Saddlepoint Approximation

Zachary R. McCaw

January 12, 2024

Background

1.1 Cumulant generating function

The **saddlepoint approximation** is a formula for approximating the density or distribution function of a statistic. Suppose X is a continuous random variable with density $f_X(\cdot)$. The **moment generating function** (MGF) is defined as:

$$M_X(t) \equiv \mathbb{E}\big(e^{tX}\big) = \int_{-\infty}^{\infty} e^{tx} f_X(x)dx, \tag{1}$$

and the **cumulant generating function** (CGF) is defined as:

$$K_X(t) \equiv \ln M_X(t). \tag{2}$$

The CGF is so named because the coefficients κ_j in the power series expansion:

$$K_X(t) = \sum_{j=0}^{\infty} \frac{\kappa_j}{j!} t^j$$

define the cumulants of X.

Example 1.1. Suppose $X_i \sim f_X$ and let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. The moment generating function of the sample mean $\bar{X}$ is:

$$M_{\bar{X}}(t) = e^{t\bar{X}} = e^{\frac{t}{n}\sum_{i=1}^{n} X_i} = \prod_{i=1}^{n} e^{\frac{t}{n}X_i} = \prod_{i=1}^{n} M_X(t/n) = \big\{M_X(t/n)\big\}^n.$$

The cumulant generating function is:

$$K_{\bar{X}}(t) = \ln M_{\bar{X}}(t) = n \ln M_X(t/n) = nK_X(t/n).$$

Note that the moment/cumulant generating function for the mean appears on the LHS, while the moment/cumulant generating function for an individual X_i appears on the RHS. ♠

1.2 Gamma function

The **gamma** function is defined by:

$$\Gamma(k) = \int_0^{\infty} x^{k-1} e^{-x} dx.$$

Using integration by parts:

$$\Gamma(k+1) = \int_0^\infty x^k e^{-x} dx = \Big[x^k e^{-x}\Big]_{x=0}^\infty + k\int_0^\infty x^{k-1} e^{-x} dx = k\Gamma(k).$$

Consequently, for $k \in \mathbb{N}$:

$$\Gamma(k) = (k-1)!$$

Definition 1.1. Stirling's approximation to the factorial is:

$$\hat{k}! = \sqrt{2\pi} k^{k+1/2} e^{-k} + \mathcal{O}(k^{-1}).$$

For the derivation, see the notes on Laplace approximation. ■

Density approximation

Definition 2.1. The **saddlepoint equation** is:

$$\dot{K}(\hat{s}) = x. \tag{3}$$

Here $\dot{K}(\cdot)$ is the first derive of the cumulant generating function (2) and $\hat{s}$ is the *saddlepoint*. Notice that $\hat{s}$ is defined implicitly, and that $\hat{s} = \hat{s}(x)$ is a function of x. ■

Theorem 2.1. Suppose X is a continuous random variable. The **saddlepoint approximation** (SPA) to the density of X at x is:

$$\tilde{f}(x) = \frac{1}{\sqrt{2\pi\ddot{K}(\hat{s})}} \exp\{K(\hat{s}) - \hat{s}x\}. \tag{4}$$

Here $\hat{s} = \hat{s}(x)$ is the solution to the saddlepoint equation (3). As written, the SPA to the density is not necessarily normalized. As such, the SPA is typically renormalized to:

$$\hat{f}(x) = \frac{1}{C}\tilde{f}(x), \qquad C = \int_{\mathcal{X}} \tilde{f}(x)dx.$$

Here $\mathcal{X} = \{x : f(x) > 0\}$ is the support of X. In addition, the SPA is only valid on the largest open neighborhood of zero within which the moment generating function (1) converges. □

Remark 2.1. The SPA to the density in (4) may also be applied to discrete random variables. However, points on the boundary of the support may require special consideration. ♦

Example 2.2 (*Exponential distribution*)**.** Consider the exponential distribution with density $f(x) = \lambda e^{-\lambda x}$, $x > 0$. The MGF is $M(t) = \frac{\lambda}{\lambda - t}$ and the CGF is $K(t) = \ln(\lambda) - \ln(\lambda - t)$. The first derivative of K is $\dot{K}(t) = \frac{1}{\lambda - t}$. The saddlepoint equation is:

$$\dot{K}(\hat{s}) = \frac{1}{\lambda - \hat{s}} \stackrel{\text{Set}}{=} x \implies \hat{s}(x) = \lambda - \frac{1}{x}.$$

