



Analisi dei Tempi in Psicologia

Una Distribuzione Sconosciuta?



ENRICO TOFFALINI enrico.toffalini@unipd.it

in collaborazione con: **PAOLO GIRARDI** paolo.girardi@unive.it

28 ottobre 2022

BANALE PREMESSA: DISTRIBUZIONE SBAGLIATA → CONCLUSIONI SBAGLIATE

→ Se i punteggi si distribuiscono come una Gamma non è appropriato analizzarli assumendo che siano Normali

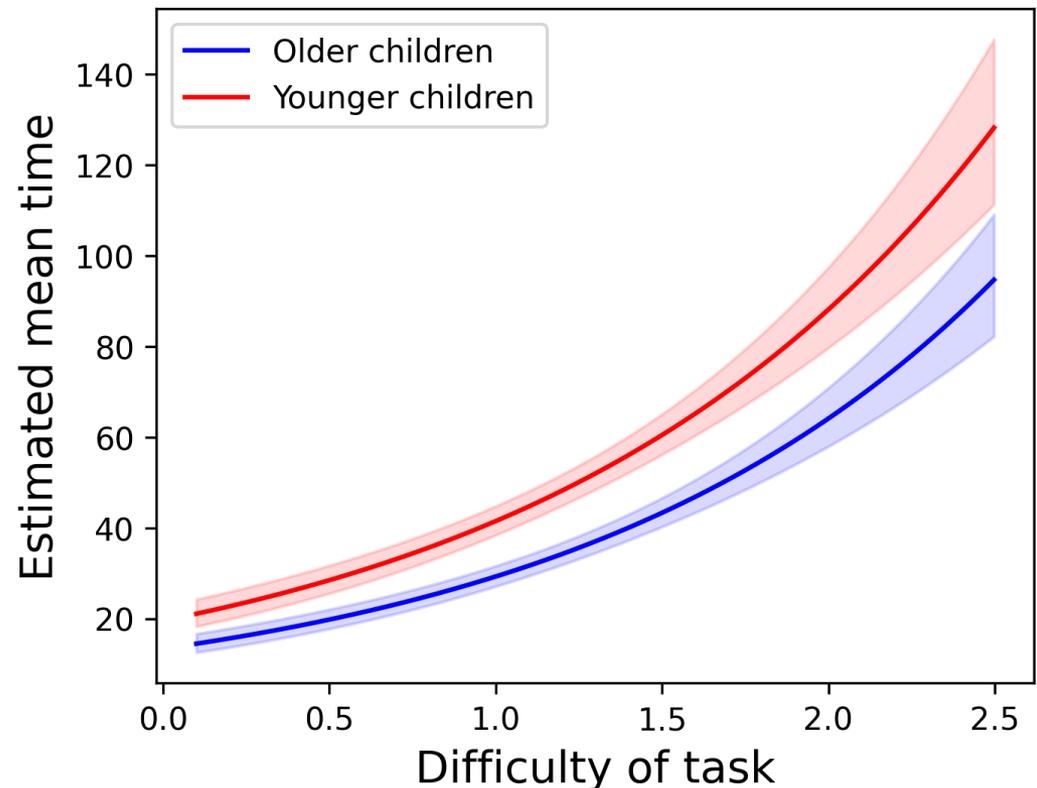
- Se modello genera dati con distribuzione Gamma e ha 2 effetti principali X + Z ma *nessuna* interazione, *linear model* «normale» troverà interazione

Gamma (link log) VERO MODELLO
 GENERATIVO

$$y \sim 2.6 + 0.8*x + 0.4*z + 0.0*x*z$$

Normal

$$y \sim 0.6 + 32.5*x + 1.3*z + 11.9*x*z$$



BANALE PREMESSA: DISTRIBUZIONE SBAGLIATA → CONCLUSIONI SBAGLIATE

→ Se i punteggi si distribuiscono come una Gamma non è appropriato analizzarli assumendo che siano Normali

- Ovviamente vale anche il contrario: una *vera* interazione sulla distribuzione Gamma può essere mascherata dal *LM* «normale»

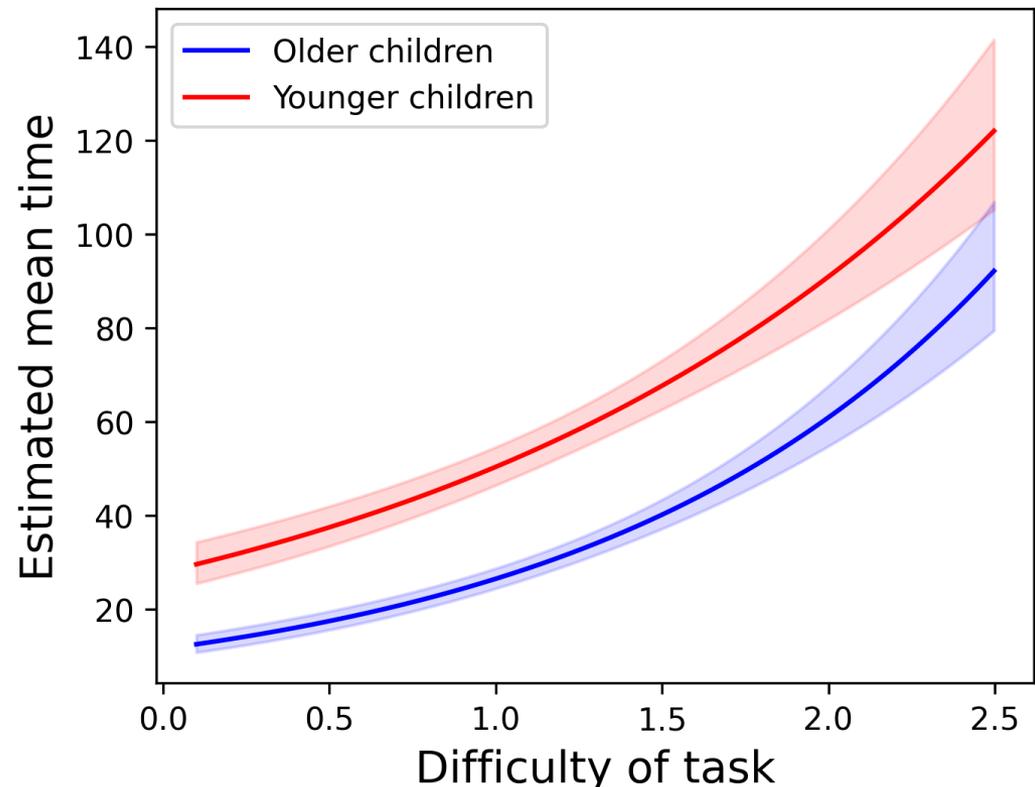
Gamma (link log) VERO MODELLO
 GENERATIVO

$$y \sim 2.4 + 0.8*x + 0.9*z - 0.2*x*z$$

Normal

$$y \sim -4.9 + 34.6*x + 25.2*z + 0.0*x*z$$

(qui con un range di osservazioni X più ampio, la curva blu avrebbe superato quella rossa, e allora anche col modello «normale» avremmo osservato l'interazione; il problema è che il modello «normale» non riproduce adeguatamente i dati!)



BANALE PREMESSA: DISTRIBUZIONE SBAGLIATA → CONCLUSIONI SBAGLIATE

Esempio

Cognitive Neuropsychology, 2014

Vol. 31, Nos. 5–6, 378–412, <http://dx.doi.org/10.1080/02643294.2014.895314>

Routledge
Taylor & Francis Group

The word-length effect in reading: A review

Jason J. S. Barton¹, Hashim M. Hanif¹, Laura Eklinder Björnström², and Charlotte Hills¹

¹Departments of Medicine (Neurology), Ophthalmology and Visual Sciences, University of British Columbia, Vancouver, BC, Canada

²Faculty of Medicine, Linköping University, Linköping, Sweden

→ Tempi di inizio lettura più lunghi per parole più lunghe ... ma «eccesso di rallentamento» per parole a bassa frequenza →

