Lab 7

Linear Algebra and Inequalities for Probability

7.1. Linear Algebra & RV's

These ideas will be explained much more in detail in BIO 235. Here are the basic ideas.

Definition 7.1.1. A *vector space* V (over the real numbers) satisfies, for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $a, b \in \mathbb{R}$,

- (Associativity of addition) $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$
- (Commutativity of addition) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
- (Additive identity) There exists $\mathbf{0} \in V$ such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for all $\mathbf{v} \in V$
- (Additive inverse) There exists $-\mathbf{v} \in V$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$
- (Associativity of scalar multiplication) $a(b\mathbf{v}) = (ab)\mathbf{v}$
- (Multiplicative identity) $1 \cdot \mathbf{v} = \mathbf{v}$
- (Distributivity of scalar sums) $(a + b)\mathbf{v} = a\mathbf{v} + b\mathbf{v}$
- (Distributivity of vector sums) $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$

We may indeed check that random variables (either univariate or multivariate) on the same probability space satisfy the properties of a vector space.

Definition 7.1.2. A *real linear map* $L : V \to \mathbb{R}$ satisfies, for all $\mathbf{u}, \mathbf{v} \in V$ and $a \in \mathbb{R}$,

- (Additivity) $L(\mathbf{u} + \mathbf{v}) = L\mathbf{u} + L\mathbf{v}$
- (Homogeneity) $La\mathbf{u} = aL\mathbf{u}$

If we restrict our space to random variables such that $\mathbb{E}|X| < \infty$, we see that the expectation operator is indeed a linear operator.

Definition 7.1.3. A *real inner product* $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ (where V is a vector space) that, for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ and $a \in \mathbb{R}$,

- (Positive definiteness) $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$, with equality iff $\mathbf{x} = \mathbf{0}$ (zero element of V)

- (Symmetry) $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$
- (Linearity) $\langle \mathbf{x} + \mathbf{x}', a\mathbf{y} \rangle = a\langle \mathbf{x}, \mathbf{y} \rangle + a\langle \mathbf{x}', \mathbf{y} \rangle$

Now, given $\mathbb{E}|X|^2, \mathbb{E}|Y|^2 < \infty$, one may indeed check that $\langle X, Y \rangle = \mathbb{E}XY$ or $= \text{Cov}(X, Y)$ satisfies the real inner product.

Theorem 7.1.1. On top of the linear algebra properties, there are additional properties of expectation and covariance. Let $\mathbf{X} \in \mathbb{R}^m, \mathbf{Y} \in \mathbb{R}^n$ be random vectors, and $A_{m\times n}, B_{n\times p}$ be constant matrices, and $\mathbf{b} \in \mathbb{R}^n$. Let $\mu_{\mathbf{X}} = \mathbb{E}X \in \mathbb{R}^m$, $\Sigma_{\mathbf{XY}} = \text{Cov}(\mathbf{X}, \mathbf{Y}) \in \mathbb{R}^{m\times n}$.

1. $\mathbb{E}(A\mathbf{Y} + \mathbf{b}) = A\mu_{\mathbf{Y}} + \mathbf{b}$
2. $\mathbb{E}A\mathbf{Y}B = A\mu_{\mathbf{Y}}B$
3. $\text{Cov}(A\mathbf{X}, B\mathbf{Y}) = A\Sigma_{\mathbf{XY}}B^\intercal$, where $A \in \mathbb{R}^{a\times m}$ and $B \in \mathbb{R}^{b\times n}$
4. $\mathbb{E}(\mathbf{X}^\intercal A\mathbf{Y}) = \text{tr}(A\Sigma_{\mathbf{YX}}) + \mu_{\mathbf{X}}^\intercal A\mu_{\mathbf{Y}}$

with property 4 giving us the immediate corollary

5. $\mathbb{E}(\mathbf{X}^\intercal C\mathbf{X}) = \text{tr}(C\Sigma_{\mathbf{X}}) + \mu_{\mathbf{X}}^\intercal C\mu_{\mathbf{X}}$

where $\text{Var}(X) = \Sigma_{\mathbf{X}}$ and $C \in \mathbb{R}^{m\times m}$. Property 4 deserves a proof.

