NESTOR CATICHA

PROBABILIDADES (NOTAS INCOMPLETAS)

Prólogo

Estas notas são complementares ao curso introdutório Probabilidades 04300xxx.

Os prérequisitos matemáticos não são muitos. Cálculo Diferencial e Integral serão a lingua franca. No nível do curso, Integração significa integral de Riemann. Não precisaremos ir além da idéia de integrar em $\mathbb{R}^n$. A simplicidade matemática não implica que os conceitos serão simples. A interpretação pode ser bastante sutil e é esse o objetivo do curso, fazer o aluno pensar e talvez modificar suas idéias sobre o que significa probabilidade. Rigor matemático não substitui rigor intelectual.

O principal objetivo do curso é que o aluno entre em contato com a idéia de probabilidade como expressão da informação disponível sobre uma possibilidade. Introdução à Teoria de Informação poderia ser um título deste livro, mas não do ponto de vista de engenharia. De um ponto de vista bem mais geral que encontrará aplicações em uma variedade muito grande de áreas da ciência. Estas incluem questões fundamentais em Física, mas também é claro em Estatística e portanto em tratamento de dados empíricos. Outras áreas como Ciência de Computação e Ciência Cognitiva tem tido uma grande influência desta forma de pensar sobre informação.

A primeira parte discute a definição de probabilidades. Todos os estudantes já foram expostos a probabilidades, tanto na linguagem coloquial quanto no curso secundário. Começaremos de forma diferente de outros cursos. Uma forma de proceder consiste em primeiro expor os princípios matemáticos e a partir deles calcular as consequência em aplicações interessantes. Faremos de outra forma. Não sabemos qual é a estrutura matemática que deve ser usada em geral, mas talvez possamos investigar se há casos simples em que podemos concordar com pessoas razoáveis como proceder. Isso dará uma lista de desejos de requisitos que a teoria deve satisfazer. Todas as estruturas que não estejam de acordo com a lista de desejos serão eliminadas. O último candidato em pé será a estrutura desejada. O aluno será convidado a procurar falhas no raciocínio, procurar exceções. Deste tipo de exercício decorrerá a confiança na estrutura final. Em ciência não devemos ser a favor de uma teoria ou sua interpretação, a não ser pelos motivos que decorrem do respeito gerado por ter resistido a todos os embates em que se tentou derrubá-la.

Há outras formas de introduzir probabilidades e aqui me refiro às idéias frequentistas. O leitor não deve esperar uma exposição neutra, onde todos tem mérito e direito a ser ouvidos. O século XX ficou para trás e somente poucos frequentistas restarão no futuro. O objetivo destas notas é apresentar aos estudantes de Física uma visão que tem se mostrado frutífera e tem conquistado cada vez mais adeptos. Aqueles que estão interessados em aplicações e análise de dados terão acesso aos métodos atuais. O uso de técnicas numéricas e do computador não podem ser deixadas de lado e mesmo que não seja o objetivo principal, um pouco desse universo será explorado. O nível do curso é introdutório e a parte formal de Probabilidades como uma parte da Matemática, um ramo da análise funcional e teoria de medida não será explorada.

A idéia de apresentar uma forma de pensar que tem aplicações em uma vasta gama de assuntos, pode levar o leitor a pensar que está na presença de alguém que com um martelo, pensa que todos os problemas são pregos. Ou que estamos apresentado dogmas, dos quais não abriremos mão. No fim talvez não saiba como me defender de tais acusações, exceto alegando que o único ponto sobre o qual serei inflexível será que só podemos acreditar naquilo que a informação e evidência permitem, e só enquanto não surgir informação contraditória. Há outras formas de pensar, por exemplo acreditar em algo porque isso me deixa mais feliz. Mas eu não saberia dar um curso sobre isso.Não faz sentido acreditar em algo que não seja respaldado por informação.

Sumário

1
Teoremas de Regraduação de Cox

Alea jacta est

Júlio César

Introdução: Determinismo Newtoniano ou aleatório?

Júlio César ao cruzar com seu exército o Rio Rubicom quebrou uma regra na República Romana. O exército deviaser mantido longe de Roma. Não havia volta. Ou conseguia o poder ou perdia tudo. Qual seria o desenlace da sua ação? Nem ele sabia e segundo Suetônio teria dito: *Alea jacta est.* A sorte está lançada. Saber estimar as consequências de uma ação é aconselhável para poder decidir que curso tomar. César talvez tenha procedido da seguinte forma. Primeiro fez uma lista das possibilidades à sua frente. Uma decisão é tomada e uma das possibilidades seguidas. Estas poderiam incluir: (Ação I) Continuar na Gália. (Ação II) Fazer uma aliança com Pompeu , (Ação III) Fugir de Pompeu, (Ação IV) Se aposentar, (Ação V) Voltar a Roma com seu exército e lutar contra Pompeu. Historiadores certamente poderiam incluir outras. Como decidir? Supomos que uma escolha foi feita. Quais as consequências? Para cada curso de ação ele deve ter feito uma lista de possibilidades. Suponha que considere tomada a Ação V. Então as consequências poderiam ser (Consequência 1 da Ação V) Vitória total, com a formação do Império e ele como Imperador. (Consequência 2 da Ação V) Derrota total levando à sua morte. (Consequência 3 da Ação V) Guerra Civíl interminável ...etc. Mas não devia acreditar que cada uma das possiblidades teria a mesma chance de ocorrer. A cada consequência de cada Ação, César poderia ter associado um valor numérico indicando sua crença na chance de ocorrer. Veremos que isto será codificado em probabilidades de ocorrência. Mas também poderia ter associado um valor numérico de quão feliz ele seria se efetivamente essa consequência ocorresse. Estes números descrevem o que se chama de utilidade, de cada possibilidade, para o agente Júlio Cesar. Parece óbvio que as utilidades dependem do agente, mas talvez não seja óbvio que as probabilidades também dependam do agente, ou melhor do que este sabe. Resumindo, Júlio César decidiu o seu curso de ação após identificar as possibilidades de ação, das consequências de cada ação, das chances de cada consequência ocorrer, e da utilidade ou felicidade que cada consequência teria. Neste curso não falaremos sobre decisão a partir das utilidades. Atualmente, em geral, este tópico não cabe em cursos de Física. Faremos um estudo sistemático sobre a chances de algo ocorrer sem importar quão feliz você fique com cada possibilidade. O ponto central será definir com cuidado o que queremos dizer com *chances*, como atribuir números e como mudá-los quando recebemos informação.

Teria Júlio César dúvidas sobre sua sorte ou saberia mais que os outros atores do drama? Se soubesse mais talvez estaria jogando um jogo de cartas marcadas enquanto os outros jogariam a cegas. A frase também implica num certo determinismo. Não há nada a fazer. O curso natural das coisas conduzirá os atores. Como observadores verão simplesmente o desenrolar da história.

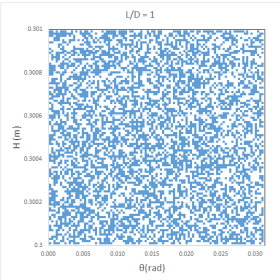


Figura 1.1: Integração numérica das equações de movimento de um modelo Newtoniano de uma "moeda"feita de massas (m) e molas (k). A figura mostra um espaço restrito de condições iniciais. H é a altura da moeda ao ser lançada e θ o ângulo com a horizontal, a moeda é solta do repouso. Nesta figura a altura é "grande"(em relação a mg/k). A estrutura é formada por quatro massas nos vértices do que seria em repouso um retângulo, ligadas por seis molas nas arestas e diagonais . O sistema está restrito a duas dimensões e a cada batida mesa há dissipação de energia. É um modelo de uma moeda ou um cubo simplificado. As simulações foram feitas por Guilherme Galanti e Osame Kinouchi, que gentilmente autorizaram o uso destas figuras.

Há alguma inconsistência em pensar que a as consequências são inevitáveis por um lado, e por outro ficar torcendo para ter sorte? Seria como torcer ao ver a gravação de um jogo de futebol que já foi jogado, mas não sabemos o resultado. Talvez seja um exercício interessante ver grandes jogos do passado sem saber qual jogo é, torcendo para seu time ganhar com direito a ficar tão feliz como quando o jogo é assistido ao vivo.

Todas estas situações são complexas. Comecemos por algo mais simples. Uma das maiores revoluções intelectuais da história da humanidade foi a introdução da Mecânica por Newton. Sabemos que caso fosse necessário temos o formalismo da Mecânica para poder calcular a trajetória de uma moeda. O determinismo Newtoniano permite fazer predições sobre o futuro a partir do estado atual. Por outro lado, os casos mais associados à sorte são o jogo de dados ou um jogo de cara ou coroa com uma moeda. Não é por acaso que a frase de César que teria sido dita em grego menciona $\kappa\upsilon\beta o\sigma$, o cubo ou dado. Estes jogos deram origem a o estudo matemático das probabilidades.

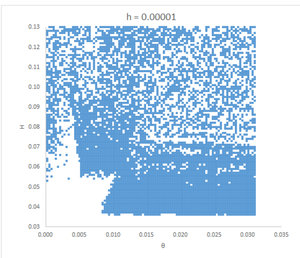


Figura 1.2: Igual à anterior, mas a moeda é solta de uma altura menor, para diferentes ângulos.

Como podemos associar a uma moeda simultaneamente as propriedades de ser um sistema determinista, governado pelas leis de Newton e a condição de exemplo mais usado ao falar de sistemas aleatórios? É necessário ter cuidado com as palavras. O que significa *aleatório*? Teremos todo este curso para atribuir-lhe significado. Em geral, ao ser usado coloquialmente, significa que não é totalmente determinado *a priori* por eventos passados.

