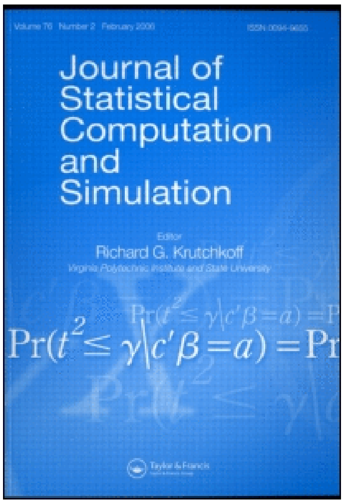
This article was downloaded by: *[University of California, Los Angeles]*
On: *10 October 2010*
Access details: *Access Details: [subscription number 918974596]*
Publisher *Taylor & Francis*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of Statistical Computation and Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713650378

A bootstrap for point processes

W. J. Branu[a]; R. J. Kulperger[b]
[a] Department of Mathematics and Statistics, The University of Winnipeg, Winnipeg, Canada [b] Department of Statistical and Actuarial Sciences, The University of Western Ontario, London, Canada

To cite this Article Branu, W. J. and Kulperger, R. J.(1998) 'A bootstrap for point processes', Journal of Statistical Computation and Simulation, 60: 2, 129 — 155
To link to this Article: DOI: 10.1080/00949659808811878
URL: http://dx.doi.org/10.1080/00949659808811878

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

J. Statist. Comput. Simul., 1998, Vol. 60, pp.129–155
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association)
Amsterdam B.V. Published under license
under the Gordon and Breach Science
Publishers imprint.
Printed in India.

Downloaded By: [University of California, Los Angeles] At: 17:34 10 October 2010

A BOOTSTRAP FOR POINT PROCESSES

W. J. BRAUN[a] and R. J. KULPERGER[b]

[a]*Department of Mathematics and Statistics, The University of Winnipeg, 515 Portage Avenue, Winnipeg, Canada R3B 2E9;* [b]*Department of Statistical and Actuarial Sciences, The University of Western Ontario, London, Canada, N6A 5B9*

(Received 25 September 1995; In final form 17 August 1997)

A nonparametric bootstrap for stationary point processes with weak dependence is proposed. Two variations of the bootstrap are described. The methods are block resampling plans which are similar in spirit to an algorithm which has been successfully applied to short-memory time series. The validity of the bootstrap for estimating first and second order intensity measures of certain Poisson cluster processes is investigated by simulation. The first order intensity bootstrap is proven to be asymptotically correct. The method is applied to some meteorological data.

Keywords: Point process; bootstrap; intensity estimate; weak dependence; Neyman-Scott process

1. INTRODUCTION

Point processes occur in a variety of contexts ranging from telecommunications networks to meteorology and ecology. Modelling point processes is a nontrivial exercise even in one dimension, unless the use of a Poisson model is justified. A number of ways of modelling point processes exhibiting dependence have been suggested (e.g. [6]), but goodness-of-fit questions do not seem to have been addressed yet. Thus, the danger of misapplying a model to given data is still very great.

Brillinger [1, 3] has investigated point processes using a model-free approach. His interest is in the estimation of product moment densities among other things, and his results depend on the assumption of weak

Downloaded By: [University of California, Los Angeles] At: 17:34 10 October 2010

dependence between distant increments. Our work here builds on Brillinger's. We propose a bootstrap method for point processes, and study its properties. In particular, we are concerned as to whether Brillinger's moment densities can be bootstrapped.

As pointed out above, the need for a nonparametric bootstrap for point processes is evident, since parametric point process models can be inadequate, or properties of relevant estimators may not be readily available. Furthermore, theoretical physicists are becoming interested in point processes or 'spike trains' resulting from time series. Their interest is primarily in determining whether a process is random or nonlinear. Therefore, methods for replicating point processes data are sought in order to nonparametrically perform this test. We will not concern ourselves with the question of randomness versus nonrandomness, but we hope our contribution is helpful in this direction. In addition, modelling network traffic is of increasing importance. The nonparametric bootstrap method described here may be of use in the inference of bursty processes which arise in this setting.

The paper by Possolo [11] on subsampling random fields, using an adaptation of a non-overlapping block-resampling scheme for time series (Carlstein [4]), represents the first attempt to resample a point process. A point process analogy to the overlapping block version [8] is proposed here and is applied to some 1-D Neyman-Scott processes.

A review of point process definitions is given in section 2, followed by the bootstrap procedure in section 3. The validity of the bootstrap for the first order intensity is demonstrated in section 6. A simulation is described in section 4, in which first and second order intensity estimators for the Neyman-Scott processes are studied. An illustrative example is given in section 5. The paper concludes with some extensions and open questions.

2. DEFINITIONS AND ASSUMPTIONS

In this section, we briefly review some standard definitions regarding point processes in R. More detail, as well as more generality, can be found in the books by Daley and Vere-Jones [6] and Reiss [12]. Here, we follow the latter reference.

Downloaded By: [University of California, Los Angeles] At: 17:34 10 October 2010

Let $\mathcal{C}$ denote the space of all countable subsets of R. Then, for any set $A \in \mathcal{C}$,

$$x_A(\cdot) = |A \cap .|$$

defines a measure on the Borel field $\mathcal{B}$ of R($|\cdot|$ denotes cardinality).

A σ-field $\mathcal{M}$ can be defined on the set M of all such measures, such that all projection functionals $\pi_B(\cdot)$ defined by

$$\pi_B(\mu_A) = \mu_A(B), \quad \text{for all} \quad B \in \mathcal{B}$$

are measurable, for each $A \in \mathcal{C}$.

A point process X is then defined as any measurable mapping from an appropriate probability space into M. Thus, X is a random measure, and $X(B)$ is the (random) cardinality of points in the Borel set B. A simple point process is a point process in which $\mathcal{C}$ consists only of those countable subsets of R in which all points are distinct.

The process X is said to be stationary if all of its marginal distributions are translation-invariant.

2.1. Parameters

Brillinger [1] describes several point process parameters that are of interest, both in the time and frequency domain. We will restrict our attention to two time domain parameters: the first and second order intensities.

The first order intensity or mean density is defined as

$$p_1(\tau) = \lim_{h \to 0} \frac{P(X((\tau, \tau + h]) > 0)}{h}$$

This corresponds to the rate at which points occur near τ. When X is stationary, $p_1(\tau) = \text{constant}$.

The second order intensity is defined as

$$p_2(\tau_1, \tau_2) = \lim_{h_1 \to 0, h_2 \to 0} \frac{P(X((\tau_1, \tau_1 + h_1]) > 0, X((\tau_2, \tau_2 + h_2]) > 0)}{h_1 h_2}.$$

This measure can be useful for detecting clustering among the points, for example. When X is stationary, $p_2(\tau_1, \tau_2)$ is a function of $\tau_2 - \tau_1$

Downloaded By: [University of California, Los Angeles] At: 17:34 10 October 2010

only. In this case, we slightly abuse notation, and denote the second order intensity by $p_2(\tau)$, where $\tau = |\tau_2 - \tau_1|$.

2.2. Inference Based on Asymptotic Normality

From now on, we assume X is a simple stationary point process which is observed on the interval $(0, T]$, and we follow Brillinger [1].

The following estimators have been suggested for p_1 and $p_2(\tau)$:

$$\hat{p}_1 = \frac{1}{T} X((0, T]) \tag{1}$$

and

$$\begin{aligned} \hat{p}_2(\tau) &= \frac{1}{hT} \int_0^{T-h} X((u+\tau, u+\tau+h])X(du) \\ &= \frac{1}{hT} \sum_{x_i} \#\{\text{points in } (x_i+\tau, x_i+\tau+h]\}. \end{aligned} \tag{2}$$

The parameter h is a window or bin width parameter. It plays the same role as the window parameter in density estimation, since $p_2(\tau)$ is a second moment density function.

Brillinger advocates the following modification to (2) when τ is large relative to T.

$$\hat{p}'_2(\tau) = \hat{p}_2(\tau) + \hat{p}_1^2 \tau h. \tag{3}$$

This would not be recommended in situations involving long-range dependence.