Proof. We have

$$\begin{aligned}
\mathbb{E}(\mathbf{X}^\intercal A\mathbf{Y}) &= \mathbb{E}\left(\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}X_iY_j\right) \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}\left[\text{Cov}(X_i, Y_j) + \mu_{X_i}\mu_{Y_j}\right] \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}\sigma_{ij} + \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}\mu_{X_i}\mu_{Y_j} \\
&= \text{tr}(A\Sigma_{\mathbf{YX}}) + \mu_{\mathbf{X}}^\intercal A\mu_{\mathbf{Y}}
\end{aligned}$$

□

Example 7.1.1. Let $\mathbf{X} = (X_1, \cdots, X_n)$ with each X_i iid with mean μ and variance σ^2. If we define J_n as a matrix of 1's of dimension $n \times n$, we see that

$$\widetilde{\mathbf{X}} \stackrel{\text{def}}{=} \underbrace{\left(\mathcal{I}_n - \frac{1}{n}J_n\right)}_{H}\mathbf{X} = \begin{pmatrix} X_1 - \overline{X} \\ X_2 - \overline{X} \\ \vdots \\ X_n - \overline{X} \end{pmatrix}$$

where $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Recall that the sample variance $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n-1}\widetilde{\mathbf{X}}^\intercal\widetilde{\mathbf{X}}$. First, it is easy to check that $H = H^\intercal$ and $H^2 = H$ (that is, H is a projection matrix). First compute

$$\mathbb{E}\widetilde{\mathbf{X}} = H\mathbb{E}\mathbf{X} = \mathbf{0} \qquad \text{and} \qquad \text{Var}(\widetilde{\mathbf{X}}) = H\text{Var}(\mathbf{X})H^\intercal = H(\sigma^2\mathcal{I}_n)H^\intercal = \sigma^2 H$$

Using property 5 with $C = \mathcal{I}_n$,

$$\mathbb{E}\widetilde{\mathbf{X}}^\intercal\widetilde{\mathbf{X}} = \mathrm{tr}(\Sigma_{\widetilde{\mathbf{X}}}) + \mu_{\widetilde{\mathbf{X}}}^\intercal \mu_{\widetilde{\mathbf{X}}} = \sigma^2 \mathrm{tr}\left(\mathcal{I}_n - \frac{1}{n}J_n\right) = \sigma^2(n-1)$$

Hence, $\mathbb{E}s^2 = \sigma^2$, and so the sample variance is an unbiased estimator of the true variance.

7.2. A Potpourri of Inequalities

Example 7.2.1. Let X, Y be RV's with finite second moment.

$$\begin{aligned} \mathbb{E}XY &\overset{(1)}{\leq} \frac{1}{4}\mathbb{E}(X+Y)^2 &&\overset{(2)}{\leq} \frac{1}{2}\mathbb{E}X^2 + \frac{1}{2}\mathbb{E}Y^2 \\ \mathbb{E}XY &\overset{(3)}{\leq} (\mathbb{E}X^2)^{\frac{1}{2}}(\mathbb{E}Y^2)^{\frac{1}{2}} &&\overset{(4)}{\leq} \frac{1}{2}\mathbb{E}X^2 + \frac{1}{2}\mathbb{E}Y^2 \end{aligned}$$

Proof. First, we have

$$0 \leq \mathbb{E}(X-Y)^2 = \mathbb{E}X^2 + \mathbb{E}Y^2 - 2\mathbb{E}XY \iff \mathbb{E}XY \leq \frac{1}{2}\mathbb{E}X^2 + \frac{1}{2}\mathbb{E}Y^2$$

Next, using the fact that $a \leq b \implies a \leq \frac{a+b}{2} \leq b$, we can easily confirm that

$$\frac{1}{4}\mathbb{E}(X+Y)^2 = \frac{\mathbb{E}XY + (\frac{1}{2}\mathbb{E}X^2 + \frac{1}{2}\mathbb{E}Y^2)}{2}$$

hence proving inequalities (1) and (2) instantly.

