The empirical failure rate of discrete reliability data

Cyril Bracquemond, Olivier Gaudoin
Institut National Polytechnique de Grenoble
Laboratoire IMAG-LMC
BP 53 - 38041 Grenoble Cedex 9
France
Cyril.Bracquemond@imag.fr, Olivier.Gaudoin@imag.fr

Abstract

This paper proposes an empirical estimator of the failure rate of discrete lifetimes. The exact distribution of the estimator is given. Asymptotic results are proved from which asymptotic confidence intervals and confidence bands for the discrete failure rate are derived.

1 Discrete lifetimes and discrete failure rate

Discrete distributions are used in reliability when lifetime measurements are taken in discrete time. This is the case for example when an equipment operates on demand and the observation consists in the number of demands successfully completed before failure, or when continuous lifetimes are grouped.

Let the random variable K be a discrete system lifetime. K is defined over the set of positive integers $\mathbb{N}^*$. Let $p(k) = P(K = k)$ be the probability that the system fails at time k. The reliability function is given by the probability that the system is still alive at time k : $R(k) = P(K > k)$. The cumulative distribution function of K is $F(k) = P(K \leq k)$.

The failure rate of a discrete distribution has been defined by Barlow-Marshall-Proschan (1963) as:

$$\forall k \in \mathbb{N}^*, \qquad \lambda(k) = P(K = k | K \geq k) = \frac{p(k)}{R(k-1)} \tag{1}$$

Several discrete reliability models have been proposed with parametric expressions of the failure rate. For example, a constant failure rate leads to the geometric distribution. Nakagawa-Osaki (1975) defined the discrete Weibull distribution. Other parametric models can be found in Salvia-Bollinger (1982), Gupta-Gupta-Tripathi (1997), Roy-Gupta (1999) and Bracquemond (2001).

We are interested here in the nonparametric estimation of λ. Although an extensive litterature is available for the nonparametric estimation of continuous failure rates (see for example Klein-Moeschberger 1997), it seems that rather few work have been done in the discrete case.

2 The discrete empirical failure rate

Let $K_1, \ldots, K_n$ be a sample of n iid random variables on $\mathbb{N}^*$, which can be understood as discrete lifetimes of identical and independent systems.

The empirical frequencies are defined as : $\forall k \in \mathbb{N}^*, \mathbb{P}_n(k) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{K_i = k\}}$, and the empirical reliability as : $\forall k \in \mathbb{N}^*, \mathbb{R}_n(k) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{K_i > k\}}$.

Then, it is natural to define the *discrete empirical failure rate* as :

$$\forall k \leq K_n^* = \max(K_1, \ldots, K_n), \quad \lambda_n(k) = \frac{\mathbb{P}_n(k)}{\mathbb{R}_n(k-1)} \tag{2}$$

The aim of this paper is to present some results on this empirical failure rate. The proofs of these results can be found in Bracquemond (2001).

Note that λ_n is not defined for $k > K_n^*$, since both $\mathbb{P}_n(k)$ and $\mathbb{R}_n(k-1)$ are null for these values of k. If k is not equal to an observation, $\lambda_n(k) = 0$. Moreover, $\lambda_n(K_n^*) = 1$.

When the number of data is too small, $\lambda_n(k)$ happens to be often equal to zero. Then, it is difficult to assess the sense of variation of the failure rate. Since this assessment is very useful in terms of wear-out or burn-in of the considered systems, the empirical failure rate will be of practical use only for reasonnably large data sets.

3 Exact distribution

$\lambda_n(k)$ is of the form i/m, where $0 \leq i \leq m \leq n$ and $m > 0$. By studying all possible cases for this ratio, the exact distribution of $\lambda_n(k)$ can be derived for fixed n and k :

Proposition 1 : *$\forall k \leq K_n^*$, the distribution of $\lambda_n(k)$ is given by :*

- $\mathbb{P}(\lambda_n(k) = 0) = (1 - p(k))^n$
- *For $1 \leq i \leq m \leq n$, where i and m are relatively prime :*
  $\mathbb{P}(\lambda_n(k) = \frac{i}{m}) = [F(k-1)]^n \sum_{j=1}^{\lfloor \frac{n}{m} \rfloor} \frac{n!}{(n-jm)!(ji)!(jm-ji)!} \left[\frac{R(k)^{m-i} p(k)^i}{F(k-1)^m}\right]^j$,
  where $\lfloor x \rfloor$ denotes the integer part of x.

