一般化エントロピーに基づく統計力学と情報幾何構造

和田達明 (T. Wada)

Ibaraki University

sul 11 genaio 2014 alla Hokudai

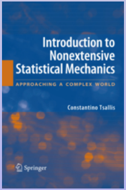
Outline

Motivation

一般化エントロピーに基づく統計力学 (の拡張版) を、情報幾何（Affine 微分幾何）の枠組みで捉え直し、幾何学的な理解を深める。

Outline

Nonextensive thermostatistics



C. Tsallis, *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World*, Springer, 2009.

Tsallis' generalized entropy:

$$S_q \equiv \frac{\sum_i p_i^q - 1}{1 - q} \quad \underset{q \to 1}{\longrightarrow} \quad S^{\mathrm{GS}} = -\sum_i p_i \ln p_i,$$

Deriving a Power law distribution based on the principle of MaxEnt

κ-entropy:

$$S_\kappa \equiv -\frac{1}{2\kappa} \sum_i \left(p_i^{1+\kappa} - p_i^{1-\kappa} \right).$$

- Another parameter (κ) extension of BGS entropy.
- it also generates power law tailed distributions.

introduced by G. Kaniadakis and A.M. Scarfone,
Politecnico di Torino, Italy.

G. Kaniadakis, Physica A **296**, 405 (2001).
G. Kaniadakis, Phys. Rev. E **66**, 056125 (2002).
G. Kaniadakis, A.M. Scarfone, Physica A **305**, 69 (2002).
G. Kaniadakis, Phys. Rev. E **72** 036108 (2005).

Outline

1. Introduction
   - Nonextensive thermostatistics
   - Generalized entropies and distributions
   - Thermostatistics
   - S_{2-q}-formalism and Legendre structures
2. Information geometry
   - Conormal map
   - dually-flat structures

Generalized entropies and distributions

A trace-form entropy functional

$$S_{\{\sigma\}}[p] = -\sum_i p_i \, \Lambda_{\{\sigma\}}(p_i) \, , \quad \underset{\sigma \to \sigma_0}{\Rightarrow} \quad S^{\mathrm{BG}}[p] = -\sum_i p_i \ln p_i \, ,$$

where $\Lambda_{\{\sigma\}}(x)$ is a generalized logarithm.

We require

- boundary condition: $\Lambda_{\{\sigma\}}(1) = 0$
- monotonic increasity: $\frac{d}{dx}\Lambda_{\{\sigma\}}(x) > 0$
- concavity $\frac{d^2}{dx^2}\Lambda_{\{\sigma\}}(x) < 0$ with $\int_0^1 \Lambda_{\{\sigma\}}(x) \, dx < \infty$

Generalized entropies and distributions

Finding $\Lambda_{\{\sigma\}}(x)$ by invoking the MaxEnt:

$$\frac{\partial}{\partial p_i}\Big(-\sum_j p_j \, \Lambda_{\{\sigma\}}(p_j) - \gamma \sum_j p_j - \beta \sum_j p_j \, E_j \Big) = 0,$$

$$\Rightarrow \; p_j^{eq} \equiv p_{\{\sigma\}}(E_j) \; .$$

We obtain the requirement:

$$\frac{d}{dx}\Big(x \, \Lambda_{\{\sigma\}}(x) \Big) = \lambda \, \Lambda_{\{\sigma\}} \left(\frac{x}{\alpha} \right) .$$

Two-parameter entropy

Kaniadakis, Lissia, Scarfone, PRE **71** (2005) 046128.

$$S_{\kappa,r} = -\sum_i p_i^{1+r} \left(\frac{p_i^{\kappa} - p_i^{-\kappa}}{2\kappa} \right).$$

- A generalization of BGS entropy:

$$\lim_{\kappa,r\to 0} S_{\kappa,r} = -\sum_i p_i \ln p_i$$

- equivalent to the entropy by Sharma-Taneja and Mittal in 1975 (information theory)

Special cases

- $\kappa = r = \frac{1-q}{2}$: Tsallis' entropy

$$S_q = \sum_i \frac{p_i^q - p_i}{1-q}$$

- $1 + \kappa^2 = (r+1)^2$, and $q_{\mathrm{A}} = r + \kappa + 1$: Abe's entropy

$$S_{q_{\mathrm{A}}} = \sum_i \frac{p_i^{\frac{1}{q_{\mathrm{A}}}} - p_i^{q_{\mathrm{A}}}}{\frac{1}{q_{\mathrm{A}}} - q_{\mathrm{A}}}$$

