THE JOINT CUMULANTS OF TRUE VALUES AND ERRORS
OF MEASUREMENT

By Frederic M. Lord

Educational Testing Service, Princeton, New Jersey

1. Introduction. This note is concerned with the situation where U fallible measurements of some single characteristic are made on each of a large number of objects. The U measurements may represent U different methods of measuring the same characteristic, each method involving a different frequency distribution of errors of measurement.

For each object, there is an unknown "true value" of the characteristic. The difference between the observed measurement and the true value is an error of measurement. The true value and the errors of measurement will be termed *latent variables*.

The results derived are currently being applied in psychometric work, but they should be applicable in almost any field where unbiased fallible measurements are made. For example, the true amount (ξ) of some chemical constituent of the blood may have been fallibly but independently measured by U different methods (or by U different laboratory technicians) for each of a large number of hospital patients. The results given here will permit the consistent estimation of the first U cumulants of ξ, the first U cumulants of the error of measurement in each of the U methods, and, further, all the multivariate cumulants of the latent variables up through order U.

In psychometric work, U strictly parallel forms of a mental test may be prepared by matching the questions assigned to the different forms on their statistical characteristics (determined by pretesting). These test forms may, in effect, all be administered "simultaneously" by the device of interspersing the questions from all forms and then scoring the questions of each form separately, counting the number answered correctly. The moments of the frequency distribution of the "true scores" (ξ) of the examinees tested and also the distribution of the errors of measurement may now be estimated by the method to be described. (In this case, the shape of the distribution of errors of measurement must be dependent on the value of ξ. This is apparent, for example, from the fact that the observed test scores cannot be negative; hence whenever ξ is near zero large negative errors of measurement cannot occur.)

Formulas illustrating the final results obtained are given in Section 2. The derivations are given in the remaining sections.

In Section 3, any multivariate cumulant of the observed measurements is expressed as a linear function of the cumulants of the joint distribution of the latent variables, no assumption being made other than the existence of the cumulants in question (the results of this section are not new; they could be di-

rectly obtained by specializing a formula given by James ([1], eq. 13), for example). In Section 4 it is assumed that each error of measurement has a mean of zero and is uncorrelated with other appropriate chance variables; it is shown that each multivariate (and univariate) cumulant of the latent variables can be expressed in closed form as a simple linear function of the cumulants of the observed measurements. Section 5 details the restrictions imposed on the cumulants of the observed measurements by the assumption about uncorrelated errors.

2. Specific formulas. There are U *observed measurements* on each object, denoted by $x_1, \cdots, x_U$. Both these and the *true value*, ξ, are random variables. The *errors of measurement* are, by definition,

$$(1) \qquad\qquad e_u = x_u - \xi, \qquad\qquad (u = 1, \cdots, U).$$

Any U-variate cumulant of the observed measurements is denoted by $\kappa_{C_1 \cdots C_U}$, where $C_1, \cdots, C_U$ are nonnegative integers referring to the variables $x_1, \cdots, x_U$, respectively. Any cumulant of the latent variables is similarly denoted by $K_{B_0, B_1 B_2 \cdots B_U}$, where the first subscript refers to variable ξ and the other U subscripts refer to the U errors of measurement. It will be notationally convenient to use a zero-order cumulant, having all zero subscripts, that is by definition equal to zero.

Explicit formulas are given below expressing all latent-variable cumulants up through the fourth order in terms of the observed-variable cumulants for the case where $U = 4$. All necessary formulas are either given or may be obtained by permutation of subscripts. The first subscript on each K, representing the true value, is not subject to permutation, but the U other subscripts on each K or κ may be permuted providing the same permutation is made on each K and κ throughout the entire formula.

$$K_{0,1000} = K_{0,0100} = K_{0,0010} = K_{0,0001} = 0 \text{ by assumption}$$

$$K_{1,0000} = \kappa_{1000} = \kappa_{0100} = \kappa_{0010} = \kappa_{0001}$$

$$K_{2,0000} = \kappa_{1100} = \kappa_{1010} = \cdots = \kappa_{0011}$$

$$K_{0,2000} = \kappa_{2000} - \kappa_{1100} = \cdots = \kappa_{2000} - \kappa_{0011}$$

$$K_{0,0200} = \kappa_{0200} - \kappa_{1100} = \cdots = \kappa_{0200} - \kappa_{0011}, \text{ etc.}$$

$$K_{3,0000} = \kappa_{1110} = \cdots = \kappa_{0111}$$

$$K_{2,1000} = \cdots = K_{2,0001} = 0, \text{ by assumption}$$

$$K_{1,1100} = K_{0,2100} = K_{0,1110} = 0, \text{ etc., by assumption}$$

$$K_{1,2000} = \kappa_{2100} - \kappa_{1110} = \cdots = \kappa_{2001} - \kappa_{0111}, \text{ etc.}$$

