International Journal of Smart Home
Vol. 10, No. 1, (2016), pp. 181-190
http://dx.doi.org/10.14257/ijsh.2016.10.1.18

Non-parameter Estimation of Failures Intensity of Tractor Based on Bootstrap

Yanfang Wang[1,2] and Changlin Ao[1※]

[1]College of Engineering, Northeast Agricultural University, Harbin 150030, China

[2]College of Management, Heilongjiang University of Science and Technology, Harbin 150028, China

aochanglin@gmail.com

Abstract

The failures data of tractors operating in field conditions were collected by failure tracking tests. A mathematical model is established with non-homogenous Poisson process to get the tractor failure intensity function. Bootstrap method is presented for constructing confidence regions for the failure intensity of a repairable system. Combined with fault data in tracking experiment, the early period failure intensity curves are drawn, revealing the failure regular in the initial stages of tractors. A confidence interval can be estimated by standard likelihood asymptotic theory in the parametric estimation, however in the non-parametric case it can be obtained by using the bootstrap. Comprising the MSE of cumulative intensity and cumulative number of failures between the two estimations, it is seen that the curve which was estimated by non-parameter estimation method is more realistic to the actual experiment process. Results provide the reference to improve tractor reliability, maintainability and maintenance strategy.

Key words*: failure intensity; bootstrap; reliability; non-homogeneous Poisson process; non-parameter estimation*

1. Introduction

Agricultural machinery is a sort of special product, which is operated by peasants with various abilities under different agricultural conditions. The reliability of machinery has an adverse effect on the annual agricultural earnings and a severe economic losses can be caused if the farm work can not be properly in the right season. At present agricultural machinery made in China in use have many problems in common such as more failures, more poor reliability and shorter service life compared to those made in developed countries. Some problems are inferior not only in performance but also in reliability in particular[1,2].

In the field of Agricultural machinery, most models for the research on the reliability are based on exponential distributions or Weibull distributions, and many methods used for parameter estimation of model are based on graphics or statistical analyses[3].In fact, the tractor failure process is a stochastic process. Non-homogeneous Poisson process as a branch of the stochastic process is widely used in the software reliability[4-6], and much work has been done on modeling the failure data of a repairable system and parameter estimation [7,8]. But not much work has been done on the reliability of agricultural machinery using the non-homogeneous Poisson process so far. In agriculture machinery reliability experiment, limited by the time and cost, the necessary and sufficient failure data can't be obtained in many cases. The reliability estimation results based on the little amount of data reflecting the actual reliability level is objective and can't be ignored. In this case, it needs to use approximate method and complex solution process in order to get model parameters and reliability interval estimation. As a result, it is difficult to get a

ISSN: 1975-4094 IJSH

effective estimation for the reliability of machinery[9]. Based on the current research a parametric estimate of the failure and a non-parametric estimate of the failure intensity of tractor are presented. In the parametric case a confidence interval can be obtained by standard likelihood asymptotic theory and in the non-parametric case a confidence region can be obtained by use of the bootstrap.

2. The Process of Tractor Failure

2.1. The Poisson Process[10][11]

Definition 1 A stochastic process $\{N(t), t \geq 0\}$ is mentioned to be a counting process. If $N(t)$ represents the total numbers of "event" that have occurred up to time t, a counting process $N(t)$ must satisfy:

(Ⅰ) $N(t) \geq 0$.
(Ⅱ) $N(t)$ is an integer value.
(Ⅲ) If $s < t$, then $N(s) \leq N(t)$
(Ⅳ) If $s < t$, then $N(t) - N(s)$ equals to the number of events that have occurred in the interval (s, t).

A counting process is mentioned to posses of independent increments if the numbers of events that occur in disjoint time intervals are independent, which means that the numbers of events that have occurred by time t (that is, $N(t)$) must be independent the numbers of events occurring between times t and $t+s$ (that is, $N(t+s) - N(t)$).

A counting process is said to possess of stationary increments if the distribution of numbers of events that occur in any interval of time only depends on the length of the time interval. In the other words, the process has stationary increments if the numbers of event in the interval (t_1+s, t_2+s) (that is, $N(t_2+s) - N(t_1+s)$ has the same distribution as the numbers of events in the interval (t_1, t_2) (that is $N(t_2) - N(t_1)$) for all $t_1 < t_2$ and $s > 0$.

