Some of my white board writing from week 12 (1)

1) I continued the discussion after formula (25) on p. 71 of the notes. I pointed out that when applying this formula for a specific random variable which is the arithmetic mean of n i.i.d. variables $X_1, X_2, \ldots, X_n$, and by utilizing the relationship between cumulant generating functions as given in Exercise 1 on p. 66 (i.e. $K_{\sum_{i=1}^n X_i}(t) = nK_{X_1}(t)$) we get the saddlepoint approximation formula for the density $f(\bar{x})$ of $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ as:

$$\hat{f}(\bar{x}) \approx \sqrt{\frac{n}{2\pi K_X''(\hat{t})}}\, e^{\{nK_X(\hat{t}) - n\hat{t}\bar{x}\}} \left\{1 + \left[\frac{1}{8n}\hat{\rho}_4 - \frac{5}{24}\hat{\rho}_3^2\right]\right\}$$

Here K_X is the cumulant generating function of a single observation X_i, $\hat{t}$ is the saddle-point (i.e. it has to be re-calculated for every value $\bar{x}$ of the argument as the solution to $K_X'(\hat{t}) = \bar{x}$,

and $\hat{\rho}_i = n^{1-\frac{i}{2}}\rho_i(\hat{t})$, with $\rho_i(t) = \frac{K_X^{(i)}(t)}{[K_X''(t)]^{i/2}}$, $i \ge 3$

Even the simpler version of the above second order saddlepoint approximation, namely the first order one:

$$\hat{f}(\bar{x}) \approx \sqrt{\frac{n}{2\pi K_X''(\hat{t})}} \cdot e^{nK_X(\hat{t}) - n\hat{t}\bar{x}}$$

is extremely accurate for sample sizes such as 5, 6, 10.

2) Then I discussed the structure (not the derivation) of a similar formula for the CDF of $\bar{X}$ (called the Lugannani–Rice formula:

$$F_{\bar{X}}(\bar{x}) = P(\bar{X} \leq \bar{x}) = \Phi(\hat{\omega}_n) + \varphi(\hat{\omega}_n)\left(\frac{1}{\hat{\omega}_n} - \frac{1}{\hat{u}_n}\right) + O\left(\frac{1}{n}\right)$$

Here $\hat{w}_n = \text{sgn}(\hat{t})\sqrt{2n[\hat{t}\bar{x} - K_X(\hat{t})]}$, $\hat{u}_n = \hat{t}\sqrt{nK_X''(\hat{t})}$, where again $\hat{t}$ is the saddlepoint, i.e. $K_X'(\hat{t}) = \bar{x}$ holds.

3.) I then plugged in the formulae above on 2 occasions:

i) Example 1: (Saddlepoint approximation for the sample mean of the standard normal: in this case the approximation is precise, i.e. no error):

$f(x) = \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. For this, the cgf is known to be $K_X(t) = \log M_X(t) = \log\left(e^{\frac{t^2}{2}}\right) = \frac{t^2}{2}$

Hence $K_X'(t) = t$; $K_X''(t) = 1$. The saddlepoint equation gives $\hat{t} = \bar{x}$, hence the first order approximation for the density becomes $\hat{f}(\bar{x}) = \sqrt{\frac{n}{2\pi}}\, e^{n\frac{(\bar{x})^2}{2} - n(\bar{x})^2} = \sqrt{\frac{n}{2\pi}}\, e^{-\frac{n(\bar{x})^2}{2}}$

which is precisely the density of $N(0, \frac{1}{n})$ (and we know that $\bar{X} \sim N(0, \frac{1}{n})$ in this case).

For the cdf we get: $\hat{w}_n = \text{sgn}(\bar{x})\sqrt{2n\left((\bar{x})^2 - \frac{(\bar{x})^2}{2}\right)} = \bar{x}\sqrt{n}$

and $\hat{u}_n = \hat{w}_n = \bar{x}\sqrt{n}$, hence $\varphi(\hat{w}_n)\left(\frac{1}{\hat{w}_n} - \frac{1}{\hat{u}_n}\right) = 0$ in this case and we get

$F_{\bar{X}}(\bar{x}) = \Phi(\bar{x}\sqrt{n})$ which is the cdf of $N(0, \frac{1}{n})$ (again precise)

ii) <u>Example 2</u> Saddlepoint approximation for the density of the sample mean of n i.i.d. observations from the Gamma$(\alpha,1)$ density.

The density of a single observation is $f(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x}$, $x > 0$.

