Math 525: Lecture 8

January 30, 2018

1 Moment inequalities

There are a few useful inequalities concerning moments of random variables we should cover. We start with Markov's inequality.

1.1 Markov's inequality

The following inequality is a special case of a more general measure theoretic result. In the measure theoretic setting, it is called Chebyshev's inequality.

Proposition 1.1 (Markov's inequality)**.** *Let $p > 0$, $\lambda > 0$, and X be a random variable with X^p integrable. Then,*

$$\mathbb{P}(\{|X| \geq \lambda\}) \leq \frac{1}{\lambda^p} \mathbb{E}\left[|X|^p\right].$$

Proof. First, note that

$$\mathbb{P}(\{|X| \geq \lambda\}) = \mathbb{P}(\{|X|^p \geq \lambda^p\}) = \mathbb{E}\left[I_{\{|X|^p \geq \lambda^p\}}\right].$$

But if $|X(\omega)|^p \geq \lambda^p$, then $1 \leq |X(\omega)|^p / \lambda^p$. Therefore,

$$\mathbb{E}\left[I_{\{|X|^p \geq \lambda^p\}}\right] \leq \mathbb{E}\left[\frac{|X|^p}{\lambda^p} I_{\{|X|^p \geq \lambda^p\}}\right] \leq \frac{1}{\lambda^p} \mathbb{E}\left[|X|^p I_{\{|X|^p \geq \lambda^p\}}\right] \leq \frac{1}{\lambda^p} \mathbb{E}\left[|X|^p\right]. \qquad \square$$

Corollary 1.2. *Let $\lambda > 0$ and Y be a square integrable (i.e., Y^2 is integrable) random variable. Then,*

$$\mathbb{P}(\{|Y - \mathbb{E}Y| \geq \lambda\}) \leq \frac{1}{\lambda^2} \operatorname{Var}(Y).$$

Proof. Take $p = 2$ and $X = Y - \mathbb{E}Y$ in Markov's inequality. $\square$

1.2 Cauchy-Schwarz(-Buniakovski) inequality

Proposition 1.3 (Cauchy-Schwarz(-Buniakovski) inequality)**.** *Let X and Y be square integrable random variables. Then,*

$$\mathbb{E}\left[XY\right] \leq \sqrt{\mathbb{E}\left[X^2\right] \mathbb{E}\left[Y^2\right]}.$$

Proof. If either X or Y is zero a.s., then the inequality is trivial. Therefore, suppose that neither is zero a.s. Let $\lambda \geq 0$. Then,

$$0 \leq \mathbb{E}\left[(X-\lambda Y)^2\right] = \mathbb{E}\left[X^2\right] - 2\lambda\mathbb{E}\left[XY\right] + \lambda^2\mathbb{E}\left[Y^2\right]$$

and hence

$$\mathbb{E}\left[XY\right] \leq \frac{1}{2}\left(\frac{1}{\lambda}\mathbb{E}\left[X^2\right] + \lambda\mathbb{E}\left[Y^2\right]\right).$$

Letting $\lambda = \sqrt{\mathbb{E}\left[X^2\right]}/\sqrt{\mathbb{E}\left[Y^2\right]}$ yieldsw

$$\begin{aligned}\mathbb{E}\left[XY\right] &\leq \frac{1}{2}\left(\frac{\sqrt{\mathbb{E}\left[Y^2\right]}}{\sqrt{\mathbb{E}\left[X^2\right]}}\mathbb{E}\left[X^2\right] + \frac{\sqrt{\mathbb{E}\left[X^2\right]}}{\sqrt{\mathbb{E}\left[Y^2\right]}}\mathbb{E}\left[Y^2\right]\right) \\ &= \frac{\mathbb{E}\left[Y^2\right]\mathbb{E}\left[X^2\right]}{\sqrt{\mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right]}} \\ &= \sqrt{\mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right]}.\end{aligned}$$

□

Example 1.4. Let $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Bernoulli}(p)$ be random variables. Then, by the Cauchy-Schwarz inequality,

$$\mathbb{E}\left[XY\right] \leq \sqrt{\mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right]} = \sqrt{\lambda\left(\lambda+1\right)p}.$$

