Statistical Inference of Discretely Observed Compound Poisson Processes

Suraj Shah
Due Date: 02/05/2019

1 Introduction

1.1 Definition & Basic Properties

We begin by introducing the definition of a CPP, illustrating its importance in modelling real world applications and providing simulations to show it's almost sure discontinuities/jumps.

We then formulate the problem and show that it is equivalent to estimating F from N observations of a Poisson random sum. We also make sure we take our observations so that we do not encounter jump sizes of 0 (i.e. condition on the event $\{X_{i\Delta} - X_{(i1)\Delta} \neq 0\}$ and making sure $F(\{0\}) = 0$.

We then describe the approaches to the density estimation problem starting from the seminal work of Buchmann & Grubel through the use of 'decompounding' and then provide more recent advances in this area through the Spectral and Bayesian approach based estimators.

2 Spectral Approach

We begin with the Spectral Approach and formulate the estimator given by Van Es (3) using the Levy-Khintchine formula and explain the care required when taking the logarithm since the exponential function is not injective for complex numbers.

We then prove asymptotic normality results under certain conditions (Theorem 2.1) for the density estimator f_{nh} and provide proofs of the required Lemmas in the Appendix.

3 Bayesian Approach

We develop the framework for applying Bayes' Theorem with prior distribution on density f being a Dirichlet mixture of normal densities and derive a posterior distribution of f as explained in (2). We then introduce the Hellinger distance and Hellinger-type neighbourhoods of f_0. We use these notions to prove that the posterior contraction rate is, up to a logarithmic factor, $(n\Delta)^{-1/2}$.

We formulate the MCMC algorithm to sample from the posterior and describe how to circumvent the intractable form of the likelihood $p(Z|f)$ by constructing a data augmentation scheme. This formulation requires care and will inevitably form a large part of the essay.

4 Numerical Simulations

We provide examples and illustrations of the feasibility of these approaches on a mixture of densities, with absolutely continuous and discrete components to show advantages and limitations of certain estimators. We also vary λ and the number of observations n and separate the scenarios of low frequency data ($\Delta > 0$ is fixed) and high frequency data ($\Delta \to 0$).

We look at the L_2 errors using Monte Carlo as outlined in (1) using a suitable approximation for the expected L_2 loss.

References

[1] C. Duval. Density estimation for compound poisson processes from discrete data. *Stochastic Processes and their Applications*, 123(11):3963–3986, 2013. URL `https://doi.org/10.1016/j.spa.2013.06.006`.

[2] S. Gugushvili, F. van der Meulen, and P. Spreij. A non-parametric bayesian approach to decompounding from high frequency data. *Statistical Inference for Stochastic Processes*, 21(1):53–79, Apr 2018. URL `https://doi.org/10.1007/s11203-016-9153-1`.

[3] B. van Es, S. Gugushvili, and P. Spreij. A kernel type nonparametric density estimator for decompounding. *Bernoulli*, 13(3):672–694, 08 2007. doi: 10.3150/07-BEJ6091. URL `https://doi.org/10.3150/07-BEJ6091`.