# Lista 1

**Questão 1: Filhos e filhas**

Considere um casal que tem dois descendentes e que as chances de cada um deles ser filho ou filha são iguais. Responda às perguntas abaixo:

1. Calcule a probabilidade dos descendentes formar um casal (ou seja, um filho e uma filha).

Posso definir o espaço amostral como $S = \{HH, HM, MM, MH\}$.

A probabilidade pedida é, então:

$$P(HM \cup MH) = P(HM) + P(MH) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

2. Calcule a probabilidade de ao menos um dos descendentes ser filho.

Resolvendo pelo complemento de duas meninas:

$$1 - P(MM) = \frac{3}{4}$$

3. Calcule a probabilidade das duas serem filhas dado que uma é filha (cuidado!).

Usando a fórmula $P(A|B) = \frac{P(A \cap B)}{P(B)}$:

$$\begin{aligned} P(MM|HM \cup MH \cup MM) &= \frac{P(MM \cap (HM \cup MH \cup MM))}{P(HM \cup MH \cup MM)} \\ &= \frac{P(MM)}{3/4} \\ &= \frac{1/4}{3/4} \\ &= \frac{1}{3} \end{aligned}$$

4. Calcule a probabilidade dos descendentes nascerem no mesmo dia (assuma que a chance de nascer em um determinado dia é igual a qualquer outro)

Dependendo das simplificações que podemos fazer, essa questão pode ser respondida de várias formas.

Se presumirmos que o primeiro descendente nasce em um ano e o próximo nasce em um outro ano, o primeiro pode nascer em qualquer dia, e o segundo irá nascer nesse mesmo dia com probabilidade $\frac{1}{365}$.

Isso desconsiderando tempos de gravidez, chance de gêmeos, etc. Acho que esse era o propósito da questão.

---

**Questão 2: Dado**

Considere um icosaedro (um sólido Platônico de 20 faces) tal que a chance de sair a face $i = 1, \ldots, 20$ seja linearmente proporcional a i. Ou seja, $P[X = i] = ci$ para alguma constante c, onde X é uma variável aleatória que denota a face do dado. Responda às perguntas abaixo:

1. Determine o valor de c.

$$\begin{aligned} \sum_{i=1}^{20} P(X = i) &= 1 \\ \sum_{i=1}^{20} ci &= 1 \\ c\sum_{i=1}^{20} i &= 1 \\ c\frac{(1+20)20}{2} &= 1 \\ 210c &= 1 \\ c &= \frac{1}{210} \end{aligned}$$

2. Calcule o valor esperado de X (obtenha também o valor numérico)

$$\begin{aligned} E[X] &= \sum_{i=1}^{20} iP(X = i) \\ &= \sum_{i=1}^{20} i \cdot ci \\ &= c\sum_{i=1}^{20} i^2 \end{aligned}$$

Para calcular o somatório, posso usar a fórmula $\frac{n(n+1)(2n+1)}{6}$, que é igual à soma dos quadrados dos primeiros $n$ números naturais.

$$E[X] = \frac{1}{210} \cdot 2870 = \frac{41}{3}$$

3. Calcule a probabilidade de X ser maior do que seu valor esperado.

$$P\left(X > \frac{41}{3}\right) = P(X > 13.666)$$

Como esse problema trata de números naturais até 20:

$$\begin{aligned}
P(X \geq 14) &= \sum_{i=14}^{20} P(X = i) \\
&= c\sum_{i=14}^{20} i \\
&= \frac{1}{210} \cdot \frac{7(14+20)}{2} \\
\\
&= \frac{17}{30}
\end{aligned}$$

4. Calcule a variância de X (obtenha também o valor numérico).

$$\begin{aligned}
\mathrm{Var}(X) &= E[X^2] - E[X]^2 \\
&= \left(\sum_{i=1}^{20} i^2 P(X = i)\right) - E[X]^2 \\
&= c\left(\sum_{i=1}^{20} i^3\right) - \left(\frac{41}{3}\right)^2 \\
&= \frac{1}{210}\left(\frac{20(20+1)}{2}\right)^2 - \frac{1681}{9} \\
&= 210 - \frac{1681}{9} \\
&= 23.222\ldots
\end{aligned}$$

Usei a esperança de X e o $c$ calculados anteriormente, além da fórmula da soma dos primeiros 20 cubos.

