



Corso Luigi Einaudi, 55 - Torino

**Appunti universitari**

**Tesi di laurea**

**Cartoleria e cancelleria**

**Stampa file e fotocopie**

**Print on demand**

**Rilegature**

NUMERO: 1846A -

ANNO: 2016

# **A P P U N T I**

STUDENTE: Aimar Mauro

MATERIA: Idrologia - prof. Claps

Il presente lavoro nasce dall'impegno dell'autore ed è distribuito in accordo con il Centro Appunti.

Tutti i diritti sono riservati. È vietata qualsiasi riproduzione, copia totale o parziale, dei contenuti inseriti nel presente volume, ivi inclusa la memorizzazione, rielaborazione, diffusione o distribuzione dei contenuti stessi mediante qualunque supporto magnetico o cartaceo, piattaforma tecnologica o rete telematica, senza previa autorizzazione scritta dell'autore.

ATTENZIONE: QUESTI APPUNTI SONO FATTI DA STUDENTIE NON SONO STATI VISIONATI DAL DOCENTE.  
IL NOME DEL PROFESSORE, SERVE SOLO PER IDENTIFICARE IL CORSO.

## INTRODUZIONE: RISORSE IDRICHE

Per iniziare, si parte da alcune opere:

→ una risorsa idrica è di sicuro una SORGENTE attorno cui si crea un'opera di captazione. In questo caso, ci si occupa di una traslazione diretta di portata da un punto all'altro x servire le persone

→ x erogare risorse idriche si può pensare anche a un lago ARTIFICIALE, pensato apposta x ricavarne acqua.  
Questo si fa quando si vuole usare l'acqua non quando è fornita spontaneamente ma quando si vuole, cioè si cambiano i tempi e si usa l'acqua anche quando non è disponibile.  
La diga stocca un volume che può essere messo da parte e confrontato con quello che mediamente arriva in 1 anno.

Es se un invaso ha volume massimo  $1250000 \text{ m}^3$  e il lago è di  $0,125 \text{ km}^3$ , tutta l'acqua invece arriva dalla porzione di territorio delimitata nella spartiacque, cioè il bacino imbrifero, di  $50 \text{ km}^2$ .  
Così la proporzione tra gli elementi è molto diversa.

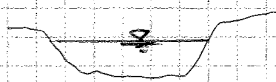
→ esiste anche un caso dove non c'è diga ma si crea una DERIVAZIONE D'ACQUA SENZA INVASO:

è una sorta di circonvallazione che prende parte della portata (se la prende tutta, si parla di DIVERDIVO).

Il prelievo avviene con una presa diretta in cui si preleva al più la portata che c'è al momento, ma non c'è possibilità di erogare la quantità nel tempo perché non c'è stoccaggio (se è in secca, non c'è nulla).



2) Quando si fa un'opera, ci si focalizza su un punto ma anche sulla superficie drenante. Tagliandola in un punto, si individua la sezione idrica.



Nelle valutazioni da lontano si guarda il corso.  
nelle valutazioni da vicino si guarda la sezione idrica.

Per sapere quanta acqua arriva, si parte dalla precipitazione

→ udm: mm

→ ordine di grandezza di interesse:  $10^3$  → in Italia la precipitazione media è di 760 mm.

Le piogge si distribuiscono su un'area, che in Italia vale  $300000 \text{ km}^2$ .

Ora,  $1 \text{ m}^3$  di pioggia corrisponde a 1000 mm di pioggia che cadono su  $1 \text{ m}^2$  di area.

$$1 \text{ m}^3 \equiv 1000 \text{ mm (su } 1 \text{ m}^2)$$

Allora su  $1 \text{ km}^2$  di area cade 1 milione  $\text{m}^3$

⇒ in Italia la pioggia annua al  $\text{km}^2$  è dell'ordine di grandezza di  $10^6 \text{ m}^3$  (valore molto grande)

OSSERVAZIONE: esistono casi in cui il coefficiente di deflusso è maggiore di 1.  
Questo non è anomalo perché la misura precisa dei liquidi in montagna è difficile (non si sa l'acqua equivalente alla neve)

+

non si sa se il ghiaccio è indifferente al deflusso

Così, se si fa una misura corretta delle precipitazioni, allora il ghiaccio sta sciogliendosi e perde massa, che si integra alla quota di precipitazione

EVENTO DI  
PROGETTO



STIMA DEL VALORE DELLA  
VARIABILE DI PROGETTO  
(portata)

①



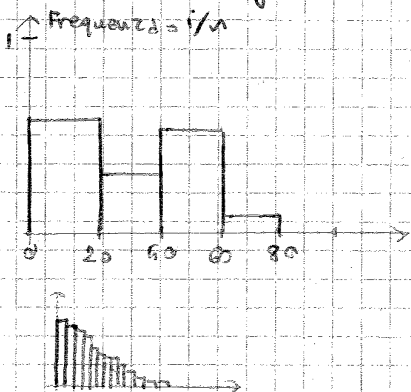
METODI DIRETTI (analisi con  
osservazione)

②



METODI INDIRETTI

Le osservazioni sono anche rappresentabili graficamente nella loro distribuzione mediante l'istogramma delle frequenze relative di classe:



qui si stabiliscono dei blocchi e li si mettono insieme delle osservazioni senza precisarle, si conta e si associa una frequenza, cioè il numero di osservazioni in una classe rispetto al numero totale.

Il diagramma può essere uniforme e tendere a occupare una certa zona, assumendo un certo trend.

il dettaglio aumenta usando un maggior numero di classi (e dunque più dati).

Per operare quantitativamente su questi dati, occorre una rappresentazione, cioè la curva di frequenza cumulata.

$i$	$z$	$i/n$
1	...	
2	...	

Questo si basa sul definire la colonna  $x$  della sequenza ordinata e una  $x$  il numero di posizione. A queste si affianca una colonna contenente un indicatore  $i/n$  che dà in termini adimensionali la posizione della grandezza in relazione al numero di osservazioni.

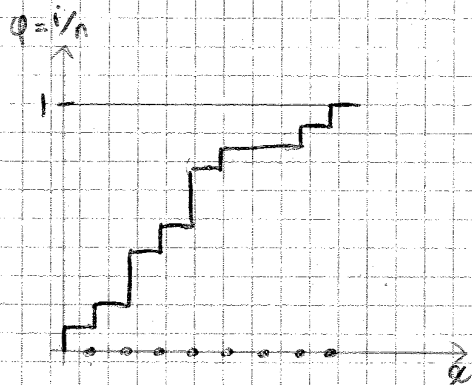
è una posizione adimensionale.

Questo indicatore si dice frequenza cumulata campionaria.

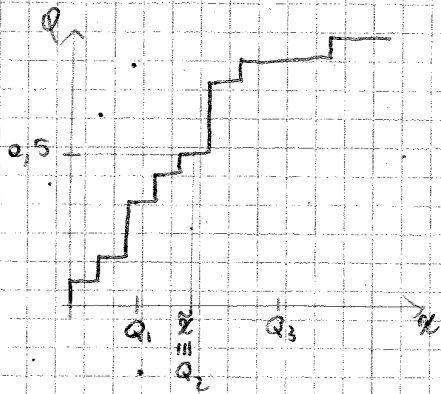
→ "frequenza cumulata" perché è distinta da quella relativa (dice solo quanto si ripete un dato) e serve a capire il numero di osservazioni inferiori a una certa entità.

PERCENTUALE DI VALORI CONTENUTA MINORE o UGUALE AL VALORE DI POSIZIONE  $z$

→ "campionaria" perché questa proviene dal campione osservato.



Questa tabella può essere rappresentata in un grafico riportando i punti osservati sull'asse delle ascisse. La curva che si ottiene presenta in genere uno scatto pari a  $1/n$ , altrimenti se ci sono salti doppi, significa che c'è 2 volte lo stesso valore.



La mediana  $\tilde{x}$  serve a costruire il grafico che indica la distribuzione dei dati:

essa infatti indica, quando incrocia la curva, la frequenza relativa di 0,5.  
 Uoto il valore  $\tilde{x}$ , si può dividere ulteriormente la curva campionaria in quartili:

essi sono i valori di  $x$  relativi alle frequenze cumulate che si ottengono dividendo in 4 l'unità.

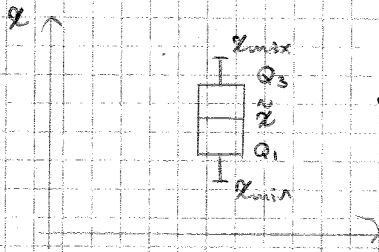
I quartile = posizione corrispondente alla frequenza 0,25

II quartile = coincide con  $\tilde{x}$

III quartile = corrisponde alla frequenza 0,75

**PROPRIETÀ DEI QUARTILI**

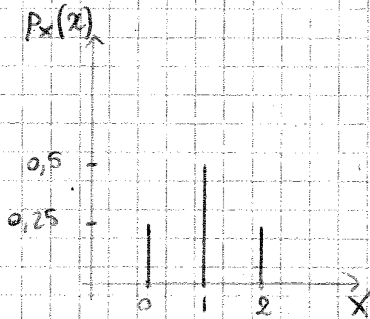
il quartile è distinto dal quantile, che è un valore  $x$  dipendente corrispondente a una generica frequenza cumulata  $\varphi$  (è la generalizzazione del concetto di quartile)



→ con i quartili si disegna un diagramma detto box plot avente in ordinata la grandezza  $x$ . In esso c'è una "scatola" avente  $\tilde{x}$  in mezzo e  $Q_1$  e  $Q_3$  agli estremi, da cui si dipartono delle stanghe con  $x_{min}$  e  $x_{max}$ . Questa è una rappresentazione schematica e utile a confrontare diverse serie di osservazioni e indica quanto distano i valori estremi rispetto al blocco centrale (stecche rispetto alla scatola).

In questo modo si è definito lo spazio campionario, che si definisce quando si circoscrivono le possibilità che derivano da questo esperimento.

Es. Nel caso della portata, ci sono infinite risposte possibili e lo spazio campionario non è circoscritto.



La situazione nella tabella può essere poi descritta con il diagramma di massa di probabilità, in cui si può calcolare con certezza la probabilità di accadimento di un evento.

Tale diagramma può essere costruito solo con **VARIABILI ALEATORIE DISCRETE**, cioè valori fissi, interi e relativi.

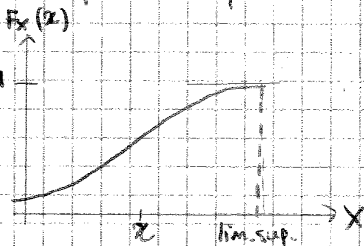
In questo diagramma si può assumere una quantità pura  $x$  esprimere la possibilità di ogni evento e da qui rispondere alle domande

→ di probabilità

→ di probabilità cumulata

Esistono però casi (v. portata) in cui si ha una grandezza continua - una variabile aleatoria continua varia con continuità e  $x$  essa non si può più definire un valore puro di probabilità perché è una grandezza continua e varia senza distribuzione e di classe.

Dunque non si può descrivere un diagramma di massa di probabilità.



Allora, assegnata una scala di variabile  $X$ , si sceglie un valore  $z$  e si valuta la **PROBABILITÀ DI NON SUPERAMENTO** di quel valore  $z$ .

Cioè ci si chiede la probabilità di ottenere un valore  $X \leq z$

$$P(X \leq z)$$

Così prima si poteva formulare la domanda  $P(X=1)$ , data dall'esperimento. Ora la variabile è continua e non finita e occorre formulare la domanda diversamente. Infatti,  $P(X=1)$  sarebbe formalmente nulla perché si chiederebbe di definire precisamente un numero reale, ma si può dare solo un'approssimazione.

Si può però chiedere se si supera o meno e  $x$  rispondere, si costruisce una curva e una funzione continua, indicata con  $F_X(z)$  e detta **funzione di ripartizione** o **funzione di probabilità cumulata**.

Con questa funzione si ottiene una probabilità, cioè un numero compreso tra 0 e 1, in modo da soddisfare la domanda.

⇒ dunque la funzione non supera mai 1.

Se si trova 1, si è allora preso il valore estremo possibile e dunque la variabile aleatoria è continua e limitata superiormente.

Tipicamente però non c'è limite superiore e la curva si perde e tende a 1.

$$F_X(z) \leq 1$$

L'obiettivo nell'ambito della stima della grandezza di progetto è costruire questa curva e questa è una curva teorica, poiché definisce il andamento di tutto il fenomeno.



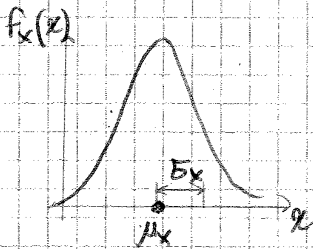
→ tutti gli altri momenti teorici sono momenti centrali di ordine  $r$ , cioè calcolati in rispetto a un valore centrale (in genere la media)

$$\mu_r' = \int_{-\infty}^{+\infty} (x - \mu_x)^r f_x(x) dx = E[(X - \mu_x)^r]$$

si applica l'operatore  $E$  alla differenza

Da qui si definisce il momento centrale di ordine 2, cioè la varianza teorica

$$\mu_2' = \text{VAR}(X) = \int_{-\infty}^{+\infty} (x - \mu_x)^2 f(x) dx$$



Si nota che il valore  $\sigma_x = \sqrt{\text{VAR}(X)}$  corrisponde al raggio d'inerzia della pdf, cioè indica dove si dovrebbe posizionare la massa rispetto al baricentro  $x$  avere lo stesso momento d'inerzia

↓

lo  $\sigma$  campionario invece non ha significato geometrico

Formulazione dell'obiettivo di progetto (definizione della condizione di progetto):

essa rappresenta chiaramente a quali condizioni sottoporre la variabile di progetto e sfrutta l'inferenza statistica.

In questa formulazione, si sa che la variabile è aleatoria e si può attribuire un valore solo se si assegna una probabilità.