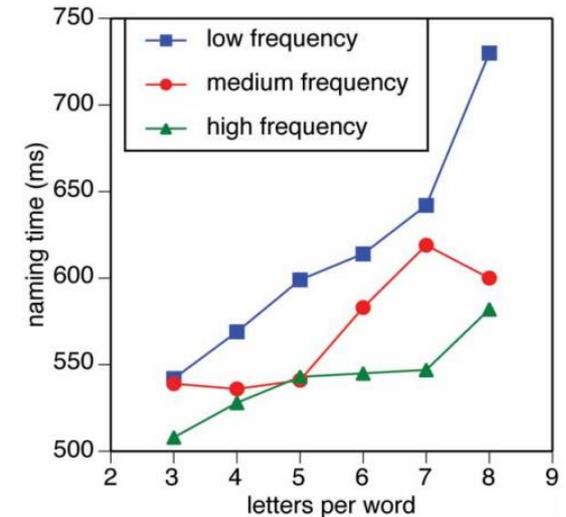


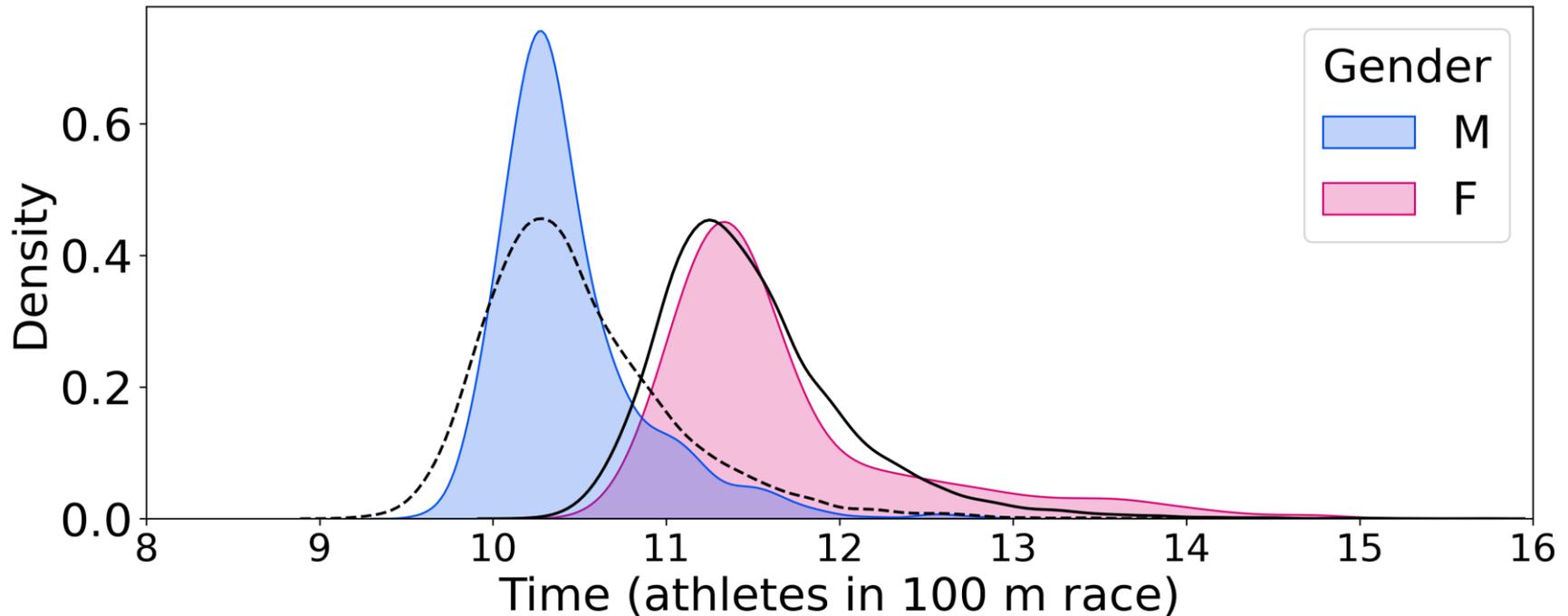
Figure 2. Frequency interactions in naming times (Cosky, 1976). These are plotted as a function of letter number, for low-frequency (20–60/10⁶), medium-frequency (61–183/10⁶), and high-frequency words (184–552/10⁶). There is a significant interaction between word length and frequency, mainly due to the excessive naming time for low-frequency 8-letter words. [To view this figure in colour, please see the online version of this Journal.]

Altri esempi in studi sulla lettura: somme di effetti principali lunghezza, frequenza, abilità di lettura (dislessia), età dei bambini, contesto di lettura (lingua opaca o trasparente) che conducono a interazioni (parziale rassegna in Toffalini et al., 2020)

EX-GAUSSIANA?

$\mu \in \mathbf{R}$ — media componente gaussiana
 $\sigma^2 > 0$ — varianza componente gaussiana
 $\tau > 0$ — rate componente esponenziale

Prima osservazione che mi aveva incuriosito: tempi dei centometristi alle olimpiadi. Distribuzione parte da punto non-zero e ha forte asimmetria destra

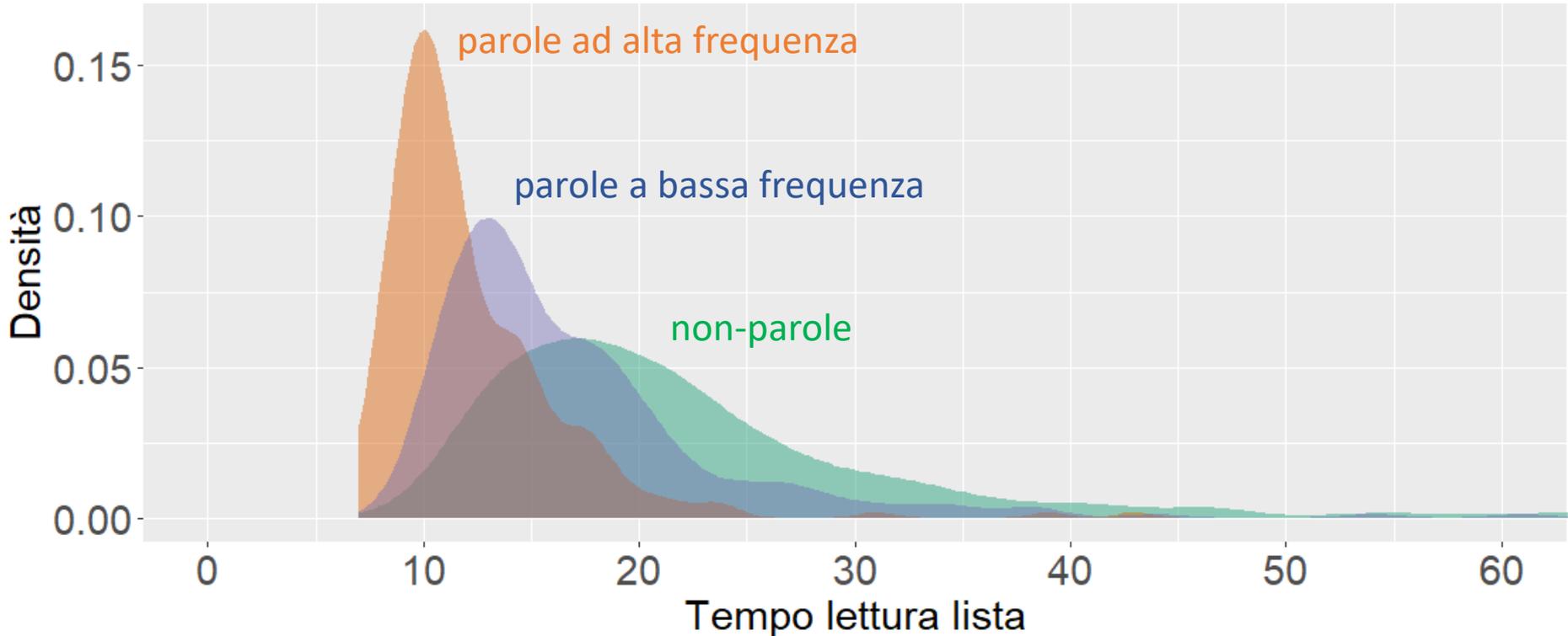


- La singola distribuzione è modellabile accettabilmente con una ex-gaussiana (ma già qui si osserva una certa dipendenza $\sigma \sim \mu$)
- Problematico processo generativo dei dati: implica forte selezione preliminare degli atleti, ma eterogenea per paese

EX-GAUSSIANA?

$\mu \in \mathbf{R}$ — media componente gaussiana
 $\sigma^2 > 0$ — varianza componente gaussiana
 $\tau > 0$ — rate componente esponenziale

Ovvia analogia: tempo di lettura – asimmetrica con inizio non-zero

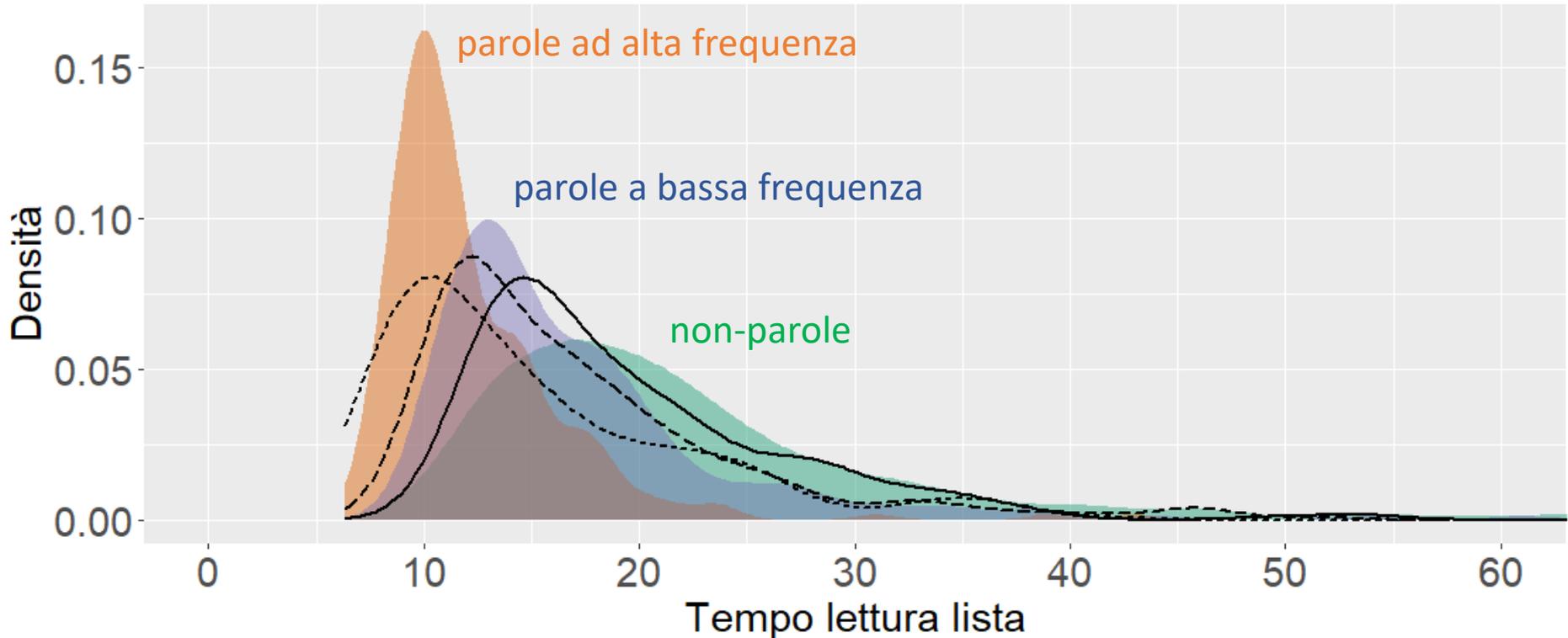


- La ex-gaussiana può modellare la singola distribuzione
- Non sembra appropriata a modellare gli effetti, perché stima μ e σ come parametri indipendenti, ma sono molto dipendenti

EX-GAUSSIANA?

