NTNU | Norwegian University of Science and Technology

SADDLEPOINT APPROXIMATION

Nora Aasen

November 30, 2022

Table of contents

Motivation

Problem

We wish to find density or p-value from unknown distribution.

Solution

Can use the cumulant generating function to approximate the distribution, if this is known.

The cumulant genertaing function is defined as follows, where $s \in (a, b)$ is an open set around 0 where the MGF is defined.

$$M(s) = E(e^{sX}) = \int_{-\infty}^{\infty} e^{xs} f(x) dx \qquad \text{(MGF)}$$

$$K(s) = \ln M(s) \qquad \text{(CGF)}$$

Cumulants

The cumulant generating function has cumulants (similar to moments for the moment generating function), and the k'th cumulant is defined as,

$$K^{(k)}(0) = \kappa_k.$$

From the first four cumulants we can compute

- The mean: $\kappa_1 = E(X) = \mu$.
- The variance: $\kappa_2 = E((X - E(X))^2) = \sigma^2$.
- The skewness: $\tilde{\kappa}_3$ which is the standardized 3rd cumulant $\kappa_3 = E((X - E(X))^3)$.
- The excess kurtosis: $\tilde{\kappa}_4$ which is the standardized 4th cumulant $\kappa_4 = E((X - E(X))^3) - 3(E((X - E(X))^2))^2$.

Density

Let $s \in (a, b)$, the open set on which the MGF converges, and $x \in \mathcal{I}_{\mathcal{X}}$, the interior of the span of the support of X. Then the saddlepoint approximation for the density/pmf of X is given as follows.

$$\hat{f}(x) = \frac{1}{\sqrt{2\pi K''(\hat{s})}} \exp(K(\hat{s}) - \hat{s}x) \qquad \text{(Continuous)}$$

$$\hat{p}(k) = \frac{1}{\sqrt{2\pi K''(\hat{s})}} \exp(K(\hat{s}) - \hat{s}k) \qquad \text{(Discrete)}$$

The *saddlepoint* $\hat{s}$ is determined from the *saddlepoint equation*

$$K'(\hat{s}) = x \quad \text{or} \quad K'(\hat{s}) = k.$$

Derivation using Laplace Approximation[1]

- Assume g is smooth, and has a global minimum on (c, d) in $\hat{x}$. Then the Laplace approximation is:

$$\int_c^d e^{-g(x)}dx \approx \frac{\sqrt{2\pi}e^{-g(\hat{x})}}{\sqrt{g''(\hat{x})}}$$

- Use Laplace approximation on $nK(s)$, which is the CGF of $n\bar{X} = \sum_{i=1}^{n} X_i$, where the X_i's are independent and all have the same CGF $K(s)$.

$$e^{nK(s)} = \int_{-\infty}^{\infty} e^{sn\bar{x}+\ln f(\bar{x})}d\bar{x} \approx \frac{\sqrt{2\pi}e^{-g(\hat{x})}}{\sqrt{g''(\hat{x})}}$$

where $g(\bar{x}, s) = -sn\bar{x} - \ln f(\bar{x})$.

[1] From chapter 2 in Butler 2007

Derivation using Edgeworth Expansions[2]

- Any density $f(x)$ can be used to create a class of densities called *the natural exponential family* indexed over s as follows,

$$e^{K(s)} = \int_{-\infty}^{\infty} e^{sx} f(x) dx$$
$$\implies 1 = \int_{-\infty}^{\infty} e^{sx-K(s)} f(x) dx = \int_{-\infty}^{\infty} f(x; s) dx$$

This technique is called *exponential tilting*.

