From ParGO - Paralelismo, Grafos e Otimização

Manuscritos: Probabilidade Discreta

Probabilidade discreta é um conceito matmeático que pode ser visto como a frequência de ocorrência de elementos com certa propriedade em um conjunto. Como o leitor verá a seguir, as técnicas de contagem vistas nas seções anteriores são muito úteis em cálculos de probabilidades discretas. Começamos a seção com o significado formal que damos ao termo *probabilidade*.

Definição *(Probabilidade discreta) Seja* $X = \{x_1, \cdots, x_n\}$, $n \geq 1$, *um conjunto finito. Uma função* $p\colon \mathcal{P}(X) \to [0,1]$ *é uma probabilidade discreta sobre* X *se as seguintes condições são satisfeitas:*

1. $p(\emptyset) = 0$ *e* $p(X) = 1$*; e*
2. *se* $X_1 \subseteq X$ *e* $X_2 \subseteq X$ *são tais que* $X_1 \cap X_2 = \emptyset$, *então* $p(X_1 \cup X_2) = p(X_1) + p(X_2)$.

Para simplificar a notação, sempre que o conjunto X estiver claro pelo contexto, uma probabilidade discreta p sobre X será referenciada somente por probabilidade. Ainda por simplicidade de notação, escrevemos p_i ou $p(x_i)$ para representar $p(\{x_i\})$, para todo $x_i \in X$. Feitas a definição formal e as simplificações convenientes na notação, podemos buscar uma interpretação apropriada as termos empregados. Em primeiro lugar, o termo "discreta" usado para classificar a probabilidade definida destaca o fato de o conjunto X possuir uma bijeção com um número natural (por conveniência, indexamos os elementos de X com números naturais de 1 a n, mas isso não nos impede de perceber a existência de uma bijeção com, por exemplo, $n-1$). Partindo, portanto, de um conjunto enumerável X, estabelece-se uma função que associa um número a cada um dos subconjuntos de X. Para que essa função seja uma probabilidade, são impostas algumas condições. Primeiramente, o número associado a cada subconjunto deve ser um número real (novamente o conjunto dos números reais, do qual continuamos a supor a existência). Mas isso não basta para caracterizar uma probabilidade, pois **as duas condições** também precisam ser satisfeitas. Em particular, observe que a segunda delas é equivalente a

$p(X') = \sum_{x_i \in X'} p_i$,

de onde decorre que $p(X) = \sum_{x_i \in X} p_i$. Como $p(X) = 1$, temos $\sum_{x_i \in X} p_i = 1$.

Fica evidente pela definição que, dado um conjunto X, há diversas probabilidades distintas que podem ser definidas sobre X. O exemplo mais imediato de uma probabilidade é a função tal que $p_i = p_j$, para quaisquer $i, j \in \{1, 2, \ldots, n\}$, o que estabelece que cada subconjunto unitário de X tem exatamente a mesma frequência de ocorrência. Naturalmente, essa função é tal que $p_i = 1/n$, para todo $i \in \{1, 2, \cdots, n\}$. Observe que, nesse caso, $p(X') = |X'|/|X|$, para todo $X' \subseteq X$. Essa probabilidade é conhecida como *probabilidade* *de Laplace*.

Funções de probabilidade encontram aplicação na determinação da "frequência de ocorrência" de elementos de cada subconjunto de X em experimentos em que um subconjunto de X é escolhido aleatoriamente. Em tal tipo de experimento, o conjunto X é formado pelos possíveis resultados do experimento. Nesse contexto, para cada $x_i \in X$, o valor p_i pode ser interpretado da seguinte forma: se realizarmos o experimento um número suficientemente grande de vezes, obteremos o resultado x_i em uma razão dada por p_i. Assim sendo, quanto maior o valor atribuído por uma função de probabilidade a um certo subconjunto, maiores são as "chances" de um elemento desse subconjunto ser escolhido no experimento, e vice-versa. Vamos considerar os seguintes exemplos ilustrativos de tal aplicação.

