Statistics 581

Problem Set 5 Solutions

Wellner; 11/2/2017

1. Suppose that $X_i = \mu + \sigma_i \epsilon_i$ where $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. with some distribution function F with $E(\epsilon_1) = 0$ and $Var(\epsilon_1) = 1 < \infty$. Consider estimators of μ of the form $T_n \equiv T_n(w) = \sum_{i=1}^n w_{ni} X_i$ where $w = w_n = (w_{n1}, \ldots, w_{nn})$ is a vector of weights with $\sum_{i=1}^n w_{ni} = 1$.
(a) Show that all the estimators $T_n(w)$ are unbiased, and that the choice of weights which minimizes $Var(T_n(w))$ is

$$w_{ni}^{opt} = \frac{1/\sigma_i^2}{\sum_{j=1}^n (1/\sigma_j^2)} \qquad \text{for } \ i = 1, \ldots, n\,. \tag{1}$$

(b) Compute $Var(T_n(w^{opt}))$ and show that $T_n(w^{opt})$ is a consistent estimator of μ if $\sum_1^n (1/\sigma_j^2) \to \infty$.
(c) Now suppose that $X_i = \mu + \sigma_i \epsilon_i$ where $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. with some distribution function F with $E(\epsilon_1) = 0$ and $Var(\epsilon_1) = 1 < \infty$ as in 2(b) above. Show that

$$\sqrt{\sum_1^n (1/\sigma_i^2)}\, \big(T_n(w^{opt}) - \mu\big) \to_d N(0,1)$$

if

$$\frac{\max_{1 \le i \le n} (1/\sigma_i^2)}{\sum_{j=1}^n (1/\sigma_j^2)} \to 0\,.$$

(d) Compute $Var[T_n(w^{opt})]/Var[\overline{X}_n]$ in the case $\sigma_i^2 = A i^r$ for $r = .25,\ .50,\ .75,\ 1$ and $n = 5, 10, 20, 50, 100$, and ∞.

Solution:
(a) Unbiasedness of the estimators $T_n(w)$ is easy:

$$E(T_n(w)) = \sum_{i=1}^n w_{ni} E(X_i) = \sum_{i=1}^n w_{ni} \mu = \mu \sum_{i=1}^n w_{ni} = \mu\,.$$

By direct calculation,

$$Var(T_n(w)) \ = \ \sum_{i=1}^n w_{ni}^2 \sigma_i^2$$

and we want to minimize this subject to $\sum_{i=1}^n w_{ni} = 1$. Thus define

$$V^2(w, \lambda) = \sum_{i=1}^n w_{ni}^2 \sigma_i^2 + \lambda \left(\sum_{i=1}^n w_{ni} - 1 \right)\,.$$

Thus

$$\frac{\partial}{\partial w_{ni}} V^2(w, \lambda) = 2 w_{ni} \sigma_i^2 + \lambda, \qquad i = 1, \ldots, n\,,$$

and setting this equal to 0 for all i yields

$$w_{ni} = -\frac{\lambda/2}{\sigma_i^2}$$

where, in order to satisfy the constraint,

$$1 = \sum_{i=1}^n w_{ni} = -\frac{\lambda}{2}\sum_{i=1}^n \frac{1}{\sigma_i^2}$$

and hence

$$\lambda = \frac{-2}{\sum_{i=1}^n (1/\sigma_i^2)}\,.$$

Thus the optimal choice of the weights w_{ni} is given by

$$w_{ni}^{opt} = \frac{1/\sigma_i^2}{\sum_{j=1}^n (1/\sigma_j^2)}\,, \qquad i = 1,\ldots,n\,.$$

