# Some non-standard asymptotic distributions for MLEs

**EXAMPLE 1.** Suppose that $X_1, \cdots, X_n$ are i.i.d. random variables with density

$$f_\theta(x) = k(\tau)\{\tau - (x-\theta)^2\} \quad \text{for } -1 \le x - \theta \le 1$$

for some known $\tau \ge 1$ (where $k(\tau) = \{2(\tau - 1/3)\}^{-1}$) and unknown $\theta$. The asymptotic distribution of the MLE of $\theta$ will depend on $\tau$ with the type of distribution depending on whether or not $\tau > 1$ or $\tau = 1$. Up to an additive constant, log-likelihood function is

$$\ell_n(\theta) = \sum_{i=1}^n \ln\left\{\tau - (X_i - \theta)^2\right\} \quad \text{for } \max_{1\le i\le n} X_i - 1 \le \theta \le \min_{1\le i\le n} X_i + 1,$$

with $\ell_n(\theta) = -\infty$ otherwise.

We will first consider the case where $\tau > 1$. In this case, the MLE of $\theta$ turns out to depend asymptotically on the sample maximum and minimum; if $\theta_0$ is the true parameter value then

$$\begin{pmatrix} n\left(\min_{1\le i\le n} X_i + 1 - \theta_0\right) \\ n\left(\max_{1\le i\le n} X_i - 1 - \theta_0\right) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} V_1 \\ -V_2 \end{pmatrix}$$

where $V_1$ and $V_2$ are independent exponential random variables with mean $\{(\tau-1)k(\tau)\}^{-1}$. From this, we can deduce that the log-likelihood function $\ell_n(\theta)$ is finite (that is, greater than $-\infty$) for values of $\theta$ in a neighbourhood of $\theta_0$ whose length is $O_p(1/n)$; therefore if $\widehat{\theta}_n$ is the MLE then $n(\widehat{\theta}_n - \theta_0) = O_p(1)$. The limiting distribution can be determined by looking at the asymptotic behaviour of the function

$$Z_n(u) = \sqrt{n}\left\{\ell_n(\theta_0 + u/n) - \ell_n(\theta_0)\right\};$$

note that $Z_n$ is maximized at $n(\widehat{\theta}_n - \theta_0)$ and that $Z_n(u) > -\infty$ for

$$n\left(\max_{1\le i\le n} X_i - 1 - \theta_0\right) \le u \le n\left(\min_{1\le i\le n} X_i + 1 - \theta_0\right).$$

If we expand $Z_n(u)$ in a Taylor series around $u = 0$, we get

$$Z_n(u) = \frac{2u}{\sqrt{n}} \sum_{i=1}^n \frac{(X_i - \theta_0)}{\tau - (X_i - \theta_0)^2} + o_p(1)$$

where the $o_p(1)$ remainder term is uniform over $u$ in compact sets. Applying the central limit theorem, we get

$$Z_n(u) \xrightarrow{d} u\,W$$

where $W$ is a normal random variable with mean 0 and variance dependent on $\tau$; $W$ is independent of both $V_1$ and $V_2$. (The variance of $W$ can be evaluated explicitly as a function of $\tau$ although its value does not affect the asymptotic distribution of $n(\widehat{\theta}_n - \theta_0)$.) This suggests that $n(\widehat{\theta}_n - \theta_0) \xrightarrow{d} U$ where $U$ maximizes the function $uW$ subject to $-V_2 \leq u \leq V_1$; in other words,

$$U = \begin{cases} V_1 & \text{if } W > 0, \\ -V_2 & \text{if } W < 0. \end{cases}$$

The convergence in distribution of $n(\widehat{\theta}_n - \theta_0)$ to $U$ can be proved rigorously by noting that $Z_n$ is a concave function whose finite dimensional limit is finite on a open set; see Geyer (1996). Thus $U$ has a Laplace distribution with density

$$g(u) = \frac{(\tau-1)k(\tau)}{2} \exp\left\{-(\tau-1)k(\tau)|u|\right\}.$$

Figure 1 shows the log-likelihood function based on 100 observations from the density with $\tau = 4$. As predicted by the asymptotics, this log-likelihood is close to linear and maximized at the endpoint so that in this case $\widehat{\theta} = \min_i X_i + 1$.

