Maximum Entropy Network Ensembles

LTCC Course
Lesson 2

Ginestra Bianconi

School of Mathematical Sciences Queen Mary University of London

Queen Mary
University of London

Second lesson

Part 1

- Introduction to Maximum Entropy Principle

Part 2

- Canonical and microcanonical network ensembles

Introduction

to

Maximum Entropy Principle

Ensemble

Definition

An ensemble X is a triple $(x, \mathscr{A}_X, \mathscr{P}_X)$ where the outcome x is the value of a random variable which takes on one of possible values $\mathscr{A}_X = \{a_1, a_2, \ldots a_M\}$ having probabilities $\mathscr{P}_X = \{p_1, p_2, \ldots p_M\}$ with $P(x = a_i) = p_i$, $p_i \geq 0$ and $\sum_{i \in \mathscr{A}_X} P(x = a_i) = 1$

Abbreviation

Briefer notation will be used. For example, $P(x = a_i)$ maybe written as $P(a_i)$ or $P(x)$

Joint ensemble

A *joint ensemble* XY is an ensemble in which each outcome is an ordered pair (x, y) with $x \in \mathcal{A}_X = \{a_1, a_2, \ldots a_M\}$ $y \in \mathcal{A}_Y = \{b_1, b_2, \ldots b_R\}$

We call $P(x, y)$ *the* ***joint probability*** *of* (x, y)

Marginal probability

We can obtain the marginal probability $P(x)$ from the joint probability $P(x, y)$ by summation $P(x) = \sum_{y \in \mathcal{A}_Y} P(x, y)$

Conditional probability

The conditional probability is defined as

$$P(x = a_i \,|\, y = b_j) = \frac{P(x = a_i, y = b_j)}{P(y = b_j)} \quad \textbf{if} \quad P(y = b_j) \neq 0$$

Shannon information content of an outcome

Definition

The *Shannon information content of an outcome* is defined to be

$$h(x) = -\log_c p(x)$$

Comment

The original definition is given in bits, i.e. the base of the logarithm is chosen to be $c = 2$. However a popular choice is also $c = e$. The Shannon information content calculated in base $c = e$ and the one calculated in base $c = 2$ differ only by a multiplicative constant. If not explicitly stated here we take $c = e$

Shannon information content of an outcome

The smaller is the probability of an outcome, the larger is its Shannon information content

$$h(x) = -\ln p(x) = \ln \frac{1}{p(x)}$$

If the Shannon information content of a constant outcome is zero

$p(x) = 1$ **then** $h(x) = 0$

Shannon information content of a joint ensemble

The Shannon information content of an outcome of a joint ensemble is given by

$$h(x, y) = -\ln p(x, y)$$

In the case in which x and y are independent we have that the Shannon information content of (x,y) is given by the sum of the information content of x and y

$$h(x, y) = -\ln p(x, y) = -\ln[p(x)p(y)] = -h(x) - h(y)$$

Entropy of an ensemble

Definition

The **entropy of an ensemble** is defined to be the average Shannon information of an outcome

$$S = -\sum_{x \in \mathcal{A}_X} P(x) \ln P(x)$$

where the following convention is adopted,

$$0 \ln 0 = 0$$

Therefore we can also write

$$S = -\sum_{x \in \mathcal{A}_X | P(x) > 0} P(x) \ln P(x)$$

Properties of the Entropy

The entropy is non negative and is zero only for deterministic outcomes

$S \geq 0$ **with** $S = 0$ **iff** $P(x) = 1$ **for one** x

- **Proof:** Given the expression for the entropy

$$S = -\sum_{x \in \mathcal{A}_X | P(x) > 0} P(x) \ln P(x)$$

- If we have a non deterministic variable the

$P(x) \in (0,1) \forall x$ **therefore** $h(x) = -\ln P(x) > 0$ **it follows that** $S > 0$

- If we have a deterministic outcome

If $P(x) > 0$ **then** $P(x) = 1$ **with** $h(x) = -\ln P(x) = 0$ **it follows that** $S = 0$

Properties of the Entropy

The entropy is maximised for uniform distribution

- If the random variable can take M distinct values, i.e.

