Z. Wahrscheinlichkeitstheorie verw. Geb. 10, 203—211 (1968)

On Products of Experiments

RICHARD SACKSTEDER *

Received February 19, 1967

Summary. An experiment in the sense of BLACKWELL is a finite parameterized set of probability distributions on a sample space. The product of a parameterized set of experiments is the experiment which describes the process in which independent elements from all of the sample spaces in the set are observed. In the paper, methods are developed for computing the properties of a product of experiments from the properties of the components. In particular, asymptotic properties (as $N \to \infty$) of the experiment describing N independent observations from a single sample space are investigated.

1. Introduction

BLACKWELL [1], [2] has defined an experiment to be a finite parameterized set of probability measures $\{m_1, \ldots, m_k\}$ on a measurable space (X, Ω). By the *product* of such an experiment with another, say $\{n_1, \ldots, n_k\}$ on (Y, Λ), is meant the experiment $\{m_1 \times n_1, \ldots, m_k \times n_k\}$ on $(X \times Y, \Omega \times \Lambda)$. Clearly, the product experiment corresponds to taking independent observations from X and Y.

In sections 2 and 3 some machinery for systematically studying experiments and their products is developed. It turns out (Theorem 3.1) that it is possible to associate to every isomorphism class (cf. [5, 9, 10, 11]) of homogeneous experiments parameterized by $1, \ldots, k$ a unique "associated measure" on Euclidean k-space R^k having certain properties. Thus the study of an experiment is reduced to the study of a single measure on R^k. The correspondence is such that products of experiments correspond to convolutions of their associated measures. Moreover, the properties of the associated measures are such that classical limit theorems can be applied to them, regardless of how irregular the measures in the original experiment may be.

The results of sections 2 and 3 are applied in later sections to a particular problem. Let $\{m_1, \ldots, m_k\}$ be as above and let a^i denote the a priori probability that the distribution m_i is the true one. Then let $B_N(a^1, \ldots, a^k)$ denote the probability of making the correct determination of the true distribution after observing N independently chosen samples from X if the optimal procedure is used. Most of the properties of B_N are classical and are well known, cf. [9, [10]. Here the properties of B_N for large N will be investigated, and it will be shown that this asymptotic behavior can be described simply in terms of a few numbers which can be computed from the measures $m_1, \ldots, m_k$.

The properties of B_N are not only of interest because of their significance from the standpoint of the Bayesian problem described above, but also because the experiment $\{m_1^N, \ldots, m_k^N\}$ is determined completely by B_N up to isomorphism in the sense of [9], where m_i^N denotes N-fold product of m^i. Thus the rate at which

* Research partially supported by National Science Foundation Grant (GP-5227).

B_N increases (and approaches 1 except in degenerate cases) in a sense measures the information obtained from observing many independently chosen elements of X.

It seems fair to say the method illustrated here is different from that usually employed in large sample statistics in the following sense: The usual method is to apply the limit theorem of probability to the individual measures m_i, if possible, and then to combine the measures m_i^N for large N into an experiment. We proceed in the opposite way, first combining measures into an experiment and then applying limit theorems to the measure associated to it. This reversal of procedure seems to have some advantages which should make it applicable to problems beyond the one illustrated here.

2. Normalization of an Experiment

Let $\{m_1, \ldots, m_k\}$ $(k \geqq 2)$ be an experiment on a measurable space (X, Ω). For any measure m on (X, Ω), $N(m)$ will denote the σ-ideal of null sets of m. The experiment will be called *homogeneous* if $N(m_i) = N(m_j)$ for all i, j. For the sake of simplicity here we shall be concerned with only homogeneous experiments, but the methods can be extended to cover the general case with a little more trouble.

