Hazard function theory for nonstationary natural hazards

By L. Read and R.M. Vogel

Nat. Hazards Earth Syst. Sci. Discuss.
MS-NR: nhessd-3-6883-2015

REMARKS

After reading the comment by Dr. Prosdocimi, I would like to share my opinion on a key point of the present manuscript. Since I had the chance to review the paper by Read and Vogel (2015b), which is still under review as far as I know, I can say that I had the same doubts about the equality $p_0=h(t)$. In particular, in that paper the Authors introduce this equality as a definition, whereas the identity $p_0=h$ is a special case holding true only for iid data (exponential inter-arrival times), whose extension to independent and "non-identically" distributed data seems to be not so straightforward. Referring to that paper (but this comment holds also for the present manuscript), my arguments on this point are as follows.

Given a known or estimated time dependent model for flow intensity X (peaks over thresholds or annual maxima), $F_X(x;\theta(t))$, where θ is a generic parameter vector, the Authors attempt to deduce the distribution of the waiting times, $F_T(t)$, for the next exceedance of a given value $X = x_0$, exploiting the hazard function to link $F_X(x;\theta(t))$ and $F_T(t)$. Now, the hazard function <u>is defined</u> as

$$h(t) = \frac{f_T(t)}{1-F_T(t)} = -\frac{\frac{dS_T(t)}{dt}}{S_T(t)} \quad (1)$$

where f_T is the probability density function of the waiting time, F_T is the corresponding cumulative distribution function, and $S_T = 1\text{-}F_T$ is the survival function, which is also known as reliability, and gives the probability that the system experiences no failure within $(0, t]$.
From the above <u>definition</u> it follows that the cumulative hazard function is

$$H(t) = \int_0^t h(s)ds = \int_0^t -\frac{dS_T(s)}{S_T(s)} = -\ln\left(S_T(t)\right) \quad (2)$$

so that

$$S_T(t) = \exp\left(-\int_0^t h(s)ds\right) = \exp\left(-H(t)\right) \quad (3)$$

For iid (independent and identically distributed data), regardless of the form of $F_X(x;\theta)$, it is known that the waiting times of exeedances over a given quantile (high) threshold x_0 are memoryless and follows an exponential distribution with rate parameter $\lambda=p_0$, where $p_0= S_X(x_0;\theta)$ is the probability of exceedance corresponding to x_0. Under exponential arrivals, the hazard function is constant,

$$h(t) = \frac{f_T(t)}{1-F_T(t)} = \frac{\lambda \exp(-\lambda t)}{\exp(-\lambda t)} = \lambda = p_0 \quad (3)$$

From this very specific result ($h(t) = \lambda = p_0 = S_X(x_0;\theta)$), which holds true under iid conditions), the Authors deduce that it still holds true in the form $h(t) = p_t = S_X(x_0;\theta(t))$ under independent and non-identically distributed conditions. This is actually the core of the paper. i.e. establishing the advocated link between $F_X(x;\theta(t))$ (or $S_X(x;\theta(t))$) and $F_T(t)$ (or $S_T(t)$). However, this is not a definition as stated in Read and Vogel (2015b), but an assumption that requires to be verified analytically or numerically, or both. It cannot be a definition because the hazard function is already defined as in Eq. (1), and a different definition would imply that the relationship between $h(t)$ and $S_T(t)$ in Eq. 3-above is no longer necessarily true, thus preventing its application. In more detail, if the Authors' assumption is true, in order Eq. 3-above to be applicable, it should be

$$S_T(t) = \exp\left(-\int_0^t h(s)ds\right) = \exp\left(-\int_0^t S_X(x_0;\theta(s))ds\right) \quad (4)$$

However, I cannot see any reason why S_X and S_T should be linked by this relationship, which holds true only if $h(t)$ is defined as in Eq. 1 so that the integral in Eq. 2 holds true.

To summarize, I think that the key point in order to make results convincing is to show the validity of Eq. 3 above, or similarly, the identity (or approximate identity) of the assumed theoretical hazard functions and the actual hazard function resulting from simulations, under non iid conditions for whatever model $F_X(x;\theta(t))$. If this hypothesis is not verified, all the framework provides no advantages, as it would require simulations from $F_X(x_0;\theta(t))$ to obtain quantities such as $F_T(t)$; however, this procedure does not need time-to-failure analysis and hazard related concepts.

Sincerely,

Francesco Serinaldi