

UNIVERSIDADE FEDERAL DE SANTA CATARINA  
CENTRO TECNOLÓGICO  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE  
PRODUÇÃO

UM ENFOQUE SEGUNDO A TEORIA DE  
CONJUNTOS DIFUSOS PARA A META-ANÁLISE

TESE SUBMETIDA À UNIVERSIDADE FEDERAL DE SANTA CATARINA PARA  
OBTENÇÃO DO TÍTULO DE DOUTOR EM ENGENHARIA DE PRODUÇÃO.



UFSC-BU

PAULO AFONSO BRACARENSE COSTA

(BU)

ORIENTADOR: PROF. DR. ROGÉRIO CID BASTOS

FLORIANÓPOLIS, OUTUBRO DE 1999

# UM ENFOQUE SEGUNDO A LÓGICA DIFUSA PARA A META-ANÁLISE

PAULO AFONSO BRACARENSE COSTA

ESTA TESE FOI JULGADA ADEQUADA PARA A OBTENÇÃO DO TÍTULO DE DOUTOR EM ENGENHARIA DE PRODUÇÃO E APROVADA EM SUA FORMA FINAL PELO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO.



---

PROF. DR. RICARDO MIRANDA BARCIA  
COORDENADOR DO CURSO

BANCA EXAMINADORA



---

PROF. DR. ROGÉRIO C. BASTOS  
ORIENTADOR



---

PROF. DR.ª ZÉLIA M. PAVÃO



---

PROF. DR. NEWTON C. A. DA COSTA



---

PROF. DR. ROBERTO PACHECO



---

PROF. DR. PEDRO BARBETTA



---

PROF. DR. ALEJANDRO MARTINS  
MODERADOR

*"Donna ti voglio cantare,  
donna la madre, donna la  
finne, donna sei roccia, donna  
sei sabbia e a volte nuvola  
sei...".*

Do cancionero italiano  
Angelo Branduardi

Dedico este trabalho à minha família. Meus pais, Oilde e Otto (*in memoriam*) pelo apoio que sempre me deram. Meus irmãos e cunhados pelo incentivo e apoio. Minhas filhas Helena e Natália e minha esposa Maria Emília, pelo carinho e paciência, e por terem compreendido que em certos momentos tive que me ausentar.

## AGRADECIMENTOS

Agradeço ao meu orientador, Professor Rogério Cid Bastos, pela sua amizade e orientação recebida, ajudando a moldar a tese apoiando minhas iniciativas.

Agradeço ao meu orientador no exterior, Professor Abe Kandel, da University of South Florida, por ter me acolhido naquela instituição americana, por sua orientação e oportunidade de trabalhar integrado com sua equipe.

Agradeço ao Dr. Roberto da Silveira Moraes, do Hospital das Clínicas da UFPR pelo auxílio como especialista na construção dos conjuntos difusos e na determinação das probabilidades de relevância.

Agradeço ao Professor João Cândido Bracarense Costa, do Departamento de Matemática e Estatística da UNIOESTE, irmão e amigo pelo apoio na discussão sobre os modelos utilizados. Pessoalmente e através de "chats" EUA-Brasil. Brasil-França.

Agradeço aos Professores Klir, DuMouchel, Verdinelli, Yusuf, Woodbury, Duckstein e Yuan pela gentileza das respostas às minhas solicitações através da cessão de materiais e softwares.

Agradeço ao meu amigo Professor Roberto Tadeu Raitz, da Escola Técnica da Universidade Federal do Paraná. Fonte inesgotável de questionamento e boas idéias. E também ao Professor João Artur de Souza da Universidade Federal de Pelotas pelo apoio sempre presente.

Através de Ricardo Cury e Marcello Thierry, companheiros de Sandwich, agradeço aos colegas de curso pelas inúmeras discussões que ajudaram a construir esse trabalho.

Agradeço aos amigos Elizabeth Gwan-Nulla, Dave e Faye Chitty, Suresh e Renuh Khator por terem ajudado a tornar a convivência nos Estados Unidos algo muito agradável.

Através da Rose Willard, Anne e Carlos Nouer, agradeço aos diversos amigos da comunidade brasileira nos Estados Unidos, por terem nos adotado imediatamente como filhos.

Agradeço à amiga Jussara Elias do Rego, responsável pelo setor de bolsas da Pró-reitoria de Pesquisa e Pós-Graduação da Universidade Federal do Paraná, pela sua competência e auxílio sempre presente em cada fase do doutorado.

Agradeço aos professores do Programa de Pós-graduação em Engenharia de Produção pelos ensinamentos recebidos.

Agradeço aos colegas do Departamento de Estatística da Universidade Federal do Paraná, por terem me liberado para a realização do curso de doutorado, apesar do aumento de suas cargas de trabalho.

Agradeço aos meus alunos através de Daniel Poit e Ricardo Piétröbon. Enquanto alunos me introduziram à Teoria de Conjuntos Difusos e aos Métodos de Meta-análise.

Agradeço aos membros da banca, cada um foi responsável de alguma forma pela minha formação intelectual.

Agradeço à Professora Eunice Passaglia do Departamento de Engenharia Civil por sua intermediação junto à Petrobrás para que a defesa pudesse ser apresentada através do sistema de video-conferência daquela empresa. Agradeço também à Petrobrás pela gentileza da cessão de sua estrutura para a apresentação da tese.

Agradeço à CAPES, pelas bolsas recebidas.

## RESUMO

Meta-análise é um procedimento estatístico para se fazer sínteses de evidências. Na produção de conhecimento científico, a despeito de saltos de qualidade paradigmáticos, o levantamento do acúmulo de conhecimento adquirido (tendo por base estudos já realizados) é, no geral, a forma como a ciência avança para realizar novas descobertas e confirmar outras já obtidas.

Os métodos para a consecução da meta-análise, análise de análises previamente realizadas, têm sido desenvolvidos ao longo desse século. Mais recentemente, a partir de trabalhos da metade da década de 70, houve um grande impulso nos métodos meta-analíticos, por terem sido estendidos a uma grande variedade de áreas do conhecimento, notadamente na educação, psicologia e medicina. Substituindo a revisão de literatura narrativa por formas mais científicas de síntese com a utilização de instrumental estatístico de significância.

Acompanhando a forte tendência de mudança paradigmática no trato com a ignorância e a incerteza, o presente trabalho propõe uma metodologia para a realização de meta-análise, incorporando ao componente aleatório da incerteza, outro componente que busca modelar incertezas cuja natureza não são aleatórias. Considera-se, então, a vaguesa, a imprecisão e a ambigüidade na modelagem estatística da meta-análise, através da incorporação de conceitos da Teoria de Conjuntos Difusos.

A meta-análise tradicional, tanto no seu enfoque clássico como bayesiano, é realizada através do agrupamento dos níveis de significância ou do chamado "effect size", dos estudos originais que são considerados como unidades amostrais. O "effect size" é uma medida de diferença entre estatísticas obtidas por diferentes tratamentos, padronizada, no geral, por uma variância conjunta. Sua necessidade decorre do fato que estudos com objetivos próximos fornecerem medidas diferentes de efeito, o que impossibilitaria a construção de uma medida comum para os diversos estudos.

Baseado no modelo de Manton para diagnósticos na área médica, propõe-se um modelo probabilístico-difuso que tenha como variáveis independentes o nível de

significância, o tamanho da amostra e o "effect size" de cada estudo. Como resposta busca-se, baseado nessas informações, a "relevância" de cada estudo através do cálculo da probabilidade de um estudo ser relevante ou não. A variável "relevância" tem um significado muito claro na pesquisa clínica, tanto no estudo de diagnóstico como de prognóstico. A relevância é por natureza uma variável com valores lingüísticos que traduz o raciocínio humano na tomada de decisões. O trato com essa variável só pode se dar em termos de ferramental matemático que possa representá-la. A Teoria de Conjuntos Difusos fornece essas ferramentas.

O modelo sugerido pretende calcular a probabilidade de relevância dos estudos que compõem a meta-análise. E a partir desses valores, buscar um valor conjunto dessas probabilidades, que resuma os achados nos estudos particulares. O modelo é intermediário entre a análise de indivíduos do procedimento médico regular e a análise de conjunto de indivíduos dos métodos estatísticos. Constrói, enfim, uma ponte entre o "único" e o "típico".

As variáveis independentes são consideradas como conjuntos difusos, cujos graus de pertinência ponderam as probabilidades subjetivas de um dado elemento pertencer aos componentes de uma partição difusa do espaço amostral. Trabalha-se portanto com duas medidas diferentes de crença: grau de pertinência difusa e probabilidade subjetiva.

O resultado obtido para cada estudo se dá em termos do grau de relevância, determinado pela probabilidade calculada. A composição dessas probabilidades de eventos difusos fornecerá o grau de relevância conjunta de todos os estudos.

O modelo foi aplicado a um conjunto clássico de dados para a realização de meta-análise em pesquisa clínica.



## ABSTRACT

Meta-analysis is a statistical procedure to make syntheses of evidences. In the production of scientific knowledge, in spite of paradigmatic skips of quality, the rising of the accumulation of acquired knowledge (based in carried studies) is, in the general, the way as science moves forward to accomplish new discoveries and to confirm other already obtained.

The methods developed for the attainment of the meta-analysis, analysis of analyses previously accomplished, have been developed during this century. More recently, starting from works of the half of the 70th, there was a great impulse in the meta-analytical methods, for have been extended to a great variety of areas of the knowledge, notably in the Education, Psychology and Medicine. Substituting the revision of narrative literature for more scientific forms of synthesis with the use of statistical instrumental of significance.

Following the strong trend of paradigmatic change in the treatment with the ignorance and the uncertainty, the present work proposes a methodology for the meta-analysis accomplishment, incorporating to the random component of the uncertainty, another component that search uncertainties whose nature is not random. It is considered, then, the vagueness, the imprecision and the ambiguity in the statistical modeling of the meta-analysis, through the incorporation of concepts of Fuzzy Set Theory.

The traditional meta-analysis, as much in its classic approach as bayesian, is carried through pooling the significance levels or of the so called "effect size", of the original studies that are considered as sample units. The effect size is a measure of difference between statistics obtained by different treatments, standardized, in general, by a joint variance. Its necessity elapses from the fact that studies with close purposes, supply different measures of effect, what would disable the construction of a common measure for the several studies.

Based on the Manton model for diagnoses in the medical research, this work considers a probabilistic-fuzzy model, which has as independent variables the significance level, the sample size and the effect size of each study. As response the "relevance" of each study is looked for, based on those information, through the calculation of the probability of a study to be "relevant" or not. The variable relevance has a very clear meaning in clinical research, as much in the diagnosis as prognostic study. It is by nature a variable with linguistic values, which translates the human reasoning in the decision making. The deal with that variable can only be given in terms of mathematical tools that can represent it. Fuzzy Set Theory supplies those tools.

The suggested model intends to calculate the probability of relevance of the studies that compose the meta-analysis. And from those values, to search a joint value of those probabilities, that summarizes the findings in the particular studies. The model is intermediate between the analysis of the individual of the regular medical procedure and the analysis of set of individuals of the statistical methods. It builds, finally, a bridge between the "unique" and the "typical".

The independent variables are considered as fuzzy sets, whose degrees of membership weight the subjective probabilities of a given element to belong to the components of a fuzzy partition of the sample space. It works therefore with two different measures of believe: fuzzy degree of membership and subjective probability.

The result obtained for each study is gave in terms of the degree of relevance, determined by the calculated probability. The composition of those probabilities of fuzzy events will supply the degree of joint relevance of all the studies.

The model was applied to a classic data set for the meta-analysis accomplishment in clinical research.

## SUMÁRIO

1 INTRODUÇÃO	1
1.1 Considerações Iniciais	1
1.2 Objetivos	6
1.3 Justificativa e Importância	7
1.4 Limitações	9
1.5 Estrutura do Trabalho	10
2 A NATUREZA DAS MEDIDAS DE INCERTEZA E IGNORÂNCIA	12
2.1 Introdução	12
2.2 Probabilidade Subjetiva	18
2.3 Teoria de Conjuntos Difusos	24
2.4 Outras Medidas de Incerteza	31
2.5 Teoria de Probabilidades e Teoria de Conjuntos Difusos	40
2.6 Medidas de Probabilidade de Eventos Difusos	47
3 META-ANÁLISE	52
3.1 Introdução - Uma nota histórica	52
3.2 Enfoque Geral para a Meta-análise	54
3.3 Métodos baseados no Nível de Significância	59
3.4 Meta-análise Exploratória e Analítica	61
3.5 Problemas em Meta-análise	63
3.6 Estado da Arte	69
4 MODELO DE PARTIÇÃO DIFUSA PARA A META-ANÁLISE	75
4.1 Introdução	75
4.2 Caracterização do Modelo Geral de Partição Difusa	77
4.3 Estimação dos parâmetros do Modelo Geral	83

	XII
4.4 Definição das Variáveis Difusas para Meta-Análise	88
4.5 Método de Agregação da Variável Relevância	95
5 APLICAÇÃO	97
5.1 Introdução	97
5.2 Apresentação dos Dados	99
5.3 Caracterização das Variáveis Independentes	101
5.4 Determinação da Partição Difusa para o Espaço Amostral	109
5.5 Determinação das Probabilidades de Relevância	111
5.6 Síntese de Evidências	113
5.7 Discussão dos Resultados	114
6 CONCLUSÕES E RECOMENDAÇÕES	118
6.1 Conclusões	118
6.2 Recomendações	120
Referências Bibliográficas	121

## LISTA DE FIGURAS

Figura 2.1.1	Taxinomia da Ignorância	16
Figura 2.3.1	Temperatura em °C no intervalo $[T_1, T_2]$ (a) variável difusa	25
Figura 2.3.2	Temperatura em °C no intervalo $[T_1, T_2]$ (b) variável "crisp"	26
Figura 2.4.1	Um exemplo da violação do axioma da aditividade da probabilidade	34
Figura 2.4.2	Uma típica "função amostral para o processo aleatório do animal	38
Figura 2.4.3	Conjunto X e suas melhores aproximações superior e inferior	39
Figura 2.5.1	Relação de "subconjunção" entre as principais medidas de incerteza	46
Figura 5.3.1	Conjuntos difusos para as variáveis "p-value"	103
Figura 5.3.2	Conjuntos difusos para as variáveis "effect size"	105
Figura 5.3.3	Conjuntos difusos para tamanho da amostra - estatística	107
Figura 5.3.4	Conjuntos difusos para tamanho da amostra - epidemiologia	108
Figura 5.4.1	Perfis extremos para as variáveis difusas	110

## LISTA DE QUADROS

Quadro 5.2.1	Número de pacientes e índices de mortalidade para 6 experimentos	100
Quadro 5.2.2	Número de pacientes e índices de mortalidade para o total de 5 exp.	101
Quadro 5.4.1	Representação dos perfis extremos	110
Quadro 5.5.1	Probabilidades subjetivas de um indivíduo no perfil k ser relevante	111
Quadro 5.5.2	Probabilidades subjetivas de um indivíduo no perfil k ser relevante	112
Quadro 5.5.3	Probabilidades de relevância para os seis estudos	113
Quadro 5.6.1	Cálculos para os modelos de efeitos fixos e de efeitos aleatórios	113
Quadro 5.7.1	Comparação entre os modelos clássico e de partição difusa	116

# 1 Introdução

## 1.1 Considerações Iniciais

Meta-análise é um procedimento estatístico que objetiva fazer sínteses de evidências. Embora o termo tenha sido definido por Glass em 1976, alguns procedimentos anteriores na tentativa de resumir evidências já haviam sido propostos por Legendre e Gauss no início do século XIX.

A sua importância está em tentar fazer revisões com a utilização de métodos que procuram se enquadrar em procedimentos científicos, no lugar de acúmulo de forma somente narrativa.

O que se busca na meta-análise é uma medida comum entre várias pesquisas, que possibilite criar uma única medida para o conjunto de pesquisas já realizadas. No início os trabalhos buscavam métodos que pudessem aglutinar os níveis de significância dos diferentes resultados de pesquisa. Mas surgiam dois problemas. O primeiro com relação à medida que se trabalha. Por exemplo, a expectativa de vida, o "odds ratio" ou risco relativo, eram medidos para os diferentes trabalhos. Então o nível de significância poderia não ajudar muito na síntese. Por outro lado, o nível de significância obtido tem uma relação direta com o tamanho da amostra. Há uma tendência de pesquisas com grandes amostras revelarem maior significância nas diferenças que pesquisas com amostras pequenas. Porque a variância para grandes amostras tende a diminuir. Uma proposta para resolver o problema é trabalhar-se com "effect size", que padroniza as medidas calculando-se a diferença entre as medidas, sejam elas quais forem, dividida pela variância comum entre as variáveis. Assim a questão do tamanho da amostra é de certa forma considerada subjacentemente pela incorporação da variância na medida.

Se fosse possível criar um modelo que considerasse o nível de significância, o "effect size" e o tamanho da amostra conjuntamente para a realização da síntese de evidências, estariam incorporados os elementos chaves para a realização da meta-análise.

Ocorre em quase todas as áreas médicas, que centros de pesquisa diferentes apresentam resultados diferentes quanto à eficácia ou não de uma certa droga ou tratamento. Quando se faz uma revisão na literatura dos diferentes achados, há uma tendência natural de se aproveitar principalmente as informações que corroboram com a hipótese de pesquisa que se está fazendo. A meta-análise procura sintetizar os diferentes achados, mesmo que contraditórios, pela incorporação de maior número de informações. Em última análise o que se deseja é uma posição da relevância de um certo tratamento. Essa decisão deve ser tomada com base no nível de significância (ou "p-value", usados indistintamente nesse trabalho) obtido em cada pesquisa, no tamanho da amostra e no "effect size".

Esse modelo teria então como variável dependente a relevância de um certo tratamento, e como variáveis independentes o nível de significância de cada pesquisa, do tamanho da amostra utilizado e do "effect size".

As definições de que nível de significância considerar como corte para significância que não fique no padrão de 5%, que tamanho de amostra pode ser considerado pequeno ou grande e que "effect size" pode ser considerado como ponto de corte são relativamente subjetivas, que o especialista acaba adotando durante o processo de pesquisa com o acúmulo de sua experiência, e que varia de pesquisador para pesquisador. Como essas, a decisão final de relevância também é subjetiva. Se fosse possível construir um modelo que incorporasse essas impressões subjetivas, ele poderia se aproximar mais da realidade diária de tomada de decisões com base nas informações obtidas dos dados aliada à experiência do pesquisador.

Já na forma tradicional, na decisão de que modelo adotar, e de que variáveis devem participar desse modelo, existe uma forte carga de opinião do pesquisador. Mesmo quando decidido o modelo a ser usado, e que variáveis devem participar dele, a interpretação dos resultados recebe a forma particular como o pesquisador analisa os resultados. Se ele retira



um "outlier" da amostra ou não. Se houver necessidade de alguma transformação de dados, que transformação escolher. Que nível de significância deve ser adotado? Quando da realização de um teste de hipóteses, que teste adotar? Paramétrico ou não paramétrico? Se não paramétrico, qual entre as múltiplas possibilidades adotar? Se paramétrico, no caso de uma ANOVA por exemplo, que teste de comparações múltiplas adotar? Como definir nesse caso os limites entre os conjuntos? O próprio delineamento experimental pode variar tendo conseqüências nos resultados encontrados. Na definição do tamanho da amostra que critério adotar? Que valores atribuir aos erros tipo I e tipo II para cálculo do tamanho da amostra? Que tipo de amostragem utilizar? O procedimento de estimação varia de pesquisador a pesquisador. Que estimador usar? De mínimos quadrados, de máxima verossimilhança, LI ou outro qualquer? A abordagem será clássica ou bayesiana? Se bayesiana que suposições adotar para a distribuição a priori dos parâmetros? Que distribuição será assumida? Como incorporar a opinião do pesquisador para a construção da distribuição a priori? Enfim, os dados não falam por si mesmos. Quando um experimento deve ser realizado ou quando um conjunto de dados precisa ser analisado uma forte carga de decisão pessoal do pesquisador acabará por ter influência no resultado obtido.

No caso da meta-análise, a influência do pesquisador sobre os resultados aumenta. Que procedimento adotar? Clássico ou bayesiano? Se bayesiano, hierárquico ou empírico? Que critérios devem ser adotados para a escolha dos artigos a participarem da amostra? Há uma variedade de propostas de como incluir um artigo, por exemplo, incluir somente artigos publicados em revistas científicas. Ou somente artigos em que o procedimento tenha sido experimental. Existe o "file drawer problem" em que as revistas tendem a publicar somente artigos que tenham dado resultado significativo. Como considerar esse problema? Há várias propostas diferentes de contorno.

Por outro lado, a meta-análise não objetiva somente a construção de medidas que forneçam um resultado único. A análise de artigo por artigo pode fornecer informações importantes. Muitas vezes uma resposta não significativa pode sugerir relações entre as variáveis em estudo. Existem dois tipos de meta-análise. Uma exploratória outra analítica com testes de significância.

Todos esses aspectos sugerem uma interferência da opinião e do raciocínio humano que embora presentes não são explicitados.

Deseja-se então construir um modelo que explicita essas escolhas. As variáveis em consideração, nível de significância, tamanho da amostra, "effect size" e a resposta "relevância" são variáveis que devem ser consideradas com certo grau de incerteza que não é simplesmente incerteza aleatória. A resposta em especial, é fundamentalmente uma variável com valores linguísticos. O que se considera de fato, é que existe um certo grau de ignorância sobre cada um dos aspectos da análise de dados. A ignorância que será considerada é a da difusidade para a construção de um modelo para a análise de um conjunto de dados.

A justificativa para o uso do nível de significância como variável difusa é que para valores entre 0.01 e 0.10 pode-se ter várias considerações de significância podendo-se então construir conjuntos difusos para a determinação da pertinência de um certo valor do nível de significância a um dos seguintes conjuntos difusos: altamente significativo, significativo e não significativo. Para o tamanho da amostra, também serão considerados conjuntos difusos para amostra pequena e amostra grande. Da mesma forma os valores para importância do "effect size" serão determinados por três conjuntos difusos. "effect size" pequeno, médio e grande. Considerando essas variáveis independentes, objetiva-se determinar as probabilidades que um certo experimento seja relevante ou não. E a partir de resultados particulares, buscar uma forma de sintetizar através de uma medida um conjunto de experimentos (artigos). Observe-se que o procedimento meta-analítico não objetiva somente encontrar essa medida mas também analisar o conjunto de resultados um a um, o componente exploratório ou substantivo da meta-análise.

MANTON EL AL.(1994) propõem para análise de diagnóstico o modelo estatístico-difuso GoM, "Grade of Membership" que busca em essência construir a ponte entre o típico, considerado para os procedimentos estatísticos usuais e o único utilizado no procedimento médico comum. Essa ponte é construída através da utilização de conjuntos difusos. Em diagnóstico, toma-se um determinado conjunto de sintomas para o cálculo da probabilidade de um paciente ter uma determinada doença. Cada paciente é único e pode ter doenças correlatas, o procedimento estatístico comum é de agrupar pacientes segundo

características básicas em grupos, particionando-se o espaço amostral de forma "crisp". Dessa forma pode aproximar pacientes de 60 aos de 70 anos considerando-os como idosos, separando pacientes de 59 e 60 anos, ao colocar pacientes de 59 anos como pertencentes ao conjunto de pacientes maduros. Se forem estabelecidos limites de 30 anos ou menos para jovens, 60 anos ou mais para idosos e intemediariamente para pacientes maduros. Já a análise médica usual, considera cada paciente como único. Se fosse possível associar ao paciente graus de pertinência a um dado conjunto estar-se-á construindo a ponte entre o único e o típico. Isso pode ser feito através de conjuntos difusos. Por exemplo um paciente de 45 anos terá um grau de pertinência de 0.1 ao conjunto dos idosos e 0.8 ao conjunto de maduros. Enquanto um paciente com 55 anos poderia ter 0.7 de pertinência ao conjunto dos idosos e 0.3 ao conjunto dos maduros. Pacientes acima de 60 anos teriam grau de pertinência 1 ao conjunto dos idosos e 0 ao conjunto dos jovens e maduros. Tornando assim o conjunto clássico como um caso particular do conjunto difuso. Dessa forma pode-se construir uma partição difusa do espaço amostral.

O modelo proposto pode ser estendido à meta-análise. Ele incorpora dois tipos de graus de crença racional que serão representados pelo graus de pertinência de um paciente a cada um dos elementos da partição difusa e por probabilidades de que um elemento de uma dada partição tenha a característica em estudo, no caso que ele seja relevante. A natureza de cada um desses graus de crença é diferente. O primeiro busca introduzir no modelo a opinião do especialista médico sobre cada uma das variáveis em consideração. Por exemplo, que tamanho de efeito tem importância epidemiológica para um certo tratamento, que pode representar de forma aproximada a percentagem de pacientes que poderiam se beneficiar com o tratamento. Essa opinião do especialista é utilizada para a construção dos conjuntos difusos que representam as três variáveis independentes. O segundo, as probabilidades subjetivas que um elemento de uma partição seja relevante, tem características de medida aleatória, apesar de ser determinada de forma subjetiva.

Então o cálculo final da probabilidade de um estudo ser relevante ou não será uma combinação linear das probabilidades em questão acima, ponderados pelo grau de pertinência de um elemento a uma partição difusa do espaço amostral. Considerando-se as

três variáveis independentes como conjuntos difusos. A resposta será dada em termos de probabilidade de relevância de cada estudo.

Os parâmetros do modelo proposto são as probabilidades e os graus de pertinência de cada indivíduo acima definidos. Então é um modelo com exigência computacional muito grande, porque para cada indivíduo (estudo) ter-se-á que estimar dois parâmetros, a probabilidade e o grau de pertinência. O modelo original faz essa estimativa através do método de máxima verossimilhança. Esse método no entanto é por vezes muito restritivo, e também impede uma maior interferência humana na estimação desses parâmetros o que é desejável. Propõe-se então, construir os conjuntos difusos para nível de significância, tamanho da amostra e "effect size" a partir da experiência de quem fará o procedimento meta-analítico, incorporando o raciocínio humano e graus de crença ao modelo. Da mesma forma as probabilidades a serem construídas para início do modelo, serão subjetivas dentro do conceito de DEGROOT (1970).

O objetivo final é calcular probabilidades de um evento difuso (relevância, não relevância) a partir dos graus de pertinência e probabilidades subjetivas associadas a um espaço amostral com partição difusa.

## 1.2 Objetivos

O objetivo geral desse trabalho é construir um modelo matemático que auxilie a realização dos procedimentos de meta-análise para a produção de conhecimento científico. Esse modelo permite a determinação da probabilidade de eventos difusos a partir da ponderação de probabilidades subjetivas pelos graus de pertinência dos elementos em consideração à uma partição difusa do espaço amostral.

Apresenta-se como objetivos específicos:

- Fazer uma revisão da discussão sobre incerteza e ignorância.
- Apresentar os conceitos básicos da Teoria de Conjuntos Difusos e dos métodos de meta-análise.

- Introduzir as principais teorias de medidas de incerteza.
- Avaliar a possibilidade de uma agenda comum entre a Teoria de Probabilidades e a Teoria de Conjuntos Difusos.
- Apresentar o modelo GoM modificado para a realização de meta-análises.
- Construir conjuntos difusos que representem as variáveis: "p-value", tamanho da amostra e "effect size".
- Apresentar método de determinação de probabilidades subjetivas, para a construção do modelo proposto.
- Determinar um método de cálculo de probabilidade de evento difuso.
- Discutir métodos de aglutinação de resultados de eventos difusos, decorrentes da análise individual de estudos feitos para a meta-análise
- Fazer uma aplicação do modelo proposto a um conjunto clássico de dados em meta-análise.

### 1.3 Justificativa e Importância

Para a produção de conhecimento científico ser cumulativa, um passo intermediário entre as pesquisas passadas e futuras deve ser realizado. Esse passo é a realização de uma síntese das evidências existentes. Por muito tempo a revisão de literatura tem sido uma rotina na preparação e apresentação de um novo estudo. Tradicionalmente, esses esforços para acumular informação têm sido realizados de forma não sistemática. E essa revisão informal freqüentemente apresenta alguns problemas. LIGHT AND PILLEMER (1984) apresentam três falhas nesse procedimento. (i) A revisão tradicional é unicamente subjetiva. Uma vez que o processo tem poucas regras formais, dois revisores podem discordar de questões tão básicas como que estudos incluir; (ii) A revisão tradicional é cientificamente problemática. Sem regras básicas formais, um pesquisador pode tirar conclusões usando métodos inconsistentes ainda que utilizando boa prática estatística, e (iii) A revisão

tradicional é um meio ineficiente para se extrair informações úteis. Isso é especialmente verdadeiro quando o número de pesquisas é grande.

Os métodos meta-analíticos buscam encontrar formas de se realizar síntese de evidências com a utilização das ferramentas estatísticas disponíveis, em substituição ao processo tradicional informal de realização de revisão de literatura. No entanto alguns princípios básicos devem nortear a realização de uma meta-análise: (i) qualquer estratégia de revisão deve vir acompanhada de questões precisas; (ii) desacordos entre achados têm seu valor e devem ser explorados; (iii) informações quantitativas e qualitativas têm papel fundamental em um boa síntese e (iv) a precisão estatística não pode substituir a clareza conceitual. Dessa forma o processo de aglutinação de informações não pode ser realizado simplesmente de forma objetiva.

O presente trabalho apresenta uma metodologia intermediária entre a revisão de literatura narrativa e informal e os métodos estatísticos "puros" de realização de meta-análise. A incorporação da experiência e do conhecimento do especialista na análise matemática de achados em diferentes pesquisas deve ser realizada com a utilização de metodologia que permita que essa incorporação seja feita de forma não informal. A Teoria de Conjuntos Difusos, por conseguir tratar com variáveis com valores lingüísticos pode ajudar na construção de tal metodologia.

A importância da construção de tal metodologia se dá pelo fato de a informação a ser obtida através de estudos anteriores dever ter incorporado a ela o conhecimento subjetivo que o especialista tem do tema a ser tratado. As várias propostas de trabalho com estatística para dados não precisos, não elaboraram uma metodologia para meta-análise. Alguns dos principais desenvolvimentos em métodos estatísticos para dados não precisos podem ser encontrados nas seguintes referências: DEMPSTER(1967), SHAFER (1976) DUBOIS AND PRADE (1986), DUMOUCHEL AND HARRIS(1983), DUMOUCHEL (1990), KLIR (1995), KRUSE AND MEYER (1987), LIANG AND SONG(1996), MANTON ET AL(1994), NGUYEN (1977), PAN ET AL (1996), PIASECKI (1985), ROMER AND KANDEL (1995 a) (1995 b) (1995 c) e (1995 d), SCHNATTER(1993), TRIBUS (1980), VIERTL (1987) e (1995), WALLEY (1991), YAGER (1980) e (1986) e ZADEH (1968) e (1984).

## 1.4 Limitações

As discussões das medidas de ignorância estão restritas a medidas de incerteza, uma abordagem mais completa implicaria na extensão da discussão a medidas de distorção (confusão e falta de acurácia), irrelevância (falta de topicalidade, tabu e indecidibilidade) e incompletude (ausência). Limitou-se a discussão às medidas de incompletude ligadas à incerteza (ambigüidade, aleatoriedade e vaguesa, e dentro de vaguesa, não especificidade e difusividade).

Dentro do contexto do trabalho, limitou-se a discussão do relacionamento e oposições entre Teoria de Probabilidades, Teoria Bayesiana, Teoria de Probabilidade Subjetiva e Teoria de Conjuntos Difusos aos elementos centrais dessas teorias. O aprofundamento dessa discussão pode diminuir a distância entre o uso dessas teorias, com a possibilidade da construção de agendas comuns. A opção por uma abordagem mais simplificada se deve ao fato dessa discussão estar ainda em aberto, havendo ainda muitas controvérsias a serem resolvidas. Não houve portanto a intenção de esgotar essa discussão, mas simplesmente pontear certos aspectos fundamentais que permitissem a construção de uma metodologia com a utilização de elementos dessas várias teorias.

Os métodos meta-analíticos, mesmo os mais clássicos, não têm ainda uma forma definitiva e são pontos de debates constantes nas comunidades científicas que os utilizam. Alguns problemas colocados não encontraram ainda solução definitiva. A questão conhecida como "file drawer problem", que diz respeito à seleção de artigos a participarem da amostra se constitui um dos problemas centrais da meta-análise. Considera-se para efeito da construção das sínteses de evidências, que cada pesquisa é uma unidade amostral de uma amostra aleatória. De fato essa suposição é muito forte e precisaria ser corrigida na consecução das análises. Esses e outros problemas foram discutidos no corpo do trabalho, mas não considerados de forma definitiva no modelo proposto.

A estimação dos parâmetros do modelo GoM se dá através do método de máxima verossimilhança. Esse método além de ser restritivo para a aplicação proposta, não incorpora de forma direta o componente do raciocínio e da experiência humana no modelo. A utilização de conjuntos difusos e probabilidades da partição difusa do espaço amostral

através do cálculo de probabilidades subjetivas deve ser aprofundada para se evitar diferenças muito marcantes entre possíveis diferentes modelos difusos para meta-análise.

O conjunto de estudos apresentados na aplicação é ainda muito restrito, embora tenha sido utilizado para expor diferentes modelos de meta-análise. Ainda assim optou-se por utilizar esse conjunto como forma de possibilitar a comparação entre os diferentes métodos. Seria útil, a incorporação de outros conjuntos para ilustração do modelo proposto.

Um procedimento de meta-análise deve se dar em dois níveis: o quantitativo e o qualitativo ou substantivo. A interpretação dos resultados, muito especialmente, para o segundo tipo de análise, deve se dar por uma equipe multidisciplinar. Como a proposta é a construção de uma metodologia, a extensão dos resultados da aplicação fica restrita mais a uma análise estatística sem o forte componente clínico ou médico que o próprio modelo exige, desde a construção dos conjuntos difusos propostos.

Os métodos meta-analíticos são gerais o suficiente para serem aplicados a várias áreas do conhecimento. O modelo proposto é específico para estudos em bio-estatística. Algumas adaptações seriam necessárias para o seu uso em outras áreas.

## 1.5 Estrutura do Trabalho

O modelo de partição difusa não teve ainda seu desenvolvimento pleno para diversas aplicações estatísticas usando conjuntos difusos. Esse trabalho propõe essa aplicação em meta-análise. Foi necessária uma discussão do relacionamento entre várias formas de se trabalhar com incerteza e a mudança paradigmática realizada do final do século passado para os dias atuais, com a incorporação da incerteza nos modelos matemáticos. Inicialmente a incerteza de caráter aleatório e a partir da segunda metade desse século pela análise de modelos levando em conta outras medidas de incerteza, que caracterizam a segunda fase dessa mudança de paradigma. Essa discussão é realizada na introdução do capítulo 2 (A Natureza das Medidas de Incerteza e Ignorância). Nesse capítulo são expostas de forma sucinta diversas teorias que tratam do assunto. Em particular, de forma mais aprofundada, são discutidos os elementos básicos da Teoria de



Conjuntos Difusos e a Teoria de Probabilidades Subjetivas, que são utilizadas na construção do modelo proposto.

No capítulo 3 apresenta-se os princípios básicos de Meta-análise e um pouco do seu desenvolvimento histórico. São expostos o método clássico e o método bayesiano, a meta-análise exploratória e a analítica, os métodos de efeito fixo e de efeito aleatório. Apresenta-se uma medida padronizada para comparação e síntese entre as várias pesquisas, o chamado "effect size". E para a consecução do modelo proposto, discute-se a importância do tamanho da amostra e do nível de significância obtido em cada trabalho. Verifica-se porque e como essas medidas podem ser compreendidas com eventos difusos. E alguns problemas que aparecem na execução dos métodos meta-analíticos são apresentados.

O modelo geral de partição difusa é apresentado no capítulo 4. Para tanto discute-se a elaboração da partição difusa do espaço amostral e a caracterização do modelo. Uma discussão da parametrização do modelo e da forma proposta para a estimação dos parâmetros é realizada. Por fim propõe-se uma adaptação desse modelo à realização da meta-análise considerando-se como variáveis difusas independentes o tamanho da amostra, o nível de significância e o "effect size". Como resposta calcula-se a probabilidade de relevância de cada estudo como ponto inicial para o procedimento meta-analítico e a aglutinação dessas medidas para a finalização da meta-análise.

Uma aplicação é realizada no capítulo 5. O conjunto de dados escolhidos para a aplicação é o da meta-análise de seis estudos multicêntricos aleatorizados conduzidos na Europa e nos Estados Unidos. Esse conjunto de dados foi utilizado em diversos estudos de métodos meta-analíticos, CANNER (1983), CANNER (1987), GOODMAN (1989), GAVER (1992), MORRIS AND NORMAND (1992) e NORMAND (1995). Esse conjunto de dados foi escolhido para se poder fazer comparações dos métodos empregados com o método em proposição.

Finalmente são apresentadas conclusões e recomendações para estudos futuros.

## 2. A Natureza das Medidas de Incerteza e Ignorância

### 2.1 Introdução

#### *Complexidade, incerteza e utilidade*

Quando se tenta modelar um sistema, tem-se que considerar três características essenciais dele: complexidade, utilidade e incerteza. Existe uma forte relação entre essas três características. Sob o ponto de vista da incerteza, pode-se falar que, em geral, se for possível trabalhar com maiores níveis de incerteza, o modelo que descreve o sistema poderia ter maior credibilidade e menor complexidade.

Quando se modela um sistema é imposta uma negociação entre essas três características. Se o nível de incerteza permitido for grande, o sistema seria possivelmente mais manipulável e útil. Depois de uma pesquisa clínica por exemplo, poder-se-ia encontrar que o tempo esperado de vida de um grupo após uma dada intervenção cirúrgica seria de dez anos mais ou menos dois anos, com 95% de confiança. Essa declaração tem mais informação útil do que quando se diz que com 99.9% de confiança, aquele grupo tem expectativa de vida de dez anos mais ou menos nove anos. Nesse caso o aumento da incerteza fornece mais credibilidade para o resultado encontrado. Considerando-se um outro exemplo com relação à complexidade. Supondo um modelo de regressão múltipla para determinar as fontes da diabete gestacional, procura-se um número limitado de variáveis para a construção do modelo, atribuindo-se mais variáveis menos importantes ao erro estatístico o que diminui possíveis problemas de multicolinearidade. Nesse caso o sistema tem sua complexidade diminuída e como custo o aumento da incerteza.

Procura-se na modelagem de sistemas desenvolver-se métodos através dos quais um nível ótimo de incerteza permitida possa ser estimada para cada problema.

#### *Mudança de paradigma*

Esses tratamentos com incerteza e outros têm sido tratados tradicionalmente no campo da Estatística. O uso de métodos estatísticos em Ciência não é casual. A

percepção da necessidade de lidar com algum nível de incerteza foi discutido por JACOB (1973) em seu livro "The Logic of Life - A History of Heredity". Ele mostrou como a termodinâmica restabelece os vínculos entre a biologia e a física. "Primeiro, porque com os conceitos de energia e sua conservação, uma das peculiaridades do mundo vivo desaparece. Segundo, porque procurando relacionar as propriedades dos corpos com sua estrutura interna, a mecânica estatística muda a forma de se ver os objetos, organismos e ainda os eventos da vida diária. O estudo de grandes populações e a introdução do método estatístico para suas análises teve ainda grandes conseqüências para a forma como a biologia e outras ciências consideram os seres e as coisas." O autor constata que durante a segunda metade do século XIX, o papel e o status da análise estatística e da teoria de probabilidades mudou. No começo, para Maxwell eles eram simplesmente ferramentas que permitiam o trabalho com grandes quantidades de dados, uma vez que era impossível considerar cada indivíduo. No entanto, para Gibbs, Boltzmann e Darwin, o método estatístico e a teoria de probabilidades eram fundamentais não importando o comportamento de um único indivíduo. "Não existe nada que um indivíduo nos conte, que a população inteira não faça melhor."

De acordo com KLIR AND YUAN (1995), existem dois estágios da transição da forma tradicional de tratar a incerteza para a moderna. O primeiro estágio começou no século XIX com a introdução de métodos estatísticos e a teoria de probabilidades no estudo das moléculas. O uso dos métodos estatísticos ultrapassaram os limites dos processos moleculares (mecânica estatística) e foram estendidos para quase todos os ramos da ciência. Eles ajudaram a providenciar o intercâmbio entre as ciências, e se tornaram importantes ferramentas para diminuir a distância que separava alguns campos do pensamento humano.

Desde Newton, os físicos baseavam-se em um rígido determinismo. O novo paradigma nasceu com a necessidade de incorporar incerteza nos sistemas com modelos mais complexos. Segundo MCNEILL AND FREIBERGER (1993) a incerteza é a matéria-prima da vida. E é ambíguo que por séculos, os estudiosos tenham procurado uma forma de escapar da incerteza, ao invés de procurar uma forma de trabalhar com ela. Esses têm preferido trabalhar com o consolo da certeza do que com a força da incerteza.

O segundo estágio da transição do ponto de vista tradicional para a moderna forma de se tratar a incerteza foi inaugurado explicitamente na literatura dos anos 60. Nesses anos varias teorias sobre incerteza que não a Teoria de Probabilidades foram desenvolvidas. Estas teorias tiveram como objetivo trabalhar com outros tipos de incertezas não tratadas pela Estatística.

O conceito de *mudança paradigmática* foi exposto por KUHN (1962). Fundamentalmente quando uma ciência começa a ter problemas difíceis ou impossíveis de serem tratados por padrões mais tradicionais, surge a crise com o aparecimento de paradoxos, e uma nova forma se impõe. A fase da mudança iniciada com o conceito de conjuntos difusos e a idéia da matemática baseada em conjuntos difusos emergiu da necessidade de construir uma ponte entre modelos matemáticos e suas interpretações empíricas. BLACK (1937) caracterizou essa necessidade, discutindo a questão de que enquanto os matemáticos constróem uma teoria em termos de objetos "perfeitos", as ciências experimentais observam objetos cujas propriedades são somente aproximadamente verdadeiras. Nesse sentido, PRIGOGINE (1997) em seu livro "The end of Certainties", apresenta os pontos fundamentais e um pouco da história da chamada "crise da matemática". O novo paradigma no trato com incerteza, proposto inicialmente por ZADEH (1965), com a apresentação dos conjuntos difusos desafiou o mais sagrado elemento dos fundamentos do trato com incerteza - a lógica bi-valorada de Aristóteles, introduzindo uma teoria cujos objetos são conjuntos com limites que não são precisos. A capacidade dos conjuntos difusos expressarem transições graduais de pertinência para não-pertinência e vice-versa tem uma grande utilidade. Permite não só uma representação poderosa das medidas de incerteza, como também uma representação com significado claro dos conceitos vagos expressos em linguagem natural. Dessa forma, um conjunto difuso pode ser definido matematicamente pelo estabelecimento de um valor que represente o grau de pertinência de um indivíduo no universo de discurso.

