

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE ESTATISTICA MATEMATICA E COMPUTAÇÃO
CIENTIFICA

DEPARTAMENTO DE ESTATISTICA

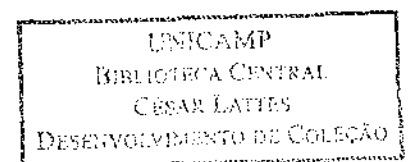
CONSIDERAÇÕES SOBRE A RELAÇÃO ENTRE
DISTRIBUIÇÕES DE CAUDA PESADA E CONFLITOS DE
INFORMAÇÃO EM INFERÊNCIA BAYESIANA

James Dean Oliveira dos Santos Junior /

Orientadora: Verónica Andrea González-López /

Co-orientadora: Laura Letícia Ramos Rifo /

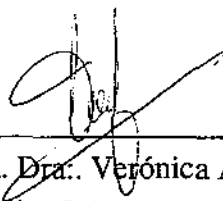
Campinas - SP 2006



CONSIDERAÇÕES SOBRE A RELAÇÃO ENTRE DISTRIBUIÇÕES DE CAUDA PESADA E CONFLITOS DE INFORMAÇÃO EM INFERÊNCIA BAYESIANA

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por James Dean Oliveira dos Santos Junior e aprovada pela comissão julgadora.

Campinas, 19 de março de 2007



Prof. Dra. Verónica Andrea
González-López
Orientador

Banca Examinadora:

1. Profa. Dra. Verónica Andrea González-López (IMECC-UNICAMP)
2. Prof. Dr. Carlos Alberto de Bragança Pereira (IME-USP)
3. Prof. Dr. Ronaldo Dias (IMECC-UNICAMP)

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica, UNICAMP, como requisito parcial para obtenção do Título de MESTRE em ESTATÍSTICA

| | |
|----------------------------|---------------------------------------|
| UNIDADE | BC |
| Nº CHAMADA: | TIUNICAMP |
| | Sa59c |
| V. _____ | Ed. _____ |
| TOMBO BCI | 72016 |
| PROC. | 16.145.07 |
| C <input type="checkbox"/> | D <input checked="" type="checkbox"/> |
| PREÇO | 4,00 |
| DATA | 18/04/07 |
| BIB-ID | 407529 |

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO IMECC DA UNICAMP**

Bibliotecária: Maria Júlia Milani Rodrigues – CRB8a / 2116

| | |
|--|---|
| <p>Santos Junior, James Dean Oliveira dos</p> <p>Sa59c</p> | <p>Considerações sobre a relação entre distribuições de cauda pesada e conflitos de informação em inferência bayesiana / -- Campinas, [S.P. :s.n.], 2007.</p> <p>Orientador : Verónica Andrea González-López</p> <p>Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.</p> <p>1. Teoria bayesiana de decisão estatística. 2. Distribuição (Probabilidades). 3. Análise de sensibilidade. I. González-López, Verónica Andrea. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.</p> |
|--|---|

Título em inglês: Considerations on the relation between heavy tailed distributions and conflict of information in bayesian inference.

Palavras-chave em inglês (Keywords): 1. Bayesian theory and statistic decision. 2. Distribution (Probability). 3. Sensibility analysis.

Área de concentração: Inferência Bayesiana

Titulação: Mestrado em Estatística

Banca examinadora: Profa. Dra. Verônica Andréa González-López (IMECC-UNICAMP)
 Prof. Dr. Carlos Alberto de Bragança Pereira (IME-USP)
 Prof. Dr. Ronaldo Dias (IMECC-UNICAMP)

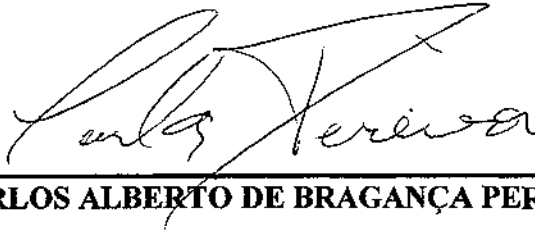
Data da defesa: 13/03/2007

Programa de Pós-Graduação: Mestrado em Estatística

Dissertação de Mestrado defendida em 13 de março de 2007 e aprovada pela Banca
Examinadora composta pelos Profs. Drs.



Prof (a). Dr (a). VERÓNICA ANDREA GONZÁLEZ LÓPEZ



Prof (a). Dr (a). CARLOS ALBERTO DE BRAGANÇA PEREIRA



Prof (a). Dr (a). RONALDO DIAS

2007.16743

*Aos meus pais James e Juventina
Juventina e aos meus irmãos
Dafny e João Paulo.*

Agradecimentos

Agradeço aos meus pais, por todo o apoio e incentivo que me deram, sem o qual não teria chegado até aqui.

Aos meus irmãos, por todo o carinho e apoio recebido nesses últimos anos.

Aos meus professores da graduação na Universidade Federal do Amazonas, que me incentivaram (e incentivam) em toda a minha história acadêmica. Agradeço especialmente a Maria Ivanilde Araújo por ter sempre me dado bronca nos momentos certos (finalmente aprendi a ser organizado Tia).

Aos meus grandes amigos de graduação: José Clelto Barros Gomes (Doutor), Antônio Balieiro (Toin), Robério Rebouças (o Robit), Francisco de Oliveira Farias (Chiquin), Francisco Alves dos Santos (Chicão) e Rômulo Costa e Silva.

Aos professores da Unicamp, por todo o conhecimento que me foi passado.

À Laura Letícia Ramos Rifo, pela paciência, pelo apoio, por todos os “minutinhos” roubados fora de horário e por ter acreditado que este trabalho daria frutos. Sem ela, este trabalho mal passaria de uma idéia.

À Verónica Andrea González-López por ter compreendido todas as minhas faltas e por ter sido tão prestativa. Sem a sua correção final este trabalho teria ficado técnico demais e pouco acessível para uma primeira leitura.

Aos membros da banca, Carlos Alberto de Bragança Pereira e Ronaldo Dias, pela atenção para com o meu trabalho e pelas valiosas sugestões para o seu melhoramento.

Aos amigos Plínio Lucas Dias Andrade, Paula Fonte Sanches, Andressa Pinheiro,

André Luiz Delai, Flávia Melo da Cunha e Carlos Gustavo Rangel por todos os momentos agradáveis e por todo apoio nesse período de mestrado.

À CAPES, a grande benfeitora, sem a qual pouco provavelmente eu teria obtido este título.

*Pode perdoar-se a um homem
a criação de alguma coisa útil
contanto que ele não a admire.
A única justificativa para a criação
de uma coisa inútil é que ela seja
admirada intensamente.
Toda arte é absolutamente inútil*

O Retrato de Dorian Gray, Oscar Wilde

Resumo

Em inferência bayesiana lidamos com informações provenientes dos dados e com informações a priori. Eventualmente, um ou mais *outliers* podem causar um conflito entre as fontes de informação. Basicamente, resolver um conflito entre as fontes de informações implica em encontrar um conjunto de restrições tais que uma das fontes domine, em certo sentido, as demais. Têm-se utilizado na literatura distribuições amplamente aceitas como sendo de cauda pesada para este fim. Neste trabalho, mostramos as relações existentes entre alguns resultados da teoria de conflitos e as distribuições de caudas pesadas. Também mostramos como podemos resolver conflitos no caso locação utilizando modelos subexponenciais e como utilizar a medida *credence* para resolver problemas no caso escala.

Palavras-Chave: inferência bayesiana, conflitos de informação, distribuições de caudas pesadas

Abstract

In bayesian inference we deal with information proceeding from the data and prior information. Eventually, one or more outliers can cause a conflict between the sources information. Basically, to decide a conflict between the sources of information implies in finding a set of restrictions such that one of the sources dominates, in certain sense, the outhter. Widely distributions have been used in literature as being of heavy tailed for this end. In this work, we show the relations between some results of the theory of conflicts and the heavy tailed distributions. Also we show how we can decide a conflicts in the location case using subexponential models and how to use the measure credence to decide problems in the scale case.

Keywords: bayesian inference, conflict of information, heavy tailed distributions

Sumário

| | | |
|----------|--|----------|
| 1 | Introdução | 1 |
| 1.1 | Fontes de informação | 1 |
| 1.2 | Conflitos de informação | 2 |
| 1.3 | Outliers | 5 |
| 1.4 | Por que distribuições de cauda pesada? | 6 |
| 1.5 | Organização | 6 |
| 2 | Distribuições de cauda pesada | 9 |
| 2.1 | Classes de distribuições de cauda pesada | 10 |
| 2.1.1 | A classe de cauda longa | 10 |
| 2.1.2 | A classe de variação dominada | 11 |
| 2.1.3 | A classe subexponencial | 12 |
| 2.1.4 | A classe de variação regular | 15 |
| 2.1.5 | Relações entre as classes de distribuições de cauda pesada | 17 |

| | | |
|----------|---|-----------|
| 2.1.6 | A classe de densidades SD | 21 |
| 2.2 | Distribuições resistentes e propensas a outliers | 22 |
| 2.2.1 | Distribuições resistentes a outliers | 22 |
| 2.2.2 | Distribuições propensas a outliers | 25 |
| 2.2.3 | Classes de distribuição segundo suas propriedades relacionadas a outliers | 28 |
| 2.3 | Relações entre distribuições PO e as distribuições de cauda pesada | 29 |
| 3 | Conflitos de informação | 37 |
| 3.1 | Mais sobre conflitos de informação | 37 |
| 3.2 | O teorema de Dawid | 40 |
| 3.2.1 | As condições de O'Hagan | 44 |
| 3.3 | O teorema de Pericchi e Sansó | 47 |
| 3.4 | Credence | 49 |
| 3.4.1 | Definições e propriedades básicas | 49 |
| 3.4.2 | O uso de credence em conflitos de informação | 56 |
| 3.5 | Resolvendo conflitos de informação em \mathcal{RV} | 59 |
| 3.5.1 | Estrutura parâmetro de escala | 59 |
| 3.5.2 | Rejeitando a informação a priori | 62 |
| 3.5.3 | Estrutura locação | 63 |
| 4 | Resultados | 65 |

| | | |
|----------|---|-----------|
| 4.1 | O Teorema de Dawid e distribuições em \mathcal{L} | 65 |
| 4.2 | Uma nota sobre o teorema de Pericchi e Sansó | 67 |
| 4.3 | Lidando com conflitos utilizando densidades em SD | 68 |
| 4.3.1 | Rejeitando a informação dos dados para uma única observação . . . | 68 |
| 4.3.2 | Rejeitando várias observações | 77 |
| 4.3.3 | Rejeitando a informação a priori | 81 |
| 4.4 | Caracterização de credence | 81 |
| 4.5 | Utilizando credence para problemas de escala | 83 |
| 5 | Considerações finais | 89 |

Lista de Figuras

| | | |
|-----|---|----|
| 2.1 | Diagrama das relações entre as classes apresentadas | 19 |
| 2.2 | Acima: estimativas para $P(X_{(n)} - X_{(n-1)} > 1)$; Abaixo: estimativas para $P(X_{(n)}/X_{(n-1)} > 1.5)$. Em ambos os gráficos, temos que $X \sim N(0, 1)$. . . | 23 |
| 2.3 | Acima: estimativas para $P(X_{(n)} - X_{(n-1)} > 1)$; Abaixo: estimativas para $P(X_{(n)}/X_{(n-1)} > 1, 5)$. Em ambos os gráficos, temos que $X \sim Gama(3, 1/2)$ | 26 |
| 2.4 | Acima: estimativas para $P(X_{(n)} - X_{(n-1)} > 1)$; Abaixo: estimativas para $P(X_{(n)}/X_{(n-1)} > 1.5)$. Em ambos os gráficos, temos que $X \sim Cauchy(0, 1)$ | 27 |
| 2.5 | Classes segundo suas propriedades relacionadas a <i>outliers</i> . Em cinza temos as distribuições <i>RRO</i> | 30 |
| 2.6 | Outra possível classificação para as distribuições <i>PO</i> | 36 |
| 3.1 | Razão entre os índices de surpresa 1 de uma verossimilhança $N(\theta, 1)$ e uma $t_{(31, \theta)}$, ambas considerando uma priori $N(0, 1)$ | 41 |
| 3.2 | Priori, Verossimilhança e Posteriori em função de θ com $x = 8$ | 41 |
| 4.1 | Priori, verossimilhança e posteriori para as densidades no Exemplo 4.2, com $x = 3$ e $x = 10$ de cima para baixo. | 72 |
| 4.2 | Razão entre a distribuição normal padrão e a distribuição t-Student com 31 graus de liberdade | 73 |

| | | |
|------|--|----|
| 4.3 | Razão entre a densidade $Beta(2, 3, 2, 10)$ e a densidade Weibull com parâmetro de forma igual a $1/2$ | 73 |
| 4.4 | Esperança a posteriori para diversos valores de x , onde $X \theta \sim Weibull(1/2, \theta)$ e $\Theta \sim Beta(2, 3, 2, 10)$ | 74 |
| 4.5 | Linha sólida: posteriori para o Exemplo 4.3, supondo $x = 20$. Linha pontilhada: densidade a priori | 75 |
| 4.6 | Esperanças de θ a posteriori com o crescimento de x para os modelos I e II respectivamente | 76 |
| 4.7 | Esperança a posteriori para valores de X e Y entre -10 e 10 , supondo que os dados possuem distribuição $N(\theta, 1)$ | 80 |
| 4.8 | Esperança a posteriori para valores de X e Y entre 0 e 50 , supondo que os dados possuem distribuição $t_{(31, \theta)}$ | 80 |
| 4.9 | Esperança de θ a posteriori para x entre 0 e 100 , considerando as distribuições dadas no Exemplo 4.7 | 88 |
| 4.10 | Priori, verossimilhança e posteriori referentes ao Exemplo 4.7 graficadas sobre θ . Acima fixamos $x = 5$ e abaixo fixamos $x = 25$ | 88 |

Capítulo 1

Introdução

1.1 Fontes de informação

Inicialmente, devemos deixar claro o que estamos considerando como sendo uma fonte de informação. Para tanto, seria necessária uma definição formal sobre o que é informação. Contudo, tal definição não existe. O seguinte conceito, apresentado por Basu (1975), parece ser o que melhor descreve o que se entende por informação (ver Pereira [37]).

“Informação é o que ela faz por você, muda a sua opinião.”

O caráter subjetivo deste conceito está no indivíduo que está tentando obter informação: um conjunto de observações poderia não alterar o conhecimento de um indivíduo, mas este mesmo conjunto poderia ser de grande valia para um outro com um paradigma diferente.

A informação que desejamos obter é sobre uma variável aleatória, aqui representada por Θ e descrita por sua função de distribuição. Esta distribuição é baseada em avaliações probabilísticas, levando em conta informações culturais e experimentais.

Quando realizamos um experimento, cujo resultado é x , estamos retirando informação sobre Θ através da função de verossimilhança $f(x|\theta)$. Em conjunto com a informação sobre

Θ que tínhamos anteriormente, extrairemos via teorema de Bayes uma nova informação sobre Θ , representada pela distribuição a posteriori.

O teorema de Bayes é um processo natural de atualização de informação, onde combinamos a nossa informação prévia ao experimento com a informação entregue por este. Diremos portanto que temos duas fontes de informação (sobre Θ): a distribuição a priori e a função de verossimilhança.

1.2 Conflitos de informação

Para introduzirmos o que é um conflito de informação, suponhamos que para um dado problema verificamos uma única observação da variável $X|\theta$ cuja distribuição é $N(\theta, 1)$ com Θ tendo distribuição $N(0, 1)$. Então temos que a posteriori, $\Theta|x$, terá distribuição $N(x/2, 1/2)$. Sendo que observamos x , é razoável supormos que verossimilhança sugere que θ está próximo do valor x observado, uma vez que $E(X|\theta) = \theta$. Por outro lado, também é bastante razoável supor que a priori sugere que θ está próximo de zero, pois $E(\Theta) = 0$. Notemos que $E(\Theta|x)$ sugere que θ está entre as informações dadas pelas duas fontes (exatamente no ponto médio). De fato, dado que $X|\theta$ possui distribuição normal, não esperamos que $|x|$ seja muito elevado e então a posteriori concilia de maneira satisfatória as informações de ambas as fontes. Contudo, se supormos que $|x|$ observado tem um valor muito elevado, temos que as fontes se desacreditarão, pois ambas dizem que θ está em vizinhanças distintas e muito distantes uma da outra. Notemos que, quando escolhemos um modelo normal para os dados, não esperamos observar um valor tão elevado para $|x|$, ou ainda, acreditamos fortemente que o valor $|x|$ estará em torno de sua moda. De maneira análoga, a escolha de uma distribuição $N(0, 1)$ para Θ sugere a nossa forte crença de que Θ está próximo de zero. A posteriori por sua vez não consegue dar maior crédito para uma fonte do que para outra. Como resultado disto temos que a posteriori é fortemente influenciada pelo *outlier*, o que resulta em $E(\theta|x) = x/2 \rightarrow \infty$ quando $x \rightarrow \infty$. Alternativamente poderíamos ter realizado uma outra escolha de distribuições para as fontes. Por exemplo, suponhamos que $X|\theta$ tenha distribuição t-Student com 31 graus de liberdade com parâmetro de locação θ e que Θ tenha distribuição $N(0, 1)$. Neste caso, ainda temos que para um valor atípico de x as fontes de informação sugerem vizinhanças distintas para Θ , mas temos que $E(\theta|x) \rightarrow E(\theta) = 0$ quando $x \rightarrow \infty$. Isto ocorre porque neste caso a posteriori converge para a priori, ou seja, a posteriori rejeita completamente o *outlier*, dando mais crédito à informação a priori (ver abaixo). Dizemos neste caso que o conflito foi resolvido a favor da informação a priori. Com esta motivação,

damos a seguinte definição.

Definição 1.1 (Conflitos de Informação). *Dizemos que existe um conflito entre as fontes de informação se, por algum motivo, as informações entregues pelas fontes se contradizem.*

A princípio, um conflito ocorre quando as fontes de informação se afastam. No exemplo, tínhamos que as modas das fontes se afastavam a medida em que o valor de $|x|$ aumentava. Contudo, ainda não existe uma medida que nos diga o quão afastadas estão as fontes a ponto de entrarem em conflito, o que dificulta a detecção de um conflito. Entretanto, existem alguns índices que podem nos auxiliar nesta questão (ver Seção 3.1).

Embora existam esforços para tentar explicar a ocorrência de conflitos entre as fontes de informação, a maior parte da literatura está voltada em resolver um possível conflito.

Definição 1.2 (Resolução de Conflitos). *Dizemos que um possível conflito entre as fontes de informação está resolvido se, de alguma forma, a posteriori se baseia mais em uma das fontes de informação*

Embora a definição acima seja bastante vaga, ela se enquadra bem nos resultados encontrados na literatura. Em geral, em um estudo de conflitos de informação, procuramos por condições suficiente para que a posteriori se baseie mais na fonte que depositamos maior credibilidade. Contudo, credibilidade é um termo bastante vago e subjetivo. A noção que adotaremos aqui será a seguinte: uma fonte tem maior credibilidade que a outra se ela depositar menos peso em suas caudas. A motivação para tal é que uma distribuição com caudas mais pesadas tende a dar maior peso para eventos mais distantes da moda. Discutiremos mais sobre distribuições de caudas pesadas no Capítulo 2. Esta noção de credibilidade está de acordo com a dada em O'Hagan [36].

Notemos que o conflito no exemplo dado foi causado pela presença de uma observação de valor elevado considerando os modelos propostos para as fontes. Embora existam outros motivos para a ocorrência de conflitos de informação, neste trabalho focaremos apenas os casos em que estes são causados por este tipo de observação, na qual passaremos a considerar como *outliers*. Veremos no Capítulo 2 que as distribuições de caudas pesadas são propensas a gerarem *outliers*.

Existem diversas publicações sobre conflitos de informação. Contudo, é interessante comentar que o conflito de informação muitas vezes está mergulhado em outros contextos, e dentre esses destacamos o tratamento de *outliers* em inferência bayesiana e a robustez bayesiana. A seguir, damos algumas referências que se encaixam no contexto de conflitos de informação.

- (1961) - de Finetti [14] comenta, sem demonstrações, como um *outlier* poderia ser rejeitado de maneira natural no escopo bayesiano. É tido como o predecessor dos estudos sobre a rejeição de *outliers* em inferência bayesiana.
- (1968) - Lindley, em resposta a uma discussão em [29], exemplifica, com uma versão do exemplo acima, como podemos fazer escolhas de modelos de tal forma que uma das fontes seja beneficiada na presença de *outliers* na amostra.
- (1973) - pensando na discussão apresentada em [29], Dawid [13] deu condições suficientes para que um conflito de informações fosse resolvido a favor de uma das fontes de informação.
- (1979) - inspirado nas definições sobre distribuições propensas e resistentes a *outliers* de Neyman e Scott [32], com posterior discussão de Green em [26] e [25], O'Hagan define em [33] distribuições propensas e resistentes a *outliers* em inferência bayesiana. Neste mesmo trabalho, o autor sugere condições suficientes para as condições dadas por Dawid [13].
- (1983) - Goldstein [24] dá condições para que a posteriori seja limitada pela priori no caso de distribuições simétricas e fortemente unimodais, tidas como sendo resistentes a *outliers* no sentido bayesiano.
- (1990) - O'Hagan [36] define uma medida de cauda, *credence*, que auxilia na resolução de conflitos de informação.
- (1993) - Lucas [30] dá condições para que, quando $x \rightarrow \infty$, a posteriori convirja para a distribuição normal, trabalhando com distribuições na família Box e Jenkins.
- (1995) - Pericchi e Sansó [40] apresentam outras condições para resolver conflitos entre as fontes de informação.
- (1999) - Haro-López e Smith [27] dão condições para que um conflito seja resolvido utilizando distribuições v -esféricas.
- (2004) Foster e O'Hagan [35] discutem sobre o que origina um conflito e quando uma posteriori falha em resolvê-lo.
- (2006) Andrade e O'Hagan [2] estudam o caso em que uma das fontes pertence à família escala e tem distribuição pertencente à classe de funções de variação regular.

1.3 Outliers

O *outlier* é um dos temas mais antigos entre os estatísticos. Segundo Anscombe [3], um *outlier* é uma observação com resíduo anormalmente alto. A presença de *outliers* em uma amostra pode influenciar alguns estimadores, como os de mínimos quadrados. Como estes estimadores foram por muito tempo os mais utilizados, diversas técnicas foram criadas para detectar *outliers*, em geral com o intuito de excluí-los. Com o avanço da estatística e da ciência, um foco diferente foi sendo dado ao *outlier*. Em alguns casos, eles são o que existe de mais representativo na amostra, como por exemplo em estudos de catástrofes naturais. A partir daí nascem ramos para lidar com este tipo de observação, como a estatística robusta e a teoria de valores extremos.

Uma maneira comum de se modelar *outliers* é através de mistura de modelos. Duas modelagens são bastante comuns. Na primeira, temos n observações independentes onde $n - k$ são centradas em μ e k são centradas em $\mu + \gamma_i$, $i = 1, \dots, k$. Na segunda, temos n observações independentes com $n - k$ com variância σ^2 e as k observações restantes com variâncias $\lambda_i^2 \sigma^2$, com $i = 1, \dots, k$.

Neyman e Scott [32] levantaram a hipótese de que algumas famílias de distribuições podem produzir *outliers* mais comumente. Posteriormente Green [25] definiu essa propensão natural para distribuições individualmente, denominando-as “distribuições propensas a *outliers*”. Este assunto será discutido posteriormente na Seção 2.2.

Neste trabalho estaremos considerando que o *outlier* é proveniente da estrutura geradora dos dados, independente da distribuição das fontes ser propensa ou não a *outliers*.

É bastante natural pensarmos que um *outlier* pode ser gerado pela verossimilhança se esta tiver uma distribuição propensa a *outliers*. Contudo, no contexto bayesiano, podemos ter que a propensão da distribuição a priori pode justificar um *outlier*. Para entendermos melhor esta noção, considere que $X|\theta \sim N(\theta, a)$, com a pequeno, e que $\Theta \sim \text{Cauchy}(0, 1)$. A verossimilhança nos diz que o valor de x observado deveria estar bastante próximo de θ . Entretanto, a priori esperamos qualquer valor para Θ . Assim, não estaríamos surpresos em observar um possível *outlier*, que neste caso seria visto como uma observação x em torno de um θ elevado em valor absoluto.

1.4 Por que distribuições de cauda pesada?

As distribuições ditas de cauda leve dão maior peso para valores próximos de suas modas, o que sugere maior crédito sobre a informação entregue por elas. Por outro lado, as distribuições de cauda pesada atribuem maior probabilidade aos eventos que ocorrem em suas caudas. Assim, a credibilidade entre as fontes poderia ser representada em termos de peso de caudas: a fonte a qual atribuímos mais crédito deve ter cauda mais leve.

Na literatura pesquisada observamos que, sob condições razoáveis, a distribuição a posteriori, em caso de conflito entre as fontes, tende a dar maior credibilidade à fonte de menor peso nas caudas. Assim, se acreditamos que a informação proveniente dos dados pode de alguma forma se distanciar da informação a priori, representamos essa opinião escolhendo uma distribuição para a verossimilhança com caudas mais pesadas do que a escolhida para a priori. Neste caso, gostaríamos de dar restrições razoáveis para que, em caso de conflito, a informação dos dados fosse rejeitada a favor da informação a priori, ou ainda, que as inferências a posteriori fossem baseadas na priori.

Por outro lado, se acreditamos que a informação a priori pode de alguma forma se distanciar dos dados, refletimos isto escolhendo para a priori uma distribuição com caudas mais pesadas do que as caudas da distribuição dos dados. Neste caso, gostaríamos que um eventual conflito entre as fontes de informação fosse resolvido a favor dos dados e, como consequência, que inferências a posteriori fossem mais influenciadas pela verossimilhança. Esta situação seria interessante, por exemplo, em situações onde o pesquisador tem uma vaga noção sobre Θ , mas não se surpreenderia caso os dados lhe revelassem uma informação diferente.

1.5 Organização

A utilização de distribuições de cauda pesada para a resolução de conflitos de informação tem aparecido esporadicamente na literatura, de modo direto ou indireto. Com o objetivo de discutir e apresentar tais utilizações da literatura, bem como apresentar nossos resultados para o mesmo fim, organizamos este trabalho em três grandes partes.

Na primeira parte, que compreende o Capítulo 2, procuramos mostrar quais as classes de distribuições de cauda pesada mais utilizadas na literatura em geral. Isto se torna importante, uma vez que existem diversas definições para o que seria o termo “distribuições

de cauda pesada”. Para tais classes, apresentamos algumas propriedades. As seguintes classes foram estudadas: cauda longa, variação dominada, subexponencial e variação regular. Neste capítulo também apresentamos as distribuições propensas e resistentes a *outliers* apresentadas por Green [25]. Aqui apresentamos nosso primeiro resultado deste trabalho, mostrando que é possível enfraquecer os teoremas de [25] e que por consequência mostrando que as distribuições de cauda pesada apresentadas são subclasses destas distribuições.

Na segunda parte, no Capítulo 3, dedicamo-nos ao estudo de conflitos de informação. Começamos com uma breve discussão na Seção 3.1 sobre a terminologia aplicada neste trabalho, bem como uma melhor caracterização sobre conflitos de informação. No restante deste capítulo apresentamos alguns resultados bastante citados na literatura que estão relacionados com distribuições de cauda pesada. Na Seção 3.2, apresentamos o teorema de Dawid, que dá condições suficientes para que um possível conflito de informação seja resolvido em favor de uma das fontes de informação. Na Seção 3.3 apresentamos o teorema de Pericchi e Sansó, que dá outras condições suficientes para a resolução de conflitos. Na Seção 3.4 apresentamos as noções de *credence* de densidade, uma medida útil para resolver conflitos de informação. Na Seção 3.5 discutimos como utilizar distribuições de variação regular para resolver conflitos.

Na terceira parte, no Capítulo 4, apresentamos nossas considerações sobre a relação entre distribuições de cauda pesada e conflitos de informação. Na Seção 4.1, discutimos o teorema de Dawid apresentado anteriormente, dando ênfase à sua relação com as distribuições de cauda longa. Na Seção 4.2 discutimos a relação entre o teorema de Pericchi e Sansó e as distribuições propensas a *outliers*, e a relação deste resultado com o obtido por Dawid. Na Seção 4.3 apresentamos nosso segundo resultado, onde conjuntamente com a suposição de que uma das fontes pertence à classe SD (uma subclasse da classe subexponencial), damos condições suficientes para um possível conflito de informação seja resolvido. Na Seção 4.4 caracterizamos a medida *credence*, mostrando que se uma distribuição possui *credence*, então deve pertencer à classe de variação regular estendida. Por último, na Seção 4.5 apresentamos nosso último resultado, o Teorema 4.10, para resolver conflitos de informação utilizando *credence* no caso em que uma das fontes possui distribuição pertencente à família escala.

Capítulo 2

Distribuições de cauda pesada

O termo distribuição de cauda pesada não tem definição universalmente aceita, ficando muitas vezes a critério da modelagem. Por exemplo alguns definem uma distribuição como sendo de cauda pesada se o decaimento de suas caudas for mais lento que um decaimento exponencial, enquanto outros exigem que certos limites ou propriedades assintóticas estejam satisfeitos.

