Maximum-entropy distributions of correlated variables with prespecified marginals

Hernán Larralde

Instituto de Ciencias Físicas, UNAM, Apartado Postal 48-3, Codigo Postal 62251, Cuernavaca, Morelos, Mexico

(Received 6 August 2012; published 13 December 2012)

The problem of determining the joint probability distributions for correlated random variables with prespecified marginals is considered. When the joint distribution satisfying all the required conditions is not unique, the "most unbiased" choice corresponds to the distribution of maximum entropy. The calculation of the maximum-entropy distribution requires the solution of rather complicated nonlinear coupled integral equations, exact solutions to which are obtained for the case of Gaussian marginals; otherwise, the solution can be expressed as a perturbation around the product of the marginals if the marginal moments exist.

DOI: 10.1103/PhysRevE.86.061117 PACS number(s): 02.50.Tt, 05.40.−a, 05.10.Gg

Consider the situation in which we are given two random variables, say $X_1 \in I_1$ and $X_2 \in I_2$, which we know to be distributed as $P_1(X_1)$ and $P_2(X_2)$, respectively. Further, assume we know the variables to be correlated; for example, assume we are given the covariance $\Gamma_{12} = \langle X_1 X_2\rangle - \langle X_1\rangle\langle X_2\rangle \neq 0$. We are now required to construct the joint probability distribution $P_{(1,2)}(X_1,X_2)$ with the prescribed marginals $P_1(X_1)$ and $P_2(X_2)$, and covariance Γ_{12}.

This and other similar problems arise in a wide variety of contexts, ranging from the description of correlated financial instruments in economics [1], electroencephalogram signals [2,3] in medicine, to systems out of equilibrium in statistical mechanics [4–7]; to name but very few. Actually, in finance and other fields of intense applied statistics [8–11], it has become popular to describe interdependent random variables with given marginals using "copulas." The idea there is that the "interdependence" of, say, N random variables described by cumulative marginal distributions $F_i(X_i)$ is encoded in the N-dimensional cumulative distribution function with uniform marginals, the copula $C(u_1,\dots,u_N): [0,1]^N \to [0,1]$, with $C(1,1,\dots,u_j,\dots,1) = u_j$. The complete description is achieved through the joint cumulative distribution $F(X_1,\dots,X_N) = C(F_1(X_1),\dots,F_N(X_N))$. This approach treats the individual statistics of the random variables, the marginals, separately from the interdependence of the said variables, allowing one, for example, to change the marginals while keeping the interdependence, the copula, fixed. These tools are extremely powerful and general, but hard to estimate directly from data. Another characterization of data interdependence relates to whether there is a *causal* relation between the variables, which can be tested along the lines originally proposed by Granger in the context of econometrics [12]. In contrast to these more sophisticated methods, and while far from a complete description of the interdependence structure of much data, the correlation between two random variables is a frequently and easily measured quantity, extensively used in many disciplines, including, of course, physics.

As it stands, however, the problem may be ill posed since there could be infinitely many distributions (or none at all) satisfying the conditions of the prespecified marginals and given covariance. To lift the ambiguity, when it arises, we follow Jaynes [13] and require the joint distribution function to be that which maximizes the relative entropy or, equivalently, minimizes the discrimination information over the product of the marginals. This choice is, as argued by Jaynes, the "least biased" distribution which is consistent with the restrictions: "the maximization of entropy is [$\cdots$] a method of reasoning which ensures that no unconscious assumptions have been introduced." [13] To this point, then, the problem is formally straightforward: we need to find the extreme of the entropy functional subject to the appropriate restrictions. That is, we require $P_{(1,2)}(x,y)$ such that

$$0 = \delta\left[\int_{I_1\times I_2} P_{(1,2)}(X_1,X_2)\ln\left(\frac{P_{(1,2)}(X_1,X_2)}{P_1(X_1)P_2(X_2)}\right)dX_1dX_2 + \lambda_{12}\int_{I_1\times I_2} X_1X_2P_{(1,2)}(X_1,X_2)dX_1dX_2 + \int_{I_1\times I_2}[a(X_1)+b(X_2)]P_{(1,2)}(X_1,X_2)dX_1dX_2\right] \quad (1)$$