Lemma 2.1. *Let* $\{m_1, \ldots, m_k\}$ *be a homogeneous experiment on* (X, Ω). *Then there is unique measure* m *on* (X, Ω) *satisfying:* (i) $N(m) = N(m_i)$, (ii) $m(X) = 1$, (iii) *there are Radon-Nikodym derivatives* $d_i = dm_i/dm$ *satisfying* $\prod_{i=1}^{k} d_i(x) \equiv A \leqq 1$ *where* A *is a positive constant. If* n *is any measure on* (X, Ω) *satisfying* $N(n) = N(m_i)$, A *is defined by*

$$A = \left[\int_X \left\{ \prod_{i=1}^{k} \varrho_i(x) \right\}^{1/k} n(dx) \right]^k, \tag{2.1}$$

where $\varrho_i = dm_i/dn$ *is a Radon-Nikodym derivative. (*A *is independent of the choices of* n *and* ϱ_i *and will be called the associated constant of the experiment* $\{m_1, \ldots, m_k\}$.*)*

Proof. Define m by

$$m(E) = c \int_E \left(\prod_{i=1}^{k} \varrho_i(x) \right)^{1/k} n(dx),$$

where c is chosen so that (ii) holds. Then $N(m) = N(n) = N(m_i)$ and

$$d_i \equiv dm_i/dm = \varrho_i c^{-1} \left(\prod_{i=1}^{k} \varrho_i(x) \right)^{-1/k}$$

so (i) and (iii) hold where $A = c^{-k}$. The inequality $A \leqq 1$ follows by applying the inequality between the arithmetic and geometric means to the integrand in (2.1).

To prove uniqueness, suppose that m' satisfies (i), (ii) and (iii) with A replaced by A'. It can be supposed that $A \geqq A'$. Now if $m \neq m'$, there is a set E such that $\varrho = dm/dm'$ satisfies $\varrho \geqq b > 1$ on E and $m(E) > 0$. Then there must be an i such that $d_i \geqq d_i' = dm_i/dm'$ on $F \subset E$, where $m(F) \neq 0$ hence $m_i(F) \neq 0$.

Then

$$m_i(F) = \int_F d_i(x)\,\varrho(x)\,m'(dx) \geqq b \int_F d_i'(x)\,m'(dx) = b\,m_i(F),$$

but this contradicts $b > 1$ and $m_i(F) \neq 0$. This proves Lemma 2.1.

Now with the assumptions and notation of Lemma 2.1, define the map $T_1\colon X \to R^k$ by letting the i-th component of $T_1(x)$ be $\log(d_i(x))$. Let μ and μ_i be the measures on R^k induced by T_1 from m and m_i, that is $\mu(E) = m(T_1^{-1}(E))$ and $\mu_i(E) = m_i(T_1^{-1}(E))$.

Lemma 2.2. *Under the assumptions of Lemma* 2.1, *the map* T_1 *is a sufficient statistic so that the experiment* $\{\mu_1, \ldots, \mu_k\}$ *is isomorphic to* $\{m_1, \ldots, m_k\}$. *Moreover, the following conditions are satisfied:*

(i) $N(\mu) = N(\mu_i)$,

(ii) $\mu(R^k) = \mu_i(R^k) = 1$,

(iii) $$S \equiv \text{support}\,\mu \subset \Pi_B \equiv \{y \in R^k : \sum_{i=1}^k y^i = B = \log A\},$$

(iv) *the Radon-Nikodym derivative* $(d\mu_i/d\mu)(y) = \exp(y^i)$,

(v) *the measure* μ *has finite absolute moments of all orders.*

Proof. That T is a sufficient statistic follows immediately from [8, Corollary 2] or [11, Theorem 2.1]. The conditions (i), (ii), (iii) follow immediately from the corresponding conditions in Lemma 2.1, where $B = \log A \leqq 0$. The definition of T_1 implies (iv). To prove (v), first note that for any integer $p \geqq 0$ and $y = (y^1, \ldots, y^k)$ in Π_B, (iii) implies that if $Z^j = y^j - B/k$, $|Z^j| \leqq k \max\{Z^i : Z^i \geqq 0\}$, and $|Z^j|^p \leqq k^p \max\{(Z^i)^p : Z^i \geqq 0\}$. Therefore (ii), (iii), and (iv) give

$$\int_{R^k} |Z^j|^p \mu(dy) \leqq k^p p! \sum_{i=1}^k \int_{R^k} \exp(y^i - B/k)\,\mu(dy) \leqq k^{p+1} p!\, A^{-1/k}.$$

