Journal of Statistical Physics, Vol. 16, No. 4, 1977

Statistical Inference and Entropy

Jean-Paul Marchand[1] **and Walter Wyss**[2]

Received September 28, 1976

We construct an entropy function such that statistical inference with respect to a partial measurement and a given a priori distribution is characterized by maximal entropy.

KEY WORDS: Statistical inference; noncommutative probability; entropy.

1. INTRODUCTION

We discuss the relation between statistical inference and entropy for a system represented by a von Neumann algebra $\mathfrak{A}$ of bounded operators on a separable Hilbert space. The Hermitian elements of $\mathfrak{A}$ are interpreted as the observables of the system and the normal states on $\mathfrak{A}$ represent physical information. An a priori distribution on $\mathfrak{A}$ is a normal state reflecting the information available prior to any actual measurement; it contains the macroscopic information about the system. A partial measurement is the outcome of a microscopic measurement on a subsystem of $\mathfrak{A}$; the subsystem is thus represented by a von Neumann subalgebra $\mathfrak{B}$ of $\mathfrak{A}$. The problem of statistical inference is to determine the most likely physical information about the system $\mathfrak{A}$ conditioned on a partial measurement on $\mathfrak{B}$ and a given a priori distribution v on $\mathfrak{A}$; this procedure is called $(\mathfrak{B}, v)$ inference. Applications to problems in classical and quantum physics may be found in Ref. 1. In this paper we construct the entropy function that characterizes the $(\mathfrak{B}, v)$ inference as a maximum entropy state.

2. PARTIAL MEASUREMENT AND $(\mathfrak{B}, v)$ INFERENCE

Let $\mathfrak{R}$ be a von Neumann algebra of bounded operators on a separable Hilbert space $\mathfrak{H}$. $F(\mathfrak{R})$ stands for the faithful normal states on $\mathfrak{R}$, $\mathfrak{R}_+$ for the positive elements in $\mathfrak{R}$, and $J(\mathfrak{R}) = \{T \in \mathfrak{R}_+ ; \exists T^{-1} \in \mathfrak{R}_+\}$.

[1] Department of Mathematics and Physics, University of Denver, Denver, Colorado.

[2] Department of Physics and Astrophysics, University of Colorado, Boulder, Colorado.

Definition 2.1. Let $\varphi, \psi \in F(\mathfrak{N})$ and $\lambda, \mu > 0$. φ is (λ, μ) *comparable with* ψ if $\varphi \leqslant \lambda\psi$ and $\psi \leqslant \mu\varphi$.

Lemma 2.1. Let φ be (λ, μ) comparable with ψ. Then there is a unique $T \in J(\mathfrak{N})$ such that $\varphi(N) = \psi(TNT)$, $\psi(N) = \varphi(T^{-1}NT^{-1})$, $\forall N \in \mathfrak{N}$, and $\|T\|^2 \leqslant \lambda$, $\|T^{-1}\|^2 \leqslant \mu$.

Proof. $\varphi \leqslant \lambda\psi$ implies $\lambda \geqslant 1$ and according to Ref. 2 there is a unique $H_1 \in \mathfrak{N}_+$ such that $\varphi(N) = (\lambda\psi)(H_1NH_1)$, $0 \leqslant H_1 \leqslant 1$. With $T_1 = \sqrt{\lambda}H_1$, we get $\varphi(N) = \psi(T_1NT_1)$, $T_1 \in \mathfrak{N}_+$, $0 \leqslant T_1 \leqslant \sqrt{\lambda}$ and $\|T_1\|^2 \leqslant \lambda$. Similarly, $\psi \leqslant \mu\varphi$ implies the unique existence of $T_2 \in \mathfrak{N}_+$ such that $\psi(N) = \varphi(T_2NT_2)$, $\|T_2\|^2 \leqslant \mu$. We now show that T_1 and T_2 are invertible. Suppose they are not invertible; then they have nontrivial null spaces. Let P_1, P_2 be the projectors onto these subspaces. Since $T_1, T_2 \in \mathfrak{N}_+$, it follows that $P_1, P_2 \in \mathfrak{N}_+$. Then $\varphi(P_1) = \psi(T_1P_1T_1) = 0$ and $\psi(P_2) = \varphi(T_2P_2T_2) = 0$. This, however, contradicts the faithfulness of φ and ψ, meaning that T_1 and T_2 are invertible. Let now

$$T_1 = \int_0^\infty x\, dE_1(x), \qquad T_2 = \int_0^\infty x\, dE_2(x)$$

be the spectral representations of T_1 and T_2.

