s)$, we find that

$$\begin{aligned} 1/\hat{\beta} &= \frac{1}{B_s - A_r} (\log \mathbb{F}_n^{-1}(s) - \log \mathbb{F}_n^{-1}(r)) \\ &\equiv a_{r,s} \log \mathbb{F}_n^{-1}(s) - a_{r,s} \log \mathbb{F}_n^{-1}(r) \end{aligned}$$

and

$$\begin{aligned}\log\hat{\alpha} &= \frac{-A_r}{B_s - A_r}\log\mathbb{F}_n^{-1}(s) + \frac{B_s}{B_s - A_r}\log\mathbb{F}_n^{-1}(r)) \\ &\equiv c_{r,s}\log\mathbb{F}_n^{-1}(s) + d_{r,s}\log\mathbb{F}_n^{-1}(r)\end{aligned}$$

where

$$a_{r,s} \equiv \frac{1}{B_s - A_r}, \qquad c_{r,s} \equiv -A_r a_{r,s} \qquad d_{r,s} \equiv B_s a_{r,s}\,.$$

Since $(\mathbb{F}_n^{-1}(r), \mathbb{F}_n^{-1}(s)) \to_{a.s.} (F_\theta^{-1}(r), F_\theta^{-1}(s))$, It follows easily by the continuous mapping theorem that

$$\frac{1}{\hat{\beta}} \to_{a.s.} a_{r,s}\log F_\theta^{-1}(s) - a_{r,s}\log F_\theta^{-1}(r) = \frac{1}{\beta}\,,$$

and

$$\log\hat{\alpha} \to_{a.s.} c_{r,s}\log F_\theta^{-1}(s) + d_{r,s}\log F_\theta^{-1}(r) = \log\alpha\,,$$

and hence by the continuous mapping theorem, $(\hat{\alpha}, \hat{\beta}) \to_{a.s.} (\alpha, \beta)$.
(c) First, we know that

$$\sqrt{n}\begin{pmatrix} \mathbb{F}_n^{-1}(r) - F^{-1}(r) \\ \mathbb{F}_n^{-1}(s) - F^{-1}(s) \end{pmatrix} \to_d \underline{Z} \sim N_2(0, \Sigma)$$

where

$$\Sigma = \begin{pmatrix} \frac{r(1-r)}{f^2(F^{-1}(r))} & \frac{r(1-s)}{f(F^{-1}(r))f(F^{-1}(s))} \\ \frac{r(1-s)}{f(F^{-1}(r))f(F^{-1}(s))} & \frac{s(1-s)}{f^2(F^{-1}(s))} \end{pmatrix}.$$

This implies that

$$\sqrt{n}\begin{pmatrix} \log\mathbb{F}_n^{-1}(r) - \log F^{-1}(r) \\ \log\mathbb{F}_n^{-1}(s) - \log F^{-1}(s) \end{pmatrix} \to_d D\underline{Z} \sim N_2(0, D\Sigma D^T)$$

where

$$D = \begin{pmatrix} 1/F^{-1}(r) & 0 \\ 0 & 1/F^{-1}(s) \end{pmatrix}.$$

Hence it follows that

$$\begin{aligned}&\sqrt{n}\begin{pmatrix} 1/\hat{\beta} - 1/\beta \\ \log\hat{\alpha} - \log\alpha \end{pmatrix} \\ &= M\sqrt{n}\begin{pmatrix} \log\mathbb{F}_n^{-1}(r) - \log F^{-1}(r) \\ \log\mathbb{F}_n^{-1}(s) - \log F^{-1}(s) \end{pmatrix} \\ &\to_d MD\underline{Z} \sim N_2(0, MD\Sigma D^T M^T)\,.\end{aligned}$$

where

$$M = \begin{pmatrix} -a_{r,s} & a_{r,s} \\ d_{r,s} & c_{r,s} \end{pmatrix} = a_{r,s}\begin{pmatrix} -1 & 1 \\ B_s & -A_r \end{pmatrix}.$$

