The Annals of Probability
1981, Vol. 9, No. 3, 498–503

THE GEOMETRIC PROGRAMMING DUAL TO THE EXTINCTION PROBABILITY PROBLEM IN SIMPLE BRANCHING PROCESSES

BY PAUL D. FEIGIN AND URY PASSY

Technion-Israel Institute of Technology

It is shown that the well-known problem of determining the probability of extinction in a simple branching process has a duality relation to the problem of determining that offspring distribution which is in a sense closest to the original one and for which the new process is subcritical (or critical). The latter problem is also considered with respect to various measures of distance.

1. Introduction. Our framework is the class of simple (Bienaymé-Galton-Watson) branching processes with single ancestor. Suppose, for a given offspring distribution $q = \{q_0, q_1, \cdots\}$, that the probability z_q^* of ultimate extinction is less than 1—the supercritical case—and consider the following problem: find an offspring distribution p^* which is closest to q and under which ultimate extinction is certain. In the sequel we interpret "close" in an appropriate sense and in fact show that the offspring distribution conditioned on ultimate extinction solves the above problem for a class of entropy-type measures of distance (closeness).

When the measure of distance is Kullback's [4] directed divergence we show that the problem described in the previous paragraph is equivalent to the geometric dual of the familiar problem of determining z_q^* given q. The solution p^* can thus be obtained either from the duality relationship or by directly solving the dual problem in a suitable form.

We initially discuss the definition of a branching measure on a suitable space. This aspect of the paper may be of independent interest.

2. Branching measures. Our processes live on the space Ω of infinite nonnegative integer sequences

$$\omega = \{\omega_1, \omega_2, \cdots\}.$$

Define $\mathscr{F}_n$, for each $n \geq 1$, as the σ-field generated by the cylinder sets $\{\omega : \omega_i = m_i;\ 1 \leq i \leq n\}$ and let $\mathscr{F} = \sigma(\cup_n \mathscr{F}_n)$. A branching process may be defined on Ω by defining the successive generation sizes Z_n and total numbers of particles Y_n as follows (see also Keiding and Lauritzen [3]):

$$Z_0(\omega) \equiv 1, \qquad Y_0(\omega) \equiv 1$$

$$Z_1(\omega) = \omega_1$$

$$Z_2(\omega) = \omega_2 + \cdots + \omega_{1+Y_1}$$

$$\vdots \qquad \vdots$$

$$Z_n(\omega) = \omega_{Y_{n-2}+1} + \cdots + \omega_{Y_{n-1}} \tag{1}$$

$$Y_n(\omega) = Y_{n-1}(\omega) + Z_n(\omega) \tag{2}$$

Received January, 1980; revised September, 1980.

AMS 1970 *subject classifications*. Primary 60J80.

Key words and phrases. Simple branching process, extinction probability, infinite geometric program, geometric duality, Kullback directed divergence, α-entropy, branching measure.

If $Y_{n-1} = Y_{n-2}$ then in (1) $Z_n = 0$, whereupon $0 = Z_{n+1} = Z_{n+2} = \cdots$.

Define $Y = \lim Y_n$; since $\{Y = n]$ is determined by $\{\omega_1, \cdots, \omega_n\}$ it follows that Y is a stopping time with respect to $\{\mathscr{F}_n]$. As far as the branching process is concerned only the history until Y is of interest—that is, we may restrict our attention to $\mathscr{F}_Y = \{F \in \mathscr{F} : F \cap \{Y = n\} \in \mathscr{F}_n; n \geq 1\}$ when defining or comparing branching measures.

Call a measure on $(\Omega, \mathscr{F})$ independent and identically distributed (i.i.d.) if the coordinate variables $[\eta_j(\omega) = \omega_j]$ are i.i.d. with respect to it. Now define a measure Q on $(\Omega, \mathscr{F})$ as *branching* if its restriction Q^Y to $\mathscr{F}_Y$ has an i.i.d. extension $\tilde{Q}$ to $(\Omega, \mathscr{F})$. This extension is unique since it is determined by the offspring distribution $\{q_j = Q^Y(\eta_1 = j)\}$. Clearly, if Q is i.i.d. it is branching; however there exist branching measures which are not i.i.d. A particular example is the subject of the following theorem which provides an analogue of the standard result that a supercritical Galton-Watson (G.W.) process conditioned on extinction is equivalent to a subcritical G.W. process.

