Brazilian Journal of Probability and Statistics
2011, Vol. 25, No. 3, 444–470
DOI: 10.1214/11-BJPS156

Contiguity and irreconcilable nonstandard asymptotics of statistical tests

Pranab K. Sen[a] and Antonio C. Pedroso-de-Lima[b]

[a]*University of North Carolina at Chapel Hill, USA*
[b]*University of São Paulo, Brazil*

Abstract. Wald-type test statistics based on asymptotically normally distributed estimators (not necessarily maximum likelihood estimation or best asymptotically normal) provides an easy access to have tests for statistical hypotheses, far beyond the parametric paradigms. The methodological perspectives rest on a basic consistent asymptotic normal (CAN) condition which is interrelated to the well-known local asymptotic normality (LAN) condition. Contiguity of probability measures facilitates the C(L)AN condition in a relatively easier way. For many regular families of distributions, when statistical hypotheses do not involve nonstandard constraints, verification of contiguity of probability measures is facilitated by the well-known LeCam's First Lemma [see Hájek, Šidák and Sen *Theory of Rank Tests* (1999), Chapter 7]. For nonregular families, though contiguity may hold under different setups, CAN estimators are not fully exploitable in the Wald type testing theory. This simple feature is illustrated by a two-parameter exponential model. Guided by this simple example, mixture of distributions are appraised in the context of Wald-type tests and the so-called $\overline{\chi}^2$- and $\overline{\mathrm{E}}$-test theory is thoroughly appraised. A general result on counter examples is presented in detail.

1 Introduction

The theory of *likelihood ratio tests* (LRT) has been extensively studied in the literature. Most of these developments have taken place in the so-called *regular case* where the *Cramér–Rao regularity conditions* hold. This development is most effective for the so-called *exponential family of densities* for which *sufficiency* plays a pioneering role. In such cases, for statistical inference, the *likelihood function* can be reduced to the distribution of *sufficient statistics* alone, and hence, exact properties of tests and estimates can be studied under relatively simpler setups. For densities not necessarily belonging to the exponential family where sufficiency may not hold, LRT works out in some cases in a manageable way. However, even for the exponential family, excepting for some simple cases, an exact treatise of LRT may not be feasible. To overcome this primary drawback, statistical inference

Key words and phrases. CAN statistics, exponential family of densities, Hellinger-distance, LA(M)N, likelihood ratio, maximum likelihood estimator, nonparametrics, Rao's score statistics, regular family of estimators, resampling plans.

Received May 2011; accepted May 2011.

has been extensively developed in an asymptotic setup where the sample size is allowed to increase indefinitely with a view to use appropriate large sample tools in examining good approximations. It is also hoped that such asymptotic methods could provide some justifications for moderate sample sizes as well. The evolution of *resampling methods* has added much more strength to this intermediate sample size situation. In this setup, nonregular cases are to be dealt with on a more or less case by case basis. With respect to the asymptotic theory of statistical inference, the methodology has incorporated several important concepts like the LAN (*local asymptotic normality*) and its variants, *regular family of estimators*, *contiguity of probability measures*, *Hellinger distance*, and many useful tools from probability theory to broaden the reach of asymptotic methods in a more general setup. In the context of hypothesis testing, in asymptotic setups, the equivalence of the LRT, Wald test and Rao's score test have been extensively studied in the literature. In a general multiparameter setup, Roy's *union-intersection principle* has also been advocated. In the present study, we like to examine the role of contiguity of probability measures in this setup without confining us to the regular case only or to simple hypotheses testing. The preliminary notion is outlined in Section 2. Section 3 deals with some nonregular cases where contiguity does not hold. Section 4 is devoted to some nonparametric inference problems and in Section 5 we briefly appraise a problem in constrained statistical inference. Section 6 deals with some general observations and concluding remarks.

2 Preliminary notions and standard asymptotics

In statistical inference, covering both estimation theory and hypothesis testing, the estimators are obtained through suitable *estimating equations* and tests of significance involve suitable test statistics. Whenever such statistics are linear the usual central limit theorems can be incorporated in providing a well manageable asymptotic theory which may also provide good indications of moderate sample size performances. However, often the estimating equations are either implicit functions or are highly nonlinear. Similarly, the likelihood ratio tests may involve highly complex statistics for which a direct adaptation of suitable central limit theorems may not be feasible. This feature invariably calls for an in-depth study of the behavior of such statistics at the true parameter point as well as local alternatives to be defined suitably. In this context, the concept of contiguity of probability measures plays a basic role. A precursor to this development [Hájek (1962)] is the *differentiability in quadratic mean* (DQM) [LeCam (1960)] and some related developments: *locally asymptotically normal* (LAN) family, *locally asymptotically quadratic* (LAQ) family, and the so-called *locally asymptotically mixed normal* (LAMN) family.

It may be remarked that DQM $\Rightarrow$ LAQ while the LAQ is more general than LAN and LAMN. Contiguity and LAQ setups go very nicely with each other and together they cater to the asymptotic theory of a much broader class of estimators and test statistics. There is also the related concept of *regular estimators* as formulated by Hájek (1970), LeCam (1979) and Inagaki (1970, 1973). Basically, the LAQ condition ensures that for a suitable sequence of *contiguous alternatives* the log-likelihood function can be well approximated by a quadratic function, and the LAMN condition specifies further structures on the discriminant of that quadratic function. A prime utility of contiguity is the simplification of the asymptotic distribution theory of statistics under such alternatives through a characterization of their distribution under the null hypothesis. Asymptotic normality of estimators plays a fundamental role in this context.

Following the lead in LeCam (1960) in laying down the concept of contiguity, with a systematic account in Hájek, Šidák and Sen (1999), Chapter 7, it is customary to consider a sequence of testing problem wherein the sample size n is regarded as a member of a sequence $\{n_\nu\}$ and a sequence of alternatives $\{H_\nu\}$ is formulated where H_ν invokes a close alternative defined by a suitable metric d_ν separating the null and the local alternatives. Let P_ν be the probability measure under the null and Q_ν under the alternative H_ν. Then, if for any sequence of events $\{A_\nu\}$, $A_\nu \in \mathcal{A}_\nu$, $[P_\nu(A_\nu) \to 0] \Rightarrow [Q_\nu(A_\nu) \to 0]$ as $\nu \to \infty$, then $\mathrm{d}Q_\nu$ is said to be contiguous to $\mathrm{d}P_\nu$. Contiguity implies that for any sequence of random variables converging to 0 in P_ν-probability converges to 0 in Q_ν-probability as well. However, Q_ν contiguous to P_ν does not necessarily imply that P_ν is contiguous to Q_ν. Further, contiguity does not imply the absolute continuity of Q_ν with respect to P_ν. Also, P_ν and Q_ν are L_1-norm equivalent if $\|P_\nu - Q_\nu\| = \sup\{|P_\nu(A) - Q_\nu(A)| : A \in \mathcal{A}_\nu\} \to 0$, then contiguity holds in both ways. If T_{n_ν} is $\mathcal{A}_\nu$-measurable, then $T_{n_\nu} \to 0 \in P_\nu$-probability implies that it does so under Q_ν-probability as well.

We define the likelihood ratio statistic L_ν as $\mathrm{d}Q_\nu/\mathrm{d}P_\nu$ wherein we let it have the value 1 when both $\mathrm{d}P_\nu$ and $\mathrm{d}Q_\nu$ are equal to 0. In this vein, LeCam (1960) characterized contiguity of $\{Q_\nu\}$ with respect to $\{P_\nu\}$ by some basic results, known as LeCam's lemmas that we now state as in Hájek, Šidák and Sen (1999).

LeCam's First Lemma. *Assume that $F_\nu(x) = P_\nu(L_\nu \le x)$ converges weakly (at continuity points) to a distribution function $F(\cdot)$ such that*

$$\int_0^\infty x \,\mathrm{d}F(x) = 1. \tag{2.1}$$

Then, $\{Q_\nu\}$ is contiguous to $\{P_\nu\}$.

For a log-normal random variable with parameters μ and σ^2, equation (2.1) will be true when $\mu = -\sigma^2/2$. We have then the following:

Corollary 1. *If under P_ν the log-likelihood ration L_ν is asymptotically $\mathcal{N}(-(1/2)\sigma^2, \sigma^2)$, then $\{Q_\nu\}$ is contiguous to $\{P_\nu\}$.*

Although the first lemma is a more general result, finding the asymptotic distribution of the likelihood ratio statistic L_ν can be difficult. In real applications it is the Corollary 1 that is widely used, specially when the alternative hypothesis is defined in a parametric fashion.

LeCam's Second Lemma. *Let* $\mathbf{x}'_\nu = (x_1, \ldots, x_{n_\nu})$,

$$\mathrm{d}P_\nu(\mathbf{x}_\nu) = \prod_{i=1}^{n_\nu} f_{\nu i}(x_i) \quad \textit{and} \quad \mathrm{d}Q_\nu(\mathbf{x}_\nu) = \prod_{i=1}^{n_\nu} g_{\nu i}(x_i),$$

so that

$$\log L_\nu = \sum_{i=1}^{n_\nu} \log[g_{\nu_i}(x_i)/f_{\nu_i}(x_i)].$$

Considering the statistic

$$W_\nu = 2\sum_{i=1}^{n_\nu}\{[g_{\nu_i}(X_i)/f_{\nu_i}(X_i)]^{1/2} - 1\},$$

if W_ν *converges to a* $\mathcal{N}(-(1/4)\sigma^2, \sigma^2)$ *under* P_ν *and*, $\forall \varepsilon > 0$

$$\lim_{\nu\to\infty} \max_{1\le i\le n_\nu} P_\nu\left\{\left|\frac{g_{\nu_i}(X_i)}{f_{\nu_i}(X_i)} - 1\right| > \varepsilon\right\} = 0,$$

then the statistic $\log L_\nu$ *satisfies*

$$\lim_{\nu\to\infty} P_\nu\left(\left|\log L_\nu - W_\nu + \frac{1}{4}\sigma^2\right| > \varepsilon\right) = 0$$

and is asymptotically $\mathcal{N}(-(1/2)\sigma^2, \sigma^2)$ *under* P_ν.

