PAPER • OPEN ACCESS

The use of fractional order statistics for estimating nonparametric confidence intervals for quantiles of the fatigue damage computed in service random loadings

To cite this article: D Benasciutti and J M E Marques 2023 *IOP Conf. Ser.: Mater. Sci. Eng.* **1275** 012020

The use of fractional order statistics for estimating nonparametric confidence intervals for quantiles of the fatigue damage computed in service random loadings

D Benasciutti[1] **and J M E Marques**[2]

[1] University of Ferrara, Department of Engineering, via Saragat 1, 44122 Ferrara, Italy
[2] Czech Technical University in Prague, Department of Mechanics, Biomechanics and Mechatronics, Faculty of Mechanical Engineering, Technická 4, 166 36 Prague, Czech Republic.

Corresponding author's email address: denis.benasciutti@unife.it

Abstract. This paper deals with the statistical variability of fatigue damage in random loadings as caused by the inherent randomness of the loading. It examines the use of a nonparametric method based on fractional order statistics as a tool for constructing confidence intervals for the quantile of fatigue damage, regardless of the knowledge of the damage probability distribution. The approach here discussed relies on a statistical method existing in the literature. After a short review of this method, with also an emphasis on possible advantages and limitations, the paper presents a numerical benchmark, followed by a practical application to time-histories measured on a mountain-bike in off-road tracks. The paper concludes with some critical remarks and an outlook.

1. Introduction

The statistical variability of the fatigue damage $D(T)$ of a random loading of length T can be characterized by the variance of damage $\mathrm{Var}[D(T)]$, which – along with the expected damage $E[D(T)]$ – is used to define the coefficient of variation (CoV) $\sqrt{\mathrm{Var}[D(T)]}/E[D(T)]$ as a dimensionless measure of uncertainty.

The interest on the study of the variance (or CoV) of fatigue damage in random loadings dates to the sixties, with the pioneering works of Bendat [1] and Crandall et al. [2,3]. They devised analytical formulas for calculating the variance $\mathrm{Var}[D(T)]$ in the case where the random time-history $z(t)$ is the output of a stationary Gaussian linear oscillator. The obtained variance expressions depend upon a few statistical properties of the random loading (e.g. variance, mean up-crossing frequency) and the oscillator (e.g. damping ratio), other than upon SN parameters.

After Bendat's and Crandall et al.'s works, the interest on the damage variance continued in subsequent decades, for example in Madsen et al. [4] and, more recently, in the studies [5] and [6,7], which extended Bendat's and Crandall et al.'s solutions to any type of narrowband random loading, either Gaussian [4-6] or non-Gaussian [7].

All the above variance solutions rely on the knowledge of the statistical properties characterizing the random loading, e.g. autocorrelation function or power spectral density. Often, these quantities are not known exactly by mathematical expressions and they must be estimated from measured data. This introduces a sampling variability into the computational model of the damage variance, the effect of which is not considered by the aforementioned methods.

This aspect motivated our attempt to develop a different approach in which the statistical variability of fatigue damage is estimated by means of confidence intervals, constructed after a direct analysis of one or more time-histories [6]. First developed for the case of stationary random loadings, the approach has been extended to non-stationary loadings of "switching type" (i.e., formed by a sequence of stationary states) [8,9]. The method has been benchmarked, first against numerically simulated time-histories, and then against more realistic time-histories measured in a mountain-bike in off-road tracks; in either case, a quite good accuracy was observed [8,9].

It is the purpose of the present article to improve further this approach by abandoning two of its hypotheses, in which the confidence interval is: i) constructed for the expected damage $E[D(T)]$ and ii) under the assumption of $D(T)$ following a normal distribution, a hypothesis introduced in early studies [3, 5,10] but recently called into question [11].

With the aim to overcome the two previous hypotheses, the method proposed herein makes use of nonparametric fractional order statistics to construct confidence intervals for any quantile of damage $D(T)$, regardless of the specific form of its probability distribution. The method thus allows for the construction of confidence intervals other than the median or the expected value. In this article, numerical and experimental examples are used to emphasize advantages and limitations of the proposed method.

It has finally to be noted that, in the following, only the statistical variability coming from the randomness of the loading will be considered. It is assumed that the scatter of constant amplitude fatigue data is taken into account e.g. by the definition of a characteristic SN line referred to prescribed levels of failure probability and confidence (e.g. P_f=2.3% with 95% confidence).

2. The uncertainty in fatigue damage

2.1. Fatigue damage and its expected value

Let $z(t)$, $0<t<T$, be a random time-history of finite length T. The fatigue damage under Palmgren-Miner (PM) rule is:

$$D(T)=\sum_{i=1}^{l}\frac{n_i}{N(s_i)} \tag{1}$$

where $N(s_i)$ is the number of cycles to failure at constant amplitude s_i, n_i is the number of cycles within the i-th class with amplitude s_i, and l is the number of amplitude classes.