Here is a more geometric proof involving a "dual" optimization problem. Write $w_i \equiv w_{ni}$. We want to minimize

$$\sum_{i=1}^n w_i^2\sigma_i^2 = w^T \mathrm{diag}(\sigma_i^2) w$$

subject to the constraint $\sum_1^n w_i = w^T 1 = 1$. Alternatively, letting $v_i \equiv w_i\sigma_i$, we want to minimize

$$v^T v \qquad \text{subject to} \qquad \langle v, 1/\sigma\rangle = 1$$

where $1/\sigma = (1/\sigma_1,\ldots,1/\sigma_n)^T$. Geometrically this amounts to finding the radius of the smallest ball which intersects the hyperplane determined by the vector $1/\sigma$ and the magnitude constant 1. Now the dual problem is:

$$\text{maximize } \ \langle v, 1/\sigma\rangle \qquad \text{subject to} \qquad v^T v \le C^2\,.$$

Geometrically this amounts to finding the hyperplane with the largest magnitude constant which intersects the ball with radius C. But by the Cauchy-Schwarz inequality,

$$\langle v, 1/\sigma\rangle \le \sqrt{v^T v (1/\sigma)^T (1/\sigma)} \le C\sqrt{\sum_{i=1}^n (1/\sigma_i^2)} = 1 \tag{2}$$

if $C = 1/\sqrt{\sum_{i=1}^n (1/\sigma_i^2)}$, and equality holds in the first inequality of (2) if $v = A(1/\sigma)$ for some A. Translating back to w gives $w_i = w_{ni}$ as in the first solution.

(b) The resulting minimal variance is

$$Var[T_n(w^{opt})] = \sum_{i=1}^n [w_{ni}^{opt}]^2\sigma_i^2 = \frac{\sum_{i=1}^n (1/\sigma_i^2)}{(\sum_{i=1}^n (1/\sigma_i^2))^2} = \frac{1}{\sum_{i=1}^n (1/\sigma_i^2)}\,.$$

Hence by Chebychev's inequality

$$P(|T_n(w^{opt}) - \mu| > \epsilon) \le \frac{Var[T_n(w^{opt})]}{\epsilon^2} = \frac{1}{\epsilon^2\sum_{i=1}^n (1/\sigma_i^2)} \to 0$$

for every $\epsilon > 0$ if $\sum_1^n (1/\sigma_i^2) \to \infty$.
(c) When $X_i = \mu + \sigma_i \epsilon_i$, and we want to show that

$$\sqrt{\sum_1^n (1/\sigma_i^2)}\,\left(T_n(w^{opt}) - \mu\right) \to_d N(0,1)\,,$$

it is convenient (for clarity in avoiding clashes with the notation of the Lindeberg-Feller CLT) to temporarily relabel the σ_i as a_i, $i = 1, \ldots, n$. Note that then the quantity on the left side in the last display becomes

$$\sqrt{\sum_1^n (1/a_i^2)}\,\left(T_n(w^{opt}) - \mu\right) = \frac{\sum_{i=1}^n (1/a_i^2) a_i \epsilon_i}{\sqrt{\sum_{i=1}^n (1/a_i^2)}} = \sum_{i=1}^n c_{ni}\epsilon_i$$

where

$$c_{ni} = \frac{1/a_i}{\sqrt{\sum_{i=1}^n (1/a_i^2)}}\,, \qquad i = 1, \ldots, n\,.$$

By the same argument we have used several times now for weighted sums of i.i.d. random variables with mean zero and finite variance, the Lindeberg condition holds if

$$\max_{1\le i\le n} |c_{ni}| = \frac{\max_{1\le i\le n}(1/a_i)}{\sqrt{\sum_{i=1}^n (1/a_i^2)}} \to 0\,,$$

and we conclude that (3) holds.
(d) When $\sigma_i^2 = Ai^r$ for $i = 1, \ldots, n$ We have

$$Var(\overline{X}_n) = \frac{\sum_{i=1}^n \sigma_i^2}{n^2} = \frac{A\sum_{i=1}^n i^r}{n^2}\,,$$

while

$$Var[T_n(w^{opt})] = \frac{1}{\sum_{i=1}^n (1/\sigma_i^2)} = \frac{A}{\sum_{i=1}^n i^{-r}}\,.$$

