Inequalities

Probability I (BST 230)

Jeffrey W. Miller

Department of Biostatistics
Harvard T.H. Chan School of Public Health

Outline

Outline

Introduction

Markov's inequality

Jensen's inequality

Introduction

- Earlier we saw Boole's inequality and Bonferroni's inequality.
- There are many useful inequalities in probability theory.
- Inequalities are useful because it is usually easier to bound some quantity of interest than to characterize it exactly.
- And often, a decent bound is all that is needed to show what you want to show.

Introduction

- For example, suppose you are manufacturing widgets.
- Each widget can be defective in one of three ways, denoted by events A_1, A_2, A_3.
- You have data on the probability of each type of defect, $P(A_k)$, but you don't have any data on the joint probability of these events.
- Fortunately, you can still bound the probability of any type of defect occurring by using Boole's inequality:

$$P(A_1 \cup A_2 \cup A_3) \leq P(A_1) + P(A_2) + P(A_3).$$

- If each $P(A_k)$ is small, then you can guarantee that the probability of any defect occuring is small.

Recall: Boole's and Bonferroni's inequalities

- *Boole's inequality (a.k.a. union bound)*: For any $A_1, A_2, \ldots$,

$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) \le \sum_{i=1}^{\infty} P(A_i).$$

- *Bonferroni's inequality*: For any $A_1, A_2, \ldots$,

$$P\Big(\bigcap_{i=1}^{\infty} A_i\Big) \ge 1 - \sum_{i=1}^{\infty} P(A_i^c).$$

Outline

Markov's inequality

- This is one of the simplest but most useful inequalities in probability theory.

- *Markov's inequality*: If X is a nonnegative random variable and $a > 0$, then
$$P(X \geq a) \leq \frac{\mathrm{E}X}{a}.$$

- Proof: Since $1 \geq \mathbb{1}(X \geq a)$,
$$\begin{aligned} \mathrm{E}X &\geq \mathrm{E}X\mathbb{1}(X \geq a) \\ &\geq \mathrm{E}a\mathbb{1}(X \geq a) \\ &= aP(X \geq a). \end{aligned}$$

  Dividing both sides by a yields the result.

Example: Investing returns

- You invest \$1000 dollars in a holding where the annual returns are $\mathrm{Pareto}(\alpha, c)$ distributed with $\alpha = 2$ and $c = 1/4$.

- More precisely, after n years, your investment is worth

$$Y_n = 1000 X_1 X_2 \cdots X_n$$

dollars, where $X_1, \ldots, X_n \sim \mathrm{Pareto}(\alpha, c)$ independently.

- Recall that the pdf of $\mathrm{Pareto}(\alpha, c)$ is

$$p(x) = \frac{\alpha c^\alpha}{x^{\alpha+1}} \mathbb{1}(x > c).$$

- Is this a good investment?

  Group exercise (5 minutes): First guess using your intuition. Then try to show something formally.

Corollaries of Markov's inequality

1. For any r.v. X and any $a > 0$,
$$P(|X| \geq a) \leq \frac{\mathrm{E}|X|}{a}.$$
2. For any r.v. X, any $a \in \mathbb{R}$, and any monotone increasing function $g(x) \geq 0$,
$$P(X \geq a) \leq \frac{\mathrm{E}g(X)}{g(a)}.$$
3. *Chebyshev's inequality*: For any r.v. X and any $a > 0$,
$$P\big(|X - \mathrm{E}X| \geq a\big) \leq \frac{\mathrm{Var}(X)}{a^2}.$$

   Chebyshev's allows us to bound the probability that a r.v. is a certain distance from its mean.

Group exercise (5 minutes): Try to show 2 and 3 using Markov's inequality.

