Achieving Accuracy and Correctness in Parametric Frequentist Inference

G. Alastair Young

Department of Mathematics
Imperial College London

Tutorial Session, ERCIM 2013, London, December 2013

Acknowledgements: Tom DiCiccio (Cornell), Todd Kuffner (Washington University in St. Louis).

Overview

Concerned with procedures of likelihood-based frequentist parametric inference. Omnibus methodology, no explicit optimality requirements imposed.

Overview

Concerned with procedures of likelihood-based frequentist parametric inference. Omnibus methodology, no explicit optimality requirements imposed.

BUT, among key desiderata of such inference are high accuracy and inferential correctness:

- Low error (e.g. high levels of coverage accuracy of CIs), particularly with small sample sizes n;

Overview

Concerned with procedures of likelihood-based frequentist parametric inference. Omnibus methodology, no explicit optimality requirements imposed.

BUT, among key desiderata of such inference are high accuracy and inferential correctness:

- Low error (e.g. high levels of coverage accuracy of CIs), particularly with small sample sizes n;
- Inferential correctness, in relation to key principles of inference, especially those involving appropriate conditioning and parameterization invariance.

Two main routes to these goals

- Analytic procedures, based on distributional approximation of likelihood-based quantities using ‘small sample asymptotics’;

Two main routes to these goals

- Analytic procedures, based on distributional approximation of likelihood-based quantities using ‘small sample asymptotics’;
- Simulation-based (‘bootstrap’) methods.

Objectives of tutorial

- Advise on effective choice of inference procedure [approach and choice of statistic] against desiderata;

Objectives of tutorial

- Advise on effective choice of inference procedure [approach and choice of statistic] against desiderata;
- Discuss relationships between analytic and bootstrap methodologies;

Objectives of tutorial

- Advise on effective choice of inference procedure [approach and choice of statistic] against desiderata;
- Discuss relationships between analytic and bootstrap methodologies;
- Special consideration to high-dimensional nuisance parameter problems;

Objectives of tutorial

- Advise on effective choice of inference procedure [approach and choice of statistic] against desiderata;
- Discuss relationships between analytic and bootstrap methodologies;
- Special consideration to high-dimensional nuisance parameter problems;
- Discuss computational issues [computational intensiveness versus analytic requirements].

Structure of Tutorial

- Background, key ideas.
- Detailed theoretical analysis.
- Further illustrations.

I: Background, key ideas

The inferential problem

Let $Y = \{Y_1, \ldots, Y_n\}$ be random sample from underlying distribution $F(y; \theta)$, indexed by d-dimensional parameter $\theta = (\theta^1, \ldots, \theta^d) = (\psi, \phi)$, ψ p-dimensional interest parameter, ϕ q-dimensional nuisance parameter, $p + q = d$. May have ϕ high-dimensional.

Wish to test $H_0 : \psi = \psi_0$, or (duality) construct confidence set for ψ.

The inferential problem

Let $Y = \{Y_1, \ldots, Y_n\}$ be random sample from underlying distribution $F(y; \theta)$, indexed by d-dimensional parameter $\theta = (\theta^1, \ldots, \theta^d) = (\psi, \phi)$, ψ p-dimensional interest parameter, ϕ q-dimensional nuisance parameter, $p + q = d$. May have ϕ high-dimensional.

Wish to test $H_0 : \psi = \psi_0$, or (duality) construct confidence set for ψ.

If $p = 1$, $\psi = \theta^1$, want one-sided inference e.g. test H_0 against (one-sided) alternative $\psi > \psi_0$ or $\psi < \psi_0$, or one-sided confidence limit.

“Break the research question of interest into simple components corresponding to strongly focused and incisive research questions.”

(D.R. Cox, ‘Principles of Statistical Inference’)

“Break the research question of interest into simple components corresponding to strongly focused and incisive research questions.”

(D.R. Cox, ‘Principles of Statistical Inference’)

Typically, $p = 1$.

Inference

Let $L(\theta) \equiv L(\theta; Y)$ be log-likelihood, $\hat{\theta} = (\hat{\psi}, \hat{\phi})$ the overall MLE of θ, $\hat{\phi}_\psi$ the constrained MLE of ϕ, for fixed value of ψ. Write $\tilde{\theta} \equiv \tilde{\theta}(\psi) = (\psi, \hat{\phi}_\psi)$.

Inference

Let $L(\theta) \equiv L(\theta; Y)$ be log-likelihood, $\hat{\theta} = (\hat{\psi}, \hat{\phi})$ the overall MLE of θ, $\hat{\phi}_\psi$ the constrained MLE of ϕ, for fixed value of ψ. Write $\tilde{\theta} \equiv \tilde{\theta}(\psi) = (\psi, \hat{\phi}_\psi)$.

Profile log-likelihood function for ψ is $M(\psi) = L\{\tilde{\theta}(\psi)\}$.

Inference

Let $L(\theta) \equiv L(\theta; Y)$ be log-likelihood, $\hat{\theta} = (\hat{\psi}, \hat{\phi})$ the overall MLE of θ, $\hat{\phi}_\psi$ the constrained MLE of ϕ, for fixed value of ψ. Write $\tilde{\theta} \equiv \tilde{\theta}(\psi) = (\psi, \hat{\phi}_\psi)$.

Profile log-likelihood function for ψ is $M(\psi) = L\{\tilde{\theta}(\psi)\}$.

Likelihood ratio statistic is $W(\psi) = 2\{M(\hat{\psi}) - M(\psi)\}$.

Inference

Let $L(\theta) \equiv L(\theta; Y)$ be log-likelihood, $\hat{\theta} = (\hat{\psi}, \hat{\phi})$ the overall MLE of θ, $\hat{\phi}_\psi$ the constrained MLE of ϕ, for fixed value of ψ. Write $\tilde{\theta} \equiv \tilde{\theta}(\psi) = (\psi, \hat{\phi}_\psi)$.

Profile log-likelihood function for ψ is $M(\psi) = L\{\tilde{\theta}(\psi)\}$.

Likelihood ratio statistic is $W(\psi) = 2\{M(\hat{\psi}) - M(\psi)\}$.

In case of scalar ψ, use signed root likelihood ratio statistic:

$$R(\psi) = \mathrm{sgn}(\hat{\psi} - \psi) W(\psi)^{1/2}.$$

Notation

Arrays and summation are denoted by using the standard conventions, for which the indices $r, s, t, \ldots$ are assumed to range over $1, \ldots, d$. Summation over the range is implied for any index appearing in an expression both as a subscript and as a superscript.

Differentiation is indicated by subscripts, so $L_r(\theta) = \partial L(\theta)/\partial\theta^r$, $L_{rs}(\theta) = \partial^2 L(\theta)/\partial\theta^r\partial\theta^s$, etc. Then $E\{L_r(\theta)\} = 0$; let $\lambda_{rs} = E\{L_{rs}(\theta)\}$, $\lambda_{rst} = E\{L_{rst}(\theta)\}$, etc.

The constants λ_{rs}, $\lambda_{rst}, \ldots$, are assumed to be of order $O(n)$. These assumptions are usually satisfied in situations involving independent observations, structured (e.g. time series) dependent data problems.

Let $\lambda_{r,s} = E(L_r L_s), \lambda_{rs,t} = E(L_{rs} L_t)$, etc.

Let (λ^{rs}) be the $d \times d$ matrix inverse of (λ_{rs}), and let $\eta = -1/\lambda^{11}$, $\tau^{rs} = \eta \lambda^{1r} \lambda^{1s}$, and $\nu^{rs} = \lambda^{rs} + \tau^{rs}$. Thus, λ^{rs}, τ^{rs}, and ν^{rs} are of order $O(n^{-1})$, while η is of order $O(n)$.

A comment

Calculation of quantities just defined requires (at most) evaluation of expectations of log-likelihood derivatives.

Other statistics

Consider, for simplicity, scalar case $p = 1$. Variants for $p > 1$ easily defined.

Other statistics

Consider, for simplicity, scalar case $p = 1$. Variants for $p > 1$ easily defined.

As alternative 'pivots' to $R(\psi)$, could use, for example:

Other statistics

Consider, for simplicity, scalar case $p = 1$. Variants for $p > 1$ easily defined.

As alternative 'pivots' to $R(\psi)$, could use, for example:

Wald statistic,

$$T_W(\psi) = (\hat{\psi} - \psi)\{-\lambda^{11}(\hat{\theta})\}^{-1/2}.$$

Other statistics

Consider, for simplicity, scalar case $p = 1$. Variants for $p > 1$ easily defined.

As alternative 'pivots' to $R(\psi)$, could use, for example:

Wald statistic,

$$T_W(\psi) = (\hat{\psi} - \psi)\{-\lambda^{11}(\hat{\theta})\}^{-1/2}.$$

Score statistic,

$$T_S(\psi) = L_1\{\tilde{\theta}(\psi)\}\{\lambda^{11}(\hat{\theta})\}^{1/2}.$$

Constructed using expected (inverse) information matrix $[\lambda^{rs}]$, evaluated at global MLE. Alternatively: use observed (inverse) information matrix $[L^{rs}]$; evaluate at constrained MLE $\tilde{\theta}(\psi)$,.......

Running Example (RE): Inverse Gaussian distribution

$Y_1, \dots, Y_n$ IID inverse Gaussian, $IG(\mu, \psi)$, with density

$$f(y; \mu, \psi) = \left(\frac{\psi}{2\pi y^3} \right)^{1/2} \exp\left(-\frac{\psi}{2\mu^2 y} (y - \mu)^2 \right), \; y > 0,$$

interest parameter is shape $\psi > 0$, mean $\mu > 0$ as nuisance.

Running Example (RE): Inverse Gaussian distribution

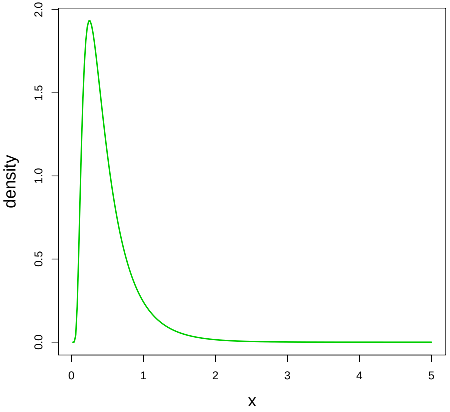
$Y_1, \ldots, Y_n$ IID inverse Gaussian, $IG(\mu, \psi)$, with density

$$f(y; \mu, \psi) = \left(\frac{\psi}{2\pi y^3} \right)^{1/2} \exp\left(-\frac{\psi}{2\mu^2 y} (y - \mu)^2 \right), \; y > 0,$$

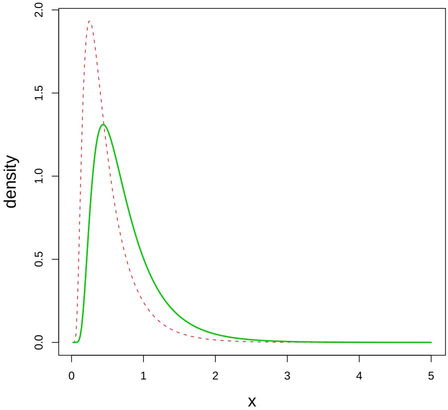
interest parameter is shape $\psi > 0$, mean $\mu > 0$ as nuisance.

First passage time of Brownian motion, widely used to model phenomena in biosciences/reliability/survival/....

