

# Aula de Erudição: Encontros com Mestres Notáveis

Julio Michael Stern <sup>1</sup>

30-05-2009, rev. 29-01-2010

## Sumário

<b>1</b>	<b>Trajetória Pessoal</b>	
	(Encontros com Mestres Notáveis)	<b>3</b>
<b>2</b>	<b>Visão Sobre o Cargo de Professor na</b>	
	<b>Universidade do Sistema de Pesquisa</b>	<b>7</b>
<b>3</b>	<b>Produção Acadêmica em 1999 - 2009</b>	<b>11</b>
3.1	FBST, Panorama Geral e Teoria Estatística . . . . .	11
3.2	FBST - Aplicações . . . . .	13
3.3	Lógica e Epistemologia . . . . .	16
3.4	Fatorações Esparsas de Redes Credais . . . . .	17
3.5	Outros Temas de Pesquisa . . . . .	18
<b>4</b>	<b>Projetos de Inovação Tecnológica</b>	<b>19</b>
4.1	Software e Algoritmos de Alto Desempenho . . . . .	19
4.2	Projetos FAPESP PITE e PIPE em 1999 - 2009 . . . . .	21
<b>5</b>	<b>MaxEnt - 2008</b>	<b>26</b>
<b>6</b>	<b>Seleção de Publicações, Softwares, Patentes, etc.</b>	<b>27</b>

---

<sup>1</sup>Julio Michael Stern *jmstern@hotmail.com*, é Graduado e Mestre pelo Instituto de Física da Universidade de São Paulo, M.Eng. e Ph.D. pela School of Operations Research and Industrial Engineering, Cornell University (Ithaca, NY, USA), Livre Docente pelo Departamento de Ciência da Computação, Professor Adjunto do Departamento de Matemática Aplicada do Instituto de Matemática e Estatística da Universidade de São Paulo, Consultor e Pesquisador Nível 1-D do CNPq na área de Pesquisa Operacional.

וַיִּשְׂאֵל יַעֲקֹב וַיֹּאמֶר הַגִּידָה נָא שְׁמֶךָ וַיֹּאמֶר לְמַה זֶה תִּשְׂאֵל לְשִׁמִּי וַיְבָרֶךְ אֹתוֹ שֵׁם:

E perguntou Jacob: Explique-me teu nome.  
“Por quê?” - Eis o que o meu nome pergunta!  
E ali ele o abençoou.

Genesis, XXXII, 30.

“Vivendo, se aprende; mas o que se aprende,  
mais, é só a fazer outras maiores perguntas.”

João Guimarães Rosa (1908-1967).  
Grande Sertão: Veredas.

Agradeço a todos aqueles que compartilharam comigo seu tempo e suas idéias na jornada em busca do saber e do entendimento, e a todos aqueles que me auxiliaram ao longo do caminho.

# 1 Trajetória Pessoal

## (Encontros com Mestres Notáveis)

Um dos requisitos de um memorial é uma descrição da trajetória (acadêmica) pessoal. Existem muitas maneiras de se contar uma história; escolhi contar a minha por algumas idéias que me influenciaram, e os mestres de quem as aprendi. Relato principalmente as idéias que ao mesmo tempo me surpreenderam e me modificaram, daí a quantidade de pontos de exclamação.

Meu interesse acadêmico começa no ginásio Bialik. Com meu professor de história, o Professor José Roberto de Souza Dias, aprendi a importância de estudar com bons livros, independentemente do livro texto “recomendado”. Aprendi também a diferença entre treinar uma habilidade e compreender uma teoria. A primeira refere-se a adquirir a capacidade de executar uma determinada tarefa. A segunda significa internalizar criticamente uma visão de mundo. Cada qual tem sua hora e sua vez.

Ainda no ginásio tive o prazer de participar das atividades dos Congressos Jovens Cientistas, organizados no IBICC pela Professora Maria Julieta Ormastroni, que sempre convidava alguns “cientistas de verdade” para conversar com a garotada. Um destes convidados foi o Professor Toledo Piza do IF-USP, o Instituto de Física da Universidade de São Paulo. Em uma palestra sobre metodologia científica, alguém opinou que, antes de atacar um problema, um cientista deveria ler toda a literatura existente sobre o assunto. O Professor Piza levantou-se e contestou: “Assim não se faz nada de original; leia o suficiente, e comece a procurar suas próprias soluções!”

Da Professora Maria Julieta recebi brinquedos maravilhosos, como os kits e livros do IBICC e dos programas PSSC (Physical Science Study Committee) e BSCS (Biological Science Curriculum Study). Ao cursar o Colégio Bandeirantes, continuei a participar dos Congressos Jovens Cientistas, onde apresentei um trabalho sobre *O Problema da Indução*, premiado no XIX Concurso Cientistas de Amanhã, realizado pelo IBICC/UNESCO em 1976.

Na graduação no IF-USP, adquiri uma base ampla e sólida para meus estudos futuros. Acho que a disciplina que mais me marcou foi o primeiro ano de Física básica, com o Professor Moyses Nussenzveig. De duas aulas me lembro até hoje:

- Numa aprendemos a usar a notação complexa para resolver a equação diferencial do oscilador harmônico forçado e amortecido. Logo em seguida, numa demonstração de bancada, vimos como as duas funções reais descritas

de forma unificada na notação complexa (exponencial e senoidal) apareciam conjuntamente na física do problema. O recurso matemático não era só um truque esperto, mas revelava algo essencial do problema em estudo!

- Noutra aula, discutimos a idéia do Teorema de Noether, deduzindo leis de conservação pela análise de choques entre partículas vistos de referenciais equivalentes (pelo grupo de Galileu). Quantidades conservadas (objetos físicos) advinham da geometria, ou de transformações invariantes da teoria!

Vários professores do IME-USP alertaram-me para sempre balancear o árido rigor formal com uma intuição viva. O Professor Seiji Hariki, de vetores e geometria, insistia para que “enxergassemos” a solução antes começar a fazer contas! Anos antes, na EMA - a escolinha de astronomia no planetário do Ibirapuera, eu já havia recebido exatamente a mesma mensagem pelo Professor Acácio Riberi (que era cego!). O Professor Edson Farah disse-me certa ocasião: Mais do que demonstrando teoremas, Topologia se aprende sabendo de cor(ação) uma boa lista de figuras com exemplos e contra-exemplos!

Nos primeiros anos da graduação, interessei-me por Epistemologia e Filosofia da Ciência. Freqüentei os seminários de fundamentos da matemática do Professor Newton da Costa. O Professor Newton sempre nos fazia pensar sobre “o que quer dizer” um teorema, exemplo ou contra-exemplo. Desta forma, passei a ver a Lógica não apenas como um formalismo útil, mas como uma linguagem importante para genuína investigação filosófica.

Por indicação do Professor Henrique Fleming, fiz uma iniciação científica sobre a história do desenvolvimento da mecânica clássica. Ainda por indicação do Professor Fleming, atento ao meu crescente interesse pela matemática e métodos formais, fiz o mestrado em Gravitação e Geometria Lorentziana com o Professor Edgard Harle do IME-USP, o Instituto de Matemática e Estatística. Neste estudo novamente me deparei com o conceito fundamental de que um objeto da teoria (geometria) deve ser caracterizado por alguma forma de invariância. O Professor Harle me dizia: “Se está complicado é porque você ainda não entendeu direito!” Segundo o Professor Harle, elegância em matemática se resume em poder e simplicidade. Nas palavras de Albert Einstein: “Explique uma idéia sempre da forma mais simples possível, mas nunca mais simples que isto!”

Em 1983 fui contratado no MAP, o Departamento de Matemática Aplicada do IME-USP. Seu catedrático, Professor Waldir Muniz Oliva, me explicou a teoria do incesto intelectual e esclareceu que eu deveria fazer um doutorado no exterior. No MAP, tive vontade de conhecer uma área mais aplicada, e assisti a disciplinas de Programação Linear, Não-Linear e Métodos

de Decomposição para problemas estruturados com o Professor Carlos Humes. Decidi que Pesquisa Operacional seria meu caminho, no espírito da máxima de Ludwig Boltzmann: “Não há nada tão prático como uma boa teoria!”

Nos seminários do Professor Carlos Humes, éramos incessantemente bombardeados com as quatro (temíveis) perguntas:

- 1- Explique (desenhe uma figurinha)!
- 2- Dê um (contra) exemplo!
- 3- Demonstre!
- 4- Mostre como o algoritmo funciona (faça a conta)!

Certamente, este foi um bom treino para a vida acadêmica em Cornell.

Fiz meu Ph.D. na Escola de Engenharia Industrial e Pesquisa Operacional da Universidade de Cornell onde, nos três primeiros semestres, cursei disciplinas obrigatórias em Análise Matemática, Otimização, Processos Estocásticos, Estatística e Simulação, além de optativas para um M.Eng. com “major” em Otimização e “minors” em Estatística e Computação (Análise Numérica). Meu orientador de Doutorado, Professor Stephen Vavasis, só queria saber de me fazer publicar. Um belo dia ele me recomendou: Passe menos tempo olhando para os livros, e mais para um papel em branco! Uma de suas máximas era: “Não seja um paiol de pólvora pronto para explodir, seja uma espingarda pronta para atirar, e vá caçar! Para ele, a pesquisa acadêmica estava no desenvolvimento da teoria e métodos originais, em oposição à erudição estéril ou à mera aplicação de técnicas conhecidas.

