Probability Inequalities

Andrew Nobel

October, 2021

Probability Inequalities

Elementary Inequalities for Probability

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

(a) $P(A^c) = 1 - P(A)$

(b) If $A \subseteq B$ then $P(A) \leq P(B)$

(c) $P(A \cup B) \leq P(A) + P(B)$

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

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

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

Concentration Inequalities

For a random variable X

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

Concentration Inequalities: Bounds on the probability that a random variable is far from its expectation

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

- Often $X = U_1 + \cdots + U_n$ sum of independent random variables
- Bounds depend on the moments (or MGF) of X
- Applications in statistics, machine learning, probability

Markov's and Chebyshev's Inequalities

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

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

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

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

- Upper bound may be larger than 1 (not useful)
- Upper bound is less than 1 if $t >$ SD(X)

Extending Chebyshev

Applying same proof idea we can show that for each $t > 0$,

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

Upshot: smaller central moments yield better upper bounds

Weak Law of Large Numbers (WLLN)

WLLN: Let $U_1, U_2, \ldots, U$ be iid with $\mathrm{Var}(U)$ finite. Then for each $t > 0$,

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

as n tends to infinity. In words, the average of $U_1, \ldots, U_n$ converges in probability to $\mathbb{E}(U)$ as n grows.

Proof: Apply Chebyshev's inequality to $X = n^{-1} \sum_{i=1}^{n} U_i$

Moment Generating Functions

Recall: The moment generating function (MGF) of a rv 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 $M_{X_i}(s)$ are finite in a neighborhood of 0 then $S_n = X_1 + \cdots + X_n$ has MGF

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

MGFs are a good way to study sums of independent random variables

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 $X \sim \chi_k^2$ 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's 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: If MGF of $(X - \mathbb{E}X)$ is at most $M(s)$ for $s \geq 0$, then for $t > 0$

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

- Inequalities for left tail $\mathbb{P}(X \leq \mathbb{E}X - t)$ established in same way
- Bound on $\mathbb{P}(|X - \mathbb{E}X| \geq t)$ obtained by adding L/R tail bounds

Hoeffding's MGF Bound and Hoeffding's Inequality

MGF bound: If $X \in [a, b]$ then for every $s \geq 0$

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

Probability 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 $t \geq 0$,

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

Also $\mathbb{P}(S_n - \mathbb{E}S_n \leq -t) \leq$ RHS and $\mathbb{P}(|S_n - \mathbb{E}S_n| \geq t) \leq$ 2 RHS

Note: Bound does not use information about variance of the X_is

Example: Bernoulli Random Variables

Let $X_1, \ldots, X_n$ be iid Bern(p). Note that $\mathbb{E}(\sum_{i=1}^n X_i) = np$

Chebyshev: Uses $\mathrm{Var}(X_i) = p(1-p)$. For each $t \geq 0$

$$\mathbb{P}\left(\sum_{i=1}^n X_i - np \geq t\right) \;\leq\; \frac{n\,p(1-p)}{t^2} \;\leq\; \frac{n}{4t^2}$$

Hoeffding: Uses $0 \leq X_i \leq 1$. For each $t \geq 0$

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

Note: Upper bounds useful only when $t \gtrsim \sqrt{n}$

Bernoulli Example, cont.

Compare bounds of Chebyshev and Hoeffding when $n = 100$

t	Chebyshev	Hoeffding
5	1	.607
10	.250	.135
12	.173	.0561
14	.128	.0198
16	.0977	.0060
20	.0625	.000335

Upshot: Once the bounds kick in, Hoeffding is better

Bernoulli Example, cont.

Bounds for sums can be converted into bounds for averages, and vice versa

Chebyshev: For each $t \geq 0$

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - p \geq t\right) \leq \frac{p(1-p)}{n\,t^2} \leq \frac{1}{4\,n\,t^2}$$

Hoeffding: For each $t \geq 0$

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

Note: Upper bounds useful only when $t \gtrsim 1/\sqrt{n}$

Other Examples of Hoeffding's Inequality

Ex: Let $X_1, \ldots, X_n \in \mathcal{X}$ be iid with distribution P and let $A \subseteq \mathcal{X}$. For $t \geq 0$,

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

Note that $n^{-1} \sum_{i=1}^{n} \mathbb{I}(X_i \in A)$ is the observed relative frequency of A, while $P(A)$ is its true probability

Ex: Let $X_1, \ldots, X_n$ iid $\sim$ U$(-\theta, \theta)$. Note that $\mathbb{E}X = 0$. For $t \geq 0$,

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