Introduction to Reliability Theory (part 2)

Frank Coolen

UTOPIAE Training School II, Durham University

3 July 2018

Outline

Statistical issues

Software reliability

Decision making

Survival signature

Imprecision

Resilience

Statistical Issues

Random failure time $T > 0$ with CDF $F(t) = P(T \leq t)$

<u>Survival Function</u> $S(t) = P(T > t) = 1 - F(t)$

pdf $f(t) = F'(t) = -S'(t)$

<u>Hazard Rate</u> $h(t) = f(t)/S(t)$

Interpretation, for small δt: $h(t)\delta t \approx P(T \leq t + \delta t \mid T > t)$

<u>Cumulative Hazard Function (CHF)</u> $H(t) = \int_0^t h(x)dx$

$H(t) = \int_0^t \frac{f(x)}{S(x)} dx = -\ln S(t)$

$S(t) = \exp\{-H(t)\} = \exp\{-\int_0^t h(x)dx\}$

Most statistical methods for estimation of parameters in assumed models (e.g. Weibull) use the Likelihood Function, either maximising it or combined with a prior distribution in Bayesian analysis.

Reliability data are often affected by *right-censoring*: an item has not failed during a particular period of time.

Let $t_1, \ldots, t_n$ be observed failure times, and $c_1, \ldots, c_m$ right-censored observations. For a parametric model with parameter θ, the likelihood function is

$$L(\theta | t_1, \ldots, t_n; c_1, \ldots, c_m) = \prod_{j=1}^{n} f(t_j|\theta) \prod_{i=1}^{m} S(c_i|\theta).$$

This assumes that the censoring mechanism is independent of the failure process.

Regression models for reliability data

For regression models in reliability applications Weibull models are often used, with the survival function, depending on a vector of covariates x, given by

$$S(t; x) = \exp\left\{-\left(\frac{t}{\alpha_x}\right)^{\eta_x}\right\}.$$

Some simple forms are often used for the shape and scale parameters as functions of x, e.g. the loglinear model for α_x, specified via $\ln \alpha_x = x^T \beta$, with β a vector of parameters, and similar models for η_x.

The statistical methodology is then pretty similar to general regression methods (e.g. fitting by MLE or Bayesian methods) and implemented in statistical software packages.

Failure processes

A system fails at certain times, there may be some actions during this period which affect failure behaviour, e.g. minimal repairs to allow the system to continue its function, or replacement of some components, or other improvements of the system. Normally this should not include full replacement of the system.

Let the random quantities $T_1 < T_2 < T_3 < \ldots$ be the failure times of the system, let $X_i = T_i - T_{i-1}$ (with $T_0 = 0$).

Rate of OCcurrence Of Failure (ROCOF)

Let $N(t)$ be the number of failures in the period $(0, t]$, then the ROCOF is:

$$v(t) = \frac{d}{dt} EN(t).$$

Increasing (decreasing) ROCOF models a system that gets worse (better) over time.

Note that the ROCOF is not the same as the hazard rate (the definitions are clearly different!), although intuitively they are a bit similar. If we consider a standard Poisson process, with iid times between failures being exponentially distributed, then the ROCOF and hazard rate happen to be identical.

Non-Homogeneous Poisson Process models (NHPP)

The crucial assumption in these models, as for standard Poisson processes which are just a special case, is that the numbers of failures in distinct intervals are independent if the process characteristics are known.

A NHPP with ROCOF $v(t)$ is easiest defined by the property that the number of failures in interval $(t_1, t_2]$ is a Poisson distributed random quantity, with mean

$$m(t_1, t_2) = \int_{t_1}^{t_2} v(t)dt.$$

Suppose we have observed the system over time period $[0, r]$, and have observed failures at times $t_1 < t_2 < \ldots < t_n \leq r$, then the likelihood function is

$$L = \left\{ \prod_{i=1}^{n} v(t_i) \right\} \exp\left[- \int_0^r v(t) dt \right].$$

This enables inference as usual.

Two basic, often used parametric ROCOFs are

$$v_1(t) = \exp(\beta_0 + \beta_1 t)$$

and

$$v_2(t) = \gamma \eta t^{\eta - 1},$$

Software Reliability

Many models that have been suggested, during the last five decades, for software reliability, are NHPPs which model the software testing process as a fault counting process.

Many models are variations to a basic model which assumes:

(1) Software contains an unknown number of bugs, N.

(2) At each failure, one bug is detected and corrected.

(3) The ROCOF is proportional to the number of bugs present.

The basic model is a NHPP with

$$v(t) = (N - i + 1)\lambda, \quad \text{for } t \in [T_{i-1}, T_i),$$

for some constant λ.

N and λ are both considered unknown, and estimated from data, where N tends to be of most interest, or in particular the number of remaining bugs.

Software reliability is a very important topic area, in particular statistical support for software testing is challenging.

Book to appear (Wiley, August 2018): *Analytic Methods in Systems and Software Testing*, edited by Keneth, Ruggeri and Faltin.

Includes chapter by Wooff, Goldstein and Coolen on the use of Bayesian Graphical Models for software testing, including many aspects of implementation.

Challenge: considering reliability of systems with hardware and software, including interaction.

