An amended MaxEnt formulation for deriving Tsallis factors, and associated issues

Jean-François Bercher

July 13, 2006

Background

Tsallis entropy

$$H_\alpha(P) = \frac{1}{1-\alpha}\left[\int P(x)^\alpha dx - 1\right],$$

was introduced in 1988 for multifractals.
It is *nonextensive* $H_\alpha(X+Y) \neq H_\alpha(X) + H_\alpha(Y)$ when X and Y are independent. Strange property?
It generalizes Shannon/Boltzmann entropy (as others):

$$\lim_{\alpha\to 1} H_\alpha(P) = S(P).$$

- Tsallis literature: 88 $\rightarrow$ now more than 1000 papers

- When maximized under mean constraint, it leads to power laws

$$\begin{cases} \max_P \;\; H_\alpha(P) \\ \text{s.t.} \;\; \overline{E} = \int \varepsilon P(\varepsilon) d\varepsilon \end{cases}$$

- When maximized under mean constraint, it leads to power laws

$$\begin{cases} \max_P \quad H_\alpha(P) \\ \text{s.t.} \quad \overline{E} = \int \varepsilon P(\varepsilon) d\varepsilon \end{cases} \Longrightarrow P = K(1+\gamma\varepsilon)^\nu \simeq K\varepsilon^\nu$$

- When maximized under mean constraint, it leads to power laws

$$\left\{\begin{array}{ll} \max_P & H_\alpha(P) \\ \text{s.t.} & \overline{E} = \int \varepsilon P(\varepsilon) d\varepsilon \end{array}\right. \implies P = K(1+\gamma\varepsilon)^\nu \simeq K\varepsilon^\nu$$

And power laws are *interesting* as they appear in turbulence, fractals, ...
Often, power laws also meet long dependence phenomena (with unclear connexions).

- When maximized under mean constraint, it leads to power laws

$$\begin{cases} \max_P & H_\alpha(P) \\ \text{s.t.} & \overline{E} = \int \varepsilon P(\varepsilon) d\varepsilon \end{cases} \Longrightarrow P = K(1+\gamma\varepsilon)^\nu \simeq K\varepsilon^\nu$$

  And power laws are *interesting* as they appear in turbulence, fractals, . . .
  Often, power laws also meet long dependence phenomena (with unclear connexions).
- Fluctuating equilibriums

$$P(\varepsilon) \propto e^{-\varepsilon/kT} \quad \xrightarrow{T\sim\gamma}$$

- When maximized under mean constraint, it leads to power laws

$$\begin{cases} \max_P \quad H_\alpha(P) \\ \text{s.t.} \quad \bar{E} = \int \varepsilon P(\varepsilon) d\varepsilon \end{cases} \Longrightarrow P = K(1+\gamma\varepsilon)^\nu \simeq K\varepsilon^\nu$$

  And power laws are *interesting* as they appear in turbulence, fractals, ...
  Often, power laws also meet long dependence phenomena (with unclear connexions).

- Fluctuating equilibriums

$$P(\varepsilon) \propto e^{-\varepsilon/kT} \quad \xrightarrow{T\sim\gamma} \quad P(\varepsilon) \propto (1+\gamma(\varepsilon-\bar{E}))^\nu$$

Constraints

- Three choices for MaxTEnt
    1. Tsallis (88)

Constraints

- Three choices for MaxTEnt
  1. Tsallis (88)

$$\begin{cases} \max_P & H_\alpha(P) \\ \text{s.t.} & \overline{E} = \int \varepsilon P(\varepsilon) d\varepsilon \end{cases}$$

Classical mean

Constraints

- Three choices for MaxTEnt
  1. Tsallis (88)

$$\begin{cases} \max_P & H_\alpha(P) \\ \text{s.t.} & \overline{E} = \int \varepsilon P(\varepsilon) d\varepsilon \end{cases}$$

Classical mean

  2. Curado-Tsallis (91)

$$\begin{cases} \max_P & H_\alpha(P) \\ \text{s.t.} & \overline{E} = \int \varepsilon P(\varepsilon)^\alpha d\varepsilon \end{cases}$$

Constraints

- Three choices for MaxTEnt

1. Tsallis (88)

$$\begin{cases} \max_P \ \ H_\alpha(P) \\ \text{s.t.} \ \ \overline{E} = \int \varepsilon P(\varepsilon) d\varepsilon \end{cases}$$

Classical mean

2. Curado-Tsallis (91)

$$\begin{cases} \max_P \ \ H_\alpha(P) \\ \text{s.t.} \ \ \overline{E} = \int \varepsilon P(\varepsilon)^\alpha d\varepsilon \end{cases}$$

3. Tsallis-Mendes-Plastino (98)

$$\begin{cases} \max_P \ \ H_\alpha(P) \\ \text{s.t.} \ \ \overline{E} = \int \varepsilon \times \frac{P(\varepsilon)^\alpha}{\int P(\varepsilon)^\alpha d\varepsilon} d\varepsilon \end{cases}$$

Generalized mean

Outline

1. Maximization of Rényi-Tsallis entropy can be argued as the minimum of Kullback-Leibler divergence (Shannon Q-entropy) under a constraint that model a displacement from conventional equilibrium

Outline

1. Maximization of Rényi-Tsallis entropy can be argued as the minimum of Kullback-Leibler divergence (Shannon Q-entropy) under a constraint that model a displacement from conventional equilibrium
2. Two scenarii for the observation constraint are relevant, that lead to (i) classical mean constraint (ii) generalized mean constraint

Outline

1. Maximization of Rényi-Tsallis entropy can be argued as the minimum of Kullback-Leibler divergence (Shannon Q-entropy) under a constraint that model a displacement from conventional equilibrium
2. Two scenarii for the observation constraint are relevant, that lead to (i) classical mean constraint (ii) generalized mean constraint
3. Determination of law parameter. We will find something like

$$P(\varepsilon) \propto (1+\gamma(\varepsilon-\overline{\varepsilon}))^{\nu}$$

That is *self-referential*. $\rightarrow$ efficient procedures for determining γ.