From

$$(1/\lambda)\varphi \leqslant \psi \qquad \text{and} \qquad (1/\mu)\psi \leqslant \varphi$$

we get

$$\begin{aligned}(1/\lambda)\varphi(E_2(y_2)) \leqslant \psi(E_2(y_2)) &= \varphi(T_2E_2(y_2)T_2) \\ &= \int_0^\infty x^2\, d\varphi(E_2(x)E_2(y_2)) \\ &= \int_0^{y_2} x^2\, d\varphi(E_2(x)) \leqslant y_2{}^2\varphi(E_2(y_2))\end{aligned}$$

Similarly, $(1/\mu)\psi(E_1(y_1)) \leqslant y_1{}^2\psi(E_1(y_1))$. This shows that the spectra of T_1 and T_2 are bounded away from zero and hence T_1^{-1} and T_2^{-1} are bounded. Then $\varphi(T_1^{-1}NT_1^{-1}) = \psi(N)$ and $\psi(T_2^{-1}NT_2^{-1}) = \varphi(N)$ and the uniqueness of T_1 and T_2 implies $T_1 = T_2^{-1}$ and $T_2 = T_1^{-1}$.

The concept of $(\mathfrak{B}, v)$ inference involves two von Neumann algebras $\mathfrak{A}, \mathfrak{B}$ with $\mathfrak{B} \subset \mathfrak{A} \subset \mathfrak{B}(\mathfrak{H})$, an a priori state v on $\mathfrak{A}$, and a state $w_{\mathfrak{B}}$ on $\mathfrak{B}$ reflecting the outcome of a partial measurement. Without loss of information we assume $v \in F(\mathfrak{A})$ and $w_{\mathfrak{B}} \in F(\mathfrak{B})$, thus avoiding redundant events.

Definition 2.2. Let $\mathfrak{A}, \mathfrak{B}, v, w_{\mathfrak{B}}$ be given such that the restriction $v|_{\mathfrak{B}}$ of v to $\mathfrak{B}$ is (λ, μ) comparable with $w_{\mathfrak{B}}$. According to Lemma 2.1, there is a unique $T \in J(\mathfrak{B})$ with

$$w_{\mathfrak{B}}(B) = v|_{\mathfrak{B}}(TBT) = v(TBT), \qquad \forall B \in \mathfrak{B}$$

The $(\mathfrak{B}, v)$ *inference w of* $w_{\mathfrak{B}}$ is then defined by

$$w(A) = v(TAT), \qquad \forall A \in \mathfrak{A}$$

Remark 2.1. The $(\mathfrak{B}, v)$ coarse-graining w of $w_{\mathfrak{B}}$ has the following properties:

(a) $w \in F(\mathfrak{A})$.
(b) w is an extension of $w_{\mathfrak{B}}$.
(c) If $w_{\mathfrak{B}} = v|_{\mathfrak{B}}$, then $w = v$.
(d) If $\mathfrak{B} = \{\lambda \cdot 1\}$, then $w = v$.

Remark 2.2. $(\mathfrak{B}, v)$ inference on the basis of global a priori information and partial microscopic measurement is compatible with the information contained in $w_{\mathfrak{B}}$ [property (b)] and such that the a priori state v is recovered if either the partial information coincides with the information already contained in v [property (c)] or if no partial measurement is made [property (d)].

