UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
CURSO DE MESTRADO EM MATEMÁTICA

RELAÇÕES ENTRE A CLÁSSICA MÉTRICA DE BERGMAN E A MÉTRICA DE INFORMAÇÃO

José Mairton Barros da Silva

Fortaleza
1994

RELAÇÕES ENTRE A CLÁSSICA MÉTRICA DE BERGMAN E A MÉTRICA DE INFORMAÇÃO

José Mairton Barros da Silva

Aprovada em 28/12/94

BANCA EXAMINADORA

Franquiberto dos Santos Pessoa
(Orientador)

Gregório Pacelli F. Bessa
Gregório Pacelli Feitosa Bessa

Raimundo Benedito do Nascimento
Raimundo Benedito do Nascimento

UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
CURSO DE MESTRADO EM MATEMÁTICA

RELAÇÕES ENTRE A CLÁSSICA MÉTRICA DE BERGMAN E A MÉTRICA DE INFORMAÇÃO

José Mairton Barros da Silva

Monografia apresentada ao Curso de Mestrado em Matemática como requisito parcial para a obtenção do grau de Mestre em Matemática.

Fortaleza
1994

À meus pais, meus filhos, a ELA que tem me suportado e convivido comigo estes anos todos do curso de mestrado e a todos aqueles que me ajudaram direta ou indiretamente.

AGRADECIMENTOS

Agradecemos a pessoas e instituições que colaboraram para a elaboração deste trabalho:

Ao Professor Franquiberto dos Santos Pessoa, pela orientação segura e pela paciência e compreensão .

Ao Professor Gregório Pacelli pela ajuda nas horas difíceis da Análise Funcional e da Geometria.

Ao Professor Raimundo Benetito pela colaboração dada na realização deste trabalho.

Ao Professor Antônio Gervásio Colares em especial.

Ao Professor Francisco Gesario pela amizade, ajuda e incentivo que sempre me deu.

Ao Professor Sebastião pela compreensão, e paciência de sempre.

Ao Professor Berniz, Chefe do Centro Tecnológico, pela ajuda, pela compreensão de sempre e pelo empréstimo de seu equipamento para conclusão deste trabalho, pois sem ele não teria sido possível.

À Coordenação da pós-graduação em Matemática.

Aos colegas no Departamento de Matemática da Universidade Federal do Maranhão, pelo incentivo.

Enfim, a todos que de forma direta, ou indiretamente nos ajudaram a realizar este trabalho.

APRESENTAÇÃO

O Trabalho que hora apresentamos constitui minha monografia de mestrado, apresentada com requisito parcial para a obtenção do grau de mestre em matemática.

Ele esta baseado no trabalho devido a **Jacob Burbea e C. Radhakishna Rao [10].**

Um dos objetivos deste trabalho é relacionar a métrica diferencial de informação com a clássica métrica de Bergman. Para este propósito, definimos um funcional ϕ-entropia sobre o espaço de distribuição de probabilidades e em seguida calculamos a matriz Hessiana ao longo de uma direção do espaço tangente. Usamos, também as ideias de medidas divergentes para calcularmos a métrica diferencial de informação. Finalmente, definimos núcleo reproduzido e analisamos a métrica diferencial de informação usando núcleos de Bergman.

NOTAS HISTÓRICAS

Em 1945, Rao notou a importância do enfoque geométrico-diferencial na estatística. Ele introduziu a métrica Riemanniana na variedade (diferencial) de um modelo estatístico e calculou a distância geodésica entre duas distribuições para alguns modelos estatísticos. Essas ideias causaram um enorme impacto na comunidade científica (matemática e estatística), porém, devido a enorme dificuldade matemática dessa teoria ela permaneceu inerte por muito tempo.

Atualmente, propriedades da variedade Riemanniana de um modelo estatistístico estão sendo estudadas por um número muito grande de pesquisadores. Mas. mesmo assim, as implicações estatísticas destes conceitos não são muito conhecidas.

SUMÁRIO

CAPÍTULO I

RELAÇÕES ENTRE A CLÁSSICA MÉTRICA DE BERGMAN E A MÉTRICA DE INFORMAÇÃO

INTRODUÇÃO

Nosso trabalho, esta baseado no trabalho de **Jacob Burbea e C. Radhakrishna Rao[10]** e tem por objetivo a construção da métrica diferencial através de funcionais entropia, analisando suas relações com a matriz de informação. Em particular, estudaremos as relações entre a métrica de Bergman e a métrica de informação.

Considere μ uma medida σ-finita sobre uma σ-álgebra de subconjuntos de um espaço amostral χ.

Define-se respectivamente o produto interno e a norma do espaço de Hilbert (separável) $L_2(\chi;\mu)$ de funções de valores complexos que têm quadrado integrável, pondo

$$(\mathbf{f},\mathbf{g})_\mu = \int_\chi \mathbf{f}(\mathbf{t})\overline{\mathbf{g}(\mathbf{t})}\mathbf{d}\mu(\mathbf{t}) \quad \mathbf{e} \quad \|\mathbf{f}\|_\mu = \sqrt{(\mathbf{f},\mathbf{f})_\mu}$$

Também define-se

$$S_\mu = \left\{f \in L_2(\chi;\mu); \|f\|_\mu = 1\right\} \qquad \text{e}$$
$$P_\mu = \left\{|f| \in L_2(\chi;\mu); |f|^{1/2} \in S_\mu\right\}$$

respectivamente a esfera unitária do $L_2(\chi;\mu)$ e o conjunto de densidades de probabilidades.

Seja D uma variedade suave mergulhada em C^n e

$$\text{(1.1)} \qquad \rho(z,w) = \left\{1 - \int_\chi \left[p(t/z)p(t/w)\right]^{1/2} d\mu(t)\right\}^{1/2} \qquad z,w \in D$$

com $p(./.) \in F(\chi/D)$, onde $F(\chi/D)$ e um subconjunto de funções densidades de probabilidades de um conjunto aberto U de um espaço de Frechet F* de funções f(./.) definidas sobre $\chi \times D$ que inclui o espaço tangente a U.

Lembramos que $\rho(z,w)$ é conhecida como a pseudo-distância de ordem-1 ou pseudo-distância de Hellinger.

Seja $\psi(./.)$ uma função onda normalizada tal que

$$(1.2)\qquad \begin{aligned} &p(t/z)=\left|\psi(t/z)\right|^2 ,\ p(./.)\in F(\chi/D) ,\ (t/z)\in\chi\times D \\ &\left\|\psi(t/z)\right\|_\mu^2=1 \end{aligned}$$

A pseudo-distância projetiva sobre D é definida como

$$(1.3)\qquad \lambda(z,w)=\left\{1-\left|\int_\chi \psi(t/z)\overline{\psi(t/w)}d\mu(t)\right|\right\}^{1/2}$$

Em alguns casos, se a função onda $g(./z)\in L_2(\chi;\mu)$, $z\in D$ não for normalizada, como em (2), consideraremos certas constantes de normalização como a seguir

$$(1.4)\qquad g(./z)=\sqrt{K(z,\bar{z})}\psi(./z) ,\ \psi(./z)\in S_\mu ,\ z\in D$$

onde

$$K(z,\bar{z})=(g(./z),g(./z))_\mu$$

é uma função suave em $z\in D$.

A abordagem acima descrita será usada no Teorema seguinte, o qual representa o ponto fundamental do nosso trabalho.

TEOREMA

Seja $p(./.)\in F(\chi/D)$ a função densidade de probabilidade, dada por

$$(1.5)\qquad p(t/z)=\left|g(t/z)\right|^2 \Big/ K(z,\bar{z}) ,\ K(z,\bar{z})>0 ,\ z\in D$$

g(t/.) é uma função holomorfa em D μ-quase todo $t\in\chi$. Então:

i) $K(z,\overline{w})$ é o núcleo de Bergman de um espaço de Hilbert de funções holomorfas em D;

ii) g(./.), definida em $\chi\times D$, é a função geradora de $L_2(\chi,\mu)$ para H(D), onde H(D) é o único espaço de Hilbert de funções holomorfas em D, tal que

$$k_z(z)=K(z,\overline{w}) ,\ z,w\in D ,$$

é o seu núcleo reproduzido;

iii) A pseudo-distância projetiva λ é a pseudo-distância de Skwarcýnski[19].

iv) Para todo $z,w\in D$, $0\le\rho(z,w)\le\lambda(z,w)\le 1$, onde ρ é a pseudo-distância de Hellinger e λé a pseudo-distância projetiva;

v) A métrica de informação ds^2 é a métrica de Bergman db^2, que pode ser expressa como

$$(1.6) \qquad db^2 = \partial\bar{\partial}\log K$$

vi)

$$\lambda^2(z,w) = d\lambda^2(z,w)\big|_{w=z} = 0 \quad \text{e} \quad d\lambda^2(z,w)\big|_{w=z} = ds^2(z) = db^2(z) \geq 0.$$

Uma importante regra em discussão de problemas de inferência são as medidas de distância entre distribuições de probabilidades **[16]**.Porém, uma grande classe de medidas, não satisfazem os requisitos de uma função distância. Estas medidas, denominadas Medidas de Divergentes, são muito usadas com frequência em problemas de taxonomia, mais especificamente nos trabalhos que tratam da classificação biológica.

Existe uma abordagem sobre a unificação aproximada para a construção de distâncias e medidas de divergência.

A distância geodésica, introduzida pela mudança de métrica diferencial quadrática em espaços de distribuição de probabilidade nos dá um método para especificar a diferença entre duas distribuições de probabilidades**[17]**.

2. MÉTRICA DIFERENCIAL DE INFORMAÇÃO.

Considere μ uma medida de probabilidade σ-finita sobre uma σ-álgebra de um espaço amostral χ.

Definimos

$$(f,g)_\mu = \int_\chi f(t)\overline{g(t)}d\mu(t)$$

e

$$\|f\|_\mu = \sqrt{(f,f)_\mu}$$

respectivamente, o produto interno e a norma do espaço de Hilbert separável $L_2(\chi,\mu)$ de funções de valores complexos que têm quadrado integrável sobre χ com respeito à medida μ.

Desta forma, a esfera unitária de $L_2(\chi,\mu)$ e o conjunto de densidades de probabilidades com parâmetros contínuos reais são, respectivamente, definidos por

$$S_\mu = \left\{f \in L_2(\chi,\mu); \|f\|_\mu = 1\right\} \text{ e}$$
$$P_\mu = \left\{|f| \in L_1(\chi,\mu); |f|^{1/2} \in S_\mu\right\}.$$

Frequentemente em estatística tratamos uma família parametrizada de distribuição de probabilidades como um modelo estatístico. Seja S = {p(t/z)} tal modelo estatístico, onde $t \in \chi$ é uma variável aleatória e p(t/z) é a função densidade de probabilidade de t, parametrizada por z, em relação à uma medida μ sobre χ. Aqui, z é um parâmetro real n-dimensional $z = (z_1, z_2, \ldots, z_n)$ pertencente a um subconjunto $D \subset R^n$. Por exemplo, o modelo normal é uma família de distribuições de probabilidade tendo a seguinte função densidade de probabilidade

$$p(t/z)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right)$$

onde o espaço amostral χ é a reta real R com a medida de Lebesgue $d\mu(t)=dx$ e o parâmetro z é bi-dimensional, ou seja

$$z=(z_1,z_2)=(\mu,\sigma);$$

onde μ e σ são os parâmetros usuais que especificam uma distribuição normal. Desta maneira, o conjunto de parâmetros D é o semi-plano

$$D=\{(\mu,\sigma);-\infty<\mu<\infty,\sigma>0\}$$

Consequentemente, o conjunto S contém todas as distribuições normais, e cada distribuição normal $N(\mu,\sigma^2)$ em S é bem definida pelo parâmetro bi-dimensional $z=(\mu,\sigma)$.

