Idraulica e Idrologia: Lezione 6

Agenda del giorno

- **Eventi estremi in idrologia: legame fra magnitudo e probabilità;**
- **Statistica e probabilità in idrologia;**
- **Tempo di ritorno.**

Analisi statistica di dati idrologici

INTRODUZIONE

- I processi idrologici evolvono, nello spazio e nel tempo, secondo modalità che sono in parte predicibili (o ***deterministiche***) ed in parte ***casuali*** ***(stocastiche - dal greco 'stochastikós' ⇒ 'congetturale')*** in quanto non adeguatamente descritti dalle leggi fisiche in nostro possesso

La incerta comprensione dei fenomeni idrologici è dovuta essenzialmente a tre fattori:

- l'implicita aleatorietà dei fenomeni che innescano i processi idrologici (es.: precipitazione);

- la presenza, in tutte le misure idrologiche, di un'elevata componente di errore dovuta a problemi di campionamento (le misure idrologiche consentono di accedere ad un campione - spesso estremamente modesto - dell'insieme dei valori assunti dalla variabile idrologica considerata);

- la mancanza di un rigoroso inquadramento fisico di tipo teorico dei processi idrologici alle diverse scale spaziali e temporali alle quali tali processi si manifestano. Questo significa che, anche se si disponesse di misure esatte non affette da errori di campionamento, la previsione del comportamento del sistema idrologico sarebbe comunque affetta da errore.

Analisi statistica degli eventi idrologici estremi

L'ANALISI STATISTICA NELL'AMBITO DELL'ANALISI IDROLOGICA

- I sistemi idrologici sono talvolta investi da eventi estremi (precipitazioni, piene, siccità). ***La magnitudo di un evento estremo è inversamente proporzionale alla sua frequenza di accadimento***: eventi molto severi accadono meno frequentemente di eventi di moderata intensità.

- Obiettivo dell'analisi statistica dei dati idrologici: ***collegare la magnitudo degli eventi estremi alla loro frequenza di accadimento tramite l'impiego di distribuzioni di probabilità***.

- ***I dati idrologici utilizzati a tale scopo sono assunti essere indipendenti e identicamente distribuiti***. In pratica, questo si consegue utilizzando per le analisi statistiche i valori massimi annuali della variabile di interesse.
  Si consideri una serie di valori di portata massima annuale: ogni valore è considerato indipendente rispetto agli altri, e ciascun valore può essere considerato come la realizzazione di un processo stocastico identico per tutti gli anni considerati (questo implica che il sistema idrologico che produce tali valori estremi - il bacino idrografico a monte della sezione dove sono stati osservati i valori massimi annuali - non subisca variazioni nel periodo considerato. Variazioni tipiche sono quelle collegate all'uso del suolo, o alla costruzione di uno sbarramento).

Analisi statistica di dati idrologici

Alcune definizioni

Statistica e probabilità

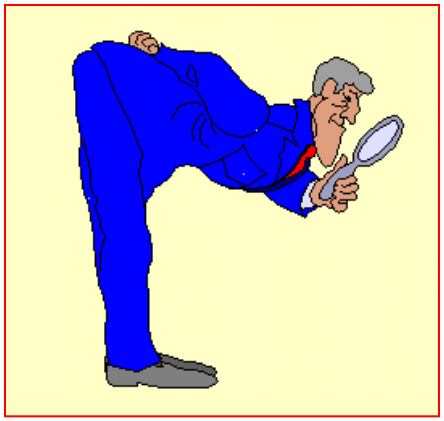
- Quando è possibile effettuare ripetuti esperimenti indipendenti, come nel lancio di un dado o di una moneta, è possibile riscontrare che la frequenza di un evento tende ad avvicinarsi ad un certo limite, anche se l'evento considerato rimane impredicibile. Questo effetto viene indicato con il termine *regolarità statistica*. Operando in laboratorio, è possibile evidenziare tale regolarità mediante la ripetizione di esperimenti indipendenti in circostanze identiche.

