Absolute Continuity of Poisson Random Fields

By

Yoichiro TAKAHASHI*

§ 0. Introduction

Let R be a locally compact Hausdorff space with countable basis. Given a nonatomic nonnegative Radon measure λ on R, we denote the Poisson measure (or random fields or point processes) with intensity λ by π_λ.

Theorem. *Let λ and ρ be two nonatomic infinite nonnegative Radon measures on R.   Then the Poisson measures π_λ and π_ρ are mutually absolutely continuous if and only if*
(a) *λ and ρ are mutually absolutely continuous and*
(b) *the Hellinger distance $d(\rho, \lambda)$ between λ and ρ is finite*:

$$(1) \qquad d(\rho, \lambda)^2 = \frac{1}{2}\int_R |\sqrt{d\rho} - \sqrt{d\lambda}|^2 < \infty.$$

Furthermore, the Hellinger distance $D(\pi_\rho, \pi_\lambda)$ between π_ρ and π_λ is then given by the formula

$$(2) \qquad D(\pi_\rho, \pi_\lambda)^2 := \frac{1}{2}\int (\sqrt{d\pi_\rho} - \sqrt{d\pi_\lambda})^2 = 1 - \exp\{-d(\rho, \lambda)^2\}.$$

The main purpose of the present note is to give a proof to Theorem above. In the last section we shall apply Theorem to the problem of giving the precise definition of the *Fisher information* or, equivalently, of giving a statistically natural Riemannian metric for an *"infinite dimensional statistical model"*, which consists of mutually absolutely continuous Poisson measures.

Remark. (i) Assume (a) and denote

$$(3) \qquad \phi = \frac{d\rho}{d\lambda}.$$

Then the Radon-Nikodym derivative ϕ is positive and finite λ-almost everywhere

and the condition (b) means that

(1') $\sqrt{\phi} - 1 \in L^2(R, \lambda)$.

In fact, by definition,

$$\int_R |\sqrt{d\rho} - \sqrt{d\lambda}|^2 = \int_R |\sqrt{\phi} - 1|^2 d\lambda = \int_R |1 - 1/\sqrt{\phi}|^2 d\rho .$$

(ii) The case where R is an Euclidean space is already studied by A. V. Skorokhod in 1957 [4]. He expresses (1') by two conditions

$$\int_{|\phi-1|>1/2} |\phi-1| d\lambda < \infty \quad \text{and} \quad \int_{|\phi-1|\leq 1/2} |\phi-1|^2 d\lambda < \infty .$$

(iii) The main idea of the proof of Theorem is the reduction to a theorem of Kakutani on the absolute continuity between product measures, though the proof will be given directly.

(iv) The motivation of the present note was the necessity to prove (1') for the treatment of a Dirichlet form in [5].

First of all let us recall a characterization of Poisson measures by the Laplace transform. (cf. [3].)

Let us denote by Q the set of integer-valued nonnegative Radon measures on R: an element ξ of Q can be written as sum of unit point masses

$$\xi = \sum_i \delta_{x_i} .$$

Since ξ is a Radon measure, the points x_i's form a *locally finite configuration* in the sense that

$$\xi(K) = \#\{i ; x_i \in K\} < \infty$$

for each compact subset K of the base space R.

The space Q is endowed with the relative topology as a closed subset of the space of Radon measures on R with vague topology, so that for each continuous function f on R with compact support the functional

(4) $\langle \xi, f \rangle = \sum_i f(x_i)$

is continuous and, conversely, such functionals generate the topology of the space Q.

The Poisson measure π_λ is characterized by its Laplace transform. Indeed, there holds the formula

(5) $\int_Q \pi_\lambda(d\xi) \exp\{-\langle \xi, f \rangle\} = \exp\left\{-\int_R (1 - e^{-f}) d\lambda\right\}$

for any nonnegative continuous function f on R with compact support. Of

course, the formula (5) remains valid for every Borel functions f provided that the integral

$$\int_R |1-e^{-f}| \, d\lambda$$

is finite.

In particular, for every Borel subset B of R

(6) $$\pi_\lambda \{\xi ; \xi(B)=0\} = \exp\{-\lambda(B)\},$$

(7) $$\pi_\lambda \{\xi ; \xi(B)>0\} = 1-\exp\{-\lambda(B)\},$$

where we understand $\exp(-\infty)=0$.

From (6) and (7) follows a preparatory observation:

Lemma 0. *If π_λ and π_ρ are mutually absolutely continuous, then for each Borel subset B of R*

(8) $$0<\lambda(B)<\infty \quad iff \quad 0<\rho(B)<\infty.$$

In particular, the measures λ and ρ are also mutually absolutely continuous.

§1. Proof of Theorem: "If" Part

Assume that the Radon measures λ and ρ are mutually absolutely continuous and set

(1) $$\phi = \frac{d\rho}{d\lambda}.$$

By the assumptions (a) and (b) of Theorem,

(2) $$0<\phi<\infty \quad \lambda\text{-almost everywhere, and}$$

(3) $$\int_R (\sqrt{\phi}-1)^2 \, d\lambda = \int_R |\sqrt{d\rho}-\sqrt{d\lambda}|^2 < \infty.$$

First of all let us prove the "if" part under a stronger assumption.

