Understanding Limit Theorems

1 Reminders

1.1 The (Weak) Law of Large Numbers

You may recall the statement of this result. In its simple form it states that

> Given a sequence of independent identically distributed random variables, with finite expectation μ and finite variance σ^2, X_1, X_2, ... , the average $\frac{1}{n}\sum_{k=1}^{n} X_k$ will approach μ as n becomes larger. More precisely, the difference $\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right|$ will be smaller than any fixed value you choose, with probability as close to 1 as you choose, provided n is large enough.

Remark 1. The general statement of this result uses much less restrictive assumptions, but this is the form which we will be using in almost all statistical applications

This is sometimes loosely stated as affirming that, for example, if we flip a fair coin 100 times, the number of heads will be, most likely, very close to 50. The only problem is that this is not quite what the result implies, at least until we make clear what we mean by "very close". A first clarification comes, maybe unexpectedly, from the proof of the statement. In fact, the proof is based on a famous result (*Chebyshev's Inequality*), which states

$$P\left[\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| > \varepsilon\right] < \frac{\sigma}{\varepsilon^2\sqrt{n}}$$

In words, we can make sure that the arithmetic mean of our random variables does not differ from their common expectation by more than ε (and it's up to us to decide how close we want them to be), provided we add enough of them - precisely, this will happen with probability higher than $1-\delta$ (we can choose δ) if $n > \frac{\sigma^2}{\delta^2\varepsilon^4}$. If ε and δ are small, this can be a really big number, but then this statement is true, no matter what the distribution of our random variables. Thus, if, say, $\sigma = 1$, $\varepsilon = 0.01$, and $\delta = 0.01$, we have

$$n > \frac{1}{10^{-4}\cdot 10^{-8}} = 10^{12}$$

which is a bit daunting.

The next theorem gives us a way to be more precise and with a lower requirement on n, for "well behaved" distributions.

1.2 The Central Limit Theorem

This is the other major tool in most statistical applications. In its simplest form it says (again, there are more general results, with weaker assumptions, but the following is good enough for most of our needs)

> Given a sequence of independent identically distributed random variables, with finite moments up to four, expectation μ, and variance σ^2, the difference
>
> $$\left|P\left[a \leqslant \frac{\frac{1}{n}\sum_{k=1}^{n} X_k - \mu}{\sigma/\sqrt{n}} \leqslant b\right] - (\Phi(b) - \Phi(a))\right|$$
>
> becomes as small as we wish, provided n is large enough

Note that the expression in the probability on the left is the average of the variables, minus its expected value, divided by its standard deviation, and is thus a random variable with expected value 0, and variance 1. The difference to the right is the probability that a standard normal random variable take values between a and b (Φ is the cumulative distribution function of such a variable: it is hard to compute "by hand", but its values are available from tables in any book on probability or statistics, and are hard coded in any modern spreadsheet or mathematical and statistical software).

Loosely speaking, the theorem tells us that

$$\frac{1}{n}\sum_{k=1}^{n} X_k - \mu$$

is, approximately, a normal random variable with mean 0, and standard deviation $\frac{\sigma}{\sqrt{n}}$, at least if n is large enough.

The problem of what "large enough" means in this case is not quite easy to discuss precisely. With a little experimentation, we can convince ourselves that n need not be too big, if the distribution of our variables is symmetric and reasonably concentrated. We will see a few examples in the next modules, when we will be using tables to evaluate certain useful distributions, and will be able to tell from the tables when that particular distribution becomes almost indistinguishable from the normal one.

In the meantime, as a famous example, if our variables are uniform over the interval $\left[-\frac{1}{2}, \frac{1}{2}\right]$, $\sum_{k=1}^{12} X_k$ will have mean zero and variance 1 (since each of the variables happens to have mean 0 and variance $\frac{1}{12}$). Even though the number of summands is pretty small, the distribution of this sum is fairly close to standard normal, and, as such, has been used for a very long time to simulate standard normal random variables on IBM mainframes (one rough way to check this is to use a random number generator to simulate this experiment).

For the binomial distribution, a common rule of thumb is that we should have $np(1-p) > 10$ (thus, if $p = \frac{1}{2}$, we would like $n > 40$ – this is the "best case": for different values of p, we need larger n: e.g., if $p = \frac{1}{10}$, this rule requires $n > \frac{1000}{9} \approx 334$). Note that if p happens to be very close to 0 or 1, there is a better approximation (the *Poisson Distribution*), which we won't go into in this course.

2 Back to the Law of Large Numbers

Thanks to the Central Limit Theorem (assuming it has kicked in, in a specific case), we can be more precise about the meaning of the Law of Large Numbers. In fact, if the difference $\frac{1}{n}\sum_{k=1}^{n} X_k - \mu$ can be considered to be a normal random variable with mean zero and variance $\frac{\sigma^2}{n}$, using tables for the standard normal distribution, we conclude that

$$P\left[-\frac{\sigma}{\sqrt{n}} < \frac{1}{n}\sum_{k=1}^{n} X_k - \mu < \frac{\sigma}{\sqrt{n}}\right] \approx 0.68$$

$$P\left[-\frac{2\sigma}{\sqrt{n}} < \frac{1}{n}\sum_{k=1}^{n} X_k - \mu < \frac{2\sigma}{\sqrt{n}}\right] \approx 0.96$$

$$P\left[-\frac{3\sigma}{\sqrt{n}} < \frac{1}{n}\sum_{k=1}^{n} X_k - \mu < \frac{3\sigma}{\sqrt{n}}\right] \approx 0.997$$

Clearly, we need a handle on σ, before we can rush into
any more specific statement.

