International Journal of Modern Research in Engineering and Technology (IJMRET)
www.ijmret.org Volume 4 Issue 2 ǁ February 2019.

# Bayes estimation of a Stochastic Index For Reliability and Quality Assessment of Transportation Systems

Elio Chiodo

Department of Industrial Engineering, University of Naples Federico II
via Claudio 21, 80125 Naples (Italy)

Luigi Pio Di Noia
DIETI
University of Naples Federico II, via Claudio 21, 80125 Naples (Italy)

**Abstract—** The paper address the topic of an efficient evaluations of the performances of transportation systems in terms of both Quality and Reliability, by means of an "ad hoc" developed mathematical index which puts the two aspects in close connection. This relationship is in accordance with the currentstate of art of the studies on the transportation Systems, and takes advantage of recent developments in the stochastic process theory, as well as risk and safety analysis. The two aspects of Quality and Reliability are indeed closely related, especially if dynamic aspects in system's behavior are taken into account. According to this philosophy, the authors discuss the main features of the proposed index, and illustrate the adoption of a suitable Bayesian method for its estimation, which is based upon prior (or "a priori") information, which should be easily available. For this purpose, the Beta and the Gamma distribution are chosen as prior models on the parameters, implying a Negative Log-Gamma for the index. Finally, the summary of a large set of numerical simulations is presented, which show the high efficiency of such Bayesian estimation methodology. In particular, its superiority with respect to the "classical" Maximum Likelihood (ML) estimation methods, traditionally adopted for such systems, is illustrated.

***Keywords— Bayes methods;Negative Log-Gamma distribution; Quality; Reliability; Transportation Systems***

## NOMENCLATURE

| | |
|---|---|
| *CDF* | cumulative distribution function |
| *DA* | Delay amplitude |
| *PDF* | probability density function |
| *SD*, $\sigma$ | standard deviation |
| *REFF* | Relative Efficiency of a Bayes estimator |
| *RV* | random variable |
| *NLG* | Negative Log-Gamma distribution |
| *D* | observed data |
| *E[X]* | mean value of the RV X |
| *EP* | exceedance probability |
| $g(\zeta)$, $g(\zeta\|D)$ | prior and posterior PDF of parameter $\zeta$ |
| $\zeta°$ | Bayes estimate of a generic parameter $\zeta$ |
| *ML* | Maximum Likelihood |
| *MSEB* | Mean Square Error of a Bayes estimator |
| *MSEL* | Mean Square Error of a ML estimator |
| *QI* | Quality Index |
| *TS* | Transportation systems |
| $\phi$ | fault frequency |
| $Z_j$ | delay amplitude at j-th fault occurrence |
| *Y(t)* | maximum delay amplitude in (0,t) |
| *Q* | Exceedance Probability = $P(W_j>z)$ |
| $\Gamma( . )$ | Euler-Gamma function |
| *z* | extremal value of the DA |

## I. INTRODUCTION

In recent times,the use of concepts of Reliability and Statistics in Transportation systems(TS)has become widespread, as also proven by a recent book issued in 2018 [1]. Key issues which were originated in that framework, such "dependability", "resilience", "vulnerability" and similar ones, have become part of the vast range of concepts used to describe, along the more traditional concept of Quality, the performances of Transportation systems [1-4].

The present paperis based upon some theoretical results derived from stochastic process theory and risk and safety analysis [5-13],and the Bayes methodology, which has become increasingly adopted in all fields of engineering, especially in Reliability [14-16]. In particular, the paper is focused on the statistical estimation of a suitable "Reliability based" Quality Index (QI) defined with respect in terms of delays suffered from the TS customers because of system faults. The approach is based on the Poisson Process model for delay occurrences, and proposes an estimation method which is, in some cases, independent of the particular PDF of delay amplitudes. This feature is particularly attractive in view of statistical estimation, here performed using a new Bayesian estimation method. In particular, the statistical characterization is based upon a probabilistic model which has been introducedin [12] and whose estimation approach is similar to the one proposed in [13],in a different framework, i.e. a study devoted to define and evaluate a safety of engineering systems subjected to random loads. Then, if the concept of safety is substituted by the one of quality, and the random loads are meant as the random delays, the quality can be approached in a way which is analogous to the one used for safety, thus taking advantage of the recent development in such field. The proposed quality index (QI) is defined as the probability that a given delay amplitude (DA) is never exceeded in a given time interval. The proposed QI is intrinsically dynamic, being dependent on the chosen value of time horizon. Once the stochastic model has been established, the fundamental problem arises of making inferences about its parameters, on the basis of observed data.