The Poisson process is one of the most important types of counting processes, which is defined as follows:

Definition 2 The counting process $\{N(t), t \geq 0\}$ is mentioned to be a Poisson process if it has the parametric λ , $\lambda > 0$, and

(Ⅰ) $N(0) = 0$.
(Ⅱ)The process has independent increments.
(Ⅲ)The numbers of events occurred in any interval of length t are Poisson distributed with the mean value λt. That is, for all $s, t \geq 0$,

$$p\{N(t+s) - N(t) = n\} = \exp(-(m(t+s) - m(t))\frac{[m(t+s) - m(t)]^n}{n!},$$

$$p\{N(t+s) - N(s) = n\} = e^{-\lambda t}\frac{(\lambda t)^n}{n!}, \quad n = 0,1,\cdots \tag{1}$$

According to condition (Ⅲ), a Poisson process has stationary increments and also that

$$E[N(t)] = \lambda t \tag{2}$$

Equation(2) explains why λ is called the rate or the intensity of the process,that is , the mean number of events in unit interval.

An alternative definition of a Poisson process is given as the following.

Definition 3 The counting process $\{N(t), t \geq 0\}$ is said to be a Poisson process if it has a parametric λ , $\lambda > 0$, and

(Ⅰ) $N(0) = 0$.
(Ⅱ)The process has stationary and independent increments.

(Ⅲ) $p(N(h)=1)=\lambda h+o(h)$.
(Ⅳ) $p(N(h)\geq 2)=o(h)$.
Definition 2 and 3 are equivalent.

Considered a Poisson process, and let x_1 denote the time of the first failure occurrence. For $n\geq 1$, let x_n denote the time between $(n-1)$-th and *n*-th failure occurrence. The sequence $\{x_n, n\geq 1\}$ is called the sequence of interval times, and x_n (n=1, 2, … n) is proved to be an exponential random variable whose mean value of independent identical distribution is $1/\lambda$.

Definition 4 The counting process $\{N(t), t\geq 0\}$ is said to be a non-stationary or non-homogeneous Poisson process if it has the intensity function $\lambda(t)$, where $t\geq 0$, and

(Ⅰ) $N(0)=0$.
(Ⅱ) $\{N(t), t\geq 0\}$ has independent increments.
(Ⅲ) $p\{N(t+h)-N(t)\geq 2\}=o(h)$.
(Ⅳ) $p\{N(t+h)-N(t)=1\}=\lambda(t)\cdot h+o(h)$.
If

$$m(t)=\int_0^t \lambda(s)ds \tag{3}$$

it can be shown that

$$p\{N(t+s)-N(t)=n\}=e^{-[m(t+s)-m(t)]}\frac{[m(t+s)-m(t)]^n}{n!},\quad n\geq 0 \tag{4}$$

that is, $N(t+s)-N(t)$ is a Poisson distribution with the mean value $m(t+s)-m(t)$. When the intensity function $\lambda(t)$ is a constant, the non-homogeneous Poisson process can be thought as a homogeneous Poisson process.

When the intensity function is

$$\lambda(t)=\lambda\beta t^{\beta-1} \tag{5}$$

This Poisson process is proved to be a Weibull process. Here, λ 、$\beta>0$, β is a shape parameter and λ is an intensity parameter.

For non-homogeneous Poisson process (NHPP), the sequence of the interval between failures is proved to be neither mutually independent nor identically distributive. That is, they are neither an exponential distribution nor independent samples from an identical distribution. Therefore any technique based on an independent identical distribution can not be used in a non-homogeneous Poisson process. But as a Poisson process, a non-homogeneous Poisson process also has independent increments.

2.2. Tractor Failure Process

Since the appearance of tractor failure at the instantaneous time $t_i(i=1,2,\cdots,n)$ is stochastic, it can be taken as a random point along the time axis. Thus the failure process of a repairable system can be described by a stochastic point process. Suppose that the reliability of a tractor after repaired be the same as before repaired, we refer to the non-homogeneous Poisson process as the model of tractor failure process[1,2].

3. The Principle of Bootstrap Method

3.1. The Kernel Estimation Method

If there is no parametric form assumed for the intensity,a non-parametric estimation method is required. One such method is the kernel estimation method. It was assumed that the system is observed for the time $(0, t_n)$,where $(t_1, t_2, \cdots, t_n)$ is the time of the tractor fault ,which is a Poisson point process.