The estimator $\hat{p}_1$ has an asymptotically normal limit (as $T \to \infty$) under the assumption of weak dependence [1, 3]. Specifically, $\sqrt{T}(\hat{p}_1 - p_1)$ has a normal limit distribution with mean 0 and variance

$$\lim_{T\to\infty} \left\{ \frac{1}{T} \int_0^T \int_0^T p_2(t_1 - t_2)\, dt_1\, dt_2 + p_1 - \frac{p_1^2}{T} \right\} = \int_{-\infty}^{\infty} c(v)dv + p_1 \tag{4}$$

where $c(v)dxdv = \mathrm{Cov}(X(dx), X(dx+dv)) = p_2(v) - p_1^2$ for $v \neq 0$. Under similar assumptions, the estimator $\hat{p}_2(\tau)$ has an asymptotically

Downloaded By: [University of California, Los Angeles] At: 17:34 10 October 2010

normal distribution with mean $p_2(\tau)$ and asymptotic variance

$$p_2(\tau)/hT \tag{5}$$

as $hT \to \infty$ and $h \to 0$.

Thus, asymptotically correct confidence intervals for both p_1 and $p_2(\tau)$ can be set up, using normal quantiles in the usual way. However, it should be pointed out that estimation of the asymptotic variance $\hat{p}_1$ is a nontrivial exercise, and the asymptotic variance of $\hat{p}_2$ is not a good approximation to the true variance, even for fairly large T (see below). Therefore, we present an alternative method of obtaining the quantiles needed for constructing confidence intervals based on a bootstrap technique.

3. THE BLOCK-RESAMPLING ALGORITHM

Let x denote a sample realization of the process X, observed on the interval $(0, T]$. Note that x is a counting measure on R. Let A denote the set of points t such that $x(t) = 1$. The point process sample x can be bootstrapped as follows.

Take b to be some positive integer.

1. Generate uniform variates $U_1, U_2, \ldots, U_b$, independently on the interval $(0, T-T/b]$.
2. For $j = 1, 2, \ldots, b$,
   (a) set
   $$A_j = (U_j, U_j + T/b] \cap A.$$
   (b) set
   $$A_j^* = A_j - U_j + \frac{(j-1)T}{b}.$$
   (c) set
   $$X_j^*(\cdot) = |A_j^* \cap .|$$
3. Set
$$X^* = \sum_{j=1}^{b} X_j^*$$

In other words, b blocks, each of length T/b, are randomly selected from the original interval $(0,T]$. These are concatenated to produce the artificial process X^*. Clearly, X^* is a point process. It can be viewed as the restriction of a stationary point process to the interval $(0, T]$.

By repeatedly simulating the X^* process, one can then calculate statistics such as the first and second order intensity estimates, $\hat{p}_1^*$ and $\hat{p}_2^*(\tau)$, and observe their bootstrap sampling distributions. One would hope that the bootstrap distribution is a good approximation to the true sampling distribution. Bootstrap confidence intervals for the parameters can then be obtained according to standard procedures (see, for example, [7]). Of course, in the case of the second order intensity, it is not advisable to consider values of τ larger than T/b.

This bootstrap method is analogous to the block-resampling bootstrap for time series suggested by Künsch [8]. Carlstein [4] proposed a method for time series using fixed blocks. Recent results of this type are given in [13]. A similar idea could be used here, but the use of random blocks gives a richer set of possible realizations. Possolo [11] has suggested subsampling point processes, using a variant of Carlstein's method. Theorem 1 says that the bootstrap method holds for first order intensities, and is proved in section 6.

THEOREM 1 *Suppose the point process X has finite third moment intensities, and is ergodic so that second sample moments converge to their expectations. Then both*

$$\sqrt{T}(\hat{p}_1 - p_1) \Rightarrow N(0,\gamma^2)$$
$$\sqrt{T}(\hat{p}_1^* - E(\hat{p}_1^*|X)) \Rightarrow N(0,\gamma^2)$$

hold provided $b \to \infty$ and $T/b^2 \to 0$ as $T \to \infty$. The limiting variance γ^2 is given by the limit in (4).

The second order intensity can be estimated as in (2) but using the bootstrap point process. It can be studied for a fixed but otherwise arbitrary window length h, instead of $h = h(T)$. At the end of this section an alternative bootstrap for the second order intensity is described. The alternative bootstrap is based on creating a marked point process which is related to the numerator in (2).

The best choice of b is unknown. Shao and Yu [14] study the choice of block size for stationary time series.

Downloaded By: [University of California, Los Angeles] At: 17:34 10 October 2010

Downloaded By: [University of California, Los Angeles] At: 17:34 10 October 2010

For time series of i. i. d. observations, Künsch's method with $b=n$ is valid, since it is equivalent to the usual i. i. d. bootstrap (e.g. [7]). Analogously, the point process bootstrap proposed here is valid for homogeneous Poisson processes, provided b is taken large enough. More precisely, we have

THEOREM 2 *If $\xi(t)$ is a measurable function taking values in* $[0, 1]$, *and $X^{*,k}$ is the bootstrap of x based on $b=k$ blocks, then*

$$\lim_{k\to\infty} E\left[e^{\int_0^T \log\xi(t)X^{*,k}(dt)}|X=x\right] = e^{-x((0,T])/T\int_0^T(1-\xi(t))\,dt}. \tag{6}$$

The proof of the theorem is not difficult, using the usual approximation to $\xi(t)$ by simple functions, and the independence of certain increments in the $X^{*,k}$ process.

The expression on the left of (6) is the probability generating functional of the $X^{*,k}$ process. The expression on the right is the probability generating functional of a Poisson process on $(0,T]$ with intensity $x((0,T])/T$. The convergence of X^* to a Poisson process then follows from Lemmas 3 and 4 of [16], which state that convergence of probability generating functionals implies convergence in distribution of the associated processes and that the probability generating functional completely determines the probability structure of the associated process.

It should be noted that the theorem does not specify that the original process is Poisson. In fact, the Poisson limit will result whenever the original process is a simple point process. The implication here is that if one suspects any dependence in the original process, block sizes must not be taken too small.

A second natural method of a point process bootstrap for the second order intensity may also be considered. One could construct a marked point process $X(dx)(x+\tau, x+\tau+h]$, that is a point process with an occurrence at x, and mark given by $X(x+\tau, x+\tau+h]$ the count in the interval $(x+\tau, x+\tau+h]$. Note that this is the numerator in the definition, of $\hat{p}_2$ (2). One could apply a block bootstrap to this marked point process. For a given h and τ, this bootstrap method estimates the distribution of an estimate of the second moment measure of X, or equivalently the first moment of the marked point

Downloaded By: [University of California, Los Angeles] At: 17:34 10 October 2010

process. This bootstrap method will be referred to as the marked point process bootstrap.

The theorem below gives a justification for the second order intensity moment measure boot-strap. For small h it says that the bootstrap method applied to the second moment density $p_2(\tau)$ gives the correct asymptotic distribution.

THEOREM 3 *Suppose the point process X has finite and integrable fourth moment densities in the sense of Brillinger* [1]. *For a given h, conditional on the observed point process X on* $[0, T]$,

$$\sqrt{hT}(\hat{p}_2^*(\tau) - E(\hat{p}_2^*(\tau)|X)) \Rightarrow N(0, \sigma_2^2)$$

as $T \to \infty$, and $\sigma^2 = p_2(\tau) + O(h)$. For small h, the limiting variance σ_2^2 is approximately the limiting variance $\hat{p}_2(\tau)$.

The proof of this theorem follows in the same manner as Theorem 1. The calculation of the variance is then carried out in a similar manner to Brillinger [1].

4. SIMULATION STUDY

In order to investigate the validity of the bootstrap procedure on non-Poisson data, some special cases of Poisson cluster processes (e.g. [6]) were simulated. In particular, Neyman-Scott processes were conside-red. Bootstrap confidence interval coverage probabilities were then estimated for the two intensity parameters and compared with coverage probabilities using approximate confidence intervals.

To simulate a Neyman-Scott process, we

1. generated a Poisson random variate n having mean λT.
2. generated points $\{s_j, j = 1, 2, \ldots, n\}$ in the interval $(0, T]$ according to a Poisson λ process (the 'parent' process). (These points were generated using a uniform random number generator on the interval $(0, T]$).
3. generated independent negative binomial (r, p) variates $\{m_j', j = 1, 2, \ldots, n\}$.
4. generated $m_j = m_j' - r$ independent variates $\{t_{s_{j,1}}, t_{s_{j,2}}, \ldots, t_{s_j,m}\}$ from the distribution $F(t)$ (see below).

Downloaded By: [University of California, Los Angeles] At: 17:34 10 October 2010

The point process realization x was defined as the measure with unit mass at points in

$$U_{j=1}^{n}\{s_j + t_{s_j,1}, s_j + t_{s_j,2}, \ldots, s_j + t_{s_j,m_j}\}.$$

Three choices of $F(t)$ were considered: exponential, normal, and uniform. Of these, we present the results obtained for the exponential (mean β) case, since the other two cases were more favorable to the bootstrap.

