DISTRIBUZIONE EV1 o DI GUMBEL.

$$F_X(x) = \exp\left[-\Lambda e^{-\alpha x}\right] = \exp\left[-e^{-\alpha(x-\varepsilon)}\right]$$ *con:* $\exp(\alpha\varepsilon) = \Lambda$ (al posto di *n)*

Significato dei parametri:

- Λ: Numero medio di eventi indipendenti in $[0, t]$, ad esempio in un anno.
- $1/\alpha$:Valore medio della grandezza dell'evento, esempio portata al colmo.

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

$$E[x] = \mu = \varepsilon + \frac{0.57722}{\alpha} \qquad \varepsilon = \mu - 0.450\sigma$$

$$\mathrm{var}[x] = \sigma^2 = \frac{\pi^2}{6\alpha^2} \qquad \frac{1}{\alpha} = \frac{\sigma\sqrt{6}}{\pi}$$

$$Cv[x] = \frac{\sigma}{\mu} = \frac{\pi}{\sqrt{6}(\alpha\varepsilon + 0.57722)} = \frac{\pi}{\sqrt{6}(\ln\Lambda + 0.57722)}$$ dipende solo da Λ

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

$f_X(\varepsilon) = \alpha e^{-1} = \max \Rightarrow \varepsilon$ è la **moda** di x $\qquad F_X(\varepsilon) = e^{-1} = 0.368$

$\frac{1}{\alpha}$: è proporzionale a σ (misura di dispersione).

$(\alpha\varepsilon), \Lambda$: dipendono solo dal coefficiente di variazione Cv.

Variabile ridotta: $Y = \alpha(x - \varepsilon)$

$E[Y] = 0.57722$ *costante di Eulero* $\qquad var[Y] = \frac{\pi^2}{6}$

$\varepsilon[Y] = 0 \qquad \frac{1}{\alpha}[Y] = 1 \qquad \Lambda[Y] = 1$

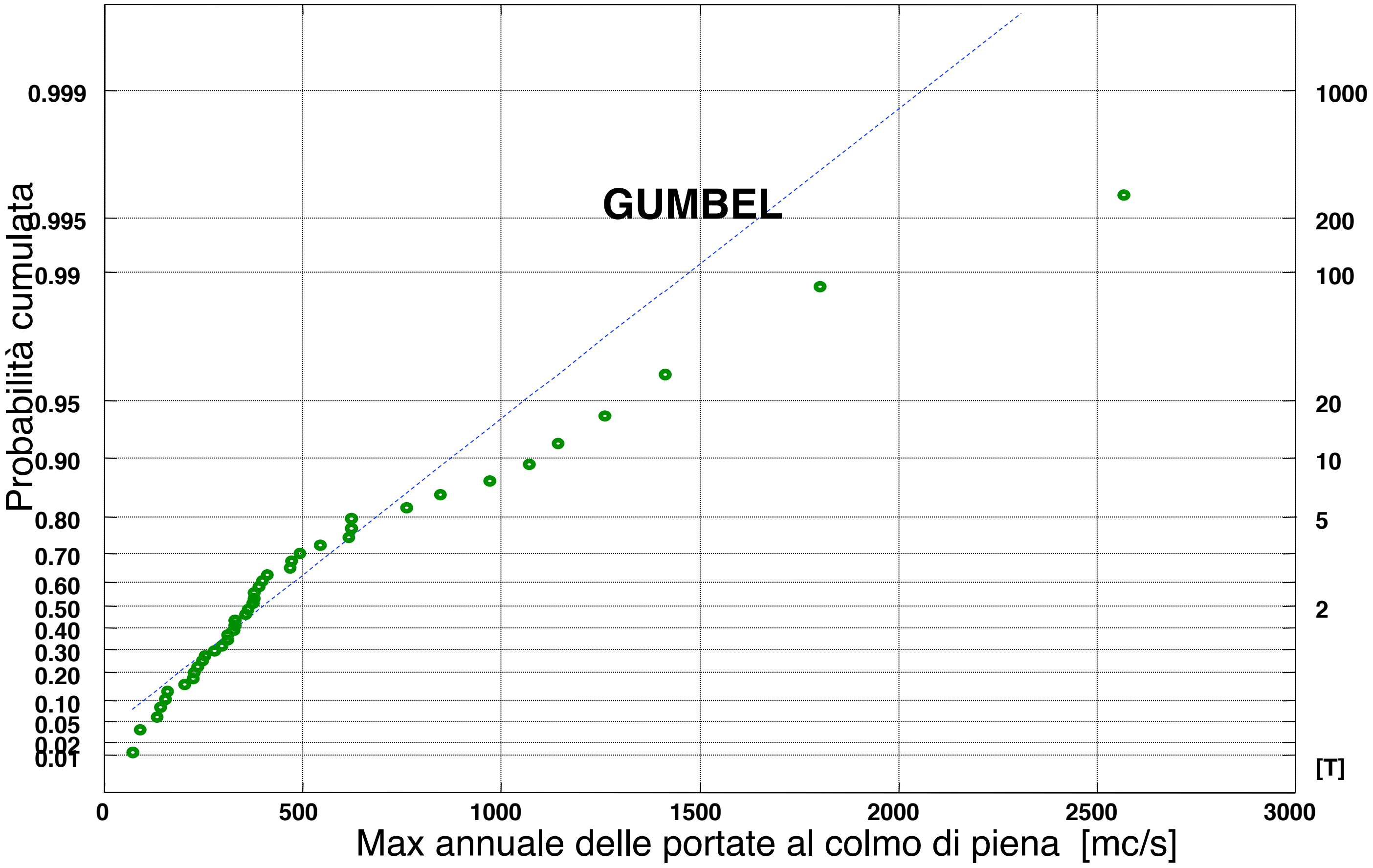
$\Phi(Y) = P[Y \le y] = \qquad F_Y(y) = \exp(-e^{-y})$

