

Probability generating functions

In all this section we will consider random variable X such that $p(X=r)=0$ for all r real number which is NOT a positive integer. Hence, the random variable X will be discrete with values in \mathbb{N} .

Def / Given the sequence $P(X=i)$, the probability generating function of X is the generating function

$$G_X(t) = \sum_{i=0}^{\infty} p(X=i) t^i = \mathbb{E}[t^X]$$

Generating function are formal power series, but can be seen as analytic objects. In the following argument both worlds (algebraic + analytic) will be combined in order to get result.

Ex / i) X is such that $p(X=c)=1$ for a certain c , and $p(X=c')=0$ if $c' \neq c$. Then $G_X(t)=t^c$.

ii) Poisson distribution: remember that $p(X=c) = \frac{\lambda^c}{c!} e^{-\lambda}$, hence,

$$G_X(t) = \sum_{c=0}^{\infty} p(X=c) t^c = \sum_{c=0}^{\infty} \frac{\lambda^c}{c!} e^{-\lambda} t^c = e^{\lambda(t-1)}.$$

Let us see the first properties of such functions:

Prop / ① Convergence: there is a radius of convergence $R \geq 0$, such that the sum converges absolutely if $|t| < R$ and diverges if $|t| > R$. In our case, $R > 1$ (possibly ∞).

② Differentiation: $G_X(t)$ can be differentiated (or integrated) term by term any number of times when $|t| < R$. In particular, this is true for $|t| < 1$.

③ Abel's Theorem: as $p(X=c) \geq 0$ for all c , $\lim_{t \rightarrow 1^-} G_X(t) = \sum_c p(X=c) = 1$.

④ Uniqueness: given $G_X(t)$, its probabilities $p(X=c)$ are completely determined by the derivatives of $G_X(t)$.

We also need the following combinatorial observation: given a sequence of numbers $\{a_n\}_{n \geq 0}$ and $\{b_n\}_{n \geq 0}$, then define $c_n = a_0 b_n + \dots + a_n b_0$. Writing then:

$$\begin{aligned} A(t) &= \sum_{n \geq 0} a_n t^n \\ B(t) &= \sum_{n \geq 0} b_n t^n \end{aligned} \Rightarrow C(t) = \sum_{n \geq 0} c_n t^n = \sum_{r \geq 0} a_r t^r \sum_{s \geq 0} b_s t^s = A(t)B(t)$$

With this comment in mind we can just prove now the first easy consequence:

Theorem / Given a random variable X and $G_X(t)$, we have that:

- a) $G_X(0) = p(X=0)$ and $G_X(1) = 1$.
- b) $\mathbb{E}[X] = G_X'(1)$
- c) More generally, $\frac{d^k}{dt^k} G_X(1) = \mathbb{E}[X(X-1)\dots(X-k+1)]$.

Proof / Just apply the definition and evaluate expression at $t=1$.

In particular, it is important to notice that $\text{Var}[X]$ can be expressed in the following way:

$$\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X(X-1)+X] - \mathbb{E}[X]^2 = G_X''(1) + G'(1) - G'(1)^2.$$

The use of generating functions is specially useful when dealing with the sum of random variables which are independent. Observe that if X and Y are independent random variables, then

$$P(X+Y=u) = \sum_{r \geq 0} P(X=r, Y=u-r) = \sum_{r \geq 0} p(X=r) p(Y=u-r)$$

This property can be exploited to prove the following

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

Proof / We use the expression $G_X(t) = \mathbb{E}[t^X]$, $G_Y(t) = \mathbb{E}[t^Y]$; then:

$$\mathbb{E}[t^{X+Y}] = \mathbb{E}[t^X t^Y] = \mathbb{E}[t^X] \mathbb{E}[t^Y]$$

This can also be proven "combinatorially" by using the previous expression for the probability distribution.

This result tells us the following: whenever we have X_1, X_2, \dots, X_n independent random variables with integer positive values, then writing $X = X_1 + \dots + X_n$, and we have that

$$G_X(t) = G_{X_1}(t) G_{X_2}(t) \dots G_{X_n}(t)$$

The question is: what do we have if x is also a random variable?

Theorem / If X_1, X_2, \dots is a sequence of independent and identically distributed random variables, and $N (\geq 1)$ is a random variable which is independent of $\{X_i\}_{i \geq 1}$, then $S = X_1 + \dots + X_N$ has generating function

$$G_S(t) = G_N(G_X(t)) \quad (G_{X_1} = G_{X_2} = \dots = G_X)$$

Proof / We use the formula of the conditional expectation: remember that, for random variable U , and B an event. Then

$$\mathbb{E}[U|B] = \frac{1}{P(B)} \mathbb{E}[U \mathbb{I}_B]$$

let A_r be the event $\{N=r\}$; then $\sum_{r \geq 0} \mathbb{I}_{A_r} = 1$, and hence:

$$\mathbb{E}[U] = \mathbb{E}[U \cdot [\mathbb{I}_{A_r}]] = \sum_{r \geq 0} \mathbb{E}[U|A_r] p(A_r).$$

