Convergence Rate Estimation Through Subsampling

Paul Popadiuk

A Thesis

in

The Department

of

Mathematics and Statistics

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Science at

Concordia University

Montreal, Quebec, Canada

September 2004

Your file Votre référence
ISBN: 0-612-94673-8
Our file Notre référence
ISBN: 0-612-94673-8

Abstract

Convergence Rate Estimation Through Subsampling

Paul Popadiuk

This thesis presents an investigation into the estimator for the rate of convergence of a sequence of distribution functions given by Bertail, Politis and Romano (1999, *Annals of Statistics*). The motivation behind the estimator is outlined in detail and subsequently a simulation is carried out in order to validate the results in the above paper. It is brought out through the simulation results that the constants involved in subsampling has to be chosen with care as subsample size affects the results wildly. In order to understand the disparity between the published results and our initial simulation results, the estimator was deconstructed and the results were reinterpreted. The basic reason for disparity is found in the fact that subsample sizes must be much smaller than the size of the sample from which the estimator is constructed, a fact not apparent in the original paper.

Acknowledgements

I would like to thank Professor Angelo Canty for accepting to guide me through my thesis and keeping a keen interest even after leaving Concordia University. Further, I would like to thank Prof. Yogendra P. Chaubey for filling in for Professor Canty after his departure. I am indebted to both of them.

Contents

Chapter 1

Introduction

Let $\delta_n = \delta_n(X_1, ..., X_n)$, based on a random sample $(X_1, ..., X_n)$ be a consistent estimator for some quantity θ and let τ_n be a sequence of positive numbers tending to ∞. Let $J_n(x|\tau) = \Pr\{\tau_n(\delta_n - \theta) \leq x\}$ and suppose that $J_n(x|\tau)$ converges weakly to a nondegenerate distribution function $J(x)$. We will call τ_n the rate of convergence of the estimator δ_n.

It is not at all evident that such a sequence τ_n should even exist, however we will take it as fact that such sequences are ubiquitous.

Lemma 1.1. *Assume that $J_n(x|\tau) \to J(x)$ and that $J(x)$ is nondegenerate. Let a_n be another sequence of positive numbers tending to ∞. Then*

- *if $a_n = o(\tau_n)$ then $J_n(x|a) \to 1$*
- *if $\tau_n = o(a_n)$ then $J_n(x|a) \to 0$*

A proof can be found in Lehmann(1998). An immediate consequence is,

Corollary 1.1.1. *If a sequence τ_n exists for which $J_n(x|\tau) \rightarrow J(x)$ and $J(x)$ is nondegenerate, then τ_n is defined uniquely only upto an order of magnitude.*

We should state, at least in passing, that the principal motivation for determining a sequence τ_n is to facilitate approximations of the quantities $J_n(x|1)$, due to the following proposition:

Proposition 1.2. $J_n(x|\tau) \rightarrow J(x)$ *with* $J(x)$ *nondegenerate* iff $J_n(x|1) \rightarrow J(\tau_n x)$.

In what follows we will attempt first, in Chapter 2, to describe how one may determine the sequence τ_n when one is given the sequence $J_n(x|1) = \Pr\{(\delta_n - \theta) \leq x\}$ and one assumes that a nondegenerate (and, in particular, a continuous and monotonic) limiting distribution $J(x)$ exists. This is foundational. It allows one to understand the methods given in Bertail *et al.* (1999) and described in Chapter 3. In Chapter 4 we undertake the study of the properties of the convergence rate estimator in the context of finite samples through simulation. We compare our results with those given in Bertail *et al.* and find that their results as stated were not reproducible. Through further simulations we provide reasons for the disparity between their results and ours. One basic finding is that the subsample sizes must be much smaller than the size of the sample from which they are constructed.This basic fact was not apparent in the original work of Bertail *et al.* (1999).

Chapter 2

Deterministic Evaluation Of Rates Of Convergence

2.1 Ad Hoc Methods

Lemma 2.1. *Assume that there is a sequence* τ_n *for which* $J_n(x|\tau) \to J(x)$ *and* $J(x)$ *is nondegenerate. Moreover assume that* $E[J_n(x|\tau)] \to E[J(x)] \neq 0$. *Then*

$$\tau_n \propto \frac{1}{E[J_n(x|1)]} \text{ as } n \to \infty.$$

Proof. Observe that $E[J_n(x|\tau)] = \tau_n E[J_n(x|1)]$. Consequently,

$$\tau_n = \frac{E[J(x)]}{E[J_n(x|1)]} + o(1), \text{ as } n \to \infty.$$

The result is immediate. □

In a not surprisingly similar vein, one also has,

Lemma 2.2. *Assume τ_n satisfies $J_n(x|\tau) \to J(x)$ and $J(x)$ is nondegenerate. Moreover assume that $Var[J_n(x|\tau)] \to Var[J(x)]$. Then*

$$\tau_n^2 \propto \frac{1}{Var[J_n(x|1)]} \quad as\ n \to \infty.$$

This Lemma appears in Lehmann(1998). The result is obtained in the same way as Lemma 1.1.

Example 2.1. *Let $X_i \overset{I.I.D.}{\sim} U(0,\theta)$. Based on a sample of size n, the MLE of θ is given by $X_{(n)}$, the largest element of the sample. Elementary computation gives $Var(X_{(n)}) = \frac{n\theta^2}{(n+1)^2(n+2)} \sim \frac{1}{n^2}$, and the above heuristic suggests we take $\tau_n = n$. Let $J_n(x) = \Pr[n(\theta - X_{(n)}) \leq x] = \Pr[X_{(n)} \geqslant \theta - \frac{x}{n}] = 1 - (1 - \frac{x}{\theta n})^n$. As $n \to \infty$, $J_n(x) \to 1 - e^{\frac{-x}{\theta}}$, an exponential distribution with mean θ.*

While the above results are nice due to their simplicity they suffers from the possibly heavy restrictions of requiring convergence of moments or, in the case of Lemma 2.1, prior knowledge that the limit law $J(x)$ has a nonzero mean. A more general approach would be helpful.

