

DISTRIBUZIONI STATISTICHE AA 2017-2018

"Statistics are like bikinis. What they reveal is suggestive, but what they conceal is vital"



Aaron Leventsein

Le parti marcate in blu non fanno parte del programma

PREMESSA

Per seguire questo corso è necessario conoscere gli elementi essenziali della statistica e della probabilità, elementi che d'altronde sono necessari in qualunque settore dell'ingegneria.

Si dà dunque per scontato che gli allievi conoscano i concetti di: frequenza, probabilità, densità di frequenza e di probabilità, distribuzione cumulata della frequenza e della probabilità, media, mediana, varianza /scarto quadratico medio. Un semplice manuale elementare di statistica¹, o al limite una lettura critica di alcune voci di Wikipedia dovrebbero bastare.

Durante il corso verranno illustrate e svolte alcune procedure per la valutazione di parametri statistici e per la costruzione delle distribuzioni empiriche di frequenza. La maggiore difficoltà per gli allievi di Idraulica Marittima è però quello di comprendere bene gli ambiti in cui si applicano queste procedure. Nel campo marittimo, infatti, si fa uso molti tipi di analisi statistiche, e bisogna bene inquadrare in che contesto e su che tipo di dati si svolgono queste operazioni.

Sulle **altezze d'acqua** $\eta(t)$, in genere campionate con frequenza che va da 1 a pochi Hz, si compiono in questo corso varie analisi, una sola delle quali può essere considerata propriamente statistica: il calcolo di $\bar{\sigma}_\eta$. Questa, e altre operazioni come l'identificazione delle singole onde H_i ed il calcolo dello spettro, sono in genere svolte direttamente dagli ondometri per caratterizzare gli stati di mare, ad intervalli che variano tra i 20' e le 3 ore.

Ha invece interesse calcolare la distribuzione statistica delle **singole onde** H_i all'interno di un dato **"stato di mare"**. Le H_i , come si è visto, seguono in genere la distribuzione di Rayleigh, (con parametro H_{rms}) che fornisce la frequenza delle onde di altezza $H_i < H$, per uno stato di mare caratterizzato da una data H_{rms} (e quindi da una data H_s poiché $H_s = 1.41 H_{rms}$; si ricordi inoltre che $H_s = 4 \bar{\sigma}_\eta$). Nel corso di queste

¹ Per esempio (ma ce ne sono molti altri):
<http://www.dipmat.univpm.it/~demeio/public/Probabilita%20e%20Statistica%20-%20Fermo,%20Ancona/Dispense%20di%20M.%20Garetto.pdf>

Da pag 1 a pag 20.

operazioni vengono dunque calcolati parametri quali H_s , il T_m , T_p , etc. che sono i dati (di carattere statistico) che contraddistinguono il singolo stato di mare.

Il motivo per cui spesso si considerano gli stati di mare e non le singole onde H_i , è il fatto che in quasi tutti i casi non è la singola onda a determinare il collasso di un'opera, ma è l'effetto cumulato di molte onde molto alte a creare problemi; questo è vero in particolare per opere non rigide come le spiagge naturali o artificiali, le gettate di massi o di calcestruzzo e talvolta i cassoni poggiati sul fondo. I valori quali H_s , T_m etc servono inoltre a ricavare informazioni per la tendenza all'erosione di un litorale, la profondità di chiusura, la non agibilità dell'ingresso in un porto etc.

Elaborazioni statistiche sulle singole onde in un determinato stato di mare

Una delle *esercitazioni mostra come si calcola la frequenza cumulata* delle singole onde di altezza maggiore di un dato valore e del loro numero in una mareggiata della durata di un'ora.