In our context, we have that:

$$\begin{aligned} G_S(t) &= \mathbb{E}[t^S] = \sum_{r \geq 0} \mathbb{E}[t^S \mathbb{I}_{A_r}] = \sum_{r \geq 0} \mathbb{E}[t^S | N=r] p(N=r) \\ &= \sum_{r \geq 0} \mathbb{E}[t^{X_1+\dots+X_r}] p(N=r) = \sum_{r \geq 0} [\mathbb{E}[t^X]^r p(N=r)] = G_N(G_X(t)) \end{aligned}$$

Ex / let N be a Poisson distribution with parameter λ , and let X_1, \dots, X_N be $\text{Ber}(p)$. Then, $S = X_1 + \dots + X_N$ has the following properties:

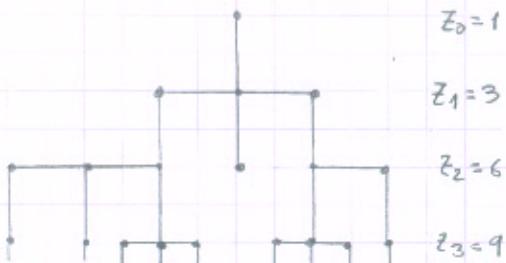
$$\begin{aligned} G_N(t) &= e^{\lambda(t-1)} \quad (\Rightarrow G_S(t) = G_N(G_X(t)) = \exp(\lambda(p t + (1-p)-1)) = \\ G_X(t) &= pt + (1-p) \quad = \exp(\lambda(pt - p)) = e^{\lambda p(t-1)} \end{aligned}$$

So, in other words, S is a $\text{Po}(\lambda p)$. Using other techniques would be difficult to be applied to get this result.

Application: Branching process in discrete time

The previous methodology can be applied successfully to study population dynamic processes. We will study now the most simple model of reproduction, and show a dichotomy by means of probability generating functions.

Assume that a population evolves in generations, and that the number of members in the n -th generation is equal to Z_n . Each member of the n -th generation gives rise to a family (possibly empty) of members of the $(n+1)$ -generation. The size of the family follows a random distribution X :



For each member, the number of birth is given by the random variable X . The size of the family of two different members of the same generation is independent.

Let us write $G_m(s) = \mathbb{E}[s^{Z_m}]$ and also $G(s) = \mathbb{E}[s^X]$. We study now $G_m(s)$ in terms only of n and of course of $G(s)$; notice that $P(X=0) \neq 0$.

Theorem / For all $n, m \geq 0$, $G_{m+n}(s) = G_m(G_n(s))$. In particular, $G_m(s) = G_0^{(m)} \circ G(s)$.

Proof / Each member of the $(m+n)$ -th generation has a single ancestor in the m -th generation. Hence,

$$Z_{m+n} = X_1 + \dots + X_{Z_m},$$

where X_i is the number of members in the $(m+n)$ -th generation which arise from member i in the m -th generation. Observe that by independence of the size of the families, we have that X_i has the same distribution as Z_m , and also X_i and X_j are independent. So, we have that:

$$G_{X_1}(s) = G_{X_2}(s) = \dots = G_{X_{Z_m}}(s) \Rightarrow G_{m+n}(s) = G_m(G_{X_1}(s)) = G_m(G_m(s)).$$

Now, we can iterate this relation: $G_{m+n}(s) = G(G_{m+n-1}(s)) = G(G(G_{m+n-2}(s))) = \dots$ which give the second statement we wanted to show.

As an easy application, let us see some parameters associated to Z_n :

Prop / Let $\mu = \mathbb{E}[X]$, $\sigma^2 = \text{Var}[X]$. Then $\mathbb{E}[Z_n] = \mu^n$ and

$$\text{Var}[Z_n] = \begin{cases} n\sigma^2 & , \mu=1 \\ \sigma^2(\mu^n - 1) \mu^{n-1} (\mu-1)^{-1} & , \mu \neq 1. \end{cases}$$

Proof / Take $G_n(s) = G(G_{n-1}(s))$, and make the derivative at $s=1$. We get that:

$$G'_n(1) = \underbrace{G'(G_{n-1}(1))}_{1} \cdot G'_{n-1}(1) \Rightarrow \mathbb{E}[Z_n] = \mu \mathbb{E}[Z_{n-1}] \Rightarrow \mathbb{E}[Z_n] = \mu^n.$$

To compute the variance, we compute $G''_n(1)$ and we apply the formula of the variance in terms of the probability generating function.

