



Toward a proper statistical description of defects

A. Cetin^{a,*}, A. Naess^b

^a Department of Engineering Design and Materials, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway

^b CeSOS & Department of Mathematical Sciences, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway

ARTICLE INFO

Article history:

Received 5 May 2011

Received in revised form 22 October 2011

Accepted 23 November 2011

Available online 23 December 2011

Keywords:

Defects

Extreme value

Sub-asymptotic

ABSTRACT

The Gumbel distribution (asymptotic theory) in conjunction with block maximum sampling is widely adopted to describe material defects. However, the appropriateness of asymptotic theory is questionable, as the block sizes are often relatively small due to practical limitations. In this paper, a recent and more appropriate sub-asymptotic extreme value distribution is presented and discussed in context of defects. Block maximum sampling scenarios with two different underlying defect size distributions (Weibull and log-normal) were simulated. It was demonstrated that the presented sub-asymptotic distribution is to a higher degree capable of capturing the actual extreme value behavior when compared to its asymptotic counterpart, i.e. Gumbel distribution. Furthermore, a comparison with the generalized extreme value distribution, which was adopted as an approximate/empirical distribution function, was also performed.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Fatigue failures originating from material defects (inclusions, voids, etc.) are widely discussed and documented in the literature. For instance, the influence of defects on the fatigue properties of various engineering metals has been studied thoroughly by Murakami [1]. In that regard, statistical methods are often applied to describe characteristics of defects.

The main purpose of statistical methods in design against fatigue is to make inference about unobserved events, e.g. estimate the largest defect in a component. Thus, proper statistical description of defect sizes is imperative.

The defect size distribution is estimated from appropriate samples of defects obtained from small reference volumes or areas, which are hereafter referred to as blocks. The sampling process usually involves various practical and analytical techniques depending on the type of defects and material in question [2–4]. Nonetheless, three main strategies are utilized; (i) direct sampling, (ii) peak over threshold sampling, and (iii) block maximum sampling.

Direct sampling is the most transparent method. The defect size distribution is estimated by counting and measuring every defect present in a block. However, this method is limited by resources and technical equipment, often rendering it unfeasible. For example, it is difficult to quantify the smallest defects due to limitations with equipment. Furthermore, there are also theoretical difficulties associated with this strategy since an explicit assumption regarding the parametrization of the distribution has to be made.

Peak over threshold sampling is an attempt to remedy some of the difficulties associated with direct sampling. In this sampling strategy, only defects larger than some threshold are considered and described by means of truncated distributions. However, the theoretical difficulties mentioned above are also valid for this strategy. Thus, asymptotic theory and assumptions are often invoked, and the generalized Pareto distribution is used to describe the defects [5,6].

In block maximum sampling, k separate and equally sized blocks are defined and only the largest defect present in each of them is sampled. In that regard, asymptotic extreme value theory is usually invoked and the Gumbel distribution (EV-I) is adopted to describe the largest defects [7]:

$$F_{EV-I}(x) = \exp \left\{ -\exp \left(-\left(\frac{x - \mu}{\sigma} \right) \right) \right\} \quad (1)$$

A valid objection against the block maximum sampling strategy is that obtainable information from a technical point of view, e.g. the 2nd largest defect in every block, is ignored. Nonetheless, this is by far the most convenient strategy. Due to this virtue, it is maybe the most common one as well. Therefore, this paper will only treat this strategy.

The use of EV-I in conjunction with block maximum sampling is to some extent theoretically justified, since samples do indeed come as extreme values. This approach is general in the sense that few assumptions of practical importance regarding the underlying defect size distributions are necessary. However, EV-I arise from asymptotic theory and great care must be shown when using it. As it was demonstrated by Cohen [8,9], convergence to the asymptotic distributions is often slow. That is, the number of defects present in a block must be very large. The practical implication is

* Corresponding author.

E-mail address: ali.cetin@ntnu.no (A. Cetin).

often that the data points do not align on a straight line in a Gumbel probability plot. Thus, the appropriateness of EV-I in describing defects is often questionable.

Recently, Naess and Gaidai [10] proposed a new distribution function to describe the extreme response of dynamical systems. Unlike EV-I, this distribution function is not exclusively based on asymptotic theory and thus to a certain extent is capable of capturing the sub-asymptotic behavior of extremes. In this paper, the proposed sub-asymptotic distribution function (SAD) is adopted as the proper distribution function of the largest defect sizes. In Section 2, the distribution function is presented in the context of describing defects. In Section 3, SAD is applied and compared to EV-I and the generalized extreme value distribution in a simulated block maximum method scenario. Although the statistical method presented in this paper has been known for some time, its application in fatigue appears to be new.

2. Extreme value theory

2.1. General

Before proceeding, fundamental aspects of extreme value theory and concepts essential for the following discussions will be presented. This is to ensure that this paper is comprehensible and self-contained. It may also be helpful to consult one of numerous textbooks on the subject.

Consider blocks containing N defects and assume that the location of the defects are randomly and independently distributed. Let the independent and identically distributed (i.i.d.) random variable X denote the size of defects. Their cumulative distribution function (CDF) is denoted $F(x) = \text{Prob}(X \leq x)$. The purpose of this paper is to give a rational statistical description and determine the distribution function of the extreme value $X_{\max}^{(N)} = \max\{X_i, i = 1, \dots, N\}$.