2.2 A General Approach

To any distribution function $J(x)$ one may associate a (generalized) inverse distribution, $J^{-1}(t)$ for $t \in (0,1)$, by setting $J^{-1}(t) = \inf\{x | J(x) \geq t\}$. In the case where $J(x)$ is continuous and monotone, the inverse distribution will be the usual Euclidean inverse of $J(x)$. Obviously, to a sequence $J_n(x)$ of distribution functions converging weakly to a distribution function $J(x)$ one may associate a sequence $J_n^{-1}(t)$ and a function $J^{-1}(t)$ and ask whether the weak convergence of distribution functions has any bearing on whether the inverse functions converge. The result is affirmative.

Theorem 2.3. $J_n(x) \to J(x)$ *iff* $J_n^{-1}(t) \to J^{-1}(t)$.

A proof of this result can be found in van der Vaart(1998).

To exploit the duality between distribution functions and their inverses we will need a simple lemma.

Lemma 2.4. $J_n^{-1}(t|\tau) = \tau_n J_n^{-1}(t|1)$.

Proof. By definition, $J_n^{-1}(t|\tau) = \inf\{x|J(x|\tau) \geq t\} = \inf\{x|\Pr(\tau_n(\delta_n - \theta) \leq x) \geq t\} = \inf\{x|\Pr((\delta_n - \theta) \leq x/\tau_n) \geq t\}$

$= \inf\{u\tau_n|\Pr((\delta_n - \theta) \leq u) \geq t\} = \tau_n J_n^{-1}(t|1)$. □

Corollary 2.3.1. *Assume that* $J_n(x|\tau) \to J(x)$ *and that* $J(x)$ *is continuous and monotone. Then*

$$\tau_n \propto \frac{1}{J_n^{-1}(t|1)}, \text{ for any } t \in (0,1) \text{ as } n \to \infty.$$

Proof. By assumption we have that $J_n^{-1}(t|\tau) \to J^{-1}(t)$ for each $t \in (0,1)$ from Theorem 2.3. Combining this with Lemma 2.4 we see that

$$\tau_n J_n^{-1}(t|1) \to J^{-1}(t) \text{ for any } t \in (0,1).$$

We can rewrite this as

$$\tau_n = \frac{J^{-1}(t)}{J_n^{-1}(t|1)} + o(1), \text{ for each t} \in (0,1), \qquad (2.1)$$

and the result follows. $\square$

It will be advantageous at this time to assume a specific form for τ_n. In many practical situations we will have $\tau_n = n^\beta$ for some $\beta > 0$, and we henceforth adopt this assumption.

Now suppose that there exists $t \in (0,1)$ for which $J^{-1}(t) > 0$. One may then take logarithms in (2.1) to obtain

$$\beta \log n = \log(J^{-1}(t)) - \log(J_n^{-1}(t|1)) + o(1).$$

If one takes positive integers $m > n$ sufficiently large, one may solve for β to obtain

$$\beta = \frac{\log(J_n^{-1}(t|1)) - \log(J_m^{-1}(t|1))}{\log m - \log n} + o(1).$$

We can generalize this result somewhat to eliminate the condition that one must find a particular t for which $J^{-1}(t) > 0$.

Theorem 2.5. *Assume that $J_n(x|\tau) \to J(x)$ and that $J(x)$ is continuous and monotone. Further suppose that $\tau_n = n^\beta$ for some $\beta > 0$. Let $0 < t_1 < \ldots < t_k < 1$ and define some linear function $u(t_1, \ldots, t_k) = \sum \alpha_i J^{-1}(t_i)$ with the α_i chosen so that $u(t_1, \ldots, t_k) > 0$. Analogously, let $u_n(t_1, \ldots, t_k|\tau, \theta) =$*

$\sum \alpha_i J_n^{-1}(t_i|\tau, \theta)$ *and let* $m > n$. *Then*

$$\beta = \frac{\log(u_n(t_1, ..., t_k|1)) - \log(u_m(t_1, ..., t_k|1)}{\log m - \log n} + o(1) \text{ as } m, n \to \infty.$$

Proof. The result follows immediately from Corollary 2.3.1 coupled with linearity. □

There are many specializations of the functions u which are of value. For instance, $u(t_1) = \pm J^{-1}(t_1) \neq 0$, depending on the polarity of $J^{-1}(t_1)$ is the most simple function, while $u(t_1, t_2) = J^{-1}(t_2) - J^{-1}(t_1)$ will always be positive if $t_1 < t_2$, assuming that $J(x)$ is monotone.

At the risk of generalizing for no apparent purpose we can extend the above to into a regression-like scenario. Although it seems rather pointless it will be of value in the next chapter.

Corollary 2.5.1. *(Least Squares) Assume everything as in the previous Theorem. Let* $b_1, ..., b_r$ *be a collection of distinct positive integers exceeding some fixed* N. *Set* $\overline{\log} = \frac{1}{r}\sum \log(b_i)$ *and let* $\overline{\log_u} = \frac{1}{r}\sum \log(u_{b_i}(t_1, ..., t_k|1))$. *Then*

$$\beta = -\frac{\sum(\log b_i - \overline{\log})(\log u_{b_i}(t_1, ..., t_k|1) - \overline{\log_u})}{\sum(\log b_i - \overline{\log})^2} + o(1) \text{ as } N \to \infty.$$

This is a standard type of result in numerical analysis. See, for example, Rudin(1976) for details.