If the two random variables are independent, then

$$\mathbb{E}\left[XY\right] = \mathbb{E}X\mathbb{E}Y = \lambda p.$$

Indeed, you can check that for all $0 \leq p \leq 1$ and $\lambda \geq 0$,

$$\lambda p \leq \sqrt{\lambda\left(\lambda+1\right)p}.$$

2 Jensen's inequality

Next, we will cover Jensen's inequality, probably one of the most useful inequalities in probability theory! To discuss Jensen's inequality, we need to recall the notion of a convex function:

Definition 2.1. Let X be a subset of $\mathbb{R}^n$. We say X is *convex* if for all points $x, y \in X$ and $\theta \in [0,1]$, we have $\theta x + (1-\theta)y \in X$.

Definition 2.2. Let X be convex and $f\colon X \to \mathbb{R}$. We call the set

$$\operatorname{epi}(f) = \{(x,\mu) \in X \times \mathbb{R} \colon f(x) \leq \mu\}$$

the *epigraph* of f. We say f is a *convex function* if its epigraph is convex. We say f is *concave* if $-f$ is convex.

Intuitively, a convex function is one whose epigraph makes a "bowl":

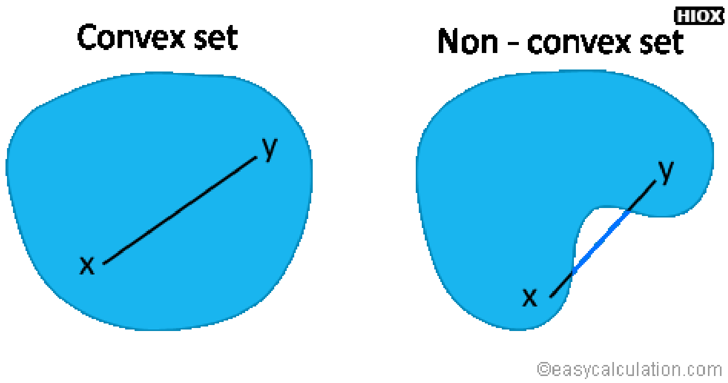


Figure 1: Examples of convex and non-convex sets

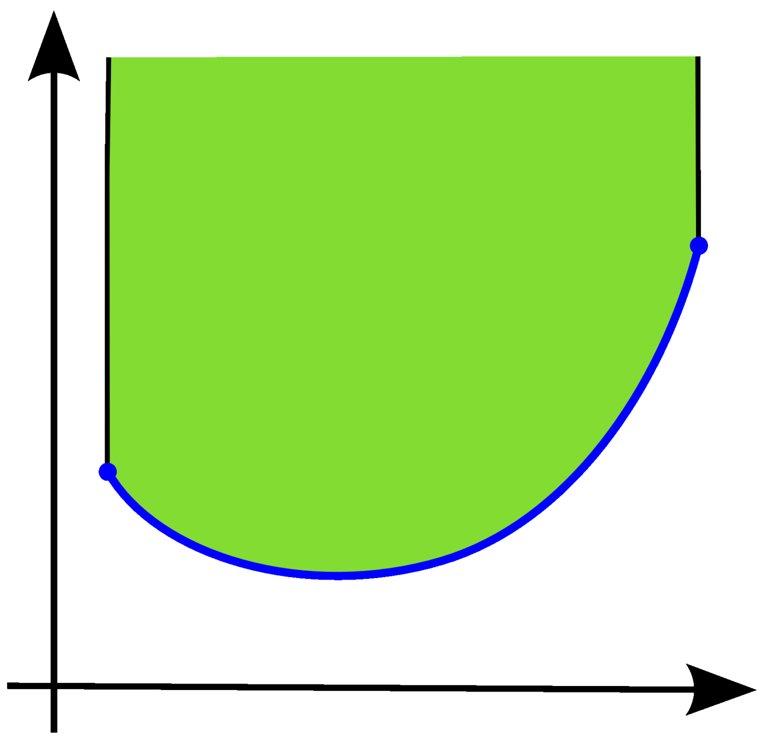


Figure 2: Epigraph (green) of a function f (blue curve)

Example 2.3. x, $|x|$, x^2, e^{-x} are convex on $\mathbb{R}$. The function f defined by

$$f(x) = \begin{cases} 1/x & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

is not convex on $\mathbb{R}$, but it is convex on $(-\infty, 0)$ and $(0, \infty)$.

