Lecture 4: Applications: random trees, determinantal measures and sampling

Robin Pemantle

University of Pennsylvania

pemantle@math.upenn.edu

Minerva Lectures at Columbia University

09 November, 2016

Sampling

Sampling problem

The most common sampling problem is how to sample from a distribution on a finite but exponentially large set.

The problem is algorithmic.

Often some type of approximation to the distribution will suffice.

Markov Chain Monte Carlo is a major tool.

See also: Metropolis algorithm; importance sampling; simulated annealing; etc.

π-p-s sampling

I will discuss a different kind of sampling problem.

We wish to sample a subset S of $[n]$ of size k, such that the probabilities $\mathbb{P}(i \in S)$ are equal to given numbers $p_1, \ldots, p_n$ summing to k. Often we extend this to $\sum_{i=1}^n p_i \neq k$ via

$$\mathbb{P}(i \in S) = \frac{kp_i}{\sum_{j=1}^n p_j} .$$

This problem is known as π-p-s sampling (apparently this stood for "probability proportional to size").

π-p-s sampling with negative correlations

When the sample is used for estimation, it is desirable that the events $\{i \in S\}$ and $\{j \in S\}$ be nonpositively correlated for all $i, j \in [n]$.

This follows whenever the binary random variables $\mathbf{X_i} := \mathbf{1_{i \in S}}$ are strong Rayleigh, and in fact this is how we establish negative correlation in all but one of the examples below.

Thus we are led to the problem:

What measures can one construct on $\mathcal{B}_\mathbf{n}$ with given marginals $\{\mathbf{p_i}\}$ and negative correlations?

Conditioned Bernoulli sampling

Let $\{\pi_i : 1 \leq i \leq n\}$ be numbers in $[0,1]$. Let $\mathbb{P}$ be the product measure making $\mathbb{E}X_i = \pi_i$ for each i. Let $\mathbb{P}' = (\mathbb{P} | S = k)$. The measure $\mathbb{P}'$ is called **conditioned Bernoulli sampling**.

Theorem (conditioned Bernoulli is SR)

For any choice of parameter values, the measure $\mathbb{P}'$ is strong Rayleigh.

PROOF: $\mathbb{P}$ is strong Rayleigh; SR is closed under conditioning on $\{S = k\}$. $\square$

Properties of conditioned Bernoulli πps sampling

- Given any probabilities $p_1, \ldots, p_n$ summing to k, there is a one-parameter family of vectors $(q_1, \ldots, q_n)$ whose conditional Bernoulli sampling law has marginals $p_1, \ldots, p_n$.
- All of these produce the same law.
- This law maximizes entropy among all laws with marginals $p_1, \ldots, p_n$.

For a discussion of the maximum entropy property, see Chen (2000), or Singh and Vishnoi (2013).

π-p-s sampling with negative correlations

This is one solution to the negative correlation sampling problem, with nice theoretical properties but poor algorithmic properties (at least in the offline sense: the quantities $q_1, \ldots, q_n$ are in general hard to compute).

A number of other schemes with better properties are reviewed by Brändén and Jonasson (2012) and shown to be strong Rayleigh.

Pivot sampling

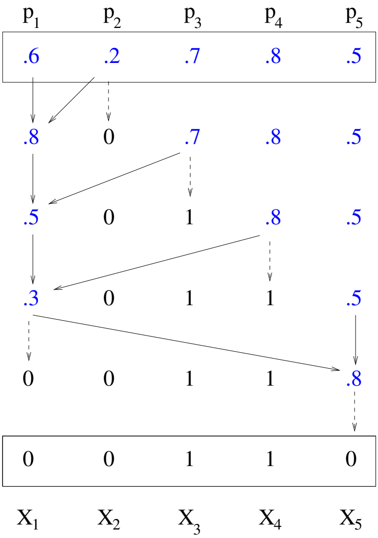
Once more let $\{p_i\}$ be probabilities summing to an integer $k < n$. Recursively, define a sampling scheme as follows.

If $p_1 + p_2 \leq 1$ then set X_1 or X_2 to zero with respective probabilities $p_2/(p_1 + p_2)$ and $p_1/(p_1 + p_2)$, then to choose the variables other than what was set to zero, run pivot sampling on $(p_1 + p_2, p_3, \ldots, p_n)$.

If $p_1 + p_2 > 1$, do the same thing except set one of X_1 or X_2 equal to 1 instead of 0 and the other to $p_1 + p_2 - 1$.

This method is very quick and does not involve having to compute auxilliary numbers such as the numbers p_i in conditional Bernoulli sampling.