(we do not need to condition the distribution to be normalized, as the marginals are assumed to be already normalized). The required distribution can be written as

$$P_{(1,2)}(X_1,X_2) = P_1(X_1)P_2(X_2)e^{-a(X_1)-b(X_2)-\lambda_{12}X_1X_2-1} \equiv P_1(X_1)P_2(X_2)\mathcal{A}(X_1)\mathcal{B}(X_2)e^{-\lambda_{12}X_1X_2}. \quad (2)$$

The Lagrange multipliers $a(X_1)$ and $b(X_2)$ [or equivalently the functions $\mathcal{A}(X_1)$ and $\mathcal{B}(X_2)$], and the constant λ_{12}, are chosen to enforce the restrictions, which results in the set of coupled nonlinear integral equations

$$\begin{aligned} 1 &= \mathcal{A}(X_1)\int_{I_2} P_2(X_2)\mathcal{B}(X_2)e^{-\lambda_{12}X_1X_2}dX_2, \quad X_1 \in I_1,\\ 1 &= \mathcal{B}(X_2)\int_{I_1} P_1(X_1)\mathcal{A}(X_1)e^{-\lambda_{12}X_1X_2}dX_1, \quad X_2 \in I_2, \end{aligned} \quad (3)$$

plus a condition on the value of λ_{12}:

$$\int_{I_1\times I_2} X_1X_2P_1(X_1)P_2(X_2)\mathcal{A}(X_1)\mathcal{B}(X_2)e^{-\lambda_{12}X_1X_2}dX_1dX_2 = \Gamma_{12} + \langle X_1\rangle\langle X_2\rangle, \quad (4)$$

where the mean values $\langle X_1\rangle$ and $\langle X_2\rangle$ are calculated from the corresponding marginals. The above equations may be rewritten in the slightly more compact form

$$\begin{aligned} P_1(X_1) &= Q_1^{(2)}(X_1)\int_{I_2} Q_2^{(2)}(X_2)e^{-\lambda_{12}^{(2)}X_1X_2}dX_2, \quad X_1 \in I_1,\\ P_2(X_2) &= Q_2^{(2)}(X_2)\int_{I_1} Q_1^{(2)}(X_1)e^{-\lambda_{12}^{(2)}X_1X_2}dX_1, \quad X_2 \in I_2, \end{aligned} \quad (5)$$

and

$$\int_{I_1\times I_2} X_1X_2Q_1^{(2)}(X_1)Q_2^{(2)}(X_2)e^{-\lambda_{12}^{(2)}X_1X_2}dX_1dX_2 = \Gamma_{12}, \quad (6)$$

where $Q_1^{(2)}(X_1) = P_1(X_1)\mathcal{A}(X_1)$ and $Q_2^{(2)}(X_2) = P_2(X_2)\mathcal{B}(X_2)$. The superscripts have been added to indicate that these quantities are elements of the joint distribution of two variables.

Though far from opening the way to a full solution, it is worthwhile noting that Eqs. (5) can be decoupled by multiplying the first one, say, by $e^{-\lambda_{12}^{(2)}X_1Y}$ and integrating over X_1. Then

$$\begin{aligned}\tilde{P}_1\left(\lambda_{12}^{(2)}Y\right) &\equiv \int_{I_1} P_1(X_1)e^{-\lambda_{12}X_1Y}dX_1\\ &= \int_{I_2} Q_2^{(2)}(X_2)\int_{I_1} Q_1^{(2)}(X_1)e^{-\lambda_{12}^{(2)}(X_2+Y)X_1}dX_1dX_2\\ &= \int_{I_2}\frac{Q_2^{(2)}(X_2)P_2(X_2+Y)}{Q_2^{(2)}(X_2+Y)}dX_2, \quad (7)\end{aligned}$$

where $\tilde{P}_1(Y)$ is the Laplace transform of $P_1(X)$. This is a rather difficult nonlinear integral equation that determines $Q_2^{(2)}(X_2)$, up to a multiplicative constant, in terms of the marginals and the covariance.

If the variables are discrete rather than continuous, similar expressions are obtained with summations instead of integrals. Either way, the above equations turn out to be extremely hard to solve for arbitrary marginals.

