Bernoulli **21**(2), 2015, 1002–1013
DOI: 10.3150/14-BEJ595

Stein factors for negative binomial approximation in Wasserstein distance

A.D. BARBOUR[1], H.L. GAN[2,*] and A. XIA[2,**]

[1]*Angewandte Mathematik, Universität Zürich, Winterthurerstr. 190, 8057 Zürich, Switzerland.*
E-mail: a.d.barbour@math.uzh.ch
[2]*Department of Mathematics and Statistics, the University of Melbourne, Parkville, VIC 3010, Australia.*
*E-mail: *ganhl@ms.unimelb.edu.au; **aihuaxia@unimelb.edu.au*

The paper gives the bounds on the solutions to a Stein equation for the negative binomial distribution that are needed for approximation in terms of the Wasserstein metric. The proofs are probabilistic, and follow the approach introduced in Barbour and Xia (*Bernoulli* **12** (2006) 943–954). The bounds are used to quantify the accuracy of negative binomial approximation to parasite counts in hosts. Since the infectivity of a population can be expected to be proportional to its total parasite burden, the Wasserstein metric is the appropriate choice.

Keywords: negative binomial approximation; Stein factors; Stein's method; Wasserstein distance

1. Introduction

The negative binomial distribution is widely used in biology to model the counts of individuals in populations, since such counts are frequently overdispersed, making the Poisson distribution an unsuitable choice. Indeed, the main advantage of the negative binomial family over the Poisson family is the extra flexibility in fitting that results because the negative binomial family has a second parameter. However, for the distribution of parasites among hosts, there are plausible mechanistic models [4,5] that predict a negative binomial distribution, and it is of interest to know whether a member of the negative binomial family would still give a reasonable approximation, if the detailed assumptions of such a model were relaxed. One of the quantities of primary interest is then the total rate of output of infective stages, which can be expected to be closely related to the total number of parasites in the population [6]. Thus the approximation needs to be good when measured by a distance that limits the differences in expectation of (not necessarily bounded) Lipschitz functionals, which makes the Wasserstein metric a natural choice. In this paper, we make negative binomial approximation using Stein's method a practical proposition, by giving bounds on the solutions of an appropriate Stein equation that correspond to Lipschitz test functions.

The negative binomial distribution $\mathrm{NB}(r, p)$ has probabilities given by

$$\mathrm{NB}(r, p)\{k\} = \frac{\Gamma(r+k)}{\Gamma(r)k!}(1-p)^r p^k, \qquad k \in \mathbb{Z}_+ := \{0, 1, \ldots\}; r > 0, 0 < p < 1.$$

One can check directly that $W \sim \mathrm{NB}(r, p)$ if and only if

$$\mathbb{E}\bigl[p(r+W)g(W+1) - Wg(W)\bigr] = 0$$

1350-7265

for a sufficiently rich class of functions $g:\mathbb{N}\to\mathbb{R}$. One such class consists of the solutions $g_f:\mathbb{N}\to\mathbb{R}$ to the equations

$$p(r+i)g_f(i+1)-ig_f(i)=f(i)-\mathrm{NB}(r,p)\{f\}, \qquad f\in\mathcal{F}_W, \tag{1.1}$$

where $\mathcal{F}_W:=\{f:|f(x)-f(y)|\le|x-y|,\forall x,y\in\mathbb{Z}_+\}$ denotes the class of Lipschitz functions on $\mathbb{Z}_+$, and $\mathrm{NB}(r,p)\{f\}:=\mathbb{E}f(Z)$ for $Z\sim\mathrm{NB}(r,p)$. Then, for any random variable W on $\mathbb{Z}_+$,

$$\mathbb{E}f(W)-\mathrm{NB}(r,p)\{f\}=\mathbb{E}\bigl[p(r+W)g_f(W+1)-Wg_f(W)\bigr], \tag{1.2}$$

and, if we can bound the right-hand side of the above equation uniformly for $f\in\mathcal{F}_W$, then we have a uniform bound for the left hand side as well; but this corresponds precisely to a bound on the Wasserstein distance between $\mathcal{L}(W)$ and $\mathrm{NB}(r,p)$.

In order to control the right-hand side of (1.2), it is typically necessary to have bounds on the quantities

$$G_1=\sup_{f\in\mathcal{F}_W}\sup_{w\in\mathbb{N}}g_f(w); \qquad G_2=\sup_{f\in\mathcal{F}_W}\sup_{w\in\mathbb{N}}\bigl|g_f(w+1)-g_f(w)\bigr|.$$

This note establishes the following result:

Theorem 1.1. *For any $r>0$ and $0<p<1$,*

$$G_1=\frac{1}{1-p}, \tag{1.3}$$

$$G_2\le\min\left\{\frac{2}{1-p},\frac{1+p}{(1-p)^2},\sqrt{\frac{r_0}{rp(1-p)^3}}\right\}, \tag{1.4}$$

where r_0 is the solution in $r>1/2$ of the equation $\Gamma(r-\frac{1}{2})/\Gamma(r)=3\sqrt{2\mathrm{e}}/8$, and satisfies $\sqrt{r_0}\le 3/2$.

