Combining imprecise Bayesian and maximum likelihood estimation for reliability growth models

Lev Utkin, Svetlana Zatenko, Frank Coolen

July 2009

What are reliability growth models?

Let $X_1, \ldots, X_n$ be a series of random variables,

$$X_i \sim p_i(x \mid \mathbf{b}, f(i, \mathbf{d})) \text{ or } X_i \sim F_i(x \mid \mathbf{b}, f(i, \mathbf{d})) \text{ (CDF)}.$$

- $\mathbf{b}$ are parameters of a probability distribution under consideration.
- $\mathbf{d}$ are parameters characterizing the growth, i.e., the growth is modelled by a function $f(i, \mathbf{d})$.

By using realization $\mathbf{K} = (k_1, \ldots, k_n)$ of $(X_1, \ldots, X_n)$, we have to find $\mathbf{b}, \mathbf{d}$ and probability measures of X_{n+1} under assumption of f and p_i.

Examples and standard methods for computing parameters

- Well-known models: Jelinski-Moranda, Schick-Wolverton model, Littlewood-Verrall, etc. (every model differs by a type of the lifetime probability distribution p and the growth function f).
- Regression models: $Y = f(\mathbf{X}, \mathbf{d}) + \epsilon$, $\mathbf{b} = (0, \sigma^2)$.
- A standard way for computing $\mathbf{b}, \mathbf{d}$ is maximization of the likelihood function

$$L(\mathbf{K} \mid \mathbf{b}, \mathbf{d}) = \prod_{i=1}^{n} p_i(k_i \mid \mathbf{b}, \mathbf{d}) \to \max_{\mathbf{b},\mathbf{d}}$$

A sketch of the main idea of new imprecise growth models (1)

1. Every X_i is governed by an unknown CDF belonging to a set $\mathcal{M}_i(\mathbf{d})$ defined by **lower and upper CDFs**:

$$\underline{F}_i(k \mid \mathbf{d}) = \inf_{F(k)\in\mathcal{M}_i(\mathbf{d})} F(k), \ \overline{F}_i(k \mid \mathbf{d}) = \sup_{F(k)\in\mathcal{M}_i(\mathbf{d})} F(k).$$

2 The likelihood function $L(\mathbf{K} \mid \mathbf{d}, F)$ is maximized over all distributions F from $\mathcal{M}_i(\mathbf{d})$ and the resulting "modified" likelihood function depends on $\mathbf{d}$:

$$L(\mathbf{K} \mid \mathbf{d}) = \max_{F\in\mathcal{M}_1(\mathbf{d}),\dots,F\in\mathcal{M}_n(\mathbf{d})} L(\mathbf{K} \mid \mathbf{d}, F).$$

Step 2 in the sketch

Proposition

Suppose that discrete random variables $X_1, \ldots, X_n$ are governed by a probability distribution $F(k)$ from sets $\mathcal{M}_i$ defined by bounds $\underline{F}_i(k)$ and $\overline{F}_i(k)$, $i = 1, \ldots, n$, respectively. If $X_1, \ldots, X_n$ are independent, then there holds

$$\begin{aligned}\max_{\mathcal{M}_1,\ldots,\mathcal{M}_n} L(\mathbf{K} \quad | \quad \mathbf{d}) \\ &= \max_{\mathcal{M}_1,\ldots,\mathcal{M}_n} \Pr\{X_1 = k_1, \ldots, X_n = k_n\} \\ &= \prod_{i=1}^{n} \left\{\overline{F}_i(k_i) - \underline{F}_i(k_i - 1)\right\}.\end{aligned}$$

A sketch of the main idea of new imprecise growth models (2)

3. The lower and upper CDFs for $\mathcal{M}_i(\mathbf{d})$ are constructed by means of **an imprecise Bayesian model with cautious parameter** s or parameters s_1, s_2 (Walley 1996, Quaeghebeur and de Cooman 2005) conditioned on the growth parameter $\mathbf{d}$ and the "growth" function $f(i, \mathbf{d})$.
4. The growth parameter $\mathbf{d}$ is determined by maximizing the "modified" likelihood function $L(\mathbf{K} \mid \mathbf{d})$ over all values of $\mathbf{d}$:

$$\widetilde{L}(\mathbf{K} \mid \mathbf{d}, s) = \prod_{i=1}^{n} \left\{ \overline{F}_i(k_i \mid \mathbf{d}, s) - \underline{F}_i(k_i - 1 \mid \mathbf{d}, s) \right\} \longrightarrow \max_{\mathbf{d}}$$

A software run reliability growth model with the imprecise beta-geometric model

- The likelihood for X_i is the geometric distribution $(1-r_i)^{k-1}r_i$ with the parameter r_i.
- The prior distribution for the parameter λ_i is the $\mathrm{Beta}(\alpha,\ \beta+f(i,d))$ distribution.
- The predictive CDF for X_i is the beta-geometric distribution.
- The imprecise model: $\alpha = s\gamma$, $\beta = s - s\gamma$, $\gamma \in [0,1]$.
- The "modified" likelihood function is

$$\widetilde{L}(\mathbf{K} \mid d,s) = \prod_{i=1}^{n} \left(\overline{F}_i(k_i \mid d,s) - \underline{F}_i(k_i - 1 \mid d,s)\right) \to \max_d .$$

A software reliability growth model with the imprecise negative binomial model

- The likelihood for X_i is the Poisson distribution with the parameter λ_i.
- The prior distribution for the parameter λ_i is the Gamma distribution $\mathrm{Gamma}(\alpha,\ \beta)$.
- The predictive CDF for X_i is the negative binomial distribution.
- The imprecise model: (α, β) in the triangle $(0, 0)$, $(s_1, 0)$, $(0, s_2)$ **(two caution parameters)**.

Regression model

Given n observations $(y_1, \mathbf{x}_1), ..., (y_n, \mathbf{x}_n)$. How to find $\mathbf{d} = (d_0, ..., d_n)^{\mathrm{T}}$ in

$$Y = d_0 + \sum_{j=1}^{m} d_j X_j + \epsilon \text{ ?}$$

- the noise $\epsilon \sim p(z \mid \sigma)$
- the prior conjugate distribution for the parameter σ is $\pi(\sigma \mid \mathbf{c})$
- the predictive CDF $F_n(z \mid s, \gamma) = F(z \mid s, \gamma)$
- Parameters $\mathbf{d}$ are computed by maximizing

$$\widetilde{L}(\mathbf{Z} \mid s) = \prod_{i=1}^{n} \left(\overline{F}(y_i - \mathbf{x}_i \mathbf{d} \mid s) - \underline{F}(y_i - \mathbf{x}_i \mathbf{d} - 1 \mid s) \right) \longrightarrow \max_{\mathbf{d}} .$$

Here $\mathbf{Z} = (y_1 - \mathbf{x}_1 \mathbf{d}, ..., y_n - \mathbf{x}_n \mathbf{d})$.

Conclusion

- The proposed imprecise growth models based on combining Bayesian inference and maximum likelihood estimation have many interesting and important properties.
- They can be regarded as a tool or framework for constructing various new models.
- Numerical experiments with the models have shown that they provide better and more cautious predictions in comparison with the well-known growth models when the number of observations is small.
- When the number of observations is large, the quality of prediction does not differ from the known models.

Questions

?