Lecture 2: Data Collection and Network Sampling

36-720, Fall 2016
Scribes: Cristobal De La Maza and Valerie Yuan
`http://www.stat.cmu.edu/~cshalizi/networks/16-1` for updates

31 August 2016

Contents

The real foundation of any branch of statistics is data collection. For the sorts of statistics we've mostly seen before, where data are IID (or IID-ish), data comes from samples or from experiments. It's hard (though not impossible) to experiment with networks, so we mostly have to deal with samples, and even there, things become much more complicated than we're used to. Unfortunately, this complexity is all too often ignored when analyzing empirical networks.

1 Sampling procedures

1.1 Ideal Data: Network Census

The ideal data would be a **census** or **enumeration** of the network. This would record every node, and every edge between nodes, with no spurious additional nodes or edges. If you are in the fortunate situation of having a complete network census, you can pretty much ignore the sampling process, and proceed to model network formation.

1.1.1 Imperfections in network censuses

Unfortunately, even studies which *try* to get a complete census may fall short of perfection. The exact failure modes depend on the nature of the network and indeed on the details of the measurement process; for concreteness, I focus here on survey-based measurements of social networks.

These surveys often work by approaching people and asking them questions like "Who are your friends?" or "From whom do you seek advice?" or "From whom have you borrowed money?" Errors can creep in here in many ways, such as varying understandings of the link (help-you-move friend, or help-you-move-a-body friend?). There can be different results depending on whether are given suggestions, or a checklist of possibilities, or are asked to spontaneously recall names. Answers may be influenced by shame, boastfulness[1], or other emotions related to the "presentation of self in everyday life" (Goffman, 1959). In older studies, it was common to frame the question as something like "name up to three colleagues you commonly go to for advice"; such **censoring by degree** necessarily prevented any recorded out-degree from being higher than three.

In, say, studies of protein interaction networks, the specific *causes* of measurement error are different — other proteins are not, presumably, ashamed to admit that they bind to cytochrome C — but there are others, which can lead both to false negatives and false positives. For a serious applied study, there is no real substitute for detailed knowledge of the domain, and of how the data is actually obtained, which is one reason the idea of the statistician as a "consultant" giving a week to a project, let alone as a "data scientist" treating the data as just another database, is so pernicious.

[1] An amusing example is asking people about the number of their heterosexual sex partners. The *mean* number must, necessarily, be equal for men and for women. (This is not true of the median; why?) However, in every study of this sort known to me, the mean number reported by men is substantially higher than the mean number reported by women.

2 Sampling Designs

If we cannot get hold of the true, "population" graph $G = (V, E)$, we may, guided by the example of IID statistics, try to measure a "sampled" graph $G^* = (V^*, E^*)$, with $V^* \subseteq V$ and $E^* \subseteq E$. Different **sampling designs** amount to different ways of obtaining such sampled subgraphs[2]. In baby statistics, our first step in understanding sampling is the concept of a **simple random sample** (SRS) of units from the population. In networks, even a simple random sample is complicated.

2.1 Induced and Incident Subgraphs

We could start with a simple random sample of *nodes*, i.e., V^* is an SRS of V. We'd then take the induced subgraph[3], i.e., $(i, j) \in E^*$ iff $(i, j) \in E$, $i \in V^*$, and $j \in V^*$. This natural procedure, **induced subgraph sampling**, turns out be very biased for even very simple network statistics, though the biases can sometimes be calculated and compensated for[4].

On the other hand, we start with a simple random sample of *edges*, i.e., E^* is an SRS of E. We'd then take the nodes which are **incident** on those edges, i.e., $i \in V^*$ if, for some $j \in V$, $(i, j) \in E^*$. Experience with conventional surveys may make incident-subgraph sampling seem odd, but there are many situations where it's actually quite natural — imagine sampling a fraction of all phone calls made through a cell-carrier (where nodes = phone numbers), or financial transactions.

Starting from the same G, induced-subgraph and incident-subgraph sampling lead to *very* different distributions over G^*, even if we adjust the sample sizes so that (e.g.) the average number of edges in G^* are comparable.