Neste capítulo revisamos algumas classes de distribuição que têm sido amplamente utilizadas como distribuições de cauda pesada em diversos ramos da estatística. Na Seção 2.1 discutimos as principais classes utilizadas neste trabalho. Na Seção 2.1.5 mostramos a relação entre tais classes. Na Seção 2.1.6 apresentamos uma importante subclasse da classe subexponencial, voltada para densidades.

Para complementar a discussão sobre as distribuições de cauda pesada, apresentamos na Seção 2.2 as distribuições propensas e resistentes a *outliers*, apresentadas por Neyman e Scott [32] e discutidas por Green [26],[25]. As distribuições propensas a *outliers* apresentam valores observados elevados como uma característica natural. Para perceber essa propensão natural para gerar *outliers*, o leitor pode simular algumas amostras de uma distribuição Cauchy e comparar os máximos e mínimos com os quantis amostrais.

Uma vez que gostaríamos de caracterizar as distribuições de caudas pesadas apresentadas como propensas a gerar *outliers*, dedicamos parte deste trabalho na elaboração de proposições que mostrassem esta relação. O resultado deste estudo é mostrado na Seção 2.3, onde enfraquecemos os Teoremas 2.28 e 2.29 sobre distribuições propensas a *outliers*

e mostramos ao longo da referida seção que as distribuições de cauda pesada apresentadas na Seção 2.1 são um subconjunto destas distribuições.

2.1 Classes de distribuições de cauda pesada

Nesta seção apresentamos as principais classes de distribuições que são consideradas na literatura como sendo de cauda pesada. São estas: a classe de cauda longa, a classe subexponencial, a classe de variação regular e a classe de variação dominada. As três primeiras classes citadas também são conhecidas como distribuições de cauda longa, uma vez que a classe de cauda longa contém ambas (ver Seção 2.1.5). A classe com a maior aplicabilidade na literatura é a classe subexponencial, e dentre seus membros destacamos as distribuições t-Student, Pareto, Weibul, lognormal, α -estáveis truncadas, loggama, Burr e Benktander tipos I e II.

Ao longo desta seção, temos que F é função de distribuição com $F(x) < 1 \forall x$ finito e $F(\infty) = 1$. Consideramos também $\bar{F}(x) = 1 - F(x)$. Convém notarmos que todos os resultados estão definidos apenas para a cauda direita. Isto porquê podemos ter distribuições com diferentes comportamentos em cada cauda. Contudo, todas as definições são válidas para ambas as caudas.

2.1.1 A classe de cauda longa

A primeira classe de distribuições de cauda pesada que apresentaremos é definida a seguir.

Definição 2.1. Dizemos que F é de cauda longa se

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(x - y)}{\bar{F}(x)} = 1, \quad \forall y \in \mathbb{R}. \quad (2.1)$$

Denotamos por \mathcal{L} a família de distribuições que verificam (2.1).

Alguns autores, como Teugels [42], denotam distribuições em \mathcal{L} como *distribuições de variação lenta quando $x \rightarrow \infty$* . Outros, como Baltrunas, Omey e Van Gulck [4], consideram \mathcal{L} como uma classe de funções mensuráveis que satisfazem (2.1). Esta classe é a maior das classes de interesse que serão discutidas nesta seção. Uma vez que, como

discutido anteriormente, o termo “distribuições de cauda pesada” não tem uma definição universalmente aceita, o termo distribuições de cauda longa pode vir a ser utilizado para este fim. De fato, muitas das distribuições que são consideradas de cauda pesada pertencem a esta classe. Veremos adiante, no Capítulo 4, que diversos resultados sobre conflitos de informação estão relacionados com esta classe.

2.1.2 A classe de variação dominada

A próxima classe de distribuições em estudo é definida a seguir.

Definição 2.2. Dizemos que F tem variação dominada se

$$\limsup_{x \rightarrow \infty} \frac{\bar{F}(x/2)}{\bar{F}(x)} < \infty. \quad (2.2)$$

A classe \mathcal{DV} destas distribuições é conhecida como classe de distribuições de variação dominada.

Esta classe bastante conhecida apresenta subclasses interessantes para o estudo sobre conflitos de informação. A mais importante delas é a de variação regular, a ser discutida na Seção 2.1.4. As subclasses mais conhecidas da classe \mathcal{DV} , bem como todas as relações entre classes serão discutidas na Seção 2.1.5.

A seguinte proposição descreve um limitante para o limite superior da razão na Definição 2.2.

Proposição 2.3. Se $F \in \mathcal{DV}$ então existem constantes A e p tais que para todo $x > x_0$ e para todo $\theta > 1$,

$$\frac{\bar{F}(x/\theta)}{\bar{F}(x)} \leq A \left(\frac{1}{\theta}\right)^{-p}. \quad (2.3)$$

Demonstração. Sabemos que, se $F \in \mathcal{DV}$ então existem $B \geq 1$ e x_0 tais que, para todo $x > x_0$, $\bar{F}(x/2)/\bar{F}(x) \leq B$. Tomando $p = \log_2 B$, podemos reescrever $\bar{F}(x/2)/\bar{F}(x) \leq 2^p$. Notemos que

$$\frac{\bar{F}(x/4)}{\bar{F}(x)} = \frac{\bar{F}(x/4)}{\bar{F}(x/2)} \frac{\bar{F}(x/2)}{\bar{F}(x)} \leq \left(\frac{1}{2}\right)^{-2p}$$

e por indução,

$$\frac{\bar{F}(2^{-n}x)}{\bar{F}(x)} \leq \left[\left(\frac{1}{2} \right)^n \right]^{-p}.$$

Seja $\theta > 1$ e n_1 o menor inteiro tal que $\theta \leq 2^{n_1}$. Seja $A = (\theta/2^{n_1})^{-p}$. Então

$$\frac{\bar{F}(x/\theta)}{\bar{F}(x)} \leq \frac{\bar{F}(2^{-n_1}x)}{\bar{F}(x)} = \frac{\bar{F}(A^{-1/p}x/\theta)}{\bar{F}(x)} \leq A \left(\frac{1}{\theta} \right)^{-p}.$$

□

Uma análise mais detalhada sobre esta classe pode ser encontrada em Feller [19] e [20].

2.1.3 A classe subexponencial

Dadas duas funções $H_1(x)$, $H_2(x)$, definimos a relação $H_1(x) \sim H_2(x)$ como

$$\lim_{x \rightarrow \infty} \frac{H_1(x)}{H_2(x)} = 1.$$

Observação: notemos que também estamos utilizando “ \sim ” para denotar a distribuição de uma variável aleatória, acreditando que o leitor não terá dificuldades em identificar os dois usos.

Sejam X, X_1, X_2, \dots variáveis aleatórias independentes e identicamente distribuídas. Suponhamos que

$$P(S_n > x) \sim nP(X > x) \sim P(M_n > x) \quad (2.4)$$

onde $S_n = X_1 + \dots + X_n$ e $M_n = \max_{i=1, \dots, n} X_i$. Uma interpretação intuitiva para (2.4) é que o máximo M_n de X_1, \dots, X_n dá a maior contribuição para a soma destas. Assim, o excesso da soma S_n acima de um alto limiar estabelecido é devido ao maior valor desta amostra. Esta interpretação sugere uma maneira de definir distribuições de cauda pesada: a cauda da soma é essencialmente determinada pela cauda do máximo. Esta abordagem intuitiva nos conduz a uma classe rica de distribuições.

Definição 2.4. *Sejam X_1, X_2 variáveis aleatórias independentes e identicamente distribuídas. Seja $F^{(2)} = F * F$, isto é, a convolução de F consigo mesma. Dizemos que $F \in \mathcal{S}$ se*

$$\lim_{x \rightarrow \infty} \frac{\bar{F}^{(2)}(x)}{\bar{F}(x)} = 2. \quad (2.5)$$

A classe \mathcal{S} é denominada classe subexponencial.

A classe \mathcal{S} foi introduzida por Chistyakov [11] e Chover, Ney e Wainger [12] no contexto de processos de ramificação. Aplicações podem ser encontradas em teoria das filas, passeio aleatório, teoria de risco, matemática financeira e teoria de probabilidade aplicada. O nome subexponencial se refere ao decaimento das caudas, que é sempre mais lento que um decaimento exponencial (a prova deste fato pode ser encontrada em Teugels [42]).

Antes de continuarmos a discussão de resultados relacionados à classe \mathcal{S} , consideremos a seguinte definição.

Definição 2.5. *Dizemos que F e G são fracamente assintoticamente equivalentes, denotado por $F \approx^w G$, se existirem m e M pertencentes à $(0, \infty)$ tais que $m \leq \bar{G}(x)/\bar{F}(x) \leq M$ para todo $x \in (0, \infty)$.*

Esta definição nos ajuda a classificar a relação entre as caudas de duas distribuições. Com esta definição enunciamos um resultado interessante que utilizaremos na Seção 4.4.

Teorema 2.6. *Suponhamos que $F \approx^w G$ e que $G \in \mathcal{L}$. Então $F \in \mathcal{S} \iff G \in \mathcal{S}$.*

Demonstração. Sejam $m, M \in (0, \infty)$ tais que $m \leq \bar{G}(x)/\bar{F}(x) \leq M \forall x \in (0, \infty)$. Suponhamos que $F \in \mathcal{S}$. Para $v > 0$ fixado e para $x > 2v$ temos

$$\begin{aligned} P(X_1 + X_2 > x) &= P(X_1 + X_2 > x, X_2 \leq v) + P(X_1 + X_2 > x, v < X_1 < x - v) \\ &\quad + P(X_1 + X_2 > x, X_1 \leq v) + P(X_1 > x - v, X_2 > v) \\ &= 2 \int_0^v \bar{G}(x - y) dG(y) + \int_v^{x-v} \bar{G}(x - y) dG(y) + \bar{G}(x - v) \bar{G}(v) \end{aligned}$$

portanto

$$\frac{\bar{G}^{(2)}(x)}{\bar{G}(x)} = 2 \int_0^v \frac{\bar{G}(x - y)}{\bar{G}(x)} dG(y) + \int_v^{x-v} \frac{\bar{G}(x - y)}{\bar{G}(x)} dG(y) + \frac{\bar{G}(x - v)}{\bar{G}(x)} \bar{G}(v). \quad (2.6)$$

Como a integral de Stieljes é linear no integrador, temos que

$$\int_v^{x-v} \frac{\bar{G}(x-y)}{\bar{G}(x)} dG(y) \leq \frac{M}{m} \int_v^{x-v} \frac{\bar{F}(x-y)}{\bar{F}(x)} dG(y) \leq \frac{M^2}{m} \int_v^{x-v} \frac{\bar{F}(x-y)}{\bar{F}(x)} dF(y)$$

e como $F \in \mathcal{S}$ temos que

$$\lim_{v \rightarrow \infty} \limsup_{x \rightarrow \infty} \int_v^{x-v} \frac{\bar{F}(x-y)}{\bar{F}(x)} dF(y) = 0.$$

Notemos agora que

$$\bar{G}(x-y) \leq \bar{G}(x-v)$$

para todo $y \in (0, v)$. Assim, pela suposição de que $G \in \mathcal{L}$ temos que

$$\int_0^v \frac{\bar{G}(x-y)}{\bar{G}(x)} dG(y) \rightarrow 1$$

por convergência dominada e que

$$\frac{\bar{G}(x-v)}{\bar{G}(x)} \bar{G}(v) \rightarrow 0$$

quando $x \rightarrow \infty$. Portanto (2.6) converge para 2, o que implica que $G \in \mathcal{S}$. \square

Com o Teorema 2.6 nos mostra que se $G \in \mathcal{S}$ e se $F \approx^w G$, então F também pertence a \mathcal{S} , uma vez que $\mathcal{S} \subset \mathcal{L}$ (ver Seção 2.1.5). Para utilização posterior na Seção 4.4, elaboramos a seguinte proposição.

Proposição 2.7. *Suponha que $F(x) < 1$ e $G(x) < 1$, para todo $x \in (0, \infty)$ tenham densidades f e g respectivamente. Suponha ainda que $g \in \mathcal{L}$. Uma condição suficiente para que $F \approx^w G$ é que existam constantes m e M pertencentes a $(0, \infty)$ tais que $m \leq f(x)/g(x) \leq M \forall x > 0$ e $g(x) \neq 0$. Neste caso diremos que $f \approx^w g$.*

Demonstração. Basta notarmos que para todo x

$$\begin{aligned} m \leq \frac{f(x)}{g(x)} \leq M &\Rightarrow mg(x) \leq f(x) \leq Mg(x) \\ &\Rightarrow m \int_{-\infty}^x g(t) dt \leq \int_{-\infty}^x f(t) dt \leq M \int_{-\infty}^x g(t) dt \Rightarrow m \leq \frac{\bar{F}(x)}{\bar{G}(x)} \leq M. \end{aligned}$$

\square

Outros resultados e maiores detalhes sobre a classe \mathcal{S} podem ser encontrados em Embrechts, Goldie e Veraverbeke [16], Kluppelberg [28], Embrechts e Omey [18], Murphree [31] e Goldie [23].

2.1.4 A classe de variação regular

A seguir, enunciaremos e discutiremos alguns resultados importantes das funções de variação regular. O primeiro teorema, dado em Feller [20] é enunciado a seguir.

Teorema 2.8. *Seja U uma função monótona positiva em $(0, \infty)$ tal que*

$$\lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = q(x) \quad (2.7)$$

em um conjunto denso de pontos. Então

$$q(x) = x^p \quad (2.8)$$

onde $-\infty \leq p \leq \infty$.

Observação: o sentido de x^∞ ou $x^{-\infty}$ é apenas a extensão do teorema para toda p .

Demonstração. Notemos que

$$\frac{U(x_1 x_2 t)}{U(t)} = \frac{U(x_1 x_2 t)}{U(tx_2)} \frac{U(tx_2)}{U(t)} \quad (2.9)$$

nos mostra que se existe um limite finito e positivo em (2.7) para $x = x_1$ e para $x = x_2$, então também existe um para $x = x_1 x_2$ e

$$q(x_1 x_2) = q(x_1) q(x_2) \quad (2.10)$$

Notemos que, por indução $q(x_1^n) = [q(x_1)]^n$. Assim, se $q(x_1) = \infty$, temos que $q(x_1^n) = \infty$ e $q(x_1^{-n}) = 0$. Basta mostrar então o teorema para um q finito. Pela monotonicidade assumida, podemos definir q em toda a parte por continuidade à direita, e por isso, (2.10) vale para todo x_1 e x_2 . Agora, fazendo $x = e^\varepsilon$ e $q(e^\varepsilon) = u(\varepsilon)$ temos que (2.10) é equivalente a $u(\varepsilon_1 + \varepsilon_2) = u(\varepsilon_1)u(\varepsilon_2)$. É fato bem conhecido que todas as soluções desta equação estão limitadas em intervalos finitos e são da forma $u(\varepsilon) = e^{p\varepsilon}$. Mas isto é o mesmo que $q(x) = x^p$, concluindo a prova. \square

Funções cujo comportamento são como no Teorema 2.8 são denominadas funções de variação regular. Uma vez que U é função monótona não crescente podemos definir diretamente distribuições de variação regular. Contudo, funções de variação regular podem ser definidas em termos mais gerais, como pode ser visto na definição a seguir.

Definição 2.9. Dizemos que uma função mensurável f é de variação regular ao infinito com índice (ordem) ρ ($f \in \mathcal{RV}_\rho$), se $\rho \in \mathbb{R}$ e

$$\lim_{x \rightarrow \infty} \frac{f(\lambda x)}{f(x)} = \lambda^\rho \quad (2.11)$$

para todo $\lambda > 0$.

As densidades de variação regular foram utilizadas por Andrade e O'Hagan [2] para resolver conflitos de informação quando o parâmetro em questão é de escala. Estes resultados serão discutidos na Seção 3.5.

Os dois resultados abaixo serão utilizados posteriormente na Seção 3.5.

Lema 2.10. Se $f \in \mathcal{RV}_p$ então para quaisquer $A > 1$ e $\delta > 0$ existe $X = X(A, \delta)$ tal que

$$\frac{f(w)}{f(z)} \leq \max \left\{ \left(\frac{w}{z} \right)^{p+\delta}, \left(\frac{w}{z} \right)^{p-\delta} \right\}, \quad \forall (w \geq X, z \geq X).$$

Proposição 2.11. Suponhamos que $f \in \mathcal{RV}_p$ e que $g \in \mathcal{RV}_q$. Então $h \in \mathcal{RV}_{p+q}$ onde $h = fg$.

Demonstração. Temos que

$$\frac{h(x\lambda)}{h(x)} = \frac{f(x\lambda)g(x\lambda)}{f(x)g(x)} \rightarrow \lambda^{p+q}$$

quando $x \rightarrow \infty$. □

A seguir, definimos uma classe ligeiramente maior que \mathcal{RV} .

Definição 2.12. Dizemos que uma distribuição é de variação regular estendida ($f \in ERV(-\alpha, -\beta)$) se para constantes $0 \leq \alpha \leq \beta < \infty$ temos que

$$y^{-\beta} \leq \liminf_{x \rightarrow \infty} \frac{\bar{F}(xy)}{\bar{F}(x)} \leq \limsup_{x \rightarrow \infty} \frac{\bar{F}(xy)}{\bar{F}(x)} \leq y^{-\alpha}$$

para qualquer $y > 1$.

A classe ERV será utilizada nas Seções 4.4 e 4.5.

2.1.5 Relações entre as classes de distribuições de cauda pesada

Discutiremos a seguir as relações existentes entre as classes de distribuições apresentadas anteriormente. Estas relações tornam-se importantes para uma melhor compreensão da abrangência de certos resultados que serão discutidos nos capítulos posteriores.

Proposição 2.13. *Se $F \in \mathcal{S}$ então $F \in \mathcal{L}$.*

Demonstração. Sejam X_1 e X_2 variáveis aleatórias positivas independentes com distribuição $F \in \mathcal{S}$. Para $x \geq y > 0$, temos que

$$\begin{aligned}
 \frac{\bar{F}^{(2)}(x)}{\bar{F}(x)} &= \frac{1 - P(X_1 + X_2 \leq x)}{\bar{F}(x)} = \frac{1 - \int_0^x P(X_1 \leq x - t | X_2 = t) dF(t)}{\bar{F}(x)} \\
 &= \frac{1 - \int_0^x F(x - t) dF(t)}{\bar{F}(x)} = \frac{\int_0^\infty dF(t) - \int_0^x F(x - t) dF(t)}{\bar{F}(x)} \\
 &= \frac{\int_x^\infty dF(t) + \int_0^x [1 - F(x - t)] dF(t)}{\bar{F}(x)} = 1 + \frac{\int_0^x \bar{F}(x - t) dF(t)}{\bar{F}(x)} \\
 &= 1 + \int_0^y \frac{\bar{F}(x - t)}{\bar{F}(x)} dF(t) + \int_y^x \frac{\bar{F}(x - t)}{\bar{F}(x)} dF(t) \\
 &\geq 1 + \int_0^y \frac{\bar{F}(x)}{\bar{F}(x)} dF(t) + \int_y^x \frac{\bar{F}(x - y)}{\bar{F}(x)} dF(t) \\
 &= 1 + F(y) + \frac{\bar{F}(x - y)}{\bar{F}(x)} [F(x) - F(y)]
 \end{aligned}$$

assim, para x grande o suficiente tal que $F(x) - F(y) \neq 0$, temos que

$$1 \leq \frac{\bar{F}(x - y)}{\bar{F}(x)} \leq \left(\frac{\bar{F}^{(2)}(x)}{\bar{F}(x)} - 1 - F(y) \right) \left(\frac{1}{F(x) - F(y)} \right) \rightarrow 1$$

quando $x \rightarrow \infty$, portanto $\mathcal{S} \subset \mathcal{L}$. □

Proposição 2.14. *Se $F \in \mathcal{L} \cap \mathcal{DV}$, então $F \in \mathcal{S}$.*

Demonstração. Sejam X_1, X_2 variáveis aleatórias positivas independentes com distribui-

ção F . Então, para $x > 0$,

$$\begin{aligned}\bar{F}^{(2)}(x) &= P(X_1 + X_2 > x) = P\left(X_1 + X_2 > x, X_1 \leq \frac{1}{2}x\right) \\ &+ P\left(X_1 + X_2 > x, X_2 \leq \frac{1}{2}x\right) + P\left(X_1 > \frac{1}{2}x, X_2 > \frac{1}{2}x\right) \\ &= 2 \int_0^{\frac{x}{2}} \bar{F}(x-y) dF(y) + [\bar{F}(x/2)]^2\end{aligned}$$

assim,

$$\frac{\bar{F}^{(2)}(x)}{\bar{F}(x)} = 2 \int_0^{\frac{1}{2}x} \frac{\bar{F}(x-y)}{\bar{F}(x)} dF(y) + \frac{[\bar{F}(x/2)]^2}{\bar{F}(x)}.$$

Como $F \in \mathcal{DV}$, temos que

$$\int_0^{\frac{1}{2}x} \frac{\bar{F}(x-y)}{\bar{F}(x)} dF(y) \leq \frac{\bar{F}(x/2)}{\bar{F}(x)} \int_0^{\frac{1}{2}x} dF(y) < \infty.$$

Logo, por convergência dominada, temos que

$$\begin{aligned}\lim_{x \rightarrow \infty} \frac{\bar{F}^{(2)}(x)}{\bar{F}(x)} &= 2 \lim_{x \rightarrow \infty} \int_0^{\frac{1}{2}x} \frac{\bar{F}(x-y)}{\bar{F}(x)} dF(y) + \lim_{x \rightarrow \infty} \frac{[\bar{F}(x/2)]^2}{\bar{F}(x)} \\ &= 2 \int_0^{\infty} \lim_{x \rightarrow \infty} \frac{\bar{F}(x-y)}{\bar{F}(x)} dF(y) = 2\end{aligned}$$

portanto temos que $\mathcal{L} \cap \mathcal{DV} \subset \mathcal{S}$. □

Proposição 2.15. *Se $F \in \mathcal{RV}_\rho$, com $\rho > 0$ e $x > 0$, então $F \in \mathcal{S}$.*

Demonstração. Sejam X_1, X_2 variáveis aleatórias independentes com distribuição $F \in \mathcal{RV}_{-\rho}$. Notemos que de $\{X_1 + X_2 > x\} \supset \{X_1 > x\} \cup \{X_2 > x\}$ temos que

$$P(X_1 + X_2 > x) \geq P(X_1 > x) + P(X_2 > x) - P(X_1 > x)P(X_2 > x) = 2\bar{F}(x) - \bar{F}(x)^2$$

Agora, tome $0 < \delta < 1/2$. Então de

$$\{X_1 + X_2 > x\} \subset \{X_1 > (1-\delta)x\} \cup \{X_2 > (1-\delta)x\} \cup \{X_1 > \delta x, X_2 > \delta x\}$$

segue que

$$\begin{aligned}P(X_1 + X_2 > x) &\leq P(X_1 > (1-\delta)x) + P(X_2 > (1-\delta)x) + P(X_1 > \delta x)P(X_2 > \delta x) \\ &= 2\bar{F}((1-\delta)x) + \bar{F}(\delta x)^2\end{aligned}$$

então temos que

$$2 - \frac{\bar{F}(x)}{\bar{F}(x)} \leq \frac{\bar{F}^{(2)}(x)}{\bar{F}(x)} \leq 2 \frac{\bar{F}((1-\delta)x)}{\bar{F}(x)} + \frac{\bar{F}(\delta x)}{\bar{F}(x)} \bar{F}(\delta x)$$

e que

$$\lim_{\delta \downarrow 0} \lim_{x \rightarrow \infty} \frac{\bar{F}^{(2)}(x)}{\bar{F}(x)} = 2$$

portanto $\mathcal{RV} \subset \mathcal{S}$. □

Proposição 2.16. *Se $F \in \mathcal{RV}_{-\rho}$, com $\rho > 0$ e $x > 0$, então $F \in \mathcal{DV}$.*

Demonstração. Basta notar que

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(xt)}{\bar{F}(x)} = t^{-\rho} \Rightarrow \limsup_{x \rightarrow \infty} \frac{\bar{F}(xt)}{\bar{F}(x)} = t^{-\rho} < \infty.$$

□

Proposição 2.17. *$\mathcal{ERV} \subset \mathcal{DV}$.*

Demonstração. Segue imediatamente da Proposição 2.3. □

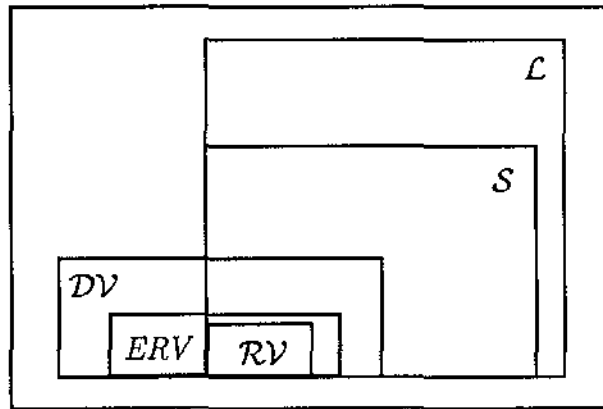


Figura 2.1: Diagrama das relações entre as classes apresentadas

As relações apresentadas nesta seção estão sumarizadas na Figura 2.1. A seguir, damos alguns exemplos de distribuições pertencentes às classes apresentadas.

Exemplo 2.1. São exemplos de distribuições em \mathcal{RV} :

- Pareto: $f(x) = \beta\alpha^\beta x^{-\beta-1}$ com $\alpha, \beta > 0$
- Cauchy: $f(x) = [\pi(1+x^2)]^{-1}$
- Burr: $\bar{F}(x) = \kappa^\alpha(\kappa+x^\tau)^{-\alpha}$ com $\alpha, \kappa, \tau > 0$
- α -Estável Truncada: $\bar{F}(x) = P(|X| > x)$ onde X é α -Estável com $1 < \alpha < 2$
- Loggama: $f(x) \propto (\log x)^{\beta-1} x^{-\alpha-1}$ com $\alpha, \beta > 0$

Exemplo 2.2 (Peter e Paul). A seguinte distribuição pertence à \mathcal{DV} mas não pertence à \mathcal{S}

$$P(X = 2^k) = 2^{-k} \text{ para } k = 1, 2, \dots$$

Exemplo 2.3. A seguinte distribuição pertence à $\mathcal{L} \cap \mathcal{DV}$ mas não pertence à \mathcal{RV}

$$\bar{F}(x) \sim x^{-1}(1 + a \sin(2\pi \log x))$$

para a pequeno.

Exemplo 2.4. Além das distribuições nos Exemplos 2.1 e 2.3 temos que as seguintes distribuições pertencem à \mathcal{S}

- Weibull: $\bar{F}(x) = \exp\{-cx^\tau\}$ com $c > 0$ e $0 < \tau < 1$
- Lognormal: $f(x) \propto x^{-1} \exp\{-(\log x - \mu)^2/2\sigma^2\}$ com $\sigma > 0$ e $\mu \in \mathbb{R}$
- Benktander Tipo I: $(1 + 2(\beta/\alpha) \log x) \exp\{-\beta(\log x)^2 - (\alpha + 1) \log x\}$ com $\alpha, \beta > 0$
- Benktander Tipo II: $\exp\{\alpha/\beta\} x^{-(1-\beta)} \exp\{-\alpha x^\beta/\beta\}$ com $\alpha > 0$ e $0 < \beta < 1$

Exemplo 2.5. Para exemplos de distribuições que pertencem à \mathcal{L} mas não pertence à \mathcal{S} , citamos Embrechts e Goldie [15] e Pitman [38].

2.1.6 A classe de densidades \mathcal{SD}

Uma das dificuldades que encontramos para relacionar conflitos de informação com distribuições de caudas pesadas é a falta de definições para densidades de caudas pesadas. Em algumas situações isto é facilmente remediado impondo condições suficientes em uma densidade que impliquem que a referida distribuição pertença a classe em questão. Em outras situações, as definições para qualquer função mensurável já existem, como nas classes \mathcal{L} e \mathcal{RV} .

Discutimos anteriormente que a classe de distribuições subexponenciais possui muitos membros famosos na literatura, conhecidos por serem de cauda pesada. Por este motivo, gostaríamos de definir densidades em \mathcal{S} . Para tanto, definimos primeiramente a taxa de risco q como sendo $q(x) = f(x)/\bar{F}(x)$. Se supormos que F é absolutamente contínua com densidade f , podemos utilizar a taxa de risco para dar o seguinte teorema de caracterização.

Teorema 2.18. *Suponha que F é absolutamente contínua com densidade f e que a taxa de risco $q(x)$ certamente decresce para 0. Então*

1. $F \in \mathcal{S}$ se e somente se

$$\lim_{x \rightarrow \infty} \int_0^x e^{yq(x)} f(y) dy = 1. \quad (2.12)$$

2. Se a função $x \mapsto \exp\{xq(x)\}f(x)$ for integrável em $[0, \infty)$ então $F \in \mathcal{S}$.

Demonstração. Ver Embrechts, Kluppelberg e Mikosh [17]. □

O Teorema 2.18 caracteriza densidades em \mathcal{S} . Entretanto, gostaríamos de lidar apenas com as densidades de \mathcal{S} que preservam as mesmas propriedades de suas caudas. Esta subclasse de densidades de \mathcal{S} é definida a seguir.

