

2015

## RATE OF CONVERGENCE OF MAXIMUM LIKELIHOOD ESTIMATORS UNDER RELAXED SMOOTHNESS CONDITIONS ON THE LIKELIHOOD FUNCTION

Brent Halonen  
*Michigan Technological University*

Follow this and additional works at: <https://digitalcommons.mtu.edu/etds>

Copyright 2015 Brent Halonen

---

### Recommended Citation

Halonen, Brent, "RATE OF CONVERGENCE OF MAXIMUM LIKELIHOOD ESTIMATORS UNDER RELAXED SMOOTHNESS CONDITIONS ON THE LIKELIHOOD FUNCTION", Master's report, Michigan Technological University, 2015.

<https://digitalcommons.mtu.edu/etds/909>

Follow this and additional works at: <https://digitalcommons.mtu.edu/etds>

RATE OF CONVERGENCE OF MAXIMUM  
LIKELIHOOD ESTIMATORS UNDER  
RELAXED SMOOTHNESS CONDITIONS  
ON THE LIKELIHOOD FUNCTION

By  
Brent Halonen

A REPORT  
Submitted in partial fulfillment of the requirements for the degree  
MASTER OF SCIENCE  
In Mathematical Sciences

MICHIGAN TECHNOLOGICAL UNIVERSITY  
2015

© 2015 Brent Halonen

This report has been approved in partial fulfillment of the requirements for the Degree of MASTER OF SCIENCE in Mathematical Sciences.

Department of Mathematical Sciences

Report Advisor:*Iosif Pinelis*

Committee Member:*Dean Johnson*

Committee Member:*Qiuying Sha*

Department Chair:*Mark Gockenbach*

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Literature Review</b>	<b>3</b>
2.1	Fisher: initial results . . . . .	3
2.2	Doob and Cramér: derivative conditions . . . . .	4
2.3	LeCam: Differentiability in Quadratic Mean . . . . .	8
2.4	Concluding Remarks on the Literature . . . . .	16
<b>3</b>	<b>Result</b>	<b>16</b>
3.1	Smoothness Conditions . . . . .	16
3.2	Theorem and Proof . . . . .	18
3.3	Discussion . . . . .	20
3.4	Simulations of the MLE for the Generalized Continuous Laplace distribution . . . . .	24
3.5	Appendix . . . . .	33
<b>4</b>	<b>Bibliography</b>	<b>34</b>

## Abstract

This report reviews literature on the rate of convergence of maximum likelihood estimators and establishes a Central Limit Theorem, which yields an  $O(\frac{1}{\sqrt{n}})$  rate of convergence of the maximum likelihood estimator under somewhat relaxed smoothness conditions. These conditions include the existence of a one-sided derivative in  $\theta$  of the pdf, compared to up to three that are classically required. A verification through simulation is included in the end of the report.

# 1 Introduction

We shall show that under somewhat relaxed smoothness conditions, a rate of convergence of  $\frac{1}{\sqrt{n}}$  may be obtained if  $f_\theta(x)$  is log-concave in  $\theta$  i.e., if  $\ln f_\theta(x)$  is concave in  $\theta$ . The usefulness of these results stems from many common families of pdf's being log-concave, and because a pdf proportional to the product of two log-concave pdf's is also log-concave. This is particularly important in Bayesian analysis, as a log-concave prior and a log-concave likelihood function implies a log-concave posterior. Examples of log-concave functions and a demonstration that log-concave functions are closed under multiplication will be shown in the remark in Section 3.3. In addition, an example of the result will be shown for a generalized Laplace distribution, and a Mathematica script will be provided demonstrating a simulation to verify the result.

Since the discovery of the maximum likelihood estimator (MLE) by R.A. Fisher in 1922, there have been many attempts to find a minimal set of conditions for a  $O(\frac{1}{\sqrt{n}})$  rate of convergence of the MLE. This is to establish the rate of convergence for as a broad of class of distributions as possible. One practical result of broadening of conditions is that it allows the characterization of the limit distribution of MLEs found through numerical methods.

An example of the convergence of the MLE that satisfies the log-concave condition can be found in the normal distribution with constant variance. The log-likelihood of the normal distribution is  $\ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \frac{(x-\theta)^2}{\sigma^2}$ , which is concave in  $\theta$ . The mean of the samples converges to the expected value,  $\theta$  at a rate of  $O(\frac{1}{\sqrt{n}})$  by the central limit theorem. A counter-example to the  $O(\frac{1}{\sqrt{n}})$  convergence can be found in the uniform distribution,  $U(0, \theta)$ , where the estimator is  $\frac{n+1}{n} \max(x_i)$  converges to  $\theta$  at a rate of  $O(\frac{1}{n})$ . The pdf violates log-concavity at the maximum of the sample, as there is no possibility of  $\theta$  being less than the  $\max(x_i)$ . This discontinuity drives the much faster rate of convergence. An example that demonstrates the lack of need for a derivative in  $\theta$  to be necessary is the Laplace distribution, with pdf  $ce^{-|\frac{x-\theta}{b}|}$ . The MLE is the median, which has a  $O(\frac{1}{\sqrt{n}})$  according to S. Kotz and Podgoński [2002]. The log-likelihood function of the Laplace distribution is  $\ln c - |\frac{x-\theta}{b}|$ , which has no derivative in  $\theta$  at  $\theta = x$ , however is concave in  $\theta$  and fits the conditions outlined in the following pages.

The log concavity in  $\theta$  of  $f_\theta(x)$  condition drives convergence by requiring that the likelihood of  $\theta$  decreases at a rate proportional to  $f_\theta(x)$ . This steep-

ness in the likelihood function enforces that for  $\theta_0$  and  $\theta_0 + k$ , the difference in the expected log likelihood function  $\ell(\theta_0) - \ell(\theta_0 + k)$  will grow at least linearly with  $k$ . Since for each  $\theta$ , the log-likelihood function converges pointwise to the expected log-likelihood function, and the value of the maximum likelihood converges to the expected maximum likelihood,  $\ell(\hat{\theta})$  will be close to  $\ell(\theta_0)$ . Then, the steepness from the requirement of log-concavity enforces  $\hat{\theta} - \theta_0$  to be  $O(\frac{1}{\sqrt{n}})$ , as larger deviations will cause  $\ell(\hat{\theta})$  to differ too much from  $\ell(\theta)$ .

## 2 Literature Review

### 2.1 Fisher: initial results

Fisher [1922] was the first to propose a method to find an asymptotic distribution of the MLE. He required that the MLE  $\hat{\theta}$  be asymptotically normally distributed near the  $\theta$ , such that  $\hat{\theta} \xrightarrow{D} N(\theta, \sigma^2)$ , where  $\sigma^2$  is the asymptotic variance of  $\hat{\theta}$ . This includes several important distributions including the normal distribution and the  $\chi^2$  distribution. Fisher also required that the first and second derivatives in  $\theta$  of the pdf exist. Using these conditions, Fisher showed that the asymptotic variance  $\sigma^2$  of  $\hat{\theta}$ , can be found as follows,

$$-\frac{1}{\sigma^2} = nE \frac{\partial^2}{\partial \theta^2} \ln f(\hat{\theta}).$$

This inverse value of the variance of the maximum likelihood estimator is now called Fisher's information, denoted as  $I(\theta_0)$ . Fisher's result was restrictive, especially the requirement of asymptotic normality of the estimator. He was aware of MLEs that was not asymptotically normal, such as the MLE of the parameter in the uniform distribution  $U(0, \theta), \theta > 0$ , where the asymptotic distribution of the MLE is exponential. Even if the condition was met, a good deal of work might be needed to show that an estimator was asymptotically normal.

Fisher [1925] also showed that under the previously mentioned conditions, the MLE was the most efficient, in the sense that it had the lowest asymptotic mean square error (MSE) of all asymptotically normal estimators. Fisher showed that for any asymptotically normal estimator  $T$ , its

asymptotic variance  $\sigma_T^2$  can be expressed as follows,

$$\frac{1}{n\sigma_T^2} = -E \frac{\partial^2}{\partial \theta^2} \ln f(\hat{\theta}) - n \left( \frac{\partial^2}{\partial \theta^2} \ln f(\hat{\theta}) \right)^2 V'(\hat{\theta}|T),$$

where  $V'(\hat{\theta}|T)$  is the variance of  $\hat{\theta}$  given the estimator  $T$ . The notation  $V'(\hat{\theta}|T)$  was chosen to be consistent with Fisher's notation. Since  $\sigma^2$  is a strictly increasing function of  $V'(\hat{\theta}|T)$  and  $V'(\hat{\theta}|T) \geq 0$ , all that is needed to minimize  $\sigma^2$  is to find the value of  $T$  such that  $V'(\hat{\theta}|T) = 0$ . This can be easily found since  $V'(\hat{\theta}|\hat{\theta}) = 0$ , which implies that  $\hat{\theta}$  has the lowest asymptotic variance of all the asymptotically normal estimators. Since  $\hat{\theta}$  is unbiased, the asymptotic MSE is also minimized. While this result did not further demonstrate the rate of convergence of the MLE, it eliminated efforts to find more efficient asymptotically normal statistics.

Hotelling [1930] attempted to prove the asymptotic normality of the MLE, with the following requirements: (i) the continuity of the pdf  $f_\theta(x)$  in  $x$  almost everywhere, (ii) the existence of the first derivative of  $f_\theta(x)$  in  $\theta$ , and (iii) that  $x^2 \frac{\partial f}{\partial \theta}$  approaches a smooth function of  $\theta$  when  $x \rightarrow \pm\infty$ . However, according to Stigler [2008], Hotelling erroneously simplified the problem by applying the arctangent transformation to the sample space and then discretizing the observed variables into a finite number of intervals, which causes issues with uniformity in convergence.

## 2.2 Doob and Cramér: derivative conditions

Doob [1934] in 1934 found a proof of the convergence of the MLE without the requirement of the the normality of the MLE and a demonstration that  $E_{\theta_0} \frac{\partial \ln f_\theta(x)}{\partial \theta} = -E_{\theta_0} \frac{\partial^2 \ln f_\theta(x)}{\partial \theta^2}$ . Doob's requirements can be summarized as follows:

1. In an  $a_1$  neighborhood of  $\theta_0$ , let,

$$\ln f_\theta(x) = \ln f_{\theta_0}(x) + (\theta - \theta_0)\alpha(x) + \frac{(\theta - \theta_0)^2}{2}\beta(x) + \gamma_\theta(x)$$

where  $E_{\theta_0}\alpha(x)$ ,  $E_{\theta_0}\alpha(x)^2$ , and  $E_{\theta_0}\beta(x)$  all exist. Of course,  $\alpha(x) = \frac{\partial \ln f_\theta(x)}{\partial \theta}$  and  $\beta(x) = \frac{\partial^2 \ln f_\theta(x)}{\partial \theta^2}$ .

