# Cálculo das Probabilidades II - Lista 1 - 2020/01 - Gabarito

Prof. Hugo Carvalho

15/01/2021

## – PARTE 1: FUNDAMENTOS –

**Questão 1:**

a) Provemos que $\bigcap_{\lambda \in \Lambda} \mathcal{F}_\lambda$ satisfaz às três condições para ser uma $\sigma$-álgebra, na ordem apresentada no *slide* 2 da aula 2.2:

i) $\Omega \in \bigcap_{\lambda \in \Lambda} \mathcal{F}_\lambda$, pois $\Omega \in \mathcal{F}_\lambda$, para todo $\lambda \in \Lambda$.

ii) Para provar a propriedade do complementar, tome $A \in \bigcap_{\lambda \in \Lambda} \mathcal{F}_\lambda$. Por definição da interseção, temos que $A \in \mathcal{F}_\lambda$, para todo $\lambda \in \Lambda$, que por serem todas $\sigma$-álgebras implica que $A^c \in \mathcal{F}_\lambda$, para todo $\lambda \in \Lambda$. Portanto, temos que $A^c \in \bigcap_{\lambda \in \Lambda} \mathcal{F}_\lambda$, e a propriedade está provada.

iii) Finalmente, para provar a propriedade chamada de *$\sigma$-aditividade*, tome $A_1, A_2, \dots \in \bigcap_{\lambda \in \Lambda} \mathcal{F}_\lambda$. Por definição da interseção, temos que $A_1, A_2, \dots \in \mathcal{F}_\lambda$, para todo $\lambda \in \Lambda$. Como cada $\mathcal{F}_\lambda$ é $\sigma$-álgebra, temos que $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}_\lambda$, para todo $\lambda \in \Lambda$, e portanto, $\bigcup_{n=1}^{\infty} A_n \in \bigcap_{\lambda \in \Lambda} \mathcal{F}_\lambda$.

b) Tal afirmação é falsa. Intuitivamente, a propriedade de $\sigma$-álgebra que tem mais chances de "falhar" no atual ceário é a iii), pois nota-se que é crucial que todos os conjuntos $(A_n)_{n \in \mathbb{N}}$ estejam em todas as $\sigma$-álgebra $\mathcal{F}_\lambda$, o que só é garantido com a interseção das $\mathcal{F}_\lambda$. Criemos agora um contra-exemplo para garantir a falsidade da afirmação. Seja $\Omega = \{1, 2, 3, 4\}$ e considere $\mathcal{F}_1 = \{\emptyset, \Omega, \{1\}, \{2, 3, 4\}\}$ e $\mathcal{F}_2 = \{\emptyset, \Omega, \{4\}, \{1, 2, 3\}\}$. Dessa forma, temos que $\mathcal{F}_1 \cup \mathcal{F}_2 = \{\emptyset, \Omega, \{1\}, \{4\}, \{1, 2, 3\}, \{2, 3, 4\}\}$, que não é uma $\sigma$-álgebra, pois $\{1\} \cup \{4\} = \{1, 4\} \notin \mathcal{F}_1 \cup \mathcal{F}_2$.

c) *(Questão de resposta livre – apresento um modelo de resposta)* Primeiramente, note que não podemos recorrer à cardinalidade de conjuntos para comparar os seus "tamanhos", pois sabemos que podemos ter conjuntos distintos porém com a mesma cardinalidade (por exemplo, o conjunto dos números pares, o conjunto dos números ímpares e o conjunto dos números primos têm todos a mesma cardinalidade). Devemos então recorrer a alguma outra forma de comparar conjuntos, e a ordem parcial induzida pela propriedade de "estar contido em" parece razoável, pois intuitivamente pensamos que: se $A \subset B$ então "faltam elementos" em $A$ para que ele seja $B$ (continuando no exemplo anterior, é razoável dizer que o conjunto de primos é "menor" que o dos ímpares, pois "faltam" números para que o primeiro se torne o segundo). Dessa forma, dentre todas as $\sigma$-álgebra que satisfaçam a propriedade de conter a coleção $\mathcal{A}$, parece razoável chamar de menor (ou *minimal*) aquela que esteja contida em todas as outras com a mesma propriedade.

d) Intuitivamente, podemos argumentar que como $\mathcal{G}$ é um dos elementos sendo intersectados na definição de $\mathcal{F}_0$ (pois $\mathcal{G}$ é $\sigma$-álgebra contendo $\mathcal{A}$), então $\mathcal{F}_0$ é ainda mais "restrita" do que $\mathcal{G}$, pois tem ainda "menos" elementos, já que a operação de interseção "restringe" mais os conjuntos em questão.