Example of pivot sampling



Probability of this was 3/4

Probability of this was 7/15

Probability of this was 8/13

Probability of this was 5/8

Probability of this was 2/10

Proof that pivot sampling is SR

Induct on n. Assume WLOG that $p_1 + p_2 \leq 1$. Let $\mathbb{P}$ denote the law of pivot sampling with probabilities $(p_1 + p_2, p_3, \ldots, p_n)$. By induction $\mathbb{P}$ is strong Rayleigh.

Let $\mathbb{P}'$ be the product of the degenerate law δ_0 with $\mathbb{P}$, that is, sample from $\mathbb{P}$ then prepend a zero. This is trivially SR as well.

Let $\mathbb{P}''$ be the law $q\mathbb{P}' + (1 - q)(\mathbb{P}')^{12}$, obtained from $\mathbb{P}'$ by transposing 1 and 2 with probability $q := p_1/(p_1 + p_2)$.

$\mathbb{P}''$ is strong Rayleigh by closure under stirring.

But $\mathbb{P}''$ is the law of pivot sampling on $(p_1, \ldots, p_n)$; this completes the induction.

Sampling without replacement

Negative dependence for conditioned Bernoulli sampling appears to come from the fact that all n elements are competing for a fixed number k of slots.

Sampling without replacement is another means of inducing negative dependence. One might wonder whether this can also be used to obtain a negatively correlated π-p-s sampling scheme.

Weighted sampling without replacement scheme

Let $r_1, \ldots, r_n$ be positive real weights. Weighted sampling without replacement is the measure on subsets of size k obtained as follows. Let $Y_1 \leq n$ have $\mathbb{P}(Y_1 = \ell)$ proportional to r_ℓ. Next, let $\mathbb{P}(Y_2 = m | Y_1 = \ell)$ be proportional to r_m for $m \neq \ell$. Continue in this way until Y_i is chosen for all $i \leq k$ and let $X_i = \sum_{j=1}^k 1_{Y_j = i}$.

There is a strong intuition that this law should be negatively associated. The law of a sample of size k is stochastically increasing in k, a property shared with strong Rayleigh measures.

Weighted sampling without replacement scheme

Let $r_1, \ldots, r_n$ be positive real weights. Weighted sampling without replacement is the measure on subsets of size k obtained as follows. Let $Y_1 \leq n$ have $\mathbb{P}(Y_1 = \ell)$ proportional to r_ℓ. Next, let $\mathbb{P}(Y_2 = m | Y_1 = \ell)$ be proportional to r_m for $m \neq \ell$. Continue in this way until Y_i is chosen for all $i \leq k$ and let $X_i = \sum_{j=1}^k 1_{Y_j = i}$.

There is a strong intuition that this law should be negatively associated. The law of a sample of size k is stochastically increasing in k, a property shared with strong Rayleigh measures.

WRONG!

In a brief note in the Annals of Statistics, K. Alexander (1989) showed that weighted sampling without replacement is not, in general, even negatively correlated.

Determinantal measures and spanning trees

Determinantal measures

The law $\mathbb{P}$ on $\mathcal{B}_n$ is *determinantal* if there is a Hermitian matrix M such that for all subsets A of $[n]$, the minor $\det(M|_A)$ computes $\mathbb{E} \prod_{k \in A} X_k$.

It is easy to see that determinantal measures have negative correlations. The diagonal elements give the marginals. By the Hermitian property, the determinant of $M|_{ij}$ must be less than $M_{ii} M_{jj}$. Thus,

$$(\mathbb{E} X_i)(\mathbb{E} X_j) = M_{ii} M_{jj} \leq M_{ii} M_{jj} - M_{ij} \overline{M_{ij}} = \mathbb{E} X_i X_j \, .$$

SR property for determinantal measures

Theorem

Determinantal measures are strong Rayleigh.

SKETCH OF PROOF: By the theory of determinantal measures, the eigenvalues of M must lie in $[0, 1]$. Taking limits later if necessary, assume they lie in the open interval.

Then $F = C \det(H - Z)$ where Z is the diagonal matrix with entries $(x_1, \ldots, x_n)$ and $H = M^{-1} - I$ is positive definite. This is a sufficient criterion for stability (Gårding, circa 1951).

Determinantal sampling

Theorem (Lyons)

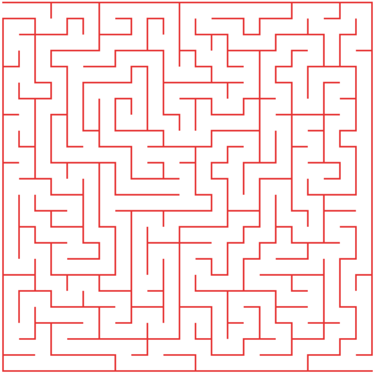
Given probabilities $p_1, \ldots, p_n$ summing to an integer $k < n$, we can always accomplish πps-sampling via a determinantal measure.

