Probability 2 - Notes 4

Branching Processes.

We can apply our results for random sums to branching processes.

We consider a single type of individual only. When there are males and females we consider one sex only, e.g. when considering the spread of family names only males (and hence male offspring) are considered.

Let Y_n be the number of individuals in generation n. Let X be a random variable specifying the number of offspring an individual has and let $E[X] = \mu$ and $Var(X) = \sigma^2$.

We assume that individuals act independent and that the number of offspring of an individual has the same distribution as X regardless of the generation the individual belongs to.

In generation zero there is one individual, the ancestor. The first generation consists of the offspring of the ancestor. Generation two consists of the offspring of generation one. In general generation $n+1$ consists of the offspring of generation n.

Therefore $Y_{n+1} = \sum_{j=1}^{Y_n} X_j$, where X_j is the number of offspring of the j^{th} individual in generation n. So Y_{n+1} is a random sum.

Finding the mean and variance for Y_n

We use the results for random sums for $n = 0, 1, ...$ to give

$$E[Y_{n+1}] = E[E[Y_{n+1}|Y_n]] = \mu E[Y_n]$$

Since $E[Y_0] = 1$, we obtain

$$E[Y_n] = \mu E[Y_{n-1}] = \mu^2 E[Y_{n-2}] = ... = \mu^n E[Y_0] = \mu^n$$

Also

$$Var(Y_{n+1}] = E[Var(Y_{n+1}|Y_n)] + Var(E[Y_{n+1}|Y_n]) = \sigma^2 E[Y_n] + \mu^2 Var(Y_n)$$

Since $E[Y^n] = \mu^n$ we obtain

$$\begin{aligned}
Var(Y_n) &= \sigma^2\mu^{(n-1)} + \mu^2 Var(Y_{n-1}) \\
&= \sigma^2\mu^{n-1} + \mu^2\left(\sigma^2\mu^{(n-2)} + \mu^2 Var(Y_{n-2})\right) \\
&= \sigma^2\left(\mu^{(n-1)} + \mu^n\right) + (\mu^2)^2 Var(Y_{n-2}) = ... \\
&= \sigma^2\left(\mu^{(n-1)} + \mu^n + ... + \mu^{(2n-2)}\right) + (\mu^2)^n Var(Y_0)
\end{aligned}$$

But $Var(Y_0) = 0$, hence if $\mu = 1$, $Var(Y_n) = n\sigma^2$ and if $\mu \neq 1$ then $Var(Y_n) = \frac{\sigma^2 \mu^{(n-1)}(1-\mu^n)}{(1-\mu)}$.

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

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. of Y_n for $n = 1, 2, ...$

Now $G_{Y_1}(t) = G_X(t)$. Also $Y_{n+1} = \sum_{j=1}^{Y_n} X_j$. Hence, using the result for the p.g.f. for a random sum, we obtain $G_{Y_{n+1}}(t) = G_{Y_n}(G_X(t))$ for $n = 1, 2, ...$

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

$G_{Y_1}(t) = G_X(t) = pt + q$ so $Y_1 \sim Bernoulli(p)$.

$G_{Y_2}(t) = G_{Y_1}(G_X(t)) = p(pt+q) + q = p^2 t + (1-p^2)$. Hence $Y_2 \sim Bernoulli(p^2)$.

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

Finding the probability of extinction

We obtain recurrence relations for the probability of extinction by generation n, which we denote by θ_n. We also obtain the probability of eventual extinction, $\theta = \lim_{n\to\infty} \theta_n$, of the branching process.

When we found the mean, variance and p.g.f. for Y_n we looked back one generation. We could use this latter result to find the probability of extinction by generation n since this is just $\theta_n = P(Y_n = 0) = G_{Y_n}(0)$. However in most cases there is not a simple form for the distribution of Y_n, so in practice this is only easily obtained for small n.

It is much simpler to find the probability of extinction by looking back to the first generation.