Inequality (3) is just Cauchy-Schwarz; here's an alternate proof different from class. Let $\widetilde{X} = X/(\mathbb{E}X^2)^{\frac{1}{2}}$ and $\widetilde{Y} = Y/(\mathbb{E}Y^2)^{\frac{1}{2}}$, we have

$$\begin{aligned} \mathbb{E}\widetilde{X}\widetilde{Y} &\leq \frac{1}{2}\mathbb{E}\widetilde{X}^2 + \frac{1}{2}\mathbb{E}\widetilde{Y}^2 \\ \implies \mathbb{E}\frac{X}{(\mathbb{E}X^2)^{\frac{1}{2}}}\frac{Y}{(\mathbb{E}Y^2)^{\frac{1}{2}}} &\leq \frac{1}{2}\mathbb{E}\frac{X^2}{\mathbb{E}X^2} + \frac{1}{2}\mathbb{E}\frac{Y^2}{\mathbb{E}Y^2} \\ \implies \mathbb{E}XY &\leq (\mathbb{E}X^2)^{\frac{1}{2}}(\mathbb{E}Y^2)^{\frac{1}{2}} \end{aligned}$$

You could use Young's inequality with $p = q = 2$ for inequality (4). Here's an alternate proof:

$$0 \leq [(\mathbb{E}X^2)^{\frac{1}{2}} - (\mathbb{E}Y^2)^{\frac{1}{2}}]^2 = \mathbb{E}X^2 + \mathbb{E}Y^2 - 2(\mathbb{E}X)^{\frac{1}{2}}(\mathbb{E}Y)^{\frac{1}{2}} \iff (\mathbb{E}X^2)^{\frac{1}{2}}(\mathbb{E}Y^2)^{\frac{1}{2}} \leq \frac{1}{2}\mathbb{E}X^2 + \frac{1}{2}\mathbb{E}Y^2$$

Remark: The middle inequalities are not always comparable. Take $X \overset{\text{a.s.}}{=} 0$ and $Y \overset{\text{a.s.}}{\neq} 0$. Clearly, $(\mathbb{E}X^2)^{\frac{1}{2}}(\mathbb{E}Y^2)^{\frac{1}{2}} < \frac{1}{4}\mathbb{E}(X+Y)^2$. For the flip, take $X \overset{\text{a.s.}}{\neq} 0$ and $Y = -X$. Then, $(\mathbb{E}X^2)^{\frac{1}{2}}(\mathbb{E}Y^2)^{\frac{1}{2}} > \frac{1}{4}\mathbb{E}(X+Y)^2$. □

Example 7.2.2. Use probabilistic Jensen's inequality $g(\mathbb{E}X) \leq \mathbb{E}g(X)$ (for convex g) to prove the summation and integral versions of Jensen's inequality:

$$g\left(\sum_{i=1}^{n} \lambda_i x_i\right) \leq \sum_{i=1}^{n} \lambda_i g(x_i) \qquad \text{and} \qquad g\left(\int_a^b \lambda(x)\varphi(x)dx\right) \leq \int_a^b \lambda(x) g(\varphi(x))dx$$

where $\lambda_i \geq 0, \sum_{i=1}^n \lambda_i = 1$ and $\lambda(x) \geq 0$ and $\int_a^b \lambda(x)dx = 1$.

Proof. For the summation case, define a random variable X with $\mathbb{P}(X = x_i) = \lambda_i$. For the integral case, let X have density function $\lambda(t)$ and use probabilistic Jensen's on $\varphi(X)$. □

Example 7.2.3 (Jensen's $\implies$ Lyapunov's)**.** In class, we used Hölder's inequality to prove Lyapunov's inequality. Instead use Jensen's to prove Lyapunov's.

