Nota de alerta sobre uso de médias

Após finalizar a análise de incerteza de um modelo, pode parecer tentador utilizar a média dos resultados do modelo como uma medida da tendência central dos resultados. No entanto, isso pode levar a um resultado equivocado, e devemos usar sempre o máximo (ou os máximos) da densidade dos resultados para representar os pontos mais prováveis da sua distribuição.

Como exemplo, considere o modelo simples

$$y = x \tag{1}$$

Onde o parâmetro x é estimado, a partir da realização de um processo Poisson, como tendo o valor $\hat{x}$. Para estudar a incerteza do resultado do modelo y, vamos utilizar a função de verossimilhança de x, normalizada de forma a representar uma probabilidade. Lembremos que a função de verossimilhança de uma distribuição Poisson é:

$$L\left(\lambda|\hat{x}\right) = C\lambda^{\hat{x}}\mathrm{e}^{-\lambda} \tag{2}$$

Onde C é uma constante multiplicativa que deve ser ajustada de forma que a integral de $L(\lambda|\hat{x})$ seja igual a 1:

$$\begin{aligned}
\int_0^\infty C\lambda^{\hat{x}}\mathrm{e}^{-\lambda} &= 1 \\
C &= \left(\int_0^\infty \lambda^{\hat{x}}\mathrm{e}^{-\lambda}\right)^{-1} \\
C &= \Gamma(\hat{x}+1)^{-1}
\end{aligned}$$

O máximo da função $L(\lambda|\hat{x})$, descrita acima, ocorre em $\lambda = \hat{x}$. Logo, o valor de $\hat{x}$ deve ser usado como estimador pontual do valor mais provável do parâmetro x. Da mesma forma, intervalos de confiança para o parâmetro x devem ser construídos ao redor de $\hat{x}$.

Após tomar amostras desta distribuição, construimos a distribuição de resultados:

$$D(y) = Cy^{\hat{x}}\mathrm{e}^{-y} \tag{3}$$

Cuja média é dada por

$$\begin{aligned}
< D(y) > &= \int_0^\infty yCy^{\hat{x}}\mathrm{e}^{-y}dy \\
&= \int_0^\infty Cy^{\hat{x}+1}\mathrm{e}^{-y}dy \\
&= \Gamma(\hat{x}+2)C \\
&= \frac{\Gamma(\hat{x}+2)}{\Gamma(\hat{x}+1)} \\
&= \hat{x}+1
\end{aligned}$$

Note então que, se escolhemos representar a distribuição dos resultados pela sua média, vamos estar usando $\hat{x} + 1$, enquanto que o valor mais provável para y é $\hat{x}$, o mesmo que para o parâmetro x. Da mesma forma, ao construir intervalos de confiança para o resultado y, estes devem ser feitos ao redor de $\hat{x}$, e não ao redor da média $\hat{x} + 1$.

Durante a análise de matrizes de Lefkovitch com denso-dependência para *E. edulis*, a população final média é de cerca de 17.000 árvores (uma estimativa muito alavancada por uma pequena parcela de simulações com população final na casa dos 30 a 40 mil). Já a densidade máxima é atingida próxima a 5.000 árvores.