In the exponential case, the underlying process X, has first order intensity

$$p_1 = \lambda r \frac{(1-p)}{p} \tag{7}$$

and second order intensity

$$p_2(\tau) = \lambda r \left(\frac{1-p}{p}\right)^2 \left[\lambda r + \frac{(r+1)}{2\beta} \, \mathrm{e}^{-|\tau|/\beta}\right]. \tag{8}$$

In addition, we have

$$\mathrm{Var}(\hat{p}_1) = \lambda r \left[\frac{1-p}{Tp} + (r+1)\left(\frac{1-p}{Tp}\right)^2 (T - \beta(1 - \mathrm{e}^{T/\beta}))\right]. \tag{9}$$

This was used in the construction of approximate confidence intervals for p_1, as a basis for comparison with the proposed bootstrap method. In a sense, the use of (9) or an appropriate estimate of it is overly optimistic, since the exact model is seldom known in practice. To construct approximate confidence intervals for $p_2(\tau)$, (5) was used.

For the simulation study, we set $\lambda = .2$, $(r, p) = (5, .5)$ and $\beta = 2$. Then

$$p_1 = 1, \tag{10}$$

and

$$p_2(\tau) = 1 + 1.5\,\mathrm{e}^{-\tau/2} \tag{11}$$

The number of process replicates was taken to be 500, which seemed adequate for our purpose. For each replicate, p_1 and $p_2(.2)$, $p_2(1.0)$

and $p_2(2.0)$ were estimated, and approximate confidence intervals were computed. The proportion of correct intervals could then be calculated using exact values obtained from (10) and (11).

Estimating p_1 was a straightforward matter. However, for the second order intensity, the value of the window parameter h had to be determined. For each value of τ, several values of h were examined. Using a mean square error criterion, it was found that the MSE was relatively constant (and minimum) from $h=0.14$ to 0.38. Thus we took $h=0.14$, in order to reduce bias as much as possible.

Initially, the value of T was chosen to be 500, large enough to see if the bootstrap method would work. With small T, the method would not be expected to work well. Some examples with smaller values of T were considered later, since the method worked well enough for large T.

For each point process realization x, 500 bootstrap replicates were obtained, using the algorithm described above. Since the optimal number of blocks is unknown, we used $b=10$, 30 and 50, in order to evaluate the effect of different block sizes.

For each bootstrap replicate, a value of $\hat{p}_1^*$ was calculated, together with valus of $\hat{p}_2^*(\tau)$ at $\tau=0.2$, 1.0 and 2.0. For $\alpha=.05$, .1, .2, bootstrap $1-\alpha$ confidence intervals were obtained using the equal-tail percentile method [7]. Here, the $\alpha/2$ and $1-\alpha/2$ quantiles of the set of 500 estimates were used as the confidence interval end points. The proportion of correct intervals was calculated, again using the true parameter values. The results are listed in Table I, and are discussed in subsection 4.1. Some adjustments were made, as well, and these are summarized in Table II, and described in subsection 4.2. Finally, some smaller samples were considered. These are summarized in Tables III and IV.

All simulations were carried out on PC's, using the Wichman and Hill [17] uniform random number generator with seeds generated by the internal clock.

4.1. Results for the Percentile Method

The results listed in Table I indicate that the bootstrap confidence intervals for p_1 perform only slightly worse than the approximate confidence intervals. It should be emphasized that the approximate confidence intervals are model-dependent, while the bootstrap is

Downloaded By: [University of California, Los Angeles] At: 17:34 10 October 2010

TABLE I Estimated confidence levels for Neyman-Scott model with $T=500$ using the bootstrap and asymptotic variance methods with $h=.14$. The last column refers to the marked point process bootstrap method

$1-\alpha$	$b=10$	$b=30$	$b=50$	*Asymptotic Approx.*	$b=30$
		p_1 (true value = 1.0000)			
.95	.888	.922	.900	.947	
.90	.832	.864	.834	.887	
.80	.732	.742	.734	.796	
		$p_2(0.2)$ (true value = 2.3573)			
.95	.880	.872	.864	.516 .522	.900
.90	.810	.806	.792	.452 .440	.844
.80	.702	.704	.680	.336 .338	.760
		$p_2(1.0)$ (true value = 1.9097)			
.95	.850	.874	.822	.530 .570	.888
.90	.796	.810	.768	.462 .498	.846
.80	.708	.700	.674	.360 .398	.744
		$p_2(2.0)$ (true value = 1.5518)			
.95	.874	.870	.842	.546 .574	.902
.90	.804	.816	.754	.480 .482	.834
.80	.696	.720	.652	.384 .406	.768

TABLE II Estimated confidence levels for Neyman-Scott model with $T=500$ using the bootstrap with an asymptotically pivotal statistic

	$h=.14$			$h=.14$	$h=.10$	$h=.05$
$1-\alpha$	$b=10$	$b=30$	$b=50$	*($b=30$,*	*marked PP*	*bootstrap)*
	$p_2(0.2)$ (true value = 2.3573)					
.95	.874 .890	.870 .880	.874 .878	.922	.926	.940
.90	.824 .826	.808 .818	.806 .816	.888	.870	.912
.80	.748 .750	.734 .736	.706 .716	.856	.820	.864
	$p_2(1.0)$ (true value = 1.9097)					
.95	.862 .906	.880 .920	.878 .928	.906	.930	.934
.90	.824 .844	.818 .870	.814 .868	.888	.886	.888
.80	.740 .760	.726 .744	.732 .766	.850	.814	.848
	$p_2(2.0)$ (true value = 1.5518)					
.95	.898 .918	.902 .962	.876 .940	.922	.912	.934
.90	.840 .846	.852 .894	.824 .894	.888	.870	.900
.80	.738 .750	.770 .764	.714 .778	.838	.818	.856

The first element of each pair in the $h=.14$ part of the table corresponds to the pivotal calculated using $\hat{p}_2$; the second element has been calculated using $\hat{p}_2'$ in the pivotal. The second part of the tables refers to the marked point process bootstrap, different h values and the pivotal method.

model-independent. Thus, it appears that for a point process with moderate dependence, the first order intensity can be estimated very well, provided there is enough data. In this case, one expects 500

Downloaded By: [University of California, Los Angeles] At: 17:34 10 October 2010

Downloaded By: [University of California, Los Angeles] At: 17:34 10 October 2010

TABLE III Estimated confidence levels for Neyman-Scott model with $T=40$ using the bootstrap and asymptotic variance methods

$1-\alpha$	$b=8$	Asymptotic Approx.
	p_1(true value = 1.0000)	
.95	.808	.952
.90	.726	.922
.80	.616	.826
	$p_2(0.2)$ (true value = 2.3573)	
.95	.700 .726 .800	.516 .578
.90	.640 .664 .718	.448 .448
.80	.532 .588 .618	.366 .366
	$p_2(1.0)$ (true value = 1.9097)	
.95	.682 .686 .826	.556 .600
.90	.636 .644 .746	.470 .510
.80	.554 .586 .662	.358. 410
	$p_2(2.0)$ (true value = 1.5518)	
.95	.744 .682 .820	.646 .788
.90	.658 .670 .670	.540 .684
.80	.554 .572. 682	.412 .452

The first element of each triple is calculated using $\hat{p}_2(\tau)$ and is non-pivotal; the second corresponds to the pivotal calculated using $\hat{p}_2$; the third element has been calculated using $\hat{p}_2'$ in the pivotal.

TABLE IV Estimated confidence levels for Neyman-Scott model with $T=80$ using the bootstrap and asymptotic varance methods

$1-\alpha$	$b=10$	Asymptotic Approx.	$b=30$
	p_1 (true value = 1.0000)		
.95	.808	.964	
.90	.752	.888	
.80	.644	.784	
	$h=.14$		$h=.10$
	$p_2(0.2)$ (true value = 2.3573)		
.95	.740 .792 .808 .844	.496 .516	.900
.90	.692 .730 .740 .782	.434 .434	.836
.80	.590 .644 .652 .736	.340 .340	.776
	$p_2(1.0)$ (true value = 1.9097)		
.95	.706 .798 .850 .846	.494 .556	.882
.90	.654 .720 .764 .794	.424 .466	.836
.80	.592 .642 .658 .720	.320 .368	.766
	$p_2(0.2)$ (true value = 1.5518)		
.95	.746 .790 .868 .830	.528 .656	.856
.90	.704 .746 .800 .774	.486 .586	.796
.80	.588 .642 .708 .714	.358 .478	.718

In the column $b=10$, $h=.14$, there are four elements: (i) uses $\hat{p}_2(\tau)$ and is non-pivotal; (ii) using the pivotal calculated $\hat{p}_2'$; (iii) using $\hat{p}_2'$ in the pivotal; and (iv) using the marked point process bootstrap with pivotal.

