Lo studio statistico delle piene

Caratteristiche diverse:

a seconda del problema (solo portata al colmo, oppure intera onda di piena)
e a seconda dei dati disponibili.

Classificazione:

analisi basate su soli dati di portata e
analisi basate sulla trasformazione afflussi-deflussi;
analisi locali (*at-site*) e analisi regionali.

A volte Q non disponibili: necessità di trasformare $q(T)$ in $Q(T)$.

Le analisi statistiche locali

Uso per massimi annuali di portata al colmo o media giornaliera.
Stima di $Q(T)$ con i dati di una sola stazione.
Piccola differenza dovuta all'uso dei soli massimi.
Uso di distribuzioni a due parametri:

Gumbel,
lognormale a due parametri,
Gamma a due parametri.

Necessità di usare distribuzioni a due parametri per la scarsità di dati.
Condizione di separazione (non spiegata dalle distribuzioni a due parametri, se si ammette l'uniformità del coefficiente di asimmetria in un'intera regione).
Necessità di usare leggi a più di due parametri per eliminare la condizione di separazione.
Piccole differenze tra leggi a due o più param. per T=100 anni
Scelta della distribuzione:

test di Pearson,
carte probabilistiche (figure Po).

Uso per portate al colmo o medie giornaliere.

Relazione tra portate massime giornaliere e al colmo

Descrizione del problema.
Obiettivo: trasformare $q(T)$ in $Q(T)$.
Considerazione del rapporto $R=Q(T)/q(T)$ oppure del rapporto $R=Q/q$ tra i valori osservati nello stesso anno.

<u>$R=Q(T)/q(T)$</u>:
R come funzione dei parametri del bacino e del tempo di ritorno.

Semplificazione del problema: dipendenza dai soli parametri del bacino o dal solo tempo di ritorno.

Esempi di formule che forniscono $R=Q(T)/q(T)$ in funzione dei parametri del bacino:

Tonini (bacini italiani),
Cotecchia (bacini italiani),
Tonini *et al.* (bacini delle Dolomiti),
Canuti e Moisello (bacini liguri e toscani)

Esempi di formule che forniscono $R=Q(T)/q(T)$ in funzione del tempo di ritorno:

Lazzari (bacini della Sardegna).
Versace e Principato (bacini della Calabria): $Q(T)/q(T)$ non dipende dall'area del bacino, è distribuito come una variabile casuale e la dipendenza da T è debole

$\underline{R=Q/q}$:
R come variabile casuale; esempio:

Canuti e Moisello (bacini italiani).

Distribuzione del massimo in N anni

Confusione sul significato di $Q(T)$.
Opportunità di usare la distribuzione del massimo in N anni.
Relazione tra distribuzione del massimo annuale e distribuzione del massimo in N anni.
Espressione asintotica della probabilità di non superamento in N anni della portata con tempo di ritorno di N anni.

Analisi statistica delle piene

Risultati richiesti
- solo portata al colmo
- intero idrogramma di piena

Osservazioni utilizzate
- solo osservazioni di portata
- osservazioni di portata e di pioggia

Uso di osservazioni di portata e di pioggia
- uso generale e indiretto della trasformazione afflussi-deflussi
- uso esplicito della trasformazione afflussi-deflussi

Serie di osservazioni considerate
- analisi locali (*at-site*)
- analisi regionali

Osservazioni di portata utilizzate
- portate al colmo
- portate medie giornaliere

