PROBABILIDADES NO ESPORTE*

Probabilities in Sport

BERNARDO NUNES BORGES DE LIMA, GILCIONE NONATO COSTA, RAFAEL NACIFE,

RENATO VIDAL MARTINS Y RODRIGO GUIMARÃES[1]

Resumo

Este trabalho versa sobre Probabilidades no Esporte, de uma forma, assim o esperamos, acessível. Começamos recordando alguns conceitos gerais sobre o assunto que é bem provável que o leitor já tenha, e tentamos argumentar que eles não funcionam se aplicados a Esportes. Isto nos leva naturalmente a uma discussão sobre o que essencialmente são probabilidades. Exibimos dois diferentes modelos de cálculo –para Futebol e Fórmula 1– que servem de exemplos para o que foi dito anteriormente no texto. Tentamos convencer o leitor de que a construção de um tal modelo não requer nenhuma habilidade especial e esperamos que estas linhas forneçam o embasamento necessário caso o leitor queira fazer seu próprio modelo.

Palavras-chave: Esporte, Probabilidades.

Abstract

This work is about Probabilities in Sport within a, hopefully, accesible shape. We start by recalling few general concepts on the subject that the reader quite likely have, and try to argue why they do not work if applied to Sports. This naturally leads us to a discussion on what essentially probabilities are. We exhibit two different models of computation –for Soccer and Formula 1– which stand for examples for what was prioly said in the text. We try to convince the reader that the construction of any such a model requires no special skills and we hope our lines provide the reader with the basic backgrounds in case he is interested in build his own model.

Keywords: Probabilities, Sport.

---

* Este trabalho é parte do Projeto Difundindo Probabilidades via Campeonatos de Futebol, que conta com apoio da FAPEMIG através do Edital de Popularização da Ciência e Tecnologia.

[1] Departamento de Matemática, Universidade Federal de Minas Gerais

O objetivo destas linhas é abordar, se possível, o vasto tema das "Probabilidades no Esporte" em nível divulgativo e, ao mesmo tempo, com um certo rigor científico. O leitor verá que, no caso presente, esta não é que seja uma tarefa tão simples. De fato, quando se diz de um texto que… é capaz de combinar na dose certa o rigor e o informal, o acessível com o hermético, pensa-se no mesmo como que composto de dois momentos: uma parte leve, de leitura fácil, possivelmente rica em conexões históricas, e uma outra restritiva, fechada aos poucos que possam se aventurar. A tal "dose certa" é então aquela arte de se fazer entender sem ser vulgar, ou, ao contrário, de se ser preciso, sem ser enfadonho.

Não. Em nosso caso, a expressão "ao mesmo tempo" é absolutamente literal. Pode-se, e mais ainda, deve-se abordar o tema que propomos de forma divulgativa e simultaneamente rigorosa. "Pode-se" pois, veremos, a matemática que subjaz o binômio "Probabilidades no Esporte" já é de per si extremamente acessível. "Deve-se", pois é importantíssimo que o leitor ponha à prova toda a ciência que julga ter sobre o assunto.

E a verdade é que quando o assunto é "Probabilidades no Esporte" todos viramos cientistas. Seja o torcedor de futebol que já antecipa a festa do título pois seu time tem chances acima de 90% de erguer a taça; ou a escuderia de Fórmula 1 que faz "jogo de equipe" pois apenas um de seus dois pilotos tem probabilidades reais de sagrar-se campeão, ou mesmo o país que planeja estrategicamente o investimento no esporte competitivo baseado em um valor esperado de desempenho olímpico entre os "Top 10"; e até o caso recente de um clube brasileiro que se auto-alcunhou "o time que desafia a Matemática", dada sua incrível capacidade de contrariar previsões probabilísticas desfavoráveis.

As probabilidades definitivamente invadiram o mundo do esporte. Os empresários, as agremiações, a mídia, os atletas, os sites de apostas, os torcedores, e até o cidadão mais comum, sem formação escolar completa, se expressam nestes termos. Sendo assim, o que poderia acrescentear um acadêmico a um tema tão disseminado, tratado com tanta propriedade por todos?

O primeiro passo seria reconhecer o desserviço que podemos estar prestando, muitas vezes sem nos darmos conta, outras vezes de forma consciente e portanto perversa, quando nós matemáticos falamos de "Probabilidades no Esporte". Isto porque probabilidades envolvem dois atos distintos: o conceito e o cálculo. E o que acontece é que a esmagadora maioria das pessoas pensa saber bem o que é probabilidade, quando na verdade não sabe, mas deixa os valores numéricos nas mãos dos matemáticos pois esta é a parte difícil, quando na verdade não é.

