**Statistics 581, Problem Set 1, Solutions**

Wellner; 10/1/2008

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

   **(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

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

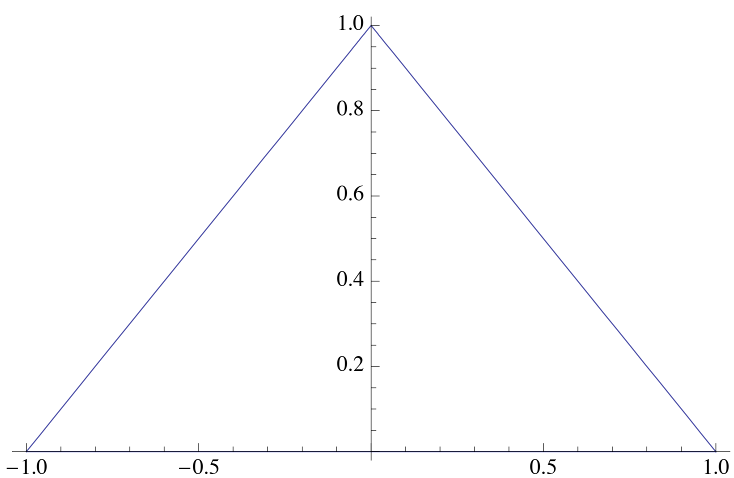


Figure 1: Figure 1: Range of $U, V$).

   (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 $X = V$, $Y = V - U$; on the set $x > y$, its inverse is given by $X = U + V$, $Y = V$. These mappings are continuously differentiable on $B^* \equiv B \setminus \{(u,v) : v = u\} = B \setminus$ a null set. On $B^*$ the Jacobian of the transformations are

$$\det \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix} = 1 \qquad \text{if } x < y, \qquad \det \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = 1 \qquad \text{if } x > y. \tag{0.1}$$

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:

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

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

$$f_U(u) = \int f_{U,V}(u,v)dv = \begin{cases} \int_0^u u+1dv = u+1, & u \in [-1,0] \\ \int_0^1 1-udv = 1-u, & u \in (0,1] \end{cases}$$

and

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

Second solution by direction calculation of the joint distribution function: Note that we can write, for $u \ge 0$,

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

(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 $F_U(u) = (1/2)(u+1)^2 1_{[-1,0)}(u) + \{1-(1/2)(1-u)^2\} 1_{[0,1)}(u)$ on $-1 \le u \le 1$ and $F_V(v) = (1-v)^2$ for $0 \le v \le 1$.

2. Lehmann & Casella, TPE, problem 5.33, page 69, (a) and (b).
(c) Find the papers by Morris (1982, 1983b). In what sense do the the normal, binomial, Poisson, gamma, and negative binomial families have "quadratic variance functions"?

**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

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

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 $\theta = 0, (1/5)(\pi/2), (2/5)(\pi/2), (3/5)(\pi/2), (4/5)(\pi/2), (9/10)(\pi/2)$

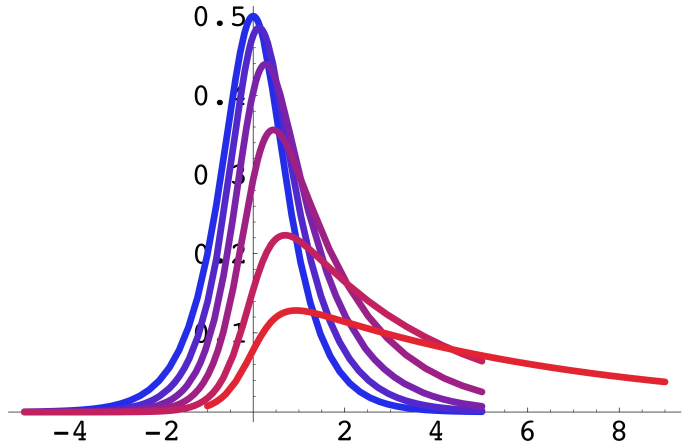


Figure 2: Figure 2: The densities $f_X(x;\theta)$, $\theta \in \{(j/5)(\pi/2), j = 0, \ldots, 4\} \cup \{(9/10)(\pi/2)\}$. ).