- $r = 0$: Kaniadakis' entropy

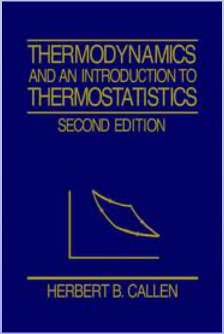
$$S_\kappa = \sum_i \frac{p_i^{1-\kappa} - p_i^{1+\kappa}}{2\kappa}$$

S. Abe et al., Phys. Lett. A 281 (2001) 126-130.

Outline

Generalized thermostatistics

H.B. Callen's book (John Wiley & Sons 1985)



From Chap. 21

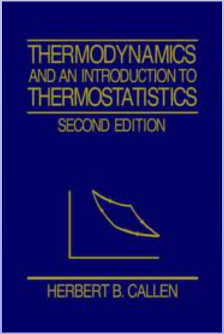
Thermostatistics characterizes the equilibrium state of microscopic systems without reference either to the specific forces or to the laws of mechanical response.

Instead thermostatistics characterizes the equilibrium state as the state that maximizes the disorder, a quantity associated with a conceptual framework ("information theory") outside of conventional physical theory.

the disorder $\Leftrightarrow$ Shannon entropy

Generalized thermostatistics

H.B. Callen's book (John Wiley & Sons 1985)



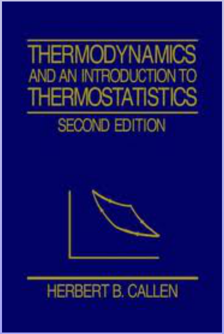
From Chap. 21

Thermostatistics characterizes the equilibrium state of microscopic systems without reference either to the specific forces or to the laws of mechanical response.
Instead thermostatistics characterizes the equilibrium state as the state that maximizes the disorder, a quantity associated with a conceptual framework ("information theory") outside of conventional physical theory.

the disorder $\Leftrightarrow$ Shannon entropy

Generalized thermostatistics

H.B. Callen's book (John Wiley & Sons 1985)



From Chap. 21

Thermostatistics characterizes the equilibrium state of microscopic systems without reference either to the specific forces or to the laws of mechanical response.
Instead thermostatistics characterizes the equilibrium state as the state that maximizes the disorder, a quantity associated with a conceptual framework ("information theory") outside of conventional physical theory.

the disorder $\Leftrightarrow$ Shannon entropy

Generalized thermostatistics

J. Naudts, Physica A **340** (2004) 32.
A generalization of Callen's thermostatistics based on a generalized entropy:

$$S_\phi = -\sum_i p_i \ln_\phi p_i \quad \underset{\phi(s)\to s}{\longrightarrow} \quad S^{\mathrm{BGS}},$$

$$\text{with} \quad \ln_\phi x \equiv \int_1^x \frac{ds}{\phi(s)}, \quad \underset{\phi(x)\to s}{\longrightarrow} \quad \ln x.$$

Examples:

- Tsallis: $\phi(s) = s^q$, $\quad \ln_q(x) = \frac{x^{1-q}-1}{1-q}$, $\quad q > 0$
- Kaniadakis:
  $\phi(s) = 2s/(s^\kappa + s^{-\kappa})$, $\quad \ln_{\{\kappa\}}(x) = \frac{x^\kappa - x^{-\kappa}}{2\kappa}$, $\quad -1 < \kappa < 1$

Generalized thermostatistics

J. Naudts, Physica A **340** (2004) 32.
A generalization of Callen's thermostatistics based on a generalized entropy:

$$S_\phi = -\sum_i p_i \ln_\phi p_i \quad \underset{\phi(s)\to s}{\longrightarrow} \quad S^{\mathrm{BGS}},$$

$$\text{with} \quad \ln_\phi x \equiv \int_1^x \frac{ds}{\phi(s)}, \quad \underset{\phi(x)\to s}{\longrightarrow} \quad \ln x.$$

Examples:

- Tsallis: $\phi(s) = s^q, \quad \ln_q(x) = \frac{x^{1-q}-1}{1-q}, \quad q > 0$
- Kaniadakis: $\phi(s) = 2s/(s^\kappa + s^{-\kappa}), \quad \ln_{\{\kappa\}}(x) = \frac{x^\kappa - x^{-\kappa}}{2\kappa}, \quad -1 < \kappa < 1$

Generalized thermostatistics

J. Naudts, Physica A **340** (2004) 32.
A generalization of Callen's thermostatistics based on a generalized entropy:

$$S_\phi = -\sum_i p_i \ln_\phi p_i \quad \underset{\phi(s)\to s}{\longrightarrow} \quad S^{\mathrm{BGS}},$$

$$\text{with} \quad \ln_\phi x \equiv \int_1^x \frac{ds}{\phi(s)}, \quad \underset{\phi(x)\to s}{\longrightarrow} \quad \ln x.$$

Examples:

- Tsallis: $\phi(s) = s^q, \quad \ln_q(x) = \frac{x^{1-q}-1}{1-q}, \quad q > 0$
- Kaniadakis:
  $\phi(s) = 2s/(s^\kappa + s^{-\kappa}), \quad \ln_{\{\kappa\}}(x) = \frac{x^\kappa - x^{-\kappa}}{2\kappa}, \quad -1 < \kappa < 1$

Outline

S_{2-q}-formalism

T. Wada, A.M. Scarfone, Phys. Lett. A **335** (2005) 351.

Starting from the q-exponential probability distribution:

$$p_i = \alpha_q \exp_q \left[-\gamma - \beta E_i\right],$$

where α_q is a q-dependent constant,

$$\alpha_q \equiv \left(\frac{1}{2-q}\right)^{1-q} = \frac{1}{\exp_q(1)}.$$

Useful relation:

$$\frac{d}{dx} x \ln_q x = \ln_q \frac{x}{\alpha_q}.$$

S_{2-q}-formalism

From the q-distribution, we have

$$-\ln_q \frac{p_i}{\alpha_q} - \beta E_i - \gamma = 0.$$

Utilizing the identity it follows

$$\frac{\partial}{\partial p_i}\Big(-\sum_j p_j \ln_q p_j - \beta \sum_j E_j p_j - \gamma \sum_j p_j \Big) = 0,$$

which is equivalent to the maximization of S_{2-q}

$$S_{2-q} = -\sum_i p_i \ln_q p_i,$$

under the constraints of

$$\sum_i E_i p_i = U \;, \quad \text{and} \quad \sum_i p_i = 1.$$

Legendre structures 1

The q-exponential distribution can be written as

$$p_i = \alpha_q \exp_q\left[-\beta E_i - \gamma\right] = \exp_q\left[-\frac{\beta}{2-q}E_i - \left(\frac{\gamma+1}{2-q}\right)\right]$$

Here we introduced

$$\beta^N \equiv \frac{\beta}{2-q}$$

$$\Phi_q^N \equiv \frac{\gamma+1}{2-q}, \qquad \text{generalized Massiue potential}$$

Legendre structures 1

$$p_i = \exp_q \left[-\beta^N E_i - \Phi_q^N \right].$$

By differentiating $\sum_i p_i = 1$ w.r.t. β^N, and using $d \exp_q(x)/dx = \exp_q(x)^q$, we have

$$0 = \sum_i \frac{dp_i}{d\beta^N} = -\sum_i \left(E_i + \frac{d\Phi_q^N}{d\beta^N}\right) p_i^q, \quad \Rightarrow \quad \frac{d\Phi_q^N}{d\beta^N} = -\frac{\sum_i E_i p_i^q}{\sum_j p_j^q}.$$

which leads to the Legendre relation:

$$\frac{d\Phi_q^N}{d\beta^N} = -U_q.$$

Note that the escort probabilities P_i are naturally appeared!

Legendre structures 1

The normalized Tsallis entropy

$$S_q^N = -\sum_i P_i \ln_q p_i.$$

Substituting

$$p_i = \exp_q\left[-\beta^N E_i - \Phi_q^N\right]$$

into S_q^N leads to

$$S_q^N = \sum_i P_i\left(\beta^N E_i + \Phi_q^N\right) = \beta^N U_q + \Phi_q^N.$$

S_q^N and Φ_q^N are Legendre duals each other.

