

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. \square

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

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

\square

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

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

so that the inequality is obtained. \square

Example 2. 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 $a \in \mathbb{R}$,

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

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 3 (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 4 (Upper tail for normal random variables). Let X be the standard normal random variable. For any $a > 1$, 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 on an interval I if $f''(x) \geq 0$ for all $x \in I$.

Jensen's inequality

If $f(x)$ is a convex function, then

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

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

Weak Law of Large Numbers (WLLN)

If X_1, X_2, \dots are independent, identically distributed random variables with mean μ , then for any $\varepsilon > 0$,

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

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

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

as $n \rightarrow \infty$. □

Strong Law of Large Numbers (SLLN)

If X_1, X_2, \dots are independent, identically distributed random variables with mean μ , then with probability 1,

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

What it means by “with probability 1” here is that

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

Remark 5. What is the difference between weak and strong LLN? The difference is their convergence types. The WLLN says that S_n converges to μ in *probability*, whereas the SLLN says that $S_n \rightarrow \mu$ *almost sure*. We say that Y_n converges to a in probability if for any $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|Y_n - a| \geq \varepsilon) = 0,$$

and almost sure if

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} Y_n = a\right) = 1.$$

It is well-known that almost sure convergence implies the convergence in probability. However, the converse does not hold in general.

Lecture 33. Central Limit Theorem (Sec 8.3)

Central Limit Theorem (CLT)

Let X_1, X_2, \dots 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

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

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

Example 6. Let X_i be independent Bernoulli random variables, then the CLT yields the normal approximation of binomial random variable.

Example 7. 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 ± 0.5 light-year?

Example 8. 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