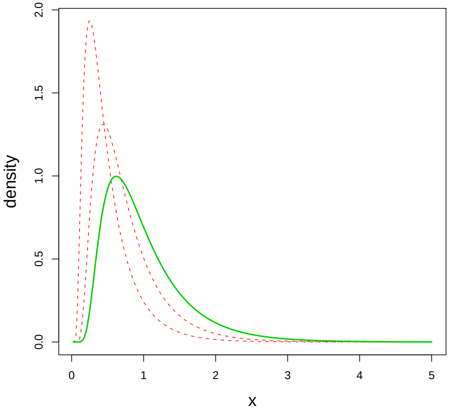
Density: $\psi = 1, \mu = 0.5$



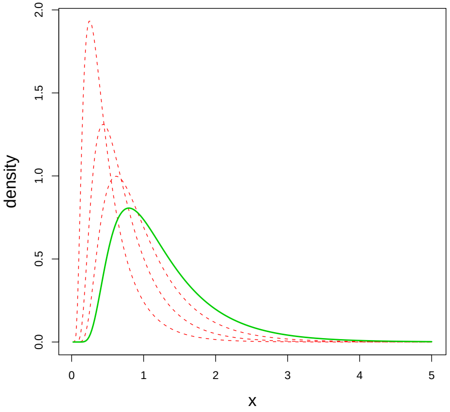
Density: $\psi = 2, \mu = 0.75$



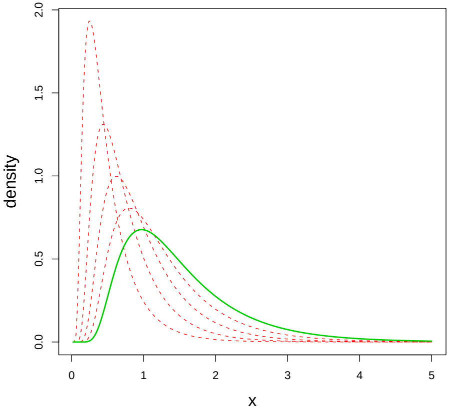
Density: $\psi = 3, \mu = 1.0$



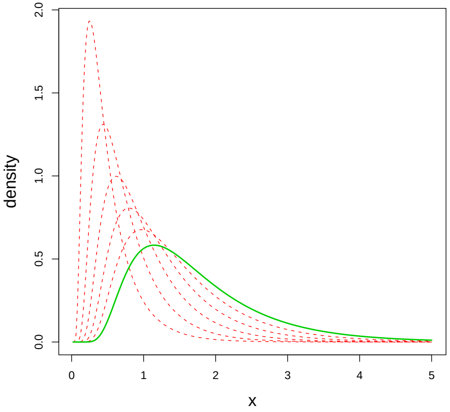
Density: $\psi = 4, \mu = 1.25$



Density: $\psi = 5, \mu = 1.5$



Density: $\psi = 6, \mu = 1.75$



MLES are:

$$\hat{\psi} = n/V,\ \hat{\mu} = \hat{\mu}_\psi = \bar{Y},$$

$$V = \sum_{i=1}^{n}(Y_i^{-1} - \overline{Y}^{-1}), \bar{Y} = n^{-1}\sum_{i=1}^{n} Y_i.$$

MLES are:

$$\hat{\psi} = n/V, \ \hat{\mu} = \hat{\mu}_\psi = \bar{Y},$$

$$V = \sum_{i=1}^{n} (Y_i^{-1} - \overline{Y}^{-1}), \bar{Y} = n^{-1} \sum_{i=1}^{n} Y_i.$$

Distribution of ψV is χ^2_{n-1}, distribution of $\hat{\mu}$ is $IG(\mu, \psi)$.

$$
\begin{aligned}
R(\psi) &= \operatorname{sgn}(\hat{\psi} - \psi)\{n(\log \hat{\psi} - 1 - \log \psi + \psi/\hat{\psi})\}^{1/2}, \\
T_W(\psi) &= \sqrt{\frac{n}{2}}\left(1 - \frac{\psi}{\hat{\psi}}\right), \\
T_S(\psi) &= \sqrt{\frac{n}{2}}\left(\frac{\hat{\psi}}{\psi} - 1\right)
\end{aligned}
$$

A data sample

Data sample size $n = 10$, generated with $\mu = 1, \psi = 2$:

```
0.435, 0.466, 1.624, 0.304, 2.165

0.936, 0.620, 0.595, 0.351, 1.688
```

A data sample

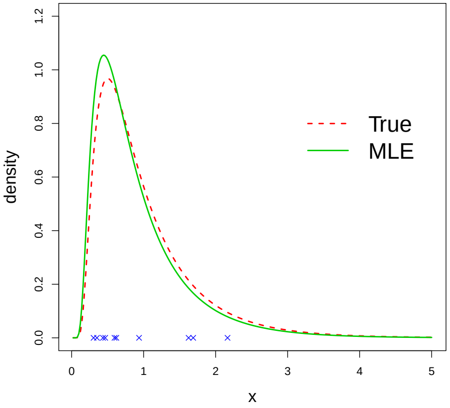
Data sample size $n = 10$, generated with $\mu = 1, \psi = 2$:

```
0.435, 0.466, 1.624, 0.304, 2.165

0.936, 0.620, 0.595, 0.351, 1.688
```

Have: $\hat{\psi} = 1.745, \hat{\mu} = 0.918$.

RE: True and estimated densities



Comment

Concentrate here on inference based on R, W, for simplicity. Most results true also for Wald and score statistics.

Parameterization invariance

Principle of parameterization invariance (PPI) important basis for choosing between different inferential procedures.

Parameterization invariance

Principle of parameterization invariance (PPI) important basis for choosing between different inferential procedures.

If θ and ζ are two alternative parameterizations and $\mathcal{P}(\cdot)$ is an inference procedure, with C_θ and C_ζ the conclusions that $\mathcal{P}(\cdot)$ leads to, expressed in the two parameterizations, then the same conclusion C_ζ should be reached by both application of $\mathcal{P}(\cdot)$ in the ζ parameterization and translation into the ζ parameterization of the conclusion C_θ.

Nuisance parameter

With nuisance parameters, parameterization invariance is restricted to mean invariance under interest respecting reparameterization.

Nuisance parameter

With nuisance parameters, parameterization invariance is restricted to mean invariance under interest respecting reparameterization.

Suppose $\theta = (\psi, \phi)$, with ψ interest parameter and ϕ nuisance parameter. An interest respecting reparameterization is of the form $\upsilon = \upsilon(\theta) = \upsilon(\psi, \phi)$ with $\upsilon = (\varphi, \chi)$, such that

$$\varphi = \varphi(\psi),\ \chi = \chi(\psi, \phi).$$

Implications of PPI

Inference based on $W(\psi)$ (or $R(\psi)$) does respect PPI.

Implications of PPI

Inference based on $W(\psi)$ (or $R(\psi)$) does respect PPI.

So does inference based on $T_S(\psi)$.

Implications of PPI

Inference based on $W(\psi)$ (or $R(\psi)$) does respect PPI.

So does inference based on $T_S(\psi)$.

Inference based on $T_W(\psi)$ does not.

Adjusted likelihood

Broadly, properties to be discussed hold also for versions of statistics based on adjusted forms of profile likelihood.

Adjusted likelihood

Broadly, properties to be discussed hold also for versions of statistics based on adjusted forms of profile likelihood.

Replace $M(\psi)$ by $\bar{M}(\psi) = M(\psi) + B(\psi)$, where (various proposals) adjustment function $B(\psi)$ introduced to take account of nuisance parameter ϕ.

Adjusted likelihood

Broadly, properties to be discussed hold also for versions of statistics based on adjusted forms of profile likelihood.

Replace $M(\psi)$ by $\bar{M}(\psi) = M(\psi) + B(\psi)$, where (various proposals) adjustment function $B(\psi)$ introduced to take account of nuisance parameter ϕ.

Intractable likelihood? Composite/pseudo-likelihood. Analysis of inference for these incomplete, predictable.

First-order theory

First-order theory

Have $W(\psi)$ distributed as χ^2_p, to error of order $O(n^{-1})$.

First-order theory

Have $W(\psi)$ distributed as χ^2_p, to error of order $O(n^{-1})$.

Also, $R(\psi)$ distributed as $N(0,1)$, to error of order $O(n^{-1/2})$.

First-order theory

Have $W(\psi)$ distributed as χ^2_p, to error of order $O(n^{-1})$.

Also, $R(\psi)$ distributed as $N(0,1)$, to error of order $O(n^{-1/2})$.

Latter true also for $T_W(\psi)$ and $T_S(\psi)$, and variants.

Inference: illustration, $p = 1$

A confidence set of asymptotic coverage $1 - \alpha$ for ψ is

$$\mathcal{I}(Y) \equiv \mathcal{I}_{1-\alpha}(Y) = \{\psi : u(Y, \psi) \leq 1 - \alpha\},$$

with $u(Y, \psi) = \Phi\{R(\psi)\}$, in terms of the $N(0, 1)$ distribution function $\Phi(\cdot)$. Call $u(Y, \psi)$ the 'significance function'.

Equivalently, the confidence set is

$$\mathcal{I}(Y) = \{\psi : R(\psi) \leq \Phi^{-1}(1 - \alpha)\}.$$

The coverage error of the confidence set is $O(n^{-1/2})$: first-order accuracy.

Have that $u(Y, \psi)$ is monotonic in ψ, so confidence set is semi-infinite interval of form $(\hat{\psi}_l(Y), \infty)$. Lower confidence limit.

Have that $u(Y, \psi)$ is monotonic in ψ, so confidence set is semi-infinite interval of form $(\hat{\psi}_l(Y), \infty)$. Lower confidence limit.

If two-sided inference is required, an equi-tailed two-sided confidence interval $\mathcal{J}(Y)$ of nominal coverage $1 - \alpha$ may be obtained by taking the set difference of two one-sided sets:

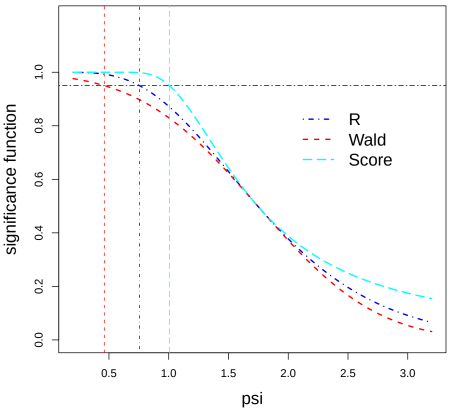
$$\mathcal{J}(Y) \equiv \mathcal{J}_{1-\alpha}(Y) = \mathcal{I}_{1-\alpha/2}(Y) \backslash \mathcal{I}_{\alpha/2}(Y).$$

Similar statements about coverage error of confidence sets true for other asymptotically $N(0, 1)$ pivots.

Similar statements about coverage error of confidence sets true for other asymptotically $N(0,1)$ pivots.

In case $p > 1$, confidence set of coverage error $O(n^{-1})$ (second-order accuracy) from χ^2_p approximation to sampling distribution of $W(\psi)$.

RE, data sample: significance functions



RE, data example: 95% confidence limits

- $R(\psi)$: interval is $(0.755, \infty)$.
- $T_W(\psi)$: interval is $(0.461, \infty)$.
- $T_S(\psi)$: interval is $(1.005, \infty)$.

Motivations for refinements

Motivations for refinements

- To obtain higher-order repeated sampling accuracy.

Motivations for refinements

- To obtain higher-order repeated sampling accuracy.

- To accommodate appropriate conditioning: multi-parameter exponential families (conditioning dictated by theory of optimal tests etc.); ancillary statistic models (relevance, by conditioning on component of minimal sufficient statistic that is approximately distribution constant).

Exponential family context

Suppose that the log-likelihood is of the form

$$L(\theta) = \psi s_1(Y) + \phi s_2(Y) - k(\psi, \phi) - d(Y),$$

so that ψ is a natural parameter of a multi-parameter exponential family.

Exponential family context

Suppose that the log-likelihood is of the form

$$L(\theta) = \psi s_1(Y) + \phi s_2(Y) - k(\psi, \phi) - d(Y),$$

so that ψ is a natural parameter of a multi-parameter exponential family.

The conditional distribution of s_1 given s_2 depends only on ψ: conditioning on s_2 eliminates the nuisance parameter.

Exponential family context

Suppose that the log-likelihood is of the form

$$L(\theta) = \psi s_1(Y) + \phi s_2(Y) - k(\psi, \phi) - d(Y),$$

so that ψ is a natural parameter of a multi-parameter exponential family.

The conditional distribution of s_1 given s_2 depends only on ψ: conditioning on s_2 eliminates the nuisance parameter.

Appropriate inference on ψ is based on the distribution of s_1, given the observed value of s_2. This is, in principle, known, since it is completely specified, once ψ fixed.

In fact, this conditional inference has unconditional (repeated sampling) optimality properties of being uniformly most powerful unbiased etc etc.

In fact, this conditional inference has unconditional (repeated sampling) optimality properties of being uniformly most powerful unbiased etc etc.

In practice, the exact inference may be difficult to construct: the relevant conditional distribution typically requires awkward analytic calculations, numerical integrations etc.

Ancillary statistic context

Fisherian proposition: inference about ψ should be based not on the original specified model $F(y;\theta)$, but instead on derived model obtained by conditioning on an ancillary statistic, when this exists.

Ancillary statistic context

Fisherian proposition: inference about ψ should be based not on the original specified model $F(y; \theta)$, but instead on derived model obtained by conditioning on an ancillary statistic, when this exists.

Suppose minimal sufficient statistic for θ can be written as

$$(\hat{\theta}, A),$$

with A (approximately) distribution constant.