2. Let  $\gamma_\theta(x)$  be differentiable in an  $a_2 < a_1$  neighborhood of  $\theta_0$ . Suppose  $\phi(x) = \sup_{|\theta - \theta_0| < a_2} \left( \frac{1}{(\theta - \theta_0)^2} \left| \frac{\partial}{\partial \theta} \gamma_\theta(x_i) \right| \right)$  and  $E_{\theta_0} \phi(x)$  exists.
3. Define  $\delta(x)$  such that

$$f_\theta(x) = f_{\theta_0}(x) \left( 1 + (\theta - \theta_0) \alpha_{\theta_0}(x) + \frac{(\theta - \theta_0)^2}{2} (\beta_{\theta_0}(x) + \alpha_{\theta_0}(x)^2) + \delta_\theta(x) \right),$$

with  $\delta_\theta(x) = o((\theta - \theta_0)^2)$ . Note that this is equivalent to stating that  $f_\theta(x)$  is twice differentiable in  $\theta$ .

4. Suppose that  $\sigma^2 := E_{\theta_0} \alpha(x)^2 > 0$ .

Doob first demonstrated that  $E_{\theta_0} \left( \frac{\partial \ln f_\theta(x)}{\partial \theta} \right)^2 = -E_{\theta_0} \frac{\partial^2 \ln f_\theta(x)}{\partial \theta^2}$  as follows. By the definition of the pdf and by Doob's requirement 3,

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f_\theta(x) dx \\ &= \int_{-\infty}^{\infty} f_{\theta_0}(x) dx + (\theta - \theta_0) E_{\theta_0} \alpha_{\theta_0}(x) + \frac{(\theta - \theta_0)^2}{2} E_{\theta_0} (\alpha_{\theta_0}(x)^2 + \beta_{\theta_0}(x)) + E_{\theta_0} \delta_\theta(x) \\ 0 &= (\theta - \theta_0) E_{\theta_0} \alpha_{\theta_0}(x) + \frac{(\theta - \theta_0)^2}{2} E_{\theta_0} (\alpha_{\theta_0}(x)^2 + \beta_{\theta_0}(x)) + E_{\theta_0} \delta_\theta(x). \end{aligned}$$

Then, dividing through by  $(\theta - \theta_0)$  and considering Doob's requirement 3 that  $\delta_\theta(x) = o((\theta - \theta_0)^2)$ ,

$$E_{\theta_0} \alpha_{\theta_0}(x) = 0. \tag{1}$$

Likewise, dividing through by  $(\theta - \theta_0)^2$ , it is seen that

$$E_{\theta_0} (\alpha_{\theta_0}(x)^2 + \beta_{\theta_0}(x)) = 0. \tag{2}$$

which, substituting from Doob's requirement 1, establishes that  $E_{\theta_0} \left( \frac{\partial \ln f_\theta(x)}{\partial \theta} \right)^2 = -E_{\theta_0} \frac{\partial^2 \ln f_\theta(x)}{\partial \theta^2}$ . Doob then proved the  $\sqrt{n}$  rate of convergence of  $\hat{\theta}$  under his requirements as follows. The log-likelihood function is defined by the formula

$$L_{\underline{x}}(\theta) := \sum_{i=1}^n \ln f_\theta(x_i),$$



where  $\mathbf{x} := (x_1, \dots, x_n)$ . Then using Doob's requirement 1,

$$\begin{aligned} L_{\mathbf{x}}(\theta) &= \sum_{i=1}^n \ln f_{\theta}(x_i) \\ &= \sum_{i=1}^n \ln f_{\theta_0}(x_i) + (\theta - \theta_0) \sum_{i=1}^n \alpha_{\theta_0}(x_i) + \frac{(\theta - \theta_0)^2}{2} \sum_{i=1}^n \beta_{\theta_0}(x_i) + \sum_{i=1}^n \gamma(x_i). \end{aligned}$$

Since  $L_{\mathbf{x}}(\theta)$  has a maximum at  $\hat{\theta}$ ,

$$\begin{aligned} 0 &= \frac{\partial}{\partial \theta} L_{\mathbf{x}}(\hat{\theta}) \\ &= \sum_{i=1}^n \alpha_{\theta_0}(x_i) + (\hat{\theta} - \theta_0) \sum_{i=1}^n \beta_{\theta_0}(x_i) + \sum_{i=1}^n \frac{\partial}{\partial \theta} \gamma_{\theta}(x_i). \end{aligned} \quad (3)$$

This expression can be further developed as follows,

$$\begin{aligned} (\hat{\theta} - \theta_0) \left( -\frac{1}{\sigma^2 n} \sum_{i=1}^n \beta_{\theta_0}(x_i) - \frac{1}{\sigma^2 n (\hat{\theta} - \theta_0)} \sum_{i=1}^n \frac{\partial}{\partial \theta} \gamma_{\theta}(x_i) \right) &= \frac{1}{\sigma^2 n} \sum_{i=1}^n \alpha_{\theta_0}(x_i) \\ \sqrt{n} \sigma (\hat{\theta} - \theta_0) &= \frac{\frac{1}{\sigma \sqrt{n}} \sum_{i=1}^n \alpha_{\theta_0}(x_i)}{-\frac{1}{\sigma^2 n} \sum_{i=1}^n \beta_{\theta_0}(x_i) - \frac{1}{\sigma^2 n (\hat{\theta} - \theta_0)} \sum_{i=1}^n \frac{\partial}{\partial \theta} \gamma_{\theta}(x_i)} \\ \sqrt{n} \sigma (\hat{\theta} - \theta_0) &= \frac{1}{\sigma \sqrt{n}} \sum_{i=1}^n \alpha_{\theta_0}(x_i) + R_n, \end{aligned}$$

where

$$R_n = \frac{\frac{1}{\sigma \sqrt{n}} \sum_{i=1}^n \alpha_{\theta_0}(x_i) \left( 1 + \frac{1}{\sigma^2 n} \sum_{i=1}^n \beta_{\theta_0}(x_i) + \frac{1}{\sigma^2 n (\hat{\theta} - \theta_0)} \sum_{i=1}^n \frac{\partial}{\partial \theta} \gamma_{\theta}(x_i) \right)}{-\frac{1}{\sigma^2 n} \sum_{i=1}^n \beta_{\theta_0}(x_i) - \frac{1}{\sigma^2 n (\hat{\theta} - \theta_0)} \sum_{i=1}^n \frac{\partial}{\partial \theta} \gamma_{\theta}(x_i)}. \quad (4)$$

From Doob's requirement 3,

$$\left| \frac{1}{\sigma^2 n (\hat{\theta} - \theta_0)} \sum_{i=1}^n \frac{\partial}{\partial \theta} \gamma_{\theta}(x_i) \right| < \frac{|\hat{\theta} - \theta_0|}{\sigma^2 n} \sum_{i=1}^n \phi_{\theta}(x_i) \xrightarrow[n \rightarrow \infty]{P} 0. \quad (5)$$

By Khintchine's law, equation (2) and Doob's requirement 4

$$\frac{1}{\sigma^2 n} \sum_{i=1}^n \beta_{\theta_0}(x_i) \xrightarrow[n \rightarrow \infty]{P} -1. \quad (6)$$

From Doob's requirement 4, (1), and the Central Limit Theorem,

$$\frac{1}{\sigma \sqrt{n}} \sum_{i=1}^n \alpha_{\theta_0}(x_i) \xrightarrow[n \rightarrow \infty]{D} N(0, 1). \quad (7)$$

Substituting (5), (6), and (7) into (4), it can be seen that  $R_n \xrightarrow[n \rightarrow \infty]{P} 0$ , which implies that  $\sqrt{n}\sigma(\theta - \theta_0) \xrightarrow[n \rightarrow \infty]{D} N(0, 1)$ .

Cramér [1978] found a proof of the convergence of the MLE which was very similar to Doob's proof, with almost identical conditions. Cramér's requirements are listed as follows:

1.  $\frac{\partial f_{\theta}(x)}{\partial \theta}$ ,  $\frac{\partial^2 f_{\theta}(x)}{\partial \theta^2}$ , and  $\frac{\partial^3 f_{\theta}(x)}{\partial \theta^3}$  exist for every  $\theta$  and for almost all  $x$ .
2. For all  $\theta$ ,  $|\frac{\partial f_{\theta}(x)}{\partial \theta}| < F_1(x)$ ,  $|\frac{\partial^2 f_{\theta}(x)}{\partial \theta^2}| < F_2(x)$  and  $|\frac{\partial^3 f_{\theta}(x)}{\partial \theta^3}| < H(x)$ , where  $F_1(x)$  and  $F_2(x)$  are integrable over the real line, and  $E_{\theta}H(x) < M$  for all  $\theta$  where  $M$  does not depend on  $\theta$ .
3. The expectation of  $\frac{\partial \ln f_{\theta}(X)}{\partial \theta}$  is finite.

Let the likelihood equation be denoted as  $L(\theta) = \prod_i f_{\theta}(x_i)$ . Cramér then showed that at the MLE,

$$\frac{1}{n} \frac{\partial \ln L(\hat{\theta})}{\partial \theta} = B_0 + B_1(\hat{\theta} - \theta_0) + \frac{1}{2} \lambda B_2(\hat{\theta} - \theta_0)^2 = 0, \quad (8)$$

where  $\lambda \in (-1, 1)$  depends on  $\theta$  and  $n$ , and where

$$B_0 = \frac{1}{n} \sum_{i=1}^n \frac{\partial \ln f_{\theta_0}(x_i)}{\partial \theta}, B_0 \xrightarrow[n \rightarrow \infty]{P} 0$$

$$B_1 = \frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \ln f_{\theta_0}(x_i)}{\partial \theta^2} \xrightarrow[n \rightarrow \infty]{P} -k^2 = I(\theta_0)$$

$$B_2 = \frac{1}{n} \sum_{i=1}^n H(x_i) \xrightarrow[n \rightarrow \infty]{P} E_{\theta}H(x) < M.$$

Cramér then showed that (8) could be rearranged so that

$$k\sqrt{n}(\widehat{\theta} - \theta_0) = \frac{\frac{1}{k\sqrt{n}} \sum_{i=1}^n \frac{\partial f_{\theta_0}(x_i)}{\partial \theta}}{-\frac{B_1}{k^2} - \lambda(\widehat{\theta} - \theta_0) \frac{B_2}{2k^2}} \xrightarrow[n \rightarrow \infty]{P} N(0, 1), \quad (9)$$

as the denominator converges to 1 and the numerator converges to  $N(0, 1)$ . More recently, Lehmann [2004] used a similar argument differing by the use of a  $\theta_n^* \in (\theta_0, \widehat{\theta}_n)$  instead of the  $\lambda$  used above to make the equality exact, with identical conditions to establish the same result.