De um modo geral , $D\subset R^k$.No entanto, quando k=2n, é mais conveniente utilizar a terminologia complexa. Neste sentido, para os parâmetros $z_j=x_j+y_j$, $j=1,2,...,n$, a diferenciação complexa nos fornece:

$$\partial_{z_j}=2^{-1}(\partial_{x_j}-i\partial_{y_j})$$
$$\partial_{\bar{z}_j}=2^{-1}(\partial_{x_j}+i\partial_{y_j}),$$

onde j = 1, ... , n, e portanto, para uma função f de classe C^1, a derivada de f ao longo de $z\in D$ é dada por:

$$df=(\partial+\bar{\partial})f,\ df=df(z),\ z\in D,$$

com

$$\partial f = \sum_{j=1}^{n} \partial_{z_j} f dz \quad e \quad \bar{\partial} f = \sum_{j=1}^{n} \partial_{\bar{z}_j} f d\bar{z}_j$$

Observe que com esta notação , $\partial_{\bar{z}_j}$ pode ser escrita como $\bar{\partial}_{z_j}$ e se todos os y_j são zeros, D é uma variedade mergulhada em R^n e a notação se reduz a notação real.

Suponhamos portanto que D seja uma variedade mergulhada em C^n e que a família $F = F(\chi / D)$ de funções densidade de probabilidades $p = p(./.) \in P_\mu$ satisfaça, para todo $z \in D$ as seguintes condições

i) $\forall t \in \chi$, $p(t|.)$ é diferenciável em D μ-quase por toda parte

ii) $\partial_{z_j} \int_\chi p(t|z) d\mu(t) = \int_\chi \frac{\partial}{\partial_{z_j}} p(t|z) d\mu(t)$ e, portanto, igual a zero.

DEFINIÇÃO 2.1

A matriz de informação de Fisher de $p = p(.|.)$, $p(.|.) \in F(\chi|D)$ é a $n \times n$ matriz hermitiana $F = [g_{k\bar{m}}]$ cujos elementos são dados por

$$g_{k\bar{m}}(z) = \int_\chi p^{-1} (\partial_{z_k} p)(\bar{\partial}_{z_m} p) d\mu(t)$$

ou

$$g_{k\bar{m}}(z) = \int_\chi p(\partial_{z_k} \log p)(\bar{\partial}_{z_m} \log p) d\mu(t)$$

Pela condição (ii), anteriormente citada, $E\{\partial_{z_k} \log p(.|z)\} = 0$, $k = 1, \ldots n$. Portanto, a matriz de informação, para $z \in D$, é uma matriz hermitiana de covariância de $\{\partial_{z_k} \log p(.|z)\}$, $k = 1, \ldots, n$.

DEFINIÇÃO 2.2

Usando a definição anterior, definimos a métrica diferencial de informação relativamente a $p(.|.) \in F(\chi|D)$ como a forma quadrática hermitiana

$$ds^2(z)=\sum_{k,m=1}^{n} g_{k\bar{m}}\, dz_k\, d\bar{z}_m \quad , \quad g_{k\bar{m}}=g_{k\bar{z}}(z),\ z\in D,$$

PROPOSIÇÃO 2.1

A forma quadrática ds^2 é definida positiva e além disso, em termos do $L_2(\chi;\mu)$ podemos escreve-la como

$$ds^2(z)=\left\|p^{1/2}\partial \log p\right\|_{\mu}^{2},\ p=p(.|z),\ z\in D.$$

PROVA

De fato,

$$\begin{aligned}
ds^2(z) &= \sum_{k,m=1}^{n}\int_{\chi} p\left(\partial_{z_k}\log p\right)\left(\bar{\partial}_{z_m}\log p\right)d\mu(t)\\
&= \sum_{k,m=1}^{n}\left\{\int_{\chi}\left(p^{1/2}\partial_{z_k}\log p\right)\left(p^{1/2}\bar{\partial}_{z_m}\log p\right)d\mu(t)\right\}dz_k\, d\bar{z}_m\\
&= \int_{\chi}\sum_{k,m=1}^{n}\left\{\left(p^{1/2}\partial_{z_k}\log p\right)\left(p^{1/2}\bar{\partial}_{z_m}\log p\right)dz_k\, d_{\bar{z}_m}\right\}d\mu(t)\\
&= \int_{\chi}\left(p^{1/2}\sum_{k,m=1}^{n}\partial_{z_k}(\log p)dz_k\right)\left(p^{1/2}\sum_{k,m=1}^{n}\bar{\partial}_{z_m}(\log p)d\bar{z}_m\right)d\mu(t)\\
&= \int_{\chi}\left(p^{1/2}\partial\log p\right)\left(p^{1/2}\bar{\partial}\log p\right)d\mu(t)\\
&= \left\|p^{1/2}\partial\log p\right\|_{\mu}^{2}
\end{aligned}$$

que é positiva definida •

PROPOSIÇÃO 2.2

A métrica diferencial de informação $ds^2(z)$ é localmente invariante sob transformações holomorfas bijetivas de z.

PROVA

Seja $\varphi: D^* \to D$ uma transformação bijetiva holomorfa. Então $p^* = p^*(.|.)$ é definida por

$$p^{*}(.|w)=p(.|\varphi(z)),\quad w\in D^{*},\quad p^{*}(.|.)\in F(\chi|D^{*}),$$

e

$$g^{*}_{k\bar{m}}(w)=\int p^{*}(\partial_{w_k}\log p^{*})(\bar{\partial}_{w_m}\log p^{*})d\mu(t)$$

ou ainda

$$g^{*}_{k\bar{m}}(w)=\int_{\chi}p^{*^{-1}}(\partial_{w_k}p^{*})(\bar{\partial}_{w_m}p^{*})d\mu(t),$$

os quais são os elementos da matriz de informação de Fisher.

Agora

$$ds^{2}_{*}(w)=\sum_{k,m=1}^{n}g^{*}_{k\bar{m}}dw_k\bar{d}w_m,\quad g^{*}_{k\bar{m}}=g^{*}_{k\bar{m}}(w),\quad w\in D^{*},$$

que é a métrica de informação correspondente.

Usando a regra da cadeia:

$$\partial_{w_k}p^{*}=\sum_{i=1}^{n}\partial_{z_i}p(.|z)\frac{\partial z_i}{\partial_{w_k}}$$

e

$$\bar{\partial}_{z_m}p^{*}=\sum_{j=1}^{n}\bar{\partial}_{z_j}p(.z)\frac{\bar{\partial}_{z_j}}{\partial_{\bar{w}_m}}$$

Como

$$g^{*}_{k\bar{m}}(w)=\int_{\chi}p^{*-1}\left(\partial_{w_k}p^{*}\right)\left(\partial_{\bar{w}_k}p^{*}\right)d\mu(t)$$

temos:

$$g^{*}_{k\bar{m}}(w)=\int_{\chi} p^{-1}\left(\sum_{i=1}^{n}\partial p(.|z)\frac{\partial z_i}{\partial w_k}\right)\left(\sum_{j=1}^{n}\partial_{\bar{z}_j} p(.|z)\frac{\partial \bar{z}_j}{\partial_{\bar{w}_m}}\right)d\mu(t)$$

Como $z_i, \bar{z}_j, w_k, \bar{w}_m$ não dependem da medida $\mu(t)$, temos

$$g^{*}_{k\bar{m}}(w)=\sum_{i,j=1}^{n}\frac{\partial z_i}{\partial w_k}\frac{\partial \bar{z}_j}{\partial \bar{w}_m}\int_{\chi} p(.|z)(\partial_{z_i} p(.|z))(\bar{\partial}_{z_j} p(.|z))d\mu(t)$$

$$=\sum_{i,j=1}^{n} g_{ij}\frac{\partial z_i}{\partial w_k}\frac{\partial \bar{z}_j}{\partial \bar{w}_m}$$

Portanto os elementos da **matriz de informação** são transformados por uma transformação bi-holomorfa em um **tensor covariante de segunda ordem**.

Além disso,

$$ds_{*}^{2}(w)=\sum_{k,m=1}^{n} g^{*}_{k\bar{m}}(w)dw_k d\bar{w}_m$$

$$=\sum_{k,m=1}^{n}\left(\sum_{i,j=1}^{n} g_{ij}\frac{\partial z_i}{\partial w_k}\frac{\partial \bar{z}_j}{\partial \bar{w}_m}\right)dw_k d\bar{w}_m$$

$$=\sum_{k,m=1}^{n}\left(\sum_{i,j=1}^{n} g_{ij}\frac{\partial z_i}{\partial w_k}dw_k\frac{\partial \bar{z}_j}{\partial \bar{w}_m}d\bar{w}_m\right)$$

$$=\sum_{i,j=1}^{n} g_{ij}dz_i d\bar{z}_j = ds^2(z)$$

•

DEFINIÇÃO 2.3

Uma pseudo-distância sobre $D\times D$ é uma função simétrica

$$\delta: D\times D \to R^{+}$$

que satisfaz a desigualdade triangular, mas podendo ter $\delta(w,z)=0$ com $w\neq z$, $w,z\in D$.

DEFINIÇÃO 2.4

A distância geodésica induzida ds^2 com respeito a $p(.|.) \in F(\chi|D)$ define uma pseudo-distância S_D sobre $D \times D$. Esta pseudo-distância é denominada **pseudo-distância de informação** com respeito a $p = p(.|.)$**[2],[10]**

Se $p(.|.)$ possui a propriedade que garanta ds^2 ser estritamente definida positiva, a pseudo-distância de informação S_D, satisfaz as condições

i) $z \in D$, $S_D(z,w) \geq 0$, $S_D(z,z) = 0 \Leftrightarrow z = 0$

ii) $S_D(z,w) = S_D(w,z)$ $\forall z, w \in D$

iii) $S_D(z,w) \leq S_D(z,\delta) + S_D(\delta,w)$ $\forall z, w, \delta \in D$

ou seja S_D satisfaz as condições de distância. Portanto, torna-se-à uma distância sobre $D \times D$ no sentido canônico. Com estas propriedades, S_D é conhecida como a **distância de informação com respeito a p=p(.|.).** Assim, D torna-se um espaço métrico (D,S_D).

O espaço métrico (D,S_D) será completo, se para cada $w \in D$ e r>0, a bola fechada

$$\{z \in D; S_D(z,w) \leq r\}$$

for um conjunto compacto.

DEFINIÇÃO 2.5

Uma métrica Riemanniana numa variedade diferenciável D é uma correspondência, que associa a cada ponto $z \in D$ um produto interno definido positivo no espaço tangente.

Como S_D é induzida pela métrica Riemanniana ds^2, a complecidade anteriormente definida é equivalente a complecidade usual, visto trabalharmos com a bola fechada e limitada.

PROPOSIÇÃO 2.3

Suponhamos que além das propriedades anteriormente exigidas, $p(.|.) \in F(\chi|D)$ também satisfaça a seguinte propriedade

iii) $z \in D$, $\partial_{z_k}\partial_{z_m} \log p(.|z)$ existe μ-quase por toda parte.