Then, A is ancillary, and the Conditionality Principle (CP) dictates that to be relevant inference on ψ should be made conditional on the observed value a of A. CP automatically respected by Bayesian inference.

Refinements: approaches

Two most established approaches:

Refinements: approaches

Two most established approaches:

- Analytic procedures, 'small sample asymptotics', saddlepoint, related methods;

Refinements: approaches

Two most established approaches:

- Analytic procedures, 'small sample asymptotics', saddlepoint, related methods;
- Simulation ('bootstrap') methods.

The third way: objective Bayes

Bayes with prior explicitly specified so (marginal) posterior for ψ yields confidence limits with correct frequentist interpretation, to high-order: 'probability matching prior'.

The third way: objective Bayes

Bayes with prior explicitly specified so (marginal) posterior for ψ yields confidence limits with correct frequentist interpretation, to high-order: 'probability matching prior'.

- conceptually simple;

The third way: objective Bayes

Bayes with prior explicitly specified so (marginal) posterior for ψ yields confidence limits with correct frequentist interpretation, to high-order: 'probability matching prior'.

- conceptually simple;
- typically awkward with high-dimensional nuisance parameter, as need to find marginal posterior of ψ;

The third way: objective Bayes

Bayes with prior explicitly specified so (marginal) posterior for ψ yields confidence limits with correct frequentist interpretation, to high-order: 'probability matching prior'.

- conceptually simple;
- typically awkward with high-dimensional nuisance parameter, as need to find marginal posterior of ψ;
- route not always open, higher-order (conditional) accuracy not necessarily obtainable.

Detail

Require prior $\pi(\psi, \phi)$ so that

$$Pr_\theta\{\psi \leq \psi^{(1-\alpha)}(\pi, Y)\} = 1 - \alpha + O(n^{-r/2}),$$

for $r = 2$ or 3, each $0 < \alpha < 1$.

Detail

Require prior $\pi(\psi, \phi)$ so that

$$Pr_\theta\{\psi \leq \psi^{(1-\alpha)}(\pi, Y)\} = 1 - \alpha + O(n^{-r/2}),$$

for $r = 2$ or 3, each $0 < \alpha < 1$.

Here:

- n is sample size;

Detail

Require prior $\pi(\psi, \phi)$ so that

$$Pr_\theta\{\psi \leq \psi^{(1-\alpha)}(\pi, Y)\} = 1 - \alpha + O(n^{-r/2}),$$

for $r = 2$ or 3, each $0 < \alpha < 1$.

Here:

- n is sample size;
- $\psi^{(1-\alpha)}(\pi, Y)$ is $(1-\alpha)$ quantile of marginal posterior, given data Y, of ψ, under prior $\pi(\psi, \phi)$;

Detail

Require prior $\pi(\psi, \phi)$ so that

$$Pr_\theta\{\psi \leq \psi^{(1-\alpha)}(\pi, Y)\} = 1 - \alpha + O(n^{-r/2}),$$

for $r = 2$ or 3, each $0 < \alpha < 1$.

Here:

- n is sample size;
- $\psi^{(1-\alpha)}(\pi, Y)$ is $(1 - \alpha)$ quantile of marginal posterior, given data Y, of ψ, under prior $\pi(\psi, \phi)$;
- Pr_θ denotes frequentist probability, under repeated sampling of Y, under parameter θ.

Probability matching priors

If condition holds with $r = 2$, speak of $\pi(\psi, \phi)$ as first-order probability matching prior.

Probability matching priors

If condition holds with $r = 2$, speak of $\pi(\psi, \phi)$ as first-order probability matching prior.

If condition holds with $r = 3$, speak of $\pi(\psi, \phi)$ as second-order probability matching prior.

Conditional probability matching

Appropriate frequentist inference to match in full exponential family or ancillary statistic context is the conditional one.

Conditional probability matching

Appropriate frequentist inference to match in full exponential family or ancillary statistic context is the conditional one.

The requirement should be 'conditional probability matching':

$$Pr_\theta\{\psi \leq \psi^{(1-\alpha)}(\pi, Y) \mid C(Y) = c\} = 1 - \alpha + O(n^{-r/2}).$$

Conditional probability matching

Appropriate frequentist inference to match in full exponential family or ancillary statistic context is the conditional one.

The requirement should be 'conditional probability matching':

$$Pr_\theta\{\psi \le \psi^{(1-\alpha)}(\pi, Y) \mid C(Y) = c\} = 1 - \alpha + O(n^{-r/2}).$$

Want the posterior $1-\alpha$ quantile to match the $1-\alpha$ conditional frequentist confidence limit for ψ.

Analytic methods: the highlights

Analytic methods: the highlights

- Bartlett correction of likelihood ratio statistic $W(\psi)$.

Analytic methods: the highlights

- Bartlett correction of likelihood ratio statistic $W(\psi)$.

- Analytically modified forms of $R(\psi)$, specifically designed to offer conditional validity, to high (asymptotic) order, in both contexts. 'Barndorff-Nielsen's R^*'.

Bartlett correction

Have

$$E_\theta\{W(\psi)\} = p\left(1 + \frac{b(\theta)}{n} + O(n^{-2})\right),$$

so modify $W(\psi)$ to

$$W_c(\psi) = W(\psi)/\{1 + b(\psi, \hat{\phi}_\psi)/n\},$$

or

$$\bar{W}_c(\psi) = W(\psi)/E_{(\psi,\hat{\phi}_\psi)}\{W(\psi)\}.$$

Bartlett correction

Have

$$E_\theta\{W(\psi)\} = p\left(1 + \frac{b(\theta)}{n} + O(n^{-2})\right),$$

so modify $W(\psi)$ to

$$W_c(\psi) = W(\psi)/\{1 + b(\psi, \hat{\phi}_\psi)/n\},$$

or

$$\bar{W}_c(\psi) = W(\psi)/E_{(\psi,\hat{\phi}_\psi)}\{W(\psi)\}.$$

Then $W_c(\psi)$ and $\bar{W}_c(\psi)$ are distributed as χ^2_p, to error of order $O(n^{-2})$. Confidence sets constructed by χ^2_p approximation have coverage error $O(n^{-2})$.

$E_{(\psi,\hat{\phi}_\psi)}\{W(\psi)\}$ constructed by (bootstrap) simulation. Estimation of expectation requires smaller MC simulation that estimation of whole sampling distribution.

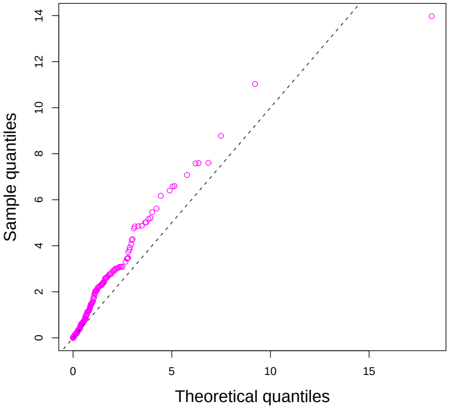
Inference by χ^2_p approximation to distribution of $\bar{W}_c(\psi)$: 'Empirical Bartlett correction'.

$E_{(\psi,\hat{\phi}_\psi)}\{W(\psi)\}$ constructed by (bootstrap) simulation. Estimation of expectation requires smaller MC simulation that estimation of whole sampling distribution.

Inference by χ^2_p approximation to distribution of $\bar{W}_c(\psi)$: 'Empirical Bartlett correction'.

Could replace χ^2_p approximation to sampling distribution of $W(\psi)$ by bootstrap distribution: sampling distribution under sampling with parameter fixed as $\theta = (\psi, \hat{\phi}_\psi)$. Confidence set will also have coverage error $O(n^{-2})$.

RE: $n = 5, \psi = 2, \mu = 1.0$, χ^2_1 QQ plot, $W(\psi)$



RE: $n = 5, \psi = 2, \mu = 1.0$

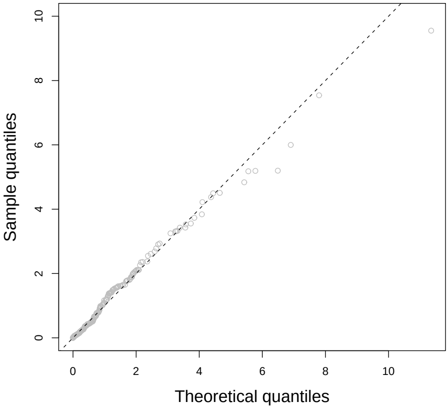
In inverse Gaussian example, $E_\theta\{W(\psi)\}$ does not depend on nuisance parameter μ.

RE: $n = 5, \psi = 2, \mu = 1.0$

In inverse Gaussian example, $E_\theta\{W(\psi)\}$ does not depend on nuisance parameter μ.

Big simulation shows, $E_\theta\{W(\psi)\} = 1.4632$.

RE: $n = 5, \psi = 2, \mu = 1.0$, χ^2_1 QQ plot, $\bar{W}_c(\psi)$



Adjusted signed root statistic R^*

Defined by

$$R^*(\psi) = R(\psi) + \log\{v(\psi)/R(\psi)\}/R(\psi)$$

Adjusted signed root statistic R^*

Defined by

$$R^*(\psi) = R(\psi) + \log\{v(\psi)/R(\psi)\}/R(\psi)$$

Here, in formulation considered, adjustment $v(\psi)$ necessitates:

Adjusted signed root statistic R^*

Defined by

$$R^*(\psi) = R(\psi) + \log\{v(\psi)/R(\psi)\}/R(\psi)$$

Here, in formulation considered, adjustment $v(\psi)$ necessitates:

- explicit specification of ancillary A in ancillary statistic (e.g. transformation) context;

Adjusted signed root statistic R^*

Defined by

$$R^*(\psi) = R(\psi) + \log\{v(\psi)/R(\psi)\}/R(\psi)$$

Here, in formulation considered, adjustment $v(\psi)$ necessitates:

- explicit specification of ancillary A in ancillary statistic (e.g. transformation) context;
- potentially awkward analytic calculations, in both ancillary/exponential family situations.

Other formulations

Other formulations of $v(\psi)$, due to Fraser and co-workers, possible: use of 'tangent exponential model' avoids need to specify transformation $Y \to (\hat{\theta}, A)$.

Other formulations

Other formulations of $v(\psi)$, due to Fraser and co-workers, possible: use of 'tangent exponential model' avoids need to specify transformation $Y \to (\hat{\theta}, A)$.

Still analytically fiddly.

RE: adjustment function

In inverse Gaussian example,

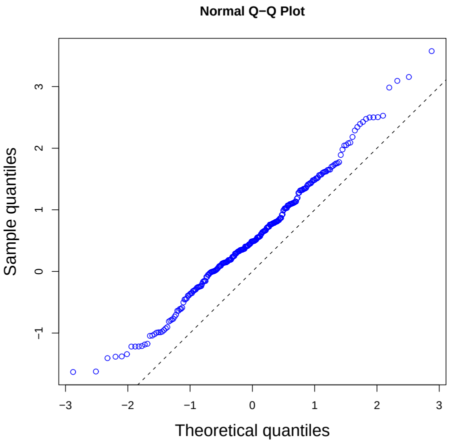
$$v(\psi) = \sqrt{\frac{n\psi}{2\hat{\psi}}}\left(1 - \frac{\psi}{\hat{\psi}}\right).$$

Sampling distribution of $R^*(\psi)$ is $N(0,1)$, to error of order $O(n^{-3/2})$, conditional on ancillary, hence unconditionally. Normal approximation to distribution of $R^*(\psi)$ yields third-order (relative) conditional accuracy in ancillary statistic setting, and confidence sets with third-order repeated sampling coverage accuracy.

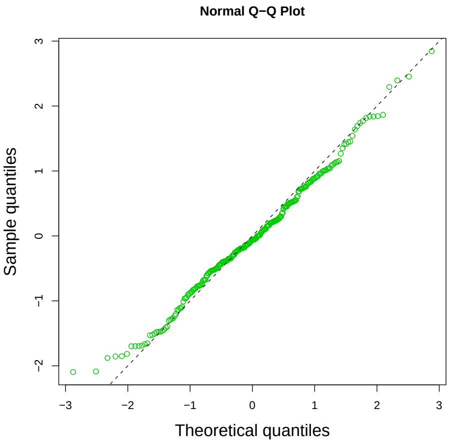
Sampling distribution of $R^*(\psi)$ is $N(0,1)$, to error of order $O(n^{-3/2})$, conditional on ancillary, hence unconditionally. Normal approximation to distribution of $R^*(\psi)$ yields third-order (relative) conditional accuracy in ancillary statistic setting, and confidence sets with third-order repeated sampling coverage accuracy.

