Statistics 3657 : Convergence

Definition 1 *Let X_n be a sequence of r.v.s. We say X_n converges in probability to a constant a if and only if for any $\epsilon > 0$*

$$P(|X_n - a| > \epsilon) \to 0$$

Theorem 1 *Suppose X_i is a sequence of iid random variables with mean μ and variance σ^2. Let*

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Then $\bar{X}_n$ converges in probability to μ.

Proof Notice that $\mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$. Therefore Chebyshev's inequality applies and $\bar{X}_n$ has finite mean and variance. Thus by Chebyshev's inequality

$$\begin{aligned}
P(|\bar{X}_n - \mu| > \epsilon) &\leq \frac{\mathrm{Var}\bar{X}_n}{\epsilon^2} \\
&= \frac{\sigma^2}{n\epsilon^2} \\
&\to 0 \text{ as } n \to \infty
\end{aligned}$$

Convergence in Distribution

Definition 2 *Let X_n be a sequence of r.v.s. with cdfs F_n. We say X_n converges in distribution to (the distribution with the cdf) F if and only if for all x such that F is continuous at x the following holds*

$$F_n(x) \to F(x) \text{ as } n \to \infty$$

We often write or state this as X_n converges in distribution to X iff and

$$F_n(x) \to F(x) \text{ as } n \to \infty$$

at all continuity points of F, and where F is the cdf of X.

The convergence is required only at continuity points of the limit cdf F. Convergence in distribution is a property about convergence of the cdfs.

In the special case where X_n and the limit cdf are both discrete and have support the integers there is an equivalent form of convergence in distribution. We state it without showing they are equivalent.

If X_n are all integer valued r.v.s and F also has support being the integers, then X_n converges in distribution to F if and only if all the point probability masses converge, that is

$$P(X_n = k) \to P(X = k) \ .$$

Aside : There is an analogous result for discrete random variables as long as the support of all the X_n and X is the same. For example this can be used to describe convergence to other discrete distributions such as multinomial.

Convergence of binomial to Poisson

Suppose X_n has Binomial(n, p_n) distribution. Suppose also that $np_n \to \lambda > 0$. ThenX_n converges in distribution to Poisson, λ. We verify this by the definition. Fix an integer $k \geq 0$. We study the sequence $P(X_n = k)$.

$$\begin{aligned}
P(X_n = k) &= \binom{n}{k} (1-p_n)^{n-k} p_n^k \\
&= \frac{n(n-1)\ldots(n-k+1)}{k!} (1-p_n)^{n-k} p_n^k \\
&= \frac{1}{k!} (1-p_n)^{n-k} (np_n)((n-1)p_n)\ldots((n-k+1)p_n) \\
&= \frac{1}{k!} \left\{ \left(1 - \frac{np_n}{n}\right)^n \right\} \left\{ (1-p_n)^{-k} \right\} \{(np_n)((n-1)p_n)\ldots((n-k+1)p_n)\}
\end{aligned}$$

In this last expression the first term is a constant, 1 over $k!$, the second term converges to $e^{-\lambda}$, the third term converges to $1^{-k} = 1$, and the fourth term converges to λ^k.

Recall from introductory calculus that if we consider a product of terms $a_n b_n$, and $a_n \to a$ and $b_n \to b$, then $a_n b_n \to ab$.

Thus for each integer $k \geq 0$

$$P(X_n = k) \to \frac{1}{k!} e^{-\lambda} \lambda^k = \frac{\lambda^k}{k!} e^{-\lambda}$$

This limit is the pmf of a Poisson λ distribution. Thus X_n converges in distribution to Poisson λ.

In Chapter 2 this result is given for $p_n = \frac{\lambda}{n}$.

Remark Usually we cannot find the limit pmf or cdf so easily. Instead there is another method that can be used to show or prove convergence in distribution. It is a result called the Continuity Theorem. We state it but do not proof this result as it is beyond the mathematical tools we have at this stage.