$$K_{0,3000} = \kappa_{3000} - 3\kappa_{2100} + 2\kappa_{1110}, \text{ etc.}$$

$$K_{4,0000} = \kappa_{1111}$$

$$K_{3,1000} = K_{2,1100} = K_{1,2100} = \cdots = K_{0,1111} = 0 \text{ by assumption}$$

$$K_{2,2000} = \kappa_{2110} - \kappa_{1111}, \text{ etc.}$$

$$K_{1,3000} = \kappa_{3100} - 3\kappa_{2110} + 2\kappa_{1111}, \text{ etc.}$$

$$K_{0,2200} = \kappa_{2200} - \kappa_{2110} - \kappa_{1210} + \kappa_{1111}, \text{ etc.}$$

$$K_{0,4000} = \kappa_{4000} - 4\kappa_{3100} + 6\kappa_{2110} - 3\kappa_{1111}, \text{ etc.}$$

3. General relations among cumulants. The characteristic function of the latent variables may be written

$$(2) \qquad F(T_0, T_1, \cdots, T_U) = E \exp i \sum_{u=0}^{U} T_u e_u,$$

where E is the expectation symbol and $e_0 = \xi$. That of the observed measurements is, by (1),

$$(3) \qquad f(t_1, \cdots, t_U) = E \exp i \sum_{u=}^{U} t_u x_u = E \exp i \sum_{u=0}^{U} t_u e_u,$$

where $t_0 = \sum_{u=1}^{U} t_u$.

It is seen that the first characteristic function may be changed to the second simply by replacing T by t. If the necessary cumulants exist, the cumulant-generating function of the latent variables is

$$\log F(T_0, T_1, \cdots, T_U)$$

$$(4) \qquad = \sum_{B_0=0}^{\infty} \sum_{B_1=0}^{\infty} \cdots \sum_{B_U=0}^{\infty} \frac{i^P T_0^{B_0} T_1^{B_1} \cdots T_U^{B_U}}{B_0! B_1! \cdots B_U!} K_{B_0,B_1\cdots B_U},$$

where $P = \sum_{u=0}^{U} B_u$. Take the right side of (4) and replace $T_1, \cdots, T_U$ by $t_1, \cdots, t_U$ and $T_0^{B_0}$ by the multinomial expansion

$$(5) \qquad t_0^{B_0} = \left(\sum_{u=1}^{U} t_u \right)^{B_0} \equiv \sum_a \frac{B_0!}{a_1! \cdots a_U!} t_1^{a_1} \cdots t_U^{a_U},$$

where $a_1, \cdots, a_U$ are nonnegative integers and Σ_a is over all sets of a such that $\sum_{u=1}^{U} a_u = B_0$. This converts the right side of (4) to the cumulant-generating function of the observed measurements:

$$\log f(t_1, \cdots, t_U)$$

$$(6) \qquad = \sum_{B_0=0}^{\infty} \sum_{B_1=0}^{\infty} \cdots \sum_{B_U=0}^{\infty} \sum_a \frac{i^P t_1^{B_1+a_1} \cdots t_U^{B_U+a_U}}{B_1! \cdots B_U . a_1! \cdots a_U!} K_{B_0,B_1\cdots B_U}.$$

The cumulant $\kappa_{C_1\cdots C_U}$ of the observed measurements is the coefficient of the term $i^{P'} t_1^{C_1} \cdots t_U^{C_U}/C_1! \cdots C_U!$ in the series at the right, where $P' = \sum_{u=1}^{U} C_u$. If $B_u + a_u$ is replaced by C_u in (6) and the terms rearranged, these cumulants are found to be

$$(7) \qquad \kappa_{C_1\cdots C_U} = \sum_B K_{B_0,B_1\cdots B_U} \prod_{u=1}^{U} \binom{C_u}{B_u},$$

where $\binom{C_u}{B_u} = C_u!/B_u!(C_u - B_u)!$, and where Σ_B is taken over all sets of non-negative integral values of the B_u subject to the restrictions that $B_u \leqq C_u$ for $u = 1, \cdots, U$ and $P' = P$.

The result given in equation (7) may also be expressed in terms of symbolic multiplication:

$$(8) \qquad \kappa_{c_1 \cdots c_U} \sim (\xi + e_1)^{c_1} (\xi + e_2)^{c_2} \cdots (\xi + e_U)^{c_U}.$$

The $\sim$ symbol may be replaced by an equals sign when each term $\xi^{B_0} e_1^{B_1} e_2^{B_2} \cdots e_U^{B_U}$ on the right has been replaced by $K_{B_0, B_1 \cdots B_U}$.[1]

Formulas (7) or (8) express any Pth-order cumulant of the observed measurements as a linear function of the Pth-order cumulants of the latent variables. Without further assumptions, it is not possible to solve any set of these equations for the unknown cumulants of the latent variables, for the reason that there are always fewer equations than unknowns.