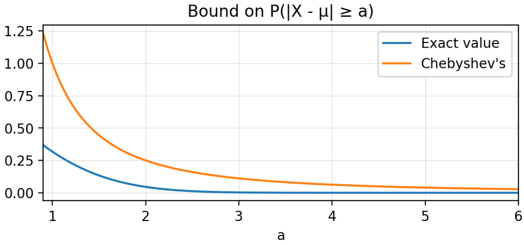
Example: Tail bounds for normal distributions

- Suppose $X \sim \mathcal{N}(\mu, \sigma^2)$ and we want to bound the probability that X is far from its mean.
- The exact expression involves the standard normal cdf $\Phi(x)$:

$$P\big(|X-\mu| \geq a\big) = P\left(\left|\frac{X-\mu}{\sigma}\right| \geq a/\sigma\right) = 2\Phi(-a/\sigma)$$

  for $a > 0$. However, $\Phi(x)$ does not have a simple closed form.
- Meanwhile, Chebyshev's inequality easily yields

$$P\big(|X-\mu| \geq a\big) \leq \frac{\mathrm{Var}(X)}{a^2} = \frac{\sigma^2}{a^2}.$$



Chernoff's bound

- This is surprisingly powerful corollary of Markov's inequality. It yields an exponentially decaying bound as a grows, compared to the $1/a$ in Markov's inequality.

- *Chernoff's bound*: For any r.v. X and any $a \in \mathbb{R}$,
$$P(X \geq a) \leq \inf_{t>0} e^{-ta} \, \mathrm{E} \exp(tX).$$

- Proof: For all $t > 0$,
$$\begin{aligned} P(X \geq a) &= P(tX \geq ta) \\ &= P(\exp(tX) \geq \exp(ta)) \\ &\leq \frac{\mathrm{E} \exp(tX)}{\exp(ta)} \\ &= e^{-ta} \, \mathrm{E} \exp(tX). \end{aligned}$$

  Since the left-hand side doesn't depend on t, the inequality holds when taking the infimum of the right-hand side over t.

Example: Tail bounds for normal distributions

- Suppose $X \sim \mathcal{N}(\mu, \sigma^2)$. By Chernoff's bound,

$$\begin{aligned} P(X - \mu \geq a) &\leq \inf_{t>0} e^{-ta}\, \mathrm{E} \exp\big(t(X-\mu)\big) \\ &= \inf_{t>0} e^{-ta} \exp\big(\tfrac{1}{2}\sigma^2 t^2\big) \\ &= \inf_{t>0} \exp\big(-ta + \tfrac{1}{2}\sigma^2 t^2\big) \end{aligned}$$

  using the formula for the mgf of $X - \mu \sim \mathcal{N}(0, \sigma^2)$.

- To minimize $f(t) = -ta + \frac{1}{2}\sigma^2 t^2$, we set

$$0 = f'(t) = -a + \sigma^2 t$$

  and solve to get $t = a/\sigma^2$. Plugging this in yields

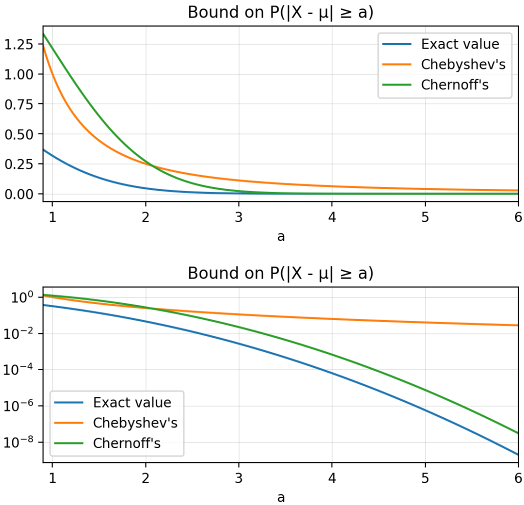
$$P(X - \mu \geq a) \leq \exp(-a^2/\sigma^2 + \tfrac{1}{2}a^2/\sigma^2) = \exp(-\tfrac{1}{2}a^2/\sigma^2).$$

- By symmetry, $P(-(X-\mu) \geq a) \leq \exp(-\frac{1}{2}a^2/\sigma^2)$. Thus,

$$P(|X - \mu| \geq a) \leq 2\exp(-\tfrac{1}{2}a^2/\sigma^2).$$

Example: Tail bounds for normal distributions

- Chebyshev's inequality: $P\big(|X-\mu| \geq a\big) \leq \sigma^2/a^2$.
- Chernoff's bound: $P(|X-\mu| \geq a) \leq 2\exp(-\frac{1}{2}a^2/\sigma^2)$.