Essa mudança permite fundamentalmente que: (i) se expresse observações irreduzíveis e medidas de incerteza em suas várias manifestações e faz essas incertezas intrínsecas aos dados, os resultados obtidos são mais carregados de significado, em termos epistemológico e pragmático; (ii) o novo paradigma tem fortes recursos para gerenciar complexidades e controlar custos computacionais; (iii) tem maior poder de expressividade,

e portanto, pode tratar de uma maior classe de problemas, notadamente aqueles que envolvem a utilização de linguagem natural e (iv) tem maior capacidade de capturar a forma de raciocínio humano do senso-comum, que quando empregado em computação permite a construção de máquinas e programas mais amigáveis.

#### *Natureza e tipos de medidas de incerteza*

SMITHSON (1988) discute incerteza no contexto da *ignorância*. Segundo o autor a ignorância tem sido tratada usualmente como ausência ou distorção do conhecimento verdadeiro, e incerteza como alguma forma de incompletude na informação ou no conhecimento. De alguma forma essas concepções do senso-comum são razoáveis, mas elas podem desviar a atenção de um conceito mais amplo de ignorância por defini-la indiretamente como não-conhecimento. E essa concepção leva a eliminação ou absorção da ignorância pelo exercício de alguma versão do "método científico". Uma simples definição de ignorância é insuficiente para contextualizá-la. No entanto o estudo de suas diversas manifestações podem deixar mais clara a forma de se lidar com ignorância, desde que se possa associar a cada uma de suas componentes, ferramentas para analisar o contexto em que ela se insere no estudo dos fenômenos do mundo real. O autor propõe a seguinte taxonomia para ignorância, apresentada na figura 2.1.1, que pode auxiliar na compreensão das diferentes naturezas das medidas de incerteza.

Estudar esses diferentes tipos de ignorância é tão importante quanto estudar os diferentes níveis de ignorância, mesmo porque a quantificação da ignorância - principalmente para termos comparativos - deve ser realizada dentro do mesmo tipo de ignorância.

Uma distinção que se faz no senso-comum é entre *ignorar* e *ser ignorante*. Ignorar, na voz ativa, denota *passar por cima* ou ainda ter deliberada falta de atenção para alguma coisa. Ser ignorante de algo, na voz passiva, por outro lado, denota conhecimento incompleto ou distorcido. Essa dualidade é uma primeira ramificação na tipologia da ignorância. SMITHSON (1988) associa o ato de ignorar como uma declaração de *irrelevância*. O estado de ignorância, por outro lado, é um estado errado de cognição e é chamado de *erro*.

A *irrelevância* pode ser manifestada de três formas: *não-topicalidade*, *tabu* e *não-decidibilidade*.

*Não-topicalidade* se refere às intuições que as pessoas carregam e negociam com outras sobre como seu domínios cognitivos se ajustam em comum. Consistência tópica é, por exemplo, uma regra não falada que guia uma conversação ordinária. Nas manifestações folclóricas existem conexões entre assuntos que parecem ser irrelevantes para a maioria das pessoas. *Tabu* é literalmente o que as pessoas não devem saber ou ainda inquirir sobre. Os tabus funcionam como guardiões da pureza e da salvaguarda através de regras socialmente sancionadas de (i)relevância. E *não-decidibilidade* é atribuída àquelas matérias que as pessoas são incapazes de designar falsa ou verdadeira porque consideram o problema insolúvel ou porque a questão da validade ou verificabilidade não é pertinente. Certos tipos de fantasias ou ficção têm essa última qualidade, como também pensamentos que são considerados não-essenciais ou sem significado.

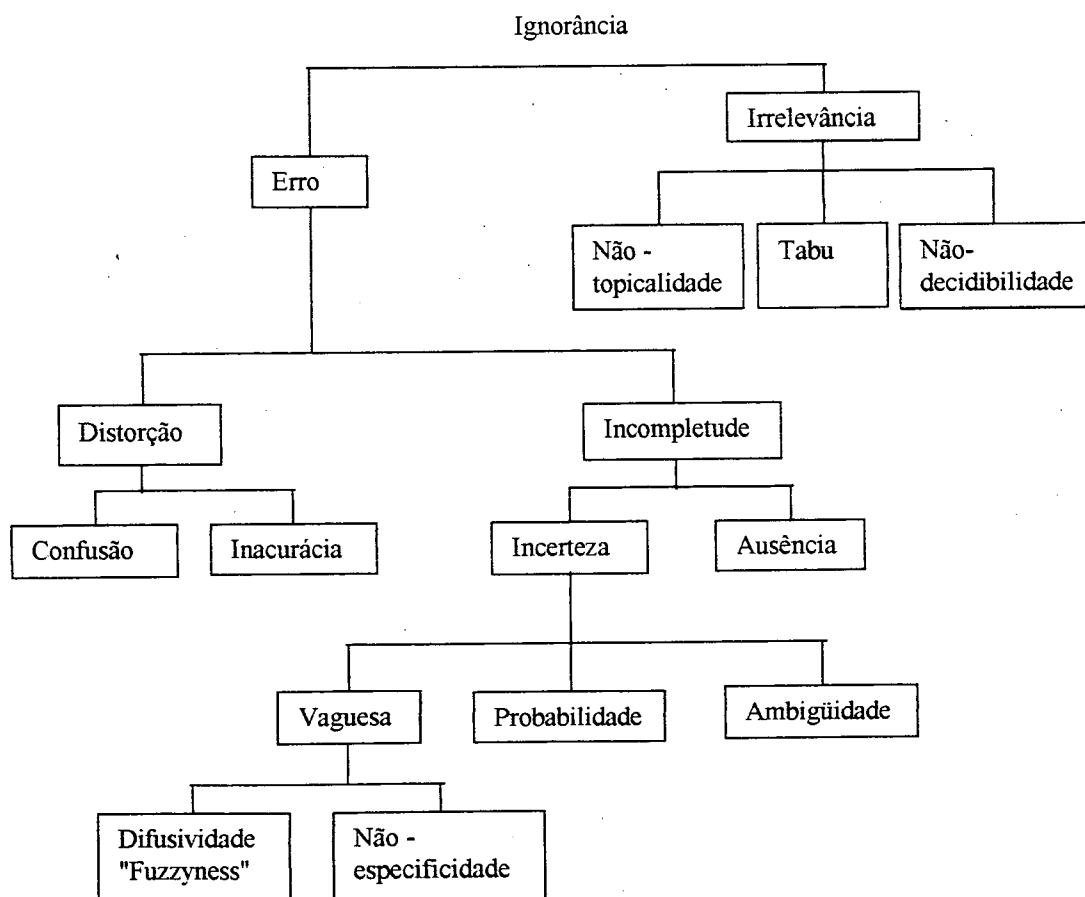


Figura 2.1.1 Taxinomia da Ignorância

A *irrelevância* é uma estratégia *negativa* para lidar com anomalias, que as ignora ou não as percebe, declarando-as irrelevantes e as banindo da realidade. A estratégia positiva, o *erro*, é revisar a estrutura da realidade, mesmo que somente através de uma declaração de ignorância. A primeira estratégia resulta em exclusão, a segunda termina em inclusão.

O *erro* pode advir de uma visão distorcida ou de uma visão incompleta (ou de ambas). *Distorção* usualmente se refere em termos de "vício" ou "falta de acurácia", de um lado, e "confusão" do outro. A primeira se refere a uma distorção de grau, enquanto a segunda indica uma substituição errada de tipo. *Incompletude* por outro lado, tem recebido considerável atenção, muito possivelmente porque ela parece ser mais corrigível do que distorção. Também a incompletude pode ser de tipo e que é chamada de *ausência*, enquanto incompletude em grau é chamada de *incerteza*. Incerteza, por sua vez, inclui os conceitos de *probabilidade*, *vaguesa* e *ambigüidade*. E finalmente a vaguesa pode ser compreendida em dois ramos, a *difusividade* ("fuzziness") e a *não-especificidade*.

KLIR AND FOLGER (1988) apresentam incertezas de quatro tipos. Eles acreditam que a probabilidade possa ser útil mas não uma "panacéia". Segundo os autores "não se deve tomar antibiótico para problemas de pressão sanguínea alta". Os quatro tipos básicos de incerteza propostos são: não-especificidade, *difusividade* ("fuzzyness"), dissonância e confusão. Para os autores não-especificidade é ambigüidade básica, uma relação entre declarações e possíveis significados. Um teste pode mostrar um paciente que tem hepatite, cirrose, pedras na vesícula ou câncer pancreático. Entre elas não se pode dizer nada. "difusividade" é vaguesa. O grau para o qual um termo como cirrose se aplica. Dissonância é conflito puro. Uma declaração é verdadeira e suas rivais são falsas. Um paciente deve sofrer um mal do fígado ou não. Algumas evidências suportam uma tese, outras evidências suportam a outra, e não se tem certeza entre as duas. Confusão é conflito puro e conflito potencial. Supor que um teste diga se um paciente tem doença no fígado e um segundo que tem doença no estômago, mas que para algumas doenças ambos os testes são positivos. Nesse caso, um teste de fígado positivo pode indicar doença no estômago. Existe não somente conflito, mas o significado da evidência é não-nuclear.

A probabilidade não pode tratar com todas essas espécies de incerteza. KLIR AND FOLGER (1988) apontam que cada um das teorias de incerteza trabalha com

diferentes setores dela: A Teoria de Conjuntos trabalha somente com não-especificidade. Qualquer que seja a doença - hepatite, cirrose, pedras no esôfago ou câncer pancreático - ela estará em um conjunto das quatro doenças. A Teoria de Probabilidade trata com dissonância. Qual das quatro doenças é a doença? A Teoria de Conjuntos Difusos trabalha com não-especificidade e nebulosidade. Quão extensa é a cirrose? A Teoria de Possibilidades trabalha com não-especificidade e confusão. Como é possível que um teste positivo do fígado, indique doença no estômago? A Teoria de Possibilidades pode medir confusão porque ela verifica quão facilmente um evento pode ocorrer, seu potencial. E a Teoria de Dempster-Shafer trata de não-especificidade, dissonância e confusão. A Teoria de Probabilidades trata só com graus de conflito porque ela foi construída com esse propósito.

## 2.2 Probabilidade Subjetiva

Uma teoria subjetiva de probabilidades define probabilidade em termos de graus de crença, que é de alguma forma uma descendente da mais antiga noção filosófica de credibilidade. As motivações dessa definição são de três ordens: (i) qualquer uso prático da teoria de probabilidades requer julgamentos subjetivos, (ii) é possível representar graus de crença em termos de preferências subjetivas com condições suficientemente fortes que possibilitem o uso de probabilidades quantitativas (e ainda o cálculo de probabilidades) e (iii) esse enfoque assegura que se forem fornecidas as regras básicas racionais aos tomadores de decisão eles se comportam racionalmente e de acordo com suas preferências subjetivas.

As principais definições apresentadas nessa seção são um estrato da apresentação completa da Teoria de Probabilidades Subjetivas em DEGROOT (1970). As provas dos teoremas apresentados podem ser encontradas nessa bibliografia.

Em muitos problemas, o assinalamento de uma distribuição de probabilidades sobre a  $\sigma$ -álgebra de eventos tem se tornado rotineira e padrão entre pesquisadores de áreas específicas. Esse assinalamento é usualmente baseado na combinação da tradição e da experiência. Em alguns outros problemas, existem modelos naturais de probabilidade que representam as situações. Então, probabilidades disponíveis podem frequentemente ser



assinaladas objetiva e rapidamente por conta da forte concordância sobre o quão apropriada é uma distribuição específica.

Por outro lado, existem certas situações para as quais pode ser muito difícil encontrar pessoas que concordem se uma específica distribuição de probabilidades é apropriada para descrever um certo fenômeno. Nessas situações, o assinalamento de probabilidades do estatístico pode ser altamente subjetivo e deve refletir suas próprias informações e crenças. Existem certas condições sob as quais o estatístico pode representar suas informações e crenças em termos de distribuição de probabilidades. Define-se então verossimilhança relativa.

### *Verossimilhança Relativa*

Seja  $\Omega$  um espaço amostral definido na  $\sigma$ -álgebra de eventos,  $\wp$ . Um conceito fundamental é que um evento possa ser mais verossímil de ocorrer que outro. Então será considerado que uma pessoa, quando considera dois eventos em  $\wp$ , pode decidir se um deles pode ser considerado mais verossímil que outro, ou se são igualmente verossímeis de ocorrer.

Quando dois eventos  $A$  e  $B$  são comparados,  $A < B$  indicará que  $B$  é mais verossímil de ocorrer que  $A$ , e  $A \sim B$  indica que os eventos são igualmente verossímeis. Em analogia com a relação de desigualdade entre dois números  $A < B$  indicará que  $B$  é pelo menos tão verossímil quanto  $A$ , ou equivalentemente que  $A$  não é mais verossímil de ocorrer que  $B$ . Então  $A < B$  é ao mesmo tempo  $A < B$  e  $A \sim B$ . Finalmente  $A > B$  tem o mesmo significado que  $B < A$ , e define-se  $A \succ B$  para significar o mesmo que  $B < A$ .

Uma vez que a probabilidade de um evento é supostamente uma medida numérica da verossimilhança que um evento ocorra, qualquer distribuição de probabilidades  $P$  que é assinalada a um evento na  $\sigma$ -álgebra de eventos  $\wp$ , deve ter a propriedade que  $P(A) \leq P(B)$  se e somente se  $A < B$ . A distribuição de probabilidades  $P$  que tem essa propriedade concorda com a relação  $<$ .

As suposições que são feitas tendo em vista a relação  $\prec\sim$  são sugeridas pelo uso intuitivo do conceito de verossimilhança relativa de ocorrência de dois eventos e pelas propriedades matemáticas das distribuições de probabilidades.

As suposições básicas são as que seguem:

**Suposição SP1** - Para quaisquer dois eventos A e B, exatamente um das seguintes relações acontece:  $A \prec B$ ,  $A \succ B$  e  $A \sim B$ .

A próxima suposição tem uma interpretação simples que a faz intuitivamente plausível:

**Suposição SP2** - Se  $A_1, A_2, B_1$  e  $B_2$  são quatro eventos tais que  $A_1 \cap A_2 = B_1 \cap B_2 = \emptyset$  e  $A_i \prec\sim B_i$  para  $i = 1, 2$ , então  $A_1 \cup A_2 \prec\sim B_1 \cup B_2$ . E ainda se  $A_1 \prec B_1$  e  $A_2 \prec B_2$  então  $A_1 \cup A_2 \prec B_1 \cup B_2$ .

Essa suposição pode ser interpretada da seguinte forma: suponha que cada um dos eventos A e B possam ocorrer de duas formas mutuamente excludentes. Se cada forma que leva a A não for mais verossímil de ocorrer do que a forma que leva a B, então  $A \prec\sim B$ . E mais, se pelo menos uma das formas que leva A a ser realmente menos verossímil que a correspondente forma que leva a B, então  $A \prec B$ .

Vários resultados podem ser derivados dessas duas suposições elementares. Um resultado interessante e importante é a transitividade da relação  $\prec\sim$ , mostrada no Teorema 2.2.1 abaixo. Antes de apresentar o teorema, deve-se apresentar o seguinte lema:

**Lema 2.2.1** - Suponha que A, B e D são eventos tais que  $A \cap D = B \cap D = \emptyset$ . Então  $A \prec\sim B$ , se e somente se  $A \cup B \succ B \cup D$ .

**Teorema 2.2.1** - Se A, B e D são eventos tais que  $A \prec\sim B$  e  $B \prec\sim D$ , então  $A \prec\sim D$ .

Segue de SP1 e da propriedade de transitividade exibida no Teorema 1, que a relação  $\prec\sim$  produz uma ordenação completa do eventos em  $\mathcal{E}$ . Nessa ordenação é possível que dois eventos distintos A e B sejam equivalentes, isto é,  $A \sim B$ .

O próximo resultado é uma extensão da Suposição SP2 para uniões de qualquer número finito de eventos disjuntos.

**Teorema 2.2.2** - Se  $A_1, \dots, A_n$ , são  $n$  eventos disjuntos e  $B_1, \dots, B_n$  são também eventos disjuntos tais que  $A_i \prec \sim B_i$  para  $i = 1, \dots, n$  então  $A_1 \cup, \dots, \cup A_n \prec \sim B_1 \cup, \dots, \cup B_n$ . E ainda se  $A_i \prec B_i$  para pelo menos um valor de  $i$  ( $i = 1, \dots, n$ ) então  $A_1 \cup, \dots, \cup A_n \prec B_1 \cup, \dots, \cup B_n$ .

O próximo teorema reflete uma propriedade básica da verossimilhança relativa que também é observada para probabilidades.

**Teorema 2.2.3** - Para quaisquer eventos  $A$  e  $B$ ,  $A \prec \sim B$  se e somente se  $A^c \succ \sim B^c$ .

A próxima suposição é a de que nenhum evento é menos verossímil que o conjunto vazio  $\phi$ , e evita-se uma situação completamente trivial pelo suposição que o espaço amostral  $\Omega$ , é realmente mais verossímil que  $\phi$ .

**Suposição SP3** - Se  $A$  é um evento, então  $\phi \prec \sim A$ . E ainda  $\phi \prec \Omega$ .

Essa suposição leva à seguinte propriedade fundamental de verossimilhança relativa, que também é observada para probabilidades.

**Teorema 2.2.4** - Se  $A$  e  $B$  são eventos tais que  $A \subset B$ , então  $A \prec \sim B$ . Em particular se  $A$  é qualquer evento, então  $\phi \prec \sim A \prec \sim B$ .

A seguinte suposição tem a propriedade especial de envolver uma seqüência infinita de eventos.

**Suposição SP4** - Se  $A_1 \supset A_2 \supset \dots$  é um seqüência decrescente de eventos e  $B$  é algum evento fixo tal que  $A_i \succ \sim B$  para  $i = 1, 2, \dots$ , então  $\bigcap_{i=1}^{\infty} A_i \succ \sim B$ .

Para ajudar a deixar mais claro o significado da suposição SP4, pode-se considerar um exemplo no qual os eventos sob consideração são subconjuntos da reta. Supondo que cada intervalo infinito da forma  $(n, \infty)$ , para  $n = 1, 2, \dots$ , é visto como mais verossímil que algum subconjunto pequeno e fixo  $B$  na reta. Uma vez que a interseção de todos os pequenos intervalos infinitos é o conjunto vazio  $\phi$ , segue então de SP4 que o menor subconjunto  $B$  deve também ser equivalente a  $\phi$ . Em outras palavras, se  $B$  é qualquer

conjunto tal que  $B \supset \phi$ , então tendo em vista o quão pequeno é o conjunto  $B$ , é impossível para qualquer intervalo infinito  $(n, \infty)$  ser pelo menos tão verossímil como  $B$ .

O teorema seguinte é um dual da suposição SP4.

**Teorema 2.2.5** - Se  $A_1 \subset A_2 \subset \dots$  é uma seqüência crescente de eventos e  $B$  é algum evento

fixo tal que  $A_i \prec \sim B$  para  $i = 1, 2, \dots$ , então  $\bigcup_{i=1}^{\infty} A_i \prec \sim B$ .

O próximo resultado estende o teorema 2 a uniões de seqüências infinitas eventos disjuntos.

**Teorema 2.2.6** - Se  $A_1, A_2, \dots$  é uma seqüência infinita de eventos disjuntos e  $B_1, B_2, \dots$  é outra seqüência infinita de eventos disjuntos tal que  $A_i \prec \sim B_i$  para  $i = 1, 2, \dots$ , então

$\bigcup_{i=1}^{\infty} A_i \prec \sim \bigcup_{i=1}^{\infty} B_i$ . Se em adição  $A_i \prec B_i$  para pelo menos um valor de  $i$  ( $i = 1, 2, \dots$ ), então

$$\bigcup_{i=1}^{\infty} A_i \prec \bigcup_{i=1}^{\infty} B_i.$$

Uma Quinta suposição é necessária com base na qual é possível construir uma distribuição de probabilidades única que concorda com a relação  $\prec \sim$ .

**Suposição SP5** - Existe uma variável aleatória que tem uma distribuição uniforme no intervalo  $[0, 1]$ .

Essa suposição perpassa todo o trabalho de construção de probabilidades subjetivas. Muitos experimentos têm somente um número finito de resultados, e eles não necessariamente são igualmente verossímeis de ocorrerem. Assim não é possível definir no espaço amostral de tal experimento uma variável aleatória que tenha uma distribuição uniforme no intervalo  $[0, 1]$ . O estatístico deve então fazer crescer seu espaço amostral considerando um experimento auxiliar no qual o valor da variável aleatória tenha uma apropriada distribuição uniforme. Dessa forma, cada ponto no espaço amostral aumentado desse experimento composto compreende um resultado do experimento original junto com um resultado do experimento auxiliar. Assume-se que as quatro primeiras suposições são válidas para esse experimento composto.

Não é necessário que um experimento auxiliar seja realmente realizado. O estatístico necessita somente ser capaz de imaginar um experimento auxiliar ideal no qual um variável aleatória uniformemente distribuída pode ser gerada e ser capaz de comparar a verossimilhança relativa de algum evento  $A$  que era originalmente de interesse. Uma variável aleatória dessa forma pode ser construída como segue: Para  $n=1,2,\dots$ , seja  $T_n$  uma amostra aleatória tal que  $T_n = 1$  se o resultado da  $n$ -ésima jogada de uma moeda for cara e  $T_n = 0$  se o resultado for coroa. Seja  $X$  a variável aleatória definida pela equação  $X = \sum_{i=1}^{\infty} 2^{-i} T_i$ . Pode ser mostrado que  $X$  tem uma distribuição uniforme no intervalo  $[0,1]$ .

### *Construção da Distribuição de Probabilidades*

Supor que o espaço amostral  $\Omega$ , a  $\sigma$ -álgebra de eventos  $\mathcal{G}$  e a relação  $\prec\sim$  satisfazem as suposições SP1 a SP5. Sob essas suposições existe uma distribuição de probabilidades única  $P$  que concorda com a relação  $\prec\sim$ .

Pela suposição SP5, existe uma variável aleatória  $X$  que tem uma distribuição uniforme no intervalo  $[0,1]$ . Para qualquer subintervalo  $(a,b)$  do intervalo unitário  $[0,1]$ , seja  $G(a,b)$  o evento que a variável aleatória  $X$  cai no intervalo  $(a,b)$ . Então para quaisquer dois eventos  $(a_1,b_1)$  e  $(a_2,b_2)$  com  $0 \leq a_i \leq b_i \leq 1$  para  $i = 1,2$ , segue que  $G(a_1,b_1) \prec\sim G(a_2,b_2)$  se e somente se,  $b_1 - a_1 \leq b_2 - a_2$ .

**Teorema 2.2.7** - Se  $A$  é um evento qualquer, existe um único número  $a^*$  ( $0 \leq a^* \leq 1$ ) tal que  $A \sim G[0,a^*]$ .

A distribuição de probabilidades  $P$  pode agora ser definida diretamente desse teorema. Se  $A$  é um evento qualquer, então  $P(A)$  é definido como sendo o número  $a^*$  definido no teorema. Então  $P(A)$  pode ser determinado pela relação  $A \sim G[0,P(A)]$ .

O teorema seguinte garante para essa definição, que  $P$  concorda com a relação  $\prec\sim$ .

**Teorema 2.2.8** - Sejam  $A$  e  $B$  dois eventos quaisquer. Então  $A \prec\sim B$ , se e somente se,  $P(A) \leq P(B)$ .

Maiores detalhamentos, questionamentos e aplicações de probabilidades subjetivas podem ser encontradas nas referências que seguem: BASU (1975), BAYES (1763),

BERGER (1985), BERRY (1997), BIRNBAUM (1962), BORCH (1977), BOUCHON (1987), BOX (1973), DEFINETTI (1972), DEFINETTI (1974), DEFINETTI (1975), DEFINETTI (1977), DEHLING ET AL. (1996), FIENBERG (1992), GOOD (1988), HODGES (1996), MORRIS (1977), SMITHSON (1989), SOOFI (1996), VON MISSES (1957), WALLEY (1991) e WASSERMAN AND KADANE (1996)

### 2.3 Teoria de Conjuntos Difusos

As principais definições apresentadas nessa seção são um estrato da apresentação completa da teoria de conjuntos difusos em KLIR AND YUAN(1995).

#### *Apresentação*

A função característica de um conjunto "crisp" (conjunto "crisp" é compreendido como o conjunto da Teoria Clássica de Conjuntos) assinala valores 1 ou 0 para cada indivíduo no conjunto universo, definindo assim se o indivíduo é membro ou não do conjunto em consideração. Essa função pode ser generalizada de tal forma que os valores assinalados aos elementos do conjunto universo se encontrem dentro de uma amplitude específica e que indique o grau de pertinência desses elementos no conjunto em questão. Valores maiores denotam maiores graus de pertinência. Tal função é chamada de *grau de pertinência*, e o conjunto é definido como *conjunto difuso*.

A amplitude mais comumente utilizada para os valores das funções de pertinência é o intervalo  $[0,1]$ . Nesse caso, cada função de pertinência leva os elementos de um dado conjunto universo  $X$ , que é sempre um conjunto "crisp", a números reais em  $[0,1]$ .

Duas notações são mais usualmente empregadas na literatura para denotar funções de pertinência. Em uma delas, a função de pertinência de um conjunto difuso  $A$  é denotado por  $\mu_A$ ; isto é

$$\mu_A: X \rightarrow [0,1] \quad (2.3.1)$$

Na outra, a função é denotada por  $A$  e tem a mesma forma:

$$A: X \rightarrow [0,1] \quad (2.3.2)$$

No modelo GoM apresentado no Capítulo 4 a função de pertinência é apresentada como  $g_{ik}$  que denota o grau de pertinência do elemento "i" ao conjunto "k". Para uma apresentação mais simplificada e mais de acordo com a literatura, nesse capítulo é utilizada a notação apresentada em 2.3.2, sem perda de generalidade. Dessa forma  $A(x)$  com  $x \in X$ , representa o grau de pertinência ao conjunto A do elemento "x" do conjunto universo X.

Cada conjunto difuso é completa e unicamente definido por uma particular função de pertinência; conseqüentemente, símbolos e funções de pertinência podem também ser usadas como notação dos conjuntos difusos associados.

Os conjuntos difusos permitem a representação de conceitos vagos expressos em linguagem natural. A representação não depende somente do conceito, mas também do contexto em que ele é utilizado. "Alta temperatura" terá conjuntos de pertinência diferentes se estiverem sendo tratadas as temperaturas do corpo humano, de condições climáticas ou de um forno em uma siderúrgica, por exemplo.

Um conjunto difuso pode ser "triangular" ou "trapezoidal". O triângulo ou o trapézio são formas fundamentais que podem sofrer alterações, conforme é visto mais adiante.

Tomando como exemplo, a variação de temperaturas entre  $T_1$  e  $T_2$ , que podem assumir os valores lingüísticos, "muito baixa", "baixa", "média", "alta" e "muito alta", as representações desses conjuntos como variáveis "crisp" e variáveis difusas são apresentadas na figura a seguir:

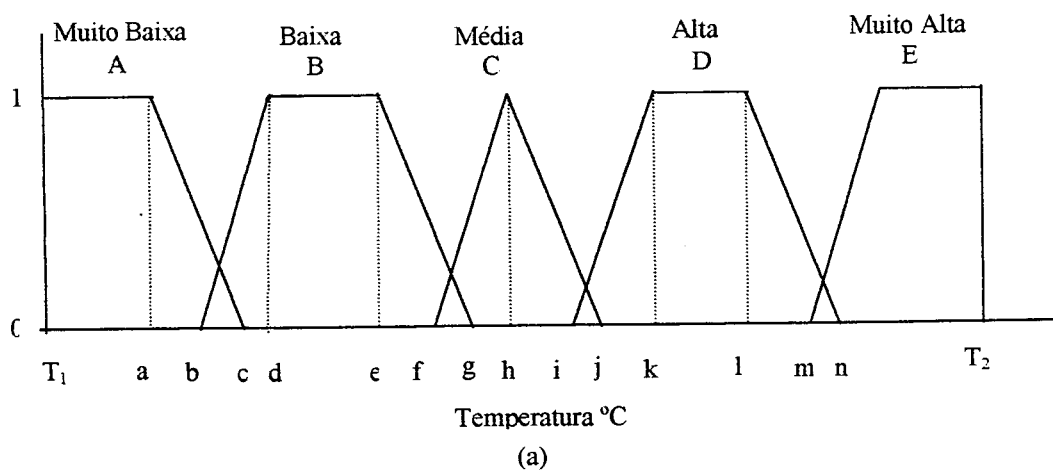


Figura 2.3.1 Temperatura em °C no intervalo  $[T_1, T_2]$  (a) variável difusa

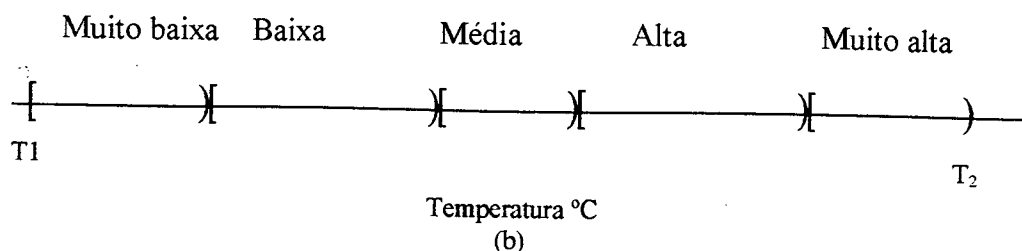


Figura 2.3.2 Temperatura em °C no intervalo  $[T_1, T_2]$  (b) variável "crisp"

A temperatura média foi apresentada na forma triangular para efeito de ilustração das duas formas básicas que os conjuntos difusos podem assumir. As expressões analíticas dos conjuntos difusos acima definidos são:

A - Temperatura muito baixa

$$\begin{aligned}
 A(x) &= 1 & \forall t, t \leq a \\
 A(x) &= \frac{c-t}{c-a} & \forall t, a \leq t \leq c
 \end{aligned} \tag{2.3.3}$$

B - Temperatura baixa

$$\begin{aligned}
 B(x) &= \frac{t-b}{d-b} & \forall t, b \leq t \leq d \\
 B(x) &= 1 & \forall t, d \leq t \leq e \\
 B(x) &= \frac{g-t}{g-e} & \forall t, e \leq t \leq g
 \end{aligned} \tag{2.3.4}$$

C - Temperatura média

$$\begin{aligned}
 C(x) &= \frac{t-f}{h-f} & \forall t, f \leq t \leq h \\
 C(x) &= \frac{j-t}{j-h} & \forall t, h \leq t \leq j
 \end{aligned} \tag{2.3.5}$$

Para as temperaturas "alta" e "muito alta" valem formas semelhantes às apresentadas para os conjuntos acima descritos.



A importância das variáveis difusas é que elas facilitam transições graduais entre os estados, e conseqüentemente, possuem uma capacidade natural de expressar e tratar com observações e medidas incertas. Então uma temperatura entre os valores "b" e "c", terá um certo grau pertinência para o conjunto "muito frio" e "frio". Poderia por exemplo, ter graus de pertinência 0.4 para "muito frio" e 0.5 para "frio". Variáveis tradicionais não tem essa capacidade. Embora a definição de estados para conjuntos "crisp" seja matematicamente correta, ela não é realística em face de erros de medidas inevitáveis. Uma medida que caia em uma vizinhança fechada para cada borda definida precisamente entre estados é tomada para um único estado, a despeito da inevitável incerteza envolvida nessa decisão.

Um interessante paradoxo é que os dados baseados em variáveis difusas fornecem, de fato, evidência mais acurada sobre o fenômeno real do que dados baseados em variáveis "crisp". Embora as matemáticas baseadas em conjuntos difusos tenham maior poder expressivo do que para conjuntos "crisp", sua utilidade depende criticamente na capacidade de se construir apropriadas funções de pertinência para os vários conceitos dados nos vários contextos.

### *Conceitos Básicos*

Um dos mais importantes conceitos em conjuntos difusos é o conceito de *corte- $\alpha$*  (" $\alpha$ -cuts") e seu variante *corte- $\alpha$  forte* ("strong  $\alpha$ -cuts"). Dado um conjunto difuso A definido em X e algum número  $\alpha \in [0,1]$ , o corte- $\alpha$ ,  ${}^\alpha A$ , e o corte- $\alpha$  forte,  ${}^{\alpha+} A$ , são os conjuntos "crisp":

$${}^\alpha A = \{x | A(x) \geq \alpha\} \quad (2.3.6)$$

$${}^{\alpha+} A = \{x | A(x) > \alpha\} \quad (2.3.7)$$

Isto é, o corte- $\alpha$  (ou o corte- $\alpha$  forte) de um conjunto difuso A é o conjunto  ${}^\alpha A$  (ou o conjunto difuso  ${}^{\alpha+} A$ ) que contém todos os elementos do conjunto universo X cujos graus de pertinência em A são maiores ou iguais (ou somente maiores) que um valor específico de  $\alpha$ .

O conjunto de todos os níveis  $\alpha \in [0,1]$  que representam distintos cortes- $\alpha$  de um dado conjunto difuso é chamado um *nível* de A. Formalmente:

$$\Lambda(A) = \{\alpha | A(x) = \alpha \text{ para algum } x \in X\} \quad (2.3.8)$$

onde  $\Lambda$  representa o nível do conjunto difuso  $A$  definido em  $X$ .

Uma importante propriedade do corte- $\alpha$  e do corte- $\alpha$  forte, que segue imediatamente de suas definições, é que a ordenação total dos valores de  $\alpha$  em  $[0,1]$  é inversamente preservada pelo conjunto inclusão dos correspondentes cortes- $\alpha$  como também dos cortes- $\alpha$  fortes. Uma consequência dessa propriedade é que todos os cortes- $\alpha$  para qualquer conjunto difuso formam uma família de conjuntos "crisp" aninhados.

O suporte de um conjunto difuso  $A$  em um conjunto universo  $X$  é um conjunto "crisp" que contém todos os elementos de  $X$  que têm graus de pertinência em  $A$  diferentes de zero. O suporte pode ser representado por  $^{0+}A$ . O corte-1,  $^1A$ , é chamado de *cerne* de  $A$ .

A altura,  $h(A)$ , de um conjunto difuso  $A$  é o maior grau de pertinência obtido por qualquer elemento naquele conjunto. Formalmente

$$h(A) = \sup A(x), \text{ para } x \in X \quad (2.3.9)$$

Um conjunto difuso é chamado *normal* quando  $h(A) = 1$ , caso contrário é chamado de *sub-normal*.

Uma importante propriedade de conjuntos difusos é a sua convexidade. Essa propriedade é vista como uma generalização do conceito clássico de convexidade para conjuntos difusos. Para fazer a convexidade generalizada consistente com a definição clássica de convexidade, é requerido que os cortes- $\alpha$  de um conjunto difuso convexo sejam convexos para todo  $\alpha \in [0,1]$ .

As três operações básicas em conjuntos "crisp" - o complemento, a interseção e a união - podem ser generalizadas para conjuntos difusos de mais de um maneira. No entanto, uma generalização particular, que resulta em operações que são usualmente chamadas de *operações padrões de conjuntos difusos*, tem um significado especial na Teoria de Conjuntos Difusos.

O complemento padrão,  $\bar{A}$ , de um conjunto difuso com respeito ao conjunto universo  $X$  é definido para todo  $x \in X$  pela equação:

$$\bar{A}(x) = 1 - A(x) \quad (2.3.10)$$

Elementos de  $X$  para os quais  $A(x) = \bar{A}(x)$  são chamados de *pontos de equilíbrio* de  $A$ . Para o complemento padrão os graus de pertinência dos pontos de equilíbrio são 0.5.

Dados dois conjuntos difusos,  $A$  e  $B$ , sua interseção padrão,  $A \cap B$ , e sua união padrão  $A \cup B$ , são definidas para todo  $x \in X$  pelas equações:

$$(A \cap B)(x) = \min [A(x), B(x)] \quad (2.3.11)$$

$$(A \cup B)(x) = \max[A(x), B(x)] \quad (2.3.12)$$

onde  $\min$  e  $\max$  denotam o operador de mínimo e o operador de máximo, respectivamente. Dada a associatividade de  $\min$  e de  $\max$ , essas definições podem ser estendidas para qualquer número finito de conjuntos difusos.

As propriedades e outras definições derivadas do complemento difuso, da interseção difusa (t-normas) e da união difusa (t-conormas) são apresentadas no textos básicos da Teoria de Conjuntos Difusos, como por exemplo DUBOIS AND PRADE (1980), KANDEL (1986), KLIR AND YUAN (1995) e ROSS (1996) entre outros.

Todas a propriedades de conjuntos "crisp" ou clássicos valem também para conjuntos difusos exceto a "lei da contradição" e a "lei do meio-excluído". Ou seja para conjuntos difusos, não necessariamente

$$A \cap \bar{A} = \emptyset \quad (2.3.13)$$

$$\bar{A} \cup A = X \quad (2.3.14)$$

Para qualquer conjunto difuso  $A$  definido no conjunto universo  $X$ , define-se cardinalidade escalar  $|A|$ , pela fórmula:

$$|A| = \sum_{x \in X} A(x) \quad (2.3.15)$$

Alguns autores se referem a  $|A|$  como "*sigma count*" de  $A$ .

O grau de sub-conjunção ("subsethood")  $S(A,B)$  é definido para qualquer par de subconjuntos definidos no conjunto universo finito  $X$  pela fórmula:

$$S(A,B) = \frac{1}{|A|} (|A| - \sum_{x \in X} \max[0, A(x) - B(x)]) \quad (2.3.16)$$

O termo  $\Sigma$  nessa fórmula descreve a soma dos graus para o qual a desigualdade  $A(x) \leq B(x)$  é violada, a diferença descreve a falta dessas violações, e a cardinalidade  $|A|$  no denominador é um fator normalizador para obter a amplitude

$$0 \leq S(A,B) \leq 1 \quad (2.3.17)$$

Pode-se converter a expressão 2.3.16 para uma forma mais conveniente:

$$S(A,B) = \frac{|A \cap B|}{|A|} \quad (2.3.18)$$

onde  $\cap$  denota a interseção difusa padrão.

Uma notação especial que é frequentemente utilizada na literatura para definir conjuntos difusos com um suporte finito é dada por:

$$A = a_1/x_1 + \dots + a_n/x_n \quad (2.3.19)$$

para um dado conjunto difuso  $A$  em um conjunto universo  $X$  e  $a_i$  denota o grau de pertinência do  $x_i$  em  $A$  para todo  $i$ . As barras são empregadas para ligar os elementos do suporte com seu graus de pertinência em  $A$ , e o sinal de mais não representa qualquer operação algébrica, mas indica que os pares de elementos listados e os graus de pertinência formam coletivamente a definição do conjunto  $A$ .

#### *Princípio da Extensão*

Dada uma função "crisp"

$$f: X \rightarrow Y \quad (2.3.20)$$

ela é "difusificada" quando é estendida para agir nos conjuntos difusos definidos em  $X$  e  $Y$ . Isto é, a função "difusificada", para a qual o mesmo símbolo  $f$  é comumente utilizado, tem a forma:

$$f: \mathfrak{F}(X) \rightarrow \mathfrak{F}(Y), \quad (2.3.21)$$

e sua função inversa,  $f^{-1}$ , tem a forma:

$$f^{-1}: \mathfrak{F}(Y) \rightarrow \mathfrak{F}(X). \quad (2.3.22)$$

Um princípio para "difusificar" funções "crisp" é chamado *Princípio de Extensão*.

### Princípio da Extensão:

Qualquer função  $f: X \rightarrow Y$  induz duas funções:

$$f: \mathfrak{F}(X) \rightarrow \mathfrak{F}(Y), \quad (2.3.24)$$

$$f^{-1}: \mathfrak{F}(Y) \rightarrow \mathfrak{F}(X). \quad (2.3.25)$$

que são definidas por:

$$[f(A)](y) = \sup_{x/y=f(x)} A(x) \quad (2.3.26)$$

para todo  $A \in \mathfrak{F}(X)$  e

$$[f^{-1}(B)](x) = B(f(x)) \quad (2.3.27)$$

para todo  $B \in \mathfrak{F}(Y)$ .

Quando os conjuntos  $X$  e  $Y$  são finitos, pode-se substituir  $\sup$  em (2.3.26) pelo  $\max$ .

As operações de Agregação são baseadas no princípio da extensão.

## 2.4 Outras Medidas de Incerteza

A *Teoria Difusa de Medidas* ("Fuzzy Measure Theory") é uma descendente da *Teoria Clássica de Medidas*. Essa última tem suas raízes na métrica geométrica, que é caracterizada por assinalamentos de números a comprimentos, áreas ou volumes. Na antigüidade, esse processo de assinalamento, ou mensuração, foi inicialmente concebido simplesmente como uma comparação com uma unidade padrão. Mas logo, o problema da incommensurabilidade revelou que mensurar é mais complicado do que esse sugestivo simples processo. Ficou claro que mensurar deve inevitavelmente envolver conjuntos e processos infinitos.

Antes da emergência e desenvolvimento suficiente do cálculo, o problema de incomensurabilidade causou ansiedade uma vez que não haviam ferramentas satisfatórias para lidar com ele. O cálculo integral, baseado na integral de Riemann, que se tornou bem desenvolvido na segunda metade do século XIX, foi a primeira ferramenta para tratar do

problema. Certas medidas que são contingentes com a existência de limites associados poderiam finalmente ser determinadas pelo uso de técnicas apropriadas de integração.

No século XIX, ocorreu uma crescente necessidade de análises matemáticas mais precisas, induzidas primeiramente pelo rápido avanço da ciência e da tecnologia. Como resultado, emergiram novas questões com relação a mensuração. Por exemplo, quando se remove os pontos finais de um intervalo unitário real, qual é a medida do conjunto remanescente? Qual é a medida do conjunto obtido pela remoção de todos os números racionais do conjunto acima?

Questões como essas e outras ainda mais difíceis, foram cuidadosamente examinadas por Émile Borel, que desenvolveu uma teoria dando um importante passo para construir um teoria mais geral que hoje é conhecida como *Teoria Clássica de Medidas*.

A teoria de Borel trata de famílias de subconjuntos do conjunto dos números reais que são fechados sob o conjunto união de muitos conjuntos contáveis e do conjunto complemento. Ele definiu uma medida que associa um número real positivo com cada conjunto limitado na família, que, no caso de um intervalo, é exatamente igual ao comprimento do intervalo. A medida é aditiva no sentido que seu valor para uma união limitada de uma seqüência de conjuntos disjuntos é igual à soma dos valores associados com conjuntos individuais.

