

Maximum Entropy Distribution of Correlated Variables: Application to Human Height and Weight

Mark P. Silverman^{1,2}

¹G. A. Jarvis Prof. of Physics, Emer., Trinity College, Hartford, USA

²Tall Pines Research, Simsbury, USA

Email: mark.silverman@trincoll.edu, jwmgibbs@gmail.com

How to cite this paper: Silverman, M.P. (2025) Maximum Entropy Distribution of Correlated Variables: Application to Human Height and Weight. *Open Journal of Statistics*, 15, 371-389. <https://doi.org/10.4236/ojs.2025.154020>

Received: July 8, 2025

Accepted: August 3, 2025

Published: August 6, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). <http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Recent investigations have shown that a bivariate lognormal probability density function predicts the statistical moments and correlations of adult human height and weight so extensively and closely as to pose an enigma regarding the underlying reason for such exactness. No genetic or environmental cause currently accounts for this distribution. In this article, it is shown that the Principle of Maximum Entropy (PME)—which is an inferential method drawn exclusively from probability theory—leads to the joint lognormal distribution of height and weight independent of any physical mechanism of biological development. The operation of the PME entails carrying out a variational procedure on a functional comprising the Shannon information entropy subject to constraints posed by known prior information in the form of expectation values. In the case of height and weight the prior information consists of the means, variances, and linear correlation of the logarithms of the two variables. When applied to a large anthropometric survey, the maximum entropy distribution resulting from this variational procedure is shown to be astronomically more probable than any other distribution consistent with the prior information. Although the PME provides *an* explanation of the enigma, the possibility is examined that an underlying stochastic mechanism may also lead to the same distribution.

Keywords

Information Entropy, Principle of Maximum Entropy, Body Mass Index, Variability of Height, Variability of Weight, Correlation of Height and Weight, Lognormal Distribution, Kullback-Leibler Divergence

1. Introduction—A Statistical Enigma

In recent publications, the author derived and discussed the exact probability density functions (PDF) of the body mass index (BMI) [1] [2] and the joint distribution of human height and weight [3]. The BMI is the most widely used medical risk factor for morbidity and mortality related to weight. Categories of risk are set by the World Health Organization (WHO) [4] or by national health organizations like the U.S. National Institutes of Health (NIH) [5]. As a mathematically defined quantity

$$\text{BMI} \equiv W/H^2, \quad (1)$$

in which W is weight in kg and H is height in meters of an individual, the functional form of the BMI PDF can be derived exactly for any and all demographics using analytical methods to transform and combine functions of random variables [6].

However, to specify the BMI distribution of a particular population, one must know the statistical distribution of height and weight for that particular group. Since the height and weight of individuals result from a complex interplay of genetic and environmental determinants, no theoretical expression corresponding to Equation (1) is known from which to deduce an exact, or even an approximate, PDF. In such cases, there are basically two different ways to proceed:

1) Hypothesize the form that the sought-for theoretical distribution should take based on prior information, such as acquired from an anthropometric survey of a sampled population.

2) Apply some fundamental physical or epistemological principle that makes use of prior information and leads by analysis to a unique statistical distribution.

In arriving at the joint probability density function for human height and weight, the author employed both methods. The first method was recently published [3] and led to a bivariate lognormal distribution that, when tested against a large anthropometric survey [7], matched the statistics so closely and so extensively as to pose an enigma regarding the underlying reason. The second method, which is based on the Principle of Maximum Entropy (PME), provides a possible explanation of the apparent exactness of the empirical distribution. This is the subject of the present paper.

1.1. Inference Based on Empirical Evidence

In the first approach, full details of which can be found in Ref. [3], various empirical features of the anthropometric data suggested that height and weight together are distributed as a bivariate lognormal random variable with probability density

$$p_{(H,W)}(h,w) = \frac{\exp\left(-\frac{1}{2(1-r^2)}\left[\left(\frac{\ln(h)-m_1}{s_1}\right)^2 - 2r\left(\frac{\ln(h)-m_1}{s_1}\right)\left(\frac{\ln(w)-m_2}{s_2}\right) + \left(\frac{\ln(w)-m_2}{s_2}\right)^2\right]\right)}{2\pi s_1 s_2 h w \sqrt{1-r^2}} \quad (2)$$

in which upper case letters (H , W) represent random variables, lower case letters (h , w) represent realizations of the variables (e.g. outcomes of measurement or inputs to calculation), and the associated parameters are defined as follows:

$$m_1 = \langle \ln(H) \rangle \quad (3)$$

$$s_1^2 = \text{var}(\ln(H)) = \langle (\ln(H) - m_1)^2 \rangle \quad (4)$$

$$m_2 = \langle \ln(W) \rangle \quad (5)$$

$$s_2^2 = \text{var}(\ln(W)) = \langle (\ln(W) - m_2)^2 \rangle \quad (6)$$

$$r = \text{cov}(\ln(H), \ln(W)) / s_1 s_2 = \left\langle \left(\frac{\ln(H) - m_1}{s_1} \right) \left(\frac{\ln(W) - m_2}{s_2} \right) \right\rangle. \quad (7)$$

Angular brackets $\langle \rangle$ symbolize the expectation value of any discrete function $f_{h,w}$ or continuous function $f(h, w)$

$$\langle f \rangle = \begin{cases} \sum_{h,w} f_{h,w} p_{h,w} \\ \iint f(h, w) p_{H,W}(h, w) dh dw \end{cases} \quad (8)$$

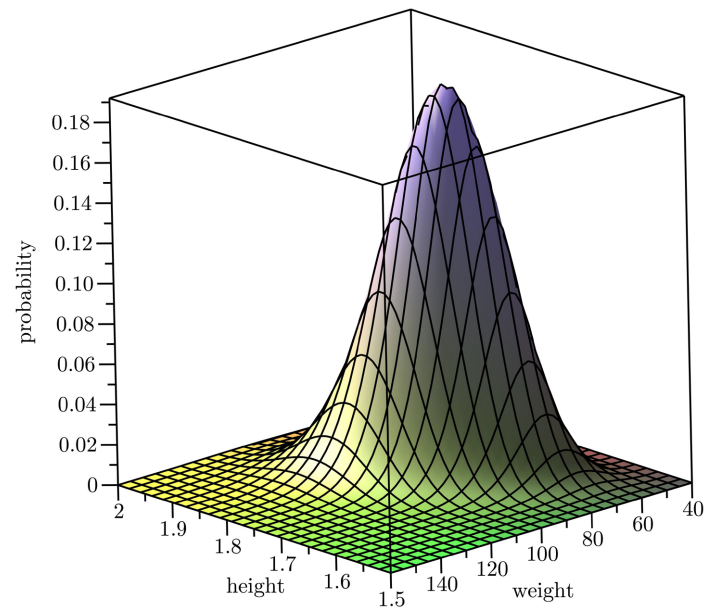
with summation or integration over a range of non-negative real numbers. The parameter r in Equation (7) is referred to as the linear correlation coefficient; it is a measure of the first-order covariance of two random variables [8].

