ISSN: 2317-0840

Modelo de distribuição de probabilidades para o número de bolas chamadas até que alguém "bata" em um bingo convencional

Pedro F. Lima[1†], Claúdio T. Cristino[2], Paulo R. A. Firmino[3], Cícero Carlos F. Oliveira[4]

[1] *Doutorando pelo Programa de Pós-Graduação em Biometria e Estatística Aplicada - UFRPE/Recife-PE, professor adjunto pela URCA/Juazeiro do Norte-CE*

[2] *Professor do Programa de Pós-Graduação em Biometria e Estatística Aplicada da UFRPE/Recife-PE*

[3] *Professor adjunto da UFCA e Colaborador do Programa de Pós-Graduação em Biometria e Estatística Aplicada da UFRPE/Recife-PE*

[4] *Doutorando pelo Programa de Pós-Graduação em Biometria e Estatística Aplicada - UFRPE/Recife-PE, professor assistente pelo IFCE - Campus Crato*

Resumo: *Este trabalho tem como objetivo a obtenção de um modelo de distribuição de probabilidade para a variável aleatória que conta o número de bolas chamadas até que alguém bata em jogos de bingos convencionais. Partindo de um modelo básico desenvolvido para um bingo com exatamente uma cartela, foi utilizada a distribuição do mínimo de* n *variáveis aleatórias independentes para expandir o modelo proposto para* $n = 200$, $n = 1000$, $n = 2000$, $n = 5000$ *e* $n = 10000$ *cartelas. A partir de uma rotina no* software *R, foram simuladas 200 execuções do bingo para os diversos números de cartelas acima citados. Com os resultados obtidos, foram realizados testes de aderência de Kolmogorov-Smirnov para verificar a qualidade do modelo desenvolvido. Todos os modelos testados foram aprovados ao nível* $\alpha = 0,05$. *O trabalho é finalizado fazendo-se uma comparação entre as principais medidas de tendência de centro: média, moda e mediana, calculadas quando são utilizadas as distribuições empíricas obtidas a partir das simulações e a teóricas construídas. Também foi observado que as diferenças entre as distribuições, tanto as diferenças absolutas, quanto as diferenças relativas, são bem pequenas, reforçando o que havia sido concluído anteriormente.*

Palavras-chave: Bingo, modelos probabilísticos, simulação, bolas chamadas.

Abstract: *This paper aims to obtain a model of the probability distribution for the random variable that counts the number of balls called until someone hit conventional bingo games. Starting from a basic model developed for a bingo with exactly one card, the distribution of the minimum of* n *independent random variables was used to expand the proposed* $n = 200$ *model,* $n = 1000$, $n = 2000$, $n = 5000$ *and* $n = 10000$ *cards. From a routine in* R, 200 plays bingo were simulated for different numbers of cards cited above. With the results, adherence Kolmogorov-Smirnov tests were performed to verify the quality of the developed model. All models tested were approved at the level $\alpha = 0.05$. The work is finished by making a comparison between the main measures of central tendency: mean, median and mode, calculated when used empirical distributions obtained from simulations and theoretical construction. It was also observed that the differences between the distributions, both the absolute differences, the relative differences are very small, reinforcing what had been previously.

Keywords: bingo games, probability models, simulation, balls called.

†Autor correspondente: limapf@yahoo.com.br.

Introdução

Considere, em um bingo, o problema de se estudar as incertezas intrínsecas ao número de bolas chamadas até que alguém tenha todos os números de sua cartela sorteados. Neste caso é comum dizer que esta pessoa "bate" o jogo, isto é, vence-o. O primeiro passo é definir a variável aleatória

$$X \equiv \text{"Número de bolas chamadas até que alguém bata no bingo"}.$$

Em bingo convencional cada cartela é formada por vinte e quatro números distribuídos em cinco colunas, conforme o exemplo na Figura 1, a seguir:

B	I	N	G	O
01	17	31	46	62
04	20	33	49	65
06	22	●	51	70
08	25	38	56	73
15	28	42	58	75

Figura 1: Cartela típica de um bingo.

Nesse tipo de bingo, na primeira coluna das cartelas, são colocados números de 1 até 15, na segunda de 16 a 30, na terceira de 31 a 45, na quarta de 46 a 60 e na última de 61 a 75. Observe que a terceira coluna só podem ser colocados quatro números.