Portanto, o matemático deve, primeiro, deixar bem claro o que exatamente quer dizer quando usa o termo “probabilidade” e, depois, mostrar que há pouquíssimo mistério no seu “cálculo”. Na verdade, os modelos que geram tais números costumam ser robustos o bastante para concluirmos que se algum deles no funciona, não é bem porque seu autor não entende de matemática; o que talvez não entenda seja esporte. Em outras palavras, este é um tema em que o matemático deve ser rigorosamente científico no que o leitor julga ser de senso comum e acessível e, por outro lado, divulgar de forma simples – porque é simples– o que o leitor julgar ser ciência quase oculta.

Este é precisamente o objetivo de nosso trabalho. Nele fazemos uma digressão inicial sobre o que vem a ser “probabilidade” e na sequência, nos valemos de dois modelos para explicar como se processa o seu “cálculo”. O primeiro trata de Futebol, e outro aborda a Fórmula 1, dois esportes com forte apelo, o que nos poupará de explicacões adicionais sobre regras e procedimentos.

1. O Conceito e o Cálculo de Probabilidades no Esporte

A probabilidade de sair cara em um lançamento de moeda é 1/2; a de sair o número 1 em um lançamento de dado é 1/6; a de sair um às em um baralho completo é 4/52, e a probabilidade do time, ou do piloto, por que torço ser campeão qual é? Ora, se conseguimos responder as primeiras perguntas, estamos, à priori, aptos a respondermos as outras também. Talvez seja apenas uma questão de tempo de cálculo, pois há mais variáveis a serem consideradas nestes casos.

Em parte isto é verdade, em parte não. Por um lado, sim, em um campeonato de Futebol, Fórmula 1, ou do que for, lidamos com bem mais elementos que em um mero lançamento de uma moeda: são vários times ou pilotos sujeitos às mais diversas combinações de resultados e situações. Por outro lado, se aplicamos ao esporte o mesmo princípio da moeda, podemos, como se verá, não chegar a lugar algum.

Com efeito, a probabilidade de se obter cara em um lançamento de moeda é 1/2, porque analisamos 1 caso –cara– entre 2 possíveis –o outro´e coroa– e faz-se o quociente. Mais precisamente temos

$$\text{probabilidade} = \frac{\text{n}^{\underline{\text{o}}} \text{ de casos favoráveis}}{\text{n}^{\underline{\text{o}}} \text{ de casos possíveis}}, \tag{1}$$

uma fórmula que talvez tenham nos ensinado ainda no ensino fundamental. Existem ao menos dois problemas sérios que surgem se resolvemos aplicar o mesmo raciocínio ao problema de determinar as chances de meu time ser campeão.

O primeiro deles é de ordem computacional, ou seja, o total de casos a serem considerados pode ser assustadoramente grande. Imagine o leitor que nos propomos a responder esta pergunta em um campeonato nacional a 15 rodadas de seu término. Tais campeonatos costumam ter o mesmo formato em diferentes países, e.g., Espanha, Inglaterra, Itália, Portugal e, mais recentemente, também no Brasil. São 20 times que se enfrentam em turno e returno, totalizando 38 rodadas. O que significaria então "casos possíveis" a 15 rodadas do final? Levando-se em conta somente o resultado de cada jogo –vitória de um, de outro, ou empate– em uma rodada são 3^{10} os cenários imagináveis. Em 15 rodadas este nu´mero sobe para 3^{150}. Bem, em tese, 3^{150} (aproximadamente 3×10^{71}) é até tratável em um computador. De fato, para sistemas criptográficos atuais são usados códigos de tamanho até 2^{1024}, como é o caso da nova codificação do principal banco estatal brasileiro. Agora, o que é absolutamente impraticável é gerar em "tempo finito"todos os universos possíveis. Assim, por exemplo, se cada caso fosse gerado em 1 trilhonésimo de segundo (processadores na faixa de Terahertz), o computador gastaria $1{,}18 \times 10^{50}$ séculos para gerar todos eles. A título de comparação, as estimativas atuais para a idade do Universo apontam para $1{,}4 \times 10^{8}$ séculos.

Mas mesmo supondo que nossos computadores fossem de uma geração vindoura, capazes de arrostar grandezas estelares, ainda assim persistiria um outro problema, agora de ordem prática: todos sabemos que nem o primeiro colocado vai perder seus 15 jogos restantes, muito menos o último colocado vencer 15 jogos em sequência. Há muitos cenários que devem ser descartados ou, ao menos, não têm direito de entrar na conta "casos possíveis"com o mesmo peso do que outros cenários bem mais viáveis.

