DISTRIBUZIONE ESPONENZIALE

Nella forma più usuale la distribuzione esponenziale assume la forma:

$$f(x) = \frac{1}{\theta_2} e^{-\frac{x}{\theta_2}} \qquad F(x) = 1 - e^{-\frac{x}{\theta_2}}$$

Le relazioni teoriche dei momenti sono:

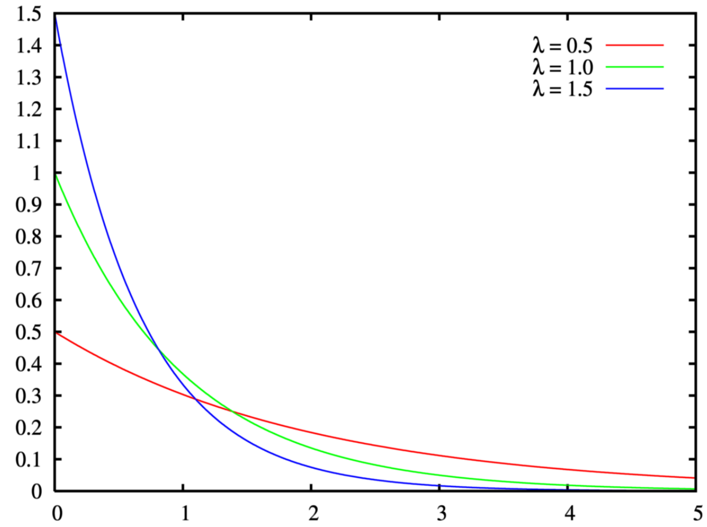
$$\mu = \theta_2$$

$$\sigma^2 = \theta_2^2$$

$$Cv = 1$$

$$\gamma = 2$$

$$k = 9$$



DISTRIBUZIONE ESPONENZIALE a 2 parametri (con soglia)

Se esiste un limite inferiore della variabile, diverso da zero, la distribuzione esponenziale assume la forma:

$$f(x) = \frac{1}{\theta_2} e^{\frac{-(x-\theta_1)}{\theta_2}} \qquad F(x) = 1 - e^{\frac{-(x-\theta_1)}{\theta_2}}$$

Le relazioni teoriche dei momenti in questo caso sono:

$$\mu = \theta_1 + \theta_2$$

$$\sigma^2 = \theta_2^2 \qquad \gamma = 2$$

$$Cv = \frac{\theta_2}{\theta_1 + \theta_2} \qquad k = 9$$

MODELLO DEL MASSIMO VALORE

Si consideri una serie di n variabili casuali $(X_1, X_2, \ldots, X_n)$ e sia $X_{(n)} = \max_{1 \le i \le n} X_i$

$$F_{X(n)}(x) = P[X_{(n)} \le x] = P[X_1 \le x, X_2 \le x, \ldots, X_n \le x]$$

Se gli X_i sono indipendenti:

$$F_{X(n)}(x) = P[X_1 \le x] \cdot P[X_2 \le x] \cdot \ldots \cdot P[X_n \le x] = F_{X1}(x) F_{X2}(x) \ldots F_{Xn}(x)$$

Se gli X_i sono identicamente distribuiti con funzione di distribuzione cumulata $F_x(x)$

$$F_{X(n)}(x) = [F_X(x)]^n \qquad \text{o semplicemente} \qquad F_n(x) = [F(x)]^n$$

Esempio:

Nota, dall' esperienza passata, la distribuzione della massima piena in un anno, o piena annuale X, si può valutare la distribuzione della massima piena in n anni, se n è la durata di progetto del sistema idrico.

DISTRIBUZIONI ASINTOTICHE

$$F_n(x) = [F(x)]^n = \left[1 - n(1-F) + \frac{n(n-1)}{2}(1-F)^2 - \frac{n(n-1)(n-2)(1-F)}{6} + \ldots\right]$$

$$\lim_{n\to\infty} F_n(x) = \left[1 - n(1-F) + \frac{n^2}{2}(1-F)^2 - \frac{n^3}{6}(1-F)^3 + \ldots\right] = e^{-n(1-F)}$$

Per n grande:

$$F_n(x) = e^{-n(1-F(x))}$$

Al crescere di n la distribuzione di $X_{(n)}$ è relativamente insensibile alla esatta forma della distribuzione delle X_i. E' stato dimostrato che, a seconda della caratteristica della distribuzione delle X_i nella coda destra, la distribuzione della massimo $X_{(n)}$ tende a solo 3 distribuzioni asintotiche:

1) $$\Phi_1 = exp\left[-e^{-\frac{1}{\theta_2}(x-\theta_1)}\right] \qquad \frac{1}{\theta_2} > 0$$

2) $$\Phi_2 = exp\left[-\left(\frac{\theta_1}{x}\right)^{\theta_2}\right] \qquad x \geq 0; \theta_1 > 0; \theta_2 > 0$$

3) $$\Phi_3 = exp\left[-\left(\frac{x}{\theta_1}\right)^{\theta_2}\right] \qquad x \geq 0; \theta_1 < 0; \theta_2 > 0$$

La distribuzione asintotica del tipo 1 o di Gumbel vale quando la coda superiore della variabile originaria cade in maniera esponenziale, cioè se, per abbastanza grande vale:

$$F_x(x) = 1 - e^{-g(x)}$$ Coda di tipo esponenziale

essendo $g(x)$ una funzione crescente di x.

Esempio:

$$g(x) = \frac{1}{\theta_2}x \qquad F_x(x) = 1 - e^{-\frac{1}{\theta_2}x}$$

Distribuzione esponenziale: $E(x) = \theta_2$

$$F_{X_{(n)}}(x) = \exp[-n(1 - F_X(x)] = \exp\left[-ne^{-\frac{1}{\theta_2}x}\right] = \exp\left[-e^{-\frac{1}{\theta_2}(x-\theta_1)}\right] = \Phi_1(x)$$

con $n = \exp\left(\frac{1}{\theta_2}\theta_1\right)$

Il massimo di n variabili esponenziali con media θ_2 è distribuito secondo la legge di Gumbel per n abbastanza grande.

