DENSITY AND FAILURE RATE ESTIMATION WITH APPLICATION TO RELIABILITY

Contribution to the Encyclopedia of Statistics in Quality and Reliability

Article ID: eqr449

February 26, 2007

Hans-Georg Müller and Jane-Ling Wang

Department of Statistics

University of California, Davis

One Shields Ave., Davis, CA 95616, USA.

e-mail: mueller@wald.ucdavis.edu

Key words: Aalen Estimator, Failure Rate, Hazard Rate, Histogram, Kernel estimator, Local Linear Fitting, Nonparametric Estimation, Smoothing

Abstract

Density and failure rate estimation are valuable tools to assess and explore the occurrence and timing of failures in reliability and quality control. This article focuses on nonparametric approaches such as histogram, kernel method and other smoothing procedures. Nonparametric density and failure rate estimation is particularly valuable for exploratory data analysis and in situations where available information is insufficient to specify a parametric model, as these methods “let the data speak for themselves”, and no assumptions are needed beyond smoothness of the functions to be estimated. Issues specific for nonparametric approaches are finite bias behavior and choice of the necessary smoothing parameter. For failure rate estimation, available data are sometimes incompletely observed (censored) or available in aggregated form only, necessitating appropriate adjustments. Shape restrictions such as increasing failure rate (IFR) are also occasionally of interest in reliability.

1. INTRODUCTION

In reliability and quality control, density and failure rate (or hazard) functions provide valuable information about the distribution of failure times. Shapes of density and failure rate functions may contain characteristics of interest, such as number, location and features of modes, monotone increasing or decreasing features, and information about tail behavior. One can distinguish two basic statistical approaches for estimating density and failure rate functions, given a sample of observed failure times: Parametric and nonparametric.

In the parametric approach to estimation and model fitting, the starting point is to assume an underlying parametric model which specifies a distribution of the observed data, typically a lifetime distribution such as Weibull, Gompertz or Gamma. The parameters are then typically identified by maximum likelihood (see eqr 237) or a Bayesian method (see eqr081).

The parameters determine the distribution uniquely, and thus also density and failure rate functions. The main advantage of the parametric approach, which we will not discuss further in this article, is ease of inference, efficiency and the absence of a bias issue, provided that the underlying parametric model assumption is correct; the latter assumption requires a leap of faith in many applied situations. In fact the main disadvantage of parametric models is that indeed the underlying distribution must be fully known: Under model mis-specification, parametric approaches result in inconsistent estimators, and the associated asymptotic bias leads to invalid inference.

The nonparametric or smoothing approach is considerably more flexible. Nothing more than basic smoothness assumptions on the distribution function are needed. Therefore this approach "lets the data speak for themselves", without imposing hard-to-verify model assumptions. It is thus particularly well suited for exploratory data analysis, and is also used for final analysis in those situations where density functions or failure rates cannot be clearly specified, or when doubts remain about these specifications. The main disadvantages of nonparametric smoothing methods are that inference, often through confidence bands, is not straightforward, and that a smoothing parameter needs to be specified. While these methods are asymptotically unbiased under weak assumptions, their application often requires to be aware of and deal with a finite bias issue.

As was already pointed out in Rosenblatt's seminal 1956 paper, nonparametric density estimates are finitely biased [1]. In parametric approaches, potential bias issues are mostly addressed indirectly through goodness-of-fit checking, and if such bias is detected, the model needs to be changed, often through a data transformation. One can argue that the explicit nature of the bias issue in nonparametric smoothing approaches is actually an advantage, as this bias is small, can be adjusted for and disappears asymptotically. While parametric models rarely fit exactly, the resulting bias is often not adequately addressed. For the multivariate case, the so-called curse of dimension leads to rapidly declining convergence rates of nonparametric methods as dimension increases. This has led to the development of various dimension reduction and semiparametric approaches [2]. In such models, smooth densities are included for low-dimensional parts of the model, while dimension reduction is achieved through a parametric part of the model.