(b) To find the mean function of this family we first use the duplication formula for the Gamma function, $\Gamma(z)\Gamma(1-z) = \pi/\sin(\pi z)$ to find that

$$\Gamma(1/2 + \theta/\pi)\Gamma(1/2 - \theta/\pi) = \frac{\pi}{\sin(\pi(1/2 + \theta/\pi))} = \frac{\pi}{\cos(\theta)}$$

to find that $A(\theta) = \cos(\theta)$ and hence that $B(\theta) = -\log(\cos(\theta))$. Then we compute

$$E_\theta(X) = B'(\theta) = \frac{\sin(\theta)}{\cos(\theta)} = \tan(\theta) \equiv \mu,$$
$$Var_\theta(X) = B''(\theta) = 1 + (\tan(\theta))^2 = 1 + \mu^2.$$

(c) *Poisson:* For the Poisson($\lambda$) distributions, $p(x;\lambda) = e^{-\lambda}\lambda^x/x! = \exp(x \log \lambda - \lambda)/x! \equiv \exp(\theta x - B(\theta))h(x)$ with $\theta = \log \lambda$, $B(\theta) = e^\theta$, and $h(x) = 1/x!$, so $E_\lambda(X) = \lambda = Var_\lambda(X)$, and we see that the variance is a linear function of $X$.
*Binomial:* For the Binomial$(n, p)$ family, $p(x;p) = \exp(x \log(p/(1-p)) + n \log(1-p))h(x)$ with $h(x) = \binom{n}{x}$, and $E_p(X) = np \equiv \mu$, $Var_p(X) = np(1-p) = \mu - \mu^2/n$.
*Negative-binomial:* For the Negative binomial$(r, p)$ distributions, $p(x;p) =$

$\exp(x \log p + r \log(1-p))h(x)$ with $h(x) = \Gamma(x+r)/(\Gamma(r)x!)$, and $E_p(X) = rp/(1-p) \equiv \mu$ while $Var_p(X) = rp/(1-p)^2 = \mu + \mu^2/r$.
*Gamma:* (Here I will use notation more consistent with my Section 1.1 and (1.1.19).) For the Gamma$(r, \lambda^{-1})$ family, $p(x;\lambda) = \exp(x(-\lambda) + r\log\lambda)h(x)$ with $h(x) = x^{r-1}$. Here $E_\lambda(X) = r/\lambda \equiv \mu$ while $Var_\lambda(X) = r/\lambda^2 = \mu^2/r$.
*Normal:* For the $N(\lambda, \sigma^2)$ family (with $\sigma^2$ fixed), $p(x;\mu) = \exp(x(\lambda/\sigma^2) - \lambda^2/(2\sigma^2))h(x)$ with $h(x) = \exp(-x^2/(2\sigma^2))$. Here $E_\lambda(X) = \lambda \equiv \mu$ and $Var_\lambda(X) = \sigma^2$, a constant function in the mean $\mu = \lambda$, and hence trivially quadratic.

*Notes:* For an interesting use of these exponential families in the study of the rates of convergence of some Gibbs sampling schemes, see the following recent article:

> Diaconis, P., Khare, K., and Saloff-Coste, L. (2008). Gibbs sampling, exponential families, and orthogonal polynomials. *Statistical Science* **23**, 151 - 178.

3. Lehmann and Casella, TPE, problem 3.8, page 64 (with $W = 1_{[X \le Y]}$). Now do the problem for $X$ with distribution function $F$ independent of $Y$ with distribution function $G$. Does the independence of $Z$ and $W$ hold in general?

   **Solution:** (b) In general, for $X \sim F$, $Y \sim G$, with $X$ and $Y$ independent,

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

   and

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

Thus

$$P(Z \le z) = 1 - P(Z > z) = 1 - P(X > z, Y > z) = 1 - (1 - F(z))(1 - G(z)),$$

while $P(W = 1) = \int_{[0,\infty)} (1 - G(x-)) dF(x)$. Thus $Z$ and $W$ are not independent in general. *Notes: The random variables $Z$ and $W$ arise naturally in biostatistics via random censoring (from the right) where $X$ represents a survival time of interest and $Y$ is a censoring time. Then these formulas give the distribution of the data $(Z, W)$ in terms of the distribution function $F$ of interest and the distribution function $G$ of the censoring variable $Y$. We will see these later when we consider the non-parametric MLE (or Kaplan-Meier estimator) of $F$ in this setting.*
(a) When $X \sim \text{Exp}(\lambda)$ and $Y \sim \text{Exp}(\mu)$,

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

and

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

In this (very special) case,

$$P(Z \le z) = P(Z \le z, W = 1) + P(Z \le z, W = 0) = 1 - \exp(-(\lambda + \mu)z)$$

while $P(W = 1) = \lambda/(\lambda + \mu) = 1 - P(W = 0)$, so $Z$ and $W$ are independent.

