Statistics 581, Solutions, Problem Set 1

Wellner; 10/4/2006

1. Let X and Y be i.i.d. Uniform$(0,1)$ random variables Define $U = X - Y$, $V = \max(X, Y) = X \vee Y$.

   (i) What is the range of (U, V)?

   (ii) Find the joint density function $f_{U,V}(u, v)$ of the pair (U, V). Are U and V independent?

   Solution: (i) The range of (X, Y) is $A = \{(x, y) : 0 \le x \le 1,\ 0 \le y \le 1\}$. The range of (U, V) is

   $$B = \{(u, v) : 0 \le u \le 1, u \le v \le 1\} \cup \{(u, v) : -1 \le u < 0, -u \le v \le 1\}.$$

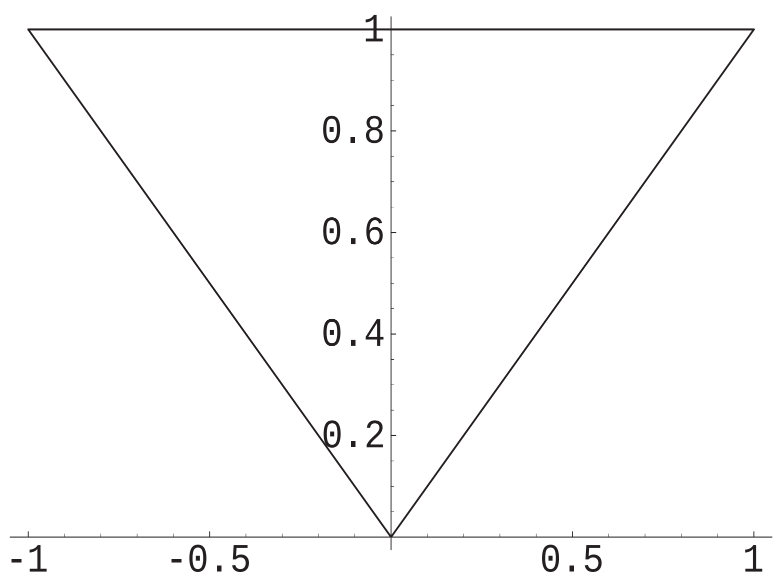

Figure 1: Range of U, V .

(ii) First solution - via Jacobians: The transformation $(X, Y) \to (U, V)$ is 1-1 and onto from A to B. On the set $x < y$, its inverse is given by $X = U + V$, $Y = V$; on the set $x > y$, its inverse is given by $X = V$, $Y = V - U$. These mappings are continuously differentiable on $B^* \equiv B \setminus \{(u,v) : (0,v)\} = B \setminus$ a null set. On B^* the Jacobian of the transformations are

$$\det \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = 1 \qquad \text{if } x < y, \qquad \det \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix} = 1 \qquad \text{if } x > y. \tag{1}$$

Thus by the usual transformation of densities formula, the joint density of (U, V) is obtained from $f_{X,Y}(x, y) = 1_{[0,1]}(x)1_{[0,1]}(y)$ as follows:

$$\begin{aligned} f_{U,V}(u,v) &= f_{X,Y}(x(u,v), y(u,v))|\det \frac{\partial(x,y)}{\partial(u,v)}|1_{[x(u,v)<y(u,v)]} \\ &\qquad + f_{X,Y}(x(u,v), y(u,v))|\det \frac{\partial(x,y)}{\partial(u,v)}|1_{[x(u,v)>y(u,v)]} \\ &= \big(1_{[0,1]}(u+v)1_{[0,1]}(v)1_{[u+v<v]} + 1_{[0,1]}(v)1_{[0,1]}(v-u)1_{[v>v-u]}\big) \\ &= 1_B(u,v)\,. \end{aligned}$$

