ASYMPTOTIC APPROXIMATIONS IN STATISTICS

JOY SONG[1]

1. INTRODUCTION

Asymptotic expansions could provide approximations to complicated integrals such as density functions or distribution functions arising in statistical calculation. This project investigates the Edgeworth expansion and saddlepoint approximation, which are the most commonly used asymptotic techniques to approximate the density function of certain statistics. Generally speaking, Edgeworth expansion is based on the Taylor expansion of the cumulant-generating function and saddlepoint approximation utilizes saddlepoints to estimate a certain type of complex integrals. The existing literature only gives an outline about how to obtain these asymptotic expansions, so following the outline and combining the descriptions in different previous work, this project forms a detailed proof for the standard Edgeworth expansion and saddlepoint approximation. In fact, we could use what we have learned in MATH 521 to obtain the one-term saddlepoint approximation, additional terms are given by a more accurate estimation of functions in the same procedure. In the end, we also verified the two approximations via numerical experiments.

2. EDGEWORTH EXPANSION

For n independent and identically distributed random variables $X_1, X_2, \dots, X_n$, let $M(X;t) = E[e^{tX}]$ and $K(X;t) = \log M(X;t)$ be the moment-generating function and cumulant-generating function of X_i. Let the standardized sample means be

$$S_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma}$$

where μ, σ are the mean and variance of X_i. In this section, following the outline in *Asymptotic Techniques for Use in Statistics* (O.E. Barndorff-Nielsen and D.R. Cox, 1989) [1], we are going to prove the following Edgeworth approximation for the density function of S_n:

$$f(S_n; x) = \Phi(x)\Big(1 + \frac{\rho_3 H_3(x)}{6\sqrt{n}} + \frac{\rho_4 H_4(x)}{24n} + \frac{\rho_3^2 H_6(x)}{72n}\Big) + O(n^{-3/2})$$

where $\Phi(x) = (2\pi)^{-1/2} e^{-x^2/2}$ is the normal density function, $\rho_r = \kappa_r / \sigma^r$ and κ_r is the r^{th} cumulant of X_i defined as the coefficient of $t^r/r!$ in the Taylor expansion of $K(X;t)$, $H_r(x)$ is the Hermite polynomial of degree r defined via

$$(\frac{d}{dx})^r \Phi(x) = (-1)^r H_r(x) \Phi(x)$$

[1] BRIGHAM YOUNG UNIVERSITY
E-mail address: `jsong@mathematics.byu.edu`.
Date: August 21, 2020.

2.1. **Expansion of Cumulant-generating Function.** The Edgeworth expansion is basically derived from the expansion of the cumulant-generating function $K(X;t)$. In this section we will find this expansion as follows

$$K(X;t) = \mu t + \frac{1}{2}\sigma^2 t^2 + \frac{1}{6}\rho_3\sigma^3 t^3 + \frac{1}{24}\rho_4\sigma^4 t^4 + \cdots$$

We first write

$$\mu_r' = E(X^r), \quad \mu_r = E(X-\mu)^r, \quad r = 1,2,\ldots$$

then we have the expansions

$$M(X;t) = 1 + \sum \frac{\mu_r'}{r!} t^r, \quad M(X-\mu;t) = e^{-t\mu} M(X;t) = 1 + \sum \frac{\mu_r}{r!} t^r$$

Therefore by definition we have

$$\begin{aligned} K(X;t) &= t\mu + K(X-\mu;t) = t\mu + \log M(X-\mu;t) \\ &= t\mu + \log(1 + \sum \frac{\mu_r}{r!} t^r) \end{aligned}$$

Let the Taylor expansion of $K(X;t)$ be

$$K(X;t) = \sum \frac{\kappa_r}{r!} t^r, \quad \kappa_r = K^{(r)}(0)$$

From the equation

$$\sum \frac{\kappa_r}{r!} t^r = t\mu + \log\left(1 + \sum \frac{\mu_r}{r!} t^r\right)$$

we obtain

$$\kappa_1 = \mu, \quad \kappa_2 = \mu_2 = E(X-\mu)^2 = \sigma^2.$$

Let $\rho_r = \kappa_r/\sigma^r, r = 3,4,\ldots$ then we have

$$K(X;t) = \mu t + \frac{1}{2}\sigma^2 t^2 + \frac{1}{6}\rho_3\sigma^3 t^3 + \frac{1}{24}\rho_4\sigma^4 t^4 + \cdots$$

2.2. **Edgeworth Expansion.** Now we could derive the Edgeworth expansion of the density function $f(S_n;x)$ from the expansion of $K(X;t)$.