PROOF: The sequence $\{p_i : 1 \leq i \leq n\}$ is majorized by the sequence which is k ones followed by $n - k$ zeros. This majorization is precisely the criterion in the Schur-Horn Theorem, for existence of a Hermitian matrix M with $p_1, \ldots, p_n$ on the diagonal and eigenvalues consisting of 1 with mulitplicity k and 0 with multiplicity $n - k$. The matrix M defines the desired determinantal processes.

Spanning tree measures

Let $G = (V, E)$ be a graph with positive edge weights $\{w(e)\}$.

The weighted spanning tree measure is the measure WST on spanning trees proportional to $\prod_{e \in T} w(e)$.



Spanning trees are strong Rayleigh

The WST is determinantal; see, e.g., Burton and Pemantle (1993).

It follows that the random variables $\{X_e := \mathit{1}_{e \in T}\}$ have the strong Rayleigh property. In particular, they are NA.

Oveis Gharan et al. (2013) use the strong Rayleigh property for spanning trees in a result concerning TSP approximation. From the strong Rayleigh property, they deduce a lower bound on the probability of a given two vertices simultaneously having degree exactly 2.

Lipschitz functionals

Concentration inequalities

We have seen that the sum $S := \sum_{k=1}^{n} X_k$ of strong Rayleigh distributions satisfies a central limit theorem.

Often what one wants from a CLT is a large deviation estimate such as a Gaussian tail bound $\mathbb{P}(S - \mu > \lambda\sigma) \leq exp(-c\lambda^2)$, holding for a more general functional f in place of S.

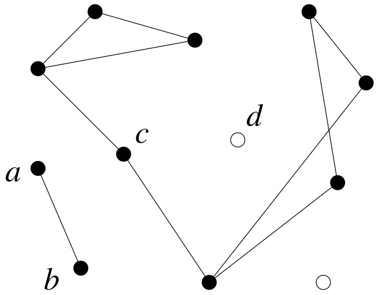
Lipschitz functionals

A Lipschitz function $f : \mathcal{B}_n \to \mathbb{R}$ is one that changes by no more than some constant c (without loss of generality $c = 1$) when a single coordinate of $\omega \in \mathcal{B}_n$ changes.

Example 1: $S := \sum_{k=1}^n X_k$ is Lipschitz-1.

Example 2: Let $\{1, \ldots, n\}$ index edges of a graph G whose degree is bounded by d. Let Y be a random subgraph of G and let $X_e := \mathit{1}_{e \in Y}$. Let f count one half the number of isolated vertices of Y. Then f is Lipschitz-1 because adding or removing an edge cannot affect the isolation of an vertex other than an endpoint of e.

Example Lipschitz function counting isolated vertices



The function counting isolated vertices is Lipschitz-2. For example, removing the edge *ab* alters the number of isolated vertices by $+2$, adding an edge *cd* alters the count by -1, and so forth.

Concentration for Lipschitz functionals

Strong tail bounds are available for Lipschitz functions of independent variables. These are based on classical exponential bounds going back to the 50's (Chernoff) and 60's (Hoeffding).

E. Mossel asked about generalizing from sums to Lipschitz functions assuming negative association. We don't know, but we can do it if we assume the strong Rayleigh property.

Theorem (Pemantle and Peres, 2015)

Let $f : \mathcal{B}_n \to R$ be Lipschitz-1. If $\mathbb{P}$ is k-homogeneous then

$$\mathbb{P}(|f - \mathbb{E}f| \geq a) \leq 2\exp\left(\frac{-a^2}{8k}\right).$$

Without the homogeneity assumption, the bound becomes $5\exp(-a^2/(16(a + 2\mu))$ *where* μ *is the mean.*

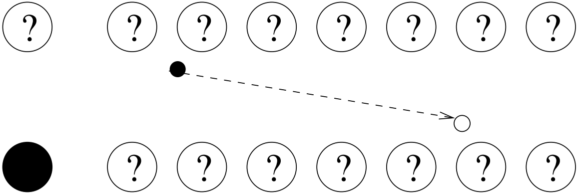
Sketch of proof

- Strong Rayleigh measures have the stochastic covering property.
- The classical Azuma martingale, $Z_k := \mathbb{E}(f \mid X_1, \ldots, X_k)$ can now be shown to have bounded differences, due to Lipschitz condition on f and coupling of the different conditional laws.