Chegamos a um impasse. Por um lado, o modo comum de se definir probabilidades não serve se aplicado ao esporte, mas, por outro lado, não temos o direito de criar uma outra definição, exclusiva ao caso de campeonatos por exemplo, posto que "probabilidade", enquanto conceito, é uma única coisa qualquer que seja o fenômeno probabilístico a ser estudado.

O interessante é que o dilema acima é solúvel, basta que nos debrucemos com mais vagar sobre a definição (1). Voltemos ao caso da moeda. Será que a tal probabilidade de cara é realmente 1/2? Para entender o problema recordamos o caso que contam do professor que perguntou a seu aluno “qual a probabilidade de sair cara no lançamento de uma moeda?”. Obteve a seguinte resposta um tanto desafiadora: “qual moeda?”. Disse então o professor: “uma moeda não-viciada”, ao que o aluno replicou “o que é uma moeda não viciada?”. E seu mestre assim o esclareceu: “aquela cuja probabilidade de sair cara é 1/2”. “Ora, então porque o senhor perguntou?” disse o rapaz com toda razão. Em suma, o aluno chamava atenção para o que, sem dúvida, foi a verdadeira pergunta de seu professor: “qual a probabilidade de sair cara no lançamento de uma moeda cuja probabilidade de sair cara é 1/2?”. A resposta, é claro, é 1/2, e de nada nos interessa.

Não é este tipo de resolução que queremos para os nossos problemas. Eles são mais complexos pois todos sabemos que a natureza tem seus graus de imperfeição: a rigor, não existem nem moedas não-viciadas, muito menos times ideais. A equação (1) não é propriamente uma definição de probabilidade. Ao contrário, trata-se de um mero exercício de “Análise Combinatória”, uma outra ciência que, em muitos casos, chega a distar da “Teoria de Probabilidades” como o nascente do poente.

Então, afinal de contas, o que´e probabilidade? O conceito exato veio bem recentemente, ao passo que a definição precisa jamais virá. Expliquemo-nos. Em 1933, Andrei Kolmogorov (1903-1987) estabelece no livro *Fundamentos da Teoria das Probabilidades*, os pilares formais de tal teoria. O enfoque dado era aos moldes do que fez o matemático e filósofo grego Euclides (360 a.C.-295 a.C) para geometria e que voltava à voga, com toda sua força, no início do século passado não apenas no caso da Teoria das Probabilidades mas na própria Matemática como um todo. Era um enfoque axiomático, ou seja, os conceitos não se definem, simplesmente são reconhecidos pelas propriedades que satisfazem.

Se isto pode parecer estranho, lembre o leitor que Euclides não definia nem ponto ou reta mas, por exemplo, dizia que por um ponto fora de uma reta passa uma única paralela, seu quinto axioma, que no porvir, acabou sendo um divisor de águas entre diferentes geometrias. O que Kolmogorov tinha em mente era exatamente a mesma tática, e queria identificar quais são as regras básicas que caracterizam um fenômeno probabilístico.

Tais regras são o que hoje se conhece por Axiomas de Kolmogorov. A título de exemplo, fazemos a tradução destes postulados para o caso do Campeonato Espanhol. Kolmogorov se limitaria a dizer:

(i) a probabilidade do seu time ser campeão é pelo menos 0%;

(ii) a probabilidade de um time madrilenho ser campeão é a soma das probabilidades de Real Madrid e Atlético de Madrid o serem;

(iii) a probabilidade de um time espanhol ser campeão é 100%.

O que é absolutamente surpreendente é que isto é mais do que bastante para inferirmos quem sera´ o campeão espanhol. O que queremos dizer é que partindo dos Axiomas de Kolmogorov –dos quais os três itens acima são mera adaptação ao caso do futebol– podemos deduzir matematicamente a seguinte fórmula

$$\text{(2)} \qquad \text{probabilidade} = \lim_{n\to\infty} \frac{\text{n}^{\underline{\text{o}}} \text{ de ocorrências em } n \text{ ensaios}}{n}.$$

conhecida por Lei dos Grandes Números. Ou seja, a probabilidade de um evento é a razão entre o número de ocorrências do mesmo e o número de ensaios quando este tende a infinito. Isto não é uma definição de probabilidade, mas apenas um modo de calculá-la ou, até melhor, inferí-la. Esta afirmação, antes dos Axiomas de Kolmogorov, era a base da *teoria frequencista*, que dava à teoria de probabilidades um caráter, a princípio, mais empírico que matemático. A dedução da Lei dos Grandes Números a partir dos Axiomas de Kolmogorov nos mostra que o “empirismo” em questão já não guarda mais eventuais resquícios pejorativos que o termo possa ter (teste, tentativa, suposição, apriorismo, etc). Ao contrário, o “frequencismo” passa a ser então a nossa grande ferramenta para se estimar probabilidades que (sempre) desconhecemos. Mais ainda, está muito bem fundamentado do ponto de vista teórico, com todo o rigor matemático.