Remark 2.2.1. *While we have chosen the functions* $u()$ *to be positive finite linear combinations of quantiles of* J^{-1}, *one can also define more exotic functions provided that the analogously defined functions* $u_n()$ *converge to* $u()$. *We* *give 2 examples. First, consider* $u(J^{-1}(t)) = \int J^{-1}(t)dt$*. By a simple change of variable we find in this case that* $u(J^{-1}(t)) = E[J(x)]$*. An application of Theorem 2.4 yields*

$$\beta = \frac{\log(E[J_n(x|1)] - \log(E[J_m(x|1)]}{\log m - \log n} + o(1). \qquad (2.2)$$

Similarly, if one defines $u(J^{-1}(t)) = \int [J^{-1}(t)]^2 dt$*, then one sees that this is the second moment of* $J(x)$*, which we denote by* $\mu_2(J(x))$*. We can apply the same methodology as above to obtain*

$$2\beta = \frac{\log(\mu_2(J_n(x|1))) - \log(\mu_2(J_m(x|1)))}{\log m - \log n} + o(1). \qquad (2.3)$$

The least squares versions are defined analogously.

2.3 Examples

Consider first the case where $X \backsim N(0,1)$, and let $\overline{X_n}^2$ be a consistent estimator of 0. It is known that $E[\overline{X_n}^2] = 1/n$. We assume that the limiting distribution $J(x)$ has a positive mean. By using the estimator in the preceding remark, that is selecting the function $u()$ to implicitly define $E[J(x)]$, we find for any choices $m > n$ that $\beta = 1 + o(1)$, so that $\beta = 1$ identically. We point out that this is less than surprising since one can determine by elementary means that $n(\overline{X_n}^2 - 0) \backsim \chi^2_{(1)}$.

One may similarly use the Least Squares analog to obtain the identical result independent of the choices b_i.

Now consider the case where $X \backsim N(2,1)$ and $\overline{X_n}^2$ is a consistent estimator of 4. For this case we will use the second formula from the preceding remark. One finds, by elementary means that $\mu_2(J_n(x|1)) = 16/n + 3/n^2$. By plugging this into the formula (2.3) we have

$$2\beta \simeq 1 + \frac{\log(1+\frac{3}{16n}) - \log(1+\frac{3}{16m})}{\log m - \log n}, \text{ for } m > n,$$

so that $\beta \to 1/2$.

Chapter 3

Stochastic Estimation Of Rates Of Convergence

3.1 Bootstrap and Subsampling

As we have seen one typically approximates the error distribution $J_n(x|\tau)$ by appealing to the limit law $J(x)$. There are however other approaches which can be used to approximate this quantity. The bootstrap estimate for $J_n(x|\tau)$ based on a sample $X_{1,...,}X_n$ is given by

$$H_n^{boot}(x) = \frac{1}{n^n} \sum 1\{\tau_n(\theta_a^* - \widehat{\theta}) \leq x\}$$

where $\widehat{\theta} = \widehat{\theta}(X_{1,...,}X_n)$ is the value of the estimator taken over the entire original sample, θ_a^* is the value of the estimator taken over a multiset of size n (that is, a sample of size n drawn with replacement from the original sample), and the sum is taken over all possible n^n multisets of the data. The bootstrap estimate is said to work if $H_n^{boot}(x)$ converges to the limit law $J(x)$ (actually this is not entirely correct - usually the supremum norm metric is used, but there is no great loss in our restriction). While in practice the bootstrap often works quite well, (that is, at least as well as the usual approach of appealing to the limit law, and sometimes better), it is rather difficult to analyze, with lots of funky analysis coming into play. Moreover, it is not a universally consistent estimator of $J_n(x|\tau)$.

Example 3.1. The extreme order statistic:

We pick up again with the previous example and let $X_{(n)}$ *be the estimator for* θ *when the observations are* $U(0,\theta)$. *Our bootstrap estimate will be* $H_n^{boot}(x) = \frac{1}{n^n}\sum 1[n(\widehat{\theta} - \theta_i^*) \leq x]$. *In particular,* $H_n^{boot}(0) = \frac{1}{n^n}$*(the number of boostrap samples that* $X_{(n)}$ *appears in)* $= \frac{n^n-(n-1)^n}{n^n} = 1-(1-\frac{1}{n})^n \to 1-e^{-1}$. *On the other hand,* $J(x) = 1 - e^{\frac{-x}{\theta}} \Rightarrow J(0) = 0$, *and it follows that* $H_n^{boot}(0) \nrightarrow J(0)$ *so that the bootstrap is not consistent in this case.*

To remedy the problem of consistency one may use Subsampling. On the face of things Subsampling looks rather like Bootstrapping with the Subsampling approximation to $J_n(x|\tau)$ is given by

$$L_{n,b}(x|\tau) = \frac{1}{\binom{n}{b}}\sum 1\{\tau_b(\widehat{\theta}_{n,b,a} - \widehat{\theta}_n) \leq x\}$$

where $\widehat{\theta}_n = \widehat{\theta}(X_1, ..., X_n)$ is the value of the statistic taken over the entire sample, $\widehat{\theta}_{n,b,a}$ is the value of the statistic taken over a subset a of size $b < n$

of distinct elements from the original sample, the sum being taken over all subsets of size b.We note that there is some ambiguity here. On the one hand we are claiming that $L_{n,b}(x)$ is an approximation to $J_n(x|\tau, \theta)$, but really it is more naturally viewed as an approximation to $J_b(x|\tau, \theta)$.

