– "A matemática pode ser definida como a ciência na qual não se sabe jamais sobre o que fala nem se o que se diz é verdade."

Bertrand Russel, 1920

A
Definições Matemáticas

CONCEITOS e definições matemáticas que possuem grande importância no decorrer da tese são descritos em mais detalhes neste apêndice.

O apêndice é dividido em duas seções. Na Seção A.1 são descritos os principais aspectos relativos aos cumulantes e momentos de uma distribuição de probabilidade qualquer. Aspectos relativos à entropia de variáveis aleatórias são exploradas na Seção A.2.

A.1 Cumulantes e momentos

A.1.1 História

Os cumulantes foram inicialmente introduzidos pelo astrônomo, contador, matemático e estaticista dinamarquês Thorvald N. Thiele (1838-1910) que os denominou *semi-invariantes*.

O termo *cumulante* surgiu pela primeira vez em 1931 no artigo "The Derivation of the

Pattern Formulæ of Two-Way Partitions from Those of Simpler Patterns", Proceedings of the London Mathematical Society, Series 2, vol. 33, pp. 195-208, publicado pelo geneticista e estaticista Sir Ronald Fisher e o estaticista John Wishart, epônimo da distribuição de Wishart.

O historiador Stephen Stigler comenta que o termo cumulante foi sugerido a Fisher numa carta de Harold Hotelling. Em um outro artigo publicado em 1929, Fisher chamou-os de funções de momentos cumulativos.

A.1.2 Cumulantes e momentos de distribuições de probabilidade

Dada uma distribuição de probabilidade $p_Y(y)$, os momentos são obtidos a partir da *função característica*, também chamada de *função geradora de momentos*, definida, para uma variável real y, como

$$\begin{aligned} \Omega_Y(\omega) &\triangleq \int_{-\infty}^{\infty} p_Y(y)\exp(\jmath\omega y)dy \\ &\triangleq \mathbb{E}\left\{\exp(\jmath\omega y)\right\}. \end{aligned} \tag{A.1}$$

Expandindo-se $\Omega_Y(\omega)$ em uma série de potências em torno da origem obtém-se [Papoulis, 1991]:

$$\Omega_Y(\omega) = \sum_{k=0}^{\infty} \frac{\kappa_k}{k!}(\jmath\omega)^k, \tag{A.2}$$

em que κ_k é o momento centrado de ordem k.

O cumulante de ordem k é definido como [Papoulis, 1991]

$$c_k = \frac{\partial^k \Upsilon_Y(\omega)}{\partial\omega^k}, \tag{A.3}$$

em que

$$\Upsilon_Y(\omega) = \ln\left[\Omega_Y(\omega)\right], \tag{A.4}$$

é a *função geradora de cumulantes.*

Para o caso de varáveis complexas, a função característica é dada por [Amblard et al., 1996a]:

$$\begin{aligned} \Omega_{Y,Y^*}(\omega,\omega^*) &\triangleq \int_{-\infty}^{\infty} p_{Y,Y^*}(y,y^*)\exp\left[j\left(\frac{\omega y^* + \omega^* y}{2}\right)\right]\, dy\; dy^* \\ &\triangleq E\left\{\exp\left[j\left(\frac{\omega y^* + \omega^* y}{2}\right)\right]\right\}. \end{aligned} \tag{A.5}$$

Neste caso, a função geradora de cumulantes é escrita como:

$$\Upsilon_Y(\omega) \triangleq \ln\left[\Omega_{Y,Y^*}(\omega,\omega^*)\right] \tag{A.6}$$

A.1.3 Algumas propriedades

Invariância e equivariância

O cumulante de ordem um é equivariante enquanto todos os demais são invariantes a deslocamentos. Então, para um cumulante de ordem k da variável Y, denotado por $c_k(Y)$, tem-se

$$\begin{aligned} c_1(Y+\alpha) &= c_1(Y)+\alpha \\ c_k(Y+\alpha) &= c_k(Y), \end{aligned} \tag{A.7}$$

para α uma constante qualquer.