Lemma 1. *Assume (2) and that the total variation of the signed measure $\rho-\lambda$ is finite:*

(4) $$\int_R |d\rho-d\lambda| = \int_R |\phi-1| \, d\lambda < \infty.$$

Then π_ρ and π_λ are mutually absolutely continuous and

(5) $$\frac{d\pi_\rho}{d\pi_\lambda}(\xi) = \exp\left\{\langle \xi, \log \phi \rangle + \int_R (1-\phi) \, d\lambda\right\}.$$

Proof. Let f be any nonnegative continuous function R with compact support. Then,

$$\int_Q \pi_\rho(d\xi)\exp\{-\langle\xi, f\rangle\}$$

$$=\exp\left\{-\int_R(1-e^{-f})d\rho\right\}=\exp\left\{-\int_R(1-e^{-f})\phi d\lambda\right\}$$

$$=\exp\left\{-\int_R(1-\phi)d\lambda\right\}\exp\left\{-\int_R(1-e^{-f+\log\phi})d\lambda\right\}$$

$$=\int_Q \pi_\lambda(d\xi)\exp\left\{\langle\xi, \log\phi\rangle-\int_R(1-\phi)d\lambda\right\}\exp\{-\langle\xi, f\rangle\}.$$

Consequently, we obtain (5).

Remark. The integrability of the expression in the right hand of (5) is guranteed by the definiteness of the integral

$$\int_R(1-e^{-f}\phi)d\lambda.$$

Lemma 2. *Assume (2) and (3). Then for every positive constant M, we obtain*

(6)
$$\int_{\phi>\exp M}(\phi+1)d\lambda<\infty, \quad and$$

(7)
$$\int_{\phi<\exp(-M)}(\phi+1)d\lambda<\infty.$$

Proof. By the symmetry in ρ and λ it suffices to prove (6) only. But it is obvious since

$$(\sqrt{x}-1)^2=x+1-2\sqrt{x}>C(x+1) \quad if \quad x>\exp M$$

for some constant C depending on M.

From now on we always assume the conditions (2) and (3).

Corollary. *Set*

(8)
$$E=\left\{x; \phi(x)\geqq 2 \ or \ \phi(x)\leqq\frac{1}{2}\right\}.$$

Then,

(9)
$$\int_E|\phi-1|d\lambda<\infty.$$

Now let us take an exhausting increasing sequence of compact subsets K_i

(i.e., K_i is contained in K_{i+1} for each i and the union of K_i's is the whole space R) and set

(10) $$B_0 = E \quad \text{and} \quad B_n = E \cup K_n \quad (n \geq 1).$$

Lemma 3. *Let ξ_B be the restriction of the measure ξ to a Borel subset B and set*

(11) $$Y_n(\xi) = \langle \xi_{B_n}, \log \phi \rangle + \int_{B_n} (1-\phi) d\lambda.$$

Then the following limit exists π_λ-almost everywhere:

(12) $$Y(\xi) = \lim_{n \to \infty} Y_n(\xi).$$

Furthermore, the convergence of (12) also holds in $L^2(\pi_\lambda)$.

Proof. Set $B_n' = B_n \setminus B_{n-1}$. Then the random variables ξ_{B_0} and $\xi_{B_n'}$, $n \geq 1$, are independent because π_λ is Poisson. Hence the random variables Z_n defined by

$$Z_n = Y_n - Y_{n-1} \ (n \geq 1) \quad \text{and} \quad Z_0 = Y_0$$

are independent. Furthermore, if

$$2 \geq \phi(x) \geq \frac{1}{2},$$

then there is a constant C such that

$$|\log \phi + 1 - \phi| \leq C(\sqrt{\phi} - 1)^2 \quad \text{and} \quad |\log \phi|^2 \leq C(\sqrt{\phi} - 1)^2.$$

Consequently, the condition (3) implies that the series

(13) $$\sum_{n=1}^{\infty} E[Z_n] = \sum_{n=1}^{\infty} \int_{B_n'} (\log \phi + 1 - \phi) d\lambda \quad \text{and}$$

$$\sum_{n=1}^{\infty} E[Z_n^2] = \sum_{n=1}^{\infty} \int_{B_n'} (\log \phi)^2 d\lambda + \sum_{n=1}^{\infty} \left\{ \int_{B_n'} (\log \phi + 1 - \phi) \right\}^2 d\lambda$$

are both absolutely convergent. Hence, by a lemma of Kolmogorov (cf., e.g., [2] p. 249; also Lemma 1 in the next section)

$$\sum_n Z_n \text{ converges almost surely}.$$

Also it is obvious from (13) that the sum converges in $L^2(\pi_\lambda)$. Consequently, we obtain

(14) $$Y = \lim_{n \to \infty} Y_n \text{ exists } \pi_\lambda\text{-almost surely and in } L^2(\pi_\lambda).$$

Remark. The almost sure convergence of Y_n follows also from the

martingale convergence theorem applied to the martingale $\exp Y_n$. But we need the tightness, which will be asserted in the next lemma.

