Predicting Extinction or Explosion in a Galton-Watson Branching Process with Power Series Offspring Distribution*

Peter Guttorp† and Michael D. Perlman‡
University of Washington

September 22, 2014

Abstract

Extinction is certain in a Galton-Watson (GW) branching process if the offspring mean $\mu < 1$, whereas explosion is possible but not certain if $\mu > 1$. Discriminating between these two possibilities is a well-studied hypothesis-testing problem. However, deciding whether extinction or explosion will occur for the *current* realization of the process is a prediction problem. This can be formulated as a different testing problem by considering the conditional distributions of the process given extinction and explosion respectively. For power series offspring distributions, fixed-sample and sequential parametric tests are presented for the prediction problem and illustrated with data on the spread of epidemics and the populations of endangered species.

*Key words and phrases: Galton-Watson branching process; extinction; explosion; subcritical; supercritical; stochastic ordering; prediction; hypothesis testing; least favorable distribution; sequential probability ratio test; epidemic; endangered species.

†peter@stat.washington.edu. Research supported in part by National Science Foundation Grant DMS-1106862.

‡michael@stat.washington.edu. Research supported in part by U.S. Department of Defense Grant H98230-10-C-0263/0000 P0004.

Table of Contents

Acknowledgement: We are grateful to Brayan Ortiz for his assistance with numerical computations.

1. Introduction: the 2012 pertussis outbreak in Washington State

In 2011 the weekly numbers of new pertussis (whooping cough) cases in Washington State remained fairly constant, but in 2012 the numbers increased rapidly (Figure 1, CDC (2012)). Faced with the possibility of a pandemic, the governor declared a state-wide health emergency in Week 14 and an inoculation/quarantine program was begun.

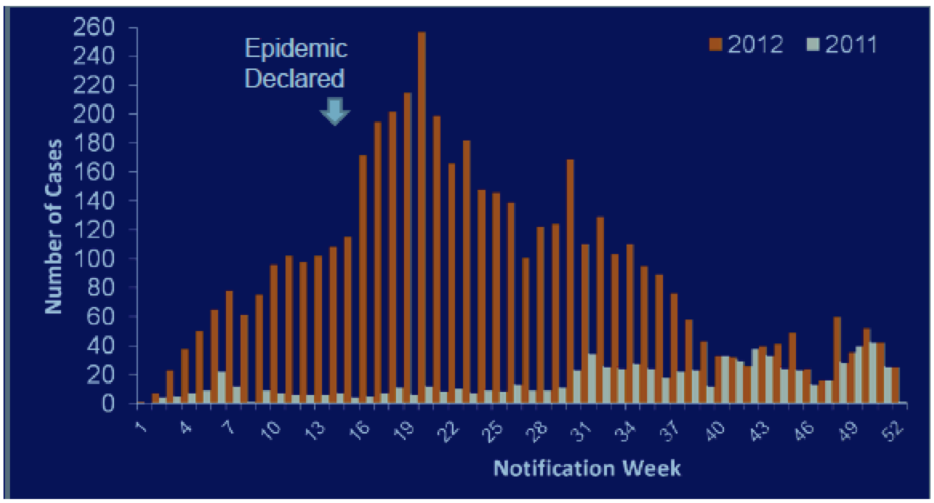

Figure 1: Weekly counts of new pertussis cases in Washington state.

The spread of an epidemic, at least in its initial stages, can be modeled as a classical Galton-Watson (GW) branching process, cf. §2. The question of predicting extinction or explosion is commonly formulated as that of testing subcriticality ($\mu < 1$) vs. supercriticality ($\mu > 1$), where μ denotes the mean number of infected offspring per individual case – cf. Becker (1974), Heyde (1979), Scott (1987).[1] Guttorp and Perlman (2014) use a decision-theoretic analysis to show, however, that this problem is more complex than previous literature suggests and that the basis of a standard test procedure is somewhat dubious.

Fortunately, this testing problem usually is not the one of actual interest, because a supercritical process still may terminate with positive probability. Of more interest is the problem of predicting whether the *current realization*

[1] Basawa and Scott (1976) and Sweeting (1978) treat a related testing problem for the supercritical case.

of a non-terminated process will terminate or explode. To our knowledge, this important problem has not been treated in the literature.

In §5-6 this prediction problem is formulated as a different hypothesis-testing problem based on the conditional distributions of the process given eventual extinction and explosion respectively. Unlike the original testing problem, this prediction problem often has relatively simple solutions in the fixed-sample (§5) and sequential sample (§6) cases, the latter based on the classical Wald sequential probability ratio test (SPRT), see §6. Using this procedure, explosion might have been predicted for the 2012 pertussis outbreak as early as Week 3; see Example 7.4.

Like the authors listed above who treated the original testing problem, we assume a parametric model for the offspring distribution, namely a *power series offspring distribution (psod)*; see §3.

The conditional distributions of a GW process given (eventual) extinction or explosion are discussed in §2, then specialized in §3 to the psod case. If the psod satisfies two total positivity conditions, these conditional distributions possess the stochastic monotonicity properties needed to justify our fixed-n and sequential prediction methods; see §4. Yaglom's (1947) well-known exponential approximation for the distribution of the population size is extended and sharpened in §5.3 and §5.4.

2. Conditional processes derived from a GW branching process

The Galton-Watson branching process is a discrete-time Markov chain that describes the growth or decline of a population that reproduces by simple branching, or splitting. Applications include nuclear chain reactions, epidemics, and the population size of endangered species. The classic reference is Harris (1963, Ch. I); also see Karlin (1966), Feller (1968), Athreya and Ney (1972), Jagers (1975), Taylor and Karlin (1984), Guttorp (1991).

For each $n = 0, 1, 2...$ let X_n denote the population size at generation n; assume that $X_0 = x_0 \geq 1$ is known. At generation $n = 0$ the ith individual is replaced by a random number $\xi_i^{(1)} \stackrel{d}{=} \xi$ of first-generation offspring, where the offspring random variable (rv) $\xi \equiv \xi_{\mathbf{p}}$ has probability distribution $\mathbf{p} \equiv (p_0, p_1, p_2, \dots)$ on $\{0, 1, 2, \dots\}$. The i-th individual in generation $n-1$ similarly is replaced by a random number $\xi_i^{(n)} \stackrel{d}{=} \xi$ of n-th generation offspring independently of its siblings. Thus the population size in the n-th generation satisfies

$$X_n = \xi_1^{(n)} + \cdots + \xi_{X_{n-1}}^{(n)}, \qquad n \geq 1, \tag{1}$$

where $\xi_1^{(n)}, \ldots, \xi_{X_{n-1}}^{(n)}$ are iid rvs, each $\stackrel{d}{=} \xi$. We assume that each $p_k < 1$ so the process is not deterministic, and that $p_0 > 0$ so extinction is possible.

Denote the probability generating function (pgf) of the offspring distribution by

$$\phi(s) \equiv \phi_{\mathbf{p}}(s) = \mathrm{E}_{\mathbf{p}}(s^{\xi}) = \sum_{k=0}^{\infty} p_k s^k, \qquad s \geq 0, \tag{2}$$

and let $1 \leq \rho \equiv \rho_{\mathbf{p}} \leq \infty$ be its radius of convergence. Note that $\phi(1) = 1$. Because $\phi(s)$ is convex and $p_1 < 1$, the equation

$$\phi(s) = s \tag{3}$$

has either one finite root or two distinct finite roots in $(0, \rho]$, one of which must be 1. If (3) has one finite root in $(0, \rho]$ denote it by $u \equiv u_{\mathbf{p}}$; if (3) has two distinct finite roots in $(0, \rho]$ denote them by $u \equiv u_{\mathbf{p}}$ and $v \equiv v_{\mathbf{p}}$, where $0 < u < v \leq \rho$.

If $x_0 = 1$, the pgf of X_n is the n-th functional iterate of ϕ, denoted by ϕ_n. For $x_0 \geq 1$ the pgf of X_n is $\phi_n^{x_0} \equiv (\phi_n)^{x_0}$. Either *extinction* ($X_n = 0$ for some $n \geq 1$) or *explosion* ($X_n \to \infty$) must occur; their probabilities are u^{x_0} and $1 - u^{x_0}$ respectively.

Denote the mean of the offspring distribution by $\mu \equiv \mu_{\mathbf{p}} = \mathrm{E}(\xi)$; then $\mu = \phi'(1)$. The GW process $\mathbf{X} \equiv \mathbf{X}_{\mathbf{p}}$ and its pgf $\phi \equiv \phi_{\mathbf{p}}$ are called *subcritical (resp., critical, supercritical)* if $\mu < 1$ ($\mu = 1$, $\mu > 1$); see Figure 2. In the subcritical case, $u = 1$ and v may or may not exist, see §2. In the critical case, $u = 1$ and v does not exist. In the supercritical case $0 < u < v = 1$, so both extinction and explosion occur with positive probability.

For a subcritical GW process, if v exists then $\mathbf{p}$, $\mathbf{X}$, and ϕ are called *extendable*; in this case $1 = u < v \leq \rho$ (see Figure 2). If $\rho = 1$ then $v > 1$ cannot exist so ϕ is not extendable, while if $\rho = \infty$ then ϕ is extendable since it grows at a quadratic rate or faster hence eventually crosses the 45° line a second time beyond $1 = u$. If $1 < \rho < \infty$ then ϕ is extendable iff $\phi(\rho) \geq \rho$.

Definition 2.1. *For a supercritical GW process* $\mathbf{X}$*, define the conditional processes*

$$\dot{\mathbf{X}} \equiv \mathbf{X}\big|\,\textit{extinction}, \tag{4}$$

$$\ddot{\mathbf{X}} \equiv \mathbf{X}\big|\,\textit{explosion}. \tag{5}$$

If $\mathbf{X}$ *is subcritical or critical, define* $\dot{\mathbf{X}} = \mathbf{X}$. □

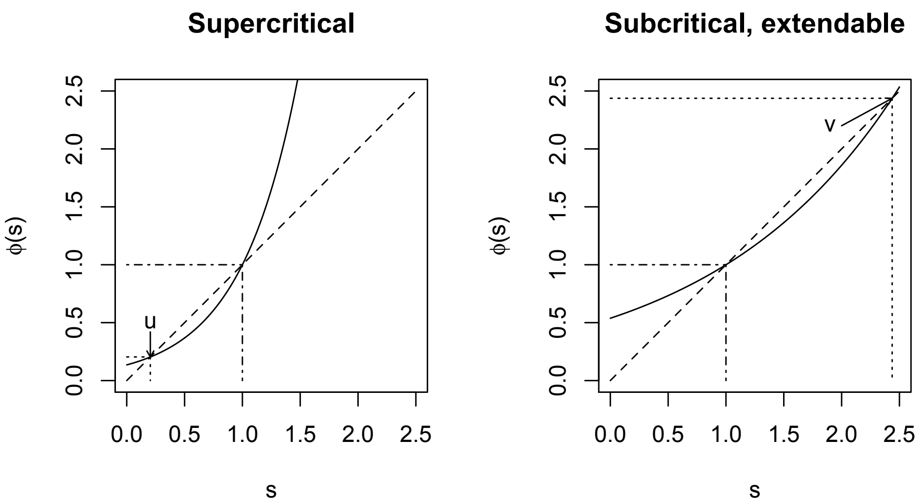

Figure 2: The duality between supercritical and extendable subcritical pgfs.

Proposition 2.1. *The set of supercritical GW processes conditional on extinction coincides with the set of subcritical extendable GW processes.*

Proof. If $\mathbf{X}$ is supercritical it is well known[2] that $\dot{\mathbf{X}}$ is a subcritical GW process with offspring pgf

$$\dot{\phi}(s) = \frac{\phi(us)}{u} \tag{6}$$

and offspring mean $\dot{\mu} = \phi'(u) < 1$. Furthermore $\dot{\phi}$ is extendable with second root $\dot{v} = 1/u$.

Now suppose that $\mathbf{X}$ is subcritical and extendable. Define

$$\tilde{\phi}(s) = \frac{\phi(vs)}{v}. \tag{7}$$

It is straightforward to verify that $\tilde{\phi}$ is a supercritical offspring pgf with offspring mean $\tilde{\mu} = \phi'(v) > 1$ and extinction probability $\tilde{u} = 1/v$. Denote the corresponding supercritical GW process by $\tilde{\mathbf{X}}$. Then

$$\dot{\tilde{\phi}}(s) = \frac{\phi(\tilde{u}vs)}{\tilde{u}v} = \phi(s). \tag{8}$$

[2]Waugh (1958, p.248), Athreya, Ney (1972, §I.12, Theorem 3), Guttorp (1991, p.101) .

Futhermore, if $\mathbf{X}$ is supercritical then

$$\tilde{\dot{\phi}}(s) = \frac{\phi(u\dot{v}s)}{u\dot{v}} = \phi(s). \tag{9}$$

This establishes the asserted result. $\square$

Successive conditioning on $X_1, \ldots, X_{n-1}$ in (1) shows that the joint probability mass function (pmf) $f \equiv f_{\mathbf{p}}$ of $\mathbf{X}_n \equiv (X_1, \ldots, X_n)$ is given by

$$f(\mathbf{x}_n) \equiv \Pr_{\mathbf{p}}[\mathbf{X}_n = \mathbf{x}_n] = \prod_{i=1}^{n} h_{\mathbf{p}}(x_{i-1}, x_i) \equiv h_{\mathbf{p}}(\mathbf{x}_n) \tag{10}$$

(e.g. Jagers (1975, eqn. (2.1.2)), where

$$h_{\mathbf{p}}(k, l) = \sum_{r_1 + \cdots + r_k = l} p_{r_1} \cdots p_{r_k} \tag{11}$$

Note that $h_{\mathbf{p}}(k, l)$ is the coefficient of s^l in the power series $[\phi_{\mathbf{p}}(s)]^k$.

From Bayes' formula, the pmf of $\dot{\mathbf{X}}_n \equiv (\dot{X}_1, \ldots, \dot{X}_n)$ is given by

$$\begin{aligned} \dot{f}(\mathbf{x}_n) \equiv \dot{f}_{\mathbf{p}}(\mathbf{x}_n) &= \Pr_{\mathbf{p}}[\,\mathbf{X}_n = \mathbf{x}_n \mid \textit{extinction}] \\ &= \frac{\Pr[\,\textit{extinction} \mid \mathbf{X}_n = \mathbf{x}_n\,] \Pr[\,\mathbf{X}_n = \mathbf{x}_n\,]}{\Pr[\,\textit{extinction}\,]} \\ &= u^{x_n - x_0}\, f(\mathbf{x}_n). \end{aligned} \tag{12}$$

Similarly the pmf of $\ddot{\mathbf{X}}_n \equiv (\ddot{X}_1, \ldots, \ddot{X}_n)$ is given by

$$\ddot{f}(\mathbf{x}_n) \equiv \ddot{f}_{\mathbf{p}}(\mathbf{x}_n) = \Big(\frac{1 - u^{x_n}}{1 - u^{x_0}}\Big) f(\mathbf{x}_n), \quad \mathbf{x}_n > 0, \tag{13}$$

where $\mathbf{x}_n > 0$ means that $x_1 > 0, \ldots, x_n > 0$. From (12) and (13), $\dot{\mathbf{X}}$ and $\ddot{\mathbf{X}}$ are Markovian with transition probabilities

$$\dot{f}(x_n | x_{n-1}) = u^{x_n - x_{n-1}} h_{\mathbf{p}}(x_{n-1}, x_n), \tag{14}$$

$$\ddot{f}(x_n | x_{n-1}) = \Big(\frac{1 - u^{x_n}}{1 - u^{x_{n-1}}}\Big) h_{\mathbf{p}}(x_{n-1}, x_n), \quad x_{n-1}, x_n > 0, \tag{15}$$

respectively. However, $\ddot{\mathbf{X}}$ is not a GW process because some individuals may die without offspring even though explosion occurs.

3. The GW process with power series offspring distribution

Following Becker (1974) we now specialize this discussion to a parametric model for the offspring distribution $\mathbf{p} \equiv (p_0, p_1, \dots)$. The *power series offspring distribution*[3] *(psod)* $\mathbf{p}_\theta \equiv (p_{\theta;0}, p_{\theta;1}, \dots)$ is given by

$$p_{\theta;k} = \frac{a_k \theta^k}{A(\theta)}, \quad k = 0, 1, \dots, \quad 0 < \theta < \psi, \tag{16}$$

where $(a_0, a_1, \dots) \equiv \mathbf{a}$ are nonnegative constants, θ is the unknown parameter, $A(\theta) = \sum a_k \theta^k$, and $0 < \psi \leq \infty$ is the radius of convergence of $A(\cdot)$. We assume that $a_0 > 0$ so extinction is possible, and that $a_k > 0$ for at least one $k \geq 2$ so growth is possible; without loss of generality we may take $a_0 = 1$. For simplicity of exposition *we limit attention to the case where* $A(\psi-) = \infty$; this includes the familiar Poisson, binomial, geometric, negative binomial, binary splitting, and logarithmic series distributions.