Now

$$|y^j|^p \leqq \sum_{q=0}^p |Z^j|^q (-B/k)^{p-q} C_q^p \quad \text{implies that} \quad \int_{R^k} |y^j|^p \mu(dy) \leqq C(p) < \infty.$$

If $(p_1, \ldots, p_k)$ is given with $p_i \geqq 0$,

$$\int_{R^k} \prod_{i=1}^k |y^i|^{p_i} \mu(dy) \leqq k^{-1} \int_{R^k} \sum_{i=1}^h |y^i|^{kp_i} \mu(dy) \leqq k^{-1} \sum_{i=1}^h C(kp_i) < \infty,$$

where the first inequality comes from the inequality between the arithmetic and geometric means. This proves (v) and Lemma 2.2.

Lemma 2.3. *Let* $\{m_1, \ldots, m_k\}$ *and* $\{n_1, \ldots, n_k\}$ *be homogeneous experiments on* (X, Ω) *and* (Y, Λ) *respectively. Let* μ *and* $\{\mu_1, \ldots, \mu_k\}$ *correspond to the first experiment as in Lemma* 2.2 *and* ν *and* $\{\nu_1, \ldots, \nu_k\}$ *correspond to the second. Then the experiments are isomorphic if and only if* $\mu = \nu$.

Proof. The first part of Lemma 2.2 shows that the experiments are isomorphic if and only if $\{\mu_1, \ldots, \mu_k\}$ and $\{\nu_1, \ldots, \nu_k\}$ are isomorphic. But part (iv) of Lemma 2.2 shows that $\{\mu_1, \ldots, \mu_k\}$ is determined by μ alone and $\{\nu_1, \ldots, \nu_k\}$ by ν alone, and the „if" part of conclusion follows.

Now suppose that $\{\mu_1, \dots, \mu_k\}$ and $\{\nu_1, \dots, \nu_k\}$ are isomorphic, say $\nu_i = T_* \mu_i$ where T_* is an isomorphism induced from a statistical operation T between two copies of R^k. By Lemma 2.2,

$$d\mu^i/d\mu^j = \exp(y^i - y^j) \quad \text{at} \quad (y^1, \dots, y^k) \in S \subset \Pi_B .$$

Then if

$$\alpha_j(y) = \prod_{i=1}^{k} (d\mu^i/d\mu^j)(y) = \exp(B - k y^j),\ y^j = k^{-1}(B - \log \alpha_j(y)),$$

and therefore, the coordinates of a point y in S can be computed from likelihood ratios $d\mu^i/d\mu^j$ at the point. Similarly, if the support of ν is $S' \subset \Pi_C$, and $z = (z^1, \dots, z^k)$ is in S', $z^j = k^{-1}(C - \log \beta_j(z))$, where $\beta_j(z) = \prod_{i=1}^{k} (d\nu^i/d\nu^j)(z)$. Theorem 2.1 of [11] implies that T preserves likelihood ratios, which now shows that T can be taken as a map from Π_B to Π_C such that y in Π_B corresponds to z in Π_C if and only if $\alpha_j(y) = \beta_j(z)$ or $B - k y^j = C - k z^j$ for $j = 1, \dots, k$. Summing over j gives $(B - C)(k - 1) = 0$. Since $k \geqq 2$, this means that $B = C$ and T is the identity map on $S = S'$ But then $(d\mu/d\mu^j)(y) = \exp(-y^j) = (d\nu/d\nu^j)(y)$ on $S = S'$, hence $\mu = \nu$. This completes the proof.

The Lemmas 2.2 and 2.3 justify the following terminology: μ will be called the *associated measure* of $\{m_1, \dots, m_k\}$, or more precisely of the isomorphism class of $\{m_1, \dots, m_k\}$.

3. Products of Experiments

The theorem below is our main tool. Except for the last statement, it is just a summary of the results of the preceding section.