DISTRIBUZIONE DEL VALORE ESTREMO DI 1° TIPO (EV1) o DI GUMBEL

Interpretazione:

1. **Asintotica** $X = \max_{1 \leq i \leq k} Z_i$

    - Z_i indipendenti ed identicamente distribuiti con leggi di tipo esponenziale.
    - $F_Z(z) = P[Z \leq z] = \left[1 - e^{-\frac{1}{\theta_2}z}\right]$
    - $E[z] = \theta_2$
    - $F_X(x) = P[X \leq x] = \lim_{k \to \infty} [F_Z(x)]^k = e^{-k[1-F_Z(x)]} = e^{-ke^{-\frac{1}{\theta_2}x}}$

2. **Esatta:** $X = \max_{1 \leq i \leq K} Z_i$

    - K numero variabile da anno ad anno con la legge di Poisson.
    - $p_k(k) = P[K = k] = \dfrac{e^{-\Lambda}\Lambda^k}{k!}$ $\quad \{k = 0, 1, 2, \ldots\}$
    - $E[K] = \Lambda$

Z_i vedi sopra ed inoltre indipendenti da K.

$$F_X(x) = P[X \leq x] = e^{-\Lambda[1-F_Z(x)]} = e^{-\Lambda e^{-\frac{1}{\theta_2}x}} \exp\left(\frac{1}{\theta_2}\theta_1\right) = \Lambda$$

X: massimo in un intervallo $[0,t]$, esempio l'anno, di un numero variabile secondo la legge di Poisson, con media Λ, di variabili indipendenti ed esponenzialmente distribuite con media θ_2.

DISTRIBUZIONE EV1 o DI GUMBEL

$$F_X(x) = \exp\left[-\Lambda e^{-\frac{1}{\theta_2}x}\right] = \exp\left[-e^{-\frac{1}{\theta_2}(x-\theta_1)}\right]$$

con: $$\exp\left(\frac{1}{\theta_2}\theta_1\right) = \Lambda$$

Significato dei parametri:

- Λ : Numero medio di eventi indipendenti in $[0,t]$, esempio in un anno.
- θ_2 :Valore medio della grandezza dell' evento, esempio portata al colmo.

$$f_X(x) = \frac{dF_X(x)}{dx} = \frac{1}{\theta_2}\exp\left[-\frac{1}{\theta_2}(x-\theta_1) - e^{\frac{1}{\theta_2}(x-\theta_1)}\right]$$

$$E[x] = \mu = \theta_1 - 0.57722\theta_2 \qquad \theta_1 = \mu - 0.450\sigma$$

$$var[x] = \sigma^2 = \frac{\pi^2}{6\left(\frac{1}{\theta_2}\right)^2} \qquad \theta_2 = \frac{\sigma\sqrt{6}}{\pi}$$

$$Cv[x] = \frac{\sigma}{\mu} = \frac{\pi}{\sqrt{6}\left(\frac{1}{\theta_2}\theta_1 + 0.57722\right)} = \frac{\pi}{\sqrt{6}(\ln\Lambda + 0.57722)}$$ dipende solo da Λ

$$Ca[x] = \frac{\mu^3}{\sigma^3} = 1.1396$$ Coefficiente di asimmetria, <u>indipendente dal valore dei parametri</u>.

$$f_x(\theta_1) = \frac{1}{\theta_2}e^{-1}$$ θ_1 è la moda di x $$F_x(\theta_1) = e^{-1} = 0.368$$

θ_2 : è proporzionale a σ (misura di dispersione).

$\left(\frac{1}{\theta_2}\theta_1\right)$, Λ : dipendono solo dal coefficiente di variazione .

Variabile ridotta: $Y = \frac{1}{\theta_2}(x - \theta_1)$

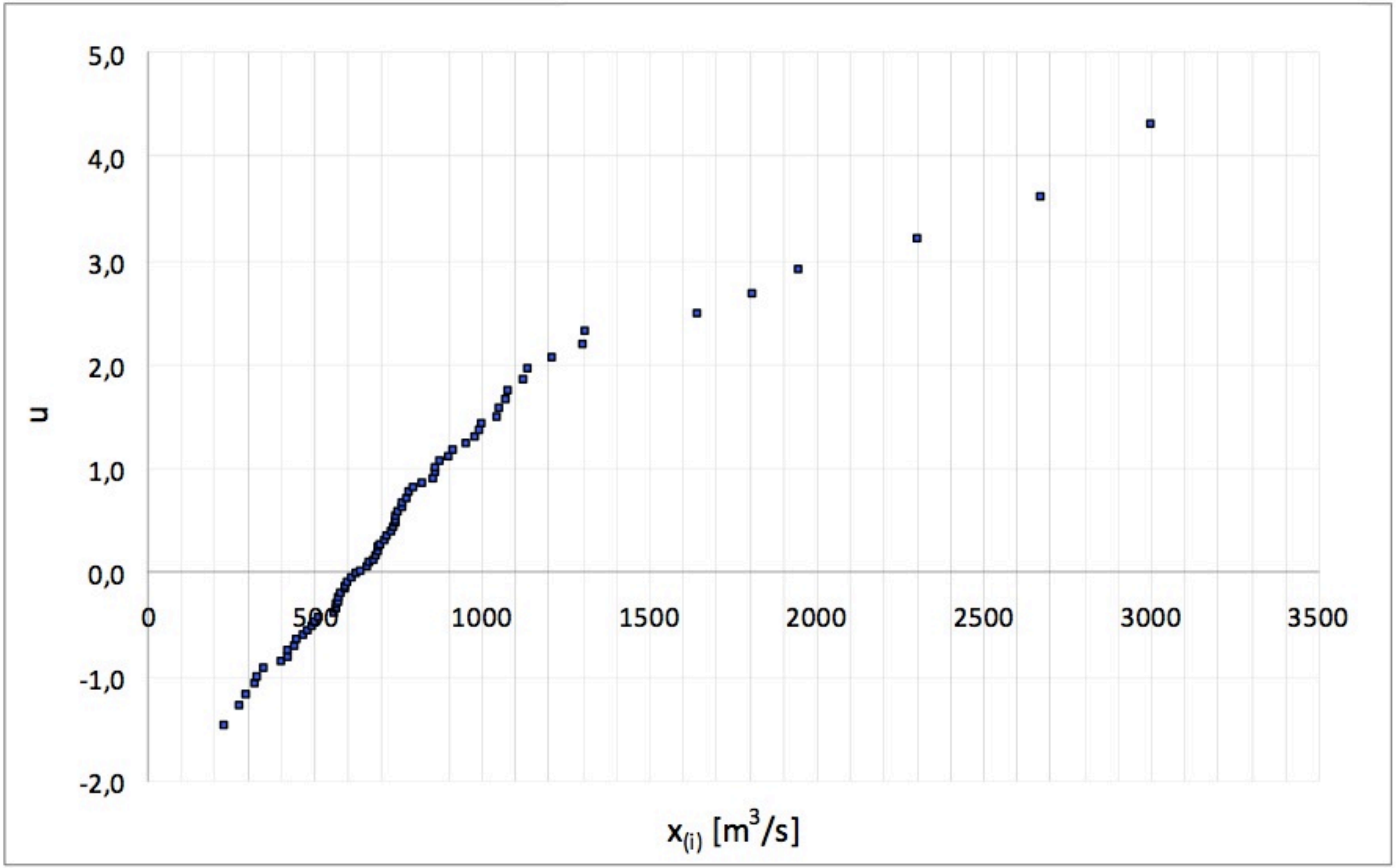
- $E[Y] = 0.57722$ *costante di Eulero*
- $var[Y] = \frac{\pi^2}{6}$
- $\theta_1[Y] = 0$
- $\theta_2[Y] = 1$
- $\Lambda[Y] = 1$

