Asymptotic normality of the sample mean and Edgeworth expansions

So far we have concentrated on showing asymptotic normality for anything under the sun. But this result, as the name suggests, is asymptotic. For finite samples, indeed small samples, this approximation can be quite poor. We recall if the original data appears to have the following features:

(i) Skewness (non-zero third order cumulant).

(ii) Thick tails (described as kurtosis - this is $\frac{1}{\sigma^4}(\kappa_4 - \sigma^4)$), where $\kappa_4$ is the fourth order cumulant of the random variable.

Then the normality result only sets in with ‘very large' sample sizes. Heuristically, this is because the features in the population distribution will influence the sampling distribution. The further the population distribution is from normal the further the sampling distribution will be too.

The above is rather heuristic, it is natural to ask how the above above effects the rate of convergence of the distribution of $T^{-1/2} \sum_{t=1}^{T} X_t$ to normality. To answer this question we recall how we usually prove normality of the sample average $T^{-1/2} \sum_{t=1}^{T} X_t$.

Remark 11.1 (A quick review of cumulants) Suppose the random variable $Y$ has distribution function $F$ (density $f$) the characteristic function of $Y$ is defined as the Fourier transform

$$\chi_Y(t) = \int \exp(itx)dF(x) = \int \exp(itx)f(x)dx.$$  

The $r$th order cumulant of a random variable, is the coefficient of $t^r/r!$ in the series expansion of $\log \chi_Y(t)$. Since cumulants are derived from the logarithm of the characteristic function, they can be represented in terms of its moments, eg.

$$\kappa_1 = \mathbb{E}(X) \quad \kappa_2 = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$  

For the normal distribution all the cumulants greater than order 2 are zero (this is unique for the normal distribution - and is usually how we prove normality of an estimator).

The joint cumulants (similar to joint moments) of a multivariate random variable can be derived in a similar way. $\kappa_r(Y_1, \ldots, Y_r)$ is the joint cumulant of $(Y_1, \ldots, Y_r)$ and is the coefficient of $t_1 \ldots t_r$ in the expansion of the log of the characteristic function of the joint distribution of $(Y_1, \ldots, Y_r)$. More general, if $(Y_1, \ldots, Y_r) = (\underbrace{X_1, \ldots, X_{r_1}}_{r_1}, \ldots, \underbrace{X_n, \ldots, X_{r_n}}_{r_n})$, where $r_1 + \ldots + r_n = r$, then $\kappa_r(Y_1, \ldots, Y_r)$ is the coefficient of $\prod_{s=1}^{r} t_s^{r_s}/r!$ in the expansion of the log of the characteristic function of $(X_1, \ldots, X_n)$.
It is interesting to note if at least one of the random variables in \((Y_1, \ldots, Y_r)\) is independent of the rest then \(\kappa_r(Y_1, \ldots, Y_r) = 0\).

The usual method to show normality is to represent the characteristic function of \(T^{-1/2} \sum_{t=1}^T X_t\) as a function of the cumulants and show that all cumulants over the second order cumulant (which is the variance) converge to zero. More precisely, the characteristic function (which is the Fourier transform of the density of \(Y\)) of a random variable \(Y\), with mean zero and variance one is approximately

\[
\chi_Y(t) = \exp\left(\frac{(it)^2}{2} + \frac{(it)^3}{3!} \kappa_3(Y) + \frac{(it)^4}{4!} \kappa_4(Y) + \ldots\right)
\]

where \(\kappa_r(Y)\) denotes the \(r\)th cumulant of \(Y\) (when this approximation is valid is beyond this course - see the recommended books for details). We recall that for standard normal random variables the characteristic function is

\[
\chi_Y(t) = \exp\left(\frac{(it)^2}{2}\right).
\]

Now let us consider what this means for the distribution of \(\bar{S}_T = T^{-1/2} \bar{X}_T = T^{-1/2} \sum_{t=1}^T X_t\). To make notation easier, we will standardise \(\bar{S}_T\), and consider the distribution of \(S_T = T^{1/2} (\bar{X} - \mu)/\sigma\), where \(X_t\) has mean \(\mu\) and variance \(\sigma^2\). Now by expanding the cumulants (which is like expanding the variance of sums of random variables) we have

\[
\kappa_r(S_T) = T^{-r/2} \sum_{s_1, \ldots, s_r=1}^T \kappa_r\left(\frac{X_{s_1} - \mu}{\sigma}, \ldots, \frac{X_{s_r} - \mu}{\sigma}\right)
\]

\[
= T^{-r/2} \sum_{s=1}^T \kappa_r\left(\frac{X_s - \mu}{\sigma}\right) = T^{-r/2+1} \tilde{\kappa}_r\left(\frac{X - \mu}{\sigma}\right),
\]

since \(\{X_t\}\) are iid random variables. We note that for \(r > 2\), \(\kappa_r\left(\frac{X}{\sigma}\right) = \kappa_r\left(\frac{X - \mu}{\sigma}\right)\) and denote \(\tilde{\kappa}_r = \kappa_r\left(\frac{X - \mu}{\sigma}\right)\) (the \(r\)th order cumulant of the standardised random variable).

