Poisson limits for empirical point processes

André Dabrowski, Gail Ivanoff and Rafał Kulik

Ottawa, 27 May 2008

Point processes

Define the scaled empirical point process on an independent and identically distributed sequence $\{Y_i : \ i \leq n\}$ as the random point measure with masses at $a_n^{-1}Y_i$:

$$N_A^{(n)} = \sum_{i=1}^{n} I_{\{Y_i \in A\}}.$$

Existing Methods for dealing with pp:

- Poisson limit approach (Resnick, ...),
- a strong approximation approach (Einmahl, Deheuvels ...)

Each of this method has some advantages and disadvantages. For example, Resnick's method is not suitable to study weak convergence of the local empirical process

$$L_{n,x}(t) = \sum_{i=1}^{n} I_{\{Y_i \in [x - ta_n, x + ta_n]\}}, \quad t \in [0,1], \tag{1}$$

as it is strong approximation method. On the other hand, this strong Poisson approximation is difficult to implement (or at least cannot be extended directly) if one wants to study the joint behavior of

$$(L_{n,x_1}(\cdot), \ldots, L_{n,x_m}(\cdot)),$$

i.e. when estimating the density of Y_1 simultaneously at $(x_1, \ldots, x_m)$.

Multiparameter martingales

The aim is to illustrate how to apply the multiparameter martingale theory of Ivanoff and Merzbach (2000) to study $weak$ Poisson limits for empirical point processes on R^d (or even more general topological spaces). This approach requires only the simple computation of so-called *-compensators to identify Poisson limits for scaled empirical point processes. The compensator method exploited here is particularly attractive since

- it is independent of the dimension of the underlying random vectors, and so easily generalizes results from the univariate to the multivariate case.

- In addition, the martingale approach allows one to handle the joint behavior at multiple points with ease through a judicious definition of the associated history (filtration).

$*$-compensator

The empirical point process has *-compensator $\Lambda^{(n)}$ where

$$\Lambda_A^{(n)} = \sum_{i=1}^{n} \int_A I_{\{u \in A_{Y_i}\}} (F(E_u))^{-1} dF(u).$$

Example: If $T = R_+$ then $A = A_t = [0, t]$ for some $t \geq 0$, $E_u = [u, \infty)$. By noting that $F(E_u) = P(Y_i \geq u) =: \bar{F}(u)$ we have the standard result.

Example: Let $Y_i = (Y_{i1}, Y_{i2})$, $i = 1, \ldots, n$. If $T = R_+^2$ then: $A = A_{\mathbf{t}} = [0, t_1] \times [0, t_2]$ for some $\mathbf{t} = (t_1, t_2)$, $E_{\mathbf{u}} = \{\mathbf{t}' : t_i' \geq u_i, i = 1, 2\}$, $\mathbf{u} = (u_1, u_2)$. By noting that $F(E_{\mathbf{u}}) = P(Y_{i1} \geq u_1, Y_{i2} \geq u_2) =: \bar{F}(\mathbf{u})$ we have

$$\Lambda_{\mathbf{t}}^{(n)} := \Lambda_{A_{\mathbf{t}}}^{(n)} = \int_{A_{\mathbf{t}}} \sum_{i=1}^{n} I\{Y_{i1} \geq u_1, Y_{i2} \geq u_2\} \frac{dF(\mathbf{u})}{\bar{F}(\mathbf{u})} \,.$$

Local empirical processes

Let $0 < q_1 < q_2 < \cdots < q_k < 1$ and $x_j = F^{-1}(q_j)$, $j = 1, \ldots, k$. Assume that

$$\lim_{x \searrow 0} \frac{P_F((x_j, x_j + xt])}{P_F((x_j, x_j + x])} = t^\alpha \quad \left(\lim_{x \searrow 0} \frac{P_F((x_j + xt, x_j])}{P_F((x_j - x, x_j])} = |t|^\alpha \right) ,$$

for all $t > 0$ and $t < 0$, respectively. Since the regular variation behavior is the same at either side of x_j, the scaling constant $a_n = a_{n,j}$ may be chosen as $F(a_n + x_j) - F(x_j) = n^{-1}$ for arbitrary $j = 1, \ldots, k$. Thus, the point processes are defined as $N^{(n)}(j) = \sum_{i=1}^n \delta_{a_n^{-1}[Y_i - x_j]}$, $j = 1, \ldots, k$. Note further, that $N_A^{(n)}(j)$ with $A = [-t, t]$ agrees with $L_{n,x_j}(t)$.

Corollary: The vector $(L_{n,x_1}(\cdot), \ldots, L_{n,x_k}(\cdot))$ of the local empirical processes converges weakly to the vector of independent Poisson processes with the same intensities $2\alpha|t|^{\alpha-1}$.

Local Density Estimation

Consider a sample $\{Y_1, Y_2, \ldots, Y_n\}$ with common marginal differentiable distribution F on $[0,1]$, and assume that its density f is positive on the range $[0,1]$. Let F_n denote the empirical distribution and define $[t]_n^+ = Y_{(k+1)}$ and $[t]_n^- = Y_{(k)}$ by $Y_{(k)} \le t < Y_{(k+1)}$. We put $[t]_n^+ = 0$ if $[t]_n^+ < Y_{(1)}$ and $[t]_n^- = 0$ if $[t]_n^+ > Y_{(n)}$. A naive nearest-neighbor estimator of the density:

$$\widehat{f}(n,t) = \frac{1}{n} / \left(([t]_n^+ - t) + (t - [t]_n^-) \right) . \tag{2}$$

We may conclude that

$$\widehat{f}(n,t)/f(t) = \frac{1}{f(t)\left(n([t]_n^+ - t) + n(t - [t]_n^-)\right)} \xrightarrow{\mathcal{D}} \frac{1}{f(t)(E_1 + E_2)} \tag{3}$$

where E_1 and E_2 are independent exponential variables of mean $1/f(t)$.

