

## Complex Networks Solutions 8

---

1. In all cases below, we assume that  $x \in [0, \infty)$ . The generating function might also be defined for  $x < 0$ , but we are not asked to compute it on that range.

(a) The generating function is given by

$$G(x) = \sum_{n=-\infty}^{\infty} \mathbb{P}(\xi = n)x^n = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda x)^n}{n!} = e^{-\lambda} e^{\lambda x} = e^{\lambda(x-1)}.$$

The sum converges, and hence  $G(x)$  is defined and finite, for all  $x \in [0, \infty)$ . Also,  $G(\cdot)$  is differentiable at 1, with derivative  $\lambda$ , which is the mean of  $\xi$ .

(b) The generating function is given by

$$G(x) = \sum_{n=-\infty}^{\infty} \mathbb{P}(\xi = n)x^n = px \sum_{n=1}^{\infty} [(1-p)x]^{n-1}.$$

Note that the sum diverges if  $(1-p)x \geq 1$ , i.e.,  $x \geq 1/(1-p)$ . Hence,  $G(x) = +\infty$  in this case. The sum converges if  $x \in [0, 1/(1-p))$ . Thus, we get

$$G(x) = \begin{cases} \frac{px}{1-(1-p)x}, & \text{if } x < \frac{1}{1-p}, \\ +\infty, & \text{otherwise.} \end{cases}$$

If  $p \in (0, 1)$ , then  $1/(1-p)$  is strictly bigger than 1, i.e., 1 is in the interior of the region where  $G(\cdot)$  is finite. It is easy to see that  $G(\cdot)$  is differentiable at 1, and  $G'(1) = 1/p$ , which is the mean of  $\xi$ .

(c) The generating function is given by

$$G(x) = \sum_{n=-\infty}^{\infty} \mathbb{P}(\xi = n)x^n = \sum_{n=0}^{\infty} \frac{x^n}{(n+1)(n+2)}.$$

It is clear that the sum diverges if  $x > 1$ , as  $x^n$  grows to infinity much faster than  $(n+1)(n+2)$ . If  $x = 1$ , then we have

$$G(1) = \sum_{n=0}^{\infty} \frac{1}{(n+1)(n+2)} = \sum_{n=0}^{\infty} \left( \frac{1}{n+1} - \frac{1}{n+2} \right) = 1,$$

as the sum telescopes. If  $x = 0$ , then

$$G(0) = \mathbb{P}(\xi = 0) = \frac{1}{2}.$$

Finally, for  $x \in (0, 1)$ , we follow the hint:  $\sum_{n=0}^{\infty} x^n = 1/(1-x)$ . Integrating once, we get

$$\sum_{n=0}^{\infty} \frac{x^{n+1}}{n+1} = -\log(1-x).$$

Integrating once more,

$$\sum_{n=0}^{\infty} \frac{x^{n+2}}{(n+1)(n+2)} = (1-x) \log(1-x) + x.$$

Putting these results together, we have

$$G(x) = \begin{cases} \frac{(1-x) \log(1-x) + x}{x^2}, & \text{if } x \leq 1, \\ +\infty, & \text{if } x > 1. \end{cases}$$

It is left as an exercise to you to check, using L'Hôpital's formula, that the above expression for  $G(x)$  gives the correct values at  $x = 0$  and  $x = 1$ .

The function  $G(\cdot)$  is finite on  $[0, 1]$  and infinity on  $(1, \infty)$ . It is not even continuous at 1, so it can't possibly be differentiable at 1. It is, however, differentiable for  $0 < x < 1$ , and we can compute

$$\begin{aligned} \lim_{x \uparrow 1} G'(x) &= \lim_{x \uparrow 1} \left( -\frac{\log(1-x)}{x^2} - \frac{2(1-x) \log(1-x) + 2x}{x^3} \right) \\ &= -\lim_{x \uparrow 1} \frac{(2-x) \log(1-x) + 2x}{x^3} = +\infty, \end{aligned}$$

because the numerator tends to  $-\infty$  and the denominator to 1, as  $x$  increases to 1. This is also the mean of  $\xi$ , as

$$\mathbb{E}[\xi] = \sum_{n=0}^{\infty} n \mathbb{P}(\xi = n) = \sum_{n=0}^{\infty} \frac{n}{(n+1)(n+2)},$$

which diverges as the summands are asymptotic to  $1/n$ , and the sum of  $1/n$  diverges.

**Remark.** The point of this problem was to illustrate the range of possible behaviours of generating functions of a non-negative random variable. The generating function  $G(\cdot)$  is always finite on  $[0, 1]$ , with  $G(1) = 1$ . On  $(1, \infty)$ ,  $G(\cdot)$  can be finite on all of it, or none of it, or a sub-interval of the form  $(1, x^*)$ . In part (b), we saw an example in which  $G$  increased continuously to infinity, whereas in part (c),  $G$  had a discontinuous jump to infinity; thus, both are possible. It was true in these problems, and it is true in general, that  $G$  is differentiable in the interior of the interval on which it is finite. Moreover, the left derivative at 1 always exists, and is equal to the mean of the random variable  $\xi$ . Note that  $\xi$  is non-negative, and hence its mean is well defined, but it could be  $+\infty$ .

