Fitting Wakeby model using maximum likelihood

Stima di massima verosimiglianza della distribuzione Wakeby

Agostino Tarsitano
Dipartimento di Economia e Statistica - Università degli Studi della Calabria
agotar@unical.it

Summary. Si affronta il problema della stima di verosimiglianza di una variabile casuale espressa attraverso la sua funzione quantile. In particolare si segure l'approccio sviluppato per i dati raggruppati in classi. L'efficacia del metodo è illustrata su due casi concreti. Si adopera poi delle simulazioni per confrontarne l'efficienza con quella dei momenti pesati.

Keywords: probability weighted moments, flood frequency analysis.

1. Introduction

Frequency curves are often used for modelling extreme floods. To be a reasonable candidate, a distribution has to be able to accommodate positively skewed histograms. Moreover, since the form of empirical distributions encountered in hydrology is found to vary considerably, it seems likely that a model with four or five parameters would be needed to describe the data. The Wakeby distribution (WAD) has gained widespread use due to its reliability in curve fitting, although the estimation of its parameters can be very difficult. For this purpose, the standard technique is the method of PW-moments which sometimes shows convergence failure. The maximum likelihood method (ML) for the Wakeby is computationally demanding because of the complex and unwieldy objective function, but offers a well established theoretical background. In fact, it is common to use the asymptotic properties of the ML estimates to construct confidence regions for parameters. The objective of this paper is to show that the ML estimates can be implemented in a straightforward manner even for a random variable described by its quantile function.

2. The Wakeby random variable

The WAD is specified by the quantile function

$$X(p,\lambda) = \lambda_1 - \left(\lambda_2 q^{\lambda_4} + \lambda_3 q^{\lambda_5}\right) \quad 0 \le p \le 1,\ q = 1 - p \tag{1}$$

where λ_1 is a location parameter, λ_2, λ_3 are linear parameters prevalently related to the scale of the variable and λ_4, λ_5 are exponential parameters determining the shape of the quantile function.The WAD has a finite lower bound $\lambda_1 - (\lambda_2 + \lambda_3)$. The upper bound is λ_1 if $\lambda_4 > 0$ and $\lambda_5 > 0$ and infinite if λ_4 or λ_5 or both are negative. The probability density function is defined implicitly by the density quantile function that is the density expressed in terms of the cumulative probability p.

$$\frac{1}{\frac{dX(p;\lambda)}{dp}} = h[X(p;\lambda)] = \left[\lambda_4\lambda_2 q^{\lambda_4-1} + \lambda_5\lambda_3 q^{\lambda_5-1}\right]^{-1} = h[p;\lambda] \tag{2}$$

The regions in which (2) is a valid density functions are

$$\begin{array}{ll} R1: \lambda_4\lambda_2 > 0, \lambda_5\lambda_3 \geq 0 & R2: \lambda_4\lambda_2 \geq 0,\ \lambda_5\lambda_3 > 0 \\ R3: \lambda_4\lambda_2 < 0, \lambda_5\lambda_3 > |\lambda_4\lambda_2|, \lambda_4 \geq \lambda_5 & R4: \lambda_5\lambda_3 < 0, \lambda_4\lambda_2 > |\lambda_3\lambda_5|, \lambda_4 \leq \lambda_5 \end{array} \tag{3}$$

The parameters λ_4 and λ_5 determine the type of tails of the WAD. For example, if $\lambda_4,\lambda_5 > 0$ then (2) has increasingly peakedness and short tails; if λ_4, $\lambda_5<0$ the tails have increasingly heaviness. The derivative of the quantile-density function (2) is

$$h'(p,\lambda) = h^3(p,\lambda)A(p,\lambda); \quad A(p,\lambda) = \lambda_2\lambda_4(\lambda_4-1)q^{\lambda_4-2} + \lambda_3\lambda_5(\lambda_5-1)q^{\lambda_5-2} \tag{4}$$

If p_c is such that $h'(p_c, \lambda)=0$ then

$$h''(p_c;\lambda) = -h^3(p_c;\lambda)D(p_c,\lambda); \quad D(p,\lambda) = \lambda_2 q^{\lambda_4-3}\prod_{i=0}^{2}(\lambda_4 - i) + \lambda_3 q^{\lambda_5-3}\prod_{i=0}^{2}(\lambda_5 - i) \tag{5}$$