3. THE ENTROPY CHARACTERIZING $(\mathfrak{B}, v)$ INFERENCE

Let $\mathfrak{A}$ be a von Neumann algebra as before and $v \in F(\mathfrak{A})$ an arbitrary a priori state. Let

$$J_v(\mathfrak{A}) = \{T \in J(\mathfrak{A}); v(T^2) = 1\}$$

and

$$F_v(\mathfrak{A}) = \{w \in F(\mathfrak{A}); w(A) = v(TAT), A \in \mathfrak{A}, T \in J_v(\mathfrak{A})\}$$

We would like to construct a relative entropy function H_v on $J_v(\mathfrak{A})$ with the following property: Let $(\mathfrak{B}, w_{\mathfrak{B}})$ be an arbitrary partial measurement and $T_0 \in J(\mathfrak{B})$ correspond to the $(\mathfrak{B}, v)$ inference of $w_{\mathfrak{B}}$. Under variations compatible with the partial measurement, H_v should be stationary at T_0.

Guided by existing information-theoretic entropy functions, we assume for H_v the general form

$$H_v(T) = v(F(T))$$

with $F(T)$ norm-analytic around any $\lambda \cdot 1$, $\lambda \neq 0$.

Theorem 3.1. Let $v \in F(\mathfrak{A})$ be a nontracial a priori state on $\mathfrak{A}$ (non-Abelian) and $(\mathfrak{B}, w_{\mathfrak{B}})$ an arbitrary partial measurement with $T_0 \in J(\mathfrak{B})$ corresponding to the $(\mathfrak{B}, v)$ inference.

Let $T(\epsilon)$ be a curve in $J_v(\mathfrak{A})$ analytic at $T(0) = T_0$ and $F(T) = \sum_{n=0}^{\infty} C_n(T - \lambda)^n$ norm-analytic at $\lambda \cdot 1$, $\lambda \neq 0$. For variations compatible with the partial measurement, i.e.,

$$(d/d\epsilon)v(T(\epsilon)BT(\epsilon))|_{\epsilon=0} = 0, \qquad \forall B \in \mathfrak{B}$$

the condition

$$(d/d\epsilon)H_v(T(\epsilon)) = 0$$

implies

$$C_n = 0, \qquad n \geqslant 3$$

Proof. Let $T(\epsilon) = T_0 + \epsilon K + O(\epsilon^2)$. Then the compatibility condition reads

$$(d/d\epsilon)v(T(\epsilon)BT(\epsilon))|_{\epsilon=0} = v(T_0BK + KBT_0) = 0, \qquad \forall B \in \mathfrak{B}$$

In particular, for $B = T_0^{-1}(T_0 - \lambda)^m$, $m \geqslant 0$, we get

$$v((T_0 - \lambda)^m K + K(T_0 - \lambda)^m) = 0, \qquad m \geqslant 0$$

The extremal property of H_v at T_0 implies

$$\frac{d}{d\epsilon} H_v(T(\epsilon))|_{\epsilon=0} = \sum_{n=0}^{\infty} C_n \frac{d}{d\epsilon} v((T(\epsilon) - \lambda)^n)|_{\epsilon=0}$$
$$= \sum_{n=1}^{\infty} C_n \sum_{l=0}^{n-1} v((T_0 - \lambda)^{n-l-1} K(T_0 - \lambda)^l) = 0$$

Using the compatibility condition, we get

$$\sum_{n=3}^{\infty} C_n v_n = 0 \qquad \text{where} \quad v_n = \sum_{l=1}^{n-2} v((T_0 - \lambda)^{n-l-1} K(T_0 - \lambda)^l)$$

Since these relations have to hold for any partial measurement, we choose $\mathfrak{B} = \{P\}''$, where P is an arbitrary nontrivial projector. Then, for suitable real λ,

$$T_0 - \lambda \cdot 1 = aP + b(1 - P)$$

for some real $a, b \neq 0$.