THEOREM 1. *Let $B = \{Y < \infty\}$. Then if Q is branching so is Q_B defined by $Q_B(\cdot) = Q(\cdot \mid B)$.*

PROOF. Since $B \in \mathscr{F}_Y$ we note that $(Q_B)^Y = (Q^Y)_B$ (in an obvious notation) so that we may take Q i.i.d. without loss of generality. Then, for $F \in \mathscr{F}_Y$,

(3) $$(Q_B)^Y(F) = Q_B(F) = \sum_{n=1}^{\infty} Q_B(F \cap \{Y = n\}) = \sum_{n=1}^{\infty} Q(F \cap \{Y = n\})/Q(B).$$

Since $G_n = F \cap \{Y = n\} \in \mathscr{F}_n$ we may write

(4) $$G_n = \cap_\nu G(n, \nu)$$

where

$$G(n, \nu) = \{\eta_1 = \nu_1, \cdots, \eta_n = \nu_n\}$$

and the union in (4) is over all $\nu = (\nu_1, \cdots, \nu_n)$ for which $G(n, \nu) \subset G_n$. For such ν, $Y = n$ implies

$$\nu_1 + \cdots + \nu_n = n - 1$$

so

$$Q(G(n, \nu)) = \prod_{i=1}^{n} q_{\nu_i} = Q(B) \prod_{i=1}^{n} \{[Q(B)]^{\nu_i - 1} q_{\nu_i}\}.$$

Taking $\tilde{Q}$ as i.i.d. on $(\Omega, \mathscr{F})$ with offspring distribution $\tilde{q} = \{\tilde{q}_j = [Q(B)]^{j-1} q_j\}$ we have shown that

$$Q(G(n, \nu))/Q(B) = \tilde{Q}(G(n, \nu))$$

which, via (4) and (3) shows that for all $F \in \mathscr{F}_Y$

$$Q_B(F) = \tilde{Q}(F)$$

so that $(Q_B)^Y = \tilde{Q}^Y$ and therefore Q_B is branching. □

REMARK. The fact that $\tilde{q}$ must be a proper distribution if $Q(B) \neq 0$, means that $z_q^* = Q(B)$ solves

(5) $$f_q(z) \equiv \sum_{j=0}^{\infty} z^j q_j = z;$$

a standard result in G.W. process theory.

The usual definition of a branching measure refers only to the history concerning successive generation sizes. In our definition it is not only the process of generation sizes that must display the "branching property" but also the offspring production of individuals in each generation must display the property. To formalize the usual definition in our framework consider the σ-fields $\mathscr{G}_n = \sigma(Z_1, \cdots, Z_n)$, $n \geqslant 1$ and $\mathscr{G} = \sigma(\cup_n \mathscr{G}_n)$. Call $Q^{\mathscr{G}}$, defined on $(\Omega, \mathscr{G})$, a *$\mathscr{G}$-branching* measure if it has an i.i.d. extension $\tilde{Q}$ to $(\Omega, \mathscr{F})$. Also, call Q on $(\Omega, \mathscr{F})$ $\mathscr{G}$-branching if its restriction $Q^{\mathscr{G}}$ is $\mathscr{G}$-branching. Since $\mathscr{G} \subset \mathscr{F}_Y$, an $\mathscr{F}_Y$-branching measure is a $\mathscr{G}$-branching measure. Athreya and Ney [1, page 52] show that Q_B is $\mathscr{G}$-branching; Theorem 1 shows that it is also $\mathscr{F}_Y$-branching.

3. The duality. If we consider the (primal) problem

$$\mathscr{A}: \min z \quad \text{subject to } z \geqslant f_q(z)$$

where f_q is defined in (5) we may express its geometric dual (see the Appendix) as:

$$\mathscr{B}: \min\{\textstyle\sum_{j=0}^{\infty} p_j \log(p_j/q_j)\}/(1-\mu_p) \tag{6}$$

$$\text{subject to } \textstyle\sum_j p_j = 1, \quad \mu_p \equiv \sum j p_j \leqslant 1. \tag{7}$$

The problem $\mathscr{B}$ suggests the following interpretation: find an offspring distribution p^* which minimizes an entropy-like distance from q and which corresponds to a branching process with certain ultimate extinction. Moreover, by duality, p^* is determined by

$$p_j^* = (z_q^*)^{j-1} q_j; \qquad j = 0, 1, \cdots \tag{8}$$

which corresponds to the offspring distribution of the branching measure $Q(\cdot \mid B)$. Via the following theorem we are able to justify this interpretation of $\mathscr{B}$ by showing that it corresponds to the constrained minimization of the more familiar Kullback directed distance between P^Y and Q^Y.

