TOPIC. Limit theorems. Here we use characteristic functions to prove versions of the weak law of large numbers and the central limit theorem, and then use these results to prove Cramer's theorem about the asymptotic distribution of maximum likelihood estimators.

We write ϕ_X for the characteristic function of a random variable X. We will need to know how $\phi_X(t)$ behaves as $t \to 0$. To see why, consider the characteristic function of S_n/n, where S_n is the sum of iid random variables $X_1, \ldots, X_n$. Since $\phi_{S_n/n}(\tau) = E(e^{i\tau S_n/n}) = \left(\phi_{X_1}(\tau/n)\right)^n$, the behavior of $\phi_{X_1}(t)$ as $t \to 0$ determines the limit behavior of $\phi_{S_n/n}(\tau)$ for each $\tau \in \mathbb{R}$, and hence the limiting distribution of S_n/n, as $n \to \infty$.

Smoothness of characteristic functions. The random variable $X = 1$ has characteristic function $\phi_X(t) = e^{it}$. The following lemma bounds the difference between $\phi_X(t)$ and its nth-order Taylor expansion $e_n(t)$ about $t = 0$; this result will be used later derive a similar bound for general X (see (4), below).

Lemma 1. *For real t and $n = 0, 1, \ldots,$*

$$\left| e^{it} - 1 - it - \frac{(it)^2}{2} - \cdots - \frac{(it)^n}{n!} \right| \leq \min\left(\frac{|t|^{n+1}}{(n+1)!}, \frac{2|t|^n}{n!} \right). \qquad (1)$$

Proof Put $\Delta_n(t) = e^{it} - \sum_{k=0}^n (it)^k/k!$ and suppose $t \geq 0$. I claim

$$|\Delta_n(t)| \leq \frac{t^{n+1}}{(n+1)!}. \qquad (2)$$

This holds for $n = 0$ because

$$\Delta_0(t) = e^{it} - e^{i0} \underset{\text{FTC}}{=} \int_0^t ie^{i\tau}\, d\tau \implies |\Delta_0(t)| \leq \int_0^t 1\, d\tau = t,$$

the second equality holding by the Fundamental Theorem of Calculus for complex-valued functions of a real variable (see Exercise 12.6). For $n \geq 1$ we have $\Delta_n'(t) = i\Delta_{n-1}(t)$ and $\Delta_n(0) = 0$, so

$$|\Delta_n(t)| \underset{\text{FTC}}{=} \left| i \int_0^t \Delta_{n-1}(\tau)\, d\tau \right| \leq \int_0^t |\Delta_{n-1}(\tau)|\, d\tau;$$

(2) follows by induction on n because $\int_0^t \tau^n/n!\, d\tau = t^{n+1}/(n+1)!$.

$$\Delta_n(t) = e^{it} - e_n(it). \qquad (1): \ |\Delta_n(t)| \leq \min\left(|t|^{n+1}/(n+1)!, 2|t|^n/n\right).$$

Now (2) for $n - 1$ implies

$$|\Delta_n(t)| = |\Delta_{n-1}(t) - i^n t^n/n!| \leq |\Delta_{n-1}(t)| + \frac{t^n}{n!} \leq 2\frac{t^n}{n!}.$$

This and (2) for n give (1). For $t < 0$, $\Delta_n(t)$ is the complex conjugate of $\Delta_n(-t)$, so $|\Delta_n(t)| = |\Delta_n(-t)|$. ∎

There is a general principle that says that the lighter are the tails of the distribution of a random variable X, the smoother is its characteristic function ϕ. The following is one instance of this principle — it states that the more moments X has, the more derivatives ϕ has.

Theorem 1. *Let X be a random variable with characteristic function ϕ, and suppose that $E(|X|^n) < \infty$ for some positive integer n. Then ϕ is n-times continuously differentiable with*

$$\phi^{(k)}(t) = i^k E(X^k e^{itX}) \qquad (3)$$

for all $k = 1, \ldots, n$ and $t \in \mathbb{R}$. Moreover

$$\left| \phi(t) - \sum_{k=0}^n \frac{E(X^k)}{k!}(it)^k \right| \leq E\left[\min\left(\frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!} \right) \right] \qquad (4)$$

for all t; the bound on the right-hand side of (4) is $o(|t|^n)$ as $t \to 0$.

Proof Fix $t \in \mathbb{R}$ and note that for nonzero real h,

$$\frac{\phi(t+h) - \phi(t)}{h} = E(Q_h)$$

where

$$Q_h := \frac{e^{i(t+h)X} - e^{itX}}{h} = \frac{e^{ihX} - 1}{h} e^{itX}.$$

As $h \to 0$, $Q_h \to iXe^{itX}$. Moreover by (1), $|Q_h| \leq |X|$ for all h. Since $|X|$ is integrable, the DCT gives $\phi'(t) = \lim_h E(Q_h) = E(iXe^{itX})$. This proves (3) for $k = 1$; the general case follows by induction on k.

(4): $\left|\phi(t) - \sum_{k=0}^{n} \frac{E(X^k)}{k!}(it)^k\right| \le E\left[\min\left(\frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!}\right)\right] = o(t^n).$

For $k \le n$, $\phi^{(k)}(t) = i^k E(X^k e^{itX})$ is continuous in t by the DCT, since $|X^k e^{itX}| \le |X|^k$ for all t and $E(|X|^k) < \infty$. Inequality (4) follows by writing

$$\left|\phi(t) - \sum_{k=0}^{n} \frac{E(X^k)}{k!}(it)^k\right| = |E(\Delta_n(tX))| \le E\left(|\Delta_n(tX)|\right)$$

and then using (1) to bound the term on the right. Finally, the DCT implies that $|t|^n E\left(|X|^n \min(|tX|, 2)\right) = o(|t|^n)$ as $t \to 0$. ∎

There is a partial converse to Theorem 1. If ϕ has an nth derivative, then X^n is integrable provided n is even; however, it may not be integrable if n is odd. See Exercises 4 and 12. Two special cases of (4) are worth emphasizing:

$$\phi(t) = 1 + i\mu t + o(t) \quad \text{as } t \to 0 \tag{4_1}$$

if X has a finite mean μ, and

$$\phi(t) = 1 - t^2 \sigma^2/2 + o(t^2) \quad \text{as } t \to 0 \tag{4_2}$$

if X has mean 0 and finite variance σ^2. We will use these estimates to prove the weak law of large numbers and the central limit theorem.

The weak law of large numbers.