From the definition $K(X;t) = \log M(X;t) = \log E(e^t X)$, we know that for

$$S_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}}$$

the cumulant-generating function is

$$K(S_n;t) = -\frac{\sqrt{n}\mu t}{\sigma} + nK(X; \frac{t}{\sigma\sqrt{n}})$$

Combined with the expansion we obtained for $K(X;t)$, we have

$$K(S_n;t) = \frac{1}{2}t^2 + \frac{\rho_3}{6\sqrt{n}} t^3 + \frac{\rho_4}{24n} t^4 + O(n^{-3/2}).$$

And the corresponding moment-generating function is

$$M(S_n;t) = e^{K(S_n;t)} = \exp(\frac{1}{2}t^2)\exp\left\{\frac{\rho_3}{6\sqrt{n}}t^3 + \frac{\rho_4}{24n}t^4 + \cdots\right\}$$
$$= \exp(\frac{1}{2}t^2)\left\{1 + \frac{\rho_3}{6\sqrt{n}}t^3 + \frac{\rho_4}{24n}t^4 + \cdots + \frac{1}{2}\left[1 + \frac{\rho_3}{6\sqrt{n}}t^3 + \frac{\rho_4}{24n}t^4 + \cdots\right]^2 + \cdots\right\}$$
$$= \exp(\frac{1}{2}t^2)\left[1 + \frac{\rho_3}{6\sqrt{n}}t^3 + \frac{\rho_4}{24n}t^4 + \frac{\rho_3^2}{72n}t^6 + O(n^{-3/2})\right].$$

Now making use of an important identity

$$\int_{-\infty}^{+\infty} e^{tx}\Phi(x)H_r(x)dx = t^r e^{\frac{1}{2}t^2} \quad \Rightarrow \quad \Phi(x)H_r(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} t^r e^{\frac{1}{2}t^2 - itx}$$

we invert $M(S_n;t)$ and obtain

$$f(S_n;x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} M(S_n;it)e^{-itx}dt$$
$$= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\left[1 + \frac{\rho_3}{6\sqrt{n}}t^3 + \frac{\rho_4}{24n}t^4 + \frac{\rho_3^2}{72n}t^6 + O(n^{-3/2})\right]e^{\frac{1}{2}t^2 - itx}dt$$
$$= \Phi(x)\left(1 + \frac{\rho_3 H_3(x)}{6\sqrt{n}} + \frac{\rho_4 H_4(x)}{24n} + \frac{\rho_3^2 H_6(x)}{72n}\right) + O(n^{-3/2})$$

which is the desired Edgeworth expansion.

2.3. **Example.** As an example, we apply the Edgeworth expansion to

$$S_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma}, \qquad X_i \text{ i.i.d. } \sim \mathrm{Exp}(\lambda).$$

For exponential distribution Exp(λ), we know that $\mu = \lambda, \sigma^2 = \lambda^2$ and the density function is

$$f(x) = \frac{1}{\lambda}e^{-\frac{x}{\lambda}}.$$

It is true for any distribution that

$$\kappa_3 = \mu_3, \quad \kappa_4 = \mu_4 - 3\mu_2^2$$

and [5]

$$H_3(x) = x^3 - 3x, \quad H_4(x) = x^4 - 6x^2 + 3, \quad H_6(x) = x^6 - 15x^4 + 45x^2 - 15.$$

This means in order to obtain the Edgeworth expansion, we only need to compute

$$\mu_2 = \sigma^2 = \lambda^2, \quad \mu_3 = 2\lambda^3, \quad \mu_4 = 9\lambda^4.$$

Finally we could use the Edgeworth expansion

$$f(S_n;x) \approx \Phi(x)\left(1 + \frac{\rho_3 H_3(x)}{6\sqrt{n}} + \frac{\rho_4 H_4(x)}{24n} + \frac{\rho_3^2 H_6(x)}{72n}\right)$$

and the results are shown in Figure 1.