As possibilidades do estado da moeda são determinados ao especificar 12 números. 3 dizem respeito à sua posição, por exemplo do centro de massa. Sua orientação é determinada por 3 ângulos.Veja, num livro de Mecânica a definição de ângulos de Euler. Ou senão, simplesmente considere 2 eixos no plano da moeda e um terceiro perpendicular ao plano e as rotações em torno deles.Esse número é duplicado ao levar em conta as suas derivadas temporais (velocidades). A dinâmica em 12 dimensões é dada pelas equações de Newton [1]. É óbvio que as equações não são suficientes para determinar como cairá a moeda. Há muitas maneiras de jogar a moeda, mas só um conjunto de equações. As mesmas equações devem ser

[1] Nem a dedução destas equações e muito menos a sua solução, serão necessárias aqui, mas cabem num curso de Mecânica.

complementadas com diferentes conjuntos de condições iniciais que parametrizam cada trajetória possível. As figuras 1.1 e 1.2 mostram porque não há incompatibilidade nessas duas caracterizações da moeda. Por simplicidade fixamos 10 parâmetros e olhamos o que ocorre quando dois parâmetros são mudados numa certa região. As figuras foram construidas de forma totalmente determinística integrando as equações diferenciais. Cada ponto é colorido de acordo com a face mostrada pela moeda. Azul se cara, branco se coroa. Vemos que a aleatoriedade não está na evolução dinâmica descrita por Newton, mas na ignôrancia que poderiamos ter sobre a condições iniciais. Se ao jogarmos a moeda não tivermos conhecimento muito preciso das condições iniciais, não teremos como prever se o ponto final será azul ou branco. Este é um indício que o conhecimento pode influenciar as probabilidades (que ainda não sabemos o que são) de que caia cara ou cora. Dois agentes apostando neste jogo terão chances diferentes de ganhar se tiverem informações diferentes sobre o modo como a moeda será jogada. Note que para alturas muito pequenas, o poder de predição fica mais forte, pois há regiões grandes com a mesma cor. Faça a experiência. Segure uma moeda com os dedos na posição horizontal. Solte a moeda, sem girá-la, de uma altura de 1 metro, 10 cm, 1 cm , 1mm. Seu poder de prever o que vai ocorrer aumenta. O determinismo é igualmente descrito pelas equações de Newton em todas as condições. A incerteza na previsão tem a ver com a forma como se solta a moeda.

Ainda isto é coloquial e não sabemos o que é probabilidade, informação ou aleatório. O objetivo do que segue é vestir isto com uma estrutura matemática. A história do desenvolvimento das ideias é complexo e não é o interesse destas notas. Porém elas estarão salpicadas de referências a grandes figuras do passado. A história contada, não é certamente como ocorreu, pois isso não sabemos. A seguir discutiremos as idéias que vem de Jakob Bernoulli, Laplace, Maxwell, Kolmogorov, Ramsey, Keynes, Pólya, Cox, Jeffreys, Jaynes entre outros. Começaremos a história no meio contando como R. Cox tentou criar uma extensão da lógica Booleana, com origens na Grécia antiga, para situações de informação incompleta. Ele poderia ter suposto de início que a estrutura matemática era a de probabilidades, mas se recusou a isso. Tentou encontrar essa estrutura e ao descobrir que era ou a teoria de probabilidade ou uma regraduação monotónica trivial, primeiro se convenceu da impossibilidade de escapar dessa estrutura e segundo forneceu uma sólida interpretação para o que queremos dizer com informação e como molda nossas crenças e para o que queremos dizer com probabilidades.

Há vários exemplos de tentativas de axiomatizar extensões da lógica a situações de informação incompleta. Savage e Lindley são exemplos importantes, mas seu objetivo é descrever o processo de tomada de decisão e isso leva a considerar utilidades. O caminho que escolhemos leva à mesma estrutura de probabilidades deixando claro que decisões é um capítulo a parte. O objetivo de um físico é descrever a natureza fazendo previsões e não tomando decisões.

Informação completa ou incompleta

Há muitas definições matemáticas possíveis que poderiam ser usadas na tentativa de formalizar o conceito coloquial de informação. Uma forma de avançar, que é bastante comum em ciência, começa por definir matematicamente algo e depois tentar interpretar as fórmulas matemáticas para mostrar que esta interpretação esta de acordo com algumas das características que podemos atribuir ao conceito coloquial de informação que temos.

Em lugar de começar por uma estrutura matemática pré-escolhida para servir de ferramenta de análise, começamos por uma interpretação e depois encontramos a estrutura matemática que se adapte à interpretação. A interpretação passa por estabelecer em alguns casos particulares suficientemente simples, tais que haja algum tipo de consenso, o quê deveria resultar da teoria. É possível que este procedimento pareça novo ao leitor e será surpreendente quantos resultados serão extraidos deste método e do rigor matemático com que a teoria se vestirá. Como este procedimento permite saber mais claramente do que estamos falando e do que não estamos, achamos que esta é atualmente a melhor maneira de introduzir a teoria de informação.

Pode parecer estranho para o estudante de Física que o elemento principal a seguir seja a idéia de asserção, isto é, uma frase que em princípio é uma proposição que se apresenta como verdadeira. Mas a matemática é um tipo de linguagem que tem a vantagem de permir a quem a usa ser muito cuidadoso com o que diz. Denotaremos asserções por letras $A, B, C...a, b, c...$ Uma frase pode ser julgada correta ou não de várias maneiras. Podemos pensar se é correta do ponto de vista da sua estrutura gramatical ou sintática. Não é isto que queremos fazer e consideraremos as asserções a seguir suficientemente bem formadas [2] . Queremos analisar seu conteúdo informacional, se realmente a podemos creer verdadeira. Mas quando se diz " a massa de Saturno está entre m_1 e m_2" ou "... entre m_3 e m_4" estamos usando asserções diferentes e a tarefa é determinar quanto acreditamos que uma ou a outra sejam verdade e aqui o estudante reconhece a linguagem científica.

[2] Embora o formalismo a ser introduzido também possa ser usado nesta direção, mas não agora.

Consideremos a asserção "Existem zumbies". Isto é verdade? Se o contexto for o de filmes gravados em Pittsburgh na década de oitenta, a resposta será uma. Se for no mundo real, outra. Nenhuma asserção sozinha pode ser analisada, no que diz respeito a ser verdadeira ou não, de forma independente do resto do universo conceitual. Ela será julgada verdadeira ou não quando analisada dentro de um contexto. A informação trazida por uma asserção C, será usada para atribuir um grau de verdade à asserção A, ou seja dentro do contexto C. Poderiamos chamar esse grau de, por exemplo, probabilidade de que A seja verdade se C for dada. Mas fazendo isto estariamos definindo de antemão que a ferramenta matemática apropriada para descrever informação é a teoria de probabilidades. Isto parece bem razoável mas não escapa às críticas acima e permite que outra ferramenta matemática seja usada por simplesmente expressar o gosto de outras pessoas ou a facilidade de uso em determinados problemas práticos com a mesma justificativa: *parece razoável, eu gosto, funciona, é prático.* Não descartamos o uso de outras ferramentas matemáticas, mas queremos deixar claro que estas poderão ser vistas como aproximações mais ou menos adequadas de uma estrutura que unifica e tem um posição diferente. O **objetivo** deste capítulo é mostrar que a escolha da teoria de probabilidades como a ferramenta matemática adequada para tratar informação é muito mais do que simplesmente conveniente. A teoria de probabilidades segue porque é a extensão da lógica a situações de informação incompleta. Mas até aqui não sabemos o que é *lógica, informação* nem *incompleta.*

A análise da lógica remonta a Aristóteles e passa por Boole no século XIX, que contribuiu para que a lógica pudesse ser representada em linguagem matemática [3]. Uma lógica envolve (i) um conjunto de proposições supostas verdadeiras, (ii) um método de dedução para estabelecer a validade de argumentos e (iii) um método para estabelecer invalidades.

[3] Veja para uma comparação: Aristotle's Prior Analytics and Boole's Laws of Thought, John Corcoran, History and Philosophy of Logics 2003.

Um argumento lógico é composto por duas partes. Um conjunto de asserções, chamadas as premissas e uma única asserção chamada de conclusão. Um argumento é válido se a conclusão pode ser obtida aplicando as regras (ii) e (iii).

Se a informação em C não permite a certeza sobre a verdade de A então diremos que a crença que temos sobre A esta baseada em informação incompleta. Em casos particulares poderá ocorrer que dado C como verdade, possa ser concluido com certeza que a asserção A é verdadeira ou ainda em outros casos que é falsa. Quando não há alternativa para a conclusão, quando ela segue por força da informação disponível, dizemos que a conclusão é racional ou lógica. Dizemos que estamos frente a casos de raciocínio dedutivo. Nestes casos a informação disponível é *completa* pois nada falta para ter certeza.

Exemplos de informação completa são dados pelos silogismos Aristotélicos: suponha que recebemos a informação contida em $C =$ "$A \rightarrow B''$, isto é, A implica B. Traduzindo, isto significa "se souber que A é certamente verdade, segue que a proposição B também o é." Dado isso, o que podemos dizer sobre B? Nada com certeza, mas se também recebemos a informação adicional A, isto é, que "A é Verdade", então segue B, ou seja "B é Verdade".

Outro caso de informação completa, novamente no contexto C, ocorre quando é dado como verdade a negação $\overline{B}$ ou seja "B é Falso". Segue $\overline{A}$, isto é, que "A é Falso" como conclusão inescapável. Note que se A não fosse falso, B não poderia sé-lo.

Nas condições que $C =$ "$A \rightarrow B''$ e "A é Falso", o quê pode ser concluido? Do ponto de vista lógico clássico nada podemos concluir sobre B. Da mesma forma se for dada a informação "B é Verdade", nada podemos concluir sobre A. Estamos frente a casos de informação incompleta e a lógica clássica não serve para chegar a uma conclusão. Não é possível deduzir nada. A indução, o que quer que isto seja, e que será discutido mais à frente, será necessária para avançar. [4]

A forma dedutiva da lógica permite somente tres tipos de respostas, *sim, não* e *não segue*[5]. A indução nos força ou permite dividir esta última em várias possibilidades e os casos extremos nesse espectro são aqueles onde havendo certeza absoluta, haverá portanto a força da dedução. Podemos falar então sobre quais das alternativas intermediárias é mais razoável acreditar com base no que sabemos. Nota-se então a necessidade de estender a lógica para poder tratar de forma racional casos de informação incompleta. Richard T. Cox, ao se defrontar com este problema por volta da década de 1940, decidiu, como dito acima, estabelecer um conjunto de desejos ou *desiderata*[6] que a teoria deveria satisfazer, e estes serão então os axiomas da extensão da lógica. Aqui podemos discordar, propor outros desejos ou axiomas, mas uma vez aceitos serão provados os teoremas de reparametrização de Cox que mostram que a teoria de probabilidade é a ferramenta para o tratamento de forma racional de situações de informação incompleta. O surpreendente disto é que surge a teoria das probabilidades como a forma para lidar de forma *racional*[7]com a informação e que corremos riscos de ser inconsistentes caso a regras de manipulação de probabilidades não sejam seguidas. Segue que **não há probabilidades que não sejam condicionais**, embora às vezes simplesmente a linguagem esqueça de deixar explícitas as relações de condicionalidade [8]. A amplidão da aplicabilidade da teoria que emerge é impressionante e por exemplo, quando o tipo de asserção for limitado àqueles entendidos em teoria de conjuntos as regras de manipulação serão não mais nem menos que aquelas ditadas pelos axiomas de Kolmogorov. Também veremos que emerge uma relação natural entre probabilidade e freqüência e ficará claro de que forma estes conceitos estão ligados e mais importante, de que forma são distintos.

