1. Considere uma variável aleatória $X \sim \text{Gompertz}(b, c)$, com função de distribuição acumulada:

$$F(x; b, c) = 1 - \exp\left(-\frac{b}{c}(e^{cx} - 1)\right), \quad x \geq 0,\, b > 0,\, c > 0.$$

   E função densidade de probabilidade (PDF) é dada por:

$$f(x; b, c) = be^{cx} \exp\left(-\frac{b}{c}(e^{cx} - 1)\right)$$

   (a) Simule uma amostra de tamanho $n = 300$ da distribuição de Gompertz com parâmetros $b = 0{,}5$ e $c = 2{,}0$, utilizando o método da inversa da função acumulada.

   (b) Construa uma função em R para a log-verossimilhança em função dos parâmetros b e c, com base na amostra simulada. Utilize essa função para estimar os parâmetros utilizando o pacote `maxLik` no R.

   (c) Construa o histograma da amostra gerada e, sobre ele, plote a curva da densidade ajustada da distribuição de Gompertz com os parâmetros estimados via máxima verossimilhança.

2. Há uma curva não linear muito importante baseada na função de Gompertz, frequentemente utilizada para modelar crescimento biológico, populacional ou tecnológico, dada por:

$$y = A \exp\Big(-Be^{-Cx}\Big),$$

   onde $A > 0$ representa o valor assintótico máximo (ou capacidade de crescimento), $B > 0$ está relacionado ao deslocamento inicial da curva, e $C > 0$ controla a taxa de crescimento.

   (a) Simule um conjunto de dados (x_i, y_i) com base na equação acima, utilizando os parâmetros verdadeiros $A = 10$, $B = 2$ e $C = 0{,}6$, adicionando um pequeno erro aleatório normal $\varepsilon_i \sim N(0, 1)$.

   (b) Ajuste o modelo não linear (curva de Gompertz) aos dados simulados utilizando a função `nls()` do R.

   (c) Compare os parâmetros nominais e os estimados obtidos no ajuste.

Tem mais na próxima página. . .

3. O conjunto de dados `colon`, disponível no pacote `survival` do R, contém informações clínicas de pacientes com câncer colorretal. Neste exercício, iremos analisar o tempo de sobrevivência (`time`) considerando apenas os eventos observados (`status == 1`) e utilizar a covariável `node4`, que indica se o paciente possui mais de 4 linfonodos afetados (`node4 = 1`) ou não (`node4 = 0`).

   O objetivo é ajustar os modelos paramétricos: Gompertz e Exponencial, assumindo que a covariável `node4` afeta apenas o parâmetro de taxa/escala b do modelo Gompertz e λ no modelo Exponencial.

   (a) Carregue o conjunto de dados `colon` e selecione apenas as observações com `status == 1`. Faça uma breve análise exploratória considerando a covariável `node4`.

   (b) Ajuste os seguintes modelos paramétricos utilizando máxima verossimilhança (pacote `maxLik` ou outro de sua preferência):

   - **Gompertz**: $f(x; b, c) = be^{cx} \exp\left[-\frac{b}{c}(e^{cx} - 1)\right]$, com $b_i = \exp(\beta_0 + \beta_1 \cdot \text{node4}_i)$ e c constante;
   - **Exponencial**: $f(x; \lambda) = \lambda \exp(-\lambda x)$, com $\lambda_i = \exp(\beta_0 + \beta_1 \cdot \text{node4}_i)$;

   (c) Para cada modelo, estime os parâmetros β_0, β_1 e interprete o efeito da covariável `node4`.

   (d) Plote os histogramas dos tempos observados e sobreponha as curvas de densidade ajustadas para cada modelo, separando os grupos `node4 = 0` e `node4 = 1`.

   (e) Compare os modelos entre si usando critérios de ajuste, como KS ou AIC, além da interpretação visual das curvas de densidade e dos efeitos da covariável.

   (f) Calcule a mediana de cada grupo node4 para a distribuição de Gompertz, utilizando a fórmula:

   $$\text{Mediana} = \frac{1}{c} \ln\left(1 - \frac{c}{b_i} \ln(0.5)\right),$$

   onde $b_i = \exp(\beta_0 + \beta_1 \cdot \text{node4}_i)$.