É importante observar que o número de cartelas efetivamente distintas que podem ser produzidas é de:

$$({}_{15}C_5)^4 \times_{15} C_4 = 111.007.923.832.370.565,$$

em que ${}_nC_k$ representa o número de combinações diferentes de k elementos selecionados dentre n possíveis.

Material e Métodos

Para construir o modelo, parte-se aqui de um caso simples considerando um bingo composto por apenas dez números e com cartelas contendo apenas três números. Será utilizado um diagrama em árvore para visualizar um modelo que descrevesse bem o caso. Então este caso simples será generalizado para o modelo de um bingo composto por 75 números e com cartelas contendo 24 números.

Uma rotina no ambiente de programação para estatística computacional R CORE TEAM (2012) foi criada para gerar grupos com diversas quantidades de cartelas e simular a execução do jogo contando o número de bolas chamadas até a batida. Com base nessas simulações, distribuições empíricas foram computadas e por meio do teste de aderência de Kolmogorov-Sminorv, a qualidade do ajuste foi medida. Em todos os casos testados, desvios insignificantes foram observados não reprovando as hipóteses de aderência do modelo testado.

Resultados e Discussão

O número mínimo de bolas sorteadas para que alguém possa bater é de: 24, que é o total de números presentes em cada cartela e o máximo é 75. Mas para se ter a certeza que alguém baterá ao sortear a vigésima quarta bola, é necessário que todas as cartelas possíveis entrem no jogo. De fato, os 24 primeiros números podem ser sorteados de ${}_{75}C_{24} = 25.778.699.578.994.555.700$ modos. Se a cartela que contém os primeiros 24 números sorteados foi produzida e vendida, então alguém baterá nesse momento. Considerando que os sorteios são completamente casuais, e que a cartela possa ter sido produzida, então a probabilidade de uma dessas cartelas ser a premiada é de:

$$P(X = 24) = \frac{{}_{24}C_{24}}{{}_{75}C_{24}} = \frac{1}{{}_{75}C_{24}}$$

Até esse momento apenas uma combinação ${}_{24}C_{24}$ foi istanciada de um total de: ${}_{75}C_{24}$. Se ninguém bateu é porque ou a cartela não foi produzida ou, no caso de ter sido produzida, não foi vendida. Esta deve então ser excluída do espaço das amostras sobre o qual X opera. Após este procedimento ainda restarão: ${}_{75}C_{24} -_{24} C_{24}$.

A probabilidade de alguém bater ao sortear 25 números é de:

$$\begin{aligned} P(X = 25) &= \frac{{}_{75}C_{24} - {}_{24}C_{24}}{{}_{75}C_{24}} \times \frac{{}_{25}C_{24} - {}_{24}C_{24}}{{}_{75}C_{24} - {}_{24}C_{24}} \\ &= \frac{{}_{25}C_{24} - {}_{24}C_{24}}{{}_{75}C_{24}} \end{aligned}$$

Deve-se observar que nesse momento podem existir até ${}_{25}C_{24} -_{24} C_{24} = 24$ combinações de 24 números com possibilidades de serem premiadas. Mas pode ocorrer de nenhuma dessas 24 combinações não ter sido vendida ou mesmo produzida. Isto ocorrendo será necessário o sorteio de mais um número. Daí:

$$\begin{aligned} P(X = 26) &= \frac{{}_{75}C_{24} - {}_{24}C_{24}}{{}_{75}C_{24}} \times \frac{{}_{75}C_{24} - {}_{25}C_{24}}{{}_{75}C_{24} - {}_{24}C_{24}} \times \frac{{}_{26}C_{24} - {}_{25}C_{24}}{{}_{75}C_{24} - {}_{25}C_{24}} \\ &= \frac{{}_{26}C_{24} - {}_{25}C_{24}}{{}_{75}C_{24}} \end{aligned}$$

Seguindo de maneira análoga obteremos:

$$\begin{aligned} P(X = k) &= \frac{{}_{75}C_{24} - {}_{24}C_{24}}{{}_{75}C_{24}} \times \frac{{}_{75}C_{24} - {}_{25}C_{24}}{{}_{75}C_{24} - {}_{24}C_{24}} \times \frac{{}_{75}C_{24} - {}_{26}C_{24}}{{}_{75}C_{24} - {}_{25}C_{24}} \times \cdots \times \frac{{}_{k}C_{24} - {}_{(k-1)}C_{24}}{{}_{75}C_{24} - {}_{(k-1)}C_{24}} \\ &= \frac{{}_{k}C_{24} - {}_{(k-1)}C_{24}}{{}_{75}C_{24}} \quad \text{se } 24 \leq k \leq 75 \ \text{ e } 0, \text{ no caso contrário .} \end{aligned}$$

onde tomamos ${}_{n}C_{p} = 0$ para $n < p$.

