

AD-A141 401

A NOTE ON AVERAGE MUTUAL INFORMATION FOR SPHERICALLY  
INVARIANT PROCESSES(U) NORTH CAROLINA UNIV AT CHAPEL  
HILL DEPT OF STATISTICS A F GUALTIEROTTI MAR 84 LAP-5  
NO0014-81-K-0373

1/1

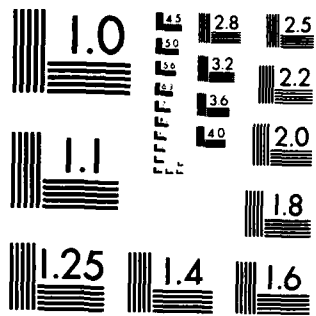
UNCLASSIFIED

F/G 12/1

NL



END  
DATE  
FILMED  
7 84  
DTIC



MICROCOPY RESOLUTION TEST CHART  
NATIONAL BUREAU OF STANDARDS-1963-A

13

AD-A141 401

A NOTE ON AVERAGE MUTUAL INFORMATION  
FOR SPHERICALLY INVARIANT PROCESSES\*

A.F. Gualtierotti

IDHEAP  
University of Lausanne  
CH 1015 Lausanne  
Switzerland

and

Department of Statistics  
University of North Carolina  
Chapel Hill, NC 27514

DTIC FILE COPY

LAP-5  
March, 1984

DTIC  
ELECTE  
MAY 23 1984  
S D  
E

\*Research supported under ONR Contract N00014-81-K-0373

This document has been approved  
for public release and sale; its  
distribution is unlimited.

84 05 21 247

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM																				
1. REPORT NUMBER	2. GOVT ACCESSION NO. <b>A141401</b>	3. RECIPIENT'S CATALOG NUMBER																				
4. TITLE (and Subtitle)  A Note on Average Mutual Information for Spherically Invariant Processes	5. TYPE OF REPORT & PERIOD COVERED  TECHNICAL																					
	6. PERFORMING ORG. REPORT NUMBER LAP-5																					
7. AUTHOR(s)  A.F. Gualtierotti	8. CONTRACT OR GRANT NUMBER(s) N00014-81-K-0373																					
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics University of North Carolina Chapel Hill, North Carolina 27514	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS  SRO 105																					
11. CONTROLLING OFFICE NAME AND ADDRESS Statistics & Probability Program Office of Naval Research Arlington, VA 22217	12. REPORT DATE March, 1984																					
	13. NUMBER OF PAGES 13																					
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report) UNCLASSIFIED																					
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE																					
16. DISTRIBUTION STATEMENT (of this Report)  Approved for Public Release: Distribution Unlimited																						
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)																						
18. SUPPLEMENTARY NOTES	<table border="1"> <tr> <td colspan="2">Accession For</td> </tr> <tr> <td>NTIS GRA&amp;I</td> <td><input checked="" type="checkbox"/></td> </tr> <tr> <td>DTIC TAB</td> <td><input type="checkbox"/></td> </tr> <tr> <td>Unannounced</td> <td><input type="checkbox"/></td> </tr> <tr> <td>Justification</td> <td></td> </tr> <tr> <td colspan="2">By _____</td> </tr> <tr> <td colspan="2">Distribution/</td> </tr> <tr> <td colspan="2">Availability Codes</td> </tr> <tr> <td>Dist</td> <td>Avail and/or Special</td> </tr> <tr> <td>A-1</td> <td></td> </tr> </table>		Accession For		NTIS GRA&I	<input checked="" type="checkbox"/>	DTIC TAB	<input type="checkbox"/>	Unannounced	<input type="checkbox"/>	Justification		By _____		Distribution/		Availability Codes		Dist	Avail and/or Special	A-1	
Accession For																						
NTIS GRA&I	<input checked="" type="checkbox"/>																					
DTIC TAB	<input type="checkbox"/>																					
Unannounced	<input type="checkbox"/>																					
Justification																						
By _____																						
Distribution/																						
Availability Codes																						
Dist	Avail and/or Special																					
A-1																						
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)  Mutual information; Spherically invariant processes																						
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)  Mutual information is considered for communication channels where the noise process is spherically invariant. An expression is obtained for noise processes having a mixing measure with finite support.																						



