Boltzmann Jaynes Inverse Problem, Maximum Entropy and Maximum Probability[*]

Marian Grendar[†]

---

[*] Presented at 'Workshop on Maximum Entropy Analysis', prepared by Robert K. Niven, UNSW at ADFA, Canberra, Sep 28-29, 2006.

[†] Bel University, Slovakia; visitor of CSE UNSW.

Overview

- Boltzmann Jaynes Inverse Problem

Overview

- Boltzmann Jaynes Inverse Problem
  - Conditional Limit Theorem

Overview

- Boltzmann Jaynes Inverse Problem
  - Conditional Limit Theorem
  - Relative Entropy Maximization (REM/MaxEnt) method

Overview

- Boltzmann Jaynes Inverse Problem
    - Conditional Limit Theorem
    - Relative Entropy Maximization (REM/MaxEnt) method
    - Maximum Probability (MaxProb) method

Overview

- Boltzmann Jaynes Inverse Problem
  - Conditional Limit Theorem
  - Relative Entropy Maximization (REM/MaxEnt) method
  - Maximum Probability (MaxProb) method
- Summary

Types etc.

- Let there be a discrete random variable which can take on finite number m of values $\mathcal{X} = [x_1, x_2, \ldots, x_m]$, with probabilities $q = [q_1, q_2, \ldots, q_m]$. $\mathcal{X}$ is called alphabet; its elements are letters. q is called source, or generator.

Types etc.

- Let there be a discrete random variable which can take on finite number m of values $\mathcal{X} = [x_1, x_2, \ldots, x_m]$, with probabilities $q = [q_1, q_2, \ldots, q_m]$. $\mathcal{X}$ is called alphabet; its elements are letters. q is called source, or generator.
- Let there be a random sample $x_1, x_2, \ldots, x_n$ of size n drawn from q. Type ν^n which the sample induces is $\nu^n = [n_1, n_2, \ldots, n_m]/n$ just the vector of relative frequencies of the m letters in the sample of size n. Usually n-type is used where it is necessary to highlight size of the underling random sample.

Types etc.

- Let there be a discrete random variable which can take on finite number m of values $\mathcal{X} = [x_1, x_2, \ldots, x_m]$, with probabilities $q = [q_1, q_2, \ldots, q_m]$. $\mathcal{X}$ is called alphabet; its elements are letters. q is called source, or generator.
- Let there be a random sample $x_1, x_2, \ldots, x_n$ of size n drawn from q. Type ν^n which the sample induces is $\nu^n = [n_1, n_2, \ldots, n_m]/n$ just the vector of relative frequencies of the m letters in the sample of size n. Usually n-type is used where it is necessary to highlight size of the underling random sample.
- There are $\Gamma(\nu^n) = \frac{n!}{\prod_{i=1}^{m} n_i!}$ sequences which induce the same type ν^n. The number $\Gamma(\cdot)$ is called multiplicity of type.

Example

Let $\mathcal{X} = [1, 2, 3, 4]$.
Let $q = [0.13, 0.09, 0.42, 0.36]$.
Let $n = 10$
and the sample let be: $3, 4, 1, 1, 4, 3, 4, 3, 2, 4$.
The sample induces type $\nu^{10} = [2, 1, 3, 4]/10$.

There is in total $\Gamma(\nu^n) = 1260$ sequences of length 10 which induce the same type.

Probability that source q generates type ν^n

What is the probability $\pi(\nu^n; q)$ that the source q generates type ν^n?
Well,

$$\pi(\nu^n; q) = \Gamma(\nu^n) \prod_{i=1}^{m} \exp(n \sum_{i=1}^{m} \nu_i^n \log q_i).$$

For q and ν^n from the Example it is $\pi(\nu^n; q) = 0.02384831$

How many n-types is there?

Let's denote the set of all n-types (on the alphabet $\mathcal{X}$), for a fixed n, by $\mathcal{P}_n(\mathcal{X})$. It is useful to view $\mathcal{P}_n$ as a subset of the set $\mathcal{P}(\mathcal{X})$ of all possible probability distributions on $\mathcal{X}$.