The generalization to more than two variables, while formally equally simple, gives rise to a rather interesting situation. To illustrate this, consider the case of three variables X_1, X_2, and X_3 with their respective marginals $P_1(X_1)$, $P_2(X_2)$, and $P_3(X_3)$; and covariance matrix elements Γ_{12}, Γ_{23}, and Γ_{13} (in passing, note that only the off-diagonal components of the covariance matrix can be introduced as constraints; the diagonal elements are fixed by the marginals). If we follow the procedure outlined above for two variables—maximizing entropy relative to the product of the marginals, constrained to the appropriate marginals and correlations—it is easy to see that the joint probability distribution should be of the following form:

$$\begin{aligned}&P_{(1,2,3)}(X_1,X_2,X_3)\\ &\quad = Q_1^{(3)}(X_1)Q_2^{(3)}(X_2)Q_3^{(3)}(X_3)e^{-\lambda_{12}^{(3)}X_1X_2-\lambda_{23}^{(3)}X_2X_3-\lambda_{13}^{(3)}X_1X_3},\end{aligned} \quad (8)$$

where the functions $Q_i^{(3)}(X_1)$ are related to the Lagrange multipliers that constrain the marginals,

$$\begin{aligned}P_1(X_1) = Q_1^{(3)}(X_1)\int_{I_2\times I_3} &Q_2^{(3)}(X_2)Q_3^{(3)}(X_3)\\ &\times e^{-\lambda_{12}^{(3)}X_1X_2-\lambda_{23}^{(3)}X_2X_3-\lambda_{13}^{(3)}X_1X_3}dX_2dX_3, \quad (9)\end{aligned}$$

and so on.

Now the question arises as to whether it must also be true that the distribution obtained by integrating $P_{(1,2,3)}(X_1,X_2,X_3)$ over X_3, say, must have the form of the constrained maximum-entropy joint distribution for two variables discussed above. That is, whether

$$\begin{aligned}&\int_{I_3} P_{(1,2,3)}(X_1,X_2,X_3)dX_3\\ &\quad = Q_1^{(3)}(X_1)Q_2^{(3)}(X_2)e^{-\lambda_{12}^{(3)}X_1X_2}\\ &\quad\quad \times\int_{I_3} Q_3^{(3)}(X_3)e^{-(\lambda_{23}^{(3)}X_2+\lambda_{13}^{(3)}X_1)X_3}dX_3\\ &\quad = Q_1^{(2)}(X_1)Q_2^{(2)}(X_2)e^{-\lambda_{12}^{(2)}X_1X_2} = P_{(1,2)}(X_1,X_2). \quad (10)\end{aligned}$$

If so, this would in turn imply that, independently of what the marginals $P_i(X_i)$ are, the integral that appears above can always be resolved as

$$\begin{aligned}&\int_{I_3} Q_3^{(3)}(X_3)e^{-(\lambda_{23}^{(3)}X_2+\lambda_{13}^{(3)}X_1)X_3}dX_3\\ &\quad = q_1^{(3)}(X_1)q_2^{(3)}(X_2)e^{-\mu_{12}^{(3)}X_1X_2}, \quad (11)\end{aligned}$$

in terms of which we can express $Q_1^{(2)}(X_1) = q_1^{(3)}(X_1)Q_1^{(3)}(X_1)$, $Q_2^{(2)}(X_2) = q_2^{(3)}(X_2)Q_2^{(3)}(X_2)$, and $\lambda_{12}^{(2)} = \lambda_{12}^{(3)} + \mu_{12}^{(3)}$.

