

ALEATORIZAÇÃO E COMPOSIÇÃO DE MEDIDAS DE PREFERÊNCIA

Annibal Parracho Sant'Anna
Universidade Federal Fluminense
Rio de Janeiro – RJ
E-mail: aparracho@uol.com.br

Recebido em 01/2002, aceito em 06/2002 após 1 revisão

Resumo

Modelos para a formação das preferências combinando preferências estabelecidas através da ordenação das opções possíveis segundo fatores isolados são aqui desenvolvidos. É, também, proposta uma sistemática de aleatorização dos postos para permitir o cálculo da probabilidade de cada opção ocupar o primeiro posto. Aplicando esta sistemática a apostas em corridas de cavalos, verifica-se que a medida da preferência pela chance relativa de a opção ocupar o primeiro posto segundo um critério qualquer, ou segundo uma combinação dos critérios, é mais correlacionada com a medida de preferência dada pela distribuição final das apostas que a medida dada diretamente pelo vetor de postos. A transformação dos postos em chances relativas permite detectar a alta correlação com as apostas de uma medida de preferência agregada que envolve a projeção sobre uma opção selecionada como a de maior preferência segundo critério dominante.

Palavras-chave: postos aleatórios, modelagem de preferências, auxílio à decisão.

Abstract

Models for preferences, established through the ranking of the options according to isolated factors, are developed here. Randomization of ranks allows us to compute the probability of each option be raised to the position of best choice. It is verified that, in horses races, measures of preference built through odds of the options to hold the first position according to some criterion or to a combination of criteria are more correlated with the measure of preference given by the final betting distribution than those given directly by the ranks.

Keywords: random ranks, modeling preferences, multicriteria decision analysis.

1. Introdução

As escolhas dos apostadores nas modalidades de apostas páreo a páreo nas corridas de cavalos constituem uma tomada de decisão coletiva a respeito de qual a correta distribuição de probabilidades das opções oferecidas. Soando o sinal que anuncia a aproximação do encerramento das apostas e a abertura dos portões de partida aos animais, centenas de pessoas dirigem-se aos guichês para realizar as apostas que, no entender de cada um, lhes oferecem a melhor relação custo/benefício, elevando deste modo a razão entre apostas efetivadas e chance nos pontos em que esta lhes pareça pequena em relação às alternativas. Na modalidade de aposta no animal vencedor do páreo, a mais simples e de maior volume de apostas, podemos concluir que o encerramento das apostas colhe uma distribuição de probabilidades que reflete muito fielmente o ponto de vista do grupo de apostadores no exato momento em que se inicia a corrida.

É claro que a distribuição de probabilidades assim gerada não precisa representar corretamente a distribuição de probabilidades de cada um dos animais ser o vencedor, de tal modo que apenas fatores aleatórios, intervindo durante a carreira conforme essa distribuição, determinem o vencedor. É possível que a formação de preferências do apostador deixe de levar em conta fatores que sistematicamente afetam os resultados das provas e que não são do conhecimento dos apostadores, assim como é possível que estes formem suas preferências com base em informações falsas e teorias errôneas, que, uma vez corrigidas, pudessem ter sido usadas para gerar, no início da corrida, mais correta distribuição de probabilidades de vitória de cada animal.

Independente disto, está estabelecido, começando com Kahneman & Tversky (1979), que, em muitas situações, a distribuição de preferências efetivamente observada se desvia da aplicação racional das informações e modelos teóricos de que o tomador de decisão dispõe. A quantificação desses desvios, catalogados de forma sistemática em McFadden (1999), vem sendo desenvolvida nas últimas décadas. Seguimos aqui a idéia de Gomes & Lima (1992) de derivar a forma funcional que melhor espelhe as distorções objetivamente encontradas.

No caso das corridas, no frêmito do fechamento das apostas é possível que fatores emocionais desviem na mesma direção o comportamento dos apostadores do objetivo de beneficiar-se das possíveis distorções na relação de custo/benefício de cada aposta. Dentre estes fatores, destacam-se, de um lado, a conveniência de fundamentar a escolha em uma comparação simples e, de outro, a insegurança inerente ao resultado de qualquer procedimento racional de escolha, dados os fatores que é sempre preciso deixar fora do modelo racional de análise, por melhor que seja. A conjugação destes dois princípios levaria a concentrar as apostas nas opções de maior probabilidade e a calcular a chance das outras opções comparativamente em relação a essas. Além disso, motivações puramente emocionais, discutidas por Ali (1977) e Thaler & Ziemba (1997), atuariam no sentido de elevar a proporção das apostas nas opções muito pouco prováveis.

Substituindo ordenações iniciais de preferência pelas probabilidades de cada opção ser a preferida, obedecemos aos dois princípios acima explicitados, de simplificação e de incerteza, e produzimos, também, aproximação entre as preferências pelas opções menos viáveis. Esta substituição é aqui aplicada a dois fatores: a preferência derivada do desempenho passado dos competidores, fornecida pelo *handicappeur* oficial do hipódromo ao elaborar o programa dos grandes prêmios, antes da assinatura dos compromissos de montaria pelos jóqueis, e a preferência exercida pelos jóqueis.

Tem sido observada, também, em diferentes situações, uma maior distância entre as opções com maior probabilidade de sucesso, conduzindo a uma distribuição de probabilidades aproximadamente exponencial. Um levantamento de situações dessa natureza é apresentado por Lootsma (1993). Como a distribuição de preferências baseada nas probabilidades de cada opção ser a preferida tem uma forma próxima da exponencial, se conseguirmos, através dela, explicar a formação de preferências, estaremos trazendo mais um elemento de convicção favorável à aproximação exponencial.