Inference which respects that of exact conditional inference in exponential family setting to same third-order.

RE: $n = 5, \psi = 2, \mu = 1.0$, QQ plot, $R(\psi)$



RE: $n = 5, \psi = 2, \mu = 1.0$, QQ plot, $R^*(\psi)$



Some comments on analytic methods

Some comments on analytic methods

- Often very awkward analytic calculations.

Some comments on analytic methods

- Often very awkward analytic calculations.
- Successfully packaged (Davison et al.) for certain classes of model, e.g. nonlinear heteroscedastic regression models.

Some comments on analytic methods

- Often very awkward analytic calculations.
- Successfully packaged (Davison et al.) for certain classes of model, e.g. nonlinear heteroscedastic regression models.
- Also, relatively unexplored is idea of using simulation to replace analytic calculations, specifically to calculate Bartlett correction.

Some comments on analytic methods

- Often very awkward analytic calculations.
- Successfully packaged (Davison et al.) for certain classes of model, e.g. nonlinear heteroscedastic regression models.
- Also, relatively unexplored is idea of using simulation to replace analytic calculations, specifically to calculate Bartlett correction.
- Versions of R^* for vector interest parameters possible, seen as less effective than in case $p = 1$, or than Bartlett correction.

(Constrained) Bootstrap

Bootstrap Principle: estimate sampling distribution of interest by that under a fitted model.

(Constrained) Bootstrap

Bootstrap Principle: estimate sampling distribution of interest by that under a fitted model.

Key: appropriate handling of nuisance parameter. Repeated sampling properties of bootstrap are [modulo Monte Carlo error from using finite simulation] entirely determined by nuisance parameter effects.

The key recommendation

Use as basis of bootstrap calculation $F(y; (\psi, \hat{\phi}_\psi))$, fitted model with nuisance parameter taken as constrained MLE, for given value of interest parameter.

Properties: repeated sampling perspective

Properties: repeated sampling perspective

- 'Essentially exact'.

Properties: repeated sampling perspective

- 'Essentially exact'.
- Estimate true sampling distribution of $W(\psi)$ to error of order $O(n^{-2})$. Confidence sets constructed from bootstrap distribution of $W(\psi)$ have coverage error of order $O(n^{-2})$.

Properties: repeated sampling perspective

- 'Essentially exact'.
- Estimate true sampling distribution of $W(\psi)$ to error of order $O(n^{-2})$. Confidence sets constructed from bootstrap distribution of $W(\psi)$ have coverage error of order $O(n^{-2})$.
- Estimate true sampling distribution of $R(\psi)$ to error of order $O(n^{-1})$.

Properties: repeated sampling perspective

- 'Essentially exact'.
- Estimate true sampling distribution of $W(\psi)$ to error of order $O(n^{-2})$. Confidence sets constructed from bootstrap distribution of $W(\psi)$ have coverage error of order $O(n^{-2})$.
- Estimate true sampling distribution of $R(\psi)$ to error of order $O(n^{-1})$.
- But, confidence sets constructed from bootstrap distribution of $R(\psi)$ have third-order coverage accuracy: coverage error of order $O(n^{-3/2})$.

Detail

The confidence set is

$$\{\psi : R(\psi) \leq \tilde{G}^{-1}(1-\alpha)\},$$

where $\tilde{G}$ denotes the sampling distribution of $R(\psi)$ under $F(y;\tilde{\theta})$, the distribution with parameter value fixed as $\tilde{\theta} = (\psi, \hat{\phi}_\psi)$.

Detail

The confidence set is

$$\{\psi : R(\psi) \leq \tilde{G}^{-1}(1-\alpha)\},$$

where $\tilde{G}$ denotes the sampling distribution of $R(\psi)$ under $F(y;\tilde{\theta})$, the distribution with parameter value fixed as $\tilde{\theta} = (\psi, \hat{\phi}_\psi)$.

Corresponds to a significance function $u(Y,\psi) = \tilde{G}(R(\psi))$.

Detail

The confidence set is

$$\{\psi : R(\psi) \leq \tilde{G}^{-1}(1-\alpha)\},$$

where $\tilde{G}$ denotes the sampling distribution of $R(\psi)$ under $F(y;\tilde{\theta})$, the distribution with parameter value fixed as $\tilde{\theta} = (\psi, \hat{\phi}_\psi)$.

Corresponds to a significance function $u(Y,\psi) = \tilde{G}(R(\psi))$.

Note: a different bootstrap calculation required for each ψ. The significance function may not be monotonic.

Other schemes, e.g. substituting global MLE of nuisance parameter, less effective, in general. If $\hat{G}$ denotes the distribution of $R(\psi)$ under sampling from $F(y;\hat{\theta})$, the confidence set

$$\{\psi : R(\psi) \leq \hat{G}^{-1}(1-\alpha)\},$$

has coverage error of order $O(n^{-1})$.

PPI

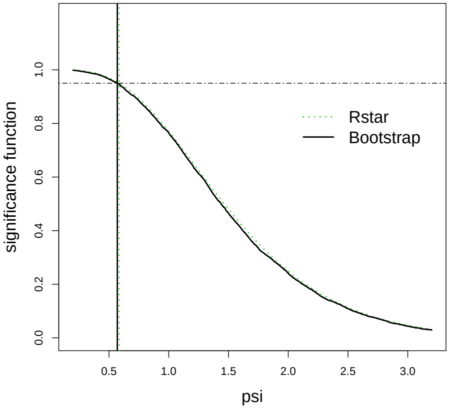
Inference based on bootstrapping distribution of $R(\psi)$ respects PPI.

PPI

Inference based on bootstrapping distribution of $R(\psi)$ respects PPI.

So does making normal approximation to sampling distribution of $R^*(\psi)$.

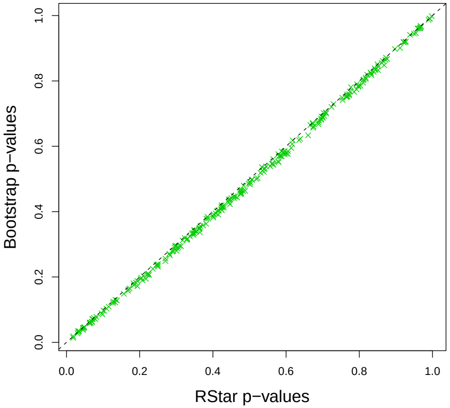
RE, data sample: significance functions



RE, data example: 95% confidence limits

- $R^*(\psi)$: interval is $(0.585, \infty)$.
- Bootstrap $R(\psi)$: interval is $(0.570, \infty)$.

RE: $n = 5$, bootstrap $p-$values vs R^* $p-$values



A practical example: signal detection

LHC: detection of signal in presence of background noise.

A practical example: signal detection

LHC: detection of signal in presence of background noise.

Set confidence limits on underlying signal, based on data from observation channel.

Observation is number of times a particular event is observed. Supposed to have Poisson distribution with mean $\psi\gamma + \beta$, where interest parameter ψ represents signal, β and γ represent respectively a background rate at which event occurs and efficiency of the measurement device.

Precise formulation

Available data is y_1, y_2, y_3. Realizations of independent Poisson random variables with means $\psi\gamma + \beta$, βt and γu respectively, where t and u are known and parameters ψ, β, γ are unknown.

Precise formulation

Available data is y_1, y_2, y_3. Realizations of independent Poisson random variables with means $\psi\gamma + \beta$, βt and γu respectively, where t and u are known and parameters ψ, β, γ are unknown.

In principle, $\psi \geq 0$, and nuisance parameters β, γ are positive.

Precise formulation

Available data is y_1, y_2, y_3. Realizations of independent Poisson random variables with means $\psi\gamma + \beta$, βt and γu respectively, where t and u are known and parameters ψ, β, γ are unknown.

In principle, $\psi \geq 0$, and nuisance parameters β, γ are positive.

Consider $y_1 = 1, y_2 = 8, y_3 = 14$, with $t = 27, u = 80$.

Inference

Appropriate inference is test of hypothesis $\psi = 0$ against one-sided alternative $\psi > 0$.

Inference

Appropriate inference is test of hypothesis $\psi = 0$ against one-sided alternative $\psi > 0$.

Significance probability is one minus significance function at $\psi = 0$.

Inference

Appropriate inference is test of hypothesis $\psi = 0$ against one-sided alternative $\psi > 0$.

Significance probability is one minus significance function at $\psi = 0$.

$R(\psi)$: $p-$value is $1 - \Phi\{R(0)\} = 0.163$.

$R^*(\psi)$: $p-$value is $1 - \Phi\{R^*(0)\} = 0.127$.

Bootstrap $R(\psi)$: $p-$value is 0.156 [10,000 bootstrap samples].

Inference

Appropriate inference is test of hypothesis $\psi = 0$ against one-sided alternative $\psi > 0$.

Significance probability is one minus significance function at $\psi = 0$.

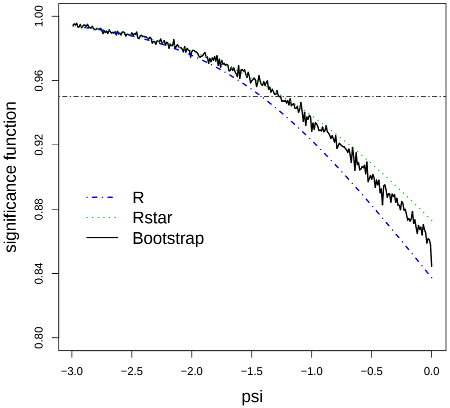
$R(\psi)$: $p-$value is $1 - \Phi\{R(0)\} = 0.163$.

$R^*(\psi)$: $p-$value is $1 - \Phi\{R^*(0)\} = 0.127$.

Bootstrap $R(\psi)$: $p-$value is 0.156 [10,000 bootstrap samples].

Weak evidence of positive signal.

Significance functions



Remarks

Remarks

- Lower confidence limits are negative. If insist on confidence limits, take lower limit as maximum, $\max\{0, \psi_\alpha\}$, of actual limit ψ_α and lower physically admissible value of zero? All lower confidence limits are zero (coherent with $p-$values for testing for a positive signal). Calculation of $p-$value more appropriate?

Remarks

- Lower confidence limits are negative. If insist on confidence limits, take lower limit as maximum, $\max\{0, \psi_\alpha\}$, of actual limit ψ_α and lower physically admissible value of zero? All lower confidence limits are zero (coherent with $p-$values for testing for a positive signal). Calculation of $p-$value more appropriate?

- Even though large simulation is carried out, bootstrap significance function here is not smooth. Smoothing required?

Remarks

- Lower confidence limits are negative. If insist on confidence limits, take lower limit as maximum, $\max\{0, \psi_\alpha\}$, of actual limit ψ_α and lower physically admissible value of zero? All lower confidence limits are zero (coherent with $p-$values for testing for a positive signal). Calculation of $p-$value more appropriate?

- Even though large simulation is carried out, bootstrap significance function here is not smooth. Smoothing required?

- Discrete distribution. Does not effect essential inferential issues, but introduces (mainly computational) complications. Not all theoretical results about rates of error etc. necessarily apply to such cases. Good practical performance.

Example: $p = 1$, $q = 20$, 'Behrens-Fisher'

Let $Y_{ij},\ i = 1, \ldots, n_g,\ j = 1, \ldots, n_i$ be independent normal rvs, $Y_{ij} \sim N(\mu, \sigma_i^2)$.

Interest parameter is μ, nuisance parameter $(\sigma_1^2, \ldots, \sigma_{n_g}^2)$.

Example: $p = 1$, $q = 20$, 'Behrens-Fisher'

Let $Y_{ij},\ i = 1, \ldots, n_g,\ j = 1, \ldots, n_i$ be independent normal rvs, $Y_{ij} \sim N(\mu, \sigma_i^2)$.

Interest parameter is μ, nuisance parameter $(\sigma_1^2, \ldots, \sigma_{n_g}^2)$.

Take case $n_g = 20$, $n_i \equiv n$, $\sigma_i^2 = i$, varying n.

Compare coverages of one-sided (upper) confidence limits for true $\mu = 0$, obtained by $N(0,1)$ approximation to distributions of R, R^*, and by (constrained) bootstrap estimation of distribution of R (based on drawing 10,000 bootstrap samples). Figures based on 50,000 MC replications.