Theorem 3.1. *Let $\{m_1, \dots, m_k\}$ be a homogeneous experiment on (X, Ω). Then there is associated to the isomorphism class of $\{m_1, \dots, m_k\}$ a measure μ on R^k with its support in $\Pi_B = \{y \in R^k: \sum_{i=1}^{h} y^i = B = \log A\}$, where A is the associated constant defined by* (2.1). *If $\mu_i(E) = \int_E \exp(y^i)\, d\mu$, μ_i is a probability measure and the experiment $\{\mu_1, \dots, \mu_k\}$ is isomorphic to $\{m_1, \dots, m_k\}$. The measure μ has moments of every order. If $\{n_1, \dots, n_k\}$ is a second homogeneous experiment on (Y, Λ) with associated measure ν, then the associated measure of the product experiment $\{m_1 \times n_1, \dots, m_k \times n_k\}$ on $(X \times Y, \Omega \times \Lambda)$ is the convolution $\mu * \nu$.*

Proof. Let $d_1, \dots, d_k$, m be as in Lemma 2.1 and $e_1, \dots, e_k$, n, the corresponding objects for the other experiment $\{n_1, \dots, n_k\}$. The objects in the lemma corresponding to the product experiment will just be $d_1 \times e_1, \dots, d_k \times e_k$, $m \times n$. The map $T_2: X \times Y \to R^k$ analogous to T_1 as defined before Lemma 2.2 will be just the map defined by setting the i-th component of $T_2(x)$ equal to $\log(d_i(x)) + \log(e_i(x))$. Then it follows from the definition of T_2 and of the convolution $\mu * \nu$ that

$$m \times n(T_2^{-1}(E)) = \mu \times \nu((\xi, \eta) \in R^k \times R^k: \xi + \eta \in E) = \mu * \nu(E).$$

This proves the last assertion and completes the proof of the theorem.

The theorem shows that one can apply the classical limit theorems of probability to μ to obtain the asymptotic properties of the N-fold product experiment

$\{m_1^N, \ldots, m_k^N\}$ on (X^N, Ω^N). Since the moments of μ are all finite the classical limit theorems apply without special assumptions on the m_i's.

4. The Function B_N

Let $a = (a^1, \ldots, a^k)$ be a k-tuple of non-negative numbers with $\sum_{i=1}^{k} a^i = 1$ and let $B_N(a)$ be as in the first section. Then it is known that $B_N(a)$ depends only on the isomorphism class of $\{m_1^N, \ldots, m_k^N\}$, cf. [9, Corollary 1]. Therefore Theorem 3.1 above and the form of the Neyman-Pearson Lemma given in [9, Section 6] show that

$$B_N(a^1, \ldots, a^k) = \int_{R^k} \max\{a^i \exp(y^i)\}\, \mu_*^N(dy), \tag{4.1}$$

where μ_*^N denotes the N-fold convolution of μ with itself. This formula is not convenient for studying the asymptotic behavior of B_N because the integrand is large for large $|y|$. This difficulty will be avoided means of the following lemma.

Lemma 4.1. *Let $\{u_1, \ldots, u_k\}$ be a set of numbers. If H is a subset of $\{1, \ldots, k\}$ let $M(H) = \min\{u_i\colon i \in H\}$ and let M_j be the sum of $M(H)$ over all subsets H with exactly j elements. Then*

$$\operatorname{Max}\{u_i\colon i = 1, \ldots, k\} = \sum_{j=1}^{k} (-1)^{j-1} M_j. \tag{4.2}$$

Proof. Suppose for the moment that $u_1 \geqq u_2 \geqq \cdots \geqq u_k$ and let i and j satisfy $1 \leqq j$, $i \leqq k$. Then if $j \leqq i$ there are exactly C_{j-1}^{i-1} subsets containing exactly j elements such that u_i is the smallest element and if $j > i$ u_i is not the (unique) smallest element of any subset containing j elements. Therefore

$$\sum_{j=1}^{k} (-1)^{j-1} M_j = \sum_{j=1}^{k} \sum_{i=j}^{k} (-1)^{j-1} C_{j-1}^{i-1} u_i = u_1.$$

This proves (4.2) under the assumed ordering of the u_i's. However, neither side of (4.2) depends on the ordering, so (4.2) holds in general.