Começamos a construção dessa distribuição ordenando as opções segundo um critério dado. A seguir, interpretamos o posto de cada opção como valor observado e estimativa do valor esperado de uma medida de preferência aleatória. Da distribuição conjunta destes postos aleatórios, construída utilizando hipóteses simplificadoras quanto à forma da distribuição do posto observado e de independência e idêntica dispersão, derivamos a probabilidade de cada opção ser a preferida. Pode-se aplicar este algoritmo à atribuição de probabilidades segundo cada um dos fatores isoladamente ou a preferências agregadas geradas combinando os fatores iniciais.

A questão de como combinar os fatores, o princípio da simplificação permite responder através da reformulação da avaliação das opções, comparando-as com aquela em maior evidência, que, no caso, é a vitória do animal favorito do *handicappeur* oficial do hipódromo. Assim, a redução de dimensionalidade é aplicada projetando todos os vetores de preferências de ordem 2, preferência do *handicappeur* e preferência dos jôqueis, na mesma direção, determinada pelo vetor correspondente ao animal preferido pelo *handicappeur*.

Estas transformações dos dados são aplicadas a seguir para explicar as preferências dos apostadores em provas para as quais se dispõe da classificação prévia oficial dos competidores. Algumas conclusões podem ser extraídas do ajustamento de modelos explicando a preferência dos apostadores refletida na distribuição final das apostas a partir das classificações segundo os dois critérios básicos.

Em primeiro lugar, verifica-se que o ajustamento do modelo de regressão linear melhora substancialmente quando se substituem os postos pelas chances relativas derivadas da probabilidade de a opção ocupar a posição de maior preferência. E melhora mais quando se agregam os fatores projetando sobre a direção da opção preferida pelo *handicappeur* oficial.

Na Seção 2, se discute a aleatorização dos postos para geração da probabilidade de escolha. Na Seção 3, se trata da composição das preferências derivadas de diferentes critérios. Na Seção 4, são estudados os exemplos de aplicação às corridas do Jôquei Clube Brasileiro. Na Seção 5, são apresentados comentários finais.

2. Adição de Componente Aleatório

Nesta Seção é desenvolvida uma sistemática para introduzir uma parcela probabilística na determinação da posição relativa de cada opção. Com a introdução deste componente aleatório, a posição inicialmente apresentada de forma determinística passa a ser tratada como uma estimativa para o ponto médio de uma distribuição de probabilidades.

A forma mais simples de fornecer esta posição inicial é através da ordenação das opções, da de menor preferência para a de maior preferência. Esta ordenação não precisa ser rigorosa, podem ser admitidos empates, ou postos vagos para representar afastamento maior que a regra geral. O importante é que, uma vez transformadas as indicações de preferência em

valores numéricos, tratando-os como variáveis aleatórias, sempre se pode derivar do conjunto de valores observados uma distribuição de probabilidades de cada opção ocupar a posição de maior preferência.

Quanto à determinação das distribuições de probabilidades dessas variáveis aleatórias, a ausência de informação, característica, pelo menos, das primeiras aplicações, acentua a exigência de parcimônia na parametrização. Para ser coerente com a idéia de aumentar as distâncias próximo à fronteira de excelência, devemos procurar trabalhar com distribuições de caudas leves. Por outro lado, para que a aleatorização seja mais efetiva é preciso que a variabilidade admitida favoreça ao máximo a possibilidade de troca de postos entre opções próximas. Estas diretrizes conduzem à escolha da distribuição uniforme.

Na família uniforme, a distribuição da posição aleatória em torno do seu valor esperado, estimado, no caso, pela posição em que a opção é deterministicamente colocada na classificação inicial, fica perfeitamente identificada se informamos o valor de um parâmetro de dispersão. Fazemos isto fixando a amplitude no menor valor que permita inversão entre duas opções quaisquer, isto é, no valor da diferença entre as medidas iniciais de preferência pelas opções de maior e menor preferência. Sendo as preferências dadas em termos de postos, esta diferença é igual ao número de opções disponíveis menos 1.

Formalmente, a transformação aplicada consiste em colocar no lugar da medida R_{ij} da opção j -ésima pelo critério i -ésimo a probabilidade de que este critério atribua a essa opção a máxima preferência, sob a hipótese de que, para todo i e todo j , a medida correspondente à j -ésima opção pelo i -ésimo critério seja uniformemente distribuída em torno do respectivo registro R_{ij} . E estas distribuições uniformes são supostas independentes, todas aquelas relativas a postos segundo um mesmo critério tendo a mesma amplitude, igual, para o critério i -ésimo, ao máximo das diferenças $R_{ik} - R_{il}$, para k e l variando ao longo de todas as opções avaliadas.

Dada a razão $(n-1)/(n+1)$ entre o valor esperado da amplitude da amostra aleatória simples de tamanho n e a amplitude da população uniformemente distribuída em um intervalo qualquer, a estimativa para a amplitude derivada da amplitude amostral deveria envolver a divisão por $(n-1)/(n+1)$, elevando para o número de opções disponíveis mais 1 a amplitude atribuída à distribuição de cada posto no caso de preferências ordinais. Esta correção pode, entretanto, tornar-se excessiva, visto que a atribuição de preferências às opções não se realiza de forma totalmente aleatória, mas, ao contrário, as perturbações aleatórias apenas destorcem ordenações de preferências subjacentes.