The features of the sample distributions of height and weight that suggested a bivariate lognormal distribution included (1) marked asymmetry about the mean as discerned graphically or measured quantitatively by the statistic skewness (related to the 3rd statistical moment), (2) a nonzero correlation coefficient of height and weight, and especially (3) the fact that histograms of the natural logarithms of the sampled height and weight passed statistical tests for normal (*i.e.* Gaussian) distributions, consistent with the definition of a lognormal distribution [9].

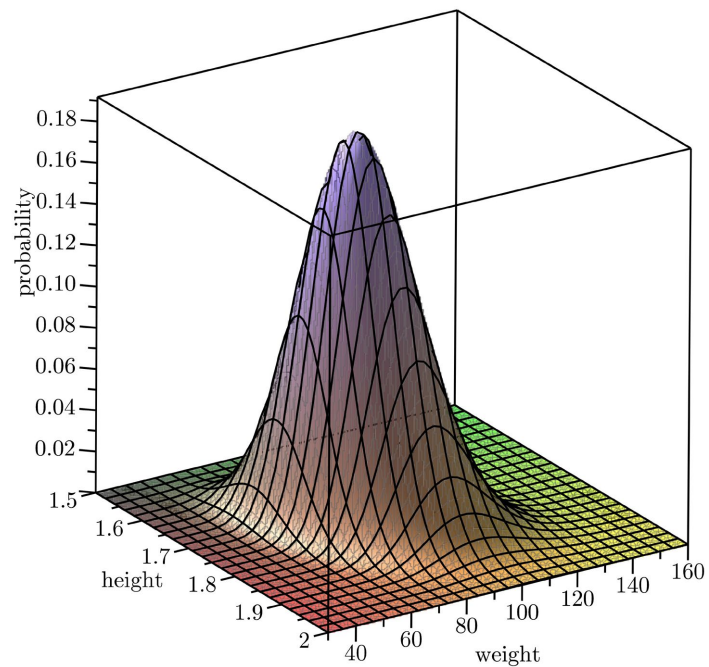
The statistical content of Equation (2) and its marginal distributions were tested comprehensively against an extensive data base of the *Anthropometric Survey of U.S. Army Personnel* (ANSUR) [7]. This survey, which included both genders, 5 categories of age ranging from younger than 20 to older than 41, 7 broad categories of race further classified into approximately 30 ethnic subpopulations, 51 US birthplaces (50 States and Washington DC), as well as approximately 30 international birthplaces, is presumed representative of a large, diverse group of basically healthy adults, since members of the military have to pass fitness requirements for acceptance. Statistical tests were carried out separately for male and female cohorts comprising respectively 4082 and 1986 subjects. **Figure 1** shows a 3-dimensional plot of the distribution function (2) for the set of parameters that apply to the ANSUR male cohort: $m_1 = 4.4351$, $m_2 = 0.5624$, $s_1 = 0.1654$, $s_2 = 0.0390$, $r = 0.4716$. The second frame of the figure shows an orientation rotated 180° relative to the first frame.

For a conjectured distribution with the foregoing characteristics, one might ex-

pect the probability density in Equation (2) to approximate a hypothetical “true” probability density reasonably well. In fact, the outcome of the tests revealed agreement between theoretical predictions and data to an astonishing degree.



(a)



(b)

Figure 1. Plot of the probability density function of the bivariate lognormal distribution of human height and weight for the male ANSUR cohort. Contour lines show lines of variable height for fixed weight and lines of variable weight for fixed height. These two sets of contours are each mutually orthogonal and lognormal in form. The views shown in (a) and (b) differ in orientation by 180° as indicated by the mirror reflection of the colors in the base plane.

Upon input of the numerical values of the five parameters of Equations (3)-(7) obtained for each gender from the ANSUR population, the statistical predictions of Equation (2) matched corresponding sample statistics over an extensive hierarchy of higher moments and correlations limited only by statistical uncertainties due to finite sample size. Especially interesting was the fact that the single *linear* correlation coefficient r , together with the other four parameters, sufficed for correctly predicting all accessible higher-order *nonlinear* correlations of the data.

The full extent of concordances of lower moments, hyperstatistics, and correlation functions of all four variables H , W , $\ln(H)$, $\ln(W)$, together with exhaustive tests for hidden nonlinear correlations beyond those intrinsic to Equation (2), are discussed in Ref. [3]. It seems improbable that such extensive agreement between data and theory is purely coincidental. Consequently, an explanation of the enigmatic perfection of the bivariate lognormal distribution of human height and weight might be attributable to some fundamental principle. Hence the second approach.

1.2. Inference From Probability Theory

The Principle of Maximum Entropy (PME), introduced by Jaynes in the 1950s as a basis for deriving and justifying the fundamental relations of equilibrium statistical mechanics (ESM) [10] [11], provides a means of finding the most probable statistical distribution compatible with known prior information. Although the initial motivation was to solve the physics problem of finding a probabilistic, rather than mechanical, explanation of the success of ESM, the PME has since been developed as a general method of inference applicable to problems as diverse as image analysis [12], detection of cheating [13], extraction of information from crowdsourcing [14], and many other applications in science, engineering, and business [15].