2.1.1 Example of a Bias from Induced-Subgraph Sampling

The canonical example of how sampling can induce a bias, even when we're just doing a simple random sample of nodes, is the mean degree. Intuitively, we don't see any edges *outside* the induced subgraph, so the degree we record for each node is at most its real degree, and the mean degree in the sampled graph should be $\leq$ the true mean degree. We can be more precise, and say by how much it's lower.

Notation: say k_i is the degree of node i, so $k_i = \sum_{j=1}^{n} A_{ij}$. The mean degree over the whole network is $\overline{k} = n^{-1} \sum_{i=1}^{n} k_i$, or $\overline{k} = n^{-1} \sum_i \sum_j A_{ij}$.

Now take a simple random sample of m nodes, so the probability of seeing node i is the same for all nodes, $\pi = m/n$. We'll write $Z_i = 1$ if node i is in the sample, and $Z_i = 0$ otherwise, i.e., Z_i is the indicator for $i \in V^*$. The observed

[2] For simplicity, we're ignoring the interaction of sampling and measurement error, but of course both can be present together.

[3] See Lecture 1 for the concept of an "induced subgraph"

[4] See problem 2 in homework 1 for examples of calculating biases due to induced-subgraph sampling.

graph G^* has an observed adjacency matrix A^*, and $A^*_{ij} = 1$ iff $A_{ij} = 1$ and both i and j are in the the sample. What's the expected value of the plug-in estimate $\overline{k}$ from G^*, say $\overline{k}^*$?

$$
\begin{aligned}
\mathbb{E}\left[\overline{k}^*\right] &= \mathbb{E}\left[\frac{1}{m}\sum_{i\in V^*} k_i^*\right] && (1)\\
&= \mathbb{E}\left[\frac{1}{m}\sum_{i\in V^*}\sum_{j\in V^*} A_{ij}^*\right] && (2)\\
&= \mathbb{E}\left[\frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij} Z_i Z_j\right] && (3)\\
&= \frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\mathbb{E}\left[Z_i Z_j\right] && (4)\\
&= \frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\pi^2 && (5)\\
&= \frac{1}{n\pi}\pi^2\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij} && (6)\\
&= \frac{\pi}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij} = \pi\overline{k} && (7)
\end{aligned}
$$

Here, the key step is replacing the expectation of the indicator variable with the sampling probability[5]. Since π is a sampling probability, and there < 1, we have that the mean degree of the induced subgraph is less than the true mean degree, by a factor of π.

2.2 "Exploratory" Sampling Designs

For both induced- and incident- subgraph sampling, the sampling frame is in some sense separate from the actual, realized graph: the population from which we draw our SRS has to include all nodes, or all edges, but doesn't use the graph beyond that. Other designs, which do make use of the graph topology, are however certainly possible, and common.

In **egocentric** designs, we sample nodes and record information about their local neighborhoods, or **ego networks**. Sometimes this is as simple as their in- and out- degrees (or degree, in undirected graphs). Other times we record edges and non-edges among the neighbors of the initial node ("ego"); this is

[5] In fact, I cheated a little; $\mathbb{E}\left[Z_i\right] = \pi$, but $\mathbb{E}\left[Z_i Z_j\right]$ would only equal π^2 if Z_i and Z_j are uncorrelated. But if we sample *exactly* m nodes, without replacement, Z_i and Z_j are (negatively) correlated. However, the correlation goes to zero as n, and you can work out the the corrections if you really want to.

sometimes called a **star** design, though I have certainly also heard it called an "egocentric design". When we deal with star designs, we collect multiple local graph neighborhoods, and an important question is whether those overlap; depending on the recording process, this information might be available (so we realized that Karl appears in the ego networks of both Irene and Joey) or not.

2.2.1 Snowball sampling

A natural extension of egocentric designs is **snowball sampling** (Goodman, 1961)[6] — known in the theory of algorithms as "breadth-first search". Here we start with a **seed** node[7], and record its immediate neighborhood. (So far, this is just a star design.) We then repeat the process for each of the neighbors, and then their neighbors, etc., etc., until either no new nodes are found or we get tired, i.e., a pre-selected size (in terms either of the number of nodes or the number of layers around the seed) is reached. Of course, there can be multiple seeds; there may then be an issue of determining when two snowballs which have formed around different seeds have over-lapped.