Decision Making

Traditionally, decisions related to reliability (e.g. planning of replacements or inspections, warranties, etc) were mainly based on *Renewal (Reward) Theorem*:

It is assumed that the same process goes on for a very long time ('infinitely'), consisting of consecutive cycli which are stochastic copies of each other.

Let X_i be the random length of cyclus i, and R_i a random reward associated with cyclus i. Assume that the X_i's are iid, and that the R_i's are iid, but allow dependence of R_i on X_i.

For a process starting at time 0, let $R(t)$ be the random cumulative reward upto time t. Then

$$\frac{R(t)}{t} \to \frac{E(R_i)}{E(X_i)} \text{ if } t \to \infty.$$

So 'long-run average reward per unit of time' is used.

Assumption underlying use of renewal reward theorem often unrealistic; optimisation over one cycle, or a few cycli, or a fixed period of time more realistic. Analysis may be harder, and decisions typically ‘less risky’.

In recent years, emphasis has shifted to service contracts of fixed length. This is a main game-changer, with possible advantages for reliability theory (e.g. wrt data ownership) and many new challenges for decision making, e.g. planning and location of spare parts.

Survival Signature

System with $K \geq 2$ types of components

m_k components of type $k \in \{1, 2, \ldots, K\}$, with $\sum_{k=1}^{K} m_k = m$

State vector $\underline{x} = (\underline{x}^1, \underline{x}^2, \ldots, \underline{x}^K)$, with $\underline{x}^k = (x_1^k, x_2^k, \ldots, x_{m_k}^k)$ the sub-vector representing the states of the components of type k with $x_i^k = 1$ if the ith component of type k functions and $x_i^k = 0$ if not. Structure function $\phi(\underline{x}) = 1$ if system functions with state $\underline{x}$ and $\phi(\underline{x}) = 0$ if not.

The Survival Signature $\Phi(l_1, l_2, \ldots, l_K)$, with $l_k = 0, 1, \ldots, m_k$, is the probability that a system functions given that *precisely* l_k of its components of type k function, for each $k \in \{1, 2, \ldots, K\}$.

There are $\binom{m_k}{l_k}$ state vectors $\underline{x}^k$ with precisely l_k of their m_k components $x_i^k = 1$, so with $\sum_{i=1}^{m_k} x_i^k = l_k$.

Let $S_{l_1,\ldots,l_K}$ denote the set of all state vectors for the whole system for which $\sum_{i=1}^{m_k} x_i^k = l_k$, $k = 1, 2, \ldots, K$. Assuming exchangeability of the failure times of the m_k components of type k

$$\Phi(l_1, \ldots, l_K) = \left[\prod_{k=1}^{K} \binom{m_k}{l_k}^{-1}\right] \times \sum_{\underline{x} \in S_{l_1,\ldots,l_K}} \phi(\underline{x})$$

Probability system functions at time t

Let $C_t^k \in \{0, 1, \ldots, m_k\}$ denote the number of components of type k in the system that function at time $t > 0$. If the probability distribution for the failure time of components of type k is known and has CDF $F_k(t)$, and we assume failure times of components of different types to be independent, then the probability that the system functions at time $t > 0$ is

$$P(T_S > t) = \sum_{l_1=0}^{m_1} \cdots \sum_{l_K=0}^{m_K} \Phi(l_1, \ldots, l_K) P(\bigcap_{k=1}^{K} \{C_t^k = l_k\}) =$$

$$\sum_{l_1=0}^{m_1} \cdots \sum_{l_K=0}^{m_K} \left[\Phi(l_1, \ldots, l_K) \prod_{k=1}^{K} \left(\binom{m_k}{l_k} [F_k(t)]^{m_k - l_k} [1 - F_k(t)]^{l_k} \right) \right]$$

Presented at 1st UTOPIAE Training School (Glasgow, November 2017).

Main challenges include upscaling to very large real-world systems (with ‘zooming in’), applications to systems with multiple functions and to networks with different routes through them.

For some questions of practical relevance other summaries of the full structure function may be required.

Imprecision

Use of imprecise probability theory (also shown at 1st UTOPIAE Training School) is attractive in reliability, particularly when considering new or updated systems or components, or when there is doubt about some model assumptions.

Inference and decision problems with imprecise probability tend to require solution of constrained optimisation problems, with the inference as target function and the set of probabilities as constraint.

There are more opportunities to use idea of imprecision to provide robustness, enabling simple models to be used (e.g. in Accelerated Life Testing).

Overall goal: Make the jobs of practitioners easier or better, preferably both!

Many opportunities and challenges!

Resilience

In recent years there has been increasing attention to ‘Resilience’: if things go wrong, how can negative effect be minimized?

Hugely important topic! For example for systems with multiple functions or multiple phases, some critical functions or phases may still be possible after a failure, possibly by re-configuring the system.

Of course, this also involves the usual aspects of maintenance, repair, etc.

Very many challenges, perhaps generalizing structure function as (imprecise) probability will prove useful.

The topic of ‘Reliability Theory’ is extremely wide, including aspects from very many traditional areas of mathematics and engineering.

For substantial progress it is important to develop theory and methods closely linked to real-world problems, as the gap between text-book problems and reality is often very large.

UTOPIAE ESRs are well placed to make good contributions!

Feel free to contact me for more information or references.