$\mu \in \mathbf{R}$ — media componente gaussiana
 $\sigma^2 > 0$ — varianza componente gaussiana
 $\tau > 0$ — rate componente esponenziale

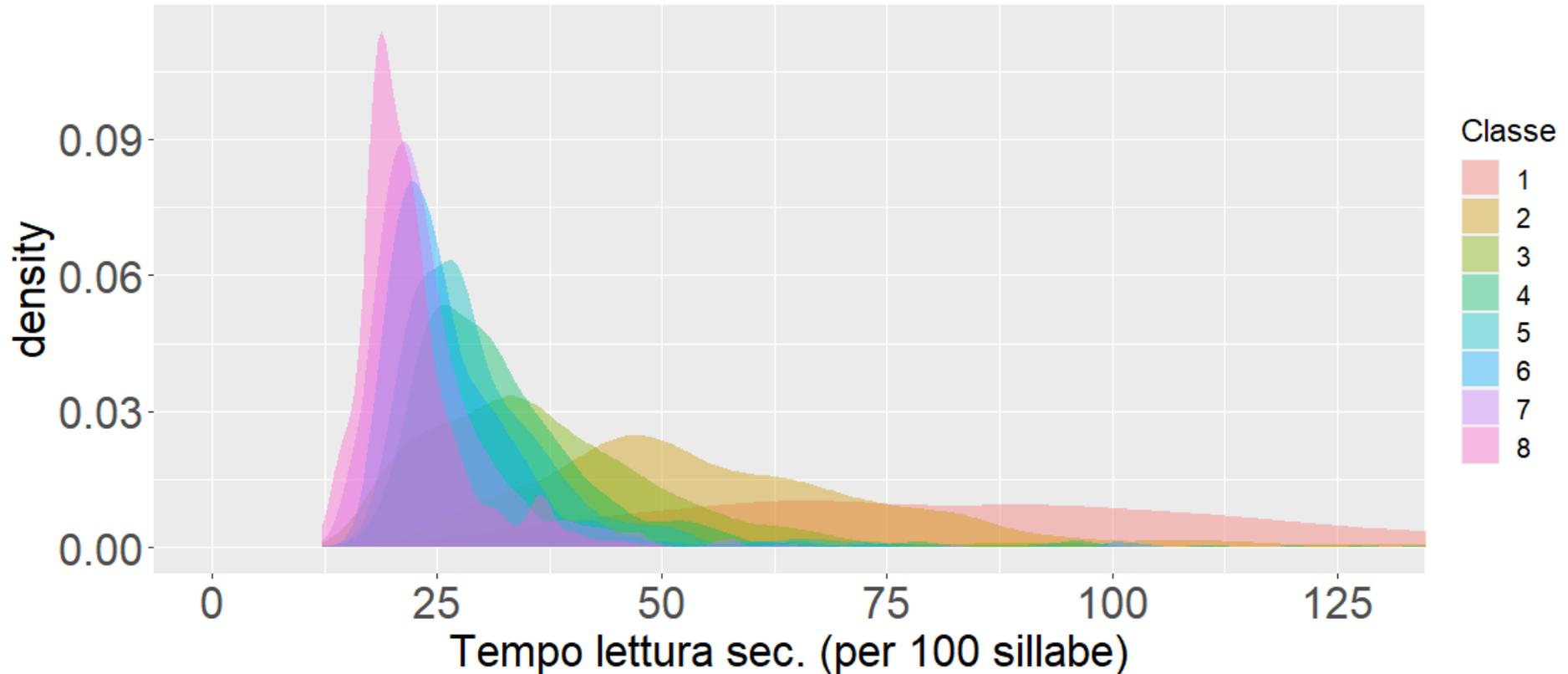
Ovvia analogia: tempo di lettura – asimmetrica con inizio non-zero



- La ex-gaussiana non appare appropriata perché stima μ e σ come parametri indipendenti, ma sono molto dipendenti

GAMMA CON OFFSET?

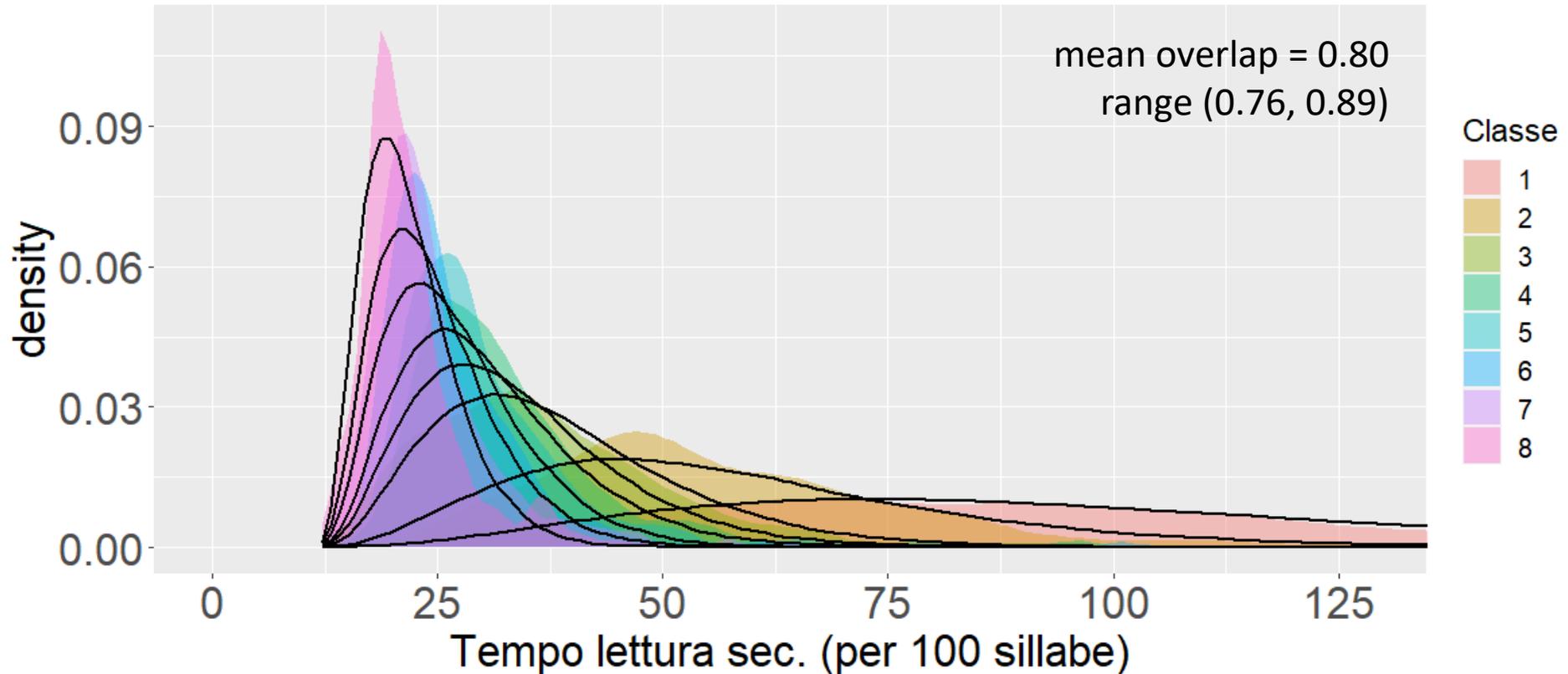
Dipendenza $\sigma \sim \mu$ molto evidente con tempi lettura a età diverse



- L'effetto ricorda molto una gamma (dove viene modellato il parametro Shape), tuttavia non parte da zero

GAMMA CON OFFSET?