Homogeneidade

O cumulante de ordem k é homogêneo[1] de grau k, ou seja, para o caso real tem-se:

$$c_k(\alpha Y) = \alpha^k \cdot c_k(Y). \tag{A.8}$$

Considerando-se o caso complexo, o k-ésimo cumulante é definido como

$$c_k(Y,Y^*) = c_k(\underbrace{Y,\ldots,Y}_{s\text{ termos}},\underbrace{Y^*,\ldots,Y^*}_{q\text{ termos}}) \quad \forall\; s+q=k. \tag{A.9}$$

Então, de acordo com a Equação (A.9), a propriedade da homogeneidade para variáveis complexas é dada por [Lacoume et al., 1997; Amblard et al., 1996b]:

$$c_k(\alpha Y,\alpha Y^*) = (\alpha)^s \cdot (\alpha^*)^q \cdot c_k(Y,Y^*). \tag{A.10}$$

Desta maneira, para os cumulantes de ordem par, pode-se definir $s = q$ que fornece a homogeneidade como

$$c_k(\alpha Y) = |\alpha|^k \cdot c_k(Y). \tag{A.11}$$

[1]Esta propriedade é algumas vezes denominada de *multilinearidade*.

Aditividade

Se X e Y são variáveis aleatórias independentes então vale a seguinte relação:

$$c_k(X+Y) = c_k(X) + c_k(Y). \tag{A.12}$$

A.1.4 Cumulantes e momentos

Os cumulantes são relacionados com os momentos através da seguinte recursão [Nikias & Petropulu, 1993]:

$$c_k = \kappa_k - \sum_{i=1}^{n-1} \left(\begin{array}{c} k-1 \\ i-1 \end{array} \right) c_i \cdot \kappa_{k-i}. \tag{A.13}$$

Desta forma, o $k-$ésimo momento é um polinômio de grau k dos k primeiros cumulantes, dados, para o caso em que $k = 6$, na seguinte forma:

$$\begin{aligned}
&\kappa_1 = c_1 \\
&\kappa_2 = c_2 + c_1^2 \\
&\kappa_3 = c_3 + 3c_2c_1 + c_1^3 \\
&\kappa_4 = c_4 + 4c_3c_1 + 3c_2^2 + 6c_2c_1^2 + c_1^4 \\
&\kappa_5 = c_5 + 5c_4c_1 + 10c_3c_2 + 10c_3c_1^2 + 15c_2^2c_1 + 10c_2c_1^3 \\
&\kappa_6 = c_6 + 6c_5c_1 + 15c_4c_2 + 15c_4c_1^2 + 10c_3^2 + 60c_3c_2c_1 + 20c_3c_1^3 + 15c_2^3 + 45c_2^2c_1^2 + 15c_2c_1^4 + c_1^6.
\end{aligned} \tag{A.14}$$

No caso de ser uma distribuição de média nula, basta anular na Equação (A.14) os termos dos polinômios nos quais c_1 aparece.

Os polinômios da Equação (A.14) possuem uma interpretação combinatorial na qual os coeficientes "contam" as partições de conjuntos. Uma fórmula geral dos polinômios é dada por

$$\kappa_k = \sum_{\aleph} \prod_{B \in \aleph} \kappa_{|B|}, \tag{A.15}$$

em que $\aleph$ contém toda a lista de partições de um conjunto de tamanho k, e $B \in \aleph$ significa que B é um dos "blocos" nos quais o conjunto é particionado, sendo $|B|$ o tamanho do conjunto B.

Com isso, cada monômio é dado por uma constante multiplicando um produto de cumulantes nos quais a soma dos índices é k, por exemplo no termo $c_3c_2^2c_1$ a soma dos índices é $3 + 2 \cdot 2 + 1 = 8$, indicando que este termo aparece no polinômio do momento de oitava ordem.

A.1.5 Cumulantes conjuntos

O cumulante conjunto de várias variáveis aleatórias $Y_1, \ldots, Y_k$ é dado por [Nikias & Petropulu, 1993]:

$$c\left(Y_1, \ldots, Y_k\right) = \sum_{\aleph} \prod_{B \in \aleph} (|B|-1)! \cdot (-1)^{|B|-1} \prod_{i \in B} \mathbb{E}\left\{Y_i\right\}, \tag{A.16}$$

em que $\aleph$ é o conjunto com todas as combinações da seqüência $\{1, \ldots, k\}$ e B é o conjunto com todas as combinações em bloco do conjunto $\aleph$. Por exemplo,

$$\begin{aligned} c(X, Y, Z) = \mathbb{E}\{XYZ\} - \mathbb{E}\{XY\}\,\mathbb{E}\{Z\} - \mathbb{E}\{XZ\}\,\mathbb{E}\{Y\} \\ - \mathbb{E}\{YZ\}\,\mathbb{E}\{X\} + 2 \cdot \mathbb{E}\{X\}\,\mathbb{E}\{Y\}\,\mathbb{E}\{Z\}. \end{aligned}$$

Se as variáveis forem independentes, o cumulante conjunto delas é nulo e se as k varáveis forem todas iguais, o cumulante conjunto é dado por $c_k(Y)$.