Lemma 4. *The limit $Y(\xi)$ defined by* (12) *satisfies*

(15) $$\int_Q \pi_\lambda(d\xi) \exp Y(\xi) = 1.$$

Proof. On one hand, by the Fatou lemma,

$$\int_Q \pi_\lambda(d\xi) \exp Y(\xi) = \int_Q \pi_\lambda(d\xi) \liminf_{n \to \infty} \exp Y_n(\xi)$$

$$\leq \liminf_{n \to \infty} \int_Q \pi_\lambda(d\xi) \exp Y_n(\xi) = 1.$$

On the other hand, by the Jensen inequality,

$$\int_Q \pi_\lambda(d\xi) \exp Y(\xi) = \int_Q \pi_\lambda(d\xi) \exp\{Y(\xi) - Y_n(\xi)\}$$

$$\geq \exp \int_Q \pi_\lambda(d\xi) \{Y(\xi) - Y_n(\xi)\}$$

$$\geq \exp\left\{-\int_Q \pi_\lambda(d\xi) |Y(\xi) - Y_n(\xi)|^2\right\} \longrightarrow 1$$

as $n \to \infty$. Consequently, we obtain (15).

Proof of "if" part. Let f be a nonnegative continuous function on R with compact support. Take an integer n so large that the support of f is contained in the set B_n defined by (10). Then,

$$\int_Q \pi_\rho(d\xi) \exp\{-\langle \xi, f \rangle\} = \int_Q \pi_\rho(d\xi) \exp\{-\langle \xi_{B_n}, f \rangle\}$$

$$= \exp\left\{-\int_{B_n} (1 - e^{-f}) d\rho\right\} = \exp\left\{-\int_{B_n} (1 - e^{-f}) \phi d\lambda\right\}$$

$$= \prod_{k=0}^{n} \exp\left\{-\int_{B_k'} (1 - \phi) d\lambda - \int_{B_k'} (1 - e^{-f}) \phi d\lambda\right\} \qquad (B_0' = B_0)$$

$$= \prod_{k=0}^{n} \int_Q \pi_\lambda(d\xi) \exp\{-\langle \xi_{B_k'}, f \rangle + Y_k(\xi) - Y_{k-1}(\xi)\} \qquad (Y_{-1} = 0)$$

$$= \int_Q \pi_\lambda(d\xi) \exp\{-\langle \xi_{B_n}, f \rangle + Y_n(\xi)\}$$

$$= \int_Q \pi_\lambda(d\xi) \exp\{Y_n(\xi)\} \exp\{-\langle \xi, f \rangle\}.$$

By Lemma 3 we can let $n\to\infty$ and find

$$\int_Q \pi_\rho(d\xi)\exp\{-\langle\xi,\,f\rangle\}=\int_Q \pi_\lambda(d\xi)\exp\{Y(\xi)\}\exp\{-\langle\xi,\,f\rangle\}\,.$$

Consequently, the measures π_ρ and π_λ are mutually absolutely continuous and the Radon-Nikodym derivative is given by

$$(16) \qquad \frac{d\pi_\rho}{d\pi_\lambda}(\xi)=\exp Y(\xi)\,.$$

The proof of the "if" part is completed.

§2. Proof of Theorem: "Only If" Part

Let ρ and λ be two nonatomic infinite nonnegative Radon measure on R and assume π_ρ and π_λ mutually absolutely continuous. As we already observed, the measure ρ and λ are mutually absolutely continuous. Set

$$(1) \qquad \phi=\frac{d\rho}{d\lambda}\in(0,\,\infty)\,.$$

We need the following well-known fact (cf., e. g., [2]).

Lemma 1 (*Three series theorem*). *Let Z_i, $i\geq1$, be independent random variables. If*

$$(2) \qquad \sum_i Z_i \ \textit{converges almost surely},$$

then for every positive constant a the three series

$$(3) \qquad \sum_i P\{|Z_i|>a\}\,, \quad \sum_i E[Z_i^a] \quad and \quad \sum_i E[(Z_i^a)^2]$$

converge, where

$$(4) \qquad Z_i^a=Z_i \quad if \quad |Z_i|\leq a\,; \ =0 \quad otherwise.$$

Conversely, if the three series (3) *converge for some positive constant a, then* (2) *holds.*

Let us use Lemma 1 to show the following:

Lemma 2. *For every positive constant M,*

$$(5) \qquad \lambda\{x\in R\,;\,\phi(x)>1+M\}<\infty$$

and

$$(6) \qquad \int_{\phi>1+M}\phi\,d\lambda<\infty\,.$$

Proof. The assertion (6) follows from (5) by virtue of Lemma 0 since

$$\rho\{x\in R \,;\, \phi(x)>1+M\}=\int_{\phi>1+M}\phi d\lambda.$$

To prove (5) let us assume the contrary. Then there would exist an infinite Borel partition $\{B_i\}$ of the set

$$\{x\in R \,;\, \phi(x)>1+M\}$$

such that for each i

(7) $\lambda(B_i)=1$

because λ is assumed nonatomic.

Now consider

(8) $X_i=\mathrm{sgn}\,\xi(B_i)=\begin{cases} 1 & if \quad \xi(B_i)>0 \\ 0 & if \quad \xi(B_i)=0. \end{cases}$

They form a Bernoulli scheme under either of π_λ and π_ρ. Let P and Q be its laws under π_λ and π_ρ, respectively. Then,

(9) $\begin{cases} Q(X_i=0)=\exp\{-\rho(B_i)\}, \quad Q(X_i=1)=1-\exp\{-\rho(B_i)\}, \\ P(X_i=0)=e^{-1} \quad and \quad P(X_i=1)=1-e^{-1}. \end{cases}$

The probability measures P and Q are mutually absolutely continuous because so are π_λ and π_ρ. Hence an argument due to Kakutani is applicable and the Radon-Nikodym derivative is given by