Outline

1. Maximization of Rényi-Tsallis entropy can be argued as the minimum of Kullback-Leibler divergence (Shannon Q-entropy) under a constraint that model a displacement from conventional equilibrium
2. Two scenarii for the observation constraint are relevant, that lead to (i) classical mean constraint (ii) generalized mean constraint
3. Determination of law parameter. We will find something like

$$P(\varepsilon) \propto (1+\gamma(\varepsilon-\overline{\varepsilon}))^{\nu}$$

   That is *self-referential*. $\rightarrow$ efficient procedures for determining γ.
4. Special cases $\rightarrow$ well known entropies

Outline

1. Maximization of Rényi-Tsallis entropy can be argued as the minimum of Kullback-Leibler divergence (Shannon Q-entropy) under a constraint that model a displacement from conventional equilibrium
2. Two scenarii for the observation constraint are relevant, that lead to (i) classical mean constraint (ii) generalized mean constraint
3. Determination of law parameter. We will find something like
$$P(\varepsilon) \propto (1+\gamma(\varepsilon-\overline{\varepsilon}))^{\nu}$$
That is *self-referential*. $\rightarrow$ efficient procedures for determining γ.
4. Special cases $\rightarrow$ well known entropies
5. Legendre structure and thermodynamics

Q-entropies and divergences

$$H(P) = -\sum_{\mathscr{D}} P(x) \log P(x)$$

do not pass easily to the continuous case (no invariance).
Correct extension (Shannon 48, Jaynes 63, Kullback 51)

$$H_Q(P) = -\int_{\mathscr{D}} P(x) \log \frac{P(x)}{Q(x)} dx = -D(P||Q)$$

Generalization

$$\begin{cases} \text{Rényi } \frac{1}{1-\alpha} \log \int P^\alpha dx & D_\alpha(P||Q) = \frac{1}{1-\alpha} \log \int P^\alpha Q^{1-\alpha} dx \\ \text{Tsallis } \frac{1}{1-\alpha} \left[\int P^\alpha dx - 1\right] & \frac{1}{1-\alpha} \left[\int P^\alpha Q^{1-\alpha} dx - 1\right] \end{cases}$$

Rényi and Tsallis entropy have the same maxima

Rationale for Rényi-Tsallis maximum Q-entropy

In statistics, Sanov theorem or entropy concentration theorem are the rationale for MaxEnt. If one has a mean constraint and generates sequences according to Q, then *the most probable (set of) distribution is the nearest to* Q*, compatible with the constraint, in the Kullback-Leibler sense.*

Rationale for Rényi-Tsallis maximum Q-entropy

In statistics, Sanov theorem or entropy concentration theorem are the rationale for MaxEnt. If one has a mean constraint and generates sequences according to Q, then *the most probable (set of) distribution is the nearest to Q, compatible with the constraint, in the Kullback-Leibler sense.*
And there exist an *overwhelmingly preponderant* distribution:

$$\hat{P}_{ME} / \begin{cases} \min_P & D(P||Q) \\ \text{s.t.} & m = E_P[X] \end{cases}$$

But minimization of Tsallis-Rényi divergence gives a different distribution $\hat{P}_\alpha$ that is absolutely improbable...

Rationale for Rényi-Tsallis maximum Q-entropy

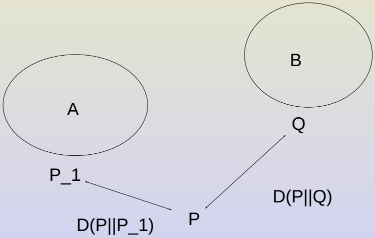
In statistics, Sanov theorem or entropy concentration theorem are the rationale for MaxEnt. If one has a mean constraint and generates sequences according to Q, then *the most probable (set of) distribution is the nearest to Q, compatible with the constraint, in the Kullback-Leibler sense.*
And there exist an *overwhelmingly preponderant* distribution:

$$\hat{P}_{ME} / \begin{cases} \min_P & D(P||Q) \\ \text{s.t.} & m = E_P[X] \end{cases}$$

But minimization of Tsallis-Rényi divergence gives a different distribution $\hat{P}_\alpha$ that is absolutely improbable...
$\rightsquigarrow$ Another probabilistic justification?

