Statistical physics models belonging to the generalised exponential family

Jan Naudts

Universiteit Antwerpen

1. The generalized exponential family
2. Theorem
3. Thermodynamic relations and duality
4. Generalised Fisher information
5. q-Gaussians
6. The porous media equation
7. The microcanonical ensemble
8. One spin of a $d = 1$-Ising model
9. Final remarks

WHY? Basic tool of statistical mechanics: calculate partition sum Z,
calculate averages by taking derivatives of Massieu's function $\ln Z$.
This works for models of the (generalised) exponential family.

1. The generalised exponential family

A *statistical model* is a probability distribution $p_\theta(x)$
which depends on one or more parameters $\theta_1, \theta_2, \cdots, \theta_n$.

For example $\quad p_{\sigma,a}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-a)^2/\sigma^2} \quad$ (the Gaussian model).

A model belongs to the *exponential family* if it can be written into the form

$$p_\theta(x) = \frac{c(x)}{Z(\theta)} e^{-\sum_j \theta_j H_j(x)}.$$

In *statistical physics,* the parameters are often the inverse temperature β
plus e.g. the external magnetic field, the $H_j(x)$ are Hamiltonians.
The Boltzmann-Gibbs distribution $p_\beta(x) = \frac{1}{Z} e^{-\beta H(x)}$, seen as a function of the
inverse temperature β, belongs to the exponential family.

The *statistical manifold* is the log-likelihood function $\theta \rightarrow \ln p_\theta(x)$
plus some additional structure to make it into a Riemannian manifold.

For a model belonging to the exponential family it is flat:

$$\ln p_\theta(x) = \ln c(x) - \ln Z(\theta) - \sum_j \theta_j H_j(x).$$

The Gaussian model is *curved* because a substitution of parameters is needed to make it *flat*:

$$\ln p_{\sigma,a}(x) = -\frac{1}{2}\ln 2\pi\sigma^2 - \frac{(x-a)^2}{\sigma^2}.$$

$$\begin{array}{ll} \theta_1 = 2a\sigma^{-2} & \qquad H_1(x) = x \\ \theta_2 = \sigma^{-2} & \qquad H_2(x) = x^2 \end{array}$$

Generalisation Fix a strictly increasing function $\Lambda(u)$ defined on $(0, +\infty)$.

A model $p_\theta(x)$ belongs to the *generalised exponential family* if real $\alpha(\theta)$ and $H_j(x)$ exist such that either $p_\theta(x) = 0$ or

$$\Lambda\left(\frac{p_\theta(x)}{c(x)}\right) = -\alpha(\theta) - \sum_j \theta_j H_j(x).$$

The exceptional treatment of $p_\theta(x) = 0$ is important! $\Lambda(0) >$r.h.s. occurs!

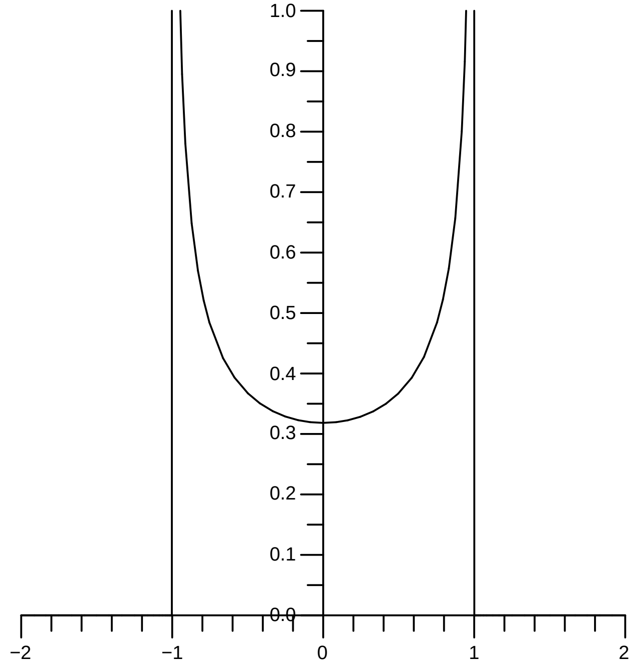
Correspondingly, the *generalised statistical manifold* is the map

$$\theta \to \Lambda\left(p_\theta(x)/c(x)\right).$$

It is defined for those θ for which $p_\theta(x) \neq 0$.
The function Λ is a *deformed logarithm*.

