ECS 253 / MAE 253
HW5b: Some analytical and semi-analytical tools for generating functions (GFs)

Created by Pierre-André Noël—University of California, Davis

Due Thurs, June 7, 2018

Note: This homework focuses on a specific aspect of generating functions (GFs). To broaden your perspective, you are recommended to read Chapter 1 of generatingfunctionology by Herbert S. Wilf (freely accessible at `http://www.math.upenn.edu/~wilf/DownldGF.html`). You may also try your hand at the exercises of the same chapter, particularly exercises 1–6 and 8. The answers are all available at the end. **This is not part of the homework.**

1 Warming up

Suppose $f(x)$ is an ordinary generating function generating the sequence $(\phi_k)_{k=0}^{\infty}$, i.e.,

$$f(x) := \sum_{k=0}^{\infty} \phi_k x^k. \tag{1}$$

We use the notation $[x^n]f(x)$ to represent the coefficient multiplying x^n in the power series of $f(x)$ (in this case $[x^n]f(x) = \phi_n$). We say that $f(x)$ is a probability generating function (PGF) if the coefficients ϕ_k are to be interpreted as the probability $\mathbb{P}(K = k)$ that the random variable K takes the value k (and thus $0 \leq \phi_k \leq 1$ for all k). We say that a PGF $f(x)$ is normalized if $\sum_{k=0}^{\infty} \phi_k = 1$.

(a) Using proof by induction, show that $[x^n]f(x) = \frac{1}{n!} \frac{\mathrm{d}^n f(x)}{\mathrm{d}x^n}\Big|_{x=0}$

(i.e., show the relation holds for $n = 0$ and $n = 1$, and then for the general case $n + 1$.)

(b) Assuming that $f(x)$ is a normalized PGF, show $f(1) = 1$.

(c) Assuming that $f(x)$ is a normalized PGF, show $\mathbb{E}(K) := \sum_{k=0}^{\infty} k\mathbb{P}(K = k) = \frac{\mathrm{d}\, f(x)}{\mathrm{d}x}\Big|_{x=1}$.

(d) Assuming that $f(x)$ is a normalized PGF, find an expression in terms of derivatives of $f(x)$ for $\mathbb{E}(K^2) := \sum_{k=0}^{\infty} k^2\mathbb{P}(K = k)$.

(e) Suppose that $\sum_{k=0}^{\infty} \mathbb{P}(K = k) = 1 - \mathbb{P}(K \text{ is infinite})$. If $\mathbb{P}(K \text{ is infinite}) > 0$, then the PGF $f(x)$ for the probability distribution $\mathbb{P}(K = k)$ is not normalized. Find the value for $\mathbb{P}(K \text{ is infinite})$. Then find the normalized PGF for the probability sequence $\mathbb{P}(K = k|K \text{ is finite})$.

2 Percolation: some analytical results

An infinite configuration model random graph has its degree distribution specified by $(p_k)_{k=0}^{\infty}$ (i.e., a node sampled uniformly at random has probability p_k to have degree k). In class, you have seen the following two generating functions, respectively for the *degree* of a node and *excess degree* of a node:

$$g_0(x) := \sum_{k=0}^{\infty} p_k x^k \tag{2a}$$

$$g_1(x) = \sum_{k=0}^{\infty} q_k x^k = \frac{g_0'(x)}{g_0'(1)} \qquad \left(\text{with } q_k := \frac{(k+1)p_{k+1}}{\sum_{k'=0}^{\infty} k' p_{k'}}\right). \tag{2b}$$

Here q_k is the probability that a node reached by following a random edge has k *other* edges than the one we followed (and thus a total degree $k+1$). Following the line of reasoning in class we can then obtain the following PGFs

$$h_1(x) := (1-T) + Txg_1\big(h_1(x)\big) \tag{2c}$$

$$h_0(x) := xg_0\big(h_1(x)\big). \tag{2d}$$

The equation you saw in class had $T = 1$. The parameter T here is the same as you saw in the last homework: it may be interpreted as either the probability for an edge of the original configuration model to be present in the percolated one, or as the probability for a spreading process to spread along an edge when it encounters it. Equation (2c) may thus be interpreted as follows: with probability $1-T$, the edge is not followed and there should be no powers of x contributing here, and with probability T the edge is followed normally (hence the term $xg_1\big(h_1(x)\big)$). In summary, the probability of reaching r nodes by following a random edge with probability T is $[x^r]h_1(x)$, and the probability of reaching s nodes by starting at a node selected uniformly at random (and including that node) is $[x^s]h_0(x)$.

In the case $T = 1$, you saw that the network contains a giant component when $g_1'(x) > 1$. In the more general case where $0 \le T \le 1$, there may be no giant component even if $g_1'(x) > 1$. In fact, there is a critical value of T, noted T_c, over which a giant component exists. Hence, for $T \le T_c$, there are only small components and the PGF $g_0(x)$ should thus be normalized. Moreover, the average size of the small components should diverge when $T = T_c$.

(a) Differentiate both sides of Eq. (2c) w.r.t. x and solve for $h_1'(x)$ for the case where T is under threshold, i.e., $h_1(1) = 1$. Find T_c in terms of average degree $\langle k \rangle := \mathbb{E}(K)$ and the second moment $\langle k^2 \rangle := \mathbb{E}(K^2)$.

(b) What is T_c if the degree distribution follows a power law $p_k \propto k^{-\gamma}$ with $\gamma = 2.5$?

