Asymptotic Relative Efficiency (ARE)

Chapter · January 2013

DOI: 10.1002/9780470057339.vnn077

1 author:

Marc Hallin
Université Libre de Bruxelles

Asymptotic relative efficiency (ARE)

Introduction

The purpose of asymptotic relative efficiency is to compare two statistical procedures by comparing the sample sizes, n_1 and n_2, say, at which those procedures achieve some given measure of performance; the ratio n_2/n_1 is called the *relative efficiency* of procedure one with respect to procedure two. Finite-sample evaluations being difficult or impossible, a sequence of measures of performances requiring that those sample sizes go to infinity is generally considered. If those measures of performance are indexed by n, say, so that n_1 and n_2 take the form $n_1(n)$ and $n_2(n)$, the limit $\lim_{n\to\infty} n_2(n)/n_1(n)$, if it exists, is called the *asymptotic relative efficiency* of procedure one with respect to procedure two.

Decision-theoretical Background and Basic Definitions

Consider a sequence of statistical models indexed by $n \in \mathbb{N}$, under which the distribution $\mathrm{P}^{(n)}_{\boldsymbol{\theta}}$ of some observation $\mathbf{X}^{(n)}$ belongs to the parametric family $\mathcal{P}^{(n)} := \{\mathrm{P}^{(n)}_{\boldsymbol{\theta}} \mid \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$, where $\boldsymbol{\Theta}$ is some (Borel) subset of $\mathbb{R}^K$ (for simplicity, we restrict to parametric models, but nonparametric ones also can be investigated).

Typically, $\mathbf{X}^{(n)}$ is of the form $(\mathbf{X}^{(n)}_1, \dots, \mathbf{X}^{(n)}_n)$, where the $\mathbf{X}^{(n)}_i$'s take values in $\mathbb{R}^k$ (equipped with the Borel σ-field $\mathcal{B}^k$), and n represents a sample size, the length of some k-dimensional vector time series, or the number of sites in an observed random field. In the sequel, we treat n as a sample size.

Associated with that sequence of models, consider two sequences of statistical procedures $\delta^{(n)}_1$ and $\delta^{(n)}_2$ addressing the same decision problem, that is, mapping the points $\mathbf{x}^{(n)}$ of the observation space to some decision space $\mathcal{D}$. The most familiar examples are

1. (point estimation) $\delta^{(n)}_i(\mathbf{X}^{(n)})$, $i = 1, 2$ are estimators of the actual value of some function $\psi(\boldsymbol{\theta})$, hence take values in $\mathcal{D} = \psi(\boldsymbol{\Theta})$. In the sequel, for simplicity, we restrict to $\psi(\boldsymbol{\theta}) = \boldsymbol{\theta}$, hence $\mathcal{D} = \boldsymbol{\Theta}$ (that is, estimation of $\boldsymbol{\theta}$ itself).
2. (hypothesis testing) $\delta^{(n)}_i(\mathbf{X}^{(n)})$, $i = 1, 2$ are tests for some null hypothesis $\mathcal{H}_0 \subsetneq \boldsymbol{\Theta}$ versus the alternative $\mathcal{H}_1 = \boldsymbol{\Theta} \setminus \mathcal{H}_0$; the traditional notation, in that case, is $\phi^{(n)}_i$ instead of $\delta^{(n)}_i$, with $\mathcal{D} = \{0, 1\}$ and $\phi^{(n)}_i(\mathbf{X}^{(n)}) = 1$ (resp. $\phi^{(n)}_i(\mathbf{X}^{(n)}) = 0$) meaning rejection (resp. nonrejection) of $\mathcal{H}_0$ (for simplicity, we only consider nonrandomized tests).

Measuring the performance, for given n and $\boldsymbol{\theta}$, of a statistical procedure $\delta^{(n)}$ involves some loss function $(d, \theta) \mapsto L(d, \boldsymbol{\theta})$, where $L(d, \boldsymbol{\theta})$ represents the loss incurred if decision $d \in \mathcal{D}$ is made while the actual value of the parameter is $\boldsymbol{\theta}$. The performance of $\delta^{(n)}$ then is evaluated via the expectation $R^{\delta^{(n)}}(\boldsymbol{\theta}) := \mathrm{E}_{\boldsymbol{\theta}}[L(\delta^{(n)}, \boldsymbol{\theta})]$, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}}$, of the loss, called the *risk* associated with $\delta^{(n)}$. That risk depends on n and $\boldsymbol{\theta}$. The smaller the risk, the better (at given n and $\boldsymbol{\theta}$) the procedure $\delta^{(n)}$. Going back to the examples of point estimation and hypothesis testing,