Legendre structure 2

$$S_{2-q} = -\sum_i p_i \ln_q p_i, \quad p_i = \alpha_q \exp_q \left[-\gamma - \beta\, E_i\right].$$

Using the identity $\frac{d}{dx} x \ln_q x = \ln_q \frac{x}{\alpha_q}$, we have

$$\frac{dS_{2-q}}{d\beta} = -\sum_i \frac{dp_i}{d\beta}\frac{d}{dp_i} p_i \ln_q p_i = -\sum_i \frac{dp_i}{d\beta} \ln_q \frac{p_i}{\alpha_q}$$
$$= \sum_i \frac{dp_i}{d\beta}\left(\beta E_i + \gamma\right) = \beta \frac{dU}{d\beta}.$$

$$\frac{dS_{2-q}}{dU} = \beta.$$

Legendre structure 2

The q-exponential distribution

$$p_i = \exp_q\left[-\frac{\beta}{2-q}E_i - \left(\frac{1+\gamma}{2-q}\right)\right], \quad S_{2-q} = -\sum_i p_i \ln_q p_i.$$

$$\Rightarrow \quad \Phi_q^N = \frac{1+\gamma}{2-q} = S_{2-q} - \frac{\beta}{2-q}U.$$

$$\Phi_{2-q} \equiv \frac{1+\gamma}{2-q} - \left(\frac{1-q}{2-q}\right)\beta U.$$

$$\Phi_{2-q} = S_{2-q} - \beta U.$$

Legendre structure 2

$$\Phi_q^N = S_{2-q} - \frac{\beta}{2-q} U, \quad \Phi_{2-q} = \Phi_q^N - \left(\frac{1-q}{2-q}\right) \beta U.$$

$$\begin{aligned}
\frac{d\Phi_q^N}{d\beta} &= \frac{dS_{2-q}}{d\beta} - \frac{\beta}{2-q}\frac{dU}{d\beta} - \frac{U}{2-q} \\
&= \left(\frac{1-q}{2-q}\right) \beta \frac{dU}{d\beta} - \frac{1}{2-q} U. \\
\frac{d\Phi_{2-q}}{d\beta} &= \frac{d\Phi_q^N}{d\beta} - \left(\frac{1-q}{2-q}\right) \beta \frac{dU}{d\beta} - \left(\frac{1-q}{2-q}\right) U = -U.
\end{aligned}$$

Legendre structures (summary)

Two different sets:

$$\Phi_q^N \equiv \frac{1+\gamma}{2-q},$$

$$\Phi_q^N = S_q^N - \beta^N U_q, \quad \frac{d\Phi_q^N}{d\beta^N} = -U_q, \quad \frac{dS_q^N}{dU_q} = \beta^N.$$

$$\Phi_{2-q} \equiv \frac{1+\gamma}{2-q} - \left(\frac{1-q}{2-q}\right)\beta U,$$

$$\Phi_{2-q} = S_{2-q} - \beta U, \quad \frac{d\Phi_{2-q}}{d\beta} = -U, \quad \frac{dS_{2-q}}{dU} = \beta.$$

Information geometry

Amari and Nagaoka, *Methods of Information Geometry*, (AMS 2001)

N-dimensional probability simplex:

$$\mathcal{S}^n \equiv \left\{ \mathbf{p} = (p_i) \middle| p_i > 0, \sum_{i=1}^{n+1} p_i = 1 \right\}.$$

The natural basis tangent vector fields are

$$\partial_i \equiv \frac{\partial}{\partial p_i} - \frac{\partial}{\partial p_{n+1}}, \quad i = 1, \cdots, n.$$

Riemannian metric g on $\mathcal{S}^n$ is Fisher information matrix:

$$g_{ij}(\mathbf{p}) \equiv \frac{1}{p_i}\delta_{ij} + \frac{1}{p_{n+1}} = \sum_{k=1}^{n+1} p_k \left(\partial_i \ln p_k\right)\left(\partial_j \ln p_k\right),$$

$$i, j, k = 1, \cdots, n.$$

A dually-flat structure in escort distributions

Ohara, Matsuzoe, Amari, J. Phys.: Conf. Series **201** (2010) 012012.

