Verifica e scelta del modello probabilistico

Introduzione

L'elaborazione statistica dei dati comporta un certo numero di ipotesi, quali ad esempio la forma della distribuzione ed il metodo utilizzato per stimare i parametri. Data una qualsiasi ipotesi statistica H_0, occorre misurarne la validità di fronte all'ipotesi alternativa H_1, anche se non esplicitamente formulata. Un test statistico è un procedimento che consente di decidere, sulla base delle osservazioni a disposizione, se accettare l'ipotesi H_0 oppure rigettarla. Si dice livello di significatività del test la probabilità α di rigettare l'ipotesi H_0 quando essa è vera (errore di tipo I), mentre si definisce potenza del test la probabilità di rigettare H_0 quando essa è falsa. Questa probabilità vale 1-β, dove β è la probabilità di commettere un errore di tipo II (accettare H_0 quando vale H_1). Si noti che, inevitabilmente, a parità di numero di osservazioni n la riduzione del livello di significatività abbassa la potenza del test.

Ovviamente esiste una varietà di test che rispecchia la varietà delle ipotesi da provare. Ci occuperemo dei test atti a provare l'ipotesi che una data variabile casuale sia distribuita secondo una assegnata funzione di probabilità, noti anche come test di adattamento. Essi consistono nel valutare l'adattamento di una legge probabilistica $F_X(x)$ ad un insieme di n osservazioni, ossia l'ipotesi H_0 che $F_X(x)$ sia la distribuzione di probabilità da cui è stato estratto il campione a disposizione.

Test basati sugli L-momenti

Questa categoria di test si basa sul fatto che, per ogni distribuzione con solo un parametro di posizione ed uno di scala, il coefficiente di L-asimmetria, $\tau_3 = \frac{L_3}{L_2}$, è univocamente definito, ossia ha un valore unico indipendentemente dai parametri. In particolare, si ha per la distribuzione normale $\tau_3 = 0$, e per la distribuzione di Gumbel $\tau_3 = \log(\frac{9}{8})/\log(2) \approx 0.1699$. Inoltre, la distribuzione dei coefficienti di L-asimmetria campionari tende asintoticamente ad una distribuzione normale, e la bontà dell'approssimazione normale con n piccoli è molto migliore di quella che si ha per il coefficiente di asimmetria classico. Queste proprietà possono essere sfruttate per costruire un test basato sulla distanza tra il coefficiente di L-asimmetria campionario $\hat{\tau}_3$ e quello teorico, nota la varianza di $\hat{\tau}_3$ per ogni distribuzione di interesse,

$$\mathrm{var}(\hat{\tau}_3) = \frac{1}{n}\left(0.1866 + \frac{0.8}{n}\right) \qquad \text{distribuzione normale}$$

$$\mathrm{var}(\hat{\tau}_3) = \frac{1}{n}\left(0.2326 + \frac{0.7}{n}\right) \qquad \text{distribuzione di Gumbel}$$

Dato un campione di dati, si determina $\hat{\tau}_3$ e si definisce la test statistic come

$$Z = \frac{\hat{\tau}_3 - \tau_3}{\sqrt{\mathrm{var}(\hat{\tau}_3)}}$$

Z avrà una distribuzione normale standardizzata (ossia con media zero e varianza unitaria) se l'ipotesi H_0 è vera. Per effettuare il test è quindi sufficiente confrontare il valore di Z con i quantili $1-\frac{\alpha}{2}$ e $1+\frac{\alpha}{2}$ di una distribuzione normale standardizzata (il test è a due code). Se $z_{1-\alpha/2} < Z < z_{1+\alpha/2}$ l'ipotesi H_0 è accettata con livello di significatività α, altrimenti H_0 è rigettata.

In particolare si hanno

$$Z_{norm} = \frac{\hat{\tau}_3}{\sqrt{\frac{1}{n}\left(0.1866 + \frac{0.8}{n}\right)}}$$

$$Z_{gumb} = \frac{\hat{\tau}_3 - 0.1699}{\sqrt{\frac{1}{n}\left(0.2326 + \frac{0.7}{n}\right)}}$$

che andranno confrontati, ad esempio, con i limiti –1.96 e +1.96 quando il livello di significatività è pari a 0.05.

Una controprova grafica della correttezza di applicazione della procedure può venire dal diagramma degli L-momenti (figura 4): se il punto corrispondente ad L-skew ed L-kurt empirici è vicino ai punti relativi alla Gumbel o alla normale è probabile che il test venga passato. Anche questa categoria di test può essere utilizzata per testare le distribuzioni log-normale ed EV2, log-trasformando preliminarmente i dati a disposizione e riapplicando la procedura vista sopra con il coefficiente di L-asimmetria dei dati campionari trasformati logaritmicamente. L'estensione a distribuzioni a tre parametri è invece complicata, dal momento che per tali distribuzioni non si avrebbe un unico valore di τ_3, ma una serie di valori diversi al variare del coefficiente di forma della distribuzione.

Il test di Pearson o del χ^2

Il test di Pearson richiede che il campo di esistenza della variabile x venga suddiviso in k intervalli che si escludono a vicenda. Se l'i-esimo intervallo è definito dagli estremi $x_{i\,\inf}$ ed $x_{i\,\sup}$, si avrà che la probabilità che un'osservazione qualsiasi ricada nell'i-esimo intervallo, se l'ipotesi H_0 è vera, vale

$p_i = F_X(x_{i\sup}) - F_X(x_{i\inf})$.

