On the Asymptotics of Convex Stochastic Optimization

Charles J. Geyer

University of Minnesota

The asymptotics of minimizers of a sequence of random convex functions on a finite-dimensional Euclidean space are described using very weak regularity conditions. If random convex functions g_n finite on some open set converge in law pointwise on a dense set to a random function g that almost surely is finite on some nonempty open set and possesses a unique minimizer, then the minimizers of the g_n converge in law to the minimizer of g. Under the same conditions, confidence sets constructed as level sets of the g_n converge in law to the corresponding level sets of g in the metric of Painlevé-Kuratowski set convergence.

1. Introduction. The asymptotics of convex stochastic optimization have been studied in recent papers by Haberman (1989), Knight (1989), Pollard (1991), and Niemiro (1992). The problem they address is finding the conditions under which a sequence g_n of random convex functions on $\mathbb{R}^d$ has minimizers t_n that converge in law to some limit t. For example, g_n could be the negative of the log likelihood for an exponential family, centered and rescaled so as to have nontrivial asymptotics

$$g_n(s) = l_n(\theta_0) - l_n(\theta_0 + n^{-1/2}s) = \sum_{i=1}^{n} \log \frac{f_{\theta_0}(x_i)}{f_{\theta_0 + n^{-1/2}s}(x_i)} \tag{1.1}$$

Other examples of convex stochastic optimization include least squares and least absolute deviation regression and M-estimation with a convex objective function.

The approach of this paper, following Geyer (1994b), is to study the convergence of the g_n to an asymptotic objective function g, where g_n and g are taken to be random elements of some complete separable metric space. In the regular case (1.1) converges in law to

$$g(s) = s'Z + \tfrac{1}{2}s'Ks \tag{1.2}$$

where K is the Fisher information matrix and Z is an $N(0, K)$ random vector.

More generally, suppose we are trying to estimate a parameter θ in $\mathbb{R}^d$ by minimizing some random convex function $F_n(\theta)$, and suppose that for some nonrandom sequences $\{\delta_n\}$ and $\{\beta_n\}$ that

$$g_n(s) = \beta_n \left[F_n(\theta_0 + \delta_n s) - F_n(\theta_0)\right] \tag{1.3}$$

AMS 1991 *subject classifications.* Primary 60F05; secondary 49J55, 62F12.

Key words and phrases. Central limit theorem, convexity, constraint, M-estimation, epi-convergence, level set, median.

has nontrivial asymptotics. Then usually $\hat{s}_n = \delta_n^{-1}(\hat{\theta}_n - \theta_0)$, the minimizer of g_n, converges in law to s, the minimizer of g.

The general theorem goes as follows. Suppose g_n is a sequence of random lower semicontinuous convex functions on $\mathbb{R}^d$ and g is another such function. Let D be a countable dense set in $\mathbb{R}^d$. If for each finite subset $\{s_1, \ldots, s_k\}$ of D, the random vector $(g_n(s_1), \ldots, g_n(s_k))$ converges in law to the random vector $(g(s_1), \ldots, g(s_k))$, and if with probability one g is finite on some nonempty open set and has a unique minimizer t, then t_n converges in law to t and $g_n(t_n)$ converges in law to $g(t)$.

This theorem makes no mention of Taylor series expansion of g_n or weaker series expansion notions such as stochastic equicontinuity and differentiability in quadratic mean. Pointwise convergence in law suffices. Moreover, it does not require asymptotic normality. If $g(s)$ is nonnormal or g is not quadratic, then the minimizer t will have a nonnormal distribution, but as Example 2 shows, this is no more difficult than the normal case. If we allow $+\infty$ as a possible value for $g_n(s)$ and $g(s)$, the same theorem deals with constraints with no additional difficulty.

This theorem can be proved using no probabilistic tools except the Prohorov and Skorohod theorems. It does need a number of facts about the convergence of nonstochastic convex functions, which are assembled in Section 2. The main theorem is proved in Section 3. A theorem on convergence of confidence sets is proved in Section 4. Section 5 gives examples.

2. Asymptotics for Nonstochastic Convex Functions. This section develops the theory of epiconvergence of extended-real-valued convex functions (allowing $\pm\infty$ as values). Readers who want to avoid some of the complexity may consider the special case of real-valued convex functions defined on all of $\mathbb{R}^d$. In this special case, epiconvergence is the same as uniform convergence on compact sets.

Following Rockafellar (1970) we say a function g from $\mathbb{R}^d$ to the extended real line $\overline{\mathbb{R}} = [-\infty, +\infty]$ is convex if

$$g\bigl(\lambda x + (1-\lambda)y\bigr) \le \lambda g(x) + (1-\lambda)g(y) \tag{2.1}$$

whenever $0 < \lambda < 1$ and neither $g(x)$ or $g(y)$ is equal to $+\infty$. Equivalently, g is convex if its *epigraph*

$$\operatorname{epi} g = \left\{\, (x, \lambda) \in \mathbb{R}^d \times \mathbb{R} : g(x) \le \lambda \,\right\},$$

the set of points on or above the graph, is a convex subset of $\mathbb{R}^d \times \mathbb{R}$.

The function g is *proper* if $g(x) > -\infty$ for all x and $g(x) < +\infty$ for at least one x. All functions of interest in applications are proper. The possibility of the value $-\infty$ is allowed only for reasons of mathematical convenience; it produces compactness in some arguments.

Allowing $+\infty$ as a value permits treatment of constrained problems as if they were unconstrained problems. Minimizing a function $g : \mathbb{R}^d \to \overline{\mathbb{R}}$ is equivalent to minimizing g over its *effective domain*

$$\operatorname{dom} g = \left\{\, x \in \mathbb{R}^d : g(x) < +\infty \,\right\}.$$

Hence the constrained problem of minimizing a real-valued convex function g defined on a convex set C in $\mathbb{R}^d$ is equivalent to the unconstrained problem of minimizing the extended-real-valued convex function $\tilde{g}$ that is equal to g on C and is equal to $+\infty$ elsewhere. Note that $\operatorname{dom} \tilde{g} = C$.

In optimization, it makes sense to restrict interest to lower semicontinuous functions. Failure to do so leads to uninteresting pathology. Consider the convex function $g : \mathbb{R} \to \overline{\mathbb{R}}$ defined by

$$g(x) = \begin{cases} x, & 0 > x \\ +\infty, & x \leq 0 \end{cases}$$

The minimum is not achieved because g is not lower semicontinuous at 0. We want to say that 0 is the minimizer, but cannot because we failed to define g to be lower semicontinuous. Any convex function is continuous on any open set on which it is finite. Thus lower semicontinuity only involves behavior on the boundary of the effective domain.

A function $g : \mathbb{R}^d \to \overline{\mathbb{R}}$ is lower semicontinuous if and only if $\operatorname{epi} g$ is a closed subset of $\mathbb{R}^d \times \mathbb{R}$ and if and only if every *level set*

$$\operatorname{lev}_\alpha g = \left\{ x \in \mathbb{R}^d : g(x) \leq \alpha \right\} \tag{2.2}$$

for $\alpha \in \mathbb{R}$ is a closed subset of $\mathbb{R}^d$. Thus there is a close connection between closed sets and lower semicontinuous functions.