A Note on Average Mutual Information  
for Spherically Invariant Processes

1. Introduction

The actual transmission capacity of a given channel is a parameter of basic importance in any communication system since it fixes limits to the rate at which information can be transmitted reliably. There has thus been an effort which started with Shannon (1948) to compute the capacity of transmission for different channel and transmission models. In the case of a continuous channel, most results have been obtained for a Gaussian noise (Baker[1978]; Hitsuda and Ihara[1975]; Kadota, Zakai and Kiv[1971]). Some attempts to steer away from the Gaussian case have also been made (Gualtierotti [1980]) and these indicate that new methods may be required. Indeed, in the Gaussian case, most quantities of interest can be explicitly obtained whereas these computations are almost always impossible in other instances. Furthermore the computation of mutual information requires that the joint law of the message and the received signal be absolutely continuous with respect to the marginals and that the Radon-Nikodým derivative be computed: though the Gaussian case is well known (Baker[1973]), this knowledge is again unavailable for most other models.

Spherically invariant distributions are mixtures of Gaussian ones and through mixing a number of well known distributions can be obtained, such as the double exponential and the student distributions (Keilson and Steutel, [1974]). There is also evidence that some real life noises can be described

through spherically invariant probabilities, particularly in underwater acoustics. It is thus natural to investigate the problems of absolute continuity, of calculation of mutual information and channel capacity for spherically invariant noises. This is the subject of the present paper.

We obtain a formula for average mutual information when the mixing measure is discrete with finite support.

## 2. Preliminaries

We give here the basic definitions and a number of useful lemmas which can be easily checked from first principles.

$H_1$  and  $H_2$  are real and separable Hilbert spaces with respective inner products  $\langle u^1, v^1 \rangle_1$ ,  $u^1, v^1 \in H_1$ , and  $\langle u^2, v^2 \rangle_2$ ,  $u^2, v^2 \in H_2$ .  $H$  is the set  $H_1 \times H_2$  and its elements are denoted  $\vec{h} = (u^1, u^2)$ . If  $\vec{h}_1 = (u^1, u^2)$  and  $\vec{h}_2 = (v^1, v^2)$ , let  $\langle \cdot, \cdot \rangle : H \times H \rightarrow \mathbb{R}$  be the map defined by the relation

$$\langle \vec{h}_1, \vec{h}_2 \rangle = \langle u^1, v^1 \rangle_1 + \langle u^2, v^2 \rangle_2.$$

$\langle \cdot, \cdot \rangle$  is an inner product on  $H$  and with this inner product it is a real and separable Hilbert space.  $p_{H_1}$  is the projection with range  $H_1 \times \{0\}$ ,  $p_{H_2}$  that with range  $\{0\} \times H_2$ .

Lemma 1: Let  $J_i : H \rightarrow H_i$  be defined by  $J_i (u^1, u^2) = u^i$ ,  $i = 1, 2$ . Then

$$1) J_1^* J_1 = p_{H_1}; J_2^* J_2 = p_{H_2}.$$

$$2) J_1^* J_1 = p_{H_1}; J_2^* J_2 = p_{H_2}.$$

$$3) J_1 J_1^* = id_{H_1}; J_2 J_2^* = id_{H_2}.$$

$$4) p_{H_1} + p_{H_2} = id_H.$$

$\mathcal{B}[H]$ ,  $\mathcal{B}[H_1]$  and  $\mathcal{B}[H_2]$  are the Borel sets of respectively  $H$ ,  $H_1$  and  $H_2$ . Then

Lemma 2:  $\mathcal{B}[H] = \mathcal{B}[H_1] \otimes \mathcal{B}[H_2]$ .

If  $P$  is a probability measure on  $H$ ,  $P^i = P \circ J_i^{-1}$ ,  $i=1,2$ , are probabilities on  $\mathcal{B}[H_i]$ ,  $i=1,2$ . They are called the marginals of  $P$ . The product of these marginals is denoted either  $P^{\otimes}$  or  $P^1 \otimes P^2$  according to convenience. It is defined on  $H$ .