A função de distribuição acumulada F_X pode ser expressada como:

$$F_X(x) = \begin{cases} 0 & \text{Se} \quad x < 24 \\ \displaystyle\sum_{k=24}^{[x]} \frac{{}_{k}C_{24} -_{k-1} C_{24}}{{}_{75}C_{24}} & \text{Se} \quad 24 \leq x \leq 75 \\ 1 & \text{Se} \quad x > 75 \end{cases} \tag{1}$$

Para o caso em estudo, F_X pode ser resumida como na Tabela 1.

Como X assume apenas valores inteiros positivos, por praticidade e de acordo com CHUNG, 2001 será usado para cálculo do valor esperado a expressão:

$$E(X) = \sum_{n=1}^{\infty} P(X \geq n) \tag{2}$$

Tabela 1: Função de distribuição acumulada do número de bolas chamadas até a batida.

$F_X(x)$	$P(X \leq x)$	$< 0{,}01$	Se $x < 63$
$F_X(x)$	$P(X \leq x)$	$\approx 0{,}01$	Se $63 \leq x < 65$
$F_X(x)$	$P(X \leq x)$	$\approx 0{,}02$	Se $65 \leq x < 67$
$F_X(x)$	$P(X \leq x)$	$\approx 0{,}04$	Se $67 \leq x < 68$
$F_X(x)$	$P(X \leq x)$	$\approx 0{,}06$	Se $68 \leq x < 69$
$F_X(x)$	$P(X \leq x)$	$\approx 0{,}09$	Se $69 \leq x < 70$
$F_X(x)$	$P(X \leq x)$	$\approx 0{,}14$	Se $70 \leq x < 71$
$F_X(x)$	$P(X \leq x)$	$\approx 0{,}21$	Se $71 \leq x < 72$
$F_X(x)$	$P(X \leq x)$	$\approx 0{,}31$	Se $72 \leq x < 73$
$F_X(x)$	$P(X \leq x)$	$\approx 0{,}46$	Se $73 \leq x < 74$
$F_X(x)$	$P(X \leq x)$	$\approx 0{,}68$	Se $74 \leq x < 75$
$F_X(x)$	$P(X \leq x)$	1,00	Se $x \geq 75$

Utilizando o *software* R e a equação (2) obtem-se o valor esperado para X, $E(X) = 72,96$. Uma simples consulta à função F_X permite concluir que o valor mais provável de X é: $\text{Moda}_X = 75$ e que o valor mediano é: $\text{Mediana}_X \approx 73,5$.

Com a finalidade de testar o modelo apresentado, duzentos bingos foram simulados, nos quais apenas uma cartela esteve em jogo. Os resultados obtidos estão resumidos na Tabela 2, a seguir:

Tabela 2: Medidas de localização para as 200 simulações realizadas de sorteios com uma única cartela.

Mínimo	Q_1	Mediana	Média	Q_3	Máximo
63	71	74	72,53	75	75

A distribuição de frequências de bolas chamadas até a batida nas simulações é dada na Tabela 3, abaixo.

Tabela 3: Frequências do número de bolas chamadas até a batida nas simulações realizadas envolvendo uma cartela em jogo.

Bolas chamadas até a batida	63	64	65	66	67	68	69	70	71	72	73	74	75
Ocorrências	1	1	2	5	6	8	10	9	13	19	24	41	61

Da Tabela 3, pode-se concluir que a moda é igual a 75, concordando com o valor encontrado teoricamente.

Como bingo com apenas uma cartela não é o mais comum, também foram simulados jogos para os seguintes números de cartelas: $n = 200$, $n = 1000$, $n = 2000$, $n = 5000$ e $n = 10000$. Na Tabela 4, os resumos dessas simulações são apresentados.