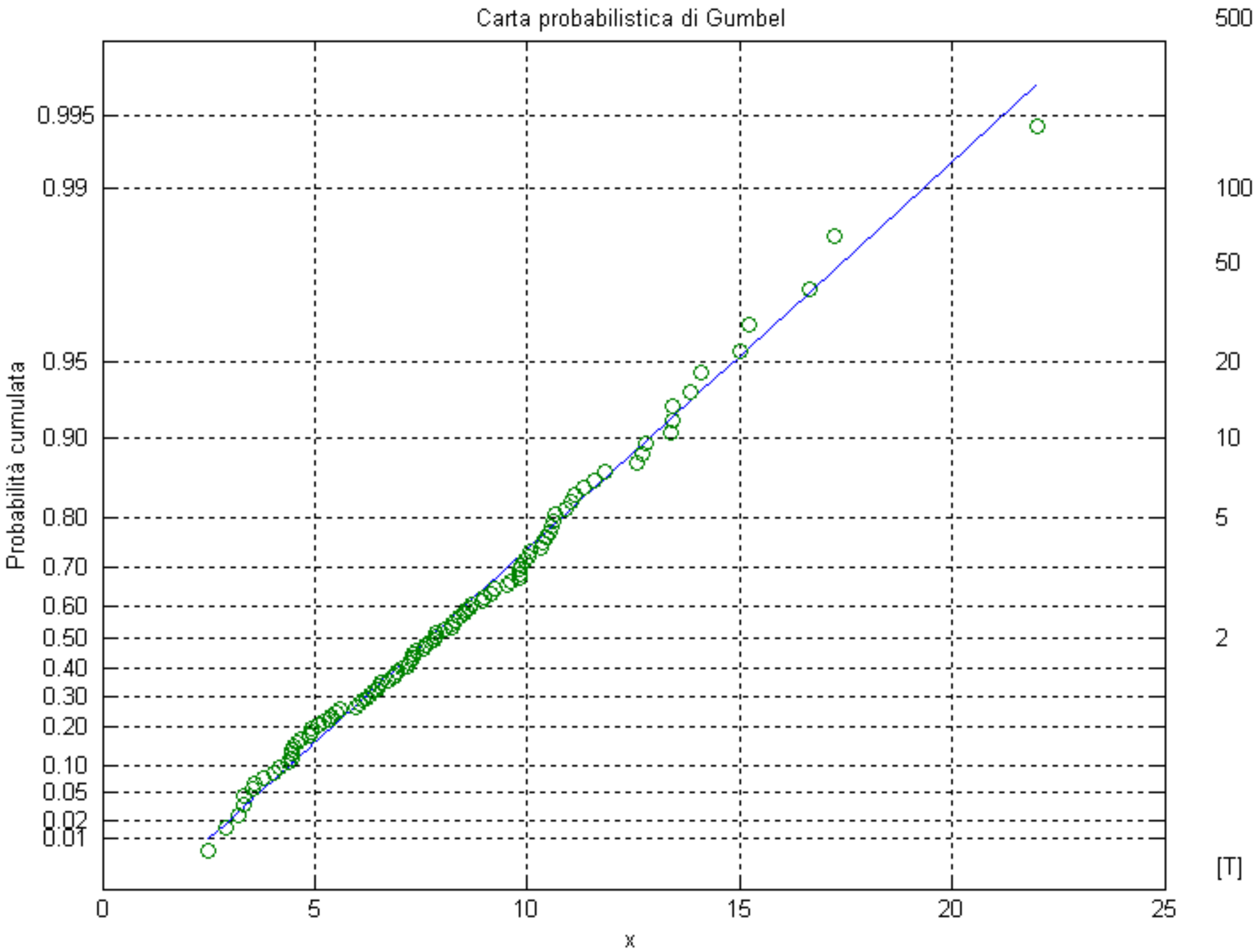
$\Phi(Y) = P[Y \leq y]$ $\qquad F_Y(y) = \exp(-e^{-y})$