Então, a matriz informação de Fisher $\left(g_{k\bar{m}}\right)$ é dada pela expressão mais simples

$$g_{k\bar{m}} = -E\left\{\partial_{z_k}\bar{\partial}_{z_m}\log p(.|z)\right\}$$

e a métrica informação, ds^2, toma a forma

$$ds^2 = -E\left\{\partial\bar{\partial}\log p(.|z)\right\}$$

PROVA

$$-E\left\{\partial_{Z_K}\bar{\partial}_{Z_M}\log p(.|z)\right\} = -\int_{\chi} p(t|z)\partial_{z_k}\bar{\partial}_{z_m}\log p(t|z)d\mu(t) =$$

$$= -\int_{\chi} p(t|z)\partial_{z_k}\left[\frac{1}{p(t|z)}\bar{\partial}_{z_m}p(t|z)d\mu(t)\right] =$$

$$= -\int_{\chi} p(t|z)\left[\frac{-1}{p^2(t|z)}\partial_{z_k}p(t|z)\bar{\partial}_{z_m}p(t|z) + \frac{1}{p(t|z)}\partial_{z_k}\bar{\partial}_{z_m}p(t|z)\right]d\mu(t) =$$

$$= \int_{\chi} p^{-1}(.|z)\partial_{z_k}p(t|z)\bar{\partial}_{z_m}p(t|z)d\mu(t) - \int_{\chi}\partial_{z_k}\bar{\partial}_{z_m}p(t|z)d\mu(t) =$$

$$= \int_{\chi} p^{-1}(t|z)\partial_{z_k}p(t|z)\bar{\partial}_{z_m}p(t|z)d\mu(t) = g_{k\bar{m}}$$

pois

$$-\int_{\chi}\partial_{z_k}\bar{\partial}_{z_m}p(t|z)d\mu(t) = 0.$$

Portanto

$$ds^2 = \sum_{k,m=1}^{n} g_{km} dz_k d\bar{z}_m =$$

$$= \sum_{k,m=1}^{n} \left[\int_{\chi} p^{-1} \partial_{z_k} p(t|z) \bar{\partial}_{z_m} p(t|z) d\mu(t) \right] dz_k d\bar{z}_m =$$

$$= \sum_{k,m=1}^{n} \int_{\chi} p(t|z) \partial_{z_k} \{\log p(t|z)\} dz_k \bar{\partial}_{z_m} \{\log p(t|z)\} d\bar{z}_m d\mu(t) =$$

$$= -\int_{\chi} \left(\sum_{k=1}^{n} p^{1/2}(t|z) \partial_{z_k} \log(t|z) dz_k \right) \left(\sum_{m=1}^{n} p^{1/2}(t|z) \bar{\partial}_{z_m} \log p(t|z) d\bar{z}_m \right) d\mu(t) =$$

$$= -\int_{\chi} p(t|z) \left(\sum_{k=1}^{n} \partial_{z_k} \log p(t|z) dz_k \right) \left(\sum_{m=1}^{n} \bar{\partial}_{z_m} \log p(t|z) d\bar{z}_m \right) d\mu(t) =$$

$$= -\int_{\chi} \left(p^{1/2}(t|z) \partial \log p(t|z) \right) \left(p^{1/2}(t|z) \bar{\partial} \log p(t|z) \right) d\mu(t) =$$

$$= -E\{\partial \bar{\partial} \log p(.|z)\}$$

CAPÍTULO III

3.ENTROPIA E MEDIDAS DE DIVERGÊNCIA

3.1 ENTROPIA

Seja $F = F(\chi / D)$ uma família de funções densidade de probabilidades contida em um conjunto aberto U, onde U é um subconjunto aberto de um espaço de Frechet $\tilde{F}$ de funções $f = f(./.)$, definidas sobre $\chi \times D$, que inclui o espaço tangente de U.

Definimos a tangente de f(t/z) em $z \in D$ na direção de $(u,v) \in C^n \times C^n$ como

$$d_{(u,v)}f(./z) = \partial_u f(./z) + \partial_v f(./z)$$

onde

$$(3.1) \qquad \begin{cases} \partial_u f(./z) = \sum_{k=1}^{n} \partial_{z_k} f(./z) u_k \\ \partial_v f(./z) = \sum_{k=1}^{n} \bar{\partial}_{z_k} f(./z) v_k \end{cases} .$$

Para $p(./.) \in F(\chi / D)$ e φ uma função de classe C^2 côncava, definimos o funcional φ-entropia

$$(3.2) \qquad \{H_\varphi(p)\}(z) = \int_\chi \varphi[p(t/z)] d\mu(t).$$

A derivada de H_φ em $p \in U$ na direção de $f \in \tilde{F}$ é dada por

$$dH_\varphi(p;f) = \frac{d}{dt} H_\varphi(p + tf)\bigg|_{t=0} = \int_\chi \varphi'[p(t/z)] f(t) d\varphi(t)$$

e a derivada segunda, por $p \in U$ na direção de $g \in \tilde{F}$ é

$$d^2 H_\varphi(p;f,g) = \int_\chi \varphi''[p(t/z)] f(t) g(t) d\mu(t)$$

O Hessiano em $p \in U$ na direção de $f \in \tilde{F}$ é definido por

$$(3.3) \qquad \Delta_f H_\varphi(p) = 4 d^2 H_\varphi(p; f, \bar{f})$$

isto é

$$\Delta_f H_\varphi(p) = 4\int_\chi \varphi''[p(t/z)]|f|^2 d\mu(t),$$

pois φ é uma função côncava, $-\varphi$ é convexa, logo

$$(3.4) \qquad -\Delta_f H_\varphi(p) \geq 0, \quad f \in \tilde{F}$$

Fazendo $f = \partial_u p$, $u = (dz_1, dz_2, \ldots, dz_n) \in C^n$, temos

$$(3.5) \qquad \Delta_{\partial_u p}\{H_\varphi(p)\}(z) = 4\int_\chi \varphi''[p(t/z)]|\partial_u p(t/z)|^2 d\mu(t)$$

onde $\partial_u p(t/z) = \sum_{k=1}^{n} \partial_{z_k} p(t/z) u_k$.

3.2 MÉTRICA ENTROPIA DE ORDEM α.

Abordaremos neste trabalho uma classe de funções, φ_α, $\alpha > 0$, dada por

$$(3.6) \qquad \varphi_\alpha(s) = \begin{cases} (\alpha-1)^{-1}(s - s^\alpha), & \alpha \neq 1 \\ -s\log s, & \alpha = 1 \end{cases}.$$

cuja derivada segunda é dada por

$$(3.7) \qquad \varphi''(s) = \begin{cases} -\alpha s^{\alpha-2}, & \alpha \neq 1 \\ -\dfrac{1}{s}, & \alpha = 1 \end{cases}$$

Estas funções são definidas para todo s>0 e serão estendidas para $s \in R_+$, usando a convenção

$$0\log 0 = 0$$

A função φ_α é côncava, pois sua derivada segunda é sempre negativa.

Para esta classe de funções, o funcional φ-entropia será denotado por H_α e será denominado funcional entropia de ordem α.

Segue-se então,

$$(3.8) \qquad \{H_\alpha(p)\}(z) = \begin{cases} (\alpha-1)^{-1}\left[1 - \int_\chi p^\alpha d\mu(t)\right], & \alpha \neq 1 \\ -\int_\chi p\log p\, d\mu(t), & \alpha = 1 \end{cases}$$

cuja matriz Hessiana será denotada por:

$$\text{(3.9a)}\qquad \Delta_{\partial_u p}\{H_\alpha(p)\}(z) = 4\begin{cases} -\alpha\int_\chi p^{\alpha-1}|\partial_u p|^2 d\mu(t), & \alpha \neq 1 \\ -\int_\chi \frac{1}{p}|\partial_u p|^2 d\mu(t), & \alpha = 1 \end{cases}$$

ou

$$\text{(3.9b)}\qquad \Delta_{\partial_u p}\{H_\alpha(p)\}(z) = -4\begin{cases} \alpha\int_\chi p^{\alpha}|\partial_u \log p|^2 d\mu(t), & \alpha \neq 1 \\ \int_\chi p^{-1}|\partial_u \log p|^2 d\mu(t), & \alpha = 1 \end{cases}$$

Portanto,

$$= -4\begin{cases} -\alpha\int_\chi\left(\sum_{m=1}^{n}\bar{\partial}_{z_m}\log p\, d\bar{z}_m\right)d\mu(t), & \alpha \neq 1 \\ \int_\chi p\left(\sum_{k=1}^{n}\partial_{z_k}\log p\, dz_k\right)\left(\sum_{m=1}^{n}\bar{\partial}_{z_m}\log p\, d\bar{z}_m\right)d\mu(t), & \alpha = 1 \end{cases}$$

$$\Delta_{\partial_u p}\{H_\alpha(p)\}(z) = -4\alpha\sum_{k,m=1}^{n}\left\{\int_\chi p^\alpha\left(\partial_{z_k}\log p\right)\left(\bar{\partial}_{z_m}\log p\right)d\mu(t)\right\}dz_k d\bar{z}_m$$

Logo, H_α está associado a uma métrica diferencial Hermitiana, a saber;

$$ds_\alpha^2(z) = -\frac{1}{4\alpha}\Delta_{\partial_u p}\{H_\alpha(p)\}(z).$$

que é definida positiva pois φ_α é côncava e portanto uma métrica da geometria Riemanniana.

Esta métrica que pode ainda ser escrita como

$$\text{(3.10)}\qquad ds_\alpha^2(z) = \sum_{k,m=1}^{n} g_{k,\bar{m}}^{(\alpha)} dz_k dz_{\bar{m}}$$

será denominada, métrica entropia de ordem α.

Os coeficientes $g_{k,\bar{m}}^{(\alpha)}$, são os elementos da matriz entropia de ordem α, são dados por

$$\text{(3.11)}\qquad g_{k,\bar{m}}^{(\alpha)}(z) = \int_\chi p^\alpha\left(\partial_{z_k}\log p\right)\left(\bar{\partial}_{z_m}\log p\right)d\mu(t).$$

Em particular, quando $\alpha = 1$, temos $H = H_1$, que é a conhecida entropia de Shannon. E neste caso, $ds^2 = ds_1^2$, $\left[g_{k,\bar{m}}\right] = \left[g^{(1)}_{k,\bar{m}}\right]$ que foram respectivamente definidos como métrica diferencial de informação e a matriz de informação.

3.3 - MEDIDAS DE DIVERGÊNCIA

Uma outra maneira de encontrarmos a métrica diferencial de informação é usando a noção de divergência que definiremos a seguir.

DEFINIÇÃO 3.3.1- Seja $F: R_+ \times R_+ \to R$ uma função de classe C^2 tal que

$$F(s,t) \geq 0 \text{ e } F(s,s) = 0 \ \forall s,t \in R_+$$

Definimos a divergência de $p,q \in P_\mu$ em relação a F como

$$D_F(p,q) = \int_\chi F(p(t),q(t))d\mu(t), \ p,q \in P_\mu$$

PROPOSIÇÃO 3.3.1

Fixando $p \in P_\mu$ e fazendo q variar, temos

$$\mathbf{D_F(p,p) = dD_F(p,q)\big|_{q=p} = 0 \ \text{ e } \ d^2D_F(p,q)\big|_{q=p} \geq 0}$$

PROVA

Como $F \in C^2$ e além disso $F(s,t) \geq 0$ e $F(s,s) = 0$, F tem um mínimo quando s=t. Logo, pela condição de mínimo o resultado segue-se. •

PROPOSIÇÃO 3.3.2

Considere $p = p(t|z)$ e $q = p(t|w)$ com $z,w \in D$ e $p,q \in F(\chi|D)$. Em analogia com (3.5), o Hessiano (complexo) é

$$\Delta_{\partial_u p}\{D_F(p,p)\}(z) = 4\int_\chi \partial_q^2 F(p,q)\big|_{q=p} |\partial p|^2 d\mu(t)$$

que é também uma forma quadrática definida positiva.

PROVA

Pela proposição anterior, temos que

$$\int_\chi \partial_q^2 F(p,q)\big|_{q=p} |\partial_u p|^2 d\mu(t)$$

é maior ou igual a zero, pois $d^2D_F(p,q)\big|_{q=p} \geq 0$. Portanto o resultado segue-se trivialmente. •

Como a medida de divergência é uma forma quadrática definida positiva, também podemos encontrar a métrica entropia de ordem α usando a notação de divergência.

Objetivando o conceito de métrica entropia de ordem α, daremos alguns exemplos de medidas de divergência.

3.4 - EXEMPLOS DE MEDIDAS DE DIVERGÊNCIA

3.4.1 - J - DIVERGÊNCIA

A J-divergência em relação a H_φ induzida pelo funcional φ-entropia H_φ é definida por

$$(3.12) \qquad J_\varphi^{(\lambda)}(p,q) = H_\varphi(\lambda p + (1-\lambda)q) - \lambda H_\varphi(p) - (1-\lambda)H_\varphi(q).$$

Por (3.5) temos

$$(3.13) \qquad J_\varphi^{(\lambda)}(p,q) = \int_\chi \{\varphi(\lambda p - (1-\lambda)q) - \lambda\varphi(p) - (1-\lambda)\varphi(q)\} d\mu(t)$$

Seja $F: R_+ \times R_+ \to R$ definida por

$$F(p,q) = \varphi(\lambda p + (1-\lambda)q) - \lambda\varphi(p) - (1-\lambda)\varphi(q)$$

cuja derivada segunda em relação a q é

$$\frac{d^2F(p,q)}{dq^2} = \varphi''(\lambda p + (1-\lambda)q)(1-\lambda)(1-\lambda)(\partial_u q)(\partial_u q) - (1-\lambda)\varphi''(q)(\partial_u q)(\partial_u q)$$

e quando $q \to p$ temos

$$\frac{d^2F(p,q)}{dq^2} = -\lambda(1-\lambda)\varphi''(p)|\partial_u p|^2$$

que é não negativa, pois φ é uma função côncava.