Snowball sampling leads to a different distribution over graphs than does either induced- or incident- subgraph sampling. Even if the seed is chosen by a simple random sample, the other nodes picked up by the snowball are *not* a random sample. Since they are nodes which can be reached by following paths from the seed, they must have degree at least 1, must be at least weakly connected to the seed, and in general tend to have higher-than-average degree.

2.2.2 Respondent-driven sampling

An important variant on snowball sampling, for social networks, is **respondent-driven sampling**. This originated as a way of studying members of hard-to-find ("hidden") sub-populations — often ones which were hidden because membership in them is stigmatized or illegal, such as intravenous drug users. The technique is to find some initial members of the group in question, and then persuade them to recruit other members whom they know as research subjects. Often, the respondents are given unique physical tokens to pass on those whom they recruit, so that links can be traced, and there may be some incentive for participation. Censoring by degree can result if, for instance, there is only a limited number of physical tokens per respondent.

2.2.3 Trace-route sampling

Trace-route sampling probes a network by, as the name suggests, tracing routes through it. The typical procedure goes as follows:

[6] Goodman (1961) did not introduce snowball sampling — the paper in fact refers to earlier work in sociology — but I have not read the earlier papers, and this is the oldest statistical analysis of the technique which I know of.

[7] Snowballs do not generally start with seeds, but the mathematical sciences are not the place to look for consistent metaphors.

1. Pick a set of source nodes.
2. Pick a set of target nodes.
3. For each source-target combination, find a path from the source to the target, and record all nodes and edges traversed along the path.

Clearly, a lot will depend on how, precisely, paths are found, but this is an application-specific issue.

While the name "trace-route" comes from a Unix utility which finds paths across the Internet, the practice is actually older, and not just limited to computer networks. The most famous instance of trace-route sampling is probably the Milgram study which led to the folklore that any two people in the US have at most "six degrees of separation". In that study, sources in the midwest were tasked to get an envelope to a target, a stock-broker living outside Boston, where at each step the envelope had to be passed on to someone known on a first-name basis[8]. Dodds *et al.* (2003) was a comparatively recent attempt to do something similar but with e-mail.

Depending on exactly how route-tracing gets done, one may or may not get information from "failed" routes, i.e., ones which didn't succeed in getting from source to target; that's long been appreciated. What was not realized until Achlioptas *et al.* (2005) is that trace-route sampling systematically distorts the degree distribution, making all kinds of graphs look like they have heavy-tailed distributions whether they do or not.

3 Coping strategies

Sampling issues with networks are real and potentially affect all aspects of inference — just like every other part of statistics. Network data analysis has therefore developed several strategies for coping with sampling problems.

3.1 Head in sand

That is, ignore distortions or biases due to sampling, and pretend that the graph we see is the whole graph. This is generally not a good idea.

For induced-subgraph sampling, the mean degree is biased from the real degree by a calculable factor. Indeed, the sample values of motif counts for all motifs are also biased (again, in calculable ways). These would be pretty easy to compensate for. But degree distribution, for example, gets distorted in very complicated, hard-to-fix ways, even with induced-subgraph sampling (Stumpf and Wiuf, 2005; Stumpf *et al.*, 2005; Lee *et al.*, 2006). As noted above, for trace-route sampling, in particular, the degree distribution of pretty much any graph ends up heavy tailed (Achlioptas *et al.*, 2005).

[8] See Kleinfeld (2002) for a detailed examination of *exactly* how the study was conducted, and what it showed.

3.2 Learn sampling theory (design based inference)

Classical sampling theory is a theory of statistical inference in which probability assumptions are only made about the sampling process. The true population is regarded as unknown but fixed, and no stochastic assumptions are made about how it is generated. (One can always regard this as conditioning on the unknown population.) Because all the probability assumptions refer to the sampling design, and the validity of the inference depends only on whether the design has been accurately modeled, this is sometimes called **design-based inference**.

As an example of how this works, consider trying to estimate the mean μ of some quantity X_i over a finite population of size n, using a sample of units S. If every unit is sampled with equal probability, the familiar sample mean $\overline{X} = |S|^{-1} \sum_{i \in S} X_i$ is a good estimate. With unequal probabilities, however, what should one do?