LeCam's Third Lemma. *If under* P_ν *we have*

$$\begin{pmatrix} T_{n_\nu} \\ \log L_\nu \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{2.2}$$

then under Q_ν,

$$T_{n_\nu} \xrightarrow{\mathcal{D}} \mathcal{N}(\mu_1 + \sigma_{12}, \sigma_{11}),$$

where $\boldsymbol{\mu}' = (\mu_1, \mu_2)$, $\boldsymbol{\Sigma} = ((\sigma_{ij}))_{i,j=1,2}$ *and* $\mu_2 = -\sigma_{22}/2$.

The third lemma avoids the need for messy derivation of the asymptotic moments of T_{n_ν} under Q_ν and simply adjusts the null distribution by a shift of the asymptotic mean. We refer to Hájek, Šidák and Sen (1999), Section 7.1, for detailed discussion of contiguity and derivation of all these powerful results. It is to be noted further that (2.2) holds in a more general context where T_{n_ν} and μ_i are

p-vectors, σ_{12} is a p-vector and σ_{11} is a $p \times p$ positive definite matrix, not necessarily of full rank. Then (2.2) with $\mathcal{N}_{p+1}$ (instead of $\mathcal{N}_2$) provides the desired result under Q_ν.

In this study, we mainly confine ourselves to the implication of contiguity in some nonstandard statistical inference problems and some nonregular cases are reviewed in the light of these developments. In the usual case of maximum likelihood estimators (MLE), excepting for the exponential family of densities, the estimating equations are complex and solutions are implicit. Therefore, there is a need to have a representation of the MLE in terms of the *Rao-score statistics*, and exploit the asymptotic normality of the latter to have parallel results for the MLE. LeCam's first lemma has an elegant use in this context. Direct approaches in deriving the asymptotic normality of MLE are often very cumbersome. In the same way, contiguity provides a justification of the classical *Wald statistics* in a variety of statistical inference problems. We shall see later on that contiguity also plays a basic role with Wald-type of statistics in a far broader domain encompassing parametric as well as *nonparametric* (and *semiparametric*) inference. Another glaring use of contiguity in statistical inference is in the area of nonparametrics comprising tests as well as estimates; typically these statistics are highly nonlinear posing cumbersome methods for studying their asymptotic properties [Chernoff and Savage (1958)] and the use of contiguity offers a tremendous simplification to the asymptotic theory [Hájek, Šidák and Sen (1999), Chapter 7]. Even under the null hypothesis, often, it is difficult to use directly some central limit theorems. Sometimes some projection into linear statistics is prescribed [Hoeffding (1948), Chernoff and Savage (1958), Hájek (1968) and others], or matching a hypothesis of invariance, some *permutational central limit theorems* are used. General theory of asymptotics for such statistics studied by Chernoff and Savage (1958), Hájek (1968) and others involve elaborate analysis, and the use of contiguity renders tremendous simplification [Sen (1981)]. In this context, there is an elegant *uniform asymptotic linearity in the parameter of interest* result that invokes contiguity of probability measures.

We conclude this section with a general result on Wald type tests beyond the parametric paradigm where contiguity (LeCam's Second Lemma) is of considerable help.

Wald-type of tests has the maximum flexibility to be adaptable in a broader situation beyond the parametric paradigms. As such, we consider here a general multiparameter hypothesis testing problem. For a t-vector $\boldsymbol{\theta}$, we frame a null hypothesis

$$H_0 : \mathbf{g}(\boldsymbol{\theta}) = \mathbf{0}, \quad \text{vs.} \quad H_1 : \mathbf{g}(\boldsymbol{\theta}) \neq \mathbf{0}.$$

Let $\mathbf{T}_n$ be an estimator of $\boldsymbol{\theta}$ such that $\sqrt{n}(\mathbf{T}_n - \boldsymbol{\theta}) \sim \mathcal{N}_t(\mathbf{0}, \boldsymbol{\Gamma})$, and let $\hat{\boldsymbol{\Gamma}}$ be a suitable consistent estimator of $\boldsymbol{\Gamma}$. Also, define $\dot{\boldsymbol{G}} = (\partial/\partial\boldsymbol{\theta})\mathbf{g}(\boldsymbol{\theta})$ at $\hat{\boldsymbol{\theta}}$ (a $k \times t$ matrix) and let

$$\boldsymbol{Q} = \hat{\dot{\boldsymbol{G}}}\hat{\boldsymbol{\Gamma}}\hat{\dot{\boldsymbol{G}}}'.$$

Then the test statistic is $\mathcal{L}_n = n[\mathbf{g}(\mathbf{T}_n)]' \boldsymbol{Q}^{-1}[\mathbf{g}(\mathbf{T}_n)]$. Under H_0, $\mathcal{L}_n$ has asymptotically chi-square distribution with k degrees of freedom. This result is based on the asymptotic multinormality of the estimator and consistency of $\boldsymbol{Q}$. In order that under local alternatives $\mathcal{L}_n$ has asymptotically noncentral chi-square distribution, we need to verify that $\boldsymbol{Q}^0$, the counterpart of $\boldsymbol{Q}$ evaluated at $\boldsymbol{\theta}$ is continuous in a neighborhood of the true value $\boldsymbol{\theta}_0$, and $\boldsymbol{Q}$ is consistent for $\boldsymbol{Q}^0$ uniformly in a neighborhood of $\boldsymbol{\theta}_0$, and further, the asymptotic (multi)normality of the estimator holds uniformly in a neighborhood of the true parameter $\boldsymbol{\theta}$. If H_1 is contiguous to H_0 (as can be verified directly without $\mathbf{T}_n$ in the picture), then LeCam's Second Lemma in the general vector case as has been presented earlier can be employed to claim this noncentral distribution directly. This scenario is particularly noteworthy when $\mathbf{T}_n$ is nonlinear or an implicit root of some complex estimating equation. This scenario crops up typically in nonparametric and robust estimation problems.

This standard scenario is based on (i) the contiguity of the alternative hypothesis with respect to the null case, and (ii) asymptotic joint normality of the log-likelihood function and the statistic $\mathbf{T}_n$. A natural question that arises in this context is what happens if either (i) or (ii) fails, and this is elaborated in the proceeding sections.

3 Nonregular family of distributions

As has been indicated earlier, contiguity hinges on some intrinsic properties of the underlying density (a bit simpler than the classical Cramér–Rao regularity conditions for the MLE); in this way, it goes beyond the treatise of the so-called exponential family of densities where sufficiency prevails. Our main emphasis here is on some nonregular cases where the Cramér–Rao regularity conditions may not hold and where asymptotically normal distributions may not pertain. In that way, the role of contiguity of probability measures in such nonregular setups is to be critically appraised. In this context, either contiguity may not hold under the usual rate of convergence or nonstandard asymptotic distributions crop up. For better motivation, we begin with a two-parameter exponential distribution, followed by a two-parameter uniform distribution.

3.1 Two-parameter exponential distribution

Let X be a random variable following a two-parameter exponential distribution, with p.d.f.

$$f(x;\theta_1,\theta_2) = \frac{1}{\theta_2} e^{-(x-\theta_1)/\theta_2} I(x \geq \theta_1), \qquad \theta_2 > 0. \tag{3.1}$$

Here, θ_1 and θ_2 are location and scale parameters, respectively. Assuming a random sample $X_1, \dots, X_n$ of X, and denoting the associated order statistics by

$X_{n:1}, \ldots, X_{n:n}$, we may rewrite the likelihood function as

$$L(\boldsymbol{\theta}) = \theta_2^{-n} \exp\left\{-\frac{1}{\theta_2}\left[\sum_{j=2}^{n}(X_{n:j} - X_{n:1}) + n(X_{n:1} - \theta_1)\right]\right\} I(X_{n:1} \geq \theta_1),$$

so that we immediately have that the statistics $[\sum_{j=2}^{n}(X_{n:j} - X_{n:1}), X_{n:1}]$ are jointly sufficient for (θ_1, θ_2), and that the maximum likelihood estimator for θ_1 is $\widehat{\theta}_1 = X_{n:1}$. The *profile* log-likelihood

$$\ell(\theta_2) = -n \log(\theta_2) - \frac{1}{\theta_2}\sum_{i=1}^{n}(X_{n:i} - X_{n:1})$$

provides the maximum likelihood estimator for θ_2 given by

$$\widehat{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n}(X_{n:i} - X_{n:1}).$$

We note that this estimator is biased; an unbiased estimator is given by

$$\widehat{\theta}_2^* = \frac{1}{n-1}\sum_{i=2}^{n}(X_{n:i} - X_{n:1}).$$

Considering the theory discussed in Sen, Singer and Pedroso-de-Lima (2010), pages 227–231, or Kendall and Stuart (1977) on extreme order statistics, we note that

$$\frac{n(X_{n:1} - \theta_1)}{\theta_2} \xrightarrow{\mathcal{D}} \exp(1),$$

or

$$n(X_{n:1} - \theta_1) \xrightarrow{\mathcal{D}} \exp(\theta_2).$$

As for the maximum likelihood estimator of θ_2, we have that $\widehat{\theta}_2$ can be written as an average of n independent and identically distributed exponential random variables so that, by the Central Limit Theorem we have

$$\sqrt{n}(\widehat{\theta}_2 - \theta_2) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \theta_2^2).$$

We note that, since the marginal distribution of $\widehat{\theta}_1$ does not converge to a normal distribution, a Wald-type statistics would not be appropriate for testing both parameters θ_1 and θ_2. We now study how the contiguity property reflects this issue, formulating the following hypotheses:

$$H_o : \boldsymbol{\theta} = \boldsymbol{\theta}_0' = (0, 1) \quad \text{and} \quad H_n : \boldsymbol{\theta} = \boldsymbol{\theta}_{1n}' = \left(a/n, 1 + b/\sqrt{n}\right),$$

where H_o could be made more general by allowing the parameters to be any constant (albeit for θ_2 the constant should be positive) and the rate of convergence (in n) in the local alternatives is chosen motivated by the rate of convergence observed for the marginal distributions of the corresponding maximum likelihood estimators of θ_1 and θ_2.