Then, it is possible to compute the moments of this distribution. The first two are :

Proposition 2 :

$$\begin{aligned}
\mathbb{E}[\lambda_n(k)] &= \lambda(k)\left[1 - F(k-1)^n\right] \\
\mathbb{E}\left[\lambda_n^2(k)\right] &= F(k-1)^n p(k) \sum_{m=1}^{n} \frac{C_n^m}{mF(k-1)^m} \left[p(k)(m-1)\left[R(k)+p(k)\right]^{m-2} + \left[R(k)+p(k)\right]^{m-1}\right]
\end{aligned}$$

So $\lambda_n(k)$ is a biased but asymptotically unbiased estimator of $\lambda(k)$. Since $\mathbb{E}[\lambda_n(k)] \leq \lambda(k)$, the failure rate is underestimated. This is probably due to the fact that $\mathbb{P}(\lambda_n(k) = 0) \neq 0$. The bias grows as k grows. In particular, $\lambda_n(K_n^*) = 1$ traduces a high bias. Then, for small sample sizes, the estimation will be of interest only for small values of k.

4 Asymptotic confidence intervals

First of all, for fixed k, the almost sure convergence of $\lambda_n(k)$ to $\lambda(k)$ is easily proved with the law of large numbers.

Then, in order to study the asymptotic normality of $\lambda_n(k)$, let us introduce the empirical process associated to the failure rate, $\mathbb{T}_n$, defined by :

$$\forall k \in \mathbb{N}^*, \quad \mathbb{T}_n(k) = \sqrt{n}[\lambda_n(k) - \lambda(k)]. \tag{3}$$

With the usual convergence results on the empirical reliability $\mathbb{R}_n$, it is possible to derive the asymptotic distribution of $\mathbb{T}_n(k)$:

Proposition 3 : *$\forall k \in \mathbb{N}^*$, $\mathbb{T}_n(k)$ converges in distribution to the normal distribution with mean zero and variance* $\Pi(k) = \frac{\lambda(k)(1-\lambda(k))}{R(k-1)} = \frac{R(k)p(k)}{R(k-1)^3}$.

Let $\widetilde{\Pi}_n(k) = \frac{\mathbb{R}_n(k)\mathbb{P}_n(k)}{\mathbb{R}_n(k-1)^3}$ be the empirical estimator of $\Pi(k)$. Then, an asymptotic $100(1-\alpha)\%$ confidence interval for $\lambda(k)$ is :

$$\left[\lambda_n(k) - z_{1-\alpha/2}\sqrt{\frac{\widetilde{\Pi}_n(k)}{n}}, \lambda_n(k) + z_{1-\alpha/2}\sqrt{\frac{\widetilde{\Pi}_n(k)}{n}}\right] \tag{4}$$

where z_α is the α-quantile of the standard normal distribution.

However, the confidence bounds can be out of $[0,1]$. To avoid this problem, it is interesting to use the logit transformation: $logit(p) = \ln \frac{p}{1-p}$. With the logit transformation, the convergence result becomes:

Proposition 4 : *$\forall k \in \mathbb{N}^*$, $\sqrt{n\lambda_n(k)\mathbb{R}_n(k)}\,(logit[\lambda_n(k)] - logit[\lambda(k)])$ converges in distribution to the standard normal distribution.*

So another asymptotic $100(1-\alpha)\%$ confidence interval for $\lambda(k)$ is :

$$\left[\frac{\lambda_n(k)}{\lambda_n(k) + (1-\lambda_n(k))exp\left[\frac{z_{1-\alpha/2}}{\sqrt{\lambda_n(k)\mathbb{R}_n(k)}}\right]}, \frac{\lambda_n(k)}{\lambda_n(k) + (1-\lambda_n(k))exp\left[-\frac{z_{1-\alpha/2}}{\sqrt{\lambda_n(k)\mathbb{R}_n(k)}}\right]} \right] \tag{5}$$

Figure 1 shows intervals (4) and (5) with $\alpha = 10\%$ for discrete data from Xie-Goh (1993). It is clear that logit intervals should be used since their bounds are in $[0,1]$ and they are narrower than the first intervals.

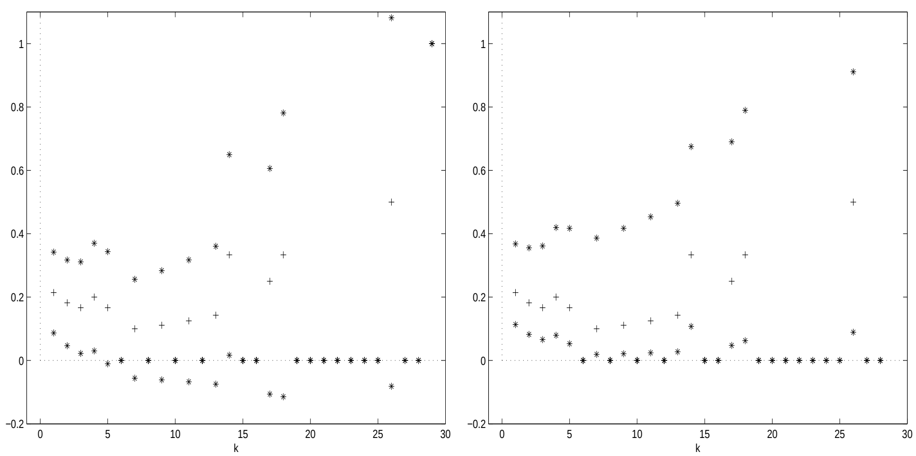

Figure 1: Asymptotic confidence intervals for $\lambda(k)$, without and with the logit transformation

