Statistics 581, Problem Set 1 Solutions
Wellner; 10/6/2017

1. Ferguson, ACILST, #2, page 6: Suppose that X_n is uniformly distributed on the set of points $\{1/n, 2/n, \ldots, 1\}$. Show that $X_n \rightarrow_d X$ where the distribution of X is Uniform$(0,1)$. Does $X_n \rightarrow_p X$?

   Solution: If $X_n \sim$ Uniformly on $\{1/n, 2/n, \ldots, 1\}$ then

$$F_n(t) \quad = \quad P(X_n \le t) = \frac{\#\text{of } j/n \le t}{n} = \frac{\lfloor nt \rfloor}{n}.$$

   Note that $|F_n(t) - t| \le 1/n$ for each fixed $0 \le t \le 1$. Then $F_n(t) \rightarrow F(t) = t$ for all $0 \le t \le 1$. That is, $X_n \rightarrow_d X \sim$ Uniform$(0,1)$. These X_n's do not necessarily converge in probability to X because the random variables X_n are not necessarily defined on the same probability space.

2. (Continuation of the previous problem). Now suppose that $U \sim$ Uniform$(0,1)$ and for each $n \ge 1$ define $V_n \equiv \sum_{j=1}^n (j/n) 1_{((j-1)/n, j/n]}(U)$.
   (a) Show that $V_n \stackrel{d}{=} X_n$ where X_n is as in problem 2.
   (b) Show that $V_n \rightarrow_p U$.

   Solution: (a) Note that $P(V_n = j/n) = P(U \in ((j-1)/n, j/n]) = 1/n = P(X_n = j/n)$. Thus $V_n \stackrel{d}{=} X_n$.
   (b) To see that $V_n \rightarrow_p U$, note that

$$\begin{aligned}
P(|V_n - U| > \epsilon) &= P(V_n - U > \epsilon) \qquad \text{since } V_n \ge U \\
&= P(\cup_{j=1}^n \{V_n - U > \epsilon\} \cap \{(j-1)/n < U \le j/n\}) \\
&= P(\cup_{j=1}^n \{j/n - U > \epsilon\} \cap \{(j-1)/n < U \le j/n\}) \\
&= nP(\{1/n - U > \epsilon\} \cap \{0 < U < 1/n\}) \\
&= \begin{cases} n(1/n - \epsilon) = 1 - n\epsilon\,, & \text{if } 1/n > \epsilon \\ 0\,, & \text{if } 1/n \le \epsilon\,. \end{cases}
\end{aligned}$$

   Hence it follows that $V_n \rightarrow_p U$.

3. Suppose that X has an inverse power distribution with distribution function $F_\alpha(x) = 1 - x^{-\alpha}$ for $x \in [1, \infty)$ where $\alpha > 0$.
   (a) Show that $E(X) = \alpha/(\alpha - 1)$ for $\alpha > 1$ and
   $Var(X) = \alpha/((\alpha-2)(\alpha-1)^2)$ for $\alpha > 2$.

(b) Let $Y_\alpha = (\alpha-1)X - \alpha$ (so that Y_α has mean 0 and variance converging to 1 as $\alpha \to \infty$). Show that $Y_\alpha \to_d Y_\infty$ for some random variable Y_∞ and find the distribution function of Y_∞.

Solution: (a) By straightforward computation, $f_\alpha(x) = \alpha x^{-\alpha-1} 1_{[1,\infty)}(x)$ and hence for $\alpha > r$ we have

$$
\begin{aligned}
E(X^r) &= \int_1^\infty x^r f_\alpha(x) dx = \alpha \int_1^\infty x^{-(\alpha+1-r)} dx \\
&= \frac{\alpha}{\alpha - r}.
\end{aligned}
$$

This yields $E(X) = \alpha/(\alpha-1)$ when $r=1$ and $E(X^2) = \alpha/(\alpha-2)$ when $r = 2$. Thus

$$
Var(X) = E(X^2) - (E(X))^2 = \alpha \left\{ \frac{1}{\alpha-2} - \frac{\alpha}{(\alpha-1)^2} \right\} = \frac{\alpha}{(\alpha-2)(\alpha-1)^2}.
$$

(b) It is easily seen that $Y_\alpha \equiv (\alpha-1)X - \alpha$ has $E(Y_\alpha) = 0$ and $Var(Y_\alpha) = \alpha/(\alpha-2) \to 1$ as $\alpha \to \infty$. Furthermore, for any fixed $y \ge -1$,

$$
\begin{aligned}
P(Y_\alpha \le y) &= P((\alpha-1)X - \alpha \le y) = P(X \le (y+\alpha)/(\alpha-1)) \\
&= 1 - \left(\frac{\alpha+y}{\alpha-1} \right)^{-\alpha} = 1 - \left(1 + \frac{y+1}{\alpha-1} \right)^{-\alpha} \\
&\to 1 - \exp(-(y+1)), \quad \text{as} \quad \alpha \to \infty.
\end{aligned}
$$

Thus $Y_\alpha \to_d Y_\infty \equiv Z - 1$ where $Z \sim$ exponential(1).

