Probability Models - Notes 4

Branching Processes

The informal definition of a branching process (BP) has been discussed in the lecture. Here we give only the formal definition.

Definition *Let X be an integer-valued non-negative r.v. with p.m.f.* $p_k = P\{X = k\},\ k = 0, 1, 2, \ldots$. *We say that a sequence of random variable* $Y_n,\ n = 0, 1, 2, \ldots$, *is a BP if*
1. $Y_0 = 1$
2. $Y_{n+1} = X_1^{(n)} + X_2^{(n)} + \ldots + X_{Y_n}^{(n)}$,
where all r.v.'s $X_j^{(n)}$ *have the same distribution as X and are independent of each other.*
We say that the distribution of X is the generating distribution of the BP.

This definition in fact describes one of the simplest models for population growth. The process starts at time 0 with one ancestor: $Y_0 = 1$. At time $n = 1$ this ancestor dies producing a random number of descendants $Y_1 = X_1^{(0)}$. Each descendant behaves independently of the others living only until $n = 2$ and being then replaced by his own descendants. This process continues at $n = 3,\ 4, \ldots$. Thus, Y_{n+1} is the number of descendants in the $(n+1)^{\text{th}}$ generation produced by Y_n individuals of generation n.

The meaning of the notations we use should by now be clear: $X_j^{(n)}$ is the number of descendants produced by the j^{th} ancestor of the n^{th} generation.

As we see, the r.v. X defined above specifies the number of offspring of an individual. We denote $E(X) = \mu$ and $\text{Var}(X) = \sigma^2$. We denote by $G(t)$ the p.g.f. of X:

$$G(t) = E(t^X) = p_0 + p_1 t + p_2 t^2 + \ldots = \sum_{j=0}^{\infty} p_j\, t^j. \tag{1}$$

Our plan now is as follows:

1. We shall find $E(Y_n)$ and $\text{Var}(Y_n)$.
2. We shall find the recurrence relations for the probability generating functions $G_n(t)$ of Y_n.
3. This will imply the recurrence relations for the probability of extinction $\theta_n \stackrel{\text{def}}{=} P\{Y_n = 0\}$ of the BP by time n.
4. We shall then find the probability of ultimate extinction $\theta = \lim_{n\to\infty} \theta_n$ of the BP.
5. In principle, the whole distribution of Y_n can be computed from $G_n(t)$ but the examples where this can be done explicitly are rare. We shall look at one such example.

1. Finding the mean and variance for Y_n Remember the formula for the expectation and variance of the random sum: if $S = X_1 + X_2 + \cdots + X_N$, where X_j are independent identically distributed random variables and N is an integer-valued non-negative random variable independent of $(X_j)_{1 \le j < \infty}$ then

$$E(S) = E(X) \times E(N), \quad \text{Var}(S) = [E(X)]^2 \times \text{Var}(N) + \text{Var}(X) \times E(N).$$

Theorem 1. $E(Y_{n+1}) = \mu^{n+1}$.

Proof. Note that Y_{n+1} is a random sum with $N = Y_n$ and $X_j = X_j^{(n)}$ being independent and having the distribution of the generating r. v. of the BP. Hence, by the just mention formula,

$$E(Y_{n+1}) = E(X) \times E(Y_n) = \mu E(Y_n)$$

. Applying this relation to $E(Y_n)$ gives $E(Y_n) = \mu E(Y_{n-1})$. Hence

$$E(Y_{n+1}) == \mu E(Y_n) = \mu^2 E(Y_{n-1}) = ... = \mu^{n+1} E(Y_0) = \mu^{n+1}$$

(because $E(Y_0) = 1$). □

Theorem 2. *If* $\mu \neq 1$, *then* $\mathrm{Var}(Y_n) = \frac{\sigma^2 \mu^{n-1}(1-\mu^n)}{(1-\mu)}$. *If* $\mu = 1$ *then* $\mathrm{Var}(Y_n) = n\sigma^2$.