(c) Suppose that the highest degree present in the network is 3 (i.e., only p_0, p_1, p_2 and p_3 may be nonzero). Obtain a closed form for $h_0(x)$. You will need the quadratic formula to obtain $h_1(x)$, and recall $h_1(1) = 1$ for $T < T_c$ to decide which of the two roots of the quadratic provides a physically valid answer.

(d) Obtain T_c in the case $p_0 = 0$, $p_1 = 0.2$, $p_2 = 0.5$, $p_3 = 0.3$, and $p_k = 0$ for $k > 3$. Obtain $h_0(1)$ for $T = 0.70$, $T = 0.75$ and $T = 0.80$. What is the size of the giant component (if any) in each of these cases?

3 Percolation: semi-analytical results

(a) Dust-off the code you created for exercise 2(e) of the last homework. Using the parameters `number_simulations`$= 10000$, $\mathbf{n} = (0, 200, 500, 300)$, and version $\mathcal{A}$ (spreading), run your code to estimate the probability distribution for the number s of reached nodes for $T = 0.70$, $T = 0.75$ and $T = 0.80$. In each case, create a log-log graph showing the probability as a function of the number of reached nodes s. Plot each value with a small dot without lines joining them.

(b) Suppose $p_0 = 0$, $p_1 = 0.2$, $p_2 = 0.5$, $p_3 = 0.3$, and $p_k = 0$ for $k > 3$. Use the DFT method with $M = 1001$ (see next page) to extract the coefficients $[x^s]h_0(x)$ from the solution you obtained in exercise 2(c) in the following three cases: $T = 0.70$, $T = 0.75$ and $T = 0.80$. Display your results on the same three log-log plots as in exercise 1, this time using a plain thin line without markers.

NOTE; Usually, you will not have access to such an analytical solution for $h_0(z)$. Fortunately, you can also build a recurrence equation (2c). Indeed, for a given z, you can estimate $h_1(z)$ as using $h_1^{(L)}(z)$ defined as follows: $h_1^{(0)}(z) = 0$, and $h_1^{(L+1)}(z) = (1 - T) + Tzg_1\big(h_1^{(L)}(z)\big)$. You can then estimate $h_0(z)$ using $h_0^{(L)}(z) = zg_0\big(h_1^{(L)}(z)\big)$. **You do not do this here, but our solutions will show how to:**

- **Create a function receiving T, $(p_k)_{k=0}^{K}$ and L and returning $h_0^{(L)}(z)$.**
- **Use the DFT method to extract the coefficients $[z^n]h_0^{(10)}$ for $p_0 = 0$, $p_1 = 0.2$, $p_2 = 0.5$, $p_3 = 0.3$, and $p_k = 0$ for $k > 3$, and in the three cases $T = 0.70$, $T = 0.75$ and $T = 0.80$. Report these results on the same 3 plots as before, this time using a doted line.**
- **Do the same for $[z^n]h_0^{(100)}$, this time using a wide dashed line.**

A stub of what could someday become a PGF tutorial

Pierre-André Noël

April 19, 2012

$$f(x) = \sum_k a_k x^k \tag{1}$$

$$a_n = \frac{1}{n!} \left. \frac{d^n f(x)}{dx^n} \right|_{x=0} \tag{2}$$

$$a_n = \frac{1}{2\pi i} \oint \frac{f(z)}{z^{n+1}} dz = r^{-n} \int_0^1 \mathrm{e}^{-2\pi i n\theta} f\left(r\,\mathrm{e}^{2\pi i\theta}\right) d\theta \tag{3}$$

We could approximate this integral by evaluating $f_m = f\left(r\,\mathrm{e}^{2\pi i \frac{m}{M}}\right)$ along the M equally spaced points $\{f_0, f_1, \ldots, f_{M-1}\}$.

$$a_n \approx \frac{1}{Mr^n} \sum_{m=0}^{M-1} f_m\,\mathrm{e}^{-2\pi i n \frac{m}{M}} \tag{4}$$

Since the f_m do not depend on n, the same points could be used in order to evaluate different a_n. In fact, the sum happens to be a discrete Fourier transform (DFT).

$$\{Ma_0, Mra_1, Mr^2a_2, \ldots, Mr^{M-1}a_{M-1}\} \approx \mathcal{F}^{-1}\{f_0, f_1, f_2, \ldots, f_{M-1}\} \tag{5}$$

Hence, all the $\{a_0, a_1, \ldots, a_{M-1}\}$ may be evaluated in the same pass of fast Fourier transform (FFT) algorithm.[1]

If $f(x)$ is a polynomial of order N such that $N < M$, this relation becomes *exact*: the discrete sum exactly evaluates the integral in the limit $M \to \infty$ and, by the Nyquist-Shannon sampling theorem, the result of the DFT for $0 \leq n \leq N$ should not depend on M when $N < M$. (Hence, a finite M greater than N gives the same result as $M \to \infty$, which is exact.)

When $N \geq M$, aliasing occur: the a_n for $n \geq M$ are "folded back" unto the lower values of n, resulting in errors. However, even when $a_n \neq 0$ in the limit $n \to \infty$, it is often possible to choose a M sufficiently high such that the error is acceptable, provided that $\sum_{n \geq M} a_n$ is small or that we have some idea of the behaviour of a_n for $n \geq M$.

[1] We here suppose that the inverse transform $\mathcal{F}^{-1}$ is defined without normalization, which is the case of the FFTW algorithm as well as most standard libraries. Note, however, that MATLAB applies a factor $\frac{1}{M}$.