Mathematically, the PME generates a variational procedure by which to maximize the Shannon information entropy function [16] augmented by constraints usually posed in the form of expectation values. The virtue of the PME is that it leads to the most objective (*i.e.* least biased) probability distribution consistent with the given constraints [17]. The information entropy, which is equivalent to the entropy function of ESM up to a universal scalar factor (the Boltzmann constant), is defined by an expression of the form

$$S(\mathbf{p}) \equiv S(p_1 \cdots p_n) = -\sum_{i=1}^n p_i \ln(p_i) \quad (9)$$

in which $p_i (i=1, \dots, n)$ is the probability of an outcome x_i . Equation (9) can be generalized to apply to the entropy of a continuous variable, an extension that will be discussed later in the paper where the distinction becomes relevant. From Equations (8) and (9) it follows that the information entropy may be thought of as the expectation

$$S(\mathbf{p}) \equiv -\langle \ln(\mathbf{p}) \rangle. \quad (10)$$

In Section 2, the maximum entropy distribution of human height and weight is derived. The statistical implications of maximum entropy are taken up in Section 3. An important matter regarding the information entropy of continuous variables is discussed in Section 4. Section 5 outlines a hypothetical stochastic mechanism of biological development that leads to the bivariate lognormal distribution. Some fine points regarding the maximum entropy explanation of the distribution of human height and weight and future steps for further testing are discussed in Section 6.

2. Maximum Entropy Distribution of Height and Weight

It is common practice in statistics to make a logarithmic transformation of data [18], especially in analyses of physical, biomedical, or economic information comprising positive values that display a marked skewness such as human height and weight. The desired outcome of such a transformation is to “normalize” the data—*i.e.* render a more symmetric distribution closer in form to Gaussian.

Consider, therefore, the bivariate random variable

$$(X, Y) \equiv (\ln(H), \ln(W)) \quad (11)$$

with realizations represented by x, y , respectively. Then the entropy of a data set comprising the logarithms of n individual measurements of H and W takes the form

$$S = -\sum_{x,y} p_{x,y} \ln(p_{x,y}). \quad (12)$$

From the transformed data set one can extract the five statistics summarized in the five Equations (3)-(7) and incorporate this information as constraints in an entropy functional

$$\begin{aligned} H(\mathbf{p}) = & -\sum_{x,y} p_{x,y} \ln(p_{x,y}) - \lambda_0 \left(\sum_{x,y} p_{x,y} - 1 \right) - \lambda_1 \left(\sum_{x,y} p_{x,y} \left(\frac{x-m_1}{s_1} \right)^2 - 1 \right) \\ & - \lambda_2 \left(\sum_{x,y} p_{x,y} \left(\frac{y-m_2}{s_2} \right)^2 - 1 \right) - \lambda_3 \left(\sum_{x,y} p_{x,y} \left(\frac{x-m_1}{s_1} \right) \left(\frac{y-m_2}{s_2} \right) - r \right) \end{aligned} \quad (13)$$

by means of the Lagrange multipliers λ_i ($i = 0, 1, 2, 3$). In this paper the symbol “S”, as used in physics, will stand for entropy; the symbol “H”, as used in communication theory, will stand for the entropy functional in the variational procedure. (The symbol “H” actually represents an upper-case Greek Eta for Entropy.)

The choice of what prior information to include in the variational analysis is somewhat arbitrary, as it depends on what prior information is available. The choice is important, for it determines the form of the statistical distribution that emerges from the analysis. This consideration will be discussed in greater detail later in the paper. For the present, suffice it to say that the author’s choice was motivated by the prior knowledge, as reported in Ref. [3], that a bivariate lognormal probability density matched the ANSUR data very closely. Therefore, the in-

initial constraints were chosen to be the five lowest statistical moments that uniquely defined a bivariate lognormal distribution. A different set of prior constraints, however, would have resulted in a different distribution, even though those constraints would have come from the same ANSUR data set.

It is to be noted that Equation (13) does not contain explicit terms for constraints on the two means, Equations (3) and (5). This is not an oversight. Rather, the information provided by those two constraints is already included through the constraints on the two variances. To add to Equation (13) additional terms for constraining the two means would be redundant, and the corresponding Lagrange multipliers would turn out to be zero. It is, in fact, a general characteristic of the maximum entropy method that it recognizes redundant information and eliminates the associated Lagrange multipliers from the PME solution [19].

A question that may arise is how uncertainties in the prior information affect the initial constraints and therefore the resulting solution. The answer is that in applying the PME, one simply treats the expectation values that serve as prior information as given known data. If, for example, the prior information is a set of mean values of the variables, and there is a need to account for the uncertainties in those variables, then *that* information is also to be included in the PME functional as constraints on the associated variances, such as implemented in Equation (13). Does one then need to consider the uncertainties of the variances, and so on up an unending ladder of higher-order uncertainties? The answer is “No”, as explained later in the paper. The PME procedure itself can indicate to the analyst when more (or different) prior information is required.

Given Equation (13), the variation with respect to \mathbf{p}

$$\delta H(\mathbf{p}) = 0 \quad (14)$$

to maximize the entropy subject to the imposed constraints then leads to the equation

$$\ln(p_{x,y}) = -(1 + \lambda_0) - \lambda_1 u(x)^2 - \lambda_2 v(y)^2 - \lambda_3 u(x)v(y) \quad (15)$$

in which

$$u(x) \equiv \frac{x - m_1}{s_1}, \quad v(y) \equiv \frac{y - m_2}{s_2} \quad (16)$$

are recognized as standard normal variables with properties

$$\langle u \rangle = \langle v \rangle = 0 \quad (17)$$

and

$$\langle u^2 \rangle = \langle v^2 \rangle = 1. \quad (18)$$

The solution to Equation (15), expressed in the variables u, v , takes the form

$$p(u, v) = \frac{\exp(-\lambda_1 u^2 - \lambda_2 v^2 - \lambda_3 uv)}{Z(\lambda_1, \lambda_2, \lambda_3)} \quad (19)$$

in which Lagrange multiplier λ_0 was incorporated into the partition function

$$Z(\boldsymbol{\lambda}) \equiv Z(\lambda_1, \lambda_2, \lambda_3) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(-\lambda_1 u^2 - \lambda_2 v^2 - \lambda_3 uv) \, dudv \tag{20}$$

$$= \frac{2\pi}{\sqrt{4\lambda_1\lambda_2 - \lambda_3^2}}$$

whose explicit evaluation in Equation (20) takes account of the continuous nature of the variables. Substitution of $Z(\boldsymbol{\lambda})$ into Equation (19) leads to the probability function

$$p(u, v) = \frac{\sqrt{4\lambda_1\lambda_2 - \lambda_3^2}}{2\pi} \exp(-\lambda_1 u^2 - \lambda_2 v^2 - \lambda_3 uv) \tag{21}$$