Para provar rigorosamente, seguimos a dica e tome $\mathcal{G}$ uma $\sigma$-álgebra que também contém $\mathcal{A}$. Podemos então escrever que

$$\mathcal{F}_0 = \left[ \bigcap_{\substack{\mathcal{F} \text{ é } \sigma\text{-álgebra} \\ \mathcal{F} \text{ contém } \mathcal{A} \\ \mathcal{F} \neq \mathcal{G}}} \mathcal{F} \right] \cap \mathcal{G},$$

que pelas propriedades básicas da operação de interseção nos dá que $\mathcal{F}_0 \subset \mathcal{G}$. Ou seja, qualquer $\sigma$-álgebra contendo $\mathcal{A}$ também conterá $\mathcal{F}_0$.

e) *(Questão de resposta livre – apresento um modelo de resposta)* Para definirmos a função de probabilidade acumulada, é razoável sabermos calcular a probabilidade de variáveis aleatórias pertencerem aos intervalos da forma $(-\infty, x]$, para todo $x \in \mathbb{R}$, justificando a motivação para tal definição. Como vimos em aula, podemos "transportar" a medida de probabilidade $\mathbb{P}$ do espaço de probabilidade $\Omega$ para $\mathbb{R}$ através de variáveis aleatórias, e para que $\mathbb{R}$ seja de fato chamado de um espaço de probabilidade é necessário que tenhamos uma $\sigma$-álgebra adequada. Motivados pelo cômputo da função de probabilidade acumulada, parece razoável pedir que tal $\sigma$-álgebra contenha todos os intervalos da forma $(-\infty, x]$, para $x \in \mathbb{R}$. É claro que $2^{\mathbb{R}}$ é uma possibilidade, porém conforme foi comentado em aula, o conjunto das partes de $\mathbb{R}^n$ pode trazer problemas técnicos[1]. A fim de evitar tais dificuldades, parece razoável tomar a $\sigma$-álgebra mais "econômica" contendo os conjuntos que temos interesse em aferir sua probabilidade. Como é bastante claro que a coleção $\mathcal{A} = \{(-\infty, x],\ \text{para } x \in \mathbb{R}\}$ por si só não forma uma $\sigma$-álgebra, tomamos então a "menor" que a contenha. Sabemos de sua existência pois ela pode ser descrita pela fórmula do item d), e é possível provar também a sua unicidade. Dessa forma, $\mathcal{B}(\mathbb{R})$ está devidamente motivada e construída, utilizando a teoria aqui apresentada.

**Questão 2:**

a) Utilizando a desigualdade de Bonferroni, temos que:

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n) = \sum_{n=1}^{\infty} 0 = 0.$$

b) Utilizando a desigualdade de Bonferroni, temos que:

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) \geq 1 - \sum_{n=1}^{\infty} \mathbb{P}(A_n^c) = 1 - \sum_{n=1}^{\infty} [1 - \mathbb{P}(A_n)] = 1 - \sum_{n=1}^{\infty} 0 = 1.$$

c) Da definição de probabilidade condicional, temos que:

$$\mathbb{P}\left(B \middle| \bigcup_{n=1}^{\infty} A_n\right) = \frac{\mathbb{P}(B \cap (\cup_{n=1}^{\infty} A_n))}{\mathbb{P}(\cup_{n=1}^{\infty} A_n)} = \frac{\mathbb{P}(\cup_{n=1}^{\infty} (B \cap A_n))}{\mathbb{P}(\cup_{n=1}^{\infty} A_n)},$$

onde na última igualdade usamos a propriedade distributiva da interseção sobre a união. Agora, como os conjuntos $A_n$ são dois-a-dois disjuntos, temos que os conjuntos $A_n \cap B$ também o são, de modo que:

$$\frac{\mathbb{P}(\cup_{n=1}^{\infty} (B \cap A_n))}{\mathbb{P}(\cup_{n=1}^{\infty} A_n)} = \frac{\sum_{n=1}^{\infty} \mathbb{P}(B \cap A_n)}{\mathbb{P}(\cup_{n=1}^{\infty} A_n)} = \frac{\sum_{n=1}^{\infty} \mathbb{P}(B|A_n)\mathbb{P}(A_n)}{\mathbb{P}(\cup_{n=1}^{\infty} A_n)},$$

onde na última igualdade usamos a definição de probabilidade condicional. Agora podemos usar a hipótese da questão, $\mathbb{P}(B|A_n) \geq c$, para concluir que:

$$\frac{\sum_{n=1}^{\infty} \mathbb{P}(B|A_n)\mathbb{P}(A_n)}{\mathbb{P}(\cup_{n=1}^{\infty} A_n)} \geq \frac{\sum_{n=1}^{\infty} c\mathbb{P}(A_n)}{\mathbb{P}(\cup_{n=1}^{\infty} A_n)} = c\frac{\sum_{n=1}^{\infty} \mathbb{P}(A_n)}{\mathbb{P}(\cup_{n=1}^{\infty} A_n)} = c,$$

onde na última igualdade utilizamos novamente o fato dos conjuntos $A_n$ serem dois-a-dois disjuntos. Portanto, temos o resultado desejado.

## – PARTE 2: VARIÁVEIS E VETORES ALEATÓRIOS –

**Questão 3:**

a) *(Questão de resposta livre – apresento um modelo de resposta)* Pensemos que $X$ e $Y$ representam o tempo de vida útil de componentes críticos para o funcionamento de determinado sistema, de modo que caso um deles falhe todo o sistema deve passar por uma manutenção. Assim, não temos a oportunidade de observar $X$ e $Y$, mas somente $Z$, que nos diz o menor tempo de vida útil dentre os dois componentes, e também $W$, que nos indica qual dos dois componentes falhou primeiro.

[1] Veja, por exemplo, o "Paradoxo de Banach-Tarski" e a existência de conjuntos ditos *não-mensuráveis*.

b) Note que a dica é razoável, pois está de acordo com o procedimento apresentado em aula para calcular a densidade do mínimo de variáveis aleatórias tendo em vista a relação entre o mínimo e a comparação *via* $\leq$. Portanto, temos que:

$$\begin{aligned}\mathbb{P}(Z \leq z, W = 0) &= \mathbb{P}(\min\{X, Y\} \leq z, W = 0) \\ &= \mathbb{P}(Y \leq z, Y \leq X) \\ &= \int_0^z \int_y^\infty \lambda e^{-\lambda x} \mu e^{-\mu y} \, dx dy \\ &= \int_0^z -\mu e^{-\mu y} e^{-\lambda x} \Big|_{x=y}^{x=\infty} \, dy \\ &= \int_0^z \mu e^{-(\mu+\lambda) y} \, dy \\ &= -\frac{\mu}{\mu+\lambda} e^{-(\mu+\lambda)y} \Big|_{y=0}^{y=z} \\ &= -\frac{\mu}{\mu+\lambda}[e^{-(\mu+\lambda)z} - 1] \\ &= \frac{\mu}{\mu+\lambda}[1 - e^{-(\mu+\lambda)z}].\end{aligned}$$

Analogamente, temos que:

$$\begin{aligned}\mathbb{P}(Z \leq z, W = 1) &= \mathbb{P}(\min\{X, Y\} \leq z, W = 1) \\ &= \mathbb{P}(X \leq z, X \leq Y) \\ &= \int_0^z \int_x^\infty \lambda e^{-\lambda x} \mu e^{-\mu y} \, dy dx \\ &= \int_0^z -\lambda e^{-\lambda x} e^{-\mu y} \Big|_{y=x}^{y=\infty} \, dx \\ &= \int_0^z \lambda e^{-(\mu+\lambda) x} \, dx \\ &= -\frac{\lambda}{\mu+\lambda} e^{-(\mu+\lambda)x} \Big|_{x=0}^{x=z} \\ &= -\frac{\lambda}{\mu+\lambda}[e^{-(\mu+\lambda)z} - 1] \\ &= \frac{\lambda}{\mu+\lambda}[1 - e^{-(\mu+\lambda)z}].\end{aligned}$$