Proof. We consider the case $\mu \neq 1$. To simplify the notation, we write V_n for $\mathrm{Var}(Y_n)$. Since $Y_n = \sum_{j=1}^{Y_{n-1}} X_j^{(n-1)}$ we can write

$$V_n = [E(X)]^2 \times \mathrm{Var}(Y_{n-1}) + \mathrm{Var}(X) \times E(Y_{n-1}) = \mu^2 V_{n-1} + \sigma^2 \mu^{n-1}.$$

We thus have derived the following recurrence relation between V_n and V_{n-1}:

$$V_n = \mu^2 V_{n-1} + \sigma^2 \mu^{n-1}. \tag{1}$$

The proof can now be completed by induction. For $n = 0$ the statement of the theorem reduces to $\mathrm{Var}(Y_1) = 0$ which obviously is true. Suppose that the statement of the theorem has been established for $\mathrm{Var}(Y_j)$, $j = 1, 2, ..., n$. It then follows from (1) that

$$\mathrm{Var}(Y_{n+1}) = \mu^2 \frac{\sigma^2 \mu^{n-1}(1-\mu^n)}{(1-\mu)} + \sigma^2 \mu^n = \frac{\sigma^2 \mu^n (1-\mu^{n+1})}{(1-\mu)}.$$

This proves the first statement of the theorem

The case $\mu = 1$ is even simpler. In this case the recurrence relation reads

$$V_n = V_{n-1} + \sigma^2. \tag{2}$$

If we already know that $V_n = n\sigma^2$ then it follows from (2) that $V_{n+1} = n\sigma^2 + \sigma^2 = (n+1)\sigma^2$. □

Example $X \sim Bernoulli(p)$, where $0 < p < 1$. Then $\mu = E[X] = p$ and $\sigma^2 = \mathrm{Var}(X) = p(1-p)$. Hence $\mu \neq 1$ so that $E[Y_n] = p^n$ and $\mathrm{Var}(Y_n) = V_{n+1} = \frac{p(1-p)p^{n-1}(1-p^n)}{(1-p)} = p^n(1-p^n)$.

2. Finding the probability generating function of Y_n

Consider again the random sum $S = X_1 + X_2 + \cdots + X_N$ and suppose that, apart of the usual assumptions, it is also given that X_j are integer-valued random variables. We can then compute the m.g.f. for S as follows: $G_S(t) = E(t^S) = E[E(t^S|N)]$, where the first equality is just the definition of the $G_S(t)$ and the second one is a part of the relevant theorem (see Notes 3). But

$$E(t^S|N = n) = E(t^{X_1+X_2+\cdots+X_n}) = E(t^{X_1} t^{X_2} \cdots t^{X_n}) = E(t^{X_1})E(t^{X_2}) \cdots E(t^{X_n}) = [G_X(t)]^n.$$

We use here the fact that the expectation of the product of independent r. v.s is equal to the product of expectations. Also $E(t^{X_j}) = G_X(t)$ by the definition of $G_X(t)$. Hence $E(t^S|N) = G_X(t)^N$ and therefore

$$G_S(t) = E[G_X(t)^N] = G_N(G_X(t)). \tag{3}$$

Formula (3) was derived in Notes 3. We repeat the proof here for the sake of completeness. We shall now use (3) in the proof of the following

Theorem 3. *The probability generating functions $G_n(t)$ of the r.v.'s Y_n satisfy the following recurrence relations:*

$$G_0(t) = t, \tag{4}$$

$$G_{n+1}(t) = G_n(G(t)) \text{ for } n \geq 0. \tag{5}$$

Proof. As before, we use the fact that Y_{n+1} is a random sum with $N = Y_n$. Hence (3) implies that

$$G_{n+1}(t) = G_{Y_n}(G(t)) = G_n(G(t)).$$

□

Corollary.

$$G_{n+1}(t) = G(G_n(t)). \tag{6}$$

Proof. We use induction. For $n = 0$ the relation obviously holds: $G_1(t) = G(G_0(t)) = G(t)$ because $G_0(t) = t$ and $G_1(t) = G(t)$. Next, if we already know that $G_n(t) = G(G_{n-1}(t))$ then (5) implies that $G_{n+1}(t) = G_n(G(t)) = G(G_{n-1}(G(t)))$. But using (5) once again we obtain that $G_{n-1}(G(t)) = G_n(t)$ and hence $G_{n+1}(t) = G(G_n(t))$. □

3. Finding the probability of extinction

Remember that (see Notes 1 or corresponding lecture) if Z is a non-negative integer-valued r.v. and $G_Z(t)$ is its p.g.f. then $P\{Z = 0\} = G_Z(0)$.

