# On Nonparametric Estimators of the Density of a Non-negative Function of Observations

**Yogendra P. Chaubey**[a] **and Pranab K. Sen**[b]

[a]Department of Mathematics and Statistics, Concordia University,
Montreal, QC, H3G 1M8, CANADA
[b]Department of Statistics and Biostatistics, University of North Carolina at Chapel Hill,
Chapel Hill, NC 27514, USA

**Abstract**

Let $\{X_1, ..., X_n\}$ be a random sample from a continuous distribution $F$ defined on the $k-$dimensional Euclidean space $\mathbf{R}^k$, for some $k \geq 1$. In many statistical applications we are interested in statistical properties of a function $q(X_1, ..., X_m)$ of $m \geq 1$ observations. Frees (1994, *J. Amer. Stat. Assoc.*) considered estimating the density function $g$ associated with the distribution function

$$G(t) = P\big(h(X_1, ..., X_m) \leq t\big)$$

using the kernel method. In many applications, though, the functions of interest are non-negative where the usual symmetric kernels applied in the kernel density estimation are not appropriate. This paper adapts the alternative density estimator developed in Chaubey and Sen (1996, *Statistics and Decisions*) by smoothing the so called *empirical kernel distribution function*:

$$G_n(t) = \binom{n}{m}^{-1} \sum_{(n,m)} \mathbf{1}\big(h(X_{i_1}, X_{i_2}, ..., X_{i_m}) \leq t\big),$$

where $\mathbf{1}(A)$ denotes the indicator of $A$ and $\sum_{(n,m)}$ denotes sum over all possible $\binom{n}{m}$ combinations. Applications and asymptotic properties of the alternative estimator are investigated.

# 1 Introduction and Background

Let $\{X_i, i \geq 1\}$ be a sequence of independent and identically distributed random vectors (i.i.d.r.v) with a distribution function (d.f.) $F$ defined on the $k-$dimensional Euclidean space $\mathbf{R}^k$, for some $k \geq 1$. Consider a functional $\theta(F)$ of the $d.f.$ $F$, for which there exists a function $h : \mathbf{R}^{k\times m} \to \mathbf{R}$, such that

$$\begin{aligned} \theta(F) &= Eh(X_1, X_2, ..., X_m) \\ &= \int_{\mathbf{R}^{k\times m}} h(x_1, x_2, ..., x_m) dF(x_1)...dF(x_m), \end{aligned} \tag{1.1}$$

for every $F$ belonging to a class $\mathcal{F}_0$ of d.f.'s on $\mathbf{R}^k$. Without loss of generality, the function $q(.)$ is assumed to be symmetric in its $m$ arguments. If $m(\geq 1)$ is the minimal sample size for which equation (1.1) holds, then $h(x_1, ..., x_m)$ is called the *kernel* of the U-statistic and $m$ the degree of $\theta(F)$. An optimal (symmetric) estimator of $\theta(F)$ is the $U-$statistic [*viz.*, Hoeffding (1948)]

$$U_n = U_n(X_1, X_2, ..., X_n) = \binom{n}{m}^{-1} \sum_{(n,m)} h(X_{i_1}, X_{i_2}, ..., X_{i_m}), \tag{1.2}$$

where the sum $\sum_{(n,m)}$ denotes the sum taken over all subsets $1 \leq i_1 < i_2 < ... < i_m \leq n$ of $\{1, 2, ..., n\}$. The readers may be referred to the texts by Serfling (1980) and Sen (1981) for basic theoretical results on $U-$ statistics whereas, the text by Lee (1990) provides an excellent introductory source.

Let us define the function $G$ on $\mathbf{R}$ for a given function $h : \mathbf{R}^{k\times m} \to \mathbf{R}$, as

$$G(t) = P\big(h(X_1, ..., X_m) \leq t\big) \tag{1.3}$$

then we can clearly write

$$\theta(F) = \int t dG(t),$$

which is a linear functional of $G$, called as the *kernel distribution function* for kernel $h$. A more flexible functional covering nonlinear cases, (such as densities and quantiles) may be considered as a general functional of $G$, denoted by $\theta^*(G)$. In this set up, the von Mises' (1947) functional estimator $\theta(F_n)$ may be replaced by $\theta^*(G_n)$, where

$$G_n(t) = \binom{n}{m}^{-1} \sum_{(n,m)} \mathbf{1}\big(h(X_{i_1}, X_{i_2}, ..., X_{i_m}) \leq t\big), \tag{1.4}$$

where $\mathbf{1}(A)$ denotes the indicator of $A$. Note that $G_n(t)$ is also a $U-$statistic, with $h$ replaced by $q$ :

$$q(x_1, x_2, ..., x_m; t) = \mathbf{1}\big(h(x_1, x_2, ..., x_m) \leq t\big), \tag{1.5}$$

which depends on an additional parameter $t$.

A large class of problems of estimation of functionals $\theta^*(G)$ are concerned with kernels which depend on some real parameter $t \in \mathbf{R}$, such as the kernel $q$ in Eq. (1.5). Some important examples are given below.

**Example 1.1: Wilcoxon score statistic** [see Sen (1963) and Hodges and Lehman (1963)] is given by

$$\hat{\theta} = \text{median}\big\{\frac{1}{2}(X_i + X_j), 1 \le i < j \le n\big\}.$$

This statistic corresponds to $G_n^{-1}(1/2)$ where

$$G_n^{-1}(u) = \inf\{x : G_n(x) \ge u\}, 0 \le u \le 1$$

is the generalized inverse of $G_n$, and $h(x, y) = (1/2)(x + y)$.

**Example 1.2: Gastwirth's (1973) modification of Gini's Coeeficient.** Gini's coefficient [see Sen (1986)] is used as a measure of income inequality defined as

$$\eta = \frac{E(|X_1 - X_2|)}{E(|X_1 + X_2|)}.$$

Gastwirth proposed a modification

$$\eta^* = E\left\{\frac{|X_1 - X_2|}{|X_1 + X_2|}\right\},$$

which corresponds to considering

$$h(x, y) = |x - y|/|x + y|,$$

where both $x, y > 0$.

**Example 1.3: Reliability of $m$ out of $n$ components in series and parallel** [see Ghosh, Mukhopadhyay and Sen (1997), §13.3] are related with the kernels $h_s$ and $h_P$ respectively, given by:

$$h_S(x_1, x_2, ..., x_m; t) = \mathbf{1}\big(\min(x_1, x_2, ..., x_k) > t\big), \quad (1.6)$$
$$h_P(x_1, x_2, ..., x_m; t) = \mathbf{1}\big(\max(x_1, x_2, ..., x_k) > t\big). \quad (1.7)$$

**Example 1.4: Generalized Gini Mean Difference** Generalized Gini mean difference is given by

$$g_s = E|X_1 - X_2|^s, s > 0 \quad (1.8)$$

which corresponds to the kernel

$$h(x, y) = |x - y|^s.$$

**Example 1.5:** In many applications interest lies in **quantiles of the kernel distribution function** (k.d.f.) [see Sen (1983)] $G(t)$. For example, Bickel and Lehman (1979) consider, median corresponding to the k.d.f. for $h(x,y) = |x-y|$ as a measure of the spread of $F$, and Choudhury and Serfling (1988) consider a $U-$ quantile in the regression context, where, the kernel is given by $h((x_1,y_1),(x_2,y_2)) = (y_2-y_1)/(x_2-x_1)$.