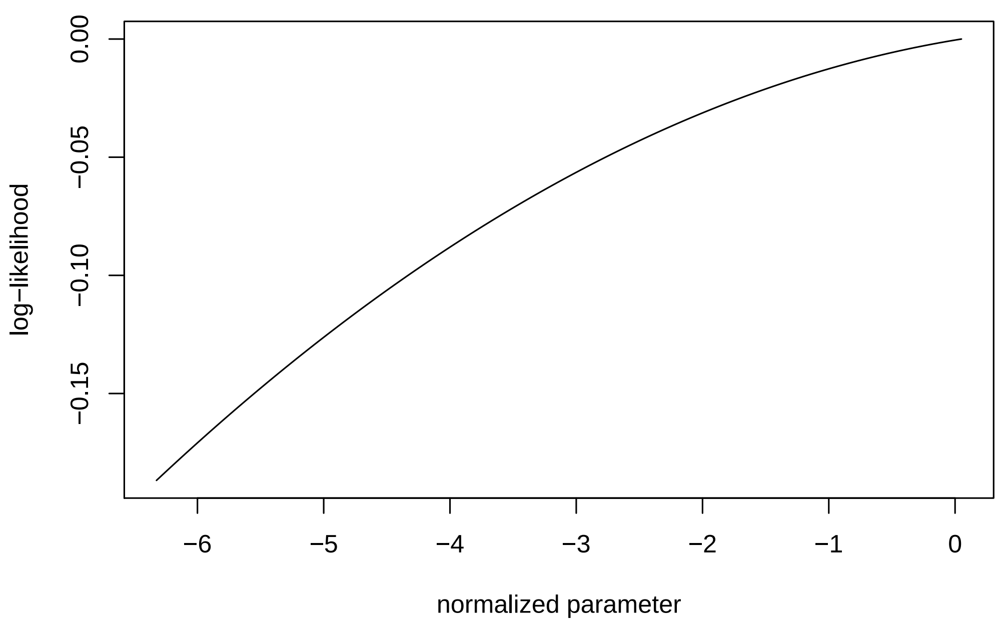


Figure 1: Log-likelihood function (normalized to have maximum 0) of the normalized parameter $n(\theta - \theta_0)$ based on 100 observations generated with $\tau = 4$ and $\theta_0 = 0$.

The case where $\tau = 1$ is somewhat more interesting. There are a number of differences between this case and the case $\tau > 1$. First of all, the convergence of the sample extremes to $\theta_0 \pm 1$ is no longer $O_p(n^{-1})$ but rather $O_p(n^{-1/2})$; when $\tau > 1$, $f_\theta(\theta \pm 1) > 0$, which leads to the $O_p(n^{-1})$ convergence while for $\tau = 1$, $f_\theta(\theta \pm 1) = 0$ with derivatives $f'_\theta(\theta \pm 1) \neq 0$, which implies the $O_p(n^{-1/2})$ convergence. More importantly, the behaviour of the log-likelihood function in a neighbourhood of $\theta_0$ is different in a subtle way.

Redefine $Z_n(u) = b_n \{\ell_n(\theta_0 + u/a_n) - \ell_n(\theta_0)\}$ where the sequences of constants $\{a_n\}$ and $\{b_n\}$ will be determined later. Expanding $Z_n(u)$ in a Taylor series around $u = 0$, we get

$$Z_n(u) = \frac{2b_n u}{a_n} \sum_{i=1}^{n} \frac{(X_i - \theta_0)}{1 - (X_i - \theta_0)^2} - \frac{b_n u^2}{a_n^2} \sum_{i=1}^{n} \frac{1 + (X_i - \theta_0)^2}{\{1 - (X_i - \theta_0)^2\}^2} + \cdots$$