Definição 2.19. *Dizemos que $f \in \mathcal{SD}$ se as seguintes condições estão satisfeitas:*

- $f \in \mathcal{L}$.
- $\frac{f^{(2)}(x)}{f(x)} \rightarrow 2$ se $x \rightarrow \infty$ onde $f^{(2)}$.

Lema 2.20. *Se $f \in \mathcal{S}$ então $F \in \mathcal{S}$ e $f^{(n)}(x) \sim n f(x)$.*

Temos então que as densidades em \mathcal{SD} preservam as características de suas caudas. Além disso essa classe é bastante rica, tendo todas as densidades dos Exemplos 2.1 e 2.4 como membros. Mais detalhes sobre classe \mathcal{SD} e demais referências podem ser encontradas em Baltrunas, Omey e Van Gulck [4].

2.2 Distribuições resistentes e propensas a outliers

Neyman e Scott [32] levanta a hipótese de que certas famílias de distribuições poderiam produzir *outliers* mais comumente. Tais famílias foram denominadas propensas a *outliers*. Posteriormente Green [25] mostrou que as definições dadas em [32] não se aplicavam à distribuições individualmente. Neste mesmo trabalho ele deu definições de distribuições propensas e distribuições resistentes a *outliers* individualmente. Nesta seção apresentaremos e discutiremos estas definições e seus subsequentes resultados.

Consideraremos ao longo desta seção que são satisfeitas as seguintes suposições:

1. $F(\infty) = 1$
2. $F(x) < 1$ para todo x finito

Dada uma amostra X_1, \dots, X_n de variáveis aleatórias independentes e identicamente distribuídas, as estatísticas de ordem serão denotadas por $X_{(1)}, \dots, X_{(n)}$.

2.2.1 Distribuições resistentes a outliers

Iniciaremos nossa discussão pelas distribuições resistentes a *outliers* (*RO*).

Definição 2.21. *Uma distribuição F será chamada de absolutamente resistente a outliers (ARO) se, para todo $\varepsilon > 0$, tivermos*

$$P(X_{(n)} - X_{(n-1)} > \varepsilon) \rightarrow 0 \text{ quando } n \rightarrow \infty. \quad (2.13)$$

Definição 2.22. *Uma distribuição F será chamada de relativamente resistente a outliers (RRO) se, para todo $k > 1$, tivermos*

$$P(X_{(n)}/X_{(n-1)} > k) \rightarrow 0 \text{ quando } n \rightarrow \infty. \quad (2.14)$$

As definições acima nos dizem que para uma amostra suficientemente grande, esperamos que as observações mais elevadas provenientes de distribuições RO estejam cada vez mais próximas entre si a medida que aumentamos o tamanho da amostra e portanto, intuitivamente, não esperamos *outliers*.

Exemplo 2.6. A distribuição normal é tanto absoluta quanto relativamente resistente a *outliers*. A Figura 2.2 mostra valores estimados através da função de distribuição empírica, para as probabilidades (2.13) e (2.14) considerando a variável X com distribuição normal padrão, $\varepsilon = 1$ e $k = 1,5$ para diversos tamanhos de amostra.

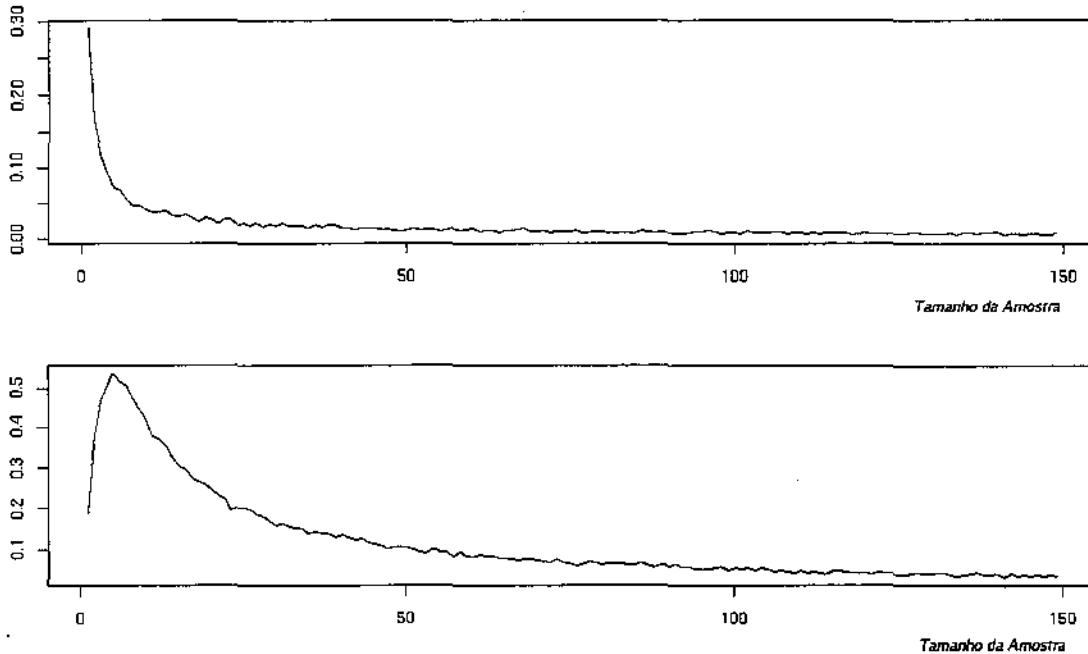


Figura 2.2: Acima: estimativas para $P(X_{(n)} - X_{(n-1)} > 1)$; Abaixo: estimativas para $P(X_{(n)}/X_{(n-1)} > 1.5)$. Em ambos os gráficos, temos que $X \sim N(0, 1)$

□

Teorema 2.23. Sob as suposições 1 e 2, a distribuição F será ARO se e somente se para todo $\varepsilon > 0$,

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(x + \varepsilon)}{\bar{F}(x)} = 0. \quad (2.15)$$

Teorema 2.24. *Sob as suposições 1 e 2, a distribuição F será RRO se e somente se para todo $k > 1$,*

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(kx)}{\bar{F}(x)} = 0. \quad (2.16)$$

Demonstração. ver Green [25] para discussão e construção das demonstrações. □

Os dois teoremas apresentados acima tornam mais fácil verificar se uma distribuição é resistente a *outliers* uma vez que não precisamos encontrar a distribuição de $X_{(n)} - X_{(n-1)}$. Contudo, em muitos casos a função de densidade é mais utilizada do que a função de distribuição da variável. O seguinte teorema nos dá condições suficientes para lidar com densidades.

Teorema 2.25. *Se as Suposições 1 e 2 estão verificadas e se a densidade $f(x)$ existe então as condições*

1. $f(x + \varepsilon)/f(x) \rightarrow 0$ quando $x \rightarrow \infty$ para todo $\varepsilon > 0$
2. $f(kx)/f(x) \rightarrow 0$ quando $x \rightarrow \infty$ para todo $k > 1$

são suficientes para F ser absoluta e relativamente resistente a outliers, respectivamente. Estas condições também são necessárias se adicionarmos a hipótese de que a densidade tem cauda direita monótona.

Demonstração. Ver Green [26]. □

Exemplo 2.7. Para a distribuição normal $N(0, \sigma^2)$ temos que

$$\frac{f(x + \varepsilon)}{f(x)} = \exp \left\{ -\frac{(x + \varepsilon)^2 - x^2}{2\sigma^2} \right\} = \exp \left\{ -\frac{2x\varepsilon + \varepsilon^2}{2\sigma^2} \right\} \rightarrow 0$$

quando $x \rightarrow \infty$. Temos também que

$$\frac{f(kx)}{f(x)} = \exp \left\{ -\frac{(kx)^2 - x^2}{2\sigma^2} \right\} = \exp \left\{ -x^2 \frac{k^2 - 1}{2\sigma^2} \right\} \rightarrow 0$$

quando $x \rightarrow \infty$ e $k > 1$. Portanto a distribuição normal é tanto ARO quanto RRO.

2.2.2 Distribuições propensas a outliers

Definição 2.26. Uma distribuição F será chamada de absolutamente propensa a outliers (APO) se existirem constantes $\varepsilon > 0$, $\delta > 0$ e um inteiro n_0 tais que

$$P(X_{(n)} - X_{(n-1)} > \varepsilon) \geq \delta \quad (2.17)$$

para todo inteiro $n \geq n_0$.

Definição 2.27. Uma distribuição F será chamada de relativamente propensa a outliers (RPO) se existirem constantes $k > 1$, $\delta > 0$ e um inteiro n_0 tais que

$$P(X_{(n)}/X_{(n-1)} > k) \geq \delta \quad (2.18)$$

para todo inteiro $n \geq n_0$.

Ambas as definições refletem que, para amostras suficientemente grandes de variáveis aleatórias com distribuições propensas a outliers (PO), sempre existe uma probabilidade não nula de que o valor observado mais elevado supere os demais valores em termos absolutos. Portanto, podemos esperar que eventualmente a distribuição gere um valor afastado dos demais e deste fato vem a denominação de distribuições propensas a outliers. Tal propriedade será de grande interesse neste trabalho pois esta reflete a opinião de que uma das fontes de informação pode eventualmente se distanciar da outra através do outlier.

Exemplo 2.8. A distribuição gama é APO e RRO. Na Figura 2.3 temos estimativas, utilizando a função de distribuição empírica, para as probabilidades (2.17) e (2.18) com $\varepsilon = 1$ e $k = 1,5$ para diversos tamanhos de amostra considerando $X \sim \text{Gama}(3, 1/2)$.

Exemplo 2.9. A distribuição Cauchy é tanto APO quanto RPO. Na Figura 2.9 temos estimativas para as probabilidades (2.17) e (2.18) com $\varepsilon = 1$ e $k = 1,5$ para diversos tamanhos de amostra considerando $X \sim \text{Cauchy}(0, 1)$.

Teorema 2.28. Sob as hipóteses 1 e 2, a distribuição F será absolutamente propensa a outliers se e somente se existirem constantes $\alpha > 0$ e $\beta > 0$ tais que

$$\frac{\bar{F}(x + \beta)}{\bar{F}(x)} \geq \alpha \quad (2.19)$$

para todo x finito.

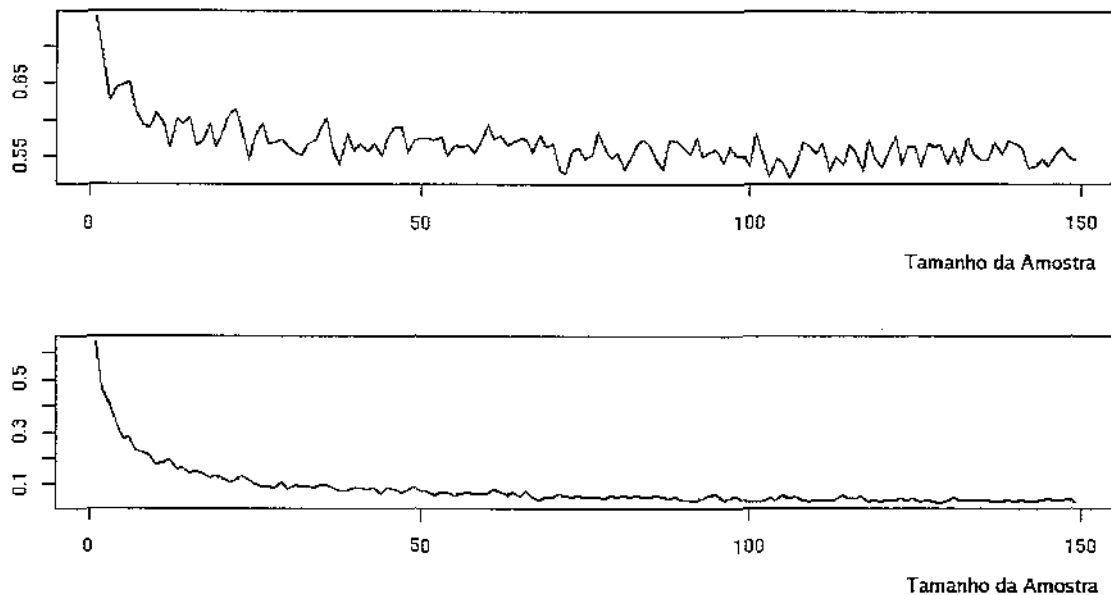


Figura 2.3: Acima: estimativas para $P(X_{(n)} - X_{(n-1)} > 1)$; Abaixo: estimativas para $P(X_{(n)}/X_{(n-1)} > 1,5)$. Em ambos os gráficos, temos que $X \sim \text{Gama}(3, 1/2)$

Teorema 2.29. *Sob as hipóteses 1 e 2, a distribuição F será relativamente propensa a outliers se e somente se existirem constantes $k > 1$ e $\delta > 0$ tais que*

$$\frac{\bar{F}(kx)}{\bar{F}(x)} \geq \delta \quad (2.20)$$

para todo x finito.

Demonstração. As demonstrações podem ser encontradas em Green [25]. □

Como posteriormente trabalharemos apenas com densidades, damos as seguintes condições suficientes para lidarmos com densidades no lugar de funções de distribuição.

Teorema 2.30. *Sob as hipóteses 1 e 2, se a densidade f existe, então as seguintes condições são suficientes para que a distribuição seja absolutamente ou relativamente propensa a outliers.*

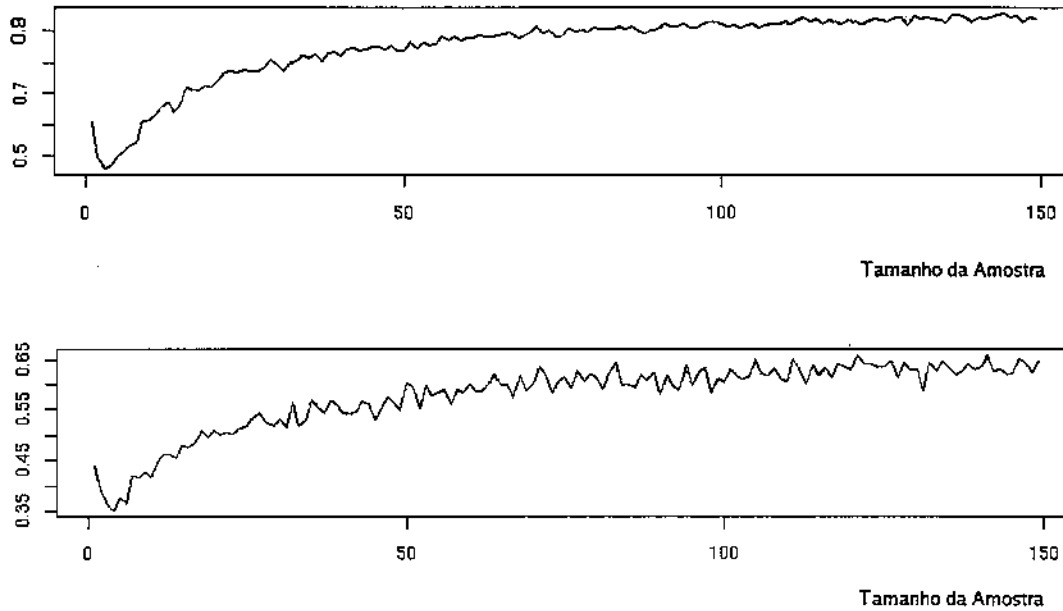


Figura 2.4: Acima: estimativas para $P(X_{(n)} - X_{(n-1)} > 1)$; Abaixo: estimativas para $P(X_{(n)}/X_{(n-1)} > 1.5)$. Em ambos os gráficos, temos que $X \sim Cauchy(0, 1)$

- Existem constantes $\varepsilon > 0$, $\delta > 0$, x_0 tais que

$$\frac{f(x + \varepsilon)}{f(x)} \geq \delta$$

para todo $x \geq x_0$.

- Existem constantes $k > 1$, $\delta > 0$, x_0 tais que

$$\frac{f(kx)}{f(x)} \geq \delta$$

para todo $x \geq x_0$.

Demonstração. A demonstração pode ser encontrada em Green [25]. □

Exemplo 2.10. As distribuições gama, logística e exponencial dupla são *APO* e *RRO*. De fato,

- para a distribuição $Gama(\alpha, \beta)$ temos que

$$\frac{\bar{F}(x + \varepsilon)}{\bar{F}(x)} = \exp\{-\beta\varepsilon\}$$

e

$$\frac{\bar{F}(kx)}{\bar{F}(x)} = \exp\{-x\beta(k-1)\} \rightarrow 0$$

quando $x \rightarrow \infty$ e $k > 1$.

- para a distribuição exponencial dupla temos que

$$\frac{f(x + \varepsilon)}{f(x)} = \exp\{-\lambda(|x + \varepsilon| + |x|)\} \geq \exp\{-\lambda\varepsilon\}$$

e

$$\frac{f(kx)}{f(x)} = \exp\{-\lambda|x|(k-1)\} \rightarrow 0$$

quando $x \rightarrow \infty$ e $k > 1$.

- para a distribuição logística, temos que

$$\frac{\bar{F}(x + \varepsilon)}{\bar{F}(x)} = \exp\{-\varepsilon\} \left\{ \frac{1 + \exp\{-x\}}{1 + \exp\{-(x + \varepsilon)\}} \right\} \geq \frac{\exp\{-\varepsilon\}}{1 + \exp\{-\varepsilon\}}.$$

Exemplo 2.11. A distribuição t-Student é tanto *APO* quanto *RPO*. De fato, pelo Lema 3.7 temos que existem constantes A e B tais que

$$\frac{f(x + \varepsilon)}{f(x)} = \left\{ \frac{v + (x + \varepsilon)^2}{v + x^2} \right\}^{-\frac{v+1}{2}} \geq A$$

e

$$\frac{f(kx)}{f(x)} = \left\{ \frac{v + (kx)^2}{v + x^2} \right\}^{-\frac{v+1}{2}} \geq B$$

e o resultado segue via Teorema 2.30.

2.2.3 Classes de distribuição segundo suas propriedades relacionadas a outliers

É possível classificar uma distribuição utilizando as propriedades relacionadas a *outliers* vistas nas seções anteriores. Existem seis classificações possíveis utilizando tais propriedades, considerando apenas a cauda direita. Estas classificações são dadas a seguir. Para informações adicionais ver Green [25].

- *Classe I*: distribuições que são absolutamente resistentes a *outliers*. Estas distribuições também são relativamente resistentes a *outliers*. A distribuição normal pertence a esta classe;
- *Classe II*: distribuições que são relativamente resistentes a *outliers* mas não são nem absolutamente resistentes nem absolutamente propensas a *outliers*. A distribuição Poisson pertence a esta classe;
- *Classe III*: distribuições que são absolutamente propensas e relativamente resistentes a *outliers*. A distribuição gama pertence a esta classe;
- *Classe IV*: distribuições que são absolutamente propensas a *outliers* mas não são nem relativamente resistentes a *outliers* nem relativamente propensas a *outliers*. Temos que X pertence à Classe IV, onde

$$P(X = f(k)) = 2^{-k}, \quad k = 1, 2, 3, \dots$$

com

$$f(k) = 10^{\lfloor \log_2 k \rfloor} + k - 2^{\lfloor \log_2 k \rfloor}.$$

- *Classe V*: distribuições que são relativamente propensas a *outliers*. Tais distribuições também são absolutamente propensas a *outliers*. A distribuição Cauchy pertence a esta classe;
- *Classe VI*: distribuições que não são nem absolutamente nem relativamente propensas a *outliers*. Temos que Y pertence à classe VI, onde

$$P(Y = 2^k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad \lambda > 0, \quad k = 0, 1, 2, \dots$$

2.3 Relações entre distribuições *PO* e as distribuições de cauda pesada

Apesar de não termos encontrado referências sobre a relação entre distribuições propensas a *outliers* e distribuições de cauda pesada, parece bastante razoável supor que haja uma relação entre ambas. Devido a esta lacuna na literatura, parte deste trabalho foi dedicada para mostrar que esta relação existe de fato, mostrando o que era intuitivo: que distribuições de caudas pesadas são propensas a gerar *outliers*. Mostramos entretanto

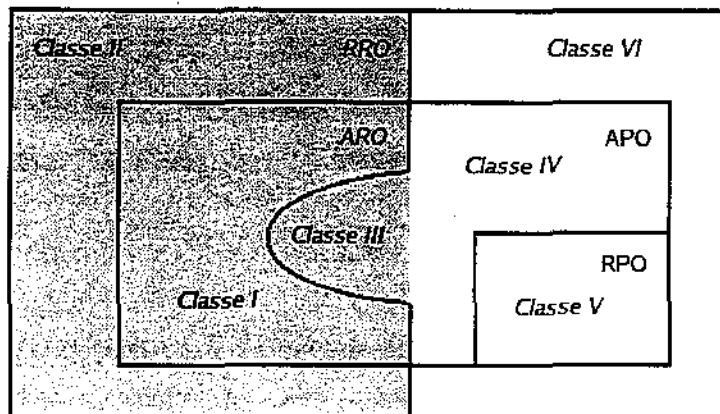


Figura 2.5: Classes segundo suas propriedades relacionadas a *outliers*. Em cinza temos as distribuições *RRO*.

que as distribuições de cauda pesada apresentadas anteriormente são uma subclasse das distribuições propensas a *outliers*, motivo pelo qual optamos por nos restringir apenas a relacionar as distribuições de cauda pesada com conflitos de informação (salvo o resultado de Pericchi e Sansó, visto na Seção 3.3 e discutido na Seção 4.2).

Discutiremos primeiramente a relação existente entre as distribuições relativamente propensas a *outliers* e as distribuições de variação dominada. Para este fim, começaremos mostrando que o Teorema 2.29 pode ser escrito de uma maneira mais geral, como mostra a proposição abaixo.

Proposição 2.31. *Se a desigualdade (2.20) for válida para algum $k > 1$ e δ , então para todo $k' > 1$ existe δ' tal que*

$$\frac{\bar{F}(k'x)}{\bar{F}(x)} > \delta',$$

para todo x finito

Demonstração. Suponha que as condições do Teorema (2.29) estão satisfeitas. Notemos primeiramente que

$$\frac{\bar{F}(k^2x)}{\bar{F}(x)} = \frac{\bar{F}(k^2x)}{\bar{F}(kx)} \frac{\bar{F}(kx)}{\bar{F}(x)} \geq \delta^2$$

e análogamente mostramos que

$$\frac{\bar{F}(k^n x)}{\bar{F}(x)} \geq \delta^n.$$

Agora, escolha qualquer $k' > 1$ e tome n_0 como sendo o menor natural tal que $n_0 \geq \log k'/\log k$. Temos então que

$$\frac{\bar{F}(k'x)}{\bar{F}(x)} \geq \frac{\bar{F}(k^{n_0}x)}{\bar{F}(x)} \geq \delta^{n_0}$$

e fazendo $\delta^{n_0} = \delta'$ obtemos o resultado desejado. \square

Proposição 2.32. *F é RPO se e somente se $F \in \mathcal{DV}$.*

Demonstração. Notemos primeiramente que, fazendo $y = x\theta$, temos que

$$\frac{\bar{F}(x\theta)}{\bar{F}(x)} \geq \delta \iff \frac{\bar{F}(y/\theta)}{\bar{F}(y)} \leq \frac{1}{\delta}$$

supondo $\delta > 0$. Agora, suponha que F é RPO. Então, pela Proposição 2.31, temos que para todo $k > 1$ existe um δ^* finito tal que

$$\frac{\bar{F}(y/k)}{\bar{F}(y)} \leq \delta^* \Rightarrow \sup_y \frac{\bar{F}(y/k)}{\bar{F}(y)} \leq \delta^* < \infty \Rightarrow \limsup_y \frac{\bar{F}(y/k)}{\bar{F}(y)} \leq \delta^* < \infty,$$

para todo y finito, onde $\delta^* = (1/\delta)$. Então temos que F ser RPO implica que $F \in \mathcal{DV}$. Suponha agora que $F \in \mathcal{DV}$. Então existem y_0 e B tais que para todo $y > y_0$ temos que $\bar{F}(y/k)/\bar{F}(y) < B$. Disto temos que

$$\begin{aligned} \frac{\bar{F}(y/k)}{\bar{F}(y)} &= \frac{\bar{F}(y/k)}{\bar{F}(y)} I_{(-\infty, y_0]}(y) + \frac{\bar{F}(y/k)}{\bar{F}(y)} I_{(y_0, \infty)}(y) \\ &\leq \frac{1}{\bar{F}(y_0)} + B < \infty \Rightarrow \sup_y \frac{\bar{F}(y/k)}{\bar{F}(y)} < \infty \end{aligned}$$

e portanto $F \in \mathcal{DV}$ implica em F ser RPO. \square

Portanto, temos que F ser RPO é equivalente a termos $F \in \mathcal{DV}$. Disto temos que a classe \mathcal{DV} está intimamente relacionada com *outliers*, no sentido de Neyman e Scott, mesmo a parte que não está contida em \mathcal{L} .

A seguir discutiremos a relação entre distribuições APO e distribuições em \mathcal{L} . Começaremos enfraquecendo o Teorema 2.28, como mostra a proposição abaixo.

Proposição 2.33. *Se a desigualdade (2.19) estiver verificada para algum β e $\alpha > 0$, então para qualquer $\beta' \in \mathbb{R}$ existe um $\alpha' > 0$ tal que (2.19) é verdadeira.*

Demonstração. Suponha que a desigualdade (2.19) é válida e considere que $\beta' < \beta$. Então,

$$\frac{\bar{F}(x + \beta')}{\bar{F}(x)} \geq \frac{\bar{F}(x + \beta)}{\bar{F}(x)} \geq \alpha$$

e neste caso basta tomar $\alpha' = \alpha$. Agora suponha que $\beta' > \beta$. Notemos que

$$\frac{\bar{F}(x + 2\beta)}{\bar{F}(x)} = \frac{\bar{F}(x + 2\beta)}{\bar{F}(x + \beta)} \frac{\bar{F}(x + \beta)}{\bar{F}(x)} \geq \alpha^2$$

e de maneira análoga temos que

$$\frac{\bar{F}(x + n\beta)}{\bar{F}(x)} \geq \alpha^n.$$

Tome n_0 como sendo qualquer inteiro que satisfaça $\beta' \leq n_0\beta$. Temos então que

$$\frac{\bar{F}(x + \beta')}{\bar{F}(x)} \geq \frac{\bar{F}(x + n_0\beta)}{\bar{F}(x)} \geq \alpha^{n_0}$$

e o resultado segue fazendo $\alpha' = \alpha^{n_0}$. □

Com a proposição acima mostramos que uma distribuição é absolutamente propensa a outliers se

$$\inf_x \frac{\bar{F}(x + \beta)}{\bar{F}(x)} > 0$$

para qualquer $\beta \in \mathbb{R}$. Isto motiva a proposição a seguir.

Proposição 2.34. *Uma distribuição é APO se e somente se*

$$\limsup_{x \rightarrow \infty} \frac{\bar{F}(x - \beta)}{\bar{F}(x)} < \infty, \tag{2.21}$$

para todo $\beta \in \mathbb{R}$.

Demonstração. Notemos que, fazendo $x = y + \beta$,

$$\frac{\bar{F}(y + \beta)}{\bar{F}(y)} > 0 \Rightarrow \frac{\bar{F}(x - \beta)}{\bar{F}(x)} < \infty.$$

Logo, supondo que F é APO e utilizando a Proposição 2.33, temos que

$$\sup_x \frac{\bar{F}(x - \beta)}{\bar{F}(x)} < \infty \Rightarrow \limsup_{x \rightarrow \infty} \frac{\bar{F}(x - \beta)}{\bar{F}(x)} < \infty$$

e portanto F ser APO implica a desigualdade (2.21). Agora, suponha que (2.21) está verificada. Então temos que existem x_0 e A tais que para todo $x > x_0$, $\bar{F}(x - \beta)/\bar{F}(x) < A$. Assim

$$\begin{aligned} \frac{\bar{F}(x - \beta)}{\bar{F}(x)} &= \frac{\bar{F}(x - \beta)}{\bar{F}(x)} I_{(-\infty, x_0]}(x) + \frac{\bar{F}(x - \beta)}{\bar{F}(x)} I_{(x_0, \infty)}(x) \\ &\leq \frac{1}{\bar{F}(x_0)} + A < \infty \end{aligned}$$

portanto

$$\sup_x \frac{\bar{F}(x - \beta)}{\bar{F}(x)} < \infty$$

e F é APO . □

A proposição acima caracteriza as distribuições APO como sendo todas as distribuições tais que

$$\limsup_{x \rightarrow \infty} \frac{\bar{F}(x - \beta)}{\bar{F}(x)} < \infty$$

análogamente à definição da classe DV . Uma subclasse das distribuições APO interessante é aquela na qual o limite

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(x - \beta)}{\bar{F}(x)}$$

existe. Este limite é bastante conhecido na literatura e uma de suas utilidades na teoria de valores extremos é caracterizar as distribuições que pertencem ao domínio de atração da distribuição Gumbel. O seguinte lema mostra que se esse limite existe, ele é da forma $\exp\{\alpha x\}$.