O significado combinatorial da expressão dos momentos em termos dos cumulantes mostra-se mais elegante, conforme mostrado abaixo [Nikias & Petropulu, 1993]:

$$\mathbb{E}\left\{Y_1 \cdots Y_k\right\} = \sum_{\aleph} \prod_{B \in \aleph} c(Y_B), \tag{A.17}$$

em que $c(Y_B)$ é o cumulante conjunto associado às variáveis aleatórias $Y_1, \ldots, Y_k$, cujos índices são incluídos no bloco B . Por exemplo,

$$\mathbb{E}\{XYZ\} = c(X, Y, Z) + c(X, Y)c(Z) + c(X, Z)c(Y) + c(Y, Z)c(X) + c(X)c(Y)c(Y).$$

A.1.6 Cumulantes condicionais

A lei de média total, que afirma que $\mathbb{E}\{Y\} = \mathbb{E}\{\mathbb{E}\{Y|X\}\}$ e a lei de variância total, na qual $\operatorname{var}(Y) = \mathbb{E}\{\operatorname{var}(Y|X)\} + \operatorname{var}(\mathbb{E}\{Y|X\})$, são naturalmente generalizadas para os cumulantes condicionais. Em geral tem-se:

$$c(Y_1, \ldots, Y_k) = \sum_{\aleph} c\left(c(Y_{\aleph_1}|X), \ldots, c(Y_{\aleph_b}|X)\right), \tag{A.18}$$

em que o somatório é tomado sobre todas as partições $\aleph$ do conjunto $\{1, \ldots, k\}$ dos índices, $\aleph_1, \ldots, \aleph_b$ são todos os blocos da partição de $\aleph$ e $c(Y_{\aleph_k})$ indica o cumulante conjunto das variáveis aleatórias cujos índices estão naquele bloco da partição.

A.2 Entropia de variáveis aleatórias

A.2.1 Definição de entropia

Seja uma variável aleatória Y multidimensional, contínua, real e centrada (média nula) com uma função de densidade de probabilidade $p_Y(\mathbf{y})$. Define-se por *entropia* a seguinte quantidade:

$$\begin{aligned}\mathcal{H}(y) &= -\mathbb{E}\left\{\ln\left[p_Y(\mathbf{y})\right]\right\} \\ &= -\int_{-\infty}^{\infty} p_Y(\mathbf{y}) \cdot \ln\left[p_Y(\mathbf{y})\right] dy.\end{aligned} \tag{A.19}$$

A.2.2 Distribuições com máxima entropia

É de grande interesse em processamento de sinais encontrar as distribuições que possuem máxima entropia. Desta maneira é interessante solucionar o seguinte problema [Cover & Thomas, 1991]:

> Maximizar $\mathcal{H}(y)$ sob todas as distribuições $p_Y(y)$ que satisfazem
>
> 1. $p_Y(y) \geq 0$, com a igualdade válida somente fora do domínio S da variável;
> 2. $\int_S p_Y(y)dy = 1$;
> 3. $\int_S p_Y(y)f_i(y)dy = \kappa_i$, para $1 \leq ileqk$;
>
> em que κ_i é o momento centrado de i-ésima e $f_i(y)$ é uma função que faz $p_Y(y)$ respeitar a restrição.

Para resolver o problema acima, é necessário utilizar os multiplicadores de Lagrange. Assim, pode-se escrever o seguinte Lagrangiano [Cover & Thomas, 1991]:

$$J\left(p_Y(y)\right) = -\int_S p_Y(y) \cdot \ln\left[p_Y(y)\right] dy + \beta_0 \cdot \left(\int_S p_Y(y)dy\right) + \sum_{i=1}^{k} \beta_i \left(\int_S p_Y(y)f_i(y)dy = \kappa_i\right), \tag{A.20}$$

em que $\beta_0, \ \ldots \ , \beta_i$ são os multiplicadores de Lagrange.

Derivando-se a Equação (A.20) em relação à distribuição $p_Y(y)$ tem-se então:

$$\frac{\partial J\left(p_Y(y)\right)}{\partial p_Y(y)} = -\ln\left[p_Y(y)\right] - 1 + \beta_0 + \sum_{i=1}^{k} \beta_i \cdot p_Y(y), \tag{A.21}$$

em que $\beta_0, \ \ldots \ , \beta_i$ são escolhidos de tal forma que $p_Y(y)$ satisfaça as restrições.

Então quais são as distribuições que maximizam[2] a Equação (A.21)?