Finally, with $g(x,y) = (g_1(x), g_2(y))$, $g_1(x) = 1/x$, $g_2(y) = \exp y$, we find, by the delta-method, that

$$\begin{aligned}\sqrt{n}\begin{pmatrix} \hat{\beta} - \beta \\ \hat{\alpha} - \alpha \end{pmatrix} \\ \to_d \quad \nabla g M D \underline{Z} \sim N_2(0, \nabla g M D \Sigma D^T M^T \nabla g^T)\end{aligned}$$

where

$$\nabla g = \begin{pmatrix} -\beta^2 & 0 \\ 0 & \alpha \end{pmatrix}.$$

We begin combining all this by noting that $D\Sigma D^T$ involves the function

$$\begin{aligned} F^{-1}(u) f(F^{-1}(u)) &= \alpha \left(\log\left(\frac{1}{1-u}\right)\right)^{1/\beta} \frac{\beta}{\alpha} \left(\log\left(\frac{1}{1-u}\right)\right)^{(\beta-1)/\beta} (1-u) \\ &= \beta(1-u)\log\left(\frac{1}{1-u}\right) \equiv \beta g(u) \end{aligned}$$

at the points $u = r$ and $u = s$. Computing $D\Sigma D^T$ yields

$$D\Sigma D^T = \beta^{-2} \begin{pmatrix} \frac{r(1-r)}{g(r)^2} & \frac{r(1-s)}{g(r)g(s)} \\ \frac{r(1-s)}{g(r)g(s)} & \frac{s(1-s)}{g(s)^2} \end{pmatrix} \equiv \beta^{-2} \begin{pmatrix} c_{11}(r,s) & c_{12}(r,s) \\ c_{12}(r,s) & c_{22}(r,s) \end{pmatrix}.$$

Since the matrix M just depends on r, s, we find that the matrix

$$M D \Sigma D^T M^T = \beta^{-2} a_{r,s}^2 \begin{pmatrix} d_{11}(r,s) & d_{12}(r,s) \\ d_{12}(r,s) & d_{22}(r,s) \end{pmatrix},$$

where

$$\begin{aligned} d_{11}(r,s) &= c_{11}(r,s) - 2c_{12}(r,s) + c_{22}(r,s) \\ d_{12}(r,s) &= B_s(c_{12}(r,s) - c_{11}(r,s)) - A_r(c_{22}(r,s) - c_{12}(r,s)) \\ d_{22}(r,s) &= A_r^2 c_{22}(r,s) - 2A_r B_s c_{12}(r,s) + B_s^2 c_{11}(r,s). \end{aligned}$$

Thus we conclude that the asymptotic covariance matrix of $(\hat{\beta}, \hat{\alpha})$ is given by

$$\nabla g M D \Sigma D^T M^T \nabla g^T = a_{r,s}^2 \begin{pmatrix} \beta^2 d_{11}(r,s) & -\alpha d_{12}(r,s) \\ -\alpha d_{12}(r,s) & (\alpha/\beta)^2 d_{22}(r,s) \end{pmatrix}.$$

(d) The asymptotic variance of $\hat{\beta}$ is

$$\beta^2 a_{r,s}^2 d_{11}(r,s) = \beta^2 \left(c_{11}(r,s) - 2c_{12}(r,s) + c_{22}(r,s)\right) a_{r,s}^2.$$

This is minimized by $r = r_0 \approx .1704$, $s = s_0 \approx .97$, and the minimum value is $\beta^2(.917) > \beta^2(6/\pi^2)$. This ad-hoc estimator $\hat{\beta}$ based on quantiles is *inefficient*; its asymptotic variance (for any value of r, s, including the minimizing r_0, s_0) is larger than the best possible asymptotic variance, which is $\beta^2(6/\pi^2)$ as we will see in Chapter 3. In fact the ARE when $r = r_0$ and $s = s_0$ is $(6/\pi^2)/.917 = .663$)

The asymptotic variance of $\hat{\alpha}$ is

$$(\alpha/\beta)^2 a_{r,s}^2 d_{22}(r,s) = (\alpha/\beta)^2 \left(B_s^2 c_{11}(r,s) - 2A_r B_s c_{12}(r,s) + B_s^2 c_{22}(r,s)\right).$$

This is minimized by $r = r_0 \approx .398$, $s = s_0 \approx .82$, and the minimum value is $(\alpha/\beta)^2(1.359) > (\alpha/\beta)^2(1.11)$. This ad-hoc estimator $\hat{\alpha}$ based on quantiles is also *inefficient*; its asymptotic variance (for any value of r, s, including the minimizing r_0, s_0) is larger than the best possible asymptotic variance, which is about $(\alpha/\beta)^2(1.11)$ as we will see in Chapter 3. For the estimator based on the optimal r_0, s_0 (for α!), the ARE is $\approx 1.109/1.359 = .816$

2. Ferguson, ACILST, problem 6, page 93, plus the following:
(d) Construct a family of estimators $\tilde{\theta}_n$ of θ based on the sample quantile function $\mathbb{F}_n^{-1}(t)$. Show that your estimators are consistent and asymptotically normal. Give a formula for the asymptotic variance of your estimators.