Example 3.2. *The subsampling analog to the previous example is*

$$L_{n,b}(x|\tau) = \frac{1}{\binom{n}{b}} \sum 1\{n(X_{(n)} - \widehat{\theta}_{n,b,a}) \leq x\}$$

so that,

$$L_{n,b}(0|\tau) = \frac{\binom{n-1}{b-1}}{\binom{n}{b}} = \frac{b}{n} \to 0, \text{ for fixed } b \text{ as } n \to \infty,$$

and the particular difficulty encountered in the bootstrap is alleviated. The fact that subsampling works is given by the following result of Politis et al. (1999).

Theorem 3.1. *Assume that $J_b(x|\tau)$ converges weakly to a nondegenerate $J(x)$. Moreover suppose that $\frac{\tau_b}{\tau_n} \to 0$ and $\frac{b}{n} \to 0$, as b and n tend to ∞. Then*

- $.L_{n,b}(x) \xrightarrow{P} J_b(x)$ *as* $n \to \infty$
- *If x is a continuity point of J then $L_{n,b}(x) \xrightarrow{P} J(x)$.*

Proof. Let $U_{n,b}(x|\tau, \theta) = \frac{1}{\binom{n}{b}} \sum 1\{\tau_b(\widehat{\theta}_{n,b,a} - \theta) \leq x\}$. Observe first that this is an idealized version of $L_{n,b}(x|\tau)$. In addition we see that it is also a $U-statistic$ of degree b with kernel $h(X_{i_1}, ..., X_{i_b}) = I\{\tau_b(\widehat{\theta}_{n,b,a}(X_{i_1}, ..., X_{i_b}) -$

$\theta) \leq x\}$,whose expectation is $J_b(x|\tau,\theta)$, and which takes values in the interval $[0,1]$. By applying Hoeffding's Theorem (see Serfling(1981))we have that $P(U_{n,b} - J_b(x) \geqslant t) \leq \exp\{-2\lfloor\frac{n}{b}\rfloor t^2\}$. If one applies the theorem to $-U_{n,b}(x)$ one obtains a similar result implying that $P\{|U_{n,b} - J_b(x)| \geqslant t\} \leq \exp\{-2\lfloor\frac{n}{b}\rfloor t^2\}$,so that $U_{n,b}(x) \xrightarrow{P} J_b(x)$. One may now write the estimator $L_{n,b}(x|\tau)$ as $\frac{1}{\binom{n}{b}}\sum 1\{\tau_b(\widehat{\theta}_{n,b,a} - \theta) + \tau_b(\theta - \widehat{\theta}_n) \leq x\}$. If x is a continuity point of $J_b(x|\tau,\theta)$ then for every $\epsilon > 0$

$$U_{n,b}(x-\epsilon)1(E_n) \leq L_{n,b}(x|\tau)1(E_n) \leq U_{n,b}(x+\epsilon)1(E_n)$$

where E_n is the event $\{\tau_b(\widehat{\theta}_n - \theta) \leq \epsilon\}$. Now this event has probability tending toward 1 as $n \to \infty$ if $\frac{\tau_b}{\tau_n} \to 0$. Thus with probability tending to 1

$$U_{n,b}(x-\epsilon) \leq L_{n,b}(x|\tau) \leq U_{n,b}(x+\epsilon) \text{ for every } \epsilon > 0 \text{ as } n \to \infty,$$

so that $L_{n,b}(x|\tau)$ converges to $u_{n,b}(x)$ and thus to $J_b(x)$ in probability.

The second assertion follows immediately by definition of convergence in distribution. □

We will need analogous results for the inverse distribution functions.

Corollary 3.1.1. *Let $L_{n,b}^{-1}(t|\tau)$ be the inverse distribution function of $L_{n,b}(x|\tau)$. Then $L_{n,b}^{-1}(t|\tau)$ converges in probability to $J_b^{-1}(t|\tau)$ and thus to $J^{-1}(t)$.*

The proof is uninteresting and can be found in Bertail *et al.*(1999) or Politis *et al.* (1999).

Remark 3.1.1. *There is a rather large practical detail to be worked out. The quantity $\binom{n}{b}$ can become enormous very quickly, making a complete enumeration impossible. Fortunately, one does not need to do a complete enumeration to know which way the wind is blowing. One may use a stochastic approximation for $L_n(x|\tau)$ by selecting some sufficiently large number B of random subsets and considering the quantity*

$$L_{n,b}(x|\tau) = \frac{1}{B}\sum 1\{\tau_b(\widehat{\theta}_{n,b,a} - \widehat{\theta}_n) \leq x\}.$$

3.2 Constructing Convergence Rate Estimators

Having determined a consistent estimator for error distributions one may adapt the methods of the previous chapter to cook up a collection of estimators. The approach is simply to use the functions $L_{n,b}^{-1}(t|1)$ in place of $J_b^{-1}(t|1)$ and hope that things work out.

