LECTURE 15 NOTES

1. The plug-in principle. A statistical functional $\tau(F)$ is a function of the CDF $F(x)$ (or the density $f(x)$). Examples include

1. the expected value $\mu = \int_{\mathbf{R}} xf(x)dx$,
2. the variance $\sigma^2 = \int_{\mathbf{R}} x^2 f(x)dx - \big(\int_{\mathbf{R}} xf(x)dx\big)^2$,
3. critical values $x_\alpha = F^{-1}(1-\alpha)$.

A general approach to estimating statistical functionals is the plug-in principle. Let $\mathbf{x}_i \overset{i.i.d.}{\sim} F$, where F is a distribution on $\mathbf{R}$. The empirical CDF is the distribution that places $\frac{1}{n}$ mass at each observation:

$$\hat{F}_n(x) := \tfrac{1}{n}\textstyle\sum_{i\in[n]} \mathbf{1}_{(-\infty,x]}(\mathbf{x}_i).$$

The empirical CDF has a "density":

$$\hat{f}_n(x) = \tfrac{1}{n}\textstyle\sum_{i\in[n]} \delta(x-\mathbf{x}_i),$$

where $\delta(x)$ is the delta "function".[1] The *plug-in estimator* of $\tau(F)$ is $\tau(\hat{F}_n)$.

Example 1.1. *Consider estimating the mean $\tau(F) = \int_{\mathbf{R}} xf(x)dx$. The plug-in estimator is the expected value of the empirical distribution function, which unsurprisingly, is the sample mean:*

$$\tau(\hat{F}_n) = \int_{\mathbf{R}} x\hat{f}_n(x)dx = \frac{1}{n}\sum_{i\in[n]} \textstyle\int_{\mathbf{R}} x\delta(x-\mathbf{x}_i)dx = \displaystyle\frac{1}{n}\sum_{i\in[n]} \mathbf{x}_i.$$

More generally, the plug-in estimator of a linear statistical functional of the form $\tau(F) = \int_{\mathbf{R}} a(x)f(x)dx$ is $\frac{1}{n}\sum_{i\in[n]} a(\mathbf{x}_i)$.

How good is the plug-in estimator? By the Glivenko-Cantelli theorem,

$$\sup_{x\in\mathbf{R}}\big|\hat{F}_n(x) - F(x)\big| \overset{a.s.}{\to} 0$$

A refinement of the Glivenko-Cantelli theorem by Dvoretzky, Kiefer, Wolfowitz shows that the rate of convergence is $\frac{1}{\sqrt{n}}$:

$$\mathbf{P}\big(\sup_{x\in\mathbf{R}}\big|\hat{F}_n(x) - F(x)\big| > \epsilon\big) \le 2\exp\big(-2n\epsilon^2\big).$$

[1] The delta function is, intuitively, a function that vanishes everywhere except at the origin, where it is infinite. It also has unit total mass: $\int_{\mathbf{R}} \delta(x)dx = 1$.

Indeed, by letting $\epsilon = \frac{a_n}{\sqrt{n}}$, we obtain

$$\mathbf{P}\big(\sup_{x\in\mathbf{R}}\big|\hat{F}_n(x) - F(x)\big| > \tfrac{a_n}{\sqrt{n}}\big) \le 2\exp\big(-2a_n^2\big),$$

which vanishes if a_n diverges. Since the empirical CDF is a consistent estimator of the CDF, the plug-in estimator is consistent as long as the statistical functional is continuous.