Dunque si fissa innanzitutto il CAMPO DELLE PROBABILITÀ:

assegnato un valore  $x_0$  di soglia della grandezza, si stabiliscono i valori

$$P(x \leq x_0)$$

$$P(x > x_0)$$

Dunque si stabiliscono 2 eventi semplici, che si superi o meno una soglia  $x_0$ . I valori sono dati da  $p$

$$P(x \leq x_0) = 1 - p$$

$$P(x > x_0) = p$$

↓ esperimento aleatorio molto semplice.

Poiché si ragiona ai fini della sicurezza,  $p$  si definisce come probabilità di insuccesso.

Ora, un esperimento aleatorio di questo tipo, con 2 soli esiti, si dice ESPERIMENTO BERNOLLIANO.

I valori di probabilità dipendono dal valore  $x_0$ .

Una possibile condizione di progetto è allora quella  $x$  cui

$$T_p = 1$$

In questo caso,  $T$  assume significato del numero medio di prove che danno luogo a un insuccesso (che può capitare anche all  $2^a$  prova).  
Dunque  $T$  vale

$$T = \frac{1}{p}$$

e si chiama tempo di ritorno.

Questo è il metodo più utilizzato a spiegare alla committenza i rischi ma non è chiaro del tutto.

Infatti, con questo si dà un significato vicino all'esperienza ma  $T$  significa scegliere quel numero di prove il cui numero medio di insuccessi è  $p$ .  
Dunque  $T$  non è altro che una modalità alternativa a definire il reciproco della probabilità semplice di insuccesso (cioè che nel prossimo esperimento c'è l'insuccesso).

In realtà, una costruisce un'opera perché duri negli anni e dunque valuta quanto debba durare, costruendo l'orizzonte progettuale  $L$  (in anni).

Allora una modalità semplice a definire la condizione di progetto è:

costruire la probabilità di  $1$  superamento in  $L$  anni consecutivi, data la probabilità semplice  $p$  di superamento (e da qui prendere decisioni progettuali).

Questo si chiama rischio residuale (o rischio in  $L$  anni) poiché, pur essendosi impegnati a un basso rischio, c'è un valore che può essere superato.  
Per costruirlo, si passa attraverso la probabilità di successo in  $L$  anni (cioè di avere sempre successo in  $L$  anni consecutivi):

si fa un esperimento composto dove il risultato è dovuto a  $L$  prove affidate e si usa l'intersezione, a valutare la probabilità di avere successo contemporaneamente in  $L$  prove

$$P(s) \wedge P(s) \wedge P(s) \wedge \dots \wedge P(s)$$

Secondo l'assioma dell'intersezione, esso vale il prodotto delle probabilità:

$$P(s) \wedge P(s) \wedge \dots \wedge P(s) = (1-p)^L \quad \rightarrow \text{SICUREZZA IN } L \text{ ANNI}$$

Allora il rischio residuale, che è l'opposto del caso di solo successi, vale

$$R_L = 1 - (1-p)^L$$

$$R_L = 1 - (1-p)^L$$

Es. Si assume un periodo di ritorno

$$T = 50 \text{ anni}$$

che si associa a una probabilità

$$p = \frac{1}{50} = 0,02$$

Ora si è svolta la 2° fase dell' inferenza statistica, cioè l'analisi preliminare delle osservazioni.

In particolare, le fasi dell' inferenza sono:

- I) ANALISI PRELIMINARE
- II) FORMULAZIONE DELLE IPOTESI SUL MODELLO PROBABILISTICO (quale funzione di probabilità)
- III) STIMA DEI PARAMETRI
- IV) VERIFICA DELLE IPOTESI, con esito positivo o riformulazione dell'ipotesi
- V) FIDUCIA NEL RISULTATO e costruzione delle fasce di confidenza (cioè margini d'incertezza quantitativi che si accompagnano al risultato)

### Ipotesi di modello

Oggi i software consentono di selezionare tra i vari modelli ma ciò non funziona con le applicazioni idrologiche, poiché hanno pochi dati e serve attenzione nella scelta.

Inanzitutto bisogna partire dalle INFORMAZIONI A PRIORI.

Ad es se si valuta la precipitazione totale annua, questa è somma di tante grandezze osservate, cioè è somma di tanti valori casuali. Sfruttando il teorema del limite centrale, si può modellare la quantità come una normale. Poi magari non va bene perché il numero di giorni di precipitazione non è così alto.

La scelta è anche aiutata dall' ESPERIENZA, cioè dai risultati già ottenuti. Inoltre, si potrebbe creare autonomamente un esperimento, ma ciò si applica poco in idrologia.

Prima di procedere a vedere come trattare informazioni ed esperienza, occorre richiami sulla distribuzione normale.

Essa in genere si scrive in forma di pdf, come

$$f_X(x) = \frac{1}{\sigma_2 \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x - \mu_1}{\sigma_2} \right)^2}$$

Notazione:  $f_X(x)$  è la funzione della variabile  $X$  calcolata in  $x$

La funzione di ripartizione si calcola come

$$F_X(x) = \int_{-\infty}^x f_X(t) dt = \int_{-\infty}^x \frac{1}{\sigma_2 \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{t - \mu_1}{\sigma_2} \right)^2} dt$$

Per essa, è possibile definire la distribuzione gaussiana in forma canonica, come

$$f_U(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} u^2}$$

La variabile  $u$  si dice VARIABILE RIDOTTA e vale

## RAPPORTI TRA PARAMETRI E MOMENTI NELLA GAUSSIANA

Si sa che il momento centrale di ordine  $r$  di una funzione vale

$$\mu_r' = \int_{-\infty}^{+\infty} f_X(x) (x - \mu_x)^r dx$$

Si nota che la VARIANZA TEORICA è legata a  $\theta_2$

$$\mu_2' = \theta_2^2$$

Invece  $\theta_1$  è la MEDIA TEORICA

$$\mu = \theta_1$$

Per vedere cosa succede a questi risultati in termini della variabile ridotta, bisogna prima vedere 2 proprietà delle funzioni di valori

→ la media di una funzione lineare è pari alla funzione applicata alla media

$$E(g(x)) = g(E(x)) \quad \text{vale se } g(x) = \frac{x - \theta_1}{\theta_2}, \text{ ad esempio.}$$

Ciò non vale × funzioni non lineari (es  $g(x) = \log x$ )

→ la varianza di una costante è nulla e la varianza di un prodotto × una costante è pari alla varianza × il quadrato della costante

$$\text{var}(c) = 0$$

$$\text{var}(cx) = c^2 \text{var} x$$

Di conseguenza, con funzione ~~usa~~ si ha

$$\text{var}\left(\frac{x - \theta_1}{\theta_2}\right) = \frac{1}{\theta_2^2} \text{var}(x)$$

Ora, nell'ambito della variabile canonica, si ha

→ media

$$\begin{aligned} E(U) &= E\left(g(x)\right) = \\ &= \frac{E(x) - \theta_1}{\theta_2} = \quad \rightarrow E(x) = \theta_1 \\ &= \frac{\theta_1 - \theta_1}{\theta_2} = 0 \end{aligned}$$

→ varianza

$$\begin{aligned} \text{var}(U) &= \text{var}(g(x)) = \\ &= \frac{1}{\theta_2^2} \text{var}(x) = \\ &= \frac{1}{\theta_2^2} - \theta_2^2 = 1 \end{aligned}$$

Ora si consideri una distribuzione gaussiana:

essa ha una forma a campana simmetrica, con centro di simmetria in  $\theta_1$ .  
Si nota che  $\theta_1$  coincide con

→ media

→ mediana

→ valore modale: valore della variabile a cui si ha il massimo valore della densità di probabilità

$$z : f(z) = \max$$

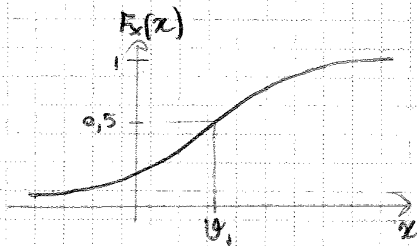
$$\theta_1 = \mu; \tilde{x}; z$$

3 significati diversi (ascissa del baricentro, limite del 50% e massimo) che si posizionano nello stesso punto

Invece il valore  $\theta_2$  si lega alla varianza teorica

$$\theta_2 = \sqrt{\text{var } x}$$

e dunque assume significato di raggio d'inerzia.  
 $\theta_2$  ha significato di dispersione, poiché indica quanto i dati sono dispersi rispetto al valore medio.



Si consideri ora la funzione di ripartizione:

il suo valore in  $\theta_1$ , poiché questo corrisponde al valore mediano (al di sotto la probabilità di non superamento è uguale a quella di superamento), è pari a 0,5

$$F_X(\theta_1) = 0,5$$

Questo vale solo x la gaussiana dove i 3 valori coincidono.  
Per questa cresce e tende all'unità x  $x \rightarrow \infty$ .

Ragionando invece in termini della variabile canonica, si sa che essa è una funzione  $g(x)$  costruita come

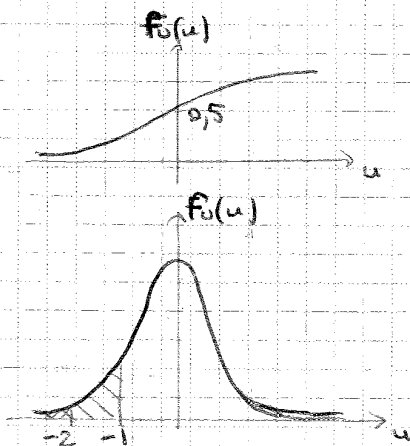
$$u = \frac{x - \theta_1}{\theta_2}$$

Si è visto che

$$F_U(u(x)) = F_X(x)$$

cioè se i valori sono gli stessi, si ottiene una funzione di ripartizione riscalata sull'asse  $x$ .  
Pertanto, tale funzione assume valore 0,5 nello zero (poiché  $\theta_1 = 0$ ).

Corrispondentemente, la pdf è una funzione simmetrica rispetto all'asse  $y$



Ora si inizia a vedere l'applicazione dell'inferenza:

Stima dei parametri, con l'ipotesi di DISTRIBUZIONE GAUSSIANA.

Un metodo molto semplice nell'applicazione è il metodo dei momenti;

i momenti teorici devono essere uguali ai momenti campionari

Da qui si forma un sistema di  $m$  equazioni in  $m$  incognite, dove le incognite sono  $\theta_1, \theta_2$  ed eventuali altri parametri (non presenti nella gaussiana)

Si ricorda che i momenti teorici sono funzioni definite tramite integrali poiché la funzione integranda dipende dai parametri (alcuni incogniti)

$$\mu_1 = h_1(\theta_1; \theta_2) = \int_{-\infty}^{+\infty} x f(x) dx$$

$$\mu_2 = h_2(\theta_1; \theta_2) = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx$$

Invece i momenti campionari sono valori numerici

$$\bar{x} = \sum \frac{x_i}{N}$$

Quando si fa l'uguaglianza, si ottiene un sistema  $2 \times 2$

$$\begin{cases} h_1(\theta_1; \theta_2) = \hat{x} & \rightarrow \text{STIMA DELLA MEDIA CAMPIONARIA} \\ h_2(\theta_1; \theta_2) = \hat{s}_x^2 \end{cases}$$

Si parla di stima perché questa operazione deve funzionare sia quando non si può scegliere il campione (le osservazioni non dipendono da chi osserva) sia quando si può sceglierlo.

Ricordando che il campione è una sequenza di osservazioni, questo può essere creato artificialmente inventandosi una distribuzione ed estraendo casualmente  $N$  ~~straneri~~ valori.

Ai strumenti dell'inferenza statistica devono dimostrare che il campione è "pulito" e dunque che ricrei la funzione originaria fedelmente

↓ devono funzionare soprattutto in questi casi

Si dimostra <sup>allora</sup> che, quando si opera il confronto tra i momenti, occorre operare delle correzioni x evitare che il risultato sia distorto.

In particolare, bisogna correggere una distorsione che è insita nelle relazioni della statistica descrittiva, in modo che i momenti campionari rappresentino la migliore stima dei momenti teorici

→ la MEDIA È UN OPERATORE INDISTORTO e dunque non necessita correzioni

$$\hat{x} = \bar{x} = \frac{1}{N} \sum x_i$$

Il problema si lega al fatto che nella curva di frequenza cumulata ci sono coppie

$$(x_i; \frac{i}{N})$$

e così l'ultimo punto ha ordinata 1. Questo è un errore sistematico perché magari la misurazione si è interrotta e ciò ha portato a dire che c'è un valore massimo non superabile.

Questo errore si elimina assumendo una CORREZIONE SULLE FREQUENZE CUMULATE  $\hat{\Phi}$ :

questa è fatta attraverso la formulazione

$$\hat{\Phi}(x_i) = \frac{i - \alpha}{n + 1 - 2\alpha}$$

dove  $\alpha$  dipende dalla distribuzione che si indaga. In assenza di informazioni, si pone

$$\alpha = 0 \Rightarrow \hat{\Phi}(x_i) = \frac{i}{n+1} \quad \text{PLOTING POSITION}$$

Questa correzione corrisponde ad assumere una posizione di plattaggio nella curva di distribuzione e sarà quella usata.