Theorem 2. *Let $X_1, X_2, \ldots$ be iid integrable random variables with mean μ. Put $S_n = X_1 + \cdots + X_n$ for $n \ge 1$. As $n \to \infty$*

$$S_n/n \to_P \mu. \tag{5}$$

Proof $S_n/n \to_P \mu$ means $P[|S_n/n - \mu| \ge \epsilon] \to 0$ as $n \to \infty$, for each $\epsilon > 0$. It is easy to see that this is equivalent to

$$P[S_n/n \le x] \to \begin{cases} 1, & \text{if } x > \mu, \\ 0, & \text{if } x < \mu, \end{cases} \qquad \overline{\quad\mu-\epsilon\quad \mu \quad \mu+\epsilon\quad}$$

and thus to $S_n/n \to_{\mathcal{D}} \mu$. By the continuity theorem (Theorem 14.10)

it suffices to show that

$$\phi_{S_n/n}(t) := E(e^{itS_n/n}) \to e^{i\mu t}$$

for each $t \in \mathbb{R}$. Fix t and set $X = X_1$. Since the X_k's are iid

$$\phi_{S_n/n}(t) = \phi_{S_n}(t/n) = \prod_{k=1}^{n} \phi_{X_k}(t/n) = \left(\phi_X(t/n)\right)^n$$

(Theorem 13.2), and since X is integrable with mean μ

$$\phi_X(\tau) = 1 + i\tau\mu + o(\tau) \tag{6}$$

as $\tau \to 0$ (equation (4_1)). Thus as $n \to \infty$

$$\phi_{S_n/n}(t) = \left(1 + \frac{it\mu + o(1)}{n}\right)^n \to e^{it\mu}$$

by (14.18): $z_n \to z$ in $\mathbb{C} \implies (1 + z_n/n)^n \to e^z$. ∎

As the proof shows, $X := X_1$ doesn't have to be integrable in order for $S_n/\sqrt{n}$ to converge in probability to a real number μ; it is enough to have (6), which is equivalent to ϕ_X being differentiable at 0 with derivative $\phi_X'(0) = i\mu$. This sufficient condition is also necessary; see Theorem 8, below.

The central limit theorem. Let $X_1, X_2, \ldots$ be iid random variables with finite mean μ and variance σ^2. Put $S_n = X_1 + \cdots + X_n$ for $n \ge 1$. We are going to study the limiting distribution of

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \left(\frac{X_k - \mu}{\sigma}\right).$$

There is no loss of generality in supposing $\mu = 0$ and $\sigma^2 = 1$. Write $\mathfrak{N}$ for the standard normal distribution and $\mathfrak{n}$ for its density:

$$\mathfrak{N}(B) = \int_B \mathfrak{n}(x)\, dx \text{ for sets } B; \quad \mathfrak{n}(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}. \tag{7}$$

Theorem 3 (The global central limit theorem). *Let X_1, X_2, ... be iid random variables with mean 0 and variance 1. Put $S_n = X_1 + \cdots + X_n$ for $n \geq 1$. As $n \to \infty$*

$$S_n/\sqrt{n} \to_{\mathcal{D}} Z \sim \mathfrak{N}. \tag{8}$$

Proof Put $X = X_1$. Since X has mean 0 and variance 1, its characteristic function satisfies

$$\phi_X(\tau) = 1 - \tau^2/2 + o(\tau^2) \tag{9}$$

as $\tau \to 0$; see (4_2). Thus for each $t \in \mathbb{R}$

$$\phi_{S_n/\sqrt{n}}(t) = \left(\phi_X(t/\sqrt{n})\right)^n = \left(1 - \frac{t^2 + o(1)}{2n}\right)^n \to e^{-t^2/2}; \tag{10}$$

since $\exp(-t^2/2)$ is the characteristic function of Z, (8) follows from the continuity theorem. ∎

For sums of iid random variables, (8) implies that the summands must have a finite mean and variance, equal to 0 and 1 respectively; see Exercise 12.

We have seen that convergence in distribution does not imply convergence of densities, even when they exist (Example 14.6). The following result strengthens the CLT by showing that under certain circumstances, $S_n/\sqrt{n}$ has a density that converges uniformly to the density of the standard normal distribution.

Theorem 4 (The local central limit theorem for densities). *Suppose X_1, X_2, ... are iid random variables with mean 0, variance 1, and characteristic function ϕ. If*

$|\phi|^k$ is integrable for some positive integer k, and (11)

$b(\delta) := \sup\{ |\phi(\tau)| : |\tau| \geq \delta \} < 1$ for all $\delta > 0$, (12)

then $S_n/\sqrt{n}$ has a bounded continuous density f_n for all $n \geq k$ and as $n \to \infty$,

$$\sup_{x \in \mathbb{R}} |f_n(x) - \mathfrak{n}(x)| \to 0 \tag{13}$$

$$\int_{-\infty}^{\infty} |f_n(x) - \mathfrak{n}(x)| \, dx \to 0. \tag{14}$$

Proof $S_n/\sqrt{n}$ has characteristic function $\phi_n(t) := \left(\phi(t/\sqrt{n})\right)^n$. For $n \geq k$, the change of variables $\tau = t/\sqrt{n}$ $(\Longleftrightarrow t = \sqrt{n}\,\tau)$ and (11) give

$$\int_{-\infty}^{\infty} |\phi_n(t)| \, dt = \sqrt{n} \int_{-\infty}^{\infty} |\phi(\tau)|^n \, d\tau \leq \sqrt{n} \int_{-\infty}^{\infty} |\phi(\tau)|^k \, d\tau < \infty;$$

by Theorem 13.4, $S_n/\sqrt{n}$ has the continuous bounded density

$$f_n(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi_n(t) \, dt.$$

The density $\mathfrak{n}$ and characteristic function $\psi(t) := \exp(-t^2/2)$ of the standard normal distribution are related by an analogous formula:

$$\mathfrak{n}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \psi(t) \, dt.$$

Thus for any $\delta > 0$

$$\sup_{x \in \mathbb{R}} |f_n(x) - \mathfrak{n}(x)| = \sup_{x \in \mathbb{R}} \left| \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \left(\phi_n(t) - \psi(t)\right) dt \right|$$

$$\leq \frac{1}{2\pi} \int_{-\infty}^{\infty} |\phi_n(t) - \psi(t)| \, dt \leq A_n(\delta) + B_n(\delta) + C_n(\delta)$$

where

$$A_n(\delta) = \int_{t:|t| \leq \sqrt{n}\delta} |\phi_n(t) - \psi(t)| \, dt,$$

$$B_n(\delta) = \int_{t:|t| > \sqrt{n}\delta} |\phi_n(t)| \, dt, \quad C_n(\delta) = \int_{t:|t| > \sqrt{n}\delta} \psi(t) \, dt.$$

I am going to show that there exists a δ such that $A_n(\delta) \to 0$, while $B_n(\delta) \to 0$ and $C_n(\delta) \to 0$ for any δ; this will give (13). The idea is that, as (10) suggests, for a fixed (large) n, the approximation $\phi_n(t) \approx \psi(t)$ is very good for t's that are sufficiently small compared to $\sqrt{n}$; for the remaining t's, (11) and (12) ensure that, like $\psi(t)$, $\phi_n(t)$ is negligibly small.