Nominal (%)		1.0	5.0	10.0	90.0	95.0	99.0
$n = 3$	R	7.6	16.3	22.5	77.6	83.7	92.4
	R^*	3.3	9.9	15.9	83.3	89.4	96.1
	boot	1.1	5.1	10.2	89.9	94.8	99.0
$n = 5$	R	3.3	9.9	15.7	84.3	90.2	96.7
	R^*	1.9	7.2	12.8	87.3	92.7	98.0
	boot	1.0	5.0	10.0	90.1	95.0	99.0
$n = 10$	R	1.8	7.0	12.5	87.6	92.9	98.1
	R^*	1.3	5.9	11.1	88.9	93.9	98.6
	boot	1.0	5.1	10.0	90.0	94.8	98.9

Example: $p = 2$, $q = 10$

Let $Y_{1ij}, Y_{2ij}, \ i = 1, \ldots, n_g, \ j = 1, \ldots, n_i$ be independent normal rvs, $Y_{1ij} \sim N(\mu_1, \sigma_i^2), Y_{2ij} \sim N(\mu_2, \sigma_i^2)$.

Interest parameter is (μ_1, μ_2), nuisance parameter $(\sigma_1^2, \ldots, \sigma_{n_g}^2)$.

Example: $p = 2$, $q = 10$

Let $Y_{1ij}, Y_{2ij}, \ i = 1, \ldots, n_g, \ j = 1, \ldots, n_i$ be independent normal rvs, $Y_{1ij} \sim N(\mu_1, \sigma_i^2)$, $Y_{2ij} \sim N(\mu_2, \sigma_i^2)$.

Interest parameter is (μ_1, μ_2), nuisance parameter $(\sigma_1^2, \ldots, \sigma_{n_g}^2)$.

Take case $n_g = 10$, $n_i \equiv n$, $\sigma_i^2 = i$, varying n.

Compare coverages of confidence regions for true $(\mu_1, \mu_2) = (1, 2)$, obtained by χ_2^2 approximation to distribution of LRS W, (empirical) Bartlett correction of W and by bootstrap estimation of sampling distribution of W (based on drawing 10,000 bootstrap samples). Figures based on 50,000 MC replications.

Nominal (%)		1.0	5.0	10.0	90.0	95.0	99.0
$n = 5$	W	0.8	4.1	8.0	83.9	90.7	97.5
	$\bar{W}_c$	1.1	5.1	10.1	90.0	95.0	99.1
	boot	0.9	5.0	9.9	89.9	94.9	99.0
$n = 10$	W	0.8	4.5	9.0	87.7	93.4	98.5
	$\bar{W}_c$	1.1	5.1	9.9	90.0	95.0	99.1
	boot	1.0	4.9	9.8	90.0	95.0	99.1
$n = 20$	W	0.9	4.6	9.6	89.2	94.3	98.7
	$\bar{W}_c$	1.0	4.9	10.1	90.3	95.0	99.0
	boot	1.0	5.0	9.6	90.0	95.1	99.1

Conditional properties of bootstrap, $p = 1$

Recall, bootstrap applied unconditionally.

Conditional properties of bootstrap, $p = 1$

Recall, bootstrap applied unconditionally.

- Multi-parameter exponential family context: inference agreeing with exact conditional inference to relative error third-order, $O(n^{-3/2})$. Same conditional accuracy as R^*. DiCiccio & Young (2008).

Conditional properties of bootstrap, $p = 1$

Recall, bootstrap applied unconditionally.

- Multi-parameter exponential family context: inference agreeing with exact conditional inference to relative error third-order, $O(n^{-3/2})$. Same conditional accuracy as R^*. DiCiccio & Young (2008).
- Same context, automatically reproduces appropriate objective ('conditional second-order probability matching') Bayesian inference to order $O(n^{-3/2})$, in many circumstances.

- Ancillary statistic models: bootstrap inference using $R(\psi)$ agrees with conditional inference to second-order, $O(n^{-1})$;

- Ancillary statistic models: bootstrap inference using $R(\psi)$ agrees with conditional inference to second-order, $O(n^{-1})$;
- Same for other asymptotically $N(0,1)$ pivots, provided these are constructed using observed information. Pivot must be 'stable' to second-order, $O(n^{-1})$: marginal and conditional distributions must agree to that order. Not true, for example, for $T_W(\psi)$ and $T_S(\psi)$.

- Ancillary statistic models: bootstrap inference using $R(\psi)$ agrees with conditional inference to second-order, $O(n^{-1})$;
- Same for other asymptotically $N(0,1)$ pivots, provided these are constructed using observed information. Pivot must be 'stable' to second-order, $O(n^{-1})$: marginal and conditional distributions must agree to that order. Not true, for example, for $T_W(\psi)$ and $T_S(\psi)$.
- Compare with third-order conditional accuracy of R^*.

- Ancillary statistic models: bootstrap inference using $R(\psi)$ agrees with conditional inference to second-order, $O(n^{-1})$;
- Same for other asymptotically $N(0,1)$ pivots, provided these are constructed using observed information. Pivot must be 'stable' to second-order, $O(n^{-1})$: marginal and conditional distributions must agree to that order. Not true, for example, for $T_W(\psi)$ and $T_S(\psi)$.
- Compare with third-order conditional accuracy of R^*.
- Third-order conditional accuracy unwarranted? Ancillary statistics typically not unique, different conditional inferences will typically only agree to second-order.

Vector interest parameter ($p > 1$)

Repeated sampling perspective: simulating the distribution of $W(\psi)$, at either global MLE or constrained MLE, produces $p-$values uniformly distributed under H_0, to error of order $O(n^{-2})$.

Vector interest parameter ($p > 1$)

Repeated sampling perspective: simulating the distribution of $W(\psi)$, at either global MLE or constrained MLE, produces $p-$values uniformly distributed under H_0, to error of order $O(n^{-2})$.

Ancillary statistic models: bootstrapping $W(\psi)$ approximates exact conditional inference given $A = a$ to third-order, $O(n^{-3/2})$.

Objective Bayes ($p = 1$)

Objective Bayes ($p = 1$)

- Exponential family context: conditional (and hence unconditional, repeated sampling) frequentist inference accurate to $O(n^{-3/2})$ achievable by any prior in a general class, provided a simple model condition holds. DiCiccio & Young (2010).

Objective Bayes ($p = 1$)

- Exponential family context: conditional (and hence unconditional, repeated sampling) frequentist inference accurate to $O(n^{-3/2})$ achievable by any prior in a general class, provided a simple model condition holds. DiCiccio & Young (2010).

- Ancillary statistics models: unconditional higher-order probability matching priors give conditional frequentist accuracy to $O(n^{-3/2})$ under some further conditions (DiCiccio, Kuffner & Young, 2012). But now, in key cases exact conditional matching priors exist and are unique. In these cases, objective Bayes is preferred route to conditional frequentist accuracy?

II: Detailed theoretical analysis

A detailed analysis of $R^*(\psi)$

The R^* statistic is defined by

$$R^*(\psi) = R(\psi) + R(\psi)^{-1} \log(v(\psi)/R(\psi)),$$

where $v(\psi)$ is given by

$$v(\psi) = \left| \begin{array}{c} L_{;\hat{\theta}}(\hat{\theta}) - L_{;\hat{\theta}}(\tilde{\theta}) \\ L_{\phi;\hat{\theta}}(\tilde{\theta}) \end{array} \right| / \{ |j_{\phi\phi}(\tilde{\theta})|^{1/2} |j(\hat{\theta})|^{1/2} \}.$$

Here, the log-likelihood function has been written as $L(\theta; \hat{\theta}, a)$, with $(\hat{\theta}, a)$ minimal sufficient and a ancillary, and

$$L_{;\hat{\theta}}(\theta) \equiv L_{;\hat{\theta}}(\theta; \hat{\theta}, a) = \frac{\partial}{\partial \hat{\theta}} L(\theta; \hat{\theta}, a),$$

$$L_{\phi;\hat{\theta}}(\theta) \equiv L_{\phi;\hat{\theta}}(\theta; \hat{\theta}, a) = \frac{\partial^2}{\partial \phi \partial \hat{\theta}} L(\theta; \hat{\theta}, a).$$

Also, j denotes the observed information matrix, $j(\theta) = (-L_{rs}(\theta))$, with $L_{rs}(\theta) = \partial^2 L(\theta)/\partial\theta^r \partial\theta^s$, and $j_{\phi\phi}$ denotes its (ϕ, ϕ) component.

A decomposition

We may decompose $R^*(\psi)$ as

$$R^*(\psi) = R(\psi) + \mathrm{NP}(\psi) + \mathrm{INF}(\psi),$$

for quantities $\mathrm{NP}(\psi)$ and $\mathrm{INF}(\psi)$, both of order $O_p(n^{-1/2})$.

Definitions

Explicitly, we have

$$\mathrm{NP}(\psi) = -\frac{1}{R(\psi)} \log C(\psi),$$

where

$$C(\psi) = \frac{\{|j_{\phi\phi}(\hat{\theta})||j_{\phi\phi}(\tilde{\theta})|\}^{1/2}}{|L_{\phi;\hat{\phi}}(\tilde{\theta})|},$$

with $L_{\phi;\hat{\phi}}(\theta) \equiv L_{\phi;\hat{\phi}}(\theta;\hat{\theta},a) = \partial^2 L(\theta;\hat{\theta},a)/\partial\phi\partial\hat{\phi}$ and $j_{\phi\phi}$ denoting, as before, the (ϕ,ϕ) component of the observed information j.

Also,

$$\mathrm{INF}(\psi) = \frac{1}{R(\psi)} \log\{u(\psi)/R(\psi)\},$$

where

$$u(\psi) = j_p(\hat{\psi})^{-1/2} \frac{\partial}{\partial \hat{\psi}} \{M(\hat{\psi}) - M(\psi)\}.$$

Here j_p is the profile observed information, $j_p(\psi) = -\partial^2 M(\psi)/\partial\psi^2$, and the derivative with respect to $\hat{\psi}$ is calculated with $M(\hat{\psi}) - M(\psi)$ considered as a function of $\psi, \hat{\psi}, \hat{\phi}_\psi$ and a.

Interpretations

$\mathrm{NP}(\psi)$ and $\mathrm{INF}(\psi)$ are interpreted as correcting respectively for presence of the nuisance parameter ϕ and deviation from standard normality of $R(\psi)$ itself.

Interpretations

$\mathrm{NP}(\psi)$ and $\mathrm{INF}(\psi)$ are interpreted as correcting respectively for presence of the nuisance parameter ϕ and deviation from standard normality of $R(\psi)$ itself.

So, broadly speaking, $\mathrm{INF}(\psi)$ represents what we can eliminate by bootstrapping to replace asymptotic approximation, $\mathrm{NP}(\psi)$ represents intrinsic difficulty of the inference.

Quantitative analysis

Quantitative analysis

- Examination of $\mathrm{NP}(\psi)$ and $\mathrm{INF}(\psi)$ provides simple quantitative method for measuring the respective effects of the two adjustments.

Quantitative analysis

- Examination of $\mathrm{NP}(\psi)$ and $\mathrm{INF}(\psi)$ provides simple quantitative method for measuring the respective effects of the two adjustments.
- Observe $\mathrm{NP}(\psi)$ and $\mathrm{INF}(\psi)$ are determined to $O_p(n^{-1})$ by their means, $E\{\mathrm{NP}(\psi)\}$ and $E\{\mathrm{INF}(\psi)\}$.

Quantitative analysis

- Examination of $\mathrm{NP}(\psi)$ and $\mathrm{INF}(\psi)$ provides simple quantitative method for measuring the respective effects of the two adjustments.
- Observe $\mathrm{NP}(\psi)$ and $\mathrm{INF}(\psi)$ are determined to $O_p(n^{-1})$ by their means, $E\{\mathrm{NP}(\psi)\}$ and $E\{\mathrm{INF}(\psi)\}$.
- Explicit approximation of $E\{\mathrm{NP}(\psi)\}$ and $E\{\mathrm{INF}(\psi)\}$ provides statistical insight, in particular to effects of high-dimensional nuisance parameter on inference and to impact of nuisance parameter on parametric bootstrap.