Il numero atteso di elementi nell'intervallo i-esimo, sempre se H_0 è vera, vale pertanto np_i. Il test di Pearson consiste nel confrontare tale numero con il numero di osservazioni che effettivamente ricadono nell'intervallo, n_i. Se si vogliono confrontare tutti gli intervalli contemporaneamente, si può utilizzare la grandezza statistica (detta "test statistic")

$$X^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}.$$

Al crescere di n la test statistic del test di Pearson è asintoticamente distribuita come una distribuzione del chi-quadrato con k-1 gradi di libertà, quando i parametri della distribuzione ipotetica non sono stati stimati dalle osservazioni. Il fatto che la distribuzione asintotica (del chi-quadrato) sia nota consente di effettuare il test andando a confrontare il valore della statistica empirica riscontrato per una data distribuzione ed un dato campione con quello della distribuzione teorica, ottenuto prendendo il quantile (1-α) della distribuzione del chi-quadrato con k-1 gradi di libertà, $\chi^2_{1-\alpha}(k-1)$. Se $X^2 < \chi^2_{1-\alpha}(k-1)$ si è nella regione di accettazione del test, altrimenti l'ipotesi H_0 viene rigettata al livello di significatività α. (vedasi Tabella 1)

Due problemi limitano l'applicazione del test di Pearson: il primo è legato alla soggettività nella scelta degli intervalli di classe, ed il secondo al fatto che i parametri sono stimati dalle stesse osservazioni che vengono sottoposte a test. Relativamente al primo punto, specifici studi hanno mostrato che la potenza del test viene massimizzata quando si scelgono classi equiprobabili, $p_1 = p_2 = ... = p_k = \frac{k}{n}$, con $k = 2 \cdot n^{0.4}$. Non conviene invece fissare il numero atteso di elementi in ogni classe a 5, $np_i = 5$, come talvolta suggerito, perché tale scelta andrebbe a scapito della potenza del test. I limiti della classe i-esima saranno allora quantili della distribuzione F,

$$x_{i\sup} = F_X^{-1}\left(\frac{i}{k}\right); \qquad x_{i\inf} = F_X^{-1}\left(\frac{i-1}{k}\right)$$

Ad esempio, per la distribuzione di Gumbel si hanno

$$x_{i\sup} = \theta_1 - \frac{1}{\theta_2}\ln\left(-\ln\left(\frac{i}{k}\right)\right); \qquad x_{i\inf} = \theta_1 - \frac{1}{\theta_2}\ln\left(-\ln\left(\frac{i-1}{k}\right)\right)$$

Si noti che in tal caso la statistica $\boldsymbol{X}^2$ assume la forma

$$X^2 = \frac{k}{n}\sum_{i=1}^{k}\left(n_i - \frac{n}{k}\right)^2$$

proporzionale alla varianza della variabile casuale n_i.

Il secondo e più importante problema del test di Pearson riguarda il caso in cui la distribuzione F abbia s parametri da stimare, e che questi siano stati stimati dalle osservazioni, come quasi sempre avviene in ambito idrologico. In tal caso la distribuzione asintotica della test statistic di Pearson non può essere definita con esattezza. Tutto quanto si può dire è che, se si utilizza il metodo della massima verosimiglianza, tale distribuzione è compresa tra quella del chi-quadrato con k-1 gradi di libertà e quella del chi-quadrato con k-s-1 gradi di libertà. Si ha quindi una ineludibile incertezza nel definire i limiti di accettazione del test.

In particolare si avrà che il test è sicuramente passato se $X^2 < \chi^2_{1-\alpha}(k-s-1)$, mentre l'ipotesi H_0 va rigettata se $X^2 > \chi^2_{1-\alpha}(k-1)$. Quando invece $\chi^2_{1-\alpha}(k-s-1) < X^2 < \chi^2_{1-\alpha}(k-1)$, si ha che il test non è in grado di fornire una risposta univoca. Quando invece della massima verosimiglianza si utilizzano altri metodi di stima dei parametri, quali il metodo dei momenti o degli L-momenti, non si può neanche più affermare che la distribuzione asintotica è compresa tra quella del chi-quadrato con k-1 gradi di libertà e quella del chi-quadrato con k-s-1 gradi di libertà, e diventa quindi assai complicato effettuare il test. Se a questo si aggiunge che il test di Pearson tende in generale ad essere molto meno potente di altri test a parità di dimensione campionaria, a causa del fatto che i dati vengono raggruppati per costruire la test statistic, si comprende che l'utilizzo di tale test andrebbe evitato quando la dimensione campionaria è minore di un centinaio di elementi ed i parametri stimati sono più di due, ossia nella quasi totalità delle applicazioni idrologiche.