The number of n-types in $\mathcal{P}_n$ is $J = \binom{n+m-1}{m-1}$.

For $m = 4$, $n = 10$ it is 286.

Simple Boltzmann Jaynes Inverse Problem

Imagine that you were told that the source q generated SOME 10-type, from the set $\mathcal{P}_{10}$ of all 286 possible 10-types. Given the available information $\{\mathcal{X}, q, n, \mathcal{P}_{10}\}$ you have to select a type.

This is an example of BJIP.

Simple Boltzmann Jaynes Inverse Problem

Imagine that you were told that the source q generated SOME 10-type, from the set $\mathcal{P}_{10}$ of all 286 possible 10-types. Given the available information $\{\mathcal{X}, q, n, \mathcal{P}_{10}\}$ you have to select a type.

This is an example of BJIP.

BJIP is an under-determined, and in this sense ill-posed inverse problem.

Simple Boltzmann Jaynes Inverse Problem

Imagine that you were told that the source q generated SOME 10-type, from the set $\mathcal{P}_{10}$ of all 286 possible 10-types. Given the available information $\{\mathcal{X}, q, n, \mathcal{P}_{10}\}$ you have to select a type.

This is an example of BJIP.

BJIP is an under-determined, and in this sense ill-posed inverse problem.

How to proceed?

LLN

Moreover, as n gets larger, J gets larger ($\pi(\nu^n; q)$ gets smaller) and the problem harder.

LLN

Moreover, as n gets larger, J gets larger ($\pi(\nu^n; q)$ gets smaller) and the problem harder.

Gets it?

LLN

Moreover, as n gets larger, J gets larger ($\pi(\nu^n; q)$ gets smaller) and the problem harder.

Gets it?

Not at all! Just recall Law of Large Numbers!

LLN

Moreover, as n gets larger, J gets larger ($\pi(\nu^n; q)$ gets smaller) and the problem harder.

Gets it?

Not at all! Just recall Law of Large Numbers!

LLN: for $\epsilon > 0$, as $n \to \infty$,

$$\pi(\nu^n \in B(q, \epsilon);\ q) = 1.$$

LLN

Moreover, as n gets larger, J gets larger ($\pi(\nu^n; q)$ gets smaller) and the problem harder.

Gets it?

Not at all! Just recall Law of Large Numbers!

LLN: for $\epsilon > 0$, as $n \to \infty$,

$$\pi(\nu^n \in B(q, \epsilon);\ q) = 1.$$

The larger n, the more types concentrate in the ball centered at the source q.

LLN

Moreover, as n gets larger, J gets larger ($\pi(\nu^n; q)$ gets smaller) and the problem harder.

Gets it?

Not at all! Just recall Law of Large Numbers!

LLN: for $\epsilon > 0$, as $n \to \infty$,

$$\pi(\nu^n \in B(q, \epsilon); q) = 1.$$

The larger n, the more types concentrate in the ball centered at the source q.
Asymptotically there is almost impossible to find an n-type outside of the q-ball.

LLN

Moreover, as n gets larger, J gets larger ($\pi(\nu^n; q)$ gets smaller) and the problem harder.

Gets it?

Not at all! Just recall Law of Large Numbers!

LLN: for $\epsilon > 0$, as $n \to \infty$,

$$\pi(\nu^n \in B(q, \epsilon);\ q) = 1.$$

The larger n, the more types concentrate in the ball centered at the source q.
Asymptotically there is almost impossible to find an n-type outside of the q-ball.

So, for very large n we know how to solve this instance of BJIP: select $\nu^n \approx q$.

Two notes

1) Note that among all probability distributions in $\mathcal{P}$, q has the highest value of the relative entropy $H(p\,||\,q) = -\sum_{i=1}^{m} p_i \log(p_i/q_i)$, with respect to q.