No entanto probabilidade segue sendo algo pré-existente às coisas e por isso mesmo um mistério, como, por exemplo, as chances de um time vencer, jogando em casa, um determinado adversário. Esta propensão é algo real e depende de uma série de fatores, desde o plantel atual ao momento na competição, do juiz à susceptibilidade do outro time à torcida alheia etc, etc. E o único modo de sabermos a probabilidade exata de vitória do time da casa é repetir, centenas, milhares, milhões de vezes o mesmo jogo, contra o mesmo adversário, com os mesmos jogadores, no mesmo campo, sob as mesmas circunstâncias, mesmo público, e o que, todos sabemos, pode de ser determinante: com o mesmo juiz! Da mesma forma, se quisermos realmente saber qual a probabilidade de um dado time ser campeão, o que temos de fazer é repetir inúmeros campeonatos e ver em quantos deles o clube leva a melhor. Quanto mais edições, mais preciso o cálculo. Um total de 3 títulos em 4 campeonatos jogados é bem menos significativo do que, digamos, 612 conquistas em 1000 disputas. A probabilidade real da equipe erguer a taça não é então de 75%, como a princípio se cogitava, mas deve girar em torno de 60%.

Ora, o leitor arrazoado dirá que é impossível repetir mil campeonatos e com razão. Mas não iremos “realizar” mil campeonatos e sim “simular” não só mil, mas milhões deles se necessário. O computador faz isto em segundos ou, quando muito, em minutos. Tudo o que temos de fazer é ensiná-lo a “sortear”de forma judiciosa o resultado de cada jogo.

2. Um modelo de Caáculo para o Futebol

Para calcularmos a probabilidade de um time ser campeão, partimos da situação atual do campeonato, e simulamos os jogos restantes. Ao final, o computador gera a classificação dos times e a registra. Depois, repete o procedimento várias vezes e no término do processo já está apto a divulgar números que aproximam muito bem as probabilidades geradas pelo modelo idealizado. Este número é o quociente entre os campeonatos (simulados) vencidos pelo time em questão e o total de disputas. Para que isto funcione, o número de simulações deve ser grande o bastante dada a precisão que exigimos em nossos cálculos. A modo de exemplo, suponha que simulando 10, 100, 1000, 10000 e 100000 vezes o campeonato, o tal time foi campeão, respectivamente, 7, 65, 677 e 6651, 66597 vezes. Já podemos, a princípio, dizer que tal probabilidade deve oscilar em torno de 66%. Se queremos precisão centesimal, devemos nos ater aos quatro primeiros algarismos do número de títulos. A medida que se aumenta o número de simulações, estes algarismos variam cada vez menos até estacionarem. Este é então o número de simulações ideal para o grau de precisão, no caso decimal, que queremos. Mais, estes algarismos pouco variam se repetirmos este processo uma ou várias vezes, fixado um número suficientemente grande de simulações, o que em linguagem matemática equivale a *convergência*.

Então, como vimos, todo trabalho fica a cargo do computador, que irá simular campeonatos. Para tal, devemos ensiná-lo a simular um jogo e o procedimento é o seguinte: para determinar o resultado de A x B, o computador recebe dois vetores que caracterizam, naquele momento, os times A e B: (PV_A, PE_A, PD_A) e (PV_B, PE_B, PD_B) que são as probabilidades de vitória, empate e derrota de cada um deles. Chamamos uma trinca destas de "vetor força" do time. A partir destes dados, formam-se as probabilidades do jogo:

$$\mathrm{P_{AxB}} = \left(\frac{\mathrm{PV_A} + \mathrm{PD_B}}{2}, \frac{\mathrm{PE_A} + \mathrm{PE_B}}{2}, \frac{\mathrm{PD_A} + \mathrm{PV_B}}{2} \right)$$

onde as três coordenadas são, na ordem, as probabilidades de vitória de A, empate, e vitória de B. Supondo, e.g., que P_{AxB} = (0.5, 0.2, 0.3), o que o computador faz é dividir o intervalo [0,1] em três partes: [0, 0.5], (0.5,0.7) e [0.7,1]. E então sorteia um número aleatório entre 0 e 1 (ele sabe como fazer isto). Se, por exemplo, o número sorteado foi 0.4579, então o resultado do jogo é vitória de A, pois 0.4579 está entre 0 e 0.5 e esta foi a parte do intervalo que reservamos para vitória de A. Se os números sorteados fossem 0.61 ou 0.9999993 o computador os identificaria como, respectivamente, empate ou vitória de B.