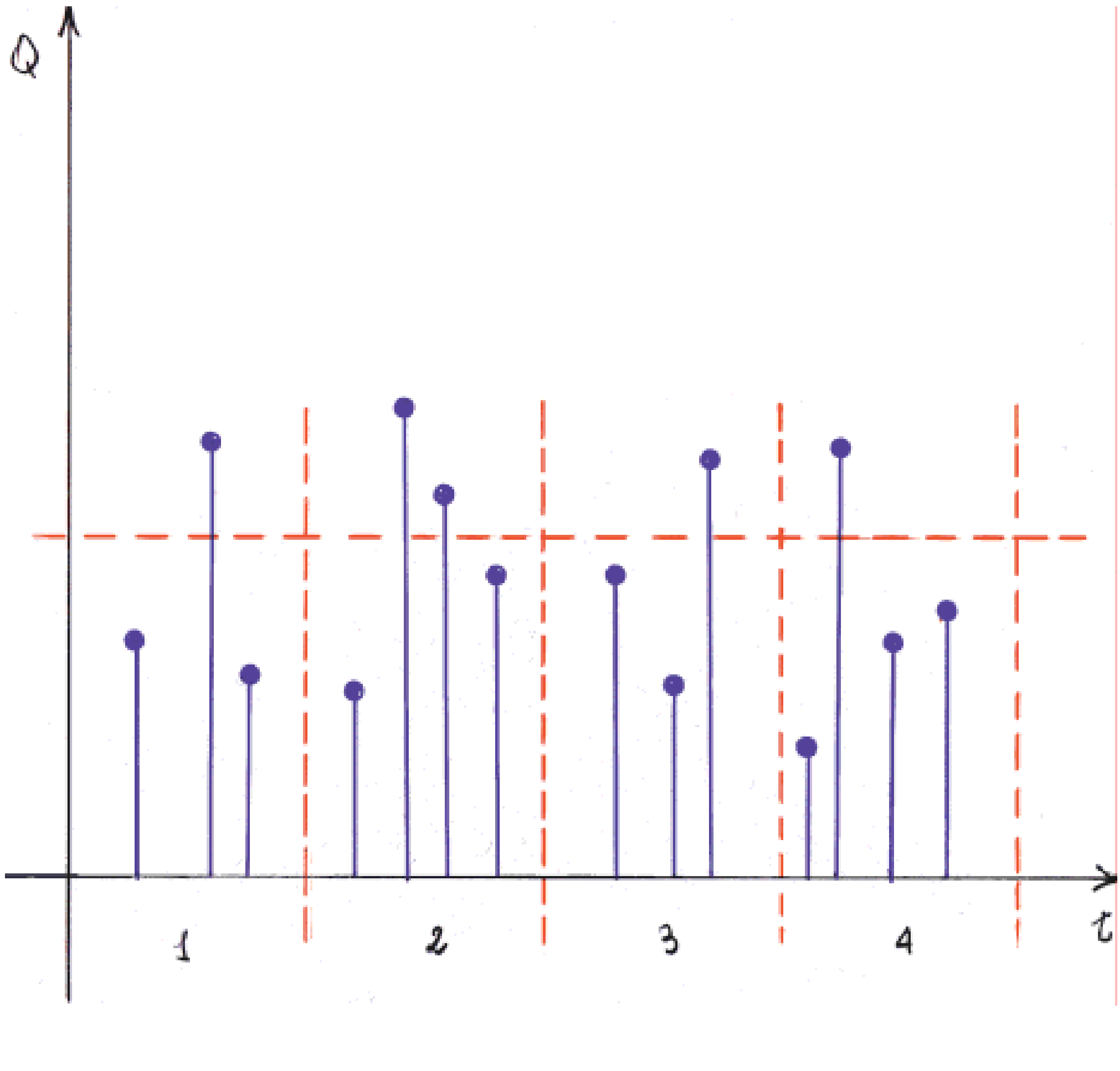
Tempo di ritorno

Due definizioni equivalenti

1) Il tempo di ritorno T di un valore x_a è la lunghezza dell'intervallo di tempo che mediamente intercorre tra due superamenti successivi del valore x_a.

2) Il tempo di ritorno T di un valore x_a è il periodo di tempo nel quale il valore x_a è superato mediamente una sola volta.

Il valore x_a con tempo di ritorno T <u>non è</u> il massimo valore della variabile x che mediamente ci si attende in un periodo di tempo di lunghezza T.


Q
1
2
3
4
t

Metodi di analisi statistica utilizzati

Analisi dei massimi annuali

Analisi delle eccedenze

Distribuzioni di probabilità per l'analisi dei massimi annuali

a due parametri
- di Gumbel
- lognormale a due parametri
- Gamma a due parametri
- . . .

a più di due parametri
- GEV (3 parametri)
- TCEV (4 parametri)
- Wakeby (5 parametri)
- . . .

Metodi di stima dei parametri
- dei momenti (MOM)
- della massima verosimiglianza (ML)
- dei momenti pesati in probabilità (PWM)

Relazione tra numero delle osservazioni e numero dei parametri della distribuzione.

Condizione di separazione:
la varianza del coefficiente di asimmetria (*skewness*)

$$\gamma = \frac{\mu_3(Q)}{\sigma^3(Q)}$$

stimata per campioni di dimensione N provenienti da una certa regione idrologica è maggiore di quella teorica corrispondente a distribuzioni a due parametri, nell'ipotesi che il coefficiente di asimmetria sia lo stesso in tutta la regione e nell'ipotesi di stazionarietà delle distribuzioni nel tempo.

La condizione di separazione è legata alla presenza di un numero di valori eccezionali (*outlier*) superiore a quello atteso.

Il problema della condizione di separazione si può risolvere adoperando certe distribuzioni a più di due parametri.

Esempio

Serie dei massimi annuali di portata al colmo del Po a Pontelagoscuro (periodo 1918-70)

53 osservazioni in tutto

Distribuzioni considerate:

- legge di Gumbel
- legge lognormale a due parametri
- legge Gamma a due parametri

Stima dei parametri: metodo dei momenti

Valori osservati della media e dello scarto quadratico medio:

$$m(Q) = 5407{,}2 \quad \mathrm{m^3\,s^{-1}}$$

$$s(Q) = 1872{,}5 \quad \mathrm{m^3\,s^{-1}}$$

Legge di Gumbel

$$P(Q) = \exp\{-\exp[-\alpha(Q - u)]\}$$

$$\alpha = \backslash\mathrm{S}\backslash\mathrm{DO3}(\backslash\mathrm{F}(1{,}283,\sigma(Q)))$$

$$u = \mu(Q) - 0{,}450\,\sigma(Q)$$

$$\alpha = \frac{1{,}283}{\sigma(Q)} = \frac{1{,}283}{1872{,}5} = 0{,}0006852 \quad \mathrm{s\,m^{-3}}$$

$$u = \mu(Q) - 0{,}450\,\sigma(Q)$$

$$= 5407{,}2 - 0{,}450 \times 1872{,}5$$

$$= 4565 \quad \mathrm{m^3\,s^{-1}}$$

Legge lognormale

$$z = a \ln Q + b$$

$$a = \frac{1}{\sqrt{\ln\left[1 + \frac{\sigma^2(Q)}{\mu^2(Q)}\right]}}$$

$$b = \frac{1}{2a} - a \ln \mu(Q)$$

$$a = \frac{1}{\sqrt{\ln\left[1 + \frac{1872{,}5^2}{5407{,}2^2}\right]}} = 2{,}971$$