Desiderata à la Cox

É interessante notar que os axiomas de Cox descritos por Jaynes não são exatamente iguais aos que Cox apresenta no seu livro *The algebra of probable inference*. A exposição de Jaynes é muito mais simples. Cox, por sua vez, esclarece sua dívida com J. M.Keynes e seu livro *A treatise on Probability*, que deve muito a Laplace e Bernoulli, a Frank P. Ramsey e George Pólya. A exposição de Jaynes teve uma grande influência, mas ainda recebeu críticas e complementos [9]. Eu seguirei a apresentação de A. Caticha, que é mais completa e clara, mas farei algumas pequenas mudanças[10].

A maneira de construir a teoria está baseada na seguinte forma de pensar bastante simples. Queremos construir uma teoria geral para a extensão da lógica nos casos de informação incompleta. Se ela for suficientemente geral, deverá ser válida em casos particulares. Se o caso for suficientemente simples, então podemos saber qual é o resultado esperado que não viole expectativas razoáveis. Poderia ocorrer

[4] Segundo Harold Jeffreys em seu livro*Theory of Probability*, Bertrand Russell disse que "induction is either disguised deduction or a mere method of making plausible guesses". Jeffreys diz que "é muito melhor trocar a ordem dos dois termos e que muito do que normalmente passa por dedução é indução disfarçada, e que até alguns dos postulados de *Principia Mathematica* foram adotados por motivações indutivas" (e adiciona, são falsos). Com o tempo o próprio Russell mudou de posição, dobrado pela evidência (?) e diz no fim da sua autobiografia: "I was troubled by scepticism and unwillingly forced to the conclusion that most of what passes for knowledge is open to reasonable doubt". Sobre indução disse ainda: "The general principles of science, such as the belief of the reign of law, and the belief that every event must have a cause, are as completeley dependent on the inductive principle as are the beliefs of daily life."(On Induction)

[5] Nem o leitor nem o autor destas notas deve neste momento ceder à tentação de discutir lógicas de um ponto de vista mais geral. Precisamos um subconjunto de Lógica proposicional, não muito mais que lógica Aristotélica, como exposta por George Boole. Talvez caiba aqui a desculpa "I have not worked out the mathematical logic of this in detail, because this would, I think, be rather like working out to seven places of decimals a result only valid to two. My logic cannot be regarded as giving more than the sort of way it might work". Frank P. Ramsey (1926) "Truth and Probability", in Ramsey, 1931, The Foundations of Mathematics and other Logical Essays, Ch. VII, p.156-198, editado por R.B. Braithwaite, 1999 electronic edition.

[6] *Desiderata*: as coisas desejadas, em Latim. Termo usado em filosofia para denotar um conjunto de propriedades essenciais de alguma estrutura. Alguns ficam tentados a chamar *axiomas*.

[7] Ao leitor que demande uma definição de racional, podemos dizer que pelo menos não queremos ser manifestamente irracionais. Não acredito que haja uma definição de consenso sobre o que é ser racional. Há consenso porém em apontar alguns casos de irracionalidade.

[8] A maior fonte de erros será devido a falhas na especificação cuidadosa das asserções condicionantes. Aparentemente a notação $a|b$ com a a a asserção a ser analisada e b a asserção condicionante é devida a John Maynard Keynes, no seu Tratado.

[9] Tribus, A. C

[10] Notem que há lugar ainda para avanços nestes primeiros passos. Tentem encontrar defeitos, generalizações, melhores argumentos.

que ao analisar um número de casos particulares sejam reveladas as inconsistências entre eles, nesse caso não poderemos chegar a uma teoria geral. Mas pode ser que os casos particulares sirvam para restringir e determinar a teoria geral. [11] Isto é o que mostraremos a seguir.

[11] Este comentário parece trivial, mas o uso que será dado a seguir é totalmente não trivial. Neste contexto de probabilidades foi colocado primeiro por J.Skilling, mas não de forma explicita. O destaque a este procedimento apareceu por primeira vez no livro de A. Caticha que o chamou de indução eliminativa. Usaremos novamente este estilo de fazer teoria ao introduzir o conceito de entropia.

Em primeiro lugar queremos falar sobre uma asserção A no caso de informação incompleta. Nos referimos então à crença ou plausibilidade de A ser verdade dado B e a denotamos pelo símbolo $A|B$ que lemos "a plausibilidade de A dado B" ou ainda de "... de A condicionada a B".

Por que não à "probabilidade de A dado B"? Porque já existe uma teoria matemática de probabilidade e não sabemos se será a estrutura matemática que emergirá desta análise. Poderiamos usar outras palavras, mas crença ou plausibilidade são conhecidas o suficiente para serem úteis neste contexto e não tem por agora o problema de ser definidas formalmente . A Desiderata que segue tem cinco desejos denotados $DP_1...DP_5$ e é um bom exercício tentar mostrar que não fazem sentido. Se você conseguir e convencer outros terá feito uma grande contribuição, se não terá mais respeito pelas conclusões que seguem.

DP_1 Representação de crenças e transitividade

Queremos analisar o primeiro caso simples que lida com o conceito de *mais plausível*. Se A dado B é mais plausível do que A dado C escrevemos $A|B \succ A|C$. Suponha ainda que $A|C \succ A|D$. Queremos, para seguir o uso cotidiano da linguagem, impor que A dado B seja mais plausível que A dado D.

Temos assim nosso primeiro desejo, a plausiblidade deverá satisfazer a transitividade:

- DP_1: Se $A|B \succ A|C$ e $A|C \succ A|D$ então deve ser o caso que $A|B \succ A|D$

Além disso, dadas duas crenças podemos imaginar que há outra asserção intermediária.

Isto é fácil de satisfazer se impusermos:

- A plausibilidade $A|B$ deverá ser representada por um número real.

Podemos satisfazer este tipo de ordenamento representado crenças com números racionais. A escolha de números reais permite usar integrais, o que não é pouco, pois fazer somas é difícil. Note que sempre usamos integrais em fisica, mesmo que o espaço tenha uma estrutura subjacente (desconhecida atualmente mas que poderia ser na escala de e.g 10^{-31} m). Não sabemos se tem, mas nos modelos para o *mundo* usados em Mecânica, os pontos do espaço e tempo vivem numa variedade real.

Dados

$$A|B > A|C$$

e

$$A|C > A|D,$$

segue imediatamente, uma vez que são números reais, que

$$A|B > A|D,$$

de acordo com o desejo DP_1. Dizer que alguma coisa é um número real nos dá imediatamente a transitividade, mas não diz nada sobre que número deve ser atribuido, nem sobre como mudá-lo se a informação condicionante passa de B para C. Também não diz que a representação das crenças seja única. Uma mudança dos números estritamente monotonica crescente não mudará a ordem. Isto levará a que há familias de atribuições de números que representam a ordem da mesma forma.

DP_2 Asserções compostas:

Através de certas operações e de diferentes asserções podemos criar asserções compostas. Exemplos de operadores são a negação, o produto e a soma lógicos.

- A **negação** de A é denotada por $\overline{A}$.
- O **produto** ou conjunção de duas asserções é uma terceira asserção, há diferentes notações equivalentes possíveis: AB, $A \wedge B$ ou ainda A **e** B.
- A **soma** ou disjunção de duas asserções é uma terceira asserção, que constuma ser denotada por $A + B$ ou $A \vee B$, ou ainda A **ou** B.

A tabela 1.1 mostra a tabela verdade para as operações de soma e produto lógico, onde V = Verdade e F = Falso. Note que as últimas duas colunas, colocadas aqui para futura referência, mostram que $\overline{A+B}$ e $\overline{A}\,\overline{B}$ são iguais.

A	$\overline{A}$	B	$A+B$	AB	$\overline{A+B}$	$\overline{A}\,\overline{B}$
V	F	V	V	V	F	F
V	F	F	V	F	F	F
F	V	V	V	F	F	F
F	V	F	F	F	V	V

Tabela 1.1
Tabela verdade para a negação e algumas asserções compostas.

Isso significa que $A + B = \overline{\overline{A}\,\overline{B}}$ portanto o conjunto de operações negação e conjunção permite construir a disjunção de asserções.

Ao falar de silogísmos introduzimos a operação $\Rightarrow$ que significa implicação. Se é verdade que $A \Rightarrow B$, significa que se A é verdade segue B. Isto não é um novo operador pois é equivalente dizer que C é verdade para $C = \overline{A \wedge \overline{B}}$.

Suponha que haja um método, usando a teoria geral que procuramos e ainda não temos, de analisar a plausibilidade de uma asserção composta por várias asserções através de conjunções ou disjunções ou negações. Esperamos que a plausibilidade possa ser expressa em termos da plausibilidade de asserções mais simples. Talvez haja mais de uma forma de realizar essa análise. Queremos então que:

- DP_2: Se a plausibilidade de uma asserção puder ser representada de mais de uma maneira, pela plausibilidade de outras asserções, todas as formas deverão dar o mesmo resultado.

Há várias formas de usar a a palavra *consistência*. Aqui a usamos da seguinte forma. Impor que duas formas de análise devam dar o mesmo resultado não garante a consistência da teoria geral, no entanto uma teoria onde isso não ocorra será inconsistente. Usamos consistência no sentido de não manifestamente inconsistente, que é o que DP_2 acima declara.

DP_3 *Informação completa*

Um tratamento geral de situações de informação incompleta deve abarcar os casos particulares de informação completa. Então olhemos para casos simples em que há informação completa.

O mais simples é $a|a$ que é a plausibilidade de algo que sabemos ser verdade, para qualquer a.