(11): $\int_{-\infty}^{\infty} |\phi(t)|^k \, dt < \infty$ (12): $b(\delta) := \sup\{\, |\phi(\tau)| : |\tau| \geq \delta \,\} < 1$

$\phi_n(t) := \phi_{S_n/\sqrt{n}}(t) = \big(\phi(t/\sqrt{n})\big)^n$ $\psi(t) := \exp(-t^2/2).$

$A_n(\delta) = \int_{t:|t|\leq\sqrt{n}\delta} |\phi_n(t) - \psi(t)| \, dt.$ (10): $\phi_n(t) \to \psi(t).$

$B_n(\delta) = \int_{t:|t|>\sqrt{n}\delta} |\phi_n(t)| \, dt.$ $C_n(\delta) = \int_{t:|t|>\sqrt{n}\delta} |\psi(t)| \, dt.$

In regard to $A_n(\delta)$, recall that $\phi(\tau) = 1 - \tau^2/2 + o(\tau^2)$ as $\tau \to 0$ by (4_2). Using Exercise 13, choose and fix $\delta > 0$ small enough to make

$$|\phi(\tau)| \leq 1 - \tau^2/4 \leq e^{-\tau^2/4} \text{ for all } |\tau| \leq \delta \tag{15}$$

$$\implies |\phi_n(t)| = |\phi(t/\sqrt{n})|^n \leq e^{-t^2/4} \text{ for all } |t| \leq \sqrt{n}\delta.$$

Thus $a_n(t) := |\phi_n(t) - \psi(t)| I_{[0,\sqrt{n}\delta]}(|t|)$ is dominated by $2\exp(-t^2/4)$. Since $a_n(t) \to 0$ for each $t \in \mathbb{R}$ by (10) and since $\int_0^\infty \exp(-t^2/4) \, dt < \infty$, we have $A_n(\delta) = \int_{-\infty}^\infty a_n(t) \, dt \to 0$ by the DCT.

In regard to $B_n(\delta)$, we have for $n \geq k$

$$B_n(\delta) = \int_{t:|t|>\sqrt{n}\delta} |\phi(t/\sqrt{n})|^n \, dt = \sqrt{n} \int_{\tau:|\tau|>\delta} |\phi(\tau)|^n \, d\tau$$

$$\leq \sqrt{n}(b(\delta))^{n-k} \int_{-\infty}^\infty |\phi(\tau)|^k \, d\tau \underset{\text{by (11)}}{=} O\big(\sqrt{n}(b(\delta))^n\big),$$

so by (12) $B_n(\delta)$ goes to 0, exponentially fast. So does $C_n(\delta)$, since

$$C_n(\delta) = 2 \int_{t=\sqrt{n}\delta}^\infty e^{-t^2/2} \, dt \leq \frac{2e^{-n\delta^2/2}}{\sqrt{n}\delta}; \tag{16}$$

see Exercise 14. That completes the proof of (13). (13) implies (14) by Theorem 14.8. ∎

Remarks. (A): By Theorem 14.8,

$$S_n/\sqrt{n} \to_s Z; \tag{17}$$

this is much stronger that (8). (B): It can be shown that (11) im-

plies (12), so (12) is actually superfluous. (C): A stronger version of (13) holds; see Theorem 9 below.

Example 1. Suppose X_1, X_2, ... are iid as $X := Y - 1$ where Y has the standard exponential distribution. X has mean 0, variance 1, and characteristic function $\phi(t) = e^{-it}/(1 - it)$. (11) and (12) hold since

$\int_{-\infty}^\infty |\phi(t)|^2 \, dt = \int_{-\infty}^\infty 1/(1 + t^2) \, dt < \infty$, and

$|\phi(t)| \leq 1/\sqrt{1 + \delta^2} < 1$ for $|t| \geq \delta > 0$.

Consequently Theorem 4 applies. Since $\sum_{k=1}^n X_k$ has a gamma distribution, it follows that the corresponding gamma densities, suitably scaled and centered, converge uniformly to the standard normal density. We proved this before, in Example 14.7, using Stirling's formula for the Gamma function and a different method. That argument can now be turned around, to give a proof of Stirling's formula; see Exercises 16 and 17, and Exercises 14.25 and 11 for other approaches. •

Similar methods (see Exercise 24) can be used to prove a local limit theorem for probability mass functions, in the case where the summands take values in a lattice, i.e., a set of the form $L_{a,h} := \{\, a + kh : k = 0, \pm 1, \pm 2, \dots \,\}$; h is called the span of $L_{a,h}$. For example, a Binomial variable takes values in $L_{0,1}$.

Theorem 5 (The local central limit theorem for probability mass functions). *Let X_1, X_2, ... be iid lattice random variables with mean 0, variance 1, and characteristic function ϕ. Put $h = 2\pi/\inf\{t > 0 : |\phi(t)| = 1\}$. For each n the distribution of $S_n/\sqrt{n}$ is supported by a lattice L_n of the form $\{a_n + kh_n : k = 0, \pm 1, \pm 2, \dots\}$ where $h_n = h/\sqrt{n}$, and*

$$\sup_{x \in L_n} \left| \frac{P[S_n/\sqrt{n} = x]}{h_n} - \mathfrak{n}(x) \right| \to 0 \tag{18}$$

as $n \to \infty$.

The asymptotic distribution of the maximum likelihood estimator. Let X_1, X_2, ... be independent random variables, each having density f_{θ_0} with respect to a given measure ν for some unknown θ_0 belonging to a given subinterval Θ of $\mathbb{R}$. Let $\hat{\theta}_n$ be the maximum likelihood estimator of θ_0 based on $X_1,\ldots,X_n$: $\hat{\theta}_n$ is that (random) $\theta \in \Theta$ which maximizes the likelihood function

$$L_n(\theta) := \prod_{m=1}^{n} f_\theta(X_m).$$

It can be shown that under appropriate regularity conditions $\hat{\theta}_n$ exists and tends to θ_0 with probability one as $n \to \infty$. Here we will use the central limit theorem and the weak law of large numbers to obtain the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$.