The proof is given in Section 2. In Section 3, we apply Theorem 1.1 to approximating the distribution of parasites in hosts.

2. The proof of Theorem 1.1

Setting $g_f(i)=h_f(i)-h_f(i-1)$, equation (1.1) becomes:

$$f(i)-\mathrm{NB}(r,p)\{f\}=p(r+i)\bigl(h_f(i+1)-h_f(i)\bigr)-i\bigl(h_f(i)-h_f(i-1)\bigr),$$

where the right-hand side is the generator of an immigration–birth–death process with constant immigration rate rp, and *per capita* birth and death rates p and 1, respectively. More generally, we let $Z_i:=Z_i^{[a,b]}$ denote an immigration–birth–death process with immigration rate a and with *per capita* birth and death rates b and 1, respectively, having $Z_i(0)=i$. We write $Y_i^{[b]}$ for $Z^{[0,b]}$. From [3],

$$h_f(i)=-\int_0^\infty \bigl[\mathbb{E}f\bigl(Z_i^{[rp,p]}(t)\bigr)-\mathrm{NB}(r,p)\{f\}\bigr]\,\mathrm{d}t. \tag{2.1}$$

We make use of the following two lemmas, proved in [4], who attributes the first to Palm. We write

$$\Lambda_t(b):=\mathrm{e}^{-(1-b)t} \quad \text{and} \quad \theta_t(b):=1-(1-b)/\bigl(1-b\Lambda_t(b)\bigr). \tag{2.2}$$

Lemma 2.1. *$Y_1^{[b]}(t)$ has a modified geometric distribution: for $0<b\neq 1$,*

$$\mathbb{P}\bigl[Y_1^{[b]}(t)=0\bigr]=b^{-1}\theta_t; \qquad \mathbb{P}\bigl[Y_1^{[b]}(t)=k\bigr]=\Lambda_t(1-\theta_t)^2\theta_t^{k-1}, \qquad k\geq 1,$$

where $\Lambda_t=\Lambda_t(b)$ and $\theta_t=\theta_t(b)$. In particular, the first two moments are given by

$$\mathbb{E}Y_1^{[b]}(t)=\Lambda_t; \qquad \mathbb{E}\bigl\{Y_1^{[b]}(t)\bigr\}^2=\frac{\Lambda_t(1+b-2b\Lambda_t)}{1-b}.$$

If $b=1$, the limiting formulae as $b\to 1$ hold true; for instance, $\theta_t(1)=t/(1+t)$ and $\mathbb{E}\{Y_1^{[1]}(t)\}^2=1+2t$.

Lemma 2.2. *$Z_0^{[a,b]}(t)$ has the negative binomial distribution $\mathrm{NB}(a/b,\theta_t)$.*

Proof of Theorem 1.1. As $g_f(i)=h_f(i)-h_f(i-1)$, it follows from (2.1) that

$$g_f(i)=-\int_0^\infty \mathbb{E}\bigl[f\bigl(Z_i(t)\bigr)-f\bigl(Z_{i-1}(t)\bigr)\bigr]\,\mathrm{d}t,$$

where, throughout the proof, we write Z_j for $Z_j^{[rp,p]}$. We now couple Z_{i-1} and Z_i by setting

$$Z_i(t)=Z_{i-1}(t)+Y_1(t),$$

where $Y_1\overset{d}{=}Y_1^{[p]}$, and $Z_{i-1}(t)$ and $Y_1(t)$ are independent. Then $g_f(i)$ can be expressed as

$$g_f(i)=-\int_0^\infty \mathbb{E}\bigl[f\bigl(Z_{i-1}(t)+Y_1(t)\bigr)-f\bigl(Z_{i-1}(t)\bigr)\bigr]\,\mathrm{d}t.$$

Now, because $f\in\mathcal{F}_W$, it follows that

$$\bigl|g_f(i)\bigr|\leq\int_0^\infty \mathbb{E}Y_1(t)\,\mathrm{d}t=\int_0^\infty \Lambda_t(p)\,\mathrm{d}t=\frac{1}{1-p},$$

using Lemma 2.1 for the first equality, and this maximal value for $|g_f|$ is attained by taking $f(x)=-x$. This completes the proof of (1.3), and also yields the bound $2/(1-p)$ in G_2.