Portanto,

$$(3.14) \qquad \int_\chi \varphi''(p)|\partial_u p|^2 d\mu(t)$$

que é uma forma quadrática definida positiva.

Em particular, quando φ é a função entropia de ordem α, temos que

$$\Delta_{\partial_u p}\{J_\varphi^{(\lambda)}(p,p)\}(z) = -4\lambda(1-\lambda)\alpha\int_\chi -4\lambda(1-\lambda)p^{\alpha-2}|\partial_u p|^2 d\mu(t)$$

$$\Delta_{\partial_u p}\{J_\varphi^{(\lambda)}(p,p)\}(z) = -4\lambda(1-\lambda)\alpha\int_\chi p^{\alpha-2}|\partial_u p|^2 d\mu(t)$$

Daí,

$$ds_{\alpha}^{2}(z)=\frac{-1}{4\alpha\lambda(1-\lambda)}\Delta_{\partial_u p}\left\{J_{\alpha}^{(\lambda)}(p,p)\right\}(z).$$

Portanto

$$ds_{\alpha}^{2}(z)=\int_{\chi}p^{\alpha-2}\left|\partial_u p\right|^2 d\mu(t)$$

que é a métrica entropia de ordem α.

3.4.2 - K - DIVERGÊNCIA

Seja ψ uma função de classe C^2 em R_+ com a propriedade

$$(3.15)\qquad s\psi(s^{-1})+\psi(s)\geq 0$$

e com a normalização, $\psi(1)\geq 0$.

Definimos a K-divergência por

$$(3.16)\qquad K_{\psi}(p,q)=\frac{1}{8\psi''(1)}\int_{\chi}\left\{p\psi\left(\frac{q}{p}\right)+q\psi\left(\frac{p}{q}\right)\right\}d\mu(t).$$

Considere

$$F(p,q)=p\psi\left(\frac{q}{p}\right)+q\psi\left(\frac{p}{q}\right)$$

que é uma função de classe C^2 cuja derivada segunda, quando $q\rightarrow p$, é

$$\left.\frac{d^2F(p,q)}{dq^2}\right|_{q=p}=\frac{1}{p}\psi''\left(\frac{q}{p}\right)\bar{\partial}_u q\partial_u q+\frac{p^2}{q^3}\psi''\left(\frac{p}{q}\right)\bar{\partial}_u q\partial_u q\bigg|_{q=p},$$

isto é;

$$\left.\frac{d^2F(p,q)}{dq^2}\right|_{q=p}=\frac{2}{p}\psi''(1)\bar{\partial}p\partial p=2\psi''(1)p\left(\bar{\partial}_u\log p\right)\left(\partial_u\log p\right)=2\psi''(1)p\left|\partial_u\log p\right|^2$$

Portanto, o Hessiano da K-divergência é

$$(3.17)\qquad \Delta_{\partial_u p}\left\{K_{\psi}(p,p)\right\}(z)=\int_{\chi}p\left|\partial_u\log p\right|^2 d\mu(t)$$

que é a métrica de informação.

Caso $\psi = -\varphi_\alpha$, onde φ_α é a função entropia de ordem α, a K-divergência resultante K_α, é denominada K-divergência de ordem α ($\alpha>0$).

Neste caso,

$$-\left[s\varphi_\alpha(s^{-1}) + \varphi_\alpha(s)\right] = \begin{cases} (\alpha-1)^{-1}(1-s^\alpha)(s^{1-\alpha}), & \alpha \neq 1 \\ (1-s)\log s^{-1}, & \alpha = 1 \end{cases}$$

a qual é não negativa, para todo $s \in R_+$ e $\varphi''(1) = \alpha > 0$.

Portanto,

$$(3.18) \qquad 8\alpha K_\alpha(p,q) = \begin{cases} (\alpha-1)^{-1}\left[\int_\chi (p^{1-\alpha}q^\alpha + q^{1-\alpha})d\mu(t) - 2\right], & \alpha \neq 1 \\ \int_\chi \left[q \log p^{-1}q + p\log q^{-1}p\right]d\mu(t), & \alpha = 1 \end{cases}$$

onde $8\alpha K_\alpha(p,q)$ é conhecida como a divergência de Kullback-Leibler **[14]**.

3.4.3 - DIVERGÊNCIA DE HELLINGER

A divergência de Hellinger é denotada por

$$(3.19) \qquad H_\psi(p,q) = \int_\chi [\psi(p) - \psi(q)]^2 d\mu(t)$$

onde ψ é uma função de classe C^2 sobre R_+.

Esta medida de divergência, $\{H_\psi(p,q)\}^{1/2}$, define uma pseudo-distância sobre $P_\mu \times P_\mu$. Em particular se p=p(t|z) e q=p(t|w) com $z,w \in D$ e $p(.|.) \in F(\chi|D)$, temos a função

$$(3.20) \qquad \rho_\psi(z,w) = \{H_\psi(p(.|z), p(.|w))\}^{1/2},$$

a qual é conhecida como a pseudo-distância de Hellinger sobre $D \times D$, $D \subset C^n$.

Podemos também escrever $\rho_\psi(z,w)$ sob a forma

$$(3.21) \qquad \rho_\psi(z,w) = \left\{\int_\chi [\psi(p(t|Z)) - \psi(p(t|w))]^2 d\mu(t)\right\}^{1/2}.$$

Cuja derivada segunda, quando $w \to z$ é dada por

$$d^2\rho_\psi(z,w)\Big|_{w=z} = 2\int_\chi [\psi'(p)]^2 |\partial_u p|^2 d\mu(t)$$

em analogia com (3.5), vemos que seu Hessiano é dado por

$$(3.22) \qquad \Delta_{\partial_u p}\{H_\psi(p,p)\}(z) = 8\int_\chi [\psi'(p)]^2 |\partial_u p|^2 d\mu(t).$$

Escolhendo como classe de funções ψ,

(3.23) $$\psi(s) = \frac{1}{\alpha\sqrt{2}} s^{\alpha/2} \ (s \in R_+, \alpha > 0),$$ tem-se

a divergência de Hellinger de ordem α dada por

(3.24) $$H_\alpha(p,q) = \frac{1}{2\alpha^2}\int_\chi [p^{\alpha/2} - q^{\alpha/2}]\mu(t)$$

e a pseudo-distância de Hellinger de ordem α por

(3.25) $$\rho_\alpha(z,w) = \frac{1}{\alpha\sqrt{2}}\left\{\int_\chi \left(p[t|z]^{\alpha/2} - q[t|w]^{\alpha/2}\right)^2 d\mu(t)\right\}^{1/2},$$

cujo Hessiano é

$$\Delta_{\partial_u p}\{H_\alpha(p,p)\}(z) = \int_\chi p^\alpha |\partial_u \log p|^2 d\mu(t)$$

visto que

$$[\psi'(s)]^2 = \frac{1}{8} s^{\alpha-2} |ds|^2 .$$

Logo,

(3.26) $$d^2\rho_\alpha^2(z,w) = \frac{1}{8}\int_\chi p^{\alpha-2}|dp|^2 d\mu(t).$$

Portanto, o Hessiano é a métrica entropia de ordem α, isto é

(3.27) $$\Delta_{\partial_u p}\{H_\alpha(p,p)\}(z) = ds_\alpha^2(z)$$

CAPÍTULO IV

4. A PSEUDO-DISTÂNCIA PROJETIVA

No caso particular $\alpha = 1$, (3.25) é denominada pseudo-distância projetiva, a qual também é conhecida como pseudo-distância de Hellinger sobre D e denotada por ρ.

PROPOSIÇÃO 4.1

A pseudo-distância de Hellinger ρ,

$$\rho(z,w) = \frac{1}{\sqrt{2}}\left\{\int_\chi \left([p(t/z)]^{1/2} - [p(t/w)]^{1/2}\right)^2 d\mu(t)\right\}^{1/2}$$

pode ser escrita como

$$(4.1) \qquad \rho(z,w) = \left\{1 - \int_\chi [p(t/z)p(t/w)]^{1/2} d\mu(t)\right\}^{1/2}$$

além disso

$$(4.2) \qquad \begin{cases} \rho^2(z,w) = d\rho^2(z,w)\big|_{w=z} = 0 & \text{e} \\ d^2\rho(z,w)\big|_{w=z} = \dfrac{1}{4}\displaystyle\int_\chi p^{-1}|dp(t/z)|^2 d\mu(t) \end{cases}$$

Em particular

$$ds^2 = \Delta_{\partial_u p}\rho^2(z,w)\big|_{w=z}$$

que é a métrica de informação

PROVA

Como

$$\rho(z,w) = \frac{1}{\sqrt{2}}\left\{\int_\chi \left([p(t/z)]^{1/2} - [p(t/w)]^{1/2}\right)^2 d\mu(t)\right\}^{1/2}$$

tem-se

$$\rho(z,w) = \frac{1}{\sqrt{2}}\left\{\int_\chi [p(t|z) - 2[p(t|z)p(t|w)] + p(t|w)]d\mu(t)\right\} =$$

$$= \frac{1}{\sqrt{2}}\left\{2 - 2\int_\chi \left[p(t|z)p(t|w)\right]^{1/2} d\mu(t)\right\}^{1/2} = \left\{1 - \int_\chi \left|p(t|z)p(t|w)\right|^{1/2} d\mu(t)\right\}.$$

Logo,

$$\rho^2(z,w) = 1 - \int_\chi \left[p(t|z)p(t|w)\right]^{1/2} d\mu(t).$$

Em particular, quando w=z, tem-se

$$\rho^2(z,z) = 1 - \int_\chi p(t|z) d\mu(t) = 0.$$

Por outro lado,

$$d\rho^2(z,w)\Big| = d\int_\chi \left[p(t/z)p(t/w)\right]^{1/2} d\mu(t) =$$

$$= \int_\chi \left[p(t/z)\right]^{1/2} d\left[p(t/w)\right]^{1/2} d\mu(t) =$$

$$= \int_\chi \left[p(t/z)\right]^{1/2} \sum_{k=1}^{n} \frac{\partial (p(t/w))^{1/2}}{\partial w_k} dw_k d\mu(t) =$$

$$= \sum_{k=1}^{n} \frac{\partial}{\partial w_k}\left\{\int_\chi \left[p(t/z)p(t/w)\right]^{1/2} . d\mu(t)\right\} dw_k \Big|_{w=k} =$$

$$= 0$$

pois $p(t/z)$ é holomorfa e $\int_\chi p(t/z) d\mu(t) = 1$, logo,

$$\frac{\partial}{\partial z_k}\int_\chi p(t/z) d\mu(t) = \int_\chi \frac{\partial}{\partial z_k} p(t/z) d\mu(t) = 0$$

Além disso,

$$d\rho^2(z,w) = \frac{1}{2\alpha^2}\int_\chi 2\left[p(t/z)^{1/2} - p(t/w)^{1/2}\right]\left(-\frac{1}{2}\right) p(t/w)^{-1/2} dp(t/w) d\mu(t)$$

e

$$d^2\rho^2(z,w) = -\frac{1}{2}\int_\chi \left(-\frac{1}{2}\right) p^{-1/2}(t/w) p^{-1/2}(t/w) dp dp d\mu(t) +$$
$$\left(-\frac{1}{2}\right)\int_\chi \left[p^{1/2}(t/z) - p^{1/2}(t/w)\right]\left(-\frac{1}{2}\right) p^{-3/2}(t/w) dp dp d\mu(t) +$$

$$+\left(-\frac{1}{2}\right)\int_{\chi}[p^{1/2}(t/z)-p^{1/2}(t/w)]p^{-1/2}(t/w)d^2p(t/w)d\mu(t).$$

Pondo w=z, tem-se

$$d^2\rho^2(z,w)\Big|_{w=z}=\frac{1}{4}\int_{\chi}p^{-1}|dp|^2d\mu(t).$$

Observando (3.22) e (3.27) concluímos que

$$\Delta_{\partial_u p}\{\rho^2(z,z)\}=ds^2(z). \qquad \bullet$$

Apesar da grande importância da pseudo-distância de Hellinger ρ na teoria da inferência estatística **[17]**, é mais conveniente para nossa proposta de trabalho, estudar uma pseudo-distância alternativa λ, definida fazendo a convenção a seguir, usada na mecânica quântica estatística

Considere $p(./.)\in F(\chi/D)$ como o módulo quadrado da função onda normalizada $\psi(./.)$.