What complicates the asymptotics of $Z_n$ is the fact that

$$\begin{aligned} \text{Var}\left\{\frac{(X_i - \theta_0)}{1 - (X_i - \theta_0)^2}\right\} &= \infty \\ \text{and} \;\; E\left\{\frac{1 + (X_i - \theta_0)^2}{\{1 - (X_i - \theta_0)^2\}^2}\right\} &= \infty, \end{aligned}$$

which means that we cannot take $a_n = \sqrt{n}$ with $b_n = 1$ and apply the central limit theorem and weak law of large numbers to the first two terms of the Taylor series expansion — at least in the standard way. However, it turns out that the central limit theorem and weak law of large numbers still hold with different normalizing constants. Defining $h_1(x) = (x - \theta_0)/\{1 - (x - \theta_0)^2\}$ and $h_2(x) = \{1 + (x - \theta_0)^2\}/\{1 - (x - \theta_0)^2\}^2$, we obtain for large values of $t$,

$$\begin{aligned} E\left\{h_1^2(X_1) I[|h_1(X_1)| \le t]\right\} &\sim \frac{3}{4}\ln(t) \\ E\left\{h_2(X_1) I[h_2(X_1) \le t]\right\} &\sim \frac{3}{4}\ln(t) \end{aligned}$$

in the sense that the ratio of the left hand to right hand sides tend to 1 as $t \to \infty$; the fact that both these truncated moments tend to infinity with $t$ very slowly (slower than $t^\epsilon$ for any $\epsilon > 0$) implies that we can still apply the central limit theorem and weak law of large numbers albeit with non-standard normalizing constants. In particular, if $c_n$ and $d_n$ satisfy

$$\frac{3}{4} n c_n^{-2} \ln(c_n) \to 1 \;\; \text{and} \;\; \frac{3}{4} n d_n^{-1} \ln(d_n) \to 1$$

then

$$\begin{aligned} \frac{1}{c_n} \sum_{i=1}^{n} \frac{(X_i - \theta_0)}{1 - (X_i - \theta_0)^2} &\xrightarrow{d} W \sim \mathcal{N}(0,1) \\ \frac{1}{d_n} \sum_{i=1}^{n} \frac{1 + (X_i - \theta_0)^2}{\{1 - (X_i - \theta_0)^2\}^2} &\xrightarrow{p} 1. \end{aligned}$$

(A good reference for this is Feller (1971), sections VII.7 and IX.8.) It follows that we can take $c_n = \sqrt{3n\ln(n)/8}$ and $d_n = 3n\ln(n)/4$. Thus setting $b_n = 1$ and $a_n = \sqrt{n\ln(n)}$, we get

$$Z_n(u) \xrightarrow{d} Z(u) = \left(\frac{3}{2}\right)^{1/2} u\, W - \frac{3}{4} u^2$$

where $W \sim \mathcal{N}(0,1)$ and finite dimensional convergence ($Z_n \stackrel{f-d}{\longrightarrow} Z$)

$$(Z_n(u_1), \cdots, Z_n(u_k)) \stackrel{d}{\longrightarrow} (Z(u_1), \cdots, Z(u_k))$$

also holds. Note that as $n \to \infty$, the set

$$\left\{u : a_n\left(\max_{1\le i\le n} X_i - 1 - \theta_0\right) \le u \le a_n\left(\min_{1\le i\le n} X_i + 1 - \theta_0\right)\right\}$$

increases to the entire real line since $a_n/\sqrt{n} \to \infty$. Thus (again using the concavity of $Z_n$), it follows that $a_n(\widehat{\theta}_n - \theta_0)$ converges in distribution to the maximizer of $Z$ and so

$$\sqrt{n\ln(n)}(\widehat{\theta}_n - \theta_0) \stackrel{d}{\longrightarrow} \left(\frac{2}{3}\right)^{1/2} W \sim \mathcal{N}(0, 2/3).$$

Figure 2 shows the log-likelihood function based on 100 observations from the density with $\tau = 1$. In this case, the asymptotic quadratic approximation to the log-likelihood function seems quite reasonable.