Solution: (a) Now $f_\theta(x) = \theta x^{\theta-1} 1_{(0,1)}(x)$, so $F_\theta(x) = x^\theta$ for $0 \le x \le 1$, and $F_\theta^{-1}(t) = t^{1/\theta}$ for $0 \le t \le 1$. This yields $m(\theta) = F_\theta^{-1}(1/2) = 2^{-1/\theta}$, $f_\theta(m(\theta)) = \theta 2^{-(\theta-1)/\theta} = 2^{-1}\theta 2^{1/\theta}$. Thus from Theorem 2.6.2,

$$\begin{aligned}
\sqrt{n}(M_n - m(\theta)) &= \sqrt{n}(\mathbb{F}_n^{-1}(1/2) - F_\theta^{-1}(1/2)) \\
&\to_d Q'_\theta(1/2)\mathbb{V}(1/2) \sim N\left(0, \frac{1/4}{f_\theta^2(m(\theta))}\right) \\
&= N\left(0, \frac{1}{\theta^2 2^{2/\theta}}\right).
\end{aligned}$$

(b) With $g(x) \equiv \log(1/2)/\log(x)$ we have $\widehat{\theta}_n = g(M_n)$. Since g is continuous on $(0,1)$ and $M_n = \mathbb{F}_n^{-1}(1/2) \to_p F_\theta^{-1}(1/2) = m(\theta)$ it follows by the Mann-Wald theorem that

$$\widehat{\theta}_n = g(M_n) \to_p g(m(\theta)) = \frac{\log(1/2)}{\log(2^{-1/\theta})} = \theta \frac{\log(1/2)}{\log(1/2)} = \theta.$$

(c) Note that

$$g'(x) = -\log(1/2)/(x(\log(x))^2) = -\frac{g^2(x)}{x \log(1/2)},$$

and hence

$$g'(m(\theta)) = -\frac{g^2(m(\theta))}{m(\theta)\log(1/2)} = -\frac{\theta^2}{2^{-1/\theta}\log(1/2)} = \frac{-\theta^2 2^{1/\theta}}{\log(1/2)}.$$

Thus it follows from the delta-method that

$$\begin{aligned}
\sqrt{n}(\widehat{\theta}_n - \theta) &= \sqrt{n}(g(M_n) - g(m(\theta))) \\
&\to_d g'(m(\theta))N(0, \theta^{-2}2^{-2/\theta}) = \frac{-\theta^2 2^{1/\theta}}{\log(1/2)} N(0, \theta^{-2}2^{-2/\theta}) \\
&= N(0, \theta^2/(\log(1/2))^2).
\end{aligned}$$

(d) For any fixed $t \in (0,1)$, $F_\theta^{-1}(t) = t^{1/\theta}$, so $\log F_\theta(t) = \theta^{-1}\log t$, and hence $\theta = \log t/\log F_\theta^{-1}(t) \equiv g_t(F_\theta^{-1}(t))$ with $g_t(x) \equiv \log t/\log x$ is consistently estimated by $\widehat{\theta}_{n,t} \equiv \log t/\log(\mathbb{F}_n^{-1}(t)) = g(\mathbb{F}_n^{-1}(t))$. Now

$$f_\theta(F_\theta^{-1}(t)) = \theta(t^{1/\theta})^{\theta-1} = \theta \cdot t \cdot t^{-1/\theta}$$

Thus by Theorem 2.6.2 we have

$$\sqrt{n}(\mathbb{F}_n^{-1}(t) - F_\theta^{-1}(t)) \to_d Q'_\theta(t)\mathbb{V}(t) \sim N\left(0, \frac{t(1-t)t^{2/\theta}}{\theta^2 t^2}\right).$$

Since

$$g'_t(x) = \frac{-\log t}{x(\log x)^2} = \frac{g^2(x)}{x \log t},$$

we have

$$g'_t(F_\theta^{-1}(t)) = \frac{g^2(F_\theta^{-1}(t))}{t^{1/\theta}\log t} = \frac{\theta^2}{t^{1/\theta}\log t}.$$

Thus, by the delta-method,

$$\begin{aligned}
\sqrt{n}(\widehat{\theta}_{n,t} - \theta) &= \sqrt{n}(g_t(\mathbb{F}_n(t)) - g_t(F_\theta(t))) \\
&\to_d g'_t(F_\theta^{-1}(t)) N\left(0, \frac{t(1-t)t^{2/\theta}}{\theta^2 t^2}\right) \\
&=_d N\left(0, \frac{\theta^2(1-t)/t}{(\log t)^2}\right).
\end{aligned}$$

Note that this reduces to the result in (c) when $t = 1/2$.