- Since the expression integrates to 1 for all $s \in (a, b)$, this is a family of densities

$$f(x; s) = \exp\{sx - K(s)\} f(x).$$

[2] From chapter 5 in Butler 2007

- Furthermore, we want to approximate $f(x;s)$ using *Edgeworth Expansions*.

$$f(x) \approx \frac{1}{\sigma}\phi(z)\left[1 + a_1\tilde{\kappa}_3 H_3(z) + a_2\tilde{\kappa}_4 H_4(z) + a_3\tilde{\kappa}_5 H_5(z) + a_2\tilde{\kappa}_3^2 H_6(z) + \dots\right],$$

  where $z = (x-\mu)/\sigma$
- Once again we get the saddlepoint equation, by choosing

$$\mu_{\hat{s}} = K'(\hat{s}) = x,$$

- and

$$\begin{aligned} f(x;s) &= \exp\{sx - K(s)\}f(x) \\ \implies f(x) &\approx \exp\left\{K(\hat{s}) - \hat{s}x\right\}\frac{1}{\sqrt{2\pi K''(\hat{s})}}\left\{1 + \left(\frac{1}{8}\tilde{\kappa}_4(\hat{s}) - \frac{5}{24}\tilde{\kappa}_3^2(\hat{s})\right)\right\} \end{aligned}$$

Some important properties

- Discrete saddlepoint approximation can only be applied directly if X is integer-valued, meaning the random variable take values $\{0, \pm 1, \pm 2, \pm 3, \ldots\}$.
- The saddelpoint approximated density is not guaranteed to integrate to 1.
- The continuous saddlepoint density is equivariant under linear transformations.
- Symmetry exhibited in the density/mass function of X will be preserved in the saddlepoint approximation.
- Convergence rate is $\mathcal{O}(n^{-1})$, and can be improved to $\mathcal{O}(n^{-3/2})$ in some circumstances (Goutis and Casella 1999). In comparison, the central limit theorem converges with a rate of $\mathcal{O}(n^{-1/2})$.

Example - Normal

Consider a random variable $X \sim \mathcal{N}(0, 1)$. Then,

$$K(s) = \frac{s^2}{2} \quad \text{and} \quad \hat{s} = x$$

It follows easily that the saddlepoint approximation $\hat{f}$ is,

$$\hat{f}_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$$

and hence, exact. To show that $Y \sim \mathcal{N}(\mu, \sigma^2)$ is also exact, we can either compute it directly or use the equivariance property which gives us

$$\hat{f}_Y(y) = \frac{1}{\sigma}\hat{f}_X(x(y)), \quad x(y) = \frac{y - \mu}{\sigma}.$$

Example - Poisson I

Consider $X \sim Poisson(\lambda)$, which has CGF $K(s) = \lambda(\exp(s) - 1)$. The saddlepoint will be

$$K'(s) = \lambda \exp(s) = x \implies \hat{s} = \ln\left(\frac{x}{\lambda}\right),$$

and the approximated saddlepoint density is

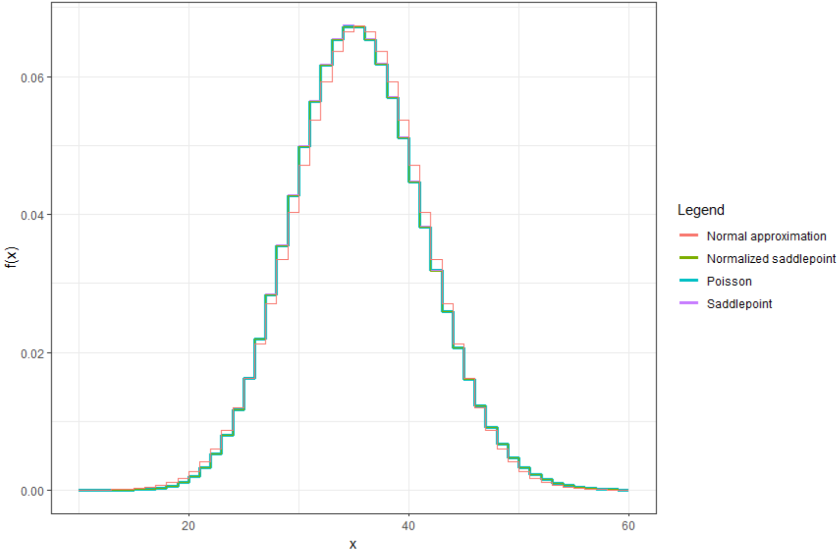
$$\begin{aligned}\hat{f}(x) &= \frac{1}{\sqrt{2\pi x}} \exp\{x - \lambda - \ln\left(\frac{x}{\lambda}\right) x\} \\ &= \frac{1}{\sqrt{2\pi x} x^x e^{-x}} \lambda^x e^{-\lambda}\end{aligned}$$

Example - Poisson II

Running the example in RStudio with $\lambda = 35$ gave the following result.

