Definition: Let X, $\{X_n\}_{n=1}^{\infty}$ be RVs on $(\Omega, \mathcal{F}, P)$. We say that X_n converges in probability to X if for every $\varepsilon > 0$,

$$\lim_n P(|X_n - X| \geq \varepsilon) = 0.$$

Example: We showed that $\{\overline{X}_n\}_{n=1}^{\infty}$ converges in probability to μ. (Recall for all $\varepsilon > 0$, $\lim_n P(|\overline{X}_n - \mu| \geq \varepsilon) = 0$.)

Definition: Let X, $\{X_n\}_{n=1}^{\infty}$ be RVs on $(\Omega, \mathcal{F}, P)$, then we say X_n converges almost surely in probability to X if

$$P(\{\omega : \lim_n X_n(\omega) = X(\omega)\}) = 1$$

We cannot (yet) that $\overline{X}_n \to \mu$ almost surely. This would be a statement about infinite sequence.

Example: Assume that $\{X_n\}_{n=1}^{\infty}$ are independent RVs such that

$$P(X_n = x) = \begin{cases} \frac{1}{n} & \text{if } x = 1 \\ 1 - \frac{1}{n} & \text{if } x = 0. \end{cases}$$

Then for all $\varepsilon > 0$, $P(|X_n - 0| \geq \varepsilon) = \frac{1}{n} \to 0$ so $\{X_n\}_{n=1}^{\infty} \to 0$ in probability. Note that $\{X_n\} \not\to 0$ almost surely because by the Borel-Cantelli Lemma,

$\sum_{n=1}^{\infty} P(X_n = 1) = \sum_{n=1}^{\infty} \frac{1}{n} = \infty$, so $P(X_n = 1) = 1$ infinitely often.

because $\{X_n\}_{n=1}^{\infty}$ are independent, so

$$P(\{\omega : \lim_n X_n(\omega) = 0\}) = 0$$

Proposition: Let X, $\{X_n\}_{n=1}^{\infty}$ be RVs. If $\{X_n\} \to X$ a.s., then $\{X_n\} \to X$ in probability.

Proof: Let $\varepsilon > 0$ be arbitrary and assume $\{X_n\} \to X$ almost surely. If $\omega \in \{\lim_n X_n(\omega) = X(\omega)\}$, then $\omega \in \bigcup_{n=1}^{\infty} A_n$ where $A_n = \{\omega : |X_N(\omega) - X(\omega)| < \varepsilon \ \forall N \geq n\}$. Then

$$1 = P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_n P(A_n) \leq \lim_n P(|X_n - X| < \varepsilon)$$

$$\Rightarrow \lim_n P(|X_n - X| \geq \varepsilon) = 0.$$

Proposition: (Cauchy-Schwarz Space) Let X, Y be RVs with $E(X^2)$ and $E(Y^2)$ finite, then

$$|E(XY)|^2 \leq E(X^2)E(Y^2)$$

Proof: For any $t \in \mathbb{R}$

$$0 \leq E((tX - Y)^2) = E(t^2X^2 - 2tXY + Y^2)$$

$$= t^2 E(X^2) - 2t E(XY) + E(Y^2)$$

Recall $ax^2 + bx + c \geq 0$ iff $\Delta = b^2 - 4ac \leq 0$. In particular, we have

$$(-2E(XY))^2 - 4E(X^2)E(Y^2) \leq 0$$

$$E(XY)^2 - E(X^2)E(Y^2) \leq 0$$

$$E(XY)^2 \leq E(X^2)E(Y^2).$$

Example: If $Y \equiv 1$, then $E(X)^2 \leq E(X^2)$

Example: For $z > 0$, let $X = \sqrt{z}$, $Y = \frac{1}{\sqrt{z}}$. Then

$$E\left(\sqrt{z} \cdot \tfrac{1}{\sqrt{z}}\right)^2 \leq E(\sqrt{z}^2)\, E\left(\tfrac{1}{\sqrt{z}^2}\right) = E(z)E\left(\tfrac{1}{z}\right)$$

$$\Rightarrow \frac{1}{E(z)} \leq E\left(\tfrac{1}{z}\right).$$

Corollary: $E(|XY|) \leq \sqrt{E(X^2)E(Y^2)}$ (exercise).

