Ex0012: Large deviation theory

Submitted by: Neta Shaul

The problem:

Consider a set of N random variables $\hat{x}_j$. Each variable can get the values 1 or 0 with probabilities p and $q = 1 - p$ respectively. Define $\hat{x} = (1/N)\sum_j \hat{x}_j$. Define $\bar{F}(x) = \text{Prob}(\hat{x} > x)$.

1) Find an approximation $\bar{F}(x) \sim exp[-S(x)]$, in terms of an explicit elementary function $S_{CLT}(x)$, based on the central limit theorem (CLT). Note that the exact result is an error function.

2) Find an approximation $\bar{F}(x) \sim exp[-S(x)]$, in terms of an explicit function $S_{LDT}(x)$, based on large deviation theory (LDT).

3) Compare the CLT and LDT approximation at the edges of the range of $\hat{x}$.

   (a) Include analytical analysis of the behaviour at $x \to 1/0$

   (b) Include a Graphical analysis using a Semi-log scale plot of the approximation of $\bar{F}(x)$

4) Compare the CLT and LDT approximation around the expected value $\langle x_1 \rangle$.

   (a) Include analytical analysis of the behaviour around $x \approx \langle x_1 \rangle$ (Taylor expansion).

   (b) Include a Graphical analysis using a plot of the approximation of $\bar{F}(x)$

5) What would be the answers if the $\hat{x}_j$ had normal probability distribution with the same average and variance?

6) Repeat section 1-4 for exponential distribution, $x_1 \sim Exp(\alpha)$.

7) Prove that for every distribution that satisfy the LDT requirements, S_{CLT} and S_{LDT} are equal up to the second order.

The solution:

1) Let $\hat{x} = \dfrac{1}{n}\displaystyle\sum_{j=1}^{n} x_j$ and let us first consider x_j to be a general i.i.d random variables and extract a more practical result from the Central Limit Theorem (CLT).
denote:

$$\mu = \langle x_1 \rangle \text{ and } \sigma^2 = Var(x_1)$$

We saw in class that the CLT states:

$$\hat{y} := \frac{\sum_{j=1}^{n} x_j - n\mu}{\sqrt{n}\sigma} \quad then \quad \lim_{n\to\infty} \hat{y} \sim N(0,1)$$

A short calculation using change of integration variables shows that for any variable $\hat{z} \sim N(0,1)$:

$$a\hat{z} + b \sim N(b, a^2) \tag{1}$$

Noticing that $\hat{x} = \frac{\sigma}{\sqrt{n}}\hat{y} + \mu$ we conclude that for large enough n :

$$\hat{x} \sim N(\mu, \frac{\sigma^2}{n}) \tag{2}$$

Back to our case where $x_1 \sim Ber(p)$, and so have:

$$\mu = p \text{ and } = \sigma^2 = p(1-p)$$

At last we note the $\bar{F}(x)$ as defined above is the tail destribution (ccdf) and we get our two approximations, the first which is excatly CLT:

(i) Exact: $F(x) \approx 1 - \frac{1}{\sqrt{\frac{p(1-p)}{n}}} \int_{-\infty}^{x} exp\left(-\frac{(t-p)^2}{\frac{2p(1-p)}{n}}\right) \frac{dt}{\sqrt{2\pi}}$

and the second which is the asymptotic approximation for *erf* and is more convenient to present it with the Folded CDF defined as $C(x) = min\{\bar{F}(x), 1-\bar{F}(x))\}$:

(ii) Elementary function: $\boxed{C(x) \sim exp\left(-\frac{(x-p)^2}{\frac{2p(1-p)}{n}}\right), \text{ where } S_{CLT}(x) = -\frac{n}{2p(1-p)}(x-p)^2}$

Note: $1 - \bar{F}(x)$ is exactly $F(x)$ the cdf.

2) Since Large Deviations theory (LDT) is not so commonly known as CLT, I will give a formal phrase of a theorem which is the main result of the theory.