Downloaded By: [University of California, Los Angeles] At: 17:34 10 October 2010

points in the interval (0, 500], and 30 blocks (which is a little large than $\sqrt{500}$) gives the best performance, with a coverage error of about 3–6%. The other block sizes give inferior performance, but the differences are not dramatic. Thus, there is not a great deal of sensitivity toward block size in this bootstrap method.

When estimating $p_2(\tau)$, the asymptotic approximation is, surprisingly, very poor, even for this sample size. It appears that the error term in the asymptotic variance expression given by (5) has a large coefficient. Thus, it gives a biased estimate of the variance of $\hat{p}_2(\tau)$. A better estimate of this variance can be obtained by using (3), at least for τ. The coverage probability estimates are given in Table I as well. There is some improvement, but the bootstrap continues to perform better.

The performance of the bootstrap when estimating $p_2(\tau)$ is almost as good as its performance when estimating p_1, with a coverage error of about 8–10%. This is somewhat surprising, given that the estimator uses larger chunks of data, so that error due to concatenating bootstrap blocks is expected to be larger than when estimating p_1. In this case, differences in performance among the different block sizes are not very noticeable, though $b = 50$ may be a little worse than the other choices.

4.2. Use of an Asymptotically Pivotal Statistic

The use of a pivotal statistic is advocated by Hall [7] in case of i. i. d. bootstrapping in order to improve the coverage properties of confidence intervals. The idea is to bootstrap a statistic whose distribution does not depend on unknown parameters. Often (but not always), such a statistic can be found by scaling the estimator with an appropriate scale statistic. An asymptotically pivotal statistic is one whose asymptotic distribution does not depend on unknown parameters. The improvement in coverage is not as dramatic with asymptotic pivotals as with exact pivotals, but can be substantial nonetheless. In case of dependent data, less is known about the use of pivotal quantities, but one might conjecture that coverage performance would improve, by analogy with the independent case, especially in the context of finite or limited range dependence.

In view of (4), an asymptotically pivotal statistic for p_1 could be constructed using an estimate of the intensity function p_2 evaluated at

Downloaded By: [University of California, Los Angeles] At: 17:34 10 October 2010

several locations in $[0,T]$. Alternatively, one could estimate the power spectrum $f(\lambda)$ from the data, and use $f(0)/2\pi$ as a scale estimate.

We have not attempted this here, but we turn our attention to the more computationally straightforward pivotal statistic for the $p_2(\tau)$ estimator. Because of (5),

$$\sqrt{hT}\,\frac{\hat{p}_2(\tau) - p_2(\tau)}{\sqrt{\hat{p}_2(\tau)}} \tag{12}$$

is a candidate asymptotically pivotal statistic for this problem. A $100(1-\alpha)$ per confidence interval for $p_2(\tau)$ is then

$$\hat{p}_2(\tau) \pm q(\alpha)\sqrt{\frac{\hat{p}_2(\tau)}{hT}}$$

where $q(\alpha)$ satisfies $P(\text{modulus } (12) \leq q(\alpha)) = 1-\alpha$. The bootstrap confidence interval is then obtained by using the bootstrap to estimte the quantile $q(\alpha)$.

Table II gives coverage proportions for the respective confidence intervals. It also gives coverage proportions for confidence intervals which use the asymptotically pivotal statistic when $p_2(\tau)$ is estimated using (3). The coverage performance is better in this case. In fact, it appears that for larger τ there is negligible coverage error when $b=30$. For smaller τ, the coverage error is still about 5–8%.

The pivotal based on the marked point process bootstrap for $p_2(\tau)$ does better than the first bootstrap method. The coverages are shown in Tables I and II. For $h=0.14$ and $b=30$ the coverages are 3–8% better for the 90% and 95% confidence intervals. The coverages improve with smaller h. For the 80% confidence intervals the coverages are about the same over the various h values.

4.3. Smaller Samples

Because of the success of the bootstrap in the case of large samples, we decided to run a simulation of the $T=40$ case in order to see if one could bootstrap the Lake Constance freeze data (consisting of 37 points), for instance. For this purpose, all other parameters were kept at the same values as in the previous simulation, but we tried it with 8 blocks. The results are listed in Table III.

Downloaded By: [University of California, Los Angeles] At: 17:34 10 October 2010

The results were much worse than for the $T=500$ case, as one would expect. Without the use of the pivotal statistic, the coverage error is about 26%. This might be viewed as disastrous, until one finds that the coverage error for the asymptotic approximation is around 45%, in the case of the second order intensity. Therefore, the bootstrap has provided a substantial improvement.

The bootstrap results with the pivotal statistic are also included in Table III in the case of the second order intensity estimates. When $\hat{p}_2$ is used, the coverage error improves to about 23%, while use of $\hat{p}'_2$ reduces the coverage error to about 13% for large τ, and 18% for small τ.

Interestingly, the normal approximation is very accurate in the case of p_1. Of course, one needs to know the exact model, to do that well in practice.

The simulation exercise was repeated once more, this time with $T=80$ and $b=10$. By effectively doubling the sample size, we see an improvement in coverage by about 3–4%.

5. EXAMPLE: ONSET OF RAIN IN WINNIPEG

We demonstrate the efficacy of the point process bootstrap in the analysis of some meteorological data. Hourly rainfall amounts at a particular site in the city of Winnipeg have been recorded by the Winnipeg Climate Centre over the relevant period (May 1 – September 30) of each year from 1960 through 1980. Such data are more precisely regarded as realizations of a marked point process (see e.g. [6]), but it is also possible to view such data as point process realization if a point is defined as the beginning of a period of rain. Cressie [5, p. 645 – 647] has treated data arising from a primate psychological experiment in a similar manner. It should be noted here, as well, that the data has been discretized, so we can not be more precise than specifying the hour that rainfall began, but this should not affect our demonstration substantially. The data is listed in Table XII.

Our interest here is in the ability of the point process bootstrap to provide variance and sampling distribution estimates for the various relevant statistics, given a single year of data. Viewing the successive years of data as essentially i. i. d. point process realizations, we have an informal way of checking the validity of the bootstrap. Crude coverage

Downloaded By: [University of California, Los Angeles] At: 17:34 10 October 2010

estimates can be obtained, but these can only serve as a rough guide, since the number of years of observed data is limited.

In this context, the first order intensity has an interpretation as the mean number of rainshower occurrences per unit time (in this case, day). This parameter was estimated for each year, using (1). Of course, in order for this estimator to be sensible, the data must be stationary. This assumption seems to be approximately satisfied, although there may be a slightly increasing trend in the mean intensity.

For each year, a 95% bootstrap confidence interval for p_1 was computed using $b = 10$. The 21 resulting intervals are given in Table V, together with a cross-validatory estimate of the true mean intensity, computed under the assumption that the mean intensity is really constant over the period considered. These mean estimates were computed by leaving out the current year's intensity estimate and averaging over all other years. From the table, it can be seen that 16/21 of the intervals are 'correct', if the assumption of mean stationarity is true. This estimate of coverage accuracy is fairly close to the coverage obtained in the simulation example with $T = 80$.

The second order intensity was similarly estimated at $\tau = .4$, 2 and 4, and $h = .3$. The results are given in Tables VI, VII and VIII. Again, a

TABLE V 95% Bookstrap confidence intervals for first order intensity for rain events (percentile method)

year	*lower conf. limit*	*upper conf. limit*	*'true' value*
60	0.209	0.572	0.546
61	0.124	0.431	0.552
62	0.294	0.693	0.541
63	0.346	0.608	0.544
64	0.199	0.510	0.549
65	0.379	0.752	0.536
66	0.271	0.533	0.548
67	0.203	0.516	0.551
68	0.340	0.709	0.538
69	0.360	0.712	0.537
70	0.353	0.598	0.539
71	0.294	0.667	0.540
72	0.242	0.477	0.549
73	0.229	0.716	0.540
74	0.360	0.614	0.540
75	0.516	0.895	0.535
76	0.203	0.559	0.548
77	0.490	0.827	0.528
78	0.278	0.556	0.540
79	0.229	0.582	0.539
80	0.216	0.719	0.538