Proof. Let $0 < r \leq s$. Then, $d \stackrel{\text{def}}{=} \frac{s}{r} \geq 1$ and $g(y) = y^d$ is convex for non-negative y. Hence

$$(\mathbb{E}Y)^d \leq \mathbb{E}Y^d \implies (\mathbb{E}Y)^{\frac{1}{r}} \leq (\mathbb{E}Y^{\frac{s}{r}})^{\frac{1}{s}}$$

Let $Y = |X|^r$ and hence we have Lyapunov's $(\mathbb{E}|X|^r)^{\frac{1}{r}} \leq (\mathbb{E}|X|^s)^{\frac{1}{s}}$ □

Example 7.2.4 (Jensen's $\implies$ Minkowski's)**.** The discrete Minkowski's inequality states that for all real a_i, b_i, and $p \geq 1$,

$$\left(\sum_{i=1}^{n} |a_i + b_i|^p\right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^{n} |a_i|^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n} |b_i|^p\right)^{\frac{1}{p}}$$

Use discrete Jensen's to prove this.

Proof. Let $g(x) = (1 + x^{1/p})^p$. We have $g''(x) = -\frac{p-1}{p}(x^{\frac{1}{p}} + 1)^{p-2}/x^{2-\frac{1}{p}} \leq 0$ for all $x \geq 0$ and $p \geq 1$, hence g is *concave* for positive x. Now define

$$\lambda_i = \frac{|a_i|^p}{\sum_{i=1}^n |a_i|^p} \qquad \text{and} \qquad x_i = \frac{|b_i|^p}{|a_i|^p}$$

Then by discrete Jensen's (with the inequality flipped because of the concavity),

$$\begin{aligned} \left(1 + \left(\frac{\sum_{i=1}^n |b_i|^p}{\sum_{i=1}^n |a_i|^p}\right)^{\frac{1}{p}}\right)^p \quad &\geq \frac{\sum_{i=1}^n |a_i|^p \left(1 + \frac{|b_i|}{|a_i|}\right)^p}{\sum_{i=1}^n |a_i|^p} \\ \implies \left(\sum_{i=1}^{n} |a_i|^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n} |b_i|^p\right)^{\frac{1}{p}} &\geq \left(\sum_{i=1}^{n} (|a_i| + |b_i|)^p\right)^{\frac{1}{p}} \end{aligned}$$

Since $|a_i| + |b_i| \geq |a_i + b_i|$ by the standard triangle inequality, our result follows. □

Example 7.2.5. (Moment and Chernoff bounds) Let X be a positive RV with finite moments of all orders. In lecture, we learned of the following variations of Markov's inequality:

$$\mathbb{P}(X \geq \varepsilon) \leq \frac{\mathbb{E}X^p}{\varepsilon^p} \qquad \text{and} \qquad \mathbb{P}(X \geq \varepsilon) \leq \frac{\mathbb{E}e^{tX}}{e^{t\varepsilon}}$$

Since these inequalities are true for any p and t, we might as well sharpen the bounds:

$$\mathbb{P}(X \geq \varepsilon) \leq \min_{p \in \mathbb{N}_0} \frac{\mathbb{E}X^p}{\varepsilon^p} \stackrel{\text{def}}{=} \mathcal{MB}_X(\varepsilon)$$
$$\mathbb{P}(X \geq \varepsilon) \leq \min_{t>0} \frac{\mathbb{E}e^{tX}}{e^{t\varepsilon}} \stackrel{\text{def}}{=} \mathcal{CB}_X(\varepsilon)$$

The two bounds are respectively known as the *moment bounds* and *Chernoff bounds*. It turns out that moments bounds are uniformly better than Chernoff bounds in the sense that

$$\mathcal{MB}_X(\varepsilon) \leq \mathcal{CB}_X(\varepsilon)$$

for any positive random variable X and any $\varepsilon > 0$. We will prove this assertion.