We begin with a heuristic version of the argument. Under appropriate conditions $\hat{\theta}_n$ will satisfy the equation $S_n(\hat{\theta}_n) = 0$, where

$$S_n(\theta) := \frac{\partial}{\partial \theta} \log L_n(\theta) = \sum_{m=1}^{n} s(\theta; X_m)$$

with $s(\theta; x) := \frac{\partial}{\partial \theta} \log f_\theta(x)$. For large n, we expect $\hat{\theta}_n$ to be close to θ_0 with high probability, in which case the slope of the secant from $\bigl(\theta_0, S_n(\theta_0)\bigr)$ to $\bigl(\hat{\theta}_n, S_n(\hat{\theta}_n)\bigr) = (\hat{\theta}_n, 0)$ ought to be close to the slope of the tangent line to S_n at θ_0:

$$\frac{0 - S_n(\theta_0)}{\hat{\theta}_n - \theta_0} \approx S_n'(\theta_0)$$
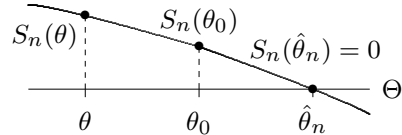
whence

$$\hat{\theta}_n - \theta_0 \approx -\frac{S_n(\theta_0)}{S_n'(\theta_0)} = -\frac{\sum_{m=1}^{n} s(\theta_0; X_m)}{\sum_{m=1}^{n} s'(\theta_0; X_m)}. \qquad (19)$$

The numerator and denominator of the ratio on the right-hand side of (19) are sums of iid random variables; how the sums behave for large n

depends on the means and variances of the summands. To learn about these, put $\Lambda_\theta(x) = \log f_\theta(x)$ and differentiate the equation

$$1 = \int f_\theta(x)\, \nu(dx) = \int e^{\Lambda_\theta(x)}\, \nu(dx)$$

twice with respect to θ to get first

$$0 = \frac{\partial}{\partial \theta} 1 = \frac{\partial}{\partial \theta} \int e^{\Lambda_\theta(x)}\, \nu(dx) \overset{*}{=} \int \frac{\partial}{\partial \theta} e^{\Lambda_\theta(x)}\, \nu(dx)$$

$$= \int \underbrace{e^{\Lambda_\theta(x)}}_{=f_\theta(x)} \underbrace{\Lambda_\theta'(x)}_{=s(\theta;x)}\, \nu(dx) = E_\theta\bigl(s(\theta; X_1)\bigr)$$

and then

$$0 = \frac{\partial}{\partial \theta} 0 = \frac{\partial}{\partial \theta} \int e^{\Lambda_\theta(x)} \Lambda_\theta'(x)\, \nu(dx) \overset{**}{=} \int \frac{\partial}{\partial \theta}\bigl(e^{\Lambda_\theta(x)} \Lambda_\theta'(x)\bigr)\, \nu(dx)$$

$$= \int e^{\Lambda_\theta(x)}\bigl((\Lambda_\theta'(x))^2 + \Lambda_\theta''(x)\bigr)\, \nu(dx)$$

$$= E_\theta\bigl(s^2(\theta; X_1)\bigr) + E_\theta\bigl(s'(\theta; X_1)\bigr),$$

assuming that the interchanges of differentiation and integration at $*$ and $**$ are legitimate. This gives

$$E_{\theta_0} s(\theta_0; X_1) = 0,$$

$$\mathrm{Var}_{\theta_0}\bigl(s(\theta_0; X_1)\bigr) = -E_{\theta_0} s'(\theta_0; X_1) := I(\theta_0).$$

Now rewrite (19) as $\sqrt{n}(\hat{\theta}_n - \theta_0) \approx N_n/D_n$ where

$$N_n = \frac{1}{\sqrt{n}} \sum_{m=1}^{n} s(\theta_0; X_m) \quad \text{and} \quad D_n = -\frac{1}{n} \sum_{m=1}^{n} s'(\theta_0; X_m).$$

By the central limit theorem, N_n will for large n be distributed nearly like $\sqrt{I(\theta_0)}Z$ with $Z \sim \mathfrak{N}$, and by the weak law of large numbers D_n will for large n be nearly equal to $I(\theta_0)$. It thus appears that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to_{\mathcal{D}} Z/\sqrt{I(\theta_0)} \qquad (20)$$

as $n \to \infty$. I am going to show that (20) is correct under assumptions C0–C4 below.

C0 (Framework) X_1, X_2, ... are iid random variables with density f_{θ_0} for some $\theta_0 \in \Theta$. Θ is a subinterval of $\mathbb{R}$ (perhaps all of $\mathbb{R}$).

C1 (Smoothness) For each $x \in \mathcal{X}$, $f_\theta(x)$ is strictly positive and twice continuously differentiable with respect to θ.

C2 (Natural moments) For each $\theta \in \Theta$, $E_\theta s(\theta; X_1) = 0$ and $0 < I(\theta) := \text{Var}_\theta\big(s(\theta; X_1)\big) = -E_\theta s'(\theta; X_1) < \infty$. Here

$$s(\theta; x) := \frac{\partial}{\partial \theta} \log f_\theta(x) \text{ and } s'(\theta; x) = \frac{\partial}{\partial \theta} s(\theta; x).$$

C3 (Integrability) For each $\theta \in \Theta$, there exists a number $\delta_\theta > 0$ such that $E_\theta\big(\sup_{\tau \in \Theta : |\tau - \theta| \le \delta_\theta} |s'(\tau; X_1)|\big) < \infty$.

C4 (Consistency) $\hat{\theta}_n$ is consistent in probability, i.e., for each $\theta \in \Theta$, $\lim_n P_\theta\big[|\hat{\theta}_n - \theta| \ge \epsilon\big] = 0$ for each $\epsilon > 0$.

Assumption **C2** can be verified by direct calculation or by using standard techniques to justify the aforementioned interchanges of differentiation and integration at $*$ and $**$. Assumption **C3** is only slightly stronger than the requirement in **C2** that $s'(\theta; X_1)$ be integrable. Assumption **C4** can be checked by direct analysis of $\hat{\theta}_n$ or by using general theorems about the consistency of MLE's.