- In idrologia, tuttavia, i dati utilizzati sono di tipo osservativo piuttosto che sperimentale (la maggior parte degli esperimenti non possono essere ripetuti nelle stesse circostanze). Per questi dati, pertanto, la regolarità statistica non può essere dimostrata tramite la ripetizione dell'esperimento.

- La giustificazione per l'uso degli strumenti statistici e probabilistici in idrologia dipende quindi dalla rappresentazione espressiva che tali metodi forniscono circa il valore atteso e la variabilità di osservazioni future.

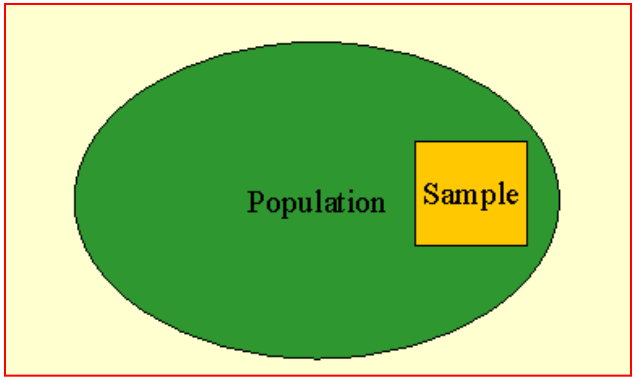
Statistica

- La statistica fornisce metodi per inferire le proprietà di una popolazione sulla base di quelle di un campione, a tale popolazione appartenente. Oltre alla semplice descrizione della popolazione, essa offre informazioni sulla incertezza relativa alla conoscenza della popolazione acquisita tramite il campione. Pertanto, la statistica consente di valutare il valore di un dato, misurato in termini di decrescita dell'incertezza. Valutare le incertezze è pertanto fondamentale per indirizzare gli sforzi di acquisizione di ulteriori dati.



Statistica

- Il concetto fondamentale in statistica è rappresentato dalla popolazione, ovvero dalla collezione completa di eventi ci cui interessa valutare le proprietà. Una popolazione può essere finita e numerabile esplicitamente (es.: la popolazione di tutti i corsi d'acqua sul pianeta terra che drenano più di 1000 km2), oppure infinita e definita in astrazione (es.: la popolazione di tutte le possibili portate che si possono verificare lungo un corso d'acqua presso una sua specifica sezione).



Statistica

Gli strumenti della statistica consentono quindi di:

- sviluppare metodologie per l'efficiente raccolta dei dati;
- individuare metodi per descrivere sinteticamente i dati raccolti;
- fornire una interpretazione dei dati;
- trarre conclusioni sulla base dei dati.

Probabilità

- La probabilità fornisce il fondamento teorico per i metodi statistici. La probabilità consente il calcolo della probabilità di un evento (ovvero, di osservare un determinato valore campionario), quando siano assegnate determinate caratteristiche della popolazione. Di converso, la statistica fornisce metodi per inferire le caratteristiche della popolazione, assegnate quelle del campione.

Stime e stimatori

- Uno stimatore è definito come una procedura per il calcolo delle caratteristiche della popolazione a partire da quelle del campione. Un valore specifico calcolato tramite uno stimatore è definito come stima.
- Uno stimatore distorto è costituito da uno stimatore che, utilizzato più volte, fornisce stime che, in media, differiscono rispetto a quelle della popolazione.

Trattamento probabilistico dei dati

DISTRIBUZIONI DI PROBABILITÀ

Il processo di inferenza statistica consente di identificare le proprietà statistiche della popolazione a partire da quelle del campione.
La funzione di probabilità così identificata consente di specificare, fra l'altro, la probabilità che un generico valore x venga (o non venga) superato.

Alcuni principi generali

La probabilità deve soddisfare una serie di assiomi:

1. La probabilità di un evento è un numero compreso fra 0 e 1
2. La probabilità dell'evento certo è 1
3. La probabilità di un evento ottenuto come somma di due eventi che mutuamente si escludono è pari alla somma delle probabilità dei due eventi.
4. La probabilità condizionata di un evento A dato che un evento B si è verificato è data dal prodotto delle probabilità di A e di B (se i due eventi sono indipendenti fra loro).