Poderíamos, também, seguir a prática usual de derivar estimativa para o desvio padrão da perturbação de cada medida do desvio padrão amostral, sendo a amostra, no caso, constituída pelas medidas de preferência atribuídas às opções. O fato de que os valores esperados das variáveis na amostra são diferentes também deve fazer este procedimento superestimar a dispersão. De fato, no caso de postos, o desvio padrão amostral é o da distribuição uniforme discreta nos inteiros de 1 a n . Esta distribuição tem variância dada por $n(n+1)(n-1)/12$, muito maior que a da distribuição uniforme em intervalo de amplitude $n-1$. E tem amplitude relativa de ordem $n^{-1/2}$, decrescente com o tamanho da amostra, contra a amplitude relativa constante das distribuições uniformes em intervalo.

Estimar o desvio padrão de cada medida pelo desvio padrão amostral não parece, portanto, apropriado no caso uniforme. Já para a distribuição normal, para a qual o desvio padrão é um parâmetro natural, o gradual decréscimo da densidade com o afastamento do centro exige, para facilitar adequadamente a inversão de postos, que a relação entre a dispersão atribuída a

cada medida e a dispersão observada entre as medidas iniciais seja maior. Neste caso, de fato, atribuindo ao desvio padrão de cada posição o desvio padrão amostral, as probabilidades de inversão obtidas são compatíveis com as esperadas. No caso de postos de 10 opções, o valor esperado para a amplitude relativa, de aproximadamente 3, implica a probabilidade de inversão entre a primeira e a décima medida da ordem de 0,1%. A superestimação pelo desvio padrão amostral pode, então, acabar por conduzir a probabilidades de inversão que representem mais fielmente a incerteza que atua nessa fase do processo.

Se o número de opções é grande, a hipótese de que todas as inversões sejam possíveis pode ser pouco realista. De qualquer modo, para o objetivo de calcular a probabilidade de ser a preferida, a partir de tamanhos de amostra de 10, pouca diferença resulta de assumir amplitude igual a 10 ou mais próxima do número total de opções.

No caso contrário, em que o número de opções seja pequeno, pode ser conveniente modelar a dispersão com uma amplitude maior que o tamanho da amostra para representar mais corretamente a possibilidade de inversão entre as opções. Isto pode ser efetuado, concretamente, acrescentando-se uma ou duas opções fictícias no extremo de menor preferência.

Pode-se, ainda, abandonar a hipótese de idêntica dispersão e ampliar ou reduzir o desvio padrão de um ou outro posto para refletir uma convicção maior ou menor sobre a posição de opções melhor ou pior conhecidas. A modelagem da dispersão é, não obstante, em geral, uma possibilidade difícil de explorar.

A independência dos fatores de aleatorização dos postos também é uma hipótese simplificadora artificial. Como a ordenação resulta de comparação entre as opções, poderíamos aperfeiçoar o modelo incluindo a correlação negativa que se deve esperar em decorrência da soma constante dos postos. Esta correlação, entretanto, decresce rapidamente, em valor absoluto, com o aumento do número de opções, também deixando de ter efeitos numéricos consideráveis a partir de 10 opções.

As decisões consideradas no presente estudo envolvem de 10 a 20 opções. Fica-se, neste caso, à vontade, para assumir independência e idêntica distribuição uniforme com amplitude igual à observada, conforme indicado inicialmente. Os resultados apresentados a seguir são todos baseados nesta hipótese. Pôde-se, entretanto, verificar que, modelando com a hipótese de normalidade e desvio padrão igual ao amostral, chega-se a resultados muito parecidos.

3. Combinação de Múltiplos Critérios

3.1 Classes de Alternativas

A determinação da preferência em termos de probabilidade de a opção ser a preferida a partir de uma classificação inicial pode ser aplicada, separadamente, a critérios simples que serão combinados a seguir, ou a um critério único que, possivelmente, resulte de anterior combinação de outros fatores. Nesta seção, apresentamos as alternativas de agregação dos critérios que serão experimentadas adiante. Estas alternativas são classificadas em dois grupos: com igual importância para todos os critérios e com ponderação dos critérios.

Há alternativas para atribuir igual importância de naturezas muito distintas. Aqui são desenvolvidas duas em que, entrando inicialmente com igual probabilidade de escolha, os

diferentes critérios acabam tendo influência muito diferente no resultado final. A primeira é baseada na composição das preferências em termos de probabilidade de ser a opção preferida. A segunda, baseada, na Análise Envoltória de Dados (DEA), mede a preferência pela proximidade de uma envoltória convexa.

Depois disto, são, também, listadas formas de agregação através de ponderação dos critérios. Destas, para manter a referência à opção de maior preferência, recebe maior atenção uma forma nova, baseada em pesos derivados da projeção sobre a opção preferida segundo o critério considerado mais importante.

3.2 Composições com Igual Importância

O princípio de igual importância para todos os critérios pode ser aplicado de várias formas. As mais simples consistem no cálculo da média aritmética das preferências ou da norma do vetor de preferências. Uma forma mais probabilística de compor com igual importância para todos os critérios consiste em usar como medida global a probabilidade de a opção ser a preferida por pelo menos um dos critérios considerados. Formalmente, representando por P_{ij} a probabilidade de a opção j -ésima ser a preferida pelo critério i -ésimo, a medida final da preferência por essa opção será $1 - \prod(1 - P_{ij})$ e a chance relativa respectiva será $[1 - \prod(1 - P_{ij})] / \prod(1 - P_{ij})$, i variando, no produto, ao longo de todos os critérios considerados.

Outra alternativa que se orienta pelo mesmo princípio de privilegiar as opções mais próximas da posição de preferida segundo algum critério consiste em medir a preferência pela proximidade à envoltória convexa do conjunto dos vetores de preferência. Este é o critério de eficiência de Farrel (1957), cujo cálculo pode ser implementado através do algoritmo de Análise Envoltória de Dados com Retornos Constantes de Escala (DEA-CRS) orientado pelo *input*, bastando para isto que avaliemos as opções em lugar de unidades de produção, considerando as preferências segundo cada critério no lugar dos produtos resultantes da utilização de um volume fixo de um *input* qualquer.