A kernel estimate of the varying intensity $\lambda(t)$ is given by

$$\lambda(t) = \frac{1}{h}\sum_{i=1}^{n} K(u)$$

Where

$$u = \frac{(t - t_i)}{h}$$

and K is a kernel function

$$K(u) = \frac{15}{16}(1 - u^2)^2 \quad \text{for } -1 \le u \le 1$$

h denotes the bandwidth. This guarantees that $\hat{\lambda}(x)$ is non-negative, an essential property of an estimate of the intensity[12].

3.2. The Bootstrap and Resampling Method

The bootstrap was introduced in 1979 by Efron as a computer-based method for estimating standard errors[13]. Bootstrap methods depend on the notion of a bootstrap sample: the bootstrap data points $T^* = (t_1^*, t_2^*, \cdots, t_{n^*}^*)$ are a random sample of size n^*,drawn with replacement from the population of n objects $T = (t_1, t_2, \cdots, t_n)$. The method is motivated by $n^* = n$ in general[14].

The interval estimation method based on bootstrap repeated sampling from the sample data which have been obtained from the population, and calculate the value of z_i in every sample .The quaitiles of z were estimated according to the sorted z_i with the given confidence level, then the confidence interval was estimated.

It is assumed that $T = (t_1, t_2, \cdots, t_n)$ is the observations .The algorithm is as follows:

(1)Calculate $\hat{\lambda} = \frac{1}{h}\sum_{i=1}^{n} K(u)\ , u = \frac{(t - t_i)}{h}$ by $T = (t_1, t_2, \cdots, t_n)$;

(2)Draw the sub sample $T_B^* = (t_1^*, t_2^*, \cdots, t_n^*)$ from $T = (t_1, t_2, \cdots, t_n)$ with the bootstrap,then $\hat{\lambda}^*_B = \frac{1}{h}\sum_{i=1}^{n} K(u)\ , u = \frac{(t - t_i^*)}{h}$;

(3)Estimate $s\hat{e}^*_B$,the standard error of $\hat{\lambda}^*_B$,as following:

①Get the sub sample with bootstrap from $T_B^* = (t_1^*, t_2^*, \cdots, t_n^*)$ of step (2) once again,then calculate $\hat{\lambda}^*_{B1} = \frac{1}{h}\sum_{i=1}^{n} K(u)\ , u = \frac{(t - t_i^{**})}{h}$;

②repeat ① $B1$ times,then get $\hat{\lambda}^*_{B1}(1), \hat{\lambda}^*_{B1}(2), \cdots, \hat{\lambda}^*_{B1}(B1)$;

③Calculate $s\hat{e}^*_B = \left\{ \frac{\sum_{i=1}^{B1} [\hat{\lambda}^*_{B1}(i) - \overline{\hat{\lambda}^*_{B1}}]}{B1 - 1} \right\}^{1/2}$,where $\overline{\hat{\lambda}^*_{B1}} = \frac{\sum_{i=1}^{B1} \hat{\lambda}^*_{B1}(i)}{B1}$;

(4)Repeat (2) and (3) B times,then get $\hat{\lambda}^*_B(1), \hat{\lambda}^*_B(2), \cdots, \hat{\lambda}^*_B(B)$ and $s\hat{e}_B^*(1), s\hat{e}_B^*(2), \cdots, s\hat{e}_B^*(B)$;

(5)Calculate $s\hat{e} = \left\{ \frac{\sum_{i=1}^{B} (\hat{\lambda}^*_B(i) - \overline{\hat{\lambda}^*_B})^2}{B - 1} \right\}$,the standard error of $\hat{\lambda}$,where

$\overline{\hat{\lambda}^*_B} = \frac{\sum_{i=1}^{B} \hat{\lambda}^*_B(i)}{B}$,and $z(i) = \frac{\hat{\lambda}^*_B(i) - \hat{\lambda}}{s\hat{e}^*_B(i)}$, $(i = 1,2,\cdots,B)$.

