2|18|25 STOR834 p. 42

1 Translation of De Haan-Stadtmüller Result to Second-Order Approximations for Extreme Value Distributions

Suppose $n(1-F(a_n x+b_n))=y$ — the objective is to derive an asymptotic expression for y as a function of x.

We assume $b_n=U(n),\ a_n=a(n)$, so this equation may be rewritten

$$a_n x+b_n \;=\; a(n)x+U(n) \;=\; U\left(\frac{n}{y}\right)$$

and hence

$$\begin{aligned} x &= \frac{U\left(\frac{n}{y}\right)-U(n)}{a(n)} \\ &= \frac{y^{-\xi}-1}{\xi}+A(n)H(y^{-1})+o(A(n)). \end{aligned} \tag{1}$$

First order solution: ignore $A(n)$, solve $x=\frac{y^{-\xi}-1}{\xi}$ to get

$$y \;=\; (1+\xi x)^{-1/\xi}.$$

Second-order solution: assume

$$\begin{aligned} y &= (1+\xi x)^{-1/\xi}+\epsilon \\ &= (1+\xi x)^{-1/\xi}\left(1+\epsilon(1+\xi x)^{1/\xi}\right) \end{aligned}$$

for some ϵ that we have to determine (asymptotically).

We calculate

$$\begin{aligned} y^{-\xi} &= (1+\xi x)\left(1-\xi\epsilon(1+\xi x)^{1/\xi}+\ldots\right) \\ &= 1+\xi x-\xi\epsilon(1+\xi x)^{1+1/\xi}+\ldots) \end{aligned}$$

where (here and subsequently) $\ldots$ denotes terms that are of smaller order than those considered. Hence

$$\frac{y^{-\xi}-1}{\xi} \;=\; x-\epsilon(1+\xi x)^{1+1/\xi}+o(\epsilon^2).$$

Comparing with (1), we deduce

$$\epsilon(1+\xi x)^{1+1/\xi} \;\sim\; A(n)H(y^{-1}) \;\sim\; A(n)H\left((1+\xi x)^{1/\xi}\right).$$

Hence

$$\epsilon \;\sim\; A(n)(1+\xi x)^{-1-1/\xi}H\left((1+\xi x)^{1/\xi}\right).$$

Thus our result is

$$n(1-F(a_n x+b_n)) \;=\; (1+\xi x)^{-1/\xi}+A(n)(1+\xi x)^{-1-1/\xi}H\left((1+\xi x)^{1/\xi}\right)+o(A(n)).$$

STOR834 p.43

By the usual argument of replacing $1-F(a_nx+b_n)$ by $-\log F(a_nx+b_n)$ (with an error of $O(n^{-1})$), we also deduce that provided $nA(n)\to\infty$,

$$F^n(a_nx+b_n) = \exp\left\{(1+\xi x)^{-1/\xi}+A(n)(1+\xi x)^{-1-1/\xi}H\left((1+\xi x)^{1/\xi}\right)+o(A(n))\right\}.$$

The stipulation $nA(n)\to\infty$ is needed because, without it, the error of $O(n^{-1})$ is as large or larger than $O(A(n))$; however, as noted earlier, cases where the error rate is $O(n^{-1})$ or smaller are rather few and generally not of interest since the rate of convergence is so fast.

2 Derivation of De Haan-Stadtmüller Result in Prior Cases

2.1 Example Model 1

Assumption:

$$1-F(y) = cy^{-\alpha}+dy^{-\alpha-\beta}+o(y^{-\alpha-\beta}),\ y\to\infty.$$

To calculate an approximation to $U(t)$, we need to solve $\frac{1}{t}=1-F(y)$ so $y=(ct)^{1/\alpha}(1+\epsilon)$ for some ϵ. Then $y^{-\alpha}=\frac{1}{ct}(1-\alpha\epsilon+\ldots)$ so $t^{-1}=t^{-1}(1-\alpha\epsilon)+d(ct)^{-1-\beta/\alpha}+\ldots$ so $\epsilon=\frac{d}{\alpha}c^{-1\beta/\alpha}t^{-\beta/\alpha}$ and hence

$$U(t) = (ct)^{1/\alpha}\left(1+\frac{d}{\alpha}c^{-1-\beta/\alpha}t^{-\beta/\alpha}+\ldots\right)$$

where $\ldots$ means smaller order terms that are omitted.