To prove the remainder of (1.4), we first observe that the function that maximizes $\Delta g_f(i)$ is $f_i(j) = -|j-i|$. This follows by using the same argument as in [2], proof of (1.4). In the rest of the proof, we write $f = f_i$. Using the couplings

$$Z_{i+1}(t) = Z_i(t) + Y_1(t); \qquad Z_i(t) = Z_{i-1}(t) + Y_1'(t),$$

where $Y_1, Y_1' \overset{d}{=} Y_1^{[p]}$ and the processes Z_{i-1}, Y_1 and Y_1' are independent, we obtain

$$\begin{aligned}
\Delta g_f(i) &= \int_0^\infty \mathbb{E}\bigl[f\bigl(Z_{i+1}(t)\bigr) - f\bigl(Z_i(t)\bigr) + f\bigl(Z_{i-1}(t)\bigr) - f\bigl(Z_i(t)\bigr)\bigr]\,\mathrm{d}t \\
&\le \int_0^\infty \mathbb{E}\bigl\{\bigl|Z_{i-1}(t) + Y_1(t) + Y_1'(t) - i\bigr| - \bigl|Z_{i-1}(t) + Y_1(t) - i\bigr| \\
&\qquad - \bigl|Z_{i-1}(t) + Y_1'(t) - i\bigr| + \bigl|Z_{i-1}(t) - i\bigr|\bigr\}\,\mathrm{d}t \\
&\le \int_0^\infty \mathbb{E}\bigl[2\min\bigl(Y_1(t), Y_1'(t)\bigr)\mathbf{1}_{\{Z_{i-1}(t) < i < Z_{i-1}(t) + Y_1(t) + Y_1'(t)\}}\bigr]\,\mathrm{d}t,
\end{aligned}$$

where the last inequality is because the quantity in the braces equals 0 if $Z_{i-1}(t) \ge i$ or $Z_{i-1}(t) + Y_1(t) + Y_1'(t) \le i$; it is bounded by $2Y_1'(t)$ if one applies the triangle inequality to $|Z_{i-1}(t) + Y_1(t) + Y_1'(t) - i| - |Z_{i-1}(t) + Y_1(t) - i|$ and $-|Z_{i-1}(t) + Y_1'(t) - i| + |Z_{i-1}(t) - i|$, and hence it is also bounded by $2Y_1(t)$ if one swaps $Y_1(t)$ and $Y_1'(t)$. This implies that

$$\begin{aligned}
\Delta g_f(i) &\le \int_0^\infty \sum_{i_1, i_2, j} 2\min(i_1, i_2)\mathbf{1}_{\{i+1-i_1-i_2 \le j \le i-1\}} \\
&\qquad \times \mathbb{P}\bigl(Y_1(t) = i_1\bigr)\mathbb{P}\bigl(Y_1'(t) = i_2\bigr)\mathbb{P}\bigl(Z_{i-1}(t) = j\bigr)\,\mathrm{d}t \\
&\le \int_0^\infty \max_j \mathbb{P}\bigl(Z_{i-1}(t) = j\bigr) \qquad (2.3) \\
&\qquad \times \sum_{i_1, i_2} 2\min(i_1, i_2)(i_1 + i_2 - 1)\mathbb{P}\bigl(Y_1(t) = i_1\bigr)\mathbb{P}\bigl(Y_1'(t) = i_2\bigr)\,\mathrm{d}t \\
&\le \int_0^\infty \max_j \mathbb{P}\bigl(Z_{i-1}(t) = j\bigr)\mathbb{E}\bigl[\bigl(Y_1(t) + Y_1'(t)\bigr)\bigl(Y_1(t) + Y_1'(t) - 1\bigr)\bigr]\,\mathrm{d}t.
\end{aligned}$$

To bound $\mathbb{P}[Z_{i-1}(t) = j]$, we decompose $Z_{i-1}(t)$ into a sum of two independent components

$$Z_{i-1}(t) \overset{d}{=} Z_0(t) + Y_{i-1}(t),$$

where $Y_{i-1} \overset{d}{=} Y_{i-1}^{[p]}$, as defined earlier. From this it follows, using Lemma 2.2, that

$$\max_j \mathbb{P}\bigl[Z_{i-1}(t) = j\bigr] \le \max_k \mathbb{P}\bigl[Z_0(t) = k\bigr] = P(r, \theta_t), \tag{2.4}$$

where $\theta_t = \theta_t(p)$ and $P(r,q) := \max_k \mathrm{NB}(r,q)\{k\}$. In [7], the representation of $\mathrm{NB}(r,q)$ as a $\Gamma(r,(1-q)/q)$ mixed Poisson distribution, where $\Gamma(r,\lambda)$ denotes the Gamma distribution with shape parameter r and scale parameter $1/\lambda$, is exploited to bound $P(r,q)$. Using the bound $\max_k \mathrm{Po}(\lambda)\{k\} \le 1/\sqrt{2\mathrm{e}\lambda}$ from [1], he shows that, if $r > 1/2$, then

$$P(r,q) \le \sqrt{\frac{1-q}{2\mathrm{e}rq}} K_r,$$

where $K_r := \sqrt{r}\,\Gamma(r-\frac{1}{2})/\Gamma(r)$ is decreasing in $r > 1/2$. Hence, since

$$\frac{1-\theta_t}{\theta_t} = \frac{1-p}{p(1-\Lambda_t)},$$

we have, for $\theta_t = \theta_t(p)$ and $\Lambda_t = \Lambda_t(p)$,

$$P(r,\theta_t) \le \begin{cases} \sqrt{\dfrac{1-p}{2\mathrm{e}rp(1-\Lambda_t)}} K_r, & \text{if } r > 1/2; \\ 1, & \text{if } r \le 1/2. \end{cases}$$