Neste caso, garantimos que todos os critérios recebem a mesma importância tanto no caso em que as medidas de preferência sejam dadas como postos quanto no caso em que já são fornecidas através das probabilidades de cada opção vir a atingir o posto de opção preferida. Além disto, o resultado da composição por este algoritmo é invariante com mudanças de escala, isto é, com mudanças que preservem a proporcionalidade entre as medidas de preferência por diferentes opções.

O problema de otimização resolvido ao aplicar este algoritmo, assumindo que as opções sob avaliação são classificadas, de acordo com cada critério, da menos preferível para a mais preferível, tem a seguinte formulação. R_{ij} denotando o posto da j -ésima opção de acordo com o critério i -ésimo, a preferência global pela o -ésima opção é dada por $e_o = \max \sum w_i R_{io}$, onde os pesos não-negativos w_i obedecem à restrição de $\sum w_i R_{ij} \leq 1$ qualquer que seja a opção j . Nas parcelas destes somatórios estarão representados todos os critérios admitidos na análise.

Ao admitir que os multiplicadores w_i possam ser nulos, permitimos que a preferência global por uma opção qualquer possa ser elevada pela exclusão de critérios em que essa opção tenha uma posição baixa. Isto pode resultar, por exemplo, na atribuição de uma preferência final máxima para uma opção que apresente a mesma classificação de outra em todos os critérios menos alguns para os quais escolha o peso zero, mesmo que, nestes critérios, a outra receba uma classificação melhor. Na abordagem baseada em postos, para que isto não possa acontecer, basta proibir empates.

A formulação dual da regra de otimização acima corresponde à formulação da envoltória no modelo de DEA de Charnes, Cooper & Rhodes (1978) orientado para o *input*. Nesta formulação, o nível de eficiência de uma unidade de produção é dado pelo mínimo dos possíveis quocientes com denominador dado pelo *input* agregado da unidade de produção avaliada e numerador dado pelo *input* agregado que uma soma de unidades de produção fictícias, obtidas reduzindo ou ampliando unidades de produção reais proporcionalmente em todos os seus *inputs* e *outputs*, precisa consumir para produzir um volume de cada *output* maior ou igual ao da unidade sob avaliação. Quando as preferências segundo os critérios particulares ocupam o lugar de *outputs* decorrentes da aplicação de um *input* fixo a cada uma das unidades sob avaliação, a preferência agregada é dada pela mínima fração desse volume padrão de *input* que, aplicado a uma fictícia combinação de opções, lhe permite atingir um posto maior ou igual ao da opção sob avaliação.

Formalmente, a preferência pela i -ésima opção será dada pelo mínimo valor de θ tal que $\theta \sum \lambda_j R_{ij} \geq R_{i0}$, para qualquer critério i , com os λ_j todos não negativos e somando 1 e o somatório realizado ao longo de todas as opções consideradas. Com todas as variáveis crescendo no mesmo sentido, é fácil visualizar a contribuição de cada uma das opções de referência que compõem a soma fictícia da qual apenas uma fração θ é suficiente para ultrapassar a posição da opção sob avaliação. Este escore θ corresponde à soma das contribuições $\theta \lambda_j$ das opções de referência.

A norma quadrática é uma forma mais simples de tratar igualmente todos os critérios do ponto de vista global e, na avaliação de cada opção, dar maior importância aos critérios em que a opção receba maior preferência. De fato, a norma mede a preferência agregada pela média ponderada das preferências segundo os diferentes critérios com o peso de cada critério dado pela própria medida de preferência pela opção segundo o critério. A classificação pela norma pode ser, então, pensada como uma simplificação do algoritmo da DEA, eliminando a busca dos preços que maximizam a eficiência relativa, substituindo-os por preços proporcionais aos volumes dos *outputs*.

3.3 Ponderação dos Critérios

Tanto a conveniência de fornecer uma explicação simples quanto a natural necessidade de desenvolver o processo de decisão a partir do conjunto de opções efetivamente disponível conduzem a pesos para os critérios que variam de um caso para outro. Alguns critérios podem não ser nunca aplicados a certos tipos de situações. Para outras decisões pode-se chegar a estabelecer uma hierarquia entre os critérios, com pesos maiores tanto para os critérios considerados mais importantes quanto para os que, por estarem menos correlacionados com os demais, oferecem uma contribuição mais destacada.

No grupo das sistemáticas de agregação que dão importância diferente aos diferentes critérios destaca-se a ordenação dos vetores de preferência segundo sua projeção sobre uma direção única. Neste caso, o princípio da simplificação leva a escolher tal direção entre as efetivamente observadas. Isto é, esta direção é determinada pelo vetor de preferências de uma opção de referência. Finalmente, chegando a este ponto, é natural escolher como opção de referência a de maior preferência segundo o critério que consideremos mais importante. Esta alternativa pode ser flexibilizada para admitir projeções sobre, não apenas uma, mas, sim, o espaço gerado por um pequeno número de direções de referência.

Em situações como a das preferências dos apostadores, em que se dispõe, além das preferências segundo cada critério isoladamente, de uma medida de preferência global, podemos obter um conjunto de pesos básicos ajustando um modelo de regressão. Mas, os pesos derivados do modelo de regressão não precisam ser aplicados uniformemente a todas as opções. Eles podem ser, por exemplo, combinados com pesos a priori associados a diferentes tipos de vetores de preferências.