Cramer's theorem: *Define* $\hat{x}$ *as above and assume that*

$$Z(\lambda) := \langle e^{\lambda x_1} \rangle < \infty \ \forall \lambda \in \mathbb{R}$$

and define:

$$I(x) := \sup_{\lambda \in \mathbb{R}} [x\lambda - log(Z(\lambda))]$$

then

$$(i) \ \lim_{n\to\infty} \frac{1}{n} log\mathbb{P}[\hat{x} \geq x] = -I(x), \ \forall x > \mu$$

$$(ii) \ \lim_{n\to\infty} \frac{1}{n} log\mathbb{P}[\hat{x} \leq x] = -I(x), \ \forall x < \mu$$

Now let us calculate $I(x)$ for our case which is $\hat{x_1} \sim Ber(p)$:

$$Z(\lambda) = \langle e^{\lambda x_1} \rangle = pe^{\lambda} + (1-p)$$

insert this result to the definition of $I(x)$ gives:

$$I(x) = \sup_{\lambda \in \mathbb{R}} [x\lambda - log(pe^{\lambda} + (1-p))]$$

finding $\lambda(x)$ which gives the supremum is done by requiring that:

$$\frac{d}{d\lambda}\left(x\lambda - log(pe^{\lambda} + (1-p))\right) = 0$$

which gives:

$$I(x) = xlog(x) + (1-x)log(1-x) - xlog(p) - (1-x)log(1-p) \tag{3}$$

At last by Cremar's theorem we get our approximation for a large enough n

(i) $\bar{F}(x) \approx e^{-nI(x)}$, $\forall x > p$

(ii) $\bar{F}(x) \approx 1 - e^{-nI(x)}$, $\forall x < p$

A more compact and maybe more illuminating way to express the above is:

$$\boxed{C(x) \sim e^{-nI(x)}, \text{ and } S_{LDT} = -nI(x)}$$

3) In this section we would like to compare the results we got from the previous sections, both graphically and analytically.
Before we begin we shall note that exact expression for $\bar{F}(x)$ is:

$$\bar{F}(x) = \mathbb{P}[\hat{x} > x] = \sum_{k=\lceil x \rceil}^{n} \binom{n}{k} p^k (1-p)^{n-k}$$

this result is easily reached using combinatorial arguments and since it doesn't contribute to the goal of this exercise we shall leave it for the reader.

Let us begin with the analytical analysis and examine the behaviour at the edges, $x \to 1$:

$$S_{CLT}(1) := \lim_{x \to 1} S_{CLT}(x) = -n\frac{1-p}{2p}$$

$$S_{LDT}(1) := \lim_{x \to 1} S_{LDT}(x) = -nlog(p)$$

so looking at the behaviour at the edge as a function of p we find that:

$$\lim_{p \to 0} \frac{S_{CLT}(1)}{S_{LDT}(1)} = \infty$$

thus we conclude that for a small p the CLT approximation would yield smaller values then the LDT approximation.
Note that a similar behaviour is obtained by letting $x \to 0$ and $p \to 1$.

Now let us look at the log-scale plot of $C(x)$ for a small p in Figure 1 (b), comparing the plots a-d, we can see that at the edges LDT gives us a much better approximation then the CLT approximation.
In addition both CLT approximation - *erf* and the asymptotic approximation, give a similar result.
Last we note that the LDT line is above the CLT line as expected from the previous analysis.

It is also interesting to look at same log-scale plot of C(x) but for a more balnced value of p as in Figure 2 (b), this time the graph is more centered and both approximation give a similar result.

4)We would like to examine behaviour around the expected value as well, again we begin with an analytical analysis. It is enlightening to expend $S_{LDT}(x)$ around $\mu = p$ in Taylor series, which in leading order is:

$$S_{LDT}(x) \approx \frac{n}{2p(1-p)}(x-p)^2 = S_{CLT}(x)$$

leading us to conclude that around the expected value S_{CLT} and S_{LDT} are equal up to a leading order thus we shall expect a similar behaviour.