Theorem 6 (Cramer). *If conditions* **C0**–**C4** *hold, then for each $\theta_0 \in$ int(Θ) the maximum likelihood estimator $\hat{\theta}_n$ satisfies*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to_{\mathcal{D}} Z/\sqrt{I(\theta_0)} \sim \mathfrak{N}\big(0, 1/I(\theta_0)\big) \tag{21}$$

under P_{θ_0}.

Warning: (21) may fail if θ_0 is an endpoint of Θ; see Exercise 26 for an example.

Proof of Theorem 6. Suppose $\theta_0 \in$ int(Θ). By the consistency condition **C4** we have $\hat{\theta}_n \in$ int(Θ) with P_{θ_0}-probability tending to one; to avoid obscuring the proof suppose for the time being that

C1: $f_\theta > 0$ is twice continuously differentiable in θ.
C2: $E_\theta s(\theta; X_1) = 0$; $0 < I(\theta) := \text{Var}_\theta\big(s(\theta; X_1)\big) = -E_\theta s'(\theta; X_1) < \infty$.
$S_n(\theta) = \frac{\partial}{\partial \theta} \log L_n(\theta) = \sum_{m=1}^n s(\theta; X_m)$. $\quad s(\theta; x) = \frac{\partial}{\partial \theta} \log f_\theta(x)$.

$\hat{\theta}_n(\omega) \in$ int(Θ) for all ω. Since the log of the likelihood function is differentiable by **C1** and has an interior maximum at $\hat{\theta}_n$, its derivative there must be zero; thus $S_n(\hat{\theta}_n) = 0$. The mean value theorem gives

$$\frac{0 - S_n(\theta_0)}{\hat{\theta}_n - \theta_0} = \frac{S_n(\hat{\theta}_n) - S_n(\theta_0)}{\hat{\theta}_n - \theta_0} = S_n'(\hat{\theta}_n^*)$$

where $\hat{\theta}_n^*$ lies between θ_0 and $\hat{\theta}_n$; this is the rigorous version of (19). It follows that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{S_n(\theta_0)/\sqrt{n}}{-S_n'(\hat{\theta}_n^*)/n} := \frac{N_n}{D_n}. \tag{22}$$

Since the iid variables $s(\theta_0; X_m)$ have mean 0 and finite variance $I(\theta_0)$ by **C2**, the central limit theorem gives $N_n = \sum_{m=1}^n s(\theta_0; X_m)/\sqrt{n} \to_{\mathcal{D}} \sqrt{I(\theta_0)}Z$ with $Z \sim \mathfrak{N}$. I claim

$$D_n = -\frac{1}{n} \sum_{m=1}^n s'(\hat{\theta}_n^*; X_m) \to_{P_{\theta_0}} I(\theta_0). \tag{23}$$

Since $I(\theta_0) > 0$ by **C2**, part (iii) of Lemma 2 below implies that $N_n/D_n \to_{\mathcal{D}} \sqrt{I(\theta_0)}Z/I(\theta_0) = Z/\sqrt{I(\theta_0)}$.

To verify (23) note that the iid variables $s'(\theta_0; X_m)$ have mean $-I(\theta_0)$ by **C2**; the weak law of large numbers gives

$$-\frac{1}{n} \sum_{m=1}^n s'(\theta_0; X_m) \to_{P_{\theta_0}} I(\theta_0).$$

C1: $f_\theta > 0$ is twice continuously differentiable in θ.

C3: $E_\theta\big(\sup_{\tau\in\Theta:|\tau-\theta|\leq\delta_\theta}|s'(\tau;X_1)|\big) < \infty$ for some $\delta_\theta > 0$.

C4: $\hat\theta_n \to_{P_\theta} \theta$ for each θ.

$$S_n(\theta) = \frac{\partial}{\partial\theta}\log L_n(\theta) = \sum_{m=1}^n s(\theta;X_m). \qquad s(\theta;x) = \frac{\partial}{\partial\theta}\log f_\theta(x).$$

It thus suffices to show that

$$A_n := \frac{1}{n}\sum_{m=1}^n \big(s'(\hat\theta_n^*;X_m) - s'(\theta_0;X_m)\big) \to_{P_{\theta_0}} 0. \tag{24}$$

For this let $\epsilon > 0$ be given and suppose δ is a positive number, to be specified further later on. Since $\hat\theta_n^*$ lies between θ_0 and $\hat\theta_n$, if $\hat\theta_n^*$ differs from θ_0 by more than δ so does $\hat\theta_n$. It follows that

$$P_{\theta_0}[\,|A_n| \geq \epsilon\,] \leq P_{\theta_0}[\,|\hat\theta_n - \theta_0| \geq \delta\,] + P_{\theta_0}\Big[\frac{1}{n}\sum_{m=1}^n Y_{m,\delta} \geq \epsilon\Big]$$

$$:= I_n + II_n$$

where

$$Y_{m,\delta} := \sup\{\,|s'(\tau;X_m) - s'(\theta_0;X_m)| : \tau \in \Theta,\ |\tau - \theta_0| \leq \delta\,\}.$$

The smoothness condition C1 implies that $Y_{1,\delta} \downarrow 0$ as $\delta \downarrow 0$. Moreover $Y_{1,\delta_{\theta_0}}$ is integrable by condition C3. The DCT thus implies that $m(\delta) := E_{\theta_0}Y_{1,\delta} \to 0$ as $\delta \to 0$. Choose and fix δ such that $m(\delta) \leq \epsilon/2$. Since the $Y_{m,\delta}$'s are iid with mean $m(\delta)$, the weak law of large numbers implies that $\frac{1}{n}\sum_{m=1}^n Y_{m,\delta} \to_{P_{\theta_0}} m(\delta) < \epsilon$, and thus that $II_n \to 0$. Since $I_n \to 0$ by C4, (24) holds.