Equation (1) considers a discrete set of amplitudes s_i referred to cycles counted in $z(t)$; another set is obtained if another time-history $z(t)$ is considered. In the limit of continuously distributed amplitudes, classes s_i become infinitely narrow and the summation approaches an integral [6]:

$$E[D(T)]=E[n]\int_0^{\infty}\frac{1}{N(s)}p(s)\mathrm{d}s \tag{2}$$

where $E[D(T)]$=expected damage, $E[n]$=expected number of cycles in $z(t)$, $p(s)$=probability distribution of amplitudes. In Eq. (2), the probability distribution $p(s)$ describes the randomness of the amplitudes of the cycles counted in $z(t)$.

The previous formulas are very general as they do not imply any mathematical expression to relate s and $N(s)$, nor do they require $z(t)$ being stationary – $z(t)$ can also be nonstationary. Often, it is assumed that amplitudes and cycles to failure are related as $N(s)=As^{-k}$ (elementary PM rule), which in a log-log diagram is a straight line with slope k and intercept $A=s_A^k N_A$, where the reference amplitude s_A is

IOP Conf. Series: Materials Science and Engineering 1275 (2023) 012020 doi:10.1088/1757-899X/1275/1/012020

computed at N_A cycles (for example, N_A=2×10^6 cycles), for a predetermined failure probability. For this single-slope SN curve, the fatigue damage in Eq. (1) and (2) writes:

$$D(T)=\frac{1}{A}\left(\sum_{i=1}^{l} n_i\, s_i^k\right) \quad , \qquad E[D(T)]=E[n]\,\frac{E[s^k]}{A} \tag{3}$$

Where $E[s^k]$ is the k-th moment of the amplitude distribution $p(s)$. Closed-form expressions can be obtained if the expression of $p(s)$ is known; for example, if $z(t)$ is a so-called narrowband loading, $p(s)$ is Rayleigh and Eq. (2) yields $E[D(T)]=(\nu_0^+ T/A)\left(\sqrt{2\mathrm{Var}[z(t)]}\right)^k \Gamma(1+k/2)$, where ν_0^+ is the number of mean value upcrossings of $z(t)$ [6].

When a single-slope SN equation is believed to be too conservative, two-slope SN models are used in which a knee point is followed by a less steep line with slope k'=2k–1 or 2k–2 (Haibach model [12]) or even by a horizontal line that represents a "endurance (fatigue) limit". In this case, it is straightforward to modify the previous damage formulas accordingly.

3. Confidence intervals

Let us consider a random variable X with cumulative distribution function $F(x)$. Further, assume that θ is a parameter of $F(x)$, e.g. it may be the expected value of a normal distribution. A two-sided confidence interval (CI) for an unknown parameter θ is an interval, defined by two end points θ_l and θ_u, that is expected to enclose the unknown parameter θ a specified percentage of cases after that an experiment is repeated under identical conditions. If the specified percentage is $100(1-\alpha)\%$, a CI is defined as [13]:

$$\Pr[\theta_l<\theta<\theta_u]=(1-\alpha) \quad \text{for } 0<\alpha<1 \tag{4}$$

(subscripts l and u mean "lower" and "upper"). The values θ_l and θ_u, which are constructed around the estimator $\hat{\theta}$ of θ, are called the lower and upper confidence limits, respectively, and the probability $(1-\alpha)$ is named coverage probability or confidence level.

Often, the parameter θ of interest is the expected value of a Gaussian probability distribution. In this case, the two-sided $100(1-\alpha)\%$ CI for μ is [13]:

$$\bar{x}-t_{\alpha/2;n-1}\frac{\hat{\sigma}}{\sqrt{n}}<\mu<\bar{x}+t_{\alpha/2;n-1}\frac{\hat{\sigma}}{\sqrt{n}} \tag{5}$$

where $t_{\alpha/2;n-1}$ is the upper $100\,\alpha/2$ percentage point of the Student t-distribution with n–1 degrees of freedom, while $\bar{x}$ and $\hat{\sigma}^2$ are the sample mean and sample variance, respectively. In other engineering applications, the parameter θ of interest is a quantile of the probability distribution. In the following, we provide an overview of some confidence interval expressions.

3.1. Confidence intervals for the expected damage E[D(T)]

The last result of previous section forms the basis of the confidence interval for the expected damage. Let $z(t)$ be a non-stationary "switching" random loading, formed by a finite number of stationary states N_S. Each stationary state has a total time length T_j, so that $T=\sum_{j=1}^{N_S} T_j$ is the total duration of $z(t)$. Note that each state needs not to appear in $z(t)$ only once in its full length T_j, but it can appear a multitude of times, each with duration shorter than T_j, provided that its total time length is T_j. It is irrelevant in which time order and how many times a state appear in $z(t)$, so at the beginning of the analysis, the states are reordered so that they appear only once in their full length T_j [8,9].