Cramér's conditions are nearly identical to Doob's, this can be seen by restating Doob's derivative of the log likelihood function (3) using Doob's requirement 2. First, Doob's requirement 2 can be rearranged as follows, slightly altered by adding a constant of one half, and a factor of  $\lambda \in (-1, 1)$  to enforce equality

$$\frac{\partial}{\partial \theta} \gamma_{\theta}(x) = \frac{1}{2} \lambda \phi(x) (\theta - \theta_0)^2. \quad (10)$$

Then (3) can be restated using (10) and dividing through by  $n$  as

$$\begin{aligned} 0 &= \frac{1}{n} \frac{\partial}{\partial \theta} L_{\mathbf{x}}(\widehat{\theta}) \\ &= \frac{1}{n} \sum_{i=1}^n \alpha(x_i) + \frac{1}{n} \sum_{i=1}^n \beta(x_i) (\theta - \theta_0) + \frac{1}{2} \lambda \frac{1}{n} \sum_{i=1}^n \phi(x_i) (\theta - \theta_0)^2. \end{aligned} \quad (11)$$

Then assigning  $B_0, B_1$  and  $B_2$  from (8),  $\frac{1}{n} \sum_{i=1}^n \alpha(x_i) = B_0$ ,  $\frac{1}{n} \sum_{i=1}^n \beta(x_i) = B_1$ , and  $\frac{1}{n} \sum_{i=1}^n \phi(x_i)$  behaves similarly to  $B_2$  as it converges to a finite number, and serves as an upper bound on the error of the Taylor expansion. This is then identical in behavior to Cramér's equation for the root of the maximum likelihood (8). The difference in the conditions lies in the requirement of the third derivative by Cramér, while Doob only requires the bounds that the existence of the third derivative imply.

### 2.3 LeCam: Differentiability in Quadratic Mean

LeCam [1986] established the important concept of local asymptotic normality, meaning a sequence of statistical models, for example, the sequence of the distribution of a maximum likelihood estimator as more samples are added, the log likelihood ratio can be approximated by a normal distribution. The

formal definition of local area normality for a distribution with one parameter with pdf  $f_\theta(x)$  can be stated as follows, if

$$\theta_n = \theta_0 + O\left(\frac{1}{\sqrt{n}}\right), \quad (12)$$

then  $f_\theta(x)$  is locally asymptotically normal if

$$\ln \frac{\prod_{i=1}^n f_{\theta_n}(x_i)}{\prod_{i=1}^n f_{\theta_0}(x_i)} = (\theta_n - \theta_0)\sqrt{n}\sqrt{I(\theta_0)}Z - \frac{n}{2}(\theta_n - \theta_0)^2 I(\theta_0) + o_p(1) \quad (13)$$

where  $Z \sim N(0, 1)$ . LeCam [1986] proved that if the above approximation converges pointwise, then

$$\sqrt{I(\theta_0)}\sqrt{n}(\theta_n - \theta) \xrightarrow{D} Z. \quad (14)$$

A simple demonstration of this can be made if the log-likelihood function is assumed to be differentiable. Then,

$$\begin{aligned} \frac{\partial \ln \frac{\prod_{i=1}^n f_{\theta_n}(x_i)}{\prod_{i=1}^n f_{\theta_0}(x_i)}}{\partial \theta_n} &= \sqrt{n}\sqrt{I(\theta_0)}Z - n(\theta_n - \theta_0)I(\theta_0) + o_p(1) = 0 \\ \sqrt{I(\theta_0)}\sqrt{n}(\theta_n - \theta) + o_p(1) &= Z. \end{aligned}$$

LeCam [1986] used the concept of differentiability in quadratic mean (DQM) defined for a univariate pdf as follows. Let  $\xi_\theta(x) = \sqrt{f_\theta(x)}$ . If  $f_\theta(x)$  is differentiable in quadratic mean in one dimension, then,

$$\xi_\theta(x) = \xi_{\theta_0}(x) + (\theta - \theta_0)\Delta_{\theta_0}(x) + r_\theta(x) \quad (15)$$

where

$$\|r_\theta(x)\| := \sqrt{\int_{\mathbb{R}} r_\theta(x)^2 dx} = o(\theta - \theta_0) \quad (16)$$

as  $\theta \rightarrow \theta_0$ .

LeCam established that the DQM condition implies local asymptotic normality, which implies the rate of convergence is of the order  $\frac{1}{\sqrt{n}}$  if the condition of differentiability in quadratic mean was met.

David Pollard [1997] explained that the reason that DQM leads to  $O(\frac{1}{\sqrt{n}})$  convergence without the requirement of a second derivative is that the square root of a pdf is an element of  $\mathcal{L}^2$  space with a norm of 1. This causes  $\langle \xi_0, r_{\theta_n} \rangle := \int_{-\infty}^{\infty} \xi_0(x)r_{\theta_n}(x)dx = O(|\theta_n - \theta_0|^2)$  without the requirement of the second derivative as stated in the following Lemma.

**Lemma 1.** *If  $f_\theta(x)$  has the DQM property, as defined in (15), then  $\langle \xi_{\theta_0}, r_{\theta_n} \rangle = -\frac{1}{2}(\theta_n - \theta_0)^2 \frac{1}{4} I(\theta_0) + o((\theta_n - \theta_0)^2)$ .*

*Proof.* Consider a sequence of  $\theta_n$  such that  $\lim_{n \rightarrow \infty} \theta_n - \theta_0 = 0$ , and a resultant sequence of  $\xi_{\theta_n}(x)$  as defined above. Then, by the fixed norm property,

$$\begin{aligned} 0 = \|\xi_{\theta_n}\|^2 - \|\xi_{\theta_0}\|^2 &= 2(\theta_n - \theta_0)\langle \xi_{\theta_0}, \Delta_{\theta_0} \rangle \\ &\quad + 2\langle \xi_{\theta_0}, r_{\theta_n} \rangle \\ &\quad + (\theta_n - \theta_0)^2 \|\Delta_{\theta_0}\|^2 \\ &\quad + 2(\theta_n - \theta_0)\langle \Delta_{\theta_0}, r_{\theta_n} \rangle \\ &\quad + \|r_{\theta_n}\|^2. \end{aligned} \tag{17}$$

Pollard [2005] showed that the definition of the score function can be extended as  $2 \frac{\Delta_\theta(x)}{\xi_\theta(x)}$  under the regularity properties of DQM. Pollard [2005] showed that to be equivalent to the usual score function under pointwise differentiability, as follows,

$$2 \frac{\Delta_\theta(x)}{\xi_\theta(x)} = \frac{2}{\sqrt{f_\theta(x)}} \frac{\partial \sqrt{f_\theta(x)}}{\partial \theta} = \frac{1}{f_\theta(x)} \frac{\partial f_\theta(x)}{\partial \theta} = \frac{\partial \ln f_\theta(x)}{\partial \theta}. \tag{18}$$

Fisher's information under the above formulation of the score function is then

$$I(\theta) = \int_{-\infty}^{\infty} f_\theta(x) \left(2 \frac{\Delta_\theta(x)}{\xi_\theta(x)}\right)^2 dx = 4 \int_{-\infty}^{\infty} \Delta_\theta(x)^2 dx. \tag{19}$$

This gives the constant in the third summand of (17),

$$\|\Delta_{\theta_0}\| = \int_{-\infty}^{\infty} \Delta_{\theta_0}(x)^2 dx = \frac{1}{4} I(\theta_0). \tag{20}$$

The order the second element of (17),  $2\langle \xi_{\theta_0}, r_{\theta_n} \rangle$  can also be found using the Cauchy-Schwarz inequality and the definition of the DQM (16) and (20) as follows,

$$\langle \xi_{\theta_0}, r_{\theta_n} \rangle \leq \|\xi_{\theta_0}\| \|r_{\theta_n}\| = o(\theta_n - \theta_0). \tag{21}$$

The order of the third element of (17),  $(\theta_n - \theta_0)^2 \|\Delta_{\theta_0}\|^2$  is  $O((\theta_n - \theta_0)^2)$  from (20). The order the fourth element of (17),  $2(\theta_n - \theta_0)\langle \Delta_{\theta_0}, r_{\theta_n} \rangle$  can also be found using the Cauchy-Schwarz inequality and the definition of the DQM (16) and (20) as follows,

$$\langle \Delta_{\theta_0}, r_{\theta_n} \rangle \leq \|\Delta_{\theta_0}\| \|r_{\theta_n}\| = o(\theta_n - \theta_0). \tag{22}$$

The definition of the DQM, (16), implies that

$$\|r_{\theta_n}\|^2 = o((\theta_n - \theta_0)^2). \quad (23)$$

Then, (21), (20), (22), and (23) imply that the second, third, fourth and fifth sums respectively are  $o(\theta_0 - \theta_n)$ . Substituting into (17) yields

$$0 = 2(\theta_n - \theta_0)\langle \xi_{\theta_0}, \Delta_{\theta_0} \rangle + o(\theta - \theta_n). \quad (24)$$

This implies that  $\langle \xi_{\theta_0}, \Delta_{\theta_0} \rangle$  is  $o(1)$  and, since it does not depend upon  $n$ ,

$$\langle \xi_{\theta_0}, \Delta_{\theta_0} \rangle = 0 \quad (25)$$

Then, (25), (20), (22), and (23) imply

$$\begin{aligned} 0 &= 0 + 2\langle \xi_{\theta_0}, r_{\theta_n} \rangle + \frac{1}{4}I(\theta_0)(\theta_n - \theta_0)^2 + o((\theta_n - \theta_0)^2) \\ \langle \xi_{\theta_0}, r_{\theta_n} \rangle &= -\frac{1}{2}(\theta_n - \theta_0)^2 \frac{1}{4}I(\theta_0) + o((\theta_n - \theta_0)^2). \end{aligned} \quad (26)$$

□

Pollard then showed how the previous lemma implies that a MLE that satisfies the DQM requirement also satisfies the LAN condition. Pollard proceeded as follows. Let  $D_{\theta_0}(x) := \frac{\Delta_{\theta_0}(x)}{\xi_{\theta_0}(x)}$  and  $R_{\theta_n}(x) := \frac{r_{\theta_n}(x)}{\xi_{\theta_0}(x)}$ . Then, if  $\theta_n = \theta_0 + \frac{t}{\sqrt{n}}$ , (15) can be restated as

$$\frac{\xi_{\theta_0 + \frac{t}{\sqrt{n}}}(x_i)}{\xi_{\theta_0}(x_i)} = 1 + \frac{t}{\sqrt{n}}D_{\theta_0}(x_i) + R_{\theta_n}(x_i). \quad (27)$$

Pollard noted that  $2D_{\theta_0}(X) = \frac{2}{\sqrt{f_{\theta_0}(x)}} \frac{\partial \sqrt{f_{\theta_0}(x)}}{\partial \theta} = \frac{\partial \ln f_{\theta_0}(x)}{\partial \theta}$ . Then

$$E_{\theta_0}D_{\theta_0}(X) = 0, \text{ and } E_{\theta_0}D_{\theta_0}(X)^2 = \frac{1}{4}I(\theta_0). \quad (28)$$

It is convenient to note here that the expectation of  $R_{\theta_n}(X)^2$  can be found as follows

$$\begin{aligned} E \sum_{i=1}^n R_{\theta_n}(X)^2 &= n \int f_{\theta_0}(x) \frac{r_{\theta_n}(x)^2}{f_{\theta_0}(x)} dx \\ &= n \int r_{\theta_n}(x)^2 dx \\ &= o(1). \end{aligned} \quad (29)$$

Pollard then stated another lemma, which provides three conditions that are necessary to the proof.