We proceed by calculating

$$\begin{aligned}U(tx)-U(t) &= (ctx)^{1/\alpha}\left\{1+\frac{d}{\alpha}c^{-1-\beta/\alpha}(tx)^{-\beta/\alpha}\right\}-(ct)^{1/\alpha}\left\{1+\frac{d}{\alpha}c^{-1-\beta/\alpha}t^{-\beta/\alpha}\right\}+o(t^{1/\alpha-\beta/\alpha})\\ &= t^{1/\alpha}\left\{(cx)^{1/\alpha}-c^{1/\alpha}\right\}+t^{1/\alpha-\beta/\alpha}\cdot c^{1/\alpha-1-\beta/\alpha}\frac{d}{\alpha}(x^{1/\alpha-\beta/\alpha}-1)+o(t^{1/\alpha-\beta/\alpha}). \quad (2)\end{aligned}$$

Side calculations: with $\xi=1/\alpha,\ \rho=-\beta/\alpha$, define

$$\begin{aligned}G(x) &= \frac{x^\xi-1}{\xi} = \alpha(x^{1/\alpha}-1),\\ H(x) &= \frac{1}{\rho}\left(\frac{x^{\xi+\rho}-1}{\xi+\rho}-\frac{x^\xi-1}{\xi}\right) = -\frac{\alpha}{\beta}\left(\frac{x^{1/\alpha-\beta/\alpha}-1}{1/\alpha-\beta/\alpha}-G(x)\right)\end{aligned}$$

and hence

$$x^{1/\alpha-\beta/\alpha}-1 = -\frac{\beta}{\alpha}(1/\alpha-\beta/\alpha)H(x)+(1/\alpha-\beta/\alpha)G(x).$$

Substituting in (2),

$$\begin{aligned}U(tx)-U(t) &= t^{1/\alpha}\frac{c^{1/\alpha}}{\alpha}G(x)+t^{1/\alpha-\beta/\alpha}\cdot c^{1/\alpha-1-\beta/\alpha}\frac{d}{\alpha}\left\{-\frac{\beta}{\alpha}(1/\alpha-\beta/\alpha)H(x)+(1/\alpha-\beta/\alpha)G(x)\right\}\\ &\quad +o(t^{1/\alpha-\beta/\alpha}). \quad (3)\end{aligned}$$

STOR 834 p. 44

We are trying to get a limit of the form $\frac{\frac{U(tx)-U(t)}{a(t)}-G(x)}{A(t)} \to H(x)$ which would imply

$$U(tx)-U(t) = G(x)a(t)+H(x)A(t)a(t)+o(A(t)a(t)).$$

Equating coefficients of $G(x)$ and $H(x)$ in (3), it suffices to take

$$a(t) = t^{1/\alpha}\frac{c^{1/\alpha}}{\alpha}+t^{1/\alpha-\beta/\alpha}\cdot c^{1/\alpha-1-\beta/\alpha}\frac{d}{\alpha}(1/\alpha-\beta/\alpha),$$
$$A(t)a(t) = -t^{1/\alpha-\beta/\alpha}\cdot c^{1/\alpha-1-\beta/\alpha}\frac{d}{\alpha}\frac{\beta}{\alpha}(1/\alpha-\beta/\alpha).$$

Simplifying a bit, an equivalent expression for $A(t)$ is

$$A(t) = -t^{-\beta/\alpha}\cdot c^{-1-\beta/\alpha}\frac{d\beta}{\alpha}(1/\alpha-\beta/\alpha).$$

It should be noted that in the case $\beta=1$ (but only that case!) the expression for $H(x)$ reduces to $\frac{\alpha}{\beta}G(x)$, in other words, it is a multiple of $G(x)$, which earlier we excluded from the theory. This confirms our earlier result that the rate of convergence can be improved when $\beta=1$, but not otherwise.