As well-known [11,15-16], Bayesian inference is rapidly becoming the most widely adopted methodology for estimation purposes, and it is often preferred to classical inference, due to its coherence and flexibility in allowing the use of experimental information and expert judgements on the parameters to be estimated. These data are converted into so-called "prior information", which is conveniently updated in the light of observed data, thus permitting to obtain more efficient estimates with respect to classical inference.In the paper an analytical technique for Bayes estimation based upon suitable prior information is described, characterized by inner robustness and flexible computational implementation, yielding an efficient Bayes estimation of the QI. The remainder of this paper is organized as follows. Section 3 provides an overview of the stochastic modeling of the TS leading to the a.m. QI. A simple case study is also reported at the end of Section 3.Section 4 applies the above said analytical technique for Bayes estimation of the QI.Finally, Section 5illustrates a large set of numerical simulations in order to assess the efficiency of such Bayesian estimation methodology.

## II. STOCHASTIC MODELING FOR RELIABILITY AND QUALITY OF TRANSPORTATION SYSTEMS

The liberalized approach that the transportation systems management has undergone recently has stimulated a big interest in issues related to the reliability, quality and safety of the service offered. These aspects characterize altogether the so-called system "dependability", which is also defined in devoted standards [1-3]. Its definition requires the introduction of the probabilistic risk and safety concept and of methodologies suitable for their evaluation. The systematic procedure proposed by the recent European legislation on the trustworthiness of constrained guide transport systems provides for the classification of dangerous events according to the severity of the possible consequences of their occurrences, along with their probability of occurrence. The combination of these two aspects defines the risk and/or safety related to the single event examined [5-13]. As here discussed, also system quality can be defined in a way which resembles the safety methodology. It should be noted, however, that while the definitions and methods for calculating reliability (and availability), as well as risk and safety, are well established, the topics related to quality have so far been addressed, even in the aforementioned standards, from a qualitative rather than quantitative point of view. In this perspective, the present work is focused on the analytical link between these quantities, with the specific aim to propose a Bayes estimation method for a new quality index that take into account the temporal evolution of the reliability features of the TS. First, it is considered appropriate, also for the purpose of clarifying the terms used below, to recall some fundamental concepts and properties related to the reliability parameters of a renewable system.

In a technological reality in which the reliability features of the systems become increasingly important, the conventional criteria for the determination of RAM (reliability, availability and maintainability) standards for the TS must undergo - in the light of what has been said before - a careful critical examination. In fact, they are based on a traditional approach to reliability, and in particular of a "static" nature. For example, a typical reliability constraint required for such systems is that on the maximum number of faults in the time unit (or per km). However, this parameter uniquely identifies the reliability function only in the Exponential case; yet, only in the latter case it is the actual "fault frequency" of the system. In the general case, the latter is a function of time. A certain imprecise in the definition of the reliability parameters in the TS field has been highlighted in the literature, deriving from inadequate regulation and insufficient tradition in the use of RAM methodologies, as discussedin [12], a discussion briefly summarized here. In particular, [12] highlights the confusion that often occurs between the following quantities, used as identical (and identified with the term "failure rate"):

- the "fault frequency" , that is the reciprocal of the MTTF (average time to failure);
- the "failure rate" function, which is the limit of a (conditional) probability of failure;
- the "failure intensity" function, i.e. the (instantaneous) frequency of faults.

Only the first of these parameters is a constant, the other two being functions of time that express the temporal evolution of the "aging" of the system, as discussed below. In the following, the "reliability function" of a given system in the time interval (0, t) is indicated with R (t). As well known [16],
this function is defined as the probability of the event (T> t), where T is the random variable (RV)) "operating time" of the system. The mean value of T- or the MTTF of the system - is indicated with m: m = E[T]. By definition [1-5], the function "failure rate" is a function of time h(t) which, multiplied by a time interval dt, provides the probability of failure in (t, t+dt), conditioned on the hypothesis that the system has been working up to the instant t. It is known that, if T is continuous, at any time t> 0 in which the probability density function (PDF), f(t), of the operating time T is defined, results:

$$h(t)=f(t)/R(t) \qquad (1)$$

This function is generally - contrary to what is sometimes reported - different from the (average) number of faults in the time unit, i.e. the "fault frequency", both in steady state and in the transient period. This may be better understood if the study is performed with respect to a succession of RV, i.e. the times of operation between two successive failures, i.e. in terms of the random "counting" process of failures [5-8,17]. This process, indicated with N(t), represents the number of faults in (0, t), for a system that starts working in t = 0, in the event of negligible repair times and of time between independent failures and identically distributed; the statistical mean of N (t) - called "renewal function" - is a function of time, indicated by M(t). The temporal derivative of this function, m(t), represents the (instantaneous) frequency of faults, calculated in the instant t. It, called "failure intensity", can be derived - in terms of Laplace-transforms - from the relation:

$$m^*(s) = f^*(s)/[1-f^*(s)] \qquad (2)$$

where $m^*(s)$ and $f^*(s)$ are the Laplace-transforms of the function m(t) and of the PDF f(t) of the time between two faults, T, of the system under examination [5]. The functions h(t) and m(t) are the same if and only if the distribution of the time to the fault turns out to be of Exponential type, in which case they are both constant, and equal to the reciprocal of the MTTF. Omitting details [5,12], in this casethe failure process N (t) follows a Poisson distribution, with parameter ϕ coincident with the three parameters mentioned above. The well-known Poisson probability law *p(k,t)* given by [5,17]:

$$p(k,t) \equiv P\left[N(t)=k\right] = e^{-\varphi t}\frac{(\varphi t)^k}{k!} \;, \qquad k=0,1,...,\infty \qquad (3)$$

In (1),$\phi$is the mean number of faults in the unit time. The mean and variance of the process N(t) are numerically equal and given by:

$$\mathrm{E[N(t)]} = \mathrm{Var[N(t)]} = \phi t \qquad (4)$$

Then, the dynamic index for quality service assessment of electrified transportation systems proposed in [12] is recalled hereafter. As previously mentioned, the reliability assessment is the major premise for facing the heterogeneous problem of the evaluation of the quality service of a generic large complex system. It is customary that the quality be regarded as a function of the customer demand of the public transportation system. More specifically, the customer satisfaction is defined in terms of service "time punctuality". At this purpose it is useful to introduce the quality concept analogously to the safety concept [5], conceived as immunity level of the system with respect to the delays. By this kind of approach, quality can be evaluated – as far as possible - in a quantitative way, avoiding the phraseological concepts which often leave the practitioner unaware whether a satisfying level of quality has been reached or not.

As well-known [5], the most widespread definition of safety, with reference to a given time horizon, is the probability that the delay is not higher than a prefixed level. In other terms the system can be considered safe if harmful events (which,in our case, are uniquely system's faults) do not occur, or if in case of fault the consequent delay is acceptable, i.e. without compromising the system safety.

Let us supposed to know (or estimate) the delay distribution related to each system's fault, i.e. the (conditional) cumulative distributionfunction (CDF) of the delay W, defined as:

$$F(w) = P(W<w) \quad (5)$$

in which w is a "waiting time". Let us focus our attention on the delay amplitude occurring at time $T_k$:such amplitude is a random variable, here indicated as $W_k$. The proposed quality function is defined in order to count the number of RV $W_k$which exceed a givenextreme value *z*of delay. This can be accomplished by associating to the stochastic process N(t) and the random variables $W_k$ ($k=1,2,\dots,N_b(t)$), the following stochastic process M(t) defined in terms of thesuccession of RV $I_k$ , each being a Bernoulli RV denoting the event ($W_k> z$), i.e.:

$$\begin{cases} I_k = 1 & \text{if } (W_k > z) \\ I_k = 0 & \text{otherwise} \end{cases} \quad (6)$$

Then, let us focus our attention only on the delay amplitudeslarger than z: these definethe following stochastic process:

$$\begin{cases} M(t) = [I_1 + I_2, \ldots + I_{N(t)}], \text{ if } N(t)>0 \\ M(t) = 0, \text{ otherwise} \end{cases} \quad (7)$$

In other words, M(t) is arrival process that counts the DA exceeding the "barrier value"z over (0,t).

In the present case of a Poisson process for N(t), let us also assume that the RV $W_k$ are assumed to be statistically independent and identically distributed with the common, time-independent, CDF of (5):

$$F(w) = F_W(w) = P(W_k \le w), \ \forall\, k = 1,2,\ldots\ldots n \quad (8)$$

Under the above hypotheses, by means of simple hypothesis holds, with the following new value of the mean frequency: $\omega = \phi[1-F(z)]$, where$\phi$ is the mean frequency of the base counting process N(t) (see (4)). So, the probabilistic distribution of M(t), i.e. the probability that a number k of extreme PLA occurs in the time interval (0,t) is:

$$p_M(k,t) \equiv P[M(t)=k] = e^{-\phi q t} \cdot \frac{(\phi q t)^k}{k!} \quad (9)$$

defining:

$\phi$ = *mean fault frequency (i.e., expected number of faults occurrences for unit time);* (10)

q= 1 - F(z) = P($W_j$>z) = *exceedance probability (EP) of the value z by any single DA* $W_j$. (11)

The proposed quality index (QI), G(t,z), is defined as the probability that z is never exceeded over (0,t): it is obtained as a particular case (k=0) of eqn. (10), i.e. the probability that no exceedance occurs in (0,t).

$$G(t,z) = e^{-\varphi t[1-F(z)]} \quad (12)$$

In the following, eq. (12) will sometimes re-written, omitting the fixed parameter z, in the following form which of course is identical to (13):

$$G(t) = e^{-\phi q t} \quad (13)$$

Another meaning of the QI may be introduced, if one definesthe following succession of RV

Y(t) = max ($W_1, W_2, \ldots, W_N$) , *if* N(t)= N>*0* (14)

Y(t) = *0, otherwise* (15)

It is readily shown that:

G(t, z) = P[Y(t) < z ] (16)

In other words, G(t, z) is the probability that the maximum delay never exceeds z over (0,t), which is consistent with theintuition. As a case study, it is reported a simple example whose figures are deduced from [15], which deals with the estimation of the rate of occurrence of failures related to the train system within the UK. For a certain kind of failure, 66 events were observed in the last 6 years, leading to an estimate of $\phi = 66/6$ = 11/year = 0.92/month. Assuming, as an example,that the DA are Exponential RV with mean value E[W]= 30 min., the CDF F(z) of (12) is expressed by (17), in which β = 1/E[W]=1/30 (min.$^{-1}$):

$$F(z) = 1 - e^{-\beta z}, \quad z > 0 \qquad (17)$$

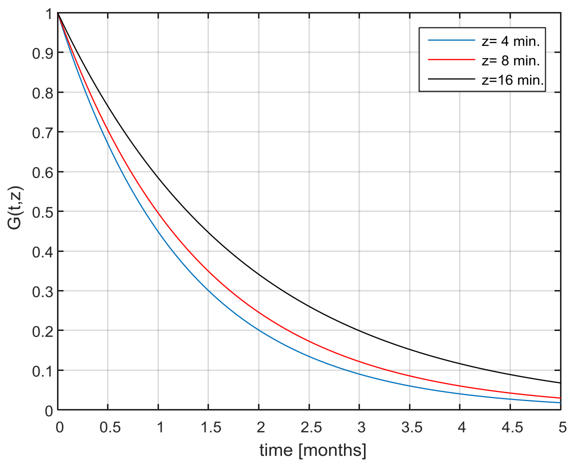


Fig. 1 The QI, G(t, z), vs. time t in months in the case study with $\phi = 0.92$/month and z=4,8,16 min. corrisponding to the three curves shown, from the lowest one (z =4, min.) to the highest one (z =16 min.).

In Fig.1, three curvesof the QI of (12) are depicted for such case, as a function of time t in moths, and three different values of delay z (here in minutes). Of course – as well apparent from the general expression of (12) - the QI always decreases with time, while it increases if z increases (accepting a greater maximum delay z implies increasing the probability that such delay is not exceeded).

## IV. A BAYES INFERENCE METHOD FOR THE STOCHASTIC QUALITY INDEX OF TRANSPORTATION SYSTEMS

Two Bayes inference methods for the QI estimation are proposed here,both by means of the "*Negative Log-Gamma*" (NLG) distribution, according to the kind of information available on the DAdistribution.The NLG PDF is the PDF of a RV in (0,1), defined by S=exp(-X), where X has a Gamma distribution [16-18]. It is recalled that the PDF of Gamma distribution with positive parameters ν and δ (shape and scale parameter respectively), is expressed, for x>0[1], by:

$$f(x) = \frac{1}{\delta^{\nu}\Gamma(\nu)} x^{\nu-1} \exp\left(-\frac{x}{\delta}\right) \qquad (18)$$

Such PDF will be denoted as *gampdf* (x,ν,δ). The PDF ofS=exp(-X), by ordinary rules of RV transformations [16-18] is expressed by ("log" denotes natural logarithm):

$$p(s) = \frac{1}{\Gamma(\nu)\delta^{\nu}} s^{(1/\delta)-1} \left(-\log s\right)^{\nu-1}, \quad 0 < s < 1 \qquad (19)$$