For the third element in the bound (1.4), we assume that $r > 1/2$, and use (2.3) to give

$$\begin{aligned} \Delta g_f(i) &\le \int_0^\infty \mathbb{E}\bigl[\bigl(Y_1(t)+Y_1'(t)\bigr)\bigl(Y_1(t)+Y_1'(t)-1\bigr)\bigr]\sqrt{\frac{1-p}{2r\mathrm{e}p}} K_r \frac{1}{\sqrt{1-\Lambda_t}}\,\mathrm{d}t \\ &= \sqrt{\frac{1-p}{2r\mathrm{e}p}} K_r \cdot 2\int_0^\infty \frac{\Lambda_t((1-3p)\Lambda_t+2p)}{(1-p)\sqrt{1-\Lambda_t}}\,\mathrm{d}t, \end{aligned}$$

using the moments given in Lemma 2.1. Direct computations now give

$$\int_0^\infty \frac{\Lambda_t}{\sqrt{1-\Lambda_t}}\,\mathrm{d}t = \frac{2}{1-p}, \qquad \int_0^\infty \frac{\Lambda_t^2}{\sqrt{1-\Lambda_t}}\,\mathrm{d}t = \frac{4}{3(1-p)},$$

leading to the result

$$\|\Delta g_f\| \le \frac{8}{3}\sqrt{\frac{1}{2r\mathrm{e}p(1-p)^3}} K_r, \qquad r > 1/2. \tag{2.5}$$

Note that, for any p, this is at least $16K_r/\{3(1-p)\sqrt{2\mathrm{e}r}\}$, which is smaller than $2/(1-p)$ whenever $r > r_0$, for $r_0^{-1/2}K_{r_0} = 3\sqrt{2\mathrm{e}}/8$. Hence,

$$\|\Delta g_f\| \le \min\left\{\frac{2}{1-p}, \sqrt{\frac{r_0}{rp(1-p)^3}}\right\}, \tag{2.6}$$

and computation gives $\sqrt{r_0} \le 1.427 < 3/2$.

Finally, for any p, r, we can simply bound $\max_j \mathbb{P}[Z_{i-1}(t) = j]$ by 1 in (2.3), giving

$$\begin{aligned}\|\Delta g_f\| &\le \int_0^\infty \mathbb{E}\bigl[\bigl(Y_1(t) + Y_1'(t)\bigr)\bigl(Y_1(t) + Y_1'(t) - 1\bigr)\bigr]\,\mathrm{d}t \\ &\le 2\int_0^\infty \frac{\Lambda_t((1-3p)\Lambda_t + 2p)}{1-p}\,\mathrm{d}t = \frac{1+p}{(1-p)^2}.\end{aligned}$$

This bound is valid irrespective of the choices of $r > 0$ and $0 < p < 1$. □

Remark. Note that the bounds in Theorem 1.1 correspond exactly to the bounds derived in [2], in the limit when $rp \to \lambda$ and $p \to 0$, giving the Poisson case.

3. An application to a parasite model

The model that we use to describe the development over time of the number of parasites in a host is based on the immigration–birth–death process $Z_0^{[a,b]}$ of the previous section, with a the rate of ingestion of parasites and b their *per capita* birth rate. This model would imply exactly negative-binomially distributed parasite numbers in any age class. However, since in reality a can be expected to be variable, both between individuals and over time, we replace it by a function a_t, and investigate how much this influences the distribution of the number W of parasites at some fixed age T. We fix any $\bar{a} > 0$, to be thought of as a typical parasite ingestion rate, and define

$$\begin{aligned}A_t &:= \int_0^t (a_{T-s} - \bar{a})\mathrm{e}^{-(1-b)s}\,\mathrm{d}s; \qquad A_T^* := \sup_{0\le t\le T} |A_t|; \\ R_T &:= \frac{1-b}{b(1-\mathrm{e}^{-(1-b)T})}\int_0^T a_{T-s}\mathrm{e}^{-(1-b)s}\,\mathrm{d}s,\end{aligned}$$

also setting $\theta_T = \theta_T(b)$ and $R_a^* := \bar{a}/b$. A_t is a measure of the amount by which the cumulative exposure at time t under an ingestion rate of a_s, $0 \le s \le t$, differs from that with constant ingestion rate $\bar{a}$, allowing for the evolution of the parasites between ingestion and time T. Thus, both $|A_t|$ and A_T^* reflect how closely the choice of $\bar{a}$ corresponds to the actual ingestion rate. If $R_T = R_a^*$, then $A_T = 0$.