2. The bootstrap. The bootstrap is a general approach to approximating the distribution of a statistic.[2] Let $\mathbf{x}_i \overset{i.i.d.}{\sim} F$. Recall the sampling distribution of a statistic $\mathbf{t}_n := \phi\big(\{\mathbf{x}_i\}_{i\in[n]}\big)$ is

$$G_F^n(t) := \mathbf{P}_{\mathbf{x}_i \overset{i.i.d.}{\sim} F}\big(\phi\big(\{\mathbf{x}_i\}_{i\in[n]}\big) \le t\big).$$

Asymptotic statistics approximates $G_F^n(t)$ by its asymptotic counterpart $G_F^\infty(t)$. The canonical example is approximating the sampling distribution of the MLE by its asymptotic distribution:

$$\sqrt{n}I(\theta)^{1/2}\big(\hat{\theta}_n - \theta\big) \approx \mathcal{N}\big(0, I_p\big).$$

The bootstrap approximates $G_F^n(t)$ by $G_{\hat{F}_n}^n(t)$, where $\hat{F}_n$ is an estimator of F. There are two possible choices:

1. the empirical CDF of the observations: $\hat{F}_n(t) = \frac{1}{n}\sum_{i\in[n]} \mathbf{1}_{(-\infty,t]}(\mathbf{x})$.
2. a parametric estimator of F. If $F = F_\theta$ is a member of a parametric family, a parametric estimator of F is $F_{\hat{\theta}_n}$, where $\hat{\theta}_n$ is an estimator of θ.

The former leads to the non-parametric boostrap, and the latter to the parametric bootstrap. The tradeoff between the two bootstraps is similar to the tradeoff between parametric and non-parametric inference: if the parametric model is correct, the parametric bootstrap is preferable.

Regardless of the choice of estimator, it is usually necessary to evaluate the approximation $G_{\hat{F}_n}^n(t)$ by simulation. Thus there are two sources of error in the bootstrap approximation:

1. the discrepancy between $G_F^n(t)$ and $G_{\hat{F}_n}^n(t)$,
2. the error due to evaluating $G_{\hat{F}_n}^n(t)$ by simulation.

[2]We are using the term statistic loosely. The bootstrap is often used to approximate the distribution of pivots, which are technically not statistics.

Since we are free to draw as many bootstrap samples as we wish, it is possible to make the second source of error negligible.

Thus far, we presented the bootstrap as a general approach to approximating the distribution of a statistic. In practice, we are usually only interested in a functional of the distribution. The obvious bootstrap estimator of $\theta\big(G_F^n\big)$ is $\theta\big(G_{\hat{F}_n}^n\big)$. To wrap up, we describe two common applications of the bootstrap: bias correction and interval estimation.

2.1. *Bootstrap bias correction.* Let $\mathbf{x}_i \overset{i.i.d.}{\sim} F$ and $\mu = \mathbf{E}\big[\bar{\mathbf{x}}_n\big]$. To keep things simple, we assume the target is a scalar $\tau(\mu)$, where $\tau : \mathcal{X} \to \mathbf{R}$ is a (known) function. A consistent estimator of θ is $\hat{\theta}_n = \tau(\bar{\mathbf{x}}_n)$. Its bias is

$$\text{bias}_n(\hat{\theta}) = \mathbf{E}\big[\tau(\bar{\mathbf{x}}_n) - \tau(\mu)\big]. \tag{2.1}$$

The bootstrap pretends the empirical distribution of the observations is the generative distribution. Thus the bootstrap analog of μ is the mean of $\hat{F}_n$, which is $\bar{\mathbf{x}}_n$, and the analog of $\tau(\mu)$ is $\tau(\bar{\mathbf{x}})$. The bootstrap estimator of the bias is

$$\text{bias}^*(\hat{\theta}) = \mathbf{E}^*\big[\tau(\bar{\mathbf{x}}_n^*) - \tau(\bar{\mathbf{x}})\big],$$

where $\mathbf{E}^*$ is the bootstrap expectation. That is,

$$\mathbf{E}^*\Big[g\big(\{\mathbf{x}_i^*\}_{i\in[n]}\big)\Big] := \mathbf{E}_{\mathbf{x}_i^* \overset{i.i.d.}{\sim} \hat{F}_n}\Big[g\big(\{\mathbf{x}_i^*\}_{i\in[n]}\big)\Big].$$

Since $\hat{F}_n$ depends on the observations, $\mathbf{E}^*$ implicitly conditions on the observations.