5. Repita os últimos três itens para o caso do dado ser uniforme, ou seja, $P[X = i] = 1/20$, $i =$ 1,..., 20. Qual dado possui maior variância? Explique intuitivamente sua descoberta.

2.

$$
\begin{aligned}
E[X] &= \sum_{i=1}^{20} iP(X=i) \\
&= \frac{1}{20}\sum_{i=1}^{20} i \\
&= \frac{1}{20}210 \\
&= \frac{21}{2}
\end{aligned}
$$

3. $P\left(X > \frac{21}{2}\right) = P(X \geq 11)$

$$
\begin{aligned}
P(X \geq 11) &= \frac{1}{20}\sum_{i=11}^{20} 1 \\
&= \frac{1}{20}\cdot 10 \\
&= \frac{1}{2}
\end{aligned}
$$

4.

$$
\begin{aligned}
\mathrm{Var}(X) &= E[X^2] - E[X]^2 \\
&= \left(\sum_{i=1}^{20} i^2 P(X=i)\right) - E[X]^2 \\
&= \left(\frac{1}{20}\sum_{i=1}^{20} i^2\right) - \left(\frac{21}{2}\right)^2 \\
&= \frac{2870}{20} - \frac{441}{4} \\
&= 33.25
\end{aligned}
$$

O dado uniforme possui maior variância. Isso é intuitivo porque o dado não uniforme introduz um viés ao resultado. Nele, a probabilidade de um valor $i$ ser o escolhido é proporcional à $i$, o que faz com que as faces maiores seja mais prováveis, reduzindo assim um pouco da variância dos resultados.

---

**Questão 3: Dado em ação**

Considere a versão uniforme do dado acima, ou seja, $P[X=i] = 1/20\; i = 1, \ldots, 20$ Seja Y uma variável aleatória indicadora da primalidade da face do dado. Ou seja, $Y = 1$

quando o X é um número primo, e $Y = 0$ caso contrário. Responda às perguntas abaixo:

1. Determine $P[Y = 1]$.

Basta contar o número de primos entre 1 e 20 e dividir por 20: 2, 3, 5, 7, 11, 13, 17, 19.

8 números primos. Então $P(Y = 1) = \frac{8}{20} = \frac{2}{5}$.

2. Considere que o dado será jogado n vezes. Seja $Y_i$ a indicadora da primalidade da i-ésima rodada, para $i = 1, \ldots, n$ e defina $Z = \sum_{i=1}^{n} Y_i$ Repare que Z é uma variável aleatória que denota o número de vezes que o resultado é primo. Determine a distribuição de Z, ou seja, $P[Z = k]$, para $k = 0, \ldots, n$. Que distribuição é esta?

$Z$ é uma soma de Bernoullis, que, por definição, é a distribuição Binomial. Mais especificamente, $Z \sim Bin\left(n, \frac{2}{5}\right)$. Sua PMF é:

$$P(Z = k) = \binom{n}{k}\left(\frac{2}{5}\right)^k\left(\frac{3}{5}\right)^{n-k}$$

3. Considere que o dado será jogado até que um número primo seja obtido. Seja $Y_i$ a indicadora da primalidade da i-ésima rodada, para $i = 1, \ldots$ e defina $Z = min\{i | Y_i = 1\}$. Repare que Z denota o número de vezes que o dado é jogado até que o resultado seja um número primo. Determine a distribuição de Z, ou seja $P[Z = k]$, para $k = 1, \ldots$ Que distribuição é esta?