While it turns out that for the simple cases considered in this paper, Eq. (10) is indeed satisfied, I cannot find any reason why it should be true in general. Actually, let me consider another joint distribution $\Pi_{(1,2,3)}(X_1,X_2,X_3)$ defined to be the maximum-entropy distribution conditioned so that integrating over any variable yields the corresponding two-point maximum-entropy distributions discussed above:

$$\begin{aligned}0 = \delta\Bigg[&\int_{I_1\times I_2\times I_3}\Pi_{(1,2,3)}(X_1,X_2,X_3)\\ &\times\ln\left(\frac{\Pi_{(1,2,3)}(X_1,X_2,X_3)}{p_1(X_1)p_2(X_2)p_3(X_3)}\right)dX_1dX_2dX_3\\ &+\int_{I_1\times I_2\times I_3}[\alpha_{(1,2)}(X_1,X_2)+\alpha_{(2,3)}(X_2,X_3)\\ &+\alpha_{(1,3)}(X_1,X_3)]\Pi_{(1,2,3)}(X_1,X_2,X_3)dX_1dX_2dX_3\Bigg],\end{aligned}$$

the solution to which can be written as

$$\begin{aligned}&\Pi_{(1,2,3)}(X_1,X_2,X_3)\\ &\quad = F_{(1,2)}(X_1,X_2)F_{(2,3)}(X_2,X_3)F_{(1,3)}(X_1,X_3), \quad (12)\end{aligned}$$

where the functions $F_{(i,j)}(X_i,X_j)$ are simply related to the Lagrange multipliers $\alpha_{(i,j)}(X_i,X_j)$ and satisfy nonlinear integral equations that enforce the conditions imposed on $\Pi_{(1,2,3)}(X_1,X_2,X_3)$, namely,

$$\begin{aligned}&P_{(1,2)}(X_1,X_2)\\ &\quad = F_{(1,2)}(X_1,X_2)\int_{I_3} F_{(2,3)}(X_2,X_3)F_{(1,3)}(X_1,X_3)dX_3,\end{aligned} \quad (13)$$

and so on. The conditions on the distributions $P_{(i,j)}(X_i,X_j)$ ensure that $\Pi_{(1,2,3)}(X_1,X_2,X_3)$ has the correct one-point marginals $P_i(X_i)$ for $i = 1,2,3$, and covariance $\Gamma_{i,j}$. It should be noted that $P_{(1,2,3)}(X_1,X_2,X_3)$, as expressed in Eq. (8), can be written in the form shown in Eq. (12). However, the conditions on $\Pi_{(1,2,3)}(X_1,X_2,X_3)$ are as restrictive or more so than those on $P_{(1,2,3)}(X_1,X_2,X_3)$; thus, it should not necesarily be the case that $P_{(1,2,3)}(X_1,X_2,X_3) = \Pi_{(1,2,3)}(X_1,X_2,X_3)$. Conversely, it does not appear to be necessarily true that

$$P_{(1,2)}(X_1,X_2) = \int_{I_3} P_{(1,2,3)}(X_1,X_2,X_3)dX_3, \qquad (14)$$

etc., so that one could end with the somewhat uncomfortable situation in which the two-point marginals obtained from the maximum-entropy three-point distribution may themselves not be maximum-entropy two-point distributions.

We now turn to simple cases for which the required maximum-entropy distributions can be calculated explicitly. First, however, for the trivial case of "uncorrelated" variables (i.e., the case in which $\Gamma_{i,j} = 0$), the required maximum-entropy joint distribution is indeed the product of the marginals. The first nontrivial example is the case of two correlated random variables with Gaussian marginal distributions, say,

$$P_1(X_1) = \sqrt{\frac{\alpha}{2\pi}} e^{-\alpha X_1^2/2}, \qquad (15)$$

$$P_2(X_2) = \sqrt{\frac{\beta}{2\pi}} e^{-\beta X_2^2/2}, \qquad (16)$$

and let the correlation parameter be $\Gamma = \langle X_1 X_2 \rangle$. Then, to determine the joint distribution we need to solve

$$\sqrt{\frac{\alpha}{2\pi}} e^{-\alpha X_1^2/2} = Q_1(X_1) \int_{-\infty}^{\infty} Q_2(X_2) e^{-\lambda X_1 X_2} dX_2, \quad (17)$$

$$\sqrt{\frac{\beta}{2\pi}} e^{-\beta X_2^2/2} = Q_2(X_2) \int_{-\infty}^{\infty} Q_1(X_1) e^{-\lambda X_1 X_2} dX_1, \quad (18)$$

where the sub- and superscripts have been dropped for notational lightness. Substituting $Q_1(X_1)$ and $Q_2(X_2)$ by Gaussians, it is easy to see that the required maximum-entropy joint distribution is

$$P_{(1,2)}(X_1,X_2) = \frac{1}{2\pi}\left(\frac{\alpha\beta}{1-\alpha\beta\Gamma^2}\right)^{1/2} \times e^{-[1/2(1-\alpha\beta\Gamma^2)][\alpha X_1^2+\beta X_2^2+2\alpha\beta\Gamma X_1 X_2]}. \qquad (19)$$