Theorem 3.2. *(Nondeterministic analog of Corollary 2.5.1) Assume that $J_n(x|\tau,\theta) \to J(x)$ and that $J(x)$ is continuous and monotone. Let $X_1, ..., X_n$ be an i.i.d. sample. Let $L_{n,b}(x|\tau)$ and $L_{n,b}^{-1}(t|\tau)$ be the subsampling approximations to $J_b(x|\tau)$ and $J_b^{-1}(t|\tau)$ respectively. Let $1 > \gamma_1 > ... > \gamma_r > 0$, and set $b_i = cn^{\gamma_i}$ for some positive constant c. Let $\overline{\log} = \frac{1}{r}\sum \log(b_i)$. Let $0 < t_1 < ... < t_k < 1$, and define $\widehat{u}_{b_i}(t_1, ..., t_k|1) = \sum \alpha_i L_{n,b_i}^{-1}(t_i|1)$.Let*

$\overline{\log_u} = \frac{1}{r}\sum \log(\widehat{u}_{b_i}(t_1, ..., t_k|1))$. *Then*

$$\widehat{\beta} = -\frac{\sum(\log(b_i) - \overline{\log})(\log(\widehat{u}_{b_i}(t_1, ..., t_k|1)) - \overline{\log_u})}{\sum(\log(b_i) - \overline{\log})^2} = \beta + o_P(\frac{1}{\log n}).$$

The proof appears in Politis *et al.* (1999).

Some comments are in order here. First we note that block sizes b_i are constructed to ensure that the requirements of Theorem 3.1 are met. Secondly, we have a potential difficulty in that the functions $\widehat{u}_{b_i}$ may take on negative values. It should be clear that if n is taken large enough that this problem will disappear. However, one is usually faced with the cumbersome detail of having a tragically small sample size. One can work around this in a couple of ways. First one can use a judicious choice for the function, such as a difference of quantiles. Alternately, one may take absolute values of either $\widehat{u}_{b_i}$ or of its individual components. In any case, there is room for discretion.

Chapter 4

Simulations

4.1 Design of Simulation and Results

Simulations were done on a Thinkpad T20 running the Solaris 9 operating system. The programs were written using the native Sun C compiler. The sundry mathematical functions used, for example random number generation, covariance calculation, and the like were from the GNU Scientific Library. Stress tests included with the GSL were done to ensure the Library's fitness.

We attempted to replicate the simulation results of Bertail *et al.*(1999). To this end we describe the simulations done in their paper and dutifully reconstructed here.

Two situations were examined. In the first instance data was generated according to a Normal Distribution having mean 0 and variance 1. We considered the estimator $\delta_n = \overline{X_n}^2$for the quantity $E[X]^2$. It is known that $J_n(x|\tau,\theta) = \Pr(n(\delta_n - 0) \leq x)$ converges to a nondegenerate limit law (in fact it is exactly distributed as a chi-squared random variable with one degree of freedom). Consequently the actual value of β is 1 in this case. The number of blocks, r, used was either 3 or 20, and the block sizes were given by the formula $b_i = [n^{\frac{1}{2}\{1+[\frac{i}{r+1}]\}}]$, where n denotes the sample size, and [] denotes the nearest integer function. For example, if $r = 3$ and $n = 100$ the block sizes were 18,32,56. Approximations to the functions $L_{n,b}(x|\tau)$ were generated by using a fixed number of 3000 points.

The number of elements in the sample were taken to be 100, 1000, and 10000. For each sample size a collection of 11 different functions $u(t_1, ..t_k)$ were used to estimate β. Ostensibly the idea was to determine what sort of function $u()$ worked best. The functions $u()$ considered were given as follows.

- $u_1 : u(t) = J^{-1}(.99)$
- $u_2 : u(t) = J^{-1}(.95)$
- $u_3 : u(t) = J^{-1}(.75)$
- $u_4 : u(t_1, ..., t_{15}) = \frac{1}{15} \sum J^{-1}(.75 + \frac{4k}{300})$, $k = 0..14$
- $u_5 : u(t_1, ..., t_{30}) = \frac{1}{30} \sum J^{-1}(.75 + \frac{2k}{300})$, $k = 0..29$
- $u_6 : u(t_1, ..., t_{600}) = \frac{1}{600} \sum J^{-1}(.75 + \frac{k}{3000})$, $k = 0..599$

- $u_7 : u(t_1, t_2) = J^{-1}(.99) - J^{-1}(.01)$
- $u_8 : u(t_1, t_2) = J^{-1}(.95) - J^{-1}(.05)$
- $u_9 : u(t_1, t_2) = J^{-1}(.75) - J^{-1}(.25)$
- $u_{10} : u(t_1, ..., t_{10}) = \frac{1}{10} \sum \{J^{-1}(.75 + \frac{k}{40}) - J^{-1}(.01 + \frac{k}{40})\}$ $k = 0..9$
- $u_{11} : u(t_1, ..., t_{600}) = \frac{1}{600} \sum \{J^{-1}(.75 + \frac{k}{2400}) - J^{-1}(.01 + \frac{k}{2400})\}$ $k = 0..599$

sample data generated were Normal with mean 2 and variance 1. In this case $J_n(x|\tau, \theta) = \Pr(n^{\frac{1}{2}}(\delta_n - 4) \leq x)$ converges to the nondegenerate law whose distribution is asymptotically Normal(0,16), a consequence of the Delta Method (see van der Vaart(1998) for details). Hence in this case $\beta = \frac{1}{2}$.

We break down the simulation algorithm as follows.

- Generate N random variables from a Normal distribution.
- Evaluate the test statistic, $\overline{X_N}^2$, based on the generated data.
- For each block size b_i, take 3000 subsamples of size b_i,and evaluate the test statistic on each subsample, and store this data in a vector.
- Sort each vector(corresponding a particular b_i) of data points in increasing order. Observe that the t^{th} quantile is just the $(3000t)^{th}$ value in our sorted arrays.

- Evaluate the estimate $\widehat{\beta}$ based on each of the u functions above
- Iterate this procedure 1000 times, and provide summary statistics for the estimates.

The results of simulations are summarized in Tables 4.1-4.4.