1. (point estimation) traditional choices, in point estimation, are (for univariate θ) the quadratic loss $L(\delta^{(n)}, \theta) = (\delta^{(n)} - \theta)^2$ or the absolute deviation $L(\delta^{(n)}, \theta) = |\delta^{(n)} - \theta|$, leading to expected squared and expected absolute deviation risks; the quadratic loss extends to the multiparameter case with positive definite quadratic forms $L(\delta^{(n)}, \boldsymbol{\theta}) = (\delta^{(n)} - \boldsymbol{\theta})'\mathbf{A}(\delta^{(n)} - \boldsymbol{\theta})$, requiring the choice of a positive definite matrix $\mathbf{A}$.
2. (hypothesis testing) the sequences of tests $\phi^{(n)}$ under consideration are limited to sequences of α-level or asymptotically α-level sequences (recall that a sequence of tests $\phi^{(n)}$ is an *α-level sequence* if $\mathrm{E}_{\boldsymbol{\theta}}[\phi^{(n)}(\mathbf{X}^{(n)})] \le \alpha$ for any $\boldsymbol{\theta} \in \mathcal{H}_0$ and $n \in \mathbb{N}$; it is an *asymptotically α-level sequence* if $\limsup_{n\to\infty} \mathrm{E}_{\boldsymbol{\theta}}[\phi^{(n)}(\mathbf{X}^{(n)})] \le \alpha$ for any $\boldsymbol{\theta} \in \mathcal{H}_0$) and the loss is, traditionally, $L(0, \boldsymbol{\theta}) = 1$ or 0 according as $\boldsymbol{\theta} \in \mathcal{H}_1$ or $\boldsymbol{\theta} \in \mathcal{H}_0$, and $L(1, \boldsymbol{\theta}) = 0$ for any $\boldsymbol{\theta}$. The risk is then the probability $1 - \mathrm{E}_{\boldsymbol{\theta}}[\phi^{(n)}]$ of not rejecting $\mathcal{H}_0$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}}$ with $\boldsymbol{\theta} \in \mathcal{H}_1$. Performance, however, is rather evaluated in terms of the *power* $\mathrm{E}_{\boldsymbol{\theta}}[\phi^{(n)}]$, $\boldsymbol{\theta} \in \mathcal{H}_1$, of $\phi^{(n)}$.

In such a context, it makes sense to state that "the performance of the sequence $\delta^{(n)}_2$, based on n_2 observations, equals, at $\boldsymbol{\theta}$, that of the sequence $\delta^{(n)}_1$ based on n_1 observations if $R^{\delta^{(n_2)}_2}(\boldsymbol{\theta}) = R^{\delta^{(n_1)}_1}(\boldsymbol{\theta})$, so that the idea of comparing procedures in terms of numbers of

Encyclopedia of Environmetrics, Online © 2006 John Wiley & Sons, Ltd.

This article was published in *Encyclopedia of Environmetrics Second Edition* in 2012 by John Wiley & Sons, Ltd.
DOI: 10.1002/9780470057339.vnn077

observations appears quite naturally. For any $n_1 \in \mathbb{N}$, denote by $n_2(\boldsymbol{\theta}; n_1)$ the number of observations it takes for the sequence of procedures $\delta_2^{(n)}$ to match the performance of the sequence $\delta_1^{(n)}$ based on n_1 observations, that is, such that $R^{\delta_2^{(n_2)}}(\boldsymbol{\theta}) = R^{\delta_1^{(n_1)}}(\boldsymbol{\theta})$. While that number $n_2(\boldsymbol{\theta}; n_1)$ carries substantial information on the respective performances of $\delta_2^{(n)}$ and $\delta_1^{(n)}$, it still depends on n_1 and $\boldsymbol{\theta}$; computing it, moreover, is difficult or impossible in practice. Therefore, consider

$$\lim_{n_1 \to \infty} n_2(\boldsymbol{\theta}; n_1)/n_1 := ARE_{\boldsymbol{\theta}}(\delta_1^{(n)}/\delta_2^{(n)}) \qquad (1)$$

That limit, if it exists, is called the *asymptotic relative efficiency (ARE) at $\boldsymbol{\theta}$ of $\delta_1^{(n)}$* with respect to $\delta_2^{(n)}$. If, for instance, $ARE_{\boldsymbol{\theta}}(\delta_1^{(n)}/\delta_2^{(n)}) = 2$, the sequence $\delta_2^{(n)}$, at $\boldsymbol{\theta}$, asymptotically requires twice as many observations as the sequence $\delta_1^{(n)}$ to achieve the same performance; in that sense, $\delta_1^{(n)}$ at $\boldsymbol{\theta}$ is twice as efficient as $\delta_2^{(n)}$.