(6) $z(1), z(2), \cdots, z(B)$ were ordered $z'(1) \le z'(2) \le \cdots \le z'(B)$,take $z'_{[B \bullet \alpha]}$ and $z'_{[B \bullet (1-\alpha)]}$ as the quantile,where $[B \bullet \alpha]$ and $[B \bullet (1-\alpha)]$ mean round numbers.

(7)Get $\overline{\hat{\lambda}^*_B} = \frac{\sum_{i=1}^{B} \hat{\lambda}^*_B(i)}{B}$,the point estimation of λ .The interval for λ is given by $(\overline{\hat{\lambda}^*_B} - z'_{[B \bullet \alpha]} s\hat{e}, \overline{\hat{\lambda}^*_B} - z'_{[B \bullet (1-\alpha)]} s\hat{e})$ with 95% confidence.

4. Research on Regularity of Tractor Failures

4.1. The Failure Data of Tractors

The failure times used for illustration are the times of 113 failures for 10 tractors from a farm in Heilongjiang province, which are given in Table 1.

Table 1. Fault Data of 10 Tractors from a Farm in Heilongjiang Province (Unit: Working Hours)

175	186	191	227	241	323	343	389	410	466
470	561	608	643	669	683	699	733	786	799
801	817	825	844	884	889	939	986	993	993
1007	1011	1021	1028	1051	1068	1072	1103	1107	1113
1121	1123	1124	1134	1143	1146	1159	1162	1172	1177
1181	1185	1384	1389	1461	1514	1536	1550	1574	1682
1727	1745	1745	1760	1766	1895	1955	1977	1996	2251
2268	2271	2317	2341	2387	2433	2551	2561	2575	2583
2599	2626	2660	2683	2772	2816	2839	2899	2927	2927

2949	2992	3049	3105	3109	3169	3218	3239	3249	3262
3326	3429	3448	3473	3538	3565	3594	3715	3835	3872
3894	3975	4083							

4.2 Parametric Estimation of the Failure Intensity

4.2.1 Point Estimation

A model for a repairable system is the NHPP with intensity function given by

$$\lambda(t) = \gamma\delta t^{\delta-1}$$

The MLE (maximum likelihood estimator) of $\lambda(t)$ is given by

$$\hat{\lambda}(t) = \hat{\gamma}\hat{\delta}t^{\hat{\delta}-1} \quad (6)$$

where $\hat{\gamma}$ and $\hat{\delta}$ are the MLE of γ and δ . The log-likelihood for a repairable system observed for the time interval $(0, t_n)$ with failures at $t_1, t_2, \cdots, t_n$ was given by Crow[15] ,and then the estimation value can be obtained :

$$\hat{\delta} = \frac{n}{\sum_{i=1}^{n} \ln \frac{t_n}{t_i}} \quad (7)$$

$$\hat{\gamma} = \frac{n}{(t_n)^{\hat{\delta}}} \quad (8)$$

To the observed information,it is possible to calculate $\gamma = 0.0401$, $\delta = 0.9555$,so $\lambda(t) = 0.0303\ t^{-0.001747}$.

The point estimation can be plotted for each value of t in $[0, t_n]$,and the plot is given in Figure 1.From this plot it is seen that the intensity function is estimated to be a decreasing function which decreases from more than 0.0305 to about 0.0265, so tractors in early failure period.

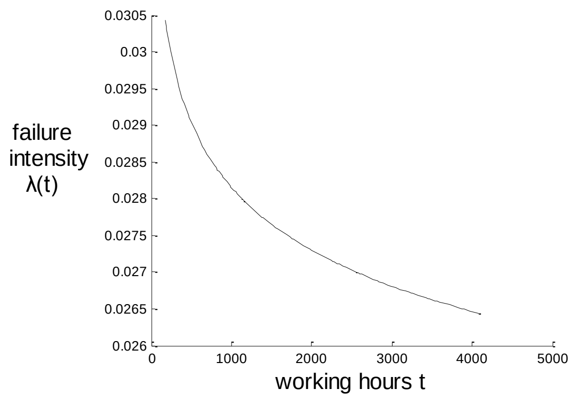

Figure 1. Point Estimation of Failure Intensity Curve of Tractors based on Parameter Estimation

The tractor failure process is NHPP and Weibull,then the cumulative failure intensity function is given by

$$m(t) = \gamma t^{\delta-1} \quad (9)$$

For $\gamma = 0.0401$, $\delta = 0.9555$,then $m(t) = 0.0401 \cdot t^{-0.0455}$.