Se $a|bc$ e $b|ac$ representam a plausibilidade de algo que sabemos ser falso com certeza, chamamos a e b de mutuamente exclusivos na condição c. Poderia ser que hajam falsidades absolutas mais falsas que outras falsidades absolutas; ou verdades absolutas mais verdadeiras que outras verdades absolutas, mas achamos razoável impor

- DP_3 : Existem dois números v_v e v_f tal que para todo a, $a|a = v_v$ e para a e b mutuamente exclusivos $a|b = v_f$.

Não sabemos que valores dar para v_v ou v_f, mas supomos o mesmo valor em todos os casos que tenhamos certeza de verdade ou falsidade. Este desejo inclui também a negação de uma asserção, pois a asserção e sua negação são mutuamente exclusivos, e estamos dizendo que $\overline{a}|a = v_f$ para qualquer a.

Usaremos frequentemente a propriedade de um conjunto de asserções $\{a_i\}_{i=1...K}$ serem mutuamente exclusivos sob condições c, que vale se para qualquer i, j diferentes $a_i|a_jc = a_j|a_ic = v_f$

DP_4 *Soma e* DP_5 *Produto*

Como sugerido na tabela 1, todo operador na álgebra Booleana pode ser representado pelas operações conjunção a *e* b (denotada ab ou $a \wedge b$) e negação de a (denotada por $\overline{a}$) [12], isto é, o produto e a negação lógicas. A soma lógica pode ser obtida usando $a \vee b = \overline{\overline{a}\,\overline{b}}$. Precisamos então analisar a plausibilidade de asserções compostas

[12] Este conjunto não é mínimo, mas é útil e claro.

usando esses operadores em termos das plausibilidade de asserções mais simples. Já que este conjunto de operadores é completo, esperamos que só tenhamos que analisar estes dois operadores, conjunção e negação. Mas é mais fácil, olhar para a conjunção e a disjunção, e junto com DP_3 obteremos a forma geral de tratar a negação.

Agora olhamos para a disjunção ou soma lógica. Novamente c se refere à informação subjacente e estamos interessados na plausibilidade $y = a \vee b|c$. Há 4 plausibilidades que serão interessantes para esta análise:

$$x_1 = a|c,\; x_2 = b|c,\; x_3 = b|ac,\; x_4 = a|bc. \tag{1.1}$$

É importante notar que todas estas plausibilidades são condicionadas a c, a informação que por hipótese é suposta verdadeira. Além disso podem ser condicionadas a outras asserções relevantes e as únicas disponíveis são a e b por separado. Não tem sentido considerar ab como parte do condicionante. Deve haver uma dependência entre $a \vee b|c$ e algum subconjunto de $\{x_i\} = \{x_1, x_2, x_3, x_4\}$, então

- DP_4: Regra da Soma: Deve existir uma função F que relaciona $a \vee b|c$ e algum subconjunto de $\{x_i\}$ e não deve tomar um valor constante, independente dos valores de $\{x_i\}$.

É claro que trocando soma por produto parece razoável desejar:

- DP_5: Regra do Produto. Deve existir uma função G que relaciona $ab|c$ e algum subconjunto de $\{x_i\}$ e não deve tomar um valor constante, independente dos valores de $\{x_i\}$.

Como F e G representam a plausibilidade de asserções (compostas), também devem tomar valores reais. Além disso não impomos nada além de que dependam em algumas, se não todas, as variáveis $\{x_i\}$. Parece natural exigir que não tenham valores constantes, pois senão a todas as asserções compostas lhes seria atribuido o mesmo número. Para facilitar as deduções também imporemos diferenciabilidade até segunda ordem com repeito a quaisquer dois argumentos. Isto não é necessário, mas as provas ficam mais longas e no fim o resultado vem na forma de funções diferenciáveis.

Porque um subconjunto? Qual subconjunto? Todos? Como decidir? Há 11 subconjuntos de dois ou mais membros: Seis ($\frac{4!}{2!2!}$) pares (x_i, x_j), quatro ($\frac{4!}{3!1!}$) triplas (x_i, x_j, x_k) e o conjunto inteiro (x_1, x_2, x_3, x_4). Analisaremos casos particulares em que é fácil ver que alguns subconjuntos levam a resultados absurdos. Do ponto de vista axiomático poderiamos adicionar estes casos particulares à lista de desejos.

Consequências da Lista de Desejos

Parece difícil que desta lista $DP_1 ... DP_5$ surja uma estrutura matemática, quanto mais única. Ou como veremos, essencialmente única a menos de regraduações montônicas que não alteram a ordem das crenças. Talvez o que será surpreendente para o leitor, é que seja a teoria de probabilidades. A estrutura matemática aparecerá analisando as restrições nas funções F e G impostas pelos desejos.

A regra da soma

Começamos com a disjunção $a \vee b|c$ e a função F. Primeiro consideramos a e b mutuamente exclusivos, mas depois veremos que isto permitirá analisar o caso geral. Sob esta restrição $a|bc = b|ac = v_f$ para qualquer c por DP_3. Logo

$$a \vee b|c = F(a|c, b|c, a|bc, b|ac) = F(a|c, b|c, v_f, v_f),$$

mas esta é uma função de apenas duas variáveis, e da constante desconhecida v_f:

$$a \vee b|c = f(a|c, b|c).$$

Para avançar olhamos para asserções compostas mais complexas, que podem ser analisadas de mais de uma maneira, que pelo desejo DP_2, devem dar o mesmo resultado. Para três asserções a, b e c mutuamente excludentes nas condições d, duas maneiras equivalentes de escrever a disjunção das três são $(a \vee b) \vee c|d = a \vee (b \vee c)|d$ o que permite usar a função f duas vezes

$$\begin{aligned} a \vee (b \vee c)|d &= f(a|d, f(b|d, c|d)) \\ (a \vee b) \vee c|d &= f(f(a|d, b|d), c|d) \end{aligned}$$

ou em notação óbvia, f satisfaz

$$f(x, f(y, z)) = f(f(x, y), z) \tag{1.2}$$

chamada equação da associatividade, primeiramente estudada por Abel no contexto de teoria de grupos. Pode se provar [13] que para toda solução de 1.2, existe um bijeção ϕ, dos reais nos reais, que tomaremos como crescente, e portanto será estritamente monotonica crescente, tal que

$$f(x, y) = \phi^{-1}(\phi(x) + \phi(y)). \tag{1.3}$$

Para o leitor bastará mostrar neste ponto, que a expressão 1.3 é uma solução da equação 1.2.

Agora um ponto central: podemos *regraduar*, usando ϕ, as atribuições de plausibilidade e não mais falar dos números do tipo $a|d$ mas de números $\phi(a|d)$. Por ser uma bijeção, resulta que a ordem de preferências não se altera, se antes as crenças sobre as asserções tinham uma certa ordem, depois da regraduação, o ordenamento da representação numérica das crenças é o mesmo. É importante ver que a função ϕ é estritamente monotonica: se $x > y$ segue que $\phi(x) > \phi(y)$, sem poder haver igualdade. Isto significa que asserções com crenças diferentes são mapeadas em valores ϕ diferentes. Caso ocorresse a possibilidade de igualdade, antes da regraduação teriamos uma separação de preferências e depois da regraduação poderiamos ter confusão entre asserções mapeadas no mesmo valor de ϕ. [14] Continuamos sem saber que números são esses, mas avançamos a ponto de poder dizer que para quaisquer eventos mutuamente exclusivos a crença da disjunção, uma asserção composta pode ser expressa em termos das crenças nas asserções mais simples:

$$\phi(a \vee b|d) = \phi(a|d) + \phi(b|d). \tag{1.4}$$

[13] Para condições em f ver Aequationes mathematicae 1989, Volume 37, Issue 2-3, pp 306-312 *The associativity equation revisited* R. Craigen, Z. Páles, ou o livro Aczél, J. (1966), *Lectures on functional equations and their applications, Mathematics in Science and Engineering* 19, New York: Academic Press,

[14] Veja A. Patriota onde as condições sobre f são relaxadas e as consequências de aceitar soluções não estritamente monotonicas são consideradas.

No caso particular que $d = \overline{a}$, isto significa

$$\begin{aligned} \phi(a \vee b|\overline{a}) &= \phi(a|\overline{a}) + \phi(b|\overline{a}) & (1.5) \\ \phi(b|\overline{a}) &= \phi(a|\overline{a}) + \phi(b|\overline{a}) & (1.6) \end{aligned}$$

pois a crença $\phi(a \vee b|\overline{a})$ é equivalente à crença $\phi(b|\overline{a})$. Segue que

$$\phi(a|\overline{a}) = \phi(v_f) = \phi_f = 0. \qquad (1.7)$$

Embora modesto, eis o primeiro resultado númerico:

O valor regraduado da certeza da falsidade é zero.

Mas e se não forem mutuamente exclusivos? O interessante é que o resultado anterior serve para o caso geral, mas precisamos usar o truque de escrever

$$a = (a \wedge b) \vee (a \wedge \overline{b}) \;\text{ e }\; b = (b \wedge a) \vee (b \wedge \overline{a}). \qquad (1.8)$$

O leitor deve mostrar que as relações acima são verdadeiras, no estilo da tabela 1. Podemos escrever $a \vee b$ como uma disjunção de asserções mutuamente exclusivas:

$$\begin{aligned} a \vee b &= \Big[(a \wedge b) \vee (a \wedge \overline{b})\Big] \vee [(b \wedge a) \vee (b \wedge \overline{a})] \\ &= (a \wedge b) \vee (a \wedge \overline{b}) \vee (b \wedge \overline{a}) \end{aligned}$$

assim a equação 1.4, que descreve a soma de asserções mutuamente exclusivas, pode ser usada, levando a

$$\begin{aligned} \phi(a \vee b|d) &= \phi(a \wedge b|d) + \phi(a \wedge \overline{b}|d) + \phi(b \wedge \overline{a}|d) \\ &= \phi(a \wedge b|d) + \phi(a \wedge \overline{b}|d) + \phi(b \wedge \overline{a}|d) + [\phi(a \wedge b|d) - \phi(a \wedge b|d)] \end{aligned}$$

onde, na última linha adicionamos e subtraimos o mesmo número. Chamamos pela ordem os termos do lado direito da equação acima de 1, 2,...5. Usando novamente a equação 1.4 para asserções mutuamente exclusivas, juntando 1 com 2 , e 3 com 4:

$$\begin{aligned} \phi(a \vee b|d) &= \phi\left((a \wedge b) \vee (a \wedge \overline{b})|d\right) + \phi\left((b \wedge \overline{a}) \vee (a \wedge b)|d\right) - \phi(a \wedge b|d) \\ &= \phi(a|d) + \phi(b|d) - \phi(a \wedge b|d), & (1.9) \end{aligned}$$

que segue das relações da equação 1.8. Temos um dos resultados principais para lidar com asserções compostas por somas de asserções

$$\boxed{\boldsymbol{\phi(a \vee b|d) = \phi(a|d) + \phi(b|d) - \phi(ab|d)}}$$

Mas ainda não acabamos pois não sabemos o que fazer com $\phi(ab|d)$, que olharemos a seguir.