Bayes Inference allows integration of two sources of knowledge: experimental data from the field and prior knowledge, which in practice always exists in TS applications. The two sources are integrated by adopting Bayes theorem [15,16]. For the purpose of present application, the Bayesian estimation of the above QI can be carried out in a relatively straightforward way, here illustrated, developing – with some innovations - a methodology already adopted for a safety index in [13], as coherent with the theoretical approach followed in the present study.In the Bayesian setting, the (generally unknown) parameters (q,ϕ) are considered as RV, so they are denoted here by capital letters: Q and Φ (except when they are arguments of functions). So, the starting point for the requested estimation process is a joint prior PDF *g(q,ϕ)* for the parameters Q and Φ. The information conveyed in such PDF can be integrated and updated with field data – denoted by D – by the below reported Bayes' theorem:

$$g(q,\phi|D) = g(q,\phi)L(D|q,\phi)/C \qquad (20)$$

where:
*L(D|q,ϕ)* denotes the Likelihood of the data D conditional to the parameters (q,ϕ);*C* is a constant (with respect to the parameter values):

$$C = \int_0^{+\infty}\int_0^{+\infty} g(q,\phi)L(D|q,\phi)\,dq\,d\phi \qquad (21)$$

As well known, the best Bayes estimate - in the mean square error sense - of a given function H=H(Q, Φ) is given by the posterior mean E[H(Q, Φ)|D], obtained by integrating the product of H with the posterior PDF of (19). First, we will introduce the case in which the DA distributionis known, and then we will briefly recall a more general methodology which can be applied, following [13], when the DA distributionis unknown.

[1]The PDF is assumed zero for negative values of x.

*- Known Delay Distribution*

Let the DA distributionbe known, i.e. the EP: q = P($W_j$> z) = 1 - F(z), being F(z) the CDF of any RV $W_j$, is known. For sake of illustration, let the CDF of the DA be an Exponential one (but the method here illustrated is in principle extensible to any kind of CDF), i.e.:

$$F(z) = 1\text{-}e^{-\beta z}, \quad z > 0 \qquad (22)$$

Being β a positive knownconstant, such as the mean DA is E[W]=1/β. So, q=$e^{-\beta z}$, and the QI is expressed as:

$$G(t,z) = e^{-\phi t e^{-\beta z}} \qquad (23)$$

In such case, the only unknown parameter, to which assign a prior PDF, is the positive parameter$\phi$, mean fault frequency.

This can be done in two ways. If the process of faults is observable, and this implies that the Exponential times between failures $T_j$are observable, then the Bayesian estimation of $\phi$ , and thus of the above QI is carried out in a straightforward way adopting the well-known "conjugate" Gamma prior PDF for $\phi$, by trivial computations [16].

If, instead, available data are only those regarding the EP, i.e. the occurrences of DA, the estimation of the QI G(t,z) can be regarded, for a fixed time interval t, as the estimation of the Gumbel [18] CDF G(t,z) seen as a function of z.Indeed, denoting by Y(t) as in(14) themaximum DA, the QI G(t,z) is the CDF of Y(t) evaluated atthe value z of delay, as apparent from (16). Let Y be the generic RV "maximum DA" of a given sample. For purpose of a straightforwardBayes estimation, it is convenient to work with the auxiliary variable X=exp(Y).It is well known [18], and easy to verify, that being Y a Gumbel RV with CDF G($t$,$z$), then X is an "Inverse Weibull" RV with CDF [14,18]:

$$F(x;\phi,\beta) = e^{-\phi t(x)^{-\beta}} \qquad (24)$$

Also in this case, inference on parameter Φ can be performed adopting the conjugate Gamma prior PDF. A more practical, but equivalent approach can be adequately adopted since generally no physical meaning can be usefully exploited in formulating prior knowledge on such parameter. Indeed, as discussed in [14], the analyst does not think in terms of parameters, but expresses his technical knowledge in terms of more practical concepts, such as some value of the CDF of (24), implying a corresponding information for the QI. For these reasons, a practical approach is to use a "*Negative Log- Gamma*" prior CDFof (24), which implies a Gamma prior PDF for the parameter$\phi$[14], as can be deduced from the more general case examined in the following sub-section.