Methods for identifying different states in $z(t)$ (fast Fourier transform, run test) are described in [14]. After state identification, each state is further subdivided into N_B blocks of equal length T_B. For each

IOP Conf. Series: Materials Science and Engineering 1275 (2023) 012020 doi:10.1088/1757-899X/1275/1/012020

block, the fatigue damage is finally computed. In summary, state identification and block subdivision yield the set of damage values $D_{B,ij}(T_B)$, with i=1,2,…,N_B and j=1,2,…,N_S. Damage $D_{B,ij}(T_B)$ refers to block i in state j. Under the assumption of $D(T)$ normally distributed, the confidence the confidence interval for the expected damage $E[D(T)]$ then is [8,9]:

$$D(T) - t_{\eta,\beta/2} \sum_{j=1}^{N_S} \sqrt{N_B \cdot \hat{\sigma}^2_{D_B,j}} \leq E[D(T)] \leq D(T) + t_{\eta,\beta/2} \sum_{j=1}^{N_S} \sqrt{N_B \cdot \hat{\sigma}^2_{D_B,j}} \tag{6}$$

where $\hat{\sigma}^2_{D_B,j}$ is the sample variance of the damage of blocks for the j-th stationary state, and $t_{\eta,\beta/2}$ is the quantile of the Student t-distribution with an equivalent number of degrees of freedom $\eta \cong (N_B - 1)\left(\sum_{j=1}^{N_S} \hat{\sigma}^2_{D_B,j}\right)^2 / \sum_{j=1}^{N_S} \left(\hat{\sigma}^2_{D_B,j}\right)^2$. When the switching time-history $z(t)$ has only one stationary state (N_S=1), the previous confidence interval expression converges to the solution given in [6] for the stationary case. Numerical and experimental results have confirmed the validity of the above confidence interval for both stationary and non-stationary random loadings [6,8,9].

It has to be emphasized that Eq. (6) relies on the hypothesis of normally distributed damage $D(T)$, which can be assumed under the validity of the central limit theorem in the limit $T \to \infty$. This hypothesis, first introduced in [3], is also mentioned in other studies [5,10], though it was not until the works [5,11] that its validity has been scrutinized by numerical simulations. In [5], we read: "*the departure from the Gaussian distribution is only marginal. These results suggest that the Gaussian hypothesis made in prior studies is valid when the CoV is small, which usually indicates that T is long enough for the central limit theorem to apply. Fortunately, this is true for the majority of practical situations*." Further investigations [11] showed that larger values of Cov (e.g. large k, very narrow spectral bandwidth of random loading) lead to larger deviations from gaussianity, with the damage distribution being skewed. In this case, the assumption of normally distributed damage is under conservative [11].

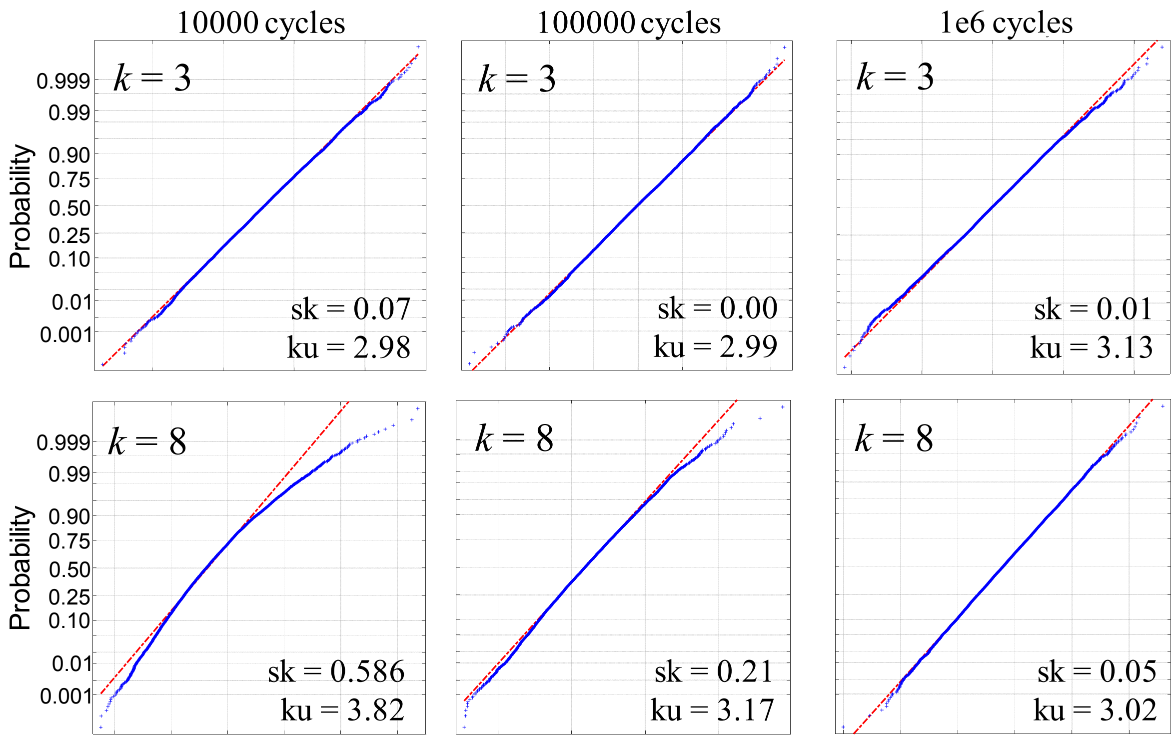

Figure 1. Normal probability plots of fatigue damage for different values of k and number of cycles.