$$b = \frac{1}{2 \times 2{,}971} - 2{,}971 \times \ln 5407{,}2$$

$$= -25{,}37$$

Legge Gamma a due parametri

$$y = \alpha Q$$

$$P(Q) = \frac{\Gamma_i(y;\gamma)}{\Gamma(\gamma)}$$

$\Gamma_i(y;\gamma)$ funzione Gamma incompleta

$\Gamma(\gamma)$ funzione Gamma completa

$$\alpha = \frac{\mu(Q)}{\sigma^2(Q)}$$

$$\gamma = \frac{\mu^2(Q)}{\sigma^2(Q)}$$

$$\alpha = \frac{5407{,}2}{1872{,}5^2} = 0{,}001542 \quad s\,m^{-3}$$

$$\gamma = \frac{5407{,}2^2}{1872{,}5^2} = 8{,}339$$

Scelta della distribuzione

Test di adattamento di Pearson

- k numero degli intervalli
- s numero dei parametri della distribuzione
- p_i probabilità che un'osservazione ricada nell'i-esimo intervallo
- N_i numero delle osservazioni che ricadono nell'i-esimo intervallo
- N dimensione del campione
- X^2 criterio del test (indice di adattamento)

$$X^2 = \sum_{i=1}^{k} \frac{(N_i - Np_i)^2}{Np_i}$$

L'adattamento della distribuzione alle osservazioni è tanto migliore quanto più basso è il valore di X^2.

La funzione di probabilità della grandezza X^2 tende, al crescere di N, a una forma asintotica compresa tra quella del χ^2 con $k-1$ gradi di libertà e quella del χ^2 con $k-1-s$ gradi di libertà.

Si può assumere che la probabilità di non superamento corrispondente a un assegnato valore di X^2 sia uguale alla media dei valori forniti dalle due distribuzioni.

Nell'esecuzione del test si assume

$$k = \mathrm{int}(N/5)$$

$$p_i = 1/k \qquad \text{(equiprobabilità)}$$

La probabilità di superamento del valore di X^2 ricavato dal campione è il livello di significatività α più alto al quale l'ipotesi che il campione provenga dalla legge considerata può ancora essere accettata.

Si sceglie la legge a cui corrisponde il livello di significatività più alto.

Numero degli intervalli, probabilità, numero dei gradi di libertà:

$$k = \text{int}(N/5) = \text{int}(53/5) = 10$$

$$p_i = 1/k = 1/10 = 0,10$$

$$s = 2$$

$$k - 1 = 10 - 1 = 9$$

$$k - s - 1 = 10 - 2 - 1 = 7$$

Valori del criterio del test:

legge di Gumbel: $X^2 = 18,51$

legge lognormale $X^2 = 16,25$

legge Gamma $X^2 = 10,21$

Massimo livello di significatività α per il quale l'ipotesi risulta accettabile:

legge di Gumbel $\alpha = 1 - P(X^2) = 0,020$

legge lognormale $\alpha = 1 - P(X^2) = 0,042$

legge Gamma $\alpha = 1 - P(X^2) = 0,256$

Relazione tra i massimi q delle portate (medie) giornaliere e i massimi Q delle portate al colmo

Per passare dalla distribuzione del massimo annuale q della portata (media) giornaliera a quella del massimo annuale Q della portata al colmo si può considerare:

a) il rapporto $Q(T)/q(T)$ a parità di tempo di ritorno T;

b) la variabile casuale costituita dal rapporto Q/q tra i massimi osservati nello stesso anno.

In generale il rapporto

$$R = \frac{Q(T)}{q(T)}$$

dipende dal tempo di ritorno T e dalle caratteristiche del bacino.

Alcune semplificazioni fanno dipendere il rapporto R solo dal tempo di ritorno T, altre solo da alcune caratteristiche del bacino.

Esempi di relazioni che esprimono R in funzione del solo tempo di ritorno

Lazzari, parte occidentale della Sardegna:

$$R = 1{,}979 \times 10^{0{,}0764 z(T)}$$

Lazzari, parte orientale della Sardegna:

$$R = 2{,}023 \times 10^{0{,}0646 z(T)}$$

Versace e Principato hanno trovato per un gruppo di bacini della Calabria che la distribuzione di R è lognormale e non dipende dall'area del bacino.

Esempi di relazioni che esprimono R in funzione delle sole caratteristiche del bacino:

Tonini:

$$R = \frac{Q(T)}{q(T)} = 1 + 68A^{-0,5}$$

Cotecchia:

$R = 32A^{-0,313}$ $\quad A > 120\text{-}140\ \ \mathrm{km}^2$

$R = 16A^{-0,190}$ $\quad A < 120\text{-}140\ \ \mathrm{km}^2$

Tonini *et al.*:

$R = 2{,}39A^{-0,112}$ (Dolomiti)

Canuti e Moisello:

$R = 11{,}72A^{-0,100}z_m{}^{-0,112}$ (Liguria e Toscana)

Il rapporto

$$R = \frac{Q}{q}$$

tra i valori di Q e di q osservati nello stesso anno è considerato ancora come una variabile casuale.

Si può assumere R indipendente da q (Canuti e Moisello). Allora

$$\mu(Q) = \mu(R)\mu(q),$$

$$\sigma^2(Q) = \sigma^2(R)[\sigma^2(q) + \mu^2(q)] + \mu^2(R)\sigma^2(q).$$

Il legame tra $\mu(R)$, $\sigma^2(R)$ e le caratteristiche del bacino (area, etc.) si può ricavare da un'analisi regionale.

