Universidade de São Paulo
Escola de Artes, Ciências e Humanidades

ACH2053 – Introdução à Estatística – 1º sem. 2025
Professor: José Ricardo G. Mendonça

1ª Prova — Data: 30 abr. 2025 — 19h00 às 20h45 — Gabarito

We are an impossibility in an impossible universe.
Ray Bradbury (1920–2012)

Na resolução dos problemas, explique seu raciocínio e o que você está fazendo de forma que eu possa acompanhá-lo(a). Soluções "mágicas" ou "geniais" não serão aceitas sem explicações.

Problemas

1. Considere o experimento aleatório de lançar uma moeda honesta repetidamente e contar o número de lançamentos efetuados até surgir uma cara.

   (a) Encontre a probabilidade de que a primeira cara apareça no k-ésimo lançamento, $k \geqslant 1$, e verifique que $P(\Omega) = 1$;

   O espaço amostral Ω é formado pelos eventos K, CK, CCK etc., onde C representa coroa e K representa cara. A probabilidade de que a primeira cara apareça no k-ésimo lançamento, $k \geqslant 1$, corresponde à probabilidade do evento $C\cdots CK = C^{k-1}K$ e vale $\mathbb{P}(C^{k-1}K) = (\frac{1}{2})^{k-1} \cdot \frac{1}{2} = (\frac{1}{2})^k$. A probabilidade total vale, portanto,

   $$P(\Omega) = \sum_{k=1}^{\infty} \mathbb{P}(C^{k-1}K) = \sum_{k=1}^{\infty} (\tfrac{1}{2})^k = \tfrac{1}{2} \sum_{k=0}^{\infty} (\tfrac{1}{2})^k = \tfrac{1}{2} \cdot \frac{1}{1-\frac{1}{2}} = 1,$$

   onde reparamos que a soma infinita $S = 1 + q + q^2 + \cdots = \dfrac{1}{1-q}$ quando $|q| < 1$.(*)

   (b) Encontre a probabilidade de que a primeira cara apareça em um lançamento ímpar.

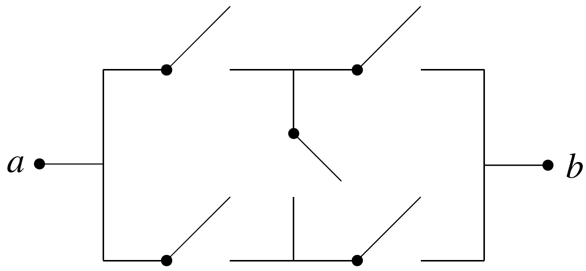
   Estamos interessados nos eventos K, CCK, $CCCCK$ etc. A soma das probabilidades desses eventos vale

   $$\mathbb{P}(K \text{ no lançamento ímpar}) = \sum_{k=0}^{\infty} \mathbb{P}(C^{2k}K) = \tfrac{1}{2} \sum_{k=0}^{\infty} (\tfrac{1}{2})^{2k} = \tfrac{1}{2} \sum_{k=0}^{\infty} (\tfrac{1}{4})^k = \frac{2}{3}.$$

   Isso nos mostra que é duas vezes mais provável observar a primeira cara num lançamento ímpar (K ou CCK ou $CCCCK$ etc.) do que num lançamento par, cuja probabilidade vale $\mathbb{P}(K \text{ no lançamento par}) = 1 - \mathbb{P}(K \text{ no lançamento ímpar}) = \frac{1}{3}$!

(*) Podemos deduzir este resultado elementar reparando que $qS = q + q^2 + q^3 + \cdots$, de forma que $S - qS = 1$.

2. Considere a malha de interruptores abaixo, conhecida como ponte de Wheatstone. Cada interruptor da malha pode abrir aleatoriamente com probabilidade $1-p$, igual e independentemente. Encontre a probabilidade de que haja um caminho fechado entre os terminais a e b da malha.