If $|\mathscr{A}_X| = M$

- then the maximum entropy over all possible distributions is

$$\max_{P(x)} S[P(x)] = S[P_U(x)] = \ln M$$

- where $P_U(x)$ is the uniform distribution

$$P_U(x) = \frac{1}{M}$$

Proof

Let us assume that our variable can take M possible values $|\mathscr{A}_X| = M$

The entropy of any distribution $P(x)$ which is naturally normalised

$$\sum_{x \in \mathscr{A}_X} P(x) = 1$$

is given by

$$S = -\sum_{x \in \mathscr{A}_X} P(x) \ln P(x)$$

In order to maximise the entropy over all normalised distributions consider the functional

$$\mathscr{F} = S - \nu \left(\sum_{x \in \mathscr{A}_X} P(x) - 1 \right) = -\sum_{x \in \mathscr{A}_X} P(x) \ln P(x) - \nu \left(\sum_{x \in \mathscr{A}_X} P(x) - 1 \right)$$

where ν is a Lagrangian multiplier.
By differentiating respect to $P(x)$ and putting the derivative to zero we get

$$\frac{\partial \mathscr{F}}{\partial P(x)} = -\ln P(x) - 1 - \nu = 0$$

Proof (continuation)

From the equations

$$\frac{\partial \mathcal{F}}{\partial P(x)} = -\ln P(x) - 1 - \nu = 0 \ \forall x \in \mathcal{A}_X$$

we get

$$P(x) = e^{-1-\nu}$$

By extremising $\mathcal{F}$ with respect to ν we get the normalization condition

$$\frac{\partial \mathcal{F}}{\partial \nu} = -\left(\sum_{x \in \mathcal{A}_X} P(x) - 1 \right) = 0$$

Since we have $|\mathcal{A}_X| = M$ the normalisation condition reads

$$\sum_{x \in \mathcal{A}_X} P(x) = e^{-1-\nu} M = 1 \text{ **or equivalently** } e^{-1-\nu} = \frac{1}{M}$$

It follows that the distribution $P(x)$ that maximised the entropy is uniform

$$P(x) = P_U(x) = \frac{1}{M}$$

and that

$$S[P_U(x)] = -\sum_{x \in \mathcal{A}_X} \frac{1}{M} \ln \frac{1}{M} = \ln M$$

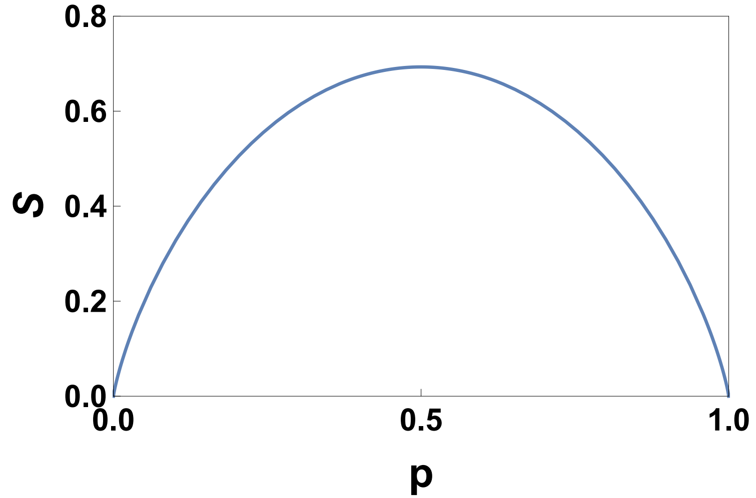
Entropy of a Bernoulli variable

Given a Bernoulli variable $x \in \{0,1\}$

with distribution $P(x) = p^x(1-p)^{1-x}$

the entropy is given by

$$S = -p\ln p - (1-p)\ln(1-p)$$



The entropy is zero for p=0 or p=1 (deterministic variable) and is maximised for p=1/2, i.e.

$$S = 0 \textbf{ for } p = 0 \textbf{ or } p = 1$$

$$S = \ln M = \ln 2 \textbf{ for } p = \frac{1}{2}$$

The entropy is a concave function

Entropy of a joint ensemble

Defintion

The entropy of a joint ensemble is defined as

$$S = -\sum_{(x,y)\in\mathcal{A}_X Y} P(x,y)\ln P(x,y)$$

with the usual convention $0\ln 0 = 0$

Uncorrelated joint ensembles

For uncorrelated variables, i.e. if $P(x,y) = P(x)P(y)$

The entropy is given by $S = -\sum_{(x,y)\in\mathcal{A}_X Y} P(x)P(y)\ln[P(x)P(y)]$

therefore we have $S = S_X + S_Y$

Quote

Everything should be made

as simple as possible, but not simpler

Einstein

Maximum entropy principle

The least biased ensemble

that satisfies a set of constraints

if the ensemble that maximises the entropy

(under the imposed constraints)

Maximum entropy principle

- Typically the constraints come from observations (data) or from previous knowledge about the ensemble.
- The maximum entropy principle is a very powerful tool to construct ensemble starting from partial information

Examples of Maximum entropy ensembles

Let us construct a maximum entropy ensemble in which we fix the expectations of some observables

$$f_\mu(x) \textbf{ for } \mu = 1,2\ldots,P$$

i.e. our constraints will be

$$\sum_{x\in\mathcal{A}_X} P(x)f_\mu(x) = C_\mu \;\; \mu = 1,2\ldots,P$$

with $C_\mu,\ \mu = 1,2\ldots,P$ being P constants.