The second derivative of K is $\ddot{K}(t) = (\lambda - t)^{-2}$; evaluating at the saddlepoint gives $\ddot{K}(\hat{s}) = x^2$. Evaluating K at $\hat{s}$ gives $K(\hat{s}) = \ln(\lambda) + \ln(x)$. Substituting these quantities into (4):

$$\tilde{f}(x) = \frac{1}{\sqrt{2\pi x^2}} \exp\{\ln(\lambda) + \ln(x) - (\lambda x - 1)\} = \frac{\lambda e^1}{\sqrt{2\pi}} \exp\{-\lambda x\}.$$

Upon renormalizing, the original density $\hat{f}(x) = f(x)$ is recovered. ♠

Example 2.3 (*Binomial distribution*)**.** Consider $X \sim \text{Binom}(n, p)$. The MGF is $M_X = \{(1 - p) + pe^t\}^n$ and the CGF is $K(t) = n \ln\{(1-p) + pe^t\}$. The first derivative of $K(t)$ is:

$$\dot{K}(t) = \frac{npe^t}{(1-p) + pe^t}.$$

The saddlepoint equation is:

$$\dot{K}(\hat{s}) = \frac{npe^{\hat{s}}}{(1-p) + pe^{\hat{s}}} \stackrel{\text{Set}}{=} x \implies \hat{s}(x) = \ln\left\{\frac{(1-p)x}{p(n-x)}\right\}.$$

The second derivative is:

$$\ddot{K}(t) = \frac{npe^t}{(1-p) + pe^t}\left\{1 - \frac{pe^t}{(1-p) + pe^t}\right\}.$$

Evaluating $\ddot{K}$ at $\hat{s}$ gives $\ddot{K}(\hat{s}) = x(1 - x/n)$. Evaluating K at $\hat{s}$ gives:

$$K(\hat{s}) = n \ln\left\{(1-p) \cdot \frac{n}{n-x}\right\}.$$

Substituting into (4) gives:

$$\begin{aligned}
\tilde{f}(x) &= \frac{1}{\sqrt{2\pi x(1-x/n)}} \cdot e^{n \ln\left\{(1-p)\cdot\frac{n}{n-x}\right\}} \cdot e^{-x \ln\left\{\frac{(1-p)x}{p(n-x)}\right\}} \\
&= \frac{1}{\sqrt{2\pi x(1-x/n)}}(1-p)^n \left(\frac{n}{n-x}\right)^n \left\{\frac{p(n-x)}{(1-p)x}\right\}^x \\
&= \widehat{\binom{n}{x}} p^x (1-p)^{n-x},
\end{aligned}$$

where $\widehat{\binom{n}{x}}$ is Stirling's approximation to the binomial coefficient. Renormalization gives:

$$\hat{f}(x) = \widehat{\binom{n}{x}}\binom{n}{x}^{-1} p^x(1-p)^{n-x}.$$

♠

Example 2.4 (*Score statistic*)**.** Suppose $X_i \stackrel{\text{IID}}{\sim} f_\theta$, where f but not θ is known. The score statistic is:

$$\Psi(\theta) = \sum_{i=1}^{n} \psi(X_i; \theta) = \sum_{i=1}^{n} \frac{\partial \ln f(X_i; \theta)}{\partial \theta}.$$

The individual-level MGF is given by:

$$M_i(t) = \mathbb{E}\big\{e^{t\psi_i(X_i;\theta)}\big\},$$

and the CGF by $K_i(t) = \ln M_i(t)$. With $K_i(t)$, a saddlepoint approximation can be developed for the sum $\Psi(\theta)$, whose CGF is $K(t) = nK_i(t)$. ♠

Distribution approximation

Theorem 3.1 (*Lugannani Rice approximation*)**.** Suppose X is a continuous random variable with mean $\mu = \mathbb{E}(X)$. The SPA to the cumulative distribution function $F(x)$ is:

$$\hat{F}(x) = \begin{cases} \Phi(\hat{w}) + \phi(\hat{w})(1/\hat{w} - 1/\hat{u}), & x \neq \mu, \\ \frac{1}{2} + \frac{K^{(3)}(0)}{6\sqrt{2\pi}\ddot{K}(0)^{3/2}}, & x = \mu. \end{cases} \tag{5}$$