Borel não conectou sua teoria com a teoria de integração. Isso foi feito alguns anos depois por Henri Lebesgue, outro matemático francês. Em um artigo publicado em 1900, ele definiu uma integral, mais geral que a de Riemann, que é baseada em uma medida generalizada que submete a medida de Borel como um caso especial. Esses conceitos generalizados de uma medida e de uma integral, que são agora conhecidos como Medida de Lebesgue e Integral de Lebesgue, são as pedras fundamentais da Teoria Clássica de Medidas.

A teoria clássica de medidas está diretamente relacionada com a *Teoria de Probabilidades*. Uma medida de probabilidade, como qualquer outra medida clássica, é um conjunto de funções que assinala 0 para o conjunto vazio e um número não negativo para qualquer outro conjunto, e que seja aditiva. No entanto, uma medida de probabilidade

requer, em adição, que 1 seja assinalado para o conjunto universo em questão. Então a teoria de probabilidades pode ser vista como uma parte da teoria clássica de medidas.

O conceito de uma medida de probabilidade (ou simplesmente probabilidade) foi formulado axiomaticamente em 1933 por Andrei N. Kolmogorov, um matemático russo (KOLMOGOROV, 1950). Este conceito de probabilidade é algumas vezes chamado de probabilidade numérica ou quantitativa para distingui-la de outros tipos de probabilidade, como por exemplo a probabilidade classificatória ou comparativa (FINE, 1973; WALLEY AND FINE, 1979 e WALLEY, 1991). Embora o termo "Teoria de Probabilidades", com nenhum outro qualificativo adicional, se refira normalmente à teoria baseada nos axiomas de Kolmogorov.

Depois de mais de 50 anos de existência e desenvolvimento estável da Teoria Clássica de Medidas, a propriedade de aditividade das medidas clássicas se tornou objeto de controvérsias. Alguns matemáticos sentiram que a exigência de aditividade é muito restritiva em alguns contextos de aplicação. Ela é muito restritiva para capturar adequadamente todo o escopo da medida. Enquanto aditividade caracteriza bem muito tipos de mensurações sob condições idealizadas livres de erro, ela não é plenamente adequada para caracterizar a maioria das mensurações de condições físicas reais, onde medidas de erros são inevitáveis. Ainda mais, algumas medidas, envolvendo por exemplo, julgamentos subjetivos ou experimentos não repetíveis, são intrinsecamente não-aditivos.

Numerosos argumentos têm sido ou podem ser levantados contra a necessidade e adequação do axioma de aditividade da teoria de probabilidades. Um desses argumentos foi apresentado por VIERTL (1987). Ele é baseado no fato de que todas as mensurações são inerentemente difusas devido aos inevitáveis erros de medida. Considerando por exemplo, dois eventos disjuntos, A e B, definidos em termos de intervalos contíguos de números reais. Observações em vizinhanças próximas (dentro de um erro de medida) de um ponto final de cada evento são inseguras e podem ser descontadas propriamente, por exemplo, de acordo com funções de razão de desconto. Isto é, observações nas vizinhanças dos pontos finais carregam menos evidências que aquelas internas. Quando medidas são tomadas pela união de dois eventos, uma das funções de razão de desconto não é aplicável. Assim, o mesmo procedimento de observação é mais evidente para esse simples evento  $A \cup B$  do

que para os dois eventos disjuntos  $A$  e  $B$ . Isso implica que o grau de crença em  $A \cup B$  (probabilidade de  $A \cup B$ ) deve ser maior que a soma dos graus de crença em  $A$  e  $B$  (probabilidades de  $A$  e  $B$ ). O axioma da aditividade é então violado.

Considerando, por exemplo, dois eventos disjuntos,  $A$  e  $B$ , definidos em termos de intervalos disjuntos de números reais, como mostrado na figura 2.4.1. As observações nas vizinhanças próximas (dentro de um erro de medida) do pontos finais de cada evento são pouco admissíveis e devem ser descontadas, por exemplo, de acordo com um função de razão de desconto mostrada na figura 2.4.1(a). Isto é, observações nas vizinhanças dos pontos finais devem carregar menos evidência que aqueles nos centros dos intervalos. Quando medidas são tomadas para a união dos dois eventos como na figura 2.4.1 (b), uma das funções de razão de desconto não é aplicável. Dessa forma, as observações produzem mais evidências para o evento simples  $A \cup B$  do que para os dois eventos disjuntos  $A$  e  $B$ . Isso implica que o grau de crença em  $A \cup B$  (probabilidade de  $A \cup B$ ) deve ser maior que a soma dos graus de crença em  $A$  e  $B$  (probabilidades de  $A$  e  $B$ ). O axioma da aditividade é então violado.

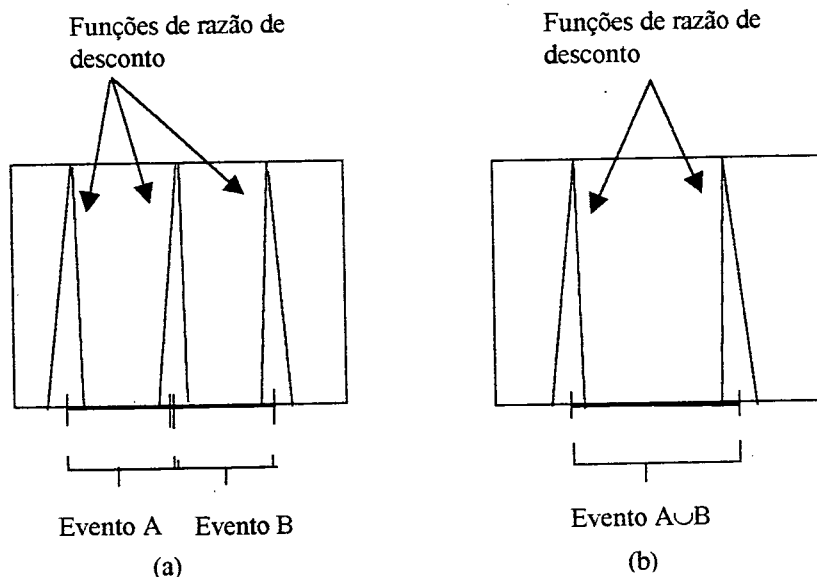


Figura 2.4.1 Um exemplo ilustrativo da violação do axioma da aditividade da teoria de probabilidade

O primeiro desafio para a teoria clássica de medidas foi proposto por G. Choquet em 1954, para o qual ele nomeou de *Teoria das Capacidades*. A capacidade de Choquet é uma função de conjunto que associa um número real (não necessariamente não negativo) a



cada subconjunto do conjunto universo empregado e é contínuo e monótono com respeito ao conjunto inclusão.

Embora a teoria de capacidades de Coquet seja uma ferramenta ampla, considerando vários tipos de medidas de não aditividade, ela é muito geral para a maioria dos problemas práticos. Outras medidas de não aditividade mais úteis emergiram mais tarde em contextos mais específicos.

No contexto da teoria de probabilidades, uma teoria generalizada baseada em dois tipos de medidas não aditivas foi originada por DEMPSTER (1967) e mais tarde, completamente desenvolvida por SHAFER (1976). Esses tipos de medidas foram obtidas pela substituição do requerimento de aditividade das medidas de probabilidades com requerimentos de superaditividade e subaditividade. As medidas de superaditividade, que são também semicontínuas superiores, são usualmente chamadas de medidas de crença. As medidas de subaditividade, que são também semicontínuas inferiores, são usualmente chamadas de medidas de plausibilidade.

Dada uma medida de cada um dos dois tipos, ela induz uma única medida de outro tipo. Tomando conjuntamente, medidas de crença e plausibilidade elas formam o que é usualmente chamado de *Teoria de Dempster-Shafer* ou *Teoria de Evidências*. Verifica-se do trabalho inicial de Dempster em 1967, que as medidas de crença e plausibilidade têm uma interpretação natural como probabilidades inferiores e superiores, respectivamente. Uma vez que medidas de crença são sempre menores ou iguais às correspondentes medidas de plausibilidade, os intervalos entre os valores de crença e plausibilidade podem ser vistos como amplitudes de probabilidades admissíveis. A Teoria de Dempster-Shafer deve então ser vista como a teoria que é capaz de tratar com probabilidades intervalares ("interval-valued").

Outra teoria baseada em medidas não aditivas, chamada de *Teoria da Possibilidade*, emergiu dos conceitos da *Teoria de Conjuntos Difusos*. Um conjunto difuso é um conjunto cujos limites não têm um ponto de corte ("sharp") completamente definido. Ou seja a mudança de pertinência para não pertinência é gradual e não abrupta. Essa mudança gradual é expressa pela função de graus de pertinência ("grade of membership function") do conjunto difuso, que assinala para cada indivíduo do conjunto universo um valor no

intervalo unitário  $[0,1]$ . Esse valor representa o grau de pertinência de um indivíduo no conjunto difuso. Um conjunto difuso é chamado de normal se o máximo da função do grau de pertinência for 1.

Dado um conjunto difuso normal, ZADEH (1978) definiu uma função de distribuição de possibilidades associada a cada conjunto como numericamente igual à sua função de grau de pertinência. Então, ele define uma medida de possibilidade tomando o supremo da função de distribuição de possibilidades para cada conjunto "crisp" concernente.

As medidas de possibilidade emergem não somente do contexto de conjuntos difusos, mas também do contexto da Teoria de Evidências, como casos especiais das medidas de plausibilidades (DUBOIS AND PRADE, 1988 e KLIR AND FOLGER, 1988). É nesse contexto que é empregado no estudo das medidas de possibilidade e a associada medida de necessidade (um caso especial de medida de crença).

YAGER (1992) discute a conexão entre a especificidade de uma distribuição de possibilidades e a entropia de uma distribuição de probabilidades. Especificidade nesse contexto é uma medida do grau para o qual a distribuição permite um e somente um elemento como sua manifestação. E como tal é uma medida da quantidade de incerteza ou informação. BOUCHON (1987) apresenta inferências difusas e distribuições condicionais de possibilidades. SANDRINI ET AL. (1995) discutem como, através de julgamento conjunto de especialistas se pode buscar e estabelecer informações usando a Teoria de Possibilidades.

A *Teoria de Probabilidades Imprecisas ou Teoria das Probabilidade Inferior e Superior* de WALLEY(1991) é baseada nos conceitos de probabilidade inferior e probabilidade superior, partindo de uma interpretação comportamental de probabilidade e previsão baseada nos princípios de coerência. Segundo o autor o raciocínio começa com o reconhecimento da ignorância e da incerteza. Em raciocínio prático, se se deseja fazer inferências ou tomar decisões, é necessário ponderar incerteza sobre fatos ou eventos e a indeterminância que vem da ignorância sobre o assunto tratado. Para medir incerteza, a teoria trabalha com o conceito de probabilidade e para medir a indeterminância necessita-se de probabilidades imprecisas. A teoria tem objetivos próximos à *Teoria Bayesiana* de

inferência e decisão estatística. O que a distingue as duas é o fato de que a Teoria Bayesiana tem como suposição central que a incerteza deve ser medida por uma simples medida aditiva de probabilidade, e os valores podem sempre ser medidos por uma função precisa de utilidade, o que é chamado de o dogma Bayesiano da precisão.

A *Teoria de Probabilidade Interverlar*, devida a DE CAMPOS AND MORAL (1994) pode ser considerada como um caso particular da teoria de probabilidades inferior e superior. Define-se  $L$  como uma família de intervalos  $L = \{[l_i, u_i], i = 1, \dots, n\}$ , onde esses intervalos podem ser interpretados como um conjunto de limites de probabilidades que definem um conjunto  $P$  de distribuições de probabilidades, tal que  $l_i \leq p_i \leq u_i$ . Então  $L$  é um conjunto de probabilidades intervalares e  $P$  um conjunto de possíveis probabilidades associadas a  $L$ . A simplicidade computacional do trato com probabilidades intervalares e a sua expressividade são ferramentas interessantes para o trato com raciocínio com incerteza. Esses intervalos podem ser obtidos diretamente da opinião de especialistas ou de informações empíricas (um conjunto de exemplos) usando técnicas estatísticas como intervalos de confiança.

Um outro enfoque para se trabalhar com incerteza probabilística é a chamada *Geometria Estocástica* que engloba duas teorias que tratam o espaço amostral como uma entidade "mais ou menos arbitrária" (KENDALL, 1974), a *Teoria de Conjuntos Aleatórios* e a *Teoria de Medidas Aleatórias*.

Deve-se pensar em um conjunto aleatório  $Z$  como um animal selvagem em uma floresta, e o caçador deseja localizá-lo. Para isso ele instala um sistema de armadilhas ao longo da floresta, conectadas eletricamente ao acampamento. Se o animal  $Z$  contatar a armadilha  $T$ , então uma luz correspondente no painel de controle se acende. O caçador inspeciona o painel de controle, nota que luzes estão acessas e desta forma deduz a posição aproximada do animal na floresta. Se " $C$ " é o espaço amostral (a floresta), então  $Z$  é algum subconjunto de  $C$ . Define-se a função  $f_z: \tau \rightarrow \{0,1\}$  como segue:

$$f_z(T) = 0, \text{ quando } Z \cap T \text{ é vazio} \quad (2.4.1)$$

$$f_z(T) = 1, \text{ caso contrário.} \quad (2.4.2)$$

A função  $f_z$  contém "toda a informação disponível ao caçador".

As questões fundamentais a serem solucionadas são: (i) Se a função define  $Z$  unicamente, e em caso negativo qual é a relação de equivalência entre os subconjuntos  $Z$  e  $C$  associados com a não-unicidade. E (ii) dado "f" como se pode falar se  $f = f_z$  para algum subconjunto de  $Z$  de  $C$ , e como se pode construir  $Z$ .

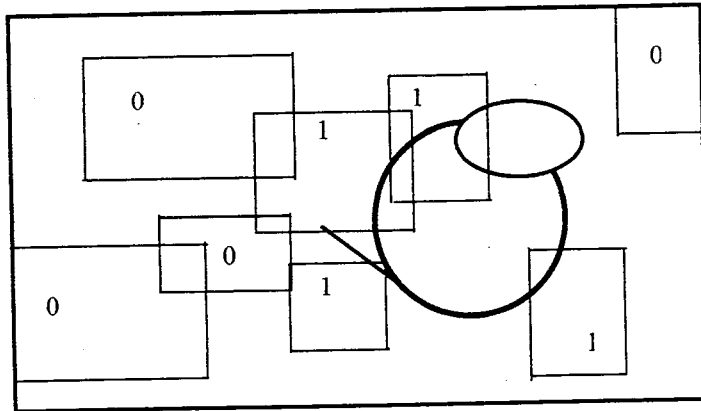


Figura 2.4.2 Uma típica "função amostral" para o processo aleatório do animal. Algumas das armadilhas  $T$  são mostradas como retângulos, e os valores  $f_z(t)$  são indicados

PENG, WANG AND KANDEL (1996 a) e PENG, WANG AND KANDEL. (1996 b) estabelecem algumas distribuições especiais de possibilidades que podem ser implementadas como ferramentas de aquisição do conhecimento através de regras para conceitos pré-definidos, essas regras em modelos difusos e controladores lógicos difusos pertencem a categorias difusas.

Uma outra forma de se tratar com conjuntos não "crisp" e complementar à Teoria de Conjuntos Difusos e à *Teoria da Tolerância* é a chamada *Teoria de Conjuntos Rústicos* ("*Rough Set Theory*") apresentada por PAWLAK (1982). Nessa teoria são investigadas a igualdade aproximada de conjuntos e a inclusão aproximada de conjuntos. Define-se  $U$  como um conjunto universo e  $X$  um subconjunto de  $U$ . O menor conjunto composto em  $A$  contendo  $X$  é chamado de melhor aproximação superior de  $X$  ( $APR_A(X)$ ) e o maior conjunto composto de  $A$  contido em  $X$  é chamado de melhor aproximação inferior de  $X$ , ( $apr_A(X)$ ), onde  $A$  é um conjunto de objetos e  $X$  um conjunto de atributos. O conjunto

$Bnd_A(X) = APR_A(X) - apr_A(X)$  é chamado de limite ("boudary") de X em A. A figura abaixo ilustra de forma geral os conceitos apresentados:

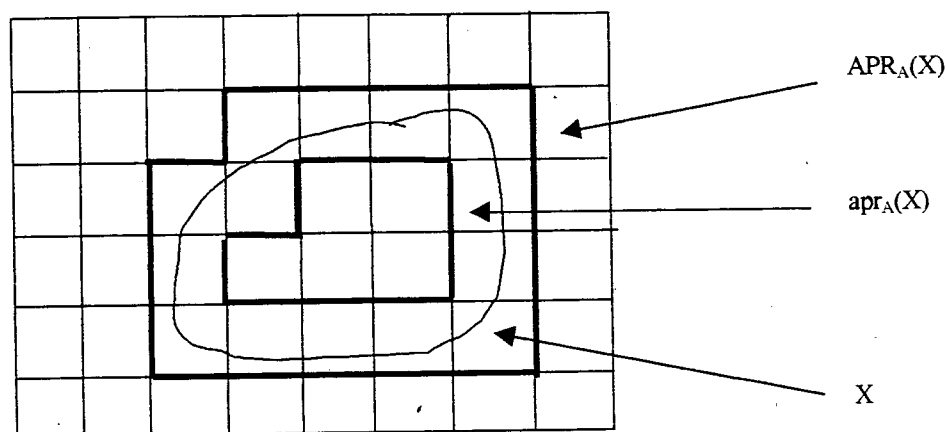


Figura 2.4.3 Conjunto X e suas melhores aproximações superior e inferior.

A chave deste enfoque é a formulação matemática exata do conceito de igualdade aproximada ("rough") de conjuntos em um dado espaço de aproximação. DUBOIS AND PRADE (1990) apresentam uma composição da incerteza "vaga" e da incerteza "grosseira". A primeira medida por Teoria de Conjuntos Difusos e a segunda pela Teoria de Conjuntos Rústicos, derivando aproximações inferiores e superiores de conjuntos difusos, na mesma linha dos conceitos de funções de crença de Shafer. Chamam essa composição de teorias de *Teoria de Conjuntos Rústicos Difusos*.

Muitas extensões e generalizações da Teoria de Probabilidades permitem análises apesar da ausência de alguma informação. A Teoria de Conjuntos Difusos procura quantificar e analisar certos tipos de vaguesa e ambigüidade., enquanto a *Teoria da Surpresa*, SMITHSON (1989) e a Teoria das Possibilidades lidam com tipos mais "folgados" de indeterminância e probabilidade.

As diversas teorias de conjuntos têm o seu dual na lógica. A Teoria Clássica de Conjuntos ou Teoria de Conjuntos de Cantor tem na Lógica Aristotélica seu dual. A base dessa teoria é que qualquer proposição ou é falsa ou é verdadeira. No entanto, estabelecida dessa forma, surgem vários paradoxos na solução de problemas lógicos. MCNEILL AND FREIBERG (1993) e KOSKO (1993) entre outros apresentam uma série desses paradoxos. RUSSEL (1923) discute a vaguesa contida nos conceitos de verdadeiro e falso, afirmando que vaguesa e precisão são características da linguagem e não realidade. Mas foi Lukasiewics quem primeiro formulou um modelo lógico tri-valorado, que permitia uma terceira condição de indeterminância entre os valores verdadeiro e falso. Outros modelos de *Lógica Tri-valorada* foram apresentados por Bochvar, Kleene, Heyting e Reichenbach (KLIR AND YUAN, 1995). Mas no entanto, foi o desenvolvimento da *Lógica Multivalorada* de Lukasiewics proposta em 1922 que permitiu o desenvolvimento posterior, na década de 60, de teorias de incerteza baseadas nessa lógica, como a Teoria de Conjuntos Difusos de ZADEH (1965), que associou essa lógica aos conceitos de vaguesa em BLACK (1937). Uma outra lógica que questiona o princípio Aristotélico da não-contradição é a *Lógica Paraconsistente* desenvolvida a partir de 1964 por Newton Carneiro da Costa como uma teoria dedutiva, cuja linguagem contenha um símbolo de negação e que o conjunto de seus teoremas contenha ao menos dois deles, um dos quais é a negação do outro (COSTA, 1993). Assim, por exemplo como as partículas atômicas se comportam ora como partículas e ora como ondas, em certos aspectos elas são e não são partículas. Essa lógica tem sido utilizada em robótica e em várias outras ciências como medicina, onde um paciente em estado grave pode receber diagnósticos bem variados feitos por vários especialistas.

## 2.5. Teoria de Probabilidades e Teoria de Conjuntos Difusos

Na discussão do relacionamento e das oposições da Teoria de Probabilidade e da Teoria de Conjuntos Difusos para caracterizar incertezas, além da questão da natureza do fenômeno e da correspondente medida de incerteza, uma outra questão básica diz respeito ao nível em que se está estabelecendo a medida.

A Teoria de Probabilidades foi construída tendo como alicerce a Teoria Clássica de Conjuntos. Dessa forma, a medida de incerteza em questão é a Medida de Probabilidade,

defina como uma função definida em uma álgebra de conjuntos que obedece os axiomas apresentados mais abaixo.

Dessa forma dois níveis de medidas de incerteza são definidos. Um para conjuntos (como se caracteriza a pertinência de um elemento a um conjunto). Na teoria clássica um elemento pertence ou não a um conjunto. Na teoria difusa um elemento pode pertencer a mais de um conjunto, com diferentes graus de pertinência.

Num segundo nível deseja-se medir incerteza associada a como um elemento pertence a um conjunto. No primeiro caso quando um elemento pertence a um conjunto não há incerteza associada à pertinência.

**Definição 2.5.1 Função Indicadora:** Seja  $\Omega$  um espaço amostral qualquer com pontos  $\omega$  e  $A$ , um subconjunto qualquer de  $\Omega$ . A função indicadora de  $A$ , denotada por  $I_A(\cdot)$ , é uma função com domínio  $\Omega$  e contradomínio igual ao conjunto consistindo de dois números reais 0 e 1 definida por:

$$\begin{aligned} I_A(\omega) &= 1 \text{ se } \omega \in A \\ I_A(\omega) &= 0 \text{ se } \omega \notin A \end{aligned} \quad (2.5.1)$$

A medida de incerteza em questão é a Medida de Probabilidade, defina como segue.

**Definição 2.5.2 Medida de Probabilidade:** Seja  $\{\Omega, \wp\}$  um espaço, onde  $\Omega$  é um espaço amostral (o conjunto de todos os possíveis resultados de um experimento aleatório) e  $\wp$  uma  $\sigma$ -álgebra de subconjuntos de  $\Omega$ . Define-se como medida de probabilidade (ou simplesmente probabilidade) a função  $P$  definida em  $\wp$ , tendo como contradomínio o intervalo unitário  $[0,1]$  e que satisfaz os seguintes axiomas:

- (i)  $P[A] \geq 0$  para todo  $A \in \wp$ .
- (ii)  $P[\Omega] = 1$ .
- (iii) Se  $A_1, A_2, \dots$  é um seqüência de eventos mutuamente exclusivos em  $\wp$  e se

$$A_1 \cup A_2 \cup \dots = \bigcup_{i=1}^{\infty} A_i \in \wp, \text{ então } P\left[\bigcup_{i=1}^{\infty} A_i\right] = \sum_{i=1}^{\infty} P A_i. \quad (2.5.2)$$

Assim sendo, na primeira definição a incerteza envolvida é a definida por KLIR (1988) como não-especificidade. Ou o elemento está no conjunto ou não está. A medida de incerteza envolvida está em 2.5.2 e é definida como dissonância. São então dois níveis distintos de incerteza.

Na Teoria de Conjuntos Difusos a incerteza já está envolvida no primeiro nível, que envolve não-especificidade e também "difusividade" ou vaguesa na própria definição da função indicadora.

**Definição 2.5.3 Medida Difusa:** Dado um conjunto  $X$  e uma família não-vazia  $\zeta$  de subconjuntos de  $X$ , uma medida difusa em  $\{X, \zeta\}$  é uma função  $g: \zeta \rightarrow [0,1]$  e que satisfaz os seguintes requerimentos:

(i)  $g(X) = 1$ .

(ii) Para todo  $A, B \in \zeta$ , se  $A \subseteq B$ , então  $g(A) \leq g(B)$ .

(iii) Para qualquer seqüência crescente  $A_1 \subset A_2 \subset \dots$  em  $\zeta$ , se  $\bigcup_{i=1}^{\infty} A_i \in \zeta$ , então

$$\lim_{i \rightarrow \infty} g(A_i) = g\left(\bigcup_{i=1}^{\infty} A_i\right)$$

(iv) Para qualquer seqüência crescente  $A_1 \supset A_2 \supset \dots$  em  $\zeta$ , se  $\bigcap_{i=1}^{\infty} A_i \in \zeta$ , então

$$\lim_{i \rightarrow \infty} g(A_i) = g\left(\bigcap_{i=1}^{\infty} A_i\right) \quad (2.5.3)$$

Verifica-se então que uma medida difusa não requer aditividade e que a função indicadora correspondente à uma medida difusa leva não em  $\{0,1\}$ , dois pontos, como na Teoria Clássica de Conjuntos mas em  $[0,1]$ , um intervalo unitário, sem que se confunda uma medida difusa com uma medida de probabilidade que também leva em  $[0,1]$ .

Sendo assim, é possível então a determinação de probabilidades de um evento difuso, conforme detalhado no item 2.6. Antes porém de se apresentar mais formalmente as possibilidades de agenda comum entre as duas teorias para se tratar a incerteza, é



apresentada ainda nessa seção algumas discussões sobre onde a Teoria de Probabilidades e a Teoria de Conjuntos Difusos são concorrentes.

Por muito tempo as incertezas dos fenômenos foram tratadas somente no campo da Teoria de Probabilidades. Mas depois do desenvolvimento de teorias específicas para modelar fenômenos que continham incerteza de outra natureza que não a aleatoriedade uma questão de fundamental importância foi colocada. Seria a Teoria de Probabilidades suficiente para lidar com incerteza? Pelo fato dessa discussão estar ainda muito em aberto, e não ser objetivo específico do trabalho contribuir de forma mais definitiva na sua solução, são apresentados argumentos das duas escolas, embora a simples utilização de modelos probabilísticos-difusos revele a opção por uma resposta negativa à pergunta formulada.

A resposta a essa pergunta tem sido positiva para probabilistas, mesmo para pensadores bayesianos, onde a probabilidade é vista, não como uma frequência ou outra quantidade objetiva, mas como um estado subjetivo do conhecimento. LINDLEY (1982) argumenta pela inevitabilidade da probabilidade através da discussão da formulação de regras de escores, afirmando que somente as descrições probabilísticas da incerteza são razoáveis. No painel de discussões que acompanha o artigo, Shafer argumenta contra essa conclusão afirmando que ela é injustificada porque Lindley confunde duas diferentes tarefas. Uma de descrever incerteza e a outra de escolher números que minimizem penalidades que se incorre sob regras de escore unidimensionais. Zadeh argumenta que para chegar às suas conclusões, o autor teve que fazer uma suposição crucial, a de que os escores são aditivos, e sugere que sob aditividade o trato com as regras de escores levam à inevitabilidade da probabilidade, e que quando a suposição de aditividade é derrubada, abre-se a porta para outras formas de se caracterizar a incerteza.

CHEESEMAN (1988) discute qual o melhor método para representar e se raciocinar sobre o senso comum (que algumas vezes é chamado de inferência plausível). O autor argumenta que as dificuldades técnicas encontradas com o uso da lógica em inteligência artificial não são encontradas com o uso da probabilidade bayesiana. E que a probabilidade substitui a mais razoável noção de crença, onde as várias peças de evidência são combinadas para se obter uma medida total de crença em uma proposição. Dubois e Prade salientam que as conclusões tiradas são baseadas na "fé na prova de Cox" sobre a unicidade

do cálculo de probabilidades. E que embora o artigo de Cox seja muito útil quando oferece um bom começo para a discussão das teorias de incerteza ele não é definitivo, como demonstrado em artigo de 1987 dos autores, onde eles afirmam que nem todos os axiomas de Cox são inquestionáveis. Em particular que o axioma da complementaridade pode ser enfraquecido, no sentido de contabilizar a ausência de crença em uma proposição tão bem quanto em sua negação. A principal discordância de Dempster com relação à Cheeseman diz respeito ao fato que o argumento de Cheeseman que a inferência bayesiana pode ser usada não somente para o assinalamento de probabilidades "a posteriori" para categorias preestabelecidas, mas também pode ser usada para estabelecer as próprias categorias, algo que hoje vai além da inferência bayesiana, ou de fato, além de qualquer teoria formal existente.

Dois importantes debates sobre as contradições e as relações entre as Teoria de Probabilidades e a Teoria de Conjuntos Difusos ocorreram em 1994 e 1995 tendo como motor artigos encabeçados por M.Laviolette, um estatístico bayesiano. Em seu primeiro artigo LAVIOLETTE AND SEAMAN JR. (1994), o autor procura demonstrar através de três aplicações, que os cinco (sic) argumentos básicos da Teoria de Conjuntos Difusos para representar incerteza não conseguem demonstrar que essa última teoria possa representar um tipo de incerteza diferente da probabilidade. Segundo o autor os argumentos são: (i) A hipótese da realidade, que assegura que imprecisão é uma propriedade inerente do mundo externo ao observador; (ii) A hipótese subjetiva, que assegura que probabilidade é uma medida de incerteza exclusivamente objetiva e que portanto incerteza subjetiva só pode ser representada por conjuntos difusos; (iii) A hipótese comportamental, que afirma que sistemas incertos devem emular mais do que prescrever o comportamento humano em face da incerteza; (iv) A hipótese da "probabilidade como ficção", que afirma que probabilidade não compreende um campo de estudo em seu sentido próprio e (v) A hipótese do superconjunto, que assegura que a Teoria de Conjuntos Difusos inclui probabilidades como um caso especial e então fornece um meio-ambiente de modelagem de incerteza mais rico.

Para cada um dos argumentos os autores respondem basicamente que (i) Rejeitar a noção que ambigüidade na classificação é uma propriedade inerente dos objetos. (ii) Embora a maioria dos estatísticos tenham sido treinados inicialmente na escola freqüentista, a estatística bayesiana trata com probabilidade subjetiva; (iii) A caracterização de que

declarações de incerteza subjetiva não são aditivas uma vez que o comportamento humano é freqüentemente contraditório, está incorreta porque os postulados de probabilidade subjetiva não assumem essa aditividade. (iv) O argumento não vale porque os próprios probabilistas afirmam, como deFinetti que "a probabilidade não existe" e (v) A inclusão de uma teoria como um caso especial de outra não implica em si, que a teoria que é subconjunto seja inferior que a que a contém.

O objetivo da Teoria de Conjuntos Difusos não é provar que a estatística, teoria de jogos e outras estabelecidas tenham se desenvolvido de forma errada até agora. Segundo Dubois e Prade, no painel de discussão, a Teoria de Conjuntos Difusos é uma ficção conveniente para modelar gradualidade no processo cognitivo humano. Encontrar formas de complementação de teorias é uma tarefa necessária. Os autores não acreditam que uma simples teoria forneça um meio ótimo e completo para resolver problemas e gerenciar incerteza. Existe um problema de epistemologia no trato com objetividade. A probabilidade seja ela freqüentista ou epistêmica (muitas vezes chamada de subjetiva) é de fato uma medida objetiva, no sentido de que pretende fornecer uma simples medida para descrever a incerteza contida em um fenômeno, a ser medida por uma função de utilidade precisa. WALLEY(1991) na apresentação de seu conceito de probabilidade subjetiva (epistêmica) critica o "dogma bayesiano da precisão", por não conseguir incorporar o que há de vago em certos fenômenos.

Do ponto de vista do relaxamento da aditividade, a Teoria de Probabilidades está inserida em outras teorias, o que de fato não a torna mais ou menos forte na descrição do mundo real como quer KOSKO (1990, 1992 e 1993). Cada teoria se matematicamente consistente (como a Teoria de Probabilidades, a Teoria de Conjuntos Difusos, a Teoria de Possibilidades e muitas outras) é somente um meio conveniente para resolver um particular tipo de problema e compreender uma faceta particular da incerteza e/ou da ignorância. O "bayesianismo subjetivo" é uma ferramenta rigorosa para tomada de decisão sob incerteza aleatória, a Teoria de Conjuntos Difusos é uma ferramenta algébrica geral para modelar imprecisão léxica e gradualidade nos predicados. A Teoria de Possibilidades é uma simples teoria ordinária das funções de conjuntos para modelar e raciocinar com conhecimento incompleto, que pode ser envolvida pela Teoria de Probabilidades Superiores.

Do ponto de vista matemático, no que diz respeito ao relaxamento da imposição da aditividade, as medidas podem ser representadas conforme esquema abaixo. KLIR (1995):

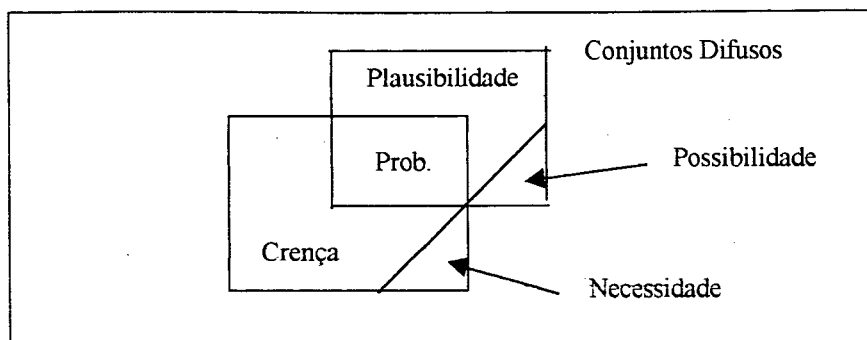


Figura 2.5.1 Relação de "subconjunção" entre as principais medidas de incerteza

No seu segundo artigo LAVIOLETTE ET AL. (1995) retomam a discussão apresentando agora dois outros exemplos para demonstrar que os métodos de probabilidade e estatística são as melhores ferramentas para resolver problemas envolvendo incerteza. Entre outros argumentos KANDEL ET AL. (1995) questionam fortes argumentos de Laviolette. Primeiro que pesquisadores na área de psicologia cognitiva revelaram que o raciocínio humano em tarefas de predição e julgamentos sob incerteza não seguem o "cálculo de chances" ou a teoria estatística. "Sob incerteza, o raciocínio humano quebra mesmo as leis fundamentais de probabilidade, como a lei de conjunção". Além de desconsiderarem o tamanho da população em estudo. Por outro lado a Teoria de Conjuntos Difusos não exige que as leis da contradição e do meio-excluído sejam obedecidas, e então categorias são definidas por elementos com graus de pertinência. Os autores afirmam que a Teoria de Conjuntos Difusos e a Teoria de Probabilidades não são excludentes. Elas podem trabalhar juntas porque medem diferentes conceitos semânticos.

A resposta à pergunta se a Teoria de Probabilidades é suficiente para lidar com incerteza, talvez pudesse ser simplesmente: "sim" e "não". O que está em jogo é como modelar um sistema, considerando-se complexidade, incerteza e utilidade. A Teoria de Conjuntos Difusos e outras teorias que buscam medir incerteza além da de natureza aleatória, são necessárias quando se trabalha com sistemas complexos, em que a sua não consideração pode diminuir drasticamente a utilidade do sistema. Nesse caso a resposta é

negativa. A Teoria de Probabilidades não é suficiente para lidar com incerteza em sistemas complexos muito presentes em Inteligência Artificial. ZADEH (1986) defende o ponto de vista que a falta de adequação da Teoria de Probabilidades reside na sua falta de expressividade com uma linguagem de incerteza para descrever eventos difusos e probabilidades difusas. No entanto, a resposta é positiva quando se trata de sistemas menos complexos em que há aleatoriedade em algum grau, e que possa ser modelado simplesmente com a utilização de probabilidade, permite-se algumas suposições e restrições que simplifiquem o modelo e resolvam satisfatoriamente o problema com economia de energia, sem perda de generalidade e da utilidade.

KLIR (1989) conclui seu trabalho "Is there more to uncertainty than some probability theorists might have us believe?" que as questões para os próximos debates não deveriam ser se a Teoria de Probabilidades e outras teorias estão certas ou não. A questão deveria ser mais que teorias são relevantes e apropriadas para conceitualizar e tratar com incerteza em cada contexto particular. E mais, a Teoria de Probabilidades e as outras teorias não deveriam ser vistas necessariamente como conflitantes umas com as outras. Algumas podem complementar outras, outras poderiam suplementar e reforçar outras. Então, é mais o caso de cooperar do que competir.

## 2.6. Medida de Probabilidade de Eventos Difusos

Um evento difuso pode ser caracterizado como um evento em que sua função indicadora a um certo conjunto pode assumir valores no intervalo unitário  $[0,1]$ . Assumindo os valores 0 ou 1 em casos extremos. Dessa forma um evento pode ter diferentes funções indicadoras (ou graus de pertinência) para diferentes conjuntos. Esse tipo de incerteza não é caracterizada pela aleatoriedade e se está trabalhando no primeiro nível de incerteza. A incerteza de pertinência, tratada em teoria de conjuntos. Em um segundo nível pode-se definir uma medida de crença dessa pertinência, sendo portanto possível se falar em probabilidade de um evento "crisp" como também em probabilidade de um evento "difuso".

Uma outra caracterização é dada pelo grau de incerteza da medida de crença. Quando se fala em alta probabilidade de um evento, caracteriza-se uma medida de probabilidade difusa. Admite-se finalmente, falar em probabilidade difusa de um evento difuso. Em linguagem corrente, em especial na língua portuguesa, há uma confusão, ou indistinção, quando se fala em verossimilhança e probabilidade, porque a primeira não é de uso corriqueiro. Então quando se diz que se um céu de inverno estrelado no sul do Brasil, em uma noite muito fria sugere que a probabilidade de geada na manhã seguinte é alta, de fato está se falando que é muito verossímil que haja geada na manhã seguinte.

Depois de seu primeiro artigo, "Fuzzy Sets", ZADEH (1965) equacionou o relacionamento entre as teorias em 1968, em um artigo intitulado "Probability measures of fuzzy sets". Segundo Zadeh, tal como definido em 2.5.2 acima, um evento é uma coleção de pontos precisamente especificada no espaço amostral. Em contraste, na experiência diária freqüentemente encontram-se situações nas quais um "evento" é uma coleção de pontos mais "difusos" do que "crisp", por causa da imprecisão do significado de palavras como: dia "quente",  $x$  é "aproximadamente" igual a 5 ou em vinte jogadas de uma moeda saem "algumas" caras mais do que coroas. "Usando o conceito de conjunto difuso, as noções de um evento e sua probabilidade podem ser estendidos de uma maneira natural para eventos difusos. É possível que tal extensão possa eventualmente fazer crescer significativamente o domínio de aplicabilidade da Teoria de Probabilidade, especialmente para aqueles campos, onde a "difusividade" é um fenômeno penetrante ZADEH (1968). O autor apresenta então as definições básicas de probabilidade de um evento difuso como o valor esperado de sua função de pertinência, e diz parecer que existem vários conceitos e resultados em Teoria de Probabilidades, Teoria da Informação e campos correlatos que admitem tais generalizações.

A partir daí muitas generalizações do relacionamento entre as duas teorias foram feitas. Na década de setenta, importantes trabalhos que buscavam relacionamento entre as duas teorias foram publicados. OKUDA ET AL. (1978) apresentam problemas de decisão do mundo real, estados da natureza, informação e ações como eventos difusos. A aplicação da Teoria de Conjuntos Difusos e da Teoria de Decisão Estatística para problemas de decisão com eventos difusos leva a uma formulação específica de problemas de decisão difusa e as definições de entropia, quantidade de informação e custo da informação em

ambiente difuso. Na esteira do relacionamento entre as duas teorias sobre incerteza KWAKERNAAK (1978 a e b) introduz o conceito de variáveis aleatórias difusas como variáveis cujos valores não são números reais mas números difusos, e subseqüentemente redefinidos como um tipo particular de conjuntos difusos (números difusos). São definidos a esperança de uma variável aleatória difusa, funções características, probabilidades conectadas a variáveis aleatórias difusas e esperanças e probabilidades condicionais relacionadas com variáveis aleatórias difusas como imagens do conjunto difuso representando a variável aleatória difusa. Vários teoremas, alguns relacionados com variáveis aleatórias difusas independentes são provados. No segundo artigo são apresentadas aplicações, que necessitam de vários algoritmos para a computação das esperanças e probabilidades das variáveis aleatórias difusas também desenvolvidas no segundo artigo. Uma aplicação para problemas de diagnóstico médico é brevemente discutida, onde a imprecisão na caracterização da extensão que um paciente tenha contraído a doença. São assumidas três possíveis caracterizações - fortemente doente, mediamente doente e não doente - que podem ser representadas por funções de pertinência. Também PURI AND RALESCU (1985, 1986) e RALESCU (1996) discutem os conceitos de variáveis aleatórias difusas, o seu conceito de normalidade e desigualdades para variáveis aleatórias difusas, que possibilitam a avaliação da função poder de testes estatísticos baseados em dados difusos.

GAINES (1978) mostra que é possível enxergar a lógica estocástica e a lógica difusa como derivadas de duas diferentes restrições da lógica probabilística: independência estatística (estocástica) e implicação lógica (difusa). A possibilidade de inferência bayesiana difusa é discutida por STALLINGS (1977) e TRIBUS (1979), mais tarde nas décadas de 80 e 90 essa discussão é aprofundada como pode-se ver um pouco mais adiante. NGUYEN (1977) discute o conceito de probabilidade lingüística introduzido por Zadeh, explorando sua relação com difusividade em teoria da medida, com também examina a aplicação do princípio da extensão a operações no espaço das funções distribuição de probabilidades.

Em 1978, Zadeh apresenta uma nova teoria baseada em conjuntos difusos, como uma tese que a imprecisão que está intrínseca nas linguagens naturais é, no principal, mais possibilística do que probabilística em natureza. Então, empregando o conceito de

distribuição de possibilidades, um proposição  $p$ , em uma linguagem natural pode ser traduzida em um procedimento que computa a distribuição de probabilidades de um conjunto de atributos que são implicados por  $p$ . Apresenta dessa forma a Teoria de Possibilidades (ZADEH, 1978).

Outras referências complementares podem ser encontradas em KLEMENT ET AL. (1981) para medidas de probabilidade difusa. YAGER (1986) apresenta valores esperados para probabilidades de subconjuntos difusos. As medidas de probabilidade de eventos difusos são discutidas por PIASECKI (1985) e TANAKA ET AL. (1993). ROMER AND KANDEL (1995) discutem variáveis aleatórias difusas e apresentam duas leis de grandes números, que vai introduzir a possibilidade de discussão de inferência estatística difusa. LIANG AND SONG (1996) propõem um modelo estatístico para melhor entender a interpretação probabilística para conjuntos difusos. O modelo mostra o relacionamento entre dois tipos de estatística: a estatística conjunto-valorada para conjuntos difusos e a estatística probabilística convencional para funções de pertinência.