JN (2004, 2008), Grünwald and Dawid (2004), Eguchi (2004)

Example



The distribution of velocities of a harmonic oscillator is given by

$$p(v) = \frac{1}{\pi}\frac{1}{\sqrt{v_0^2 - v^2}}.$$

Let $\Lambda(u) = \frac{1}{2}(1 - u^{-2})$ and $c(v) = \dfrac{\sqrt{2}}{\pi|v|}$.

Then $$\Lambda\left(\frac{p(v)}{c(v)}\right) = -\frac{1}{2} + \frac{v^2}{v_0^2}.$$

Note: $p(v) = 0$ holds for $|v| > |v_0|$.

Note: This is an example of a non-extensive model with $q = 3$.

Universiteit Antwerpen

2. Theorem

Introduce a generalised entropy functional $I(p)$ by

$$I(p) = -\int \mathrm{d}x \, c(x) \, F\left(\frac{p_\theta(x)}{c(x)}\right), \qquad \text{with } F(u) = \int_1^u \mathrm{d}v \, \Lambda(v) + A$$

Note that $F(u)$ is strictly convex.

Theorem

The model $p_\theta(x)$ belongs to the generalised exponential family if and only if it satisfies the *variational principle*: for any pdf p is

$$I(p_\theta) - \sum_j \theta_j \langle p_\theta | H_j \rangle \geq I(p) - \sum_j \theta_j \langle p | H_j \rangle .$$

This theorem generalises (Ruelle, 1967).

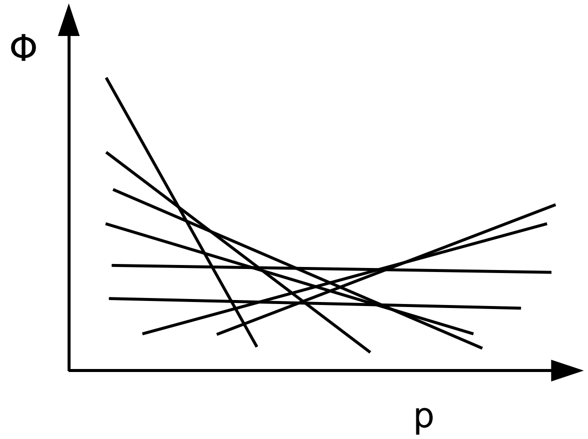
It is known to physicists as "The free energy is minimal in equilibrium".

The proof that $p_\theta(x)$ satisfies the variational principle uses positivity of the Bregman-type of divergence

$$\begin{aligned} D(p||p^*) &= I(p^*) - I(p) - \int \mathrm{d}x\, [p(x) - p^*(x)]\, \Lambda\left(\frac{p^*(x)}{c(x)}\right) \\ &= \int \mathrm{d}x\, c(x) \int_{p^*(x)/c(x)}^{p(x)/c(x)} \mathrm{d}u\, \left[\Lambda(u) - \Lambda\left(\frac{p^*(x)}{c(x)}\right)\right] \end{aligned}$$

The proof in the other direction is standard, but is complicated by the possibility that $p_\theta(x)$ may vanish.

3. Thermodynamic relations and duality



The function $\Phi(\theta) = I(p_\theta) - \sum_j \theta_j \langle p_\theta | H_j \rangle$

is the contact transformation of the entropy functional $I(p)$.

(This follows from the variational principle.)

Hence it is a convex function.

It satisfies $\dfrac{\partial \Phi}{\partial \theta_j} = -U_j$ with $U_j = \langle p_\theta | H_j \rangle$.

Let $S(U) \equiv I(p_\theta)$.

Then $\Phi(\theta) = S(U) - \sum_j \theta_j U_j = \sup_{U'} \{ S(U') - \sum_j \theta_j U'_j \}$

is a Legendre-Fenchel transform.

The inverse Legendre transform is

$$S(U) = \Phi(\theta) + \sum_j \theta_j U_j = \inf_{\theta'} \{\Phi(\theta') + \sum_j \theta'_j U_j\}.$$

It automatically satisfies $\quad \dfrac{\partial S}{\partial U_j} = \theta_j.$

In thermodynamics, the Legendre transform $\Phi(\theta)$ of $S(U)$ is Massieu's function. Introduced around 1860, shortly after Carnot invented his principle.