Em Cornell obtive interessantes resultados nas áreas de otimização numérica e álgebra linear computacional para problemas descritos por matrizes esparsas e estruturadas. Tive a oportunidade de implementar estes algoritmos em máquinas paralelas, como convidado do Argonne National Laboratory.

De meu retorno ao Brasil em 1991 até 1998, dediquei-me intensamente a serviços de extensão e consultoria na área de Pesquisa Operacional. Com a BM&F, a Bolsa de Mercadorias e Futuros, firmamos um convênio com a USP e implantamos na FEA, a Faculdade de Economia Administração e Contabilidade, a primeira disciplina de graduação sobre mercados financeiros derivativos. Desenvolvemos ainda o software acadêmico e comercial *CriPo - Critical Point for Windows*, para análise e otimização de carteiras de investimento, incluindo contratos de opção. Representando no Brasil o software de otimização GAMS (General Algebraic Modeling System), do Banco Mundial (World Bank) tive muitas oportunidades interessantes para consultoria em planejamento econômico, financeiro industrial, social e tributário.

Destas experiências da prática de Pesquisa Operacional, creio ter compreendido a máxima de Richard Hamming: “O propósito da modelagem é intuição, não números!” Mais ainda, creio ter compreendido que um mod-

elo só realiza plenamente seu benefício potencial para uma empresa se for incorporado à vida e cultura da organização, sendo rotineiramente utilizado para exercícios de simulação, modelagem de decisões, processos de negociação (interna ou externa), etc. com o envolvimento dos formadores de opinião e responsáveis pela empresa. Assim, há muito mais entre a abstração matemática e a realidade do chão de fábrica que um modelo de Pesquisa Operacional ou, como me disse o executivo Eduardo Bassi do grupo Bunge y Born, “Da próxima vez que você disser que (a solução) é trivial, eu te enforco!”

Paralelamente, continuei minha atividade acadêmica de ensino e pesquisa, nas áreas de modelagem econômica e financeira, otimização e estatística aplicada. Obtivemos também a patente de um algoritmo para compensação multilateral em tempo real de ordens de pagamento. Nesta experiência confirmei a importância de definir, prévia e claramente, o escopo e a responsabilidade de um trabalho especializado. Sigo sempre a política dos colegas Fábio Nakano e Marcelo Lauretto: “Entregamos (à empresa integradora) e mantemos o algoritmo como um código claro, limpo, bem documentado e portátil, com funções de I/O igualmente simples e portáteis!”

A partir de 1999 dediquei-me prioritariamente ao programa de pesquisa FBST, ou Teste Bayesiano Completo de Significância, uma teoria alternativa para o teste estatístico de hipóteses precisas. Desenvolvemos o FBST com o Professor Carlos Alberto de Bragança Pereira (Carlinhos), que me ensinou quase toda a estatística que sei (isto foi fácil) depois de me des-ensinar quase toda a estatística que eu pensava que sabia (bem mais difícil). O Professor Carlos Bragança é um mestre da maiêutica, o método socrático que, pelo diálogo, leva o interlocutor a descobrir a (sua) verdade. Uma de suas frases prediletas é do Professor Oswaldo Frota Pessoa: “O mundo se divide entre os cretinos, que agem por conveniência, e os genuínos, que agem por convicção!”

A teoria do FBST deu-me a oportunidade de olhar novamente, agora de outra perspectiva, para muitos assuntos pelos quais me interessei ao longo de minha trajetória acadêmica. Esta trajetória foi bastante diversificada e, embora não tenha experimentado em nenhum momento uma quebra de continuidade, tive por vezes a sensação de relegar ao esquecimento interesses passados. A teoria do FBST possibilitou-me uma visão autônoma e elegante de muitos destes temas, o que é motivo de grande satisfação intelectual.

Escrever estes textos de memorial forçou-me a uma longa reflexão retrospectiva. As sensações que me ficam deste exercício são a de que fazer esta jornada foi um grande privilégio, e uma profunda gratidão para com tantos mestres notáveis, que comigo compartilharam seu tempo e suas idéias.

## 2 Visão Sobre o Cargo de Professor na Universidade do Sistema de Pesquisa

Professor (titular) = Posição de Liderança na Universidade.

A equação acima é (quase) uma tautologia.

Resta portanto responder à pergunta: Qual Universidade?

O conceito que denomino Universidade do Sistema de Pesquisa (USPe) é um paradigma acadêmico cuja concepção básica remonta a Leibnitz, von Humboldt, e outros, sendo caracterizada como uma instituição que tem por diretrizes os seguintes princípios:

1- *Pesquisa Autônoma*. A USPe faz parte de um sistema autônomo de pesquisa científica, isto é, dedicado a criação e produção de conhecimento científico.

2- *Theoria cum Praxi*. Esta máxima Leibniziana indica a possibilidade de um ciclo virtuoso de mútua inspiração e fomento entre a pesquisa teórica e o desenvolvimento tecnológico, sem que haja qualquer forma de subordinação que viole os demais princípios.

3- *Indissociabilidade de Ensino e Pesquisa*. Na USPe as atividades de ensino e pesquisa são mutuamente engajadas, referindo-se ao mesmo corpo dinâmico de saber. Este princípio, que von Humboldt enunciava como “ensino através da pesquisa”, oferece ao aluno a oportunidade de elevar-se, com a orientação dos professores, a co-participante do esforço de pesquisa, propiciando-lhe, no processo, um ensino de excelência.

4- *Liberdade Acadêmica*. Todas as atividades referidas nos princípios 1, 2 e 3 devem ser exercidas em um ambiente de respeito à liberdade acadêmica individual dos pesquisadores.

5- *Meritocracia*. Avaliação de desempenho acadêmico deve ser o fator preponderante para o acesso, ingresso e progresso do aluno / pesquisador dentro da USPe.

6- *Transdisciplinaridade*. Disciplinas são unidades operacionais mínimas na estrutura de articulação de um corpo de conhecimento cada vez maior e mais complexo. Esta estrutura operacional decorre da necessidade de atribuir responsabilidades pedagógicas, administrar recursos eficientemente, etc. No entanto, a USPe deve evitar o perigo de que esta estrutura operacional leve a uma especialização estanque, onde as disciplinas são erroneamente consideradas como unidades cognitivas fechadas e independentes. Pelo contrário, a USPe deve caracterizar-se por uma grande abertura interdisciplinar, pro-

movendo a pesquisa transdisciplinar.

O paradigma USPe, na formulação supra e suas muitas variações, é reconhecido como o fundamento da Universidade moderna, e já foi objeto de profunda reflexão e vários artigos. O termo *Universidade moderna* indica sobretudo uma instituição voltada à criação e produção do conhecimento (inovação), em contraponto ao termo *Universidade medieval*, que indica uma instituição voltada primordialmente à fiel transmissão do conhecimento (preservação) e, secundariamente, à assimilação de conhecimentos provenientes de fontes externas (organização).

Minha formulação do paradigma USPe pouco acrescenta à idéia básica. Creio contudo ser oportuno tecer algumas considerações sobre inserção da USPe no presente contexto econômico, histórico, social e político, deste início do século XXI. Objetivando manter o foco do texto, apresentarei estas considerações como normativas referentes à inserção social da USPe.

a- *Diversidade das Instituições de Ensino Superior*. O paradigma USPe, com toda a sua comprovada eficiência, elegante simplicidade e idealismo acadêmico não é, e nem deve ser, o único modelo implementado de instituição de ensino superior. Em particular, é imprescindível que haja instituições dedicadas a cursos de caráter técnico ou profissionalizante.

b- *Diferenciação das Instituições de Ensino Superior*. As diversas modalidades de instituição referidas em (a) devem ser diferenciadas. Apenas uma instituição diferenciada pode almejar a excelência em sua modalidade. O contrário, a de-diferenciação, implica uma instituição amorfa e incapaz de coordenar coerentemente seus recursos.

c- *Mérito Acadêmico e Mobilidade Social*. A USPe deve funcionar como um mecanismo de mobilidade social, onde o mérito acadêmico (e não a influência econômica ou política) determinam o sucesso pessoal. É obvio que, sendo a educação um processo contínuo, este aspecto depende da existência de um bom sistema público de ensino (fundamental e médio).

d- *Importância Estratégica da Competência Científica e Tecnológica*. Nenhuma nação pode manter um desenvolvimento sustentável sem dispor de quadros de alta competência em ciência e tecnologia. A USPe é o celeiro natural para esta competência.

e- *Importância Estratégica da Competência Transdisciplinar*. A especialização fechada leva à perda da capacidade de pensar relações amplas, o que por sua vez leva à perda da capacidade de orientação em contextos amplos. Esta capacidade é estratégica em uma sociedade complexa e, novamente, a USPe é o celeiro natural para esta competência.



f- *Multiplicidade de Fontes de Financiamento*. Historicamente o estado é o financiador primário da USPe, e assim o continuará sendo em todo cenário previsível. Todavia, a capacidade de financiamento do estado parece ter encontrado seu limite, em virtude de restrições tributárias, que por sua vez decorrem de restrições econômicas e demográficas. Assim, faz-se necessário encontrar fontes complementares de financiamento.