Lema 2.35. *Se*

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(x - y)}{\bar{F}(x)} = \phi(y) < \infty \tag{2.22}$$

para todo $y \in \mathbb{R}$ então $\phi(y) = \exp\{h(y)\}$ onde h é um operador linear.

Demonstração. Supondo que o limite em (2.22) existe, temos que

$$\frac{\bar{F}(x-2y)}{\bar{F}(x)} = \frac{\bar{F}(x-2y)}{\bar{F}(x-y)} \frac{\bar{F}(x-y)}{\bar{F}(x)} = \phi(y)^2$$

e análogamente mostramos que

$$\frac{\bar{F}(x-ny)}{\bar{F}(x)} = \phi(y)^n$$

mas por (2.22) temos que

$$\frac{\bar{F}(x-ny)}{\bar{F}(x)} = \phi(ny)$$

logo, temos que $\phi(y)^n = \phi(ny)$. Com um raciocínio semelhante, mostramos que

$$\frac{\bar{F}(x-y_1-y_2)}{\bar{F}(x)} = \phi(y_1)\phi(y_2)$$

e de maneira similar, temos que

$$\phi\left(\sum_{i=1}^n y_i\right) = \left(\prod_{i=1}^n \phi(y_i)\right).$$

A solução para as relações acima é dada por $\phi(y) = e^{h(y)}$ e disto temos que $h(\sum_{i=1}^n y_i) = \sum_{i=1}^n h(y_i)$ e $h(ny) = nh(y)$, portanto h é linear. Além disso, como $\phi(0) = 1 \Rightarrow h(0) = 0$ temos que $h(y) = \alpha y$. \square

Assim, podemos relacionar as distribuições *APO* com as distribuições em \mathcal{L} como segue:

Corolário 2.36. *Seja \mathcal{A} o conjunto de todas as distribuições *APO*. Então temos que $\mathcal{L} \subset \mathcal{A}$.*

Demonstração. Basta notar que, se $F \in \mathcal{L}$ então

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(x-\beta)}{\bar{F}(x)} = 1 = \exp\{0\}$$

e portanto $F \in \mathcal{A}$. \square

Com estes resultados, podemos classificar as distribuições propensas a *outliers* nas seguintes classes:

- \mathcal{A} : como definida anteriormente, é a classe que contém as distribuições absolutamente propensas a *outliers*;
- \mathcal{DV} : como definida anteriormente, é a classe de distribuições de variação dominada;
- \mathcal{G} : diremos que $F \in \mathcal{G}$ se

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(x - y)}{\bar{F}(x)}$$

existe;

- \mathcal{L} : como definida anteriormente, é a classe de distribuições de cauda longa;

As relações entre as classes apresentadas estão representadas na Figura 2.3. Quase todas já foram discutidas, com exceção da apresentada na proposição abaixo.

Proposição 2.37. $\mathcal{G} \cap \mathcal{L}^c \cap \mathcal{DV} = \emptyset$

Demonstração. Suponha que $F \in \mathcal{G} \cap \mathcal{L}^c$. Temos então que para qualquer $\theta \in \mathbb{R}$

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(x - \theta)}{\bar{F}(x)} = \exp\{\beta\theta\}$$

com $\beta > 0$. Seja $L(x) = \bar{F}(x) \exp\{x\}$ e suponha que $\theta > 1$. Temos então que

$$\frac{\bar{F}(x/\theta)}{\bar{F}(x)} = \frac{L(x/\theta)}{L(x)} \exp\left\{-x \left(\frac{1}{\theta} - 1\right)\right\}$$

Notemos que a equação acima só terá limite superior finito se $L(x/\theta)/L(x)$ tender a zero mais rápido do que $\exp\{-x(1/\theta - 1)\}$ tende ao infinito, e neste caso

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(x/\theta)}{\bar{F}(x)} = 0.$$

Suponha que a relação acima é verdadeira. Como $\theta > 1$ temos que $x - \theta > (x - \theta^2)/\theta$ e então

$$0 \leq \frac{\bar{F}(x - \theta)}{\bar{F}(x)} \leq \frac{\bar{F}((x - \theta^2)/\theta)}{\bar{F}(x)} = \frac{\bar{F}((x - \theta^2)/\theta)}{\bar{F}(x/\theta)} \frac{\bar{F}(x/\theta)}{\bar{F}(x)} \rightarrow 0$$

quando $x \rightarrow \infty$, o que é absurdo, uma vez que $\bar{F}(x - \theta)/\bar{F}(x) \rightarrow \exp(\theta\beta)$ quando $x \rightarrow \infty$. Portanto temos que $F \notin \mathcal{DV}$ e $\mathcal{G} \cap \mathcal{L}^c \cap \mathcal{DV} = \emptyset$. \square

Seguem abaixo alguns exemplos de distribuições propensas a *outliers* segundo as classes apresentadas nesta seção:

- Distribuição em $\mathcal{A} \cap \{\mathcal{G} \cup \mathcal{DV}\}^c$:

$$f(x) = \frac{2}{5}(2 + \sin x) \exp\{-x\}$$

- Distribuições em $\mathcal{G} \cap \mathcal{L}^c$: gama, exponencial dupla, logística
- Distribuições em $\mathcal{DV} \cap \mathcal{G}^c$:
 - Distribuição de Peter e Paul
 - $f(x) = (1/\pi) \times (2 + \sin x)/(1 + x^2)$

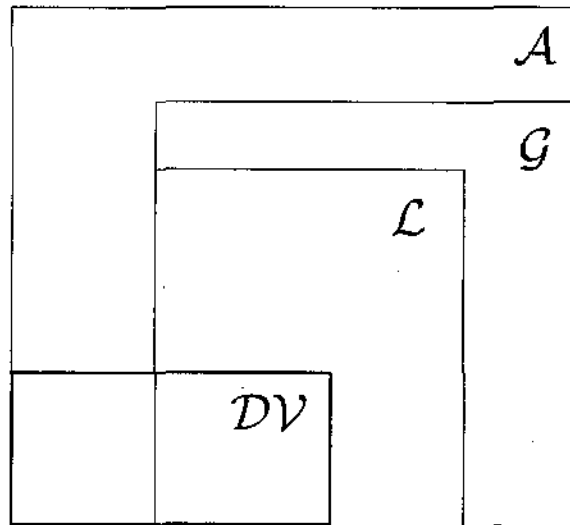


Figura 2.6: Outra possível classificação para as distribuições *PO*

Capítulo 3

Alguns estudos sobre conflitos de informação

Neste capítulo apresentamos resultados da literatura de conflitos de informação. Esta, por sua vez, está intimamente relacionada com o tratamento de *outliers* em inferência bayesiana. Dos diversos resultados, selecionamos aqueles que atribuem, direta ou indiretamente, distribuições de cauda pesada para pelo menos uma das fontes de informação.

Na Seção 3.1 discutimos o que é um conflito de informação e o que significa resolvê-lo. Na Seção 3.2 discutimos o teorema de Dawid - resultado muito citado na literatura revisada. Apresentamos também o conjunto das condições de O'Hagan, que implicam o teorema de Dawid. Na Seção 3.3 apresentamos os resultados obtidos em Pericchi e Sansó [40] para resolução de conflitos entre as fontes. Na Seção 3.4 discutimos a medida *credence* e sua utilização em conflitos de informação quando as fontes possuem uma estrutura de locação. Por último, na Seção 3.5 discutimos o uso de distribuições de variação regular no estudo de conflitos de informação.

3.1 Mais sobre conflitos de informação

Suponhamos que estamos interessados em realizar inferências sobre a variável aleatória Θ . Com a função de verossimilhança, podemos obter informações sobre Θ através dos dados, enquanto que a priori reflete uma opinião prévia sobre este. Essas duas fontes de

informação são combinadas para gerar inferências a posteriori sobre Θ . Eventualmente, na presença de um *outlier*, uma das fontes pode gerar uma informação que desacredite a outra, gerando um conflito entre as fontes.

Contudo, o interesse maior nos estudos de conflitos entre as fontes de informação não está em identificar um conflito, mas sim em dar suporte para que um possível conflito seja resolvido. Dizemos que um conflito de informação está resolvido se, de alguma forma, uma das fontes de informação domina as demais. Verificamos isto via o comportamento assintótico da distribuição a posteriori. Considere que f e g são as densidades das fontes de informação e seja p a densidade a posteriori para um determinado problema. Seja x uma observação. Na literatura estudada, predominam quatro tipos de situações específicas relacionadas aos conflitos de informação:

1. **Rejeição Completa:** neste caso temos que $p \rightarrow f$ quando $|x| \rightarrow \infty$. A fonte de informação g é ignorada e inferências a posteriori são baseadas unicamente em f .
2. **Rejeição Parcial:** neste caso temos que $p \rightarrow \psi f$ quando $|x| \rightarrow \infty$. A fonte de informação g não foi totalmente rejeitada, estando toda esta condensada em ψ . Diz-se então que g se tornou pouco informativa.
3. **Dominância:** neste caso, existem constantes $K > k > 0$ tais que para todo $|x|$ suficientemente grande $kf < p < Kf$. A posteriori é limitada por múltiplos positivos de f , e dizemos que f domina g .
4. **Falha:** a distribuição a posteriori falha em conseguir conciliar as informações em uma única moda.

Notemos que deixamos em aberto a escolha entre f e g para a priori e verossimilhança. O motivo disto é que podemos ter mais credibilidade na informação proveniente dos dados do que na informação a priori, resolvendo um possível conflito a favor dos dados. O contrário também é válido: podemos ter mais credibilidade na informação a priori do que nos dados, e neste caso, quando as fontes conflitarem, optaremos pela informação a priori.

Com o intuito de explicitarmos melhor quando a posteriori não consegue resolver um conflito, consideremos $f(x|\theta)$ a função de verossimilhança, $\pi(\theta)$ a densidade a priori e $p(\theta|x)$ a densidade a posteriori. Do teorema de Bayes temos

$$p(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta} = \frac{f(x|\theta)\pi(\theta)}{m(x)} \quad (3.1)$$

onde $m(x) = \int f(x|\theta)\pi(\theta)d\theta$. Notemos que $m(x)$ será pequena sempre que o produto $f(x|\theta)\pi(\theta)$ for pequeno para todo θ . Isto caracteriza um conflito de informação que a posteriori não consegue resolver, uma vez que para todo θ no qual a verossimilhança é alta a priori é baixa e sempre que a densidade a priori é alta a verossimilhança de θ é baixa, de modo que a forma da densidade a posteriori pode mudar dramaticamente em resposta a qualquer mudança nos valores das caudas de $f(x|\theta)$ ou $\pi(\theta)$.

Contudo, não existe medida absoluta para o quão pequeno deveria ser $m(x)$, a ponto de caracterizar um conflito entre as fontes de informação. Três possíveis índices, chamados de *índices de surpresa*, apresentados em O'Hagan e Foster [35], são dados a seguir:

1. $\sup_{\theta \in \Theta} \frac{f(x|\theta)}{m(x)}$
2. $\frac{\sup_x m(x)}{m(x)}$
3. $\int_Y m(y)dy$ onde $Y = \{y : m(y) \leq m(x)\}$

O primeiro mensura, em certo sentido, o quão “surpreso” estaríamos com relação ao valor de x observado de acordo com a priori. A Figura 3.1 mostra a razão entre o índice de surpresa 1 de uma verossimilhança $N(\theta, 1)$ e o de uma verossimilhança $t_{(31, \theta)}$, ambas com priori $N(0, 1)$. Notemos que a razão aumenta a medida que nos afastamos de zero, implicando que estamos mais “surpresos” em observar um valor absoluto elevado no caso em que a distribuição é normal em comparação com a t-Student. O segundo índice tenta mensurar a surpresa para outros valores de x considerando a priori utilizada, enquanto que o terceiro analisa a probabilidade de observar um dado com densidade marginal menor ou igual ao valor de $m(x)$ para o dado observado. Notemos que os dois últimos índices violam o princípio da verossimilhança, uma vez que se referem a dados que não foram observados.

A seguir, analisamos dois exemplos. No primeiro, consideramos o exemplo iniciado na introdução. No segundo consideramos o caso no qual o parâmetro de interesse é de escala e mostramos que a posteriori consegue resolver o conflito parcialmente.

Exemplo 3.1. Consideremos que para um determinado problema tenhamos que $X|\theta$ tem distribuição $N(\theta, 1)$ e que Θ tem distribuição $N(0, 1)$. Suponha que observamos $x = 8$, que neste caso é um valor atípico. Dissemos anteriormente que a posteriori falhava em resolver o conflito, uma vez que esta era fortemente influenciada pelo *outlier*. De fato, se observarmos a Figura 3.2, notaremos que a posteriori se encontra em uma região afastada

das modas das fontes e sua região de mais alta densidade corresponde à região das caudas das fontes. Podemos ver, na Figura 4.1, que o mesmo não acontece quando uma das fontes tem distribuição t-Student.

Ainda na introdução, discutimos que a escolha da distribuição dos dados como sendo uma t-Student faria com que a posteriori convergisse para a priori quando $|x| \rightarrow \infty$. Na seção seguinte mostraremos que isto é verdadeiro com a utilização do teorema de Dawid.

Exemplo 3.2. Consideremos agora que $f(x|\theta) = 2\theta^2 x^{-3} I_{(\theta, \infty)}$ e que $\pi(\theta) = e^{-\theta} I_{(0, \infty)}(\theta)$. Notemos que $E(X|\theta) = 2\theta$ e que $E(\Theta) = 1$. Assim, para um valor x observado, temos que a verossimilhança sugere que Θ esteja próximo de $x/2$ enquanto que a priori sugere que Θ esteja próximo de 1. Temos que se $x \rightarrow \infty$ as fontes entrarão em conflito. Neste caso,

$$\lim_{x \rightarrow \infty} p(\theta|x) = \lim_{x \rightarrow \infty} \frac{\theta^2 e^{-\theta} I_{(0, x)}(\theta)}{2 - e^{-x}(x+1)^2} = \frac{1}{2} \theta^2 e^{-\theta} I_{(0, \infty)}(\theta) = \frac{1}{2} \theta^2 \pi(\theta)$$

Observemos que a informação proveniente dos dados não foi completamente rejeitada. Pode-se dizer que o termo θ^2 contém toda a informação do *outlier*. Este comportamento, onde a posteriori é proporcional à fonte dominante vezes um polinômio é característico das distribuições de variação regular, como será discutido na Seção 3.5.

3.2 O teorema de Dawid

Na seção anterior vimos exemplos de situações de conflito entre as fontes de informação. Vimos que a escolha de certas distribuições acarreta decisões capazes de resolver o conflito, no sentido de que uma das fontes de informação será mais privilegiada que as demais. Apresentamos nesta seção um resultado bastante referenciado na literatura pesquisada, que nos permite dar condições suficientes para escolher distribuições com as quais um conflito estará resolvido em favor de uma das fontes. Considere um problema com verossimilhança $g(x|\theta) = f(x - \theta)$ conhecida. Temos então o seguinte teorema:

Teorema 3.1 (Condições de Dawid). *Sejam g a verossimilhança tal que $g(x|\theta) = f(x - \theta)$, π a densidade a priori para Θ e suponha que a esperança a priori de $M(\Theta)$ seja finita. Então as seguintes condições são suficientes para que a esperança a posteriori $E\{M(\Theta)|X = x\}$ convirja para a esperança a priori $E_{\pi}\{M(\Theta)\}$ quando $x \rightarrow \infty$*

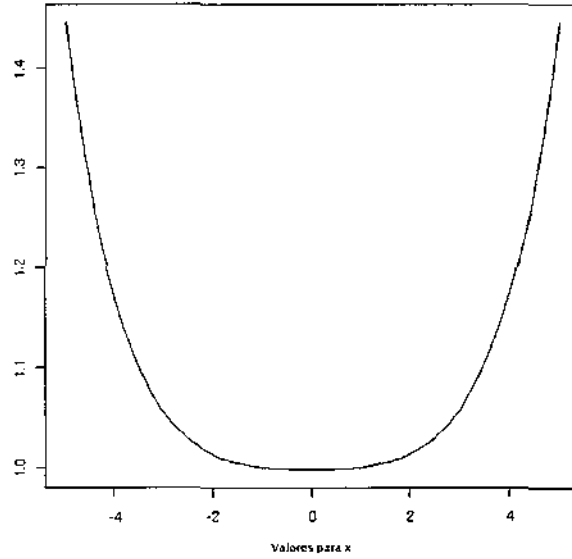


Figura 3.1: Razão entre os índices de surpresa 1 de uma verossimilhança $N(\theta, 1)$ e uma $t_{(31, \theta)}$, ambas considerando uma priori $N(0, 1)$.

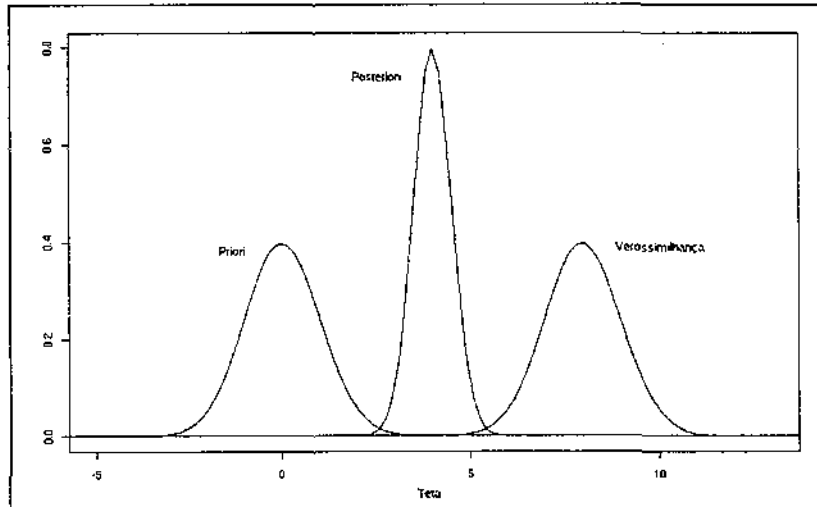


Figura 3.2: Priori, Verossimilhança e Posteriori em função de θ com $x = 8$

1. Dados $\varepsilon > 0$, $y' > 0$ existe x_0 tal que para todo $x > x_0$

$$|f(x') - f(x)| < \varepsilon f(x)$$

sempre que $|x' - x| < y'$.

2. Para algumas constantes B, C ,

$$0 < f(x') < Cf(x)$$

sempre que $x' > x > B$.

3. (a) $\int_{-\infty}^{\infty} k(\theta)\pi(\theta)d\theta < \infty$

(b) $\int_{-\infty}^{\infty} |M(\theta)|k(\theta)\pi(\theta)d\theta < \infty$

onde $k(\theta) = \sup_x \{f(x - \theta)/f(x)\}$.

Demonstração. Suponha que valem as três condições do teorema. Sejam $n(x) = \int_{\mathbb{R}} M(\theta)f(x - \theta)\pi(\theta)d\theta$ e $m(x) = \int_{\mathbb{R}} f(x - \theta)\pi(\theta)d\theta$. Temos que

$$E\{M(\Theta)|X = x\} = \frac{\int_{\mathbb{R}} M(\theta)f(x - \theta)\pi(\theta)d\theta}{m(x)} = \frac{n(x)}{m(x)}.$$

Então

$$\begin{aligned} \left\{ \frac{n(x)}{f(x)} - E_{\pi}\{M(\Theta)\} \right\} &= \int_{\mathbb{R}} \frac{M(\theta)f(x - \theta)\pi(\theta)}{f(x)} - M(\theta)\pi(\theta)d\theta \\ &= \int_{\mathbb{R}} M(\theta)\pi(\theta) \left\{ \frac{f(x - \theta)}{f(x)} - 1 \right\} d\theta. \end{aligned}$$

Suponha que $x > B$. Então, para $h > 0$,

$$\begin{aligned} &\left| \left(\int_{-\infty}^{-h} + \int_h^{\infty} \right) M(\theta)\pi(\theta) \left\{ \frac{f(x - \theta)}{f(x)} - 1 \right\} d\theta \right| \\ &\leq \int_{-\infty}^{-h} |M(\theta)|\pi(\theta) \left| \frac{f(x - \theta)}{f(x)} - 1 \right| d\theta + \int_h^{\infty} |M(\theta)|\pi(\theta) \left| \frac{f(x - \theta)}{f(x)} - 1 \right| d\theta \\ &\leq \int_{-\infty}^{-h} |M(\theta)|\pi(\theta) \left(\frac{f(x - \theta)}{f(x)} + 1 \right) d\theta + \int_h^{\infty} |M(\theta)|\pi(\theta) \left(\frac{f(x - \theta)}{f(x)} + 1 \right) d\theta \\ &\leq (C + 1) \int_{-\infty}^{-h} |M(\theta)|\pi(\theta)d\theta + \int_h^{\infty} |M(\theta)|\pi(\theta)(k(\theta) + 1)d\theta, \end{aligned}$$

que tende a zero quando h tende ao infinito em virtude de (3)(b) e pela existência de $E_\pi\{M(\Theta)\}$. Além disso, para h fixado, temos que

$$\left| \int_{\mathbb{R}} M(\theta)\pi(\theta) \left\{ \frac{f(x-\theta)}{f(x)} - 1 \right\} d\theta \right| \leq \int_{\mathbb{R}} M(\theta)\pi(\theta)(k(\theta) + 1)d\theta < \infty,$$

logo, pela condição (1)

$$\int_{-h}^h M(\theta)\pi(\theta) \left\{ \frac{f(x-\theta)}{f(x)} - 1 \right\} d\theta \rightarrow 0$$

quando $x \rightarrow \infty$. Assim, fazendo $h \rightarrow \infty$ e $x \rightarrow \infty$, temos que

$$\int_{\mathbb{R}} M(\theta)\pi(\theta) \left\{ \frac{f(x-\theta)}{f(x)} - 1 \right\} d\theta \rightarrow 0$$

e portanto

$$\frac{n(x)}{f(x)} \rightarrow E_\pi\{M(\Theta)\}$$

quando $x \rightarrow \infty$. Agora, fazendo $M(\theta) \equiv 1$, temos que

$$\frac{m(x)}{f(x)} \rightarrow 1$$

quando $x \rightarrow \infty$. Portanto, temos que

$$E\{M(\Theta)|X = x\} = \frac{n(x)}{m(x)} = \frac{n(x)}{f(x)} \frac{f(x)}{m(x)} \rightarrow E_\pi\{M(\Theta)\}$$

quando $x \rightarrow \infty$. □

Agora, seja Λ o espaço paramétrico. Tomemos $M(A) = I(A)$, onde $A \subset \Lambda$. Temos então que $E_\pi\{M(A)\} = \int_A \pi(\theta)d\theta$ e, desde que as condições do Teorema 3.1 estejam verificadas, temos que

$$E\{M(A)|X = x\} = \int_A p(\theta|x)d\theta \rightarrow \int_A \pi(\theta)d\theta$$

quando $x \rightarrow \infty$. O resultado acima mostra que, quando satisfeitas as condições de Dawid, a distribuição a posteriori converge para a distribuição a priori. Temos então uma

completa rejeição do *outlier*, ou ainda, temos que os dados se tornam pouco informativos. Neste caso dizemos que o conflito foi resolvido a favor da distribuição a priori.

Agora, suponhamos que $g(x|\theta) = f(x - \theta)$ é a verossimilhança e que $\pi(\theta)$ é a priori. Se tomarmos $\Phi = X - \Theta$ teremos que $g(\phi + \theta|\theta) = f(\phi)$ e que $\pi(\theta) = \pi(x - \phi)$ e disto temos que

$$\int_{\mathbb{R}} f(x - \theta)\pi(\theta)d\theta = \int_{\mathbb{R}} f(\phi)\pi(x - \phi)d\phi = m(x).$$

Portanto podemos enunciar o Teorema 3.1 impondo as mesmas condições sobre a densidade da distribuição a priori $\pi(\theta)$. Assim, se as condições estiverem satisfeitas,

$$\begin{aligned} E\{M(x - \Theta)|X = x\} &= \frac{\int_{\Theta} M(x - \theta)f(x - \theta)\pi(\theta)d\theta}{m(x)} \\ &= \frac{\int_{\Phi} M(\phi)f(\phi)\pi(x - \phi)d\phi}{m(x)} \rightarrow E_f\{M(\phi)\} \end{aligned}$$

quando $x \rightarrow \infty$, onde $E_f(\cdot)$ representa a esperança com relação a densidade f . Neste caso temos que a informação a priori foi rejeitada e a informação a posteriori é baseada na verossimilhança.

Podemos agora concluir o Exemplo 3.1 Notemos que para a distribuição t-Student com v graus de liberdade, temos que

$$k(z) \propto \left\{ \frac{z}{2} + \left(c + \frac{z^2}{4} \right)^{1/2} \right\}^{v+1}$$

que se comporta como z^{v+1} no infinito. Assim, o teorema será válido sempre que $E(Z^{v+1})$ existir. Se assumirmos que $X|\theta$ tem distribuição t-Student com parâmetro de locação θ e com v graus de liberdade e que Θ tem distribuição normal, teremos que a posteriori $\Theta|x$ irá convergir para a distribuição normal quando $x \rightarrow \infty$. Contudo, se assumirmos que $X|\theta$ tem distribuição $N(\theta, \sigma^2)$ e que Θ tem distribuição t_v , teremos que a posteriori convergirá para a distribuição $N(x, \sigma^2)$ e portanto $E(\Theta|x) - x \rightarrow 0$ quando $x \rightarrow \infty$.

3.2.1 As condições de O'Hagan

O'Hagan em [33] estabeleceu um conjunto alternativo de condições suficientes para as condições de Dawid. Este conjunto é dado a seguir:

Teorema 3.2. *As seguintes condições são suficientes para que $E(M(\theta)|x) \rightarrow E_\pi(M(\theta))$ quando $x \rightarrow \infty$:*

1. Dados $\varepsilon > 0$, $y' > 0$ existe x_0 tal que para todo $x > x_0$

$$|f(x') - f(x)| < \varepsilon f(x)$$

sempre que $|x' - x| < y'$

2. (a) $f(y)$ é contínua e positiva para todo $y \in \mathbb{R}$.
 (b) Existe B tal que para todo $y \geq B$,
 i. $f(y)$ é decrescente em y ,
 ii. $b(y) = d \log f(y)/dy$ existe e é crescente em y .
 (c) Existe um $C \leq B$ tal que, para todo $y \leq C$, $f(y)$ é crescente em y .
3. (a) $\int f(\theta)^{-1} \pi(\theta) d\theta < \infty$
 (b) $\int |M(\Theta)| f(\theta)^{-1} \pi(\theta) d\theta < \infty$

Demonstração. Notemos primeiramente que as condições (1) e (2)(b)(i) de O'Hagan implicam, respectivamente, as condições (1) e (2) de Dawid. Mostraremos que, dado este novo conjunto de condições, vale a terceira condição de Dawid. Demonstraremos isto em três partes:

Parte 1: Mostraremos que existe um $x_0 < c$ tal que se $\theta > B - x_0$ teremos que $k(\theta)$ é atingido no intervalo $C + \theta \leq x \leq B + \theta$. Primeiramente façamos $r(x, \theta) = f(x - \theta)/f(x)$. Temos então que

$$\log r(x, \theta) = \log \frac{f(x - \theta)}{f(x)} = \log(f(x - \theta)) - \log(f(x))$$

é decrescente em $\theta > 0$ e $x > B + \theta$ por (2)(b)(ii).

Se $x \leq C$, temos que $r(x, \theta) \leq 1$. Como existem valores de $r(x, \theta) > 1$ o supremo de $r(x|\theta)$ é atingido em $x > C$. Agora, por (2)(a) e (2)(c) e pelo fato de que f é densidade, existe uma constante $x_0 < C$ tal que $f(x_0) \leq \inf_{C \leq x \leq B} f(x) \leq f(x)$, para $C \leq x \leq B$, e $\theta \geq B - x_0$ e $C \leq x \leq B$, temos que $x - \theta \leq x_0$, $f(x - \theta) \leq \inf_{C \leq x \leq B} f(x) \leq f(x)$, e então $r(x|\theta) \leq 1$. O supremo de $r(x, \theta)$ deve portanto ser atingido em $x > B$.

Finalmente, se $\theta > B - x_0$ e $B < x < C + \theta$ o numerador em $r(x, \theta)$ é crescente e o denominador é decrescente. Assim, neste conjunto, $r(x, \theta)$ é função crescente. O supremo deve portanto ser atingido em $x \geq C + \theta$.

Parte 2: Mostraremos que existem $D > 0$ e $\theta^* > 0$ tais que se $\theta > \theta^*$ teremos que $k(\theta) \leq D[f(\theta)]^{-1}$. Para tanto, notemos da *Parte 1* que se $\theta > B - x_0$ então $k(\theta) = f(z(\theta))/f(\theta + z(\theta))$, onde $C \leq z(\theta) \leq B$. Tomando $f_1 = \sup_{C \leq x \leq B} f(x)$ temos que se $\theta > B - x_0$, $k(\theta) \leq f_1/f(\theta + z(\theta))$ e então, de (2)(b)(i), $f(\theta)/f(\theta + B) < 1 + \varepsilon$. Disto, basta tomar $D = f_1(1 + \varepsilon)$.