4. Determining the cumulants of the latent variables. The result in (7) and (8) was obtained without any assumption about the distribution of the latent variables other than the existence of the cumulants. It will now be assumed that each error of measurement has a mean value of 0 and is uncorrelated with every product of the remaining latent variables. Thus $K_{B_0, B_1 \cdots B_U} = 0$ whenever any B_u ($u > 0$) is equal to 1. This is much less restrictive than the usual assumption that the errors of measurement are distributed independently of the true value and of each other. *The present assumption, for example, permits the variance of the errors of measurement and all higher moments to be dependent on ξ—it is only the mean error of measurement that is independent of ξ.*

With this assumption, we may proceed to prove

THEOREM 1. *Given that $K_{B_0, B_1 \cdots B_U} = 0$ whenever any $B_u = 1$ ($u > 0$), all equations (7) for which $\sum_{u=1}^{U} C_u = P$ is constant can be ranked so that the right side of each contains at most one nonzero K appearing in no preceding equation; thus, given that the equations are consistent, they may be solved so as to express any K of order $\leqq U$ as a linear function of κ's.*

Let $U - T$ be the number of subscripts on the left side of (7) that are equal to 1, so that the observed-variable cumulant may be written $\kappa_{c_1 c_2 \cdots c_T 11 \cdots 1}$. For $B_0 < U - T$, there must be at least one value of $u > 0$ for which $B_u = 1$ on the right side of (7), so every $K_{B_0, B_1 \cdots B_U}$ will vanish whenever $B_0 < U - T$. For $B_0 = U - T$, there is on the right side of (7) one and only one cumulant, $K_{B_0, B_1 \cdots B_U}$, without unit subscripts; this unique nonvanishing cumulant has T subscripts that are the same as those of the observed-variable cumulant and (at least) $U - T$ zero subscripts, so it may be written $K_{(U-T), c_1 c_2 \cdots c_T 00 \cdots 0}$; it will be spoken of as the latent-variable cumulant to which $\kappa_{c_1 c_2 \cdots c_T 11 \cdots 1}$ (with

---

[1] As pointed out by a referee, formula (8) shows that the relation between observed-variable and latent-variable cumulants is exactly the same as the relation between observed-variable and latent-variable moments about an arbitrary origin.

$U - T$ unit subscripts) *corresponds*. For $B_0 > U - T$, there may be on the right side of (7) a number of cumulants $K_{B_0, B_1 \cdots B_U}$ without unit subscripts; each of these, however, is a latent-variable cumulant to which some other observed-variable cumulant $\kappa_{B_1 B_2 \cdots B_U 11 \cdots 1}$ having B_0 unit subscripts corresponds.

Consider all the observed-variable cumulants of a given order (P) and suppose them to be grouped according to $U - T$, the number of unit subscripts, and the groups ranked on $U - T$ starting with the cumulant for which $T = 0$. The results of the preceding paragraph show that equation (7) expresses any κ as a linear function of K's, one of these being the K to which the given κ corresponds, all the others being K's to which correspond some other κ's of lower rank. This completes the proof of Theorem 1.

5. Restrictions on the observed cumulants. If none of the C's are zero in $\kappa_{c_1 c_2 \cdots c_T 11 \cdots 1}$, then this is the only κ that corresponds to $K_{(U-T), c_1 c_2 \cdots c_T 00 \cdots 0}$. If V of the C's in $\kappa_{c_1 c_2 \cdots c_T 11 \cdots 1}$ are zero, then every κ obtained by permuting the zero and unit subscripts on κ also corresponds to the same K.

It follows from Theorem 1 that any K may be expressed as equal to any one of the $\begin{pmatrix} V + U - T \\ U - T \end{pmatrix}$ corresponding κ's plus other κ's of lower rank (unless the first κ already is of lowest rank). For the κ's of lowest rank, $T = V$ and there are $\begin{pmatrix} U \\ U - V \end{pmatrix}$ equally good equations, such as $K_{(U-T), 00 \cdots 0} = \kappa_{11 \cdots 100 \cdots 0}$, there being U zero subscripts on the left side of the equation, T zero subscripts and $U - T$ unit subscripts on the right. These $\begin{pmatrix} U \\ U - T \end{pmatrix}$ different κ's must thus be equal. Proceeding to the case of next higher rank, another set of κ's are found that must be equal to each other. Mathematical induction now shows that all κ's corresponding to a given K must be equal. Thus,

THEOREM 2. *Given that* $K_{B_0, B_1 \cdots B_U} = 0$ *for any* $B_u = 1$ $(u > 0)$, *any two* κ's *will be equal if their subscripts are the same except for a permutation that involves zero and unit subscripts only.*

Theorem 2 states a restriction on the observed-variable κ's that is implicit in the assumption made in Section 4 about uncorrelated errors. Since the matrix of the large-sample sampling variances and covariances of the κ's could be computed if desired, the assumption made in Section 4 can be submitted to statistical test, at least in large samples, to determine whether or not any given set of observed data is compatible with it.

REFERENCE

[1] G. S. JAMES, "Cumulants of a transformed variate," *Biometrika*, Vol. 42 (1955), pp. 529-531.