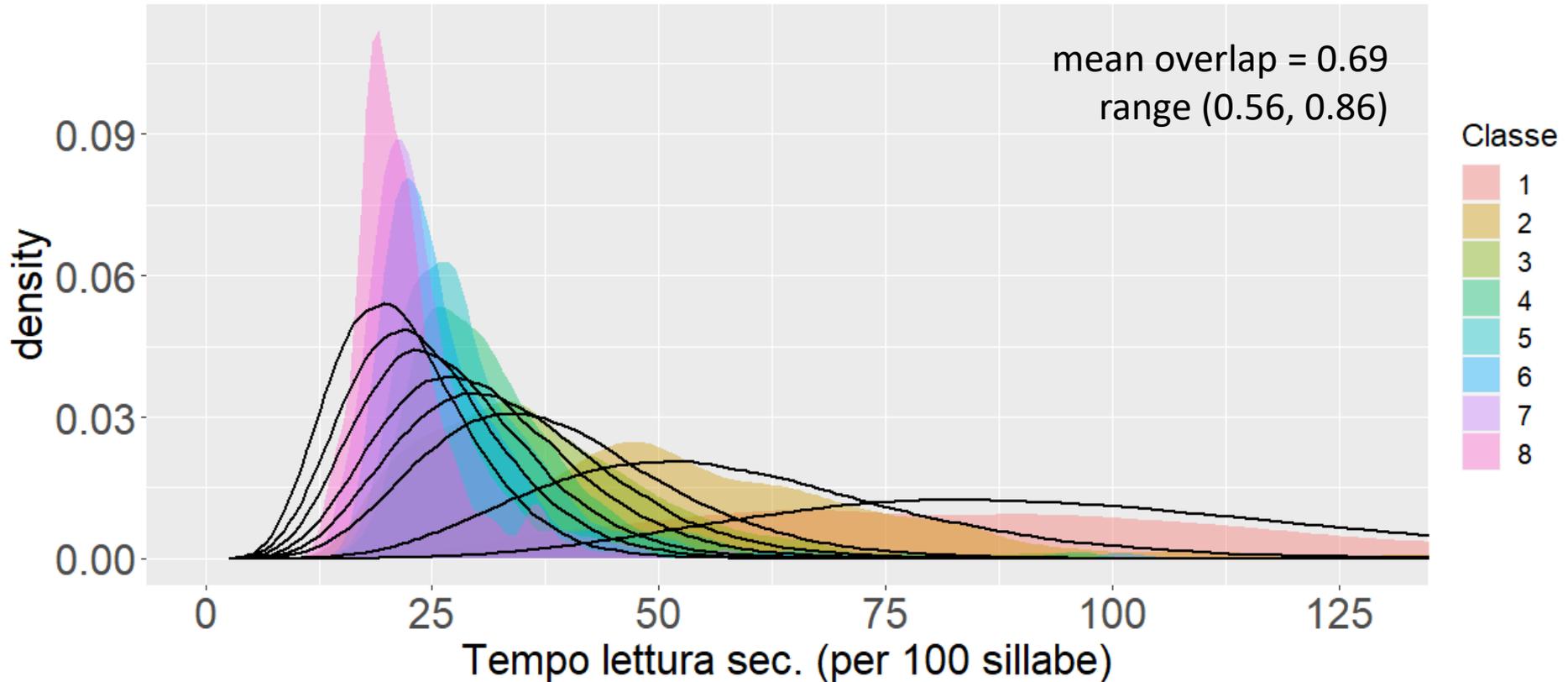
GLM con family Gamma, ma con dati «shiftati» di +12



- L'effetto ricorda molto una gamma (dove viene modellato il parametro Shape), tuttavia non parte da zero

GAMMA CON OFFSET?

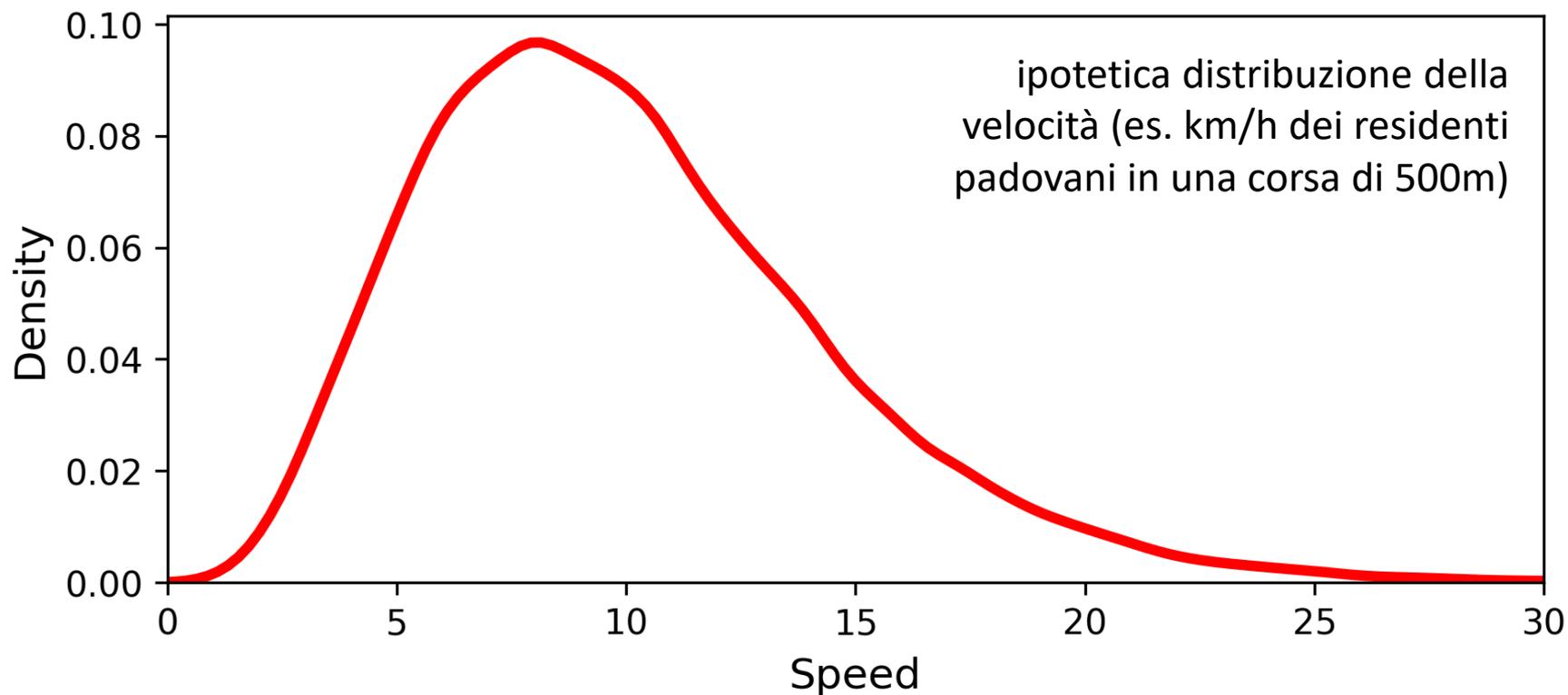
GLM con family Gamma *senza* offset



- L'effetto ricorda molto una gamma (dove viene modellato il parametro Shape), tuttavia non parte da zero

PROCESSO GENERATIVO?

Supponiamo che il parametro cruciale sottostante sia la «velocità» di un dato processo. Potrebbe essere distribuito come una classica Gamma: dai più lenti prossimi a zero, al dato di moda, fino a una coda di «alieni» super-efficienti



Seguici su:  

SALUTE

CERCA



FESTIVAL DI SALUTE 2022

COVID

SPORTELLLO CUORE

TUMORI

PSICOLOGIA

ALIMENTAZIONE

LONGFORM

VIDEO

PODCAST

SCRIVICI

LA CURIOSITA'



Velocità massima 65 km all'ora. Il record di Bolt sarà un ricordo

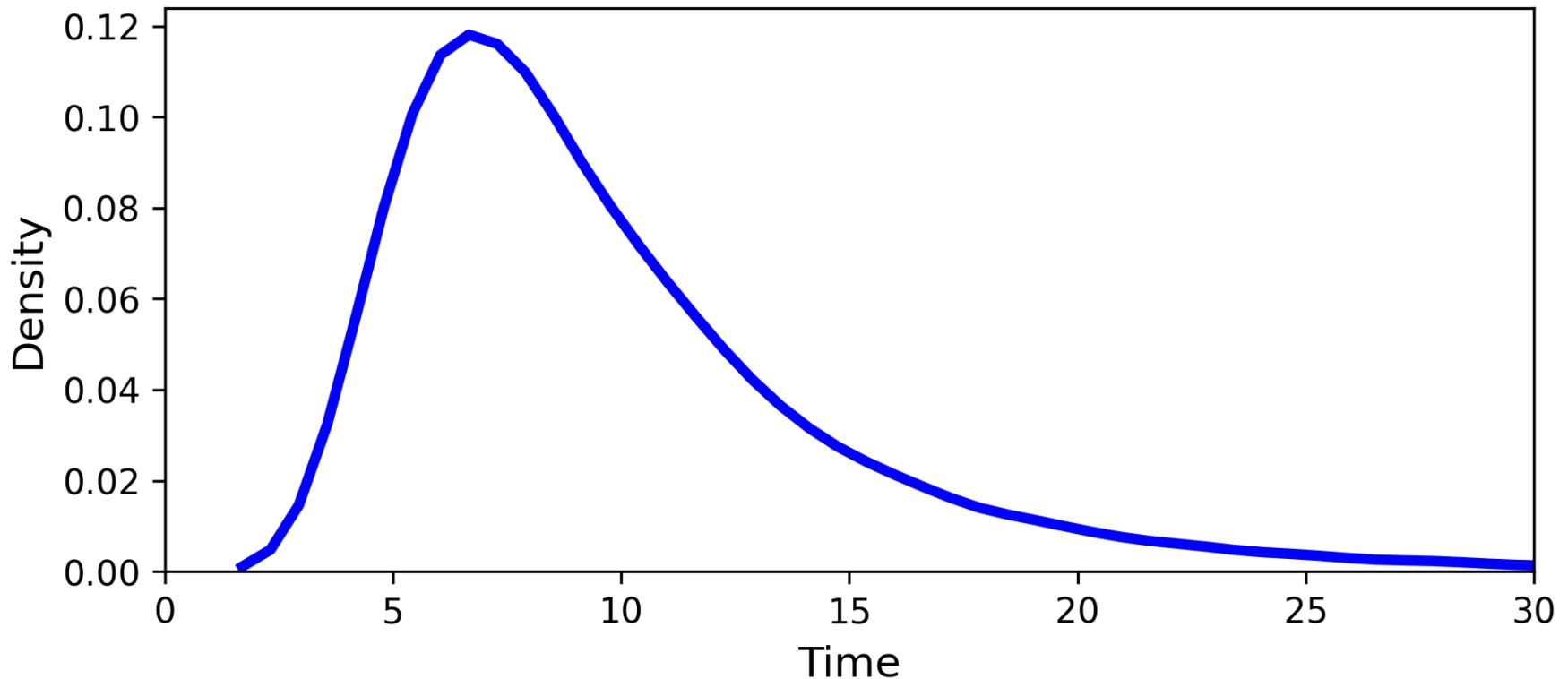
Una ricerca texana sostiene che l'uomo potrebbe raggiungere la velocità di 65 chilometri orari, ben 20 in più rispetto al primato dello sprinter giamaicano. Non sono i muscoli a fare la differenza ma il tempo di contatto tra il piede e il terreno, che negli atleti è troppo breve per sprigionare tutta la forza possibile