4.2 Discussion

Our results do not mirror those of Bertail *et al.* (1999).In every instance our estimates tended to overstate the actual value of β while the original published results only overstated in the case where $E[X] = 2$, and even then their results showed marked improvement as the sample size N grew. On the other hand, in the case where $E[X] = 0$ their results tended to understate the actual value of β.

To explain the disparity between our results and those of BPR we need to understand the possible sources of error in the estimator.

- The first is the intrinsic error of the approach. In the idealized situation of Chapter 2 the estimate of β was realized by passing to a limit. In the stochastic version one is bound generally by the amount of available data and specifically by the block sizes used. In the smallest case considered $N = 100$, the block sizes b_i were $18, 32, 56$. While this might explain some of the difficulty in the small sample behavior we note that

Table 4.1: A comparison of simulation results found in Bertail *et al.*(1999)

	B=3 E[X] = 2					
	N=100		N=1000		N=10000	
u	BPR	US	BPR	US	BPR	US
u_1	$.567 \pm .052$	$.830 \pm .040$	$.526 \pm .032$	$.660 \pm .152$	$.509 \pm .022$	$.586 \pm .020$
u_2	$.552 \pm .042$	$.817 \pm .032$	$.520 \pm .024$	$.655 \pm .019$	$.507 \pm .018$	$.582 \pm .015$
u_3	$.546 \pm .042$	$.792 \pm .048$	$.516 \pm .025$	$.642 \pm .030$	$.505 \pm .018$	$.580 \pm .022$
u_4	$.516 \pm .042$	$.802 \pm .032$	$.503 \pm .015$	$.648 \pm .020$	$.499 \pm .011$	$.580 \pm .014$
u_5	$.541 \pm .044$	$.803 \pm .099$	$.514 \pm .025$	$.648 \pm .019$	$.505 \pm .018$	$.580 \pm .014$
u_6	$.551 \pm .042$	$.803 \pm .031$	$.513 \pm .025$	$.648 \pm .019$	$.505 \pm .019$	$.580 \pm .014$
u_7	$.517 \pm .023$	$.771 \pm .026$	$.504 \pm .021$	$.637 \pm .017$	$.499 \pm .016$	$.576 \pm .012$
u_8	$.516 \pm .025$	$.773 \pm .020$	$.504 \pm .016$	$.637 \pm .013$	$.499 \pm .012$	$.575 \pm .010$
u_9	$.518 \pm .032$	$.773 \pm .027$	$.504 \pm .022$	$.636 \pm .017$	$.499 \pm .016$	$.576 \pm .013$
u_{10}	$.517 \pm .023$	$.768 \pm .017$	$.504 \pm .015$	$.631 \pm .011$	$.499 \pm .011$	$.575 \pm .008$
u_{11}	$.517 \pm .023$	$.766 \pm .017$	$.504 \pm .015$	$.633 \pm .012$	$.499 \pm .011$	$.574 \pm .009$

BPR: The results given in Bertail *et al.*(1999)

US: Our simulation results

Table 4.2: A comparison of simulation results found in Bertail *et al.*(1999)

B=3, E[X]=0						
	N=100		N=1000		N=10000	
u	BPR	US	BPR	US	BPR	US
u_1	.933 ± .109	1.34 ± .153	.932 ± .083	1.16 ± .102	.943 ± .062	1.08 ± .071
u_2	.928 ± .107	1.32 ± .169	.924 ± .085	1.15 ± .113	.941 ± .063	1.07 ± .076
u_3	.927 ± .113	1.33 ± .208	.924 ± .088	1.17 ± .126	.942 ± .066	1.10 ± .071
u_4	.931 ± .111	1.32 ± .189	.928 ± .087	1.15 ± .121	.943 ± .064	1.08 ± .076
u_5	.932 ± .114	1.32 ± .188	.929 ± .088	1.15 ± .121	.950 ± .061	1.08 ± .076
u_6	.928 ± .119	1.32 ± .187	.912 ± .089	1.15 ± .121	.947 ± .064	1.08 ± .076
u_7	.908 ± .128	1.28 ± .202	.912 ± .098	1.13 ± .130	.931 ± .073	1.06 ± .086
u_8	.889 ± .139	1.24 ± .235	.895 ± .111	1.10 ± .154	.921 ± .082	1.05 ± .101
u_9	.852 ± .176	1.16 ± .294	.869 ± .143	1.06 ± .200	.908 ± .103	1.03 ± .130
u_{10}	.871 ± .156	1.21 ± .262	.882 ± .125	1.08 ± .173	.915 ± .091	1.04 ± .111
u_{11}	.864 ± .167	1.19 ± .273	.873 ± .136	1.08 ± .180	.912 ± .095	1.04 ± .115

BPR: The results given in Bertail *et al.*(1999)

US: Our simulation results

Table 4.3: A comparison of simulation results found in Bertail *et al.*(1999)