Il valore che si ottiene non è altro che una stima del valore relativo alla popolazione in  $x_i$  e dunque, più correttamente, si scrive

$$\hat{F}(x_i) = \frac{i}{N+1}$$

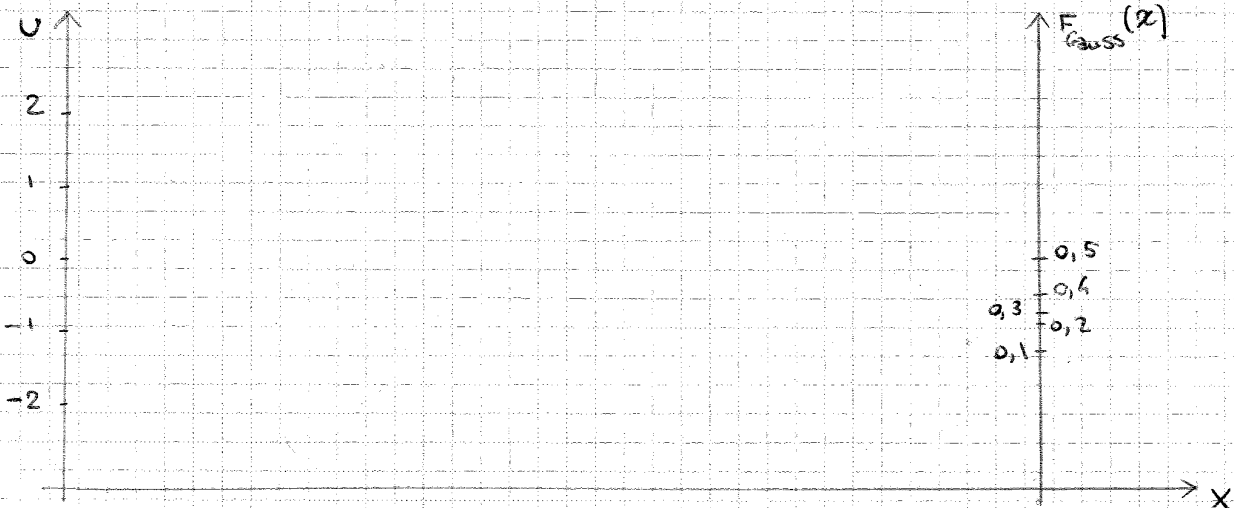
Stima della probabilità cumulata della popolazione in corrispondenza di  $x_i$  (con questo si può fare il confronto)

Da qui deriva che

$$\hat{F}(x_{\max}) = \frac{N}{N+1} < 1$$

### Verifica

Il metodo più speditivo è il metodo grafico: esso si fa attraverso la carta probabilistica



1 Si sa che tipicamente la distribuzione non è normale

2 Si parte da un problema reale:

una stazione idrometrica, inanzitutto, è uno strumento per misurare e trasmettere informazioni sugli ingressi d'acqua in tempo reale.

Tipicamente essa presenta un pluviometro, un pannello solare x ricaricare le batterie e lo strumento analogico-digitale x convertire i dati e salvarli in una memoria solida. C'è anche un trasmettitore.

Il problema consiste nel vedere se il ponte è sormontato o, peggio, se fa da diga. Dunque bisogna valutare la probabilità di superamento della portata massima che può essere convogliata nel ponte.

Il ponte è ad arco e in quel tratto il corso d'acqua è assimilabile a un canale rettangolare vegetato sulla sponda (alta scabrezza).

Nota la sezione, si può fare la valutazione di stato critica cioè di portata massima.

Prima però occorre vedere l'asta del torrente perché così si valuta il bacino di raccolta e si può ottenere una grande idea

⇒ bacino e sezione servono a valutare se può esserci una situazione di pericolo

Il problema consiste in un PROBLEMA DI VERIFICA poiché il ponte esiste e si vuole studiare la sicurezza associata.

Si assume di conoscere la massima portata che fluisce liberamente sotto il ponte. A tal proposito, si considera il massimo storico perché comunque è il valore massimo che è passato senza problemi.

$$Q_{max} = 1650 \text{ m}^3 \text{ s}^{-1}$$

Il problema consiste nello stimare la probabilità di superamento

$$P(x > x_{max})$$

cioè valutare la quantità

$$1 - F(x_{max}) = p$$

Si sa che da  $p$  si può ricavare il periodo di ritorno di  $x_{max}$

$$T(x_{max}) = \frac{1}{p}$$



Dunque, visto che la gaussiana spesso non è adatta, si riprende l'ipotesi sul modello probabilistico, cioè si sceglie una nuova funzione di distribuzione di probabilità teorica

### Distribuzione lognormale

Essa è una pdf imparentata con la normale, che può essere utile x il calcolo dei parametri.

Essa nasce attraverso una variabile di trasformazione

$$y = \ln x$$

tale x cui la pdf è una normale

$$F_Y(y) \sim N(\mu_Y; \sigma_Y)$$

In questo senso, Y è la distribuzione di una TRASFORMATA

Nel caso delle funzioni trasformate, si ricorda che la probabilità cumulata ~~non~~ è la stessa

$$F(y(x)) = F(x)$$

In base a questa relazione, si può trovare la pdf della variabile x. Infatti, la relazione si estende ai differenziali

$$dF_Y(y) = dF_X(x)$$

Di conseguenza, si può riscrivere l'uguaglianza in termini di derivate

$$\frac{dF_Y}{dy} dy = \frac{dF_X}{dx} dx$$

La derivata della funzione di ripartizione corrisponde però alla pdf

$$\begin{aligned} f_X(x) &= \frac{dF_X}{dx} = \\ &= \frac{dF_Y}{dy} \frac{dy}{dx} = \\ &= f_Y(y) \frac{dy}{dx} \end{aligned}$$

Si sa che Y è gaussiana e, x ignorare la parentela tra X e Y, si esplicita la funzione  $f_Y(y)$ , sostituendo a y il  $\log x$

$$f_X(x) = \frac{1}{\sigma_{2,y} \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{\log x - \mu_{1,y}}{\sigma_{2,y}} \right)^2} \cdot \frac{1}{x}$$

↓  
derivata di  $\log x$

Ora si prova ad applicare la lognormale nell'inferenza statistica.

Un metodo x la stima dei parametri è il METODO DEI MOMENTI:

si costruisce un sistema di equazioni in  $n$  incognite a partire dalla funzione di momento.

La lognormale ha 2 parametri incogniti e si sfruttano le equazioni:

$$\begin{cases} \mu(x) = \bar{x} \\ \sigma^2(x) = \frac{s_x^2}{\bar{x}^2} \end{cases}$$

La risoluzione è un po' più complessa ma si può procedere x sostituzione e si ottengono le formule x la stima dei parametri:

$$\begin{cases} \theta_1 = \ln \bar{x} - \frac{1}{2} \theta_2^2 \\ \theta_2^2 = \log \left( 1 + \frac{s_x^2}{\bar{x}^2} \right) \end{cases}$$

$$\begin{cases} \theta_1 = \ln \bar{x} - \frac{1}{2} \theta_2^2 \\ \theta_2^2 = \log \left( 1 + \frac{s_x^2}{\bar{x}^2} \right) \end{cases}$$

In realtà esistono altri metodi, anche perché il metodo dei momenti va in crisi già in distribuzioni a 3 parametri.

Qui servirebbe una III equazione dove interviene il coefficiente di asimmetria, che è dato dalla somma di distanze al cubo. ciò significa che, avendo anche un solo valore molto lontano da un punto, questa distanza al cubo pesa moltissimo. Infatti tale parametro dipende molto dalla distanza anche di un solo punto.

Così questo metodo va in crisi con formulazioni più complesse perché, con più parametri, bisogna usare dei momenti il cui valore dipende molto da occasionalità.

⇒ il metodo non è molto robusto e basta modificare un dato x stravolgere tutto.

Un altro metodo è il metodo della massima verosimiglianza:

essa si basa sulla definizione di una funzione detta funzione di verosimiglianza.

Se si vuole stimare una generica distribuzione a più parametri

$$f(x; \theta_1; \theta_2; \dots; \theta_r)$$

o, meglio, stimare i parametri  $\theta_1, \dots, \theta_r$  sulla base di  $n$  osservazioni:

$$\{x_1; x_2; \dots; x_n\}$$

si definisce allora la funzione

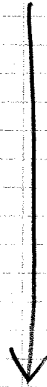
$$V = f_x(x_1) f_x(x_2) \dots f_x(x_n)$$

Dal calcolo del massimo, si ottiene la II relazione

$$\frac{1}{2\theta_2^2} \sum_i (x_i - \theta_1)^2 + \frac{n}{2\theta_2} = 0$$

Da qui si ottiene

$$\theta_2^2 = 5\theta_1^2$$



Di conseguenza, i parametri ottenuti con il metodo di verosimiglianza con riferimento alla gaussiana sono

$$\theta_1 = \bar{x}$$

$$\theta_2 = \sqrt{\frac{\sum x_i^2}{5n}}$$

cioè corrisponde allo stesso risultato ottenuto con il metodo dei momenti

↓ solo x la gaussiana, questo metodo è uguale dal punto di vista delle formule al metodo dei momenti

→  $L_3$  corrisponde a  $6b_2 - 6b_1 - b_0$   
 Il termine  $b_2$  vale

$$b_2 = \frac{1}{n} \sum_{i=3}^n \frac{(j-1)(i-2)}{(i-1)(i-2)} x_i$$

$$L_1 = b_0$$

$$L_2 = 2b_1 - b_0$$

Si nota che nel sistema ci sono differenze

→ non c'è la quantità  $x_i - \bar{x}$  ma solo  $x_i$ .  
 Dunque i momenti sono somma delle posizioni e danno idea della dispersione e dell'asimmetria.

→ non ci sono distanze al quadrato, al cubo, etc. ma le  $x_i$  pesano tutte con ordine 1.

Così tutti gli  $L$ -momenti dipendono da  $x$  e così non c'è più il problema che si ottengano valori (varianza, asimmetria) molto grandi e diversi tanto che non si ha sensibilità nel percepire errori.

Invece gli  $L$ -momenti hanno l'unità di misura delle  $x$  e dunque è più facile percepire gli errori (e si evita così il rischio di errore perché non si riesce a percepire il risultato).

Ora si scrivono le relazioni teoriche tra la stima dei parametri con il metodo degli  $L$ -momenti.

→ LEGGE NORMALE

$$\hat{\theta}_1 = L_1$$

→ perché  $L_1$  è rappresentazione della media campionaria  
 →  $L_1$  è il momento campionario di ordine 1

$$\hat{\theta}_2 = \sqrt{L_2}$$

→ DISTRIBUZIONE LOG-NORMALE

$$\hat{\theta}_1 = \ln L_1 - \frac{\theta_2^2}{2}$$

→ formulazione analoga a quella del metodo dei momenti

$$\hat{\theta}_2 = \sqrt{2} \cdot \text{INV. NORM. ST.} \left( \frac{L_2(x)}{L_1(x)} + 1 \right)$$

→ interviene l'inversa della pdf standard, che è una funzione che analiticamente non esiste e si risolve numericamente

A questo punto, si può procedere alla VERIFICA QUALITATIVA IN CARTA PROBABILISTICA.  
 Si sfrutta sempre la proprietà

$$y = \ln x \sim N(\mu_{1y}; \sigma_{2y})$$

In fatti, basta tracciare la variabile  $y$ , cioè  $\ln x$ , correlata alla variabile  $r$  data.

$$u = \frac{y - \mu_{1y}}{\sigma_{2y}}$$

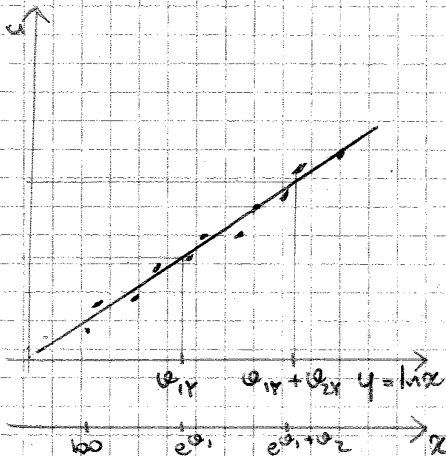
↓ si modificano i dati e si fa il diagramma

La scala ora è un po' più piccola ma, x tracciare la retta, servono 2 punti, cioè

$$\rightarrow \mu_{1y}$$

$$\rightarrow \mu_{1y} + \sigma_{2y}$$

Poi si piazzano i punti nella carta.



In genere, in ascissa si usa una scala distorta logaritmica, x cui le posizioni corrispondenti ai punti di tracciamento della retta ~~debbono~~ essere

$$e^{\mu_1}$$

$$e^{\mu_1 + \sigma_2}$$

La distribuzione di questa grandezza derivata si ricava tramite gli assiomi di probabilità:

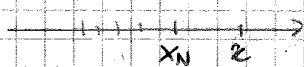
x definizione, la Funzione di ripartizione è una probabilità cumulata

$$F_{X_N}(x) = P(X_N \leq x)$$

Il valore  $X_N$  è però il massimo di una sequenza ordinata e dunque si può scrivere

$$F_{X_N}(x) = P[(X_1 \leq x) \wedge (X_2 \leq x) \wedge \dots \wedge (X_N \leq x)]$$

↓



se  $X_N \leq x$ , sono minori anche gli altri e dunque la probabilità è intersezione delle probabilità

Se i valori sono estratti in modo indipendente (EVENTI INDIPENDENTI), si può scrivere

$$\begin{aligned} F_{X_N}(x) &= P(X_1 \leq x) P(X_2 \leq x) \dots P(X_N \leq x) = \\ &= F_{X_1}(x) F_{X_2}(x) \dots F_{X_N}(x) \end{aligned}$$

→ prodotto delle probabilità cumulate

Ora, la probabilità cumulata dei vari eventi  $X_i$  è la stessa (non c'è motivo che la piena dell'esperimento  $X_i$  cambi).  
In questo caso, si parla di PROBABILITÀ CUMULATE IDENTICHE e si ottiene

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

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

Funzione di ripartizione del massimo valore

Questa funzione di ripartizione è stata studiata da Gumbel, che valuta una serie di alternative alla funzione di distribuzione di massimo sulla base di una serie di ipotesi:

→ si riescono a ottenere soluzioni analitiche se il numero  $N$  è grande e si parla di soluzioni asintotiche.

