A Kernel Type Nonparametric Density Estimator for Decompounding

Bert van Es
Shota Gugushvili
Peter Spreij

Korteweg-de Vries Instituut voor Wiskunde
Universiteit van Amsterdam

2nd General AMaMef Conference
Bedlewo, 30 April – 5 May 2007

Model

We observe independent copies of the random variable

$$X = \sum_{j=1}^{N} Y_j,$$

where Y's are i.i.d. and N is Poisson(λ) distributed.

Assumption on Y

The Y_j have unknown density f.

Assumption on the observations

We actually observe $X_1, X_2, \ldots, X_{T_n}$, where the stopping time T_n is such that there are exactly n nonzero observations, which will be denoted by $Z_1, \ldots, Z_n$.

Aim

Estimate the unknown density f.

Extra assumption

The intensity λ is known.

Related work

Estimation of the distribution function F of the Y_j has been considered by Buchmann and Grübel, Ann. Statist. (2003).

Motivation

Let N be Poisson process with intensity μ, and the Y_j as before. Let ξ be the *compound Poisson process*

$$\xi_t = \sum_{j=1}^{N_t} Y_j.$$

Let $t_k = kh$ be observation instants and $X_k = \xi_{t_k} - \xi_{t_{k-1}}$. Then the X_k are iid and in distribution equal to

$$X = \sum_{j=1}^{N_h} Y_j,$$

where N_h has a Poisson distribution with parameter $\lambda = h\mu$.

Conclusion

The $X_1, \ldots, X_{T_n}$ can be viewed as observations from a discretely sampled compound Poisson process.

Financial application

In insurance mathematics, the Y_j are the sizes of incoming claims. We estimate the density of the claim sizes.

A Nonparametric Estimation

Tools: Kernel smoothing and Fourier inversion.

The estimator is based on the non zero observations $Z_1, \ldots, Z_n$. Let g denote the density of the Z_j, which coincides with the conditional density of X given $N > 0$.

Its characteristic function equals

$$\phi_g(t) = \frac{1}{e^\lambda - 1}\left(e^{\lambda\phi_f(t)} - 1\right).$$

We solve this equation for $\phi_f(t)$.

Then

$$\phi_f(t) = \frac{1}{\lambda} \mathsf{Log}\left((e^\lambda - 1)\phi_g(t) + 1\right).$$

Here Log denotes the *distinguished logarithm*.

If ϕ_f is integrable, then by Fourier inversion we have

$$\begin{aligned} f(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi_f(t) dt \\ &= \frac{1}{2\pi\lambda} \int_{-\infty}^{\infty} e^{-itx} \mathsf{Log}\left((e^\lambda - 1)\phi_g(t) + 1\right) dt. \end{aligned}$$

If we have an estimator of g (and hence of ϕ_g) we obtain an estimator for f by the *plug-in device*.

Step 1: Kernel estimator of g

Let w denote a *kernel function* with characteristic function ϕ_w and let h denote a positive number, the *bandwidth*.

We estimate the density g by the kernel estimator

$$g_{nh}(x) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{h}w\left(\frac{x - Z_j}{h}\right)$$

The characteristic function of g_{nh}, given by

$$\phi_{g_{nh}}(t) = \phi_{emp}(t)\phi_w(ht),$$

serves as an estimator of ϕ_g.

Here

$$\phi_{emp}(t) = \frac{1}{n}\sum_{j=1}^{n} e^{itZ_j},$$

is the *empirical characteristic function*.

Step 2: Adjusted plug-in estimator

Define the plug-in estimator

$$f_{nh}(x) = \frac{1}{2\pi\lambda} \int_{-1/h}^{1/h} e^{-itx} \, \mathrm{Log} \Big((e^{\lambda} - 1)\phi_{g_{nh}}(t) + 1 \Big) \, dt.$$

For technical reasons we use the modified estimator

$$\widehat{f}_{nh}(x) = (M_n \wedge f_{nh}(x)) \vee (-M_n),$$

where $M = (M_n)_{n \geq 0}$ is a sequence of positive numbers converging to infinity at a suitable rate. If, for certain ω, the argument of Log equals zero for some t, then we define $\widehat{f}_{nh}(x)$ as zero.

Condition W. *The kernel function* w *satisfies*

1. $w(u) = w(-u)$;
2. $\int_{-\infty}^{\infty} w(u)du = 1$;
3. $\sup_u w(u) \leq C < \infty$;
4. $\int_{-\infty}^{\infty} u^2|w(u)|du < \infty$;
5. $\lim_{|u|\to\infty} |uw(u)| = 0$.
6. $\phi_w(t)$ *has support* $[-1, 1]$.