A Gaussian probability  $P$  on any real and separable Hilbert space  $H$  is determined by its characteristic function  $\phi_P$  or its mean  $m$  and covariance  $R$ .  $m$  belongs to  $H$  and is identified by the relation

$$\langle h, m \rangle = \int_H \langle h, x \rangle P(dx).$$

$R$  belongs to the nonnegative and self-adjoint operators on  $H$  which have finite trace and is identified by the relation

$$\langle h, Rk \rangle = \int_H \langle h, x-m \rangle \langle k, x-m \rangle P(dx).$$

$m$ ,  $R$  and  $\phi_P(h) = \exp\{i\langle h, m \rangle - \frac{1}{2} \langle h, Rh \rangle\}$ .

Lemma 3: Let again  $H=H_1 \times H_2$  and  $P$  be a Gaussian probability on  $\mathcal{B}[H]$  with mean  $\vec{m} = (m^1, m^2)$  and covariance  $R$ . Then:

- 1)  $P^i$  is Gaussian with mean  $m^i$  and covariance  $R_i = J_i R J_i^*$ ,  $i=1,2$ .
- 2)  $P^{\otimes}$  is Gaussian with mean  $\vec{m}^{\otimes}$  and covariance  $R^{\otimes} = P_{H_1} R P_{H_1} + P_{H_2} R P_{H_2}$ .

If  $H$  is any real and separable Hilbert space and  $a > 0$ ,  $T_a: H \rightarrow H$  defined by  $T_a h = ah$  is a homeomorphism of  $H$ , so that  $P \circ T_a^{-1}$  is well defined. This measure is written  $P_a$  or  $P(a, \cdot)$  according to convenience.

Lemma 4:  $P_a$  is Gaussian with mean  $am$  and covariance  $a^2 R$ .  $P(a, B) = P_a(B)$ ,  $B \in \mathcal{B}[H]$ , is a transition function defined on  $[0, \infty[ \times \mathcal{B}[H]$ .

Let  $F$  be a probability measure on  $\mathcal{B}[\mathbb{R}_+]$ . A spherically invariant measure on  $H$  is a probability  $Q$  of the form

$$Q(B) = \int_0^\infty P(a,B) F(da) , B \in \mathcal{B}[H].$$

Lemma 5: Let again  $H=H_1 \times H_2$ . Then:

1)  $P_a^i = P_a^i \circ T_a^{-1}$ .

2)  $P_a^1 \otimes P_a^2 = P_a^\otimes \circ T_a^{-1}$ , written  $P_a^\otimes$ .

3) The covariance  $R_{a,b}^\otimes$  of the measure  $P_a^1 \otimes P_b^2$  written  $P_{a,b}^\otimes$  is given by the formula

$$R_{a,b}^\otimes = a^2 P_{H_1} R_{P_{H_1}} + b^2 P_{H_2} R_{P_{H_2}} .$$

4)  $Q^i(B) = \int_0^\infty P^i(a,B) F(da) , B \in \mathcal{B}[H_i]$ .

5)  $Q^\otimes(B) = \int_0^\infty \int_0^\infty P_{a,b}^\otimes(B) F \otimes F(da,db) , B \in \mathcal{B}[H]$ .

Lemma 6: Let  $P_1, \dots, P_m$  and  $Q_1, \dots, Q_n$  be probability measures on  $(\Omega, \mathcal{A})$  such that  $P_i \perp Q_j$ ,  $1 \leq i \leq m$ ,  $1 \leq j \leq n$ . Then there exists  $A \in \mathcal{A}$  such that

$$P_i(A) = Q_j(A^c) = 0, \quad 1 \leq i \leq m, \quad 1 \leq j \leq n.$$

proof: For measures  $\lambda, \mu, \nu$  such that  $\lambda \perp \nu$  and  $\mu \perp \nu$ ,  $\lambda + \mu \perp \nu$  (Ash, p.67). By recursion, one obtains that

$$\sum_{i=1}^m P_i \perp \sum_{j=1}^n Q_j.$$

$A$  is a set such that  $(\sum_{i=1}^m P_i)(A) = (\sum_{j=1}^n Q_j)(A^c) = 0$ .