Theorem 3.1. *Under the above circumstances, we have*

$$\begin{aligned}&d_W\bigl(\mathrm{NB}\bigl(R_a^*, \theta_T\bigr), W\bigr) \\ &\quad\le |A_T| + 16\theta_T A_T^*\bigl(1 + \ln\bigl\{1/(1-\theta_T)\bigr\}\bigr)\min\left\{\frac{2}{1-\theta_T}, \frac{3}{2\sqrt{R_a^*\theta_T(1-\theta_T)^3}}\right\}.\end{aligned}$$

Remark. If $\{a_s\}$ includes a random component, $|A_T|$ and A_T^* should be replaced by their expectations in the bound given in the theorem.

Proof of Theorem 3.1. We define $N := \{N_s, 0 \le s \le T\}$ to be a Poisson process with mean function $\mathbb{E}N_t = \int_0^t a_u \,\mathrm{d}u$. Given that the points of N in $[0, T]$ are $\tau_1 < \tau_2 < \cdots$, we sample values $(X_j, j \ge 1)$ independently from the distributions $\mathcal{L}(Y_1^{[b]}(T - \tau_j))$, and let Ξ be the point process with $\Xi\{(0, s)\} := \sum_{j\,:\,\tau_j<s} X_j$. Then $W \overset{d}{=} \int_0^T \Xi(\mathrm{d}s)$. For each $f \in \mathcal{F}_W$, let $g := g_f$ be a solution to the Stein equation (1.1) with $p = \theta_T$ and $r = R_a^*$. Since

$$\mathcal{L}\bigl(W|N\{s\} = 1\bigr) = \mathcal{L}\bigl(W + Y_1^*(T - s)\bigr),$$

where $Y_1^* \overset{d}{=} Y_1^{[b]}$ is independent of W, we have

$$\begin{aligned}\mathbb{E}Wg(W) &= \mathbb{E}\biggl\{\int_0^T g\bigl(\Xi\bigl\{[0, T]\bigr\}\bigr)\Xi(\mathrm{d}s)\biggr\} = \sum_{j\ge1} j\mathbb{E}g(W + j)\int_0^T \mathbb{P}\bigl[Y_1^{[b]}(T - s) = j\bigr]a_s \,\mathrm{d}s\\ &= \sum_{j\ge1} \mathbb{E}g(W + j)jC_j^T,\end{aligned}$$

where $C_j^T := \int_0^T \mathbb{P}[Y_1^{[b]}(T - s) = j]a_s \,\mathrm{d}s$. Hence, for any r,

$$\begin{aligned}&\mathbb{E}\bigl(\theta_T(r + W)g(W + 1) - Wg(W)\bigr)\\ &\quad= r\theta_T\mathbb{E}g(W + 1) + \theta_T\sum_{j\ge1}\mathbb{E}g(W + j + 1)jC_j^T - \sum_{j\ge1}\mathbb{E}g(W + j)jC_j^T \qquad (3.1)\\ &\quad= \bigl(r\theta_T - C_1^T\bigr)\mathbb{E}g(W + 1) + \sum_{j\ge2}\mathbb{E}g(W + j)\bigl(\theta_T(j - 1)C_{j-1}^T - jC_j^T\bigr).\end{aligned}$$

Using Lemma 2.1, we can verify that

$$\begin{aligned}\sum_{j\ge1} jC_j^T &= \int_0^T a_s\sum_{j\ge1} j\mathbb{P}\bigl[Y_1^{[b]}(T - s) = j\bigr]\,\mathrm{d}s = \int_0^T a_s\mathbb{E}Y_1^{[b]}(T - s)\,\mathrm{d}s\\ &= \int_0^T a_s\mathrm{e}^{-(1-b)(T-s)}\,\mathrm{d}s,\end{aligned}$$

which in turn implies that

$$-\sum_{j\ge2}\bigl(\theta_T(j - 1)C_{j-1}^T - jC_j^T\bigr) = (1 - \theta_T)\sum_{j\ge1} jC_j^T - C_1^T = r\theta_T - C_1^T,$$

if $r = R_T$. Thus, it follows from (3.1) that

$$\begin{aligned}&\mathbb{E}\bigl(\theta_T(R_T + W)g(W + 1) - Wg(W)\bigr)\\ &\quad= \sum_{j\ge2}\bigl(\mathbb{E}g(W + j) - \mathbb{E}g(W + 1)\bigr)\bigl(\theta_T(j - 1)C_{j-1}^T - jC_j^T\bigr).\end{aligned} \qquad (3.2)$$