This means that in our case $\theta_n = G_n(0)$ for $n = 1, 2, ...$ Hence putting $t = 0$ in (6) gives

$$\theta_{n+1} = G_{n+1}(0) = G(G_n(0)) = G(\theta_n) \tag{7}$$

Since $\theta_1 = p_0$ we can then iteratively obtain θ_n for $n = 2, 3, ...$

Remark (7) holds also for $n = 0$ since $G_0(t) = t$. In fact we could start with $n = 0$. Indeed, $\theta_0 = 0$ and hence $p_0 = G_1(0) = \theta_1$.

Example. Suppose $X \sim Bernoulli(p)$ so that $G(t) = pt + q$. Then

$$\begin{aligned}
\theta_1 &= G(\theta_0) = G_X(0) = q = 1 - p \\
\theta_2 &= G(\theta_1) = G_X(q) = pq + q = 1 - p^2 \\
\theta_3 &= G(\theta_2) = G_X(1 - p^2) = p(1 - p^2) + (1 - p) = 1 - p^3
\end{aligned}$$

and you can show using induction that $\theta_n = 1 - p^n$ for $n = 0, 1, 2, ...$ Taking the limit as n tends to infinity gives the probability of eventual extinction θ. Here $\theta = \lim_{n\to\infty} \theta_n = 1$.

Note that for this very simple example you can obtain the result directly. In each generation there can at most be one individual and $P(Y_n = 1)$ is just the probability that the individual in each generation has one offspring, so that $P(Y_n = 1) = p^n$ and therefore $\theta_n = P(Y_n = 0) = 1 - p^n$.

4. Finding the probability of ultimate (eventual) extinction

We only consider the case where $0 < P(X = 0) < 1$ since the other two cases are trivial. If $P(X = 0) = 1$ then the process is certain to die out by generation 1 so that $\theta = 1$. If $P(X = 0) = 0$ then the process cannot possibly die out and $\theta = 0$.

Theorem 4. *If $0 < P(X = 0) < 1$ then the probability of eventual extinction is the smallest positive solution of the equation $t = G(t)$.*

Proof. We first establish that $\theta_{j+1} > \theta_j$ for all $j = 1, 2, \ldots$. Indeed, $G(t)$ is a strictly increasing function of t. Now $\theta_1 = G(0) > 0$. Hence $\theta_2 = G(\theta_1) > G(0) = \theta_1$. Assume that $\theta_j > \theta_{j-1}$ for all $j = 2, \ldots, n$. Then $\theta_{n+1} = G(\theta_n) > G(\theta_{n-1}) = \theta_n$. Hence the statement follows by induction. Thus θ_n is a strictly increasing function of n which is bounded above by 1. Hence it must tend to a limit as n tends to infinity. We shall call this limit $\theta = \lim_{n\to\infty} \theta_n$. Since $\theta_{n+1} = G(\theta_n)$, it immediately follows that $\lim_{n\to\infty} \theta_{n+1} = \lim_{n\to\infty} G(\theta_n)$ and hence $\theta = G(\theta)$.

Let z be any positive solution of $z = G(z)$. It remains to prove that prove that $\theta \leq z$. Now $z > 0$ and so $z = G(z) > G(0) = \theta_1$. Then $z = G(z) > G(\theta_1) = \theta_2$. Now assume that $z > \theta_j$ for all $j = 1, \ldots, n$. Then $z = G(z) > G(\theta_n) = \theta_{n+1}$. Hence by induction $z > \theta_j$ for all $j = 1, 2, \ldots$ and therefore $z \geq \theta$. Since the last inequality holds for any positive solution z to $z = G(z)$, θ must be the smallest positive solution.
□

Note that $t = 1$ is always a solution to $G(t) = t$.