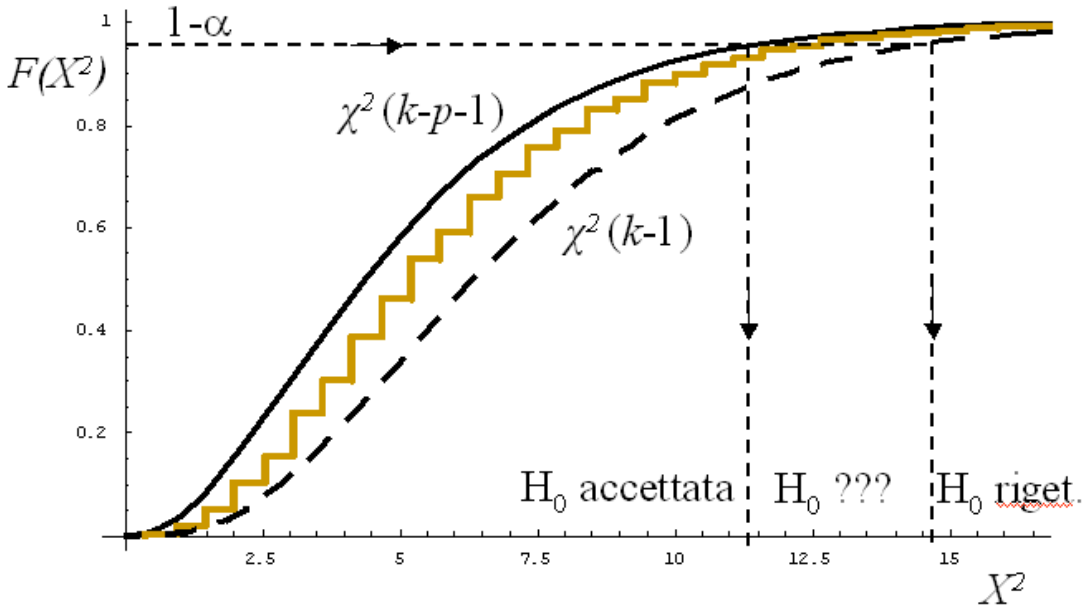

Figura 1: Rappresentazione grafica dei limiti di accettazione del test di Pearson

Tab. 1.2. Funzione di probabilità della distribuzione χ^2: valori della variabile χ^2 in funzione dei valori della probabilità di non superamento P e del numero di gradi di libertà f. (La virgola decimale è sostituita dal punto.)

f \ P	.995	.99	.975	.95	.90	.75	.50	.25	.10	.05	.025	.01	.005
1	7.88	6.63	5.02	3.84	2.71	1.32	.455	.102	.0158	.0039	.0010	.0002	.0000
2	10.6	9.21	7.38	5.99	4.61	2.77	1.39	.575	.211	.103	.0506	.0201	.0100
3	12.8	11.3	9.35	7.81	6.25	4.11	2.37	1.21	.584	.352	.216	.115	.072
4	14.9	13.3	11.1	9.49	7.78	5.39	3.36	1.92	1.06	.711	.484	.297	.207
5	16.7	15.1	12.8	11.1	9.24	6.63	4.35	2.67	1.61	1.15	.831	.554	.412
6	18.5	16.8	14.4	12.6	10.6	7.84	5.35	3.45	2.20	1.64	1.24	.872	.676
7	20.3	18.5	16.0	14.1	12.0	9.04	6.35	4.25	2.83	2.17	1.69	1.24	.989
8	22.0	20.1	17.5	15.5	13.4	10.2	7.34	5.07	3.49	2.73	2.18	1.65	1.34
9	23.6	21.7	19.0	16.9	14.7	11.4	8.34	5.90	4.17	3.33	2.70	2.09	1.73
10	25.2	23.2	20.5	18.3	16.0	12.5	9.34	6.74	4.87	3.94	3.25	2.56	2.16
11	26.8	24.7	21.9	19.7	17.3	13.7	10.3	7.58	5.58	4.57	3.82	3.05	2.60
12	28.3	26.2	23.3	21.0	18.5	14.8	11.3	8.44	6.30	5.23	4.40	3.57	3.07
13	29.8	27.7	24.7	22.4	19.8	16.0	12.3	9.30	7.04	5.89	5.01	4.11	3.57
14	31.3	29.1	26.1	23.7	21.1	17.1	13.3	10.2	7.79	6.57	5.63	4.66	4.07
15	32.8	30.6	27.5	25.0	22.3	18.2	14.3	11.0	8.55	7.26	6.26	5.23	4.60
16	34.3	32.0	28.8	26.3	23.5	19.4	15.3	11.9	9.31	7.96	6.91	5.81	5.14
17	35.7	33.4	30.2	27.6	24.8	20.5	16.3	12.8	10.1	8.67	7.56	6.41	5.70
18	37.2	34.8	31.5	28.9	26.0	21.6	17.3	13.7	10.9	9.39	8.23	7.01	6.26
19	38.6	36.2	32.9	30.1	27.2	22.7	18.3	14.6	11.7	10.1	8.91	7.63	6.84
20	40.0	37.6	34.2	31.4	28.4	23.8	19.3	15.5	12.4	10.9	9.59	8.26	7.43
21	41.4	38.9	35.5	32.7	29.6	24.9	20.3	16.3	13.2	11.6	10.3	8.90	8.03
22	42.8	40.3	36.8	33.9	30.8	26.0	21.3	17.2	14.0	12.3	11.0	9.54	8.64
23	44.2	41.6	38.1	35.2	32.0	27.1	22.3	18.1	14.8	13.1	11.7	10.2	9.26
24	45.6	43.0	39.4	36.4	33.2	28.2	23.3	19.0	15.7	13.8	12.4	10.9	9.89