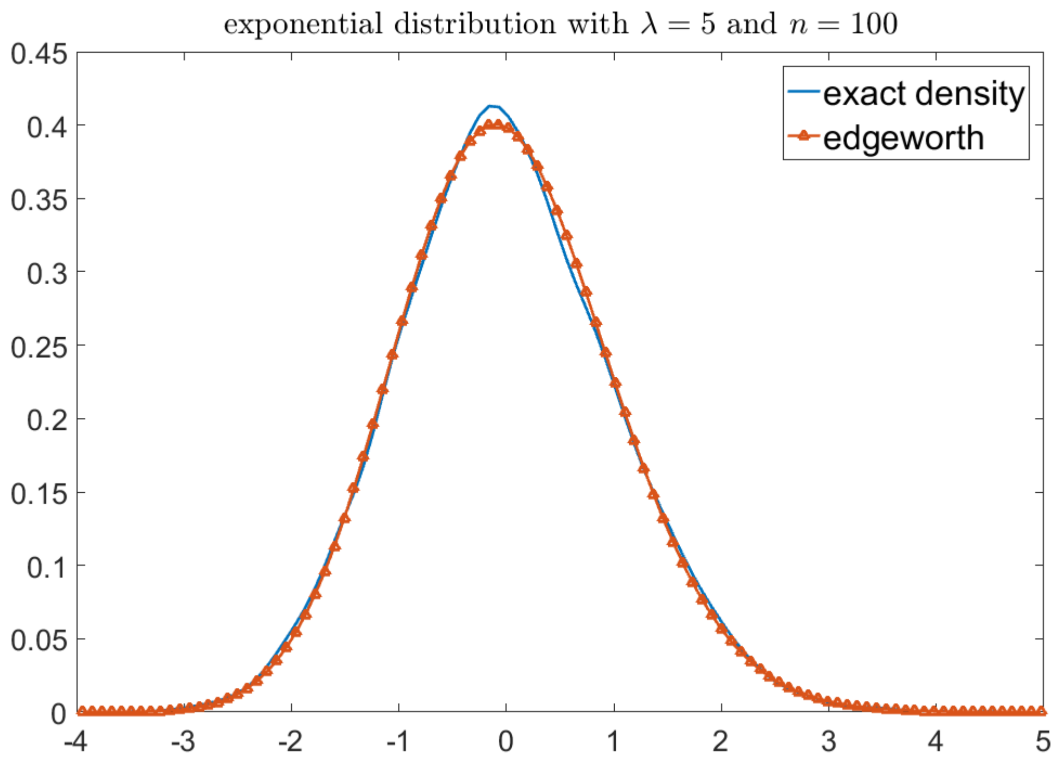

Figure 1. For n independent and identically distributed random variables from the exponential distribution family $\mathrm{Exp}(\lambda)$, use Edgeworth expansion to approximate the density function of the sandardized sample means S_n

3. Saddlepoint Approximation

For n independent and identically distributed random variables $X_1, X_2, \ldots, X_n$, we are going to prove in this section the following saddlepoint approximation for the density function of the sample means $\bar{X}$:

$$f_n(\bar{x}) = \sqrt{\frac{n}{2\pi K''(t_0)}} e^{n[K(t_0) - t_0\bar{x}]} \Big(1 + \frac{3\lambda_4(t_0) - 5\lambda_3^2(t_0)}{24n} + O(n^2)\Big),$$

where t_0 is the saddlepoint of $h(t) = K(t_0) - t_0\bar{x}$ which satisfies $K'(t_0) = \bar{x}$, $K(t)$ is the cumulant generating function of X_i, $\lambda_r(t) = K^{(r)}(t)/[K''(t)]^{r/2}$ is the r^{th} cumulant function.

3.1. **Preliminary.**

Before applying the saddlepoint techniques, we introduce some preliminary notions as well as analyse their properties, thereby formulating the density function $f_n(x)$ and guaranteeing the validity of saddlepoint approximation.

3.1.1. *Properties of Generating Functions.*

We are going to investigate the convergence and the analyticity of moment-generating function and cumulant-generating function, which equips us with the foundation of further analysis.

The moment-generating function of a random variable X is defined as

$$M(X;t) = \int_{-\infty}^{+\infty} e^{tx} dF(x) = \int_{-\infty}^{+\infty} e^{tx} f(x) dx, \quad x \in \mathbb{R}, t \in \mathbb{C}$$

where $F(x)$ is the distribution function of X and $f(x)$ is the density function of X. Using Theorem 2 in Chapter VI of the book *The Laplace Transform*[6] (Widder, David Vernon, 1946), we know that if $M(X;t)$ converges for two points $t_1 = \sigma_1 + i\tau_1$ and $t_2 = \sigma_2 + i\tau_2$ with $\sigma_1 < \sigma_2$, then it converges in the vertical strip $\sigma_1 < \mathscr{R}(t) < \sigma_2$. Since $M(X;t)$ converges at $t = 0$, from now on we assume that $M(X;t)$ converges in some vertical strip $-\sigma_1 < \mathscr{R}(t) < \sigma_2$ with $\sigma_1 + \sigma_2 > 0, 0 \leqslant \sigma_1 \leqslant +\infty, 0 \leqslant \sigma_2 \leqslant +\infty$.