(10) $\dfrac{dQ}{dP}=\exp\sum_i Z_i\in(0,\,\infty)$ *P-almost surely,*

where

(11) $Z_i=1[X_i=0]\{1-\rho(B_i)\}+1[X_i=1]\log\dfrac{1-e^{-\rho(B_i)}}{1-e^{-1}}.$

Thus the random variables Z_i are mutually independent and

(12) $\sum_i Z_i$ *converges.*

In particular,

(13) $P\{Z_i<-M$ *infinitely often*$\}=0$

and so, by the second Borel-Cantelli lemma,

(14) $\sum_i P\{Z_i<-M\}<\infty.$

On the other hand, $Z_i<-M$ if $X_i=0$ since $\phi>1+M$ on each B_i. Hence we would obtain

(15) $$\sum_i P\{Z_i < -M\} \geqq \sum_i P\{X_i = 0\} = \sum_i e^{-1} = +\infty,$$

a contradiction to (14). Consequently, we can conclude (5).

Lemma 3. *For every positive constant* M,

(16) $$\lambda\{x \in R ; \phi(x) < 1 - M\} < \infty.$$

Proof. Interchange the role of ρ and λ in Lemma 2.

Proof of "only if" part. Take an arbitrary positive constant M. From Lemmas 2 and 3 we already know that

(17) $$\int_{\phi > e^M \text{ or } \phi < e^{-M}} |\phi - 1| d\lambda < \infty$$

and, therefore, that

(18) $$\int_{\phi > e^M \text{ or } \phi < e^{-M}} |\sqrt{\phi} - 1|^2 d\lambda < \infty.$$

To complete the proof of "only if" part it remains to prove

(19) $$\int_{1 < \phi \leqq e^M} |\sqrt{\phi} - 1|^2 d\lambda < \infty \qquad \text{and}$$

(20) $$\int_{1 > \phi \geqq e^{-M}} |\sqrt{\phi} - 1|^2 d\lambda < \infty.$$

If the set $\{x \in R ; 1 < \phi(x) \leqq e^M\}$ has λ-measure finite, (19) is trivial. Otherwise, we can take a Borel partition $\{B_i\}$ of this set with $\lambda(B_i) = 1$ such that

(21) $$B_i \text{ is contained in } \{x ; b_i \geqq \sqrt{\phi}(x) - 1 \geqq b_{i+1}\}$$

for some nonincreasing sequence $b_0 \geqq b_1 \geqq \cdots \geqq b_n \to 0$. Let us use the same notation as in the proof of Lemma 3. Now Z_i are bounded random variables and the three series theorem implies

$$\sum_i E[Z_i^2] < \infty.$$

Since $E[Z_i^2] \geqq E[Z_i^2 ; X_i = 0] = e^{-1}\{\rho(B_i) - \lambda(B_i)\}^2$, we obtain

(22) $$\sum_i \left\{\int_{B_i} (\phi - 1) d\lambda\right\}^2 < \infty.$$

(By the way, (22) follows also from $\sum_i E[Z_i] < \infty$.) Noting that $\lambda(B_i) = 1$ and that $\phi - 1 \geqq b_{i+1}^2 + 2b_{i+1}$ on B_i, it follows from (22) that

(23) $$\sum_i b_i^2 \leqq \sum_i (b_i^2 + 2b_i) < \infty.$$

But (23) implies, using $\lambda(B_i)=1$ again, that

$$\int_{1<\phi\leq e^M}|\sqrt{\phi}-1|^2 d\lambda \leq \sum_i b_i^2 < \infty.$$

Hence we obtain (19) and, automatically, (20) since the latter is nothing but (19) if ρ and λ are interchanged.

Consequently, from (17)–(20) we obtain

$$\int_R |\sqrt{d\rho}-\sqrt{d\lambda}|^2 = \int_R |\sqrt{\phi}-1|^2 d\lambda < \infty,$$

as is desired. The proof of the "only if" part is completed.

§3. The Hellinger Distance: Proof of the Final Part of Theorem

Let us compute the Hellinger distance between mutually absolutely continuous Poisson measures π_ρ and π_λ:

$$D(\pi_\rho, \pi_\lambda)=\frac{1}{2}\int_Q (\sqrt{d\pi_\rho}-\sqrt{d\pi_\lambda})^2.$$

Lemma. *If the total variation* $\int_R |d\rho-d\lambda|$ *is finite, then*

$$D(\pi_\rho, \pi_\lambda)=1-\exp\{-d(\rho, \lambda)\},$$

where $d(\rho, \lambda)$ *is the Hellinger distance between* ρ *and* λ.

Proof. Recall Lemma 1 of Section 1. Under the assumption of Lemma the Radon-Nikodym derivative can be written explicitly:

$$d\pi_\rho/d\pi_\lambda=\exp\{\langle\xi, \log\phi\rangle-\langle\lambda, \phi-1\rangle\},$$

where $\phi=d\rho/d\lambda$. Consequently,

$$D(\pi_\rho, \pi_\lambda)=\frac{1}{2}\int\pi_\lambda(d\xi)\left[\left\{1-\exp\frac{1}{2}\{\langle\xi, \log\phi\rangle-\langle\lambda, \phi-1\rangle\}\right]^2\right.$$

$$=1-\int\pi_\lambda(d\xi)\exp\frac{1}{2}\{\langle\xi, \log\phi\rangle-\langle\lambda, \phi-1\rangle\}$$

$$=1-\exp\left\{-\frac{1}{2}\langle\lambda, \phi-1\rangle+\langle\lambda, \sqrt{\phi}-1\rangle\right\}$$

$$=1-\exp\left\{-\frac{1}{2}\langle\lambda, (\sqrt{\phi}-1)^2\rangle\right\}$$

$$=1-\exp\{-d(\rho, \lambda)\}.$$

Proof of Theorem: the final part. As was proved in Section 2 (14), the Radon-Nikodym derivative $d\pi_\rho/d\pi_\lambda$ can be expressed as $\exp Y$, $Y = \lim Y_n$ and