Podemos fazer variar a importância dos critérios em outras sistemáticas. Se agregarmos através da probabilidade de ocupar o posto de maior preferência segundo algum critério, podemos aplicar um expoente diferente a cada probabilidade de não ser a opção preferida. Com o mesmo fim, no caso de aplicarmos o critério de proximidade à fronteira de excelência, podemos atribuir limites às relações entre os preços-sombra, isto é, às participações de cada critério na medida final.

4. Aplicação ao Caso das Corridas

Esta seção apresenta os resultados de uma investigação empírica da influência da matriz de preferências efetivamente observada sobre a estrutura de pesos dos critérios eventualmente adotada. Os dados são das preferências dos apostadores, em 2001, nos páreos de grandes prêmios da semana em que se realiza, no Rio de Janeiro, o Grande Prêmio Brasil.

Destes páreos foi analisada, em conjunto e isoladamente, a relação entre a distribuição final de apostas e as preferências fornecidas pelo *handicappeur* oficial, ao listar os animais inscritos, e pelos jóqueis. A preferência segundo os jóqueis foi determinada ordenando-os pelo número de vitórias na última temporada hípica e ordenando os animais em concordância com o número de vitórias do jôquei respectivo.

Estes são dois critérios importantes para os apostadores, mas, em vez de dois, múltiplos critérios poderiam ser combinados, para tornar mais realista o modelo. Como todas as alternativas de dados consideradas se estendem trivialmente a mais de dois critérios, preferimos usar neste exemplo o modelo mais simples.

Inicialmente é ajustado um modelo de regressão tendo como variáveis dependentes os vetores de chances relativas dos competidores aos páreos escolhidos e tendo como variáveis explicativas os vetores de preferências segundo os dois critérios: preferência do *handicappeur* e preferência do jôquei. Estas preferências são dadas, no primeiro ajuste, na forma de postos. A seguir são dadas na forma de chances relativas de a opção ser a preferida, calculadas como indicado na Seção 2. A hipótese a investigar é se é possível identificar pesos mais uniformes ao explicar as chances relativas determinadas pelas apostas efetivamente realizadas medindo as preferências através das referidas transformações dos vetores de postos que medindo-as através desses vetores originais de postos. A confirmação desta hipótese torna possível desenvolver uma estratégia de cálculo dos pesos atribuídos aos critérios levando em conta as distribuições de preferência efetivamente observadas.

Depois são investigadas formas de agregação interna anteriores à atribuição de pesos aos critérios. É considerada inicialmente a agregação atribuindo igual importância aos dois critérios e estabelecendo as preferências em termos de proximidade da fronteira da excelência. Duas transformações de variáveis, baseadas nas duas formas de agregação desenvolvidas na Seção 3.2, são consideradas para medir a preferência nesse sentido: chance

relativa de a opção ser a preferida por algum dos dois critérios e proximidade do vetor de preferências pela opção segundo os dois critérios da envoltória determinada pelas opções de maior preferência.

Finalmente, outro modelo é ajustado, considerando a agregação conforme desenvolvido na Seção 3.3, através da projeção dos vetores de postos segundo os dois critérios sobre a direção determinada pelos postos da opção preferida pelo critério mais importante, no caso, o do *handicappeur* oficial. Uma vez determinada a norma ou o módulo desta projeção, aplica-se o mesmo procedimento aplicado na construção do segundo modelo, de derivar da medida de preferência, suposta sujeita a perturbações independentes com distribuição uniforme de amplitude determinada pela distância entre as medições mais afastadas, a chance relativa de cada opção ser a de maior preferência.

Os resultados do ajustamento dos modelos de regressão são apresentados na Tabela 1 abaixo.

Tabela 1.
Regressão da Chance Relativa Extraída das Apostas em Duas Variáveis Explicativas

	Postos	Chance	Excelência	Projeção
R²	15%	49%	59%	70%
F	8,0	43,2	64,5	105,6

Em todas as regressões os coeficientes das variáveis explicativas são significativos ao nível de 1%, exceto o posto segundo os jôqueis na primeira e a chance relativa derivada do módulo da projeção na última. O coeficiente desta última variável é negativo, assim como o do escore de proximidade da envoltória de excelência na regressão sobre as variáveis de agregação com igual importância para os critérios. O valor p correspondente a esta última estimativa embora bem pequeno é, também, consideravelmente superior ao da outra variável explicativa. Isto sugere que, entre as duas variáveis explicativas correlacionadas empregadas, se prefira, na primeira regressão, a indicação do *handicappeur*, na seguinte a chance relativa de ser preferida segundo algum critério e, na última, a chance relativa derivada da preferência medida pela norma da projeção.

A análise dos resíduos das quatro equações é esclarecedora. Examinando ponto a ponto, é fácil perceber que o ajustamento melhora da primeira para a última regressão, à medida que o modelo abre mão de ajustar precisamente o grande número de pontos com a variável dependente próxima da origem, isto é com pequeno volume de apostas. As predições para estes pontos nas últimas regressões sistematicamente superestimam os valores observados. Em contrapartida, a aproximação linear melhor para os pontos representativos de opções de maior preferência eleva as correlações.

A regressão simples da chance relativa derivada das apostas sobre a chance relativa derivada da norma da projeção do vetor de postos sobre o da opção preferida pelo critério principal apresenta um R² de 69% e estatística F igual a 207,6. Quando a variável explicativa da regressão simples é a própria norma da projeção, não se aplicando o cálculo da chance relativa de a opção ser a melhor, o R² cai para 14% e a estatística F para 14,5. O coeficiente de correlação amostral cai de 83% para 37%.