Exemplo *Lança-se uma moeda, que pode cair em cara ou coroa. Definindo* $X = \{cara, coroa\}$, *podemos verificar que a função* $p: X \to [0,1]$ *tal que* $p(\emptyset) = 0$, $p(cara) = 1/2$, $p(coroa) = 1/2$ *e* $p(X) = p(cara) + p(coroa)$ *é uma probabilidade. De fato, trata-se da probabilidade de Laplace. Sempre que a chance de obtenção de cara ou coroa no lançamento da moeda for dada por essa função p, dizemos que a moeda é não-viciada. Caso contrário, estamos diante de uma moeda viciada.*

Exemplo *Ao jogar um dado, obtemos um número do conjunto* $\{1,2,3,4,5,6\}$. *Se p é uma função que atribui 0 ao conjunto vazio, o valor* $p_i = \frac{1}{6}$ *a cada número i entre 1 e 6, e o valor* $\sum_{x_i \in X'} p_i$ *para cada subconjunto* X' *de* $\{1,2,3,4,5,6\}$, *então p é uma probabilidade. Para convencer-se desse fato, basta verificar que* $p(\{1,2,3,4,5,6\}) = 1$. *Assim como no exemplo da moeda, a probabilidade assim definida (que é a probabilidade de Laplace para este caso) identifica um dado não-viciado, enquanto qualquer outra probabilidade corresponde a um dado viciado.*

Exemplo *Temos* C_5^{52} *possíveis jogos de pôquer em um baralho. A probabilidade de Laplace de se obter uma mão com quatro ases é de* $\frac{C_1^{48}}{C_5^{52}} = 0.0000185$, *pois são 48 os jogos que são formados com quatro ases.*

Algumas propriedades das probabilidades podem ser derivadas diretamente da definição. Vejamos dois exemplos onde $X = \{x_1, \cdots, x_n\}$ e $p: \mathcal{P}(X) \to [0,1]$ é uma probabilidade.

Lema *Se* $X'' \subseteq X$ *e* $X' \subset X''$, *então* $p(X') = p(X'') - p(X'' - X')$.

Demonstração: Consequência direta das **propriedades da definição de probabilidade** e do fato $X'' = X' \cup (X'' - X')$. $\square$

Lema *Se $X' \subseteq X$, então $p(X') = 1 - p(X - X')$*
.

Demonstração: Como $X' \cap (X - X') = \emptyset$, podemos usar **a segunda propriedade da definição de probabilidade** para obter $p(X) = p(X') + p(X - X')$. Logo, como $p(X) = 1$, temos o resultado desejado. $\square$

Uma outra propriedade de probabilidade complementa a definição, no sentido que indica a probabilidade da união de subconjuntos não-disjuntos. Mais especificamente, tomemos dois subconjuntos X_1 e X_2 de X, supondo que $X_1 \cap X_2 \neq \emptyset$. De forma semelhante ao raciocínio adotado nas demonstrações que acabamos de apresentar, escrevemos $X_1 \cup X_2 = X_1 \cup (X_2 - X_1)$, obtendo $p(X_1 \cup X_2) = p(X_1) + p(X_2 - X_1)$. Em seguida, escrevemos $X_2 - X_1 = X_2 - (X_2 \cap X_1)$, do que deduzimos, à luz do **Lema**, que $p(X_2 - X_1) = p(X_2) - p(X_2 \cap X_1)$. Juntando as peças, chegamos a $p(X_1 \cup X_2) = p(X_1) + p(X_2) - p(X_1 \cap X_2)$, igualdade que também pode ser derivada usando-se um raciocínio análogo ao Princípio da Inclusão e Exclusão. Vejamos como. Sabendo que $p(X_1) = \sum_{x_i \in X_1} p_i$ e $p(X_2) = \sum_{x_i \in X_2} p_i$, quando fazemos $p(X_1) + p(X_2)$, os termos referentes aos elementos de $X_1 \cap X_2$ aparecem duas vezes no somatório resultante. Assim sendo, $p(X_1 \cup X_2) = p(X_1) + p(X_2) - p(X_1 \cap X_2)$. Usando qualquer desses raciocínios, de maneira indutiva, para r subconjuntos de X, chegamos ao seguinte resultado.

Teorema *Se $X_1, X_2, \cdots, X_r$ são subconjuntos de X, então*

$$p(X_1 \cup X_2 \cup \cdots \cup X_r) = \sum_{1 \leq j \leq r} \left((-1)^{j+1} \sum_{1 \leq i_1 < \cdots < i_j \leq r} p(X_{i_1} \cap X_{i_2} \cap \cdots \cap X_{i_j}) \right)$$