For each given $x \in \mathbb{R}$, we know that e^{tx} is analytic in the convergence region $D = \{t = \sigma + i\tau \in \mathbb{C} : -\sigma_1 < \sigma < \sigma_2\}$, and since

$$\int_{-\infty}^{+\infty} |e^{tx}| dF(x) = \int_{-\infty}^{+\infty} |e^{\sigma x}| dF(x) \leqslant \int_{-\infty}^{+\infty} e^{\sigma_2 x} dF(x)$$

we know that $\int_{-\infty}^{+\infty} |e^{tx}| dF(x)$ is uniformly bounded on compact subsets of D. This means for any closed C^1 curve Γ in D, we have

$$\oint_\Gamma \int_{-\infty}^{+\infty} e^{tx} dF(x) dt = \int_{-\infty}^{+\infty} \oint_\Gamma e^{tx} dt dF(x) = 0$$

which by Morera's theorem gives us $M(X;t)$ is an analytic function with respect to $t \in D$.

Now we move on to the cumulant-generating function of X, which is defined as

$$K(X;t) = \log M(X;t).$$

It is always true that $\mathscr{R}(M(X;t)) > 0$, therefore $K(X;t)$ has the same convergence region D as $M(X;t)$ and is also an analytic function with respect to $t \in D$.

3.1.2. *Density Function.* Now we are able to use moment-generating function and cumulant-generating function to formulate the desired density function

For n independent and identically distributed random variables $X_1, X_2, \ldots, X_n$, let $f_n(\bar{x})$ be the density function of the sample means $\bar{X}$, then we have

$$M(\bar{X}; int) = \int_{-\infty}^{+\infty} e^{int\bar{x}} f_n(\bar{x}) d\bar{x},$$

which is assumed to converge in some neighborhood $D = \{t \in \mathbb{C} : -\sigma_1 < \mathscr{R}(int) < \sigma_2\}$. Applying the inversion formula [4] we get

$$\begin{aligned} f_n(\bar{x}) &= \frac{n}{2\pi} \int_{-\infty}^{+\infty} M(\bar{X}; int) e^{-int\bar{x}} dt \\ &= \frac{n}{2\pi} \int_{-\infty}^{+\infty} M^n(X; it) e^{-int\bar{x}} dt \\ &= \frac{n}{2\pi} \int_{-\infty}^{+\infty} e^{n[K(X;it) - it\bar{x}]} dt \end{aligned}$$

which under the change of variable $t = it$ becomes

$$f_n(\bar{x}) = \frac{n}{2\pi i} \int_{\tau - i\infty}^{\tau + i\infty} e^{n[K(X;t) - t\bar{x}]} dt$$

where $\tau \in (-\sigma_1/n, \sigma_2/n) = (-c_1, c_2)$.

3.2. **Steepest Descents.** Now we could approximate the density function using the steepest descents method we learned in MATH 521. From now on we write $M(X;t), K(X;t)$ simply as $M(t), K(t)$, let

$$h(t) = K(t) - t\bar{x}, \quad t \in D = \{t \in \mathbb{C} : \mathscr{R}(t) \in (-c_1, c_2)\},$$

let C be the contour that extends from $\tau - i\infty$ to $\tau + i\infty$. Then $h(t)$ is analytic in the region D which contains the contour C. As $n \to \infty$, we aim to find an asymptotic approximation to the integral

$$f_n(\bar{x}) = \frac{n}{2\pi i}\int_{\tau - i\infty}^{\tau + i\infty} e^{n[K(t) - t\bar{x}]}\,dt = \frac{n}{2\pi i}\int_C e^{nh(t)}\,dt$$

The saddlepoint of $h(t)$ is given by solving

$$h'(t) = K'(t) - \bar{x} = 0,$$

hence we need to investigate the existence and properties of the roots of this equation. For $M(t) = \displaystyle\int_{-\infty}^{+\infty} e^{tx}\,dF(x)$, let the largest region of convergence be $\mathscr{R}(t) \in (-c_1, c_2)$, and suppose

$$F(x)\begin{cases} = 0 & x \leqslant a \\ \in (0,1) & x \in (a,b) \\ = 1 & x \geqslant b \end{cases}.$$

It has been proved in *Saddlepoint Approximations in Statistics* (Daniels, 1954) [2] that if

$$\lim_{t \to c_2} K'(t) = b, \quad \lim_{t \to -c_1} K'(t) = a,$$

then for each $\bar{x} \in (a,b)$, the equation $K'(t) = \bar{x}$ has a simple root $t_0 \in (-c_1, c_2)$ with $K''(t_0) > 0$.

