Lecture 33

Andrei Antonenko

April 20, 2005

1 Limit Theorems

Consider a sequence $X_1, X_2, \ldots$ of independent identically distributed random variables with mean μ and variance σ^2. Let

$$S_n = X_1 + \cdots + X_n \tag{1}$$

be the sum of the first n of them. Limit theorems are mostly concerned with the properties of S_n and related random variables, as n becomes very large. We have:

$$\mathbf{E}\left[S_n\right] = \mathbf{E}\left[X_1\right] + \cdots + \mathbf{E}\left[X_n\right] = n\mu. \tag{2}$$

Moreover, because of independence, we have

$$\mathbf{var}\left(S_n\right) = \mathbf{var}\left(X_1\right) + \cdots + \mathbf{var}\left(X_n\right) = n\sigma^2. \tag{3}$$

Thus, the distribution of S_n spreads out as n increases, and does not have a meaningful limit.

The situation is different if we consider the sample mean

$$M_n = \frac{X_1 + \cdots + X_n}{n} = \frac{S_n}{n}. \tag{4}$$

A quick calculation yields

$$\mathbf{E}\left[M_n\right] = \mu, \quad \mathbf{var}\left(M_n\right) = \mathbf{var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}. \tag{5}$$

In particular, the variance of M_n decreases to zero as n increases, and the bulk of its distribution must be very close to the mean μ. This phenomenon is the subject of certain laws of large numbers, which generally assert that the sample mean M_n (a random variable) converges to the true mean μ (a number), in a precise sense. These laws provide a mathematical basis for the loose interpretation of an expectation $\mathbf{E}\left[X\right] = \mu$ as the average of a large number of independent samples drawn from the distribution of X. We will also consider a quantity which is intermediate between S_n and M_n. We first subtract $n\mu$ from S_n, to obtain the zero-mean random variable $S_n - n\mu$ and then divide by $\sigma\sqrt{n}$, to obtain

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}. \tag{6}$$

We have:

$$\mathbf{E}\left[Z_n\right] = \frac{\mathbf{E}\left[S_n\right] - n\mu}{\sigma\sqrt{n}} = 0 \tag{7}$$

and

$$\mathbf{var}\left(Z_n\right) = \mathbf{var}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}}\right) = \frac{1}{\sigma^2 \cdot n} \cdot \mathbf{var}\left(S_n\right) = \frac{1}{\sigma^2 \cdot n} \cdot n\sigma^2 = 1. \tag{8}$$

Since the mean and the variance of Z_n remain unchanged as n increases, its distribution neither spreads, nor shrinks to a point. The central limit theorem is concerned with the asymptotic shape of the distribution of Z_n and asserts that it becomes the standard normal distribution. Limit theorems are useful for several reasons:

(a) Conceptually, they provide an interpretation of expectations (as well as probabilities) in terms of a long sequence of identical independent experiments.

(b) They allow for an approximate analysis of the properties of random variables such as S_n. This is to be contrasted with an exact analysis which would require a formula for the PMF or PDF of S_n, a complicated and tedious task when n is large.

2 Inequalities

2.1 Markov Inequality

If X is a nonnegative random variable, than the following inequality holds:

$$P(X \geq a) \leq \frac{\mathbf{E}\left[x\right]}{a} \qquad \text{for all } a \geq 0. \tag{9}$$

This inequality is called **Markov inequality**. The meaning of the Markov inequality is that the probability that the random variable will take the value far from it's mean is small.

Proof. Let's fix a positive number a and consider the random variable Y_a, defined as follows:

$$Y_a = \begin{cases} 0, & \text{if } X < a \\ a, & \text{if } X \geq a. \end{cases}$$

We see that always $Y_a \leq X$, and therefore

$$\mathbf{E}\left[Y_a\right] \leq \mathbf{E}\left[X\right].$$

But

$$\mathbf{E}\left[Y_a\right] = 0 \cdot P(Y_a = 0) + a \cdot P(Y_a = a) = a \cdot P(X \geq a),$$

and therefore

$$a \cdot P(X \geq a) \leq \mathbf{E}\left[X\right],$$

from where we get Markov inequality by dividing both parts by a. □

Example 2.1. Let X be uniformly distributed on the interval $[0, 4]$ and note that $\mathbf{E}\left[X\right] = 2$. Then, the Markov inequality asserts that

$$P(X \geq 2) \leq \frac{2}{2} = 1, \quad P(X \geq 3) \leq \frac{2}{3} = 0.67, \quad P(X \geq 4) \leq \frac{2}{4} = 0.5.$$

By comparing with the exact probabilities

$$P(X \geq 2) = 0.5, \quad P(X \geq 3) = 0.25, \quad P(X \geq 4) = 0,$$

we see that the bounds provided by the Markov inequality can be quite loose.

2.2 Chebyshev's Inequality

If X is a random variable with mean μ and variance σ^2, then

$$P(|X-\mu| \geq c) \leq \frac{\sigma^2}{c^2}, \text{for all } c > 0. \tag{10}$$

This inequality is called **Chebyshev's Inequality**.

Loosely speaking, asserts that if the variance of a random variable is small, then the probability that it takes a value far from its mean is also small.

Proof. To justify the Chebyshev inequality, we consider the nonnegative random variable $(X-\mu)^2$ and apply the Markov inequality with $a = c^2$. We obtain

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

The derivation is completed by observing that the event $(X-\mu)^2 \geq c^2$ is identical to the event $|X-\mu| \geq c$ and

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

□

Example 2.2. As in the previous example, let X be uniformly distributed on $[0,4]$. Let us use the Chebyshev inequality to bound the probability that $|X-2| \geq 1$. We have $\sigma^2 = 16/12 = 4/3$, and

$$P(|X-2| \geq 1) \leq \frac{4}{3},$$

which is not particularly informative.

Example 2.3. For another example, let X be exponentially distributed with parameter $\lambda = 1$, so that $\mathbf{E}[X] = \mathbf{var}(X) = 1$. For $c > 1$, using Chebyshevs inequality, we obtain

$$P(X \geq c) = P(X-1 \geq c-1) \leq P(|X-1| \geq c-1) \leq \frac{1}{(c-1)^2}.$$

This is again conservative compared to the exact answer

$$P(X \geq c) = e^{-c}.$$