Em resumo, encontramos forte indicação de que a transformação das ordenações de preferência em chances relativas de a opção ser a preferida eleva a precisão das estimativas dos coeficientes dos modelos lineares explicando as chances relativas encontradas na distribuição das apostas. A aplicação desta transformação a variáveis explicativas que combinam os postos, como aquelas construídas projetando os postos sobre direções consideradas mais importantes, também eleva a precisão do ajustamento linear. Fazendo uso desta transformação, se consegue estabelecer sustentação para a conjectura de que o mecanismo de agregação envolve a norma da projeção.

O uso do módulo do vetor de preferências projetadas não resulta em correlações tão altas quanto as obtidas com a norma. Outros critérios de agregação considerados, pela proximidade da posição de excelência segundo algum critério, também se mostram menos aptos a explicar a preferência observada nas apostas.

A Tabela 2 abaixo apresenta as correlações páreo a páreo, entre as probabilidades finais oferecidas pelos apostadores e cada uma das variáveis explicativas consideradas. À medida que se avança da esquerda para a direita nesta tabela, cada par de colunas de correlações corresponde a transformações mais complexas. Nas duas primeiras, as preferências segundo cada critério são dadas apenas pelos postos. Nas duas seguintes, já são dadas pelas probabilidades de a opção ser a preferida. As variáveis das colunas seguintes já são resultado da composição dos dois critérios usando os dois algoritmos desenvolvidos na Seção 3.2. Finalmente, as duas últimas são o resultado da composição por projeção sobre a opção preferida, desenvolvida na Seção 3.3.

Tabela 2.

Correlação da Chance Relativa Derivada de Apostas com Outras Medidas de Preferência

Páreo	Posto oficial	Posto jóquei	Chance oficial	Chance jóquei	Pelo menos um critério	Escore DEA	Norma da projeção	Módulo da projeção
Total	37%	23%	65%	35%	71%	48%	83%	77%
Breno	54%	53%	85%	83%	93%	68%	98%	93%
Tirolesa	43%	30%	64%	58%	71%	44%	83%	71%
Sukow	64%	9%	92%	-16%	62%	52%	86%	80%
Seabra	52%	46%	26%	20%	39%	49%	54%	62%
Presidente	47%	36%	41%	28%	50%	51%	52%	62%
Delegações	82%	-19%	92%	-13%	53%	66%	84%	83%
Mossoró	70%	68%	62%	36%	70%	71%	79%	83%
Brasil	60%	19%	83%	1%	68%	63%	93%	91%
Cidade	42%	48%	39%	44%	50%	42%	49%	49%

O total referido na primeira linha da tabela é dos sete páreos em distâncias habituais para os quais houve ordenação pelo *handicappeur*. Verifica-se, nesse total, uma correlação de 83% entre o vetor de chances relativas derivadas das apostas e chances relativas derivadas da distribuição de probabilidades da norma da projeção sobre a direção da opção preferida pelo

critério dominante vir a ser a maior norma. Este coeficiente de correlação não é atingido por nenhuma outra medida de preferência que se tenha podido derivar das classificações pelos dois critérios básicos considerados.

Os dois grandes prêmios centrais da semana, o próprio Grande Prêmio Brasil e o Prêmio Taça Cidade Maravilhosa, realizado logo em seguida, foram analisados separadamente porque o trabalho do *handicappeur* na ordenação dos concorrentes a estes páreos é realizado conjuntamente, resultando, nesta oportunidade, na designação de dois animais para o número 1 do Prêmio Cidade Maravilhosa e na dúvida sobre a efetiva composição do campo de cada um desses páreos até poucas horas antes da realização dos mesmos. Além disso, um comportamento diferente é esperado do apostador nesses páreos em resposta às indicações do *handicappeur* e à classificação dos jôqueis porque eles se realizam em distância longa, pouco praticada, onde fatores como a experiência do jôquei e a estrutura genética do animal podem ser, excepcionalmente, considerados de maior relevância pelo apostador. Incluem-se no final da tabela as correlações observadas nestas duas provas.

No Grande Prêmio Brasil, a grande favorita dos apostadores, a égua *Canzone*, empinou no momento da largada, ferindo-se, e foi retirada. As apostas foram reabertas por alguns minutos, mas, dada a superlotação do hipódromo e as circunstâncias da retirada, muito poucas das apostas a ela dirigidas foram redirecionadas pelos apostadores. As correlações relativas a este páreo foram, então, calculadas mantendo *Canzone* entre os concorrentes e atribuindo-lhe 30% das apostas. A norma da projeção continua produzindo uma variável explicativa com coeficiente de correlação entre os mais altos observados.

Usando os logaritmos naturais das chances relativas, os resultados obtidos são semelhantes, a vantagem da projeção não sendo, entretanto, tão acentuada. Esta perda de correlação pode ser explicada por resultar da aplicação do logaritmo um afastamento das opções de baixa probabilidade que a aleatorização dos postos e o cálculo da chance relativa tinham aproximado. A Tabela 3 apresenta as correlações, no conjunto dos sete páreos considerados nos totais apresentados na primeira linha da Tabela 2, dos logaritmos das chances relativas derivadas das apostas com as mesmas variáveis explicativas dessa Tabela, substituídas, também, as que representam chances relativas pelos seus logaritmos naturais.

Tabela 3.
Correlações dos Logaritmos das Chances Relativas de Apostas com Outras Medidas

Posto oficial	Posto jôquei	ln chance oficial	ln chance jôquei	ln norma da projeção	ln módulo da projeção	ln pelo menos um critério	Score DEA
58%	34%	66%	41%	71%	66%	65%	65%

5. Conclusão

A estratégia de ordenar as opções e, em seguida, derivar as probabilidades de cada opção ser classificada em primeiro lugar conduziu a distribuições de probabilidades de cada opção ser finalmente escolhida mais correlacionadas com as distribuições de apostas observadas que os vetores de postos. Esta transformação da medida de preferência revela-se efetiva mesmo quando não se parte dos postos, mas de uma medida de preferência resultante da projeção de vetores de postos sobre uma direção escolhida.