**Example 1.6: Liu's (1990) simplicial depth** function for $x \in \mathbf{R}^p$ is defined by

$$D_n(x) = \binom{n}{p+1}^{-1} \sum_{(n,(p+1))} \mathbf{1}(x \in [S(X_{i_1}, X_{i_2}, ..., X_{i_{p+1}})]), \tag{1.9}$$

where $S(x_1, x_2, ..., x_{p+1})$ denotes the $p-$dimensional simplex determined by the points $x_1, x_2, ..., x_{p+1}$ and $\mathbf{1}(A)$ denotes the indicator function of the set $A$. Obviously, $D_n(x)$ is a $U-$ statistics indexed by $x \in \mathbf{R}^p$ with

$$h(x_1, ..., x_{p+1}; x) = \mathbf{1}(x \in [S(x_1, x_2, ..., x_{p+1})]).$$

Other measures of depth can also be cast in this set up. [see *e.g.*, Zuo and Serfling (2000)].

**Example 1.7: "Ley Hunting".** Silverman and Brown (1978) propose the following statistic for testing randomness against some collinearties in the data;

$$T_n(\epsilon) = \binom{n}{3}^{-1} \sum_{(n,3)} \mathbf{1}(\alpha(X_i, X_j, X_k) > \pi - \epsilon).$$

**Example 1.8: Correlation-Dimension.** Grassberger and Proccacia (1983) propose estimating the correlation dimension of a dynamical system by considering the statistic,

$$C_n(r) = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \mathbf{1}(\|X_i - X_j\| \le r).$$

Here, also, the kernel function is of the general form $h(x,y;t)$.

In all these examples, the $U-$statistic

$$U_n = U_n(t) = \binom{n}{m}^{-1} \sum_{(n,m)} h(X_{i_1}, X_{i_2}, ..., X_{i_m}; t),$$

is a random function of the real parameter $t$, rather than a random variable. We will denote

$$U(t) = \int_{\mathbf{R}^{k \times m}} h(x_1, x_2, ..., x_m; t) dF(x_1)...dF(x_m),$$

and consider the following process,

$$W_n(t) = \sqrt{n}(U_n(t) - U(t)),$$

which is called *the empirical process of U-statistics structure.*

$W_n(t)$ can be regarded as a generalization of the classical empirical process. Construct the empirical distribution $G_n$ from the set of all random vectors

$$\{(X_{i_1}, ..., X_{i_m}) : 1 \leq i_1 < i_2 < ... < i_m \leq n\},$$

*i.e.*

$$G_n := \binom{n}{m}^{-1} \sum_{(n,m)} \delta_{(X_{i_1}, X_{i_2}, ..., X_{i_m})},$$

where $\delta_{(X_{i_1}, X_{i_2}, ..., X_{i_m})}$ puts mass 1 on the $m-$tuple $(X_{i_1}, X_{i_2}, ..., X_{i_m})$. Let $G = F \times F \times ... \times F$. Then the empirical process of $U-$ statistics structure is the process

$$\begin{aligned} W_n(t) &= \sqrt{n} \int_{\mathbf{R}^{k\times m}} h(x_1, x_2, ..., x_m; t) d(G_n - G) \\ &= \sqrt{n} \left[ \binom{n}{m}^{-1} \sum_{(n,m)} h(X_{i_1}, X_{i_2}, ..., X_{i_m}, t) - U(t) \right]. \end{aligned}$$

If $X_i \in \mathbf{R}$ and $h(x;t) = \mathbf{1}(x \leq t)$, then $W_n(t)$ is the ordinary empirical process. Silverman (1983) studied the weak-convergence of such processes whereas independently, Sen (1983) considered these processes corresponding to kernels of the type as given in Eq. (1.5) and obtained weak and strong convergence using martingale methods. These processes have been further studied by Serfling (1984), Dehling, Denker and Philipp (1987), Helmers, Janssen and Serfling (1988), Nolan and Pollard (1987,1988), Schneemeier (1993), Arcones and Giné (1993), Arcones (1993, 1996) and many others.

In this paper, we are concerned with the kernels of the form given in (1.5) and interested in estimating

$$\theta(F) = G(t) = \int \mathbf{1}\big(h(x_1, ..., x_m) \leq t\big) dF(x_1)...dF(x_m)$$

or functionals of $F$ expressed as regular functionals of $G$, i.e.

$$\theta_f(G) = \int f dG, \quad f \in \mathcal{F}.$$

Or, we may want to estimate a non-regular functional of $G$ such as the density $g(t) = (d/dt)G(t)$ or quantile function $\theta(F) = Q_H(u) = \int_0^1 G^{-1} d\mu_\theta$, for some finite signed Borel measure $\mu_\theta$. Estimation of $G(t)$ and associated quantiles are of specific interest in many applications, such as in examples 1.1, 1.3, 1.5 and 1.7. Frees (1994) considered smooth kernel estimator of the density $g(t)$. He cites many areas of applications, including reliability and actuarial science. A large subclass of these applications concerns with the situation where the kernel function for the $U-$ statistic is non-negative. This may happen because the random variables involved are non-negative, such as in insurance claims (see Chaubey, Garrido and Trudeau (1987)) or the kernel $h$ may represent some sort of metric which is necessarily non-negative. In fact in all three examples explored in detail in Frees' paper, the kernal function happens to be non-negative. Hence, we are going to pay special consideration to this case here. The natural estimator of $\theta(F)$ given by $G_n(t)$ is another $U-$ statistic as explained earlier. It is proved in Sen (1983) [see also Helmers, Janssen and Serfling (1988)] that

$$\sup_{x\in\mathbf{R}} |G_n(t) - G(t)| \to 0 \ a.s. \text{ as } \ n \to \infty.$$

but $G_n$ is not smooth, whereas $G$ may be absolutely continuous with density $h$ with respect to Lebesgue measure. Hence, interest lies in its its smooth version. Following Chaubey *et al.* (2012), we propose a smooth estimator of $G(t)$ as

$$\tilde{G}_n(t) = \int G_n(x) dK_{n,t}(x), \tag{1.10}$$

where $K_{n,t}$ represents a distribution function, continuous in $t$ such that

$$(i) \int x dK_{n,t}(x) = t, (ii) \lim_{n\to\infty} \int (x-t)^2 dK_{n,t}(x) = 0.$$

It can be easily shown that the proposed smooth estimator is also *almost sure* consistent and moreover,

$$\sup_{t\in\mathbf{R}} |\tilde{G}_n(t) - G(t)| \leq \sup_{t\in\mathbf{R}} |G_n(t) - G(t)|$$

Furthermore, $\int |d(G_n - G)| \to 0$ in probability implies that $|\theta_f(\tilde{G}_n) - \theta_f(G)| \to 0$ in probability for all uniformly bounded functions $f$ [see Radulović and Wegakamp (2003)].

The plan of the paper is as follows. In Section 2, we present a general class of smooth estimators of $G(t)$ derived from $G_n(t)$. This method has an important feature that it can incorporate the support of the density effectively and may avoid the boundary value problem in general. As a special case, it provides the popular kernel method of smoothing. For specific examples of estimating densities on bounded support we may refer to Bouezmarni and Rolin (2003), Chaubey and Sen (1996), Bagai and Prakasa Rao (1996) and Babu, Chaubey and Canty (2002). Here we generalize the method given in Chaubey and Sen (1996), proposed for smooth estimation of density and distribution functions for non-negative data. Section 3 considers the special case of the non-negative kernels, where we adapt the method in Chaubey and Sen (1996) that uses Poisson weights for smoothing $G_n$ as an estimator of $G(t)$. Sections 4 and 5 study some asymptotic properties of the new estimator and Section 6 presents some applications to some well-known examples. Finally, Section 7 presents a summary and some additional remarks.

# 2 A General Smooth Estimator of the Kernel Distribution Function

The following theorem is key to the motivation of the proposal in this paper.