Denote $\mathbf{X}_{\mathbf{p}_\theta}$, $\xi_{\mathbf{p}_\theta}$, $f_{\mathbf{p}_\theta}$, $\phi_{\mathbf{p}_\theta}$, $\rho_{\mathbf{p}_\theta}$, $u_{\mathbf{p}_\theta}$, $v_{\mathbf{p}_\theta}$, $\mu_{\mathbf{p}_\theta}$ by $\mathbf{X}_\theta$, ξ_θ, f_θ, ϕ_θ, ρ_θ, u_θ, v_θ, μ_θ respectively. By (2) and (16), ϕ_θ has radius of convergence $\rho_\theta = \psi/\theta$ and

$$\phi_\theta(s) = \frac{A(\theta s)}{A(\theta)} = \frac{B(\theta s)}{B(\theta)} s, \quad 0 < s < \rho_\theta, \tag{17}$$

where $B(\theta) = A(\theta)/\theta$ (see Figure 3). Here $B(\theta)$ is a strictly convex positive function on $(0, \psi)$ with $B(0+) = B(\psi-) = \infty$, so $B(\cdot)$ has a unique minimum at some $\tau \in (0, \psi)$ with $B'(\tau) = 0$; $B(\theta)$ is strictly decreasing for $\theta < \tau$ and strictly increasing for $\theta > \tau$.

It follows from (17) that for $\theta \in (0, \psi)$,

$$\mathrm{E}_\theta(\xi) \equiv \mu_\theta = \frac{\theta A'(\theta)}{A(\theta)}, \tag{18}$$

$$\frac{\mu_\theta - 1}{\theta} = \frac{B'(\theta)}{B(\theta)} = \frac{d \log B(\theta)}{d\theta}, \tag{19}$$

$$\mathrm{Var}_\theta(\xi) \equiv \sigma_\theta^2 = \theta \frac{d\mu_\theta}{d\theta}. \tag{20}$$

By (19), $\mu_\tau = 1$ so $\mathbf{X}_\tau$ is critical. By (20), μ_θ is strictly increasing in θ, hence

[3] This terminology apparently is due to Noack (1950); this is just the general one-parameter exponential family, introduced in 1935-6 by Darmois, Pitman, and Koopmans, on the nonnegative integers.

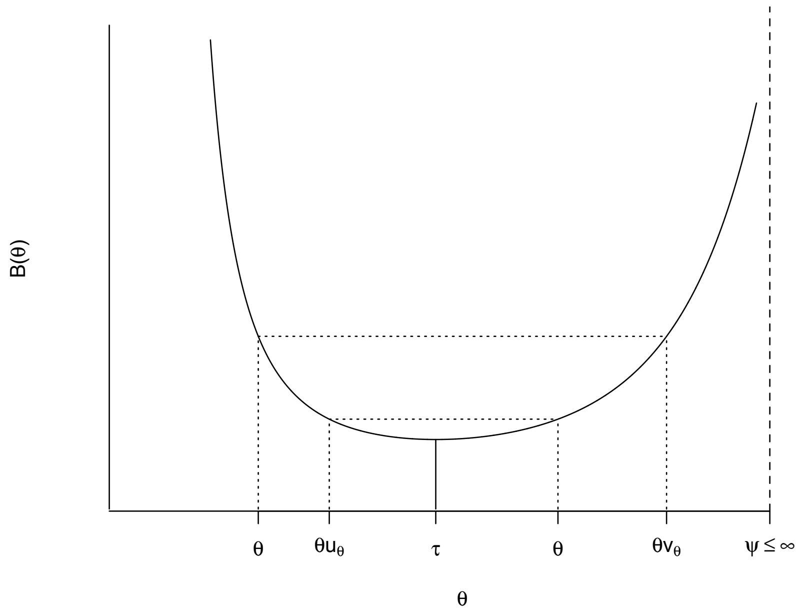

Figure 3: The function $B(\theta) = A(\theta)/\theta$.

the subcritical and supercritical parameter spaces are both open intervals:

$$\{\theta \,|\, \mu_\theta < 1\} \quad = \quad (0, \tau) \quad \text{(subcritical)}, \tag{21}$$

$$\{\theta \,|\, \mu_\theta > 1\} \quad = \quad (\tau, \psi) \quad \text{(supercritical)}. \tag{22}$$

If $\theta \in (\tau, \psi)$ then from (3) and (17), u_θ is the unique solution to

$$B(\theta u_\theta) = B(\theta), \quad 0 < \theta u_\theta < \tau. \tag{23}$$

If $\theta \in (0, \tau)$ then v_θ is the unique solution to

$$B(\theta) = B(\theta v_\theta), \quad \tau < \theta v_\theta < \psi. \tag{24}$$

Thus each subcritical $\mathbf{X}_\theta$ is extendable. It follows from the uniqueness of the solutions of (23) and (24) that

$$v_{\theta u_\theta} \quad = \quad u_\theta^{-1} \qquad \text{for } \theta \in (\tau, \psi), \tag{25}$$

$$u_{\theta v_\theta} \quad = \quad v_\theta^{-1} \qquad \text{for } \theta \in (0, \tau). \tag{26}$$

$\square$

Proposition 3.1. (i) *For $\theta \in (\tau, \psi)$, θu_θ strictly decreases from τ to 0; u_θ strictly decreases from 1 to 0.*

(ii) *For* $\theta \in (0, \tau)$, θv_θ *strictly decreases from* ψ *to* τ*;* v_θ *strictly decreases from* ∞ *to 1.*

Proof. (i) It follows from (23) and (19) that for $\theta \in (\tau, \psi)$,

$$d_\theta \equiv \frac{d(\theta u_\theta)}{d\theta} \quad = \quad \left(\frac{\mu_\theta - 1}{\mu_{\theta u_\theta} - 1}\right) u_\theta. \tag{27}$$

Thus $d_\theta < 0$ because $\theta u_\theta < \tau < \theta$, so θu_θ is strictly decreasing, a fortiori u_θ is strictly decreasing. As $\theta \downarrow \tau$, $B(\theta) \downarrow B(\tau)$, its unique minimum, hence $\theta u_\theta \uparrow \tau$ by (23), so $u_\theta \uparrow 1$. As $\theta \uparrow \psi$, $B(\theta) \uparrow \infty$, hence $\theta u_\theta \downarrow 0$ by (23), so $u_\theta \downarrow 0$.

(ii) It follows from (24) and (19) that for $\theta \in (0, \tau)$,

$$\frac{d(\theta v_\theta)}{d\theta} = \left(\frac{\mu_\theta - 1}{\mu_{\theta v_\theta} - 1}\right) v_\theta, \tag{28}$$

which is < 0 because $\tau < \theta v_\theta < \psi$. The remaining results are verified as in (i). Alternatively, (25) and (26) can be applied to obtain $v_{\tau-}$ and v_{0+}. $\square$

Proposition 3.1, (25), and (26) establish analytically a 1-1 relation between the subcritical $(0, \tau)$ and supercritical (τ, ψ) parameter spaces. The corresponding probabilistic relation between the subcritical and supercritical processes themselves is now presented.

Proposition 3.2. *The set of supercritical processes* $\{\dot{\mathbf{X}}_\theta \mid \theta \in (\tau, \psi)\}$ *conditional on extinction coincides with the set of subcritical processes* $\{\mathbf{X}_\theta \mid \theta \in (0, \tau)\}$*. Specifically,*

$$\dot{\mathbf{X}}_\theta \stackrel{d}{=} \mathbf{X}_{\theta u_\theta}, \qquad \theta \in (\tau, \psi), \tag{29}$$

$$\mathbf{X}_\theta \stackrel{d}{=} \dot{\mathbf{X}}_{\theta v_\theta}, \qquad \theta \in (0, \tau). \tag{30}$$

(Note too that $\dot{\mathbf{X}}_\tau \stackrel{d}{=} \mathbf{X}_\tau$.)

Proof. Suppose first that $\mathbf{X}_\theta$ is supercritical, i.e., $\theta \in (\tau, \psi)$. From (6), $\dot{\mathbf{X}}_\theta$ is a subcritical GW process with offspring pgf in the same psod family (16):

$$\dot{\phi}_\theta(s) = \frac{A(\theta u_\theta s)}{A(\theta) u_\theta} = \frac{A(\theta u_\theta s)}{A(\theta u_\theta)} = \phi_{\theta u_\theta}(s); \tag{31}$$

cf. Becker (1974, p.394). Since $\theta u_\theta < \tau$, $\mathbf{X}_{\theta u_\theta}$ is subcritical.

Suppose next that $\mathbf{X}_\theta$ is subcritical, i.e., $\theta \in (0, \tau)$. A similar argument using (7) shows that $\tilde{\mathbf{X}}_\theta$ is a supercritical GW process with offspring pgf

$$\tilde{\phi}_\theta(s) = \phi_{\theta v_\theta}(s). \tag{32}$$

Now apply (8) to obtain $\phi_\theta = \dot{\phi}_{\theta v_\theta}$; since $\theta v_\theta > \tau$, $\mathbf{X}_{\theta v_\theta}$ is supercritical. $\square$

Example 3.1: the Poisson(θ) psod. Here $p_{k;\theta} = e^{-\theta}\theta^k/k!$, $0 < \theta < \infty$, so $a_k = 1/k!$, $A(\theta) = e^\theta$, $\psi = \infty$, $A(\psi-) = \infty$, $B(\theta) = e^\theta/\theta$, and $\phi_\theta(s) = e^{\theta(s-1)}$. Then $\mu_\theta = \sigma_\theta^2 = \theta$, $\tau = 1$, and u_θ and v_θ satisfy the equation

$$e^{\theta(s-1)} = s. \tag{33}$$

by (23) and (24). This cannot be solved explicitly, but necessarily

$$\begin{cases} u_\theta = 1,\ v_\theta > 1 & \text{if } \theta < 1 \text{ (subcritical);} \\ u_\theta < 1,\ v_\theta = 1 & \text{if } \theta > 1 \text{ (supercritical).} \end{cases} \tag{34}$$

Example 3.2: the negative binomial NB(r, θ) and geometric GM(θ) psods. For fixed $r > 0$, the NB(r, θ) psod has $p_{k;\theta} = \frac{\Gamma(r+k)}{\Gamma(r)k!}(1-\theta)^r\theta^k$, $0 < \theta < 1$. Here $a_k = \frac{\Gamma(r+k)}{\Gamma(r)k!}$, $A(\theta) = \frac{1}{(1-\theta)^r}$, $\psi = 1$, $B(\theta) = \frac{1}{[\theta(1-\theta)^r]}$, and $\phi_\theta(s) = \frac{(1-\theta)^r}{(1-\theta s)^r}$. Also $\mu_\theta = \frac{r\theta}{1-\theta}$, $\sigma_\theta^2 = \frac{r\theta}{(1-\theta)^2}$, $\tau = \frac{1}{1+r}$, and u_θ and v_θ satisfy the equation

$$(1-\theta)^r = (1-\theta s)^r s. \tag{35}$$

This can be solved explicitly for the GM$(\theta) \equiv$ NB$(1, \theta)$ psod where $r = 1$ and $\tau = 1/2$:

$$\begin{cases} u_\theta = 1, & v_\theta = \frac{1-\theta}{\theta} \quad \text{if } \theta < \frac{1}{2} \text{ (subcritical);} \\ u_\theta = \frac{1-\theta}{\theta}, & v_\theta = 1 \quad \text{if } \theta > \frac{1}{2} \text{ (supercritical).} \end{cases} \tag{36}$$

Here the relations (25) and (26) can be verified directly. $\square$

Example 3.3: binary splitting. Take $a_0 = a_2 = 1$ and $a_k = 0$ for $k \neq 2$. Thus $A(\theta) = 1 + \theta^2$ for $0 < \theta < \infty \equiv \psi$, so $p_{0;\theta} = \frac{1}{1+\theta^2}$, $p_{2;\theta} = \frac{\theta^2}{1+\theta^2}$, and $p_{k;\theta} = 0$ for $k \neq 0, 2$. Here $B(\theta) = \theta^{-1} + \theta$, $\phi_\theta(s) = \frac{1+\theta^2 s^2}{1+\theta^2}$, $\mu_\theta = \frac{2\theta^2}{1+\theta^2}$, $\sigma_\theta^2 = \frac{4\theta^2}{(1+\theta^2)^2}$, $\tau = 1$, and

$$\begin{cases} u_\theta = 1, & v_\theta = \frac{1}{\theta^2} \quad \text{if } \theta < 1 \text{ (subcritical);} \\ u_\theta = \frac{1}{\theta^2}, & v_\theta = 1 \quad \text{if } \theta > 1 \text{ (supercritical).} \end{cases} \tag{37}$$

Again the relations (25) and (26) can be verified directly. $\square$

4. Stochastic orderings for a psod GW process

Let W, Z, $\mathbf{W}_n \equiv (W_1, \dots, W_n)$, $\mathbf{Z}_n \equiv (Z_1, \dots, Z_n)$, and $\mathbf{W} \equiv (W_1, \dots)$, $\mathbf{Z} \equiv (Z_1, \dots)$ be nonnegative-integer-valued random variables, random vectors, and discrete-time stochastic processes, respectively. We say W is *stochastically smaller* than Z, written $W \prec Z$, if $\mathrm{E}[g(W)] \le \mathrm{E}[g(Z)]$ for all increasing bounded nonnegative functions g on the nonnegative integers $\mathbb{Z}_0$ with strict inequality for at least one g. It is straightforward to show that if U, V, W, Z are independent, then

$$U \prec V \text{ and } W \prec Z \implies U + W \prec V + Z. \tag{38}$$

Similarly, we write $\mathbf{W}_n \prec \mathbf{Z}_n$ if

$$\mathrm{E}[g(\mathbf{W}_n)] \le \mathrm{E}[g(\mathbf{Z}_n)] \tag{39}$$

for all increasing bounded nonnegative functions g on $\mathbb{Z}_0^n$ with strict inequality for at least one g. Finally, we write $\mathbf{W} \prec \mathbf{Z}$ if $\mathbf{W}_n \prec \mathbf{Z}_n$ for all $n = 1, 2, \dots$. The next lemma follows directly from (1) and (38).

Lemma 4.1. *Let $\mathbf{X}$ and $\mathbf{X}'$ be GW processes with offspring rv's ξ and ξ' respectively. If $\xi \prec \xi'$ then $\mathbf{X} \prec \mathbf{X}'$.* $\square$

Stochastic orderings satisfied by a GW process $\mathbf{X}_\theta$ with psod (16) and by the conditional processes $\dot{\mathbf{X}}_\theta$ and $\ddot{\mathbf{X}}_\theta$ are now developed. These will be useful for the testing and prediction problems treated below.