On the other hand, Lemma 2.1 shows that

$$\mathbb{P}\big[Y_1^{[b]}(t) = j\big] = (1-\theta_t)^2 \mathrm{e}^{-(1-b)t}\theta_t^{j-1}. \tag{3.3}$$

Hence, defining $\bar{C}_j^T := \bar{a}\int_0^T \mathbb{P}[Y_1^{[b]}(T-s) = j]\,\mathrm{d}s$, it follows that $\bar{C}_j^T = \bar{a}\theta_T^j/(jb)$, $j \geq 1$, which in turn gives

$$(j-1)\theta_T \bar{C}_{j-1}^T - j\bar{C}_j^T = 0. \tag{3.4}$$

Combining (3.2) and (3.4) and using Lemma 2.1 yields

$$\begin{aligned}
&\big|\mathbb{E}\big(\theta_T(R_T+W)g(W+1) - Wg(W)\big)\big| \\
&\quad \leq \|\Delta g\| \sum_{j\geq 2}(j-1)\big|\theta_T(j-1)\big(C_{j-1}^T - \bar{C}_{j-1}^T\big) - j\big(C_j^T - \bar{C}_j^T\big)\big| \\
&\quad = \|\Delta g\| \sum_{j\geq 2}(j-1) \\
&\qquad\qquad \times \left|\int_0^T (a_{T-s} - \bar{a})\big(\theta_T(j-1)\mathbb{P}\big[Y_1^{[b]}(s) = j-1\big] - j\mathbb{P}\big[Y_1^{[b]}(s) = j\big]\big)\,\mathrm{d}s\right|,
\end{aligned} \tag{3.5}$$

which, with (3.3), allows concrete estimates to be undertaken.

The simplest and most direct strategy is to impose bounds on $|a_{T-s} - \bar{a}|$. However, this may not lead to practically useful results. For instance, animals may sleep at night and graze during the day, so that a_s can have substantial variation, but over time scales typically much faster than the life history of the parasite. Instead, we prefer to formulate bounds expressed in terms of differences between cumulative exposure, which may more reasonably be expected to be small. For this reason, we write the quantity within the moduli in (3.5) as

$$\int_0^T (a_{T-t} - \bar{a})\big(\theta_T(j-1) - j\theta_t\big)\theta_t^{j-2}(1-\theta_t)^2\mathrm{e}^{-(1-b)t}\,\mathrm{d}t,$$

write $f_j(\theta) := (\theta_T(j-1) - j\theta)\theta^{j-2}(1-\theta)^2$ and integrate by parts, giving

$$A_T f_j(\theta_T) - \int_0^T A_t f_j'(\theta_t)\frac{\mathrm{d}\theta_t}{\mathrm{d}t}\,\mathrm{d}t, \tag{3.6}$$

where $A_t := \int_0^t (a_{T-s} - \bar{a})\mathrm{e}^{-(1-b)s}\,\mathrm{d}s$. Now the first term in (3.6) can easily be bounded, because $|f_j(\theta_T)| = (1-\theta_T)^2\theta_T^{j-1}$. For the second, we use the bound

$$\left|\int_0^T A_t f_j'(\theta_t)\frac{\mathrm{d}\theta_t}{\mathrm{d}t}\,\mathrm{d}t\right| \leq A_T^* \int_0^{\theta_T} \big|f_j'(\theta)\big|\,\mathrm{d}\theta. \tag{3.7}$$

Observe that

$$f_j'(\theta) = \theta^{j-3}\big\{(1-\theta_T) + (\theta_T - \theta)\big\}Q_j\big(\theta_T, j(\theta_T - \theta), j(1-\theta_T)\big), \tag{3.8}$$

where $Q_j(x,y,z)$ is a homogeneous multinomial of degree 2 in its arguments and has coefficients that are uniformly bounded in $j \geq 2$, with the coefficient of z^2 being zero. Hence, $|f_j'(\theta)|$ can be bounded above by replacing Q_j by $\widehat{Q}_j$ in (3.8), where $\widehat{Q}_j$ is obtained from Q_j by taking the absolute values of its coefficients. Integrating any of the terms from 0 to θ_T gives a bounded multiple of either $j^{-2}\theta_T^{j+1}$, $j^{-1}\theta_T^j(1-\theta_T)$ or $\theta_T^{j-1}(1-\theta_t)^2$ to go into (3.7), and multiplying each of these by $(j-1)$ and adding over $j \geq 2$, as required by (3.5), gives a multiple of $\theta_T^2 \log\{1/(1-\theta_T)\}$, θ_T^2 or θ_T, respectively. Hence, it follows that

$$\begin{aligned}
&\bigl|\mathbb{E}\bigl(\theta_T(R_T+W)g(W+1)-Wg(W)\bigr)\bigr| \\
&\quad\leq \|\Delta g\|\theta_T\bigl\{|A_T|+A_T^*\bigl(K_1+K_2\log\bigl\{1/(1-\theta_T)\bigr\}\bigr)\bigr\} \\
&\quad\leq K\|\Delta g\|\theta_T A_T^*\bigl(1+\log\bigl\{1/(1-\theta_T)\bigr\}\bigr),
\end{aligned} \tag{3.9}$$

for suitable constants K_1, K_2 and K. Careful computation in the Appendix shows that $K_1 \leq 34/3$ and $K_2 \leq 16$, giving $K \leq 16$.