B=20, E[X]=0						
	N=100		N=1000		N=10000	
u	BPR	US	BPR	US	BPR	US
u_1	$.927 \pm .100$	$1.49 \pm .166$	$.934 \pm .084$	$1.26 \pm .109$	$.944 \pm .054$	$1.15 \pm .075$
u_2	$.909 \pm .097$	$1.46 \pm .189$	$.931 \pm .083$	$1.24 \pm .125$	$.941 \pm .060$	$1.14 \pm .084$
u_3	$.910 \pm .101$	$1.46 \pm .233$	$.927 \pm .085$	$1.26 \pm .139$	$.957 \pm .056$	$1.16 \pm .087$
u_4	$.900 \pm .100$	$1.45 \pm .212$	$.935 \pm .083$	$1.24 \pm .135$	$.948 \pm .056$	$1.15 \pm .088$
u_5	$.902 \pm .103$	$1.45 \pm .212$	$.927 \pm .084$	$1.24 \pm .134$	$.941 \pm .060$	$1.15 \pm .088$
u_6	$.906 \pm .105$	$1.45 \pm .211$	$.923 \pm .084$	$1.24 \pm .134$	$.933 \pm .062$	$1.15 \pm .088$
u_7	$.881 \pm .106$	$1.41 \pm .228$	$.905 \pm .108$	$1.21 \pm .146$	$.931 \pm .068$	$1.12 \pm .098$
u_8	$.858 \pm .130$	$1.35 \pm .261$	$.871 \pm .129$	$1.18 \pm .171$	$.920 \pm .081$	$1.10 \pm .115$
u_9	$.820 \pm .170$	$1.27 \pm .307$	$.858 \pm .143$	$1.13 \pm .207$	$.904 \pm .114$	$1.07 \pm .139$
u_{10}	$.840 \pm .149$	$1.31 \pm .286$	$.875 \pm .123$	$1.16 \pm .189$	$.913 \pm .090$	$1.09 \pm .127$
u_{11}	$.835 \pm .154$	$1.30 \pm .295$	$.872 \pm .127$	$1.15 \pm .195$	$.905 \pm .103$	$1.08 \pm .131$

BPR: The results given in Bertail *et al.*(1999)

US: Our simulation results

Table 4.4: A comparison of simulation results found in Bertail *et al.*(1999)

B=20, E[X]=2						
	N=100		N=1000		N=10000	
u	BPR	US	BPR	US	BPR	US
u_1	.549 ± .018	.940 ± .020	.510 ± .011	.731 ± .008	.501 ± .006	,638 ± .006
u_2	.537 ± .017	.924 ± .013	.509 ± .009	.724 ± .006	.501 ± .005	.635 ± .005
u_3	.534 ± .018	.898 ± .015	.504 ± .009	.713 ± .009	.501 ± .006	.630 ± .008
u_4	.519 ± .022	.910 ± .009	.502 ± .009	.718 ± .006	.500 ± .005	.632 ± .005
u_5	.527 ± .025	.910 ± .009	.503 ± .009	.718 ± .006	.501 ± .006	.632 ± .005
u_6	.534 ± .024	.911 ± .009	.513 ± .009	.718 ± .006	.505 ± .006	.632 ± .005
u_7	.507 ± .008	.879 ± .008	.502 ± .007	.706 ± .005	.500 ± .005	.627 ± .004
u_8	.507 ± .007	.879 ± .006	.502 ± .007	.706 ± .004	.500 ± .004	.628 ± .003
u_9	.508 ± .009	.880 ± .008	.502 ± .007	.706 ± .006	.500 ± .004	.627 ± .004
u_{10}	.507 ± .006	.874 ± .005	.502 ± .004	.703 ± .003	.500 ± .003	.626 ± .002
u_{11}	.507 ± .007	.872 ± .005	.502 ± .004	.703 ± .003	.500 ± .003	.626 ± .002

BPR: The results given in Bertail *et al.*(1999)

US: Our simulation results

the results when $N = 10000$, while in the appropriate neighborhood, are still not entirely satisfactory.

- The random error, or noise inherent in the random data. While one can not control this we note that this is likely a rather small component of the problem since there is very little variability in our estimates. Moreover with a sample size of 10000 it appears unlikely that this should be a significant issue.

- The last possibility is that the estimator $L_{n,b}(x|1)$ for $J_b(x|1)$ may be at fault. We shall see that this is precisely where the problem lies.

We will examine the 2 cases corresponding to the different means of the generated data.

Case 1, $X \sim N(2,1)$

If $X \sim N(0,1)$ then $bJ_b(x|1)$ is distributed as a χ_1^2 random variable. By appealing to Theorem we find that $\widehat{\beta} = 1$ identically for any choices t_i and any pair m, n of positive integers. Consequently any problem with the estimator will be either due to the estimates $L_{n,b}(x|1)$ or random error. We can examine the behavior of $L_{n,b}(x|1)$ by equivalently looking at the values of $L_{n,b}^{-1}(t|1)$ for various values of b, and t and fixed n, chosen for our example to be 100. We obtain approximate values by running 100 simulations and taking the mean of these values given in Table 4.5.

Table 4.5: A comparison of approximation of $F^{-1}(t)$ by simulation

$F^{-1}(t)$	$t = .01$	$t = .05$	$t = .5$	$t = .95$	$t = .99$
$\chi_1^{2(inv)}(t)$	.0002	.004	.455	3.84	6.64
$L_{100,2}^{-1}(t\|1)$	−.02	−.02	.45	3.85	6.61
$L_{100,3}^{-1}(t\|1)$	−.03	−.03	.44	3.84	6.57
$L_{100,4}^{-1}(t\|1)$	−.04	−.04	.43	3.82	6.55
$L_{100,16}^{-1}(t\|1)$	−.17	−.16	.31	3.63	6.21
$L_{100,17}^{-1}(t\|1)$	−.18	−.17	.31	3.61	6.18
$L_{100,18}^{-1}(t\|1)$	−.18	−.18	.30	3.59	6.14
$L_{100,19}^{-1}(t\|1)$	−.20	−.19	.29	3.58	6.11
$L_{100,32}^{-1}(t\|1)$	−.32	−.31	.21	3.30	5.60
$L_{100,56}^{-1}(t\|1)$	−.55	−.52	.10	2.59	4.29

The result is apparent. Even with only 100 points we see a reasonably good fit between the actual inverse chi-squared distribution and the estimated versions, $bL_{n,b}^{-1}(t|1)$, for $b = 2, 3, 4$ although even at this point we begin to spot a trend. The quantile estimates are systematically understated as the block size increases. While this exhibits the difficulty it is instructive to view the problem in a way that quantifies how the estimate $\widehat{\beta}$ begins to blow up as the block size increases.