Consequently, the density is unimodal if $(\lambda_4, \lambda_5>2)\cap(\lambda_2, \lambda_3>0)$ or if $(\lambda_4, \lambda_5<0)\cap(\lambda_2, \lambda_3<0)$ or if $(0<\lambda_4, \lambda_5<1)\cap(\lambda_2, \lambda_3>0)$. The density is zeromodal if $(Min\{\lambda_4, \lambda_5\}<0)\cap(\lambda_2, \lambda_3>0)$ or if $(Max\{\lambda_4\ \lambda_5\}>1)\cap(\lambda_2, \lambda_3>0)$ or if $(0<\lambda_4,\lambda_4<1)\cap(\lambda_2\lambda_3<0)$. Consider the linear transformation $Z=X-\lambda_1$. The expected value of Z^i is given by

$$\mu_i(Z) = \sum_{j=0}^{i}\binom{i}{j}\left[i\lambda_4 + 1 - j(\lambda_4-\lambda_5)\right]^{-1}\lambda_2^{i-j}\lambda_3^{j} \tag{6}$$

Since $Z-E(Z)=X-E(X)$ the central moments of X coincide with the central moments of Z. In particular, the first moment exists finite if $\lambda_4 \neq -1$ and $\lambda_5 \neq -1$; $\sigma^2(X)<\infty$ if $(\lambda_4, \lambda_5>-0.5)\cap(\lambda_4+\lambda_5>1)$. The i-th probability-weighted moment $E\{X[1-F(x)]^i\}$ has the simple expression

$$\tau_i(\lambda) = (i+1)^{-1}\lambda_1 - (i+1+\lambda_4)^{-1}\lambda_2 - (i+1+\lambda_5)^{-1}\lambda_3, \quad i = 0, 1, \ldots \tag{7}$$

3. Method of maximum likelihood (ML)

Let $\{X_1, X_2, \ldots, X_n\}$ be a simple random sample drawn from a WAD. The negative log-likelihood function is $S_1(\lambda) = -\sum n_i Ln[h(p_i, \lambda)]$ where $p_i=F(X_i,\lambda)$ and F is cumulative distribution function of the WAD. The criterion $S_1(\lambda)$ can be minimized over λ through application of the scoring method. An alternative could be the minimization over for λ_2, λ_3, λ_4, λ_5; an estimate for λ_1 is then obtained by $\lambda_1 - (\lambda_2 + \lambda_3) = X_{min}$. However, we have left this as a check of the results. The gradient and the information matrix are:

$$g_r(\lambda)=\sum_{i=1}^{n} h^3(p_i,\lambda)A(p_i,\lambda)X_r^{'}(p_i,\lambda) \quad where \quad X_r^{'}(p_i,\lambda)=\left[\partial X_r^{'}(p,\lambda)/\partial\lambda_r\right]_{p=p_i}$$

$$W_{r,s}(\lambda)=-\sum_{i=1}^{k} h^4(p_i,\lambda)X_s^{'}(p_i,\lambda)\left\{X_r^{'}(p_i,\lambda)\left[3A^2(p_i,\lambda)h^2(p_i,\lambda)-D(p_i,\lambda)\right]+\right. \tag{8}$$

$$\left.+A(p_i,\lambda)B_r(p_i,\lambda)\right\} \quad with \quad B_r(p_i,\lambda)=\left[dX_r^{'}(p,\lambda)/dp\right]_{p=p_i}$$

Given a preliminary estimate $\lambda^{\mathbf{0}}$ for λ then to obtain ML estimates we use the recursion $\lambda^{m+1}=\lambda^{m}-B_m\,\delta(\lambda^{m})$ with $\delta(\lambda^{m})=[W(\lambda^{m})+\gamma_m I]^{-1}g(\lambda^{m})$. B_m is a diagonal matrix of correction factors which fixes the step length of each parameter and γ_m is a positive scalar large enough to make $[W(\lambda^{m})+\gamma_m I]$ positive definite when $W(\lambda^{m})$ it is not. The nonzero elements of B_m are selected performing a systematic search along the Newton direction by means of Faure sequences. More precisely, sequences of 1'000 points in $[-\delta(\lambda^{m}),\delta(\lambda^{m})]$ are examined and the search is interrupted whenever $S_1(\lambda^{m+1})<0.9999*S_1(\lambda^{m})$. Since the computation of p_i cannot be directly carried out we have used a combination of bisection and Newton's method to solve $X_i=X(p_i,\lambda)$ for $p_i>p_{i-1}$ and $p_0=0$. The algorithm has been applied to two data sets reported in Haktanir and Bozduman (1995) and Haktanir (1997) of $n=50$ and $n=51$ annual flood peaks values respectively. It must be said that when two different data were recorded as the same value we have added 0.01 to one of the observations. The results have been compared with those obtained by the PWM method. These estimates were determined by computing the first five sample PW-moments

$$t_i=n^{-1}\sum_{j=1}^{n}\left(1-P_j\right)^i x_{(j)},\; i=0,1,\ldots,\; with\; P_j=\prod_{r=1}^{i}(j-r)\Big/\prod_{r=1}^{i}(n-r) \tag{9}$$

setting them equal to (7) and solving $\tau_i(\lambda)=t_i;\ i=0,1,\ldots,4$ for λ. To this end, we have applied the downhill simplex minimization to $S_2(\lambda)=Max\{\,|\tau_i(\lambda)-t_i|\ i=0,\ldots,4\}$. This is a well-known derivative-free optimization algorithm due to Nelder and Mead. It requires only function evaluations and has a wide applicability for general function minimization. Table 1 below shows estimates obtained by the ML method and by the PWM method. The last column reports the mean standard error of prediction $MSE(\lambda)=\{n^{-1}\sum[x_i-X(\lambda,p_i)]^2\}^{0.5}$. The starting value $\lambda^{\mathbf{0}}$ for the two methods has been determined by examining Faure sequences of 100'000 points over the relevant parameter intervals.

