1 Convergence of random variables

We discuss here two notions of convergence for random variables: convergence in probability and convergence in distribution.

1.1 Convergence in Probability

We begin with a very useful inequality.

Proposition 1 (Markov's Inequality). *Let X be a non-negative random variable, that is, $P(X \geq 0) = 1$. Then*

$$P(X \geq c) \leq \frac{1}{c} E(X) . \tag{1}$$

Proof. We have

$$E(X) \geq E(X \mathbf{1}\{X \geq c\}) \geq cE(\mathbf{1}\{X \geq c\}) = cP(X \geq c) .$$

□

Corollary 2 (Chebyshev's Inequality). *Let X be a random variable with mean μ and variance σ^2. Then*

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

Proof. Notice that the random variable $(X - \mu)^2$ is non-negative and has expectation equal to σ^2. Thus applying Proposition 1 to this random variable yields

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

□

A random variable is *concentrated* about its mean if the probability of being far away from its mean is small. Chebyshev's inequality says that if the variance of a random variable is small, then the random variable is concentrated about its mean. This is a very important property, especially if we are using X as an estimator of $E(X)$. In the case of a random variable with small variance, it is a good estimator of its expectation.

In many situations, we have a sequence of random variables $X_1, X_2, \ldots$, often denoted by $\{X_n\}$, which as n becomes large, become more and more concentrated around a common mean value.

Theorem 3 (Weak Law of Large Numbers). *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables, each with mean μ and variance σ^2. Then for every $\epsilon > 0$,*

$$\lim_{n\to\infty} P\left(\left|n^{-1}\sum_{i=1}^{n} X_i - \mu\right| > \epsilon\right) = 0\,. \tag{3}$$

Proof. Notice that

$$E\left(n^{-1}\sum_{i=1}^{n} X_i\right) = n^{-1}\sum_{i=1}^{n} E(X_i) = \mu$$
$$\mathrm{Var}\left(n^{-1}\sum_{i=1}^{n} X_i\right) = n^{-2}\sum_{i=1}^{n} \mathrm{Var}(X_i) = \frac{\sigma^2}{n}\,.$$

Applying Corollary 2 to the random variable $n^{-1}\sum_{i=1}^{n} X_i$ yeilds

$$P\left(\left|n^{-1}\sum_{i=1}^{n} X_i - \mu\right| > \epsilon\right) \le \frac{1}{\epsilon^2}\frac{\sigma^2}{n}\,.$$

Letting $n \to \infty$ in the above inequality proves the Theorem. □

The interpretation of the Weak Law of Large Numbers is the following: If we average a large number of independent random variables, each with the same expecatation μ, then this average is with high probability very close to μ.

For example, if with toss a coin a large number of times, then the percentage of these tosses which will land "heads" is with large probability close to 1/2, for a fair coin.

The Weak Law of Large of Numbers gives an example where a sequence of random variables *converges in probability*:

Definition 1. Let $\{X_n\}$ be a sequence of random variables, and let X be a random variables. Then $\{X_n\}$ is said to *converge in probability* to X if for every $\epsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0\,. \tag{4}$$

We write $X_n \xrightarrow{\text{Pr}} X$ to indicate convergence in probability. Thus, the Weak Law says that $n^{-1}\sum_{i=1}^n X_i$ converges in probability to μ, provided $\{X_i\}$ is a sequence of i.i.d. random variables with expectation μ.

1.2 Convergence in Distribution

Example 1. Let X_n be a Geometric random variable with parameter $p = \lambda n^{-1}$. Consider the random variable $n^{-1}X_n$. Notice that $E(n^{-1}X_n) = \lambda^{-1}$, which does not depends on n.

Let us compute $1 - F_n(t)$, where F_n is the distribution function of $n^{-1}X_n$.

$$\begin{aligned}
P\big(n^{-1}X_n > t\big) &= P(X_n > tn) \\
&= P(X_n > \lfloor tn \rfloor) \\
&= \left(1 - \frac{\lambda^{-1}}{n}\right)^{\lfloor tn \rfloor} \\
&= \left(1 + \frac{-\lambda^{-1}}{n}\right)^{n(\lfloor tn \rfloor/n)} \\
\log P\big(n^{-1}X_n > t\big) &= \frac{\lfloor tn \rfloor}{n} \log \left(1 + \frac{-\lambda^{-1}}{n}\right)^{n} .
\end{aligned}$$

Thus,

$$\begin{aligned}
\lim_{n\to\infty} \log P\big(n^{-1}X_n > t\big) &= -\lambda^{-1}t \\
\lim_{n\to\infty} P\big(n^{-1}X_n > t\big) &= e^{-\lambda^{-1}t} .
\end{aligned}$$

We conclude that for any t,

$$\lim_{n\to\infty} F_n(t) = F(t) ,$$

where $F(t)$ is the distribution function for an Exponential random variable with mean λ^{-1}. This is useful, because then we can approximate any probability involving $n^{-1}X_n$ for large values of n:

$$P\big(a < n^{-1}X_n \leq b\big) \approx \int_a^b \frac{1}{\lambda} e^{-\frac{x}{\lambda}} dx ,$$

where the error in the approximation goes to 0 as $n \to \infty$.