$Y \approx EV1(0,1)$ $\qquad X \approx EV1(\theta_1, \theta_2)$

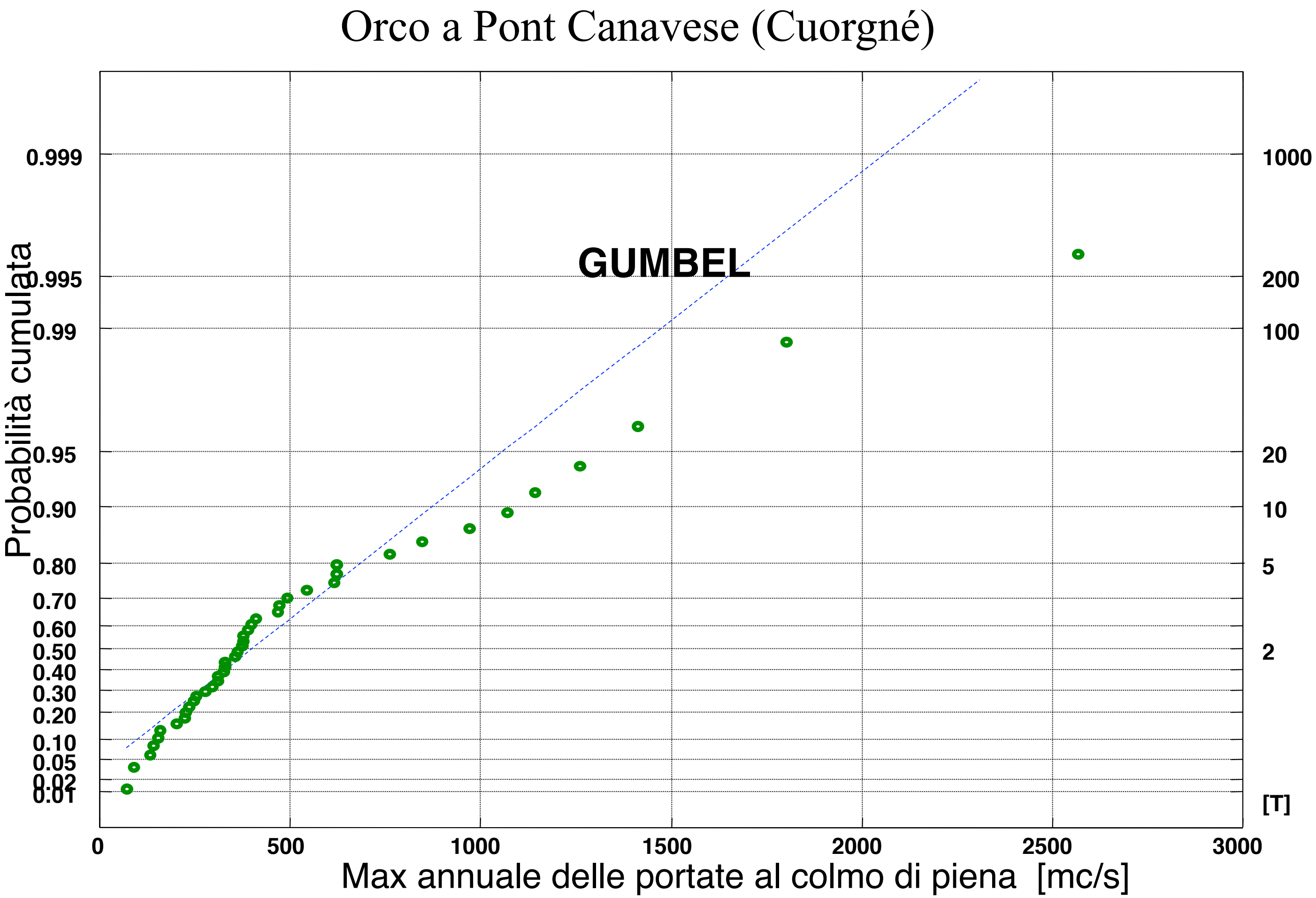
$$Y_F = -\ln\ln\frac{1}{F} \qquad X_F = \theta_1 + Y_F\theta_2 = \theta_1 - \theta_2\ln\ln\frac{1}{F} = \theta_2\left(\ln\Lambda - \ln\ln\frac{1}{F}\right)$$

Carta probabilistica di Gumbel




Carta probabilistica di Gumbel
Probabilità cumulata
x
0.995
0.99
0.95
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.05
0.02
0.01
0
5
10
15
20
25
500
100
50
20
10
5
2
[T]