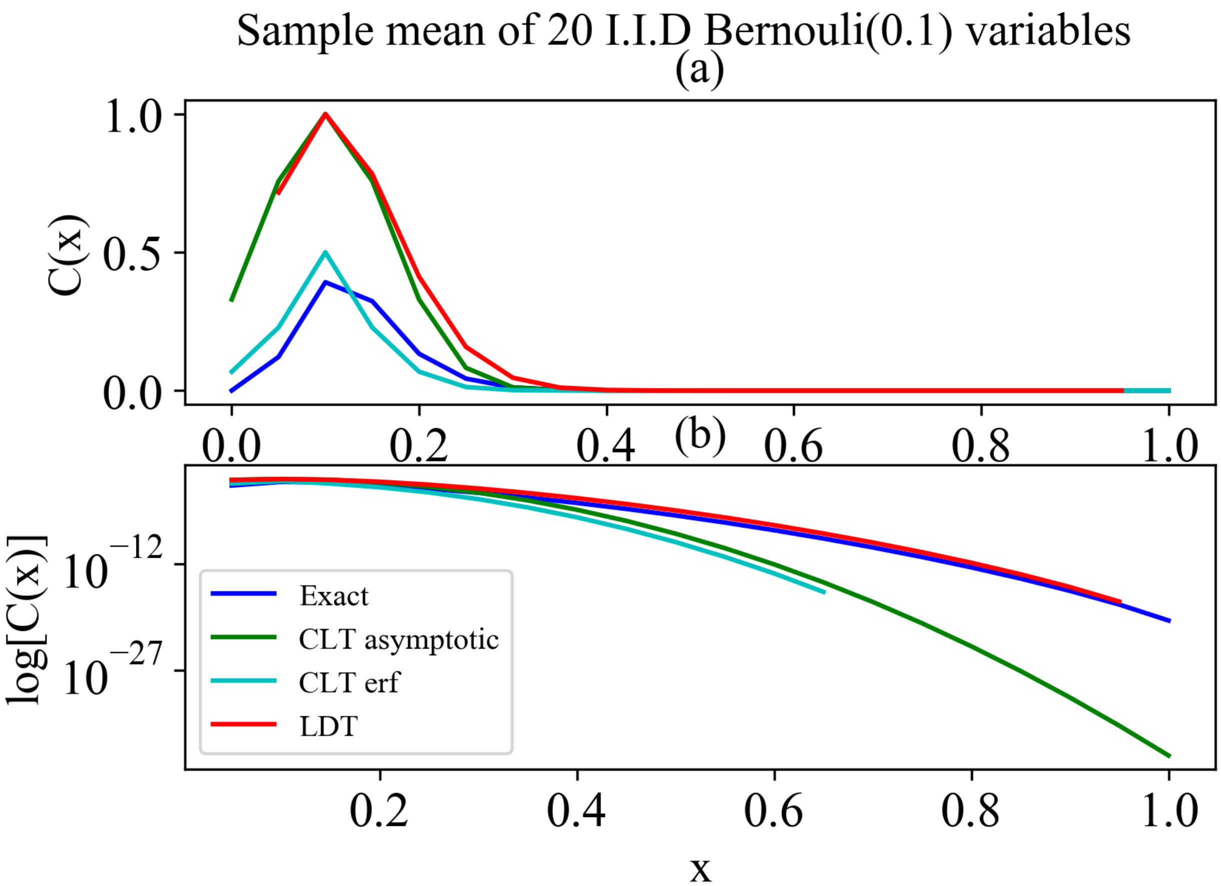

Figure 1: (a): Normal plot of C(x) and the CLT, LDT approximation for $x_1 \sim Ber(0.1)$ (b): Semi-log scale of plot a

Note: This behaviour of LDT approximation unifying with the asymptotic approximation for the *erf* is not by chance, and we shall give a proof that this is always the case at the end of this exercise.

Let us now look at the plot of $C(x)$ in Figure 1 (a), first we see that as expected, around the mean value $\mu = p$, LDT and the asymptotic approximation of CLT are almost unified.
Second notice that *erf* gives a much better result then both asymptotic result, but a closer look would suggests this is only up to a factor of $\frac{1}{2}$.

5)In this section we would like show that for $x_1 \sim N(\mu, \sigma^2)$ the results for sections 1-4 are trivial. We shall skip all technical calculation and focus on the results.