Thus the joint density of (U, V) is uniform on B. The random variables U and V are clearly *not* independent since the range of (U, V) is not a product set in R^2; moreover, the joint density of (U, V) does not factor into the product of its marginal densities. [The marginal densities are given by

$$f_U(u) = \int f_{U,V}(u,v)dv = \begin{cases} \int_u^1 dv = 1 - u, & u \in [0,1] \\ \int_{-u}^1 dv = 1 + u, & u \in [-1, 0) \end{cases}$$

and

$$f_V(v) = \int f_{U,V}(u,v)du = \int_{-v}^{v} du = 2v1_{[0,1]}(v).]$$

Second solution by direction calculation of the joint distribution function: Note that we can write

$$\begin{aligned} &P(U \le u, V \le v) \\ &\quad = P(X - Y \le u, X \vee Y \le v) = P(X - Y \le u, X \le v, Y \le v) \\ &\quad = P(Y \ge X - u, X \le v, Y \le v) \\ &\quad = \begin{cases} v^2 - \frac{1}{2}(v-u)^2, & \text{if } \ 0 \le u \le v \le 1, \\ \frac{1}{2}(v+u)^2, & \text{if } \ -1 \le u < 0, \ 0 < -u \le v \le 1\,. \end{cases} \end{aligned}$$

(This is easy by pictures!) Computing $(\partial^2/\partial u \partial v) P(U \le u, V \le v)$ on each of these pieces separately again yields $f_{U,V}(u,v) = 1_B(u,v)$. Also note that the marginal distribution functions of U and V are given by $F_U(u) = (1/2)(1+u)^2 1_{[-1,0)}(u) + \{1 - \frac{1}{2}(1-u)^2\} 1_{[0,1]}(u)$ on $-1 \le u \le 1$ and $F_V(v) = v^2$ for $0 \le v \le 1$.

2. Lehmann & Casella, TPE, problem 5.33, page 69.
Morris (1982, 1983b) investigated the properties of natural exponential families with quadratic variance functions. There are only six such families: normal, binomial, gamma, Poisson, negative binomial, and the lesser-known generalized hyperbolic secant distribution, which is the density of $X = \pi^{-1} \log(Y/(1-Y))$ when $Y \sim \text{Beta}((1/2) + \theta/\pi, (1/2) - \theta/\pi)$ $|\theta| < \pi/2$. For this sixth family:
(a) Find the density of X and show that it constitutes an exponential family.
(b) Find the mean and variance of X, and show that the variance equals $1 + \mu^2$ where μ is the mean.
(Subsequent work on quadratic and other power variance families has been done by Bar-Lev and Enis (1986, 1988), Bar-Lev and Bshouty (1989), and Letac and Mora (1990).)
Solution: (a) Since $Y \sim \text{Beta}((1/2) + \theta/\pi, 1/2 - \theta/\pi)$,

$$f_Y(y;\theta) = C_\theta y^{\theta/\pi - 1/2} (1-y)^{-\theta/\pi - 1/2}$$

for $|\theta| < \pi/2$ where $C_\theta = \Gamma(1)/[\Gamma(1/2 + \theta/\pi)\Gamma(1/2 - \theta/\pi)]$. Thus $X = \pi^{-1} \log(Y/(1-Y))$ has distribution function

$$\begin{aligned} P_\theta(X \le x) &= P_\theta(\frac{1}{\pi} \log \frac{Y}{1-Y} \le x) \\ &= P_\theta(\frac{Y}{1-Y} \le e^{\pi x}) \\ &= P_\theta(Y \le e^{\pi x}/(1+e^{\pi x})) = F_Y\left(\frac{e^{\pi x}}{1+e^{\pi x}}; \theta\right). \end{aligned}$$