Exemplo *Considere um conjunto de 100.000 pessoas, onde 51.500 são homens e 48.500 são mulheres. Entre os homens, suponha que 9.000 são loiros e entre as mulheres 30.200 são loiras. Suponha que devemos escolher uma pessoa aleatoriamente. A probabilidade de Laplace p de cada um ocorrer é $\frac{1}{100000}$. Considere os conjuntos formados por todos os homens loiros, todas as mulheres loiras, todos os homens morenos e todas as mulheres morenas, codificados como hl, ml, hm e mm, respectivamente. As probabilidades dos conjuntos acima são:*

$$p(mm) = 0.183, p(hl) = 0.090, p(hm) = 0.425, p(ml) = 0.302.$$

Por outro lado, sejam os conjuntos A, formado pelas pessoas loiras, e B, formado por todas as mulheres. Temos que $p(A \cup B)$ é a probabilidade de se escolher uma pessoa que é loira ou mulher; $p(A \cap B)$ é a probabilidade de se escolher uma mulher loira; $p(A - B)$ é probabilidade de se escolher um homem loiro; e $p(A \oplus B)$ é a probabilidade de se escolher um homem loiro ou uma mulher morena. Os valores dessas probabilidades são, respectivamente,

$p(A) = 0.090 + 0.302$, $p(B) = 0.183 + 0.302$, $p(A \cup B) = 0.183 + 0.090 + 0.302$, $p(A \cap B) = 0.302$, $p(A - B) = 0.090$ *e* $p(A \oplus B) = 0.090 + 0.183$.

Funções de probabilidade são também usadas em contextos mais gerais de experimentos consistindo de uma sequência de eventos aleatórios. Nesse tipo de experimento, um experimento básico, como o lançamento de uma moeda, é especificado, assim como uma função de probabilidade associada a esse experimento básico. O experimento propriamente dito consiste em repetir o experimento básico um certo número de vezes. Uma função de probabilidade associada a essa sequência é a seguinte.

Definição *(Probabilidade sequencial) Sejam* $X = \{x_1, x_2, \cdots, x_n\}$ *um conjunto finito e* p *uma probabilidade discreta sobre* X*. Seja também* X^r *o conjunto de* r*-uplas de elementos de* X*,* $r > 0$*. A probabilidade sequencial é a função* $p_s : \mathcal{P}(X^r) \to [0, 1]$ *tal que*

1. $p_s(\emptyset) = 0$;
2. *para cada* r*-upla* $x = (x_{i_1}, x_{i_2}, \cdots, x_{i_r})$ *em* X^r*,* $p_s(x) = p(x_{i_1})p(x_{i_2})\cdots p(x_{i_r})$*; e*
3. *para todo* $Y \subseteq X^r$*,* $p(Y) = \sum_{x \in Y} p_s(x)$*.*

Podemos facilmente verificar que a probabilidade sequencial, conforme definida acima, é de fato uma probabilidade. Para isso, basta verificar que $p_s(X^r) = \sum_{x \in X^r} p_s(x) = 1$ por indução em r. Inicialmente observamos que quando $r = 1$, temos $X^1 = X$ e $p_s = p$, e portanto é uma probabilidade. Tomando $r > 1$, temos que

$$\sum_{x \in X^r} p_s(x) = \sum_{x_i \in X} \left(p_i \sum_{x' \in X^{r-1}} p_s{}'(x')\right)$$

pois, para cada $x_i \in X$, o conjunto X^r inclui todas as r-uplas formadas por x_i na primeira coordenada e uma $r-1$-upla de X. Além disso, $p_s{}'$ é a probabilidade sequencial definida para X^{r-1}. Supondo, como hipótese de indução, que $p_s{}'$ é uma probabilidade, podemos escrever

$$\sum_{x' \in X^{r-1}} p_s{}'(x') = 1.$$

Assim, obtemos

$$\sum_{x \in X^r} p_s(x) = \sum_{x_i \in X} p_i.$$

Portanto, como p é uma probabilidade, chegamos ao resultado desejado.

Uma interpretação da probabilidade sequencial no contexto de experimentos aleatórios é a seguinte. Suponhamos que o experimento de realizar uma sequência de r eventos aleatórios seja repetido um número suficientemente grande de vezes. Nesses experimentos, obtemos o resultado x_{i_1} no primeiro evento na razão p_{i_1}. Tomemos apenas esses experimentos. A razão p_{i_2} deles têm o resultado x_{i_2} no segundo evento, o que nos a dizer que os resultados x_{i_1} e x_{i_2} ocorrem para o primeiro e segundo evento, respectivamente, na razão $p_{i_1} p_{i_2}$. Seguindo com esse raciocínio, chegamos a conclusão que a probabilidade sequencial expressa a razão com que obtemos a r-upla $x = (x_{i_1}, x_{i_2}, \cdots, x_{i_r})$ como resultado do experimento.