Vamos denominar os dois interruptores de cima do circuito de A e B, o interruptor central de C e os dois interruptores de baixo de D e E. Teremos um circuito fechado entre os pontos a e b, que denotamos por $a \leftrightarrow b$, se qualquer um dos seguintes eventos acontecerem (onde indicamos o interruptor fechado pela sua letra): $X_1 = A \cap B$, $X_2 = D \cap E$, $X_3 = A \cap C \cap E$ e $X_4 = D \cap C \cap B$. Assim, temos $\mathbb{P}(a \leftrightarrow b) = \mathbb{P}(X_1 \cup X_2 \cup X_3 \cup X_4)$. Empregando o princípio da inclusão-exclusão e a independência dos eventos A, B, C, D e E encontramos

$$\mathbb{P}(a \leftrightarrow b) = \sum_i \mathbb{P}(X_i) - \sum_{i<j} \mathbb{P}(X_i \cap X_j) + \sum_{i<j<k} \mathbb{P}(X_i \cap X_j \cap X_k) - \mathbb{P}(X_1 \cap X_2 \cap X_3 \cap X_4),$$

onde os índices variam de 1 até 4. Como $\mathbb{P}(X_1) = \mathbb{P}(A \cap B) = p^2$ e similarmente para os outros termos, encontramos

$$\begin{aligned}
\sum_i \mathbb{P}(X_i) &= p^2 + p^2 + p^3 + p^3; \\
\sum_{i<j} \mathbb{P}(X_i \cap X_j) &= p^4 + p^4 + p^4 + p^4 + p^4 + p^5; \\
\sum_{i<j<k} \mathbb{P}(X_i \cap X_j \cap X_k) &= p^5 + p^5 + p^5 + p^5; \\
\mathbb{P}(X_1 \cap X_2 \cap X_3 \cap X_4) &= p^5.
\end{aligned}$$

Assim, a probabilidade de que haja um caminho fechado entre os terminais a e b da malha vale $\mathbb{P}(a \leftrightarrow b) = 2p^2 + 2p^3 - 5p^4 + 2p^5$.

3. O exame de Papanicolaou é um procedimento usado para detectar alterações nas células do colo do útero que podem indicar a presença de um câncer. Para mulheres que possuem esse câncer, existe 16% de probabilidade do exame resultar em falso negativo, enquanto para mulheres que não possuem o câncer existe 19% de probabilidade do exame resultar

em falso positivo. No Brasil, em 2023 a incidência dessa forma de câncer é de 13,25 casos a cada 100 mil mulheres.[†] Qual é a probabilidade de que uma mulher seja portadora de câncer do colo do útero caso seu exame de Papanicolaou tenha dado resultado positivo?

Segundo os dados do problema, a probabilidade do exame de Papanicolaou resultar em falso negativo vale $P(-|C) = 16\%$, donde $P(+|C) = 84\%$, enquanto a probabilidade do exame resultar em falso positivo vale $P(+|\overline{C}) = 19\%$, donde $P(-|\overline{C}) = 81\%$. Queremos calcular a probabilidade $P(C|+)$ de que uma mulher seja portadora de câncer do colo do útero caso seu exame de Papanicolaou tenha dado resultado positivo. Pela identidade de Bayes temos que

$$P(C|+)P(+) = P(+|C)P(C) \quad \Rightarrow \quad P(C|+) = \frac{P(+|C)P(C)}{P(+)}.$$

Dados epidemiológicos indicam que, no Brasil, $P(C) = 13{,}25/100.000 = 0{,}01325\%$, e podemos calcular a probabilidade $P(+)$ em termos da probabilidade total

$$P(+) = P(+|C)P(C) + P(+|\overline{C})P(\overline{C}) = 84\% \cdot 0{,}01325\% + 19\% \cdot 99{,}98675\%$$

Dessa forma, a probabilidade procurada vale

$$P(C|+) = \frac{P(+|C)P(C)}{P(+)} = \frac{84\% \cdot 0{,}01325\%}{84\% \cdot 0{,}01325\% + 19\% \cdot 99{,}98675\%} \simeq 0{,}000586$$

Esse número significa que de cada um milhão de resultados positivos no exame de Papanicolaou, apenas 586 indicam real presença de câncer de colo do útero, o que lança dúvidas sobre a efetividade do teste. De nossa análise fica claro que a razão de sua ineficácia reside na relativa baixa incidência da doença (0,01325% da população feminina) e das elevadas taxas de falsos negativos (16%) e falsos positivos (19%) do teste. Para melhorar a confiabilidade do teste deve-se procurar reduzir sua probabilidade de resultados falso positivos, pois $P(C|+) \to 1$ conforme $P(+|\overline{C}) \to 0$.

4. A probabilidade de determinada fruta apresentar defeitos que a tornam imprópria para produção de sucos é de 15%. Calcule a probabilidade de que numa amostra de 40 frutas escolhidas ao acaso haja 4 ou mais frutas com defeitos usando a aproximação normal com a correção de continuidade.