Table 1: parameter estimation for Wakeby distribution

Dataset	Method	λ_1	λ_2	λ_3	λ_4	λ_5	M.S.E.
2001,Kilavuzlu	M.L.	154.91543	-17.68096	-9.48047	-2.10804	-0.14043	0.0000016
	P.W.M.	683.17209	545.34247	-42.80659	1.19854	-0.63382	0.0000017
902,Beskonak	M.L.	210.75671	-139.11622	69.88650	-1.34346	1.07023	0.0000003
	P.W.M.	487.09605	653.79794	-441.41528	1.46537	-0.33784	0.0000007

To limit the evaluations of $S_1(\lambda)$ and $S_2(\lambda)$ in the required region, the algorithm handled the constraints (3) by setting the objective function equal to 10^{42} whenever an inequality was violated.ML represents a slight improvement over the PW for the two datasets. The most obvious point to be noticed is that we can have multiple solutions (with both the methods). When we analyze data which conform to different type of curves, the estimates of the parameters can differ according to which method of estimation is used. This is not surprising since the large number of parameters in WAD leaves a considerable variation possible. However, the possibility of multiple solutions it is not necessarily a problem, but can be an opportunity.

4. A Monte Carlo experiment

N=2'000 random samples of size $n \in \{30,50,70\}$ were obtained by first generating a pseudo random number $q \in \{0,1\}$ and then inserting q in (1) to determine the "observed" peak flow X. The parameter combination is $\lambda_{1,0}=200$, $\lambda_{2,0}= -70$, $\lambda_{3,0}=50$, $\lambda_{4,0}= -0.5$, $\lambda_{5,0}=1.5$. For each generated sample the ML and PWM estimates of λ were compared to the true value. To initialize the iteration process from a sufficiently good value we took λ_0.Two simple coefficients of performance have been considered for comparison: the mean relative bias: $Mrb(\lambda_i)=\{N^{-1}\sum|\lambda_{i,j}-\lambda_{i,0}|\lambda_{i,0}\}$ and the standard deviation: $Std(\lambda_i)=\{N^{-1}\sum[\lambda_{i,j}-m_i]^2\}^{0.5}$ where $m_i=\{N^{-1}\sum\lambda_{i,j}\}$. Table 2 shows our findings.

Table 2: simulation results

	n		λ_1	λ_2	λ_3	λ_4	λ_5
Mrb	30	ML	0.016918	0.000001	0.000002	0.076525	0.196644
		PWM	0.098008	0.306295	0.325835	0.231257	0.345402
	50	ML	0.012316	0.000002	0.000003	0.059768	0.147888
		PWM	0.080877	0.249158	0.268090	0.197745	0.277388
	70	ML	0.007839	0.000002	0.000003	0.043824	0.119267
		PWM	0.068840	0.211989	0.222257	0.171248	0.239008
Std	30	ML	0.021383	0.000005	0.000007	0.106755	0.325108
		PWM	0.140863	0.339405	0.342661	0.291862	0.623882
	50	ML	0.056726	0.000010	0.000016	0.157737	0.232244
		PWM	0.116547	0.289995	0.301015	0.249937	0.450909
	70	ML	0.015156	0.000006	0.000008	0.061116	0.193209
		PWM	0.098444	0.247174	0.248242	0.213486	0.387857

The ML method has given results which are decidedly better than those obtained by PWM both with respect to the bias and with respect to the standard deviation for all the type of parameters (linear and exponential). It must be said that the ML estimates are computationally demanding, but not by an amount of likely practical importance in the age of ample availability of fast computers.

It is well known that ML does not always give satisfactory results for models in which one or more of the parameters corresponds to a limit on the range of the variable. The estimation problem for the WAD when the regularity conditions for the ML estimates do no apply has not yet been rewieved and discussed.

References

Haktanir T., Bozduman A. (1995): A study on sensitivity of the probability-weighted moments method on the choice of the plotting position formula. Journal of Hydrology, 168, 265-281.

Haktanir T. (1997): Self-determined probability-weighted moments method on the choice of the plotting position formula. Journal of Hydrology, 194, 180-200.