$N$  alto  $\Rightarrow$  soluzioni asintotiche

→  $x \rightarrow \infty$ , la funzione di ripartizione del massimo ha espressione

$$F_{X_N}(x) = e^{-N(1 - F_X(x))}$$

Conviene ora assumere che la grandezza  $N$  sia sostituita dall'espressione

$$N = e^{\frac{\psi_1}{\psi_2}}$$

Il termine  $\psi_2$  è noto e si inventa un termine  $\psi_1$  a far scomparire  $N$ .  
Si sostituisce e si ottiene

$$F_{X_N}(x) = e^{-e^{-\frac{x-\psi_1}{\psi_2}}}$$

Si nota che è comparso il termine di standardizzazione e dunque si introduce la variabile ridotta  $y$  della distribuzione  $X_N$ , che è detta variabile ridotta di Gumbel

$$y = \frac{x - \psi_1}{\psi_2}$$

Si ottiene così

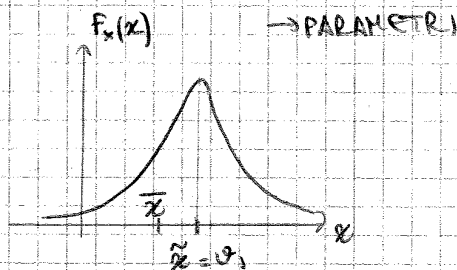
$$F_x = e^{-e^{-y}}$$

Questa funzione è così semplice che si può costruire l'inversa

$$y(F_y) = -\ln\left(\ln\left(\frac{1}{F_y}\right)\right) \quad \text{Distribuzione di Gumbel}$$

anche qui si può ricorrere alla carta probablistica per la combinazione di v.d. che porta a una retta

PROPRIETÀ → pur essendo asimmetrica, essa non è vincolata a non assumere valori negativi. Ciò interessa poco perché si sta valutando il massimo



$$\psi_1 = \tilde{x}$$

$$F(\psi_1) = 0,368$$

$$c_a \approx 1,16$$

→ se nella lognormale, l'asimmetria si poteva "stiracchiare" agendo sul coef. efficiente di variazione; ora è costante perché deriva dall'esponenziale (dove  $c_a \approx \text{cost}$ )

→ MOMENTI ORDINARI

$$E(x) = \mu = \psi_1 + 0,577\psi_2$$

$$\text{Var}(x) = \frac{\pi^2\psi_2^2}{6}$$

→ L-MOMENTI

$$L_1 = \psi_1 + 0,577\psi_2 \quad \sim L_1 = \mu$$

$$L_2 = \psi_2 \cdot \log 2$$

Si ha così un'alternativa semplice alle altre distribuzioni, che però appartiene a una nuova famiglia di distribuzioni di valore estremo.

*Handwritten notes:*  
 - distribuzione di Gumbel  
 - distribuzione di valore estremo  
 - distribuzione di Gumbel  
 - distribuzione di Gumbel

Per facilitare lo studio della distribuzione di Gumbel, si può rappresentare in forma di distribuzione dei quantili:

essa in pratica è la relazione inversa della distribuzione.

Si sa che

$$F(x) = e^{-e^{-\frac{x-\theta_1}{\theta_2}}}$$

La sua inversa è

$$x = -\theta_1 - \theta_2 \ln\left(\frac{1}{F}\right)$$

Si estende la ridotta

$$\frac{x-\theta_1}{\theta_2} = -\ln\left(\frac{1}{F}\right)$$

Si ottiene

$$x(F) = \theta_1 - \theta_2 \ln\left(\frac{1}{F}\right)$$

In genere, x comodità, si esprime la distribuzione dei quantili in funzione del periodo di ritorno

$$x_T = \theta_1 - \theta_2 \ln\left(\frac{T}{T-1}\right)$$

Quando si inseriscono  $\hat{\theta}_1$  e  $\hat{\theta}_2$  (le stime), si ottiene la stima  $\hat{x}_T$  del quantile x un assegnato periodo di ritorno T.

$$\hat{x}_T = \hat{\theta}_1 - \hat{\theta}_2 \ln\left(\frac{T}{T-1}\right)$$

A questo punto, si possono apportare delle semplificazioni, giocando sulle funzioni dei parametri

→ nel metodo dei momenti ordinati le funzioni erano

$$\theta_1 = \bar{x} - 0,45 S_x$$

$$\theta_2 = \frac{\sqrt{6} S_x}{\pi}$$

Dividendo entrambi i parametri x la media, si ottiene

$$\frac{\theta_1}{\bar{x}} = 1 - 0,45 C_v$$

$$\frac{\theta_2}{\bar{x}} = \frac{\sqrt{6}}{\pi} C_v$$

Può essere utile scrivere il quantile in una forma dove la media  $\bar{x}$  è a fattore

In particolare, con questa scrittura il quantile assume espressione

$$x_T = \bar{x} \left\{ 1 - C_v \left[ 0,45 + \frac{\sqrt{6}}{\pi} \ln\left(\frac{T}{T-1}\right) \right] \right\}$$



Si ottiene così, in scala logaritmica, una retta che curva leggermente x valori piccoli ( $x \ln T < 3$ )

Il vantaggio è che, qualunque distribuzione si abbia, la descrizione della funzione quantile può essere rappresentata su carta probabilistica.  
Se poi si mette  $\log T$ , si hanno le stesse proprietà indicate quando si usa la legge di Gumbel poiché c'è un unico andamento rettilineo.

Si introduce ora una distribuzione a 3 parametri.  
Le distribuzioni a 2 parametri sono comode perché danno sempre un'idea di adattabilità ai dati osservati ma spesso non sono sufficienti e dunque occorre passare a distribuzioni più complicate.

I modelli più complessi si basano sulla presenza di un parametro di più.  
Così, se nei modelli a 2 parametri si aveva

$\theta_1$  = parametro di posizione

$\theta_2$  = parametro di scala

se si poteva scrivere la rappresentazione

$$x_T = \theta_1 + \theta_2 u(F)$$

Il nuovo parametro  $\theta_3$  è il parametro di forma, cioè influenza la forma.  
Con questo parametro, si mantengono posizione e scala ma si varia la forma.

Attraverso questo parametro, si introduce la funzione generalizzata del valore estremo (GEV).

$$F(x) = F(x; \theta_1; \theta_2; \theta_3)$$

→ assegnando un valore, si ottiene una certa funzione

Essa ha forma

$$F(x) = e^{-\left[1 - \theta_3 \frac{x - \theta_1}{\theta_2}\right]^{1/\theta_3}}$$

CASI →  $\theta_3 \rightarrow 0$  si usa la funzione

$$e^{-\frac{x - \theta_1}{\theta_2}}$$

$$F(x) = e^{-\frac{x - \theta_1}{\theta_2}}$$

→ funzione di Gumbel

→  $\theta_3 < 0$  si ottiene una distribuzione EU2

→  $\theta_3 > 0$  si ottiene una distribuzione EU3

Così la GEV consente di muoversi con regolarità tra i 3 casi in funzione del valore  $\theta_3$ .

## Verifica quantitativa dell'ipotesi:

si parla in questo caso di TEST STATISTICI, che consente di capire la correttezza di un'ipotesi.

Per orientarsi nei test, innanzitutto si identificano le ipotesi

$H_0$  = ipotesi sottoposta a verifica

$H_1$  = ipotesi alternativa

Dunque si sottopongono delle ipotesi a test e queste ipotesi  $H_0$  si formulano come test di adattamento

$H_0$  = il campione osservato può essere considerato come estratto dalla popolazione, che è di fatto una funzione  $F(x)$  di ripartizione

Questo fatto si può testare in più modi e i modi definiscono degli INDICATORI che fanno capire se un aspetto del campione è assimilabile a un aspetto della popolazione.

Per testare  $H_0$ , ci sono 4 modalità

I RIFIUTARE  $H_0$  SE VERA

II ACCETTARE  $H_0$  SE FALSA

III ACCETTARE  $H_0$  SE VERA

IV RIFIUTARE  $H_0$  SE FALSA

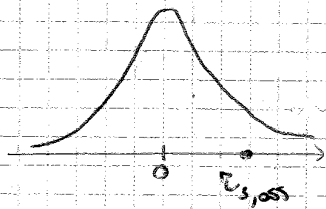
Assegnare una probabilità a questi casi significa creare le condizioni di applicazione del test.

Al caso, si parla di livello di significatività  $\alpha$  (probabilità di rifiutare  $H_0$  se vera) e rappresenta il rischio che si vuole correre nel rigettare un'ipotesi tenendo si in sicurezza.

Se poi  $\beta$  è la probabilità di accettare  $H_0$  se falsa, allora  $1-\beta$  si dice potenza del test e corrisponde alla probabilità di rigettare un'ipotesi  $H_0$  falsa. E cioè la capacità del test di essere selettivo.

In questo ambito, interessa il valore  $\alpha$  perché implica una certa decisione presa d'accordo con il committente (si sa quanto è stretto)

Si nota che la deviazione standard relativa all'indicatore diventa sempre più piccola all'aumentare della numerosità delle osservazioni nel campione.



Ora si disegna la distribuzione dell'indicatore, ponendo in ascissa l'insieme dei coefficienti di asimmetria calcolati da un campione di numerosità  $n$ , estratto dalla pdf normale.

Non si sa se si può considerare il campione estratto dalla normale.

Per capire la reale distanza del valore dallo zero, si disegna la pdf dello  $Z_3$ .

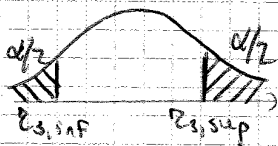
Dal disegno, si nota che si è lontani dalla zero.

III) Si seleziona l'intervallo di accettazione, scegliendo un valore  $\alpha$  arbitrario:

l'intervallo di accettazione è costituito dai valori all'esterno in cui si trova il punto (si rigetta l'ipotesi vera). Dunque come estremi ci sono

$$Z_{3,inf} = Z_3 \left( F = \frac{\alpha}{2} \right)$$

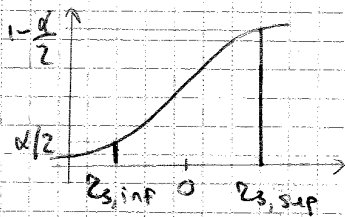
$$Z_{3,sup} = Z_3 \left( F = 1 - \frac{\alpha}{2} \right)$$



Infatti, la probabilità che l'indicatore cada nell'intervallo di accettazione coincide con la probabilità di appartenenza all'intervallo sia maggiore o minore di  $\alpha/2$  o minore di  $1 - \alpha/2$ . In altri termini, si ha

$$P(Z_{3,camp} \in [Z_{3,inf}; Z_{3,sup}]) > 1 - \alpha$$

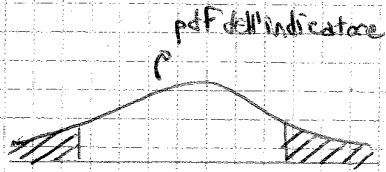
$$F(Z_{3,sup}) - F(Z_{3,inf})$$



Così, anziché confrontare i valori osservati con tutti i dati (che non si hanno), si fornisce la distribuzione e si valuta se il valore è marginale o dentro la distribuzione. In pratica, si specifica la coda definendo il fattore  $\alpha$  (che rappresenta la probabilità di sbagliare) e che il valore si trovi nella coda della distribuzione e si guarda dove cade l'indicatore.

⇒ si stabilisce se il campione appartiene alla distribuzione (è complicato perché non si può dire con certezza) a base di un intervallo definito da  $\alpha$

Si è visto che, fissato il livello di significatività  $\alpha$ , cioè il rischio di incorrere nel rifiuto di un'ipotesi quando questa è corretta, si può determinare l'intervallo di accettazione.

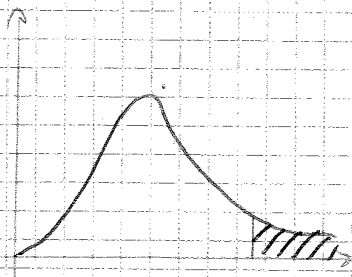


A volte, l'intervallo è simmetrico rispetto a un valore medio e dunque l'indicatore si può discostare dal valore teorico sia in senso positivo che negativo, con uguale probabilità. In questo caso, si parla di test a 2 code:

la zona di non accettazione è indifferentemente molto lontana a sinistra e a destra.

Dunque l'intervallo si costruisce trovando un valore limite a destra e a sinistra e la probabilità di errore è equamente distribuita a destra e a sinistra,  $\alpha$  cui l'area in ciascuna zona di non accettazione vale

$$\frac{\alpha}{2} \rightarrow \text{metà errore}$$

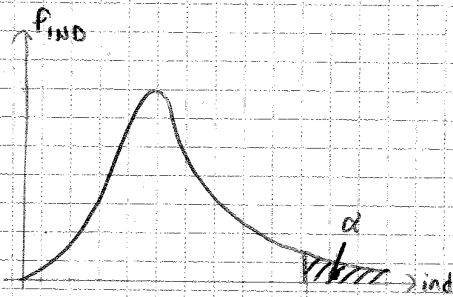


Esistono anche test a 1 coda, dove ad es. non ha senso cercare errori in senso negativo ( $x$  variabili definite positive). L'area di non accettazione si trova tutta nella coda destra e l'area vale

$\alpha$

Si è anche notato che spesso si vuole valutare QUANTO LARGAMENTE UN TEST È SUPERATO.

Si può allora calcolare  $\alpha_{max}$ , cioè il livello di significatività limite perché il test sia superato con un dato indicatore osservato.



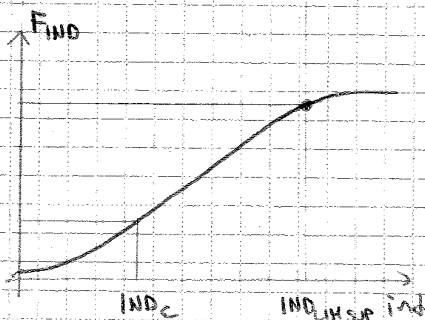
Per fare ciò, basta osservare a cosa corrisponde la posizione limite nella probabilità cumulata dell'indicatore teorico: se la regione di esclusione ha area  $\alpha$ , il valore corrispondente al limite superiore dell'indicatore nella curva cumulata dà proprio l'area, cioè

$$F(\text{IND}_{\text{LIM SUP}}) = 1 - \alpha$$

Allora è facile trovare  $\alpha_{max}$  perché, se  $\text{IND}_c$  ha una certa posizione nella regione di accettazione, si può trovare la probabilità cumulata e dire che

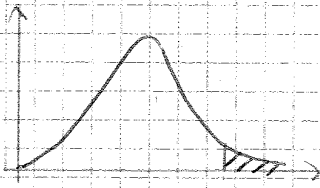
$$\alpha_{max} = 1 - F(\text{IND}_c) \rightarrow \text{vale solo x test a 1 coda}$$

Si nota che  $\alpha_{max} \gg \alpha$  se il test è ben riuscito, cioè si sarebbero potuto superare test molto più selettivi e con ipotesi molto forti perché



III) Il test è immediato perché il calcolo dell'intervallo è semplice:

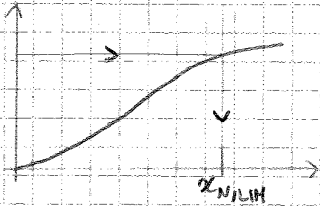
l'intervallo di accettabilità deriva dalla scelta del livello di significatività e dal procedere a un test a 1 o 2 code (a seconda delle condizioni in cui ci si trova)



In genere, Gumbel è inefficace x la stima dei massimi x cui si cerca di sopravvalutare la rarità di un evento osservato.

