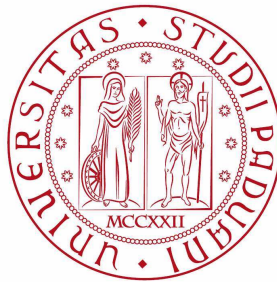


UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI FISICA E ASTRONOMIA
"Galileo Galilei"

CORSO DI LAUREA IN FISICA



TESI DI LAUREA

**Amplificazione esponenziale di
fluttuazioni esponenzialmente rare:
un meccanismo universale**

Relatore: Prof. ANTONIO TROVATO

Laureando: ROBERTO SARTORI

ANNO ACCADEMICO 2015/2016

Indice

1	Introduzione	3
1.1	Distribuzione di Gumbel	4
2	Amplificazione esponenziale	6
3	Introduzione ai test numerici	9
4	Studio del limite asintotico	11
4.1	Parametro di scala β	11
4.2	Parametro di posizione α	12
5	Confronto tra i risultati dei due metodi	18
5.1	Regime asintotico	18
5.2	Regime non asintotico	19
6	Conclusioni	21

1 Introduzione

La teoria dei valori estremi è una branca della statistica che si occupa di analizzare e sviluppare modelli che descrivono comportamenti estremi. Formalmente si occupa di studiare le deviazioni estreme dalla mediana di una distribuzione di probabilità: in linea generale, quindi, analizza il comportamento delle code della distribuzione di probabilità di un particolare fenomeno. Di questa teoria si trovano applicazioni in molti ambiti, tra i quali la finanza, l'idrologia o la meteorologia [1]. Oltre a questi, negli ultimi anni la statistica dei valori estremi ha suscitato interesse in quanto in grado di descrivere con le sue leggi fenomeni nei quali, apparentemente, non si presentano comportamenti estremi. Un caso particolare, che verrà analizzato in seguito, è quello della bioluminescenza di colonie di batteri [2].

Innanzitutto saranno introdotte le basi della teoria dei valori estremi, per poi analizzare in particolare la distribuzione di Gumbel.

Siano $X_1, X_2, \dots, X_N$ variabili casuali indipendenti e identicamente distribuite secondo una distribuzione con cumulativa $G(x)$. Sia quindi $M_N = \max\{X_1, \dots, X_N\}$. M_N sarà distribuito come:

$$\begin{aligned} \Pr\{M_N \leq x\} &= \Pr\{X_1 \leq x, \dots, X_N \leq x\} \\ &= \Pr\{X_1 \leq x\} \cdot \dots \cdot \Pr\{X_N \leq x\} \\ &= [G(x)]^N \end{aligned} \tag{1.1}$$

Tuttavia, non conoscendo $G(x)$, non è possibile stimare questa probabilità.

Si può comunque studiare cosa succede con $N \rightarrow \infty$. In questo caso si può utilizzare il teorema dei valori estremi: se esistono sequenze di costanti a_N e $b_N > 0$ tali che, con $N \rightarrow \infty$:

$$\Pr\left\{\frac{M_N - a_N}{b_N} \leq x\right\} \rightarrow P(x) \tag{1.2}$$

allora $P(x)$ è dello stesso tipo di una delle seguenti:

$$\text{Gumbel} \quad I : P_G(x) = e^{-e^{-x}} \quad -\infty < x < +\infty \quad (1.3)$$

$$\text{Fréchet} \quad II : P_F(x) = \begin{cases} 0 & x \leq 0 \\ e^{-(x)^{-\gamma}} & x > 0, \gamma > 0 \end{cases} \quad (1.4)$$

$$\text{Weibull} \quad III : P_W(x) = \begin{cases} e^{-(-x)^\gamma} & x < 0, \gamma > 0 \\ 1 & x \geq 0. \end{cases} \quad (1.5)$$

In questa tesi, ci si occuperà del caso I .

1.1 Distribuzione di Gumbel

La distribuzione di Gumbel ha PDF (*Probability Density Function*) data da:

$$p_G(x) = \frac{1}{\beta} (e^{-\frac{x-\alpha}{\beta}} e^{-e^{-\frac{x-\alpha}{\beta}}}) \quad (1.6)$$

definita su tutto $\mathbb{R}$, e ha come CDF (*Cumulative Distribution Function*) la funzione di Gompertz, definita da:

$$P_G(x) = e^{-e^{-\frac{x-\alpha}{\beta}}} \quad (1.7)$$

tale che:

$$\frac{dP_G(x)}{dx} = p_G(x) \quad (1.8)$$

$$P_G(x) \xrightarrow{x \rightarrow -\infty} 0 \quad (1.9)$$

$$P_G(x) \xrightarrow{x \rightarrow +\infty} 1. \quad (1.10)$$

Affinché sia una CDF, inoltre, è richiesto che la (1.7) sia monotona crescente e continua a destra. La Gompertz è comunque continua in $\mathbb{R}$.

I parametri α e β sono reali con l'unica condizione $\beta > 0$: il primo è il parametro di posizione, il secondo il parametro di scala. α rappresenta inoltre la moda, mentre la media è data da $\alpha + \beta\gamma$, con $\gamma \approx 0.57..$ costante di Eulero.

In particolare, i massimi di una specifica classe di distribuzioni definita da funzioni con coda esponenziale si distribuiscono come una Gumbel. Denotando le CDF di queste distribuzioni con $G(x)$ si intende quindi:

$$G(x) \stackrel{x \gg 1}{\approx} 1 - \lambda e^{-\frac{(x-\mu)}{s}} \quad (1.11)$$

con λ e $\tilde{s}$ reali positivi.

Infatti (per semplicità $\mu = 0$ e $\lambda = \tilde{s} = 1$):

$$\begin{aligned}
 [G(x + \ln(N))]^N &= (1 - e^{-(x + \ln(N))})^N \\
 &= \left(1 - \frac{1}{N}e^{-x}\right)^N \\
 &\xrightarrow{N \rightarrow +\infty} e^{-e^{-x}}
 \end{aligned} \tag{1.12}$$

che è una Gompertz.

Formalmente si richiede che $G(x)$ sia definita su tutto $\mathbb{R}$ ma in pratica si osserva una Gumbel anche nel caso ci sia un limite inferiore finito.