Hence if $t<1$: $M_X(t) = \frac{1}{\Gamma(\alpha)} \int_0^\infty e^{-(1-t)x} x^{\alpha-1} dx = \frac{1}{\Gamma(\alpha)(1-t)^\alpha} \int_0^\infty e^{-y} y^{\alpha-1} dy$

set $(1-t)x = y$

$dx = \frac{dy}{1-t}$

$= \frac{\Gamma(\alpha)}{\Gamma(\alpha)} \cdot \frac{1}{(1-t)^\alpha} = (1-t)^{-\alpha}$

Hence $K_X(t) = -\alpha \log(1-t)$, $K_X'(t) = \frac{\alpha}{1-t}$, $K_X''(t) = \frac{\alpha}{(1-t)^2}$

Saddlepoint equation is: $\frac{\alpha}{1-\hat{t}} = \bar{x}$, i.e. $\hat{t} = 1 - \frac{\alpha}{\bar{x}}$.

Hence $K_X''(\hat{t}) = \frac{(\bar{x})^2}{\alpha}$ and substituting in the first-order saddlepoint approximation formula we get

$$\hat{f}(\bar{x}) = \sqrt{\frac{n\alpha}{2\pi(\bar{x})^2}} \exp\left[-n\alpha \log\left(\frac{\alpha}{\bar{x}}\right) - n(\bar{x}-\alpha)\right] =$$

$$= \left(\sqrt{\frac{2\pi}{n\alpha}}(n\alpha)^{n\alpha} e^{-n\alpha}\right)^{-1} (n\bar{x})^{n\alpha-1} e^{-n\bar{x}} \cdot n$$

Now: $\frac{1}{\Gamma(n\alpha)}(n\bar{x})^{n\alpha-1} e^{-n\bar{x}} \cdot n$ is the <u>exact</u> density of $\bar{X}$

(because $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$, and $\sum_{i=1}^n X_i \sim$ Gamma$(n\alpha, 1)$).

Hence the difference between the <u>exact</u> and the approximation is that the norming $\frac{1}{\Gamma(n\alpha)}$ has been replaced by $\frac{1}{\sqrt{\frac{2\pi}{n\alpha}}(n\alpha)^{n\alpha} e^{-n\alpha}}$

However the <u>famous Stirling</u> approximation of the Gamma function tells us precisely that

$$\Gamma(n\alpha) \approx \sqrt{\frac{2\pi}{n\alpha}} (n\alpha)^{n\alpha} e^{-n\alpha}$$

Hence the difference is only that the norming constant in the true density $\left(\frac{1}{\Gamma(n\alpha)}\right)$ has been replaced by its Very accurate approximation given by Stirling.

3) In the lecture about robustness, I first discussed the fact that $\bar{X}$ is a disastrously bad "estimator" for the location parameter of the Cauchy distribution.
Indeed, it is known that the characteristic function of a single observation $X_1 \sim$ Cauchy (θ) with density $f(x,\theta) = \frac{1}{\pi} \cdot \frac{1}{1+(x-\theta)^2}$ is $\varphi_{X_1}(t) = e^{i\theta t - |t|}$
Hence for the characteristic function of $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ we have

$$\varphi_{\bar{X}}(t) = E e^{i\frac{t}{n}\left(\sum_{i=1}^{n} X_i\right)} \underset{\uparrow \text{ using independence}}{=} \left[E e^{i\left(\frac{t}{n}\right)X_1}\right]^n = \left(e^{i\frac{\theta}{n}t - \frac{|t|}{n}}\right)^n =$$

$$= e^{i\theta t - |t|} = \varphi_{X_1}(t)$$

that is, $\bar{X}$ has the <u>same</u> distribution as a single X_1 <u>no matter</u> what the sample size! Hence $\bar{X}$ <u>can not</u> be consistent for θ.
With respect to the remaining discussions, I believe that my derivations in the lecture notes are detailed enough to reproduce them again here. Finally, I note that the asymptotic variance for the sample quantile as given by $V(T,F) = \frac{p(1-p)}{f(q_p)^2}$ on p. 78 of the notes,

specialises to

$$\frac{\frac{1}{2}\left(1-\frac{1}{2}\right)}{f(0)^2}$$ when applied for the median $(p=\frac{1}{2})$ of the location family $f(x-\theta)=f(x,\theta)$. In particular this justifies the result

$$\sqrt{n}(\tilde{\theta}_n-\theta) \xrightarrow{d} N\left(0, \frac{1}{4f^2(0)}\right)$$ which I

was using on p.75 of the notes (formula (29)).

For the normal location family this gave us:

$$\sqrt{n}(\tilde{\theta}_n-\theta) \xrightarrow{d} N\left(0, \frac{\pi}{2}\right)$$, i.e. the empirical median for the standard normal is asympotically normally distributed with asymptotic variance $\frac{\pi}{2} > 1$ and (as we know), is less efficient than the empirical (sample) mean for which we have precisely

$$\sqrt{n}(\bar{X}-\theta) \sim N(0,1).$$

This illustrates the claim that we "pay for robustness by a slight loss of efficiency".