Ad es, se si distrugge una fabbrica, un perito può valutare l'evento con legge di Gumbel e stimare il periodo di ritorno a 1500 anni.

Per andare sul sicuro, si fa un altro test che dimostra che Gumbel non è adeguata e, con una distribuzione diversa, si ottengono 120 anni.



⇒ in generale, preoccupa quando una distribuzione fornisce un tempo di ritorno troppo alto e dunque conviene scegliere TEST A 1 CODA, in modo che si abbia una regione di accettabilità più stretta (l'area è tutta concentrata a destra).

Si fa dunque il test a 1 coda sulla Gumbel, entrando nella funzione cumulativa con  $F = 1 - \alpha$  ( $\alpha = 0,05$ ) e trovando il valore (quantile)  $x_{N,lim}$

$$x_{N,lim} = \theta_1 + \theta_2 \ln \ln \left( \frac{1}{0,95} \right)$$

Se, data la distribuzione, si ottiene un valore massimo estremo appartenente alla distribuzione teorica, il test è positivo.

OSSERVAZIONE: se la distribuzione non è di Gumbel, si gioca sul valore limite ricavato da  $F = 1 - \alpha$ . Infatti, si sa che

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

Nelle distribuzioni a una coda, il valore limite superiore ha probabilità cumulata  $1 - \alpha$ .

$$F_{X_N}(x_{N,lim}) = 1 - \alpha$$

Il test è positivo se si trova un indicatore (un massimo)  $x_{N,oss}$  inferiore a  $x_{N,lim}$ , cioè se

$$F_{X_N}(x_{N,oss}) < 1 - \alpha$$

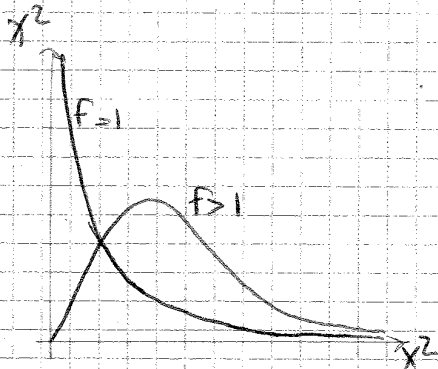
Ciò equivale però a scrivere

$$[F_X(x_{N,oss})]^N < 1 - \alpha$$

Dunque, in questo test del massimo valore, non bisogna determinare la funzione analitica ma si ha

$$F_X(x_{N,oss}) < (1 - \alpha)^{1/N} \Rightarrow \text{test passato}$$

Per quanto riguarda la distribuzione, se ne è ottenuta una che approssima la normale  $\times k \rightarrow \infty$  ed è la distribuzione  $\chi^2$ .



Essa corrisponde a una somma finita di  $n$  variabili casuali normali ed è definita positiva. Essa dipende da un parametro  $F$ .

$$F = k - s - 1 \quad s = \text{numero di parametri della distribuzione iniziale (} s=2 \text{ se a 2 parametri)}$$

Se  $F=1$ , si ottiene una distribuzione esponenziale. Se  $F>1$ , si ottiene una distribuzione simile a quella di un'onda di piena. Infatti, nella pdf è compresa una funzione speciale, detta funzione gamma.

III) Per valutare i limiti, bisogna ricorrere a una tabella di valori.

Per vedere se fare il test a 1 o 2 code, si nota che  $\chi^2$  è una distanza quadratica del numero di osservazioni in una classe rispetto al numero teorico  $Np_i$  e dunque

$$\chi^2 > 0$$

La gaussiana però è una distribuzione definita positiva

$\Rightarrow$  il test è a 1 coda e così si può sfruttare che

$$\chi_{LIM} = F^{-1}(1 - \alpha)$$

La funzione  $F$  dipende dal numero di osservazioni (da cui si ricava  $k$ ). Nel caso di S. Martino Chi Sone, si ha

$$N = 33 \Rightarrow k = 8$$

Come pdf di partenza, si usa la Gumbel a 2 parametri, a cui

$$F = k - s - 1 = 8 - 2 - 1 = 5$$

I quantili di interesse  $\chi_{LIM}$  sono quelli corrispondenti a diversi valori di  $\alpha$ :

In questo caso, l'ipotesi  $H_0$  è forte se la distanza è piccola, cioè ci si aspetta che la distribuzione fornisca un valore limite via via più piccolo man mano che  $\alpha$  cresce (intervallo di accettazione più stretto).

Un intervallo di accettazione buono si ottiene con  $\alpha=1$  (e  $F=0,9$ ), con

$$\chi_{LIM} = 9,246$$

Se invece si è molto efficienti nel trovare la distribuzione, si può scegliere  $\alpha=0,25$ , che dà

$$\chi_{LIM} = 6,63$$

→ test di Cramer-Von Mises;

esso si basa sulla distanza quadratica tra distribuzione teorica e frequenza cumulata osservata.

$$Q^2 = N \int [\hat{F}(x) - F_x(x)]^2 \cdot f_x(x) dx \quad \rightarrow \text{di fatto, diventa una somma}$$

La presenza della pdf  $f_x(x)$  fa sì che l'integrale sia come  $f_x(x)dx$  e cioè un numero puro, perché è 1.

$$[F] = \frac{1}{[x]}$$

Altrimenti, esso dipenderebbe dalle dimensioni della grandezza.

Questo test è più credibile come test di scostamento ( $Q^2 \Rightarrow$  curve più vicine).  
D'altra parte, non interessa tanto che la pdf si adatti al centro perché interessano i valori estremi, e che dunque ci sia l'adattamento agli estremi.

⇒ si apporta una correzione, passando al test di Anderson:

x valutare l'adattamento alla coda rispetto al centro, si usa una funzione peso che dia maggior peso alle differenze agli estremi che al centro.

$$A^2 = N \int \frac{[\hat{F}(x) - F_x(x)]^2}{F_x(x)[1 - F_x(x)]} \cdot f_x(x) dx$$

Con il denominatore, si incrementa il valore della differenza x valori  $F_x$  piccoli e grandi, cioè agli estremi.

Il problema è che non si può imporre qudi code pesare, ma si è visto che il test è comunque affidabile x verificare un'ipotesi.

Con questa formula, si trova il valore limite di  $\bar{x}$  in funzione del numero di osservazioni.  
 Infatti, si può ricavare  $\mu_x$

$$\mu_x = - \frac{t \cdot S_{N-1, x}}{\sqrt{N}} + \bar{x}$$

e si nota che corrisponde a  $\bar{x}$  più una quantità con valore medio nullo. Così, in funzione del livello di significatività, cioè l'ampiezza data all'intervallo di confidenza (ampiezza data all'intervallo di confidenza, si ottiene

$$\mu_{sup} = \bar{x} - \frac{t_{\alpha=0,025}}{\sqrt{N}} \cdot S_{N-1, x}$$

$$\mu_{inf} = \bar{x} - \frac{t_{\alpha=0,975}}{\sqrt{N}} \cdot S_{N-1, x}$$

→  $\alpha = 0,025$   
 è  $\mu_{sup}$ , nonostante il segno "-", perché  $t_{\alpha=0,025} < 0$

Si è scelta quel valore di  $\alpha$  perché l'intervallo di confidenza deve comprendere una variazione della media entro il 95% del suo valore (altrimenti, si va fuori dall'intervallo di naturale oscillazione)

→ distribuzione della varianza

Facendo la stessa operazione, si nota che la varianza è una variabile casuale distribuita secondo la distribuzione  $\chi^2$ .  
 Da qui, si trovano i limiti di variabilità di  $\hat{\sigma}$  entro il 95% del suo valore

$$[\hat{\sigma}_{inf, \alpha}; \hat{\sigma}_{sup, \alpha}]$$

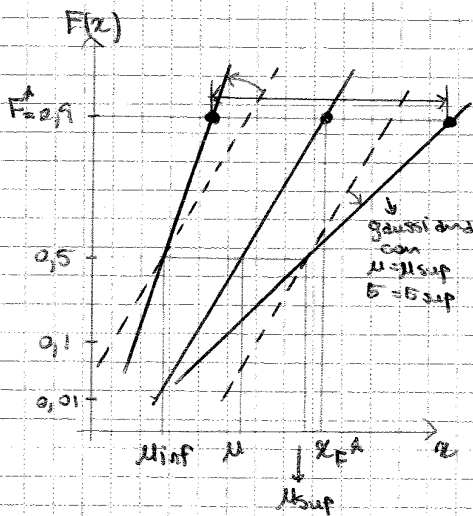
In base a questo, come varia il quantile  $\chi_F$ ?

Si sa che

$$\chi_F = \mu + u_F \hat{\sigma}$$

Applicando l'oscillazione dei parametri a  $\chi_F$ , si ottiene innanzitutto che il limite inferiore è somma dell'oscillazione minima della media e massima oscillazione della  $\sigma$ .

$$\chi_{F, inf} = \hat{\mu}_{inf} - \hat{\sigma}_{sup} u_F$$



Per capire ciò, si considerino nella carta probabilistica la retta teorica e il quantile riferito alla probabilità di riferimento  $F^A$ .  
 Il quantile  $\chi_F^A = \mu + \hat{\sigma}_{sup} u_F$  corrisponde alla posizione media del vero quantile, che oscilla poiché  $\mu$  e  $\hat{\sigma}$  sono calcolati con un errore.  
 La variabilità della media fa  
 La variabilità della media fa sì che la curva teorica trasli a sinistra e a destra rispetto a quella centrale e l'intervallo spazzato a  $F = 0,5$  definisce l'intervallo di confidenza della media.  
 Così, se  $\mu$  varia e  $\hat{\sigma}$  no, l'intervallo di confidenza è definito come

$$[\mu_{inf}; \mu_{sup}]$$



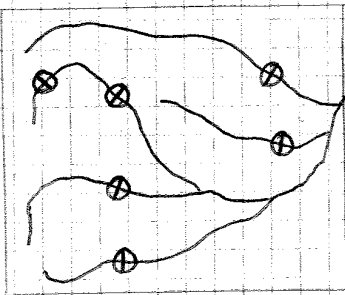
## Metodi diretti senza osservazioni

1) Finora si sono visti i metodi diretti di stima della variabile di progetto con metodi probabilistici a partire da una serie di dati osservati.

2) Si sa che nei metodi diretti la variabile usata è sempre la stessa (portata di colmo  $x$  e un assegnato periodo di ritorno).  
Ora però non ci sono osservazioni e non si può applicare l'inferenza statistica.  
In particolare, si vuole prendere in prestito i dati da un'altra parte e  $x$  elaborare la distribuzione.

↳ questo è il caso più frequente

Questi metodi sono detti metodi diretti regionali poiché, siccome non ci sono dati nella specifica sezione di interesse, si cercano i dati in sezioni vicine e ci si muove in una regione che è maggiore del bacino d'interesse.



Infatti, dato un corso d'acqua con il suo bacino, se non ci sono dati, si allarga la visione a un'area estesa e si definisce una regione idrografica con più aste. In essa, si nota che esistono sezioni con dati in più punti.

⇒ la logica dell'analisi regionale è l'uso del trasferimento dell'informazione idrologica:

in presenza di informazioni in altre zone, si assume che il comportamento del fenomeno sia analogo al punto considerato.

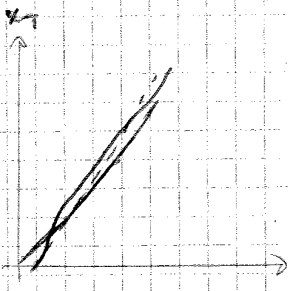
Ciò significa ricercare una serie di casi dove si può trovare il comune denominatore del comportamento e ciò corrisponde a trovare una regione idrologicamente omogenea:

« omogenea » si intende dire che c'è qualcosa all'interno della regione e a che avviene sempre allo stesso modo.

Nel caso di interesse, si può ipotizzare la funzione adimensionale di probabilità di colmo di piena costante

$$\frac{x_T}{x} = k(T)$$

$k$  = funzione di crescita



Essendo adimensionali, si possono sovrapporre le varie funzioni e vedere la parentela tra esse. Se c'è omogeneità, si prende la curva media e la si attribuisce alle sezioni che non hanno dati.

↓

funzioni adimensionali perché  
bacino  $\Rightarrow$  piena  $\Rightarrow$

Si dimostra che questa operazione funziona, cioè esistono test che dimostrano che  $n$  stazioni appartengono a un'unica regione omogenea.

## Metodi indiretti: metodo di trasformazione afflussi/deflussi

1 Tale metodo prevede l'uso di ENTRATE ESTERNE alla portata x determinarne il valore e, in genere, queste sono delle forzanti che danno luogo al superamento della portata.

2 Afflusso = afflusso meteorico

Deflusso = deflusso di piena

3 Metodo di trasformazione afflussi/deflussi:

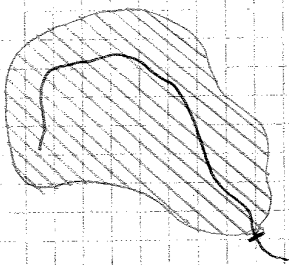
in esso, il concetto fondamentale è la trasformazione della pioggia di progetto in portata di progetto, poiché la portata di progetto è prodotta dalla pioggia di progetto.

La pioggia di progetto è però qualcosa di complesso perché è un evento di progetto e non può essere ridotta a una sola variabile di progetto.

In una piena, infatti, si modella con un'ondata di piena della piena e così si identifica con un unico valore istantaneo, x cui il dato è univocamente determinato (il dato in tutto l'anno).

Invece, con una pioggia, bisogna ricostruire l'ondata di piena e c'è bisogno dell'intero evento e non del singolo valore. Si parla così di evento pluviometrico.