Displaced equilibriums

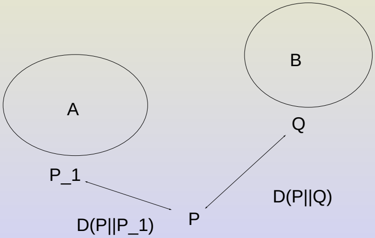
Fluctuations of an intensive parameter
$\equiv$ modified/perturbated "classical" equilibrium.
Instead of selecting the nearest distribution to Q, one selects the nearest to Q but also to P_1: the equilibrium distribution is somewhere between P_1 and Q



With $D(P||Q) = D(P||P_1) + \theta$

Displaced equilibriums

Fluctuations of an intensive parameter
$\equiv$ modified/perturbated "classical" equilibrium.
Instead of selecting the nearest distribution to Q, one selects the nearest to Q but also to P_1: the equilibrium distribution is somewhere between P_1 and Q

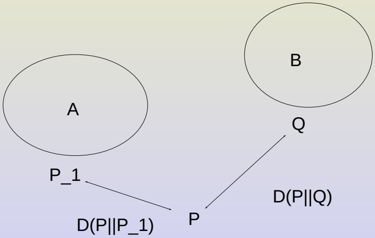


With $D(P||Q) = D(P||P_1) + \theta$

$$\begin{cases} \min_P \; D(P||Q) \\ s.t. \quad \theta = D(P||Q) - D(P||P_1) \end{cases}$$

Displaced equilibriums

Fluctuations of an intensive parameter
$\equiv$ modified/perturbated “classical” equilibrium.
Instead of selecting the nearest distribution to Q, one selects the nearest to Q but also to P_1: the equilibrium distribution is somewhere between P_1 and Q



With $D(P||Q) = D(P||P_1) + \theta$

$$\begin{cases} \min_P \; D(P||Q) \\ s.t. \quad \theta = D(P||Q) - D(P||P_1) \end{cases}$$

$$\theta = \int P(x) \log \frac{P_1(x)}{Q(x)} dx$$

is the mean log-likelihood.

Observables

We also have an observable

1. $m = E_{P_1}[X] = \int x P_1(x) dx$ mean of subsystem (A)
2. $m = E_{P^*}[X] = \int x P^*(x) dx$ mean of global system (A,B)

$$\begin{cases} \min_P \; D(P||Q) = \min_P \; \int P(x) \log \frac{P(x)}{Q(x)} dx \\ \text{subject to: } \; \theta = \int P(x) \log \frac{P_1(x)}{Q(x)} dx \end{cases},$$

Observables

We also have an observable

1. $m = E_{P_1}[X] = \int x P_1(x) dx$ mean of subsystem (A)
2. $m = E_{P^*}[X] = \int x P^*(x) dx$ mean of global system (A,B)

$$K = \begin{cases} \min_{P_1} \begin{cases} \min_P \; D(P||Q) = \min_P \; \int P(x) \log \frac{P(x)}{Q(x)} dx \\ \text{subject to: } \; \theta = \int P(x) \log \frac{P_1(x)}{Q(x)} dx \end{cases} \\ \text{subject to: } m = E_{P_1}[X] \text{ or } m = E_{P^*}[X] \end{cases},$$

Solution to the first problem

$$\begin{cases} \min_P \; D(P||Q) \\ s.t \;\; \theta = D(P||Q) - D(P||P_1) \end{cases}$$

Solution to the first problem

$$\begin{cases} \min_P \; D(P||Q) \\ s.t \;\; \theta = D(P||Q) - D(P||P_1) \end{cases}$$

Solution: (Kullback59)

$$P^*(x) = \frac{P_1(x)^\alpha Q(x)^{1-\alpha}}{\int P_1(x)^\alpha Q(x)^{1-\alpha} dx},$$

$\rightarrow$ Escort distribution of nonextensive statistics

- P^* which is the geometric mean between P_1 and Q realizes a trade-off, governed by α, between the two references.
- Note that $m = E_{P^*}[X]$ is the 'generalized α-expectation' and has a clear meaning now!

Optimum Lagrange parameter

The parameter α is simply the Lagrange parameter associated to the constraint θ, $\alpha \leq 1$, and is given by

$$\alpha^* / \quad \sup_{\alpha} \left\{ \alpha\theta - \log\left(\int P_1(x)^{\alpha} Q(x)^{1-\alpha} dx \right) \right\}$$

and

$$K_1 = \begin{cases} \min_P \; D(P||Q) \\ s.t \;\; \theta = D(P||Q) - D(P||P_1) \end{cases}$$

$$= \alpha^* \theta - \log\left(\int P_1(x)^{\alpha^*} Q(x)^{1-\alpha^*} dx \right)$$

And the maximization of Rényi Q-entropy...

$$K = \begin{cases} \min_{P_1} \begin{cases} \min_P \; D(P||Q) = \min_P \; \int P(x) \log \frac{P(x)}{Q(x)} dx \\ \text{subject to: } \; \theta = \int P(x) \log \frac{P_1(x)}{Q(x)} dx \end{cases} \\ \text{subject to: } m = E_{P_1}[X] \text{ or } m = E_{P^*}[X] \end{cases}$$

and

And the maximization of Rényi Q-entropy...

$$K = \left\{ \begin{array}{l} \min_{P_1} \left\{ \begin{array}{c} \sup_{\alpha} \left\{ \alpha\theta - \log\left(\int P_1(x)^{\alpha} Q(x)^{1-\alpha} dx \right) \right\} \\ \text{by dual attainment} \end{array} \right. \\ \quad \text{subject to: } m = E_{P_1}[X] \text{ or } m = E_{P^*}[X] \end{array} \right.$$

and

And the maximization of Rényi Q-entropy...