$Y_n = \sum_{j=1}^{Y_1} U_j$, where U_j is the number of descendants in generation n who descended from the j^{th} individual in generation 1. So each U_j has the same distribution as Y_{n-1} since it gives the number of $(n-1)^{th}$ generation descendants having the j^{th} individual in generation 1 in the original process as ancestor. Also Y_1 has the same distribution as X. Hence

$$G_{Y_n}(t) = G_{Y_1}\left(G_{Y_{(n-1)}}(t)\right) = G_X\left(G_{Y_{(n-1)}}(t)\right).$$

Now $\theta_n = G_{Y_n}(0)$ for $n = 1, 2, ...$ Hence putting $t = 0$ in the relation above gives

$$\theta_n = G_{Y_n}(0) = G_X\left(G_{Y_{(n-1)}}(0)\right) = G_X(\theta_{(n-1)})$$

Since $\theta_0 = 0$ we can then iteratively obtain θ_n for $n = 1, 2, ...$

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

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

and you can show using induction that $\theta_n = 1 - p^n$ for $n = 0, 1, 2, \ldots$. 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 $\theta_n = P(Y_n = 0) = 1 - p^n$.

A simple result for the probability of 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. When $0 < P(X = 0) < 1$ the probability of eventual extinction is the smallest positive solution of $t = G_X(t)$.

Proof. $G_X(t)$ is a strictly increasing function of t. Now $\theta_1 = G_X(0) > 0$. Hence $\theta_2 = G_X(\theta_1) > G_X(0) = \theta_1$. Assume that $\theta_j > \theta_{(j-1)}$ for all $j = 2, \ldots, n$. Then $\theta_{n+1} = G_X(\theta_n) > G_X(\theta_{n-1}) = \theta_n$. Hence by induction $\theta_{j+1} > \theta_j$ for all $j = 1, 2, \ldots$. Then θ_n is a strictly increasing function of n which is bounded above by 1. Hence it must tend to a limit θ as n tends to infinity. Since $\theta_{n+1} = G_X(\theta_n)$, it immediately follows that $\theta = G_X(\theta)$.

Let z be any positive solution of $z = G_X(z)$. Now $z > 0$ so that $z = G_X(z) > G_X(0) = \theta_1$. Then $z = G_X(z) > G_X(\theta_1) = \theta_2$. Now assume that $z > \theta_j$ for all $j = 1, \ldots, n$. Then $z = G_X(z) > G_X(\theta_n) = \theta_{n+1}$. Hence by induction $z > \theta_j$ for all $j = 1, 2, \ldots$ and therefore $z \geq \theta$. Since θ is less than or equal to any positive solution z to $z = G_X(z)$ it must be the smallest positive solution.

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

Example $P(X = x) = 1/4$ for $x = 0, 1, 2, 3$. Therefore $G_X(t) = (1 + t + t^2 + t^3)/4$. We need to solve $t = G_X(t)$, 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 $\theta = \sqrt{2} - 1$.

A note when there are k ancestors

Each ancestor generates its own independent branching process. If we let W_j be the number 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 $Var(W) = kVar(Y_n)$.

If the branching process is extinct by generation n, then each of the k branching 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.

An important joint distribution

Consider a sequence of n independent trials of an experiment. The binomial distribution arises if each trial can result in 2 outcomes, success or failure, with fixed probability of success p at each trial. If X counts the number of successes, then $X \sim Binomial(n, p)$.

Now suppose that at each trial there are 3 possibilities (A, B or neither) with corresponding probabilities $p, \theta, 1-p-\theta$, which are the same for all trials. Let X and Y count the number of times that A and B occur. The sample space consists of all sequences length n of A's, B's and N's, where N indicates neither A nor B has occurred. If a specific sample point u has x A's and y B's then $P(u) = p^x\theta^y(1-p-\theta)^{n-x-y}$. There are ${}^nC_x{}^{n-x}Cy = \frac{n!}{x!y!(n-x-y)!}$ different sample points with x A's and y B's. Hence

$$P(X=x, Y=y) = \frac{n!}{x!y!(n-x-y)!} p^x\theta^y(1-p-\theta)^{n-x-y}$$

for x and y non-negative integers with $x+y \le n$. This is the trinomial distribution.