The formula $\dfrac{\mathrm{d}S}{\mathrm{d}U} = \beta = \dfrac{1}{T}$ is often used as the definition of temperature T.

The dual relations $\quad \dfrac{\partial \Phi}{\partial \theta_j} = -U_j, \quad \dfrac{\partial S}{\partial U_j} = \theta_j \quad$ generalise the geometric duality discussed by Amari in 1985.

4. Generalised Fisher information

A *theoretical* physicist assumes to know the parameters θ_j of the model and tries to calculate quantities like $U_j \equiv \langle p_\theta | H_j \rangle$.

An *experimental* physicist measures quantities such as U_j and makes a best guess of the model parameters θ_j.

How much information about θ_j is contained in U_j?

It is high when $U_j = \langle p_\theta | H_j \rangle$ varies strongly with θ.

Note that $\dfrac{\partial \Phi}{\partial \theta_j} = -U_j$. This is the tangent plane condition.

Hence, information content depends on the metric tensor $g_{i,j}(\theta) = \dfrac{\partial^2 \Phi}{\partial \theta_i \partial \theta_j}$.

A classical result is that the Fisher information matrix $I(\theta)$ equals the metric tensor $g(\theta)$. In addition, models belonging to the exponential family optimise the inequality of Cramér and Rao. These results can be generalised.

Let $$I_{ij}(\theta) = \int \mathrm{d}x\, c(x)\Lambda'\left(\frac{p_\beta(x)}{c(x)}\right)\left(\frac{\partial}{\partial\theta_i}\frac{p_\beta(x)}{c(x)}\right)\left(\frac{\partial}{\partial\theta_j}\frac{p_\beta(x)}{c(x)}\right).$$

The derivative of $$\Lambda\left(\frac{p_\theta(x)}{c(x)}\right) = -\alpha(\theta) - \sum_j \theta_j H_j(x)$$

is $$\Lambda'\left(\frac{p_\theta(x)}{c(x)}\right)\left(\frac{\partial}{\partial\theta_i}f_\beta(x)\right) = -\frac{\partial\alpha}{\partial\theta_i} - H_i(x).$$

Hence, $$\begin{aligned} I_{ij}(\theta) &= \int \mathrm{d}x\, c(x)\left[-\frac{\partial\alpha}{\partial\theta_i} - H_i(x)\right]\left(\frac{\partial}{\partial\theta_j}p_\beta(x)\right) \\ &= -\frac{\partial}{\partial\theta_j}\langle p_\theta | H_i\rangle = g_{i,j}(\theta). \end{aligned}$$

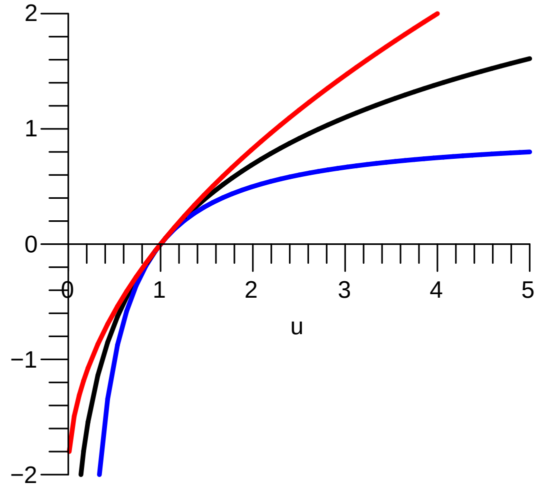
5. q-Gaussians

An important subclass is formed by the models belonging to the q-*exponential family.* These are the models of nonextensive thermostatistics (Tsallis 1988).

They correspond with the q-deformed logarithm $\Lambda(u) = \frac{1}{1-q}(u^{1-q} - 1) \equiv \ln_q(u)$.

$q = 1$ is the natural logarithm $\Lambda_1(u) = \ln u$.

The inverse function is the deformed exponential $\exp_q(u) = [1 + (1-q)u]_+^{1/(1-q)}$.