2. NONPARAMETRIC DENSITY ESTIMATION

We begin with basic definitions and relations. Denoting a non-negative random variable that represents failure time or lifetime by T, the survival function is

$$\bar{F}(t) = P\left(T \geq t\right), \qquad t \geq 0, \tag{1}$$

where $F(t) = 1 - F(t)$ is the cumulative distribution function of the lifetime distribution. If this function is differentiable, we may define the probability density function

$$f(t) = \lim_{\Delta \to 0} \frac{1}{\Delta} P\left(t \leq T < t + \Delta\right), \qquad t \geq 0, \tag{2}$$

and the hazard or failure rate function

$$\lambda(t) = \lim_{\Delta \to 0} \frac{1}{\Delta} P\left(t + \Delta > T \geq t | T \geq t\right) = \frac{f(t)}{\bar{F}(t)}. \tag{3}$$

The importance of the failure rate λ in reliability is due to its interpretation as a risk function. Specifically, $\lambda(t)$ quantifies imminent risk of failure at time t, and is related with the survival function $\bar{F}$ through

$$\bar{F}(t) = \exp[-\int_0^t \lambda(u) du], \quad \lambda(t) = -\frac{d}{dt} \log\left[\bar{F}(t)\right]. \tag{4}$$

Both density and failure rate function characterize the failure time distribution. The transformations from density to failure rate and vice versa are as follows [3]:

$$\lambda(t) = \frac{f(t)}{1 - \int_0^t f(u) du}, \quad f(t) = \lambda(t) \exp[-\int_0^t \lambda(u)\, du]. \tag{5}$$

Basic properties of densities f, failure rates λ and the cumulative hazard rate $\Lambda(t) = \int_0^t \lambda(s)\, ds$ are

$$f \geq 0, \quad \int f(u)\, du = 1, \quad \lambda \geq 0, \quad \Lambda(t) \to \infty \text{ as } t \to \infty. \tag{6}$$

Turning to the nonparametric estimation of these functions, historically the first density estimate is the histogram which is still much in use. Other density estimates can be viewed as extensions of the histogram. Besides being of interest in assessing shape, especially symmetry, multimodality and tails of a distribution, density estimates are also useful for a variety of specific statistical tasks, such as the smoothed bootstrap, the construction of confidence intervals of estimated quantiles, e.g., the median, for obtaining Fisher information, or for classification, as centers of clusters may correspond to modes in a density estimate and often multimodality in a density is of interest.

We proceed with a formal definition of the two most common nonparametric density estimates, histograms and kernel estimators, starting with a sample $X_1, ..., X_n$ of i.i.d. univariate data whose distribution has density f. To construct a *histogram*, one divides the range of the data into m bins $B_1, B_2, ..., B_m$ which are usually chosen to be of the same size but can also vary in size, and then obtains the histogram

$$\hat{f}_{\text{hist}}(x) = \frac{1}{n}\sum_{j=1}^{m} \frac{(\text{no. of data } X_i \text{ falling into bin } B_j)}{(\text{size of bin } B_j)} \quad . \tag{7}$$

The smoothing parameters are the leftmost endpoint of the first bin (which matters not so much) and the bin size (which matters). Advantages of the histogram are that it is straightforward to construct and to understand, and its ubiquity in statistical packages. Disadvantages are relatively slow convergence with increasing sample size and the discontinuity of these density estimates which is at variance with the assumed smoothness of the underlying density.

Both disadvantages are remedied by kernel density estimates. These can be conceived as generalizations of sliding histograms, or as convolution of the empirical distribution function with a smooth kernel function,

$$\hat{f}_K(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) = \int \frac{1}{h}K\left(\frac{x-u}{h}\right) dF_n(u), \tag{8}$$

where $h = h(n)$ is a sequence of bandwidths or smoothing parameters, K is a kernel function, and dF_n is the empirical measure, which assigns mass $\frac{1}{n}$ to every observation point X_i. The kernel method for density estimation was introduced by Fix and Hodges [4], Rosenblatt [1] and Parzen [5]. The best source for basic asymptotic and finite sample properties of kernel density estimators remains the monograph of Silverman [6], along with the monograph by Scott [7] for the multivariate case. For general background on kernel methods, see [8].

