Il test del massimo valore

Per dimostrare che in alcuni casi la legge del valore estremo EV1 non è adeguata a descrivere le scie degli estremi idrologici, si può far ricorso ad un test molto semplice, nel quale la statistica di riferimento è il valore massimo tra i dati del campione.
L'ipotesi H_0 da sottoporre a verifica è: "Il valore massimo X_n appartiene alla distribuzione dei massimi di una variabile di Gumbel"?
Se la risposta è positiva ne deriva la necessaria conseguenza che quel valore appartiene anche alla distribuzione di partenza (Gumbel) che lo avrebbe generato.
Per la statistica X_n è semplice ricavare per via diretta la distribuzione. Infatti, come risulta dal procedimento di costruzione della distribuzione del massimo di n osservazioni di una variabile casuale, si ha:

$$F_{X_{(N)}} = [F_X(x)]^N$$

Se la variabile $F_X(x)$ è una Gumbel, la distribuzione del massimo di un campione di N dati estratti da una Gumbel è pertanto:

$$F_{X_{(N)}}(x) = \left[e^{-e^{-\frac{1}{\theta_2}(X-\theta_1)}} \right]^N$$

L'espressione precedente è ancora una legge di Gumbel. Si vede infatti che:

$$e^{-Ne^{-\frac{1}{\theta_2}(X-\theta_1)}} = e^{-e^{\ln N}e^{-\frac{1}{\theta_2}(x-\theta_1)}}$$

per cui, ipotizzando una distribuzione di Gumbel con un nuovo valore modale ϑ_1^*, questo risulta dall'equivalenza:

$$e^{-e^{-\frac{1}{\theta_2}\left(X-\theta_1^*\right)}} = e^{-e^{-\frac{1}{\theta_2}(X-\theta_1)+\ln N}}$$

da cui risulta ancora:

$$-\frac{1}{\theta_2}\left(X-\theta_1^*\right) = -\frac{1}{\theta_2}\left(x-\theta_1\right)+\ln N$$

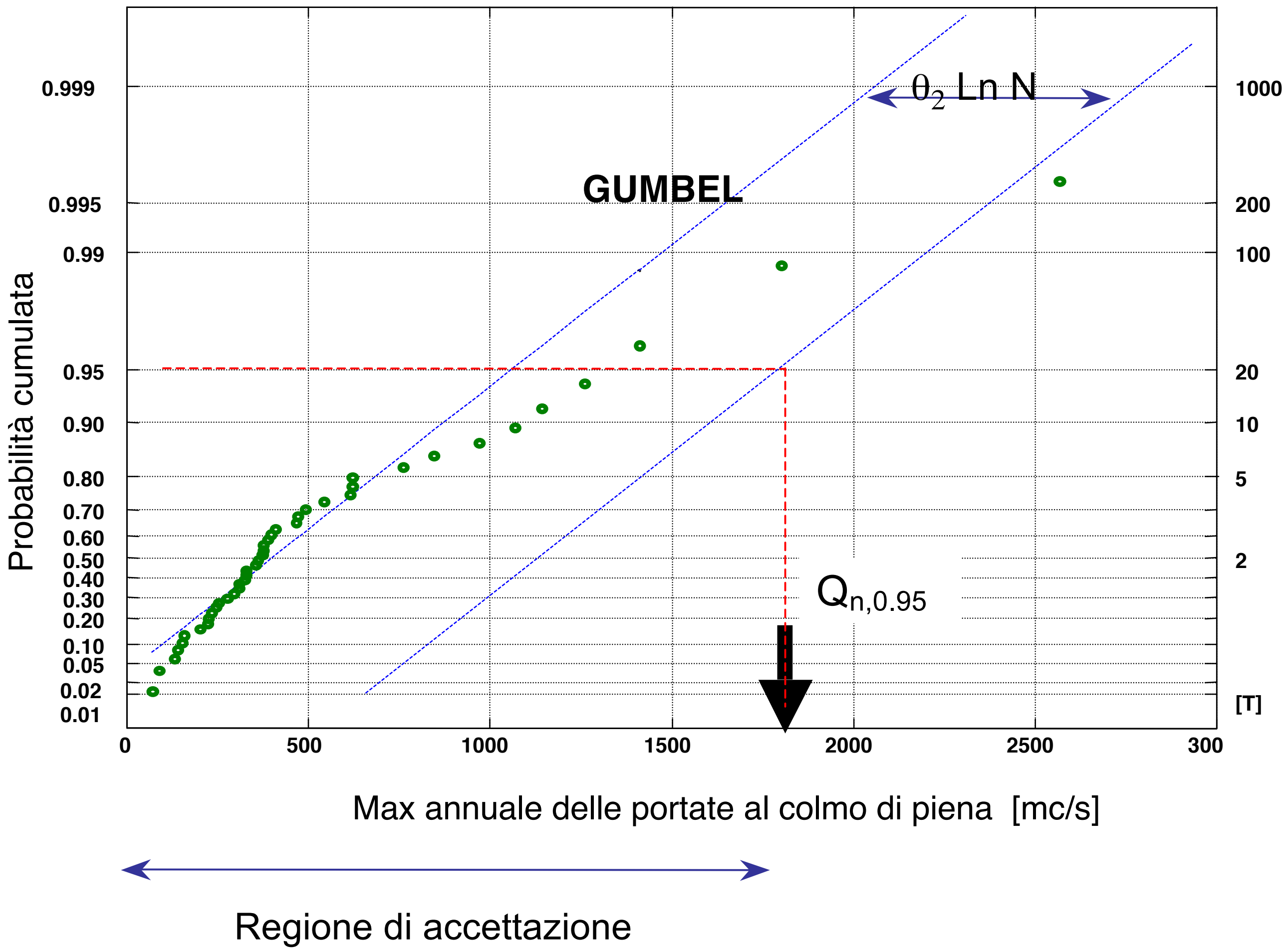
$$x-\theta_1^* = x-\theta_1-\theta_2 \ln N$$