A resposta depende das restrições impostas. A título de exemplo considera-se dois casos:

1. **Suporte fixo** ($S = [a, b]$)

Neste caso, não há nenhuma restrição quanto aos momentos, logo os multiplicadores $\beta_1, \ \ldots \ , \beta_k = 0$ uma vez que não há necessidade de restrição. Deste modo, igualando-se a Equação (A.21) a zero tem-se:

$$\begin{aligned} -\ln\left[p_Y(y)\right] - 1 + \beta_0 &= 0 \\ \ln\left[p_Y(y)\right] &= \beta_0 - 1 \\ p_Y(y) &= \exp\left[\beta_0 - 1\right]. \end{aligned} \tag{A.22}$$

Resolvendo a integral sobre o suporte determinado, tem-se:

$$\begin{aligned} \int_a^b p_Y(y)dy &= 1 \\ \int_a^b \exp\left[\beta_0 - 1\right] dy &= 1 \\ \exp\left[\beta_0 - 1\right] \cdot (b - a) &= 1 \\ p_y(y) = \exp\left[\beta_0 - 1\right] &= \frac{1}{b - a}. \end{aligned} \tag{A.23}$$

Assim, sob a restrição de um suporte fixo, a distribuição com máxima entropia é a *distribuição uniforme*.

2. **Média e variância fixas**

Sob estas restrições, $\beta_0, \beta_1, \beta_2 \neq 0$ e $S =]-\infty, \infty[$. Assim tem-se a seguinte solução para a distribuição ao tomar-se $\frac{\partial J(p_Y(y))}{\partial p_Y(y)} = 0$:

$$p_Y(y) = \exp\left[\beta_0 + \beta_1 y + \beta_2 y^2 - 1\right]. \tag{A.24}$$

Logo, necessita-se encontrar os valores de β_0, β_1 e β_2 através do seguinte sistema de

[2]A rigor deve-se tomar a segunda derivada da Equação (A.20) para mostrar que é um valor de máximo.

equações:

$$\begin{aligned}\int_{-\infty}^{\infty} \exp\left[\beta_0 + \beta_1 y + \beta_2 y^2 - 1\right] dy = 1 \\ \int_{-\infty}^{\infty} y \cdot \exp\left[\beta_0 + \beta_1 y + \beta_2 y^2 - 1\right] dy = \kappa_1 \\ \int_{-\infty}^{\infty} y^2 \cdot \exp\left[\beta_0 + \beta_1 y + \beta_2 y^2 - 1\right] dy = \sigma^2 = \kappa_2.\end{aligned} \tag{A.25}$$

A solução do sistema na Equação (A.25) fornece os seguintes valores para os multiplicadores de Lagrange:

$$\begin{aligned}\beta_0 &= -\ln\left[\sqrt{2\pi\sigma}\right] \\ \beta_1 &= \kappa_1 \\ \beta_2 &= -\frac{1}{2\sigma^2}.\end{aligned} \tag{A.26}$$

Desta maneira, substituindo-se os valores da Equação (A.26) na Equação (A.24) obtém-se

$$p_Y(y) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(y-\kappa_1)^2}{2\sigma^2}\right]. \tag{A.27}$$

Logo, para a restrição de média e variância fixas, a *distribuição gaussiana* apresenta a máxima entropia.

É possível mostrar esta propriedade da distribuição gaussiana sob outra abordagem, conforme descrito na seção a seguir.

A.2.3 Entropia de uma variável gaussiana: abordagem alternativa

Seja Y uma variável aleatória gaussiana multidimensional e de média nula cuja densidade é escrita como:

$$p_G(\mathbf{y}) = \frac{1}{\sqrt[n]{2\pi} \cdot |\det(\mathbf{R}_y)|^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2}\mathbf{y}^T \mathbf{R}_y^{-1} \mathbf{y}\right), \tag{A.28}$$

em que n é a dimensão do vetor $\mathbf{y}$ e $\mathbf{R}_y$ é a matriz de autocorrelação de $\mathbf{y}$.