Regra do produto: quais as variáveis relevantes?

Queremos expressar

$$y = \phi(ab|c) \tag{1.10}$$

em termos da função ainda por determinar G e de algum dos subconjuntos de $\{x_i\}$. Lembramos a notação $x_1 = a|c$, $x_2 = b|c$, $x_3 = b|ac$, $x_4 = a|bc$. Tribus sugeriu a análise das 11 possibilidades para verificar que só há duas que sobrevivem a casos extremos. Seguimos A. Caticha, pois corrige vários erros anteriores. Os dois conjuntos sobreviventes são (x_1, x_3) e (x_2, x_4). Note que se o primeiro deles fosse um dos sobreviventes, o segundo também deveria ser pela simetria trazida pela comutatividade do produto lógico. Cox já parte da conclusão de que estes dois subconjuntos são os adequados. O exercício que segue mostra que ele tinha razão, mas retira a arbitrariedade aparente, de fazer a escolha sem analisar outros candidatos.
Vejamos como chegar a esta conclusão (novamente seguimos AC) Os 11 casos são reduzidos a 7 por simetria:

1. $y = G(\phi(a|I), \phi(b|I))$ (1 possibilidade)
2. $y = G(\phi(a|I), \phi(a|bI))$ (2 possibilidades $a \leftrightarrow b$)
3. $y = G(\phi(a|I), \phi(b|aI))$ (2 possibilidades $a \leftrightarrow b$)
4. $y = G(\phi(a|bI), \phi(b|aI))$ (1 possibilidade)
5. $y = G(\phi(a|I), \phi(b|I), \phi(a|bI))$ (2 possibilidades $a \leftrightarrow b$)
6. $y = G(\phi(a|I), \phi(a|bI), \phi(b|aI))$ (2 possibilidades $a \leftrightarrow b$)
7. $y = G(\phi(a|I), \phi(b|I), \phi(a|bI), \phi(b|aI))$ (1 possibilidade)

Caso 1 Mostraremos que $y = a \wedge b|I = G(\phi(a|I), \phi(b|c)) = G(x_1, x_2)$ não funciona pois nao satisfaz o esperado em um caso simples. Porque não serve o subconjunto mais óbvio (x_1, x_2)? Primeiro vejamos que não segue o bom senso. Seja *a*= 'Helena usa um tenis esquerdo vermelho' enquanto que *b*= ' Helena usa um tenis direito preto' . A plausibilidade dessas duas asserções será julgada dada a seguinte informação *c*= 'Helena gosta de tenis pretos e de tenis vermelhos', e talvez seja possível concluir que as duas asserções são bastante plausíveis. Mas se tivessemos $y = G(x_1, x_2)$ poderiamos ser levados a pensar que 'Helena usa um tenis esquerdo vermelho e um tenis direito preto' é bastante plausível. Posso acreditar bastante nas duas asserções, mas não que seja muito plausível que use um tenis de cada cor ao mesmo tempo. Devemos rejeitar esta forma para G.
Para convencer os incrédulos no exposto acima, um argumento mais formal: Suponha que $a|d = a'|d'$ e $b|d = b'|d'$, mas que embora a e b sejam mutuamente exclusivos, a' e b' não o sejam. Neste caso teriamos que

$$\phi(a'b'|d') = G(\phi(a'|d'), \phi(b'|d')) = G(\phi(a|d), \phi(b|d)) = \phi(ab|d) = \phi_F = 0.$$

E isto ocorreria para qualquer par de asserções não mutuamente exclusivas (a', b'), pois sempre poderiamos supor um caso auxiliar (a, b) adequado e portanto teria um valor constante, independende das asserções sob consideração. Insistindo, suponha que Bruno joga uma moeda contra o teto, bate no ventilador e cai. A Helena pega outra moeda e faz o mesmo. Temos a mesma crença que saia cara ou coroa nas duas situações. Chamamos c_B a asserção que saiu cara no primeiro experimento e c_H no segundo. Achamos razoável escrever

$$\phi(c_B|I) = \phi(c_H|I) \quad \text{e} \quad \phi(\bar{c}_B|I) = \phi(\bar{c}_H|I)$$

E também achamos impossível que $c_B\bar{c}_B|I$ seja verdade, não pode ser verdade que Bruno obteve cara e coroa nessa única jogada. Mas seriamos levados a pensar que

$$\begin{aligned} \phi(c_B\bar{c}_H|I) &= G(\phi(c_B|I), \phi(\bar{c}_H|I)) \\ &= G(\phi(c_B|I), \phi(\bar{c}_B|I)) = \phi(c_B\bar{c}_B|I) = 0 \end{aligned} \quad (1.11)$$

que significaria que se Bruno obteve cara, Helena não poderia ter obtido coroa.

Caso 2

Para qualquer asserção $b|I$, sob quaisquer condições teriamos

$$\phi(b|I) = \phi(Ib|I) = G(\phi(I|I), \phi(I|bI)) = G(\phi_V, \phi_V) = \text{constante.}$$

Um método que atribui o mesmo valor numérico a todas as asserções não pode ser aceitável.

Caso 3 Para o caso $y = G(\phi(a|I), \phi(b|aI))$ e a alternativa $G(\phi(b|I), \phi(a|bI))$ ninguém tem encontrado casos que se oponham ao bom senso. Este será o único candidato a sobreviver e será a pedra de sustentação a toda a teoria que segue. Não analisaremos as consequências disto agora. Ainda falta eliminar os outros candidatos e posteriormente encontrar a forma específica de G.

Caso 4 Se $y = G(\phi(a|bI), \phi(b|aI))$ somos levados a algo inaceitável considerando que para qualquer asserção b teriamos

$$\phi(b|I) = \phi(bb|I) = G(\phi(b|bI), \phi(b|bI)) = G(\phi_v, \phi_v) = \text{constante}$$

independente de b. Novamente a crença sobre a plausibilidade de uma asserção seria independente da asserção.

Caso 5 $y = G(\phi(a|I), \phi(b|I), \phi(a|bI))$. Este caso é mais complicado de analisar. Mostraremos, no entanto que se reduz a algum dos casos anteriores. Ainda consideraremos a conjunção de mais de duas asserções , $abc|I$, que pode ser escrito de duas formas diferentes $(ab)c|I = a(bc)|I$, portanto, considerando a primeira forma obtemos

$$\begin{aligned} \phi((ab)c|I) &= G(\phi(ab|I), \phi(c|I), \phi(ab|cI)) \\ &= G(G(\phi(a|I), \phi(b|I), \phi(a|bI)), \phi(c|I), G(\phi(a|cI), \phi(b|cI), \phi(a|bcI)) \\ &= G(G(x, y, z), u, G(v, w, s)). \end{aligned} \quad (1.12)$$

Para a segunda, com as mesmas definições das variáveis $x, y...$, obtemos

$$\begin{aligned}\phi(a(bc)|I) &= G(\phi(a|I), \phi(bc|I), \phi(a|bcI)) \\ &= G(\phi(a|I), G(\phi(b|I), \phi(c|I), \phi(b|cI), \phi(a|bcI)) \\ &= G(x, G(y, u, w), s) \qquad (1.13)\end{aligned}$$

Notamos as duas maneiras de escrever a mesma coisa. Repetimos que por DP_2 que declarava que não queremos ser manifestamente inconsistentes, devemos ter

$$G(G(x, y, z), u, G(v, w, s)) = G(x, G(y, u, w), s).$$

Ainda notamos que embora éstas variáveis possam ter quaisquer valores, não ocorre o mesmo conjunto dos dois lados: Lado esquerdo $\{x, y, z, u, v, w, s\}$, lado direito $\{x, y, u, w, s\}$. Portanto o lado esquerdo nao deve depender de $z = \phi(a|bI))$ nem de $v = \phi(a|cI)$ explicitamente. Para que essa expressão não dependa de z nem v , podemos impor que G não dependa do terceiro argumento o que levaria a eliminar o que foi riscado na equação abaixo:

$$G(G(x, y, \cancel{z}), u, \cancel{G(v, w, s)}) = G(x, G(y, u, \cancel{w}), \cancel{s})$$

levando a que G tem só dois argumentos e uma expressão sem z nem v:

$$G(G(x, y), u) = G(x, G(y, u))$$

e portanto somem todas as variáveis exceto x, y e u. Lembrando suas definições

$$G(G(\phi(a|I), \phi(b|I)), \phi(c|I)) = G(\phi(a|I), G(\phi(b|I), \phi(c|I)))$$

que equivale ao **Caso 1** e portanto já foi eliminado.
Mas também podemos dizer que não depende do primeiro argumento, que também elimina z e v:

$$G(\cancel{G(x, y, z)}, u, G(\cancel{v}, w, s)) = G(\cancel{x}, G(\cancel{y}, u, w), s)$$

que leva à expressão

$$G(u, G(w, s)) = G(G(u, w), s)$$

que voltando às variáveis originais toma a forma

$$G(\phi(c|I), G(\phi(b|cI), \phi(a|bcI))) = G(G(\phi(c|I), \phi(b|cI)), \phi(a|bcI))$$

e mostra ser equivalente ao que teriamos obtido se partissemos do **Caso 3** e portanto aceitável.
Fica como exercício mostrar que

1. o **Caso 6** pode ser reduzido ao **Caso 2,** ao **Caso 3** ou ao **Caso 4**
2. o **Caso 7** pode ser reduzido aos **Caso 5** ou **Caso 6.**

Concluimos portanto que

$$\boxed{\begin{array}{rcl} \phi(ab|c) & = & G(\phi(a|c), \phi(b|ac)) \\ & = & G(\phi(b|c), \phi(a|bc)) \end{array}}$$

Cox coloca isto como um axioma, mas não precisamos fazer isto, basta dizer que existe uma função G mas que não sabemos *a priori* quais seus argumentos. A eliminação dos casos que contradizem o bom senso em casos suficientemente simples, mostra de forma satisfatória (o leitor pode pular e reclamar, mas terá que encontrar argumentos) que as equações 1.3.2 refletem a única opção. Outra queixa e que introduzimos casos simples onde os casos diferentes do 3 se mostraram contrários ao bom senso. Isto significa que o DP_5 é mais complexo do que parecia inicialmente.
Note que agora será possível concluir que 'Helena usa um tenis esquerdo vermelho e um tenis direito preto' pode ser pouco plausível por que precisamos saber a plausibilidade de 'Helena usa um tenis esquerdo vermelho dado que Helena usa um tenis direito preto' e isto pode ser pouco plausível.
Mas ainda não acabamos. Precisamos determinar a função especifica G, com a vantagem que pelo menos sabemos seus argumentos.