	Poisson	Saddlepoint	Normalized sp	Normal approx
Abs error	0	0.002456	0.000340	0.042681
Tail: x = 250	2.0282e-122	2.0289e-122	2.0239e-122	1.0950e-288

Example - Poisson III



Saddlepoint approximation of a p-value

Idea

Since we now have a quite good approximation $f \approx \hat{f}$, then perhaps a good approximation for F is simply the numerical integral,

$$F(x) \approx \int_{-\infty}^{x} \hat{f}(t)dt.$$

However, there has been derived one approximation for continuous random variables, denoted $\hat{F}$, and three continuity-corrected approximations for discrete random variables, denoted $\hat{P}r_1, \hat{P}r_2$ and $\hat{P}r_3$.

Continuous approximation

This approximation to a CDF where X is a continuous random variable was first presented in Lugannani and Rice 1980.

Continuous

$$\hat{F}(x) = \begin{cases} \Phi(\hat{\omega}) + \phi(\hat{\omega})(1/\hat{\omega} - 1/\hat{u}) & x \neq \mu \\ \frac{1}{2} + \frac{K'''(0)}{6\sqrt{2\pi}K''(0)^{3/2}} & x = \mu \end{cases}$$

$$\hat{\omega} = \text{sgn}(\hat{s})\sqrt{2(\hat{s}x - K(\hat{s}))}, \quad \hat{u} = \hat{s}\sqrt{K''(\hat{s})}$$

First continuity correction

These approximations to a CDF where X is a discrete random variable was first presented in Daniels 1987.

Discrete 1

$$\hat{P}r_1(X \geq k) = \begin{cases} 1 - \Phi(\hat{\omega}) - \phi(\hat{\omega})(1/\hat{\omega} - 1/\tilde{u}_1) & k \neq \mu \\ \frac{1}{2} - \frac{1}{\sqrt{2\pi}}\left[\frac{K'''(0)}{6K''(0)^{3/2}} - \frac{1}{2\sqrt{K''(0)}}\right] & k = \mu \end{cases}$$

$$\hat{\omega} = \operatorname{sgn}(\hat{s})\sqrt{2(\hat{s}k - K(\hat{s}))}, \quad \tilde{u}_1 = (1 - \exp(-\hat{s}))\sqrt{K''(\hat{s})}$$

Second continuity correction

Discrete 2

$$\hat{P}r_2(X \geq k) = \begin{cases} 1 - \Phi(\tilde{\omega}_2) - \phi(\tilde{\omega}_2)(1/\tilde{\omega}_2 - 1/\tilde{u}_2) & k^- \neq \mu \\ \frac{1}{2} - \frac{K'''(0)}{6\sqrt{2\pi}K''(0)^{3/2}} & k^- = \mu \end{cases}$$

$$k^- = k - 1/2, \quad K'(\tilde{s}) = k^-$$

$$\tilde{\omega}_2 = \operatorname{sgn}(\tilde{s})\sqrt{2(\tilde{s}k - K(\tilde{s}))}, \quad \tilde{u}_2 = 2\sinh(\tilde{s}/2))\sqrt{K''(\tilde{s})}$$

Tails of discrete distributions

- Notice that the approximations for the discrete case give us the survival function $P(X \geq k)$. This is to avoid “Difficult notational problems” Butler 2007.
- In general, the first and second continuity correction are most accurate. However, the first lack certain theoretical properties and therefore might be less preferable.
- The left tail approximations for a random variable X can be shown to be $\hat{F}_i(k) = 1 - \hat{P}r_i(X \geq k+1)$, for $i = 2, 3$. However, for the first continuity correction we get instead a slightly different expression, called

$$\hat{F}_{1a}(k) \neq \hat{F}_1(k) = 1 - \hat{P}r_1(X \geq k+1).$$

Properties

	Preserves symmetry	**Linear transformations**
Continuous	✓	✓
$\hat{P}r_1$	Yes, if using (1)	Only increasing
$\hat{P}r_2$	✓	✓
$\hat{P}r_3$	✓	✓