It has to be noticed, though, that the numerical example in [11] considered a small number of cycles (3200 cycles) that, while accentuating the skewness of damage distribution, appears to be too small for any engineering application. On the other hand, the rate of convergence to the normal distribution for the damage depends on the number of random variables being summed (i.e. number of counted cycles) and to which extent each random variable departs from the normal distribution (i.e. effect of inverse slope k). This trend is exemplified in Figure 1, which compares the normal probability papers for different combinations of k and number of counted cycles. The graphs refer to 10000 damage values computed from samples of independent Rayleigh distributed stress amplitudes (corresponding to narrowband loadings); independence among amplitudes neglects the time correlation between adjacent cycles, which in fact corresponds to the output of any counting method in which the relative time position of cycles is lost. From Figure 1, it is apparent that, for high k, far more cycles are required for the damage to become normally distributed.

3.2. Confidence intervals for quantile estimates based on normal approximation

For an assigned probability P, let $\hat{x}_P=\widehat{F}^{-1}(P)$ be the estimator of the Pth quantile x_P, where $\widehat{F}^{-1}$ denotes the inverse of the probability distribution $\widehat{F}(x)$ that approximates $F(x)$. A method for constructing CIs for quantiles is to assume that $\hat{x}_P$ is normally distributed, with mean x_P and variance $\mathrm{Var}[\hat{x}_P]=\frac{1}{n}\frac{P(1-P)}{\hat{f}^2(\hat{x}_P)}$ [15,16], where $\hat{f}(x)$ is the probability density corresponding to $\widehat{F}(x)$, and n is the sample size. For large sample size (validity of central limit theorem), an approximate CI is:

$$\hat{x}_P - z_{1-\alpha/2}\sqrt{\mathrm{Var}[\hat{x}_P]} \le x_P \le \hat{x}_P + z_{1-\alpha/2}\sqrt{\mathrm{Var}[\hat{x}_P]} \tag{7}$$

where $z_{1-\alpha/2}$ is the upper $1-\alpha/2$ percentage point of the standard normal random variable. This approach is simple as it only requires the knowledge of the estimator variance $\mathrm{Var}[\hat{x}_P]$, which in turn depends on the probability density $\hat{f}(x)$ – in literature, analytical expressions are available for the most common probability distributions [15,16].

3.3. Confidence intervals for quantiles based on order statistics

An alternative approach to construct CIs for quantiles, which is especially useful when $F(x)$ or its estimation $\widehat{F}(x)$ are not known in advance, is by means of order statistics. Let $X_{1:n}\le X_{2:n}\le\ldots\le X_{n:n}$ be the order statistics from a sample of size n of a random variable X with probability distribution $F(x)$. It is well known that [13]:

$$\pi(q,r,n,P)=\Pr\left[X_{q:n}\le x_P\le X_{r:n}\right]=\sum_{i=q}^{r-1}\binom{n}{i}P^i(1-P)^{n-i} \quad , \quad q<r \tag{8}$$

where $P=F(x_P)$ is the probability associated to the quantile. Thus, any choice of q and r such that $\pi(q,r,n,P)=1-\alpha$ provides a $100(1-\alpha)\%$ nonparametric CI for the Pth quantile x_P. The approach followed in practice is to try different combinations of q and r that make π as much close to $(1-\alpha)$ as possible. However, if the sample size n is small − as usually happens in practice – the probability π may substantially deviate from the prescribed coverage probability $(1-\alpha)$.

The use of fraction order statistics [17] allows the definition of an *exact* confidence interval for x_P for any value of confidence. The first step is to rewrite Eq. (8) as [18]:

$$\Pr[x_l\le x_P\le x_u]=I_P\left(n'P_l, n'(1-P_l)\right)-I_P\left(n'P_u, n'(1-P_u)\right) \tag{9}$$

where $0<P_l<P_u<1$, $n'=n+1$ and $I_P(a,b)=\int_0^P t^{a-1}(1-t)^{b-1}\mathrm{d}t/I(a,b)$ denotes the incomplete beta function ratio (it is the beta cumulative distribution function). The *exact* $100(1-\alpha)\%$ CI for the Pth quantile x_P

IOP Conf. Series: Materials Science and Engineering 1275 (2023) 012020 doi:10.1088/1757-899X/1275/1/012020

is defined by $(x_l, x_u)=\left(X_{n'P_l:n}, X_{n'P_u:n}\right)$ by using fractional order statistics, where the two probabilities P_l and P_u are determined numerically by solving:

$$I_P\left(n'P_l, n'(1-P_l)\right)=1-\alpha/2 \quad , \quad I_P\left(n'P_u, n'(1-P_u)\right)=\alpha/2 \tag{10}$$

Fractional order statistics is also used to define the estimator of the Pth quantile as $\hat{x}_P=X_{n'P:n}$.