Algorithm 1 Bootstrap bias estimation

Require: $(\mathbf{x}_i)_{i\in[n]} \in \mathcal{X}^n$
1: **for** $b = 1, 2, \dots$ **do**
2: Draw n *i.i.d.* samples from $\hat{F}_n$: $\mathbf{x}_{b,i}^* \overset{i.i.d.}{\sim} \hat{F}_n$. If $\hat{F}_n$ is the empirical CDF, drawing n *i.i.d.* samples from $\hat{F}_n$ amounts to resampling the observations with replacement.
3: Evaluate $\hat{\theta}_b^* \leftarrow \tau(\bar{\mathbf{x}}^*)$ and the its bias $\hat{\theta}_b^* - \tau(\bar{\mathbf{x}})$.
4: **end for**
5: $\text{bias}^*(\hat{\theta}) \leftarrow \frac{1}{B}\sum_{b\in[B]} \hat{\theta}_b^* - \tau(\bar{\mathbf{x}})$.

Armed with the bootstrap estimator of bias, a bias-corrected estimator is

$$\tilde{\theta} = \hat{\theta} - \text{bias}^*(\hat{\theta}).$$

It is possible to show that, under suitable conditions, the bias of the bias-corrected estimator is of smaller order than that of its uncorrected counterpart. However, because the correction term depends on the observations, the variance of the bias-corrected estimator is generally larger than that of its uncorrected counterpart. Thus the MSE of the bias-corrected estimator is not necessarily smaller.

2.2. *Bootstrap interval setimation.* Consider forming a $1-\alpha$-confidence interval for $\tau(\mu)$. If we knew the distribution of $\tau(\bar{\mathbf{x}}_n)+\tau(\mu)$, a $1-\alpha$-confidence interval is $\mathcal{C}-\tau(\bar{\mathbf{x}}_n)$, where $\mathcal{C}\subset\mathbf{R}$ is a set of $1-\alpha$ mass under the distribution of $\tau(\bar{\mathbf{x}}_n)-\tau(\mu)$. Indeed,

$$\mathbf{P}\big(\tau(\mu)\in\mathcal{C}-\tau(\bar{\mathbf{x}}_n)\big)=\mathbf{P}\big(\tau(\bar{\mathbf{x}}_n)+\tau(\mu)\in\mathcal{C}\big)=1-\alpha.$$

The bootstrap approximates the distribution of $\tau(\bar{\mathbf{x}}_n)+\tau(\mu)$ by that of $\tau(\bar{\mathbf{x}}_n^*)+\tau(\bar{\mathbf{x}}_n)$, which allows us to choose a set $\mathcal{C}^*\subset\mathbf{R}$ that has $1-\alpha$ mass under the bootstrap distribution. We "invert" the set to obtain a confidence interval:

$$\begin{aligned}\mathbf{P}\big(\tau(\mu)\in\mathcal{C}^*-\tau(\bar{\mathbf{x}}_n)\big)&=\mathbf{P}\big(\tau(\bar{\mathbf{x}}_n)+\tau(\mu)\in\mathcal{C}^*\big)\\&\approx\mathbf{P}^*\big(\tau(\bar{\mathbf{x}}_{n,b}^*)+\tau(\bar{\mathbf{x}}_n)\in\mathcal{C}^*\big)\\&=1-\alpha,\end{aligned}$$

where $\mathbf{P}^*$ is the bootstrap probability.

We remark that $\tau(\bar{\mathbf{x}}_n)-\tau(\mu)$ is not the only statistic whose distribution allows us to obtain a confidence interval. As we shall see, the bootstrap is better at approximating the distribution of asymptotic pivots. Since the coverage of the bootstrap confidence intervals depend on the accuracy of the bootstrap approximation, we should bootstrap asymptotic pivots.