Note that the range of $\ln_q(u)$ is *not* all of $\mathbb{R}$.

The multivariate q-Gaussian is defined by

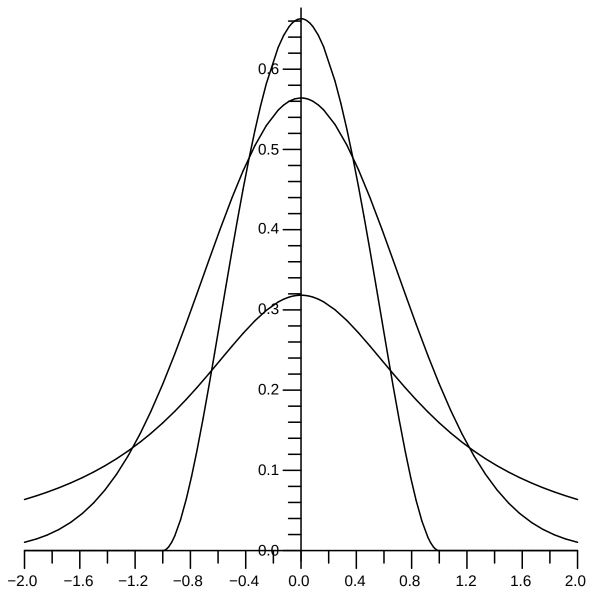
$$p(\mathbf{x}) = \frac{1}{c_q}(\det\Sigma)^{1/2}\exp_q\left(-(\mathbf{x}-\mathbf{a})^{\mathrm{T}}\Sigma(\mathbf{x}-\mathbf{a})\right).$$

Σ is the covariance matrix.

$\mathbf{a}$ is the position of the maximum.

The normalisation c_q does not depend on $\mathbf{a}$ or Σ.

Drawing for 1 variable and $q = 2, 1, 0.5$.
Note the cutoff for $q < 1$,
the fat tail for $q > 1$.
$q < 3$ is needed because $p(\mathbf{x}) \sim |x|^{\frac{2}{q-1}}$.



The q-Gaussian belongs to the q-exponential family,
with new parameters replacing $\mathbf{a}$ and Σ.

Interest in q-Gaussians: a central limit theorem for correlated systems.

(Moyano, Tsallis, Gell-Mann, 2006) brought numerical evidence for a central limit theorem with convergence towards a q-Gaussian distribution in a simple model of correlated variables.

(Hilhorst, Schehr, 2007) showed by analytic computation that in the given example the limiting distribution slightly deviates from the q-Gaussian.

(Vignat, Plastino, 2007) showed that if $X_1, X_2, \cdots$ are iid with $\mathbb{E}X = 0$ and $\mathbb{E}XX^{\mathrm{T}} = K$ and Y is chi-distributed with m degrees of freedom then

$$Z_n = \frac{\frac{1}{\sqrt{n}}\sum_{j=1}^{n} X_j}{Y}$$

converges weakly to a q-Gaussian with covariance matrix K and with $q = 1 + \frac{2}{m+N}$.

6. The porous media equation (PME)

based on "A. Ohara and T. Wada, Geometric aspects of a certain type of nonlinear diffusion equation, unpublished (2008)," and papers cited there.

The PME $\quad \dfrac{\partial u}{\partial t} = \Delta_{\mathbf{x}} u^m$, $m > 1$, $x \in \mathbb{R}^N$,
is a generalised diffusion equation, describing the flow of gas in a porous medium.

It has scaling solutions, known as Barenblatt-Pattle solutions (Barenblatt, 1952).

$$u_\gamma(\mathbf{x}) = \gamma^N \left[c^2 - \frac{m-1}{2m}\gamma^2|\mathbf{x}|^2 \right]_+^{1/m-1} = \gamma^N c^{2/(m-1)} \exp_q\left(-\frac{\gamma^2}{2mc^2}|\mathbf{x}|^2 \right),$$