Proposition 2.4. *Let X be convex and $f\colon X \to \mathbb{R}$. f is* a *convex function if and only if for all $x, y \in X$ and $\theta \in [0,1]$,*

$$f(\theta x + (1-\theta)\, y) \leq \theta f(x) + (1-\theta)\, f(y).$$

Proof. Suppose f is a convex function. Let $x, y \in X$ and $\theta \in [0,1]$. Note that $(x, f(x))$ and $(y, f(y))$ are both points in $\operatorname{epi}(f)$. By convexity,

$$\theta(x, f(x)) + (1-\theta)(y, f(y)) = (\theta x + (1-\theta)\, y, \theta f(x) + (1-\theta)\, f(y)) \in \operatorname{epi}(f)$$

and hence

$$f(\theta x + (1-\theta)\, y) \leq \theta f(x) + (1-\theta)\, f(y),$$

as desired.

Suppose f satisfies the convexity inequality. Let (x, μ_x) and (y, μ_y) be points in $\operatorname{epi}(f)$ and $\theta \in [0,1]$. By the convexity inequality,

$$f(\theta x + (1-\theta)\, y) \leq \theta f(x) + (1-\theta)\, f(y) \leq \theta \mu_x + (1-\theta)\, \mu_y$$

and hence

$$\theta(x, \mu_x) + (1-\theta)\,(y, \mu_y) = (\theta x + (1-\theta)\, y, \theta \mu_x + (1-\theta)\, \mu_y) \in \operatorname{epi}(f),$$

as desired. $\square$

Proposition 2.5. *A convex set in $\mathbb{R}$ is an interval.*

Proof. Suppose $X \subset \mathbb{R}$ is convex and not an interval. Let $x, z \in X$ and y be such that $x < y < z$. Pick

$$\theta = \frac{z-y}{z-x}.$$

Then,

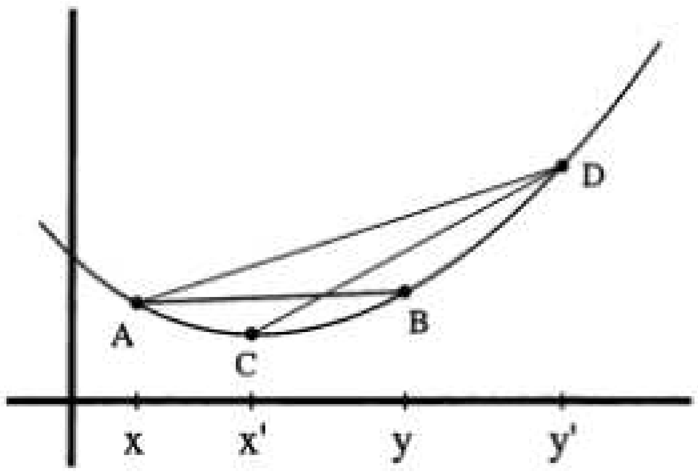
$$\theta x + (1-\theta)\, y = \frac{z-y}{z-x} x + \frac{y-x}{z-x} z = \frac{xz - xy}{z-x} + \frac{yz - xz}{z-x} = \frac{yz - xy}{z-x} = y.$$

$\square$

Proposition 2.6. *Let I be an interval and $f\colon I \to \mathbb{R}$. Then, f is convex if and only if for all points $x, y, x', y' \in I$ such that $x \leq x' < y'$ and $x < y \leq y'$,*

$$\frac{f(y) - f(x)}{y - x} \leq \frac{f(y') - f(x')}{y' - x'}.$$

Proof. Suppose f is convex and let $A = (x, f(x))$, $B = (y, f(y))$, $C = (x', f(x'))$, and $D = (y', f(y'))$. Then... (proof by picture)



For the converse, let $x_1, x_2 \in I$ and $\theta \in [0,1]$. Take $x = x_1$, $y' = x_2$, and $y = x' = \theta x_1 + (1-\theta)x_2$ to get

$$\frac{f(\theta x_1 + (1-\theta)x_2) - f(x_1)}{\theta x_1 + (1-\theta)x_2 - x_1} \leq \frac{f(x_2) - f(\theta x_1 + (1-\theta)x_2)}{x_2 - \theta x_1 - (1-\theta)x_2}$$

and hence

$$\theta\left(f(\theta x_1 + (1-\theta)x_2) - f(x_1)\right)(x_2 - x_1) \leq (1-\theta)\left(f(x_2) - f(\theta x_1 + (1-\theta)x_2)\right)(x_2 - x_1).$$