See the Rice text, Section 5.3, Theorem A.

Theorem 2 (Continuity Theorem) *Suppose F_n is a sequence of cdfs, each with moment generating function (mgf), and let M_n be the mgf corresponding to cdf F_n. Let F be a cdf with mgf M. If $M_n(t) \to M(t)$ for all t in this open interval about 0, then F_n converges in distribution to F.*

Remark Theorem 2 gives us another tool to show that a sequence of random variables, or equivalently their sequence of cdf-s converge in distribution, that is F_n converges to F in the sense of Definition 2.

Sometimes this calculation of working with the mgf is *much easier* than the cdf F_n directly. We may not even be able to easily calculate the cdf F_n easily, for example in the case of the cdf of $\sqrt{n}\left(\bar{X}_n - \mu\right)$ that comes up in the Central Limit Theorem.

Example This is a continuation of the example of above. Suppose X_n has Binomial(n, p_n) distribution. Suppose also that $np_n \to \lambda > 0$. ThenX_n converges in distribution to Poisson, λ.

First recall that the Poisson mgf is obtained as follows. Suppose $Y \sim$ Poisson λ.

$$\begin{aligned} M(t) &= \mathrm{E}(e^{tY}) \\ &= \sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \sum_{k=0}^{\infty} \frac{(e^t\lambda)^k}{k!} e^{-\lambda} \\ &= e^{\lambda e^t} e^{-\lambda} \\ &= e^{\lambda\left(e^t-1\right)} \end{aligned}$$

X_n has mgf

$$M_n(t) = \left(1 - p_n + p_n e^t\right)^n = \left(1 + p_n(e^t - 1)\right)^n$$

We need to show for each t in an open neighbourhood of 0, that $M_n(t) \to M(t)$. Since M_n and M are defined for all t, we just show for each t that $M_n(t) \to M(t)$.

$$\begin{aligned} M_n(t) &= \left(1 + p_n(e^t - 1)\right)^n \\ &= \left(1 + \frac{np_n}{n}(e^t - 1)\right)^n \\ &= \left(1 + \frac{np_n(e^t - 1)}{n}\right)^n \end{aligned}$$

Since for sequences $a_n \to a$ we have

$$\left(1 + \frac{a_n}{n}\right)^n \to e^a$$

and using the property that

$$np_n(e^t - 1) \to \lambda(e^1 - 1)$$

we then obtain

$$M_n(t) \to e^{\lambda\left(e^t-1\right)} = M(t)$$

Thus by the Continuity Theorem (Theorem 2) X_n converges in distribution to Poisson λ.

Remark This is only a little easier than the direct method.

End of Example

Example Refer back to the handout on moment generating function, specifically the last calculation. There we considered the example of X_i iid mean 0 and variance 1, and then considered the random variable

$$Z_n = \sqrt{n}\bar{X}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$$

If X has mgf M_X then Z_n has mgf

$$M_n(t) = \left(M_X(\frac{t}{\sqrt{n}})\right)^n$$

We also showed that

$$M_n(t) \to e^{\frac{1}{2}t^2}$$

Recall that the standard normal distribution has mgf $M(t) = e^{\frac{1}{2}t^2}$. Thus we showed that $M_n(t) \to M(t)$. Therefore by the Continuity Theorem we conclude that Z_n converges in distribution the standard normal distribution.

End of Example

Theorem 3 (Central Limit Theorem) *Suppose that X_i are iid random variables with mean μ and variance σ^2. Let*

$$Z_n = \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma}$$

Then Z_n converges in distribution to $N(0,1)$.