Following the discussion of the CLT in section one, which its conclusion was eq 2, the CLT states that for large enough n:

$$\hat{x} \sim N(\mu, \frac{\sigma^2}{n})$$

and the approximations we get are:

(i) Exact: $\bar{F}(x) \approx \frac{1}{2} - \frac{1}{2} erf\left(\frac{x-\mu}{\sqrt{2}\frac{\sigma}{\sqrt{n}}}\right)$

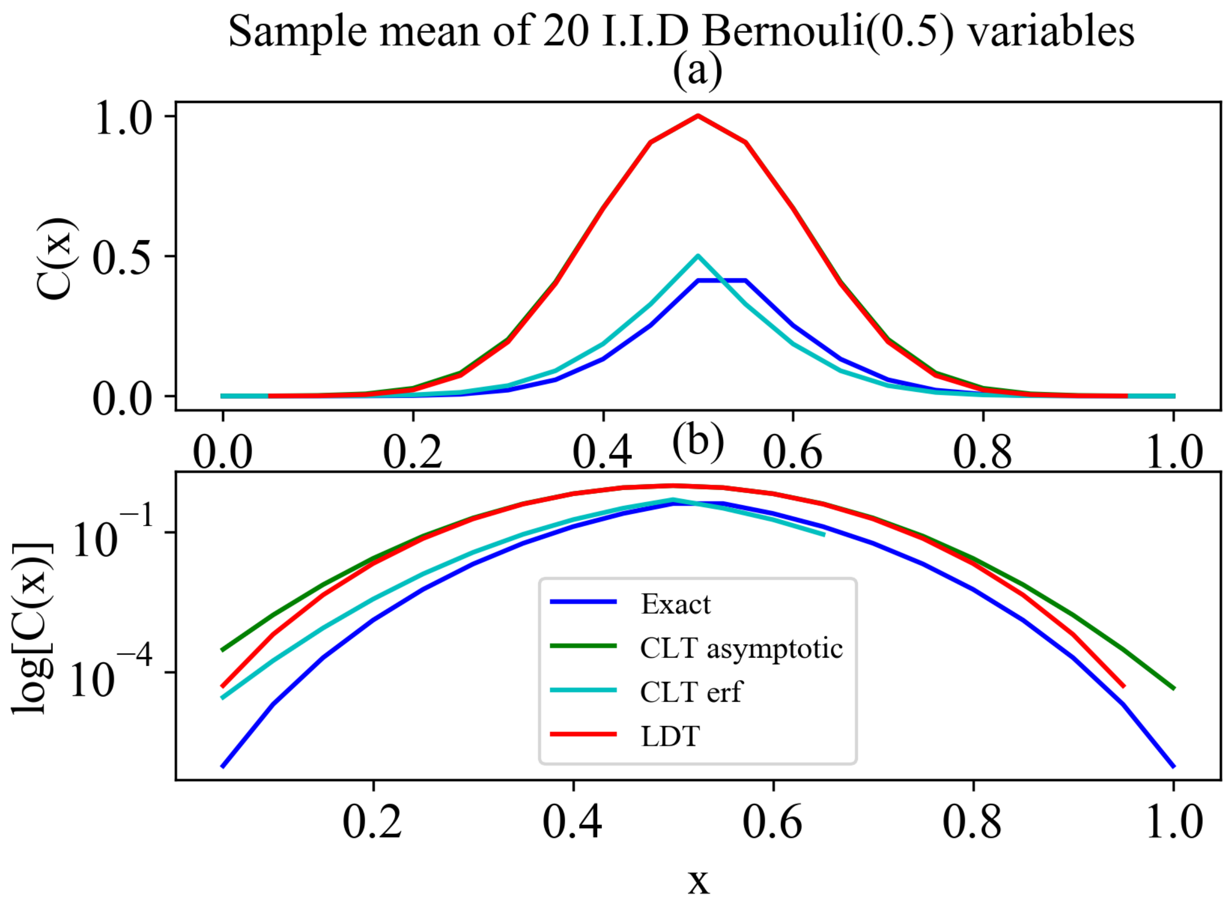

Figure 2: (a): Normal plot of C(x) and the CLT, LDT approximation for $x_1 \sim Ber(0.5)$ (b): Semi-log scale of plot a

and its asymptotic approximation which again is presented with the Folded CDF:

(ii) Elementary function: $\boxed{C(x) \sim exp\left(-\frac{(x-p)^2}{\frac{2\sigma^2}{n}}\right), \text{ where } S_{CLT}(x) = \frac{n}{2\sigma^2}(x-\mu)^2}$