Hence

$$\frac{Var[T_n(w^{opt})]}{Var[\overline{X}_n]} = \frac{n^2}{\left(\sum_{i=1}^n i^{-r}\right)\left(\sum_{i=1}^n i^r\right)}\,.$$

Computing this for $r = .25,\ .50,\ .75,\ 1$ and $n = 5, 10, 20, 50, 100$, and ∞, and noting that the above ratio equals

$$\begin{aligned}\frac{1}{\left(n^{-1}\sum_1^n (i/n)^{-r}\right)\left(n^{-1}\sum_1^n (i/n)^r\right)} \quad &\to \quad \frac{1}{\int_0^1 x^{-r}dx \int_0^1 x^r dx} \\ &= \quad (1-r)(1+r) = 1 - r^2\end{aligned}$$

(where the convergence holds for $0 < r \le 1$ even though the first integral in the denominator is infinite for $r = 1$), yields the following table.

Table 1: Efficiencies of $\overline{X}_n$ relative to $T_n(w^{opt})$

n	5	10	20	50	100	∞
$r = .25$	.980	.970	.961	.952	.948	.9375
$r = .50$	.923	.886	.854	.820	.801	.7500
$r = .75$	.836	.763	.700	.633	.595	.4375
$r = 1.0$	.730	.621	.529	.436	.382	0

2. Suppose that $X_1, \ldots, X_n$ are i.i.d. with density given by $f(x) = rx^{-(r+1)}1_{[1,\infty)(x)}$ for some $r > 1$.
(a) Compute the distribution function F of the X_i's.
(b) Compute and plot the inverse distribution function F^{-1} corresponding to F.
(c) For what values of $s > 0$ is $E|X_1|^s < \infty$?
(d) Find the distribution function of $M_n \equiv \max_{1\le i\le n} X_i$.
(e) For what values of s is $E|M_n|^s < \infty$?
(f) Find a sequence of constants b_n so that $M_n/b_n \to_d$ and find the limiting distribution. [Hint: see Ferguson, ACLST, Theorem 14, page 95.]

Solution: (a) $F(x) = 0$ for $-\infty < x < 1$, and $F(x) = 1 - x^{-r}$ for $1 \le x < \infty$.
(b) Setting $F(x) = u$ and solving for $x = F^{-1}(u)$ yields $F^{-1}(u) = (1-u)^{-1/r}$ for $0 < u < 1$. Note that $F^{-1}(1/2) = 2^{1/r}$; $F^{-1}(1) = \infty$, and $F^{-1}(0) = 1$. See Figure **??** for plots of F^{-1} for $r \in \{1, 5, 20\}$.
(c) We compute

$$\begin{aligned} E|X_1|^s &= \int_1^\infty x^s f(x)dx = \int_1^\infty x^s \cdot rx^{-r-1}dx \\ &= r\int_1^\infty x^{s-r-1}dx = \frac{r}{r-s} < \infty \ \text{ if } s < r. \end{aligned}$$

Alternatively,

$$\begin{aligned} EX_1^s &= \int_0^\infty sx^{s-1}(1-F(x))dx = \int_0^1 sx^{s-1}dx + \int_1^\infty sx^{s-1}x^{-r}dx \\ &= 1 + \frac{s}{r-s} = \frac{r}{r-s} < \infty \ \text{ if } \ s < r. \end{aligned}$$

