1 Introduction

We will give yet another special case of Stein's Method and use it to give a quantitative bound on the convergence of the number of isolated vertices of an Erdős-Rényi graph to a normal distribution. These notes are based on section 3.4 of Nathan Ross's survey *Fundamentals of Stein's Method*.

2 Size-Bias Distribution

Given a nonnegative random variable X with fininte mean μ, we say that X^s has the *size-bias* distribution with respect to X if

$$\mathbb{E}[Xf(X)] = \mu\mathbb{E}[f(X^s)]$$

for every f such that the left-hand side is well-defined. If F and F^s are the CDFs of X and X^s respectively, this condition can be rewritten as

$$\int xf(x)\,dF = \mu \int f(x)\,dF^s,$$

and so the condition is equivalent to $\frac{dF^s}{dF} = \frac{x}{\mu}$. In particular, the distribution of X^s must be absolutely continuous with respect to the distribution of X, and one can read from the Radon-Nikodym derivative that X^s has a bias to higher values, proportional to the size of the value. Hence the name size-bias.

Important Example: The size-bias distribution with respect to any indicator is simply the constant 1.

Now we come to the main theorem of this section:

Theorem 1. *Let $X \geq 0$ be a random variable with $\mu = \mathbb{E}[X] < \infty$ and $\sigma^2 = \mathrm{Var}[X] < \infty$. Let X^s have the size-bias distribution with respect to X. If $W = \frac{X-\mu}{\sigma}$ and $Z \sim \mathcal{N}(0,1)$, then*

$$d_{\mathrm{Was}}(W,Z) \leq \frac{\mu}{\sigma^2}\sqrt{\frac{2}{\pi}}\sqrt{\mathrm{Var}[\mathbb{E}[X^s - X|X]]} + \frac{\mu}{\sigma^3}\mathbb{E}[(X^s - X)^2].$$

Notice that this only requires a second moment assumption (as well as the existence of a suitable size-biased random variable), which is better than the previous special cases of Stein's method we've seen, which required third or even fourth moment assumptions.

I'm not going to give the proof here, you can find it in Ross's notes. It is basically the same idea as the previous special cases of Stein's method. We want to bound the quantity $|\mathbb{E}[f'(W) - Wf(W)]|$ for f bounded with two bounded derivatives, since the supremum of all such quantities is an upper bound for $d_{\mathrm{Was}}(W,Z)$. The idea is to bound the $Wf(W)$ term by Taylor expanding that function around W and plugging in the value $\frac{X^s-\mu}{\sigma}$, and then move things around and obtain the bound.

Notice that if we want this bound to be good, we want X^s to be as closely coupled to X as possible (while still having the size-bias distribution). Here is one way to construct a decently closely coupled size-bias distributed random variable in general, if X is a sum of n nonnegative random variables.

Theorem 2. *Suppose that $X = \sum_{i=1}^n X_i$, where $X_i \geq 0$ and $\mathbb{E}[X_i] = \mu_i$. Here is a construction of a size-bias distributed random variable X^s.*

(1) Choose a random index $I \in \{1,\ldots,n\}$ proportional to μ_i and indpendent of all else.

(2) Sample X_I^s from the size-bias distribution with respect to X_I.

(3) Sample $(X_j)_{j\neq I}$ conditioned on $X_I = X_I^s$.

(4) Let $X^s = \sum_{j\neq I} X_j + X_I^s$, the sum involving the sampled values of the random variables.

Now we have "kicked the can down the road" and we leave the details of the way to couple X_i^s and X_i up to the implementation/application. But once that is given, we can use it to create a size-bias version of X as above. Here is the proof that it actually works.

Proof. Write $\mu = \mathbb{E}[X]$. For any (valid) function f, we have

$$\begin{aligned}
\mathbb{E}[f(X^s)] &= \sum_{i=1}^{n} \frac{\mu_i}{\mu} \mathbb{E}\left[f\left(\sum_{j \neq i} X_j + X_i^s \right) \right] \\
&= \sum_{i=1}^{n} \frac{\mu_i}{\mu} \mathbb{E}\left[\mathbb{E}\left[f\left(\sum_{j \neq i} X_j + X_i^s \right) \middle| X_i^s \right] \right] && (*) \\
&= \sum_{i=1}^{n} \frac{1}{\mu} \mathbb{E}\left[X_i \mathbb{E}\left[f\left(\sum_{j \neq i} X_j + X_i \right) \middle| X_i \right] \right] && (**) \\
&= \frac{1}{\mu} \sum_{i=1}^{n} \mathbb{E}[X_i f(X)] \\
&= \frac{1}{\mu} \mathbb{E}[X f(X)].
\end{aligned}$$