Então,

$$(4.3)\qquad p(t/z)=|\psi(t/z)|^2,\ (t,z)\in\chi\times D$$

onde $\psi(./z)$ é uma função de valores complexos em S_μ, a esfera unitária de $L_2(\chi;\mu)$.

Assim,

$$(4.4)\qquad \|\psi(./z)\|_{\mu}^{2}=1,\ z\in D.$$

Usando esta nova notação, a pseudo-distância de Hellinger é da por

$$(4.5)\qquad \rho(z,w)=\left\{1-\int_{\chi}|\psi(t/z)\psi(t/w)|d\mu(t)\right\}^{1/2}.$$

Seja $\lambda(z,w)$ a pseudo-distância alternativa, definida por

$$(4.6)\qquad \lambda(z,w)=\frac{1}{\sqrt{2}}\min_{\theta_1,\theta_2\in[0,2\pi]}\left\|e^{i\theta_1}\psi(./z)-e^{i\theta_2}\psi(./w)\right\|_{\mu}$$

onde θ_1,θ_2 percorre o intervalo $[0,2\pi]$.

Desde que $|e^{i\theta_1}|=1$, fazendo $\theta=\theta_2-\theta_1$, temos

$$(4.7)\qquad \lambda(z,w)=\frac{1}{\sqrt{2}}\min_{0\le\theta\le2\pi}\left\|\psi(./z)-e^{i\theta}\psi(./w)\right\|_{\mu}$$

Conseqüentemente, λ indica uma pseudo-distância sobre D.

PROPOSIÇÃO 4.2

i) A pseudo-distância $\lambda(z,w)$, pode ser escrita como

$$(4.8) \qquad \lambda(z,w)=\left\{1-\left|\int_\chi \psi(t/z)\psi(t/w)d\mu(t)\right|\right\}^{1/2}$$

a qual comparando com a expressão (4.3), conclui-se que

$$(4.9) \qquad \rho(z,w)\leq\lambda(z,w),\ z,w\in D$$

ii) Sabendo que

$$(4.10) \qquad \lambda^2(z,w)=1-\left|(\psi(./z),\psi(./w))_\mu\right|$$

tem-se

$$\lambda^2(z,z)=d\lambda^2(z,w)\big|_{w=z}=0,$$

e consequentemente

$$(4.11) \qquad d^2\lambda^2(z,z)=\|d\psi\|_\mu^2-\left|(\psi,d\psi)_\mu\right|^2,\ \psi=\psi(./z).$$

PROVA

i) Tem-se

$$\lambda(z,w)=\frac{1}{\sqrt{2}}\min_{\theta_1,\theta_2\in[0,2\pi]}\left\|e^{i\theta_1}\psi(./z)-e^{i\theta_2}\psi(./w)\right\|_\mu=$$

$$=\frac{1}{\sqrt{2}}\min_{\theta_1,\theta_2\in[0,2\pi]}\left\|e^{i\theta_1}\right\|_\mu\left\|\psi(./z)-e^{i\theta}\psi(./w)\right\|_\mu$$

onde $\theta=\theta_2-\theta_1$.

Logo,

$$\lambda(z,w)=\frac{1}{\sqrt{2}}\min_{\theta\in[0,2\pi]}\left\{\int_\chi(\psi(t/z)-e^{i\theta}\psi(t/w))\overline{(\psi(t/z)-e^{i\theta}\psi(t/w))}d\mu(t)\right\}^{1/2}$$

$$\lambda(z,w)=\frac{1}{\sqrt{2}}\min_{\theta\in[0,2\pi]}\left\{\int_\chi\left[|\psi(t/z)|^2-e^{i\theta}\psi(t/w)\overline{\psi(t/z)}-e^{i\theta}\overline{\psi(t/w)}\psi(t/z)+|\psi(t/w)|^2\right]d\mu(t)\right\}$$

$$\lambda(z,w)=\frac{1}{\sqrt{2}}\min_{\theta\in[0,2\pi]}\left\{2-2\,\mathrm{Re}\int_{\chi}e^{i\theta}\psi(t/z)\overline{\psi(t/w)}d\mu(t)\right\}^{1/2}=$$

$$=\min_{\theta\in[0,2\pi]}\left\{1-\mathrm{Re}\int_{\chi}\psi(t/z)\overline{\psi(t/w)}d\mu(t)\right\}^{1/2}.$$

Por outro lado,

$$e^{i\theta}\int_{\chi}\psi(t/z)\overline{\psi(t/w)}d\mu(t)=\rho\cos\alpha+i\rho\,\mathrm{sen}\,\alpha$$

onde α é o argumento de $e^{i\theta}\int_{\chi}\psi(t/z)\overline{\psi(t/w)}d\mu(t)$ e

$$\rho=\left|e^{i\theta}\int_{\chi}\psi(t/z)\overline{\psi(t/w)}d\mu(t)\right|=\left|e^{i\theta}\right|\left|\int_{\chi}\psi(t/z)\overline{\psi(t/w)}\right|=\left|\int_{\chi}\psi(t/z)\overline{\psi(t/w)}d\mu(t)\right|.$$

Daí,

$$\lambda^2(z,w)=\min_{\theta\in[0,2\pi]}\left\{1-\mathrm{Re}\int_{\chi}\psi(t/z)\overline{\psi(t/w)}d\mu(t)\right\}=1-\max_{\theta\in[0,2\pi]}\mathrm{Re}\left[\int_{\chi}\psi(t/z)\overline{\psi(t/w)}d\mu(t)\right]$$

cujo máximo será atingido quando $\alpha=0$.

Logo,

$$\lambda^2(z,w)=\left\{1-\left|\int_{\chi}\psi(t/z)\overline{\psi(t/w)}d\mu(t)\right|\right\}.$$

isto é,

$$\lambda(z,w)=\left\{1-\left|\int_{\chi}\psi(t/z)\overline{\psi(t/w)}d\mu(t)\right|\right\}^{1/2}$$

Comparando a expressão (4.8) com a (4.3) tem-se

$$\rho^2(z,w)-\lambda^2(z,w)=\left\{1-\int_{\chi}\left|\psi(t/z)\overline{\psi(t/w)}\right|d\mu(t)\right\}-\left\{1-\left|\int_{\chi}\psi(t/z)\overline{\psi(t/w)}d\mu(t)\right|\right\}=$$

$$=-\int_{\chi}\left|\psi(t/z)\overline{\psi(t/w)}\right|d\mu(t)+\left|\int_{\chi}\psi(t/z)\overline{\psi(t/w)}d\mu(t)\right|.$$

Ora, $\left|\int_{\chi}f d\mu(t)\right|\le\int_{\chi}|f|d\mu(t)$, daí tem-se $\rho^2(z,w)-\lambda^2(z,w)\le 0$. Portanto,

$$\rho^2(z,w)\le\lambda^2(z,w)\quad\text{isto é}\quad\rho(z,w)\le\lambda(z,w)$$

ii) Vamos mostrar que $d^2\lambda^2(z,w)\big|_{w=z} = \|d\psi\|_\mu^2 - \left|(\psi, d\psi)_\mu\right|^2$, visto que a verificação da igualdade $\lambda^2(z,z) = d\lambda^2(z,w)\big|_{w=z} = 0$ é trivial.

Com efeito,

$$\lambda^2(z,w) = 1 - \left\{(\psi(./z)\psi(./w))_\mu . \overline{(\psi(./z)\psi(./w))_\mu}\right\}^{1/2} =$$

$$= 1 - \left\{\int_\chi \psi(t/z)\overline{\psi(t/w)}d\mu(t) . \int_\chi \overline{\psi(t/z)}\psi(t/w)d\mu(t)\right\}^{1/2}.$$

Logo,

$$d\lambda^2(z,w) = -\frac{1}{2}\left\{\int_\chi \psi(t/z)\overline{\psi(t/w)}d\mu(t) . \int_\chi \overline{\psi(t/z)}\psi(t/w)d\mu(t)\right\}^{1/2}.$$

$$.\left\{\int_\chi \psi(t/z)\frac{\partial}{\partial w}\overline{\psi(t/w)}d\mu(t) . \int_\chi \overline{\psi(t/z)}\psi(t/w)d\mu(t) + \int_\chi \psi(t/z)\overline{\psi(t/w)}d\mu(t) . \int_\chi \overline{\psi(t/z)}\frac{\partial}{\partial w}\psi(t/w)d\mu($$

Portanto,

$$d^2\lambda^2(z,w)\big|_{w=z} = -\frac{1}{2}\left(-\frac{1}{2}\right)\left\{\int_\chi \psi(t/z)\overline{\psi(t/w)}d\mu(t) . \int_\chi \overline{\psi(t/z)}\psi(t/w)d\mu(t)\right\}^{-3/2}.$$

$$.\left\{\int_\chi \psi(t/z)\frac{\partial}{\partial w}\overline{\psi(t/w)}d\mu(t) . \int_\chi \overline{\psi(t/z)}\psi(t/w)d\mu(t) + \int_\chi \overline{\psi(t/z)}\frac{\partial}{\partial w}\psi(t/w)d\mu(t)\right\}.$$

$$\left\{\int_\chi \psi(t/z)\frac{\partial}{\partial w}\overline{\psi(t/w)}d\mu(t) . \int_\chi \overline{\psi(t/z)}\psi(t/w)d\mu(t) + \int_\chi \overline{\psi(t/z)}\frac{\partial}{\partial w}\psi(t/w)d\mu(t) . \int_\chi \psi(t/z)\overline{\psi(t/w)}d\mu(t)\right\}$$

$$+\frac{1}{2}\left\{\int_\chi \psi(t/z)\overline{\psi(t/w)}d\mu(t) . \int_\chi \overline{\psi(t/z)}\psi(t/w)d\mu(t)\right\}^{1/2} . \left\{\int_\chi \psi(t/z)\frac{\partial}{\partial w}\left(\frac{\partial}{\partial w}\overline{\psi(t/w)}\right)d\mu(t) . \int_\chi \overline{\psi(t/z)}\psi(t/w)d\mu(t) +\right.$$

$$\left. + \int_\chi \psi(t/z)\overline{\psi(t/w)}d\mu(t) . \int_\chi \overline{\psi(t/z)}\frac{\partial}{\partial w}\left(\frac{\partial}{\partial w}\psi(t/w)\right)d\mu(t)\right\}.$$

Quando w=z, temos

$$d^2\lambda^2(z,w)\big|_{w=z} = -\frac{1}{2}\left\{\int_\chi \psi(t/z)\frac{\partial^2}{\partial w^2}\overline{\psi(t/w)}d\mu(t) + \int_\chi \psi(t/z)\frac{\partial}{\partial w}\overline{\psi(t/w)}d\mu(t).\right.$$

$$.\int_\chi \overline{\psi(t/z)}\frac{\partial}{\partial w}\psi(t/w)d\mu(t) + \int_\chi \psi(t/z)\frac{\partial}{\partial w}\overline{\psi(t/w)}d\mu(t) . \int \overline{\psi(t/z)}\frac{\partial}{\partial w}\psi(t/w)d\mu(t) +$$

$$+\int_\chi \overline{\psi(t/z)}\frac{\partial^2}{\partial^2 w}\psi(t/w)d\mu(t)\Big\} = -\frac{1}{2}\Big\{2\left|(\psi,d\psi)_\mu\right|^2 + \int_\chi \overline{\psi(t/z)}\frac{\partial^2}{\partial^2 w}\psi(t/w)d\mu(t) +$$

$$+\int_\chi \psi(t/z)\frac{\partial^2}{\partial^2 w}\overline{\psi(t/z)}d\mu(t) + \int_\chi \frac{\partial}{\partial w}\overline{\psi(t/z)}.\frac{\partial}{\partial w}\psi(t/z)d\mu(t) = -\frac{1}{2}\Big\{\partial\Big(\int_\chi \overline{\psi}\partial\psi d\mu(t) +$$

$$\int_\chi \psi\partial\overline{\psi}d\mu(t)\Big) - 2\int_\chi \partial\overline{\psi}\partial\psi d\mu(t) + 2\left|(\psi,d\psi)_\mu\right|^2\Big\}$$

Mas

$$\int_\chi \overline{\psi}\partial\psi d\mu(t) = \int_\chi \psi\partial\overline{\psi}d\mu(t) = 0$$

Portanto,

$$d^2\lambda^2(t) = \|d\psi\|_\mu^2 - \left|(\psi,d\psi)_\mu\right|^2$$

a qual é não negativa. •

CAPÍTULO V

INTRODUÇÃO

Iniciamos este capítulo definindo núcleo reproduzido e mostrando algumas de suas propriedades básicas.