I parametri $\mu(Q)$, $\sigma^2(Q)$ si stimano dalle relazioni sopra riportate. Il tipo di legge probabilistica secondo cui è distribuita la variabile Q si assume in base all'esperienza.

Relazione tra $q(T)$ e $Q(T)$

Con buona approssimazione è

$$Q(T) = a\,q(T) + b$$

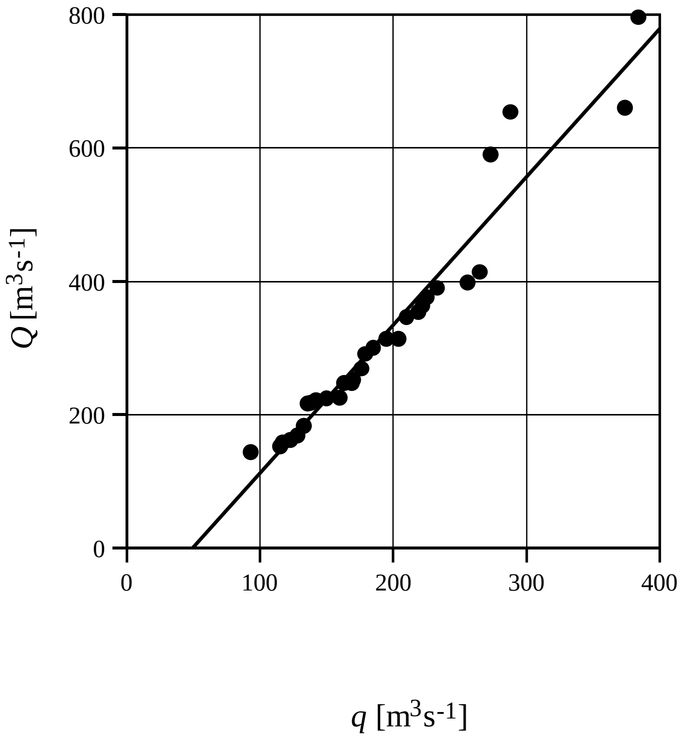
a è sempre positivo e non minore di uno

b può essere

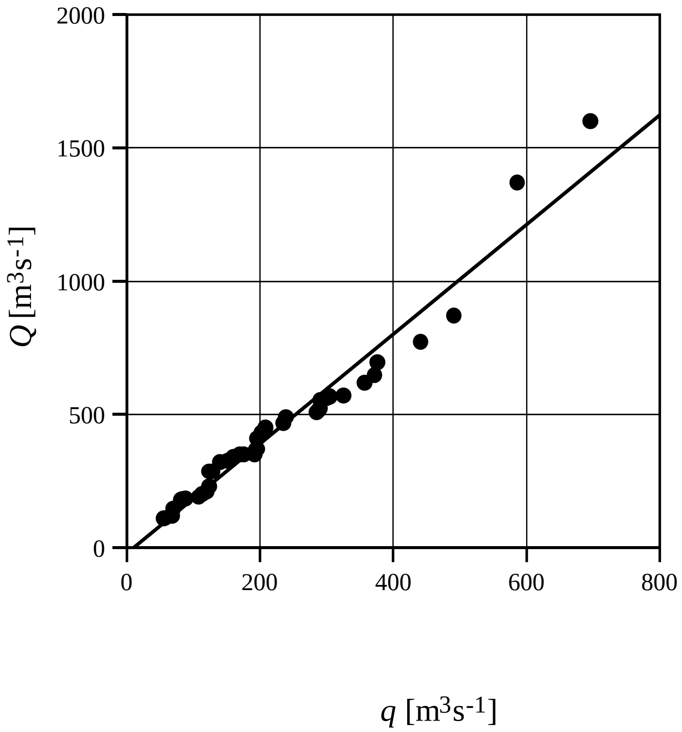
positivo: il rapporto $Q(T)/q(T)$ decresce al crescere di $q(T)$