**Lemma 2.** *Given that  $f_{\theta_0}(x)$  fulfills the DQM property,*

$$(a) \quad |\max_{i < n} D_{\theta_0}(x_i)| = o_P(\sqrt{n})$$

$$(b) \quad |\max_{i < n} R_{\theta_n}(x_i)| = o(1)$$

$$(c) \quad \sum_{i=1}^n 2R_{\theta_n}(x_i) = -\frac{n(\theta_n - \theta_0)^2}{4} I(\theta_0) + o\left(\frac{1}{n}\right).$$

*Proof.* Part (a) of the Lemma can be found as follows,

$$\begin{aligned} P_{\theta_0}(|\max_{i < n} D_{\theta_0}(X_i)| \geq \epsilon\sqrt{n}) &< \sum_{i=1}^n P_{\theta_0}(|D_{\theta_0}(X_i)| \geq \epsilon\sqrt{n}) \\ &= nP_{\theta_0}(D_{\theta_0}(X_1) \geq \epsilon\sqrt{n}) \text{ as all } X_i \text{ are identically distributed} \\ &\leq \epsilon^{-2} E_{\theta_0} D_{\theta_0}(X)^2 I D_{\theta_0}(X_1) \geq \epsilon\sqrt{n} \\ &\xrightarrow[n \rightarrow \infty]{} 0, \end{aligned}$$

by dominated convergence. Part (b) of the Lemma can be found as using Chebyshev's inequality as follows,

$$\begin{aligned} P_{\theta_0}(\max_{i < n} R_{\theta_n}(x_i) > \epsilon) &< P_{\theta_0}\left(\sum_{i=1}^n R_{\theta_n}(x_i) > \epsilon\right) \\ &< \epsilon^{-2} \text{Var}_{\theta_0} \sum_{i=1}^n R_{\theta_n}(X) \text{ from Chebyshev's inequality} \\ &< \epsilon^{-2} n E_{\theta_0} R_{\theta_n}(X)^2 \quad \text{as all } X_i \text{ are identically distributed} \\ &= o_p(1) \quad \text{by (29).} \end{aligned}$$

Part (c) of the Lemma is where the property of the fixed norm comes into effect. First it can be noted that the variance of  $2 \sum_{i=1}^n R_{\theta_n}(x_i)$  is bounded above by  $4E_{\theta_0} \sum_{i=1}^n R_{\theta_n}(x_i)^2$ , which converges to zero by (29). This implies that  $2 \sum_{i=1}^n R_{\theta_n}(x_i)$  converges to its expected value, which can be found as follows,

$$\begin{aligned} 2nE_{\theta_0} R_{\theta_n} &= 2n \int_{-\infty}^{\infty} \xi_{\theta_0}(x) r_{\theta_n}(x) dx \\ &= -\frac{n(\theta_n - \theta_0)^2}{4} I(\theta_0) + o\left(\frac{1}{n}\right) \text{ by (26).} \end{aligned}$$

□

Then, using (27) the log likelihood ratio can be re-expressed as

$$\ln \frac{\prod_{i=1}^n f_{\theta_n}(x_i)}{\prod_{i=1}^n f_{\theta_0}(x_i)} = \sum_{i=1}^n 2 \ln \frac{\xi_{\theta_0 + \frac{t}{\sqrt{n}}}(x_i)}{\xi_{\theta_0}(x_i)} = \sum_{i=1}^n 2 \ln \left( 1 + \frac{t}{\sqrt{n}} D_{\theta_0}(x_i) + R_{\theta_n}(x_i) \right). \quad (30)$$

Conditions (a) and (b) from Pollard's second Lemma assure that (30) is bounded. A Taylor expansion of  $\ln(\cdot)$  at 1 can be stated as

$$\ln(1 + y) = y - \frac{1}{2}y^2 + \frac{1}{2}\beta(y) \quad (31)$$

where  $\beta(y) = o(y^2)$  near  $y=0$ . Applying this expansion to (30) case results in

$$2 \sum_{i=1}^n \left( \frac{t}{\sqrt{n}} D_{\theta_0}(x_i) + R_{\theta_n}(x_i) \right) - \sum_{i=1}^n \left( \frac{t}{\sqrt{n}} D_{\theta_0}(x_i) + R_{\theta_n}(x_i) \right)^2 + \sum_{i=1}^n \beta \left( \frac{t}{\sqrt{n}} D_{\theta_0}(x_i) + R_{\theta_n}(x_i) \right). \quad (32)$$

Which can be rearranged to

$$\begin{aligned} & \frac{2t}{\sqrt{n}} \sum_{i=1}^n D_{\theta_0}(x_i) + 2 \sum_{i=1}^n R_{\theta_n}(x_i) - \frac{1}{n} \sum_{i=1}^n (t D_{\theta_0}(x_i))^2 \\ & - \frac{2}{\sqrt{n}} \sum_{i=1}^n t D_{\theta_0}(x_i) R_{\theta_n}(x_i) - \sum_{i=1}^n R_{\theta_n}(x_i)^2 \\ & + o(1) \sum_{i=1}^n \frac{t}{n} D_{\theta_0}(x_i)^2 + R_{\theta_n}(x_i)^2. \end{aligned} \quad (33)$$

The first sum converges in distribution by the Central Limit Theorem and (28) to

$$\sqrt{n}(\theta_n - \theta_0) \sqrt{I(\theta)} Z, \quad (34)$$

the second sum converges to  $-\frac{n(\theta_n - \theta_0)^2}{4} I(\theta_0)$  by Pollard's second Lemma part c and the third sum converges to

$$\frac{n}{4} (\theta_n - \theta_0)^2 I(\theta_0) \quad (35)$$



by the law of large numbers. To show the  $o(1)$  asymptotic behavior of the last three sums of (33) the following two facts will be used. First,  $\frac{\sum_{i=1}^n D_{\theta_0}(x_i)^2}{n} \xrightarrow[n \rightarrow \infty]{} \frac{1}{4}I(\theta_0)$ , a constant, and is therefore  $O_p(1)$ . Second, the expectation of  $\sum_{i=1}^n R_{\theta_n}(x_i)^2$  is  $o(1)$  by (29).

The second fact directly establishes the behavior of expected value of the fifth sum of (33). The variance can be shown to be  $o(1)$  as,  $Var_{\theta_0} \sum_{i=1}^n R_{\theta_n}(X)^2 < E \sum_{i=1}^n R_{\theta_n}(X_i)^4 = nE_{\theta_0} R_{\theta_n}(X)^4$ . Since  $|\max_{i < n} R_{\theta_n}(x_i)| = o(1)$ , then, because of Pollard's second Lemma part a, for some  $m$ ,  $nE_{\theta_0} R_{\theta_n}(X)^4 < nE_{\theta_0} m^2 R_{\theta_n}(X)^2 = o(1)$ , which implies that the fourth sum is  $o(1)$ . The  $o(1)$  behavior of the fourth sum can be established using (22) by first showing that the expectation of the sum is  $o(1)$ , then showing that the variance is also  $o(1)$ . The expectation can be found as

$$\begin{aligned} \frac{2}{\sqrt{n}} E_{\theta_0} \sum_{i=1}^n t D_{\theta_0}(x_i) R_{\theta_n}(x_i) &= 2t\sqrt{n} E_{\theta_0} D_{\theta_0}(X) R_{\theta_n}(X) \\ &= 2t\sqrt{n} \int_{-\infty}^{\infty} f_{\theta_0}(x) \frac{\Delta_{\theta_0}(x)}{\xi_{\theta_0}} \frac{r_{\theta_n}(x)}{\xi_{\theta_0}} dx \\ &= 2t\sqrt{n} \int_{-\infty}^{\infty} \Delta_{\theta_0}(x) r_{\theta_n}(x) dx \\ &= 2t\sqrt{n} o\left(\frac{1}{\sqrt{n}}\right) \text{ by (22)} \\ &= o(1). \end{aligned}$$

The variance can be shown to be  $o(1)$  as follows:

$$Var_{\theta_0} \frac{2}{\sqrt{n}} \sum_{i=1}^n t D_{\theta_0}(x_i) R_{\theta_n}(x_i) \leq 4t^2 Var_{\theta_0} D_{\theta_0}(X) R_{\theta_n}(X)$$

as the covariance of residuals is negative. Since the expectation of  $D_{\theta_0}(X) R_{\theta_n}(X)$  is  $o(1)$  the variance can be considered equal to the second moment. Then, using the Cauchy-Schwarz inequality,

$$4t^2 E_{\theta_0} (D_{\theta_0}(X) R_{\theta_n}(X))^2 \leq 4t^2 E_{\theta_0} D_{\theta_0}(X)^2 E_{\theta_0} R_{\theta_n}(X)^2.$$

Using (16) we can show that  $E_{\theta_0} R_{\theta_n}(X)^2 = o(\frac{1}{n})$  and (28) states that  $E_{\theta_0} D_{\theta_0}(X)^2 = O(1)$ . This implies that the variance of the fourth sum converges to zero, which implies the fourth sum is  $o(1)$ . The last sum of (33), can be shown

to be  $o(1)$  by noting that  $\sum_{i=1}^n R_{\theta_n}(x_i)^2 = o(1)$  and  $\sum_{i=1}^n \frac{t}{n} D_{\theta_0}(x_i)^2 = O(1)$ , which multiplied by  $o(1)$  is  $o(1)$ .

Then, the log likelihood ratio can be expressed as

$$\ln \frac{\prod_{i=1}^n f_{\theta_n}(x_i)}{\prod_{i=1}^n f_{\theta_0}(x_i)} = \sqrt{n}(\theta_n - \theta_0) \sqrt{I(\theta)} Z - \frac{n}{2}(\theta_n - \theta_0)^2 I(\theta_0) + o(1), \quad (36)$$

which delivers the LAN condition, implying asymptotic normality.

The Laplace distribution is an interesting case because its pdf lacks a derivative in  $\theta$  at  $\theta = \theta_0$ . However, the Laplace distribution meets the DQM condition. This can be seen as follows, where  $f_{\theta}(x) = \frac{1}{2}e^{-|x-\theta|}$ ,  $\xi_{\theta}(x) = \frac{1}{\sqrt{2}}e^{-|x-\theta|/2}$ , and  $\Delta_{\theta}(x) = \frac{1}{2\sqrt{2}}e^{-|x-\theta|/2} \text{sgn}(x - \theta)$ . If  $\theta = \theta_0 + \epsilon$ , then

$$\begin{aligned} r_{\theta_0+\epsilon}(x) &= \xi_{\theta_0+\epsilon}(x) - \xi_{\theta_0}(x) - \epsilon \Delta_{\theta_0}(x) \\ r_{\theta_0+\epsilon}(x) &= \frac{1}{\sqrt{2}}e^{-|x-(\theta_0+\epsilon)|/2} - \frac{1}{\sqrt{2}}e^{-|x-\theta_0|/2} - \frac{\epsilon}{2\sqrt{2}}e^{-|x-\theta_0|/2} \text{sgn}(x - \theta_0). \end{aligned}$$