From (10), (11), and (16), the pmf of $(\mathbf{X}_\theta)_n$ is

$$f_\theta(\mathbf{x}_n) = \frac{\theta^{y_n - x_0}}{(A(\theta))^{y_{n-1}}} h_{\mathbf{a}}(\mathbf{x}_n), \qquad \mathbf{x}_n \in R'_{\mathbf{a},n}, \tag{40}$$

where $y_n = x_0 + x_1 + \cdots + x_n$ and $R'_{\mathbf{a},n} = \{\mathbf{x}_n \mid h_{\mathbf{a}}(\mathbf{x}_n) > 0\}$. Then (12), (13), and (40) give the following:

$$\text{for } \theta > 0 : \qquad \dot{f}_\theta(\mathbf{x}_n) = u_\theta^{x_n - x_0} \frac{\theta^{y_n - x_0}}{(A(\theta))^{y_{n-1}}} h_{\mathbf{a}}(\mathbf{x}_n), \tag{41}$$

$$\text{for } \theta > \tau : \qquad \ddot{f}_\theta(\mathbf{x}_n) = \Big(\frac{1 - u_\theta^{x_n}}{1 - u_\theta^{x_0}}\Big) \frac{\theta^{y_n - x_0}}{(A(\theta))^{y_{n-1}}} h_{\mathbf{a}}(\mathbf{x}_n), \quad \mathbf{x}_n > 0. \tag{42}$$

The transition probabilities are obtained from (40)-(42) (recall (14)-(15)):

$$f_\theta(x_n|x_{n-1}) = \frac{\theta^{x_n}}{(A(\theta))^{x_{n-1}}} h_{\mathbf{a}}(x_{n-1}, x_n), \tag{43}$$

$$\dot{f}_\theta(x_n|x_{n-1}) = u_\theta^{x_n - x_{n-1}} \frac{\theta^{x_n}}{(A(\theta))^{x_{n-1}}} h_{\mathbf{a}}(x_{n-1}, x_n), \tag{44}$$

$$\ddot{f}_\theta(x_n|x_{n-1}) = \Big(\frac{1-u_\theta^{x_n}}{1-u_\theta^{x_{n-1}}}\Big) \frac{\theta^{x_n}}{(A(\theta))^{x_{n-1}}} h_{\mathbf{a}}(x_{n-1}, x_n), \quad x_{n-1}, x_n > 0. \tag{45}$$

The definitions of $\ddot{f}_\theta(\cdot)$ and $\ddot{f}_\theta(\cdot|\cdot)$ can be extended to the critical case $\theta = \tau$:

$$\begin{aligned} \ddot{f}_\tau(\mathbf{x}_n) &= \lim_{\theta\downarrow\tau} \ddot{f}_\theta(\mathbf{x}_n) && (46)\\ &= \frac{x_n}{x_0} \frac{\tau^{y_n - x_0}}{(A(\tau))^{y_{n-1}}} h_{\mathbf{a}}(\mathbf{x}_n), \quad \mathbf{x}_n > 0; && (47)\\ \ddot{f}_\tau(x_n|x_{n-1}) &= \frac{x_n}{x_{n-1}} \frac{\tau^{x_n}}{(A(\tau))^{x_{n-1}}} h_{\mathbf{a}}(x_{n-1}, x_n), \quad x_{n-1}, x_n > 0. && (48) \end{aligned}$$

Denote the resulting Markov process by $\ddot{\mathbf{X}}_\tau$.[4] By (46) and Scheffe's Theorem,

$$\ddot{\mathbf{X}}_\theta \xrightarrow{L_1} \ddot{\mathbf{X}}_\tau \quad \text{as } \theta \downarrow \tau. \tag{49}$$

Proposition 4.1. (i) $\mathbf{X}_\theta$ *is stochastically increasing for* $\theta \in (0, \psi)$*, that is,* $\theta < \theta' \Rightarrow \mathbf{X}_\theta \prec \mathbf{X}_{\theta'}$.
(ii) $\dot{\mathbf{X}}_\theta$ *is stochastically decreasing for* $\theta \in [\tau, \psi)$*, that is,* $\theta < \theta' \Rightarrow \dot{\mathbf{X}}_\theta \succ \dot{\mathbf{X}}_{\theta'}$.

Proof. (i) follows from Lemma 4.1 since $\theta < \theta' \Rightarrow \xi_\theta \prec \xi_{\theta'}$ by the strict monotone likelihood ratio (MLR) property of the psod family.[5] (ii) follows from (i), (29), and Proposition 3.1(i). □

The verifications of the stochastic orderings of $\mathbf{X}_\theta$ and $\dot{\mathbf{X}}_\theta$ are straightforward because these are GW processes. However, $\ddot{\mathbf{X}}_\theta$ is not a GW process so its stochastic ordering properties if any are not apparent. Although it might appear that $\ddot{\mathbf{X}}_\theta$ should inherit the stochastic increasing property of $\mathbf{X}_\theta$, upon

[4]This is not to be interpreted as $\mathbf{X}_\tau$ | explosion, which is vacuous.
[5]Lehmann and Romano (2005) Lemma 3.4.2 and Problem 3.39; Karlin (1968) Proposition 3.1 and the discussion following Proposition 3.3, both in Chapter 1.

closer examination this is not obvious. Conditional on ultimate explosion, as θ increases above the critical value τ those trajectories of $\mathbf{X}_\theta$ with relatively small initial values might have increasing likelihood of survival, hence for fixed n, $(\ddot{\mathbf{X}}_\theta)_n$ might tend to decrease stochastically, not increase.

In Proposition 4.2(iii) it will be shown, however, that $\ddot{\mathbf{X}}_\theta$ is indeed stochastically increasing for $\theta \geq \tau$ provided that two additional conditions are imposed, namely TP2a and/or TP2b (see below) based on total positivity of order 2 (TP2). Also, it is shown in Proposition 4.2(i) that under TP2a alone, the conditional random vector $(\mathbf{X}_\theta)_n \mid X_{\theta,n} > 0$ is stochastically increasing for $\theta \leq \tau$.

Karlin (1968) is the primary reference for total positivity. The TP2 property is equivalent to MLR, cf. Lehmann and Romano (2005, Problem 50)). The following results for the TP2 and FKG properties appear in Kemperman (1977) and Perlman and Olkin (1980).

Definition 4.1. *Let $f(\mathbf{x})$ be a nonnegative function defined on a measurable rectangle $\mathbf{R} = \prod_{i=1}^n R_i \subseteq \mathbb{R}^n$. Then f satisfies the FKG condition on $\mathbf{R}$ if*

$$f(\mathbf{x}_n)f(\mathbf{y}_n) \leq f(\mathbf{x}_n \wedge \mathbf{y}_n)f(\mathbf{x}_n \vee \mathbf{y}_n) \quad \forall \mathbf{x}_n, \mathbf{y}_n \in \mathbf{R}, \tag{50}$$

where $\mathbf{x}_n \wedge \mathbf{y}_n = (x_1 \wedge y_1, \ldots, x_n \wedge y_n)$ and $\mathbf{x}_n \vee \mathbf{y}_n = (x_1 \vee y_1, \ldots, x_n \vee y_n)$; we say that f is FKG on $\mathbf{R}$. TP2 is defined as FKG for $n = 2$. $\square$

Some properties of TP2 and FKG: If $h(x_i, x_j)$ is TP2 on $R_i \times R_j$ in a single pair (x_i, x_j) then $f(\mathbf{x}_n) \equiv h(x_i, x_j)$ is FKG on $\mathbf{R}$. If $f_1, \ldots, f_m$ are FKG on $\mathbf{R}$ then so is $\prod_{i=1}^m f_i$. If $f(\mathbf{x}_n) = h_i(x_i)$ for a single i then f is trivially FKG on $\mathbf{R}$, so $f(\mathbf{x}_n) = \prod_{i=1}^n h_i(x_i)$ is also trivially FKG. If f is FKG on $\mathbf{R}^* \equiv \prod R_i^*$ and if, for each $i = 1, \ldots, n$, $\beta_i : R_i \to R_i^*$ is increasing in x_i, then $f(\beta_1(x_1), \ldots, f(\beta_n(x_n))$ is FKG on $\mathbf{R} \equiv \prod R_i$.

Lemma 4.2. *(The FKG Inequality). Let $\mathbf{Z}$ be a random vector with an FKG pdf f on $\mathbf{R}$ w.r.to a product measure ν and let g, h be component-wise increasing nonnegative functions on $\mathbf{R} \cap \{f > 0\}$. Then*

$$\mathrm{E}[g(\mathbf{Z})h(\mathbf{Z})] \geq \mathrm{E}[g(\mathbf{Z})]\mathrm{E}[h(\mathbf{Z})]. \tag{51}$$

Strict inequality holds in (51) *if g is nonconstant w.r.to f* ($\Pr[g(Z) = c] < 1$ *for all constants c) and h is strictly increasing.*

Proof. Perlman and Olkin (1980, Propositions 2.4, 2.6, and Remark 2.5.) $\square$

Condition TP2a: $h_{\mathbf{a}}(x,y)$ *is TP2 in* (x,y) *for* $x,y=1,2,\ldots$. *(Note that* $h_{\mathbf{a}}(x,y)$ *is the coefficient of* θ^y *in the power series* $[A(\theta)]^x$*.)*

Condition TP2b: $(1-u_\theta^x)\theta^x$ *is TP2 in* (x,θ) *for* $x=1,2,\ldots$ *and* $\tau<\theta<\psi$.

A sufficient condition for TP2a to hold is that $\{a_k|k=0,1,\ldots\}$ is a one-sided Polya frequency sequence of order 2 (PF2); cf. Karlin (1968, (ii) on pp.142-3, also Ch.8).

Let $(\mathbf{X}_\theta)_n^+$ denote the conditional random vector $(\mathbf{X}_\theta)_n \mid X_{\theta,n}>0$. For notational convenience the subscript θ sometimes will be omitted. The conditional pmf of $(\mathbf{X}_\theta)_n^+ \equiv \mathbf{X}_n^+$ is given by

$$f_\theta^+(\mathbf{x}_n)=\Pr_\theta[\mathbf{X}_n=\mathbf{x}_n|X_n>0]=b_{\theta,n}f_\theta(\mathbf{x}_n),\quad \mathbf{x}_n>0, \tag{52}$$

where $b_{\theta,n}^{-1}=\Pr_\theta[X_n>0]$. Note that $X_n>0 \Rightarrow \mathbf{X}_n>0$. Clearly $\mathbf{X}_n^+\succ\mathbf{X}_n$ and $\dot{\mathbf{X}}_n^+\succ\dot{\mathbf{X}}_n$ for all $\theta>0$, while $\ddot{\mathbf{X}}_n^+\equiv\ddot{\mathbf{X}}_n$ for $\theta\geq\tau$.

Proposition 4.2. (i) *If TP2a holds then for each* $n\geq 1$, $(\mathbf{X}_\theta)_n^+$ *is stochastically increasing for* $\theta\in(0,\tau]$. *Therefore, by Propositions 3.1 and 3.2,* $(\dot{\mathbf{X}}_\theta)_n^+ \stackrel{d}{=} (\mathbf{X}_{\theta u_\theta})_n^+$ *is stochastically decreasing for* $\theta\in[\tau,\psi)$.
(ii) *If TP2a holds then for each* $n\geq 1$, $(\dot{\mathbf{X}}_\tau)_n \prec (\dot{\mathbf{X}}_\tau)_n^+ \prec (\ddot{\mathbf{X}}_\tau)_n^+ \equiv (\ddot{\mathbf{X}}_\tau)_n$.
(iii) *If TP2a and TP2b hold,* $\ddot{\mathbf{X}}_\theta$ *is stochastically increasing for* $\theta\in[\tau,\psi)$.

Proof. (i) We will show that $\mathrm{E}_\theta[g(\mathbf{X}_n^+)]$ is strictly increasing in $\theta\in(0,\tau]$ for any increasing bounded nonconstant $g\geq 0$ on $\mathbb{Z}_+^n$, where $\mathbb{Z}_+$ denotes the positive integers. The FKG inequality will yield the required result as follows: for $0<\theta_1<\theta_2\leq\tau$,

$$\begin{aligned}
\mathrm{E}_{\theta_2}[g(\mathbf{X}_n^+)] &= \mathrm{E}_{\theta_1}\left[g(\mathbf{X}_n^+)\frac{f_{\theta_2}^+(\mathbf{X}_n^+)}{f_{\theta_1}^+(\mathbf{X}_n^+)}\right]\\
&> \mathrm{E}_{\theta_1}[g(\mathbf{X}_n^+)]\,\mathrm{E}_{\theta_1}\left[\frac{f_{\theta_2}^+(\mathbf{X}_n^+)}{f_{\theta_1}^+(\mathbf{X}_n^+)}\right]\\
&= \mathrm{E}_{\theta_1}[g(\mathbf{X}_n^+)].
\end{aligned}$$

To apply the FKG inequality (51) with strict inequality it must be shown that (a) $f_{\theta_1}^+(\mathbf{x}_n)$ is FKG on $\mathbb{Z}_+^n$; and (b) the ratio $r(\mathbf{x}_n)\equiv\frac{f_{\theta_2}^+(\mathbf{x}_n)}{f_{\theta_1}^+(\mathbf{x}_n)}$ is strictly increasing on $\mathbb{Z}_+^n\cap\{f_{\theta_1}^+(\mathbf{x}_n)\}=\mathbb{Z}_+^n\cap\{h_{\mathbf{a}}(\mathbf{x}_n)>0\}$. First, for all $\theta>0$ and $\mathbf{x}_n>0$, it follows from (40) and (52) that

$$f_\theta^+(\mathbf{x}_n)=\frac{b_{\theta,n}\theta^{x_1+\cdots+x_n}}{(A(\theta))^{x_0+\cdots+x_{n-1}}}\prod_{i=1}^n h_{\mathbf{a}}(x_{i-1},x_i),\quad \mathbf{x}_n>0. \tag{53}$$

By TP2a each factor $h_{\mathbf{a}}(x_{i-1}, x_i)$ in (53) is TP2, hence their product is FKG, thus so is $f_\theta^+(\mathbf{x}_n)$; this gives (a). Next, $0 < \theta_1 < \theta_2 \le \tau \Rightarrow B(\theta_1) > B(\theta_2)$, so

$$r(\mathbf{x}_n) \equiv \frac{b_{\theta_2,n}}{b_{\theta_1,n}} \left(\frac{A(\theta_1)}{A(\theta_2)}\right)^{x_0} \left(\frac{B(\theta_1)}{B(\theta_2)}\right)^{x_1+\cdots+x_{n-1}} \left(\frac{\theta_2}{\theta_1}\right)^{x_n} \tag{54}$$

is strictly increasing in $x_1, \ldots, x_{n-1}, x_n$, which establishes (b).

(ii) The first inequality is immediate. For the second, apply the FKG inequality as follows:

$$\begin{aligned} \mathrm{E}_\tau[g(\ddot{\mathbf{X}}_n)] &= \mathrm{E}_\tau\left[g(\dot{\mathbf{X}}_n^+)\frac{\ddot{f}_\tau(\dot{\mathbf{X}}_n^+)}{\dot{f}_\tau^+(\dot{\mathbf{X}}_n^+)}\right] \\ &\ge \mathrm{E}_\tau[g(\dot{\mathbf{X}}_n^+)]\,\mathrm{E}_\tau\left[\frac{\ddot{f}_\tau(\dot{\mathbf{X}}_n^+)}{\dot{f}_\tau^+(\dot{\mathbf{X}}_n^+)}\right] \\ &= \mathrm{E}_\tau[g(\dot{\mathbf{X}}_n^+)]. \end{aligned} \tag{55}$$

As in (i), FKG is applicable in (55) because (a) $\dot{f}_\tau^+(\mathbf{x}_n) \equiv f_\tau^+(\mathbf{x}_n)$ is FKG on $\mathbb{Z}_+^n$ (by (53) with $\theta = \tau$); and (b) the ratio

$$r(\mathbf{x}_n) \equiv \frac{\ddot{f}_\tau(\mathbf{x}_n)}{\dot{f}_\tau^+(\mathbf{x}_n)} = \frac{x_n}{b_{\tau,n}x_0}, \tag{56}$$

(obtained from (47) and (52) with $\theta = \tau$) is increasing on $\mathbb{Z}_+^n \cap \{h_{\mathbf{a}}(\mathbf{x}_n) > 0\}$.

To show that $\mathrm{E}_\tau[g(\ddot{\mathbf{X}}_n)] > \mathrm{E}_\tau[g(\dot{\mathbf{X}}_n^+)]$ for at least one increasing g, take $g(\mathbf{x}_n) = 1_{\{2,3,\ldots\}}(x_1)$. Because $\ddot{X}_{\tau,1} \ge 1$ and $\dot{X}_{\tau,1}^+ \ge 1$, it follows from (47) and (52) that

$$\begin{aligned} 1 - \mathrm{E}_\tau[g(\ddot{\mathbf{X}}_n)] &= \Pr_\tau[\ddot{X}_1 = 1] \\ &= \frac{\tau}{x_0(A(\tau))^{x_0}} h_{\mathbf{a}}(x_0, 1); \end{aligned}$$

$$\begin{aligned} 1 - \mathrm{E}_\tau[g(\dot{\mathbf{X}}_n^+)] &= \Pr_\tau[\dot{X}_1^+ = 1] \\ &= \frac{\dot{b}_{1,\tau}\,\tau}{(A(\tau))^{x_0}} h_{\mathbf{a}}(x_0, 1) \\ &= \frac{\tau}{\{1 - \Pr_\tau[\dot{X}_1 = 0]\}(A(\tau))^{x_0}} h_{\mathbf{a}}(x_0, 1) \\ &= \frac{\tau}{\left\{1 - \left(\frac{a_0}{A(\tau)}\right)^{x_0}\right\}(A(\tau))^{x_0}} h_{\mathbf{a}}(x_0, 1). \end{aligned}$$

Because $a_0 > 0$ and $x_0 \geq 1$, we conclude that $\mathrm{E}_\tau[g(\ddot{\mathbf{X}}_n)] > \mathrm{E}_\tau[g(\dot{\mathbf{X}}_n^+)]$.

(iii) Since $B(\theta_1) < B(\theta_2)$ when $\tau \leq \theta_1 < \theta_2$, the FKG inequality is not applicable here (recall (54)). Instead we use induction on n to show that

$$\mathrm{E}_\theta[g(\ddot{\mathbf{X}}_n)] \equiv \sum\nolimits_{\mathbf{x}_n > 0} g(\mathbf{x}_n) \ddot{f}_\theta(\mathbf{x}_n) \tag{57}$$

is increasing for $\theta \in [\tau, \psi)$.