Serie di dati molto asimmetriche non si allineano in carta di Gumbel



dipendenza del valore di progetto dal metodo di stima

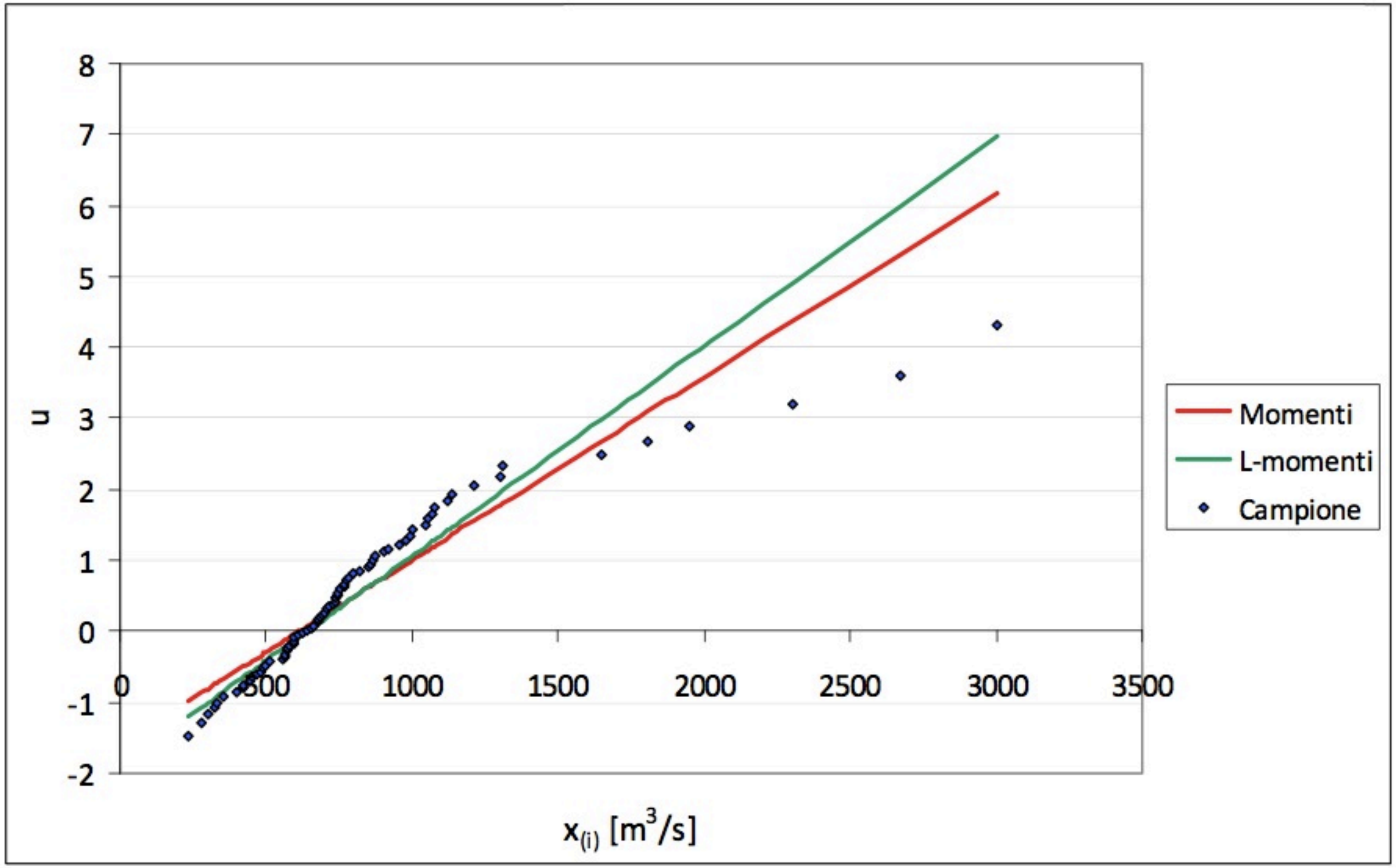


Figura 2.6 Confronto rette a parametri stimati su carta probabilistica di Gumbel.

IL METODO DELLA MASSIMA VEROSIMIGLIANZA

Consiste nel determinare i valori dei parametri in modo da massimizzare la funzione di verosimiglianza:

$$V(x_1, x_2, \dots, x_n; \vartheta_1, \vartheta_2, \dots, \vartheta_r) = f_x(x_1) f_x(x_2) \dots f_x(x_n)$$

che è la funzione di densità congiunta del campione per osservazioni indipendenti.

La stima dei parametri $\vartheta_1, \vartheta_2, \dots, \vartheta_n$ si ottiene dal sistema di r equazioni in r incognite.

$$\frac{\delta V}{\delta \vartheta_1}(x_1, x_2, \dots, x_n; \vartheta_1, \vartheta_2, \dots, \vartheta_r) = 0 \qquad i = 1, 2, \dots, r$$

Stima di Massima Verosimiglianza: Distribuzione normale

$$V(x_1, x_2, ..., x_n; \theta_1, \theta_2) = \frac{1}{(2\pi\theta_2)^{\frac{n}{2}}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i - \theta_1}{\sqrt{\theta_2}}\right)^2\right]$$

$$\ln V = -\ln[2\pi\theta_2]^{\frac{n}{2}} - \frac{1}{2\theta_2}\sum_{i=1}^{n}(x_i - \theta_1)^2$$

$$\max(V) \Rightarrow \max(\ln V) \Rightarrow \min(-\ln V)$$

$$\frac{\delta}{\delta\theta_1}\left[\frac{1}{2\theta_2}\sum_{i=1}^{n}(x_i - \theta_1)^2 + \ln(2\pi\theta_2)^{\frac{n}{2}}\right] = -\left[\frac{1}{\theta_2}\sum_{i=1}^{n}(x_i - \theta_1)\right] = 0$$ verificata se $\theta_1 = \mu$

$$\frac{\delta}{\delta\theta_2}\left[\frac{1}{2\theta_2}\sum_{i=1}^{n}(x_i - \theta_1)^2 + \ln(2\pi\theta_2)^{\frac{n}{2}}\right] = -\frac{1}{2\theta_2^2}\sum_{i=1}^{n}(x_i - \theta_1)^2 + \frac{n}{2\theta_2} = 0$$ verificata se $\theta_2 = \sigma^2$

Stima di Massima Verosimiglianza: Distribuzione di Gumbel

$$\theta_2 = \bar{x} - \frac{\sum_{i=1}^{n} x_1 \cdot e^{-\frac{1}{\theta_2} x_i}}{\sum_{i=1}^{n} e^{-\frac{1}{\theta_2} x_i}}$$

$$e^{-\frac{\theta_1}{\theta_2}} = \frac{1}{n} \cdot \sum_{i=1}^{n} e^{-\frac{1}{\theta_2} x_i}$$

con n=numero di dati del campione.

Il valore a primo membro è determinato sulla base di un valore di tentativo assegnato al secondo membro. Il procedimento si arresta quando i due valori differiscono di una quantità inferiore alla tolleranza assegnata (per es. 0.001 per θ_2).

Il primo valore di tentativo potrà essere quello derivante dalla stima dei momenti.

La correzione che si può apportare ad ogni tentativo per accelerare la convergenza può essere quella suggerita da Gumbel: $\theta_2(3) = \theta_2(2) + (\theta_2(1) - \theta_2(2))/3$

L-momenti

Gli L-momenti sono un modo alternativo di descrivere la forma delle distribuzioni di probabilità.
Derivano dai *probability weighted moments* (PWM) o momenti pesati in probabilità.