Let $D_N^j (1 \leqq j \leqq k)$ be defined by

$$D_N^j(a^1, \ldots, a^k) = \sum_H \int_{R^k} \min\{a^i \exp(y^i)\colon\ i \in H\}\, \mu_*^N(dy) \tag{4.3}$$

where the summation is over all subsets H of $\{1, \ldots, k\}$ containing exactly j elements. Each term in the summation on the right side of (4.3) can be written in the form

$$\int_{R^k} \operatorname{Min}\{a^i \exp(y^i)\colon i \in H\}\, \mu_*^N(dy) = \int_{R^j} \operatorname{Min}\{a^i \exp(z^i)\colon i \in H\}\, \mu_*^N(dz, H) \tag{4.4}$$

where $\mu_*^N(\ , H)$ denotes the measure formed in a way analogous to μ_*^N but with respect to the subexperiment $\{\mu^i\colon i \in H\}$. The support of this measure lies in some hyperplane $\sum_{i \in H} z^i = \text{const.}$, hence the integrand on the right of (4.4) is bounded there. Now (4.1) can be written

$$B_N(a^1, \ldots, a^k) = \sum_{j=1}^{k} (-1)^{j-1} D_N^j(a^1, \ldots, a^k), \tag{4.5}$$

which in view of (4.3) and (4.4) is a finite sum of integrals with bounded integrands. Since obviously $D^1_N(a^1, \ldots, a^k) = \sum_{i=1}^{k} a^i = 1$, (4.5) can also be written as

$$(4.6) \qquad 1 - B_N(a^1, \ldots, a^k) = \sum_{j=2}^{k} (-1)^j D^j_N(a^1, \ldots, a^k).$$

5. Bounds for B_N

The following lemma will make it possible to give bounds for B_N.

Lemma 5.1. *Let $\{m_1, \ldots, m_k\}$ be a homogeneous experiment on (X, Ω). Let A be the associated constant and let μ be the associated measure. Then if $a^i \geqq 0$, $\sum a^i \leqq 1$*

$$(5.1) \qquad \int_{R^k} \mathrm{Min}\,(a^i \exp(y^i))\,\mu(dy) \leqq A^{1/k}\, G(a^1, \ldots, a^k),$$

where $G(a^1, \ldots, a^k)$ denotes the geometric mean of $a^1, \ldots, a^k$.

Proof. The left side of (5.1) can be written as

$$\sum_{i=1}^{k} J_i, \quad \text{where} \quad J_i = a^i \int_{S_i} \exp(y^i)\,\mu(dy).$$

$S_i = \{y \in \Pi_B : a^j \exp(y^j) > a^i \exp(y^i) \leqq a^h \exp(y^h),\ j < i \leqq h\}$, and

$$\Pi_B = \{y \in R^k : \sum_{i=1}^{k} y^i = B = \log A\}.$$

Now if y is in S_i, $y^i \leqq y^j + \log(a^j/a^i)$ for $1 \leqq j \leqq k$. Summing over j gives

$$z_i = \mathrm{Sup}(\{y^i : y \in S_i\}) \leqq B/k + \log\{G(a^1, \ldots, a^n)/a^i\}.$$

Therefore $J_i \leqq a^i \mu(S_i) \exp(z_i) = \mu(S_i) A^{1/k} G(a^1, \ldots, a^n)$. Since the support of μ is in Π_B, summing over i gives (5.1).

Theorem 5.1. *Let $\{m_1, m_2\}$ be a homogeneous experiment on (X, Ω). Then for every N, and $a^i \geqq 0$, $\sum_{i=1}^{2} a^i = 1$*

$$(5.3) \qquad B_N(a^1, a^2) \geqq 1 - A^{N/2}\, G(a^1, a^2)$$

where A and G are as in Lemma 5.1.

