Lecture Note: Week 14

MATH 461: Probability Theory, Spring 2021
Daesung Kim

Lecture 31. Chebyshev inequality (Sec 8.2)

Markov's inequality
If X is a nonnegative random variable, then for any $a > 0$,

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Proof. Let $A = \{X \geq a\}$ and I_A be the indicator random variable, then $I_A \leq \frac{X}{a}$. Taking the expectation, we get the inequality. □

Chebyshev's inequality
If X is a random variable with mean μ and variance σ^2, then for any $a > 0$,

$$\mathbb{P}(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}.$$

Proof. Applying Markov's inequality for a nonnegative random variable $Y = (X - \mu)^2$, we have

$$\mathbb{P}(|X - \mu| \geq a) = \mathbb{P}(|X - \mu|^2 \geq a^2) \leq \frac{\mathbb{E}[|X - \mu|^2]}{a^2} = \frac{\sigma^2}{a^2}.$$

□

Example 1. Suppose that it is known that the number of items produced in a factory during a week is a random variable with mean 50.

(i) What can be said about the probability that this week's production will exceed 75?

(ii) If the variance of a week's production is known to equal 25, then what can be said about the probability that this week's production will be between 40 and 60?

One-sided Chebyshev's inequality
If X is a random variable with mean 0 and variance σ^2, then for any $a > 0$,

$$\mathbb{P}(X \geq a) \leq \frac{\sigma^2}{\sigma^2 + a^2}.$$

Proof. Applying Markov's inequality, we have

$$\mathbb{P}(X \geq a) = \mathbb{P}(X + t \geq a + t) \leq \mathbb{P}((X + t)^2 \geq (a + t)^2) \leq \frac{\mathbb{E}[(X + t)^2]}{(a + t)^2} = \frac{\sigma^2 + t^2}{(a + t)^2} =: f(t)$$

for any $t \in \mathbb{R}$. If $t = \frac{\sigma^2}{a}$, then

$$f(t) = \frac{\sigma^2}{\sigma^2 + a^2}, \qquad f'(t) = \frac{2(at - \sigma^2)}{(a+t)^3} = 0, \qquad f''(t) > 0$$

so that the inequality is obtained. $\square$

Remark 2. If we apply one-sided Chebyshev's inequality for $X - \mathbb{E}[X]$, then for $a > \mathbb{E}[X]$,

$$\mathbb{P}(X \geq a) \leq \frac{\mathrm{Var}(X)}{\mathrm{Var}(X) + (a - \mathbb{E}[X])^2}.$$

Example 3. If the number of items produced in a factory during a week is a random variable with mean 100 and variance 400, compute an upper bound on the probability that this week's production will be at least 120.

Chernoff inequality
If X is a random variable with mgf $M_X(t)$, then for any $a \in \mathbb{R}$,

$$\begin{aligned} \mathbb{P}(X \geq a) &\leq e^{-at} M_X(t), \qquad \text{for all } t > 0, \\ \mathbb{P}(X \leq a) &\leq e^{at} M_X(-t), \qquad \text{for all } t > 0. \end{aligned}$$

Proof. By Markov's inequality,

$$\mathbb{P}(X \geq a) = \mathbb{P}(e^{tX} \geq a^{at}) \leq \frac{\mathbb{E}[e^{tX}]}{e^{at}}$$

for $t > 0$ and

$$\mathbb{P}(X \leq a) = \mathbb{P}(e^{tX} \geq a^{at}) \leq \frac{\mathbb{E}[e^{tX}]}{e^{at}}$$

for $t < 0$, which completes the proof. $\square$

Example 4 (Upper tail for Poisson random variables)**.** Let X be a Poisson random variable with parameter λ. For any $a > 1$, we have

$$\mathbb{P}(X \geq a\lambda) \leq e^{-\lambda(a \log a - a + 1)}.$$

Example 5 (Upper tail for normal random variables)**.** Let X be the standard normal random variable. For any $a > 0$, we have

$$\mathbb{P}(X \geq a) \leq e^{-\frac{a^2}{2}}.$$

We say that a twice differentiable real-valued function $f(x)$ is convex (concave) on an interval I if $f''(x) \geq 0$ ($f''(x) \leq 0$) for all $x \in I$.

Jensen's inequality
If $f(x)$ is a convex function on an interval I and $\mathbb{P}(X \in I) = 1$, then

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

Proof. Since f is convex, there exists a linear function $\varphi(x) = ax + b$ such that $f(x) \geq \varphi(x)$ for all x and $\varphi(c) = f(c)$ where $c = \mathbb{E}[X]$. Then,

$$\mathbb{E}[f(X)] \geq \mathbb{E}[aX + b] = a\mathbb{E}[X] + b = \varphi(c) = f(\mathbb{E}[X]).$$

$\square$

Example 6. Let $a, b \in \mathbb{R}$ and f be a convex function on an interval I such that $a, b \in I$. Consider a random variable X such that $\mathbb{P}(X = a) = \mathbb{P}(X = b) = \frac{1}{2}$. Then, Jensen's inequality yields

$$\frac{f(a) + f(b)}{2} = \mathbb{E}[f(X)] \geq f(\mathbb{E}[X]) = f\left(\frac{a+b}{2}\right).$$