Un'altra motivo x cui l'analisi è più complessa è la scala spaziale:



ad es. a San Martino Chisone si sa che la portata valutata arriva da un bacino.

Se lì si trova una stazione di misura, si raccolgono in un unico punto le informazioni relative all'acqua che arriva da tutto il territorio, cioè l'acqua rappresenta il contributo di tutto il territorio.

Se ciò non accade, quando si ricostruisce il fenomeno dalle precipitazioni, non si possono più trascurare le caratteristiche di tutto il territorio, perché le piogge cadono su tutta l'area. Così, se si misura in un punto, si ottiene l'informazione relativa a quel punto ma la misura è poco rappresentativa rispetto al territorio.

⇒ quando si usa il metodo indiretto, bisogna Ricostruire LA PRECIPITAZIONE SULL'INTERA AREA e bisogna usare più stazioni.

Dunque bisogna tenere conto di

→ misura delle precipitazioni

→ distribuzione nello spazio (come interpolare le stazioni)

→ DISPONIBILITÀ DI DATI STORICI

In tal modo, l'informazione non è più una strisciata pluviometrica ma è un'informazione simile al diagramma a punti, dove ci sono gli scatti sull'asse del tempo  $t$ .

↓ se c'è precipitazione, ci sono più scatti

Ogni scatto corrisponde a un volume

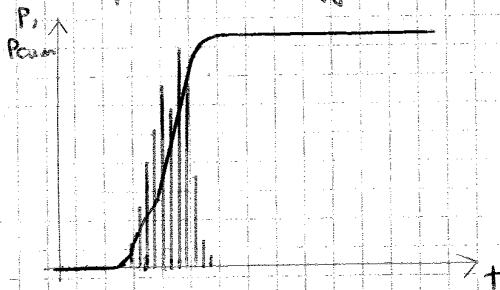
$$h = 0,2 \text{ mm}$$

e così, contando il numero di scatti in un intervallo, si ottiene il volume corrispondente.



⇒ l'informazione è un lungo vettore di orari in cui, preso un periodo, si conta il numero di scatti.

Da questo conteggio, si ottiene il diagramma pluviometrico (o PLUVIOGRAMMA):



× costruirla, si fissa la RISOLUZIONE TEMPORALE, cioè l'intervallo entro cui si somma il volume di pioggia (es. 10 min.). Da qui si definiscono delle colonne che danno il volume misurato in quell'intervallo.

In genere poi si sommano cumulativamente queste grandezze (si sommano i valori precedenti), e così si valuta il totale di precipitazione.

In genere, nell'evento osservato si distinguono 2 fasi

→ all'inizio piove pochi millimetri

→ poi piove una grande quantità

Per tenere conto di questo, si può definire l'intensità di precipitazione, come variazione della precipitazione nel tempo

$$I = \frac{\Delta P}{\Delta t} \quad [I] = \text{mmh}^{-1}$$

Essa è la pendenza della precipitazione cumulata calcolata nell'intervallo di interesse.

Si nota che l'intensità varia nel tempo e dunque si può valutare

→ un'INTENSITÀ ISTANTANEA (ad es. se c'è un'alta risoluzione)

→ un'INTENSITÀ MEDIA

Esse sono diverse perché il valore istantaneo è maggiore o minore del valore medio.

L'intensità è importante nella valutazione dell'evento di progetto

## 5 Distribuzione spaziale delle precipitazioni:

bisogna capire come ripartire le informazioni dei singoli strumenti nello spazio.

Qual è però la risoluzione temporale della misura di uno strumento pluviografico?  
Cioè, qual è il dettaglio minimo di tempo entro cui si può rappresentare la misura?

Ad es un termometro dà la misura di temperatura in ogni istante e la risoluzione temporale è di frazioni di secondo (si può dire che la misura è continua).

Nel pluviometro, invece, non si può fare una misura continua perché, anche se esiste un'intensità istantanea, bisogna aspettare un tempo minimo  $\Delta t$  di risoluzione.

Infatti, la risoluzione dipende dal tempo minimo di riempimento della vaschetta (anche se bastano 0,2 mm) e dunque non c'è continuità.

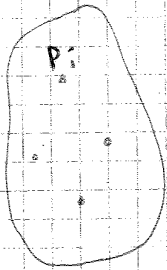
Così, in sede di rappresentazione del pluviogramma, la scelta dell'intervallo minimo  $\Delta t$  e l'istogramma è sostanzialmente arbitraria perché lo strumento consente di scegliere valori piccoli, ma spesso inutili e le incertezze legate alla qualità dell'informazione iniziale.

In anzitutto, a distinguere dalla misura puntuale, si parla di afflusso meteorico es:

esso si definisce quando si considera una precipitazione su un'area e rappresenta l'ALTEZZA DI PRECIPITAZIONE EQUIVALENTE, come se tutta l'area di interesse fosse uniformemente ricoperta.

Poiché si hanno sempre misure in punti distinti, si usano metodi di interpolazione spaziale sull'area di interesse (area del bacino idrografico)

→ media aritmetica (metodo più semplice)



Noti l'area di interesse e i punti di misura, si vuole definire un valore uniforme.

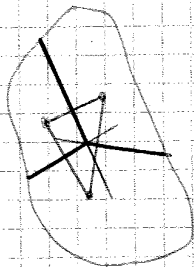
Si può allora assumere che non ci sia motivo  $\chi$  cui nei vari punti la precipitazione sia sempre diversa dagli altri e ogni punto rappresenti  $1/N$  della quota dell'area.

si dà peso uniforme al valore di ogni dato.

L'ipotesi è attendibile se le stazioni sono uniformemente distribuite e, in tal caso, si può fare la media

$$\bar{P} = \frac{1}{N} \sum_{i=1}^N P_i$$

→ metodo dei poligoni di Thiessen (o dei TORRETTI = poligoni disegnati attorno a ogni stazione):



Per valutare l'area di rappresentatività, si fa una costruzione geometrica:

si uniscono i punti, si tracciano le perpendicolari e queste, con il perimetro di chiusura, creano i poligoni. Da qui si assume che ogni stazione è rappresentativa della porzione di area in cui è contenuta.

Di conseguenza, il valore osservato in una stazione rispetto al totale ha peso  $A_i/A$  e dunque non c'è più distribuzione uniforme ma è pesata sull'area.

In tal modo, se i punti sono vicini, questi restituiscono la stessa informazione e hanno poco peso.

Questo è calcolato x ogni cella e, poiché le celle hanno ugual area e i valori sono uniformemente distribuiti su ogni area e hanno lo stesso peso, il valore areale di afflusso è la media aritmetica.

$$\bar{P} = \frac{1}{N} \sum_{i=1}^N \hat{P}_i$$

Questo è l'evento più semplice che si possa immaginare come pioggia di progetto.  
 Questa rappresentazione è utile a definire la variabile casuale da elaborare:

infatti, in un evento di piena si sa definire il valore  $q_i$  da elaborare

$$q_i = \text{portata massima di piena nell'anno } i$$

Nelle piogge, definire la variabile è più complessa, anche perché ora si ragiona sull'evento di progetto (e non sulla variabile di progetto).

In virtù della semplificazione nell'evento rettangolare, però, si nota che ora c'è una variabile casuale (l'altezza) che è parametrica nella durata  $d$ .

⇒ prendendo tutti eventi di durata  $d$ , si definisce la variabile casuale come

$$h_d = \text{altezza massima annua per gli eventi di durata } d$$

↓  
 valore massimo rispetto a tutti gli eventi di quella durata.

Esistono però tante durate utili (30', 1 h, 12 h, etc.) e quindi ci sono tante variabili  $x$  le possibili durate.

Si definisce allora come grandezza di progetto la quantità:

$$h_{d,T} = \text{altezza massima annua per gli eventi di durata } d \text{ e tempo di ritorno } T$$

↓  
 si legano durata e tempo di ritorno all'altezza

Così la relazione sembra più complessa, ma in realtà la variazione con la durata è regolare e allora si può dire che la variabile casuale  $h_{d,T}$  è una variabile parametrica (non univoca ma legata a  $d$ ).

In particolare, poi, il metodo indice aiuta a scomporre il problema in media e variazione (funzione di crescita), separando così i contributi di durata e tempo di ritorno.

$$h_{d,T} = \bar{h}(d) \cdot k_d(T)$$

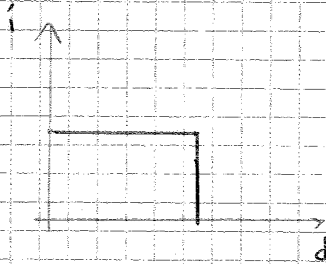
Questa si dice rappresentazione delle curve intensità - durata - frequenza (lega queste quantità).

Questa è la rappresentazione usata a descrivere la media di precipitazione in funzione della durata e si dice relazione altezza-durata per la media delle precipitazioni.

Da qui si può trovare l'evento di precipitazione medio rettangolare;

nella curva di prima, nota la durata  $d$ , si ricava  $\bar{h}(d)$ .  
Da qui si ricava

$$T(d) = \frac{\bar{h}(d)}{d}$$



In tal modo si è ottenuto un sistema che permette di valutare le caratteristiche medie (non gli estremi) e si ha

$$T(d) = \alpha d^{n-1} \quad \text{Relazione intensità media-durata}$$

↓  
consente di valutare l'evento di precipitazione media

### II) RICERCA DELLA FUNZIONE DI CRESCITA

Indovinito, si assume che per qualunque durata le precipitazioni siano descritte dalla distribuzione di Gumbel (l'unica in cui si ha la rappresentazione analitica di  $k(T)$ ).

Per fissare le idee, si pone

$$d = 1 \text{ h}$$

in modo da lavorare su un'unica variabile casuale  $(h_{1,T})$ .  
Si cerca una parametrizzazione a la quale vale la relazione

$$h_{1,T} = \bar{h}_1 - k_1(T)$$

Però si sa a ipotesi che

$$h_{1,T} \sim \text{Gumbel}(\theta_1; \theta_2)$$

Ora, si ricorda che il quantile si esprime come

$$x(T) = \theta_1 - \theta_2 \ln \left( \frac{T}{T-1} \right)$$

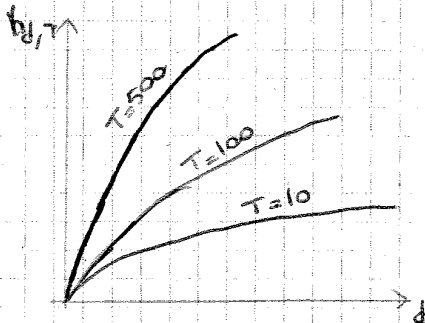
e, in esso, i momenti valgono

$$\theta_1 = \bar{x} - 0,45 S_x$$

$$\theta_2 = \frac{\sqrt{6}}{\pi} S_x$$

### III COSTRUZIONE DELLE IDF

Svolte tutte queste operazioni, si può costruire l'intera rappresentazione della probabile altezza di pioggia a diverse durate e diversi tempi di ritorno e la relazione è rappresentata dalle curve IDF, che sono curve di possibilità pluviometrica.



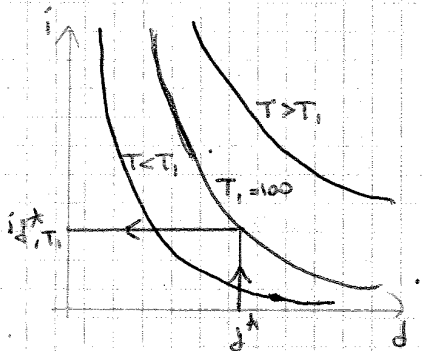
Per ottenere questa famiglia di curve, si fissa un certo tempo di ritorno, ad es.

$$T=10$$

Si ottiene allora un certo valore  $k(T)$ .  
Moltipliandolo  $\times ad^m$ , si ottiene una certa curva nel diagramma durata-altezza.  
Modificando  $T$ , si possono disegnare diverse curve.

Questo rappresenta lo strumento operativo  $\times$  creare le piogge di progetto, e che dunque permette di ottenere l'altezza di precipitazione  $\times$  ogni tempo di ritorno e ogni durata.

In particolare, è immediato determinare la pioggia di progetto rettangolare:



fissato un periodo di ritorno ad es  $T=100$ , si vuole determinare la pioggia  $\times$  una durata  $d$ .  
Si sa allora che  $\times$  avere una pioggia rettangolare, l'intensità è costante e vale

$$i_{d,T=100} = \frac{h_{d,T=100}}{d}$$

Per questo valore di  $T$ , la curva intensità-durata ha espressione

$$i_{d,T=100} = ad^{m-1} k(T)$$

Si nota che, nel diagramma intensità-durata

- $\rightarrow$  la curva decresce perché il valore  $n$  stimato è sempre minore di 1 e dunque l'intensità media decresce con la durata (più si aspetta, più si notano intensità minori)
- $\rightarrow$   $\times d \rightarrow 0$ , l'intensità va a  $\infty$  e così questa curva è poco realistica  $\times$  durate piccole. Questo problema non riguarda l'altezza di precipitazione perché,  $\times d \rightarrow 0$ ,  $h_{d,T} \rightarrow 0$

Allora da qui, inoltre, scelta una durata  $d^*$ , il pluviogramma di progetto diventa un rettangolo di base  $d^*$  e altezza  $i_{d^*,100}$  (quindi è ricavabile direttamente dall'IDF - vantaggio)

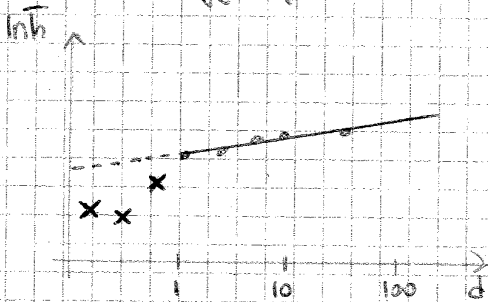
- $\Rightarrow$  la curva IDF dà il massimo di precipitazione prevedibile in  $T$  anni e con durata  $d^*$ .  
La curva intensità-durata invece fornisce l'intensità media in quella durata (cioè la pendenza della retta cumulata nel pluviogramma)



→ precipitazioni con  $d < 1h$ . Di queste, ci sono solo indicazioni perché è difficile trovare i massimi di precipitazione e, un tempo, questi erano dettati dai salti del pennino. A volte, a l'incertezza, ci sono 2 dati a la stessa stazione (e si sceglie il massimo / valore medio).