(3) $$Y_n(\xi) = \langle \xi, \log \phi_n \rangle - \langle \lambda, \phi_n - 1 \rangle$$

where

(4) $$\phi_n = \phi \quad \text{on } B_n; \ = 1 \quad \text{otherwise}.$$

Since $\int |\phi_n - 1| d\lambda$ is finite, we can apply Lemma to find

$$D(\pi_\rho, \pi_\lambda)^2 = 1 - \int \pi_\lambda(d\xi) \exp \frac{1}{2} Y(\xi)$$

$$= 1 - \lim_{n \to \infty} \int \pi_\lambda(d\xi) \exp \frac{1}{2} Y_n(\xi)$$

$$= 1 - \lim_{n \to \infty} \exp \left\{ -\frac{1}{2} \int (\sqrt{\phi_n} - 1)^2 d\lambda \right\}.$$

Noting that $|\sqrt{\phi_n} - 1| \leqq |\sqrt{\phi} - 1|$ and that $|\sqrt{\phi} - 1|$ is square integrable, one obtains

$$D(\pi_\rho, \pi_\lambda)^2 = 1 - \exp \left\{ -\frac{1}{2} \int (\sqrt{\phi} - 1)^2 d\lambda \right\} = 1 - \exp \{ -d(\rho, \lambda)^2 \}.$$

The proof is completed.

§4. Application: Hilbert-Riemannian Manifold Associated with Fisher Information

The geometrical structure of statistical models has been investigated intensively; cf. [1]. Since we obtained the explicit formula for the Radon-Nikodym derivative $d\pi_\rho/d\pi_\lambda$ between Poisson measures π_ρ and π_λ, we can compute the Fisher information for the "infinite dimensional statistical model"

(1) $$Psn(\lambda) = \{ \pi_\rho; \ \rho \sim \lambda, \ d(\rho, \lambda) < \infty \}.$$

The goal of the present section is to show the following:

Proposition. *The space $L^2(\rho)$ can be regarded as the tangent space of $Psn(\lambda)$ at π_ρ and the Fisher information evaluated at π_ρ is equal to the symmetric bilinear form*

(2) $$F(\rho)[u, v] := \int_R uv \, d\rho$$

defined on the space $L^2(\rho)$.

Furthermore, the statistical model $Psn(\lambda)$ equipped with the Fisher information as Riemannian metric is a Hilbert-Riemannian manifold modelled by $L^2(\lambda)$.

Remark. The former part of Proposition above may rather be thought of as a definition justified by Lemmas 1–2 below. The latter part is a usual mathematical assertion. By Proposition above one may take a domain of $L^2(\lambda)$ as the parameter space for $Psn(\lambda)$.

One of the implication of the Fisher information is the *Cramer-Rao inequality*, which takes the following form in the present case.

Corollary. *Let $T(\xi)$ be a Radon measure valued random variable defined on the space Q and assume that it is an unbiased estimator for ρ : namely, assume that for each π_ρ there holds the equality*

$$(3) \qquad \int_Q \pi_\rho(d\xi)\langle T(\xi), f\rangle = \int_R f\,d\rho$$

for any continuous function f on R with compact support, where

$$(4) \qquad \langle T(\xi), f\rangle = \int_R f(x)T(\xi)(dx).$$

Let us denote the variance of T under π_ρ by

$$(5) \qquad \mathrm{Var}_\rho(\langle T, f\rangle) := \int_Q \pi_\rho(d\xi)\left\{\langle T(\xi), f\rangle - \int_R f\,d\rho\right\}^2 \quad (\leqq \infty).$$

Then there holds the inequality for the average T_n of n independent copies of T :

$$(6) \qquad \mathrm{Var}_\rho(\langle T_n, f\rangle) \geqq \frac{1}{n}\int_R f^2\,d\rho.$$

The minimum is attained by the random variable $T(\xi)=\xi$.

Let us recall some basic terminology. In [1] a family M of probability measures on a given measurable space $(\Omega, \mathcal{B}_\Omega)$ is called a *statistical model* if each member is absolutely continuous with respect to a fixed measure, say μ, on $(\Omega, \mathcal{B}_\Omega)$. Usually it is assumed in statistics that a statistical model is parametrized by a finite number of coordinates, although we are going to remove this restriction. So we temporally set

$$(7) \qquad M=\{P_\theta \,;\, \theta \in \Theta\}$$

and assume that the parameter space Θ is a domain of some Euclidean space R^n.