For $n = 1$, (42) gives

$$\ddot{f}_\theta(x_1) = \Big(\frac{1 - u_\theta^{x_1}}{1 - u_\theta^{x_0}}\Big) \frac{\theta^{x_1}}{(A(\theta))^{x_0}} h_{\mathbf{a}}(x_0, x_1), \quad x_1 > 0,$$

which is TP2 in (θ, x_1) by TP2b, so

$$\mathrm{E}_\theta[g(\ddot{X}_1)] \equiv \sum\nolimits_{x_1 > 0} g(x_1) f_\theta(x_1) \tag{58}$$

is increasing for $\theta \in (\tau, \psi)$ by the monotonicity-preserving property of a TP2 $\equiv$ MLR kernel (Karlin (1968, Ch.1, Proposition 3.1)). For $n \geq 2$,

$$\begin{aligned} \mathrm{E}_\theta[g(\ddot{\mathbf{X}}_n)] &= \mathrm{E}_\theta[\mathrm{E}_\theta[g(\ddot{\mathbf{X}}_{n-1}, \ddot{X}_n) \mid \ddot{\mathbf{X}}_{n-1}]] && (59)\\ &= \mathrm{E}_\theta\Big[\sum\nolimits_{x_n > 0} g(\ddot{\mathbf{X}}_{n-1}, x_n) \ddot{f}_\theta(x_n \mid \ddot{X}_{n-1})\Big] && (60)\\ &\equiv \mathrm{E}_\theta\,[g_\theta^*(\ddot{\mathbf{X}}_{n-1})]. && (61) \end{aligned}$$

From TP2a, TP2b, and (45), the transition probability $\ddot{f}_\theta(x_n \mid x_{n-1})$ of the Markov process $\ddot{\mathbf{X}}_\theta$ is TP2 in (x_n, θ) and in (x_n, x_{n-1}), so the monotonicity-preserving property implies that $g_\theta^*(\ddot{\mathbf{X}}_{n-1})$ is increasing in θ and in $\ddot{\mathbf{X}}_{n-1}$. Thus by (60)-(61) and the induction hypothesis, $\mathrm{E}_\theta[g(\ddot{\mathbf{X}}_n)]$ is increasing for θ for $\theta \in (\tau, \psi)$. Lastly, these results extend to $[\tau, \psi)$ by (49) and continuity.

To show that $\mathrm{E}_\theta[g(\ddot{\mathbf{X}}_n)]$ is strictly increasing in θ for at least one increasing g, take $g(\mathbf{x}_n) = 1_{\{2,3,\ldots\}}(x_1)$. Because $\ddot{X}_1 \geq 1$, it follows from (42) that

$$\begin{aligned} 1 - \mathrm{E}_\theta[g(\ddot{\mathbf{X}}_n)] &= \mathrm{Pr}_\theta[\ddot{X}_1 = 1]\\ &= \Big(\frac{1 - u_\theta}{1 - u_\theta^{x_0}}\Big) \frac{\theta}{(A(\theta))^{x_0}} h_{\mathbf{a}}(x_0, 1)\\ &= \left[\frac{(1 - u_\theta)\theta}{(1 - u_\theta^{x_0})\theta^{x_0}}\right] \left[\frac{1}{B(\theta)}\right]^{x_0} h_{\mathbf{a}}(x_0, 1). && (62) \end{aligned}$$

Because $x_0 \geq 1$ the first factor in (62) is decreasing in θ by TP2b, while the second factor is strictly decreasing because $B(\theta)$ is strictly increasing for $\theta \in [\tau, \psi)$. □

Lemma 4.3. *Let $\mathbf{X}_\theta$ be a GW branching process with psod offspring distribution* (16)*. Each of the following two conditions is equivalent to Condition TP2b: for $\theta \in (\tau, \psi)$,*

$$\frac{\mu_\theta - 1}{1 - \mu_{\theta u_\theta}} \leq \frac{1}{u_\theta}; \tag{63}$$

$$B'(\theta u_\theta) + B'(\theta) \leq 0. \tag{64}$$

Proof. Let $\delta_\theta = du_\theta/d\theta$. Then $(1-u_\theta^x)\theta^x$ is TP2 iff for $x = 1, 2, \ldots$, the ratio $\frac{(1-u_\theta^{x+1})\theta^{x+1}}{(1-u_\theta^x)\theta^x}$ is increasing in θ for $\theta \in (\tau, \psi)$, equivalently, iff

$$\frac{d}{d\theta} \log \left[\frac{(1-u_\theta^{x+1})\theta}{(1-u_\theta^x)} \right] \equiv \frac{-(x+1)u_\theta^x \delta_\theta}{1-u_\theta^{x+1}} + \frac{x u_\theta^{x-1}\delta_\theta}{1-u_\theta^x} + \frac{1}{\theta} \geq 0. \tag{65}$$

After some algebra, we find that this is equivalent to the inequality

$$[(1-u^x) - xu^x(1-u)] + d_\theta u^{x-1}[x(1-u) - u(1-u^x)] \geq 0, \tag{66}$$

where we use the relation $d_\theta = \theta\delta_\theta + u_\theta$ and abbreviate u_θ by u. Because both terms in square brackets are positive and $d_\theta < 0$, this is in turn equivalent to the inequality

$$-d_\theta \leq \frac{(1-u^x) - xu^x(1-u)}{u^{x-1}[x(1-u) - u(1-u^x)]} \equiv \Delta(u, x). \tag{67}$$

But $\Delta(u, 1) = 1$ and $\Delta(u, x) \geq 1$ for $x \geq 2$:

$$\begin{aligned}
\Delta(u,x) - 1 &= \frac{(1-u^x)(1+u^x) - u^{x-1}x(1-u)(1+u)}{u^{x-1}[x(1-u) - u(1-u^x)]} \\
&= \frac{(1-u^{2x}) - u^{x-1}x(1-u^2)}{u^{x-1}[x(1-u) - u(1-u^x)]} \\
&= \frac{(1-u^2)[(1+u^2+\cdots+u^{2(x-1)}) - u^{2(x-1)/2}x]}{u^{x-1}[x(1-u) - u(1-u^x)]} \\
&\geq 0
\end{aligned}$$

because u^{2x} is convex in x. Thus TP2b is equivalent to the simple relation

$$-d_\theta \leq \Delta(u, 1) \equiv 1, \tag{68}$$

which, by (27), is equivalent to (63). Lastly, differentiate (23) with respect to θ to establish the equivalence of (68) and (64). $\square$

Example 4.1 (= 3.1 continued). For the Poisson(θ) psod, the coefficient of θ^y in the power series $[A(\theta)]^x = e^{x\theta}$ is $h_{\mathbf{a}}(x,y) = x^y/y!$, which is TP2 in (x,y) so TP2a is satisfied. Furthermore $\mu_\theta = \theta$, $\tau = 1$, and from (33),

$$-\frac{\log u_\theta}{1-u_\theta} = \theta \tag{69}$$

for $\theta \geq 1$, so (63) is equivalent to the inequality

$$-2u\log u \leq 1-u^2, \tag{70}$$

where $u = u_\theta$. This inequality holds for all $u \in [0,1]$, hence TP2b is also satisfied. Thus by Proposition 4.2, $(\dot{\mathbf{X}}_\theta)_n^+$ is stochastically decreasing and $(\ddot{\mathbf{X}}_\theta)_n$ is stochastically increasing for $\theta \geq 1$, while $(\dot{\mathbf{X}}_\tau)_n^+ \prec (\ddot{\mathbf{X}}_\tau)_n$. $\square$

Example 4.2 (= 3.2 continued). For the negative binomial(r,θ) psod, the coefficient of θ^y in the power series $[A(\theta)]^x = 1/(1-\theta)^{rx}$ is

$$h_{\mathbf{a}}(x,y) = \frac{\Gamma(rx+y)}{\Gamma(rx)y!} = \frac{(rx+y-1)\cdots(rx)}{y!}, \tag{71}$$

which is TP2 in (x,y), so NB($r,\cdot$) satisfies TP2a for all $r>0$. Next, $\mu_\theta = \frac{r\theta}{(1-\theta)}$ and $\tau = \frac{t}{1+t}$, where $t = \frac{1}{r}$. Set $u = u_\theta$ and apply (35) to obtain

$$\frac{1-\theta}{1-\theta u} = u^t, \tag{72}$$

$$\frac{1-u^t}{1-u^{t+1}} = \theta. \tag{73}$$

After some algebra it is seen that (63) is equivalent to each of the inequalities

$$\frac{1-u^t}{1-u^{t+1}} \leq \tau\left(\frac{1+u^{t-1}}{1+u^t}\right), \tag{74}$$

$$\frac{v^t-v^{-t}}{t} \leq v-v^{-1}, \tag{75}$$

where $v = u^{-1} \geq 1$. Because $h(t) \equiv v^t - v^{-t}$ is convex in t and $h(0)=0$, (75) holds iff $t \leq 1$. Thus the NB($r,\cdot$) psod family satisfies TP2b iff $r \geq 1$. This includes the geometric psod family ($r=t=1$) where equality holds in (75).

Thus by Proposition 4.2, if $r \geq 1$ then $(\dot{\mathbf{X}}_\theta)_n^+$ is stochastically decreasing and $(\ddot{\mathbf{X}}_\theta)_n$ is stochastically increasing for $\tau \leq \theta < 1$, while $(\dot{\mathbf{X}}_\tau)_n^+ \prec (\ddot{\mathbf{X}}_\tau)_n$, where $\tau = 1/(1+r)$. $\square$

Example 4.3 (= 3.3 continued). For the binary splitting GW process, the coefficient of θ^y in the power series $[A(\theta)]^x = (1+\theta^2)^x$ is

$$h_{\mathbf{a}}(x,y) = \begin{cases} \binom{x}{y/2} & \text{for } y = 0, 2, \ldots, 2x, \\ 0 & \text{otherwise,} \end{cases} \tag{76}$$

which is TP2 in (x, y), so TP2a is satisfied. Furthermore $B(\theta) = \theta + \theta^{-1}$ and $u_\theta = \theta^{-2}$ for $\theta \geq \tau = 1$, so (64) is equivalent to the valid inequality $2 - \theta^2 - \theta^{-2} \leq 0$, hence TP2b is satisfied. Thus by Proposition 4.2, $(\dot{\mathbf{X}}_\theta)_n^+$ is stochastically decreasing and $(\ddot{\mathbf{X}}_\theta)_n$ is stochastically increasing for $\theta \geq 1$, while $(\dot{\mathbf{X}}_\tau)_n^+ \prec (\ddot{\mathbf{X}}_\tau)_n$. $\square$

Remark 4.1. The maximum likelihood estimate (MLE) $\hat{\theta}$ is derived by differentiating (40), then applying (18) to obtain the relation

$$\hat{\mu} \equiv \mu_{\hat{\theta}} = \frac{Y_n - x_0}{Y_{n-1}}, \tag{77}$$

from which $\hat{\theta}$ can be obtained. Here $\hat{\mu}$ denotes the MLE of the mean μ_θ. $\square$

5. Predicting extinction or explosion: the fixed sample size case

Based on observed data $\mathbf{x}_n \equiv (x_1, \ldots, x_n)$ from a non-terminated psod GW process $\mathbf{X} \equiv \mathbf{X}_\theta$ with initial size x_0 and fixed n, predict whether extinction or explosion will occur for the current realization of the process.

By the Markov property for $\mathbf{X}$,

$$\Pr_\theta[\,\textit{extinction} \,|\, \mathbf{X}_n = \mathbf{x}_n] = u_\theta^{x_n} = 1 - \Pr_\theta[\,\textit{explosion} \,|\, \mathbf{X}_n = \mathbf{x}_n\,]. \tag{78}$$

The MLE $\hat{u}$ of u_θ is given by $\hat{u} = u_{\hat{\theta}}$ where $\hat{\theta}$ is obtained from (77), so the estimated extinction probability is

$$\Pr_{\hat{\theta}}[\,\textit{extinction} \,|\, \mathbf{X}_n = \mathbf{x}_n] = \hat{u}^{x_n} \begin{cases} = 1 & \text{if } x_n \leq x_0, \\ < 1 & \text{if } x_n > x_0. \end{cases} \tag{79}$$

The value of $\hat{u}^{x_n}$ can be used to predict extinction or explosion.

However, this procedure may reach unwarranted conclusions. For example, if $x_n = x_0 - 1$ then extinction will be predicted with certainty even though the population has declined only slightly. Whereas (79) is based solely on the value of x_n, the prediction procedures derived in this section compare x_n to n in order to predict whether the observed process is on a trajectory toward extinction or toward explosion.

5.1. Prediction as a testing problem. We reformulate the prediction problem as a hypothesis-testing problem to which Neyman-Pearson theory can be applied. As in Proposition 4.2, let $\mathbf{X}_n^+ \equiv (\mathbf{X}_\theta)_n^+$ denote the conditional random vector $\mathbf{X}_n \mid X_n > 0$. Then conditional on non-termination at time n, extinction will occur iff either

$$H_{\leq}^+ : \mathbf{X}_n^+ \stackrel{d}{=} (\mathbf{X}_\theta)_n^+,\ \theta \leq \tau \quad \text{or} \quad \dot{H}_{\geq}^+ : \mathbf{X}_n^+ \stackrel{d}{=} (\dot{\mathbf{X}}_\theta)_n^+,\ \theta \geq \tau$$

holds, while explosion will occur iff $\ddot{H}_{>}^+ : \mathbf{X}_n^+ \stackrel{d}{=} (\ddot{\mathbf{X}}_\theta)_n^+,\ \theta > \tau$ holds. (Recall that $(\ddot{\mathbf{X}}_\theta)^+ = \ddot{\mathbf{X}}_\theta$.) However, $H_{\leq}^+ = \dot{H}_{\geq}^+$ by Proposition 3.2 while by (49) the L_1-closure of $\ddot{H}_{>}^+$ is

$$\ddot{H}_{\geq}^+ : \mathbf{X}_n^+ \stackrel{d}{=} (\ddot{\mathbf{X}}_\theta)_n^+,\ \theta \geq \tau, \tag{80}$$

so the prediction problem can be formulated as the following testing problem:

Based on observing $\mathbf{X}_n^+ = \mathbf{x}_n > 0$, *test*

$$\dot{H}_{\geq}^+ \ (\textit{eventual extinction}) \quad \textit{vs.} \quad \ddot{H}_{\geq}^+ \ (\textit{eventual explosion}). \tag{81}$$

Either $\dot{H}_{\geq}^+$ or $\ddot{H}_{\geq}^+$ may be taken to be the null hypothesis. Note that $\theta \geq \tau$ under both $\dot{H}_{\geq}^+$ and $\ddot{H}_{\geq}^+$. The conditional pmfs of $\mathbf{X}_n^+$ under $\dot{H}_{\geq}^+$ and $\ddot{H}_{\geq}^+$ are

$$\dot{f}_\theta^+(\mathbf{x}_n) \equiv \Pr_\theta[\dot{\mathbf{X}}_n^+ = \mathbf{x}_n] = \dot{b}_{\theta,n} \dot{f}_\theta(\mathbf{x}_n), \quad \mathbf{x}_n > 0, \tag{82}$$

and $\ddot{f}_\theta^+(\mathbf{x}_n) \equiv \ddot{f}_\theta(\mathbf{x}_n)$, respectively, where $\ddot{f}_\theta$ is given by (42) and (47) and

$$\dot{b}_{\theta,n}^{-1} = \Pr_\theta[\dot{X}_n > 0] = \Pr_{\theta u_\theta}[X_n > 0] = b_{\theta u_\theta,n}^{-1}. \tag{83}$$

A version of the generalized LR criterion for (81) is

$$\lambda^+(\mathbf{x}_n) \equiv \frac{\sup_{\tau \leq \theta < \psi} \ddot{f}_\theta^+(\mathbf{x}_n)}{\sup_{\tau \leq \theta < \psi} \dot{f}_\theta^+(\mathbf{x}_n)}, \tag{84}$$

but the numerator and denominator may be difficult to evaluate.

5.2. The least favorable distributions for fixed sample size. When the psod satisfies Conditions TP2a and TP2b the testing problem (81) has a convenient solution. Proposition 4.2 implies that $\dot{H}^+_{\geq}$ and $\ddot{H}_{\geq}$ are separated families and that $(\dot{f}^+_\tau, \ddot{f}^+_\tau)$ is a pair of least favorable distributions for (81). By Theorem 3.8.1 of Lehmann and Romano (2005), a test of the form

$$\begin{cases} \textit{accept } \dot{H}^+_{\geq} \ (\textit{predict extinction}) & \text{if } \lambda^+_\tau(\mathbf{X}^+_n) \leq d, \\ \textit{accept } \ddot{H}_{\geq} \ \ (\textit{predict explosion}) & \text{if } \lambda^+_\tau(\mathbf{X}^+_n) > d, \end{cases} \tag{85}$$

is the UMP test of its size for (81), where d is a nonnegative constant and, from (47) and (82),

$$\lambda^+_\tau(\mathbf{x}_n) \equiv \frac{\ddot{f}^+_\tau(\mathbf{x}_n)}{\dot{f}^+_\tau(\mathbf{x}_n)} = \frac{x_n}{x_0\dot{b}_{\tau,n}} = \frac{x_n}{x_0 b_{\tau,n}}, \quad \mathbf{x}_n > 0. \tag{86}$$

Because $\lambda^+_\tau(\mathbf{x}_n)$ is strictly increasing in x_n, the test (85) has the form

$$\begin{cases} \textit{accept } \dot{H}^+_{\geq} \ (\textit{predict extinction}) & \text{if } X^+_n \leq c, \\ \textit{accept } \ddot{H}^+_{\geq} \ \ (\textit{predict explosion}) & \text{if } X^+_n \geq c+1, \end{cases} \tag{87}$$

where c is a nonnegative integer.