Regra do produto: qual é a função G?

Novamente olhamos para um caso simples, onde podemos escrever o resultado de duas maneiras. Considere a, b, c e d com $b|d$ e $c|d$ mutuamente exclusivos, e a asserção $a(b \vee c)$ uma conjunção que pode ser escrita como uma disjunção:

$$a(b \vee c) = (ab) \vee (ac). \tag{1.14}$$

Podemos usar o resultado para a soma para estudar o produto $\phi(a(b \vee c)|d)$:

$$\begin{aligned} \phi(a(b \vee c)|d) &= G\left(\phi(a|d), \phi(b \vee c)|ad)\right) \\ &= G\left(\phi(a|d), \phi(b|ad) + \phi(c|ad)\right) \qquad (1.15) \\ \phi((ab) \vee (ac))|d) &= \phi(ab|d) + \phi(ac|d) \\ &= G\left(\phi(a|d), \phi(b|ad)\right) + G\left(\phi(a|d), \phi(c|ad)\right) \qquad (1.16) \end{aligned}$$

onde a equação 1.15 usa primeiro que $a(b \vee c)$ é um produto e em segundo lugar a regra da soma para asserções mutuamente exclusivas $b|d$ e $c|d$. A equação 1.16 mostra o resultado de considerar a soma $(ab) \vee (ac)$. Mas devido à equação 1.14 e DP_2, estas duas formas devem dar o mesmo resultado:

$$G(x, y+z) = G(x, y) + G(x, z). \tag{1.17}$$

Para obter a solução geral desta equação notemos que o primeiro argumento é o mesmo nos três termos, é portanto um parâmetro que podemos manter fixo em qualquer valor arbitrário. Não é necessário supor diferenciabilidade, mas requerindo que G seja duas vezes diferencíavel, e definindo $w = y + z$ obtemos a equação diferencial

$$\frac{\partial^2 G(x,w)}{\partial w^2} = 0 \tag{1.18}$$

que tem solução geral $G(x,w) = A(x)w + B(x)$ em termos de duas funções desconhecidas, mas fáceis de determinar.
Substituindo esta forma em 1.17 obtemos

$$A(x)(y+z) + B(x) = A(x)y + B(x) + A(x)z + B(x), \tag{1.19}$$

portanto $B(x) = 0$, ou seja $G(x,w) = A(x)w = G(x,1)w$ [15]. Agora olhamos para $a|d$ e usamos $a|d = ad|d$ para a e d quaisquer.

$$\begin{aligned}\phi(a|d) &= \phi(ad|d) = G(\phi(a|d), \phi(d|ad)) \\ &= G(\phi(a|d), \phi_v) = A(\phi(a|d))\phi_v\end{aligned} \tag{1.20}$$

onde $\phi(d|ad) = \phi_v$ pois, obviamente d é informação completa para d. Ou seja $x = A(x)\phi_v$, logo

$$G(x,w) = \frac{xw}{\phi_v} \tag{1.21}$$

isto significa que, para $e = b \vee c$, b e c mutuamente exclusivos

$$\phi(ae|d) \quad = \quad \frac{\phi(a|d)\phi(e|ad)}{\phi_v}. \tag{1.22}$$

Mas resta um problema: e se retirarmos a restrição de b e c mutuamente exclusivos? É simples de considerar pois novamente usamos a equação 1.8

$$e = (e \wedge d) \vee (e \wedge \overline{d}), \tag{1.23}$$

agora para qualquer asserção d, de tal forma que $b = e \wedge d$ e $c = e \wedge \overline{d}$ [16]. Portanto não ha restrições para o resultado que obtivemos.
Se não usarmos esse atalho deveriamos usar a equação 1.9 para obter:

$$G(x, y + z - G(y,w)) = G(x,y) + G(x,z) - G(x, G(x,w))$$

e sabemos que a solução é dada pela equação 1.21. Sem usar esse atalho é mais difícil mostrar que esta é a única forma se G for diferenciável duas vezes em cada argumento. O leitor interessado deverá consultar Áczel. Temos assim uma possibilidade de uma prova muito mais simples.
Da equação 1.22, dividindo por ϕ_v obtemos

$$\frac{\phi(ae|d)}{\phi_v} \quad = \quad \frac{\phi(a|d)}{\phi_v}\frac{\phi(e|ad)}{\phi_v} \tag{1.24}$$

o que permite regraduar mais uma vez os números associados as crenças sem mudar sua ordem. Crenças regraduadas, de forma

[15] Suponha a equação $h(x+y) = h(x) + h(y)$, para qualquer x, y. Em particular, para $n \neq 0$ e m inteiros, vale que $h(nx) = h((n-1)x + x) = h((n-1)x) + h(x) = h((n-2)x + x) + h(x) = h((n-2)x) + 2h(x) = \cdots = nh(x)$. Considere $x = x'/n$. Segue que $h(x') = nh(x'/n)$. Tome $x' = m$, portanto $h(x') = h(m) = mh(1) = nh(m/n)$. Logo $h(m/n) = (m/n)h(1)$. Basta supor continuidade que podemos passar dos racionais para os reais e obter $h(x) = xh(1)$.

[16] Agradeço a ...e a ... por me lembrar deste truque.

bijetora representam o mesmo ordenamento e portanto podem ser ainda chamados de crenças. Definimos os novos números

$$p(a|b) = \frac{\phi(a|b)}{\phi_v} \tag{1.25}$$

que serão daqui a pouco chamados de probabilidade de a dado b. E a regra do produto em termos destes novos números regraduados é

$$\boxed{\boldsymbol{p(ab|c) = p(b|c)p(a|bc) = p(a|c)p(b|ac)}}$$

Temos uma regra para o produto e para soma lógicas de asserções. Como fica a negação? Apesar de não ter introduzido nada específico sobre ela veremos que com os desejos impostos podemos deduzir a plausibilidade regraduada ou probabilidade da negação de uma asserção a partir da probabilidade de sua afirmação.
A regra do produto e a consistência permitem escrever

$$p(a|bc) = \frac{p(a|c)p(b|ac)}{p(b|c)} \tag{1.26}$$

que é chamado de Teorema de Bayes, mas que foi escrito pela prmeira vez por Laplace. A contribuição de Bayes foi apontar a relação chamada de inversão

$$p(a|bc) \propto p(b|ac) \tag{1.27}$$

onde a probabilidade de uma asserção a condicionada a outra b é proporcional à probabilidade de b condicionada a a. Não podemos exagerar a importância desta afirmação que ficara clara à luz da variedade de aplicações tanto teóricas quanto experimentais que veremos adiante.

Negação

A lista de desejos inclui a menção de algo sobre a negação. A crença em asserções condicionadas à sua negação constituem casos de informação completa: $\phi(a|\bar{a}) = p(a|\bar{a}) = 0$. Também sabemos que $a \vee \bar{a}$ deve ser verdade, pois não resta alternativa. Portanto

$$\begin{aligned} \phi(a|\bar{a}d) &= p(a|\bar{a}d) = 0 \\ p(a \vee \bar{a}|d) &= \frac{\phi(a \vee \bar{a}|d)}{\phi_v} = 1 \end{aligned} \tag{1.28}$$

$$\begin{aligned} 1 &= p(a \vee \bar{a}|d) \\ &= p(a|d) + p(\bar{a}|d) - p(a\bar{a}|d) \\ &= p(a|d) + p(\bar{a}|d) - p(\bar{a}|d)p(a|\bar{a}d) \\ &= p(a|d) + p(\bar{a}|d), \end{aligned} \tag{1.29}$$

$$\boxed{\boldsymbol{p(\bar{a}|d) = 1 - p(a|d)}}$$

ou a soma das crenças regraduadas de uma asserção e da sua negação é 1. Essencialmente chegamos ao fim do começo.

Estrutura matemática sobrevivente

Em termos destes números, reescrevemos os resultados até aqui obtidos:

$p(a\|a)$	$=$	$p_v = 1$	Certeza da veracidade
$p(a\|\overline{a})$	$=$	$p_f = 0$	Certeza da falsidade
$p(a \vee b\|c)$	$=$	$p(a\|c) + p(b\|c) - p(ab\|c)$	regra da soma
$p(ab\|c)$	$=$	$p(a\|c)p(b\|ac)$	regra do produto
$p(ab\|c)$	$=$	$p(b\|c)p(a\|bc)$	regra do produto
$p(\overline{a}\|d)$	$=$	$1 - p(a\|d)$	regra da negação

Tabela 1.2
Probabilidades

Não falaremos mais em números $a|b$, nem na sua regraduação $\phi(a|b)$ mas somente na última transformação $p(a|b)$ que chamaremos a probabilidade de a dado b, ou a probabilidade de a condicionada à informação que b é verdadeira. O motivo disto é que ao longo de séculos estas regras from destiladas pelo bom senso de vários matemáticos e cientistas. Por volta de 1930, Kolmogorov formalizou, sem incluir a regra do produto nem condicionantes, usando linguagem de teoria de medida ou integração, mas já eram conhecidas desde Laplace. O que não estava claro é porque essas e não outras. Está completa a identificação das crenças ou plausibilidade regraduadas em números que satisfazem as regras da probabilidade. Concluimos que a estrutura matemática adequada, e que usaremos nestas notas, para descrever situações de informação incompleta é a teoria de probabilidades. O leitor, caso deseje usar outras regras para manipular informação deverá responder quais dos desejos considerados acima não é razoável e portanto ao ser evitado, justificar essas outras regras.

