Statistics 581, Problem Set 1 Solutions

Wellner; 9/30/2009

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

   (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

$$\begin{aligned} B &= \{(u,v) : 0 \le v \le 1, 2v \le u \le 1+u\} \\ &= \{(u,v) :\ 0 \le u \le 1,\ 0 \le v \le u/2\} \\ &\qquad \cup \{(u,v) :\ 1 < u \le 2,\ u-1 \le v \le v/2\}. \end{aligned}$$

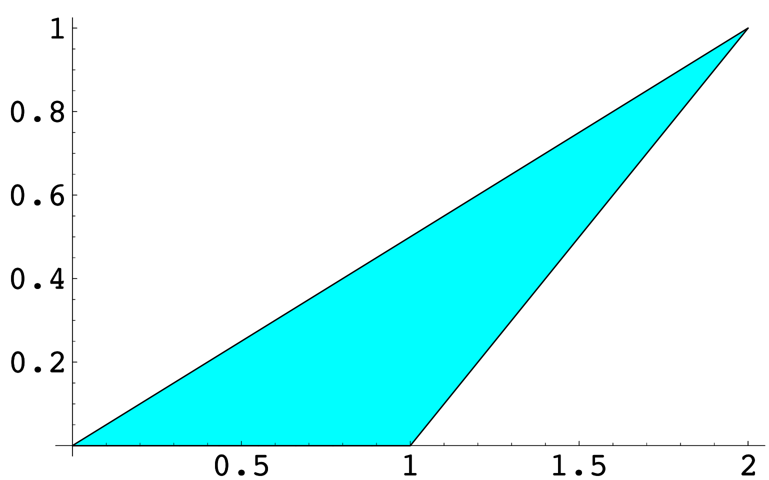

Figure 1: Range of U, V .

(ii) First solution - via Jacobians: The transformation $(X, Y) \to (U, V)$ is 1-1 and onto from $A \cap \{(x, y) : x < y\}$ to B and from $A \cap \{(x, y) : x \geq y\}$ to B. On the set $x < y$, its inverse is given by $X = V$, $Y = U - V$; on the set $x > y$, its inverse is given by $X = U - V$, $Y = V$. These mappings are continuously differentiable on $B^* \equiv B \setminus \{(u, v) : \ v = u/2\} = B \setminus$ 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:

$$\begin{aligned} f_{U,V}(u, v) &= f_{X,Y}(x(u, v), y(u, v))|\det \frac{\partial(x, y)}{\partial(u, v)}|1_{[0<x(u,v)<y(u,v)<1]} \\ &\quad + f_{X,Y}(x(u, v), y(u, v))|\det \frac{\partial(x, y)}{\partial(u, v)}|1_{[1>x(u,v)>y(u,v)>0]} \\ &= \left(1_{[0,1]}(v)1_{[0,1]}(u - v)1_{[0<v<u-v<1]} + 1_{[0,1]}(u - v)1_{[0,1]}(v)1_{[1>u-v>v>0]}\right) \\ &= 2 \cdot 1_{B^*}(u, v) \, . \end{aligned}$$

Thus the joint density of (U, V) is uniform on B^* (or 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

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

the Beta$(1, 2)$ density, and

$$f_U(u) = \int f_{U,V}(u, v)du = \begin{cases} \int_0^{u/2} 2dv = u & 0 \leq u \leq 1 \\ \int_{u-1}^{u/2} 2dv = 2 - u & 1 < u \leq 2, \end{cases}$$

the triangular density on $[0, 2]$.

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

$$P(U \leq u, V > v)$$

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

(This is easy by pictures!) Since

$$F_{U,V}(u,v) = P(U \le u, V \le v) = P(U \le u) - P(U \le u, V > v)$$

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

2. (a) Ferguson, ACILST, #2, page 6.
(b) Now suppose that $U \sim \text{Uniform}(0,1)$ and for each $n \ge 1$ define $V_n \equiv \sum_{j=1}^n (j/n) 1_{((j-1)/n, j/n]}(U)$. Show that $V_n \stackrel{d}{=} X_n$ where X_n is as in (a).
(c) Show that $V_n \to_p U$.

Solution: (a) 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 $F_n(t) \to F(t) = t$ for all $0 \le t \le 1$. That is, $X_n \to_d X \sim \text{Uniform}(0,1)$. 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.
(b) Note that $P(V_n = j/n) = P(U \in ((j-1)/n, j/n]) = 1/n = P(X_n = j/n)$. Thus $V_n \stackrel{d}{=} X_n$.
(c) To see that $V_n \to_p U$, 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 $V_n \to_p U$.

3. 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

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

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

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

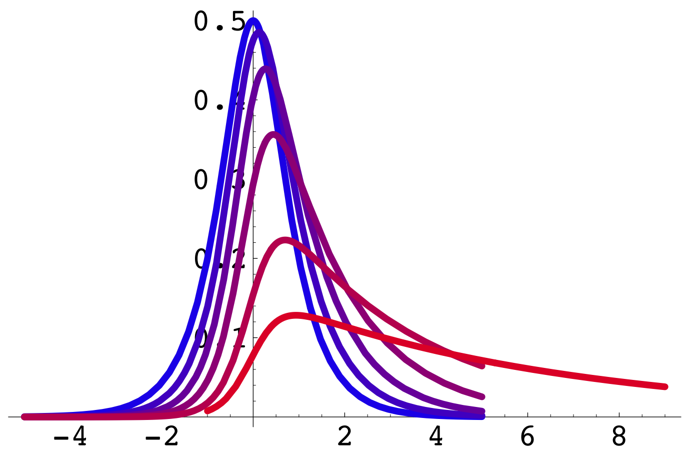

Figure 2: Figure 2: The densities $f_X(x;\theta)$, $\theta \in \{(j/5)(\pi/2), j = 0, \ldots, 4\} \cup \{(9/10)(\pi/2)\}$.).

(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

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

(c) This was not assigned, but here we verify the quadratic variance property for the other five families:
Poisson: For the Poisson(λ) 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 $E(X)$; i.e. a quadratic $aE(x) + bE(X)^2$ with $a = 1$ and $b = 0$.
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,λ^{-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 σ^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.

4. (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 $n \ge N_\epsilon$. 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 $P(A) \ge P(B(x_0,\epsilon) > 0$. 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 $P(A_x) = 0$. 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

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

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 $\|y\| = \max_{1\le i\le d}|x_i|$) with $P(A) \le P(C^c) = 0$. 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.

5. (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. Show that among all linear estimators $\sum_1^n \alpha_i X_i$ of θ satisfying $\sum_1^n \alpha_i = 1$, 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 $\sum_1^n \alpha_i X_i$ over α_i's with $\sum_1^n \alpha_i = 1$ when: (a') the variance of X_i ic σ^2/c_i, $c_i > 0$ and known.
(b') the X_i have common variance σ^2 but are correlated with common correlation coefficient ρ.

Solution: (a) (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 diag$(1/c_i)$ is just diag(c_i). This implies that $\alpha_0'X = (\sum_1^n c_iX_i)/(\sum_1^n c_i)$ 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*!