$$K = \begin{cases} \min_{P_1} \begin{cases} \sup_\alpha \{\alpha\theta - \underbrace{\log\left(\int P_1(x)^\alpha Q(x)^{1-\alpha} dx\right)}_{(\alpha-1)D_\alpha(P_1||Q)}\} \\ \end{cases} \\ \quad \text{subject to: } m = E_{P_1}[X] \text{ or } m = E_{P^*}[X] \end{cases}$$

and

And the maximization of Rényi Q-entropy…

$$K = \left\{ \min_{P_1} \left\{ \sup_\alpha \{ \alpha\theta - \underbrace{\log\left(\int P_1(x)^\alpha Q(x)^{1-\alpha} dx\right)}_{(\alpha-1)D_\alpha(P_1||Q)} \} \right. \right. \quad \text{subject to: } m = E_{P_1}[X] \text{ or } m = E_{P^*}[X]$$

and

$$K = \sup_\alpha \left\{ \alpha\theta + (1-\alpha)\min_{P_1} \left\{ \begin{array}{c} D_\alpha(P_1||Q) \\ \text{subject to: } m = E_{P_1}[X] \text{ or } m = E_{P^*}[X] \end{array} \right. \right.$$

Amounts to the minimization of Rényi/Tsallis divergence!

And the maximization of Rényi Q-entropy...

$$K = \left\{ \begin{array}{l} \min_{P_1} \left\{ \sup_\alpha \{ \alpha\theta - \underbrace{\log\left(\int P_1(x)^\alpha Q(x)^{1-\alpha} dx \right)}_{(\alpha-1) D_\alpha(P_1||Q)} \} \right. \\ \qquad\qquad \text{subject to: } m = E_{P_1}[X] \text{ or } m = E_{P^*}[X] \end{array} \right.$$

and

$$K = \sup_\alpha \left\{ \alpha\theta + (1-\alpha) \min_{P_1} \left\{ \begin{array}{l} D_\alpha(P_1||Q) \\ \text{subject to: } m = E_{P_1}[X] \text{ or } m = E_{P^*}[X] \end{array} \right. \right.$$

Amounts to the minimization of Rényi/Tsallis divergence!

And the maximization of Rényi Q-entropy...

$$K = \left\{ \min_{P_1} \left\{ \sup_\alpha \{ \alpha\theta - \underbrace{\log\left(\int P_1(x)^\alpha Q(x)^{1-\alpha} dx \right)}_{(\alpha-1)D_\alpha(P_1||Q)} \} \right. \right.$$

subject to: $m = E_{P_1}[X]$ or $m = E_{P^*}[X]$

and

$$K = \sup_\alpha \left\{ \alpha\theta + (1-\alpha) \right. \qquad \mathscr{F}_\alpha^{(1 \text{ or } \alpha)}(m)$$

Entropy functionals

Entropy functionals in the domain of observables.
'Contractions' of Rényi information divergence or of Kullback-Leibler information divergence for given constraints.
Level-one entropy functionals.

$$\mathscr{F}_\alpha^{(1 \text{ or } \alpha)}(m) = \begin{cases} \min_{P_1} D_\alpha(P_1||Q) \\ \text{subject to: } m = E_{P_1}[X] \text{ or } m = E_{P^*}[X] \end{cases},$$

Original problem reads

$$K = \sup_\alpha \left[\alpha\theta + (1-\alpha)\mathscr{F}_\alpha^{(1 \text{ or } \alpha)}(m) \right].$$

Properties: These entropy functionals are nonnegative, with an unique minimum at m_Q , the mean of Q. Furthermore, $\mathscr{F}_\alpha^{(1)}(m)$ is strictly convex for $\alpha \in [0,1]$.

A general 'Levy' distribution

Distribution $P^{\#}_{\nu}(x)$ is defined by:

$$P^{\#}_{\nu}(x) = [\gamma(x-\bar{x})+1]^{\nu}\, Q(x) e^{D_{\alpha}(P^{\#}_{\nu}||Q)},$$

on domain $\mathscr{D} = \mathscr{D}_Q \cap \mathscr{D}_{\gamma}$, where $\mathscr{D}_Q = \{x : Q(x) \geq 0\}$ and $\mathscr{D}_{\gamma} = \{x : \gamma(x-\bar{x})+1 \geq 0\}$.

A general 'Levy' distribution

Distribution $P^{\#}_{\nu}(x)$ is defined by:

$$P^{\#}_{\nu}(x) = [\gamma(x-\bar{x})+1]^{\nu}\, Q(x) e^{D_{\alpha}(P^{\#}_{\nu}||Q)},$$

on domain $\mathscr{D} = \mathscr{D}_Q \cap \mathscr{D}_{\gamma}$, where $\mathscr{D}_Q = \{x : Q(x) \geq 0\}$ and $\mathscr{D}_{\gamma} = \{x : \gamma(x-\bar{x})+1 \geq 0\}$.

- $\bar{x}$ is either (a) a fixed parameter, say m, and $P^{\#}_{\nu}(x)$ is a two parameters distribution, (b) or some statistical mean with respect to $P^{\#}_{\nu}(x)$, e.g. its "classical" or "generalized" mean, and as such a function of γ.
- $P^{\#}_{\nu}(x)$ is not necessarily normalized to one.
- Partition function $Z_{\nu}(\gamma,\bar{x}) = \int_{\mathscr{D}} [\gamma(x-\bar{x})+1]^{\nu}\, Q(x) dx.$

Normalization of ‘Levy’ distribution

Theorem

- *The Levy distribution* $P_{\xi}^{\#}(x)$ *with exponent* $\nu = \xi = \frac{1}{\alpha - 1}$ *is normalized to one if and only if* $\bar{x} = E_{\xi}[x]$, *the statistical mean of the distribution, and*
  $D_{\alpha}(P_{\xi}^{\#}||Q) = -\log Z_{\xi+1}(\gamma, \bar{x}) = -\log Z_{\xi}(\gamma, \bar{x})$.