This motivates the following definition:

Definition 2. Let $\{X_n\}$ be a sequence of random variables, and let X be a random variable. Suppose that X_n has distribution function F_n, and X has distribution function X. We say that $\{X_n\}$ *converges in distribution* to the random variable X if

$$\lim_{n\to\infty} F_n(t) = F(t)\,,$$

at every value t where F is continuous. We write $X_n \xrightarrow{\mathrm{d}} X$ to indicate convergence in distribution.

It should be clear what we mean by $X_n \xrightarrow{\mathrm{d}} F$: the random variables X_n converge in distribution to a random variable X having distribution function F. Similarly, we have $F_n \xrightarrow{\mathrm{d}} F$ if there is a sequence of random variables $\{X_n\}$, where X_n has distribution function F_n, and a random variable X having distribution function F, so that $X_n \xrightarrow{\mathrm{d}} X$.

An example explains why we require that the distribution functions converge only at continuity points for the limiting distribution function.

Example 2. Let $X_n = 1 + \frac{1}{n}$ be a constant random variable. Then X_n has the distribution function

$$F_n(t) = \begin{cases} 0 & \text{if } t < 1 + \frac{1}{n} \\ 1 & \text{if } t \geq 1 + \frac{1}{n}\,. \end{cases}$$

Notice that $\lim_n F_n(t) = \tilde{F}(t)$, where

$$\tilde{F}(t) = \begin{cases} 0 & \text{if } t \leq 1 \\ 1 & \text{if } t > 1\,. \end{cases}$$

Notice that $\tilde{F}$ is not a distribution function, as it is not right-continuous at $t = 1$. However, a modification of $\tilde{F}$ is a distribution function: let

$$F(t) = \begin{cases} 0 & \text{if } t < 1 \\ 1 & \text{if } t \geq 1\,. \end{cases}$$

Notice that F is the distribution function for the random variable $X = 1$. Thus $X_n \xrightarrow{\mathrm{d}} X$, even though $F_n(1) \not\to F(1)$. This is fine, because the definition of convergence in distribution requires only that the distribution functions converge at the continuity points of F, and F is discontinuous at $t = 1$.

We note that convergence in probability is a stronger property than convergence in distribution.

Proposition 4. *Suppose that $X_n \xrightarrow{\text{Pr}} X$. Then $X_n \xrightarrow{\text{d}} X$.*

Proof.

$$\begin{aligned} P(X_n \le t) &= P(\{X_n \le t\} \cap \{|X_n - X| \le \epsilon\}) + P(\{X_n \le t\} \cap \{|X_n - X| > \epsilon\}) \\ &\le P(X \le t + \epsilon) + P(|X_n - X| > \epsilon) \end{aligned}$$

First pick ϵ small enough so that $P(X \le t + \epsilon) \le P(X \le t) + \eta/2$. (Since F is right-continuous.) Then pick n large enough so that $P(|X_n - X| > \epsilon) < \eta/2$. Then, for n large enough, we have

$$F_n(t) \le F(t) + \eta\,.$$

Similarly, for n large enough, if F is continuous at t, we have $F_n(t) \ge F(t) - \eta$. This shows that $\lim_n F_n(t) = F(t)$ at continuity points of F. □

The converse is not true, but there is one special case where it is. We leave the proof to the reader.

Proposition 5. *Suppose that $X_n \xrightarrow{\text{d}} c$, where c is a constant. Then $X_n \xrightarrow{\text{Pr}} c$.*

Thus, when the limit is a constant, convergence in probability and convergence in distribution are equivalent.

2 Central Limit Theorem

The Weak Law says that a sample mean, $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$, is close to μ in probability, when $X_1, X_2, \ldots.$ is and i.i.d. sequence. But this says nothing about the actual distribution of $\bar{X}_n$. What does it look like for large values of n.

This is what the Central Limit Theorem answers.

Theorem 6 (Central Limit Theorem). *Let $X_1, X_2, \ldots$ be an i.i.d. sequence of random variables with mean μ and variance σ^2. Let*

$$Z_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma}\,.$$

Then $Z_n \to Z$, where Z is a standard normal random variable.