We now use (1.4) of Theorem 1.1 to bound $\|\Delta g\|$ for all $g = g_f$, where $f \in \mathcal{F}_W$ and g_f satisfies (1.1) with $r = R_a^*$ and $p = \theta_T$; in particular, this gives

$$\|\Delta g\| \leq \min\biggl\{\frac{2}{1-\theta_T}, \frac{3}{2\sqrt{R_a^*\theta_T(1-\theta_T)^3}}\biggr\}.$$

Therefore, it follows from (3.9) that

$$\begin{aligned}
&\bigl|\mathbb{E}\bigl(\theta_T(R_T+W)g(W+1)-Wg(W)\bigr)\bigr| \\
&\quad\leq 16\theta_T A_T^*\bigl(1+\ln\bigl\{1/(1-\theta_T)\bigr\}\bigr)\min\biggl\{\frac{2}{1-\theta_T}, \frac{3}{2\sqrt{R_a^*\theta_T(1-\theta_T)^3}}\biggr\}.
\end{aligned} \tag{3.10}$$

But it is immediate from (1.2) and (1.3) that

$$d_W\bigl(\mathrm{NB}(R_T,\theta_T), \mathrm{NB}\bigl(R_a^*,\theta_T\bigr)\bigr) \leq \frac{\theta_T}{1-\theta_T}\bigl|R_T-R_a^*\bigr| = |A_T|,$$

completing the proof of the theorem. □

Remark. Note also that, if $b \to 0$ while a is held fixed, then $\theta_T \asymp b \to 0$, so that the upper bound in (3.10) approaches 0. In this limiting case, the number of parasites has precisely a Poisson distribution, even for time varying a, with mean $\mu_T := \int_0^T \mathrm{e}^{-(T-s)}a_s\,\mathrm{d}s$.

Similar considerations can be applied to the distribution of the total parasite burden $W := \sum_{i=1}^n W^{(i)}$ among n independent individuals, with their own functions $a^{(i)}$, $1 \leq i \leq n$, but all with the same b. First, defining $\overline{R} := n^{-1}\sum_{i=1}^n R_T^{(i)}$, it follows easily from (3.9) that

$$d_W\bigl(\mathcal{L}(W), \mathrm{NB}(n\overline{R},\theta_T)\bigr) \leq \sup_{f\in\mathcal{F}_W}\|\Delta g_f\|16\theta_T\sum_{i=1}^n\bigl(A_T^*\bigr)^{(i)}\bigl(1+\ln\bigl\{1/(1-\theta_T)\bigr\}\bigr),$$

where $g = g_f$ satisfies (1.1), with $r = n\overline{R}$ and $p = \theta_T$, also because $A_T^{(i)} = 0$ when approximating by $\mathrm{NB}(R_T^{(i)}, \theta_T)$. Hence, for example, from Theorem 1.1, if $n\overline{R} > r_0$,

$$d_W\big(\mathcal{L}(W), \mathrm{NB}(n\overline{R}, \theta_T)\big) \le \frac{24(1 + \log\{1/(1-\theta_T)\})\sqrt{\theta_T}}{(1-\theta_T)^{3/2}\sqrt{n\overline{R}}} \sum_{i=1}^{n} \big(A_T^*\big)^{(i)},$$

where r_0 is as for (2.6). Defining $\sigma := n^{-1}\sum_{i=1}^n (A_T^*)^{(i)}$, the bound grows with n roughly as $\sigma\sqrt{n/\overline{R}}$. However, the variability of the distribution $\mathrm{NB}(n\overline{R}, \theta_T)$ is also on the scale $\sqrt{n}$, so that the relevant measure of distance is $n^{-1/2} d_W(\mathcal{L}(W), \mathrm{NB}(n\overline{R}, \theta_T))$, which is small provided that $\sigma \ll \overline{R}$. If $\mathrm{NB}(n\overline{R}, \theta_T)$ is replaced by $\mathrm{NB}(nR_a^*, \theta_T)$, the additional term $|\sum_{i=1}^n A_T^{(i)}|$ in $d_W(\mathcal{L}(W), \mathrm{NB}(nR_a^*, \theta_T))$ is also roughly of order $\sigma\sqrt{n}$, if, for instance, the $A_T^{(i)}$ are independent random variables with mean zero.