Examples of Maximum entropy ensembles

The maximum entropy ensemble satisfying these constraints is given by the Gibbs measure

$$P(x) = \frac{e^{-\sum_{\mu=1}^{P} \lambda_\mu f_\mu(x)}}{Z}$$

where Z is the normalisation constant also called partition function

$$Z = \sum_{x \in \mathscr{A}_X} e^{-\sum_{\mu=1}^{P} \lambda_\mu f_\mu(x)}$$

and λ_μ are the Lagrangian multipliers fixed by the constraints or equivalently

$$-\frac{\partial \ln Z}{\partial \lambda_\mu} = C_\mu$$

Proof

We consider the maximum entropy ensemble of distribution $P(x)$
satisfying the constraints

$$\sum_{x \in \mathscr{A}_X} P(x) f_\mu(x) = C_\mu \;\; \mu = 1,2\ldots, P$$

and the normalisation constraint

$$\sum_{x \in \mathscr{A}_X} P(x) = 1$$

Therefore we need to maximise the entropy $S = -\sum_{x \in \mathscr{A}_X} P(x) \ln P(x)$

Under this constraints.
To this end we consider the functional

$$\mathscr{F} = -\sum_{x \in \mathscr{A}_X} P(x) \ln P(x) - \sum_{\mu=1}^{P} \lambda_\mu \left(\sum_{x \in \mathscr{A}_X} P(x) f_\mu(x) - C_\mu \right) - \nu \left(\sum_{x \in \mathscr{A}_X} P(x) - 1 \right)$$

where $\{\lambda_\mu\}, \nu$ are Lagrangian multipliers.
By differentiating respect to $P(x)$ and to each Lagrangian multiplier putting
the derivative to zero we can determine the maximum entropy ensemble distribution.

Proof (continuation)

These equations read

$$\frac{\partial \mathcal{F}}{\partial P(x)} = -\ln P(x) - \sum_{\mu=1}^{P} \lambda_\mu f_\mu(x) - 1 - \nu = 0$$

$$\frac{\partial \mathcal{F}}{\partial \lambda_\mu} = -\left(\sum_{x \in \mathcal{A}_X} P(x) f_\mu(x) - C_\mu \right) = 0$$

$$\frac{\partial \mathcal{F}}{\partial \nu} = -\left(\sum_{x \in \mathcal{A}_X} P(x) - 1 \right) = 0$$

From the first equation we get

$$P(x) = e^{-1-\nu} e^{-\sum_{\mu=1}^{P} \lambda_\mu f_\mu(x)}$$

From the normalisation condition we get

$$e^{\nu+1} = Z = \sum_{x \in \mathcal{A}_X} e^{-\lambda_\mu f_\mu(x)}$$

Finally $\{\lambda_\mu\}$ are fixed by the conditions

$$C_\mu = \sum_{x \in \mathcal{A}_X} f_\mu(x) P(x) = \frac{1}{Z} \sum_{x \in \mathcal{A}_X} f_\mu(x) e^{-\sum_{\tilde{\mu}=1}^{P} \lambda_{\tilde{\mu}} f_{\tilde{\mu}}(x)} = -\frac{\partial \ln Z}{\partial \lambda_\mu}$$

Entropy of the ensemble

- The entropy of this ensemble is given by

$$S = \sum_{\mu=1}^{P} \lambda_\mu C_\mu + \ln Z$$

- (left as an exercise)

Log-likelihood of an outcome

Consider an outcome x of a random variable with unknown distribution $P(x)$

We assume that the unknown distribution is coming from a family

of distributions $P_{\vec{\lambda}}(x)$ dependent on the parameters $\vec{\lambda}$

Definition

The *log-likelihood* of a parameters $\vec{\lambda}$ is defined as

$$\mathscr{L}(\vec{\lambda} \,|\, x) = \ln P_{\vec{\lambda}}(x)$$

Likelihood of a set of data

- Consider a set of data formed by independent outcomes of the random variable $\mathbf{x}$

$$\mathbf{x} = \{x_1, x_2, \ldots, x_N\}$$

- The log-likelihood of this set of data is

$$\mathscr{L}(\overrightarrow{\lambda} | \mathbf{x}) = \sum_{i=1}^{N} \ln P_{\overrightarrow{\lambda}}(x_i)$$