Downloaded By: [University of California, Los Angeles] At: 17:34 10 October 2010

TABLE VI 95% Bootstrap confidence intervals for $p_2(0.4)$ for rain events (percentile method)

year	*lower conf. limit*	*upper conf. limit*	*'true' value*
60	0.044	0.414	0.612
61	0.000	0.523	0.610
62	0.174	1.035	0.589
63	0.109	0.566	0.602
64	0.087	0.632	0.600
65	0.261	1.264	0.581
66	0.065	0.458	0.608
67	0.044	0.283	0.613
68	0.174	0.817	0.588
69	0.174	0.915	0.590
70	0.120	0.392	0.589
71	0.207	0.697	0.592
72	0.087	0.632	0.608
73	0.109	1.340	0.589
74	0.174	0.828	0.586
75	0.305	1.024	0.592
76	0.065	0.490	0.612
77	0.305	0.872	0.581
78	0.174	0.610	0.589
79	0.098	0.708	0.585
80	0.174	1.002	0.588

TABLE VII 95% Bootstrap confidence intervals for $p_2(2.0)$ for rain events (percentile method)

year	*lower conf. limit*	*upper conf. limit*	*'true' value*
60	0.000	0.196	0.330
61	0.000	0.338	0.330
62	0.109	0.632	0.321
63	0.044	0.414	0.327
64	0.000	0.174	0.330
65	0.131	0.719	0.316
66	0.044	0.327	0.328
67	0.000	0.392	0.325
68	0.087	0.523	0.312
69	0.087	0.490	0.315
70	0.044	0.501	0.322
71	0.109	0.458	0.320
72	0.044	0.370	0.326
73	0.000	0.643	0.321
74	0.087	0.436	0.319
75	0.218	0.850	0.309
76	0.044	0.534	0.325
77	0.349	0.959	0.300
78	0.065	0.261	0.321
79	0.044	0.414	0.313
80	0.044	0.458	0.321

Downloaded By: [University of California, Los Angeles] At: 17:34 10 October 2010

TABLE VIII 95% Bootstrap confidence intervals for $p_2(4.0)$ for rain events (percentile method)

year	*lower conf. limit*	*upper conf. limit*	*'true' value*
60	0.000	0.490	0.349
61	0.000	0.087	0.362
62	0.065	0.588	0.346
63	0.153	0.566	0.341
64	0.000	0.283	0.355
65	0.055	0.414	0.344
66	0.011	0.349	0.357
67	0.000	0.425	0.353
68	0.044	0.523	0.345
69	0.087	0.458	0.340
70	0.055	0.697	0.334
71	0.044	0.414	0.349
72	0.022	0.283	0.356
73	0.000	0.719	0.348
74	0.044	0.436	0.346
75	0.229	1.100	0.330
76	0.055	0.370	0.351
77	0.153	0.632	0.337
78	0.022	0.283	0.348
79	0.044	0.468	0.345
80	0.044	0.697	0.341

cross-validation idea was used as a check on the coverage accuracy of the intervals, assuming stationarity as before. The coverage estimates range from 14/21 for small τ to 17/21 for large τ. This may be slightly better than the coverage obtained in the simulation example with $T = 80$.

The pivotal statistic was used as well in the case of p_2. The results are given in Tables IX, X and XI. Interestingly, the coverage performance for $\tau = .4$ and $\tau = 4$ was worse than for the nonpivotal case, but the coverage for $\tau = 2$ was better. This may possibly be explained by the small sample sizes involved, but it may also have to do with the slight violation of the stationarity assumption alluded to above.

6. PROOF OF THEOREM 1

The bootstrap estimator of p_1 is

$$\hat{p}_1^* = \frac{1}{T} X^*(0, T) = \frac{1}{T} \sum_{j=1}^{b} Z_j^* \tag{13}$$

TABLE IX 95% Bootstrap confidence intervals for $p_2(0.4)$ for rain events (percentile-t method)

year	*lower conf. limit*	*upper conf. limit*	*'true' value*
60	−0.234	0.421	0.612
61	−3.153	0.490	0.610
62	−0.325	1.031	0.589
63	−0.206	0.601	0.602
64	−0.459	0.654	0.600
65	−0.216	1.280	0.581
66	−0.265	0.474	0.608
67	−0.189	0.307	0.613
68	−0.351	0.874	0.588
69	−0.279	0.950	0.590
70	−0.641	0.345	0.589
71	−0.121	0.774	0.592
72	−0.128	0.596	0.608
73	−0.810	1.254	0.589
74	−0.441	0.879	0.586
75	0.195	1.023	0.592
76	−0.101	0.464	0.612
77	−0.006	0.976	0.581
78	−0.345	0.657	0.589
79	−1.169	0.753	0.585
80	−0.392	1.022	0.588

TABLE X 95% Bootstrap confidence intervals for $p_2(2.0)$ for rain events (percentile-t method)

year	*lower conf. limit*	*upper conf. limit*	*'true' value*
60	−0.085	0.371	0.330
61	−0.416	0.364	0.330
62	0.203	0.787	0.321
63	0.183	0.612	0.327
64	0.000	0.349	0.330
65	0.264	1.048	0.316
66	0.119	0.450	0.328
67	−0.466	0.500	0.325
68	−0.064	0.821	0.313
69	0.081	0.792	0.315
70	0.106	0.825	0.322
71	0.252	0.850	0.320
72	−0.006	0.508	0.326
73	−0.598	0.949	0.321
74	0.142	0.699	0.319
75	0.379	1.137	0.309
76	−0.075	0.626	0.325
77	0.654	1.479	0.230
78	0.049	0.547	0.321
79	−0.657	0.726	0.313
80	−0.119	0.716	0.321

Downloaded By: [University of California, Los Angeles] At: 17:34 10 October 2010

TABLE XI 95% Bootstrap confidence intervals for $p_2(4.0)$ for rain events (percentile-t method)

year	*lower conf. limit*	*upper conf. limit*	*'true' value*
60	−0.156	0.817	0.349
61	0.067	0.263	0.362
62	0.285	1.007	0.346
63	0.393	1.021	0.341
64	−0.047	0.626	0.355
65	0.445	1.142	0.344
66	0.206	0.604	0.357
67	−0.055	0.655	0.353
68	0.333	1.112	0.345
69	0.353	1.065	0.340
70	0.025	1.378	0.334
71	0.394	1.186	0.349
72	0.203	0.592	0.356
73	−0.070	1.389	0.348
74	0.267	0.968	0.346
75	0.647	1.679	0.330
76	0.165	0.682	0.351
77	0.855	1.684	0.337
78	0.136	0.859	0.348
79	0.044	1.123	0.345
80	−0.268	1.225	0.341

where $Z_j^* = X((U_j, U_j + T/b)) = |X_j^*|$ from the bootstrap method, item 3 above.

$$\begin{aligned}
E(Z_j^*|X) &= \frac{1}{T\left(1-\frac{1}{b}\right)} \int_0^{T(1-\frac{1}{b})} X\left(\left(u, u+\frac{T}{b}\right]\right) du \\
&= \frac{b}{T(b-1)} \int_0^{T(1-\frac{1}{b})} \int_u^{u+\frac{T}{b}} X(dx) du \\
&= \frac{b}{T(b-1)} \int_0^{\frac{T}{b}} \int_0^x du X(dx) + \frac{b}{T(b-1)} \int_{\frac{T}{b}}^{T-\frac{T}{b}} \int_{x-\frac{T}{b}}^x du X(dx) \\
&\quad + \frac{b}{T(b-1)} \int_{T-\frac{T}{b}}^T \int_{x-\frac{T}{b}}^{T-\frac{T}{b}} du X(dx) \\
&= \frac{1}{(b-1)} \frac{b}{T} \int_0^{\frac{T}{b}} x X(dx) + \frac{1}{b-1} X\left(\left(\frac{T}{b}, T-\frac{T}{b}\right]\right) \\
&\quad + \frac{b}{T(b-1)} \int_{T-\frac{T}{b}}^T (T-x) X(dx)
\end{aligned}$$