(d) For $M_n = \max_{1\le i\le n} X_i$ we have

$$\begin{aligned} F_{M_n}(x) &= P(\max_{1\le i\le n} X_i \le x) = P(X_1 \le x, \ldots, X_n \le x) \\ &= F(x)^n = (1-x^{-r})^n \ \text{ if } \ 1 \le x < \infty, \end{aligned}$$

and 0 otherwise.
(e) Since $M_n^s = \max_{1\le i\le n} X_i^s \le \sum_{i=1}^n X_i^s$, it follows that $EM_n^s \le \sum_{i=1}^n E(X_i^s) = nE(X_1^s) < \infty$ if $s < r$ by (c). Alternatively,

$$\begin{aligned} P(M_n > x) &= P(\max_{1\le i\le n} X_i > x) = P(\cup_{i=1}^n [X_i > x]) \\ &\le \sum_{i=1}^n P(X_i > x) = n(1-F(x)) = nx^{-r}, \quad \text{for} \ \ 1 \le x < \infty. \end{aligned}$$

Thus

$$\begin{aligned}EM_n^s &= \int_0^\infty P(M_n^s > x)dx \\ &= \int_0^1 dx + \int_1^\infty P(M_n > x^{1/s})dx \\ &\le 1 + \int_1^\infty n x^{-r/s} dx = 1 + n\frac{1}{r/s - 1} < \infty \ \text{ if } \ s < r.\end{aligned}$$

(f) Now from (d), for any $x > 0$,

$$\begin{aligned}P(M_n/b_n \le x) &= P(M_n \le x b_n) = (1 - b_n^{-r} x^{-r})^n \\ &= \left(1 - \frac{x^{-r}}{n}\right)^n \quad \text{if } \ b_n^{-r} = 1/n, \ \text{ or } \ b_n = n^{1/r}, \\ &\to \exp(-x^{-r}) \equiv G_r(x) \ \text{ as } \ n \to \infty.\end{aligned}$$

Thus $M_n/b_n \to_d G_r$. In fact, G_r is a member of the Weibull family with shape parameter $-r$, and is one of the three different families that can arise as limit distributions of maxima of independent rv's; see e.g. Ferguson (1996), *A Course in Large Sample Theory*, chapter 14.

3. Suppose that $X_1, \ldots, X_n$ are i.i.d. random vectors with values in R^k with $E(X_1) = \mu$ and $E(X_1^T X_1) < \infty$ so that $\Sigma = E(X_1 - \mu)(X_1 - \mu)^T$ is well-defined. Thus

$$Z_n \equiv \sqrt{n}(\overline{X}_n - \mu) \to_d Z \sim N_k(0, \Sigma)\,.$$

Suppose that $g : R^k \to R$ is a function, and suppose that $\nabla g = (g')^T$ exists at μ. Then the delta-method (or g' theorem) tells us that

$$\text{(3)} \qquad \sqrt{n}(g(\overline{X}_n) - g(\mu)) \to_d \nabla g(\mu)^T Z \sim N(0, \nabla g(\mu)^T \Sigma \nabla g(\mu))\,.$$

(a) Show that we can strengthen (3) as follows: Suppose that $\nabla g = (g')^T$ is continuous at μ. Then $\sqrt{n}(g(\overline{X}_n) - g(\mu))$ is *asymptotically linear* at μ:

$$\begin{aligned}\sqrt{n}(g(\overline{X}_n) - g(\mu)) &= \nabla g(\mu)^T \sqrt{n}(\overline{X}_n - \mu) + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(X_i) + o_p(1)\end{aligned}$$

where

$$\text{(4)} \qquad \psi(x) = \nabla g(\mu)^T (x - \mu)$$

which is called the *influence function* of $g(\overline{X}_n)$ as an estimator of $g(\mu)$, has mean $E\psi(X_i) = 0$ and $Var(\psi(X_i)) = \nabla g(\mu)^T \Sigma \nabla g(\mu)$.
(b) Does the result of (a) apply to the situation considered in problem 3(b) of problem set #4? If so, what is the resulting influence function?
(c) Does the result of (a) apply to Example 2.3.6 (the correlation coefficient) on pages 18-19 of the course notes?