With all this information we can go to the study of the central question in this part. Write $A_n = \{Z_n = 0\}$. The population extints iff for some n , $Z_n = 0$. So we define an event E_X as

$$E_X = \bigcup_n A_n.$$

Clearly, $A_n \subseteq A_{n+1}$, and so $\mathbb{P}(Z_n=0) = p(X)$. We have then a very easy characterization:

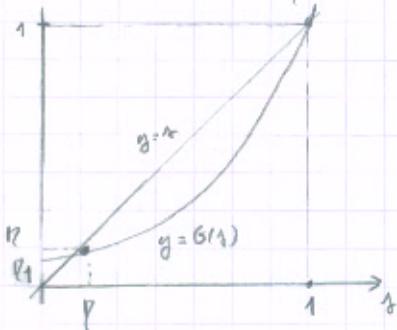
Theorem / let $E[X]=\mu$, and write η for the smallest (non-negative) solution of the equation $\eta = G(\eta)$. Then, $p(X)=\eta$. In particular,

- $\circledcirc \quad \mu < 1 \Rightarrow \eta = 1$ (and the population extincts with prob 1)
- $\circledcirc \quad \mu > 1 \Rightarrow \eta < 1$
- $\circledcirc \quad \mu = 1 \Rightarrow \eta = 1$ as long as the family size distribution has strictly positive variance

Proof / Write $\eta_n = p(Z_n=0)$. Then, in particular,

$$\eta_n = (s_n(0)) = G(s_{n-1}(0)) = G(\eta_{n-1}).$$

As we have said, $\eta = \lim_{n \rightarrow \infty} \eta_n$, so we need to study the last solution of the equation $\eta = G(\eta)$. As $G(1)$ is a continuous function, this guarantees that $\eta = G(\eta)$. Let us study then this equation:



Some point to have in mind:

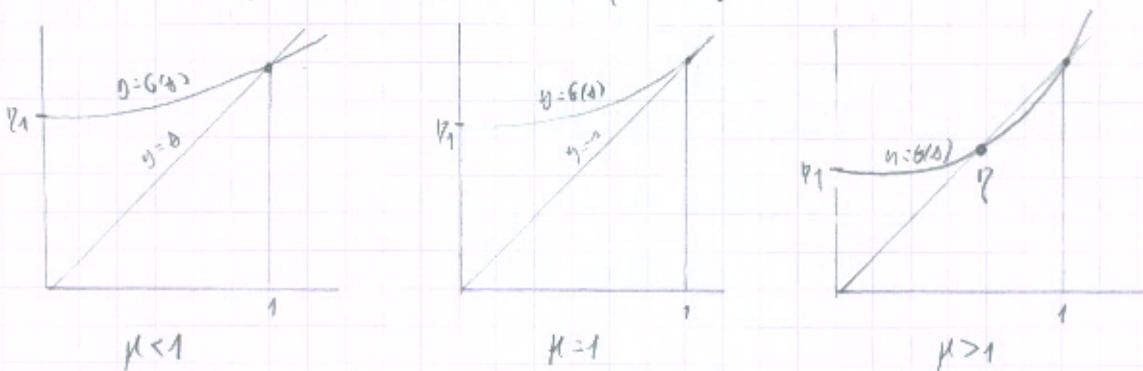
- $\circledcirc \quad G(s)$ is an increasing function in s , with $G(0)=\eta_1$, $G(1)=1$.

Let us show now that if $\bar{\eta}$ is any non-negative root to the equation $\eta = G(\eta)$, then $\eta \leq \bar{\eta}$:

$$\eta_1 = G(0) \leq G(\bar{\eta}) = \bar{\eta}.$$

Iterating this, we have that $\eta_2 = G(\eta_1) \leq G(\bar{\eta}) = \bar{\eta}$, so by induction $\eta_n \leq \bar{\eta} \forall n \Rightarrow \eta \leq \bar{\eta}$.

Let us now go to the study of the three possible cases. Observe that $G(s)$ is a concave function because $G''(s) \geq 0$ if $\alpha \geq 0$ (coefficient of G are positive). This tells us that the curves $y = G(s)$ and $y = s$ intersect only at $s = \eta$ and in $s = 1$. Also, $\eta_1 = p(X=0) \neq 0$, otherwise all is trivial (there is no selection). The 3 pictures we obtain are the following:



Observe that the value of μ is precisely $G'(1)$, and this determines 2 possible behaviours for the curves under study. In the case $\mu=1$, we must say something more: if $\mu=1$, $\sigma^2=0$, then $p(X=1)=1$, and so with probability 0 there is extinction, otherwise ($\sigma^2>0$), $\eta=1$.

Def / The branching process is called subcritical, critical and supercritical if $\mu < 1$, $\mu = 1$ and $\mu > 1$.

Def / The extinction time is $\tau = \min\{n \geq 1 : Z_n=0\}$.

In particular, $p(\tau > n) = p(Z_n > 0) = 1 - p(Z_n=0) = 1 - G_n(0)$. This can be exploited to estimate the extinction time in both the subcritical and the critical case, which differ in terms of its speed.