That completes the proof under the supplementary assumption that $\hat\theta_n(\omega) \in \text{int}(\Theta)$ for all ω. To handle the general case, put $G_n := \{\hat\theta_n \in \text{int}(\Theta)\}$, set $\hat\theta_n^* = \theta_0$ on G_n^c, and note that $\sqrt{n}(\hat\theta_n - \theta_0) = N_n/D_n$ on G_n. Since $N_n/D_n \to_\mathcal{D} Z/\sqrt{I(\theta_0)}$ as before and since C4 and $\theta_0 \in \text{int}(\Theta)$ entail $P_{\theta_0}[G_n] \to 1$, we get $\sqrt{n}(\hat\theta_n - \theta_0) \to_\mathcal{D} Z/\sqrt{I(\theta_0)}$. ∎

For future reference it is worth extracting two facts from the proof:

Theorem 7. *Let $\theta_0 \in \Theta$. If conditions C0–C2 hold, then under P_{θ_0}*

$$S_n(\theta_0)/\sqrt{n} \to_\mathcal{D} \sqrt{I(\theta_0)}Z \tag{25}$$

with $Z \sim \mathfrak{N}$. If in addition condition C3 holds and $T_n \to_{P_{\theta_0}} \theta_0$, then

$$S_n'(T_n)/n \to_{P_{\theta_0}} -I(\theta_0). \tag{26}$$

The following lemma was referred to in the proof of Cramer's theorem:

Lemma 2. *Suppose X_n and Y_n are real random variables such that $X_n \to_\mathcal{D} X$ and $Y_n \to_P c$, where c is a finite constant. Then*

$$(X_n, Y_n) \to_\mathcal{D} (X, c) \ \text{in}\ \mathbb{R}^2. \tag{27}$$

In particular

  (i) $X_n + Y_n \to_\mathcal{D} X + c$,

  (ii) $X_n Y_n \to_\mathcal{D} cX$, and

  (iii) $X_n/Y_n \to_\mathcal{D} X/c$ provided $c \neq 0$.

It doesn't matter how X_n/Y_n is defined in (iii) when $Y_n = 0$, since $Y_n \to_P c \neq 0$ implies $P[Y_n = 0] \to 0$.

Proof of Lemma 2. To get (27) it suffices to show that $E\big(g(X_n, Y_n)\big) \to E\big(g(X, c)\big)$ when $g:\mathbb{R}^2 \to \mathbb{R}$ is bounded, say by C, and Lipschitz continuous, say with Lipschitz norm L. For this, write

$$\big|E\big(g(X_n, Y_n)\big) - E\big(g(X, c)\big)\big| \leq I_n + II_n$$

where $I_n = \big|E\big(g(X_n, Y_n)\big) - E\big(g(X_n, c)\big)\big|$ and $II_n = \big|E\big(g(X_n, c)\big) - E\big(g(X, c)\big)\big|$. One has $I_n \to 0$ since $I_n \leq E\big(\min(L|Y_n - c|, 2C)\big)$ and $Y_n \to_\mathcal{D} c$. Moreover $II_n \to 0$ since $X_n \to_\mathcal{D} X$ and $g(\bullet, c):\mathbb{R} \to \mathbb{R}$ is bounded and continuous. That completes the proof of (27). (i)–(iii) follow since (27) implies that $h(X_n, Y_n) \to_\mathcal{D} h(X, c)$ for all (X, c)-continuous mappings h. ∎

Exercise 1 (*On l'Hospital's rule for complex-valued functions*). (a) Consider the complex valued functions f and g defined on $(0,1)$ by $g(x) = xe^{i/x}$ and $f(x) = g(x) + x$. Show that $\lim_{x\downarrow 0} f(x) = 0 = \lim_{x\downarrow 0} g(x)$ and $\lim_{x\downarrow 0} f'(x)/g'(x) = 1$, but that $\lim_{x\downarrow 0} f(x)/g(x)$ does not exist. (b) Show that l'Hospital's rule for evaluating the limit of a ratio of two complex-valued functions of a real variable is valid if the denominator is real-valued. ◇

Exercise 2. Let X be a random variable with mean 0, variance 1, and characteristic function ϕ. For $t \in \mathbb{R}$, set $\alpha(t) = \phi'(t)$, $\beta(t) = \phi''(t)$, and $\gamma(t) = \phi'(t)/t$ (take $\gamma(0)$ to be 1). For $\delta = \alpha$, β, and γ, show that

$$|\delta(t)| \leq 1 \text{ for all } t \text{ and } \lim_{t\to 0} \delta(t) = \begin{cases} 0, & \text{if } \delta = \alpha, \\ -1, & \text{if } \delta = \beta \text{ or } \gamma. \end{cases} \tag{28}$$

Show moreover that

$$|\alpha(t)| \leq |t| \text{ for all } t. \tag{29}$$

[Hint: Use Theorem 1.] ◇

Exercise 3. Use (4) to prove that $Z \sim N(0,1)$ has characteristic function $\phi_Z(t) = \exp(-t^2/2)$. ◇

Exercise 4. Let X be a random variable with characteristic function ϕ. Show that if ϕ is differentiable near 0 and ϕ' is differentiable at 0, then $E(X^2) < \infty$. [Hint: use l'Hospital's rule (carefully, see Exercise 1) to show that $\lim_{t\downarrow 0} \big(\phi(t) + \phi(-t) - 2\big)/t^2 = \phi''(0)$ and then use Fatou's Lemma (Theorem 11.4).] ◇

In the next two exercises, $f \colon \mathbb{R} \to \mathbb{C}$ is a continuous integrable function and $\hat{f}$ is its Fourier transform, defined by

$$\hat{f}(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\, dx$$

for $-\infty < t < \infty$.

Exercise 5. Let f and $\hat{f}$ be as above and let k be a positive integer. (a) Show that if $x^k f(x)$ is an integrable function of x, then $\hat{f}$ is k times continuously differentiable and

$$\hat{f}^{(k)}(t) = i^k \int_{-\infty}^{\infty} e^{itx} x^k f(x)\, dx \tag{30}$$

for all $t \in \mathbb{R}$. [Hint: this can be proved by the argument used for (3); alternatively, it can be deduced from (3) itself.] (b) Show that if $x^k f(x)$ is an integrable function of x and $\hat{f}^{(k)}(t)$ is an integrable function of t, then

$$x^k f(x) = \frac{i^{-k}}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \hat{f}^{(k)}(t)\, dt \tag{31}$$

for all $x \in \mathbb{R}$. [Hint: apply the Fourier inversion formula of Exercise 13.14 to (30).] ◇

Exercise 6. Let X be a (not necessarily integrable) random variable with characteristic function ϕ.. Show that if $\phi'(0)$ exists, then $\phi'(0)/i \in \mathbb{R}$. [Hint: Think about $\lim_{t\to 0}(\phi(t) - \phi(-t))/(2t)$.] ◇

The next three exercises deal with the following situation. Let X_1, X_2, … be independent random variables, each distributed like a random variable X with distribution function F and characteristic function ϕ. Put $S_n = X_1 + \cdots + X_n$ for $n \geq 1$. For $x \geq 0$ set

$$p(x) = xP[\,|X| \geq x\,]$$

$$\mu(x) = E(XI_{\{|X|\leq x\}}) = \int_{\xi : |\xi| \leq x} \xi\, F(dx).$$

Finally, let μ be a real number. The goal of the exercise is to prove (34) $\implies$ (33) $\implies$ (32) in the following version of the weak law of the large numbers. It can be shown that (32) $\implies$ (34), but we don't explore that here.

