Probability Inequalities

Andrew Nobel

March, 2020

Review: Elementary Inequalities for Probability

Recall: If A, B are events, the axioms of probability ensure that

- If $A \subseteq B$ then $P(A) \leq P(B)$
- $P(A \cup B) \leq P(A) + P(B)$

Example: Let X, Y be random variables and $a, b > 0$

- $\mathbb{P}(|X + Y| \geq a + b) \leq \mathbb{P}(|X| \geq a) + \mathbb{P}(|Y| \geq b)$
- $\mathbb{P}(|XY| \geq a) \leq \mathbb{P}(|X| \geq a/b) + \mathbb{P}(|Y| \geq b)$

Overview of Probability Inequalities

Recall: For a random variable $X \in \mathbb{R}$

- $\mathbb{E}X$ tells us about the center of its distribution
- $\mathrm{Var}(X)$ tells us about the spread of its distribution

Goal: Upper bounds on the probability that X is far from $\mathbb{E}X$

$$\mathbb{P}(|X - \mathbb{E}X| \geq t) \qquad \mathbb{P}(X \geq \mathbb{E}X + t) \qquad \mathbb{P}(X \leq \mathbb{E}X - t)$$

Markov's Inequality: If $X \geq 0$ and $t > 0$ then

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}X}{t}$$

First Steps: Moment Based Inequalities

Chebyshev's Inequality: If X has $\mathbb{E}X^2 < \infty$ then for all $t > 0$

$$\mathbb{P}(|X - \mathbb{E}X| \geq t) \;\leq\; \frac{\mathrm{Var}(X)}{t^2}$$

Note: Upper bound may be larger than 1 (not useful)

Extension: Applying same idea we can show for each $t > 0$,

$$\mathbb{P}(|X - \mathbb{E}X| \geq t) \;\leq\; \min_{s>0} \frac{\mathbb{E}|X - \mathbb{E}X|^s}{t^s}$$

Smaller moments yield better upper bounds

Chebyshev and Averages

Given: $X_1, X_2, \ldots$ i.i.d with $\mathbb{E}X_i^2 < \infty$. For each $n \geq 1$ let

$$\overline{X}_n \; = \; \frac{1}{n}\sum_{i=1}^{n} X_i \; = \; \text{Average of } X_1, \ldots, X_n$$

It is easy to see that $\mathbb{E}(\overline{X}_n) = \mathbb{E}X$ and $\mathrm{Var}(\overline{X}_n) = n^{-1}\mathrm{Var}(X)$. Applying Chebyshev, we find that for each $t > 0$

$$\mathbb{P}\left(\left| \overline{X}_n - \mathbb{E}X \right| \geq t \right) \; \leq \; \frac{\mathrm{Var}(X)}{nt^2}$$

Note: Upper bound is non-trivial only when $t \geq \mathrm{SD}(X)/\sqrt{n}$

Weak Law of Large Numbers (WLLN)

Theorem (WLLN): For every $t > 0$, as n tends to infinity

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mathbb{E}X \right| \geq t \right) \to 0$$

The average of $X_1, \ldots, X_n$ *converges in probability* to $\mathbb{E}X$

Moment Generating Functions

Recall: Moment generating function (MGF) of r.v. X is defined by

$$M_X(s) = \mathbb{E}\left[e^{sX}\right] \quad \text{for } s \in \mathbb{R}$$

Note that $M_X(s) \geq 0$ and that $M_X(s)$ may be $+\infty$.

Fact: if $X_1, \ldots, X_n$ are independent and each MGF $M_{X_i}(s)$ is well defined in a neighborhood of 0 then $S_n = X_1 + \cdots + X_n$ has MGF

$$M_{S_n}(s) = \prod_{i=1}^{n} M_{X_i}(s)$$

MGF Examples

1. Normal. If $X \sim \mathcal{N}(0, \sigma^2)$ then $M_X(s) = e^{s^2\sigma^2/2}$
2. Poisson. If $X \sim$ Poiss(λ) then $M_X(s) = e^{\lambda(e^s-1)}$
3. Chi-squared. If $X \sim \chi^2_k$ then $M_X(s) = (1-2s)^{-k/2}$ for $s < 1/2$
4. Sign. If $X = 1, -1$ with probability $1/2$ then $M_X(s) = (e^s + e^{-s})/2$

Chernoff Inequality

Chernoff Bound: For any random variable X and $t \in \mathbb{R}$

$$\mathbb{P}(X \geq t) \;\leq\; \min_{s>0} e^{-st}\, \mathbb{E}e^{sX} \;=\; \min_{s>0} e^{-st}\, M_X(s)$$

Corollary: Suppose $M_{(X-\mathbb{E}X)}(s) \leq M(s)$ for $s \geq 0$. Then for each $t > 0$,

$$\mathbb{P}(X \geq \mathbb{E}X + t) \;\leq\; \inf_{s>0} e^{-st}\, M(s)$$

Inequalities for $\mathbb{P}(X \leq \mathbb{E}X - t)$, $\mathbb{P}(|X - \mathbb{E}X| \geq t)$ established in same way

Hoeffding's MGF Bound and Inequality

MGF bound: If $X \in [a, b]$ has $\mathbb{E}X = 0$ then for every $s \geq 0$

$$\mathbb{E}e^{sX} \leq e^{s^2(b-a)^2/8}$$

Hoeffding's Inequality: Let $X_1, \ldots, X_n$ be independent with $a_i \leq X_i \leq b_i$ and let $S_n = X_1 + \cdots + X_n$. For every $t \geq 0$,

$$\mathbb{P}(S_n - \mathbb{E}S_n \geq t) \;\leq\; \exp\left\{ \frac{-2t^2}{\sum_{i=1}^n (b_i - a_i)^2} \right\}$$

and also

$$\mathbb{P}(|S_n - \mathbb{E}S_n| \geq t) \;\leq\; 2\exp\left\{ \frac{-2t^2}{\sum_{i=1}^n (b_i - a_i)^2} \right\}$$

Example

Bernoulli: If $U_1, \ldots, U_n$ are iid $\sim$ Bern(p) then for each $t \geq 0$

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i - np \geq t\right) \leq \exp\left\{\frac{-2t^2}{n}\right\}$$

Corollary: If $X_1, \ldots, X_n \in \mathcal{X}$ iid $\sim P$ then for each $A \subseteq \mathcal{X}$ and $t \geq 0$

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(X_i \in A) - P(A)\right| \geq t\right) \leq 2\exp\left\{-2nt^2\right\}$$

Example

Uniform: Let $X_1, \ldots, X_n$ iid $\sim$ U$(-\theta, \theta)$. Note that $\mathbb{E}X = 0$ and $\mathrm{Var}(X) = \theta^2/3$. By Hoeffding's inequality, for $t \geq 0$

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq t\right) \leq \exp\left\{\frac{-t^2}{2n\theta^2}\right\}$$