Dessa forma, tomando a derivada em relação a $z$, temos que a função densidade de probabilidade e massa de probabilidade conjunta do par aleatório $(Z, W)$ é dada por:

$$f_{Z,W}(z, w) = \begin{cases} \mu e^{-(\mu+\lambda)z}, & \text{para } z \geq 0 \text{ e } w = 0 \\ \lambda e^{-(\mu+\lambda)z}, & \text{para } z \geq 0 \text{ e } w = 1. \end{cases}$$

c) Para encontrar a marginal em $Z$, usamos o resultado do item b) e encontramos primeiramente a sua função de probabilidade acumulada:

$$\begin{aligned}\mathbb{P}(Z \leq z) &= \mathbb{P}(Z \leq z, W = 0) + \mathbb{P}(Z \leq z, W = 1) \\ &= \frac{\mu}{\mu+\lambda}[1 - e^{-(\mu+\lambda)z}] + \frac{\lambda}{\mu+\lambda}[1 - e^{-(\mu+\lambda)z}] \\ &= 1 - e^{-(\mu+\lambda)z},\end{aligned}$$

implicando que $Z \sim \text{Exp}(\mu + \lambda)$. A fim de completude, sua densidade é dada por $f_Z(z) = (\mu + \lambda)e^{-(\mu+\lambda)z}$, para $z \geq 0$.

Para calcular a marginal em $W$, note que:

$$\begin{aligned}
\mathbb{P}(W=0) &= \mathbb{P}(Y \leq X) \\
&= \int_0^\infty \int_y^\infty \lambda e^{-\lambda x} \mu e^{-\mu y} \; dxdy \\
&= \int_0^\infty -\mu e^{-\mu y} e^{-\lambda x} \Big|_{x=y}^{x=\infty} \; dy \\
&= \int_0^\infty \mu e^{-(\mu+\lambda)y} \; dy \\
&= -\frac{\mu}{\mu+\lambda} e^{-(\mu+\lambda)y} \Big|_{y=0}^{y=\infty} \\
&= -\frac{\mu}{\mu+\lambda}[0-1] \\
&= \frac{\mu}{\mu+\lambda},
\end{aligned}$$

e também que:

$$\mathbb{P}(W=1) = 1 - \mathbb{P}(W=0) = \frac{\lambda}{\mu+\lambda}.$$

Assim, temos que $W \sim \text{Bern}\left(\frac{\lambda}{\mu+\lambda}\right)$.

d) As condicionais são dadas por:

$$\begin{aligned}
f_{Z|W}(z|1) &= \frac{f_{Z,W}(z,1)}{f_W(1)} = \frac{\lambda e^{-(\mu+\lambda)z}}{\lambda/(\mu+\lambda)} = (\mu+\lambda)e^{-(\mu+\lambda)z}, \\
f_{Z|W}(z|0) &= \frac{f_{Z,W}(z,0)}{f_W(0)} = \frac{\mu e^{-(\mu+\lambda)z}}{\mu/(\mu+\lambda)} = (\mu+\lambda)e^{-(\mu+\lambda)z}.
\end{aligned}$$

e) Uma forma de concluir a independência é notar que as distribuições condicionais $Z|(W=1)$ e $Z|(W=0)$ são iguais. Outra forma é notarmos que $f_{Z,W}(z,w) = f_Z(z)f_W(w)$, para $z \geq 0$ e $w = 0, 1$.

f) *(Questão de resposta livre – apresento um modelo de resposta)* Prosseguindo no cenário proposto no item a), podemos interpretar a independência entre $Z$ e $W$ das seguintes formas:

   i) Se conhecemos o tempo até algum dos dois componentes apresentar a primeira falha, nada de novo sabemos sobre qual componente falhou primeiro;

   ii) Se sabemos qual dos dois componentes falhou primeiro, não ganhamos informação alguma sobre o tempo até tal falha.