The partition function, defined in the first line of Equation (20), is seen to serve as the normalization factor ensuring that $p(u, v)$ obeys the completeness requirement of a probability density

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(u, v) \, dudv = 1 \tag{22}$$

irrespective of the values of the Lagrange multipliers. However, one sees from Equations (8) and (19) that $Z(\boldsymbol{\lambda})$ is also the function that relates the values of the Lagrange multipliers to the values of the constraints through the relations

$$-\frac{\partial \ln(Z(\boldsymbol{\lambda}))}{\partial \lambda_1} = \frac{2\lambda_2}{4\lambda_1\lambda_2 - \lambda_3^2} = \langle u^2 \rangle = 1 \tag{23}$$

$$-\frac{\partial \ln(Z(\boldsymbol{\lambda}))}{\partial \lambda_2} = \frac{2\lambda_1}{4\lambda_1\lambda_2 - \lambda_3^2} = \langle v^2 \rangle = 1 \tag{24}$$

$$-\frac{\partial \ln(Z(\boldsymbol{\lambda}))}{\partial \lambda_3} = \frac{-\lambda_3}{4\lambda_1\lambda_2 - \lambda_3^2} = \langle uv \rangle = r \tag{25}$$

From Equations (23) and (24), it follows that $\lambda_1 = \lambda_2$. And from Equations (23) and (25), it follows that $\lambda_3 = -2r\lambda_1$. Substitution of the preceding expressions for λ_2 and λ_3 into Equation (24) determines λ_1 , leading to the solution of all three multipliers

$$\lambda_1 = \lambda_2 = \frac{1}{2(1-r^2)}; \quad \lambda_3 = \frac{-r}{1-r^2} \tag{26}$$

and the partition function, Equation (20)

$$Z = 2\pi\sqrt{1-r^2} \tag{27}$$

The exact probability function (21) is then determined to be

$$p(u, v) = \frac{1}{2\pi\sqrt{1-r^2}} \exp\left(-\frac{1}{2(1-r^2)}(u^2 + v^2 - 2ruv)\right) \tag{28}$$

or

$$p(x, y) = \frac{\exp\left(-\frac{1}{2(1-r^2)}\left(\left(\frac{x-m_1}{s_1}\right)^2 + \left(\frac{y-m_2}{s_2}\right)^2 - 2r\left(\frac{x-m_1}{s_1}\right)\left(\frac{y-m_2}{s_2}\right)\right)\right)}{2\pi s_1 s_2 \sqrt{1-r^2}} \tag{29}$$

in terms of the original coordinates x, y . That the solution $p(x)$ of the variational Equation (14) actually leads to the absolute *maximum* entropy when substituted into expression (9) has been established in general arguments by Jaynes [20].

Equation (29), which is the maximum entropy solution to the problem of two correlated variables (X, Y) with constraints on their means, variances, and correlation, is recognized to be a bivariate *normal* probability function. This means that this is the most probable density function of (X, Y) that can be constructed subject to the same set of constraints. Upon carrying out the inverse logarithmic transformation to return to the original variables (H, W) of height and weight, one obtains Equation (2), the bivariate *lognormal* probability density. The appearance of the factor $s_1 s_2$ in the denominator of Equation (29) and $h w s_1 s_2$ in the denominator of Equation (2) results from the Jacobians of the transformations derived from

$$p(h, w) dh dw = p(x, y) dx dy = p(u, v) du dv . \quad (30)$$

Since the logarithmic transformation of coordinates and its inverse transformation do not introduce new empirical information, it follows that Equation (2) is the maximum entropy density in coordinates (h, w) , given that Equation (29) is the maximum entropy density in coordinates $(x = \ln(h), y = \ln(w))$. Although the argument is sound, there is a subtle matter regarding the entropy of continuous distributions that will be taken up in Section 4.

3. Implications of Maximum Entropy

The foregoing section employing the PME has established that the bivariate lognormal distribution of human height and weight, previously deduced on the basis of empirical sampling, is derivable by a theoretical procedure. The only empirical input in this procedure is the prior information serving as constraints. Apart from this information, the resulting probability density (2) incorporates no additional assumptions, explicit or implicit, introduced through any hypothetical model. In this regard, the PME is said to generate the most objective solution.

To address the question of whether the PME can actually account for the extraordinary match between the theoretical and sampling distributions of human height and weight, one must understand the probabilistic implications of the PME. For illustrative purposes, consider the male cohort (sample size $N > 4000$) of the *Anthropometric Survey of U.S. Army Personnel* (ANSUR) [7], which was the sampling distribution used in Reference [3] to test the statistical predictions of the hypothesized lognormal distribution of height and weight. For simplicity, divide the x - y plane (with x and y the variables in Equation (29)) into a rectilinear grid of n^2 cells of equal size labeled by indices $(i = 1, \dots, n; j = 1, \dots, n)$. Each cell represents a specified range of heights and weights into which an observed number $n_{i,j}$ of sampled individuals fall. Conservation of participants requires that

$$\sum_{i,j=1}^n n_{i,j} = N . \quad (31)$$

In what follows it will be more convenient to relabel cells sequentially by a single index, rather than by a Cartesian double index. This can be accomplished by labeling a Cartesian cell (x_i, y_j) by the index $k = (j-1)n + i$ in which $k = 1, 2, \dots, n^2$. For examples, cell (x_1, y_1) becomes cell $k = 1$, cell (x_1, y_n) becomes cell $k = n^2 - n + 1$, cell (x_n, y_1) becomes cell $k = n$, and cell (x_n, y_n) becomes cell $k = n^2$. Then n_k is the number of individuals falling into the k^{th} cell, and Equation (31) becomes

$$\sum_{k=1}^K n_k = N \tag{32}$$

with $K = n^2$ the maximum number of cells. The number of ways Ω of distributing the possible values $\{n_k\}$ over the K cells is described by a multinomial distribution

$$\Omega = \frac{N!}{n_1! n_2! \dots n_K!} \tag{33}$$

In physics, Ω is referred to as the volume of phase space. Of the vast number of possible ways to partition $N \gg 1$ values over K cells, the partition most likely to occur is the one that maximizes Ω . Applying to Equation (33) the simplest form of Stirling's approximation [21] of $z!$ for some integer z

$$\ln(z!) \approx z \ln(z) - z, \tag{34}$$

one can show that

$$\frac{1}{N} \ln(\Omega) \approx - \sum_{k=1}^K \frac{n_k}{N} \ln\left(\frac{n_k}{N}\right). \tag{35}$$