A simple, classic solution is the **Horvitz-Thompson** estimator:

$$\hat{\mu}_{HT} \equiv \frac{1}{n} \sum_{i \in S} \frac{X_i}{\pi_i} \tag{8}$$

where π_i is the (assumed-known) **inclusion probability** of unit i, i.e., the probability of unit i being included in the sample[9]. Notice that if all inclusion probabilities are equal, $\pi = |S|/n$, we get back the sample mean $\overline{X}$.

The intuition here is that if we saw one unit with inclusion probability π_i, there are probably about $1/\pi_i$ others that we didn't see. More formally, we can show that this is an unbiased estimator. To see this, let's work out the expectation of $\hat{\mu}_{HT}$, introducing indicator variables Z_i, $i \in 1 : n$, which are 1 if $i \in S$ and 0 otherwise.

[9] Using $1/\pi_i$ here is an example of a general trick in estimation known as **inverse probability weighting**.

$$
\begin{aligned}
\mathbb{E}\left[\hat{\mu}_{HT}\right] &= \mathbb{E}\left[\frac{1}{n}\sum_{i\in S}\frac{X_i}{\pi_i}\right] && (9)\\
&= \mathbb{E}\left[\frac{1}{n}\sum_{i\in 1:n}\frac{X_i}{\pi_i}Z_i\right] && (10)\\
&= \frac{1}{n}\sum_{i\in 1:n}\frac{X_i}{\pi_i}\mathbb{E}\left[Z_i\right] && (11)\\
&= \frac{1}{n}\sum_{i\in 1:n}\frac{X_i}{\pi_i}\mathbb{P}\left(Z_i = 1\right) && (12)\\
&= \frac{1}{n}\sum_{i\in 1:n}\frac{X_i}{\pi_i}\pi_i && (13)\\
&= \frac{1}{n}\sum_{i\in 1:n}X_i && (14)\\
&= \mu && (15)
\end{aligned}
$$

One can show (Exercise 2a) that the variance of the estimator is

$$
\mathrm{Var}\left[\mu_{\hat{HT}}\right] = \frac{1}{n^2}\sum_{i\in 1:n}\sum_{j\in 1:n}X_i X_j\left(\frac{\pi_{ij}}{\pi_i\pi_j} - 1\right) \tag{16}
$$

with π_{ij} being the joint inclusion probability, i.e., the probability of including both i and j in the sample (with $\pi_{ii} = \pi_i$). Notice that if all the $\pi_i \to 1$, the variance goes to 0, as is reasonable (and as is required if the estimator is to be consistent). We can't actually calculate this true variance, since we can't sum over all the unknown units in the population, but there is a (consistent) empirical counter-part:

$$
\widehat{\mathrm{Var}}\left[\hat{\mu}_{HT}\right] = \frac{1}{n^2}\sum_{i\in S}\sum_{j\in S}X_i X_j\left(\frac{1}{\pi_i\pi_j} - \frac{1}{\pi_{ij}}\right) \tag{17}
$$

Horvitz-Thompson is a basic tool of design-based inference, but hardly the only one; this deliberately barely scratches the surface.

3.2.1 Strengths and Weaknesses

The sampling-theory approach works well for stuff you can express as averages (or totals) of population quantities, *and* where you can work out inclusion probabilities from knowledge of the sampling design. Many network statistics can be expressed as averages (sometimes by defining the "unit" as, e.g., a dyad of nodes), but exact calculation of inclusion probabilities is harder. Kolaczyk (2009) collects many examples.

3.3 Missing data tools

Another approach is to treat the unobserved part of the network as missing data, and try to infer it. This can range from simple imputation strategies, to complex model-based strategies for inference, such as the EM algorithm. Successful imputation or EM is not design-based but model-based, and requires a model both of the network, and of the sampling process. It is very, very rare for anything to be "missing at random", let alone "missing completely at random". Perhaps for this reason, comparatively little has been done on this direction (Handcock and Gile, 2010)[10]; more should be.

3.4 Model the Effective Network

A final strategy is to model the *observed* network. This means modeling both the observation/sampling process and the actual network, but combining them so that we get a family of probability distributions over the observed graph. That observed network is (or can be) still informative about the parameters of the underlying generative model. If *that* is all that's of interest, it may be possible to short-circuit the use of EM or imputation, which are more about recovering the full graph.

