My research

Martin Bladt

January 6, 2024

I am broadly interested in statistical methodology and applied probability, with motivation from insurance problems. I have worked in various areas, and recently have focused on extreme value theory and survival analysis. Two lines of research that I find interesting within these areas are the following.

Extremes and incomplete data

Extreme value theory is a branch of statistics which models catastrophic events. The literature on extreme events is vast, and while many of these techniques apply for fully observed data, there is a large knowledge gap when it comes to data where some observations are missing or partially observed, also known as censored. Despite the well-known fact that ignoring censoring effects leads to serious underestimation of the risks associated with calamitous phenomena, the mathematical and computational aspects of censored extreme events are still in their infancy.

Under some definitions, a random variable X is called heavy-tailed if its tail is regularly varying, that is if

$$\mathbb{P}(X > x) = 1 - F(x) = x^{-1/\gamma_F}\ell_F(x),$$

where ℓ_F is a slowly-varying function at infinity, and $\gamma_F > 0$ is called the tail index. When X is right-censored by another heavy-tailed variable C, then the observable variable $Z = \min\{X, C\}$ has a heavy-tailed behaviour which incorrectly estimates the heaviness of X. Thus, a main problem of interest is recovering the tail behaviour of X from that of Z and $\delta = I(X = Z)$.

There has been interest in tail index estimation under random right-censorship which further generated important research in this direction, e.g. using covariates or truncation. I have focused on approaches to capture the full tail behaviour, including but also beyond the estimation of tail indices. For instance, most recently by studying *extreme Kaplan-Meier integrals:*

$$S_{k,n}(\varphi) = \int \varphi \; \mathrm{d}\mathbb{F}_{k,n},$$

where $\mathbb{F}_{k,n}(x) = 1 - \prod_{i=1}^{k} \left[1 - \frac{\delta_{[n-i+1:n]}}{i}\right]^{I(Z_{n-i+1,n}/Z_{n-k,n} \leq x)}$ is the product-limit estimator – a consistent estimator of censored samples – evaluated at normalized upper order statistics. Such integrals allow to capture a variety of functionals of the tail distribution, and also generalize to more complex data structures that arise naturally in insurance data.

Survival analysis and contamination

Survival analysis is a branch of statistics which deals with time-to-event data, and has classically found applications in biostatistics and medicine. However, very often insurance data also requires such techniques, though not without requiring adaptation. For instance, a key problem arising in insurance beyond censoring is the re-opening of closed claims, which from a survival analysis perspective can be understood as the contamination of data. Furthermore, expert information about the severity of the contamination may be available.

In a general setting, let Z be a Markov jump process on a finite state space. Let N be a multivariate counting process with components given by

$$N_{jk}(t) = \#\{s \in (0,t] : Z_{s-} = j, Z_s = k\},$$

which can be used to effectively estimate the transition probabilities $p_{jk}(t,s) = \mathbb{P}(Z_s = k \mid Z_t = j)$. If N is subject to censoring and perturbations and we instead observe $\widetilde{N}$, together with an expert guess $\widehat{N}$, then learning the distribution of Z from the latter two quantities can greatly decrease estimation bias and variance. I am currently working on this kind of problem, in connection with disability or liability insurance data. An additional line of my research is incorporating covariate information with modern machine-learning tools.

The Markovianity of the process Z is an assumption that can be relaxed, and the formal detection of when this relaxation is possible is an exciting and relevant problem that has recently caught my attention. In this connection, empirical process methods for proving asymptotic convergence of estimators of $p_{jk}(t,s)$ and related quantities seem to pave the way.