Common cause failure with no observation

PSAM9

A.Barros*, C.Bérenguer*, L.Dieulle*, M.Fouladirad*, , A.Grall*, D.Vasseur** and J.Dewailly**

*Université Technologie de Troyes and **EDF, MRI

23/05/07

Outlines

1. CCF parametric models
   - Basic parameter (BP) model
   - α factor model
   - Multiple Greek Letter model
2. Parameter Estimation
   - Point Estimation
   - Bayesian Estimation

CCF parametric models: Basic parametric model

- Observations: n_k number of failure of k component.
- Each component contributes uniformly to the failure
- CCF events : the failure of k components among the m-components set,

$$Q_k = \Pr(\text{failure of } k \text{ specific components}), \quad 1 \leq k \leq m.$$

- The global probability of failure of any component of the CCF-set :

$$Q_p = \sum_{k=1}^{m} \binom{m}{k} Q_k.$$

α factor model

- Observations: n_k.
- The parameters of this model are :

$$Q_p = \Pr(\text{failure of at least one component})$$

$$\alpha_k = \Pr(k \text{ components are failed/failure has occurred})$$

for $1 \leq k \leq m$

- The parameters $(\alpha_k)_{1 \leq k \leq m}$ satisfy the relation

$$\sum_{k=1}^{m} \alpha_k = 1.$$

- The α factor model is equivalent to the BP model :

$$\alpha_k = \frac{Q_k^*}{Q_P}$$

Multiple Greek Letter model

- The parameters are

$$Q_t = \Pr(\text{failure of at least } 1 \text{ specified component})$$

$\varrho_k^{(m)} = \Pr($failure is shared by at least k including the specific one / a common cause)

- For a group of three components the parameters are :
$\beta = \Pr($failure is shared by at least two components including the specific one/ a common cause failure has affected at least a specific component)
$\gamma = \Pr($failure is shared by all the three components / a common cause failure has affected at least two components including a specific one)
$Q_t = \Pr($failure of at least one specific component).

Multiple Greek Letter model

- The expression of the likelihood complicated and not a known distribution
- Study the case: a group of three components
- Bijective relation between MGL parameters and α factor model parameters

$$\begin{cases} \beta = \frac{2\alpha_2+3\alpha_3}{\alpha_1+2\alpha_2+3\alpha_3} \\ \gamma = \frac{3\alpha_3}{2\alpha_2+3\alpha_3} \\ Q_t = \frac{(\alpha_1+2\alpha_2+3\alpha_3)Q_P}{3} \end{cases} \Leftrightarrow \begin{cases} \alpha_1 = \frac{6(1-\beta)}{6-\beta(3+\gamma)} \\ \alpha_2 = \frac{3\beta(1-\gamma)}{6-\beta(3+\gamma)} \\ \alpha_3 = \frac{2\beta\gamma}{6-\beta(3+\gamma)} \\ Q_P = Q_t(3 - \frac{\beta}{2}(3+\gamma)) \end{cases}$$

Likelihood function

- Basic Model:
  - The likelihood function of $(n_1, \ldots, n_m)$ defined by $L(n_0, n_1, \ldots, n_m / Q_P, \alpha_1, \ldots, \alpha_m)$ is a $Multi(n_0, n_1, \ldots, n_m)$
- α model:
  - $N_D = \sum_{i=0}^{m} n_i$ number of solicitations of the system
  - The likelihood function of $(n_0, n_1, \ldots, n_m)$ with respect to $(Q_k)_{1 \leq k \leq m}$: $Multi(n_0, n_1, \ldots, n_m)$
  - each number n_k follows a binomial distribution
- MGL model:
  - Replacing the parameters α_i, using the one to one transformation: likelihood unknown
  - Directly $\Rightarrow$ complicated and unknown

Point Estimation

- To maximise the likelihood function of the observed data $(n_0, n_1, \ldots, n_m)$.
- In the basic model

$$\widehat{Q^*}_k^{MV} = \frac{n_k}{N_D} \quad 1 \leq k \leq m.$$

- In the α factor model:

$$\widehat{\alpha}_k^{MV} = \frac{n_k}{n_P}, \quad 1 \leq k \leq m.$$

- In the MGL model (Likelihood obtained by the α factor model) :

$$\widehat{\varrho}_k^{(m)^{MV}} = \frac{\sum_{i=k}^{m} i n_i}{\sum_{i=k-1}^{m} i n_i}, \quad 2 \leq k \leq m.$$