  (See illustration)

Note: this actually proves that any law with the SCP satisfies the same tail bounds for Lipschitz-1 functionals.

Illustration



There is a coupling such that the upper row samples from $\mathbb{P}$, the lower row samples from $(\mathbb{P}|X_1 = 1)$, and the only difference is in the X_1 variable and at most one other variable.

A similar picture holds for $(\mathbb{P}|X_1 = 0)$.

Therefore, f varies by at most 2 from the upper to the lower row, hence $|\mathbb{E}f - \mathbb{E}(f|X_1)| \leq 2$.

Example (number of leaves)

The proportion of vertices in a uniform spanning tree in $\mathbb{Z}^2$ that are leaves is known to be $8/\pi^2 - 16/\pi^3 \approx 0.2945$. Let us bound from above the probability that a UST in an $N \times N$ box has at least $N^2/3$ leaves. Letting f count half the number of leaves, we see that f is Lipschitz-1. The law of $\{X_e := 1_{e \in T}\}$ is SR and $N^2 - 1$ homogeneous. Therefore,

$$\mathbb{P}(f - \mathbb{E}f \geq a) \leq 2\exp(-a^2/(8N^2 - 8)) \,.$$

The probability of a vertex being a leaf in the UST on a box is bounded above by the probability for the infinite UST. Plugging in $a = N^2(1/3 - 8\pi^{-2} + 16\pi^{-3})$ and replacing the denominator by $8N^2$ therefore gives an upper bound of

$$2\exp\left[\left(\frac{1}{3} - \frac{8}{\pi^2} + \frac{16}{\pi^3}\right)^2 N^2\right] \approx 2e^{-0.0015N^2} \,.$$

Open questions

1. Are there natural measures that are negatively associated but not strong Rayleigh?
2. What are the most interesting measures for which negative association is conjectured, and is it possible they are strong Rayleigh?

1. NA but not SR?

Negative association does not imply strong Rayleigh. The only published examples, however, are highly contrived.

Problem

Find natural measures which are NA but not SR.

There are some candidates: classes of measures that are known to be NA but for which it is not known whether they are SR. I will briefly discuss two of these, namely Gaussian threshhold measures and sampling via Brownian motion in a polytope.

Gaussian threshhold measures

Let M be a positive semi-definite matrix with no positive entries off the diagonal. The multivariate Gaussian Y with covariances $\mathbb{E} Y_i Y_j = M_{ij}$ has pairwise negative (or zero) correlations. It is well known, for the multivariate Gaussian, that this implies negative association.

Let $\{a_j\}$ be arbitrary real numbers and let $X_j = 1_{Y_j \geq a_j}$. For obvious reasons, we call the law $\mathbb{P}$ of X on $\mathcal{B}_n$ a Gaussian threshhold measure.

Any monotone function f on $\mathcal{B}_n$ lifts to a function $\tilde{f}$ on $\mathbb{R}_n$ with $\mathbb{E}\tilde{f}(Y) = \mathbb{E}f(X)$. Thus, negative association of the Gaussian implies negative association of $\mathbb{P}$.

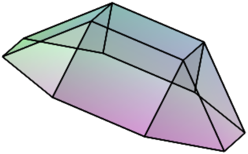
Problem

Is the Gaussian threshhold sampling law, $\mathbb{P}$, strong Rayleigh?

Sampling and polytopes

Sampling k elements out of n chooses a random k-set. Suppose we wish to restrict the set to be in a fixed list, $\mathcal{M}$. If we embed $\mathcal{B}_n$ in $\mathbb{R}^n$, then each set in $\mathcal{M}$ becomes a point in the hyperplane $\{\omega : \sum_{i=1}^{n} \omega_i = k\}$.

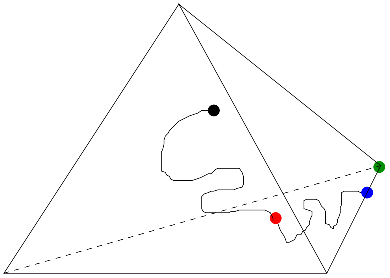
The set of probability measures on $\mathcal{M}$ maps to the convex hull of $\mathcal{M}$. The inverse image of p is precisely the set of measures with marginals given by $(p_1, \ldots, p_n)$. Thus, the πps-sampling problem is just the problem of choosing a point in the inverse image of p.



A point p in the polytope is a mixture of vertices in many ways, each corresonding to a measure supported on $\mathcal{M}$ that has mean p.