Test basati sui probability plot

Un'ulteriore categoria di test di adattamento si basa sulla valutazione dell'allineamento dei punti in carta probabilistica. Il test utilizza il coefficiente di correlazione r tra la serie ordinata delle osservazioni x_i ed i corrispondenti quantili ipotetici w_i, definiti come

$$w_i = F_X^{-1}(\Phi_i)$$

dove p_i è la plotting position dell'i-esimo valore nella serie ordinata. Il coefficiente di correlazione è definito come

$$r = \frac{\sum_{i=1}^{n}\left(x_{(i)} - \bar{x}\right)\left(w_i - \bar{w}\right)}{\left[\sum_{i=1}^{n}\left(x_{(i)} - \bar{x}\right)^2\left(w_i - \bar{w}\right)^2\right]^{0.5}}$$

dove $\bar{x}$ e $\bar{w}$ sono i valori medi delle osservazioni e dei quantili. Più elevato è il coefficiente di correlazione, tanto migliore è l'allineamento dei punti in carta probabilistica. Il test è quindi basato sul confronto di r con opportuni valori tabellati in funzione di α e di n per la normale e la EV1.

Tabella 2

TABLE 18.3.3 Lower Critical Values of the Probability Plot Correlation Test Statistic for the Normal Distribution Using $p_i = (i - \frac{3}{8})/(n + \frac{1}{4})$

	Significance level		
n	0.10	0.05	0.01
10	0.9347	0.9180	0.8804
15	0.9506	0.9383	0.9110
20	0.9600	0.9503	0.9290
30	0.9707	0.9639	0.9490
40	0.9767	0.9715	0.9597
50	0.9807	0.9764	0.9664
60	0.9835	0.9799	0.9710
75	0.9865	0.9835	0.9757
100	0.9893	0.9870	0.9812
300	0.99602	0.99525	0.99354
1000	0.99854	0.99824	0.99755

Source: Refs. 101, 152, 153. Used with permission.

TABLE 18.3.4 Lower Critical Values of the Probability Plot Correlation Test Statistic for the Gumbel and Two-Parameter Weibull Distributions Using $p_i = (i - 0.44)/(n + 0.12)$

	Significance level		
n	0.10	0.05	0.01
10	0.9260	0.9084	0.8630
20	0.9517	0.9390	0.9060
30	0.9622	0.9526	0.9191
40	0.9689	0.9594	0.9286
50	0.9729	0.9646	0.9389
60	0.9760	0.9685	0.9467
70	0.9787	0.9720	0.9506
80	0.9804	0.9747	0.9525
100	0.9831	0.9779	0.9596
300	0.9925	0.9902	0.9819
1000	0.99708	0.99622	0.99334

Source: Refs. 152, 153. See also Table 18.3.2.

Se r è inferiore al valore limite indicato, l'allineamento è peggiore di quello che ci si aspetterebbe e l'ipotesi H_0 viene rigettata con significatività α. Si noti ancora che le tabelle sono costruite per essere utilizzate con le seguenti plotting positions:

$$\Phi_i = \frac{i - 3/8}{n + 1/4}$$ distribuzione Normale

$$\Phi_i = \frac{i - 0.44}{n + 0.12}$$ distribuzione di Gumbel

I quantili vanno quindi calcolati come

$$w_i = \hat{\theta}_1 + \hat{\theta}_2 \cdot NORMINV\left(\frac{i - 3/8}{n + 1/4}\right)$$ distribuzione Normale

$$w_i = \hat{\theta}_1 - \hat{\theta}_2 \ln\left(-\ln\left(\frac{i - 0.44}{n + 0.12}\right)\right)$$ distribuzione di Gumbel

Anche in questo caso, possono essere costruiti opportuni test per la log-normale ed EV2 previa log-trasformazione dei dati. Esistono invece sostanziali ostacoli per l'estensione del metodo alle distribuzioni con tre parametri.

I test basati sulla funzione di frequenza cumulata

Un'altra categoria di test di adattamento è basata sul confronto tra la distribuzione di probabilità corrispondente all'ipotesi H_0 e la funzione di frequenza cumulata, che è una funzione a gradini definita come

$$F_n(x_i) = 0, \qquad x < x_{(1)}$$

$$F_n(x_i) = \frac{i}{n}, \qquad x_{(i)} \le x < x_{(i+1)}$$

$$F_n(x_i) = 1, \qquad x_{(n)} \le x$$

dove $x_{(i)}$ indica l'i-esima order statistic, ossia l'i-esimo elemento della serie campionaria ordinata in ordine crescente. Questa categoria di test di adattamento è basata sulla valutazione dello scostamento tra la distribuzione ipotetica, $F_X(x)$ e la funzione di frequenza cumulata $\Phi(x) = F_n(x)$ (vedere ad esempio la Figura).