Trattamento probabilistico dei dati

ESEMPI

Nel caso del lancio di un dado, la probabilità che si verifichi uno dei 6 possibili risultati è pari ad 1/6.

La probabilità che si verifichi: 1 o 2 o 3 o 4 o 5 o 6 = 1/6+1/6+..1/6=1

La probabilità che, in due lanci successivi, si verifichi la prima volta 2 e la seconda 5 si può calcolare come una probabilità condizionata:
1/6 * 1/6 = 1/36

La probabilità che, in due lanci successivi, non si verifichi né 2 né 5 si può calcolare utilizzando il principio della probabilità totale e quello della probabilità condizionata:
p(A)= probabilità che in due lanci successivi si verifichino 2 e 5=1/36
p(B)= probabilità che in due lanci successivi non si verifichino né 2 né 5

$\Rightarrow P(A)+P(B)=1 \Rightarrow P(B)=1-P(A) \Rightarrow P(B)=1-1/36=35/36$

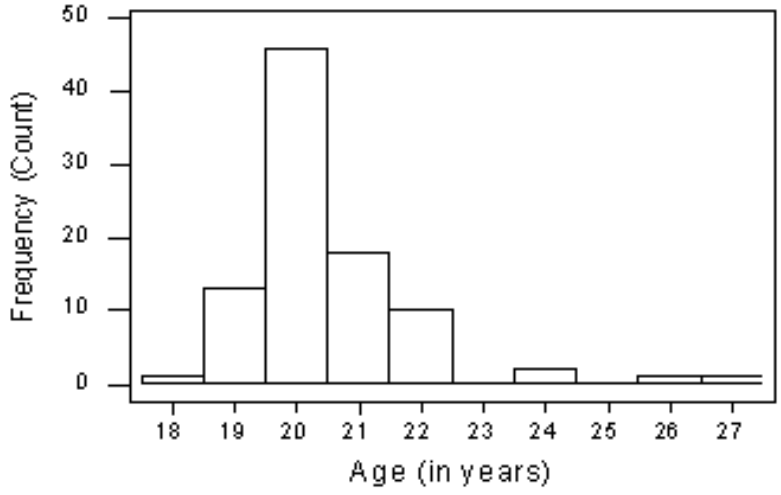
Statistica descrittiva

Passiamo agli strumenti

Sintesi e descrizione del campione di dati: istogramma

L'istogramma rappresenta la distribuzione di frequenza del campione, ottenuta:

- suddividendo l'intervallo di misura in **k** parti uguali (intervalli);
- determinando il numero di dati e la percentuale di dati che ricadono in ciascun intervallo (nel primo caso si ha la distribuzione di frequenza assoluta, nel secondo di frequenza relativa);
- riportando una linea a barre la cui altezza è proporzionale al numero ovvero alla percentuale di dati che ricadono in ciascun intervallo.



Il numero degli intervalli

Per ottenere una buona rappresentazione grafica della frequenza, il numero degli intervalli va scelto accuratamente.

Per un campione di numerosità **n**, il numero di intervalli **k** dovrebbe essere eguale al più piccolo numero intero tale che $\mathbf{2^k \geq n}$.

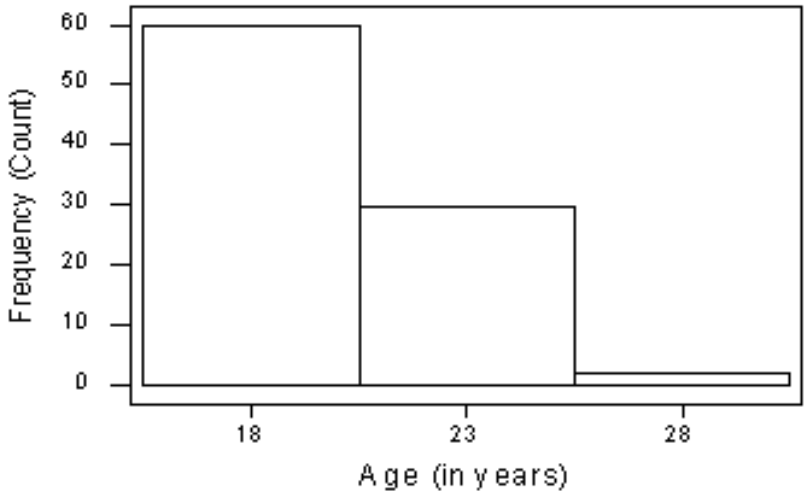
Esempio:

n=30

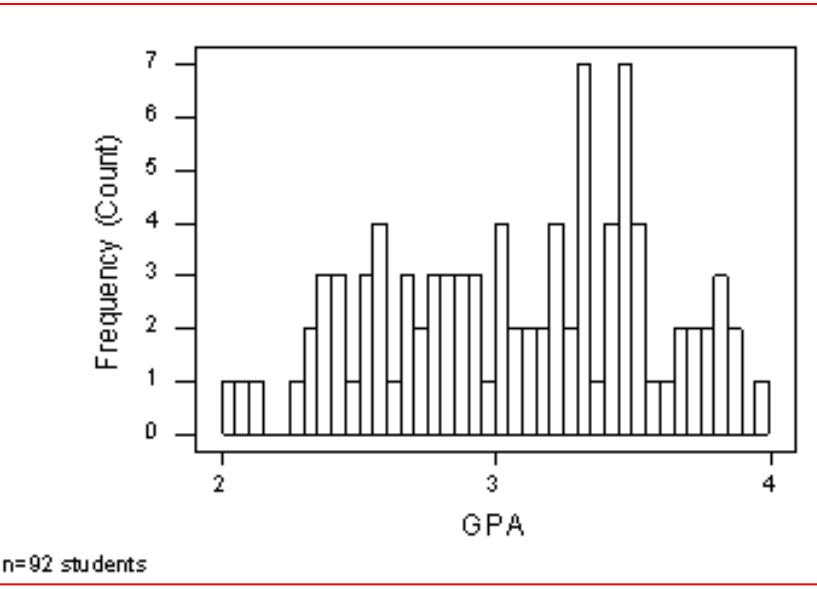
allora:

k=5

Istogramma



Numero intervalli eccessivamente ridotto



Intervalli eccessivamente numerosi

Distribuzioni di frequenza e di probabilità

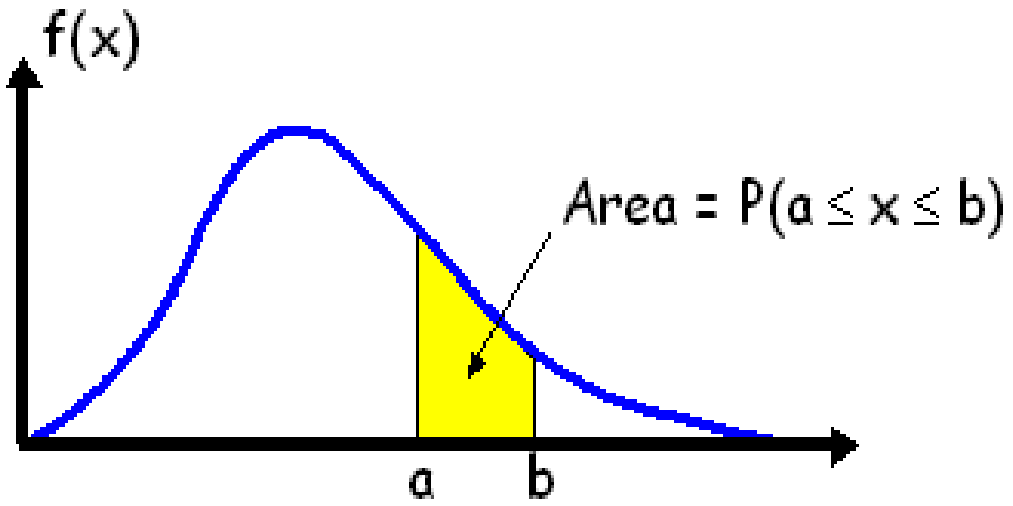
Le funzioni di **frequenza relativa** e di **frequenza cumulata** sono definite per il **campione**.