Questo è utile ad esempio per calcolare il numero di singole onde – all'interno di un determinato stato di mare – la cui altezza sia al sopra di una certa soglia in un determinato tempo - **parametro richiesto da alcune formule per il dimensionamento delle mantellate. Analogamente si potrebbe procedere a calcolare (ma non verrà fatto in questo corso) la singola onda massima di uno stato di mare, dato utile per la progettazione di alcune opere "rigide", la cui sopravvivenza cioè dipende dal valore della massima sollecitazione che verrà incontrata nel corso dell' esistenza.**

Elaborazioni statistiche dei parametri degli stati di mare ("clima del moto ondoso")

Sui parametri degli stati di mare (H_s , T_m , T_p etc) che possono essere triorarii, orarii, semi-orari etc si compiono poi ulteriori elaborazioni statistiche. In questo caso, l'universo statistico è proprio quello degli H_s o dei T_p , o dei T_m , degli stati di mare di una lunga registrazione, ad es. di un anno, o più

Dunque gli H_s o gli T_p , o gli T_m , che sono, come si è visto, parametri statistici di un periodo di misura, diventano a loro volta le variabili casuali delle analisi statistiche, *che saranno svolte ad esercitazione.*

Distribuzioni di frequenza

Si costruiscono le distribuzioni di frequenza degli eventi e delle energie in funzione delle altezze d'onda significative, e le stesse distribuzioni in funzione delle direzioni; tutto ciò ripetiamo, facendo riferimento agli stati di mare che possono avere diversa lunghezza: ad esempio 30' per i dati di boa, 6h o 12h per i dati di modello.

Utili esercizi:

distribuzione e cumulata per classi di altezza delle mareggiate di altezza significativa H_s

distribuzione per classi di direzione delle mareggiate; per vari valori di soglia di altezza significativa H_s

distribuzione per classi di direzione del flusso di energia delle mareggiate

Valori estremi

Un'informazione molto importante è quella della probabilità di stati di mare molto alti (estremi). Si usano le stesse tecniche (adattamento delle distribuzioni di valori estremi) impiegate ad esempio nelle Costruzioni Idrauliche. Nel seguito ne vengono brevemente esposti i concetti essenziali

Lo strumento essenziale è il c.d "tempo di ritorno" T_r relativo ad un'altezza significativa² H_r : l'intervallo medio tra due eventi con $H \geq H_r$.

Supponiamo di conoscere la probabilità cumulata³ $P(H_r)$ degli eventi di una distribuzione, cioè la probabilità che un evento H sia minore di H_r (probabilità di non superamento).

Si ha che $1-P(H_r)$ è la probabilità che un evento H sia $\geq H_r$. (probabilità di superamento).

Si definisce "tempo di ritorno di un evento di intensità H_r " :

$$T_r(H_r) = 1/(1-P(H_r)) ,$$

cioè l'inverso della probabilità di superamento. Tale probabilità è relativa all'unità di tempo considerata (di solito un anno). Il T_r è dunque l'intervallo medio (in anni) tra due eventi maggiori o eguali ad H_r .

Prima di considerare le probabilità $P(H)$, occorre lavorare e familiarizzarsi con le frequenze empiriche F_i dei dati a disposizione. Si devono seguire dunque diversi passi:

- 1) Estrazione degli eventi estremi indipendenti dalla successione storica degli stati di mare
- 2) Costruzione della funzione di distribuzione empirica F_i degli eventi estremi indipendenti e stima del tempo di ritorno sempre relativo ai dati empirici
- 3) Adattamento delle funzioni di probabilità $P(H_r)$ e quindi della funzione $H_t(T_r)$ che lega il tempo di ritorno all'altezza significativa

1)

Gli eventi considerati sono i valori estremi della successione temporale delle altezze significative indipendenti degli stati di mare H_i . "indipendenti" va qui inteso in senso statistico: vanno presi i valori massimi della successione di dati, escludendo quelli tra loro collegati. Bisogna considerare che gli eventi indipendenti, in realtà, non sono gli stati di mare bensì le mareggiate, ovvero i fenomeni meteorologici che le causano: "perturbazioni", "fronti" "uragani", etc. Il parametro H da considerare è H_r , altezza significativa più alta della perturbazione.

Una delle tecniche più comuni si chiama POT (Peak Over Threshold = massimi sopra soglia)

Si procede come segue:

Si parte dalla registrazione degli stati di mare di un determinato punto : possono essere dati ogni 6 ore, ogni 3 ore, ogni ora, ogni 30' o 20', etc. Si hanno N anni ⁴ di osservazione

Si individuano i dati al di sopra di una certa soglia: ad es. 4 metri.