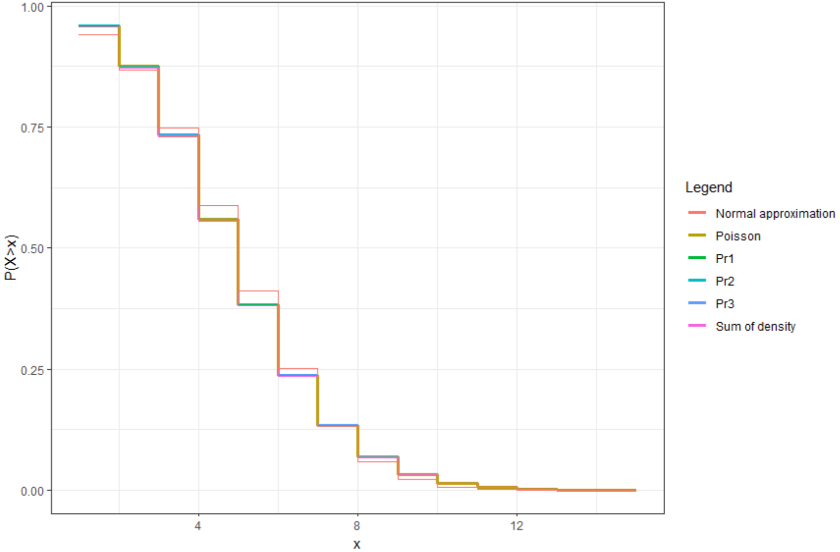
$$\begin{array}{ll} \hat{F}_{1a}(k) & \textit{if} \quad k \leq \mu \\ 1 - \hat{P}r_1(X \geq k+1) & \textit{if} \quad k+1 \geq \mu \end{array} \tag{1}$$

Example 2 - Poisson I

Once again we consider $X \sim Poisson(\lambda)$, and implement the different approximation methods in RStudio, for $\lambda = 5$.

k	$P(X \geq k)$	$\hat{P}r_1(X \geq k)$	$\hat{P}r_2(X \geq k)$	$\hat{P}r_3(X \geq k)$	$\sum_{i=k}^{\infty} \hat{p}(i)$	$P(Z \geq k - \frac{1}{2})$
2	0.9595	0.9594	0.9601	0.9572	0.9576	0.9412
11	0.0136	0.0136	0.0137	0.0140	0.0135	0.0069
26	3.04e-11	3.05e-11	3.05e-11	3.41e-11	2.99e-11	2.41e-20

Example 2 - Poisson II



Multivariate case

Density

For a continuous random m-vector X with CGF $K(s)$, the saddlepoint density is defined as,

$$\hat{f}(x) = (2\pi)^{-m/2}|K''(\hat{s})|^{-1/2}\exp\{K(\hat{s}) - \hat{s}^T x\},$$

where $\hat{s}$ uniquely solves $K'(\hat{s}) = x$, with little to no conceptual difference from the univariate setting.

CDF

For multivariate CDFs there existed no general computing scheme for the probability $P(X \in A)$ for arbitrary sets A, although there exists methods for special cases (in particular when A is rectangular).

Conditional distribution

Density

For both continuous, discrete or a mixture of these two, the conditional saddlepoint approximation given by,

$$\hat{f}_{Y|X=x}(y) = \frac{\hat{f}_{X,Y}(x,y)}{\hat{f}_X(x)}$$

is well defined and the same properties as before holds.

CDF

There exists no general scheme for computing $P(Y \in A|X = x)$, where A is any set. However, when $\dim(Y) = 1$, there exists an expression for both the continuous and discrete case, derived by Skovgaard 1987, and based on the expressions we have seen earlier.

Exponential family

Consider an exponential family on the form

$$f(z;\theta) = \exp\{\theta x(z) - c(\theta) + d(x)\}$$

where θ is a canoncial sufficient parameter, $x(z) = X$ is a canonical sufficient statistic, and $\theta \in \Theta$, where Θ is such that the family is regular. Then the following is true.