Normalization of 'Levy' distribution

Theorem

- *The Levy distribution* $P^{\#}_{\xi}(x)$ *with exponent* $\nu = \xi = \frac{1}{\alpha-1}$ *is normalized to one if and only if* $\bar{x} = E_{\xi}[x]$, *the statistical mean of the distribution, and* $D_{\alpha}(P^{\#}_{\xi}||Q) = -\log Z_{\xi+1}(\gamma,\bar{x}) = -\log Z_{\xi}(\gamma,\bar{x})$.
- *The Levy distribution* $P^{\#}_{-\xi}(x)$ *with exponent* $\nu = -\xi = \frac{1}{1-\alpha}$ *is normalized to one if and only if* $\bar{x} = E_{-\xi-1}[x] = E^{(\alpha)}_{-\xi}[x]$, *the generalized* $\alpha-$*expectation of the distribution, and* $D_{\alpha}(P^{\#}_{-\xi}||Q) = -\log Z_{-(\xi+1)}(\gamma,\bar{x}) = -\log Z_{-\xi}(\gamma,\bar{x})$.

Normalization of ‘Levy’ distribution

Theorem

- *The Levy distribution* $P^{\#}_{\xi}(x)$ *with exponent* $\nu = \xi = \frac{1}{\alpha - 1}$ *is normalized to one if and only if* $\bar{x} = E_{\xi}[x]$, *the statistical mean of the distribution, and* $D_{\alpha}(P^{\#}_{\xi}||Q) = -\log Z_{\xi+1}(\gamma, \bar{x}) = -\log Z_{\xi}(\gamma, \bar{x})$.
- *The Levy distribution* $P^{\#}_{-\xi}(x)$ *with exponent* $\nu = -\xi = \frac{1}{1-\alpha}$ *is normalized to one if and only if* $\bar{x} = E_{-\xi-1}[x] = E^{(\alpha)}_{-\xi}[x]$, *the generalized* $\alpha-$*expectation of the distribution, and* $D_{\alpha}(P^{\#}_{-\xi}||Q) = -\log Z_{-(\xi+1)}(\gamma, \bar{x}) = -\log Z_{-\xi}(\gamma, \bar{x})$.

$\Rightarrow$ *When* $\bar{x}$ *is a fixed parameter* m, *this will be only true for a special value* γ^{*} *of* γ *such that* $E_{\xi}[x] = m$ *or* $E^{(\alpha)}_{-\xi}[x] = m$.

Sketch of proof

1. If $P(x) = K(x)Q(x)e^{D_\alpha(P||Q)}$, then $D_\alpha(P||Q) = -\log \int K(x)^\alpha Q(x)dx$ and $D_\alpha(P||Q) = -\log \int K(x)Q(x)dx$ if $P(x)$ is normalized to one.

Sketch of proof

1. If $P(x) = K(x)Q(x)e^{D_\alpha(P||Q)}$, then
$D_\alpha(P||Q) = -\log\int K(x)^\alpha Q(x)dx$ and
$D_\alpha(P||Q) = -\log\int K(x)Q(x)dx$ if $P(x)$ is normalized to one.
2. For distribution $P_\nu^\#(x)$, and any parameter γ we have
$D_\alpha(P_\nu^\#||Q) = -\log Z_{\alpha\nu}(\gamma,\bar{x})$ and
$D_\alpha(P_\nu^\#||Q) = -\log Z_\nu(\gamma,\bar{x})$ if $P_\nu^\#(x)$ is normalized to one.

Sketch of proof

1. If $P(x) = K(x)Q(x)e^{D_\alpha(P||Q)}$, then
$D_\alpha(P||Q) = -\log \int K(x)^\alpha Q(x)dx$ and
$D_\alpha(P||Q) = -\log \int K(x)Q(x)dx$ if $P(x)$ is normalized to one.
2. For distribution $P^\#_\nu(x)$, and any parameter γ we have
$D_\alpha(P^\#_\nu||Q) = -\log Z_{\alpha\nu}(\gamma,\bar{x})$ and
$D_\alpha(P^\#_\nu||Q) = -\log Z_\nu(\gamma,\bar{x})$ if $P^\#_\nu(x)$ is normalized to one.
3. If $\nu = \pm\xi$, we have $Z_{\alpha\nu}(\gamma,\bar{x}) = Z_{\pm(\xi+1)}(\gamma,\bar{x}) = Z_{\pm\xi}(\gamma,\bar{x})$ if
$P^\#_{\pm\xi}(x)$ is normalized to one.