Maximum likelihood estimation

The maximum likelihood estimation of the parameters $\overrightarrow{\lambda^\star}$

corresponding to the distribution $P_{\overrightarrow{\lambda^\star}}(x)$

that best approximate the data

(according to maximum likelihood estimation) takes the form

$$\overrightarrow{\lambda}^\star = \textbf{argmax}_{\overrightarrow{\lambda}} \mathscr{L}(\overrightarrow{\lambda} | \mathbf{x}) = \textbf{argmax}_{\overrightarrow{\lambda}} \left[\sum_{i=1}^{N} \ln P_{\overrightarrow{\lambda}}(x_i) \right]$$

Relation between maximum entropy and maximum likelihood

Assuming that $P_{\vec{\lambda}}(x)$ is the Gibbs measures of the type

$$P_{\vec{\lambda}}(x) = \frac{e^{-\sum_{\mu=1}^{P} \lambda_\mu f_\mu(x)}}{Z}$$

Maximum likelihood estimation of the parameters $\vec{\lambda}^\star$

$$\vec{\lambda}^\star = \textbf{argmax}_{\vec{\lambda}} \mathscr{L}(\vec{\lambda} \,|\, \mathbf{x})$$

Implies that $P_{\vec{\lambda}}(x)$ is the maximum entropy ensemble with constraints fixed by the data

$$\langle f_\mu(x) \rangle_{DATA} = \langle f_\mu(x) \rangle_{ENSEMBLE} = \sum_{x \in \mathscr{A}_X} P_{\vec{\lambda}} f_\mu(x)$$

Proof

Consider a set of data formed by independent outcomes of the random variable X

$$D = \{x_1, x_2, \ldots, x_N\}$$

The log-likelihood of this set of data is

$$\mathscr{L}(\overrightarrow{\lambda} | \mathbf{x}) = \sum_{i=1}^{N} \ln P_{\overrightarrow{\lambda}}(x_i)$$

assuming

$$P_{\overrightarrow{\lambda}}(x) = \frac{e^{-\sum_{\mu=1}^{P} \lambda_\mu f_\mu(x)}}{Z}$$

We have

$$\mathscr{L}(\overrightarrow{\lambda} | \mathbf{x}) = \sum_{i=1}^{N} \ln P_{\overrightarrow{\lambda}}(x_i) = - \sum_{\mu} \lambda_\mu \sum_{i=1}^{N} f_\mu(x_i) - N \ln Z$$

Proof

Maximising the log-likelihood

$$\mathscr{L}(\overrightarrow{\lambda}\,|\,\mathbf{x}) = \sum_{i=1}^{N} \ln P_{\overrightarrow{\lambda}}(x_i) = -\sum_{\mu} \lambda_\mu \sum_{i=1}^{N} f_\mu(x_i) - N \ln Z$$

The log-likelihood of this set of data is

$$0 = \frac{\partial \mathscr{L}(\overrightarrow{\lambda}\,|\,\mathbf{x})}{\partial \lambda_\mu} = -\sum_{i=1}^{N} f_\mu(x_i) - N\frac{\partial \ln Z}{\partial \lambda_\mu} \textbf{ for } \mu = 1,2\ldots,P$$

We get

$$\frac{1}{N}\sum_{i=1}^{N} f_\mu(x_i) = -\frac{\partial \ln Z}{\partial \lambda_\mu} = \sum_{x\in\mathscr{A}_X} P_{\overrightarrow{\lambda}}(x) f_\mu(x) \textbf{ for } \mu = 1,2,\ldots,P$$

Therefore we have

$$\langle f_\mu(x)\rangle_{DATA} = \langle f_\mu(x)\rangle_{ENSEMBLE} = \sum_{x\in\mathscr{A}_X} P_{\overrightarrow{\lambda}} f_\mu(x) \textbf{ for } \mu - 1,2\ldots,P$$

What we have covered so far

In this first lesson we have covered

A. *Maximum entropy principle*

B. *Uniform distribution maximised the entropy*

C. *Exponential families (Gibbs distributions) maximise the entropy given a set of soft constraints*

D. *Relation between maximum entropy and maximum likelihood*

In the next lesson we will introduce

maximum entropy ensembles of networks

Microcanonical

and

Canonical

Network Ensembles

References

Books

- Mark Newman ***Networks: An introduction*** (Oxford University Press,2010)
- Ginestra Bianconi ***Multilayer networks:Structure and Function*** (Oxford University Press,2018)

Articles

- Park, J. and Newman, M.E., 2004. Statistical mechanics of networks. *Physical Review E*, *70*(6), p.066117.
- Bianconi, G., 2007. The entropy of randomized network ensembles. *EPL (Europhysics Letters)*, *81*(2), p.28005.
- Bianconi, G., 2009. Entropy of network ensembles. Physical Review E, 79(3), p.036114.
- Anand, K. and Bianconi, G., 2009. Entropy measures for networks: Toward an information theory of complex topologies. Physical Review E, 80(4), p.045102.
- Anand, K. and Bianconi, G., 2010. Gibbs entropy of network ensembles by cavity methods. Physical Review E, 82(1), p.011116.