Expectations

We have:

$$E\{\mathrm{INF}(\psi)\} = \eta^{1/2}\lambda^{1r}\tau^{st}(\tfrac{1}{2}\lambda_{rs,t} + \tfrac{1}{6}\lambda_{rst}) + O(n^{-1});$$

$$E\{\mathrm{NP}(\psi)\} = -\eta^{1/2}\lambda^{1r}\nu^{st}(\lambda_{rs,t} + \tfrac{1}{2}\lambda_{rst}) + O(n^{-1}).$$

Interpretation

If there is no nuisance parameter, then $\lambda^{11} = (\lambda_{11})^{-1}$, $\eta = -\lambda_{11}$, $\tau^{11} = (-\lambda_{11})^{-1}$, and $\nu^{11} = 0$, and it follows that

$$E\{R(\psi)\} = (-\lambda_{11})^{-3/2}(\tfrac{1}{2}\lambda_{11,1} + \tfrac{1}{6}\lambda_{111}) + O(n^{-1}).$$

Suppose there is a vector nuisance parameter ϕ, but assume that the interest parameter ψ and ϕ are orthogonal [always achievable in principle]; then $\lambda^{11} = (\lambda_{11})^{-1}$, $\eta = -\lambda_{11}$, $\lambda^{1a} = 0$ ($a = 2, \ldots, d$), $\tau^{rs} = 0$ except when $r = s = 1$, in which case $\tau^{11} = (-\lambda_{11})^{-1}$, and

$$E\{\mathrm{INF}(\psi)\} = -(-\lambda_{11})^{-3/2}(\tfrac{1}{2}\lambda_{11,1} + \tfrac{1}{6}\lambda_{111}) + O(n^{-1}).$$

Therefore, to error of order $O(n^{-1})$, $E\{\mathrm{INF}(\psi)\}$ corresponds to a mean adjustment for the signed root statistic $R(\psi)$ in the problem where the orthogonal nuisance parameter ϕ is known.

The $N(0,1)$ approximation to the distribution of $R(\psi)$ is typically rather accurate in scalar parameter problems, so the mean adjustment should be generally small, so we can anticipate that $\mathrm{INF}(\psi)$ is in some generality small.

Further analysis

In principle, we can always reparameterize so that ψ and the nuisance parameter ϕ are orthogonal.

Further analysis

In principle, we can always reparameterize so that ψ and the nuisance parameter ϕ are orthogonal.

Invariance of adjustments to reparameterization allows nuisance parameter effects to be quantified by

$$E\{\mathrm{NP}(\psi)\} = -\tfrac{1}{2}(-\lambda_{11})^{-1/2}\lambda^{ab}\lambda_{ab1} + O(n^{-1}).$$

Multiple sum over nuisance parameter, in accordance with intuition that $\mathrm{NP}(\psi)$ can be anticipated to be large when dimensionality of nuisance parameter is large. How large?

Explicit approximations

We may explicitly approximate $E\{\mathrm{INF}(\psi)\}$ to $O(n^{-1})$ by

$$g_{\mathrm{INF}}(\theta) = \eta^{1/2}\lambda^{1r}\tau^{st}(\tfrac{1}{2}\lambda_{rs,t} + \tfrac{1}{6}\lambda_{rst})$$

and $E\{\mathrm{NP}(\psi)\}$ to the same order by

$$g_{\mathrm{NP}}(\theta) = -\eta^{1/2}\lambda^{1r}\nu^{st}(\lambda_{rs,t} + \tfrac{1}{2}\lambda_{rst}).$$

Remarks 1

Quantities $g_{\mathrm{INF}}(\theta)$ and $g_{\mathrm{NP}}(\theta)$ are related to asymptotic quantities detailed by Efron (1987, JASA) in description of the 'bias corrected accelerated', BC_a, method of construction of bootstrap confidence intervals.

Remarks 1

Quantities $g_{\mathrm{INF}}(\theta)$ and $g_{\mathrm{NP}}(\theta)$ are related to asymptotic quantities detailed by Efron (1987, JASA) in description of the ‘bias corrected accelerated’, BC_a, method of construction of bootstrap confidence intervals.

Specifically, we have $g_{\mathrm{INF}}(\theta) = a_c$ and $g_{\mathrm{NP}}(\theta) = z_0 - a_c$, where a_c and z_0 are respectively acceleration and bias-correction quantities.

The quantity a_c satisfies

$$a_c = -\frac{1}{6}\{\mathrm{skew}(U) + \mathrm{skew}(T)\} + O(n^{-1}),$$

where $U = (\hat{\psi} - \psi)/\sigma$, with σ^2 the asymptotic variance of $\hat{\psi}$, so that $\sigma^2 \equiv \sigma^2(\theta) = \lambda^{1,1} + O(n^{-2})$, and $T = (\hat{\psi} - \psi)/\hat{\sigma}$, with $\hat{\sigma}^2 = \sigma^2(\hat{\theta})$.

The quantity a_c satisfies

$$a_c = -\frac{1}{6}\{\mathrm{skew}(U) + \mathrm{skew}(T)\} + O(n^{-1}),$$

where $U = (\hat{\psi} - \psi)/\sigma$, with σ^2 the asymptotic variance of $\hat{\psi}$, so that $\sigma^2 \equiv \sigma^2(\theta) = \lambda^{1,1} + O(n^{-2})$, and $T = (\hat{\psi} - \psi)/\hat{\sigma}$, with $\hat{\sigma}^2 = \sigma^2(\hat{\theta})$.

Also, z_0 is interpreted by

$$\Phi(z_0) = \mathrm{Pr}(\hat{\psi} \leq \psi) + O(n^{-1}),$$

where Φ is the standard normal distribution function.

Remarks 2

Quantities $g_{\mathrm{NP}}(\theta)$ and $g_{\mathrm{INF}}(\theta)$ are both of order $O(n^{-1/2})$.

Calculation of the individual values provides valuable statistical insight to importance of nuisance parameter effects and likely operational performance of bootstrap (which is completely determined by nuisance parameter).

Remarks 3

In general, $g_{\mathrm{NP}}(\theta)$ and $g_{\mathrm{INF}}(\theta)$ depend on the unknown parameter θ.

Remarks 3

In general, $g_{\mathrm{NP}}(\theta)$ and $g_{\mathrm{INF}}(\theta)$ depend on the unknown parameter θ.

Bootstrap principle: they may be estimated by $g_{\mathrm{NP}}(\tilde{\theta})$ and $g_{\mathrm{INF}}(\tilde{\theta})$ respectively.

Remarks 3

In general, $g_{\mathrm{NP}}(\theta)$ and $g_{\mathrm{INF}}(\theta)$ depend on the unknown parameter θ.

Bootstrap principle: they may be estimated by $g_{\mathrm{NP}}(\tilde{\theta})$ and $g_{\mathrm{INF}}(\tilde{\theta})$ respectively.

A simple adjustment of the signed root statistic $R(\psi)$, is given by $R_a(\psi) = R(\psi) + g_{\mathrm{NP}}(\tilde{\theta}) + g_{\mathrm{INF}}(\tilde{\theta})$.

Remarks 3

In general, $g_{\mathrm{NP}}(\theta)$ and $g_{\mathrm{INF}}(\theta)$ depend on the unknown parameter θ.

Bootstrap principle: they may be estimated by $g_{\mathrm{NP}}(\tilde{\theta})$ and $g_{\mathrm{INF}}(\tilde{\theta})$ respectively.

A simple adjustment of the signed root statistic $R(\psi)$, is given by $R_a(\psi) = R(\psi) + g_{\mathrm{NP}}(\tilde{\theta}) + g_{\mathrm{INF}}(\tilde{\theta})$.

Since $g_{\mathrm{NP}}(\tilde{\theta}) - g_{\mathrm{NP}}(\theta) = O_p(n^{-1})$, we have that $R_a(\psi) = R^*(\psi) + O_p(n^{-1})$, and therefore that $R_a(\psi)$ is $N(0,1)$ to error of order $O(n^{-1})$.

Methodological Issues

- ‘Uniqueness of inference’.
- Computational considerations.
- Relationship between analytic and bootstrap approaches.
- Choice of a ‘good pivot’.

When do inferences agree?

In general, $p-$values from different asymptotically $N(0,1)$ pivots will agree only to first-order, $O(n^{-1/2})$.

When do inferences agree?

In general, $p-$values from different asymptotically $N(0,1)$ pivots will agree only to first-order, $O(n^{-1/2})$.

However, establish simple sufficient conditions, under which $p-$values from two statistics will agree to second-order, $O(n^{-1})$, provided approximations to distributions accurate to $O(n^{-1})$ are employed. Such accurate approximation obtained quite generally by bootstrap.

Consequences

- $T_W(\psi)$ and $T_S(\psi)$ in general do not provide $p-$values that agree with those from $R(\psi)$ to order $O_p(n^{-1})$.

Consequences

- $T_W(\psi)$ and $T_S(\psi)$ in general do not provide $p-$values that agree with those from $R(\psi)$ to order $O_p(n^{-1})$.
- But, versions of Wald and score statistics constructed using observed information will yield $p-$values agreeing with those from $R(\psi)$ to $O_p(n^{-1})$.

Consequences

- $T_W(\psi)$ and $T_S(\psi)$ in general do not provide $p-$values that agree with those from $R(\psi)$ to order $O_p(n^{-1})$.
- But, versions of Wald and score statistics constructed using observed information will yield $p-$values agreeing with those from $R(\psi)$ to $O_p(n^{-1})$.
- Etc., etc.

Computational considerations

Use of $W(\psi)$ and $R(\psi)$ requires calculation of both global and constrained MLEs. Potentially unattractive compared to Wald statistic, $T_W(\psi)$ [or multivariate version]. Latter routinely employed in statistical packages etc., but not stable or parameterization invariant.

Computational considerations

Use of $W(\psi)$ and $R(\psi)$ requires calculation of both global and constrained MLEs. Potentially unattractive compared to Wald statistic, $T_W(\psi)$ [or multivariate version]. Latter routinely employed in statistical packages etc., but not stable or parameterization invariant.

Bootstrap: must recalculate for a series of B bootstrap samples. General guideline: B of order of few 1000's to reduce Monte Carlo variability to acceptable levels, to 'capture' good theoretical properties. In small samples or with high-dimensional nuisance parameter solution of likelihood equations can be a worry.

Computational considerations

Use of $W(\psi)$ and $R(\psi)$ requires calculation of both global and constrained MLEs. Potentially unattractive compared to Wald statistic, $T_W(\psi)$ [or multivariate version]. Latter routinely employed in statistical packages etc., but not stable or parameterization invariant.

Bootstrap: must recalculate for a series of B bootstrap samples. General guideline: B of order of few 1000's to reduce Monte Carlo variability to acceptable levels, to 'capture' good theoretical properties. In small samples or with high-dimensional nuisance parameter solution of likelihood equations can be a worry.

$R^*(\psi)$: computationally simple, potentially awkward analytic calculations/coding. (Highly) stable, parameterization invariant.

General pivot

For general, asymptotically $N(0,1)$ pivot $T(\psi)$, producing same $p-$values as $R(\psi)$ to $O_p(n^{-1})$, normalized (Cornish-Fisher) version of $T(\psi)$, $N(0,1)$ to error of order $O(n^{-1})$, is:

$$T(\psi) - \frac{1}{6}\kappa_3\{T(\psi)\}^2 + NP(\psi) + INF(\psi),$$

in terms of third cumulant κ_3 of $T(\psi)$.

When does bootstrap work?

Normalizing transformation is automatically incorporated by bootstrap refinement of the asymptotic $N(0,1)$ approximation.

When does bootstrap work?

Normalizing transformation is automatically incorporated by bootstrap refinement of the asymptotic $N(0,1)$ approximation.

- 'Primary effect of bootstrap is to estimate skewness'. Key requirement for bootstrap to perform well is that third cumulant of $T(\psi)$ can be estimated accurately. Difficult if n is small, or number of nuisance parameters is large.

When does bootstrap work?

Normalizing transformation is automatically incorporated by bootstrap refinement of the asymptotic $N(0,1)$ approximation.

- 'Primary effect of bootstrap is to estimate skewness'. Key requirement for bootstrap to perform well is that third cumulant of $T(\psi)$ can be estimated accurately. Difficult if n is small, or number of nuisance parameters is large.
- If skewness is small, as with $R(\psi)$, where it is of order $O(n^{-1})$, estimation of skewness less crucial, explaining why bootstrap works extraordinarily well with $R(\psi)$.

- If skewness is constant with respect to nuisance parameter, bootstrap should work well, inaccuracy in estimating nuisance parameter does not translate into inaccuracy in estimating skewness.

- If skewness is constant with respect to nuisance parameter, bootstrap should work well, inaccuracy in estimating nuisance parameter does not translate into inaccuracy in estimating skewness.

- Previous focus on variance-stabilizing transformations to improve bootstrap accuracy: variance stabilizing transformations typically reduce skewness of parameterization dependent pivot $T(\psi)$. DiCiccio, Monti & Young (2006).

Relationship between Bootstrap and $R^*(\psi)$

Conceptually related, not distinct methodologies.

Details

Specifically:

Details

Specifically:

- $p-$ values calculated from $N(0, 1)$ approximation to distribution of $R^*(\psi)$ will quite generally agree with those from bootstrap to order $O_p(n^{-1})$.