Remarks Z_n is the standardized random variable obtained from $\bar{X}_n$. Let

$$S_n = X_1 + X_2 + \ldots X_n \ .$$

Then Z_n can be rewritten in equivalent forms.

$$\begin{aligned} Z_n &= \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n\sigma^2}} \\ &= \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - \mu}{\sigma} \end{aligned}$$

Proof of Theorem 3 : Let M_{CX} be the moment generating function of the standardized X_i, that is the mgf of $\frac{X_i - \mu}{\sigma}$. Let M_n be the mgf of Z_n. Then using the calculation in the previous example we have

$$M_n(t) = \left(M_{CX}(\frac{t}{\sqrt{n}})\right)^n \to e^{\frac{1}{2}t^2}$$

Note also that $M(t) = e^{\frac{1}{2}t^2}$ is the mgf of the $N(0,1)$ distribution. Therefore by the the Continuity Theorem (Theorem 2) Z_n converges in distribution to $N(0,1)$.

End of Proof

We could also have worked the proof by studying

$$M_n(t) = \mathrm{E}\left(\exp\{\frac{t}{\sqrt{n}\sigma}(X-\mu)\}\right)^n = \left(M\left(\frac{t}{\sqrt{n}\sigma}\right)e^{-\frac{t\mu}{\sqrt{n}\sigma}}\right)^n \ .$$

Thus we would then have to study the limit, as $n \to \infty$, of the expression

$$\log\left(M_n(t)\right) = n\log\left(M\left(\frac{t}{\sqrt{n}\sigma}\right)\right) - n\frac{t\mu}{\sqrt{n}\sigma}$$

Using the l'Hôpital's method as we did in class we have

$$\begin{aligned} \lim_{n\to\infty} \log\left(M_n(t)\right) &= \lim_{n\to\infty}\left\{n\log\left(M\left(\frac{t}{\sqrt{n}\sigma}\right)\right) - \sqrt{n}\frac{t\mu}{\sigma}\right\} \\ &= \lim_{n\to\infty} n\left\{\log\left(M\left(\frac{t}{\sqrt{n}\sigma}\right)\right) - \frac{1}{\sqrt{n}}\frac{t\mu}{\sigma}\right\} \end{aligned}$$

$$
\begin{aligned}
&= \lim_{n\to\infty} \frac{1}{\frac{1}{n}} \left\{ \log\left(M\left(\frac{t}{\sqrt{n}\sigma} \right) \right) - \frac{1}{\sqrt{n}} \frac{t\mu}{\sigma} \right\} \\
&= \lim_{\theta\to 0} \frac{1}{\theta^2} \left\{ \log\left(M\left(\theta \frac{t}{\sigma} \right) \right) - \theta \frac{t\mu}{\sigma} \right\}
\end{aligned}
$$

The student should finish this at home. Notice that one cannot separate the two terms in the numerator and take limits separately as these individual limits do not exist. For the second of these we *would have*

$$\lim_{\theta\to 0} \frac{\theta \frac{t\mu}{\sigma}}{\theta^2} = \lim_{\theta\to 0} \frac{\frac{t\mu}{\sigma}}{\theta}$$

and this last limit does not exist.

Example

In order to help the student understand this calculation of working with the MGF of the standardized sum consider the case of X_i being iid Gamma(α, λ). Recall this distribution has MGF

$$M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha} .$$

From this we obtain

$$\mathrm{E}(X) = \frac{\alpha}{\lambda} \ , \ \mathrm{Var}(X) = \frac{\alpha}{\lambda^2} \ .$$

Notice also that

$$\begin{aligned} Z_n &= \frac{\sqrt{n}\left(\bar{X}_n - \frac{\alpha}{\lambda}\right)}{\sqrt{\frac{\alpha}{\lambda^2}}} \\ &= \frac{\lambda}{\sqrt{\alpha}}\sqrt{n}\bar{X}_n - \frac{\lambda\alpha}{\sqrt{\alpha}\lambda}\sqrt{n} \\ &= \frac{\lambda}{\sqrt{\alpha}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i - \sqrt{\alpha}\sqrt{n} \end{aligned}$$