The compatibility condition implies

$$v(KP + PK) = 0, \qquad v(K) = 0$$

and the extremal property then reads

$$v_n = \sum_{l=1}^{n-2} v([a^{n-l-1}P + b^{n-l-1}(1 - P)]K[a^l P + b^l(1 - P)])$$
$$= v(PKP) \sum_{l=1}^{n-2} (a^{n-1} - 2a^{n-l-1}b^l + b^{n-1})$$
$$= v(PKP)[(n - 2)(a^{n-1} + b^{n-1}) - 2 \sum_{l=1}^{n-2} a^{n-l-1}b^l]$$

or, with $x = a/b$,

$$v_n = v(PKP)b^{n-1}\left[(n-2)(x^{n-1}+1) - 2\sum_{l=1}^{n-2} x^l\right]$$

Then

$$\sum_{n=3}^{\infty} C_n v_n = v(PKP)\sum_{k=0}^{\infty} a_k x^k = 0$$

with

$$a_0 = \sum_{n=3}^{\infty} (n-2)C_n b^{n-1}$$

$$a_k = (k-1)C_{k+1}b^k - 2\sum_{n=k+2}^{\infty} C_n b^{n-1}, \qquad k \geqslant 1$$

Varying $w_{\mathfrak{B}}$ implies that we can assume x to range over a compact set in $\mathbb{R}$. Therefore $a_k = 0$, $k \geqslant 0$, or equivalently $C_n = 0$, $n \geqslant 3$.

Remark 3.1. Since C_0, C_1, C_2 are arbitrary, it follows for v nontracial that

$$H_v(T) = C_0 + C_1 v(T - \lambda) + C_2 v([T - \lambda]^2) = a_0 + a_1 v(T)$$

is the most general entropy function. On the other hand, if v is tracial, all C_n are arbitrary. The class of admissible entropy functions is then much larger and contains, in particular, the standard form of the entropy.

Definition 3.1. The normalized relative entropy function defined on $J_v(\mathfrak{A})$ by

$$H_v(T) = v(T)$$

is called the *v-entropy*.

Theorem 3.2. Let $(\mathfrak{B}, w_{\mathfrak{B}})$ be any partial measurement and $J_v(\mathfrak{A}, \mathfrak{B}, w_{\mathfrak{B}}) = \{T \in J_v(\mathfrak{A}); v(TBT) = w_{\mathfrak{B}}(B), \forall B \in \mathfrak{B}\}$ the T's compatible with the partial measurement. Then the v-entropy $H_v(T) = v(T)$ restricted to $J_v(\mathfrak{A}, \mathfrak{B}, w_{\mathfrak{B}})$ has a unique absolute maximum at $T_0 \in J(\mathfrak{B})$ corresponding to the $(\mathfrak{B}, v)$ inference of $w_{\mathfrak{B}}$.

Proof. Let $T \in J_v(\mathfrak{A}, \mathfrak{B}, w_{\mathfrak{B}})$. Then

$$v(TBT) = v(T_0 B T_0) = w_{\mathfrak{B}}(B)$$

implies, for $B = T_0^{-1} \in J(\mathfrak{B})$,

$$v(TT_0^{-1}T) = v(T_0)$$

With

$$A = T_0^{\frac{1}{2}}, \qquad B = T_0^{-\frac{1}{2}}T, \qquad AB = T$$

the Schwarz inequality reads

$$\begin{aligned} |v(T)|^2 &= |v(AB)|^2 \leqslant v(AA^*)v(B^*B) \\ &= v(T_0)v(TT_0^{-1}T) \\ &= [v(T_0)]^2 \end{aligned}$$

or

$$v(T) \leqslant v(T_0)$$

Since v is faithful, the equality holds if $B = \lambda A$ or $T = \lambda T_0$, and from $v(T^2) = v(T_0{}^2) = 1$ and T, T_0 positive it follows that $T = T_0$.

