Compound Poisson approximation for Markov chains

Torkel Erhardsson

Ph. D. Thesis
Department of Mathematics
Royal Institute of Technology
S–100 44 STOCKHOLM
Sweden
ter@math.kth.se

Akademisk avhandling som med tillstånd av Kungliga Tekniska Högskolan i Stockholm framlägges till offentlig granskning för avläggande av teknisk doktorsexamen fredagen den 5 december 1997 kl. 10.00 i Kollegiesalen, Administrationsbyggnaden, Kungliga Tekniska Högskolan, Valhallavägen 79, Stockholm.

Abstract

Let η be a stationary Harris recurrent Markov chain on a Polish state space $(S, \mathcal{F})$, with stationary distribution μ. Let $\Psi_n := \sum_{i=1}^{n} I\{\eta_i \in S_1\}$ be the number of visits to $S_1 \in \mathcal{F}$ by η, where S_1 is "rare" in the sense that $\mu(S_1)$ is "small". We want to find an approximating *compound Poisson* distribution for $\mathcal{L}(\Psi_n)$, such that the approximation error, measured using the *total variation* distance, can be *explicitly* bounded with a bound of order not much larger than $\mu(S_1)$. This is motivated by the observation that approximating *Poisson* distributions often give larger approximation errors when the visits to S_1 by η tend to occur in *clumps*, and also by the compound Poisson *limit theorems* of classical *extreme value theory.*

We here propose an approximating compound Poisson distribution which in a natural way takes into account the *regenerative* properties of Harris recurrent Markov chains. A total variation distance error bound for this approximation is derived, using the compound Poisson *Stein equation* of Barbour, Chen and Loh (1992), and certain *couplings.* When the chain has an *atom* (e.g., a singleton), the bound is particularly simple, and depends only on much studied quantities like *hitting probabilities* and *expected hitting times*, which satisfy *Poisson's equation* and can in some cases be bounded using *Lyapunov* functions. Adding a few terms to the bound gives *upper and lower* bounds for the error in the approximation with *arbitrary* compound Poisson, or *normal*, distributions. Applications of the above results are given, to the number of *overlapping occurrences* of fixed sequences in finite-state Markov chains (in particular, the sequence 111...111 in a Markov chain on $\{0, 1\}$), and to the visits by *birth-death* chains and arbitrary *finite-state* Markov chains to "rare" sets.

The most important application concerns the *Johnson-Mehl* crystallization model of stochastic geometry. In this model, for each point (x_i, y_i) of a Poisson point process on $R \times [0, \infty)$, an interval starts to grow in R from x_i with constant speed in both directions at time y_i. Let $S(z, L)$ be the number of components of the *uncovered set* (= the complement of the union of growing random intervals) which intersect $(0, L]$, at time $z > 0$. $S(z, L)$ can be interpreted as the number of visits to a "rare" set by a Markov chain. We first give an approximating Poisson distribution for $\mathcal{L}(S(z, L))$, and a total variation distance error bound, using the *Stein-Chen method* and the existence of *monotone increasing* couplings. Then, using the general results described above and suitable couplings, we give an approximating compound Poisson distribution, and an error bound for this approximation. Under a mild condition, the latter bound is of considerably smaller order than the former. These results are illustrated numerically.

Keywords: Compound Poisson approximation, error bound, stationary Harris chain, "rare" set, Stein equation, coupling, expected hitting times, overlapping occurrences of patterns, Johnson-Mehl model, uncovered set.

AMS 1991 subject classification: Primary 60E15, 60J05. Secondary 60D05, 60G70.