Até aqui, o texto analisa o modelo USPe em caráter geral. Cabe agora um pequeno comentário histórico-crítico sobre a Universidade de São Paulo. A USP é a maior instituição de ensino superior no Brasil a seguir (ao menos em parte) o paradigma USPe. Sua construção, a partir década de 30, tinha entre diversos objetivos (declarados ou não), três que gostaria de destacar:

i) Formar profissionais qualificados em um projeto nacionalista e desenvolvimentista para um “Brasil País do Futuro”, como diria Stefan Zweig.

ii) Formar um núcleo intelectual com identidade própria, capaz de contribuir ao estudo da realidade Brasileira e sua inserção internacional.

iii) Atender aos filhos da elite econômica de São Paulo impedidos de estudar na Europa, cuja crise abria também novas oportunidades.

No mundo bipolar pós segunda guerra, este projeto recebeu grandes incentivos de programas do Banco Mundial, Fundação Ford, USAID e outros, em uma convergência de interesses geopolíticos nacionais e internacionais.

Uma leitura superficial da realidade presente pode levar à seguinte falácia:

*Falácia da (In) Competência Global*: A indústria tradicional é uma atividade poente, neste mundo globalizado em geral, e no Brasil em particular. A boa ciência sempre foi universal e, em um mundo globalizado, é globalmente acessível. Assim há pouco interesse em manter uma instituição no paradigma USPe no Brasil. Finalmente, os poucos indivíduos de uma elite intelectual que queiram se dar ao luxo da criação científica, podem praticar seu hobby em alguma instituição no exterior.

A falácia da (in) competência global, em todo ou em parte, e obviamente de forma mais dissimulada, nos é hoje reiteradamente apresentada no discurso de alguns indivíduos. Entre eles,

- Há aqueles que representam toda sorte de interesses escusos, desde o mais reles entreguismo, até o interesse econômico de grandes corporações cujos interesses conflitam com a existência de uma USPe. A estes nos cabe combater, abertamente, sem medo e sem trégua.

- Há também estes bem intencionados inocentes (úteis), que vêm na diferenciação da USPe, uma forma de atender a legítimos anseios populares de acesso a educação. A estes nos cabe esclarecer, mostrando que esta é uma

forma imediatista, ineficiente e demagógica de atender a estes anseios, que tem (a médio prazo) um custo imenso: O colapso da USPe, pelo uso de seus recursos de forma incompatível com sua autonomia.

Esta demagogia faz lembrar no Brasil outro processo que, embora diverso, apresenta algumas semelhanças: A degradação do sistema público de ensino (fundamental, médio e técnico), cuja qualidade foi deliberada, inútil e sistematicamente sacrificada em nome de ideais pouco claros e interesses obscuros. Este processo primeiro abusou da resiliência de instituições virtuosas – embora incapazes de ser a panacéia do mundo – para depois encontrar o desespero de um sistema aprisionado em um ciclo vicioso de baixa qualidade.

Preservar a USPe em sua autonomia não implica de forma alguma a indiferença em relação aos problemas da sociedade, uma USPe na “torre de marfim”. Pelo contrário, a USPe deve, através de convênios e serviços de assessoria, consultoria e extensão, assumir sua responsabilidade social, desenvolvendo e fornecendo cursos especiais, orientação e materiais pedagógicos, métodos e processos tecnológicos e conhecimentos científicos, contribuindo assim ativamente com os demais sistemas da sociedade.

### **Referências:**

- G.Casper (1996). *Eine Welt ohne Universitäten?* München: Bayerische Akademie der Wissenschaften. Em português: *Um Mundo sem Universidades?* Editora UERJ, 2003.
- W. von Humboldt (1810). *Über die innere und äussere Organisation der höheren wissenschaftlichen Anstalten in Berlin.* Werke, Stuttgart: Klett-Cotta, 1993. In english: *On the Spirit and the Organizational Framework of Intellectual Institutions in Berlin.* Minerva, 8 4/1970, 243-260. Em português: *Sobre a Organização Interna e Externa das Instituições Científicas Superiores em Berlin.* em Casper (1996).
- W.Krohn, G.Kueppers, H.Nowotny (1990). *Selforganization. Portrait of a Scientific Revolution.* Dordrecht: Kluwer.
- N.Luhmann (1989). *Ecological Communication.* Chicago University Press.
- N.Luhmann (1995). *Social Systems.* Stanford University Press.
- J.Mittelstrass (1995). *Transdisciplinarity.* Panorama, 5/1995, 45-53.
- J.Mittelstrass (2002). *Leibniz’s World: Calculation and Integration.* Vienna University.
- R.Stichweh (1990). *Self-Organization and Autopoiesis in the Development of Modern Science.* in Krohn et al. (1990).

## 3 Produção Acadêmica em 1999 - 2009

Nesta seção faço um rápido apanhado de meus trabalhos de pesquisa no período de 1999 a 2009. Neste período, concentrei-me no desenvolvimento de uma nova teoria para teste de hipóteses precisas, o FBST. O FBST pode ser caracterizado por várias e importantes propriedades teóricas. A primeira sub-seção lista os principais resultados que estabelecem a teoria estatística para o novo método. A segunda sub-seção lista as principais aplicações e métodos numéricos desenvolvidas neste programa de pesquisa. A terceira sub-seção lista resultados correlatos pertinentes às áreas de Lógica e Epistemologia.

### 3.1 FBST, Panorama Geral e Teoria Estatística

O FBST, Full Bayesian Significance Test, ou Teste de Significância Bayesiano Completo, foi especialmente desenhado para dar um valor credal, o e-valor  $ev(H|X)$ , refletindo a crença na hipótese precisa  $H$ , dadas as observações  $X$ . O e-valor é ainda interpretado como o *valor epistêmico* da hipótese  $H$  (em função dos dados  $X$ ), ou o *valor da evidência*  $X$  (suportando a hipótese  $H$ ).

O principais objetivos do programa de pesquisa do FBST são:

1- Fornecer uma teoria intuitiva, coerente e conveniente para acessar a significância estatística de hipóteses precisas;

2- Fornecer os métodos matemáticos e algoritmos numéricos necessários à implementação do FBST;

3- Fornecer um arcabouço epistemológico adequado para a teoria do FBST, e seu uso em ciência e tecnologia;

4- Fornecer uma variedade de problemas de teste prototípicos, em uma larga gama de áreas de aplicação;

5- Fornecer os correspondentes programas computacionais, na forma de módulos em um ambiente integrado de programação e software estatístico aberto (licença GNU, como Octave, R ou Python).

A função de suporte do FBST, o e-valor,  $ev(H|X)$ , tem muitas propriedades desejáveis ou necessárias a uma função de suporte estatístico a hipóteses precisas, como:

(I) Fornecer uma função de suporte à (ou medida de significância estatística da) hipótese em teste, idealmente uma medida de probabilidade no espaço paramétrico original ou natural do problema; vide *PS99* e *MPS03*;

(II) Ter uma definição intrinsecamente geométrica, independente de qualquer aspecto não geométrico, como a particular parametrização da (variedade

representando a) hipótese sendo testada, ou o particular sistema de coordenadas escolhido para o espaço paramétrico, i.e., ser um procedimento invariante; vide *MPS03* e *Ste07a*;

(III) Fornecer uma função de suporte que seja suave, i.e. contínua e diferenciável, nos parâmetros da hipótese e nas estatísticas da amostra, dentro de condições apropriadas de regularidade do modelo; vide *MPS03*, *LPSZ03* e *Ste07a*;

(IV) Obedecer ao princípio da verossimilhança, i.e. a informação obtida das observações deve ser representada pela, e apenas pela, função de verossimilhança; vide *MPS03*, *LPSZ03* e *Ste07a*;

(V) Não requerer qualquer artifício ad hoc, como dar probabilidades positivas a conjuntos de medida nula, ou estabelecer razões de crença iniciais arbitrárias entre hipóteses; vide *MPS03*, *LPSZ03* e *Ste07a*;

(VI) Ser uma função de suporte possibilística, onde o suporte a uma disjunção lógica é o máximo entre o suporte dos disjuntos; vide *Ste03*, *Ste04* e *BS07*;

(VII) Ser um procedimento exato, i.e., não utilizar na definição do e-valor qualquer aproximação assintótica;

(VIII) Fornecer um teste consistente para uma dada hipótese precisa; vide *Ste04* e *BS07*;

(IX) Fornecer operações de composicionalidade para modelos complexos; vide *BS07*;

(X) Permitir a incorporação de experiência prévia ou opiniões de especialistas via distribuições a priori, vide *ILPS02*.