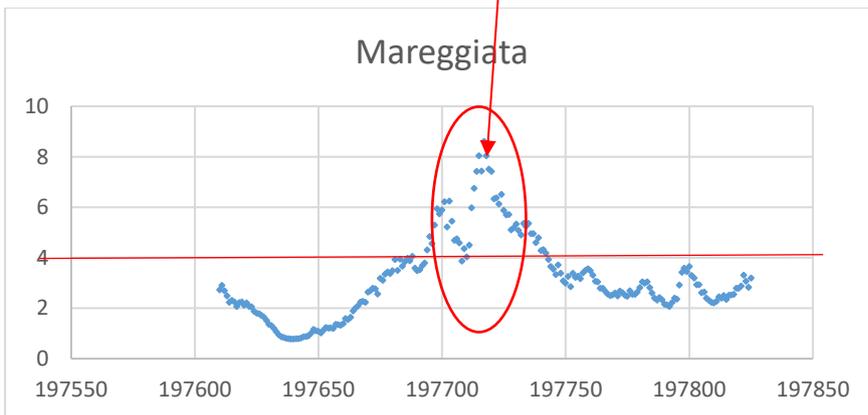
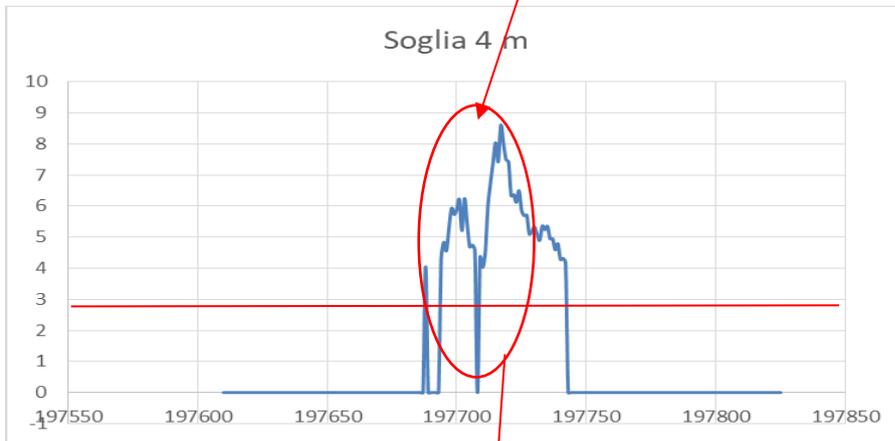
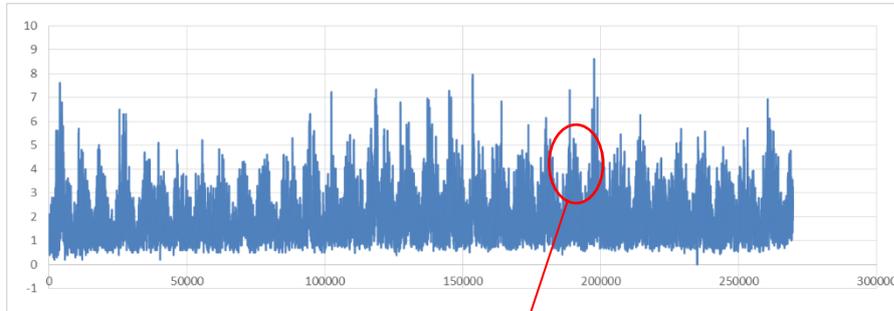
² Da qui in poi, con "H" si intende l'altezza significativa di uno stato di mare; non l'altezza della singola onda. Fare molta attenzione

³ Ricordare il concetto di Probabilità (o frequenza cumulata)

