Lecture 5: Generating Graphs from Graphons, and Converging Back to the Graphon

36-781, Fall 2016

8 November 2016

Contents

1 w-random graphs

Start with a graphon w; let $G_n(w)$ be the n-node random graph generated from it as follows:

- Pick $U_1, \ldots U_n$ IIDly from $Unif(0,1)$
- For each dyad (i, j), pick ξ_{ij} IIDly from $Unif(0,1)$
- Set $A_{ij} = \mathbf{1}(\xi_{ij} \leq w(U_i, U_j))$

2 Homomorphism Densities Have Exponentially Small Deviations

We would like to show that as n grows, $G_n(w)$ converges to w. This is embodied in the following theorem.

Theorem 1 *Fix any motif f and any $\epsilon \in (0,1)$. Then*

$$\Pr\left(|t(f, G_n(w)) - t(f, w)| > \epsilon\right) \leq 2\exp\left\{-\frac{\epsilon^2 n}{4k^2}\right\} \quad (1)$$

so $t(f, G_n(w)) \xrightarrow{P} t(f, w)$.

Remark: Requiring $\epsilon < 1$ isn't a substantive limitation, because the homomorphism density $t(f, \cdot)$ is, as we saw in lecture 3, a probability, and the difference between two probabilities is at most 1. But the restriction does play a role in the proof (§2.3.1).

PROOF: Lengthy when spelled out. We'll need to combine two types of components:

1. A general exponential-deviation inequality for functions of independent random variables, the **bounded-difference inequality** (§2.1).

2. Results on the expected homomorphism density (§2.2).

§2.3 puts these together to finish the proof.

2.1 The Bounded-Difference Inequality

This was mentioned in the notes for Lecture 1.

Proposition 1 *Suppose $Z_1, Z_2, \ldots Z_n \equiv Z$ are independent (not necessarily IID) random variables, and f is any real-valued function of them. We say that f has the bounded-difference property if $|f(z) - f(z')| \leq c$ when z and z' differ in only one coordinate. Then*

$$\Pr(|f(Z) - \mathbb{E}[f(Z)]| > \epsilon) \leq 2\exp\left\{-\frac{\epsilon^2}{2c^2 n}\right\} \quad (2)$$

REFERENCE: Boucheron *et al.* (2013, Theorem 6.2, p. 171).

This is a key result in the field of deviation and concentration inequalities, and accordingly has lots of variations and refinements (many of them covered in Boucheron *et al.* 2013). Notably, if f is more sensitive to changes in some coordinates than others, the constant $c^2 n$ can be improved; also, the requirement that the Z_i be independent can be weakened, at the cost of further assumptions on f (e.g., van de Geer 2002).

2.1.1 Independent Variables in the w-Random Graph

To apply the bounded-difference inequality, we need to divide the random variables used to generate the w-random graph into ones where tweaking each one has a bounded influence on $t(f, G_n(w))$. Here is one: $Z_1 = U_1$; $Z_2 = (U_2, \xi_{12})$; $\ldots Z_n = (Z_n, \xi_{1n}, \xi_{2n}, \ldots \xi_{nn-1})$. These Zs are independent, though not identically distributed.

2.1.2 Deviation Inequality for $t(f, G_n(w))$

We assert (Exercise 2) that changing any one of the Z_i can change $t(f, G_n(w))$ only by at most k/n. Thus

$$\Pr\left(\left|t(f, G_n(w)) - \mathbb{E}\left[t(f, G_n(w))\right]\right| \geq \eta\right) \leq 2\exp\left\{-\frac{\eta^2 n}{2k^2}\right\} \tag{3}$$

Clearly, we're making progress towards our desired theorem, but we're going to have to relate $\mathbb{E}\left[t(f, G_n(w))\right]$ to $t(f, w)$.

2.2 Expected Homomorphism Density

It is fairly straight-forward to verify that (Exercise 3)

$$\mathbb{E}\left[t_{injective}(f, G_n(w))\right] = t(f, w) \tag{4}$$

and we already know (from lecture 4) that

$$\left|t_{injective}(f, w) - t(f, w)\right| < \frac{k^2}{2n} \tag{5}$$

for any w (including one which comes directly from a graph). Therefore

$$\left|\mathbb{E}\left[t(f, G_n(w))\right] - t(f, w)\right| < \frac{k^2}{2n} \tag{6}$$

(Can you explain why we can't just show $\mathbb{E}\left[t(f, G_n(w))\right] = t(f, w)$?)