3. The case of a mixing measure F with finite support

A. Absolute continuity

Lemma 7: Suppose that neither  $R_1$  nor  $R_2$  are zero and that b and c are not both equal to a. Then

$$P_a^\otimes \perp P_{b,c}^\otimes.$$

proof: If  $a=0$ , or if  $a>0$  and either b or c is zero, one has in the first case  $R_a^\otimes = 0$ ,  $R_{b,c}^\otimes \neq 0$  and, in the second,  $R_a^\otimes = a^2 \{p_{H_1} R_{p_{H_1}} + p_{H_2} R_{p_{H_2}}\}$  and  $R_{b,c}^\otimes = b^2 p_{H_1} R_{p_{H_1}}$ , or  $R_{b,c}^\otimes = c^2 p_{H_2} R_{p_{H_2}}$ , or  $R_{b,c}^\otimes = 0$ .  $\sqrt{R_a^\otimes}$  and  $\sqrt{R_{b,c}^\otimes}$  do not then have the same range and, since  $P_a^\otimes$  and  $P_{b,c}^\otimes$  are Gaussian, they must be orthogonal (Rao-Varadarajan, 1963). If a, b and c are all positive and if  $P_a^\otimes$  and  $P_{b,c}^\otimes$  are not orthogonal, they must be equivalent since they are Gaussian (Rao-Varadarajan, 1963). But then, from the same reference, we have that

$$\sqrt{R_{b,c}^\otimes} = \sqrt{R_a^\otimes} (I+T) \sqrt{R_a^\otimes},$$

where T is Hilbert-Schmidt with spectrum  $\sigma(T) > -1$ . However T can be identified as the "diagonal" operator with diagonal elements  $([b^2/a^2]-1)id_{H_1}$  and  $([c^2/a^2]-1)id_{H_2}$  leading to a contradiction.

Lemma 8: Let a, b and c be positive. Then, if P and  $P^\otimes$  are orthogonal, so are  $P_a$  and  $P_{b,c}^\otimes$ .

proof: Since  $P_a$  and  $P_{b,c}^\otimes$  are Gaussian, if they are not orthogonal, they must be equivalent (Rao-Varadarajan, 1963). But

$$P_a \sim N(0, a^2 R)$$

and  $P_{b,c}^\otimes \sim N(0, R_{b,c}^\otimes)$ ,

so that (Rao-Varadarajan, 1963),

$$(1) \quad R_{b,c}^{\otimes} = a^2 R^{1/2} (I+T) R^{1/2},$$

T Hilbert-Schmidt,  $\sigma(T) > -1$ .

By Lemma 5, 3)

$$P_{H_1} R_{b,c}^{\otimes} P_{H_2} = 0,$$

and, equivalently, using (1),

$$(2) \quad P_{H_1} R P_{H_2} + P_{H_1} R^{1/2} T R^{1/2} P_{H_2} = 0.$$

Furthermore, by Lemma 3, 2),

$$R^{\otimes} - R = -\{P_{H_1} R P_{H_2} + P_{H_2} R P_{H_1}\},$$

so that, using (2), one gets

$$(3) \quad R^{\otimes} - R = P_{H_1} R^{1/2} T R^{1/2} P_{H_2} + P_{H_2} R^{1/2} T R^{1/2} P_{H_1}.$$

By Lemma 5, 3), and the assumption  $b > 0$ ,

$$\langle P_{H_1} R P_{H_1} \vec{h}, \vec{h} \rangle \leq \frac{1}{b^2} \langle R_{b,c}^{\otimes} \vec{h}, \vec{h} \rangle,$$

and, by (1),

$$\langle R_{b,c}^{\otimes} \vec{h}, \vec{h} \rangle \leq a^2 \|I+T\| \langle R \vec{h}, \vec{h} \rangle.$$

Consequently, if K is an appropriate constant,

$$\langle P_{H_1} R P_{H_1} \vec{h}, \vec{h} \rangle \leq K \langle R \vec{h}, \vec{h} \rangle.$$

Thus (Douglas, 1966)

$$(4) \quad P_{H_1} R P_{H_1} = R^{1/2} U R^{1/2},$$

where U is bounded, nonnegative and self-adjoint.