Parte 3: Nas partes 1 e 2 mostramos que $k(\theta)$ é atingido e majorado por $D[f(\theta)]^{-1}$. Portanto a condição (3) de Dawid estará satisfeita se a condição (3) de O'Hagan o estiver. \square

Uma característica interessante deste teorema é troca da condição (3)(a) de Dawid pela condição (3)(a) de O'Hagan.

Exemplo 3.3. Suponha que X dado Θ tenha distribuição Box e Tiao

$$f(x|\theta) = \frac{\exp\{-|x - \theta|^{\frac{2}{1+\beta}}\}}{2\Gamma\left(\frac{3+\beta}{2}\right)}$$

com $x \in \mathbb{R}$, $\beta > 1$. Suponha ainda que Θ tenha distribuição $N(0, 1)$. Mostraremos, utilizando as condições de O'Hagan que em caso de conflito a esperança a posteriori tenderá para a esperança a priori. Notemos primeiramente que

$$\frac{f(x - \theta)}{f(x)} = \exp\left\{-|x - \theta|^{\frac{2}{1+\beta}} + |x|^{\frac{2}{1+\beta}}\right\} \rightarrow 1$$

quando $|x| \rightarrow \infty$, estando satisfeita a primeira condição. Notemos agora que

- f é contínua
- $f(y)$ é decrescente para todo $y > 0$
- Para todo $y > 0$, temos que

$$\frac{d}{dy} \log f(y) = -\frac{2}{1+\beta} y^{\frac{1-\beta}{1+\beta}}$$

é crescente em y

- $f(y)$ é decrescente para todo $y \leq 0$

e portanto a segunda condição de O'Hagan está satisfeita. Por último, notemos que

$$\int_{\mathbb{R}} \exp\{-\theta^2/2 + |x - \theta|^{\frac{2}{1+\beta}}\} d\theta < \infty$$

estando satisfeita a terceira condição de O'Hagan. Portanto

$$E(\Theta|x) \rightarrow 0$$

quando $x \rightarrow \infty$.

3.3 O teorema de Pericchi e Sansó

As condições de Dawid e as de O'Hagan conduzem-nos, no caso de uma única observação, ao problema no qual a distribuição a posteriori para um parâmetro de locação converge para a priori quando a observação converge para o infinito. Em tais casos, a informação da observação é perdida, no sentido de que a verossimilhança converge para a densidade (imprópria) uniforme na reta. Entretanto, o resultado a seguir, enunciado por Pericchi e Sansó em [40], mostra que é possível que uma das fontes de informação convirja para outra densidade (também imprópria). Nesta seção, considere uma verossimilhança $g(x|\theta) = f(x - \theta)$ e uma densidade a priori $\pi(\theta)$. Enunciamos a seguir os resultados principais de [40]:

Teorema 3.3. *Seja $f(x - \theta)$ a verossimilhança de uma observação x , onde θ é um parâmetro de locação, e seja $\pi(\theta)$ a densidade a priori para θ tal que $\pi(\theta) > 0$ para todo $\theta \in \mathbb{R}$. Defina*

- $k(t) = \sup_x \pi(x - t)/\pi(x)$
- $h(t) = \lim_{x \rightarrow \infty} \pi(x - t)/\pi(x)$

onde assumimos que o limite existe para todo $t \in \mathbb{R}$. Então se para todo $t \in \mathbb{R}$,

1. $\int |t|k(t)f(t)dt < \infty$

2. $\int k(t)f(t)dt < \infty$, $f(t) > 0$, para todo $t \in \mathbb{R}$

temos que $\lim_{x \rightarrow \infty} p(x - \theta \in A|x) = Q(A)$ para qualquer A mensurável, onde Q é uma função de distribuição com densidade proporcional a $e^{\lambda t} f(t)$ para $\lambda = \log h(1) \geq 0$. Disto temos que $E(\Theta|x) - x$ é limitado para $x \in \mathbb{R}$.

Demonstração. Temos que

$$\begin{aligned} p(\theta - x \in A|x) &= \frac{\int_{A+x} \pi(\theta) f(x - \theta) d\theta}{\int_{\mathbb{R}} \pi(\theta) f(x - \theta) d\theta} \\ &= \frac{\int_A \{\pi(x - t)/\pi(x)\} f(t) dt}{\int_{\mathbb{R}} \{\pi(x - t)\pi(x)\} f(t) dt} \rightarrow \frac{\int_A h(t) f(t) dt}{\int_{\mathbb{R}} h(t) f(t) dt} = Q(A) \end{aligned}$$

por convergência dominada. O limite $h(t)$ é o mesmo dado no Lema 2.35 sendo portanto igual a $h(t) = e^{\lambda t}$ e como π é densidade, temos que $\lambda \geq 0$. Assim, Q é proporcional a $e^{\lambda t} f(t)$. Para mostrar que a esperança é limitada, observemos que

$$\begin{aligned} E(\Theta|x) - x &= x - \frac{\int_{A+x} (\theta - x) \pi(\theta) f(x - \theta) d\theta}{\int_{\mathbb{R}} \pi(\theta) f(x - \theta) d\theta} \\ &= \frac{\int_A \{\pi(x - t)/\pi(x)\} t f(t) dt}{\int_{\mathbb{R}} \{\pi(x - t)\pi(x)\} f(t) dt} \rightarrow \frac{\int_A t h(t) f(t) dt}{\int_{\mathbb{R}} h(t) f(t) dt} = E_Q(t) \end{aligned}$$

terminando a demonstração. □

Neste caso, temos que a influência a priori sobre a posteriori é limitada. Intuitivamente podemos dizer que toda a informação a priori sobre o *outlier* está condensada em $h(t) = e^{\lambda t}$, ocorrendo uma rejeição parcial da fonte informação conflitante.

Corolário 3.4. *Se F é simétrica em torno de zero, então*

$$\lim_{x \rightarrow \infty} E(\Theta|x) - x = 0$$

se e somente se $\lambda = 0$

Corolário 3.5. *Se a verossimilhança é normal com média θ e variância σ^2 então*

$$\lim_{x \rightarrow \infty} E(\Theta|x) - x = -\lambda \sigma^2$$

O Teorema 3.3 pode ser reformulado sem maiores dificuldades para o caso em que desejamos rejeitar a informação proveniente dos dados, de maneira análoga ao que foi feito com as condições de Dawid.

Exemplo 3.4. Primeiramente suponha que X dado $\Theta = \theta$ tenha distribuição $N(\theta, 1)$ e que Θ tenha distribuição exponencial dupla. Neste caso, temos que

- $k(t) = \sup_x \frac{\pi(x-t)}{\pi(x)} = \exp\{-|x-t| + |x|\} = e^t$
- $h(t) = \lim_{x \rightarrow \infty} \frac{\pi(x-t)}{\pi(x)} = e^t$

e as integrais

- $\int_{\mathbb{R}} \exp\{t - t^2/2\} dt$
- $\int_{\mathbb{R}} |t| \exp\{t - t^2/2\} dt$

são finitas e portanto são válidas as condições do Teorema 3.3. Além disso, pelo Corolário 3.5, temos que

$$\lim_{x \rightarrow \infty} E(\Theta|x) - x = -1$$

Agora, se supormos que X dado $\Theta = \theta$ tem distribuição normal $N(\theta, \sigma^2)$ e que Θ tem distribuição logística, teremos que $k(t) = h(t) = e^t$ e portanto poderemos realizar uma análise análoga à feita no caso anterior.

3.4 Credence

3.4.1 Definições e propriedades básicas

Criada por O'Hagan em [36], a *credence* é uma medida de cauda que pode auxiliar na resolução de um possível conflito de informação. Damos a definição a seguir:

Definição 3.6. Uma densidade f em \mathbb{R} tem credence c , denotado por $\text{cred}(f) = c$, se existirem $K \geq k > 0$ tais que para todo $x \in \mathbb{R}$

$$k \leq (1 + x^2)^{c/2} f(x) \leq K. \quad (3.2)$$

Dada uma variável aleatória X , dizemos que $\text{cred}(X) = c$ se a distribuição de X possuir uma densidade com credence c .

A definição anterior foi motivada pelo uso da distribuição t -Student em problemas de conflitos de informação. Notemos que uma t com d graus de liberdade tem credence $1 + d$. No entanto, uma densidade com credence não necessariamente satisfaz qualquer condição de regularidade, como monotonicidade e continuidade nas caudas. Notemos ainda que credence não é uma medida geral para o peso de caudas e que $\text{cred}(\cdot)$ não está definida para várias densidades. Se $\text{cred}(\cdot)$ está definido, então esta é única.

Lema 3.7. Para quaisquer $a \neq 0$ e b , existem $L \geq l > 0$ tais que para todo $x \in \mathbb{R}$

$$l \leq \frac{1 + (ax + b)^2}{1 + x^2} \leq L.$$

Demonstração. A demonstração deste lema será construída em três partes.

Parte 1: Mostraremos que para qualquer $a \neq 0$ existem $L_a \geq l_a > 0$ tais que

$$l_a \leq \frac{1 + (ax)^2}{1 + x^2} \leq L_a.$$

Seja

$$\sigma(x)_a = \frac{1 + (ax)^2}{1 + x^2}.$$

Então, temos que $\lim_{x \rightarrow \infty} \sigma(x)_a = \lim_{x \rightarrow -\infty} \sigma(x)_a = a^2$, que

$$\frac{d}{dx} \log \sigma(x)_a = \frac{2a^2 x}{1 + (ax)^2} - \frac{2x}{1 + x^2} \quad (3.3)$$

e que

$$\frac{d^2}{dx^2} \log \sigma(x)_a = \frac{2a^2(1 - a^2 x^2)}{[1 + (ax)^2]^2} - \frac{2(1 - x^2)}{[1 + x^2]^2}. \quad (3.4)$$

Fazendo (3.3) igual a zero, encontramos o ponto crítico $x = 0$. Substituindo este ponto em (3.4) temos que:

- o ponto crítico será ponto de máximo se $a \in (-1, 1)$. Neste caso $a^2 \leq \sigma(x)_a \leq 1$

- o ponto crítico será ponto de mínimo se $a \in (-\infty, -1) \cup (1, \infty)$. Neste caso $1 \leq \sigma(x)_a \leq a^2$
- o ponto crítico será ponto de inflexão se $a \in \{-1, 1\}$ e $\sigma(x)_a = 1$ para qualquer $x \in \mathbb{R}$

Faça $l_a = \min\{1, a^2\}$ e $L_a = \max\{1, a^2\}$. Então $l_a \leq \sigma(x)_a \leq L_a$, completando esta parte da demonstração.

Parte 2: Mostraremos que para qualquer b existem $L_b \geq l_b > 0$ tais que

$$l_b \leq \frac{1 + (x + b)^2}{1 + x^2} \leq L_b. \quad (3.5)$$

Tome

$$\sigma(x)_b = \frac{1 + (x + b)^2}{1 + x^2}.$$

Notemos que $\lim_{x \rightarrow \infty} \sigma(x)_b = \lim_{x \rightarrow -\infty} \sigma(x)_b = 1$, que

$$\frac{d}{dx} \log \sigma(x)_b = \frac{2(x + b)}{1 + (x + b)^2} - \frac{2x}{1 + x^2} \quad (3.6)$$

e que

$$\frac{d^2}{dx^2} \log \sigma(x)_b = \frac{2 - 2(x + b)^2}{[1 + (x + b)^2]^2} - \frac{2 - 2x^2}{(1 + x^2)^2}. \quad (3.7)$$

Fazendo (3.6) igual a zero, temos

$$(x - b)(1 + x^2) - x - x(x + b)^2 = 0 \Rightarrow -x^2 - bx + 1 = 0$$

assim, temos os seguintes pontos críticos

$$x' = \frac{b \pm \sqrt{b^2 + 4}}{-2}.$$

Para não carregar a notação, tomemos $\Delta = \sqrt{b^2 + 4}$. Aplicando em (3.7) os pontos críticos encontrados temos que

$$\frac{d^2}{dx^2} \log \sigma(x') \propto \frac{1 - (x' + b)^2}{[1 + (x' + b)^2]^2} - \frac{1 - x'^2}{(1 + x'^2)^2}.$$

Notemos que

$$\begin{aligned}
 1 - (x' + b)^2 &= 1 - \left(\frac{b \pm \Delta}{-2} + b \right)^2 = \frac{-b^2 \pm b\Delta}{2} \\
 (1 + (x' + b)^2)^2 &= \left(\frac{4 + b^2 + \Delta^2 \mp 2b\Delta}{4} \right)^2 = \left(\frac{\Delta^2 \mp b\Delta}{2} \right)^2 \\
 1 - x'^2 &= 1 - \left(\frac{b \pm \Delta}{-2} \right)^2 = \frac{-b^2 \mp b\Delta}{2} \\
 (1 + x'^2)^2 &= \left(\frac{4 + b^2 + \Delta^2 \pm 2b\Delta}{4} \right)^2 = \left(\frac{\Delta^2 \pm b\Delta}{2} \right)^2
 \end{aligned}$$

então temos que

$$\begin{aligned}
 \frac{d^2}{dx^2} \log \sigma(x') &\propto (\Delta \pm b)^2(-b^2 \pm b\Delta) - (\Delta \mp b)^2(-b^2 \mp b\Delta) \\
 &\propto \pm(b\Delta^3 - b^3\Delta) = \pm 4b\Delta
 \end{aligned}$$

portanto, para $b \neq 0$, teremos que um dos pontos críticos será de máximo e o outro será de mínimo, dependendo do sinal de b . As possibilidades são:

- Se $b < 0$, então $-(b + \Delta)/2$ será ponto de máximo e $-(b - \Delta)/2$ será ponto de mínimo. Neste caso, $L_b = (\Delta - b)/(\Delta + b)$ e $l_b = (\Delta + b)/(\Delta - b)$
- Se $b > 0$, então $-(b + \Delta)/2$ será ponto de mínimo e $-(b - \Delta)/2$ será ponto de máximo. Neste caso, $L_b = (\Delta + b)/(\Delta - b)$ e $l_b = (\Delta - b)/(\Delta + b)$
- Se $b = 0$, então ambos os pontos serão de inflexão e, neste caso, $l = L = 1$

Assim, concluímos esta parte da demonstração tomando

- $l_b = \min\{(\Delta + b)/(\Delta - b), (\Delta - b)/(\Delta + b)\}$
- $L_b = \max\{(\Delta + b)/(\Delta - b), (\Delta - b)/(\Delta + b)\}$.

Parte 3: Para concluirmos a demonstração, notemos que

$$\frac{1 + (ax + b)^2}{1 + x^2} = \frac{1 + (ax + b)^2}{1 + (ax)^2} \frac{1 + (ax)^2}{1 + x^2} = \frac{1 + (y + b)^2}{1 + y^2} \frac{1 + y^2}{1 + (y/a)^2}$$

onde $y = ax$. Sabemos que das partes 1 e 2 existem $l_{1/a}, l_b, L_{1/a}$ e L_b tais que

$$\frac{l_b}{L_{1/a}} \leq \frac{1 + (ax + b)^2}{1 + (y/a)^2} \leq \frac{L_b}{l_{1/a}}$$

assim, basta tomarmos $l = l_b/L_{1/a}$ e $L = L_b/l_{1/a}$. \square

Teorema 3.8 (Invariância). *Seja X uma variável aleatória com distribuição contínua. Se $\text{cred}(X) = c$, então para quaisquer a e b ,*

$$\text{cred}(aX + b) = c.$$

Demonstração. Seja f a densidade de X e seja g a densidade de $aX + b$. Então $g(x|a, b) = a^{-1}f(a^{-1}(x - b))$. Assim

$$\begin{aligned} (1 + x^2)^{c/2}g(x) &= \frac{1}{a}(1 + x^2)^{c/2}f\left(\frac{x - b}{a}\right) = \frac{1}{a}(1 + (ay + b)^2)^{c/2}f(y) \\ &= \frac{1}{a}\left(\frac{1 + (ay + b)^2}{1 + y^2}\right)^{c/2}(1 + y^2)^{c/2}f(y) \end{aligned}$$

onde $y = a^{-1}(x - b)$. Utilizando o Lema 3.7 e o fato de que f possui *credence*, temos que

$$\frac{kl^{c/2}}{a} \leq (1 + y^2)^{c/2}f(y) \leq \frac{KL^{c/2}}{a}$$

portanto, $\text{cred}(g) = c$. \square

O Lema 3.7 estabelece diretamente que a distribuição t-Student padrão com d graus de liberdade possui *credence* $d + 1$. De fato, supondo que f é uma densidade t-Student com parâmetros de locação e escala, temos que

$$\begin{aligned} \frac{h(d+1)}{\sigma} L^{-\frac{d+1}{2}} &\leq (1 + x^2)^{\frac{d+1}{2}} f(x|\mu, \sigma) \\ &= (1 + x^2)^{\frac{d+1}{2}} h(d+1) \frac{1}{\sigma} \left[1 + \left(\frac{x - \mu}{\sigma} \right)^2 \right]^{-\frac{d+1}{2}} \leq \frac{h(d+1)}{\sigma} l^{-\frac{d+1}{2}}. \end{aligned}$$

Em geral, *credence* independe de locação e escala.

Teorema 3.9 (Convolução). *Sejam X e Y variáveis aleatórias independentes com distribuições contínuas, tais que $\text{cred}(X) = c$ e $\text{cred}(Y) = c'$. Então*

$$\text{cred}(X + Y) = \min\{c, c'\}. \quad (3.8)$$

Demonstração. Sejam f a densidade de X e g a densidade de Y . Sejam $c^* = \min\{c, c'\}$ e seja $m(x)$ a densidade de $X + Y$, isto é, $m(x) = \int_Y f(x-y)g(y)dy$. Como X e Y possuem credence existem constantes $K_f \geq k_f > 0$ e $K_g \geq k_g > 0$ tais que

$$k_f \leq (1 + (x - y)^2)^{-c/2} f(x - y) \leq K_f$$

e que

$$k_g \leq (1 + y^2)^{-c'/2} g(y) \leq K_g.$$

Para o limitante inferior, notemos que

$$f(x - y)g(y) = f(x - y)g(y) \left\{ \frac{1 + (x - y)^2}{1 + (x - y)^2} \right\}^{c/2} \left\{ \frac{1 + y^2}{1 + y^2} \right\}^{c'/2}.$$

Tomemos $\varepsilon > 0$. Se supormos que $y \in [0, \varepsilon]$, teremos

$$f(x - y)g(y) \geq k_f k_g (1 + x^2)^{-c/2} (1 + \varepsilon^2)^{-c'/2} > 0.$$

Agora, se supormos que $y \in [x - \varepsilon, x]$, teremos

$$f(x - y)g(y) \geq k_f k_g (1 + x^2)^{-c/2} (1 + \varepsilon^2)^{-c'/2} > 0.$$

Portanto

$$(1 + x^2)^{c^*/2} \int_Y f(x - y)g(y)dy \geq k_m > 0$$

onde $k_m = k_f k_g (1 + \varepsilon^2)^{-c^*/2}$. Para o limitante superior, observemos que

$$\begin{aligned} m(x) &= \int_Y f(x - y)g(y)dy \\ &= \left(\int_{-\infty}^{x/2} + \int_{x/2}^{\infty} \right) f(x - y)g(y) \left\{ \frac{1 + (x - y)^2}{1 + (x - y)^2} \right\}^{c/2} \left\{ \frac{1 + y^2}{1 + y^2} \right\}^{c'/2} dy \\ &\leq \int_{-\infty}^{x/2} K_f g(y) (1 + (x - y)^2)^{-c/2} dy + \int_{x/2}^{\infty} K_g f(x - y) (1 + y^2)^{-c'/2} dy \\ &\leq K_f \left(1 + \frac{x^2}{4} \right)^{-c/2} \int_{-\infty}^{x/2} g(y) dy + K_g \left(1 + \frac{x^2}{4} \right)^{-c'/2} \int_{x/2}^{\infty} f(x - y) dy \\ &\leq (K_f + K_g) \left(1 + \frac{x^2}{4} \right)^{-c^*/2} = (4 + x^2)^{-c^*/2} \left\{ \frac{K_f + K_g}{2^{-c^*}} \right\} \leq (1 + x^2)^{-c^*/2} K_m \end{aligned}$$

onde $K_m = (K_f + K_g)2^{c^*}$. Assim, tomando ε suficientemente pequeno teremos que $K_m \geq k_m > 0$. \square

O conceito de *credence* é facilmente estendido para o caso em que temos várias fontes de informação. Suponhamos que X_1, \dots, X_n são independentemente distribuídas dado $\Theta = \theta$ e que a densidade de X_i seja $g_i(x_i - \theta)$. Com isto, temos o seguinte teorema.

Teorema 3.10 (Agrupamento). *Para $i = 1, \dots, m$, seja $\text{cred}(g_i) = c_i$. Seja*

$$f(t) = \prod_{i=1}^m g_i(x_i + t)$$

Então $\text{cred}(f) = \sum_{i=1}^m c_i$.

Demonstração. Suponhamos que o teorema é válido para $i = 1, \dots, m-1$. Temos então que existem constantes $K \geq k > 0$ e $K_m \geq k_m > 0$ tais que

$$k \leq (1 + t^2)^{\sum_{i=1}^{m-1} c_i/2} \prod_{i=1}^{m-1} g_i(x_i + t) \leq K$$

e que

$$k_m \leq (1 + t^2)^{c_m/2} g_m(x_m + t) \leq K_m.$$

Disto temos que

$$\begin{aligned} (1 + t^2)^{\sum_{i=1}^m c_i/2} f(t) &= (1 + t^2)^{\sum_{i=1}^m c_i/2} \prod_{i=1}^m g_i(x_i + t) \\ &= \left\{ (1 + t^2)^{\sum_{i=1}^{m-1} c_i/2} \prod_{i=1}^{m-1} g_i(x_i + t) \right\} (1 + t^2)^{c_m/2} g_m(x_m + t) \end{aligned}$$

portanto temos que

$$kk_m \leq (1 + t^2)^{\sum_{i=1}^m c_i/2} f(t) \leq KK_m$$

e portanto $\text{cred}(f) = \sum_{i=1}^m c_i/2$. \square

Definição 3.11 (Credence Infinita). *Dizemos que uma densidade possui credence infinita se para quaisquer c e $k > 0$ existir um A tal que para todo $|x| > A$ tenhamos*

$$(1 + x^2)^{c/2} f(x) < k. \quad (3.9)$$

Neste caso, podemos considerar a distribuição normal e a família Box-Tiao com densidades proporcionais a $\exp(-|x|^b)$ como tendo *credence* infinita.

Veremos posteriormente, na Seção 4.4, que a medida *credence* está relacionada com a classe *ERV*.

3.4.2 O uso de credence em conflitos de informação

Utilizando-se a *credence* das fontes de informação, O'Hagan [36] criou um novo conceito de rejeição de *outliers* para o caso onde uma das fontes (ou ambas) pertencem à família locação. Neste conceito, temos que uma das fontes domina a outra, no sentido de que a posteriori está limitada por múltiplos positivos da fonte dominante para $|x|$ suficientemente grande. Neste caso dizemos que um conflito é resolvido a favor da fonte com maior *credence* (como será mostrado mais adiante no Teorema 3.13). Nesta seção discutiremos este conceito e apresentaremos seus principais resultados.

Consideremos que foi realizada uma observação de X com densidade $g(x|\theta) = f(x-\theta)$. Suponhamos que a densidade para Θ é dada por $\pi(\theta)$ e que a densidade a posteriori é dada por $p(\theta|x)$. Sejam $\text{cred}(f) = c'$, $\text{cred}(\pi) = c$ e $\text{cred}(m) = c^*$, onde $m(x) = \int_{\Theta} f(x-\theta)\pi(\theta)d\theta$. Assim, sabemos que existem constantes $K_f \geq k_f > 0$ tais que

$$k_f \leq (1 + (x - \theta)^2)^{c'/2} f(x - \theta) \leq K_f,$$

existem constantes $K_\pi \geq k_\pi > 0$ tais que

$$k_\pi \leq (1 + \theta^2)^{c/2} \pi(\theta) \leq K_\pi,$$

e existem constantes $K_m \geq k_m > 0$ tais que

$$k_m \leq (1 + x^2)^{c^*/2} m(x) \leq K_m$$

onde $c^* = \min\{c, c'\}$.

Teorema 3.12 (Credence da Posteriori). *Para todo x , $\text{cred}(p) = c + c'$.*

Demonstração. Como $m(x)$ não depende de θ , notemos primeiramente que

$$\begin{aligned} (1 + \theta^2)^{\frac{c+c'}{2}} p(\theta|x) &= (1 + \theta^2)^{\frac{c+c'}{2}} \frac{f(x-\theta)\pi(\theta)}{m(x)} \\ &= (1 + \theta^2)^{\frac{c+c'}{2}} \frac{f(x-\theta)\pi(\theta)}{m(x)} \left\{ \frac{1 + (x-\theta)^2}{1 + (x-\theta)^2} \right\}^{c'/2} \left\{ \frac{1 + \theta^2}{1 + \theta^2} \right\}^{c/2}. \end{aligned}$$

Portanto, utilizando o fato de que f e π possuem *credence* temos que

$$(1 + \theta^2)^{\frac{c+c'}{2}} p(\theta|x) \leq \frac{K_f K_\pi}{m(x)} \left\{ \frac{1 + (x - \theta)^2}{1 + \theta^2} \right\}^{-c'/2} \leq \frac{K_f K_\pi}{m(x)} l^{-c'/2}$$

e de maneira análoga,

$$(1 + \theta^2)^{\frac{c+c'}{2}} p(\theta|x) \geq \frac{k_f k_\pi}{m(x)} L^{-c'/2}$$

onde L e l são dados pelo Lema 3.7. □

Teorema 3.13 (Rejeição). *Se $c > c'$ então:*

1. *para quaisquer $d > 0$ e $\varepsilon > 0$ dados, existe um x_0 tal que para todo $|x| > x_0$ e para todo $|\theta| < d$,*

$$\frac{k_f}{K_m} (1 - \varepsilon) \pi(\theta) \leq p(\theta|x) \leq \frac{K_f}{k_m} (1 + \varepsilon) \pi(\theta) \quad (3.10)$$

2. *para toda $t : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ tal que $t(x) \rightarrow \infty$ quando $x \rightarrow \infty$,*

$$\lim_{|x| \rightarrow \infty} P(|\Theta| > t(|x|)|x) = 0 \quad (3.11)$$

Demonstração. Notemos que

$$\frac{p(\theta|x)}{\pi(\theta)} = \frac{f(x - \theta)(1 + (x - \theta)^2)^{c'/2}}{m(x)(1 + x^2)^{c'/2}} \left\{ \frac{1 + (x - \theta)^2}{1 + x^2} \right\}^{-c'/2}$$

Como f e m possuem *credence*, temos que

$$\frac{p(\theta|x)}{\pi(\theta)} \leq \frac{K_f}{k_m} \left\{ \frac{1 + (x - \theta)^2}{1 + x^2} \right\}^{-c'/2}$$

e que

$$\frac{p(\theta|x)}{\pi(\theta)} \geq \frac{k_f}{K_m} \left\{ \frac{1 + (x - \theta)^2}{1 + x^2} \right\}^{-c'/2}$$

Notemos agora que o termo entre chaves nas últimas duas inequações converge para 1 quando $|x| \rightarrow \infty$, logo para qualquer $\varepsilon > 0$ existe um x_0 tal que para todo $|x| > x_0$

$$\frac{k_f}{K_m} (1 - \varepsilon) \pi(\theta) \leq p(\theta|x) \leq \frac{K_f}{k_m} (1 + \varepsilon) \pi(\theta)$$

concluindo a demonstração de (1). Para demonstrar (2) consideraremos, por simetria, somente o caso onde $x > 0$. Temos que

$$\begin{aligned}
P(\Theta < -t(x)|x) &= \int_{-\infty}^{-t(x)} \frac{f(x-\theta)\pi(\theta)}{m(x)} d\theta \\
&= \int_{-\infty}^{-t(x)} \frac{f(x-\theta)\pi(\theta)}{m(x)} \left\{ \frac{1+(x-\theta)^2}{1+(x-\theta)^2} \right\}^{c'/2} \left\{ \frac{1+x^2}{1+x^2} \right\}^{c'/2} d\theta \\
&\leq \frac{K_f}{k_m} (1+x^2)^{c'/2} \int_{-\infty}^{-t(x)} \pi(\theta) (1+(x-\theta)^2)^{-c'/2} d\theta \\
&\leq \frac{K_f}{k_m} \left\{ \frac{1+(x+t(x))^2}{1+x^2} \right\}^{-c'/2} \int_{-\infty}^{-t(x)} \pi(\theta) d\theta \rightarrow 0
\end{aligned}$$

quando $x \rightarrow \infty$. Agora, consideremos que $a > 0$. Temos então que

$$\begin{aligned}
P(\Theta > ax|x) &= \int_{ax}^{\infty} \frac{f(x-\theta)\pi(\theta)}{m(x)} d\theta = \int_0^{\infty} \frac{f(x-ax-\mu)\pi(ax+\mu)}{m(x)} d\mu \\
&\leq \frac{K_\pi}{k_m} (1+x^2)^{-c'/2} (1+a^2x^2)^{-c'/2} \int_0^{\infty} f(x-ax-\mu) d\mu \\
&\leq \frac{K_\pi}{k_m} (1+x^2)^{-c'/2} (1+a^2x^2)^{-c'/2} \rightarrow 0
\end{aligned}$$

quando $x \rightarrow \infty$. Esta demonstração é suficiente para completar a prova no caso em que $t(x) \rightarrow \infty$ mais rápido que ax para algum a . Em caso contrário, calculemos

$$\begin{aligned}
P(t(x) < \Theta \leq ax|x) &= \int_{t(x)}^{ax} \frac{f(x-\theta)\pi(\theta)}{m(x)} d\theta \\
&\leq \frac{K_f}{k_m} \left\{ \frac{1+(1-a)^2x^2}{1+x^2} \right\}^{-c'/2} \int_{t(x)}^{ax} \pi(\theta) d\theta \rightarrow 0
\end{aligned}$$

pois o termo antes da integral converge para $(1-a)^2$ e a integral converge para 0 quando $t(x) \rightarrow \infty$, completando a demonstração. \square

Antes de interpretar o teorema acima, convém discutir o caso onde f e g são densidades normais. Neste caso a posteriori também será normal, sua variância independará de x e sua média será ax , onde $0 < a < 1$. Quando $|x|$ cresce, notemos primeiramente que nas vizinhanças de x e 0 a probabilidade a posteriori tende a 0, ignorando o fato de que são nestas duas vizinhanças que as fontes de informação sugerem onde Θ deveria estar. Em vez disso, a probabilidade a posteriori se concentra em torno de ax . Em contraste,

no fenômeno de rejeição de *outliers*, quando $|x| \rightarrow \infty$ uma fonte de informação domina a outra e a probabilidade a posteriori se concentra em torno da informação dominante, tendendo a zero em qualquer outra parte. O resultado (1) do teorema acima nos diz que as probabilidades a posteriori para qualquer θ fixados estão limitadas por múltiplos positivos da priori. Assim, a fonte de informação que possuir a maior *credence* dominará a outra. O resultado (2) diz que as probabilidades a posteriori aplicadas em uma função de x convergem para zero em qualquer parte quando x tende ao infinito. Em particular, a sugestão da verossimilhança de que Θ está próximo da vizinhança de x é ignorada no limite. No mais, não há nenhuma probabilidade a posteriori na vizinhança de ax , o que contrasta com o caso normal.