3. Consistency of the bootstrap.

DEFINITION 3.1. *We say the bootstrap approximation to $G_F^n(t)$ is consistent if*

$$\sup_{t\in\mathbf{R}^n}\big|G_{\hat{F}_n}^n(t)-G_F^n(t)\big|\xrightarrow{p}0.$$

It is possible to show that the bootstrap is consistent under fairly general conditions. Intuitively, the discrepancy between $G_F^\infty(t)$ and its bootstrap approximation consists of two parts:

$$\big|G_{\hat{F}_n}^n(t)-G_F^n(t)\big|\le\big|G_{\hat{F}_n}^n(t)-G_F^\infty(t)\big|+\big|G_F^n(t)-G_F^\infty(t)\big|. \tag{3.1}$$

By the Glivenko-Cantelli theorem, the second term is negligible. As long as $\hat{F}_n$ is a consistent approximation of F, i.e.

$$\sup_{t\in\mathbf{R}}\big|\hat{F}_n(t)-F(t)\big|\xrightarrow{p}0,$$

and $G_F^n(t)$ is continuous in F, the first term should also be negligible. Here is a theorem that formalize the preceding argument.

THEOREM 3.2. *Assume*

1. $\rho(F, \hat{F}_n) \sim o_P(1)$, *where* ρ *is the Mallows metric,*[3]
2. $G_F^\infty(t)$ *is continuous in* t *for any* $F \in \mathcal{F}$,
3. $G_{F_n}^n(t) \to G_F^\infty(t)$ *for any sequence* $\{F_n\} \in \mathcal{F}$ *that converges to* F *in the Mallows metric.*

Under the preceding assumptions,

$$\sup_{t\in\mathbf{R}}\big|G_{\hat{F}_n}^n(t) - G_F^n(t)\big| \xrightarrow{p} 0.$$

Although the bootstrap is consistent under fairly general conditions, there are non-pathological examples of its inconsistency. Generally speaking, the bootstrap tends to fail if $G_F^\infty(t)$ is discontinuous in F.

3.1. *Higher-order consistency.* The preceding section shows that bootstrapping leads to a good approximation of $G_F^n(t)$. However, as remarked in Section 2, asymptotic statistics provide similar benefits. As we shall see, the bootstrap approximation is more accurate than its asymptotic counterpart.

The theory behind the higher-order consistency of the boostrap is based on *Edgeworth expansions*. These are series expansions (similar in spirit to Taylor expansions) of the CDF of a statistic around the CDF of its asymptotic distribution:

$$G_F^n(t) = G_F^\infty(t) + \frac{g_F(t)}{\sqrt{n}} + \frac{g_F'(t)}{n} + O\big(\tfrac{1}{n^{3/2}}\big). \tag{3.2}$$

We recognize the first term as the asymptotic approximation to $G_F^n(t)$.

Assuming $G_F^n(t)$ and its bootstrap approximation $G_{\hat{F}_n}^n(t)$ both admit an expansion of the form (3.2),

$$\begin{aligned} G_{\hat{F}_n}^n(t) - G_F^n(t) &= G_{\hat{F}_n}^\infty(t) - G_F^\infty(t) + \tfrac{1}{\sqrt{n}}\big(g_{\hat{F}_n}(t) - g_F(t)\big) \\ &\quad + \tfrac{1}{n}\big(g_{\hat{F}_n}'(t) - g_F'(t)\big) + O\big(\tfrac{1}{n^{3/2}}\big). \end{aligned}$$

By the Glivenko-Cantelli theorem,

$$\sup_{x\in\mathbf{R}}\big|\hat{F}_n(x) - F(x)\big| \sim O_P\big(\tfrac{1}{\sqrt{n}}\big).$$

[3]The Mallows metric is $\rho(P_1, P_2) := \inf\big\{\mathbf{E}_P\big[\|x-y\|_2^2\big]\big\}$. The infimum is over all joint distributions P of $(\mathbf{x}, \mathbf{y})$ whose marginals are P_1 and P_2.