Now following the same steps of section 2 of calculating the legandre transform of $z(\lambda)$ we obtain:

$$S_{LDT}(x) = \frac{n}{2\sigma^2}(x-\mu)^2$$

and so the approximation we get from LTD is:

$$\boxed{C(x) \sim exp\left(-\frac{(x-p)^2}{\frac{2\sigma^2}{n}}\right)}$$

Last in order to calculate the exact expression for $\bar{F}(x)$ we will use the following rule:

Let $X \sim N(\mu_x, \sigma_x^2)$ *and* $Y \sim N(\mu_y, \sigma_y^2)$ *then* $X + Y \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$

There are a few proofs for this proposition, the cleanest one I founnd is with Fourier transform, but they are all completely technical and we will not go over them.

By the above claim and eq 1 it is straight forward that the exact expression is:

$$\bar{F}(x) = \frac{1}{2} - \frac{1}{2} erf\left(\frac{x-\mu}{\sqrt{2}\frac{\sigma}{\sqrt{n}}}\right)$$

To summarise we compare the four results above, showing that in the case where x_j are normally distributed, the CLT is exact and the asymptotic approximation for *erf* unified exactly with the LDT.

6)So far we have treated two cases, for the first we chose a simple distribution - Bernouli, for the seconed we chose a trivial one - Normal, and now we would like to give one last example, for exponential distribution.

Assume $x_j \sim Exp(\alpha)$, by following the same steps as before the approximations we obtain from CLT are:

(i) Exact: $\bar{F}(x) \approx \frac{1}{2} - \frac{1}{2} erf\left(\frac{x-\frac{1}{\alpha}}{\sqrt{2}\frac{1}{\alpha\sqrt{n}}}\right)$

(ii) Elementary function: $\boxed{C(x) \sim exp\left(-\frac{(x-\frac{1}{\alpha})^2}{2\frac{1}{\alpha^2 n}}\right) \text{, where } S_{CLT}(x) = \frac{n}{2\frac{1}{\alpha^2}}(x-\frac{1}{\alpha})^2}$

From LDT we obtain:

$$\boxed{C(x) \sim exp\left(-n(\alpha x - 1 - log(\alpha x))\right) \text{ and so } S_{LDT}(x) = n(\alpha x - 1 - log(\alpha x))}$$

The exact expression for $\bar{F}(x)$ requires some calculation, but the result is that $\hat{x} \sim Gamma(n, \alpha n)$ and thus:

$$\bar{F}(x) = 1 - \frac{\gamma(n, n\alpha x)}{(n-1)!} = 1 - \frac{1}{(n-1)!}\int_0^{n\alpha x} t^{n-1}e^{-t}dt$$

Note: In the literature $\gamma(s, x)$ is called *lower incomplete gamma function.*

The fact that we were able to get an approximation from LDT is not obvious, notice that Cremar's theorem requires:

$$Z(\lambda) := \langle e^{\lambda x_1}\rangle < \infty \ \forall\lambda \in \mathbb{R}$$

yet for the case $x_1 \sim Exp(\alpha)$ we have:

$$Z(\lambda) = \begin{cases} \frac{\alpha}{\alpha-\lambda} & \lambda < \alpha \\ \infty & \lambda \geq \alpha \end{cases}$$

Nevertheless, we are interested only in $\bar{\lambda}(x)$ such that:

$$I(x) := \sup_{\lambda\in\mathbb{R}}[x\lambda - log(Z(\lambda))] = x\bar{\lambda}(x) - log(Z(\bar{\lambda}(x)))$$

carrying out the calculation we find:

$$\bar{\lambda}(x) = \alpha - \frac{1}{x} < \alpha \ \forall x > 0$$

and indeed we shall see the LDT approximation works in this case.

As before, let us begin by analyze the behaviour at the edges:

(i) $\lim_{x\to\infty} \frac{S_{CLT}(x)}{S_{LDT}(x)} = \lim_{x\to\infty} \frac{\frac{\alpha^2}{2}(x-\frac{1}{\alpha})^2}{(\alpha x - 1) - log(\alpha x)} = \infty$

(ii) $\lim_{x\to 0} \frac{S_{CLT}(x)}{S_{LDT}(x)} = \lim_{x\to 0} \frac{\frac{\alpha^2}{2}(x-\frac{1}{\alpha})^2}{(\alpha x - 1) - log(\alpha x)} = 0^+$

thus for large x CLT is above LDT and for small x the opposite. This is interesting since it was not the case for the Bernouli distribution, where for both small and large x CLT was above LDT.