The α-covariant derivative is given by

$$\nabla^{(\alpha)}_{\partial_i}\partial_j = \sum_{k=1}^{n} \Gamma^{(\alpha)k}_{ij}, \quad i,j = 1,\cdots,n,$$

$$\text{where} \quad \Gamma^{(\alpha)k}_{ij}(\mathbf{p}) = \frac{1+\alpha}{2}\left(-\frac{1}{p_k}\delta^k_{ij} + p_k g_{ij}\right).$$

α-immersion

The α-immersion f of $\mathcal{S}^n$ into $\mathbb{R}_+^{n+1}$ with $q = (1-\alpha)/2$:

$$f : \mathbf{p} = (p_i) \mapsto \mathbf{x} = (x^i) = L^{(\alpha)}(p_i) = \frac{p_i^q}{q}.$$

$f(\mathcal{S}^n)$ is a level hypersurface in $\mathbb{R}_+^{n+1}$:

$$\Psi(\mathbf{x}) = \frac{1}{1-q}\sum_{i=1}^{n+1}(qx_i)^{\frac{1}{q}}.$$

Choosing a transversal vector ξ on the level hypersurface by

$$\xi \equiv \sum_{i=1}^{n+1}\xi^i\frac{\partial}{\partial x^i}, \quad \xi^i = -q(1-q)x^i = -\kappa x^i,$$

then the affine immersion (f, ξ) realizes the α-geometry on $\mathcal{S}^n$.

Outline

Conormal map

Introducing the dual vector $\mathbf{x}^*$ by

$$x_i^*(\mathbf{p}) = L^{(-\alpha)}(p_i) = \frac{1}{1-q} p_i^{1-q},$$

which satisfies

$$x_i^*(\mathbf{p}) = \frac{\partial \Psi}{\partial x^i}(\mathbf{x}(\mathbf{p})),$$

$$-\sum_{i=1}^{n+1} \xi^i(\mathbf{p}) x_i^*(\mathbf{p}) = 1, \quad \sum_{i=1}^{n+1} x_i^*(\mathbf{p})\, X^i = 0,$$

for an arbitrary vector $X = \sum_i X^i \partial/\partial x^i$ at $\mathbf{x}(\mathbf{p})$ tangent to $f(\mathcal{S}^n)$. Hence $-\mathbf{x}^*(\mathbf{p})$ is a conormal map.

The escort probability is

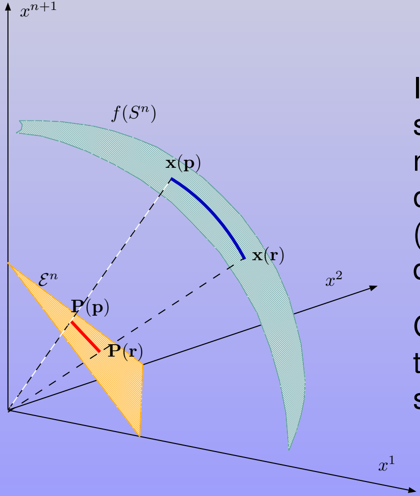
$$P_i(\mathbf{p}) = \frac{x^i}{Z_q}, \quad Z_q(\mathbf{p}) \equiv \sum_{i=1}^{n+1} x^i(\mathbf{p}) = \sum_{i=1}^{n+1} \frac{p_i^q}{q},$$
$$\lambda(\mathbf{p}) \equiv \frac{1}{Z_q(\mathbf{p})}.$$

The simplex

$$\mathcal{E}^n \equiv \left\{ \mathbf{x} = (x^i) \middle| x^i > 0, \sum_{i=1}^{n+1} x^i = 1 \right\}.$$

Outline

Conformal transformation



Introducing another immersion $\tilde{f} \equiv \lambda f$ and Riemannian metric $h \equiv \lambda g$, we can obtain a statistical manifold $(\mathcal{S}^n, h, \nabla, \nabla^*)$, which has a dually flat structure.

Conformal transformation to the dually-flat structure on the space of escort distributions.