Thus we expect

$$G^{\infty}_{\hat{F}_n}(t) - G^{\infty}_F(t) \sim O_P\big(\tfrac{1}{\sqrt{n}}\big),$$

which suggests the bootstrap incurs an error of size $O\big(\frac{1}{\sqrt{n}}\big)$. Turning our attention to the asymptotic approximation, we rearrange (3.2) to obtain

$$G^n_F(t) - G^{\infty}_F(t) = \frac{1}{\sqrt{n}} g_F(t) + O\big(\tfrac{1}{n}\big) \sim O\big(\tfrac{1}{\sqrt{n}}\big).$$

Thus the bootstrap approximation is generally as accurate as its asymptotic counterpart.

The preceding analysis shows that the error of the bootstrap is dominated by the discrepancy between $G^{\infty}_{\hat{F}_n}(t)$ and $G^{\infty}_F(t)$. We observe that the discrepancy vanishes if $\mathbf{t}_n$ is an asymptotic pivot. For example, let

$$\mathbf{t}_n = \frac{\sqrt{n}(\bar{\mathbf{x}}_n - \mu)}{\sigma},$$

where μ and σ^2 are the mean and variance of $\mathbf{x}_i$. The bootstrap analog of $\mathbf{t}_n$ is

$$\mathbf{t}^*_n = \frac{\sqrt{n}(\bar{\mathbf{x}}^*_n - \bar{\mathbf{x}}_n)}{\mathbf{s}_n},$$

By the CLT, the limiting distribution of $\mathbf{t}_n$ and $\mathbf{t}^*_n$ are both standard normal! In such cases, the error of the bootstrap is dominated by the subsequent terms in the CDF expansion:

$$\frac{1}{\sqrt{n}}\big(g_{\hat{F}_n}(t) - g_F(t)\big).$$

By the Glivenko-Cantelli theorem, we expect

$$g_{\hat{F}_n}(t) - g_F(t) \sim O_P\big(\tfrac{1}{\sqrt{n}}\big),$$

which suggests the error of the bootstrap is $O_P\big(\frac{1}{n}\big)$. On the other hand, the error of the asymptotic approximation remains $O\big(\frac{1}{\sqrt{n}}\big)$ regardless of whether $\mathbf{t}_n$ is an asymptotic pivot. Thus the bootstrap approximation is more accurate than its asymptotic counterpart if the statistic is an asymptotic pivot.

Example 3.3. *Let $\mathbf{x}_i \overset{\text{i.i.d.}}{\sim} f(x-\mu)$, where $f(x)$ is the density of a mean zero, variance σ^2 random variable. We wish to infer the location parameter μ. Consider two $1-\alpha$-confidence intervals for μ:*

1. *the Wald interval:* $\Big[\bar{\mathbf{x}}_n - \frac{z_{\alpha/2}}{\sqrt{n}}\hat{\sigma}_n, \bar{\mathbf{x}}_n + \frac{z_{\alpha/2}}{\sqrt{n}}\hat{\sigma}_n\Big]$.

2. *the bootstrap interval:* $\Big[\bar{\mathbf{x}}_n - \frac{t^*}{\sqrt{n}}\hat{\sigma}_n, \bar{\mathbf{x}}_n + \frac{t^*}{\sqrt{n}}\hat{\sigma}_n\Big]$, *where* t^* *is the* $1-\frac{\alpha}{2}$ *quantile of the bootsrap approximation to the distribution of the pivot* $\frac{\sqrt{n}(\bar{\mathbf{x}}_n-\mu)}{\sigma}$.