Network Ensemble

Definition (for simple networks)

A network ensemble $\mathcal{G}$ is a triple $(G, \Omega_G, P(G))$ where G is any possible network $G = (E . V)$ belonging to the set of all simple networks with N nodes Ω_G and $P(G) \geq 0$ with $\sum_{G \in \Omega_G} P(G) = 1$ is the probability associate to each graph G

Generalization

The definition can be extended to non simple networks such as directed, weighted networks and also to generalised network structures by suitably changing the definition of Ω_G

Entropy of network ensembles

Definition

The *entropy of a network ensemble* is given by

$$S = -\sum_{G \in \Omega_G} P(G) \ln P(G)$$

It can be thought as the logarithm of the typical number of networks in the ensemble.

Here we have chosen the natural logarithm for simplicity

Constraints

We distinguish between soft constraints and hard constraints.

The ***soft constraints*** are the constraints satisfied in average over the ensemble of networks.

$$\sum_{G\in\Omega_G} F_\mu(G)P(G) = C_\mu \ \textbf{for}\ \mu = 1,2\ldots,P$$

The ***hard constraints*** are the constraints satisfied by each network in the ensemble.

$$F_\mu(G) = C_\mu \ \textbf{for}\ \mu = 1,2\ldots,P$$

Anand Bianconi 2009

Examples of hard constraints

$$F_\mu(G) = C_\mu \text{ for } \mu = 1,2\ldots,P$$

- Example 1:We can fix the total number of links L

$$\sum_{i<j} a_{ij} = L \quad \Rightarrow \quad \begin{cases} P = 1 \\ F_1(G) = \sum_{i<j} a_{ij} \\ C_1 = L \end{cases}$$

- Example 2: We can fix the entire degree sequence

$$\sum_{j=1}^{N} a_{ij} = k_i \text{ for } i = 1,2,\ldots,N \quad \Rightarrow \quad \begin{cases} P = N \\ F_i(G) = \sum_{j=1}^{N} a_{ij} \\ C_i = k_i \end{cases}$$

Examples of soft constraints

$$\sum_{G \in \Omega_G} F_\mu(G) P(G) = C_\mu \textbf{ for } \mu = 1,2\ldots,P$$

- Example 1:We can fix the expected total number of links $\bar{L}$

$$\sum_{G \in \Omega_G} \left(\sum_{i<j} a_{ij} \right) P(G) = \bar{L} \quad \Rightarrow \quad \begin{cases} P = 1 \\ F_1(G) = \sum_{i<j} a_{ij} \\ C_1 = \bar{L} \end{cases}$$

- Example 2: We can fix the expected degree sequence

$$\sum_{G \in \Omega_G} \left(\sum_{j=1}^{N} a_{ij} \right) P(G) = \bar{k}_i \textbf{ for } i = 1,2,\ldots,N \quad \Rightarrow \quad \begin{cases} P = N \\ F_i(G) = \sum_{j=1}^{N} a_{ij} \\ C_i = \bar{k}_i \end{cases}$$

Canonical and microcanical ensembles

- The **microcanonical ensemble** is the maximum entropy ensemble satisfying a given set of hard constraints of the type

$$F_\mu(G) = C_\mu \text{ for } \mu = 1,2\ldots,P$$

- The **canonical ensemble** is the maximum entropy ensemble satisfying a given set of soft constraints of the type

$$\sum_{G\in\Omega_G} F_\mu(G)P(G) = C_\mu \text{ for } \mu = 1,2\ldots,P$$

Anand Bianconi 2009

Conjugated ensembles

A microcanonical ensemble and a canonical ensemble

are **conjugated**

when they satisfy corresponding constraints,

i.e. when they satisfy

$$F_\mu(G) = C_\mu \textbf{ for } \mu = 1,2\ldots,P$$

$$\sum_{G\in\Omega_G} F_\mu(G)P(G) = C_\mu \textbf{ for } \mu = 1,2\ldots,P$$

with the same choice of $F_\mu(G)$ and C_μ respectively.

Canonical network ensemble

Proposition

The canonical ensemble satisfying the set of soft constraints

$$\sum_{G \in \Omega_G} F_\mu(G) P(G) = C_\mu \text{ for } \mu = 1,2\ldots,P$$

is determined by a probability given by

$$P(G) = \frac{1}{Z} e^{-\sum_{\mu=1}^{P} \lambda_\mu F_\mu(G)}$$

where Z is a normalisation constant $H(G) = \sum_{\mu=1}^{P} \lambda_\mu F_\mu(G)$ is called the Hamiltonian

and the Lagrangian multipliers λ_μ are fixed by the constraints.

