Convergence of Random Variables

1 Review of Definitions

Let \( X_i, i = 1, \ldots \), be a collection of random variables. The sample space on which \( X_i \) is defined will be denoted by \( \Omega_i \). Let \( X \) be a random variable on a sample space \( \Omega \). We will consider ways to make meaning of the statement “\( X_i \) converges to \( X \).”

The two following definitions assume \( \Omega = \Omega_1 = \Omega_2 = \cdots \).

**Almost sure convergence.** We will say that \( X_i \) converges to \( X \) almost surely if \( X_i(\omega) \) approaches \( X(\omega) \) for all \( \omega \in \Omega \), except possibly in a set of measure zero.

**Convergence in probability.** We will say that \( X_i \) converges to \( X \) in probability if \( P(|X_i - X| > \epsilon) \) approaches 0 as \( i \) goes to infinity, for any \( \epsilon > 0 \).

The next definition does not require \( \Omega_i \) to be identical.

**Convergence in distribution.** We will say that \( X_i \) converges to \( X \) in distribution if the function \( F_{X_i} \) converges to the function \( F_X \) at all points where \( F_X \) is continuous.

2 The relationship between convergence almost surely and convergence in probability

**Theorem.** Suppose \( X_i \) converges to \( X \) almost surely. Then, \( X_i \) converges to \( X \) in probability.

**Proof.** Fix \( \epsilon > 0 \). Define \( A_n(\epsilon) \) to be the set where \( X_n \) differs from \( X \) by at least \( \epsilon \):

\[
A_n(\epsilon) = \{ w \in \Omega : |X_n(w) - X(w)| > \epsilon \}
\]
Let $A(\epsilon)$ be the set of $\omega$ which are in some $A_n(\epsilon)$ infinitely often:

$$A(\epsilon) = \cap_{k=1}^{\infty} \cup_{n=k}^{\infty} A_n(\epsilon).$$

If $\omega \in A(\epsilon)$, then $X_n(\omega)$ cannot converge to $X(\omega)$; this means that $A(\epsilon)$ is a subset of a set of measure 0, and therefore

$$P(A(\epsilon)) = 0.$$  

However, $A(\epsilon)$ is the intersection of a decreasing sequence of sets; applying the continuity of probability,

$$\lim_{k \to \infty} P(\cup_{n=k}^{\infty} A_n(\epsilon)) = 0$$

Since $A_k \subset \cup_{n=k}^{\infty} A_n(\epsilon)$, this implies

$$\lim_{k \to \infty} P(A_k(\epsilon)) = 0,$$

which means that $X_k$ converges to $X$ in probability. \(\square\)

**Remark:** The converse of the above theorem is not true. Suppose $X_i$ converges to $X$ in probability. It may be that $X_i$ does not approach $X$ almost surely.

Indeed, let $X_n$ be the random variable which takes value 1 with probability $1/n$, and value 0 with probability $1-1/n$. Let $X$ be the random variable that identically zero. We have that $X_n$ converges to $X$ in probability:

$$P(|X_n - X| > \epsilon) \leq \frac{1}{n},$$

for any positive $\epsilon$. As $n$ approaches infinity, $P(|X_n - X| > \epsilon)$ will approach zero.

On the other hand, by the Borel-Cantelli lemma, $X_n = 1$ infinitely often with probability 1, so that $P(A(\epsilon)) = 1$ for any $\epsilon$. If $X_n$ approached $X$ almost surely, then we would have $P(A(\epsilon)) = 0$.

3 The relationship between convergence in probability and convergence in distribution

**Theorem.** Suppose $X_i$ converges to $X$ in probability. Then $X_i$ converges to $X$ in distribution.
**Proof:** Let $F_i(x)$ denote the distribution function of $X_i$ and $F(x)$ denote the distribution function of $X$. We can write

\[
F_n(x) = P(X_n \leq X) = P(X_n \leq X, X \leq x + \epsilon) + P(X_n \leq x, X > x + \epsilon)
\leq F(x + \epsilon) + P(|X_n - X| > \epsilon).
\]

This inequality holds for all $n$ and $\epsilon$. It gives us an upper bound on $F_n$ in terms of $F$. To obtain a lower bound, we argue as:

\[
F(x - \epsilon) = P(X \leq x - \epsilon) = P(X \leq x - \epsilon, X_n \leq x) + P(X \leq x - \epsilon, X_n > x)
\leq F_n(x) + P(|X_n - X| > \epsilon)
\]

The last part can be rewritten as

\[
F_n(x) \geq F(x - \epsilon) - P(|X_n - X| > \epsilon).
\]

Let us now combine the upper and lower bounds:

\[
F(x - \epsilon) + P(|X_n - X| > \epsilon) \leq F_n(x) \leq F(x + \epsilon) + P(|X_n - X| > \epsilon).
\]

Again, note this equation holds for all $\epsilon$ and for all $n$. Let us take the limit of both sides as $n$ approaches infinity, and then as $\epsilon \to 0$; we obtain that if $F$ is continuous at $x$, then

\[
\lim_{n} F_n(x) = F(x).
\]

\[\square\]

**Remark:** The converse of this theorem does not hold. Indeed, even assuming $X_i$ approach $X$ in distribution, they may not even be defined on the same space.

We can, however, refine the question as follows. Suppose $X_i$ approach $X$ in distribution and $\Omega = \Omega_1 = \Omega_2 = \cdots$. Will it always be true that $X_i$ approach $X$ in probability?

The answer is no. This was discussed in class: suppose $X, X_1, X_2, \ldots$ are all independent $N(0, 1)$ Gaussians. Certainly, $X_i$ converges to $X$ in distribution, since all the distributions are equal. However, $X_i - X = N(0, 2)$, which does not become concentrated around 0 as $i$ grows.
4 Some special cases

We now catalog some special cases when stronger statements can be made about the relationship between various types of convergence.

**Theorem:** Suppose $X_i$ converges to $X$ in probability. Then there exists a sequence of integers $n_1, n_2, \ldots$ such that $X_{n_i}$ converges to $X$ almost surely.

**Proof:** We know that $P(|X_k - X| > \frac{1}{i})$ approaches 0 as $k$ approaches $\infty$; pick $n_i$ with the property that

$$P(|X_{n_i} - X| > \frac{1}{i}) < \frac{1}{i^2}.$$ 

Let $A_i$ be the event that $|X_{n_i} - X| > 1/i$ and let $A$ be the event “$A_i$ occurs infinitely often.” Note that $X_{n_i}$ converges to $X$ on $A^c$. But the Borel-Cantelli lemma says that the probability of $A$ is zero.

**Theorem:** Suppose $X_i$ converges to a constant $c$ in distribution. Then, $X_i$ converges to $X$ in probability.

**Remark:** Observe that since the constant random variable can be defined on any space, we do not run into problems when writing expressions like $P(|X_i - c| > \epsilon)$.

**Proof:** We have that

$$P(|X_i - c| > \epsilon) = P(X_i > c + \epsilon) + P(X_i < c - \epsilon) \leq (1 - F_i(c + \epsilon)) + F_i(c - \epsilon).$$

We know that $F_i(x)$ converges to the function $1_{[c, +\infty)}(x)$ for all $x \neq c$. This means that $F_i(c + \epsilon)$ approaches 1 and $F_i(c - \epsilon)$ approaches 0 as $i$ approaches infinity. Thus $P(|X_i - c| > \epsilon)$ is sandwiched between 0 and a sequence that approaches 0 as $i$ approaches infinity; therefore, it must approach zero.
6.436J / 15.085J Fundamentals of Probability
Fall 2008

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.