The MGF of Z_n is M_n given by

$$\begin{aligned} M_n(t) &= \mathrm{E}\left(e^{tZ_n}\right) \\ &= \left(M_X\left(t\frac{\lambda}{\sqrt{\alpha}}\frac{1}{\sqrt{n}}\right)\right)^{n} e^{-t\sqrt{\alpha}\sqrt{n}} \end{aligned}$$

Next notice that

$$\begin{aligned} M_X\left(t\frac{\lambda}{\sqrt{\alpha}}\frac{1}{\sqrt{n}}\right) &= \left(\frac{\lambda}{\lambda - t\frac{\lambda}{\sqrt{\alpha}}\frac{1}{\sqrt{n}}}\right)^{\alpha} \\ &= \left(\frac{1}{1 - \frac{t}{\sqrt{\alpha}}\frac{1}{\sqrt{n}}}\right)^{\alpha} \end{aligned}$$

We now study the natural log of the MGF M_n, and take limits as $n \to \infty$.

$$\begin{aligned} \log M_n(t) &= -n\alpha \log\left(1 - \frac{t}{\sqrt{\alpha}\sqrt{n}}\right) - t\sqrt{\alpha}\sqrt{n} \\ &= -\alpha n \left\{\log\left(1 - \frac{t}{\sqrt{\alpha}\sqrt{n}}\right) + \frac{t}{\sqrt{\alpha}\sqrt{n}}\right\} \\ &= -\alpha \frac{1}{\left(\frac{1}{\sqrt{n}}\right)^2}\left\{\log\left(1 - \frac{t}{\sqrt{\alpha}\sqrt{n}}\right) + \frac{t}{\sqrt{\alpha}\sqrt{n}}\right\} \end{aligned}$$

Notice this is of the form where, as $n \to \infty$, the numerator and the denominator both converge to 0. Thus we can embed the limit into the form of a limit in terms of $\theta \to 0$, where θ corresponds to $\frac{1}{\sqrt{n}}$, and then use L'Hopital's Rule to evaluate the limit in terms of $\theta \to 0$.

Aside: Review the use of this technique in the handout on MGFs. In the calculation below we will use L'Hopital's Rule in the third and fourth lines.

Thus

$$\begin{aligned}
\lim_{n\to\infty} \log M_n(t) &= -\alpha \lim_{n\to\infty} \frac{1}{\left(\frac{1}{\sqrt{n}}\right)^2} \left\{ \log\left(1 - \frac{t}{\sqrt{\alpha}\sqrt{n}}\right) + \frac{t}{\sqrt{\alpha}\sqrt{n}} \right\} \\
&= -\alpha \lim_{\theta\to 0} \frac{1}{\theta^2} \left\{ \log\left(1 - \frac{t}{\sqrt{\alpha}}\theta\right) + \frac{t}{\sqrt{\alpha}}\theta \right\} \\
&= -\alpha \lim_{\theta\to 0} \frac{1}{2\theta} \left\{ \left(1 - \frac{t}{\sqrt{\alpha}}\theta\right)^{-1} (-1)\frac{t}{\sqrt{\alpha}} + \frac{t}{\sqrt{\alpha}} \right\} \\
&= -\alpha \lim_{\theta\to 0} \frac{1}{2} \left\{ \left(1 - \frac{t}{\sqrt{\alpha}}\theta\right)^{-2} (-1)^3 \left(\frac{t}{\sqrt{\alpha}}\right)^2 \right\} \\
&= \alpha \frac{t^2}{2\alpha} \\
&= \frac{t^2}{2}
\end{aligned}$$

Thus

$$\log M_n(t) \to \frac{t^2}{2}$$

and hence

$$M_n(t) \to e^{\frac{t^2}{2}} \ .$$

Since the limit is the MGF of the standard normal distribution, therefore by the Continuity Theorem Z_n converges in distribution to $N(0,1)$.