Point Estimation and No Observation

- When no observation available ($n_k = 0$, for $k = 0, \ldots m$).
- In the basic model

$$\widehat{Q^*}_k^{MV} = 0 \quad 1 \leq k \leq m.$$

- In the α factor model:

$$\widehat{\alpha}_k^{MV} = 0, \quad 1 \leq k \leq m.$$

- In the MGL model:

$$\widehat{\varrho}_k^{(m)^{MV}} = \frac{\sum_{i=k, n_i \neq 0}^{m} i n_i}{\sum_{i=k-1}^{m} i n_i}, \quad 2 \leq k \leq m, \quad \hat{\varrho}_k^{(m)^{MV}} = 0, \text{ if } \forall k, n_k = 0.$$

Existing Alternative Methods

- Point Estimate with Expert judgement:
  - preserving the expression of the ML estimator and replacing the number of failures by a value q between 0 and 1.
  - For a component with an exponential lifetime with parameter λ, if n failures are observed in $[0, T]$, $\widehat{\lambda} = \frac{n}{T}$
  - If no failure observed, $n = 0$:

$$\widehat{\lambda}_e = \frac{q}{T}, \quad 0 \leq q \leq 1$$

  - IEC 605-4 standard recommends to set q equal to $1/3$,
- Non parametric Inference
  - Failure if $U_i \geq S$ where U_i is a random variable
  - If $(U_{(1)}, \ldots, U_{(n)})$ ordered r.v U_i after n tests then

$$\Pr(U_{(j)} \leq U_{n+1} \leq U_{(j+1)}) = \frac{1}{n+1}, \;\; 0 \leq j \leq n$$

Uncertainties

- uncertainties resulting from the estimation procedures on small sized samples
- uncertainties related to the model assumptions
  - uncertainty about the test procedure for gathering data,
  - assumption of homogeneity of the data
  - calculations of mean impact vectors
- uncertainties due to the data-gathering and the constitution of the data bases
  - absence or insufficiency of information on the procedure of collection
- Solution: to treat the uncertainty on the CCF events frequency is to determine the probability distribution (nureg5485)

Bayesian Estimation

- Let be $\delta_1, \ldots, \delta_m$ parameters
- prior law $\pi_0(\delta_1, \ldots, \delta_m)$
- $C_1, C_2, \ldots, C_m$ observations
- $L(C_1, C_2, \ldots, C_m/\delta_1, \ldots, \delta_m)$ likelihood
- Posterior law:

$$\frac{L(C_1, C_2, \ldots, C_m/\delta_1, \ldots, \delta_m)\pi_0(\delta_1, \ldots, \delta_m)}{\int L(C_1, C_2, \ldots, C_m/\delta_1, \ldots, \delta_m)\pi_0(\delta_1, \ldots, \delta_m)d\delta_1 \ldots d\delta_m}$$

- Bayesian estimator: mean of the a posterior law

$$\hat{\delta}_k = \mathbb{E}(\delta_k/C_1, C_2, \ldots, C_m)$$

Bayesian estimation and basic model

- $1 - Q_p + \sum_{k=1}^{m} Q_k^* = 1$
- The prior law on Q_k^* such that $(1 - Q_p, Q_1^*, \ldots, Q_m^*)$ follows a Dirichlet distribution $\mathcal{D}(A_0, \ldots, A_m)$
- $Q_k^* \sim \beta(A_0, \sum_{i \neq k} A_i)$
- Posterior distribution is a Dirichlet distribution $\mathcal{D}(B_0, \ldots, B_m)$ where $B_k = A_k + n_k$.
- The Bayesien estimates of Q_k^* :

$$\hat{Q}_k^* = \frac{A_k + n_k}{\sum_{k=0}^{m} A_k + N_D}.$$

- If no observation $\hat{Q}_k^* = \frac{A_k}{\sum_{k=0}^{m} A_k + N_D}$.

Bayesian estimation and α Factor model

- Dirichlet prior on α_k with parameters A_k, $1 \leq k \leq m$,

$$\widehat{\alpha}_k = \frac{A_k + n_k}{\sum_{k=1}^{m}(A_k + n_k)} = \frac{A_k + n_k}{\sum_{k=1}^{m}(A_k) + n_P}.$$

$$\alpha_{k,mode} = \frac{A_k + n_k - 1}{\sum_{k=1}^{m}(A_k) + n_P - m}.$$

- If for a fixed k, $n_k = n_P$ then:

$$\widehat{\alpha}_k = \frac{n_P + 1}{m + n_P} \simeq 1 \quad \text{if } n_P >> m, \; \widehat{\alpha}_{k,mode} = 1.$$

- If no observation

$$\hat{\alpha}_k = \frac{A_k}{\sum_{k=1}^{m}(A_k)}, \quad \widehat{\alpha}_{k,mode} = \frac{A_k - 1}{\sum_{k=1}^{m}(A_k) - m}.$$