   Ao meu ver tal resultado parece um pouco contra-intuitivo, pois $\lambda$ e $\mu$ podem ser parâmetros distintos, de modo que o tempo médio de vida útil de cada componente pode ser diferente. Porém, o fato de tais tempos serem modelados por distribuições exponenciais torna tal informação irrelevante, conforme está explicitado nos cálculos realizados nos itens anteriores.

**Questão 4:**

a) Como a variável aleatória $X$ pode assumir valores positivos e negativos, não podemos usar o teorema apresentado na *slide* 1 da aula 13.2, pois ele assume que a função $g$, nesse caso $g(x) = x^2$, é bijetiva no suporte da variável aleatória $X$. Calculemos então primeiro a função de probabilidade acumulada de $Y$:

$$\begin{aligned}
F_Y(y) &= \mathbb{P}(Y \leq y) \\
&= \mathbb{P}(X^2 \leq y) \\
&= \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) \\
&= F_X(\sqrt{y}) - F_X(-\sqrt{y}),
\end{aligned}$$

para $y \geq 0$, e onde usamos na última igualdade o fato de $X$ ser uma variável aleatória contínua. Derivando para encontrar a função densidade de probabilidade de $Y$, temos que:

$$f_Y(y) = \frac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})], \text{ para } y \geq 0.$$

Note que se a variável aleatória $X$ fosse positiva, encontraríamos somente a primeira parcela da soma acima. Podemos então interpretar que a segunda parcela da soma é análoga à primeira, porém incorpora a "redundância" da possível não bijetividade da função $g$.

b) Comecemos com a explicação intuitiva do teorema a ser enunciado. Na notação do item a), assuma por simplicidade que a variável aleatória $X$ assume valores em todo $\mathbb{R}$, de modo que a função $g : \mathbb{R} \to [0, \infty)$ não é bijetiva. Porém, podemos particionar $\mathbb{R}$ nos conjuntos $(-\infty, 0)$, $\{0\}$ e $(0, \infty)$, de modo que a restrição de $g$ tanto a $(-\infty, 0)$ quanto a $(0, \infty)$ satisfaz as hipóteses do teorema do *slide* 1 da aula 13.2, ou seja, é bijetiva e diferenciável. O conjunto $\{0\}$ é onde $g$ tem derivada zero, e em torno de tal ponto ela deixa de ser inversível. Porém, como tal conjunto tem probabilidade zero de ser assumido por $X$, ele pouco irá interferir no processo. Portanto, parece razoável pedir que o domínio da função $g$ possa ser particionado em conjuntos onde a restrição de $g$ a eles seja bijetiva e diferenciável, além de um conjunto "irrelevante", que poderá ser desconsiderado. Formalmente, podemos fazer o seguinte enunciado:

**Teorema**: Seja $X$ uma variável aleatória contínua assumindo valores no conjunto $A \subset \mathbb{R}$ e com função densidade de probabilidade $f_X$, e seja $Y = g(X)$, onde $g : A \to \mathbb{R}$ é uma função diferenciável. Assuma que exista uma partição $A_0, A_1, \ldots, A_k$ de $A$ tal que:

i) $\mathbb{P}(X \in A_0) = 0$

ii) As restrições de $g$ aos conjuntos $A_i$, denotadas por $g_i$ para $i = 1, \ldots, k$, são estritamente crescentes ou estritamente decrescentes

iii) A imagem de cada $A_i$ por $g$ é a mesma para todo $i = 1, \ldots, k$, e será denotada por $B$.

Então temos que $Y$ é uma variável aleatória contínua com densidade dada por

$$f_Y(y) = \begin{cases} \sum_{i=1}^{k} f_X(g_i^{-1}(y)) \left| \frac{d}{dy} g_i^{-1}(y) \right|, & \text{para } y \in B \\ 0, & \text{caso contrário.} \end{cases}$$

*(O rascunho da prova é uma parte de resposta livre – apresento um modelo de resposta)* Podemos enxergar um esboço da demonstração notando que a contribuição para a massa de probabilidade de $Y$ em torno de $y$ deve vir de todas as possibilidades para "$g^{-1}(y)$", que denotamos por $g_i^{-1}(y)$, para $i = 1, \ldots, k$, assumido que a imagem de cada $A_i$ por $g$ é a mesma para todo $i = 1, \ldots, k$. Isso justifica somarmos tais parcelas. Como a restrição de $g$ a cada $A_i$ é uma função bijetiva, podemos aplicar o teorema do *slide* 1 da aula 13.2, justificando assim as parcelas sendo somadas.