Example 7. Let X be a nonnegative random variable and $f(x) = x^k$ for $k \geq 1$. Since f is convex on $(0, \infty)$, we have

$$\mathbb{E}[X^k] \geq (\mathbb{E}[X])^k.$$

Remark 8. If f is concave on I, then $g(x) = -f(x)$ is convex. Applying Jensen's inequality for g, we get

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

Example 9. Let X be a nonnegative random variable and $f(x) = \log x$. Since f is concave on $(0, \infty)$, we have

$$\mathbb{E}[\log X] \leq \log(\mathbb{E}[X]).$$

Lecture 32. The Law of Large Numbers (Sec 8.2, 4)

Weak Law of Large Numbers (WLLN)
If $X_1, X_2, \cdots$ are independent, identically distributed random variables with mean μ, then for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| \geq \varepsilon\right) = 0.$$

Proof. Let $S_n = (X_1 + \cdots + X_n)/n$, then $\mathbb{E}[S_n] = \mu$ and $\mathrm{Var}(S_n) = \frac{\sigma^2}{n}$. By Chebyshev's inequality,

$$\mathbb{P}(|S_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \to 0.$$

as $n \to \infty$. □

Strong Law of Large Numbers (SLLN)
If $X_1, X_2, \cdots$ are independent, identically distributed random variables with mean μ, then with probability 1,

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to \mu.$$

What it means by "with probability 1" here is that

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{n} = \mu\right) = 1.$$

Idea of Proof. Assume that $\mathbb{E}[X_i^4] < \infty$ and $\mu = 0$. (This can be relaxed with truncation argument and centering.) Then, we claim that

$$\mathbb{E}\left[\sum_{n=1}^{\infty} \frac{S_n^4}{n^4}\right] < \infty.$$

This implies that $\sum_{n=1}^{\infty} \frac{S_n^4}{n^4} < \infty$ with probability 1. Since each summand is positive, $\frac{S_n}{n}$ converges to 0 with probability 1 as desired. □

Definition 10. Let $X_1, X_2, \cdots$ be a sequence of random variables and X a random variable. We say X_n converges to X in probability if for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0.$$

Definition 11. Let $X_1, X_2, \cdots$ be a sequence of random variables and X a random variable. We say X_n converges to X almost surely if

$$\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1.$$

Equivalently, for any $\varepsilon > 0$,

$$\lim_{m\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon \text{ for some } n \geq m) = 0.$$

Remark 12. What is the difference between weak and strong LLN? The WLLN says that S_n converges to μ in *probability*, whereas the SLLN says that $S_n \to \mu$ *almost sure*.

Remark 13. From the definitions, it is clear that almost sure convergence implies convergence in probability. Thus, the SLLN is *stronger* result than the WLLN. In general, convergence in probability does not implies almost sure convergence. Suppose U is a uniform random variable on $(0,1)$. Consider a sequence of functions $\varphi_1(x), \varphi_2(x), \cdots$ such that

$$\varphi_1(x) = 1_{(0,\frac{1}{2})}(x), \quad \varphi_2(x) = 1_{(\frac{1}{2},1)}(x), \quad \varphi_3(x) = 1_{(0,\frac{1}{3})}(x), \quad \varphi_4(x) = 1_{(\frac{1}{3},\frac{2}{3})}(x), \quad \cdots.$$

Define $X_i = \varphi_i(U)$ for $i \geq 1$. Then, it is easy to see that X_n converges to 0 in probability but not almost surely.

Lecture 33. Central Limit Theorem (Sec 8.3)

Central Limit Theorem (CLT)

Let $X_1, X_2, \cdots$ be independent, identically distributed random variables with mean μ and variance σ^2. Then,

$$S_n = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

converges to the standard normal random variable Z in distribution, meaning that

$$\lim_{n\to\infty} \mathbb{P}(S_n \leq a) = \mathbb{P}(Z \leq a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{x^2}{2}}\, dx = \Phi(a)$$

for any $a \in \mathbb{R}$.

Example 14. Let X_i be independent Bernouli random variables, then the CLT yields the normal approximation of binomial random variable.

Example 15. An astronomer is interested in measuring the distance, in light-years, from his observatory to a distant star. Although the astronomer has a measuring technique, he knows that because of changing atmospheric conditions and normal error, each time a measurement is made, it will not yield the exact distance, but merely an estimate. As a result, the astronomer plans to make a series of measurements and then use the average value of these measurements as his estimated value of the actual distance. If the astronomer believes that the values of the measurements are independent and identically distributed random variables having a common mean d (the actual distance) and a common variance of 4 (light-years), how many measurements need he make to be reasonably sure that his estimated distance is accurate to within $\pm.5$ light-year?

Example 16. The number of students who enroll in a psychology course is a Poisson random variable with mean 100. The professor in charge of the course has decided that if the number enrolling is 120 or more, he will teach the course in two separate sections, whereas if fewer than 120 students enroll, he will teach all of the students together in a single section. What is the probability that the professor will have to teach two sections?

References

[SR] Sheldon Ross, *A First Course in Probability*, 9th Edition, Pearson

Department of Mathematics, University of Illinois at Urbana-Champaign
E-mail address: daesungk@illinois.edu