negativo: il rapporto $Q(T)/q(T)$ cresce al crescere di $q(T)$

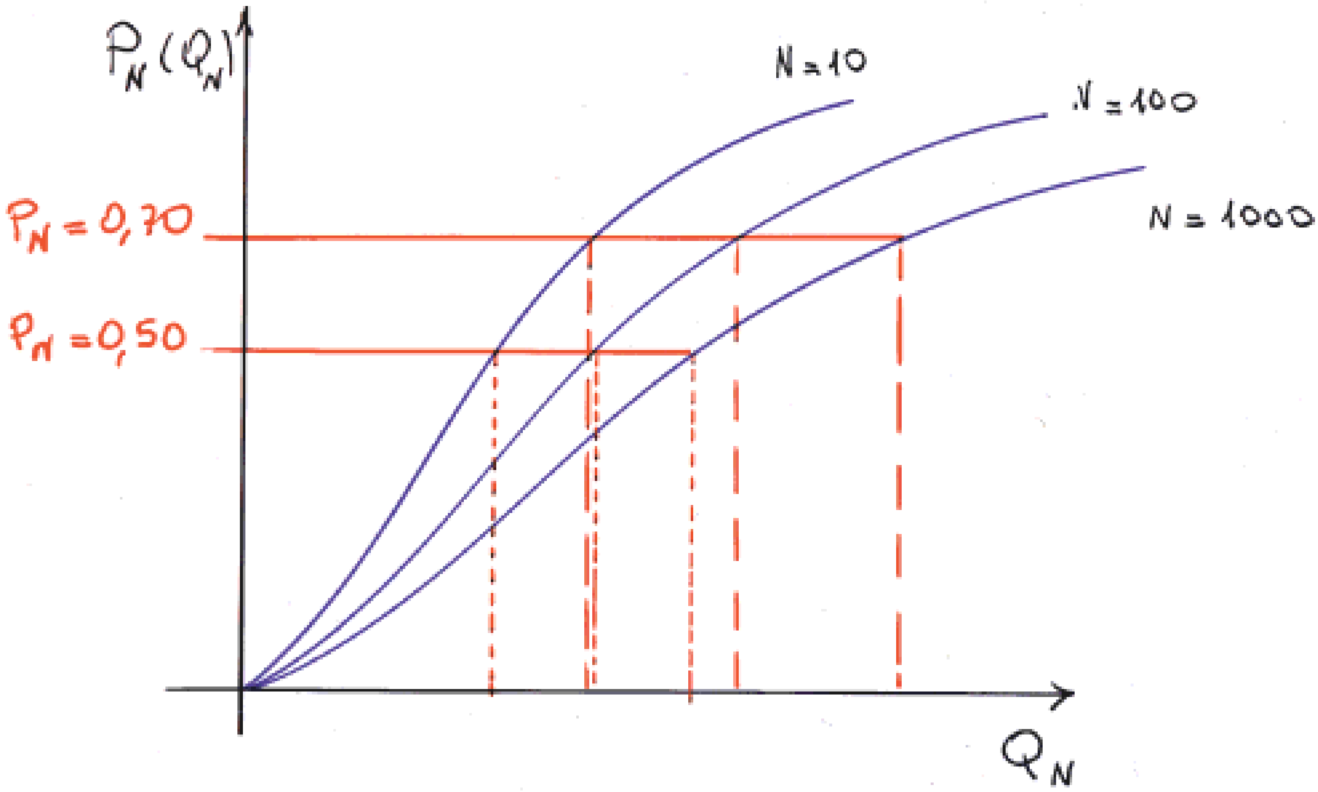
I parametri a e b si ricavano dalle osservazioni



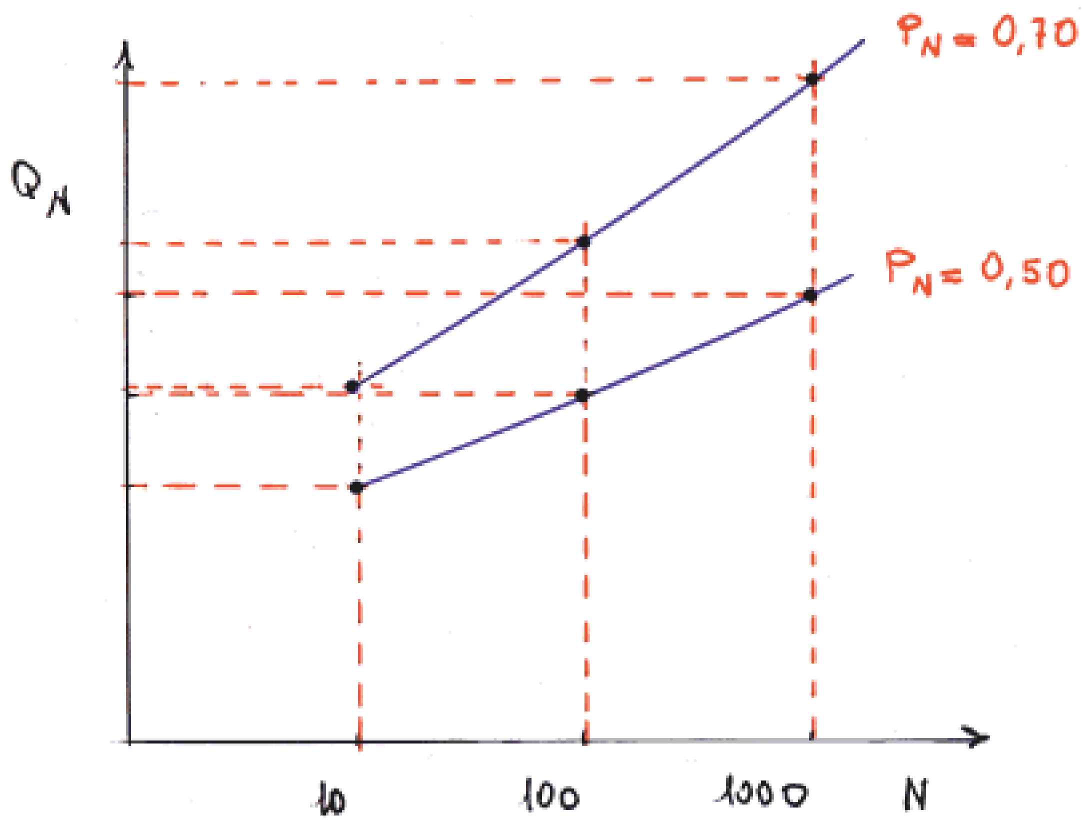
Pescara a S. Teresa: relazione tra le distribuzioni di frequenza osservate del massimo annuale q della portata media giornaliera e del massimo annuale Q della portata al colmo