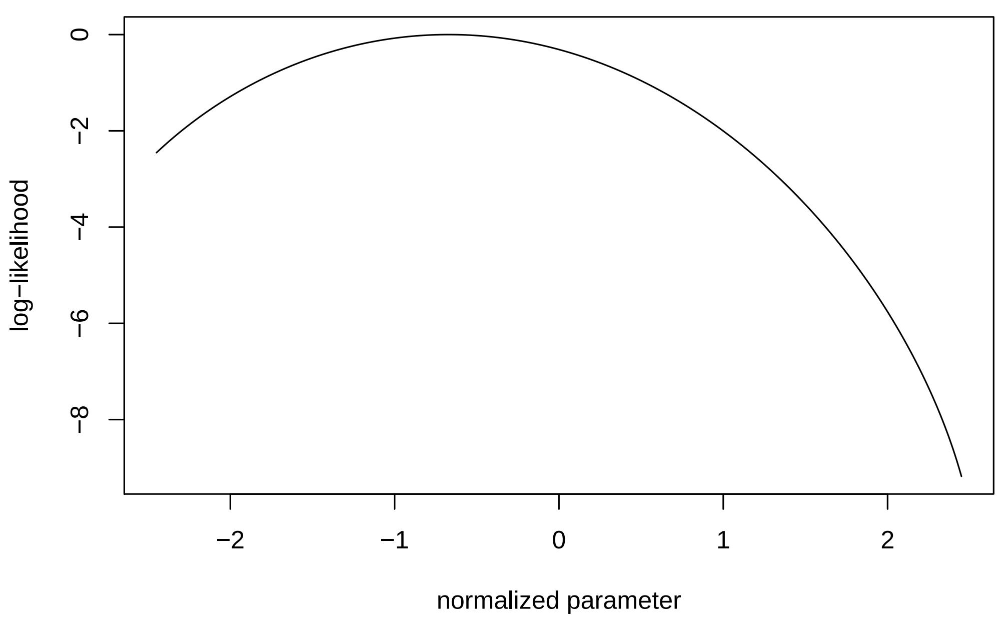


Figure 2: Log-likelihood function (normalized to have maximum 0) of the normalized parameter $\sqrt{n\ln(n)}(\theta - \theta_0)$ based on 100 observations generated with $\tau = 1$ and $\theta_0 = 0$.

**EXAMPLE 2.** Suppose that $X_1, \cdots, X_n$ are i.i.d. with density $f_\theta(x) = f(x - \theta)$ where

$$\lim_{t\downarrow 0} f(t) = \lambda^+ > 0 \quad \text{and} \lim_{t\uparrow 0} f(t) = \lambda^- > 0$$

where $\lambda^+ \neq \lambda^-$; in addition, assume that $f(t)$ is continuous and differentiable elsewhere. The asymptotics of the log-likelihood function are determined by those of a point process defined in terms of the observations close to the true parameter value $\theta_0$. Define the random measure

$$M_n(A) = \sum_{i=1}^{n} I[n(X_i - \theta_0) \in A].$$

For a given set $A$, it is straightforward to show that $M_n(A)$ converges in distribution to a Poisson random variable. This "pointwise" convergence can be extended to $\{M_n\}$ as a sequence of measures; it is straightforward to show that $M_n$ converges in distribution (with respect to the so-called vague topology) to a Poisson random measure $M$ ($M_n \xrightarrow{v-d} M$) where the mean measure of $M$ is

$$E[M(A)] = \lambda^-|A \cap (-\infty, 0)| + \lambda^+|A \cap (0, \infty)|$$

with $|B|$ is the Lebesgue measure of the set $B$. (More precisely, $M_n \xrightarrow{v-d} M$ if

$$\int h(x)\, M_n(dx) \xrightarrow{d} \int h(x)\, M(dx)$$

for all bounded continuous functions $h$ with compact support.)

Define the log-likelihood process

$$\ell_n(u) = \sum_{i=1}^{n} \ln \left\{ f(X_i - \theta_0 - u/n)/f(X_i - \theta_0) \right\}.$$

The limit of $\ell_n(u)$ is a sum of two parts, one a linear term, the other a point process term. The linear term arises from the fact that if $x - \theta_0$ is bounded away from 0 then for large $n$,

$$\ln \left\{ f(x - \theta_0 - u/n)/f(x - \theta_0) \right\} \approx -\frac{u}{n} \frac{f'(x - \theta_0)}{f(x - \theta_0)}.$$