Posteriormente, definimos núcleo sesqui-holomorfo e o núcleo de Bergman, para finalmente enunciar e demonstrar o teorema central do trabalho.

5.1 TEORIA DOS NÚCLEOS REPRODUZIDOS

5.1.1 - Definição de núcleos reproduzidos.

Considere um espaço vetorial F de funções f(x) definidas em um conjunto E. Suponhamos que F seja um espaço complexo, isto é, admite a multiplicação por um número complexo.

Suponhamos também que para $f \in F$, seja definida uma norma, $\|f\|$, isto é, um número real satisfazendo as seguintes propriedades:

i) $\|f\| \geq 0$, $\|f\| = 0$ se somente se $f = 0$;

ii) $\|cf\| = |c|\|f\|$;

iii) $\|f + g\| \leq \|f\| + \|g\|$, $\forall f, g \in F$;

Suponhamos ainda que a norma acima seja dada pela forma quadrática, Q(f);

$$\|f\|^2 = Q(f).$$

O funcional Q(f) é denominado forma quadrática hermitiana se $\forall \varepsilon_1, \varepsilon_2$ constantes, e funções f_1, f_2 de F tem-se

$$Q(\varepsilon_1 f_1 + \varepsilon_2 f_2) = |\varepsilon_1|^2 Q(f_1) + \varepsilon_1 \bar{\varepsilon}_2 Q(f_1, f_2) + \bar{\varepsilon}_1 \varepsilon_2 Q(f_2, f_1) + |\varepsilon_2|^2 Q(f_2)$$

onde $Q(f_1, f_2) = \overline{Q(f_1, f_2)}$ é a forma bilinear hermitiana unicamente determinada, correspondendo a forma quadrática Q(f). Esta forma quadrática será denotada por $(f_1, f_2) \equiv Q(f_1, f_2)$ e chamada o produto interno, correspondente a norma $\|f\|^2$, onde $\|f\|^2 = (f, f)$.

O espaço F com a norma $\| \ \|$, forma um espaço vetorial normado. Se este espaço for completo, ele será um espaço de Hilbert.

Se F for um espaço de funções com valores reais, formando um espaço vetorial real, temos que, a norma, $\| \ \|$, em F, é dada por

$$\|f\|^2 = Q(f),$$

com a forma quadrática canônica, isto é,

$$\forall \varepsilon_1, \varepsilon_2 \in R,\ Q(\varepsilon_1 f_1 + \varepsilon_2 f_2) = |\varepsilon_1|^2 Q(f_1) + 2\varepsilon_1\varepsilon_2 Q(f_1, f_2) + |\varepsilon_2|^2 Q(f_2),$$

onde $Q(f_1, f_2)$ corresponde a forma quadrática bilinear simétrica, e se F for completo, ele será um espaço de Hilbert real e o produto interno, será dado por $(f_1, f_2) = Q(f_1, f_2)$.

Todo espaço de Hilbert real F determina um espaço de Hilbert complexo da seguinte forma

Considere todas as funções $f_1 + if_2$, $f_1, f_2 \in F$. Estas funções formam um espaço vetrorial complexo F_c sobre o qual definimos a norma como

$$\|f_1 + if_2\| = \|f_1\|^2 + \|f_2\|^2.$$

Assim, F_c é um espaço de Hilbert complexo.

Os espaços de Hilbert complexos F determinados pelos espaços de Hilbert reais são caracterizados por duas propriedades

i) Se $f \in F$, ent $\bar{f} \in F$ ($\bar{f}$ é a função complexa conjugada de f);

ii) $\|f\| = \|\bar{f}\|$, $\forall f \in F$.

DEFINIÇÃO 5.1

Seja F um espaço de funções definidas em E, formando um espaço de Hilbert (real ou complexo). A função K(x,y), definida sobre E×E, é denominada um núcleo reproduzido de F se

i) Para cada y, K(x,y) como função de x pertence a F;

ii) *A propriedade de reprodução*: para $y \in E$ e $f \in F$,

$$f(y) = (f(x), K(x,y))_x$$

onde o subscrito x no produto interno, indica que o produto é aplicado na função x.

Se uma classe real F possui um núcleo reproduzido K(x,y), então verifica-se imediatamente que o espaço complexo F_c possui o mesmo núcleo que tem valores reais. Por isto, consideraremos apenas espaços de Hilbert complexos.

Introduziremos uma distinção entre o termo subclasse e subespaço. Quando F_1 e F_2 são duas classes de funções no mesmo conjunto E, F_1 é subclasse de F_2 se para cada $f \in F_1$, f pertence também a F_2, e $\|f\|_1 = \|f\|_2$ ($\| \ \|_1$ e $\| \ \|_2$ são as normas de F_1 e F_2 respectivamente). A notação $F_1 \subset F_2$ significa que F_1 é um subclasse de F_2.

5.1.2-PROPRIEDADES BÁSICAS DOS NÚCLEOS REPRODUZIDOS

Seja F um espaço de funções f(x) definidas em E, formando um espaço de Hilbert com a norma $\|f\|$ e o produto interno (f_1, f_2). Denotaremos K(x,y) o núcleo reproduzido. Então

i) Se existe um núcleo reproduzido K(x,y), ele é único.

De fato, se K*(x,y) é um outro núcleo reproduzido de F, então para o mesmo y tem-se

$$0 < \|K(x,y) - K^*(x,y)\|^2 = (K - K^*, K - K^*) = (K - K^*, K) - (K - K^*, K^*) = 0,$$

pela propriedade de reprodução de K e K*.

ii) *Existência.* Para a existência de um núcleo reproduzido K(x,y), a condição necessária e suficiente é que para $y \in E$, f(y) seja um funcional contínuo de f para o espaço de Hilbert F.

Com efeito, se K existe, então

$$|f(y)| \leq \|f\|(K(x,y), K(x,y))^{1/2} = K(y,y)^{1/2}\|f\|.$$

Por outro lado, se f(y) é um funcional continuo, então pelo teorema geral dos espaços de Hilbert, existe uma função $g_y(x) \in F$, tal que

$$g_y(x) = (f(y), g_y(x)).$$

Então, se tomarmos $K(x,y) = g_y(x)$ como o núcleo reproduzido, o resultado segue-se.

iii) K(x,y) é uma matriz positiva, isto é, a forma quadrática em $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$

$$(1) \qquad \sum_{i,j=1}^{n} K(x,y)\bar{\varepsilon}_i . \varepsilon_j$$

é não negativa para todo $y_1, y_2, \ldots, y_n$ em E. É claro que a expressão (1) é igual a

$$\left\| \sum_{i=1}^{n} K(x,y)\varepsilon_i \right\|^2$$

pela propriedade de reprodução. Em particular segue-se que

$$K(x,y) \geq 0; \quad K(x,y) = \overline{K(x,y)}$$

$$|K(x,y)|^2 \leq K(x,x)K(y,y)$$

iv) Para a matriz K(x,y) existe um único espaço de funções que determina uma única forma quadrática, formando um espaço de Hilbert e admitindo K(x,y) como o núcleo reproduzido.

Este espaço de funções é gerado por todas as funções da forma $\sum \alpha_k K(x,y_k)$. A norma destas funções é definida pela forma quadrática

$$\left\|\sum \alpha_k K(x,y_k)\right\|^2 = \sum\sum K(y_i,y_j)\bar{\varepsilon}_i\varepsilon_j.$$

Funções com esta norma não formam um espaço de Hilbert completo, porém podem ser completado adicionando as funções que são limites de sequências de Cauchy. Formando assim um espaço de Hilbert completo

v) Se o espaço F possui um núcleo reproduzido K(x,y), cada sequência de funções $\{f_n\}$ que converge fortemente para uma função f no espaço de Hilbert F, converge também ponto a ponto no sentido canônico, ou seja

$$\lim f_n(x) = f(x).$$

Esta convergência é uniforme em um subconjunto de E, em que H(x,x) é uniformemente limitada.

Isto segue do fato,

$$|f(y) - f_n(y)| = |(f(x) - f_n(x), K(x,y))| \leq \|f - f_n\|\|K(x,y)\| = \|f - f_n\|(K(y,y))^{1/2}$$

Se f_n converge fracamente para f, temos novamente $f_n(y) \to f(y)$ para cada y, visto que por definição, a convergência fraca é dada por

$$(f_n(x), K(x,y)) \to (f(x), K(x,y)).$$

Há em geral uma sequência decrescente de conjuntos $E_1 \subset E_2 \subset \ldots \to E$, em que cada f_n converge uniformemente para f.

Se uma topologia (uma noção de limite) é definida em E e se a correspondência

$$y \leftrightarrow K(x,y)$$

transforma E continuamente em um subconjunto do espaço F, então a convergência fraca da sequência $\{f_n\}$ implica na convergência uniforme em cada conjunto compacto $E_1 \subset E$.

De fato, E_1 é transformado em um subconjunto compacto de espaço F. Para cada $\varepsilon > 0$, podemos escolher um conjunto finito

$$(y_1, y_2, \ldots, y_n) \subset E_1$$

tal que, para cada $y \in E_1$, existe pelo menos um y_k com

$$\|K(x,y) - K(x,y_k)\| \leq \varepsilon / 4M$$

onde $M = \sup_n \|f\|_n$.

Mas, se escolhermos n_0 tal que $n > n_0$,

$$|f(y_n) - f_n(y_k)| < \varepsilon / 4.$$

Obtemos $y \in E_1$,

$$|f(y) - f_n(y)| = |f(y) - f(y_k) + f(y_k) + f(y_k) - f_n(y_k) + f_n(y_k) - f_n(y)| \leq$$
$$\leq |f(y_k) - f_n(y_k)| + |(f(x) - f_n(x), K(x,y) - K(x,y_k))| \leq$$
$$\leq \varepsilon / 4 + \|f - f_n\| \|K(x,y) - K(x,y_k)\| \leq \varepsilon / 4 + 2M\varepsilon / 4M$$

A continuidade da correspondência $y \leftrightarrow K(x,y)$ é equivalente a equicontinuidade de todas as funções de F com $\|f\| \leq M$, $\forall M > 0$. Esta propriedade é satisfeita para vários espaços que têm núcleos reproduzidos, que são usualmente considerados, tais como o espaço das funções analíticas, harmônicas, soluções de equações diferenciais parciais etc.

vi) Se F é um espaço com o núcleo reproduzido K, então F é um subespaço de um espaço de Hilbert, **H**, neste caso a igualdade

$$f(y) = K(h, K(x,y))_x$$

verifica-se a partir da projeção de elementos $h \in \mathbf{H}$ sobre F.

Com efeito, considere h=f+g, onde g é ortogonal ao espaço F. A função K(x,y), considerada como uma função de x, pertence a F. Então

$$(h, K(x,y))_x = (f + g, K(x,y))_x = (f, K(x,y))_x = f(y),$$

pela propriedade de reprodução.

5.2 - NÚCLEOS SESQUI-HOLOMORFOS

Consideremos uma função onda

$$g(./z) \in L_2(\chi;\mu),\ z \in D$$

não necessariamente normalizada como em (4.4).

Em alguns casos, estas funções envolvem certas constantes de normalização, detalhadas abaixo

$$(5.1) \qquad g(./z) = \sqrt{K(z,\bar{z})}\,\psi(./z)$$

com $\psi(./z) \in S_\mu,\ z \in D$ e a constante de normalização dada por

$$(5.2) \qquad K(z,\bar{z}) = (g(./z), g(./z))_\mu$$

pois $K(z,\bar{z})$ é uma função de z e de $\bar{z}$.

Isto se torna mais aparente quando requeremos que g(t/.) seja holomorfa em $D, \forall t \in \chi$, a menos de um conjunto de medida nula.

PROPOSIÇÃO 5.1

A forma (4.5) com a notação (5.1) e (5.2) torna-se

$$(5.3) \qquad d^2\lambda^2(z,w)\Big|_{w=z} = K^{-2}\left[K\|dg\|_\mu^2 - \left|(g,dg)_\mu\right|^2\right]$$

onde $K = K(z,\bar{z})$ e $g = g(./z)$.