3. Compute and plot the score for location $-f'/f(x)$ when:
(a) $f = \phi$;
(b) $f(x) = \exp(-x)/(1+\exp(-x))^2$ (logistic);
(c) $f(x) = (1/2)\exp(-|x|)$ (double exponential);
(d) $f(x) = t_k$, the $t-$density with $k-$degrees of freedom
(e) $f(x) = \exp(-x)\exp(-\exp(-x))$.

Solution: (a) and 5(a): For $f(x) = (2\pi)^{-1/2}\exp(-x^2)$, we have $\log f(x) = -x^2/2 - (1/2)\log(2\pi)$, so that $-(f'/f)(x) = x$ and $-1 - x(f'(x)/f(x)) = x^2 - 1$.
(b) and 5(b): For $f(x) = e^{-x}/(1+e^{-x})^2$, we have $\log f(x) = -x - 2\log(1+e^{-x})$ and hence

$$-\frac{f'(x)}{f(x)} = \frac{1-e^{-x}}{1+e^{-x}},$$

while

$$-1 - x\frac{f'(x)}{f(x)} = x\frac{1-e^{-x}}{1+e^{-x}} - 1 \sim |x| - 1 \quad \text{as} \quad x \to \infty.$$

(c) and 5(c): For $f(x) = (1/2)\exp(-|x|)$ we have $\log f(x) = |x| - \log 2$, and hence

$$-\frac{f'(x)}{f(x)} = \begin{cases} -1, & x < 0, \\ \text{undefined}, & x = 0 \\ +1, & x > 0. \end{cases}$$

and

$$-1 - x\frac{f'(x)}{f(x)} = |x| - 1, \quad \text{for} \quad x \neq 0.$$

(d) and 5(d): For the t_k density

$$f_(x) = \frac{\Gamma(\frac{k+1}{2})}{\Gamma(k/2)} \frac{1}{\sqrt{\pi k}} \frac{1}{(1+x^2/k)^{(k+1)/2}},$$

$$\log f(x) = -\frac{k+1}{2} \log\left(1 + \frac{x^2}{k}\right) + \text{a constant},$$

and hence

$$-\frac{f'(x)}{f(x)} = \frac{k+1}{k} \frac{x}{1+\frac{x^2}{k}},$$

while

$$-1 - x\frac{f'(x)}{f(x)} = k\frac{x^2-1}{k+x^2}.$$

(e) and 5(e): For $f(x) = \exp(-x)\exp(-\exp(-x))$ we find that

$$\log f(x) = -x - \exp(-x)$$

and hence

$$-\frac{f'(x)}{f(x)} = 1 - \exp(-x)$$

and

$$-1 - x\frac{f'(x)}{f(x)} = -1 + x(1-\exp(-x)).$$

4. Compute the information for location for each of the densities f in problem 3 above.

Solution: (a) For the normal density $f = \phi$, $I_f = \int x^2\phi(x)dx = Var(Z) = 1$ where $Z \sim N(0,1)$.
(b) For the logistic density the information for location is

$$\begin{aligned} I_f &= \int_{-\infty}^{\infty} \left(\frac{1-e^{-x}}{1+e^{-x}}\right)^2 dF(x) \\ &= \int_{-\infty}^{\infty} (2F(x)-1)^2 dF(x) = \int_0^1 (2u-1)du = 4Var(U) \\ &= 4/12 = 1/3 \end{aligned}$$

where $U \sim \text{Uniform}(0,1)$.
(c) For the double exponential density $-(f'/f)^2(x) = 1$, so $I_f = 1$.
(d) For the t_k density, by using a change of variables and letting T_r denote a random variable with the t_r distribution with r degrees of freedom,

$$\begin{aligned} I_f &= \int_{-\infty}^{\infty} \left(\frac{k+1}{k}\right)^2 \frac{x^2}{(1+x^2/k)^2} \frac{\Gamma((k+1)/2)}{\Gamma(k/2)\sqrt{\pi k}} \frac{1}{(1+x^2/k)^{(k+1)/2}} dx \\ &= \frac{(k+1)(k+2)}{(k+4)(k+3)} Var(T_{k+4}) \\ &= \frac{(k+1)(k+2)}{(k+4)(k+3)} \frac{k+4}{k+2} \\ &= \frac{k+1}{k+3} \end{aligned}$$

since $Var(T_r) = r/(r-2)$ for $r > 2$.
(e) For the extreme value distribution, $F(x) = \exp(-\exp(-x))$, and therefore if $X \sim F$ the random variable $Y = \exp(-X) \sim$ exponential(1):

$$P(Y \geq y) = P(\exp(-X) \geq y) = P(X \leq -\log(y)) = \exp(-\exp(\log y)) = \exp(-y).$$

Since $-(f'/f)(x) = -1 + e^{-x}$, it is easily seen that

$$I_f = E\{(-f'/f)^2(X)\} = E\{(e^{-X} - 1)^2\} = E(Y-1)^2 = Var(Y) = 1.$$

5. Compute and plot the score for scale for the all the densities f in the previous problem and (f) $f(x) = \exp(-x)1_{(0,\infty)}(x)$.