$Y \approx EV1(0,1) \qquad X \approx EV1\left(\varepsilon, \frac{1}{\alpha}\right)$

$Y_F = -\ln\ln\frac{1}{F} \qquad X_F = \varepsilon + \frac{Y_F}{\alpha} = \quad \varepsilon - \frac{1}{\alpha}\ln\ln\frac{1}{F} \quad = \frac{1}{\alpha}(\ln\Lambda - \ln\ln\frac{1}{F})$

(***Quantile della distribuzione per assegnato F***)

Orco a Pont Canavese (Cuorgné)



FORMA CARATTERISTICA

$X_T = \varepsilon\left(1 - k'\log\ln\frac{T}{T-1}\right)$ con $k' = \frac{1}{0{,}4343\alpha\varepsilon}$ (caratteristica della distribuzione)

Per T abbastanza grande, in genere superiore ai 20 anni, vale che $\ln\frac{T}{T-1} \cong \frac{1}{T}$;
Di conseguenza X_T può essere espressa come:

$X_T \cong \varepsilon(1 + k'\log T)$ dove $k' = \frac{1}{\log\Lambda}$ => **X_T varia linearmente con Log T**

LEGGE DEI VALORI ESTREMI A 3 PARAMETRI (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* $\theta_1+\theta_2/\theta_3$

1.2 Stimatori

Nella pratica spesso si assume che la forma di una qualche distribuzione di probabilità sia conosciuta a medo di un set di parametri incogniti $\theta_1, \ldots, \theta_p$. Sia $x(u; \theta_1, \ldots, \theta_p)$ la funzione dei quantili di una distribuzione con p parametri incogniti. In molte applicazioni tra i parametri incognti si possono identificare un parametro di posizione ed un parametro di scala. Un parametro ξ di una distribuzione è un *parametro di posizione* se per la funzione dei quantili vale l'eguaglianza

$$x(u; \xi, \theta_2, \ldots, \theta_p) = \xi + x(u; 0, \theta_2, \ldots, \theta_p) \ . \tag{1.9}$$

Si dice, invece, che α è un *parametro di scala* della funzione dei quantili della distribuzione se