The name of the distribution comes from the trinomial expansion

$$\begin{aligned}
(a+b+c)^n &= (a+(b+c))^n = \sum_{x=0}^{n} {}^nC_x a^x (b+c)^{n-x} \\
&= \sum_{x=0}^{n}\sum_{y=0}^{n-x} {}^nC_x{}^{n-x}C_y a^x b^y c^{n-x-y} \\
&= \sum_{x=0}^{n}\sum_{y=0}^{n-x} \frac{n!}{x!y!(n-x-y)!} a^x b^y c^{n-x-y}
\end{aligned}$$

The joint p.g.f.

We define $G_{X,Y}(s,t) = E[s^X t^Y]$.

Properties

(i) $G_X(s) = G_{X,Y}(s,1)$ and $G_Y(t) = G_{X,Y}(1,t)$ so that it is simple to find the marginal distributions and moments for each of X and Y from the joint distribution.

(ii) It is simple to show that

$$E[XY] = \frac{\partial^2 G_{X,Y}(s,t)}{\partial s \partial t}\Big|_{s=1,t=1}$$

(iii) As for the univariate p.g.f., the joint p.g.f. identifies the joint distribution.

If X and Y are independent then $G_{X,Y}(s,t) = G_X(s)G_Y(t)$.

Similarly if $G_{X,Y}(s,t)$ can be written as the product of (a p.g.f with entry s) times (a p.g.f. with entry t), then X and Y are independent with distributions corresponding to those p.g.f.'s.

Use of the joint p.g.f. for the trinomial

Using the trinomial expansion it is simple to find the joint p.g.f.

$$G_{X,Y}(s,t) = E[s^X t^Y] = \sum_{x=0}^{n}\sum_{y=0}^{n-x} \frac{n!}{x!y!(n-x-y)!}(ps)^x(\theta t)^y(1-p-\theta)^{n-x-y} = (ps+\theta t+(1-p-\theta))^n$$

Hence $G_X(s) = G_{X,Y}(s,1) = (ps+(1-p))^n$ and hence $X \sim Binomial(n,p)$ and $G_Y(t) = G_{X,Y}(1,t) = (\theta t + (1-\theta))^n$ and so $Y \sim Binomial(n,\theta)$.

$$\frac{\partial G_{X,Y}(s,t)}{\partial s \partial t} = n\theta(n-1)p(ps+\theta t+(1-p-\theta))^{n-2}$$

and therefore $E[XY] = n(n-1)p\theta$. Hence

$$Cov(X,Y) = E[XY] - E[X]E[Y] = n(n-1)p\theta - npn\theta = -np\theta$$

and so

$$\rho(X,Y) = \frac{-np\theta}{\sqrt{np(1-p)n\theta(1-\theta)}} = -\sqrt{\frac{p\theta}{(1-\theta)(1-p)}}$$

Note that when $p+\theta=1$ then $\rho(X,Y)=-1$. This corresponds to the case when only A or B can occur so that $X+Y=n$, i.e. $Y=n-X$. So there is an exact linear relation for Y in terms of X, with the coefficient of X negative.

The multinomial distribution

Now suppose that there are k outcomes possible at each of the n independent trials (the outcomes are $A_1, A_2, ..., A_k$) with probabilities $p_1, ..., p_k$ where $\sum_{j=1}^{k} p_j = 1$. Let X_j count the number of times A_j occurs. Then

$$P(X_1 = x_1, ..., X_{k-1} = x_{k-1}) = \frac{n!}{x_1!x_2!...x_{k-1}!(n-\sum_{j=1}^{k-1} x_j)!} p_1^{x_1} p_2^{x_2} ... p_{k-1}^{x_{k-1}} p_k^{n-\sum_{j=1}^{k-1} x_j}$$

where $x_1, x_2, ..., x_{k-1}$ are non-negative integers with $\sum_{j=1}^{k-1} x_j \leq n$.