The curve of $m(t)$ and the scatter diagram of cumulative failure times which is the accumulation of fault times in tracking experiment on tractors are given in figure 2.

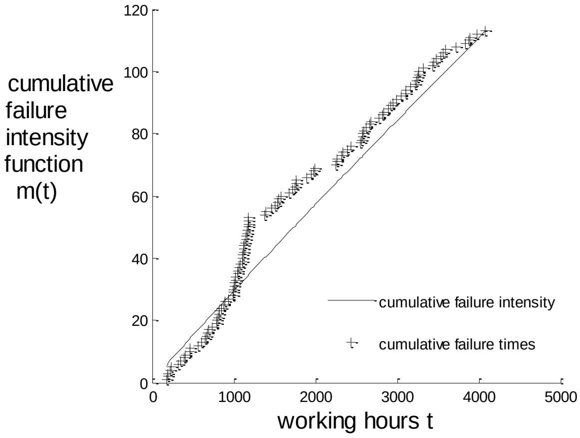

Figure 2. Cumulative Failure Intensity Curve and Cumulative Failure of Tractors based on Parameter Estimation

From this figure, it is seen that the trends between the curve of cumulative failure intensity and the scatter diagram of cumulative failure times in practice are consistent in $[0, t_n]$,however the deviation exists. Calculate the mean square error between the cumulative failure intensity and the cumulative failure times, then it is calculated that $MSE = 8.2601$.

4.2.2 Interval Estimation

The intensity function for each value of t in $[0, t_n]$ is Estimated and hence a 95% confidence interval for $\lambda(t)$ is calculated, which is given by

$$(\hat{\gamma}\hat{\delta}t^{\hat{\delta}-1}e^{(-\frac{1.96}{\sqrt{n}}\sqrt{1+(1+\hat{\delta}\ln\frac{t}{t_n})^2})}, \hat{\gamma}\hat{\delta}t^{\hat{\delta}-1}e^{(\frac{1.96}{\sqrt{n}}\sqrt{1+(1+\hat{\delta}\ln\frac{t}{t_n})^2})})$$

$$=(0.0235\ t^{-0.001747}, 0.0389\ t^{-0.001747})$$

The limits of these individual confidence intervals can then be plotted in figure 3.

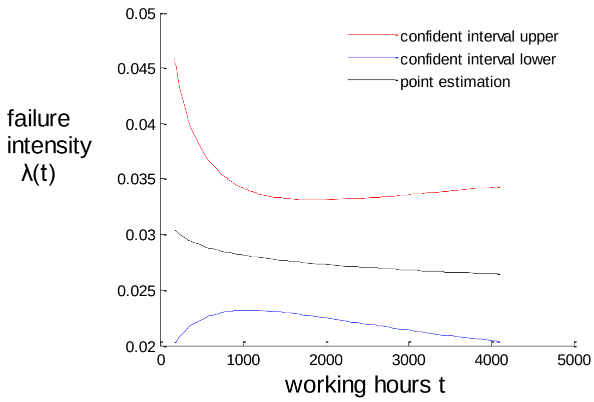

Figure 3. Interval Estimation Curve of Failure Intensity of Tractors base on Parameter Estimation

The black curve in figure 3 is the point estimation value of tractors fault intensity, same with figure 1; the red and blue curves are the upper and lower limit of confident interval of tractor failure intensity.

4.3 Non-parametric Estimation of the Failure Intensity

If no parametric form is assumed for the intensity, a non-parametric estimation method is required [9,10].Hence a non-parametric method will be subscribed for $\lambda(t)$.

By the non-parametric estimation algorithm from 3.2, tractors fault intensity curve can be plotted in figure 4,where $B = 400$, $B_1 = 20$.