Proof. Recall this inequality from Lab 5, Additional Problem 1, endearingly known as *Cauchy's third inequality*:

$$\min_{1 \leq i \leq n} \frac{a_i}{b_i} \leq \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i} \leq \max_{1 \leq i \leq n} \frac{a_i}{b_i}$$

for any positive $a_1, \cdots, a_n, b_1, \cdots, b_n$. Using the lower bound of this inequality,

$$\min_{0 \leq p \leq n} \frac{t^p \mathbb{E}X^p/p!}{t^p \varepsilon^p/p!} \leq \frac{\sum_{p=0}^n t^p \mathbb{E}X^p/p!}{\sum_{p=0}^n t^p \varepsilon^p/p!}$$

Taking the limit as $n \to \infty$,

$$\min_{p \in \mathbb{N}_0} \frac{\mathbb{E}X^p}{\varepsilon^p} \leq \frac{\mathbb{E}e^{tX}}{e^{t\varepsilon}}$$

which holds for all t, hence take the minimum on the RHS, and we arrive at our conclusion. □

Example 7.2.6 (Hardy's inequality)**.** Let $f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ and $\int_0^\infty f^p(x)dx < \infty$ for $p \geq 1$. Define the averaging function $\overline{f}$ as

$$\overline{f}(x) = \frac{1}{x} \int_0^x f(t)dt$$

Prove *Hardy's inequality*

$$\int_0^\infty \overline{f}^p(x)dx \leq \left(\frac{p}{p-1}\right)^p \int_0^\infty f^p(x)dx$$

Proof. Letting $u = (\int_0^x f(t)dt)^p$ and $dv = \frac{1}{x^p}$, integration by parts gives us

$$\int_0^\infty \overline{f}^p(x)dx = \underbrace{\left\{\frac{1}{-p+1} x^{-p+1} \left(\int_0^x f(t)dt\right)^p\right\}_{x=0}^{x=\infty}}_{A} + \underbrace{\frac{p}{p-1} \int_0^\infty x^{-p+1} f(x) \left(\int_0^x f(t)dt\right)^{p-1} dx}_{B}$$

Quick Lemma. We have the following inequality

$$0 \le x^{-p+1}\left(\int_0^x f(t)dt\right)^p \le \int_0^x f^p(x)dx$$

The expression is clearly greater than 0 for $x \ge 0$, and the upper bound comes from

$$x^{-p+1}\left(\int_0^x f(t)dt\right)^p = x\left(\int_0^x \frac{1}{x}\cdot f(t)dt\right)^p \le x\left(\int_0^x \frac{1}{x}\cdot f^p(t)dt\right) = \int_0^x f^p(x)dx$$

where the inequality follows from integral Jensen's with $g(y) = y^p$, $\lambda(t) = \frac{1}{x}$, and $\varphi(t) = f(t)$. Using the Quick Lemma,

$$0 \le \lim_{x\to 0} x^{-p+1}\left(\int_0^x f(t)dt\right)^p \le \lim_{x\to 0}\int_0^x f^p(x)dx = 0$$
$$0 \le \lim_{x\to \infty} x^{-p+1}\left(\int_0^x f(t)dt\right)^p \le \lim_{x\to \infty}\int_0^x f^p(x)dx < \infty$$

Implying that $A \le 0$. Hence $\int_0^\infty \overline{f}^p(x)dx \le B$. The integral version of Hölder's inequality states that

$$\int_0^\infty |a(x)b(x)|dx \le \left(\int_0^\infty |a(x)|^p dx\right)^{\frac{1}{p}}\left(\int_0^\infty |b(x)|^q dx\right)^{\frac{1}{q}}$$

with $p^{-1} + q^{-1} = 1$. Let $a(x) = f(x), b(x) = \overline{f}^{p-1}(x)$ to get

$$B \le \frac{p}{p-1}\left(\int_0^\infty f^p(x)dx\right)^{\frac{1}{p}}\left(\int_0^\infty \overline{f}^{q(p-1)}(x)dx\right)^{\frac{1}{q}}$$

