Approximation by Geometric Sums: Markov chain passage times and queueing models

Fraser Daly

Heriot–Watt University

LMS–EPSRC Durham Symposium
Markov Processes, Mixing Times and Cutoff
July 2017

Outline

- Geometric sums
- Approximation results
  - Application to Markov chain passage times
- Approximations using bounded failure rate
  - Applications to the M/G/1 queue

Geometric sums

Let $Y = X_1 + X_2 + \cdots + X_N$, where

- $X, X_1, X_2, \ldots$ are i.i.d. positive integer-valued random variables, and
- N has a geometric distribution (independent of the X_i), supported on $\{0, 1, 2, \ldots\}$, with $\mathbb{P}(N = 0) = p$.

Applications include

- Records processes (time until a new record is set)
- Random walks, including the ruin problem (maximum of the random walk)
- Reliability systems (time to failure)

Geometric sums

We will consider situations where a (non-negative, integer-valued) random variable of interest W has a distribution *close* to that of a geometric sum, measured by total variation distance:

$$\begin{aligned} d_{TV}(\mathcal{L}(W), \mathcal{L}(Y)) &= \sup_{A \subseteq \mathbb{Z}^+} |\mathbb{P}(W \in A) - \mathbb{P}(Y \in A)| \\ &= \inf_{(W,Y)} \mathbb{P}(W \neq Y), \end{aligned}$$

where the infemum is taken over all couplings of W and Y.

Geometric sums

Rényi's theorem:

$$\lim_{p \to 0} \mathbb{P}\left(\frac{1}{\mathbb{E}N} Y \leq x\right) = 1 - \exp\left\{-\frac{x}{\mathbb{E}X}\right\}$$

Explicit bounds in exponential approximation are available, as are explicit bounds for geometric approximation of the geometric sum Y.

A characterisation

Define the random variable V by

$$V + X \stackrel{d}{=} (W | W > 0) \, .$$

Note that if W is a geometric sum (with summand X), then $V \stackrel{d}{=} W$. The converse is also true.

We may expect to be able to quantify the distance of W from our geometric sum Y by assessing the distance of W from V.

Approximation by geometric sums

Theorem

If $p = \mathbb{P}(W = 0)$,

$$d_{TV}(\mathcal{L}(W), \mathcal{L}(Y)) \leq \frac{1-p}{p} d_{TV}(\mathcal{L}(W), \mathcal{L}(V)) .$$

The proof, using Stein's method, is based on the above characterisation for geometric sums.

Proof

Write $q = 1 - p$. Given a function $h : \mathbb{Z}^+ \mapsto \mathbb{R}$, define the function $f = f_h$ as the solution to the following 'Stein equation':

$$h(k) - \mathbb{E}h(Y) = q\mathbb{E}[f(k + X)] - f(k)$$

(with $f(0) = 0$), so that

$$\begin{aligned} \mathbb{E}h(W) - \mathbb{E}h(Y) &= q\mathbb{E}[f(W + X)] - q\mathbb{E}[f(W)|W > 0] \\ &= q\mathbb{E}[f(W + X) - f(V + X)]\,. \end{aligned}$$

Hence

$$|\mathbb{E}h(W) - \mathbb{E}h(Y)| \leq q\mathbb{P}(W \neq V) \sup_{j,k} |f(j) - f(k)|\,.$$

With $h(k) = I(k \in A)$, taking the supremum over $A \subseteq \mathbb{Z}^+$ gives us d_{TV} on the LHS. The final term on the RHS may be shown to be bounded by $1/p$.

Markov chain passage times

Let $\{\xi_i : i \geq 0\}$ be an ergodic discrete-time Markov chain with transition matrix P, started according to its stationary distribution π, and let $W = \min\{i : \xi_i \in B\}$ for some set of states B.

Let B_i be the set of states j from which a move to B requires a minimum of i steps, i.e., for which $\mathbb{P}(\xi_i \in B | \xi_0 = j) > 0$ but $\mathbb{P}(\xi_k \in B | \xi_0 = j) = 0$ for $k < i$.

Let

$$\begin{aligned} p &= \mathbb{P}(W = 0) = \pi(B)\,, \\ \mu_i &= \mathbb{P}(\xi_0 \in B_i | \xi_0 \notin B) = \frac{\sum_{j \in B_i} \pi(j)}{1-p}\,. \end{aligned}$$

Define the random variable X by $\mathbb{P}(X = i) = \mu_i$.