Relacionados com as teorias de incerteza estão duas questões adicionais que têm sido tratadas tradicionalmente por ramos da ciência que tiveram desenvolvimento paralelo ao da Teoria de Probabilidades. São a análise exploratória de dados e a inferência estatística através da estimação de parâmetros e testagem de hipóteses. Também nesses campos existem contribuições importantes de como organizar dados e fazer inferências podem ser desenvolvidas com o auxílio da Teoria de Conjuntos Difusos.

O trabalho de DUBOIS AND PRADE (1986) sobre conjuntos difusos e dados estatísticos procura descobrir uma forma como dados estatísticos e a Teoria de Possibilidades podem ser casados. Como resultado, são desenvolvidos procedimentos para construir possibilidades fracas substituindo medidas de probabilidades e procedimentos para processar dados estatísticos imprecisos, que fornecem novas introspeções sobre o relacionamento entre conjuntos difusos e as teorias de probabilidades. Também KRUSE (1987), SCHNATTER (1991) e VIERTL (1995 e 1996) apresentam formas de se trabalhar com dados não precisos.

VIERTL (1990), SCHNATTER (1992) e ROMER AND KANDEL (1994) apresentam uma teoria completa para a realização de testes estatísticos para dados difusos.



Uma discussão especial é apresentada no artigo de Romer e Kandel, quando os autores defendem a natureza difusa dos testes de hipóteses estatísticos. Além da imprecisão na tomada de decisões a partir da determinação do nível de significância associado a um teste estatístico, que fornece um valor de confiabilidade não preciso, não se tem conhecimento do erro que se comete pela aceitação de uma hipótese nula quando ela não é verdadeira. Tal erro, medido pela probabilidade de aceitação do espaço sobe a suposição que a hipótese nula é verdadeira, é chamado de erro tipo II. Quando existe mais de uma hipótese alternativa, tal erro não é unicamente definido. Por essa razão a teoria de testes de hipóteses fornece o termo "poder do teste contra a hipótese alternativa", que voltará a ser discutido na construção do modelo de partição difusa do espaço amostral no Capítulo 4. Além do mais, todo procedimento de análise de resíduos, tão importante nos modernos métodos estatísticos, e que substituem os tradicionais testes de significância sobre as suposições realizados para a modelagem como em Análise de Regressão e Análise de Variância exigem um forte grau de poder de interpretação do especialista e que é portanto não preciso e depende fortemente de análises subjetivas.

Também ROMER AND KANDEL (1995) apresentam o conceito de partição difusa do espaço amostral e hipóteses paramétricas difusas. Conceitos que são apresentados com maiores detalhes no Capítulo 4 e que servem de base para a construção do modelo de partição difusa do espaço amostral que é empregado para a realização de meta-análise, objeto central do presente estudo. Fundamentalmente o conceito de hipóteses usualmente leva à partição do espaço amostral em dois subconjuntos complementares, chamados de espaço de aceitação e espaço de rejeição, respectivamente, de uma correspondente hipótese nula contra uma hipótese competitiva. O conceito de hipótese é visto então mais como uma predição imprecisa buscando o parâmetro desconhecido do que uma predição exata e com limites claros fornecendo asserções sobre sua localização no espaço paramétrico.

Uma discussão sobre a necessidade de se desenvolver uma Inferência Bayesiana-difusa é realizada por PIASECKI (1986), SCHNATTER (1993), SPIEGELHALTER AND FREEDMAN (1988), TRIBUS (1980), VIERTL (1987), VIERTL AND HULE (1991), YANG AND KLIR (1995), KRUSE (1982), PAN ET AL. (1996) e PAN AND KLIR (1996) e não são objetos de maior aprofundamento nesse trabalho.

## 3 Meta-Análise

### 3.1 Introdução - Uma nota histórica

Combinar informações de dados coletados sob diferentes condições e mesmo com diferentes níveis de precisão para produzir conclusões mais fortes do que aquelas disponíveis em cada fonte de informação. Esta atividade tem uma longa história. STIGLER (1986) descreve a criação dos mínimos quadrados por Legendre em 1805 e Gauss 1809 como uma tentativa de resolver o problema do uso de observações astronômicas coletadas de diferentes observatórios para estimar a órbita de cometas e determinar arcos meridianos em Geodesia.

Durante o século XX cientistas em vários campos combinaram informações através de experimentos para aumentar a acurácia de estimativas ou do poder de significância de testes estatísticos. Em 1904, Pearson, tentou determinar a relação entre mortalidade e inoculação com uma vacina para uma certa febre, usando dados de cinco amostras pequenas independentes. Ele computou estimativas da correlação entre inoculação e mortalidade em cada uma das amostras e então estimou uma função de correlação média entre as amostras, uma vez que as amostras individuais eram muito pequenas para permitir que qualquer opinião definitiva pudesse ser formada, tendo em vista o tamanho provável do erro envolvido. Pearson usou essa estimativa combinada para avaliar a eficácia da vacina, comparando-a com valores típicos de correlações similares computadas de outras vacinas. GAVER ET AL. (1992).

GAVER ET AL. (1992) informam que em ciências físicas, esforços para estimar os valores de constantes físicas pela combinação de resultados através de experimentos também começou cedo nesse século. Em 1920, Clark produziu uma determinação muito influente de pesos atômicos para cada elemento separadamente - reunindo todos os experimentos que mediram o peso atômico para cada elemento - e então combinou as estimativas por mínimos quadrados ponderados. Em 1932, Birge estava entre os primeiros a combinar estimativas através de experimentos de diferentes laboratórios para

estabelecer valores de referência para as constantes fundamentais em física. Ele tratou o problema da relação funcional não-linear entre as constantes usando mínimos quadrados ponderados em uma forma linearizada de estimação de equações, que eram resolvidas simultaneamente.

DICKERSIN AND BERLIN(1992) citam como trabalhos importantes na combinação de resultados de experimentos agrônômicos os trabalhos realizados em 1931 por Tippett, em 1932 por Fisher, em 1933 por Pearson, em 1937 por Cochran e em 1938 por Yates e Cochran. Nesses trabalhos o enfoque era a apresentação de métodos para combinar níveis de significância ("p-values") de diferentes estudos. Outros estudos que recebem menção são os estudos realizados por Mosteller em 1948, Mosteller e Bush em 1954 e Jeffreys em 1961.

Embora a busca de síntese de evidências através de métodos estatísticos tenha se dado ao longo da história nos últimos dois séculos, o termo meta-análise foi cunhado por GLASS (1976) no seu artigo "Primary, secondary and meta-analysis of research" publicado na *Educational Researcher*, em que define o conceito de meta-análise como uma técnica específica estimulada pelo uso de métodos de síntese quantitativa por cientistas sociais, particularmente por psicólogos.

Somente mais tarde a meta-análise foi descoberta pela Medicina, Saúde Pública e campos de pesquisa para serviços de saúde, como estudos epidemiológicos em JONES (1993). Em uma revisão da história da meta-análise nas áreas de saúde, DICKERSIN AND BERLIN (1992) descrevem os primeiros importantes artigos na área. O número de trabalhos na área médica é hoje incomensurável, dada a enormidade de pesquisas publicadas nas diversas revistas médicas, com resumos publicados nos mais importantes bancos de dados disponíveis na Internet como o MedLine e o BioMedNet por exemplo.

Na seção 3.2 apresenta-se os principais métodos, clássico e bayesiano para a realização da meta-análise. Os métodos baseados no nível de significância, chamados de Omnibus testes são apresentados em 3.3. A discussão dos métodos exploratórios e analíticos são apresentados na seção 3.4. Os principais problemas encontrados em meta-análise, são discutidos em 3.5. E em 3.6 são feitas considerações de ordem prática e teórica para a realização da meta-análise, com a apresentação de diversas aplicações.

### 3.2 Enfoque geral para a Meta-Análise

O termo meta-análise tem sido adotado de forma mais definitiva nos recentes trabalhos que buscam fazer combinação de informações entre diferentes estudos. No entanto outros termos ainda são utilizados para descrever o conjunto de técnicas com esse mesmo objetivo. Meta-análise ou síntese quantitativa de pesquisas ou achados são termos com largo emprego em Ciências Sociais, Educação e Psicologia. Em medicina, apesar do crescimento do uso do termo meta-análise, o uso de métodos estatísticos para combinar estimativas de pesquisas clínicas que investigam o efeito dos mesmos tratamentos é freqüentemente chamado de "agrupamento de resultados" ou "criação de uma revisão ("overview)". Em Física, encontra-se o termo "revisão de resultados de pesquisas". Em Química e Termodinâmica, o processo de gerar valores de referência para constantes físicas usando os resultados de existentes estudos empíricos é chamada de "avaliação crítica".

Na área médica a meta-análise freqüentemente combina os resultados de estudos comparativos tais como pesquisas clínicas aleatorizadas, estudos de coorte e estudos de caso-controle. Nesses estudos, algum resultado  $Y$  (por exemplo, um índice de mortalidade, a proporção de pacientes que mostram melhora ou a mudança de percentagem em uma escala quantitativa) é medida nos grupos controle e experimental.

Considerando uma série de  $k$  estudos independentes, os dados para o  $i$ -ésimo estudo são sumarizados pela estatística  $d_i$  ( $i = 1, \dots, k$ ), uma estimativa consistente do efeito verdadeiro do tratamento  $\delta_i$  para o estudo  $i$ . O parâmetro  $\delta_i$  pode ser definido de várias formas, dependendo do tipo de resultados e da informação disponível para o estudo  $i$ . Por exemplo, para resultados contínuos pode-se usar a diferença entre a média do grupo experimental e a média do grupo controle,  $\delta_i = \mu_{ti} - \mu_{ci}$ , ou a diferença padronizada das médias (também chamada de "effect size"),  $\delta_i = (\mu_{ti} - \mu_{ci})/\sigma_{ci}$ , onde  $\sigma_{ci}$  denota o desvio padrão do grupo controle.

"Effect sizes" são mais úteis quando os resultados dos estudos são obtidos através de diferentes instrumentos e medidos de maneira contínua. "Effect sizes" são também úteis como uma alternativa para se combinar "p-values" quando os autores de estudos relevantes reportam somente os testes estatísticos que obtiveram. ROSENTHAL (1984)

fornece alguns exemplos do relacionamento entre testes de significância e "effect sizes". COHEN (1977) fornece as linhas gerais para a interpretação dos termos "pequeno", "médio" e "grande" no contexto de "effect size".

Para resultados dicotômicos  $Y$ , se uma distribuição binomial ou hipergeométrica é assumida, onde  $\pi_{ti} = P(y = 1/ \text{tratamento})$  e  $\pi_{ci} = P(y = 1/ \text{controle})$ , o efeito populacional do tratamento é usualmente definido em uma das seguintes maneiras:

- A diferença de razões:  $\delta_i = \pi_{ti} - \pi_{ci}$
- O risco relativo  $RR_i = \delta_i = \pi_{ti}/\pi_{ci}$
- O "odds ratio"  $OR_i = (\pi_{ti}/1-\pi_{ti})/(\pi_{ci}/1-\pi_{ci})$
- O logaritmo do "odds ratio"  $\delta_i = \log(OR_i)$

### *Modelos de efeito fixo*

Um das questões importantes em meta-análise é a verificação da homogeneidade dos estudos. A escolha de uma medida comum para se realizar a síntese é ponto chave da análise. Uma das propostas e que será discutida em seção subsequente é o agrupamento dos níveis de significância. Outra é utilizar estudos com a mesma medida. Por exemplo, se os estudos puderem ser comparados e sumarizados por uma estatística  $d_i$ , como a diferença entre índices de mortalidade ou riscos relativos para se comparar dois ou mais tratamentos, a meta-análise pode ser realizada de uma forma simples. Supondo que  $p_{it}$  represente o índice de mortalidade para cada estudo  $i$ , com  $i$  variando de 1 a  $k$ , para o tratamento  $t$ , variando de 1 a 2. Então as diferenças  $(p_{i1} - p_{i2})$  representadas por  $d_i$  serão as diferenças entre os índices de mortalidade entre os dois tratamentos. Se os estudos forem suficientemente homogêneos em seus delineamentos e protocolos de amostragem dos pacientes, de forma que estejam medindo a mesma coisa, pode-se construir um modelo estatístico que pode ser expresso por

$$d_i = \delta + e_i \quad (3.2.1)$$

onde  $e_i$  é o erro estocástico com uma certa distribuição, freqüentemente assumida como normal com média 0 e variância  $V_i$ , de tal forma que os  $e_i$  sejam independentes. Na meta-análise freqüentista esse modelo é chamado de modelo de efeitos fixos (EF), e assumindo

por simplicidade que as variâncias sejam conhecidas, esse modelo implica em um regra de agrupamento baseada nos mínimos quadrados ponderados, que fornece a seguinte estimativa conjunta para o efeito

$$\hat{\delta}_w = \frac{\sum_{i=1}^k w_i d_i}{\sum_{i=1}^k w_i} \quad (3.2.2)$$

onde  $\hat{\delta}_w$  é tomado como a média ponderada dos  $d_i$ , na qual os pesos  $w_i$  são proporcionais a  $1/V_i$ , a informação no  $i$ -ésimo estudo. Assim  $V(\hat{\delta}_w) = 1/\sum w_i$ . Uma vez que as variâncias são usualmente desconhecidas, estimativas apropriadas para os  $w_i$ 's devem ser encontradas. Nesse caso a variância de  $\hat{\delta}_w$  pode ser maior que  $1/\sum \hat{w}_i$  por causa da variabilidade crescente dada pela estimativa dos pesos. No entanto, no caso simples onde  $d_i = \bar{Y}_i$  é a média das  $n_i$  observações do resultado  $Y$  e  $V(Y) = \sigma^2$  para todo  $i$ ,  $V(d_i) = V_i = \sigma^2/n_i$ . Aqui não importa que  $\sigma^2$  seja desconhecido uma vez que

$$\hat{\delta}_w = \frac{\sum_{i=1}^k n_i \bar{Y}_i}{\sum_{i=1}^k n_i} \quad (3.2.3)$$

é ótimo mesmo com  $\sigma^2$  desconhecido. Quando os modelos para meta-análise são mais complicados estimativas dos  $w_i$ 's devem acompanhar as suposições realizadas.

A suposição de homogeneidade,  $H_0 = \delta_1 = \delta_2 = \dots = \delta_k$ , é comumente testada através da estatística proposta por Cochran em 1938:

$$Q_w = \sum_{i=1}^k w_i (d_i - \hat{\delta}_w)^2 = \sum_{i=1}^k (d_i - \hat{\delta}_w)^2 / \sigma_i^2 \quad (3.2.4)$$

No caso mais simples apresentado acima, se os  $d_i$ 's são distribuídos normalmente,  $Q_w$  tem uma distribuição qui-quadrado com  $(k - 1)$  graus de liberdade sob  $H_0$ . Testes de homogeneidade para definições mais complicadas de  $d_i$  têm definições análogas. (DERSIMONIAN & LAIRD, 1986).

### *Modelos de efeito aleatório*

Se ao contrário concluir-se que os estudos são heterogêneos, as diferenças nos  $d_i$  provêm tanto do erro experimental como de diferenças reais entre os estudos (tais como as variações nos protocolos). Se informações explícitas da fonte de heterogeneidade são disponíveis, pode ser possível contabilizar essas diferenças no nível dos estudos. Mas na ausência de tal informação um modo útil e simples de expressar a heterogeneidade entre estudos é através do modelo hierárquico:

$$\begin{aligned}d_i &= \delta_i + e_i \\ \delta_i &= \delta + \varepsilon_i\end{aligned}\tag{3.2.5}$$

onde os  $e_i$  tem distribuição normal com média 0 e variância  $V_i$  e os  $\varepsilon_i$  tem também distribuição normal com média 0 e variância  $\tau^2$ , e todos os  $e_i$  e  $\varepsilon_i$  são considerados independentes. Esse modelo é conhecido em meta-análise freqüentista como modelo de efeito aleatório (EA). Em contraste com o modelo em 3.2.1, os modelos EA concebem os estudos observados como tendo sido retirados de uma grande população de estudos. O modelo de EF é um caso especial do modelo em 3.2.3 com  $\tau = 0$ . Nesse caso, todos os  $\delta_i$  são constantes e iguais a  $\delta$ .

A segunda equação no modelo hierárquico fornece uma perspectiva interessante na estimação dos efeitos subjacentes de  $\delta_i$  no estudo  $i$ . Essa equação implica que existe informação sobre  $\delta_i$  em todos os estudos, e não somente no estudo  $i$ , dessa forma a simples estimativa de  $d_i$  pode ser melhorada pelo "empréstimo de força" de outros estudos. Em sua forma mais simples emprestar força envolve o cálculo de uma média linearmente ponderada de  $d_i$  e a média geral ponderada  $\hat{\delta}_w$  de tal forma que:

$$\hat{\delta}_i = (1 - \hat{C}_i)d_i + \hat{C}_i \hat{\delta}_w.\tag{3.2.6}$$

O estimador 3.2.6 é um exemplo da classe de procedimentos chamado de estimativa empírica de Bayes ("Empirical Bayes") que tem tido seu uso aumentado nos últimos anos. Alguns importantes trabalhos com Enfoque Bayesiano e Modelos Hierárquicos são: CARLIN (1992), CONSONNI AND VERONESSE(1995), DRAPER (1995), DUMOUCHEL AND HARRIS (1983), DUMOUCHEL AND GROER (1989),

DUMOUCHEL (1990), DUMOUCHEL (1994), GUREVITCH ET AL. (1992), GUREVITCH ET AL. (1993), MOSTELLER AND CHALMERS (1992), TWEEDIE ET AL. (1994), VERDINELLI ET AL. (1996), YUSUF ET AL. (1985) e YUSUF ET AL. (1994)

Para os modelos de efeito aleatório, o valor esperado de  $\delta$  dos  $\delta_i$ 's é também estimado como uma média ponderada dos efeitos dos tratamentos:

$$\hat{\delta}_w^* = \frac{\sum_{i=1}^k w_i^* d_i}{\sum_{i=1}^k w_i^*} \quad (3.2.7)$$

onde  $w_i^* = 1 / (V_i + \hat{\tau}^2)$  e  $V_i$  estima  $E(d_i - \delta_i)^2$ . Aqui  $w_i^*$  estima o peso ótimo, que é dado pela inversa da variância marginal de  $d_i$ ,  $V(d_i) = V_i + \tau^2$ .

Observe-se aqui que as estimativas de  $\delta$  para os modelos de efeito fixo e efeito aleatório diferem somente na escolha dos pesos  $w_i$ . Essas estimativas de  $\delta$  freqüentemente não são diferentes. O erro-padrão  $EP(\hat{\delta})$ , no entanto, é muito sensível para o método de agrupamento e modelo escolhidos. Em geral, os modelos de efeito aleatório produzem estimativas maiores de  $EP(\hat{\delta})$  do que os modelos de efeito fixo porque  $\tau^2$  é tipicamente diferente de zero. Os erros-padrões do modelo de efeito fixo não conta para a heterogeneidade dos estudos e então o verdadeiro valor dos erros-padrões é subestimado quando a heterogeneidade está presente. Para valores grandes de  $\tau^2$  o erro-padrão pode crescer de forma dramática. DERSIMONIAN AND LAIRD(1986) comparam quatro enfoques para estimar a variância entre estudos: máxima verossimilhança, máxima verossimilhança restrita e dois estimadores não iterativos baseados no método dos momentos. Um dos estimadores não iterativos recomendado é

$$\hat{\tau}^2 = \max\left\{0, \frac{Q_w - (k-1)}{\sum w_i - \frac{\sum w_i^2}{\sum w_i}}\right\} \quad (3.2.8)$$



A maneira na qual os termos "efeito fixo" e "efeito aleatório" são usados na literatura de meta-análise é algo diferente das definições clássicas usadas em outras técnicas da Estatística tais com análise de variância, onde "efeito fixo" é um termo requerido para negar o conceito de um distribuição dos efeitos verdadeiros,  $\delta_1, \dots, \delta_k$ , e "efeito aleatório" supõe que os  $\delta_i$  's são amostradas de populações e portanto têm uma distribuição.

Os "modelos mistos" (HEDGES,1990) combinam as idéias dos modelos EF e EA. Nesses modelos, não se assume que todos os  $\delta_i$  's são iguais. Algumas variações entre estudos são vistas como explicáveis e outras aleatórias. STRAM (1996) descreve o uso de um modelo de regressão linear com efeitos mistos para a meta-análise de dados publicados.

### 3.3 Métodos baseados no nível de significância ("p-value")

O primeiro teste para combinar significância baseado nos "p-values" foi proposto por TIPPETT(1931). Para tanto,  $p_1, \dots, p_k$  são níveis de significância independentes, cada um tendo distribuição uniforme sob  $H_0$ . Tippett propôs um teste baseado em  $p_{[1]}$ , o mínimo entre os "p-values" ( a estatística de primeira ordem de  $p_1, \dots, p_k$ ). Para obter um teste das hipóteses nulas combinadas ao nível  $\alpha$ , ele propôs o seguinte procedimento de teste

$$\text{Rejeitar } H_0 \text{ se } p_{[1]} < 1 - (1 - \alpha)^{1/k} \quad (3.3.1)$$

Uma generalização do procedimento de Tippett é dada por WILKINSON (1951), que sugeriu usar os  $r$ -ésimos menores valores dos "p-values". Então se  $p_{[1]} < p_{[2]} < \dots < p_{[k]}$  são estatísticas de ordem de  $p_1, \dots, p_k$ , o procedimento de Wilkinson ao nível  $\alpha$  toma a forma:

$$\text{Rejeitar } H_0 \text{ se } p_{[r]} < C_{\alpha,k,r} \quad (3.3.2)$$

Onde  $C_{\alpha,k,r}$  é uma constante que pode ser obtida da distribuição beta. A vantagem desse método para  $r > 1$  é que ele não depende de resultados mais extremos. Conseqüentemente ele pode ser mais resistente aos efeitos de conjuntos de dados com "outliers".

A combinação mais fortemente utilizada foi a proposta por Fisher em 1932, e de uma forma um pouco diferente por Pearson em 1933. O método de Fisher usa o produto  $p_1 \cdot p_2 \dots p_k$  para combinar os "p-values". Para obter valores críticos da estatística proposta, Fisher determinou que:

$$-2 \log (p_1 \cdot p_2 \dots p_k) = \sum_{i=1}^k -2 \log p_i \quad (3.3.3)$$

Quando a hipótese nula  $H_0$  for verdadeira, cada  $p_i$  tem distribuição uniforme e  $-2 \log p_i$  tem distribuição qui-quadrado com 2 graus de liberdade, então a soma tem uma distribuição qui-quadrado com  $2k$  graus de liberdade. O procedimento de teste recomendado por Fisher se torna:

$$\text{Rejeitar } H_0 \text{ se } -2 \sum_{i=1}^k \log p_i > C_\alpha \quad (3.3.4)$$

Onde  $C_\alpha$  é o  $100(1 - \alpha)$  ponto percentil da distribuição qui-quadrado com  $2k$  graus de liberdade.

Um outro método vastamente utilizado para combinar "p-values" é o método da normal inversa proposto por Stouffer et al. em 1949 e descrito por HEDGES & OLKIN (1985) e GAVER ET AL. (1992). O procedimento envolve transformar cada  $p_i$  utilizando a função distribuição normal cumulativa e então tirar a média. Especificamente define-se  $z_i = \Phi^{-1} (1 - p_i)$ , onde  $\Phi(x)$  é a função distribuição normal cumulativa. Sob  $H_0$  a estatística  $z = (z_1 + \dots + z_k) / \sqrt{k}$  tem distribuição normal padrão. Então um teste de combinação ao nível  $\alpha$  é:

$$\text{Rejeitar } H_0 \text{ se } z > C_\alpha \quad (3.3.5)$$

Onde  $C_\alpha$  é um valor crítico da distribuição normal padrão.

Outros métodos baseados no produto  $(1 - p_1) \dots (1 - p_k)$  e na soma dos  $p_i$ , também chamados de "omnibus tests" são apresentados e criticados por HEDGES & OLKIN (1985) e GAVER ET AL. (1992). Nenhum teste não-viciado e uniformemente mais poderoso baseado em combinações dos "p-values" existe sem maiores especificações do

modelo para testes individuais. Conseqüentemente comparações entre procedimentos de combinação são restritos a modelos específicos.

Também procedimentos chamados de "vote-counting" ou "box-score" são apresentados por HEDGES & OLKIN (1985) e mesmos testes baseados na construção de estimadores apesar de simples e com grande apelo intuitivo têm, segundo os autores acima severas limitações que restringem sua aplicação na prática, e que portanto não serão considerados no escopo desse trabalho.

Um estudo comparativo entre os principais métodos apresentados nesse capítulo com o modelo difuso proposto é realizado no Capítulo 5 de aplicação.

### 3.4 Meta-análise exploratória e analítica

Meta-análise tem sido definida como a análise estatística que combina ou integra os resultados de diferentes estudos independentes. No entanto, seu uso inclui: aumentar o poder de um teste estatístico, melhorar as estimativas de efeitos, conflitar conclusões ou fazê-las ter sentido e achar e responder novas questões. Esses usos não refletem especificamente a natureza secundária da meta-análise, isto é, de ser um estudo onde as unidades de observação são os resultados obtidos de outros estudos, que são chamados de estudos primários. Ao contrário do investigador de estudos primários, o meta-analista está restrito à disponibilidade de estudos publicados e não publicados.

Um estudo observacional tem duas características principais, a primeira é que o objetivo é estudar efeitos causais de certos agentes, procedimentos ou programas e o segundo é que o investigador não pode usar experimentação controlada (COCHRAN,1983). Uma vez que a meta-análise tem as duas características, ela deve ser vista como estudo observacional onde as unidades a serem observadas são outros estudos.

Na área médica a meta-análise poder ter como estudos primários, experimentos clínicos aleatorizados ou estudos epidemiológicos. (PETO,1987), (FLEISS,1985). Segundo ANTMAN (1992) quando as unidades são estudos epidemiológicos, a meta-análise está mais sujeita a vícios, o que não significa que para experimentos aleatorizados não possam haver também alguns problemas. E a questão central a ser determinada em

ambos os casos é qual o principal objetivo da realização da síntese. Essa distinção pode ser importante quando o objetivo é estimar um efeito comum ou aumentar o poder estatístico ou ainda aumentar a precisão das estimativas dos efeitos. Quando os objetivos são classificar variações nas respostas ou levantar novas questões essa distinção não é fundamental. Fazendo um paralelo com os métodos estatísticos, ANELLO AND FLEISS (1995) procuram alterar o foco das preocupações da realização da meta-análise para a questão de ela melhorar as estimativas pontuais e reduzir os erros padrões de alguma medida de efeito de interesse ou de ela buscar causas da variabilidade dos efeitos medidos. BEARD ET AL. (1997) concentram a discussão nos problemas de buscar medir a variabilidade que existe entre estudos e entender as causas dessa variabilidade.

O primeiro objetivo acima exposto caracteriza o que os autores chamam *de meta-análise analítica*. Esse tipo de análise é caracterizado por aquelas revisões que buscam estimativas comuns através da estimação pontual com seus erros aleatórios. O segundo tipo de meta-análise é referida pelos autores como *meta-análise exploratória ou causal*, que está associada com a variação observada dos efeitos. Com base nessa orientação ANELLO AND FLEISS (1995) sugerem protocolos especiais dependendo do tipo de meta-análise a ser realizado.

Embora HOLZMANN (1995) discorde dos nomes sugeridos pelos autores acima, preferindo chamar o primeiro tipo de *descritivo* e o segundo de *analítico*, a questão central levantada por ambos trabalhos diz respeito a como trabalhar com a heterogeneidade entre estudos, seja na decisão do critério de inclusão e exclusão das pesquisas a serem incluídas, seja no modelo a ser adotado.

A importância dessa discussão no contexto da proposta de construção de um modelo fuzzy para meta-análise se dá pelo fato de que em se considerando o procedimento meta-analítico como observacional, em que existirá inevitavelmente uma heterogeneidade entre os estudos, mais forte ou menos forte dependendo do estudo a ser realizado, o procedimento deve incorporar essa fonte de incerteza. Em última análise, mesmo que o objetivo da meta-análise seja inferencial - para fugir da discussão epistemológica - o procedimento de meta-análise deve também explorar os estudos individuais, o que caracterizaria uma meta-análise descritiva ou exploratória.

No estudo inferencial o pesquisador busca sintetizar os achados dos diferentes estudos através de procedimentos indutivos, ou seja busca uma síntese através de procedimentos estatísticos usuais como cálculo de estatísticas pontuais para inferência, determinação de intervalos de confiança ou realização de testes de hipótese. Considerar os estudos individuais como elementos de uma amostra aleatória, com capacidade de ajudar a fazer inferências e previsões sobre determinados procedimentos médicos será discutido no próximo tópico. Mas do ponto de vista da metodologia que está sendo proposta, mesmo havendo essa distinção de objetivos, sob a perspectiva da realização do estudo eles são mais complementares que exclusivos. É essa a única forma de incorporar ao modelo os elementos de subjetividade retirados da opinião do especialista que pode ser baseada também na análise descritiva ou exploratória, para ajudar a construir um procedimento inferencial onde os possíveis vícios são explicitados e incorporados a esta etapa da meta-análise.

### 3.5 Problemas em Meta-análise

Embora os processos de meta-análise estejam bem desenvolvidos, ainda há uma série de questões levantadas que questionam as formas mais tradicionais de sua realização. Novas metodologias surgem a todo momento procurando superar os principais entraves para a realização de sínteses quantitativas. A discussão que segue objetiva apresentar esses problemas procurando alternativas em um modelo difuso para equacionar o problema e apontar sugestões de contorno. O desenvolvimento do modelo, levando em conta esses problemas é apresentado no Capítulo 5.

Os principais problemas apresentados são: (i) a exigência de precisão matemática; (ii) seleção dos estudos ou critérios de inclusão ou exclusão; (iii) vícios de publicação que gera o chamado "file drawer problem" ou problema de arquivamento de resultados; (iv) que estatística considerar e como combiná-las levando em conta o tamanho da amostra e (v) definição da população objeto que é em última análise a discussão da extensão do procedimento inferencial derivado da análise quantitativa para revisão de literatura.

### *A exigência de precisão matemática*

LIGHT AND PILLEMER (1984) definem meta-análise como "qualquer tipo de sumário quantitativo de dados". Colocado dessa forma, sem distinção entre as operações puramente matemáticas da estatística descritiva e as especiais ambições inferenciais da estatística indutiva, pode-se compreender a sua definição como uma composição dessas duas características. Se forem consideradas como duas etapas do procedimento meta-analítico supõe-se procedimentos que refletem simplesmente incertezas devidas ao acaso, em especial na segunda etapa. De uma forma ou de outra, os requerimentos de precisão matemática são impositivos de tal forma que possam garantir graus de crença objetivos em proposições com incerteza.

Segundo OAKES (1993) a intolerância à ambigüidade é parte da condição humana. No entanto, ele enfatiza que ela não é moralmente indefensável nem logicamente incoerente. Um dos temas que perpassa a meta-análise é a extensão que os julgamentos devem ser exercitados para que a análise estatística resultante seja válida ou que pelo menos tenha valido a pena. O autor sugere então, que se deve refletir antes de se fazer o elogio da "objetividade" da meta-análise e denegrir a "subjetividade" do enfoque alternativo da revisão de literatura. Essas digressões não consideram explicitamente qualquer enfoque intermediário entre a chamada "revisão de literatura narrativa e portanto, subjetiva" e a "meta-análise clássica".

Mesmo o enfoque Bayesiano, que admite em seus modelos matemáticos um salto para o exercício de um processo mais intuitivo de julgamentos pessoais, (PEARSON, 1962) - como a escolha de classes mais prováveis de hipóteses admissíveis, a apropriada interpretação de nível de significância, a magnitude de efeitos que valham a pena e o equilíbrio de utilidades - a medida de probabilidade ainda é um ente matemático preciso e objetivo.

Um modelo intermediário, que admita trabalhar com variáveis lingüísticas, incorporando o julgamento do especialista através de ferramentas matemáticas próprias para tratar com incertezas de outra natureza é fundamentalmente não preciso e com ponderável nível de subjetividade. O que não implica em ser menos útil ou correto, mas

simplesmente mais abrangente por buscar incorporar toda a complexidade dos fenômenos da vida real, sobre os quais se deseja opinar e ter como base para a tomada de decisões.

### *Seleção de estudos - o que incluir.*

Talvez o primeiro problema que se defronta alguém que deseja sumarizar um corpo de pesquisas é a identificação do corpo de trabalhos que de alguma forma dizem respeito às questões levantadas. Isso requer que o propósito da meta-análise seja claramente articulado. Algumas vezes o interesse recai na eficácia de uma particular formulação no tratamento de uma indicação muito específica, em outras ocasiões deseja-se procurar uma confirmação. Mas seja qual for o objetivo da revisão, a seleção dos estudos a serem integrados será o suporte da conclusão procurada.

A seleção é uma questão de inclusão ou exclusão, e os julgamentos requeridos são problemáticos. Existem várias posições com relação a que tipo de estudo deve ser incluído. Se só experimentos aleatorizados devem ser escolhidos, ou se pesquisas com pequena qualidade devem ser excluídos, ou se somente estudos com simples pontos finais devem ser analisados, ou se relatórios "secundários" devem ser incluídos. Uma grande atenção tem sido dada a esses problemas e não há um consenso dos pontos-de-vista. A tendência nos últimos anos tem sido a de adotar uma regra de decisão explícita e objetiva YUSUF ET AL. (1985). No entanto segundo OAKES (1993) é necessário lembrar-se que objetividade não é sinônimo de validade e que embora em um sentido formal a primeira seja condição necessária para a última, isso não significa que uma implica na outra.

Um dos problemas mais difíceis de serem tratados é o que fazer com pesquisas de qualidade duvidosa. FLEISS (1987) é plenamente favorável à inclusão de todos os estudos. Outros como PETO (1987) sugerem um requerimento mínimo para a entrada de um estudo, que ele seja uma pesquisa aleatorizada. O autor pondera que a inclusão de qualquer outro requerimento pode levar a dificuldades impossíveis de serem resolvidas. No entanto há um problema ético a ser considerado. ROYALL (1991) considerando o problema de que paciente deve receber qual tratamento, pondera que estudos que utilizam controles correntes históricos ou não-aleatorizados não têm menor ou nenhum valor científico. A questão é como considerá-los quando da construção do modelo estatístico.

CHALMERS (1991) foi quem mais esforço realizou para construir uma escala de qualidade de pesquisa com o propósito de ponderar os estudos a entrarem em uma meta-análise. Ainda segundo OAKES (1993), alguns cuidados devem ser tomados para se evitar o uso de definições derivadas de dados, mas que em última análise se tentada a adoção de qualquer regra de inclusão e exclusão pode-se deparar com dificuldades intransponíveis.

### *Vícios de publicação - "the file drawer problem"*

Os artigos de revistas científicas não são uma amostra representativa dos trabalhos que foram endereçados a qualquer particular área de pesquisa (SMITH, 1980). Achados de pesquisas que foram significativos, têm maior chance de serem publicados do que resultados não significativos. Dessa forma, qualquer método integrativo que recaia somente em dados publicados está sujeito a ser viciado.

Meta-análises têm procurado superar essa dificuldade tentando localizar pesquisas não publicadas e incorporando-as à sua análise. Ocorre que esse método pode tomar muitos anos de pesquisa e ainda assim se teria dúvidas que tal esforço reduziria o perigo de superestimar a eficácia de um dado tratamento. Um enfoque meta-analítico menos freqüente para o inevitável vício de publicação é o emprego de métodos estatísticos que contornem o problema.

ROSENTHAL (1979) é freqüentemente citado como tendo endereçado o problema de se arquivar artigos sem significância estatísticas em gavetas. O chamado "file drawer problem". Ele afirma que existe uma amostra viciada de estudos que são realmente consideradas para a consecução da meta-análise. A solução de Rosenthal é combinar os resultados experimentais fazendo uma soma de variáveis normal padrão. OAKES (1993) critica o método através da simulação de alguns resultados básicos apresentados por Rosenthal, e sugere o método de Halperin que sugere que tentativas exploratórias devem ser realizadas com uma perspectiva de análise de dados censurados que possam fazer descontos severos no "effect size".

Existem alguns métodos estatísticos que podem acessar retrospectivamente o risco de vício de publicação. Esses métodos estatísticos são delineados para testar a possibilidade da existência de alguma hipótese de pesquisa que reporta resultados



negativos, mas não são identificados pela revisão de literatura. Esses métodos só podem ser aplicados se a meta-análise fornecer um resultado positivo, estatisticamente significativo. MESSORI (1996) cita como exemplos de tais testes os métodos de L'Abbè, Klein et al. e o chamado "funnel plot method" devido a Vandembroucke.

Um aprofundamento da discussão pode ser feita com o auxílio das discussões contidas nos seguintes trabalhos: BEGG AND BERLIN (1988), DEAR AND BEGG(1992), EGGER ET AL. (1997), HEDGES (1992), HEDGES AND VEVEA (1996), HUNTER AND SCHMIDT (1990) e IYENGAR AND GREENHOUSE (1998)

MANTON ET AL. (1994), propõem em seu modelo para aplicações estatísticas utilizando conjuntos difusos uma modificação na função de verossimilhança para observações incompletas através da inclusão de um ponderador dos efeitos que estão sendo estudados. Na aplicação desse processo ao procedimento meta-analítico, o problema é a estimação desses ponderadores. O uso do julgamento de especialistas será a forma de contornar o problema dessa estimação.

### *Significância, e "effect size"*

Nos primeiros anos o principal resultado de uma meta-análise era a probabilidade associada com a testagem de uma hipótese ( o "p-value"). Mais tarde, maior atenção foi devotada à estimação do "effect size". Esse desenvolvimento foi uma consequência natural do fato de se estar ciente que para qualquer pesquisa, um pequeno valor de "p" pode ser obtido tomando-se uma amostra suficientemente grande. Por outro lado, o fato de em um único experimento clínico obter-se um resultado significativo não implica em absoluto que o resultado seja definitivamente significativo e mesmo que tenha importante relevância clínica. Pequenas mudanças no sorteio da amostra podem revelar importantes mudanças no valor do nível de significância, em particular para pequenas amostra. OAKES (1993) apresenta um exemplo com 250 pacientes asmáticos. Verifica que para um nível de significância de 5%, somente 134 pacientes podem ter contribuído para a eficácia de um certo tratamento. E mais surpreendentemente que o teste foi significativo a 1% quando o número de pacientes crescem para 137.

Por outro lado, o relacionamento entre o "effect size" - que não depende do tamanho da amostra - e o "p-value" não tem, no geral, um relacionamento linear, mas

assintótico. Um crescimento relativamente trivial no "effect size" pode resultar em um efeito dramático no "p-value".

Dessa forma, um modelo mais completo deveria incluir as três variáveis em questão, tamanho da amostra, "effect size" e "p-value", e de forma não absoluta. As considerações sobre a forma de se trabalhar com essas três variáveis incorporando-se questões como: se o experimento é aleatorizado ou não, a forma como o experimento foi delineado, as conseqüências de se considerar um maior ou menor erro tipo I e um maior ou menor erro tipo II na determinação do tamanho da amostra e a relação dessas determinações sobre o valor do nível de significância são considerados no tópico 4 do Capítulo 4, quando se apresenta a extensão do modelo GoM para a realização da meta-análise.

#### *Quem é a população - "problem of apples and oranges"*

A questão que se levanta nesse ponto não é diretamente com relação à superioridade de uma revisão quantitativa sobre uma revisão narrativa, mas entre a revisão quantitativa e o propósito de combinação inferencial.

Algumas condições devem estar postas, ou suposições devem ser assumidas para que uma inferência estatística de qualquer tipo possa ser propriamente legitimada. A principal característica da inferência estatística é o argumento da amostra para a população. A amostra deve ser resultado de um procedimento aleatório, e a população deve poder ser indistinguível. No caso da meta-análise, as duas características não são dadas diretamente.

Para análises paramétricas, as inferências realizadas - estimação por ponto e intervalo e testes de hipótese - para o efeito do tratamento, devem ser aplicadas para implícita ou explicitamente para pacientes "como esses" em circunstâncias "como essas". Fazendo isso suprime-se o fato que tais inferências não são estritamente estatísticas. Quando se considera a meta-análise, está-se tratando com dados observacionais, cujas unidades são os resultados de pesquisas individuais e que portanto não são objetos nem de aleatorização nem de processo de amostragem aleatória.

O coração do problema é em que extensão dados de diferentes estudos podem legitimamente ser combinados - o usualmente chamado "problem of apples and oranges" - para se tirar uma conclusão para uma dada população. Para GLASS (1978) essa resposta é simples: "maçãs e laranjas podem legitimamente ser combinadas no estudo de frutas". Ou seja, em estudo sobre a relevância de um dado tratamento, pode-se combinar estudos que tenham por objetivo diferentes estatísticas, como por exemplo, se a administração de aspirina para pacientes pós-infarto medirem diferentes resultados, como expectativa de vida ou risco relativo, a combinação através da diferença de médias padronizadas, do "p-value" obtido e do tamanho das amostras de cada estudo podem ser combinadas se puderem responder se é relevante para cada um dos estudos a administração da aspirina. A hipótese formulada não é propriamente estatística, mas que admite um certo grau de vaguesa na sua formulação.

Segundo OAKES (1993) estatísticos desde Fisher já abandonaram a aspiração de requerer que a população seja "real", mas reclama que a imprecisão da descrição das populações em meta-análise são alarmantes se confrontadas com a precisão das análises. Essa questão de fato é relevante. E dois caminhos podem ser apontados para a sua solução. Um é o de buscar fazer a meta-análise tendo estudos individuais bastante homogêneos, com os cuidados de que as suposições de amostra aleatória sejam cumpridas. Outra é o de admitir que o procedimento meta-analítico não necessariamente precise trabalhar com amostras aleatórias, e que as inferências a serem feitas sejam mais restritas e que também admitam um certo grau de vaguesa e imprecisão, através do trabalho com variáveis lingüísticas. De fato a simples determinação do "effect size" combinado de estudos, ou de um "p-value" combinado, podem não trazer contribuições importantes para as questões levantadas. Podem até não dizerem nada sobre os efeitos de um dado tratamento. Para efeitos clínicos, no diagnóstico ou prognóstico de certa intervenção, o que importa é se um tratamento é eficaz ou não, ou ainda em que grau pode-se acreditar que ele é eficaz. Quanto de "crença" um médico pode ter sobre o tratamento a ponto de aplicá-lo a certos pacientes. A que ponto os resultados obtidos são relevantes que sugiram a utilização de um dado tratamento.