Quanto à agregação de critérios, obteve-se evidência a favor do uso da norma da projeção sobre a direção determinada pelo vetor da opção preferida segundo o critério dominante. É interessante observar que para que esta evidência surja faz-se necessário aplicar a transformação das variáveis explicativas em chances relativas.

Os resultados obtidos parecem bastante claros no contexto da formação de preferência de apostadores em corridas. Em particular, a evidência de que a classificação prévia e a classificação dos jôqueis explicam substancialmente a preferência final dos apostadores está fortemente estabelecida. Mecanismos de simplificação e a influência da incerteza também foram tornados explícitos. A extensão das idéias aqui apresentadas a outros contextos exige a suposição de que as motivações para a escolha da melhor opção sejam semelhantes e deve ser, por sua vez, objeto de investigação empírica.

Agradecimentos

Agradecemos as contribuições dos três árbitros anônimos que, de diferentes formas, ajudaram a tornar o trabalho mais legível. Erros que eventualmente permaneçam continuam a ser, é claro, de plena responsabilidade do autor.

Referências Bibliográficas

- (1) Ali, M. (1977). Probability and Utility Estimates for Racetrack Bettors. *Journal of Political Economy*, **85**, 803-815.
- (2) Charnes, A.; Cooper, W.W. & Rhodes, E. (1978). Measuring the Efficiency of Decision Making Units. *European Journal of Operations Research*, **2**, 429-444.
- (3) Farrell, M.J. (1957). The Measurement of Productive Efficiency. *JRSS, A*, **CXX**, **3**, 253-290.
- (4) Gomes, L.F.A.M. & Lima, M.M.P.M. (1992). From Modeling Individual Preferences to Multicriteria ranking of Discrete Alternatives: a Look at Prospect Theory and the Additive Difference Model. *Foundations of Computing and Decision Sciences*, **17**, 172-184.
- (5) Kahneman, D. & Tversky, A. (1979). Prospect Theory: An Analysis of Decisions Under Risk. *Econometrica*, **47**, 263-291.
- (6) Lootsma, F.A. (1993). Scale Sensitivity in the Multiplicative AHP and SMART. *Journal of Multicriteria Decision Analysis*, **2**, 87-110.
- (7) McFadden, D. (1999). Rationality for Economists? *Journal of Risk and Uncertainty, Special Issue on Preference Elicitation*, **19**, 73-105.
- (8) Thaler, R. & Ziemba, W.T. (1997). Parimutual Betting Markets: Racetracks and Lotteries. *Journal of Economic Perspectives*, **2**, 161-174.

ANEXO

As tabelas abaixo apresentam os valores, páreo a páreo, das variáveis básicas da análise realizada.

Chances Relativas de Apostas, Postos, Proximidade da Fronteira, Chances Relativas de Preferência por algum Critério e Derivadas da Norma da Projeção.

Grande Prêmio Breno Caldas

Apostas	Posto oficial	Posto jóqueis	Proximidade da envoltória	Maior preferência por algum critério	Maior preferência pela norma da projeção dos postos
1,41	11	11	1,00	1,71	1,73
0,07	10	9	0,68	0,64	0,33
0,01	9	2	0,42	0,20	0,01
0,12	8	6	0,24	0,13	0,03
0,08	7	7	0,12	0,10	0,03
0,06	6	8	0,24	0,13	0,03
0,02	5	4	0,02	0,01	0,00
0,01	4	5	0,02	0,01	0,00
0,02	3	10	0,68	0,36	0,02
0,04	2	3	0,00	0,00	0,00
0,01	1	1	0,00	0,00	0,00

Prova Tirolesa

Apostas	Posto oficial	Posto jóqueis	Proximidade da envoltória	Maior preferência por algum critério	Maior preferência pela norma da projeção dos postos
0,45	17	17	1,00	1,12	1,16
0,04	16	12	0,75	0,37	0,13
0,06	15	3	0,54	0,20	0,00
0,03	14	15	0,54	0,36	0,19
0,01	13	16	0,75	0,41	0,19
0,02	12	5	0,14	0,05	0,00
0,11	11	10	0,08	0,04	0,01
0,01	10	1	0,04	0,01	0,00
0,17	9	11	0,08	0,03	0,00
0,03	8	14	0,37	0,13	0,01
0,19	7	2	0,00	0,00	0,00
0,02	6	13	0,24	0,08	0,00
0,04	5	9	0,02	0,01	0,00
0,01	4	7	0,00	0,00	0,00
0,01	3	4	0,00	0,00	0,00
0,00	2	6	0,00	0,00	0,00
0,01	1	8	0,01	0,00	0,00

Grande Prêmio Major Sukow

Apostas	Posto oficial	Posto jóqueis	Proximidade da envoltória	Maior preferência por algum critério	Maior preferência pela norma da projeção dos postos
0,65	11	7	1,00	0,73	0,77
0,14	10	8	0,84	0,51	0,57
0,09	9	6	0,43	0,23	0,11
0,11	8	1	0,24	0,10	0,00
0,07	7	9	0,53	0,26	0,10
0,05	6	2	0,05	0,02	0,00
0,06	5	3	0,02	0,01	0,00
0,02	4	4	0,01	0,00	0,00
0,04	3	10	0,68	0,36	0,01
0,04	2	11	1,00	0,65	0,00
0,03	1	5	0,02	0,01	0,00