By (3.1), the log-likelihood ratio is given by

$$\begin{aligned}
\log L_n &= \sum_{i=1}^{n} \log \frac{f(X_i, \boldsymbol{\theta}_{1n})}{f(X_i, \boldsymbol{\theta}_0)} \\
&= \log\Big[1 - \Big(\frac{b}{\sqrt{n}} + \frac{b^2}{n}\Big) + O(n^{-3/2})\Big] \sum_{i=1}^{n} I\Big(X_i \geq \frac{a}{n}\Big) \\
&\quad + \frac{a}{n} \sum_{i=1}^{n} I\Big(X_i \geq \frac{a}{n}\Big) \qquad (3.2)\\
&\quad + \Big(\frac{b}{\sqrt{n}} - \frac{b^2}{n}\Big) \sum_{i=1}^{n} X_i I\Big(X_i \geq \frac{a}{n}\Big) \\
&\quad + O_p(n^{-3/2}).
\end{aligned}$$

If we further consider the expansion of $\log(1+x) = x - x^2/2 + O(x^3)$, where in expression (3.2) we take $x = b/\sqrt{n} + b^2/n$, then we have

$$\begin{aligned}
\log L_n &= n\Big(-\frac{b}{\sqrt{n}} + \frac{b^2}{n} - \frac{b^2}{2n} + O(n^{-3/2})\Big) \frac{1}{n} \sum_{i=1}^{n} I\Big(X_i \geq \frac{a}{n}\Big) \\
&\quad + a \frac{1}{n} \sum_{i=1}^{n} I\Big(X_i \geq \frac{a}{n}\Big) \\
&\quad + n\Big(\frac{b}{\sqrt{n}} - \frac{b^2}{n}\Big) \frac{1}{n} \sum_{i=1}^{n} X_i I\Big(X_i \geq \frac{a}{n}\Big) \\
&\quad + O_p(n^{-3/2}).
\end{aligned}$$

Considering the *empirical* (or *sample*) distribution function as $F_n(x) = n^{-1} \times \sum_{i=1}^{n} I(X_i \leq x)$, $x \in \mathbb{R}$, we have

$$\begin{aligned}
\log L_n &= \Big(-\sqrt{n} b + \frac{b^2}{2}\Big)\Big[1 - F_n\Big(\frac{a^-}{n}\Big)\Big] \\
&\quad + a\Big[1 - F_n\Big(\frac{a^-}{n}\Big)\Big] \\
&\quad + (\sqrt{n} b - b^2) \frac{1}{n} \sum_{i=1}^{n} X_i I\Big(X_i \geq \frac{a}{n}\Big) \\
&\quad + O_p(n^{-3/2}).
\end{aligned}$$

Rearranging the terms and noting that $\sum_{i=1}^n X_i I(X_i \geq a/n) = \sum_{i=1}^n X_i - \sum_{i=1}^n X_i I(X_i < a/n)$, we have

$$\begin{aligned}\log L_n &= b\sqrt{n}(\overline{X}_n - 1) - \sqrt{n} b F_n\left(\frac{a^-}{n}\right) + \frac{b^2}{2}\left[1 - F_n\left(\frac{a^-}{n}\right)\right] \\ &\quad + a\left[1 - F_n\left(\frac{a^-}{n}\right)\right] - b^2\overline{X}_n - (\sqrt{n}b - b^2)\frac{1}{n}\sum_{i=1}^n X_i I\left(X_i < \frac{a}{n}\right) \\ &\quad + O_p(n^{-3/2}).\end{aligned} \tag{3.3}$$

Now we note that, as discussed earlier, for the first term on the r.h.s. of (3.3),

$$b\sqrt{n}(\overline{X}_n - 1) \xrightarrow{\mathcal{D}} \mathcal{N}(0, b^2).$$

As for the second term on the r.h.s. of (3.3), we note that, under H_0,

$$\begin{aligned}\mathbb{E}\left[\sqrt{n} F_n\left(\frac{a}{n}\right)\right] &= \sqrt{n}\mathbb{E}\left[F_n\left(\frac{a^-}{n}\right)\right] = \sqrt{n} F\left(\frac{a}{n}\right) \\ &= \sqrt{n}(1 - e^{-a/n}) = \sqrt{n}\left[\frac{a}{n} + O(n^{-2})\right] = \frac{a}{\sqrt{n}} + O(n^{-3/2}) \\ &= O(n^{-1/2}),\end{aligned}$$

implying (by the Chebyshev inequality) that $\sqrt{n} b F_n(a/n) = O_p(n^{-1/2})$. Also, noting that $\sqrt{n}(\overline{X}_n - 1) = O_p(1)$, it follows that the 3rd, 4th and 5th terms on the r.h.s. of (3.3) converges in probability to $(1/2)b^2$, a and $-b^2$ respectively. The 6th term is $O_p(n^{-1/2})$. Therefore, we have that, as $n \to +\infty$,

$$\log L_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, b^2) + \frac{b^2}{2} + a - b^2 \equiv \mathcal{N}\left(-\frac{b^2}{2} + a, b^2\right).$$

We therefore conclude that unless $a = 0$, contiguity does not hold for this simple nonregular case. However, if we treat θ_1 as a nuisance parameter, contiguity will be intact for θ_2 variation in the usual rate $n^{-1/2}$. If we want to test for θ_1 alone treating θ_2 as a nuisance parameter, the above result implies that the contiguity of Q_ν to P_ν does not hold even for the adjusted rate n instead of $n^{-1/2}$.

We also note that, for testing hypotheses involving both parameters, one could think of using the sum of the two test statistics described earlier, that is, taking a n rate of convergence for $\widehat{\theta}_1$ with corresponding exponential asymptotic distribution and $\sqrt{n}$ rate of convergence for $\widehat{\theta}_2$. In this case, $X_{n:1} \geq \theta_1$ with probability 1, and hence, under the null hypothesis, $nX_{n:1}$ is nonnegative while under H_{1n}, if a is negative, $X_{n:1}$ could be negative with a positive probability. This will distort the nonnull distribution and affect the efficiency properties (inspite of being based on sufficient statistics).

3.2 Two-parameter uniform distribution

Let $X_1, \ldots, X_n$ be a collection of independent random variables, with the same two-parameter Uniform distribution as X, with p.d.f.

$$f(x;\theta_1,\theta_2) = \frac{1}{\theta_2} I(\theta_1 - \theta_2/2 \le x \le \theta_1 + \theta_2/2), \qquad \theta_2 > 0.$$

The likelihood may then be written as

$$L(\boldsymbol{\theta}) = \theta_2^{-n} I(\theta_1 - \theta_2/2 \le X_{n:1} \le X_{n:n} \le \theta_1 + \theta_2/2),$$

for which $(X_{n:1}, X_{n:n})$ are known to be sufficient for $\boldsymbol{\theta}$, yielding the MLE as

$$\widehat{\boldsymbol{\theta}} = \begin{pmatrix} 1/2 & 1/2 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} X_{n:1} \\ X_{n:n} \end{pmatrix}.$$

In this nonregular case, it is known that $X_{n:1}$ and $X_{n:n}$ are asymptotically independent, and $U_n = n[(\theta_1 + \theta_2/2) - X_{n:n}]$ and $V_n = n[X_{n:1} - (\theta_1 - \theta_2/2)]$ have asymptotically exponential distributions with parameter θ_2. Therefore, it follows that

$$n(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} \begin{pmatrix} -1/2 & 1/2 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} U \\ V \end{pmatrix},$$

where U and V are independent exponential random variables with the common parameter θ_2; none of which is normally distributed. Note that whereas for $U + V$ the chi-square distribution holds, for $U - V$ that is untenable too. Further, we show that in this case too, contiguity does not hold even for the rate n instead of $n^{1/2}$, as formulated in a simplified form below.

Consider the hypothesis

$$H_o : \theta_1 = 0;\ \theta_2 = 1 \quad \text{and} \quad H_{1n} : \theta_1 = a/n;\ \theta_2 = 1 + b/n.$$

The log-likelihood ratio is

$$\begin{aligned} \log L_n &= \sum_{i=1}^{n} \log\left(\frac{f(X_i \mid \theta_1)}{f(X_i \mid \theta_0)}\right) \\ &= \log\left(\frac{1}{1 + b/n}\right) \\ &\quad \times \sum_{i=1}^{n} I\left(\max\left\{-\frac{1}{2} + \frac{2a - b}{2n};\ \frac{1}{2}\right\} \le X_i \le \min\left\{\frac{1}{2} + \frac{2a + b}{2n};\ \frac{1}{2}\right\}\right). \end{aligned}$$

Note that while $a \in \mathbb{R}$, b/n needs to be greater than -1, thus yielding the four cases: $a \ge$ or ≤ 0, $b \ge 0$ or $b \in (-n, 0)$. In this context, it may be noted that for $X \sim \text{Unif}[-1/2, 1/2]$, by the Glivenko–Cantelli theorem (Sen, Singer and Pedroso-de-Lima (2010), p. 157), $F_n(-1/2) \overset{\text{a.s.}}{\to} 0$ and $F_n(1/2) \overset{\text{a.s.}}{\to} 1$ as $n \to \infty$. Thus, depending on the combination of a and b, $n(1 - F_n(1/2 + (2a - b)/2n))$ and

$nF_n((-1/2)+(2a-b)/2n)$ will a.s. converge to 0 and $(a-b/2)$, respectively, if $a-b/2$ is greater than 0, and to $(a-b/2)$ and 0 if $a-b/2$ is less than 0. As such, some routine computation leads to

$$\log L_n \xrightarrow{\mathcal{P}} -b$$

as $n \to \infty$.

Hence for the two parameter uniform distribution, contiguity does not hold, as was to be expected. Sans contiguity, we may not be able to use standard results on nonnull distribution.

4 Statistical functionals and contiguity

Statistical functionals, traditionally arising in nonparametric inference, are typically nonlinear functions, where traditional central limit theorems may not directly apply. There are two principal approaches to the study of asymptotic properties of such functionals: (i) To use suitable projection of such a statistic onto a linear functional, on which standard asymptotics hold, and show that the residual component is negligible in a well-defined statistical sense, and (ii) formulate some asymptotically equivalent linear functionals and use contiguity to strengthen this equivalence to local alternatives too. In the first approach contiguity is not that essential, but the regularity assumptions may often be a bit more stringent. For example, for U-statistics, the kernel needs to be of a finite degree so that the parameter is estimable in a nonparametric way—a condition which does not hold even for the simple case of population medians or quantiles in general. In the second approach, both contiguity and projection are used in a conjugate way to derive parallel asymptotics, often under somewhat less stringent regularity assumptions.