2) Let $\hat{\nu}^n$ denote a type in $\mathcal{P}_n$, for which $\pi(\nu^n; q)$ is maximal. Formally, $\hat{\nu}^n = \arg\sup_{\nu^n \in \mathcal{P}_n} \pi(\nu^n; q)$. It holds true that as $n \to \infty$, $\hat{\nu}^n$ converges to q.

Finite n?

Yet, how to approach the above simplest BJIP when n is finite?

Finite n?

Yet, how to approach the above simplest BJIP when n is finite?
For now, we have to leave it open.

Finite n?

Yet, how to approach the above simplest BJIP when n is finite?
For now, we have to leave it open.

Harder BJIP

Let's try a harder BJIP. Now, we are given a set $\Pi \subset \mathcal{P}$ which does not contain q. Again, the only available information is $\{\mathcal{X}, q, n, \Pi\}$. How to select n-type from Π???

Harder BJIP

Let's try a harder BJIP. Now, we are given a set $\Pi \subset \mathcal{P}$ which does not contain q. Again, the only available information is $\{\mathcal{X}, q, n, \Pi\}$. How to select n-type from Π???

Example (cont'd): Mean value for q is 3.01. Let $\Pi = \{p : \sum_{i=1} p_i x_i = 3.2\}$, so that q is not in Π. Let $n = 10^{30}$. How to select n-type from Π?

Harder BJIP

Let's try a harder BJIP. Now, we are given a set $\Pi \subset \mathcal{P}$ which does not contain q. Again, the only available information is $\{\mathcal{X}, q, n, \Pi\}$. How to select n-type from Π???

Example (cont'd): Mean value for q is 3.01. Let $\Pi = \{p : \sum_{i=1} p_i x_i = 3.2\}$, so that q is not in Π. Let $n = 10^{30}$. How to select n-type from Π?

LLN is of little help here.

Harder BJIP

Let's try a harder BJIP. Now, we are given a set $\Pi \subset \mathcal{P}$ which does not contain q. Again, the only available information is $\{\mathcal{X}, q, n, \Pi\}$. How to select n-type from Π???

Example (cont'd): Mean value for q is 3.01. Let $\Pi = \{p : \sum_{i=1} p_i x_i = 3.2\}$, so that q is not in Π. Let $n = 10^{30}$. How to select n-type from Π?

LLN is of little help here. Indeed, LLN can be equivalently stated as:
$\pi(\nu^n \in B(q, \epsilon) \,|\, \nu^n \in \mathcal{P};\, q) = 1.$

Harder BJIP

Let's try a harder BJIP. Now, we are given a set $\Pi \subset \mathcal{P}$ which does not contain q. Again, the only available information is $\{\mathcal{X}, q, n, \Pi\}$. How to select n-type from Π???

Example (cont'd): Mean value for q is 3.01. Let $\Pi = \{p : \sum_{i=1} p_i x_i = 3.2\}$, so that q is not in Π. Let $n = 10^{30}$. How to select n-type from Π?

LLN is of little help here. Indeed, LLN can be equivalently stated as: $\pi(\nu^n \in B(q, \epsilon) \,|\, \nu^n \in \mathcal{P};\ q) = 1.$

We would like to have a statement of the following form: $\pi(\nu^n \in B(?, \epsilon) \,|\, \nu^n \in \Pi;\ q) = 1.$

Harder BJIP

Let's try a harder BJIP. Now, we are given a set $\Pi \subset \mathcal{P}$ which does not contain q. Again, the only available information is $\{\mathcal{X}, q, n, \Pi\}$. How to select n-type from Π???

Example (cont'd): Mean value for q is 3.01. Let $\Pi = \{p : \sum_{i=1} p_i x_i = 3.2\}$, so that q is not in Π. Let $n = 10^{30}$. How to select n-type from Π?

LLN is of little help here. Indeed, LLN can be equivalently stated as:
$\pi(\nu^n \in B(q, \epsilon) \,|\, \nu^n \in \mathcal{P};\ q) = 1.$

We would like to have a statement of the following form:
$\pi(\nu^n \in B(?, \epsilon) \,|\, \nu^n \in \Pi;\ q) = 1.$

The problem is to determine the point ? upon which the types conditionally concentrate.