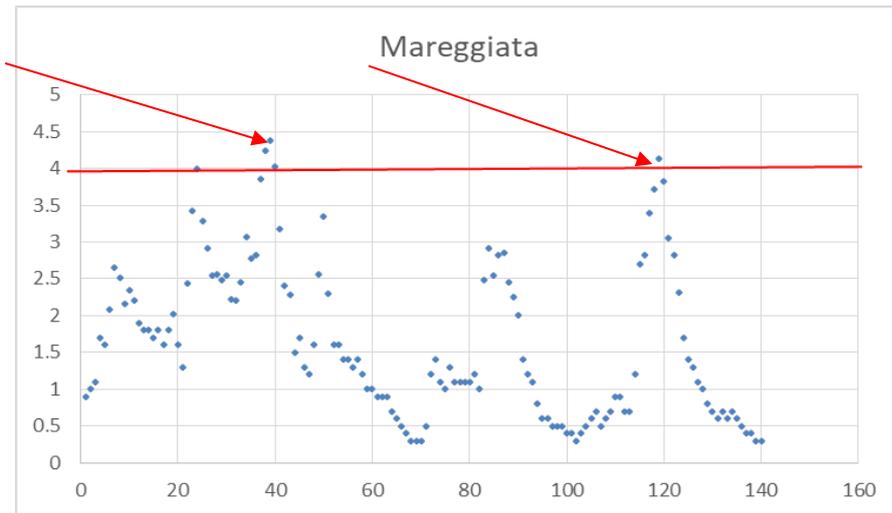
⁴ N puo' anche non essere intero

Si eliminano i dati che pure sopra soglia – appartengono alla stessa mareggiata (= perturbazione atmosferica, tempesta, fronte). In sostanza, si considera solo il singolo valore massimo per ciascuna mareggiata.

Ciò detto si procede come segue: dalla successione degli stati di mare da una successione di N anni, si estraggono tutti quelli superiori alla soglia SH prefissata.

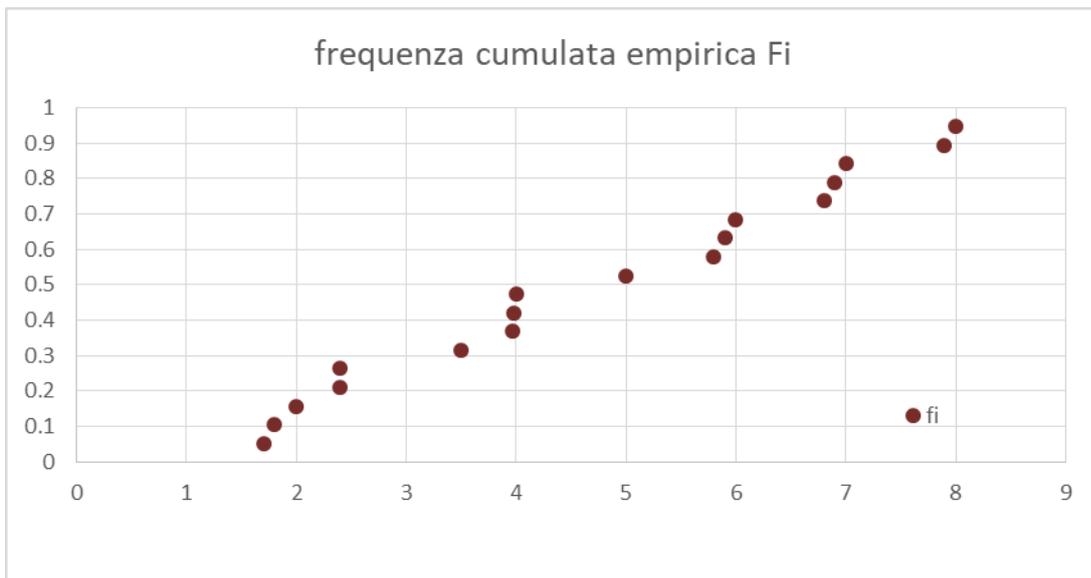


Si escludono quindi gli stati di mare che appartengono alla stessa perturbazione (quindi , vicini tra di loro) e si considera per ciascun evento solo il valore H_i più alto.