Downloaded By: [University of California, Los Angeles] At: 17:34 10 October 2010

TABLE XII Winnipeg rain events recorded in number of days after April 30

- 1960: 17.08 17.29 18.00 18.25 18.54 25.42 25.58 25.88 29.08 30.88 31.50 31.58 31.75 32.46 32.54 32.67 34.88 38.83 39.38 39.58 43.63 44.21 44.33 47.29 47.96 48.33 48.67 51.25 51.38 51.88 52.83 52.96 58.00 58.75 58.83 59.13 70.88 80.29 84.21 89.21 92.25 98.58 98.83 104.04 107.50 108.50 108.63 108.88 112.58 112.71 114.75 116.29 116.75 118.25 118.50 127.04 127.17 127.29 128.58 130.38 130.63 130.83 134.08 138.21 138.54 138.67 145.96 150.00 150.13
- 1961: 5.25 5.38 5.63 6.29 6.63 7.00 10.08 38.29 40.88 60.46 70.83 71.83 72.00 72.25 72.58 72.92 73.13 73.46 73.96 74.17 74.50 77.96 81.63 82.54 83.54 86.08 86.33 91.83 94.96 108.46 115.00 115.08 115.21 123.58 124.33 124.92 125.00 129.83 130.08 130.21 131.96 132.58 140.96 141.04 141.13 141.58 141.75 141.92 145.58 145.96 147.92
- 1962: 6.46 6.79 8.58 8.75 8.88 8.96 12.79 14.33 16.25 16.54 18.21 18.58 21.25 21.71 22.00 22.08 22.29 22.46 28.58 29.00 29.29 29.50 29.63 29.96 30.08 30.29 34.04 37.13 37.25 37.58 40.13 40.42 44.83 45.50 54.08 63.92 66.33 67.79 68.04 68.46 69.67 70.04 80.04 80.46 82.29 82.54 83.67 83.83 84.21 84.50 84.63 84.83 88.67 89.42 89.50 90.54 94.04 94.17 94.33 94.96 95.58 96.38 96.58 102.17 102.71 105.88 106.88 109.75 109.92 113.29 113.88 113.96 119.00 122.13 122.38 122.71 125.08 130.54 130.92 131.04 132.04 132.17 132.29 137.92
- 1963: 6.04 11.46 11.54 11.75 12.04 12.13 12.46 12.67 16.54 16.75 16.96 25.63 25.79 29.83 29.92 30.50 30.71 30.92 33.29 33.79 34.04 36.29 36.63 38.38 38.50 40.17 40.67 42.67 42.92 43.21 48.54 49.00 53.17 55.25 56.96 59.38 67.83 70.67 71.08 71.88 73.42 75.96 76.88 78.79 78.92 84.38 85.92 86.29 96.96 99.13 99.38 99.46 103.58 106.63 106.83 106.96 109.38 112.00 112.13 112.25 114.50 118.75 119.75 119.83 120.04 120.29 120.46 120.58 132.92 133.46 136.96 139.96 140.13 141.83 146.38 149.88
- 1964: 2.21 2.79 5.25 5.92 6.46 6.96 7.92 9.71 9.96 14.67 17.17 38.54 38.63 38.79 41.75 42.21 42.29 42.38 46.92 47.13 47.63 48.42 48.71 49.08 53.08 53.33 53.71 62.83 66.71 68.67 68.92 86.54 86.71 86.79 91.46 91.54 93.17 101.33 101.46 101.67 101.88 106.04 112.79 113.46 114.58 115.79 116.04 119.21 120.63 120.75 123.04 124.38 124.50 125.67 126.08 131.42 132.54 132.75 137.83 140.96 148.21
- 1965: 0.79 4.04 4.63 4.88 5.00 5.25 7.17 7.29 16.42 19.21 19.88 23.04 23.58 23.75 23.92 24.38 24.54 24.67 24.92 25.04 25.17 25.29 25.71 25.79 30.50 30.75 32.33 32.42 33.29 33.71 33.79 33.88 37.71 38.58 50.17 51.42 55.21 55.58 55.88 56.50 56.67 56.96 57.67 57.88 58.33 58.63 58.71 61.29 61.75 63.83 66.67 68.92 71.50 72.29 72.79 73.25 73.54 73.63 76.29 77.17 79.46 80.92 82.79 89.38 95.75 96.00 96.21 104.29 107.83 109.75 117.08 117.46 117.54 117.71 120.13 120.67 120.88 121.04 121.75 121.83 122.58 125.88 126.08 126.17 126.38 130.88 131.33 136.63 138.29 138.50 139.04 139.21 139.38 139.58 144.58 144.83 144.92 146.75 151.58
- 1966: 14.21 14.71 14.79 16.29 16.42 16.58 17.38 18.13 18.33 21.92 22.04 22.50 22.75 22.96 23.58 31.63 35.25 35.42 37.04 40.96 42.58 43.29 43.54 43.79 48.88 51.21 53.42 54.88 55.46 60.33 61.33 61.42 62.17 62.25 65.63 65.71 66.17 68.96 77.04 78.67 81.17 83.13 83.21 92.04 96.00 96.75 98.00 103.50 103.75 106.88 108.25 113.42 113.63 113.75 122.00 122.25 122.75 125.46 125.79 134.79 135.04 150.29 150.63
- 1967: 6.42 21.33 21.54 21.67 22.96 23.04 23.17 24.13 38.75 42.46 42.75 49.42 49.67 50.58 50.71 51.00 56.71 58.63 58.88 59.75 61.71 61.92 62.08 65.83 65.96 66.25 67.33 67.75 82.63 83.50 83.75 90.50 92.21 92.54 92.63 93.88 94.42 94.50 96.17 96.92 97.00 98.67 98.75 99.04 99.67 107.58 107.88 132.29 134.79 134.92 135.50 136.83 148.33
- 1968: 2.46 5.63 5.83 6.79 7.13 13.63 13.92 14.04 14.13 14.54 26.13 26.75 28.38 29.46 29.67 31.46 33.88 34.00 37.42 38.08 40.71 42.96 47.67 47.88 48.13 50.83 58.00 58.21 60.46 61.42 67.88 70.96 73.13 73.46 74.83 76.33 76.63 80.88 81.50 82.92 89.25 89.75 90.00 90.25 90.88 90.96 94.21 96.00 99.71 99.92 101.00 103.17 104.54 104.63 106.17

Downloaded By: [University of California, Los Angeles] At: 17:34 10 October 2010

TABLE XII (Continued)

106.42 106.67 106.96 107.21 107.29 107.50 109.08 109.29 109.75 110.67 110.83 114.38 114.88 115.21 115.50 120.54 122.58 124.63 124.79 124.96 125.25 125.71 125.92 126.67 126.96 127.29 128.83 129.21 138.08 138.25 138.79 138.88 139.33 144.04 144.50 144.58 146.79 147.25 147.33