Depois de simular uma rodada, atualizam-se todos os vetores força de todos os times, de acordo com os resultados da rodada, e simula-se a rodada seguinte. O modo de atualizar tais vetores é precisamente o que caracteriza um modelo. De fato, o leitor já deve ter notado que hoje há inúmeros sites que ofertam probabildades para o futebol. O normal é que todos sigam o roteiro acima descrito, porém a maneira de medir a força de um time pode variar bastante de modelo para modelo. E a força de um time, quantificada através de seu vetor força, é algo que se mede pelos resultados que vem obtendo, em especial o da partida que acabou de disputar, daí ser este o ponto chave na especificação de um modelo.

Em 2005, os autores deste artigo, e mais os Profs. Fábio Brochero e Marcelo Terra Cunha, iniciamos no Dep. Mat. da UFMG, um projeto tendo em vista a divulgação da Matemática através de probabildades no esporte. A idéia era subsidiar quem quer que se interesasse a tentar criar um modelo de cálculo, inicialmente, para o futebol. Achamos por bem então desenvolver nosso próprio modelo –cujos cálculos efetivos encontram-se disponíveis em www.mat.ufmg.br/futebol– que descrevemos na sequência. É muito mais um exercício teórico do que propriamente científico pois, como dissemos, nosso interesse é bem mais divulgativo. Ou até melhor, a ciência que um modelo destes exige é acessível o bastante a que cada qual possa, se quiser, desenvolver o modelo que mais lhe aprouver.

Pois bem, voltando aos vetores força, o procedimento que padronizamos para atualizá-los após uma rodada é o seguinte. Primeiro atribuimos a cada time participante do campeonato, na verdade, dois vetores de força

PC=(PVC,PEC,PDC) e PF=(PVF,PEF,PDF)

onde PVC, PEC e PDC ser˜ao utilizados, respectivamente, no cálculo das probabilidades de vitória, empate e derrota do clube em uma partida jogando como mandante (em “casa”), e PVF, PEF e PDF seguem a mesma lógica, desta feita sendo a equipe em questão visitante (joga “fora de casa”).

A cada rodada, os vetores de força são realimentados. Se o time jogou como mandante, muda-se o seu vetor mandante, e o mesmo vale para seu vetor visitante, caso o jogo tenha sido fora de casa. A ideia de se tratar um mesmo time como dois diferentes – um em casa, e outro fora dela– reflete o modo como a torcida, ou o próprio campo onde a equipe habitualmente joga, podem influenciar no desempenho do clube.

A primeira premissa da qual partimos é a de que, se um time vence uma partida, aumenta sua probabilidade de vencer a partida seguinte. Além disso, este aumento na probabilidade de vitória é tanto maior quanto melhor for o time derrotado. Analogamente, se o time perde uma partida, aumenta sua probabilidade de perder o próximo jogo e o aumento é tanto maior quanto pior for seu adversário.

Podemos entender o modo como modificamos os tais vetores de probabilidade, através de um exemplo. Vamos supor que houve o jogo Real Madrid x Barcelona, disputado no Santiago Bernabeu, e que os merengues foram os vencedores do confronto. Então o novo vetor de força do Real Madrid como mandante, digamos , é obtido a partir do vetor anterior, digamos , da seguinte forma[2]

[2] Vetores tridimensionais são objetos que apresentam 3 coordenadas. Existem duas operações básicas, a saber, soma de vetores e multiplicação de um vetor por um número real definidas dessa maneira: se **u** = (a,b,c), **v** = (x,y,z) e *k* um número real então **u** + **v** = (a + x, b + y, c + z) e *k* · **v** = (*ka*, *kb*, *kc*).

$$\mathrm{PC}^{\mathbf{1}}_{\mathrm{Real}} = \frac{p\cdot \mathrm{PC}^{\mathbf{0}}_{\mathrm{Real}} + r_{\mathrm{Barça}}\cdot(1,0,0)}{\mathrm{PC}^{\mathbf{0}}_{\mathrm{Real}} \qquad p + r_{\mathrm{Barça}}} \ \mathrm{PC}^{\mathbf{1}}_{\mathrm{Real}}$$

onde p é o “peso” que damos ao passado e $r_{\mathrm{Barça}}$ é o rendimento do Barcelona (que também deve ser modificado para o próximo jogo) dado pela seguinte fórmula que vale para todos os clubes:

$$r = \frac{\text{nº de pontos conquistados}}{\text{nº de pontos disputados}}.$$

Por exemplo, se $p = 0$ significa que o passado não tem importância alguma, e o fato de que o Real Madrid venceu o Barcelona no Santiago Bernabeu daria aos merengues “força máxima” para a próxima partida que jogasse em casa. Por outro lado, se damos a p um valor muito alto, os vetores de força seguem quase que inalterados jogo a jogo, e perdemos o “fator moral” trazido pelas vitórias. Ou seja, p indica a sensibilidade dos times em relação ao resultado de uma partida. Quanto ao rendimento r do adversário, o que a fórmula acima diz é que, se o Barcelona tivesse, por exemplo, um rendimento muito baixo, então a vitória do Real não seria tão significativa a ponto de alterar substancialmente suas chances de vitória.