Here $\hat{w} = \text{sign}(\hat{s})\sqrt{2\{\hat{s}x - K(\hat{s})\}}$, $\hat{u} = \hat{s}\sqrt{\ddot{K}(\hat{s})}$, and $\hat{s}$ is the saddlepoint from (3). $\phi(\cdot)$ and $\Phi(\cdot)$ denote the standard normal density and distribution functions, respectively. □

Example 3.5 (*Gamma distribution*)**.** Suppose $X \sim \text{Gamma}(\alpha, 1)$, with $f(x) = \frac{1}{\Gamma(\alpha)}x^{\alpha-1}e^{-x}$, for $x > 0$. The MGF is $M(t) = (1-t)^{-\alpha}$ and the CGF is $K(t) = -\alpha\ln(1-t)$. The first derivative of the CGF is $\dot{K}(t) = \frac{\alpha}{1-t}$. The saddlepoint equation is:

$$\dot{K}(\hat{s}) = \frac{\alpha}{1-\hat{s}} \overset{\text{Set}}{=} x \implies \hat{s}(x) = 1 - \frac{\alpha}{x}. \tag{6}$$

The second derivative of the CDF is $\ddot{K}(t) = \frac{\alpha}{(1-t)^2}$. Evaluating at $\hat{s}$ gives $\ddot{K}(\hat{s}) = x^2/\alpha$. Evaluating K at $\hat{s}$ gives $K(\hat{s}) = -\alpha\ln(\alpha/x)$. Calculating $\hat{w}$:

$$\hat{w} = \text{sign}(\hat{s})\sqrt{2\{\hat{s}x - K(\hat{s})\}} = \text{sign}(x-\alpha)\sqrt{2\{(x-\alpha) + \alpha\ln(\alpha/x)\}}.$$

Calculating $\hat{u}$:

$$\hat{u} = \hat{s}\sqrt{\ddot{K}(\hat{s})} = \left(1 - \frac{\alpha}{x}\right)\frac{x}{\sqrt{\alpha}} = \frac{(x-\alpha)}{\sqrt{\alpha}}.$$

With $\hat{w}$ and $\hat{u}$, the CDF may be approximated as in (5). ♠

Theorem 3.2 (*Daniels approximation*)**.** Suppose X is a discrete random variable with mean $\mu = \mathbb{E}(X)$. The SPA to the distribution function is:

$$\hat{\mathbb{P}}(X \geq x) = \begin{cases} 1 - \Phi(\hat{w}) - \phi(\hat{w})(1/\hat{w} - 1/\hat{u}), & x \neq \mu, \\ \frac{1}{2} - \frac{1}{\sqrt{2\pi}}\left\{\frac{K^{(3)}(0)}{6\ddot{K}(0)^{3/2}} - \frac{1}{2\sqrt{\ddot{K}(0)}}\right\}, & x = \mu. \end{cases} \tag{7}$$

Here $\hat{w} = \text{sign}(\hat{s})\sqrt{2\{\hat{s}x - K(\hat{s})\}}$, $\hat{u} = (1 - e^{-\hat{s}})\sqrt{\ddot{K}(\hat{s})}$, and $\hat{s}$ is the saddlepoint from (3). $\phi(\cdot)$ and $\Phi(\cdot)$ denote the standard normal density and distribution functions, respectively. In comparison to the approximation for the continuous case, notice that (i) the discrete approximation is given in terms of the right-tail, (ii) the definition of $\hat{u}$ is modified, and (iii) the approximation at the mean is modified. □

Example 3.6 (*Poisson distribution*)**.** Suppose $X \sim \text{Poisson}(\lambda)$, with $f(x) = e^{\lambda}\lambda^x/x!$ and $x \in \mathbb{Z}^{\geq 0}$. The MGF is $M(t) = e^{-\lambda+\lambda e^t}$ and the CGF is $K(t) = -\lambda + \lambda e^t$. The first derivative of the CGF is $\dot{K}(t) = \lambda e^t$. The saddlepoint equation is:

$$\dot{K}(\hat{s}) = \lambda e^{\hat{s}} \overset{\text{Set}}{=} x \implies \hat{s} = \ln\left(\frac{x}{\lambda}\right).$$