Example. $P(X = x) = 1/4$ for $x = 0, 1, 2, 3$. Therefore $G(t) = (1 + t + t^2 + t^3)/4$. We need to solve $t = G(t)$, i.e. $t^3 + t^2 - 3t + 1 = 0$ i.e. $(t-1)(t^2 + 2t - 1) = 0$. The solutions are $t = 1, \sqrt{2} - 1, -\sqrt{2} - 1$. Hence the smallest positive root is $\sqrt{2} - 1$ so the probability of eventual extinction $\theta = \sqrt{2} - 1$.

Remark. Very often it may be important to know whether $\theta = 1$. It turns out that this is the case if and only if $\mu \leq 1$.

5. Finding the distribution of Y_n

If we know the distribution for X (i.e. the offspring distribution) then we can use the p.g.f. of X to successively find the p.g.f. $G_n(t)$ of Y_n for $n = 1, 2, \ldots$ as formulae (4), (5), and (6) suggest. In principle, once $G_n(t)$ has been found, we can compute $P(Y_n = k) = \frac{1}{k!} G_n^{(k)}(0)$. However, this may be a difficult thing to do.

Example 1. $X \sim Bernoulli(p)$. $G(t) = pt + q$. Then

$G_1(t) = G(t) = pt + q$ so $Y_1 \sim Bernoulli(p)$.

$G_2(t) = G(G_1(t)) = p(pt + q) + q = p^2 t + (1 - p^2)$. Hence $Y_2 \sim Bernoulli(p^2)$.

It is easily shown by induction that, for this very simple example, $Y_n \sim Bernoulli(p^n)$.

Example 2. $X \sim Geometric(\frac{1}{2})$. In this case $G_1(t) = G(t) = \sum_{k=0}^{\infty} \frac{1}{2^{k+1}} t^k = \frac{1}{2-t}$. Then

$G_2(t) = G(G_1(t)) = \frac{1}{2-\frac{1}{2-t}} = \frac{2-t}{3-2t}$.

$G_3(t) = G(G_2(t)) = \frac{1}{2-\frac{2-t}{3-2t}} = \frac{3-2t}{4-3t}$.

It is reasonable to conjecture that $G_n(t) = \frac{n-(n-1)t}{n+1-nt}$. The fact that this is true can now be easily verified by induction. It is also not difficult to see that

$$P(Y_n = 0) = \frac{n}{n+1} \quad \text{and} \quad P(Y_n = k) = \frac{n^{k-1}}{(n+1)^{k+1}} \quad \text{for } k \geq 1.$$

In particular we see that $\theta_n = \frac{n}{n+1}$ and $\theta = \lim_{n\to\infty} \theta_n = 1$.

6. A note on the case of *k* ancestors

Each ancestor generates its own independent branching process. If we let W_j be the number of descendants in generation n generated by the j^{th} ancestor, then the total number in generation n is $W = \sum_{j=1}^{k} W_j$. The W_j are independent identically distributed random variables. Each W_j has the same distribution as Y_n, the number in generation n from one ancestor (i.e. with $Y_0 = 1$).

Therefore $E[W] = kE[Y_n]$ and $\text{Var}(W) = k\text{Var}(Y_n)$.

If the branching process is extinct by generation n, then each of the k branches generated by the k ancestors must be extinct by generation n, so

$$P(W = 0) = P(W_1 = 0, W_2 = 0, ..., W_k = 0) = \prod_{j=1}^{k} P(W_j = 0) = \theta_n^k$$

So the probability of extinction by generation n when there are k ancestors is just θ_n^k.

The probability of eventual extinction is just the probability that each of the k independent branching processes eventually become extinct. Since the branching processes are independent, this is just the product of the probabilities that each of individual branching processes eventually become extinct, which is θ^k.