In this section we consider some cases to illustrate how contiguity may (or not) play an important role in the asymptotics of statistical functionals.

4.1 Hoeffding's U-statistics

Following Hoeffding (1948), we define a general U-statistic

$$U_n = \binom{n}{m}^{-1} \sum_{1\le i_1<\cdots<i_m\le n} \phi(X_{i_1},\ldots,X_{i_m}),$$

where the kernel ϕ of degree $m \ge 1$ unbiasedly estimates a parameter $\theta(F)$, a *statistical functional* or *estimable parameter*. If $\phi(\cdot)$ is square integrable and we let $\phi_1(x) = \mathbb{E}[\phi(X_1,\ldots,X_m)|X_1 = x]$, then defining

$$U_n^{(1)} = n^{-1}\sum_{i=1}^{n} m[\phi_1(X_i) - \theta(F)],$$

we note that the projection of $[U_n - \theta(F)]$ to a linear statistic is $U_n^{(1)}$ and the remainder term has variance $O(n^{-2})$ [Hoeffding (1948)] so that whenever ζ_1, the variance of $\phi_1(X_1)$, is positive, $n^{1/2}[U_n - \theta(F)]$ is asymptotically normal with zero mean and variance $m^2\zeta_1$. In this derivation, contiguity is neither needed nor has any additional advantage in dealing with asymptotic results. On the other hand, ζ_1 actually is a functional of the distribution function F and under suitable null hypothesis, it may be explicitly known. If we consider local alternatives of the Pitman-type, whenever $\zeta_1(F)$ is continuous in F in a neighborhood of F_o, the asymptotic normality holds for local alternatives. Invoking the asymptotic normality of the estimator of ζ_1 and contiguity it could be easier to verify this uniformity of the convergence of the estimator and obtain a result parallel to that of LeCam's Third Lemma. Otherwise, we need to verify the uniform convergence of the variance estimator without invoking contiguity. If in the above context, $\zeta_1 = 0$ but higher order variances are nonzero, the asymptotic normality of U_n will not hold but contiguity may still give the uniform consistency of the variance estimator. To illustrate a simple case of $\zeta_1 = 0$, vitiating the asymptotic normality of U_n, we consider the following example:

Suppose that $X_1, \dots, X_n$ are independent and identically distributed random variables with distribution function F such that $\mathbb{E}(X_i) = \mu(F)$ and $\mathbb{V}\mathrm{ar}(X_i) = \sigma^2(F)$. To estimate of $\theta(F) = \mu^2(F)$, take the kernel $\phi(X_i, X_j) = X_i X_j$, $i \neq j$. We have that $\mathbb{E}_F[\phi(X_i, X_j)] = \mathbb{E}_F(X_i)\mathbb{E}_F(X_j) = \mu^2(F)$ so that the corresponding U-statistic is unbiased for $\theta(F)$. The first-order kernel is then

$$\phi_1(X_i) = \mathbb{E}_F[\phi(X_i, X_j)|X_i] = X_i \mathbb{E}_F(X_j) = X_i \mu(F),$$

so that when $\mu(F) = 0$, $\phi_1(x) = 0$ a.e., and hence, $\zeta_1(F) = 0$. Thus, the parameter $\mu^2(F)$ at $\mu(F) = 0$ yields a U_n which is stationary of order 1. Also,

$$\phi_2(X_i, X_j) = \mathbb{E}_F[\phi(X_i, X_j)|X_i, X_j] = X_i X_j,$$

and, hence, when $\mu(F) = 0$, the variance of the second-order kernel is

$$\zeta_2 = \mathbb{E}_F[(X_i X_j)^2] = \mathbb{E}_F(X_i^2)\mathbb{E}_F(X_j^2) = \sigma^4(F).$$

Noting that

$$U_n = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} X_i X_j = \frac{1}{n-1}\left(n\overline{X}_n^2 - \frac{1}{n}\sum_{i=1}^n X_i^2 \right),$$

if $\mu(F) = 0$, if follows that

$$n\frac{\overline{X}_n^2}{\sigma^2(F)} \xrightarrow{\mathcal{D}} \chi_1^2 \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{\text{a.e.}} \sigma^2(F),$$

so that

$$\frac{(n-1)}{\sigma^2(F)} U_n \xrightarrow{\mathcal{D}} \chi_1^2 - 1.$$

In the same example, let F be a continuous distribution function with a probability density function $f(x)$ a.e., and consider the alternative $H_n: F$ has the mean $\mu_n = n^{-1/2}\delta$, for some fixed δ. Then the log-likelihood ratio statistic for this sequence of alternative is asymptotically normal with mean equal to $(-1/2)$ times the variance, so that by LeCam's Second Lemma, contiguity holds. On the other hand, using the above asymptotic result (that $nU_n/\sigma^2 + 1 \xrightarrow{\mathcal{D}} \chi_1^2$ under $H_0: \delta = 0$), we could extend the asymptotic distributional result to noncentral χ_1^2 distribution using a direct extension of LeCam's Third Lemma. We present this extension of LeCam's Third Lemma as follows:

Lemma 4.1. *Suppose that a statistic S_n is expressible as $g(\mathbf{T}_n)$ where $\mathbf{T}_n$ satisfies LeCam's Third Lemma regularity conditions. Then the asymptotic distribution function of normalized S_n under alternative can be obtained by appealing to the joint distribution of normalized $\mathbf{T}_n$ under alternative.*

In the above example, the simple χ_1^2 distribution under $H_0: \mu = 0$ goes over to noncentral chi-square distribution with 1 degree of freedom and suitable noncentrality parameter. However, in a general case of U-statistics stationary of order 1 or more, the limiting distribution of normalized U-statistics becomes more complex von-Mises type distributions (involving multiple Wiener integrals) and such a simple contiguity based approach may not render significant simplifications for the asymptotics. This difficulty in a more general setup of constrained statistical inference will be appraised in a later subsection. In passing, we may remark that for U-statistics or in general, for statistical functionals, alternatives are defined in terms of such functionals and may not be of a specific parametric form. This creates an impasse to LeCam's lemmas insuring contiguity, although alternative ways of establishing contiguity may work out in some situation [viz., Pinheiro, Sen and Pinheiro (2011) dealing with Hamming distance in high-dimensional categorical data models].

4.2 Rank tests for Lehmann alternatives

In survival analysis a first step frequently involves testing homogeneity of groups. The problem can be posed as follows. Let $X_1, \ldots, X_n$ be failure times, that is, independent random variables such that X_i has a survival function denoted by $\overline{F}_i = 1 - F_i$, $i = 1, \ldots, n$. Consider the hypotheses

$$H_0: \overline{F}_i(x) = \overline{F}(x) \qquad \forall x \in \mathbb{R}, i = 1, \ldots, n$$

against the *Lehmann alternatives* [Lehmann (1953)] given by

$$H_n: \overline{F}_i(x) = [\overline{F}(x)]^{1+((c_i - \bar{c}_n)/C_n)\lambda} \qquad \forall x \in \mathbb{R}, \tag{4.1}$$

where $\lambda > 0$, $c_1, \dots, c_n$ are known constants, $\overline{c}_n = n^{-1}\sum_{i=1}^n c_i$, $C_n^2 = \sum_{i=1}^n (c_i - \overline{c}_n)^2$ and the following *Noether condition* holds:

$$\max_{1\le i\le n} \frac{(c_i - \overline{c}_n)^2}{C_n^2} \overset{n\to+\infty}{\longrightarrow} 0.$$

The hypothesis H_n is in the context of linear regression model, and it has been generalized in proportional hazards models, introduced by Cox (1972) and other multiplicative intensity models in semiparametric setups. The role of contiguity for theses cases will be addressed in a later communication.

In the case of the two-sample procedure, where we have samples of sizes n_1 and n_2, $n_1 + n_2 = n$, we take $c_i = 1/n_1$, $i = 1, \dots, n_1$ and $c_i = -1/n_2$, $i = n_1 + 1, \dots, n$ and $\overline{F}_i = \overline{F}$ for $i = 1, \dots, n_1$ and $\overline{F}_i = \overline{G}$ for $i = 1, \dots, n_2$.

Contiguity may be established in this case. Considering absolutely continuous failure times, it follows that under H_0, $f_i(x) = f(x)$, $i = 1, \dots, n$ whereas for H_n,

$$f_i(x) = \left(1 + \frac{c_i - \overline{c}_n}{C_n}\lambda\right)[\overline{F}(x)]^{((c_i-\overline{c}_n)/C_n)\lambda} f(x),$$

so that the log-likelihood ratio is

$$\begin{aligned}
\log L_n &= \sum_{i=1}^n \left\{\log\left(1 + \frac{c_i - \overline{c}_n}{C_n}\lambda\right) + \frac{c_i - \overline{c}_n}{C_n}\lambda \log[1 - F(X_i)]\right\} \\
&= \lambda \sum_{i=1}^n \left\{\frac{c_i - \overline{c}_n}{C_n} - \frac{\lambda^2}{2}\sum_{i=1}^n \frac{(c_i - \overline{c}_n)^2}{C_n^2} + o(1)\right. \\
&\qquad \left. + \lambda \sum_{i=1}^n \frac{c_i - \overline{c}_n}{C_n} \log[1 - F(X_i)]\right\} \\
&= -\frac{\lambda^2}{2} - \lambda \sum_{i=1}^n \frac{c_i - \overline{c}_n}{C_n}\bigl(-1 - \log[1 - F(X_i)]\bigr) + o(1).
\end{aligned}$$

If we write

$$W_i = \frac{c_i - \overline{c}_n}{C_n}\bigl(-1 - \log[1 - F(X_i)]\bigr),$$

we have that $W_1, \dots, W_n$ are independent random variables with $\mathbb{E}(W_i) = 0$ and $\mathbb{V}\mathrm{ar}(W_i) = (c_i - \overline{c}_n)^2/C_n^2$. Applying the Hájek–Šidák Central Limit theorem [see, e.g., Sen, Singer and Pedroso-de-Lima (2010)], we have

$$\sum_{i=1}^n W_i \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

so that

$$\log L_n \xrightarrow{\mathcal{D}} -\frac{\lambda^2}{2} - \lambda\mathcal{N}(0, 1) \equiv \mathcal{N}(-\lambda^2/2, \lambda^2).$$

Thus, by Corollary 1 of LeCam's First Lemma [see Hájek, Šidák and Sen (1999)], contiguity holds for the considered Lehmann alternatives.