For any Hausdorff topological space S, let $K(S)$ denote the family of closed subsets of S. The Painlevé-Kuratowski topology for $K(S)$ has a subbasis consisting of sets of the form

$$\{ D \in K(S) : D \cap K = \varnothing \},$$

where K is a fixed compact set in S, and

$$\{ D \in K(S) : D \cap G \neq \varnothing \},$$

where G is a fixed open set in S (Theorem 2.75 in Attouch, 1984). Every open set in $K(S)$ is a union of finite intersections of such sets. This is sometimes called the "hit and miss criterion." A sequence C_n converges to C if C_n eventually misses every compact set that misses C and eventually hits every open set that hits C.

Let $F(S)$ denote the set of lower semicontinuous functions from S to $\overline{\mathbb{R}}$. The epiconvergence topology for $F(S)$ is inherited from the Painlevé-Kuratowski topology for epigraphs. The epigraph of $g \in F(S)$ is

$$\operatorname{epi} g = \{ (x, \lambda) \in S \times \mathbb{R} : g(x) \leq \lambda \},$$

and the subbasis consists of sets of the form

$$\{ g \in F(S) : \operatorname{epi} g \cap K = \varnothing \},$$

where K is a fixed compact set in $S \times \mathbb{R}$, and

$$\{ g \in F(S) : \operatorname{epi} g \cap G \neq \varnothing \},$$

where G is a fixed open set in $S \times \mathbb{R}$. A sequence of functions g_n epiconverges to a function g, written $g_n \xrightarrow{e} g$, if and only if $\operatorname{epi} g_n$ converges to $\operatorname{epi} g$ in the Painlevé-Kuratowski sense. For more on epiconvergence see Attouch (1984) or Rockafellar and Wets (forthcoming) and for applications in statistics Geyer (1994a, 1994b).

If S is locally compact and second countable, then $K(S)$ and $F(S)$ are compact and metrizable (Attouch, 1984, Corollary 2.79) and hence metrics can be chosen to make each complete separable metric spaces. We shall be interested in two different choices of the space S, either $\mathbb{R}^d$ or its one-point compactification $\overline{\mathbb{R}^d} = \mathbb{R}^d \cup \{\infty\}$.

Any $g \in F(\mathbb{R}^d)$ can be extended to $\overline{\mathbb{R}^d}$ by defining

$$g(\infty) = \sup_{W \in \mathcal{N}(\infty)} \inf_{s \in W} g(s)$$

where $\mathcal{N}(\infty)$ is the family of open neighborhoods of ∞ (complements of compact sets in $\mathbb{R}^d$). This is the unique lower semicontinuous extension. Thus we can consider a lower semicontinuous function on $\mathbb{R}^d$ as an element of either $F(\mathbb{R}^d)$ or $F(\overline{\mathbb{R}^d})$.

Let $K_c(\mathbb{R}^d)$ denote the topological subspace of $K(\mathbb{R}^d)$ consisting of convex sets, and let $F_c(\mathbb{R}^d)$ and $F_c(\overline{\mathbb{R}^d})$ denote the topological subspaces of $F(\mathbb{R}^d)$ and $F_c(\overline{\mathbb{R}^d})$ consisting of convex functions. The epilimit of a sequence of convex sets is convex (Aubin and Frankowska, 1990, p. 26). Hence each of these spaces $K_c(\mathbb{R}^d)$, $F_c(\mathbb{R}^d)$, and $F_c(\overline{\mathbb{R}^d})$ is a closed subspace of its parent, and each is compact and metrizable.

The following five propositions, which are minor variants of known results in optimization theory, provide the basis for our use of epiconvergence theory. Proofs are given in the appendix.

The usefulness of epiconvergence in optimization problems is clear from the following proposition. For any function g on a set S, the notation $\inf g$ means $\inf \{\, g(s) : s \in S \,\}$. We say that t_n is an *approximate minimizing sequence* for a sequence of functions g_n if for some sequences $\eta_n \downarrow 0$ and $r_n \downarrow -\infty$

$$g_n(t_n) \le \begin{cases} \inf g_n + \eta_n, & \inf g_n > -\infty \\ r_n, & \inf g_n = -\infty \end{cases} \tag{2.3}$$

Proposition 2.1. *Suppose $g_n \xrightarrow{e} g$, $t_n \to t$, and t_n is an approximate minimizing sequence for g_n. Then*

$$g(t) = \inf g \qquad \text{and} \qquad g_n(t_n) \to g(t).$$

Let $\Gamma(\mathbb{R}^d)$ denote the topological subspace of $F_c(\mathbb{R}^d)$ consisting of functions that are finite on some open set. An alternative characterization is functions that are proper and have effective domains with nonempty interiors. Let $\Gamma(\overline{\mathbb{R}^d})$ be the corresponding subspace of $F_c(\overline{\mathbb{R}^d})$. Epiconvergence and convergence of minimizers are easily established for convex functions when the limit is in $\Gamma(\mathbb{R}^d)$ from the following propositions.

Proposition 2.2. *Suppose g_n is a sequence in $F_c(\mathbb{R}^d)$ converging pointwise on a dense set to a limit g that is finite on some open set. Then g is convex and $g_n \xrightarrow{e} g$ on $\mathbb{R}^d$. Moreover, $g_n \in \Gamma(\mathbb{R}^d)$ for all sufficiently large n.*

For convex functions, the value at infinity is closely connected to the behavior of level sets. For any extended-real-valued convex function g, the level sets of g are defined by (2.2). A particular level set of interest is the "argmin"

$$\operatorname{argmin} g = \{\, x : g(x) = \inf g \,\}$$

The argmin is empty if the infimum is not achieved, but $\operatorname{lev}_\alpha g$ is nonempty for any $\alpha > \inf g$.

PROPOSITION 2.3. *For a lower semicontinuous proper convex function g on $\mathbb{R}^d$, the following conditions are equivalent*

(a) $g(\infty) = +\infty$.

(b) $\operatorname{argmin} g$ *is compact and nonempty.*

(c) *Some level set of g is compact and nonempty.*

(d) $\operatorname{lev}_\alpha g$ *is compact and nonempty for every $\alpha \geq \inf g$.*

A special case of (b) is that the minimizer exists and is unique, i. e. that $\operatorname{argmin} g$ is a singleton. From now on, condition (a) will usually be the one explicitly mentioned; the reader should remember the other equivalent conditions.

