Convergence

1. Overview

This note will discuss the probability concepts associated with "convergence," along with the applications and results that these concepts permit.

The first section dicussions convergence *in probability* (also known as "weak convergence") before moving onto the stronger concept of *almost sure convergence* (also known as "strong convergence"). It concludes with convergence *in distribution*.

The remainder of the note uses these concepts to define the famous *Law of Large Numbers* (LLN), which gives the conditions under which sample moments converge to population moments as $n \to \infty$. Both the weak and strong versions will be covered, which will each use the concepts of weak convergence and strong convergence, respectively, discussed above and defined below.

Beyond that, the *Central Limit Theorem* (CLT) provides a refinement of the LLN via the concept of convergence in distribution. The CLT describes the *rate at which* sample moments converge to population moments as $n \to \infty$.

2. Types of Convergence

2.1. Preliminary Definition

A sequence $\{x_n\}$ has a limit x, written

$$\lim_{n\to\infty} x_n = x \tag{1}$$

if $\forall \epsilon > 0$, there exists an $n_\epsilon < \infty$ such that for all $n \geq n_\epsilon$,

$$|x_n - x| \leq \epsilon$$

In words: "You tell me how arbitrarily close x_n should be to x. I'll tell you an index n_ϵ, past which, that will happen."

But this, of course, is for a sequence of numbers, $\{x_n\}$. There's no randomness there. So what about a sequence of Random Variables—something non-deterministic, like an average $\bar{X}$? For that, we turn to the topics of the next few subsections.

2.2. Convergence in Probability

Consider a sequence of random variables $\{X_n\}$, each with corresponding distribution function F_n. Then a random variable X_n *converges in probability* to X if

$$\lim_{n\to\infty} \Pr\left(|X_n - X| \leq \epsilon\right) = 1 \qquad \forall \epsilon > 0 \tag{2}$$

This is denoted $\text{plim}_{n\to\infty} X_n = X$ or $(X_n \xrightarrow{p} X)$, while X is called the *probability limit* (or *plim*) of X_n. Convergence in probability is also known as *weak convergence*.

To get the intuition, consider the Definition given in 2. Notice that it uses the traditional definition of a limit, but applied to a sequence of *probabilities*. It does **not** say that realizations equal the plim (i.e. $X_n = X$) as $n \to \infty$. Instead, it describes the distribution of $|X_n - X|$ and stipulates that the realizations cluster very close to X as $n \to \infty$.

Now for some final notes. Convergence in probability is **not** convergence in expectation. The former concerns a sequence of probabilities, while the latter a sequence of expectations. Finally, the probability limit X must be free of all dependence upon the sample size n.

2.3. Almost Sure Convergence

Now, we turn to a concept stronger than convergence in probability, *almost sure convergence*—also known as "strong convergence."[1] A random variable, X_n converges *almost surely* to X if

$$\Pr\left(\lim_{n\to\infty} |X_n - X|\right) = 1 \qquad \forall \epsilon > 0 \tag{3}$$

We denote this form of convergence by $X_n \xrightarrow{a.s.} X$. It is stronger than convergence in probability because it computes the probability of a limit, rather than the limit of a probability.

2.4. Relationships

The concepts just defined are related in the following way:

- Almost sure $\Rightarrow$ In Probability.
- In Probability $\Rightarrow$ there's a deterministic subsequence that converges Almost Surely.
- In p-norm $\Rightarrow$ In Probability.
- Almost Surely and In p-norm, undecidable.
- Almost Surely, In Probability, and In p-norm each $\Rightarrow$ In Distribution.

[1] In probability terminology, a random event which occurs with probability one is called "almost sure."

3. Law of Large Numbers (LLN)

We now use the concept of convergence in probability along with the following relation, *Chebyshev's Inequality*, to define the Weak LLN:

$$\Pr\left(|X_n - \mathbb{E}X_n| > \delta\right) \leq \frac{\mathrm{Var}(X_n)}{\delta^2} \tag{4}$$

See the appendix for a proof of the inequality, which is refreshingly simple given how useful Chebyshev's Inequality is.

4. Related Concepts

Consistency A sequence of estimators $\{\hat{\theta}_n\}$ where $n = 1, 2, \ldots$ is *consistent* for parameter θ if $\hat{\theta}_n$ converges *In Probability* to θ.

Strongly Consistent If convergence of $\hat{\theta}_n$ to θ holds with probability 1.

A. Proof of Chebyshev's Inequality

Recall that we want to prove

$$\Pr\left(|X_n - \mathbb{E}X_n| > \delta\right) \leq \frac{\mathrm{Var}(X_n)}{\delta^2} \tag{5}$$

Proof. Assume that X has finite variance, σ^2, and let $f_X(x)$ denote the density function for X. Then

$$\Pr\left(|X - \mathbb{E}X| > \delta\right) = \Pr\left((X - \mathbb{E}X)^2 > \delta^2\right) \tag{6}$$

We now define a new variable Z, with density function $f_Z(z)$:

$$Z = (X - \mathbb{E}X)^2 \tag{7}$$

Note that, by definition,

$$\mathbb{E}Z = \mathbb{E}\left(X - \mathbb{E}X\right)^2 = \mathrm{Var}(X)$$

Using this fact, we now write out Equation 6 more explicitly in terms of the density function for Z:

$$\begin{aligned}\Pr\left((X - \mathbb{E}X)^2 > \delta^2\right) &= \Pr\left(Z > \delta^2\right) \\ &= \int_{\delta^2}^{\infty} f_Z(z)\, dz \end{aligned} \tag{8}$$

Now because our support is $(\delta^2,\ \infty)$, it's clear that $z > \delta^2$ over that support. This allows us to deduce

$$z > \delta^2 \quad \Rightarrow \quad 1 \leq \frac{z}{\delta^2}$$

We can now use this inequality along with Equation 8:

$$\begin{aligned}\int_{\delta^2}^{\infty} 1 \cdot f_Z(z)\, dz &\leq \int_{\delta^2}^{\infty} \frac{z}{\delta^2}\, f_Z(z)\, dz \\ &\leq \frac{1}{\delta^2} \int_{\delta^2}^{\infty} z \cdot f_Z(z)\, dz \\ \text{From Eq. 7} \qquad &\leq \frac{1}{\delta^2}\mathbb{E}Z = \frac{\mathrm{Var}(X)}{\delta^2}\end{aligned}$$

$$\Rightarrow \quad \Pr\left((X - \mathbb{E}X)^2 > \delta^2\right) \leq \frac{\mathrm{Var}(X)}{\delta^2}$$

which is exactly what we want.

□