O teorema ainda é válido quando uma das fontes possui *credence* finita e a outra não, mas ele não necessariamente será verdadeiro se ambas possuírem *credence* infinita (temos que ter uma *credence maior que a outra*). Como já discutimos, a rejeição não ocorre no caso de distribuições normais. É interessante notarmos que a rejeição ocorre nas distribuições Box-Tiao ($f(y) \propto \exp(-|y|^{2/(\beta+1)})$) com $\beta > 1$, mesmo esta tendo *credence* infinita.

3.5 Resolvendo conflitos de informação em \mathcal{RV}

Muito se tem discutido na literatura sobre condições suficientes para resolver conflitos de informação quando uma das fontes assume uma estrutura de locação. Contudo, são poucos os resultados que se preocuparam com outras estruturas. Nesta seção apresentaremos os resultados obtidos por Andrade e O'Hagan [2] supondo que pelo menos uma das fontes de informação possui distribuição de variação regular, assumindo ou uma estrutura de escala ou uma de locação.

3.5.1 Estrutura parâmetro de escala

Rejeitando a informação dos dados

Considere um modelo escala puro, ou seja, $f(x|\theta) = (1/\theta)h(x/\theta)$, e apenas uma observação. O seguinte resultado dá condições para que um conflito possa ser resolvido em favor da priori.

Teorema 3.14. *Seja $X|\theta \sim g$ tal que g é da forma $g(x|\theta) = (1/\theta)f(x/\theta)$ e $\Theta \sim \pi$. Se para alguns $0 < \delta < p$*

1. $f \in \mathcal{RV}_{-p-1}$
2. $\int_{\Theta} \theta^{p+\delta} \pi(\theta) d\theta < \infty$

Então

$$p(\theta|x) \rightarrow \frac{\theta^p \pi(\theta)}{\int_{\Theta} \theta^p \pi(\theta) d\theta}$$

quando $x \rightarrow \infty$.

Demonstração. A distribuição a posteriori é dada por

$$p(\theta|x) = \frac{g(x|\theta)\pi(\theta)}{\int_{\Theta} g(x|\theta)\pi(\theta)d\theta} = \frac{(1/\theta)f(x/\theta)\pi(\theta)}{\int_{\Theta} (1/\theta)f(x/\theta)\pi(\theta)d\theta}$$

Dividindo o numerador e o denominador por $f(x)$ e tomando o limite de $p(\theta|x)$ quando $x \rightarrow \infty$, temos

$$\lim_{x \rightarrow \infty} p(\theta|x) = \frac{\lim_{x \rightarrow \infty} (1/\theta)f(x/\theta)\pi(\theta)/f(x)}{\lim_{x \rightarrow \infty} \int_{\Theta} (1/\theta)f(x/\theta)\pi(\theta)d\theta/f(x)} \quad (3.12)$$

supondo que os dois limites existam e o denominador seja não nulo. Notemos que o numerador em (3.12) é $\theta^p \pi(\theta)$. Para encontrar o limite do denominador, observemos do Lema 2.10 que para qualquer $c > 1$ e $\delta > 0$ e para todo x suficientemente grande

$$\frac{1}{\theta} \frac{f(x/\theta)}{f(x)} \pi(\theta) \leq \frac{1}{\theta} \pi(\theta) \max\{\theta^{p+1-\delta}, \theta^{p+1+\delta}\}.$$

Agora, notemos que

$$\begin{aligned} \int_0^{\infty} \frac{1}{\theta} \max\{\theta^{p+1-\delta}, \theta^{p+1+\delta}\} \pi(\theta) d\theta &= \int_0^1 \max\{\theta^{p-\delta}, \theta^{p+\delta}\} \pi(\theta) d\theta \\ &+ \int_1^{\infty} \max\{\theta^{p-\delta}, \theta^{p+\delta}\} \pi(\theta) d\theta \\ &= \int_0^1 \theta^{p-\delta} \pi(\theta) d\theta + \int_1^{\infty} \theta^{p+\delta} \pi(\theta) d\theta \\ &\leq \int_0^1 \pi(\theta) d\theta + \int_1^{\infty} \theta^{p+\delta} p(\theta) d\theta \end{aligned}$$

que é finito pela condição 2. Disto, podemos utilizar o teorema da convergência dominada na integral em (3.12) e obter

$$p(\theta|x) \rightarrow \frac{\theta^p \pi(\theta)}{\int_{\Theta} \theta^p \pi(\theta) d\theta}$$

quando $x \rightarrow \infty$

□

O teorema acima nos diz que se escolhermos uma distribuição de variação regular para os dados e uma distribuição com a cauda mais leve para a distribuição a priori, a distribuição a posteriori se tornará independente do *outlier* quando este estiver suficientemente afastado da distribuição a priori. Notemos que, diferente dos resultados obtidos por Dawid [13] e O'Hagan[33], não existe uma completa rejeição do *outlier*. Apesar da distribuição a posteriori limite não depender de x , esta é uma combinação entre θ^p e a distribuição a priori. Pode-se pensar em θ^p como uma distribuição imprópria para os dados que carrega pouca informação sobre estes, ou mais formalmente podemos dizer que os dados se tornam pouco informativos. Neste caso, dizemos que houve uma rejeição parcial do *outlier*.

Várias observações

O teorema 3.14 pode ser generalizado para o caso onde temos várias observações. Neste caso podemos ter um ou mais *outliers* na amostra. Para nossos propósitos considere que o grupo de *outliers* está concentrado em torno de um centro afastado dos demais dados. Isto é equivalente a pensar que o i -ésimo *outlier* é da forma $x_i = x + \varepsilon_i$, para algum $|\varepsilon_i|$ fixado.

Para mostrar que o teorema 3.14 se aplica para várias observações, suponhamos que x_1, \dots, x_n é amostra aleatória independente distribuída de acordo com $g_i(x_i|\theta) = (1/\theta) f_i(x_i/\theta)$, com $i = \{1, \dots, n\}$. Assim, teremos que a distribuição conjunta dos dados será $g(\mathbf{x}|\theta) = (1/\theta)^n \times \prod_{i=1}^n g(x_i|\theta)$. Suponhamos também que θ tenha distribuição $\pi(\theta)$. Suponhamos ainda que as observações x_j com $j \in I$ são *outliers* onde $I \cup I^c = \{1, \dots, n\}$. Assim, podemos escrever a verossimilhança como sendo $g(\mathbf{x}|x) = (1/\theta)^n \times \prod_{j \in I} f((x + \varepsilon_j)/\theta) \times \prod_{j \in I^c} f(x_j/\theta) \times \pi(\theta)$. Supondo que cada $f_i \in \mathcal{RV}_{-p_i}$, temos pela Proposição 2.11 que $\prod_{j \in I} f((x + \varepsilon_j)/\theta) \in \mathcal{RV}_{-\sum_i p_i}$. Assim, pelo teorema 3.14, teremos que se estiverem satisfeitas as condições

1. $\prod_{j \in I} f(x + \varepsilon_j/\theta) \in \mathcal{RV}_{-\sum_{j \in I} p_j - 1}$

$$2. \int_{\Theta} \theta^{p+\delta} \times \prod_{j \in I^c} g(x_j|\theta) \times \pi(\theta) d\theta < \infty \text{ para algum } \delta > 0$$

então temos que

$$p(\theta|\mathbf{x}) \rightarrow \frac{\theta^{\sum_{j \in I} p_j} \times \prod_{j \in I^c} g(x_j|\theta) \times \pi(\theta) d\theta}{\int_{\Theta} \theta^{\sum_{j \in I} p_j} \times \prod_{j \in I^c} g(x_j|\theta) \times \pi(\theta) d\theta}$$

quando $x \rightarrow \infty$.

3.5.2 Rejeitando a informação a priori

Os resultados vistos anteriormente nesta seção lidam com o caso no qual estamos interessados em resolver conflitos a favor da distribuição a priori. Consideraremos agora o caso em que temos condições suficientes para rejeitar a informação a priori a favor da distribuição dos dados em caso de conflito entre as fontes. Neste caso, acreditamos que, de alguma forma, a informação disponível para a formulação da distribuição a priori pode produzir uma densidade que discorde dos dados.

Considere a transformação de Θ para $\Phi = |Y|/\Theta$. Utilizando o método do Jacobiano, obtemos que

$$\begin{aligned} p_{\Phi}(\phi|y) &= \frac{(1/\phi)f(y\phi/|y|)\pi(|y|/\phi)}{\int_{\Phi} (1/\phi)f(y\phi/|y|)\pi(|y|/\phi)d\phi} \\ &= \frac{(1/\phi)f(\text{sgn}(y)\phi)\pi(|y|/\phi)}{\int_{\Phi} (1/\phi)f(\text{sgn}(y)\phi)\pi(|y|/\phi)d\phi} \end{aligned}$$

onde $\text{sgn}(y) = 1$ se $y \geq 0$ e $\text{sgn}(y) = -1$ se $y < 0$.

Notemos que $p_{\Phi}(\theta|y)$ é a distribuição a posteriori de um parâmetro de escala, onde $f(\text{sgn}(y)\phi)$ e $q(y|\phi) = (1/\phi) \times \pi(|y|/\phi)$ fazem o papel de priori e verossimilhança respectivamente. Assim, supondo que f tem a cauda mais leve que a de π e $\pi \in \mathcal{RV}_{-(\rho+1)}$ ($\rho > 0$), podemos aplicar o teorema 3.14 e obter

$$\lim_{|y| \rightarrow \infty} p_{\Phi}(\phi|y) = \frac{\phi^{\rho} f(\text{sgn}(y)\phi)}{\int_{\Phi} \phi^{\rho} f(\text{sgn}(y)\phi) d\phi}$$

Em outras palavras, se escolhermos uma distribuição de variação regular para a priori e uma distribuição com a cauda mais leve para os dados, a distribuição a posteriori se

tornará aproximadamente independente da informação a priori se esta estiver distante dos dados, e então a distribuição a posteriori será baseada principalmente na verossimilhança. Entretanto, não ocorrerá uma completa rejeição da informação a priori. Notemos que as estimativas a posteriori serão baseadas na informação proveniente dos dados.

3.5.3 Estrutura locação

Mostraremos aqui que as condições, no caso locação, estabelecidas por Dawid [13] e O'Hagan [33] podem ser simplificadas pela utilização de distribuições de variação regular. Para o caso escala, podemos garantir a existência da distribuição limite da posteriori lidando com o peso das caudas das distribuições envolvidas. Isto não é possível no caso locação, uma vez que a integração é feita sobre todo \mathbb{R} e não temos como garantir o bom comportamento da distribuição dos dados. Para tanto, Andrade e O'Hagan [2] combinaram as distribuições de variação regular com algumas suposições comportamentais para estabelecer novas condições para o caso locação. Estas condições são apresentadas no teorema abaixo.

Teorema 3.15. *Seja $X|\theta$ com densidade $f(x - \theta)$ e $\Theta \sim \pi$ tal que*

1. $f \in \mathcal{RV}_p$ ($p > 0$)
2. $f > 0$ é contínua em \mathbb{R}
3. existem $C_1 \leq C_2 \geq C_1$ tais que $f(x)$ é decrescente para $x \geq C_2$ e crescente em $x \leq C_1$. Também, $d \log f(x)/dx$ existe e é crescente para $x \geq C_2$.
4. $\pi \notin \mathcal{RV}$ ou $\pi \in \mathcal{RV}_{-c}$ com $c > p + 1$

então

$$p(\theta|x) \rightarrow \pi(\theta).$$

Demonstração. Ver Andrade e O'Hagan [2]. □

Basicamente, temos que as condições do Teorema 3.15 implicam nas condições de O'Hagan. Diferente do caso escala, aqui temos uma rejeição completa do *outlier* e o conflito é resolvido a favor da priori.

Capítulo 4

Discussões e Resultados

Neste capítulo apresentamos os nossos resultados obtidos ao longo deste estudo no que se refere a relação entre distribuições de cauda pesada e conflitos de informação em inferência bayesiana. Na Seção 4.1 discutimos sobre a relação do teorema de Dawid e as distribuições de cauda longa. Na Seção 4.2 discutimos a relação entre o teorema de Pericchi e Sansó entre as distribuições em \mathcal{G} , além da sua relação com o teorema de Dawid. Na Seção 4.3 damos condições suficientes para que um conflito entre as fontes de informação seja resolvido quando utilizamos distribuições pertencentes à classe SD . Na Seção 4.4 caracterizamos a medida *credence*, mostrando que se uma distribuição tem *credence* então esta pertence à ERV . Por último, na Seção 4.5 mostramos como a medida *credence* pode ser utilizada para resolver conflitos assumindo uma estrutura de parâmetro de escala.

4.1 O Teorema de Dawid e distribuições em \mathcal{L}

Como foi visto anteriormente, o teorema de Dawid foi um dos primeiros envolvidos na resolução de conflitos de informação. Nesta seção discutiremos a sua relação com as distribuições de cauda longa.

Proposição 4.1. *Se $f \in \mathcal{L}$ então f satisfaz a condição (1) de Dawid.*

Demonstração. Suponhamos que $f \in \mathcal{L}$. Então para $\varepsilon > 0$ fixo, existe x_0 tal que para



todo $x > x_0$

$$|f(x - y) - f(x)| < \varepsilon f(x) \quad (4.1)$$

para todo $y \in \mathbb{R}$. Seja $y' \in \mathbb{R}$. Tomando $y < y'$ e $x' = x - y$ temos que

$$|f(x') - f(x)| < \varepsilon f(x), \quad \forall x > x_0 \quad (4.2)$$

onde x' é tal que $|x - x'| < y < y'$, o que implica na condição (1). \square

Com isto, temos um perfil do tipo de distribuição para a qual vale o teorema de Dawid, uma vez que as densidades em \mathcal{L} satisfazem a condição (1) e portanto estas são bons candidatos a obedecerem às demais condições.

De posse deste resultado, seria bastante útil se a condição (3)(a) de Dawid fosse válida para qualquer distribuição em \mathcal{L} . Infelizmente isto não é verdade como podemos ver no seguinte exemplo.

Exemplo 4.1. Suponhamos que $g(x|\theta) = (x - \theta)^{-2} I_{(\theta+1, \infty)}(x)$ e que $\pi(\theta) = \theta^{-2} I_{(1, \infty)}(\theta)$ sejam respectivamente verossimilhança e priori para um determinado problema. Notemos que, para todo θ ,

$$\lim_{x \rightarrow \infty} \frac{f(x - \theta)}{f(x)} = \lim_{x \rightarrow \infty} \left(\frac{x}{x - \theta} \right)^2 = 1$$

para todo $y \in \mathbb{R}$, portanto temos que $f \in \mathcal{L}$. Agora, notemos que

$$k(\theta) = \sup_x \frac{f(x - \theta)}{f(x)} = \sup_x \left(\frac{x}{x - \theta} \right)^2 I_{(\theta+1, \infty)}(x) = (\theta + 1)^2.$$

Assim,

$$\int_{\Theta} k(\theta) \pi(\theta) d\theta = \int_1^{x-1} \left(\frac{\theta + 1}{\theta} \right)^2 d\theta = x - 1 - \frac{1}{x-1} + 2 \log(x-1) \rightarrow \infty$$

quando $x \rightarrow \infty$. Portanto não satisfaz a condição (3)(a) de Dawid.

O exemplo acima nos mostra que nem todas as distribuições em \mathcal{L} satisfazem a condição (3)(a) de Dawid. O teorema seguinte ilustra uma situação especial na qual, para densidades em \mathcal{SD} , o teorema de Dawid não pode ser empregado.

Teorema 4.2. *Suponhamos que $g(x|\theta) = f(x - \theta)$ e $f(\theta)$ sejam respectivamente verossimilhança e priori para um problema, com $f \in \mathcal{SD}$. Então temos que a condição (3)(a) do teorema de Dawid não é válida.*

Demonstração. Suponhamos que $\int_{\Theta} k(\theta)f(\theta)d\theta$ seja finita. Neste caso, pelo teorema da convergência dominada, teremos que

$$\begin{aligned}\lim_{x \rightarrow \infty} \frac{f^{(2)}(x)}{f(x)} &= \lim_{x \rightarrow \infty} \int_{\Theta} \frac{f(x-\theta)}{f(x)} f(\theta) d\theta = \int_{\Theta} \lim_{x \rightarrow \infty} \left\{ \frac{f(x-\theta)}{f(x)} \right\} f(\theta) d\theta \\ &= \int_{\Theta} f(\theta) d\theta = 1 \neq 2\end{aligned}$$

o que contraria a hipótese de que $f \in \mathcal{S}$. Portanto $k(\theta)$ não é integrável em θ . □

Com isto, podemos ver que o teorema de Dawid não se aplica naturalmente a qualquer distribuição em \mathcal{L} .

4.2 Uma nota sobre o teorema de Pericchi e Sansó

Na Seção 3.3 apresentamos os resultados em Pericchi e Sansó [40]. Nesta seção discutiremos a relação entre o Teorema 3.3 e a classe \mathcal{G} bem como sua relação com o teorema de Dawid.

Primeiramente, notemos que uma das exigências do Teorema 3.3 é que

$$\lim_{x \rightarrow \infty} \frac{\pi(x-t)}{\pi(x)} = \exp\{\lambda t\}.$$

Já vimos, pelo Lema 2.35, que se $\bar{F}(x-t)/\bar{F}(x) \rightarrow \phi(t)$ quando $x \rightarrow \infty$, então este limite é da forma $\exp\{\lambda t\}$. Notemos entretanto que não utilizamos nenhuma propriedade de \bar{F} para a demonstração do lema. De fato, podemos demonstrar que o lema também é verdadeiro trocando \bar{F} por f . Damos então a seguinte proposição.

Proposição 4.3. *Se $\pi(x-t)/\pi(x) \rightarrow \exp\{\lambda t\}$ quando $x \rightarrow \infty$, então $F_{\pi} \in \mathcal{G}$, onde F_{π} é a função de distribuição de Θ .*

Demonstração. Sabemos que, dado $\varepsilon > 0$ existe x_0 tal que para todo $x > x_0$,

$$\begin{aligned}\left| \frac{\pi(x-t)}{\pi(x)} - e^{\lambda t} \right| \leq \varepsilon &\Rightarrow |\pi(x-t) - e^{\lambda t} \pi(x)| \leq \pi(x) \varepsilon \\ &\Rightarrow |\bar{F}_{\pi}(x-t) - e^{\lambda t} \bar{F}_{\pi}(x)| \leq \bar{F}_{\pi(x)} \varepsilon\end{aligned}$$

portanto

$$\frac{\bar{F}_\pi(x-t)}{\bar{F}_\pi(x)} \rightarrow e^{\lambda t}$$

quando $x \rightarrow \infty$, implicando em $F_\pi \in \mathcal{G}$. □

Assim, notemos que esta condição do teorema de Pericchi e Sansó tem forte relação com distribuições propensas a *outliers*, em especial com a classe \mathcal{G} . Sabemos que a classe de distribuições de cauda longa está contida em \mathcal{G} . Assim, um corolário interessante surge quando $\pi \in \mathcal{L}$.

Corolário 4.4. *Suponha satisfeitas as condições do Teorema 3.3. Suponha também que $\pi \in \mathcal{L}$. Então as condições do Teorema 3.3 implicam no teorema de Dawid (Teorema 3.1).*

Demonstração. Basta fazer $\lambda = 1$ na demonstração do Teorema 3.3. □

Assim, a condição (2) do teorema de Dawid será desnecessária sempre que $\pi \in \mathcal{L}$. Contudo, não podemos afirmar que o teorema de Pericchi e Sansó implica no teorema de Dawid. Sabemos que a classe de densidades que atendem a condição (1) de Dawid contém \mathcal{L} mas não temos certeza se todas essas densidades estão em \mathcal{G} .

4.3 Lidando com conflitos utilizando densidades em \mathcal{SD}

Vimos anteriormente que a classe \mathcal{SD} é uma vasta classe bastante utilizada em diversos ramos da estatística. Neste seção mostraremos como, sob certas condições, podemos utilizar distribuições nesta classe para resolver conflitos entre as fontes de informação, supondo uma estrutura de locação.

4.3.1 Rejeitando a informação dos dados para uma única observação

A seguir enunciamos o principal resultado desta seção.

Teorema 4.5. *Suponha que para um determinado problema temos que $g(x|\theta) = f(x - \theta)$, $\pi(\theta)$ e $p(\theta|x)$ sejam, respectivamente, verossimilhança, priori e posteriori, onde f e π possuem formas distintas. Suponha ainda que*

1. $f \in \mathcal{SD}$
2. $A(\pi) \subseteq A(f)$, onde $A(\cdot)$ é o suporte da distribuição.
3. $1/2 \leq \sup_{\theta} \pi(\theta)/f(\theta) < \infty$.

Então temos que

$$\lim_{x \rightarrow \infty} p(\theta|x) = \pi(\theta) \quad (4.3)$$

Demonstração. Primeiramente, suponha que ambas as fontes de informação possuem o mesmo suporte. Sabemos que

$$p(\theta|x) = \frac{g(x|\theta)\pi(\theta)}{\int_{\Theta} g(x|\theta)\pi(\theta)d\theta} = \frac{f(x - \theta)\pi(\theta)}{\int_{\Theta} f(x - \theta)\pi(\theta)d\theta}.$$

Dividindo ambos numerador e denominador por $f(x)$ teremos

$$\lim_{x \rightarrow \infty} p(\theta|x) = \frac{\lim_{x \rightarrow \infty} f(x - \theta)\pi(\theta)/f(x)}{\lim_{x \rightarrow \infty} \int_{\Theta} f(x - \theta)\pi(\theta)d\theta/f(x)}$$

desde que os dois limites acima existam e que o denominador seja não nulo. Para o numerador observemos que, pelo fato de $f \in \mathcal{L}$,

$$\lim_{x \rightarrow \infty} \frac{f(x - \theta)}{f(x)} \pi(\theta) = \pi(\theta)$$

restando apenas mostrar que o limite do denominador existe e é não nulo. Notemos que pela condição (2) do teorema temos que $\sup_{\theta} \pi(\theta)/f(\theta) = M$, com $1/2 \leq M < \infty$. Portanto, pela suposição (1) temos que

$$\int_{\Theta} \frac{f(x - \theta)\pi(\theta)}{f(x)} d\theta = \int_{\Theta} \frac{f(x - \theta)f(\theta)}{f(x)} \frac{\pi(\theta)}{f(\theta)} d\theta < M \int_{\Theta} \frac{f(x - \theta)f(\theta)}{f(x)} d\theta \rightarrow 2M$$

quando $x \rightarrow \infty$. Logo, pelo teorema da convergência dominada, temos que

$$\lim_{x \rightarrow \infty} \int_{\Theta} \frac{f(x - \theta)}{f(x)} \pi(\theta) d\theta = \int_{\Theta} \lim_{x \rightarrow \infty} \frac{f(x - \theta)}{f(x)} \pi(\theta) d\theta = \int_{\Theta} \pi(\theta) d\theta = 1$$

e portanto

$$p(\theta|x) \rightarrow \pi(\theta)$$

quando $x \rightarrow \infty$. Agora suponha que $A(\pi) \subset A(f)$. Teremos então que

$$p(\theta|x) = \frac{f(x-\theta)\pi(\theta)}{\int_{A(\pi)} f(x-\theta)\pi(\theta)d\theta}.$$

Notemos agora que

$$\begin{aligned} \int_{A(\pi)} \frac{f(x-\theta)}{f(x)} \pi(\theta) d\theta &\leq M \int_{A(\pi)} \frac{f(x-\theta)}{f(x)} f(\theta) d\theta \\ &\leq \int_{A(f)} \frac{f(x-\theta)}{f(x)} f(\theta) d\theta \rightarrow 2M \end{aligned}$$

logo podemos utilizar o teorema da convergência dominada e o resultado segue como visto no caso anterior. \square

O resultado acima nos dá condições suficientes para que um conflito seja resolvido a favor da priori. A condição (3) pede que o supremo da razão entre a priori e a verossimilhança, sob o mesmo suporte, seja finito. Isto é bastante razoável se supormos que a priori tem caudas mais leves que a verossimilhança, o que pode evitar que o denominador decresça rápido demais, fazendo a referida razão tender ao infinito. A suposição de que essa razão é maior que 1/2 serve apenas para garantir que a posteriori convergirá para uma densidade. A exigência de que f e π tenham formas distintas também serve para garantir que a posteriori convergirá para uma densidade (se as duas fontes tiverem a mesma densidade h , então a posteriori convergirá para $h/2$).

O teorema acima é adaptável para o caso no qual existam condições para descartarmos a informação a priori. Para tanto basta trocar os papéis da verossimilhança e da priori. A seguir damos alguns exemplos onde conflitos são resolvidos empregando o teorema 4.5.

Exemplo 4.2. O caso no qual X dado Θ tem distribuição t-Student com ν graus de liberdade e Θ possui distribuição normal é facilmente tratado à luz do Teorema 4.5. Primeiramente, notemos que a distribuição t-Student pertence à família subexponencial e que possui o mesmo suporte da normal, satisfazendo portanto os itens 1 e 2 do referido teorema. Em seguida, basta notar que o supremo pedido no item 3 existe para qualquer ν , uma vez que a distribuição normal concentra mais massa ao redor de sua moda e suas caudas têm decaimento mais rápido do que a distribuição t-Student. Na Figura 4.2, mostramos a razão entre a distribuição normal padrão e a distribuição t-Student com 31

graus de liberdade e nesta podemos notar que o referido supremo é finito e maior que meio. Na Figura 4.1, temos dois gráficos mostrando a densidade a priori e a posteriori e a função de verossimilhança, com $x = 3$ e $x = 10$, respectivamente. Notemos que com o aumento de x a distribuição a posteriori tende a se aproximar da distribuição a priori.

Exemplo 4.3. Suponha que para um determinado problema temos que

$$\begin{cases} f(x|\theta) = \frac{1}{6}(x - \theta)^{-1/2} \exp\left\{-\frac{(x-\theta)^{1/2}}{3}\right\} \\ \pi(\theta) = \frac{1}{8} \frac{\Gamma(5)}{\Gamma(2)\Gamma(3)} \left(\frac{\theta-2}{8}\right) \left(\frac{10-\theta}{8}\right)^2 \end{cases}$$

Notemos que neste caso depositamos forte credibilidade na informação a priori, representada pela distribuição beta com parâmetros de forma $\alpha = 2$ e $\beta = 3$ e com suporte entre 2 e 10. Em particular, temos que $E(\Theta) = 5,2$. Para a distribuição dos dados temos uma Weibull com parâmetro de forma igual a $1/2$. Com estes modelos, podemos esperar que eventualmente uma observação entre em desacordo com as demais devido a natureza da distribuição dos dados. Para mostrar que esta modelagem descarta a informação dos dados em caso de conflito, basta notarmos que estão satisfeitas as três condições do teorema 4.5, isto é:

- $f \in \mathcal{S}$
- $(2, 10) \subset (0, \infty)$
- o supremo da razão $\pi(\theta)/f(\theta)$ é finito e maior que meio, como mostra a Figura 4.3.