Solution: (a) By Taylor's theorem, for some Y_n^* satisfying $|Y_n^* - \mu| \le |\overline{X}_n - \mu| \to_p 0$ it follows that

$$\begin{aligned}\sqrt{n}(g(\overline{X}_n) - g(\mu)) &= \nabla g(Y_n^*)\sqrt{n}(\overline{X}_n - \mu) \\ &= \nabla g(\mu)\sqrt{n}(\overline{X}_n - \mu) \\ &\qquad + \{\nabla g(Y_n^*) - \nabla g(\mu)\}\sqrt{n}(\overline{X}_n - \mu) \\ &= \nabla g(\mu)\sqrt{n}(\overline{X}_n - \mu) + o_p(1)\end{aligned}$$

since $\nabla g(Y_n^*) \to_p \nabla g(\mu)$ by continuity of ∇g at μ and since $\sqrt{n}(\overline{X}_n - \mu) = O_p(1)$. Now note that

$$\nabla g(\mu)\sqrt{n}(\overline{X} - \mu) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \nabla g(\mu)(X_i - \mu) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi(X_i)$$

with ψ as in (4).

In fact, the hypothesis of continuity of ∇g can be dropped: consider a new function $h(x) = g(x) - \nabla g(\mu)x$. Then $\nabla h(\mu) = \nabla g(\mu) - \nabla g(\mu) = 0$, and we can write

$$\begin{aligned}\sqrt{n}(g(\overline{X}_n) - g(\mu) - \nabla g(\mu)(\overline{X}_n - \mu)) &= \sqrt{n}(h(\overline{X}_n) - h(\mu)) \\ &\to_d \nabla h(\mu)Z = 0 \cdot Z = 0\end{aligned} \tag{5}$$

by the delta-method applied to the function h. Since convergence in distribution to a constant implies convergence in probability to the same constant, we conclude from (5) that the left side of (5) converges in probability to 0. But this is just the claimed asymptotic linearity with $\psi(x) = \nabla g(\mu)(x - \mu)$.

(b) The result in (a) does not quite apply since

$$Z_n \equiv \sqrt{n}(\overline{X}_n - \mu, S_n^2 - \sigma^2)'$$

is not exactly an average of i.i.d. random vectors. But the key features of the proof in (a) do carry through since $n^{-1/2}Z_n \to_p 0$ and $Z_n = n^{-1/2}\sum_{i=1}^n \underline{Y}_i + o_p(1)$ where $\underline{Y}_i = (X_i - \mu, (X_i - \mu)^2 - \sigma^2)'$ are i.i.d. with mean 0 and finite second moment under the assumptions of problem 3(b) of problem set #4. Thus the conclusion continues to hold. Here is an argument: Thus with $g(u, v) = v/u$

$$\begin{aligned}\sqrt{n}\left(\frac{S_n^2}{\overline{X}_n} - \frac{\sigma^2}{\mu}\right) &= \nabla g(\mu_n^*, \sigma_n^{*2})\sqrt{n}(\overline{X}_n - \mu, S_n^2 - \sigma^2) \\ &\qquad \text{where } \|(\mu_n^* - \mu, \sigma_n^{*2} - \sigma^2)\| \le \|(\overline{X}_n - \mu, S_n^2 - \sigma^2)\| \to_p 0, \\ &= \nabla g(\mu_n^*, \sigma_n^{*2})\left(\sqrt{n}\underline{Y}_n + o_p(1)\right) \\ &= \nabla g(\mu, \sigma^2)\sqrt{n}\underline{Y}_n + \left(\nabla g(\mu_n^*, \sigma_n^{*2}) - \nabla g(\mu, \sigma^2)\right)\sqrt{n}\underline{Y}_n \\ &\qquad + \nabla g(\mu_n^*, \sigma_n^{*2})o_p(1) \\ &= \frac{1}{\mu}(-\sigma^2/\mu, 1)\sqrt{n}\underline{Y}_n + o_p(1)O_p(1) + o_p(1)o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{1}{\mu}\left\{(X_i - \mu)^2 - \sigma^2 - \frac{\sigma^2}{\mu}(X_i - \mu)\right\} + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi(X_i) + o_p(1)\end{aligned}$$