Typical basic results are rates of convergence of (Integrated) Mean Squared Error (IMSE or MSE) and asymptotic distributions. Optimal rates of MSE convergence are achieved by kernel estimators and are slower than the parametric rate of n^{-1}, due to the nonparametric nature of these estimators. For example, if the underlying density is twice continuously differentiable, the sequence of bandwidths $h = h(n)$ satisfies $h \to 0$ and $nh \to \infty$ as $n \to \infty$, and the kernel satisfies $\int K(x)dx = 1$, $\int K(x)xdx = 0$, and some other basic regularity conditions, then using an IMSE-minimizing bandwidth sequence, the IMSE of the kernel density estimator satisfies

$$\begin{aligned}&\int E\left(\hat{f}_K(x) - f(x)\right)^2 dx \\ &\quad = n^{-4/5}\left[\left(\int f^{(2)}(x)^2 dx\right)\left\{\left(\int K(u)u^2 du\right)\left(\int K(u)^2 du\right)^2\right\}\right]^{1/5} + o\left(n^{-4/5}\right). \end{aligned} \tag{9}$$

The IMSE-rate of convergence of $n^{-4/5}$ is better than the corresponding rate $n^{-2/3}$ for histograms.

A main issue for nonparametric smoothing methods is the choice of the smoothing parameter. There exists a large literature on bandwidth selection for kernel density estimation (cf.

the overview in Sheather [9] and also in [10]); for the related problem of kernel choice see [11]). Smoothing parameters need to be chosen such that both undersmoothing, associated with large variance and small bias, and oversmoothing, associated with large bias and small variance, are avoided and instead the optimal trade-off between variance and bias is found.

A common selection method remains visual choice, often done interactively by increasing a clearly undersmoothing bandwidth, until the first sign of oversmoothing appears. When comparing to a histogram with small bin width, "residuals" defined by differences between histogram values and corresponding fitted density values at bin midpoints start to show signs for lack-of-fit when oversmoothing sets in. Popular automatic methods are cross-validation and plug-in methods. The main obstacle for data-adaptive selection methods is that the theoretically optimal bandwidth depends on the unknown density, as seen for a special case in Equation (3). The best currently known methods are plug-in methods, where these unknown quantities are estimated and substituted. Preliminary estimates of the unknown quantities include estimates of density derivatives. These preliminary estimates are also nonparametric, and their smoothing bandwidths are obtained in turn by estimating further unknown density-dependent quantities with pre-preliminary estimators; the bandwidths of the pre-estimators ultimately are obtained by making parametric (usually Gaussian) assumptions on the unknown densities.

A special problem for estimating densities with compact support are boundary effects that arise when estimating at or near the endpoints. These can cause problems when structure of interest occurs near such endpoints. One remedy is the use of special boundary adjustments such as boundary kernels as described in [12]. The basic density estimation schemes can be generalized to cover the estimation of multivariate densities, of density derivatives, and of densities obtained from incomplete or length-biased data. Other variants that have been studied include density estimation combined with data transformations, density estimation with locally varying bandwidths, density estimation under shape constraints such as unimodality, and the estimation of functionals of densities and of conditional densities. For reliability, applications of density estimation in renewal theory are of particular interest ([13]-[15]).

Alternative density estimation schemes include orthogonal series estimators, penalized maximum likelihood estimators and various estimation schemes based on the smoothing of histograms. Typically the histograms subjected to a subsequent smoothing step will he strongly undersmoothed and are constructed with very small bins, so that many of the bins remain empty. The midpoints of the m bins x_j and corresponding heights of the histogram bars y_j constitute a scatterplot $\{(x_j, y_j),\ j = 1, \ldots, m\}$ to which a nonparametric regression type smoother is applied. This smoother can be based on smoothing splines or B-splines. An easy-to-implement and explicit smoothing scheme which also automatically adjusts for boundary effects is provided by local linear fitting [16], where one uses a nonnegative kernel function K and a bandwidth h as defined above to obtain the minimizers of

$$\sum_{j=1}^{m} K\left(\frac{t - x_j}{h}\right) [y_j - \{a_0 + a_1(t - x_j)\}]^2 \tag{10}$$

w.r. to a_0, a_1. Then the density estimate is $\hat{f}_L(t) = \hat{a}_0(t)$, the estimated intercept at each t. The problem of adjusting density estimates that do not necessarily display the properties of a density (nonnegativity and integrating to 1) have been well studied (e.g., in [17]).