Details

Specifically:

- $p-$ values calculated from $N(0,1)$ approximation to distribution of $R^*(\psi)$ will quite generally agree with those from bootstrap to order $O_p(n^{-1})$.

- Multi-parameter exponential family models: (unconditional) bootstrap $p-$values agree with those from $R^*(\psi)$ to $O_p(n^{-3/2})$.

Details

Specifically:

- $p-$ values calculated from $N(0,1)$ approximation to distribution of $R^*(\psi)$ will quite generally agree with those from bootstrap to order $O_p(n^{-1})$.
- Multi-parameter exponential family models: (unconditional) bootstrap $p-$values agree with those from $R^*(\psi)$ to $O_p(n^{-3/2})$.
- Ancillary statistic models: normal approximation to $R^*(\psi)$ is an $O(n^{-3/2})$ (saddlepoint) approximation to conditional bootstrap [which could use if we could simulate the conditional distribution of $R(\psi)$ given $A=a$].

The bottom line

If likelihood equations can be reliably solved, analytic simplicity indicates bootstrapping of $R(\psi)$ or $W(\psi)$ as a highly effective methodology.

- Competitive in terms of accuracy with analytic alternatives.

- Competitive in terms of accuracy with analytic alternatives.

- Unlikely to be computationally prohibitive [moderate B adequate to ensure MC variability does not impair good theoretical properties].

- Competitive in terms of accuracy with analytic alternatives.
- Unlikely to be computationally prohibitive [moderate B adequate to ensure MC variability does not impair good theoretical properties].
- Stable (respects CP to high-order) and parameterization invariant: ‘inferentially correctness is OK’.

- Competitive in terms of accuracy with analytic alternatives.

- Unlikely to be computationally prohibitive [moderate B adequate to ensure MC variability does not impair good theoretical properties].

- Stable (respects CP to high-order) and parameterization invariant: 'inferentially correctness is OK'.

- Vector ψ: use bootstrap calculation to estimate mean of $W(\psi)$, then base inference on χ^2 approximation to empirically Bartlett-corrected statistic $\bar{W}_c(\psi)$.

III: Further illustrations

Further Illustration 1: resampling accuracy, RE (ctd)

$Y_1, \ldots, Y_n$ IID inverse Gaussian, with density

$$f(y; \mu, \psi) = \left(\frac{\psi}{2\pi y^3}\right)^{1/2} \exp\left(-\frac{\psi}{2\mu^2 y}(y-\mu)^2\right), \ y > 0,$$

interest parameter is shape ψ, mean μ as nuisance.

Further Illustration 1: resampling accuracy, RE (ctd)

$Y_1, \ldots, Y_n$ IID inverse Gaussian, with density

$$f(y; \mu, \psi) = \left(\frac{\psi}{2\pi y^3} \right)^{1/2} \exp\left(-\frac{\psi}{2\mu^2 y}(y-\mu)^2 \right), \ y > 0,$$

interest parameter is shape ψ, mean μ as nuisance.

20,000 replications, $n = 5$, true $\mu = 1, \psi = 2$. Compare coverages of confidence limits of different nominal coverages obtained by: normal approximation to $R(\psi)$; normal approximation to $R^*(\psi)$; bootstrap of $R(\psi)$; three objective Bayes priors ($OB1$, $OB2$ and $OB3$). Each replication: 5,000 bootstrap samples, MC construction of Bayes posterior quantile.

$OB1$ has $\pi(\psi, \mu) \propto \psi^{-1}\mu^{-2}$, $OB2$ has $\pi(\psi, \mu) \propto \psi^{-1/2}\mu^{-3/2}$ and $OB3$ has $\psi(\psi, \mu) \propto \psi^{-1}\mu^{-3/2}$.

In theory, $OB1$ and $OB3$ should give $O(n^{-3/2})$ coverage accuracy, but not $OB2$. Typical of non-uniqueness of second-order ($O(n^{-3/2})$) matching prior.

$OB1$ has $\pi(\psi, \mu) \propto \psi^{-1}\mu^{-2}$, $OB2$ has $\pi(\psi, \mu) \propto \psi^{-1/2}\mu^{-3/2}$ and $OB3$ has $\psi(\psi, \mu) \propto \psi^{-1}\mu^{-3/2}$.

In theory, $OB1$ and $OB3$ should give $O(n^{-3/2})$ coverage accuracy, but not $OB2$. Typical of non-uniqueness of second-order ($O(n^{-3/2})$) matching prior.

Actually exponential family, appropriate frequentist inference is conditional, but provides instructive example where repeated sampling properties should, in principle, be very similar.

Nominal (%)	1.0	5.0	10.0	90.0	95.0	99.0
$\Phi(R)$	0.3	1.7	3.8	74.6	84.4	95.0
Bootstrap R	1.1	5.1	10.1	90.1	95.1	99.0
$\Phi(R^*)$	1.0	4.8	9.6	89.4	94.8	98.9
$OB1$	1.0	4.9	10.0	90.3	95.1	99.0
$OB2$	3.0	10.6	18.4	94.1	97.2	99.5
$OB3$	1.0	4.7	9.3	89.6	94.7	98.9

Comments

- Normal approximation to distribution of $R(\psi)$ is inaccurate.

Comments

- Normal approximation to distribution of $R(\psi)$ is inaccurate.
- Bootstrap and normal approximation to $R^*(\psi)$ both highly accurate.

Comments

- Normal approximation to distribution of $R(\psi)$ is inaccurate.
- Bootstrap and normal approximation to $R^*(\psi)$ both highly accurate.
- Objective Bayes yields good repeated sampling accuracy, with OB1 or OB3.

Conditional inference: detail

With $S = n^{-1}\sum_i Y_i^{-1}$, $C = n^{-1}\sum_i Y_i$, correct inference is conditional, based on conditional distribution of S, given $C = c$.

Conditional inference: detail

With $S = n^{-1}\sum_i Y_i^{-1}$, $C = n^{-1}\sum_i Y_i$, correct inference is conditional, based on conditional distribution of S, given $C = c$.

Equivalent to inference based on the marginal distribution of $V = \sum_i (Y_i^{-1} - \overline{Y}^{-1})$. Distribution of ψV is χ^2_{n-1}.

Have

$$R(\psi) = \mathrm{sgn}(\hat{\psi} - \psi)\{n(\log \hat{\psi} - 1 - \log \psi + \psi/\hat{\psi})\}^{1/2},$$

$\hat{\psi} = n/V.$

Have

$$R(\psi) = \mathrm{sgn}(\hat{\psi} - \psi)\{n(\log \hat{\psi} - 1 - \log \psi + \psi/\hat{\psi})\}^{1/2},$$

$\hat{\psi} = n/V$.

Distribution of $R(\psi)$ is free of nuisance parameter μ: infinite simulation bootstrap will approximate sampling distribution exactly, no coverage error.

Have

$$R(\psi) = \mathrm{sgn}(\hat{\psi} - \psi)\{n(\log \hat{\psi} - 1 - \log \psi + \psi/\hat{\psi})\}^{1/2},$$

$\hat{\psi} = n/V$.

Distribution of $R(\psi)$ is free of nuisance parameter μ: infinite simulation bootstrap will approximate sampling distribution exactly, no coverage error.

Also, since $R(\psi)$ is a monotonic function of V, bootstrap inference will actually replicate the appropriate exact conditional inference without error.

Further Illustration 2: RE (ctd), extended

Let Y_{ij}, for $i = 1, \ldots, n$ and $j = 1, \ldots, q$ be independent, inverse Gaussian random variables, with Y_{ij} having probability density

$$f(y; \psi, \phi_j) = \{\psi/(2\pi)\}^{1/2} y^{-3/2} \exp\{-\tfrac{1}{2}(\psi y^{-1} + \phi_j y) + (\psi\phi_j)^{1/2}\},$$

$y > 0$, $\psi, \phi_j > 0$, so that $\theta = (\psi, \phi_1, \ldots, \phi_q)$. Here ψ and $(\phi_1, \ldots, \phi_q)$ are non-orthogonal.

Here, irrespective of the parameter value θ,

$$n^{1/2} g_{\mathrm{INF}}(\theta) \equiv -1/\{3(2q)^{1/2}/2\},$$

and

$$n^{1/2} g_{\mathrm{NP}}(\theta) \equiv -(q/2)^{1/2},$$

so that

$$g_{\mathrm{NP}}(\theta)/g_{\mathrm{INF}}(\theta) \equiv 3q/2.$$

The adjusted statistic $R_a(\psi)$ may be constructed in this example without the need to estimate the nuisance parameters $\phi_1, \ldots, \phi_q$.

Implications for bootstrap

To order $O(n^{-1})$ expectations of adjustments do not depend on values of nuisance parameters, only dimension.

Implications for bootstrap

To order $O(n^{-1})$ expectations of adjustments do not depend on values of nuisance parameters, only dimension.

Parametric bootstrap ought to be accurate? Inference, at least to order $O(n^{-1})$, not governed by nuisance parameter values: bootstrap substitution should be OK.

Implications for bootstrap

To order $O(n^{-1})$ expectations of adjustments do not depend on values of nuisance parameters, only dimension.

Parametric bootstrap ought to be accurate? Inference, at least to order $O(n^{-1})$, not governed by nuisance parameter values: bootstrap substitution should be OK.

In fact, here $R(\psi)$ readily seen to be exactly 'pivotal': its sampling distribution is completely free of nuisance parameter. (Infinite simulation) bootstrap gives exact inference. In practice, finite Monte Carlo simulation: exactness is compromised by finiteness of simulation.

Numerical results

Consider sample size $n = 5$, and two values of q, $q = 5, 20$. For parameter settings $\psi = 2, \phi_i = i, i = 1, \ldots, q$, 100,000 datasets were generated.

Accuracy of inference based on $N(0,1)$ approximation to the distributions of the three statistics $R(\psi), R^*(\psi)$ and $R_a(\psi)$ expressed in terms of observed coverages over the 100,000 samples of confidence sets for ψ, obtained by the normal approximation, for different nominal coverages.

Numerical results

Consider sample size $n = 5$, and two values of q, $q = 5, 20$. For parameter settings $\psi = 2, \phi_i = i, i = 1, \ldots, q$, 100,000 datasets were generated.

Accuracy of inference based on $N(0,1)$ approximation to the distributions of the three statistics $R(\psi), R^*(\psi)$ and $R_a(\psi)$ expressed in terms of observed coverages over the 100,000 samples of confidence sets for ψ, obtained by the normal approximation, for different nominal coverages.

Also, coverages when bootstrapping used to approximate sampling distribution of $R(\psi)$, using 5000 bootstrap samples.

Nominal (%)	1·0	2·5	5·0	10·0	90·0	95·0	97·5	99·0
$q = 5$								
$R(\psi)$	0·1	0·3	0·8	1·9	65·4	77·2	85·3	91·9
$R^*(\psi)$	0·9	2·3	4·7	9·5	89·4	94·7	97·4	98·9
$R_a(\psi)$	1·1	2·6	5·0	9·7	87·8	93·6	96·6	98·5
Boot	1·0	2·5	5·0	10·0	90·2	95·1	97·6	99·0
$q = 20$								
$R(\psi)$	0·0	0·0	0·0	0·0	38·8	52·6	64·4	76·5
$R^*(\psi)$	0·8	2·2	4·4	8·8	88·8	94·3	97·2	98·8
$R_a(\psi)$	0·9	2·3	4·4	8·7	86·8	92·9	96·2	98·3
Boot	1·0	2·5	4·9	9·8	90·0	95·0	97·6	99·0

Discussion

Discussion

- $R^*(\psi)$ easily constructed in this full exponential family model, gives accurate results.

Discussion

- $R^*(\psi)$ easily constructed in this full exponential family model, gives accurate results.
- Normal approximation to the distribution of $R(\psi)$ itself is highly inaccurate, and the nuisance parameter effect is substantial.

Discussion

- $R^*(\psi)$ easily constructed in this full exponential family model, gives accurate results.
- Normal approximation to the distribution of $R(\psi)$ itself is highly inaccurate, and the nuisance parameter effect is substantial.
- Coverage figures for the simple adjusted statistic $R_a(\psi)$ are decent.

Discussion

- $R^*(\psi)$ easily constructed in this full exponential family model, gives accurate results.
- Normal approximation to the distribution of $R(\psi)$ itself is highly inaccurate, and the nuisance parameter effect is substantial.
- Coverage figures for the simple adjusted statistic $R_a(\psi)$ are decent.
- Bootstrap is, however, highly accurate.