La funzione di **densità di probabilità** è rappresentata dal rapporto fra la funzione di frequenza cumulata e Δx, per n (numero di elementi del campione) tendente ad infinito e Δx tendente a zero.

La **funzione di probabilità** F(x) rappresenta la probabilità che la variabile X assuma un valore compreso fra l'estremo inferiore ed x. (Alcune distribuzioni sono definite per valori negativi di x).
F(x) indica la probabilità di non superamento

Distribuzione di probabilità

L'area sotto la curva rappresenta la probabilità.



La probabilità che x cada nell'intervallo [a,b] è pari a

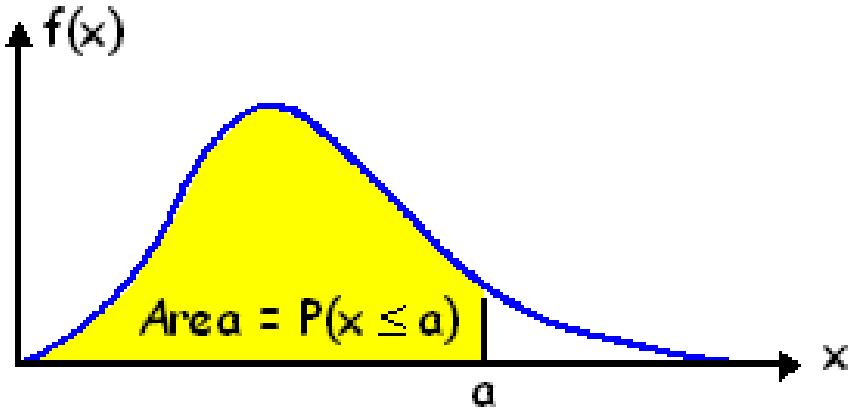
$$P(a \le x \le b) = \int_a^b f(x)dx$$

Distribuzione di probabilità cumulata

Definizione:
La probabilità cumulata di non superamento relativa al valore 'a' è la probabilità che un evento sia compreso fra il limite inferiore ed a, ovvero

$$F(a) = P(x \leq a)$$

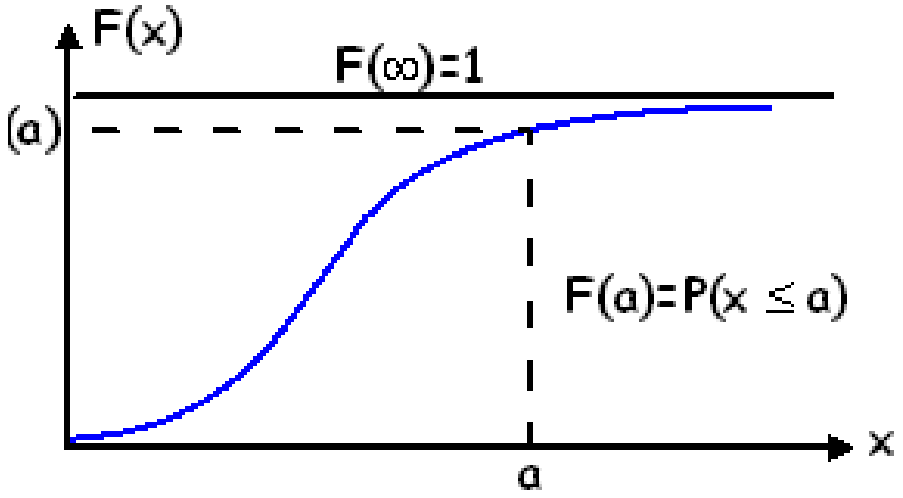
$$= \int_0^a f(x)\,dx$$



La probabilità cumulata di superamento relativa al valore 'a' è la probabilità che un evento sia superiore ad a, ovvero