Then DQM condition requires that  $\sqrt{\int_{\mathbb{R}} r_{\theta}(x)^2 dx} = o(\theta - \theta_0)$ . This can be checked directly as follows for  $\epsilon > 0$ , with the case of  $\epsilon < 0$  being similar

$$\begin{aligned} \int_{\mathbb{R}} r_{\theta}(x)^2 dx &= \int_{\mathbb{R}} r_{\theta_0+\epsilon}(x)^2 dx \\ &= \int_{\mathbb{R}} \left( \frac{1}{\sqrt{2}}e^{-|x-(\theta_0+\epsilon)|/2} - \frac{1}{\sqrt{2}}e^{-|x-\theta_0|/2} - \frac{\epsilon}{2\sqrt{2}}e^{-|x-\theta_0|/2} \text{sgn}(x - \theta_0) \right)^2 dx \\ &= \int_{-\infty}^{\theta_0} \left( \frac{1}{\sqrt{2}}e^{-((\theta_0+\epsilon)-x)/2} - \frac{1}{\sqrt{2}}e^{-(\theta_0-x)/2} + \frac{\epsilon}{2\sqrt{2}}e^{-(\theta_0-x)/2} \right)^2 dx \\ &+ \int_{-\theta_0}^{\theta_0+\epsilon} \left( \frac{1}{\sqrt{2}}e^{-((\theta_0+\epsilon)-x)/2} - \frac{1}{\sqrt{2}}e^{-(x-\theta_0)/2} - \frac{\epsilon}{2\sqrt{2}}e^{-(x-\theta_0)/2} \right)^2 dx \\ &+ \int_{\theta_0+\epsilon}^{\infty} \left( \frac{1}{\sqrt{2}}e^{-(x-(\theta_0+\epsilon))/2} - \frac{1}{\sqrt{2}}e^{-(x-\theta_0)/2} - \frac{\epsilon}{2\sqrt{2}}e^{-(x-\theta_0)/2} \right)^2 dx \\ &= 2\left(1 + \epsilon^2/8 - e^{-\frac{\epsilon}{2}}\left(1 + \frac{\epsilon}{2} + \frac{\epsilon^2}{4}\right)\right) \\ &= \frac{\epsilon^3}{6} + O(\epsilon^4) \\ &= o(\epsilon^2). \end{aligned}$$

## 2.4 Concluding Remarks on the Literature

Fisher’s result has several more restrictive assumptions than the later results, with the requirement of two derivatives in  $\theta$  of the pdf and asymptotic normality of the MLE. These requirements work fairly well, especially since many MLEs can be shown to be the mean of values that have an expected value. Doob’s and Cramér’s results removed the requirement of the asymptotic normality of the MLE, but added a third derivative in  $\theta$  condition. This was a marked improvement over Fisher’s result in that it could be used to find the asymptotic distribution of a MLE versus the asymptotic distribution needed to be found first. Also, Cramér’s result made on Doob’s result more communicateable by stating Doob’s assumptions as a requirement for the third derivative of the pdf in  $\theta$ , which is equivalent to Doob’s requirement in most situations. LeCam’s result replaces the other derivative requirements with the DQM. The DQM is a much reduced requirement, but does not include distributions with discontinuities in the pdf. The current result only requires one directional derivative in  $\theta$ , but also has the condition of log-concavity of the pdf. The current formulation causes it to be fairly restrictive, however, the strict log-concavity could be replaced by local log-concavity near the MLE, which, with some further work, may be able to handle discontinuities in the pdf.

## 3 Result

### 3.1 Smoothness Conditions

The following assumptions and definitions on the properties of  $f_\theta(x)$  will be required for the convergence conditions.

- Cond. 1 The parameter space  $\Theta$  is a nonempty open interval in  $\mathbb{R}$ . Let  $\theta_0 \in \Theta$  denote the “true” value of the parameter.
- Cond. 2 Suppose that the set  $\mathcal{X} := \{x \in \mathbb{R} : f_\theta(x) > 0\}$  does not depend on  $\theta$ .
- Cond. 3 Suppose that  $\ln f_\theta(x)$  is concave in  $\theta$  for each  $x \in \mathcal{X}$ ; so, for such  $x$ ,  $\ln f_\theta(x)$  has left and right derivatives in  $\theta$ , and  $f_\theta(x)$  is absolutely

continuous in  $\theta$ . Let the score function be stated as

$$Y_{i,\theta} := \frac{\partial^-}{\partial\theta} \ln f_\theta(X_i),$$

where  $\frac{\partial^-}{\partial\theta}$  denotes the left derivative.

The scaled log-likelihood function is defined by the formula

$$\ell_{\underline{\mathbf{x}}}(\theta) := \frac{1}{n} \sum_{i=1}^n \ln f_\theta(x_i),$$

where  $\underline{\mathbf{x}} := (x_1, \dots, x_n)$ . Since  $\ln f_\theta(x)$  is concave in  $\theta$ , the log-likelihood formula  $\ell_{\underline{\mathbf{x}}}(\theta)$  is concave in  $\theta$  for each  $\underline{\mathbf{x}} \in \mathcal{X}^n$ .

Cond. 4 Let  $\widehat{\Theta}_{\max} := \widehat{\Theta}_{\max}(\underline{\mathbf{x}}) := \{\theta \in \widehat{\Theta} : \ell_{\underline{\mathbf{x}}}(\theta) \geq \ell_{\underline{\mathbf{x}}}(\widetilde{\theta}) \quad \forall \widetilde{\theta} \in \Theta\}$ , with  $\underline{\mathbf{x}} \in \mathcal{X}^n$ . Assume that  $\widehat{\Theta}_{\max}(\underline{\mathbf{x}})$  is nonempty and bounded for each  $\underline{\mathbf{x}} \in \mathcal{X}^n$ . Note that  $\widehat{\Theta}_{\max}(\underline{\mathbf{x}})$  is also closed, since the function  $\ell_{\underline{\mathbf{x}}}(\theta)$  is concave on an open interval  $\Theta$  and hence continuous. Therefore, the definition

$$\widehat{\theta} := \widehat{\theta}_{MLE} := \widehat{\theta}_{MLE}(\underline{\mathbf{x}}) := \max \widehat{\Theta}_{\max}(\underline{\mathbf{x}})$$

is valid; that is,  $\widehat{\theta} = \max \widehat{\Theta}_{\max}(\underline{\mathbf{x}})$  exists in  $\mathbb{R}$  for each  $\underline{\mathbf{x}} \in \mathcal{X}^n$ . Note that  $\widehat{\theta}$  is the largest of all MLE's.

Cond. 5 Suppose that  $m := M_{\theta_0}(\theta_0)$  exists, where  $M_{\theta_0}(\theta) := \frac{\partial}{\partial\theta} E_{\theta_0} Y_{i,\theta}$  and  $m \in (-\infty, 0)$ . This requires that the expectation of  $Y_{i,\theta}$  exist, and the derivative in  $\theta$  of  $E_{\theta_0} Y_{i,\theta}$  also exist at  $\theta = \theta_0$ .

Cond. 6 Suppose that  $\sigma_0 := \sqrt{\text{Var}_{\theta_0}(Y_{i,\theta})} \in (0, \infty)$ .

Cond. 7 Suppose that  $\int_{\theta_1}^{\theta_2} d\theta \int_{\mathcal{X}} dx |\frac{\partial^-}{\partial\theta} f_\theta(x_1)| < \infty$  for any  $\theta_1$  and  $\theta_2$  in  $\Theta$  such that  $\theta_1 < \theta_2$ . This condition is needed for an application of Fubini's theorem.

Cond. 8 Let  $G(\theta) := \int_{\mathbb{R}} dx \frac{\partial^-}{\partial\theta} f_\theta(x_1)$  and assume that  $G(\theta)$  is continuous.

### 3.2 Theorem and Proof

**Theorem 1.** *Let  $X_1, X_2, \dots$ , be independent identically distributed (iid) random variables (r.v.'s) with pdf  $f_{\theta_0}(x)$ , and assume that Conditions 1-8 hold. Then  $(\hat{\theta} - \theta_0) \xrightarrow[\sigma_0]{-m\sqrt{n}} D N(0, 1)$ .*

*Proof.*

$$P_{\theta_0}(\hat{\theta} < \theta) = P_{\theta_0}\left(\frac{\partial^-}{\partial \theta} \ell_{\mathbf{x}}(\theta) < 0\right). \quad (37)$$

Concerning the probability on the right hand side of the above identity, one can write:

$$\begin{aligned} P_{\theta_0}\left(\frac{\partial^-}{\partial \theta} \ell_{\mathbf{x}}(\theta) < 0\right) &= P_{\theta_0}\left(\sum_{i=1}^n \frac{\partial^-}{\partial \theta} \ln(f_{\theta}(X_i)) < 0\right) \\ &= P_{\theta_0}\left(\sum_{i=1}^n Y_{i,\theta} < 0\right) \\ &= P_{\theta_0}\left(Z_n < -\frac{\sqrt{n}E_{\theta_0}Y_{1,\theta}}{\sigma_0}\right), \end{aligned} \quad (38)$$

where

$$Z_n := \frac{\sum_{i=1}^n (Y_{i,\theta} - E_{\theta_0}Y_{1,\theta})}{\sqrt{n}\sigma_0} \xrightarrow{D} N(0, 1), \quad (39)$$

by the central limit theorem. This means that  $P_{\theta_0}(Z_n < z) \xrightarrow[n \rightarrow \infty]{} P(Z < z)$  for each  $z \in \mathbb{R}$ , where  $Z \sim N(0, 1)$ . Moreover ([Chow and Teicher, 1962, p. 265]), this convergence is uniform in  $z$ :

$$\sup_{z \in \mathbb{R}} |P_{\theta_0}(Z_n < z) - P(Z < z)| \xrightarrow[n \rightarrow \infty]{} 0.$$

So,

$$P_{\theta_0}\left(Z_n < -\frac{\sqrt{n}E_{\theta_0}Y_{1,\theta}}{\sigma_0}\right) - P_{\theta_0}\left(Z < -\frac{\sqrt{n}E_{\theta_0}Y_{1,\theta}}{\sigma_0}\right) \xrightarrow[n \rightarrow \infty]{} 0 \quad (40)$$

where  $Z \sim N(0, 1)$ .

By Condition 5,

$$E_{\theta_0}Y_{1,\theta} = E_{\theta_0}Y_{1,\theta_0} + \frac{\partial}{\partial \theta} E_{\theta_0}Y_{1,\theta}|_{\theta=\theta_0}(\theta - \theta_0) + o(|\theta - \theta_0|). \quad (41)$$

$E_{\theta_0}Y_{1,\theta_0}$  can be shown to be zero, and a proof of this is given in a technical appendix, see section 3.5.