The point process term arises from the fact that if $x_n - \theta_0$ close to 0 (so that $n(x_n - \theta_0)$ stays bounded), we have for large $n$,

$$\begin{aligned} \ln \left\{ f(x_n - \theta_0 - u/n)/f(x_n - \theta_0) \right\} &\approx \begin{cases} 0 & \text{if } n(x_n - \theta_0) > 0,\ u < n(x_n - \theta_0); \\ \ln(\lambda^-/\lambda^+) & \text{if } n(x_n - \theta_0) > 0,\ u > n(x_n - \theta_0); \\ 0 & \text{if } n(x_n - \theta_0) < 0,\ u > n(x_n - \theta_0); \\ \ln(\lambda^+/\lambda^-) & \text{if } n(x_n - \theta_0) < 0,\ u < n(x_n - \theta_0). \end{cases} \\ &= \tau(n(x_n - \theta_0), u). \end{aligned}$$

From this, it follows (see, for example, Chernozhukov and Hong (2004) who consider an extension of this model to regression) that

$$\begin{aligned} \ell_n(u) &= -u\, E\{f'(X_1 - \theta_0)/f(X_1 - \theta_0)\} + \int \tau(x, u)\, M_n(dx) + o_p(1) \\ &\xrightarrow{f-d} (\lambda^+ - \lambda^-)u + \int \tau(x, u)\, M(dx) \\ &= \ell(u) \end{aligned}$$

since

$$
\begin{aligned}
E\{f'(X_1-\theta_0)/f(X_1-\theta_0)\} &= \int_{-\infty}^{\infty} f'(x)\,dx \\
&= \int_{-\infty}^{0} f'(x)\,dx + \int_{0}^{\infty} f'(x)\,dx \\
&= \lambda^- - \lambda^+.
\end{aligned}
$$

The asymptotic behaviour of the log-likelihood function suggests that we can find estimators $\widehat{\theta}_n$ such that $n(\widehat{\theta}_n - \theta_0) = O_p(1)$. Intuitively, this is a consequence of the discontinuity in the density function; assuming, for example, that $\lambda^+ > \lambda^-$, in a given neighbourhood of $\theta_0$, there are more observations in this neighbourhood greater than $\theta_0$ than there are less than $\theta_0$. This difference in intensity provides a lot of information about the value of the parameter. Because of the discontinuities in the likelihood function, the maximum likelihood estimator is not particularly useful nor easily computed.

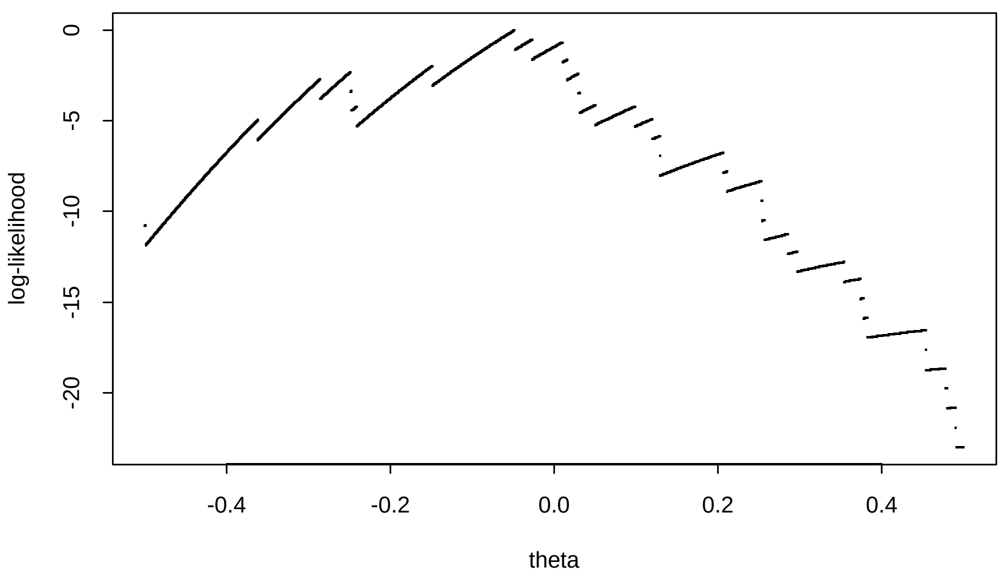


Figure 3: Log-likelihood function for $\theta$ (normalized to have maximum 0) based on 100 observations generated according to the normal density jump model.