As características teóricas do FBST enumerados neste decálogo coletam as melhores características encontradas quer nos testes de hipótese da Estatística freqüentista baseados em p-valores, quer nos testes Bayesianos baseados em fatores de Bayes. Ademais, algumas destas das propriedades, como as operações de composicionalidade para modelos complexos, não são encontradas em nenhum dos antigos concorrentes.

Finalmente, após 10 anos de árduas e duras batalhas, foi publicado o artigo *PSW08 - Can a Significance Test be Genuinely Bayesian?* - na revista *Bayesian Analysis*, a mais prestigiada revista da área. Considero este um evento que marca a aceitação definitiva da teoria estatística e da metodologia do FBST pela comunidade Bayesiana internacional.

## 3.2 FBST - Aplicações

Vários problemas aplicados foram resolvidos com sucesso com o FBST. Algumas destas soluções são, em nossa opinião, (muito) melhores que as abordagens alternativas existentes. Por exemplo, em *SZ02* comparamos o FBST com todos os testes de independência anteriormente publicados para a distribuição de Holgate, e mostramos que o FBST é consistentemente o teste mais poderoso.

Em outras aplicações, o FBST resolveu problemas que não podiam ser resolvidos por técnicas alternativas. Por exemplo, em *LPSZ03* usamos o FBST para testar hipóteses gerais na forma de restrições de proporcionalidade entre os parâmetros de duas distribuições normais bivariadas. Este artigo generaliza uma vasta literatura de testes para uma série de casos (muito) particulares de nossa hipótese geral.

Como já mencionamos anteriormente, o FBST fornece um arcabouço teórico unificado e homogêneo para testar hipóteses precisas. O FBST não requer nenhum “truque especial”, como um sistema de coordenadas esperto que permita a eliminação analítica de parâmetros *nuisance*, uma priori especial feita sob medida para prevenir singularidades de integração ou o paradoxo de Lindley, massas de probabilidade atribuídas de maneira *ad hoc* a conjuntos de medida (de Lebesgue) nula, etc.

O FBST requer, isto sim, algoritmos eficientes para otimização e integração numérica; mas estes algoritmos, uma vez desenvolvidos e implementados para um problema, podem ser usados em uma grande gama de problemas similares em tamanho e complexidade.

Nosso modelo de demonstração favorito é a hipótese de equilíbrio genético de Hardy Weinberg. Este modelo multinomial tem espaço paramétrico bidimensional e limitado com uma hipótese unidimensional não linear. Assim, é um problema não trivial de tamanho minimal e ao mesmo tempo fácil de visualizar. O problema de Hardy Weinberg é uma das aplicações usadas para introduzir o FBST em *PS99b*, cuja extensão ao caso multi-alélico encontra-se em *NPSW06* e *LNFSP09*. *PS99b* contém ainda outras aplicações, como um problema de engenharia de confiabilidade testando a igualdades de duas funções de sobrevivência. O artigo também apresenta o FBST para outras hipóteses em modelos multinomiais, como testes de independência e homogeneidade.

Problemas de teste em modelos multinomiais, como os apresentados em *PS99b*, foram a chave para procedimentos de certificação e verificação de

software apresentados em *PS99a*. Estes problemas de teste em engenharia de software surgiram em nossa prática de consultoria, e motivaram vários dos requisitos teóricos para o FBST.

Em *ILPS02*, apresentamos um problema de teste complexo em engenharia de confiabilidade. Este problema ressalta a importância de utilizar, no modelo estatístico, informação a priori proveniente de fenômenos básicos da engenharia do problema, como a caracterização da distribuição Weibull como distribuição de vida limite para um equipamento com grande número de componentes em série, e o efeito de procedimentos de pré-teste (burn-in) para limitar o parâmetro de forma da distribuição de vida dos componentes individuais.

*PS01a* apresenta casos particulares do teste geral de estrutura para problemas multivariados apresentado em *LPS03*. Esta é uma hipótese de equivalência proporcional, afirmando a proporcionalidade de um par de medidas de respostas multivariadas a dois estímulos distintos. Neste artigo, testamos a equivalência até segunda ordem, i.e., a equivalência dos vetores de média e das matrizes de covariância. Este problema surgiu em nossa prática de consultoria como um problema de formulação em engenharia química ou farmacológica, tratando da bioequivalência de dois produtos farmacêuticos.

*PS01b* é o primeiro artigo aplicando o FBST ao importante problema de seleção de modelos. Um dos casos de teste no artigo surgiu na nossa prática de consultoria, como um problema de engenharia ambiental referente à administração de um aterro desativado de resíduos sólidos industriais, onde modelamos a concentração química de um contaminante tóxico. A concentração e o volume de contaminante eram os aspectos chave para a decisão de (se ou onde) limpar (por incineração) ou deixar o solo contaminado em sua condição presente.

Em *SZ02* apresentamos problemas de teste para a distribuição de Holgate, uma versão bivariada da distribuição de Poisson. Este problema foi motivado por uma aplicação desenvolvida para estimar chances de aposta em campeonatos de futebol. Modelos relacionados, com distribuições bivariadas ou multivariadas, discretas ou contínuas, surgem em muitas outras áreas de aplicação.

*LS05* aplica o FBST a modelos de mistura. Esta classe de modelos é usada em uma variedade de aplicações práticas, como para análise de cluster. *LFPS07* utiliza este modelo para resolver o importante problema estatístico de *hipóteses separadas*. Aplicações práticas destes trabalhos incluem modelos de clusterização e análise de expressão gênica, vide *LPS08*. O Pro-

fessor Marcelo de Souza Lauretto, da USP-Leste, está desenvolvendo novas aplicações nas áreas de genética e bio-tecnologia.

*DPS07* e *DPS08* implementam o FBST para testes de raiz unitária e co-integração. Estes problemas são notórios pela dificuldade que impõem à modelagem Bayesiana ortodoxa. Mais uma vez, a performance do FBST é melhor que a alternativa ortodoxa, e sua formulação é elegante, simples e direta, contrastando com os elaborados truques e tecnicidades requeridas para a formulação de fatores de Bayes. O Professor Márcio Alves Diniz, do Departamento de Estatística da Universidade Federal de São Carlos, está desenvolvendo novas aplicações nas áreas de econometria e finanças.

*HLS08* implementa o FBST para uma classe de distribuições Lagrangianas que generaliza a distribuição de Poisson para taxas não homogêneas no número de eventos ocorridos. Esta classe de distribuições permite a modelagem de riscos crescentes, cumulativos ou efeitos de fragilização, e tem aplicações importantes nas áreas de seguro de acidentes e risco operacional.

Ao nosso ver, já desenvolvemos um número suficiente de aplicações para ter confiança de que:

- O FBST é uma ferramenta útil e poderosa, que pode ter um desempenho melhor que técnicas rivais em muitos problemas, ou que pode ser implementada em muitas aplicações onde as técnicas rivais simplesmente não podem ser utilizadas;

- Os algoritmos e programas de otimização e integração numérica de que dispomos até agora formam uma base eficiente, estável e robusta para uma biblioteca que pode ser utilizada para desenvolver muitas outras aplicações.

Todavia, o ritmo de desenvolvimento de novas aplicações tem sido retardado pela falta de um ambiente de programação amigável e bem organizado. O desenvolvimento de uma caixa de ferramentas para o FBST em uma linguagem de programação de alto nível com uma biblioteca de funções de uso geral (numérico e estatístico) e facilmente expansível, integrada a um ambiente computacional de alta performance, é uma de nossas prioridades atuais. Temos a intenção de usar um ambiente aberto (licença GNU) como Octave, R ou Python como plataforma deste projeto.

### 3.3 Lógica e Epistemologia

No programa de pesquisa do FBST, hipóteses precisas são consideradas de importância fundamental em ciência e tecnologia. Em contraste, as duas maiores escolas de estatística, a freqüentista (clássica) e a Bayesiana ortodoxa, tradicionalmente vêem hipóteses precisas apenas como aproximações grosseiras de hipóteses imprecisas. Esta visão tradicional tem duas explicações:

a) As dificuldades técnicas impostas por hipóteses precisas para testes de significância, quer seja na forma tradicional da teoria freqüentista (p-valores), quer seja na teoria Bayesiana (com maior gravidade em fatores de Bayes);

b) As dificuldades teóricas impostas por hipóteses precisas, tanto no arcabouço epistemológico freqüentista, o Falsificacionismo Popperiano, como no arcabouço Bayesiano ortodoxo, baseado na Teoria da Decisão.

Assim, foi necessário desenvolver um arcabouço epistemológico coerente para o FBST. Este arcabouço foi baseado no Construtivismo Cognitivo, em propriedades essenciais de auto-soluções, e na teoria de sistemas autopoieticos. Os artigos *Ste06*, *Ste07a*, *Ste07b* e *Ste08a*, publicados ou aceitos para publicação na revista *Cybernetics & Human Knowing*, a melhor revista de filosofia nesta área específica, estabelecem este arcabouço epistemológico.