5 Asymptotic simultaneous confidence bands

Finally, the convergence of fini-dimensional vectors of the empirical process $\mathbb{T}_n$ is studied. The following result is proved using the convergence results on the empirical reliability :

Proposition 5 : *For all integer $m \geq 1$ and all sequence $(x_1, \ldots, x_m)$ of m ordered integers, the random vector $(\mathbb{T}_n(x_j) \,:\, 1 \leq j \leq m)^t$ converges in distribution to a random gaussian vector $\underline{\mathbb{T}}^t$ in $\mathbb{R}^m$, with mean zero and covariance matrix Π defined by :*

$$\begin{cases} \Pi(j,k) & = 0 \text{ for } j \neq k \\ \Pi(k,k) & = \Pi(k) \end{cases}$$

Since Π is a diagonal matrix, asymptotic simultaneous confidence bands for λ can be built. If a and b are two integers such that $a < b$, let $Z^n_{a,b} = \max_{\{a \leq k \leq b\}} \sqrt{n}\frac{\lambda_n(k) - \lambda(k)}{\sqrt{\widetilde{\Pi}_n(k)}}$. $Z^n_{a,b}$ converges in distribution to the distribution of the maximum of $b - a + 1$ independent standard normal random variables.

Then, an asymptotic $100(1-\alpha)\%$ confidence band for λ when $k \in [a, b]$, is :

$$\left[\lambda_n(k) - q_\alpha\sqrt{\frac{\widetilde{\Pi}_n(k)}{n}}, \lambda_n(k) + q_\alpha\sqrt{\frac{\widetilde{\Pi}_n(k)}{n}}\right] \tag{6}$$

where q_α is such that $\phi^{b-a+1}(q_\alpha) - [1-\phi(q_\alpha)]^{b-a+1} = 1-\alpha$, and ϕ is the cdf of the standard normal distribution.

Figure 2 presents this asymptotic confidence band with $\alpha = 10\%$ for data in Xie-Goh (1993) and $k \in [1, 10]$. This band is too large to be of practical use. This is not a surprise since the studied sample has a too small size ($n = 28$).

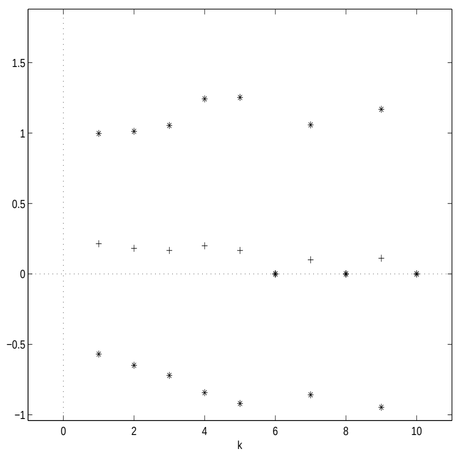

Figure 2: Asymptotic simultaneous confidence band for $\lambda(k)$

References

Barlow, R.E., Marshall, A.W. and Proschan, F. (1963). Properties of probability distributions with monotone hazard rate. *Annals of Mathematical Statistics 34*, 375-389.

Bracquemond, C. (2001). *Modélisation stochastique du vieillissement en temps discret.* PhD Thesis, INP Grenoble (in French).

Gupta, P.L., Gupta, R.C. and Tripathi, R.C. (1997). On the monotonic properties of the discrete failure rate. *Journal of Statistical Planning and Inference 65*, 255–268.

Klein, J.P. and Moeschberger, M.L. (1997). *Survival Analysis : Techniques for Censored and Truncated Data.* Springer.

Nakagawa, T. and Osaki, S. (1975). The discrete Weibull distribution. *IEEE Transactions on Reliability R-24*(5), 300–301.

Roy, D. and Gupta, R.P. (1999). Characterizations and model selections through reliability measures in the discrete case. *Statistics and Probability Letters 43*, 197–206.

Salvia, A.A. and Bollinger, R.C. (1982). On discrete hazard functions. *IEEE Transactions on Reliability R-31*(5), 458–459.

Xie, M. and Goh, T.N. (1993). SPC of a nearzero defect process subject to random shock. *Quality and Reliability Engineering International 9*, 89–93.