4.3 General linear rank statistics

A simple linear rank statistic $T_n = T_n(X_1, \ldots, X_n)$ depends on $X_1, \ldots, X_n$ only through their ranks $R_{n1}, \ldots, R_{nn}$, where

$$R_{ni} = \sum_{j=1}^{n} I(X_i \geq X_j), \qquad i = 1, \ldots, n.$$

Let

$$T_n = \sum_{i=1}^{n} (c_i - \overline{c}_n) a_n(R_{ni}), \qquad n \geq 1, \tag{4.2}$$

where $c_1, \ldots, c_n$ are known constants and $a_n(1), \ldots, a_n(n)$ are scores, often expressed as $a_n(i) = \mathbb{E}[\phi(U_{n:i})]$ or $\phi[i/(n+1)]$, $1 \leq i \leq n$, where $U_{n:1} \leq \cdots \leq U_{n:n}$ are ordered random variables of a sample $U_1, \ldots, U_n$ from the Uniform(0, 1) distribution, so that $\mathbb{E}(U_{n:i}) = i/(n+1)$, $1 \leq i \leq n$. For the score function $\phi(\cdot)$ we will assume that, without loss of generality,

$$\overline{\phi} = \int_0^1 \phi(u)\, \mathrm{d}u = 0.$$

Based on (4.2), we define

$$T_n^* = \sum_{i=1}^{n} (c_i - \overline{c}_n) \phi(U_i), \tag{4.3}$$

where $\phi(\cdot)$ is the same score function defined earlier. If we take $\mathbf{R}_n = (R_{n1}, \ldots, R_{nn})$, and denote the distribution function of X_i by $F_i(\cdot)$, $i = 1, \ldots, n$, then, under $H_0 : F_1 = \cdots = F_n = F$,

$$\mathbb{E}(T^* | \mathbf{R}_n) = T_n, \tag{4.4}$$

since $(R_{n1}, \ldots, R_{nn})$ is independent of the order statistics $U_{n:j}$, $j = 1, \ldots, n$. Expression (4.4) shows that T_n may be thought of as a projection of T_n^* based on the vector of ranks. In order to establish the equivalence of T_n and T_n^*, note that

$$\mathbb{E}(T_n) = \sum_{i=1}^{n} (c_i - \overline{c}_n) \mathbb{E}[\phi(U_{n:R_{ni}})].$$

Since, under H_0, $\mathbb{E}[\phi(U_{n:R_{ni}})] = n^{-1} \sum_{i=1}^{n} \mathbb{E}[\phi(U_i)] = \bar{\phi} = 0$, it follows that $\mathbb{E}(T_n) = 0$ and $\mathbb{E}(T_n^*) = 0$. Also, using the projection of T_n^* based on $\mathbf{R}_n$, we have that under H_0,

$$\mathbb{E}(T_n - T_n^*) = \mathbb{E}(T_n^*)^2 - \mathbb{E}(T_n)^2. \tag{4.5}$$

After some algebraic computation, we get that

$$\mathbb{E}(T_n^2) = v_n^2 = A_n^2 C_n^2 \frac{n}{n-1}, \tag{4.6}$$

and

$$\mathbb{E}(T_n^*)^2 = C_n^2 A_\phi^2, \tag{4.7}$$

where $A_n^2 = n^{-1}\sum_{i=1}^n [a_n(i) - \overline{a}_n]^2$, $C_n^2 = \sum_{i=1}^n (c_i - \overline{c}_n)^2$ and

$$A_\phi^2 = \int_0^1 \phi^2(u)\,\mathrm{d}u.$$

Plugging (4.6) and (4.7) into (4.5), we have that, under H_0,

$$\mathbb{E}\left[\frac{T_n - T_n^*}{v_n^*}\right]^2 = \left(\frac{n-1}{n}\right)\frac{A_\phi^2}{A_n^2} - 1 \longrightarrow 0, \qquad n \to \infty,$$

since $A_n^2 \to A_\phi^2$, as $n \to \infty$. Therefore,

$$\frac{T_n - T_n^*}{v_n} \xrightarrow{\mathcal{P}} 0. \tag{4.8}$$

As

$$\frac{T_n}{v_n} = \frac{T_n - T_n^*}{v_n} + \frac{T_n^*}{v_n}, \tag{4.9}$$

the convergence in distribution for T_n/v_n may be derived from the convergence in distribution for T_n^*/v_n.

The asymptotic distribution of T_n^*/v_n may be derived based on LeCam's Third Lemma. For that, consider the Lehmann alternatives (4.1) discussed in the previous section and the corresponding log-likelihood function, that may be written as

$$\log L_n = \ell_n = -\frac{\lambda^2}{2} - \lambda \sum_{i=1}^n \frac{c_i - \overline{c}_n}{C_n}\phi^*(U_i) + o(1),$$

where $\phi^*(U_i) = -1 - \log(U_i)$, with $U_i \sim \text{Uniform}[0, 1]$.

For any real-valued constants α and β,

$$\begin{aligned}\alpha\frac{T_n^*}{v_n^*} + \beta\ell_n &= \sum_{i=1}^n \frac{c_i - \overline{c}_n}{C_n}\frac{1}{A_n^*}\alpha\phi(U_i) - \beta\frac{\lambda^2}{2} - \lambda\sum_{i=1}^n \frac{c_i - \overline{c}_n}{C_n}\beta\phi^*(U_i) + o(1) \\ &= -\frac{\lambda^2}{2}\beta + \sum_{i=1}^n \frac{c_i - \overline{c}_n}{C_n} W_i + o(1),\end{aligned} \tag{4.10}$$

where $W_1, \ldots, W_n$ are independent and identically distributed random variables, such that, under H_0,

$$\mathbb{E}(W_i) = \frac{\alpha}{A_n^*}\mathbb{E}[\phi(U_i)] - \lambda\beta\mathbb{E}[\phi^*(U_i)] = 0$$

and

$$\mathbb{V}\mathrm{ar}(W_i) = \alpha^2 + \lambda^2\beta^2 - 2\frac{\lambda}{A_n^*}\alpha\beta\sigma_{12},$$

where $\sigma_{12} = \mathrm{Cov}[\phi(U_i), \phi^*(U_i)] = \int_0^1 \phi(u)\phi^*(u)\,\mathrm{d}u = \langle \phi, \phi^* \rangle$.

Applying the Hájek–Šidák Central Limit Theorem in (4.10), we conclude that for all α and β,

$$\alpha\frac{T_n^*}{v_n^*} + \beta\ell_n \xrightarrow{\mathcal{D}} \mathcal{N}\Big(-\beta\frac{\lambda}{2}, \alpha^2 + \lambda^2\beta^2 - 2\alpha\beta\lambda\frac{\sigma_{12}}{A_\phi}\Big) \qquad \text{as } n \to \infty.$$

Therefore, by Cramér–Wold, under H_0,

$$\begin{pmatrix} T_n^*/v_n \\ \ell_n \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}_2\left[\begin{pmatrix} 0 \\ -\lambda/2 \end{pmatrix}; \begin{pmatrix} 1 & -\lambda\sigma_{12}/A_\phi \\ -\lambda\sigma_{12}/A_\phi & \lambda^2 \end{pmatrix}\right].$$

Then, if we apply Lemma 3 of LeCam's Third Lemma [see Hájek, Šidák and Sen (1999), p. 259], we have that, under H_n,

$$\frac{T_n^*}{v_n} \xrightarrow{\mathcal{D}} \mathcal{N}\Big(-\lambda\frac{\sigma_{12}}{A_\phi}; 1\Big). \tag{4.11}$$

The result in (4.11) shows that, asymptotically under H_n, the distribution of T_n^* (and hence, T_n) has noncentrality parameter given by

$$\nu_{T_n^*} = \frac{\lambda^2\sigma_{12}^2}{A_\phi^2} = \frac{\lambda^2}{A_\phi^2}[\langle \phi, \phi^* \rangle]^2.$$

Inspecting the expression of the log-likelihood ℓ_n more closely, we see that it involves the linear statistic

$$T_{\ell_n} = \sum_{i=1}^n \frac{c_i - \overline{c}_n}{C_n}\phi^*(U_i),$$

where $\phi^*(U_i) = -1 - \log(1 - U_i)$. Then, if we follow the same steps as in the case of T_n^*, we obtain

$$\nu_{T_{\ell_n}} = \frac{\lambda^2}{A_{\phi^*}^2}[\langle \phi^*, \phi^* \rangle]^2.$$

Since

$$\langle \phi^*, \phi^* \rangle = \int_0^1 \phi^*(u)\phi^*(u)\,\mathrm{d}u = \int_0^1 \phi^{*2}\,\mathrm{d}u = 1$$

and

$$A_{\phi^*}^2 = 1,$$

it turns out the noncentrality parameter for the asymptotic distribution of T_{ℓ_n}, under the alternative hypothesis, is

$$\nu_{T_{\ell_n}} = \lambda^2.$$

In order to compare both statistics, we note that the Pitman *asymptotic relative efficiency* (PARE) is

$$\begin{aligned}\mathrm{PARE}(T_n^*|T_{\ell_n}) &= \frac{[\langle\phi^*, \phi^*\rangle]^2}{A_{\phi^*}^2 \times 1} = \frac{(\int_0^1 \phi(u)\phi^*(u)\,\mathrm{d}u)^2}{(\int_0^1 \phi^2(u)\,\mathrm{d}u)(\int_0^1 \phi^{*2}(u)\,\mathrm{d}u)} \\ &= [\rho(\phi, \phi^*)]^2 \le 1,\end{aligned}$$

where the equality holds when $\phi = \phi^*$, that is, the (generalized) log-rank statistic. In other words, for Lehmann alternatives, the log-rank statistic is asymptotically optimal (or the most efficient).