Perhaps not unexpectedly, the maximum-entropy joint distribution for more variables with Gaussian marginals will be again a Gaussian distribution with appropriate correlations. At this point it is worth mentioning that if we restrict ourselves to the class of continuous functions, then the set of joint distributions having prespecified marginals and covariance is convex, and the concavity of the entropy functional guarantees that the distribution at which it is maximized is unique. However, a more difficult problem concerns whether distributions satisfying the requirements exist at all [note, for example, that for large enough Γ in Eq. (19), the argument of the exponential changes sign, in which case $P_{(1,2)}(X_1,X_2)$ cannot be interpreted as a probability distribution]. Unfortunately, the general conditions under which the set of distributions having the prescribed marginals and covariance is not empty are not easy to establish [14]. Finally, as upon integration Gaussians beget Gaussians, for these distributions Eq. (14), as well as generalizations to more variables, will always hold.

Explicit expressions for maximum-entropy joint distributions corresponding to non-Gaussian marginal distributions appear to be very hard to obtain. However, a perturbation expansion in powers of the parameter λ can be carried out rather easily. Writing

$$Q_1(X_1) = P_1(X_1)\left[1 + f_1^{(1)}(X_1)\lambda + f_1^{(2)}(X_1)\lambda^2 + \cdots\right], \quad (20)$$

$$Q_2(X_2) = P_2(X_2)\left[1 + f_2^{(1)}(X_2)\lambda + f_2^{(2)}(X_2)\lambda^2 + \cdots\right] \quad (21)$$

in Eqs. (5) and grouping powers of λ, after some rather messy algebra, one can write that, correct to order λ^2,

$$P_{(1,2)}(X_1,X_2) = P_1(X_1)P_2(X_2)e^{-\{\lambda(X_1-\langle X_1\rangle)(X_2-\langle X_2\rangle)+(\lambda^2/2)[(\langle X_2^2\rangle-\langle X_2\rangle^2)(X_1-\langle X_1\rangle)^2+(\langle X_1^2\rangle-\langle X_1\rangle^2)(X_2-\langle X_2\rangle)^2-(\langle X_1^2\rangle-\langle X_1\rangle^2)(\langle X_2^2\rangle-\langle X_2\rangle^2)]\}},$$

where the averages $\langle X_1\rangle$, $\langle X_1^2\rangle$, etc., are taken over the marginal distributions, assuming the moments exist, and λ is calculated using Eq. (6):

$$\lambda \approx \frac{\langle X_1 X_2\rangle - \langle X_1\rangle\langle X_2\rangle}{\left(\langle X_1^2\rangle - \langle X_1\rangle^2\right)\left(\langle X_2^2\rangle - \langle X_2\rangle^2\right)} + \cdots. \qquad (22)$$

Clearly, writing the joint distribution as an exponential in Eq. (22) is not really warranted, except by the fact that it guarantees both positivity and integrability, and that it turns out to be slightly more compact than might have been expected. Further, it also highlights the fact that the approximation for $P_{(1,2)}(X_1,X_2)$ has the form required by Eq. (2), corresponding to a maximum-entropy distribution. From this expression, approximate conditional distributions can be derived immediately, as well as conditional expectations. Thus, for example, the conditional expectation of X_1 given X_2, to linear order in λ, is

$$\langle X_1|X_2\rangle \approx \langle X_1\rangle - \lambda\left(\langle X_1^2\rangle - \langle X_1\rangle^2\right)(X_2 - \langle X_2\rangle) + \cdots = \langle X_1\rangle - \frac{(\langle X_1X_2\rangle - \langle X_1\rangle\langle X_2\rangle)(X_2-\langle X_2\rangle)}{\langle X_2^2\rangle - \langle X_2\rangle^2} + \cdots. \qquad (23)$$