↓ da qui comunque si hanno indicazioni su come varia l'intensità con la durata, anche se i punti non sono certi dati

Se si aggiungono i dati  $x d \leq 1h$ , cosa succede?



Se si riportano i valori medi di altezza  $x d > 1h$  nel diagramma logaritmico, ci si aspetta un andamento lineare. Si prolunga poi idealmente a sinistra l'andamento.

Riportando gli altri punti, si nota che questi però si dispergono secondo un andamento più pendente e non regolare.

Visto che allora in molte applicazioni idrauliche bisogna riferirsi alle brevi durate ma ci sono pochi dati e questi non sono regolari, occorre MODIFICARE L'ANDAMENTO DELLA CURVA.

A seconda della disponibilità dei dati, la modifica può avvenire in 2 modi.

① definizione di una nuova funzione parametrica  $h(d)$ :

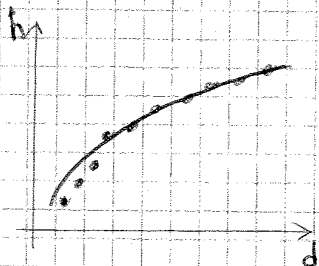
Si cambia completamente la legge della media, in modo che sia compatibile con l'intensità reale,  $x d \rightarrow 0$ . Poiché la funzione a due parametri non è esaustiva, quella nuova è a 3 parametri.

$$i_d = \frac{i_0}{(1 + B d)^A} \quad \rightarrow \text{funzione iperbolica a 3 parametri}$$

Si nota che,  $x d \rightarrow 0$ ,

$$i_d \rightarrow i_0$$

Dunque  $i_0$  è il valore tendenziale di intensità  $x d \rightarrow 0$ .

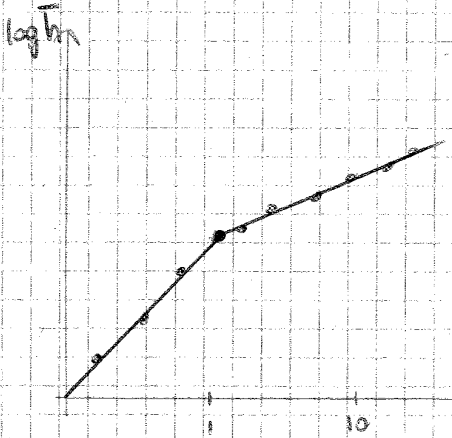


Con questa funzione, si ottiene un andamento abbastanza compatibile con la curva IDF  $x$  durate brevi, cui non si perdono i suoi vantaggi.

Poi si scosta dalla linearità  $x d \leq 1h$ .

Si nota poi che i valori medi sotto 1h non sono troppo vicini alla curva, a causa della qualità del dato e la complessità del fenomeno da esaminare.

II Spezzata di relazioni  $ad^n$



Si è visto che, nella relazione  $h-d$ , si può mantenere quella classica  $x > 1 h$ .  
 Dopo di che, inserite le medie  $x < 1 h$ , si può cercare di ADATTARE LA RETTA a questi valori.  
 In tal modo, si ottiene una migliore precisione, anche se non è idoneo a valori molto piccoli.

La nuova relazione, che è lineare in scala logaritmica, è

$$\bar{h}_d = ad^{n'}$$

Si nota che il parametro  $a$  è lo STESSO perché il valore medio  $x < 1 h$  è lo stesso  $x$  tutte e due le curve (x continuità)

$$a = \bar{h}_1$$

Pertanto, nel definire la spezzata, bisogna imporre il valore  $a$  dal valore  $\bar{h}_1$ , e bisogna stimare solo la pendenza

↓ l'intercetta è bloccata e si parla di stima vincolata

In assenza di osservazione, si suggerisce in genere di adottare

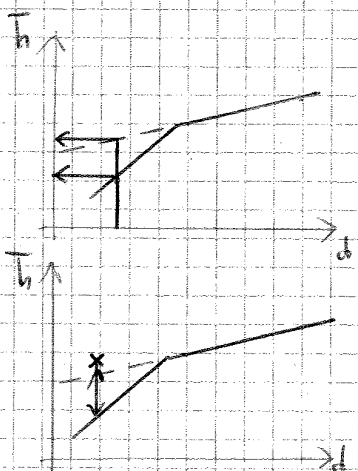
$$n' = \frac{4}{3} n$$

ERRORI LEGATI ALLE BREVI DURATE

Cosa succede se in progetto e verifica si trascura questa variazione?

L'entità dell'errore dipende da quanto ci si allontana da  $d = 1 h$ .  
 Se la durata è pochi minuti, l'errore è alta; se la durata è poco meno di  $1 h$  ( $45 \div 50'$ ), si può prolungare la relazione.

Che tipo di errore ci si aspetta?



PROGETTO: in progetto si entra con la durata, si legge il valore medio sulla curva e poi si moltiplica x il fattore di crescita.  
 Se si si basa sul prolungamento della curva, si sovrastima l'altezza

⇒ A FAVORE DI SICUREZZA

VERIFICA: partendo dal valore osservato, sul prolungamento si legge un valore più grande del valor medio reale e si ottiene così un  $K_{ass}$  maggiore di 1.  
 Dunque lavorare con la spezzata porta a dire che le osservazioni hanno  $K_{ass} > 1$  e dunque che gli eventi osservati hanno un  $T$  maggiore rispetto a quello ottenuto con il prolungamento della retta.  
 Dunque si assegna maggiore rarità, ma è noto che dire che un evento è più frequente è cautelativo, perché spinge x interventi.

⇒ A FAVORE DI SICUREZZA

## Problemi legati alle precipitazioni

**PREMESSA**: tipicamente, in presenza di un evento eccezionale, bisogna classificarlo in base alla pdf relativa a esso.  
In genere, non ci sono stazioni di rilevazione e si usano metodi indiretti x fare l'analisi:

si considera un certo numero di stazioni, valutandone l'attendibilità (si confronta la misura del bacino con la misura della zona = x piccoli bacini, il dato di una stazione pluviometrica è già rappresentativo).

Poiché si è in un metodo indiretto, bisogna studiare l'intero evento (non solo una variabile).

In presenza di un evento e dell'uso di un metodo indiretto, si possono avere due problemi

→ problema di progetto:

nei metodi indiretti, inanzitutto, si stabilisce che

$$T_{piena} = T_{precipitazione associata} \rightarrow \text{ISOFREQUENZA}$$

Portanto, il coefficiente di crescita  $K_T$  è assunto valido anche x la piena (è comune a pioggia e piena).

Così, nota la pioggia è stabilito a monte il tempo di ritorno  $T$  (poiché è un requisito progettuale legato all'importanza del progetto), si trova la portata di progetto.

$$h_{d,T} \xrightarrow{K_T} q \quad \rightarrow \text{in generale, il passaggio è } T \rightarrow x(T)$$

→ problema di verifica:

in questo caso, si stabilisce il tempo di ritorno di una precipitazione  $h$  osservata (l'evento è stato osservato)

$$h \rightarrow T(h) \quad \rightarrow \text{in generale, il passaggio è } x \rightarrow F(x) \rightarrow T(x)$$

Assumendo l'isofrequenza, questo può diventare il periodo di ritorno di una piena  $q$  (il cui valore non è usato).

**Verifica del periodo di ritorno di un evento pluviometrico (QUANTITATIVA):**

in generale, nelle verifiche si procede secondo lo schema

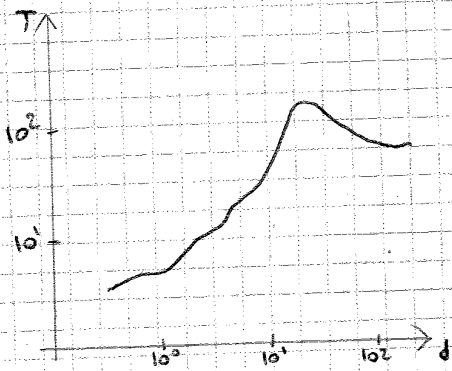
$$Z(\text{moto}) \rightarrow F(Z) \rightarrow T(Z)$$

Se quell'evento è assunto come evento rettangolare, si sfrutta la forma

$$Z_T = Z \left\{ 1 - \text{erfc} \left[ 0,45 + \frac{\sqrt{5}}{\pi} \ln \left( \frac{T}{T-1} \right) \right] \right\}$$

Però, usando l'evento rettangolare, si rischia di sottostimare il risultato

Il risultato della fase di verifica può essere sintetizzato nel diagramma di severità:



esso è un diagramma con

→  $x$  = durata  $d$

→  $y$  = periodo di ritorno  $T$

e fornisce  $x$  ogni durata la rarità (in termini di  $T$ ) del valore osservato.

Ciò aiuta a classificare un evento osservato in fase di verifica (in genere l'analisi è spinta fino al termine dell'evento, riferendosi anche a più stazioni), e ci possono essere grandi durate (100 h).

Dal grafico, si possono capire le caratteristiche salienti

→ un evento temporalesco si caratterizza di alti livelli e piccoli tempi di ritorno

→ nel caso in esempio, la curva minima ha un picco a 50 h e dunque l'aspetto importante di questo esempio evento è la quantità di pioggia riversata in 2 giorni, poiché ha un tempo di ritorno di 300 anni (è importante)

⇒ da questo grafico si capisce la NATURA PIÙ SIGNIFICATIVA DI UN EVENTO (in questo caso, la precipitazione a 48 h), che raramente si vedono nelle curve di probabilità pluviometrica poiché in genere analizzano intervalli più piccoli.

## Progetto

Questa fase rientra nei metodi indiretti ma ora è più complicato perché non conviene ragionare sulle piogge rettangolari.

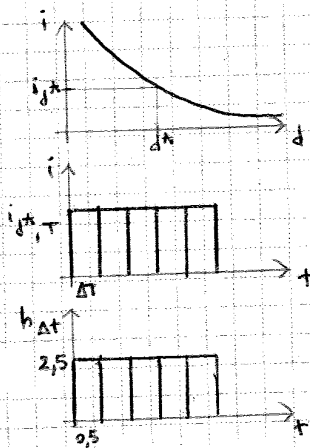
Finora, con l'evento rettangolare, si era ragionato in un certo modo:

l'altezza di precipitazione si può scrivere come

$$x_{d,T} = \bar{x}_d \cdot k(T)$$

Questa espressione però dà solo il totale in tutte le durate, mentre si vuole valutare l'andamento nel tempo. Pertanto, si divide  $x$  la durata e si ricava l'intensità

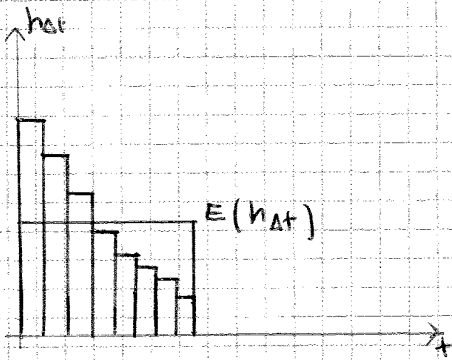
$$i_{d,T} = \frac{x_{d,T}}{d} = a d^{m-1} \cdot k(T)$$



Fissata una certa durata  $d^*$ , si può individuare l'intensità media  $i_{d^*}$  corrispondente e si assume che questa sia l'intensità rappresentativa del pluviogramma, poiché rappresenta cosa accade in un pluviogramma rettangolare.

Per semplicità, si usa un pluviogramma a stogramma. Il vantaggio è che così si può lavorare indifferentemente con altezza o intensità poiché si divide l'intervallo in porzioni di ampiezza  $\Delta t$  e si può introdurre l'intensità media nell'intervallo oppure l'altezza corrispondente, espressa come  $T \cdot \Delta t$ .

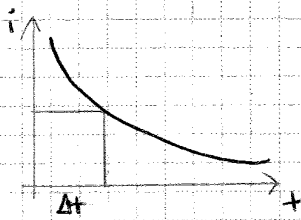
Se ad es.  $\Delta t = 0,5$  h  $i = 5$  mm/h,  
si pone 2,5 mm in ogni mezz'ora



Così sulla curva cumulata si prendono delle differenze, che rappresentano l'altezza della stecca ~~piu~~ del pluviogramma. Si ottiene così un pluviogramma decrescente.

Per controllare il vincolo, basta calcolare la media. Si nota che il vincolo è soddisfatto perché si sommano valori mai sovrapposti (le differenze nella IDF non si sovrappongono) e, al tempo  $t_p$ , la somma dei pezzi dà  $h_{t_p, T}$ .

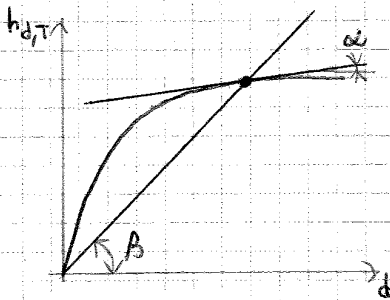
PROBLEMA:



Se si sceglie un intervallo  $\Delta t$  piccolo, c'è il problema che

$$i \rightarrow \infty \text{ se } \Delta t \rightarrow 0$$

cioè, se non ci sono problemi con  $\Delta t$  grandi, con  $\Delta t$  piccoli si ottiene un valore di intensità che non esiste in natura.



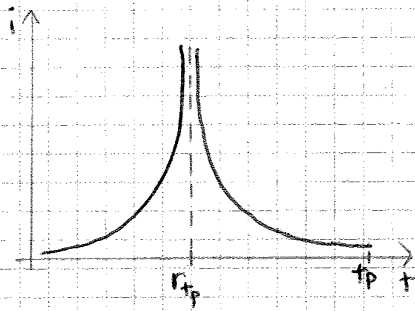
È però possibile ottenere una funzione continua poiché la IDF è derivabile e si può allora considerare la forma istantanea dell'intensità come

$$i(d) = \frac{d}{dt} h(d) = n \cdot d^{n-1}$$

Essa è un'espressione diversa dall'intensità media e rappresenta la pendenza della retta tangente alla curva IDF, mentre la definizione

$$i(d) = \frac{h(d)}{d}$$

cioè l'intensità media, corrisponde alla pendenza della congiungente con l'origine.