Since

$$\mu \in \Big[\bar{\mathbf{x}}_n - \tfrac{z_{\alpha/2}}{\sqrt{n}}\hat{\sigma}_n, \bar{\mathbf{x}}_n + \tfrac{z_{\alpha/2}}{\sqrt{n}}\hat{\sigma}_n\Big] \iff \frac{\sqrt{n}(\bar{\mathbf{x}}_n-\mu)}{\hat{\sigma}_n} \in \big[-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\big],$$

the coverage of the Wald intervals depends on the accuracy of the asymptotic approximation to the distribution of the pivot. It is possible to show that the CDF of the pivot admits an Edgeworth expansion:

$$G_F^n(t) = \Phi(t) - \phi(t)\Big(\frac{\kappa_3}{6\sqrt{n}}h_2(t) + \frac{\kappa_4}{24n}h_3(t) + \frac{\kappa_3^2}{72n}h_5(t) + O\big(\tfrac{1}{n^{3/2}}\big)\Big), \tag{3.3}$$

Thus the coverage of the Wald interval is

$$\begin{aligned}&\mathbf{P}\Big(\mu \in \Big[\bar{\mathbf{x}}_n - \tfrac{z_{\alpha/2}}{\sqrt{n}}\hat{\sigma}_n, \bar{\mathbf{x}}_n + \tfrac{z_{\alpha/2}}{\sqrt{n}}\hat{\sigma}_n\Big]\Big)\\ &\quad= \mathbf{P}\big(\tfrac{\sqrt{n}(\bar{\mathbf{x}}_n-\mu)}{\hat{\sigma}_n} \in \big[-z_{\alpha/2}, z_{\alpha/2}\big]\big)\\ &\quad= G_F^n\big(z_{\frac{\alpha}{2}}\big) - G_F^n\big(-z_{\frac{\alpha}{2}}\big),\end{aligned}$$

which, by the Edgeworth expansion of $G_F^n(t)$, *is*

$$= \underbrace{\Phi\big(z_{\frac{\alpha}{2}}\big) - \Phi\big(-z_{\frac{\alpha}{2}}\big)}_{1-\alpha} + O\big(\tfrac{1}{\sqrt{n}}\big).$$

By a similar argument, it is possible to show the coverage of the bootstrap interval depends on the accuracy of the bootstrap approximation to $G_F^n(t)$:

$$\mathbf{P}\Big(\mu \in \Big[\bar{\mathbf{x}}_n - \tfrac{t^*}{\sqrt{n}}\hat{\sigma}_n, \bar{\mathbf{x}}_n + \tfrac{t^*}{\sqrt{n}}\hat{\sigma}_n\Big]\Big) = G_F^n\big(t^*\big) - G_F^n\big(-t^*\big).$$

It is also possible to show that the CDF of the bootstrap distribution $G_{\hat{F}_n}^n(t)$ *admits an Edgeworth expansion. Since* $\frac{\sqrt{n}(\bar{\mathbf{x}}_n-\mu)}{\sigma}$ *is an asymptotic pivot,*

$$G_F^\infty(t) = G_{\hat{F}_n}^\infty(t) + O_P\big(\tfrac{1}{n}\big).$$

Thus the coverage of the bootstrap interval is

$$\mathbf{P}\Big(\mu \in \Big[\bar{\mathbf{x}}_n - \tfrac{t^*}{\sqrt{n}}\hat{\sigma}_n, \bar{\mathbf{x}}_n + \tfrac{t^*}{\sqrt{n}}\hat{\sigma}_n\Big]\Big) = \underbrace{G_{\hat{F}_n}^n\big(t^*\big) - G_{\hat{F}_n}^n\big(-t^*\big)}_{1-\alpha} + O_P\big(\tfrac{1}{n}\big).$$

We remark that the second-order consistency of the bootstrap intervals depends crucially on $\frac{\sqrt{n}(\bar{\mathbf{x}}_n-\mu)}{\sigma}$ *being an asymptotic pivot.*