Já os artigos *Ste03 - Significance Tests, Belief Calculi, and Burden of Proof in Legal and Scientific Discourse*, e *BS07 - The Rules of Logic Composition for the Bayesian Epistemic e-Values*, investigam as propriedades lógicas do e-valor, demonstrando suas regras algébricas de composição intra ou inter modelos. Como nem o p-valor da estatística clássica, nem o fator de Bayes da estatística Bayesiana ortodoxa, possuem propriedades semelhantes, optamos por publicar estes resultados em revista e eventos de Lógica ou Inteligência Artificial, onde resultados de composicionalidade para cálculos de credais são corriqueiros, ainda que geralmente baseados em formalismos discretos, e não em medidas contínuas.

Como resultado desta linha de investigação chegamos também à caracterização do e-valor como uma medida *possibilística* no espaço das hipóteses de um modelo estatístico, construído sobre a densidade *a posteriori*, que é uma medida *probabilística* no espaço paramétrico do mesmo modelo.

Finalmente, o artigo *Ste04a - Paraconsistent Sensitivity Analysis for Bayesian Significance Tests* explora alguns paralelos entre as noções de consistência / inconsistência em Lógica e Estatística, estabelecendo novas inter-relações entre os cálculos credais possibilístico e probabilístico no contexto do FBST.



### 3.4 Fatorações Esparsas de Redes Credais

No artigo *Ste08a* discutimos a importância do princípio de desacoplamento na ciência em geral, dando exemplos concretos na Física (auto-funções da corda vibrante) e na Estatística (estrutura e esparsidade da matriz de covariância). Estudamos também o papel dos operadores de desacoplamento destes modelos e sua representação por operações de fatoração em álgebra linear computacional (no caso dos exemplos citados, respectivamente, fatoração SVD e fatoração de Cholesky).

Ocorre que a estrutura lógica ou combinatória da operação de desacoplamento de modelos de covariância é muito geral, aplicando-se imediatamente a redes de propagação para cálculos credais abstratos ou ABCs - Abstract Belief Calculi. Dentre estas redes credais, as redes Bayesianas são certamente as mais conhecidas e utilizadas. Este fato já havia sido notado por alguns pesquisadores mas, o quanto eu saiba, ainda não havia sido aplicado ou utilizado para nenhum fim prático, como o desenho de algoritmos eficientes.

Em *SC09*, mostramos que o fato da estrutura lógica de acoplamento / desacoplamento ser comum a todos estes problemas permite separar completamente a solução dos mesmos problema (nos exemplos, a fatoração de Cholesky da matriz de covariância ou a fatoração de potenciais em redes Bayesianas) em duas fases completamente separadas: Uma primeira fase puramente combinatória, e uma segunda fase puramente numérica.

Este fato é amplamente conhecido da comunidade de métodos computacionais para matrizes esparsas, sendo a base de todo e qualquer algoritmo eficiente. Ademais, todas as estruturas de dados podem ser previamente alocadas (ou ter a alocação agendada) na fase combinatória, de modo que a fase numérica utilize apenas estruturas de dados estáticas. Esta separação do trabalho nas fases combinatória e numérica é também essencial para implementação eficiente de algoritmos paralelos, vide *Ste92*, *SV93*, *SV94* e *Ste94*.

Todavia, todas as implementações previamente existentes para inferência em redes Bayesianas ignoravam estas possibilidades para desenvolver algoritmos eficientes e facilmente paralelizáveis. Esperamos que a publicação destes resultados contribua significativamente para melhorar a performance de futuras implementações de programas de produção para inferência em redes credais de grande porte.

### 3.5 Outros Temas de Pesquisa

*RSP01* - Algoritmo randomizado para concretização e compensação de mensagens de pagamento em tempo real. Este algoritmo foi patenteado pela FINANTEC, nossa parceira comercial.

*BPRS02* - Estudo ecológico de manejo ambiental apresentando o modelo Beta-Poisson, inédito na literatura estatística. A solução analítica em um suporte compacto é obtida através de séries de funções especiais hipergeométricas confluentes de Kummer.

*GHS03* - CLG, Critério de Linha Generalizado, para convergência de métodos tipo Gauss-Seidel. Esta é uma condição suficiente de convergência para métodos tipo GS baseada no valor absoluto da matriz de coeficientes. Este critério é significativamente mais geral e poderoso que os anteriormente conhecidos. Obter um resultado assim é algo surpreendente em uma área de pesquisa com três séculos de história.

*SZ03* - Estimador tipo Monte Carlo para quocientes funcionais utilizando uma distribuição,  $g$ , de amostragem por importância. Através do método delta, derivamos uma variância assintótica,  $V(g, m)$ , que, após algumas simplificações, pode ser usada para controlar a duração da simulação de Monte Carlo, ou quantidade de pontos,  $m$ . Em trabalhos posteriores mostramos ainda vários exemplos de como utilizar duas estratégias para acelerar a convergência do integrador: a) Estratificação, b) Aproximações assintóticas da posteriori como variável de controle.

*Ste92* - Método de Simulated Annealing acelerado através de uma função objetivo com um termo heurístico de resfriamento rápido. A idéia original foi apresentada neste artigo, de 1992. A demonstração de convergência para este tipo de cadeia de Markov inhomogênea foi apresentada em Pflug (1996), veja também *Ste08b, H.1*.

*LNPS09* - Modelos hierárquicos para previsão de demanda. Nestes modelos, uma primeira camada implementa modelos padrão de séries temporais SVARMA sobre dados econométricos, enquanto uma segunda camada baseada em redes neurais ou polinomiais incorpora dados qualitativos. A síntese destas redes é feita por Programação Genética, utilizando idéias em *Ste92*. Trabalhos em andamento focalizam a emergência de módulos ou blocos construtivos, e sua interpretação semântica.

## 4 Projetos de Inovação Tecnológica

### 4.1 Software e Algoritmos de Alto Desempenho

Projetos na área de computação científica ou alto desempenho têm características próprias. Faremos a seguir - *Algumas Considerações sobre o Desenvolvimento de Software Baseado em Algoritmos de Alto Desempenho, e sua Inserção em Programas de Inovação Tecnológica*. Estas reflexões foram importantes para a instituição de bolsas para técnicos de alto nível, especialmente programadores, dentro dos programas PITE e PIPE da FAPESP.

#### Etapas de Desenvolvimento do Projeto

Existe um grande caminho a ser percorrido desde um algoritmo fruto de um projeto de pesquisa acadêmica até o desenvolvimento de um produto comercial. A seguir, detalharemos alguns aspectos deste caminho:

O *programa de alto nível* para efeito de teste numérico de um algoritmo é usualmente o trabalho de pesquisa de um aluno de pós-graduação levando a uma tese de mestrado ou a um artigo conforme a sofisticação do algoritmo e sua implementação, e o grau de inovação introduzido. Obviamente, isto não inclui o desenvolvimento do algoritmo em si, usualmente resultado da pesquisa de um aluno de doutorado ou do orientador.

A transformação do programa para testes numéricos em um *protótipo* inclui a definição e uso de estruturas de dados eficientes, gerenciamento eficiente de memória, entrada e saída de dados, etc. Tipicamente esta transformação também inclui a troca de uma linguagem de alto nível na qual foi programado o algoritmo para testes numéricos (ex. Octave, R, Python) para linguagens adequadas à implementação do protótipo (ex. Fortran, C, C++). Nesta fase todo o código deve ser mantido, na medida do possível, portátil, i.e., independente de especificidades e particularidades de plataforma de hardware ou sistema. Também as funções de entrada e saída devem ser modulares e portáteis. Nesta fase, há alto envolvimento do pesquisador sênior para definição e orientação das linhas mestras da implementação. Devido à sofisticação existente nestas definições e sua codificação, o trabalho de programação decorrente pode ser parte central de uma dissertação de mestrado, ou parte do trabalho em um programa de doutorado.

Finalmente, temos que considerar a transformação do protótipo em *produto*. Nesta fase, o código do protótipo deve ser otimizado, considerando

inclusive os detalhes de arquitetura das plataformas (hardware, sistema operacional e software). Devem ainda ser definidas e desenvolvidas as diversas interfaces entre módulos que compõem o produto final (interface de usuário, recursos para manipulação de dados, visualização e análise, etc.). Esta é a fase mais demorada e de maior custo de programação. Esta etapa requer programadores de um perfil distinto do das etapas anteriores, porém de alta qualificação. Esta é também a fase em que ocorre a maior transferência tecnológica da instituição de pesquisa para o cliente. Esta fase demanda um grande esforço de integração, de modo a tornar harmoniosa a absorção da nova tecnologia ao produto ou processo previamente existente no cliente.