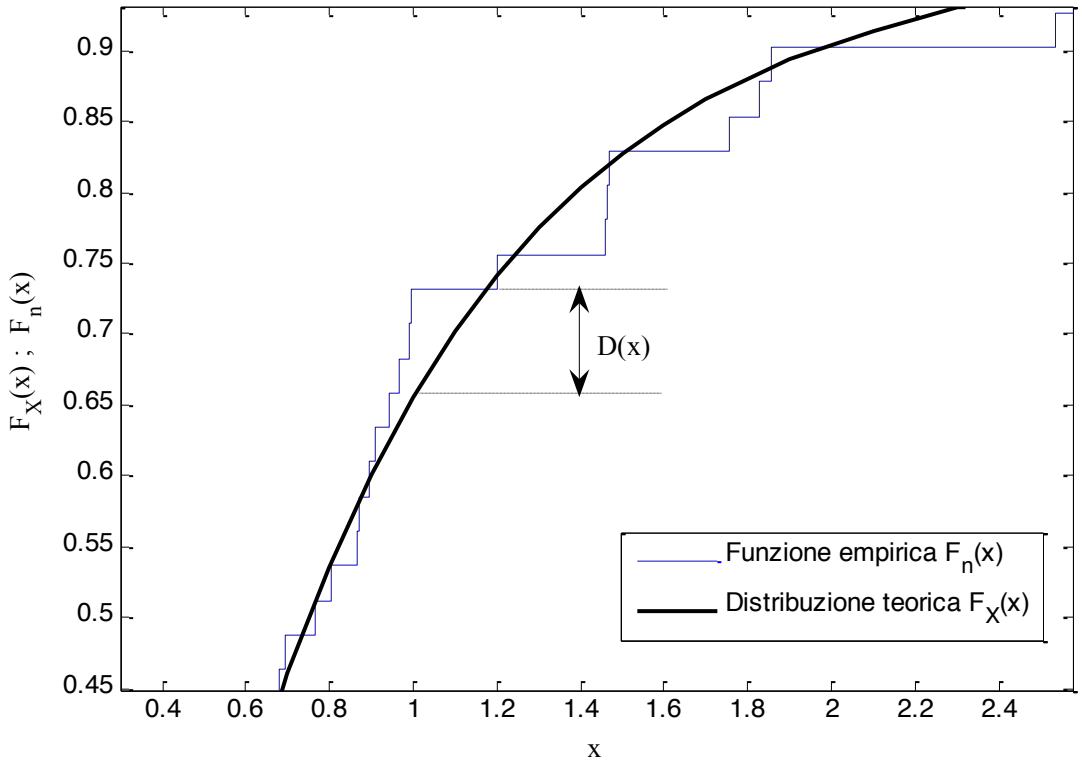

Figura 2: Scostamento tra la distribuzione di frequenza empirica cumulata $\Phi(x_i) = F_n(x)$ e la distribuzione teorica ipotizzata.

Come misura dello scostamento si può utilizzare la distanza massima tra le due funzioni in valore assoluto, $D_n = \max\left|F_X(x) - F_n(x)\right|$, nel qual caso si ha il test di Smirnov-Kolmogorov. Quando l'ipotesi H_0 è vera, la distribuzione di D_n tende asintoticamente ad una forma nota, il che consente

di determinare la regione di accettazione del test utilizzando opportune tabelle in funzione del livello di significatività del test. Per esempio, per α = 0.05 ed n> 50 si ha che la regione di accettazione coincide con l'insieme dei valori D_n per cui $D_n \leq \frac{1.3581}{\sqrt{n}}$. Occorre però prestare molta attenzione all'applicazione del test di Smirnov-Kolmogorov, dal momento che i limiti di accettazione visti sopra valgono solo nel caso in cui i parametri di $F_X(x)$ non siano stimati utilizzando le osservazioni. Nel caso opposto, che è il più comune in ambito idrologico, si ha che la stima dei parametri tende ad "avvicinare" la funzione di frequenza cumulata alla distribuzione ipotetica . A titolo di esempio, si riporta nella Figura il caso di una distribuzione empirica dei massimi annui dei colmi di piena sui quali è stata adattata una distribuzione a 2 parametri (in questo caso una Gumbel) ed una a 3 parametri (in questo caso una GEV, che corrisponde ad una generalizzazione della Gumbel). Si noti come la distribuzione GEV si adatti certamente meglio ai dati, al prezzo però di un ulteriore parametro da stimare.