4. Suppose that $X \sim$ Uniform$(0, 1)$ and $Y = (1 - X^2)^{-1}$. Find the joint distribution function $F(x, y) = F_{X,Y}(x, y)$ of $(X, Y)$. Does the pair $(X, Y)$ have a joint density function (with respect to Lebesgue measure)?

**Solution:** Here, for $0 \le x \le 1$, $y \ge 1$, we compute

$$\begin{aligned} F(x, y) &= P(X \le x, Y \le y) = P(X \le x, (1 - X^2)^{-1} \le y) \\ &= P(X \le x,\ 1 - X^2 \ge 1/y) = P(X \le x,\ 1 - 1/y \ge X^2) \\ &= P(X \le x,\ X \le \sqrt{1 - 1/y}) = x \wedge \sqrt{1 - 1/y}. \end{aligned}$$

This distribution is concentrated (or puts all its mass) on the set $C = \{(x, (1-x^2)^{-1}) : \ 0 \le x < 1\}$ in the positive orthant. This set has $P(C) = 1$, but $\lambda(C) = 0$ where $\lambda$ is two-dimensional Lebesgue measure on $R^2$. Thus $F$ is *singular* with respect to $\lambda$ and it does not have a density $f$.

5. Ferguson, ACILST, #1, page 6.

**Solution:** (a) Since $\Gamma(1+1/n) = (1/n)\Gamma(1/n)$, $\Gamma(1/n) = n\Gamma(1+1/n) \sim n$ (noting that $\Gamma(1) = 0! = 1$). Similarly, $\Gamma(1+2/n) = (2/n)\Gamma(2/n)$, and hence $\Gamma(2/n) = (n/2)\Gamma(1+2/n) \sim n/2$. Thus the Beta$(1/n, 1/n)$ densities are

$$\begin{aligned} f_n(x) &= \frac{\Gamma(2/n)}{\Gamma(1/n)^2} x^{1/n-1}(1-x)^{1/n-1} 1_{(0,1)}(x) \\ &\sim \frac{1}{2n} x^{-1}(1-x)^{-1} 1_{(0,1)}(x) \\ &\to 0 \qquad \text{as } n \to 0 \end{aligned}$$

for any fixed $x \in (0,1)$. Thus for each fixed $\epsilon > 0$, $P(\epsilon \le X_n \le 1-\epsilon) \to 0$ and, by symmetry, $P(X_n \le \epsilon) = P(X_n \ge 1-\epsilon) \to 1/2$. Thus $X_n \to_d X$ with $X \sim$ Bern$(1/2)$.

(b) When $X_n \sim$ Beta$(\alpha/n, \beta/n)$, then the same sort of computation as in (a) yields

$$\begin{aligned} f_n(x) &= \frac{\Gamma(1+(\alpha+\beta)/n)/((\alpha+\beta)/n)}{\Gamma(1+\alpha/n)/(\alpha/n)\Gamma(1+\beta/n)/(\beta/n)} x^{\alpha/n-1}(1-x)^{\beta/n-1}, \quad 0 < x < 1 \\ &\sim \frac{\alpha\beta}{n(\alpha+\beta)} x^{-1}(1-x)^{-1} \to 0, \end{aligned}$$

so $P(\epsilon < X_n < 1-\epsilon) \to 0$ for every $0 < \epsilon \le 1/2$ as before. But now

$$\begin{aligned} P(X_n \le x) &= \int_0^x f_n(y)dy \\ &\sim \frac{\alpha\beta}{n(\alpha+\beta)} \int_0^x y^{\alpha/n-1}(1-y)^{\beta/n-1} dy \\ & \begin{cases} \le \frac{\alpha\beta}{n(\alpha+\beta)} \int_0^x y^{\alpha/n-1} dy (1-x)^{\beta/n-1} = \frac{\beta}{\alpha+\beta} x^{\alpha/n}(1-x)^{\beta/n-1} \\ \ge \frac{\alpha\beta}{n(\alpha+\beta)} \int_0^x y^{\alpha/n-1} dy = \frac{\beta}{\alpha+\beta} x^{\alpha/n} \end{cases} \\ &\to \begin{cases} \frac{\beta}{\alpha+\beta}(1-x)^{-1} \\ \frac{\beta}{\alpha+\beta} \end{cases} . \end{aligned}$$