Ao tomar-se o logaritmo natural da Equação (A.28), obtém-se

$$\ln\left[p_G(\mathbf{y})\right] = -\frac{n}{2} \cdot \ln[2\pi] - \frac{1}{2}\ln\left[|\det(\mathbf{R}_y)|\right] - \frac{1}{2}\mathbf{y}^T \mathbf{R}_y^{-1} \mathbf{y}. \tag{A.29}$$

Como pode-se escrever [Picinbono & Barret, 1990]

$$\mathbf{y}^T\mathbf{R}_y^{-1}\mathbf{y} = \mathrm{tr}\left(\mathbf{y}\mathbf{y}^T\mathbf{R}_y^{-1}\right),$$

em que tr(·) é o traço da matriz, tem-se que

$$\mathbb{E}\left\{\mathbf{y}^T\mathbf{R}_y^{-1}\mathbf{y}\right\} = \mathbb{E}\left\{\mathrm{tr}\left(\mathbf{y}\mathbf{y}^T\mathbf{R}_y^{-1}\right)\right\} = \mathrm{tr}(\mathbf{I}) = n. \tag{A.30}$$

Então, substituindo os resultados obtidos nas Equações (A.29) e (A.30) na Equação (A.19) tem-se:

$$\mathcal{H}_G(\mathbf{y}) = \frac{n}{2}\cdot\{\ln[2\pi]+1\} + \frac{1}{2}\cdot\ln\left[\left|\det(\mathbf{R}_y)\right|\right], \tag{A.31}$$

em que $\mathcal{H}_G(\mathbf{y})$ é a entropia da distribuição gaussiana de média nula.

Um aspecto importante a ser demonstrado é que a distribuição gaussiana apresenta a maior entropia entre todas as distribuições. Para tal, considera-se uma função de densidade de probabilidade qualquer sobre a variável representada por $p_Y(\mathbf{y})$.

A média da v.a. $\ln\left[p_G(\mathbf{y})\right]$ é a mesma tanto considerando-a com uma distribuição qualquer $p_Y(y)$ como no caso particular de uma distribuição gaussiana $p_G(y)$. Isto porque a matriz de autocorrelação para as duas distribuições é a mesma, ou seja, a restrição é de que a distribuição tenha uma variância definida [Picinbono & Barret, 1990].

Daí, pode-se escrever

$$\int_{-\infty}^{\infty} p_Y(\mathbf{y})\cdot\ln\left[p_G(\mathbf{y})\right]d\mathbf{y} = \int_{-\infty}^{\infty} p_G(\mathbf{y})\cdot\ln\left[p_G(\mathbf{y})\right]d\mathbf{y} = -\mathcal{H}_G(\mathbf{y}). \tag{A.32}$$

A partir da definição da divergência de Kulback-Leibler pode-se escrever:

$$\begin{aligned}
\int_{-\infty}^{\infty} p_Y(\mathbf{y})\cdot\ln\left[\frac{p_Y(\mathbf{y})}{p_G(\mathbf{y})}\right]d\mathbf{y} &= \int_{-\infty}^{\infty} p_Y(\mathbf{y})\cdot\ln\left[p_Y(\mathbf{y})\right]d\mathbf{y} - \int_{-\infty}^{\infty} p_Y(\mathbf{y})\cdot\ln\left[p_G(\mathbf{y})\right]d\mathbf{y}\\
\int_{-\infty}^{\infty} p_Y(\mathbf{y})\cdot\ln\left[\frac{p_Y(\mathbf{y})}{p_G(\mathbf{y})}\right]d\mathbf{y} &= -\mathcal{H}_Y(\mathbf{y}) - \int_{-\infty}^{\infty} p_Y(\mathbf{y})\cdot\ln\left[p_G(\mathbf{y})\right]d\mathbf{y} \qquad \text{(A.33)}\\
\int_{-\infty}^{\infty} p_Y(\mathbf{y})\cdot\ln\left[\frac{p_G(\mathbf{y})}{p_Y(\mathbf{y})}\right]d\mathbf{y} &= \mathcal{H}_Y(\mathbf{y}) + \int_{-\infty}^{\infty} p_Y(\mathbf{y})\cdot\ln\left[p_G(\mathbf{y})\right]d\mathbf{y},
\end{aligned}$$

e substituindo-se na Equação (A.32) obtém-se a seguinte relação

$$\mathcal{H}_Y(\mathbf{y}) - \mathcal{H}_G(\mathbf{y}) = \int_{-\infty}^{\infty} p_Y(\mathbf{y})\cdot\ln\left[\frac{p_G(\mathbf{y})}{p_Y(\mathbf{y})}\right]d\mathbf{y}. \tag{A.34}$$

Ao utilizar-se então a desigualdade $\ln[x] \leq x - 1$, a igualdade só ocorre quando $x = 1$, obtendo-se

$$\mathcal{H}_Y(\mathbf{y}) - \mathcal{H}_G(\mathbf{y}) \leq 0, \tag{A.35}$$

obtendo-se a igualdade somente quando $p_Y(\mathbf{y}) = p_G(\mathbf{y})$.

Desta maneira mostra-se que a entropia da variável gaussiana é máxima.