Cálculo das Probabilidades II - Lista 3 - Gabarito - 2021/02

Prof. Hugo Carvalho

27/02/2022

Questão 1:

a) Como, à medida que n cresce a probabilidade se concentra cada vez mais em $\{X_n = -1\}$, parece razoável admitir que a convergência em probabilidade se dá para tal valor. Verifiquemos então este fato, a partir da definição de convergência em probabilidade. Tomando $\varepsilon > 0$, temos que:

$$\begin{aligned}\mathbb{P}(|X_n - (-1)| > \varepsilon) &= \mathbb{P}(X_n \neq -1)\\ &= \mathbb{P}(X_n - -(n+4)) + \mathbb{P}(X_n = n+4)\\ &= \frac{4}{n+4}\\ &\to 0,\ \text{quando } n \to \infty.\end{aligned}$$

Portanto, a convergência $X_n \xrightarrow{p} X$ se verifica, onde X assume o valor -1 com probabilidade 1. No entanto, temos que:

$$\begin{aligned}\mathbb{E}[X_n] &= -(n+4)\frac{1}{n+4} + (n+4)\frac{3}{n+4} - \left(1 - \frac{4}{n+4}\right)\\ &= -1 + 3 - 1 + \frac{4}{n+4}\\ &= 1 = \frac{4}{n+4}\\ &\to 1,\ \text{quando } n \to \infty,\end{aligned}$$

um valor diferente de $\mathbb{E}[X] = -1$.

b) Para estudar a convergência quase-certa da sequência $(X_n)_{n\in\mathbb{N}}$, fixado $\varepsilon > 0$, precisamos analisar o evento

$$\mathbb{P}(\{|X_n - (-1)| > \varepsilon\} \text{ para infinitos } n).$$

Pelo Lema de Borel-Cantelli, tal probabilidade é zero se e somente se

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n - (-1)| > \varepsilon) < \infty.$$

Porém, pelo raciocínio do item a), temos que $\mathbb{P}(|X_n - (-1)| > \varepsilon) = \dfrac{4}{n+4}$, de modo que o somatório acima é divergente, e portanto, concluímos que X_n não converge quase-certamente para -1, e portanto, para nenhuma outra variável aleatória (pois se os limites quase-certo e em probabilidade existem, eles devem ser iguais).

Questão 2: Analogamente à Questão 1, independente do valor de $\alpha > 0$, a probabilidade torna-se cada vez mais concentrada em 0 à medida que n cresce. Estudemos, portanto, a convergência quase-certa para tal limite. Fixado $\varepsilon > 0$, devemos estudar a probabilidade

$$\mathbb{P}(|X_n| > \varepsilon \text{ para infinitos } n),$$

e tal probabilidade é zero se e somente se o somatório $\displaystyle\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > \varepsilon)$ é convergente. É imediato notar que

$$\begin{aligned}\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > \varepsilon) &= \sum_{n=1}^{\infty} \mathbb{P}(X_n = n) + \mathbb{P}(X_n = -n)\\ &= \sum_{n=1}^{\infty} \frac{1}{n^{\alpha}},\end{aligned}$$

que é convergente se $\alpha > 1$ e divergente se $\alpha \in (0, 1]$. Portanto, para $\alpha > 1$ a sequência $(X_n)_{n\in\mathbb{N}}$ converse quase-certamente para zero, e para $\alpha \in (0, 1]$, não o é. Finalmente, note que 0 é o único candidato ao limite quase-certo de $(X_n)_{n\in\mathbb{N}}$ para $\alpha \in (0, 1]$, pois 0 é o limite em probabilidade de tal sequência, para todo $\alpha > 0$.