Essa é definição da distribuição Geométrica. $Z \sim Geom\left(\frac{2}{5}\right)$. PMF:

$$P(Z = k) = \left(\frac{3}{5}\right)^{k-1}\frac{2}{5}$$

---

**Questão 4: Cobra**

Considere três imagens tiradas em uma floresta, $I_1$, $I_2$, e $I_3$ Em apenas uma das imagens existe uma pequena cobra. Um algoritmo de detecção de cobras em imagens detecta a cobra na imagem i com probabilidade $\alpha_i$. Suponha que o algoritmo não encontrou a cobra na imagem $I_1$. Defina o espaço amostral e os eventos apropriados e use regra de Bayes para determinar:

1. A probabilidade da cobra estar na imagem $I_1$.
2. A probabilidade da cobra estar na imagem $I_2$.

Digamos que a função $Y_j$ é a indicadora de detecção de cobra, e que $C_j$ indica que a cobra está na imagem $j$.

Vou definir a distribuição de $Y$ dado $C$ como:

$$P(Y_j = 1|C_j = 1) = \alpha_j$$
$$P(Y_j = 0|C_j = 1) = 1 - \alpha_j$$

Vou assumir que $P(Y_j = 1|C_j = 0) = 1$, isto é, o algoritmo não pode detectar uma cobra se ela não estiver presente na imagem.

Probabilidade da cobra estar na imagem $I_1$:

$$P(C_1 = 1|Y_1 = 0) = \frac{P(Y_1 = 0|C_1 = 1)P(C_1 = 1)}{P(Y_1 = 0)}$$

O primeiro termo já está disponível de acordo com minha definição de $Y$. O denominador pode ser encontrado com a lei da probabilidade total. Porém, não temos a *prior* $P(C_1 = 1)$, que é a probabilidade inerente de a cobra estar na imagem 1. Como a cobra definitivamente está em uma das imagens, vou assumir uniformidade: $P(C_j = 1) = \frac{1}{3}$.

$$P(Y_1 = 0|C_1 = 1) = 1 - \alpha_1$$

$$\begin{aligned} P(Y_1 = 0) &= P(Y_1 = 0|C_1 = 0)P(C_1 = 0) + P(Y_1 = 0|C_1 = 1)P(C_1 = 1) \\ &= 1 \cdot P(C_2 = 1 \cup C_3 = 1) + P(Y_1 = 0|C_1 = 1)P(C_1 = 1) \\ &= \frac{2}{3} + (1 - \alpha_1)\frac{1}{3} \end{aligned}$$

Então:

$$\begin{aligned} P(C_1 = 1|Y_1 = 0) &= \frac{(1 - \alpha_1)\frac{1}{3}}{\frac{2}{3} + (1 - \alpha_1)\frac{1}{3}} \\ &= \frac{\frac{1}{3} - \frac{\alpha_1}{3}}{\frac{2}{3} + \frac{1}{3} - \frac{\alpha_1}{3}} \\ &= \frac{1 - \alpha_1}{3 - \alpha_1} \end{aligned}$$

Probabilidade da cobra estar na imagem $I_2$:

Da mesma forma, e reutilizando o resultado de $P(Y_1 = 0)$:

$$\begin{aligned} P(C_2 = 1|Y_1 = 0) &= \frac{P(Y_1 = 0|C_2 = 1)P(C_2 = 1)}{P(Y_1 = 0)} \\ &= \frac{1 \cdot \frac{1}{3}}{\frac{1}{3}(3 - \alpha_1)} \\ &= \frac{1}{3 - \alpha_1} \end{aligned}$$

---

**Questão 5: Sem memória**

Seja $X \sim Geo(p)$ uma variável aleatória Geométrica com parâmetro p. Mostre que a distribuição geométrica não tem memória. Ou seja dado que $X > k$, o número de rodadas adicionais até que o evento de interesse ocorra possui a mesma distribuição (dica: formalize esta afirmação).

Formalizando a afirmação:

$$P(X > k + d | X > d) = P(X > k)$$

Vou tentar chegar na segunda expressão a partir da primeira.

$$\begin{aligned} P(X > k + d | X > d) &= \frac{P(X > k + d \cap X > d)}{P(X > d)} \\ &= \frac{P(X > k + d)}{P(X > d)} \\ &= \frac{1 - P(X \leq k + d)}{1 - P(X \leq d)} \\ &= \frac{1 - (1 - (1 - p)^{k+d})}{1 - (1 - (1 - p)^{d})} \\ &= \frac{(1 - p)^{k+d}}{(1 - p)^{d}} \\ &= (1 - p)^{k} \\ &= P(X > k). \end{aligned}$$

---

**Questão 6: Ônibus**

Considere que o processo de chegada do ônibus 485 no ponto do CT seja bem representado por um processo de Poisson. Ou seja, $X \sim Poi(\lambda, t)$ denota o número (aleatório) de ônibus que chegam ao ponto em um intervalo de tempo t com taxa média de chegada igual a $\lambda$. Assuma que $\lambda = 10$ ônibus por hora.