In the limit of large (technically infinite) N , the ratio (n_k/N) approaches the probability p_k , whereupon one has to an excellent approximation the relation

$$\frac{1}{N} \ln(\Omega) \xrightarrow{N \gg 1} S(\mathbf{p}) \tag{36}$$

between sampling frequency and entropy. Therefore, finding the distribution of frequencies $\{n_k\}$ that maximizes Ω , Equation (33), is equivalent to finding the distribution of probabilities $\{p_{x,y}\}$ that maximizes the information entropy S , Equation (12). One can therefore rearrange Equation (36) to write

$$\Omega = e^{N S}. \tag{37}$$

Suppose next that Ω_{ME} is the phase space volume of frequencies $\{n_k\}$ with maximum entropy S_{max} , and Ω is the phase space volume of some other distribution having the same constraint Equation (32) with lower entropy S . Then from Equation (37) the ratio $\Omega_{\text{ME}}/\Omega$ in terms of the entropy difference ΔS is given by

$$\frac{\Omega_{\text{ME}}}{\Omega} = \exp(N(S_{\text{max}} - S)) = \exp(N \Delta S). \tag{38}$$

The implications of Equation (38) can be astounding. Consider a sample size comparable to that of the ANSUR male cohort approximated to be 4000, and two distributions that differ in entropy by only 0.01. Then evaluation of (38) leads to

$(\Omega_{\text{ME}}/\Omega) > 10^{17}$. In words, the maximum entropy distribution Ω_{ME} is one hundred thousand million million times more likely to occur by chance than the distribution Ω .

The foregoing estimate is a prediction. If the height and weight of individuals in some group of people are measured, and the only prior information that an analyst has are the means, variances, and correlation coefficient of the logarithms of the two variables, then the most likely distribution to account for the sample is the bivariate lognormal distribution, such as was initially found empirically [3]. The larger the sample size, the more probable is the maximum entropy distribution to be observed in any sampling. As just illustrated, a maximum entropy distribution can be overwhelmingly more probable than any other distribution subject to the same constraints. In such cases, how a probability density with just a few parameters can suffice to predict correctly an extensive array of testable moments and correlations is no longer a mystery.

4. Maximum Entropy of Continuous Random Variables

In Section 2 it was shown that a bivariate lognormal PDF, Equation (2), is the maximum entropy solution to the variational equation for the distribution of human height and weight. The procedure entailed first showing that a logarithmic transformation of the variables led to a bivariate normal PDF, Equation (29). It was then argued that a deterministic (in contrast to stochastic) coordinate transformation and its inverse do not change the uncertainty in knowledge about the state of a system. From this line of reasoning, it followed that the inverse transformation of the PDF back to the original variables likewise described a maximum entropy distribution.

The preceding reasoning is not wrong, but there is a subtle underlying issue that should not be glossed over. Since the two distributions, bivariate normal (BN) and bivariate lognormal (BLN), are both maximum entropy distributions of the *same* system, although expressed in different coordinates, one might have expected that they would have the same value of maximum entropy. There is no fundamental requirement that this be the case, and, in fact, it is *not* the case. Straightforward calculations using PDFs (29) and (2) lead to the entropy expressions:

$$\begin{aligned} \text{Bivariant Normal} \quad S_{\text{BN}} &= \ln \left(2\pi e s_1 s_2 \sqrt{1-r^2} \right) \\ \text{Bivariant Lognormal} \quad S_{\text{BLN}} &= \ln \left(2\pi e s_1 s_2 \sqrt{1-r^2} e^{m_1+m_2} \right) \end{aligned} \quad (39)$$

where it is seen that the BLN entropy depends on all five parameters in contrast to the BN entropy which is independent of the two means.

One can understand the reason for this curious discrepancy—although the expressions in Equation (39) are correct—by examining the general relation between two entropy expressions connected by a coordinate transformation. For illustrative purposes consider a coordinate transformation (which can be multidimensional) $\mathbf{x} \rightarrow \mathbf{y}(\mathbf{x})$ leading to the two entropy expressions $S_{\mathbf{x}}$, $S_{\mathbf{y}}$ defined

by

$$\begin{aligned} S_x &= -\int p_x(\mathbf{x}) \ln(p_x(\mathbf{x})) d\mathbf{x} \\ S_y &= -\int p_y(\mathbf{y}) \ln(p_y(\mathbf{y})) d\mathbf{y} \end{aligned} \quad (40)$$

where the probability densities satisfy

$$p_x(\mathbf{x}) d\mathbf{x} = p_y(\mathbf{y}) d\mathbf{y} \quad (41)$$

and are therefore connected by the transformation

$$p_y(\mathbf{y}(\mathbf{x})) = p_x(\mathbf{x}) \left| \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right| = J p_x(\mathbf{x}) \quad (42)$$

with Jacobian J . Substitution of Equation (42) into Equation (40) leads to the relation

$$S_y = S_x - \int p_x(\mathbf{x}) \ln(J(\mathbf{x})) d\mathbf{x} \quad (43)$$

or more succinctly,

$$S_y = S_x - \langle \ln(J) \rangle. \quad (44)$$

Applied to a bivariate coordinate transformation $(x_1, x_2) \rightarrow (y_1, y_2)$ as a concrete example, Equations (40), (42), and (43) reduce to the explicit form

$$S_y = -\int p_x(x_1, x_2) \ln \left(\frac{p_x(x_1, x_2) dx_1 dx_2}{dy_1 dy_2} \right) dx_1 dx_2 \quad (45)$$

or

$$S_y - S_x = -\int p_x(x_1, x_2) \ln \left(\frac{dx_1 dx_2}{dy_1 dy_2} \right) dx_1 dx_2. \quad (46)$$

