Title	Closure to " Problems in the extreme value analysis" (Struct. Safety 2008:30:405-419)
Author(s)	Makkonen, Lasse; Pajari, Matti; Tikanmäki, Maria
Citation	Structural Safety. Elsevier. Vol. 40 (2013) No: January, Pages 65 - 67
Date	2013
URL	http://dx.doi.org/10.1016/j.strusafe.2012.09.007
Rights	Post-print version of the article. This article may be downloaded for personal use only.

Closure to “Problems in the extreme value analysis” (Struct. Safety 2008:30:405-419)

Lasse Makkonen[1)], **Matti Pajari**[2)] **and Maria Tikanmäki**[1)]

[1)] *VTT Technical Research Centre of Finland*
[2)] *VTT Expert Services Ltd*

Corresponding author address:

Lasse Makkonen, VTT Technical Research Centre of Finland, Box 1000, 02044 VTT, Espoo, Finland

Email: lasse.makkonen@vtt.fi

Keywords: Extreme value analysis; Plotting positions; Order-statistics; Probability theory

ABSTRACT

Extreme value analysis (EVA) is an essential part of the statistical methodology for securing structural safety. Makkonen [3] argued that the estimators of the plotting position in EVA should be abandoned and replaced by the Weibull expression $m/(N+1)$. In a rebuttal Cook [5] challenged this development. Here we prove by the probability theory that the Weibull expression provides the rank probability exactly. This shows that no estimators of the plotting positions are necessary. We also comment on the specific criticism to [3] in [5] and outline the foundations of the correct methodology.

1. Introduction

The first step in the extreme value analysis (EVA) is to associate the observed variable values with their probability. For this many so called plotting position formulas have been proposed [1].

Makkonen [1-4] concluded based on the classical definition of probability and a sampling exercise that the Weibull plotting position formula $m/(N+1)$ uniquely provides the non-exceedance probability of the mth smallest value in order-ranked N observations. Cook [5] challenged this conclusion and called for a rigorous and comprehensive proof.

Here we provide such a proof. We also reply to the criticism on [1] made in Cook's rebuttal [5].

2. Proof of $P\{x \leq x_{m:N}\} = m/(N+1)$

A compact proof of the uniqueness of Weibull's plotting position has been published in a textbook by Madsen et al. [6]. Another rigorous proof is given in the following.

Consider variate x with cumulative distribution function F and a sample of N observations $X_{1:N}$, ... , $X_{n:N}$ ranked in ascending order. Values $X_{m:N}$ in different samples of size N are random values of variate $x_{m:N}$ for which the probability density function $f_{m:N}$ in terms of F and $f = F'$ is given by [6]

$$f_{m:N}(x_{m:N}) = m\binom{N}{m}F(x_{m:N})^{m-1}[1-F(x_{m:N})]^{N-m}f(x_{m:N})$$

To estimate the distribution function, we associate a probability $P\{x \le x_{m:N}\}$ to each observed rank m. Since the probability of event $\{x \le x_{m:N}\}$ is controlled by *two* variates, $x_{m:N}$ and x, $P\{x \le x_{m:N}\}$ is obtained by integrating their joint density function, $f_{x_{m:N},x}$, over the area $x \le x_{m:N}$, see Fig. 1.

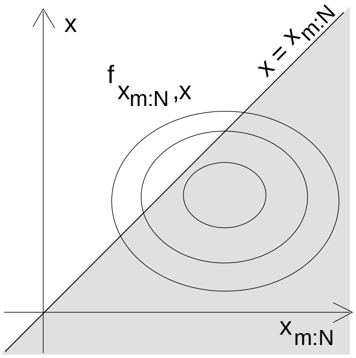

Fig. 1. Contours of $f_{x_{m:N},x}$ in $x_{m:N}$ -x plane. Integration of $f_{x_{m:N},x}$ over the grey half plane gives probability $P\{x \le x_{m:N}\}$.