2.2.1 Connecting Expected and Limiting Homomorphism Densities

Just use the triangle inequality:

$$\begin{aligned}
&|t(f, G_n(w)) - t(f, w)| \\
&\leq \left|t(f, G_n(w)) - \mathbb{E}\left[t(f, G_n(w))\right]\right| + \left|\mathbb{E}\left[t(f, G_n(w))\right] - t(f, w)\right| && (7) \\
&\leq \left|t(f, G_n(w)) - \mathbb{E}\left[t(f, G_n(w))\right]\right| + \frac{k^2}{2n} && (8)
\end{aligned}$$

2.3 Final Steps in the Proof of Theorem 1

To keep $t(f, G_n(w))$ close to $t(f, w)$ while using the triangle inequality, we need to ensure that $t(f, G_n(w))$ is extra close to its expectation. The difference is controlled by k^2/n, so we need to somehow relate that to ϵ.

2.3.1 Relating k^2/n to ϵ

The relationship comes from wanting the inequality in Eq. 1 to be non-trivial, i.e., to bound the probability by a number < 1. This requires:

$$\begin{aligned} 2\exp\left\{-\frac{\epsilon^2}{4k^2}\right\} &< 1 && (9)\\ -\frac{\epsilon^2 n}{4k^2} &< \ln 1/2 && (10)\\ \frac{\epsilon^2 n}{4k^2} &> \ln 2 && (11)\\ \frac{k^2}{n} &< \frac{\epsilon^2}{4\ln 2} && (12)\end{aligned}$$

Now, notice that the theorem only claims to apply to $\epsilon < 1$. But then $\epsilon^2 < \epsilon$, and we have

$$\frac{k^2}{n} < \frac{\epsilon}{4\ln 2} \tag{13}$$

and (for later use)

$$\frac{k^2}{2n} < \frac{\epsilon}{8\ln 2} \tag{14}$$

2.3.2 Final Steps

Using the triangle inequality (Eq. 8), we see that if we want $|t(f, G_n(w)) - t(f, w)| < \epsilon$, we'll need $\eta < \epsilon - k^2/2n$ in Eq. 3. But, by Eq. 14, this is equivalent to requiring

$$\eta < \epsilon\left(1 - \frac{1}{8\ln 2}\right) \tag{15}$$

Now, plugging in to Eq. 3,

$$\Pr\left(|t(f, G_n(w)) - t(f, w)| > \epsilon\right) \leq 2\exp\left\{-\frac{\epsilon^2\left(1 - \frac{1}{8\ln 2}\right)^2 n}{2k^2}\right\} \tag{16}$$

The numerical constant $\left(1 - \frac{1}{8\ln 2}\right)^2 \approx 0.8197$, which, in the denominator, is a factor of ≈ 1.22; rounding twice this to 4 is simplifying and harmless[1]. This completes the proof of Eq. 1, and so of Theorem 1. □

[1] See §4 on the possibility of improving the constant.

2.4 Strengthening Theorem 1 to Almost-Sure Convergence

Corollary 1 *Under the conditions of Theorem 1,* $t(f, G_n(w)) \stackrel{a.s.}{\to} t(f, w)$.

PROOF: The probabilities in the theorem are summable, so, by the Borel-Cantelli lemma, $|t(f, G_n(w)) - t(f, w)| > \epsilon$ only finitely often with probability 1. Taking $\epsilon = 2^{-m}$, so we have a *countable* sequence of ϵs tending to zero, we have

$$\Pr\left(\bigcap_{m=0}^{\infty} \left\{|t(f, G_n(w)) - t(f, w)| > 2^{-m} \text{finitely often}\right\}\right) = 1 \quad (17)$$

and so we have convergence almost surely. □

(This pattern of argument is extremely common; it's one reason why exponentially-small deviation probabilities are important.)

The next theorem is actually also a corollary, but the conclusion is important enough to merit a more dignified name.

Theorem 2 *With probability 1, all homomorphism densities* $t(f, G_n(w))$ *converge on* $t(f, w)$*, i.e.,* $G_n(w) \stackrel{a.s.}{\to} w$.

Comment: The difference between this theorem and the previous corollary is about the quantifiers go. The corollary says (in symbols)

$$\forall f \ \left(\Pr\left(t(f, G_n(w)) \to t(f, w)\right) = 1\right) \quad (18)$$

while the theorem says

$$\Pr\left(\forall f \ \left(t(f, G_n(w)) \to t(f, w)\right)\right) = 1 \quad (19)$$

which is stronger.