Now we have obtained the saddlepoint $t_0 \in (-c_1, c_2)$ of order one. The Taylor expansion of $h(t)$ around t_0 is given by

$$\begin{aligned} h(t) &= h(t_0) + \frac{1}{2}h''(t_0)(t - t_0)^2 + O(t - t_0)^3 \\ &= K(t_0) - t_0\bar{x} + \frac{1}{2}K''(t_0)(t - t_0)^2 + O(t - t_0)^3 \end{aligned}$$

In the vicinity of the saddlepoint t_0, let

$$t = t_0 + re^{i\theta}, \quad r > 0, \quad \theta \in [0, 2\pi)$$

Therefore when t is close to t_0 we have

$$h(t) \approx K(t_0) - t_0\bar{x} + \frac{1}{2}K''(t_0)(t - t_0)^2 \approx K(t_0) - t_0\bar{x} + \frac{1}{2}K''(t_0)r^2e^{2i\theta}$$

Combined with $K''(t_0) > 0$, in the vicinity of t_0 we have that

$$\mathscr{I}h(t) = \mathscr{I}(K(t_0) - t_0\bar{x}) + \frac{1}{2}K''(t_0)r^2\sin 2\theta$$

$$\mathscr{R}h(t) = \mathscr{R}(K(t_0) - t_0\bar{x}) + \frac{1}{2}K''(t_0)r^2\cos 2\theta$$

Then the steepest descents path is given by

$$\mathscr{I}h(t) \quad \text{constant} \quad \Longleftrightarrow \quad \sin 2\theta = 0 \quad \Longleftrightarrow \quad \begin{cases} \theta = 0, \quad \theta = \pi \\ \theta = \pi/2, \quad \theta = 3\pi/2 \end{cases}$$

When $\theta = 0, \theta = \pi$,

$$\mathscr{R}h(t) = \mathscr{R}(K(t_0) - t_0\bar{x}) + \frac{1}{2}K''(t_0)r^2 > \mathscr{R}(K(t_0) - t_0\bar{x}) = \mathscr{R}h(t_0).$$

When $\theta = \pi/2, \theta = 3\pi/2$,

$$\mathscr{R}h(t) = \mathscr{R}(K(t_0) - t_0\bar{x}) - \frac{1}{2}K''(t_0)r^2 < \mathscr{R}(K(t_0) - t_0\bar{x}) = \mathscr{R}h(t_0).$$

Therefore the steepest descents path through t_0 is given by

$$t = t_0 + re^{i\theta}, \quad \theta = \pi/2, \theta = 3\pi/2.$$

Since $t_0 \in (-c_1, c_2)$, we are able to deform the original contour C into C_1 that lies along the steepest descents path in the vicinity of the saddlepoint t_0, and the integral becomes

$$f_n(\bar{x}) = \frac{n}{2\pi i}\int_C e^{nh(t)}dt = \frac{n}{2\pi i}e^{nh(t_0)}\int_{C_1} e^{n[h(t)-h(t_0)]}dt$$

Now we introduce $\tau \in \mathbb{R}$ such that

$$-\tau^2 = h(t) - h(t_0) \approx \frac{1}{2}K''(t_0)(t-t_0)^2,$$

which gives us

$$t = t_0 \pm i\tau\sqrt{\frac{2}{K''(t_0)}}.$$

In the vicinity of t_0, we have

$$\theta = \pi/2, \quad \Rightarrow \quad t = t_0 + ir$$
$$\theta = 3\pi/2, \quad \Rightarrow \quad t = t_0 - ir$$

In order to make $\tau > 0$ when $\theta = \pi/2$ and $\tau < 0$ when $\theta = 3\pi/2$, we choose

$$t = t_0 + i\tau\sqrt{\frac{2}{K''(t_0)}} \quad \Rightarrow \quad \tau = -\sqrt{\frac{K''(t_0)}{2}}(t-t_0)i.$$

Under this change of variable, the integral becomes

$$f_n(\bar{x}) = \frac{n}{2\pi i}e^{nh(t_0)}\int_{C_1} e^{n[h(t)-h(t_0)]}dt \approx \frac{n}{2\pi i}e^{nh(t_0)}\int_{C_1'} e^{-n\tau^2}\frac{dt}{d\tau}d\tau$$

where C_1' is the contour of τ under the transformation

$$\tau = -\sqrt{\frac{K''(t_0)}{2}}(t-t_0)i$$

which recenters C_1 from t_0 to the origin and then rotate it by $-i = e^{3\pi i/2}$. Therefore, when t is along the steepest descents path through t_0, the corresponding τ is along the real axis in some neighborhood of the origin. Let $n \to \infty$, the asymptotic approximation is as follows

$$f_n(\bar{x}) \approx \frac{n}{2\pi i}e^{nh(t_0)}\int_{C_1'} e^{-n\tau^2}\frac{dt}{d\tau}d\tau \sim \frac{n}{2\pi i}e^{nh(t_0)}\int_{-\infty}^{+\infty} e^{-n\tau^2}\frac{dt}{d\tau}d\tau$$

Substituting

$$h(t_0) = K(t_0) - t_0\bar{x}, \quad \frac{dt}{d\tau} = i\sqrt{\frac{2}{K''(t_0)}}$$

we finally obtain

$$f_n(\bar{x}) \sim \sqrt{\frac{n}{2\pi K''(t_0)}} e^{n[K(t_0)-t_0\bar{x}]}, \quad \text{as} \quad n \to \infty$$

3.3. **Additional Terms.** The steepest descents method we learned in MATH 521 only gives us a one-term approximation of the density function $f_n(\bar{x})$. To obtain additional asymptotic terms, we need to have a more accurate Taylor expansion of $h(t)$, as outlined in *Saddlepoint Approximations in Statistics* (Daniels, 1954) [2].

