Demonstração da Variância do Estimador de Kaplan-Meier

Pretendemos demonstrar a fórmula da variância do estimador de Kaplan-Meier:

$$\text{Var}[\hat{S}(t)] = [\hat{S}(t)]^2 \sum_{j:t_j<t} \frac{d_j}{n_j(n_j - d_j)}$$

Demonstração

O estimador de Kaplan-Meier, $\hat{S}(t)$, é dado por:

$$\hat{S}(t) = \prod_{j:t_j<t} \left(1 - \frac{d_j}{n_j}\right).$$

Definindo $Q_j = \left(1 - \frac{d_j}{n_j}\right)$, temos:

$$\hat{S}(t) = \prod_{j:t_j<t} Q_j.$$

Aplicando o logaritmo natural:

$$\log(\hat{S}(t)) = \sum_{j:t_j<t} \log(Q_j).$$

Supondo independência entre intervalos, a variância é:

$$\text{Var}[\log(\hat{S}(t))] = \sum_{j:t_j<t} \text{Var}[\log(Q_j)].$$

Aplicando o método Delta para a função $g(Q_j) = \log(Q_j)$:

$$\text{Var}[\log(Q_j)] \approx \frac{\text{Var}[Q_j]}{(E[Q_j])^2} \approx \frac{\text{Var}[Q_j]}{Q_j^2}.$$

Sabemos que $Q_j = 1 - \frac{d_j}{n_j}$, portanto:

$$\text{Var}[Q_j] = \text{Var}\left[\frac{d_j}{n_j}\right] = \frac{1}{n_j^2}\text{Var}[d_j].$$

O número de falhas d_j segue aproximadamente uma distribuição binomial:

$$d_j \sim \text{Binomial}(n_j, p_j), \quad \text{Var}[d_j] = n_j p_j (1 - p_j).$$

Estimando p_j por $\frac{d_j}{n_j}$, obtemos:

$$\text{Var}[Q_j] \approx \frac{d_j(n_j - d_j)}{n_j^3}.$$

Substituindo na expressão anterior:

$$\text{Var}[\log(Q_j)] \approx \frac{\frac{d_j(n_j - d_j)}{n_j^3}}{\left(1 - \frac{d_j}{n_j}\right)^2} = \frac{d_j}{n_j(n_j - d_j)}.$$

Assim,

$$\text{Var}[\log(\hat{S}(t))] = \sum_{j:t_j<t} \frac{d_j}{n_j(n_j - d_j)}.$$

Para retornar à escala original, seja $Y = \log(\hat{S}(t))$. Então:

$$\text{Var}[\hat{S}(t)] \approx \left(e^{E[Y]}\right)^2 \text{Var}[Y].$$

Aproximando $E[Y] \approx \log(\hat{S}(t))$:

$$\text{Var}[\hat{S}(t)] \approx [\hat{S}(t)]^2 \text{Var}[\log(\hat{S}(t))].$$

Finalmente, substituindo a expressão encontrada para $\text{Var}[\log(\hat{S}(t))]$:

$$\text{Var}[\hat{S}(t)] = [\hat{S}(t)]^2 \sum_{j:t_j<t} \frac{d_j}{n_j(n_j - d_j)}.$$