25 GENNAIO 2010 ALLE 18:50

🕒 1 MINUTI DI LETTURA

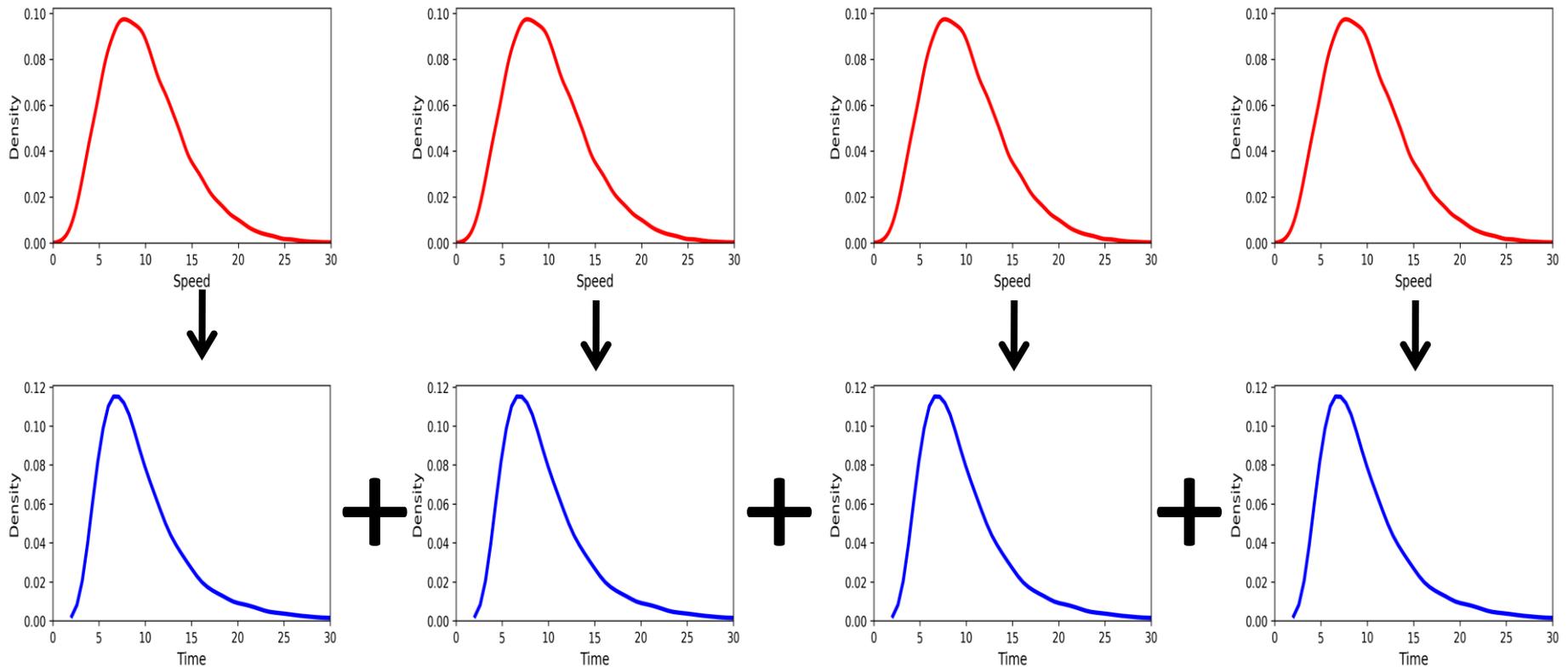
PROCESSO GENERATIVO?

Il tempo è il reciproco della velocità. Ovviamente non parte da zero perché nessuno aveva velocità ∞ , ma non abbiamo comunque ancora riprodotto il fenomeno visto in precedenza



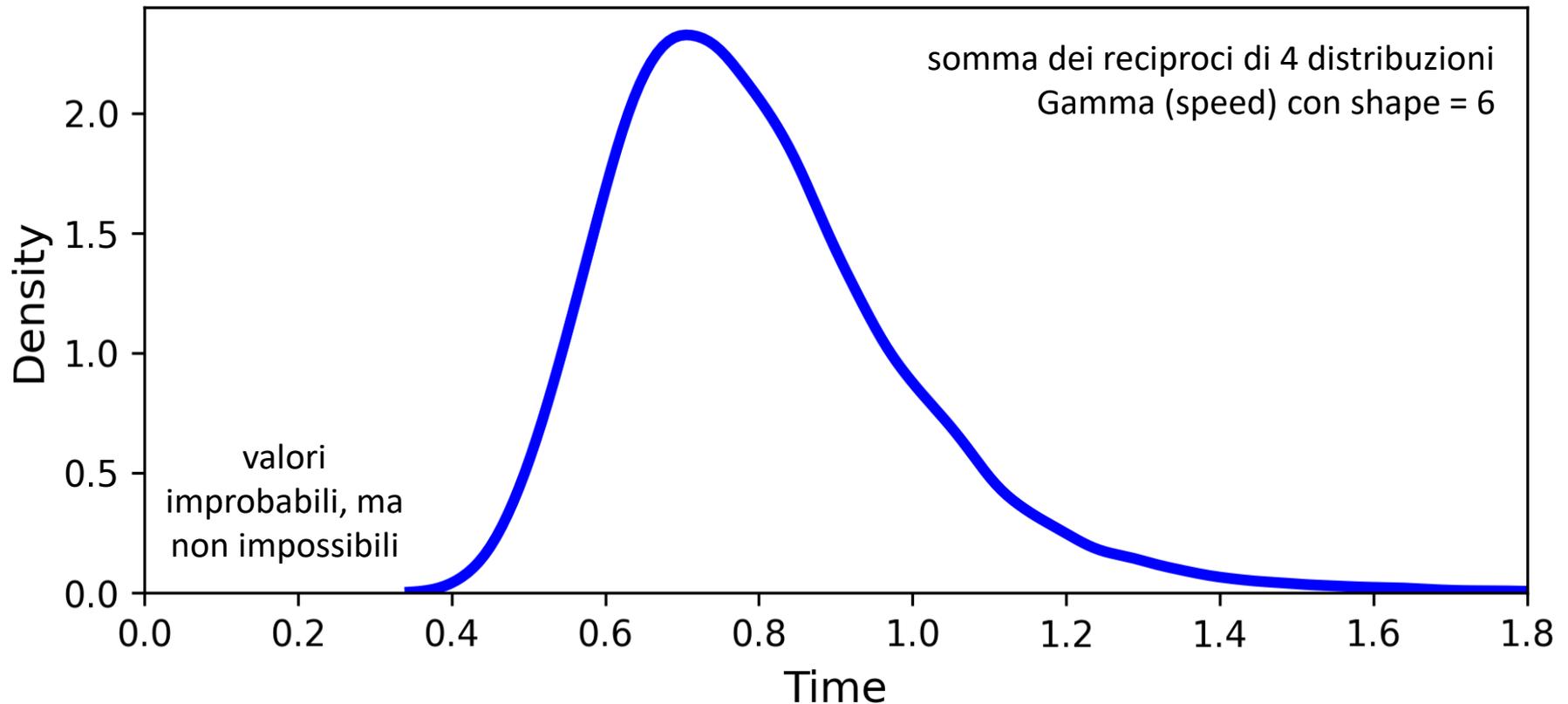
PROCESSO GENERATIVO?

Ma il fenomeno reale potrebbe dipendere da una SOMMA di processi sequenziali. Esempio **lettura**: velocità saccadi \rightarrow velocità trasmissione impulso nervoso \rightarrow velocità accesso lessicale \rightarrow velocità articolazione, ecc.