- Fisher's factorization theorem gives us

$$\mathcal{L}(\theta) = f_Z(z;\theta) = f_X(x;\theta)g(z;x).$$

- The log-likelihood is given by

$$\ell(\theta) \propto \theta x - c(\theta)$$

- $K_X(s) = c(s+\theta) - c(\theta)$.
- $E(X) = c'(\theta)$ and $\mathrm{Var}(X) = c''(\theta)$.
- MLE of θ is uniquely (if it exists) given by $c'(\theta) = x$.

Example - Poisson

Consider $X \sim Poisson(\lambda)$. This has density

$$\begin{aligned} f(x;\lambda) &= \frac{1}{x!}e^{-\lambda}\lambda^x \\ &= \exp\{\ln(\lambda)x - \lambda - \ln(x!)\} \end{aligned}$$

and CGF

$$\begin{aligned} K(s) &= \exp(s+\theta) - \exp(\theta) \\ &= \exp(s + \ln\lambda) - \lambda \\ &= \lambda(\exp(s) - 1). \end{aligned}$$

Saddlepoint approximation I

Notice first that

$$K'(\hat{s}) = c'(\hat{s} + \theta) = x \implies \hat{s} = \hat{\theta} - \theta.$$

Then, inserting this into the saddlepoint approximated density, we get

$$\begin{aligned}
\hat{f}(x;\theta) &= \frac{1}{\sqrt{2\pi K''(\hat{s})}} \exp(K(\hat{s}) - \hat{s}x) \\
&= \frac{1}{\sqrt{2\pi c''(\hat{\theta})}} \exp\{c(\hat{\theta}) - c(\theta) - (\hat{\theta} - \theta)x\} \\
&= (2\pi j(\hat{\theta}))^{-1/2} \frac{\mathcal{L}(\theta)}{\mathcal{L}(\hat{\theta})}.
\end{aligned}$$

Saddlepoint approximation II

It is also possible to show something similar using tilting, as follows.

$$f(x;\theta) = \frac{\mathcal{L}(\theta)}{\mathcal{L}(\theta+\hat{s})} f(x;\theta+\hat{s}).$$

The Edgeworth expansion used the standardized cumulants $\tilde{\kappa}_i$, which for the exponential is on the form

$$\tilde{\kappa}_i = \frac{K^{(i)}(0)}{\sigma^i} = \frac{c^{(i)}(\hat{\theta})}{(c''(\hat{\theta}))^{i/2}} = \frac{-\ell^{(i)}(\hat{\theta})}{(-\ell''(\hat{\theta}))^{i/2}}.$$

Since ℓ is analytic on Θ, we can rewrite $\ell(\theta)$ as a Taylor polynomial where the coefficients will be the derivatives in $\hat{\theta}$. The conclusion becomes that $\hat{f}(x;\theta)$ depends only on the likelihood.

References I

- Butler, Ronald W., *Saddlepoint Approximations with Applications*. Cambridge UP, 2007, https://doi.org/10.1017/CBO9780511619083. Cambridge Series in Statistical and Probabilistic Mathematics.
- Goutis, Constantino, and George Casella, "Explaining the Saddlepoint Approximation". *The American Statistician*, vol. 53, no. 3, 1999, Publisher: [American Statistical Association, Taylor & Francis, Ltd.], pp. 216–24. *JSTOR*, https://doi.org/10.2307/2686100.
- Lugannani, Robert, and Stephen Rice, "Saddle Point Approximation for the Distribution of the Sum of Independent Random Variables". *Advances in Applied Probability*, vol. 12, no. 2, 1980, Publisher: Applied Probability Trust, pp. 475–90. *JSTOR*, https://doi.org/10.2307/1426607.

References II

- Daniels, H. E., "Tail Probability Approximations". *International Statistical Review / Revue Internationale de Statistique*, vol. 55, no. 1, 1987, Publisher: [Wiley, International Statistical Institute (ISI)], pp. 37–48. *JSTOR*, https://doi.org/10.2307/1403269.
- Skovgaard, Ib M., "Saddlepoint Expansions for Conditional Distributions". *Journal of Applied Probability*, vol. 24, no. 4, 1987, Publisher: Applied Probability Trust, pp. 875–87. *JSTOR*, https://doi.org/10.2307/3214212.