$$\theta_1^* = \theta_1 + \theta_2 \ln N$$

La nuova distribuzione è quindi ancora una Gumbel, ma con parametro ϑ_1 incrementato:

$$F_{X_{(N)}}(x) = e^{-e^{-\frac{1}{\theta_2}(x-[\theta_1+\theta_2 \ln N])}}$$

In carta probabilistica di Gumbel ciò equivale a traslare la retta di Gumbel originaria della quantità $\theta_2 \ln N$.



Nota quindi la distribuzione della statistica test se ne deduce l'intervallo di accettazione, in funzione del livello di significatività α. Se α = 5%, trattandosi di un controllo per estremi positivi, si lascia l'errore solo da un lato, per cui risulta che il limite di accettazione positivo per $X_{(N)}$ è X($F_{X_{(N)}}$ =0.95). Di conseguenza, se il valore max osservato su N dati risulta essere superiore al valore limite al 95% dei massimi su N dati estratti da una Gumbel, tale distribuzione mostra di essere inadatta a rappresentare il campione, almeno nella sua coda positiva. Si rende quindi necessario far ricorso a distribuzioni più asimmetriche.

Applicazione del test a distribuzioni qualsiasi

Si immagina che il valore osservato di $X_{(N)}$ corrisponda al valore X_{LIM} . La condizione limite richiesta dal test è che

$$F_{X_{(N)}}(X_{LIM}) = 1 - \alpha$$

Essendo $F_{X_{(N)}} = \left[F_X(x)\right]^N$ vale allora

$F_X(X_{LIM}) = (1-\alpha)^{1/N}$, ovvero:

per $X=X_{LIM}$ deve valere:

$$[F_X(X_{LIM})]^N = 1 - \alpha \quad .$$

Di conseguenza, dato il valore $X_{(N)}$ osservato del massimo, il test corrisponde a verificare se:

$F_X(X_{(N)})]^N < 1-\alpha$ Ipotesi sul test soddisfatta

contro l'ipotesi alternativa:

$F_X(X_{(N)})]^N > 1-\alpha$ Ipotesi sul test non soddisfatta

Valutazione del periodo di ritorno di un nuovo evento "eccezionale"

Un'applicazione pratica del test del massimo valore consente di riconoscere in modo corretto il periodo di ritorno associabile ad un nuovo evento di intensità particolarmente elevata. La circostanza è quanto mai frequente, dal momento che si presenta ogni volta che un nubifragio viene definito 'eccezionale', e se ne voglia determinare la rarità.

In sostanza, quindi, osservato un nuovo evento'eccezionale' $X_{(N+1)}$, tale da essere di entità piu' elevata del precedente valore massimo, si vuole assegnare ad esso un valore oggettivo di rarità, attraverso l'attribuzione del periodo di ritorno.

Si fa riferimento ad una serie storica registrata nella stessa stazione nella quale si è registrata la nuova osservazione. La procedura usuale prevede l'impiego della distribuzione di probabilità della variabile di interesse X ed il suo impiego nella determinazione di $T(X_{(N+1)})$.

Ci sono due possibilità:

a) $X_{(N+1)}$ appartiene a $F_X(x)$

b) $X_{(N+1)}$ non appartiene a $F_X(x)$

Per stabilirlo si applica il test del massimo valore al nuovo dato, utilizzando la distribuzione $F(X)$ ottenuta **senza includere** nella serie storica analizzata il dato $X_{(N+1)}$. Questo è il caso più tipico nell'ambito dell'attribuzione del T ad una nuova osservazione. E' ciò che si mette in atto quando si estraggono le curve di probabilità pluviometrica da una cartografia già realizzata (es. AdB Po, Linee Segnalatrici).

Affinchè il test sia positivo (caso 1) deve quindi risultare :

$$\left[F_X\left(X_{N+1}\right)\right]^{N+1} < 1-\alpha$$

Se ciò accade il periodo di ritorno potrà essere determinato come:

$$T(X_{N+1}) = \frac{1}{1-F_X\left(X_{N+1}\right)}$$

Se invece il test ha esito negativo, la $F_X(x)$ non può dirsi rappresentativa per il valore $X_{(N+1)}$. Di conseguenza si dovranno nuovamente stimare i valori dei parametri della $F_X(x)$ includendo il nuovo valore. Ne risulterà una nuova funzione di probabilità cumulata $F^*_X(x)$, i cui parametri saranno stimati su N+1 osservazioni.

A questo punto si riapplica il test di adattamento (del massimo valore) e sono possibili due esiti:

i) Il test è soddisfatto. Di conseguenza è possible attribuire in modo corretto il periodo di ritorno ad $X_{(N+1)}$ mediante la relazione: $T(X_{N+1}) = \dfrac{1}{1-F_X^*\left(X_{N+1}\right)}$

ii) Il test non è soddisfatto. Ciò significa che la distribuzione utilizzata non è più adatta alla rappresentazione dell'intero campione. E' necessario a questo punto formulare una nuova ipotesi di modello probabilistico che porti il test ad avere esito positivo.