   Solution: See the solution for problem 3 above.

6. **Optional bonus problem 1:** (a) Let $\mathbb{U}_X$ and $\mathbb{U}_Y$ be two independent Brownian bridge processes on $[0,1]$, and let $\lambda \in [0,1]$. Show that the process $\mathbb{U}$ defined by $\mathbb{U} = \sqrt{1-\lambda}\mathbb{U}_X - \sqrt{\lambda}\mathbb{U}_Y$ is also a Brownian bridge process.
   (b) Suppose that $X_1, \ldots, X_m$ are i.i.d. F and $Y_1, \ldots, Y_n$ are i.i.d. G, with the X's and Y's independent. Let $\mathbb{F}_m$ and $\mathbb{G}_n$ denote the empirical df's of the X's and Y's respectively. Suppose that $\lambda_N \equiv m/N \to \lambda \in (0,1)$ where $N \equiv m+n$. Show that

$$\begin{aligned} \mathbb{X}_{m,n} &\equiv \sqrt{\frac{mn}{N}}(\mathbb{F}_m - F) - \sqrt{\frac{mn}{N}}(\mathbb{G}_n - G) \\ &\Rightarrow \sqrt{1-\lambda}\mathbb{U}_X(F) - \sqrt{\lambda}\mathbb{U}_Y(G) \equiv \mathbb{X}. \end{aligned}$$

   (c) A distribution function F is said to be *stochastically smaller than another distribution function* G, and we write $F <_s G$, if $F(x) \geq G(x)$ for all $x \in \mathbb{R}$ with strict inequality for some $x \in \mathbb{R}$. Note that this means $F^{-1}(u) \leq G^{-1}(u)$ for all $0 < u < 1$ so that the random variables resulting from a Skorokhod construction with one uniform random variable ξ satisfy satisfy $X^* \equiv F^{-1}(\xi) \leq G^{-1}(\xi) \equiv Y^*$. Consider testing $H_0: F = G$ continuous versus $H_1: F <_s G$ based on the one-sided Kolmogorov-Smirnov statistic

$$D^+_{m,n} \equiv \sqrt{\frac{mn}{N}}\|(\mathbb{F}_m - \mathbb{G}_n)^+\|_\infty = \sqrt{\frac{mn}{N}}\sup_{x\in\mathbb{R}}(\mathbb{F}_m(x) - \mathbb{G}_n(x));$$

   here the notation f^+ is the *positive part* of the function f: $f^+(x) \equiv \max\{f(x), 0\}$. Use the result of (b) to show that under H_0 it follows that

$$D^+_{m,n} \to_d \|\mathbb{U}^+\|_\infty = \sup_{0\leq t\leq 1} \mathbb{U}(t).$$

   (d) To test the effectiveness of vitamin B_1 in stimulating growth in mushrooms, vitamin B_1 was applied to 13 mushrooms selected at random from a group of 24, while the remaining 11 did not receive this treatment. The weights of the mushrooms at the end of the period of observation were:

$\underline{X} = (18, 14.5, 13.5, 12.5, 23, 24, 21, 17, 18.5, 9.5, 14)$, $m = 11$;
$\underline{Y} = (27, 34, 20.5, 29.5, 20, 28, 20, 26.5, 22, 24.5, 34, 35.5, 19)$, $\quad n = 13$.

Plot the two empirical df's and compute $D^+_{m,n}$. What is the approximate p-value for testing H_0 versus $H_1 : F <_s G$? You may use your favorite tables of the distribution of $D^+_{m,n}$, or the asymptotic distribution.

