

Algoritmo sintetizado do código `random_cure`

Objetivo Gerar n observações (t_i, δ_i) sob um modelo de fração de cura, controlando a *proporção total de zeros* q (curados + censurados) e permitindo uma CDF de censura arbitrária ou uma regra uniforme simples.

Símbolos-chave

- $S_T(t)$ — função de sobrevivência teórica (do usuário).
- $\pi = S_T(\infty)$ — fração de cura.
- q — `prop_zeros` no código. Exige $\pi \leq q \leq 1$.
- $p_c = (q - \pi)/(1 - \pi)$ — fração de censura *entre os suscetíveis*.
- $Q_T(u)$ — quantil de T (opcional).
- $F_C(c)$ — CDF de censura fornecida (opcional).
- M_i — indicador: 1 suscetível, 0 curado.
- $|I|$ — cardinalidade (número de elementos) do conjunto I .
- `round` — arredondamento ao inteiro mais próximo (R).

Fluxo do algoritmo

1. Cura teórica

$$\pi \leftarrow S_T(\infty).$$

2. Frações desejadas

$$p_c \leftarrow \frac{q - \pi}{1 - \pi}.$$

3. Gerar tempos de falha verdadeiros T_i^{true} :

- Amostrar $U_i^{(1)} \sim \mathcal{U}(0, 1 - \pi)$.
- Se Q_T disponível, $T_i^{\text{true}} = Q_T(U_i^{(1)})$; caso contrário, resolver $S_T(t) = \pi + U_i^{(1)}$ numericamente.

4. Atribuir cura ou suscetibilidade

$$M_i \sim \text{Bernoulli}(1 - \pi), \quad T_i^{\text{true}} \leftarrow \infty \text{ se } M_i = 0.$$

5. Escolher quais suscetíveis serão censurados

$$I = \{i \mid M_i = 1\}, \quad m = |I|, \quad k = \text{round}(mp_c).$$

Escolher aleatoriamente $\mathcal{C} \subset I$ com $|\mathcal{C}| = k$.

6. Gerar tempos de censura C_i (apenas $i \in \mathcal{C}$)

- *Caso exista CDF de censura F_C :*

$$V_i \sim \mathcal{U}(0, F_C(T_i^{\text{true}})), \quad C_i = F_C^{-1}(V_i).$$

Esta escolha garante $C_i < T_i^{\text{true}}$.

- *Caso contrário (regra uniforme interna):*

$$C_i \sim \mathcal{U}(0, T_i^{\text{true}}).$$

Isto faz o tempo de censura ser escolhido uniformemente dentro do intervalo permitido $[0, T_i^{\text{true}})$ de *cada* suscetível selecionado, sem depender de um t_{max} global.

Para índices não censurados, definir $C_i = \infty$.

7. Construir observações

$$t_i = \min(T_i^{\text{true}}, C_i), \quad \delta_i = \begin{cases} 1, & M_i = 1 \text{ e } T_i^{\text{true}} \leq C_i, \\ 0, & \text{caso contrário.} \end{cases}$$

Por que $C_i \sim \mathcal{U}(0, T_i^{\text{true}})$? Sem CDF externa, o código impõe a censura apenas *dentro* do tempo de falha do próprio indivíduo. Isso:

- garante automaticamente $C_i < T_i^{\text{true}}$ (requisito de coerência);
- evita calcular um t_{max} comum; cada suscetível tem seu próprio limite natural T_i^{true} ;
- mantém a fração p_c exatamente porque o conjunto censurado foi escolhido antes (tamanho k).