Definition: Let $I \subseteq \mathbb{R}$ be an interval. A function $f: I \to \mathbb{R}$ is convex if for all $x, y \in I$, $\lambda \in [0,1]$

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$$

If f is twice differentiable, then convexity is same as f'' nonnegative on I.

Proposition: (Jensen's Inequality) If ϕ is a convex function on I and X is a RV on I, i.e. $P(X \in I) = 1$. Then

$$E(\phi(X)) \geq \phi(E(X)).$$

Corollary: Let Z be a nonnegative RV. Assume $0 < p < q$. Then $E(Z^p)^{\frac{1}{p}} \leq E(Z^q)^{\frac{1}{q}}$

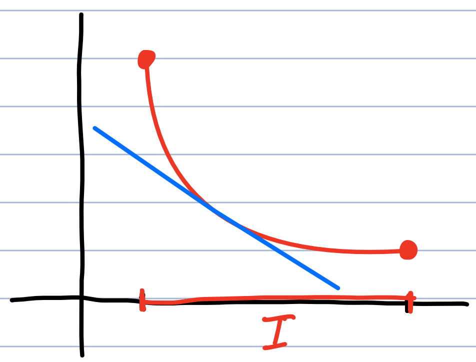
Proof: Let $\phi(x) = x^{\frac{q}{p}}$ convex on $[0, \infty)$. Let $X = Z^p$. The

$$E(Z^q) = E(X^{\frac{q}{p}}) = E(\phi(X)) \geq \phi(E(X)) = E(Z^p)^{\frac{q}{p}}$$

Exercise: Use Jensen to prove for all $x_1, \dots, x_n$

$$\frac{x_1 + \cdots + x_n}{n} \geq (x_1 x_2 \cdots x_n)^{\frac{1}{n}} \qquad (\text{Hint: } -\log(x))$$

Proof of Jensen: Because I is an interval $E(X) \in I$. Convexity implies that ϕ has a tangent line at $(E(X), \phi(E(X)))$



$$L(t) = m(t - E(X)) + \phi(E(X)) \quad \text{such that} \quad \phi(t) \geq L(t)$$

$\forall t \in J$

$$E(\phi(X)) \geq E(L(X)) = E(\cancel{mX} - m\cancel{E(X)} + \phi(E(X))) = \phi(E(X))$$

Lemma: Let $X, \{X_n\}_{n=1}^{\infty}$ be RVs. TFAE.

(a) $\{X_n\} \to X$ almost surely to X

(b) For all $\varepsilon > 0$, $P(|X_n - X| \geq \varepsilon) = 0$ infinitely often.

Proof: (a) $\Leftrightarrow$ (b) : $P(\{w : |X_n(w) - X(w)| < \varepsilon \text{ a.a}\}) = 1$

Put $\varepsilon = \frac{1}{M}$, $P(\{w : |X_n(w) - X(w)| < \frac{1}{M} \text{ a.a } \forall M \in \mathbb{N}\}) = 1$

so $P(\bigcap_{M=1}^{\infty} A_M) = 1$, where $A_M = \{|X_n - X| < \frac{1}{M} \text{ a.a}\}$

$\Leftrightarrow$ $P(A_M) = 1 \ \forall M \Leftrightarrow 0 = P(A_M^c) = P(\{|X_n - X| \geq \frac{1}{M} \text{ i.o}\})$

Theorem: (First Version of Law of Large Numbers)

Assume $\{X_n\}_{n=1}^{\infty}$ are independent RVs, $E(X_n) = \mu$ and

$E((X_n - \mu)^4) \leq A < \infty \ \forall n$. Let $\forall N \in \mathbb{N}$

$$\overline{X}_N = \frac{X_1 + \cdots + X_n}{N}.$$

Then $\overline{X}_N \to \mu$ almost surely.