I have seen very few examples of this applied to networks, but it's actually rather common for univariate data, which is often recorded in grouped or binned form. One *could* try to tackle that by EM, but it's often much more straightforward to derive (if only numerically) a likelihood for the binned data, and then estimate parameters of the unbinned distribution based on that (e.g., Virkar and Clauset 2014). I think this is a seriously under-explored topic for network modeling.

4 Big Data Solves Nothing

Even when, as the promoters say, "n = all", and the data are automatically recorded (voluntarily or involuntarily), almost all the network sampling issues we've gone over remain. After all, as the promoters do *not* say, you're getting all of a biased convenience sample, not all of the truth[11]. Three issues are particularly prominent for networks: entity resolution, diffusion, and performativity.

Entity resolution, or **record linkage**, is a pervasive problem for data analysis. Generally speaking, it's the problem of determining when multiple data points all record information about the same thing (or records which are apparently co-referent really are about different things). In networks, this is usually about determining when two (or more) apparent nodes really refer to the same underlying entity. In building collaboration networks, for instance, one has to determine whether "Mark Newman", "M. E. J. Newman" and "M.

[10] Studies demonstrating the importance of not ignoring the fact that some data are missing include Borgatti *et al.* (2006); Kossinets (2006).

[11] I know I borrowed the phrase "big data is a biased convenience sample" from someone, but I can't recall who.

Newman" refer to the same researcher (or, rather, when they do); in citation networks, papers may be cited differently; the same person may have multiple phone numbers; etc., etc. Network structure can in fact be an important clue in entity resolution (Bhattacharya and Getoor, 2007), but that gets a bit circular...

Diffusion refers to the way that many of the automatically-recorded networks which provide us with our big data have themselves *spread* over other, older social networks. What we see when we look at the network of (say) Facebook ties is a combination of the pre-Facebook social network and the results of the diffusion process. Comparatively little has been done to understand the results. One of the best studies is that of Schoenebeck (2013), who showed how even if the diffusion process treats all nodes homogeneously, the network-as-diffused can differ radically in its properties from the underlying network. If you say "I only care about Facebook, not about the social network", this may not matter, but even then it can change your understanding of why Facebook looks the way it does.

The third issue is **performativity**, the way theories can become (partially) self-fulfilling prophecies. The companies which run online social networks are all very invested in getting very big, very dense networks of users[12]. This is why they all offer link suggestion or link recommendation services. The algorithms behind these recommendations implement theories about how social networks form, and what sort of link patterns they should have. To the extent that people follow these recommendations, then, the recorded network will seem to conform to the theory. The only worthwhile study of this I know of is Healy (2015).

5 Exercises

To think through, not to hand in.

1. *Induced-subgraph sampling and degree* The **density** of a graph is defined as the ratio between the number of edges it has, and the maximum number it could have, i.e., the ratio $\frac{\sum_i k_i}{\binom{n}{2}}$ for a simple undirected graph. §2.1.1 showed that the mean degree is biased when we take a random sample of nodes. Modify the same argument to show that the *density* is unbiased.

2. *Horvitz-Thompson*

[12] This is because they want to lock you in to using their network. Economically, the amount they can charge a user, either monetarily or in hassle, time spent watching ads, etc., is just less than the "switching cost" to the user of changing over to a different network. Given a choice between two equally-functional websites, users will typically prefer ones with more of their friends / contacts / peers, so switching costs will increase as the network gets larger and denser. (Said slightly differently, large groups of users would have to *coordinate* switching to a different service, and coordination itself creates switching costs.) The economics of lock-in, switching costs and network externalities were all laid out very lucidly long ago by Shapiro and Varian (1998) — from the point of view of the companies running the networks, not of the users who are, as the saying goes, the product being sold. (Note that Varian is now the chief economist at Google.)

(a) Derive the variance of $\hat{\mu}_{HT}$. *Hint:* Repeat the trick with the Z_i indicators used to show $\hat{\mu}_{HT}$ is unbiased.

(b) To use Eq. 8, we need to know n, the total population size. Suppose we replace n by $\sum_{i \in S} 1/\pi_i$. Will this still be an unbiased estimate? Will the variance be larger or smaller than that of Eq. 8?