4. (Lehmann and Casella, TPE, problems 1.2 and 1.3, page 62.)
(a) Let $X_1, \ldots, X_n$ be uncorrelated random variables with common expectation θ and variance σ^2. Show that, among all linear estimators $\sum \alpha_i X_i$ of θ satisfying $\sum \alpha_i = 1$, the mean $\overline{X}_n$ has the smallest variance.
(b) In the preceding problem, minimize the variance of $\sum \alpha_i X_i$ ($\sum \alpha_i = 1$)
(i) When the variance of X_i is σ^2/c_i (c_i known).
(ii) When the X_i have common variance σ^2 but are correlated with common correlation coefficient ρ.

Solution: (i) First solution – via the Cauchy-Schwarz inequality: First recall the Cauchy-Schwarz inequality in R^n: if $u, v \in R^n$, then $(u'v)^2 \leq (u'u)(v'v)$ with equality iff $u = cv$ for some real number c. Now extend this as follows: if Σ is positive definite and $x, y \in R^n$, then

$$(x'y)^2 = (\Sigma^{1/2}x)'(\Sigma^{-1/2}y) \leq (x'\Sigma x)(y'\Sigma^{-1}y)$$

with equality iff $\Sigma^{1/2}x = c\Sigma^{-1/2}y$; i.e. iff $x = c\Sigma^{-1}y$.

Now consider X, a random vector in R^n, with $E(X) = \mathbf{1}\theta$ and $Cov(X, X) = E[(X - E(X))(X - E(X))'] = \Sigma$, where $\mathbf{1} = (1, \ldots, 1)' \in R^n$. A linear estimator $\alpha' X = \alpha_1 X_1 + \ldots + \alpha_n X_n$ is unbiased for θ iff $\theta = E(\alpha' X) = \alpha' E(X) = (\alpha'\mathbf{1})\theta$ for all θ; i.e., iff $\alpha'\mathbf{1} = 1$. The variance of $\alpha' X$ is $Var(\alpha' X) = \alpha'\Sigma\alpha$. To find the best such estimator, we must find

$$\min\{\alpha'\Sigma\alpha : \alpha'\mathbf{1} = 1\}.$$

But by the Cauchy-Schwarz inequality, if $\alpha'\mathbf{1} = 1$, then

$$\alpha'\Sigma\alpha \geq 1/(\mathbf{1}'\Sigma^{-1}\mathbf{1})$$

with equality iff $\alpha = c\Sigma^{-1}\mathbf{1}$. The condition $\alpha'\mathbf{1} = \mathbf{1}$ then implies that $c = 1/(\mathbf{1}'\Sigma^{-1}\mathbf{1})$, so the optimal α is $\alpha_0 \equiv \Sigma^{-1}\mathbf{1}/(\mathbf{1}'\Sigma^{-1}\mathbf{1})$, and the optimal linear unbiased estmator is $\alpha_0' X = (\mathbf{1}'\Sigma^{-1}X)/(\mathbf{1}'\Sigma^{-1}\mathbf{1})$ whose variance is $Var(\alpha_0' X) = 1/(\mathbf{1}'\Sigma^{-1}\mathbf{1})$.
The solutions to 1.2, 1.3(a), and 1.3(b) now follow:
(a) In this case $\Sigma = \sigma^2 I$, so $\alpha_0 = \mathbf{1}(1/\sigma^2)/(\mathbf{1}'I\mathbf{1}/\sigma^2) = \mathbf{1}(1/n)$.
(b)(i): The inverse of the matrix $\mathrm{diag}(1/c_i)$ is just $\mathrm{diag}(c_i)$. This implies that $\alpha_0' X = (\sum_1^n c_i X_i)/(\sum_1^n c_i)$ and $Var(\alpha_0' X) = \sigma^2 / \sum c_i$.
(b)(ii): The inverse of the matrix with 1 on the diagonal and ρ off the diagonal is of the form a in the diagonal entries and b in the off-diagonal entries for some a, b. Hence $\Sigma^{-1}\mathbf{1} = \sigma^{-2}(a + (n-1)b)\mathbf{1}$, which leads to $\mathbf{1}'\Sigma^{-1}X = \sigma^{-2}(a + (n-1)b)(X_1 + \ldots + X_n)$, and $\mathbf{1}'\Sigma^{-1}\mathbf{1} = \sigma^{-2}(a + (n-1)b)n$. Hence we find that $\alpha_0' X = \sum_1^n X_i/n$. But $\Sigma\mathbf{1} = \sigma^2(1 + (n-1)\rho)\mathbf{1}$, so $\mathbf{1} = \sigma^2(1 + (n-1)\rho)\Sigma^{-1}\mathbf{1} = (1 + (n-1)\rho)(a + (n-1)b)\mathbf{1}$, and hence $[a + (n-1)b] = [1 + (n-1)\rho]^{-1}$. Therefore