For this reason the canonical network ensembles are also called exponential random graphs

Proof

We consider the maximum entropy network ensemble of distribution $P(G)$ satisfying the constraints

$$\sum_{G \in \Omega_G} P(G) F_\mu(G) = C_\mu \;\; \mu = 1,2\ldots,P$$

and the normalisation constraint

$$\sum_{G \in \Omega_G} P(G) = 1$$

Therefore we need to maximise the entropy

$$S = -\sum_{G \in \Omega_G} P(G) \log P(G)$$

Under this constraints.
To this end we consider the functional

$$\mathscr{F} = -\sum_{G \in \Omega_G} P(G) \log P(G) - \sum_{\mu=1}^{P} \lambda_\mu \left(\sum_{G \in \Omega_G} P(G) F_\mu(G) - C_\mu \right) - \nu \left(\sum_{G \in \Omega} P(G) - 1 \right)$$

where $\{\lambda_\mu\}, \nu$ are Lagrangian multipliers.
By differentiating respect to $P(G)$ and to each Lagrangian multiplier putting the derivative to zero we can determine the maximum entropy ensemble distribution.

Proof (continuation)

By maximising the functional

$$\mathscr{F} = -\sum_{G\in\Omega_G} P(G)\log P(G) - \sum_{\mu=1}^{P}\lambda_\mu\left(\sum_{G\in\Omega_G} P(G)F_\mu(G) - C_\mu\right) - \nu\left(\sum_{G\in\Omega} P(G) - 1\right)$$

We obtain the equations

$$\frac{\partial\mathscr{F}}{\partial P(G)} = -\ln P(G) - \sum_{\mu=1}^{P}\lambda_\mu F_\mu(G) - 1 - \nu = 0$$

$$\frac{\partial\mathscr{F}}{\partial\lambda_\mu} = -\left(\sum_{G\in\Omega_G} P(G)F_\mu(G) - C_\mu\right) = 0$$

$$\frac{\partial\mathscr{F}}{\partial\nu} = -\left(\sum_{G\in\Omega_G} P(G) - 1\right) = 0$$

From the first equation we get

$$P(G) = e^{-1-\nu}e^{-\sum_{\mu=1}^{P}\lambda_\mu F_\mu(G)}$$

Proof (continuation)

Given the Gibbs measure $P(G) = e^{-1-\nu} e^{-\sum_{\mu=1}^{P} \lambda_\mu F_\mu(G)}$

by using the normalisation condition $\sum_{G \in \Omega_G} P(G) = 1$

we get

$$e^{\nu+1} = Z = \sum_{G \in \Omega_G} e^{-\sum_\mu \lambda_\mu F_\mu(G)}$$

The other Lagrangian multipliers $\{\lambda_\mu\}$ are fixed by the conditions

$$\sum_{G \in \Omega_G} P(G) F_\mu(G) = C_\mu$$

Obtaining

$$C_\mu = \sum_{G \in \Omega_G} F_\mu(G) P(G) = \frac{1}{Z} \sum_{G \in \Omega_G} F_\mu(G) e^{-\sum_{\tilde{\mu}=1}^{P} \lambda_{\tilde{\mu}} F_{\tilde{\mu}}(G)} = -\frac{\partial \ln Z}{\partial \lambda_\mu}$$

$$C_\mu = -\frac{\partial \ln Z}{\partial \lambda_\mu}$$

Entropy of canonical ensemble

Proposition

The entropy of a canonical ensemble enforcing the constraints

$$\sum_{G \in \Omega_G} F_\mu(G)P(G) = C_\mu \textbf{ for } \mu = 1,2\ldots,P$$

is given by

$$\boxed{S = \sum_{\mu=1}^{P} \lambda_\mu C_\mu + \ln Z}$$

Proof

The maximum entropy distribution of a canonical network ensemble is given by

$$P(G) = \frac{1}{Z} e^{-\sum_{\mu=1}^{P} \lambda_\mu F_\mu(G)}$$

This ensemble has entropy

$$S = -\sum_{G \in \Omega_G} P(G) \ln P(G)$$

The entropy can be calculated explicitly as

$$S = -\sum_{G \in \Omega_G} P(G) \ln P(G) = -\sum_{G \in \Omega_G} P(G) \ln \left[\frac{1}{Z} e^{-\sum_{\mu=1}^{P} \lambda_\mu F_\mu(G)} \right]$$

$$S = -\sum_{G \in \Omega_G} P(G) \left[-\ln Z - \sum_{\mu=1}^{P} \lambda_\mu F_\mu(G) \right] = \ln Z + \sum_{\mu=1}^{P} \lambda_\mu \left[\sum_{G \in \Omega_G} P(G) F_\mu(G) \right] = \ln Z + \sum_{\mu=1}^{P} \lambda_\mu C_\mu$$

where we have used the constraints that the ensemble satisfies.