Per una variabile casuale x con funzione di probabilità cumulata $F(x)$ si ha

$$M_{p,r,s} = E\left[x^p \{F(x)\}^r \{1 - F(x)\}^s\right]$$

L-momenti vs momenti

$$\beta_r = M_{1,r,0} = E\left[x\{F(x)\}^r\right]$$

$$M_r = E\left[(x-\mu(x))^r\right]$$

$$\int_{-\inf}^{\inf} x\{F(x)\}^r \cdot f(x)\mathrm{d}x$$

$$\int_{-\inf}^{\inf} (x-\mu(x))^r \cdot f(x)\mathrm{d}x$$

PWM $\rightarrow$ lineari nella variabile casuale (la x non è mai elevata a potenza, ma lo è solo $F(x)$)
Momenti $\rightarrow$ la x è elevata a potenza per $r > 1$

L-momenti

I PWM possono essere usati per stimare i parametri di una distribuzione, ma sono difficili da interpretare direttamente come misure di dispersione, forma, ...

Tali informazioni sono però presenti in particolari combinazioni di PWM

$\lambda_1 = \beta_0 \rightarrow$ media

$\lambda_2 = 2\beta_1 - \beta_0 \rightarrow$ dispersione

$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 \rightarrow$ asimmetria

$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0$

Coefficienti adimensionali

$$\tau = \frac{\lambda_2}{\lambda_1} \rightarrow \text{L-CV}$$

$$\tau_3 = \frac{\lambda_3}{\lambda_2} \rightarrow \text{L-skewness}$$

$$\tau_4 = \frac{\lambda_4}{\lambda_2} \rightarrow \text{ L-kurtosis}$$

in generale $\tau_r = \frac{\lambda_r}{\lambda_1}$

L-momenti

Notazione:

$x_1, x_2, \ldots, x_i \ldots x_n$ campione

$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(i)} \leq \ldots \leq x_{(n)}$ campione ordinato in senso crescente

$x_{1:n} \leq x_{2:n} \leq \ldots \leq x_{i:n} \leq \ldots \leq x_{n:n}$ campione ordinato in senso crescente

L-momenti

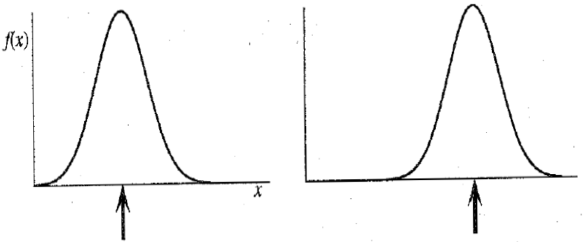


Fig. 2.1. Definition sketch for first *L*-moment.

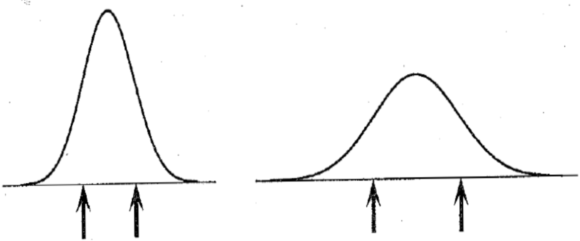


Fig. 2.2. Definition sketch for second *L*-moment.

L-momenti

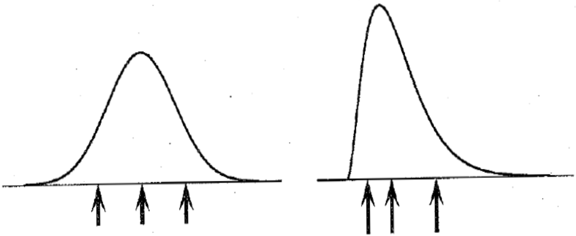


Fig. 2.3. Definition sketch for third L-moment.

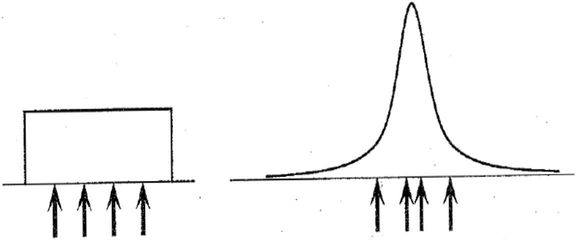


Fig. 2.4. Definition sketch for fourth L-moment.

Proprietà

esistenza se la media della distribuzione esiste, allora esistono tutti gli L-momenti

unicità se la media della distribuzione esiste, allora gli L-momenti definiscono in maniera univoca la distribuzione (non ci sono due distribuzioni con gli stessi L-momenti)

- $\lambda_2 \geq 0$
- $|\tau_r| < 1$ per tutti $r \geq 3$
- $\frac{1}{4}(5\tau_3^2 - 1) \leq \tau_4 < 1$
- per distribuzioni che assumono solo valori positivi τ_3 è limitato in funzione di τ $2\tau - 1 \leq \tau_3 < 1$

PWM campionari

$$b_0 = \frac{1}{n}\sum_{j=1}^{n} x_{j:n}$$

$$b_1 = \frac{1}{n}\sum_{j=2}^{n} \frac{(j-1)}{(n-1)} x_{j:n}$$

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

$$b_r = \frac{1}{n}\sum_{j=r+1}^{n} \frac{(j-1)(j-2)\dots(j-r)}{(n-1)(n-2)\dots(n-r)} x_{j:n}$$