Then, (41) can be restated as  $E_{\theta_0}Y_{1,\theta} = \frac{\partial}{\partial\theta_0}E_{\theta_0}Y_{1,\theta_0}(\theta - \theta_0) + o(|\theta - \theta_0|)$ . Since  $m = \frac{\partial}{\partial\theta}E_{\theta_0}Y_{i,\theta}|_{\theta=\theta_0}$  by Condition 5, one has

$$E_{\theta_0}Y_{1,\theta} = m(\theta - \theta_0) + o(|\theta - \theta_0|) = (m + o(1))(\theta - \theta_0). \quad (42)$$

Let  $\theta - \theta_0$  vary on the order of  $\frac{1}{\sqrt{n}}$ . More precisely, let

$$\theta = \theta_0 - \frac{\sigma_0 z}{m\sqrt{n}}, \quad (43)$$

where  $z$  is a fixed real number. Then by (42)

$$\sqrt{n}E_{\theta_0}Y_{1,\theta} = -\sigma_0 z \frac{(m + o(1))}{m} \xrightarrow{n \rightarrow \infty} -\sigma_0 z.$$

Using this approximation and the continuity of the standard normal cdf, we have

$$P_{\theta_0}\left(Z < -\frac{\sqrt{n}E_{\theta_0}Y_{1,\theta}}{\sigma_0}\right) \xrightarrow{n \rightarrow \infty} P_{\theta_0}(Z < z). \quad (44)$$

The probability on the LHS of (37) can be written as

$$\begin{aligned} P_{\theta_0}(\widehat{\theta} < \theta) &= P_{\theta_0}\left(Z < -\frac{\sqrt{n}E_{\theta_0}Y_{1,\theta}}{\sigma_0}\right) \\ &\quad + \left(P_{\theta_0}(\widehat{\theta} < \theta) - P_{\theta_0}\left(Z < -\frac{\sqrt{n}E_{\theta_0}Y_{1,\theta}}{\sigma_0}\right)\right). \end{aligned} \quad (45)$$

The first term of the above sum converges to  $P_{\theta_0}(Z < z)$  by (44), and the second term converges to 0 by (37), (38), and (40). Then,

$$P_{\theta_0}(\widehat{\theta} < \theta) \xrightarrow{D} P(Z < z). \quad (46)$$

$$P_{\theta_0}(\widehat{\theta} < \theta) = P_{\theta_0}(\widehat{\theta} - \theta_0 < \theta - \theta_0) = P_{\theta_0}\left(\widehat{\theta} - \theta_0 < -\frac{\sigma_0 z}{m\sqrt{n}}\right) \quad (47)$$

$$= P_{\theta_0}\left(\left(\widehat{\theta} - \theta_0\right) \frac{-m\sqrt{n}}{\sigma_0} < z\right). \quad (48)$$

This, together with (46) implies that

$$(\hat{\theta} - \theta_0) \frac{-m\sqrt{n}}{\sigma_0} \xrightarrow{D} N(0, 1). \quad (49)$$

□

### 3.3 Discussion

Recall that our MLE  $\hat{\theta}$  was defined as the maximum of the set of  $\hat{\Theta}_{\max}$ . However, the proof could be similarly constructed using the definition  $\hat{\theta} := \hat{\theta}_{MLE} := \hat{\theta}_{MLE}(\mathbf{x}) := \min \hat{\Theta}_{\max}(\mathbf{x})$ .

The usefulness of the above result stems from two sources. Log-concave pdf's are rather common. In particular, the pdf of the normal distribution, the pdf of the Laplace distribution, the pdf of  $\chi_d := \sqrt{\chi_d^2}$ , the pdf of  $\chi_k^2$ ,  $k \geq 2$ , and the logistic pdf  $e^x(1 + e^x)^{-2}$  are log-concave; see e.g. Pinelis [1999].

$\ln h_\theta(x) = \ln f_\theta(x) + \ln g_\theta(x)$ . So, if  $\ln f_\theta(x)$  and  $\ln g_\theta(x)$  are concave in  $\theta$ ,  $\ln h_\theta(x)$  is also concave in  $\theta$  as the sum of concave functions is also concave.

In particular in the Bayesian context, if the prior density  $\pi(\theta)$  and the likelihood function  $f_\theta(x)$  are both log-concave in  $\theta$ , the resultant posterior pdf  $f(\theta|x)$  is log-concave. This can be shown as follows. It is known that the posterior distribution is proportional to the product of the prior and the likelihood function:  $f(\theta|x) \propto \pi(\theta)f_\theta(x)$ . Therefore,  $\ln f(\theta|x) = \ln \pi(\theta) + \ln f_\theta(x) + c(x)$ , then  $\ln f(\theta|x)$  is concave in  $\theta$ .

**Example 1.** *For a generalized continuous version of the Laplace distribution, with pdf*

$$f_\theta(x) = ce^{-\frac{\theta-x}{b_1}} I(x \leq \theta) + ce^{-\frac{x-\theta}{b_2}} I(x > \theta) \quad (50)$$

where  $c, b_1, b_2$  are known,  $c, b_1,$  and  $b_2$  in  $\mathbb{R}^+$ ,  $c(b_1 + b_2) = 1$  and  $\theta \in \mathbb{R}$  we will find the limit distribution of the maximum likelihood estimator for the parameter  $\theta$  using the shown result.

Then Conditions 1-8 on pages 2 and 3 are satisfied:

Cond. 1: The parameter space is  $\mathbb{R}$ .

Cond. 2: The support of the distribution is  $\mathbb{R}$  for all  $\theta$ .

Cond. 3:  $\ln f_\theta(x) = -\frac{\theta-x}{b_1}I(x \leq \theta) + -\frac{x-\theta}{b_2}I(x > \theta)$  is concave because for  $\theta < x$ ,  $\frac{\partial}{\partial \theta} \ln f_\theta(x) = \frac{1}{b_2} > 0$  and for  $\theta > x$ ,  $\frac{\partial}{\partial \theta} \ln f_\theta(x) = -\frac{1}{b_1} < 0$ , then  $\frac{\partial}{\partial \theta} \ln(f_\theta(x))$  is nondecreasing in  $\theta$ , and since  $\ln(f_\theta(x))$  is continuous in  $\theta$ ,  $\ln(f_\theta(x))$  is concave in  $\theta$

Cond. 4: The condition that  $\widehat{\Theta}_{\max}(\mathbf{x})$  is nonempty and bounded is met can be shown as follows. For  $\theta > \max_i x_i$ ,  $\frac{\partial}{\partial \theta} \ln(f_\theta(x)) < 0$  and for  $\theta < \min_i x_i$ ,  $\frac{\partial}{\partial \theta} \ln(f_\theta(x)) > 0$ . This leaves  $\theta \in [\min_i x_i, \max_i x_i]$ , and since  $\ln(f_\theta(x))$  is continuous in  $\theta$ , there is a maximum in a bounded set.

Cond. 5: The constant  $m = \frac{\partial}{\partial \theta_0} E_{\theta_0} \frac{\partial}{\partial \theta} \ln(f_\theta(X_1))$  can be found as follows,

$$\begin{aligned}
m &= \frac{\partial}{\partial \theta} E_{\theta_0} \frac{\partial}{\partial \theta} \ln(f_\theta(x_1))|_{\theta=\theta_0} \\
&= \frac{\partial}{\partial \theta} \int \frac{\partial}{\partial \theta} \ln (ce^{-\frac{\theta-x}{b_1}} I(x \leq \theta) + ce^{-\frac{x-\theta}{b_2}} I(x > \theta)) f_{\theta_0}(x) dx |_{\theta=\theta_0} \\
&= \frac{\partial}{\partial \theta} \int (-\frac{1}{b_1} I(x \leq \theta) + \frac{1}{b_2} I(x > \theta)) f_{\theta_0}(x) dx |_{\theta=\theta_0} \\
&= \frac{\partial}{\partial \theta} (-\frac{1}{b_1} \int_{-\infty}^{\theta_0} f_{\theta_0}(x) dx + \frac{1}{b_2} \int_{\theta_0}^{\infty} f_{\theta_0}(x) dx) |_{\theta=\theta_0} \\
m &= -c(\frac{1}{b_1} + \frac{1}{b_2}) = -\frac{1}{b_1 b_2} < 0.
\end{aligned}$$

Cond. 6: The constant  $\sigma_0 = \sqrt{Var(E_{\theta_0} \frac{\partial}{\partial \theta} \ln(f_\theta(X_1)))}$ , can be found as follows.



$$\begin{aligned}
\text{Var}_{\theta_0} \frac{\partial^-}{\partial \theta} \ln(f_{\theta}(X_1)) &= E_{\theta_0} \left( \frac{\partial^-}{\partial \theta} \ln(f_{\theta}(X_1)) \right)^2 \\
&= \int \left( \frac{\partial^-}{\partial \theta} \ln \left( ce^{-\frac{\theta-x}{b_1}} I(x \leq \theta) \right. \right. \\
&\quad \left. \left. + ce^{-\frac{x-\theta}{b_2}} I(x > \theta) \right) \right)^2 f_{\theta_0}(x) dx \\
&= \int \left( -\frac{1}{b_1} I(x \leq \theta) \frac{1}{b_2} I(x > \theta) \right)^2 f_{\theta_0}(x) dx \\
&= \left( \frac{1}{b_1} \right)^2 \int_{-\infty}^{\theta_0} f_{\theta_0}(x) dx + \left( \frac{1}{b_2} \right)^2 \int_{\theta_0}^{\infty} f_{\theta_0}(x) dx \\
&= c \left( \frac{1}{b_1} + \frac{1}{b_2} \right) = \frac{1}{b_1 b_2}.
\end{aligned}$$

Then,  $\sqrt{\text{Var}_{\theta_0} \frac{\partial^-}{\partial \theta} \ln(f_{\theta}(X_1))} = \sigma_0 = \sqrt{\frac{1}{b_1 b_2}} > 0$ .

Cond. 7: To establish that Fubini's theorem could be applied, we must show that  $\int_{\theta_1}^{\theta_2} d\theta \int_{-\infty}^{\infty} dx \left| \frac{\partial^-}{\partial \theta} f_{\theta_0}(x_1) \right| < \infty$ . This can be demonstrated as follows:

$$\begin{aligned}
\frac{\partial^-}{\partial \theta} f_{\theta_0}(x_1) &= c \left( -\frac{1}{b_1} e^{-\frac{\theta-x}{b_1}} I(x \leq \theta) + \frac{1}{b_2} e^{-\frac{x-\theta}{b_2}} I(x > \theta) \right) \\
\left| \frac{\partial^-}{\partial \theta} f_{\theta_0}(x_1) \right| &= c \left( \frac{1}{b_1} e^{-\frac{\theta-x}{b_1}} I(x \leq \theta) + \frac{1}{b_2} e^{-\frac{x-\theta}{b_2}} I(x > \theta) \right)
\end{aligned}$$

Then, for any  $\theta_2 > \theta_1$ ,

$$\begin{aligned}
\int_{\theta_1}^{\theta_2} d\theta \int_{-\infty}^{\infty} dx \left| \frac{\partial^-}{\partial \theta} f_{\theta_0}(x_1) \right| &= c \int_{\theta_1}^{\theta_2} d\theta \int_{-\infty}^{\infty} dx \left( \frac{1}{b_1} e^{-\frac{\theta-x}{b_1}} I(x \leq \theta) \right. \\
&\quad \left. + \frac{1}{b_2} e^{-\frac{x-\theta}{b_2}} I(x > \theta) \right) \\
&= c \int_{\theta_1}^{\theta_2} d\theta \left( \int_{-\infty}^{\theta} \frac{1}{b_1} e^{-\frac{\theta-x}{b_1}} dx \right. \\
&\quad \left. + \int_{\theta}^{\infty} \frac{1}{b_2} e^{-\frac{x-\theta}{b_2}} dx \right) \\
&= 2c \int_{\theta_1}^{\theta_2} d\theta \\
&= 2c(\theta_2 - \theta_1) < \infty.
\end{aligned}$$

Cond. 8: To establish the continuity of  $G(\theta) = \int_{\mathbb{R}} dx \frac{\partial^-}{\partial \theta} f_{\theta}(x_1)$ , for any  $\theta \in \Theta$

$$G(\theta) = c \left( \int_{-\infty}^{\theta} -\frac{1}{b_1} e^{-\frac{\theta-x}{b_1}} dx + \int_{\theta}^{\infty} \frac{1}{b_2} e^{-\frac{x-\theta}{b_2}} dx \right) = c(-1 + 1) = 0$$

Also, in S. Kotz and Podgoński [2002] the value of the MLE is found to be the quantile,  $\hat{\theta} = Q_{cb_1}(\mathbf{x})$ , where  $Q_{cb_1}$  is the  $cb_1$  quantile. The previous statements establish that the generalized continuous Laplace distribution meets the assumptions for this theorem. Then, according to Theorem 1,  $\hat{\theta}$  in this setting is asymptotically normal with mean  $\theta_0$  and asymptotic variance  $\frac{b_1 b_2}{n}$ .

$$(\hat{\theta} - \theta_0) \frac{-m\sqrt{n}}{\sigma_0} = (\hat{\theta} - \theta_0) \frac{\sqrt{n}}{\sqrt{b_1 b_2}} \xrightarrow{n \rightarrow \infty} N(0, 1). \quad (51)$$

This asymptotic variance  $\frac{b_1 b_2}{n}$  coincides with the Cramér–Rao lower bound on the variance.