c) Seguindo o particionamento sugerido na dica, denotemos por $\arccos(y)$ a inversa de $g = \cos$ em $D_1 = (0, \pi]$. Ao se analisar o gráfico da função cos, notamos que nos intervalos $D_i$, para $i$ ímpar, a inversa de $g$ restrita a tais intervalos é a função arccos porém deslocada de $(i-1)\pi$ para adequar-se ao domínio $D_i$. Portanto, na notação introduzida no item b), temos que

$$g_i^{-1}(y) = \arccos(y) + (i-1)\pi, \text{ para } y \in [-1, 1] \text{ e } i \text{ ímpar.}$$

Analogamente, nos intervalos $D_i$, para $i$ par, a inversa de $g$ restrita a tais intervalos é a função $-\arccos$ porém deslocada de $(i-1)\pi$, também para adequar-se ao domínio $D_i$, de modo que

$$g_i^{-1}(y) = -\arccos(y) + (i-1)\pi, \text{ para } y \in [-1, 1] \text{ e } i \text{ par.}$$

Assim, podemos juntar ambas as expressões e dizer simplesmente que

$$g_i^{-1}(y) = (-1)^{i-1} \arccos(y) + (i-1)\pi, \text{ para } y \in [-1, 1] \text{ e } i = 1, 2, 3, 4, \ldots.$$

Do Cálculo, temos que

$$\left| \frac{d}{dy} g_i^{-1}(y) \right| = \frac{1}{\sqrt{1-y^2}}, \text{ para } y \in [-1, 1] \text{ e } i = 1, 2, 3, 4, \ldots.$$

Portanto, pelo teorema enunciado no item b), temos que

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{1-y^2}} \sum_{i=1}^{\infty} e^{(-1)^i \arccos(y) - (i-1)\pi}, & \text{para } y \in [-1, 1] \\ 0, & \text{caso contrário.} \end{cases}$$

**Questão 5:**

a) A intuição aqui é bastante semelhante àquela apresentada no item 4 b): particionamos o conjunto $A$ onde o vetor aleatório $\mathbf{X}$ toma seus valores em conjuntos $A_0, A_1, \ldots, A_k$ de modo que em cada $A_1, \ldots, A_k$ a função $g$ seja bijetiva, diferenciável e com inversa diferenciável, e o conjunto $A_0$ – que são os pontos aonde $g$ tem derivada zero ou outra coisa "esquisita" acontece – tem probabilidade zero. O teorema enuncia-se então como:

**Teorema**: Seja $\mathbf{X}$ um vetor aleatório contínuo assumindo valores no conjunto $A \subset \mathbb{R}^m$ e com função densidade de probabilidade $f_{\mathbf{X}}$, e seja $\mathbf{Y} = g(\mathbf{X})$, onde $g : A \to \mathbb{R}^m$ é uma função diferenciável. Assuma que exista uma partição $A_0, A_1, \ldots, A_k$ de $A$ tal que:

i) $\mathbb{P}(\mathbf{X} \in A_0) = 0$

ii) As restrições de $g$ aos conjuntos $A_i$, denotadas por $g_i$ para $i = 1, \ldots, k$, são inversíveis, com inversa (denotada por $h_i$) diferenciável

iii) A imagem de cada $A_i$ por $g$ é a mesma para todo $i = 1, \ldots, k$, e será denotada por $B$.

Então temos que $\mathbf{Y}$ é um vetor aleatório contínuo com densidade dada por

$$f_{\mathbf{Y}}(\mathbf{y}) = \begin{cases} \sum_{i=1}^k f_{\mathbf{X}}(h_i(\mathbf{y}))|J_i|, & \text{para } \mathbf{y} \in B \\ 0, & \text{caso contrário,} \end{cases}$$

onde $|J_i|$ representa o módulo do determinante jacobiano associado a função $h_i$.