Stura di Lanzo a Lanzo: relazione tra le distribuzioni di frequenza osservate del massimo annuale q della portata media giornaliera e del massimo annuale Q della portata al colmo



$P_N(Q_N)$
N = 10
N = 100
N = 1000
$P_N = 0,70$
$P_N = 0,50$
Q_N



Q_N
$P_N = 0,70$
$P_N = 0,50$
10
100
1000
N

x variabile originaria (massimo annuale)

$P(x)$ probabilità di non superamento di x in un anno

$P_N(x)$ probabilità di non superamento di x in N anni

Legame tra $P(x)$ e $P_N(x)$:

$$P_N(x) = P(x)^N$$

Consideriamo il valore $x(T)$ con tempo di ritorno di T anni:

$$P_N[x(T)] = P[x(T)]^N$$

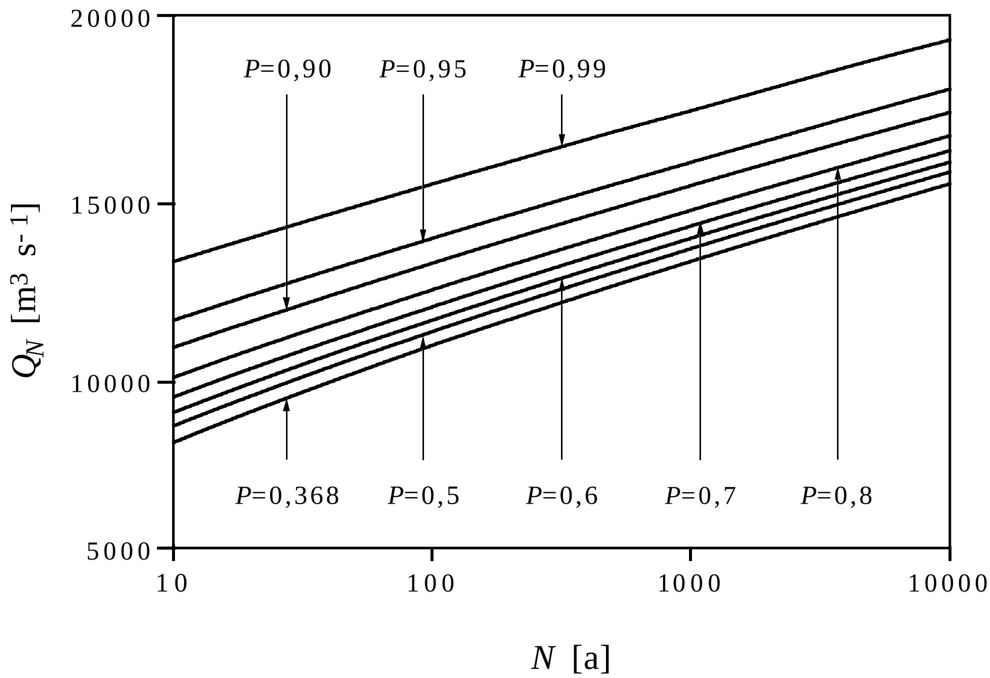
$$P[x(T)] = 1 - 1/T$$

$$P_N[x(T)] = [1 - 1/T]^N$$

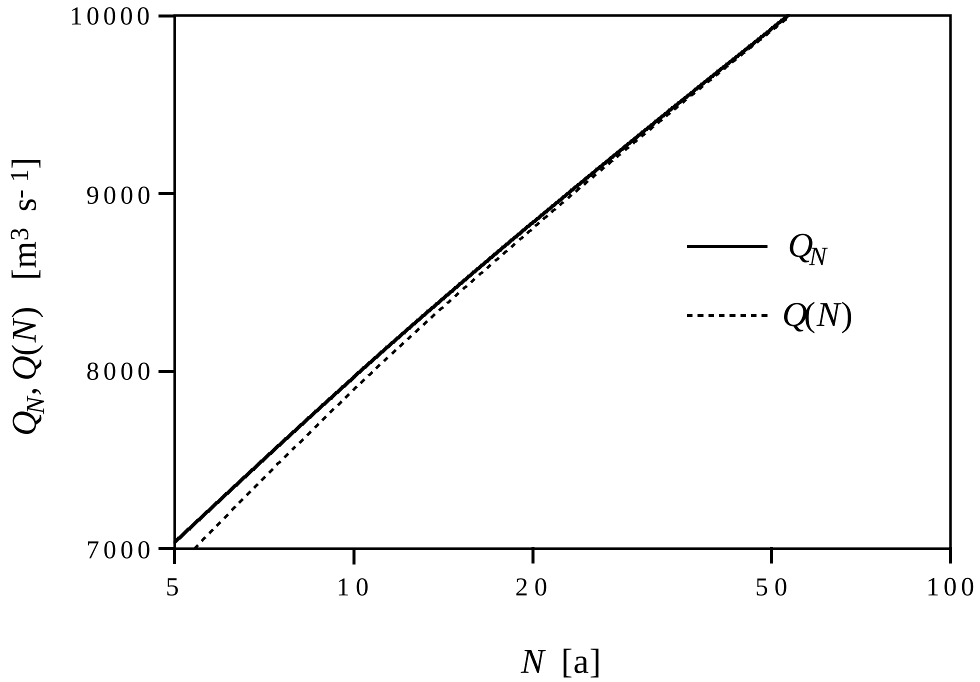
e ponendo

$$T = N$$

$$P_N[x(N)] = [1 - 1/N]^N \approx e^{-1} = 0,368\ldots$$



Po a Pontelagoscuro. Distribuzione del massimo della portata al colmo Q_N in un periodo di N anni.