Em cada uma das simulações realizadas, foram geradas aleatoriamente n cartelas, obedecendo aos mesmos critérios de regularidade impostos na produção das cartelas. Este procedimento foi repetido duzentas vezes, simulando igual número de repetições do jogo nos mesmos moldes. É apresentado um resumo dos resultados nos histogramas dados na Figura 2 a seguir:

Tabela 4: Resumo dos resultados obtidos em 200 simulações contendo n cartelas em jogo.

n	Mínimo	Q_1	Mediana	Moda	Média	Q_3	Máximo
200	54	61	62	64	62,19	64	67
1000	50	58	59	59	59,03	61	63
2000	50	56	58	58	57,31	59	63
5000	46	54	56	57	55,86	58	61
10000	49	54	55	55	54,94	56	59

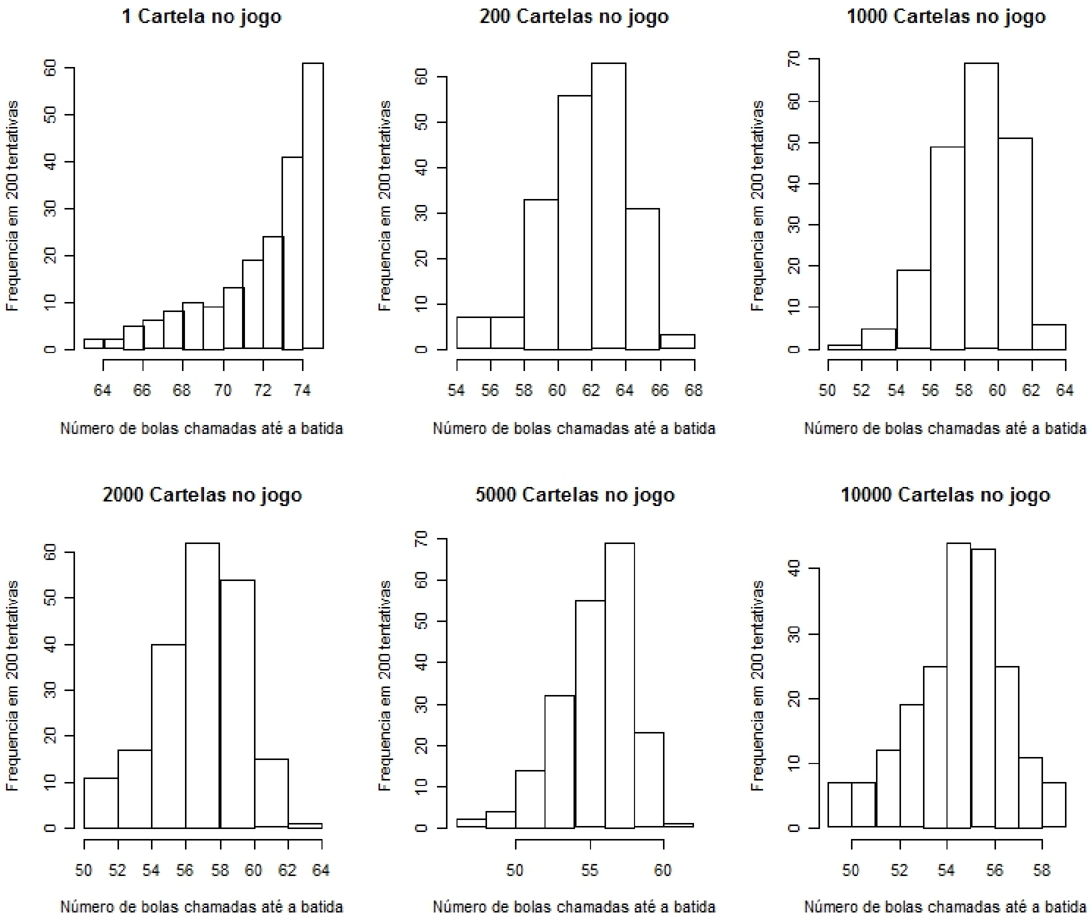

Figura 2: Histogramas das distribuições empíricas.