Markov chain passage times

Theorem

$$d_{TV}(\mathcal{L}(W), \mathcal{L}(Y)) \leq \frac{1}{p}\mathbb{E} \sum_{i,j \in B} \pi_i \sum_{n \geq 0} \left| P_{ij}^{(n+X)} - \pi_j \right| .$$

This is proved using our earlier bound: construct copies of the original chain, but started according to π restricted to B_j for each j. Randomising j according to X gives us the random variable V by considering the first passage time to the set of states B.

We now use the maximal coupling of these processes with copies of our original Markov chain to estimate $\mathbb{P}(W \neq V)$. Some analysis yields the stated bound.

Application: Sequence patterns

Consider a sequence of IID Bernoulli trials (of 0s and 1s), and let I_i be the indicator that a given k-digit pattern (B, say) appears, starting at position i in our sequence. We are interested in $W = \min\{i : I_i = 1\}$, the time we have to wait to observe our pattern.

Define a 2^k state Markov chain such that at time n, the state of the process is the outcome of the k Bernoulli trials from time n to time $n + k - 1$.

Our theorem gives the bound

$$d_{TV}(\mathcal{L}(W), \mathcal{L}(Y)) \leq \sum_{i=1}^{k-1} \sum_{n=1}^{i} \mu_n |c_i - p| ,$$

where $c_i = \mathbb{P}(I_i = 1 | I_0 = 1)$. This represents the i-step transition probability from B to B, for $i \leq k - 1$.

This is sharper than available geometric approximation bounds.

Application: Sequence patterns

For example, let $k = 3$ and consider the pattern 010.

We have that $p = r(1-r)^2$, where r is the expected value of each of the original Bernoulli variables.

We have the following partition of the state space of our Markov chain:

$$B = \{010\}\,, \;\; B_1 = \{001, 101\}\,, \;\; B_2 = \{000, 100, 110\}\,, \;\; B_3 = \{011, 111\}\,.$$

We can easily calculate $c_1 = 0$, $c_2 = r(1-r)$,

$$\mu_1 = \frac{r(1-r)}{1-p}\,, \;\; \mu_2 = \frac{(1-r)(1-r+r^2)}{1-p}\,, \;\; \mu_3 = \frac{r^2}{1-p}\,.$$

Approximations using bounded failure rate

Our general bound above can be tricky to evaluate: estimating $d_{TV}(\mathcal{L}(W), \mathcal{L}(V))$ is not always easy.

However, simpler upper bounds for the approximation of W by a geometric sum are available, under assumptions on the structure of the underlying random variables.

Notation

Let W be a non-negative, integer-valued random variable. Define the failure (or hazard) rate of W by

$$r_W(j) = \frac{\mathbb{P}(W = j)}{\mathbb{P}(W > j)} .$$

For future use, we also define (for $0 \leq p < 1$ and X a positive, integer-valued random variable)

$$H_p(X) = \min\left\{ p + (1-p)\mathbb{P}(X > 1), p\left(1 + \sqrt{\frac{-2}{u\log(1-p)}}\right)\right\} ,$$

where $u = 1 - d_{TV}(\mathcal{L}(X), \mathcal{L}(X+1))$, a measure of the 'smoothness' of X. This is small (close to p) when X is 'smooth' or $\mathbb{P}(X = 1)$ is large.

Approximations using bounded failure rate

Theorem

Let $\mathbb{P}(W = 0) = p$ and $r_W(j) > \delta > 0$ for all j. If

$$\mathbb{E}X \geq \frac{p}{(1-p)\delta},$$

then

$$d_{TV}(\mathcal{L}(W), \mathcal{L}(Y)) \leq H_p(X)\,(\mathbb{E}Y - \mathbb{E}W),$$

where $Y = X_1 + \cdots + X_N$ is our geometric sum.

Natural candidates for applications are when W has increasing failure rate (IFR) or decreasing failure rate (DFR). The first setting leads to simple geometric approximation bounds ($X = 1$ almost surely).

Proof

- Let X be as in the theorem, and V be as before. We can show that $\mathbb{P}(V + X > j) \leq \mathbb{P}(W + X > j)$.
- Let f be the solution to the same 'Stein equation' as before:

$$h(k) - \mathbb{E}h(Y) = (1-p)\mathbb{E}[f(k+X)] - f(k)\,.$$

When $h(k) = I(k \in A)$ for $A \subseteq \mathbb{Z}^+$, we can show that

$$\sup_j |\Delta f(j)| \leq \frac{1}{p} H_p(X)\,,$$

where $\Delta f(j) = f(j+1) - f(j)$.