In this case, we have a possibly unbounded score function, such as the normal scores where $\phi(u) = \Phi^{-1}(u)$, $u \in (0, 1)$ where $\Phi(x)$ is the standard normal distribution function. Chernoff and Savage (1958) initiated a general line of attack on the asymptotic theory (before the concept and use of contiguity become popular). In their approach whereas contiguity was not needed, it has some more stringent regularity conditions on the score function and its first and second derivatives. On the other hand, Hájek (1962), invoking contiguity, was able to relax these regularity conditions to practically square integrability of the score function, provided the underlying density admits a finite Fisher information. Of course, there are some densities (such as a mixture of two densities each having finite Fisher information, as arising in error contamination models) where the finite Fisher information clause is not tenable. This led Hájek (1968) to explore the general case without imposing contiguity and to show how contiguous alternatives can be conceived in that context as well. For details of these developments, we may refer to Hájek, Šidák and Sen (1999), Chapter 7.

4.4 Wilcoxon two-sample rank sum test

This rank test statistic is expressible in terms of a generalized U-statistics and its asymptotic normality does not need contiguity of probability measures. The closeness of the two distributions allow the uniform consistency of the variance in a neighborhood of the null hypothesis. This is expected because the score function is bounded and continuous.

4.5 Contiguity and differentiable statistical functionals

Unlike the case of U-statistics, statistical functionals in general may not be of a finite degree. For example, the population median of a distribution F, defined on $\mathbb{R}$, is a functional $\theta(F) = F^{-1}(1/2)$. In general, a p-quantile of F is defined as

$\theta_p(F) = F^{-1}(p)$, $0 < p < 1$. The corresponding sample counterpart, based on the empirical distribution function $F_n(\cdot)$, is defined as $F_n^{-1}(p)$, with some refinements when $(n+1)p$ is not an integer. Since F_n is a step function, the population functional $\theta(F)$ and its sample counterpart $\widehat{\theta}_n = \theta(F_n)$ may not share the same continuity and other properties. Moreover, often the alternative hypotheses is formulated in a nonparametric or nonstandard way wherein the usual way of constructing the likelihood ratio test statistic and verifying LeCam's First Lemma may be difficult. In some cases, the asymptotic distribution may not be even normal or chi-square distributions. In that way, the role of contiguity is somewhat subdued in the study of asymptotic properties of these statistics. In the case of sample quantiles, whenever the probability density function at the population quantile is continuous and positive, asymptotic normality holds without recourse to contiguity arguments. This asymptotic results is based on a representation, known as the *Bahadur representation of sample quantile* [see Bahadur (1966)], which enables a first-order representation in terms of a linear statistics with a reminder term that stochastically converges to 0 at a faster rate. In a general framework, we could write under appropriate (Hadamard, Fréchet or Gâteaux) differentiability conditions:

$$\theta(F_n) - \theta(F) = n^{-1} \sum_{i=1}^{n} T_1(X_i; F) + o_p(n^{-1/2}), \tag{4.12}$$

where $T_1(X_i; F)$, $i = 1, \ldots, n$ are centered at 0 and have a finite variance σ_T^2. Thus, the first term on the r.h.s. of (4.12) represents a linear functional where the central limit theorem applies and that paves the way to asymptotic normality, based on the Slutsky theorem. The functional $T_1(\cdot; F)$ is often termed the *influence function*, and their average is the so called *Hadamard derivative* of $\theta(F_n)$ at F. Thus, basically, here contiguity is bypassed by verifying the *first-order representation* of $\theta(F_n)$. In general, we may consider under more stringent differentiability conditions, a higher order representation for $\theta(F_n)$ similar to the case of U-statistics or von Mises functions, treated earlier. In that way, if $T_1(x; F) = 0$ a.e., $\theta(F_n) - \theta(F)$ is dominated by the second-order term, a quadratic functional whose asymptotic distribution may not be chi-square but a von Mises type distribution involving a convolution of scalar multiples of independent variables having chi-square distributions with one degree of freedom. In any case, the corresponding nonnull distribution may not be representable by the same linear combination of independent noncentral chi-square variables. Let us illustrate this feature with the following example:

The Cramér–von Mises statistic: Let $X_1, \ldots, X_m$ be independent and identically distributed random variables with a continuous distribution $F(x)$, defined on $\mathbb{R}$, and let $Y_1, \ldots, Y_n$ be a second independent sample from a continuous distribution $G(x)$, also defined on $\mathbb{R}$. Let $F_m(x)$ and $G_n(x)$ be the respective empirical distributions and let $N = m + n$, $H_N(x) = (m/N)F_m(x) + (n/N)G_n(x)$ be the combined sample empirical distribution. Similarly, let $H(x) = H_{(N)}(x) = (m/n)F(x) + (n/N)G(x)$. Note that under the null hypothesis $H_0 : F = G$, the functional $\Delta(F, G) = \int_{\mathbb{R}} (F(x) - G(x))^2 \, \mathrm{d}H(x) = 0$, and it is nonnegative for any $F \neq G$. Consider the sample counterpart

$$T_N = (mn/N) \int_{\mathbb{R}} [F_m(x) - G_n(x)]^2 \, \mathrm{d}H_N(x). \tag{4.13}$$

Under the null hypothesis, the asymptotic distribution of T_N agrees with the distribution of $\sum_{j=1}^{\infty} Z_j^2/(j^2\pi^2)$ where the Z_j are independent and identically distributed random variables with the standard normal distribution [viz., Theorem 3 of Hájek, Šidák and Sen (1999), p. 221]. If, however, we consider a local alternative where $\Delta(F, G) = \Delta_N = N^{-1/2}\gamma$ for some fixed $\gamma > 0$, irrespective of being a contiguous alternative or not, the asymptotic distribution of T_N will not be of the same form but involving noncentral counterparts of the Z_j^2. If in defining such local alternatives, some norms (like the Kolmogorov) are chosen (viz., $d(F, G) = \sup_x |F(x) - G(x)|$), LeCam's lemmas may not be applicable in deriving their nonstandard distributions.

5 Contiguity in constrained statistical inference

In constrained statistical inference (CSI) with inequality, ordered, or shape restraints on parameters or sample spaces, standard asymptotics on statistical inference may not generally hold [Silvapulle and Sen (2005)]. In fact, in some most simple CSI cases arising in parametric as well as nonparametric approaches, variants of chi-squared or beta distributions, known as the chi-bar square or E-bar distributions relate to the null hypothesis case whereas their nonnull (even for contiguous alternatives) counterparts may not have the noncentral chi-bar square or non-central E-bar distributions. In more complex CSI setups, it may be difficult to formulate contiguity and simple statistical tests. Thus, contiguity seems to have a somewhat limited role in the derivation of asymptotic distribution theory of tests and estimates in CSI. Going back to the three basic lemmas of LeCam [Hájek, Šidák and Sen (1999), Chapter 7], we can comment that whereas Lemmas 1 and 2 characterize contiguity, it is Lemma 3 which enables contiguity to provide a simple asymptotic nonnull distribution in standard cases. It is the third lemma that may not serve the same prime role in CSI. In this respect, under null hypotheses, both restricted likelihood ratio tests (RLRT) and union-intersection tests (UIT) share the nonstandard asymptotics to a certain extent, albeit their nonnull distributions even under contiguous (constrained) alternatives are mostly difficult to formulate. We illustrate this feature with a couple of examples.

5.1 Tests for a binomial population under mixture

Consider the model

$$\begin{aligned} P\{X = k\} &= \binom{n}{k} [(1-\alpha)\pi_1^k (1-\pi_1)^{(n-k)} + \alpha\pi_2^k (1-\pi_2)^{n-k}], \\ &\quad k = 0, 1, \ldots, n, \end{aligned} \tag{5.1}$$

where $\alpha \in (0, 1)$ and π_1, π_2 are in $(0, 1)$, all unknown. The null hypothesis H_0 specifies that it is a single binomial law and the alternative hypothesis is that it is the mixture in (5.1). Note that in the above model, we have a single binomial population when either α is 0 or 1, or $\pi_1 = \pi_2$, irrespective of $\alpha \in (0, 1)$. Thus, if we consider the parameter space $\boldsymbol{\Theta} = [0, 1]^3$, the null hypothesis refers to a union of a subspace of dimension 1, namely the vertexes of $\boldsymbol{\Theta}$ with $\alpha = \{0, 1\}$ and a two-dimensional subspace where $\pi_1 = \pi_2$, $\alpha \in (0, 1)$. This violates the usual regularity condition that the parameter space under H_0 is an inner subspace of $\boldsymbol{\Theta}$, and the log-likelihood ratio test statistic under the null hypothesis fails to have the usual asymptotic chi-square distribution with 2 degrees of freedom. Contiguity rests on a shrinking neighborhood of the null parameter space but lack of identifiability precludes an easy verification of that through LeCam's First and Second Lemmas. In this particular case, the asymptotic null distribution is a chi-bar distribution, though under alternatives, it may not be a noncentral chi-bar distribution. The difference comes from the fact that the mixing coefficients in the nonnull case themselves may depend on the noncentrality which is no longer homogeneous across the alternative parameter space.

5.2 Multivariate normal mean: Positive orthant alternative

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent p-dimensional random vectors such that

$$\mathbf{X}_j \sim \mathcal{N}_p(\boldsymbol{\theta}, \boldsymbol{\Sigma}),$$

where both the p-vector $\boldsymbol{\theta}$ and the $p \times p$ matrix $\boldsymbol{\Sigma}$ are unknown. Consider that our primary interest is on $\boldsymbol{\theta}$ (i.e., $\boldsymbol{\Sigma}$ is a nuisance parameter) and the hypotheses

$$H_0 : \boldsymbol{\theta} = \mathbf{0} \quad \text{against} \quad H_1 : \boldsymbol{\theta} \geq \mathbf{0}, \tag{5.2}$$

where the inequality in the alternative hypothesis is taken elementwise.