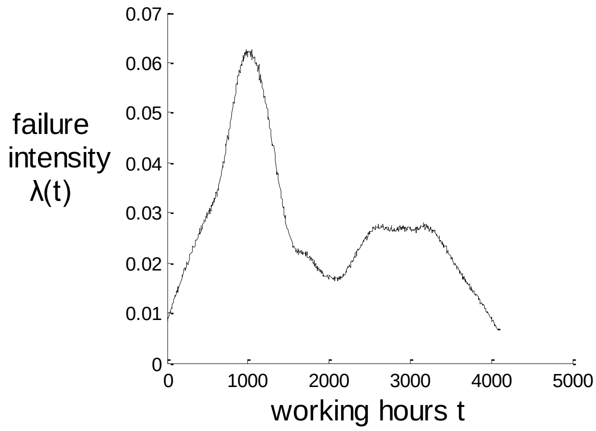

Figure 4. Point Estimation of Failure Intensity Curve of Tractors based on Non-parameter Estimation

From this plot it is seen that the non-parameter estimation describes the overall tractors failure intensity trend, comparing with the parameter estimation,the details of the failure intensity such as the little fluctuation of failure intensity ascending and decreasing are exhibited obviously.

In the parameter estimation, $\lambda(t)$ is decreasing from 0.0265 to 0.0305 over time for assuming that the failure process is Weibull process.While in the non-parameter,the curve being obtained by kernel estimation reveals that the values of $\lambda(t)$ between 0.0068 and 0.0630 is not monotone decreasing in [0,4083] ,in fact it ascends in [0,1000].Therefor it should be paid attention to the failure process and record the failure component in early period from 0 to 1000 hour of tractor in using, to provide guidelines for improving the reliability of tractors.

The figure 5 is given,where $m(t)$ is the value of estimation,and cumulative failure times is the accumulation of the actual failure times in tractors test.

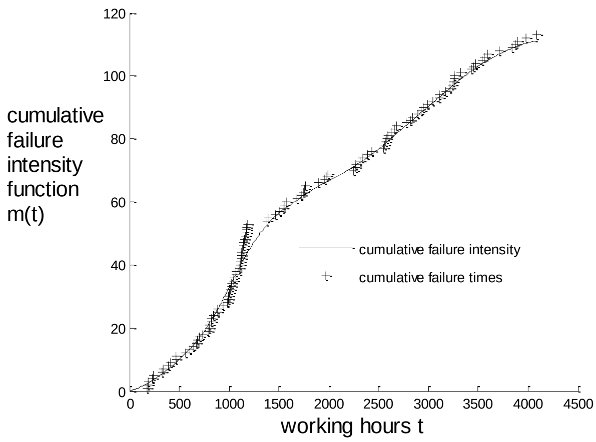

Figure 5. Cumulative Failure Intensity Curve and Cumulative Failure of Tractors based on Non-parameter Estimation

From figure 5,the curve tend of cumulative failure intensity by non-parameter estimation is fitting well to the scatter diagram of cumulative failure times in practice.Then the MSE between the values of cumulative failure intensity and cumulative failure times can be calculated:

$$MSE = 2.3197 .$$

The $MSE = 2.3197$ in non-parameter estimation is obviously less than the $MSE = 8.2601$ in parameter estimation .Therefore the failure intensity curve in non-parameter estimation is more close to the actual test process.

By the non-parametric estimation algorithm from 3.2,Using the resampling method, it is possible to produce a confidence region for the tractors data in Figure 6.

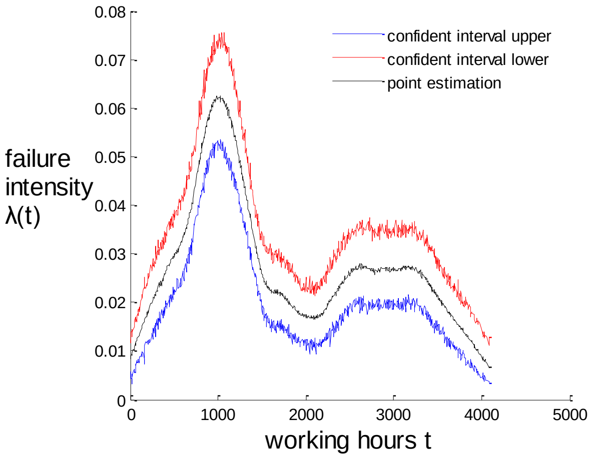

Figure 6. Interval Estimation Curve of Failure Intensity of Tractors based on Non-parameter Estimation

The black curve in figure 6 is the tractors fault intensity point estimation, similar to figure 4; The red and blue curves are the interval upper and lower limits of tractor failure intensity with 95% confident based on bootstrap.