Questão 3: Denote por X_i o deslocamento do bêbado em cada passo. Pelas hipóteses do enunciado, temos que

$$\mathbb{P}(X_n = -1) = \mathbb{P}(X_n = 1) = \frac{1}{2},$$

e além disso tais variáveis aleatórias são independentes. Dessa forma, temos que $D_n = |X_1 + \cdots + X_n|$, e queremos estudar o comportamento assintótico de tal variável aleatória. Primeiramente, analisemos o comportamento assintótico de $S_n = X_1 + \cdots + X_n$, ou seja, D_n mas sem o valor absoluto. Como a sequência $(X_n)_{n\in\mathbb{N}}$ é independente e identicamente distribuída, temos, pelo Teorema Central do Limite, que

$$\frac{\sqrt{n}}{\sigma}\left(\frac{S_n}{n} - \mu\right) \xrightarrow{d} \mathcal{N}(0, 1).$$

Sendo $\mu = \mathbb{E}[X_n] = 0$ e $\sigma^2 = \mathbb{V}(X_n) = 1$, tal resultado simplifica-se para

$$\frac{S_n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1),$$

e portanto, informalmente temos que $S_n \approx \sqrt{n}\mathcal{N}(0, 1)$. Assim, concluímos que

$$D_n = |S_n| \approx \sqrt{n}|\mathcal{N}(0, 1)|,$$

de onde concluímos os resultados desejados, usando a dica.

Questão 4:

$$\mathcal{H}(X) = -\sum_{x\in\mathcal{X}} p(x)\log_2 p(x) = \mathbb{E}\left[\log_2\left(\frac{1}{p(X)}\right)\right].$$

a) Por um lado, temos que:

$$\begin{aligned}
\mathbb{P}(X = X') &= \sum_{x\in\mathcal{X}} \mathbb{P}(X = x \text{ e } X' = x) \\
&= \sum_{x\in\mathcal{X}} \mathbb{P}(X = x)\mathbb{P}(X' = x) \\
&= \sum_{x\in\mathcal{X}} p(x)^2,
\end{aligned}$$

o que prova a igualdade (1). Para provar (2), use o ponto (1) e note que:

$$\begin{aligned}
\mathbb{P}(X = X') &= \sum_{x\in\mathcal{X}} p(x)^2 \\
&= \sum_{x\in\mathcal{X}} p(x)p(x) \\
&= \sum_{x\in\mathcal{X}} p(x)2^{\log_2 p(x)} \\
&= \mathbb{E}[2^{\log_2 p(X)}] \\
&\overset{*}{\geq} 2^{\mathbb{E}[\log_2 p(X)]} \\
&\overset{**}{=} 2^{-\mathcal{H}(X)},
\end{aligned}$$

onde em $(*)$ usamos a desigualdade de Jansen para a função convexa $g(x) = 2^x$ e em $(**)$ usamos a definição de entropia.

b) Note que:

$$
\begin{aligned}
-\frac{1}{n}\log_2 p(X_1,\dots,X_n) &= -\frac{1}{n}\log_2[p(X_1)\dots p(X_n)] \\
&= -\frac{1}{n}\sum_{i=1}^{n}\log_2 p(X_i) \\
&= -\overline{\log_2 p(X_i)} \\
&= \overline{\log_2\left(\frac{1}{p(X_i)}\right)} \\
&\xrightarrow{qc} \mathcal{H}(X),
\end{aligned}
$$

pela Leis Forte de Kolmogorov, e usando o fato de que as variáveis aleatórias $\log_2\left(\frac{1}{p(X_i)}\right)$ são iid e com valor esperado igual a $\mathcal{H}(X)$. Portanto, também se dá a convergência em probabilidade.

c) É imediato da definição do conjunto típico que:

$$
\begin{aligned}
(x_1,\dots,x_n)\in A_\varepsilon^{(n)} &\Leftrightarrow 2^{-n(\mathcal{H}(X)+\varepsilon)} \le p(x_1,\dots,x_n) \le 2^{-n(\mathcal{H}(X)-\varepsilon)} \\
&\Leftrightarrow -(\mathcal{H}(X)+\varepsilon) \le \frac{1}{n}\log_2 p(x_1,\dots,x_n) \le -(\mathcal{H}(X)-\varepsilon) \\
&\Leftrightarrow \mathcal{H}(X)+\varepsilon \ge -\frac{1}{n}\log_2 p(x_1,\dots,x_n) \ge \mathcal{H}(X)-\varepsilon.
\end{aligned}
$$