**Theorem 2.1** (Lemma 1, §VII.1, Feller 1965). *Let $u$ be any bounded and continuous function. Let $K_{n,t}$ be a distribution function (continuous in $t$) with mean $t$ and variance $\delta_n^2$ then we have for $\delta_n \to 0$*

$$\tilde{u}(t) = \int_{-\infty}^{\infty} u(x) dK_{n,t}(x) \to u(t). \tag{2.1}$$

*The convergence extends to the entire range if $u(t)$ is monotone.*

Replacing $u(.)$ by $G_n(.)$ that is bounded but not continuous, in the above theorem motivates the following smooth estimator of $G(t)$,

$$\tilde{G}_n(t) = \int_{-\infty}^{\infty} G_n(x) dK_{n,t}(x) \tag{2.2}$$

Chaubey *et al.* (2012) have recently used this general approach for estimating the density and distribution function in the context of survival analysis and established the strong convergence property. We can establish the same in the context of the kernel distribution function $G(t)$, using the Glivenko-Cantelli type result established in Sen (1983) for $G_n(t)$.

**Theorem 2.2** *If $\delta \equiv \delta(n,t) \to 0$ for every fixed $t$ as $n \to \infty$ we have*

$$\sup_t |\tilde{G}_n(t) - G(t)| \overset{a.s.}{\to} 0 \tag{2.3}$$

*as $n \to \infty$.*

**Proof:** We have

$$|\tilde{G}_n(t) - G(t)| \leq |\tilde{G}_n(t) - \tilde{G}(t)| + |\tilde{G}^*(t) - G(t)|. \tag{2.4}$$

Also for every $t$

$$|\tilde{G}_n(t) - \tilde{G}(t)| \leq \max_x |G_n(t) - G(t)| \int_{-\infty}^{\infty} dK_{n,t}(x). \tag{2.5}$$

From Sen (1983) Eq. (2.2) [see also Helmers, Janssen and Serfling (1988)] we have $\sup_t |\tilde{G}_n(t) - G(t)| \to 0,\ a.s.$ as $n \to \infty$. Hence, the result follows. □

**Remark 2.1** Technically, $K_{n,t}$ can have any support but it may be prudent to choose it so that it has the same support as the random variable under consideration; because this will rid of the problem of the estimator assigning positive mass to undesired regions. This approach was adapted in Babu, Chaubey and Canty (2002) using Bernstein polynomials for

estimating the density and distribution function with support $[0, 1]$ and by Chaubey and Sen (1996) for random variables with support $[0, \infty)$, which is considered later in more detail.
**Remark 2.2** For $\tilde{G}_n(t)$ to be a proper distribution function, $K_{n,t}$ must be decreasing function of $t$. This can be easily demonstrated for the ordinary empirical distribution function (*i.e.* $m = 1, h(x, t) = \mathbf{1}(x \leq t)$.) One can write $\tilde{G}_n(t)$ as

$$\tilde{G}_n(t) = \sum_{i=1}^{n} \int_{x_{i:n}}^{x_{(i+1):n}} \frac{i}{n} dK_{n,t}(x) \tag{2.6}$$

$$= \sum_{i=1}^{n} \frac{i}{n} \big( K_{n,t}(x_{(i+1):n}) - K_{n,t}(x_{i:n}) \big) \tag{2.7}$$

$$= 1 - \frac{1}{n} \sum_{i=1}^{n} K_{n,t}(x_{i:n}), \tag{2.8}$$

$x_{i:n}$ denotes the $i$th order statistic from $(x_1, x_2, ..., x_n)$. A smooth estimator of the density function $g(t)$ as derived from this expression becomes,

$$\tilde{g}_n(t) = \frac{d\tilde{G}_n(t)}{dt} = -\frac{1}{n} \sum_{i=1}^{n} \frac{d}{dt} K_{n,t}(X_i)., \tag{2.9}$$

which shows that $K_{n,t}(x)$ must be a decreasing function of $t$.
**Remark 2.3** The representation given by Eq. (2.9) can also be used to have another look at the popular kernel estimator as follows. Let $K_{n,t}(.)$ be given by

$$K_{n,t}(x) = K\left(\frac{x - t}{\delta_n}\right),$$

which has mean $t$ and variance $\delta_n^2$, where $K(.)$ is a distribution function with mean zero and variance 1. Then, the estimator $\tilde{g}_n(t)$ becomes

$$\tilde{g}_n(t) = \frac{1}{n\delta_n} \sum_{i=1}^{n} k\left(\frac{X_i - t}{\delta_n}\right),$$

which is the well known kernel estimator with kernel $k(x) = \frac{d}{dx}K(x)$ and window width, which has been vigorously studied in literature. It has been studied in the context of the present paper by Frees (1984), hence we will concentrate more on non- negative $U-$ functionals.

# 3 Smooth Estimator of the Kernel Distribution and Density Function for Nonnegative Support

If the kernel function $h$ is defined on $\mathbf{R}^+$, then we use the following lemma which is a special case of theorem (2.1), where $K_{n,t}$ is obtained by attaching a probability $p_k(t\lambda_n) = e^{-\lambda_n t} \frac{(\lambda_n t)^k}{k!}$ to the point $k/\lambda_n$).

**Theorem 3.1** (Lemma 1, §VII.1, Feller 1965) *Let $u(t)$ be a bounded function on $[0,\infty)$, then the function $\tilde{u}(t)$ defined by*

$$\tilde{u}(t) = e^{-\lambda_n t} \sum_{j=0}^{\infty} u\left(\frac{k}{\lambda_n}\right) \frac{(\lambda_n t)^j}{j!} \tag{3.1}$$

*converges uniformly to $u(x)$ in any finite sub-interval of $[0,\infty)$, as $\lambda \to \infty$. This convergence extends to the whole interval if the function $u(x)$ is monotone.*

Since, $G_n(x)$ is bounded and monotone, we hope to adopt the above lemma in a stochastic set-up, *i.e.* using $G_n(x)$ in place of $u(x)$. This motivates the following estimator of $G(t)$

$$\tilde{G}_n(t) = \sum_{j=0}^{\infty} p_j(t\lambda_n) G_n(j/\lambda_n), \; x \in \mathbf{R}^+, \tag{3.2}$$

where,

$$p_j(\mu) = \mathrm{e}^{-\mu} \frac{(\mu)^j}{j!}, j = 0, 1, 2, ... \tag{3.3}$$

By allowing $\{\lambda_n\}$ to be possibly stochastic, *e.g.*, $\lambda_n = \max(X_1, ... X_n) = X_{n:n}$ and noting that $G_n(\cdot)$ is itself a random function, we gather from (3.2) that $\tilde{G}_n$ is generally a stochastic convex combination of $H_n(\cdot)$. With this choice of $\lambda_n$, the infinite sum in Eq. (3.2) is actually finite, since, in this case, $S_n(j/\lambda_n) = 0$, for $j \geq n$. where, $S_n(x) = 1 - G_n(x)$. In general also, for any choice of $\lambda_n$, let $n^* = [\lambda_n X_{n:n}]$, then again, $S_n(j/\lambda_n) = 0$, for $j \geq n^*$. In the finite sum the weights do not add to unity, and due to this reason, Chaubey and Sen (1996) considered truncated weights. However, in the following exposition we will dispense with un-truncated weights.

The estimator $\tilde{G}_n(t)$, is infinitely differentiable and therefore, it provides a very smooth estimator of the *kernel distribution function* $G(t)$. Moreover, it is a proper distribution function as can be easily demonstrated. First, it is clear that $0 \leq \tilde{G}_n(t) \leq 1$. Next, we show below that it is monotone. To see this define

$$P_j(t) = \sum_{i=j}^{\infty} p_i(t), j = 0, 1, 2, .... \tag{3.4}$$

then we can write

$$\tilde{G}_n(t) = \sum_{k=1}^{\infty} P_j(t\lambda_n) a_{nj}, \tag{3.5}$$

where

$$a_{nj} = G_n\left(\frac{j}{\lambda_n}\right) - G_n\left(\frac{j-1}{\lambda_n}\right), \; j = 1, 2, ... \; . \tag{3.6}$$

Now, since, $G_n(.)$ is non-decreasing, $a_{nj} \geq 0$ for all $j \geq 1$. Furthermore, since for an integer $\alpha > 0$

$$e^{-x} \sum_{j=0}^{\alpha} \frac{x^j}{j!} = \frac{1}{\Gamma(\alpha+1)} \int_x^\infty e^{-y} y^\alpha dy, \tag{3.7}$$

we can write

$$P_j(\lambda_n t) = \frac{1}{\Gamma(j)} \int_0^{\lambda_n t} e^{-y} y^{j-1} dy, \tag{3.8}$$

and it becomes clear that $P_j(\lambda_n t)$ is increasing in $t$. □

Recently, Chaubey, *et al.* (2012) have used a generalization of the Hille's Lemma as given in Feller (1968) (Chapter V, pp. 229) where the discrete weights have been replaced by non-negative density functions satisfying some regularity properties. In this paper we will deal only with Poisson weights; the alternative method of generating weights from a continuous asymmetric distribution will be discussed elsewhere.