5.3. Exponential-type approximations for $\dot{X}^+_n$ when $\theta = \tau$. Suppose first that $\dot{H}^+_{\geq}$ is taken to be the null hypothesis. If $X^+_n = x_n > 0$ is observed, the p-value $\Pr_\tau[\dot{X}^+_n \geq x_n]$ for test (87) is determined by the distribution of $\dot{X}^+_n$ under $\dot{f}^+_\tau$. For large n the mean and variance of $\dot{X}^+_n$ can be approximated via (93) as follows:

$$\mathrm{E}_\tau(\dot{X}^+_n) \ = \ x_0 b_{\tau,n} \approx \frac{n\sigma^2_\tau}{2}\,, \tag{88}$$

$$\mathrm{Var}_\tau(\dot{X}^+_n) \ = \ x_0 b_{\tau,n}[n\sigma^2_\tau - x_0(b_{\tau,n}-1)] \approx \frac{n\sigma^2_\tau}{2}\Big(\frac{n\sigma^2_\tau}{2} + x_0\Big)\,. \tag{89}$$

Unfortunately the conditional rv $\dot{X}^+_n \,(\equiv \dot{X}_n \mid X_n > 0)$ is not the sum of x_0 i.i.d. copies each with initial size 1: conditional on $X_n > 0$, some of the initial x_0 family lines may have terminated by time n. Therefore a normal approximation is not available for $\dot{X}^+_n$, even when x_0 is large.

Fortunately, however, when n is large Yaglom's classical exponential approximation can be applied. For a critical GW process (not necessarily psod) with offspring variance $\sigma^2 < \infty$, Yaglom (1947) showed that if $x_0 = 1$ then

$$\lim_{n\to\infty} \Pr[X_n^+ \geq nz] = e^{-2z/\sigma^2} \tag{90}$$

for any $z > 0$. This result appears under progressively weaker moment conditions in Harris (1963, §I.10), Kesten, Ney, and Spitzer (1966, p.582), Athreya and Ney (1972, §9), and Jagers (1975, Theorem 2.4.2), but only for the case $x_0 = 1$. When $x_0 \geq 2$ it might be expected that the limiting exponential (EXP) distribution in (90) should be replaced by the distribution of the sum of x_0 independent exponential rvs, i.e., a gamma distribution, but (90) continues to hold without change, cf. (92). However, we will also present a more accurate gamma (GAM) approximation (94) that does depend on x_0.

Let G_r denote a gamma rv with shape parameter $r > 0$ and scale parameter 1 and let $G_r(z)$ denote its cdf, that is,

$$G_r(z) = \frac{1}{\Gamma(r)} \int_0^z t^{r-1} e^{-t} dt. \tag{91}$$

Proposition 5.1. (i) *Let $\{X_n\}$ be a critical GW process with offspring pgf ϕ, offspring variance $\sigma^2 < \infty$, and initial size $x_0 \geq 1$. For any $z > 0$,*

$$\lim_{n\to\infty} \Pr[X_n^+ \geq nz] = e^{-2z/\sigma^2}, \tag{92}$$

$$\lim_{n\to\infty} n\Pr[X_n > 0] = 2x_0/\sigma^2. \tag{93}$$

(ii) Let $\bar{G}_r(z) = 1 - G_r(z)$. For $x_0 \geq 1$ and large n,

$$\Pr[X_n^+ \geq nz] \approx \frac{1}{1 - \left(1 - \frac{2}{n\sigma^2}\right)^{x_0}} \sum_{r=1}^{x_0} \tbinom{x_0}{r} \bar{G}_r(\tfrac{2z}{\sigma^2})(\tfrac{2}{n\sigma^2})^r (1 - \tfrac{2}{n\sigma^2})^{x_0 - r}. \tag{94}$$

Proof. (i) The existing results for the case $x_0 = 1$ are based on the following fact, cf. Jagers (1975, Lemma 2.4.1):

$$\lim_{n\to\infty} \frac{1}{n} \left[\frac{1}{1 - \phi_n(s)} - \frac{1}{1 - s} \right] = \frac{\sigma^2}{2} \quad \text{uniformly for } 0 \leq s < 1. \tag{95}$$

Set $s = 0$ to obtain

$$\lim_{n\to\infty} n(1 - \phi_n(0)) = \frac{2}{\sigma^2}. \tag{96}$$

For $x_0 \geq 2$, X_n has pgf $\phi_n^{x_0}$ so by (96),

$$
\begin{aligned}
&(97) \qquad & n\Pr[X_n > 0] &= n(1-\phi_n^{x_0}(0)) \\
&(98) & &= n(1-\phi_n(0))[1+\phi_n(0)+\cdots+\phi_n^{x_0-1}(0)] \\
&(99) & &\to \frac{2x_0}{\sigma^2}
\end{aligned}
$$

because $\phi_n(0) \uparrow 1$; this confirms (93). Furthermore, the Laplace transform of X_n^+/n is, for $t \geq 0$,

$$
\begin{aligned}
& \mathrm{E}[e^{-tX_n^+/n}] \\
= \;& \frac{\phi_n^{x_0}(e^{-t/n}) - \phi_n^{x_0}(0)}{1-\phi_n^{x_0}(0)} \\
= \;& 1 - \frac{1-\phi_n^{x_0}(e^{-t/n})}{1-\phi_n^{x_0}(0)} \\
= \;& 1 - \frac{n(1-\phi_n(e^{-t/n}))[1+\phi_n(e^{-t/n})+\cdots+\phi_n^{x_0-1}(e^{-t/n})]}{n(1-\phi_n(0))[1+\phi_n(0)+\cdots+\phi_n^{x_0-1}(0)]} \\
\to \;& 1 - \frac{\frac{\sigma^2}{2}}{\frac{1}{t}+\frac{\sigma^2}{2}}\frac{x_0}{x_0} = \frac{1}{1+\frac{t\sigma^2}{2}} \quad \text{as } n \to \infty
\end{aligned}
$$

by (95) and the inequalities $\phi_n(0) < \phi_n(e^{-t/n}) < 1$. This is the Laplace transform of the exponential distribution in (92), confirming that result.

(ii) Represent $X_n \stackrel{d}{=} U_1 + \cdots + U_{x_0}$, where the U_i are i.i.d. copies of X_n but each with initial size $x_0 = 1$. Then

$$
\begin{aligned}
& \Pr[X_n > 0] \Pr[X_n^+ \geq nz] \\
& \equiv \Pr[U_1 + \cdots U_{x_0} > 0] \Pr\big[\big(U_1+\cdots+U_{x_0}\big)^+ \geq nz\big] \\
& = \Pr\big[U_1+\cdots+U_{x_0} \geq nz\big] \\
& = \sum\nolimits_{\omega \in 2^{x_0}\setminus\emptyset} \Pr\big[\sum\nolimits_{i\in\omega} U_i \geq nz,\ U_i > 0 \text{ for } i \in \omega,\ U_i = 0 \text{ for } i \notin \omega\big] \\
& = \sum_{r=1}^{x_0} \tbinom{x_0}{r} \Pr[U_1+\cdots+U_r \geq nz,\ U_1 > 0, \ldots, U_r > 0, U_{r+1} = \cdots = U_{x_0} = 0] \\
& = \sum_{r=1}^{x_0} \tbinom{x_0}{r} \Pr[U_1^+ + \cdots + U_r^+ \geq nz] \Pr[U_1 > 0]^r \Pr[U_{x_0} = 0]^{x_0-r} \\
& \approx \sum_{r=1}^{x_0} \tbinom{x_0}{r} \bar{G}_r(\tfrac{2z}{\sigma^2})(\tfrac{2}{n\sigma^2})^r(1-\tfrac{2}{n\sigma^2})^{x_0-r}
\end{aligned}
$$

for large n, by (92) and by (93) with $x_0 = 1$. Furthermore, by (97) and (96),

$$\Pr[X_n > 0] = (1 - \phi_n^{x_0}(0)) \approx 1 - \big(1 - \tfrac{2}{n\sigma^2}\big)^{x_0}, \tag{100}$$

which yields (94). □

From (92) and (94) and a continuity correction we obtain exponential-type approximations for the p-value of the test (87) when $\dot{H}^+_{\geq}$ and $\ddot{H}^+_{\geq}$ are taken to be the null and alternative hypothesis, respectively:

$$\begin{aligned}\Pr_\tau[\dot{X}^+_n \geq x_n] &\approx e^{-\frac{2(x_n-1)}{n\sigma_\tau^2}} \\ &\equiv \dot{\pi}^{\mathrm{EXP}}(x_n; n), \qquad (101)\\ \Pr_\tau[\dot{X}^+_n \geq x_n] &\approx \frac{1}{1-\big(1-\frac{2}{n\sigma_\tau^2}\big)^{x_0}} \sum_{r=1}^{x_0} \tbinom{x_0}{r}\big(\tfrac{2}{n\sigma_\tau^2}\big)^r\big(1-\tfrac{2}{n\sigma_\tau^2}\big)^{x_0-r}\bar{G}_r\big(\tfrac{2(x_n-1)}{n\sigma_\tau^2}\big) \\ &\equiv \dot{\pi}^{\mathrm{GAM}}(x_n; n, x_0). \qquad (102)\end{aligned}$$

The EXP and GAM approximations coincide when $x_0 = 1$. The approximate p-value $\dot{\pi}^{\mathrm{EXP}}(x_n; n)$ does not depend on the value of x_0; it conveys significance for $\ddot{H}^+_{\geq}$ (explosion) iff $x_n \gg n\sigma_\tau^2$, but convergence to the exact p-value is slow, see Remark 5.1. The $\dot{\pi}^{\mathrm{GAM}}(x_n; n)$ approximation is noticeably better.

Remark 5.1. The accuracy of the EXP and GAM approximations can be assessed for the geometric psod (cf. Example 3.2). Here the pgf of $X_n \equiv \dot{X}_n$ in the critical case can be obtained explicitly[6] and expanded in a power series, from which the exact distribution of X_n can be recovered. By (82) and (83) this yields the exact distribution of $X^+_n \equiv \dot{X}^+_n$ in the critical case. The exact p-values and their approximations $\dot{\pi}^{\mathrm{EXP}}$ and $\dot{\pi}^{\mathrm{GAM}}$ are shown in Tables 1 and 2, from which the superiority of GAM is apparent. □

5.4. Exponential-type approximations for $\ddot{X}^+_n$ when $\theta = \tau$. Suppose next that $\ddot{H}^+_{\geq}$ is the null hypothesis. The p-value $\Pr_\tau[\ddot{X}^+_n \leq x_n]$ for test (87) is determined by the distribution of $\ddot{X}^+_n \equiv \ddot{X}_n$ under $\ddot{f}^+_\tau \equiv \ddot{f}_\tau$. Again a normal approximation is not available for large x_0 because $\ddot{X}_n$ is not the sum of x_0 i.i.d. copies each with initial size 1: conditional on explosion, some of the initial x_0 family lines nonetheless may become extinct. However, an exponential-type approximation is available for large n, based on the following representation for the process $\ddot{\mathbf{X}}_\tau \equiv \{\ddot{X}_{\tau,n} | n \geq 1\}$. We shall abbreviate $\ddot{X}_{\tau,n}$ to $\ddot{X}_n$.

[6]cf. eqn.(8.32) in Taylor and Karlin (1998) for the case $x_0 = 1$.

			$x_0 = 1$	$x_0 = 2$	
n	x_n	$\dot{\pi}^{\mathrm{EXP}}$	Exact	$\dot{\pi}^{\mathrm{GAM}}$	Exact
5	10	0.165	0.194	0.198	0.226
5	20	0.022	0.031	0.032	0.042
5	30	0.003	0.005	0.005	0.008
5	40	0.000	0.001	0.001	0.001
10	20	0.150	0.164	0.165	0.178
10	40	0.020	0.024	0.024	0.029
10	60	0.003	0.004	0.004	0.005
15	20	0.282	0.293	0.294	0.305
15	40	0.074	0.081	0.081	0.087
15	60	0.020	0.022	0.022	0.025
100	150	0.225	0.227	0.227	0.229

Table 1: Exact and approximate p-values $\Pr_\tau[\dot{X}_n^+ \geq x_n]$ for the geometric psod when $x_0 = 1$ and $x_0 = 2$.

			$x_0 = 8$		$x_0 = 14$	
n	x_n	$\dot{\pi}^{\mathrm{EXP}}$	$\dot{\pi}^{\mathrm{GAM}}$	Exact	$\dot{\pi}^{\mathrm{GAM}}$	Exact
5	10	0.165	0.420	0.428	0.631	0.618
5	20	0.022	0.129	0.141	0.286	0.288
5	30	0.003	0.035	0.042	0.108	0.115
5	40	0.000	0.009	0.011	0.036	0.041
10	20	0.150	0.262	0.273	0.369	0.375
10	40	0.020	0.058	0.064	0.108	0.115
10	60	0.003	0.012	0.014	0.029	0.032
10	80	0.000	0.002	0.003	0.007	0.008
15	20	0.282	0.369	0.378	0.444	0.450
15	40	0.074	0.126	0.133	0.178	0.184
15	60	0.020	0.042	0.046	0.069	0.073
15	80	0.005	0.014	0.015	0.026	0.028
100	150	0.225	0.237	0.239	0.248	0.249

Table 2: Exact and approximate p-values $\Pr_\tau[\dot{X}_n^+ \geq x_n]$ for the geometric psod when $x_0 = 8$ and $x_0 = 14$.

Proposition 5.2 *Define* $Z_n = \ddot{X}_n - 1$, $n = 1, 2, \ldots$, $z_0 = x_0 - 1$. *When* $\theta = \tau$, $\mathbf{Z} \equiv \{Z_n | n \geq 1\}$ *is a critical GW process with immigration (GWI). Specifically,*

$$Z_n \mid Z_{n-1} = \xi_1^{(n)} + \cdots + \xi_{Z_{n-1}}^{(n)} + W_n, \tag{103}$$

where $\xi_1^{(n)}, \ldots, \xi_{Z_{n-1}}^{(n)}, W_n$ *are independent rvs,* $\xi_j^{(n)} \stackrel{d}{=} \xi_\tau$, *and* W_n *is a nonnegative integer-valued rv with pgf* $\phi'_\tau(s)$. *(This is a pgf since* $\phi'_\tau(1) = \mu_\tau = 1$.*)*

Proof. From (47),

$$\begin{aligned} \ddot{f}_\tau(\mathbf{x}_n) &= \prod_{i=1}^n \frac{x_i}{x_{i-1}} \frac{\tau^{x_i}}{(A(\tau))^{x_{i-1}}} h_{\mathbf{a}}(x_{i-1}, x_i) \\ &\equiv \prod_{i=1}^n g_\tau(x_i | x_{i-1}), \end{aligned} \tag{104}$$

so $\ddot{\mathbf{X}}_\tau$ is a Markov chain with transition probability $g_\tau(x_i|x_{i-1})$. The conditional pgf corresponding to $g_\tau(x_i|x_{i-1})$ is

$$\begin{aligned} \mathrm{E}_\tau(s^{\ddot{X}_i} \mid \ddot{X}_{i-1} = \ddot{x}_{i-1}) &= \frac{1}{x_{i-1}} \sum_{x_i} x_i s^{x_i} \frac{\tau^{x_i}}{(A(\tau))^{x_{i-1}}} h_{\mathbf{a}}(x_{i-1}, x_i) \\ &= \frac{s}{x_{i-1}} \frac{d}{ds} \sum_{x_i} s^{x_i} \frac{\tau^{x_i}}{(A(\tau))^{x_{i-1}}} h_{\mathbf{a}}(x_{i-1}, x_i) \\ &= \frac{s}{x_{i-1}} \frac{d}{ds} \left[(\phi_\tau(s))^{x_{i-1}}\right] \\ &= s(\phi_\tau(s))^{x_{i-1}-1} \phi'_\tau(s). \end{aligned} \tag{105}$$

The third equality holds since $\frac{\tau^{x_i}}{(A(\tau))^{x_{i-1}}} h_{\mathbf{a}}(x_{i-1}, x_i)$ is the pmf of $\xi_1^{(i)} + \cdots + \xi_{x_{i-1}}^{(i)}$. Thus (105) implies that

$$\ddot{X}_i \mid \ddot{X}_{i-1} = 1 + \xi_1^{(i)} + \cdots + \xi_{X_{i-1}-1}^{(i)} + W_i, \tag{106}$$

where the $\xi_j^{(i)}$'s and W_i are mutually independent rvs, the $\xi_j^{(i)}$'s have common pgf ϕ_τ, and W_i is the nonnegative integer-valued rv with pgf $\phi'_\tau(s)$. Now set $i = n$ in (106) to obtain (103). $\square$

By the theorem of Seneta (1970), $2Z_n/n\sigma_\tau^2 \xrightarrow{d} G_2$ (cf. (91)) if $z_0 = 1$ ($x_0 = 2$). Since $\ddot{X}_n = Z_n + 1$, we obtain the following approximation when $x_0 = 2$:

$$\Pr_\tau[\ddot{X}_n \le x_n] \approx G_2\left(\frac{2x_n}{n\sigma_\tau^2}\right) \equiv \ddot{\pi}^{\mathrm{G2}}(x_n; n) \quad \text{for large } n. \tag{107}$$

We now show that if n is sufficiently large, $\ddot{\pi}^{\mathrm{G2}}(x_n; n)$ remains a valid approximation for $\Pr_\tau[\ddot{X}_n \le x_n]$ for all $x_0 \ge 2$. In the process we derive a sharper approximation $\ddot{\pi}^{\mathrm{G23}}(x_n; n, x_0) \le \ddot{\pi}^{\mathrm{G2}}(x_n; n)$ that depends on x_0 as well as n. The case $x_0 = 1$ is treated separately.

