Modes of convergence and the two big limit theorems

Jason Swanson

May 3, 2007

1 Almost sure convergence

The notation: "$Y_n \xrightarrow{\text{a.s.}} Y$" or "$Y_n \to Y$ a.s."

How to say it: "Y_n converges to Y almost surely."

The idea: The actual values of Y_n that you get when you perform your experiments will converge to the actual value of Y.

The definition: We say that $Y_n \xrightarrow{\text{a.s.}} Y$ if $P\left(\lim_{n\to\infty} Y_n = Y\right) = 1$.

2 Convergence in probability

The notation: "$Y_n \xrightarrow{\mathcal{P}} Y$" or "$Y_n \to Y$ in probability."

How to say it: "Y_n converges to Y in probability."

The idea: If n is large, then the probability that the actual values of Y_n are not close to the actual value of Y is very small.

The definition: We say that $Y_n \xrightarrow{\mathcal{P}} Y$ if, for all $\varepsilon > 0$, $\lim_{n\to\infty} P\left(|Y_n - Y| > \varepsilon\right) = 0$.

3 Convergence in distribution

The notation: "$Y_n \xrightarrow{\mathcal{D}} Y$" or "$Y_n \Rightarrow Y$" or "$Y_n \to Y$ in distribution."

How to say it: "Y_n converges to Y in distribution."

The idea: The actual values of Y_n may not be close to the actual value of Y at all. But the distribution functions of Y_n get close to the distribution function of Y.

The definition: Let $F_n(y) = P(Y_n \le y)$ and $F(y) = P(Y \le y)$. We say that $Y_n \xrightarrow{\mathcal{D}} Y$ if $\lim_{n\to\infty} F_n(y) = F(y)$ for all y such that F is continuous at y.

4 The relationship between the three

Almost sure convergence implies convergence in probability, and convergence in probability implies convergence in distribution. In other words, if $Y_n \xrightarrow{\text{a.s.}} Y$, then $Y_n \xrightarrow{\mathcal{P}} Y$. And if $Y_n \xrightarrow{\mathcal{P}} Y$, then $Y_n \xrightarrow{\mathcal{D}} Y$. None of the reverse implications are true, in general.

5 The law of large numbers

Let $X_1, X_2, \ldots$ be iid random variables with a (finite) mean μ. Let $\overline{X}_n = \frac{1}{n}\sum_{j=1}^n X_j$. The weak law of large numbers says that $\overline{X}_n \xrightarrow{\mathcal{P}} \mu$. The strong law of large numbers says that $\overline{X}_n \xrightarrow{\text{a.s.}} \mu$. Here is the formal statement of the weak law of large numbers.

Theorem 5.1 *Let $X_1, X_2, \ldots$ be a sequence of iid random variables with a (finite) mean $\mu = E[X_1]$. Then for any $\varepsilon > 0$,*

$$P\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| > \varepsilon\right) \to 0$$

as $n \to \infty$.

And here is the formal statement of the strong law of large numbers.

Theorem 5.2 *Let $X_1, X_2, \ldots$ be a sequence of iid random variables with a (finite) mean $\mu = E[X_1]$. Then*

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

6 The central limit theorem

Consider for the moment an example that has nothing to do with probability. You should know from calculus that

$$\lim_{n\to\infty} \cos(1/n) = 1.$$

So when n is large, $\cos(1/n) \approx 1$. This is what is sometimes called a "first-order approximation." If we wanted to more precise than this, then we might calculate (using L'Hôpital's rule) that

$$\lim_{n\to\infty} n^2(\cos(1/n) - 1) = -1/2.$$

In other words, when n is large, $n^2(\cos(1/n) - 1) \approx -1/2$. Doing a little algebra on this tells us that $\cos(1/n) \approx 1 - 1/(2n^2)$. This is a "second-order approximation." It is more accurate than the first-order approximation.

Now, according to the law of large numbers, $\overline{X}_n \xrightarrow{\text{a.s.}} \mu$ as $n \to \infty$. In other words, when n is large, $\overline{X}_n \approx \mu$. This is a first-order approximation. The central limit theorem gives us a second-order approximation. The central limit theorem says that

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{\mathcal{D}} \sigma Z,$$

where $\sigma^2 = \mathrm{Var}(X_1)$ and $Z \sim N(0,1)$. In other words, when n is large,

$$\overline{X}_n \overset{\mathcal{D}}{\approx} \mu + (\sigma/\sqrt{n})Z.$$

The symbol "$\overset{\mathcal{D}}{\approx}$" means that the distribution function of the random variable on the left is approximately equal to the distribution function of the random variable on the right. This is a second-order approximation.

We can rewrite $\sqrt{n}(\overline{X}_n - \mu)$ as

$$\begin{aligned}
\sqrt{n}(\overline{X}_n - \mu) &= \sqrt{n}\left(\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right) \\
&= \sqrt{n}\left(\frac{X_1 + X_2 + \cdots + X_n - n\mu}{n}\right) \\
&= \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sqrt{n}}
\end{aligned}$$

In other words, the central limit theorem says that

$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\mathcal{D}} Z.$$

Here is the formal statement of the central limit theorem.

Theorem 6.1 *Let $X_1, X_2, \ldots$ be a sequence of iid random variables with a (finite) mean $\mu = E[X_1]$ and a (finite) variance $\sigma^2 = \mathrm{Var}(X_1)$. Then, for all real numbers a,*

$$P\left(\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq a\right) \to \Phi(a)$$

as $n \to \infty$, where Φ is the cumulative distribution function of a standard normal random variable.