Thus the density $f_X(\cdot;\theta)$ of X is given by

$$\begin{aligned} f_X(x;\theta) &= f_Y\left(\frac{e^{\pi x}}{1+e^{\pi x}}; \theta\right) \cdot \frac{d}{dx}\left(\frac{e^{\pi x}}{1+e^{\pi x}}\right) \\ &= C_\theta \left(\frac{e^{\pi x}}{1+e^{\pi x}}\right)^{\theta/\pi - 1/2} \left(\frac{1}{1+e^{\pi x}}\right)^{-\theta/\pi - 1/2} \left\{\frac{\pi e^{\pi x}}{1+e^{\pi x}} - \frac{e^{\pi x}\pi e^{\pi x}}{(1+e^{\pi x})^2}\right\} \end{aligned}$$

$$
\begin{aligned}
&= C_\theta \frac{e^{(\theta-1/2)x}}{(1+e^{\pi x})^{-1}} \cdot \frac{\pi e^{\pi x}}{(1+e^{\pi x})^2} \\
&= C_\theta \exp((\theta-\pi/2)x)\frac{\pi e^{\pi x}}{1+e^{\pi x}} \\
&= \pi C_\theta \exp(\theta x)\frac{e^{\pi x/2}}{1+e^{\pi x}} \\
&\equiv A_\theta \exp(\theta x) h(x) \equiv \exp(\theta x - B(\theta)) h(x)
\end{aligned}
$$

where $A_\theta \equiv \pi/(\Gamma(1/2+\theta/\pi)\Gamma(1/2-\theta/\pi))$, $B(\theta) \equiv -\log A_\theta$, and $h(x) = e^{\pi x/2}/(1+e^{\pi x})$. Here is a plot of these densities for $\theta = 0, (1/5)(\pi/2), (2/5)(\pi/2), (3/5)(\pi/2), (4/5)(\pi/2), (9/10)(\pi/2)$

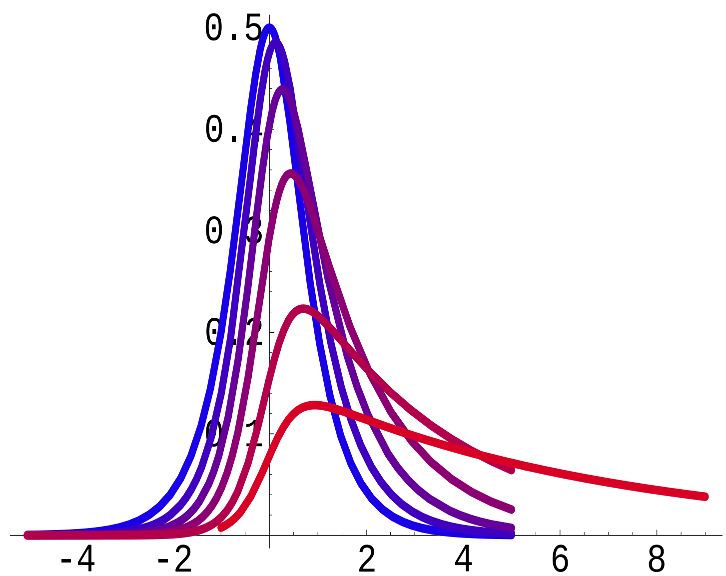

Figure 2: Densities $f_X(x;\theta)$, $\theta \in \{(j/5)(\pi/2), j = 0,\ldots,4\} \cup \{(9/10)(\pi/2)\}$.

(b) Now from exponential family theory

$$E_\theta X = B'(\theta) \qquad \text{and} \qquad Var_\theta(X) = B''(\theta);$$

see e.g. problem 5.6 (a), Lehmann and Casella, page 66. Furthermore

$$
\begin{aligned}
A_\theta &= \frac{\pi}{\Gamma(1/2+\theta/\pi)\Gamma(1/2-\theta/\pi)} \\
&= \sin(\pi(1/2+\theta/\pi)) = \sin(\pi/2+\theta)
\end{aligned}
$$

by the reflection formula for the Gamma function:

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)}$$

for all complex z. Thus $B(\theta) = -\log(\sin(\pi/2 + \theta))$, and we calculate

$$\mu(\theta) \equiv E_\theta X = B'(\theta) = -\frac{\cos(\pi/2+\theta)}{\sin(\pi/2+\theta)} = -\cot(\pi/2+\theta),$$

$$\begin{aligned} Var_\theta(X) &= B''(\theta) = 1 + \frac{\cos^2(\pi/2+\theta)}{\sin^2(\pi/2+\theta)} \\ &= 1 + \mu(\theta)^2. \end{aligned}$$

3. (a) Lehmann & Casella, TPE, problem 3.5, page 64.
Let S be the support of a distribution on a Euclidean space $(\mathcal{X}, \mathcal{A})$. Then, (i) S is closed; (ii) $P(S) = 1$; (iii) S is the intersection of all closed sets C with $P(C) = 1$. (The *support* S of a distribution P on $(\mathcal{X}, \mathcal{A})$ is the set of all points x for which $P(A) > 0$ for all open rectangles $A = \{(x_1, \ldots, x_n) : \ a_i < x < b_i, \ i = 1, \ldots, n\}$ for numbers $a_i < b_i$ in R.)
(b) Lehmann & Casella, TPE, problem 3.6, page 64.
If P and Q are two probability measures over the same Euclidean space which are equivalent, then they have the same support.
(c) Lehmann & Casella, TPE, problem 3.7, page 64.
Let P and Q assign probabilities

$$P: \quad P(X = 1/n) = p_n > 0, \quad n = 1, 2, \ldots \quad (\sum_n p_n = 1),$$

$$Q: \quad P(X = 0) = 1/2; \quad P(X = 1/n) = q_n > 0, \quad n = 1, 2, \ldots (\sum_n q_n = 1/2).$$

Then, show that P and Q have the same support but are not equivalent.

Solution: (a) (i) Suppose that S is not closed. Then there exists a sequence $\{x_n\} \subset S$ such that $x_n \to x_0 \in S^c$. But then, for every $\epsilon > 0$ there is an open ball $B(x_0, \epsilon)$ such that $x_n \in B(x_0, \epsilon)$ for $n \geq N_\epsilon$. Since each x_n is a support point, $P(B(x_0, \epsilon)) > 0$ for each $\epsilon > 0$. But for any open set A with $x_0 \in A$, $B(x_0, \epsilon) \subset A$ for some $\epsilon > 0$, and hence

$P(A) \geq P(B(x_0, \epsilon) > 0$. But this implies $x_0 \in S$. Contradiction. Thus S is closed.
(ii) $P(S) = 1$. From (i) S is closed, so S^c is open. Since $x \in S^c$ if and only if $x \in A_x$ with A_x an open rectangle satisfying $P(A) = 0$. Thus $S^c \subset \cup_x A_x$. By the Lindelöf theorem, for any such open covering $\{A_x\}_{x \in S^c}$ of $S^c \subset R^d$, there is a countable subcollection $\{A_{x_n}\}$ which covers S^c: $S^c \subset \cup_n A_{x_n}$. Then we have

$$P(S^c) \leq P(\cup_n A_{x_n}) \leq \sum_n P(A_{x_n}) = \sum_n 0 = 0\,.$$

Hence $P(S) = 1$.
(iii) We want to show that $S = \cap\{C : \ C \text{ closed}, P(C) = 1\}$. From (i) and (ii) we know that S is in the collection of sets on the right side, so it follows that $S \supset \cap\{C : \ C \text{ closed}, P(C) = 1\}$. Thus it remains to show that $S \subset \cap\{C : \ C \text{ closed}, P(C) = 1\}$. Equivalently, it remains to show that $S^c \supset \cup\{C^c : C^c \text{ open}, \ P(C^c) = 0\}$. But if $x \in \cup\{C^c : C^c \text{ open}, \ P(C^c) = 0\}$, then $x \in C^c$ for some C^c open with $P(C^c) = 0$, and hence also $x \in A \subset C^c$ for some open rectangle A (an open ball centered at x for the metric $\|y\| = \max_{1 \leq i \leq d} |x_i|$) with $P(A) \leq P(C^c) = 0$. Hence $x \in S^c$.
(b) Suppose that P and Q are equivalent: i.e. $Q \prec\prec P$ and $P \prec\prec Q$. Then for any open set A, $P(A) = 0$ if and only if $Q(A) = 0$. This implies that for any closed set A^c,