In particular, we have identified the limiting distribution of $\widehat{f}(n,t)/f(t)$ as Inverse Gamma, $\Gamma^{(-1)}(2,1)$. The mode, mean and variance of an Inverse Gamma density of parameters (α,β) are $\beta/(\alpha+1)$, $\beta/(\alpha-1)$ (for $\alpha>1$) and $\beta^2/((\alpha-1)(\alpha-2))$ (for $\alpha>2$), respectively. Thus we see that this naive estimator of $f(t)$ has mode $f(t)/3$, mean $f(t)$ and infinite variance.

This development can be easily extended to estimators based on the k lower nearest neighbors and k upper nearest neighbors. As above, asymptotically the spacings between consecutive neighbors are independent exponential variables with mean $1/f(t)$. The asymptotic joint density is the product of $2k$ exponentials, and the sufficient statistic is just the total distance from the lower kth-nearest neighbor of t, $[t]_n^{-k}$, to the upper kth-nearest neighbor, $[t]_n^{+k}$.

Corollary: The asymptotically uniformly minimum variance unbiased estimator based on k nearest neighbors $(k > 1)$ is

$$\widehat{f}_k(n,t) = \frac{(2k-1)/n}{[t]_n^{+k} - [t]_n^{-k}},$$

and $\widehat{f}_k(n,t)/f(t)$ has an asymptotic $\Gamma^{(-1)}(2k,1)$ density. Moreover, the limiting distribution of

$$\left\langle \frac{\widehat{f}_k(n,t_1)}{f(t)}, \frac{\widehat{f}_k(n,t_2)}{f(t)}, \ldots, \frac{\widehat{f}_k(n,t_m)}{f(t)} \right\rangle$$

is given by a vector of m independent $\Gamma^{(-1)}(2k,1)$ Gamma variables.

Consequently we can obtain the limiting distribution of expressions such as approximate integrals,

$$\widehat{E}(g(Y)) = \sum_{i=1}^{m} g(t_i)(\widehat{f}_k(n, t_i)),$$

even for arbitrary dimension with appropriate norming.

Remark: We see that $\widehat{f}_k(n,t)/f(t)$ still has an Inverse Gamma distribution, but with finite variance for $k > 1$. It has asymptotic variance $1+1/(2k-2)$, and so remains inherently random regardless of the fixed number of nearest neighbors used in the estimate. The above discussion shows that even in highly regular cases, the best k-nearest-neighbor density estimate will not converge in probability to the desired limit, and remains random.

Multivariate extremes

Let $\{(Y_{n1}, Y_{n2})\}_{n\geq 1}$ be an i.i.d. sequence of bivariate random vectors. To focus on the bivariate dependence structure rather than the marginal distributions, we assume that (Y_{11}, Y_{12}) has a copula C and standard uniform marginals. We want to characterize

$$P(Y_{11} > 1 - xt_1, Y_{12} > 1 - xt_2)$$

as $x \searrow 0$. If Y_{11} and Y_{12} are dependent but the maxima are asymptotically independent then the extreme value methods fail; see Fougeres (2004) for a general discussion of this problem. For most known families of copulas which have the asymptotic independence property of maxima, we have

$$P(Y_{11} > 1 - xt_1, Y_{12} > 1 - xt_2) \sim cx^2. \tag{4}$$

By the results of this paper the appropriate scaling to obtain a point process limit for the joint extremes is $a_n = n^{-1/2}$, and not the $a_n = n^{-1}$ that would be used to normalize the marginal variables individually. Note, moreover, that the methods of this paper are "dimension free", and so we can address multivariate copulas of any dimension.

Further we can address the joint extreme value behavior of copulas with the asymptotic independence property but where (4) is not satisfied. As an example, consider the case when C is the bivariate normal copula with correlation $\rho \in (0, 1]$ – i.e. $C(x, y)$ is given by a joint normal distribution function at $(\Phi^{-1}(x), \Phi^{-1}(y))$ with standard marginals and correlation ρ. We have

$$P(Y_{11} > 1 - xt_1, Y_{12} > 1 - xt_2) \sim x^{2/(1+\rho)} g(t_1, t_2).$$

Define a_n to be such that

$$P(Y_{11} > 1 - a_n, Y_{12} > 1 - a_n) = n^{-1},$$

so that $a_n = n^{-(1+\rho)/2}\ell(n)$, where $\ell(n)$ is slowly varying at infinity.

Corollary: Assume that $\{\mathbf{Y}_n = (Y_{n1}, Y_{n2})\}_{n\geq 1}$ are independent, have a common normal copula of parameter ρ and uniform marginals. Then for $\mathbf{u} = (1, 1)$ and ,

$$N^{(n)} = \sum_{i=1}^{n} \delta_{a_n^{-1}[\mathbf{Y}_i - \mathbf{u}]}$$

converges to a Poisson process on R_-^2 with a mean measure $W(\cdot, \cdot)$.

Some further remarks

Let Y_i be iid random vectors and define U-statistics

$$U_n = \sum_{i \neq j}^{n} h(Y_i, Y_j).$$

Dabrowski, Dehling, Mikosch and Sharipov studied behaviour of U_n by assuming that h is *regularly varying* with infinite variance. The point process limit is Poisson. Open question: is it possible to generalize it to dependent sequences.