PROVA

Como $\psi(./z) = \left[K(z,\bar{z})\right]^{-1/2} g(./z)$, tem-se

$$d^2\lambda^2(z,w)\Big|_{w=z} = \left\|d\left(\left[K(z,\bar{z})\right]^{-1/2} g(./z)\right)\right\|_\mu^2 - \left|\left(\left[K(z,\bar{z})\right]^{-1/2} g(./z), d\left(\left[K(z,\bar{z})\right]^{-1/2} g(./z)\right)\right)_\mu\right|^2 =$$

$$= \left\|\frac{1}{K(z,\bar{z})}dg\right\|_\mu^2 - \left(\frac{g(./z)}{\sqrt{K(z,\bar{z})}}, \frac{dg(./z)}{\sqrt{K(z,\bar{z})}}\right)_\mu \left(\overline{\frac{g(./z)}{\sqrt{K(z,\bar{z})}}, \frac{dg(./z)}{\sqrt{K(z,\bar{z})}}}\right)_\mu$$

$$= K^{-2}(z,\bar{z})\left[K\|dg\|_{\mu}^{2} - \left|(g,dg)_{\mu}\right|^{2}\right]$$

pois $dK(z,\bar{z}) = d(g(./z),g(./z))_{\mu} = 0$

Um caso de grande importância ocorre quando a não normalizada função onda $g(./z)$ é holomorfa em D, μ-quase todo $t \in \chi$.

Neste caso, por (5.2) e pelo teorema de Hartogs, $K(z,\bar{z})$ é holomorfa em $(z,\bar{z})$, $z \in D$. Portanto, por polarização

$$K(z,\bar{w}) = (g(./z),g(./w))_{\mu}, \; z,w \in D. \tag{5.4}$$

Então, $\forall w \in D$ fixo, $K(.,\bar{w})$ é holomorfa em D.

Como o núcleo é hermitiano, isto é;

$$K(z,\bar{w}) = \overline{K(w,\bar{z})}, \; \forall z,w \in D$$

$K(w,.)$ é anti-holomorfo em D.

E novamente pelo teorema de Hartogs, $K(z,\bar{w})$ é holomorfo em $(z,\bar{w})$, $(z,w) \in D$.

DEFINIÇÃO 5.1

Um núcleo satisfazendo as propriedades anteriores é chamado núcleo sesqui-holomorfo sobre $D \times D$

DEFINIÇÃO 5.2

Chama-se núcleo de Bergman sobre $D \times D$, qualquer núcleo sesqui-holomorfo $K(z,\bar{z})$, que é definido positivo, isto é, qualquer sistema fixo de pontos $z_1, z_2, \ldots, z_N \in D$ e quaisquer escalares $\alpha_1, \alpha_2, \ldots, \alpha_N \in C$, tal que,

$$\sum_{k,m=1}^{N} K(z_k,\bar{z}_m)\alpha_k\bar{\alpha}_m \geq 0$$

PROPOSIÇÃO 5.2

$K(z,\bar{z})$ é um núcleo de Bergman sobre $D \times D$.

PROVA

De fato,

$$\sum_{k,m=1}^{N} K(z_k,\bar{z}_m)\alpha_k\bar{\alpha}_m = \left\|\sum_{k=1}^{N} \alpha_k g(./z_k)\right\|_{\mu}^{2} \geq 0 \tag{5.5}$$

e como $K(z,\overline{w})$ é sesqui-holomorfo, o resultado segue-se. •

Pela teoria clássica de núcleo reproduzidos[1] e [12] existe um único espaço de Hilbert, H(D), de funções que são holomorfas em D, tal que

$$k_w(z) = K(z,\overline{w}), \quad z,w \in D$$

seja seu núcleo reproduzido.

Seja $(./.)$ o produto interno de H(D). Então $\forall w \in D$, temos que $k_w \in H(D)$ e

$$f(w) = (f,k_w), \quad f \in H(D).$$

Em particular,

$$K(z,\overline{w}) = k_w(z) = (k_w,k_z), \quad z,w \in D.$$

PROPOSIÇÃO 5.3

Assumindo que nos seja dado um núcleo reproduzido do espaço H(D) de funções holomorfas em D com o produto interno $(.,.)_\mu$, seja $k_w = K(z,\overline{w})$, $w \in D$ o núcleo reproduzido de H(D). Então, $K(z,\overline{w})$ é um núcleo de Bergman.

PROVA

De fato, considere qualquer isometria de H(D) em $L_2(\chi;\mu)$

(5.6) $\qquad g(./z) = \overline{Tk_w}, \quad z \in D$

Então, para $z,w \in D$, tem-se

$$K(z,\overline{w}) = (k_w,k_z)_\mu = (Tk_w,Tk_z)_\mu = \left(\overline{g(./w)}.\overline{g(./z)}\right)_\mu,$$

isto é

$$K(z,\overline{w}) = (g(./z),g(./w))_\mu = \int_\chi g(t/z)\overline{g(t/w)}d\mu(t)$$

que é realmente a equação (5.4).

Logo, existe uma correspondência biunívoca canônica entre funções ondas holomorfas $g(./z) \in L_2(\chi;\mu)$ e o núcleo de Bergman k_z, $z \in D$ de espaços de Hilbert H(D), de funções holomorfas em D. Esta correspondência de funções holomorfas é determinada por (5.6) e por esta razão $g(./.)$ definida sobre $\chi \times D$ é chamada função geradora do $L_2(\chi;\mu)$ para $H(D)$, [5].

PROPOSIÇÃO 5.4

Sob as condições anteriores,

$$d^2\lambda^2(z,w)\big|_{w=z} = K^{-2}\left[K\|dg\|_\mu^2 - \left|(g,dg)_\mu\right|^2\right]$$

é equivalente

$$d^2\lambda^2(z,w)\big|_{w=z} = K^{-2}\left[K\partial\bar{\partial}K - |\partial K|^2\right]$$

isto é

$$d^2\lambda^2(z,w)\big|_{w=z} = \partial\bar{\partial}\log K$$

PROVA

Observe que

$$\partial\bar{\partial}K = \partial\bar{\partial}\left(g(./z),g(./z)\right)_\mu = \partial\left[\left(\bar{\partial}g(./z),g(./z)\right)_\mu + \left(g(./z),\bar{\partial}g(./z)\right)_\mu\right] = \partial\left(g(./z),\bar{\partial}g(./z)\right)$$

$$\left(\bar{\partial}g(./z),g(./z)\right)_\mu = \int_\chi \bar{\partial}g(./z)\overline{g(./z)}d\mu(t) = 0, \text{pois} \quad \bar{\partial}g(./z) = 0,$$

e

$$|\partial K|^2 = \partial K.\bar{\partial}K = \partial\left(g(./z),g(./z)\right)_\mu.\bar{\partial}\left(\overline{g(./z)},\overline{g(./z)}\right)_\mu = \left(g(./z),\partial g(./z)\right)_\mu.\left(\overline{g(./z)},\bar{\partial}g(./z)\right)_\mu = \left|\left(g(./z),dg(./z)\right)_\mu\right|^2$$

Consequentemente,

$$K^{-2}\left[K\partial\bar{\partial}K - |\partial K|^2\right] = K^{-2}\left[\|dg\|_\mu^2 - \left|(g,dg)_\mu\right|^2\right],$$

e

$$\partial\bar{\partial}\log K = \partial\left\{\frac{1}{K}\bar{\partial}K\right\} = -\frac{1}{K^2}\partial K\bar{\partial}K + \frac{1}{K}\partial\bar{\partial}K = K^{-2}\left[K\partial\bar{\partial}K - |\partial K|^2\right] = d^2\lambda^2(z,w)\big|_{w=z} \quad \bullet$$

Como a métrica de Bergman é definida por

$$(5.7) \qquad db^2(z) = \partial\bar{\partial}\log K = \sum_{k,m=1}^{N} T_{k,m}dz_k d\bar{z}_m$$

onde $T_{k,m} = \partial_{z_k}\bar{\partial}_{z_m}\log K(z,\bar{z})$, $z \in D$

temos que

$$d^2\lambda^2(z,w)\big|_{w=z} = db^2(z)$$

Um caso especial desta métrica foi inicialmente estudada por Bergman e [4], [6] e [8].

Uma definição geométrica desta métrica é dada como se segue:

Sejam

$$S(D) = \{f \in H(D); \|f\|_\mu < 1\}$$

e

$$S_w(D) = \{f \in S(D); f(w) = 0\}$$

.

para $w \in D$ fixo.

Dada uma direção $v \in C^n$, definimos

$$b(w,v) = \sqrt{K(w,\overline{w})}\max\{|\partial_v f(w)|; f \in S_w(D)\} \tag{5.8}$$

onde

$$\partial_v f(z) = \sum_{k=1}^{n} v_k \partial_{z_k} f(z), \quad v = (v_1, v_2, \ldots, v_n) \in C^n$$

Como em espaços de Hilbert canônicos, o problema (5.8) tem uma solução única, a menos de uma rotação, isto é

$$4b^2(w{:}\,v) = 4\partial_v\overline{\partial}_v \log K = \Delta_v \log K. \tag{5.9}$$

Então, quando $v = dz = (dz_1, dz_2, \ldots, dz_n)$

$$b^2(z; dz) = db^2(z),$$

e obtemos dessa maneira a métrica de Bergman, que é métrica Kähler, isto é, a dois forma

$$\sum_{k,\overline{m}=1}^{n} T_{k,\overline{m}} dz_k \wedge d\overline{z}_m$$

é fechada.

Assim, a (μ,ψ)-pseudo-distância projetiva (4.4) assume uma expressão mais simples, indicada como (4.4), (5.1), (5.2) e (5.4);

$$\lambda(z,w) = \left\{1 - \left[\left|K(z,\overline{w})\right|^2 K(z,\overline{z})^{-1} K(w,\overline{w})^{-1}\right]^{1/2}\right\}^{1/2} \quad z, w \in D \tag{5.10}$$

Portanto, quando $K(z,\overline{w})$ é o clássico núcleo de Bergman, isto é, o núcleo reproduzido de $H(D) \equiv H_2(D)$, onde $H_2(D)$ representa o espaço das funções holomorfas no domínio limitado $D \subset C^n$ que são quadrado integrável em relação a medida de Lebesgue μ, sobre D, a pseudo-distância $\lambda(z,w)$, (5.10), torna-se uma distância (no sentido canônico de distância).

Esta distância foi inicialmente estudada por Skwarzynski **[19]**. Por isto, (5.10) é também chamada a pseudo-distância de Skwarzynski de D **[9]**. Uma condição de suficiência para λ ser uma distância é dada pelo seguinte teorema.

TEOREMA 5.1

Suponha que $1, f_1, f_2, \ldots, f_n \in H(D)$, onde $f_j(z) = z_j,\ 1 \leq j \leq n$ e $z = (z_1, z_2, \ldots, z_n) \in D$. Então λ é uma distância sobre D.

PROVA

Assuma que $\lambda(z,w) = 0$, para $z, w \in D$. Portanto, por (5.10)

$$\left|K(z,\overline{w})\right|^2 = K(z,\overline{z})K(w,\overline{w}),$$

que significa que $k_w = \overline{\alpha}k_z$, para algum $\alpha \in C$. Daí segue-se que $f(w) = \overline{\alpha}f(z),\ \forall f \in H(D)$. Tomando f=1 e $f = f_j,\ j = 1,2,\ldots,n$, obtemos w=z e portanto a prova esta completa. •

Na presente colocação, uma relação de grande importância é o fato que a métrica de informação ds^2 nada mais é senão a métrica de Bergman db^2, indicada anteriormente por (4.1), (4.2) e (5.1), logo

$$\log p(t/z) = \log g(t/z) + \log \overline{g(t/z)} - \log K(z,\overline{z}).$$

Em virtude das equações de Cauchy-Riemann, deduzimos

$$\partial z_k \overline{\partial}_{z_m} \log p(t/z) = \partial_{z_k} \overline{\partial}_{z_m} \log K(z,\overline{z})$$

que é valida $\forall t \in \chi$ μ-quase por toda parte $\forall z \in D$.