3. FAILURE RATE ESTIMATION

The failure rate $\lambda(t)dt$ represents the instantaneous chance that equipment fails in the interval $(t, t + dt)$, given that it works at age t. The hazard or failure rate function $\lambda(t)$ then provides a trajectory of risk. Nonparametric failure rate estimation often involves the smoothing of an initial failure rate estimate. Available data may be either grouped, in which case failures are reported aggregated over time intervals, or may consist of continuously observed failure times. For continuous data, the analogy of the relationships $F(t) = \int_0^t f(x)dx$ between distribution function F and density f, and $\Lambda(t) = \int_0^t \lambda(x)dx$ between cumulative hazard rate Λ and failure rate function λ motivates kernel estimates for λ and demonstrates the closeness of failure rate and density estimation problems.

One complication that one often faces with lifetime data is that available data are incomplete, due to loss to follow-up or termination of a study. The most commonly encountered type of incompleteness is random censoring [3]. In random censoring, we assume there is a censoring random variable C_i such that observed data are $X_i = \min(T_i, C_i)$, the minimum of the lifetime T_i and censoring time C_i of the ith individual. One also observes the censoring indicator $\delta_i = 1_{\{X_i = T_i\}}$, which is one if the actual lifetime is observed and zero otherwise. Let $(X_{(i)}, \delta_{[i]})$, $i = 1, 2, ..., n$, be the ordered sample with respect to lifetimes X_i, i.e., $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$, with $\delta_{[i]}$ the concomitant censoring indicator of $X_{(i)}$.

Hazard estimators can then be obtained by smoothing the increments of the Nelson-Aalen estimator [18]

$$\Lambda_n(t) = \sum_{i=1}^{n} \frac{\delta_{[i]} 1_{\{X_{(i)} \leq t\}}}{n - i + 1} \tag{11}$$

of the cumulative hazard function $\Lambda(t)$ (if there are no tied observations). Using the kernel method, we arrive at the kernel failure rate function estimator

$$\hat{\lambda}(t) = \sum_{i=1}^{n} \frac{1}{h} K(\frac{t - X_{(i)}}{h}) \frac{\delta_{[i]}}{n - i + 1}, \tag{12}$$

if there are no tied observations. Bias and variance issues can be similarly analyzed as for density estimates.

Boundary corrected kernel estimators and practical bandwidth choices have been considered in [12] and bootstrap confidence intervals in [19]. For the closely related problem of intensity function estimation with applications in reliability see [20],[21]. Other smoothing methods that have been considered include spline fitting with penalized likelihood [22].

For grouped data, one problem is the aggregation which can pose problems for nonparametric failure rate estimation. The reason is that the failure rate is defined as a limit. However, nonparametric estimation of failure rates works reasonably well when based on a combination of smoothing central rates of failure and a suitable transformation [23]. Assume the data are aggregated into intervals or bins of length Δ, the number of failures is d_i in the i-th such interval, and the number of items in working order and at risk of failure is n_i at the beginning and n_{i+1} at the end of the i-th such interval. Then the central failure rate for the i-th interval is defined as $q_{c_i} = 2d_i/(n_i + n_{i+1})$.

Denoting the midpoint of the i-th interval or bin by t_i, the smoothed central failure rate $\hat{q}_{c_i}$ is obtained by smoothing the data $\{t_i, q_{c_i}\}$, for example with weighted least squares smoothers as in equation 10 or with spline smoothers (see eqr 482). Then a recommended transformation estimate to obtain the (continuous) hazard or failure rate is [24]

$$\psi(\hat{q}_c(t)) = \frac{1}{\Delta} \log \frac{2 + \Delta \hat{q}_c(t)}{2 - \Delta \hat{q}_c(t)}. \tag{13}$$

It can be shown that this smoothing and transformation approach reduces the unavoidable discretization bias.