Assuming that each P_θ is absolutely continuous with respect a measure μ on $(\Omega, \mathcal{B}_\Omega)$, we denote its Radon-Nikodym derivative by $dP_\theta/d\mu$ and set

$$(8) \qquad l(\theta)=\log dP_\theta/d\mu.$$

The random variable $l(\theta)$ on $(\Omega, \mathcal{B}_\Omega)$ is called the *(logarithmic) likelihood.*

Now assume that $l(\theta)$ is differentiable in θ and denote the partial derivatives by

$$(9) \qquad \partial_i l(\theta) = \frac{\partial}{\partial \theta_i} l(\theta).$$

Moreover, assume that the random variables (9) are square-integrable with respect to P_θ for each θ and set

$$(10) \qquad F_{ij}(\theta) = E_\theta [\partial_i l(\theta) \partial_j l(\theta)].$$

Then one obtains the matrix $F(\theta) = (F_{ij}(\theta))$ called the *Fisher matrix*. Obviously it is nonnegative definite and is associated with the nonnegative definite symmetric bilenear form

$$(10') \qquad F(\theta)[u, v] = \sum F_{ij}(\theta) u_i v_j, \quad u = (u_i), \ v = (v_i),$$

which we refer as the *Fisher information*.

The third assumption is that $F(\theta)$ is positive definite for each θ. If we take

$$ds^2 = \sum F_{ij}(\theta) d\theta_i d\theta_j$$

as the Riemannian metric on M, then the statistical model M becomes a Riemannian manifold. This assumption is automatic if there is an *unbiased estimator* T of θ, i.e., a random variable T defined on $(\Omega, \mathcal{B}_\Omega)$ such that $E_\theta[T] = \theta$ for all θ.

Let us generalize the above construction of statistical manifold to our model $Psn(\lambda)$. Since it is infinite dimensional, one cannot define the derivative of the likelihood nor the Fisher information in a direct way. One has to extend those notions to "prove" Proposition.

We shall take the space Q as Ω and the following function p as the coordinate of π_ρ in $Psn(\lambda)$:

$$(11) \qquad p = \log(d\rho/d\lambda).$$

In other words, we set $P_p = \pi_\rho$ with ρ defined by $d\rho = e^p d\lambda$. Then it follows from Theorem that

$$(12) \qquad Psn(\lambda) = \{P_p ; p \in M\}$$

where M is the infinite dimensional space

$$(13) \qquad M = \{p \in \boldsymbol{L}^0(\lambda), \ e^{p/2} - 1 \in \boldsymbol{L}^2(\lambda)\}.$$

Here $\boldsymbol{L}^p(\lambda)$ stands for the $\boldsymbol{L}^p$-space for $p > 0$ and $\boldsymbol{L}^0(\lambda)$ denotes the space of (equivalence classes of) λ-measurable functions on R.

Remark. Under the choice of the coordinate stated above the model $Psn(\lambda)$

turns out to be an exponential family in statisticians' terminology. Precisely to say, it is so in the sense that, at least for p such that e^p-1 belongs to $L^1(\lambda)$ the likelihood $l(p)$ is linear in p up to a nonlinear deterministic additive constant, namely,

$$l(p)(\xi)=\log dP_p/dP_0(\xi)=\langle \xi,\ p\rangle-\int_R(e^p-1)d\lambda.$$

as is shown in Lemma 1 of Section 1.

What we are to do is to introduce the structure of differentiable manifold modelled by a certain topological vector space, say E, which satisfies the following requirements:

(a) The likelihood $l(p)$ is differentiable in each direction u from E.

(b) The directional derivative $D_u l(p)$ is a square integrable random variable under P_p:

$$D_u l(p)\in L^2(P_p) \qquad \text{if} \quad u\in E.$$

Since we are concerned with the density functions it is natural to consider the weak topology (precisely to say, weak* topology) on $Psn(\lambda)$. Hence *we say that*

$$p_n \text{ converges to } p \text{ in } M$$

if for any continuous function f on R with compact support

$$\lim\int f\exp p_n d\lambda=\int f\exp p\ d\lambda.$$

Definition. Let us call a curve $c(t)$ in M *smooth* if it is continuously differentiable with respect to the weak topology and if

(14)
$$\lim_{h\to\infty}\int_R\left\{\frac{e^{c(t+h)}-e^{c(t)}}{h}-\frac{dc}{dt}(t)\right\}d\lambda=0.$$

Remark. (i) It turns out that a condition such as (14) is inavoidable for the likelihood to be differentiable along $c(t)$.

(ii) Another possible condition in place of (14) is that $\exp\{c(t)/2\}-1$ is differentiable in $L^2(\lambda)$ and the integral $\int_R c'(t)\exp\{c(t)/2\}d\lambda$ is meaningful. But we do not discuss it.

Lemma 1. *Let* $p\in M$ *and* $u\in L^2(\lambda_p)$. *Then there is a smooth curve* $c(t)$, $0\leq t\leq 1$, *in* M *such that*
$$c(0)=p \qquad and \qquad c'(0)=u.$$

Proof. Keeping in mind that $|t|e^{-|t|}\leq e^{-1}$ for any real t, set

$$c(t) = p + 2\log\left[1 + \frac{t}{2}ue^{-t|u|}\right].$$

Then,

$$e^{c(t)/2} - 1 = (e^{p/2} - 1) + \frac{t}{2}ue^{-t|u|}e^{p/2}.$$

Since $ue^{p/2} \in L^2(\lambda)$, one obtains

$$e^{c(t)/2} - 1 \in L^2(\lambda).$$

Furthermore, $e^{c(t)/2}$ is differentiable λ-almost everywhere and in $L^2(\lambda)$ and

$$c'(t) = (1 - t|u|)e^{-t|u|}e^{p/2 - c(t)/2}u \in L^2(\lambda_{c(t)}).$$

In particular, $c'(0) = u$.