PROPOSITION 2.4. *Suppose g_n is a sequence in $F_c(\mathbb{R}^d)$ epiconverging to a proper limit g, and suppose $g(\infty) = +\infty$. Then g_n epiconverges to g on $\overline{\mathbb{R}^d}$, and* $\inf g_n \to \inf g$. *Moreover, if t_n is an approximate minimizing sequence for g_n, then $g_n(t_n) \to \inf g$.*

PROPOSITION 2.5. *Suppose g_n is a sequence in $F_c(\mathbb{R}^d)$ epiconverging to a proper limit g, and suppose the infimum of g is achieved at a unique point t. Then $g(\infty) = +\infty$ and g_n epiconverges to g on $\overline{\mathbb{R}^d}$. If t_n is an approximate minimizing sequence for g_n, then $t_n \to t$ and $g_n(t_n) \to g(t)$.*

3. Asymptotics for Stochastic Convex Functions. Now consider a sequence g_n of random elements of $F_c(\mathbb{R}^d)$. That is, for some probability space $(\Omega, \mathcal{A}, P)$, there is a sequence of measurable maps $g_n : \Omega \to F_c(\mathbb{R}^d)$. So $g_n(\omega)$ is a lower semicontinuous convex function taking the value $g_n(\omega)(s)$ at the point s. One criterion for measurability is that $(\omega, s) \mapsto g_n(\omega)(s)$ be a jointly measurable function from $\Omega \times \mathbb{R}^d \to \overline{\mathbb{R}}$ and that $(\Omega, \mathcal{A}, P)$ be a complete measure space. For other criteria, see Aubin and Frankowska (1990, pp. 303 ff.)

$F_c(\mathbb{R}^d)$ and $F_c(\overline{\mathbb{R}^d})$ are compact and metrizable, hence Polish spaces. $\Gamma(\mathbb{R}^d)$ and $\Gamma(\overline{\mathbb{R}^d})$ are not closed subspaces of their parent spaces, hence not complete, though they are separable. By Lemma A.3 in the appendix $\Gamma(\mathbb{R}^d)$ and $\Gamma(\overline{\mathbb{R}^d})$ are measurable subsets of their parent spaces, hence a random element of $\Gamma(\mathbb{R}^d)$ can also be considered a random element of the Polish space $F_c(\mathbb{R}^d)$, and similarly for $\Gamma(\overline{\mathbb{R}^d})$ and $F_c(\overline{\mathbb{R}^d})$.

We are interested in pointwise convergence in law and epiconvergence in law. Because of Proposition 2.2, it is sufficient to consider pointwise convergence on a countable dense set D in $\mathbb{R}^d$. The space of all functions from D to $\overline{\mathbb{R}}$ with the topology of pointwise convergence is denoted $\overline{\mathbb{R}}^D$. By epiconvergence in law we mean convergence in law of random elements of the Polish space $F_c(\mathbb{R}^d)$, and by pointwise convergence in law we mean convergence in law of random elements of the Polish space $\overline{\mathbb{R}}^D$.

Because open sets in $\overline{\mathbb{R}}^D$ only restrict a finite number of coordinates, the portmanteau theorem says that convergence in law in $\overline{\mathbb{R}}^D$ is the same as convergence of finite-dimensional distributions: g_n converges in law to g if and only if for each finite subset $\{s_1, \ldots, s_k\}$ of D, the random vector $(g_n(s_1), \ldots, g_n(s_k))$ converges in law to the random vector $(g(s_1), \ldots, g(s_k))$.

Convergence in law in $F_c(\mathbb{R}^d)$ means, of course, that for every bounded continuous function H on $F_c(\mathbb{R}^d)$ that $EH(g_n) \to EH(g)$, but we will not need to use this definition, because we will derive everything from the nonstochastic case using the Prohorov and Skorohod theorems.

LEMMA 3.1. *Suppose $g_1, g_2, \ldots$ and g are random elements of $\Gamma(\mathbb{R}^d)$, and suppose for some countable dense subset D in $\mathbb{R}^d$ that $g_n|D$ converges in law to $g|D$ in $\overline{\mathbb{R}}^D$. Then g_n epiconverges in law to g.*

Suppose, in addition, that $g(\infty) = +\infty$ almost surely. Then g_n epiconverges in law to g on $\overline{\mathbb{R}^d}$.

PROOF. Define $u : \Gamma(\mathbb{R}^d) \to \overline{\mathbb{R}}^D$ by $g \mapsto g|D$. By Lemma A.4 in the appendix u is a measurable bijection onto its range. Apply the Skorohod theorem to the convergence of $h_n = u(g_n)$ to $h = u(g)$ in $\overline{\mathbb{R}}^D$ constructing random functions h_n^* and h^* on another probability space with the same laws as h_n and h such that the convergence is almost sure. Then $g_n^* = u^{-1}(h_n^*)$ and $g^* = u^{-1}(h^*)$ have the same laws as g_n and g, and by Proposition 2.2 $g_n^* \xrightarrow{e} g^*$ almost surely. Since almost sure convergence implies convergence in law, we have epiconvergence in law of g_n to g.

If $g(\infty) = +\infty$ almost surely, then $g^*(\infty) = +\infty$ almost surely, and by Proposition 2.4 $g_n^* \xrightarrow{e} g^*$ on $\overline{\mathbb{R}^d}$ almost surely, which implies epiconvergence in law of g_n to g on $\overline{\mathbb{R}^d}$. □

THEOREM 3.2. *Suppose $g_1, g_2, \ldots$ and g are random elements of of $\Gamma(\mathbb{R}^d)$, and g_n epiconverges in law to g, and suppose that the infimum of g is achieved at a unique point t with probability one. If t_n is an approximate minimizing sequence for g_n, then $t_n \xrightarrow{\mathcal{L}} t$ and $g_n(t_n) \xrightarrow{\mathcal{L}} g(t)$.*

That there exist random vectors t_n satisfying (2.3) follows from measurable selection theorems (see, e. g., Niemiro, 1992, or Aubin and Frankowska, 1990, Section 8.1).

PROOF. Since $\overline{\mathbb{R}^d}$ and $F_c(\mathbb{R}^d)$ are both compact, their Cartesian product is compact and (t_n, g_n) is a tight sequence of random elements of $\overline{\mathbb{R}^d} \times F_c(\mathbb{R}^d)$. Hence by the Prohorov theorem every subsequence has a further subsequence that converges in law to a limit (t', g). By the Skorohod theorem there are random elements on some other probability space with the same laws such that the convergence is almost sure. That is, suppressing multiple subscripts for the subsequences, $t_n \to t'$ and $g_n \xrightarrow{e} g$ with probability one. By Propositions 2.3 and 2.4 $g_n \xrightarrow{e} g$ on $\overline{\mathbb{R}^d}$, and by Proposition 2.1 $t' = t$ and $g_n(t_n) \to g(t)$. Since almost sure convergence implies convergence in law, every subsequence of t_n has a further subsequence converging in law to t and every subsequence of $g_n(t_n)$ has a further subsequence converging in law to $g(t)$, which implies convergence in law of both sequences. □

If there are no constraints (functions are everywhere finite), pointwise convergence of g_n to g on dense set implies convergence everywhere, so the convergence can be checked everywhere rather than just on a dense set. If there are constraints, however, $\operatorname{dom} g$ will have a nonempty interior and exterior. Convergence on a dense set does imply (Theorem 7.12 in Rockafellar and Wets, forthcoming) convergence everywhere in the interior and exterior of $\operatorname{dom} g$, but in general g_n will not converge to g pointwise on the boundary of $\operatorname{dom} g$. Thus one checks pointwise convergence on a dense set that almost surely avoids the boundary of the effective domain of the limit function g.

In most applications there is no need to check convergence of all finite-dimensional distributions, one-dimensional distributions suffice.