Solution: (a) Let $\mathbb{Z} \equiv \sqrt{1-\lambda}\mathbb{U}_X - \sqrt{\lambda}\mathbb{U}_Y$. Then

$$\begin{aligned} E\mathbb{U}(t) &= \sqrt{1-\lambda}E\mathbb{U}_X(t) - \sqrt{\lambda}E\mathbb{U}_Y(t) = \sqrt{1-\lambda}\cdot 0 - \sqrt{\lambda}\cdot 0 = 0, \\ Cov[\mathbb{U}(s), \mathbb{U}(t)] &= (1-\lambda)Cov(\mathbb{U}_X(s), \mathbb{U}_X(t)) + \lambda Cov(\mathbb{U}_Y(s), \mathbb{U}_Y(t)) \\ &\qquad \text{since } \mathbb{U}_X, \mathbb{U}_Y \text{ are independent} \\ &= (1-\lambda)(s \wedge t - st) + \lambda(s \wedge t - st) = s \wedge t - st. \end{aligned}$$

Thus $\mathbb{U}$ is a mean zero Gaussian process with covariance $s \wedge t - st$; it follows that $\mathbb{U}$ is a Brownian bridge process on $[0, 1]$.
(b) This follows from the continuous mapping (Mann - Wald) theorem: with $\mathbb{Z}_m \equiv \sqrt{m}(\mathbb{F}_m - F)$, $\mathbb{W}_n \equiv \sqrt{n}(\mathbb{G}_n - G)$, and $\lambda_N \equiv m/(m+n) \equiv m/N$, we have $(\mathbb{Z}_m, \mathbb{W}_n, \lambda_N) \Rightarrow (\mathbb{Z}, \mathbb{W}, \lambda)$ where $\mathbb{Z} \equiv \mathbb{U}_X(F)$ and $\mathbb{W} \equiv \mathbb{U}_Y(G)$ are independent F and G Brownian bridge processes on $\mathbb{R}$ respectively. Since the function $g : D(\mathbb{R}) \times D(\mathbb{R}) \times \mathbb{R} \mapsto D(\mathbb{R})$ defined by

$$g(z, w, \lambda) = \sqrt{1-\lambda}z - \sqrt{\lambda}w$$

is continuous (with respect to $\|\cdot\|_\infty \vee \|\cdot\|_\infty \vee |\cdot|$), it follows that $g(\mathbb{Z}_m, \mathbb{W}_n, \lambda_N) \Rightarrow g(\mathbb{Z}, \mathbb{W}, \lambda)$.

Here is an alternative argument (using Skorokhod constructions) that makes the continuity of g more explicit: For special constructions we have

$$\sqrt{m}(\mathbb{F}_m - F) \stackrel{d}{=} \mathbb{U}^*_{X,m}(F) \qquad \text{and} \qquad \sqrt{n}(\mathbb{G}_n - G) \stackrel{d}{=} \mathbb{U}^*_{Y,n}(G)$$

where the starred processes satisfy

$$\|\mathbb{U}^*_{X,m}(F) - \mathbb{U}^*_X(F)\|_\infty \to_{a.s.} 0, \qquad \text{and} \qquad \|\mathbb{U}^*_{Y,n}(G) - \mathbb{U}^*_X(G)\|_\infty \to_{a.s.} 0;$$

here $\mathbb{U}^*_X$ and $\mathbb{U}^*_Y$ are independent Brownian bridge processes Thus, with

$$\begin{aligned} \mathbb{X}^*_{m,n} &\equiv \sqrt{1-\lambda_N}\mathbb{U}^*_{X,m}(F) - \sqrt{\lambda_N}\mathbb{U}_{Y,n}(G), \\ \mathbb{X}^*_N &\equiv \sqrt{1-\lambda_N}\mathbb{U}^*_X(F) - \sqrt{\lambda_N}\mathbb{U}^*_Y(G), \text{and} \\ \mathbb{X}^* &\equiv \sqrt{1-\lambda}\mathbb{U}^*_X(F) - \sqrt{\lambda}\mathbb{U}^*_Y(G), \end{aligned}$$

it follows that

$$\begin{aligned} \|\mathbb{X}^*_{m,n} - \mathbb{X}^*_N\|_\infty &= \|\sqrt{1-\lambda_N}(\mathbb{U}^*_{X,m}(F) - \mathbb{U}^*_X(F)) - \sqrt{\lambda_N}(\mathbb{U}^*_{Y,n}(G) - \mathbb{U}^*_Y(G)\|_\infty \\ &\leq \sqrt{1-\lambda_N}\|\mathbb{U}^*_{X,m}(F) - \mathbb{U}^*_X(F)\|_\infty + \sqrt{\lambda_N}\|\mathbb{U}^*_{Y,n}(G) - \mathbb{U}^*_Y(G)\|_\infty \\ &\to_{a.s.} 0 + 0 = 0. \end{aligned} \tag{0.4}$$