### 3.6 Estado da Arte

Algumas questões adicionais têm sido colocadas para discussão no meio meta-analítico. São aspectos "secundários" do problema, não por serem menos importantes que os já comentados até esse ponto, mas pela sua especificidade. O objetivo de comentá-los aqui é o de ampliar o espectro da discussão, tomando-se como base as principais referências na área com maior número de citações em artigos publicados sobre a meta-análise.

Essas questões dizem respeito à história da meta-análise, métodos de implementação e construção de metodologia adequada para a síntese delineamentos de pesquisa, vantagens e desvantagens do procedimento, meta-análise de literatura versus meta-análise de pacientes, apresentação de aplicações em algumas áreas científicas, análise cumulativa, técnicas estatísticas para a consecução de revisão quantitativa para certos casos especiais e uma breve discussão dos "softwares" disponíveis para a realização da meta-análise. A apresentação das aplicações à Medicina se restringiram aqui a problemas gerais. Os exemplos nas diferentes especialidades médicas permeiam toda a discussão e são apresentados ao longo do trabalho. Dada a atual extensão da aplicação de meta-análise nas áreas médicas, optou-se por não apresentá-las no corpo do trabalho. Uma pesquisa na literatura ou em algum banco de dados citados pode cobrir o particular interesse de aplicação em alguma especialidade em particular.

#### *História*

Uma revisão histórica crítica mais sistematizada da meta-análise é apresentada no livro "How Science Take Stock. The History of Meta-Analysis", HUNT (1997). Que pode ser complementada pelos livros de LIGHT AND PILLEMER (1984), "Summing Up - The Science of Reviewing Research", cuja abordagem é mais metodológica que estatística e de WACHTER AND STRAFF (1990), "The Future of Meta-Analysis", que busca na revisão historiográfica, perspectivas para o avanço do procedimento. Em vários outros artigos e livros citados nas Referências Bibliográficas há extratos da história da meta-análise, alguns deles já citados em seções anteriores.

### ***Problemas Metodológicos de Implementação***

ABRAMI AND COHEN (1995) discutem os primeiros problemas para implementação de um processo de meta-análise, começando pela especificação dos critérios de inclusão, a localização dos estudos, a codificação das características a serem estudadas, o cálculo de resultados dos estudos individuais e a análise de dados.

A principal fonte de aquisição de informações sobre artigos a serem colecionados para a realização de revisão de literatura é a rede internacional. HEWITT (1991) em um importante artigo "Taming the Literature" ("Domesticando a Literatura") traça o planejamento básico para a realização de tal consulta apresentando uma forma sistematizada de assédio aos principais bancos de dados disponíveis na rede. LEFERING AND NEUGEBAUER (1991), NEUGEBAUER ET AL. (1991) e PAGLIARO (1997) apresentam métodos sistemáticos de preparar a meta-análise, discutindo linhas gerais de como organizar o material de referência.

As vantagens e desvantagens do uso da meta-análise são discutidas por: FAGARD ET AL. (1996) e KASSIRER (1992). E LECKY ET AL. (1996) e SCHECHTER ET AL. (1991) apresentam discussões sobre o mau uso da meta-análise.

LAVALLEY (1997) e OLKIN (1992) propõem em seus respectivos trabalhos um guia completo para a realização da meta-análise, que vão desde os aspectos de seleção e organização da literatura até a construção de algumas ferramentas estatísticas de análise.

Um debate sobre a confiabilidade foi realizado sob o nome: "Can Meta-analysis be Trusted" por importantes pesquisadores bio-estatísticos, apresentada nos artigos: CHALMERS (1991), STEWART ET AL. (1991) e THOMPSON AND POCOCK (1991)

### ***Meta-análise da Literatura e Meta-análise de Dados de Pacientes***

A informação apresentada na literatura que descreve os resultados de pesquisas clínicas é a principal fonte de dados clínicos necessários para a condução da meta-análise. Mas existe uma outra estratégia de recuperar essas informações clínicas através de contatos do pesquisador com os autores de cada uma das pesquisas clínicas e obter deles as informações de cada paciente inscrito em seus estudos. Essa segunda estratégia recai então em dados dos pacientes individuais como fonte de informação clínica. Essa

estratégia se aproxima da pesquisa clínica multi-centro. MESSORI (1996) cita alguns importantes trabalhos utilizando essa segunda estratégia.

### *Aplicações*

As aplicações de meta-análise se estendem a vários campos do conhecimento científico. Em alguns deles há artigos com aplicações de métodos já conhecidos, outros desenvolvem métodos propostos para fins específicos e outros apresentam métodos mais universais. Os trabalhos que seguem são de desenvolvimento de métodos mais universais com aplicações a algumas áreas. Não há qualquer pretensão de cobrir todos os artigos e todas as áreas de aplicação nem tão pouco discuti-los nesse trabalho. Como são referências básicas em suas áreas, pelo número de citações em outros artigos eles são apresentados aqui como meros guias para uma primeira abordagem.

Em *Meio-Ambiente*, os trabalhos de BLAIR ET AL. (1995) e OSENBURG ET AL. (1997) são guias gerais para a aplicação do método em Meio-Ambiente e Ecologia.

Na área de *Ciências Sociais*, onde o método foi primeiramente desenvolvido, além do artigo de GLASS (1976), o primeiro a usar o nome *meta-análise* existe uma enormidade de artigos, mas pela extensão e importância que tiveram os artigos de COOPER AND DOOR (1995) e DRAPER (1995) merecem referência a parte.

Também em *Educação*, importantes trabalhos são apresentados nas principais publicações da área, alguns artigos merecem destaque, principalmente por serem de épocas diferentes, o que pode dar uma idéia da evolução da meta-análise na área. GLASS (1978), RUBIN (1981) e GRAHAM (1994).

ROSENTHAL (1979) levantou o problema do "file drawer problem" nesse artigo com aplicações em *Psicologia*. Existem outros artigos e livros do mesmo autor na área. Uma outra aplicação em Psicologia com a preocupação de desenvolvimento de metodologia pode ser encontrado em SCHRAM (1996).

Um artigo importante por discutir a replicação e o relacionamento da **Parapsicologia** com a Estatística e o método meta-analítico é o artigo de UTTS (1991) acompanhado de um painel de comentários e discussão de alguns dos principais pesquisadores da área de meta-análise.

Em **Medicina**, mais particularmente em experimentos clínicos aleatorizados, principal fonte de trabalhos na área, os artigos de debate metodológico que servem de base para a discussão, com ampla citação bibliográfica são: ANTMAN ET AL. (1992), CHALMERS (1987), DETSKY ET AL. (1992), GELIJNS (1990), GOODMAN (1989), NONY ET AL. (1995), PETO (1987), RAVNSKOV (1992) e SACKS ET AL (1987).

### *Análise Cumulativa*

O estudo meta-analítico não pretende ainda ser definitivo. Novas peças de evidência podem ser acumuladas na medida em que novos estudos são realizados. ANTMAN ET AL. (1992) e LAU ET AL. (1992) desenvolveram métodos para examinar a relação temporal entre o acúmulo de dados e pesquisas controladas e aleatorizadas, sob o nome de meta-análise cumulativa. O processo consiste basicamente na realização de análises seqüenciais que acumulam novos resultados ao resultado da última meta-análise realizada.

### *Métodos Estatísticos*

Os métodos estatísticos desenvolvidos para a meta-análise permeiam toda a discussão realizada até agora. Alguns textos clássicos não citados anteriormente podem complementar uma bibliografia básica e introdutória dos métodos estatísticos apresentados. HEDGES (1981), HEDGES AND OLKIN (1985), LAW (1995), WEST (1993) e WACHTER AND STRAF (1990)

Em algumas áreas de aplicação específica, foram desenvolvidos métodos próprios para meta-análise. Como por exemplo para o modelo logístico ver COOK AND WALTER (1997), para meta-análise de risco relativo e odds ratio dois artigos específicos são: GART (1992) e MEHTA ET AL. (1985). Para ajuste de curva de sobrevivência ver GELBER ET AL. (1995) e para análise de opinião viciada de especialistas ver MARTZ AND BRYSON (1984). Além dessas referências e das citadas anteriormente, existe mais uma série de outras metodologias desenvolvidas para realizar meta-análise como em modelos lineares, regressão e anova e análise multivariada, no entanto por não terem correspondência direta ou indireta com esse trabalho não foram citadas.

### *Software*

Existem no mercado uma série de pacotes estatísticos destinados exclusivamente para a solução de problemas de meta-análise. NORMAND (1995) discute comparativamente, a partir do exemplo que servirá de aplicação para esse trabalho, três desses programas. O DSTAT, o TRUE EPISTAT e o FAST\*PRO. Todos três são para PC e são dirigidos por menu. Os dois primeiros são similares em suas concepções e só realizam análises de efeitos fixos. O FAST\*PRO é diferente dos outros dois e suplementam o usuário com análises de efeitos fixos e efeitos aleatórios. Uma cópia do FAST\*PRO acompanha o livro de EDDY ET AL. (1992), que apresenta seu "Confidence Profile Method" para a meta-análise com aplicações naquele software.

Outro pacote específico para a consecução da meta-análise é o EasyMA apresentado por CUCHERAT (1997). O programa pretende ajudar a realização de pesquisas médicas com resultados dicotômicos, e permite a realização dos vários delineamentos de meta-análise inclusive a cumulativa.

Também os programas SAS, GLIM e MINITAB permitem a realização de meta-análise com certas adaptações. Um programa para ambiente UNIX desenvolvido por DUMOUCHEL (1994) da BBN Software Products Corporation permite a realização de meta-análise para modelos lineares bayesianos hierárquicos.



## 4 Modelo Difuso para Meta-Análise

### 4.1 Introdução

A motivação original para a construção de modelos estatísticos com conjuntos difusos deriva de problemas de diagnóstico formal para doenças crônicas, especialmente diagnósticos Bayesianos de doenças crônicas. Os procedimentos Bayesianos, que assumem pequeno número de processos discretos de doenças, frequentemente têm sido ineficientes no problema de diagnóstico formal para doenças crônicas, especialmente aquelas com histórias naturais prolongadas em populações idosas (WOODBURY, 1963). As doenças interagem com outras doenças, e com a fisiologia dos pacientes ( e mudanças nas doenças e na fisiologia do paciente com o tempo). Então os diagnósticos produzem manifestações sintomáticas que quase sempre têm elementos únicos para cada elemento - pelo menos em alguma amostra de tamanho razoável.

MANTON ET AL. (1994) apresentam métodos para modelar dados de resposta discretas com alta dimensão usando a teoria de conjuntos difusos e partições. A maioria dos métodos estatísticos são baseados na suposição que conjunto de pessoas ou objetos são organizados em conjuntos "crisp". Isto é, uma observação vem de um conjunto identificado com características específicas ou valores de parâmetros (ou não). A estimação das características do conjunto, diante de incerteza no processo de observação, é a tarefa dos estatísticos ou do analista de dados. Quando não se faz a suposição que os objetos têm pertinência "crisp" em um conjunto (i. é, tudo ou nada), mas permite um conjunto de pertinência difusa (i.e., uma pessoa pode ter graus de pertinência em vários conjuntos), é possível usar famílias mais gerais de modelos de probabilidade para modelar as observações. Uma vez que as suposições teóricas dos conjuntos clássicos "crisp" podem ser vistas como casos especiais de suposições usadas na teoria de conjuntos difusos, muitos métodos estatísticos padrões em uso podem também ser mostrados como casos especiais de métodos estatísticos baseados em conjuntos difusos - ou, inversamente, que métodos baseados em conjuntos difusos generalizam modelos estatísticos padrões baseados na

lógica dos conjuntos "crisp" no sentido de melhor tratar com fontes de heterogeneidade nos dados adicionais e mais complexas.

O uso de métodos estatísticos com "partição difusa", baseados nos princípios matemáticos da teoria de conjuntos difusos, foram usados como meio de analisar dados com alta dimensionalidade com respostas categóricas. Os procedimentos estatísticos como estimação, teste de hipóteses e bondade de ajustamento foram baseados nos princípios de máxima verossimilhança adaptados para a matemática e a lógica de conjuntos difusos. (p.ex. KAUFMANN AND GUPTA, 1985). Esse métodos são computacionalmente mais complexos que aqueles de aplicações estatísticas mais padrão por causa da necessidade de tratar com parâmetros para indivíduos. Outras possíveis técnicas de estimação como métodos de mínimos quadrados, mínima informação e distância mínima não foram considerados.

Um modelo geral de partição difusa do espaço amostral foi desenvolvido por MANTON ET AL. (1994). Os autores usaram o modelo como uma ferramenta para combinar múltiplas fontes de dados, ilustrado como método potencial para esses problemas. Como o modelo não foi completamente construído para esse propósito, propõe-se nesse trabalho a construção de tal modelo baseado no contexto da Análise do Poder Estatístico de COHEN (1988). O poder de um teste estatístico depende de três parâmetros: o critério de significância, a confiabilidade nos resultados amostrais e no "effect size". Dessa forma o poder do teste pode medir o grau para o qual o fenômeno em estudo existe. Nesse trabalho, esse grau é medido através do cálculo de probabilidades de *relevância*, uma variável com valores lingüísticos.

As probabilidades de relevância de cada estudo são utilizadas como peças de evidência difusa para a realização da meta-análise. Foram consideradas tanto a meta-análise exploratória ou substantiva como a meta-análise analítica com propósito de fazer síntese de evidências e determinar a relevância geral de um dado tratamento.

Dessa forma, na seção 4.2 o modelo geral de partição difusa é apresentado. Na seção 4.3 são desenvolvidos os tópicos necessários para a parametrização do modelo geral e o método de estimação dos parâmetros do modelo empregado. Nas seções 4.4 são feitas as adaptações da parametrização e método de estimação para o particular emprego do modelo

em meta-análise. Em 4.5 é construído o modelo de parametrização difusa para a meta-análise. E na seção 4.6 é apresentado o método utilizado para a agregação das probabilidades da variável difusa relevância.

## 4.2 Caracterização Do Modelo Geral De Partição Difusa

### *Caracterização dos Dados*

Assume-se que os dados provêm de  $J$  variáveis-resposta discretas com somente um de finitos números ( $L_j$ ) de categorias para a  $j$ -ésima variável ocorrento ou que as variáveis contínuas podem ser recodificadas em intervalos discretamente definidos para gerar uma variável categórica sem perda de informação significativa. Para variáveis de natureza intrinsecamente discreta (sexo, corrente estado marital) essa codificação é direta. Nesse caso pode-se ver os dados como consistindo de  $J$  variáveis multinomiais  $x_{ij}$  com  $L_j$  níveis de resposta para a  $j$ -ésima variável, ou equivalentemente,  $L_j$  variáveis binárias  $y_{ijl}$  definida para a  $j$ -ésima variável (i.e. um total de  $\sum_{j=1}^J L_j$  variáveis binárias).

Para usar variáveis contínuas o esforço analítico e computacional será maior porque os pontos de corte que formam categorias intervalares representando todos as informações com significado em distribuição contínua de resultados da variável original têm que ser selecionadas. Felizmente, pode ser mostrado que, para tipos mais padrões de variáveis aleatórias contínuas, o número de categorias  $L_j$  necessárias para representar uma variável  $j$  raramente precisam ser maiores que 15 (TERREL AND SCOTT, 1985). Em muitos casos 5 a 10 categorias intervalares aproximarão satisfatoriamente a informação contida em uma variável contínua. Na prática esse esforço adicional de recodificação tem a vantagem potencial que (i) frequentemente dados de variável contínua são realmente reportados acumulados, então as categorias discretas filtram ou suavizam variações estranhas dos processos de medidas vindos da distribuição e (ii) sem a necessidade de usar transformações de escala não linear, os efeitos das observações exteriores, ou anomalias distribucionais (e.g. distribuições bimodais) são reduzidas.

### *Conjuntos Discretos versus Conjuntos Difusos*

O conceito de conjunto é familiar em matemática. Cursos elementares em matemática definem conjunto como uma coleção de objetos, com cada objeto na coleção definido como um elemento do conjunto. Muito da matemática, e conseqüentemente muito da base teórica da estatística, é determinado por restrições adicionais, suposições e características que são definidas por conjuntos "crisp" no modelos desenvolvidos para analisar dados gerados de planos observacionais específicos.

Durante os últimos 25 anos ou mais, o conceito de conjunto difuso tem sido desenvolvido e implementado para resolver muitos problemas de engenharia. Em contraste aos conjuntos difusos, o tipo clássico de conjunto é chamado de conjunto "crisp". A maior diferença entre conjuntos "crisp" e difusos é na natureza essencial da pertinência de elementos individuais aos conjuntos. Por exemplo, pertinência para conjuntos "crisp" requer que elementos individuais sejam membro do conjunto ou não. Dois conjuntos são disjuntos se eles não têm elementos em comum. Em conjuntos difusos, um elemento pode ser um membro parcial de múltiplos conjuntos. Como conseqüência, a pertinência de qualquer elemento particular relativo a um coleção de conjuntos difusos não é "crisp" mas difusa.

De uma forma mais formal pode-se definir:

**Definição 4.2.1** Para cada elemento em um conjunto difuso existe um escore de Grau de pertinência ( $g_{ij}$ ) que representa o grau que o elemento  $i$  pertence ao  $k$ -ésimo conjunto. Assume-se que esses escores variam entre 0 e 1 inclusive. O 0 indica que o elemento não tem pertinência ao conjunto, e o 1 indica que o elemento tem pertinência completa, ou "crisp", ao conjunto.

Se o escore do Grau de Pertinência para cada elemento no conjunto é 1 ou 0, o conjunto é "crisp". De outra forma o conjunto é difuso. Então, a presença dos escores do Grau de Pertinência que variam continuamente entre 0 e 1 implicitamente definem um conjunto difuso. O grau com o qual elementos são membros dos conjuntos difusos é representado pelos escores dos Graus de Pertinência. A natureza dos métodos estatísticos difusos são determinados pela definição das propriedades desses escores. Escores dos Graus de Pertinência são essenciais para métodos com conjuntos difusos uma vez que eles

representam o estado para o parâmetro e para os elementos e devem ser explicitamente definidos e manipulados. Nos cálculos com conjuntos "crisp", onde os escores do Grau de Pertinência podem ser apenas 0 ou 1, eles são geralmente manipulados implicitamente pelo uso de modelos de probabilidade onde o conjunto de pertinência é desconhecido ou medido com erro, e probabilidades são usadas para se fazer declarações sobre a incerteza da pertinência do conjunto "crisp". Uma vez que a pertinência é freqüentemente implícita nos modelos estatísticos baseados em conjuntos "crisp" existe freqüentemente uma confusão sobre a estrutura matemática desses dois tipos de modelos por causa da coincidência que  $g_{ik}$  varia entre 0 e 1 sobre k. Então, à primeira vista  $g_{ik}$  lembra probabilidades. Como mostrado abaixo essa interpretação não é correta.

Primeiro é preciso ser visto que existe pequena necessidade de conjuntos difusos em áreas onde as regras de comportamento ou organização podem ser claramente e "crispmente" definidas. No entanto, no contexto de engenharia e ciências da computação, onde a representação matemática é a norma, está se tornando crescentemente claro que conjuntos difusos são necessários para descrever o comportamento de sistemas físicos que se comportam de forma complexa.

Além de ser útil como um modelo matemático em aplicações de engenharia, os métodos de conjuntos difusos tem sido úteis em áreas onde os objetos são intrinsecamente membros difusos de conjuntos.

As diferenças não estocásticas de alta dimensionalidade de indivíduos de uma população será chamada de heterogeneidade implicando diferenças em múltiplos momentos estatísticos, e não pode ser confundido com o uso do termo variação que implica somente em momentos de segunda ordem (i.é., variâncias).

Quando a heterogeneidade está presente em uma população em estudo, é necessário aproximar essa heterogeneidade com um modelo de conjunto "crisp" (usando uma variedade de variáveis explanatórias ligadas ao fenômeno por coeficientes de regressão para condicionar a heterogeneidade gerando uma distribuição marginal média sobre a heterogeneidade) ou usando um modelo de conjunto difuso onde a heterogeneidade dos indivíduos é explicitamente parametrizada no modelo. Quando um modelo de conjunto difuso é usado, a "fuzzificidade" não representa simplesmente limitações do sistema de

classificação, mais do que isso é um modelo de indivíduos cuja pertinência é difusa em algum conjunto definido. Com o crescimento do tamanho da amostra, as características comuns dos indivíduos faz crescer a certeza de perfis definidos para cada conjunto difuso, que cresce com o número de variáveis, as quais, por seu turno, ajudam a descrever com mais confiança o indivíduo ou as características difusas do modelo. Então o modelo de conjunto difuso usa os dois tipos de crescimento de informação ( i.e., cresce na dimensionalidade do espaço de observações do caso e da variável.).

#### *Uma Notação para a Classe de Modelos Difusos*

Um exemplo genérico começa com a suposição que uma amostra de  $I$  indivíduos é retirada de uma população. Para cada indivíduo existem  $J$  diferentes questões ou respostas medidas. Para a questão  $j$ , o indivíduo tem uma escolha de  $L_j$  diferentes respostas. Cada indivíduo fornece somente uma resposta para cada questão.

Esse paradigma introduz a seguinte notação. A variável  $I$  representa o número de indivíduos na amostra. A variável  $J$  representa o número de variáveis medidas para cada indivíduo.

#### *Conjuntos "Crisp" nos Métodos Estatísticos*

O uso de conjuntos "crisp" no desenvolvimento de modelos estatísticos é correntemente a prática comum. Esses modelos assumem que uma pessoa é um membro de um e somente um dos conjuntos que medem o espaço de estado para o fenômeno de interesse. Por espaço de estado significa o conjunto de condições definindo o estado de cada indivíduo na população. Então, o espaço de estado é coberto por todas as variáveis relevantes na explicação do fenômeno em estudo, onde explanação significa que eventos passados e futuros são distribuídos independentemente após condicionamento nas variáveis de estado. As restrições implícitas no espaço de estado "crisp" obriga a assumir que o conjunto ao qual uma pessoa pertence é homogêneo. Assume-se que cada conjunto para o qual um elemento deve ser assinalado ocupa um conjunto distinto de coordenadas ou ponto no espaço de estado  $J$  dimensional definido pelos valores nas  $J$  variáveis medidas. Então todos os membros de um dado conjunto ocupam identicamente o mesmo ponto físico (isto é, tem

as mesmas  $J$  coordenadas) no espaço de estados e não existe qualquer heterogeneidade individual dentro de qualquer conjunto "crisp". Qualquer variação entre os indivíduos no mesmo conjunto é atribuída a um erro estocástico gerado pelo instrumento de medida ou experimento amostral.

As implicações dos conceitos de conjuntos "crisp" são algumas vezes ambíguas em modelos estatísticos específicos porque a interpretação geométrica de conjuntos discretos é deixada implícita, isto é, é uma suposição não estabelecida nas declarações matemáticas e estatísticas do modelo. É de particular interesse a distinção entre modelos onde a pertinência em cada conjunto é assumida conhecida a priori, como oposto a modelos onde a pertinência não é assumida como conhecida e é identificada probabilisticamente.

#### *Partição Difusa.*

No modelo de conjuntos difusos, a função de assinalamento em um conjunto atribui um grau de pertinência para cada elemento para cada e todo conjunto definido. Esse escore do grau de pertinência  $g_{ik}$  pode tomar qualquer valor entre 0 e 1 para um ou mais conjuntos ou classe de objetos ou pessoas. Uma vez que um pessoa pode ser um membro parcial de múltiplos conjuntos ou classes, o modelo é então chamado de modelo de conjunto difuso. Permitindo escores de graus de pertinência para cada indivíduo que varia em magnitude para representar o grau de pertinência em  $k$  conjuntos difusos, a variação ou heterogeneidade dos indivíduos é representada dentro de cada conjunto. Nessa situação, a classificação é generalizada pelo assinalamento de casos a conjuntos discretos para identificar características das estruturas da função de densidade multivariada dos escores do grau de pertinência dos indivíduos para a resolução que é estatisticamente possível com os dados disponíveis. A modelagem da heterogeneidade dos indivíduos é então análoga a identificar várias características da função de densidade multivariada que descreve a distribuição dos escores dos graus de pertinência através das populações de indivíduos de interesse. Essa função densidade multivariada é crucial para a análise de dados heterogêneos. Uma metodologia baseada em conjuntos difusos requer a identificação e a estimação dos escores dos graus de pertinência para cada indivíduo relativo a qualquer conjunto. Isto é acompanhado pela construção de uma partição difusa.

**Definição 4.2.2** Uma partição difusa de uma coleção de objetos ou pessoas consiste de um número de conjuntos "crisp" de um único elemento chamados de *perfis extremos* e um conjunto de escores de Graus de Pertinência para cada elemento na coleção relativa a esses *perfis extremos*.

Uma partição difusa é simples de construir uma vez que os escores dos graus de pertinência, isto é, posições no espaço de estados são construídas relativamente a esses perfis extremos. Perfis extremos têm um papel similar à base de vetores em espaços lineares, que permitem a representação de pontos no espaço como combinação linear dos vetores da base. Em particular, todos os elementos de uma partição difusa podem ser expressos como um combinação convexa de padrões extremos. Uma vez que os perfis extremos são conjuntos "crisp" de um único elemento, eles são definidos sem ambigüidade. De uma maneira prática, os perfis extremos são escolhidos de tal forma que uma grande variedade de características da coleção sejam representados pelos perfis. Por exemplo, na representação da saúde de uma população humana de homens idosos deve-se escolher "frágil, 100 anos" e "ativo, 50 anos" para serem dois conjuntos de perfis extremos. Um escore de grau de pertinência individual poderia representar o quão ativo fisicamente e quanto idoso é relativo a esses dois perfis extremos. As definições de perfis extremos podem englobar características em um grande número de dimensões. A performance dos modelos de conjuntos difusos geralmente melhoram quando a dimensionalidade do espaço de medidas cresce.

Um importante problema com vista à partição difusa é determinar quantos perfis extremos são necessário para representar os indivíduos em um dado conjunto finito de dados. Continuando com a analogia com a base de vetores, os perfis extremos varrem o espaço de indivíduos no espaço de medidas. A partição baseada nos perfis extremos deveria incluir o número mínimo de perfis extremos necessários para construir esse varrimento. Diferentemente do que na análise vetorial, no entanto, pode não ser fácil a determinação de quando o espaço está completamente varrido.

Formalizando então algumas das notações apresentadas:

**Notação.** A variável  $g_{ik}$  é usada para representar o escore do grau de pertinência para o  $i$ -ésimo indivíduo para o  $k$ -ésimo perfil extremo em uma partição difusa.



Um valor 0 de  $g_{ik}$  para um indivíduo para algum  $k$  indica que o indivíduo não tem pertinência no conjunto representado pelo perfil extremo  $k$ . Se o valor  $g_{ik}$  for 1, o indivíduo é um membro completo do conjunto difuso representado pelo perfil extremo  $k$ . Um valor entre 0 e 1 indica que o indivíduo é um membro parcial do conjunto considerado. Em um sentido o valor  $g_{ik}$  representa a proporção ou intensidade de pertinência em cada perfil extremo. Isso não é o mesmo que a probabilidade que o indivíduo seja do perfil extremo  $k$  porque a proporção ou intensidade é uma medida matemática de quantidade - por exemplo, quantos dos  $J$  atributos uma pessoa de um dado perfil extremo manifesta - e não uma declaração sobre a frequência de um resultado específico e "crisp" em um experimento repetido. O indivíduo não é de algum simples perfil extremo a não ser que ele tenha um valor unitário para algum dos  $g_{ik}$ . No limite, quando  $J$  tende para infinito, a verossimilhança de  $g_{ik} = 1.0$  deveria ser 1.0. Em todos os outros casos ele tem um mistura de atributos ponderada convexamente associada com os perfis extremos.

**Notação.** A variável inteira  $k$  é usada para representar o número de perfis extremos usados na partição difusa.

A maneira que foi descrita a partição difusa impõe algumas restrições a  $g_{ik}$ . Explicitamente, elas devem ser todas não negativas e somar 1 sobre os valores de  $k$ . Essa restrição será chamada de Condição 1.

**Condição 1** (Restrições básicas para  $g_{ik}$ )

$$g_{ik} \geq 0 \quad \text{para cada } i \text{ e } k \quad (4.2.1)$$

$$\sum_{k=1}^K g_{ik} = 1 \quad \text{para cada } i. \quad (4.2.2)$$

### 4.3 Estimação Dos Parâmetros Do Modelo Geral

#### *Suposições do Modelo*

Embora bem aceitos em engenharia e ciências da computação e empregados em delineamentos de "chips" eletrônicos, os princípios de conjuntos difusos não tem sido ainda extensivamente usados em modelos estatísticos a serem empregados na análise de dados.

Uma falha de tentativas anteriores para implementar modelos estatísticos de conjuntos difusos tem sido a dificuldade em acoplar mecanismos computacionais e estatísticos padrões baseados na teoria de probabilidades e na ótica de conjuntos discretos com a base conceitual diferente da lógica de conjuntos difusos e os cálculos de possibilidades.

Conjuntos difusos podem ser empregados em análises estatísticas se os princípios subjacentes dos modelos estatísticos e algoritmos numéricos forem feitos consistentes como os princípios matemáticos dos modelos de conjuntos difusos. Uma das possíveis dificuldades encontradas no esforço de fazer a tradução é que, mesmo para aqueles que aplicam conjuntos difusos, a heterogeneidade das medidas ainda tende a ser vista como um artefato de quebra ("artifact of flaws") na habilidade para representar (por exemplo, sistemas lingüísticos incompletos) ou medir heterogeneidade mais do que sendo uma propriedade física real (determinística) de sistemas multidimensionais altamente complexos. Essa diferença de ponto de vista leva a representações geométricas diferentes que geram espaços de parâmetros com diferentes propriedades matemáticas e estatísticas.

Uma forma básica desse modelo foi definida por WOODBURY AND CLIVE (1974) e WOODBURY ET AL. (1978). Essa implementação estatística de um modelo de conjunto difuso foi apresentado como modelo GoM ("Grade of Membership Model"), onde como foi definido antes, graus de pertinência são parâmetros que representam o grau para o qual um dado caso ou indivíduo  $i$  é um membro do  $k$ -ésimo perfil extremo. Esses são estimados de observações.

**Notação.** Para cada questão ou medida  $j$  para o indivíduo  $i$  seja  $y_{ijl}$  a resposta representada por um conjunto de  $L_j$  variáveis aleatórias binárias Denota-se  $y_{ijl}$  como as realizações empíricas da variável aleatória  $Y_{ijl}$ .

Para formular o modelo são necessárias as seguintes suposições:

**Suposição 1:** As variáveis aleatórias  $Y_{ijl}$  são independentes para os diferentes valores de  $i$ .

**Suposição 2:** Os valores  $g_{ik}$ ,  $K=1, \dots, k$ , são realizações dos componentes do vetor aleatório  $\xi_i = (\xi_{i1}, \dots, \xi_{ik})$  com função de distribuição  $H(\mathbf{x}) = \Pr(\xi_i \leq \mathbf{x})$ .

**Suposição 3:** Condicional ao valores realizados de  $g_{ik}$  para o indivíduo  $I$ , os  $Y_{ijl}$  são independentes para os diferentes valores de  $j$ .

A suposição 1 estabelece que as respostas para diferentes indivíduos são independentes. Inicialmente foi assumido que os  $Y_{ijl}$  são amostrados aleatoriamente da população. No entanto, diferentes modelos de amostragem podem ser usados para modificar apropriadamente a forma da verossimilhança.

A suposição 2 estabelece que os escores dos graus de pertinência ( $g_{ik}$ ) são variáveis aleatórias que são realizadas quando um indivíduo é selecionado da população. A distribuição amostral das realizações (os  $g_{ik}$  ou escores na amostra) fornecem estimativas da função de distribuição  $H(\mathbf{x})$ . Embora muito da discussão seja baseado nos  $g_{ik}$ , é importante perceber que uma amostra diferente de indivíduos produzirá um diferente conjunto de valores de  $g_{ik}$  para as partições difusas.

A suposição 3 é a suposição de independência condicional comum nos modelos de classe latente e na análise fatorial discreta. Essa suposição, para modelos de conjuntos difusos, estabelece que se os escores dos graus de pertinência  $g_{ik}$  são conhecidos, as respostas do indivíduo  $i$  para as várias questões  $Y_{ijl}$  são independentes dos valores de  $j$ .

O papel das partições difusas do modelo é definido pelas suposições 4 e 5.

**Suposição 4:** A probabilidade de uma resposta  $l$  para a  $j$ -ésima questão pelo indivíduo com o  $k$ -ésimo perfil extremo é  $\lambda_{kjl}$ .

Embora muitos indivíduos possam ser um membro parcial do  $k$ -ésimo tipo, pela suposição existe pelo menos um indivíduo, teoricamente possível, que é um membro "crisp" do  $k$ -ésimo tipo. A suposição 4 fornece as probabilidades de resposta para esse indivíduo para os vários níveis de cada questão. Claramente, as probabilidades  $\lambda_{kjl}$ , satisfazem as restrições requeridas das probabilidades discretas. Isso é estabelecido formalmente pela Condição 2.

**Condição 2** (Restrições de  $\lambda$ )

$$\lambda_{kjl} \geq 0 \quad \text{para cada } k, j \text{ e } l \quad (4.3.1)$$

$$\sum_{l=1}^{L_j} \lambda_{kjl} = 1 \quad \text{para cada } k \text{ e } j. \quad (4.3.2)$$

**Suposição 5.** A probabilidade de uma resposta do nível 1 para a  $j$ -ésima questão para indivíduo  $i$ , condicional nos escores  $g_{ik}$ , é dado pela forma bilinear

$$\Pr(Y_{ijl} = 1.0) = \sum_{k=1}^K g_{ik} \lambda_{kjl} \quad (4.3.3)$$

A suposição 5 constrói a ponte para a definição da partição difusa com a expressão paramétrica para a probabilidade de uma resposta específica discreta. Implícito a essa suposição é que o valor de  $K$  é fixo. Inicialmente é assumido que  $K$  é conhecido. É importante notar que a partição difusa irá necessariamente requerer um conjunto de valores de  $g_{ik}$  para cada elemento da coleção. No entanto, não é necessário que tal partição requeira a probabilidade de resposta a ser dada pela expressão acima.

As suposições 1 a 5 fornecem a base para formar um modelo probabilístico para construir os procedimentos de estimação de verossimilhança. Explicitamente, o modelo de probabilidade para uma amostra aleatória é o produto do modelo multinomial com células de probabilidade dadas por

$$E(Y_{ijl}) = \sum_{k=1}^K g_{ik} \lambda_{kjl} \quad (4.3.4)$$

Ou, equivalentemente,

$$P_{ijl} = \sum_{k=1}^K g_{ik} \lambda_{kjl} \quad (4.3.5)$$

Onde assume-se que os  $g_{ik}$  são conhecidos e satisfazem a Condição 1. Diferentes modelos de probabilidade podem ser formados por outros delineamentos de amostragem usando as suposições de 1 a 5.

#### *Distribuição de Probabilidades do Modelo Básico*

Da suposição 1 e do fato que cada uma das  $J$  questões contém um número finito ( $L_j$ ) de níveis, as contagens observadas para cada questão seguirá o modelo de probabilidade multinomial. Da Suposição 3 sabe-se que o modelo geral para os dados, através das questões, será um modelo multinomial de produto, dados os valores de  $g_{ik}$ . As suposições 4 e 5 são usadas para determinar as probabilidades das células para cada questão. Então, dessas suposições, a verossimilhança desse modelo pode ser escrita como

$$L(y) = \prod_{i=1}^I \prod_{j=1}^J \prod_{l=1}^{L_j} \left( \sum_{k=1}^K g_{ik} \lambda_{kjl} \right)^{y_{ijl}} \quad (4.3.6)$$

Essa notação no lugar da mais comum com fatoriais foi usada por razões didáticas. Para todas as combinações de  $i$  e  $j$ , os  $y_{ijl}$  serão zero para todos os valores de  $l$  exceto um, para o qual será unitário.

### *Estimação dos Parâmetros*

A estimação dos parâmetros do modelo geral é baseada nos princípios da máxima verossimilhança, apresentados no Capítulo 3 de MANTON ET AL. (1988). A identificabilidade do modelo tem em vista a especificação matemática única da verossimilhança. A identificabilidade estatística das estimativas dos parâmetros diz respeito à habilidade que o método de estimação tem para gerar parametrização única a partir dos dados disponíveis. Ou seja, a verossimilhança e a função densidade de probabilidades devem ser identificáveis no sentido de que elas são unicamente especificadas pela distribuição das variáveis aleatórias observáveis.

Ocorre que na prática, nem sempre se tem dados suficientes para estimar todos os parâmetros estruturais  $\lambda_{kjl}$ . No caso da partição difusa a diferença entre os dados que se tem e que se precisa podem ser devida ao fato de que o número de "perfis extremos"  $K$  no modelo seja muito grande, o número de questões  $J$  relacionada a cada indivíduo na amostra seja muito pequena, ou o número de indivíduos seja muito pequeno. (MANTON ET AL., 1988). Nesse caso o método de estimação por máxima verossimilhança tem restrições em seu emprego, e algum outro método indireto de estimação dos parâmetros deve ser empregado.

O número máximo de equações para  $J$  variáveis cada uma com  $L_j$  respostas é  $\prod_{j=1}^J L_j - 1$ . Para uma solução com  $K$  classes, o número de parâmetros  $\lambda_{kjl}$  a serem estimados é  $k \sum_{j=1}^J (L_j - 1)$ . O número de momentos  $H(x)$  de grau  $J$  ou menos é

$$[(J+K-1)!/J!(K-1)!]-1 \quad (4.3.7)$$

Então para se ter um número adequado de equações para estimar os parâmetros, uma condição necessária para identificabilidade é:

$$\prod_{j=1}^J L_j > k \sum_{j=1}^J (L_j - 1) + \frac{(J + K - 1)!}{J!(K - 1)!} - 1 \quad (4.3.8)$$

Isso compara o número de equações formadas pela variação dos  $y_{ijl}$  com o número de  $\lambda_{kjl}$  e momentos para o modelo com K classes.

Uma condição suficiente para que os parâmetros sejam estimáveis é que  $K < J/2$ , onde K é número de "perfis extremos" e J o número de variáveis.

#### 4.4 Definição das Variáveis Difusas para a Meta-Análise

*A lógica do procedimento de testagem estatística de hipóteses.*

A inferência estatística tem um caráter necessariamente probabilístico. Ele é dado pelo requerimento que um pesquisador estabelece certos padrões de probabilidade para o resultado de sua pesquisa que fornece uma base para a rejeição da hipótese nula e assim para a prova da existência do fenômeno que está sendo testado.

Os resultados de uma amostra aleatória retirada de uma população somente aproximará as características da população. Então, ainda que a hipótese nula seja de fato verdadeira, não se espera que um dado resultado amostral reflita exatamente esse fato.

Existem dois métodos fundamentais de acessar hipótese estatísticas. Uma devida a Fisher, através do uso do chamado "p-value", que consiste fundamentalmente em determinar o valor de significância pós-experimentação para indicar a força da evidência contra a hipótese nula. Outra devida a Neyman e Pearson, contemporânea à primeira, onde o experimentador deve tomar a decisão de rejeição da hipótese nula com base em um nível de significância  $\alpha$ , pré-estabelecido. Nos dois enfoques é questionada a utilização simples do nível de significância para a tomada de decisões estatísticas. Outros elementos como o poder do teste, a dimensão dos parâmetros considerados, o tamanho da amostra e a confiabilidade/variabilidade do experimento devem também ser considerados.

*O enfoque de Neyman-Pearson*

Quando se estabelece um valor baixo para o nível de significância, 5 ou 1%, deseja-se ser capaz de dizer que se a hipótese nula é falsa, a probabilidade do resultado amostral

obtido não é maior que o valor do nível de significância em questão. Se se puder fazer essa declaração, uma vez que o nível de significância é pequeno, pode-se dizer que a hipótese nula foi rejeitada a esse nível. Se, por outro lado, a probabilidade encontrada for maior que o nível estabelecido, não se pode fazer a declaração acima e o teste falhou em rejeitar a hipótese nula.

Essa forma de tratar os testes estatísticos isola um elemento desse enfoque de inferência estatística, a prova padrão que o fenômeno existe, ou equivalentemente, uma "disprova" padrão da hipótese nula que estabelece que o fenômeno de fato existe. O padrão da prova é um resultado amostral que ocorrerá em menos de 5%, por exemplo, das vezes se a hipótese nula for verdadeira. O nível de significância é a probabilidade de se rejeitar a hipótese nula incorretamente. Toda essa discussão é condicionada na "verdade da hipótese nula". A esperança usual do pesquisador, que estabelece a hipótese nula com propósitos táticos, é rejeitar a hipótese nula para concluir que o fenômeno existe. Mas o fato de o fenômeno existir na população não garante um resultado estatisticamente significativo. E é nesse ponto que o conceito de poder estatístico deve ser considerado.

O poder de um teste estatístico de uma hipótese nula é a probabilidade que ele conduzirá à rejeição da hipótese nula. Isto é, a probabilidade que ele resultará na conclusão que o fenômeno existe. Dadas as características do teste realizado e o comportamento da população, o poder do teste pode ser determinado. Ele representa uma peça vital de informação sobre um teste estatístico aplicado aos dados de uma pesquisa. Tanto durante a fase de planejamento da pesquisa em que se pode mudar o seu delineamento para a obtenção de um maior poder de teste, como na fase pós pesquisa, quando se encontra um pequeno poder e então os resultados negativos são vistos de forma ambígua, uma vez que falhar em rejeitar a hipótese nula não pode ter significado mais substantivo, ainda que o fenômeno exista (a um dado grau).

O poder de um teste estatístico depende de três parâmetros: o critério de significância, a confiança nos resultados amostrais e o "effect size" (COHEN, 1988). Em última análise, o que importa é a determinação da existência de um fenômeno tendo por base os dados experimentais. O nível de significância é então apenas um dos parâmetros

para a determinação da relevância dos resultados amostrais obtidos, que é de fato, a informação que se utiliza na tomada de decisões.