   Neste problema usamos os subscritos B e N para binomial e normal, respectivamente. A probabilidade de encontrar k defeitos em n frutas é dada pela distribuição binomial $\mathrm{Bin}(n,p)$, com $\mathbb{P}_B(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 1, \ldots, n$. Assim, a probabilidade de que numa amostra de n frutas escolhidas ao acaso haja 4 ou mais frutas com defeitos vale

$$\mathbb{P}_B(X \geqslant 4) = 1 - \mathbb{P}_B(X < 4) = 1 - \left[\mathbb{P}_B(X=0) + \mathbb{P}_B(X=1) + \mathbb{P}_B(X=2) + \mathbb{P}_B(X=3)\right].$$

[†]Fonte: M. de O. Santos, F. C. da S. de Lima, L. F. L. Martins, J. F. P. Oliveira, L. M. de Almeida, M. de C. Cancela, "Estimativa de incidência de câncer no Brasil, 2023–2025", **Rev. Bras. Cancerol. [Internet]**, vol. 69, n. 1, art. e-213700 (2023). Disponível em: DOI: 10.32635/2176-9745.RBC.2023v69n1.3700.

Com $n = 40$ e $p = 0{,}15$ encontramos $\mathbb{P}_B(X \geqslant 4) \simeq 1 - 0{,}1302 \simeq 87{,}0\%$.(‡) O problema, no entanto, pede para calcularmos essa probabilidade usando a aproximação normal à distribuição binomial. Sabemos que $\mathrm{Bin}(n,p) \sim N(\mu,\sigma^2)$ com $\mu = np$ e $\sigma^2 = np(1-p)$ quando $\min(np, n(1-p)) \gtrsim 5$. Aplicando a correção de continuidade à aproximação normal encontramos

$$\mathbb{P}_B(X \geqslant 4) \simeq \mathbb{P}_N(X \geqslant 3{,}5) = 1 - \mathbb{P}_N(X < 3{,}5).$$

Transformando essa última probabilidade em uma probabilidade envolvendo a distribuição normal padrão encontramos, com $n = 40$ e $p = 0{,}15$,

$$\mathbb{P}_N(X < 3{,}5) = \mathbb{P}_N\Big(Z < \frac{3{,}5 - np}{\sqrt{np(1-p)}}\Big) = \mathbb{P}_N\Big(Z < -\frac{2{,}5}{\sqrt{5{,}1}}\Big) \simeq \mathbb{P}_N(Z < -1{,}107),$$

de forma que

$$\mathbb{P}_B(X \geqslant 4) \simeq \mathbb{P}_N(X \geqslant 3{,}5) = 1 - \mathbb{P}_N(Z < -1{,}107) = \mathbb{P}_N(Z < 1{,}107),$$

onde na última igualdade usamos a identidade $\mathbb{P}_N(Z < -z) = 1 - \mathbb{P}_N(Z < z)$. Consultando uma "tabela z" da distribuição normal padrão encontramos que $\mathbb{P}_N(Z < 1{,}107) \simeq 0{,}8665 \simeq 86{,}7\%$, um valor bastante próximo do valor obtido a partir da distribuição binomial.(§)

5. Uma v.a. $X \sim \mathrm{Pareto}(\alpha,\beta)$ tem distribuição de Pareto de parâmetros α, $\beta > 0$ se sua pdf é dada por $f(x;\alpha,\beta) = 0$ se $x < \beta$ e $f(x;\alpha,\beta) = Cx^{-\alpha-1}$ para $x \geqslant \beta$, onde $C = C(\beta) > 0$ é uma constante de normalização. Essa distribuição é muito empregada em economia, finanças e atuária em geral. Dada uma v.a. $X \sim \mathrm{Pareto}(\alpha,\beta)$, calcule:

   (a) A cdf $F(x;\alpha,\beta)$ para todo $x \in \mathbb{R}$;

   Devemos calcular a cdf $F(x;\alpha,\beta) = \mathbb{P}(X \leqslant x)$ quando $X \sim \mathrm{Pareto}(\alpha,\beta)$. Pela definição de $f(x;\alpha,\beta)$ temos que $F(x;\alpha,\beta) = 0$ se $x < \beta$. Se $x \geqslant \beta$, encontramos

   $$F(x;\alpha,\beta) = \int_{-\infty}^{x} f(u;\alpha,\beta)\,\mathrm{d}u = C\int_{\beta}^{x} u^{-\alpha-1}\,\mathrm{d}u = -C\frac{u^{-\alpha}}{\alpha}\bigg|_{\beta}^{x} = \frac{C}{\alpha}(\beta^{-\alpha} - x^{-\alpha}).$$

   O valor de C pode ser obtido de $\lim\limits_{x\to\infty} F(x;\alpha,\beta) = 1$, que fornece $C = \alpha\beta^{\alpha}$. Assim, encontramos que $F(x;\alpha,\beta) = 0$ se $x < \beta$ e $F(x;\alpha,\beta) = 1 - (\beta/x)^{\alpha}$ se $x \geqslant \beta$.