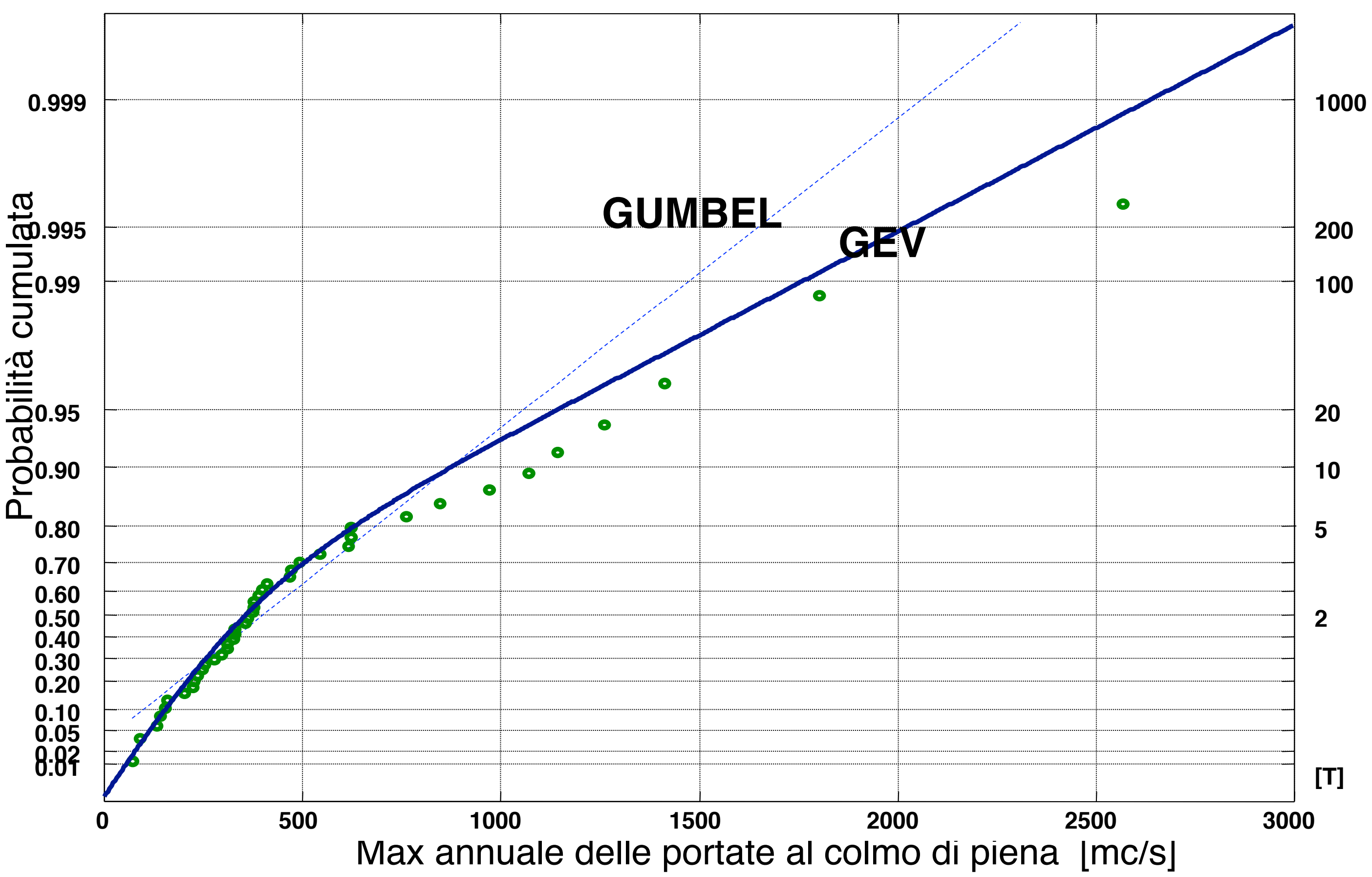
$$x(u; \alpha, \theta_2, \ldots, \theta_p) = \alpha \times x(u; 1, \theta_2, \ldots, \theta_p) \ . \tag{1.10}$$