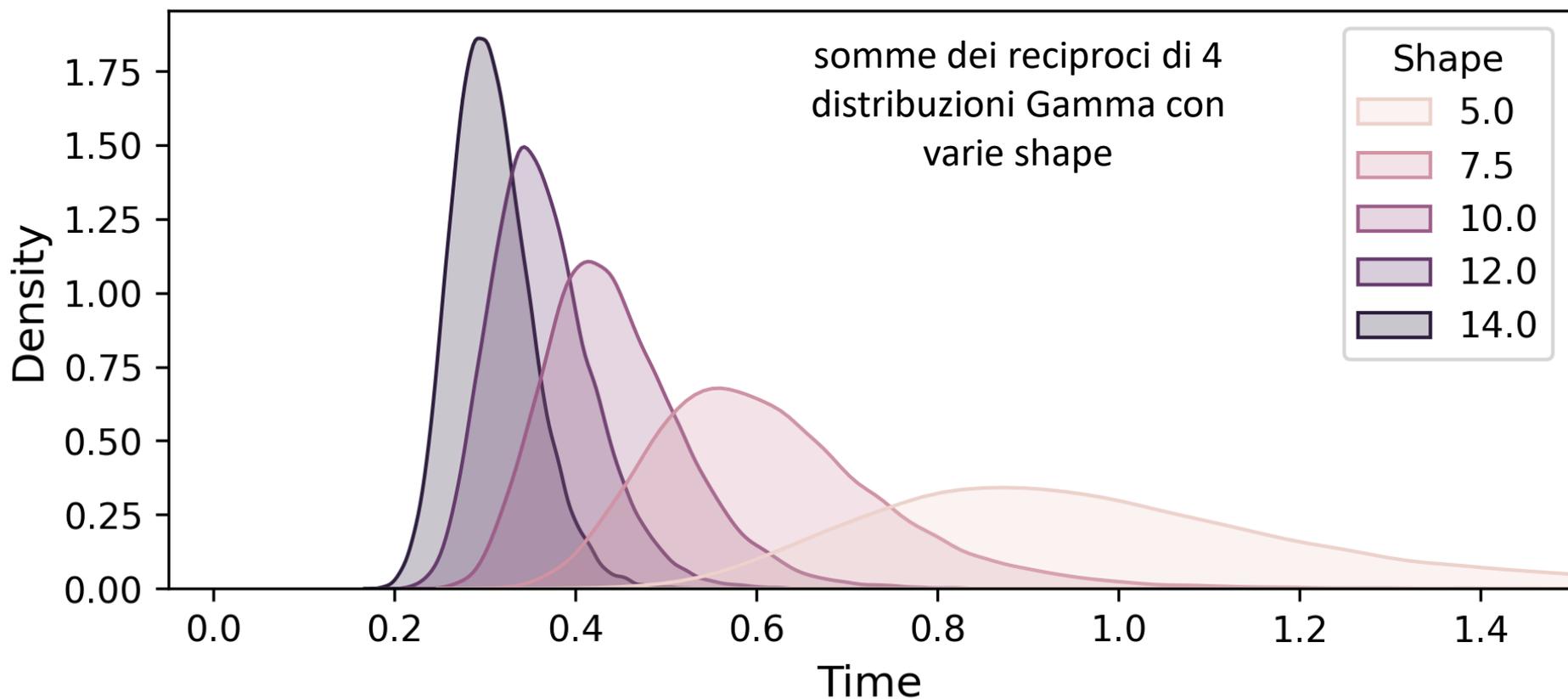
PROCESSO GENERATIVO?

Il risultato assomiglia a quello che ci aspettiamo



PROCESSO GENERATIVO?

Questo processo rende conto abbastanza bene anche degli effetti modellando la Shape



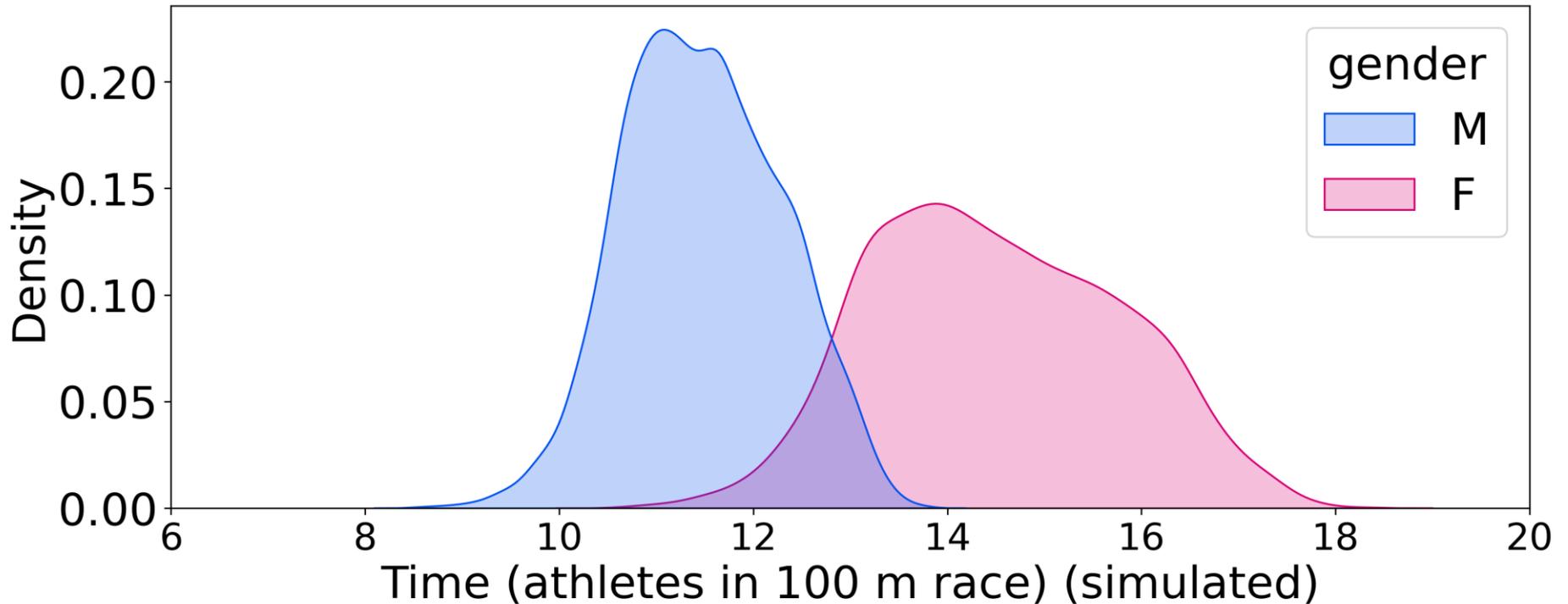
PROCESSO GENERATIVO?

Centometristi olimpiadi: mettiamo insieme le cose

- Supponiamo che il tempo finale dipenda dalla somma di 5 velocità \sim Gamma con Shape diverso per maschi e femmine (8 vs 6) e variabile (± 1) per nazione
- Le nazioni sono centinaia (poniamo $k = 1000$ per smoothing)
- Ogni nazione manda i suoi atleti migliori alle olimpiadi ($N = 10$)
- Ogni nazione ha un pool variabile di atleti candidati tra cui può selezionare effettivamente (da qualche migliaio a centinaia di migliaia)

PROCESSO GENERATIVO?

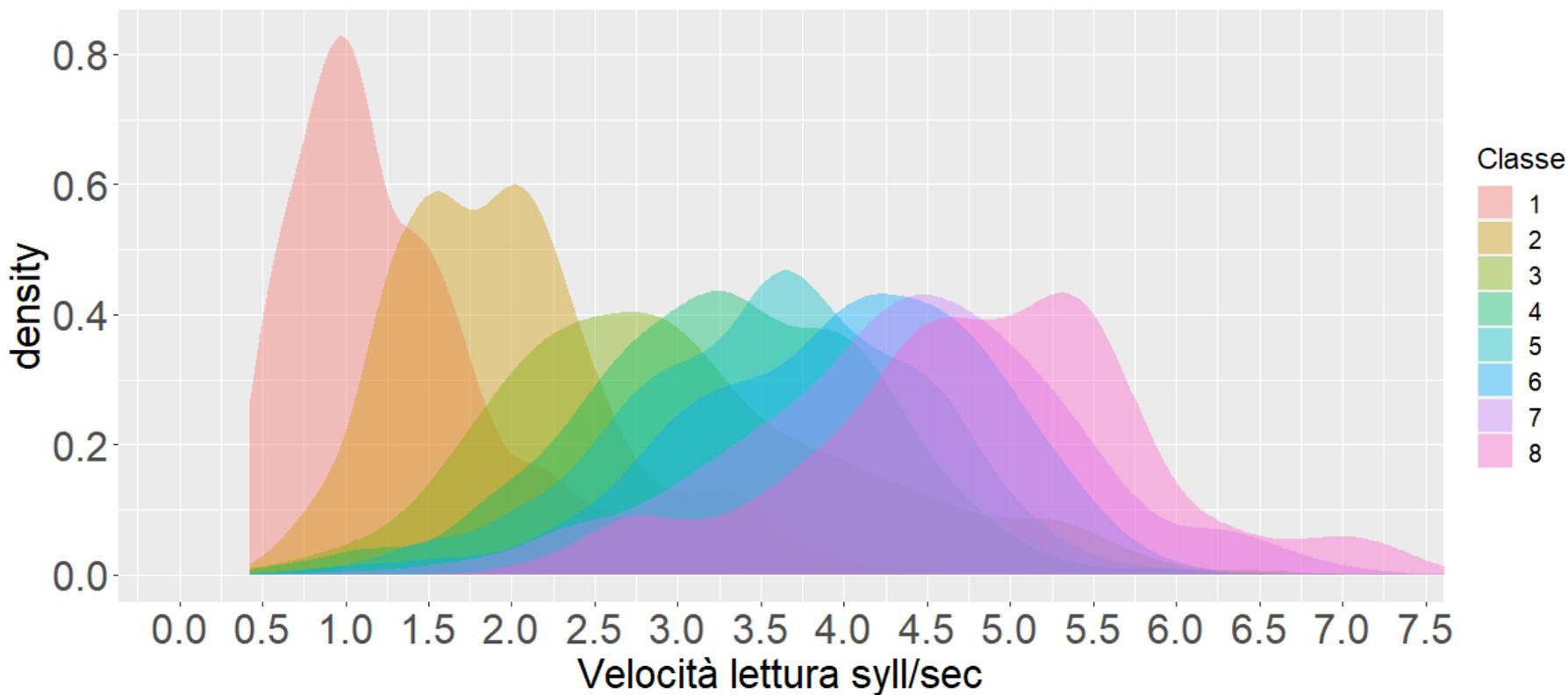
Centometristi olimpici: mettiamo insieme le cose