2.2 Example Model 2

Assumption:

$$1-F(y) = c|y|^{\alpha}+d|y|^{\alpha+\beta}+o(|y|^{\alpha+\beta}),\ y\uparrow 0.$$

In this case if we solve $1-F(y)=1/n$ we find

$$U(n) = -(nc)^{-1/\alpha}-\frac{d}{\alpha}n^{-1/\alpha-\beta/\alpha}c^{-1-1/\alpha-\beta/\alpha}+o(n^{-1/\alpha-\beta/\alpha})$$

The same formula holds if n is replaced by real $t\to\infty$ so

$$U(tx)-U(t) = -(tc)^{-1/\alpha}(x^{-1/\alpha}-1)-\frac{d}{\alpha}t^{-1/\alpha-\beta/\alpha}c^{-1-1/\alpha-\beta/\alpha}(x^{-1/\alpha-\beta/\alpha}-1)+o(t^{-1/\alpha-\beta/\alpha}) \qquad (4)$$

In this case with $\xi=-1/\alpha$, $\rho=-\beta/\alpha$, we have

$$G(x) = -\alpha(x^{-1/\alpha}-1),\quad H(x) = \frac{\alpha^2}{\beta(1+\beta)}(x^{-1/\alpha-\beta/\alpha}-1)-\frac{\alpha^2}{\beta}(x^{-1/\alpha}-1)$$

and hence

$$x^{-1/\alpha}-1 = -\frac{1}{\alpha}G(x),\quad x^{-1/\alpha-\beta/\alpha}-1 = \frac{\beta(1+\beta)}{\alpha}H(x)-\frac{1+\beta}{\alpha}G(x).$$

Substituting into (4),

$$U(tx)-U(t) = \frac{1}{\alpha}(tc)^{-1/\alpha}G(x)-\frac{d}{\alpha}t^{-1/\alpha-\beta/\alpha}c^{-1-1/\alpha-\beta/\alpha}\left\{\frac{\beta(1+\beta)}{\alpha}H(x)-\frac{1+\beta}{\alpha}G(x)\right\}+o(t^{-1/\alpha-\beta/\alpha}). \qquad (5)$$

To write this in the form $a(t)G(x)+a(t)A(t)H(x)+o(a(t)A(t))$, it would suffice to take

$$a(t) = \frac{1}{\alpha}(tc)^{-1/\alpha}+\frac{d(1+\beta)}{\alpha^2}t^{-1/\alpha-\beta/\alpha}c^{-1-1/\alpha-\beta/\alpha},\quad A(t) = \frac{d\beta(1+\beta)}{\alpha}t^{-\beta/\alpha}c^{-1-\beta/\alpha}. \qquad (6)$$

STOR 834 p.45

2.3 Example Model 3

Assumption:

$$F(t) \quad = \quad \Phi(t) \quad \text{(standard normal cdf)}, \quad -\infty < t < \infty.$$

Recall the formula that $1 - \Phi(b_n) = \frac{1}{n}$ leads to

$$b_n \quad = \quad \sqrt{2\log n} - \frac{1}{2\sqrt{2\log n}}(\log 4\pi + \log\log n) + o\left(\frac{\log\log n}{\sqrt{2\log n}}\right),$$

where we write t in place of n and $U(t) = b_t$, we deduce

$$U(t) \quad = \quad \sqrt{2\log t} - \frac{\log 4\pi + \log\log t}{2\sqrt{2\log t}} + o\left(\frac{\log\log t}{\sqrt{2\log t}}\right),$$

and hence

$$U(tx) - U(t) \quad = \quad \sqrt{2\log tx} - \sqrt{2\log t} - \frac{\log\log tx - \log\log t}{2\sqrt{(2\log t}} + o\left(\frac{\log\log t}{\sqrt{2\log t}}\right).$$

Side calculations: as $t \to \infty$ for fixed x,

$$\log\log tx - \log\log t \quad = \quad \frac{\log x}{\log t} - \frac{1}{2}\frac{\log^2 x}{\log^2 t} + O\left(\frac{1}{\log^3 t}\right),$$

$$\sqrt{2\log(tx)} - \sqrt{2\log(t)} \quad = \quad \frac{1}{\sqrt{2\log t}}\left\{\log x - \frac{1}{4}\frac{\log^2 x}{\log t} + O\left(\frac{1}{\log^2 t}\right)\right\}.$$