Fatte queste ipotesi, in tutta generalità i parametri della Gumbel che descrive la distribuzione dei massimi di N campioni della (1.11) sono dati da [3]:

$$\beta = \lim_{N \rightarrow +\infty} b_N = \tilde{s} \tag{1.13}$$

$$\alpha = \lim_{N \rightarrow +\infty} a_N = \mu + \tilde{s} \cdot \ln(N) + \tilde{s} \cdot \ln(\lambda). \tag{1.14}$$

2 Amplificazione esponenziale

Come accennato nell'introduzione, un caso interessante dove si è osservata una Gumbel è quello relativo alla bioluminescenza di colonie di batteri [2]. Brevemente, nel lavoro citato è stata misurata l'energia emessa sotto forma di radiazione nell'unità di tempo: moltiplicando questa per il tempo di acquisizione si ottiene una stima del numero di fotoni emessi dal campione di batteri in analisi.

Viene riportato in figura (1) il fit dei dati sperimentali relativi:

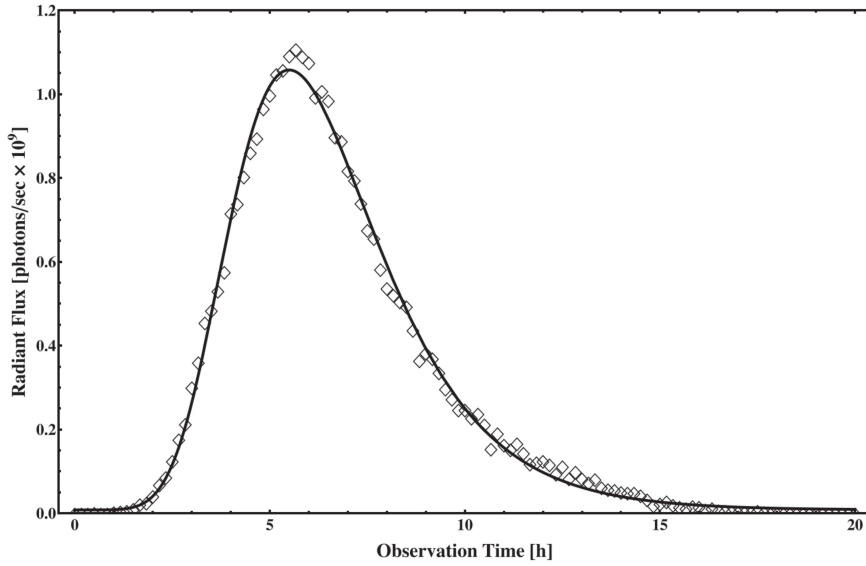


Figura 1: Fit con una Gumbel dei dati sull'emissione bioluminescente [2]

Si può supporre che l'energia in funzione del tempo possa essere scritta come:

$$E(t) \propto n(t) = n_0 \cdot 10^{K e^{-e^{[1+e \cdot (\lambda_m - t)] \frac{\mu m}{K}}}} \quad (2.1)$$

$$= n_0 \cdot 10^{K \cdot G_b(t)} \propto e^{K \cdot G_b(t)} \quad (2.2)$$

dove $n(t)$ è la concentrazione di batteri nella colonia in funzione del tempo, e gli altri termini sono costanti non importanti per questa analisi, eccetto K : questo controlla l'amplificazione esponenziale delle colonie batteriche raggiunta asintoticamente per $t \rightarrow \infty$.

La $G_b(t)$ introdotta nella (2.2) è in effetti una Gompertz [4] ma ciò che è interessante studiare è il caso generale dell'amplificazione esponenziale di una generica $G(x)$ che rispetta la (1.11), i cui massimi quindi sono distribuiti come una Gumbel.

Per farlo, conviene innanzitutto riscrivere $n(t)$ come una generica $F_k(x)$, del tipo:

$$F_k(x) = \frac{e^{kG(x)} - 1}{e^k - 1} \quad (2.3)$$

Gli altri termini della (2.3) sono necessari affinché, se $G(x)$ è una CDF con le condizioni del paragrafo (1), allora lo sia anche la rispettiva $F_k(x)$:

$$G(x) \rightarrow 1 \Rightarrow F_k(x) \rightarrow 1 \quad (2.4)$$

$$G(x) \rightarrow 0 \Rightarrow F_k(x) \rightarrow 0. \quad (2.5)$$

Inoltre, con $G(x)$ monotona crescente lo è anche $F_k(x)$.

L'ipotesi da verificare è che i campioni delle PDF corrispondenti $f_k(x) = \frac{dF_k(x)}{dx}$ si distribuiscano effettivamente come una Gumbel per $k \gg 1$. Si può vedere questo sfruttando una analogia con il metodo della trasformata inversa: quest'ultimo si usa per campionare una generica distribuzione $g(x) = \frac{dG(x)}{dx}$. Se y è una variabile casuale uniforme continua in $[0, 1]$ allora $G^{-1}(y)$ è distribuita secondo $g(x)$.

Nel nostro caso partendo dalla (2.3) e ponendo $F_k(x) = y$:

$$G(x) = \frac{\ln[y \cdot (e^k - 1) + 1]}{k}. \quad (2.6)$$

In pratica, verrà usata la (2.6) per campionare le $f_k(x)$ e verificare il loro comportamento nel limite $k \gg 1$, dato $y \in [0, 1]$.

Per k abbastanza grande si può approssimare la (2.6) a:

$$G(x) \stackrel{k \gg 1}{\approx} 1 + \frac{\ln(y)}{k}. \quad (2.7)$$

È possibile arrivare a questo risultato anche partendo dalla (1.1), usando k al posto di N :

$$[G(x)]^k = \tilde{F}(x) = y \quad (2.8)$$

$$\Rightarrow k \cdot \ln(G(x)) = \ln(y) \quad (2.9)$$

$$\Rightarrow G(x) = e^{\left(\frac{\ln(y)}{k}\right)} \quad (2.10)$$

$$\xrightarrow{k \gg 1} G(x) \approx 1 + \frac{\ln(y)}{k}.$$

Perciò i valori campionati saranno del tipo:

$$x \stackrel{k \gg 1}{\approx} G^{-1} \left(1 + \frac{\ln(y)}{k} \right). \quad (2.11)$$

Con $k \gg 1$:

$$1 + \frac{\ln(y)}{k} \lesssim 1 \quad (2.12)$$

$$\Rightarrow x \stackrel{k \gg 1}{\approx} G^{-1} \left(1 + \frac{\ln(y)}{k} \right) \rightarrow +\infty \quad (2.13)$$

dove nell'ultimo passaggio si è usata la proprietà delle CDF data dalla (1.10). Essendo quindi nella condizione $x \gg 1$ e sapendo che tutte le $G(x)$ in analisi rispettano la (1.11) si può usare quella approssimazione, che è facilmente invertibile con:

$$x_e = G^{-1}(y_e) = \mu - \tilde{s} \cdot \ln \left(\frac{1 - y_e}{\lambda} \right). \quad (2.14)$$

