

A Maximum Entropy Framework for Anthropometric Tail Modeling via Gaussian–Weibull Mixtures

Ken Kiyono

kiyono.ken.es@osaka-u.ac.jp
The University of Osaka
Toyonaka, Osaka, JAPAN

Syarifah Nabilah Syed Taha

ananabilah@bpe.es.osaka-u.ac.jp
The University of Osaka
Toyonaka, Osaka, JAPAN

Human anthropometric data such as height and body weight in a general population are commonly modeled using Gaussian distributions or Box-Cox-type power transformations designed to improve normality. However, in pediatric and adolescent populations, deviations from normality are not merely statistical artifacts; they are closely linked to clinically meaningful health conditions and growth dynamics. Notably, even after applying such transformations, residual non-Gaussian structures—particularly in the distribution tails—are frequently observed, suggesting that these deviations reflect intrinsic biological heterogeneity rather than insufficient transformation.

Both undernutrition and excess adiposity during critical growth periods carry significant long-term health consequences. Thinness is associated with impaired growth and reduced peak bone mass, while obesity increases the risk of metabolic disorders such as type 2 diabetes and cardiovascular disease. Accurate assessment of growth abnormalities—particularly in the distributional tails—is therefore essential for both clinical evaluation and public health intervention.

We propose a parametric framework for estimating asymmetric tail probabilities in anthropometric distributions using a Gaussian–Weibull mixture model. From a maximum entropy perspective, a Gaussian distribution arises naturally when only the mean and variance are constrained, consistent with the aggregate influence of numerous independent factors on anthropometric traits in healthy populations. In contrast, nutritional abnormalities introduce systematic departures from Gaussianity, manifesting as heavy or asymmetric tails that encode latent information beyond what the mean and variance alone can capture. Such deviations are modeled using Weibull distributions, which provide flexible control over tail shape and skewness.

Parameter estimation is performed using a profile likelihood-guided EM algorithm. Boundary parameters separating the central and tail regions are first optimized via profile likelihood, after which the remaining parameters are estimated via EM with latent component assignments. This procedure can be interpreted as minimizing the Kullback–Leibler divergence between the empirical distribution and the model. Specifically, the EM step maximizes a variational lower bound (evidence lower bound) that balances

the log-likelihood term with the entropy of the latent assignment distribution. This entropy term quantifies the uncertainty in assigning observations to either the Gaussian or Weibull component, reflecting the entropy-based nature of the estimation framework.

Applications to large-scale anthropometric data demonstrate that the model accurately captures deviations from Gaussianity, particularly in the distribution tails. By explicitly modeling tail structure, the method enables direct estimation of latent at-risk subpopulations (e.g., thinness and obesity) and reveals asymmetries not detectable under standard Gaussian assumptions. Figure 1 presents the results of fitting the proposed model to body weight data from 429,885 Japanese females aged 14 years. After logarithmic transformation, the central portion of the distribution is well approximated by a Gaussian, whereas the tail regions exhibit clear and systematic deviations. In the proposed framework, these deviations are captured by Weibull components. As shown in the right panel of Fig. 1, where the vertical axis is on a logarithmic scale, the model provides an accurate fit in both the central region and the tails. From a nutritional perspective, the left tail corresponds to individuals with undernutrition (e.g., growth impairment), while the right tail reflects excess adiposity (e.g., obesity), indicating the presence of subpopulations deviating from physiological normality. The estimated proportions of these tail components are approximately 0.6% for the left tail and 6.0% for the right tail, both of which appear to be within a plausible range.

This framework provides a principled and interpretable approach for assessing growth abnormalities, with direct relevance to public health monitoring and early intervention. The proposed approach addresses the limitations of conventional tail probability estimation methods that rely on normality assumptions.

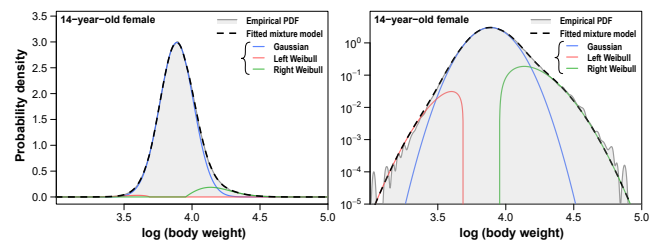


Figure 1: Fit of the proposed Gaussian–Weibull mixture model to log-transformed body weight data of Japanese females aged 14 years. The central region is approximated by a Gaussian component, while the tails are captured by Weibull components. The right panel uses a logarithmic vertical scale to highlight the tail fit.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ACM-BCB 2026, Calabria, Italy

© 2026 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-XXXX-X/2026/06

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>