As pointed out in [18], the fractional quantile $X_{n'P:n}$ represents only a technical device for defining a continuum of order statistics and it cannot be calculated from the sample, since $n'P$ may not be integer. The same drawback also applies to $X_{n'P_l:n}$ and $X_{n'P_u:n}$. They are then approximated by a linear interpolation, as proposed in [18]. Indeed, it has been shown that the distribution of $X_{n'P:n}$ is well approximated by the distribution of the following linear combination of order statistics:

$$\hat{x}_P^{int}=(1-\varepsilon)X_{\lfloor n'P\rfloor:n}+\varepsilon X_{\lfloor n'P\rfloor+1:n} \quad , \quad \frac{1}{n+1}<P<\frac{n}{n+1} \tag{11}$$

where $\lfloor\cdot\rfloor$ denotes the floor function, and $\varepsilon=n'P-\lfloor n'P\rfloor$, with $0<\varepsilon<1$. The same linear interpolation in Eq. (11) is also used to compute the values x_l^{int} and x_u^{int}, provided that P is replaced by P_l and P_u, respectively; this introduces two new interpolation weights $\varepsilon_l=n'P_l-\lfloor n'P_l\rfloor$ and $\varepsilon_u=n'P_u-\lfloor n'P_u\rfloor$. The two latter values define the lower and upper limits of the interval $\left(x_l^{int}, x_u^{int}\right)$, which represents the *approximate* $100(1-\alpha)\%$ nonparametric CI for the Pth quantile x_P.

An illustrative example is now discussed. Suppose to have an ordered sample $X_{i:n}$ of size n=10 and consider the quantile P=0.75. Since Pn'=0.75×11=8.25, the fractional order statistics gives the quantile estimator $\hat{x}_{0.75}=X_{8.25:10}$, which clearly does not exist in the ordered sample. The nearest sample values, $X_{8:10}$ and $X_{9:10}$, are then used to compute the linear interpolation estimator as $\hat{x}_P^{int}=0.75\cdot X_{8:10}+0.25\cdot X_{9:10}$, weight parameter being $\varepsilon=n'P-\lfloor n'P\rfloor=8.25-8=0.25$.

A point worth to mention is that the linear interpolation in Eq. (11) is only valid for quantiles P in the region between $1/(n+1)$ and $n/(n+1)$, which are the lowest and highest quantiles corresponding to a sample size n. Owing to this condition, the lower and upper confidence limits cannot be computed if $\lfloor n'P_l\rfloor=0$ and $\lfloor n'P_u\rfloor=n$ (see [19]), which is equivalent to the conditions $P_l<1/(n+1)$ and $P_u>n/(n+1)$. For example, assume we want to construct a 99% CI for a sample of size n=50, and for two extreme quantiles P=0.1 and P=0.9. According to Eq. (10), for P=0.1 it is P_l=0.019371 and so $\lfloor n'P_l\rfloor=0$ (note that P_l=0.019371<1/51=0.0196). Similarly, for P=0.9, it is P_u=0.980629 and $\lfloor n'P_u\rfloor=50$, so the required value $\lfloor n'P_u\rfloor+1=51$ exceeds the last available sample value (again, note that P_u=0.980629>50/51=0.9804). In both cases, the CI limits do not exist.

These conditions restrict the admissible combinations of sample size, confidence and quantile levels. As an example, Table 1 lists the minimum sample size n required for some combinations of quantile and confidence levels. For P=0.75 and α=0.05 (confidence 95%), a minimum sample size of 13 is needed, but this threshold almost triples if the quantile level becomes 0.90, the confidence being kept unchanged. The table therefore emphasizes how a simultaneous large quantile level and confidence require an equally large sample size.

Table 1. Minimum sample size n for several combinations of quantile and confidence

	Quantile level, P		
$(1-\alpha)$	P=0.5	P=0.75	P=0.90
0.90	5	11	29
0.95	6	13	36
0.99	8	19	50

A possibility to partly overcome the above limitations is to use the extrapolation formula proposed in [20], which extends the definition in Eq. (10) to extreme quantiles:

$$\hat{x}_P(P)=\begin{cases} X_{1:n}+(X_{2:n}-X_{1:n})\ln(n'P) & , \quad 0<P\leq\frac{1}{n+1} \\ (1-\varepsilon)X_{\lfloor n'P\rfloor:n}+\varepsilon X_{\lfloor n'P\rfloor+1:n} & , \quad \frac{1}{n+1}<P<\frac{n}{n+1} \\ X_{n:n}+(X_{n:n}-X_{n-1:n})\ln\big(n'(1-P)\big) & , \quad \frac{n}{n+1}\leq P<1 \end{cases} \tag{12}$$

where the second expression in Eq. (12) is the linear interpolation introduced previously. For a random variable with values X>0, and sufficiently near zero, the first expression modifies as $\varepsilon X_{1:n}$ [20].