- 1969: 0.04 0.63 3.92 4.08 4.25 4.79 5.50 5.58 6.83 7.46 7.63 11.96 17.88 25.92 29.88 30.08 30.88 30.96 31.13 33.67 40.13 40.21 42.33 48.08 55.54 55.63 55.71 56.17 56.42 56.67 57.13 57.29 57.50 58.92 59.29 63.75 63.92 64.42 64.58 66.71 67.63 67.75 68.13 68.29 69.75 70.67 73.21 74.29 75.63 78.46 79.71 83.79 86.21 86.38 86.83 87.17 87.29 90.29 90.79 94.79 96.96 97.38 97.71 98.08 103.21 104.13 104.33 104.50 108.08 108.17 117.25 120.13 120.54 122.54 124.63 125.08 126.63 126.83 127.08 128.83 128.92 129.63 132.42 137.25 143.08 143.79 144.00 146.21 146.42 146.67 146.92 147.38 150.58 150.71 150.83 152.88
- 1970: 3.50 8.21 8.33 9.25 10.71 11.04 12.42 13.33 13.83 14.29 14.67 14.75 19.04 19.63 23.67 23.88 23.96 26.04 27.58 27.75 27.88 28.00 28.08 29.17 30.58 30.83 32.13 38.71 38.79 39.00 40.88 41.46 41.96 42.04 42.21 46.83 53.13 53.25 55.17 55.42 55.54 57.29 60.29 60.42 62.67 62.83 63.08 66.00 67.29 71.92 72.21 73.63 74.04 78.42 82.75 83.42 84.29 85.96 89.25 89.46 89.54 106.54 106.67 106.83 108.92 112.63 120.13 120.38 120.63 120.71 124.04 127.92 129.33 130.83 130.96 131.33 133.54 133.96 134.50 134.63 137.21 142.92 143.13 143.29 143.67 143.75 143.96 147.08 147.63 147.92 148.08 148.17
- 1971: 6.08 6.17 9.21 13.67 15.88 21.17 21.29 21.71 27.96 34.54 35.08 35.71 40.00 40.17 40.42 40.75 41.67 41.79 41.92 42.67 46.08 47.00 49.08 49.42 49.67 52.67 53.21 56.17 57.75 57.92 60.00 60.42 60.50 63.29 63.79 67.00 70.00 72.21 72.58 72.71 72.83 72.96 75.83 77.88 78.04 79.04 79.29 80.92 82.38 82.50 83.79 83.88 84.08 84.54 85.17 85.25 87.25 87.63 88.50 88.58 88.83 89.08 91.25 104.63 107.54 107.88 110.25 115.00 115.79 123.50 126.29 127.08 127.21 127.38 127.71 128.21 129.71 137.38 137.50 138.54 143.08 143.58 144.88 148.83 152.67
- 1972: 4.71 10.83 10.92 22.08 23.71 23.83 24.13 25.38 25.83 25.92 26.38 34.88 41.00 41.08 43.17 43.50 48.21 48.33 48.50 48.96 56.75 57.75 66.75 68.63 69.00 69.79 70.04 70.13 71.92 73.50 73.83 74.63 78.96 79.38 82.88 83.71 90.13 92.83 92.96 96.38 98.33 98.54 105.42 106.71 111.04 122.92 123.38 125.83 125.96 128.04 130.04 138.21 140.04 141.88 146.00 146.17 149.46 149.63 150.00 150.75
- 1973: 4.25 4.67 10.33 11.67 20.79 22.67 24.33 24.42 29.29 29.42 29.50 32.63 33.00 33.17 33.33 33.42 33.50 33.58 34.25 37.71 37.96 39.08 39.21 39.38 40.04 44.00 44.08 44.63 46.96 47.38 47.54 48.58 48.67 49.13 49.21 49.38 49.46 49.63 49.79 50.04 50.17 50.71 56.83 61.75 62.29 65.38 67.08 73.54 74.38 77.92 78.42 83.17 84.08 84.21 86.63 86.79 86.88 89.25 89.42 98.04 98.13 98.42 99.96 100.33 103.42 109.92 110.08 115.58 115.71 115.79 115.96 122.17 123.83 124.00 125.25 125.63 126.46 126.58 135.13 135.21 136.71 137.08 137.21 143.67 144.13 146.25 146.50
- 1974: 4.00 6.17 8.08 8.17 9.58 10.83 14.67 15.46 18.29 18.29 18.79 19.17 19.25 19.67 19.75 19.92 21.71 22.00 26.42 26.67 27.92 32.92 34.88 35.29 35.46 37.50 43.88 44.00 44.21 44.63 44.75 44.96 50.17 59.42 61.17 64.88 66.83 68.13 68.29 71.92 72.25 76.83 84.25 85.92 91.75 91.83 92.08 102.08 103.67 105.50 105.79 105.92 107.04 107.79 111.25 112.50 112.83 115.13 115.29 115.54 115.96 119.33 119.67 120.33 120.96 121.42 121.67 122.13 122.21 123.50 123.63 124.33 127.92 128.08 132.08 132.71 132.83 133.17 133.33 133.79 138.13 140.04 141.29 141.54 142.08 142.50 146.38 151.29 151.63
- 1975: 0.08 0.25 1.63 12.54 12.71 12.83 13.29 16.13 16.25 16.79 16.96 17.83 17.96 19.13 19.33 19.42 22.17 22.25 22.50 23.08 23.17 26.21 26.38 26.63 26.88 26.96 28.21 29.71 30.46 32.54 32.71 34.46 34.67 38.33 38.46 39.38 39.54 40.08 40.46 43.00 50.00 50.75 52.25 52.79 56.63 58.33 60.08 72.29 77.54 78.29 78.50 78.63 78.79 79.63 80.92 81.04 82.38 83.75 85.92 91.33 91.54 91.67 92.00 94.04 94.58 98.75 99.00 102.13 103.17

Downloaded By: [University of California, Los Angeles] At: 17:34 10 October 2010

Downloaded By: [University of California, Los Angeles] At: 17:34 10 October 2010

TABLE XII (Continued)

103.79 103.92 105.13 106.46 106.96 111.54 113.96 115.63 116.13 116.21 116.54 116.67 117.08 117.42 118.42 118.54 120.50 122.58 124.58 124.71 125.71 126.75 127.46 128.21 128.46 128.83 128.96 133.58 140.25 140.46 141.58 141.75 142.88 152.17 152.63

- 1976: 12.04 13.54 13.79 13.88 14.25 25.54 32.08 35.00 35.88 36.00 36.17 36.67 37.63 38.71 39.58 42.00 42.92 43.29 43.46 43.75 44.04 44.42 44.54 46.71 51.17 54.54 54.75 54.96 56.38 56.46 56.63 56.75 56.92 58.71 68.92 72.96 75.25 79.46 79.79 85.13 87.96 89.63 89.79 90.50 100.71 101.96 102.46 109.75 110.54 111.08 111.29 117.79 117.92 118.17 118.96 129.71 130.88 134.79 134.88 135.67 144.00 147.67
- 1977: 2.17 2.29 2.50 2.71 3.21 3.75 4.21 4.38 4.88 12.46 13.63 13.79 14.75 14.83 17.04 17.96 18.17 18.63 21.71 22.25 22.33 25.58 26.79 27.58 27.96 28.83 29.00 29.13 29.21 33.17 35.38 35.58 40.04 40.42 42.75 42.83 43.50 44.83 45.75 47.17 47.42 47.67 48.71 49.79 57.75 58.63 59.58 59.71 60.50 60.83 62.63 62.75 65.42 65.75 67.83 70.50 70.63 71.08 71.25 71.33 71.46 71.63 73.33 73.54 73.79 88.58 90.13 90.71 90.92 91.04 93.79 95.29 95.58 96.17 96.42 97.88 98.63 99.08 100.67 101.50 101.83 102.88 103.75 106.04 106.67 107.54 107.63 111.00 111.46 112.67 112.79 113.29 116.33 116.54 119.50 120.71 120.92 122.38 122.67 123.96 125.79 126.13 127.21 127.58 130.21 130.71 135.04 141.29 141.42 142.75 142.83 143.00 144.13 144.33 144.67 145.04 145.79 146.13 146.79 146.88 147.13 147.67 147.96 148.38 148.83
- 1978: 4.71 7.04 7.88 9.33 9.58 9.83 9.96 10.13 10.38 18.29 22.63 22.88 23.92 24.08 24.79 30.08 30.67 31.17 31.42 31.58 36.13 36.25 44.04 46.96 49.46 49.79 50.29 50.38 54.29 54.46 54.96 56.42 57.75 57.83 57.92 62.63 62.79 64.38 64.58 65.13 65.29 68.46 72.13 72.33 72.63 75.29 77.00 77.67 79.17 82.75 84.38 88.13 88.75 88.96 89.04 91.71 92.21 93.21 93.50 94.25 97.92 101.00 104.88 114.54 114.83 117.13 118.29 123.17 134.21 134.50 135.29 135.83 135.92 136.25 136.58 136.92 138.08 138.25 140.79 140.96 141.04 141.58 148.50 150.50 150.67 150.96 151.08
- 1979: 0.25 0.75 0.83 3.92 5.63 6.13 6.29 6.67 11.58 16.00 20.88 21.25 22.38 24.50 24.58 24.75 28.25 28.33 28.46 28.83 28.92 31.04 31.42 31.58 31.71 32.33 33.88 34.29 37.75 43.88 44.75 49.79 50.04 50.17 50.58 50.71 57.79 58.08 73.04 73.29 73.88 74.71 74.88 79.08 83.21 83.38 85.25 89.83 92.67 92.79 94.42 94.58 94.71 95.00 100.33 103.04 103.25 103.79 103.88 112.50 112.71 112.96 114.00 114.33 114.58 114.67 114.79 114.92 119.13 122.13 122.46 122.63 122.88 122.96 123.46 123.71 125.54 125.63 128.79 129.88 131.71 132.83 132.96 133.13 133.38 133.88 134.92 135.71 139.83 149.63
- 1980: 9.63 10.50 10.58 25.17 25.38 26.00 27.38 27.46 27.58 27.92 30.71 42.46 50.04 52.17 54.13 57.42 57.54 57.67 58.17 58.63 60.38 63.92 64.33 64.54 71.38 74.25 78.00 80.54 81.29 90.08 91.29 94.75 94.92 95.25 95.54 95.75 96.00 96.21 97.42 97.67 97.92 98.79 98.96 101.46 101.75 102.17 102.54 103.00 104.50 104.58 104.67 104.92 107.54 107.83 108.08 108.17 108.46 111.42 112.17 112.25 114.50 115.63 115.92 119.42 120.33 120.67 120.88 121.17 123.83 125.46 125.71 126.21 126.29 126.71 130.29 134.63 137.08 137.96 138.38 139.25 139.46 141.21 141.63 141.79 143.50 144.46 144.71 145.29 145.50 146.92 151.29 151.38 152.83

$$= \frac{1}{(b-1)}\int_0^{\frac{T}{b}}\left(\frac{bx}{T}-1\right)X(dx) + \frac{1}{b-1}X((0,T]) + \frac{1}{(b-1)}\int_{T-\frac{T}{b}}^{T}\left(\frac{b}{T}(T-x)-1\right)X(dx). \tag{14}$$