Since, $\tilde{G}_n(t)$, is a proper smooth distribution function, we propose the following smooth estimator of the density function $h(x)$;

$$\begin{aligned} \tilde{g}_n(t) &= (d/dt)\tilde{G}_n(t) \\ &= \lambda_n \sum_{j=0}^{n*} p_j(t\lambda_n) \left[G_n((j+1)/\lambda_n) - G_n(j)/\lambda_n)\right], \ t \in \mathbf{R}^+. \end{aligned} \tag{3.9}$$

In the following section we prove the asymptotic properties of the resulting smooth processes.

# 4 Asymptotic Properties of $\tilde{G}_n(.)$

For a given (non-degenrate) $U$ statistic with kernel $h$ of order $m$, denoted by $U_n^h$, define a stochastic process

$$\{U_n^h - P^m h; h \in \mathcal{H}\}, \tag{4.1}$$

where the random variables $X_1, ..., X_n$ are defined on the probability space $(S, \mathcal{S}, P)$, and $P^m h$ denotes the expectation as in (1.1). Arcones and Giné (1993) provide necessary and sufficient conditions for limit theorems for general $U-$processes as given above. For the class of functions, $\mathcal{F}_1 = \{q(X_1, ..., X_m) = I[h(X_1, ..., X_m) \leq x], x \in R\}$, we have

$$\sup_{\mathcal{F}_1} |U_n^q - P^m q| = \sup_{t \in R} |G_n(t) - G(t)|.$$

Hence, using their theorem 3.6 (see also their example 3.10), since indicator functions are totally bounded, we get,

**Theorem 4.1**

$$\sup_{x\in R} |G_n(t) - G(t)| \to 0 \ \ a.s \ \ as \ n \to \infty. \tag{4.2}$$

The above theorem aids in proving the following theorem about the smooth estimator $\tilde{G}_n(t)$.

**Theorem 4.2** *Let $\lambda_n$ be a sequence of positive constants converging to $\infty$ as $n \to \infty$, then*

$$\sup_{x\in R} |\tilde{G}_n(t) - G(t)| \to 0 \ \ almost\ sure \ \ as \ n \to \infty. \tag{4.3}$$

**Proof:** Consider

$$\tilde{G}(t) = \sum_{j=0}^{\infty} p_j(t\lambda_n) G(j/\lambda_n), \tag{4.4}$$

then we note that by Hille's theorem,

$$\sup_{t\in R^+} |\tilde{G}(t) - G(t)| \to 0 \ \ \text{as} \ \ n \to \infty. \tag{4.5}$$

Furthermore,

$$|\tilde{G}_n(t) - G(t)| \leq |\tilde{G}_n(t) - \tilde{G}(t)| + |\tilde{G}(t) - G(t)| \tag{4.6}$$

Also for every $x$

$$|\tilde{G}_n(t) - \tilde{G}(t)| \leq \max_t |G_n(t) - G(t)|, \tag{4.7}$$

converges to zero using theorem 3.1, we claim from Eq. (4.6) and Eq. (4.5) that

$$\sup_t |\tilde{G}_n(t) - G(t)| \to 0, \ a.s.,$$

the result follows. □

**Remark 4.1:** Convergence of $|G_n(t) - G(t)|$ in the sup norm may also be demonstrated in a simpler way by adhering to the reverse sub martingale property of $U-$statistics (see Lee (1990), Chap. 3, Theorem 3.or Sen (1981), Theorem 3.2.1). For every $t > 0$, $G_n(t)$ is nondegenerate, however for $t \leq 0, G_n(t) = 0$ Therefore, we can not use this theorem directly. Note, however, that for any $t \in (0, t^*), t^* > 0$ using martingale convergence theorem for nondegenarate bounded kernels, we have

$$\lim_{n\to\infty} \sup_{0<t\leq t^*} |G_n(t) - G(t)| = 0, \quad a.s..$$

For, $x = 0$, we have $G_n(0) = G(0) = 0$, hence, the convergence in the above equation can be extended to the compact set $[0, x^*]$. Further, since,

$$|G_n(t) - G(t)| = |S_n(t) - S(t)|,$$

and both $S_n(t)$ and $S(t)$ are decreasing functions (to 0) for any given $\epsilon > 0$, there exists $X_\epsilon$, such that $S(t) < \epsilon/2$ as well as $S_n(t) < \epsilon/2$ for all $t > t_\epsilon$, *i.e.*,

$$|G_n(t) - G(t)| < \epsilon \text{ for all } t > t_\epsilon.$$

Since, $\epsilon$ is arbitrary, this implies that

$$\sup_{t \in R^+} |G_n(t) - G(t)| \to 0 \ \ a.s. \text{ as } \ n \to \infty.$$

**Remark 4.2:** The above theorem just proves the almost convergence of the smooth estimator. To get an idea about the rate of convergence, we may use the result in Silverman (1976) or Sen (1983) and conclude that

$$\sup_{t \in R^+} |\tilde{G}_n(t) - G(t)| = O_p(n^{-1/2}).$$

This rate can be improved due the result of Dehling *et al*. (1987) (see their Corollary 2), we have with probability 1, as $n \to \infty$,

$$\sup_{t \in R^+} |\tilde{G}_n(t) - G(t)| = O\left(\frac{(\log\log n)^{1/2}}{n^{1/2}}\right). \tag{4.8}$$

Note also that the above rate is better than that reported in Sen (1981), we claim that the same (or better) rate holds for the smoothed estimator, because

$$\begin{aligned}\sup_{t \in R^+} |\tilde{G}_n(t) - G(t)| &= \sup_{t \in R^+} |\int (G_n(y) - G(y)) dP_{\lambda_n t}(y)| \\ &\leq \sup_{t \in R^+} |G_n(t) - G(t)|,\end{aligned}$$

where $P_{\lambda_n t}$ denotes the measure induced by the Poisson-weights. Because of the fact that the process

$$\{\sqrt{n}(G_n(t) - G(t)), t \in R^+\} \to Gaussian,$$

(see Theorem 4.10 of Arcones and Giné (1983)), we can claim that

**Theorem 4.3** *For $\lambda_n \to \infty$, as $n \to \infty$, we have*

$$\{\sqrt{n}(\tilde{G}_n(t) - G(t)), t \in R^+\} \to Gaussian.$$

**Remark 4.3:** The above theorem may also be established using a result parallel to that established in Chaubey and Sen (1996) using Bahadur-Kiefer representation for U-quantiles (see Choudhury and Serfling (1988)) as given below. For $\lambda_n \to \infty, n^{-1}\lambda_n \to 0$ as $n \to \infty$,

$$\sup_{t \in R^+} |\tilde{G}_n(t) - G_n(t)| = O(n^{-3/4}(\log n)^{3/4}),$$

that implies $\sqrt{n}(\tilde{G}_n(t) - G_n(t)) \to 0$ a.s. as $n \to \infty$ and the asymptotic normality of $\sqrt{n}(\tilde{G}_n(t) - G(t))$, consequently follows from that of $\sqrt{n}(G_n(t) - G(t))$, since

$$\sqrt{n}(\tilde{G}_n(t) - G(t)) = \sqrt{n}(\tilde{G}_n(t) - G_n(t)) + \sqrt{n}(G_n(t) - G(t)).$$

# 5 Asymptotic Properties of $\tilde{g}_n(.)$

We can claim almost sure convergence of the derived density estimator as well, however, the rate at which $\lambda_n \to \infty$ has to be controlled. We can establish the following.