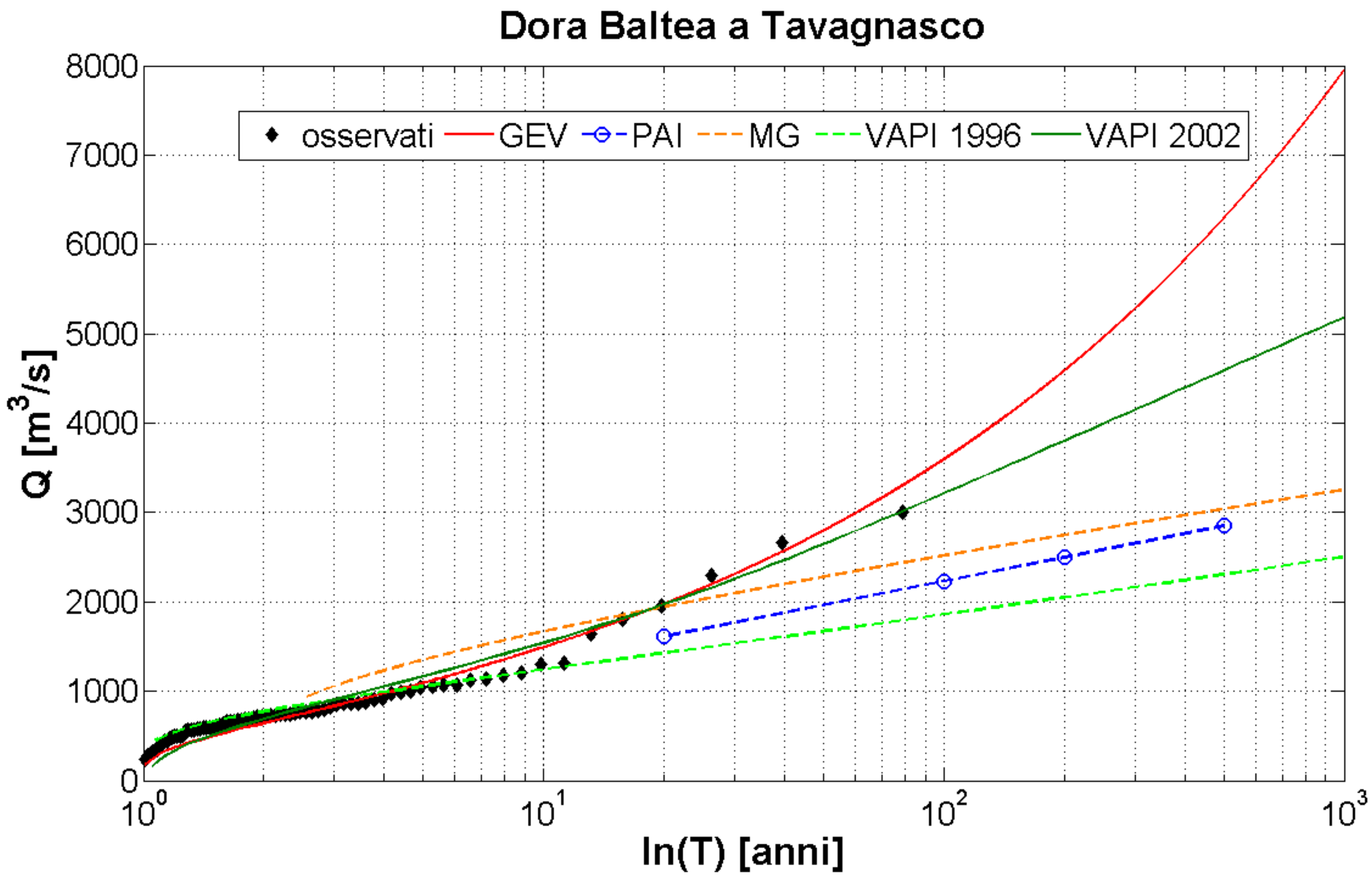
Se per la distribuzione esistono entrambi questi parametri, allora vale l'eguaglianza

$$x(u; \xi, \alpha, \theta_3, \ldots, \theta_p) = \xi + \alpha \times x(u; 0, 1, \theta_3, \ldots, \theta_p) \ . \tag{1.11}$$

I parametri incogniti sono stimati a partire dai dati osservati. Dato un set di osservazioni, una funzione $\hat{\theta}$ di queste deve essere scelta come *stimatore* di θ. Lo stimatore $\hat{\theta}$ è a sua volta una variabile casuale ed ha una distribuzione di probabilità. La bontà di $\hat{\theta}$ come stimatore di θ tipicamente dipende da quanto $\hat{\theta}$ si avvicina a θ. La deviazione di $\hat{\theta}$ da θ può essere scomposta

Orco a Pont Canavese (Cuorgné)





Dora Baltea a Tavagnasco
osservati
GEV
PAI
MG
VAPI 1996
VAPI 2002
Q [m3/s]
ln(T) [anni]
0
1000
2000
3000
4000
5000
6000
7000
8000
10^0
10^1
10^2
10^3