5. Results

1. According to theoretical analyses and statistical testing, the tractor failure process is a non-homogeneous Poisson process; The failure intensity curves in working hours give a complete description of the regular of tractors failures in early period by the parameter estimation, further more the non-parameter estimation is given by bootstrap.

2. it is seen that the failure intensity function is estimated to be a decreasing function with time in parameter estimation, and the bathtub curve is the early period of mechanics in using .Wherever in non-parameter estimation, it is found that the failure intensity of the tractors is increasing during 0 to 1000 hour, therefore the process of design and manufacture of tractor need to be further improved.

3. Comparing cumulative failure intensity and actual cumulative failure times of the mean square error between the estimation of parameter and non-parameter, it is found that the MSE is much smaller in non-parametric estimation. Thus it is validated that the non-parametric estimation of tractor failure intensity is much closer to the actual test process.

Acknowledgments

The authors would like to acknowledge the financial support from the National Natural Science Foundation of China (71171044).

References

[1] C. Ao, Y. Li, X. Yan and F. Chu, “Operational reliability of tractor engines based on non-homogeneous poisson process”, Chinese Journal of Mechanical Engineering, vol. 43, no. 10, **(2007)**, pp. 206-209.
[2] C. Ao, X. Zheng, L. Xie and Y. Dai, “Parameter estimation of a tractor reliability model based on artificial neural network”, Journal of agricultural machinery, vol. 35, no. 3, **(2004)**, pp. 31-33.
[3] T. Bucar, M. Nagode and M. Faidiga, “Reliability approximation using finite Weibull mixture distributions”, Reliability Engineering and System safety, vol. 84, **(2004)**, pp. 241-251.
[4] M. Kimura, “A study on bootstrap confidence intervals of software reliability measures based on an incomplete gamma function model”, World Scientific, **(2006)**, pp. 419-426.
[5] D. R. Jeske, X. Zhang and L. Pham, “Adjusting software failure rates that are estimated from test data”, IEEE Transactions on Reliability, vol. 54, no. 3, **(2005)**, pp. 107-114.
[6] S. Inoue and S. Yamada, “Discrete software reliability assessment with discretized NHPP models”, Computers & Mathematics with Applications: An International Journal, vol. 51, no. 2, **(2006)**, pp. 161-170.
[7] S. T. Jiang, T. L. Landers and T. R. Rhoads, “Semi-parametric proportional intensity models robustness for right-censored recurrent failure data”, Reliability Engineering &system safety, vol. 90, no. 1, **(2005)**, pp. 91-98.
[8] M. M. B. Othman, A. Mohamed and A. Hussain, “Determination of transmission reliability margin using parametric bootstrap technique”, IEEE Transactions on Power Systems, vol. 23, no. 4, **(2008)**, pp. 1689-1700.
[9] D. N. Politis and J. P. Romano, “Non-parametric resampling for homogeneous strong mixing random fields”, Journal of Multivariate Analysis, vol. 47, **(1993)**, pp. 301-328.
[10] Y. Lin, “Applied random process”, Tsinghua University press, **(2012)**.
[11] V. V. Krivtsov, “Practical extensions to NHPP application in repairable system reliability analysis”, Reliability Engineering and System Safety, vol. 92, no. 5, **(2007)**, pp. 560-562.
[12] Z. Wang and J. Yang, “Bootstrap method of parameter estimation in reliability model”, Journal of Harbin institute of technology, vol. 42, no. 5, **(2010)**, pp. 820-823.
[13] B. Efron, “Bootstrap methods:another look at the jackknife”, The Annals of Statistics, vol. 7, no. 1, **(1979)**, pp. 1-26.
[14] T. Kaneishi and T. Dohi, “Parametric bootstrapping for assessing software reliability measures”, Dependable Computing(PRDC), IEEE 17th Pacific Rim International Symposium, **(2011)**.
[15] L. H. Crow, “Confidence interval procedures for the Weibull process with applications to reliability growth”, Technometrics, vol. 24, **(1982)**, pp. 67–72.

Author

Yanfang Wang (1979-), female, native place:Liaoning province,China. She is a lecturer at college of management in Heilongjiang University of science and technology, PhD student at College of Engineering in the Northeast Agricultural University. Research area: agricultural system engineering, system reliability, modeling and simulation.