For $x$ arbitrarily close to zero the upper and lower bounds become equal (in the limit) and the common value is $\beta/(\alpha+\beta)$. Combined with $P(\epsilon < X_n < 1-\epsilon) \to 0$, this implies that $P(X_n \le x) \to \beta/(\alpha+\beta)$ for $0 < x < 1$, and hence $X_n \to_d$ Bernoulli$(\alpha/(\alpha+\beta))$. [Note the slight difference with Ferguson's solution at this point; see ACILST page 172, line -7.]

6. (a) Lehmann and Casella, TPE, problem 1.4, page 62.
(b) Lehmann and Casella, TPE, problem 1.10, page 62.

**Solution:**
(a) (a') Now

$$\sigma^2 = Var(X) < Var\left(\frac{X+Y}{2}\right) = \frac{1}{4}(\sigma^2 + \tau^2 + 2\rho\sigma\tau)$$

if and only if

$$\frac{3}{4}\sigma^2 - \frac{\rho\tau}{2}\sigma - \frac{\tau^2}{4} \le 0.$$

If we view $\rho$ and $\tau$ as fixed, then the quadratic equation

$$\frac{3}{4}\sigma_0^2 - \frac{\rho\tau}{2}\sigma_0 - \frac{\tau^2}{4} = 0$$

has solutions

$$\sigma_0^{\pm} = \frac{\rho\tau/2 \pm \sqrt{\rho^2\tau^2/4 + 4(3/4)(\tau^2/4)}}{2(3/4)} = \frac{\rho\tau}{3} \pm \sqrt{\frac{\rho^2\tau^2}{9} + \frac{\tau^2}{3}}.$$

since $\sigma > 0$, only the solution with $+$ is relevant, and we conclude that the desired inequality holds if

$$\sigma < \sigma_0^{+} = \frac{\rho\tau}{3} + \sqrt{\frac{\rho^2\tau^2}{9} + \frac{\tau^2}{3}}.$$

(b') Define

$$V(\alpha) \equiv Var(\alpha X + (1-\alpha)Y) = \alpha^2\sigma^2 + (1-\alpha)^2\tau^2 + 2\alpha(1-\alpha)\rho\sigma\tau.$$

Then

$$V'(\alpha) = 2\alpha\sigma^2 - 2(1-\alpha)\tau^2 + 2(1-2\alpha)\rho\sigma\tau = 0$$

if

$$\alpha = \frac{\tau^2 - \rho\sigma\tau}{\sigma^2 + \tau^2 - 2\rho\sigma\tau}.$$

Note that this corresponds to a minimum of the function $V$ since $\rho \leq 1$ implies that

$$V''(\alpha) = 2(\sigma^2 + \tau^2 - 2\rho\sigma\tau) \geq 2(\sigma - \tau)^2 > 0.$$

and hence the value of $\alpha$ minimizing $V$ is negative if $\tau < \rho\sigma$.

(b) First note that that $H(a)$ is minimized by any value of $a$, say $a_{min}$, satisfying

$$h(a_{min}) = \frac{1}{n}\sum_1^n h(x_i)\,.$$

This follows by noting that

$$H(a) = H(a_{min}) + (h(a_{min}) - h(a))^2\,.$$

Since we have assumned that $h$ is (strictly) monotone, it follows easily that

$$a_{min} = h^{-1}\left(\frac{1}{n}\sum_1^n h(x_i)\right)\,.$$

When $h(x) = x$,

$$a_{min} = \frac{1}{n}\sum_1^n (x_i) = \text{the arithmetic mean}\,.$$

When $h(x) = 1/x$,

$$a_{min} = \frac{1}{\frac{1}{n}\sum_1^n \frac{1}{x_i}} = \text{the harmonic mean}\,.$$

When $h(x) = \log(x)$,

$$\begin{aligned} a_{min} &= \exp\left(\frac{1}{n}\sum_1^n \log(x_i)\right) = \exp\left(\frac{1}{n}\log(\prod_1^n x_i)\right) \\ &= \exp\left(\log\left((\prod_1^n x_i)^{1/n}\right)\right) = \left(\prod_1^n x_i\right)^{1/n} \\ &= \text{the geometric mean}\,. \end{aligned}$$