Following the same procedure as 3.2 above, when we introduce $\tau \in \mathbb{R}$ such that

$$h(t) - h(t_0) = -\tau^2,$$

we now use more terms to approximate $h(t) - h(t_0)$, that is

$$-\tau^2 = h(t) - h(t_0) = \frac{1}{2}K''(t_0)(t-t_0)^2 + \frac{1}{6}K^{(3)}(t_0)(t-t_0)^3 + \frac{1}{24}K^{(4)}(t_0)(t-t_0)^4 + \cdots$$

Let $z = (t-t_0)\sqrt{K''(t_0)}$ then we have

$$-\tau^2 = h(t) - h(t_0) = F(z) = \frac{1}{2}z^2 + \frac{1}{6}\lambda_3(t_0)z^3 + \frac{1}{24}\lambda_4(t_0)z^4 + \cdots$$

where

$$\lambda_r(t) = K^{(r)}(t)/[K''(t)]^{r/2}$$

is the r^{th} cumulant function. To obtain the expression of z in terms of τ, let $z = a_1\tau + a_2\tau^2 + a_3\tau^3 + \cdots$ and substitute it into the equation:

$$\begin{aligned}-\tau^2 = &\frac{1}{2}(a_1\tau + a_2\tau^2 + a_3\tau^3 + \cdots)^2 + \frac{1}{6}\lambda_3(t_0)(a_1\tau + a_2\tau^2 + a_3\tau^3 + \cdots)^3 \\ &+ \frac{1}{24}\lambda_4(t_0)(a_1\tau + a_2\tau^2 + a_3\tau^3 + \cdots)^4 + \cdots\end{aligned}$$

Collecting the coefficients of different powers of τ gives us

$$\begin{aligned}\frac{1}{2}a_1^2 = -1, &\quad \Rightarrow \quad a_1 = \sqrt{2}i \\ a_2 + \frac{1}{6}\lambda_3(t_0)a_1^2 = 0 &\quad \Rightarrow \quad a_2 = \frac{1}{3}\lambda_3(t_0) \\ \frac{1}{6}\lambda_4(t_0) - \frac{5}{18}\lambda_3^2(t_0) + \sqrt{2}i a_3 = 0 &\quad \Rightarrow \quad a_3 = \frac{\sqrt{2}}{12}\lambda_4(t_0)i - \frac{5\sqrt{2}}{36}\lambda_3^2(t_0)i\end{aligned}$$

Therefore the inversion of the series

$$-\tau^2 = F(z) = \frac{1}{2}z^2 + \frac{1}{6}\lambda_3(t_0)z^3 + \frac{1}{24}\lambda_4(t_0)z^4 + \cdots$$

yields an expansion

$$z = G(\tau) = \sqrt{2}i\tau + \frac{1}{3}\lambda_3(t_0)\tau^2 + \Big(\frac{\sqrt{2}}{12}\lambda_4(t_0) - \frac{5\sqrt{2}}{36}\lambda_3^2(t_0)\Big)i\tau^3 + \cdots$$

The integral now becomes

$$\begin{aligned}
f_n(\bar{x}) &= \frac{n}{2\pi i} e^{nh(t_0)} \int_{C_1} e^{n[h(t)-h(t_0)]} dt \\
&= \frac{n}{2\pi i} e^{nh(t_0)} \int_{C_2} e^{nF(z)} \frac{dt}{dz} dz \\
&= \frac{n}{2\pi i} e^{nh(t_0)} \int_{C_3} e^{-n\tau^2} \frac{dt}{dz} \frac{dz}{d\tau} d\tau \\
&= \frac{n}{2\pi i} \frac{e^{nh(t_0)}}{\sqrt{K''(t_0)}} \int_{C_3} e^{-n\tau^2} \frac{dz}{d\tau} d\tau
\end{aligned}$$