Theorem 8. *In the situation described just above, the following are equivalent:*

$$S_n/n \to_P \mu \ \text{as } n \to \infty; \tag{32}$$

$$\phi \ \text{is differentiable at 0 with derivative } \phi'(0) = i\mu; \tag{33}$$

$$\lim_{x \to \infty} p(x) = 0 \ \text{and} \ \lim_{x \to \infty} \mu(x) = \mu. \tag{34}$$

[handwritten: $P(\frac{x}{2^i}) = \frac{x}{2^i} \cdot P(|X| \geq \frac{x}{2^i})$ dominated function $\leq \frac{x}{2^i} \cdot \frac{|E|x|}{\frac{x}{2^i}} = |E|X|$]

Exercise 7. Show that if $p(x) \to 0$ as $x \to \infty$, then

[handwritten: $E_i(x) = \{w: \frac{x}{2^{i+1}} < |X(w)| * \frac{x}{2^i}\}$ $x^{-1} E(X^2 I_{|X| \leq x}) = x^{-1} \int_{\cup E_i} X^2(w) P(dw) \leq x^{-1} \sum_{i=0}^{\infty} \frac{x^2}{2^i \cdot 2^i} P(E_i)$]

$$\frac{1}{x} E(X^2 I_{\{|X| \leq x\}}) \to 0 \ \text{as } x \to \infty. \tag{35}$$

[handwritten: $\leq \sum_{i=0}^{\infty} \frac{x}{2^i} P(|X| \geq \frac{x}{2^{i+1}})$ $\leq 2 \sum_{i=0}^{\infty} \frac{x}{2^i} P(\frac{x}{2^{i+1}})/2^i$]

[Hint: Show that $x^{-1} E(X^2 I_{\{|X| \leq x\}}) \leq 2 \sum_{i=0}^{\infty} p(x/2^i)/2^i$ and use the DCT. Alternatively, write X^2 as $\int_0^{|X|} 2\xi \, d\xi$ and use Fubini's theorem.]◇

Exercise 8. Show that (34) implies (33). [Hint: for nonzero $h \in \mathbb{R}$ write

[handwritten: $x^{-1} E(X^2 \mathbb{1}_{|X| \leq x}) \to 0$]

$$\frac{\phi(h) - \phi(0)}{h} - i\mu(x)$$

[handwritten: $-\frac{h^2 X^2}{2} + o(h^2)$]   [handwritten: $ihX + o(hX)$]

$$= \frac{1}{h} E\big((e^{ihX} - 1 - ihX) I_{\{|X| \leq x\}}\big) + \frac{1}{h} E\big((e^{ihX} - 1) I_{\{|X| > x\}}\big)$$

[handwritten: $\| ihX + o(hX)$; $\to 0 \quad x \to \infty$]

with $x := 1/h$ and use (1) and (35).]   ◇

Exercise 9. Show that (33) implies (32).   ◇

Exercise 10. (a) Suppose that X is an integrable random variable. Show directly, without using Theorem 8, that (34) holds with $\mu = E(X)$. (b) Let X be a random variable with density f of the form $f(x) = [C/(x^2 \log(|x|))] I_{[2,\infty)}(|x|)$ for an appropriate constant C. Show that X is not integrable, but that (34) holds with $\mu = 0$.   ◇

Exercise 11. Prove Stirling's formula for $n!$ by the following method. Let $S_n = X_1 + \cdots + X_n$, where the X_n's are independent and each has a Poisson distribution with mean 1. Show that:

(a) $E\big((S_n - n)^-/\sqrt{n}\big) = n^{n+1/2} e^{-n}/n!$;

(b) $(S_n - n)^-/\sqrt{n} \to_{\mathcal{D}} Z^-$;

(c) $E\big((S_n - n)^-/\sqrt{n}\big) \to EZ^- = 1/\sqrt{2\pi}$;

(d) $n! \sim \sqrt{2\pi n}\, n^n e^{-n}$.   ◇

Exercise 12. Let X_1, X_2, ... be iid random variables, each distributed like X, and suppose that $S_n/\sqrt{n} \to_{\mathcal{D}} Z \sim \mathfrak{N}$. Show that $E(X^2) = 1$ and $E(X) = 0$ by performing the steps below. Put $\psi = |\phi_X|^2$; note that ψ is the characteristic function of $D = X_1 - X_2$, and that ψ is real and positive.

(a) Show that $\big(\psi(1/\sqrt{n})\big)^n \to e^{-1}$.

(b) Use (a) to show that $n(1 - \psi(1/\sqrt{n})) \to 1$.

(c) Use (b) and Fatou's Lemma to show that $E(D^2) \leq 2$.

(d) Use (b) and (c) to show that $E(D^2) = 2$.

(e) Use (d) to show first that $E(X^2) < \infty$ and then $\text{Var}(X_1) = 1$.

(f) Show that $E(X_1) = 0$.   ◇

Exercise 13. Show that $\big||z_2| - |z_1|\big| \leq |z_2 - z_1|$ for any two complex numbers z_1 and z_2. Deduce that (15) holds.   ◇

Exercise 14. Let $\mathfrak{n}$ and $\mathfrak{N}$ be defined by (7). Show that

$$\frac{1}{z} - \frac{1}{z^3} \leq \frac{1 - \mathfrak{N}(z)}{\mathfrak{n}(z)} \leq \frac{1}{z} \tag{36}$$

for each $z > 0$. [Hint: integrate by parts.] Deduce that (16) holds.   ◇

Exercise 15. Give an example where (8) holds, but (17) does not. [Hint: use discrete summands.]   ◇

Exercise 16. Let ϕ be a characteristic function having the following properties: for each real number $r > 0$, $\phi_r(t) := (\phi(t/\sqrt{r}))^r$ is the characteristic function of some random variable X_r; X_1 has mean 0 and variance 1; and ϕ satisfies (11) and (12). Show that: (a) for all large r, X_r has a bounded continuous density f_r; (b) one has $|f_r(x) - \mathfrak{n}(x)| \to 0$, uniformly in $x \in \mathbb{R}$, as $r \to \infty$. $\diamond$