Furthermore

$$\|\mathbb{X}_N^* - \mathbb{X}^*\|_\infty \leq |\sqrt{1-\lambda_N} - \sqrt{1-\lambda}|\|\mathbb{U}_X\|_\infty + |\sqrt{\lambda_N} - \sqrt{\lambda}|\mathbb{U}_Y\|_\infty \to_{a.s.} 0. \quad (0.5)$$

Combining (0.4) and (0.5) using the triangle inequality yields

$$\|\mathbb{X}_{m,n}^* - \mathbb{X}^*\|_\infty \to_{a.s.} 0.$$

(c) Under $H_0: F = G$ continuous, it follows that

$$\begin{aligned} D_{m,n} &= \|\sqrt{\frac{mn}{N}}(\mathbb{F}_m - \mathbb{G}_n)^+\|_\infty \\ &= \|\left\{\sqrt{\frac{mn}{N}}(\mathbb{F}_m - F) - \sqrt{\frac{mn}{N}}(\mathbb{G}_n - G)\right\}^+\|_\infty \\ &\to_{a.s.} \|(\mathbb{X}^*)^+\|_\infty = \|(\mathbb{Z}^*)^+\|_\infty \stackrel{d}{=} \|\mathbb{U}^+\|_\infty \end{aligned}$$

where $\mathbb{U}$ is a Brownian bridge process. Hence $D_{m,n}^+ \to_d \|\mathbb{U}^+\|_\infty = \sup_{0<t<1} \mathbb{U}(t)$.
(d) A quick look at a plot of the empirical distributions $\mathbb{F}_m$ and $\mathbb{G}_n$ convinces one that there is indeed a difference, so there is relatively little need for the formal test. Nevertheless, the observed significance level or $P-$value gives some indication of the magnitude of the difference. I compute $D_{m,n}^+ = \sqrt{(11 \cdot 13)/24}(8/11) \approx 1.7752\ldots$ for the observed data; the supremum is achieved on the interval $[18.5, 19) = [X_{(8)}, Y_{(1)})$. This yields an approximate $P-$value of $\exp(-2(1.7752)^2) \approx .0018\ldots$.

Unfortunately, tables of the exact distribution of $D_{m,n}^+$ for small sample sizes $m \neq n$ do not seem to exist. The best available tables for $D_{m,n}$ for $m \neq n$ with $m \leq n \leq 100$ are those of Kim and Jennrich in *Selected Tables in Mathematical Statistics, Volume I*, pages 79 - 170, American Mathematical Society, Providence, 1970, H. L. Harter and D. B. Owen, editors. The original tables of Massey (1952), *Ann. Math. Statist.* **23**, for $D_{m,n}^+$ and $D_{m,n}$ only go through $m, m \leq 10$. It is also true that the convergence in distribution of the two-sample Kolmogorov-Smirnov statistic occurs quite quickly for $m = n \to \infty$, but slowly for $m \neq m \wedge n \to \infty$; this was studied carefully by J. L. Hodges (1958), "The significance probability of the Smirnov two-sample test", *Archiv for Mathematik* **3**, 469 - 486. What seems to be needed is a simple and reliable algorithm implementing the graphical recursion methods of Hodges for small $m \neq n$; see e.g. J. D. Gibbons (1971), *Nonparametric Statistical Inference*, pages 128 - 131.

7. **Optional bonus problem 2:** Course notes, Chapter 2, Exercise 6.6, page 37: Let $\mathbb{U}$ be a Brownian bridge process. For $0 \leq t < \infty$ define a process $\mathbb{B}$ by $\mathbb{B}(t) = (1+t)\mathbb{U}(t/(1+t))$. Show that $\mathbb{B}$ is a Brownian motion process on $[0, \infty)$.

Solution: $\mathbb{B}$ is clear a Gaussian process since $\mathbb{U}$ is Gaussian. For any fixed t we have $E(\mathbb{B}(t)) = E\{(1+t)\mathbb{U}(t/(1+t))\} = (1+t)E\{\mathbb{U}(t/(1+t))\} = (1+t) \cdot 0 = 0$. Moreover, for $0 \leq s \leq t < \infty$ we find that

$$\begin{aligned} Cov(\mathbb{B}(s), \mathbb{B}(t)) &= (1+s)(1+t)E\left\{\mathbb{U}\left(\frac{s}{1+s}\right)\mathbb{U}\left(\frac{t}{1+t}\right)\right\} \\ &= (1+s)(1+t)\left\{\frac{s}{1+s}\left(1 - \frac{t}{1+t}\right)\right\} \\ &= s(1+t) - st = s = s \wedge t. \end{aligned}$$