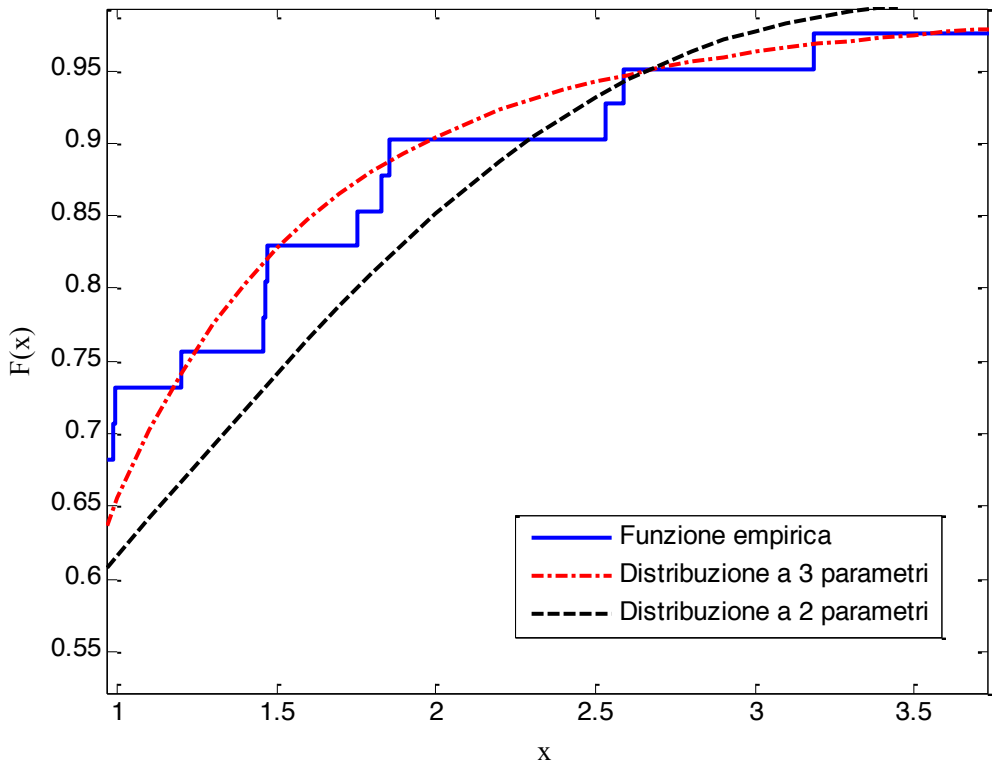

Figura 3: Esempio di adattamento alla curva empirica di una distribuzione a 2 parametri e di una a 3 parametri, con parametri stimati dal campione.

Incrementando il numero di parametri della distribuzione stimati direttamente dal campione si ottengono quindi dei valori di D_n molto inferiori rispetto a prima e non si possono utilizzare le tabelle ottenute nel caso non parametrico per effettuare il test: il valore limite $\frac{1.3581}{\sqrt{n}}$, che nel caso

non parametrico corrisponde ad $\alpha = 0.05$, nel caso in cui si stimano i parametri viene invece a corrispondere ad un livello di significatività intorno a 0.005. Questo significa che l'ipotesi H_0 verrebbe quasi sempre accettata, rendendo sostanzialmente inutile il test. Esistono tabelle che riportano i limiti di accettazione corretti da utilizzare nel caso in cui si stimino i parametri dalle osservazioni: il problema è che però i limiti di accettazione dipendono in questo caso anche dalla forma della distribuzione ipotetica, e questo comporta la necessità di avere una diversa tabella per ogni distribuzione che si vuole sottoporre a test. In particolare esistono tabelle per la distribuzione normale e la distribuzione di Gumbel (che consentono di sottoporre a test anche la log-normale e la EV2), mentre per la Gamma a tre parametri e la GEV non esistono tabelle altrettanto semplici.

Tabella 3

TABLE 4.7 Modifications and Percentage Points for a Test for Normality with μ and σ^2 Unknown (Section 4.8.1, Case 3)

Statistic	Modified statistic	Significance level α .50	.25	.15	.10	.05	.025	.01	.005
					Upper tail				
D	$D(\sqrt{n} - 0.01 + 0.85/\sqrt{n})$		-	0.775	0.819	0.895	0.995	1.035	
V	$V(\sqrt{n} + 0.05 + 0.82/\sqrt{n})$		-	1.320	1.386	1.489	1.585	1.693	
W^2	$W^2(1.0 + 0.5/n)$	.051	.074	.091	.104	.126	.148	.179	.201
U^2	$U^2(1.0 + 0.5/n)$	.048	.070	.085	.096	.117	.136	.164	.183
A^2	$A^2(1.0 + 0.75/n + 2.25/n^2)$	.341	.470	.561	.631	.752	.873	1.035	1.159
					Lower tail				
W^2	$W^2(1.0 + 0.5/n)$	.051	.036	.029	.026	.022	.019	.017	
U^2	$U^2(1.0 + 0.5/n)$	.048	.033	.027	.025	.021	.018	.016	
A^2	$A^2(1.0 + 0.75/n + 2.25/n^2)$	.341	.249	.226	.188	.160	.139	.119	

Adapted, with additions, from Table 54 of Pearson and Hartley (1972) and from Stephens (1974b), with permission of the Biometrika Trustees and of the American Statistical Association.

TABLE 4.17 Modifications and Upper Tail Percentage Points for Statistics W^2, U^2, and A^2 for the Extreme-Value or Weibull Distributions (Sections 4.10, 4.11)

Statistic	Modification	Significance level α .25	.10	.05	.025	.01
W^2						
Case 1	$W^2(1 + 0.16/n)$	.116	.175	.222	.271	.338
Case 2	None	.186	.320	.431	.547	.705
Case 3	$W^2(1 + 0.2/\sqrt{n})$	.073	.102	.124	.146	.175
U^2						
Case 1	$U^2(1 + 0.16/n)$	.090	.129	.159	.189	.230
Case 2	$U^2(1 + 0.15/\sqrt{n})$	.086	.123	.152	.181	.220
Case 3	$U^2(1 + 0.2/\sqrt{n})$	.070	.097	.117	.138	.165
A^2						
Case 1	$A^2(1 + 0.3/n)$	.736	1.062	1.321	1.591	1.959
Case 2	None	1.060	1.725	2.277	2.854	3.640
Case 3	$A^2(1 + 0.2/\sqrt{n})$	.474	.637	.757	.877	1.038

Taken from Stephens (1977), with permission of the Biometrika Trustees.