1. Determine a probabilidade de não chegar nenhum ônibus em um intervalo de 30 minutos (inclusive numericamente).

Nesse intervalo de 30 minutos, teremos em média $\lambda \cdot t = 10 \cdot \frac{1}{2} = 5$ ônibus chegando. Definindo a variável aleatória $Y \sim Poi(5)$, podemos calcular a probabilidade desejada.

$$P(Y = k) = \frac{\mu^k e^{-\mu}}{k!}$$

$$\begin{aligned} P(Y=0) &= \frac{5^0 e^{-5}}{0!} \\ &= e^{-5} \approx 0.0067 \end{aligned}$$

2. Determine a probabilidade da média ocorrer, ou seja, de chegarem exatamente 10 ônibus em uma hora (inclusive numericamente).

$Y \sim Poi(10)$:

$$P(Y=10) = \frac{10^{10} e^{-10}}{10!} \approx 0.125$$

3. Determine a taxa $\lambda$ tal que a probabilidade de chegar ao menos um ônibus em um intervalo de 5 minutos seja maior que 90% (inclusive numericamente).

Tenho que encontrar $\lambda$ tal que $1 - P(Y=0) > 0.9$.

$$\begin{aligned} 1 - P(Y=0) &> 0.9 \\ -P(Y=0) &> -0.1 \\ P(Y=0) &< 0.1 \\ \frac{\mu^0 e^{-\mu}}{0!} &< 0.1 \\ e^{-\mu} &< 0.1 \\ -\mu &< \ln 0.1 \\ \mu &> -\ln 0.1 \\ \lambda \cdot \frac{5}{60} &> -\ln 0.1 \\ \lambda &> 12 \cdot (-\ln 0.1) \\ \lambda &> 27.63 \end{aligned}$$

---

**Questão 7: Propriedades**

Seja X e Y duas variáveis aleatórias discretas. Mostre as seguintes equivalências usando as definições:

1. $E[X] = E[E[X|Y]]$, conhecida como regra da torre da esperança.

$$\begin{aligned} E[X] &= \sum_y E[X|Y=y]P(Y=y) \quad \text{(condicionando em } Y=y\text{)} \\ E[E[X]] &= E\left[\sum_y E[X|Y=y]P(Y=y)\right] \text{(aplicando esperança dos dois lados)} \\ E[X] &= E[E[X|Y]] \qquad \text{(definição de esperança condicional)} \end{aligned}$$

2. $Var[X] = E[X^2] - E[X]^2$

Vou começar por $Var(X) = E[X - E[X])^2]$ e tentar chegar no lado direito da equação acima.

$$\begin{aligned} E[X - E[X]]^2 &= E[X^2 - 2XE[X] + E[X]^2] \\ &= E[X^2] - 2E[X]E[X] + E[X]^2 \quad \text{(linearidade)} \\ &= E[X^2] - 2E[X]^2 + E[X]^2 \\ &= E[X^2] - E[X^2] \end{aligned}$$

---

**Questão 8: Paradoxo do Aniversário**

Considere um grupo com n pessoas e assuma que a data de nascimento de cada uma é uniforme dentre os 365 dias do ano. Vamos calcular a chance de duas ou mais pessoas fazerem aniversário no mesmo dia.

1. Seja $c(n)$ a probabilidade de duas ou mais pessoas fazerem aniversário no mesmo dia. Determine explicitamente $c(n)$.

Pode ser mais fácil calcular o inverso da probabilidade não haver uma 'colisão' de aniversários, isto é, a probabilidade de todos nascerem em dias diferentes.