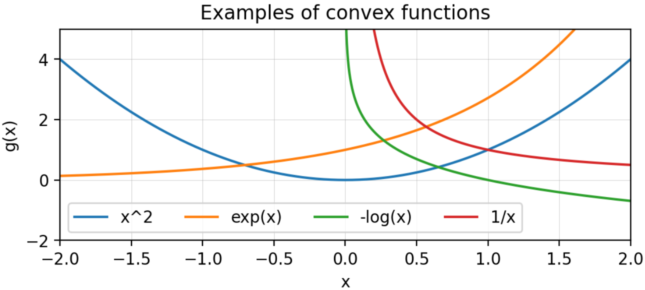
Outline

Convex functions

- A function $g : \mathcal{X} \to \mathbb{R}$ is *convex* if
$$g(tx + (1-t)y) \leq tg(x) + (1-t)g(y)$$
for all $x, y \in \mathcal{X}$ and all $t \in (0,1)$.

- A function $g : \mathcal{X} \to \mathbb{R}$ is *concave* if $-g$ is convex.

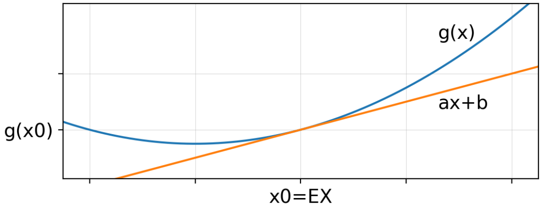
- Intuition: Convex functions curve upwards, concave functions curve downwards.



Jensen's inequality

- This is an extremely useful inequality in many situations.
- *Jensen's inequality*: Let X be a r.v. with range $\mathcal{X}$.
  If $g : \mathcal{X} \to \mathbb{R}$ is a convex function then

$$g(\mathrm{E}X) \leq \mathrm{E}g(X).$$



- Proof: Define $x_0 = \mathrm{E}X$. Since g is convex, there exist $a, b \in \mathbb{R}$ such that $ax + b \leq g(x)$ for all $x \in \mathcal{X}$ and $ax_0 + b = g(x_0)$. Therefore,

$$g(\mathrm{E}X) = g(x_0) = ax_0 + b = \mathrm{E}(aX + b) \leq \mathrm{E}g(X).$$

Jensen's inequality

- Examples of Jensen's inequality:
  - $|\mathrm{E}X| \leq \mathrm{E}|X|$.
  - $(\mathrm{E}X)^2 \leq \mathrm{E}X^2$.
  - $\exp(t\mathrm{E}X) \leq \mathrm{E}\exp(tX)$ for $t > 0$.
  - If $X > 0$ then $1/\mathrm{E}X \leq \mathrm{E}(1/X)$.
  - If $X > 0$ then $-\log(\mathrm{E}X) \leq -\mathrm{E}\log(X)$.

Hoeffding's inequality

- A particularly interesting use of Jensen's inequality is Hoeffding's inequality.

- *Hoeffding's inequality*: Suppose $X_1, \ldots, X_n$ are independent r.v.s such that $a_i \leq X_i \leq b_i$, and denote $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$. Then for all $a > 0$,

$$P(|\overline{X} - \mathrm{E}\overline{X}| > a) \leq 2\exp\left(-\frac{2a^2 n}{\frac{1}{n}\sum_{i=1}^n (b_i - a_i)^2}\right).$$

- Hoeffding's shows that for bounded random variables, the sample mean converges exponentially quickly to the mean.

- For instance, if $X_1, \ldots, X_n \sim \mathrm{Bernoulli}(q)$, then

$$P(|\overline{X} - q| > a) \leq 2\exp(-2a^2 n).$$