Grande Prêmio Seabra

Apostas	Posto oficial	Posto jóqueis	Proximidade da envoltória	Maior preferência por algum critério	Maior preferência pela norma da projeção dos postos
0,08	14	4	1,00	0,53	0,49
0,07	13	3	0,72	0,33	0,16
0,25	12	10	0,67	0,29	0,38
0,31	11	11	0,61	0,26	0,21
0,04	10	1	0,18	0,07	0,00
0,12	9	12	0,57	0,25	0,05
0,03	8	14	1,00	0,56	0,03
0,02	7	2	0,02	0,01	0,00
0,16	6	13	0,72	0,34	0,00
0,01	5	5	0,00	0,00	0,00
0,03	4	9	0,10	0,04	0,00
0,01	3	8	0,05	0,02	0,00
0,02	2	6	0,01	0,00	0,00
0,03	1	7	0,02	0,01	0,00

Grande Prêmio Presidente da República

Apostas	Posto oficial	Posto jóqueis	Proximidade da envoltória	Maior preferência por algum critério	Maior preferência pela norma da projeção dos postos
0,10	16	13	1,00	0,67	0,83
0,15	15	6	0,74	0,32	0,04
0,72	14	14	0,86	0,45	0,47
0,06	13	4	0,35	0,13	0,00
0,05	12	10	0,22	0,10	0,04
0,02	11	3	0,13	0,04	0,00
0,02	10	11	0,17	0,07	0,02
0,06	9	16	1,00	0,49	0,11
0,02	8	5	0,01	0,00	0,00
0,05	7	15	0,74	0,32	0,02
0,05	6	12	0,22	0,08	0,00
0,01	5	8	0,01	0,00	0,00
0,02	4	1	0,00	0,00	0,00
0,01	3	2	0,00	0,00	0,00
0,01	2	7	0,01	0,00	0,00
0,01	1	9	0,03	0,01	0,00

Grande Prêmio Delegações

Apostas	Posto oficial	Posto jóqueis	Proximidade da envoltória	Maior preferência por algum critério	Maior preferência pela norma da projeção dos postos
0,46	10	2	1,00	0,70	0,60
0,23	9	5	0,68	0,40	0,37
0,12	8	10	1,00	1,04	0,37
0,26	7	7	0,34	0,20	0,06
0,05	6	9	0,66	0,43	0,03
0,05	5	3	0,04	0,02	0,00
0,01	4	6	0,10	0,05	0,00
0,03	3	1	0,00	0,00	0,00
0,02	2	4	0,01	0,01	0,00
0,03	1	8	0,40	0,20	0,00

Prova Mossoró

Apostas	Posto oficial	Posto jóqueis	Proximidade da envoltória	Maior preferência por algum critério	Maior preferência pela norma da projeção dos postos
0,12	15	9	1,00	0,53	0,29
0,24	14	11	0,85	0,42	0,33
0,21	13	14	1,00	0,60	0,53
0,08	12	10	0,40	0,17	0,07
0,11	11	13	0,62	0,29	0,11
0,01	10	5	0,11	0,04	0,00
0,19	9	12	0,36	0,15	0,02
0,02	8	6	0,03	0,01	0,00
0,02	7	8	0,03	0,01	0,00
0,02	6	1	0,00	0,00	0,00
0,04	5	15	1,00	0,50	0,00
0,03	4	3	0,00	0,00	0,00
0,05	3	7	0,01	0,00	0,00
0,01	2	2	0,00	0,00	0,00
0,01	1	4	0,00	0,00	0,00

Grande Prêmio Brasil

Apostas	Posto oficial	Posto jóqueis	Proximidade da envoltória	Maior preferência por algum critério	Maior preferência pela norma da projeção dos postos
0,43	20	13	1,00	0,52	0,77
0,04	19	8	0,73	0,33	0,15
0,14	18	11	0,51	0,21	0,17
0,09	17	15	0,44	0,17	0,24
0,09	16	7	0,20	0,07	0,02
0,02	15	3	0,11	0,04	0,00
0,04	14	14	0,11	0,04	0,03
0,04	13	17	0,36	0,14	0,04
0,06	12	2	0,01	0,00	0,00
0,01	11	16	0,20	0,07	0,01
0,03	10	18	0,51	0,21	0,00
0,11	9	20	1,00	0,50	0,00
0,02	8	4	0,00	0,00	0,00
0,01	7	9	0,00	0,00	0,00
0,01	6	5	0,00	0,00	0,00
0,01	5	1	0,00	0,00	0,00
0,01	4	19	0,73	0,33	0,00
0,01	3	12	0,01	0,00	0,00
0,01	2	10	0,00	0,00	0,00
0,01	1	6	0,00	0,00	0,00

Taça Cidade Maravilhosa

Apostas	Posto oficial	Posto jóqueis	Proximidade da envoltória	Maior preferência por algum critério	Maior preferência pela norma da projeção dos postos
0,20	19	19	1,00	1,02	1,55
0,02	18	1	0,77	0,30	0,00
0,09	17	4	0,57	0,20	0,00
0,14	16	13	0,40	0,17	0,08
0,03	15	18	0,77	0,41	0,28
0,02	14	2	0,17	0,05	0,00
0,03	13	15	0,27	0,12	0,06
0,08	12	7	0,05	0,02	0,00
0,11	11	14	0,17	0,06	0,02
0,09	10	9	0,01	0,01	0,00
0,01	9	10	0,01	0,01	0,00
0,01	8	11	0,03	0,01	0,00
0,22	7	16	0,40	0,13	0,01
0,05	6	12	0,05	0,02	0,00
0,00	5	6	0,00	0,00	0,00
0,01	4	3	0,00	0,00	0,00
0,01	3	8	0,00	0,00	0,00
0,01	2	5	0,00	0,00	0,00
0,01	1	17	0,57	0,20	0,00