4. EXAMPLE

The simplest nontrivial illustration of a $(\mathfrak{B}, v)$ inference is the quantum mechanical single-spin system. In this case $\mathfrak{A} = \mathfrak{B}(\mathfrak{H})$ and $\mathfrak{H}$ is two-dimensional. If the system is placed into a magnetic field in the 3-direction, the a priori state v on $\mathfrak{A}$ is the Gibbs state, which has the general form

$$v(A) = \tau(VA), \qquad v = \tfrac{1}{2}\sigma_0 + \tfrac{1}{2}a\sigma_3$$

where τ is the trace, σ_0 and σ_k are the Pauli matrices, and V is the density matrix corresponding to v. Suppose that the partial measurement consists in measuring the 1-component of the spin:

$$w_{\mathfrak{B}}(\sigma_1) = b$$

The measured subalgebra $\mathfrak{B} = \{\sigma_1\}''$ consists of the projectors

$$\mathfrak{B} = \{0, P, 1 - P, 1\}, \qquad P = \tfrac{1}{2}\sigma_0 + \tfrac{1}{2}\sigma_1$$

and $w_{\mathfrak{B}}$ assumes on $\mathfrak{B}$ the values $\{0, \frac{1}{2}(1 + b), \frac{1}{2}(1 - b), 1\}$. For $T_0 \in J(\mathfrak{B})$ one obtains

$$\begin{aligned} T_0 &= \left[\frac{w_{\mathfrak{B}}(P)}{v(P)}\right]^{1/2} P + \left[\frac{w_{\mathfrak{B}}(1 - P)}{v(1 - P)}\right]^{1/2} (1 - P) \\ &= \tfrac{1}{2}[(1 + b)^{1/2} + (1 - b)^{1/2}]\sigma_0 + \tfrac{1}{2}[(1 + b)^{1/2} - (1 - b)^{1/2}]\sigma_1 \end{aligned}$$

and the $(\mathfrak{B}, v)$ inference w_0 of $w_{\mathfrak{B}}$ is

$$w_0(A) = \tau(W_0 A), \qquad W_0 = T_0 V T_0 = \tfrac{1}{2}\sigma_0 + \tfrac{1}{2}b\sigma_1 + \tfrac{1}{2}a(1 - b^2)^{1/2}\sigma_3$$

For a general $T \in J_v(\mathfrak{A}, \mathfrak{B}, w_{\mathfrak{B}})$ we have

$$TVT = W = \tfrac{1}{2}\sigma_0 + \tfrac{1}{2}b\sigma_1 + \tfrac{1}{2}W_2\sigma_2 + \tfrac{1}{2}W_3\sigma_3$$

and it can be verified that within this class of T's the entropy $H_v(T) = v(T)$ has a unique maximum at T_0.

5. DISCUSSION

Within the framework of inference $\{\mathfrak{A}, v, \mathfrak{B}, w_{\mathfrak{B}};\ v \text{ nontracial}\}$ we have constructed a unique entropy function characterizing the coarse-grained state as the maximum entropy state.

If $\mathfrak{A} = \mathfrak{B}(\mathfrak{H})$ and v the trace τ, then every $w \in F(\mathfrak{A})$ can be written as $w(A) = \tau(WA)$. In this case our entropy is

$$H_\tau(T) = \tau(T) = \tau(\sqrt{W})$$

and differs from the standard entropy

$$\hat{H}_\tau(W) = -\tau(W \log W)$$

For our inference they are, however, equivalent since they lead to the same maximum entropy states. On the other hand, the straightforward extension[3]

$$\hat{H}_v(T) = -v(T^2 \log T^2)$$

of the standard entropy to the case where v is nontracial does not characterize $(\mathfrak{B}, v)$ inference.[4]

ACKNOWLEDGMENT

One of the authors (W.W.) would like to thank Dr. F. Straumann, Waldenburg, for helpful remarks.

REFERENCES

1. J. P. Marchand, *Found. Phys.* **7**:35 (1977).
2. M. Takesaki, *Tomita's Theory of Modular Hilbert Algebras*, Springer Lecture Notes in Mathematics, No. 128 (1970), p. 104.
3. J. Naudts, *Comm. Math. Phys.* **37**:175 (1974).
4. R. Benoist, Thesis, University of Denver (1977).