L-momenti campionari

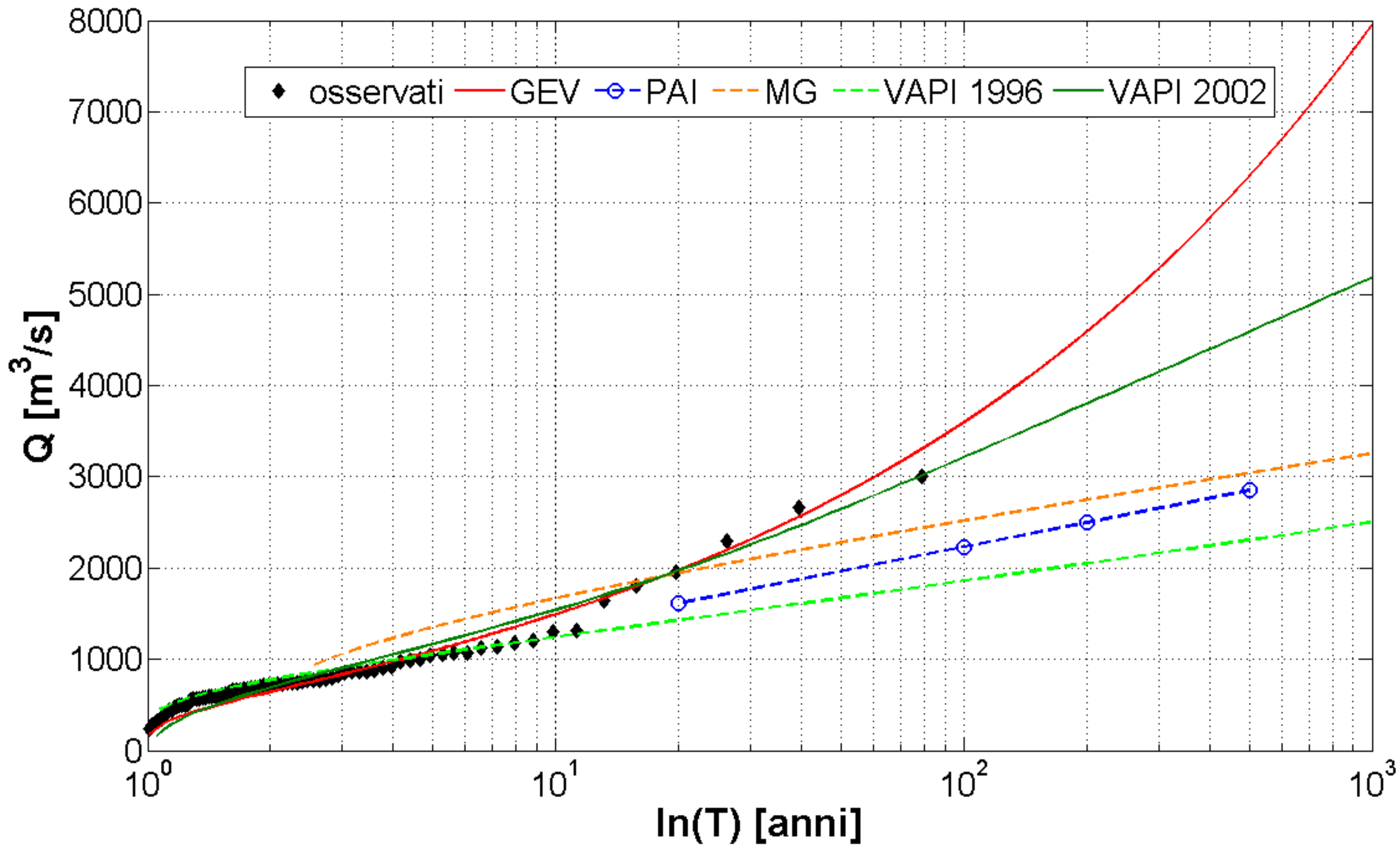
$$l_1 = b_0$$

$$l_2 = 2b_1 - b_0$$

$$l_3 = 6b_2 - 6b_1 + b_0$$

$$l_4 = 20b_3 - 30b_2 + 12b_1 - b_0$$

Distribuzioni degli estremi a 3 parametri



Per T abbastanza grande, in genere superiore ai 20 anni, vale: $\ln \frac{T}{T-1} \cong \frac{1}{T}$

Di conseguenza una X_T GUMBEL varia linearmente con Ln T (linea blu)

LEGGE GENERALIZZATA DEI VALORI ESTREMI (GEV)

L' espressione della GEV è :

$$F(x) = \exp\left[-\left(1-\theta_3\frac{(x-\theta_1)}{\theta_2}\right)^{1/\theta_3}\right] \qquad \theta_3 \neq 0$$

$$F(x) = \exp\left[-exp\left(-\frac{(x-\theta_1)}{\theta_2}\right)\right] \qquad \theta_3 = 0$$