In the coin-tossing experiment, we are in luck: the variables in question have a Bernoulli distribution, with parameter p, in general, and $p = \frac{1}{2}$ for a fair coin. Their sum has a Binomial distribution, with mean np, and variance $np(1-p)$, or, in the fair case, $\frac{n}{2}$, $\frac{n}{4}$ respectively (this is as high as the variance can go for a binomial distribution). Hence, the inequalities above become ($\mu = \frac{1}{2}$, $\sigma = \frac{1}{2}$), after multiplying throughout by n,

$$P\left[-\frac{1}{2}\sqrt{n} < \sum_{k=1}^{n} X_k - \frac{n}{2} < \frac{1}{2}\sqrt{n}\right] \approx 0.68$$

$$P\left[-\sqrt{n} < \sum_{k=1}^{n} X_k - \frac{n}{2} < \sqrt{n}\right] \approx 0.96$$

$$P\left[-\frac{3}{2}\sqrt{n} < \sum_{k=1}^{n} X_k - \frac{n}{2} < \frac{3}{2}\sqrt{n}\right] \approx 0.997$$

For $n = 1{,}000{,}000$, $\sqrt{n} = 1{,}000$. Hence, the number of heads will be in the range between 499,500 and 500,500 with 68% probability, within 499,000 and 501,000 with 96% probability, and within 498,500 and 501,500 with 99.7% probability. This is probably "close enough" for most of us. For $n = 100$, $\sqrt{n} = 10$, and the respective ranges are now $[45, 55]$, $[40, 60]$, and $[35, 65]$. This is rougher than one might expect.

As should be obvious by now, what becomes very small is the *relative difference* from the expected value, while the *absolute difference* increases with n, but at a slower pace — like $\sqrt{n}$.

3 What Does the Central Limit Theorem Imply?

This theorem is sometimes called "the second fundamental theorem of statistics" (after the Law of Large Numbers), and it is indeed a cornerstone of much of statistical practice. However, its scope is wider, and we should be aware that this theorem has further implications.

The strict and main implication affects our analysis of sample means. The theorem is fairly clear in this respect: if we have a collection of independent, identically distributed variables, their average is going to be approximately normally distributed, provided the sample is large enough. How large is "enough" is a different issue, but, broadly speaking, this is a big argument for assuming that sample means can be treated as being normally distributed – a large portion of statistical day-to-day work.

There is a second implication that is less precise, but extremely important in applications. Let's rephrase the theorem for this particular purpose. The expression that is approximately distributed as a standard normal variable is

$$\frac{\frac{1}{n}\sum_{k=1}^{n}(X_k - \mu)}{\sigma/\sqrt{n}} = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\frac{X_k - \mu}{\sigma}$$

This can be read as the sum of n zero-mean $(X_k - \mu)$ terms (the division by σ reduces the variance to 1, but is not crucial), "rescaled" by a factor of $\frac{1}{\sqrt{n}}$. An important point is that these terms are assumed to be independent. In other words, without trying to be very precise, on an intuitive basis, this theorem tells us that if we add up the effect of many small **independent** contributions, the result will be approximately normally distributed.

Now, a situation similar to the one sketched above is fairly common. The "poster child" is provided by measurements of physical quantities. After making sure we have eliminated all extraneous perturbations, measuring a physical quantity (for example think of measuring the voltage of a battery) is going to be affected by many uncontrollable small actions that are way outside our control. It makes sense to assume that the result will be a random addition to the basic measure, as in $m + \varepsilon$, where m is the "true" value, while ε is a zero-mean normal random variable.

Not the two basic assumptions in the argument above:

- The components adding up are small
- They are independent

How can this go wrong? It is too tempting to refer to financial models as examples where these assumptions may fail – and where not being aware of this could cause very significant consequences. As far as the first point goes, we need to assume that no large disruption is expected to occur: all uncontrolled influences must not be large compared to the scale of what we are observing (things like the "flash crash" in the stock market in 2010 are not supposed to happen). As far as the second point goes, there should be no "herd effect", or "snowball effect": however you want to call it, a swing in one direction should not cause further pushes in the same direction.

It is fair to say, at least regarding many financial markets, that both assumptions may make sense in "normal" times. However, if an external event should suddenly barge into the picture, the "normal" model is not likely to be useful. More ominously, if an event is going to have a ripple effect, disrupting the independence assumption (in 2008 the drop in housing prices started a stampede), the result will be anything but normal.

What this short discussion is leading to is that the "universal" nature of the Gaussian distribution has a solid foundation, but it also has its limits. Depending on what we are trying to model a normally distributed model might be easy to manage, but it may not necessarily be appropriate.