Now we obtain the characteristic function of the sample mean \(S_T\). By substituting (55) into (54) we have

\[
\chi_{S_T}(t) = \exp\left(\frac{(it)^2}{2}\right) \exp\left(\frac{(it)^3}{3!} T^{-1/2} \tilde{\kappa}_3 + \frac{(it)^4}{4!} T^{-1} \tilde{\kappa}_4 + \ldots\right).
\]

Since \(T^{-r/2+1} \tilde{\kappa}_r \to 0\) as \(T \to \infty\) we have that \(\chi_{S_T}(t) \to \exp\left(\frac{(it)^2}{2}\right)\), which is the normal distribution. It can be shown if the characteric function converges to the characteristic distibution of
a normal, so the distribution must converge to the normal. Hence we have the CLT - though we note that we do not require all moments to exist, in fact it is sufficient that only the second moment exist. However, for this heuristic discussion we shall assume that at least four moments exist.

The above is a heuristic proof of the CLT, but already we get a feeling of how fast this convergence should be. The leading term in the above expansion is \( \frac{(it)^3}{3!}T^{-1/2}\tilde{\kappa}_3 \), which suggests that the ‘error’ in the normal approximation should be of order \( T^{-1/2} \), ie.

\[
|P(S_T \leq x) - \Phi(x)| \approx O(T^{-1/2}),
\]

where \( \Phi(\cdot) \) is the cumulative function. Indeed this is the case, and it can be shown by using an Edgeworth expansion. The Edgeworth expansion, is effectively an expansion of the distribution of \( G_T(x) = P_T(S_T \leq x) \) in terms of the normal distribution and higher order terms. To see how it arises, use the series expansion of \( \exp(\frac{(it)^3}{3!}T^{-1/2}\tilde{\kappa}_3 + \frac{(it)^4}{4!}T^{-1}\tilde{\kappa}_4 + \ldots) \) to rewrite \( \chi_{S_T}(t) \) as

\[
\chi_{S_T}(t) = \exp\left(\frac{(it)^2}{2}\right)\exp\left(\frac{(it)^3}{3!}T^{-1/2}\tilde{\kappa}_3 + \frac{(it)^4}{4!}T^{-1}\tilde{\kappa}_4 + \ldots\right).
\]

Now, we recall that the characteristic function is the Fourier transform of the distribution function, hence inverting the Fourier transform of the above we have

\[
G_T(x) = P(S_T \leq x) = \Phi(x) + \frac{1}{T^{1/2}}p_1(x)\phi(x) + \frac{1}{T}p_2(x)\phi(x) + \frac{1}{T^{3/2}}p_3(x)\phi(x) + \ldots, \quad (57)
\]

where \( \Phi \) is the distribution of the normal, \( \phi(x) \) is the normal distribution and

\[
p_1(x) = -\frac{1}{6}\tilde{\kappa}_3(x^2 - 1) \quad \text{and} \quad p_2(x) = -x\left\{\frac{1}{24}\tilde{\kappa}_4(x^2 - 3) + \frac{1}{72}\tilde{\kappa}_3^2(x^4 - 10x^2 + 15)\right\}.
\]

Hence, as expected we have

\[
\frac{P(S_T \leq x) - \Phi(x)}{G_T(x)} = \frac{1}{T^{1/2}}p_1(x)\phi(x) + \frac{1}{T}p_2(x)\phi(x) + \ldots.
\]

Therefore, the error in this approximation is of order \( O(T^{-1/2}) \) if the there is a skew in the distribution of \( X_t \) and of order \( T^{-1} \) if there isn’t a skew. The technical details can be found in Hall (1992), Chapter 2.
This means that when we construct confidence intervals there will be some errors. For example, if we construct a 95% CI for the mean using the normal approximation, it may in reality be less than a 95% CI.

Now in the same way that the errors in the probabilities using the normal approximation, can be calculated by using Edgeworth expansion, so can the errors in the quantile (which give the CI), by using what is known as a Cornish-Fisher expansion, can be calculated. To understand this, let us recall what the CI for the mean $\mu$ actually means. We recall if we want to construct a 95% CI we try to find the 2.5% and 97.5% quantiles, $\xi_{0.025}$ and $\xi_{0.975}$, such that

$$P(\xi_{0.025} \leq \frac{\sqrt{T}(\bar{X} - \mu)}{\sigma} \leq \xi_{0.975}) = 0.95.$$ 

$\xi_\alpha$ correspond to the $\alpha$ quantile of the distribution $G_T$, which is the distribution of $\sqrt{T}\frac{\bar{X} - \mu}{\sigma}$. Since $\frac{\sqrt{T}(\bar{X} - \mu)}{\sigma}$ is asymptotically normally distributed we approximate $\xi_{0.025}$ and $\xi_{0.975}$ with $z_{0.025}$ and $z_{0.975}$ (which is $-1.96$ and $1.96$) we would approximate the true CI for the mean

$$[\bar{X} + \xi_{0.025}\sigma/\sqrt{T}, \bar{X} + \xi_{0.975}\sigma/\sqrt{T}],$$

with its normal approximation

$$[\bar{X} + z_{0.025}\sigma/\sqrt{T}, \bar{X} + z_{0.975}\sigma/\sqrt{T}].$$

Hence it is interesting to see (a) how close $z_\alpha$ and $\xi_\alpha$ are and (b) what the difference between the two CIs are. But the Edgeworth expansion can be inverted to go from probabilities to quantiles and it can be shown that (see Hall (1992), Chapter 2.5) that

$$\xi_\alpha = z_\alpha + \frac{1}{T^{1/2}} \tilde{p}_1(z_\alpha) + \frac{1}{T} \tilde{p}_2(z_\alpha) + \ldots,$$

where

$$\tilde{p}_1(z) = -p_1(z) \quad \tilde{p}_2(z) = p_1(z)p_1'(z) - \frac{1}{z}p_1(z)^2 - p_2(z).$$

74