Po a Pontelagoscuro. Confronto tra il massimo della portata al colmo Q_N in un periodo di N anni (con probabilità di non superamento uguale a 0,368) e il massimo della portata al colmo $Q(N)$ con tempo di ritorno di N anni.

1.

Nel lavoro di Matalas *et al*. sulla condizione di separazione (1975) si dà la dipendenza della media e dello scarto quadratico medio della stima del coefficiente di asimmetria g dalla dimensione N del campione, determinata per via numerica con il metodo Montecarlo per un certo numero di distribuzioni di probabilità.

Le distribuzioni di probabilità considerate sono sette: normale, uniforme, lognormale a tre Parametri, Gamma a tre parametri, Gumbel, Weibull e Pareto.

Per ogni distribuzione si assume che la media della variabile sia uguale a zero e che lo scarto quadratico medio sia uguale a uno. (In effetti la normalizzazione della variabile originaria lascia inalterato il valore del coefficiente di asimmetria.)

Per le distribuzioni a tre parametri (lognormale a tre Parametri, Gamma a tre parametri, Weibull e Pareto) si prendono in considerazione diversi valori del coefficiente di asimmetria.

Si effettua inoltre un'indagine sperimentale, utilizzando più di 1000 stazioni di portata degli Stati Uniti. L'indagine è condotta suddividendo le stazioni in gruppi, secondo la regione di appartenenza. Per ogni regione si calcolano la media e lo scarto quadratico medio della stima del coefficiente di asimmetria.

Per le serie storiche lo scarto quadratico medio della stima del coefficiente di asimmetria cresce con la dimensione N del campione. I risultati dell'analisi effettuata con il metodo Montecarlo mostrano lo stesso andamento per valori del coefficiente di asimmetria maggiori di due.

Il confronto tra i risultati dell'analisi effettuata con il metodo Montecarlo e quelli ricavati dalle osservazioni è fatto per mezzo di grafici. Ogni grafico corrisponde a un particolare valore della dimensione N del campione. Nel grafico sono riportate la media della stima di g in ascisse e lo scarto quadratico medio della stima di g in ordinate. I risultati dell'analisi Montecarlo sono completati tracciando delle linee, ciascuna delle quali si riferisce a una delle sette distribuzioni considerate.

I punti sperimentali, ciascuno dei quali ha come ascissa e come ordinata la media e lo scarto quadratico medio della stima di g corrispondenti a una particolare regione, si trovano sempre più in alto delle linee teoriche (*condizione di separazione*).

E` importante osservare (come fatto da Klemeš) che l'assunzione di costanza del coefficiente di asimmetria è arbitraria e non corrisponde alla realtà.

2.

Le distribuzioni a due parametri - pur largamente utilizzate - non si mostrano del tutto adatte a interpretare le serie di massimi annuali di portata. Questa limitazione è dovuta all'esistenza della cosiddetta *condizione di separazione* (Matalas *et. al.*, 1975) che cerchiamo di illustrare brevemente.

Data una qualunque distribuzione di probabilità, la stima del coefficiente di asimmetria ricavata da un campione di dimensione assegnata costituisce una variabile casuale, la cui distribuzione ha ovviamente una propria media e una propria varianza.

Matalas, Slack e Wallis hanno considerato un gran numero di serie di osservazioni di massimi annuali di portata provenienti da stazioni diverse situate in una regione idrologica omogenea, le hanno suddivise in diverse serie parziali più brevi di lunghezza assegnata N, hanno stimato per ciascuna serie parziale il valore del coefficiente di asimmetria e hanno quindi stimato la varianza delle stime. Hanno infine confrontato il valore empirico della varianza del coefficiente di asimmetria dei campioni di dimensione N con quello teorico, determinato facendo l'ipotesi che in tutta la regione la distribuzione dei massimi sia dello stesso tipo e che il coefficiente di asimmetria abbia ovunque lo stesso valore (pur essendo ovviamente diverse le medie e le varianze), e hanno ripetuto le elaborazioni per 14 regioni omogenee degli Stati Uniti.

Dall'analisi, nella quale sono stati considerati sei diversi tipi di distribuzione (a due e a tre parametri, tra le quali la distribuzione di Gumbel e quella lognormale a tre parametri), è risultato che la varianza dei valori osservati del coefficiente di asimmetria risulta sempre maggiore di quella teorica. Questa circostanza prende il nome di *condizione di separazione*.
E` stato successivamente verificato che alcune distribuzioni a più di due parametri - essenzialmente la distribuzione Wakeby e la distribuzione del valore estremo a due componenti (TCEV) - permettono invece di ottenere valori teorici della varianza del coefficiente di asimmetria che non dànno luogo alla condizione di separazione. Per utilizzare queste distribuzioni, che interpretano meglio la realtà, occorre però necessariamente fare ricorso alle analisi regionali, a cui si è fatto cenno nel paragrafo precedente.