(...potrebbe andare meglio)

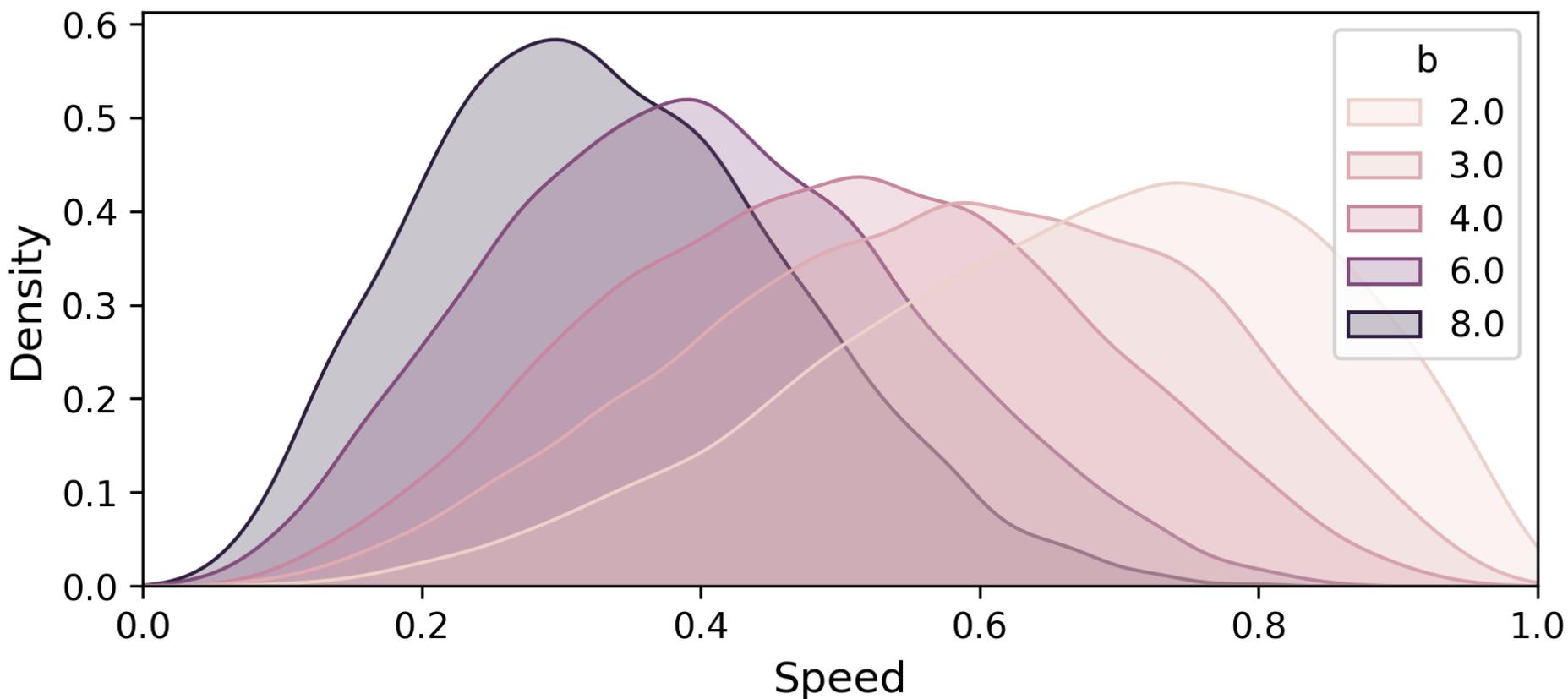
PROCESSO GENERATIVO?

La velocità non è davvero una Gamma: se guardiamo le velocità nella lettura di testo, appaiono limitate tra un minimo e (verso) un massimo



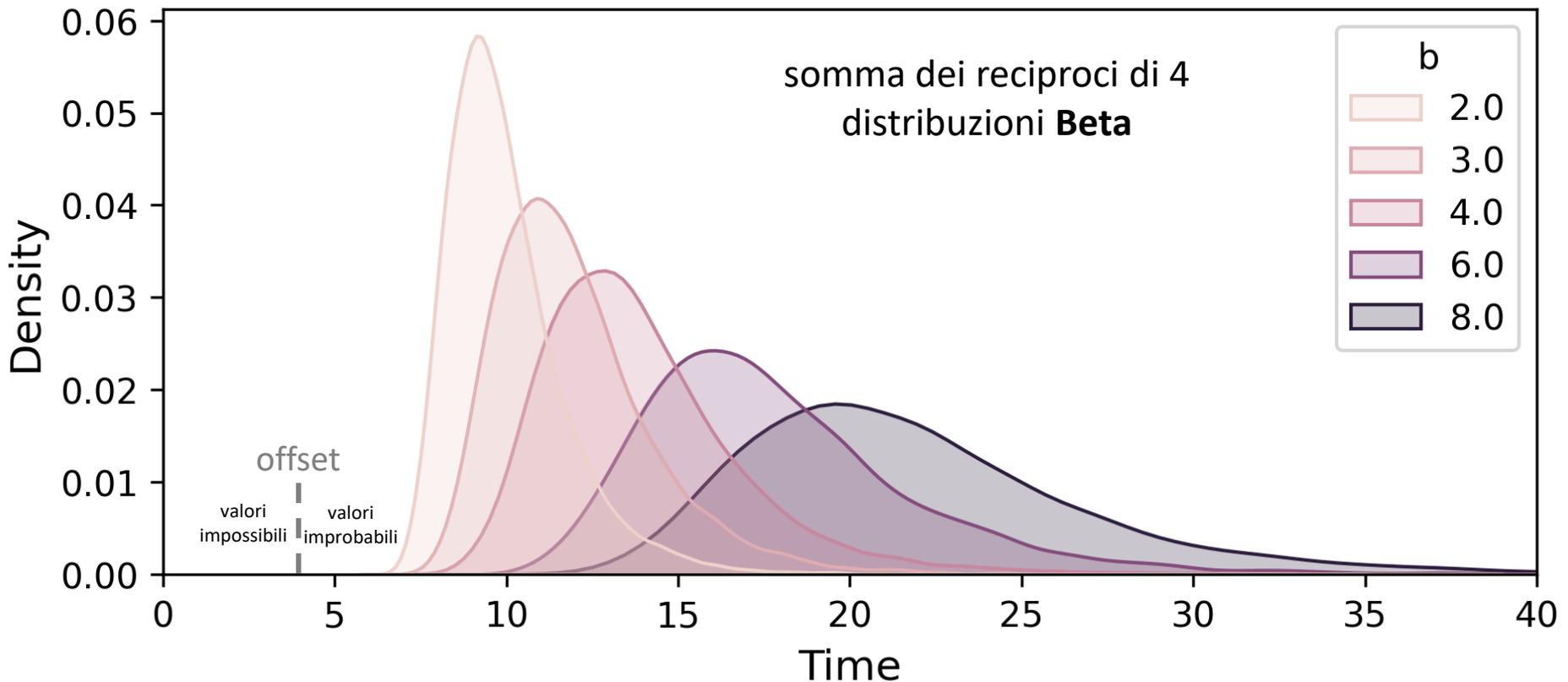
PROCESSO GENERATIVO?

Supponiamo che le velocità sottostanti si distribuiscano come una \sim Beta e modelliamo il parametro b (tengo fisso $a = 4$)



PROCESSO GENERATIVO?

Supponiamo che le velocità sottostanti si distribuiscano come una \sim Beta e modelliamo il parametro b (tengo fisso $a = 4$)



PROPOSTE DI MODELLAZIONE

L'uso della ex-Gaussiana è ben convalidato da importanti assunzioni strutturali sui tempi di reazione*

Per i dati di lettura, proponiamo altre alternative:

- Gamma mean regression;
- Gamma mean regression con offset;
- Median regression
- Laplace regression;

Misuriamo l'adattamento della regressione ai dati con:

- RMSE (Root Mean Squared Error)
- MAE (Mean Absolute Error)

**Analysis of response time distributions: An example using the Stroop task. Heathcote, Andrew; Popiel, Stephen J.; Mewhort, D. J. Psychological Bulletin, Vol 109(2), Mar 1991, 340-347.*

Vossel, S., Thiel, C. M., & Fink, G. R. (2006). Cue validity modulates the neural correlates of covert endogenous orienting of attention in parietal and frontal cortex. Neuroimage, 32(3), 1257-1264.