Maximum entropy micro canonical ensembles

- The **microcanonical ensemble** is the maximum entropy ensemble satisfying a given set of hard constraints of the type

$$F_\mu(G) = C_\mu \textbf{ for } \mu = 1,2\ldots,P$$

- In other words the micro canonical ensemble is the ensemble which satisfies the constraint

$$P(G) > 0 \textbf{ iff } F_\mu(G) = C_\mu \textbf{ for } \mu = 1,2\ldots,P$$

Therefore the entropy of this ensemble can be written as

$$S = -\sum_{G \in \Omega_G | \{F_\mu(G)=C_\mu\}_{\mu=1,2\ldots,P}} P(G) \ln P(G)$$

Maximum entropy micro canonical ensembles

The **microcanonical ensemble** satisfying a given set of hard constraints of the type

$$F_\mu(G) = C_\mu \textbf{ for } \mu = 1,2\ldots,P$$

has uniform distribution over all the networks satisfying the above constraints i.e.

$$P(G) = \frac{1}{Z_M}\prod_{\mu=1}^{P}\delta\left(F_\mu(G), C_\mu\right)$$

or where

$$Z_M = \sum_{G\in\Omega_G}\prod_{\mu=1}^{P}\delta\left(F_\mu(G), C_\mu\right)$$

Proof

The proof follows directly from the fact that maximum entropy distribution over a set of possible outcomes

$$\{G \in \Omega_G \,|\, F_\mu(G) = C_\mu \;\; \forall \mu\}$$

of cardinality

$$Z_M = |\{G \in \Omega_G \,|\, F_\mu(G) = C_\mu \;\; \forall \mu\}| = \sum_{G \in \Omega_G} \prod_{\mu=1}^{P} \delta\left(F_\mu(G), C_\mu\right)$$

is the uniform distribution

$$P(G) = \frac{1}{Z_M} \prod_{\mu=1}^{P} \delta\left(F_\mu(G), C_\mu\right)$$

Entropy of the microcanonical ensemble

Proposition

The entropy of the micro canonical ensemble is given by

$$\Sigma = -\sum_{G \in \Omega_G | \{F_\mu(G) = C_\mu\}_{\mu=1,2\ldots,P}} P(G) \ln P(G) = \ln Z_M$$

Proof

In fact we have

$$P(G) = \frac{1}{Z_M} \prod_{\mu=1}^{P} \delta\left(F_\mu(G), C_\mu\right) \quad \textbf{with} \quad Z_M = \sum_{G \in \Omega_G} \prod_{\mu=1}^{P} \delta\left(F_\mu(G), C_\mu\right)$$

Therefore

$$S = -\sum_{G \in \Omega_G | \{F_\mu(G) = C_\mu\}_{\mu=1,2\ldots,P}} \frac{1}{Z_M} \ln\left(\frac{1}{Z_M}\right) = \ln Z_M$$

Entropy of conjugated ensembles

Proposition

The entropy of a micro canonical ensemble Σ and the entropy S of the conjugated canonical ensemble are related by

$$\Sigma = S - \Omega$$

where

$$\Omega = -\ln \sum_{G \in \Omega_G} P_C(G) \prod_{\mu=1}^{P} \delta\left(F_\mu(G), C_\mu\right) \qquad P_C(G) = \frac{1}{Z} e^{-\sum_{\mu=1}^{P} \lambda_\mu F_\mu(G)}$$

and where $\delta(x, y)$ indicates the Kronecker delta.

Anand Bianconi 2010

Proof

Our aim is to calculate

$$\Omega = -\ln \sum_{G\in\Omega_G} P_C(G) \prod_{\mu=1}^{P} \delta\left(F_\mu(G), C_\mu\right)$$

where

$$P_C(G) = \frac{1}{Z} e^{-\sum_{\mu=1}^{P} \lambda_\mu F_\mu(G)}$$

By inserting this explicit expression we obtain

$$\Omega = -\ln\left[\sum_{G\in\Omega_G} \frac{1}{Z} e^{-\sum_{\mu=1}^{P} \lambda_\mu F_\mu(G)} \prod_{\mu=1}^{P} \delta\left(F_\mu(G), C_\mu\right)\right] = -\ln\left[\sum_{G\in\Omega_G} \frac{1}{Z} e^{-\sum_{\mu=1}^{P} \lambda_\mu C_\mu} \prod_{\mu=1}^{P} \delta\left(F_\mu(G), C_\mu\right)\right]$$

$$\Omega = -\ln\left[\frac{1}{Z} e^{-\sum_{\mu=1}^{P} \lambda_\mu C_\mu} \sum_{G\in\Omega_G} \prod_{\mu=1}^{P} \delta\left(F_\mu(G), C_\mu\right)\right] = -\ln\left[e^{-S} Z_M\right] = -\ln e^{-S+\Sigma} = S - \Sigma$$