*- Unknown Delay Distribution*

This is a more general case, of which the previous one can be considered a particular subcase, so the mathematics is illustrated here with more details, as meant to illustrate both cases. The Bayes inference here proposed, following [13], uses well known "conjugate" [15-16] priors for the RV Q and Φ, i.e. the Beta prior PDFfor Q and the Gamma prior PDF for Φ. The two RV are moreover assumed to be statistically independent, so that the prior joint PDF $g(q,\phi)$ is given by[2]:

$$g(q,\phi) = betapdf\,(q; r_0, s_0) \cdot gampdf\,(\phi; n_0, \delta_0) \qquad (25)$$

(see Appendix for the definition of the function *betapdf.* Function *gampdf* is defined after (18). According to the Bayes method, the parameters' values ($r_0$,$s_0$,$n_0$,$\delta_0$) are deduced from prior information, i.e. previously collected wind data or experts' judgements. Such prior data are updated by means of field data.As shown in [13], the posterior PDF are easily seen to be again Beta and Gamma, and the product QΦ is still Gamma distributed (or approximately so), so that also the posterior PDF of the QI is again NLG.

## V. COMPARISON OF CLASSICAL AND BAYESIAN ESTIMATION TROUGH NUMERICAL SIMULATION

For brevity, here only the more general case of unknown delay distribution is illustrated.A large set of numerical experiments have been performed by generating simulatedsamples - obtained by means of Monte Carlo simulation [19]- with the purpose of showing the efficiency of the proposed Bayes estimation of the QI.The simulations were conducted for various sample sizes n (number of faults), and various input data values. Data were generated as follows:

- Data on the observed number of faults in a given time interval were generated by a Poisson Process of mean frequency Φ (randomly generated previously according its prior Gamma PDF).

[2] The suffix "0" is used to denote prior PDF parameters. See also [13] for more details

- Data on the observed exceedance number m were generated by a Binomial RV with parameters (n,Q), being also Q randomly generated according its prior Beta PDF.

For each sample size n, a number of $N=10^4$ replications has been effected; in particular, the results for a wide range of sample-sizes (from n=10 to n=120) are reported in Tab. 1, in terms of MSEB (Mean Square Error of the Bayes estimator), MSEL (Mean Square Error of the ML estimator) and REFF (defined as MSEL/MSEB). All above indexes are dimensionless (being the QI to be estimated a probability).The "REFF" index is the Relative Efficiency of the Bayes estimator with respect to the ML estimator. The above "Mean Square Errors" have been obtained at the end of each simulation as the averages over the N sampled estimator's square errors. The above-defined quantities are based on the concept of Mean Square Error (MSE) of an estimator. Given an estimator, ζ °, of the parameter ζ , its theoretical MSE, is – as well known [16-18] - defined as the expected value of $[(\zeta° -\zeta)^2]$. Such quantitywas evaluated at the end of each simulation case study by means of the ordinary large-sample estimator, i.e. the "observed" MSE (briefly, MSE in the sequel), evaluated on the estimated $(\zeta°_j)$and "true" $(\zeta_j)$ values of the parameter ζ of the N simulated samples as:

$$MSE = \frac{1}{N}\sum_{j=1}^{N}(\zeta°_j - \zeta_j)^2 \qquad (26)$$

TABLE I.SIMULATION RESULTS SHOWING THE EFFICIENCY OF THE BAYES ESTIMATOR OF THE QI.

| n | *MSEB* | *MSEL* | *RABB* | *RABL* | *MREB* | *MREL* | $R_1$ | $R_2$ | REFF |
|---|---|---|---|---|---|---|---|---|---|
| 10 | 0.0372 | 0.0832 | 0.2750 | 0.4022 | 2.3479 | 1.730 | 1.463 | 0.730 | **2.237** |
| 50 | 0.0343 | 0.2501 | 0.2632 | 0.5503 | 1.9407 | 7.326 | 2.091 | 3.623 | **7.292** |
| 80 | 0.0302 | 0.2452 | 0.2275 | 0.4448 | 1.9729 | 8.019 | 1.955 | 3.831 | **8.116** |
| 100 | 0.0294 | 0.2097 | 0.2469 | 0.3630 | 1.7987 | 7.058 | 1.470 | 3.908 | **7.133** |
| 120 | 0.0243 | 0.1680 | 0.1912 | 0.3498 | 1.3914 | 9.451 | 1.829 | 6.827 | **6.914** |

**Legenda**: *n= sample size; other terms are defined in the text.*

In the above eqn. and the following ones, $\zeta_j$ (j=1,..N) is the "true" parameter value correspondent to the j-th n-dimensional simulated sample generated by the simulated values of (Q, Φ) for that sample, and $(\zeta°_j)$its Bayes (or ML) estimate.The ratio REFF is the most widely adopted measure of the "relative efficiency" of the Bayes estimator with respect to the ML estimator: the more it exceeds unity, the more efficient is the Bayes estimate when compared with the ML estimate.Other significant quantities here evaluated, in order to assess the performances of the estimates, are:

- RABB: Relative Average Bias of the Bayes estimator;
- RABL: Relative Average Bias of the ML estimator;
- MREB: Maximum Relative Error of the Bayes estimator;
- MREL: Maximum Relative Error of the ML estimator.