### **Competitividade, Software e Inovação Tecnológica**

A indústria de software pode, usando em seu benefício alguns efeitos gerados por sua própria tecnologia de produção, superar dificuldades da conjuntura local e ganhar competitividade, tanto no mercado doméstico quanto no mercado externo. Entre estes efeitos destacamos:

- Desintermediação: A eliminação na cadeia produtiva, de elos que não agregam valor. Por exemplo, é possível comercializar mundialmente um software especializado a partir de uma página na (world wide) web, eliminando distribuidores e representantes comerciais;

- Globalização: A permeabilidade das fronteiras nacionais ao capital e à informação. Em contraste, a indústria de software requer pouca permeabilidade para bens tangíveis (hardware), e praticamente nenhuma permeabilidade para o deslocamento físico de recursos humanos;

- Diferenciação: A capacidade de ter ganho de escala com artigos diferenciados. Por exemplo, é possível customizar as interfaces e até algumas funcionalidades de um software às necessidades de vários clientes, usando os mesmos algoritmos básicos de solução;

- Integração: A capacidade de integrar partes produzidas em unidades separadas. Com boa engenharia de projeto, é possível contratar a programação de módulos específicos de várias empresas independentes (e sem intercomunicação), para depois facilmente integrá-los em um produto final.

Finalmente, vale ressaltar que a indústria de software é “labor intensive” e que o Brasil tem boa disponibilidade de recursos humanos nesta área. Conseqüentemente, existe um interesse estratégico na aplicação de recursos para inovação tecnológica em projetos de desenvolvimento de software.

## 4.2 Projetos FAPESP PITE e PIPE em 1999 - 2009

### FAPESP PITE 96/2341-2

Projeto: Solver de Alto Desempenho para Problemas na Forma Angular Blocada Aninhada.

Pesquisadores: Julio Michael Stern, Miguel Taube Netto, Fábio Nakano.

Este projeto teve como parceira a empresa UniSoma, representada pelo Prof. Miguel Taube Netto do IMECC-UNICAMP e Presidente da UniSoma. Parte deste trabalho constituiu o cerne da Tese de Mestrado de Fábio Nakano, defendida no IME-USP em 1999.

O objetivo do projeto foi o desenvolvimento de solvers de Programação Matemática especialmente adaptados para Otimização de problemas com estrutura (forma) angular bloqueada por linhas (RBAF - Row Block Angular Form), e forma angular bloqueada por linhas aninhada (NRBAF - Nested Row Block Angular Form); vide *Ste92*, *SV93*, *SV94* e *Ste94*. As formas NRBAF, e sua dual NCBFAF bloqueada por colunas, aparecem em problemas estruturados de grande porte, comuns nas áreas de programação estocástica, otimização e controle de sistemas distribuídos ou hierárquicos, etc.

O principal cliente da UniSoma no projeto foi o grupo Sadia Alimentos, onde um modelo de Pesquisa Operacional foi usado para determinar os percentuais dos ingredientes de rações sujeitas a restrições de qualidade (níveis de nutrientes e toxinas), para otimização do valor (custo) de operação multi-período, aplicação clássica de programação linear. Problemas agroindustriais típicos de multi-rações apresentam da ordem de 25 restrições por ração para 40 rações, perfazendo 1.000 restrições. Quando formulamos o problema multi-período temos tipicamente 12 períodos, ou da ordem de 12.000 restrições. Considerando um número de 30 ingredientes, diferenciados por origem em 10 regiões, temos da ordem de 150.000 variáveis. Um problema de grande porte altamente estruturado.

No projeto desenvolvido pela UniSoma na Sadia Alimentos, o gasto anual com rações para aves na empresa era da ordem de meio bilhão de Dólares Americanos. A formulação multi-mistura acarretou reduções de custo da ordem de 3%, e a formulação multi-período da ordem de 3% adicionais, ou seja uma economia anual de US\$ 30 milhões.

## FAPESP PIPE 02/07887-6

Projeto: Análise Computacional de Exame Genético de Paternidade

Pesquisadores: Martin Ritter Whittle, Julio Michael Stern, Carlos Alberto de Bragança Pereira, Fábio Gagliardi Cozman, Fábio Nakano.

O projeto visa desenvolver algoritmos para cálculo de paternidade. Esta aplicação visa estabelecer se um determinado indivíduo, o Demandante, é filho de um segundo indivíduo, o Demandado, também denominado Pai Putativo ou Pai Alegado na literatura forense. Os testes hoje mais comuns nos tribunais utilizam análises de DNA (Ácido Desoxirribonucleico) para avaliar a probabilidade de relações familiares nestas disputas judiciais. Este é o chamado cálculo de (probabilidade de) paternidade ou de vínculo genético.

Os algoritmos desenvolvidos para o projeto são baseados em Redes Bayesianas, e permitem relaxar três das hipóteses básicas pré-supostas por métodos alternativos:

- 1- Homogeneidade, ou equilíbrio genético de Hardy-Weinberg,
- 2- Independência, ou ausência de consangüinidades espúrias, e
- 3- Ausência de mutações.

Cabe ressaltar que a hipótese 1, assumida como verdadeira em outros métodos, provavelmente está distante da realidade de uma população geneticamente heterogênea, como a população brasileira. Também já tivemos que analisar alguns casos, envolvendo casamento entre primos, em que a hipótese 2 é falsa. Na análise de pedigrees em genética animal, uma aplicação colateral do projeto, esta condição é bastante comum.

Quando à hipótese 3, a Genomic (empresa parceira) utiliza marcadores em micro-satélites (STRs - Short Tandem Repeats). Estas pequenas seqüências de DNA são sabidamente suscetíveis a mutações, com taxa da ordem de uma a cinco por mil em cada geração. Esta taxa de mutação é muito maior que a probabilidade de erro no resultado final exigida em análises forenses. Portanto a possibilidade de mutações não pode ser desprezada na análise de paternidade, sob o risco de obter-se “falsas exclusões”.

Felizmente, as técnicas de modelagem por nós desenvolvidas com o uso de Redes Bayesianas torna possível a análise de paternidade, relaxando todas as três hipóteses supra mencionadas, desde que tenhamos disponível um software de alto desempenho para propagação de probabilidades em redes Bayesianas, como o JavaBayes do Professor Fabio Cozman.

## FAPESP PIPE 02/12864-5

Projeto: Otimizador de Inserções em Mídia através do Modelo de Média-Variância

Pesquisadores: Pedro Jesus Fernandez, Julio Michael Stern, Carlos Alberto de Bragança Pereira, José Afonso Mazzon, Carlos Humes Jr, Fábio Nakano, Marcelo Lauretto.

O produto desenvolvido é um sistema integrado de otimização e análise de planos de mídia (no escopo deste projeto, mídia impressa). O profissional de mídia é a pessoa responsável em agências de propaganda ou em anunciantes por desenvolver uma estratégia de alocar um dado orçamento de propaganda em diferentes mídias, meios e veículos de comunicação, de tal modo que maximize a atenção, o interesse, a compreensão e a retenção da mensagem dirigida a uma audiência-alvo ou target, por meio de uma campanha publicitária concebida por um profissional de criação.

A analogia entre a otimização de portfólios de investimento e a otimização de grades de inserções é bastante intuitiva: em ambos os problemas, os recursos disponíveis devem ser aplicados entre os diversos ativos disponíveis (ou veículos, no mercado de mídia), vide *Ste9608*. A principal diferença conceitual é que o retorno esperado de uma campanha publicitária não se dá em termos monetários (ao menos não diretamente), mas em termos de indicadores de penetração (alcance, hits, hits/indivíduo, etc).

Os otimizadores de mídia ora em uso no mercado são baseados em otimizadores lineares, que não contemplam a correlação entre os diversos canais de mídia, a não linearidade dos efeitos de exposição, etc. O principal objetivo deste projeto foi o desenvolvimento de um sistema de planejamento ótimo de mídia que incorporasse o modelo de média-variância, utilizando seu potencial e suas vantagens como paradigma de otimização.

O artigo *FSL07* apresenta a base teórica do projeto. Este artigo foi apresentado em 2005 na *Advertising Research Foundation Conference*, um dos mais importantes eventos internacionais da área.

## FAPESP PIPE 06/60831-0

Projeto: Sistema de Suporte à Decisão para Manejo de Bovinos.

Pesquisadores: Flávio Augusto Portela Santos (ESALQ), Julio Michael Stern, Carlos Alberto de Bragança Pereira, Fábio Nakano, Marcelo Lauretto.

Este projeto implementa um sistema de suporte à decisão para o planejamento estratégico e tático da produção de bovinos. O sistema incorpora bancos de dados com o histórico de dois grandes grupos de variáveis explicativas relevantes ao problema: BD-VE- Banco de Dados de Variáveis Econômicas. BD-VDAI- Banco de Dados de Variáveis Descritivas de Animais Individuais.

A modelagem matemática do sistema é baseada em três modelos integrados, a saber: 1- MACRECO: Modelo Agregado de Crescimento e Conversão. 2- PRODINE: Modelo de Programação Dinâmica Econômica e Financeira. 3- MOCLIFE: Modelo de Classificação Individual de Acabamento e Peso.