Similarly, one has

$$(5) \quad P_{H_2} R P_{H_2} = R^{1/2} V R^{1/2},$$

where V is bounded, nonnegative and self-adjoint.

The polar decomposition (Weidmann, p.197) yields

$$(6) \quad R^{1/2} P_{H_1} = A (P_{H_1} R P_{H_1})^{1/2}$$

where A is a partial isometry such that:

$$AA^* = P_{L_1}, \quad L_1 = \text{closure of range of } R^{1/2} P_{H_1},$$

$$A^*A = P_{L_2}, \quad L_2 = \text{closure of range of } (P_{H_1} R P_{H_1})^{1/2}.$$

Similarly, one has

$$(7) \quad U^{1/2} R^{1/2} = B (R^{1/2} U R^{1/2})^{1/2}.$$

$$(6) \text{ yields } A^* R^{1/2} P_{H_1} = (P_{H_1} R P_{H_1})^{1/2} \text{ and } (7) \text{ yields } B^* U^{1/2} R^{1/2} = (R^{1/2} U R^{1/2})^{1/2}.$$

Consequently, one has from (4) that

$$A^* R^{1/2} P_{H_1} = B^* U^{1/2} R^{1/2}.$$

This can be rewritten as

$$(8) \quad P_{H_1} R^{1/2} = R^{1/2} U^{1/2} B A^*.$$

Similarly, one has

$$(9) \quad P_{H_2} R^{1/2} = R^{1/2} V^{1/2} D C^*.$$

Using (8) and (9) in (3), one gets, if  $\tilde{T} = U^{1/2} B A^* T C D^* V^{1/2}$ ,

$$(10) \quad R^\circ = R^{1/2} (I + \tilde{T} + \tilde{T}^*) R^{1/2}.$$

Let us show that

$$(11) \quad \sigma(\tilde{T} + \tilde{T}^*) > -1$$

or equivalently, that

$$(12) \quad \sqrt{R^\circ} \text{ and } \sqrt{R} \text{ have the same range.}$$

One has

$$0 \leq \min(b^2, c^2) R^\circ \leq R_{b,c}^\circ \leq \max(b^2, c^2) R^\circ,$$

so that  $\sqrt{R^\circ}$  and  $\sqrt{R_{b,c}^\circ}$  have the same range (Douglas, 1966). By assumption,

$P_a$  and  $P_{b,c}^\circ$  are equivalent and Gaussian so, (Rao-Varadarajan, 1963),

$\sqrt{R_{b,c}^{\otimes}}$  and  $\sqrt{a^2 R}$  have the same range. Then (12) is established.

From (10), (11) and (Rao-Varadarajan, 1963), one has that  $P$  and  $P^{\otimes}$  are equivalent, contradicting the assumptions.

Let now  $0 \leq a_1 < a_2 < \dots < a_n$  be the support of  $F$  and suppose  $F$  has mass  $p_i \in ]0, 1[$  at  $a_i$ ,  $1 \leq i \leq n$ . We write  $P_i, P_i^{\otimes}, P_{i,j}^{\otimes}$  for, respectively,  $P_{a_i}, P_{a_i}^{\otimes}, P_{a_i, a_j}^{\otimes}$ .

Proposition 1: If  $P \ll P^{\otimes}$ ,  $Q \ll Q^{\otimes}$

proof: Since  $P \ll P^{\otimes}$ ,  $P_i \ll P_i^{\otimes}$ ,  $1 \leq i \leq n$ . So, if  $Q^{\otimes}(B) = 0$ ,  $P_i^{\otimes}(B) = 0$ ,  $1 \leq i \leq n$ , using Lemma 5, 5). But then  $P_i(B) = 0$ ,  $1 \leq i \leq n$ , that is  $Q(B) = 0$ .