Lemma 3.3. *Suppose*

$$
\begin{aligned}
g_n(s) &= Z_n' s + h_n(s) \\
g(s) &= Z' s + h(s)
\end{aligned}
$$

where Z_n and Z are random vectors, h_n are random functions, and h is a nonrandom function. If $Z_n \xrightarrow{\mathcal{L}} Z$ and $h_n(s) \xrightarrow{P} h(s)$, for all $s \in D$, then g_n converges in law to g in $\overline{\mathbb{R}}^D$.

Proof. Fix a finite subset $\{s_1, \ldots, s_k\}$ of D. The Cramér-Wold device implies $(Z_n' s_1, \ldots Z_n' s_k) \xrightarrow{\mathcal{L}} (Z' s_1, \ldots Z' s_k)$. By definition of convergence in probability, $(h_n(s_1), \ldots h_n(s_k)) \xrightarrow{\mathcal{L}} (h(s_1), \ldots h(s_k))$. Hence $(g_n(s_1), \ldots g_n(s_k)) \xrightarrow{\mathcal{L}} (g(s_1), \ldots g(s_k))$. $\square$

4. Confidence Sets. The theory of epiconvergence also yields the following result about confidence sets constructed as level sets of the objective function. For any $\gamma > 0$ define

$$
A_{n,\gamma} = \left\{ s \in \mathbb{R}^d : g_n(s) \le \inf g_n + \gamma \right\} \tag{4.1a}
$$

$$
A_{\gamma} = \left\{ s \in \mathbb{R}^d : g(s) \le \inf g + \gamma \right\} \tag{4.1b}
$$

Since the objective functions are lower semicontinuous, $A_{n,\gamma}$ and A_γ are closed sets. If $\inf g_n = -\infty$ and g_n is proper, then $A_{n,\gamma}$ will be empty, but A_γ will never be empty if we impose the requirement that $g(\infty) = +\infty$.

In the usual case, where the argument of g_n represents deviation from the true parameter value, the level set $A_{n,\gamma}$ covers the true parameter value if it contains the origin. Moreover, as in (1.1), (1.2), and (1.3) $g_n(0) = g(0) = 0$ almost surely. Hence $A_{n,\gamma}$ covers whenever the random variable $U_n = -\inf g_n$ is less than or equal to γ. In general, we cannot show that the coverage of $A_{n,\gamma}$ converges to the coverage of A_γ (more on this later), but we can show that $A_{n,\gamma}$ converges in law to A_γ considered as random elements of the space $K(\mathbb{R}^d)$.

As with Theorem 3.2, this follows simply from the following proposition about convergence of level sets of nonstochastic convex functions. The exact statement is from Rockafellar and Wets (forthcoming), but the proof is contained in the proof of Theorem 3.1 in Beer, Rockafellar and Wets (1992).

Proposition 4.1. *If $g_n \to g$ in $F(\mathbb{R}^d)$, then for any real α*

$$
\limsup_n \operatorname{lev}_{\alpha_n} g_n \subset \operatorname{lev}_\alpha g, \qquad \text{for any sequence } \alpha_n \to \alpha \tag{4.2a}
$$

$$
\liminf_n \operatorname{lev}_{\alpha_n} g_n \supset \operatorname{lev}_\alpha g, \qquad \text{for some sequence } \alpha_n \to \alpha \tag{4.2b}
$$

The limits superior and inferior here refer to Painlevé-Kuratowski convergence. The limsup of a sequence C_n is the set of all cluster points x defined by $x_{n_k} \to x$ for $x_{n_k} \in C_{n_k}$, and the liminf is the set of all limit points x defined by $x_n \to x$ for $x_n \in C_n$, $n \ge n_0$.

Corollary 4.2. *If $g_n \to g$ in $F_c(\overline{\mathbb{R}^d})$ and $\inf g$ is finite, then $A_{n,\gamma} \to A_\gamma$ for any $\gamma > 0$, where $A_{n,\gamma}$ and A_γ are defined by (4.1).*

PROOF. Since $\overline{\mathbb{R}^d}$ is compact, $\inf g_n > R$ for all sufficiently large n and any $R < \inf g$ by the hit and miss criterion. Any approximate minimizing sequence t_n has a convergent subsequence $t_{n_k} \to t$. By Proposition 2.1 for all sufficiently large n

$$R < \inf g_{n_k} \le g_{n_k}(t_{n_k}) \to g(t) = \inf g.$$

Hence $\inf g_n \to \inf g$. Since $\operatorname{lev}_\alpha g \subset \operatorname{lev}_\beta g$ whenever $\alpha \le \beta$

$$\begin{aligned} \limsup\nolimits_n A_{n,\gamma} &\subset A_\gamma \\ \liminf\nolimits_n A_{n,\gamma} &\supset A_{\gamma-\epsilon}, \qquad \epsilon > 0 \end{aligned}$$

Hence

$$\liminf\nolimits_n A_{n,\gamma} \supset \operatorname{cl}\left(\bigcup\nolimits_{\epsilon>0} A_{\gamma-\epsilon}\right)$$

because a Painlevé-Kuratowski limit is closed. The right hand side is equal to A_γ by Theorem 7.6 in Rockafellar (1970). Hence $A_{n,\gamma} \to A_\gamma$. □

THEOREM 4.3. *Suppose g_1, g_2, ... and g are random elements of of $\Gamma(\overline{\mathbb{R}^d})$ and g_n epiconverges in law to g on $\overline{\mathbb{R}^d}$ and $\inf g > -\infty$ almost surely. If $A_{n,\gamma}$ and A_γ are defined by (4.1), then $A_{n,\gamma}$ converges in law to A_γ in $K_c(\mathbb{R}^d)$ for any $\gamma > 0$.*

PROOF. The set

$$W = \left\{\, g \in \Gamma(\mathbb{R}^d) : g(\infty) > -\infty \,\right\}$$

is open in $\Gamma(\overline{\mathbb{R}^d})$ by the hit and miss criterion.

Consider the map $H : \Gamma(\overline{\mathbb{R}^d}) \to K(\mathbb{R}^d)$ defined by $H(g) = A_\gamma$ when g is in W and $H(g) = \varnothing$ elsewhere. By assumption g_n converges in law to g in $\Gamma(\overline{\mathbb{R}^d})$ and g is concentrated on W. By the corollary, H is continuous on W. Hence $A_{n,\gamma} = H(g_n)$ converges in law to $A_\gamma = H(g)$ in $K(\mathbb{R}^d)$ by the continuous mapping theorem. □

This limit theorem means, of course, that $EH(A_{n,\gamma}) \to EH(A_\gamma)$ for any real-valued bounded continuous function H on $K(\mathbb{R}^d)$. By the continuous mapping theorem, we also have $H(A_{n,\gamma}) \xrightarrow{\mathcal{L}} H(A_\gamma)$ for any continuous H, bounded or not.

One might think we would want to take $H(A) = 1\{0 \in A\}$, the indicator of whether the set covers the true parameter value. But this function is not continuous as is shown by the following simple example. Defining g_a by

$$g_a(x) = \begin{cases} 0, & x \ge a \\ +\infty, & x < a \end{cases}$$

Then $g_{1/n} \xrightarrow{e} g_0$, but for every $\gamma > 0$, the set $A_{n,\gamma} = [1/n, \infty)$ does not contain the origin, whereas $A_\gamma = [0, \infty)$ does.