Brownian vertex sampling

Based on an idea of Lovett and Meka, M. Singh proposed to sample by running Brownian motion started from p. At any time, the Brownian motion is in the relative interior of a unique face, in which it constrained to remain thereafter. It stops when a vertex is reached. The martingale property of Brownian motion guarantees that this random set has marginals p.



Brownian vertex sampling restricted to a matroid

It turns out that the resulting scheme is negatively associated as long as the set system $\mathcal{M}$ is a **matroid**. This notion generalizes many others, such as spanning trees and vector space bases. For **balanced** matroids it is known that the uniform measure is strong Rayleigh, but even negative correlation can fail for other matroids.

Nevertheless...

Negative association of Brownian vertex sampling

Theorem (Peres and Singh, 2014)

For any matroid $\mathcal{M}$, the random k-set chosen by Brownian vertex sampling is negatively associated.

SKETCH OF PROOF: Let f and g be monotone functions depending on different sets of coordinates. Then $f(B_t)g(B_t)$ can be seen to be a supermartingale. At the stopping time, one gets

$$\int fg \, d\mathbb{P} = \mathbb{E}f(B_\tau)g(B_\tau) \leq \mathbb{E}f(B_0)\,\mathbb{E}g(B_0) = \int f \, d\mathbb{P} \int g \, d\mathbb{P} \, .$$

Is Brownian sampling strong Rayleigh?

Problem

Is the Brownian sampler strong Rayleigh?

2. conjectured NC/NA/SR

I will conclude by mentioning some measures where negative dependence is conjectured but it is not known whether any of the properties from negative correlation to strong Rayleigh holds.

A spanning tree is a connected acyclic graph. A seemingly small perturbation of the uniform or weighted spanning tree is the uniform or weighted acyclic subgraph – we simply drop the condition that the graph be connected.

Problem

Is the uniform acyclic graph strong Rayleigh? Is it even NC?

Note: one might ask the same question for the dual problem, namely uniform or weighted connected subgraphs. This problem is open. I don't know offhand whether the two are equivalent.

The random cluster model

The random cluster model is a statistical physics model in which a random subset of the edges of a graph G is chosen. The probability of $H \subseteq G$ is proportional to a product of edge weights $\prod_{e \in H} \lambda_e$, times q^N where N is the number of connected components of H. When $q \geq 1$, it is easy to check the positive lattice condition, hence positive association. When $q \leq 1$, is is conjectured to be negatively dependent (all properties from negative correlation to strong Rayleigh being equivalent for this model).

The random cluster model

The random cluster model is a statistical physics model in which a random subset of the edges of a graph G is chosen. The probability of $H \subseteq G$ is proportional to a product of edge weights $\prod_{e \in H} \lambda_e$, times q^N where N is the number of connected components of H. When $q \geq 1$, it is easy to check the positive lattice condition, hence positive association. When $q \leq 1$, is is conjectured to be negatively dependent (all properties from negative correlation to strong Rayleigh being equivalent for this model).

Problem (random cluster model)

Prove that the random cluster model has negative correlations.

Warning: this one has withstood a number of attacks.

End of Lecture 4

References I

K. Alexander.
A counterexample to a correlation inequality in finite sampling.
Ann. Statist., 17:436–439, 1989.

P. Brändén and J. Jonasson.
Negative dependence in sampling.
Scand. J. Stat., 39:830–838, 2012.

R. Burton and R. Pemantle.
Local characteristics, entropy and limit theorems for uniform spanning trees and domino tilings via transfer-impedances.
Ann. Probab., 21:1329–1371, 1993.

H. Chernoff.
A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations.
Ann. Math. Statist., 23:493–507, 1952.

S. X. Chen.
General properties and estimation of conditional Bernoulli models.
J. Multivariate Anal., 74:69–87, 2000.

References II

L. Gårding.
Linear hyperbolic partial differential equations with constant coefficients.
Acta Math., 85:1–62, 1951.

S. Gharan, A. Saberi, and M. Singh.
A randomized rounding approach to the Traveling Salesman Problem.
In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*, pages 550–559. IEEE Computer Science: Los Alamitos, CA, 2011.

W. Hoeffding.
Probability inequalities for sums of bounded random variables.
J. Amer. Stat. Assoc., 58:13–30, 1963.

R. Pemantle and Y. Peres.
Concentration of Lipschitz functionals of determinantal and other strong Rayleigh measures.
Comb. Prob. Comput., 23:140–160, 2014.

M. Singh and N. Vishnoi.
Entropy, optimization and counting.
In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 50–59, New York, 2014. ACM Press.