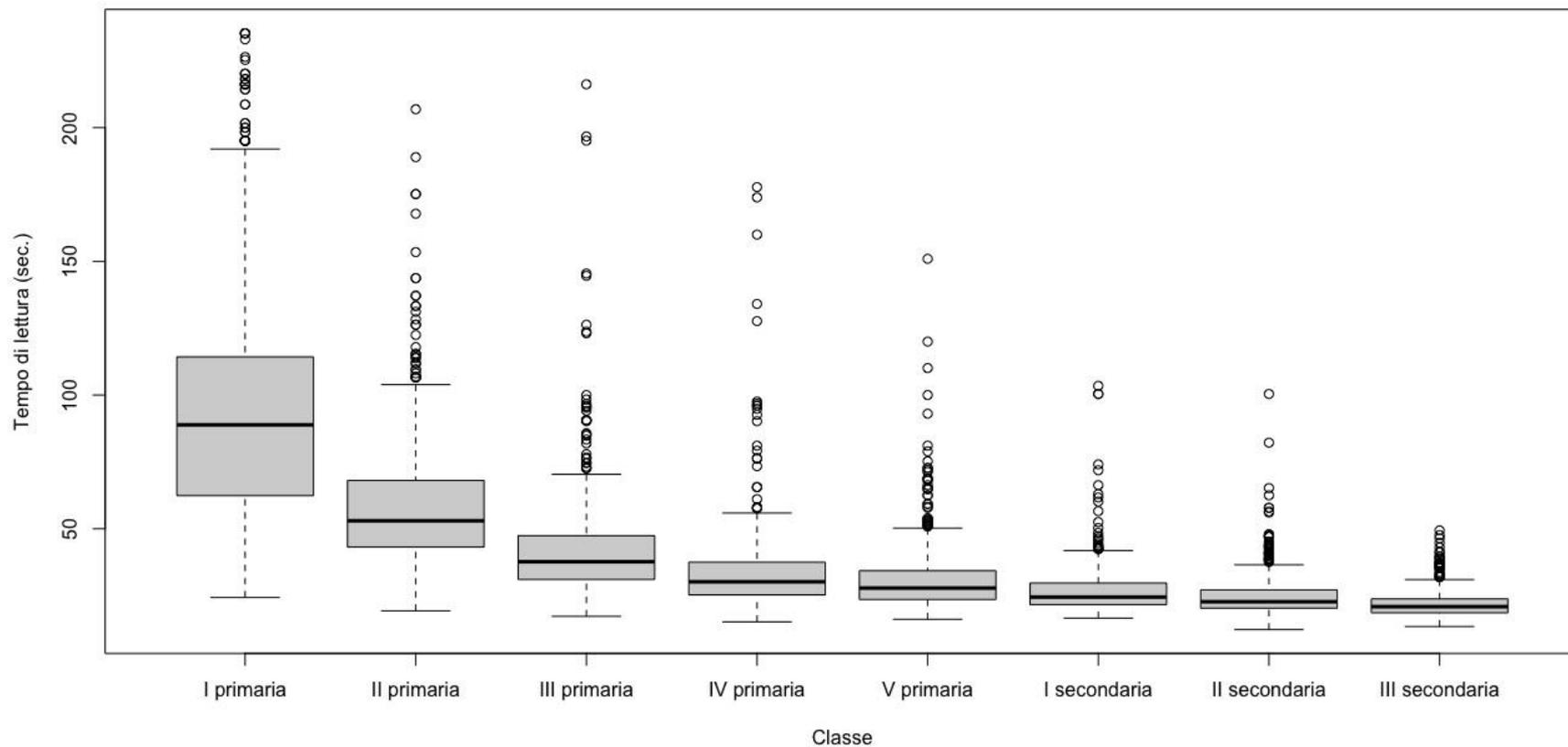
TEMPI DI LETTURA - DATASET

| | Overall, N = 5,397 ¹ | Male, N = 2,732 ¹ | Female, N = 2,665 ¹ | p-value ² |
|----------------|---------------------------------|------------------------------|--------------------------------|----------------------|
| Classe | | | | >0.9 |
| I primaria | 739 (14%) | 380 (14%) | 359 (13%) | |
| II primaria | 770 (14%) | 392 (14%) | 378 (14%) | |
| III primaria | 811 (15%) | 412 (15%) | 399 (15%) | |
| IV primaria | 620 (11%) | 310 (11%) | 310 (12%) | |
| V primaria | 772 (14%) | 398 (15%) | 374 (14%) | |
| I secondaria | 459 (8.5%) | 234 (8.6%) | 225 (8.4%) | |
| II secondaria | 638 (12%) | 312 (11%) | 326 (12%) | |
| III secondaria | 588 (11%) | 294 (11%) | 294 (11%) | |
| Tempo_Lettura | 32 (24, 50) | 33 (24, 50) | 32 (23, 51) | 0.040 |

¹ n (%); Median (IQR)

² Pearson's Chi-squared test; Wilcoxon rank sum test

TEMPI DI LETTURA - DATASET



Il modello proposto:

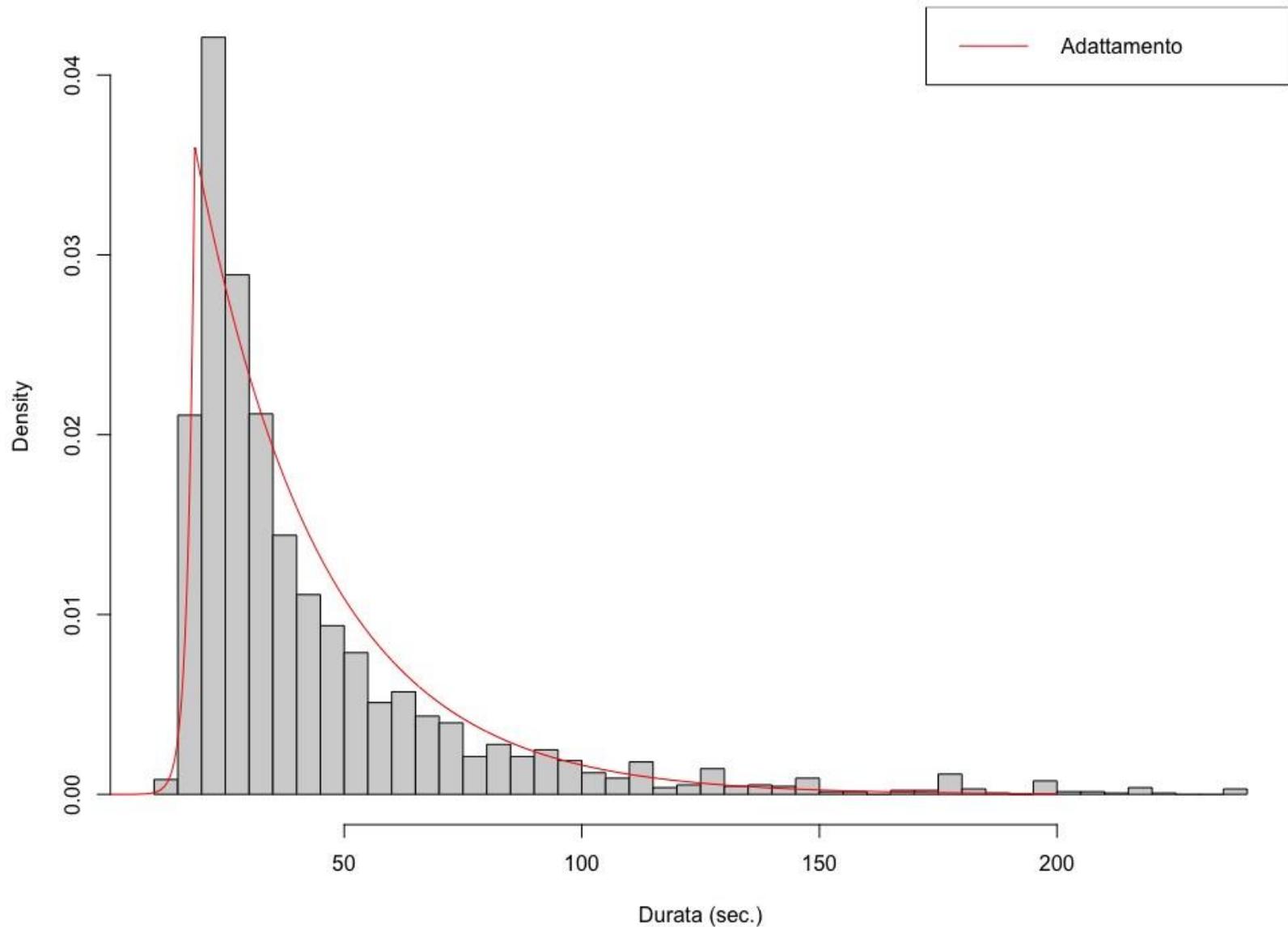
$$E[Y_i] = \alpha + \beta * Genere_i + \gamma * Classe_i$$

TEMPI DI LETTURA - Risultati

| Modello | Pacchetto | RMSE | MAE |
|-------------------------------|-----------------------|-------|-------|
| ExGAUSS | BRMS su R | 30.07 | 15.95 |
| Gamma (log link) | glm() su R | 21.81 | 12.72 |
| Gamma con offset (log link) | glm() su R | 21.83 | 12.73 |
| Median regression | Quantreg su R | 22.21 | 12.25 |
| Asymmetric Laplace regression | ald su R con optim()* | 23.04 | 12.70 |

*Metodo costruito con funzione sulla base della densità di una Laplace asimmetrica. NB: Andrebbe costruito un algoritmo EM o catene Bayesiane.

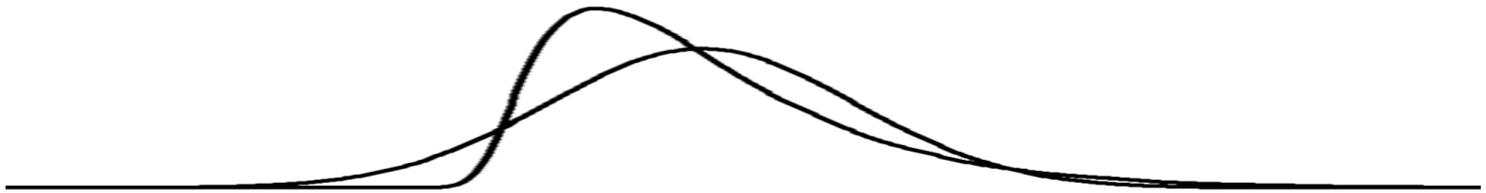
TEMPI DI LETTURA – LAPLACE ASYMMETRICA





STATISTICS IS  ART

GRAZIE PER L'ASCOLTO E PER L'AIUTO



28 ottobre 2022