Also, if we define $\mu_n = -\frac{1}{b_n^2}$, $\psi_n = 1 - \frac{1}{b_n^2}$, $\xi_n = -\frac{1}{b_n^2}$, we have

$$\left(1 + \xi_n \frac{x - \mu_n}{\psi_n}\right)^{-1/\xi_n} \quad = \quad e^{-x}\left\{1 + \xi_n\left(1 + x - \frac{x^2}{2}\right) + O(\xi_n^2)\right\}.$$

probably not needed

Therefore,

$$U(tx) - U(t) \quad = \quad \frac{1}{\sqrt{2\log t}}\left(\log x - \frac{1}{4}\frac{\log^2 x}{\log t} + \dots\right) - \frac{1}{2\sqrt{2\log t}}\frac{\log x}{\log t} + \dots$$

$$= \frac{1}{\sqrt{2\log t}}\left\{\left(1 - \frac{1}{2\log t}\right)\log x - \frac{1}{4}\frac{\log^2 x}{\log t} + \dots\right\}$$

so we set

$$a(t) \quad = \quad \frac{1}{\sqrt{2\log t}}\left(1 - \frac{1}{2\log t}\right),$$

$$A(t) \quad = \quad -\frac{1}{2\log t},$$

$$G(x) \quad = \quad \log x,$$

$$H(x) \quad = \quad \frac{1}{2}\log^2 x.$$

STOR 834 p.46

Last part of 2/18/25 class

Covered previously

It can be readily checked that this implies

$$y = (ct)^{1/\alpha}\left\{1+\frac{d}{\alpha}c^{-1-\beta/\alpha}t^{-\beta/\alpha}+o(t^{-\beta/\alpha})\right\}. \tag{2.38}$$

Therefore, $U(t)$ satisfies the right hand side of (2.38).

Hence,

$$\begin{aligned} U(tx)-U(t) &= (cxt)^{1/\alpha}\left\{1+\frac{d}{\alpha}c^{-1-\beta/\alpha}(xt)^{-\beta/\alpha}\right\}-(ct)^{1/\alpha}\left\{1+\frac{d}{\alpha}c^{-1-\beta/\alpha}t^{-\beta/\alpha}\right\}+o(t^{1/\alpha-\beta/\alpha}) \\ &= (ct)^{1/\alpha}(x^{1/\alpha}-1)+\frac{d}{\alpha}c^{1/\alpha-1-\beta/\alpha}t^{1/\alpha-\beta/\alpha}(x^{1/\alpha-\beta/\alpha}-1)+o(t^{1/\alpha-\beta/\alpha}). \end{aligned} \tag{2.39}$$

If we define $a(t)=\alpha^{-1}(ct)^{1/\alpha}$, we get

$$\frac{U(tx)-U(t)}{a(t)}-\frac{x^{1/\alpha}-1}{1/\alpha} = dc^{-1-\beta/\alpha}t^{-\beta/\alpha}(x^{1/\alpha-\beta/\alpha}-1)+o(t^{-\beta/\alpha}),$$

which, however, does not give the form of limit function we are aiming at.

Therefore, we return to (2.39) and rewrite

$$\begin{aligned} U(tx)-U(t) &= \left\{(ct)^{1/\alpha}+\frac{(1-\beta)d}{\alpha}c^{1/\alpha-1-\beta/\alpha}t^{1/\alpha-\beta/\alpha}\right\}(x^{1/\alpha}-1) \\ &+\frac{(1-\beta)d}{\alpha^2}c^{1/\alpha-1-\beta/\alpha}t^{1/\alpha-\beta/\alpha}\cdot\frac{\alpha}{1-\beta}\left\{x^{1/\alpha-\beta/\alpha}-1-(1-\beta)(x^{1/\alpha}-1)\right\}+o(t^{1/\alpha-\beta/\alpha}). \end{aligned}$$

This seems to be wrong.

Now define $a(t)=\alpha^{-1}\left\{(ct)^{1/\alpha}+\frac{\beta d}{\alpha}c^{1/\alpha-1-\beta/\alpha}t^{1/\alpha-\beta/\alpha}\right\}$, $A(t)=-\frac{(1-\beta)d}{\beta}c^{-1-\beta/\alpha}t^{-\beta/\alpha}$, then

$$\lim_{t\to\infty}\frac{\frac{U(tx)-U(t)}{a(t)}-\frac{x^{1/\alpha}-1}{(1/\alpha)}}{A(t)} = -\frac{\alpha}{\beta}\left(\frac{x^{1/\alpha-\beta/\alpha}-1}{1/\alpha-\beta/\alpha}-\frac{x^{1/\alpha}-1}{1/\alpha}\right).$$

This is precisely of the form (2.37) with $\xi=\frac{1}{\alpha}$, $\rho=-\beta/\alpha$.