Sketch of proof

1. If $P(x) = K(x)Q(x)e^{D_\alpha(P||Q)}$, then
$D_\alpha(P||Q) = -\log \int K(x)^\alpha Q(x)dx$ and
$D_\alpha(P||Q) = -\log \int K(x)Q(x)dx$ if $P(x)$ is normalized to one.
2. For distribution $P^{\#}_\nu(x)$, and any parameter γ we have
$D_\alpha(P^{\#}_\nu||Q) = -\log Z_{\alpha\nu}(\gamma,\bar{x})$ and
$D_\alpha(P^{\#}_\nu||Q) = -\log Z_\nu(\gamma,\bar{x})$ if $P^{\#}_\nu(x)$ is normalized to one.
3. If $\nu = \pm\xi$, we have $Z_{\alpha\nu}(\gamma,\bar{x}) = Z_{\pm(\xi+1)}(\gamma,\bar{x}) = Z_{\pm\xi}(\gamma,\bar{x})$ if
$P^{\#}_{\pm\xi}(x)$ is normalized to one.
4. Partition functions of successive exponents are linked by
$Z_{\nu+1}(\gamma,\bar{x}) = E_{\nu+1-k}\left[(\gamma(x-\bar{x})+1)^k\right] Z_{\nu+1-k}(\gamma,\bar{x})$.
For k=1: $Z_{\nu+1}(\gamma,\bar{x}) = E_\nu\left[\gamma(x-\bar{x})+1\right] Z_\nu(\gamma,\bar{x})$,

Sketch of proof

1. If $P(x) = K(x)Q(x)e^{D_\alpha(P||Q)}$, then
   $D_\alpha(P||Q) = -\log\int K(x)^\alpha Q(x)dx$ and
   $D_\alpha(P||Q) = -\log\int K(x)Q(x)dx$ if $P(x)$ is normalized to one.
2. For distribution $P^\#_\nu(x)$, and any parameter γ we have
   $D_\alpha(P^\#_\nu||Q) = -\log Z_{\alpha\nu}(\gamma,\bar{x})$ and
   $D_\alpha(P^\#_\nu||Q) = -\log Z_\nu(\gamma,\bar{x})$ if $P^\#_\nu(x)$ is normalized to one.
3. If $\nu = \pm\xi$, we have $Z_{\alpha\nu}(\gamma,\bar{x}) = Z_{\pm(\xi+1)}(\gamma,\bar{x}) = Z_{\pm\xi}(\gamma,\bar{x})$ if $P^\#_{\pm\xi}(x)$ is normalized to one.
4. Partition functions of successive exponents are linked by
   $Z_{\nu+1}(\gamma,\bar{x}) = E_{\nu+1-k}\left[(\gamma(x-\bar{x})+1)^k\right] Z_{\nu+1-k}(\gamma,\bar{x})$.
   For k=1: $Z_{\nu+1}(\gamma,\bar{x}) = E_\nu\left[\gamma(x-\bar{x})+1\right] Z_\nu(\gamma,\bar{x})$,
5. And $Z_{\nu+1}(\gamma,\bar{x}) = Z_\nu(\gamma,\bar{x})$ iff $\bar{x} = E_\nu[X]$

Solutions

Procedure: (i) minimize the Lagrangian in $P(x) \to P_{\lambda,\mu}(x)$, (ii) maximize the dual function in order to exhibit the optimum Lagrange parameters.

Solutions

Procedure: (i) minimize the Lagrangian in $P(x) \rightarrow P_{\lambda,\mu}(x)$, (ii) maximize the dual function in order to exhibit the optimum Lagrange parameters.
Taking into account normalization conditions described above,

$$(C) \quad P_C(x) = \frac{[\gamma(x-\bar{x})+1]^{\xi}}{Z_{\xi}(\gamma,\bar{x})} Q(x), \text{ with } \bar{x} = E_{P_C}[X] = E_{\xi}[X]$$

$$(G) \quad P_G(x) = \frac{(1+\gamma(x-\bar{x}))^{-\xi}}{Z_{-\xi}(\gamma,\bar{x})} Q(x) \text{ with } \bar{x} = E_{P_G}[X] = E_{-(\xi+1)}[X]$$

where $\bar{x}$ is a statistical mean, function of γ, and NOT a fixed value *(long-time mistake)*; $\xi = \frac{1}{\alpha-1}$.

Solutions

Procedure: (i) minimize the Lagrangian in $P(x) \to P_{\lambda,\mu}(x)$, (ii) maximize the dual function in order to exhibit the optimum Lagrange parameters.
Taking into account normalization conditions described above,

$$(C) \quad P_C(x) = \frac{[\gamma(x-\bar{x})+1]^{\xi}}{Z_{\xi}(\gamma,\bar{x})} Q(x), \text{ with } \bar{x} = E_{P_C}[X] = E_{\xi}[X]$$

$$(G) \quad P_G(x) = \frac{(1+\gamma(x-\bar{x}))^{-\xi}}{Z_{-\xi}(\gamma,\bar{x})} Q(x) \text{ with } \bar{x} = E_{P_G}[X] = E_{-(\xi+1)}[X]$$

where $\bar{x}$ is a statistical mean, function of γ, and NOT a fixed value *(long-time mistake)*; $\xi = \frac{1}{\alpha-1}$.
Optimum distributions $P_{C,G}(x)$ are self referential (implicitely defined) and associated dual functions are intractable.

Optimum distributions $P_{C,G}(x)$ are self referential (implicitely defined) and associated dual functions are intractable
$\rightsquigarrow$ alternate dual functions?

Optimum distributions $P_{C,G}(x)$ are self referential (implicitely defined) and associated dual functions are intractable

$\rightsquigarrow$ alternate dual functions?