where

$$\begin{aligned}\psi(x) &= \frac{1}{\mu}\{(x-\mu)^2 - \sigma^2 - (\sigma^2/\mu)(x-\mu)\} \\ &= \frac{\sigma^2}{\mu}\left\{\left(\frac{x-\mu}{\sigma}\right)^2 - 1 - \frac{\sigma}{\mu}\left(\frac{x-\mu}{\sigma}\right)\right\}\end{aligned}$$

has $E\psi(X_i) = 0$ and

$$Var(\psi(X_i)) = \frac{\sigma^4}{\mu^2}\left\{2 + \gamma_2 - 2\frac{\sigma\gamma_1}{\mu} + \frac{\sigma^2}{\mu^2}\right\} = V^2$$

as in the solution of problem 3(b), Problem Set 4.
(c) As in (b), (a) does not apply directly, but it does apply once we take care of smaller order terms. In this case $r_n = S_{XY}/\sqrt{S_{XX}S_{YY}} = g(S_{XY}, S_{XX}, S_{YY})$ where $g(u,v,w) = u/\sqrt{vw}$ is differentiable at $(\rho, 1, 1)$ with derivative $\nabla g(\rho, 1, 1) = (1, -\rho/2, -\rho/2)$ and where (S_{XY}, S_{XX}, S_{YY}) is the vector of (average) cross-products and squares of the X's and Y's as in the notes. As shown in the notes,

$$\begin{aligned}\sqrt{n}\begin{pmatrix} S_{XY} - \rho \\ S_{XX} - 1 \\ S_{YY} - 1\end{pmatrix} &= \sqrt{n}\begin{pmatrix} \overline{XY} - \rho \\ \overline{X^2} - 1 \\ \overline{Y^2} - 1\end{pmatrix} + o_p(1) \\ &\equiv \underline{Z}_n + o_p(1)\end{aligned}$$

where $\underline{Z}_n \to_d \underline{Z} \sim N_3(0, \Sigma)$ and Σ is as on page 19 of the chapter 2 notes. Thus it follows by arguing as in (b) above, that

$$\begin{aligned}\sqrt{n}(r_n - \rho) &= \sqrt{n}\left(g(S_{XY}, S_{XX}, S_{YY}) - g(\rho, 1, 1)\right) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \left\{(\tilde{X}_i\tilde{Y}_i - \rho) - \frac{\rho}{2}(\tilde{X}_i^2 - 1) - \frac{\rho}{2}(\tilde{Y}_i^2 - 1)\right\} + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi(X_i, Y_i) + o_p(1)\end{aligned}$$

in terms of the "standardized" data $(\tilde{X}_i\tilde{Y}_i)) = ((X_i - \mu_X)/\sigma_X, (Y_i - \mu_Y)/\sigma_Y)$. Thus

$$\psi(x,y) = \frac{(x-\mu_X)}{\sigma_X}\cdot\frac{(y-\mu_Y)}{\sigma_Y} - \frac{\rho}{2}\frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{\rho}{2}\frac{(y-\mu_Y)^2}{\sigma_Y^2}.$$

4. (a) Write out a proof of (17) on page 27 of the Chapter 2 notes. Compare the result with what you get by combining (11) and (16).
(b) Write out a proof of the corresponding fact (38) concerning the general empirical process $\mathbb{G}_n$: $\mathbb{G}_n \to_{f.d.} \mathbb{G}$ where $\mathbb{G}_n$ and $\mathbb{G}$ are as defined on page 21 of the chapter 2 notes; i.e. for any $f_1, \ldots, f_k \in L_2(P)$, $(\mathbb{G}_n(f_1), \ldots, \mathbb{G}_n(f_k)) \to_d (\mathbb{G}(f_1), \ldots, \mathbb{G}(f_k))$.