The likelihood in this case is

$$L(\mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\Sigma}) = (2\pi)^{-np/2} |\boldsymbol{\Sigma}|^{-n/2} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} (\mathbf{X}_i - \boldsymbol{\theta})' \boldsymbol{\Sigma}^{-1} (\mathbf{X}_i - \boldsymbol{\theta}) \right\},$$

where we assume that $\boldsymbol{\Sigma}$ is a nonsingular matrix. Under H_0, such likelihood is

$$L(\mathbf{X}, \mathbf{0}, \boldsymbol{\Sigma}) = (2\pi)^{-np/2} |\boldsymbol{\Sigma}|^{-n/2} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} \mathbf{X}_i' \boldsymbol{\Sigma}^{-1} \mathbf{X}_i \right\},$$

and the corresponding maximum likelihood estimator for $\boldsymbol{\Sigma}$ is readily computed as

$$\widehat{\boldsymbol{\Sigma}}^0 = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i'.$$

The unrestricted MLE of $\boldsymbol{\theta}$ is the sample mean vector $\overline{\mathbf{X}}_n$. Hence, even if we consider local alternatives $H_{1n}: \boldsymbol{\theta} = n^{-1/2}\boldsymbol{\lambda}$, $\boldsymbol{\lambda} > \mathbf{0}$, treating $\boldsymbol{\Sigma}$ as nuisance parameter, direct computation leads to the following expression for the log-likelihood ratio statistic

$$\sqrt{n}\overline{\mathbf{X}}_n'\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda} - (1/2)\boldsymbol{\lambda}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda},$$

which is exactly normally distributed with mean equal to $(-1/2)$ variance, and hence, by LeCam's Second Lemma, contiguity holds.

Let us next consider tests for H_0 against the positive orthant alternative considering the union-intersection test (UIT) approached in Silvapulle and Sen (2005). For testing the hypotheses in (5.2) we may write, for a p-vector $\mathbf{a}$,

$$H_0: \bigcap_{\mathbf{a}\in\mathbb{R}^{+p}} H_{0\mathbf{a}} \quad \text{against} \quad H_1: \bigcup_{\mathbf{a}\in\mathbb{R}^{+p}} H_{1\mathbf{a}}^{+},$$

where $H_{0\mathbf{a}}: \mathbf{a}'\boldsymbol{\theta} = 0$ and $H_{1\mathbf{a}}^{+}: \mathbf{a}'\boldsymbol{\theta} > 0$. For $\mathcal{P} = \{1, \ldots, p\}$, consider $\varnothing \subseteq \ell \subseteq \mathcal{P}$, denoting the complement by ℓ' and the cardinality by $0 \le |\ell| \le p$. For a given ℓ, we partition $\overline{\mathbf{X}}_n$ and $\boldsymbol{\Sigma}$ in the following way

$$\overline{\mathbf{X}}_n = \begin{pmatrix} \overline{\mathbf{X}}_{n\ell} \\ \overline{\mathbf{X}}_{n\ell'} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{\ell\ell} & \boldsymbol{\Sigma}_{\ell\ell'} \\ \boldsymbol{\Sigma}_{\ell'\ell} & \boldsymbol{\Sigma}_{\ell'\ell'} \end{pmatrix}$$

and then denote

$$\overline{\mathbf{X}}_{n\ell:\ell'} = \overline{\mathbf{X}}_{n\ell} - \boldsymbol{\Sigma}_{\ell\ell'}\boldsymbol{\Sigma}_{\ell'\ell'}^{-1}\overline{\mathbf{X}}_{n\ell'}$$

and

$$\boldsymbol{\Sigma}_{\ell\ell:\ell'} = \boldsymbol{\Sigma}_{\ell\ell} - \boldsymbol{\Sigma}_{\ell\ell'}\boldsymbol{\Sigma}_{\ell'\ell'}^{-1}\boldsymbol{\Sigma}_{\ell'\ell}.$$

Also, we define

$$I_{n\ell} = I(\overline{\mathbf{X}}_{n\ell:\ell'} > \mathbf{0},\ \boldsymbol{\Sigma}_{\ell'\ell'}^{-1}\overline{\mathbf{X}}_{n\ell'} \le \mathbf{0}).$$

The UIT statistic is then

$$T_n^{*2} = n \sum_{\ell\subseteq\mathcal{P}} \overline{\mathbf{X}}_{n\ell:\ell'}'\boldsymbol{\Sigma}_{\ell\ell:\ell'}^{-1}\overline{\mathbf{X}}_{n\ell:\ell'} I_{n\ell}.$$

Under the null hypothesis, the quadratic form $\bar{\mathbf{X}}_{n\ell:\ell'}'\boldsymbol{\Sigma}_{\ell\ell:\ell'}^{-1}\bar{\mathbf{X}}_{n\ell:\ell'}$ and $I_{n\ell}$ are stochastically independent; for any given $\ell \in \mathcal{P}$, the quadratic form has the chi-square distribution with degrees of freedom equals to $|\ell|$; we note that $I_{n\ell} = 1$ with probability equal to the corresponding orthant probability. As such, T_n^* has the chi-square bar distribution with coefficients given by the 2^p quadrant probabilities.

We consider now the situation where $\boldsymbol{\Sigma}$ is unspecified, being replaced by $\mathbf{S}_n = n^{-1}\sum_{i=1}^{n}(\mathbf{X}_i - \overline{\mathbf{X}}_n)(\mathbf{X}_i - \overline{\mathbf{X}}_n)'$, so that in the previous expressions we consider the partitioned matrix

$$\mathbf{S}_n = \begin{pmatrix} \mathbf{S}_{n\ell\ell} & \mathbf{S}_{n\ell\ell'} \\ \mathbf{S}_{n\ell'\ell} & \mathbf{S}_{n\ell'\ell'} \end{pmatrix}$$

and, also, write

$$\mathbf{S}_{n\ell:\ell'} = \mathbf{S}_{n\ell\ell} - \mathbf{S}_{n\ell\ell'}\mathbf{S}_{n\ell'\ell'}^{-1}\mathbf{S}_{n\ell'\ell}$$

so that

$$I_{n\ell} = I(\overline{\mathbf{X}}_{n\ell:\ell'} > \mathbf{0},\ \mathbf{S}_{n\ell'\ell'}^{-1}\overline{\mathbf{X}}_{n\ell'} \leq \mathbf{0}).$$

Now, the UIT statistic is given by

$$T_n^{*2} = n\sum_{\ell\subseteq\mathcal{P}} \overline{\mathbf{X}}'_{n\ell:\ell'}\mathbf{S}_{n\ell\ell:\ell'}^{-1}\overline{\mathbf{X}}_{n\ell:\ell'}I_{n\ell}.$$

In this setup, the likelihood ratio test statistic will be given by [see Silvapulle and Sen (2005)]

$$L_n = \sum_{\ell\subseteq\mathcal{P}} I_{n\ell}(n\overline{\mathbf{X}}'_{n\ell:\ell'}\mathbf{S}_{n\ell\ell:\ell'}^{-1}\overline{\mathbf{X}}_{n\ell:\ell'})(1+\overline{\mathbf{X}}'_{n\ell'}\mathbf{S}_{n\ell'\ell'}^{-1}\overline{\mathbf{X}}_{n\ell'})^{-1}$$

so that

$$L_n \leq T_n^{*2},$$

where the equality holds when the vector $\overline{X}_n$ lies in the positive orthant, which occurs with probability less than one.

It follows that the UIT and LRT are then different and none has strictly a specified null distribution; even their asymptotic distributions are complicated, not chi-bar square in the true sense.

The complications arising in this restricted alternative problem for the nonnull distribution is due to the fact that $I_{n\ell}$ and the associated quadratic form for different $\ell \in \mathcal{P}$ are no longer stochastically independent, so that Lemma 3.2 of Perlman (1969) no longer holds, and further, the orthant probabilities $P\{X \in \ell\}$, $\ell \in \mathcal{P}$ also depend on the unspecified $\boldsymbol{\lambda}$. Note that even if in (5.2) we replace $\boldsymbol{\theta}$ by $n^{-1/2}\boldsymbol{\lambda}$, the distribution of $n^{1/2}\overline{X}_n$ will be multinormal with mean vector $\boldsymbol{\lambda}$ and same covariance matrix as $\mathbf{X}$. This will change the orthant probabilities for $n^{-1/2}\overline{X}_n$, which will affect the independence of $I_{n\ell}$ and the RMLE of $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$, invalidating Lemma 3.2 of Perlman (1969). As such this does not yield a noncentral chi-bar square distribution, even for contiguous alternatives.

6 Concluding remarks

There are a few important results where contiguity plays a basic role, and we briefly outline some of these perspectives in this concluding section.

6.1 Uniform asymptotic linearity in probability

In the context of robust R-estimation of location and regression parameters, aligned rank statistics were incorporated in the formulation of suitable estimating equations. Similarly, in robust M-estimation of location and regression, suitable aligned M-statistics are used. There are other situations where such alignments are incorporated to formulate suitable estimating equations yielding appropriate estimators. Even the maximum likelihood estimators, mostly based on implicit equations, are based on such alignment; we refer to Chapter 3 of Jurečková and Sen (1996). Basically, asymptotic properties of such estimators are more conveniently studied by some appropriate *uniform asymptotic linearity in probability result* wherein contiguity plays a basic role. We explain this with the simple case of Wilcoxon score estimation of location, and a similar result follows for other examples too.

We define

$$W_n(b) = \sum_{i=1}^{n} \operatorname{sign}(X_i - b) R_{ni}^{+}(b),$$

where $R_{ni}^{+}(b)$ is the rank of $|X_i - b|$ among the n values $|X_j - b|$, $j = 1, \ldots, n$. If θ is the true location parameter (point of symmetry of the distribution F), then $W_n(\theta)$ is distribution free with mean 0 and variance $n(n+1)(2n+1)/6$. Further, $W_n(b)$ is a nonincreasing (step) function of $b \in \mathbb{R}$. Hence, equating $W_n(b)$ to 0, one gets the R-estimator of θ. The uniform asymptotic linearity result states that, for every $C < \infty$,

$$\sup_{\sqrt{n}|b-\theta| \le C} \{|W_n(b) - W_n(\theta) + n(n+1)\gamma(b-\theta)|\} = o_p\left(\sqrt{\frac{n(n+1)(2n+1)}{6}}\right),$$

where the variance of $W_n(\theta)$ is equal to $n(n+1)(2n+1)/6$. A similar result holds for the log-likelihood ratio statistics under the Cramér–Rao regularity conditions. Basically, in deriving the above result, pointwise (in b), one can use contiguity of the probability measures and then the conclusion is extended to the interval by invoking the nonincreasing property of $W_n(b)$ in b. For general rank scores or likelihood ratio scores statistics, this contiguity greatly simplifies the process [Jurečková and Sen (1996), Chapters 3–6].

6.2 Contiguity and compactness of derived stochastic processes

In the context of statistical inference on stochastic processes with special emphasis on invariance principles, it has been observed that pointwise a stochastic function, say $S(t)$ may have a first-order representation by a simpler function which is more amenable to stochastic analysis. In such a case, it may be simpler to verify the tightness of the process $S(t)$ under suitable measures $\{P_n\}$ (usually related to suitable null hypothesis), and weak invariance principles can be formulated under $\{P_n\}$.