In the preceding example, the bootstrap intervals are, in fact, even more accurate: their coverage is $1-\alpha+o_P\big(\frac{1}{n}\big)$. The reason is the $O\big(\frac{1}{\sqrt{n}}\big)$ term in the expansion of $G_F^\infty(t)-G_F^\infty(-t)$ vanishes. Indeed,

$$\begin{aligned}
&G_F^n(t)-G_F^n(-t)\\
&\quad= G_F^\infty(t)+\phi(t)\big(\tfrac{\kappa_3}{6\sqrt{n}}h_2(t)+\tfrac{\kappa_4}{24n}h_3(t)+\tfrac{\kappa_3^2}{72n}h_5(t)+O\big(\tfrac{1}{n^{3/2}}\big)\big)\\
&\qquad-\big(G_F^\infty(-t)+\phi(-t)\big(\tfrac{\kappa_3}{6\sqrt{n}}h_2(-t)+\tfrac{\kappa_4}{24n}h_3(-t)+\tfrac{\kappa_3^2}{72n}h_5(-t)+O\big(\tfrac{1}{n^{3/2}}\big)\big)\big).
\end{aligned}$$

Since $\phi(t)$ and h_2 are even functions, and h_3, h_5 are odd, the expansion simplifies to

$$= \Phi(t)-\Phi(-t)+\phi(t)\big(\tfrac{\kappa_4}{12n}h_3(t)+\tfrac{\kappa_3^2}{36n}h_5(t)+O\big(\tfrac{1}{n^{3/2}}\big)\big).$$

The error term is, in fact, $O\big(\frac{1}{n^2}\big)$ because the $O\big(\frac{1}{n^{3/2}}\big)$ terms vanish by a similar argument. The takeaway is to reap the benefits of the higher-order consistency of the bootstrap, it is necessary to bootstrap asymptotic pivots.

APPENDIX A: EDGEWORTH EXPANSIONS

Recall the proof of the central limit theorem. We begin with the characteristic function of a real-valued random variable $\mathbf{x}$: $\mathbf{E}\big[e^{it\mathbf{x}}\big]$, where i is the imaginary unit. Although the characteristic function is generally complex-valued, by the fact that

1. $e^{itx}=\cos(tx)+i\sin(tx)$,
2. sin is an odd function,

the characteristic function of a random variable whose density is even is real-valued. As long as the k-th moment of $\mathbf{x}$ is finite, it is possible to show that

$$\log\mathbf{E}\big[e^{it\mathbf{x}}\big]=\sum_{j\in[k]}\frac{\kappa_j}{k!}(it)^k+o\big(|t|^r\big),$$

where κ_j's are cumulants of $\mathbf{x}$. The first two cumulants are the mean $\mathbf{E}\big[\mathbf{x}\big]$ and the variance $\mathbf{var}\big[\mathbf{x}\big]$, and we assume they are finite.

The characteristic function of a sum of *i.i.d.* random variables is

$$\mathbf{E}\Big[e^{it\sum_{j\in[n]}\mathbf{x}_j}\Big]=\mathbf{E}\Big[\textstyle\prod_{j\in[n]}e^{it\mathbf{x}_j}\Big]=\displaystyle\prod_{j\in[n]}\mathbf{E}\Big[e^{it\mathbf{x}_j}\Big].$$

Its logarithm is

$$\log\mathbf{E}\Big[e^{it\sum_{j\in[n]}\mathbf{x}_j}\Big]=\sum_{j\in[n]}\log\mathbf{E}\Big[e^{it\mathbf{x}_j}\Big],$$

which, by its Taylor expansion, is

$$= \sum_k \frac{\kappa_j}{k!} n(it)^k + o\big(n|t|^k\big),$$

If $\mathbf{E}\big[\mathbf{x}_j\big] = 0$, the first term vanishes. If also assume $\mathbf{var}\big[\mathbf{x}_j\big] = 1$,

$$\log \mathbf{E}\Big[e^{\frac{it}{\sqrt{n}}\sum_{j\in[n]}\mathbf{x}_j}\Big] = \sum_{j\in[n]} -\frac{t^2}{2n} + o\big(\tfrac{|t|^2}{n}\big) = -\frac{t^2}{2} + o\big(|t|^2\big).$$