Rather than choose some simple function $u()$ of quantiles we look instead at an implicitly defined $u()$ on $J_b^{-1}(t|1)$ which corresponds to the expected value $E[J_b(x|1)]$. We find by elementary means that $E[L_{n,b}(x|1)] = 1/b - 1/n$. We can compute the exact values of our estimators based on these values. Table 4.6 illustrates the results.

Case 2, $X \sim N(2, 1)$

This case is significantly tougher to analyze than the previous since we don't have a nice closed form for the distributions $J_b(x|1)$, although we do know that $\sqrt{b}J_b(x|1)$ is asymptotically $N(0, 16)$.Unlike the previous example we can't appeal to taking the function $u()$ as implicitly defining expectation since the limiting law has mean 0 and our method only works when $u()$ is a positive function of $J(x)$. We can however consider defining $u()$ so that its action on $J(x)$ gives the second moment of $J(x)$. In this way we obtain that

Table 4.6: A comparison of simulation approximation for $\hat{\beta}$ by the first moment

Case	$\widehat{\beta}$ by 1^{st} Moment of $L_{n,b}(x\|1)$	$\widehat{\beta}$ by 1^{st} Moment of $J_b(x\|1)$
$N = 100, B = 3$	1.547	1.0
$N = 100, B = 20$	1.759	1.0
$N = 1000, B = 3$	1.272	1.0
$N = 1000, B = 20$	1.412	1.0
$N = 10000, B = 3$	1.15	1.0
$N = 10000, B = 20$	1.254	1.0

Table 4.7: A comparison of simulation approximation for $\hat{\beta}$ by the second moment

Case	$\widehat{\beta}$: 2^{nd} Moments of $L_{n,b}(x\|1)$	$\widehat{\beta}$: 2^{nd} Moments of $J_b(x\|1)$
$N = 100, B = 3$	.776	.503
$N = 100, B = 20$	.882	.503
$N = 1000, B = 3$	.637	.501
$N = 1000, B = 20$	.706	.501
$N = 10000, B = 3$	.576	.500
$N = 10000, B = 20$	.627	.500

the second moment of $J_b(x|1)$ is given by $16/b + 3/b^2$. On the other hand one can compute the second moment of $L_{n,b}(x|1)$ via the theory of symmetric functions to give $(16/b + 3/b^2) + (-16/n - 1/n^2 - 2/bn + 112/bn^2)$. We summarize the results in Table 4.7.

For the sake of comparison consider what happens when one evaluates $\widehat{\beta}$ based on the 2^{nd} moment of $L_{n,b}(x|1)$ when one selects $B = 2$, based on block sizes $\{2, 3\}$. Even with only 100 points one arrives at an estimate $\widehat{\beta} = .548$. For 1000 or more points one stabilizes at the value .536, which is the value of the estimator $\widehat{\beta}$ evaluated on the 2^{nd} moments of $J_b(x|1)$.

This is somewhat surprising for the simple reason that even though one would think that the smallest block size used should be at least a little larger than 2, we find that the estimator based on the 2 smallest block sizes yields the best result.

4.3 Conclusion

It should be clear from the examples that the approximations $L_{n,b}(x|1)$, although consistent for $J_b(x|1)$, can behave quite poorly even if $b << n$, so that their use in estimating the rate of convergence must be a cautious one. This fact was apparently known by Bertail *et al.*(1999) so it is curious that they discount this behavior in both their choices of block sizes and the number of blocks used. Indeed, the idea that one should use many blocks with increasing block sizes is specious precisely because the errors in the approximations greatly exceed any benefit one might hope to obtain by removing "noise" in the regression equations. In both cases examined one saw that the summary statistics corresponding to the first and second moments depended on both the sample size n as well as the block size b used.

In the end one has to be aware that the family of estimators given by Bertail et. al (1999) are not of the "plug and chug" variety where one indiscriminately selects arbitrary block sizes and functions $u()$. In the examples given it was possible to obtain reasonable estimates with just 2 blocks with smallest possible size. Although we doubt that this will be uniformly applicable it seems a more prudent initial first step than blindly bashing away and coming up with a dubious result. It should be stated, somewhat as an afterthought, that among the functions $u()$ considered in the simulations that the function u_9, given by $u(t_1, t_2) = J^{-1}(.75) - J^{-1}(.25)$, performed the best in our simulations. However we can't conceive why this would be a universally good choice for any instance of the problem.

Bibliography

[1] Bertail, P., Politis, D.N., Romano, J.P. (1999) On Subsampling Estimators With Unknown Rate of Convergence. *Journal of the American Statistical Association* **94**, 569-579.

[2] Lehmann, E.L.(1998) *Elements of Large Sample Theory.* New York: Springer Verlag, 1998

[3] Politis, D.N, Romano, J.P., Wolf, M.(1999) *Subsampling.* New York: Springer Verlag

[4] Rudin, W. (1976) *Principles of Mathematical Analysis.* New York: McGraw-Hill

[5] Serfling, R.J., (1980) *Approximation Theorems Of Mathematical Statistics.* Toronto: John Wiley & Sons

[6] van der Vaart, A.W. (1998) *Asymptotic Statistics.* New York: Cambridge University Press.