Da mesma forma, após a derrota no Santiago Bernabeu, o vetor visitante do Barcelona´ e modificado do seguinte modo

$$\mathrm{PF}^{\mathbf{1}}_{\mathrm{Barça}} = \frac{p\cdot \mathrm{PF}^{\mathbf{0}}_{\mathrm{Barça}} + (1 - r_{\mathrm{Real}})\cdot(0,0,1)}{p + (1 - r_{\mathrm{Real}})}.$$

Note que aqui aparece $1 - r_{\mathrm{Real}}$ (e não r_{Real}, como na anterior) pois o raciocínio se inverte: perder, por exemplo, para um time hipotético que sempre vence ($r = 1$) não significa absolutamente nada ($1 - r = 0$ e o vetor força fica intacto).

Como se viu, vencer uma partida aumenta a probabilidade de vencer a próxima, uma vez que a componente PV aumenta e as demais componentes diminuem. Simetricamente, a probabilidade do time derrotado ser novamente derrotado aumenta. Contudo, o mesmo não se aplica em caso de empate entre as equipes. Ou seja, empatar uma partida não aumenta as chances de empatar a próxima e diminuam as chances de vencer ou perder. De fato, suponhamos que um time, jogando como visitante, empatasse com o líder do campeonato. É razoável supor que a sua probabilidade de vencer uma partida, como visitante, também aumente. Da mesma forma, se o empate é com o lanterna do campeonato, a probabilidade de perder a próxima partida´e que deve ser aumentada. A grosso modo, empate com time bom é "meia vitória", empate com time ruim´e "meia derrota".

Para expressar este raciocínio em termos numéricos, montamos as fórmulas que se seguem. Para entendê-las, voltemos ao clássico Real Madrid x Barcelona no Santiago Bernabeu, e suponha que houve empate. Se $r_{\text{Barça}} \leq 1/2$ então o Barcelona não está tão forte no campeonato, e atualizamos o vetor força do Real Madrid como mandante da seguinte forma

$$\mathrm{PC}^{\mathbf{1}}_{\mathrm{Real}} = \frac{p \cdot \mathrm{PC}^{\mathbf{0}}_{\mathrm{Real}} + (1 - 2\,r_{\mathrm{Barça}}) \cdot (0, \frac{1}{2}, \frac{1}{2}) + 2\,r_{\mathrm{Barça}} \cdot (0,1,0)}{p+1}.$$

Ou seja, $(0, \frac{1}{2}, \frac{1}{2})$ é a tal "meia derrota" que dissemos acima e (0, 1, 0) é o empate puro. Entendemos melhor a fórmula considerando os casos extremos. Se os azul-grenás tiverem rendimento $r_{\text{Barça}} = 1/2$, então o segundo termo do numerador se anula e o terceiro é (0, 1, 0), ou seja, apenas a tendência ao empate do Real Madrid em casa será aumentada. Mas supondo, hipoteticamente, que o Barcelona sequer pontuou no campeonato, seu rendimento é nulo e o terceiro termo do numerador desaparece, sendo o segundo igual $(0, \frac{1}{2}, \frac{1}{2})$ .Ou seja, neste caso o Real Madrid teve uma "meia derrota" para o Barcelona apesar de ter, na prática, empatado o jogo, o que deve fazer crescer tanto sua tendência a futuros empates quanto derrotas. Os casos intermediários $0 < r_{\text{Barça}} < 1/2$ correspondem à média ponderada entre os vetores $(0, \frac{1}{2}, \frac{1}{2})$ e (0, 1, 0), dependendo do rendimento do time catalão.

Agora, se o Barcelona vem forte no campeonato, digamos com rendimento $r_{\text{Barça}} > 1/2$, então o empate do Real Madrid está mais para "meia vitória" e seu vetor probabilidade é atualizado do seguinte modo

$$\mathrm{PC}^{\mathbf{1}}_{\mathrm{Real}} = \frac{p \cdot \mathrm{PC}^{\mathbf{0}}_{\mathrm{Real}} + (2\, r_{\mathrm{Bar}}\, (0, \frac{1}{2}, \frac{1}{2})\ (\frac{1}{2}, \frac{1}{2}, 0) + 2(1 - r_{\mathrm{Barça}}) \cdot (0, 1, 0)}{p+1}.$$

Após o empate, o vetor probabilidade do Barcelona como visitante é atualizado de forma absolutamente análoga levando em conta o rendimento r_{Real} do Real Madrid.