Exercise 17. Prove Stirling's formula for the Gamma function, to wit:

$$\Gamma(r + 1) = (1 + o(1)) \sqrt{2\pi r}\, r^r\, e^{-r} \quad \text{as } r \to \infty. \tag{37}$$

[Hint: Apply the preceding exercise to the characteristic function ϕ_X of $X := Y - 1$, where Y is a standard exponential random variable.]$\diamond$

The next several exercises concern the following result:

Theorem 9. *Let X_1, X_2, ... be iid random variables with mean $E(X_1) = 0$, variance $E(X_1^2) = 1$, characteristic function $\phi(t) = E(e^{itX_1})$, and partial sums $S_n = X_1 + \cdots + X_n$. Suppose that*

$$|\phi|^k \text{ is integrable for some integer } k \geq 1, \text{ and} \tag{38}$$
$$b(\delta) := \sup\{\, |\phi(t)| : |t| \geq \delta \,\} < 1 \quad \text{for all } \delta > 0. \tag{39}$$

Then as $n \to \infty$ the density f_n of $S_n/\sqrt{n}$ tends to the density $\mathfrak{n}$ of $Z \sim N(0,1)$ in such a way that

$$x^2 f_n(x) \text{ tends to } x^2 \mathfrak{n}(x) \text{ uniformly in } x \in \mathbb{R} \tag{40}$$
and
$$\int_{-\infty}^{\infty} x^2 |f_n(x) - \mathfrak{n}(x)|\, dx \to 0. \tag{41}$$

This is a strengthening of Theorem 4, which asserted that for all $n \geq k$, f_n exists and can be chosen to be continuous, and that (40) and (41) hold without the x^2's. In each of the next five exercises, suppose that the conditions of Theorem 9 hold.

Exercise 18. Show that (41) holds. [Hint: By (13) $x^2|f_n(x) - \mathfrak{n}(x)| \to 0$ for each $x \in \mathbb{R}$. Use the Sandwich Theorem.] $\diamond$

Exercise 19. Let C be a real positive number and let G_C be the collection of (Borel-measurable) functions g mapping $\mathbb{R}$ to $\mathbb{R}$ such that $|g(x)| \leq C(1 + x^2)$ for all $x \in \mathbb{R}$. Show that as n tends to ∞, $E(g(S_n/\sqrt{n}))$ tends to $E(g(Z))$, uniformly for all $g \in G_C$. In what ways is this an improvement over the conclusion $S_n/\sqrt{n} \to_{\mathcal{D}} Z$ of the global CLT? [Hint: Use (14) and (41).] $\diamond$

Exercise 20. Let $\psi_n(t) := (\phi(t/\sqrt{n}))^n$ be the characteristic function of $S_n/\sqrt{n}$. Show that for $n \geq k + 2$,

$$\psi_n''(t) = (n-1)(\phi(t/\sqrt{n}))^{n-2}(\phi'(t/\sqrt{n}))^2 + (\phi(t/\sqrt{n}))^{n-1}\phi''(t/\sqrt{n})$$

is integrable in t and deduce that

$$x^2 f_n(x) = -\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \psi_n''(t)\, dt \tag{42}$$

for all $x \in \mathbb{R}$. [Hint: Use (28) and (31).] $\diamond$

Exercise 21. Let $\psi(t) := e^{-t^2/2}$ be the characteristic function of Z. Show that

$$\lim_{n\to\infty} \psi_n''(t) = \psi''(t) \tag{43}$$

for all $t \in \mathbb{R}$. [Hint: Use (28) again.] $\diamond$

Exercise 22. Show that

$$\sup\{\, x^2|f_n(x) - \mathfrak{n}(x)| : x \in \mathbb{R} \,\}$$
$$\leq \frac{1}{2\pi} \int_{-\infty}^{\infty} |\psi_n''(t) - \psi''(t)|\, dt \to 0 \quad \text{as } n \to \infty. \tag{44}$$

[Hint: Imitate the argument for (13); use (29).] $\diamond$

Exercise 23. Let X be a random variable with characteristic function ϕ. Show that: (a) if X has a lattice distribution with span h, then $|\phi(2\pi/h)| = 1$; (b) conversely, if $|\phi(t)| = 1$ for some $t \neq 0$, then X has a lattice distribution with span $h = 2\pi/|t|$. ◇

Exercise 24. Prove Theorem 5. [Hint: use the preceding exercise and the inversion theorem in Exercise 13.7. Most of the argument of Theorem 4 applies without change; just give the part that is different.] ◇

Exercise 25. In the context of Cramer's theorem, suppose that $\mathcal{X} = (0, \infty) = \Theta$, ν is Lebesgue measure, and $f_\theta(x) = \theta e^{-\theta x}$. Show that conditions C1–C4 hold and use Theorem 6 to deduce that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to_{\mathcal{D}} \theta_0 Z \tag{45}$$

under P_{θ_0}. Then deduce (45) anew, by finding an explicit expression for $\hat{\theta}_n$ and deriving (45) from the central limit theorem and the Δ method (see Theorem 14.6). ◇

Exercise 26. In the context of Cramer's theorem, suppose that $\mathcal{X} = \mathbb{R}$, $\Theta = [0, \infty)$, and $f_\theta = \exp\left(-(x-\theta)^2/2\right)/\sqrt{2\pi}$. Show that when $\theta_0 = 0$, the maximum likelihood estimator $\hat{\theta}_n$ of θ satisfies $\sqrt{n}(\hat{\theta}_n - \theta_0) \to_{\mathcal{D}} \max(0, Z)$ where $Z \sim \mathfrak{N}$. Why doesn't this contradict Cramer's theorem? ◇

Exercise 27. Prove the following theorem. [Hint: use a second order Taylor expansion of $L_n(\theta)$ about $\hat{\theta}_n$ in conjunction with (21), (26), and Lemma 2.] ◇

Theorem 10 (Wilk's theorem). *Suppose conditions* C0–C4 *of Cramer's theorem hold and* $\theta_0 \in \text{int}(\Theta)$. *Let*

$$\lambda_n := \frac{L_n(\theta_0)}{\sup_{\theta \in \Theta} L_n(\theta)} = \frac{L_n(\theta_0)}{L_n(\hat{\theta}_n)} \tag{46}$$

be the likelihood ratio statistic for testing the hypothesis $H: \theta = \theta_0$ *against the alternative* $A: \theta \neq \theta_0$. *Under* P_{θ_0} *one has*

$$-2 \log(\lambda_n) \to_{\mathcal{D}} Z^2, \tag{47}$$

with $Z \sim \mathfrak{N}$.