Apart from a minus sign, the right side of Equation (46) takes the form of a Kullback-Leibler (K-L) divergence [22] [23], which quantifies the difference between two probability distributions. It is thus seen that the difference in entropies in Equation (46) is attributable to different prior distributions of the two sets of volume elements. S_x is calculated on the presumption that each volume element $d\mathbf{x} \equiv dx_1 dx_2 \cdots dx_n$ (of an n -dimensional system) is equally likely, whereas S_y is calculated presuming each volume element $d\mathbf{y} \equiv dy_1 dy_2 \cdots dy_n$ is equally likely. However, since \mathbf{y} is a function of \mathbf{x} , the volume elements of the two systems of coordinates cannot in general both be uniformly distributed. Hence the coordinate transformation $\mathbf{x} \rightarrow \mathbf{y}(\mathbf{x})$ leads to a difference in entropies.

The fact that entropy S_y in Equation (43) or (44) differs in *value* from entropy S_x does not mean that *variations* of the corresponding entropy functionals H_x , H_y —from which the maximum entropy PDFs are obtained—are necessarily different. Consider the logarithmic transformation $(X = \ln(H), Y = \ln(W))$ of the variables (H, W) and the inverse transformation $(H = e^X, W = e^Y)$. The Jacobian of the transformation is

$$J = \left| \frac{\partial(X, Y)}{\partial(H, W)} \right| = \frac{1}{HW} = \exp(-(X + Y)). \quad (47)$$

Substitution of Equation (47) into Equation (44) leads to the relation

$$S_{H,W} = S_{X,Y} - \langle \ln(J) \rangle = S_{X,Y} - \langle X \rangle - \langle Y \rangle. \quad (48)$$

Recall that the two expectation values on the right side of Equation (48) are respectively m_1 and m_2 , which are part of the prior information, Equations (3) and (5). Given the entropy functional $H_{X,Y}$ of Equation (13), the functional $H_{H,W}$ associated with entropy $S_{H,W}$ of Equation (48) would then include two additional constraints and take the form

$$H_{H,W} = H_{X,Y}(p_{x,y}) - \lambda_4 \sum_{x,y} x p_{x,y} - \lambda_5 \sum_{x,y} y p_{x,y} \quad (49)$$

where $H_{X,Y}(p_{x,y})$ includes the terms with Lagrange multipliers λ_j ($j = 0, 1, 2, 3$). However, information on the mean values of X and Y is already included in the constraints on the variances of X and Y in $H_{X,Y}(p_{x,y})$. Thus the two additional terms in Equation (49) constitute redundant information, whereupon the Lagrange multipliers λ_4 and λ_5 would vanish. It is then apparent that the variation $\delta H_{H,W} = 0$ since $\delta H_{X,Y} = 0$, and, as expected, the bivariate lognormal PDF is the maximum entropy solution in coordinates (h, w) given that the bivariate normal PDF is the maximum entropy solution in coordinates $(x = \ln(h), y = \ln(w))$.

As a general procedure consistent with Bayesian principles, Jaynes [17] proposed a *relative* entropy of the form

$$S_{\text{rel}} = \begin{cases} -\int p(\mathbf{x}) \ln(p(\mathbf{x})/m(\mathbf{x})) d\mathbf{x} & \text{(continuous)} \\ -\sum_x p_x \ln(p_x/m_x) & \text{(discrete)} \end{cases} \quad (50)$$

for either continuous or discrete random variables. This again is the form of a K-L divergence. The background distribution $m(\mathbf{x})$ is a measure that transforms under a change of coordinates in the same way as the probability function $p(\mathbf{x})$, whereupon expression (50) is invariant under a coordinate change.

As employed in applications of the PME, the function $m(\mathbf{x})$ can be interpreted as a Bayesian prior, *i.e.* the distribution representing a state of complete ignorance before any empirical information has been acquired [24]. How one determines the appropriate measure for an arbitrary statistical problem is, in general, still a work in progress. In some cases, especially for a finite system of discrete random variables (see [13] for a worked example), the correct measure may be obvious. In more complex cases involving continuous random variables, Jaynes proposed the use of group theoretical methods to deduce $m(\mathbf{x})$ from the appropriate transformation group that describes the invariances of a statistical problem [24].

It might be thought that the obvious choice of $m(\mathbf{x})$ would be a uniform distribution over the full range of possible outcomes \mathbf{x} , since that certainly depicts a state of total ignorance. However, a uniform distribution is not always suitable or even possible, especially for a system of continuous random variables. First, a uniform distribution is not normalizable if the range of the variables is infinite. And

second, being independent of coordinates, a uniform distribution does not change under a coordinate transformation, so that the expression (50) would again not be invariant.

It is rarely the case, however, that an analyst is in a state of *total* ignorance prior to data collection. For example, in the problem of the distribution of human height and weight, one can be reasonably assured that no adult human ever had a height exceeding 10 m or a weight exceeding 1000 kg. Thus, a suitable approximation to a uniform distribution could be a Gaussian measure $m(\mathbf{x})$ with uncertainty sufficiently broad that the center is irrelevant and the function varies very slowly over the practical range of the variables.

5. A Stochastic Model of Growth

Demonstration that an analysis based on the Principle of Maximum Entropy leads naturally to a bivariate lognormal PDF of height and weight does not mean there can be no stochastic physical mechanism that can also generate this statistical distribution. If future anthropometric surveys of larger cohorts are implemented and continue to sustain the bivariate lognormal as the effectively exact distribution of height and weight, there may well be an underlying physical reason. Here is one possible mechanism by which such a distribution can develop.