Tabella 4

TABLE 4.18 Upper Tail Percentage Points for Statistics $\sqrt{n}D^+$, $\sqrt{n}D^-$, $\sqrt{n}D$, and $\sqrt{n}V$, for Tests for the Extreme-Value or Weibull Distributions (Sections 4.10, 4.11)

		Significance level α			
Statistic	n	.10	.05	.025	.01
$\sqrt{n}D^+$	10	.872	.969	1.061	1.152
Case 1	20	.878	.979	1.068	1.176
	50	.882	.987	1.070	1.193
	∞	.886	.996	1.094	1.211
$\sqrt{n}D^-$	10	.773	.883	.967	1.103
Case 1	20	.810	.921	1.013	1.142
	50	.840	.950	1.031	1.171
	∞	.886	.996	1.094	1.211
$\sqrt{n}D$	10	.934	1.026	1.113	1.206
Case 1	20	.954	1.049	1.134	1.239
	50	.970	1.067	1.148	1.263
	∞	.995	1.094	1.184	1.298
$\sqrt{n}V$	10	1.43	1.55	1.65	1.77
Case 1	20	1.46	1.58	1.69	1.81
	50	1.48	1.59	1.72	1.84
	∞	1.53	1.65	1.77	1.91
$\sqrt{n}D^+$	10	.99	1.14	1.27	1.42
Case 2	20	1.00	1.15	1.28	1.43
	50	1.01	1.17	1.29	1.44
	∞	1.02	1.17	1.30	1.46
$\sqrt{n}D^-$	10	1.01	1.16	1.28	1.41
Case 2	20	1.01	1.15	1.28	1.43
	50	1.00	1.14	1.29	1.45
	∞	1.02	1.17	1.30	1.46

Una seconda categoria di statistiche, sempre appartenenti a questa famiglia, si basa su una misura dello scostamento medio quadratico tra la funzione di frequenza cumulata e la distribuzione ipotetica (Test di Cramer – Von Mises),

$$Q^2 = n\int_{allx}\left(\Phi(x) - F_X(x)\right)^2 f_X(x)dx$$

Se invece della media semplice si utilizza una funzione di peso atta ad attribuire maggiore importanza agli scostamenti sulle code delle due distribuzioni, si ottiene l'importante test statistic di **Anderson – Darling**:

$$A^2 = n\int_{allx}\frac{\left(\Phi(x) - F_X(x)\right)^2}{F_X(x)\left(1 - F_X(x)\right)} f_X(x)dx$$

La statistica di Anderson-Darling viene calcolata utilizzando la seguente espressione, che consente di evitare di dover calcolare l'integrale visto sopra ogni volta che si vuole fare il test:

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}\left\{(2i-1)\ln\left(F_X(x_{(i)})\right) + (2n+1-2i)\ln\left(1 - F_X(x_{(i)})\right)\right\}.$$

Anche per il test di Anderson esistono gli stessi problemi riscontrati per il test di Kolmogorov, ossia si deve fare riferimento ad una diversa tabella per ogni distribuzione considerata quando i parametri sono stimati in base alle osservazioni. Nelle applicazioni pratiche conviene trasformare la variabile A^2 tramite le relazioni

$$\omega = 0.0403 + 0.116\left(\frac{A^2 - \xi_p}{\beta_p}\right)^{\frac{\eta_p}{0.851}} \qquad \text{se} \qquad 1.2\xi_p \le A^2$$

$$\omega = \left[0.0403 + 0.116\left(\frac{0.2\xi_p}{\beta_p}\right)^{\frac{\eta_p}{0.851}}\right]\frac{A^2 - 0.2\xi_p}{\xi_p} \qquad \text{se} \qquad 1.2\xi_p > A^2$$

dove ξ_p, β_p, ed η_p sono coefficienti, diversi per ogni distribuzione, tabulati in Laio (2004) e riportati più sotto (Tabella 5). Nella tabella dove $\theta_3 = k$, ovvero è il parametro di forma della distribuzione.

Tabella 5

Distribution[b]	ξ_p	β_p	η_p
EV1 and EV2	0.169	0.229	1.141
NORM and LN	0.167	0.229	1.147
GEV[c]	$0.147\,(1 + 0.13\,\hat{\theta}_3 + 0.21\,\hat{\theta}_3^2 + 0.09\,\hat{\theta}_3^3)$	$0.189\,(1 + 0.20\,\hat{\theta}_3 + 0.37\,\hat{\theta}_3^2 + 0.17\,\hat{\theta}_3^3)$	$1.186\,(1 - 0.04\,\hat{\theta}_3 - 0.04\,\hat{\theta}_3^2 - 0.01\,\hat{\theta}_3^3)$
GAM and LP3[d]	$0.145\,(1 + 0.17\,\hat{\theta}_3^{-1} + 0.33\,\hat{\theta}_3^{-2})$	$0.186\,(1 + 0.34\,\hat{\theta}_3^{-1} + 0.30\,\hat{\theta}_3^{-2})$	$1.194\,(1 - 0.04\,\hat{\theta}_3^{-1} - 0.12\,\hat{\theta}_3^{-2})$

[a]Here $\hat{\theta}_3$ is an asymptotic efficient estimator (usually maximum likelihood) of the shape parameter of the distribution.
[b]For tests of the EV2, LN, and LP3 distributions the data must be preliminarily log transformed.
[c]For the GEV distribution, if $\hat{\theta}_3 > 0.5$, $\hat{\theta}_3 = 0.5$ must be set in the regressions.
[d]For the GAM and LP3 distributions, if $\hat{\theta}_3 < 2$, $\hat{\theta}_3 = 2$ must be set in the regressions.

Una volta effettuata la trasformazione di cui sopra, l'ipotesi può essere accettata se ω è minore di 0.347 (con $\alpha = 0.10$), 0.461 (con $\alpha = 0.05$) e 0.743 (con $\alpha = 0.01$)

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:

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

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:

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

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(=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.