PROOF of Theorem 2: As is often the case when proving that something happens with probability 1, we'll show that the complementary event, call it B for "bad", has probability 0. B is the event where not all of the homomorphism densities converge. Writing B_f for the failure-to-converge event for motif f, then clearly $B \subseteq \bigcup_f B_f$. (Can you strengthen this to $B = \bigcup_f B_f$?) Since there is only a countable infinity of motifs (Exercise 1),

$$\Pr(B) \leq \Pr\left(\bigcup_f B_f\right) \leq \sum_f \Pr\left(B_f\right) \quad (20)$$

By Corollary 1, $t(f, G_n(w)) \stackrel{a.s.}{\to} t(f, w)$ for each f, so $\Pr\left(B_f\right) = 0$ for each f. Thus

$$0 \leq \Pr(B) \leq 0 \quad (21)$$

and $\Pr(B) = 0$. Therefore the complementary event has probability 1. □

Remark: Each individual convergence $t(f, G_n(w))$ has probability 1, but to get the union of all of them to have probability 1, it's important that we're only taking a *countable* union. Probabilities add over countable unions, but not over uncountable ones[2]. In other branches of probability, similarly, we often want to show convergence-as-a-whole by finding a "convergence-determining class" of events or test functions which is merely countably infinite.

3 Mean-square convergence and variance bound

We've just seen that $t(f, G_n(w))$ converges in probability and almost surely to $t(f, w)$. We can also show that it converges in mean-square, i.e., that $t(f, G_n(w)) \xrightarrow{L_2} t(f, w)$. That is, we show that

$$\mathbb{E}\left[\left(t(f, G_n(w) - t(f, w)\right)^2\right] \to 0 \tag{22}$$

As usual, there's a bias-variance decomposition; the left-hand side is

$$\left(\mathbb{E}\left[t(f, G_n(w)\right] - t(f, w)\right)^2 + \mathbb{V}\left[t(f, G_n(w)\right] \tag{23}$$

We saw, in Eq. 6, that $\mathbb{E}\left[t(f, G_n(w))\right] \to t(f, w)$, so it remains to control the variance.

Proposition 2 $\mathbb{V}\left[t(f, G_n(w))\right] \leq 3k^2/n$.

PROOF: The variance is of course $\mathbb{E}\left[t^2(f, G_n(w))\right] - \mathbb{E}\left[t(f, G_n(w))\right]^2$. Let's start with the square of the expectation.

$$\begin{aligned}
\mathbb{E}\left[t(f, G_n(w))\right]^2 \quad & geq \quad \left(\mathbb{E}\left[t_{injective}(f, G_n(w))\right] - \frac{k^2}{2n}\right)^2 & (24)\\
& \geq \quad \mathbb{E}\left[t_{injective}(f, G_n(w))\right]^2 - \frac{k^2}{n} & (25)\\
& = \quad t^2(f, w) - \frac{k^2}{n} & (26)
\end{aligned}$$

Now let's turn to the expectation of the square. The key technique is to observe that if a motif f consists of two *disjoint* subgraphs, say that $f = f_1 \oplus f_2$, then $t(f, w) = t(f_1, w)t(f_2, w)$ for any graphon w (Exercise 4). Thus define a graph f^2 which consists of two disjoint copies of f, so that

$$\begin{aligned}
t(f^2, G_n(w)) &= t^2(f, G_n(w)) & (27)\\
t(f^2, w) &= t^2(f, w) & (28)
\end{aligned}$$

[2]This is what allows a uniformly-distributed random number U on the unit interval to have both $\Pr(U = x) = 0$ for any x, i.e., $\forall x\ (\Pr(U \neq x) = 1)$ and $\Pr(U \in [0,1]) = 1$, i.e., $\Pr(\forall x\ (U \neq x)) = 0$. It is still possible to find papers by engineers on arxiv.org which in essence deny this (e.g., arxiv:0606635). On the other hand, this theorem, and the previous corollary, give a hint of how much we'd have to give up if we restricted probability to merely *finite* unions of events, rather than countable ones. Finitely-additive probability does have its advocates, or at least sympathizers, such as our own Profs. Kadane and Seidenfeld, but on balance it's an unattractive idea.