**Theorem 5.1** *Assume that $g(x)$ is bounded and absolutely continuous. Also, let $g(x)$ admit a bounded derivative $g'(\cdot)$ a.e. on $\mathbf{R}^+$ and let $\lambda_n \to \infty$ such that $n^{-1/2}\lambda_n \to 0$, then we have*

$$\|\tilde{g}_n - g\| \;\to\; 0 \quad a.s. \;\; as\; n \to \infty. \tag{5.1}$$

**Proof:** First, note $\tilde{S}_n(t) = 1 - G_n(t)$ is non–increasing in $t \in \mathbf{R}^+$, and $\tilde{g}_n(t)$ is continuous a.e. thus, for every $\eta > 0$, there exists a $c(= c_\eta < \infty)$, such that

$$\tilde{S}_n(t) - \tilde{S}_n(t+y) < \eta \;\; \text{a.s.}, \;\forall\; x \geq c,\; y \geq 0. \tag{5.2}$$

Since, the left hand side of (5.2) is equal to $\int_t^{t+y} \tilde{g}_n(u)du$, a direct application of the first mean value theorem (of calculus) yields that by choosing $y$ such that $\eta/y$ is small, as $n \to \infty$,

$$\tilde{g}_n(t) \leq \eta' \;\; \text{a.s. for every x} \geq \text{c}, \tag{5.3}$$

where $\eta' \to 0$ as $\eta \to 0$. Also, repeating the same argument with $S(x)$, we have $h(x) < \eta'$, $\forall x \geq c$. Consequently, we have $\sup\{|\tilde{f}_n(x) - f(x)| : x \geq c\} \leq 2\eta'$, a.s. as $n \to \infty$. Thus to prove (5.1), it suffices to show that

$$\sup\{|\tilde{g}_n(t) - g_n^*(t)| : \; 0 \leq t \leq c\} \to 0 \;\; \text{a.s., as } n \to \infty. \tag{5.4}$$

To see this, for any $x : 0 \leq x \leq c$, consider the function $g_n^*(t)$ defined by

$$g_n^*(t) = \lambda_n \sum_{j=0}^{\infty} p_j(t\lambda_n) \left(G((j+1)/\lambda_n) - G(j/\lambda_n)\right). \tag{5.5}$$

First we see by expanding $G((j+1)/\lambda_n)$ in a Taylor series, that

$$g_n^*(t) = \sum_{j=0}^{\infty} p_j(t\lambda_n) \left[g(j/\lambda_n) + O(\lambda_n^{-1})\right] = \tilde{g}(t) + O(\lambda_n^{-1}), \tag{5.6}$$

hence, by the use of the Hille's theorem

$$|g_n^*(t) - g(t)| \to 0 \;\; a.s. \;\; \text{as} \;\; n \to \infty. \tag{5.7}$$

Next, we see that,

$$\begin{aligned} |\tilde{g}_n(t) - g(t)| &\leq |\tilde{g}_n(t) - g_n^*(t)| + |g_n^*(t) - g(t)| \\ &\leq \sup |G_n(t) - G(t)|\lambda_n + |g_n^*(t) - g(t)|. \end{aligned} \tag{5.8}$$

Now the result follows using (5.7), (4.8) and the condition on $\lambda_n$ given in the theorem. □

The following theorem establishes the asymptotic normality of $\tilde{g}_n(t)$. We would like to remark that the asymptotic limit is same as that achieved by using the kernel method of density estimation as investigated in Frees (1994).

**Theorem 5.2** *Assume that $g(t)$ is bounded and absolutely continuous. Also, let $g(t)$ admit a bounded derivative $g'(\cdot)$ a.e. on $\mathbf{R}^+$ and let $\lambda_n \to \infty$ such that $n^{1/2}\lambda_n^{-1} \to 0$. Further let*

$$G_1(x;t) = P[g(x, X_2, ..., X_m) \leq t | X_1 = x]$$

*and assume that $g_1(x;t) = \frac{d}{dt}G_1(x;t)$ exists and is bounded with $E[g_1(X_1;t)] < \infty$, then we have*

$$\sqrt{n}(\tilde{g}_n(t) - g(t)) \to^{\mathcal{D}} N(0, m^2\xi_1) \tag{5.9}$$

*where*

$$\xi_1 = \mathrm{Var}(g_1(X_1)).$$

**Proof:** The basic step of the proof is the following theorem on the asymptotic distribution of a general U-statistic [see Theorem 12.3, van der Vaart (1998), pp. 162], where $h(.)$ the general kernel of the $U-$ statistic for estimating $\theta \equiv \theta(F)$.

**Theorem:** *If $Eh^2(X_1, ..., X_m) < \infty$ then $\sqrt{n}(U - \theta - \hat{U}) \xrightarrow{P} 0$, where $\hat{U}$ is the projection of $U - \theta$ onto the set of all statistics of the form $\sum_{i=1}^n g_i(X_i)$, that is given by*

$$\hat{U} = \sum_{i=1}^n E(U - \theta | X_i) = \sum_{i=1}^n \frac{m}{n} h_1(X_i),$$

*where*

$$h_1(x) = Eh(x, X_2, ..., X_m) - \theta.$$

*Consequently, $\sqrt{n}(U - \theta) \to^{\mathcal{D}} N(0, m^2\mathrm{Var}h_1(X_1))$.*

Using this result and realizing that $\tilde{g}_n(t)$ is of a $U-$statistic structure, we can claim that

$$\sqrt{n}(\tilde{g}_n(t) - g_n^*(t)) \to^{\mathcal{D}} N(0, m^2\sigma^2),$$

where

$$\sigma^2 = \lim_{n\to\infty} \mathrm{Var}\lambda_n \sum_{k=0}^\infty p_k(\lambda_n t) P[g(X_1, ..., X_m) \in (k/\lambda_n, (k+1)/\lambda_n] | X_1],$$

and

$$g_n^*(t) = \lambda_n \sum_{j=0}^\infty p_j(t\lambda_n) \left[ G((j+1)/\lambda_n) - G(j)/\lambda_n) \right].$$

Defining $G_1(x;t) = P[g(X_1, ..., X_m) \leq t | X_1 = x]$ as the conditional distribution function of $g$ given $X_1 = x$ and the corresponding density by $G_1(x;t) = (d/dt)G_1(x;t)$, we have

$$\begin{aligned} \sum_{k=0}^\infty p_k(\lambda_n t) P\left[ g(X_1, ..., X_m) \in (\frac{k}{\lambda_n}, \frac{k+1}{\lambda_n}] | X_1 \right] &= \sum_{k=0}^\infty p_k(\lambda_n t)[H_1(G_1; \frac{k+1}{\lambda_n}) - G_1(X_1; \frac{k}{\lambda_n})] \\ &= (1/\lambda_n) \sum_{k=0}^\infty p_k(\lambda_n t) g_1(X_1; \frac{k}{\lambda_n}) + O(1/\lambda_n^2) \end{aligned}$$

Next, it is easy to see that we show that

$$g_n^*(t) = \tilde{g}(t) + O(1/\lambda_n). \tag{5.10}$$

Further we can show that

$$\tilde{g}(t) = g(t) + O(n^{-1}\log n). \tag{5.11}$$

Using (5.11) we have

$$\sigma^2 \;=\; \mathrm{Var}(g_1(X_1;t))$$

and using (5.10) we find that since $n^{1/2}\lambda_n^{-1} \to 0$

$$\sqrt{n}(\tilde{g}_n(t) - g_n^*(t)) \to^{\mathcal{D}} \sqrt{n}(\tilde{g}_n(t) - g(t))$$

and the result stated in the theorem follows. □

**Remark 5.1.** Similar properties as studied for $G_n(t)$ hold for the von-Mises' differentiable functional estimator of $G(t)$ given by

$$G_{nV}(t) = \frac{1}{n^m}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\dots\sum_{i_m=1}^{n} I[h(X_{i_1}, ..., X_{i_m}) \le t].$$

As commented in Frees (1994), "the choice between the two estimators depend on the application on hand." Jones and Sheather (1991) provide arguments in favor of $G_{nV}(t)$ for estimating integrated squared density derivative, however, Frees (1994) considers the use of $G_n(t)$ more appropriate in studying the distribution of spatial statistics.