4. Numerical examples

The accuracy of the nonparametric confidence interval described in Sec. 3.3 has already been checked in [18] for the median and third quartile, by considering five different standardized probability distributions (uniform, normal, Cauchy, exponential, Gumbel). The method was proved to yield coverage probabilities very close to the prescribed nominal value of 95%.

Here, the aim is to perform another simulation study that considers a higher quantile (P=0.90) and also relatively "small" sample sizes for which $P_u>n/(n+1)$, a condition that requires the use of the extrapolation formula in Eq. (12). In the present numerical study, 100000 pseudo-random samples are generated from four standardized probability distributions (uniform, normal, exponential, Gumbel); for each sample, the CI for the quantile x_P of the distribution is constructed by the method in Sec. 3.3. The coverage probability is estimated by counting the percentage fraction of confidence intervals that enclose the true quantile x_P, and it is finally compared with the required nominal value 95%. Note that for the considered combination of quantile level (P=0.90) and confidence (95%), the limit sample size is n=36; for smaller values, the extrapolation in Eq. (12) must be invoked. The comparison of results is summarized shown in Figure 2.

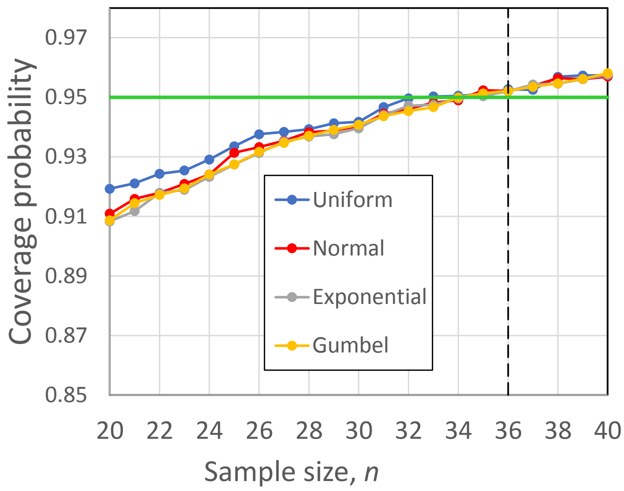

Figure 2. Coverage probability as a function of sample size for different probability distributions. Quantile P=0.90, nominal confidence 95%.

The figure emphasizes that, for a sample size n larger than the limit value 36 (i.e. extrapolation formula not used), the coverage probability is close to the nominal value 0.95, with a slightly increasing trend against n (a tendency already observed in [15]), towards a horizontal asymptote for very large n (result not shown in Figure 2). By contrast, the coverage probability diminishes as n becomes lower than 36, a condition in which $P_u \gg n/(n+1)$ and the upper confidence limit is approximated by the extrapolation formula Eq. (12). The figure nevertheless shows that it is possible to use a sample size smaller than that dictated by the condition $P_u<n/(n+1)$, if a small loss of confidence is accepted.

5. Case study: Mountain-bike data

This section considers a set of time-histories recorded in a mountain-bike on off-road tracks. The aim is not that of evaluating the safety of the bicycle, but to show an engineering application of the statistical methods described in the previous Sections.

The mountain-bike is a 2010 Scott Sportster P6, with a frame in 6061 aluminum alloy and a Unicrown front fork in carbon steel. Two Rigida Cyber 10 size 700C wheels are coupled with 700×37c S207 semi-slick tires. Handlebar and saddle are from Scott Sports; transmission, chain, and crankset are from Shimano, Inc.

The loadings acting on the bicycle front fork were measured by two strain gages, placed symmetrically on the left tube and connected in half-bridge configuration; the system was calibrated in laboratory static loading tests. Strain gage are from HBM, model LY Linear strain gauge with one measuring grid (only one direction). Time-histories were recorded by a Dewesoft data acquisition system (model Minitaurs Dewe-101 with 8 channels). This model has an industrial power computer built directly into the unit. Sampling frequency was 1000 Hz. A filter was set up with cut-off frequency of 300 Hz, which was above the maximum frequency of interest. The data acquisition system was fixed at the inclined tube of the main "triangle" frame of the bicycle: the rechargeable supply battery was allocated behind the seat, on a welded support [21].

The bicycle speed was monitored by a speedometer (Marwi Group, model Union 8 Cycling Computer), formed by a sensor and magnet fixed on a wheel rod. When fully equipped, the mountain-bike weighed about 12.2 kg.

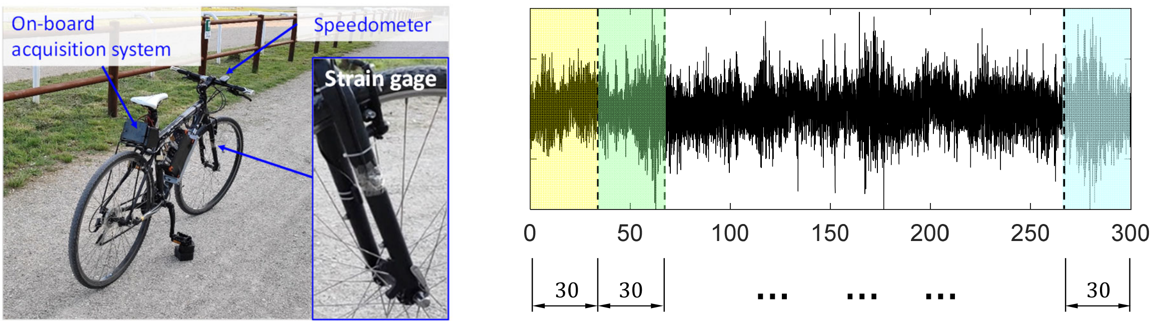

Figure 3. Mountain-bike with strain gages and onboard acquisition system, and example of measured time-history subdivided into segments of 30 seconds.