Thus the characteristic function of $\frac{1}{\sqrt{n}}\sum_{i\in[n]}\mathbf{x}_i$ is

$$\mathbf{E}\Big[e^{\frac{it}{\sqrt{n}}\sum_{j\in[n]}\mathbf{x}_j}\Big] = \exp\big(-\tfrac{t^2}{2} + o\big(|t|^2\big)\big).$$

We recognize $\exp\big(-\frac{t^2}{2}\big)$ as the characteristic function of a $\mathcal{N}(0,\kappa_2)$ random variable. By Lévy's continuity theorem,

$$\frac{1}{\sqrt{n}}\sum_{j\in[n]}\mathbf{x}_j \xrightarrow{d} \mathcal{N}(0,\kappa_2),$$

which is the familiar CLT.

It is possible to approximate the log-characteristic function more accurately by incorporating higher order terms of the Taylor expansion (assuming the higher cumulants are finite):

$$\begin{aligned}\log \mathbf{E}\Big[e^{\frac{it}{\sqrt{n}}\sum_{j\in[n]}\mathbf{x}_j}\Big] &= \sum_{j\in[n]} -\frac{\kappa_2}{2}\frac{t^2}{n} + \frac{\kappa_3}{6}\frac{(it)^3}{n^{3/2}} + \frac{\kappa_4}{24}\frac{(it)^4}{n^2} + o\big(\tfrac{|t|^4}{n^2}\big)\\ &= -\frac{t^2}{2} + \frac{\kappa_3}{6}\frac{(it)^3}{\sqrt{n}} + \frac{\kappa_4}{24}\frac{(it)^4}{n} + o\big(\tfrac{|t|^4}{n}\big).\end{aligned} \tag{A.1}$$

The improved approximation of the characteristic function is

$$\begin{aligned}\mathbf{E}\Big[e^{\frac{it}{\sqrt{n}}\sum_{j\in[n]}\mathbf{x}_j}\Big] &= \exp\Big(-\frac{t^2}{2} + \frac{\kappa_3}{6}\frac{(it)^3}{\sqrt{n}} + \frac{\kappa_4}{24}\frac{(it)^4}{n} + o\big(\tfrac{|t|^4}{n}\big)\Big)\\ &= e^{-\frac{t^2}{2}}\exp\Big(\frac{\kappa_3}{6}\frac{(it)^3}{\sqrt{n}} + \frac{\kappa_4}{24}\frac{(it)^4}{n} + o\big(\tfrac{|t|^4}{n}\big)\Big),\end{aligned}$$

which, by the Taylor expansion $e^x = 1 + x + \frac{x^2}{2} + o\big(x^2\big)$, is

$$= e^{-\frac{t^2}{2}}\Big(1 + \frac{\kappa_3}{6}\frac{(it)^3}{\sqrt{n}} + \frac{\kappa_4}{24}\frac{(it)^4}{n} + \frac{\kappa_3^2}{72}\frac{(it)^6}{n} + o\big(\tfrac{|t|^4}{n}\big)\Big).$$

We invert the Fourier transform to obtain a higher-order approximation to the density of $\frac{1}{\sqrt{n}}\sum_{i\in[n]}\mathbf{x}_j$:

$$g_n(x) = \phi(x)\Big(1 + \frac{\kappa_3}{6\sqrt{n}}h_3(x) + \frac{\kappa_4}{24n}h_4(x) + \frac{\kappa_3^2}{72n}h_6(x) + o\big(\tfrac{1}{n}\big)\Big),$$

where h_k is the k-th Hermite polynomial. We integrate to obtain the (second-order) Edgeworth expansion of the CDF:

$$G_n(x) = \Phi(x) - \phi(x)\Big(\frac{\kappa_3}{6\sqrt{n}}h_2(x) + \frac{\kappa_4}{24n}h_3(x) + \frac{\kappa_3^2}{72n}h_5(x) + o\big(\tfrac{1}{n}\big)\Big).$$

If we keep higher order terms in (A.1), the preceding derivation yields higher order Edgeworth expansions.

Yuekai Sun
Berkeley, California
December 5, 2015