Due to the mutual independence of $x_{m:N}$ and x, their joint density function $f_{x_{m:N},x}(x_{m:N},x)$ equals $f_{m:N}(x_{m:N})f(x)$. Consequently,

$$P\{x \le x_{M:N}\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{x_{m:N}} f_{m:N}(x_{m:N})f(x)\,dx\,dx_{m:N}$$

$$= \int_{-\infty}^{+\infty}\left[\int_{-\infty}^{x_{m:N}} f(x)\,dx\right] f_{m:N}(x_{m:N})dx_{m:N}$$

$$= \int_{-\infty}^{+\infty} F(x_{m:N})f_{m:N}(x_{m:N})dx_{m:N}$$

$$= \int_{-\infty}^{+\infty} F(x_{m:N})m\binom{N}{m}F(x_{m:N})^{m-1}\left(1-F(x_{m:N})\right)^{N-m}f(x_{m:N})dx_{m:N}$$

$$= \binom{N}{m}\int_0^1 F^m(1-F)^{N-m}dF = \cdots = \frac{m}{N+1}$$

3. Sample error and plotting positions

The term *rank probability* used by Makkonen is equivalent to $P\{x \leq x_{m:N}\} = m/(N+1)$. This is different from $P\{x \leq X_{m:N}\} = F(X_{m:N})$ which equals $P_{m:N}$ in Cook's notation [5]. Figure 2 illustrates the difference. In x-direction, a point $(X_{m:N}, \frac{m}{N+1})$ deviates from the curve $F(x)$ by

$$X_{m:N} - F^{-1}\left(\frac{m}{N+1}\right)$$

and the sampling error in $X_{m:N}$ completely explains the deviation. There is no reason to acknowledge the standard error in $m/(N+1)$ as required by Cook [5] because such an error does not exist.

To conclude, Makkonen's [1-4] claim that $m/(N+1)$ is an exact, unique and distribution free expression for the probability $P\{x \leq x_{m:N}\}$ is correct. The sampling error in the plot is completely attributable to the dispersion of x which is reflected to $x_{m:N}$ and $X_{m:N}$.

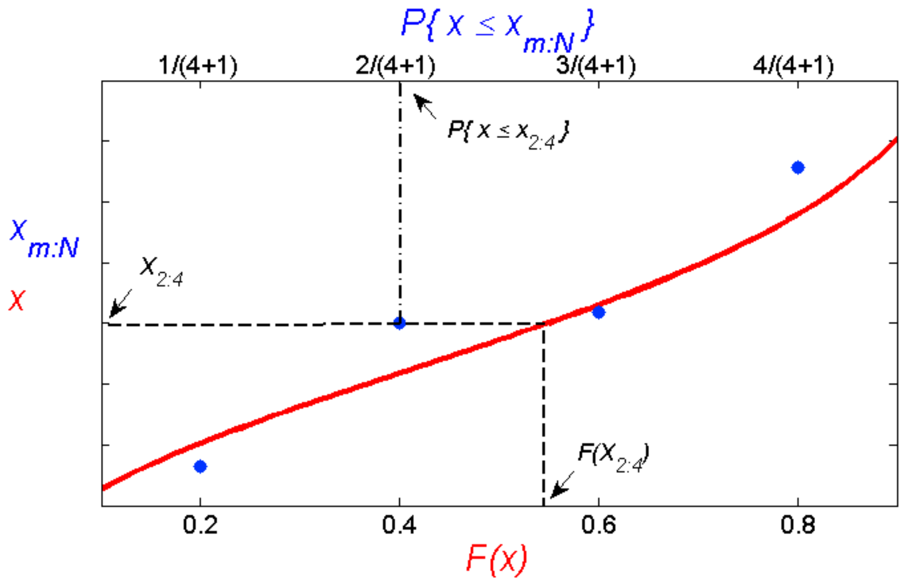

Fig. 2. Example of probability plotting with $N = 4$. The unknown distribution $F(x)$ is shown in red colour. The observations $X_{m:4}$ (in blue) are plotted at their rank probability. $F(X_{m:N}) = P\{x \leq X_{m:N}\}$ is random. However, the rank probability equals $m/(N+1)$ and is deterministic.

It follows that any other plotting position is a biased estimate of $P\{x \leq x_{m:N}\}$. This is demonstrated in Figure 3 by Monte-Carlo simulations. Figure 3 confirms that using Weibull's plotting positions the relative frequency converges to the underlying distribution. Using Gringorten's estimates results in a systematic error particularly at the tails. A similar result is obtained for all underlying distributions.