2.5 Estimation theory based on second-order asymptotics

We focus here on a paper by Dombry and Ferreira [58], but this is just one of a series of papers going back to the 1980s [224, 61, 56, 74, 57, 174].

Consider an IID random sequence $\{X_i,\ i=1,2,\ldots\}$ where the common distribution function is F. Suppose the observations are grouped into blocks of length m, and let $M_{k,m}=\max\{X_i:(k-1)m+1,\ldots,km\}$ be the maximum of the k'th block. We assume F is in the domain of attraction of the GEV, so that

$$\Pr\left\{\frac{M_{k,m}-b_m}{a_m}\le x\right\} = F^m(a_mx+b_m) \to G_{\xi_0}(x) = \exp\left\{-(1+\xi_0x)_+^{-1/\xi_0}\right\}. \tag{2.40}$$

STOR 834 p.47

for some "true value" ξ_0 which we write that way to distinguish it from the unknown parameter ξ in the following likelihood analysis. We define $g_{\xi_0}(x) = \frac{dG_{\xi_0}(x)}{dx} = (1+\xi_0 x))^{-1/\xi_0-1}\exp\left\{-(1+\xi_0 x))^{-1/\xi_0}\right\}$ defined whenever $1+\xi_0 x > 0$ to be the density of G_{ξ_0} and let

$$\ell(\mu,\psi,\xi;x) = \log\psi + \log g_\xi\left(\frac{x-\mu}{\psi}\right) \tag{2.41}$$

be the log density for arbitrary ξ when the distribution is extended to include a location and scale parameter. The idea is that we treat the block maxima $M_{i,m}$ for $1 \le i \le k$ as if their exact distribution was GEV with parameters $\boldsymbol{\theta} = (\mu,\psi,\xi)$ though we know that for finite m this is only an approximation. Define the log likelihood

$$L_{k,m}(\boldsymbol{\theta}) = \sum_{i=1}^{k} \ell(\boldsymbol{\theta}, M_{i,m}) \tag{2.42}$$

In the following, we shall consider a sequence of sample sizes and block lengths k_n, m_n where both k_n and M_n are indexed by n. We define $\hat{\boldsymbol{\theta}}_n = (\hat{\mu}_n, \hat{\psi}_n, \hat{\xi}_n)$ to be a local maximizer of the log likelihood function, or just the MLE for short, if it satisfies the likelihood equations

$$\frac{\partial L_{k,m}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0 \tag{2.43}$$

and if the hessian matrix $\frac{\partial^2 L_{k,m}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}$ is positive definite at $\hat{\boldsymbol{\theta}}_n$.

Dombry and Ferreira differ slightly from the notation of the previous section by defining $V = (-1/\log F)^{\leftarrow}$ (instead of $U = (1/(1-F))^{\leftarrow}$ as previously, though in most cases the two definitions will lead to the same asymptotics). In that context they assume, first, that there exists a_m such that

$$\lim_{m\to\infty} \frac{V(mx)-V(m)}{a_m} = \frac{x^{\xi_0}-1}{\xi_0} \tag{2.44}$$

and, second, that for some positive function $a(t)$ as $t\to\infty$ and some positive or negative function $A(t)$ as $t\to\infty$ with $\lim_{t\to\infty} A(t) = 0$,

$$\lim_{t\to\infty} \frac{\frac{V(tx)-V(t)}{a(t)} - \frac{x^{\xi_0}-1}{\xi_0}}{A(t)} = \int_1^x \int_1^s s^{\xi_0-1} u^{\rho-1}\,du\,ds = H_{\xi_0,\rho}(x),\ x>0, \tag{2.45}$$

where $\xi_0 > -\frac{1}{2}$, $\rho \le 0$, the function A is regularly varying with index ρ, and $H_{\xi_0,\rho}$ is given by (2.37) with $\xi = \xi_0$. As noted previously, in any case where a limit of the form (2.45) exists, we can without loss of generality, redefining the functions $a(t)$ and $A(t)$ is necessary, assume that the right hand side is $H_{\xi_0,\rho}(x)$ for suitable $\rho \le 0$.