Downloaded By: [University of California, Los Angeles] At: 17:34 10 October 2010

By Markov's inequality, as $T/b \to \infty$,

$$E(Z_j^*|X) = O_p\left(\frac{T}{b^2}\right) + \frac{1}{b-1}X(0,T]). \tag{15}$$

From (14) and (15)

$$\begin{aligned} T\,\mathrm{Var}(\hat{p}_1^*|X) &= \frac{b}{T}\mathrm{Var}(Z_1^*|X) \\ &= \frac{b}{T}\{E(\{Z_1^*\}^2|X) - E(Z_1^*|X)^2\} \\ &= \frac{b}{T}\frac{1}{T(1-\frac{1}{b})}\int_0^{T(1-\frac{1}{b})}\int_u^{u+\frac{T}{b}}\int_u^{u+\frac{T}{b}}(X(dx_2)X(dx_1) \\ &\quad - p_1^2\,dx_2\,dx_1)\,du \\ &\quad + \frac{b}{T}\left(\frac{T}{b}p_1\right)^2 - \frac{b}{T}\left(O_p\left(\frac{T}{b^2}\right) + \frac{T}{b-1}\frac{1}{T}\,X((0,T])\right)^2 \end{aligned}$$

The sum of the last two terms is

$$\frac{\sqrt{T}}{b}\sqrt{T}\{p_1^2 - \hat{p}_1^2\} + \frac{T}{b}O\left(\frac{1}{b}\right) = O\left(\frac{\sqrt{T}}{b}\right) + O\left(\frac{T}{b^2}\right).$$

The first term is

$$\begin{aligned} &\frac{b}{T}\frac{1}{T(1-\frac{1}{b})}\int_0^{T(1-\frac{1}{b})}\int_u^{u+\frac{T}{b}}\int_u^{u+\frac{T}{b}}(X(dx_2))X(dx_1) - p_1^2dx_1dx_1)du \\ &= \frac{b}{T^2(1-\frac{1}{b})}\int_0^{T(1-\frac{1}{b})}\int_u^{u+\frac{T}{b}}\int_{-\infty}^{\infty} I\left(v \neq 0, u < v + x_1 \le u + \frac{T}{b}\right) \\ &\quad (X(d(v+x_1))X(dx_1) - p_1^2dvdx_1)du \\ &\quad + \frac{b}{T^2(1-\frac{1}{b})}\int_0^{T(1-\frac{1}{b})}\int_u^{u+\frac{T}{b}} X(dx)du \\ &\to \int_{-\infty}^{\infty} c(v)dv + p_1 \end{aligned}$$

as $T \to \infty$, $b \to \infty$ and $T/b^2 \to 0$, for an ergodic point process with second moments, and $c(v)$ is given in (4).

Note that this limiting normalized conditional variance of $\hat{p}_1^*$ (the bootstrap estimate) equals the normalized limiting variance of $\hat{p}_1$.

Downloaded By: [University of California, Los Angeles] At: 17:34 10 October 2010

This allows one to apply a Berry-Esseen theorem to the conditionally i.i.d. array Z_j^*, $j = 1,\ldots,b$. Since the bootstrap distribution of $\sqrt{T}(\hat{p}_1^* - E[\hat{p}_1^*|X])$ has the same normal limit distribution as $\sqrt{T}(\hat{p}_1 - p_1)$, it follows that the bootstrap distribution approximates asymptotically the distribution of $\hat{p}_1$. Note also that the rates for the block size b is $b \to \infty$ and $Tb^2 \to 0$, the same as the block bootstrap method for time series as given in Künsch [8].

7. OPEN QUESTIONS

The algorithm is readily generalized to d-dimensions. For point processes data on a d-dimensional hypercube $(0,T]^d$,

1. Generate b^d independent uniform random variates U on the hypercube

$$(0, T(1 - 1/b)]^d.$$

2. For $j \in \{1, 2, \ldots, b\}^d$,
   (a) Set

$$A_j = \prod_{i=1}^{d} (U_j + e_i T/b].$$

   (b) Set

$$A_j^* = A_j - U_j + (j-1)\frac{T}{j}b.$$

   (c) Set

$$X_j^*(\cdot) = |A_j^* \cap .|$$

3. Set

$$X^* = \sum_j X_j^*.$$

The result is a d-dimensional point process observed on the original hypercube.

We conjecture that this method should be suitable for inference involving the first order intensity, in the presence of weak dependence for the original process. However, the situation is less certain in the case of second order intensity estimation.

There is a geometric consideration. The borders for concatenating small blocks are of size $(T/b)^{d-1}$, which is as big or bigger than the rate of convergence for the normal limit theorem for $\hat{p}_2$. This does not matter for the first order intensity, since this can be viewed as either the mean for a concatenated series or the mean of smaller i.i.d. blocks.

As mentioned in the rain example, a marked point process may be a more precise model. Here, the marks correspond to the recorded amount of rainfall for each hour. Can the method of block resampling be successfully applied to such marked point processes?

Politis and Romano [9] have suggested a circular variant to correct for bias in the Künsch time series bootstrap. The point process version (in 1-D), as well as a possible toroidal version (in 2-D) requires investigation. Furthermore, the blocks of blocks method of Politis and Romano [10] can be adapted to point processes in one dimension. Potentially, this adaptation offers a substantial improvement over the original block-resampling method. Whether the method can be generalized to higher dimensions is another question to be addressed.

8. CONCLUSIONS

In this paper, we have proposed a method for bootstrapping point processes which is inspired by the block-resampling algorithm used in time series bootstrapping. We have considered the first and second order intensities and the confidence intervals proposed by Brillinger [1]. We used the bootstrap to estimate confidence intervals for these parameters, and have found that the bootstrap performs well, in the case of a weakly dependent point process. For first order intensities, the bootstrap works very well, and the bootstrap works adequately in the case of second order intensities. In the latter case, we find that the use of a pivotal statistic is helpful. When the sample sizes are small (e.g., expected number of points < 80), the confidence interval coverages are poor for second order intensities, unless the estimator is modified to adjust for small T.

Downloaded By: [University of California, Los Angeles] At: 17:34 10 October 2010

An alternative to using the pivotal statistic might be to try bootstrap iteration [7]. We have not tried that here, since it is very computationally intensive, but as computing machines become faster, this is another possibility to explore.

Downloaded By: [University of California, Los Angeles] At: 17:34 10 October 2010

Acknowledgements

This work has been supported by grants from the Natural Sciences and Engineering Research Council of Canada. The authors thank the referee for some helpful comments and questions which have led to improvements in the paper.

References

[1] Brillinger, D. R. (1975a). Statistical inference for stationary point processes. In *Stochastic Processes and Related Topics.* (M. Puri, Ed.) 55–99, Academic Press, New York.
[2] Brillinger, D. R. (1975b). *Time Series: Theory and Data Analysis.* McGraw-Hill.
[3] Brillinger, D. R. (1978). Comparative aspects of the study of ordinary time series and of point processes. In *Developments in Statistics,* **1**, 33–133.
[4] Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Ann. Statist.,* **14**, 1171–9.
[5] Cressie, N. (1991). *Statistics for Spatial Data,* Wiley, New York.
[6] Daley, D. J. and Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes.* Springer-Verlag, New York.
[7] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion.* Springer-Verlag, New York.
[8] Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.,* **17**, 1217–41.
[9] Politis, D. N. and Romano, J. P. (1992a). A circular block-resampling procedure for stationary data. In *Exploring the Limits of Bootstrap* (R. Lepage and L. Billand, Eds.) 263–270. Wiley, New York.
[10] Politis, D. N. and Romano, J. P. (1992b). A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation. *Ann. Statist.,* **20**, 1985–2007.
[11] Possolo, A. (1991). Subsampling a random field. *IMS Lecture Notes–Monograph Series: Spatial Statistics and Imaging.* (A. Possolo, Ed.) **20**, 286–294.
[12] Reiss, R.-D. (1993). *A Course on Point Processes.* Springer-Verlag, New York.
[13] Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap.* Springer, New York.
[14] Shao, J. and Yu, H. (1993). Bootstrapping the sample means for stationary mixing sequences. *Stoch. Proc. and their Appl.,* **48**, 175–90.
[15] Steinijans, V. W. (1976). A stochastic point process model for the occurrence of major freezes in Lake Constance. *Appl. Statist.,* **25**, 58–61.
[16] Westcott, M. (1972). The probability generating functional. *J. Aust. Math. Soc.,* **14**, 448–466.
[17] Wichman and Hill (1982). An efficient and portable pseudo-random number generator. *Appl. Statist.,* **31**, AS 183, 188–190.