Com isto mostramos que, quando a observação x tende ao infinito, a distribuição a posteriori tende à priori, descartando a informação conflitante. A Figura 4.5 mostra essa aproximação supondo $x = 20$. A Figura 4.4 mostra a esperança a posteriori para diversos valores de x .

Exemplo 4.4. Mostraremos um exemplo no qual as condições de Dawid não se aplicam mas a convergência das distribuições é garantida pelo teorema 4.5. Suponha que X dado Θ tem distribuição Cauchy e que Θ seja parâmetro de locação com densidade Pareto dada por $\pi(\theta) = \theta^{-2} I_{(1, \infty)}(\theta)$. Notemos que, pelo Lema 3.7 temos que

$$\sup_x \frac{f(x - \theta)}{f(x)} = \sup_x \frac{1 + x^2}{1 + (x - \theta)^2} = \frac{\sqrt{\theta^2 + 4} + \theta}{\sqrt{\theta^2 + 4} - \theta} \geq \theta$$

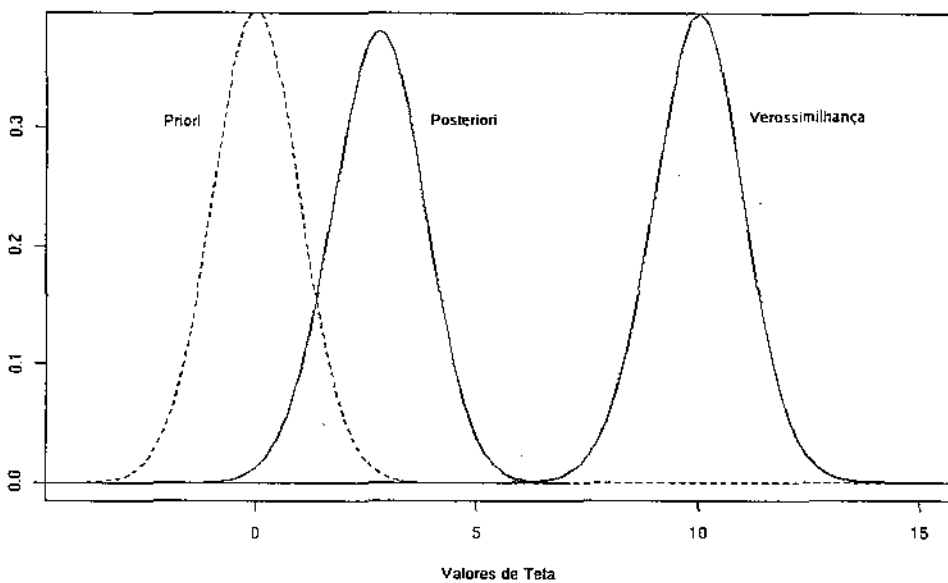
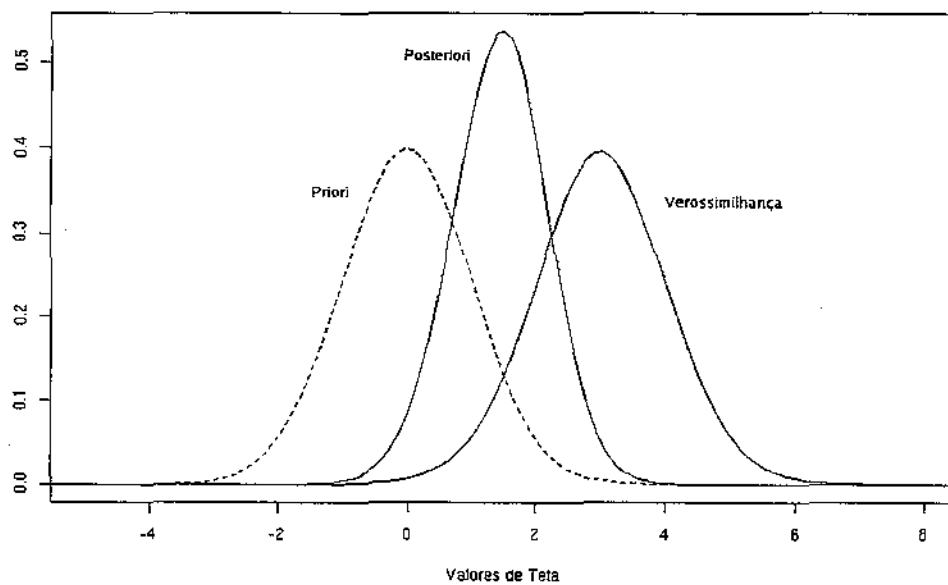


Figura 4.1: Priori, verossimilhança e posteriori para as densidades no Exemplo 4.2, com $x = 3$ e $x = 10$ de cima para baixo.

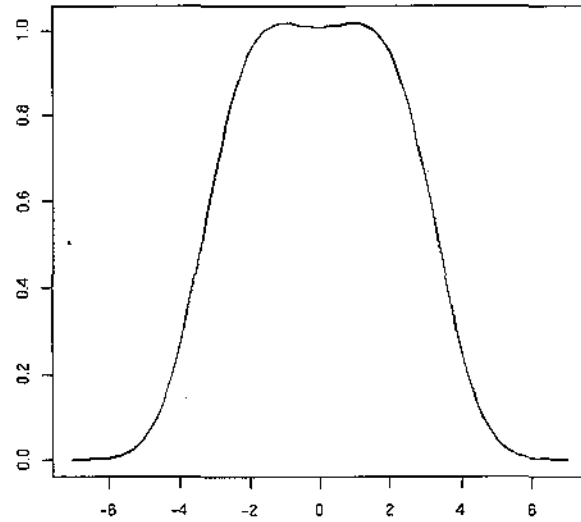


Figura 4.2: Razão entre a distribuição normal padrão e a distribuição t-Student com 31 graus de liberdade

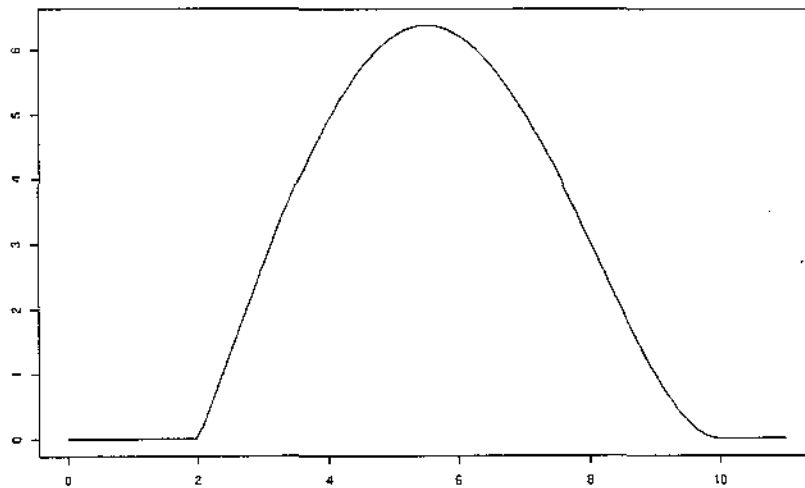


Figura 4.3: Razão entre a densidade $Beta(2, 3, 2, 10)$ e a densidade Weibull com parâmetro de forma igual a $1/2$

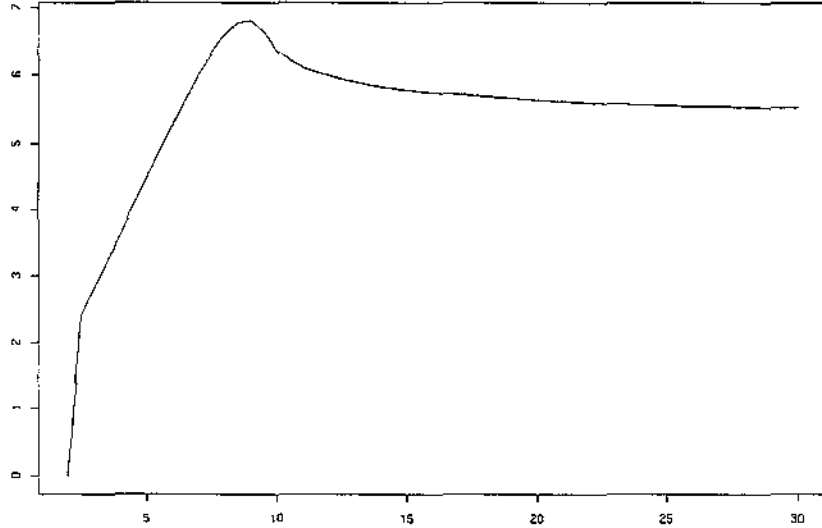


Figura 4.4: Esperança a posteriori para diversos valores de x , onde $X|\theta \sim Weibull(1/2, \theta)$ e $\Theta \sim Beta(2, 3, 2, 10)$

para todo $\theta > 1$. Assim, temos que

$$\int_{\Theta} \sup_x \frac{f(x - \theta)}{f(x)} \pi(\theta) d\theta \geq \int_1^{\infty} \frac{1}{\theta} d\theta = \infty$$

e portanto a terceira condição de Dawid não está satisfeita. Entretanto, é imediato que as condições 1 e 2 do teorema 4.5 estão satisfeitas. Observemos agora que

$$\frac{\pi(\theta)}{f(\theta)} = \pi \left(\frac{1 + \theta^2}{\theta^2} \right) I_{(1, \infty)}$$

logo

$$\pi \leq \sup_{\theta} \frac{\pi(\theta)}{f(\theta)} \leq 2\pi$$

e portanto a terceira condição do Teorema 4.5 também está satisfeita, logo temos que em caso de conflito entre as fontes de informação, a distribuição a posteriori deveria convergir

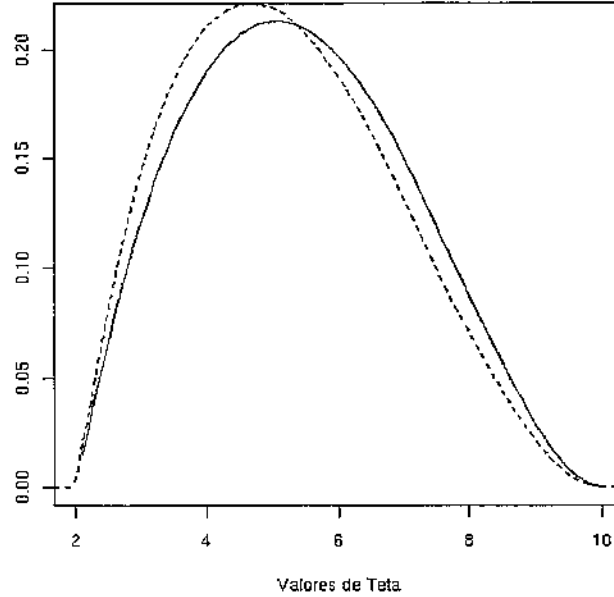


Figura 4.5: Linha sólida: posteriori para o Exemplo 4.3, supondo $x = 20$. Linha pontilhada: densidade a priori

para a distribuição a priori. De fato, temos que

$$\begin{aligned}
 p(\theta|x) &= \frac{f(x-\theta)}{f(x)}\pi(\theta) \left(\int_{\Theta} \frac{f(x-\theta)}{f(x)}\pi(\theta)d\theta \right)^{-1} \\
 &= \frac{1+x^2}{1+(x-\theta)^2}\frac{1}{\theta^2} \left(\int_1^{\infty} \frac{1+x^2}{1+(x-\theta)^2}\frac{1}{\theta^2}d\theta \right)^{-1}.
 \end{aligned}$$

Temos que, para o numerador da expressão acima,

$$\frac{1+x^2}{1+(x-\theta)^2}\frac{1}{\theta^2} \rightarrow \frac{1}{\theta^2}$$

quando $x \rightarrow \infty$. Para o denominador, notemos que

$$\begin{aligned}
 &\int_1^{\infty} \frac{1+x^2}{1+(x-\theta)^2}\frac{1}{\theta^2}d\theta \\
 &= \frac{2-\pi+(2+\pi)x^2-2(x^2-1)\arctan(1-x)+2x\log[2+(x-2)x]}{2(1+x^2)} \\
 &\rightarrow 1
 \end{aligned}$$

quando $x \rightarrow \infty$, e portanto

$$\lim_{x \rightarrow \infty} p(\theta|x) = \pi(\theta),$$

conforme o esperado.

Exemplo 4.5. Consideremos uma observação x e que estamos interessados em um parâmetro de locação θ . Dois modelos são propostos a seguir:

- **Modelo I:** $g(x|\theta) \propto (x - \theta) \exp\{-2.4x\}$ e $\pi(\theta) \propto \theta \exp\{-\theta\}$
- **Modelo II:** $g(x|\theta) \propto (x - \theta)^{-1} \exp\{-\frac{(\log(x - \theta))^2}{0.5}\}$ e $\pi(\theta) \propto \theta \exp\{-\theta\}$

O Modelo I considera a distribuição gama, de cauda leve. O Modelo II considera uma lognormal que possui densidade satisfazendo a condição 1 do teorema 4.5. Notemos que as condições 2 e 3 também estão satisfeitas. Calculamos para ambos a esperança a posteriori. Para o Modelo I, obtivemos uma esperança a posteriori sempre crescente com o aumento de x enquanto que para o Modelo II, esta se aproximou cada vez mais da esperança a priori, como pode ser visto na Figura 4.6

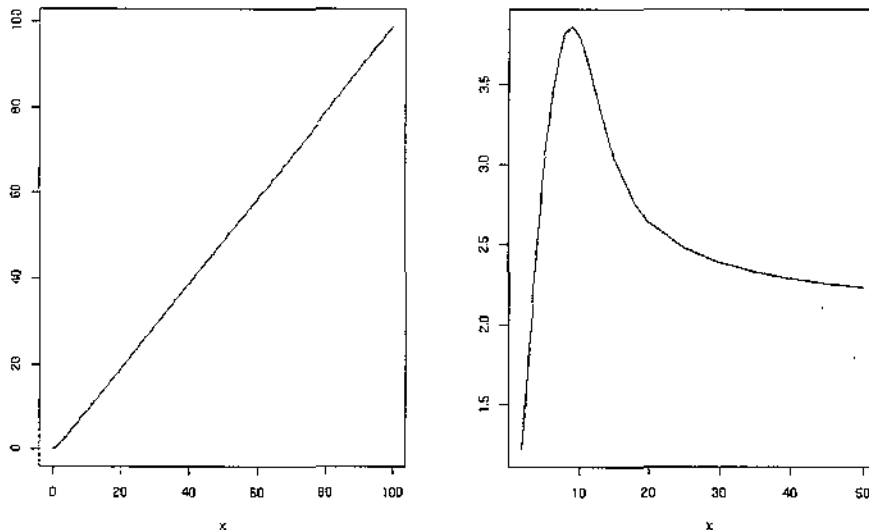


Figura 4.6: Esperanças de θ a posteriori com o crescimento de x para os modelos I e II respectivamente

4.3.2 Rejeitando várias observações

Suponha que X_1, \dots, X_n sejam variáveis aleatórias independentes e identicamente distribuídas dado Θ , com função densidade $f \in \mathcal{SD}$. Seja π a função densidade a priori. Mostraremos como o Teorema 4.5 pode ser utilizado para o caso em que temos um grupo de observações *outliers*.

Suponha que observamos $\mathbf{x} = \{x_1, \dots, x_n\}$. Considere que as observações x_{J_1} , com $i \in I$, sejam *outliers*, onde $I = \{1, \dots, n_o\}$. Suponha ainda que

- $\sup_{\theta} \pi(\theta)/f(\theta) = M < \infty$
- $\sup_{\theta} f(x_i - \theta) = C < \infty$
- $A(\pi) \subseteq A(f)$, onde $A(\cdot)$ é o suporte da distribuição

Inicialmente, suponha que temos apenas um *outlier*, identificado por x_{J_1} . Temos que

$$p(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)\pi(\theta)/f(x_{J_1})}{\int_{\Theta} f(\mathbf{x}|\theta)\pi(\theta)d\theta/f(x_{J_1})} = \frac{\prod_{i=1}^n f(x_i|\theta)\pi(\theta)/f(x_{J_1})}{\int_{\Theta} f(\prod_{i=1}^n f(x_i|\theta)\pi(\theta)d\theta/f(x_{J_1}))}$$

Como $f \in \mathcal{SD}$, temos que

$$\prod_{i=1}^n \frac{f(x_i - \theta)}{f(x_{J_1})} = \frac{f(x_{J_1} - \theta)}{f(x_{J_1})} \prod_{i \neq J_1} f(x_i - \theta) \rightarrow \prod_{i \neq J_1} f(x_i - \theta)$$

quando $x_{J_1} \rightarrow \infty$. Também temos que

$$\begin{aligned} \int_{A(\pi)} \prod_{i=1}^n \frac{f(x_i - \theta)\pi(\theta)}{f(x_{J_1})} d\theta &\leq \int_{A(f)} \prod_{i=1}^n \frac{f(x_i - \theta)f(\theta)\pi(\theta)}{f(x_{J_1})f(\theta)} d\theta \\ &\leq \int_{A(f)} \frac{f(x_{J_1} - \theta)}{f(x_{J_1})} f(\theta) \prod_{i \neq J_1} f(x_i - \theta) M d\theta \\ &\leq MC^{n-1} \int_{A(f)} \frac{f(x_{J_1} - \theta)}{f(x_{J_1})} f(\theta) d\theta \rightarrow 2MC^{n-1}. \end{aligned}$$

Notemos agora que

$$\int_{A(\pi)} \prod_{i=1}^{n-1} f(x_i - \theta)\pi(\theta) d\theta \leq C^{n-1} \int_{A(\pi)} \pi(\theta) d\theta = C^{n-1}.$$

Portanto, desde que $M > 1/2$, teremos que

$$\begin{aligned} \lim_{x_{J_1} \rightarrow \infty} \int_{A(\pi)} \prod_{i=1}^n \frac{f(x_i - \theta)\pi(\theta)}{f(x_{J_1})} d\theta &= \int_{A(\pi)} \lim_{x_{J_1} \rightarrow \infty} \prod_{i=1}^n \frac{f(x_i - \theta)\pi(\theta)}{f(x_{J_1})} d\theta \\ &= \int_{A(\pi)} \prod_{i \neq J_1}^n f(x_i - \theta)\pi(\theta) d\theta \end{aligned}$$

o que implica em

$$p(\theta|\mathbf{x}) \rightarrow \frac{\prod_{i \neq J_1} f(x_i - \theta)\pi(\theta)}{\int_{A(\pi)} \prod_{i \neq J_1} f(x_i - \theta)\pi(\theta) d\theta}$$

Suponha agora que temos dois *outliers*, x_{J_1} e x_{J_2} . De maneira análoga ao que foi feito acima, teremos que

$$\lim_{x_{J_2} \rightarrow \infty} \lim_{x_{J_1} \rightarrow \infty} p(\theta|\mathbf{x}) = \lim_{x_{J_2} \rightarrow \infty} p(\theta|x_i, i \neq J_1) = p(\theta|x_i \neq J_1, J_2)$$

e a extensão segue naturalmente para todos os *outliers*. Assim, desde que as condições do Teorema 4.5 estejam satisfeitas para uma única observação, teremos que

$$p(\theta|\mathbf{x}) \rightarrow p(\theta|x_i, i \in I^c)$$

quando $x_j \rightarrow \infty$, para todo $j \in I$.

Exemplo 4.6. Para ilustrar a utilização do Teorema 4.5, suponha que para um determinado problema observamos X_1, \dots, X_n , uma amostra de variáveis aleatórias independentes identicamente distribuídas dado Θ . Suponha que Θ é parâmetro de locação com distribuição $Beta(\alpha, \beta, a, b)$, onde α e β são os parâmetros de forma e a e b são os de locação, com seus respectivos valores sendo 3, 3, -3 e 3. Para fins de comparação, propomos duas distribuições para os dados:

- $X_i|\theta \sim N(\theta, 1)$: neste caso não esperamos que a informação gerada pelos dados possa discordar com a informação a priori e portanto *outliers* não são esperados;
- $X_i|\theta \sim t_{(31, \theta)}$: neste caso suspeitamos que a informação dos dados pode discordar da informação a priori;

Suponha que achamos a distribuição t-Student com 31 graus de liberdade apropriada para a situação e gostaríamos de saber se, em caso de conflito, a inferência a posteriori será

influenciada pelas observações discrepantes. Notemos que as condições 1 e 2 do Teorema 4.5 são de verificação imediata. Para mostrar que a condição 3 é verdadeira, podemos mostrar que

$$\left\{ \prod_{j \in I} \sup_{\theta} f(x_j - \theta) \right\} \sup_{\theta} \frac{\pi(\theta)}{f(\theta)} < \infty.$$

Suponha que I tenha n_0 elementos. Temos então que

$$\prod_{j \in I} \sup_{\theta} f(x_j - \theta) = f(0)^{n_0} < \infty$$

Como a razão $\pi(\theta)/f(\theta)$ é finita no suporte de θ temos que $\sup_{\theta} \pi(\theta)/f(\theta) < \infty$ e portanto o Teorema 4.5 pode ser aplicado. Assim, na ocorrência de *outliers* teremos que a inferência a posteriori dará maior peso para a nossa crença a priori. Para ilustrarmos a aplicação, considere X e Y independentes e identicamente distribuídos dado Θ . As Figuras 4.7 e 4.8 mostram a esperança a posteriori para diversos valores de X e Y assumindo que Θ é parâmetro de locação com distribuição $Beta(3, 3, -3, 3)$, com X e Y tendo distribuição $N(\theta, 1)$ e $t_{(31, \theta)}$. Observemos que para a distribuição normal, pequenas variações de X e Y fazem com que a esperança a posteriori se aproxime de -3 ou 3 , sugerindo que θ deveria estar próximo dos extremos da distribuição a priori, causando um conflito entre as fontes. Contudo, para a distribuição t-Student temos que a posteriori tende a rejeitar a informação conflitante à medida que esta aumenta em valor absoluto.

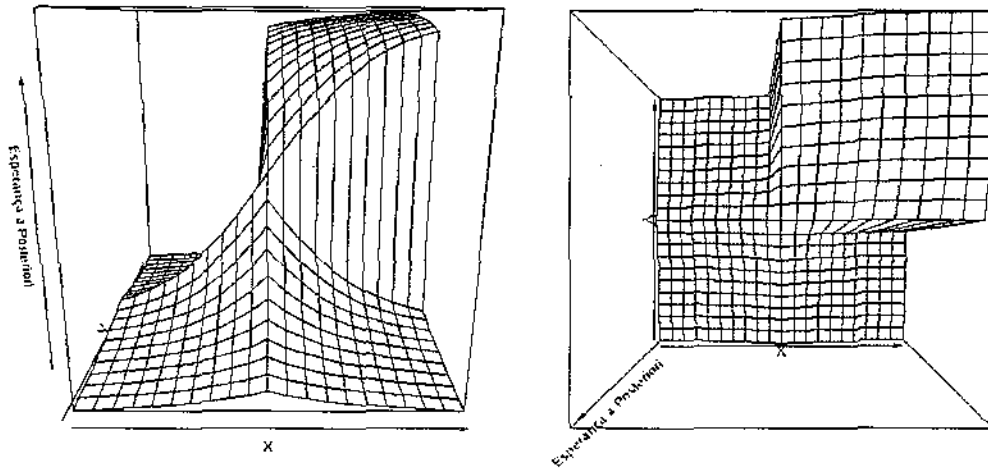


Figura 4.7: Esperança a posteriori para valores de X e Y entre -10 e 10, supondo que os dados possuem distribuição $N(\theta, 1)$

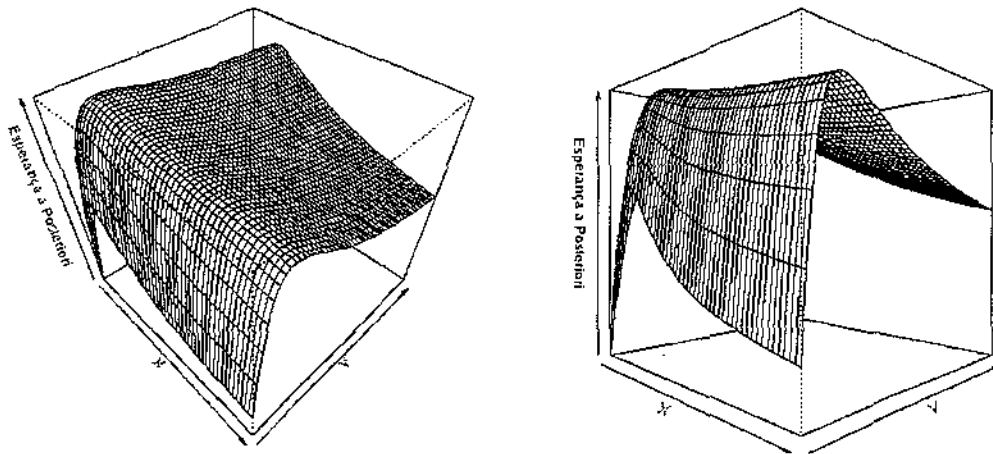


Figura 4.8: Esperança a posteriori para valores de X e Y entre 0 e 50, supondo que os dados possuem distribuição $t_{(31, \theta)}$

4.3.3 Rejeitando a informação a priori

Suponha agora que acreditamos que a informação a priori pode, eventualmente, se distanciar da informação proveniente dos dados. Mais especificamente, suponha que X dado Θ tenha densidade $g(x|\theta) = f(x - \theta)$ e que Θ tenha densidade $\pi(\theta)$ com $\pi \in \mathcal{SD}$. Temos então que

$$p_\theta(\theta|x) \propto f(x - \theta)\pi(\theta).$$

Sejam $\Phi = X - \Theta$ e $Y = X$. Então teremos que

$$p_\phi(\phi|y) \propto f(\phi)\pi(x - \phi)$$

Assim, se supormos que $A(f) \subset A(\pi)$ e que

$$\sup_\phi \frac{f(\phi)}{\pi(\phi)} < \infty$$

poderemos aplicar o Teorema 4.5 para mostrar que

$$p_\phi(\phi|y) \rightarrow f(\phi)$$

quando $y \rightarrow \infty$, rejeitando completamente a informação a priori.

4.4 Caracterização de credence

Discutimos na Seção 3.4 sobre a medida denominada *credence*, utilizada para a resolução de conflitos de informação. O autor comenta que muitas distribuições de caudas pesadas comportam-se de maneira semelhante. Nesta seção mostraremos que se uma densidade possui *credence* então ela pertence à *ERV*. Em seguida mostraremos a relação desta medida com a distribuição t-Student.

Proposição 4.6. *Se f possui credence então $F \in ERV$*

Demonstração. Suponhamos que $g(x|\theta) = (1/\theta) \times f(x/\theta)$. Veremos mais adiante, na Proposição 4.9 que para todo $x \in \mathbb{R}$

$$\frac{k \min\{1, \theta^c\}}{\theta K} \leq \frac{g(x|\theta)}{f(x)} \leq \frac{K \max\{1, \theta^c\}}{\theta k}$$

Façamos $m(\theta) = (k/K) \times \min\{1, \theta^c\}$ e $M(\theta) = (K/k) \times \max\{1, \theta^c\}$. Agora, notemos que para todo $x \in \mathbb{R}$

$$\begin{aligned} \frac{m(\theta)}{\theta} \leq \frac{g(x|\theta)}{f(x)} \leq \frac{M(\theta)}{\theta} &\Rightarrow m(\theta)f(x) \leq f(x/\theta) \leq M(\theta)f(x) \\ &\Rightarrow \frac{m(\theta)}{\theta} \int_x^\infty f(t)dt \leq \int_x^\infty \frac{1}{\theta} f(t/\theta)dt \leq \frac{M(\theta)}{\theta} \int_x^\infty f(t)dt \\ &\Rightarrow \frac{m(\theta)}{\theta} \leq \frac{\bar{F}(x/\theta)}{\bar{F}(x)} \leq \frac{M(\theta)}{\theta} \end{aligned}$$

e para todo θ teremos que

$$\frac{k}{K} \min\{\theta^{-1}, \theta^{c-1}\} \leq \frac{\bar{F}(x/\theta)}{\bar{F}(x)} \leq \frac{K}{k} \max\{\theta^{-1}, \theta^{c-1}\}$$

logo $F \in ERV$. □

Com base nisto, seria interessante que todas as distribuições de variação regular tivessem *credence*. Infelizmente isto não é sempre verdade, uma vez que as distribuições de variação regular não precisam estar definidas em toda a reta, nem precisam possuir densidade para todo x . Podemos notar, por exemplo, que a distribuição Pareto pertence a \mathcal{RV}_2 mas sua *credence* é infinita, recebendo portanto a mesma classificação que uma distribuição normal.