$$Var(\alpha_0' X) = \frac{\sigma^2}{n}[1 + (n-1)\rho] \begin{cases} > \sigma^2/n & \text{if } \rho > 0 \\ < \sigma^2/n & \text{if } -1/(n-1) \leq \rho < 0 \end{cases}.$$

[Note that if $\rho < -1/(n-1)$, the matrix Σ of this form *is not a covariance matrix*!]

(ii) Second solution via Lagrange multipliers:
(a) When the X_i's are uncorrelated with common variance σ^2, the variance of $\alpha' X$ is just $\sigma^2 \sum_1^n \alpha_i^2$, so we want to minimize $\sum_1^n \alpha_i^2$ subject to $\sum_1^n \alpha_i = 1$. Thus we set

$$S(\alpha, \lambda) = \sum_{i=1}^{n} \alpha_i^2 + \lambda(\sum_{1}^{n} \alpha_i - 1), \quad \alpha \in \mathbb{R}^n, \quad \lambda \in \mathbb{R}.$$

To find the stationary points we look for solutions to

$$\begin{aligned} \nabla_\alpha S(\alpha, \lambda) &= 2\alpha + \lambda \mathbf{1} = 0, \\ \frac{\partial}{\partial \lambda} S(\alpha, \lambda) &= \alpha \mathbf{1} - 1 = 0. \end{aligned}$$

Solving the first for α yields $\alpha = -(\lambda/2)\mathbf{1}$. Substitution of this in the second equation gives $1 = \alpha' \mathbf{1} = -n\lambda/2$. This yields $\lambda = -2/n$ and hence $\alpha = n^{-1}\mathbf{1} \equiv \alpha_0$ and hence $Var(\alpha_0 X) = \sigma^2/n$, in agreement with the first argument above via Cauchy-Schwarz.
(b)(i) When $Var(X_i) = \sigma^2/c_i$ for $1 \le i \le n$, $Var(\alpha' X) = \sigma^2 \sum_1^n \alpha_i^2/c_i$, and now we want to minimize $\sum_1^n \alpha_i^2/c_i$ subject to $\mathbf{1}'\alpha = 1$. Thus we let

$$S(\alpha, \lambda) = \sum_{i=1}^{n} c_i^{-1} \alpha_i^2 + \lambda(\sum_{1}^{n} \alpha_i - 1),$$

and to find the stationary points we look for solutions to

$$\begin{aligned} \nabla_\alpha S(\alpha, \lambda) &= 2\mathrm{diag}(c^{-1})\alpha + \lambda \mathbf{1} = 0, \\ \frac{\partial}{\partial \lambda} S(\alpha, \lambda) &= \alpha \mathbf{1} - 1 = 0. \end{aligned}$$

Solving the first of these yields $\alpha = -(\lambda/2)\mathrm{diag}(c)\mathbf{1}$, and plugging this into the second equation yields $\lambda = -2/(\mathbf{1}'\mathrm{diag}(c)\mathbf{1}) = -2/\sum_{i=1}^n c_i$. Thus α_0 is given by $\alpha_{0,i} = c_i/\sum_{j=1}^n c_j$ and the minimal variance is just $\sigma^2/\sum_{i=1}^n c_i$.

b(ii) In this case $Var(\alpha' X) = (\alpha'\Sigma_0\alpha)\,\sigma^2$ where Σ_0 is the matrix with 1's on the diagonal and ρ everywhere off the diagonal. Thus we want to minimize $\alpha'\Sigma_0\alpha$ subject to $\alpha'\mathbf{1} = 1$, and we define

$$S(\alpha, \lambda) = \alpha'\Sigma_0\alpha + \lambda(\mathbf{1}'\alpha - 1).$$

Then we look for solutions to

$$\begin{aligned} \nabla_\alpha S(\alpha, \lambda) &= 2\Sigma_0\alpha + \lambda\mathbf{1} = 0, \\ \frac{\partial}{\partial\lambda} S(\alpha, \lambda) &= \alpha\mathbf{1} - 1 = 0. \end{aligned}$$