Now instead of finding expectations of $t^2(f, G_n(w))$, we find the expectation of $t(f^2, G_n(w))$ along the lines we just did:

$$\begin{aligned}
\mathbb{E}\left[t^2(f, G_n(w))\right] &= \mathbb{E}\left[t(f^2, G_n(w))\right] && (29)\\
&\leq \mathbb{E}\left[t_{injective}(f^2, G_n(w)) + \frac{(2k)^2}{2n}\right] && (30)\\
&= t(f^2, w) + \frac{2k^2}{n} && (31)\\
&= t^2(f, w) + \frac{2k^2}{n} && (32)
\end{aligned}$$

Subtracting,

$$\begin{aligned}
\mathbb{V}\left[t(f, G_n(w))\right] &\leq t^2(f, w) + \frac{2k^2}{n} - \left(t^2(f, w) - \frac{k^2}{n}\right) && (33)\\
&= \frac{3k^2}{n} && (34)
\end{aligned}$$

Remark: Mean-square convergence implies convergence in probability, by means of Chebyshev's inequality. This gives only a quadratic rate of decay for the deviation probabilities in ϵ, unlike the exponential rate we got using the bounded-difference inequality.

4 Source Notes

The crucial Theorem 1 seems to have first appeared as as Lovász and Szegedy (2006, Theorem 2.5), though with a worse (i.e., larger) constant in the exponent (18 in the denominator instead of 4). The version given here, and the proof, follows Borgs *et al.* (2006, Lemma 4.4). (Cf. also Lovász 2012, Corollary 10.4, which gives the constant as 8.) The constant in the denominator could be improved without changing the proof technique, but not by much (Exercise 5); I don't know whether another approach could deliver a tighter inequality.

The variance bound in Proposition 2, and its proof, is from Lovász and Szegedy (2006, Lemma 2.4c).

5 Exercises

1. Prove that the set of all finite, simple graphs or motifs f is countably infinite, by constructing an explicit enumeration scheme for such graphs, showing that they are in 1-1 correspondence with the natural numbers. Remember that this means not only assigning a unique natural number to each graph, but also checking that every number corresponds to a graph. (If your first cut at numbering scheme doesn't have a graph for each number, try proving that the set of numbers with graphs is itself countably infinite.)

2. Prove that changing one, but only one, of the Z_i can change $t(f, G_n(w))$ by at most k/n.

3. Prove Eq. 4, that the expected injection density is equal to the limiting homomorphism density.

4. Prove that if f consists of two disjoint graphs, $f = f_1 \oplus f_2$, then $t(f, w) = t(f_1, w)t(f_2, w)$. *Hint:* use the integral form of $t(f, w)$.

5. Improve (i.e., reduce) the constant in the denominator of the exponent in Eq. 1.

   (a) Verify that an entirely parallel argument to the one gone through for Eq. 1 would also prove that the deviation probability is $\leq 2\exp\left\{-\epsilon^2 n/8k^2\right\}$.

   (b) Verify that the same technique could *not* prove that the deviation probability is $\leq 2\exp\left\{-\epsilon^2 n/3k^2\right\}$.

   (c) Show that the technique of the proof can establish that the deviation probability is $\leq 2\exp\left\{-\epsilon^2 n/\varkappa k^2\right\}$, so long as $\varkappa \geq 2/(1-1/(2\varkappa \ln 2))^2$, and so that the smallest $\varkappa$ attainable by this method is ≈ 3.2842703.

References

Borgs, Christian, Jennifer T. Chayes, László Lovász, Vera T. Sós, Balázs Szegedy and Katalin Vesztergombi (2006). "Graph Limits and Parameter Testing." In *Proceedings of the 38th Annual ACM Symposium on the Theory of Computing [STOC 2006]*, pp. 261–270. New York: ACM. URL `http://research.microsoft.com/en-us/um/people/jchayes/Papers/TestStoc.pdf`.

Boucheron, Stéphane, Gábor Lugosi and Pascal Massart (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford: Oxford University Press.

Lovász, László (2012). *Large Networks and Graph Limits*. Providence, Rhode Island: American Mathematical Society.

Lovász, László and Balázs Szegedy (2006). "Limits of Dense Graph Sequences." *Journal of Combinatorial Theory B*, **96**: 933–957. URL `http://arxiv.org/abs/math.CO/0408173`. doi:10.1016/j.jctb.2006.05.002.

van de Geer, Sara A. (2002). "On Hoeffding's Inequality for Dependent Random Variables." In *Empirical Process Techniques for Dependent Data* (Herold Dehling and Thomas Mikosch and Michael Sorensen, eds.), pp. 161–169. Boston: Birkhäuser. URL `http://stat.ethz.ch/~geer/hoeffding2.pdf`.