We can take $H(A)$ to be the distance from the set A to the origin, that is, either zero if A covers or the amount by which it fails to cover otherwise. This is continuous by the hit and miss criterion

$$\left\{\, A \in K(\mathbb{R}^d) : H(A) < \epsilon \,\right\} = \left\{\, A \in K(\mathbb{R}^d) : A \cap B_\epsilon \ne \varnothing \,\right\}$$

where B_ϵ is the open ball of radius ϵ centered at the origin. The portmanteau theorem then says for any $\epsilon > 0$

$$\liminf_{n\to\infty} \Pr\left(H(A_{n,\gamma}) < \epsilon\right) \geq \Pr\left(H(A_\gamma) < \epsilon\right),$$

that is, the probability of the finite sample confidence interval covering the true parameter value or missing by less than ϵ is asymptotically bounded below by same quantity in the asymptotic problem. This does not hold in general for $\epsilon = 0$. The "or missing by less than ϵ" necessary.

The theorem does not require that g have a unique minimizer or even that $g(\infty) = +\infty$. It only requires $\inf g > -\infty$ almost surely. The only tools developed in this paper for proving $g_n \xrightarrow{e} g$ on $\overline{\mathbb{R}^d}$ use the condition $g(\infty) = +\infty$, but this condition is not necessary for convergence of level sets.

Note also that the theorem does not say anything about convergence of argmins, the case $\gamma = 0$. It is not true in general that $A_{n,0} = \operatorname{argmin} g_n$ converges to $A_0 = \operatorname{argmin} g$.

5. Examples. For simple examples of the method, consider asymptotics for the sample median.

Example 1. First consider the usual case where the true probability distribution has a density f that is continuous and strictly positive at the median. Let m be the true median, then the objective function for the convex minimization problem for calculating the median from an i. i. d. sample $X_1, \ldots, X_n$ is

$$g_n(s) = \sum_{i=1}^n \left(|X_i - m - \delta_n s| - |X_i - m|\right) \tag{5.1}$$

where $\delta_n = n^{-1/2}$. This is minimized at $t_n = \delta_n^{-1}(\hat{m}_n - m)$ where $\hat{m}_n$ is the sample median.

Without loss of generality assume that $m = 0$ and write $f(0) = c$. Take first the case $s \geq 0$, where (5.1) becomes

$$g_n(s) = \delta_n s \sum_{i=1}^n \left\{1_{[X_i<0]} - 1_{[X_i>0]}\right\} + 2\sum_{i=1}^n (\delta_n s - X_i)\, 1_{[0<X_i<\delta_n s]} \tag{5.2}$$

The first term on the right hand side converges in law to Zs where Z is a standard normal random variable. Define $Y = 2(u - X)1_{[0<X<u]}$. Then $E(Y) = cu^2 + o(u^2)$ and $\operatorname{Var}(Y) = \frac{4}{3}cu^3 + o(u^3)$, so the second term in (5.2) has expectation $cs^2 + o(1)$ and variance $O(n^{-1/2})$ and converges in probability to cs^2. Thus (5.1) converges in law to

$$g(s) = Zs + cs^2 \tag{5.3}$$

for $s \geq 0$. By symmetry (5.1) also converges in law to (5.3) for $s < 0$. By Lemma 3.3 this implies convergence for all finite-dimensional distributions.

Hence t_n converges to the minimizer of (5.3), which is $t = -Z/2c$, and this gives the usual asymptotic distribution for the median: $\delta_n^{-1}(\hat{m}_n - m)$ converges to $N(0, 1/4c^2)$.

Example 2. Now consider cases where the density is either zero or infinite at the median. To keep the discussion simple assume that the density is symmetric about zero and of the form $f(x) = c|x|^\alpha$ (with $\alpha > -1$) in some neighborhood of zero. The case ($\alpha > 0$) has been done in more generality by Kiefer (1970). Now Y defined in Example 1 has

$$
\begin{aligned}
E(Y) &= \frac{2c}{(\alpha+1)(\alpha+2)} u^{\alpha+2} \\
E(Y^2) &= \frac{8c}{(\alpha+1)(\alpha+2)(\alpha+3)} u^{\alpha+3}
\end{aligned}
$$

Then a rough calculation makes (5.2) approximately

$$
\delta_n s n^{1/2} Z + n \frac{2c}{(\alpha+1)(\alpha+2)} (\delta_n s)^{\alpha+2} \tag{5.4}
$$

which will have both terms about the same size if we choose $\delta_n = n^{-1/2(\alpha+1)}$. Then the size of (5.4) is $n^{\alpha/2(\alpha+1)}$, so we need to scale by this to get nontrivial asymptotics, choosing our objective function to be

$$
g_n(s) = n^{-\alpha/2(\alpha+1)} \sum_{i=1}^{n} (|X_i - \delta_n s| - |X_i|)
$$

Then considering the case $s \geq 0$, this becomes (much like Example 1)

$$
\begin{aligned}
g_n(s) = n^{-1/2} s \sum_{i=1}^{n} &\left\{ 1_{[X_i < 0]} - 1_{[X_i > 0]} \right\} \\
&+ 2n^{-\alpha/2(\alpha+1)} \sum_{i=1}^{n} (\delta_n s - X_i)\, 1_{[0 < X_i < \delta_n s]}
\end{aligned} \tag{5.5}
$$

As in Example 1 the first term converges in law to Zs. The second term now has expectation $2cs^{\alpha+2}/(\alpha+1)(\alpha+2)$ and variance $O(\delta_n)$ so (5.5) converges in law to

$$
g(s) = Zs + \frac{2c}{(\alpha+1)(\alpha+2)} |s|^{\alpha+2}, \tag{5.6}
$$

and as in Example 1 we also have $g_n(s) \xrightarrow{\mathcal{L}} g(s)$ for $s < 0$ by symmetry. Again Lemma 3.3 now implies convergence of all finite-dimensional distributions. Since (5.6) achieves its minimum at

$$
t = -\operatorname{sign}(Z) \left(\frac{|Z|(\alpha+1)}{2c} \right)^{1/(\alpha+1)}
$$

this gives us the asymptotic distribution of $t_n = \delta_n^{-1} \hat{m}_n$. Note that for $\alpha \neq 0$ the distribution of t is not normal, although the distribution of $g(s)$ for fixed s is normal. For example, if $\alpha = -\frac{1}{2}$, the rate is $\delta_n = n^{-1}$ and the asymptotic distribution is

$$
t = -\operatorname{sign}(Z) \frac{Z^2}{16c^2}
$$

which is symmetric about zero with each side a χ^2_1 random variable divided by $16c^2$.