THEOREM 2. *The problem $\mathscr{B}$ is equivalent to the problem*

$$\mathscr{B}': \min\left\{E_{P^Y} \log\left(\frac{dP^Y}{dQ^Y}\right)\right\}$$

subject to $P^Y(B) = 1$, and P branching.

PROOF. Since P and Q are branching and $\mathscr{B}'$ only refers to their restrictions to $\mathscr{F}_Y$, we may take P and Q i.i.d. on $(\Omega, \mathscr{F})$. If $P^Y(B) = 1$ it is not hard to show that the Radon-Nikodym derivative satisfies

$$\frac{dP^Y}{dQ^Y} = \textstyle\sum_{n=1}^{\infty} \dfrac{dP^n}{dQ^n} I(Y = n)$$

where P^n, Q^n are restrictions to $\mathscr{F}_n$ and, due to the i.i.d. property,

$$\frac{dP^n}{dQ^n}(\omega) = \textstyle\prod_{j=1}^{n} [p_{\eta_j}(\omega)/q_{\eta_j}(\omega)].$$

Therefore,

$$\log\left(\frac{dP^Y}{dQ^Y}\right) = \textstyle\sum_{j=1}^{Y} \log[p_{\eta_j}/q_{\eta_j}] I(Y < \infty) \tag{9}$$

and since

$$E_{P^Y} \log\left(\frac{dP^Y}{dQ^Y}\right) = E_P \log\left(\frac{dP^Y}{dQ^Y}\right)$$

we may apply Wald's equation to (9) and conclude

$$E_{P^Y}\log\left(\frac{dP^Y}{dQ^Y}\right) = E_P(Y)E_P\log(p_{\eta_1}/q_{\eta_1})$$
$$= \left(\frac{1}{1-\mu_p}\right)\sum_{j=0}^{\infty} p_j\log(p_j/q_j)$$

where p denotes the offspring distribution corresponding to P. Thus the objective functions in $\mathscr{B}$ and $\mathscr{B}'$ are the same for P branching, and the constraints are also since it is well known that $P^Y(B) = 1$ if and only if $\mu_p \leqslant 1$. □

Now consider α-entropies or measures of distance between P and Q defined by

$$D_\alpha(P, Q) = E_P\left[f_\alpha\left(\frac{dP}{dQ}\right)\right]$$

where $f_\alpha(x) = \log x$, $\alpha = 0$, and $f_\alpha(x) = x^\alpha$, $0 < \alpha \leqslant 1$. We show in the following theorem how to solve a class of problems $\mathscr{C}_\alpha$, and so indicate how $\mathscr{B}'$ may be solved directly. Define, for a given probability measure Q on an arbitrary $(\Omega', \mathscr{F}')$ and for fixed $A \in \mathscr{F}'$

$$\mathscr{C}_\alpha\text{: } \min D_\alpha(P, Q)$$
$$\text{subject to } P(A) = 1,$$

THEOREM 3. *For all $0 \leqslant \alpha \leqslant 1$, the solution to $\mathscr{C}_\alpha$ is given by $P(\cdot) = Q(\cdot \mid A)$, provided $Q(A) > 0$.*

PROOF. For $0 < \alpha \leqslant 1$ and X a nonnegative integrable random variable it follows by an application of Jensen's inequality that

$$Ef_\alpha(X) \geqslant \{Ef_\alpha(X^{-1})\}^{-1}$$

whereas

$$Ef_0(X) = -Ef_0(X^{-1}).$$

Since $f_\alpha(x)$ is concave for $0 \leqslant \alpha \leqslant 1$ we apply Jensen's inequality, while keeping in mind that $P(A) = 1$, to obtain

$$D_\alpha(P, Q) \geqslant \left\{f_\alpha\left(E_P\left[\frac{dQ}{dP}I_A\right]\right)\right\}^{-1} = f_\alpha(Q(A))^{-1}, \qquad 0 < \alpha \leqslant 1$$
$$D_0(P, Q) \geqslant -f_0(Q(A)).$$

The minima are attained for $P(\cdot) = Q(\cdot \mid A)$, unless $Q(A) = 0$. □

COROLLARY. *The solution to $\mathscr{B}'$ is $P^* = Q_B$ and that of $\mathscr{B}$ is given by p^* in* (8).