Proof. The case $N = 1$ follows immediately from Lemma 5.1, (4.3) and (4.6). The general case follows from the case $N = 1$ on observing that the associated constant of the experiment $\{m_1^N, \ldots, m_k^N\}$ is A^N. The next theorem contains Theorem 5.1.

Theorem 5.2. *Let $\{m_1, \ldots, m_k\}$ be a homogeneous experiment on (X, Ω). For any subset $H \subset \{1, \ldots, k\}$ let A_H denote the associated constant of the subexperiment $\{m_i : i \in H\}$, let $G_H(a^1, \ldots, a^k)$ be the geometric mean of the numbers $\{a^i : i \in H\}$,* and *$k(H)$ the number of elements of H. Then if $a^i \geqq 0$, $\sum_{i=1}^{k} a^i = 1$,*

$$B_N(a^1, \ldots, a^k) \geqq 1 - \sum_H (A_H)^{N/k(H)}\, G_H(a^1, \ldots, a^k),$$

where the summation is over all subsets H such that $k(H)$ is even.

Theorem 5.2 also follows immediately from (4.3), (4.5), (4.6) and Lemma 5.1. A still more general theorem is:

Theorem 5.3. *Let* $\{m_{11}, \ldots, m_{1k}\}$, $\{m_{21}, \ldots, m_{2k}\}, \ldots$ *be a sequence of experiments on* (X_1, Ω_1), $(X_2, \Omega_2), \ldots$ *respectively. Let the number* A_H *as defined in Theorem* 5.2 *associated to the j-th experiment be denoted by* A_H^j *and let* $B^N(a^1, \ldots, a^n)$ *be the function corresponding to the experiment*

$$\left\{\prod_{j=1}^{N} m_{j1}, \ldots, \prod_{j=1}^{N} m_{jk}\right\} \quad on \quad \left(\prod_{j=1}^{N} X_j, \prod_{j=1}^{N} \Omega_j\right).$$

Then

$$B^N(a^1, \ldots, a^k) \geqq 1 - \sum_H \left(\prod_{j=1}^{N} A_H^j\right)^{1/k(H)} G_H(a^1, \ldots, a^k),$$

where a^i, G_H *and the summation have the same meaning as in Theorem* 5.2.

The proof of this theorem is the same as the proof of Theorem 5.2.

An application of the case $k = 2$ of Theorem 5.3 shows that if $\prod_{j=1}^{\infty} A^j$ diverges to zero, $B^N(a^1, a^2) \to 1$, where A^j denotes the associated constant of the experiment $\{m_{j1}, m_{j2}\}$. The convergence of $B^N(a^1, a^2)$ to one means, of course, that one can make estimates of the „true" parameter with arbitrarily low probability of error by observing from enough of the X_i's.

6. Use of Limit Theorems

The theorems of the last section did not exploit the possibility of using central limit theorems to estimate the asymptotic properties of B_N. This possibility will now be illustrated in the case $k = 2$ by employing the Berry-Essen Theorem [4], [6], [7, Chapter 8]. For $k > 2$ similar results can be obtained from the multivariate form of this theorem, which is due to Bergström [3]. In this case the explicit formulae become much more complicated.

Now let $\{m_1, m_2\}$ be an experiment with associated measure μ on R^2. Define

$$Y^i = \int_{R^2} y^i \mu(dy), \qquad \alpha_{ij} = \int_{R^2} (y^i - Y^i)(y^j - Y^j)\mu(dy).$$

Since the support of μ is in $\Pi_B = \{y \in R^2 \colon y^1 + y^2 = B \leqq 0\}$, $Y^1 + Y^2 = B$ and $\alpha_{11} + 2\alpha_{12} + \alpha_{22} = 0$. Also

$$\alpha_{11} - 2\alpha_{12} + \alpha_{22} = \int_{R^2} (y^1 - y^2 - Y^1 + Y^2)^2 \mu(dy) \geqq 0$$

where equality holds only in the trivial case where $m_1 = m_2$, which we shall exclude here. Let σ be the positive square root of $\frac{1}{4}(\alpha_{11} - 2\alpha_{12} + \alpha_{22})$ so that $\sigma^2 = \frac{1}{4}(\alpha_{11} - 2\alpha_{12} + \alpha_{22}) = \frac{1}{2}(\alpha_{11} + \alpha_{22}) = -\alpha_{12}$.