Equivalence of the ensembles

If

$$\Omega = \mathcal{O}(1)$$

the entropies of conjugated ensembles

are asymptotically equal in the large network limit, i.e.

$$\Sigma \simeq S \qquad N \gg 1$$

In this case we say that we have equivalence of the ensembles.

This implies that the two resembles have the same statistical properties.

Preview:
Non-equivalence of the ensembles with extensive number of constraints

If the number of constraints is extensive

$$P = \mathcal{O}(N) \qquad \text{then} \qquad \Omega = \mathcal{O}(N)$$

Therefore the conjugated ensembles are not equivalent

In this case the entropy of the micro canonical ensemble is given by

$$\Sigma = S - \Omega$$

i.e. it is significantly lower than the entropy of the canonical ensemble.

Anand & Bianconi 2009 Anand & Bianconi 2010

Log-likelihood

Consider a network G coming from an unknown network ensemble $P(G)$

We assume that the unknown distribution of the ensemble is coming from an

ensemble with distribution $P_{\vec{\lambda}}(G)$ dependent on the parameters $\vec{\lambda}$

Definition

The *log-likelihood* of a parameters $\vec{\lambda}$ is defined as

$$\mathcal{L}(\vec{\lambda}\,|\,G) = -\ln P_{\vec{\lambda}}(G)$$

Maximum likelihood estimation

The maximum likelihood estimation of the parameters $\overrightarrow{\lambda^\star}$

corresponding to the distribution $P_{\overrightarrow{\lambda^\star}}(G)$

that best approximate the observed network

(according to maximum likelihood estimation) takes the form

$$\overrightarrow{\lambda}^\star = \textbf{argmax}_{\overrightarrow{\lambda}} \mathcal{L}(\overrightarrow{\lambda} \,|\, G) = \textbf{argmin}_{\overrightarrow{\lambda}} \left[-\ln P_{\overrightarrow{\lambda}}(G) \right]$$

Relation between maximum entropy and maximum likelihood

Assuming that $P_{\vec{\lambda}}(G)$ is the Gibbs measures of the type

$$P_{\vec{\lambda}}(G) = \frac{e^{-\sum_{\mu=1}^{P} \lambda_\mu F_\mu(G)}}{Z}$$

Maximum likelihood estimation of the parameters $\vec{\lambda}^{\star}$

$$\vec{\lambda}^{\star} = \textbf{argmax}_{\vec{\lambda}} \mathscr{L}(\vec{\lambda} \mid G)$$

Implies that $P_{\vec{\lambda}}(G)$ is the maximum entropy ensemble with constraints fixed by the data

$$F_\mu(G) = \langle F_\mu(G) \rangle_{ENSEMBLE} = \sum_{G' \in \Omega_G} P_{\vec{\lambda}}(G') F_\mu(G')$$

Proof

Minimising the negative log-likelihood

$$-\mathscr{L}(\overrightarrow{\lambda}\,|\,G) = -\ln P_{\overrightarrow{\lambda}}(G) = \sum_{\mu} \lambda_\mu F_\mu(G) + \ln Z$$

We get

$$0 = \frac{\partial \mathscr{L}(\overrightarrow{\lambda}\,|\,G)}{\partial \lambda_\mu} = F_\mu(G) + \frac{\partial \ln Z}{\partial \lambda_\mu} \textbf{ for } \mu = 1,2\ldots, P$$

Therefore

$$F_\mu(G) = -\frac{\partial \ln Z}{\partial \lambda_\mu} = \sum_{G' \in \Omega_G} P_{\overrightarrow{\lambda}}(G')F_\mu(G') \textbf{ for } \mu = 1,2,\ldots, P$$

Therefore we have

$$F_\mu(G) = \langle F_\mu(G) \rangle_{ENSEMBLE} = \sum_{G' \in \Omega_G} P_{\overrightarrow{\lambda}}(G')F_\mu(G')$$

Final remarks

In this second part of the second lesson we have covered

A. Canonical and microcanonical network ensembles

B. Non-equivalence of the ensembles in presence of extensive number of constraints

In the next lesson we will introduce

Random graphs and

Canonical ensembles with given expected degree sequence