Bayesian estimation and MGL model

Problem : Choice of the prior and the calculation of the posterior law Two methods :

- Use the likelihood of α model parameters / change of variable
- approximate method :
  - Consider kn_k as independent observations
  - Suppose kn_k failure of k components occurred independently (false hypothesis)
  - Associate multinomial distribution to $(n_1, \ldots, mn_m)$

Bayesian estimation of MGL model parameters

- Use the α model priors .
  - $\varrho_k^{(m)} = f_k(\alpha_1, \ldots, \alpha_m) \Leftrightarrow \alpha_k = g_k(\varrho_1^{(m)}, \ldots, \varrho_m^{(m)})$
  - $\hat{\varrho}_k^{(m)} = f_k(\hat{\alpha}_1, \ldots, \hat{\alpha}_m) = \frac{\sum_{i=k}^m iA_i + in_i}{\sum_{i=k-1}^m iA_i + in_i}$
- Approximate method
  - Multinomial likelihood for $(n_1, 2n_2, 3n_3, \ldots)$
  - The prior on $Z_1 = 1 - \beta$, $Z_2 = \beta(1-\gamma)$ et $Z_3 = \beta\gamma$ is a Dirichlet $\mathcal{D}(A_1, A_2, A_3) A_k$, $1 \leq k \leq 3$
  - Dirichlet Posterior with parameters $A_k + kn_k - k$

$$\hat{\varrho}_k^{(m)} = \frac{\sum_{i=k}^m A_i + in_i}{\sum_{i=k-1}^m A_i + in_i}$$

$$\hat{\varrho}_{k,mode}^{(m)} = \frac{\sum_{i=k}^m A_i + in_i - i}{\sum_{i=k-1}^m A_i + in_i - i}$$

Bayesian estimation of MGL model parameters

If no information available

- Use the α model priors .

$$\hat{\varrho}_k^{(m)} = \frac{\sum_{i=k}^{m} iA_i}{\sum_{i=k-1}^{m} iA_i}$$

- Approximate method

$$\hat{\varrho}_k^{(m)} = \frac{\sum_{i=k}^{m} A_i}{\sum_{i=k-1}^{m} A_i}$$

$$\hat{\varrho}_{k,mode}^{(m)} = \frac{\sum_{i=k}^{m} A_i - i}{\sum_{i=k-1}^{m} A_i - i}$$

Bayesian estimates of the MGL model parameters

Dirichlet prior with parameters equal to 0.5 (non-informative) :

	β	γ
Maximum likelihood	$\frac{2n_2+3n_3}{n_1+2n_2+3n_3}$	$\frac{3n_3}{2n_2+3n_3}$
posterior mean approx	$\frac{2n_2+3n_3+1}{n_1+2n_2+3n_3+1.5}$	$\frac{3n_3+0.5}{2n_2+3n_3+1}$
posterior mean/ alpha model	$\frac{2n_2+3n_3+2.5}{n_1+2n_2+3n_3+3}$	$\frac{3n_3+1.5}{2n_2+3n_3+2.5}$

Example: Case of 3-components set

$n_1 \neq 0$, $n_2 = n_3 = 0$ with 5000 tests.

$n_3 = 0$	β	γ	Q_t
prior	3.6010^{-2}	1.6410^{-3}	2.1010^{-5}
Estimate	3.3310^{-2}	γ	2.0710^{-5}
Empirical var	2.4710^{-5}	0	
Estimate2	3.0310^{-2}	γ	2.0410^{-5}
Empirical var	2.3810^{-5}	0	

Example: Case of 3-components set

$n_1 \neq 0$, $n_2 \neq 0$, $n_3 = 0$ with 5000 tests.

$n_3 = 0$	β	γ	Q_t
prior	3.3310^{-2}	1.6410^{-3}	2.0710^{-5}
Estimate	3.3610^{-2}	1.6310^{-3}	2.0710^{-5}
Empirical var	1.3510^{-5}	1.8310^{-8}	

$n_3 \neq 0$	β	γ
real	0.08	0.25
MV	0.0628	0.4286
approximate	0.0730	0.4545
alpha	0.0629	0.4333

Example: Case of 4-components set

$n_1 \neq 0$, $n_2 = n_3 = n_4 = 0$ with 5000 tests.

$n_3 = 0$	β	γ	δ	Q_t
parameters	5.4910^{-2}	3.2410^{-3}	7.9410^{-3}	2.1010^{-5}
Estimate	5.4810^{-2}	3.2410^{-3}	7.9410^{-3}	2.0110^{-5}
Empirical var	1.4310^{-8}	0	0	

Thank you for your attention