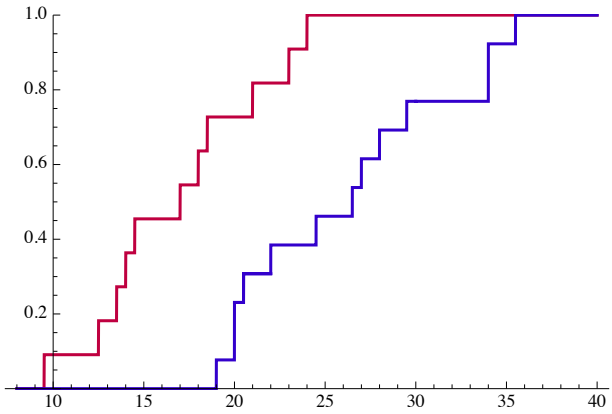

Figure 1: The empirical distribution $\mathbb{F}_m$ (of X's, red – to the left), and empirical distribution $\mathbb{G}_n$ of (Y's, blue – to the right)

It follows that $\mathbb{B}$ is a standard Brownian motion process on $[0, \infty)$.

8. **Optional bonus problem 3:** Chapter 2, Exercise 6.3, page 35. [Hint: One approach uses the fact that $\mathbb{S}_n(t_j) - \mathbb{S}_n(t_{j-1}) = n^{-1/2} \sum_{i=[nt_{j-1}]+1}^{[nt_j]} X_i$, $j = 1, \ldots, k$ with $t_0 \equiv 0$ are independent random variables.]

Solution: Note that

$$\begin{pmatrix} \mathbb{S}_n(t_1) \\ \mathbb{S}_n(t_2) \\ \vdots \\ \mathbb{S}_n(t_k) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \mathbb{S}_n(t_1) \\ \mathbb{S}_n(t_2) - \mathbb{S}_n(t_1) \\ \vdots \\ \mathbb{S}_n(t_k) - \mathbb{S}_n(t_{k-1}) \end{pmatrix}.$$

where the components of the vector on the right side are independent and

$$\begin{aligned} \mathbb{S}_n(t_j) - \mathbb{S}_n(t_{j-1}) &= n^{-1/2} \sum_{i=[nt_{j-1}]+1}^{[nt_j]} X_i \\ &= \sqrt{\frac{[nt_j] - [nt_{j-1}]}{n}} \frac{1}{\sqrt{[nt_j] - [nt_{j-1}]}} \sum_{i=[nt_{j-1}]+1}^{[nt_j]} X_i \\ &\to_d \mathbb{S}(t_j) - \mathbb{S}(t_{j-1}) \sim \sqrt{t_j - t_{j-1}} N(0,1) = N(0, t_j - t_{j-1}). \end{aligned}$$

Thus it follows that

$$\begin{pmatrix} \mathbb{S}_n(t_1) \\ \mathbb{S}_n(t_2) - \mathbb{S}_n(t_1) \\ \vdots \\ \mathbb{S}_n(t_k) - \mathbb{S}_n(t_{k-1}) \end{pmatrix} \to_d \begin{pmatrix} \mathbb{S}(t_1) \\ \mathbb{S}(t_2) - \mathbb{S}(t_1) \\ \vdots \\ \mathbb{S}(t_k) - \mathbb{S}(t_{k-1}) \end{pmatrix}$$

where the coordinates of the vector on the right side are independent. (This justifies the notation, since Brownian motion has independent increments.) Then

the continuous mapping (or Mann-Wald) theorem yields

$$\begin{aligned}
\begin{pmatrix} \mathbb{S}_n(t_1) \\ \mathbb{S}_n(t_2) \\ \vdots \\ \mathbb{S}_n(t_k) \end{pmatrix} &= \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \mathbb{S}_n(t_1) \\ \mathbb{S}_n(t_2) - \mathbb{S}_n(t_1) \\ \vdots \\ \mathbb{S}_n(t_k) - \mathbb{S}_n(t_{k-1}) \end{pmatrix} \\
&\to_d \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \mathbb{S}(t_1) \\ \mathbb{S}(t_2) - \mathbb{S}(t_1) \\ \vdots \\ \mathbb{S}(t_k) - \mathbb{S}(t_{k-1}) \end{pmatrix} \\
&= \begin{pmatrix} \mathbb{S}(t_1) \\ \mathbb{S}(t_2) \\ \vdots \\ \mathbb{S}(t_k) \end{pmatrix} \sim N_k(0, (t_j \wedge t_{j'})_{j,j'=1}^k).
\end{aligned}$$