PROOF. From Theorem 3, Q_B will be a solution of $\mathscr{B}'$ if it is branching, but this follows from Theorem 1. The second half of the assertion follows immediately from the relationship between $\mathscr{B}$ and $\mathscr{B}'$ established in Theorem 2. □

We remark that if $Q^Y(B) = 0$ (i.e., $q_0 = 0$) any P^Y satisfying $P^Y(B) = 1$ is orthogonal to Q^Y and so $D_\alpha(P^Y, Q^Y) = \infty$. Thus, we see the futility in trying to convert a supercritical branching process with certain explosion into a "related" subcritical one.

Acknowledgement. The authors are grateful to the referee and associate editor for pointing out an error in the original version.

APPENDIX

Geometric programming deals with a pair of programs.

(a) The primal program:

$$\mathscr{P}: \inf g_0(t)$$

$$\text{subject to } g_k(t) \leq 1;\ k = 1, \cdots, p$$

$$t_1 > 0, \cdots, t_m > 0$$

where $g_k(t) = \sum_{i\in J[k]} c_i \prod_{j=1}^{m} t_j^{\alpha_{ij}}$; $k = 0, 1, \cdots, p$, and there exists $\tilde{t} > 0$ such that $g_k(\tilde{t}) < 1$, $k = 0, 1, \cdots, p$. The integer sets $J[k]$ are defined by

$$J[k] = \{m_k, m_k + 1, \cdots, n_k\}; \qquad k = 0, 1, \cdots, p$$

with

$$m_0 = 1, \quad m_1 = n_0 + 1, \cdots, \quad m_p = n_{p-1} + 1; \quad n_p = n.$$

The exponents α_{ij} are arbitrary constants, the coefficients c_i are positive and $\{n_0 < n_1 < \cdots < n_p\}$ is a set of integers.

(b) The dual program:

$$\mathscr{D}: \sup \prod_{k=0}^{p} \prod_{i\in J[k]} \left(\frac{c_i \gamma_k}{\delta_i}\right)^{\delta_i}$$

$$\text{subject to } \delta_1 \geq 0, \cdots, \delta_n \geq 0$$

$$\sum_{i\in J[0]} \delta_i = 1$$

$$\sum_{i=1}^{n} \alpha_{ij}\delta_i = 0; \qquad j = 1, \cdots, m$$

where

$$\gamma_k = \sum_{i\in J[k]} \delta_i; \qquad k = 0, 1, \cdots, p.$$

Assume there exists $\tilde{\delta} > 0$ which satisfies the constraints.

Duffin, Peterson and Zener [2] give the following results for these programs.

Given the problems $\mathscr{P}$ and $\mathscr{D}$:

(i) $\mathscr{P}$ attains its constrained minimum value at a point t^* which satisfies the primal constraints;

(ii) the corresponding dual program $\mathscr{D}$ attains its maximum value at a point δ^* which satisfies the dual constraints and

$$\text{(A1)} \qquad \delta_i^* = \begin{cases} c_i \prod_{j=1}^{m} (t_j^*)^{\alpha_{ij}}/g_0(t^*) & i \in J[0] \\ \gamma_k^* c_i \prod_{j=1}^{m} (t_j^*)^{\alpha_{ij}} & i \in J[k];\ k = 1, \cdots, p \end{cases}$$

(iii) the constrained maximum of $\mathscr{D}$ is equal to the constrained minimum of $\mathscr{P}$.

Consider now the following program (see (4)) for the supercritical case, $f_q'(1-) > 1$:

$$\mathscr{A}(N): \min\{z \mid f_q(z, N)z^{-1} \leq 1, z > 0\}$$

where

$$f_q(z, N) = \sum_{j=0}^{N} q_j z^j \leq f_q(z, N + j) \leq f_q(z).$$

The supercriticality condition ensures the existence of $z_0 < 1$ with $f_q(z_0) < z_0$.

