EBA3500: CONVERGENCE IN DISTRIBUTION AND THE CENTRAL LIMIT THEOREM

JONAS MOSS

Intuitively, convergence in distribution means "convergence of histograms". The definition, however, is quite different.

Definition 1. A sequence of real-valued random variables X_n with distribution functions F_n converges in distribution to a random variable X with distribution function F if $F_n(x) \to F(x)$ for all points x where F is continuous. We denote convergence in distribution by $X_n \xrightarrow{d} X$.

We include this definition mostly for completeness – I won't ask you to do anything with it. But you should be aware that convergence in distribution has a precise, mathematical definition and that it involves the distribution function F.

There are some trivial examples of convergence in distribution. Again, think in terms of histograms, not distribution functions:

(1) Let X be normal and $X_n = X + \frac{1}{n}$. Then X_n converges in distribution to X.
(2) Let $Y_1, Y_2, ...$ be independent standard normal variables. Then Y_n still converges in distribution to a standard normal.

As these two examples show, convergence in distribution has nothing to do with the dependence of the variables in the sequence.
Convergence of estimators. We usually deal with estimators

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \sigma)$$

The term $\sqrt{n}$ is called the *rate of convergence*. In this case it's the square root of n. This is by far the most common rate of converges, and virtually every interesting quantity that converges with this reate converges to a normal distribution. The variance σ^2 is frequently called the *asymptotic variance*.

Theorem 2 (Central Limit Theorem)**.** *Let $X_1, ..., X_n$ be independent and identically distributed random variables with finite variance σ^2 and mean μ. Then $\sqrt{n}(\overline{X} - \mu) \to N(0, \sigma)$.*

Remark 3. The rate of convergence is $\sqrt{n}$ for any sequence $X_1, X_2, ..., X_n$, but the speed of convergence will vary. For instance, if all X_i are normal, their mean will be exactly normal, for any n. But if the variables have fat tails, such as the log-normal distribution, the convergence will be slow.

Remark 4. The assumption that σ^2 is finite is necessary. There are distributions with infinite variance, such as the t-distribution with 2 degrees of freedom. See this post on CrossValidated.

Remark 5. The most common proof of the central limit theorem is based on so-called *moment-generating functions* or on Fourier transforms. This material is beyond the scope of this lecture, but I would recommend reading the wikipedia page on moment-generating functions, as it does not require much mathematical background.

What is the point of this? We care about convergence in distribution for three reasons:

(1) **(Qualitative)**. It gives us qualitative information about how much we can expect to know, and how much we can expect it to help to get more data.
(2) **(Inference)**. We can use the convergence to construction confidence intervals and do hypothesis tests.

Qualitative aspects. Think about any problem where you want to measure something, for instance the mean income among Norwegian 30-year olds. Suppose you have sampled $n_1 = 100$ people from this population. How accurate is your estimated mean? You can answer that question using the standard deviation. But how much will sampling yet another 100 people reduce your standard deviation? Sampling 100 more people, so that you have $n_2 = 200$, will reduce your standard deviation by a factor of

$$\frac{\sigma}{\sqrt{n_1}} / \frac{\sigma}{\sqrt{n_2}} = \sqrt{\frac{n_2}{n_1}} = \sqrt{\frac{200}{100}} = \sqrt{2} \approx 1.414.$$

Why would you care about this? The result that most estimators converge at the same rate, to the same limit distribution, allows you to do the same thing with other estimators too! You can reason as if the estimator was the mean! Examples include the regression coefficients in virtually any regression model, the median, the standard deviation, and so on.

Review question: You have calculated the median based on $n = 100$ samples. How much more precise will you be when using $n = 300$ samples instead?

Answer: You can reason just as you would for the mean, hence the ratio is $\sqrt{3} \approx 1.73$.

Let's consider another example, namely that of estimating the mode of a distribution. The most famous estimator of the mode is Chernoff's mode estimator, which converges at the rate $n^{1/3}$. Then you can't reason as you would for the mean anymore.

Review question: You have calculated the mode, using Chernoff's estimator, based on $n = 100$ samples. How much more precise will you be when using $n = 300$ samples instead?

Answer: Now you can't reason in the same way anymore, but nearly in the same way.

$$\frac{\sigma}{n_1^{1/3}} / \frac{\sigma}{n_2^{1/3}} = \left(\frac{n_2}{n_1}\right)^{1/3} = \left(\frac{300}{100}\right)^{1/3} = 3^{1/3} \approx 1.44.$$

Inference. The central limit theorem allows you pretend that everything is normal all the time. Importantly, you can use this to construct approximate confidence intervals and hypothesis tests.

Remember the definition of a confidence interval: An interval CI is a level $1-\alpha$ confidence interval for θ if it includes θ with probability at least $1-\alpha$, no matter what the true θ is. If we pretend the central limit theorem holds exactly, we might make

Proposition 6. *Suppose that* $\sqrt{n}(\hat{\theta} - \theta) \to N(0, \sigma)$ *and let* $\hat{\sigma}$ *be a consistent estimator of the asymptotic variance. Then an approximate* $(1 - \alpha)\%$ *confidence interval can be constructed,*

$$CI = [\hat{\theta} - z_{1-\alpha/2}\hat{\sigma}/\sqrt{n}, \hat{\theta} + z_{1-\alpha/2}\hat{\sigma}/\sqrt{n}],$$

where z_q *is the* q*-quantile of the normal distribution.*

Remark 7. The quntile function of the normal distribution is called `scipy.stats.ppf(q)`in Python.