To avoid discretization biases it is preferable to record continuous failure times if either the density or the failure rate of the lifetime distribution is of interest. Extensions of the methods discussed here are available for the case where one is interested to determine which features of products influence lifetime. This leads to semiparametric hazard regression models, such as the Cox model or the accelerated failure time model (see [3]). Such models combine a baseline hazard function as a nonparametric component with a parametric part that models the influence of the covariates.

In reliability applications shape-restricted failure rates sometimes play a role. A common assumption is increasing failure rate (IFR). Monotonizing kernel estimators can be done in various ways, from applying the simple pool-adjacent violators algorithm (PAVA) to more sophisticated schemes [25]. Other shape restrictions such as bathtub-shape can also be implemented [21].

Related Entries

eqr 232, eqr 363, eqr 475, eqr 482

REFERENCES

1. Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27**, 832-837.

2. Liebscher, E. (2005). A semiparametric density estimator based on elliptical distributions. *J. Multiv. Analysis* **92**, 205-225.

3. Cox, D.R. and Oaks, D.R. (1990). *Analysis of Survival Analysis.* London: Chapman and Hall.

4. Fix, E. and Hodges, J.L. (1951). Discriminatory analysis and nonparametric estimation: consistency properties. *Rept. No. 4, Proj. No. 21-49-004,* USAF School of Aviation Medicine, Randolph Field, Texas.

5. Parzen, E. (1962). On estimating of probability density function and mode. *Ann. Math. Statist.* **33**, 1065-1076.

6. Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis.* London:Chapman and Hall.

7. Scott, D.W. (1992). *Multivariate Density Estimation.* New York:Wiley.

8. Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing.* Chapman and Hall: London.

9. Sheather, S.J. (2004). Density estimation. *Statistical Science* **19**, 588-597.

10. Jones, M.C., Marron, J. and Sheather, S. (1996). A brief survey of bandwidth selection for density estimation. *J. Amer. Statist. Assoc.* **91**, 401-407.

11. Granovsky, B. and Müller, H.G. (1991). Optimizing kernel methods: A unifying variational principle. *International Statistical Review* **59**, 373-388.

12. Müller, H.G. and Wang, J.L. (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics* **50**, 61-76.

13. Watelet, L., Winter, B.B. (1991). Nonparametric estimation of nonincreasing densities and use of data from renewal processes. *Communications in Statistics – Theory and Methods* **20**, 2073-2094.

14. Winter, B.B. (1989). Joint simulation and forward recurrence times in a renewal process. *J. Appl. Probability* **26**, 404-407.

15. Müller, H.G., Wang, J.L., Carey, J.R., Caswell-Chen, E.P., Chen, C., Papadopoulos, N., Yao, F. (2004). Demographic window to aging in the wild: Constructing life tables and estimating survival functions from marked individuals of unknown age. *Aging Cell* **3**, 125-131.

16. Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and Its Applications.* London: Chapman and Hall.

17. Gajek, L. (1986). On improving density estimators which are not bona fide functions. *Ann. Statist.* **14**, 1612-1618.

18. Nelson, W. (2000). Theory and applications of hazard plotting for censored failure data. *Technometrics* 42, 12-25.

19. Cheng, M.Y., Hall, P., Tu, D. (2006). Confidence bands for hazard rates under random censorship. *Biometrika* **93**, 357-366.

20. Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* **11**, 453-466.

21. Reboul, L. (2005). Estimation of a function under shape restrictions. Applications to reliability. *Ann. Statist.* **33**, 1330-1356.

22. O'Sullivan, F. (1988). Fast computation of fully automated log-density and log-hazard estimators. *SIAM J. Sci. Statist. Comput.* **9**, 363-379.

23. Wang, J.L., Müller, H.G., Capra, W.B. (1998). Analysis of oldest-old mortality: Lifetables revisited. *Ann. Statist.* **26**, 126-163.

24. Müller, H.G., Wang, J.L., Capra, W.B. (1997). From lifetables to hazard rates: The transformation approach. *Biometrika* **84**, 881-892.

25. Hall, P., Huang, L.-S., Gifford, J. A. and Gijbels, I. (2001). Nonparametric estimation of hazard rate under the constraint of monotonicity. *J. Comput. Graph. Statist.* **10**, 592–614.