CLLN

Conditional Law of Large Numbers, aka Conditional Limit Theorem (CoLT):
Provided that Π is convex, for $\epsilon > 0$, and $n \to \infty$,

CLLN

Conditional Law of Large Numbers, aka Conditional Limit Theorem (CoLT):
Provided that Π is convex, for $\epsilon > 0$, and $n \to \infty$,

$$\pi(\nu^n \in B(\hat{q}, \epsilon) \,|\, \nu^n \in \Pi; \, q) = 1,$$

CLLN

Conditional Law of Large Numbers, aka Conditional Limit Theorem (CoLT):
Provided that Π is convex, for $\epsilon > 0$, and $n \to \infty$,

$$\pi(\nu^n \in B(\hat{q}, \epsilon) \,|\, \nu^n \in \Pi;\, q) = 1,$$

where $\hat{q} = \arg\sup_{p \in \Pi} H(p \,||\, q)$ is the Relative Entropy Maximizing distribution, over Π.

CLLN

Conditional Law of Large Numbers, aka Conditional Limit Theorem (CoLT):
Provided that Π is convex, for $\epsilon > 0$, and $n \to \infty$,

$$\pi(\nu^n \in B(\hat{q}, \epsilon) \,|\, \nu^n \in \Pi;\, q) = 1,$$

where $\hat{q} = \arg\sup_{p \in \Pi} H(p \,||\, q)$ is the Relative Entropy Maximizing distribution, over Π.

This also solves the small n problem!

In order to comply with CoLT, one thus has to solve BJIP by REM method; in other words: to select that n-type(s) from Π_n for which the value of $H(\nu^n \,||\, q)$ is maximal.

CLLN

Conditional Law of Large Numbers, aka Conditional Limit Theorem (CoLT):
Provided that Π is convex, for $\epsilon > 0$, and $n \to \infty$,

$$\pi(\nu^n \in B(\hat{q}, \epsilon) \,|\, \nu^n \in \Pi;\, q) = 1,$$

where $\hat{q} = \arg\sup_{p \in \Pi} H(p \,||\, q)$ is the Relative Entropy Maximizing distribution, over Π.

This also solves the small n problem!

In order to comply with CoLT, one thus has to solve BJIP by REM method; in other words: to select that n-type(s) from Π_n for which the value of $H(\nu^n \,||\, q)$ is maximal.

... or by any other method which asymptotically obeys CoLT.

Maximum Probability method

Set Π_{10} of 10-types with mean value $\sum_{i=1}^{4} \nu_i^{10} x_i = 3.2$; and their probabilities (wrt q):

2	1	0	7	0.000429091	
2	0	2	6	0.00817656	
1	2	1	6	0.00242601	
0	4	0	6	2.99919e-05	
1	1	3	5	0.0264166	
0	3	2	5	0.00195947	
1	0	5	4	0.0359559	*
0	2	4	4	0.0133353	
0	1	6	3	0.0193609	
0	0	8	2	0.00564692	

Type denoted by asterisk has the highest $\pi(\nu^n; q)$ in this subset.

Maximum Probability (cont'd)

Table 1: MaxProb and REM/MaxEnt

n	J	$\hat{\nu}^n; q$
10	10	0.1000 0.0000 0.5000 0.4000
50	154	0.0800 0.0600 0.4400 0.4200
100	574	0.0800 0.0700 0.4200 0.4300
500	13534	0.0820 0.0700 0.4140 0.4340
1000	53734	0.0830 0.0700 0.4110 0.4360
$\hat{\mathbf{p}}$		0.0826 0.0709 0.4103 0.4361

As $n \to \infty$ MaxProb type(s) converges to REM distribution.

MaxProb and REM

There are two options:

1) REM can be viewed as an asymptotic instance of MaxProb.

MaxProb and REM

There are two options:

1) REM can be viewed as an asymptotic instance of MaxProb.

2) REM is a self-standing method (i.e., when n is finite choose the type(s) with highest value of relative entropy). REM and MaxProb asymptotically coincide.