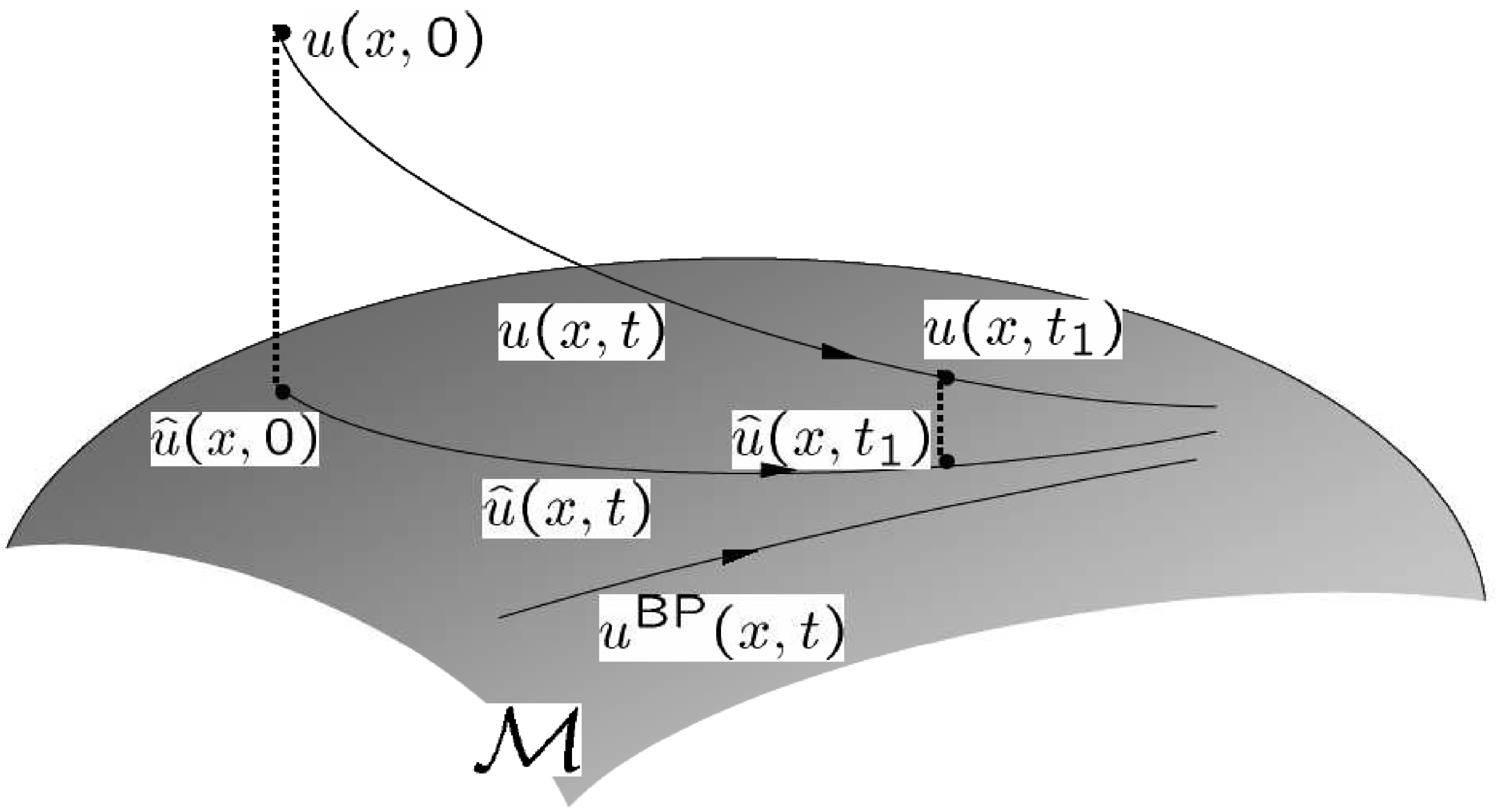
with $\gamma = \dfrac{1}{(\mu + \lambda t)^{1/\lambda}}$ and $q = 2 - m$. This is clearly a q-Gaussian.

Many results are known about this equation, for instance existence and uniqueness of the solutions, and asymptotic convergence to the Barenblatt solutions.

See (Vázquez, 2003) and (Toscani, 2005)

Main result of (Ohara, Wada, 2008): the projection of $u(x, t)$ onto the statistical manifold of q-Gaussian distributions follows a geodesic.

Previous result (Otto, 2001) involved infinite dimensional manifold and analysis involving Brownian motion.