Here C_1 is the contour of t that lies along the steepest descents path in the vicinity of the saddle-point t_0. C_2 is the contour of z under the transformation $z = (t - t_0)\sqrt{K''(t_0)}$, which recenters C_1 from t_0 to the origin. C_3 is the contour of τ under the transformation $-\tau^2 = F(z)$, which rotates C_2 by $-i = e^{3\pi i/2}$. Therefore, when t is along the steepest descents path through t_0, the corresponding τ is along the real axis in some nieghborhood of the origin. Let $n \to \infty$, the asymptotic approximation is as follows

$$f_n(\bar{x}) = \frac{n}{2\pi i} \frac{e^{nh(t_0)}}{\sqrt{K''(t_0)}} \int_{C_3} e^{-n\tau^2} \frac{dz}{d\tau} d\tau \sim \frac{n}{2\pi i} \frac{e^{nh(t_0)}}{\sqrt{K''(t_0)}} \int_{-\infty}^{+\infty} e^{-n\tau^2} \frac{dz}{d\tau} d\tau$$

where

$$\frac{dz}{d\tau} = G'(\tau) = \sqrt{2} i + \frac{2}{3}\lambda_3(t_0)\tau + \left(\frac{\sqrt{2}}{4}\lambda_4(t_0) - \frac{5\sqrt{2}}{12}\lambda_3^2(t_0)\right) i\tau^2 + \cdots$$

Therefore

$$\begin{aligned}
\int_{-\infty}^{+\infty} e^{-n\tau^2} \frac{dz}{d\tau} d\tau &\approx \int_{-\infty}^{+\infty} e^{-n\tau^2} \left[\sqrt{2} i + \left(\frac{\sqrt{2}}{4}\lambda_4(t_0) - \frac{5\sqrt{2}}{12}\lambda_3^2(t_0)\right) i\tau^2\right] d\tau \\
&= \sqrt{\frac{2\pi}{n}}\, i + \frac{1}{2n}\sqrt{\frac{2\pi}{n}}\, i \left(\frac{1}{4}\lambda_4(t_0) - \frac{5}{12}\lambda_3^2(t_0)\right)
\end{aligned}$$

which gives us

$$\begin{aligned}
f_n(\bar{x}) &\sim \frac{n}{2\pi i} \frac{e^{nh(t_0)}}{\sqrt{K''(t_0)}} \int_{-\infty}^{+\infty} e^{-n\tau^2} \frac{dz}{d\tau} d\tau \\
&= \sqrt{\frac{n}{2\pi K''(t_0)}}\, e^{n[K(t_0) - t_0\bar{x}]} \left(1 + \frac{3\lambda_4(t_0) - 5\lambda_3^2(t_0)}{24n}\right)
\end{aligned}$$

3.4. **Example.** We also apply the saddlepoint approximation to

$$S_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma}, \qquad X_i \text{ i.i.d. } \sim \mathrm{Exp}(\lambda).$$

For exponential distribution $\mathrm{Exp}(\lambda)$ with the density function

$$f(x) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}},$$

we have

$$M(X;t) = \frac{1}{\lambda t}, \quad K(X;t) = -\log(1 - \lambda t)$$

which are convergent for $|t| < 1/\lambda$. We also have

$$K'(X;t) = \frac{\lambda}{1-\lambda t}, \quad K''(X;t) = \frac{\lambda^2}{(1-\lambda t)^2}, \quad K^{(3)}(X;t) = \frac{2\lambda^3}{(1-\lambda t)^3}, \quad K^{(4)}(X;t) = \frac{6\lambda^4}{(1-\lambda t)^4}.$$

Solving $K'(X;t) = \bar{x}$ gives us the saddlepoint

$$t_0 = \frac{1}{\lambda} - \frac{1}{\bar{x}}$$

and

$$K''(X;t_0) = \bar{x}^2, \quad K^{(3)}(X;t_0) = 2\bar{x}^3, \quad K^{(4)}(X;t_0) = 6\bar{x}^4.$$

Therefore

$$\lambda_r(t_0) = K^{(r)}(X;t_0)/(K''(X;t_0))^{r/2} \quad \Rightarrow \quad \lambda_3(t_0) = 2, \quad \lambda_4(t_0) = 6$$

Finally we obtain the saddlepoint approximation

$$\begin{aligned} f_n(\bar{x}) &\approx \sqrt{\frac{n}{2\pi K''(t_0)}} e^{n[K(t_0)-t_0\bar{x}]} \Big(1 + \frac{3\lambda_4(t_0) - 5\lambda_3^2(t_0)}{24n}\Big) \\ &= \sqrt{\frac{n}{2\pi}} e^{-\frac{n\bar{x}}{\lambda}} (\bar{x})^{n-1} (\frac{e}{\lambda})^n \left(1 - \frac{1}{12n}\right). \end{aligned}$$

We also know that the exact distribution of the sample means is $\Gamma(n, \lambda/n)$, and the results are shown in Figure 2.