Some care needs to be taken at the step going from $(*)$ to $(**)$. Remember that X_j for $j \neq i$ are conditioned on the value of the last random variable. In line $(*)$ that is X_i^s and in line $(**)$ that is X_i, so the distributions of the X_j are also changing from one step to the next. ∎

3 Application: Isolated Vertices

Theorem 3. *Fix α with $1 \leq \alpha < 2$. Let $G = G(n,p)$ be an Erdős-Rényi graph with $p = \Theta(n^{-\alpha})$, and let X be the number of isolated (degree-zero) vertices in G. Let $\mu = \mathbb{E}[X]$ and $\sigma^2 = \mathrm{Var}[X]$.*

$$d_{\mathrm{Was}}\left(\frac{X - \mu}{\sigma}, Z \right) \leq \frac{C}{\sigma} = \Theta(n^{\frac{1}{2}\alpha - 1})$$

for some constant C independent of n, where $Z \sim \mathcal{N}(0,1)$.

We can write $X = \sum_{i=1}^{n} X_i$, where X_i is the indicator that the ith vertex is isolated. Notice that there is absolutely no independence among the X_i's. They are all positively correlated because if vertex i is isolated then in particular it does not share an edge with vertex j, so this increases the odds for vertex j to be isolated. So the previous results which involved random variables with a sparse dependance graph do not apply here. Nevertheless, we'll be able to apply Theorem 1 to X to obtain the bound indicated in the theorem statement.

Proof. First, we will need a size-biased version of X for which we appeal to Theorem 2. First, choose a random index $I \in \{1, \ldots, n\}$ uniformly (since every X_i has the same mean). Then, since X_I is an indicator, $X_I^s = 1$. So we must sample $(X_j)_{j \neq I}$ conditioned on $X_I = 1$, or in other words conditioned on the Ith vertex being isolated. Since all of the edges in $G(n,p)$ are independent, this means we simply sample $G(n-1,p)$ on the vertices $\neq I$. In other words, X^s should be the number of isolated vertices in G after erasing all edges connected to the Ith vertex, where I is chosen uniformly at random.

In order to apply Theorem 1, we need to compute $\mathbb{E}[X]$, $\mathrm{Var}[X]$, $\mathrm{Var}[\mathbb{E}[X^s - X | X]]$, and $\mathbb{E}[(X^s - X)^2]$. Since any given vertex has probability $(1-p)^{n-1}$ to be isolated, we have

$$\mu = \mathbb{E}[X] = n(1-p)^{n-1} = \Theta(n)$$

by the assumption $p = \Theta(n^{-\alpha})$ which implies that $(1-p)^n$ tends to a finite positive constant.

Since $\mathbb{E}[X_i X_j] = (1-p)^{2n-3}$ for $i \neq j$, we also have

$$\begin{aligned}
\sigma^2 = \mathrm{Var}(X) &= \sum_{i,j=1}^{n} \mathbb{E}[X_i X_j] - \mu^2 \\
&= \sum_{i=1}^{n} \mathbb{E}[X_i] + \sum_{i \neq j} \mathbb{E}[X_i X_j] - \mu^2 \\
&= n(1-p)^{n-1} + n(n-1)(1-p)^{2n-3} - n^2(1-p)^{2n-2} \\
&= n(1-p)^{n-1}\left(1 + (np-1)(1-p)^{n-2}\right) \\
&= \Theta(n^{2-\alpha})
\end{aligned}$$

Now to calculate the terms involving $X^s - X$. Notice that when we remove all edges attached to the Ith vertex, we will increase the number of isolated vertices by one if vertex I is not already isolated. Also, if there were any vertices which were only attached to vertex I, then they will also become isolated. In other words,

$$X^s - X = \mathbf{1}_{\{d_I > 0\}} + D_I$$

where d_i is the degree of the ith vertex, and D_i is the number of vertices connected to the ith vertex which have degree 1. Therefore

$$\begin{aligned}
\mathrm{Var}[\mathbb{E}[X^s - X | X]] &\leq \mathrm{Var}[\mathbb{E}[X^s - X | G]] \\
&= \frac{1}{n^2} \mathrm{Var}\left[\sum_{i=1}^{n} (\mathbf{1}_{\{d_i > 0\}} + D_i)\right] \\
&\leq \frac{2}{n^2}\left(\mathrm{Var}\left[\sum_{i=1}^{n} D_i\right] + \mathrm{Var}\left[\sum_{i=1}^{n} \mathbf{1}_{\{d_i > 0\}}\right]\right).
\end{aligned}$$