1- O MACRECO avalia o potencial agregado de conversão do plantel disponível na Fazenda. Este modelo ajusta uma função não linear de crescimento (curva de referência), escolhida entre os modelos clássicos de Bertalanfy, Brody, Gompertz, Logística, Michaelis-Menten e Richards, ou ainda as generalizações de Birch, Damgaard, e Wang e Jackson. Foi dada especial atenção às curvas de referência cuja parametrização permite uma interpretação biológica, de modo a propiciar um acompanhamento fácil e intuitivo dos lotes em função das VDAI.

2- O PRODINE fornece os parâmetros ótimos para o planejamento estratégico do negócio. A tecnologia do PRODINE é baseada em Programação Dinâmica em Cadeias de Markov, vide *Ste0608*. As principais incertezas no modelo são preços futuros, de boi gordo e dos insumos, e a disponibilidade de pasto. O PRODINE utiliza os parâmetros das curvas de referência fornecidas pelo MACRECO, e fornece metas agregadas de abate e entrada em confinamento para cada período de produção.

3- O MOCLIFE fornece os parâmetros ótimos para a operação tática do negócio. A operação tática baseia-se na classificação, isto é destinação, de cada animal individual para: A- Abate, C- Confinamento, ou P- Retorno ao Pasto. A tecnologia estatística do classificador do MOCLIFE é o REAL, ou Real Attribute Classification Tree, vide *LNRS98*. O MOCLIFE gera uma solução tática a partir dos parâmetros correspondentes à solução estratégica fornecida pelo PRODINE que, por sua vez, baseia-se em avaliações agregadas do potencial de conversão fornecidas pelo MACRECO.



## FAPESP PIPE 06/156505-0

Projeto: Análise Atuarial por Processos de Ramificação.

Pesquisadores: Cláudio Paiva, Julio Michael Stern, Carlos Alberto de Bragança Pereira, Fábio Nakano, Marcelo Lauretto.

Este projeto implementa um software para análise e simulação para gestão de fundos de pensão do tipo benefício definido. Em *PNS03* apresentamos uma metodologia de modelagem atuarial para análise dos fluxos financeiros em fundos de pensão do tipo benefício definido, tipo de fundo que é muito comum no Brasil, com peculiaridades raramente encontradas em fundos no exterior. O núcleo numérico do software é um módulo de cálculo para os fluxos de encargos e de contribuições em fundos deste tipo.

Um participante do fundo é denominado ativo no período em que realiza suas contribuições para o plano previdenciário, sem o recebimento dos benefícios. Deixa sua condição de ativo no momento em que se torna aposentado, inválido, ou quando falece, iniciando ou interrompendo nestas transições fluxos de benefícios (aposentadoria, pensão, pecúlio, etc.) para seu grupo familiar.

A modelagem dos parâmetros globais dos fluxos de encargos e de contribuições incluem a modelagem de processos de sobrevivência, incluindo a correção de tabelas de vida oficiais, modelagem da dinâmica de grupo familiar típico, crescimento salarial, crescimento da empresa (gerações futuras), etc. Cabe salientar que, neste projeto, nosso esforço de fazer uma modelagem correta do ponto de vista matemático ou estatístico, colidiu mais de uma vez com uma determinação, legal, regulamentação ou sentença judicial “definindo” procedimentos de cálculo *ad hoc*. As discrepâncias resultantes tiveram que ser incorporadas às ferramentas de simulação, aumentando sobremaneira a complexidade do projeto.

O processo estocástico subjacente às transições entre os estados possíveis no fundo de cada participante e seu grupo familiar é modelado com um processo de ramificação, governado por várias taxas de risco dependentes do tempo. Os fluxos de caixa esperados são computados por funções recursivas descrevendo o processo de ramificação. Assim evitamos uma série de aproximações rotineiramente utilizadas na prática atuarial, cuja precisão é difícil de estimar ou controlar. Estas funções recursivas também fornecem um cálculo direto da variância do processo e outras estatísticas úteis.

## 5 MaxEnt - 2008

O MaxEnt 2008, 28th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, foi realizado entre 6 e 11 de Julho de 2008, em Boracéia, litoral de São Paulo, Brasil. Este evento é (provavelmente) o mais prestigiado encontro de Estatística Bayesiana da comunidade de Engenharia, Física e Instrumentação, sendo esta a primeira vez, em mais de vinte anos, que este se realiza fora do eixo Estados Unidos - Europa Ocidental. Foi deste pesquisador a proposta de realizar este evento no Brasil, feita ao comitê organizador do MaxEnt em 2006, confirmada e aceita pelo comitê organizador do MaxEnt em 2007. Para maiores detalhes, vide página do evento em <http://www.brastex.info/maxent2008/>

Para realizar este evento tivemos o apoio de empresas privadas e órgãos de fomento nacionais e estrangeiros, entre os quais destacamos (taxa de câmbio 1.6 R\$/US\$): 1- E.T. Jaynes Foundation / Boise State University: R\$ 29.368,00 e US\$ 8.000,00; 2- FAPESP: R\$ 34.693,00; 3- CAPES: R\$ 12.780,00; 4- IM-AGIMB (CNPq): R\$ 5.400,00; 5- PROEX Estatística: R\$ 5.175,00; 6- PROAP Matemática Aplicada: R\$ 5.073,00; 7- FIA (FEA-USP): R\$ 3.115,00; 8- FIPE (FEA-USP): R\$ 2.011,00; 9- IME-USP: R\$ 4.000,00; 10- IF-USP: R\$ 3.000,00 (xerox, gráfica, perua e motorista, valor estimado); 11- Universia (Santander): R\$ 500,00 (brindes, valor estimado).

O MaxEnt 2008 teve 4 tutoriais, 9 palestrantes convidados, 26 apresentações orais e 35 posters, com participantes de instituições de 13 países, a saber: Argentina, Bélgica, Brasil, Colômbia, França, Alemanha, Israel, Itália, Japão, Holanda, Suíça Estados Unidos e Reino Unido, com cidadãos de mais 8 países, a saber, Bulgária, China, Iran, Nepal, Rússia, Espanha, Síria e Turquia. Ao contrário de anos anteriores, nenhum participante teve problemas para obtenção de visto ou permissão para entrar no país.

O MaxEnt 2008 contou com a apresentação de trabalhos teóricos nas áreas de Estatística Bayesiana, formalismos de máxima entropia e fundamentos de probabilidade. Foram apresentadas ainda soluções Bayesianas para tratamento de dados científicos incluindo, não exclusivamente, as áreas de Física Geral, Instrumentação, Física de Plasmas, Física de Materiais e da Matéria Condensada, Astronomia e Astrofísica, Tratamento de Imagens, Diagnóstico Médico, Economia, Econometria e Finanças, etc. Como de hábito, o MaxEnt 2008 publicou proceedings de alta qualidade através do AIP - American Institute of Physics.

## Seleção de Livros e Artigos Publicados, Softwares, Patentes, etc.

Apresento a seguir uma seleção com alguns itens de minha produção acadêmica entre os que considero mais importantes ou representativos de minhas diversas linhas de pesquisa. Dou preferência a itens mais recente, que continuem a repercutir fortemente na minha produção atual, ou citados diretamente neste memorial.

### Livros

-*Ste08b*- J.M.Stern (2008). *Cognitive Constructivism and the Epistemic Significance of Sharp Statistical Hypotheses*. Santa Cruz do Rio Pardo: Editora Viena. Este livro texto foi preparado para o curso de verão Inferência Estatística Comparada, MAE-5747, em 2006 e 2007, e impresso para um dos tutorias do MaxEnt 2008.

-*LPS08*- M.S.Lauretto, C.A.B.Pereira, J.M.Stern, eds. (2008). *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*. Melville, NY: American Institute of Physics. *AIP Conference Proceedings*, 1073. Este é o volume de proceedings do MaxEnt 2008.

-*ST96*- J.M.Stern, R.Terada (1996-2008). *Vade-Me-Cum, Um Guia para Programação em C*. IME-USP. Este texto foi preparado pelo primeiro autor para a disciplina de introdução à programação no IF-USP em 1994. Em 1996 o texto foi revisado e ampliado com a colaboração do segundo autor, sendo utilizado desde então na maioria dos cursos introdutórios de computação na USP. Trata-se pois, de um dos textos didáticos mais utilizados na história da Universidade de São Paulo.

-*Ste9608*- J.M.Stern et al (1996-2008). *Métodos de Otimização em Economia e Finanças*. Goiânia: XIX Cong.Nac. de Matemática Aplicada e Computacional, 1996. IME-USP / BM&F, 2008. Texto preparado no Convênio USP - BM&F em 1994, sendo utilizado nas disciplinas MAC/MAP-419 (graduação), MAC/MAP-5796 e EAE-866 (pós), bem como nos mestrados profissionalizantes da USP e BM&F.

-*Ste94*- J.M.Stern (1994). *Esparsidade, Estrutura, Estabilidade e Escalonamento em Álgebra Linear Computacional*. Recife: IX Escola de Computação. Texto preparado visando criar as disciplinas MAC/MAP-300 (graduação) e MAC/MAP-5795 (pós), pois não havia na USP, até então, disciplinas dedicadas especificamente aos métodos numéricos da álgebra linear.

