Stat206: Random Graphs and Complex Networks

Lecture 13: Gibbs Distributions and Entropy

Lecturer: David Aldous *Scribe: Raul Etkin*

13.1 Gibbs distributions and entropy

13.1.1 Definition of Entropy

Finite set S. Any probability distribution $\tilde{p} = (p_i)$ has entropy $\mathrm{Ent}(\tilde{p}) =_{\mathrm{def}} -\sum_{i \in S} p_i \log p_i \geq 0$.

Note: S need not have any structure for Ent to make sense. This is not true for expectations, for example.

13.1.2 Properties

Entropy is maximized by the uniform distribution: $\mathrm{Ent}(\mathrm{Unif}) = \log |S|$.

Entropy is minimized by a distribution (call it δ_{i_0}) concentrated at one element i_0: $\mathrm{Ent}(\delta_{i_0}) = 0$.

Entropy is one (of many) ways to measure "how random" is a distribution. This is separate from any structure on S.

13.1.3 Constrained entropy maximization problem

Given a function $H : S \to \mathcal{R}$ and given a constant h, any $\tilde{p}$ has an average $\sum_{i \in S} p_i H(i) = E_{\tilde{p}} H$.

Problem: what $\tilde{p}$ maximizes $\mathrm{Ent}(\tilde{p})$ subject to $E_{\tilde{p}} H = h$?

We can solve this problem by introducing a Lagrange multiplier λ. The resulting problem is:

Problem*: what $\tilde{p}$ maximizes $\mathrm{Ent}(\tilde{p}) - \lambda(E_{\tilde{p}} H - h)$?

Point: get solution to problem*, $p^{(\lambda)}$ which has some value $E_{p^{(\lambda)}} H = h_\lambda$. Choose λ such that $h_\lambda = h$. Then this $p^{(\lambda)}$ solves the original constrained problem.

In problem we also have a constraint $\sum_i p_i = 1$. We need to add another Lagrange multiplier μ to the objective function:

$$\mathrm{Ent}(\tilde{p}) - \lambda(E_{\tilde{p}} H - h) - \mu\left(\sum_i p_i - 1\right) = \sum_i (-p_i \log p_i) - \lambda\left(\sum_i p_i H(i) - h\right) - \mu\left(\sum_i p_i - 1\right)$$

To maximize this function, we use calculus and optimize over p_i separately for each i:

$$\begin{aligned} \frac{\partial}{\partial p_i}(\text{Objective function}) &= 0 = -\log p_i - 1 - \lambda H(i) - \mu \\ \log p_i &= -1 - \mu - \lambda H(i) \\ p_i &= \text{const.} * \exp[-\lambda H(i)] \end{aligned}$$

This is the Gibbs distribution: $p_i^{(\lambda)} = z_\lambda^{-1} \exp[-\lambda H(i)]$ where $z_\lambda = \sum_i \exp[-\lambda H(i)]$ is a normalizing constant.

We choose λ such that $E_{p^{(\lambda)}} H = h$.

13.1.4 Gibbs principle

Given S, K $(\leq \infty)$, functions $H_j : S \to \mathcal{R}$, and constants (h_j), the distribution $\tilde{p}$ which maximizes $\mathrm{Ent}(\tilde{p})$ subject to the K constraints $E_{\tilde{p}} H_j = h_j$, $1 \leq j \leq K$, is of the form:

$$p_i = \frac{\exp\left[\sum_{j=1}^{K} \lambda_j H_j(i)\right]}{z_{\lambda_1,\ldots,\lambda_K}}$$

where $\lambda_1, \ldots, \lambda_K$ are chosen to make the constraints hold.

Note: even though we worked with a finite set S, one can extend the results for countable and uncountable S. We can work with continuous distributions by using densities.

13.1.5 Examples

- $S = \{0, 1, \ldots, n\}$, constraints on $\tilde{p}$: "mean of $\tilde{p} = \beta$", i.e. $H(i) = i$. Then Gibbs distribution is $Bin(n, \beta/n)$.
- $S = [0, \infty)$, constraint "mean $= \beta$", i.e. $H(x) = x$. Then Gibbs distribution is exponential: $\lambda \exp(-\lambda x) = \frac{1}{\beta} \exp(-x/\beta)$.
- $S = (-\infty, \infty)$, constraints "mean $= \beta$" and "variance $= \sigma^2$". Then Gibbs distribution is $Normal(\beta, \sigma^2)$.

13.2 Relation to random graphs

Let $\mathcal{G}$ be the set of all graphs on the vertices $1, 2, \ldots, n$. Note that $|\mathcal{G}|$ is finite. Define:

$$H(\mathcal{G}) = \frac{\#\text{ of edges of } \mathcal{G}}{\binom{n}{2}}$$

Then the Gibbs distribution with constraint $EH(\cdot) = \beta$ is Erdös-Renyi $\mathcal{G}(n, \beta)$. This can be seen by noting that each graph in $\mathcal{G}$ can be represented as a sequence of $\binom{n}{2}$ zeros and ones. With this representation $H(\cdot)$ corresponds to the sample mean of the sequence. Since we are constraining the mean the corresponding Gibbs distribution is the Binomial, and therefore each edge appears independently with constant probability.

We want to extend this relation even further by recalling the random graphs with prescribed degree distribution $\{d_1, d_2, \ldots\}$ that we studied in Lecture 3. Define:

$$H_j(\mathcal{G}) = \frac{\#\text{ of degree-j vertices in } \mathcal{G}}{n}$$

Then the Gibbs distribution with constraints $EH_j = h_j$, $j = 1, 2, \ldots$, is $p(\mathcal{G}) = \text{const.}{*}\exp\left[-\sum_{j=1}^{\infty} \lambda_j H_j(\mathcal{G})\right]$. One can show that the construction from Lecture 3 gives approximately this.

Finally consider a real graph with some edge connecting a $degree-i$ with a $degree-j$ vertex. For each i and j define

$$m_{ij} = \frac{\#\text{ of such edges}}{n}$$

We could specify the matrix with entries m_{ij} as a constraint and obtain the corresponding Gibbs distribution.

13.2.1 All-purpose scheme for statistical description of a single complex network $\mathcal{G}_0$

- Choose some "statistic" H_1.
- Calculate $H_1(\mathcal{G}_0) = h_1$
- Look at Gibbs distribution with $EH_1 = h_1$. If $\mathcal{G}_0$ is like typical realization $\rightarrow$ done.
- If not, find some $H_2(\mathcal{G}_0) = h_2$ that is different from the prediction.
- Look at Gibbs distribution with $EH_1 = h_1$ and $EH_2 = h_2$.
- Continue ...

Good and bad features of the scheme

- $(\ddot\smile)$ All-purpose.
- $(\ddot\frown)$ Difficult to study Gibbs distribution analitically (hard statistical physics !).
- $(\ddot\smile)$ Can study via MCMC (Metropolis algorithm) simulations.
- $(\ddot\frown)$ Hard to tell if simulations give the correct answer.

13.2.2 Class projects

Big project: Specify 3 statistics of complex network : λ = scaling exponent, κ= cluster coefficient, $\bar{l}$= average vertex to vertex distance. Obtain the Gibbs distribution. This may be hard because sequential construction models do not have easy maximum entropy representations.

Little project: Gibbs distribution with $\bar{d} = 5$ (average degree) and $\kappa = 1/100$. For small κ we can guess that an Erdös-Renyi model with some extra edges to form triangles would work.