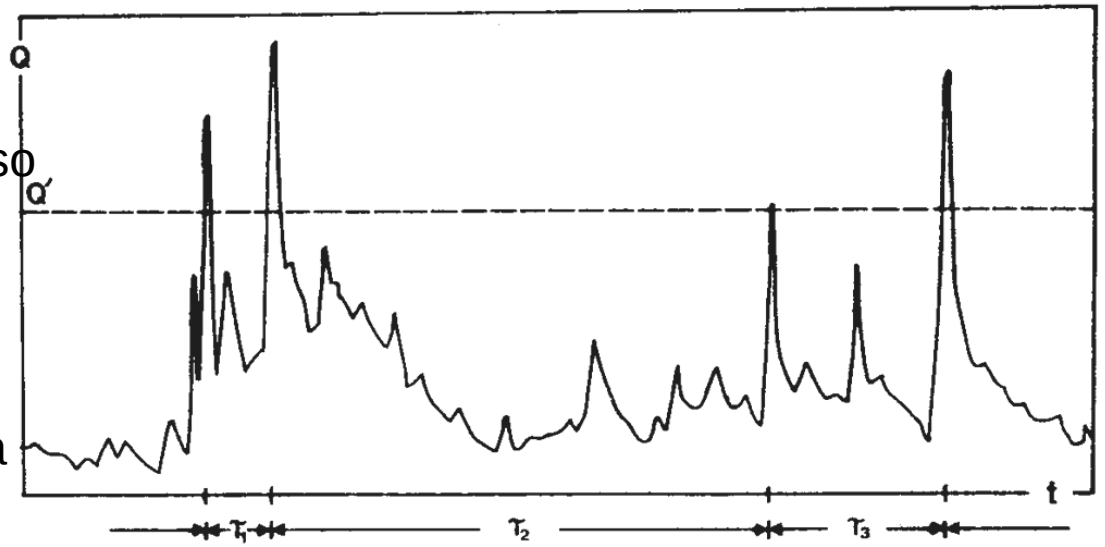
$$P(x > a) = 1 - P(x \leq a)$$

$$= 1 - F(a)$$



Tempo di ritorno - 1

Il concetto di probabilità di non superamento (associato ad un certo evento idrologico) è spesso sostituito (per ragioni di comodità di rappresentazione) da quello di tempo di ritorno T.
In realtà, il concetto di tempo di ritorno può essere introdotto senza far ricorso alla probabilità.



Si consideri l'idrogramma in figura, dove sono indicati gli intervalli τ compresi fra successi istanti associati a valori di portata maggiori di Q'. Se si considera un periodo di tempo molto esteso, per ogni valore arbitrario di portata Q' il tempo di ritorno corrispondente è definito come la media degli intervalli τ

$$T(Q') = media(\tau_1, \tau_2, \tau_3, \ldots, \tau_n)$$

Tempo di ritorno - 2

DEFINIZIONI

$$T = \frac{1}{P(X \geq x_T)}$$

dove x_T è la variabile caratterizzata da un tempo di ritorno T
$\Rightarrow$ se P=0.01, T=100 anni

Il tempo di ritorno di un evento di assegnata intensità è quindi:

- Numero di anni che in media separa il verificarsi di due eventi di intensità eguale o superiore a quella assegnata.
- Numero di anni in cui l'evento di intensità assegnata viene eguagliato o superato in media 1 volta.

In queste definizioni, la parola chiave è "in media". Infatti, il tempo di ritorno non è il numero di anni che separa due eventi di intensità eguale o superiore a quella assegnata. Secondo tale ultima definizione, dopo il verificarsi di un evento T-ennale (ovvero di probabilità di superamento 1/T), occorrerebbe attendere T anni affinché l'evento si ripeta (con certezza). Questo non è vero: infatti, la probabilità di un tale evento rimane pari ad 1/T in ciascun anno, indipendentemente dal verificarsi di un simile evento nell'anno precedente o in anni recenti.