Para encontrar a distribuição de probabilidades da vairiável aleatória: $X \equiv$ "número de bolas chamadas até que alguém bata num bingo com n cartelas, "foram consideradas as n variáveis aleatórias X_i com $i = 1, \ldots n$, onde $X_i \equiv$ número de bolas que precisam ser chamadas até que a cartela i seja completamente preenchida". Assim, o interesse é estudar a distribuição de probabilidades da variável aleatória $X = \text{mínimo}(X_1, X_2, \ldots, X_n)$. Supondo independência entre quaisquer variáveis X_i e X_j, $i \neq j = 1, \ldots n$, de maneira a simplificar o problema, pode-se testar, para cada n, a hipótese de que a distribuição de X (ver MAGALHÃES, 2006) é dada por:

$$H_0 : F_X(x) = 1 - [1 - F_{X_1}(x)]^n, \ \ \forall x \tag{3}$$

Nas Tabelas 5 e 6, dadas a seguir, estão as distribuições empíricas F_0 e as teóricas F_X para os valores de n anteriormente mencionados.

Tabela 5: Distribuições empíricas e teóricas para n igual a 1, 200 e 1000 cartelas

	$n=1$			$n=200$			$n=1000$		
i	x_i	$F_X(x_i)$	$F_0(x_i)$	x_i	$F_X(x_i)$	$F_0(x_i)$	x_i	$F_X(x_i)$	$F_0(x_i)$
1	63	0,005	0,006	54	0,005	0,011	50	0,005	0,005
2	64	0,010	0,010	55	0,015	0,019	51	0,005	0,009
3	65	0,020	0,015	56	0,035	0,033	52	0,005	0,016
4	66	0,045	0,024	57	0,045	0,057	53	0,010	0,030
5	67	0,075	0,038	58	0,070	0,095	54	0,030	0,053
6	68	0,115	0,058	59	0,125	0,155	55	0,080	0,092
7	69	0,165	0,089	60	0,235	0,244	56	0,125	0,155
8	70	0,210	0,136	61	0,360	0,370	57	0,235	0,253
9	71	0,275	0,206	62	0,515	0,529	58	0,370	0,392
10	72	0,370	0,308	63	0,655	0,705	59	0,545	0,568
11	73	0,490	0,459	64	0,830	0,858	60	0,715	0,753
12	74	0,695	0,680	65	0,935	0,955	51	0,875	0,901
13	75	1,000	1,000	66	0,985	0,993	62	0,970	0,977
14				67	1,000	1,000	63	1,000	0,998
15							64	1,000	1,000
	$d_0 = 0,076$			$d_0 = 0,050$			$d_0 = 0,038$		

Tabela 6: Distribuições empíricas e teóricas para diversos números de cartelas

	$n=2000$			$n=5000$			$n=10000$		
i	x_i	$F_X(x_i)$	$F_0(x_i)$	x_i	$F_X(x_i)$	$F_0(x_i)$	x_i	$F_X(x_i)$	$F_0(x_i)$
1	50	0,005	0,009	46	0,005	0,002	49	0,020	0,024
2	51	0,025	0,018	47	0,010	0,003	50	0,035	0,046
3	52	0,055	0,033	48	0,010	0,006	51	0,070	0,085
4	53	0,085	0,059	49	0,015	0,012	52	0,130	0,152
5	54	0,140	0,103	50	0,030	0,023	53	0,225	0,261
6	55	0,210	0,176	51	0,050	0,044	54	0,350	0,420
7	56	0,340	0,287	52	0,100	0,079	55	0,570	0,619
8	57	0,470	0,442	53	0,160	0,140	56	0,785	0,815
9	58	0,650	0,631	54	0,260	0,238	57	0,910	0,946
10	59	0,810	0,813	55	0,385	0,383	58	0,965	0,993
11	60	0,920	0,939	56	0,535	0,570	59	1,000	1,000
12	61	0,985	0,990	57	0,730	0,768			
13	62	0,995	0,999	58	0,880	0,917			
14	63	1,000	1,000	59	0,970	0,985			
15				60	0,995	0,999			
16				61	1,000	1,000			
	$d_0 = 0,053$			$d_0 = 0,038$			$d_0 = 0,070$		

Ao realizar testes de Kolmogorov-Smirnov com $n = 200$, já que em cada caso o jogo foi repetido esse número de vezes. Com $\alpha = 5\%$, de acordo com Bussab e Morettin (2006), é necessário observar que em todos os casos, tem-se $d_0 < D_c$. Onde:

$$D_c = \frac{\sqrt{-ln\left(\frac{0,05}{2}\right)}}{\sqrt{2 \times 200}} \approx 0,096$$

e d_0, é a maior diferença, em valor absoluto, entre $F_X(x_i)$ e $F_0(x_i)$. Os valores omissos tiveram frequencias nulas.