Solution: (a) $\sqrt{n}(\mathbb{F}_n - F) \to_{f.d.} \mathbb{U}(F)$. To see this, let $x_1 < \cdots < x_k \in \mathbb{R}$. Then the vector $(\sqrt{n}(\mathbb{F}_n(x_1) - F(x_1)), \ldots, \sqrt{n}(\mathbb{F}_n(x_k) - F(x_k)))^T$ can be written as

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \begin{pmatrix} 1_{(-\infty, x_1]}(X_i) - F(x_1) \\ \vdots \\ 1_{(-\infty, x_k]}(X_i) - F(x_k)\end{pmatrix} \equiv n^{-1/2}\sum_{i=1}^n \underline{Y}_i$$

where the $\underline{Y}_i$'s are i.i.d. with $EY_{1,j} = F(x_j) - F(x_j) = 0$ for $j = 1, \ldots, k$ and

$$\Sigma \equiv E\underline{Y}_1\underline{Y}_1^T = (F(x_j \wedge F(x_{j'}) - F(x_j)F(x_{j'}))_{j,j'=1}^k .$$

It follows from the multivariate CLT that $n^{-1/2}\sum_{i=1}^k \underline{Y}_i \to_d N_k(0,\Sigma)$. But this is just the distribution of $(\mathbb{U}(F(x_1)), \ldots, \mathbb{U}(F(x_k)))^T$. Thus $\sqrt{n}(\mathbb{F}_n - F) \to_{f.d.} \mathbb{U}(F)$ holds.

To see that this agrees with the result of (11) and (16) note that $\sqrt{n}(\mathbb{F}_n - F) \stackrel{d}{=} \mathbb{U}_n(F)$ where $\mathbb{U}_n \to_{f.d.} \mathbb{U}$ by the argument detailed below. Thus $\sqrt{n}(\mathbb{F}_n - F) \to_{f.d.} \mathbb{U}(F)$.

Details for $\mathbb{U}_n \to_{f.d.} \mathbb{U}$: To see this, let $0 < t_1 < t_2 < \cdots < t_k < 1$. Then define random vectors $\underline{Y}_i$ by

$$\underline{Y}_i = (1_{[0,t_1]}(\xi_i) - t_1, \ldots, 1_{[0,t_k]}(\xi_i) - t_k),$$

for $i = 1, \ldots, n$. Note that $E\underline{Y}_1 = 0$ and

$$\begin{aligned} E\underline{Y}_1\underline{Y}_1' &= \begin{pmatrix} t_1(1-t_1) & t_1 - t_1t_2 & \cdots & t_1 - t_1t_k \\ t_1 - t_1t_2 & t_2(1-t_2) & \cdots & t_2 - t_2t_k \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ t_1 - t_1t_k & t_2 - t_2t_k & \cdots & t_k(1-t_k) \end{pmatrix} \\ &= (t_i \wedge t_j - t_it_j)_{i,j=1}^k \equiv \Sigma. \end{aligned}$$

Thus it follows from the multivariate central limit theorem that

$$(\mathbb{U}_n(t_1), \ldots, \mathbb{U}_n(t_k))' = \sqrt{n}\underline{Y}_n \to_d N_k(0,\Sigma).$$

But for a Brownian bridge process $\mathbb{U}$, $(\mathbb{U}(t_1), \ldots, \mathbb{U}(t_k))' \sim N_k(0,\Sigma)$, so we have shown that $(\mathbb{U}_n(t_1), \ldots, \mathbb{U}_n(t_k))' \to_d (\mathbb{U}(t_1), \ldots, \mathbb{U}(t_k))'$. But since this holds for every k and every choice of $t_1, \ldots, t_k$, it follows that $\mathbb{U}_n \to_{f.d.} \mathbb{U}$.