Using the conormal map we can define the α-divergence as a contrast function on $(\mathcal{S}^n, g, \nabla^{(\alpha)})$:

$$D^{(\alpha)}(\mathbf{p},\mathbf{r}) = -\sum_{i=1}^{n+1} x_i^*(\mathbf{r})\big(x^i(\mathbf{p}) - x^i(\mathbf{r})\big) = \frac{1}{q}\sum_{i=1}^{n+1} r_i \ln_q \frac{p_i}{r_i}$$

$$= \frac{1}{q}\sum_{i=1}^{n+1} r_i \left(-\ln_{2-q}\frac{r_i}{p_i}\right). \quad \text{f-divergence}$$

The conformal divergence is a contrast function of $(\mathcal{S}^n, h, \nabla)$:

$$\rho(\mathbf{p},\mathbf{r}) = \lambda(\mathbf{r})D^{(-\alpha)}(\mathbf{p},\mathbf{r}) = \langle \mathbf{P}(\mathbf{r}), \mathbf{x}^*(\mathbf{p}) - \mathbf{x}^*(\mathbf{r})\rangle$$

$$= \sum_{i=1}^{n+1} P_i\big(\ln_q p_i - \ln_q r_i\big).$$

Using the conormal map we can define the α-divergence as a contrast function on $(\mathcal{S}^n, g, \nabla^{(\alpha)})$:

$$D^{(\alpha)}(\mathbf{p},\mathbf{r}) = -\sum_{i=1}^{n+1} x_i^*(\mathbf{r})\big(x^i(\mathbf{p}) - x^i(\mathbf{r})\big) = \frac{1}{q}\sum_{i=1}^{n+1} r_i \ln_q \frac{p_i}{r_i}$$

$$= \frac{1}{q}\sum_{i=1}^{n+1} r_i \left(-\ln_{2-q}\frac{r_i}{p_i}\right). \quad \text{f-divergence}$$

The conformal divergence is a contrast function of $(\mathcal{S}^n, h, \nabla)$:

$$\rho(\mathbf{p},\mathbf{r}) = \lambda(\mathbf{r})D^{(-\alpha)}(\mathbf{p},\mathbf{r}) = \langle \mathbf{P}(\mathbf{r}), \mathbf{x}^*(\mathbf{p}) - \mathbf{x}^*(\mathbf{r})\rangle$$

$$= \sum_{i=1}^{n+1} P_i\big(\ln_q p_i - \ln_q r_i\big).$$

Using the conormal map we can define the α-divergence as a contrast function on $(\mathcal{S}^n, g, \nabla^{(\alpha)})$:

$$D^{(\alpha)}(\mathbf{p}, \mathbf{r}) = -\sum_{i=1}^{n+1} x_i^*(\mathbf{r})\big(x^i(\mathbf{p}) - x^i(\mathbf{r})\big) = \frac{1}{q}\sum_{i=1}^{n+1} r_i \ln_q \frac{p_i}{r_i}$$

$$= \frac{1}{q}\sum_{i=1}^{n+1} r_i \left(-\ln_{2-q} \frac{r_i}{p_i}\right). \quad \text{f-divergence}$$

The conformal divergence is a contrast function of $(\mathcal{S}^n, h, \nabla)$:

$$\rho(\mathbf{p}, \mathbf{r}) = \lambda(\mathbf{r}) D^{(-\alpha)}(\mathbf{p}, \mathbf{r}) = \langle \mathbf{P}(\mathbf{r}), \mathbf{x}^*(\mathbf{p}) - \mathbf{x}^*(\mathbf{r})\rangle$$

$$= \sum_{i=1}^{n+1} P_i\big(\ln_q p_i - \ln_q r_i\big).$$

The mutually dual affine coordinate systems:

$$\eta_i = \frac{\partial \Psi}{\partial \theta^i}, \quad \theta^i = \frac{\partial \Psi^*}{\partial \eta_i}, \quad i = 1, \cdots, n,$$

for the potential function $\Psi(\theta)$ and its conjugate $\Psi^*(\eta)$.

An important basic concepts in information geometry is:

- α-representation (or α-immersion) of probability distributions,

$$\ell^{\alpha}(p_i) = \frac{2}{1-\alpha} p_i^{\frac{1-\alpha}{2}} = \frac{p_i^q}{q},$$

$$\ell^{-\alpha}(p_i) = \frac{2}{1+\alpha} p_i^{\frac{1+\alpha}{2}} = \frac{p_i^{1-q}}{1-q},$$

with $\alpha = 1 - 2q$.

dually-flat structures