Suppose, for example, a study of the genetic basis of human height were to reveal that the variable H could be represented by the geometric mean

$$H = \left(\prod_{i=1}^n X_i \right)^{\frac{1}{n}} \quad (51)$$

of a product of independent random variables $\{X_i\}$ ($i=1, \dots, n$) of finite non-vanishing means and variances. Then the logarithm of H would take the form of a *sum* of independent variables $\{\ln(X_i)\}$

$$\ln(H) = \frac{1}{n} \sum_{i=1}^n \ln(X_i) \rightarrow N(M_H, \Sigma_H^2), \quad (52)$$

which, under conditions for which the Central Limit Theorem [25] is applicable, can be approximated by a normal distribution of resulting mean and variance

$$M_H = \frac{1}{n} \sum_{i=1}^n \langle \ln(X_i) \rangle, \quad \Sigma_H^2 = \text{var}(M_H) \quad (53)$$

as indicated by the arrow in Equation (52). From Equations (52) and (53) it would then follow that the marginal distribution of H is lognormal in form,

$H \rightarrow \Lambda(M_H, \Sigma_H^2)$. (Upper-case lambda is symbolic of a lognormal distribution.)

Suppose further that the variable W is likewise found to be representable as the geometric mean of independent variables $\{Y_i\}$ ($i=1, \dots, n$) with each factor Y_i correlated with the corresponding variable X_i . (For example, suppose that the genes associated with each pair of variables (X_i, Y_i) are located on the same chromosomes.) Then, by the previous argument, the marginal distribution of W would also be lognormal in form, and the full statistics of the two correlated vari-

ables (H, W) would be described by a bivariate lognormal distribution, as is currently found to be the case.

The foregoing mechanism, entailing both a physical cause (biological growth as a product of independent random factors) and a probabilistic reduction (the Central Limit Theorem) is at present entirely hypothetical. As of this writing, the author is not aware of any published studies examining the influence of genetics and/or environment that could account for the statistical distribution of adult human height and weight.

6. Conclusions: Current Status and Future Steps

The objective of this paper has been to account by means of the Principle of Maximum Entropy (PME) for the extraordinary predictive capacity of the bivariate lognormal probability density function (PDF) of human height and weight [3]. Currently, there is no known physical reason rooted in the biology of human development for the exactness of this distribution. However, the maximum entropy distribution, derived in Section 2 and elaborated in Section 3, is the most probable distribution consistent with constraints imposed by prior information. In the case of human height and weight, the prior information comprised the means, variances, and linear correlation of the logarithms of height and weight.

The PME constitutes a general inferential procedure derived from probability theory and not associated with any specific physical agency. Loosely summarized in words, the PME says that the distribution that maximizes the entropy, subject only to the known information provided at the outset, is the distribution that is most likely to occur in any sampling of the chosen variables. As illustrated in Section 3, the probability of this occurrence can be astronomically greater than the occurrence of any other distribution likewise subject to the same prior information. Thus, the maximum entropy distribution can appear to be an effectively exact distribution for no discernible physical reason.

A question might arise as to whether one can improve the predictive capacity of a maximum entropy distribution by supplying a more inclusive set of prior information. Actually, the PME itself informs a user whether more or different information is needed or not. If the resulting PME solution satisfactorily accounts for a given set of data, then providing more prior information is not likely to yield a statistically significant improvement. However, more or different information is needed if the PME solution does not adequately account for the data. Moreover, if the PME variational procedure leads to no solution, then the prior information is either mutually inconsistent, or else does not reflect the conditions of the experiment or observations that generated the data. Further scrutiny of the physical nature of the problem in question could then suggest to the user what revisions to make.

As illustration of the preceding comments, consider first the case of equilibrium statistical mechanics (ESM) initially derived by J. Willard Gibbs in the 19th Century and given a modern PME justification by E. T. Jaynes [10]. Application of the

PME with prior information comprising just the *mean values* of the internal energy and the number of particles in the system leads to a probability density function of exponential form. From this PDF one can predict correctly a wide range of macroscopic equilibrium properties of both classical and quantum systems, including the fluctuations (*i.e.* predictive uncertainties) of the variables. Since ESM is ordinarily applied to systems with an enormous number N of degrees of freedom (e.g. $N \sim 10^{24}$ for 1 mol of a monatomic gas), no further prior information beyond the expectation of energy and particle number is required, since the variances of the means decrease as $1/N$.

However, for the problem of the distribution of human height (H) and weight (W), prior information consisting of just the mean values $(\langle H \rangle, \langle W \rangle)$ will not suffice. This would lead to an *exponential* PDF for (H, W) , which does not match the data. Including in the prior information the variances $(\text{var}(H), \text{var}(W))$ will also not suffice. This would lead to a *normal* PDF for (H, W) , whose sample distribution is *not* symmetric about the means. Only when the prior information comprises the means and variances and linear correlation of $(\ln(H), \ln(W))$ does the resulting bivariate *lognormal* probability density permit accurate prediction of all testable statistics extractable from the ANSUR data.

It may at first thought seem strange that so small a set of prior information makes it possible to predict correctly the *asymmetries* of the (H, W) sample distribution without explicit inclusion in the entropy functional of additional terms proportional to the third moments (skewness) of any of the variables. To have included such terms would have greatly complicated the analysis, since the resulting distributions would no longer be lognormal in form or even be reducible to a closed form expression. However, the extra work would have achieved little if anything, since contributions of skewness to the entropy functional decrease as $1/N^2$ for a sample size N . Recall that $N > 4000$ for the ANSUR male cohort. *Thus, for large sample size N the PME analysis of height and weight no more needed prior information on skewness than did the PME analysis of equilibrium statistical mechanics need prior information on variance.*

Consequently, at this present stage in the statistical analysis of human height and weight, there would be no compelling reason to modify the prior information without first having more extensive data. Because the sampling variance of a moment depends on the population moment of *twice* the order [26], the uncertainties of higher-order statistical moments (hyperstatistics) increase rapidly with order, and tests of the maximum entropy PDF to predict such statistics will require correspondingly larger sample sizes than the large ANSUR data base used in References [2] and [3].