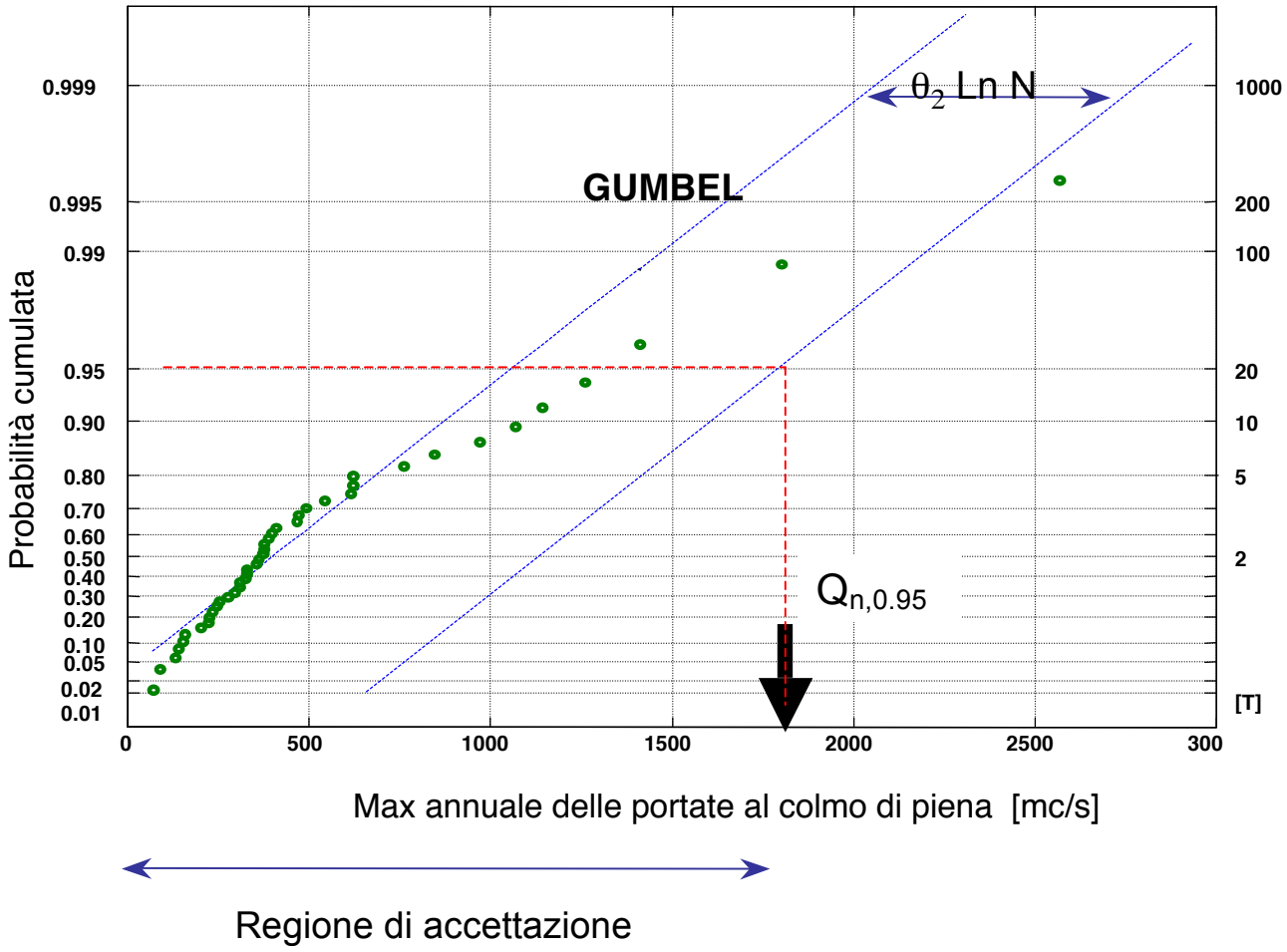

Figura 4: limiti di accettazione dell'ipotesi nel test del massimo valore

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:

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

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: $T(X_{N+1}) = \frac{1}{1 - F_X^*(X_{N+1})}$

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.

Considerazioni finali

Il problema della verifica e scelta del modello probabilistico può essere affrontato in maniera appropriata per le distribuzioni a due parametri, per le quali si hanno 4 test di adattamento affidabili: il test di Kolmogorov-Smirnov, quello di Anderson-Darling, quello basato sull'L-skewness e quello basato sull'allineamento dei punti in carta probabilistica. Per le distribuzioni a tre parametri non esistono invece test altrettanto facilmente utilizzabili. Una possibile strategia ai fini della scelta della distribuzione potrebbe pertanto essere la seguente: sottoporre ai 4 test suddetti le distribuzioni normale, lognormale, EV1 ed EV2: se una distribuzione passa tutti i test, adottarla come distribuzione di probabilità nel seguito dell'analisi; se nessuna distribuzione passa i quattro test, scegliere una distribuzione a tre parametri, eventualmente basando la scelta sul diagramma degli L-momenti. Se più di una distribuzione a due parametri passa i test, scegliere tra esse in base all'adattamento visuale su carta probabilistica, oppure portare avanti le successive elaborazioni con più di una distribuzione.

Riferimenti Bibliografici

D'Agostino R.B. e M.A. Stephens (eds) *Goodness of fit techniques,* Dekker, New York, 1986.

Laio F., Cramer-von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters. *Water Resources Research*, 40, W09308, doi:10.1029/2004WR003204, 2004