Example of w

A kernel w satisfying these conditions is

$$w(t) = \frac{48t(t^2 - 15)\cos t - 144(2t^2 - 5)\sin t}{\pi t^7}.$$

It has the simple characteristic function

$$\phi_w(t) = (1 - t^2)^3 1_{\{|t|<1\}}.$$

More conditions

Condition H. *The bandwidth* h *depends on* n *and is of the form* $h = Cn^{-\alpha}$ *for* $0 < \alpha < 1$, *where* C *is some constant.*

Condition M. *The truncating sequence* $M = (M_n)_{n \geq 1}$ *is given by* $M_n = n^{\alpha}$ *with* $\alpha > 0$.

Condition F. *The function*

1. f *is a* $\mathcal{C}^2$ *function and*
2. f', f'' *and* $t^2 \phi_f(t)$ *are integrable.*

Mean squared error, Asymptotic Bias and Variance

As a criterion for the quality of the estimator we use the MSE

$$\mathsf{E}[\hat{f}_{nh}(x) - f(x)]^2,$$

which has the decomposition

$$MSE = (\mathsf{E}[\hat{f}_{nh}(x)] - f(x))^2 + \mathsf{Var}[\hat{f}_{nh}(x)].$$

Expansion of the Bias

Proposition 1. *The bias of $\widehat{f}_{nh}(x)$ admits the expansion*

$\mathsf{E}[\widehat{f}_{nh}(x)] - f(x) =$

$$h^2 \frac{\sigma^2(e^\lambda - 1)}{4\pi\lambda} \int_{-\infty}^{\infty} e^{-itx} \frac{t^2 \phi_g(t)}{(e^\lambda - 1)\phi_g(t) + 1} dt + o(h^2) + O\left(\frac{1}{nh}\right).$$

- In ordinary kernel estimation under the same conditions the bias is of order h^2. We have an additional term of order $\frac{1}{nh}$.
- Under the conditions $h \to 0, nh \to \infty$ standard in ordinary kernel estimation the bias vanishes.

Expansion of the Variance

Proposition 2. *Assume additionally* $nh^9 \to 0$. *Then the variance of* $\widehat{f}_{nh}(x)$ *admits the expansion*

$$\mathrm{Var}[\widehat{f}_{nh}(x)] = \frac{1}{nh}\frac{(e^{\lambda}-1)^2}{\lambda^2}\, g(x) \int_{-\infty}^{\infty} (w(u))^2 du + o\left(\frac{1}{nh}\right).$$

- The condition $nh^9 \to 0$ appears to handle the remainder term.
- Under the conditions $h \to 0, nh \to \infty$ the variance vanishes.

Further results

Corollary 3. *The MSE of the estimator $\widehat{f}_{nh}$ satisfies*

$$MSE = C_1 h^4 + \frac{C_2}{nh} + o(h^4) + o(\frac{1}{nh}),$$

for certain constants C_1 and C_2. Hence the estimator is consistent under the previous conditions.
The optimal bandwidth is of order $n^{-1/5}$ and the MSE is then of order $n^{-4/5}$.

Minimax results

Proposition 4. *Let f satisfy Condition F and let $\lambda < \log 2$. Then for an arbitrary $x_0 \in \mathbb{R}$ the inequality*

$$\liminf_{n\to\infty} \inf_{f_{T_n}} \sup_{f\in\mathcal{F}} \mathsf{E}\left[n^{4/5}|f_{T_n}(x_0) - f(x_0)|^2\right] > 0, \qquad (1)$$

is valid. Here the infimum is taken over all estimators based on nonzero discrete observations $Z_1, \ldots, Z_n$.
Hence, the minimax convergence rate for the quadratic loss function for the decompounding problem is lower bounded by $n^{-2/5}$.

We conjecture that the estimator $\widehat{f}_{nh}$ attains it for the bandwidth h satisfying the condition $h = Cn^{-1/5}$, where C denotes some constant.

Asymptotic Normality

Proposition 5. *Assume all previous conditions, and that the bandwidth* h *satisfies an additional condition* $nh^5 \to 0$ *and* $g(x) \neq 0$*. Then*

$$\left(\frac{\widehat{f}_{nh}(x) - f(x)}{\sqrt{\mathsf{Var}[\widehat{f}_{nh}(x)]}} \right) \xrightarrow{\mathfrak{D}} N(0,1).$$

Proposition 6. *Assume all previous conditions, and that the bandwidth* h *satisfies an additional (weaker) condition* $nh^9 \to 0$ *and* $g(x) \neq 0$*. Then we have*

$$\left(\frac{\widehat{f}_{nh}(x) - \mathsf{E}[\widehat{f}_{nh}(x)]}{\sqrt{\mathsf{Var}[\widehat{f}_{nh}(x)]}} \right) \xrightarrow{\mathfrak{D}} N(0,1).$$

Extensions and refinements

It is possible to derive results for the case where

- w is a higher order kernel
- f belongs to the Hölder class $\mathcal{H}(\beta, L_1)$ and to the Nikol'skii class $\mathcal{N}(\beta, L_2)$.