Simplifying,

$$f(\theta x_1 + (1-\theta)x_2) \leq \theta f(x_1) + (1-\theta)\, f(x_2). \qquad \square$$

Corollary 2.7. *Let $I = (a,b)$ be an open interval and $f\colon I \to \mathbb{R}$ be a convex function. Then, f is continuous and the left and right derivatives*

$$D_- f(x) = \lim_{h \downarrow 0} \frac{f(x) - f(x-h)}{h} \qquad \text{and} \qquad D_+ f(x) = \lim_{h \downarrow 0} \frac{f(x+h) - f(x)}{h}$$

exist at each point $x \in I$. Moreover, $D_- f$ and $D_+ f$ are nondecreasing with $D_- f \leq D_+ f$.

Proof. Let x be a point in I. By Proposition 2.6, for all $h > 0$ such that $x - h$ and $x + h$ are points in I,

$$\frac{f(x) - f(x-h)}{h} \leq \frac{f(x+h) - f(x)}{h}. \tag{1}$$

We would like to take limits and conclude

$$D_- f(x) = \lim_{h \downarrow 0} \frac{f(x) - f(x-h)}{h} \leq \lim_{h \downarrow 0} \frac{f(x+h) - f(x)}{h} = D_+ f(x).$$

But first, we have to show these limits exist: note that Proposition 2.6 implies that the left hand side of (1) increases while the right hand side of (1) decreases as h is made smaller. That is, for a decreasing sequence of $(h_n)_n$ with $h_n \downarrow 0$,

$$\frac{f(x) - f(x-h_1)}{h} \le \frac{f(x) - f(x-h_2)}{h} \le \cdots \le \frac{f(x+h_2) - f(x)}{h} \le \frac{f(x+h_1) - f(x)}{h}.$$

Then, the limits exist by the monotone convergence theorem (recall that the monotone convergence theorem for sequences says that if a sequence is nondecreasing and bounded above, it must have a limit). $\square$

Proposition 2.8. *Let $I = (a, b)$ be an open interval and $f\colon I \to \mathbb{R}$ be a convex function. Then, for each $x_0 \in I$, there exists m such that for all $x \in I$,*

$$f(x) \ge f(x_0) + m\,(x - x_0)\,.$$

That is, at each point x_0, there exists a ***supporting line****.*

This fact generalizes to higher dimensions, in which case the supporting line becomes a **supporting hyperplane**.

Proof. Choose m such that

$$D_-f(x_0) \le m \le D_+f(x_0).$$

Now, if $x > x_0$,

$$m \le D_+f(x_0) = \lim_{h\downarrow 0} \frac{f(x_0+h) - f(x_0)}{h} \le \frac{f(x) - f(x_0)}{x - x_0}$$

and hence

$$f(x_0) + m\,(x - x_0) \le f(x).$$

The case of $x < x_0$ is identical (use $D_-f(x_0)$ in the argument). $\square$

Proposition 2.9 (Jensen's inequality)**.** *Let $I = (a, b)$ be an open interval and $f\colon I \to \mathbb{R}$ be a convex function. Let X be a random variable which takes values in (a, b) a.s. If X and $f \circ X$ are both integrable,*

$$\mathbb{E}\,[f(X)] \ge f(\mathbb{E}X).$$

Proof. Let $x_0 = \mathbb{E}X$. Now, we can find some supporting line parameterized by m:

$$f(x) \ge f(x_0) + m(x - x_0).$$

Substitute $x = X$ to get

$$f(X) \ge f(x_0) + m(X - x_0)$$

(this inequality holds only a.s.). Take expectations of both sides to get

$$\mathbb{E}\,[f(X)] \ge \mathbb{E}\,[f(x_0) + m(X - x_0)] = \mathbb{E}\,[f(x_0)] = f(x_0) = f(\mathbb{E}X).$$ $\square$