O que foi obtido vai ser comparado com os axiomas de Kolmogorov na próxima secção. Vemos uma diferença importante. Na formulação da teoria de probabilidade como um capítulo da teoria da medida, as probabilidades são medidas e não há menção a condicionais. Rao adicionou mais tarde a complementação introduzindo, como uma idéia tardia, razoável mas *ad hoc*, a probabilidade condicional definida a partir da regra do produto e portanto colocando com a mão o teorema de Bayes, que Cox obteve como uma consequência direta da consistência em particular e dos outros membros da desiderata.

Este é o conteúdo dos teoremas de Cox: uma atribuição de números para descrever as crenças em asserções, dada a informação, que satisfaça os casos particulares, pode ser mudada de forma a não alterar o ordenamento das crenças e preferências e a satisfazer as regras da probabilidade. Tem cheiro e cor de probabilidade e tem todas as propriedades das probabilidades. Não falaremos mais sobre plausibilidade. Não sabiamos o que era, e a abandonamos como a um andaime, após ter construido o edifício da teoria de probabilidades. Obviamente este exercício não forneceu os valores das probabilidades. Que bom, senão fechariam os institutos dedicados ao estudo e às aplicações das probabilidades. Mais sérios, podemos dizer que a nossa grande preocupação agora será dirigida à busca de técnicas que baseadas na informação disponível permitam atribuições ou talvez o problema associado mas diferente, de atualização dos números associados a probabilidades dos eventos ou asserções de interesse quando recebemos nova informação. Esta é a preocupação central da inferência e da teoria de aprendizado e nos levará à introdução da idéia de entropia. A entropia no sentido de teoria de informação está intimamente ligada à idéia de entropia termodinâmica e mais ainda à de Mecânica Estatística como veremos mais tarde. Poderemos afirmar que a Mecânica Estatística foi a primeira teoria de informação, embora não seja costumeiro colocá-la nessa luz.

Exercícios

Mostre, construindo a tabela verdade as seguines propriedades da Algebra Booleana a partir da tabela verdade para a soma, produto e negação

- Idempotência do produto $AA = A$
- Idempotência da soma $A + A = A$
- Comutatividade do produto $AB = BA$
- Comutatividade da soma $A + B = B + A$
- Associatividade do produto $A(BC) = (AB)C$
- Associatividade da soma $(A + B) + C = A + (B + C)$
- Distributividade $A(B + C) = AB + AC$
- Dualidade $C = AB \Rightarrow \overline{C} = \overline{A} + \overline{B}$ e $C = A + B \Rightarrow \overline{C} = \overline{AB}$

Mostre que $(A + B)A = A$ e portanto $A + BC = (A + B)(A + C)$

Exercícios Propostos

- Mostre que a conjunção e a disjunção não formam um conjunto de operadores completo para a álgebra booleana. Por exemplo mostre que não há combinação de estes operadores que permitam obter a negação. Mas nos propusemos uma função F e uma G e obtivemos uma forma de lidar com a negação. Como isso é possível? A resposta será achada ao ver que que o desejo DP_3 sobre informação completa introduz a noção de negação mas só parcialmente ao dizer que a e sua negação são mutuamente exclusivos e que $a|\overline{a} = v_f$ como o mesmo v_f para todo a. Outra forma de proceder poderia ser introduzir um desejo do tipo: Deve existir uma função H, desconhecida tal que $a|c = H(\overline{a}|c)$. Isto codifica o desejo de encontrar uma teoria em que conhecimento sobre a implica conhecimento sobre $\overline{a}$. Claro que nesta altura sabemos que $H(x) = 1 - x$. Tente deduzir a as consequências ao trocar disjunção F por negação H no Desiderata para lidar com informação incompleta.

- Mostre a relação da equação 1.8. Desenhe o diagrama de Venn.

- A equação 1.9 relaciona a crença da disjunção às crenças nas asserções primitivas, mas inclui a subtração da crença na conjunção. Desenhe o diagrama de Venn adequado a esta situação. Discuta a origem do term subtraido.

- Voltemos ao **Caso 5** e suponhamos que G seja diferenciável com respeito a qualquer argumento. As derivadas parciais com respeito a z ou v devem dar zero. Use a regra da cadeia para mostrar que

$$\begin{aligned} 0 &= \frac{\partial}{\partial z} G(G(x,y,z),u,G(v,w,s)) \\ &= \frac{\partial}{\partial r} G(r,u,G(v,w,s))_{r=G(x,y,z)} \frac{\partial}{\partial z} G(x,y,z) \end{aligned} \tag{1.30}$$

  Se um produto é zero, pelo menos um dos fatores é zero, de onde concluimos que ou G não depende do primeiro argumento ou não depende do terceiro. Se não depende do primeiro mostre que voltamos ao **Caso 3** . Se não depende do terceiro mostre que voltamos ao **Caso 1.**

- **2** Para a função G da regra do produto mostrar que o **Caso 6** pode ser reduzido ao **Caso 3** ou ao **Caso 4** e que o **Caso 7** aos **Caso 5** ou **Caso 6.**

- Mostre que a forma produto (eq. 1.21) é solução da equação funcional. Mostre que esta é a única forma se G for diferenciável duas vezes em cada argumento.

- Escreva a regra do produto $P(AB|I)$, da soma $P(A+B|I)$ e da negação de $A|I$, de A no contexto I em termos das Chances, percentagem e Logprob definidos abaixo. Mostre que cada uma dessas é uma transformação monotónica ϕ das probabilidades e portanto uma regraduação possível da representação numérica das crenças.

1. Chances: Defina as chances (odds em inglés) como $O(A|I) = \frac{P(A|I)}{P(\overline{A}})$.

2. Percentagem é o que chamariamos a probabilidade se em lugar de estar confinada ao intervalo $[0,1]$ estivesse no intervalo$[0,100]$.

3. Logprob $L_P(A|I) = \log P(A|I)$.

4. Logit ou log-odds: $\text{Logit}(P(A|I)) = \log(\frac{P(A|I)}{P(\bar{A})})$.

5. Exprob $\text{Exp}_P(A|I) = \exp P(A|I)$ (Essa acabei de inventar).

6. Sineprob $\text{Sen}_P(A|I) = \sin\frac{P(A|I)\pi}{2}$ (Posso continuar.)

Em algum caso as regras escritas em termos das regraduações são mais simples do que a regraduação que leva às probabilidades? [17]

[17] Não é verdade que neste caso "What's in a name? that which we call a rose By any other name would smell as sweet"

Exercício Problema de Linda 1. Amos Tversky and Daniel Kahneman colocaram a questão a seguir, chamada de Problema de Linda, sobre probabilidades. Considere as asserções a seguir:

- I : Linda tem 31 anos, é solteira, assertiva, e muito inteligente. Ela se formou em filosofia. Quando estudante, estava profundamente preocupada com questões de discriminação e justiça social, e também participou de manifestações anti-nucleares.
- A : Linda é bancária .
- B : Linda é bancária e participa do movimento feminista .

Responda rapidamente qual das duas asserções é mais provável?

Exercício Problema de Linda 2. Não continue lendo até ter respondido à pergunta anterior.
Responda após pensar.O problema é atribuir números a $P(A|I)$ e $P(B|I)$. Qual é maior? Responda usando a regra do produto e use o fato que qualquer probabilidade tem uma cota superior 1. Este problema também é chamado de Falácia da conjunção. [18] Introduza a asserção

[18] These long-term studies have provided our finest insight into "natural reasoning" and its curious departure from logical truth. Stephen Jay Gould, sobre Tversky and Kahneman

- C : Linda é bancária e não participa do movimento feminista .

Qual seria o ordenamento das três probabilidades $P(A|I)$, $P(B|I)$ e $P(C|I)$? Procure alguém feminista e faça a pergunta, faça o mesmo com alguém machista. Divirta-se com a percepção que as pessoas são irracionais. O que você acha que as pessoas acham que respondem quando tem que ser rápidas? Note que muitas vezes ao fazer uma pergunta, quem responde está respondendo a uma pergunta parecida mas não exatamente aquela demandada.

Exercício Problema de Linda 3. Mostre usando a regra do produto que $P(A|I) \geq P(B|I)$. Tente inferir o que as pessoas fazem quando acham que está certo que $P(A|I) \leq P(B|I)$. Encontre asserções $A'|I'$ e $B'|I'$ *parecidas* com $A|I$ e $B|I$ tal que seja razoável supor mais provável supor o ordenamento contrário.

Exercício

- I : O preço do petróleo cai a 10 dolares o barril

- A : A Russia invade a Ucrânia
- B : A Russia invade a Ucrânia e os Estados Unidos corta relações diplomáticas com a Russia

Dado I qual é mais provável, A ou B? Note que as pessoas que cometem o erro de Falácia da Conjunção agem aparentemente como se estivessem comparando $P(A|I)$ com $P(C|AI)$, onde $B = AC$. Se você fosse presidente, manteria como assesor em política internacional alguém que ache $A|I$ menos provável que $B|I$?

Exercício

- I : Sou estudante da USP;
- A : Não estudei probabilidades
- B : Não estudei probabilidades e cometo a Falácia da conjunção

Dado I qual é mais provável, A ou B?