Proposition 2: If  $P \perp P^{\otimes}$  and  $a_1 > 0$ ,  $Q \perp Q^{\otimes}$ . If  $a_1 = 0$ ,  $Q$  and  $Q^{\otimes}$  cannot be orthogonal.

proof: If  $a_1 = 0$ ,  $Q(\{0\}) > 0$  and  $Q^{\otimes}(\{0\}) > 0$ . So, if  $Q(B) = 0$ ,  $B^c \ni \{0\}$  and  $Q^{\otimes}(B^c) > 0$ . If  $a_1 > 0$ , then, by Lemma 8,  $P_i \perp P_{j,k}^{\otimes}$ ,  $1 \leq i, j, k \leq n$ . The result then follows by Lemma 6.

Proposition 3: If  $Q \perp Q^{\otimes}$ ,  $P \perp P^{\otimes}$ .

proof: Let  $Q(B) = Q^{\otimes}(B^c) = 0$ . Then  $P_i(B) = P_i^{\otimes}(B^c) = 0$ ,  $1 \leq i \leq n$ . But, by Lemma 5, 1) and 2,  $P_i(B) = P\{T_{a_i}^{-1}(B)\}$  and  $P_i^{\otimes}(B^c) = P^{\otimes}\{[T_{a_i}^{-1}(B)]^c\}$ .

Proposition 4: If  $Q \ll Q^{\otimes}$  and  $a_1 > 0$ ,  $P \ll P^{\otimes}$ .

proof: Suppose  $P$  is not absolutely continuous with respect to  $P^{\otimes}$ . Since  $P$  and  $P^{\otimes}$  are Gaussian,  $P$  and  $P^{\otimes}$  are then orthogonal (Rao-Varadarajan, 1973). But, by Proposition 2, one has that  $Q \perp Q^{\otimes}$ . So, if  $Q(B) = Q^{\otimes}(B^c) = 0$ ,  $Q(B^c) = 0$  by assumption and  $Q$  is identically zero and cannot be a probability.

Remark:  $Q \ll Q^{\otimes}$  does not imply  $Q \equiv Q^{\otimes}$ .

proof: Let  $Q = \sum_{i=1}^n p_i P_i^{\otimes}$ . Then  $Q^{\otimes} = \sum_{i=1}^n \sum_{j=1}^n p_i p_j P_{i,j}^{\otimes}$ .

Use Lemma 7 and 6 to find a Borel set  $B$  such that

$$P_i^{\otimes}(B) = P_{j,k}^{\otimes}(B^c) = 0, \quad 1 \leq i, j, k \leq n, \quad j \neq k.$$

Then  $Q(B)=0$ , but  $Q^{\otimes}(B) = \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n p_i p_j > 0$ .

B. The Radon-Nikodym derivative of  $Q$  with respect to  $Q^{\otimes}$

Proposition 5: If  $a_1 > 0$  and  $Q \ll Q^{\otimes}$ , there exist Borel sets  $B_i$ ,  $1 \leq i \leq n$ , such that

$$\frac{dQ}{dQ^{\otimes}} = \sum_{i=1}^n \frac{1}{p_i} \{1 - I_{B_i}\} \frac{dP_i}{dP_i^{\otimes}}$$

proof: For each fixed  $i \in \{1, \dots, n\}$  choose, with the help of Lemma 7 and Lemma 6, a Borel set  $B_i$  such that

$$P_i(B_i) = P_{j,k}^{\otimes}(B_i^c) = 0, \quad 1 \leq j, k \leq n, \quad (j, k) \neq (i, i).$$

From Proposition 4 and Lemma 5, 1) and 2), we also have that  $P_i \ll P_i^{\otimes}$ ,  $1 \leq i \leq n$ .

Set 
$$\Delta = \sum_{i=1}^n \frac{1}{p_i} \{1 - I_{B_i}\} \frac{dP_i}{dP_i^{\otimes}}.$$

Then 
$$\int_B \Delta dQ^{\otimes} = \sum_{i=1}^n \sum_{j=1}^n p_i p_j \int_B \Delta dP_{i,j}^{\otimes}.$$

But 
$$\int_B \Delta dP_{i,j}^{\otimes} = \sum_{k=1}^n \frac{1}{p_k} \int_{B \cap B_k^c} dP_{i,j}^{\otimes} \frac{dP_k}{dP_k^{\otimes}}$$

and  $P_{i,j}^{\otimes}(B_k^c) = 0$  for  $(i, j) \neq (k, k)$ .