Per ottenere ora gli x distribuiti come le $F_k(x)$, non resta che porre:

$$y_e = 1 + \frac{\ln(y)}{k} \quad (2.15)$$

ottenendo da (2.14):

$$x = \mu - \tilde{s} \cdot \ln \left(\frac{1 - 1 - \frac{\ln(y)}{k}}{\lambda} \right) = \mu - \tilde{s} \cdot \ln \left(-\frac{\ln(y)}{k\lambda} \right) \quad (2.16)$$

$$\Rightarrow x = \mu - \tilde{s} \cdot \ln[-\ln(y)] + \tilde{s} \cdot \ln(k) + \tilde{s} \cdot \ln(\lambda). \quad (2.17)$$

È necessario ora invertire la (1.7) per vedere la forma dei valori distribuiti come una Gumbel tramite il metodo della trasformata inversa, essendo appunto la Gompertz la CDF della Gumbel:

$$x_g = -\beta \cdot \ln[-\ln(y)] + \alpha. \quad (2.18)$$

I valori x che si distribuiscono seguendo le $F_k(x)$ hanno lo stesso andamento degli x_g , perciò seguono effettivamente una Gumbel, con:

$$\beta = \tilde{s} \quad (2.19)$$

$$\alpha = \mu + \tilde{s} \cdot \ln(k) + \tilde{s} \cdot \ln(\lambda). \quad (2.20)$$

In definitiva, lo scopo di questo lavoro di tesi è di verificare con test numerici che $f_k(x)$ per $k \gg 1$ segue una Gumbel con α e β dati dalle (2.19) e (2.20) e confrontare i modi in cui si raggiunge il regime asintotico rispetto al metodo 'standard' dei massimi.

3 Introduzione ai test numerici

Per l'analisi numerica, sono stati generati numeri pseudocasuali distribuiti secondo le distribuzioni di prova. Le $G(x)$ considerate sono invertibili, perciò è stato possibile sfruttare il metodo della trasformata inversa per campionare le relative distribuzioni. Lo stesso vale per le $F_k(x)$, dove comunque si è usata la (2.6) conoscendo già le inverse delle $G(x)$.

In particolare, si è studiato il comportamento delle seguenti funzioni (e relative $F_k(x)$) caratterizzate da due parametri reali, μ ed $s > 0$:

$$\text{Esponenziale, } G(x) = 1 - e^{-\frac{x-\mu}{s}} \quad (3.1)$$

$$\text{Tanh, } G(x) = \frac{\tanh(\frac{x-\mu}{s}) + 1}{2} \quad (3.2)$$

$$\text{Laplace, } G(x) = \begin{cases} 1 - \frac{1}{2}e^{-\frac{x-\mu}{s}}, & \text{se } x \geq \mu \\ \frac{e^{\frac{x-\mu}{s}}}{2}, & \text{se } x < \mu \end{cases} \quad (3.3)$$

$$\text{Sigmoide, } G(x) = \frac{1}{1 + e^{-\frac{x-\mu}{s}}} \quad (3.4)$$

$$\text{Gompertz, } G(x) = e^{-e^{-\frac{x-\mu}{s}}}. \quad (3.5)$$

Tutte queste distribuzioni hanno PDF definita in $\mathbb{R}$ eccetto la (3.1) (esponenziale), che è invece definita in $[\mu, +\infty[$.

Con $x \gg 1$, si possono poi ricavare i termini λ e $\tilde{s}$ per la (1.11) nei vari casi. Si ha $\lambda = 1$ in tutti i casi eccetto per Laplace, per il quale $\lambda = \frac{1}{2}$.

Invece, si ha $\tilde{s} = s$ per tutte le distribuzioni eccetto la tangente iperbolica, dove si ottiene $\tilde{s} = \frac{s}{2}$. Infatti sfruttando l'esponenziale:

$$\tanh\left(\frac{x-\mu}{s}\right) = \frac{1 - e^{\frac{2(x-\mu)}{s}}}{1 + e^{\frac{2(x-\mu)}{s}}} \stackrel{x \rightarrow +\infty}{\approx} 1 - 2e^{-\frac{2(x-\mu)}{s}} + o\left(e^{-\frac{2(x-\mu)}{s}}\right), \quad (3.6)$$

ovvero $\tilde{s} = \frac{s}{2}$.

Come primo test, si è verificato che effettivamente sia i massimi di N campioni delle $G(x)$ sia i valori dei singoli campioni delle $F_k(x)$ fossero distribuiti come una Gumbel. Ciò è risultato vero per le $G(x)$ per $N \gg 1$ (figura (2)) e per le $F_k(x)$ nel caso $k \gg 1$ (figura (3)):