O nível de significância,  $\alpha$ , é a probabilidade de se rejeitar a hipótese nula quando ela é verdadeira. Essa rejeição indevida é chamada de erro tipo I. Uma vez que ele representa a razão de se rejeitar uma hipótese nula verdadeira, ele é tomado como um valor pequeno. Segue então, que quanto menor for esse valor, mais rigoroso é o padrão de rejeição da hipótese nula, ou equivalentemente da prova de que o fenômeno em estudo existe. Supondo que o fenômeno de fato exista na população, quanto menor for o valor de  $\alpha$ , mais pobres são as chances que a amostra fornecerá resultados que encontrem esse padrão, ou seja, menor o poder. (COHEN, 1988).

A prática de se tomar um valor muito pequeno para  $\alpha$  resulta em valores de poder também relativamente pequenos. No entanto, o complemento do poder ( $1 - \text{poder}$ ), simbolizado por  $\beta$ , é também a probabilidade de se cometer um erro, chamado de erro tipo II, uma vez que ele representa a razão de se falhar em rejeitar uma hipótese nula falsa. Dessa forma a inferência estatística pode ser vista como uma ponderação entre esses dois tipos de erros. Um pesquisador pode estabelecer que o risco da rejeição de uma hipótese nula ser falsa seja bem pequeno, por exemplo  $\alpha = 0.001$ , mas fazendo isso, ele pode reduzir o poder do teste para 0.10 (com uma probabilidade  $\beta = 1 - 0.10 = 0.90$ ). Essa concepção implica que a seriedade relativa do erro tipo I para o erro tipo II será de  $\beta/\alpha = .90/0.0001 = 900$ . Ou seja, o pesquisador implicitamente acredita que erradamente rejeitar a hipótese nula é 900 vezes mais sério que erradamente aceitá-la. Em outra situação, com  $\alpha = 0.05$ , poder = 0.80 e  $\beta = 1 - 0.80 = 0.20$ , a seriedade relativa do erro tipo I sobre o tipo II é  $\beta/\alpha = 0.20/0.05 = 4$  para 1, então errar na rejeição da hipótese nula é considerada quatro vezes mais sério do que a aceitação.

A direcionalidade do critério de significância também afeta o poder do teste estatístico. Quando a hipótese nula pode ser rejeitada em ambas direções então a região crítica é nas duas caldas da distribuição amostral do teste estatístico, o teste resultante terá menor poder que o teste com mesmo nível só que direcional. Uma vez que o teste unilateral não pode, por definição, conduzir à rejeição da hipótese nula no sentido oposto do que o



predito, esses testes não tem poder para detectar tais efeitos, e aquela conclusão é inadmissível. Nesse tipo de análise,  $\alpha$  deve ser determinado antes da consecução do teste.

A confiabilidade (ou precisão) de um valor amostral está relacionado com o que pode ser esperado como aproximação do valor populacional relevante. Ele é necessariamente na prática um valor estimado, uma vez que o valor na população não é conhecido. Depende da estatística em questão e do específico modelo estatístico no qual o teste foi baseado. E sempre depende do tamanho da amostra.

Um meio convencional para acessar a confiabilidade de uma estatística é o erro padrão da estatística. E ele depende fortemente do tamanho da amostra. Quanto maior for o tamanho da amostra, mantendo-se os outros valores iguais, menor será o erro e maior a confiabilidade ou precisão dos resultados, e maior será a probabilidade de se detectar diferenças significativas (COHEN, 1988). Dessa forma pode-se formular diretamente a relação entre o tamanho da amostra e o poder do teste. Crescimento da amostra implica em crescimento do poder, o probabilidade de se detectar o fenômeno sob teste, embora fontes de variabilidade no transcurso do próprio experimento possam afetar o poder do teste, uma vez que o aumento de variabilidade não controlada reduz a precisão dos resultados amostrais.

O grau para o qual o fenômeno está presente na população é estimado pelo tamanho do efeito de um certo tratamento sobre as unidades amostrais observadas. Essa diferença é convenientemente chamada de "effect size", o grau para o qual a hipótese nula é falsa. Não importando a maneira da representação do fenômeno em uma particular pesquisa, a hipótese nula sempre significa que o tamanho do efeito é zero. Quando a hipótese nula é falsa em algum grau, o "effect size" é algum valor específico diferente de zero na população. Quanto maior for esse valor, maior será o grau para o qual o fenômeno em estudo é manifestado.

De qualquer forma como ele é medido, o "effect size" pode ser tratado como parâmetro que toma valor zero quando a hipótese nula é verdadeira, e serve como um índice do grau de afastamento da hipótese nula. O "effect size" é na prática o mais importante determinante do poder ou da tamanho da amostra requerido. Existe uma relação entre o nível de significância, o "effect size", o tamanho da amostra e o poder. Se mantidos

os mesmos nível de significância e o tamanho da amostra, quanto maior o "effect size" maior será o poder. Se mantidos os mesmos nível de significância e o poder, quanto maior for o "effect size" menor será o tamanho da amostra necessário para detectá-lo.

O "effect size" a ser utilizado depende do tipo de modelo estatístico em estudo. Uma discussão geral é apresentada no Capítulo 3 e maiores detalhamentos sobre modelos estatísticos e "effect size" podem ser encontrados em COHEN (1988), HEDGES AND OLKIN (1983) e HUNTER AND SCHMIDT(1990).

### *O enfoque de Fisher*

O enfoque *fisheriano* é um pouco diferente do apresentado acima. Considerando uma situação de testagem de hipóteses onde a apropriada região crítica para o teste caia em uma particular cauda da distribuição amostral, o valor observado na amostra pode ser usado para determinar a probabilidade sob a hipótese nula de um resultado amostral ser igual ou mais extremo do que aquele observado. Essa probabilidade é chamada de "p-value".

Do ponto de vista da tomada de decisões esse procedimento contém mais informação se ou não as observações são estatisticamente significativas. Reportar um "p-value" se exato ou dentro de um intervalo, de fato permite que cada indivíduo escolha seu próprio nível de significância como a probabilidade máxima tolerável para o erro tipo I. Isso é particularmente importante quando o pesquisador não tem uma razão real justificável ou não tem um sentimento sobre os custos e conseqüências do erro estatístico para a particular situação experimental.

Em muitas situações a decisão a ser tomada é mais prática do que estatística. Então o resultado estatístico deve ser considerado não mais do que um auxílio para a formação de uma decisão subjetiva (GIBBONS AND PRATT ,1975) Significância estatística não implica relevância prática. O processo de tomada de decisão é freqüentemente influenciada por muitos fatores em adição ao nível de significância. Alguns desse fatores como confiança no procedimento amostral ou validade do procedimento de teste são estatísticos, enquanto outros como implicações práticas e econômicas da decisão são fatores puramente não estatísticos relevantes para uma particular decisão.

Quando a decisão é relacionada a uma população alvo diferente da população amostrada, por motivos práticos ou pela impossibilidade de se amostrar da população alvo,

o processo de tomada de decisão se torna ainda mais subjetivo. Algum ou todos os outros fatores podem ser mais importantes que o "p-value".

Se o tamanho da amostra é moderado ou grande, então o teste é relativamente mais poderoso. Então se o valor do "p-value" é moderado ou grande, o teste não só falhou em "disprovar" a declaração contida na hipótese nula, ele também fornecerá evidência substancial que a hipótese nula é verdadeira ou quase verdadeira. E quanto menor for o "p-value" mais extemado é o resultado do fenômeno em estudo.

Comparações entre "p-values" entre tamanhos amostrais e entre experimentos, diferentemente do que no trato com o poder do teste têm significado limitado. Argumentos dentro e fora da escola freqüentista de probabilidade fortalecem a idéia de que a extensão da contradição da hipótese nula é, em geral, menos uma função do "p-value" e mais da função de verossimilhança. (BIRNBAUM, 1962).

Se um teste é muito poderoso, então é muito provável a rejeição de  $H_0$  e então o "p-value" tende a ser menor ainda quando a distância da situação nula é pequena. Similarmente ocorre o inverso. No enfoque Bayesiano, onde as probabilidades não são usadas somente com significado "objetivo" mas também como *graus de crença*, a probabilidade que a hipótese nula seja verdadeira, dadas as observações (isto é a probabilidade "a posteriori" da hipótese nula, deve variar fortemente, dependendo do tamanho da amostra e do problema, para um valor fixo do "p-value" e uma probabilidade fixa da hipótese nula antes da observação (isto é, uma probabilidade "a priori" fixa da hipótese nula). Nesse enfoque, se o valor de um teste estatístico é significativo a 5%, existe ainda uma chance que a hipótese nula seja verdadeira.

*"p-value", "effect size" e tamanho da amostra como variáveis difusas*

SCHERVISH (1996) apresenta interpretações de "p-value" além dos testes unilaterais e bilaterais, e sustenta em última análise que a interpretação de um particular valor de significância obtida, tal como o 0.05 varia fortemente com as hipóteses estabelecidas. HUNG ET AL. (1997) discutem como o comportamento da distribuição do "p-value" depende do tamanho da amostra para o experimento e da magnitude do parâmetro testado na hipótese alternativa. Em ambos os casos, os autores corroboram com a

discussão previamente feita sobre a dificuldade de se tomar os "p-value" segundo uma interpretação puramente objetiva.

Na consecução do procedimento meta-analítico as considerações acima são levadas em conta para a formulação de um modelo GoM para a síntese de evidências. O "p-value", o "effect size" e o tamanho da amostra são consideradas como variáveis independentes difusas para a determinação da probabilidade de relevância de cada estudo. A relevância é uma variável lingüística que receberá também um tratamento como variável difusa. O que se deseja em última análise para a tomada de decisão clínica é se um estudo ou um conjunto de estudos semelhantes fornecem evidências suficientemente fortes para a adoção de um dado tratamento.

Em Inteligência Artificial, relevância pode ser entendida de várias formas, como por exemplo, que atributos são relevantes para diferentes tarefas, que tarefas são relevantes para a implementação de um sistema ou ainda como a produção de algoritmos de aprendizagem produzem classificadores se treinados somente por subconjuntos de relevantes exemplos de treinamento. SUBRAMANIAN ET AL. (1997) no editorial "The relevance of relevance" do número especial da revista Artificial Intelligence apresenta uma série de artigos que tratam de diferentes aspectos da relevância. Nesse número especial Baum e Smith estabelecem um critério de relevância que os força a se moverem do modelo tradicional de fazer estimativas pontuais para o uso de distribuições paramétricas que podem aprender durante o curso de um jogo. Eles mostram que a distinção entre linhas importantes e não importantes de jogar pode ser capturada usando medidas de relevância em distribuições sobre escores de avaliação. É aproximadamente essa última concepção do raciocínio com base na relevância que serão construídas probabilidades de relevância de um tratamento em um particular estudo que participará de uma meta-análise. Ou seja, baseando-se nas três características básicas do procedimento de testagem de hipóteses estatísticas - o "p-value", o "effect size" e o tamanho da amostra calcula-se a probabilidade que um fenômeno seja relevante, sem tomar o nível de significância como último decisor da existência do fenômeno em estudo.

#### 4.5 Método de Agregação da Variável Relevância

A variável probabilidade de relevância é uma medida de probabilidade difusa, e portanto e pode ser compreendida como números difusos. Uma das formas de agregação de números difusos pode ser feita através do cálculo da média ou média ponderada de números difusos. Para tanto seria necessário definir os suportes de cada um desses números e então operar a soma de números difusos e depois dividir esse resultado pelo total de parcelas dessa soma. No caso de média ponderada, seria necessário multiplicar-se cada número obtido pelo seu peso.

O resultado final seria um número difuso - a probabilidade de relevância conjunta dos estudos - e então "difusificar" o resultado para obtenção de um resultado "crisp" de relevância conjunta dos estudos, para comparação com um valor padrão estabelecido, que possibilitaria a tomada de decisão final, sobre se um efeito é relevante ou não.

O estabelecimento do valor de corte dependerá do grau de rigor que o especialista que usará os resultados, deseje impor para a tomada de decisões. Sugere-se como ponto de corte padrão em torno do qual o especialista deve considerar a aceitação de relevância, valores em torno de 0.8, que é o valor de poder de teste sugerido nos procedimentos clássicos de tomada de decisão em estatística.

Os resultados obtidos dessas operações alcançam valores muito próximos dos resultados obtidos pela consideração que as probabilidades de relevância obtidas nos estudos particulares são números "crisp", embora tenha-se que levar em conta que esses números são ainda valores aproximados. É possível então operar os valores das probabilidades de relevância de forma a obter uma média desses valores como forma de obtenção da probabilidade de relevância conjunta dos estudos que participam da meta-análise.

Conforme visto na seção 3.2, existem duas formas de se proceder a meta-análise, a primeira chamada de "meta-análise de efeito fixo", que considera que o conjunto de estudos é homogêneo entre si, e que a ponderação dos valores obtidos deve ser feita considerando o inverso das variâncias de cada estudo. Assim define-se  $w_i = 1/V_i$ , a informação contida no  $i$ -ésimo estudo, como os ponderadores para o cálculo da média final (GAVER, 1992). A suposição de homogeneidade pode ser verificada através do cálculo da estatística de

Cochran ( $Q_w$ ), conforme 3.2.4 apresentada na página 56. Caso os estudos não sejam homogêneos o modelo deve ser considerado de "efeito aleatório", e nesse caso a média deve ser obtida pela ponderação através de  $w_i^* = 1/(V_i + \hat{\tau}^2)$ . Aqui  $w_i^*$  estima o peso ótimo.

A necessidade de estimar  $\hat{\tau}^2$ , além de  $V_i$ , torna o estabelecimento das incertezas nos modelos de efeito aleatórios mais complicados do que nos modelos de efeito fixo. Muitos enfoques foram propostos. GAVER (1992) cita os procedimentos de componentes de variância propostos por Gilbert et al. em 1977, Laird e DerSimonian em 1982, Hedges e Olkin em 1985 e Goodman em 1989. DERSIMONIAN AND LAIRD (1986) comparam quatro enfoques para estimar a variância entre estudos: método de máxima verossimilhança, método de máxima verossimilhança restrita (REML) e dois estimadores não iterativos baseados no método dos momentos. Um dos estimadores não iterativos que eles recomendam é:

$$\hat{\tau}^2 = \max \left\{ 0, \frac{Q_w - (k-1)}{\sum w_i - \frac{\sum w_i^2}{\sum w_i}} \right\} \quad (4.5.1)$$

Em síntese, caso os estudos possam ser considerados homogêneos empregar-se-á o modelo de efeito fixo utilizando como ponderadores das probabilidades de relevância os inversos das variâncias ( $w_i$ ), caso contrário, utilizar-se-á como ponderadores os valores de  $w_i^*$  conforme definido acima. Assim o valor encontrado para a probabilidade de relevância conjunta dos estudos que participaram da meta-análise, deverá ser comparado com um valor padrão (em torno de 0.80, dependendo do rigor do especialista), de tal forma que se a probabilidade de relevância conjunta encontrada for maior que esse valor padrão, considerar-se-á que as diferenças obtidas entre os tratamentos é relevante, e que portanto pode-se empregar aquele procedimento clínico que está sob consideração.

## 5 Aplicação

### 5.1 INTRODUÇÃO

Na década de 70 seis grandes experimentos clínicos aleatorizados multicêntricos sobre a eficiência do uso da aspirina no tratamento de pacientes após um ataque cardíaco (infarto do miocárdio, IM), foram conduzidos na Europa e nos Estados Unidos.

As seis pesquisas em consideração incluem duas realizadas na Inglaterra dirigidas por ELWOOD ET. AL. (1974) denominada de UK-1 e por ELWOOD AND SWETNAN (1979), UK-2. Uma realizada na Alemanha e Áustria, o "German-Autrian Multicenter Study", GAMS, BREDDIN ET. AL. (1979). Outra pesquisa é o "Coronary Drug Project Aspirin Study", CDPA (1976). A Quinta é a "Persantine-Aspirin Reinfarction Study Research Group", PARIS (1980). E a última "Aspirin Myocardial Infarction Study Research Group", AMIS (1980).

Todas as pesquisas foram conduzidas durante o período de 1970 a 1979. Foram aleatorizadas, com controle de placebo, duplo-cegas e multiclinicas. Uma síntese dos delineamentos experimentais utilizados pode ser encontrado em CANNER (1983).

Muitos estudos têm demonstrado que as plaquetas do sangue têm um papel fundamental em iniciar a formação de trombose nas artérias. Outros estudos têm demonstrado a habilidade do ácido acetil-salicílico (aspirina) em produzir uma inibição prolongada da agregação de plaquetas. Essa inibição pode prolongar a vida do paciente pelo efeito profilático da aspirina. Segundo ELWOOD ET. AL. (1974) a evidência clínica do benefício pode vir somente através de pesquisas clínicas aleatorizadas em que um grupo recebe o tratamento com aspirina e outro de controle com placebo.

Embora uma série grande de estudos sobre o efeito da aspirina tenham sido realizados de 1980 para cá, e que comporiam um procedimento meta-analítico mais atualizado e completo, optou-se por empregar a metodologia difusa utilizando somente os seis estudos citados, como forma de comparar os resultados obtidos através dessa

metodologia com outras meta-análises já realizadas que utilizaram diferentes métodos com o mesmo grupo de pesquisa. Também o pequeno número de pesquisas incorporadas, podem revelar como as diferentes metodologias adquirem informação de poucas peças, e de que formam buscam tirar evidências relevantes delas.

Uma rápida apresentação de seis trabalhos em meta-análise realizadas com o conjunto de dados será feita nessa introdução. A discussão dos resultados obtidos nelas é realizada no último item desse capítulo.

CANNER (1983) utilizou como principal estatística o "odds ratio", onde  $p_i$  denota a proporção observada de eventos (óbitos) nos grupos de tratamento ( $i=1$ , para aspirina e  $i=2$ , para placebo).

CANNER (1987) não encontrando razões definitivas que explicassem fontes de heterogeneidade em seu estudo anterior, repetiu o estudo realizado para comparar oito diferentes possíveis métodos meta-analíticos para a análise: (i) o método de Fisher da soma dos logaritmos dos "p-values", (ii) o método da soma dos "p-values", (iii) o método da testagem da média dos "p-values", (iv) o método da soma de estatísticas "z" ponderadas, (v) o mesmo método ponderado pelo número de pacientes, (vi) o mesmo método ponderado pela raiz quadrada do número de pacientes, (vii) o método de Cochran-Peto, com a utilização do teste de log-rank e (viii) o teste utilizando o logaritmo do "odds ratio" ponderado.

GOODMAN (1989) utilizou esse conjunto de pesquisas como forma de verificar que o procedimento meta-analítico clássico (de efeito fixo com interpretação frequentista) não fornece uma medida de evidência por não estabelecer claramente uma definição de evidência, adotando o nível de significância como tal medida. O autor propõe o uso da função de verossimilhança como medida de evidência, através do desenvolvimento do modelo de "efeito aleatório", procurando formas de utilização dos resultados da análise estatística para o procedimento clínico em face de um novo paciente.

GAVER ET. AL. (1992) são os editores do "Panel on Statistical Issues and Opportunities for Research in the Combination of Information" patrocinado pelo "US - National Research Council" no início da década de 90. Nesse painel todo procedimento meta-analítico foi revisado e as várias técnicas discutidas. Exemplos em diferentes áreas



foram apresentados e entre eles a meta-análise dos seis experimentos sobre a eficiência da aspirina, que é apresentado sob a ótica de cada uma das técnicas de meta-análise.

MORRIS AND NORMAND (1992) apresentam os enfoques "hierárquico bayesiano" e "empírico bayesiano" para a consecução da meta-análise. Embora o exemplo utilizado para o desenvolvimento desses métodos não seja o dos seis experimentos aleatorizados em questão, nas conclusões os autores fazem uma breve discussão dos achados de CANNER (1987). Outra justificativa da inclusão desse estudo para comparação é o fato dele ser discutido por seis painelistas com pesquisas em meta-análise, cinco dos quais foram referenciados em outras partes desse trabalho. (Carlin, DuMouchel, Fienberg, Lindley e West).

Em trabalho mais recente, NORMAND (1995) retoma essa discussão apresentando um estudo comparativo dos resultados obtidos por diferentes metodologias através de três programas para a realização da meta-análise do problema proposto.

O enfoque difuso para a análise das seis pesquisas pretende cobrir em parte algumas das críticas levantadas pelos autores acima referenciados, buscando comparar alguns resultados obtidos. A aplicação é feita através da (2) apresentação dos dados, com alguma discussão dos resultados obtidos, (3) caracterização das variáveis independentes - "p-value", "effect-size" e "tamanho da amostra" como variáveis difusas, (4) determinação da partição difusa do espaço amostral, (5) determinação das probabilidades de "relevância" com a utilização do modelo de "partição difusa do espaço amostral", (5) síntese dos resultados individuais obtidos e finalmente (6) uma discussão comparativa dos resultados obtidos pela metodologia difusa com os resultados antecessores.

## 5.2 APRESENTAÇÃO DOS DADOS

O quadro 5.1.1 mostra o número de pacientes e o número de mortes para os dois tratamentos em cada estudo e os resultados dos testes de significância para os estudos individuais e para o resultado conjunto das seis pesquisas.

Estudo	Aspirina		Placebo		Comparação		$Z_i^{(1)}$	$p_i^{(2)}$
	nº de pacientes	Índice de mortalidade (%)	nº de pacientes	Índice de mortalidade (%)	Diferença (%)	Erro padrão Da diferença (%)		
UK-1	615	7.97	624	10.74	2.77	1.65	1.68	0.047
CDPA	758	5.80	771	8.30	2.50	1.31	1.91	0.028
GAMS	317	8.52	309	10.36	1.84	2.34	0.79	0.216
UK-2	832	12.26	850	14.82	2.56	1.67	1.54	0.062
PARIS	810	10.49	406	12.81	2.31	1.98	1.17	0.129
AMIS	2267	10.85	2257	9.70	-1.15	0.90	-1.27	0.898
Total	5599	9.88	5217	10.73	0.86	0.59	1.47	0.072

(1)  $Z_i$  representa a razão da diferença nos índices de mortalidade sobre seus erros-padrões, onde é assumida uma distribuição binomial.

(2)  $p_i$  é o nível de significância unilateral associado com  $Z_i$ , onde a aproximação normal é usada.

Quadro 5.2.1 Número de pacientes e índice de mortalidade por todas as causas, para seis experimentos clínicos comparando o uso de Aspirina e Placebo para pacientes após um ataque cardíaco

A inspeção dos dados mostra índices de mortalidade semelhantes para os dois grupos e para as diferenças entre os índices dos dois grupos nos cinco primeiros estudos, havendo uma diferença em sentido contrário no último estudo (AMIS), onde o índice de mortalidade para aspirina é mais ou menos próximo de outros resultados obtidos, em particular com os resultados obtidos em PARIS. No entanto, embora o índice de mortalidade para placebo em AMIS não apresente também uma grande diferença em relação aos outros obtidos, a menos de UK-2, ele é menor do que o seu índice obtido para aspirina. O que leva à conclusão que o placebo resultou em menor índice de mortalidade do que a aspirina, embora essa diferença não seja estatisticamente significativa. Em termos de diferenças significativas, somente o CDPA apresenta um nível mais fortemente significativo ( $p = 0.028$ ), sendo que os estudos ingleses (UK-1 e UK-2), revelam alguma significância (0.047 e 0.062, respectivamente). O resultado total, se considerado o método de "efeito-fixo" não ponderado pelas variâncias apresenta uma significância de 0.072, revelando que a aspirina propicia uma certa diminuição no índice de mortalidade pela

administração de aspirina, embora o resultado não possa ser considerado definitivo. Se forem considerados somente os cinco primeiros estudos, o nível de significância obtido de  $p = 0.0014$  revelaria uma forte tendência pró-aspirina, conforme quadro 5.2.2:

Estudo	Aspirina		Placebo		Comparação		$Z_i$	$p_i$
	nº de pacientes	Índice de mortalidade (%)	nº de pacientes	Índice de mortalidade (%)	Diferença (%)	Erro padrão Da diferença (%)		
Sub-Total	3332	9.21	2960	11.52	2.31	0.77	2.98	0.0014

Quadro 5.2.2 Número de pacientes e índice de mortalidade por todas as causas, para o total de cinco experimentos clínicos comparando o uso de Aspirina e Placebo para pacientes após um ataque cardíaco

A questão colocada é então: se um médico deseja basear a sua decisão sobre o uso de aspirina para um paciente em tratamento pós-infarto do miocárdio, como ele poderia usar a informação desses seis estudos para orientá-lo na escolha do tratamento.

### 5.3 CARACTERIZAÇÃO DAS VARIÁVEIS INDEPENDENTES

O que acreditar com base nas evidências dos testes estatísticos depende em parte de evidências que não estão neles e que tem como ponto de apoio os eventos observados. Existe uma distinção clara entre a filosofia de evidência e o pensamento freqüentista. O "p-value" é uma medida de raridade com respeito a uma hipótese. Em termos de evidências o uso simples do "p-value" é irrelevante para propósitos indutivos. O mesmo "p-value" tem diferentes significados inferenciais para amostras de tamanhos diferentes. Sob a hipótese nula, um pequeno efeito em uma grande pesquisa pode ser tão raro como um efeito grande em uma pesquisa pequena., produzindo o mesmo escore "z" e o mesmo "p-value". O valor absoluto do "effect size", uma parte essencial da evidência, não é refletida nem no valor de "z" nem no "p-value". Por outro lado, dois experimentos com o mesmo valor absoluto de efeito observado, tem bases similares, a evidência fornecida para qualquer parâmetro será aproximadamente proporcional aos seus tamanhos de amostra.

O modelo difuso para meta-análise considera três variáveis com valores lingüísticos para determinar valores de probabilidades de "relevância" de cada estudo, e que devem ser compostas para uma análise de evidência dos estudos em conjunto.

### *"p-value"*

Existem quatro valores de significância que são usados como padrões aproximados para a tomada de decisões. O valor 0.01 (ou 1%) é normalmente utilizado como ponto de referência para determinar se a rejeição da hipótese pode ser considerada altamente significativa. Valores menores que 0.01 indicam que o resultado é "altamente significativo". Quanto mais esse valor se aproxima de 0 maior é a "evidência" da diferença entre os valores dos parâmetros entre dois tratamentos. A variável difusa que tem o valor lingüístico "altamente significativo" deve refletir o fato que valores entre 0.01 e 0 devem ter aumentado o seu grau de pertinência muito próximos de 1 para valores muito próximos de 0, mantendo ainda valor elevado de grau de pertinência para o valor 0.01. Na medida em que esses valores se afastam desse ponto eles vão diminuindo o seu grau de pertinência até chegar a um grau de pertinência 0 quando o valor de "p" for igual a 0.05.

O valor 0.05 (5%) é um ponto de corte importante na análise tradicional. Comparações que obtêm valores menores que 0.05 são considerados significativos enquanto valores maiores são considerados não significativos. Então 0.049 e 0.051 levam a conclusões diferentes e conseqüente mudança na decisão tomada. Tomando por base esse valor, a variável significativa é construída como uma variável difusa triangular que levará a graus de pertinência à esquerda menores até atingir o nível 0.01 quando o valor do grau de pertinência para "significativo" é 0. À direita esse valor irá diminuindo de 1, para 0.05 até o valor 0 quando o nível de significância for 0.10.

Níveis de significância superiores a 10% até 50%, importam em não significância. Valores muito próximos de 0.10 já redundam em um grau elevado de não-significância e que portanto terão valores de graus de pertinência para essa variável bastante próximas de .1, atingindo esse valor para o nível de significância de 50%. Valores maiores do que 50% implicam em incorreção no sentido na hipótese formulada, e portanto o grau de "evidência" para a hipótese alternativa será nula, e portanto tem grau de pertinência 1 no conjunto "não-significativo".

As variáveis linguísticas acima descritas podem ser desenvolvidas a partir de conjuntos difusos trapezoidais e triangulares apresentados na figura 5.3.1, com as modificações necessárias, realizadas pela aplicação de "hedges", pela imposição das características acima.

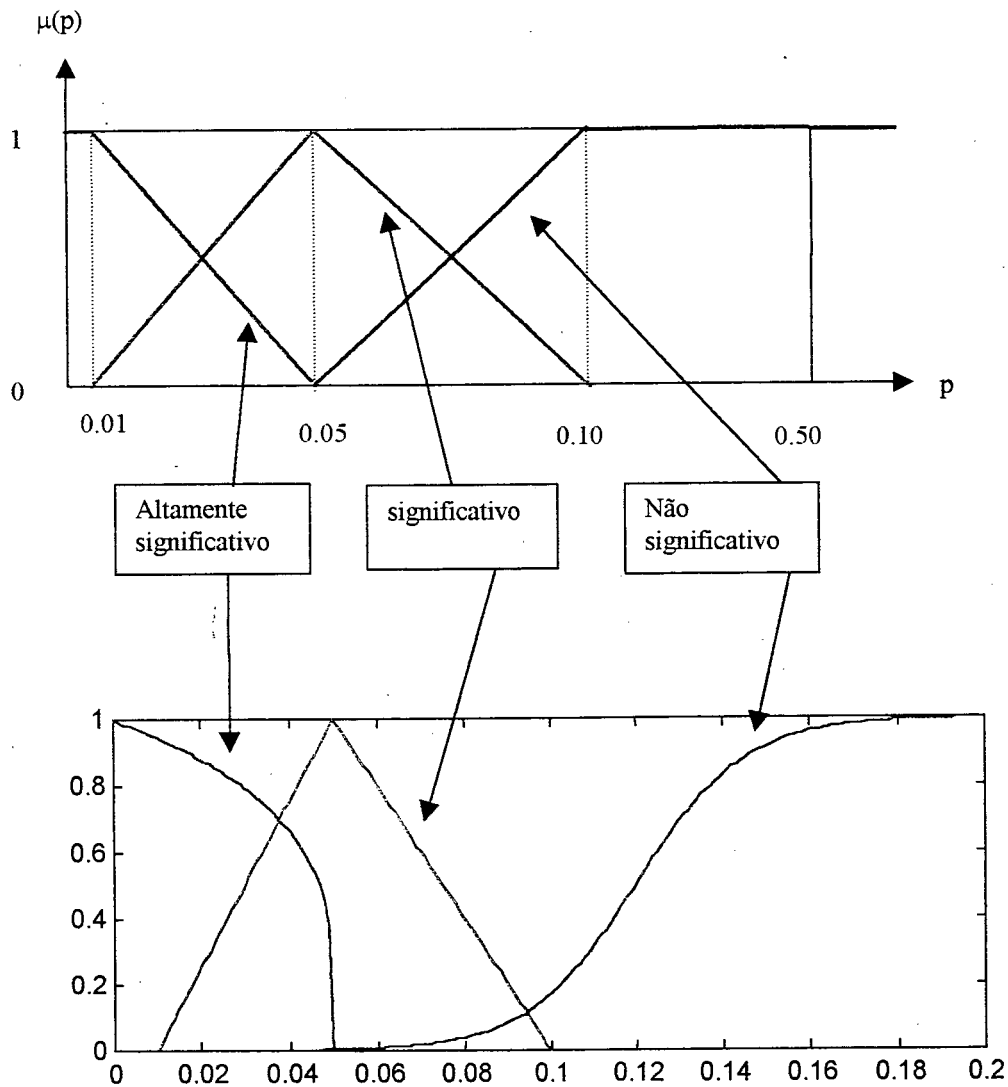


Figura 5.3.1 Conjuntos difusos para as variáveis "p-value".

As funções de pertinência dos três conjuntos definidos acima podem ser expressas por:

A - Altamente significativo

$$A(p) = \left( \frac{0.05 - p}{0.05} \right)^{0.25} \quad \forall p, 0 \leq p \leq 0.05 \quad (5.3.1)$$

B - Significativo

$$B(p) = \frac{p - 0.01}{0.04} \quad \forall p, 0.01 \leq p \leq 0.05$$

$$B(p) = \frac{0.10 - p}{0.05} \quad \forall p, 0.05 \leq p \leq 0.10 \quad (5.3.2)$$

C - Não significativo

$$C(p) = \frac{1}{1 + \exp(-80(p - 0.12))} \quad \forall p, p \geq 0.10 \quad (5.3.3)$$

*"effect size"*

O "effect size" é o tamanho do efeito de um certo tratamento sobre um outro. Ele é um parâmetro natural usado para sintetizar os resultados experimentais. Quando se trabalha com médias a interpretação mais direta do tamanho do efeito na população é que  $\delta$  (o "effect size") - que é determinado por  $\delta = (\mu^e - \mu^c) / \sigma$ , onde "e" é experimental e "c" é controle - é a diferença média que se obteria se a variável dependente fosse escalonada para ter variância unitária dentro dos grupos do experimento. Então o "effect size" é somente a diferença de médias re-expressa em unidades de uma escala com  $\sigma = 1$  para remover a dependência de uma escala arbitrária ao fator  $\sigma$ .

Quando se trabalha com proporções, a forma de se retirar o efeito de escala é a aplicação da função arco seno da raiz da proporção. Então, por exemplo, a diferença entre

as proporções 0.4 e 0.1, 0.3 é melhor representada pela diferença dos arcos-seno das raízes das proporções que é 0.26, quando se deseja comparar com as diferenças entre as proporções 0.5 e 0.8, que é também 0.3, mas cuja diferença entre os arcos-seno das raízes é 0.32. (COHEN, 1988). Quanto maior for o valor do "effect size" maior será o grau para o qual o fenômeno em estudo é manifestado. Cohen propõe para diferenças entre os arcos-seno de proporções três valores que determinam se o "effect size" é pequeno, médio ou grande. Esses valores são respectivamente: 0.30, 0.50 e 0.80. Esses valores servem de base para a construção dos três conjuntos difusos considerados nesse trabalho, conforme a figura 5.3.2:

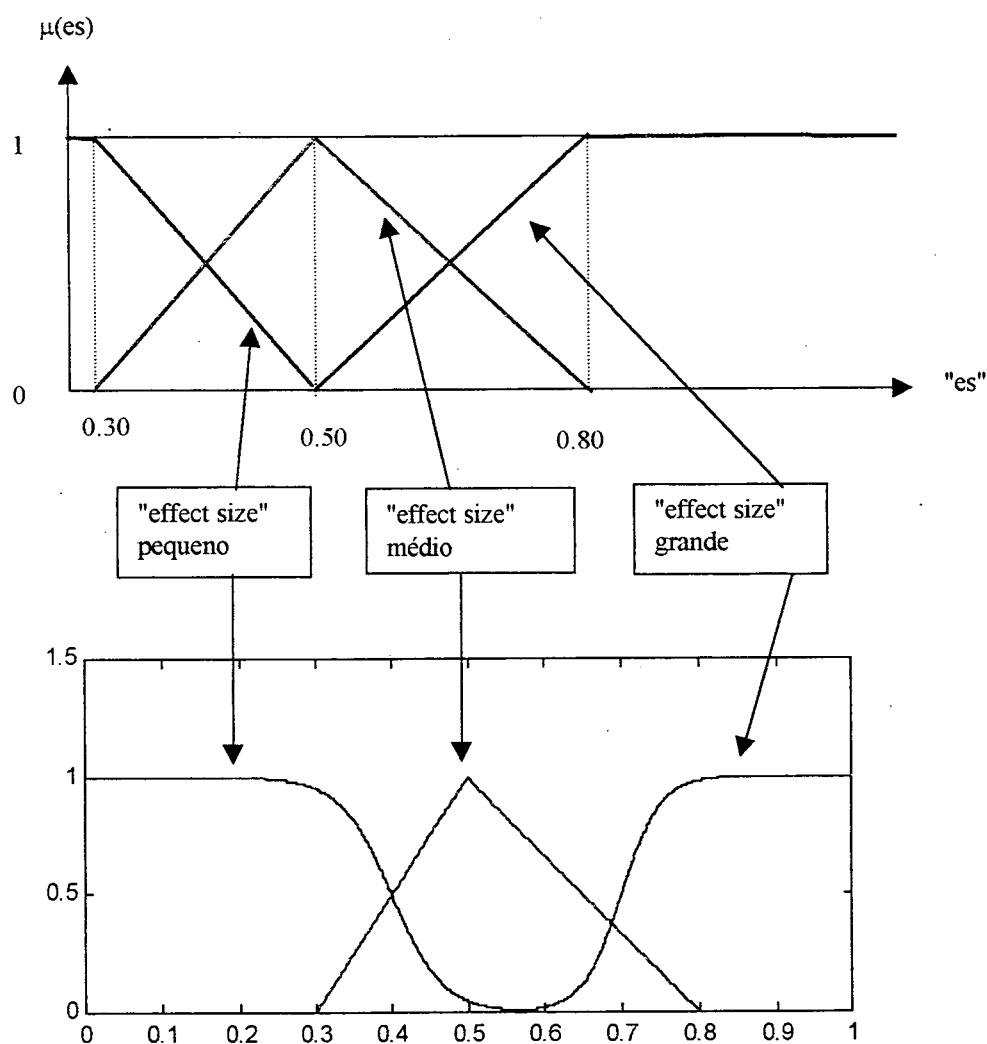


Figura 5.3.2. Conjuntos difusos para as variáveis "effect size"

As funções de pertinência dos três conjuntos definidos acima podem ser expressas por:

A - "Effect size" pequeno

$$A(es) = \left( \frac{1}{1 + \exp(30(es - 0.4))} \right) \forall es, 0 \leq es \leq 0.50 \quad (5.3.4)$$

B - "Effect size" médio

$$B(es) = \frac{es - 0.30}{0.20} \forall es, 0.30 \leq es \leq 0.50$$

$$B(es) = \frac{0.80 - es}{0.30} \forall es, 0.50 \leq es \leq 0.80 \quad (5.3.5)$$

C - "Effect size" grande

$$C(es) = \frac{1}{1 + \exp(-80(p - 0.12))} \forall es, es \geq 0.50 \quad (5.3.6)$$

### *Tamanho da amostra*

Do ponto de vista estatístico, dois valores são mais ou menos formalmente estabelecidos para a consideração do tamanho da amostra com relação ao seu tamanho. Esses valores foram estabelecidos pela proximidade de uma distribuição "t" de Student para uma distribuição normal. Com uma amostra de tamanho 30 o valor da estatística "t" se aproxima muito do valor de "z". E com uma amostra de tamanho 100 esses valores são praticamente iguais. Então pelo Teorema Central do Limite, 30 já pode ser considerado como um valor de grande amostra. No entanto com uma amostra de tamanho 100 essa consideração é bem mais tranqüila. Esses valores, na realidade, são usados como parâmetros práticos quando da realização dos procedimentos estatísticos de inferência.



Pode-se, dessa forma, considerar-se que uma amostra menor que 30 é uma amostra pequena. E que uma amostra maior que 100 é uma amostra grande. Assim, os seguintes conjuntos difusos podem ser construídos com base nessas informações. Figura 5.3.3.

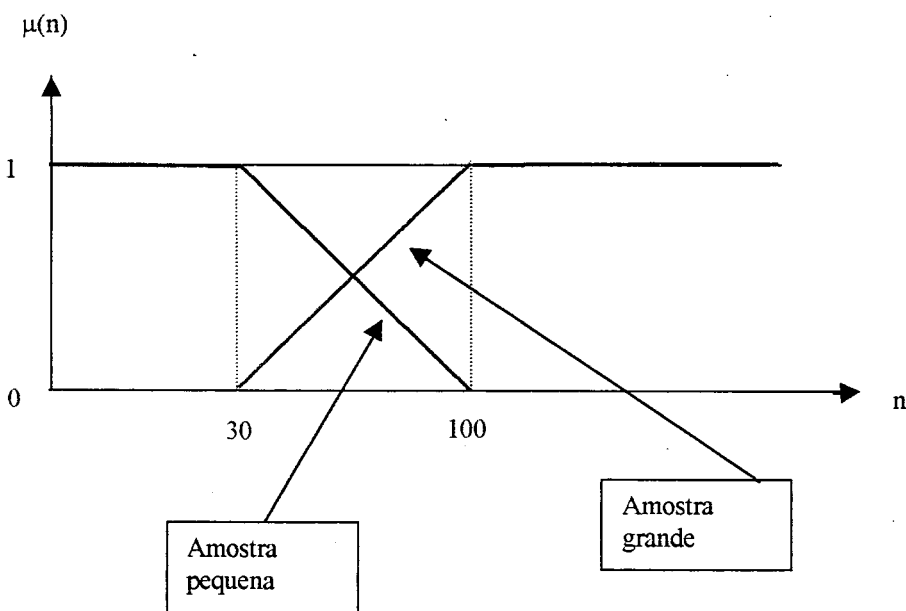


Figura 5.3.2 Conjuntos difusos para tamanho da amostra do ponto de vista estatístico.

Para efeito de um primeiro estudo os conjuntos difusos considerados foram baseados nessas definições "estatísticas" de tamanho pequeno e grande de uma amostra. Essa definição é também útil quando se faz análises de experimentos aleatórios em que se restrinjam a amostras nessa amplitude. No entanto, uma consideração epidemiológica é mais apropriada para a definição do que seria uma grande amostra, quando os estudos considerados tiverem grandes amostras de fato. A consideração epidemiológica é importante quando se considera a administração de um certo tratamento que poderá resultar em diminuir mortalidade, como é o caso do presente estudo. A mudança de 1% para uma amostra de 50 pessoas pode estar-se referindo a nenhuma pessoa, enquanto que 1% de 5000, está-se falando em 50 pessoas. Na construção do modelo, quando as pesquisas que servem de unidade amostral tiverem então tamanhos de amostra muito grandes, como é o caso do presente estudo onde o menor tamanho de amostra é 626 e o maior é de 4524 pacientes, a construção dos conjuntos para tamanho de amostra devem ser modificados.