The corresponding dual program $\mathscr{D}(N)$ is given by

$$\mathscr{D}(N): \max\left\{v(\delta, N) = \prod_{i=0}^{N} \left(\frac{q_i \gamma}{\delta_i}\right)^{\delta_i} \,\middle|\, \sum_{i=0}^{N} (i - 1)\delta_i + 1 = 0\right\}$$

where

$$\gamma = \sum_{i=0}^{N} \delta_i,$$

and we may choose $\tilde{\delta}$ as follows:

$$\tilde{\delta}_0 = 2 - \frac{1}{N}, \qquad \tilde{\delta}_1 = 1, \qquad \tilde{\delta}_i = \{N(i-1)\}^{-1}; \qquad i = 2, \cdots, N.$$

Thus, for all $N \geqslant 1$ the above results apply. Transforming to the variables

$$p_j = \delta_j/\gamma, \qquad \mu_p = \sum_{j=0}^{N} jp_j$$

the dual becomes

$$\mathscr{D}(N): \max_{\{\mu_p<1\}} \{V(p, N) = [\prod_{i=0}^{N} (q_i/p_i)^{p_i}]^{1/(1-\mu_p)}\}$$

$$\text{subject to } \sum_{i=0}^{N} p_i = 1,\ \mu_p = \sum_{i=0}^{N} ip_i.$$

Let $z^*(N+j)$ and $p^*(N)$ be the solutions of $\mathscr{A}(N+j)$ and $\mathscr{D}(N)$ respectively. Then $z^*(N+j)$ satisfies the constraints of $\mathscr{A}(N)$ and $p^*(N)$ satisfies those of $\mathscr{D}(N+j)$ and therefore

$$\min \mathscr{A}(N) \leqslant \min \mathscr{A}(N+j)$$
$$\max \mathscr{D}(N+j) \geqslant \max \mathscr{D}(N).$$

Moreover, for any N, $\min \mathscr{A}(N) \leqslant 1$ so that we have

$$1 \geqslant \min \mathscr{A}(N+j) = \max \mathscr{D}(N+j) \geqslant \max \mathscr{D}(N) = \min \mathscr{A}(N); \qquad N > 0, j > 0$$

and so by monotonicity

$$z^* = \lim_N z^*(N) = \lim_N V(p^*(N), N).$$

Following (A1), we define for each $N > 0$, the sequence $(p_i^{**}(N))$ by

$$p_i^{**}(N) = \begin{cases} q_i\{z^*(N)\}^{i-1}; & i = 0, \cdots, N \\ 0; & i > N. \end{cases}$$

Clearly,

$$p_i^{**} = \lim_N p_i^{**}(N) = q_i\{z^*\}^{i-1}$$

exists for each i and so we may define the pair of *infinite* geometric programs:

$$\mathscr{A} = \mathscr{A}(\infty): \min\{z \mid f_q(z)z^{-1} \leqslant 1\}$$

$$\mathscr{D} = \mathscr{D}(\infty): \max\{V(p) = [\prod_{i=0}^{\infty} (q_i/p_i)^{p_i}]^{1/(1-\mu_p)} \mid \sum p_i = 1, \mu_p \leqslant 1, p_i \geqslant 0\}$$

with

$$z^* = \min \mathscr{A} = \max \mathscr{D} = V(p^{**}).$$

Finally, minimizing $\{-\log V(p)\}$ in $\mathscr{D}$ yields the form $\mathscr{B}$ of the dual program (see (6) and (7)), and we see from the forms $\mathscr{A}$ and $\mathscr{D}$ that the duality relation also holds when $f'_q(1-) \leqslant 1$ whereupon $z^* = 1$ and $p^{**} = q$. Note that for the case $f'_q(1-) = 1$ (i.e., $\mu_q = 1$) continuity considerations also yield this latter solution for $\mathscr{D}$.

REFERENCES

[1] Athreya, K. B. and Ney, P. E. (1972). *Branching Processes*, Springer, Berlin.
[2] Duffin, R. S., Peterson, E. L. and Zener, C. (1967). *Geometric Programming Theory and Application.* Wiley, New York.
[3] Keiding, N. and Lauritzen, S. L. (1978). Marginal maximum likelihood estimates and estimation of the offspring mean in a branching process. *Scand. J. Statist.* **5** 106–110.
[4] Kullback, S. (1959). *Information Theory and Statistics.* Wiley, New York.

Technion-Israel Institute of Technology
Faculty of Industrial Engineering and Management
Haifa 32 000, Israel