It can be shown that γ^* is solution of the *equivalent* pbs

$$\max_{\gamma}\left[-\log Z_{\xi+1}(\gamma,m)\right]$$

Classical

$$\max_{\gamma}\left[-\log Z_{-\xi}(\gamma,m)\right]$$

Generalized

Optimum distributions $P_{C,G}(x)$ are self referential (implicitely defined) and associated dual functions are intractable
$\rightsquigarrow$ alternate dual functions?

It can be shown that γ^* is solution of the *equivalent* pbs

$$\max_{\gamma}\left[-\log Z_{\xi+1}(\gamma,m)\right] \qquad \max_{\gamma}\left[-\log Z_{-\xi}(\gamma,m)\right]$$

Classical | Generalized

$$\max_{\gamma}\widetilde{D}(\gamma)=\begin{cases}\min_{P_1} D_\alpha(P_1||Q)\\ \text{s.t. } m=E_{.}[X]\end{cases} = D_\alpha(\hat{P}_1||Q) = \mathscr{F}_\alpha^{(.)}(m)$$

dual attainment | entropy functional

Optimum distributions $P_{C,G}(x)$ are self referential (implicitely defined) and associated dual functions are intractable
$\rightsquigarrow$ alternate dual functions?

It can be shown that γ^* is solution of the *equivalent* pbs

$$\max_{\gamma}\left[-\log Z_{\xi+1}(\gamma,m)\right] \qquad \max_{\gamma}\left[-\log Z_{-\xi}(\gamma,m)\right]$$

Classical — Generalized

$$\max_{\gamma}\widetilde{D}(\gamma) = \left\{\begin{array}{c}\min_{P_1} D_\alpha(P_1||Q)\\ \text{s.t. } m = E_{.}[X]\end{array}\right. = D_\alpha(\hat{P}_1||Q) = \mathscr{F}_\alpha^{(.)}(m)$$

dual attainment — entropy functional

$\mapsto$ two practical numerical schemes for the identification of the distributions parameters ($Z_{\xi+1}(\gamma,m)$ and $Z_{-\xi}(\gamma,m)$ are two convex functions for $\alpha\leq 1$) + subtilities

Entropy functionals in special cases for Q

e.g. uniform, Bernoulli, gamma, Poisson, Gauss, ...

$$\mathscr{F}_{\alpha}^{(1 \text{ or } \alpha)}(m) = \begin{cases} \min_{P_1} D_{\alpha}(P_1||Q) \\ \text{subject to: } m = E_{P_1}[X] \text{ or } m = E_{P^*}[X] \end{cases},$$

Computation of entropies $\mathscr{F}_{\alpha}^{(1 \text{ or } \alpha)}(x)$ can then be carried in the following way:

Entropy functionals in special cases for Q

e.g. uniform, Bernoulli, gamma, Poisson, Gauss, ...

$$\mathscr{F}_\alpha^{(1 \text{ or } \alpha)}(m) = \begin{cases} \min_{P_1} D_\alpha(P_1||Q) \\ \text{subject to: } m = E_{P_1}[X] \text{ or } m = E_{P^*}[X] \end{cases},$$

Computation of entropies $\mathscr{F}_\alpha^{(1 \text{ or } \alpha)}(x)$ can then be carried in the following way:

(a) compute $Z_\nu(\gamma, m)$ for the reference measure Q considered,

Entropy functionals in special cases for Q

e.g. uniform, Bernoulli, gamma, Poisson, Gauss, ...

$$\mathscr{F}_{\alpha}^{(1 \text{ or } \alpha)}(m) = \left\{ \begin{array}{c} \min_{P_1} D_{\alpha}(P_1||Q) \\ \text{subject to: } m = E_{P_1}[X] \text{ or } m = E_{P^*}[X] \end{array} \right. ,$$

Computation of entropies $\mathscr{F}_{\alpha}^{(1 \text{ or } \alpha)}(x)$ can then be carried in the following way:

(a) compute $Z_{\nu}(\gamma, m)$ for the reference measure Q considered,

(b) solve (or approximate the solution to) $\frac{d}{d\gamma} Z_{\nu+1}(\gamma, m) = 0$ in terms of γ,

Entropy functionals in special cases for Q

e.g. uniform, Bernoulli, gamma, Poisson, Gauss, ...

$$\mathscr{F}_\alpha^{(1 \text{ or } \alpha)}(m) = \left\{ \begin{array}{c} \min_{P_1} D_\alpha(P_1||Q) \\ \text{subject to: } m = E_{P_1}[X] \text{ or } m = E_{P^*}[X] \end{array} \right. ,$$

Computation of entropies $\mathscr{F}_\alpha^{(1 \text{ or } \alpha)}(x)$ can then be carried in the following way:

(a) compute $Z_\nu(\gamma, m)$ for the reference measure Q considered,

(b) solve (or approximate the solution to) $\frac{d}{d\gamma} Z_{\nu+1}(\gamma, m) = 0$ in terms of γ,

(c) $\mathscr{F}_\alpha^{(1)}(m) = -\log Z_{\xi+1}(\gamma^*, m)$ and $\mathscr{F}_\alpha^{(\alpha)}(m) = -\log Z_{-\xi}(\gamma^*, m)$, where γ^* realizes the maximum of the function.