Os conjuntos triangulares abaixo, servem de ponto de partida para a construção dos conjuntos difusos pertinentes. Uma alteração através da aplicação de um "hedges" foi realizado para aproximar as funções para a forma de raciocínio humano usual, conforme figura 5.3.4:

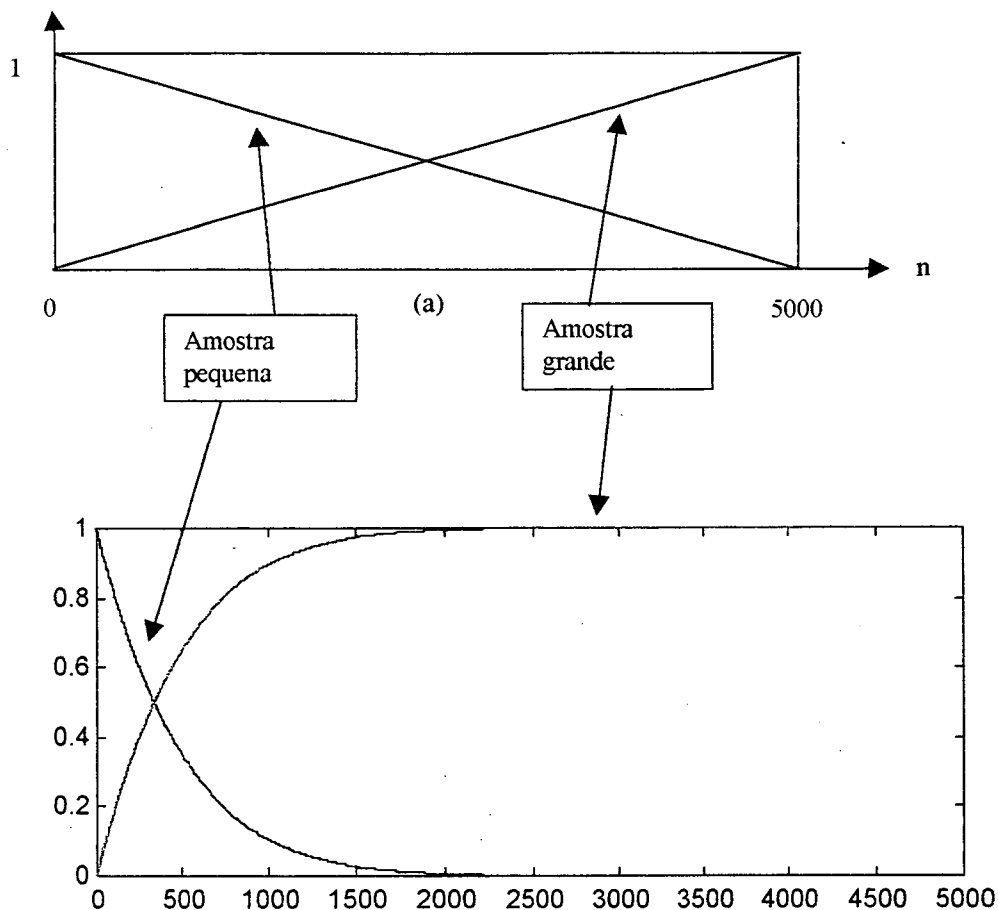


Figura 5.3.4. Conjuntos difusos para tamanho da amostra do ponto de vista epidemiológico (a) conjunto base, (b) conjunto modificado pela aplicação de "hedges".

As funções de pertinência dos três conjuntos definidos acima podem ser expressas por:

A - Amostra pequena

$$A(n) = \left( \frac{5000 - n}{5000} \right)^{10} \quad \forall n, 0 \leq n \leq 5000 \quad (5.3.7)$$

B - Amostra grande

$$A(n) = 1 - \left( \frac{5000 - n}{5000} \right)^{10} \quad \forall n, 0 \leq n \leq 5000 \quad (5.3.8)$$

#### 5.4 Determinação da Partição Difusa do Espaço Amostral

No modelo de conjuntos difusos, a função de assinalamento do conjunto atribui um grau de pertinência para cada elemento para cada e todo conjunto definido. Esses escores de grau de pertinência  $g_{ik}$  pode tomar qualquer valor entre 0 e 1 para um ou mais conjuntos ou classes. Então um elemento pode ser um membro parcial de múltiplos conjuntos ou classes. Uma partição difusa de uma coleção de objetos consiste de um número de conjuntos "crisp" com um elemento chamado de "perfil extremo" e um conjunto de graus de pertinência para cada elemento da coleção relativa a esses perfis extremos. Um problema importante para construir a partição difusa é determinar quantos perfis extremos são necessários para representar os indivíduos em um conjunto de dados finito. A partição baseada nos perfis extremos deve incluir um número mínimo de perfis extremos necessários, uma vez que ele tem um papel semelhante à base em espaços vetoriais lineares.

No presente caso, trata-se de construir o espaço amostral das possíveis realizações de três variáveis com valores lingüísticos. O tamanho da amostra, o "p-value" e o "effect size". Cada uma delas com duas, três e três repostas lingüísticas, respectivamente. Então o número máximo de perfis extremos da partição do espaço amostral é de dezoito, conforme esquema na figura 5.4.1.

A questão colocada então é a seguinte: é possível diminuir o número de perfis extremos e que ainda preserve as categorias em que se deseja alocar os estudos? Uma possível solução para o problema seria colocar as variáveis "effect size" e "p-value" em termos de "grande" e "pequeno" e o número de perfis extremos se reduziria a nove. Perde-se um pouco do detalhamento mais as condições seriam mais favoráveis para o emprego do método de máxima verossimilhança. Mas ainda assim em termos de identificabilidade do modelo, o número de equações seria bem maior do que o número de variáveis resposta, e maior do que o número de indivíduos (pesquisas) que se pretende modelar. Dessa forma, não é conveniente diminuir o número de perfis extremos.

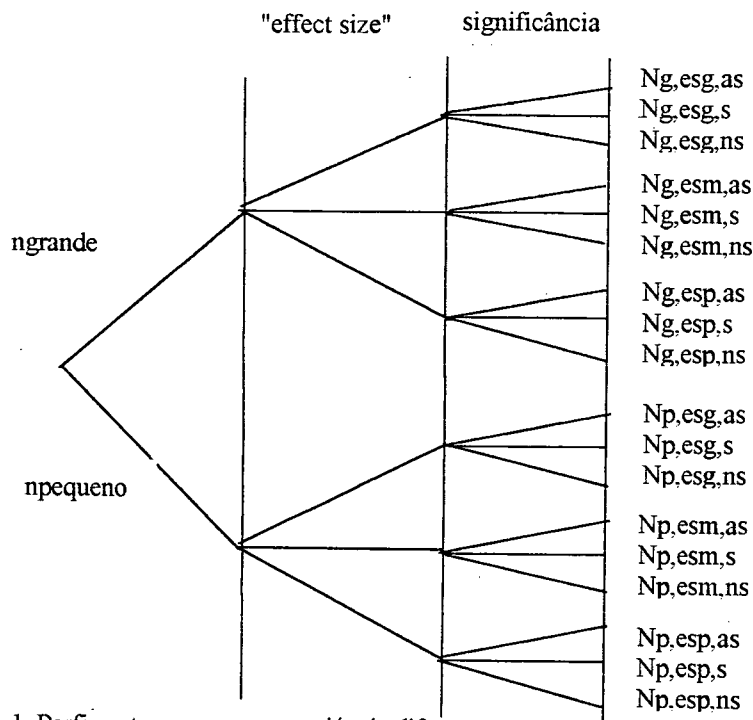


Figura 5.4.1. Perfis extremos para as variáveis difusa

Sendo assim, a partição do espaço amostral permanece com dezoito perfis extremos. Duas medidas devem então ser realizadas. O grau de pertinência de um dado indivíduo a cada um dos perfis extremos e a probabilidade de um indivíduo em um dado perfil extremo ser relevante.

O quadro abaixo representa a partição do espaço amostral conforme delineado acima, com perfis extremos de 1 a 18:

Perfis	Amostra Grande			Amostra Pequena		
	Alt. sign.	Sign.	Não sign.	Alt. sign.	Sign.	Não sign.
ES grande	1	2	3	10	11	12
ES médio	4	5	6	13	14	15
ES pequeno	7	8	9	16	17	18

Quadro 5.4.1 Representação dos perfis extremos

Então uma pesquisa que tenha um tamanho de amostra muito grande, por exemplo 5000, um "effect size" grande, de 1.0, e um "p-value" de 0.0001, terá pertinência 1.0 para o primeiro perfil extremo e 0 para os demais. Se uma pesquisa tem  $n = 600$ , "es" = 0.40 e

"p-value = 0.04", então os seus graus de pertinência para cada uma das variáveis seria, para amostra grande (perfis extremos de 1 a 9), graus de pertinência de 0.62, para os outros graus de pertinência de 0.52. Para "effect size", graus de pertinência de 0.67 para médio (perfis extremos (4,5,6,13,14 e 15), graus de pertinência 0.33 para pequeno (perfis extremos 7,8,9,16,17 e 18) e grau de pertinência zero para os demais. Para nível de significância, graus de pertinência 0.33 para "altamente significativo" (perfis extremos, 1,4,7,10,13 e 16), 0,67 para "significativo"( perfis extremos 2,5,8,11,14 e 17) e zero para os demais. Compondo os graus de pertinência pela operação de mínimo, os graus de pertinência para os perfis extremos, 4,5,7,8,14,16 e 17, são respectivamente, 0.33,0.66,0.33,0.33,0.52,0.33 e 0.33 respectivamente. Para todos os demais o grau de pertinência é zero. Como o modelo tem como restrição que a soma dos graus de pertinência seja unitário é necessário então fazer-se uma padronização, para que a soma resulte em 1. Obtendo-se então os valores de graus de pertinência de 0.11,0.20,0.11,0.20,0.17,0.11 e 0.17.

### 5.5 Determinação da Probabilidade de Relevância

A probabilidade de relevância de um dado estudo será a média das probabilidades de um elemento em um dado perfil extremo ponderada pelos graus de pertinência determinados conforme item 5.4.

É necessário então determinar as probabilidades dos elementos nos perfis extremos resultarem em relevância. O método utilizado para elucidar e conferir valores a essas probabilidades é o de elucidação de probabilidades subjetivas pelo julgamento de especialistas.

Os resultados encontrados são os que seguem:

Perfis Extremos	Amostra Grande			Amostra Pequena		
	Alt. sign.	Sign.	Não sign.	Alt. sign.	Sign.	Não sign.
ES grande	1 0.99	2 0.95	3 0.50	10 0.98	11 0.90	12 0.25
ES médio	4 0.79	5 0.76	6 0.40	13 0.62	14 0.58	15 0.16
ES pequeno	7 0.50	8 0.48	9 0.25	16 0.25	17 0.23	18 0.06

Quadro 5.5.1 Probabilidades Subjetivas de um indivíduo do perfil extremo K ser relevante

Os valores do quadro acima devem ser ponderados pelos graus de pertinência da pesquisa em estudo a cada um dos perfis extremos. No exemplo em discussão a probabilidade de significância de um estudo com  $n = 600$ ,  $es = 0.4$  e  $p\text{-value} = 0.04$  é dado por:

$$P(\text{relevância}) = \sum g_{ik} \lambda_{kjl}, K=1, k$$

$$P(\text{relevância}) = (g_{i4} \lambda_{4rel.})(g_{i5} \lambda_{5rel.})(g_{i7} \lambda_{7rel.})(g_{i8} \lambda_{8rel.})(g_{i14} \lambda_{14rel.})(g_{i16} \lambda_{16rel.})(g_{i17} \lambda_{17rel.}) =$$

$$= (0.11 \times 0.79)(0.20 \times 0.76)(0.11 \times 0.50)(0.20 \times 0.48)(0.17 \times 0.58)(0.11 \times 0.25)(0.17 \times 0.23) = 0.56$$

$$P(\text{relevância} / n = 600, es = 0.4, p\text{-value} = 0.04) = 0.56$$

Conclui-se dessa simulação, que apesar de o estudo apresentar um "p-value" igual a 4%, o que tradicionalmente redundaria em significância, ou seja que o tratamento em teste é significativamente superior ao tratamento controle, o fato de o "effect size" ser relativamente pequeno faz com que a probabilidade de relevância seja de aproximadamente 56%. O que para um nível padrão de probabilidade de relevância de 0.8 pode ser considerada pequena.

O quadro abaixo mostra algumas probabilidades de relevância para resultados de estudos padrões:

Estudos Padrões	Amostra = 1000			Amostra = 100			Amostra = 30		
	p=0.01.	p=0.05	p=0.10	p=0.01.	p=0.05	p=0.10	P=0.01	p=0.05	p=0.10
ES = 0.8	0.99	0.94	0.40	0.98	0.91	0.38	0.98	0.90	0.31
ES = 0.5	0.74	0.70	0.27	0.63	0.59	0.25	0.61	0.57	0.20
ES = 0.3	0.47	0.45	0.17	0.30	0.27	0.15	0.26	0.24	0.10

Quadro 5.5.2 Probabilidades Subjetivas de um indivíduo do perfil extremo K ser relevante

Os resultados obtidos dos seis estudos são apresentados no quadro 5.5.3.

Estudo	nº de pacientes	"effect size"	"p-value"	Probabilidade de relevância
UK-1	1239	.05	.0047	.49
CDPA	1529	.05	.028	.484
GAMS	626	.03	.216	.20
UK-2	1682	.04	.062	.47
PARIS	1216	.04	.129	.23
AMIS	4524	0	.898	0

Quadro 5.5.3 Probabilidades de relevância para os seis estudos

## 5.6 Síntese de Evidências

O método de agregação utilizado depende da consideração de homogeneidade entre os estudos. Caso os estudos possam ser considerados homogêneos emprega-se a síntese pelo modelo de efeitos fixos, caso contrário o modelo de efeitos aleatórios.

Empregando-se a estatística de Cochran dada em 3.2.4 na página 43, para os seis estudos obtém-se o valor de  $Q_w = 9.57$  (com 5 graus de liberdade e  $p = 0.089$ ), que leva a dúvidas substanciais sobre se o modelo de efeitos fixos é apropriado para combinar esses estudos.

Sob a suposição de modelo de efeitos aleatórios e usando a estimativa de DerSimonian e Laird em 4.5.1 na página 96, para  $\tau^2$  obtém-se do quadro abaixo:

Estudo	$d_i$	$S_i$	$w_i$	$w_i^*$
UK-1	2.77	1.65	0.367	0.191
CDPA	2.50	1.31	0.583	0.237
GAMS	1.84	2.34	0.183	0.125
UK-2	2.56	1.67	0.359	0.189
PARIS	2.31	1.98	0.255	0.156
AMIS	-1.15	0.90	1.235	0.301

Quadro 5.6.1 Cálculos para os modelos de efeitos fixos ( $w_i$ ) e de efeitos aleatórios ( $w_i^*$ )

$Q_w = 9.605$ , onde  $Q_w$  tem distribuição qui-quadrado com 5 graus de liberdade, o que leva à conclusão que os estudos não podem ser considerados homogêneos. Dessa forma os  $w_i^*$  devem ser utilizados como ponderadores e o modelo é de efeito aleatório. O valor de  $\tau^2$  obtido foi de 2.509.

A probabilidade de relevância conjunta  $P(\text{rel}_\Sigma)$  é determinada então por:

$$P(\text{rel}_\Sigma) = \frac{\sum_{i=1}^6 w_i^* P(\text{rel}_i)}{\sum_{i=1}^6 w_i^*} \quad (5.6.1)$$

Dessa forma a probabilidade de relevância conjunta para as seis pesquisas é de 35.7%, o que implica que apesar do valor de significância estar próximo dos 5% e do tamanho da amostra ser bastante grande, o "effect size" é muito pequeno, o que implica que a relevância da aspirina no trato de pacientes pós infarto, não altera de forma importante a sobrevida dos pacientes.

É curioso notar que se forem tomados somente os 5 primeiros estudos que resultam em um "p-value" de 0.0014, portanto altamente significativo, o valor da probabilidade de relevância é de 0.723. Ou seja somente 72.3% o que também não recomenda a aspirina como um tratamento extremado para o caso dos pacientes em questão, se for considerado o ponto de corte de 0.80 para relevância. Nesse segundo caso, como existe alta homogeneidade entre os estudos, verificada pelo teste de Cochran, utiliza-se os ponderadores  $w_i$  no lugar dos  $w_i^*$ , por ter sido utilizado o modelo de efeitos fixos.

## 5.7 Discussão dos Resultados

A questão central a ser respondida pelos seis estudos é: como podem os dados nos seis estudos serem combinados de forma a fornecerem linhas para que um médico possa fazer uma escolha adequada do tratamento a ser administrado a pacientes que sofreram Infarto do Miocárdio?

Supondo que seja razoável julgar os pacientes dentro de um dado estudo como suficientemente similares, independentemente de eles terem recebido aspirina ou placebo,



de forma que se possa sumarizar com sensibilidade suficiente cada estudo pelo número  $d_i$ , que representa o efeito da mortalidade no estudo  $i$ , ( $i = 1, \dots, k=6$ ). Esse valor pode representar por exemplo a diferença entre as proporções ( $p_{i1} - p_{i2}$ ) entre as taxas de mortalidade entre aspirina e placebo. É natural nesse ponto considerar a combinação entre os seis estudos. Como isso pode ser feito depende de uma primeira consideração: São os estudos suficientemente homogêneos em delineamentos e protocolos de amostragem dos pacientes que indique que eles estão medindo a mesma coisa? Ou não?

Os modelos de efeito fixo e efeito aleatório foram empregados nesse estudo por GAVER ET AL.(1992), sumarizando o que era conhecido sobre a heterogeneidade dos seis estudos e descreveram os resultados dados pelo método de Fisher de combinação dos "p-values". Sob a hipótese nula de que não há diferenças na mortalidade entre aspirina e placebo, se for assumida homogeneidade entre os estudos, a estatística de Fisher  $-2\sum \log(p_i)$  tem distribuição qui-quadrado com  $2k$  graus de liberdade, onde  $p_i$  é o "p-value" do  $i$ -ésimo estudo. Quando aplicado aos cinco primeiros estudos esse método produz um valor de qui-quadrado de 26.1 com 10 graus de liberdade, levando a um valor combinado dos "p-values" de 0.004, e a hipótese nula poderia então ser fortemente rejeitada em favor do ponto de vista que a aspirina tem um efeito benéfico. No entanto, quando se adiciona o sexto estudo, não importando que ele seja contraditório com os demais, ele vai fazer crescer o valor do qui-quadrado aumentando somente em dois os graus de liberdade da distribuição qui-quadrado. Então, uma vez que o valor da estatística para os cinco estudos é muito grande, o efeito da aspirina não pode ser melhorado pelo método de Fisher. Com o acréscimo do último estudo o valor obtido para a estatística qui-quadrado é de 27.3 com 12 graus de liberdade, que fornece um valor de  $p = 0.008$ . Em outras palavras, se esse último estudo não tivesse sido realizado produzir-se-ia ainda um resultado positivo para a aspirina, não importando as dimensões e os resultados desse último estudo.

Alguns autores, como CANNER (1983 e 1987) e GOODMAN (1989) consideraram a questão da combinabilidade dos dados dos seis estudos. Seus artigos contém informações detalhadas sobre os protocolos e características de delineamento dos estudos. CANNER (1987) examinou as diferenças de delineamentos (programa de doses e dosagens) e as características dos pacientes ( gênero, idade média e tempo médio do infarto até a entrada no estudo), na tentativa de explicar as discrepâncias observadas no maior estudo (AMIS), e

não encontrou nada claro que pudesse explicá-las. Mostrou também que embora houvessem diferenças entre os índices de mortalidade entre esse estudo e os outros, havia uma forte concordância com respeito a uma grande variedade de resultados não-fatais. Canner concluiu que os seis estudos eram homogêneos. Se for assumida homogeneidade e for adotado o modelo de efeito fixo, a estimativa combinada de  $\delta$ , 0.86%, tem um erro padrão de 0.59%, correspondendo a um julgamento de efeito benéfico modesto para a aspirina com forte incerteza sobre seu valor numérico (GAVER, 1992). Essa estimativa ignora qualquer ajustamento para as características de dosagem e dos pacientes.

GAVER (1992) considera o modelo de efeito aleatório para a meta-análise que inclui os seis primeiros estudos e o modelo de efeito fixo para a meta-análise que envolve somente os cinco primeiros estudos, como o realizado nesse estudo. No entanto, as conclusões que chega são um pouco diferentes. O autor conclui que existe para os dois casos alguma significância embora não forte a favor da aspirina. Os valores de  $z$  para cada um dos estudos foi de 1.69 e 1.73 respectivamente, que fornece valores de significância de 0.091 e 0.084. A retirada do último estudo - AMIS, embora seja o maior dos estudos e tenha resultado opinião favorável ao placebo, não produziu grande modificação no resultado obtido. Pelo fato dos valores estarem muito próximos de 0.05, sugere-se alguma relevância para a aspirina em ambos os casos. Os resultados obtidos através do modelo de partição difusa da meta-análise são mais consistentes na medida em que fornecem valores bem diferentes para os dois casos (com a inclusão de AMIS e sem ele) respectivamente de 36% para o primeiro caso e 72% para o segundo. No entanto, nenhum deles é relevante, e isso ocorre porque os "effect size" de todos os estudos foram muito pequenos não importando em relevância clínica. Outro fato que chama atenção nessa comparação são as distâncias entre os achados para os modelos de meta-análise clássicos não ponderado, ponderado e para o modelo de partição difusa, que ficam mais claros na análise que segue:

Meta-análise	Modelo Clássico Não-ponderado	Modelo Clássico Ponderado	Modelo de Partição Difusa
6 estudos	$p = 0.072$	$p = 0.09$	$p^* = 0.36$
5 estudos	$p = 0.0014$	$p = 0.08$	$p^* = 0.72$

Quadro 5.7.1 - Comparação entre os modelos clássicos não-ponderado, ponderado e de partição difusa.

Os primeiros quatro valores (p) devem ser comparados com algum padrão de nível de significância. No geral toma-se o valor de 0.05 como um padrão aproximado para a análise. Sendo assim, para o modelo não ponderado, existe uma forte evidência a favor da aspirina, em particular quando se toma somente os cinco primeiros estudos. Mesmo o valor de 0.072 se aproxima de 0.05 de tal forma que confirma as conclusões acima. Existe uma certa coerência nesses estudos em razão de que a retirada do último estudo faz com que a diferença entre as mortalidades para aspirina e placebo chegue a 2.31% , embora a comunidade médica estime que o efeito de um certo tratamento sobre outro passa a ser relevante a partir de pelo menos 5% de diferença entre o efeito observado. Mas, enfim, todos os estudos concordam que a aspirina é superior na diminuição da percentagem de óbitos do que o placebo.

No caso do modelo clássico ponderado, que é no geral mais recomendável que o anterior, os valores obtidos a favor da aspirina são medianamente significativos, em torno de 8 e 9%. Aqui a coerência se dá em outro sentido, como as variâncias são consideradas na ponderação, há uma tendência que os estudos se aproximem, uma vez que as amostras são muito grandes, o que naturalmente diminui o valor das variâncias aproximando os achados. Assim os dois estudos são fracamente significativos, mas têm alguma significância porque os valores estão mais ou menos próximos dos 5%. O problema da consideração da relevância clínica no entanto, permanece.

O modelo de partição difusa consegue, pelo menos aproximadamente, resolver os dois problemas. Em primeiro lugar ele tem capacidade de perceber as diferenças entre o trabalho com os 5 e os 6 estudos, fornecendo diferenças importantes entre eles. A medida de relevância do estudo menor é o dobro da do estudo maior. E mais, ele também é capaz de revelar que embora ambos os estudos não tenham forte relevância, o menor tem uma forte tendência a relevância (0.72 em 0.80) enquanto o segundo não a tem (0.36 em 0.80). O que reforça a necessidade de se trabalhar com o tamanho da amostra e do tamanho do efeito ("effect size") compostos com o "p-value" e ponderados pela informação que é medida pelo inverso da variância.

## 6. CONCLUSÕES E RECOMENDAÇÕES

As conclusões e recomendações são baseadas na revisão de literatura realizada, no desenvolvimento da metodologia proposta e nos resultados da aplicação proposta.

### 6.1 Conclusões

- Quando se modela um sistema, há que se negociar complexidade, incerteza e utilidade.
- O maior número de manifestações de incerteza possível deve ser incorporado explicitamente a um modelo que busca compreender um fenômeno do mundo real.
- Um segundo estágio da transição da forma tradicional de se tratar incerteza para a moderna tem sido desenvolvido nesse século, em particular com o desenvolvimento de teorias para tratar com a ignorância e a incerteza desenvolvidas a partir da década de 60.
- As diferentes manifestações da ignorância e da incerteza devem ser tratadas por uma teoria específica que tenha por objetivo o trato com a sua natureza. Em particular, a incerteza é apenas uma das manifestações da ignorância. E também a aleatoriedade é apenas uma das manifestações da incerteza.
- É possível representar graus de crença em termos de preferências subjetivas com condições suficientemente fortes que possibilitem o uso de probabilidades quantitativas.
- Alguns fenômenos podem ser melhor compreendidos se forem baseados em teorias que têm suporte em lógicas não-aristotélicas.
- É possível a construção de agendas comuns entre as diversas teorias de incerteza, com o desenvolvimento de medidas de probabilidade difusas, e inferência bayesiana difusa entre outras.
- Para minimizar o risco de vício de publicação, a meta-análise deve procurar não somente por pesquisas publicadas como artigos científicos, mas também por pesquisas publicadas na forma de "abstract". Nesse contexto, uma revisão sistemática de bancos de "abstracts" pode reduzir consideravelmente o risco de vício de publicação.

- Formas de cooperação internacional delineadas para construir bancos de dados são também úteis para diminuir o risco de se perder importantes informações clínicas.
- O planejamento do delineamento das pesquisas com o auxílio de profissional em estatística é de fundamental importância para a realização de pesquisas que envolvam procedimentos estatísticos de significância e muito particularmente para possibilitar a realização futura de meta-análises.
- A meta-análise deve ser realizada levando-se em conta seu aspecto exploratório e seu aspecto analítico com objetivos de inferência estatística.
- A estimação dos parâmetros do modelo GoM de Manton é muito restritivo para aplicação em meta-análise.
- Para a construção de um modelo de partição difusa para a meta-análise três variáveis independentes devem ser consideradas: o tamanho da amostra, o "effect size" e o "p-value", como variáveis difusas.
- É possível matematizar o conceito de relevância como medida de evidência, construído a partir de uma combinação linear das variáveis difusas independentes acima citadas.
- O método de partição difusa para a meta-análise é mais consistente na análise de um conjunto de experimentos clínicos que os métodos tradicionais, por conseguir incorporar o conhecimento do especialista através de variáveis com valores linguísticos.
- O método de partição difusa, por incorporar o tamanho da amostra e o "effect size" ao "p-value", consegue discriminar melhor estudos com diferentes graus de relevância.
- A inclusão da quantidade de informação, dada por funções da variância, ajuda a discriminar estudos com diferentes graus de relevância.
- As evidências não apontam a aspirina como um fator determinante no aumento da sobrevivência para pacientes após infarto do miocárdio.
- A meta-análise realizada para determinação da evidência da aspirina como benéfica no prolongamento de vida de pacientes após infarto do miocárdio foi de 36%, enquanto esperava-se um grau próximo a 80% de relevância.

## 6.2 Recomendações

- Ampliar a discussão sobre a natureza das medidas de ignorância, de forma a possibilitar a matematização de outras manifestações de ignorância não restrita a medidas de incerteza.
- Realizar um grande aprofundamento entre o relacionamento das teorias de probabilidade freqüentista, bayesiana e subjetiva, como forma de suporte para a ampliação do relacionamento dessas teorias com outras que procuram medir incerteza de natureza diferente da aleatoriedade. Sugere-se os trabalhos de POPPER (1959), HOWSON AND URBACH (1993) e MAYO (1996) e outros artigos desses autores referenciados nesse trabalho como ponto inicial para essa discussão, em especial seus aspectos filosóficos que não foram possíveis tratar em profundidade no presente trabalho.
- A questão conhecida como "file drawer problem", que consiste no "engavetamento" de artigos que não encontram significância estatística, em razão de as revistas científicas resistirem em sua publicação é um dos maiores problemas em meta-análise. Uma vez que os estudos que participam do procedimento de síntese de evidências serem considerados como uma amostra aleatória de todos os estudos realizados, isso consiste em uma suposição muito forte, para um estudo que na verdade é observacional. O método de partição difusa não faz essa consideração, mas no entanto não trata de forma mais pormenorizada do problema. Seria interessante na continuidade de formulação do método que esse problema fosse melhor detalhado.
- O estabelecimento das probabilidades subjetivas foram realizadas com a participação de um único especialista. Seria conveniente que essas probabilidades fossem estabelecidas por técnica que permitisse a participação de vários especialistas.
- A interpretação dos resultados da meta-análise devem ser realizadas por um grupo multi-disciplinar, para que se possa tirar toda informação que os artigos originais possuem e então procurar incluí-la no modelo proposto.
- O método proposto é geral o suficiente para aplicação em várias áreas do conhecimento. O trabalho foi realizado para uma específica aplicação em cardiologia. Ter-se-ia que

pensar em ampliar o procedimento para outras áreas de bio-estatística e de outras áreas do conhecimento como economia, administração, meio-ambiente, agronomia, reconhecimento de padrões e engenharia só para ficar nas áreas mais evidentes de aplicação.

- Na revisão sobre incerteza seria importante incluir mais especificamente estudos sobre o tratamento de incerteza na área de Inteligência Artificial. Por ser uma discussão muito ampla optou-se por fazer somente algumas referências à área. Um estudo mais aprofundado é necessário incluindo entre outros tópicos a Teoria dos Fatores de Certeza, a Teoria de Paul Cohen sobre Avaliação não-numérica da Incerteza e a Teoria de Probabilidades Comparativas.
- É possível o desenvolvimento de "software" que resolva problemas colocados para a realização da meta-análise através da partição difusa do espaço amostral. Tal "software" dependeria da solução de alguns problemas não completamente resolvidos nesse trabalho.
- Solucionados tais problemas, essa metodologia deveria ser apresentada a órgãos de controle de saúde, como forma de economia da alta quantidade de energia e recursos que são empregados em pesquisas isoladas.
- Verificar de que forma a lógica modal pode servir de ambiente comum às várias teorias que tratam com diferentes manifestações de incerteza.
- Procurar uma forma de desenvolver a meta-análise através do uso da lógica paraconsistente.

## Referências Bibliográfica

1. ABRAMI, P.C., P.A. COHEN AND S. D'APOLLONIA (1988). Implementation problems in meta-analysis. *Review of Educational Research*. Summer 1988; **58(2)**: 151-79. (MA)
2. ALGINA, J., R.C. BLAIR AND W.T. COOMBS (1995). A maximum test for scale: type 1 error rates and power. *Journal of Educational and Behavioral Statistics*. Spring 1995; **20(1)**: 27-39. (P)
3. ANELLO, C. AND J.L. FLEISS (1995). Exploratory or analytic meta-analysis: should we distinguish between them? (with discussion). *J.Clin.Epidemiol.* **48(1)**: 109-116. (MA)
4. ANTMAN, E.M., J. LAU, B. KUPELNICK, F. MOSTELLER AND T.C. CHALMERS (1992). A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. *Journal of the American Medical Association*. July 1992; **268(2)**: 240-8. (MA/ME)
5. ASPIRIN MYOCARDIAL INFARCTION STUDY RESEARCH GROUP (1980). A randomized controlled trial of aspirin in persons recovered from myocardial infarction. *Journal of Amer. Med. Assoc.* **243**: 661-9. (MA)
6. BARDOSSY, A., L. DUCKSTEIN AND I. BOGARDI (1993). Combination of fuzzy numbers representing expert opinions. *Fuzzy Sets and Systems*. **59**: 173-81. (F)
7. BARNES, E.C. (1998). Probabilities and epistemic pluralism. *British Journal of Philosophy of Science*. **49**: 31-47. (P/U)
8. BASU, D. (1975). Statistical information and likelihood. *Sankhya: The Indian Journal of Statistics. Serie A*. **37**: 1-71. (BP)
9. BAUMANN, P. (1998). Can reliabilists believe in subjective probability? *The philosophical quarterly*. **48 (191)**: 199-201. (BP/U)
10. BAYES, F.R.S. (1763). An essay towards solving a problem in the doctrine of chances. *Reprinted in Biometrika 1958*. **45**: 296-315. (BP)
11. BEARD, M.T., E.L. CURRY, K. EDWARDS, and B. NADAMS (1997). Advances in meta-analysis as a research method. *The ABNF Journal*. **8(5)**: 92-7. (MA)
12. BEGG, C.B. AND J.A. BERLIN (1988) Publication bias: a problem in interpreting medical data. *R. Statist. Soc., Serie A. Part 3*: 419-63. (MA/ME)
13. BELLMAN R. AND M. GIERTZ (1973). On the analytic formalism of the theory of fuzzy sets. *Information Sciences*. **5**: 149-56. (F)



14. BERLIN, J.A., N.M. LAIRD, H.C. SACKS AND T.C. CHALMERS (1989). A comparison of statistical methods for combining event rates from clinical trials. *Statistics in Medicine*. **8**: 141-51. (MA)
15. BERGER, J.O. (1985). In defense of the likelihood principle: axiomatics and coherency. *Bayesian Statistics*. **2**: 33-66. (BP)
16. BERRY, D.A. (1997). Teaching elementary bayesian statistics with real applications in science (with discussion). *The American Statistician*. **51(3)**: 241-74. (BP)
17. BETTENCOURT, B.A. AND N. MILLER (1996). Gender differences in aggression as a function of provocation: a meta-analysis. *Psychological Bulletin*. **119 (3)**: 422-47. (MA)
18. BEZDEK, J.C. (1976). Feature selection for binary data - medical diagnosis with fuzzy sets. *Proceedings of National Computer Conference*. pp. 1057-68. (F/ME)
19. BEZDEK, J.C. (1993). Fuzzy models – What are they, and why? *IEEE Transactions on Fuzzy Systems*. **1(1)**: 1-5. (F)
20. BIRNBAUM, A.(1962). On the foundations of statistical inference. *Journal of the American Statistical Association*. **57**: 269-326. (S/U)
21. BLACK, M. (1937). An exercise in logical analysis. *Philosophy of Science*. **4**: 427-55. Reprinted in *Int. J. General Systems* 1990. **17**: 107-28. (U)
22. BLAIR, A., J. BURG, J. FORAAN, H. GIBBS, S. GREENLAND, R. MORRIS, G. RAABE, D. SAVITZ, J. TETA, D. WARTENBERG, O. WONG AND R. ZIMMERMAN (1995). Guidelines for application of meta-anlaysis in environmental epidemiology. *Regulatory Toxicology and Pharmacology*. **22**: 189-97. (MA/ME)
23. BLINOWSKA A. AND L. DUCKSTEIN (1993). Medical applications of fuzzy logic - fuzzy patient state in arterial hypertension analysis. *SPIE*. 1-10. (F/ME)
24. BLOCKLEY, D.I., B.W. PILSWORTH, AND J.F. BALDWIN (1983) Measures of Uncertainty. *Civ. Eng. Syst.* **1**: 3-9. (U)
25. BORCH, K. (1977). The monster in Loch Ness. In: *New Developments in the applications of bayesian methods*. Edited by A. Aykac and C. Brumat. North-Holland Publishing Company. Amsterdam. **Cap. 16**: 273-8. (BP)
26. BOUCHON, B. (1987). Fuzzy inferences and conditional possibility distributions. *Fuzzy Sets and Systems*. **23**: 33-41. (F)
27. BOX, G.E.P. (1973). *Bayesian inference in statistical analysis*. Addison-Wesley Publishing Company. Massachusetts. 588p. (BP/S)
28. BOX, J.F. (1987). Guinness, Gosset, Fisher, and small samples. *Statistical Science*. **2(1)**: 45-52. (S)

29. BREDDIN K., D. LOEW, D. LECHNER, K. UBERLA AND E. WALTER (1979). Secondary prevention of myocardial infarction. Comparison of acetylsalicylic two-year prospective study. *Thomb. Haemost.* **40**: 225-36. (MA)
30. CANNER, P.L. (1983). Aspirin in coronary heart disease. Comparison of six clinical trials. *Israel J. Medical Sciences.* **19**: 413-23. (MA/ME)
31. CANNER, P.L. (1987). An overview of six clinical trials of aspirin in coronary heart disease. *Statistics in Medicine.* **6**: 255-63. (MA/ME)
32. CARLIN, J.B. (1992). Meta-analysis for 2x2 tables: a bayesian approach. *Statistics in Medicine.* **11**: 141-58. (MA/S)
33. CASELLA, G. AND M.T. WELLS (1996). Comparing p-values to Neyman-Pearson tests. In: *Bayesian Analysis in Statistics and Econometrics*. Edit by D.A. Berry, K.M. Chaloner and J.K. Geweke. John Wiley, New York. **Cap. 43**: 507-14. (P)
34. CHALMERS, I. (1991). Can meta-analyses be trusted? *The Lancet.* Dec. 1991; **338**: 1464-5. (MA/ME)
35. CHALMERS, T.C. (1987). Meta-analysis in clinical medicine. *Transactions of The American Clinical and Climatological Association.* **99**: 144-50. (MA/ME)
36. CHALMERS, T.C., H. LEVIN, H.S. SACKS, D. REITMAN, J. BERRIER AND R. NAGALINGAM (1987). Meta-analysis of clinical trials as a scientific discipline. I: Control of bias and comparison with large co-operative trials. *Statistics in Medicine.* **6**: 315-25. (MA/ME)
37. CHALMERS, T.C. (1993). Meta-analytic stimulus for changes in clinical trials. *Statistical Methods in Medical Research.* **2**: 161-72. (MA)
38. CHARLTON, B.G. (1996). The uses and abuses of meta-analysis. *Family Practice.* **13(4)**: 397-401. (MA)
39. CHEESEMAN, P. (1988). An inquiry into computer understanding (with comments). *Comp. Intell.* **4**: 58-78. (F/P)
40. CHEN, Y.H., W.J. WANG AND C.H. CHIU (1996). Membership values determination for fuzzy sets. *Proceedings of IEEE 5<sup>th</sup> International Conference on Fuzzy Systems. New Orleans, L.A., USA. Sept. 1996.* pp 1706-11. (F)
41. CHIANG, C.L., J.L. HODGES, JR. AND J. YERUSHALMY (1956). Statistical problems in medical diagnoses. *Proceedings of the 3<sup>rd</sup> Berkeley Symposium on Mathematical Statistics and Probability. Volume IV.* Edited by J. Neyman. pp. 121-33. (ME/S)
42. COCHRAN, W.G. (1983). *Planning and analysis of observational studies.* L.E. Moses and F. Mosteller, eds. John Wiley. New York. (S)

43. COHEN, J. (1988) *Statistical power analysis for the behavioral sciences*. Academic Press. New York. 570 p. (MA/S)
44. COLTON, T. (1974). *Statistics in medicine*. Little, Brown and Company. Boston. 372 p. (ME/S)
45. CONSONNI, G. AND P. VERONESSE (1995). A bayesian method for combining results from several binomial experiments. *Journal of the American Statistical Association*. **90(431)**: 935-44. (BP/MA)
46. COOK, R.J. AND S.D. WALTER (1997). A logistic model for trend in  $2 \times 2 \times K$  tables with applications to meta-analyses. *Biometrics*. **53**: 352-357. (MA/S)
47. COOPER, H. AND N. DORR (1995). Race comparisons on need for achievement: a meta-analytic alternative to Graham's narrative review. *Review of Educational Research*. Winter 1995; **65(4)**: 483-508. (MA)
48. CORNFIELD, J. (1956). A statistical problem arising from retrospective studies. *Proceedings of the 3<sup>rd</sup> Berkeley Symposium on Mathematical Statistics and Probability Volume IV*. Edited by J. Neyman. pp.135-48. (ME/S)
49. CORONARY DRUG PROJECT RESEARCH GROUP (1976). Aspirin in coronary heart disease. *J. Chron. Dis.* **29**:625-42. (MA/ME)
50. COSTA, N.C. A. DA (1993). *Sistemas formais paraconsistentes*. Ed. UFPR. 66p.
51. DEAR, K.B.G. AND C.B. BEGG (1992). An approach for assessing publication bias prior to performing a meta- analysis. *Statistical Science*. **7(2)**: 237-45. (MA)
52. DEBOECK, G.J. (1994). Why use fuzzy modeling? *AI in Finance*. Summer 1994; **1(2)**: 17-24. (F)
53. DE CAMPOS, L.M., J.F. HUETE AND S. MORAL (1994). Probability intervals: a tool for uncertain reasoning. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. **2(2)**: 167-96. (E/U)
54. DE FINETTI, B. (1972). *Probability, induction and statistics*. Wiley. London. 266p. (BP/ST)
55. DE FINETTI, B. (1974). *Theory of probability, a critical introductory treatment. Vol.1*. John Wiley. New York. (BP/S)
56. DE FINETTI, B. (1975). *Theory of probability, a critical introductory treatment. Vol.2*. John Wiley. New York. (BP/S)
57. DE FINETTI, B. (1977). Probability: beware of falsifications. In: *New developments in the applications of bayesian methods*. Edited by A. Aykac and C. Brumat. North Holland Publishing Company. Amsterdam. **Cap. 21**: 347-79. (BP/S)
58. DEGROOT, M.H. (1970). *Optimal statistical decisions*. McGraw Hill. New York. 489 p. (BP/P).
59. DEHLING, H.G., T.K. KIJKSTRA, J.J. GUICHELAAR, W. SCHAAFSMA, A.G.M. STEERNEMAN, T.J. WANSBEEK AND J.T. VAN DER ZEE

- (1996). Structuring the inferential contest. In: *Bayesian Analysis in Statistics and Econometrics*. Edited by D.A. Berry, K.M. Chaloner, and J.K. Geweke. John Wiley. New York. **Cap. 46**: 539-47. (BP)
60. DEMPSTER, A.P. (1967). Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*. **38**: 325-39. (S/U)
  61. DERSIMONIAN, R. AND N. LAIRD (1986). Meta-analysis in clinical trials. *Contr. Clin. Trials*. **7**: 177-88. (MA/ME)
  62. DETSKY, A.S., C.D. NAYLOR, K. O'ROURKE, A.J. MCGEER AND K.A. L'ABBÉ (1992). Incorporating variations in the quality of individual randomized trials into meta-analysis. *J. Clin Epidemiol.* **45(3)**: 255-65. (MA/ME)
  63. DICKERSIN, K. AND J.A. BERLIN (1992). Meta-analysis: state-of-the art. *Epidemiologic Reviews*. **14**: 154-76. (MA)
  64. DOLBI, G.R. (1982). The role of statistics in the methodology of the life sciences. *Biometrics*. **38**: 1069-83. (S/U)
  65. DOMBI, J. (1990) Membership function as an evaluation. *Fuzzy sets and systems* **35**: 1-21 (F)
  66. DRAPER, D. (1995). Inference and hierarchical modeling in the social sciences. *Journal of Educational and Behavioral Statistics*. Summer 1995; **20(2)**: 115-147. (MA)
  67. DUBOIS, D. AND H. PRADE (1980). *Fuzzy sets and systems: Theory and applications*. Academic Press. Boston. 393p. (F/P/U)
  68. DUBOIS, D. AND H. PRADE (1986). Fuzzy sets and statistical data. *European Journal of Operational Research*. **25**: 345-56. (F/S/U)
  69. DUBOIS, D. AND H. PRADE (1989). Fuzzy sets, probability and measurement. *European Journal of Operational Research*. **40**: 135-54. (F/P/U)
  70. DUBOIS, D. AND H. PRADE (1990). Rough fuzzy sets and fuzzy rough sets. *International Journal of General Systems*. **17**: 191-209. (F/U)
  71. DUBOIS, D. AND H. PRADE (1993). Fuzzy sets and probability: Misunderstandings, bridges and gaps. *2<sup>nd</sup> IEEE International Conference on Fuzzy System. San Francisco - CA, USA. Volume II*. pp. 1059-68. (F/P/U)
  72. DUBOIS, D. AND H. PRADE (1994a). A survey of belief revision and updating rules in various uncertainty models. *International Journal of Intelligent Systems*. **9**: 61-100. (U)
  73. DUBOIS, D. AND H. PRADE (1994b). Fuzzy sets - A convenient Fiction for modeling vagueness and possibility. *IEEE Transactions on Fuzzy Systems*. **2(1)**: 16-21. (F/P/U)
  74. DUBOIS, D. AND H. PRADE (1996). Focusing vs. revision in possibility theory. *Proceedings of IEEE 5<sup>th</sup> International Conference on Fuzzy Systems. New Orleans, LA, USA. Sept. 1996*. pp. 1700-5. (F/U)