Also, the excess entropy over the product of marginals is found to be

$$\Delta S = \int_{I_1\times I_2} P_{(1,2)}(X_1,X_2)\ln\left(\frac{P_{(1,2)}(X_1,X_2)}{P_1(X_1)P_2(X_2)}\right)dX_1dX_2 \approx -\lambda^2\left(\langle X_1^2\rangle - \langle X_1\rangle^2\right)\left(\langle X_2^2\rangle - \langle X_2\rangle^2\right) + \cdots = -\frac{(\langle X_1X_2\rangle - \langle X_1\rangle\langle X_2\rangle)^2}{\left(\langle X_1^2\rangle - \langle X_1\rangle^2\right)\left(\langle X_2^2\rangle - \langle X_2\rangle^2\right)} + \cdots.$$

The three-point distribution can be obtained in the same way, but the result is too long and unenlightening to include here. Nevertheless, it should be mentioned that at least to second order in the perturbation parameter, Eq. (14) still holds (assuming that all the correlation coefficients can be considered to be of linear order in the perturbation parameter).

In summary, the construction of maximum-entropy joint probability distributions with the prescribed marginals and covariance has been discussed. It should be noted that while there are other convenient methods for constructing joint probability distributions with the prescribed marginals and covariance [14], only when the entropy is maximized can we be sure that no extra, uncontrolled assumptions have been introduced.

Extensions to even more variables are straightforward in principle, but the set of coupled equations that results from the maximization of entropy is larger and harder to solve, the exception being, as mentioned earlier, the case of Gaussian marginals with fixed correlations, for which the maximum-entropy distribution is the appropriate correlated Gaussian distribution. Also, the whole discussion can be extended to the case in which the inter-relation among the variables is not encoded in the linear correlation constant, but rather by other more general moments; for example, for the case of random variables X_1 and X_2 with given marginals, with the constriction $\langle f(X_1,X_2)\rangle = 0$. Another interesting extension pertains to approximation schemes for the joint distribution. For example, for the case in which the marginals do not have second moment, the perturbation expansion as presented above is not possible.

I am grateful to Ana María Contreras for her valuable comments on the manuscript, and to F. Leyvraz and S. Majumdar for useful discussions. Partial support through Grant No. DGAPA-UNAM IN109111 is gratefully acknowledged.

[1] J.-P. Bouchaud and M. Potters, *Theory of Financial Risk and Derivative Pricing From Statistical Physics to Risk Management*, 2nd ed. (Cambridge University Press, Cambridge, 2003).
[2] C. Rummel, G. Baier, and M. Mãijller, J. Neurosci. Methods **166**, 138 (2007).
[3] K. Schindler, H. Leung, C. E. Elger, and K. Lehnertz, Brain **130**, 65 (2007).
[4] G. Nicolis and M. M. Mansour, Phys. Rev. A **29**, 2845 (1984).
[5] J. R Dorfman, T. R. Kirkpatrick, and J. V. Sengers, Annu. Rev. Phys. Chem. **45**, 213 (1994).
[6] J. M. O. de Zarate and J. V. Sengers, *Hydrodynamic Fluctuations in Fluids and Fluid Mixtures* (Elsevier, Amsterdam, 2006).
[7] H. Larralde and D. P. Sanders, J. Phys. A **42**, 335002 (2009).
[8] T. Schmidt, in *Copulas—From Theory to Application in Finance*, edited by J. Rank (Risk Books, London, 2007), pp. 3–34.
[9] C. Genest and J. MacKay, Am. Stat. **40**, 280 (1986).
[10] J. Piantadosi, P. Howlett, and J. Boland, J. Indust. Manage. Optim. **3**, 305 (2007).
[11] J. Piantadosi, P. Howlett, and J. Borwein, Optim. Lett. **6**, 99 (2012).
[12] C. W. J. Granger, Econometrica **37**, 424 (1969).
[13] E. T. Jaynes, Phys. Rev. **106**, 620 (1957).
[14] Soumyadip Ghosh, Ph.D. thesis, Cornell University, 2004, http://people.orie.cornell.edu/shane/theses/GhoshThesis.pdf.