**Remark 5.2.** It is clear that for the density estimation here a large number of computations may be required. To circumvent the problem of such large scale computations, we may use the idea described in Blom (1976) that is described in Frees (1994). This involves choosing a positive integer $B = B(n)$ such that $B \to \infty$ as $n \to \infty$. Based on the observed sample, $B$ independent draws are made and for $b = 1, ..., B$, $m$ draws are made without replacement to get the observations $(X_1^{*b}, ..., X_m^{*b})$. The proposed estimator of $G(t)$ is given by

$$G_{nR}(t) = \frac{1}{R}\sum_{b=1}^{B} I[h(X_1^{*b}, ..., X_m^{*b}) \le t]$$

# 6 Examples

## 6.1 Redwood Locations

This example concerns the density of locations of 62 redwood seedlings in a unit square as reported in Diggle (1983). The data is now freely available in R-package `statspat Reference`. It is commonly believed that the locations are not randomly scattered over the unit square as it is apparent from Figure 1. It was recommended in Diggle (1983) to examine the distribution of the interpoint distances in order to evaluate the degree of spatial randomness. Frees (1994) used kernel density estimator based on the $\binom{62}{2=1,891}$ interpoint distances, $d((x_1, y_1), (x_2, y_2)) = ((x_1 - x_2)^2 + (y_1 - y_2)^2)^{1/2}$ that is superimposed on the corresponding histogram in Figure 2, labeled as hard line curve. This curve was compared to the reference distribution $g_0(t)$ given by Bartlett (1964) as depicted in Figure 3:

$$g_0(t) = \begin{cases} (2t)(\pi - 4t + t^2) & \text{for} \quad 0 \leq t \leq 1, \\ (2t)(-2 - t^2 + 4(t^2 - 1)^{1/2} + 2sin^{-1}(2t^{-2}2 - 1) & \text{for} \quad 1 < t \leq \sqrt{2} \end{cases}$$

The kernel density estimator shows a clear departure from the reference distribution, however, it shows stretching below zero that is not a desirable feature as the distances are nonnegative. On the other hand, the estimator based on Poisson weights produces almost the same estimator, except that the undesirable feature near zero is removed. It is natural to ask if the Poisson weights based estimator is consistent. One can see $\tilde{g}_n(0) = \lambda_n G_n(1/\lambda_n)$ that approximates the density $g(t)$ near zero and is not necessarily zero unless $1/\lambda_n$ is smaller than the $\min(g(X_{i_1}, ..., X_{i_m})$ over all possible combinations $1 \leq i_1 < i_2 < ... < i_m \leq n$. We have selected $\lambda_n = 180$, a choice obtained by trial and error,so as to be close to the kernel estimator. Alternatively, data based optimal value of the smoothing parameter may be obtained using the cross-validation as explored in Chaubey and Sen (2009). It is clear that the density estimator is in sharp contrast to the reference distribution, however it may not be visually as clear by comparing the distribution functions.

## 6.2 Inter State Centroid Distances

Inter-population distances are of interest in geographical studies in order to quantify the separation between two populations. Frees (1994) fitted the kernel density estimator and concluded the log-normal shape commonly assumed in disciplines studying with population movements. We collected the data on centroids of 51 US states from `MAPTECH` `http://www.maptechnavigation.com/` website and a SAS program was used to compute the geodesic distance between the pairwise centroids. The histogram along with the kernel density estimator and Poisson weights based estimator are given in Figure 4. It is surprising to see that the density does not resemble lognormal as claimed in Frees (1994). In any case for the present data, the kernel density estimator is not adequate at all. In trying to allocate

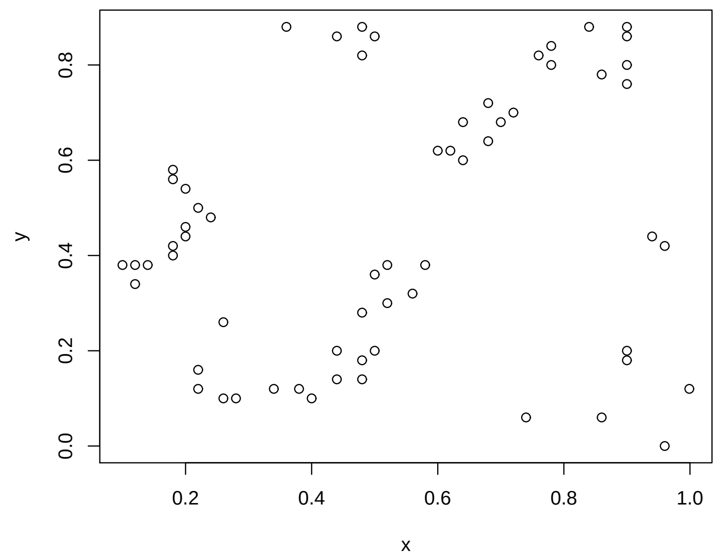


Figure 1: Scatterplot of the Locations of 62 Redwood Seedlings, Rescaled to Unit Square

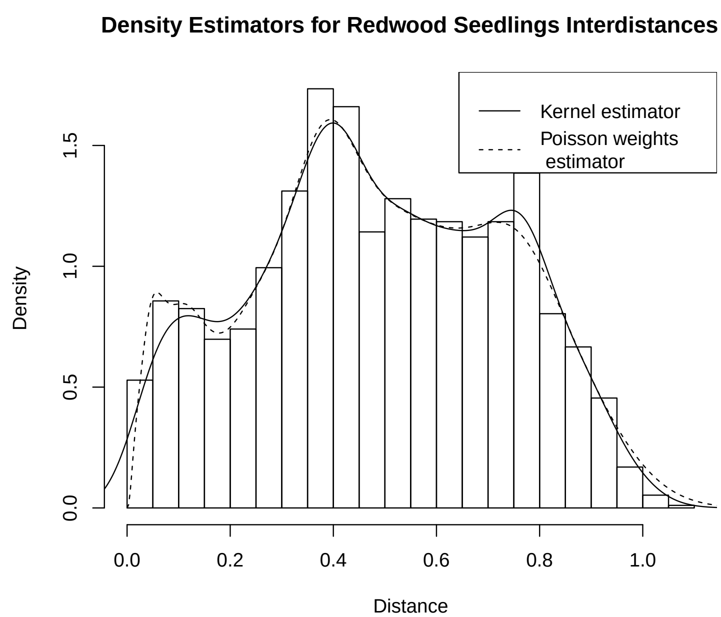


Figure 2: Histogram of 1,891 Interpoint Distances of 62 Redwood Seedlings Locations. Kernel Density Estimator and the New Estimator are Superimposed on the Histogram