1.2 Stimatori

Nella pratica spesso si assume che la forma di una qualche distribuzione di probabilità sia conosciuta a medo di un set di parametri incogniti $\theta_1, \ldots, \theta_p$. Sia $x(u; \theta_1, \ldots, \theta_p)$ la funzione dei quantili di una distribuzione con p parametri incogniti. In molte applicazioni tra i parametri incognti si possono identificare un parametro di posizione ed un parametro di scala. Un parametro ξ di una distribuzione è un *parametro di posizione* se per la funzione dei quantili vale l'eguaglianza

$$x(u; \xi, \theta_2, \ldots, \theta_p) = \xi + x(u; 0, \theta_2, \ldots, \theta_p) \ . \tag{1.9}$$

Si dice, invece, che α è un *parametro di scala* della funzione dei quantili della distribuzione se

$$x(u; \alpha, \theta_2, \ldots, \theta_p) = \alpha \times x(u; 1, \theta_2, \ldots, \theta_p) \ . \tag{1.10}$$

Se per la distribuzione esistono entrambi questi parametri, allora vale l'eguaglianza

$$x(u; \xi, \alpha, \theta_3, \ldots, \theta_p) = \xi + \alpha \times x(u; 0, 1, \theta_3, \ldots, \theta_p) \ . \tag{1.11}$$

I parametri incogniti sono stimati a partire dai dati osservati. Dato un set di osservazioni, una funzione $\hat{\theta}$ di queste deve essere scelta come *stimatore* di θ. Lo stimatore $\hat{\theta}$ è a sua volta una variabile casuale ed ha una distribuzione di probabilità. La bontà di $\hat{\theta}$ come stimatore di θ tipicamente dipende da quanto $\hat{\theta}$ si avvicina a θ. La deviazione di $\hat{\theta}$ da θ può essere scomposta in *distorsione* (tendenza di dare stime sistematicamente più alte, o più basse, del valore vero) e *variabilità* (deviazione casuale dal valore vero, che si verifica anche per gli stimatori che non presentano distorsione).

La performance di uno stimatore $\hat{\theta}$ può essere valutata con due misure, il *bias* ("distorsione") e la *radice dell'errore quadratico medio* (RMSE), definite come

$$\text{bias}(\hat{\theta}) = \mathrm{E}(\hat{\theta} - \theta) \ , \quad \text{RMSE}(\hat{\theta}) = \{\mathrm{E}(\hat{\theta} - \theta)^2\}^{1/2} \ , \tag{1.12}$$

e caratterizzate dall'avere la stessa unità di misura del parametro θ. Si dice che lo stimatore $\hat{\theta}$ è *indistorto* se $\text{bias}(\hat{\theta}) = 0$ ovvero se $\mathrm{E}(\hat{\theta}) = \theta$. Diversi stimatori indistorti dello stesso parametro possono essere paragonati in termini della loro varianza: il rapporto $\text{var}(\hat{\theta}^{(1)})/\text{var}(\hat{\theta}^{(2)})$ si dice *efficienza* dello stimatore $\hat{\theta}^{(2)}$ rispetto allo stimatore $\hat{\theta}^{(1)}$. La radice dell'errore quadratico medio può essere anche scritta come

$$\text{RMSE}(\hat{\theta}) = [\{\text{bias}(\hat{\theta})\}^2 + \text{var}(\hat{\theta})]^{1/2} \ , \tag{1.13}$$

da cui si vede come l'RMSE combina distorsione e variabilità di $\hat{\theta}$ e dà una misura globale dell'accuratezza della stima. Nei classici problemi di statistica in cui la stima dei parametri è basata su un campione di lunghezza n, sia il bias che la varianza di $\hat{\theta}$ sono asintoticamente proporzionali a n^{-1} per n grandi (v.es. **?**), per cui l'RMSE di $\hat{\theta}$ è proporzionale a $n^{-1/2}$.

STIMA DEI PARAMETRI.

Esistono diversi metodi per la stima dei parametri:

- Metodo grafico di interpolazione su carta probabilistica.
- Metodo dei momenti.
- Metodo della massima verosimiglianza.
- Metodo degli L-Momenti.

METODO DEI MOMENTI.

Consiste semplicemente nell'eguagliare i momenti della popolazione ai momenti della serie.

Poiché i momenti della popolazione sono funzione dei parametri, basta considerare tanti momenti quanti sono i parametri da stimare e risolvere un sistema di r equazioni in r incognite (r: numero dei parametri da stimare) dando ovviamente preferenza ai momenti di ordine minore.