$$P(A^c) = 1 \qquad \text{if and only if} \qquad Q(A^c) = 1\,.$$

This implies that the minimal closed set S_P with $P(S_P) = 1$ is also the minimal closed set S_Q with $Q(S_Q) = 1$; i.e. $S_P = \text{supp}(P) = \text{supp}(Q) = S_Q$.
(c) Since $P(X = 1/n) = p_n > 0$ for $n = 1, 2, \ldots$ with $\sum_1^\infty p_n = 1$, it follows that $\text{supp}(P) = \{0, \ldots, 1/n, \ldots, 1/2, 1\}$, which is closed. Similarly, Since $Q(X = 1/n) = q_n > 0$ for $n = 1, 2, \ldots$ with $\sum_1^\infty q_n = 1/2$, and $Q(X = 0) = 1/2$, it follows that $\text{supp}(Q) = \{0, \ldots, 1/n, \ldots, 1/2, 1\} = \text{supp}(P)$. But $P(\{0\}) = 0$ while $Q(\{0\}) = 1/2$, so $Q \prec\prec P$ fails. Thus Q and P are not equivalent.

4. Suppose that $X \sim$ Uniform$(0, 1)$ and $Y = -\log(1 - X)$. Find the joint distribution function $F(x, y) = F_{X,Y}(x, y)$ of (X, Y).

Solution: The joint distribution function F of (X, Y) is given by

$$\begin{aligned} F(x,y) &= P(X \le x, Y \le y) = P(X \le X, -\log(1-X) \le y) \\ &= P(X \le x, X \le 1 - e^{-y}) = P(X \le x \wedge 1 - e^{-y}) \\ &= \begin{cases} x, & x \le 1 - e^{-y}, \\ 1 - e^{-y}, & 1 - e^{-y} \le x. \end{cases} \end{aligned}$$

Note that $F(x, \infty) = x$, $0 \le x \le 1$, while $F(1, y) = 1 - e^{-y}$ for all $0 < y < \infty$, so the marginal distributions of X and Y are Uniform$(0,1)$ and exponential(1) respectlvely. But X and Y are very dependent since $Y = -\log(1-x)$ with probability one. Thus the joint distribution is not absolutely continuous with respect to Lebesgue measure on R^2, and the joint distribution function F does not have a density with respect to Lebesgue measure.

5. Ferguson, ACILST, #6, page 7. (This is known as the Polya-Cantelli lemma; see Chapter 2, Propostion 2.11, page 10.)

   Solution: See Ferguson,ACILST, #6, page 173. [Why does this proof fail if the hypothesis of continuity of the limit df F is violated?]

6. (a) Lehmann and Casella, TPE, problem 1.2, page 62.
   Let $X_1, \ldots, X_n$ be uncorrelated random variables with common expectation θ and variance σ^2. Then, 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) Lehmann and Casella, TPE, problem 1.3, page 62.
   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/α_i (α_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 $(u'v)^2 \le (u'u)(v'v)$ 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

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

   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:
1.2: In this case $\Sigma = \sigma^2 I$, so $\alpha_0 = \mathbf{1}(1/\sigma^2)/(\mathbf{1}' I \mathbf{1}/\sigma^2) = \mathbf{1}(1/n)$.
1.3(a): The inverse of the matrix $\text{diag}(1/c_i)$ is just $\text{diag}(c_i)$. This implies that $\alpha_0' X = (\sum_1^n a_i X_i)/(\sum_1^n c_i)$ and $Var(\alpha_0' X) = \sigma^2 / \sum c_i$.
1.3(b): 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*!