Discussion

- $R^*(\psi)$ easily constructed in this full exponential family model, gives accurate results.
- Normal approximation to the distribution of $R(\psi)$ itself is highly inaccurate, and the nuisance parameter effect is substantial.
- Coverage figures for the simple adjusted statistic $R_a(\psi)$ are decent.
- Bootstrap is, however, highly accurate.
- Simulation allows estimation of $E\{\mathrm{NP}(\psi)\}$ and $E\{\mathrm{INF}(\psi)\}$: we have, $q = 5$, $E\{\mathrm{NP}(\psi)\}/g_{\mathrm{NP}}(\theta) = 1{\cdot}05$, with $E\{\mathrm{INF}(\psi)\}/g_{\mathrm{INF}}(\theta) = 1{\cdot}02$, so that the approximations to the means of the two adjustments are highly accurate even for $n = 5$.

Further Illustration 3: Curved exponential family model

Let Y_{ij}, for $i = 1, \ldots, n$ and $j = 1, \ldots, q$ be independent normal random variables with means $\mu_j > 0$ and variances $\psi \mu_j^{\zeta}$, with ζ a known constant.

If $\zeta = 0$ or $\zeta = 1$ the model is a full exponential family, otherwise it is curved. The parameter of interest is ψ, with $\mu_1, \ldots, \mu_q$ as nuisance parameters, $\theta = (\psi, \mu_1, \ldots, \mu_q)$. Fix $\zeta = 1/2$. Again, ψ and $(\mu_1, \ldots, \mu_q)$ are non-orthogonal.

Calculation of $R^*(\psi)$ is intractable. Construction of $R_a(\psi)$ is no more complex than in Further Illustration 2.

Now the ratio $g_{\mathrm{NP}}(\theta)/g_{\mathrm{INF}}(\theta)$ does depend (weakly?) on the value of the parameter θ. Illustrative values are given below, for two cases: case (a) has $\psi = 1, \mu_i = i, i = 1, \ldots, q$, while case (b) has $\psi = 2, \mu_i = i, i = 1, \ldots, q$.

q	1	2	5	10	20	50
(a)	1·11	2·45	6·77	14·17	29·09	74·01
(b)	0·82	2·04	6·19	13·49	28·33	73·19

Numerical results

Obtain empirical estimates, based on 20,000 replications, for case (a), with sample size $n = 15$, of coverages of confidence sets obtained by normal approximation to the distributions of $R(\psi)$ and $R_a(\psi)$ in this problem, as before for two values of nuisance parameter dimension, $q = 5, 20$.

Again, compare with bootstrap using 5000 bootstrap samples for each estimation.

Nominal (%)	1·0	2·5	5·0	10·0	90·0	95·0	97·5	99·0
$q = 5$								
$R(\psi)$	4·0	8·2	14·2	23·4	96·5	98·5	99·3	99·8
$R_a(\psi)$	1·1	2·9	5·6	11·1	90·3	95·1	97·6	99·0
Boot	1·0	2·5	5·0	10·2	90·4	95·3	97·7	99·1
$q = 20$								
$R(\psi)$	10·5	18·8	28·0	40·7	98·8	99·6	99·7	99·9
$R_a(\psi)$	1·4	3·1	6·0	11·4	90·4	94·9	97·5	99·0
Boot	1·1	2·6	5·2	10·2	90·2	94·8	97·5	99·0

The distribution of the unadjusted statistic $R(\psi)$ is very far from $N(0,1)$: the empirical adjustment leads to a statistic $R_a(\psi)$ whose distribution is satisfactorily approximated as $N(0,1)$.

Bootstrap is again best.

Further Illustration 4: an example of conditional inference

$Y_1, \ldots, Y_n$ IID gamma, mean μ, shape parameter ν and density

$$f(y; \mu, \nu) = \frac{\nu^\nu}{\Gamma(\nu)} \exp\left[-\nu\left\{\frac{y}{\mu} - \log\left(\frac{y}{\mu}\right)\right\}\right]\frac{1}{y}, \quad y > 0.$$

Further Illustration 4: an example of conditional inference

$Y_1, \ldots, Y_n$ IID gamma, mean μ, shape parameter ν and density

$$f(y; \mu, \nu) = \frac{\nu^\nu}{\Gamma(\nu)} \exp\left[-\nu\left\{\frac{y}{\mu} - \log\left(\frac{y}{\mu}\right)\right\}\right]\frac{1}{y}, \quad y > 0.$$

Appropriate inference on ν, with μ as nuisance, is conditional, based on conditional distribution of $Q = \prod Y_i$, given observed value, c, of $C = \sum Y_i$.

Data configuration $q = 1.0, c = 20.0$, varying n.

Evaluate conditional frequentist confidence levels of bootstrap, analytic and specific objective Bayes limits, against exact conditional inference.

Bootstrap limits based on 5 million samples. MC construction of objective Bayes (OB) limits.

Model condition is not satisfied here, so objective Bayes here does not achieve theoretical $O(n^{-3/2})$ accuracy.

n	Method	5%	(quantile)	95%	(quantile)
5	OB	5.18	(0.122)	95.03	(0.820)
	boot	5.07	(0.121)	95.01	(0.819)
	R^*	5.67	(0.126)	95.35	(0.832)
10	OB	5.11	(0.357)	95.01	(1.370)
	boot	5.00	(0.355)	95.00	(1.369)
	R^*	5.19	(0.358)	95.11	(1.374)
15	OB	5.05	(0.912)	95.01	(2.908)
	boot	4.98	(0.909)	95.00	(2.907)
	R^*	5.06	(0.913)	95.06	(2.912)

Further Illustration 5: conditional inference, Weibull

Let $\{T_1, \ldots, T_n\}$ be random sample from the Weibull density

$$f(t; \nu, \lambda) = \lambda\nu(\lambda t)^{\nu-1} \exp\{-(\lambda t)^{\nu}\}, \ t > 0,$$

interest parameter ν.

Take $Y_i = \log T_i$: the Y_i are random sample from extreme value distribution $EV(\mu, \psi)$, location-scale family, with scale and location parameters $\psi = \nu^{-1}, \ \mu = -\log\lambda$.

Exact conditional inference for ψ conditions on $a = (a_1, \ldots, a_n)$, with $a_i = (y_i - \hat{\mu})/\hat{\psi}$.

5000 replications from $\nu = \lambda = 1$.

One-sided inference: test $H_0 : \psi = 1$, against $\psi > 1$. Inference based on: $N(0, 1)$ approximation to distribution of R; $N(0, 1)$ approximation to distribution of R^*; bootstrapping (marginal) distribution of R.

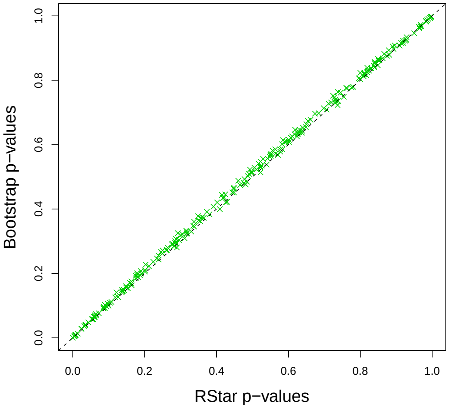
Two-sided inference: test H_0 against $\psi \neq 1$. Inference based on: χ^2_1 approximation to distribution of W; empirical (marginal) Bartlett correction; bootstrapping (marginal) distribution of W.

Compare the average absolute percentage relative error of different approximations to the exact conditional p-values over the 5000 replications.

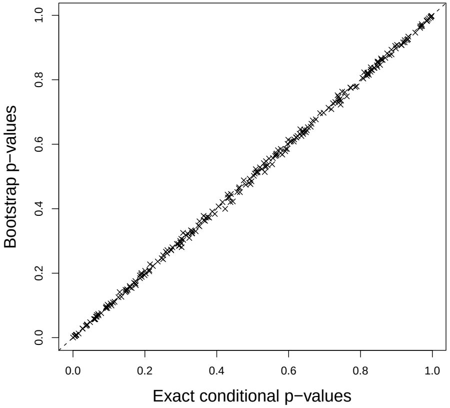
Bootstrap results are based on 5,000,000 samples, same simulation being used for empirical Bartlett correction.

	One-sided			Two-sided		
n	R	R^*	boot	W	$\bar{W}_c$	boot
10	37.387	1.009	0.674	12.318	0.666	0.611
	(0.0%)	(17.1%)	(82.9%)	(0.0%)	(43.9%)	(56.1%)
20	25.473	0.388	0.397	6.118	0.185	0.227
	(0.0%)	(46.2%)	(53.8%)	(0.0%)	(63.4%)	(36.6%)
30	20.040	0.252	0.307	4.158	0.131	0.200
	(0.0%)	(60.9%)	(39.1%)	(0.0%)	(68.7%)	(31.3%)
40	17.865	0.250	0.273	3.064	0.117	0.177
	(0.0%)	(70.1%)	(29.9%)	(0.0%)	(69.7%)	(30.3%)

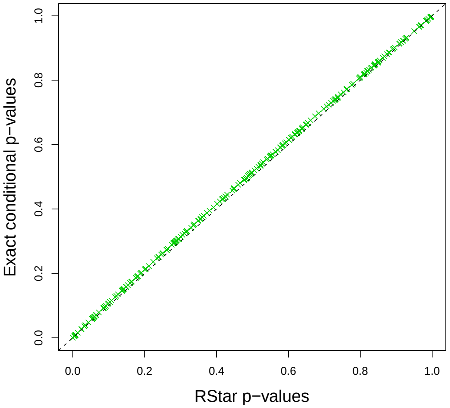
Weibull: $n = 5$, R^* p−values vs bootstrap p−values



Weibull: $n = 5$, exact conditional $p-$values vs bootstrap $p-$values



Weibull: $n = 5$, R^* $p-$values vs exact conditional $p-$values



Concluding remarks

Concluding remarks

- In some circumstances, objective Bayes may be judged as most effective route to frequentist accuracy and correctness, but not always available.

Concluding remarks

- In some circumstances, objective Bayes may be judged as most effective route to frequentist accuracy and correctness, but not always available.
- Analytic and bootstrap approaches to inference highly comparable.

Concluding remarks

- In some circumstances, objective Bayes may be judged as most effective route to frequentist accuracy and correctness, but not always available.
- Analytic and bootstrap approaches to inference highly comparable.
- Strong theoretical basis for use of signed root statistic (and likelihood ratio statistic).

Concluding remarks

- In some circumstances, objective Bayes may be judged as most effective route to frequentist accuracy and correctness, but not always available.
- Analytic and bootstrap approaches to inference highly comparable.
- Strong theoretical basis for use of signed root statistic (and likelihood ratio statistic).
- Discrete data problems. Broad operational conclusions OK, detail of theory less certain.

Concluding remarks

- In some circumstances, objective Bayes may be judged as most effective route to frequentist accuracy and correctness, but not always available.
- Analytic and bootstrap approaches to inference highly comparable.
- Strong theoretical basis for use of signed root statistic (and likelihood ratio statistic).
- Discrete data problems. Broad operational conclusions OK, detail of theory less certain.
- Non-regular problems?

- Robustness to model (mis-)specification important. Lu & Young (2012): work with a robustified version of $R(\psi)$ or $W(\psi)$. Bootstrapping the robust statistic is strikingly effective: simulation of the statistic under wrong distribution can nevertheless yield accurate inference with small n, even if theoretical order of error does not improve on normal approximation. By contrast, methods such as $R^*(\psi)$ are highly non-robust.

- Robustness to model (mis-)specification important. Lu & Young (2012): work with a robustified version of $R(\psi)$ or $W(\psi)$. Bootstrapping the robust statistic is strikingly effective: simulation of the statistic under wrong distribution can nevertheless yield accurate inference with small n, even if theoretical order of error does not improve on normal approximation. By contrast, methods such as $R^*(\psi)$ are highly non-robust.

- Conclusions valid for common adjusted forms of likelihood. Intractable or complex likelihood: theory necessary for composite and pseudo-likelihood, but practical effectiveness striking.

- Robustness to model (mis-)specification important. Lu & Young (2012): work with a robustified version of $R(\psi)$ or $W(\psi)$. Bootstrapping the robust statistic is strikingly effective: simulation of the statistic under wrong distribution can nevertheless yield accurate inference with small n, even if theoretical order of error does not improve on normal approximation. By contrast, methods such as $R^*(\psi)$ are highly non-robust.

- Conclusions valid for common adjusted forms of likelihood. Intractable or complex likelihood: theory necessary for composite and pseudo-likelihood, but practical effectiveness striking.

- Stratification of a bootstrap simulation by values of appropriate conditioning statistic effective as means of reducing error, but awkward to implement.