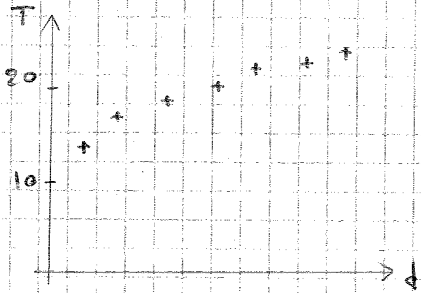


Con questa nuova interpretazione, la ietogramma di Chicago diventa una partizione di questa curva con una fase ascendente, un tempo di picco e una fase discendente.

In tal modo, non si ottiene un pluviogramma con uno scroscio all'inizio ma ci si avvicina alla forma reale, con un picco a circa  $\frac{1}{3}$  della durata totale



in ogni caso, si ha un volume compatibile con la IDF.



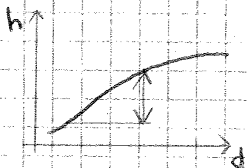
Per questo motivo, si considerano i valori di pioggia cumulata dal minimo fino al massimo valore utile e si prendono gli intervalli di misura (di 10 min).  
Dopo che, ai valori di pioggia caratteristici di questi intervalli, si associa il periodo di ritorno.

Il diagramma che si ottiene si dice diagramma di severità.

Come si costruisce?

Si realizza una tabella con i valori misurati  $\Delta h_{10}$  in 10 min. adiacenti  
Dopo che si crea una lista dei massimi in 10', 20', 30', etc., sommando i valori e considerando il risultato più alto.  
Questi valori rappresentano dei massimi di evento, cioè i massimi valori che si sono avuti in tot minuti.

	$\Delta h_{10}$	$\Delta t$	$\Delta h_{max, \Delta t}$
12:30	8,4	10	25 → il valore massimo in 10' è 25
12:40	25	20	47,4 → il valore massimo in 20' è 25 + 21,4
12:50	21,4	30	68,8
13:00	22,4	40	83,7
13:10	14,4	50	92,6
13:20	0,4	60	93

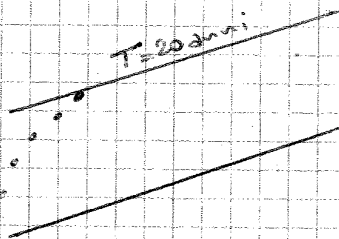


La tabella rappresenta le massime distanze verticali tra punti distanziati nel grafico delle piogge ~~dei~~ cumulate.  
Non c'è rimescolamento, bensì ci sono dei totali di piogge che si sono verificate in sequenza (ma le sequenze hanno un punto d'inizio stabilito da chi fa la verifica).

A questo punto, si definiscono le curve IDF, ricorrendo all'espressione indice

$$h_{d,T} = \bar{h}(d) \cdot k(T) \quad \text{in questo caso, } \bar{h}(d) = 47,5 d^{0,27} \quad \text{e } \overline{CV} = 0,50$$

$h_{d,T}$



In realtà, si può notare che nel piano bilogarithmico le curve IDF sono rette e dunque è immediato tracciarle.  
Dopo che si posizionano in grafico i valori tabellati, in modo tale da coprire qualitativamente i periodi di ritorno.  
Si può notare che tutti i valori sono sotto  $T = 20$ .

Per costruire invece il diagramma di severità, occorre valutare analiticamente il periodo di ritorno, mediante la pdf

$$T(h_d) = \frac{1}{1 - F(h_d)}$$

In questo caso,

$$F(h_d) = \text{Gumbel}(\omega_1; \omega_2)$$

dove  $\omega_1, \omega_2$  dipendono da  $\bar{h}_d$  e  $\overline{CV}$  (poiché nel metodo dei momenti  $\omega_2 = F(\bar{h}_d + \overline{CV} \cdot \bar{h}_d)$ ).

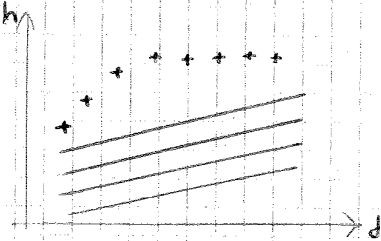
Questo approccio di verifica è semplice poiché, nel metodo di Gumbel-momenti, noto il valore  $k_{ass}$ , è immediato invertire la funzione di crescita  $k(T)$  e ricavare il periodo di ritorno, in quanto questa non dipende dalla durata.

↓ c'è un'unica funzione di crescita  
x tutte le durate

È però possibile che la Gumbel non sia la distribuzione adatta, in quanto non va bene nella zona Appenninica e la Sudare.

In queste zone, se si costruisce la curva di possibilità pluviometrica (IDF), si otterrebbero dei valori sperimentali ben al di sopra della curva relativa a  $T=500$  anni, anche se l'evento non è stato im-

portante. In particolare, la distanza è maggiore nelle durate comprese tra 6h e 12h (qui c'è la situazione peggiore in termini di periodo di ritorno).



Pertanto, almeno x quelle zone, occorre costruire la funzione  $k(T)$  con un'altra distribuzione, cercando però di far sì che questa la funzione di crescita  $k(T)$  comunque sia costante con la durata.

Per fare ciò, si esprime la funzione  $k(T)$  come

$$k(T) = f(\omega_1'; \omega_2'; \omega_3')$$

Essendo  $k(T)$  una funzione adimensionale (e' il quantile diviso media), essa dovrà dipendere da dei parametri di posizione e scala adimensionali.

Allora, x avere la funzione  $k(T)$  costante con la durata, si impone

$$\begin{aligned} \omega_1' &= E[\omega_1'(d)] & \omega_2' &= E[\omega_2'(d)] \\ \omega_3' &= E[\omega_3'(d)] \end{aligned}$$

Una distribuzione molto usata è la GEV (efficace in quelle zone sopra citate), x la quale si ha

$$\omega_1' = \bar{\varepsilon}(d) \quad \omega_2' = \bar{\alpha}(d) \quad \omega_3' = \omega_3(d)$$

Si osserva che la media può essere fatta sulle sole durate disponibili (non sulla serie storica, perché è sufficiente).

Visto che poi la media è applicata su parametri adimensionali,  $k(T)$  non potrà mai variare con la durata (se avesse una tendenza a crescere, significherebbe che ha una dimensione).

In tal modo, si usa un'unica espressione x tutte le durate e non ci sono problemi d'inversione.

→ il bacino contiene delle aste fluviali, cioè pezzi di corsi d'acqua che seguono concavità naturali.  
L'insieme delle aste si dice rete idrografica ed è il sistema che raccoglie le acque superficiali.

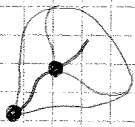
In questa rete, a comodità, si definisce l'asta principale, cioè il percorso più lungo all'interno della rete idrografica.

↳ è un'asta ma di fatto è un percorso

Questo elemento fornisce una dimensione lineare all'interno di un'entità bidimensionale e ha senso perché dà l'idea del percorso che l'acqua deve necessariamente fare.

### PROPRIETÀ

→ lo spartiacque passa nei punti di convessità, tranne nelle selle (= punti topografici sia convessi che concavi), dove passa x i punti singolari.



→ a seconda del numero di sezioni d'interesse, si definiscono più sottobacini.  
In genere infatti, si lavora con più sezioni ma, in questo caso, se ne considera solo una.



→ lo spartiacque che si è definito è detto spartiacque superficiale e con quello si traccia il contorno del bacino idrografico.  
Così, quando la precipitazione incontra il territorio, se il terreno è perfettamente impermeabile, tutta l'acqua che impatta all'interno della spartiacque segue un percorso che porta alla sezione di chiusura.

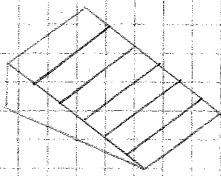
Inizialmente, si supponerà un TERRENO IMPERMEABILE, cosa poco realistica ma vale bene se il terreno è poco permeabile o se i percorsi sotterranei si ~~raccolgono~~ raccolgono nella sezione di chiusura.

Esistono però casi dove il percorso sotterraneo non porta alla sezione di chiusura e si parla allora di spartiacque sotterraneo. Esso si costruisce mediante analisi idrogeologiche ed è importante in presenza di rocce fortemente fratturate e di zone carbonatiche (carsismo).

↓  
in caso contrario, lo spartiacque è sotterraneo e solo la proiezione dello spartiacque superficiale.



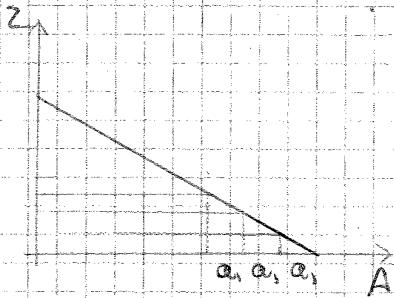
### CASO SEMPLICE : BACINO RETTANGOLARE



Essa è non realistica perché è piano (non ha concavità), ha curve di livello piane e le aree  $A_i$  sono uguali.

Spostandosi dello stesso livello, cioè tra 2 curve di livello, si aggiunge sempre la stessa area

⇒ la curva ipsografica è una RETTA

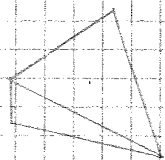


Altezza media vale

$$\begin{aligned} \bar{z} &= \frac{\sum a_i z_i}{A} = \\ &= \frac{\sum a_i z_i}{\sum a_i z} \quad \rightarrow \text{le } a_i \text{ sono tutte uguali} \\ &= \frac{\sum z_i}{z} = \frac{z_{\max} + z_{\min}}{2} \end{aligned}$$

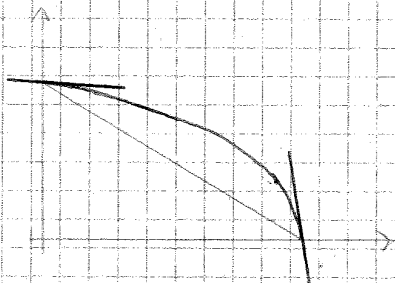
$$\bar{z} = \frac{z_{\max} + z_{\min}}{2}$$

### CASO SEMPLICE : BACINO TRIANGOLARE



Per portare un valore medio  $\bar{z}$  diverso da quello lineare, sapendo che questo è una media pesata, basta variare i pesi e dunque le aree. Si può fare dunque un bacino triangolare.

In tal modo, i pesi si modificano perché aumentano ad alta quota e diminuiscono a bassa quota e dunque l'altezza media si sposta verso l'alto.



A livello di curva ipsografica, è sufficiente notare che la curva lineare di prima era l'integrale di una costante. Se si ha una funzione lineare, allora l'ipsogramma è una PARABOLA tale x cui i punti più in alto hanno area maggiore.

↓ si nota che la derivata  $dA/dz$  varia ed è alta a quote alte.

Si definisce poi la curva ipsometrica, che è la curva ipsografica normalizzata.

Nella normalizzazione, si riportano le variabili da un certo grado di evoluzione generica a un grado di evoluzione tra 0 e 1.

→ x le quote, si toglie  $z_{\min}$  e, avendo ottenuto un dislivello, si divide x il dislivello totale

$$q = \frac{z - z_{\min}}{\Delta z} \quad \text{quota relativa}$$

→ x le aree, si divide l'area parziale x l'area totale

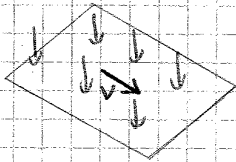
$$a(z) = \frac{a(z)}{A}$$

## Metodi di trasformazione afflussi - deflussi

1 Sono degli algoritmi che prevedono di ignorare la statistica, poiché si usa un sistema concettuale e deterministico in cui si fa una forte schematizzazione della realtà, fatta su base fisica.

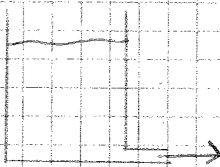
Esistono 2 metodi

→ metodo cinematico: è una classe di metodi in cui, dato un sistema fisico (x semplicità, una superficie piana rettangolare) su cui cade pioggia, si ottiene il risultato in base alla conoscenza della VELOCITÀ  $v$  delle gocce d'acqua caduta.



↓ le gocce si muovono con una certa velocità e i risultati dipendono dalla distribuzione di queste velocità.

→ metodo dell'invaso: è uno schema molto diverso, in cui si ignorano la conformazione spaziale del bacino e la velocità ma si vuole sapere il VOLUME INVASATO in un certo momento.



Infatti, si immagina che, all'aumentare del volume invasato  $W$ , aumenti la velocità d'uscita e dunque aumenti la portata.

$$Q = F(W)$$

↓ velocità vs volume

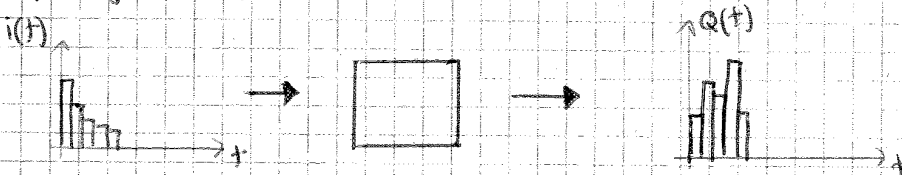
## Metodo cinematico di trasformazione afflussi - deflussi

Innanzitutto, si tiene conto dell'aspetto dimensionale:

l'obiettivo è trovare la variabile  $Q$  che, dimensionalmente, è

$$[Q] = \frac{[L]^3}{[t]}$$

Il sistema di trasformazione è fatto da un ingresso (pluriogramma) e da un'uscita (idrogramma).



La variabile in ingresso, cioè l'intensità, ha dimensioni

$$[i] = \frac{[L]}{[t]}$$

ed è una grandezza distribuita.

Perché diventa portata, occorre moltiplicarla x un'area

$$[i \cdot A] = \frac{[L]}{[t]} \cdot [L]^2 = \frac{[L]^3}{[t]}$$