(b) $\mathbb{G}_n \to_{f.d.} \mathbb{G}$. To see this, let $f_1, \ldots, f_k \in L_2(P)$. Then define random vectors $\underline{Y}_i$ by

$$\underline{Y}_i = (f_1(X_i) - Pf_1, \ldots, f_k(X_i) - Pf_k)$$

for $i = 1, \ldots, n$. Note that $E\underline{Y}_i = 0$ and

$$\begin{aligned} E\underline{Y}_1\underline{Y}_1' &= \begin{pmatrix} P(f_1^2) - (P(f_1)^2 & P(f_1f_2) - Pf_1Pf_2 & \cdots & P(f_1f_k) - Pf_1Pf_k \\ P(f_1f_2) - Pf_1Pf_2 & P(f_2^2) - (Pf_2)^2 & \cdots & P(f_2f_k) - Pf_2Pf_k \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ P(f_1f_k) - Pf_1Pf_k & P(f_2f_k) - Pf_2Pf_k & \cdots & P(f_k^2) - (Pf_k)^2 \end{pmatrix} \\ &= (P(f_if_j) - Pf_iPf_j)_{i,j=1}^k \equiv \Sigma. \end{aligned}$$

Thus it follows from the multivariate central limit theorem that

$$(\mathbb{G}_n(f_1), \ldots, \mathbb{G}_n(f_k))' = \sqrt{n}\underline{Y}_n \to_d N_k(0,\Sigma).$$

But for a $P-$Brownian bridge process $\mathbb{G}_P$, $(\mathbb{G}(f_1), \ldots, \mathbb{G}(f_k))' \sim N_k(0,\Sigma)$, so we have shown that $(\mathbb{G}_n(f_1), \ldots, \mathbb{G}_n(f_k))' \to_d (\mathbb{G}(f_1), \ldots, \mathbb{G}(f_k))'$. But since this holds for every k and every choice of $f_1, \ldots, f_k \in L_2(P)$, it follows that $\mathbb{G}_n \to_{f.d.} \mathbb{G}$.

5. Suppose that $X_1, \ldots, X_n$ are i.i.d. exponential(θ); i.e. with density $p_\theta(x) = \theta \exp(-\theta x) 1_{[0,\infty)}(x)$. Let $X_{(n)} = X_{n:n}$ be the largest order statistic of $X_1, \ldots, X_n$.
(a) Find constants c_n so that $Y_n = X_{(n)} - c_n \to_d Y$ for some random variable Y and find the limiting distribution of F_Y.
(b) Compute the density of Y_n and show that it converges to the density f_Y of Y.
(c) What can you conclude from the result of (b) and Scheffé's theorem (chap. 2 notes, prop. 1.14, page 9)?

Solution: Let $c_n = \theta^{-1} \log n = F_\theta^{-1}(1 - 1/n)$. Then

$$\begin{aligned}
F_n(y) = P(Y_n \le y) &= P(X_{(n)} - c_n \le y) = P(X_{(n)} \le y + c_n) \\
&= P(X_j \le y + c_n \text{ for all } 1 \le j \le n) \\
&= P(X_1 \le y + c_n)^n = (1 - \exp(-\theta(y + c_n)))^n \\
&= \left(1 - \frac{e^{-\theta y}}{n}\right)^n \to \exp(-e^{-\theta y}) \\
&\equiv F_Y(y);
\end{aligned}$$

this is an extreme - value distribution of the the "double-exponential" or "Gumbel" type; see part (c) of Theorem 14, Ferguson, ACILST page 95.
(b) The density of Y_n is found easily by differentiating in the previous display. The result is that

$$f_n(y) = (1 - n^{-1} \exp(-\theta y))^{n-1} \exp(-\theta y) \to \exp(-e^{-\theta y}) \exp(-\theta y) = f_Y(y) \equiv f(y).$$

(c) Since the densities f_n converge pointwise to the limiting density, we conclude by Scheffé's theorem that with P_n begin the probability measure on $\mathbb{R}$ corresponding to F_n and P the corresponding probability measure on $\mathbb{R}$ corresponding to F_Y,

$$d_{TV}(P_n, P) = \frac{1}{2} \int |f_n(y) - f(y)| dy \to 0.$$