Universiteit Antwerpen

7. The microcanonical ensemble

Consider a classical particle with Hamiltonian $H(\mathbf{x},\mathbf{p}) = T + V(\mathbf{x})$ with $T = \frac{1}{2m}|p|^2$.

The microcanonical density distribution is $f_U(\mathbf{x},\mathbf{p}) = \dfrac{1}{\rho(U)}\delta(H(\mathbf{x},\mathbf{p}) - U)$.

$\rho(U)$ is the density of states.

The configurational density distribution $f_U(\mathbf{x})$ is obtained by integrating out the momenta $\mathbf{p}$. One finds

$$f_U(\mathbf{x}) = 2mh^{-1}\sqrt{2\pi m}\,\exp_{-1}\left(-\frac{1}{2} + \frac{U}{2\rho^2} - \frac{1}{2\rho^2}V(\mathbf{x})\right).$$

$\exp_{-1}(u) = \sqrt{[1+2u]_+}$ (q-deformed exponential with $q = -1$).

This distribution belongs to the q-exponential family, with $q = -1$ and $H(\mathbf{x}) = V(\mathbf{x})$.

The measured values of $\langle f_U | V \rangle$ can be used to estimate $\theta = \dfrac{1}{2\rho^2}$.

Assume that the density of states $\rho(U)$ is strictly increasing. Then the model parameter U is uniquely determined by θ: $U = \rho^{-1}\left(\frac{1}{\sqrt{2\theta}}\right)$.

Experiments measure kinetic energy T. Note that $\langle f_U | V \rangle = U - \langle f_U | T \rangle$. Hence, $\langle f_U | T \rangle$ can be used to estimate U, given the knowledge of $\rho(U)$.

8. One spin of a $d = 1$-Ising model

The probability distribution of a single spin of the $d = 1$-Ising model is

$$p_\beta(\pm) = \frac{1}{2}\left(1 \pm \frac{y}{\sqrt{1+y^2}}\right) \quad \text{with} \quad y = \sinh(\beta\epsilon)e^{\beta J}$$

ϵ and J are fixed constants; assume $\epsilon \geq 0$ and $J > 0$; then $+$ is the ground state.

The internal energy equals $U = \langle p_\beta | H \rangle = -\epsilon(p_\beta(+) - p_\beta(-))$ with $H(\pm) = \mp\epsilon$.

Does there exist $\Lambda(u)$ increasing and $\alpha(\beta)$, such that

$$\Lambda\left(p_\beta(\pm)\right) = -\alpha(\beta) - \beta H(\pm)?$$

Does it satisfy $\dfrac{\mathrm{d}S}{\mathrm{d}U} = \beta$?

... automatically! The temperature of the single spin is that of the whole chain.

Universiteit Antwerpen

Define $\lambda(u)$ by $\lambda(p_\beta(-)) = \epsilon\beta$. Note that $\lambda(1/2) = 0$.

Try $\Lambda(u) = -\ln 2(\cosh(\lambda(u)) - \lambda(u)$, $0 \leq u \leq \frac{1}{2}$.

For $\frac{1}{2} \leq u \leq 1$ let $\Lambda(u) = \Lambda(1-u) + 2\lambda(1-u)$

Then $\Lambda\left(p_\beta(\pm)\right) = -\ln 2\cosh(\epsilon\beta) - g(\epsilon\beta) \mp \epsilon\beta$.

This is of the form $-\alpha(\beta) - \beta H(\pm)$.

Finally note that $\Lambda(u)$ is strictly increasing.

a) $0 \leq u \leq \frac{1}{2}$: $\dfrac{\mathrm{d}\Lambda}{\mathrm{d}u} = [-\tanh(\lambda) - 1]\dfrac{\mathrm{d}\lambda}{\mathrm{d}u} > 0$.

b) $\frac{1}{2} \leq u \leq 1$: $\dfrac{\mathrm{d}\Lambda}{\mathrm{d}u} = [\tanh(\lambda) - 1]\dfrac{\mathrm{d}\lambda}{\mathrm{d}u} > 0$.

J. Naudts, *Generalized thermostatistics and mean-field theory,* Physica A332, 279 – 300 (2004).

9. Final remarks

Other examples: persistent random walk, percolation model