Since $\sum_{i=1}^{n} \mathbf{1}_{\{d_i > 0\}}$ is the number of non-isolated vertices, it is $n - X$ and so this sum has the same variance as X, calculated above as $\Theta(n^{2-\alpha})$. As for the first term, notice that $\sum_{i=1}^{n} D_i$ is actually the number of vertices in G with degree 1, since each such vertex is counted exactly one time in the sum. So write this sum as $\sum_{i=1}^{n} Y_i$, where Y_i is the indicator that vertex i has degree 1 in G. Now

$$\begin{aligned}
\mathrm{Var}\left[\sum_{i=1}^{n} Y_i\right] &= \sum_{i,j=1}^{n} \mathbb{E}[Y_i Y_j] - \left(\sum_{i=1}^{n} \mathbb{E}[Y_i]\right)^2 \\
&= \sum_{i=1}^{n} \mathbb{E}[Y_i] + \sum_{i \neq j} \mathbb{E}[Y_i Y_j] - \left(\sum_{i=1}^{n} \mathbb{E}[Y_i]\right)^2 \\
&= n(n-1)p(1-p)^{n-2} + n(n-1)[p(1-p)^{2n-4} + (n-2)^2 p^2 (1-p)^{2n-5}] - n^2(n-1)^2 p^2 (1-p)^{2n-4} \\
&= n(n-1)p(1-p)^{n}[1 - (n-1)p(1-p)^{n-2} + (1-p)^{n-2} + (n-1)^2 p^2 (1-p)^{n-3}] \\
&= \Theta(n^{2-\alpha})
\end{aligned}$$

as well. In total, we obtain that $\mathrm{Var}[\mathbb{E}[X^s - X|X]] = O(n^{-\alpha})$. Finally, we must calculate

$$\begin{aligned}
\mathbb{E}[(X^s - X)^2] &= \mathbb{E}[\mathbb{E}[(X^s - X)^2 | G]] \\
&= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[(D_i + \mathbf{1}_{\{d_i > 0\}})^2] \\
&\leq \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[(D_i + 1)^2] \\
&= \mathbb{E}[D_1^2] + 2\mathbb{E}[D_1] + 1 \\
&\leq 3\mathbb{E}[D_1^2] + 1,
\end{aligned}$$

since D_1 takes nonnegative integer values. Write $D_1 = \sum_{i=2}^n A_i$, where A_i is the indicator that vertex i is connected to vertex 1 and has degree 1. Then

$$\begin{aligned}
\mathbb{E}[D_1^2] &= \sum_{i,j=2}^n \mathbb{E}[A_i A_j] \\
&= \sum_{i=2}^n \mathbb{E}[A_i] + \sum_{i \neq j} \mathbb{E}[A_i A_j] \\
&= (n-1)p(1-p)^{n-2} + (n-1)(n-2)p^2(1-p)^{2n-5} \\
&= \Theta(n^{1-\alpha}).
\end{aligned}$$

Putting it all together, we apply Theorem 1 and get

$$\begin{aligned}
d_{\mathrm{Was}}(W, Z) &\leq \frac{\mu}{\sigma^2}\sqrt{\frac{2}{\pi}}\sqrt{\mathrm{Var}[\mathbb{E}[X^s - X|X]]} + \frac{\mu}{\sigma^3}\mathbb{E}[(X^s - X)^2] \\
&= O\left(\frac{n}{n^{2-\alpha}}\sqrt{n^{-\alpha}} + \frac{n}{n^{3-\frac{3}{2}\alpha}} n^{1-\alpha}\right) \\
&= O\left(\frac{1}{n^{1-\frac{1}{2}\alpha}} + \frac{1}{n^{1-\frac{1}{2}\alpha}}\right) \\
&= O\left(\frac{1}{\sigma}\right).
\end{aligned}$$

This finishes the proof. ■