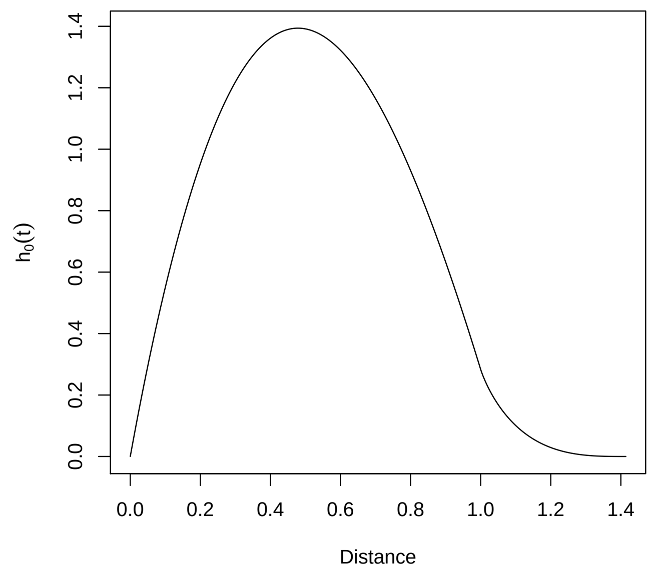


Figure 3: Bartlett's Reference Density of Interpoint Distances on a Unit Square

some density below zero, it completely distorts the picture, where as the proposed estimator adequately picks the high mass near zero.

## 6.3 Convolution of Insurance Claims

Frees (1994) considered estimating the densities of $m$ convolutions for $m = 2, 3, 4$ and 5 of insurance claims collected from 33 female patients to illustrate the effect of an additional expected claim. We reproduce these along with the estimator studied in this paper in Figures 6, 7, 8 and 9. For reference purpose, the original 33 claims are plotted in Figure 5. As expected the bimodal nature of the original distribution flattens a bit. In practice, the risk manager can use these figures with the best guess for expected number of claims. Through all these figures, the the new estimator emerges as correcting the boundary bias of the kernel estimator near lower tail. In case the observations are far from zero, the two methods seem to provide almost identical shapes. The kernel estimator does not integrate to unity that can be corrected through various methods (see Silverman (1976)), however the new estimator takes care of this in a natural way.

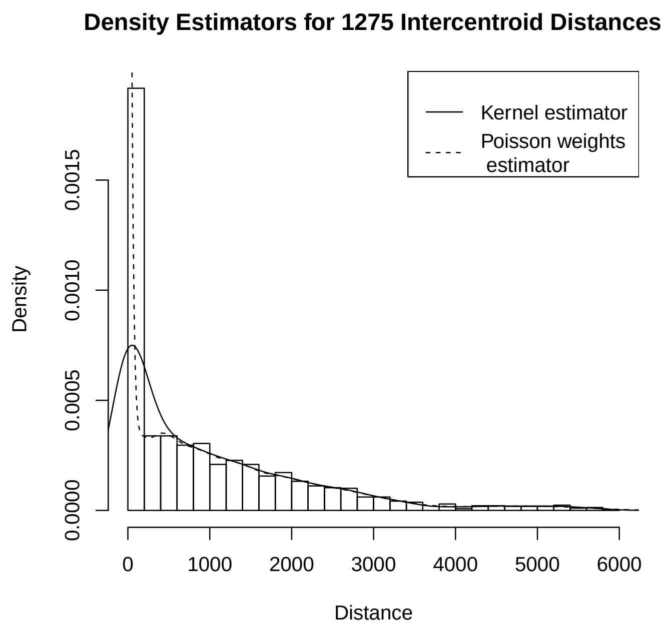


Figure 4: Histogram of 1275 Intercentroid Distances of 51 States

# 7 Summary and Concluding Remarks

Here we have considered the use of Poisson weights in smoothing the kernel distribution function for non-negative kernels. This simple method has the same asymptotic properties as the kernel method that may be inappropriate for the non-negative kernel involved in the U-statistic. This shortcoming is naturally taken care of the new-estimator. Where as the kernel estimator may give an impression of a unimodal density due to its nature to stretch in the direction of negative values, the new estimator may be able to capture the peak of the density properly near zero. Another alternative that has been recently investigated by Chaubey *et al.* (2012), namely that of using asymmetric kernels in the context of density estimation of non-negative random variables, can be also adapted in the present context. This method however has to be specifically tailored to provide correct behaviour near zero for the densities which may not be zero near zero and therefore requires two smoothing parameters. The Poisson estimator does not have this difficulty and the determination of the single smoothing parameter can be easily handled using modern optimising software as discussed in Chaubey and Sen (2009) in the context of density estimation for the *i.i.d.* setup.