4. Conclusions

This project gives a detailed proof of the two commonly used asymptotic approximations in statistics. Given n i.i.d. random variables, the Edgeworth expansion of the density function of the standardized sample means $S_n = (\sum_{i=1}^{n} X_i - n\mu)/\sqrt{n}\sigma$ is

$$f(S_n;x) = \Phi(x)\Big(1 + \frac{\rho_3 H_3(x)}{6\sqrt{n}} + \frac{\rho_4 H_4(x)}{24n} + \frac{\rho_3^2 H_6(x)}{72n}\Big) + O(n^{-3/2}).$$

The saddlepoint approximation of the density function of the sample means $\bar{X} = \sum_{i=1}^{n} X_i/n$ is

$$f_n(\bar{x}) = \sqrt{\frac{n}{2\pi K''(t_0)}} e^{n[K(t_0)-t_0\bar{x}]} \Big(1 + \frac{3\lambda_4(t_0) - 5\lambda_3^2(t_0)}{24n} + O(n^2)\Big).$$

These two asymptotic approximations have numerous applications [3] since many statistical inferences are based on the estimation of the density functions.

When doing numerical experiments, in order to compare the two expansions in a single plot, the author tried but failed to derive the Edgeworth expansion for the sample means $\bar{X}$ or the saddlepoint approximation for the standardized sample means S_n. It seems that rescaling or translation of the statistics will result in much more complicated form of integrals.

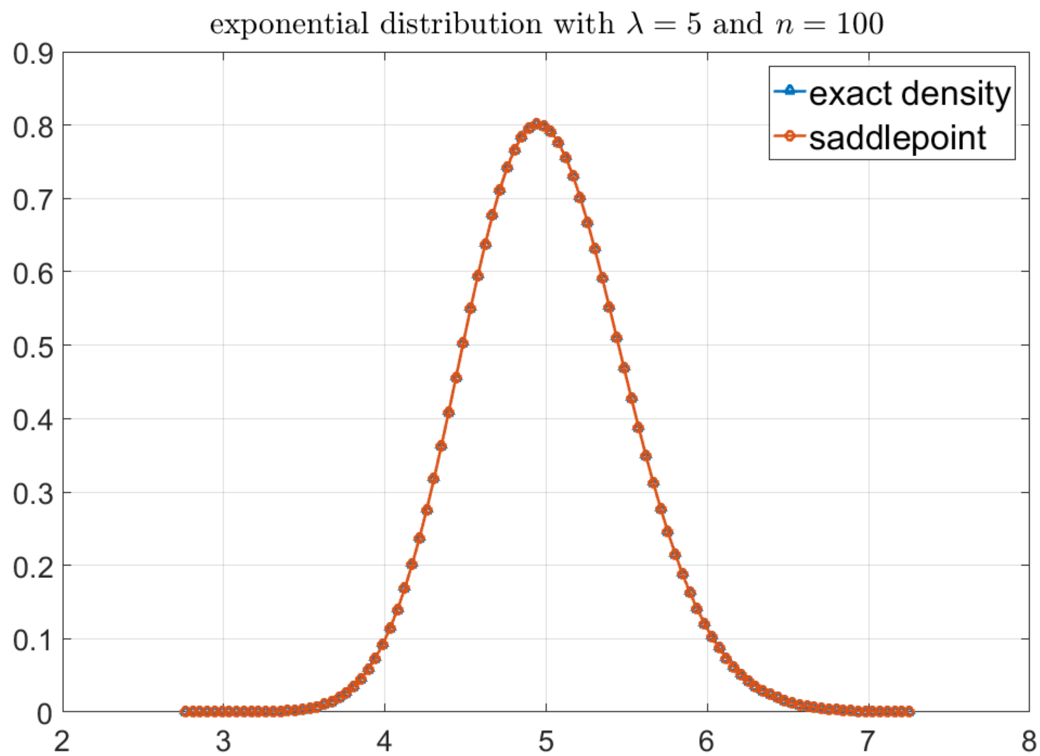

Figure 2. For n independent and identically distributed random variables from the exponential distribution family $\mathrm{Exp}(\lambda)$, use saddlepoint approximation to approximate the density function of the sample means $\bar{X}$

References

[1] O.E. Barndorff-Nielsen and D.R. Cox. Asymptotic techniques for use in statistics. pages 90–91, 1989.

[2] H. E. Daniels. Saddlepoint approximations in statistics. *Ann. Math. Statist.*, 25(4):631–650, 12 1954.

[3] Anirban DasGupta. Asymptotic theory of statistics and probability. 2008.

[4] David Freedman and William Feller. An introduction to probability theory and its applications. *Econometrica*, 35:509–510, 07 1967.

[5] Jens Ledet Jensen. Saddlepoint approximations. pages 18–19, 1995.

[6] David Vernon Widder. The laplace transform. 1946.