The second derivative of the CGF is $\ddot{K}(t) = \lambda e^t = \dot{K}(t)$, thus $\ddot{K}(\hat{s}) = x$. Evaluating K at $\hat{s}$ gives $K(\hat{s}) = x - \lambda$. Calculating $\hat{w}$:

$$\hat{w} = \text{sign}(\hat{s})\sqrt{2\{\hat{s}x - K(\hat{s})\}} = \text{sign}(x-\lambda)\sqrt{2\{x\ln(x/\lambda) - (x-\lambda)\}}$$

Calculating $\hat{u}$:

$$\hat{u} = (1 - e^{-\hat{s}})\sqrt{\ddot{K}(\hat{s})} = \left(1 - \frac{\lambda}{x}\right)\sqrt{x}.$$

With $\hat{w}$ and $\hat{u}$, the CDF may be approximated as in (7). ♠

Conditionals

4.1 Density

Consider approximating the conditional density of Y given $X = x$:

$$f(y|x) \equiv \frac{f(x,y)}{f(x)}.$$

A saddlepoint approximation to $f(y|x)$ is obtained by separately approximating the joint density $f(x,y)$ and the marginal density $f(x)$. Let $K(s,t)$ denote the joint cumulant generating function:

$$K(s,t) = \ln M(s,t) = \ln \mathbb{E}\{e^{sX+tY}\} = \int e^{sx+ty} f(x,y)dxdy.$$

Let $\dot{K}$ denote the gradient of K with respect to both s and t:

$$\dot{K}(s,t) \equiv \frac{\partial K(s,t)}{\partial(s,t)}.$$

The *unconstrained* saddlepoints $\hat{s}$ and $\hat{t}$ satisfy:

$$\dot{K}(\hat{s},\hat{t}) = (x,y)$$

where (x,y) are is the tuple at which evaluation of $f(y|x)$ is sought. The approximation to the joint density $f(x,y)$ is:

$$\hat{f}(x,y) = (2\pi)^{-m/2}\det\{\ddot{K}(\hat{s},\hat{t})\}^{-1/2}e^{K(\hat{s},\hat{t})-\hat{s}x-\hat{t}y},$$

where $m = \dim(x) + \dim(y)$ and $\ddot{K}$ is the Hessian of K:

$$\ddot{K}(s,t) = \frac{\partial}{\partial(s,t)'}\dot{K}(s,t) = \frac{\partial}{\partial(s,t)'}\frac{\partial K(s,t)}{\partial(s,t)}.$$

To approximate the marginal density $f(x)$, the *constrained* saddlepoint $\tilde{s}$ is required:

$$\dot{K}_s(\tilde{s},0) = x.$$

Here $\dot{K}_s = \partial_s K(s,t)$ denotes the partial of $K(s,t)$ with respect to s only. The saddlepoint approximation ot $f(x)$ is:

$$\hat{f}(x) = (2\pi)^{-m_x/2}\det\{\ddot{K}_{ss}(\tilde{s},0)\}^{-1/2}e^{K(\tilde{s},0)-\tilde{s}x},$$

where $m_x = \dim(x)$ and $\ddot{K}_{ss}$ is the Hessian of K with respect to s only.

Theorem 4.1 (*Barndorff-Nielsen and Cox*)**.** The double saddlepoint approximation to the conditional density of Y given $X = x$ is:

$$\hat{f}(y|x) = (2\pi)^{-m_y/2} \left(\frac{\det\{\ddot{K}_{ss}(\tilde{s}, 0)\}}{\det\{\ddot{K}(\hat{s}, \hat{t})\}} \right)^{1/2} \exp\left[\left\{ K(\hat{s}, \hat{t}) - \hat{s}x - \hat{t}y \right\} - \left\{ K(\tilde{s}, 0) - \tilde{s}x \right\} \right].$$