Proof

Rearrange our Stein equation to get

$$\mathbb{E}h(W) - \mathbb{E}h(Y) = (1-p)\sum_{j=0}^{\infty} \Delta f(j) \left[\mathbb{P}(W+X>j) - \mathbb{P}(V+X>j)\right].$$

With $h(k) = I(k \in A)$, we then obtain

$$|\mathbb{P}(W \in A) - \mathbb{P}(Y \in A)| \leq \frac{1-p}{p} H_p(X)\mathbb{E}[W-V] = H_p(X)\mathbb{E}[Y-W].$$

Taking the supremum over $A \subseteq \mathbb{Z}^+$ gives us $d_{TV}(\mathcal{L}(W), \mathcal{L}(Y))$.

Applications to the M/G/1 queue

Consider a single server queue with

- customers arriving at rate λ, and
- i.i.d. customer service times with the same distribution as S.

Let $\rho = \lambda \mathbb{E} S$ and assume $\rho < 1$.

We will approximate

1. the number of customers in the system in equilibrium by a geometric distribution (so $X = 1$ almost surely), and
2. the number of customers served during a busy period by a geometric sum.

Number of customers in the system

Let W be the equilibrium number of customers in the system. It is well-known that

$$\mathbb{P}(W=0)=1-\rho\,, \qquad \mathbb{E}W=\rho+\frac{\rho^2\mathbb{E}[S^2]}{2(1-\rho)(\mathbb{E}S)^2}\,,$$

$$r_W(j)\geq\frac{1-\rho}{\lambda\sup_{t\geq0}\mathbb{E}[S-t|S\geq t]}\,, \qquad \forall j\in\mathbb{Z}^+\,.$$

S is said to be New Better than Used in Expectation (NBUE) if $\mathbb{E}[S-t|S\geq t]\leq\mathbb{E}S$ for all $t\geq0$. In that case our theorem gives

$$d_{TV}(\mathcal{L}(W),\mathrm{Ge}(1-\rho))\leq\rho^2\left(1-\frac{\mathbb{E}[S^2]}{2(\mathbb{E}S)^2}\right)\,.$$

As expected, this upper bound is zero if S has an exponential distribution.

Customers served during a busy period

Let $W + 1$ be the number of customers served during a busy period of our system. Then $\mathbb{P}(W = 0) = \mathbb{E}e^{-\lambda S} = p$ and $\mathbb{E}W = \rho(1-\rho)^{-1}$.

If S is IFR, it is known that W is DFR. We may then find a lower bound on the failure rate as follows.

Let ψ be the Laplace transform of the density of S and let ξ be the real root of $1 + \lambda\psi'(s)$ nearest the origin. Let

$$\theta = \frac{\xi - \lambda + \lambda\psi(\xi)}{(\xi - \lambda)\psi(\lambda)} \,.$$

Then if $Y = \sum_{i=1}^{N} X_i$, where $N \sim \text{Ge}(p)$ and $(1-p)\theta\mathbb{E}X \geq 1 - p\theta$,

$$d_{TV}(\mathcal{L}(W), \mathcal{L}(Y)) \leq H_p(X) \left(\frac{(1-p)\mathbb{E}X}{p} - \frac{\rho}{1-\rho} \right) .$$

References

- S. Asmussen (2003). *Applied Probability and Queues* 2nd ed. Springer, New York.
- F. Daly (2010). Stein's method for compound geometric approximation. *J. Appl. Prob.* **47**: 146–156.
- F. Daly (2016). Compound geometric approximation under a failure rate constraint. *J. Appl. Prob.* **53**: 700–714.
- V. Kalashnikov (1997). *Geometric Sums: Bounds for Rare Events with Applications*. Kluwer, Dordrecht.
- E. K. Kyprianou (1972). The quasi–stationary distributions of queues in heavy traffic. *J. Appl. Prob.* **35**: 821–831.
- E. Peköz (1996). Stein's method for geometric approximation. *J. Appl. Prob.* **33**: 707–713.
- E. Peköz and A. Röllin (2011). New rates for exponential approximation and the theorems of Rényi and Yaglom. *Ann. Prob.* **39**: 587–608.
- E. Peköz, A. Röllin and N. Ross (2013). Total variation error bounds for geometric approximation. *Bernoulli* **19**: 610–632.
- A. Rényi (1957). A characterization of Poisson processes. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **1**: 519-527.
- S. Ross (2006). Bounding the stationary distribution of the M/G/1 queue size. *Probab. Engrg. Inform. Sci.* **20**: 571–574.
- J. G. Shanthikumar (1988). DFR property of first-passage times and its preservation under geometric compounding. *Ann. Prob.* **16**: 397–406.