3. *Degree-proportional sampling* Write the degree distribution of your favorite graph in terms of its probability mass function as p, i.e., the fraction of nodes of degree k is p_k.

(a) Suppose that nodes are included in the sample with a probability proportional to their degree. Find the

(b)

References

Achlioptas, Dimitris, Aaron Clauset, David Kempe and Cristopher Moore (2005). "On the Bias of Traceroute Sampling (or: Why almost every network looks like it has a power law)." In *Proceedings of the 37th ACM Symposium on Theory of Computing*. URL `http://arxiv.org/abs/cond-mat/0503087`.

Bhattacharya, Inrajit and Lise Getoor (2007). "Collective Entity Resolution In Relational Data." *ACM Transactions on Knowledge Discovery from Data*, **1(1)**: 5. doi:10.1145/1217299.1217304.

Borgatti, Stephen P., Kathleen M. Carley and David Krackhardt (2006). "On the robustness of centrality measures under conditions of imperfect data." *Social Networks*, **28**: 124–136. doi:10.1016/j.socnet.2005.05.001.

Dodds, Peter Sheridan, Roby Muhamad and Duncan J. Watts (2003). "An Experimental Study of Search in Global Social Networks." *Science*, **301**: 827–829. doi:10.1126/science.1081058.

Goffman, Erving (1959). *The Presentation of Self in Everyday Life*. New York: Anchor Books.

Goodman, Leo A. (1961). "Snowball Sampling." *Annals of Mathematical Statistics*, **32**: 147–170. doi:10.1214/aoms/1177705148.

Handcock, Mark S. and Krista J. Gile (2010). "Modeling Social Networks from Sampled Data." *Annals of Applied Statistics*, **4**: 5–25. URL `http://arxiv.org/abs/1010.0891`.

Healy, Kieran (2015). "The Performativity of Networks." *European Journal of Sociology*, **56**: 175–205. URL `http://kieranhealy.org/files/papers/performativity.pdf`. doi:10.1017/S0003975615000107.

Kleinfeld, Judith (2002). "Could It Be a Big World After All? What the Milgram Papers in the Yale Archive Reveal About the Original Small World Study." *Society*, **39**: 61–66. URL http://www.uaf.edu/northern/big_world.html.

Kolaczyk, Eric D. (2009). *Statistical Analysis of Network Data*. New York: Springer-Verlag.

Kossinets, Gueorgi (2006). "Effects of Missing Data in Social Networks." *Social Networks*, **28**: 247–268. URL http://arxiv.org/abs/cond-mat/0306335. doi:10.1016/j.socnet.2005.07.002.

Lee, Sang Hoon, Pan-Jun Kim and Hawoong Jeong (2006). "Statistical Properties of Sampled Networks." *Physical Review E*, **73**: 016102. URL http://arxiv.org/abs/cond-mat/0505232. doi:10.1103/PhysRevE.73.016102.

Schoenebeck, Grant (2013). "Potential Networks, Contagious Communities, and Understanding Social Network Structure." In *Proceedings of the 22nd International World Wide Web Conference [WWW 2013]* (Daniel Schwabe and Virgilio Almeida and Hartmut Glaser and Ricardo Baeza-Yates and Sue Moon, eds.), pp. 1123–1132. Geneva, Switzerland: International World Wide Web Conferences Steering Committee. URL http://arxiv.org/abs/1304.1845.

Shapiro, Carl and Hal R. Varian (1998). *Information Rules: A Strategic Guide to the Network Economy*. Boston: Harvard Business School Press, 1st edn.

Stumpf, Michael P. H. and Carsten Wiuf (2005). "Sampling Properties of Random Graphs: The Degree Distribution." *Physical Review E*, **72**: 036117. URL http://arxiv.org/abs/cond-mat/0507345.

Stumpf, Michael P. H., Carsten Wiuf and Robert M. May (2005). "Subnets of Scale-free Networks are not Scale-free: Sampling Properties of Networks." *Proceedings of the National Academy of Sciences (USA)*, **102**: 4221–4224. doi:10.1073/pnas.0501179102.

Virkar, Yogesh and Aaron Clauset (2014). "Power-law distributions in binned empirical data." *Annals of Applied Statistics*, **8**: 89–119. URL http://arxiv.org/abs/1208.3524. doi:10.1214/13-AOAS710.