Asymptotic Relative Efficiency in Point Estimation

The general definition (1) essentially applies to point estimation problems. Even so, its practical implementation is typically restricted to estimators $\delta_i^{(n)}$ of $\boldsymbol{\theta}$ (more general cases follow with obvious changes) such that, for some $\alpha > 0$,

$$n^{\alpha}(\delta_i^{(n)} - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \sigma_i \boldsymbol{\xi}, \quad i = 1, 2$$

where the random vector $\boldsymbol{\xi}$ has an absolutely continuous distribution (which may depend on $\boldsymbol{\theta}$) over $(\mathbb{R}^k, \mathcal{B}^k)$. Then, one can show that, irrespective of $\boldsymbol{\theta}$ and the loss function considered,

$$ARE_{\boldsymbol{\theta}}(\delta_1^{(n)}/\delta_2^{(n)}) = (\sigma_2/\sigma_1)^{1/\alpha}$$

In particular, if $\delta_1^{(n)}$ and $\delta_2^{(n)}$ are such that $\sqrt{n}(\delta_i^{(n)} - \boldsymbol{\theta})$ under $\mathrm{P}_{\boldsymbol{\theta}}^{(n)}$ is asymptotically normal, with mean $\mathbf{0}$ and covariance matrix $\sigma_i^2 \boldsymbol{\Sigma}(\boldsymbol{\theta})$, the ARE at $\boldsymbol{\theta}$ of $\delta_1^{(n)}$ with respect to $\delta_2^{(n)}$ is

$$ARE_{\boldsymbol{\theta}}(\delta_1^{(n)}/\delta_2^{(n)}) = \sigma_2^2/\sigma_1^2 \qquad (2)$$

and does not depend on $\boldsymbol{\theta}$.

Consider, for example, a sample $X_1, \ldots, X_n$ of iid univariate observations with density f symmetric with respect to $\mu \in \mathbb{R}$, and four root-n-consistent estimators of μ: the mean $\bar{X}^{(n)} := \frac{1}{n}\sum_{i=1}^{n} X_i$, the median $X_{1/2}^{(n)}$, the Hodges–Lehmann R-estimator $\hat{\mu}_{\mathrm{HL}}^{(n)}$ (minimizing Wilcoxon's one-sample signed rank test statistic), and the normal-score or van der Waerden R-estimator $\hat{\mu}_{\mathrm{vdW}}^{(n)}$ (minimizing the normal-score one-sample signed rank test statistic); see Ref. 1 for details. We have

$$n^{1/2}(\bar{X}^{(n)} - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_f^2) \qquad \text{provided that } \int (x-\mu)^2 f(x)\mathrm{d}x = \sigma_f^2 < \infty$$

$$n^{1/2}(X_{1/2}^{(n)} - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1/4f^2(\mu)) \qquad \text{provided that } f \text{ continuous at } \mu \text{ and } f(\mu) > 0$$

$$n^{1/2}(\hat{\mu}_{\mathrm{HL}}^{(n)} - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1/12[\textstyle\int f^2(z)\mathrm{d}z]^2) \qquad \text{provided that } \int f^2(z)\mathrm{d}z < \infty$$

$$n^{1/2}(\hat{\mu}_{\mathrm{vdW}}^{(n)} - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_f^2 \mathcal{I}_f/[\textstyle\int_0^1 \Phi^{-1}(u)\varphi_f(F^{-1}(u))\mathrm{d}u]^2) \qquad \text{provided that } \mathcal{I}_f := \int_0^1 \varphi_f^2(x) f(x)\mathrm{d}x < \infty$$

where f is assumed to be differentiable and $\varphi_f := -\dfrac{\mathrm{d}\log f(x)}{\mathrm{d}x}$; Φ as usual stands for the standard normal distribution function. Applying (2), we obtain the AREs in Table 1 (which, as we shall see, coincide with the AREs of the corresponding tests).