b) Primeiramente, notemos que podemos escrever a densidade conjunta de $X_1$ e $X_2$ como

$$f(x_1, x_2) = \frac{1}{2\pi\sigma^2} e^{-(x_1^2+x_2^2)/2\sigma^2}, \text{ para } -\infty < x_1, x_2 < \infty,$$

e a função $g : \mathbb{R}^2 \to (0, \infty) \times (-1, 1)$ como

$$g(x_1, x_2) = \left( x_1^2 + x_2^2, \frac{x_1}{\sqrt{x_1^2 + x_2^2}} \right).$$

Segundo a observação, o problema da não bijetividade se dá por não conseguirmos determinar o sinal de $X_2$ a partir de $Y_1$ e $Y_2$. Portanto, consideramos a partição de $\mathbb{R}^2$ dada por

$$\begin{aligned} A_0 &= \{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 = 0\} \\ A_1 &= \{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 > 0\} \\ A_2 &= \{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 < 0\}, \end{aligned}$$

de modo que a restrição de $g$ a $A_1$ e $A_2$ são bijetivas e $A_0$ tem probabilidade zero por $(X_1, X_2)$. Denotando por $h_i = g_i^{-1}$ as inversas de $g$ quando restritas a $A_i$, para $i = 1, 2$ respectivamente, temos que

$$\begin{aligned} h_1(y_1, y_2) &= (y_2\sqrt{y_1}, \sqrt{y_1 - y_1 y_2^2}), \text{ para } (y_1, y_2) \in (0, \infty) \times (-1, 1), \\ h_2(y_1, y_2) &= (y_2\sqrt{y_1}, -\sqrt{y_1 - y_1 y_2^2}), \text{ para } (y_1, y_2) \in (0, \infty) \times (-1, 1). \end{aligned}$$

Após calcular os Jacobianos, notamos que

$$|J_1| = |J_2| = \frac{1}{2\sqrt{1 - y_2^2}}.$$

Finalmente, usamos o teorema enunciado no item a) para concluir que a densidade conjunta de $Y_1$ e $Y_2$ é dada por

$$\begin{aligned} f(y_1, y_2) &= \left[ \frac{1}{2\pi\sigma^2} e^{-y_1/2\sigma^2} \frac{1}{2\sqrt{1 - y_2^2}} \right] + \left[ \frac{1}{2\pi\sigma^2} e^{-y_1/2\sigma^2} \frac{1}{2\sqrt{1 - y_2^2}} \right] \\ &= \frac{1}{2\pi\sigma^2} e^{-y_1/2\sigma^2} \frac{1}{\sqrt{1 - y_2^2}}, \text{ para } (y_1, y_2) \in (0, \infty) \times (-1, 1). \end{aligned}$$

c) Nota-se que $Y_1$ e $Y_2$ são independentes pois podemos fatorar a sua densidade conjunta da seguinte forma:

$$f(y_1, y_2) = \widetilde{f}_1(y_1)\widetilde{f}_2(y_2), \forall\ (y_1, y_2) \in (0, \infty) \times (-1, 1)$$

sendo

$$\widetilde{f}_1(y_1) = \begin{cases} \frac{1}{2\pi\sigma^2} e^{-y_1/2\sigma^2}, & \text{se } 0 < y_1 < \infty \\ 0, & \text{caso contrário,} \end{cases}$$

$$\widetilde{f}_2(y_2) = \begin{cases} \frac{1}{\sqrt{1-y_2^2}}, & \text{se } -1 < y_2 < 1 \\ 0, & \text{caso contrário.} \end{cases}$$

*Obs.: Denotou-se por* $\widetilde{f}_i$ *pois não necessariamente elas representam as distribuições marginais.*

Quanto à interpretação geométrica da independência, note que $Y_1 = X_1^2 + X_2^2$ representa o quadrado da distância do ponto $(X_1, X_2)$ até a origem, e que $Y_2 = X_1/\sqrt{X_1^2 + X_2^2}$ representa o cosseno do ângulo entre o eixo horizontal e o segmento ligando o ponto $(X_1, X_2)$ até a origem. Portanto, a independência entre $Y_1$ e $Y_2$ diz, geometricamente, que se $X_1$ e $X_2$ são sorteadas por distribuições normais independentes e identicamente distribuídas, então a distância de tal ponto até a origem e o seu ângulo são variáveis aleatórias independentes.