For the graphical analysis we look at the semi-log scale plot of the three approximation as in Figure 3 (b), and as before we can see that the LDT gives a much better approximation at the edges.

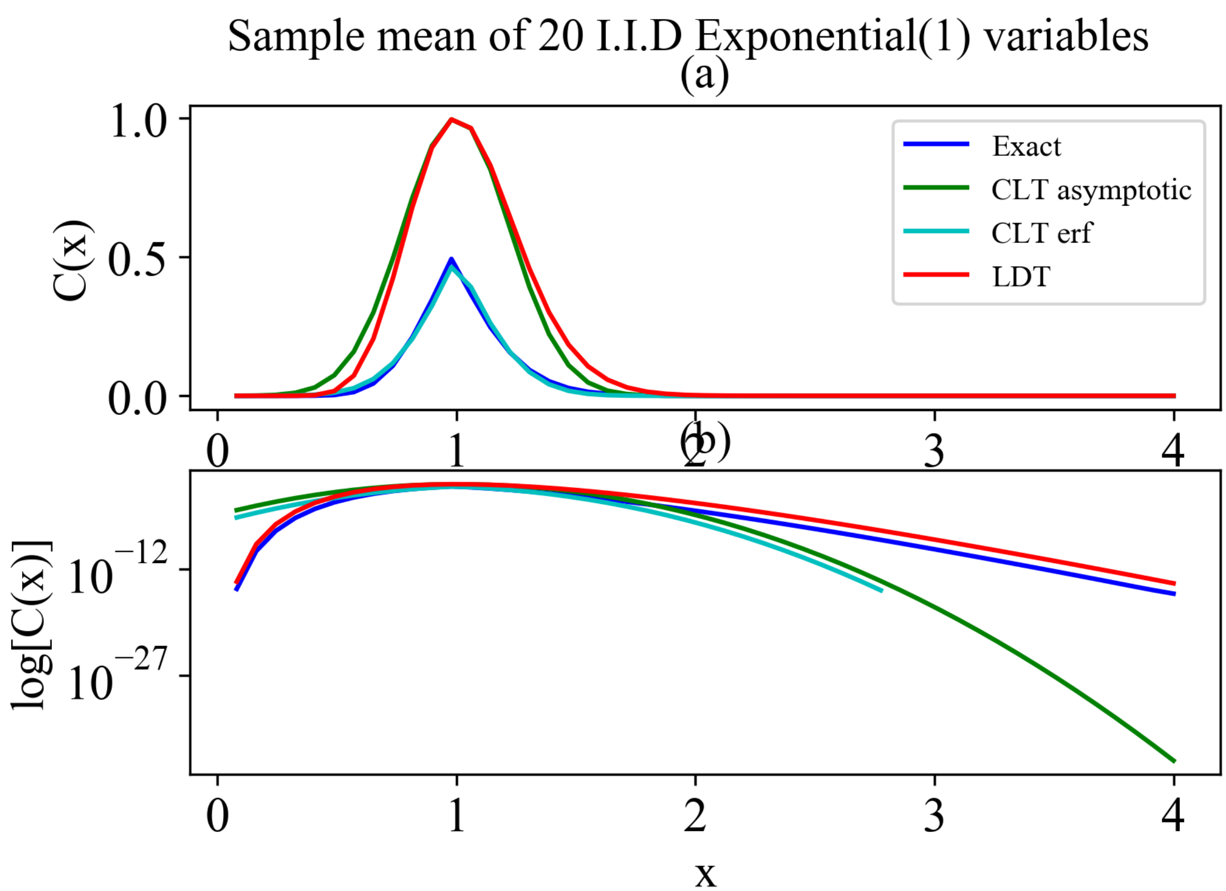

Figure 3: (a): Plot of C(x) and the CLT, LDT approximation for $x_1 \sim Exp(1)$, (b): Semi-log scale of plot a

We now continue with examine the behaviour around the expected value $\mu = \frac{1}{\alpha}$. Expending $S_{LDT}(x)$ in Taylor series up to the second order we find:

$$S_{LDT}(x) \approx \frac{n\alpha^2}{2}(x-\frac{1}{\alpha})^2 = S_{CLT}(x)$$

as expected from previous sections. And indeed looking at the plot in Figure 3 (a), we see that LDT and the asymptotic approximation for CLT behave similarly around $x = \frac{1}{\alpha}$.