Table 1 Asymptotic relative efficiencies (AREs) of various estimators of location (the median $X_{1/2}^{(n)}$, the Hodges–Lehmann estimator $\hat{\mu}_{\mathrm{HL}}^{(n)}$ and the van der Waerden or normal-score estimator $\hat{\mu}_{\mathrm{vdW}}^{(n)}$) with respect to the mean $\bar{X}^{(n)}$, and the corresponding tests (the sign test, the Wilcoxon and van der Waerden or normal-score signed rank tests and their linear model extensions) with respect to the Student or Fisher tests, under normal, logistic, and double exponential densities, respectively.

Estimators (tests)	Normal	Logistic	Double exponential
Median (sign-test score or Laplace test)	$2/\pi \approx 0.637$	$\pi^2/12 \approx 0.822$	2.000
Hodges–Lehmann (Wilcoxon test)	$3/\pi \approx 0.955$	$\pi^2/9 \approx 1.098$	1.500
van der Waerden (normal-score test)	1.000	$\pi/3 \approx 1.049$	$4/\pi \approx 1.280$

Encyclopedia of Environmetrics, Online © 2006 John Wiley & Sons, Ltd.
This article is © 2013 John Wiley & Sons, Ltd.
This article was published in *Encyclopedia of Environmetrics Second Edition* in 2012 by John Wiley & Sons, Ltd.
DOI: 10.1002/9780470057339.vnn077

Asymptotic Relative Efficiency in Hypothesis Testing

The case of hypothesis testing is more delicate. While the choice of the performance criterion (power) here is clear, its evaluation also depends on the choice of a significance level α and an alternative. Fixed-$\boldsymbol{\theta}$ alternatives moreover are not always the most interesting ones as, for large n, all consistent tests basically have power close one. The most standard framework is associated with the name of Pitman and his concept of *local alternatives*; see Ref. 2 and the short historic comment at the end of this article. A sequence $\boldsymbol{\theta}^{(n)} \in \mathcal{H}_1$ is called *local* if $\boldsymbol{\theta}^{(n)} \to \boldsymbol{\theta}$ as $n \to \infty$ for some $\boldsymbol{\theta} \in \mathcal{H}_0$, in such a way that, for any sequence $A^{(n)}$ of events, $\lim_{n\to\infty} \mathrm{P}^{(n)}_{\boldsymbol{\theta}^{(n)}}(A^{(n)}) = 1$ implies $\lim_{n\to\infty} \mathrm{P}^{(n)}_{\boldsymbol{\theta}}(A^{(n)}) = 1$. A sequence of tests of $\mathcal{H}_0$ with asymptotic level $\alpha < 1$ thus cannot have asymptotic power one against the sequence of alternatives $\boldsymbol{\theta}^{(n)}$ (a property which is also known as *contiguity*; see, for instance, Chapter 6 in Ref. 3). Efficiency comparisons for two tests $\phi_1^{(n)}$ and $\phi_2^{(n)}$ with the same asymptotic significance level α then are based on their *local powers*, that is, on functions of the form $\mathrm{E}_{\boldsymbol{\theta}^{(n)}}[\phi_i^{n_i(n)}]$, where $n_i(n) \uparrow \infty$ as $n \to \infty$, $i = 1, 2$. Let the sequences $n_i(n)$ be chosen such that, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}^{(n)}}$, the powers of $\phi_2^{n_2(n)}$ and $\phi_1^{n_1(n)}$ coincide: $\mathrm{E}_{\boldsymbol{\theta}^{(n)}}[\phi_i^{n_2(n)}] = \mathrm{E}_{\boldsymbol{\theta}^{(n)}}[\phi_i^{n_1(n)}]$ (an equation that may have multiple solutions). The ARE at $\boldsymbol{\theta}$, in the sense of Pitman, of the sequence $\phi_1^{(n)}$ with respect to the sequence $\phi_2^{(n)}$ is then defined as

$$ARE_{\boldsymbol{\theta}}(\phi_1^{(n)}/\phi_2^{(n)}) := \lim_{n\to\infty} n_2(n)/n_1(n) \qquad (3)$$

provided that the limit exists and does not depend on the choice of $n_1(n)$ and $n_2(n)$.