Entropy functionals in special cases for Q

e.g. uniform, Bernoulli, gamma, Poisson, Gauss, ...

$$\mathscr{F}_\alpha^{(1 \text{ or } \alpha)}(m) = \left\{ \begin{array}{c} \min_{P_1} D_\alpha(P_1||Q) \\ \text{subject to: } m = E_{P_1}[X] \text{ or } m = E_{P^*}[X] \end{array} \right. ,$$

Computation of entropies $\mathscr{F}_\alpha^{(1 \text{ or } \alpha)}(x)$ can then be carried in the following way:

(a) compute $Z_\nu(\gamma, m)$ for the reference measure Q considered,

(b) solve (or approximate the solution to) $\frac{d}{d\gamma} Z_{\nu+1}(\gamma, m) = 0$ in terms of γ,

(c) $\mathscr{F}_\alpha^{(1)}(m) = -\log Z_{\xi+1}(\gamma^*, m)$ and $\mathscr{F}_\alpha^{(\alpha)}(m) = -\log Z_{-\xi}(\gamma^*, m)$, where γ^* realizes the maximum of the function.

Limit case $\alpha \to 1$:

- $\nu = \xi + 1 \to -\infty$ or $\nu = -\xi \to +\infty$

Bernoulli reference

$$Q(x) = \beta\delta(x) + (1-\beta)\delta(x-1).$$

$$Z_{\nu+1}(\gamma, m) = \beta\,(1-m\gamma)^{\nu+1} + (1-\beta)\,(\gamma - m\gamma + 1)^{\nu+1}$$

$$\mathscr{F}^{(.)}_{\alpha\to 1}(x) = x\ln\left(\frac{x}{1-\beta}\right) + (1-x)\ln\left(\frac{1-x}{\beta}\right).$$

Fermi-Dirac entropy.

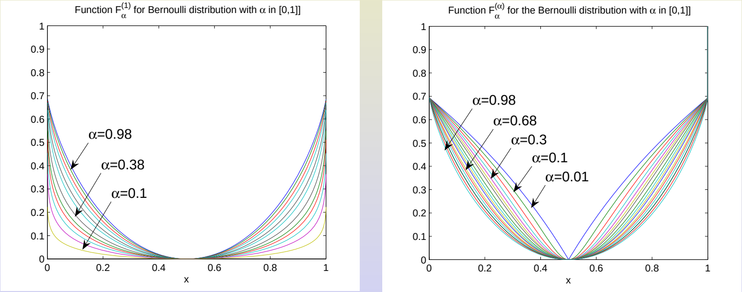


Figure: Entropy functionals $\mathscr{F}^{(1)}_{\alpha}(x)$ and $\mathscr{F}^{(\alpha)}_{\alpha}(x)$.

Other references

- Exponential reference (β): leads to a family of functions that converge to
$$(\beta x - 1) - \log(\beta x)$$
(Burg entropy for $\beta = 1$).
- Poisson reference (μ): Leads to a family of functions that converge to
$$x \ln \frac{x}{\mu} + (\mu - x)$$
cross-entropy between x and μ or Kullback-Leibler (Shannon) entropy functional with respect to μ.
- ...

The $\alpha \leftrightarrow 1/\alpha$ duality

- We will have pointwise equality of dual functions

$$-\log Z_{\xi_1+1}(\gamma,m) \text{ and } -\log Z_{-\xi_2}(\gamma,m)$$

if $\xi_1+1=-\xi_2$, that is if $\alpha_1=1/\alpha_2$.
In the general case, it can be checked that we *always* have the equality $D_{\frac{1}{\alpha}}(P^*||Q)=D_\alpha(P_1||Q)$ so that

$$\left\{\begin{array}{l}\inf_{P_1} D_\alpha(P_1||Q) \\ s.t\ \ E_{P^*}[X]=m\end{array}\right. = \left\{\begin{array}{l}\inf_{P^*} D_{\frac{1}{\alpha}}(P^*||Q) \\ s.t\ \ E_{P^*}[X]=m\end{array}\right. ,$$

so that generalized and classical mean constraints can always be swapped, if $\alpha \leftrightarrow 1/\alpha$, and

$$\mathscr{F}_\alpha^{(\alpha)}(x)=\mathscr{F}_{1/\alpha}^{(1)}(x).$$

Legendre structure

Entropies: general form $S = \log Z_{\nu+1}(\gamma, \bar{x})$.
We obtain the Euler formula:

$$\frac{dS}{d\lambda} = \frac{dS}{d\gamma}\frac{d\gamma}{d\lambda} = \lambda \frac{d\bar{x}}{d\lambda}.$$

The derivative of the entropy with respect to the mean is

$$\frac{dS}{d\bar{x}} = \frac{dS}{d\lambda}\frac{d\lambda}{d\bar{x}} = \lambda.$$

Massieu potential $\phi(\lambda) = S - \lambda\bar{x}$ ($\equiv$ free energy).

$$\frac{d\phi}{d\lambda} = -\bar{x}, \text{ and } \frac{d\phi}{d\bar{x}} = -\bar{x}\frac{d\lambda}{d\bar{x}}.$$

These four relations show that S and ϕ are conjugated with variables $\bar{x}$ and λ : $S\,[\bar{x}] \rightleftharpoons \phi\,[\lambda]$, so that the basic Legendre structure of thermodynamics is preserved.

Summary

- suggested a link between classical ME and maximization of Rényi-Tsallis entropy,
- derived expression of solutions,
- proposed numerical schemes,
- worked out special cases,

Todo

- aspects in practical computation of $\mathscr{F}_\alpha^{(1 \text{ or } \alpha)}(m)$,
- $\bar{x} \leftrightarrow \gamma$ mapping,
- extensions to many constraints m, θ
- ...