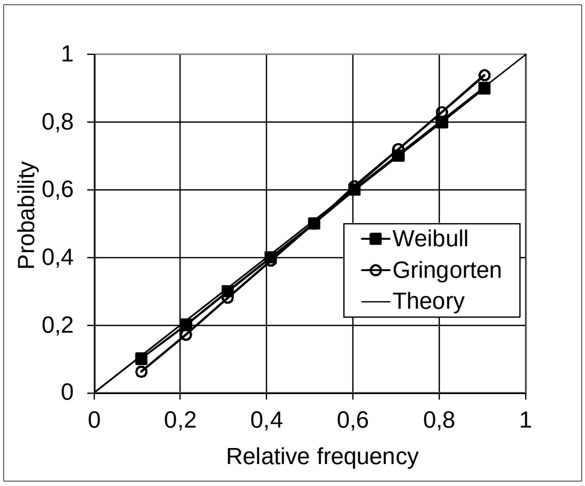

Fig. 3. Frequency of non-exceedance of $x_{m:9}$ in Monte-Carlo simulation with underlying Gumbel distribution. A sample of nine values has been order-ranked and a tenth random value has been taken and compared with the order-ranked values. This procedure has been repeated for 10 000 times. Finally, the observed cumulative relative frequencies of non-exceedance have been plotted by Weibull's and Gringorten's [7] plotting positions.

In the Section "Establishing a consistent notation" Cook [5] writes: "... probability, *P*, has a cumulative distribution **P**(*x*)". Such a concept is alien to the theory of probability. It follows directly from the axioms of the probability theory that, there is a unique real number assigned to every event, called the probability *P*. Thus, *P* is not a variable but an *invariable property of the random process*. When *x* is a variate, *F(x)* can be considered a variate, too. Nevertheless, the probability function *F* and the rank probabilities remain unchanged (see Fig. 2).

In connection with his Fig. 1 Cook [5] uses a concept of "expectation of *R*". However, the return period $R = 1/(1\text{-}P)$ is a non-additive variable so that *E(R)* has no meaningful interpretation. Moreover, the data given in Fig. 1 of [5] are not the same as in Makkonen's

[3] Fig. 1, but are obtained by substituting Gringorten's plotting positions to the Gumbel distribution and then solving for the wind speed. On Gumbel paper *any points* (P,-ln(-ln$P$)) where P is a random number between 0 and 1 result in a straight line. Therefore, Fig. 1 in [5] is unrelated to the correctness of plotting positions.

In Fig. 3 in [5], an example on what happens when the sample size N increases is considered. This example proves nothing of the problem at hand because the expression $m/(N+1)$ already shows that the plotting positions change with N.

Consequently, Cook's conclusions based on the figures and table in [5] are invalid. We also note that his criticism on Fig. 1 in Makkonen [3] is unsubstantiated. The probabilities in that figure are based on N= 21 and are, therefore, correctly plotted also for the six top ranks.

4. Conclusions

Cook [5] claimed to have demonstrated that "the prescriptive approach to EVA advocated in Makkonen [1-4] is based on a fundamental fallacy". Here it was shown that this claim is invalid. A proof, based on the probability theory, was given for $P\{x \le x_{m:N}\} = m/(N+1)$. This shows that the Weibull expression $m/(N+1)$ is both exact and unique and that the sampling error only exists in the observed variate value $X_{m:N}$.

Acknowledgements

We wish to thank Academy of Finland for support via the Recast-project.

Notations

$E(z)$	expectation of variate z
F	cumulative distribution function of x (underlying distribution)
$F(X_{m:N})$	$P\{x \le X_{m:N}\}$, value of F calculated at $x = X_{m:N}$
$F_{m:N}$	$F(X_{m:N})$, variate, the values of which are determined by the $X_{m:N}$ values in successive order-ranked samples of size N

$P\{A\}$	probability of event A
X	observed value of x in a sample
$X_{m:N}$	mth order-ranked value in a sample including N observations of x
f	probability density function of x
$f_{x:y}$	joint probability density function of variates x and y
$f_{m:N}$	probability density function of $x_{m:N}$
m	index of m^{th} order-ranked value , $1 \leq m \leq N$
N	size of a sample
$x_{m:N}$	variate, the sampled values of which are denoted by $X_{m:N}$

References

[1] Makkonen L. Plotting positions in extreme value analysis. J Appl Meteorol Climatol 2006;45:334-340.

[2] Makkonen L. Bringing closure to the plotting position controversy. Comm Stat – Theor Meth 2007;37(3):460-7.

[3] Makkonen L. Problems in the extreme value analysis. Struct Safety 2008;30:405-19.

[4] Makkonen, L. Reply to “Comments on “Plotting positions in extreme value analysis” by N. Cook”. J Appl Meteorol Climatol 2011;50:267-70.

[5] Cook, NJ. Rebuttal of “Problems in the extreme value analysis. Struct Safety 2011;34:418-23.

[6] Madsen HO, Krenk S, Lind NC. Methods of structural safety. Prentice-Hall; 1986.

[7] Gringorten II. A plotting rule for extreme probability paper. J Geophys Res 1963;68:813-814.