Practical computation of (3) is, in general, impossible unless the test statistics involved exhibit comparable structures and behaviors. This, however, often happens to be the case. Letting $\boldsymbol{\theta}^{(n)} = \boldsymbol{\theta} + n^{-1/2}\boldsymbol{\tau}$, the tests $\phi_1^{(n)}$ and $\phi_2^{(n)}$ typically are based on K-dimensional statistics $\mathbf{S}_1^{(n)}(\boldsymbol{\theta})$ and $\mathbf{S}_2^{(n)}(\boldsymbol{\theta})$ such that, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}^{(n)}}$, $\mathbf{S}_i^{(n)}(\boldsymbol{\theta})$ is asymptotically multinormal, with mean $\nu_i \boldsymbol{\Gamma}(\boldsymbol{\theta})\boldsymbol{\tau}$ and full-rank covariance matrix $\sigma_i^2\boldsymbol{\Gamma}(\boldsymbol{\theta})$, $i = 1, 2$ ($\nu_i\gamma(\theta)\tau$ and $\sigma_i^2\gamma(\theta)\tau$ in the univariate case). The form of the tests depends on $\boldsymbol{\Theta}$ and $\mathcal{H}_0$:

1. if $\boldsymbol{\Theta}$ is one-dimensional ($K = 1$) and $\mathcal{H}_0$ is one-sided, of the form $\theta \le \theta_0$ versus $\theta > \theta_0$, $\phi_i^{(n)}$, $i = 1, 2$ rejects for large values of $S_i^{(n)}(\theta_0)/\sigma_i\gamma^{1/2}(\theta_0)$ which, under $\mathrm{P}^{(n)}_{\theta_0}$, is asymptotically standard normal, and, under $\mathrm{P}^{(n)}_{\theta^{(n)}}$ (with $\theta^{(n)} = \theta_0 + n^{-1/2}\tau$), asymptotically $\mathcal{N}((\nu_i/\sigma_i)\gamma^{1/2}(\theta_0)\tau, 1)$ (the squared shift $(\nu_i/\sigma_i)^2\gamma(\theta_0)\tau^2$ plays the role of a noncentrality parameter);
2. if $\mathcal{H}_0$ is of the form $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, $\phi_i^{(n)}$, $i = 1, 2$ rejects for large values of $\sigma_i^{-2}\mathbf{S}_i^{(n)\prime}(\boldsymbol{\theta}_0)\boldsymbol{\Gamma}^{-1}(\boldsymbol{\theta}_0)\mathbf{S}_i^{(n)}(\boldsymbol{\theta}_0)$ which, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_0}$, is asymptotically chi-square with K degrees of freedom and, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}^{(n)}}$ (with $\boldsymbol{\theta}^{(n)} = \boldsymbol{\theta}_0 + n^{-1/2}\boldsymbol{\tau}$), asymptotically noncentral chi-square, with noncentrality parameters $(\nu_i/\sigma_i)^2\boldsymbol{\tau}'\boldsymbol{\Gamma}(\boldsymbol{\theta}_0)\boldsymbol{\tau}$;
3. if $\mathcal{H}_0$ puts m independent linear constraints on $\boldsymbol{\theta}$, rejection for $\phi_i^{(n)}$, $i = 1, 2$ takes place for large values of quadratic forms $Q_i^{(n)}(\hat{\boldsymbol{\theta}})$, where

$$Q_i^{(n)}(\boldsymbol{\theta}) = \sigma_i^{-2}\mathbf{S}_i^{(n)\prime}(\boldsymbol{\theta})\boldsymbol{\Gamma}^{-1/2}(\boldsymbol{\theta})[\mathbf{I} - \mathbf{M}(\boldsymbol{\theta})] \boldsymbol{\Gamma}^{-1/2}(\boldsymbol{\theta})\mathbf{S}_i^{(n)}(\boldsymbol{\theta}) \qquad (4)$$

$[\mathbf{I} - \mathbf{M}(\boldsymbol{\theta})]$ is some projection matrix depending on the null hypothesis to be tested, and $\hat{\boldsymbol{\theta}}$ some appropriate estimator such that $Q_i^{(n)}(\hat{\boldsymbol{\theta}}) - Q_i^{(n)}(\hat{\boldsymbol{\theta}}) \xrightarrow{\mathrm{P}^{(n)}_{\boldsymbol{\theta}}} 0$ as $n \to \infty$. Those quadratic forms, under $\boldsymbol{\theta} \in \mathcal{H}_0$ are asymptotically chi-square with m degrees of freedom and, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}^{(n)}}$, asymptotically noncentral chi-square, with noncentrality parameters $(\nu_i/\sigma_i)^2\boldsymbol{\tau}'\boldsymbol{\Gamma}^{1/2}(\boldsymbol{\theta})[\mathbf{I} - \mathbf{M}(\boldsymbol{\theta})]\boldsymbol{\Gamma}^{1/2}(\boldsymbol{\theta})\boldsymbol{\tau}$.