Finally, it is immediate to see

$$c'(t) = (1 - t|u|)e^{-t|u|}e^{p/2 - c(t)/2}u \in L^2(\lambda_{c(t)}).$$

In particular, $c'(0) = u$.

Finally, it is immediate to see

$$e^{c(t)} - e^p - he^p u = e^p\left\{-h(1 - e^{-h|u|})u + \frac{1}{4}h^2u^2e^{-h|u|}\right\} = O(h^2) \quad \text{in} \quad L^1(\lambda).$$

The condition (14) follows from this and similar estimates for $t > 0$.

Now let us use the following conventional notation. We denote

(15) $$:\langle \xi, f\rangle + \langle \lambda, g\rangle:$$

for functions f and g on R if, for any measurable partition $\{B_n\}$ by relatively compact subsets, the random variables

$$X_n = \int_{B_n} f(x)\xi(dx) + \int_{B_n} g(x)\lambda(dx)$$

are well-defined and their sum $\sum X_n$ converges in $L^2(\lambda_p)$ and is independent of the choice of the partition $\{B_n\}$, p being given.

Lemma 2. *Let $p \in M$ and $u \in L^2(\lambda_p)$. Then for any smooth curve $c(t)$ with $c(0) = p$ and $\dfrac{dc}{dt}(0) = u$ the likelihood $l(\cdot)$ is differentiable at $t = 0$ along $c(t)$ in $L^2(\lambda_p)$ and the derivative is given by*

(16) $$\frac{d}{dt}\Big|_{t=0} l(c(t)) = D_u l(p)(\xi) = :\langle \xi, u\rangle - \langle \lambda_p, u\rangle:.$$

(In particular, it depends only on p and u.) Moreover,

(17) $$E_p[(D_u l(p))^2]=\int_R u^2 d\lambda_p=\int_R u^2 e^p d\lambda.$$

Proof. For the first moment let us assume that u belongs to the set

(18) $K=\{u\in L^0(\lambda)\,;\,u$ *is bounded and with compact support*$\}$.

Then, $p+tu\in M$ for any real t. Let us show that the Gateaux derivative exists and is given by

(19) $$\frac{d}{dt}\Big|_{t=0} l(p+tu)=\langle\xi,\,u\rangle-\langle\lambda,\,u\rangle.$$

If p is such that $e^p-1\in L^1(\lambda)$, then so is $c(t)=p+tu$ for any t and

$$l(c(t))(\xi)=\langle\xi,\,c(t)\rangle-\langle\lambda,\,e^{c(t)}-1\rangle$$

$$=l(p)(\xi)+t\langle\xi,\,u\rangle-\langle\lambda_p,\,e^{tu}-1\rangle.$$

Consequently, we obtain (19). Now let p be arbitrary. Still we have $c(t)=p+tu\in M$ but we must take the operation $:\cdot:$ and

$$l(c(t))(\xi)=\,:\langle\xi,\,c(t)\rangle-\langle\lambda,\,e^{c(t)}-1\rangle:.$$

Let $\{B_n\}$ be a measurable partition $\{B_n\}$ by relatively compact subsets and set

(20) $$X_n(t)=\int_{B_n} c(t)(x)\xi(dx)-\int_{B_n}(e^{c(t)(x)}-1)\lambda(dx).$$

Then,

(21) $$X_n(t)=X_n(0)-\int_{B_n}(e^{tu(x)}-1)\lambda_p(dx)$$

and they are well-defined and are mutually independent. Since u is assumed to have compact support, $X_n(t)=X_n(0)$ except for a finite number of n's. Furthermore,

$$\frac{1}{t}\{X_n(t)-X_n(0)\}-\int_{B_n} u(x)\{\xi(dx)-\lambda_p(dx)\}$$

$$=-\frac{1}{t}\int_{B_n}\{e^{tu(x)}-1-tu(x)\}\lambda_p(dx)\longrightarrow 0$$

as $t\to0$. Hence the proof of (19) is completed.

Now let $u\in L^2(\lambda_p)$ be arbitrary. Then, $p+tu$ is not necessarily a curve in M and we cannot take the Gateaux derivative any more. We take a smooth curve $c(t)$ in M with $c(0)=p$ and $c'(0)=u$ and compute the derivative along $c(t)$. Define $X_n(t)$ by (20), as before. Then,

$$\frac{1}{t}\{X_n(t)-X_n(0)\}-\int_{B_n} u(x)\{\xi(dx)-\lambda_p(dx)\}$$

$$=-\frac{1}{t}\int_{B_n}\{e^{c(t)(x)}-e^{p(x)}-tu(x)e^{p(x)}\}\lambda(dx).$$

Consequently,

$$\frac{1}{t}\{l(c(t))-l(p)\}-:\langle\xi-\lambda_p,\ u\rangle:$$

$$=-\frac{1}{t}\int_R\{e^{c(t)}-e^p-tue^p\}\lambda(dx)\longrightarrow 0$$

as $t\to 0$ by the condition (14). Hence, $D_u l(p)(\xi)=:\langle\xi-\lambda_p,\ u\rangle:$

Finally, (14) follows from the direct computation.

Now it will be quite natural to give the following definition.