Proposition 5.3. *As in Proposition 5.2 let $Z_n = \ddot{X}_n - 1$, $z_0 = x_0 - 1$, $\theta = \tau$.*

(i) Assume that $x_0 \ge 2$, so $z_0 \ge 1$. Then if n is large and $z > 0$,

$$\Pr_\tau[Z_n \le nz] \approx \left(1 - \frac{2(z_0 - 1)}{n\sigma_\tau^2}\right) G_2\left(\frac{2z}{\sigma_\tau^2}\right) + \frac{2(z_0 - 1)}{n\sigma_\tau^2} G_3\left(\frac{2z}{\sigma_\tau^2}\right), \tag{108}$$

so

$$\begin{aligned} &\Pr_\tau[\ddot{X}_n \le x_n] \\ &\approx \left(1 - \frac{2(x_0 - 2)}{n\sigma_\tau^2}\right) G_2\left(\frac{2x_n}{n\sigma_\tau^2}\right) + \frac{2(x_0 - 2)}{n\sigma_\tau^2} G_3\left(\frac{2x_n}{n\sigma_\tau^2}\right) \\ &\equiv \ddot{\pi}^{\mathrm{G23}}(x_n; n, x_0). \end{aligned} \tag{109}$$

This reduces to (107) *if $x_0 = 2$ or $n\sigma_\tau^2 \gg 2(x_0 - 2)$.*

(ii) *Assume that $x_0 = 1$, so $z_0 = 0$, and define*

$$K = \min\{k | \ddot{X}_k \ge 2\} = \min\{k | Z_k \ge 1\}. \tag{110}$$

Then if $n - K$ is large,

$$\begin{aligned} &\Pr_\tau[Z_n \le (n-K)z \mid K, Z_K] \\ &\approx \left(1 - \frac{2(Z_K - 1)}{(n-K)\sigma_\tau^2}\right) G_2\left(\frac{2z}{\sigma_\tau^2}\right) + \frac{2(Z_K - 1)}{(n-K)\sigma_\tau^2} G_3\left(\frac{2z}{\sigma_\tau^2}\right), \end{aligned} \tag{111}$$

so the conditional p-value given K and $\ddot{X}_K$ can be approximated as follows:

$$\begin{aligned} &\Pr_\tau[\ddot{X}_n \le x_n | K, \ddot{X}_K] \\ &\approx \left(1 - \frac{2(\ddot{X}_K - 2)}{(n-K)\sigma_\tau^2}\right) G_2\left(\frac{2x_n}{(n-K)\sigma_\tau^2}\right) + \frac{2(\ddot{X}_K - 2)}{(n-K)\sigma_\tau^2} G_3\left(\frac{2x_n}{(n-K)\sigma_\tau^2}\right) \\ &\equiv \ddot{\pi}^{\mathrm{G23}}(x_n; n-K, \ddot{X}_K). \end{aligned} \tag{112}$$

Proof. (i) First assume that $x_0 \geq 3$, so $z_0 \geq 2$. Rewrite (103) as follows. For $n = 1$,

$$Z_1 \mid z_0 = (\xi_1^{(1)} + W_1) + (\bar{\xi}_1^{(1)} \cdots + \bar{\xi}_{z_0-1}^{(1)}) \equiv U_1 + V_1, \tag{113}$$

where the ξ's and $\bar{\xi}$'s are i.i.d. copies of ξ_τ. For $n \geq 2$,

$$Z_n \mid Z_{n-1} = (\xi_1^{(n)} + \cdots + \xi_{U_{n-1}}^{(n)} + W_n) + (\bar{\xi}_1^{(n)} \cdots + \bar{\xi}_{V_{n-1}}^{(n)}) \equiv U_n + V_n. \tag{114}$$

(If $V_{n-1} = 0$, $V_n = 0$.) Then $\{U_n\}$ is a critical GWI process with immigration rvs $\{W_n\}$ and initial size $u_0 = 1$, $\{V_n\}$ is a critical GW process with initial size $v_0 = z_0 - 1 = x_0 - 2 \geq 0$, and $\{U_n\}$ is independent of $\{V_n\}$. Therefore

$$\begin{aligned} &\Pr_\tau[2Z_n \leq n\sigma_\tau^2 z] \\ &\quad = \Pr_\tau[2U_n \leq n\sigma_\tau^2 z] \Pr_\tau[V_n = 0] + \Pr_\tau[2(U_n + V_n^+) \leq n\sigma_\tau^2 z] \Pr_\tau[V_n > 0]. \end{aligned} \tag{115}$$

Since $u_0 = 1$, Seneta's result applies to give $2U_n/n\sigma_\tau^2 \xrightarrow{d} G_2$, while by (92) $2V_n^+/n\sigma_\tau^2 \xrightarrow{d} G_1$. Because U_n and V_n are independent,

$$\Pr_\tau[2(U_n + V_n^+) \leq n\sigma_\tau^2 z] \to G_3(z) \quad \text{as } n \to \infty. \tag{116}$$

Furthermore, $n\Pr_\tau[V_n > 0] \to 2(z_0 - 1)/\sigma_\tau^2$ by (93), so (115) yields (108), which, applying the continuity correction, yields (109) since $Z_n = \ddot{X}_n - 1$.

If $x_0 = 2$ so $z_0 = 1$, then all $V_n = 0$ and (108) reduces to Seneta's result for $\{U_n\}$.

(ii) The case $x_0 = 1$ differs because when $z_0 = 0$ the first nonzero value for the GWI process $\{Z_n\}$ is $Z_K = W_K$ and occurs when $n = K$. By conditioning on K and Z_K or $\ddot{X}_K$, however, (111) and (112) follow directly from (108) and (109) by replacing z_0 by Z_K, x_0 by $\ddot{X}_K$, and n by $n - K$. $\square$

Like $\dot{\pi}^{\mathrm{EXP}}(x_n; n)$, the approximate p-value $\ddot{\pi}^{\mathrm{G2}}(x_n; n)$ does not depend on x_0 (≥ 2); it conveys significance for $\dot{H}_{\geq}^{+}$ (eventual extinction) iff $x_n \ll n\sigma_\tau^2$. We expect that $\ddot{\pi}^{\mathrm{G2}}$, like $\dot{\pi}^{\mathrm{EXP}}$, will converge only slowly to the exact p-value, but that $\ddot{\pi}^{\mathrm{G23}}$ will perform noticeably better. Note that $\ddot{\pi}^{\mathrm{G23}}$ requires $n\sigma_\tau^2 > 2(x_0 - 2)$; otherwise the weight assigned to G_2 in (109) is negative.

Remark 5.2. The accuracy of the G2 and G23 approximations can be assessed for the geometric psod. The pmf of $\ddot{X}_n$ in the critical case can be obtained from (47) as follows: for $x_n > 0$,

$$\Pr_\tau[\ddot{X}_n = x_n] = \frac{x_n}{x_0} \Pr_\tau[X_n = x_n], \tag{117}$$

and $\Pr_\tau[X_n = x_n]$ can be obtained explicitly as in Remark 5.1. Exact p-values and the G2 and G23 approximations are shown in Table 3, from which the superiority of G23 is apparent. □

			$x_0 = 2$	$x_0 = 8$		$x_0 = 14$	
n	x_n	$\ddot{\pi}^{G2}$	Exact	$\ddot{\pi}^{G23}$	Exact	$\ddot{\pi}^{G23}$	Exact
5	2	0.062	0.062		0.022		0.008
5	4	0.191	0.169		0.069		0.028
5	6	0.337	0.291		0.134		0.060
5	8	0.475	0.411		0.209		0.102
5	10	0.594	0.520		0.290		0.153
10	4	0.062	0.063	0.029	0.038		0.022
10	8	0.191	0.180	0.105	0.115		0.073
10	12	0.337	0.313	0.207	0.213		0.143
10	16	0.475	0.441	0.320	0.317		0.224
10	20	0.594	0.555	0.432	0.418		0.310
20	8	0.062	0.063	0.045	0.048	0.029	0.037
20	12	0.122	0.120	0.092	0.094	0.063	0.074
20	16	0.191	0.186	0.148	0.148	0.105	0.118
20	20	0.264	0.255	0.209	0.207	0.154	0.168
20	24	0.337	0.325	0.272	0.268	0.207	0.220
20	28	0.408	0.393	0.336	0.329	0.263	0.274
50	16	0.041	0.042	0.037	0.038	0.033	0.034
50	24	0.084	0.084	0.076	0.076	0.067	0.069
50	32	0.135	0.134	0.122	0.122	0.109	0.111
50	40	0.191	0.189	0.174	0.173	0.157	0.158
50	48	0.250	0.246	0.228	0.226	0.207	0.208
50	56	0.308	0.304	0.284	0.281	0.259	0.259

Table 3: Exact and approximate p-values $\Pr_\tau[\ddot{X}_n^+ \leq x_n]$ for the geometric psod when $x_0 = 2, 8, 14$.

Remark 5.3. Moments of $\ddot{X}_n$ can be obtained from (106) by recursion, e.g.,

$$\text{(118)} \qquad \mathrm{E}_\tau(\ddot{X}_n) = x_0 + n\sigma_\tau^2,$$

$$\text{(119)} \qquad \mathrm{Var}_\tau(\ddot{X}_n) = n\left[\omega_\tau + \left(\frac{n-3}{2}\right)\sigma_\tau^4 + (x_0 - 3)\sigma_\tau^2 - 1\right],$$

where $\omega_\tau = \mathrm{E}(\xi_\tau^3)$. □

6. Predicting extinction or explosion: sequential sampling

6.1. Sequential probability ratio tests (SPRT). The SPRT (Barnard (1946), Wald (1947), Ghosh (1970), Stuart and Ord (1991)) is well suited for the following sequential version of the prediction problem:

Based on sequential data $\mathbf{x} \equiv (x_0, x_1, x_2, \dots)$ *from a psod GW process* $\mathbf{X} \equiv \mathbf{X}_\theta$ *with initial size* x_0*, predict whether extinction or explosion will occur for the current realization of the process.*

Unlike Section 5, non-termination need not be assumed. This prediction problem can be formulated as the following testing problem:

Based on observing $\mathbf{X}$ *sequentially, test*

$$\begin{aligned} &\dot{H}_{\geq} : \mathbf{X} \overset{d}{=} \dot{\mathbf{X}}_\theta,\ \theta \geq \tau\ (\textit{eventual extinction}) \\ \textit{vs.}\quad &\ddot{H}_{\geq} : \mathbf{X} \overset{d}{=} \ddot{\mathbf{X}}_\theta,\ \theta \geq \tau\ (\textit{eventual explosion}). \end{aligned} \tag{120}$$

For fixed $\theta \geq \tau$, the SPRT for testing $\dot{f}_\theta$ vs. $\ddot{f}_\theta$ has the following form:

The SPRT$(\theta; B, A)$: fix $0 < B < 1 < A < \infty$. For $n = 1, 2, \dots,$

$$\begin{cases} \textit{stop and accept } \dot{H}_{\geq}\ (\textit{predict extinction}) & \text{if } \lambda_\theta(\mathbf{x}_n) \leq B, \\ \textit{stop and accept } \ddot{H}_{\geq}\ (\textit{predict explosion}) & \text{if } \lambda_\theta(\mathbf{x}_n) \geq A, \\ \textit{continue sampling} & \text{if } B < \lambda_\tau(\mathbf{x}_n) < A, \end{cases}$$

where $\lambda_\theta(\mathbf{x}_n) = \ddot{f}_\theta(\mathbf{x}_n)/\dot{f}_\theta(\mathbf{x}_n)$.

The stopping time for the SPRT$(\theta; B, A)$ is a random variable $N(\theta; B, A)$. Because $\mathrm{Pr}_{\theta'}[X_n \to 0 \text{ or } \infty] = 1$ for all $\theta' \geq \tau$, $N(\theta; B, A)$ is finite with probability 1. As B decreases and A increases, $\mathrm{E}_{\theta'}[N(\theta; B, A)]$ increases under both $\dot{H}_{\geq}$ and $\ddot{H}_{\geq}$, but it is not necessarily true that the first and second error probabilities $\alpha_{\theta'} \equiv \alpha_{\theta'}(\theta; B, A)$ and $\beta_{\theta'} \equiv \beta_{\theta'}(\theta; B, A)$ both decrease (cf. Wald (1947, p.45)). Here,

$$\begin{aligned} \alpha_{\theta'}(\theta; B, A) &\equiv \mathrm{Pr}_{\theta'}[\mathrm{SPRT}(\theta; B, A) \textit{ accepts } \ddot{H}_{\geq} \mid \dot{H}_{\geq}] \\ &= \mathrm{Pr}_{\theta'}[\lambda_\theta(\dot{\mathbf{X}}) \textit{ hits } A \textit{ before } B], \\ \beta_{\theta'}(\theta; B, A) &\equiv \mathrm{Pr}_{\theta'}[\mathrm{SPRT}(\theta; B, A) \textit{ accepts } \dot{H}_{\geq} \mid \ddot{H}_{\geq}] \\ &= \mathrm{Pr}_{\theta'}[\lambda_\theta(\ddot{\mathbf{X}}) \textit{ hits } B \textit{ before } A]. \end{aligned}$$

Wald (1947, §3.2) derived the following upper bounds: for any $\theta \geq \tau$,

$$\alpha_\theta(\theta; B, A) \leq \frac{1-\beta_\theta(\theta; B, A)}{A} \leq \frac{1}{A}, \tag{121}$$

$$\beta_\theta(\theta; B, A) \leq (1-\alpha_\theta(\theta; B, A))B \leq B. \tag{122}$$

Thus if α and β are prespecified, we may choose $B = \beta$ and $A = \frac{1}{\alpha}$ to guarantee that SPRT$(\theta; \beta, \frac{1}{\alpha})$ satisfies the error bounds

$$\alpha_\theta(\theta; \beta, \tfrac{1}{\alpha}) \leq \alpha \quad \text{and} \quad \beta_\theta(\theta; \beta, \tfrac{1}{\alpha}) \leq \beta. \tag{123}$$

Wald also derived the following approximations: if $\alpha + \beta < 1$ then SPRT$\big(\theta; \frac{\beta}{1-\alpha}, \frac{1-\beta}{\alpha}\big)$ more nearly attains the specified error probabilities α and β than does SPRT$\big(\theta; \beta, \frac{1}{\alpha},\big)$, i.e.,

$$\alpha_\theta\Big(\theta; \tfrac{\beta}{1-\alpha}, \tfrac{1-\beta}{\alpha}\Big) \approx \alpha, \quad \beta_\theta\Big(\theta \tfrac{\beta}{1-\alpha}, \tfrac{1-\beta}{\alpha}\Big) \approx \beta, \tag{124}$$

$$\alpha_\theta\Big(\theta; \tfrac{\beta}{1-\alpha}, \tfrac{1-\beta}{\alpha}\Big) + \beta_\theta\Big(\theta; \tfrac{\beta}{1-\alpha}, \tfrac{1-\beta}{\alpha}\Big) \leq \alpha + \beta. \tag{125}$$

6.2. The least favorable distribution for sequential sampling. Because θ is unknown, the SPRT$(\theta; \cdot, \cdot)$ cannot be applied directly (but see Section 6.3.) When the psod satisfies TP2a and TP2b, however, like (81) the testing problem (120) has a convenient solution, namely the SPRT$(\tau; \cdot, \cdot)$. Propositions 4.1 and 4.2 imply that $\dot{H}_\geq$ and $\ddot{H}_\geq$ are separated families and that $(\dot{f}_\tau, \ddot{f}_\tau)$ is a pair of least favorable distributions for (120). Furthermore, by Propositions 4.1 and 4.2, $\alpha_{\theta'}$ and $\beta_{\theta'}$ both decrease as θ' increases. Therefore SPRT$(\tau; \beta, \frac{1}{\alpha})$ (respectively, SPRT$\big(\tau; \frac{\beta}{1-\alpha}, \frac{1-\beta}{\alpha}\big)$) provides an optimal test for $\dot{f}_\tau$ vs. $\ddot{f}_\tau$ for which $\alpha_{\theta'} \leq \alpha$ and $\beta_{\theta'} \leq \beta$ (resp., approximately) for $\theta' \geq \tau$.