Solving the first of these for α yields $\alpha_0 = -(\lambda/2)\Sigma_0^{-1}\mathbf{1}$. Substitution of this in the second equation yields $\lambda = -2/(\mathbf{1}'\Sigma_0^{-1}\mathbf{1})$, and it follows that $\alpha_0 = \Sigma_0^{-1}\mathbf{1}/(\mathbf{1}^T\Sigma_0^{-1}\mathbf{1})$ in agreement with the first solution above via Cauchy-Schwarz.

5. (a) First suppose that X has distribution function F on $\mathbb{R}^+ = [0, \infty)$ and Y with distribution function G on $\mathbb{R}^+$ are independent random variables. Find the joint distribution of $Z = \min\{X, Y\}$ and $W = 1\{X \le Y\}$. That is, find $P(Z \le z, W = 1)$ and $P(Z \le z, W = 0)$ in terms of F and G.
(b) Show that if $F = \text{exponential}(\lambda)$ and $G = \text{exponential}(\mu)$, then Z and W are independent.

Solution: (a) Rewriting by using indicator functions and then conditioning yields

$$\begin{aligned} P(Z \le z, W = 1) &= P(X \le z, X \le Y) = E\{1_{[X\le z]}1_{[X\le Y]}\} \\ &= E[E\{1_{[X\le z]}1_{[X\le Y]}|X\}] = E[1_{[X\le z]}E\{1_{[X\le Y]}|X\}] \\ &= E[1_{[X\le z]}(1 - G(X-))] = \int_{[0,z]} (1 - G(x-))dF(x). \end{aligned}$$

Similarly,

$$\begin{aligned} P(Z \le z, W = 0) &= P(Y \le z, X > Y) = E\{1_{[Y\le z]}1_{[X>Y]}\} \\ &= E[E\{1_{[Y\le z]}1_{[X>Y]}|Y\}] = E[1_{[Y\le z]}E\{1_{[X>Y]}|Y\}] \\ &= E[1_{[Y\le z]}(1 - F(Y)] = \int_{[0,z]} (1 - F(y))dG(y). \end{aligned}$$

Also note that by independence of X and Y

$$\begin{aligned} 1 - H(z) &\equiv P(Z > z) = P(X > z, Y > z) \\ &= P(X > z)P(Y > z) = (1 - F(z))(1 - G(z)), \quad (1) \end{aligned}$$

and on the other hand, by an application of the first identity in Proposition 1.4.1 (course notes, page 17),

$$\begin{aligned} P(Z \le z) &= P(Z \le z, W = 1) + P(Z \le z, W = 0) \\ &= \int_{[0,z]} (1 - G(x-))dF(x) + \int_{[0,z]} (1 - F(y))dG(y) \\ &= (1 - G(z))F(z) - \int_{[0,z]} FdG + \int_{[0,z]} (1 - F(y))dG(y) \\ &= F(z) + G(z) - F(z)G(z) \\ &= 1 - (1 - F(z))(1 - G(z)) = 1 - (1 - H(z)) \end{aligned}$$

in agreement with the result in (1).
(b) When $X \sim \exp(\lambda)$ and $Y \sim \exp(\mu)$, $1 - G(x-) = \exp(-\mu x)$ and $1 - F(y) = \exp(-\lambda y)$ so

$$\begin{aligned} P(Z \le z, W = 1) &= \int_0^z e^{-\mu x} \lambda e^{-\lambda x} dx \\ &= \frac{\lambda}{\lambda + \mu} \int_0^z (\lambda + \mu) e^{-(\lambda+\mu)x} dx \\ &= \frac{\lambda}{\lambda + \mu} (1 - e^{-(\lambda+\mu)z}). \end{aligned}$$

Similarly,

$$\begin{aligned} P(Z \le z, W = 0) &= \int_0^z e^{-\lambda x} \mu e^{-\mu x} dx \\ &= \frac{\mu}{\lambda + \mu} \int_0^z (\lambda + \mu) e^{-(\lambda+\mu)y} dy \\ &= \frac{\mu}{\lambda + \mu} (1 - e^{-(\lambda+\mu)z}). \end{aligned}$$

Thus Z and W are independent with $Z \sim$ Exponential$(\lambda + \mu)$ and $W \sim$ Bernoulli$(\lambda/(\lambda + \mu))$.