In the theorem below it is of interest note that Y^1 and Y^2 are both negative. This follows by applying Jensen's inequality to $\int_{R^2} \exp(y^i)\mu(dy) = 1$. The possibility that $Y^i = 0$ is excluded by the assumption that $m_1 \neq m_2$.

Theorem 6.1. *Let* $\{m_1, m_2\}$ *be a homogeneous experiment with associated constant* A *and* $m_1 \neq m_2$. *Let* $\Phi(t) = (2\pi)^{-1/2} \int_{-\infty}^{t} \exp(-s^2/2)ds$, $W_1 = \frac{1}{2}N^{-1/2}\sigma^{-1}\{\log(a^2/a^1)$

$+N(Y^1-Y^2)\}-N^{1/2}\sigma$, $W_2=W_1+2N^{1/2}\sigma$. *Let* $a^i>0$, $a^1+a^2=1$, $G(a^1, a^2) = (a^1a^2)^{1/2}$. *Then*

$$\begin{aligned}(6.1)\quad &|1 - B_N(a^1, a^2) - a^1 \exp(N(Y^1+\tfrac{1}{2}\sigma^2))\Phi(W_1)\\ &\quad - a^2\exp(N(Y^2+\tfrac{1}{2}\sigma^2))(1-\Phi(W_2))| \leqq 4CN^{-1/2}A^{N/2}G(a^1,a^2),\end{aligned}$$

where C is a constant depending on the moments of μ up to the third order.

Proof. $1 - B_N(a^1, a^2)$ is just $I^N + J^N$, where

$$I^N = a^1\int_{S^N}\exp(y^1)\,\mu_*^N(dy), \quad S^N = \{y\in\Pi_{NB}!\quad y^1\leqq y^2+\log(a^2/a^1)\}$$

$$J^N = a^2\int_{T^N}\exp(y^2)\,\mu_*^N(dy), \quad T^N = \{y\in\Pi_{NB}!\quad y^1 > y^2+\log(a^2/a^1)\},$$

since the support of μ_*^N is in $\Pi_{NB} = \{y\in R^2\colon y^1+y^2 = NB\}$, $B=\log A$. Change coordinates in I^N as follows:

$$y^1 = NY^1 + N^{1/2}\sigma(u+v), \quad y^2 = NY^2+N^{1/2}\sigma(u-v).$$

Let ν_N denote the measure induced on R^2 from μ_*^N by the map sending y to (u,v). The support of ν_N is clearly in the line $u=0$ and ν_N can be described by a distribution F_N on this line. Then

$$(6.2)\qquad I^N = a^1\exp(NY^1)\int_{-\infty}^{w}\exp(N^{1/2}\sigma v)\,dF_N,$$

where $w = \frac{1}{2}N^{-1/2}\sigma^{-1}\{\log(a^2/a^1) - N(Y^1-Y^2)\}$. Moreover, F_N is the distribution function obtained from F_1 by N-fold convolution followed by contraction by the factor $N^{-1/2}$ and $\int_{-\infty}^{+\infty} v\,dF_N = 0$. F_1 has moments of all orders, hence the Berry-Essen theorem asserts that if $\Delta_N = F_N - \Phi$, $|\Delta_N(v)|\leqq CN^{-1/2}$, where C is a constant depending on the moments of F_N (hence of μ) up to the third order. If F_N is replaced by $\Delta_N+\Phi$ in (6.2), two integrals are obtained. The first can be estimated by an integration by parts:

$$\begin{aligned}&\left|a^1\exp(NY^1)\int_{-\infty}^{w}\exp(N^{1/2}\sigma v)\,d\Delta_N\right| \leqq \left|a^1\exp(NY^1+N^{1/2}\sigma w)\Delta_N(w)\right|\\ &+\left|a^1N^{1/2}\sigma\exp(NY^1)\int_{-\infty}^{w}\Delta_N(v)\exp(N^{1/2}\sigma v)\,dv\right| \leqq 2CN^{-1/2}A^{N/2}G(a^1a^2).\end{aligned}$$