Specifically, by (41) and (47),

$$\lambda_\tau(\mathbf{x}_n) \equiv \frac{\ddot{f}_\tau(\mathbf{x}_n)}{\dot{f}_\tau(\mathbf{x}_n)} = \frac{x_n}{x_0}, \tag{126}$$

so the SPRT$(\tau; B, A)$ assumes the simple form

$$\begin{cases} \textit{stop and accept } \dot{H}_\geq \ (\textit{predict extinction}) & \text{if } x_n \leq x_0 B, \\ \textit{stop and accept } \ddot{H}_\geq \ (\textit{predict explosion}) & \text{if } x_n \geq x_0 A, \\ \textit{continue sampling} & \text{if } x_0 B < x_n < x_0 A. \end{cases}$$

Note that this is a *universal* prediction procedure, that is, it is valid for any psod, in particular it does not depend on σ_τ^2. As a consequence, however, it is somewhat conservative.

6.3. A less conservative sequential prediction procedure. If θ were known ($\theta > \tau$), the SPRT$(\theta; \beta, \frac{1}{\alpha})$ (respectively, SPRT$\big(\theta; \frac{\beta}{1-\alpha}, \frac{1-\beta}{\alpha}\big)$) provides an optimal test for $\dot{f}_\theta$ vs. $\ddot{f}_\theta$ for which $\alpha_\theta \le \alpha$ and $\beta_\theta \le \beta$ (resp., $\alpha_\theta \approx \alpha$ and $\beta_\theta \approx \beta$). From (12) and (13),

$$\lambda_\theta(\mathbf{x}_n) \equiv \frac{\ddot{f}_\theta(\mathbf{x}_n)}{\dot{f}_\theta(\mathbf{x}_n)} = \frac{u_\theta^{-x_n} - 1}{u_\theta^{-x_0} - 1}. \tag{127}$$

Because $\lambda_\theta(\mathbf{x}_n)$ is strictly increasing in x_n, the SPRT$(\theta; B, A)$ assumes the following form:

$$\begin{cases} \textit{stop and accept } \dot{H}_\theta \textit{ (predict extinction)} & \text{if } X_n \le x_0 \ell\big(u_\theta^{x_0}, B\big), \\ \textit{stop and accept } \ddot{H}_\theta \textit{ (predict explosion)} & \text{if } X_n \ge x_0 \ell\big(u_\theta^{x_0}, A\big), \\ \textit{continue sampling} & \text{if } x_0 \ell\big(u_\theta^{x_0}, B\big) < X_n < x_0 \ell\big(u_\theta^{x_0}, A\big), \end{cases}$$

where

$$\ell(u, \eta) = \frac{\log\big[(\frac{1-u}{u})\eta + 1\big]}{\log(\frac{1}{u})}, \quad 0 < u < 1, \quad 0 \le \eta < \infty. \tag{128}$$

For fixed u, $\ell(u, \eta)$ increases strictly and continuously from 0 to ∞ as η ranges from 0 to ∞; also $\ell(u, 0) = 0$ and $\ell(u, 1) = 1$. Define $\ell(1, \eta) = \ell(1-, \eta) = \eta$.

Lemma 6.1. *If* $0 < u < 1$ *and* $0 < \eta < 1$ *(resp.,* $1 < \eta < \infty$*), then* $\ell(u, \eta)$ *is strictly decreasing (resp., strictly increasing) in* u *and*

$$\eta < \ell(u, \eta) < 1 \quad \big(\text{resp., } 1 < \ell(u, \eta) < \eta\big). \tag{129}$$

Proof. Set $v = \frac{1-u}{u}$, so that $0 < v < \infty$ and

$$\ell(u, \eta) = \frac{\log\big[v\eta + 1)\big]}{\log(v + 1)} \equiv \bar{\ell}(v, \eta).$$

For $1 < \eta < \infty$, to show that $\ell(u, \eta)$ is strictly increasing for $0 < u < 1$, it suffices to show that $\bar{\ell}(v, \eta)$ is strictly decreasing for $0 < v < \infty$, that is, $\partial \bar{\ell}(v, \eta)/\partial v < 0$. This is equivalent to showing that

$$\frac{\eta \log(v + 1)}{\eta v + 1} - \frac{\log(\eta v + 1)}{v + 1} < 0,$$

equivalently, that

$$\Delta(v,\eta) \equiv (v+1)\log(v+1) - (v+\frac{1}{\eta})\log(\eta v+1) < 0. \tag{130}$$

But $\Delta(v,1) = 0$ for $\eta = 1$ and

$$\begin{aligned} \frac{\partial \Delta(v,\eta)}{\partial \eta} &= -\frac{(v+\frac{1}{\eta})v}{\eta v+1} + \frac{\log(\eta v+1)}{\eta^2} \\ &= -\frac{1}{\eta^2}\Big[\eta v - \log(\eta v+1)\Big] < 0, \end{aligned}$$

hence (A.1) holds. Then by L'Hospital's rule,

$$\eta = \bar{\ell}(0+,\eta) > \bar{\ell}(v,\eta) > \bar{\ell}(\infty-,\eta) = 1,$$

which yields the desired inequalities for $\ell(u,\eta)$. The results for $0 < \eta < 1$ are established in similar fashion. □

From Lemma 6.1, the lower (resp., upper) stopping boundary for the SPRT$(\theta; B, A)$ strictly increases (resp., strictly decreases) as θ increases on $[\tau,\psi)$, hence the stopping region decreases and $N(\theta; B, A)$ decreases. In particular,

$$x_0 B < x_0\ell\big(u_\theta^{x_0}, B\big) < x_0 < x_0\ell\big(u_\theta^{x_0}, A\big) < x_0 A. \tag{131}$$

This difference can be substantial (see Table 4) and implies that

$$\mathrm{E}_{\theta'}[N(\theta; B, A)] < \mathrm{E}_{\theta'}[N(\tau; B, A)] \quad \text{for all } \theta' \geq \tau. \tag{132}$$

Thus if one is willing to assume a fairly unrestrictive upper bound $u_\theta \leq \bar{u} < 1$ for the extinction probability u_θ (e.g., $\bar{u} = 0.90$, 0.95, or 0.99), corresponding to an unrestrictive lower bound $\theta \geq \underline{\theta} \equiv \theta_{\bar{u}} > \tau$ for θ itself, then by using the SPRT$(\underline{\theta}; \beta, \frac{1}{\alpha})$ or SPRT$\big(\underline{\theta}; \frac{\beta}{1-\alpha}, \frac{1-\beta}{\alpha}\big)$), by Proposition 4.1(ii) one would control the first error probability for problem (120), i.e., $\alpha_\theta \leq \alpha$ for all $\theta \geq \underline{\theta}$, while substantially reducing the expected stopping time. If TP2a and TP2b hold, then by Proposition 4.2(iii) the second error probability also would be controlled, i.e., $\beta_\theta \leq \beta$ for all $\theta \geq \underline{\theta}$.

Remark 6.1. For the Poisson(θ) psod, it follows from (33) that

$$\theta_{\bar{u}} = -\frac{\log(\bar{u})}{1-\bar{u}}, \tag{133}$$

	u_θ	x_0	$x_0\ell\big(u_\theta^{x_0}, 0.05\big)$	$x_0\ell\big(u_\theta^{x_0}, \frac{1}{0.05}\big)$
$(\theta = \tau)$	1.0	5	0.25	100
	0.99	5	0.26	70.5
	0.95	5	0.28	37.5
	0.9	5	0.32	25.6
	0.7	5	0.62	12.9
	0.5	5	1.35	9.3
$(\theta = \tau)$	1.0	10	0.50	200
	0.99	10	0.53	113
	0.95	10	0.64	52.0
	0.9	10	0.85	34.6
	0.7	10	2.80	18.4
	0.5	10	5.70	14.3
$(\theta = \tau)$	1.0	20	1.00	400
	0.99	20	1.10	169
	0.95	20	1.67	70.3
	0.9	20	2.93	47.3
	0.7	20	11.6	28.4
	0.5	20	15.7	24.3

Table 4: Stopping boundaries for SPRT$(\theta; 0.05, \frac{1}{0.05})$.

so $\theta_{.90} = 1.0536$, $\theta_{.95} = 1.0259$, $\theta_{.99} = 1.0050$, which lower bounds are close to the critical value $\tau = 1$. For the negative binomial(r, θ) psod, (35) yields

$$\theta_{\bar{u}} = \frac{1 - \bar{u}^{\frac{1}{r}}}{1 - \bar{u}^{\frac{r+1}{r}}}, \tag{134}$$

which reduces to $\theta_{\bar{u}} = \frac{1}{1+\bar{u}}$ for the geometric(θ) psod when $r = 1$. Here again this lower bound will be close to the critical value $\tau = \frac{1}{1+r}$ if $\bar{u}$ is close to 1. In these cases, therefore, the assumption that $u_\theta \leq \bar{u}$ is not very restrictive for $\bar{u} = .90, .95, .99$. $\square$

Remark 6.2. Unlike the fixed-n prediction procedures derived in §5 (cf. Remarks 5.1 and 5.2), the SPRTs compare x_n to x_0 rather than to n in order to predict whether the observed process is on a trajectory toward extinction or toward explosion. Note that if x_0 is small, the SPRTs are useful for predicting explosion but not for predicting extinction. For example,

$$x_0\,\ell\big(u_\theta^{x_0},\,\beta\big) < 1 \iff x_0 < \ell\big(u_\theta,\,\tfrac{1}{\beta}\big), \tag{135}$$

so if $x_0 < \ell\big(u_\theta, \frac{1}{\beta}\big)$ then the SPRT$(\theta; \beta, \frac{1}{\alpha})$ reduces to the SPRT$(\theta; 0, \frac{1}{\alpha})$:

$$\begin{cases} \textit{stop and accept } \dot{H}_\theta \text{ (\textit{predict extinction})} & \text{if } X_n = 0, \\ \textit{stop and accept } \ddot{H}_\theta \text{ (\textit{predict explosion})} & \text{if } X_n \geq x_0 \ell\big(u_\theta^{x_0}, \frac{1}{\alpha}\big), \\ \textit{continue sampling} & \text{if } 1 \leq X_n < x_0 \ell\big(u_\theta^{x_0}, \frac{1}{\alpha}\big), \end{cases}$$

hence will predict extinction only if extinction actually occurs. If $x_0 < \frac{1}{\beta}$, the SPRT$(\tau; \beta, \frac{1}{\alpha})$ also reduces to SPRT$(\theta; 0, \frac{1}{\alpha})$ hence behaves similarly. $\square$

Remark 6.3. Note that the SPRT$(\theta; B, A)$ depends on θ only through the value of the extinction probability u_θ, not on the specific offspring distribution, whether a psod or not. Therefore, hereafter we shall use the notation SPRT$(u_\theta; B, A)$, or simply SPRT$(u; B, A)$. In particular, the universal SPRT$(\tau; B, A)$ in §9.2 is now designated as SPRT$(1; B, A)$. $\square$

7. Examples

Six examples are presented to illustrate the fixed-n (§5) and sequential (§6) procedures for predicting extinction or explosion from the current realization of a GW process. Because the Poisson, negative binomial, and geometric psods are assumed, conditions TP2a and TP2b are satisfied, so these prediction procedures possess the properties asserted in §5.2, 6.2, and 6.3.

Example 7.1: Smallpox in Sao Paolo, Brazil. An outbreak of *variola minor* in Sao Paolo occurred in 1956 (see Table 5). This outbreak was caused by a single infectious individual and lasted four generations before the schools closed; see Becker (1972), Guttorp (1991, p.59). Becker (1977) and Heyde (1979) modeled these data by a GW process; also see Guttorp (1991, p.58). Like Heyde we assume a Poisson(θ) psod; here $\mu_\theta = \theta$, $\tau = 1$, $\sigma_\tau^2 = 1$.

n	0	1	2	3	4
x_n	1	5	3	12	24

Table 5: Occurrences of *variola minor* in Sao Paolo, Brazil, 1956.

These data suggest a trajectory toward explosion. To assess the strength of this prediction, first consider the fixed-n testing problem (81) with $\dot{H}^+_\geq$

(eventual extinction) taken to be the null hypothesis and $\ddot{H}^+_{\geq}$ (explosion) the alternative. Here $x_0 = 1$, $n = 4$, $x_n = 24$ so the exponential approximation (101) for the p-value of the fixed-n prediction procedure (87) is

$$\dot{\pi}^{\mathrm{EXP}}(24;4) = e^{-\frac{47}{4}} \approx 7.89\times 10^{-6}, \tag{136}$$

which strongly supports the prediction of explosion. Because $n = 4$ is not large, this approximation is not entirely reliable. (Because $x_0 = 1$, the EXP and GAM approximations coincide.)

x_0	$\bar{u}$	$x_0\ell\big(\bar{u}^{x_0}, 0.05\big)$	$x_0\ell\big(\bar{u}^{x_0}, \frac{1}{0.05}\big)$	$x_0\ell\big(\bar{u}^{x_0}, 0.01\big)$	$x_0\ell\big(\bar{u}^{x_0}, \frac{1}{0.01}\big)$
1	1.00	0.05	20	0.01	100
	0.99	0.05	18.3	0.01	69.5
	0.95	0.05	14.0	0.01	35.8
	0.90	0.05	11.1	0.01	23.7
8	1.00	0.40	160	0.08	800
	0.99	0.42	97.9	0.08	222.7
	0.95	0.49	47.0	0.10	76.9
	0.90	0.61	31.4	0.12	46.4
14	1.00	0.70	280	0.14	1400
	0.99	0.75	138.5	0.15	276.5
	0.95	0.99	60.3	0.20	90.9
	0.90	1.48	40.1	0.31	55.3
38	1.00	1.90	760	0.38	3800
	0.99	2.29	232.1	0.46	384.2
	0.95	5.13	93.6	1.14	124.8
	0.9	12.39	66.3	4.09	81.5

Table 6: Stopping boundaries for SPRT$(\bar{u};\beta,\frac{1}{\alpha})$, $\alpha = \beta = 0.05$ and 0.01.

Next we apply the sequential testing approach. Here $x_0 = 1$, so the stopping boundaries for the sequential prediction procedure SPRT$(\bar{u};\beta,\frac{1}{\alpha})$ (cf. Remark 6.3) appear in the first tier of Table 6 for $\alpha = \beta = .05, .01$ and $\bar{u} = 1.0, .99, .95, 0.90$. As $\bar{u}$ decreases, SPRT$(\bar{u};\beta,\frac{1}{\alpha})$ becomes less conservative, stopping more quickly. For example, SPRT$(\bar{u} = 1.0; .05, \frac{1}{.05})$ stops and predicts explosion when $x_n \geq 20$, which here occurs when $n = 4$, while

SPRT($\bar{u} = .90; .05, \frac{1}{.05}$) stops and predicts explosion when $x_n \geq 11.1$, which occurs when $n = 3$.

The SPRT($\bar{u} = .90; .05, \frac{1}{.05}$) requires the assumption that $u_\theta \leq \bar{u} = .90$, equivalently $\theta \geq \theta_{\bar{u}} = 1.0536$, see Remark 6.1. The reliability of this assumption can be assessed in two ways. First, $y_{n-1} = 21$ and $y_n = 45$ so $\hat{\theta} = \hat{\mu} = (45 - 1)/21 \approx 2.095$ from (77), which is substantially larger than 1.0536. Second, an estimate of u_θ could be obtained from the nonparametric MLE $\hat{\mathbf{p}}$ of the offspring distribution $\mathbf{p}_\theta$ (cf. Guttorp (1991, Proposition 3.4), also Stigler (1971)), but this would require knowledge of the family histories of each infected individual, which is unavailable. Here, however, $\hat{\mathbf{p}}$ can be obtained from the EM algorithm because n is small, cf. Guttorp (1991, pp. 119-120). For $n = 3$, $\hat{\mathbf{p}}$ puts masses (0.239, 0.428, 0.206, 0.127) on 0, 1, 5, 6, and from (3) the estimated extinction probability for this distribution is found to be 0.424. For $n = 4$ the estimated distribution puts masses (0.332, 0.147, 0.219, 0.302) on 0, 1, 2, and 5, yielding an estimated extinction probability 0.447. Both estimates fall well below the assumed upper bound $\bar{u} = 0.9$. □

Example 7.2: Smallpox in Nigeria. Becker (1976) presented generational data for a smallpox outbreak in Nigeria (see Table 7). Heyde (1979) modeled these data by a GW process, using the Poisson(θ) psod as suggested by Becker (1976, p.776), so again $\sigma_\tau^2 = 1$.

n	0	1	2	3	4	5	6	7
x_n	1	1	7	6	3	8	4	0

Table 7: Occurrences of smallpox in Abakaliki, Nigeria.