Here $m_y = \dim(y)$ and the remaining terms are defined above. □

4.2 Distribution

Theorem 4.2 (*Skovgaard*)**.** Suppose Y is a continuous, scalar random variable. The saddlepoint approximation to $F(y|x) = \mathbb{P}(Y \leq y|X = x)$ is:

$$\hat{F}(y|x) = \Phi(\hat{w}) + \phi(\hat{w})(1/\hat{w} - 1/\hat{u}),$$

where:

$$\hat{w} = \text{sign}(\hat{t}) \sqrt{2\{K(\tilde{s}, 0) - \tilde{s}x\} - 2\{K(\hat{s}, \hat{t}) - \hat{s}x - \hat{t}y\}},$$
$$\hat{u} = \hat{t} \left(\frac{\det\{\ddot{K}_{ss}(\tilde{s}, 0)\}}{\det\{\ddot{K}(\hat{s}, \hat{t})\}} \right)^{-1/2},$$

$(\hat{s}, \hat{t})$ are the unconstrained saddlepoints, satisfying $\dot{K}(\hat{s}, \hat{t}) = (x, y)$, $\tilde{s}$ is the constrained saddlepoint, satisfying $\dot{K}_s(\tilde{s}, 0) = x$, and $\hat{t} \neq 0$. □

Empirical

Example 5.7 (*Sum*)**.** Suppose $X_i \overset{\text{IID}}{\sim} f$, where f is unknown. Consider approximating the distribution of $T = \sum_{i=1}^{n} X_i$. The individual-level MGF is $M_i(t) = \mathbb{E}(e^{tX_i})$, which can be estimated by

$$\hat{M}_i(t) = \frac{1}{n} \sum_{i=1}^{n} e^{tx_i}.$$

The empirical, individual-level CGF is:

$$\hat{K}_i(t) = \ln \hat{M}_i(t) = \ln \left\{ \frac{1}{n} \sum_{i=1}^{n} e^{tx_i} \right\}.$$

Now, the MGF of the sum $T = \sum_{i=1}^{n} X_i$ is:

$$M(t) = \mathbb{E}(e^{tT}) = \mathbb{E}(e^{t\sum_{i=1}^{n} X_i}) = \prod_{i=1}^{n} \mathbb{E}(e^{tX_i}) = \prod_{i=1}^{n} M_i(t),$$

and the CGF is:

$$K(t) = \ln M(t) = \sum_{i=1}^{n} \ln M_i(t) = \sum_{i=1}^{n} K_i(t).$$

Thus, the empirical CGF of the sum is:

$$\hat{K}_n(t) = \sum_{i=1}^{n} \hat{K}_i(t) = \sum_{i=1}^{n} \ln\left\{\frac{1}{n}\sum_{i=1}^{n} e^{tx_i}\right\} = n\ln\left\{\frac{1}{n}\sum_{i=1}^{n} e^{tx_i}\right\}.$$

The first derivative is:

$$\dot{\hat{K}}_n(t) = n\frac{\sum_{j=1}^{n} x_j e^{tx_j}}{\sum_{j=1}^{n} e^{tx_j}}$$

The second derivative is:

$$\ddot{\hat{K}}_n(t) = n\frac{(\sum_{j=1}^{n} x_j^2 e^{tx_j})(\sum_{j=1}^{n} e^{tx_j}) - (\sum_{j=1}^{n} x_j e^{tx_j})^2}{(\sum_{j=1}^{n} e^{tx_j})^2}$$

Finally, the saddlepoint equation $\dot{\hat{K}}_n(\hat{t}) \overset{\text{Set}}{=} x$ may be solved empirically for $\hat{t}(x)$. ♠

References

[1] RW Butler. *Saddlepoint Approximation with Applications.* Cambridge University Press, 2007. https://doi.org/10.1017/CBO9780511619083.

[2] C Goutis and G Casella. Explaining the saddlepoint approximation. *The American Statistician*, 53(3):216 – 224, 1999. https://www.jstor.org/stable/2686100.

[3] S Huzurbazar. Practical saddlepoint approximations. *The American Statistician*, 53(3):225 – 232, 1999. https://www.jstor.org/stable/i326505.

[4] N Reid. Saddlepoint methods and statistical inference. *Statistical Science*, 3(2):213 – 227, 1988. https://doi.org/10.1214/ss/1177012906.