A medida *credence* foi idealizada a partir do comportamento da distribuição t-Student. De fato, se supormos que g tem distribuição t-Student e portanto

$$g(x) \propto (v + x^2)^{-\frac{1}{2}(v+1)}$$

e se supormos que $cred(f) = v + 1$, teremos que

$$\frac{f(x)}{(1 + x^2)^{-\frac{1}{2}(v+1)}} = \frac{f(x)}{(v + x^2)^{-\frac{1}{2}(v+1)}} \left\{ \frac{v + x^2}{1 + x^2} \right\}^{-\frac{1}{2}(v+1)} = \frac{f(x)}{g(x)} h(v) \left\{ \frac{v + x^2}{1 + x^2} \right\}^{-\frac{1}{2}(v+1)}$$

onde $g(x) = h(v)(v + x^2)^{-\frac{1}{2}(v+1)}$. Pelo fato de f possuir *credence* e por existirem constantes $b \geq a > 0$ que limitam o termo entre chaves na última igualdade acima, temos que $g \approx^w f$. Em especial, pelo Lema 2.6, se $f \in \mathcal{L}$, então $F \in \mathcal{S}$.

Portanto, concluímos nossa caracterização de *credence* em termos de caudas pesadas: se f possui *credence*, então $f \in ERV$ e, se supormos ainda que $f \in \mathcal{L}$ então $F \in \mathcal{S}$.

4.5 Utilizando credence para problemas de escala

Vimos anteriormente que a medida *credence* poderia ser utilizada para resolver conflitos entre as fontes de informação no caso onde pelo menos uma das fontes tinha uma estrutura de locação. Neste caso um conflito estaria resolvido a favor da fonte que possuísse maior *credence*. Nesta seção discutiremos como essa mesma medida pode ser utilizada para resolver conflitos de informação quando uma das fontes assume a estrutura escala.

Neste momento se torna oportuno recordar o Lema 3.7. Este dizia que para quaisquer $a \neq 0$ e b , existem $L \geq l > 0$ tais que para todo $x \in \mathbb{R}$

$$l \leq \frac{1 + (ax + b)^2}{1 + x^2} \leq L$$

Notemos que as constantes l e L dadas pelo Lema 3.7 são funções de a e b . De posse deste resultado, podemos caracterizar melhor as constantes dadas no teorema da invariância, como sugere a seguinte proposição.

Proposição 4.7. *Seja $g(x|a, b) = (1/a) \times f((x - b)/a)$ e seja $\text{cred}(f) = c$. Então para quaisquer a e b existem $K_{(a,b)} \geq k_{(a,b)} > 0$ tais que*

$$k_{(a,b)} < (1 + x^2)^c g(x|\theta) < K_{(a,b)}. \quad (4.4)$$

Demonstração. Da demonstração do teorema da invariância temos que

$$\frac{kl^{c/2}}{a} \leq (1 + x^2)^{c/2} g(x|\theta) \leq \frac{KL^{c/2}}{a}$$

onde l e L são obtidos pelo Lema 3.7. Mas isto implica em

$$k_{(a,b)} \leq (1 + x^2)^{c/2} g(x|a, b) \leq K_{(a,b)}$$

onde $k_{(a,b)} = (k/a) \times (l_b/L_{1/a})^{c/2}$ e $K_{(a,b)} = (K/a) \times (L_b/l_{1/a})^{c/2}$ concluindo a demonstração. \square

Corolário 4.8. *Suponhamos que $g(x|\theta) = (1/\theta) \times f(x/\theta)$ onde $\text{cred}(f) = c$. Então*

$$k \min\{\theta^{-1}, \theta^{c-1}\} \leq (1 + x^2)^{c/2} g(x|\theta) \leq K \max\{\theta^{-1}, \theta^{c-1}\}. \quad (4.5)$$

Demonstração. Da demonstração anterior temos que

$$k_{(\theta,0)} \leq (1+x^2)^{c/2} g(x|\theta) \leq K_{(\theta,0)}$$

pois não temos o parâmetro de locação. Assim, $L_0 = l_0 = 1$ e portanto $k_{(\theta,0)} = k \min\{1, \theta^c\}/\theta$ e $K_{(\theta,0)} = K \max\{1, \theta^c\}/\theta$. \square

Proposição 4.9. *Seja $g(x|\theta) = (1/\theta)f(x/\theta)$. Então existem $\kappa_2(\theta) \geq \kappa_1(\theta) > 0$ tais que*

$$\kappa_1(\theta) \leq \frac{g(x|\theta)}{f(x)} \leq \kappa_2(\theta).$$

Demonstração. Pelo teorema da invariância, sabemos que g também possui credence e portanto,

$$\frac{g(x|\theta)}{f(x)} = \frac{g(x|\theta)(1+x^2)^{c/2}}{f(x)(1+x^2)^{c/2}} \leq \frac{K \max\{1, \theta^c\}}{\theta k}.$$

Por outro lado

$$\frac{g(x|\theta)}{f(x)} = \frac{g(x|\theta)(1+x^2)^{c/2}}{f(x)(1+x^2)^{c/2}} \geq \frac{k \min\{1, \theta^c\}}{\theta K}$$

portanto

$$\frac{k \min\{1, \theta^c\}}{\theta K} \leq \frac{g(x|\theta)}{f(x)} \leq \frac{K \max\{1, \theta^c\}}{\theta k}.$$

Agora, basta tomar $\kappa_1(\theta) = (k/K) \min\{\theta^{-1}, \theta^{c-1}\}$ e $\kappa_2(\theta) = (K/k) \max\{\theta^{-1}, \theta^{c-1}\}$. \square

Teorema 4.10. *Sejam $g(x|\theta) = (1/\theta)f(x/\theta)$ a verossimilhança, $\pi(\theta) > 0$ a distribuição a priori e $p(\theta|x)$ a distribuição a posteriori para um determinado problema. Suponhamos que*

1. $\text{cred}(f) = c(c > 1)$
2. $\int_{\Theta} \theta^c \pi(\theta) d\theta < \infty$

então, para todo $x \in \mathbb{R}$ temos que

$$\frac{\kappa_1(\theta)\pi(\theta)}{\int_0^\infty \kappa_2(\theta)\pi(\theta)d\theta} \leq p(\theta|x) \leq \frac{\kappa_2(\theta)\pi(\theta)}{\int_0^\infty \kappa_1(\theta)\pi(\theta)d\theta}$$

e que $E(\theta|x) < \infty$, onde $\kappa_1(\theta)$ e $\kappa_2(\theta)$ são como na Proposição 4.9.

Demonstração. Temos que

$$p(\theta|x) = \frac{g(x|\theta)\pi(\theta)}{\int_0^\infty g(x|\theta)\pi(\theta)d\theta} = \frac{g(x|\theta)\pi(\theta)/f(x)}{\int_0^\infty \{g(x|\theta)\pi(\theta)/f(x)\}d\theta}.$$

Para o numerador, temos que para todo $x \in \mathbb{R}$,

$$\kappa_1(\theta)\pi(\theta) \leq \frac{g(x|\theta)}{f(x)}\pi(\theta) \leq \kappa_2(\theta)\pi(\theta).$$

Agora, notemos que

$$\begin{aligned} \int_0^\infty \kappa_2(\theta)\pi(\theta)d\theta &= \int_0^\infty \frac{K}{k} \max\{\theta^{-1}, \theta^{c-1}\}\pi(\theta)d\theta \\ &= \int_0^1 \frac{K}{k} \max\{\theta^{-1}, \theta^{c-1}\}\pi(\theta)d\theta + \int_1^\infty \frac{K}{k} \max\{\theta^{-1}, \theta^{c-1}\}\pi(\theta)d\theta \\ &= \frac{K}{k} \int_0^1 \theta^{-1}\pi(\theta)d\theta + \frac{K}{k} \int_1^\infty \theta^{c-1}\pi(\theta)d\theta \end{aligned}$$

que é finito pela condição (2). Notemos também que

$$\begin{aligned} \int_0^\infty \kappa_1(\theta)\pi(\theta)d\theta &= \int_0^\infty \frac{k}{K} \min\{\theta^{-1}, \theta^{c-1}\}\pi(\theta)d\theta \\ &= \int_0^1 \frac{k}{K} \min\{\theta^{-1}, \theta^{c-1}\}\pi(\theta)d\theta + \int_1^\infty \frac{k}{K} \min\{\theta^{-1}, \theta^{c-1}\}\pi(\theta)d\theta \\ &= \frac{k}{K} \int_0^1 \theta^{c-1}\pi(\theta)d\theta + \frac{k}{K} \int_1^\infty \theta^{-1}\pi(\theta)d\theta > 0 \end{aligned}$$

assim, temos que para todo $x \in \mathbb{R}$,

$$0 < \int_0^\infty \kappa_1(\theta)\pi(\theta)d\theta \leq \int_0^\infty \frac{g(x|\theta)}{f(x)}\pi(\theta)d\theta \leq \int_0^\infty \kappa_2(\theta)\pi(\theta)d\theta < \infty$$

portanto

$$\frac{\kappa_1(\theta)\pi(\theta)}{\int_0^\infty \kappa_2(\theta)\pi(\theta)d\theta} \leq p(\theta|x) \leq \frac{\kappa_2(\theta)\pi(\theta)}{\int_0^\infty \kappa_1(\theta)\pi(\theta)d\theta}.$$

Para mostrar que a esperança a posteriori é limitada, note que

$$E(\theta|x) = \frac{\int_0^\infty \{\theta g(x|\theta)\pi(\theta)/f(x)\}d\theta}{\int_0^\infty \{g(x|\theta)\pi(\theta)/f(x)\}d\theta}.$$

Já vimos que o denominador da razão acima é limitado e não nulo. Para o numerador, observe que

$$\begin{aligned} \int_0^{\infty} \theta \kappa_2(\theta) \pi(\theta) d\theta &= \int_0^{\infty} \theta \frac{K}{k} \max\{\theta^{-1}, \theta^{c-1}\} \pi(\theta) d\theta \\ &= \int_0^1 \theta \frac{K}{k} \max\{\theta^{-1}, \theta^{c-1}\} \pi(\theta) d\theta + \int_1^{\infty} \theta \frac{K}{k} \max\{\theta^{-1}, \theta^{c-1}\} \pi(\theta) d\theta \\ &= \frac{K}{k} \int_0^1 \pi(\theta) d\theta + \frac{K}{k} \int_1^{\infty} \theta^c \pi(\theta) d\theta \end{aligned}$$

que é finito pela condição (2) portanto

$$E(\theta|x) \leq \frac{\int_0^{\infty} \theta \kappa_2(\theta) \pi(\theta) d\theta}{\int_0^{\infty} \kappa_1(\theta) \pi(\theta) d\theta} < \infty$$

terminando a prova. □

Assim, se as condições deste teorema forem satisfeitas, podemos determinar a maior influência que um valor de x terá sobre a a posteriori. Este resultado mostra que a fonte de informação proveniente dos dados se torna pouco informativa, mas não há rejeição completa desta. Podemos dizer que toda a informação do *outlier* está contida em $\kappa_1(\theta)$ e $\kappa_2(\theta)$ e que a informação a priori domina a informação entregue pelos dados.

Exemplo 4.7. Podemos observar que os Teoremas 4.10 e 3.14 se aplicam em situações parecidas. Neste exemplo mostraremos uma situação na qual o teorema 3.14 não pode ser aplicado. Assim, para um determinado problema suponhamos que a densidade dos dados seja

$$g(x|\theta) = \frac{1}{\theta} f(x/\theta) = \frac{1}{\theta} \frac{\Gamma(\frac{v}{2})}{\Gamma(\frac{v-1}{2})} \frac{1}{2\sqrt{\pi}} \frac{2 + \sin(x/\theta)}{\{1 + (x/\theta)^2\}^{v/2}}$$

e que a distribuição a priori seja $Gama(3, 2)$. Temos que $E(X|\theta) = 0,210512\theta$ (para $v > 2$) e que $E\Theta = 1,5$. Notemos que o teorema 3.14 não pode ser aplicado uma vez que a densidade de g não é de variação regular para todo θ (observemos contudo que g pertence à *ERV*). Notemos que

- f possui *credence* igual a v
- $\int_0^{\infty} \theta^c \pi(\theta) d\theta = \Gamma(3+v)/(2^v \Gamma(3)) < \infty$

e portanto podemos aplicar o teorema 4.10. Assim, damos condições suficientes para que um possível conflito seja resolvido a favor da priori. A informação *outlier* não será completamente rejeitada mas exercerá pouca influência sobre a posteriori. Em especial podemos garantir que a esperança a posteriori será finita.

Podemos ainda calcular o limitante superior da esperança a posteriori. Para tanto, suponhamos, por comodidade, que v é igual a 3. Então temos que:

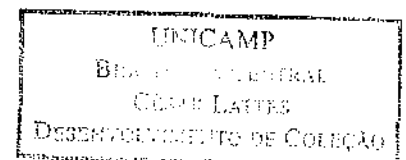
- $k = 0,25$ e $K = 0,75$ são as constantes encontradas através da definição de *credence*
- $\int_0^\infty \kappa_1(\theta)\pi(\theta)d\theta = 0,187988$
- $\int_0^\infty \kappa_2(\theta)\pi(\theta)d\theta = 10,3081$
- $\int_0^\infty \theta\kappa_2(\theta)\pi(\theta)d\theta = 23,0973$

portanto, para x suficientemente grande, temos que

$$E(\theta|x) \leq 122,866$$

Com este intervalo podemos afirmar que, na presença de um *outlier*, a esperança da posteriori será menor que 122,866. Contudo, para valores observados muito elevados, a esperança a posteriori apresenta oscilações bem inferiores ao limitante superior dado em função de $\kappa_1(\theta)$ e $\kappa_2(\theta)$, como pode ser visto na Figura 4.9.

Na Figura 4.10 mostramos o gráfico das fontes de informação e da posteriori sobre θ em duas situações. Na primeira, fixamos $x = 5$. Notemos que a posteriori se concentra próxima das regiões de maior densidade de ambas as fontes. Na segunda situação, fixamos $x = 25$. Notemos que neste caso, a verossimilhança pouco influencia a posteriori. Entretanto, a posteriori oscila, não se aproximando perfeitamente da priori. Apesar de não haver convergência para a priori, sabemos, pelo teorema 4.10, que existem múltiplos positivos da priori que dominam a posteriori. Neste caso dizemos que a informação a priori domina a informação dos dados e portanto o conflito é resolvido a favor da priori.



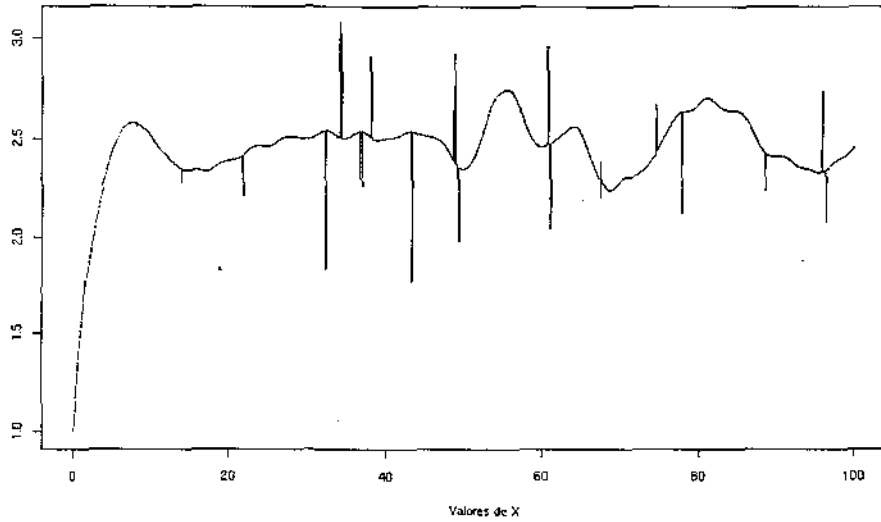


Figura 4.9: Esperança de θ a posteriori para x entre 0 e 100, considerando as distribuições dadas no Exemplo 4.7

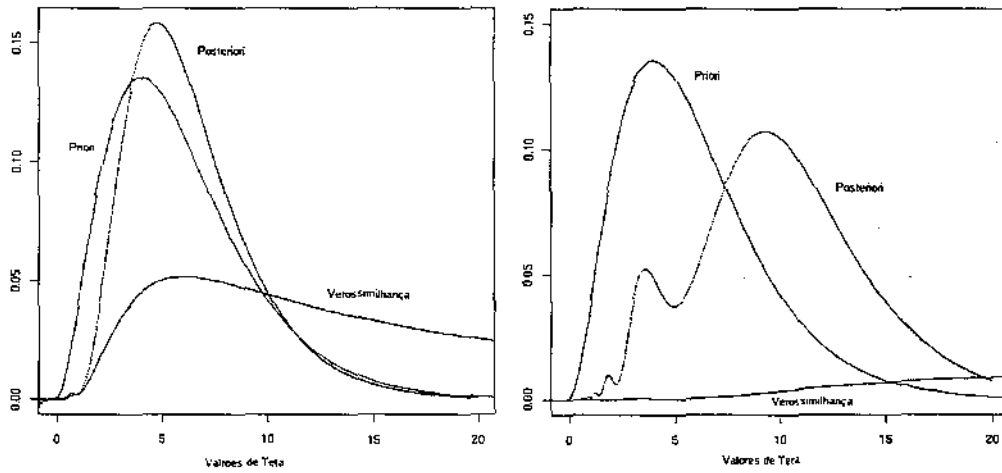


Figura 4.10: Priori, verossimilhança e posteriori referentes ao Exemplo 4.7 graficadas sobre θ . Acima fixamos $x = 5$ e abaixo fixamos $x = 25$.

Capítulo 5

Considerações finais

Ao longo deste discutimos os conceitos básicos de conflitos de informação, bem como algumas possíveis ações que evitam que a posteriori seja influenciada pelo agente causador do conflito, que neste caso é o *outlier*. Dentre estas possíveis ações, destacamos em especial o uso de distribuições de cauda pesada. Assim, atribuindo-se distribuições com pesos diferentes para as fontes de informação, vimos que em caso de conflito a posteriori, sob certas condições, rejeitava a informação com a cauda mais pesada. Vimos também os diferentes tipos de rejeição encontrados na literatura.

No Capítulo 2, vimos que as distribuições de cauda longa são um subconjunto da classe \mathcal{G} , que por sua vez é um subconjunto da classe de distribuições propensas a *outliers*. Embora não tenhamos discutido muitos resultados sobre resolução de conflitos de informação utilizando distribuições que não fossem de cauda pesada, não descartamos a hipótese de que outras distribuições propensas a *outliers* não possam ser utilizadas para tal fim. Desde que utilizamos o peso da cauda da distribuição como um dos critérios para resolver conflitos, ou ainda, a propensão da distribuição para gerar *outliers*, a distribuições propensas a *outliers* podem refletir nossa opinião sobre a credibilidade da fonte.

No Capítulo 3, mostramos diversos resultados da literatura sobre conflitos de informação, muitos deles clássicos. Como dissemos anteriormente, algum destes resultados estão mergulhados em contextos diferentes. O teorema de Dawid, por exemplo, simplesmente tenta dar condições suficientes para a conjectura de Lindley [29], sobre como a escolha de distribuições pode tornara posteriori mais ou menos sensível à *outliers*. Neste ponto, temos que muitos resultados que surgiram a partir de Dawid [13] estão relacionados com

análise de sensibilidade em inferência bayesiana, ou mesmo em apenas dar condições suficiente para que a esperança a posteriori não seja influenciada por um *outlier*, como vimos na Seção 3.3 que trata dos resultados obtidos por Pericchi e Sansó [40] e em Haro-Lopéz e Smith [27] que elaboraram um resultado análogo ao de Dawid utilizando distribuições v-esféricas. Contudo, os jargões da teoria de conflitos de informação foram, aparentemente, introduzido por O'Hagan [33]. De fato, muitos dos trabalhos que lidam conflitos de informação de fato têm [33] entre as suas referências. Dentre estes trabalhos, citamos Lucas [30], Goldstein [24], O'Hagan [34], [36] e Andrade e O'Hagan [2]. Em especial, Andrade e O'Hagan [2] e O'Hagan [36] trazem boas discussões sobre o que é um conflito e o que significa resolvê-lo.

No Capítulo 4 discutimos sobre a relação entre as distribuições de cauda pesada e conflitos de informação em inferência bayesiana. Para o teorema de Dawid, mostramos que existe uma relação entre sua primeira condição e as distribuições de cauda longa. Apresentamos ainda duas pequenas discussões. A primeira dizia sobre como resultado e Pericchi e Sansó está relacionado com a classe \mathcal{G} e com o teorema de Dawid e a segunda sobre como a medida *credence* está relacionada com a classe *ERV*. Também demonstramos que a medida *credence* poderia ser aplicada para resolver conflitos nos quais uma das fontes assumem uma estrutura de escala (Teorema 4.10). Embora este resultado possa ser utilizado em situações análogas as do Teorema 3.14, discutimos e exemplificamos situações onde o Teorema 3.14 não poderia ser aplicado.

Ainda no Capítulo 4, apresentamos o Teorema 4.5. Se comparado aos resultados de Dawid [13] e Pericchi e Sansó [40], o Teorema 4.5 é aplicado a uma classe menor de distribuições (a classe *SD*). Contudo, esta classe é bastante utilizada na prática. Além disso, as condições do Teorema 4.5 são mais fáceis de serem verificadas do que as dos Teorema 3.1 e 3.3, uma vez que trocamos os cálculos de integrais de supremos por um gráfico da razão entre as fontes.

Por último, para trabalhos posteriores, gostaríamos de sugerir:

- estudar como as distribuições propensas a *outliers* podem refletir diferentes graus de credibilidade para as fontes;
- verificar se existe uma relação entre as distribuições propensas a *outliers* multivariadas apresentadas em Barne-Delcorix e Gather [6] e conflitos de informação em inferência bayesiana;
- verificar as relações entre conflitos de informação e distribuições multivariadas subexponenciais e de variação regular;

Referências Bibliográficas

- [1] ATHREYA, K. B., NEY V. P. (1972). Branching Processes. Spriger, Berlin
- [2] ANDRADE, J. A., O'HAGAN, A. (2006). Bayesian robustness modeling using regular varying distributions. *Bayesian Analysis*, Vol. 1, No. 1, pp. 169-188
- [3] ANSCOMBE, F. J. (1960). Rejection of outliers. *Technometrics*, Vol. 2, No. 2, pp. 123-147
- [4] BALTRUNAS, A., OMEY, E., VAN GULK, S. (2006). Hazard rates and subexponential distributions, *Publications de l'institut mathématique, Nouvelle série*, Tomo 80(94), pp. 29-46
- [5] BERGER, J. O. (1985). Statistical decision theory and bayesian analysis. New York: Springer-Verlag New York Inc., 2nd edition
- [6] BARMÉ-DELCROIX, M., GATHER, U. (2002). An isobar-surfaces approach to multidimensional outlier-proneness. *Extremes*, Vol. 5, No. 2, pp. 131-144
- [7] BARTLE, R. G. (1964). The elements of real analysis, New York: Wiley
- [8] BOX, G. E. P. (1980). Samplig and Bayes inference in scientific modelling and robustness. *Journal of the Royal Statistical Society. Series A*, Vol. 143, pp. 383-430
- [9] BOX, G. E. P., TIAO, G. C. (1968). A bayesian approach to some outlier problems *Biometrika*, Vol. 55, No. 1, pp. 119-129
- [10] CARLIN, B. P., LOUIS, T. A. (2000). Bayes and empirical methods for data analysis, Chapman & Hall/CRC, 2nd edition
- [11] CHISTYAKOV, V. P. (1964). A theorem on sums of independent positive random variables and its application to branch processes, *Theor. Probability Appl.*, Vol. 9, 640-648.

- [12] CHOVER, J. NEY, P., WAINGER, S. (1972). Functions of probability measures. *J. Analysis Math* **26**, 255-302.
- [13] DAWID, A. P. (1973). Posterior expectations for large observations. *Biometrika*, **60**, 664-667
- [14] de FINETTI, B. (1961). The bayesian approach to the rejection of outliers. *Proceedings of the Fourth Berkeley Symposium on Probability and Statistics*, Vol. 1, pp. 199-210. Berkeley, California: University Press.
- [15] EMBRECHTS P., GOLDIE, C. M. (1980). On closure and factorization theorems for subexponential and related distributions. *Journal of Austral. Math. Soc. Ser. A* Vol.29, pp 243-278
- [16] EMBRECHTS P., GOLDIE, C. M., VERAVERBEKE, N. (1979). Subexponential distribution function and some applications, *Advances in Applied Probability*, Vol. 11, No. 2, pp. 283
- [17] EMBRECHTS P., KLUPPELBERG. C., MIKOSH, T. (1997). Modelling extremal events for insurance and finance. Springer, Berlin
- [18] EMBRECHTS P., OMEY E. (1984). A property of longtailed distributions, *Journal of Applied Probability*, Vol. 21, No. 1, pp. 80-87
- [19] FELLER, W. (1967). On regular variation and local limits theorems, *Fifth Berkeley sympos. math. statist. probab.* V.2. Contributions to probability theory. Part I. Berkeley: University of California Press, 373-388
- [20] FELLER, W. (1971). An Introduction to Probability and its Applications, vol II. New York: Wiley, 2nd edition
- [21] FERGUNSON, T. S. (1961). Rules for rejection of outliers. *Revue de l'Institut International de Statistique*, Vol. 29, No. 3, pp. 29-43
- [22] FENÁNDEZ, C., OSIEWALSKI, J., STEEL, M. F. (1995). Modelling and inference with v -spherical distributions, *Journal of American Society Serie B*, Vol. 90, 1331-1340
- [23] GOLDIE, C. M. (1978). Subexponential distributions and dominated-variation tails. *Journal of Applied Probability*, Vol. 5, No. 2, pp. 440-442
- [24] GOLDSTEIN, M. (1983). Outlier resistant distributions: where does the probability go? *Journal of the Royal Statistical Society, Series B(Methodological)*, Vol. 45, No. 3, pp. 355-357

- [25] GREEN, R. F. (1976). Outlier-prone and outlier-resistant distributions, *Journal of the American Statistical Association*, Vol. 71, No. 354, pp.502-505
- [26] GREEN, R. F. (1974). A note on outlier-prone families of distributions, *The Annals of Statistics*, Vol. 2, No. 6, pp.1293-1295
- [27] HARO-LOPÉZ, R. A., SMITH, A. F. M. (1999). On robust bayesian analysis for location and scale parameters, *Journal of Multivariate Analysis*, No. 70, pp. 30-56
- [28] KLUPPELBERG, C. (1988). Subexponential distributions and integrated tails, *Journal of Applied Probability*, Vol. 25, No. 1, pp. 132-141
- [29] LINDLEY, D. V. (1968). The choice of variables in multiple regression (with discussion) *Journal of the Royal Statistical Society. Series B* Vol. 30, pp. 31-66
- [30] LUCAS, T. W. (1993). When is conflict normal? *Journal of the American Statistical Association* Vol. 88, No. 424, pp. 1433-1437
- [31] MURPHREE, E. S. (1989). Some new results on the subexponential class. *Journal of Applied Probability*, Vol. 26, No. 4, pp. 892-897.
- [32] NEYMAN, J. SCOTT, E. L. (1971). Outlier proneness of phenomena and of related distributions, *Optimizing Methods in Statistics*, New York: Academic press, Inc., 413-430
- [33] O'HAGAN, A. (1979). On outlier rejection phenomena in Bayes Inference. *Journal of the Royal Statistical Society. Series B*, Vol. 41, No. 3, 358-367
- [34] O'HAGAN, A. (1981). A moment of indecision. *Biometrika*, Vol. 68, No. 1, 329-330
- [35] O'HAGAN, A., FORSTER, J. (2004). *Bayes Inference. Kendall's Advance Theory of Statistics*, 2nd edition. Arnold
- [36] O'HAGAN, A. (1990). Outliers and credence for location parameter Inference. *Journal of the American Statistical Association*, Vol. 85, No. 409, pp. 172-176
- [37] PEREIRA, C. A. B. (2006). Estatística aplicada: uma visão genuinamente bayesiana. 51ª *Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria*, 24 a 26 de maio de 2006, Botucatu: Instituto de Biociências, Universidade Estadual Paulista.
- [38] PITMAN, E. J. G. (1980). Subexponential distributions functions. *Journal of Austral. Math. Soc. Ser. A*, Vol. 29, pp. 337-347

- [39] PRATT, J. W. (1960). On interchanging limits and integrals, *The Annals of Mathematical Statistics* Vol. 31, No. 1, pp. 74-77
- [40] PERICCHI, L. R., SANSÓ, B. (1995). A note on bounded influence in bayesian analysis. *Biometrika*, Vol. 82, No. 1, 223-225
- [41] TANG, Q. TSITSIASHVILI, G. (2003). Randomly weighted of subexponential random variables with application to ruin theory. *Extremes*, **6**, 171-188
- [42] TEUGELS, J. L. (1975). The class of subexponential distributions, *The Annals of Probability* , Vol. 3, No. 6, pp. 1000-1011.