θ_1 parametro di posizione

θ_2 parametro di scala

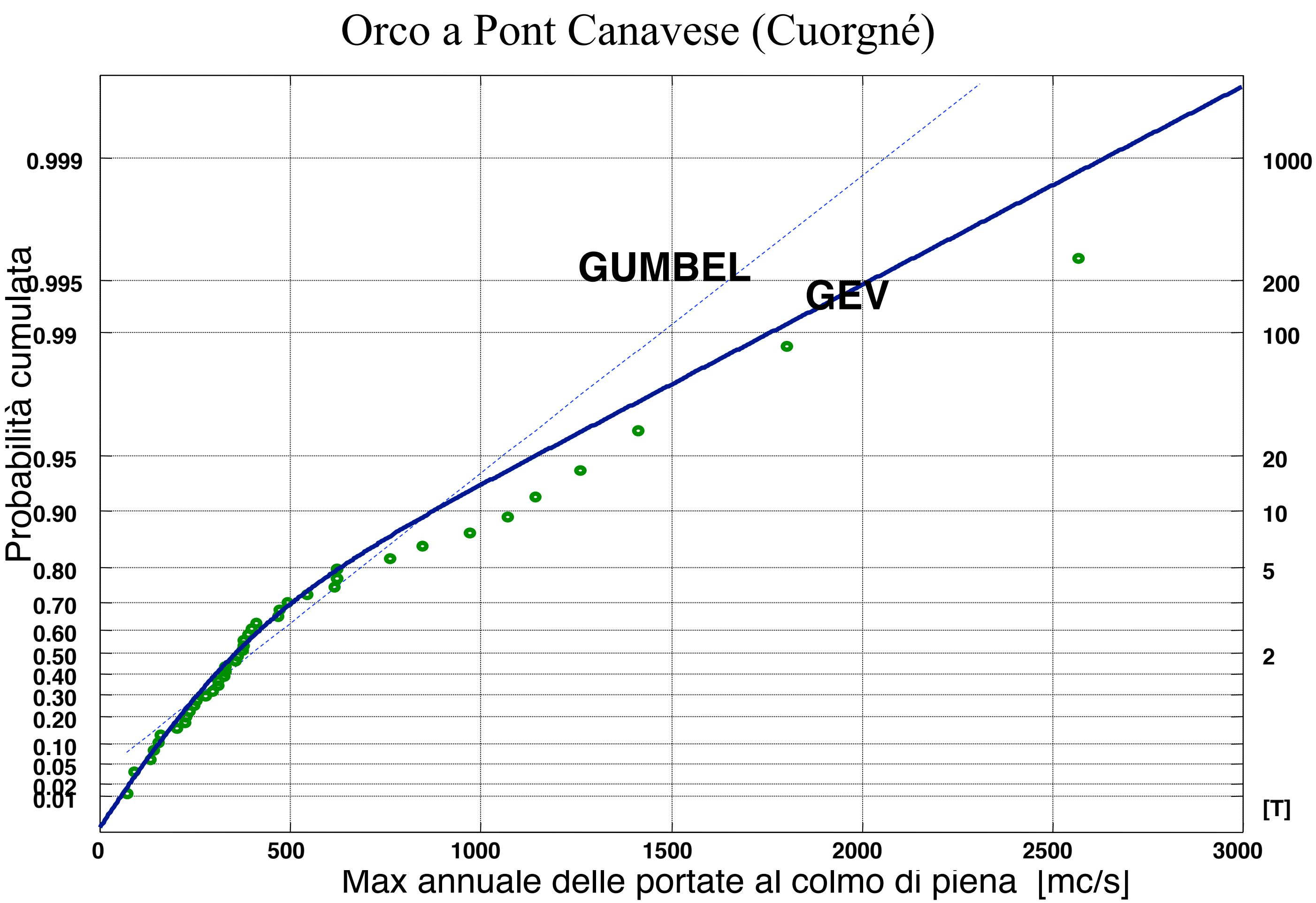
θ_3 parametro di forma

$\theta_3 = 0$ → *distribuzione EV I (Gumbel)*

$\theta_3 < 0$ → *EV II (Fréchet) - limitata sup. da* $\theta_1+\theta_2/\theta_3$

$\theta_3 > 0$ → *EV III (Weibull) - limitata inf. da* $\theta_1+\theta_2/\theta_3$

La GEV si adatta a serie con elevata asimmetria



Formulazioni per la stima della probabilità cumulata della popolazione dipendenti dalla distribuzione che si sta testando:

$$\phi(x_i) = P(\mu(x_i)) \qquad\qquad \phi(x_i) = \frac{i-\alpha}{n+1-2\alpha}$$

Si hanno ad esempio:

- Distribuzioni debolmente asimmetriche (*Cunnane*)

$$\phi(x_i) = \frac{i-0.4}{n+0.2} \qquad\qquad (\alpha = 0.4)$$

- Distribuzioni debolmente asimmetriche (*Gringorten*)

$$\phi(x_i) = \frac{i-0.44}{n+0.12} \qquad\qquad (\alpha = 0.44)$$

- Distribuzioni fortemente asimmetriche (*Hazen*)

$$\phi(x_i) = \frac{i-0.5}{n} \qquad\qquad (\alpha = 0.5)$$

Scelta della distribuzione

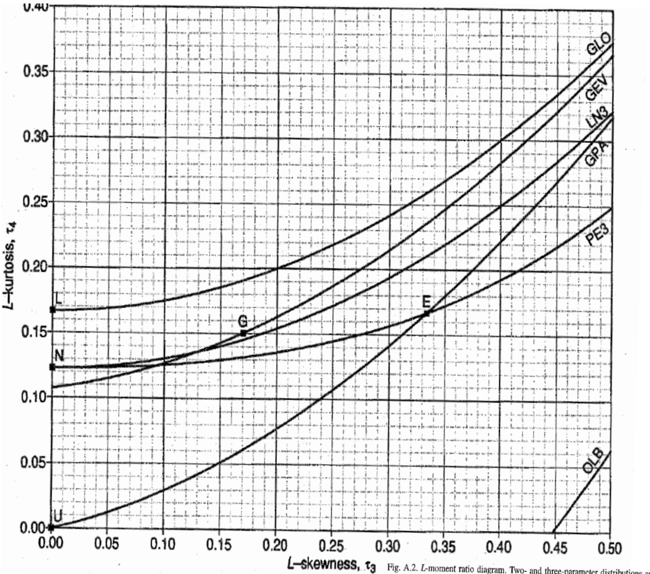
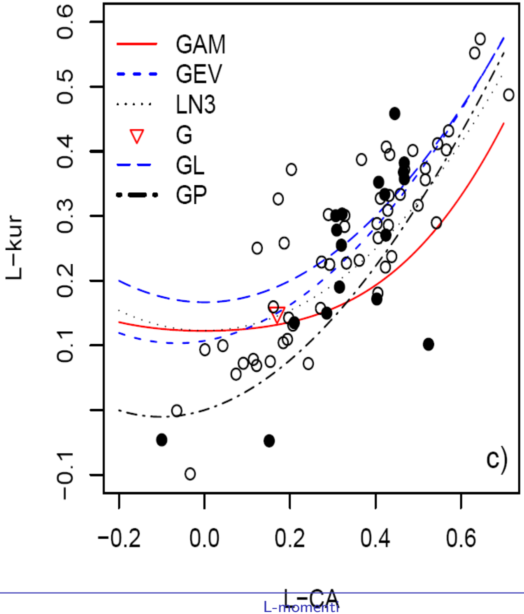


Fig. A.2. *L*-moment ratio diagram. Two- and three-parameter distributions are shown as points and lines, respectively. Key to distributions: E – exponential, G – Gumbel, L – logistic, N – Normal, U – uniform, GLO – generalized logistic, GEV – generalized extreme-value, GPA – generalized Pareto, LN3 – lognormal, and PE3 – Pearson type III. OLB is the overall lower bound of τ_4 as a function of τ_3, given by Eq. (2.45).

Scelta della distribuzione



Scelta della distribuzione