Dombry and Ferreira consider limiting cases as $k = k_n \to \infty$, $m = m_n \to \infty$ where

$$\lim_{n\to\infty} \sqrt{k_n} A(m_n) = \lambda \in \mathbb{R}. \tag{2.46}$$

STOR834 p.48

They define $\boldsymbol{\theta}_0 = (0, 1, \xi_0)$ and then

$$\begin{aligned}
Q_{\xi_0}(s) &= \frac{(-\log s)^{-\xi_0} - 1}{\xi_0}, \ s \in (0,1) \\
\mathbf{b}(\xi_0, \rho) &= \int_0^1 \frac{\partial^2 \ell}{\partial x \partial \boldsymbol{\theta}}(\boldsymbol{\theta}_0, Q_{\xi_0}(s)) H_{\xi_0,\rho}\left(\frac{1}{-\log s}\right) ds, \\
I_{\xi_0} &= -\int_0^1 \frac{\partial^2 \ell}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}(\boldsymbol{\theta}_0, Q_{\xi_0}(s)) ds.
\end{aligned}$$

Note that I_{ξ_0} is the Fisher information for the GEV evaluated at $\boldsymbol{\theta}_0$; this is the same matrix as was shown in Chapter 1 following [185].

With these preliminaries, Theorem 2.2 of [58] states:

(a) There exists a sequence of estimators $\hat{\boldsymbol{\theta}}_n = \hat{\mu}_n, \hat{\psi}_n, \hat{\xi}_n$ such that

$$\begin{aligned}
\lim_{n\to\infty} \Pr\left\{\hat{\boldsymbol{\theta}}_n \text{ is a MLE}\right\} &= 1, \\
\sqrt{k_n}\left(\frac{\hat{\mu}_n - b_{m_n}}{a_{m_n}}, \frac{\hat{\psi}_n}{a_{m_n}} - 1, \hat{\xi}_n - \xi_0\right) &\xrightarrow{d} \mathcal{N}\left(\lambda I_{\xi_0}^{-1}\mathbf{b}, I_{\xi_0}^{-1}\right).
\end{aligned}$$

end 2/18/25

(b) If $\hat{\boldsymbol{\theta}}_n^i = (\hat{\mu}_n^i, \hat{\psi}_n^i, \hat{\xi}_n^i)$, $i = 1, 2$ are two sequences of estimators satisfying

$$\begin{aligned}
\lim_{n\to\infty} \Pr\left\{\hat{\boldsymbol{\theta}}_n^i \text{ is a MLE}\right\} &= 1, \\
\lim_{n\to\infty} \Pr\left\{\sqrt{k_n}\left(\frac{\hat{\mu}_n^i - b_{m_n}}{a_{m_n}}, \frac{\hat{\psi}_n^i}{a_{m_n}} - 1, \hat{\xi}_n^i - \xi_0\right) \in H_n\right\} &= 1,
\end{aligned}$$

where H_n is a ball in $\mathbb{R}^3$ of center 0 and radius r_n, where $r_n = O(k_n^\delta)$, $0 < \delta < \min\left(\frac{1}{2}, \xi_0 + \frac{1}{2}\right)$ as $n \to \infty$, then

$$\lim_{n\to\infty} \Pr\left\{\hat{\boldsymbol{\theta}}_n^1 = \hat{\boldsymbol{\theta}}_n^2\right\} = 1.$$

2.5.1 *Side Section 1: A heuristic on biased estimation*

Suppose we have a sequence of experiments indexed by n, where in the nth experiment there are k_n observations $X_1, \ldots, X_{k_n}$ whose true joint density is g_n, but for reasons of convenience or because we don't know how to exactly calculate g_n, we replace g_n by a known joint density f_n indexed by a parameter vector $\boldsymbol{\theta}_n$. The examples of interest to us include the X_i's being either block maxima or exceedances over a threshold and their density f_n being approximated by a GEV or GPD density. We will always want $f_n - g_n \to 0$ under some suitable metric (e.g. total variation norm or Hellinger distance) but we won't worry about precise modes of convergence for the moment — that can come later.

Suppose we estimate $\boldsymbol{\theta}$ by defining a set of equations

$$\sum_{i=1}^{k_n} \mathbf{T}(X_i; \boldsymbol{\theta}) = 0$$