Under the i.i.d. assumption, that is, the X_i are independent copies of a common random variable X , the exact CDF of the extreme values is obtained from order statistics [7]:

$$F_{X_{\max}^{(N)}}(x) = (1 - S(x))^N \quad (2)$$

where $S(x) = 1 - F(x)$. Exploiting the series definition of the exponential function, Eq. (2) can be expressed as:

$$F_{X_{\max}^{(N)}}(x) = q(x, N) \exp\{-NS(x)\} \quad (3)$$

where $q(x, N) = \left(1 - \frac{S^2(x)}{2N} + \frac{S^3(x)(8+3S(x))}{24N^2} + \dots\right)$ and is related to the remainder term in the Taylor series. Obviously, $q(x, N)$ will approach unity as $N \rightarrow \infty$. More importantly, $q(x, N)$ will be fairly close to unity even for intermediate values of N .

Assuming blocks with equal number of defects may appear to be a stretch in practical applications. A more realistic scenario would be blocks of equal spatial size with the number of defects in each block varying stochastically as a Poisson process of a certain intensity. Curiously, the resulting CDF would be similar to Eq. (3), where the value of $q(x, N)$ would be unity regardless of N . Therefore, the effect of choosing either of the discussed assumption would be insignificant in practice. (The current assumption was chosen since it is more transparent and comprehensible).

2.2. Asymptotic distribution functions

Asymptotic extreme value theory states that as $N \rightarrow \infty$, $F_{X_{\max}^{(N)}}(x)$ approaches one of three possible limiting distribution functions often referred to as Type I (Gumbel), Type II (Fréchet) and Type III (Weibull) extreme value distributions. It is said that $F(x)$ is in the domain of attraction of a Type I/Type II/Type III extreme value distribution. However, this is not merely a formal mathematical

statement. The authors would like to emphasize that the purpose is to model physical reality. Each of the three limiting extreme value distributions describe distinctively different behaviors.

The asymptotic extreme value distributions can be jointly parametrized with GEV (Type I $\xi \rightarrow 0$), Type II ($\xi > 0$) and Type III ($\xi < 0$):

$$F_{\text{GEV}}(x) = \exp\left\{-\left[1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right]^{-1/\xi}\right\} \quad (4)$$

The most common distribution function (such as normal, log-normal, gamma and Weibull) are in the domain of attraction of a Type I extreme value distribution. Thus, in order to avoid unnecessary complications, the underlying defect size distribution is assumed to be in the domain of attraction of a Type I extreme value distribution; that is, the resulting extreme value distribution will asymptotically approach EV-I. (Distribution functions in the domain of attraction of a Type II and a Type III extreme value distributions do in most cases possess properties that make them unlikely candidates for defect size distribution). Makkonen [11] discusses plausible candidates for the underlying defect size distribution and concludes that Weibull and log-normal distributions are the most likely ones. Although the discussion is somewhat limited, it is consistent with the assumption made here.

The main concern is that the size of blocks is often relatively small due to practical limitations and N is not sufficiently large for the asymptotic approximation to be valid. The result is often that the data points do not align on a straight line in a Gumbel probability plot. Therefore, the appropriateness of the asymptotic extreme value theory in defect size description is questionable.

The poor fit is often misinterpreted as EV-I not being the correct distribution. However, the poor performance of EV-I in approximating the distribution of extremes in practical problems was already demonstrated empirically by Fisher and Tippet [12] in 1928. Cohen [8,9] demonstrated that Type II or Type III extreme value distributions would be better approximations for the observed data. Therefore, GEV with unconstrained parameters will fit the observed data better even if EV-I is the asymptotically correct one. That said, extrapolation in order to make inference beyond the observed data should be avoided as the tail behavior of GEV may be significantly different from the actual behavior. Notice also that even if GEV offers better fit for the observed data it will asymptotically diverge from the actual behavior.

In that regard, SAD is developed and presented as an appropriate extreme value distribution.

2.3. Sub-asymptotic distribution function

The motivation behind SAD is the following. The distribution of $X_{\max}^{(N)}$ is well defined for two limiting cases; (i) $N = 1$, where the extreme value distribution is the underlying distribution and (ii) $N \rightarrow \infty$, where the extreme value distribution approaches EV-I. However, most practical applications of block maximum sampling is bound to end up in a region between these two cases. The aim is to describe this intermediate region appropriately while ensuring consistency with the asymptotes.

Consider Eq. (3); $NS(x)$ comes directly from the underlying defect size distribution, but its exact form is in general unknown. Examining it for the relevant underlying distribution functions, reveals that it often takes the form of an exponential function. Encouraged by this and the asymptotic form (EV-I), Eq. (3) may be expressed as:

$$F_{X_{\max}^{(N)}}(x) \simeq \begin{cases} \exp\{-\exp(g(x))\} & x > x_T \\ Z(x) & x \leq x_T \end{cases} \quad (5)$$

where $g(x)$ and $Z(x)$ are suitable functions, and x_T is an appropriate tail marker. The function $Z(x)$ is often not of practical interest and

Download English Version:

<https://daneshyari.com/en/article/778415>

Download Persian Version:

<https://daneshyari.com/article/778415>

[Daneshyari.com](https://daneshyari.com)