Uma interpretação probabilística do conjunto típico é que ele representa as sequências de observações $(x_1,\dots,x_n)$ tais que retornam uma estimativa ε-próximas da entropia de X_n. Uma interpretação no contexto de Teoria da Informação é que o conjunto típico é um conjunto relativamente pequeno e de alta probabilidade, cujos elementos podem ser codificados em poucos *bits*, sendo portanto, fundamental para a teoria de códigos.

d) Do item c), temos que o conjunto típico está bastante associado à convergência em probabilidade demonstrada no item b), a saber:

$$
\begin{aligned}
&\mathbb{P}\left(\left|-\frac{1}{n}\log_2 px_1,\dots,x_n) - \mathcal{H}(X)\right| > \varepsilon\right) = \mathbb{P}\left((x_1,\dots,x_n)\notin A_\varepsilon^{(n)}\right) = \mathbb{P}\left(A_\varepsilon^{(n)^c}\right) \\
\Leftrightarrow\ &\mathbb{P}\left(\left|-\frac{1}{n}\log_2 p(x_1,\dots,x_n) - \mathcal{H}(X)\right| \le \varepsilon\right) = \mathbb{P}\left((x_1,\dots,x_n)\in A_\varepsilon^{(n)}\right) = \mathbb{P}\left(A_\varepsilon^{(n)}\right).
\end{aligned}
$$

A convergência em probabilidade do item b) nos implica, portanto, que:

$$
\begin{aligned}
&\mathbb{P}\left(A_\varepsilon^{(n)^c}\right) \to 0,\ \text{quando } n\to\infty \\
\Leftrightarrow\ &\mathbb{P}\left(A_\varepsilon^{(n)}\right) \to 1,\ \text{quando } n\to\infty.
\end{aligned}
$$

Da definição de limite, temos que existe n suficientemente grande tal que $\mathbb{P}\left(A_\varepsilon^{(n)}\right) \ge 1-\varepsilon$, como queríamos demonstrar.

e) Começando usando a dica nos dois primeiros passos, temos que:

$$
\begin{aligned}
1 &= \sum_{(x_1,\dots,x_n)\in\mathcal{X}^n} p(x_1,\dots,x_n) \\
&\ge \sum_{(x_1,\dots,x_n)\in A_\varepsilon^{(n)}} p(x_1,\dots,x_n) \\
&\overset{*}{\ge} \sum_{(x_1,\dots,x_n)\in A_\varepsilon^{(n)}} 2^{-n(\mathcal{H}(X)+\varepsilon)} \\
&= 2^{-n(\mathcal{H}(X)+\varepsilon)} \sum_{(x_1,\dots,x_n)\in A_\varepsilon^{(n)}} 1 \\
&= 2^{-n(\mathcal{H}(X)+\varepsilon)} \# A_\varepsilon^{(n)},
\end{aligned}
$$

onde na desigualdade marcada com $(*)$ usamos a definição do conjunto típico. Daí concluímos que

$$\#A_\varepsilon^{(n)} \leq 2^{n(\mathcal{H}(X)+\varepsilon)}.$$

f) Seguindo a dica, tome n suficientemente grande para que tenhamos $\mathbb{P}\left(A_\varepsilon^{(n)}\right) \geq 1-\varepsilon$. Dessa forma:

$$\begin{aligned}
1-\varepsilon &\leq \mathbb{P}\left(A_\varepsilon^{(n)}\right) \\
&= \sum_{(x_1,\ldots,x_n)\in A_\varepsilon^{(n)}} p(x_1,\ldots,x_n) \\
&\overset{*}{\leq} \sum_{(x_1,\ldots,x_n)\in A_\varepsilon^{(n)}} 2^{-n(\mathcal{H}(X)-\varepsilon)} \\
&= 2^{-n(\mathcal{H}(X)-\varepsilon)} \sum_{(x_1,\ldots,x_n)\in A_\varepsilon^{(n)}} 1 \\
&= 2^{-n(\mathcal{H}(X)-\varepsilon)} \#A_\varepsilon^{(n)},
\end{aligned}$$

de modo que $\#A_\varepsilon^{(n)} \geq (1-\varepsilon)2^{n(\mathcal{H}(X)-\varepsilon)}$, para n suficientemente grande.