Appendix

The constant K in (3.9) can be shown to satisfy $K \le 16$ as follows. Expression (3.8) can be written in a neat form:

$$\begin{aligned} f_j'(\theta) = \theta^{j-3}(1-\theta)\big\{ & j(j-1)\big((\theta_T-\theta)^2 + (\theta_T-\theta)(1-\theta_T)\big) \\ & - 2j\big(\theta_T - \theta^2\big) + 2\theta_T\big\}, \end{aligned}$$

from which it follows that

$$\big|f_2'(\theta)\big| \le 2(1-\theta)(3\theta + 1 + \theta_T) \tag{3.11}$$

and, for $j \ge 3$ and $0 \le \theta \le \theta_T$,

$$\begin{aligned} \big|f_j'(\theta)\big| \le \theta^{j-3}(1-\theta)\big\{ & j(j-1)\big((\theta_T-\theta)^2 + (\theta_T-\theta)(1-\theta_T)\big) \\ & + 2j\big(\theta_T - \theta^2\big) + 2\theta_T\big\}. \end{aligned} \tag{3.12}$$

Now, (3.11) yields

$$\int_0^{\theta_T} \big|f_2'(\theta)\big|\,\mathrm{d}\theta \le \int_0^{\theta_T} 2(1-\theta)(3\theta + 1 + \theta_T)\,\mathrm{d}\theta = 4\theta_T^2 + 2\theta_T - 3\theta_T^3,$$

and, for $j \ge 3$, integrating (3.12) gives

$$\int_0^{\theta_T} \big|f_j'(\theta)\big|\,\mathrm{d}\theta \le 3\theta_T^{j-1}(1-\theta_T)^2 + 4\theta_T^{j-1}\bigg(\frac{2}{j-2} - \frac{\theta_T^2}{j+1} - \frac{\theta_T}{j-1}\bigg).$$

Hence,

$$
\begin{aligned}
&\sum_{j\ge 2}(j-1)\int_0^{\theta_T}\bigl|f_j'(\theta)\bigr|\,\mathrm{d}\theta \\
&\quad\le \sum_{j\ge 3}(j-1)\biggl\{3\theta_T^{j-1}(1-\theta_T)^2+4\theta_T^{j-1}\biggl(\frac{2}{j-2}-\frac{\theta_T^2}{j+1}-\frac{\theta_T}{j-1}\biggr)\biggr\} \\
&\qquad + 4\theta_T^2+2\theta_T-3\theta_T^3 \qquad (3.13)\\
&\quad= -6\theta_T+14\theta_T^2-(14/3)\theta_T^3-8(\theta_T+1)\ln(1-\theta_T) \\
&\quad\le 2\theta_T+14\theta_T^2-(14/3)\theta_T^3-16\theta_T\ln(1-\theta_T)
\end{aligned}
$$

and

$$
\sum_{j\ge 2}(j-1)\bigl|A_T f_j(\theta_T)\bigr| = |A_T|\sum_{j\ge 2}(j-1)(1-\theta_T)^2\theta_T^{j-1} = |A_T|\theta_T. \qquad (3.14)
$$

Combining (3.5), (3.6), (3.7), (3.13) and (3.14) yields

$$
\begin{aligned}
&\bigl|\mathbb{E}\bigl(\theta_T(R_T+W)g(W+1)-Wg(W)\bigr)\bigr| \\
&\quad\le \|\Delta g\|\bigl\{3\theta_T+14\theta_T^2-(14/3)\theta_T^3-16\theta_T\ln(1-\theta_T)\bigr\} \\
&\quad\le \|\Delta g\|\theta_T A_T^*\bigl(37/3+16\ln\bigl\{1/(1-\theta_T)\bigr\}\bigr).
\end{aligned}
$$

Acknowledgements

Work supported in part by Australian Research Council Grants Nos DP120102728 and DP120102398.

References

[1] Barbour, A.D. and Jensen, J.L. (1989). Local and tail approximations near the Poisson limit. *Scand. J. Statist.* **16** 75–87. MR1003970

[2] Barbour, A.D. and Xia, A. (2006). On Stein's factors for Poisson approximation in Wasserstein distance. *Bernoulli* **12** 943–954. MR2274850

[3] Brown, T.C. and Xia, A. (2001). Stein's method and birth–death processes. *Ann. Probab.* **29** 1373–1403. MR1872746

[4] Kendall, D.G. (1948). On some modes of population growth leading to R. A. Fisher's logarithmic series distribution. *Biometrika* **35** 6–15. MR0026282

[5] Kendall, D.G. (1952). Les processus stochastiques de croissance en biologie. *Ann. Inst. H. Poincaré* **13** 43–108. MR0057526

[6] Kretzschmar, M. (1989). A renewal equation with a birth–death process as a model for parasitic infections. *J. Math. Biol.* **27** 191–221. MR0991050

[7] Phillips, M.J. (1996). Stochastic process approximation and network applications. Ph.D. thesis, Univ. Melbourne.

Received October 2013 and revised December 2013