Indichiamo con N_t il numero totale di massimi così individuato e con $\lambda = N_t/N$ il numero medio di eventi estremi per anno. Si costruisce la distribuzione empirica di non superamento, associando ad ogni evento H_i la frequenza F_i di non superamento (cioè il numero di di eventi minori o uguali ad H_i , diviso in numero totale di eventi N_t). Cio' si fa facilmente ordinando gli eventi in ordine di altezza, ed assegnando a ciascuno la frequenza⁵ F_i :

$$F_i = i / (N_t + 1) \quad (1)$$

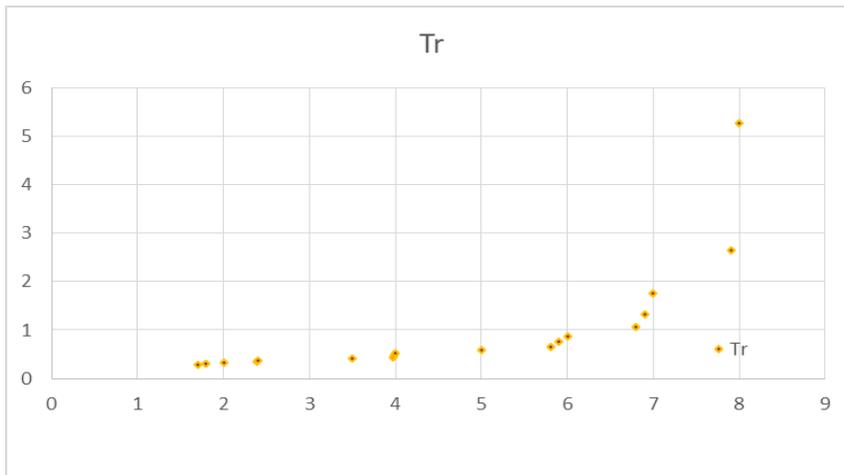


2)

Da questa si può ricavare la distribuzione empirica di superamento: $1-F_i$ e quindi il tempo di ritorno

$$T_r = 1 / [\lambda(1-F_i)]$$

⁵ Esistono varie formulazioni empiriche di F_i ; quella riportata è la più semplice e intuitiva disponibile, tenendo presente che al valore più alto non può corrispondere il valore $F=1$



Il tempo di ritorno è il tempo medio intercorrente tra mareggiate con onde significative maggiori o eguali di un dato valore. E' un dato essenziale per la progettazione o l'analisi del rischio.

https://it.wikipedia.org/wiki/Tempo_di_ritorno

3)

Fin qui abbiamo considerato solo dati empirici, il che comporta un ovvio limite: i valori di T_r non possono superare quelli della durata dell'osservazione N ; per superare questo limite occorre interpolare ai dati empirici una qualche distribuzione analitica di probabilità $P(H)$. https://en.wikipedia.org/wiki/Extreme_value_theory

Si cercano cioè i parametri di particolari funzioni (distribuzioni dei valori estremi) che meglio si adattano ai valori empirici

La distribuzione più comunemente usata è la distribuzione di Weibull, che è data da

$$P(H) = 1 - \exp \left[- \left(\frac{H-B}{A} \right)^k \right] \quad (2)$$

Dove A , B e k sono noti rispettivamente come parametri di scala, posizione e forma.

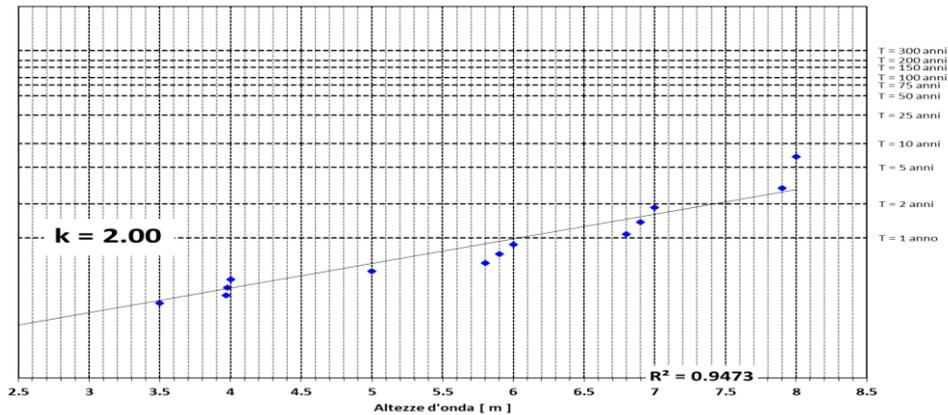
$P(H)$ è la probabilità che l'altezza significativa superi il valore H durante periodo di osservazione N .