2. (a) Wooster has probability  $1 - p$  of losing in any round, independent of previous rounds. When he does so, he loses his entire fortune and has to stop. Hence, the number of rounds he plays has a Geometric( $1 - p$ ) distribution, with mean  $1/(1 - p)$ .
- (b) For each pound that Jeeves bets, he gets \$2 with probability  $p$  and \$0 with probability  $1 - p$ , independent of his other bets. Thus, you can think of each pound as having two children with probability  $p$  and zero children with probability  $1 - p$ . In other words, Jeeves' fortune evolves as a branching process with offspring distribution  $2 \times \text{Bern}(p)$ , where  $\text{Bern}(p)$  denotes a Bernoulli random variable with parameter  $p$ .

We can work out the extinction probability for this branching process. The generation function of the offspring distribution is given by

$$G(x) = px^2 + (1-p)x^0 = 1 - p + px^2.$$

Solving the equation  $G(x) = x$ , i.e.,  $(1-p) - x + px^2 = 0$ , we obtain the roots  $x = 1$  and  $x = \frac{1-p}{p}$ . The latter is the smallest solution if  $p > 1/2$ , and so the extinction probability is  $(1-p)/p$ . No matter how long the evening, Jeeves will still be playing at the end of it with probability at least  $1 - (1-p)/p = 2 - (1/p)$ , which is the probability of non-extinction.

3. (a) The number of potential Hamiltonian paths is the number of permutations of  $n$  nodes, which is  $n!$ . Any such path is present if all  $n - 1$  edges on it are present, which has probability  $p^{n-1}$ . Letting  $i$  denote the index of one such path, and  $\chi_i$  its indicator, the number of Hamiltonian paths is given by

$$N_H = \sum_{i=1}^{n!} \chi_i,$$

which is a random variable. The expected number of such paths is given by

$$\mathbb{E}[N_H] = \sum_{i=1}^{n!} \mathbb{E}[\chi_i] = n!p^{n-1} \approx \left(\frac{n}{e}\right)^n \sqrt{2\pi n} \left(\frac{\lambda}{n}\right)^{n-1} = \text{const.} n^{3/2} \left(\frac{\lambda}{e}\right)^{n-1}.$$

We have used the linearity of expectation to get the first equality and the fact that all  $\chi_i$  have the same mean  $p^{n-1}$  to get the second. The approximate equality comes from Stirling's formula for  $n!$  and substituting the expression for  $p$ . It is now obvious that  $\mathbb{E}[N_H]$  tends to infinity as  $n$  tends to infinity if  $\lambda \geq e$ .

- (b) The random graph  $G(n, p)$  is disconnected with high probability if  $p < (1 - \epsilon)(\log n)/n$  for some fixed  $\epsilon > 0$  (by which we mean that the inequality holds for all  $n$  and the corresponding  $p$ . The dependence of  $p$  on  $n$  has not been made explicit in the notation; we should really write  $p_n$ , but have chosen not to in order not to make the notation too heavy. Make sure you keep in mind that  $p$  is really  $p_n$  when interpreting inequalities like the above.)
- (c) The quantity  $(\log n)/n$  is much bigger than  $\lambda/n$  for all  $n$  sufficiently large. Thus, we can pick a  $p$  that is bigger than  $\lambda/n$  but smaller than  $((1 - \epsilon) \log n)/n$ . For any such  $p$ , part (a) tells us that  $G(n, p)$  has infinitely many Hamiltonian paths in expectation, while part (b) tells us that  $G(n, p)$  is disconnected whp. But if  $G(n, p)$  is disconnected, it cannot have even a single Hamiltonian path because such a path will have to visit every vertex, and some vertices are unreachable! This is the paradox.

The resolution is similar to a counterexample we saw in lectures. The fact that the expected number of Hamiltonian paths tends to infinity is not sufficient to guarantee that there is at least one whp. It may be the case that much of the time there is no Hamiltonian path, but when there is one, there is a very large number. This would be consistent with the expected number of Hamiltonian paths tending to infinity, even as the probability that there is at least one tends to zero!

4. (a) By Cayley's formula, the number of spanning trees in the complete graph on  $n$  nodes is  $n^{n-2}$ . For any one of these, the probability that it is present in  $G(n, p)$  is the probability that all  $n - 1$  of its edges are present, which is  $p^{n-1}$ . Hence, by the linearity of expectation, the expected number of spanning trees present in  $G(n, p)$  is  $n^{n-2}p^{n-1}$ .
- (b) We saw in lectures that  $G(n, p)$  is disconnected if it contains an isolated node, which happens with high probability if  $p = (\log n - \alpha(n))/n$ , where  $\alpha(n)$  is any sequence tending to infinity, however slowly, as  $n$  tends to infinity.
- (c) From part (a), the expected number of spanning trees tends to infinity if  $p = c/n$  for any  $c > 1$ . In this case, the expected number of spanning trees is  $n^{n-2}p^{n-1} = c^{n-1}/n$ , which tends to infinity as  $n$  tends to infinity because  $c^{n-1}$  grows much faster than  $n$ . (A more careful analysis shows that the expected number of spanning trees tends to infinity if we take  $p = (1 + \beta(n))/n$  for any sequence  $\beta(n)$  tending to infinity fast enough that  $\frac{\log n}{n\beta(n)}$  tends to zero. But such a precise answer is not expected.)
- Hence, if  $p > c/n$  for some  $c > 1$ , but  $p < (c' \log n)/n$  for some  $c' < 1$ , then  $G(n, p)$  is disconnected with high probability, but the expected number of spanning trees it contains tends to infinity!