The second integral is

$$\begin{aligned}&(2\pi)^{-1/2}a^1\exp(NY^1)\int_{-\infty}^{w}\exp(N^{1/2}\sigma v - \tfrac{1}{2}v^2)\,dv\\ &\qquad = a^1\exp(N(Y^1+\tfrac{1}{2}\sigma^2))\Phi(w-N^{1/2}\sigma).\end{aligned}$$

Similar estimates apply to J^N and lead to integrals, one of which is in absolute value not greater than $2CN^{-1/2}A^{N/2}G(a^1, a^2)$, while the other is

$$a^2\exp(N(Y^2+\tfrac{1}{2}\sigma^2))(1-\Phi(w+N^{1/2}\sigma)).$$

This proves Theorem 6.1.

There are other forms of the remainder in the central limit theorem which can be used in a similar way to estimate $B_N(a^1, a^2)$. Theorems of this type can be found in [7, chapter 8]. However the multivariate forms of these theorems do not seem to have been worked out. One can employ the asymptotic series for Φ to estimate B_N by formula (6.1). However, there are different cases to the considered depending on the relative size of $Y^2 - Y^1$ and σ. We have therefore not listed these results.

It is also possible to generalize Theorem 6.2 to sequences of experiments $\{m_{11}, m_{12}\}$, $\{m_{21}, m_{22}\}, \ldots$; on (X_1, Ω_1), $(X_2, \Omega_2), \ldots$ respectively. To obtain a result analogous to (6.1), one needs the assumption that if $B_i^N = \frac{1}{N} \sum_{j=1}^{N} \beta_{ij}, \beta_{ij} = i$-th absolute moment of associal measure of $\{m_{j1}, m_{j2}\}$ (view as a measure on a line), then

$$\limsup B_3^N / B_2^N)^{3/2} < \infty \quad \text{as} \quad N \to \infty .$$

The proof is essentially the same as the proof of Theorem 6.1, except that the Berry-Essen theorem must be replaced by a theorem of Essen [6, p. 43]. We omit the exact statement of this generalization of Theorem 6.1.

References

1. Blackwell, D.: Comparison of experiments. Proc. Second Berkeley Sympos. math. Statist. Probability 93—102 (1951).
2. — Equivalent comparisons of experiments. Ann. math. Statistics **29**, 265—272 (1953).
3. Bergström, H.: On the central limit theorem in the space R_k, $k > 1$. Skand. Aktuarietidskr. **28**, 106—127 (1945).
4. Berry, A. C.: The accuracy of the Gaussian approximation to the sum of independent variates. Trans. Amer. math. Soc. **49**, 122—136 (1941).
5. Cencov, N. N.: The categories of mathematical statistics. Doklady Akad. Nauk S.S.S.R. u. Ser. **165**, 511—514 (1965).
6. Essen, C. G.: Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law. Acta math. **77**, 1—125 (1945).
7. Gnedenko, B. V., and A. N. Kolmogorov: Limit distributions of sums of random variables. (Trans. by K. L. Chung.) Cambridge, Mass.: Addison-Wesley 1954.
8. Halmos, P. R., and L. J. Savage: Application of the Radon-Nikodym theorem to the theory of sufficient statistics. Ann. math. Statistics **20**, 225—241 (1949).
9. Morse, N., and R. Sacksteder: Statistical isomorphism. Ann. math. Statistics **37**, 203—214 (1966).
10. Sacksteder, R.: On the structure and properties of statistical systems. Trans. N.Y. Acad. Sci. 28, 998—1004 (1966).
11. — A note on statistical equivalence. Ann. math. Statistics **38**, 784—794 (1967).

Dr. Richard Sacksteder
The City University of New York
Department of Mathematics
Graduate Center: 33, West 42nd Street
New York, N.Y. 10036
USA