7) In this section we shall show that for a general i.i.d random variables x_j, the asymptotic approximation for CLT and the LDT are equal up to the second order in the neighborhood of $\mu = \langle x_1 \rangle$.

Keeping the notation as before the asymptotic approximation for CLT always gives:

$$S_{CLT}(x) = \frac{n}{2\sigma^2}(x-\mu)^2$$

We now want to expand $I(x)$ as defined in Cremar's theorem in Taylor series up to the second oreder. We remind that by definition:

$$I(x) := \sup_{\lambda \in \mathbb{R}}[x\lambda - log(Z(\lambda))]$$

which is exactly the Legandre transform of $log(Z(\lambda))$, thus we know λ is given by solving the equation:

$$x = \frac{1}{Z(\lambda)}\frac{dZ}{d\lambda} \tag{4}$$

In order to solve this equation we shall calculate the Taylor expansion of $Z(\lambda)$:

$$Z(\lambda) := \langle e^{\lambda x_1} \rangle = \langle \sum_{k=0}^{\infty} \frac{(\lambda x_1)^k}{k!} \rangle = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \langle x^k \rangle$$

and a more illuminating way to right it, is:

$$Z(\lambda) = 1 + \mu\lambda + (\mu^2 + \sigma^2)\frac{\lambda^2}{2} + ...$$

and the information we will need from this calculation is:

$$Z(0) = 1,\ Z'(0) = \mu,\ Z''(0) = \mu^2 + \sigma^2 \tag{5}$$

Let us note $\bar{\lambda}(x)$ such that:

$$I(x) = x\bar{\lambda}(x) - log(Z(\bar{\lambda}(x)))$$

then in order to find the Taylor expansion of $I(x)$ around $x = \mu$ we have to solve the following 3 equation:

(i) $I(\mu) = \mu\bar{\lambda}(\mu) - log(Z(\bar{\lambda}(\mu)))$

(ii) $I'(\mu) = \bar{\lambda}(\mu)$

(iii) $I''(\mu) = \frac{d\bar{\lambda}}{dx}(\mu)$

where we have used eq 4 to calculate $I'(x)$.
Again looking at eq 4, and inserting the values for $\lambda = 0$ from eq 5 we find that:

$$x(\lambda = 0) = \frac{1}{Z(0)}Z'(0) = \mu$$

so we conclude:

$$\bar{\lambda}(\mu) = 0$$

which gives us the solutions for (i) and (ii):

(i) $I(\mu) = 0$

(ii) $I'(\mu) = 0$

For (iii) we will have to work little harder, once again we look at eq 4 which we know is true for $\bar{\lambda}$ and differentiate both side with respect to x:

$$1 = -\frac{1}{(Z(\bar{\lambda}))^2}\left(\frac{dZ}{d\bar{\lambda}}\right)^2 + \frac{1}{Z(\bar{\lambda})}\frac{d^2 z}{d\bar{\lambda}^2}\frac{d\bar{\lambda}}{dx}$$

inserting $x = \mu$ and the values from eq 5 we get the third solution:

(iii) $I''(\mu) = \frac{d\bar{\lambda}}{dx}(\mu) = \frac{1}{\sigma^2}$

At last we can right the Taylor series up to the second order for $I(x)$:

$$I(x) \approx \frac{1}{2\sigma^2}(x - \mu)^2$$

and so:

$$S_{LDT}(x) = nI(x) \approx \frac{n}{2\sigma^2}(x - \mu)^2 = S_{CLT}(x)$$

as we claimed.

Summery: To summarise, we haven't prove so rigorously but the above discussion suggest that for $|x| >> \mu$ the LDT approximation is better then both exact CLT approximation (erf) and its asymptotic approximation (Gaussain), while for $x \approx \mu$ it is still better then the asymptotic approximation for CLT but falls in compare to the *erf*.