The relative bias of an estimator, ζ°, of the parameter ζ can be defined as: (E[ζ°]-E[ζ])/E[ζ]. Therefore, as an overall "sample" measure of the relative bias, the "Relative Average Bias", RAB, is used here, in which the quantities E[ζ°] and E[ζ] are evaluated through their estimated values – i.e. their averages - at the end of each simulation case study.

Similar computations wereperformed for the MRE indexes. Both RAB and MRE are evaluated in terms of absolute values. Also, the following (dimensionless) ratios are reported, as useful synthetic measures in evaluating the "precision" of the Bayes estimates:

R1=RABL/ RABB; R2= MREL/ MREB

Among the many performed simulations, the results shown in table I refer to theafore mentioned available ST data (see end of section 3), so that the prior PDF of this application were chosen as follows: *Φ* has a Gamma PDF with mean $\mu=0.920 month^{-1}$and $\sigma = 0.092 month^{-1}$; (i.e. the SD is 10% if the mean); Q, the EP of a given threshold level z, has a Beta probability density function with $\mu=0.080$ and$\sigma = 0.027$. These values imply the following prior parameters in (25)[3]:

$n_0 = 100; \delta_0 = 0.0092; r_0 = 8; s_0 = 92.$

In Table I, the values of MSEB, MSEL, RABB, RABL, MREB, MREL, R1, R2, REFF, relevant to the Bayes and ML estimates of the QI are reported as a function of the sample size n considered in the simulations. In the first column, the sample size is reported, i.e. the number of (simulated) fault events, n. The reported results point out the efficiency of the proposed Bayesian approach, evidencing that the Bayes estimate errors, both in terms of mean square

[3] A greater uncertainty generally exists on Q: such values imply indeed that the its SD is about 33% if the mean. In terms of prior percentiles: the 0.05 and 0.95 percentiles of Q are: 0.04 and 0.13, so that: P(0.04<Q<0.13)= 0.90, a large "credibility interval"

error (as measured by MSEB), and of relative errors (RABB, MREB) are reasonably limited. This last aspect, regarding the "precision" of the Bayes estimates, is shown also by the relative indices R1 and R2, always larger than 1 except for a case with small sample size (n=10). The relative efficiency with respect to the ML estimate is always larger than 1, not only for small sample sizes as generally happens [16], so that the ML estimates are outperformed by the Bayes ones.

## VII. CONCLUSIONS

In the paper, a stochastic model is adopted in order to describe and estimate reliability and quality assessment of TS. The model has two basic features: 1) as a probabilistic model, it exploits results from stochastic process theoryrelatedto safety analyses; 2) from the point of view of statistical estimation, it assumes no particular probability model for the delay amplitude, thus proving to be "robust". Bayesian estimation of the above QI is the novel feature of the paper, based upon Gamma and Beta prior distributions respectively for fault frequency and "exceedance probability", leading to a Negative Log-Gamma distribution for the QI. A large set of Monte Carlo numerical experiments have been performed, yielding excellent results both in terms of efficiency and precision, with reference to very different sample sizes. Further studies seem opportune to highlight the TS system reliability and quality, also in terms of "capacity"[22-26], as well as exploring the use of the empirical Bayes method, which has found recent railway application [15].

## APPENDIX – THE BETA PDF

The Beta PDF [16-18] is defined as follows for the argument $q$ in the range *(0,1)*, which is indeed a probability here:

$$\text{betapdf}\,(q;a,b)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}q^{a-1}(1-q)^{b-1},\ 0<q<1 \qquad \text{(A.1)}$$

## REFERENCES

[1] I. Kabashkin and I. Yatskiv, O. Prentkovskis,"Reliability and Statistics in Transportation and Communication",*17th International Conference RelStat'17*,18-21 October, 2017, Riga, LatviaSpringer, 23 gen 2018

[2] M. G. H. Bell and C. Cassir, Reliability of Transport Networks. Baldock, Hertfordshire, U.K.: Research Studies Press, 2000.

[3] L.G. Mattsson, E. Jenelius, "Vulnerability and resilience of transport systems – A discussion of recent research", *Transportation Research Part A*, Vol. 81, pp. 16–34, 2015.