Por (2.1) e (5.7), $g_{k\overline{m}}(z) = T_{k\overline{m}}(z),\ z \in D$, portanto $ds^2(z)db^2(z)$.

Resumindo estes resultados, temos o seguinte teorema;

TEOREMA 5.2

Assuma que a função densidade de probabilidade $p(./.) \in F(\chi/D)$ é dada por

TEOREMA 5.2

Assuma que a função densidade de probabilidade $p(./.) \in F(\chi / D)$ é dada por

$$p(t/z) = \frac{|g(t/z)|^2}{K(z,\bar{z})}, \quad k(z,\bar{z}) > 0, \quad z \in D$$

onde g(t/.) é holomorfa em D μ-quase todo $t \in \chi$. Então

i) $K(z,\bar{w})$ é o núcleo do Bergman de um espaço de Hilbert H(D), de funções holomorfa em D

ii) g(./.), definida sobre $\chi \times D$ é a função geradora do $L_2(\chi;\mu)$ para H(D)

iii) A pseudo-distância projetiva λ em (4.4) é a pseudo-distância de Skwarzynski em (5.10)

iv) $\forall z, w \in D$, $0 \le \rho(z,w) \le \lambda(z,w) \le 1$, onde ρ é a pseudo-distância de Hellinger em (4.3)

v) A métrica de informação ds^2 é a métrica de Bergman db^2, que pode ser expressa como (5.6)-(5.9)

vi) $\lambda^2(z,z) = d\lambda^2(z,w)\big|_{w=z} = 0$ e $d^2\lambda^2(z,w)\big|_{w=z} = ds^2(z) = db^2(z) \ge 0$.

APÊNDICE

A-1 A GEOMETRIA DIFERENCIAL E A ESTATÍSTICA

A-1.1 - Variedades de um modelo estatístico.

Usualmente define-se um modelo estatístico como uma família de distribuições de probabilidades.

Seja $S = \{p(t/z)\}$ um modelo estatístico, onde t é uma variável aleatória pertencente ao espaço amostral χ e p(t/z) a função densidade de probabilidade de t, parametrizada por z, em relação a alguma medida μ e

$$z = (z_1, z_2, \ldots, z_n)$$

um parâmetro n-dimensional pertencente a um subconjunto $D \subset R^n$ ou $D \subset C^n$.

DEFINIÇÃO A.1

Uma variedade n-dimensional S, é um espaço de Hausdorff que é localmente homeomorfa a um espaço euclidiano n-dimensional.

Seja $U \subset S$, um aberto, que é homeomorfo ao R^n (ou C^n) com o homeomorfismo ϕ. Então, $p \in U$ é transformado num ponto $z = (z_1, z_2, \ldots, z_n) \in R^n$ (ou C^n), isto é,

$$\phi(p) = z = (z_1, z_2, \ldots, z_n)$$

DEFINIÇÃO A.2

A transformação ϕ, acima citada, é chamada uma função coordenada na vizinhança coordenada de U.

No presente trabalho trataremos apenas de propriedades locais de variedades de modelos estatísticos. Portanto, assumiremos que a variedade S sempre tem um sistema de coordenadas globais cobrindo inteiramente S. Nossa teoria é válida apenas sobre algumas vizinhanças U de S.

Podemos portanto definir uma transformação

$$\phi : S \to R^n \quad (\text{ou } C^n)$$

dada por $\phi(p(t/z)) = z$.

Quando esta função tem o

papel de uma função coordenada, o vetor z é usado com o nome de coordenada ou de distribuição $p(t/z)$ e uma estrutura diferencial é introduzida em S para esta função coordenada. Logo, S é uma variedade diferenciável.

As propriedades seguintes são requeridas na teoria geométrica:

i) $\forall p(t/z)$ tem um suporte comum, que é $p(t/z) > 0, \quad \forall t \in \chi$, onde χ é o suporte;

ii) $l(t/z) = \log(p(t/z))$, e para cada z fixo, as n funções em t

$$\frac{\partial l(t/z)}{\partial z_i}, \quad i=1,2,\ldots,n$$

são linearmente independentes;

iii) O momento da variável aleatória $\frac{\partial l(t/z)}{\partial z_i}$, existe para todo i=1,2,...,n;

iv) A derivada parcial $\frac{\partial}{\partial z_i}$, e a integral com respeito a medida μ sempre pode ser mudada, ou seja

$$\frac{\partial}{\partial z_i}\int_\chi f(t/z)d\mu(t) = \int_\chi \frac{\partial f(t/z)}{\partial z_i}d\mu(t),$$

para toda $f(t/z)$ e χ o espaço amostral.

A.2. O ESPAÇO TANGENTE

Grosseiramente falando, o espaço tangente T_p, por p pertencente a uma variedade S, é um espaço vetorial, obtido pela linearização local de S em torno de p. É composto do vetor tangente a curva passando por p.

Seja $\theta: [a,b] \rightarrow S$ uma transformação continua, ou seja $\theta(t) \in S$ é a imagem de $t \in [a,b]$.

Se usarmos um sistema de coordenadas $\theta = \theta(p)$, a imagem $\theta(t)$ de t é definida por

$$\theta(t) = \{\theta_1(t), \theta_2(t), \ldots, \theta_n(t)\}.$$

E a equação $\theta = \theta(t)$ é a representação paramétrica da curva.

Matematicamente, definimos o espaço tangente da seguinte forma

Seja F o conjunto de todas as funções suaves sobre S. Usando um sistema de coordenadas θ, $f \in F$ é uma função suave $f(\theta_1, \theta_2, \ldots, \theta_n)$ em θ. Dada uma curva suave $\theta = \theta(t)$ e uma função $f \in F$, definimos uma função

$$f \circ \theta: [a,b] \rightarrow R$$

que é escrita como $f(\theta(t))$ em expressão coordenadas.

Seja Cf a derivada desta função,

$$Cf = \frac{d}{dt}(f \circ \theta) = \sum_{i=1}^{n} \frac{d\theta_i}{dt} \frac{\partial f}{\partial \theta_i}.$$

Obviamente, a derivada de f ao longo da curva θ ou na direção da tangente de θ. Então, o operador derivada direcional C é associado com cada curva, e intuitivamente C depende somente do vetor tangente $\frac{d\theta_i}{dt}$ da curva θ. Alem disso, para cada ponto $\theta(t_0)$, tem-se

i) C é uma transformação linear

ii) $C(f.g)=(Cf).g + f(Cg) \quad \forall f,g \in F$.

Logo, C é o operador derivada direcional de uma curva.

O conjunto dessas Transformações C é um espaço vetorial n-dimensional uma variedade C^{∞}.

Este espaço vetorial é chamado o espaço tangente T_p de S por p.

Sabemos que os n vetores $\frac{\partial}{\partial \theta_i} = \partial_i$ são linearmente independentes, logo eles formam uma base natural associada ao sistema de coordenadas θ. Assim, para qualquer vetor tangente $A \in T_p$, tem-se

$$A = \sum_{i=1}^{n} A_i \partial_i \ , \ \partial_i = \frac{\partial}{\partial \theta_i}$$

onde A_i, i=1,2,...,n, são as coordenadas de A com respeito a base natural $\{\partial_i\}$.

Considere $S = \{p(t/\theta)\}$, uma variedade de um modelo estatístico, tomando

$$l(t/\theta) = \log p(t/\theta)$$

e suas n derivadas parciais $\partial_i l(t/\theta)$, i = 1,2,...,n. Como são funções linearmente independentes em t, para cada θ, fixo, podemos construir o espaço n-dimensional

$$T_\theta^{(1)} = \left\{ A(t);\ A(t) = \sum_{i=1}^{n} \partial_i l(t/\theta) \right\}$$

que é a representação usual do espaço tangente. Isto é, $A(t) \in T_\theta^{(1)}$ pode ser escrito como combinação linear de $\partial_i l(t/\theta)$, com $A(t) = \sum_{i=1}^{n} A_i \partial_i l(t/\theta)$, onde A_i são as componentes de A(t) em relação a base $\partial_i l(t/\theta)$. Logo, t é uma variável aleatória e $T_\theta^{(1)}$ é um espaço vetorial caracterizado pela base $\{\partial_i l(t/\theta)\}$.

Como existe um isomorfismo natural entre os espaços vetoriais T_θ e $T_\theta^{(1)}$, caracterizado por

$$\partial_i \in T_\theta \leftrightarrow \partial_i l(t/\theta),$$

um vetor tangente $A(t)=\sum_{i=1}^{n} A_i\partial_i \in T_\theta$, corresponde a uma variável aleatória $A(t)=\sum_{i=1}^{n} A_i\partial_i l(t/\theta)$, tendo a mesma componente A_i. Desta forma, podemos identificar T_θ com $T_\theta^{(1)}$, lembrando que T_θ é o operador diferencial representando o espaço tangente, enquanto $T_\theta^{(1)}$ é a variável aleatória representando o mesmo espaço tangente.

DEFINIÇÃO A-3

O espaço $T_\theta^{(1)}$ é denominado a 1-representação do espaço tangente.

Seja E(.) a esperança com relação a distribuição $p(t/\theta)$, definida por

$$E(f(t))=\int_\chi f(t)p(t/\theta)d\mu(t).$$

Diferenciando a identidade $\int_\chi p(t/\theta)d\mu(t)=1$ com relação θ_i, temos

$$0=\partial_i\int_\chi p(t/\theta)d\mu(t)=\int_\chi p(t/\theta)d\mu(t)=$$
$$=\int_\chi p(t/\theta)\partial_i l(t/\theta)d\mu(t)=E[\partial_i l(t/\theta)].$$

Então, para qualquer variável aleatória $A(t)\in T_\theta^{(1)}$, temos

$$E[A(t)]=0.$$

BIBLIOGRAFIA

[1] Shun-ichi Amari, Differencial Geometrical Methods in Statistiocs, Lecture Notes in Statistics vol. 28, Springer-Verlag, Heidelberg,(1985).

[2] N. Aronszajn, Theory of kernels reproducing, Trans. Amer. Math. Soc. 68 (1950), p. 337-404.

[3] C. Atkinson and A. F. S. Mitchel, Rao's distance measure, Sankhyã 43 (1081), p.345-365.

[4] S. Bergman, The kernel function and conformal mapping, Math. Survey 5, Amer. Math. Soc.,Providence, R. I., 1970.

[5] J. Burbea, Total positivity of certain reproducing kernels, Pacific J. Math. 67 (1967), p.101-130.

[6] - The Carathéodory metric and its majorant metrics, Canad. J. Math. 29 (1977), p. 771-780

[7] - On the hessian of the Carathéodory metrics, Rocky Mountain J. Math. 8 (1978), p. 555-559.

[8] - A generalization of Pick`s theorem and its applications to intrinsics metrics, Ann. Polon. Math. 39 (1981), p. 49-61.

[9] - On metrics and distortion theorems, recent developments in several complex variables, Ann. Math. Studies 100 (1981), p. 65-92.

[10] - and C. R. Rao, Differencial Metric in Probability Space, Probability and Mathematical Statistics, vol 3, fase 2 (1984), p. 241-258.

[11] - Entropy differencial metric, distance and divergence measure in probability spaces - a unifield approach, T. Multivariate Anal. 12 (1982), p. 575-596.

[12] - On the convexity of some divergence measure based on entropy functions, IEEE Trans. Information Theory 28 (1982), p. 489-495.

[13] W. F. Donoghue. Jr, Monotone matrix functions and analitic continuation, Springer-Verlag, Berlin 1974.

[14] S. Kobayashi. Hyperbolyc manifolds and holomorphic mapping, M. Dekker, New York 1970

[15] S. Kullback and R. A. Leibler, On information and sufficiency, Ann. Math. Statist. 22 (1951).

[16] K. Matusita, Decision rules based on distance for problem of fit, two samples and estimation, Ann. Math. Staitst. 26 (1955), p. 631-640.

[17] C. R. Rao, Information and accurracy attainable in the estimation of statistical parametrs. Bull. Calcutta Math. Soc. 37 (1945), p. 81-91.

[18] - Linear statistical inference and its applications. J. Wiley. New York 1973.

[19] M. Skwarcyznski, The invariant distance in the theory of pseudo-conformal transformation and the Lu Qi Keng conjecture, Proc. Amer. Math Soc. 22 (1969), p. 305-310.