Definition. Let us call the space $L^2(\rho)$ the *tangent space* at π_ρ of $Psn(\lambda)$, denote

(22) $$T_\rho Psn(\lambda)=L^2(\rho)$$

and call the following bilinear form the *Fisher information* at π_ρ:

(23) $$F(\rho)[u,\ v]=\int_Q D_u l(p)D_v l(p)d\pi_\rho \qquad (u,\ v\in T_\rho Psn(\lambda)).$$

It follows from Lemma 2 that

(24) $$F(\rho)[u,\ v]=\int_R uvd\rho \qquad (u,\ v\in T_\rho Psn(\lambda)).$$

The proof of Proposition will be completed if we prove the following.

Lemma 3. *Take the Fisher information $F(p)$ as Riemannian metric on $Psn(\lambda)$. Then it is a trivial Hilbert-Riemannian manifold modelled by $L^2(\lambda)$. Moreover, it is an open submanifold of the manifold $L^2(\lambda)$.*

Proof. The space $Psn(\lambda)$ is identified with the space

$$M=\{p\in L^0(\lambda)\,;\ e^{p/2}-1\in L^2(\lambda)\}\,.$$

The latter can be identified with the open subset

$$\{f\in L^2(\lambda)\,;\ f+1>0\ a.e.\}$$

of the space $L^2(\lambda)$ by the injection $\phi:M=Psn(\lambda)\to L^2(\lambda)$ defined by

(25) $$\phi(p)=e^{p/2}-1\,.$$

Denote $\rho=\lambda_p$. Then the map ϕ induces the following map $(d\phi)_p$ from $T_\rho Psn(\lambda)=L^2(\lambda_p)$ onto $L^2(\lambda)$ in a natural way:

$$(d\phi)_p u = \frac{1}{2} e^{p/2} u \qquad (u \in L^2(\lambda_p)).$$

Consequently, ϕ gives a global chart of $Psn(\lambda)$ and $(d\phi)_p$ is an isomorphism of $T_p Psn(\lambda)$ onto $L^2(\lambda)$ up to the constant factor $1/2$. Hence, the Hilbert-Riemannian manifold $(Psn(\lambda), F(\rho))$ can be identified with the subset $\{f \in L^2(\lambda);$ $f + 1 > 0 \ a.e.\}$ of $L^2(\lambda)$.

Proof of Corollary. Let $\rho = \lambda_p$, i.e., $d\rho = e^p d\lambda$ and f be a continuous function on R with compact support. By the assumption

$$E_p[\langle T, f \rangle] = E_0[\langle T, f \rangle \exp l(p)] = \int_R f e^p d\lambda = \int_R f d\rho.$$

For each $u \in K$ we can take the Gateaux derivatives of both sides and obtain

$$E_p[\langle T, f \rangle D_u l(p)] = \int_R f u d\rho.$$

On the other hand, from the identity $E_0[\exp l(p)] = 1$ we obtain

$$E_p[D_u l(p)] = 0.$$

Hence,

$$\left\{ \int_R f u d\rho \right\}^2 = E_p[\{\langle T, f \rangle - E_p[\langle T, f \rangle]\} D_u l(p)]^2$$

$$\leq E_p[\{\langle T, f \rangle - E_p[\langle T, f \rangle]\}^2] E_p[\{D_u l(p)\}^2]$$

$$= \text{Var}_\rho(\langle T, f \rangle) F(\rho)[u, u]$$

by Schwartz' inequality. Consequently,

$$\text{Var}_\rho(\langle T, f \rangle) \geq \sup_{u \in K} \left\{ \int_R f u d\rho \right\}^2 / F(\rho)[u, u]$$

$$= \sup_{u \in K} \left\{ \int_R f u d\rho \right\}^2 \Big/ \int_R u^2 d\rho = \int_R f^2 d\rho$$

because K is dense in $L^2(\rho)$. Hence follows (6) and the proof is completed.

Remark. In our statistical model $Psn(\lambda)$ the geodesic $P_{p(s)}$ joining two Poisson measures $\pi_\sigma = P_q$ and $\pi_\rho = P_p$ with coordinate q and p, i.e., $\sigma = \lambda_q$ and $\rho = \lambda_p$, is given by

(26) $$e^{p(s)} = \{(1-s)e^{q/2} + s e^{p/2}\}^2 \qquad (0 \leq s \leq 1).$$

In other words, the geodesic in our Riemannian metric is linear in the square root of the density of intensity measure and the geodesic distance $dist(\pi_\rho, \pi_\lambda)$ coincides with the Hellinger distance (up to the constant factor $1/2$):

$$(27) \qquad dist(\pi_\rho, \pi_\lambda) = \left\{ \int_R (\sqrt{d\rho} - \sqrt{d\lambda})^2 \right\}^{1/2}.$$

References

[1] Amari, S., *Differential-Geometrical Methods in Statistics, Lecture Notes in Statistics*, Vol. 28, Springer, 1985.

[2] Billingsley, P., *Probability and Measures*, John Wiley, 1979.

[3] Matthes, K., Kerstan, J. and Mecke, J., *Infinitely Divisible Point Processes*, John Wiley, 1978.

[4] Skorohod, A.V., On the differentiability of measures which correspond to stochastic processes I. Processes with independent increments, *Teoria Verojat. Primen.* **2** (1957), 418–444.

[5] Takahashi, Y., Scattering length, capacity and large deviation for the equilibrium process of infinitely many independent transient Markovian particles, *in preparation*.