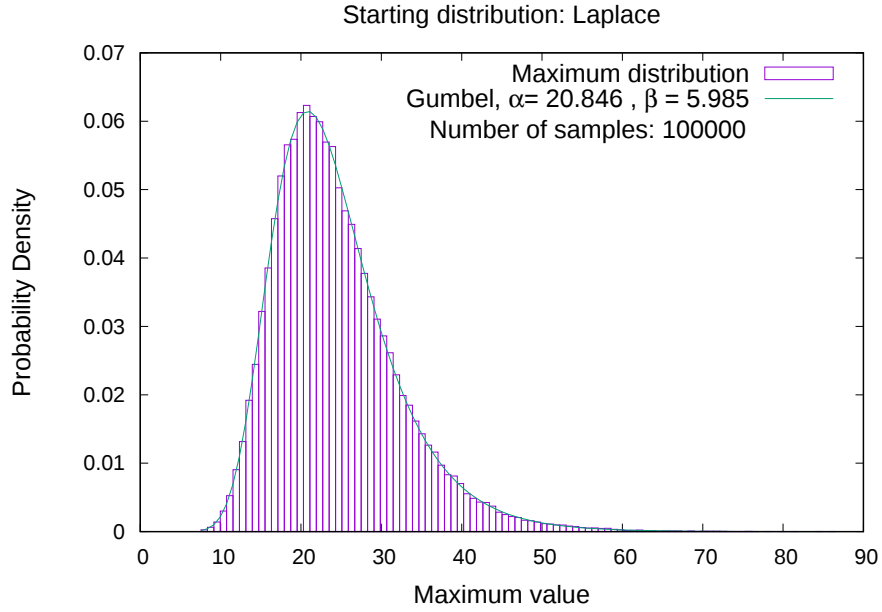


Figura 2: Gumbel ottenuta dai massimi di $G(x)$, $s = \mu = 6$, $N = 200$

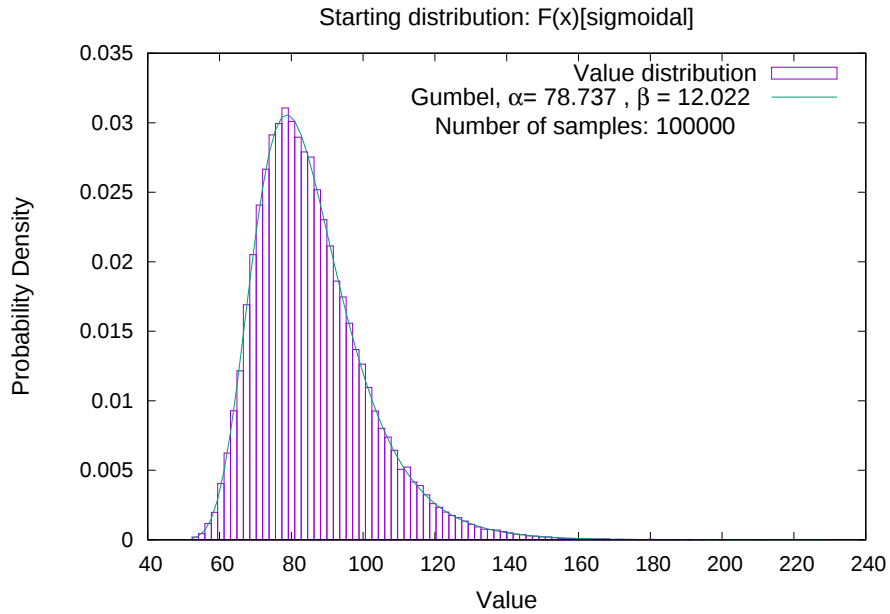


Figura 3: Gumbel ottenuta da $F_k(x)$, $s = 12$, $\mu = 6$, $k = 500$

Aumentando il numero di campionamenti delle distribuzioni considerate migliora la stima dei parametri. Questo porta però ad un costo computazionale maggiore, soprattutto nel caso dei massimi: anche in questo caso, comunque, restando nell'ordine delle centinaia di migliaia di campioni di $G(x)$ generati si ha una buona stima per gli scopi di questo lavoro.

4 Studio del limite asintotico

4.1 Parametro di scala β

Come previsto dalla (1.13) e dalla (2.19), il parametro β della Gumbel risulta essere in tutti i casi simile al parametro $\tilde{s}$ determinato da (1.11). Perciò risulta $\beta \approx s$ in generale, mentre per la (3.2) (tangente) si ha $\beta \approx \frac{s}{2}$.

Ciò è valido solo per k ed N abbastanza grandi. La stima infatti migliora all'aumentare di questi:

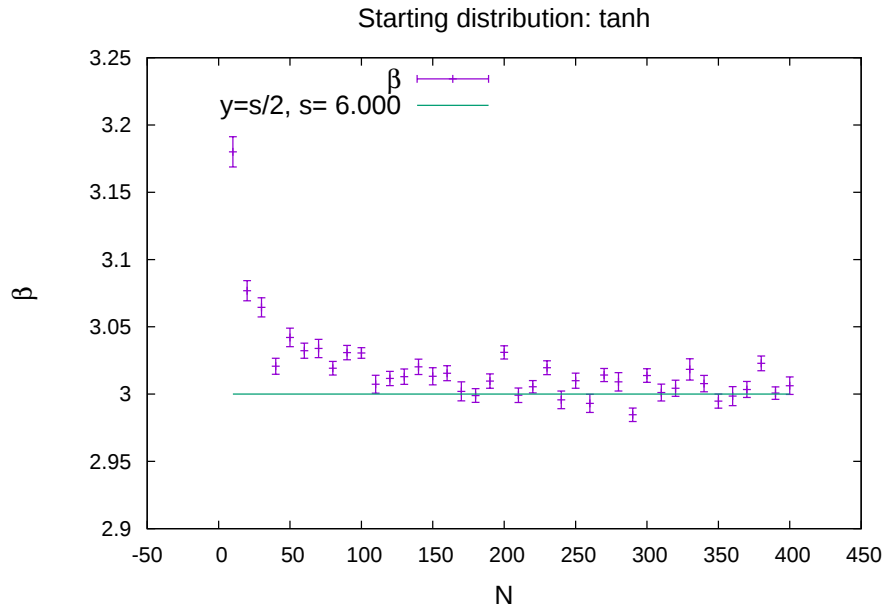


Figura 4: β in funzione di N , $s = 6$, $\mu = 6$

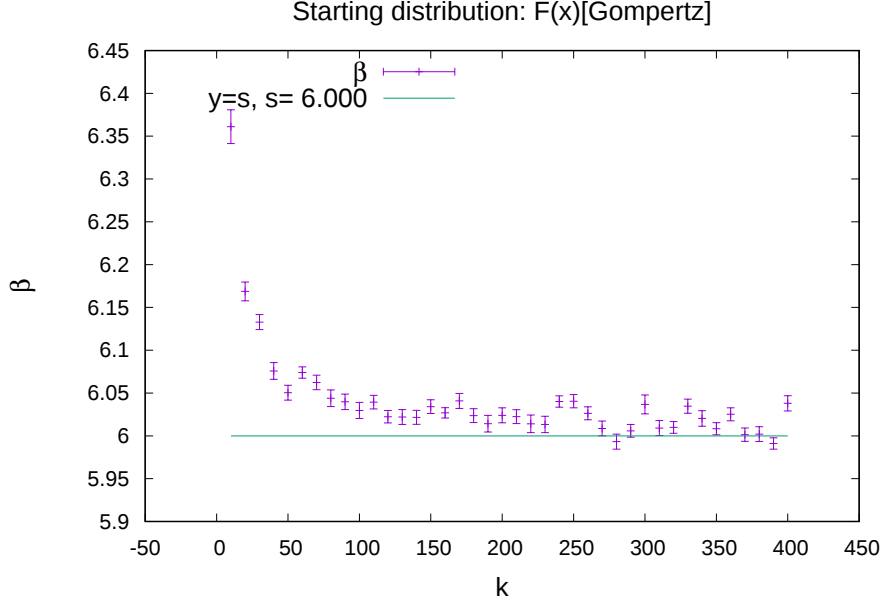


Figura 5: β in funzione di k , $s = 6$, $\mu = 6$

Nel caso della $G(x)$ relativa a (3.3) (Laplace) β tende ad $\tilde{s}$ (per N che va all'infinito) dal basso.