From the foregoing comments on prior information, sample size, and PME solutions, one can draw an important epistemological lesson. Expressed loosely, the probability density function does not describe attributes of the real world; rather, it describes one's *state of knowledge* about the real world. Thus, the idea that there

exists a unique “true” probability density function of human height and weight which may emerge from surveys of ever larger populations is illusory. What can emerge, of course, is a more refined PDF that attains greater predictive capability. Nevertheless, it is conceivable that more than one PDF, each a PME solution for a different set of prior information, can account for the same set of data with comparable success. This does not appear to be the case for human height and weight, although the author has encountered such a situation in his analysis of the probability of infectivity of SARS-Cov 2 (COVID) [27].

Author

Dr. M. P. Silverman is the G. A. Jarvis Professor of Physics Emeritus at Trinity College and senior scientist at Tall Pines Research. His areas of research are in nuclear and medical physics. This article was conceived and written by him alone with no reliance on artificial intelligence.

Disclaimer

The author is not affiliated with any commercial companies or organizations and has received no compensation for the composition of this article.

Acknowledgements

The author expresses his gratitude to Dr. S. B. Brachwitz for numerous discussions on matters relating to biological development. He would also like to thank the reviewers for their helpful comments.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Silverman, M.P. and Lipscombe, T.C. (2022) Exact Statistical Distribution of the Body Mass Index (BMI): Analysis and Experimental Confirmation. *Open Journal of Statistics*, **12**, 324-356. <https://doi.org/10.4236/ojs.2022.123022>
- [2] Silverman, M.P. (2025) Perspective on the Body Mass Index (BMI) and Variability of Human Weight and Height. *Journal of Biosciences and Medicines*, **13**, 309-320. <https://doi.org/10.4236/jbm.2025.136026>
- [3] Silverman, M.P. (2022) Exact Statistical Distribution and Correlation of Human Height and Weight: Analysis and Experimental Confirmation. *Open Journal of Statistics*, **12**, 743-787. <https://doi.org/10.4236/ojs.2022.125044>
- [4] World Health Organization, Body Mass Index (BMI). <https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/body-mass-index>
- [5] Weir, C. and Jan, A. (2023) BMI Classification and Cut-Off Points. <https://www.ncbi.nlm.nih.gov/books/NBK541070/>
- [6] Silverman, M.P. (2014) *A Certain Uncertainty: Nature’s Random Ways*. Cambridge University Press. <https://doi.org/10.1017/cbo9781139507370>

- [7] Gordon, C.C., *et al.* (2012) Technical Report Natick/TR-15/007, Anthropometric Survey of U.S. Army Personnel: Methods and Summary Statistics. U.S. Army Natick Soldier Research, Development and Engineering Center.
- [8] Mood, A.M., Graybill, F.A. and Boes, D.C. (1974) Introduction to the Theory of Statistics. 3rd Edition, McGraw-Hill, 155-156.
- [9] Forbes, C., Evans, M., Hastings, N. and Peacock, B. (2010) Statistical Distributions. Wiley. <https://doi.org/10.1002/9780470627242>
- [10] Jaynes, E.T. (1957) Information Theory and Statistical Mechanics. *Physical Review*, **106**, 620-630. <https://doi.org/10.1103/physrev.106.620>
- [11] Jaynes, E.T. (1957) Information Theory and Statistical Mechanics. II. *Physical Review*, **108**, 171-190. <https://doi.org/10.1103/physrev.108.171>
- [12] Skilling, J. and Bryan, R.K. (1984) Maximum Entropy Image Reconstruction: General Algorithm. *Monthly Notices of the Royal Astronomical Society*, **211**, 111-124. <https://doi.org/10.1093/mnras/211.1.111>
- [13] Silverman, M.P. (2015) Cheating or Coincidence? Statistical Method Employing the Principle of Maximum Entropy for Judging Whether a Student Has Committed Plagiarism. *Open Journal of Statistics*, **05**, 143-157. <https://doi.org/10.4236/ojs.2015.52018>
- [14] Silverman, M.P. (2019) Extraction of Information from Crowdsourcing: Experimental Test Employing Bayesian, Maximum Likelihood, and Maximum Entropy Methods. *Open Journal of Statistics*, **9**, 571-600. <https://doi.org/10.4236/ojs.2019.95038>
- [15] Levine, R.D. and Tribus, M. (1978) The Maximum Entropy Formalism. MIT Press.
- [16] Shannon, C.E. and Weaver, W. (1964) The Mathematical Theory of Communication. The University of Illinois Press.
- [17] Jaynes, E.T. (1989) Where Do We Stand on Maximum Entropy? (1978). In: Rosenkrantz, R.D., Ed., *E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics*, Kluwer Academic Publishers, 210-314. https://doi.org/10.1007/978-94-009-6581-2_10
- [18] Altman, D.G. (1999) Practical Statistics for Medical Research. Chapman & Hall/CRC, 36-37, 143-146.
- [19] Jaynes, E.T. (1989) Prior Probabilities (1968). In: Rosenkrantz, R.D., Ed., *E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics*, Kluwer Academic Publishers, 114-130. https://doi.org/10.1007/978-94-009-6581-2_7
- [20] Jaynes, E.T. (1963) Information Theory and Statistical Mechanics, in the Brandeis University Summer Institute Lectures in Theoretical Physics. W. A. Benjamin, 188-189.
- [21] Arfken, G.B. and Weber, H.J. (2005) Mathematical Methods for Physicists. 6th Edition, Elsevier, 489-497.
- [22] Kullback, S. (1968) Information Theory and Statistics. Dover Publications, 1-31.
- [23] Lang, N. (2024) What Is the Kullback-Leibler Divergence? <https://databasecamp.de/en/statistics/kullback-leibler-divergence>
- [24] Jaynes, E.T. (2003) Probability Theory: The Logic of Science. Cambridge University Press, 374-386.
- [25] Martin, B.R. (1971) Statistics for Physicists. Academic Press, 42-49.
- [26] Kendall, M.G. and Stuart, A. (1963) The Advanced Theory of Statistics Vol 1: Distribution Theory. 2nd Edition, Hafner, 234.

- [27] Silverman, M.P. (2023) Probability Distribution of SARS-CoV-2 (COVID) Infectivity Following Onset of Symptoms: Analysis from First Principles. *Open Journal of Statistics*, **13**, 233-263. <https://doi.org/10.4236/ojs.2023.132013>