A total of n=40 time-histories was measured and used in the subsequent analysis. Each time-history is first subdivided into 10 segments of 30 seconds each. The fatigue damage is computed for each segment, by assuming an SN curve $N(s)=As^{-k}$, with A=1 and k=3 and 5. Therefore, 10 damage values were obtained from each time-history; their maximum was finally computed. In summary, the initial set of measured time-histories returned a sample of 40 maxima of damage $D_{max,i}$, i=1,…,40, which represents the random sample X_i used to construct the nonparametric confidence interval of quantile described in Sec. 3.3. Subsets with n=10, 20, 30 damage values were also considered to study the behavior of the confidence interval computed from smaller samples.

Figure 4(a) compares the 95% confidence intervals for 0.90 quantile, obtained for different sample sizes. Damage values are normalized to the sample mean. Black circles (•) represent the sample damage values, the star (∗) is the quantile estimator, while the vertical lines are the quantile confidence intervals. For a sample size n<36 (limit value), the extrapolation formula Eq. (12) is used (confidence intervals are red colored). Note that the nonparametric confidence interval needs not to be symmetric around the quantile estimator. For n=40, it is $P_u<n/(n+1)$ and therefore the upper limit of CI falls below the last

sample value in the ordered statistics. Instead, for sample size <36, the confidence interval has an increasing width, while its upper limit exceeds the last samples value (also by a large amount for small n) and falls in the region of validity of the extrapolation formula.

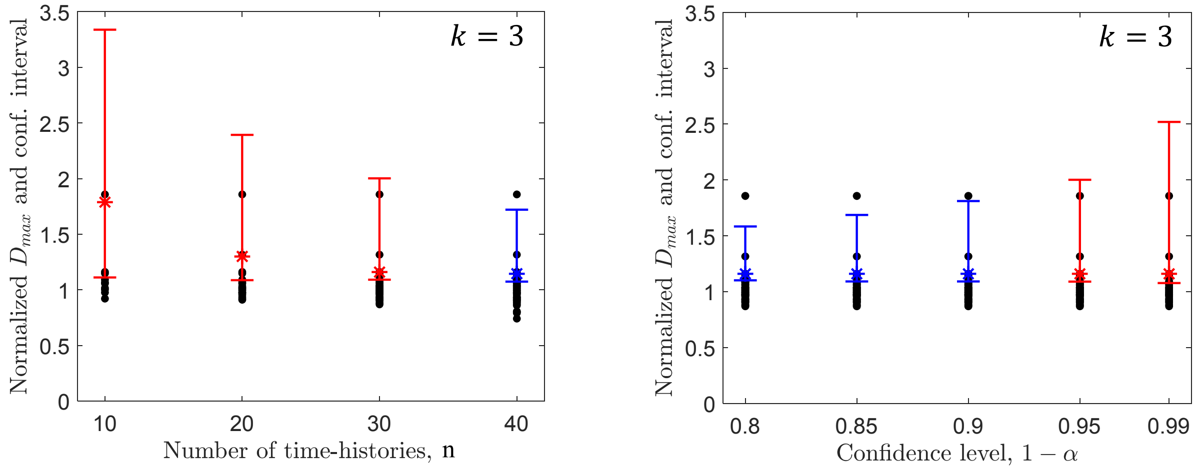

Figure 4. Confidence intervals of the 0.90 quantile, in the case of (a) fixed confidence 95% and different sample size, or (b) fixed sample size n=40 and different confidence.

Figure 4(b) replots the same results of Figure 4(a), where CIs are now shown as a function of the confidence level. It can be seen that the extrapolation formula is used only for confidence levels higher than 0.90; in this case, the upper confidence limit exceeds the highest value in the ordered statistics.

6. Conclusions

The paper studied the applicability of a nonparametric method based on fractional order statistics − and developed in two literature papers [18,20] − as a tool for estimating the confidence interval of quantiles of fatigue damage computed in random loadings. After a brief theoretical summary of the method, a numerical study and an experimental case study have been examined, the results of which allow for the following considerations:

- the nonparametric method has the undoubtful advantage of making no prior assumption on the probability distribution of fatigue damage, contrary to our previous method that hypothesized a normally distributed damage to construct the confidence interval;
- in the original formulation of the nonparametric method [18], there is a constraint in the admissible combinations of sample size n, quantile level P and confidence $(1-\alpha)$, which is dictated by the limit quantile levels, $1/(n+1)$ and $n/(n+1)$, established by sample size n;
- in a subsequent formulation [20], an approximated extrapolation formula allows the previous constraint to be overcome, which however occurs – as shown by our simulation results – at the expense of a loss in confidence, which becomes more pronounced if the sample size decreases far below the "limit" value;
- in order for the extrapolation formula not to be applied, a large sample size n must be used. However, for very high quantiles (e.g. >0.99) and confidence levels (e.g. 95%), the required n may be so large that the method could become impractical (indeed, n represents the number of time-histories from which the sample of damage values are computed). For very high quantile levels (e.g. >0.99), the use of the extreme value theory may be more appropriate [22].