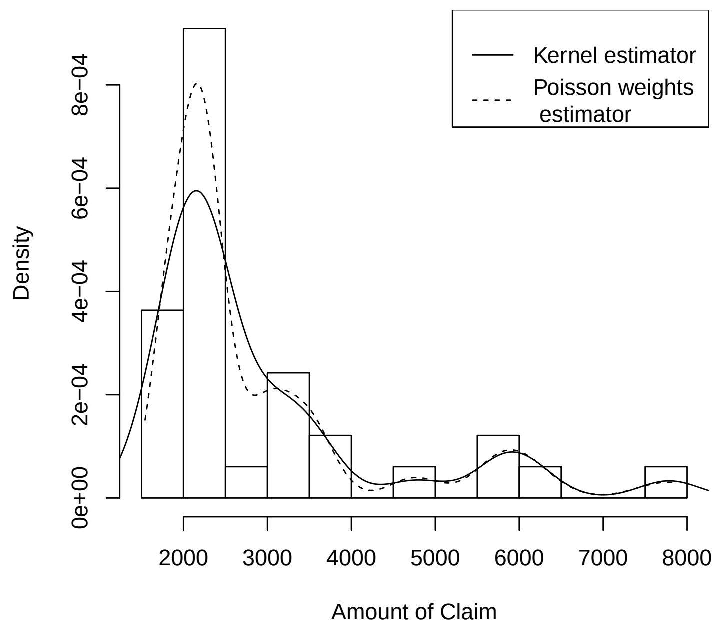


Figure 5: 1989 Hospital Charges for 33 Female Patients

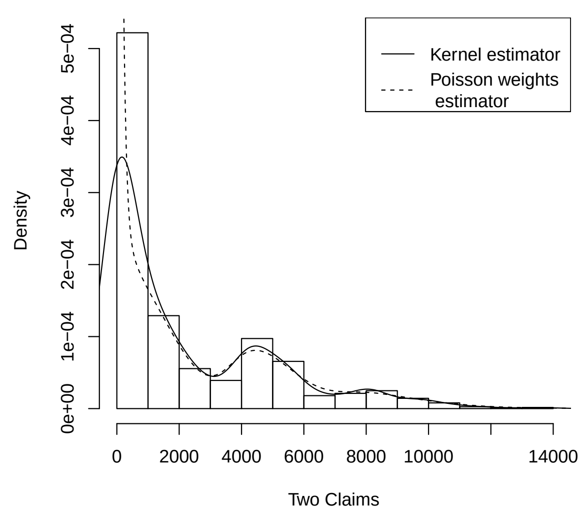


Figure 6: Distribution of Sum of Two Claims of Hospital Charges

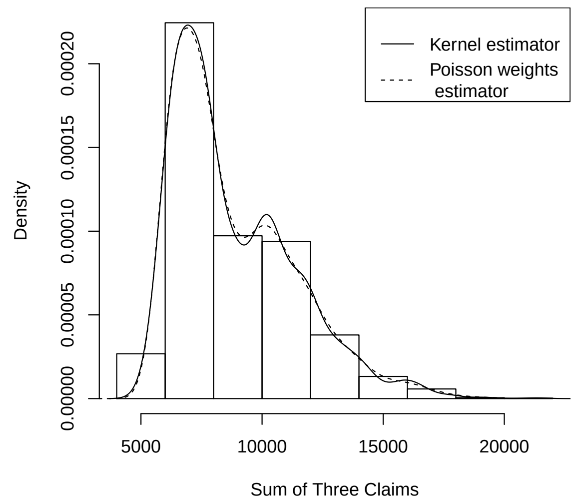


Figure 7: Distribution of Sum of Three Claims of Hospital Charges

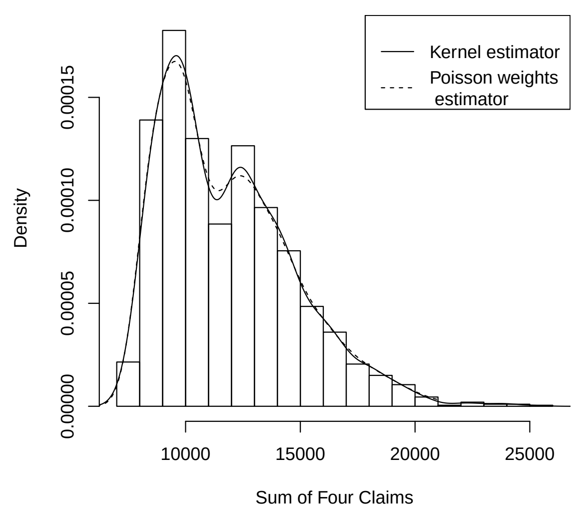


Figure 8: Distribution of Sum of Four Claims of Hospital Charges

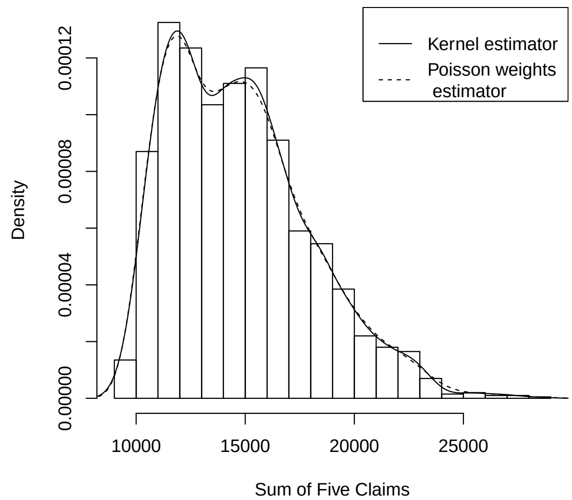


Figure 9: Distribution of Sum of Five Claims of Hospital Charges

# References


[1] Arcones, M.A. (1993). The law of iterated logarithm for $U-$ processes. *Journal of Multivariate Analysis* **47,** 139-151.

[2] Arcones, M.A. (1996). The Bahadur-Kiefer representation for $U-$quantiles. *Annals of Statistics* **24,** 1400-1422.

[3] Arcones, M.A. and Giné, E. (1993). Limit theorems for $U-$ processes. *The Annals of Statistics* **21,** 1494-1542.

[4] Babu, G.J., Canty, A.J. and Chaubey, Y.P. (2002). Application of Bernstein Polynomials for smooth estimation of a distribution and density function. *Jour. Statist. Plann. and Inference* **105** 377-392.

[5] Bagai, I. and Prakasa Rao, B.L.S. (1996). Kernel Type Density Estimates for Positive Valued Random Variables. *Sankhya* **A57** 56-67.ibitemarc93

[6] Bickel, P.J. and Lehmann, E.L. (1979). Descriptive statistics for nonparametric models IV. Spread. In *Contributions to Statistics (Hajék Memorial Volume)* (J. Jurečová, ed.) 33-40, Academia, Prague.

[7] Bouezmarni, T. and Rolin, J.-M. (2003). Consistency of the beta kernel density function estimator. *The Canadian Journal of Statistics* **31,** 89-98.

[8] Chaubey, Y.P., Garrido, J. and Trudeau, S. (1998). On the computation of aggregate claims distribution: Some new approximations. *Insurance: Mathematics and Economics* **23,** 215-230.

[9] Chaubey, Y.P. and Sen, P.K. (1996). On smooth estimation of survival and density functions. *Statistics and Decisions* **14**, 1-22.

[10] Chaubey, Y.P. and Sen, P.K. (2009). On the selection of the smoothing parameter in poisson smoothing of histogram estimator: Computational aspects. *Pakistan Jour. Stat.* **25,** 385–401.

[11] Chaubey, Y. P., Li, J., Sen, A. and Sen, P.K. (2002). A new smooth density estimator for non-negative random variables. *Jour. Ind. Statist. Assoc.* **50,** 83–104.

[12] Choudhury, J. and Serfling, R. (1988). Generalized order statistics, Bahadur representations, and sequential fixed-width confidence intervals. *J. Statist. Plann. Inf.* **19,** 269–282.

[13] Dehling, H., Denker, M. and Philipp, W. (1987). The almost sure invariance principle for the empirical process of U-statistic structure. *Annales de l'Inststitute Henri Poincaré* **23,** 121-134.

[14] Feller, W. (1965). *An Introduction to Probability Theory and its Applications, Vol. II.* John Wiley & Sons, New York.

[15] Frees, E.W. (1994). Estimating densities of functions of observations. *Jour. Amer. Statist. Assoc.* **89**, 517-525.

[16] Gastwirth, J.L. (1973). A new index of income inequality , *Proceedings of the I.S. I. Meeting,* Vienna.

[17] Ghosh, M., Mukhopadhyay, N. and Sen, P.K. (1997). *Sequential Estimation.* John Wiley & Sons, New York.

[18] Grassberger, P. and Procaccia, I. (1983). Characterization of strange attractors. *Phys. Rev. Lett.* **50,** 346-349.

[19] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.* **19,** 293-325.

[20] Helmers, R., Janssen, P. and Serfling, R. (1988). Glivenko-Cantelli properties of some generalized empirical DF's and strong convergence of generalized $L$-statistics. *Probab. Theory Related Fields* **79,** 75–93.

[21] Lee, A.J. (1990). *U-Statistics: Theory and Practice.* Marcel Dekker, New York.

[22] Liu, R.Y. (1990). On a notion of data depth based upon random simplices. *Annals of Statistics* **14,** 405-414. Stat

[23] Nolan, D. and Polalrd, D. (1987). $U-$ processes: rates of convergence. *The Annals of Statistics* **15,** 780-799.

[24] Nolan, D. and Polalrd, D. (1988). Functional limit theorems for $U-$ processes. *The Annals of Probability* **16,** 1291-1298.

[25] Radulović, D. and Wegakamp, M. (2003). Necessary and sufficient conditions for weak convergence of smooth empirical processes. *Statistics and Probability Letters* **61,** 321-336.

[26] Schneemeier, W. (1993). Weak convergence and Glivenko-Cantelli results for weighted empirical $U-$ processes. *The Annals of Statistics* **21,** 1170-1184.

[27] Sen, P.K. (1963). On the estimation of relative potency in dilution (-direct) assays by distribution free methods. *Biometrics* **19,**532- 552.

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

[29] Sen, P.K. (1983). On the limiting behaviour of empirical kernel distribution function. *Calcutta Statist. Assoc. Bul.***32,**1-8.

[30] Sen, P.K. (1986). The Gini coefficient and poverty indexes: some reconciliations. *Journal of American Statistical Assiociation* **81,** 1050-1057.

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

[32] Serfling, R.J. (1984). Generalized $L-$, $M-$ and $R-$ statistics. *Annals of Statistics* **12,** 76-86.

[33] Silverman, B.W. (1976). Limit theorems for dissociated random variables. *Adv. App. Prob.* **8** 806-819.

[34] Silverman, B.W. (1983). Convergence of a class of empirical distributions functions of dependent random variables. *Ann. Prob.* **11**745-751.

[35] Silverman, B.W. and Brown, T. C. (1978). Short distances, flat triangles and Poisson limits *J. Appl. Prob.* **15** 815-825.

[36] Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function. *Ann. Statist.* **28**, 461-482.