A questo comportamento fanno eccezione la $G(x)$ relativa a (3.5) (Gompertz), la $F_k(x)$ relativa a (3.1) (esponenziale) e la $F_k(x)$ relativa a (3.3) (Laplace), per le quali $\beta \approx \tilde{s}$ anche con k bassi: nel primo caso, ciò può essere dovuto al fatto che la Gompertz è la CDF della Gumbel, perciò i massimi saranno distribuiti come quest'ultima per ogni N . Il secondo e il terzo caso, invece, possono essere giustificati dalla forma dell'esponenziale: si è usata infatti la stessa funzione, a parte costanti, per le approssimazioni necessarie a ricavare la forma dei campioni delle $F_k(x)$ nel paragrafo (2), quindi è lecito aspettarsi questa distribuzione per qualsiasi $k > 1$.

Per quanto riguarda gli errori, si è lasciato l'errore indicato dal software utilizzato per l'analisi, Gnuplot.

4.2 Parametro di posizione α

La moda della Gumbel, indicata come α , deve crescere linearmente con il logaritmo naturale di N nel caso delle $G(x)$ secondo (1.14). Oltre a verificare questo, si è verificato come lo stesso andamento si presenti nel caso delle $F_k(x)$ in funzione del logaritmo di k , come previsto dalla (2.20).

In entrambi i casi si è quindi effettuata la regressione lineare di α in funzione

del logaritmo di k (N):

$$\begin{cases} \alpha = a + l \cdot \ln(N) \\ \alpha = a + l \cdot \ln(k). \end{cases} \quad (4.1)$$

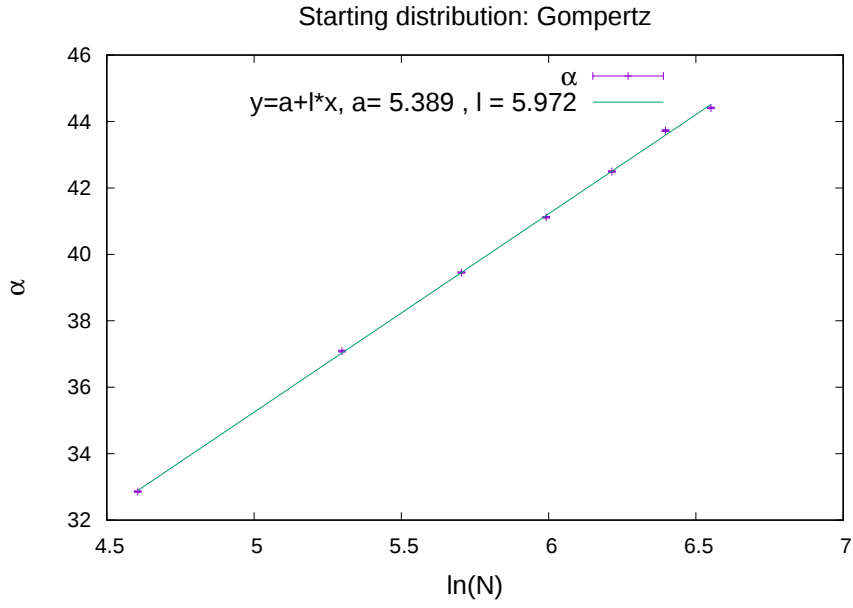


Figura 6: α in funzione di $\ln(N)$, $s = 6$, $\mu = 6$

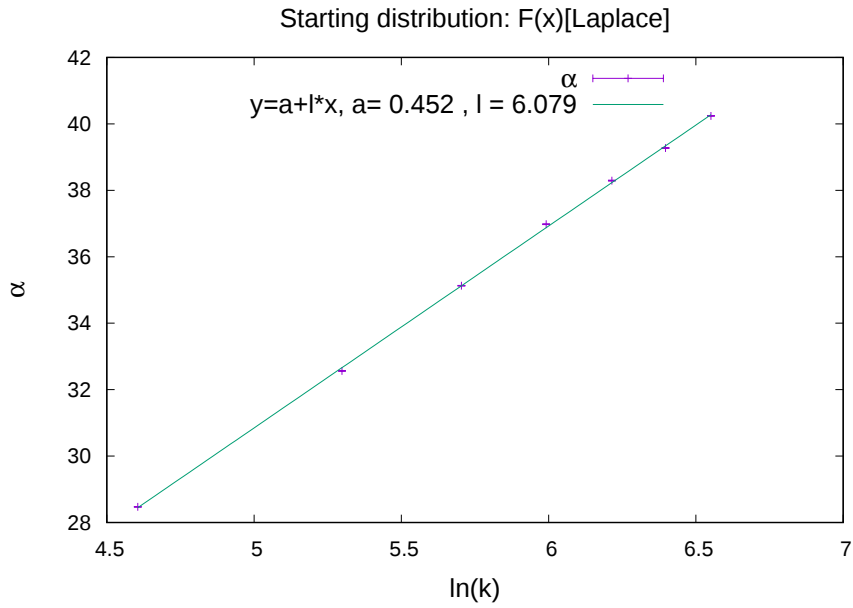


Figura 7: α in funzione di $\ln(k)$, $s = 6$, $\mu = 6$

Nei fit in funzione dei logaritmi di N e k (figure (6) e (7)), la prima cosa che si nota è che, come previsto, la pendenza l è simile al termine $\tilde{s}$. Anche in questo caso la (3.2) (tangente) ha la pendenza l dimezzata rispetto alle altre distribuzioni avendo $\tilde{s} = \frac{s}{2}$.

Dalle (1.14) e (2.20) ci si aspetta che nel caso $\lambda = 1$ l'intercetta a dipenda da μ e non da $\tilde{s}$, tuttavia in generale risulta una dipendenza da entrambi i parametri: questa è comunque lineare, anche se in presenza di un forte rumore. La regressione lineare $a = b + c \cdot \tilde{s}$ (figure (8) e (9)) dà risultati simili nei vari casi, sia per le $G(x)$ che per le $F_k(x)$:

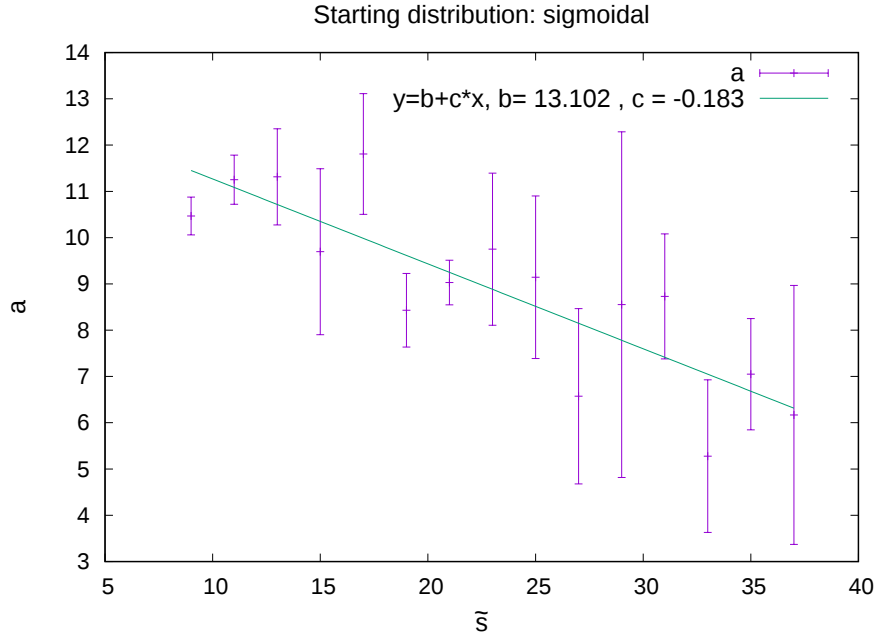


Figura 8: a in funzione di $\tilde{s}$, $\mu = 13$

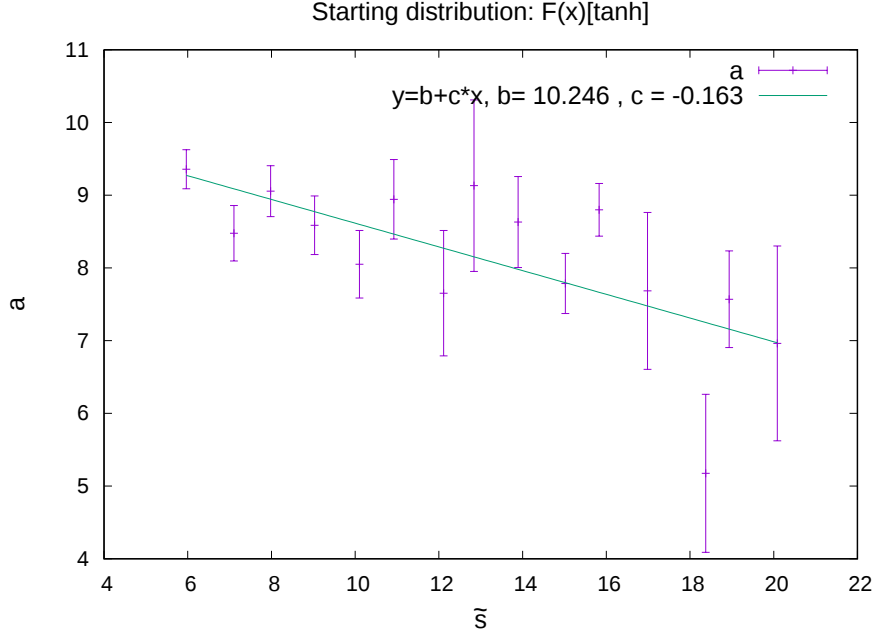


Figura 9: α in funzione di $\tilde{s}$, $\mu = 10$

Gli errori nei fit delle figure (8) e (9) sono dati dalla regressione lineare dalla quale si ottiene α .

Si può supporre che il comportamento osservato sia dovuto al fatto che la distribuzione di un campione da una $F_k(x)$ tende ad una Gumbel solo per $k \gg 1$, mentre si comporta diversamente per k piccoli (lo stesso vale per le $G(x)$ e N). Tale comportamento segue comunque una regola generale data da:

$$b + c \approx \mu. \quad (4.2)$$

Per verificare questo escludendo i comportamenti che si hanno per k (N) bassi si è fissato un k (N) alto e si è osservata la dipendenza di α da $\tilde{s}$ effettuando dei fit lineari (figure (10) e (11)) del tipo:

$$\alpha = d + m \cdot \tilde{s}. \quad (4.3)$$

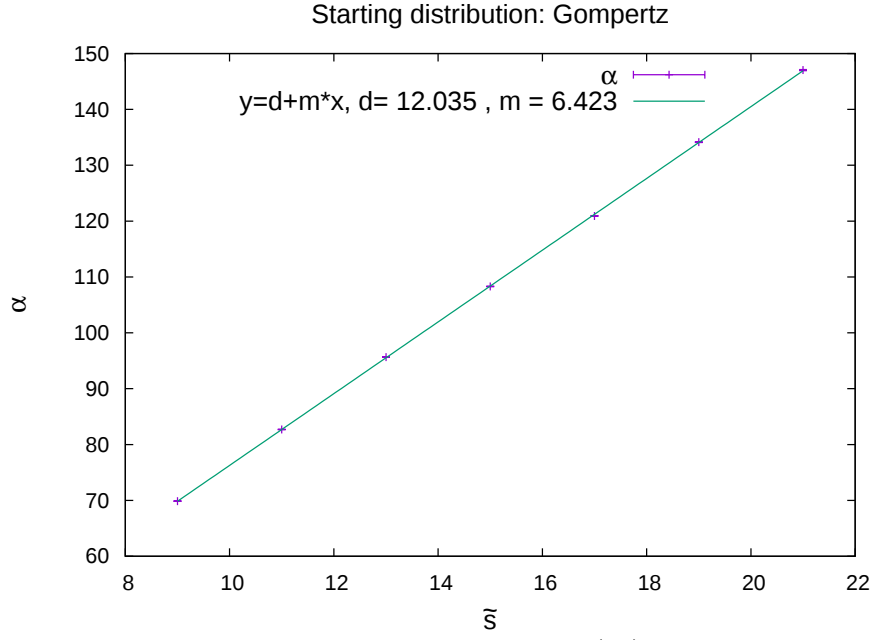


Figura 10: $N = 700, \mu = 12, \ln(N) = 6.55$

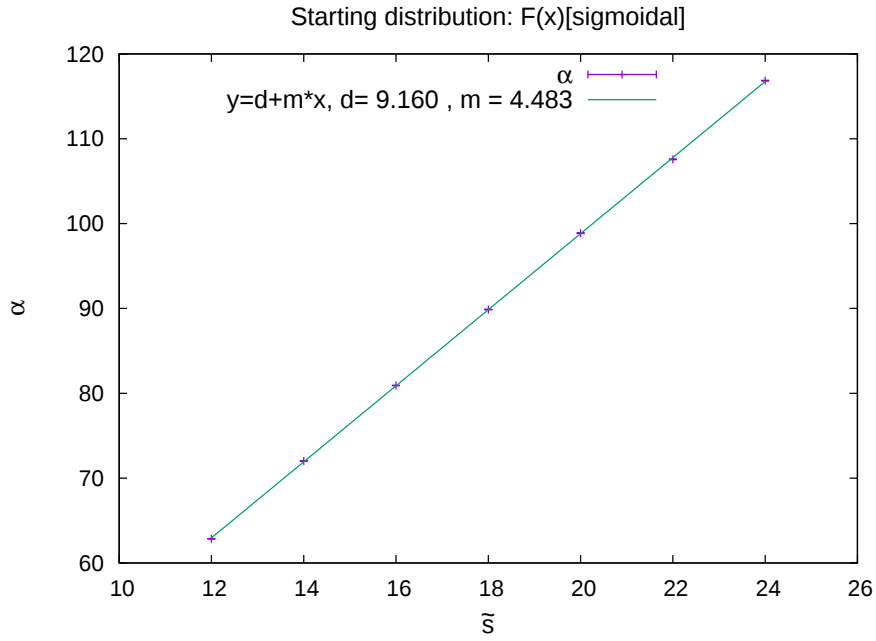


Figura 11: $k = 100, \mu = 9, \ln(k) = 4.60$

Dalla (2.20) ci si aspetta che $m \approx \ln(k) + \ln(\lambda)$ e che $d \approx \mu$ ed in effetti è quello che si osserva.

Se in generale $\lambda = 1$ e perciò $m \approx \ln(k)$, questo non vale per la (3.3) (Laplace), nel qual caso m ha un valore diverso da $\ln(k)$ ma che aumenta comunque

logaritmicamente con k (N). La (3.3) (Laplace) è infatti l'unica distribuzione considerata per la quale $\lambda \neq 1$, bensì $\lambda = \frac{1}{2}$. Perciò, al valore di α contribuisce anche un termine $\propto \ln(\lambda)$. Quindi, riscrivendo la (4.3) tenendo conto della (2.20):

$$\alpha_{laplace} = d + m \cdot \tilde{s} = d + (\ln(k) + \ln(\lambda)) \cdot \tilde{s} \quad (4.4)$$

$$\left(\lambda = \frac{1}{2}\right) \Rightarrow m \approx \ln(k) - 0.7. \quad (4.5)$$

Questi valori sono effettivamente quelli osservati.

Analogamente, lo stesso comportamento si verifica con le $G(x)$ in funzione del logaritmo di N .

5 Confronto tra i risultati dei due metodi

Come ultima analisi, si è confrontato il comportamento delle Gumbel ottenute con i due diversi procedimenti descritti in questa tesi con gli stessi parametri.

5.1 Regime asintotico

Ponendo $k = N$ e usando dei valori 'abbastanza alti' ($k \gg 10$) nonché gli stessi parametri μ ed s iniziali, si ottiene la stessa distribuzione di Gumbel sia prendendo i massimi di diversi campioni da una $G(x)$ sia prendendo i valori di un campione dalla rispettiva $F_k(x)$:

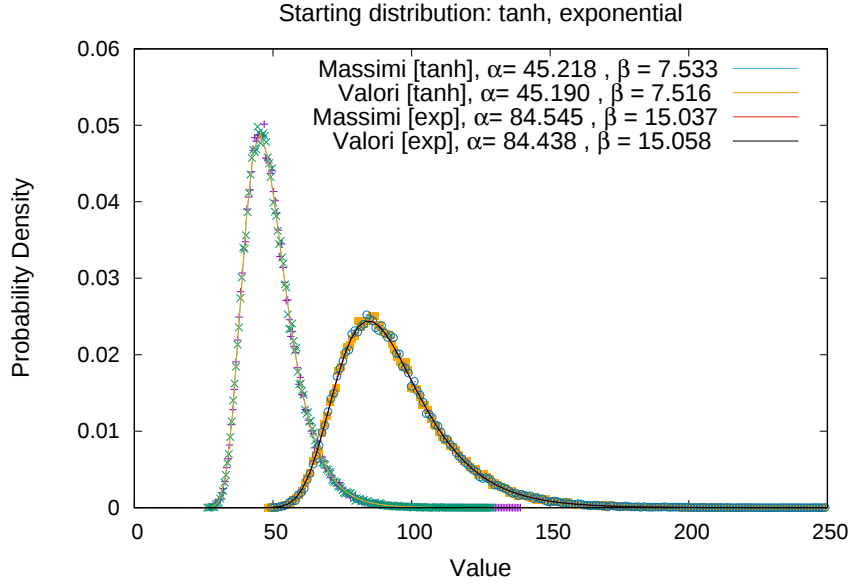


Figura 12: $\mu = 6$, $s = 15$, $k = N = 200$

Una volta scelta la distribuzione di partenza il valore del parametro β , come ci si aspetta essendo facilmente prevedibile dai parametri iniziali, è lo stesso in entrambi i casi (dimezzato per la tangente).

Anche il parametro α rispetta le previsioni che si ottengono dalle (1.14) e (2.20).

		α	β	α atteso	β atteso
Esponenziale	$G(x)$	84.545	15.037	85.475	15
	$F_k(x)$	84.438	15.058	85.475	15
Tanh	$G(x)$	45.218	7.533	45.737	7.5
	$F_k(x)$	45.190	7.516	45.737	7.5
Laplace	$G(x)$	74.037	15.003	75.078	15
	$F_k(x)$	73.975	15.049	75.078	15
Sigmoide	$G(x)$	84.176	15.006	85.475	15
	$F_k(x)$	84.128	15.128	85.475	15
Gompertz	$G(x)$	84.247	14.956	85.475	15
	$F_k(x)$	84.136	15.056	85.475	15

Tabella 1: $\mu = 6$, $s = 15$, $N = k = 200$, 10^5 campionamenti

5.2 Regime non asintotico

La situazione cambia con $k = N \lesssim 10$, ovvero in un regime in cui si riesca ad osservare una differenza della Gumbel rispetto ai dati fittati:

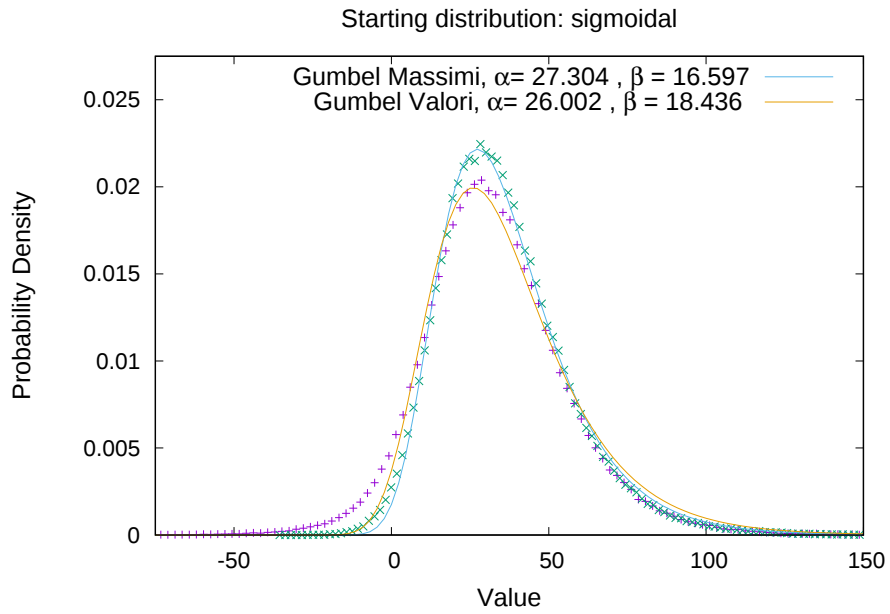


Figura 13: $k = N = 5$

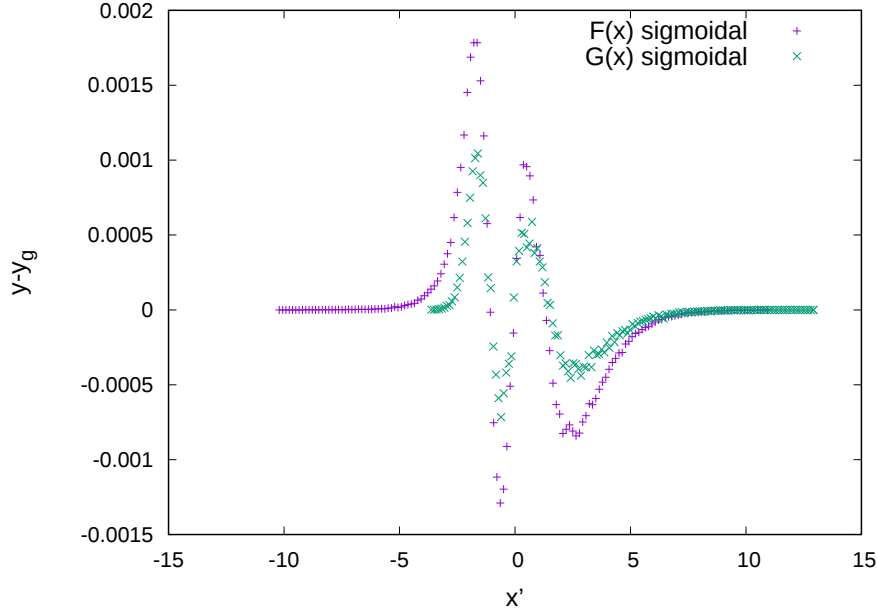


Figura 14: $k = N = 5$

Tale differenza ($y - y_g$) si mostra in figura (14), dove y e y_g denotano rispettivamente i valori della distribuzione campionata e della distribuzione di Gumbel fittata a parità di ascissa.

Per effettuare il confronto, la coordinata x di ogni punto è stata traslata usando i dati della Gumbel relativa:

$$x' = \frac{x - \alpha}{\beta}. \quad (5.1)$$

Anche in questo caso, sia i punti ottenuti dalle $G(x)$ che quelli ottenuti dalle $F_k(x)$ mostrano lo stesso andamento, seppur con ampiezze diverse. Fanno eccezione la $G(x)$ relativa a (3.5) (Gompertz), la $F_k(x)$ relativa a (3.1) (esponenziale) e la $G(x)$ relativa alla (3.3) (Laplace).

Dei primi due casi si è discusso nel paragrafo (4.1), e le motivazioni possono essere le stesse: in questo caso i punti sembrano distribuirsi attorno allo zero ($y - y_g \approx 0$) senza un andamento specifico. Nel caso della $G(x)$ della distribuzione di Laplace, l'andamento è simile con segno opposto rispetto agli altri casi: anche questo è coerente con le osservazioni del paragrafo (4.1) riguardo la stima di β .

6 Conclusioni

Le analogie tra i due casi studiati sono piuttosto evidenti. Questo nonostante le metodologie indagate fossero concettualmente diverse: nel caso delle $F_k(x)$ non c'è infatti bisogno di considerarlo un problema di valori estremi, come invece è necessario nel caso dei massimi. Questa potrebbe essere una possibile spiegazione al caso della bioluminescenza da cui è partita l'analisi delle amplificazioni delle code esponenziali: si osserva in effetti una Gumbel anche se non sembra trattarsi di un problema ai valori estremi. In quel caso $k \approx 5$, che è la situazione in figura (13) nel caso della Gompertz.

Da notare, comunque, che il metodo studiato risulta universale: i risultati ottenuti sono validi per qualsiasi distribuzione, con l'unica condizione che questa abbia coda esponenziale.

Uno degli sviluppi che potrebbe essere interessante per il futuro potrebbe essere quello di estendere questo approccio allo studio delle variabili correlate: negli ultimi anni infatti la distribuzione di Gumbel è stata utilizzata per descrivere la distribuzione delle medie di queste [5]. Il legame tra correlazione e valori estremi non è ancora chiaro [6], tuttavia la metodologia proposta in questo lavoro di tesi potrebbe risultare comune anche in questo ambito.

In conclusione, un'altra prospettiva per dei lavori futuri consiste nello studio di distribuzioni che non hanno coda esponenziale: in questo lavoro si è analizzata solo questa classe di distribuzioni, tuttavia considerando i massimi si arriva ad una Gumbel anche in altri casi. Alcuni esempi sono la gaussiana, la lognormale o anche un'esponenziale con il dominio limitato superiormente [3].

Riferimenti bibliografici

- [1] Stuart Coles, Anthony Davison. *Statistical Modelling of Extreme Values*, Basato su ‘An Introduction to Statistical Modeling of Extreme Values’, di Stuart Coles, Springer, 2001 (2008).
- [2] D. Delle Side, M.S.Tredici. *Bacterial bioluminescence and gumbel statistics: From quorum sensing to correlation*, Applied Physics Letters, 2015.
- [3] P. Embrechts, C. Klüppelberg, T. Mikosch. *Modelling Extremal Events*, Insurance and Finance, February 12, 1997.
- [4] M. H. Zwietering, I. Jongenburger, F. M. Rombouts, K. van’t Riet. *Modeling of the Bacterial Growth Curve*, Appl. Environ. Microbiol. 56, 1875 (1990).
- [5] Steven T. Bramwell. *The distribution of spacially averaged critical properties*, Nature Physics 5, 2009.
- [6] Eric Bertin, Maxime Clusel. *Generalised extreme value statistics and sum of correlated variables*, Journal of Physics A 39, 2006.