L'adattamento può essere fatto con diverse tecniche: nel foglio EXCEL fornito con le esercitazioni è implementata una procedura dei minimi quadrati. Il parametro di forma k può venire assegnato per tentativi, o essere utilizzato per migliorare l'adattamento.

Grazie all'interpolazione, valori di altezza d'onda $H(T_r)$ si possono quindi ottenere per valori di T_r molto alti, tenendo presente che tra T_r e la probabilità cumulata di non superamento $P(H)$ esiste la relazione:

$$P(H) = 1 - \frac{1}{\lambda T_r} \quad (1bis)$$

Il diagramma che lega T_r con H è essenziale per le analisi di rischio e la progettazione delle opere; i risultati vengono forniti spesso su carte distorte, per favorirne la lettura.



$$H(T_r) = B + A [\ln(\lambda T_r)]^{\frac{1}{k}} \quad (2)$$

Considerazioni operative

Una parola di cautela va aggiunta comunque, in particolare quando – come spesso accade – si cercano i valori di altezza d'onda per tempi di ritorno molto alti, quelli che cioè dovrebbero essere impiegati per le verifiche estreme delle opere. Il problema è che si cerca un'informazione relativa a fenomeni, per esempio, centennali, mentre si dispone al più di 30-40 anni di dati (nel caso delle onde; nel caso dei venti la situazione è più favorevole). Tutto quello che in realtà si fa è estrapolare una distribuzione statistica, e quanto più ci si allontana dall'intervallo temporale per cui sono disponibili i dati tanto più aumenta l'arbitrarietà dell'estrapolazione; in altre parole i valori relativi a tempi di ritorno di qualche anno sono abbastanza affidabili, quelli relativi a tempi molto lunghi lo sono molto meno.

Poiché in molti casi l'onda massima che può raggiungere un'opera costiera è determinata dal fondale e non dal valore al largo (come è visto), questo problema è meno grave di quel che sembra, almeno per le opere che sono imbasate a bassa profondità (è il caso di molti porti turistici, e certamente di quasi tutte le opere di protezione dei litorali); questo spiega il perché della relativa varietà di conclusioni di molti studi meteomarinari, che portano poi a dimensionamenti diversi ma generalmente senza conseguenze catastrofiche.

Questo non deve però, al contrario, condurre alla erronea sensazione che l'analisi del clima sia un esercizio inutile ai fini del dimensionamento delle opere su bassi fondali, e ciò per il motivo essenziale che una corretta progettazione non si fa solo sulla singola mareggiata estrema, con tempi di ritorno molto alti ma anche sulla frequenza e l'intensità di molti eventi minori che, seppure con un tempo di ritorno più basso, sollecitano comunque le strutture o ne limitano l'utilità, oppure l'agibilità. Si pensi per esempio al cumulo del danno su una scogliera, o alla frequenza di inagibilità di un porto, o agli effetti sul trasporto solido di una protezione litoranea.

Per tutti questi motivi il progetto di un'opera non è basato su una singola "onda di progetto" con un determinato ed alto tempo di ritorno, ma anche una serie di dati statistici relativi ad eventi con probabilità più bassa. Sono quindi importanti i valori dell'altezza d'onda con tempi di ritorno annuali, o anche meno che annuali, così come la distribuzione direzionale media. Fortunatamente – e come è ovvio - il livello di incertezza su questi eventi di maggior frequenza è molto più basso. E' quindi possibile ed utile in questi casi operare direttamente sulle distribuzioni empiriche dei dati anziché su distribuzioni standard adattate.