Because the outbreak terminated at the 7th generation, our fixed-n prediction methods (§5) are not relevant. Instead, beause $x_0 = 1$ the stopping boundaries of SPRT($\bar{u}; \beta, \frac{1}{\alpha}$) for $\bar{u} = 1.0, 0.99, 0.95, 0.90$ and $\alpha = \beta = .05, .01$ again appear in the first tier of Table 6. Because $1 \leq x_n \leq 7$ for $n = 1, \ldots, 6$ in this example, none of these SPRTs would stop sampling until the extinction observed at $n = 7$. Note that $\hat{\theta} = \hat{\mu} = (30 - 1)/30 = 0.967 < 1$ by (77) ($y_{n-1} = y_n = 30$). □

Example 7.3: Pertussis in Washington State, 2011. The WA State Department of Health reported no cases of pertussis (whooping cough) in Week 8 of 2011. The numbers of new cases in Weeks 9, 10,..., 20 are shown in Table 8 (also see Figure 1). Here $x_0 = 8$, $n = 11$, $x_n = 11$, $y_{n-1} = 66$, and $y_n = 77$.

Week	9	10	11	12	13	14	15	16	17	18	19	20
n	0	1	2	3	4	5	6	7	8	9	10	11
x_n	8	6	5	5	5	7	3	4	7	10	6	11

Table 8: Weekly occurrences of pertussis in Washington State, 2011.

As in Examples 7.1 and 7.2, these data are assumed to arise from a GW process with a Poisson(θ) psod. As a rough check, the MLE of the offspring mean is $\hat{\theta} = \hat{\mu} = (77 - 8)/66 = 1.04551$, which agrees fairly well with Dion's estimate of the offspring variance (cf. Guttorp (1991, p.109)):

$$\tilde{\sigma}^2 \equiv \frac{1}{n} \sum_{\nu=1}^{n} x_{\nu-1} \Big(\frac{x_\nu}{x_{\nu-1}} - \hat{\mu} \Big)^2 = 1.2047. \tag{137}$$

The population counts in Table 8 remain fairly constant, suggesting eventual extinction. To assess this prediction, first consider the fixed-n testing problem (81) with $\ddot{H}^+_{\geq}$ (eventual explosion) taken to be the null hypothesis and $\dot{H}^+_{\geq}$ (eventual extinction) the alternative. Again $\tau = 1$ and $\sigma^2_\tau = 1$, so $n\sigma^2_\tau < 2(x_0 - 2)$, hence the approximations $\ddot{\pi}^{\mathrm{G2}}$ in (107) and $\ddot{\pi}^{\mathrm{G23}}$ in (109) for the prediction procedures (87) are unreliable (cf. Proposition 5.3(i)).

Because $x_0 = 8$, we next compare the data in Table 8 to the stopping boundaries in the second tier of Table 6 for the sequential prediction procedures SPRT($\bar{u}; \beta, \frac{1}{\alpha}$) for $\bar{u} = 1.0, 0.99, 0.95, 0.90$ and $\alpha = \beta = .05, .01$. None of these stop by generation $n = 11$ so neither extinction nor explosion is predicted.

In fact, for the remainder of 2011 the weekly numbers of new cases remained in the range 6-11 for Weeks 21-29, then increased into the range 12-41 for Weeks 30-51. However only 1 new case was reported in Week 52. $\square$

Example 7.4: Pertussis in Washington State, 2012. The number of new cases of pertussis for Weeks 1–12 of 2012 increased dramatically, suggesting possible explosion (Table 9). Here $x_0 = 1$, $n = 11$, $x_n = 98$, $y_{n-1} = 594$, and $y_n = 692$. The MLE $\hat{\mu} = 691/594 = 1.1633$ and $\tilde{\sigma}^2 = 8.0342$.

Because $\hat{\mu} \ll \tilde{\sigma}^2$, the Poisson distribution does not fit these data. Instead, since $\hat{\mu}$ and $\tilde{\sigma}^2$ agree with the mean and variance of the negative binomial NB($\hat{r}, \hat{\theta}$) psod with $\hat{r} = 0.1970$ and $\hat{\theta} = 0.8552$ (cf. Example 3.2), we shall assume the model NB($r = 0.1970, \theta$) with θ unknown ($0 < \theta < 1$), so $\tau = r(1+r)^{-1} = 0.1646$ and $\sigma_\tau^2 = 6.0761$.

Week	1	2	3	4	5	6	7	8	9	10	11	12
n	0	1	2	3	4	5	6	7	8	9	10	11
x_n	1	7	22	38	50	65	78	61	74	96	102	98

Table 9: Weekly occurrences of pertussis in Washington State, 2012.

To assess the evidence for a prediction of explosion, first consider the fixed-n testing problem (81) with $\dot{H}_{\geq}^{+}$ (eventual extinction) taken to be the null hypothesis and $\ddot{H}_{\geq}^{+}$ (eventual explosion) the alternative. From (101), the exponential approximation to the p-value of the fixed-n prediction procedure (87) is

$$\dot{\pi}^{\text{EXP}}(98; 11) = e^{-\frac{195}{11(6.0761)}} \approx 0.054, \tag{138}$$

which moderately supports the prediction of explosion.

By contrast, from Tables 9 and 6 the conservative SPRT$(1.0; \beta, \frac{1}{\alpha})$ with $\alpha = \beta = 0.05$ would have stopped and predicted explosion as early as Week 3! With $\alpha = \beta = 0.01$ this SPRT would not have stopped until Week 11, but the SPRT$(0.90; 0.01, \frac{1}{0.01})$ would have stopped and predicted explosion by Week 4.

In fact a state of health emergency was declared after Week 14 and an innoculation program begun. The number of new cases[7] continue to increase to a peak of 254 in Week 20, then declined to 23 new cases in Week 52. Had these sequential prediction procedures been applied, this program could have begun much earlier, possibly greatly reducing the number of occurrences of the disease. □

Note that the predictions for Examples 7.3 and 7.4 differ even though the change in $\hat{\mu}$ from 2011 to 2012 is small, namely from 1.0455 to 1.1633. □

Example 7.5: California condors. Wilbur (1978) gives the annual population counts of the threatened California condor from 1968 through 1976 (see Table 10). Here $x_0 = 38$, $n = 8$, $x_n = 19$, $y_{n-1} = 183$, and $y_n = 202$; the MLE $\hat{\mu} = 164/183 = 0.8962$ and $\tilde{\sigma}^2 = 2.2755$. Because $\tilde{\sigma}^2$ is not greatly

[7]The weekly data shown have since been revised. We have used the unrevised data because it was those upon which public health decisions were based.

different from the estimated variance $\sigma^2_{\hat{\theta}} = \hat{\theta}/(1-\hat{\theta})^2 = 1.6992$ under the geometric GM($\hat{\theta}$) distribution with $\hat{\theta} = \hat{\mu}/(1+\hat{\mu}) = .4726$ (see Example 3.2), we will assume the GM(θ) psod model ($0 < \theta < 1$) to illustrate its ease of application. For this model $A(\theta) = 1/(1-\theta)$, $\tau = 1/2$, and $\sigma^2_\tau = 2$.

Year	1968	1969	1970	1971	1972	1973	1974	1975	1976
n	0	1	2	3	4	5	6	7	8
Count x_n	38	26	27	18	25	19	19	11	19

Table 10: Annual counts of California condors 1968-1976.

The data in Table 5 suggest a declining population, hence possible extinction. To evaluate this prediction, first consider the fixed-n testing problem (81) with $\ddot{H}^+_{\geq}$ (eventual explosion) as the null hypothesis and $\dot{H}^+_{\geq}$ (eventual extinction) the alternative. Here $x_0 = 38$ and $n = 8$ so $n\sigma^2_\tau < 2(x_0 - 2)$, hence the approximations $\ddot{\pi}^{\mathrm{G2}}$ in (107) and $\ddot{\pi}^{\mathrm{G23}}$ in (109) for the fixed-n prediction procedures (87) are inapplicable (cf. Proposition 5.3(i)).

By contrast, the sequential prediction procedure SPRT$(0.9; 0.05, \frac{1}{0.05})$ would have stopped in 1975 and predicted extinction. (Compare the data in Table 10 to the stopping boundaries in the last row of Table 6.)

In fact, by the mid 1980's all remaining wild condors were captured and moved to zoos, where a breeding program was begun, followed by relocation back to the wild. By 2011 the total wild population had grown to 191, in addition to 178 remaining in captivity. □

Example 7.6: North American whooping cranes. Miller et al. (1974) give the annual counts of migrating whooping cranes, an endangered species, arriving in Texas from 1938 ($n = 0$) through 1972 ($n = 34$); see Figure 4 and Guttorp (1991, p.190)). Here $x_0 = 14$, $n = 34$, $x_n = 51$, $y_{n-1} = 1072$, and $y_n = 1123$; the MLE $\hat{\mu} = 1109/1072 = 1.0345$. Since $\hat{\mu}$ does not differ greatly from Dion's estimate $\tilde{\sigma}^2 = 0.84$, the Poisson(θ) psod model is assumed.

The counts in Figure 4 show an increasing trend, suggesting explosion. To evaluate this prediction, first consider the fixed-n testing problem (81) with $\dot{H}^+_{\geq}$ (eventual extinction) as the null hypothesis and $\ddot{H}^+_{\geq}$ (eventual explosion) as the alternative. Here $\sigma^2_\tau = 1$, so the EXP and GAM approximations (101) and (102) for the p-values of the fixed-n prediction procedure (87) are

$$(139) \qquad \dot{\pi}^{\mathrm{EXP}}(51; 34) = e^{-\frac{101}{34}} \approx 0.051,$$

$$(140) \qquad \dot{\pi}^{\mathrm{GAM}}(51; 34, 14) = \tfrac{17}{14}\left(\tfrac{16}{17}\right)^{14} \textstyle\sum_{r=1}^{14} \binom{14}{r} \frac{1}{16^r} \bar{G}_r\left(\frac{101}{34}\right) \approx 0.086,$$

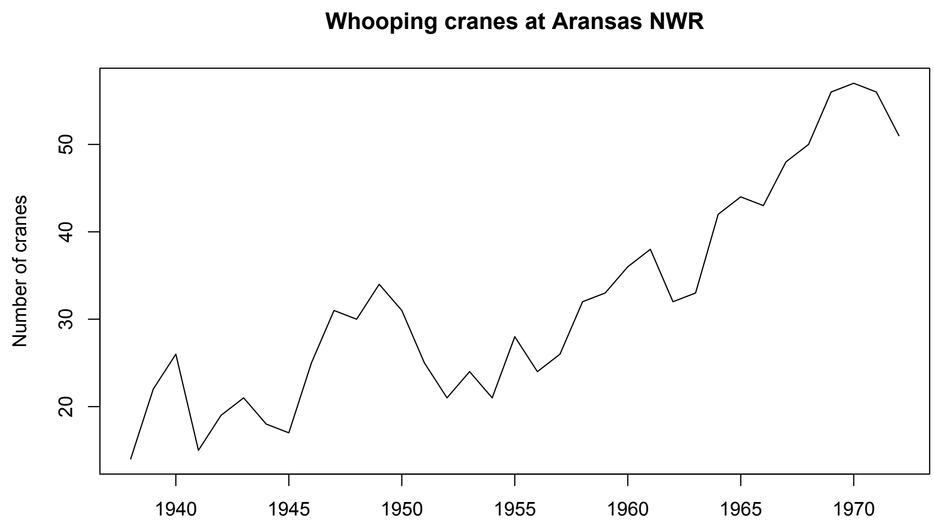

Figure 4: North American whooping crane population counts 1938-1972.

respectively, with $\dot{\pi}^{\text{GAM}}$ expected to be more accurate. This provides modest support for a prediction of explosion.

By contrast, the sequential prediction procedure SPRT$(0.9; 0.05, \frac{1}{0.05})$ would have stopped in 1964 ($n = 26$, $x_n = 42 > 40.1$) and predicted explosion, while SPRT$(0.9; 0.01, \frac{1}{0.01})$ would have stopped in 1969 ($n = 31$, $x_n = 56 > 55.3$) and predicted explosion. (The values 40.1 and 55.3 appear in the third tier of Table 6.) □

References

Athreya, K. B. and Ney, P. E. (1972). *Branching Processes.* Springer-Verlag, Berlin.

Barnard, G. A. (1946). Sequential tests in industrial statistics (with discussion). *J. Royal Statist. Soc.* Supplement **8**, 1-26.

Becker, N. (1972). Vaccination programmes for rare infectious diseases. *Biometrika* **59** 443-453.

Becker, N. (1974). On parametric estimation for mortal branching processes. *Biometrika* **61** 393-399.

Becker, N. (1976). Estimation for an epidemic model. *Biometrics* **32** 769-777.

Becker, N. (1977). Estimation for discrete time branching processes with application to epidemics. *Biometrics* **33** 515-522.

Centers for Disease Control and Prevention (2012). Pertussis epidemic - Washington, 2012; Morbidity and Mortality Weekly Report. Online at: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6128a1.htm

Dion, J.-P. (1975). Estimation of the variance of a branching process. *Ann. Statist.* **3** 1183-1187.

Feller, W. (1968). *An Introduction to Probability Theory and its Applications, Third Edition.* Wiley, New York.

Ghosh, B. K. (1970). *Sequential Tests of Statistical Hypotheses.* Addison-Wesley, Reading, PA.

Guttorp, P. (1991). *Statistical Inference for Branching Processes.* Wiley, New York.

Guttorp, P. and Perlman, M. D. (2014). Testing subcriticality vs. supercriticality in a Galton-Watson branching process with power series offspring distribution. In preparation.

Harris, T. E. (1963). *The Theory of Branching Processes.* Springer-Verlag, Berlin.

Heyde, C. C. (1979). On assessing the potential severity of an outbreak of a rare infectious disease: a Bayesian approach. *Australian J. Statist.* **21** 282-292.

Jagers, P. (1975). *Branching Processes with Biological Applications.* Wiley, New York.

Karlin, S. (1966). *A First Course in Stochastic Processes.* Academic Press, New York.

Karlin, S. (1968). *Total Positivity.* Stanford University Press, Stanford, CA.

Kemperman, J. H. B. (1977). On the FKG-inequality for measures on a partially ordered space. *Indag. Math.* **39** 313- 331.

Kesten, H., Ney, P., and Spitzer, F. (1966). The Galton-Watson process with mean one and finite variance. *Teor. Veroyatnost. i Primenen.* **11** 579-611.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses, Third Edition.* Springer, New York.

Miller, R. S., Botkin, D. B., and Mendelssohn, R. (1974). The whooping crane (*Grus americana)* population of North America. *Biol. Cons.* **6** 106-111.

Noack, A. (1950). A class of random variables with discrete distributions. *Ann. Math. Statist.* **21** 127- 132.

Perlman, M. D. and Olkin, I. (1980). Unbiasedness of invariant tests for MANOVA and other multivariate problems. *Ann. Statist.* **8** 1326-1341.

Seneta, E. (1970). An explicit-limit theorem for the critical Galton-Watson process with immigration. *J. Royal Statist. Soc. Series B* **32** 149-152.

Scott, D. (1987). On posterior asymptotic normality and asymptotic normality of estimators for the Galton-Watson process. *J. Royal Statist. Soc. Series B* **49** 209-214.

Stigler, S. M. (1971). The estimation of the probability of extinction and other parameters associated with branching processes. *Biometrika* **58** 499-508.

Stuart, A. and Ord, J. K. (1991). *Kendall's Advanced Theory of Statistics, Vol. 2.* Oxford University Press, New York.

Taylor, H. and Karlin, S. (1998). *An Introduction to Stochastic Modeling, Third Edition.* Academic Press, Orlando.

Waugh, W. A. O'N. (1958). Conditioned Markov processes. *Biometrika* **45** 241-249.

Wilbur, S. R. (1978). The California Condor, 1966-76: A Look at its Past and Future. North American Fauna No. 72. 136 pp. U.S. Dept. of the Interior, Fish and Wildlife Services, Washington D. C. 1978. Online at: http://library.sandiegozoo.org/journal_list/sWilbur_CAcondor.pdf.