End of Example

In the above example we used L'Hopital's Rule to obtain the limit. We can also use Taylor's formula. This Taylor's formula will involve the expansion of $g(x) = \log(1-x)$ about the value $x_0 = 0$. Notice this is because we have to work with

$$\log\left(1 - \frac{t}{\sqrt{\alpha}\sqrt{n}}\right)$$

and for large n the function g is evaluated at $\frac{1}{\sqrt{\alpha}\sqrt{n}}$, that is a value near 0. Next notice that we have to study

$$n\log(1 - \frac{1}{\sqrt{\alpha}\sqrt{n}}) = ng\left(\frac{1}{\sqrt{\alpha}\sqrt{n}}\right) \ .$$

Thus for the Taylor's approximation formula we will need to go out to order high enough, that is to degree k so that the terms we ignore or drop in our approximation are small, that is k so that

$$n\left(\frac{1}{\sqrt{\alpha}\sqrt{n}}\right)^{k+1}$$

converges to 0 as $n \to \infty$. Notice that when $k = 2$ and so $k + 1 = 3$, then this term is

$$\frac{1}{\sqrt{\alpha}^{3/2}} \frac{1}{\sqrt{n}} \to 0 \text{ as } n \to \infty \ .$$

Thus we will need to use $k = 2$, and hence use a Taylor's formula or Taylor's approximation of degree 2.

We have

$$\begin{aligned}
g'(x) &= -\frac{1}{1-x} \\
g''(x) &= -\frac{1}{(1-x)^2}
\end{aligned}$$

Therefore the second order Taylor's approximation for g about $x_0 = 0$ is

$$\begin{aligned} g_2(x) &= g(0) + g'(0)x + \frac{1}{2}g''(0)x^2 \\ &= -x - \frac{1}{2}x^2 \end{aligned}$$

Therefore

$$\begin{aligned} &-n\alpha \log\left(1 - \frac{t}{\sqrt{\alpha}\sqrt{n}}\right) - t\sqrt{\alpha}\sqrt{n} \\ \approx \quad & -n\alpha\left\{-\frac{t}{\sqrt{\alpha}\sqrt{n}} - \frac{1}{2}\frac{t^2}{\alpha n}\right\} - t\sqrt{\alpha}\sqrt{n} \\ = \quad & t\sqrt{\alpha}\sqrt{n} + \frac{t^2}{2} - t\sqrt{\alpha}\sqrt{n} \\ = \quad & \frac{t^2}{2} \end{aligned}$$

Thus

$$M_n(t) \to e^{\frac{1}{2}t^2} \text{ as } n \to \infty \ .$$

Remark : Taylor's approximation is mathematically more insightful in studying these limits than L'Hopital's Rule, but is does require a little more care. In particular this is the case if one were to show that the remainder term in the approximation converges to 0 as $n \to \infty$, but we not consider this step in detail in this course. It is also worth remarking here that the type of Taylor's analysis above is one of the problems on the first assignment.

How many games will the Toronto Maple Leafs win this season?

Let X_i be a Bernoulli random variable, 1 if the Leafs win the i-th game, and 0 if they loose the game. There are $n = 82$ games. Suppose that X_i are iid Bernoulli with parameter θ. Base on the CLT the 0.95 prediction interval for $\sum_{i=1}^n X_i$ is obtained by solving

$$0.95 = P\left(-a \leq \frac{\sum_{i=1}^n X_i - n\theta}{\sqrt{n\theta(1-\theta)}} \leq a\right)$$

giving $a = 1.96$. Thus with probability 0.95

$$-1.96 \leq \frac{\sum_{i=1}^n X_i - n\theta}{\sqrt{n\theta(1-\theta)}} \leq 1.96$$

If we know θ then with probability 0.95

$$X_n \in \left[n\theta - 1.96 * \sqrt{n\theta(1-\theta)}, n\theta + 1.96 * \sqrt{n\theta(1-\theta)}\right]$$

This notion is called a prediction interval or prediction set.