## Métodos Esparsos e Estruturados

- SC09*- J.M.Stern, E.C.Colla (2009). Factorization of Sparse Bayesian Networks. p.275-294 in: K.Nakamatsu, G.Phillips-Wren, L.C.Jain, R.J.Howlett, eds. *New Advances in Intelligent Decision Technologies*. Heidelberg: Springer.
- LNPS09*- M.Lauretto, F.Nakano, C.A.B.Pereira, J.M.Stern (2009). Hierarchical Forecasting with Polynomial Nets. In Nakamatsu (2009), 305-315.
- GHS03*- M.V.P.Garcia, C.Humes Jr, J.M.Stern (2003). Generalized Line Criterion for Gauss-Seidel Method. *Computat.and Applied Math.*, 22, 91-97.
- SV94*- J.M.Stern, S.A.Vavasis (1994). Active Set Methods for Problems in Column Block Angular Form, *Computational and Applied Mathematics*, 12, 199-226.
- SV93*- J.M.Stern, S.A.Vavasis (1993). Nested Dissection for Sparse Nullspace Bases. *SIAM Journal of Matrix Analysis and Applications*, 14, 766-775.
- Ste92*- J.M.Stern (1992). Simulated Annealing with a Temperature Dependent Penalty Function. *ORSA Journal on Computing*, 4, 311-319.

## Lógica e Epistemologia

- Ste09a*- J.M.Stern (2009). The Living and Intelligent Universe. RT-MAP-09-04. Accepted, MBR09 - The Internaternational Conference on Model-Based Reasoning in Science and Tecnology 2009, Dec 17-19, Unicamp, Brazil.
- Ste08a*- J.M.Stern (2008). Decoupling, Sparsity, Randomization, and Objective Bayesian Inference. *Cybernetics and Human Knowing*, 15, 2, 49-68.
- BS07*- W.Borges, J.M.Stern (2007). The Rules of Logic Composition for the Bayesian Epistemic e-Values. *Logic J.of the IGPL*, 15, 5-6, 401-420.
- Ste07b*- J.M.Stern (2007). Language and the Self-Reference Paradox. *Cybernetics and Human Knowing*, 14, 4, 71-92.
- Ste07a*- J.M.Stern (2007). Cognitive Constructivism, Eigen-Solutions, and Sharp Statistical Hypotheses. *Cybernetics and Human Knowing*, 14, 1, 9-36.
- Ste06*- J.M.Stern (2006). Language, Metaphor and Metaphysics: The Subjective Side of Science. RT-MAC-06-09 (submitted).
- Ste04a*- J.M.Stern (2004). Paraconsistent Sensitivity Analysis for Bayesian Significance Tests. SBIA'04, *Lecture Notes Artificial Intell.*, 3171, 134-143.
- Ste03a*- J.M.Stern (2003). Significance Tests, Belief Calculi, and Burden of Proof in Legal and Scientific Discourse. Laptec-2003, *Frontiers in Artificial Intelligence and its Applications*, 101, 139-147.

## Estatística, Teoria e Métodos Numéricos

- HLS09- P.C.Hubert Jr., M.Lauretto, J.M.Stern (2009). FBST for a Generalized Poisson Distribution. *AIP Conference Proceedings*, accepted.
- PSW08- C.A.B.Pereira, J.M.Stern, S.Wechsler (2008). Can a Significance Test be Genuinely Bayesian. *Bayesian Analysis*, 3, 79-100.
- PS08- C.A.B.Pereira, J.M.Stern (2008). Special Characterization of Standard Discrete Models. *RevStat - Statistical Journal*, 6, 199-230.
- DPS08- M.Diniz, C.A.B.Pereira, J.M.Stern (2008). FBST for Cointegration Problems. *AIP Conference Proceedings*, 1073, 157-164.
- LFPS07- M.Lauretto, S.R.Faria, C.A.B.Pereira, J.M.Stern (2007). The Problem of Separate Hypotheses via Mixtures Models. *AIP Conference Proceedings*, 954, 268-275.
- DPS07- M.Diniz, C.A.B.Pereira, J.M.Stern (2007). FBST for Unit Root Problems. *AIP Conference Proceedings*, 954, 260-267.
- LS05- M.Lauretto, J.M.Stern (2005). Testing Significance in Bayesian Classifiers. *Frontiers in Artificial Intelligence and Applications*, 132, 34-41.
- LPSZ03- M.Lauretto, C.A.B.Pereira, J.M.Stern, S.Zacks (2003). Full Bayesian Significance Test Applied to Multivariate Normal Structure Models. *Brazilian Journal of Probability and Statistics*, 17, 147-168.
- MPS03- M.R.Madruga, C.A.B.Pereira, J.M.Stern (2003). Bayesian Evidence Test for Precise Hypotheses. *Journal of Statistical Planning and Inference*, 117, 185-198.
- SZ03- J.M.Stern, S.Zacks (2003). Sequential Estimation of Ratios, with Applications to Bayesian Analysis. RT-MAC-03-10.
- SZ02- J.M.Stern, S.Zacks (2002). Testing Independence of Poisson Variates under the Holgate Bivariate Distribution. The Power of a New Evidence Test. *Statistical and Probability Letters*, 60, 313-320.
- PS01b- C.A.B.Pereira, J.M.Stern (2001). Model Selection and Regularization: Full Bayesian Approach. *EnvironMetrics*, 12, 6, 559-568.
- PS01a- C.A.B.Pereira, J.M.Stern (2001). Full Bayesian Significance Tests for Coefficients of Variation. p. 391-400 in: E.I.George ed. *Bayesian Methods with Applications to Science, Policy, and Official Statistics*.
- PS99b- C.A.B.Pereira, J.M.Stern (1999). Evidence and Credibility: Full Bayesian Significance Test for Precise Hypotheses. *Entropy Journal*, 1, 69-80.

## Modelagem e Pesquisa Operacional

- LNFP09*- M.Lauretto, F.Nakano, S.R.Faria, C.A.B.Pereira, J.M.Stern (2009). A Straightforward Multiallelic Significance Test for the Hardy-Weinberg Equilibrium Law. *Genetics and Molecular Biology*, 32, 3, 619-625.
- LPS08*- M.Lauretto, C.A.B.Pereira, J.M.Stern (2008). The Full Bayesian Significance Test for Mixture Models: Results in Gene Expression Clustering. *Genetics and Molecular Research*, 7, 883-897, 2008.
- FSL07*- P.J.Fernandes, J.M.Stern, M.S.Lauretto (2007). A New Media Optimizer Based on the Mean-Variance Model. Presented at ARF'05 - Advertising Research Foundation Conference. *Pesquisa Operacional*, 27, 427-456.
- NPSW06*- F.Nakano, C.A.B.Pereira, J.M.Stern, M.R.Whittle (2006). Genuine Bayesian Multiallelic Significance Test for the Hardy-Weinberg Equilibrium Law. *Genetics and Molecular Research*, 4, 619-631.
- PNS03*- C.A.B.Pereira, F.Nakano, J.M.Stern (2003). A Model for Defined Benefit Plans. p.176-181 in J.Dhaene, N.Kolev, P.Morettin eds. *Proceedings of the First Brazilian Conference on Statistical Modelling in Insurance and Finance*. IME-USP.
- BPRS02*- A.M.S.Bueno, C.A.B.Pereira, M.N.Rabelo-Gay, J.M.Stern (2002). Environmental Genotoxicity Evaluation: Bayesian Approach for a Mixture Statistical Model. *SERRA - Stochastic Environmental Research and Risk Assessment*, 16, 267-278.
- ILPS02*- T.Z.Irony, M.Lauretto, C.A.B.Pereira, and J.M.Stern (2002). A Weibull Wearout Test: Full Bayesian Approach. p. 287-300 Y.Hayakawa, T.Irony, M.Xie, eds. *Systems and Bayesian Reliability, Quality, Reliability and Engineering Statistics*, 5, Singapore: World Scientific.
- PS99a*- C.A.B.Pereira, J.M.Stern (1999). A Dynamic Software Certification and Verification Procedure. *ISAS-SCI'99 Proceedings*, 2, 426-435.
- LNRS98*- M.Lauretto, F.Nakano, C.Ribeiro, J.M.Stern (1998). Real Attribute Learning Algorithm. *ISAS-SCI'98 Proceedings*, 2, 315-321.

## Softwares e Patentes:

- RSP01*- G.F.G.Rafare, J.M.Stern, C.A.B.Pereira, F.Nakano (2001). *Algoritmo de Concretização e Compensação de Mensagens de Pagamento*. Tipo de Programa: AT03, AT04, TC02; Campos de Aplicação: EC04, EC06, FN03, FN04. Número de registro no INPI: 00042036 (11 envelopes).
- SN93*- J.M.Stern, F.Nakano (1993). *Critical Point for Windows*.