3. Um Modelo de Cálculo para a Fórmula 1

No início de 2010, resolvemos estender nosso projeto a Fórmula 1. A idéia era montar um modelo que calculasse as probabilidades de cada piloto vencer o campeonato (ou terminar em uma determinada posição), e a partir deste fornecer informações adicionais como, por exemplo, a pontuação esperada de cada um.

Como era de se esperar, o primeiro passo era ver, do ponto de vista probabilístico, o que a Fórmula 1 tem em comum com o Futebol posto que já tínhamos um modelo em mãos. E, de fato, a teoria acima descrita se aplica da mesma forma se substituímos times por pilotos, jogos por corridas e a pontuação do Futebol por partida (V=3, E=1, D=0), pela pontuação da Fórmula 1 por prova (1º=25, 2º=18, 3º=15,…). E para calcular as probabilidades que queremos, o computador irá simular inúmeros campeonatos, e, para tal devemos, como antes, ensiná-lo a simular uma corrida.

O problema começa aí. Uma corrida de Fórmula 1 é bem mais complexa do que um jogo de Futebol. Não são apenas 3 cenários cabíveis (vitória de um, de outro, e empate) como no Futebol, mas supondo que há 20 pilotos, este número sobe para 20! e é muito grande. Com efeito, o computador teria de montar o vetor probabilidade da corrida, nos mesmos moldes do vetor probabilidade de um jogo, ponderando as chances de cada um dos pilotos em cada uma dos 20! cenários, depois dividir o intervalo de 0 a 1 em 20! pedaços e então fazer o sorteio, gerando um número aleatório. O tempo de processamento, para tal, é enorme e descartamos esta possibilidade.

Nossa primeira tentativa foi criar um modelo que fosse o mais simples possível e analisar sua viabilidade. Em vez de “vetor força”, cada piloto teria um “valor força”, i.e., bastaria um único número para descrever a força de cada piloto. Esta variava de 0 a 10, dividida em uma parte fixa (de 0 a 7) dependendo da equipe, e outra variável (de 0 a 3) de acordo com os resultados do piloto. Dividíamos então o intervalo [0, 1] em 20 partes proporcionais ao desempenho de cada piloto, e fazíamos o sorteio para o primeiro colocado. Na sequência, o computador excluía o piloto sorteado, redimensionava o intervalo de acordo com as forças dos restantes, sorteava o segundo colocado, e assim sucessivamente.

O modelo definitivamente não funcionou. Há vários modos de se verificar se os resultados matemáticos contrastam com o mundo real e aí é fundamental entender o que se pretende modelar. No ano de 2010, três equipes dominaram o circo da Fórmula 1 –BAR, Ferrari e McLaren– ao passo que três delas –Lotus, Hispania e Virgin– estavam muito aquém de qualquer outra, às vezes 4 segundos mais lentas por volta do que as demais, o que é uma enormidade. Portanto, um sorteio que tenha um piloto da Virgin è frente de um ferrarista deve ser muito raro. O problema é que nas nossas simulações, isto não acontecia com a raridade prevista. Mesmo aumentando a força fixa das equipes grandes e reduzindo a das pequenas o problema persistia, embora em menor escala. Um modo de se ver isto era através da pontuação esperada. O normal seria que a pontuação final dos pilotos da Lotus, Hispania e Virgin oscilasse entre 0 e 2 pois só pontuariam em uma corrida muito atípica com chuva e desastres, sendo que na prática estes eram os carros menos preparados para as “wet races” além de, em geral, serem os primeiros a bater. Ao contrário, em nossas simulações tais pilotos terminavam o campeonato, como esperado, nas últimas colocações, mas com pontuação que variava entre 6 e 10. Detalhe: todos eles terminaram o campeonato de 2010 sem ponto algum.

O interessante é que o melhor modo de se chegar a um modelo condizente com a realidade é justamente se afastando dela. Vendo as entrevistas dos pilotos antes das provas, temos a impressão que todos querem ganhar e, dada a largada, todos disparam em busca da vitória. Porém, do ponto de vista meramente formal e abstrato –que é como o fenômeno será modelado– a história de uma corrida pode ser muito diferente. Suponhamos, por exemplo que o nosso referencial é o primeiro colocado e convencionamos que o mesmo está parado, sendo que os outros se movem em relação a ele. Nos atendo somente ao movimento do último colocado, temos a impressão que ele se afasta da primeira colocação como se estivesse fugindo de um bandido. A última coisa que nos passaria pela cabeça é que um tal piloto entrou para vencer. Ao contrá rio, talvez o modo certo de se encarar uma corrida de Fórmula 1 seja compará-la a um jogo em que 20 coelhos partem em direção a 20 casas numeradas de 1 a 20. Há coelhos que disputam as primeiras casas, outros com tendência às casas intermediárias, e alguns coelhos que estão muito felizes nas últimas casas e não há nada nesse mundo que os tire de lá.

