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 $1 \leq i \leq k$;

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 $\mathrm{tr}(\cdot)$ 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:

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

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. $\ln[p_G(\mathbf{y})]$ 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.