1. DISTRIBUZIONE NORMALE.

Due parametri da stimare μ e σ.

$$E[x] = \mu$$

$$\mathrm{var}[x] = \sigma^2$$

A partire dai valori $(X_1, X_2, \ldots\ldots X_n)$ della serie si calcola:

- $\bar{x} = \dfrac{\sum x_i}{n}$
- $s^2 = \dfrac{\sum (x_i - \bar{x})^2}{n-1}$

Sistema di due equazioni in due incognite:

$\hat{\mu} = \bar{x}$ $\qquad\qquad$ $\hat{\sigma}^2 = s^2$

2. **DISTRIBUZIONE LOG – NORMALE.**

$Y = \ln X$

Due parametri: μ_y e σ_y

Due momenti della popolazione: μ_X e σ_X^2:

- $$\mu_Y = \ln \mu_X - \frac{1}{2}\sigma_Y^2$$
- $$\sigma_Y^2 = \ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right)$$

Dalla serie di dati $(X_1, X_2, \ldots\ldots X_n)$:

- $$\bar{x} = \frac{\sum x_i}{n} = \hat{\mu}_X$$
- $$s_X^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \hat{\sigma}_x^2$$

Sistema di due equazioni in due incognite:

$$\hat{\mu}_Y = \ln \bar{x} - \frac{1}{2}\hat{\sigma}_Y^2 \qquad \hat{\sigma}_y^2 = \ln\left(1 + \frac{s_x^2}{\bar{x}^2}\right)$$

DISTRIBUZIONE EV1 o DI GUMBEL.

Due parametri da stimare: ε ed α

Due momenti della popolazione:

$$E[x] = \mu = \varepsilon + \frac{0.57722}{\alpha}$$

$$\mathrm{var}[x] = \sigma^2 = \frac{\pi^2}{6\alpha^2}$$

Dal sistema di 2 equazioni in 2 incognite:

$$\varepsilon = \mu - 0.450\sigma$$

$$\frac{1}{\alpha} = \frac{\sigma\sqrt{6}}{\pi}$$

METODO DELLA MASSIMA VEROSIMIGLIANZA.

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

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

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

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

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

Distribuzione normale:

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

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

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

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

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

Distribuzione di Gumbel

$$\frac{1}{\alpha} = x - \frac{\sum_{i=1}^{n} x_i \cdot e^{-\alpha \cdot x_i}}{\sum_{i=1}^{n} e^{-\alpha \cdot x_i}} \qquad e^{-\alpha \cdot \varepsilon} = \frac{1}{n} \cdot \sum_{i=1}^{n} e^{-\alpha \cdot x_i}$$

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 α).

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: $\alpha_3 = \alpha_2 + (\alpha_1 - \alpha_2)/3$

Momenti pesati in probabilità

(Probability Weighted Moments - PWM)

Assunto: $F(x(u)) = u$

Si ha: $$\mathrm{E}(X) = \int_{-\infty}^{\infty} x dF(x) = \int_{-\infty}^{\infty} x f(x) dx$$

ovvero: $$\mathrm{E}(X) = \int_0^1 x(u) du \ .$$

Essendo $x(u)$ il quantile della variabile x in corrispondenza della probabilità cumulata u, il PWM di ordine r è definito come:

$$\beta_r = \int_0^1 x(u) u^r du$$

Si ha pertanto: $\beta_0 = E(X)$

Lo stimatore campionario di β_r, detto b_r è:

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

dove $x_{j:n}$ è il j-esimo elemento del campione ordinato in senso crescente.

L-momenti

Gli L-momenti sono combinazioni lineari di PWM e sono statistiche basate su combinazioni lineari dei dati.

$$l_1 = b_0 \ , \quad = \text{media campionaria}$$
$$l_2 = 2b_1 - b_0 \ ,$$
$$l_3 = 6b_2 - 6b_1 + b_0 \ ,$$
$$l_4 = 20b_3 - 30b_2 + 12b_1 - b_0$$

Di particolare interesse sono i rapporti adimensionali di L-momenti:

- L-CV o $\tau = l_2/l_1$
- L-CA o L-skewness o $\tau_3 = l_3/l_2$
- L-kurtosis o $\tau_4 = l_4/l_2$