Tendo em mente o dito acima, montamos um modelo que se mostrou bastante satisfatório após vários testes e se baseia na fórmula de Poisson. Criamos, para cada corredor X, um vetor força

$$P_X = (P_1, P_2, P_3, \ldots, P_i, \ldots, P_{20})$$

onde P_i indica a probabilidade do piloto X chegar na i-ésima colocação. O valor de P_i é dado pela equação

$$\mathrm{P}_i = \frac{e^{-\mu_X} \cdot (\mu_X)^{i-1}}{(i-1)!}$$

em que μ_X é a média das colocações de X nas corridas até então disputadas e e = 2, 7128… é a base do logaritmo natural. Se por exemplo, após 5 provas, X venceu as duas primeiras, depois chegou em quarto, sexto e último na quinta prova após bater na largada temos

$$\mu_X = \frac{1+1+4+6+20}{5} = 6,4$$

e deverá ser atualizado após a próxima prova. Repare o leitor que quanto mais forte for o piloto X, menor será o μ_X que varia entre 1 (X venceu todas as provas) a 20 (X chegou em último em todas elas). Portanto, pode-se verificar que quanto melhor o piloto maior será Pi pois $\mu_X \geq 1$. Nos raciocínios acima sempre supomos que são 20 os pilotos que disputam o campeonato, mas é lógico que podemos ajustálo ao número exato de pilotos na temporada em questão[3]. Ao tomarmos a média simples da classificação de cada piloto, situações adversas como uma corrida sob forte chuva ou forte calor, que afeta consideravelmente o desempeho dos pneus, são consideradas indiretamente pela média μ_X, uma vez que não teríamos um retrospecto considerável de cada piloto em cada uma dessas situações.

[3] Verifica-se matematicamente que à medida que o número de corridas aumenta, a soma das componentes de P_X vai se aproximando de 1, ou, em linguagem matemática, $\sum_{i=1}^{\infty} \mathrm{P}_i = 1$. Por isso os P_i são, de fato, probabilidades.

No início de qualquer campeonato de Fórmula 1, certos pilotos são considerados favoritos enquanto outros meros coadjuvantes. E essas expectativas são descritas nos valores iniciais de P_X de cada piloto. Logicamente, nem sempre essas expectativas se confirmam ao longo do campeonato. Portanto, por uma questão de equidade, poderíamos considerar que todos os pilotos iniciam com a mesma probabilidade de ser campeão, e o modelo levaria algumas corridas para aferir, com melhor precisão, o desempenho de cada piloto. Além disso, as probabilidades são calculadas sem a consideração do *grid* de largada de cada corrida. Em certas provas, como a de Mônaco, esse fator é fundamental na classificação final. Ao ser determinado o *grid*, pode-se, por exemplo, balancear a média μ_X de cada piloto com o restrospecto médio dos pilotos que largam na mesma posição do *grid*.

Fazemos então a simulação para a primeira colocação vendo a componente P_1 de cada piloto e dividindo o intervalo [0,1] em 20 partes proporcionais a estas componentes. O computador sorteia um número entre 0 e 1 e, de acordo com a parte do intervalo em que cair, escolherá o vencedor da prova. Na sequência, exclui este piloto e redivide o intervalo de acordo com a componente P_2 de cada um dos 19 pilotos restantes e obtém o segundo colocado após novo sorteio. Depois, repete o mesmo processo até que sobre apenas um piloto que será decretado o último na prova. Com esses cálculos, obtemos valores como a probabilidade para o piloto ganhar a próxima corrida, ganhar o campeonato, a pontuação esperada, e outros dados que podem ser de interesse do público em geral.

Referências

[1] B. James, *Probabilidade: um Curso em Nível Itermediário*, Rio de Janeiro, Projeto Euclides, 3a edição (2008).

[2] F. Brochero, G. N. Costa, M. Terra Cunha, B. N. B. de Lima, R. V. Martins, “Futebol: uma Caixinha de… Sorteios”, *Ciência Hoje*, 254 (2008), pp. 24-29.

[3] I. Richard, *The Pleasures of Probability*, Nova Iorque, Springer-Verlag (1995).