Thus, if  $i \neq j$ ,  $\int_B \Delta dP_{i,j}^{\otimes} = 0$ , and, if  $i=j$ ,

$$\begin{aligned} \int_B \Delta dP_{i,i}^{\otimes} &= \frac{1}{p_i} \int_{B \cap B_i^c} dP_i^{\otimes} \frac{dP_i}{dP_i^{\otimes}} \\ &= \frac{1}{p_i} P_i(B \cap B_i^c) \\ &= \frac{1}{p_i} P_i(B). \end{aligned}$$

Consequently,  $\int_B \Delta dQ^{\otimes} = \sum_{i=1}^n p_i P_i(B) = Q(B)$ , or  $\Delta = \frac{dQ}{dQ^{\otimes}}$ .

C. Calculation of mutual information

If  $P$  is any probability measure on  $H$ ,  $I(P)$ , the average mutual information of  $P$ , is given by the formula

$$I(P) = \int_H \log \frac{dP}{dP^{\otimes}} dP,$$

provided  $P \ll P^{\otimes}$ . The entropy of  $F$ ,  $H(F)$ , is the quantity  $H(F) = - \sum_{i=1}^n p_i \log p_i$ .

One has:

Proposition 6:  $I(Q) = H(F) + I(P)$

proof: Let  $\vec{e}_n$  be the eigenvector of  $R$  with associated eigenvalue  $\lambda_n$ . Let:

$$\Psi_n(\vec{h}) = \frac{1}{n} \sum_{i=1}^n \frac{\langle \vec{h}, \vec{e}_n \rangle^2}{\lambda_n},$$

$$\Psi(\vec{h}) = \overline{\lim}_n \Psi_n(\vec{h}),$$

$$A_i = \Psi^{-1}(a_i^2), \quad 1 \leq i \leq n.$$

Since  $\Psi_n$  is measurable,  $\Psi$  is, so that the  $A_i$ 's are Borel sets which are disjoint.

Let  $X_n = \langle \cdot, \vec{e}_n \rangle / \sqrt{\lambda_n}$ . With respect to  $P_i$ ,  $\{X_n, n \in \mathbb{N}\}$  is a family of independent  $N(0, a_i^2)$ -random variables.  $\{X_n^2, n \in \mathbb{N}\}$  is thus a family of independent random variables with mean  $a_i^2$  and variance  $2a_i^4$ , still with respect to  $P_i$ . By the law of large numbers:

$$P_i(A_i) = 1, \quad 1 \leq i \leq n.$$

We can thus assume that  $\frac{dP_i}{dP_i^{\otimes}}$  is zero outside of  $A_i$

Consequently,  $\int_H \log \frac{dQ}{dQ^{\otimes}} dP_i = \int_H \log \frac{1}{p_i} \frac{dP_i}{dP_i^{\otimes}} dP_i$ .

But  $\int_H \log \frac{dP_i}{dP_i^\otimes} dP_i = \int_H \log \frac{dP}{dP^\otimes} dP = I(P).$

Finally,  $\int_H \log \frac{dQ}{dQ^\otimes} dQ = \sum_{i=1}^n p_i \int_H \log \frac{1}{p_i} \frac{dP_i}{dP_i^\otimes} dP_i$   
 $= \sum_{i=1}^n p_i \{-\log p_i + I(P)\} = H(F) + I(P).$

4. The case of a general F

Lemma 9: Let  $B = \{\vec{x} \in H: \|\vec{x} - \vec{h}\| < \alpha\}$  and  $I = \{a \geq 0: a\vec{x} \in B\}$ .  $I$  is an open interval.

proof: Suppose  $a_1 < a < a_2$  and  $a_1, a_2 \in I$ . Let

$$\lambda = (a_2 - a) / (a_2 - a_1) \text{ and } 1 - \lambda = (a - a_1) / (a_2 - a_1).$$