Acknowledgments
Research has partially been funded by CTU Global Postdoc Fellowship Program (research topic #2-11).

IOP Conf. Series: Materials Science and Engineering 1275 (2023) 012020 doi:10.1088/1757-899X/1275/1/012020

References

[1] Bendat J S 1964. *Probability functions for random responses: Prediction of peaks, fatigue damage, and catastrophic failures*. *National Aeronautics and Space Administration contract report 33* (Washington, D.C., USA) p 57
[2] Mark W D 1961 *The inherent variation in fatigue damage resulting from random vibration. Ph.D. Thesis Department of Mechanical Engineering MIT* (Cambridge, Massachusetts, USA) p 93
[3] Crandall S H, Mark W D and Khabbaz G R 1962 *The variance in Palmgren-Miner damage due to random vibration*. *Proc. of 4th US National Congress of Applied Mechanics* vol 1 (California, Berkeley, USA) pp 119–26
[4] Madsen H O, Krenk S and Lind N C 1986 *Methods of Structural Safety* (Englewood Cliffs, NJ, USA: Prentice-Hall) p 416
[5] Low Y M 2012 Variance of the fatigue damage due to a Gaussian narrowband process. *Struct. Saf.* **34**(1) 381–9
[6] Marques J M E, Benasciutti D and Tovo R 2020 Variability of the fatigue damage due to the randomness of a stationary vibration loading. *Int. J. Fatigue* **141** 105891
[7] Marques J M E, Benasciutti D and Tovo R 2021 Variance of the fatigue damage in non-Gaussian stochastic processes with narrow-band power spectrum. *Struct. Saf.* **93** 102131
[8] Benasciutti D and Marques J M E 2021 Methods for estimating the statistical variability of the fatigue damage computed in one single stationary or non-stationary random time-history record. *Proc. of 3rd Int. Conf. on Structural Integrity for Offshore Energy Industry STRUCTURAL INTEGRITY 2021* (15–16 November 2021, Aberdeen, UK)
[9] Marques J M E and Benasciutti D 2021 Evaluating confidence interval of fatigue damage from one single measured non-stationary time-history. *Proc. of VIRTUAL 20th Int. Colloq. on Mechanical Fatigue of Metals* (15–17 September 2021, Wrocław, Poland)
[10] Desmond A F 1987 On the distribution of the time to fatigue failure for the simple linear oscillator. *Probab. Eng. Mech.* **2**(4) 214–8
[11] Low Y M 2012 An analytical formulation for the fatigue damage skewness relating to a narrowband process. *Struct. Saf.* **35**(1) 18–28
[12] Haibach E 2002 *Betriebsfestigkeit* 2006 (Heidelberg, Germany: Springer-Verlag Berlin)
[13] Montgomery D C and Runger G C 2014. *Applied Statistics and Probability for Engineers* 6th ed. (Hoboken, NJ, USA: John Wiley & Sons) p 284
[14] Marques J M E, Solazzi L and Benasciutti D 2022 Fatigue analysis of nonstationary random loadings measured in an industrial vehicle wheel: Uncertainty of fatigue damage. *Metals* **12**(4) 616
[15] Serinaldi F 2009 Assessing the applicability of fractional order statistics for computing confidence intervals for extreme quantiles. *J. Hydrol.* **376**(3–4) 528–41
[16] Su S 2009 Confidence intervals for quantiles using generalized lambda distributions. *Comput. Stat. Data Anal.* **53**(9) 3324–33
[17] Stigler S M 1977 Fractional order statistics, with applications. *J. Am. Stat. Assoc.* **72**(359) 544–50
[18] Hutson A D 1999 Calculating nonparametric confdence intervals for quantiles using fractional order statistics. *J. Appl. Stat.* **26**(3) 343–53
[19] Nagaraja C H and Nagaraja H N 2020 Distribution-free approximate methods for constructing confidence intervals for quantiles. *Int. Stat. Rev.* **88**(1) 75–100
[20] Hutson A D 2002 A semi-parametric quantile function estimator for use in bootstrap estimation procedures. *Stat. Comput.* **12** 331–8
[21] Marques J M E 2021 Confidence intervals for the expected damage in random loadings: Application to measured time-history records from a mountain-bike. *IOP Conf. Ser. Mater. Sci. Eng.* **1038**, 012025
[22] Castillo E *Extreme value theory in engineering* 1988 (San Diego, USA: Academic Press)