3.
A proposito della variabilità del coefficiente di asimmetria calcolato, è importante ricordare la scoperta della così detta *condizione di separazione* (Matalas *et al*., 1975). La condizione di separazione consiste nel fatto che nessuna delle distribuzioni comunemente usate nella pratica idrologica al tempo in cui l'indagine è stata eseguita (nessuna legge a due parametri, per esempio) produce - nell'assunzione che una data regione idrologica omogenea sia caratterizzata dall'uniformità spaziale del coefficiente di asimmetria e dalla stazionarietà delle distribuzioni nel tempo - campioni che si comportano nello stesso modo delle serie di massimi annuali osservati della stessa dimensione: la variabilità delle stime del coefficiente di asimmetria ricavate dalle osservazioni risulta sempre maggiore di quella attesa. Solo la legge Wakeby, la legge dei valori estremi a due componenti (*Two Components Extreme Value*, TCEV) e la legge logistica generalizzata (*Generalized Logistic*, GLG) (oppure la sua forma logaritmica, *Log-logistic*, LLG) emergono come possibili candidate a rappresentare le serie di massimi annuali della portata al colmo (Cunnane, 1989). Si presenta però il problema della stima dei parametri, il cui numero è troppo alto per poter fare affidamento sui risultati ottenuti attraverso un'analisi locale.
In realtà, se si definisce l'omogeneità dei bacini di una stessa regione come l'esistenza di un legame di scala, la condizione di separazione si può spiegare come la conseguenza di un'ipotesi arbitraria (quella dell'uniformità spaziale del coefficiente di asimmetria) (Gupta e Dawdy, 1995). Il coefficiente di asimmetria resta infatti effettivamente costante passando da un bacino all'altro della stessa regione omogenea, quando l'omogeneità è definita come l'esistenza di un legame di scala semplice tra le portate delle diverse stazioni, cioè quando il rapporto tra le portate con uguale tempo di ritorno è una potenza con esponente costante del rapporto tra le aree dei rispettivi bacini idrografici. Ma non è più così, quando l'omogeneità è definita come l'esistenza di un legame di scala multipla, quando cioè il rapporto tra le portate con uguale tempo di ritorno è ancora una potenza del rapporto tra le aree dei rispettivi bacini, ma con esponente che a sua volta dipende dal tempo di ritorno. Ora, l'ipotesi dell'esistenza di un legame di scala semplice non appare corretta, perchè implica la costanza del coefficiente di variazione, in contrasto con l'esperienza, che ne mostra la dipendenza dall'area (per aree non piccolissime, il coefficiente di variazione decresce al crescere dell'area) (Dawdy, 1961) (Benson, 1962) (Gupta e Dawdy, 1995). Negando l'uniformità del coefficiente di asimmetria, l'ipotesi di un legame di scala multipla ne può spiegare la variabilità della stima.

4.
Circa il metodo della portata indice occorre fare un'osservazione importante. L'ipotesi che l'unico parametro che varia, passando da una stazione all'altra della stessa regione, sia la portata indice (comunemente individuata nella media) implica che tutti gli altri parametri restino costanti. Ciò è in contrasto con l'esperienza: come si è già avuto modo di rilevare, per esempio, il coefficiente di variazione varia al variare dell'area. Proprio la dipendenza del coefficiente di variazione dall'area, messa in evidenza da Dawdy (1961) e da Benson (1962), ha spinto l'USGS (United States Geological Survey) ad abbandonare sin dai primi anni '60 il metodo della portata indice per quello della regressione dei quantili (Gupta *et al*., 1994), che fa

dipendere i quantili della portata di piena considerata direttamente da un certo numero di parametri caratteristici del bacino. Il metodo della portata indice è stato recentemente criticato, sostanzialmente per le stesse ragioni, anche da autori italiani (Maione, 1997), che hanno proposto per i bacini italiani una forma particolare di modello parametrico, che utilizza una distribuzione di probabilità derivata come modifica di quella di Gumbel, appositamente costruita in modo da rispettare il comportamento medio osservato dei bacini (Maione *et al.*, 1999a) (Maione *et al.*, 1999b)

5.

Le osservazioni della portata al colmo sono pubblicate più raramente (almeno in Italia) di quelle della portata media giornaliera, e spesso mancano del tutto. E` dunque a volte opportuno stimare innanzi tutto la portata media giornaliera $q(T)$ con il tempo di ritorno T assegnato e quindi ricavare dalla stima di $q(T)$ quella della portata al colmo $Q(T)$ con lo stesso tempo di ritorno, moltiplicando $q(T)$ per una stima del rapporto R tra $Q(T)$ e $q(T)$, ricavata da un'analisi statistica regionale (Moisello, 1998).
Vale la pena di ricordare che molto spesso si assume come portata media giornaliera la media delle portate corrispondenti a poche letture dell'asta idrometrica, effettuate a ore fisse, o addirittura la portata corrispondente a un'unica lettura. Però tra gli errori, che qualche volta sono in difetto e qualche volta in eccesso, si ha una certa compensazione. Alla relazione tra portate al colmo e portate medie giornaliere sono stati dedicati diversi studi, basati su diverse ipotesi.
Un'ipotesi è che il rapporto R, assunto costante per una data sezione, dipenda dalle caratteristiche del bacino (Tonini, 1939) (Cotecchia, 1965) (Canuti e Moisello, 1980) (Todini e Serraglini, 1981). Un'altra ipotesi è che R dipenda solo dal tempo di ritorno (Lazzari, 1968). Un'altra ancora è che R sia una variabile casuale, con distribuzione indipendente dall'area del bacino (Versace e Principato, 1977).
Un altro modo di affrontare il problema è di considerare il rapporto R tra le portate massime Q e q osservate nello stesso anno come una variabile casuale, la cui distribuzione dipende dalle caratteristiche principali del bacino (Canuti e Moisello, 1987), ma non dal valore di q.