EType

There is yet another method which asymptotically converges to REM distribution: Expected Type method (EType). The method selects:

$$\tilde{\nu}^n = \frac{\sum j=1^J \pi(\nu_j^n;q)\nu_j^n}{\sum_{j=1}^J \pi(\nu_j^n;q)}.$$

EType

There is yet another method which asymptotically converges to REM distribution: Expected Type method (EType). The method selects: $\tilde{\nu}^n = \frac{\sum j=1^J \pi(\nu_j^n;q)\nu_j^n}{\sum_{j=1}^J \pi(\nu_j^n;q)}$.

It can be shown that for set Π with unique REM distribution, EType converges to REM.

EType

There is yet another method which asymptotically converges to REM distribution: Expected Type method (EType). The method selects:

$$\tilde{\nu}^n = \frac{\sum j=1^J \pi(\nu_j^n;q)\nu_j^n}{\sum_{j=1}^J \pi(\nu_j^n;q)}.$$

It can be shown that for set Π with unique REM distribution, EType converges to REM.

The asymptotic identity of EType and REM however breaks up when there are several REM distributions!

Summary

- Questions are more important than answers. Problems are more important than their solutions. When thinking about MaxEnt/REM, it is worth formulating the problem which is to be solved.

Summary

- Questions are more important than answers. Problems are more important than their solutions. When thinking about MaxEnt/REM, it is worth formulating the problem which is to be solved.
- Boltzmann Jaynes Inverse Problem. Application of REM/MaxEnt for solving BJIP is justified by CoLT. No need to rely upon axiomatic arguments, etc. CoLT implies that BJIP can be solved by selecting REM type also when n is finite.

Summary

- Questions are more important than answers. Problems are more important than their solutions. When thinking about MaxEnt/REM, it is worth formulating the problem which is to be solved.
- Boltzmann Jaynes Inverse Problem. Application of REM/MaxEnt for solving BJIP is justified by CoLT. No need to rely upon axiomatic arguments, etc. CoLT implies that BJIP can be solved by selecting REM type also when n is finite.
- Maximum Probability method can be as well justified by CoLT. MaxProb converges to REM/MaxEnt. When n is finite, it is thus also possible to select MaxProb type.

Summary (cont'd)

- BJIP is not the only problem where one can ask a question of the form: 'what is the most probable type...'. There is also more complicated Jeffreys Inverse Problem, where MaxProb asymptotically converges to Jeffreys Entropy Maximization method. And one could take inspiration from Robert Niven and think also about Fermi Dirac Inverse Problem and Bose Einstein Inverse Problem.

Summary (cont'd)

- BJIP is not the only problem where one can ask a question of the form: 'what is the most probable type...'. There is also more complicated Jeffreys Inverse Problem, where MaxProb asymptotically converges to Jeffreys Entropy Maximization method. And one could take inspiration from Robert Niven and think also about Fermi Dirac Inverse Problem and Bose Einstein Inverse Problem.
- Though everything is possible, not everything is allowed (Roger Bacon). There are many entropies flying around. Some of them are even maximized! However, not each of the entropy maximization method can be given a shelter, in form of an associated inverse problem and CoLT. Whether there is such a chance for Renyi-Tsallis Entropy Maximization method is still an open problem[1].

[1] However, check Bercher, arXiv:math-ph/0609077

Some references

Cover and Thomas, Elements of Information Theory, Ch. 12

Csiszár, The Method of Types, IEEE IT, 1998.

Ellis, Entropy, Large Deviations and Statistical Mechanics.

Grendár and Grendár, What is the question that MaxEnt answers? A probabilistic interpretation, Proc of MaxEnt2000

Niven, Exact Maxwell-Boltzmann, Bose-Einstein and Fermi-Dirac Statistics, Physics Letters A

Thank you!

Thank you!

Special thanks to Robert Niven!

Thank you!

Special thanks to Robert Niven!

Hospitality of CSE and ADFA of UNSW is gratefully acknowledged.