Example 3. Consider a sequence of exponential families with the same minimal canonical parameter space $\mathbb{R}^d$. For each $n \in \mathbb{N}$ there is a probability space

$(\Omega_n, \mathcal{A}_n, \mu_n)$ and a random vector $t_n : \Omega_n \to \mathbb{R}^d$, the canonical statistic of the nth family, such that $\mu_n \circ t_n^{-1}$ is not concentrated on a proper affine subspace of $\mathbb{R}^d$. Define

$$
\begin{aligned}
c_n(\theta) &= \int e^{t_n(x)'\theta} \mu_n(dx), \qquad \theta \in \mathbb{R}^d \\
\Theta_n &= \operatorname{dom} c_n = \left\{ \theta \in \mathbb{R}^d : c_n(\theta) < +\infty \right\} \\
f_{n,\theta}(x) &= \frac{1}{c_n(\theta)} e^{t_n(x)'\theta}, \qquad x \in \Omega_n,\ \theta \in \Theta_n \\
\mathcal{F}_n &= \{ f_{n,\theta} : \theta \in \Theta_n \}
\end{aligned}
$$

Then each $\mathcal{F}_n$ is a full exponential family of densities with respect to μ_n. The families are not necessarily defined on the same sample space and there has been no mention of i. i. d. sampling. There is at this point no connection between the $\mathcal{F}_n$ except that they have natural parameter spaces of the same dimension.

Write $k_n = \log c_n$, and suppose θ is a point in the interior of Θ_n. Then, as is well known from the theory of exponential families,

$$
\tau_n(\theta) = \nabla k_n(\theta) = E_{n,\theta} t_n(X) = \frac{1}{c_n(\theta)} \int t_n(x) e^{t_n(x)'\theta} \mu_n(dx)
$$

For a numerical sequence $\delta_n \to 0$, consider the moment generating function of $\delta_n \left[t_n(X) - \tau_n(\theta) \right]$ where X has density $f_{n,\theta}$ with respect to μ_n

$$
\begin{aligned}
M_{n,\theta}(s) &= \int e^{\delta_n [t_n(x) - \tau_n(\theta)]' s} \frac{1}{c_n(\theta)} e^{t_n(x)'\theta} \mu_n(dx) \\
&= c_n(\theta + \delta_n s) e^{-\delta_n \tau_n(\theta)' s} / c_n(\theta)
\end{aligned}
\tag{5.7}
$$

Theorem 5.1. *Suppose θ_0 lies in the interior of Θ_n for all n and we observe X_n having density f_{n,θ_0} with respect to μ_n for each n. Suppose M_{n,θ_0} defined by (5.7) converges pointwise on a dense set to $M : s \mapsto \exp(\frac{1}{2} s' K s)$ for some positive definite matrix K. Let $\hat{\theta}_n$ be the maximizer of the log likelihood*

$$
l_n(\theta) = t_n(X_n)'\theta - k_n(\theta)
$$

or an approximate maximizing sequence if the maximizer is not well defined for all n. Then

$$
\delta_n^{-1}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} N(0, K^{-1}).
$$

Proof. The assumption of convergence of moment generating functions implies $\delta_n \left[t_n(X) - \tau_n(\theta_0) \right] \xrightarrow{\mathcal{L}} N(0, K)$. Write

$$
\begin{aligned}
g_n(s) &= l_n(\theta_0) - l_n(\theta_0 + \delta_n s) \\
&= -\delta_n t_n(X_n)' s + \log \frac{c_n(\theta_0 + \delta_n s)}{c_n(\theta_0)} \\
&= -\delta_n \left[t_n(X_n) - \tau_n(\theta_0) \right]' s + \log M_{n,\theta}(s) \\
&\xrightarrow{\mathcal{L}} Z' s + \tfrac{1}{2} s' K s \\
&= g(s)
\end{aligned}
$$

Lemma 3.3 implies convergence of all finite-dimensional distributions, and the minimizer $t = K^{-1} Z$ of g is unique. Thus g_n epiconverges to g on $\overline{\mathbb{R}^d}$, and the minimizer $t_n = \delta_n^{-1}(\hat{\theta}_n - \theta_0)$ of g_n converges in law to t, which has the distribution $N(0, K^{-1})$. □

The theorem is trivial but does simplify otherwise elaborate calculations. In the context of Markov random fields it implies Theorems 2 and 3 of Gidas (1993). In Theorem 2 Gidas uses a different condition than the moment generating function convergence assumed here, but the moment generating function convergence is established during his proof, in his equation (43), and hence is no stronger than the assumptions for his theorem (in nice situations, the assumptions are equivalent).

Theorem 3 in Gidas (1993) involves the maximum likelihood estimate $\hat{\beta}_n$ of the canonical parameter β of a one-parameter Ising model observed in a square window containing $N_n = n^2$ pixels and having the true parameter $\beta_c = \frac{1}{2}\sinh^{-1} 1 \approx .44$, the critical value where the infinite-volume Gibbs distribution undergoes phase transition. Gidas's theorem says

$$\sqrt{\frac{N_n}{\log N_n}}(\hat{\beta}_n - \beta_c) \log |\hat{\beta}_n - \beta_c| \xrightarrow{\mathcal{L}} N(0, \sigma^2) \tag{5.8}$$

for some $\sigma^2 > 0$ as $n \to \infty$. Gidas's proof cites a central limit theorem for the canonical statistic due to Pickard (1977). But Pickard's proof uses moment generating function convergence, so Theorem 5.1 says the limit must be normal

$$\sqrt{N_n \log N_n}(\hat{\beta}_n - \beta_c) \xrightarrow{\mathcal{L}} N\left(0, \tfrac{\pi}{4}\right). \tag{5.9}$$

(The variance $\frac{\pi}{4}$ is derived below.) Although (5.8) and (5.9) look very different, they are actually both correct. If X_n is the left hand side of (5.8) and Y_n is the left hand side of (5.9), then

$$X_n = Y_n \left(-\frac{1}{2} - \frac{\log \log N_n}{2 \log N_n} + \frac{\log Y_n}{\log N_n} \right) = -\frac{1}{2} Y_n + o_p(1).$$

Pickard actually gives a theorem for a two-parameter Ising model that has different coupling in the horizontal and vertical directions. The canonical statistic $t_n(x)$ is bivariate, the first component being the sum of products of spins for horizontal nearest neighbor pairs and the second component being the sum for vertical pairs. The canonical parameter β is also bivariate and is critical if $\beta = (\frac{1}{2}\tanh^{-1}\cos\psi, \frac{1}{2}\tanh^{-1}\sin\psi)$ for $0 < \psi < \frac{\pi}{2}$. The case $\psi = \frac{\pi}{4}$ gives the symmetric interaction $\beta_1 = \beta_2 \approx .44$.

If $t_n(x)$ is the canonical statistic for a model observed in a rectangular window with N_n pixels and the ratio of the sides of the window is bounded away from zero and infinity as $n \to \infty$, then Pickard shows that the moment generating function convergence that is the condition for Theorem 5.1 holds with $\delta_n = (N_n \log N_n)^{-1/2}$ and

$$K = \frac{1}{\pi} \begin{pmatrix} \tan\psi & 1 \\ 1 & \cot\psi \end{pmatrix} \tag{5.10}$$

This is equation (49) in Pickard (1977). The one-parameter case is obtained from the two-parameter case by setting $t(x) = t_1(x) + t_2(x)$ and $\beta = \beta_1 = \beta_2$. If $M_{n,\theta}(s_1, s_2)$ is the moment generating function (5.7) for the two-parameter case, then $M_{n,\theta}(s, s)$ is the moment generating function for the one-parameter case, which converges to $s \mapsto \frac{1}{2}Ks^2$ with $K = \frac{4}{\pi}$. This implies (5.9). This result was presumably understood by Pickard, although he did not state it formally. Pickard also pointed out that his result (5.10) does not lead to a central limit theorem for $\hat{\beta}_n$ in the two-parameter problem because K is singular.