In all three cases, it can be shown that the ARE of $\phi_1^{(n)}$ with respect to $\phi_2^{(n)}$ is the ratio of the corresponding squared shifts or noncentrality parameters; hence (with $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ for (1) and (2) and $\boldsymbol{\theta}$ such that $\mathrm{P}^{(n)}_{\boldsymbol{\theta}} \in \mathcal{H}_0$ for (3)),

$$ARE_{\boldsymbol{\theta}}(\phi_1^{(n)}/\phi_2^{(n)}) = (\nu_1/\sigma_1)^2/(\nu_2/\sigma_2)^2 \qquad (5)$$

a quantity which does not depend on $\boldsymbol{\theta}$, nor on the choice of $\boldsymbol{\theta}^{(n)}$ (that is, on $\boldsymbol{\tau}$).

As an illustration, consider the model under which $X_1, \ldots, X_n$ satisfy the linear regression equation $X_i = \mu + \sum_{j=1}^K c_{ij}^{(n)}\beta_j + \varepsilon_i$, $i = 1, \ldots, n$, where the ε_i's are iid with density f and median zero, say, and $\mathbf{c}_i^{(n)} := (c_{i1}^{(n)}, \ldots, c_{iK}^{(n)})'$ are regression constants. Assume that for all n, $\bar{c}_j^{(n)} := n^{-1}\sum_{i=1}^n c_{ij}^{(n)} = 0$, that $\lim_{n\to\infty} \max_{1\le i\le n}(c_{ij}^{(n)})^2/\sum_{i=1}^n(c_{ij}^{(n)})^2 = 0$,

Encyclopedia of Environmetrics, Online © 2006 John Wiley & Sons, Ltd.
This article is © 2013 John Wiley & Sons, Ltd.
This article was published in *Encyclopedia of Environmetrics Second Edition* in 2012 by John Wiley & Sons, Ltd.
DOI: 10.1002/9780470057339.vnn077

$j = 1, \ldots, K$ (the so-called *Noether condition* ; see **Ranks**), and that $\mathbf{R}^{(n)} := n^{-1}\sum_{i=1}^{n}\mathbf{c}_i^{(n)}\mathbf{c}_i^{(n)\prime}$ converges, as $n \to \infty$, to a positive definite matrix $\mathbf{R}$. Let $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)'$, and consider testing a set of $1 \le m \le K$ linear constraints on $\boldsymbol{\beta}$. The classical test for this is Fisher's F-test (see, e.g., Chapter 12 in Ref. 4), the asymptotic version of which is based on a quadratic form $Q_{\mathcal{N}}^{(n)}$ (the numerator of the Fisher statistic divided by the variance σ^2 of f) which does have the special structure (4). Denoted by $R_i^{(n)}$ the rank of the estimated residual $e_i^{(n)} := X_i - \hat{\mu} - \sum_{j=1}^{K} c_{ij}^{(n)}\hat{\beta}_j$ among $e_1^{(n)}, \ldots, e_n^{(n)}$. A **sign-test** score (or Laplace) test statistic $Q_{\mathrm{L}}^{(n)}$, a Wilcoxon test statistic $Q_{\mathrm{W}}^{(n)}$, or a van der Waerden (normal score) test statistic $Q_{\mathrm{vdW}}^{(n)}$ are obtained by replacing, in $Q_{\mathcal{N}}^{(n)}$, the residual $e_i^{(n)}$ with $\mathrm{sign}(R_i^{(n)} - (n+1)/2)$, $R_i^{(n)}$ or $\Phi^{-1}(R_i^{(n)}/(n+1))$, respectively, then dividing by some suitable scaling constant. Asymptotically, the resulting quadratic forms, under $\mathrm{P}_{\boldsymbol{\theta}^{(n)}}^{(n)}$, are noncentral chi-square, with m degrees of freedom and the noncentrality parameters