However, we can still use the likelihood function to estimate $\theta$ using Bayes estimators such as the posterior mean or posterior median. Given a prior density $\pi(\theta)$, the posterior density of the rescaled parameter $n(\theta - \theta_0)$ is proportional to $\pi(\theta_0 + u/n)\exp\{\ell_n(u)\}$. Under fairly mild conditions on the prior density, the limit of the posterior density (of the transformed

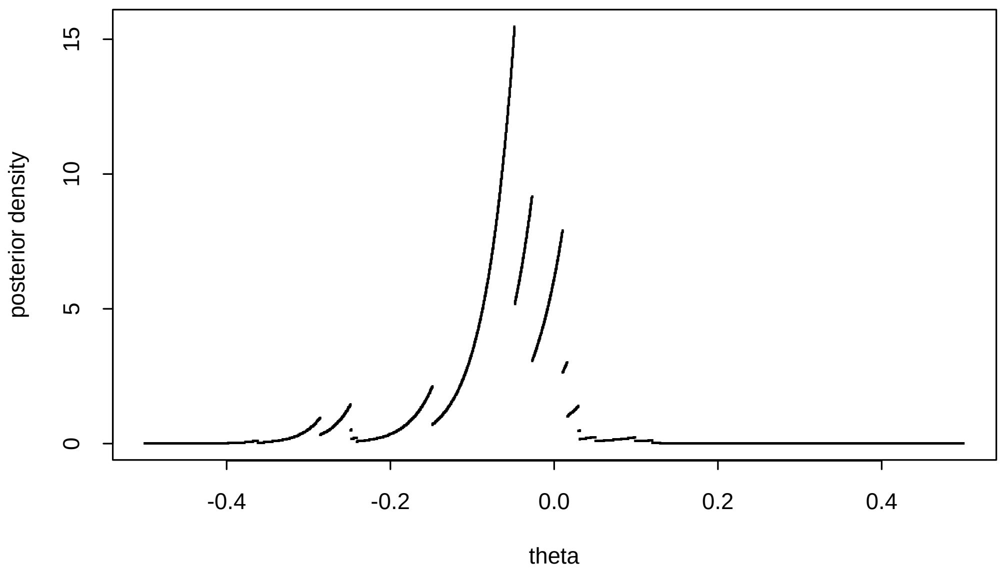


Figure 4: Posterior density (with standard Cauchy prior) for $\theta$ based on 100 observations generated according to the normal density jump model.

parameter $n(\theta - \theta_0))$ is given by the normalized asymptotic likelihood:

$$\Lambda(u) = \frac{\exp\{\ell(u)\}}{\int_{-\infty}^{\infty} \exp\{\ell(s)\}\, ds}.$$

$\Lambda$ can be used to determine the limiting distribution of Bayes estimators; note that $\Lambda$ is a random density function since it depends on the Poisson process $M$. For example, for the posterior mean, we have

$$n(\widehat{\theta}_n - \theta_0) \xrightarrow{d} \int_{-\infty}^{\infty} u\, \Lambda(u)\, du.$$

As an illustration, suppose that

$$f(t) = \begin{cases} \phi(t) & \text{for } t \geq 0 \\ \phi(t/3)/3 & \text{for } t < 0 \end{cases}$$

where $\phi(t)$ is the standard normal density function. Figure 3 shows the log-likelihood function (normalized so that its maximum is 0) based on 100 observations with $\theta_0 = 0$; the MLE is approximately $-0.048$. Figure 4 shows the posterior density for $\theta$ when the prior density is a standard Cauchy density; the posterior mean is $-0.068$ while the posterior median is $-0.056$.

# REFERENCES

Chernozhukov, V. and Hong, H. (2004) Likelihood inference in a class of nonregular econometric models. *Econometrica.* **72**, 1445-1480.

Feller, W. (1971) *An Introduction to Probability Theory and its Applications, Volume II (2nd edition).* New York: Wiley.

Geyer, C.J. (1996) On the asymptotics of convex stochastic optimization. Unpublished manuscript.