[4] A. Chen, H. Yang, H. K. Lo and W. H. Tang, "Capacity reliability of a road network: an assessment methodology and numerical results", *Transportation Research Part B: Methodological*, Vol. 36, Issue 3, pp. 225-252, 2002

[5] W. A. Thompson, Point Process Models with Applications to Safety and Reliability, Chapman and Hall, London, 1988.

[6] V.V. Krivtsov, "Recent advances in theory and applications of stochastic point process models in reliability engineering",*Reliability Engineering and System Safety*, Vol. 92, pp.549 -551, 2007

[7] T. Aven , Risk management. Berlin, Heidelberg: Springer; 2010

[8] T. Aven, "On some recent definitions and analysis frameworks for risk, vulnerability, and resilience", *Risk Analysis.*, vol. 31, no. 4, pp. 515–522, 2011.

[9] T. Aven andE. Zio, "Foundational issues in risk assessment and risk management",*Risk Analysis*, Vol. 34, Issue:7, pp.1164–1172, 2014

[10] T. Aven, "Risk assessment and risk management: Review of recent advances on their foundation", *European Journal of Operational Research*, Vol. 253, Issue:1, pp.1-13, 2016.

[11] M. Kalantarnia, F. Khanand K. Hawboldt, "Dynamic risk assessment using failure assessment and Bayesian theory", *Journal of Loss Prevention in the Process Industries*, vol. 22, no. 5, pp. 600-606, 2009.

[12] E. Chiodo and D. Lauria, "Stochastic index definition and estimation for reliability and quality assessment of transportation systems",*2012 Electrical Systems for Aircraft, Railway and Ship Propulsion*, Bologna, 2012, pp. 1-5.

[13] E. Chiodo and D. Lauria, "Probabilistic description and prediction of electric peak power demand", *2012 Electrical Systems for Aircraft, Railway and Ship Propulsion*, Bologna, 2012, pp. 1-7.

[14] R. Calabria and G. Pulcini, "Bayes 2-sample prediction for the inverse weibull distribution",*Communications in Statistics - Theory and Methods*, vol. 23, n.6, pp. 1811-1824, 1994.

[15] J. Quigley, T. Bedford and L. Walls,"Estimating

rate of occurrence of rare events with empirical Bayes: a railway application", *Reliability Engineering & System Safety*, Vol.92, Issue:5, pp. 619–627,2007.

[16] H. F. Martz and R. A. Waller, Bayesian Reliability Analysis, Krieger Publishing, Malabar, USA, 1991.

[17] A. Papoulis, Probability, Random Variables, Stochastic Processes, Mc Graw Hill, 2002.

[18] N. L. Johnson, S. Kotz, N. Balakrishnan. "Continuous Univariate Distributions", J. Wiley & Sons, 2002.

[19] M. H. Kalos and P. A. Whitlock, Monte Carlo methods, John Wiley & Sons,2008.

[20] A. Chen, H. Yang, H.K. Lo and W.Tang, "A capacity related reliability for transportation networks", *Journal of Advanced Transportation* , 1999,Vol. 33, pp. 183– 200.

[21] Y. Tyrinopoulos and C. Antoniou. "Public transit user satisfaction: Variability and policy implications.",*Transport Policy*, Vol:15, Issue:4, pp.260-272, 2008.

[22] J. W. C. van Lintand H. J. van Zuylen, "Monitoring and predicting freeway travel time reliability: Using width and skew of day-to-day travel time distribution",*Transp. Res. Rec.*, Vol. 1917, pp. 54-62, 2005.

[23] H. Al-Deek and E. B. Emam, "New methodology for estimating reliability in transportation networks with degraded link capacities", *Journal of Intelligent Transportation Systems*, Vol. 10, no. 3, pp. 117-129, 2006.

[24] M.A. Taylor, "Critical Transport Infrastructure in Urban Areas: Impacts of Traffic Incidents Assessed Using Accessibility-Based Network Vulnerability Analysis",*Growth and Change*, Vol.:39, Issue:4, pp. 593-616, 2008.

[25] M. A. Taylor and G. M. D'Este, "Transport network ulnerability: a method for diagnosis of critical locations in transport infrastructure systems", *Critical infrastructure*, pp. 9-30,2007, Springer, Berlin, Heidelberg.

[26] M. Sharov and A. Mikhailov, "Urban Transport System Reliability Indicators",*Transportation Research Procedia*,Vol. 20, pp. 591-595, 2017.