75. DUCKSTEIN, L. AND A. BLINOWSKA (1995). Fuzzy classification of patient state with application to electrodiagnosis of peripheral polyneuropathy. *IEEE Transaction on Biomedical Engineering*. **42(8)**: 786- 92. (F/ME).
76. DUMOUCHEL, W. AND J.E. HARRIS (1983). Bayes methods for combining the results of cancer studies in humans and other species (with comments). *Journal of the American Statistical Association*. June 1983; **78(382)**: 293-315. (BP/MA)
77. DUMOUCHEL, W. AND P.G. GROER. (1989). A bayesian methodology for scaling radiation studies from animal to man. *Health Physics*. **57(supp.1)**: 411-18. (BP/MA/ME)
78. DUMOUCHEL, W. (1990). Bayesian metaanalysis. In: *Statistical Methodology in the pharmaceutical sciences*. Edited by D.A. Berry. Marcel Dekker, Inc. New York. **Cap. 15**: 509-28. (BP/MA)
79. DUMOUCHEL, W. (1994). Hierarchical bayes linear models for meta-analysis. *Technical Report #27*. Sept. 1994. National Institute of Statistical Sciences. pp. 1-26. (BP/MA/ME)
80. EDDY, D.M., V. HASSELBLAD AND R. SHACHTER (1992). *Meta-analysis by the confidence profile method*. Academic Press, Inc. Boston. 428p. (MA)
81. EFFRON, B. (1998). R.A. Fisher in the 21<sup>st</sup> century (with discussion). *Statistical Science*. **13(2)**: 97-122. (BP/P/U)
82. EGGER, M., G.D. SMITH, M. SCHNEIDER AND C. MINDER (1997). Bias in meta-analysis detected by a simple graphical test. *BMJ*. Sept. 1997; **315**: 629-34. (MA)
83. ELKAN, C. (1994). The paradoxical success of fuzzy logic. (with discussion). *IEEE Expert*. Pp. 1-45. (F/U)
84. ELLENBERG, J.H., M.H. GAIL AND N.L. GELLER (1997). Conversations with NIH statisticians: interviews with the pioneers of biostatistics at the United States National Institutes of Health. *Statistical Science*. **12(2)**: 77-335. (ME/S)
85. ELWOOD, P.C., A.L. COCHRANE, M.L. BURR, P.M. SWEETNAM, G. WILLIAMS, E.WELSBY, S.J. HUGHES AND R. RENTON (1974). A randomized controlled trial of acetyl salicylic acid in the secondary prevention of mortality from myocardial infarction. *Br. Med. Journal*. **1**: 436-40. (MA)
86. ELWODD, P.C. AND P.M. SWEETNAM (1979). Aspirin and secondary mortality after myocardial infarction. *Lancet* **ii**: 1313-15. (MA)
87. FAGARD, R.H., J.A. STAESSEN AND L. THIJIS (1996). Advantages and disadvantages of meta-analysis approach. *Journal of Hypertension*. **114(supl.2)**: 9-13. (MA)

88. FERRIS, C.D., F.E. GRUBBS AND C.L. WEAVER (1946). Operating characteristics for the common statistical tests of significance. *Annals of Mathematical Statistics*. **17**: 178-97. (S)
89. FIENBERG, S.E. (1992). A brief history of statistics in three and one-half chapters: a review essay. *Statistical Science*. **7(2)**: 208-25. (S)
90. GAINES, B.R. (1978). Fuzzy and probability uncertainty logics. *Information and Control*. **38**: 154-69. (F/P/U)
91. GARDNER, M. (1996). *Weird water and fuzzy logic*. Prometheus Book. New York. 260p. (F/U)
92. GART, J.J. (1992). On the combination of relative risk. *Biometrics*. **18**: 601-10. (MA/S)
93. GAVER JR, D.P. , D.DRAPER, P.K. GOEL, J.B.GREENHOUSE, L.V. HEDGES, C.N. MORRIS AND C. WATERNAUX (1992). *Panel on statistical issues and opportunities for research in the combination of information*. National Research Council. National Academic Press. Washington, D.C..217p. (MA)
94. GELBER, R.D., B.F. COLE, A. GOLDBIRSCH, G. BONADONA, A. HOWELL, C.S. MCARDLE, H.T. MOURIDSEN, R.D. RUBENS, AND K. WELVAART (1995). Adjuvant chemotherapy for premenopausal breast cancer: a meta-analysis using quality-adjusted survival. *Technical Reports Jan 1995. Dana-Farber Cancer Institute*. pp. 1-22. (MA/ME)
95. GELIJNS, A.C. (1990). *Modern methods of clinical investigation*. National Academy Press. Washington, D.C. 222p. (MA/ME)
96. GIBBONS, J.D. AND J.W. PRATT (1975). P-values: interpretation and methodology. *The American Statistician*. **29(1)**: 20-25. (P)
97. GIERE, R.N. (1997). Scientific inference: two points of view. *Philosophy of Science*. **Supplement 64(4)**: 180-4. (U)
98. GIEVENS, G.H., D.D. SMITH AND R.L. TWEEDIE (1997). Publication bias in meta-analysis: A Bayesian data-augmentation approach to account for issues exemplified in the passive smoking debate. *Statistical Science*. **12(4)**: 221-50. (MA/ME)
99. GLASS, G.V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher* **6**: 3-8. (MA)
100. GLASS, G.V. (1978). Integrating findings: the meta-analysis of research. *Review of Research in Education*. **5**: 351-79. (MA)
101. GLEISSER, M. (1998). A ciência precisa de crise para poder avançar. Folha de Sao Paulo. 21 de Junho de 1998. (U)
102. GOODMAN, S.N. (1989). Meta-analysis and evidence. *Controlled Clinical Trials*. **10**: 188-204. (MA)

103. GOOD, I.J. (1988). The interface between statistics and philosophy of science (with comments). *Statistical Science*. **3(4)**: 386-412. (S/U)
104. GRAHAM, S. (1994). Motivation in african americans. *Review of Educational Research*. Spring 1994; 64(1): 55-117. (MA)
105. GUEYFFIER, F., F. BOUTITIE, J.P. BOISSEL, J. COOPE, J. CUTLER AND T. EKBON (1995). Indiana: a meta-analysis on individual patient data in hypertension. Protocol and preliminary results. *Therapie* **50**: 353-62. (MA/ME)
106. GUEYFFIER, F., F. BOUTITIE, J.P. BOISSEL, S. POCOCK, J. COOPE, J. CUTLER, T. EKBON, R. FAGARD, L. FRIEDMAN, M. PERRY, R. PRIENEAS AND E. SCHRON (1997). Effect of antihypertensive drug treatment on cardiovascular outcomes in women and men. *Ann. Int. Med.* **126**: 761-7. (MA/ME)
107. GUREVITCH, J., L.L. MORROW, A. WALLACE AND J. S. WALSH (1992). A meta-analysis of competition in field experiments. *The American Naturalist*. October 1992; **140(4)**: 539-72. (MA)
108. GUREVITCH, J. AND L.V. HEDGES (1993). Meta-analysis: combining the results of independent experiments. In: *Design and analysis of Ecological Experiments*. Edited by S.M. Schneider and J. Gurevitch. Chapman and Hall. New York. **Cap. 17**: 378-98. (MA)
109. HEDGES, L.V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*. Summer 1981; **6(2)**: 107-28. (MA)
110. HEDGES, L.V. AND I. OLKIN (1985). *Statistical methods for meta-analysis*. Academic Press, Inc. Boston. 369 p. (MA)
111. HEDGES, L.V. (1990). *Directions for future methodology*. In: The future of Meta-analysis, pp.11-26. K.W. Wachter and M.L. Straf, eds. Russel Sage Foundation. New York. (MA)
112. HEDGES, L.V. (1992). Modeling publication selection effects in meta-analysis. *Statistical Science*. **7(2)**: 246-55. (MA)
113. HEDGES, L.V. AND J.L. VEVEA (1996). Estimating effect size under publication bias: small sample properties and robustness of a random effect selection model. *Journal of Educational and Behavioral Statistics*. Winter 1996; **21(4)**: 299-332. (MA)
114. HELLMAN, G. (1997). Bayes and Beyond. *Philosophy of Science*. **64**: 191-221. (BP/U)
115. HEWITT, P. (1991). Taming the literature. In: *Principles and Practice of Research. Strategies for Surgical Investigators*. 2<sup>nd</sup> ed.. Edited by H. Troidl, W.O. Spitzer, B. McPeck, D.S. Mulder, M.F. McKneally, A.S. Wechsler and C.M. Balch. Spring Verlag. New York. **Cap. 7**: 49-62. (MA/ME)

116. HISDAL, E. (1988) The philosophical issues raised by fuzzy set theory. *Fuzzy Sets and Systems*. **25**: 349-56. (F/P/U)
117. HODGES, J.S. (1996). The effect of partial-ordering utilities on bayesian design of sequential experiments. In: *Bayesian analysis in statistics and econometrics*. Edited by D.A. Berry, K.M. Chaloner and, J.K. Geweke. John Wiley. New York. **Cap. 44**: 515-25. (BP)
118. HOLZMANN, I.G. (1995). "Exploratory or analytic meta-analysis: should we distinguish between them": discussion. *J. Clin. Epidemiol.* **48(1)**: 117-8. (MA/S)
119. HOLDER, R.D. (1998). Hume on miracles: Bayesian interpretation, multiple testimony, and the existence of God. *British Journal of Philosophy of Science*. **49**: 49-65. (BP/U)
120. HOWELL, W.C. AND S.A.BURNETT (1978). Uncertainty measurement: a cognitive taxonomy. *Organizational Behavior and Human Performance*. **22**: 45-68. (P/U)
121. HOWSON, C. AND P. URBACH (1989). *Scientific reasoning: the bayesian approach*. 2<sup>nd</sup>. Ed.. Open Court Publishing Company. Chicago. 470p. (BP/P/U)
122. HOWSON, C. (1997a). A logic of induction. *Philosophy of Science*. **64**: 268-90.
123. HOWSON, C. (1997b). Error probabilities in error. *Philosophy of Science. Supplement 64(4)*: 185-94. (BP/P/U)
124. HOWSON, C. (1997c). Logic and probability. *British Journal of Science*. **48**: 517-31. (P/U)
125. HUNG, H.M.J., R.T. O'NEILL, P. BAUER, AND K. KOHNE (1997). The behavior of the p-value when the alternative hypothesis is true. *Biometrics* **53**: 11-22. (P)
126. HUNT, M. (1997). *How science takes stock. The story of meta-analysis*. Russel Sage Foundation. New York. 210p. (MA)
127. HUNTER, J.E. AND F.L. SCHMIDT (1990). *Methods of meta-analysis: correction error and bias in research findings*. Sage Publications. Newbury Park. 592p. (MA)
128. IYENGAR, S. AND J.B. GREENHOUSE (1998). Selections models and the file drawer problem (with comments). *Statistical Science*. **3(1)**: 109-35. (MA)
129. JACOB, F. (1973). *The logic of life. A history of heredity*. Princeton Science Library. New Jersey. 348 p. (P/U)
130. JONES, D.R. (1993). El metanálisis en los estudios epidemiológicos observacionales. *Bol. of Sanit. Panam.* **115(5)**: 438-45. (MA)
131. KALMANSON, D. AND H.F STEGALL (1975). Cardiovascular investigations and fuzzy sets theory. *The American Journal of Cardiology*. **35**: 80-4. (F/ME)



132. KANDEL, A. (1978). Fuzzy sets, fuzzy algebra, and fuzzy statistics. *Proceedings of the IEEE*. **66(12)**: 1619-39. (F/P/S)
133. KANDEL, A. (1979). On fuzzy statistics. *Advances in fuzzy set theory and applications*. Edited by M.M. Gupta, R.K. Ragade and R.R. Yager. North-Holland Publishing Company. 181-199. (F/S)
134. KANDEL, A. (1986). *Fuzzy mathematical techniques with applications*. Addison-Wesley Publishing Company. Massachusetts. 274p. (F/P)
135. KASSIRER, J.P. (1992). Clinical Trials and meta-analysis: What do they do for us? (editorial). *The New England Journal of Medicine*. July 1992. **327(4)**: 273-4. (MA/ME)
136. KNEEFE, R. (1998). Vagueness by numbers. *Mind. A quarterly review of philosophy*. **107(427)**: 566-79. (U)
137. KENDALL, D.G. (1974). Foundations of a theory of random sets. In: *Stochastic Geometry. (A tribute to the memory of Rollo Davidson)*. Edited by E.F. Harding and D.G. Kendall. John Wiley. **Cap. 6.2**: 322-76. (RS)
138. KLEMENT, E.P., W. SCHWYHLA. AND R. LOWEN.(1981). Fuzzy probability measures. *Fuzzy Sets and Systems*. **5**: 21-30. (F/P)
139. KLIR, G.J. (1987). Where do we stand on measures of uncertainty, ambiguity, fuzziness, and the like? *Fuzzy Sets and Systems*. **24**: 141-60. (F/U)
140. KLIR, G.J. (1988). *Fuzzy sets, uncertainty and information*. Prentice Hall, Englewood Cliffs, New Jersey. (F/U).
141. KLIR, G.J. (1989). Is there more to uncertainty than some probability theorists might have us believe? *International Journal of General Systems*. **15**: 347-78. (F/P/U)
142. KLIR, G.J. (1990a). An introduction to the especial issue on a quarter century of fuzzy systems. *Int. J. General Systems*. **17**: 89-93. (F)
143. KLIR, G. J. (1990b). A principle of uncertainty and information invariance. *Int. J. General Systems*. **17**: 249-75. (U)
144. KLIR, G.J. (1994). Multivalued logics versus modal logics: alternative frameworks for uncertainty modelling. In: *Advances in Fuzzy Theory and Technology - Volume II*. Edited by P.P. Wang. Duke University Press. Durham, North Carolina. Pp1-47. (F/U)
145. KLIR, G.J. (1995a). *Fuzzy sets and fuzzy logic. Theory and applications*. Prentice Hall. Englewood Cliffs. 574p. (F/P/S/U)
146. KLIR, G.J. (1995b). Principles of uncertainty: What are they? Why do we need them? *Fuzzy Sets and Systems*. **74**: 15-31. (F/U)
147. KLIR, G.J. (1995c). From classical sets to fuzzy sets: a grand paradigm shift. In: *Advances in fuzzy theory and technology. Volume III*. Edited by Paul P. Wang. Duke University -USA. pp. 5-30. (F/U)

148. KLIR, G.J. AND J.A. COOPER (1996). On constrained fuzzy arithmetic. *Proceedings of 5th IEEE International Conference on Fuzzy Systems*. pp. 285-90. (F)
149. KOCH, G.G., W.D. JOHNSON AND H.D. TOLLEY (1972). A linear models approach to the analysis of survival and extent of disease in multidimensional contingency tables. *Journal of American Statistical Association*. **67(340)**: 783-96. (ME/S)
150. KOSKO, B. (1992). *Neural networks and fuzzy systems*. Prentice Hall. Englewood Cliffs. 449p. (F)
151. KOSKO, B. (1993). *Fuzzy thinking: the new science of fuzzy logic*. Hyperion. NewYork. 318p. (F/P/U)
152. KOSKO, B. (1990). Fuzzines vs. probability. *Int. J. General Systems*. **17** : 211-40. (F/P/U)
153. KRUSE, R.F. (1982). The strong law of large numbers for fuzzy random variables. *Information Sciences*. **28**: 233-41. (F/P)
154. KRUSE, R.F. AND K.D. MEYER (1987). *Statistics with vague data*. D. Reidel Publishing Company. Dordrecht, Holland. 279p. (F/P/U)
155. KWAKERNAAK, H. (1978a). Fuzzy random variables – I. Definitions and theorems. *Information Sciences*. **15**: 1-29. (F/P)
156. KWAKERNAAK, H. (1978b). Fuzzy random variables – II. Algorithms and examples for the discrete case. *Information Sciences*. **17**: 253-78. (F/P)
157. LAPORTE, J. (1997). Essential membership. *Philosophy of Science*. **64**: 96-112. (U)
158. LAU, J., E.M. ANTMAN, J. JIMENEZ-SILVA J., B. KUPELNICK, F. MOSTELLER, AND T.C. CHALMERS (1992). Cumulative meta-anlysis of therapeutic trials for myocardial infarction. *The New England Journal of Medicine*. July 1992; **327(4)**: 248-54. (MA/ME)
159. LAVIOLETTE, M. AND J.W. SEAMAN JR. (1994). The efficacy of fuzzy representations of uncertainty (with discussion). *IEEE Transactions of Fuzzy Systems*. **2(1)**: 4-42. (F/P/S/U)
160. LAVIOLETTE, M., J.W. SEAMAN JR., J.D. BARRET AND W.H. WOODALL.(1995). A probabilistic and statistical view or fuzzy methods (with discussion). *Technometrics*. **37(3)**: 249-92. (F/P/S)
161. LAVALLEY, M. (1997). A consumer's guide to meta-analysis. *Arthritis Care and Research*. June 1997; **10(3)**: 208-13. (MA)
162. LAW, K.S. (1995). The use of Fisher's Z in Schmidt-Hunter-type meta-analysis. *Journal of Educational and Behavioral Statistics*. Fall 1995. **20(3)**: 287-306. (MA)

163. LECKY, F.E., R.A. LITTLE, AND P. BRENNAN (1996). The use and misuse of meta-analysis. *Journal of Accident and Emergency Medicine*. **13**: 373-8. (MA)
164. LEE, P.M. (1997). *Bayesian statistics*. 2<sup>nd</sup> ed.. John Wiley. New York. 344p. (BP)
165. LEFERING, R. AND E.A.M. NEUGEBAUER (1991). How to organize your reference material. In: *Principles and Practice of Research. Strategies for Surgical Investigators*. 2<sup>nd</sup> ed.. Edited by H. Troidl, W.O. Spitzer, B. McPeck, D.S. Mulder, M.F. McKneally, A.S. Wechsler and C.M. Balch. Spring Verlag. New York. **Cap. 8**: 63-68 (MA/ME)
166. LIANG, P. AND F. SONG (1996). What does probabilistic interpretation of fuzzy sets mean? *IEEE Transactions of Fuzzy Systems*. **4(2)**: 200-5. (F/P)
167. LINDLEY, D.V. (1982). Scoring rules and the inevitability of probability (with discussion). *International Statistical Review*. **50**: 1-26. (F/P/U)
168. LIGHT, R.J. AND D.B. PILLEMER (1984). *Summing up. The science of reviewing research*. Harvard University Press. Cambridge. 191p. (MA)
169. LIGHT, R.J. (1987). Accumulating evidence from independent studies: what we can win and what we can lose. *Statistics in Medicine*. **6**: 221-28. (MA/ME)
170. MANTEL, N. AND E. HAENSZEL (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer Inst.*. **22**: 719-48. (S)
171. MANTON, D.G., M.A. WOODBURY AND H.D. TOLLEY. (1994). *Statistical applications using fuzzy sets*. John Wiley. New York. 312p. (F/S)
172. MARINOFF, L. (1994). A resolution of Bertrand's paradox. *Philosophy of Science*. **61**: 1-24. (P)
173. MARTZ, H.F. AND M.C. BRYSON (1984). A statistical model for combining biased expert opinions. *IEEE Transactions on Reliability*. August 1984; **33(3)**: 227-32. (MA/S)
174. MAYO, D.G. (1996). *Error and growth of experimental knowledge*. The University of Chicago Press. Chicago. 493 p.(BP/P/U)
175. MAYO, D.G. (1997). Duhem's problem, the bayesian way, and error statistics, or "What's belief got to do with it?". *Philosophy of Science*. **64**: 222-44. (BP/P/U)
176. MAYO, D.G. (1997). Error statistics and learning from error: making a virtue of necessity. *Philosophy of Science - Supplement*. **64(4)**:195-212. (S/U)
177. MAYO, D.G. (1997). Response to Howson and Laudan. *Philosophy of Science*. **64(4)**: 323-333. (S/U)
178. MCNEILL, D. AND P. FREIBERGER (1993). *Fuzzy logic: the revolutionary computer technology that is changing our world*. Touchstone. New York. 295 p. (F/U)

179. MEHTA, C.R., N.R. PATEL AND R. GRAY (1985). Computing an exact confidence interval for the common odds ratio in several 2x2 contingency tables. *Journal of the American Statistical Association*. **80(392)**: 969-73. (MA)
180. MESSORI, A. (1997). Current controversies in the application of meta-analysis, with special reference to oncological treatments. *Pharm. World Sci.* **19(3)**: 152-8. (MA)
181. MORRIS, C.N. AND S.L. NORMAND (1992). Hierarchical models for combining information and for meta-analyses. *Bayesian Statistics*. **4**: 321-344. (MA)
182. MORRIS, P.A. (1977). Combining expert judgments: a bayesian approach. *Management Science*. **23(7)**: 679-92. (BP)
183. MOSTELLER, F. AND T.C. CHALMERS (1992). Some progress and problems in meta-analysis of clinical trials. *Statistical Science*. **7(2)**: 227-36. (MA)
184. NAKAMURA, A. AND J.M. GAO (1991). A logic for fuzzy data analysis. *Fuzzy Sets and Systems*. **39**: 127-32. (F)
185. NELSON, W. (1982). *Applied life data analysis*. John Wiley. New York. 634p. (S)
186. NEUGEBAUER, E.A.M., B. MCPEEK, AND V.A. LAWRENCE (1991). Systematically reviewing previous work. In: *Principles and Practice of Research. Strategies for Surgical Investigators*. 2<sup>nd</sup> ed.. Edited by H. Troidl, W.O. Spitzer, B. McPeek, D.S. Mulder, M.F. McKneally, A.S. Wechsler and C.M. Balch. Spring Verlag. New York. **Cap. 9**: 69-80. (MA/ME)
187. NEYMAN, J. AND B. TOKARSKA (1936). Errors of the second kind in testing "Student's" hypothesis. *Journal of the American Statistician Association*. **31**: 318-26 (S)
188. NGUYEN, H.T. (1977). On fuzziness and linguistic probabilities. *Journal of Mathematical Analysis and Applications*. **61**: 658-71. (F/P)
189. NGUYEN, H.T., KOSHELEVA, O.M., AND KREINOVICH, V. (1994). Is the success of fuzzy logic really paradoxical? Or What is the actual logic behind expert systems. (F/U)
190. NICKEERSON, R.S. (1996). Ambiguities and unstated assumption in probabilistic reasoning. *Psychological Bulletin*. **120(3)**: 410-33. (P/U)
191. NONY, P. M. CUCHERAT, M.C. HAUGH, AND J.P. BOISSEL (1995). Critical reading of the meta-analysis of clinical trials. *Therapie*. **50**: 339-51. (MA)
192. NORMAND, S.L.T. (1995). Meta-analysis software: a comparative review. *The American Statistician*. **49(3)**: 298-309. (MA)
193. OAKES, M. (1993). The logic and role of meta-analysis in clinical research. *Statistical Methods in Medical Research*. **2**: 147-60. (MA)

194. OKUDA, T., H. TANAKA AND K. ASAI (1978). A formulation of fuzzy decision problems with fuzzy information using probability measures of fuzzy events. *Information and Control*. **38**: 135-47. (F/P)
195. OLKIN, I. (1992). Meta-analysis: methods for combining independent studies. *Statistical Science*. **7(2)**: 226. (MA)
196. OSENBERG, C.W., O. SARNELLE AND S. COOPER (1997). Effect size in ecological experiments: the application of biological models in meta-analysis. *The American Naturalist*. Dec. 1997; **150(6)**: 798-812. (MA)
197. PACHECO, R, A. MARTINS, AND A. KANDEL (1997). On the power of fuzzy logic. (F/U)
198. PAGLIARO, L. (1997). Preparing and interpreting meta-analysis in clinical research. *Arch. It. Urol.* **LXIX(4)**: 217-25. (MA/ME)
199. PAN, Y., G.J. KLIR AND B. YUAN (1996). Bayesian inference based on fuzzy probabilities. *Proceedings of IEEE 5<sup>th</sup> International Fuzzy Systems New Orleans*. pp.1693-9. (BP/F)
200. PAN, Y. AND G.J. KLIR (1996). Bayesian inference based on interval-valued prior distributions and likelihoods. *Technical report for Binghamton University - New York*. pp. 1-19. (BP/F)
201. PAWLAK, Z. (1982). Rough sets. *International Journal of Computer and Information Sciences*. **11(5)**: 341-56. (U).
202. PENG, X.T., P. WANG AND A. KANDEL (1996a). Knowledge acquisition by random sets. *International Journal of Intelligent Systems*. **11**:149-83. (RS)
203. PENG, X.T., P. WANG AND A. KANDEL (1996b). On random sets, knowledge acquisition, and pattern recognition. *International Journal of Intelligent Systems*. **11**: 149-83. (RS)
204. PERSANTINE-ASPIRIN REINFARCTION STUDY RESEARCH GROUP (1980). Persantine and aspirin in coronary heart disease. *Circulation*. **62**: 449-61. (MA)
205. PETO, R. (1987). Why do we need systematic overviews of randomized trials? (with discussion). *Statistics in Medicine*. **6**: 233-40. (MA/ME)
206. PIASECKI, K. (1985). Probability of fuzzy events defined as denumerable additivity measure. *Fuzzy Sets and Systems*. **17**: 271-84. (F/FM/P)
207. PIASECKI, K. (1986). On the Bayes formula for fuzzy probability measures. *Fuzzy Sets and Systems*. **18**: 183-85. (BP/F/FM/P)
208. PIEGORSCH W.W., E.P. SMITH, D. EDWARDS, AND R.L. SMITH (1998). Statistical advances in environmental science. *Statistical Science*. **13(2)**: 186-208. (MA)
209. POIRIER, D.J. (1996). Window washing: a subjective bayesian perspective of diagnostic checking. In: *Bayesian Analysis in Statistics and Econometrics*.

- Edited by D.A. Berry, K.M. Chaloner, and J.K. Geweke. John Wiley. New York. **Cap. 45:** 527-37. (BP)
210. POPPER, K.R. (1959) *The logic of scientific discovery*. Routledge Ed. 480 p. (P/U)
  211. PRIGOGINE, I. (1997). *The end of certainty. Times, chaos, and the new laws of nature*. The Free Press. New York. 228 p. (P/U)
  212. PURI, M.L. AND D.A. RALESCU (1985). The concept of normality for fuzzy random variables. *The Annals of Probability*. **13(4)**: 1373-79. (FR)
  213. PURI, M.L. AND D.A. RALESCU.(1986). Fuzzy random variables. *Journal of Mathematical Analysis and Applications*. **114**: 409-22. (FR)
  214. RALESCU, D. (1994). Inequalities for fuzzy random variables. *Proceedings of the Brazil-Japan Joint Symposium on Fuzzy Systems, July 1994*. pp.52-4. (FR)
  215. RAVNSKOV, U. (1992). Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. *BMJ*. **305**: 15-9. (MA/ME)
  216. RAU, G., K. BECKER, R. KAUFMANN, AND H.J. ZIMMERMANN (1995). Fuzzy logic and control: principal approach and potential applications in medicine. *Artif. Organs.* **19(1)**: 105-12. (F/ME)
  217. RAY, K. (1993). A probabilistic approach to approximate reasoning. *Journal off the IETE*. **39(4)**: 225-34. (E/P)
  218. ROGER, M. AND Y. HOSHIAI (1990). The future looks "fuzzy". *Newsweek*. May 28, 1990; pp. 46-7. (F)
  219. ROMER C., A. KANDEL AND M. FRIEDMAN. On fuzzy random variables. (FR)
  220. ROMER, C. AND A. KANDEL (1995a). Constraints on belief functions imposed by fuzzy random variables. *IEEE Transactions on Systems, Man and Cybernetics*. **25(1)**: 86-99. (FR)
  221. ROMER, C. AND A. KANDEL (1995b). Fuzzy partitions of the sample space and fuzzy parameter hypothesis. *IEEE Transactions on Systems, Man and Cybernetics*. **25(9)**: 1314-22. (E/F/FR)
  222. ROMER, C. AND A. KANDEL (1995c). Statistical tests for fuzzy data. *Fuzzy Sets and Systems*. **72**: 1-26. (FM/S)
  223. ROMER, C. AND A. KANDEL (1995d). Applicability analysis of fuzzy inference by means of generalized Dempster-Shafer Theory. *IEEE Transactions on Fuzzy Systems*. **3(4)**: 448-79. (E/F)
  224. ROSENTHAL, R. (1979). The "file drawer problem" and tolerance for null hypothesis. *Psychological Bulletin*. **86**: 638-41. (MA,S)
  225. ROSENTHAL, R. (1984). Meta-analysis procedures for social research. Sage. Beverly Hills, CA. (MA)

226. ROSS, T.J. (1996). *Fuzzy logic with engineering applications*. McGraw Hill. New York. 600p. (F/P/U)
227. ROYALL, R.M. (1991). Ethics and statistics in randomized clinical trials. *Statistical Science*. 6(1): 52-88. (ME/S)
228. RUBIN, D.B. (1981). Estimation in parallel randomized experiments. *Journal of Educational Statistics*. Winter 1981; 6(4): 377-400. (MA)
229. SACKS, H.S., J. BERRIER, D. REITMAN, V.A. ANCONA-BERK, AND T.C. CHALMERS (1987). Meta-analyses of randomized controlled trials. *The New England Journal of Medicine*. Feb. 1987; 316(8): 450-5. (MA/ME)
230. SANDRI, S.A., D. DUBOIS AND H.W. KALFSBEEK (1995). Elicitation, assessment, and pooling of expert judgments using possibility theory. *IEEE Transactions on Fuzzy Systems*. 3(3): 313-35. (F/U)
231. SANFORD, D.H. (1995). Uses and abuses of fuzziness in philophy: A selective survey of how recent philosophical logic and metaphysics treat vagueness. *International Journal of General Systems*. 23: 271-7. (F)
232. SCHECHTER, M.T., F.E. LEBLANC, AND V.A. LAWRENCE (1991). Critical appraisal of published research. In: : *Principles and Practice of Research. Strategies for Surgical Investigators*. 2<sup>nd</sup> ed.. Edited by H. Troidl, W.O. Spitzer, B. McPeck, D.S. Mulder, M.F. McKneally, A.S. Wechsler and C.M. Balch. Spring Verlag. New York. Cap. 10: 81-7. (MA/ME)
233. SCHERVISH, M.J.(1996). P-values: what they are and what they are not. *The American Statistician*. 50(3): 203-206. (P)
234. SCHNATTER, S.F. (1991). On the propagation of fuziness of data. *Environmetrics*. 2(2): 241-52. (F/S)
235. SCHNATTER, S.F. (1992). On statistical inference for fuzzy data with applications to descriptive statistics. *Fuzzy Sets and Systems*. 50: 143-65. (F/S)
236. SCHNATTER, S.F. (1993). On fuzzy bayesian inference. *Fuzzy Sets and Systems*. 60: 41-58. (F/BP)
237. SCHNEIDER, M. AND A. KANDEL (1994). On uncertainty management in fuzzy inference procedures. *Information Sciences*. 79: 181-90. (F/U)
238. SCHRAM, C.M. (1996). A meta-anlaysia of gender differences in applied statistics achievement. *Journal of Educational and Behavioral Statistics*. Spring 1996; 21(1): 55-70. (MA)
239. SHAFER, G. (1976). *A mathematical theory of evidence*. Princeton University Press. Princeton. (E,U).
240. SHAFER, G. (1978). Non-additive probabilities in the work of Bernoulli and Lambert. *Archieve for History of Exact Sciences*. 19: 309-70. (B/P/S/U)
241. SHAFER, G. (1987). Probability judgment in artificial intelgence and expert systems. *Statistical Science*. 2(1): 3-44.(P)

242. SMETS, P. (1995). Non standart probabilistic and non probabilistic representations of uncertainty. In: *Advances in fuzzy theory and technology. Volume III*. Edited by Paul P. Wang - Duke University - USA. pp 125-54. (P/U)
243. SMITH, T.C., D.J. SPIEGELHALTER, AND A. THOMAS (1995). Bayesian approaches to random-effects meta-anlaysis: a comparative study. *Statistics in Medicine*. **14**: 2685-99. (ME/S)
244. SMITHSON, M. (1989). *Ignorance and uncertainty: emerging paradigms*. Springer-Verlag. New York. 393p. (U)
245. SOOFL, E.S. (1996). Information theory and bayesian statistics. In: *Bayesian Analysis in Statistics and Econometrics*. Edited by D.A. Berry, K.M. Chaloner, and J.K. Geweke. John Wiley. New York. **Cap. 15**: 179-89. (BP)
246. SPIEGELHALTER, D.J. AND L.S. FREEDMAN (1988). Bayesian approaches to clinical trials. *Bayesian Statistics*. **3**: 453-77 (BP)
247. STALLINGS, E. (1977). Fuzzy set theory versus bayesian statistics. *IEEE Transactions on Systems, Man, and Cybernetics*. 216-9. (BP/F)
248. STEIMANN, F. (1997). Fuzzy set theory in medicine (editorial). *Artificial Intelligence in Medicine*. **11**: 1-7. (F/ME)
249. STEWART, L.A., D.J. GIRLING, M.K.B PARMAR, AND D. MACHIN (1991). Can meta-analyses be trusted? *The Lancet*. Dec. 1991; **338**:1465. (MA/ME)
250. STIGLER, S.M. (1986). *The history of meta-analysis: The measurement of uncertainty before 1900*. Harvard University Press. Cambridge. (MA)
251. STRAM, D.O. (1996). Meta-analysis of published data using a linear mixed-effects model. *Biometrics*. **52**: 536-44. (MA/S)
252. STREVENS, M. (1998). Inferring probabilities from symmetries. *Nous*. **32(2)**: 231-46. (P/U)
253. SUBRAMANIAN, D., R. GREINER, AND J. PEARL (1997). The relevance of relevance. *Artificial Inteligence*. **97**: 1-5. (U)
254. TANAKA H., H. ISHIBUCHI, AND I. HAYASHI (1993). Identification method of possibility distributions and its application to discriminant analysis. *Fuzzy Sets and Systems*. **58**: 41-50. (F/S)
255. THOMPSON, S.G. AND S.J. POCOCK (1991). Can meta-analyses be trusted? *The Lancet* Nov. 1991. **338**:1127-30. (MA/ME)
256. THOMPSON, S.G. (1993). Controversies in meta-analysis: the case of the trials of serum cholesterol reduction. *Statistical Methods in Medical Research*. **2**: 173-92. (MA/ME)
257. TRIBUS, M. (1979). Comments on 'fuzzy sets, fuzzy algebra, and fuzzy statistics'. *Proceedings of the IEEE*. **67(8)**: 1168-9. (F/S)



258. TRIBUS, M. (1980) Fuzzy sets and bayesian methods applied to the problem of literature search. *IEEE Transactions on Systems, Man, and Cybernetics*. **10(8)**: 501-2. (BP/F)
259. TWEEDIE, R. (1998). Consulting: real problems, real interactions, real outcomes. *Statistical Science*. **13(1)**: 1-29. (S)
260. TWEEDIE, R.L., D.J. SCOTT, B.J. BIGGERSTAFF AND K.L. MENGENSEN (1994). Bayesian meta-analysis, with application to studies of ETS and lung cancer. *Technical Reports Nov. 1994*. Colorado State University. pp.1-23. (BP/MA/ME)
261. UTTS, J. (1991). Replication and meta-analysis in parapsychology (with comments). *Statistical Science*. **6(4)**: 363-403. (MA)
262. VERDINELLI, I, K. ANDREWS, K. DETRE AND P. PEDUZZI (1996). The bayesian approach to meta-analysis: a case-study. *Technical Report*. Carnegie Mello University – Pittsburgh. pp.1-16. (BP/MA)
263. VIERTL, R. (1987). Is it necessary to develop a fuzzy bayesian inference? In: *Probability and bayesian statistics*. Edited by R. Viertl. Plenum Press. New York. pp.471- 5. (BP/F)
264. VIERTL, R. (1990). Statistical inference for fuzzy data in environmetrics. *Environmetrics* **1(1)**: 37:42. (F/S)
265. VIERTL, R. (1995). Statistics with fuzzy data. In: *Mathematical and statistical methods in artificial intelligence*. Edited by G.D. Riccia, R. Kruse, and R. Viertl. Springer-Verlag. New York. pp. 33-49. (F/S)
266. VIERTL, R. (1996). *Statistical methods for non-precise data*. Library of Congress Cataloging-in-Publication Data. Boca Raton, Florida. 191p. (S/U)
267. VIERTL, R. AND H. HULE (1991). On Bayes' theorem for fuzzy data. *Statistical Papers*. **32**: 115-22. (BP/F)
268. VITEZ, T.S., R. WADA AND A. MACARIO (1996). Fuzzy logic: theory and medical applications. *Journal of Cardiothoracic and Vascular Anesthesia*. October 1996; **10(6)**: 800-8. (F/ME)
269. VON MISSES, R. (1957). *Probability, statistics and truth*. 2th ed.. Pitman Press. London. 244p. (P/S/U)
270. WACHTER, K.W. AND M.L. STRAF (1990). *The future of meta-analysis*. Russel Sage Foudation. New York. 210 p. (MA)
271. WALLEY, P. (1991). *Statistical reasoning with imprecise probabilities*. Chapman and Hall. London. 706p. (BP/E/F/FM/FR/P/RS/S/U)
272. WANG, Z. AND G. KLIR (1992). *Fuzzy measure theory*. Plenum Press. New York. 353 p. (F/M)
273. WASSERMAN, L. AND J.B KADANE (1996). Entropy when probabilities are imprecise. In: *Bayesian Analysis in Statistics and Econometrics*. Edited by

- D.A. Berry, K.M. Chaloner, and J.K. Geweke. John Wiley. New York. **Cap. 47: 549-55. (BP/U)**
274. WEST, R.R. (1993). A look at the statistical overview (or meta-analysis). *Journal of The Royal College of Physicians of London*. **27(2): 111-5. (MA/ME)**
275. WOODBURY, M.A. (1963). Inapplicabilities of Bayes' theorem to diagnosis. *Proc. of the 5<sup>th</sup> International Conference on Medicin and Eletronics*. 861-8. (BP/ME)
276. WOODBURY, M.A.(1974). Clinical pure types as a fuzzy partition. *Journal of Cybernetics*. **4(3): 111-21. (F/ME)**
277. WOODBURY, M.A., J. CLIVE AND A. GARSON JR. (1978). Mathematical thypology: a grade of membership technique for obtaining disease definition. *Computers and Biomedical Research*. **11: 277-98. (F/ME)**
278. YAGER, R.R. (1980). Fuzzy sets, probabilities, and decision. *Journal of Cybernetics*. 10: 1-18. (F/P)
279. YAGER, R.R. (1986). Expected values from probabilities of fuzzy subsets. *European Journal of Operational Research*. **25: 336-44. (F/P)**
280. YAGER, R.R. (1992). On the specificity of a possibility distribution. *Fuzzy Sets and Systems*. **50: 279-92. (P/F)**
281. YOUNG, P., S. PARKINSON, AND M. LEES (1996). Simplicity out of complexity in environmental modeling: Occam's razor revisited. *Journal of Applied Statistics*. **3(2 and 3): 165-210. (S)**
282. YUAN, B. AND G.J. KLIR (1995). Bayesian inference based upon fuzzy events. *IEEE Proceedings of ISUMA-NAFIPS 1995*. pp. 90-4. (BP/F)
283. YUSUF, S., R. PETO, J. LEWIS, R. COLLINS, AND P. SLEIGHT (1985). Beta blockage during and after myocardial infarction: An overview of the randomized trials. *Progress in Cardiovascular Diseases*. **XXVII(5): 335-71. (MA)**
284. YUSUF, S., D. ZUCKER, P. PEDUZZI, L.D. FISHER, T. TAKARO, J.W. KENNEDY, K. DAVIS, T. KILLIP, E. PASSAMANI, R. NORRIS, C. MORRIS, V. MATHUR, E. VARNAUSKAS, AND T.C. CHALMERS (1994). Effect of coronary artery bypass graft surgery on survival: overview of 10-year results from randomized trial by the coronary artery bypass surgery trialists collaboration. *The Lancet*. **344: 563-70. (MA/ME)**
285. ZADEH, L.A. (1965) Fuzzy sets. *Information and Control*. **8: 338-53. (F)**
286. ZADEH, L.A. (1968). Probability measures of fuzzy events. *Journal of mathematical Analysis and Applications*. **23: 421-7. (F/P)**
287. ZADEH, L.A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*. **1: 3-28. (F/P)**
288. ZADEH, L.A. (1983). The role of fuzzy logic in the management of uncertainty in expert systems. *Fuzzy Sets and Systems*. **11: 199-227. (F/U)**

289. ZADEH, L.A. (1984). Fuzzy probabilities. *Information Processing and Management*. **20(3)**: 363-72. (F/P)
290. ZADEH, L.A. (1986). Is probability theory sufficient for dealing with uncertainty in AI: A negative view. *Uncertainty in Artificial Intelligence*. Edited by L.N. Kanal and J.F. Lemmer. Elsevier Science Publishers B. B., North-Holland. 103-116. (F/P/U)
291. ZADEH, L.A. (1988). Fuzzy logic. *Proceedings of IEEE*. 0018-9162/88/0400-0083\$01.00 IEEE. pp 83-93. (F)
292. ZADEH, L.A. (1990). The birth and evolution of fuzzy logic. *International Journal of General Systems*. **17**: 95-105. (F)
293. ZHANG, X. AND L. DUCKSTEIN. (1993). Combination of radiologists' diagnosis by a fuzzy number approach. *2<sup>nd</sup> IEEE International Conference on Fuzzy Systems- S. Francisco - California. Volume II*. pp.1293-97. (F)

(BP) - Bayesian Probability, (E) - Evidence Theory, (F) - Fuzzy, (FM) - Fuzzy Measure Theory, (FR) - Fuzzy Random Variable, (MA) - Meta-Analysis, (ME) - Medicine, (P) - Probability, (RS) - Random Sets, (S) - Statistics, (U) - Uncertainty