Then  $a = \lambda a_1 + (1 - \lambda) a_2.$

So  $\|a\vec{x} - \vec{h}\| = \|\lambda a_1 + (1 - \lambda) a_2\vec{x} - [\lambda + (1 - \lambda)]\vec{h}\| \leq \lambda \|a_1\vec{x} - \vec{h}\| + (1 - \lambda) \|a_2\vec{x} - \vec{h}\| < \alpha.$

Thus  $a \in I$  and  $I$  is an interval. Furthermore if  $a \in I$  and  $\beta = \alpha - \|a\vec{x} - \vec{h}\|$ , then, for  $b < \beta / \|\vec{x}\|$ ,

$$\|(a+b)\vec{x} - \vec{h}\| \leq \|a\vec{x} - \vec{h}\| + b \|\vec{x}\| < \|a\vec{x} - \vec{h}\| + \beta = \alpha. \text{ } I \text{ is thus open.}$$

Lemma 10: Suppose

$\{F_n, n \in \mathbb{N}\}$  converges weakly to  $F$ . Let, for  $B \in \mathcal{B}[H]$ ,

$$Q_n(B) = \int_0^\infty P_a(B) F_n(da)$$

$$Q(B) = \int_0^\infty P_a(B) F(da).$$

Then  $\{Q_n, n \in \mathbb{N}\}$  converges weakly to  $Q$ .

proof: Let  $g(\vec{x}) = \int_0^\infty I_B(a\vec{x}) F(da)$ .  $g_n$  is defined similarly,  $F_n$  replacing  $F$ .

Then:  $Q(B) = \int_H P(d\vec{x}) g(\vec{x})$

and  $Q_n(B) = \int_H P(d\vec{x}) g_n(\vec{x}).$

Let  $B$  be open: it is a union of open balls, so that  $I_B(\vec{a}_n)$ , as a function of  $a$ , is an open set of the real line (Lemma 9). Consequently, by weak convergence,  $g(\vec{x}) \leq \underline{\lim}_n g_n(\vec{x})$  and thus, by Fatou's lemma,  $Q(B) \leq \underline{\lim}_n Q_n(B)$ .

Proposition 7 Let  $\{F_n, n \in \mathbb{N}\}$  be a sequence of discrete probabilities with no mass at the origin which converges weakly to  $F$ . Then

$$I(Q) \leq \underline{\lim}_n H(F_n) + I(P)$$

proof:  $I$  is lower semicontinuous for weak convergence.

Remark: Taking  $F$  to be the uniform distribution, one sees that the bound of Proposition 7 will not in general be useful. This seems to indicate that there is little hope to study the general case of  $F$  starting with finite dimensional approximations. This is due to the form taken by  $I(Q)$  in Proposition 6: indeed, in general,  $H(F_n)$  does not tend to  $H(F)$ .



References

- R.B. Ash (1972)  
Real Analysis and Probability  
Academic Press, New York
- C.R. Baker (1973)  
Joint measures and cross covariance operators  
Trans. Amer. Math. Soc. 186, 273-289.
- C.R. Baker (1978)  
Capacity of the Gaussian channel without feedback  
Information and Control 37, 70-89.
- R.C. Douglas (1966)  
On majorization, factorization and range inclusion of operators in  
Hilbert space, Proc.Amer.Math.Soc.17, 413-415.
- A.F. Gualtierotti (1980)  
On average mutual information and capacity for a channel without  
feedback and contaminated Gaussian noise  
Information and Control 46, 46-70.
- M. Hitsuda and S. Ihara (1975)  
Gaussian channels and the optimal coding  
J. Multivariak Anal.5, 106-118.
- T.T. Kadota, M. Zakai and J. Ziv (1971)  
Mutual information of the white Gaussian channel with and without feedback,  
IEEE Trans.Information Theory 17, 368-371.
- J. Keilson and F.W. Steutel (1974)  
Mixtures of distributions, moment inequalities and measures of  
exponentiality and normality,  
Ann. Prob.2, 112-230.
- C.R. Rao and V.S. Varadarahan (1968)  
Discrimination of Gaussian processes  
Sankhyā A 25, 303-330.
- J. Weidmann (1980)  
Linear Operators in Hilbert Spaces  
Springer Verlag, New York.