2

Outras definições de probabilidade

Kolmogorov e as probabilidades

Kolmogorov introduziu na década dos trinta [1] os seus famosos axiomas para a teoria das probabilidades. No seu livro ele declara que não vai entrar no debate filosófico sobre o significado de probabilidades e depois dá uma pequena justificativa dos axiomas com base na interpretação freqüentista de von Mises. Já descreveremos alguns dos motivos que nos levam a achar a posição freqüentista, incompleta e até, como mostraremos abaixo, insuficiente e errada. Pelo contrário, os axiomas de Kolmogorov, que codificam o bom senso da área já existente no trabalho de Laplace, podem ser vistos como não antagonicos aos resultados obtidos no capítulo 1. Interessante ler Kolmogorov. Ele não tem outro objetivo que

[1] Foundations of the Theory of Probability
http://www.mathematik.com/Kolmogorov/index.htr

> "... colocar no seu lugar natural, entre as noções gerais de matemática moderna, os conceitos básicos da teoria de probabilidade - conceitos que até recentemente eram considerados bastante peculiares.
>
> Esta tarefa não teria tido esperança de sucesso antes da introdução das teorias de medida e integração de Lebesgue..."
>
> A. N. Kolmogorov

Ele está organizando uma área após ficar claro como fazé-lo graças ao trabalho de Lebesgue e também Fréchet e admite que este ponto de vista era comum entre certos matemáticos mas merecia uma exposição concisa e livre de algumas complicações desnecessárias. Kolmogorov começa for considerar E uma coleção de elementos A, B, C.... que são eventos elementares [2] e em nossa discussão anterior chamamos de asserções. $\mathcal{F}$ é o conjunto de subconjuntos de E. Um tal sistema de conjuntos é chamado um campo se a soma, produto, interseção de dois elementos quaisquer pertencem ao sistema. Os axiomas de Kolmogorov para a teoria de Proabilidades são

[2] Em física E é conhecido como espaço de fases

- AK1) $\mathcal{F}$ é um campo de conjuntos fechado ante um número de uniões (disjunções) e interseções (conjunções) enumeráveis e se $A \in \mathcal{F}$ e $\overline{A} = E - A$, então $\overline{A} \in \mathcal{F}$

ou seja $\mathcal{F}$ é um σ-campo,

- AK2) $\mathcal{F}$ contém o conjunto E.
- AK3) A cada conjunto $A \in E$ é atribuido um número real não negativo, chamado de probabilidade do evento A denotado por $P(A)$.
- AK4) $P(E) = 1$
- AK5) Se $A \wedge B = \emptyset$, então $P(A \vee B) = P(A) + P(B)$

Vejamos se estes axiomas estão de acordo com os resultados da seção anterior. Em primeiro lugar uma definição que não será necessária neste curso, a de σ-campo. É uma coleção de subconjuntos fechado ante um número contável de operações de conjunto, tais como disjunção, conjunção, negação. Esta noção só é necessária ao falar de conjuntos com infinitos elementos. Vimos que a coleção de asserções também permite tais operações. Portanto estamos lidando com o mesmo tipo de coleção de eventos que Kolmogorov. [3] Um exemplo de um σ-campo é o conjunto de conjuntos abertos nos reais. Neste curso usaremos asserções do tipo: "A variável X tem valor no aberto $(x, x + dx)$"e extensões a $\mathbb{R}^N$. A ideia de σ-campo é essencial na teoria de integração de Lebesgue e aparecerá em tratamentos matematicamente mais sofisticados de probabilidade. Neste curso não iremos além de integrais de Riemann e somas infinitas.

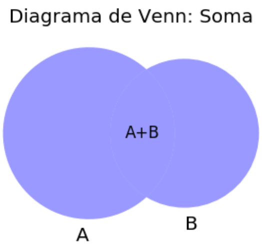
[3] Talvez a queixa é que as provas do capítulo 1 são para número finito de conjunções e disjunções. Isto porém não deve ser motivo de preocupação agora pois não é um empecilho irremovível.

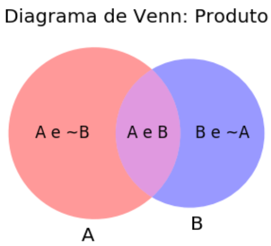
A probabilidade da *certeza* é 1 por AK4; a probabilidade está entre zero e um e a probabilidade da disjunção de asserções que não tem elementos em comum é a soma das probabilidades. Notamos, na introdução aos axiomas no livro de Kolmogorov, porém a falta de uma regra para o produto. Kolmogorov não a introduziu incialmente e só em trabalhos posteriores foi incluida por sugestão de Rao. No livro (página 6) ele introduz, como um adendo aos axiomas, as probabilidades condicionais através de

$$p(A|B) = \frac{P(AB)}{P(B)} \tag{2.1}$$

de onde segue para a prova do teorema de Bayes, usando a comutação de A e B, portanto $P(AB) = P(BA)$ e a simetria ante troca $A \leftrightarrow B$.

Se uma vez estabelecidos os axiomas e dados valores numéricos para as probabilidades partirmos para as aplicações matemáticas, não haverá nenhuma diferença de resultados pois será a mesma estrutura matemática. Enfatizamos que as diferenças que temos são sobre a motivação dos axiomas e com a interpretação da idéia de probabilidades. Isso tem importância em inferência e portanto em aplicações. Em muitos livros o estudante encontrará uma diferença entre probabilidades e probabilidades condicionais. Deve ficar claro que no ponto de vista destas notas, não há probabilidade que não seja condicional.





Ainda outras definições de Probabilidade

Outra proposta de definição de probabilidades é a frequentista, que tem mais chances de ser a que o leitor já viu. A definição parece muito simples: é o limite da razão entre o número de vezes que um evento é verdade e o número de tentativas, quando este último vai para infinito.
Esta definição veio no esteio de uma colocada por Jacob Bernoulli e Laplace. Para eles é às vezes conveniente definir a teoria de chances pela redução de eventos do mesmo tipo a certo número de casos, igualmente possíveis e a

> "...probabilidade, que é então simplesmente a fração cujo numerador é o número de casos favoráveis e cujo denominador é o número de todos os casos possíveis."[4]

O que significa "do mesmo tipo"? O físico verá aqui a uso da ideia de simetria. Se diferentes estados são tais que somos indiferentes ou incapazes de distingui-los então os colocamos na mesma categoria. Idéias de simetria são extremamente frutíferas. Mas quando não há simetria ou simplesmente não temos informação sobre ela é preciso estender a definição. Na época de Laplace as coisas não estavam muito claras, embora este tipo de regra seja útil e como veremos adiante não é uma regra nova a ser adicionada à *Desiderata* mas a ser deduzida do que já obtivemos. Além disso Laplace e Bernoulli deixaram claro em outros lugares que a probabilidade era uma manifestação numérica de crenças a partir de informação, portanto foram predecessores do exposto aqui. Considere, como Laplace há mais de duzentos anos, M_S a massa de Saturno. Ele fez asserções do tipo: "A probabilidade que $M_S < M_0$ ou $M_S > M_0 + \Delta m$ é menor que 10^{-4}", que ele colocou em linguagem de apostas. Em linguagem atual é algo como $P(M_0 < M_S < M_0 + \Delta m|I) > 1 - 10^{-4}$. A informação de fundo condicionante I representa a teoria de Newton e os dados experimentais [5]. Ele não está dizendo que a massa de Saturno é uma grandeza que apresenta variações e se for medida exatamente apresentará diferentes valores. Esqueça meteoritos, que poderiam mudar sua massa. Por exemplo, ao jogar um dado, se medirmos qual é o número de pontos na face que está para cima,

[4] The theory of chances consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining the number of cases favorable to the event whose probabilty is sought. The ratio of this number to that of all the cases possible is the measure of this probability which is thus simply a fraction whose numerator is the number of favorable cases and whose denominator is the number of all te cases possible
A Philosophical Essay on Probabilities, Pierre Simon, Marquis de Laplace. 6a ed. F.W.Truscott e F.L. Emory trads.

[5] A *incerteza* Δm que Laplace tem é da ordem de 1% de M_0. O erro da estimativa de Laplace em relação ao valor estimado moderno é de aproximadamente 0.6%. Ou seja, ele teria ganho a aposta. O valor numérico $P(M_0 < M_S < M_0 + \Delta m|I)$ representa a crença que M_s esteja dentro do intervalo $(M_0, M_0 + \Delta m)$

este terá variações para diferentes jogadas. Alguns autores acham que só este tipo de variável merece ser descrito por probabilidades. Mas não a massa de Saturno à qual se atribui a propriedade de ter um valor *real*[6]. O que Laplace quer dizer é sobre o valor que atribuimos, com base nos dados e teoria, à crença que a massa está em um ou em outro intervalo. Quem acredita na definição de probabilidades como frequência, no pode falar da massa de Saturno em termos de probabilidade, pois não há um conjunto de Saturnos com diferentes massas. Falam em lugar disto, da probabilidade de que o conjunto de medidas seja observado para o caso em que a massa seja M_0. Em alguns casos isto dará o mesmo resultado, mas em outros não. Se você for acuado a definir a maior diferença entre alguém que define probabilidades através de frequências e quem as usa para expressar graus de crença, poderá responder de forma simplificada que este último não hesita em falar da distribuição de probabilidades de um parâmetro, como a massa de Saturno, enquanto o primeiro não admite tal linguagem [7].

[6] Real no sentido de ter existência independente do observador. Procure o significado de ontologico e de epistemico.

[7] Em linguagem mais técnica, ao espaço de parâmetros também é atribuido um $\sigma-$campo.

Algumas definições

Nesta altura podemos identificar os elementos formais principais para falar de probabilidades na linguagem de Kolmogorov. Primeiro é necessãrio deixar claro sobre o que se está falando:

- E a coleção de elementos $A, B, C....$ eventos elementares ou de asserções. Em alguns meios é chamado de espaço amostral.

- $\mathcal{F}$ o campo: o sistema de conjuntos de asserções. Espaço de eventos

- $P(A)$ a atribuição de um número positivo a cada elemento de $\mathcal{F}$;

Desta forma é costumeiro chamar a trinca (Espaço amostral, Espaço de eventos, Probabilidade de cada evento).

$$\boxed{(E, \mathcal{F}, \mathcal{P})}$$

de Espaço de probabilidades.
A apresentação do capítulo 1 não discorda disto, a não ser pelo ponto essencial que as probabilidades serão sempre condicionais e esquecer isso será a maior fonte de erros nas aplicações. Quando alguém se refere a uma probabilidade tipicamente tem em mente detalhes que se recusa a deixar explícitos pois esse exercício pode parecer cansativo. Outras vezes, e isso é mais perigoso, age como se tivesse em mente certos detalhes de informação, mas ao não perceber pode achar que não há alternativas. Além disso quando a teoria tem parâmetros, como será discutido em mais detalhes no próximo capítulo, queremos poder falar das probabilidades de que os parâmetros tenham valores em uma dada região. Isto não está em desacordo com a posição de quem adota os axiomas de

Kolmogorov. Basta aumentar o espaço amostral e o σ-campo e atribuir probabilidades aos elementos do novo campo. Isto porém não está de acordo com uma visão frequentista pois a massa de Saturno, ou qualquer outro parâmetro de uma teoria tem uma natureza ontológica que não lhes permite ser descrito em termos de frequência.