Uma outra propriedade da probabilidade sequencial diz respeito a $p_s(X_i^r)$, onde $X_i^r \subseteq X^r$ é o subconjunto de r-uplas que contém o elemento x_i apenas na primeira coordenada. Essa probabilidade pode ser escrita na forma

$p_s(X_i^r) = \sum_{x \in X_i^r} p_s(x)$.

Se escrevermos $x = (x_{i_1}, x_{i_2}, \cdots, x_{i_r})$ e, sem perda de generalidade, supusermos $i = 1$, então

$p_s(X_1^r) = p_1 \sum_{x \in X^r} p_{i_2} p_{i_3} \cdots p_{i_r}$.

Na igualdade acima, observamos que o somatório pode ser re-escrito na forma

$(p_2 + p_3 + \cdots + p_r) \sum_{x \in X^r} p_{i_3} \cdots p_{i_r}$.

Aplicando sucessivamente essa transformação, chegamos a

$p_s(X_i^r) = p_i (1 - p_i) \cdots (1 - p_i) = p_i (1 - p_i)^{r-1}$.

Vejamos alguns exemplos de aplicação da função de probabilidade sequencial na determinação da frequência de ocorrência de determinadas sequências de resultados em experimentos aleatórios.

Exemplo *Retomando* ***o exemplo da probabilidade de Laplace para o lançamento de uma moeda****, a igualdade anterior implica que a probabilidade de obter cara somente no primeiro lançamento de uma sequência de r lançamentos é* $p_i (1 - p_i)^{r-1} = (1/2)^r$.

Exemplo *Considere* ***o exemplo do lançamento de um dado com seis faces****. Em uma sequência de r lançamentos, a probabilidade de obter 1 somente no primeiro lançamento é* $(1/6)(5/6)^{r-1}$. *Observe que podemos interpretar o lançamento do dado de 6 faces como o lançamento de uma moeda (2 faces) se imaginarmos ver cara sempre que o dado cai com a face 1 e ver coroa quando qualquer o resultado é qualquer outra face. Dessa forma, essa moeda imaginária não é uma moeda não-viciada, mas sim uma moeda em que* $p(cara) = 1/6$ *e* $p(coroa) = 1 - p(cara) = 5/6$.

Exemplo *Oito estudantes serão entrevistados. Qual a probabilidade de Laplace de cada dois deles terem as mesmas aptidões entre as seguintes opções:* $\{$*iniciante, com pouca experiência, com experiência, com muita experiência*$\}$ *? Todas as possibilidades são* 4^8. *O número de permutações em que duas aptidões de cada tipo têm que ser atribuídas é* $8!/2!2!2!2!$ *(É como se tivéssemos que guardar oito bolas, duas de cada cor, em oito caixas com capacidade de apenas uma bola). Logo, a probabilidade procurada é dada por* $\frac{8!}{2!2!2!2!4^8}$.

Generalizando as observações dos exemplos acima, chegamos ao seguinte resultado.

Teorema *Se* $Y \subseteq X^r$ *é o conjunto de todas as* r*-uplas de* X *contendo* k *vezes o elemento* $x_i \in X$, *então*

$$p_s(Y) = C^r_k p_i^k (1 - p_i)^{r-k}.$$

Demonstração: Por indução em k, pode-se facilmente verificar que a probabilidade do subconjunto das r-uplas nas quais as k primeiras coordenadas são compostas pelo elemento i é

$$p_i^k (1 - p_i)^{r-k}.$$

Como essa é também a probabilidade para as r-uplas em que o elemento i aparece em um conjunto fixo, mas arbitrário, de k coordenadas (e não necessariamente as k primeiras coordenadas). Então, o teorema decorre do fato de existirem C^r_k possibilidades de se escolher k coordenadas dentre as r disponíveis. $\square$

Exemplo *Suponha que, dados* n *vértices* $v_1, v_2, \cdots, v_n$ *e uma moeda não-viciada, usamos o seguinte mecanismo para construir um grafo simples* G *com esses vértices. Inicialmente, o conjunto* E *de arestas de* G *é vazio. Para cada par* $v_{i_1} v_{i_2}$ *de vértices distintos, lançamos a moeda e observamos o resultado. Caso este seja cara, incluimos a aresta* $v_{i_1} v_{i_2}$ *no conjunto* E. *Então, a probabilidade de o grafo obtido ao final do processo possuir* ℓ *arestas,* $0 \leq \ell \leq C^n_2$, *é*

$$C_\ell^{C^n_2} (1/2)^{C^n_2}.$$