Under these assumptions the expansion of the bias of $\widehat{f}_{nh}(x)$ changes,only an order bound is available for arbitrary β. The order of the variance stays the same.

Two Simulated Examples

The kernel we used is the example that we have given before.

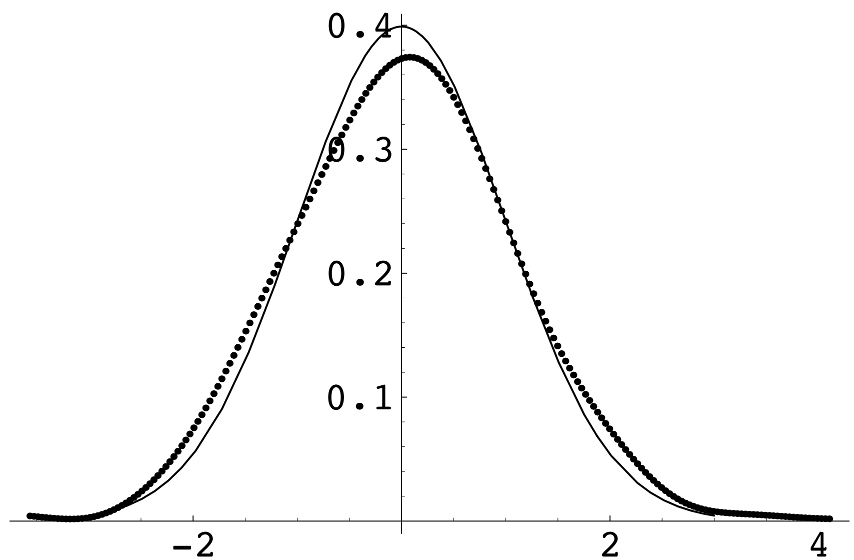
$$w(t) = \frac{48t(t^2 - 15)\cos t - 144(2t^2 - 5)\sin t}{\pi t^7},$$

To compute the estimator we used the fast Fourier transform.

Example 1

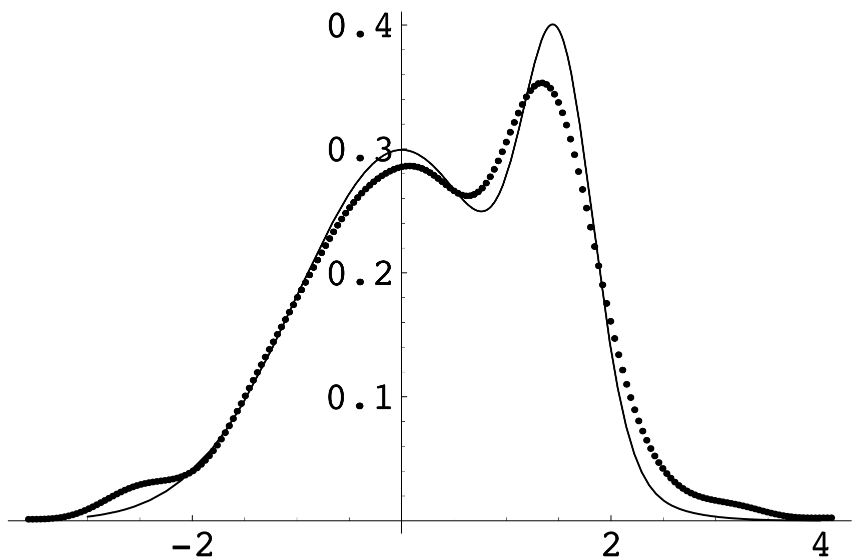
The true density f is the standard normal density and $\lambda = 0.3$.

The estimate is based on 1000 observations and the bandwidth was selected to be 0.14.



Example 2

The density f is the mixture of two normal $N(0,1)$ and $N(\frac{3}{2},\frac{1}{9})$ densities with mixing probabilities $\frac{3}{4}$ and $\frac{1}{4}$ respectively. The intensity $\lambda = 0.3$. The estimate is based on 1000 observations and the bandwidth equals 0.1.



References

This presentation is based on

1. Bert van Es, Shota Gugushvili, Peter Spreij, A kernel type nonparametric density estimator for decompounding, *arXiv:math/0505355* (preprint, full paper).

2. Bert van Es, Shota Gugushvili, Peter Spreij, A kernel type nonparametric density estimator for decompounding, forthcoming in *Bernoulli* (short version).