Em palavras, isto quer dizer que os dados não refutam as hipóteses nulas, ou seja, que o mínimo modela bem o problema.

Tabela 7: Diferenças entre as principais medidas de tendência central, empíricas e teóricas.

Número de	**Média**			**Moda**			**Mediana**		
cartelas	Teórica	Empírica	d_0	Teórica	Empírica	d_0	Teórica	Empírica	d_0
$n = 1$	72,96	72,53	0,43	75	75	0	73,0	74	-1,0
$n = 200$	61,96	62,19	-0,23	63	64	-1	61,5	62	-0,5
$n = 1000$	58,79	59,03	-0,24	60	59	1	58,5	59	-0,5
$n = 2000$	57,49	57,31	0,18	58	58	0	57,5	58	-0,5
$n = 5000$	55,83	55,86	-0,03	57	57	0	55,5	56	-0,5
$n = 10000$	54,61	54,94	-0,33	56	55	1	54,5	55	-0,5

Observa-se que em todos os casos, veja a tabela 7, as diferenças absolutas são bem pequenas, implicando numa diferença relativa inferior a 0,5%. Este fato, reforça a ideia do bom ajuste do modelo proposto.

Conclusões

O modelo desenvolvido e testado, assim como os resultados obtidos pelas simulações, mostram que a medida que o número de cartelas cresce, diminuem os valores das medidas de posição, tendência central e separatrizes, veja Tabela 4. Os testes realizados com os dados simulados, reforçam a ideia de que o modelo está coerente.

Considerando a hipótese de adequação do modelo, a produção e distribuição de todas cartelas, mesmo aquelas que não obedecem as leis de produção, ou seja, todas as combinações possíveis de 24 números tomados dentre 75 possíveis, ao chamar a vigésima quarta bola, alguém deveria bater, ou seja $F_X(x)$ seria igual a zero, para $x < 24$ e 1, para $x \geq 24$. A partir dos cálculos usando o modelo proposto, obtêm-se que se $X = \min(X_1, \ldots, X_{_{75}C_{24}})$, então F_X é igual a zero, para $x < 27$ e 1, para o caso contrário, discordando um pouco do que deveria ocorrer. Este problema foi atribuído em parte a deficiência da máquina utilizada para efetuar os cálculos e, principalmente, devido à suposição de independência entre as variáveis que representam o número de sorteios até alguém vencer o jogo. De fato, cada cartela em jogo concorre com as demais na busca pelo seu total preenchimento. Em outros termos, a vitória de um jogador muito provavelmente implicá na derrota de oponentes, refletindo a dependência entre as variáveis sob estudo. Isto ilustra o potencial do modelo desenvolvido para o estudo de situações envolvendo riscos competitivos Langseth e Lindqvist (2005).

Excluindo casos extremos, como o discutido no parágrafo anterior, o modelo descreve bem a natureza do jogo e pode ser usando para se ter uma ideia do tempo esperado até alguém bater, usando o número esperado de bolas a chamar até que alguém bata, de acordo com o número de cartelas que serão distribuídas.

Nos trabalhos em andamento, os autores tem investigado aplicações para o modelo na área de riscos competitivos e teoria das filas.

Referências:

BUSSAB, W. O.; MORETTIN, P. A. *Estatística Básica* 5ª ed. São Paulo: Saraiva, 2006.

CHUNG, K. L. *A course in probability theory.* Third edition. San Diego: Standford University Academic Press, 2001.

LANGSETH, H.; LINDQVIST, B. *Competing Risks for Repairable Systems: A data study.* Journal of Planning and Inference, vol 136(5), 1687-1700, 2005.

MAGALHÃES, M. N. *Probabilidade e Variável Aleatória*, 2ª ed. São Paulo: EDUSP, 2006.

R CORE TEAM. *R*: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2012. ISBN 3-900051-07-0, URL http://www.R-project.org/.