### 3.4 Simulations of the MLE for the Generalized Continuous Laplace distribution

This section illustrates the CLT (51) through the inverse-transform method of simulation. The Laplace distribution is an interesting case because its pdf lacks a derivative in  $\theta$  at  $\theta = \theta_0$ . In Table 1, the asymptotic variance  $\frac{b_1 b_2}{n}$  and the simulated variance of  $\hat{\theta}$  are compared for several sample sizes  $n$  for the example of  $c = 1, \theta_0 = 0, b_1$  and  $b_2 = .5$  in (50). The simulated variance of the MLE is found from 1000 simulations of the MLE from samples of size  $n$ , where  $n = 100, 250, 500, 1000$  in Mathematica (see attached code).

$n$	Asymptotic Variance of MLE	Simulated Variance of MLE
100	.00250	.00304
250	.00100	.00108
500	.000500	.000537
1000	.000250	.000266

Table 1: Simulation of MLE variance for various sample sizes  $n$

The simulated variance appears to converge to the asymptotic variance as  $n$  increases.

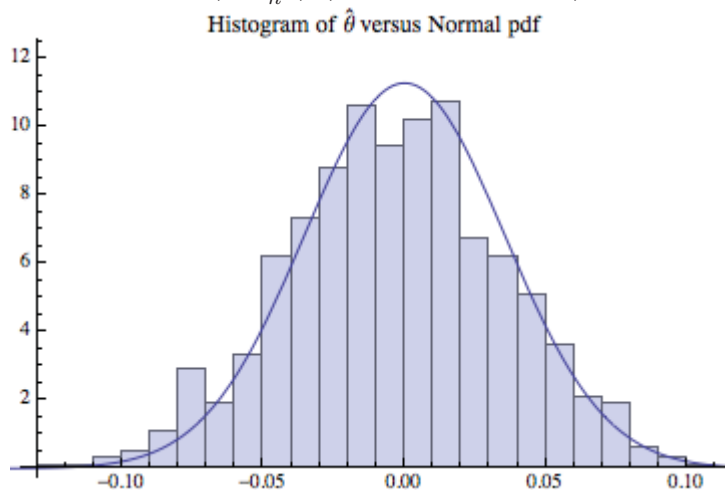
In Table 2 the asymptotic variance  $\frac{b_1 b_2}{n}$  and the simulated variance of  $\hat{\theta}$  are compared for different  $b_1$  and  $b_2$  with  $c = 1, \theta_0 = 0, n = 1000$  in (50). The simulated variance of the MLE is found from 1000 simulations of the MLE in Mathematica (see attached code).

$b_1$	$b_2$	Asymptotic Variance of MLE	Simulated Variance of MLE
.5	.5	.000250	.000281
.66	.33	.000222	.000228
.75	.25	.000188	.000193
.8	.2	.000160	.000175

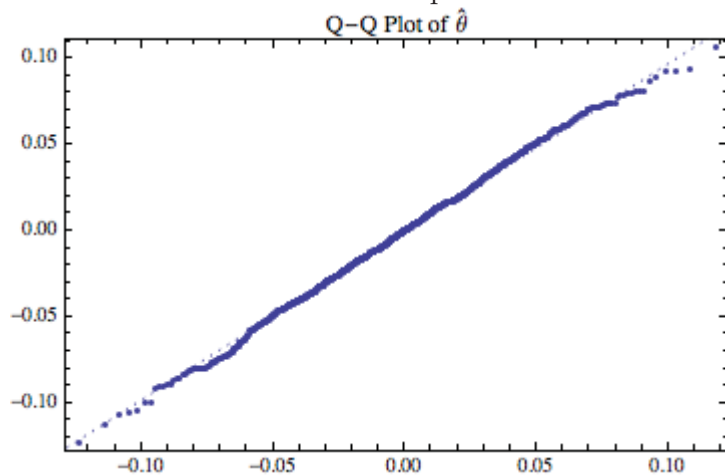
Table 2: Simulation of MLE variance for various  $b_1$  and  $b_2$  with  $n = 1000$

The simulated variance appears to be close to the asymptotic variance for different choices of  $b_1$  and  $b_2$ .

The convergence of  $\hat{\theta}$  to the normal distribution can also be seen from the convergence of the histogram of  $\hat{\theta}$  to the asymptotic normal pdf. The histogram is of 1000 estimates of the MLE of with  $c = 1, b_1 = .5, b_2 = .5, n = 200$  and  $\theta = 0$ , in (50). There appears to be strong convergence to the normal distribution  $N(0, \frac{b_1 b_2}{n})$  (see attached code).



The below normal Q-Q plot uses the same data set as the above histogram. The close linear association implies the distribution of  $\hat{\theta}$  is normal.



# A Simulation to Verify the Normality of the Maximum Likelihood Estimate of the location parameter of the Generalized Laplace Distribution

```
(* Please note that text between (**) is a
comment. *)
(* The generalized Laplace distribution has
the pdf  $f(x) = c e^{-\frac{\theta-x}{b_1}} I(x < \theta) + c e^{-\frac{x-\theta}{b_2}} I(x \geq \theta)$ ,
where  $\theta \in (-\infty, \infty)$ ,  $b_1 = b_1 \in (0, \infty)$ ,  $b_2 = b_2 \in (0, \infty)$ , and *)
c :=  $\frac{1}{b_1 + b_2}$ ;
(* Let F =
Fb1,b2, $\theta$  be the cdf of this distribution and
let F-1 = F-1b1,b2, $\theta$  be the inverse of F;
the latter is computed by the following
command: *)
FInverse[u_,  $\theta$ _, b1_, b2_] :=
If [u < c b1, b1 Log [ $\frac{u}{(c b1)}$ ] +  $\theta$ ,
( $\theta$  - b2 Log [ $\frac{1-u}{c b2}$ ])];
```

The following script finds the simulated variance of the maximum likelihood estimate (MLE) of the generalized Laplace distribution which will be defined in the comments of the script, for several different sample sizes  $n$ , where  $n = 100, 250, 500, \text{ and } 1000$ . The results produced are used in Table 1 of the Report.

```
(* The variance of the MLE  $\hat{\theta} =$ 
 $\hat{\theta}_{b_1, b_2, \theta, n}$  will be estimated for each of the
sample sizes stored in following array: *)
SampleSizes = {100, 250, 500, 1000};
```

```

(* The parameters of the Laplace distribution
   will be  $b_1=.5$ ,  $b_2=.5$ , and  $\theta=0$ . *)
(* The variance of  $\hat{\theta}$  estimated for each of
   the four sample sizes will be stored in
   the following array of length 4: *)
Variances = Array[1 &, 4];
(* In the following loop,
   for each of the sample sizes  $n = 100, 250,$ 
   500 and 1000,
   1000 independent samples are simulated and
   for each of 4x1000 samples the value of
   the MLE is computed. Then for each sample
   size an estimated variance of the MLE is
   obtained. *)
Do[
  (* The array *)
  MLEs = Array[1 &, 1000];
  (* serves as a placeholder of 1000 values of  $\hat{\theta}$ ,
   with  $n = \text{SampleSizes}[[i]]$ , with  $i = 1, 2,$ 
   3, 4. *)
  Do[
    (* With the parameter values, *)
    { $\theta$ , b1, b2} = {0, .5, .5};
    (* and  $c = \frac{1}{b_1+b_2}=1$ , the array *)
    Sample = Array[FInverse[Random[],  $\theta$ , b1, b2] &,
      SampleSizes[[i]]];
    (* is a pseudorandom sample of size
     SampleSizes[[i]] drawn from the generalized
     Laplace distribution where Random[]
     produces a pseudorandom value drawn from
      $U(0,1)$ . *)
    (* From equation 15 in the report,
      $\hat{\theta} = Q_{c b_1}(\text{Sample}) =$  *)
    MLEs[[j]] = Quantile[Sample, c b1];
    (* where the Quantile function returns
     the  $c*b_1$  quantile from the Sample array. *)
    Clear[ $\theta$ , b1, b2];
    , {j, 1000}];
  Variances[[i]] = Variance[MLEs];

```

```

(* Variances[[i]] is the sample variance
of the the 1000 values of  $\hat{\theta}$  generated
using using sample SampleSizes[[i]] *)
, {i, 4}]
(* The following command prints out a table
of the simulated variances and associated
sample sizes,
these numbers are used in Table 1 in the
report. *)
ScientificForm[
TableForm[
Transpose[ArrayFlatten[
{SampleSizes, Variances}]],
TableHeadings →
{None, {"sample size",
"simulated variance"}}]]]

```

sample size	simulated variance
100	$2.89208 \times 10^{-3}$
250	$1.1556 \times 10^{-3}$
500	$5.72624 \times 10^{-4}$
1000	$2.90022 \times 10^{-4}$

The following script find the simulated variance of the MLE for several different combinations of  $b_1$  and  $b_2$ . The results produced are used in Table 2 of the Report.

```

(* Using the aforementioned generalized
Laplace distribution, the variance of  $\hat{\theta}$ ,
will be estimated for various values of  $b_2$ ,
stored in following array: *)
b2array = {.5, .66666, .75, .8};
(* The value of c will be set to 1,
which implies that corresponding values of
b1 can be found as *)
blarray = 1 - b2array;
(* Also, in this script we will set n=
1000 and  $\theta=0$ . *)
(* The variance of  $\hat{\theta}$  estimated for each of the