Since $p^{-1} + q^{-1} = 1 \implies q = \frac{p}{p-1}$, we have

$$\begin{aligned}
\int_0^\infty \overline{f}^p(x)dx &\le \frac{p}{p-1}\left(\int_0^\infty f^p(x)dx\right)^{\frac{1}{p}}\left(\int_0^\infty \overline{f}^p(x)dx\right)^{\frac{1}{q}} \\
\implies \left(\int_0^\infty \overline{f}^p(x)dx\right)^{1-\frac{1}{q}} &\le \frac{p}{p-1}\left(\int_0^\infty f^p(x)dx\right)^{\frac{1}{p}} \\
\implies \int_0^\infty \overline{f}^p(x)dx &\le \left(\frac{p}{p-1}\right)^p \int_0^\infty f^p(x)dx
\end{aligned}$$

as desired. □

7.3. Additional Problems

1. Suppose X satisfies $\mathbb{E}X^2 < \infty$. Find the maximum value of $\mathbb{E}ZX$ subject to $Z \geq 0$ and $\mathbb{E}Z^2 \leq 1$.

2. Let X be a nonnegative random variable with finite and strictly positive second moment. Prove the *Paley–Zygmund inequality*: for all $\varepsilon \in [0,1]$,

$$\mathbb{P}(X > \varepsilon\mathbb{E}X) \geq (1-\varepsilon)^2 \frac{(\mathbb{E}X)^2}{\mathbb{E}X^2}$$

3. This is a variation of problem B4 from the 2013 Putnam Mathematical Competition. Let X and Y be bounded RV's, say $|X| \leq M_X$ and $|Y| \leq M_Y$. They are not necessarily independent. Prove that

$$\operatorname{Var}(XY) \leq 2M_X^2 \operatorname{Var}(Y) + 2M_Y^2 \operatorname{Var}(X)$$

   Perhaps the best way to prove this is through the following steps:

$$\begin{aligned} \operatorname{Var}(XY) &\leq \mathbb{E}(XY - \mathbb{E}X\mathbb{E}Y)^2 && \text{(a)} \\ &\leq 2\mathbb{E}(XY - X\mathbb{E}Y)^2 + 2\mathbb{E}(X\mathbb{E}Y - \mathbb{E}X\mathbb{E}Y)^2 && \text{(b)} \\ &\leq 2M_X^2 \operatorname{Var}(Y) + 2M_Y^2 \operatorname{Var}(X) && \text{(c)} \end{aligned}$$

   Prove steps (a) – (c).

4. Let X, U, V be RV's with finite second moments. We know that $\rho_{UX} \leq 1$ and $\rho_{VX} \leq 1$, and hence we can instantly conclude $\rho_{UX} + \rho_{VX} \leq 2$ and $\rho_{UX}\rho_{VX} \leq 1$. These bound are actually quite weak; if U and V are not completely correlated, then it is impossible for both ρ_{UX} and ρ_{VX} to simultaneously equal 1. We shall derive improved bounds in this problem.

   (a) Let $\widetilde{X} = (X - \mu_x)/\sigma_X$. Show that $\mathbb{E}(\widetilde{X} - \widetilde{U})^2 = 2 - 2\rho_{UX}$.[1]

   (b) Prove the identity

$$2\mathbb{E}A^2 + 2\mathbb{E}B^2 = \mathbb{E}(A+B)^2 + \mathbb{E}(A-B)^2$$

   (c) Using (a) and (b), prove that

$$\rho_{UX} + \rho_{VX} \leq \frac{3 + \rho_{UV}}{2}$$

   Hint: Let $A = \widetilde{X} - (\widetilde{U} + \widetilde{V})/2$ and $B = (\widetilde{U} - \widetilde{V})/2$.

   (d) Define $W = 2\frac{\sigma_{UX}}{\sigma_X^2}X - U$. Show that $\sigma_W^2 = \sigma_U^2$; hence $\sigma_{WV} \leq \sigma_U\sigma_V$. Reorganize the latter to show that

$$\rho_{UX}\rho_{VX} \leq \frac{1 + \rho_{UV}}{2}$$

[1] Note that this also gives us an immediate proof of $\rho_{UX} \leq 1$. Why?