Example 4. This is a simple counterexample due to Bill Sudderth that shows that Lemma 3.3 is not the whole story. Let $X_n \xrightarrow{\mathcal{L}} X$ and $Y_n \xrightarrow{\mathcal{L}} Y$ be sequences of nonnegative random variables, with the limits X and Y being almost surely strictly positive. Define

$$g_n(s) = \begin{cases} -X_n s, & s \leq 0 \\ Y_n s, & s \geq 0 \end{cases}$$

and g by the same formula without the n's. Then g_n and g are convex, and $g(\infty) = +\infty$ with probability one. For each $s \in \mathbb{R}$, we have $g_n(s) \xrightarrow{\mathcal{L}} g(s)$, but this does not imply that g_n converges weakly to g in $\overline{\mathbb{R}}^D$. That requires

$$(g_n(-1), g_n(1)) = (X_n, Y_n) \xrightarrow{\mathcal{L}} (X, Y) = (g(-1), g(1)).$$

So checking only one-dimensional marginals is not enough.

6. Discussion. This paper and Geyer (1994a, 1994b) show the value of thinking of asymptotics of optimization problems using epiconvergence. One first shows epiconvergence of objective functions for optimization problems, either almost sure epiconvergence when discussing consistency or epiconvergence in law when discussing convergence in law. Then the relevant convergence of minimizers or approximate minimizers follows as a simple corollary using Proposition 2.1, using the Prohorov and Skorohod theorems to reduce the stochastic case to the nonstochastic case. One might think that this method only adds an extra step to the proofs and does nothing that could not be done by ad hoc methods, but in practice epiconvergence can greatly clarify and simplify proofs. It can also lead to sharper results by focusing attention on minimal conditions for convergence of optimizers.

In the case of asymptotics of convex objective functions studied in this paper, the method achieves remarkable simplicity. The only probabilistic calculations required to apply Theorem 3.2 are verifying convergence in law of $(g_n(s_1), \ldots, g_k(s_k))$ to $(g(s_1), \ldots, g(s_k))$ for fixed $s_1, \ldots, s_k$. There is no need to consider derivatives, even weak derivative notions like differentiability in quadratic mean. The theorem does not require asymptotic normality or independent sampling and also handles constraints.

Epiconvergence theory also suggests new results like Theorem 4.3, which would be difficult even to state without using the terminology of Painlevé-Kuratowski convergence and epiconvergence and even more difficult to prove. Using epiconvergence, however, the theorem is a simple application of known results of variational analysis (Proposition 4.1 and its corollary).

APPENDIX

Proof of Proposition 2.1. The proposition stated is almost the same as Theorem 1.10 in Attouch (1984). The only difference is that Attouch requires $g_n(t_n) \leq \inf g_n + \eta_n$, making no allowance for the case $\inf g_n = -\infty$. The version here is obtained from Attouch's version by applying Lemma 2.70 in Attouch: $g_n \xrightarrow{e} g$ if and only if $h \circ g_n \xrightarrow{e} h \circ g$ where h is any increasing homeomorphism from $\overline{\mathbb{R}}$ to some closed interval in $\mathbb{R}$. Define h by $h(x) = 1 + x$, $x \geq 0$ and $h(x) = e^x$, $x < 0$. Then $h(g_n(t_n)) \leq \inf h \circ g_n + \epsilon_n$, where $\epsilon_n = \max(e^{r_n}, \eta_n) \to 0$. This implies that $h(g_n(t_n)) \to h(g(t))$, which implies $g_n(t_n) \to g(t) = \inf g$. $\square$

Proof of Proposition 2.2. That g is convex follows from Theorem 10.8 in Rockafellar (1970). That $g_n \xrightarrow{e} g$ follows from Theorem 7.12 in Rockafellar and Wets (forthcoming). By the same theorem g_n converges uniformly to g on every compact set K contained in $\operatorname{int}(\operatorname{dom} g)$. Hence g_n is eventually in $\Gamma(\mathbb{R}^d)$. □

Proof of Proposition 2.3. Since g is lower semicontinuous, every level set is closed. (a) is equivalent to every level set being bounded, hence compact. Since $g(x) < +\infty$ for some x, some level set is nonempty. Thus (a) implies (c). By Theorems 27.1 and 8.4 and in Rockafellar (1970), all nonempty level sets have the same recession cone, so if any is nonempty and compact, all are compact. Since g is lower semicontinuous, its infimum is achieved if any level set is nonempty and compact. Thus (c) implies (d), which trivially implies (b). Conversely (b) trivially implies (c) which we have already established implies (d), and (d) implies that every level set is compact, which is equivalent to (a). □

Proof of Proposition 2.4. Proposition 4.1 implies

$$\begin{aligned} &\limsup_n \operatorname{lev}_\alpha g_n \subset \operatorname{lev}_\alpha g \\ &\limsup_n \operatorname{lev}_\alpha g_n \supset \operatorname{lev}_\beta g, \qquad \beta < \alpha \end{aligned}$$

By the hit and miss criterion (Theorem 2.75 in Attouch, 1984) $\operatorname{lev}_\alpha g_n$ eventually misses any compact set that misses $\operatorname{lev}_\alpha g$. The assumption $g(\infty) = +\infty$ implies that $\operatorname{lev}_\alpha g$ is compact, hence it misses the surface of any large enough ball, hence so eventually does $\operatorname{lev}_\alpha g_n$. By the same criterion $\operatorname{lev}_\alpha g_n$ must eventually hit every neighborhood of $\operatorname{lev}_\beta g$ for $\beta < \alpha$ hence, $\operatorname{lev}_\alpha g_n$, being convex, must eventually be inside every sufficiently large ball. In other words g_n is ultimately *equi–level bounded:* for every $R \in \mathbb{R}$ there is a compact set K such that

$$\liminf_{n\to\infty} \inf_{s \notin K} g_n(s) > R. \tag{A.1}$$

Using the definition of epilimit inferior at a point (Attouch, 1984, p. 26) relation (A.1) implies

$$\text{e-}\liminf_{n\to\infty} g_n(\infty) = \sup_{W \in \mathcal{N}(\infty)} \liminf_{n\to\infty} \inf_{s \in W} g_n(s) = +\infty = g(\infty). \tag{A.2}$$

So g_n epiconverges to g on $\mathbb{R}^d$ by Proposition 2.2 and it epiconverges to g at ∞ by (A.2), thus $g_n \xrightarrow{e} g$ on $\overline{\mathbb{R}^d}$. Let t_n be any sequence satisfying (2.3), then since $\overline{\mathbb{R}^d}$ is compact there exists a convergent subsequence $t_{n_k} \to t^*$, and by Proposition 2.1 $g_{n_k}(t_{n_k}) \to g(t^*) = \inf g$. Since g is proper, level bounded, and lower semicontinuous, it achieves its infimum. Hence $\inf g$ is finite, and we cannot have $g_n(t_n) \to \inf g$ unless $\inf g_n$ is eventually finite and $|g_{n_k}(t_{n_k}) - \inf g_n| \to 0$. Thus $\inf g_n \to \inf g$, and $g_n(t_n) \to \inf g$. □