```

combination of values of  $b_1$  and  $b_2$  will be stored in the following array of length four: \*)

```

Variances = Array[1 &, 4];
(* In the following loop,
for each of the parameter combinations of
b1 and b2, 1000 independent samples are
simulated and for each of 4x1000 samples
the value of the MLE is computed. Then
for each parameter combination an estimated
variance of the MLE is obtained. *)
Do [
  (* The array *)
  MLEs = Array[1 &, 1000];
  (* serves as a placeholder of 1000 values of  $\hat{\theta}$ ,
  with  $b_1=b1[[i]]$  and  $b_2 = b2[[i]]$ , with  $i = 1,$ 
  2, 3, 4. *)
  Do [
    Clear[b1, b2,  $\theta$ ];
    (* With the parameter values, *)
    {b1, b2,  $\theta$  = {b1array[[i]], b2array[[i]], 0};
    (* and  $c = \frac{1}{b_1+b_2}$ , the array *)
    Sample = Array[FInverse[Random[],  $\theta$ , b1, b2] &,
      1000];
    (* is a pseudorandom sample of size 1000
    drawn from the generalized Laplace
    distribution where Random[] produces a
    psuedorandom value drawn from  $U(0,1)$ . *)

    (* From equation 15 in the report,
     $\hat{\theta} = Q_c b_1(\text{Sample}) = *$ )
    MLEs[[j]] = Quantile[Sample, c b1];
    (* where the Quantile function returns
    the  $c*b_1$  quantile from the Sample array. *)
    Clear[b1, b2,  $\theta$ ];
    , {j, 1, 1000}];
  Variances[[i]] = Variance[MLEs];
  (* Variances[[i]] is the sample variance
  of the the 1000 values of  $\hat{\theta}$  generated

```



```

using using sample SampleSizes[[i]] *)
, {i, 1, 4}]
(* The following command prints out a table
of the simulated variances and associated
values of  $b_2$  and  $b_1$ ,
these numbers are used in Table 2 in the
report. *)
ScientificForm[
TableForm[
Transpose[ArrayFlatten[
{b1array, b2array, Variances}]],
TableHeadings →
{None, {"b1", "b2", "simulated variance"}}]]

```

$b_1$	$b_2$	simulated variance
$5. \times 10^{-1}$	$5. \times 10^{-1}$	$2.56534 \times 10^{-4}$
$3.3334 \times 10^{-1}$	$6.6666 \times 10^{-1}$	$2.13623 \times 10^{-4}$
$2.5 \times 10^{-1}$	$7.5 \times 10^{-1}$	$1.85677 \times 10^{-4}$
$2. \times 10^{-1}$	$8. \times 10^{-1}$	$1.62442 \times 10^{-4}$

The following script simulates 1000 MLE's of the generalized Laplace distribution, and compares the histogram of the MLE's to the expected normal distribution, and finds the Q-Q plot of the results.

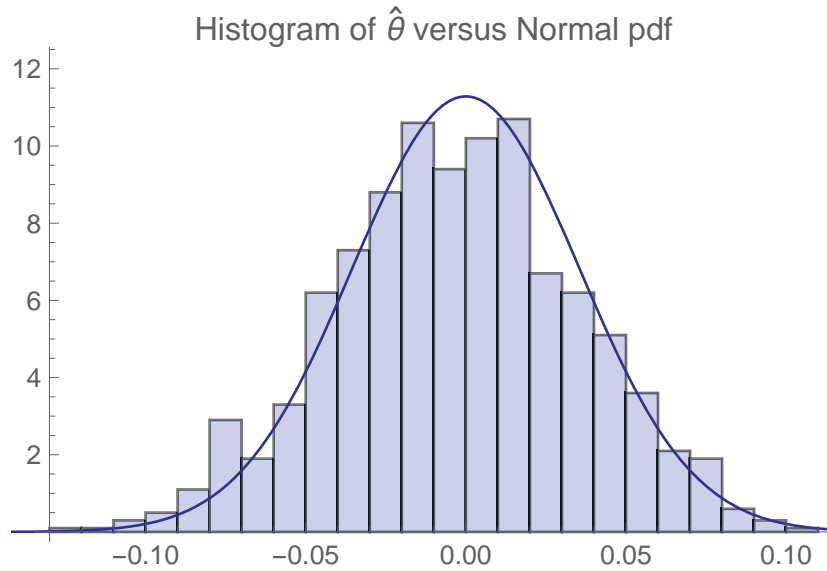
```

(* The array *)
MLEs = Array[1 &, 1000];
(* serves as a placeholder of 1000 values
of  $\hat{\theta}$ . *)

(* In the following loop,
for a sample size n = 200,
1000 independent samples are are simulated
and for each of 1000 samples the value of
the MLE is computed. *)
Do[
  (* With the parameter values, *)
  { $\theta$ , b1, b2} = {0, .5, .5};
  (* and  $c = \frac{1}{b_1+b_2}=1$ , the array *)
  Sample = Array[Finverse[Random[],  $\theta$ , b1, b2] &,
    200];
  (* is a pseudorandom sample of size 200
drawn from the generalized Laplace
distribution where Random[] is a
psuedorandom sample drawn from U(0,1). *)
  (* From equation 15 in the report,
 $\hat{\theta} = Q_{c b_1}(\text{Sample}) =$  *)
  MLEs[[j]] = Quantile[Sample, c b1];
  (* where the Quantile function returns
the  $c*b_1$  quantile from the Sample array. *)
  Clear[ $\theta$ , b1, b2];
  , {j, 1000}];

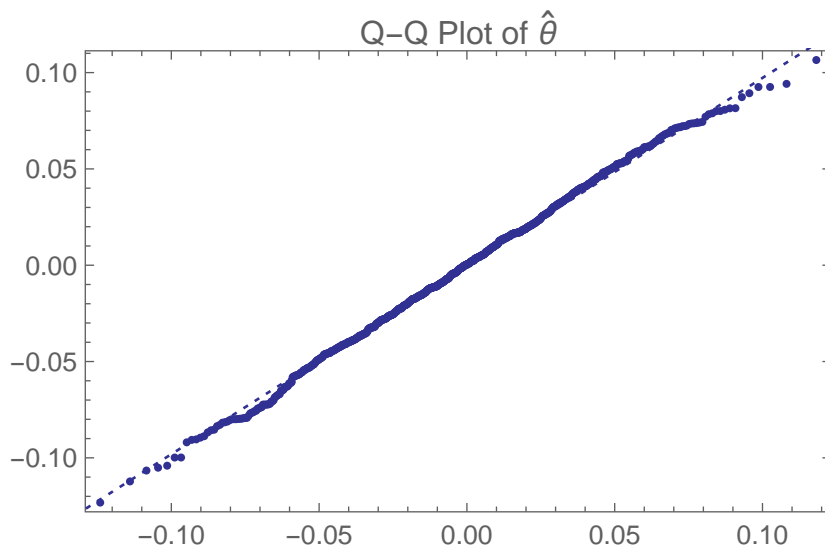
(* An overlaid graphic,
with a histogram of the above MLEs and a
plot of the pdf of the normal distribution
associated with the asymtotic limit of the
 $\hat{\theta}$  is created by the following command: *)
Show[Histogram[MLEs, 25, "PDF"],
  Plot[PDF[NormalDistribution[0, (.25 / 200) ^ .5],
    x], {x, -.3, .3}],
  PlotLabel  $\rightarrow$  "Histogram of  $\hat{\theta}$  versus Normal pdf"]

```



(\* The following command `QuantilePlot` produces a Q-Q Plot of the 1000 estimates of  $\hat{\theta}$  stored in the MLEs array from the above cell. \*)

```
QuantilePlot[MLEs, PlotLabel -> "Q-Q Plot of  $\hat{\theta}$ "]
```



### 3.5 Appendix

We are going to show that  $E_{\theta_0} Y_{1,\theta_0} = 0$ , using Conditions 3, 7, and 8. Since  $f_\theta(x)$  is log-concave in  $\theta$  by Condition 3,  $f_\theta(x)$  is absolutely continuous in  $\theta$ , whence

$$f_{\theta_2}(x) - f_{\theta_1}(x) = \int_{\theta_1}^{\theta_2} \frac{\partial^-}{\partial \theta} f_\theta(x) d\theta \quad (52)$$

for any  $\theta_1$  and  $\theta_2$  in  $\Theta$  such that  $\theta_1 < \theta_2$ .

Then

$$\begin{aligned} 0 &= \int_{\mathbb{R}} f_{\theta_2}(x) dx - \int_{\mathbb{R}} f_{\theta_1}(x) dx \\ &= \int_{\mathbb{R}} (f_{\theta_2}(x) - f_{\theta_1}(x)) dx \\ &= \int_{\mathbb{R}} dx \int_{\theta_1}^{\theta_2} d\theta \frac{\partial^-}{\partial \theta} f_\theta(x) \quad (\text{by (52)}) \\ &= \int_{\theta_1}^{\theta_2} d\theta \int_{\mathbb{R}} dx \frac{\partial^-}{\partial \theta} f_\theta(x) \quad (\text{by Condition 7 and Fubini's Theorem}) \\ &= \int_{\theta_1}^{\theta_2} d\theta G(\theta). \end{aligned}$$

Because the integral of the function  $G(\theta)$  is zero over every interval  $(\theta_1, \theta_2)$ , and since  $G(\theta)$  is continuous by Condition 8,  $G(\theta) = 0$  zero everywhere. Then, the expectation of  $E_{\theta_0} Y_{1,\theta_0}$  can be shown to be zero as follows,

$$\begin{aligned} E_{\theta_0} Y_{1,\theta_0} &= \int_x \frac{\partial^-}{\partial \theta} (\ln(f_\theta(x))) \Big|_{\theta_0} f_{\theta_0}(x) dx \\ &= \int_x \frac{1}{f_{\theta_0}(x)} \frac{\partial^-}{\partial \theta} (f_{\theta_0}(x)) \Big|_{\theta_0} f_{\theta_0}(x) dx \\ &= \int_x \frac{\partial^-}{\partial \theta} (f_\theta(x)) \Big|_{\theta_0} dx = G(\theta_0) = 0. \end{aligned}$$

## 4 Bibliography

- Y.S. Chow and H Teicher. *Probability Theory Independence, Interchangeability, and Martingales (Second Edition)*. Springer-Verlang, New York, 1962.
- Harald Cramér. *Stochastic processes. Distribution (Probability theory) Random variables*. Cambridge University Press, Cambridge, 1978.
- Grace L. Yang David Pollard, Erik Torgersen. *Festschrift for Lucien Le Cam*. Springer-Verlang, New York, 1997.
- J.L. Doob. Probability and statistics. *Transactions of the American Mathematical Society*, 36:759–775, 1934.
- R.A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London*, 222:309–368, 1922.
- R.A. Fisher. Theory of statistical estimation. *Proceedings of the Cambridge Philosophical Society*, 22:700–725, 1925.
- Harold Hotelling. The consistency and ultimate distribution of optimum statistics. *Transactions of the American Mathematical Society*, 32:847–859, 1930.
- Lucien LeCam. *Asymptotic Methods in Statistical Decision Theory*. Springer-Verlang, 1986.
- E.L. Lehmann. *Elements of Large Sample Theory*. Springer-Verlang, New York, 2004.
- Iosif Pinelis. Fractional sums and integrals of  $r$ -concave tails and applications to comparison probability inequalities. *Contemporary Mathematics*, 234: 149–168, 1999.
- David Pollard. *Asymptopia*. Manuscript in Progress, 2005.
- T.J. Kozubowski S. Kotz and K. Podgoński. Maximum likelihood estimation of asymmetric laplace parameters. *Ann. Inst. Statist. Math.*, 54:816–826, 2002.
- Stephen Stigler. The epic story of maximum likelihood. *ARXIV math.ME*, 2008.