$$
\begin{array}{lll}
Q_{\mathcal{N}}^{(n)} : \sigma^{-2}\boldsymbol{\tau}'\mathbf{R}^{-1}[\mathbf{I}-\mathbf{M}]\mathbf{R}^{-1}\boldsymbol{\tau} & & (\int x f(x)\mathrm{d}x = \mu, \quad \int (x-\mu)^2 f(x)\mathrm{d}x = \sigma_f^2 < \infty) \\
Q_{L}^{(n)} : 4f^2(0)\boldsymbol{\tau}'\mathbf{R}^{-1}[\mathbf{I}-\mathbf{M}]\mathbf{R}^{-1}\boldsymbol{\tau} & & (f \text{ differentiable}, \ \int \varphi_f^2(x) f(x)\mathrm{d}x =: \mathcal{I}_f < \infty) \\
Q_{W}^{(n)} : 12[\int f^2(z)\mathrm{d}z]^2\boldsymbol{\tau}'\mathbf{R}^{-1}[\mathbf{I}-\mathbf{M}]\mathbf{R}^{-1}\boldsymbol{\tau} & & (f \text{ differentiable}, \ \int \varphi_f^2(x) f(x)\mathrm{d}x =: \mathcal{I}_f < \infty); \\
Q_{vdW}^{(n)} : \dfrac{\left[\displaystyle\int_0^1 \Phi^{-1}(u)\varphi_f(F^{-1}(u))\mathrm{d}u\right]^2}{\sigma_f^2\mathcal{I}}\boldsymbol{\tau}'\mathbf{R}^{-1}[\mathbf{I}-\mathbf{M}]\mathbf{R}^{-1}\boldsymbol{\tau} & & (f \text{ differentiable}, \ \int \varphi_f^2(x) f(x)\mathrm{d}x =: \mathcal{I}_f < \infty)
\end{array}
$$

where $\varphi_f(x) := -\dfrac{\mathrm{d}\log f(x)}{\mathrm{d}x}$ and Φ as usual stands for the standard normal distribution function (*see* **Ranks**, Refs 5 or 6 for details). Taking ratios of those noncentrality parameters yield asymptotic relative efficiencies (with respect to Fisher or, in case (1), Student) which coincide with those of their estimation counterparts displayed in Table 1.

Historical note

The concept of ARE described here was proposed by Pitman in the unpublished lecture notes [2] he prepared for a 1948–1949 course at Columbia University. The first published rigorous treatment of the subject was by Noether [7] in 1955. Other concepts of efficiency have been proposed by Bahadur [8] and by Hodges and Lehmann [1], to where we refer for details; they require asymptotic large deviation results under the null and the alternative, respectively, which generally implies nontrivial calculation. An in-depth presentation of those concepts can be found in Chapter 10 of Serfling [9], Chapter 14 of van der Vaart [3], or in the monograph by Nikitin [10].

References

[1] Hodges, J. & Lehmann, E.L. (1956). The efficiency of some nonparametric competitors of the t test, *Annals of Mathematical Statistics* **26**, 324–335.

[2] Pitman, E.J.G. (1949). *Notes on Nonparametric Statistical Inference*, Columbia University, mimeographed.

[3] van der Vaart, A.W. (1998). *Asymptotic Statistics*, Cambridge University Press, Cambridge.

[4] Rohatgi, V.K. (1998). *An Introduction to Probability and Mathematical Statistics*, John Wiley & Sons, Inc., New York.

[5] Hájek, I., Sidák, Z., & Sen, P.K. (1999). *Theory of Rank Tests*, 2nd Edition, Academic Press, New York.

[6] Puri, M.L. & Sen, P.K. (1985). *Nonparametric Methods in General Linear Models.*, John Wiley & Sons, Inc., New York.

[7] Noether, G.E. (1955). On a theorem of Pitman, *Annals of Mathematical Statistics* **26**, 64–68.

[8] Bahadur, R.R. (1960). Stochastic comparison of tests, *Annals of Mathematical Statistics* **31**, 276–295.

[9] Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, Inc., New York.

[10] Nikitin, Ya (1995). *Asymptotic Efficiency of Nonparametric Tests*, Cambridge University Press, Cambridge.

Marc Hallin

Encyclopedia of Environmetrics, Online © 2006 John Wiley & Sons, Ltd.

This article was published in *Encyclopedia of Environmetrics Second Edition* in 2012 by John Wiley & Sons, Ltd.
DOI: 10.1002/9780470057339.vnn077