Proof of Proposition 2.5. By Proposition 2.3 the assumption that g has a unique minimizer implies $g(\infty) = +\infty$. Every subsequence of t_n has a subsubsequence that converges in $\overline{\mathbb{R}^d}$, so by Proposition 2.1 the limit must minimize g and hence must be equal to t. Since every subsequence of t_n has a convergent subsubsequence that converges to t, the whole sequence converges to t. Hence also $g_n(t_n) \to g(t)$. □

Lemma A.1. *For $x \in \mathbb{R}^d$ and U an open set in $\overline{\mathbb{R}}$, the set*

$$W = \{\, g \in F(\mathbb{R}^d) : g(x) \in U \,\} \tag{A.3}$$

is a measurable subset of $F(\mathbb{R}^d)$.

Proof. Since U is a countable union of open intervals, it is enough to prove

$$\begin{aligned} W &= \{\, g \in F(\mathbb{R}^d) : a < g(x) \,\} \\ &= \{\, g \in F(\mathbb{R}^d) : \operatorname{epi} g \cap \{(x, a)\} = \varnothing \,\} \end{aligned}$$

is measurable. The form in the second line shows it is actually open. □

Corollary A.2. *The restriction map $u : F(\mathbb{R}^d) \to \overline{\mathbb{R}}^D$ defined by $g \mapsto g|D$ is measurable.*

Proof. Sets of the form

$$W' = \{\, g \in \overline{\mathbb{R}}^D : g(x) \in U \,\}$$

for $x \in D$ and U open in $\overline{\mathbb{R}}$ form a subbasis for the topology of $\overline{\mathbb{R}}^D$, and $u^{-1}(W') = W$ with W given by (A.3). □

Lemma A.3. *$\Gamma(\mathbb{R}^d)$ is a measurable subset of $F_c(\mathbb{R}^d)$, and $\Gamma(\overline{\mathbb{R}^d})$ is a measurable subset of $F_c(\overline{\mathbb{R}^d})$.*

Proof. A function $g \in \Gamma(\mathbb{R}^d)$ is actually bounded on some open set, because $\operatorname{int}(\operatorname{dom} g)$ is nonempty, and g is continuous on $\operatorname{int}(\operatorname{dom} g)$ (Rockafellar, 1970, Theorem 10.1), hence bounded on compact subsets of $\operatorname{int}(\operatorname{dom} g)$.

Let D be a countable dense set in $\mathbb{R}^d$. For any compact set K in $\mathbb{R}^d$ with nonempty interior and any real number R, define

$$W = \{\, g \in F_c(\mathbb{R}^d) : -R < g(x),\ x \in K \text{ and } g(x) < R,\ x \in K \cap D \,\}$$

W is measurable by Lemma A.1 and the hit and miss criterion. By lower semicontinuity for $y \in \operatorname{int}(K)$ and $g \in W$

$$g(y) = \sup_{\epsilon \to 0} \inf_{\substack{x \in \mathbb{R}^d \\ \|x-y\| < \epsilon}} g(x) \le R$$

So $|g(x)| \le R$ for $x \in \operatorname{int}(K)$. The union of a countable family of sets of this form with rational R and balls K with rational radii and centers with rational coordinates is measurable and is equal to $\Gamma(\mathbb{R}^d)$.

The same proof works for $\Gamma(\overline{\mathbb{R}^d})$ because every compact set in $\mathbb{R}^d$ is also compact in $\overline{\mathbb{R}^d}$. □

Lemma A.4. *Let $u : \Gamma(\mathbb{R}^d) \to \overline{\mathbb{R}}^D$ be the restriction map $g \mapsto g|D$. Then u is a measurable bijection onto its range.*

Proof of Lemma A.4. First, we need to show that u is injective, that every $g \in \Gamma(\mathbb{R}^d)$ is determined by its values on D. Since $W = \operatorname{int}(\operatorname{dom} g)$ is nonempty, W is determined by $W \cap D$, and since g is continuous on W (Rockafellar, 1970, Theorem 10.1) $g|W$ is determined by its values on D, and since $g|W$ determines g (Rockafellar, 1970, Theorem 7.4), we conclude that u is injective. By Lemma A.2 u is measurable, and by Proposition 2.2 u^{-1} is actually continuous (pointwise convergence implying epiconvergence). □

Acknowledgement. This paper owes much to the forthcoming book by Rockafellar and Wets and the course taught from it by Terry Rockafellar. Roger Wets supplied the citations to the current version. Joe Eaton and Bill Sudderth helped with the examples. Jon Wellner supplied the reference to Kiefer (1970). The referees pointed out the necessity for the convergence of all finite-dimensional distributions for pointwise convergence in law. Basilis Gidas explained the equivalence of (5.8) and (5.9).

REFERENCES

Attouch, H. (1984). *Variational Convergence of Functions and Operators.* Pitman, Boston.

Aubin, J. P. and Frankowska, H. (1990). *Set-Valued Analysis.* Birkhäuser, Boston.

Beer, G., Rockafellar, R. T. and Wets, R. J.-B. (1992). A characterization of epi-convergence in terms of convergence of level sets. *Proc. Amer. Math. Soc.* bf 116 753–761.

Geyer, C. J. (1994a). On the convergence of Monte Carlo maximum likelihood calculations. *J. Roy. Statist. Soc. Ser. B* **56** 261–274.

Geyer, C. J. (1994b). On the asymptotics of constrained M-estimation. *Ann. Statist.* **22** 1993–2010.

Gidas, B. (1993). Parameter estimation for Gibbs distributions from fully observed data. In *Markov Random Fields: Theory and Application* (R. Chellappa and A. Jain, eds.) 471–498. Academic Press, New York.

Haberman, S. J. (1989). Concavity and estimation. *Ann. Statist.* **17** 1631–1661.

Kiefer, J. (1970). Old and new methods for studying order statistics and sample quantiles. In *Nonparametric Techniques in Statistical Inference* (M. L. Puri, ed.) 349–357. Cambridge University Press.

Knight, K. (1989). Limit theory for the autoregressive estimates in an infinite-variance random walk. *Canad. J. Statist.* **17** 261–278.

Niemiro, W. (1992). Asymptotics for M-estimators defined by convex minimization. *Ann. Statist.* **20** 1514–1533.

Pickard, D. K. (1977). Asymptotic inference for an Ising lattice, II *Adv. in Appl. Probab.* **9** 476–501.

Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer-Verlag, New York.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory* **7** 186–199.

Rockafellar, R. T. (1970). *Convex Analysis.* Princeton University Press.

Rockafellar, R. T. and Wets, R. J.-B. (forthcoming). *Variational Analysis.* Springer-Verlag, New York.

School of Statistics
University of Minnesota
270 Vincent Hall
206 Church St. S. E.
Minneapolis, MN 55455