Cada pessoa nasce independentemente em um dos 365 dias do ano. Então, existem $365^n$ combinações diferentes de datas de nascimento. Esse é o denominador.

Para não haver colisão, a primeira pessoa tem 365 opções de dias, a segunda tem 364 (não pode nascer no mesmo dia da primeira), a terceira tem 363... e a n-ésima pessoa tem 365-n+1. Então o numerador é $365 \cdot 364 \cdot 363 \cdot \ldots \cdot (365 - n + 1) = \frac{365!}{(365-n)!}$.

Resposta:

$$c(n) = 1 - \frac{365!}{(365 - n)! \cdot 365^n}$$

2. Usando a aproximação $e^x \approx 1 + x$, determine o valor aproximado para $c(n)$.

???

3. Usando o valor aproximado de $c(n)$, determine o menor valor de n tal que a chance de colisão na data de aniversário seja maior do que $1/2$. Você considera este número alto ou baixo?

---

**Questão 9: Caras em sequência**

Considere uma moeda enviesada, tal que o probabilidade do resultado ser cara é p (e coroa 1-p). Considere o número de vezes que a moeda precisa ser jogada para obtermos k caras consecutivas. Por exemplo, na sequência "COCOCCOOCOCCC" a moeda teve que ser jogada 13 vezes até o aparecimento de $k = 3$ caras consecutivas, onde $C =$ cara e $O =$ coroa.

1. Seja $N_k$ a variável aleatória que denota esta quantidade. Qual é o numero médio de vezes que a moeda precisa ser jogada para obtermos k caras consecutivas, ou seja, qual é o valor esperado de $N_k$? Dica: comece com $k = 1$, monte uma recursão em k, e use a regra da torre da esperança.

Para $k = 1$, $N_k$ é uma variável geométrica $N_1 \sim Geom\left(\frac{1}{2}\right)$. O valor esperado de uma geométrica é $\frac{1}{p}$.

Para $k = 2$, primeiro deve ocorrer $N_1$, e então é preciso que ocorra mais uma cara. Senão, o processo recomeça recursivamente.

$$\begin{aligned}
E[N_2] &= E[N_1] + (1 \cdot p + (1 + E[N_2])(1-p)) \\
E[N_2] &= \frac{1}{p} + p + 1 - p + E[N_2] - pE[N_2] \\
E[N_2] - E[N_2] + pE[N_2] &= \frac{1}{p} + 1 \\
E[N_2] &= \frac{1}{p^2} + \frac{1}{p} \\
&= \frac{p+1}{p^2}
\end{aligned}$$

Para $k = 3$:

$$\begin{aligned}
E[N_3] &= E[N_2] + (1 \cdot p + (1 + E[N_3])(1-p)) \\
pE[N_3] &= \frac{p+1}{p^2} + 1 \\
E[N_3] &= \frac{p+1}{p^3} + \frac{1}{p} \\
&= \frac{p^2+p+1}{p^3}
\end{aligned}$$

Mais uma vez...

$$
\begin{aligned}
E[N_4] &= E[N_3] + (1 \cdot p + (1 + E[N_4])(1-p)) \\
pE[N_4] &= \frac{p^2+p+1}{p^3} + 1 \\
E[N_4] &= \frac{p^2+p+1}{p^4} + \frac{1}{p} \\
&= \frac{p^3+p^2+p+1}{p^4}
\end{aligned}
$$

Agora é possível enxergar o padrão. Generalizando para $k$:

$$
\begin{aligned}
E[N_k] &= \frac{p^{k-1}+p^{k-2}+\cdots+1}{p^k} \\
&= \frac{1}{p^k}\sum_{i=0}^{k-1} p^i
\end{aligned}
$$

O somatório é a soma da série geométrica com razão $p$. Se a razão for $r$ e o número de elementos for $r$, a soma da série geométrica é igual à $\frac{1-r^{n+1}}{1-r}$. No nosso caso:

$$
\begin{aligned}
E[N_k] &= \frac{1}{p^k} \cdot \frac{1-p^k}{1-p} \\
&= \frac{1-p^k}{p^k(1-p)}.
\end{aligned}
$$