   (b) O valor esperado de X.

   O valor esperado de uma v.a. $X \sim \mathrm{Pareto}(\alpha,\beta)$ é dado por

   $$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x;\alpha,\beta)\,\mathrm{d}x = \alpha\beta^{\alpha}\int_{\beta}^{+\infty} x^{-\alpha}\,\mathrm{d}x = \alpha\beta^{\alpha}\,\frac{x^{-\alpha+1}}{-\alpha+1}\bigg|_{\beta}^{+\infty} = \frac{\alpha\beta}{\alpha-1}.$$

⋆ — ⋆ — ⋆

(‡) O comando `1-pbinom(3,40,0.15)` em R fornece o valor mais preciso $1 - \mathbb{P}_B(X \leqslant 3) = 0{,}8698312$.

(§) O comando `pnorm(2.5/sqrt(5.1))` em R fornece o valor mais preciso $\mathbb{P}(Z < 2{,}5/\sqrt{5{,}1}) = 0{,}8658571$.

Formulário

Axiomas da teoria das probabilidades

Para quaisquer eventos (subconjuntos) A e B de um espaço amostral Ω valem $P(A) \geqslant 0$, $P(\Omega) = 1$ e $P(A \cup B) = P(A) + P(B)$ se $A \cap B = \varnothing$.

Como consequências dos axiomas da teoria das probabilidade valem $P(\bar{A}) = 1 - P(A)$, $P(\varnothing) = 0$ e $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ para quaisquer eventos $A, B \subseteq \Omega$.

Identidades de Bayes

$$P(A|B)P(B) = P(B|A)P(A), \quad P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i),$$

onde $A_i \cap A_j = \varnothing$ se $i \neq j$ e $A_1 \cup A_2 \cup \cdots \cup A_n = \Omega$.

Algumas distribuições de probabilidade

- Binomial $\mathrm{Bin}(n,p)$: $p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0, 1, \ldots, n$;
  $\mathbb{E}(X) = np$, $\mathrm{Var}(X) = np(1-p)$.
- Poisson(λ): $p_X(k) = e^{-\lambda} \dfrac{\lambda^k}{k!}$, $k = 0, 1, \ldots$; $\mathbb{E}(X) = \lambda$, $\mathrm{Var}(X) = \lambda$.
- Uniforme $U(a,b)$: $f_X(x) = \dfrac{1}{b-a}$, $a \leqslant x \leqslant b$; $\mathbb{E}(X) = \frac{1}{2}(a+b)$, $\mathrm{Var}(X) = \frac{1}{12}(b-a)^2$.
  Às vezes, por conveniência o domínio da distribuição uniforme é dado como $a < x < b$.
- Exponencial $\mathrm{Exp}(\beta)$: $f_X(x) = \dfrac{1}{\beta} e^{-x/\beta}$, $x \geqslant 0$; $\mathbb{E}(X) = \beta$, $\mathrm{Var}(X) = \beta^2$.
- Normal $N(\mu, \sigma^2)$: $f_X(x) = \dfrac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\dfrac{(x-\mu)^2}{2\sigma^2}\right]$, $-\infty < x < +\infty$;
  $\mathbb{E}(X) = \mu$, $\mathrm{Var}(X) = \sigma^2$.

Transformação entre distribuições normais

Se $F_X(x)$ é a cdf de uma v.a. $X \sim N(\mu, \sigma^2)$ e $\Phi(z)$ é a cdf de uma v.a. padrão $Z \sim N(0,1)$, então

$$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right), \quad \Phi(-z) = 1 - \Phi(z).$$

Alguns valores de $\Phi(z) = \mathbb{P}(Z < z) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{z} \exp\left(-\tfrac{1}{2}u^2\right) \mathrm{d}u$ aparecem na tabela abaixo.

z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998

Tabela: Valores da distribuição cumulativa normal padrão $\Phi(z) = \mathbb{P}(Z < z)$ para $z \geqslant 0$.