However, if some alternative measures $\{Q_n\}$ are considered, direct verification of such *tightness* or compactness condition may often be very cumbersome [viz., Sen (1981), Chapters 3–7]. In this context, the following theorem [Sen (1981), p. 100] seems to be very useful.

If $\{Q_n\}$ is contiguous to $\{P_n\}$ and a process is tight under the P_n-measure, then it remains tight under the Q_n-measure as well.

This result is not confined to asymptotic Gaussian processes and can be used in more complex situations too.

6.3 Contiguity, regular estimators and convolution theorems

Following the notion in Section 3, we elaborate here the role of contiguity in the so-called regular estimators in a general setup as adapted from Hájek (1970, 1972) and Inagaki (1970, 1973), both exploiting the basic idea of LeCam (1960), in a parametric setup.

Consider a family of probability measures $\{P_{\boldsymbol{\theta},n} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ on some measure space $\{(\mathcal{X}_n, \mathcal{A}_n);\ n \geq 1\}$ where $\boldsymbol{\theta}$ is a finite-dimensional parameter lying in the interior of $\mathbb{R}^k$, for some $k \geq 1$. Following LeCam (1960), we consider a bounded set $\mathcal{B}$ and a sequence $\{\delta_n\}$ of real numbers, such that $\{P_{\boldsymbol{\theta}+\delta_n \mathbf{t}_n}\}$ is contiguous to $\{P_{\boldsymbol{\theta}}\}$, for all $\mathbf{t}_n \in \mathcal{B}$. Typically, $\delta_n = O(n^{-1/2})$. We denote the log-likelihood function by $\Lambda(\boldsymbol{\theta} + \delta_n \mathbf{t}_n; \boldsymbol{\theta}) = \log\{\mathrm{d}P_{\boldsymbol{\theta}+\delta_n\mathbf{t}_n}/\mathrm{d}P_{\boldsymbol{\theta},n}\}$. Recall that the model $\{P_{\boldsymbol{\theta}_n}, \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ is LAN at $\boldsymbol{\theta}_0$ if (i) for $\boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + \delta_n \boldsymbol{h}$, $\boldsymbol{h} \in R^k$,

$$\Lambda(\boldsymbol{\theta}_n, \boldsymbol{\theta}_0) - \mathbf{t}_n' \mathbf{U}_n + (1/2)\boldsymbol{h}'\mathcal{I}(\boldsymbol{\theta}_0)\boldsymbol{h} \to 0,$$

in $P_{\boldsymbol{\theta}_0,n}$-probability, where $\mathbf{U}_n = \mathbf{U}_n(\boldsymbol{\theta}_o)$ is the Rao score statistic, $\mathcal{I}(\boldsymbol{\theta}_0)$ is the (Fisher) information (matrix) at $\boldsymbol{\theta}_0$, and (ii) under $P_{\boldsymbol{\theta}_o,n}$,

$$\mathbf{U}_n \xrightarrow{\mathcal{D}} \mathcal{N}[0, \mathcal{I}(\boldsymbol{\theta}_0)].$$

Let $\{\mathbf{T}_n\}$ be a sequence of estimators of $\boldsymbol{\theta}$, and let $\boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + n^{-1/2}\boldsymbol{h}$, $\boldsymbol{h} \in \mathbb{R}^k$. Denote by $H_n(\cdot; \boldsymbol{\theta})$ the probability law for $n^{1/2}(\mathbf{T}_n - \boldsymbol{\theta})$ under $P_{\boldsymbol{\theta},n}$. Then $\mathbf{T}_n$ is said to be regular at $\boldsymbol{\theta}_0$, if $H_n(\cdot; \boldsymbol{\theta}_n) \Rightarrow H(\cdot; \boldsymbol{\theta}_0)$, where $H(\cdot.; \boldsymbol{\theta}_0)$ does not depend on $\boldsymbol{h}$. It is possible to replace *regular* by *essentially regular* estimators wherein we replace "for all $\boldsymbol{h} \in \mathbb{R}^k$" by "for all $\boldsymbol{h} \in D_k \subset \mathbb{R}^k$, for some unique D_k." The Hájek convolution theorem may then be presented as follows:

Suppose that $\{P_{\boldsymbol{\theta},n}\}$ is LAN at $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$, and $\{\mathbf{T}_n\}$ is a sequence of regular estimators at $\boldsymbol{\theta}_0$. Then under $\{P_{\boldsymbol{\theta},n}\}$, $n^{1/2}(\mathbf{T}_n - \boldsymbol{\theta}_0) = \mathcal{I}^{-1}(\boldsymbol{\theta}_0)\mathbf{U}_n + \boldsymbol{\psi}_n + o_p(1)$, where $\boldsymbol{\psi}_n$ is (asymptotically) independent of $\mathbf{U}_n$. Thus, $H(\boldsymbol{\theta}_0) = \Phi(0, \mathcal{I}^{-1}(\boldsymbol{\theta}_0)) * G(\cdot; \boldsymbol{\theta}_0)$ where $*$ stands for the convolution, $\Phi(\cdot)$ for the Gaussian distribution and $G(\cdot; \boldsymbol{\theta}_0)$ for the distribution of $\boldsymbol{\psi}_n$.

As a corollary, it follows that if $\mathbf{T}_n$ is BAN, then $\boldsymbol{\psi}_n$ is $\mathbf{0}$ a.e. so that $\mathbf{T}_n$ is asymptotically equivalent to $\mathbf{U}_n$. Actually, Hájek (1970) established a stronger result on the asymptotic limits of mean squared error (or dispersion matrix) of regular estimators under some extra regularity conditions. As far as the convolution theorem is concerned, the above result suffices. Following Sen (2000), we may remark that contiguity along with the first-order asymptotic representation paves the way for the convolution theorem for Bayes estimators and Bayes version of some robust estimators. In conclusion, we may note that the contiguity as assumed in the above setup essentially relates to a parametric setup. It can also be developed for a semiparametric model when the null hypothesis and the alternative relate to the parametric component, treating the nonparametric component as a nuisance functional. This was the case with the Lehmann–Cox proportional hazards model treated earlier. We intend to explore this further in more complex semiparametric models arising in survival and reliability analysis, extending the notion to some conditional model and incorporating suitable martingale theory.

Acknowledgments

This work was partially supported by Fundação de Apoio à Pesquisa do Estado de São Paulo—FAPESP, Brazil, Grant number 2010/05967-9 and by the Boshamer Foundation at The University of North Carolina, USA.

References

Bahadur, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37**, 577–580. MR0189095

Chernoff, H. and Savage, I. R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. *Ann. Math. Statist.* **29**, 972–994. MR0100322

Cox, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34**, 187–220. MR0341758

Hájek, J. (1962). Asymptotically most powerful rank-order tests. *Ann. Math. Statist.* **33**, 1124–1147. MR0143304

Hájek, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives. *Ann. Math. Statist.* **39**, 325–346. MR0222988

Hájek, J. (1970). A characterization of limiting distributions of regular estimates. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **14**, 323–330.

Hájek, J. (1972). Local asymptotic minimax and admissibility in estimation. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. I: Theory of Statistics* 175–194. Berkeley: Univ. California Press. MR0400513

Hájek, J., Šidák, Z. and Sen, P. K. (1999). *Theory of Rank Tests*, 2nd ed. San Diego, CA: Academic Press Inc. MR1680991

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19**, 293–325. MR0026294

Inagaki, N. (1970). On the limiting distribution of a sequence of estimators with uniformity property. *Ann. Inst. Math. Statist.* **22**, 1–13. MR0273725

Inagaki, N. (1973). Asymptotic relations between the likelihood estimating function and the maximum likelihood estimator. *Inst. Statist. Math.* **25**, 1–26. MR0339388

Jurečková, J. and Sen, P. K. (1996). *Robust Statistical Procedures. Asymptotics and Interrelations*. New York: John Wiley & Sons, Inc. MR1387346

Kendall, M. and Stuart, A. (1977). *The Advanced Theory of Statistics*, Vol. 1, *Distribution Theory*, 4th ed. New York: Macmillan Publishing Co., Inc. MR0467977

LeCam, L. (1960). Locally asymptotically normal families of distributions. Certain approximations to families of distributions and their use in the theory of estimation and testing hypotheses Convergence of random processes and limit theorems in probability theory. *Univ. California Publ. Statist.* **3**, 37–98. MR0126903

LeCam, L. (1979). On a theorem of J. Hájek. In *Contributions to Statistics* 119–135. Dordrecht: Reidel. MR0561265

Lehmann, E. L. (1953). The power of rank tests. *Ann. Math. Statist.* **24**, 119–135. MR0054208

Perlman, M. D. (1969). One-sided testing problems in multivariate analysis. *Ann. Math. Statist.* **40**, 549–567. MR0239689

Pinheiro, A., Sen, P. K. and Pinheiro, H. P. (2011). A class of asymptotically normal degenerate quasi U-Statistics. *Ann. Inst. Statist. Math.* DOI 10.1007/s10463-010-0271-z. To appear.

Sen, P. K. (1981). *Sequential Nonparametrics: Invariance Principles and Statistical Inference*. New York: John Wiley & Sons Inc. MR0633884

Sen, P. K. (2000). The Hájek convolution theorem and empirical Bayes estimation: Parametrics, semiparametrics and nonparametrics analysis. *J. Statist. Plann. Inference* **91**, 541–556. MR1814800

Sen, P. K., Singer, J. M. and Pedroso-de-Lima, A. C. (2010). *From Finite Sample to Asymptotic Methods in Statistics*. New York: Cambridge Univ. Press. MR2566692

Silvapulle, M. J. and Sen, P. K. (2005). *Constrained Statistical Inference. Inequality, Order, and Shape Restrictions*. New York: Wiley-Interscience. MR2099529

Department of Biostatistics and
Department of Statistics and Operational Research
University of North Carolina
Chapel Hill, North Carolina 27599-7420
USA
E-mail: pksen@bios.unc.edu

Department of Statistics
University of São Paulo
São Paulo, SP, 05508-090
Brazil
E-mail: acarlos@ime.usp.br