MATHEMATIK-ARBEITSPAPIERE

SOME RESULTS ON CONVERGENCE OF MOMENTS AND CONVERGENCE IN DISTRIBUTION WITH APPLICATIONS IN STATISTICS

GERHARD OSIUS

PREPRINT NR. 33 MARCH 1989

UNIVERSITÄT BREMEN

MATHEMATIK-ARBEITSPAPIERE

A: MATHEMATISCHE FORSCHUNGSPAPIERE

SOME RESULTS ON CONVERGENCE OF MOMENTS AND CONVERGENCE IN DISTRIBUTION WITH APPLICATIONS IN STATISTICS

GERHARD OSIUS

PREPRINT NR. 33 MARCH 1989

FACHBEREICH MATHEMATIK/INFORMATIK
UNIVERSITÄT BREMEN

Bibliothekstraße
28334 Bremen
Germany

Some Results on Convergence of Moments and Convergence in Distribution with Applications in Statistics

Gerhard Osius

University of Bremen

Contents

Summary

Many asymptotical results in statistics state only convergence in distribution although convergence of certain moments (like expectation and variance) are also of interest and sometimes no more difficult to derive. We consider notions of convergence which require convergence in distribution together with convergence of certain moments or even convergence of the moment-generating function of the norm. The common tools - like the central limit theorem, Slutzky's rules, Cramèr-Wold's device and the delta-method - for establishing convergence in distribution are extended to the stronger notions. For illustration, selected statistical applications are given concerning transformations of binomial rates, the log-odds-ratio, the linear model and sample quantiles (e.g. the median).

AMS 1980 subject clasifications. Primary 60B10, 60F05, secondary 62F12, 62E20.

Key words and phrases. Convergence of moments and moment generating functions, central limit theorem, delta-method, transformation of binomial rates, log-odds ratio, linear model, sample quantiles, median.

0. Introduction

Many asymptotical results in statistics state that the distribution of a random variable X_n (e.g. a properly scaled test statistic or parameter) converges to the distribution of some variable X as the number n of observations tends to infinity. Often it is desirable to know whether the expectation and variance of X_n converge to those of X. More generally one may ask for which $r \geq 1$ one has convergence of the absolute moments of order r:

$$E(\|X_n\|^r) \longrightarrow E(\|X\|^r)$$

Bickel & Freedman (1981) introduced a Mallows metric d_r for probability measures, such that *d_r-convergence* is equivalent to convergence in distribution *and* convergence of absolute moments of order r. An even stronger concept of *e_r-convergence*, with $r > 0$, requires convergence in distribution together with convergence of the moment generating function of $\|X_n\|$ to the one of $\|X\|$ at the point r:

$$E(\exp(r\|X_n\|)) \longrightarrow E(\exp(r\|X\|)) .$$

It turns out, that many important results concerning convergence in distribution may be generalized to the stronger concepts of d_r- or e_r-convergence provided the relevant moments exists. The main purpose of this paper is to point out how derivations of statistical results concerning convergence in distribution may be generalized to d_r- or e_r-convergence. To allow such generalizations in a straight forward manner, we first extend the most common tools for establishing convergence in distribution - namely the central limit theorem, Slutzky's rules, Cramèr-Wold's device and the stochastic Taylor formula or Delta-method - to the above concepts of convergence. To illustrate the general method of generalization, some examples of limit theorems relevant in statistical applications are outlined concerning the asymptotic distribution of estimators for transformations of binomial rates (e.g. probit or logit), the log-odds-ratio for 2x2-tables, parameters in linear models, and sample quantiles (e.g. the median).

The notions of d_r- and e_r-convergence are introduced on separable Banach spaces with particalur emphasis on finite dimensional Euclidean spaces $\mathbb{R}^K$, where all concrete examples are studied. To illustrate the relative strength of the above concepts of convergence, a number of counterexamples are provided.

1. The Concepts of d- and e-Convergence

Let B be a separable Banach space with norm $\|\cdot\|$, e.g. a finite-dimensional Euclidean space $\mathbb{R}^K$ with Euklidean norm. For $r \geq 0$ let $\mathfrak{D}_r = \mathfrak{D}_r(B)$ be the set of probability measures P on the Borel sets of B, such that the r-th absolute moment is finite, i.e. $\int \|x\|^r \, P(dx) < \infty$. Thus the distribution $\mathfrak{L}(X)$ of a random element X in B belongs to $\mathfrak{D}_r$ iff $E(\|X\|^r) < \infty$.

We say that a sequence $P_n \in \mathfrak{D}_r$ *converges (weakly) of order r* to $P \in \mathfrak{D}_r$, abbreviated $P_n \xrightarrow{d_r} P$, iff P_n converges weakly to P and

(1) $$\int \|x\|^r \, P_n(dx) \longrightarrow \int \|x\|^r \, P(dx) \quad \text{as } n \longrightarrow \infty.$$

For notational convenience and an intuitive understanding we prefer in the sequel formulations in terms of random elements (and their distributions) rather than probability measures. Thus for random elements X_n, X with distributions P_n, $P \in \mathfrak{D}_r$ we say that X_n *converges in distribution of order r* iff $P_n \xrightarrow{d_r} P$, i.e. $X_n \xrightarrow{\mathfrak{L}} X$ converges in distribution and the r-th absolute moment of X_n converges to the one of X:

(2) $$E(\|X_n\|^r) \longrightarrow E(\|X\|^r) \quad \text{as } n \longrightarrow \infty .$$

Bickel & Freedman (1981) originally introduced the concept of convergence of order $r \geq 1$ via a Mallows metric d_r on $\mathfrak{D}_r$, such that $d_r(P_n, P) \to 0$ is equivalent to $P_n \xrightarrow{d_r} P$ above. They also gave the following important characterizations of convergence of order r, which will be used freely later .

Lemma 1: For random elements X_n, X with distributions in $\mathfrak{D}_r$ the following conditions are equivalent

(i) $X_n \xrightarrow{d_r} X$

(ii) $X_n \xrightarrow{\mathfrak{L}} X$ and $\|X_n\|^r$ is uniformly integrable for $n \in \mathbb{N}$.

(iii) $E(f(X_n)) \longrightarrow E(f(X))$ for any continuous function $f: B \longrightarrow \mathbb{R}$ such that $|f(x)|/(1+\|x\|^r)$ is bounded.

Proof: For (i) $\Leftrightarrow$ (ii) see Billingsley (1968), Thm. 5.4.

(ii) $\Rightarrow$ (iii): Since $f(X_n) \xrightarrow{\mathfrak{L}} f(X)$ it suffices to show that $f(X_n)$ is uniformly integrable, which follows from the condition on f, since $\|X_n\|^r$ is uniformly integrable.

(iii) $\Rightarrow$ (i): Application of (iii) to all bounded continuous functions f gives $X_n \xrightarrow{\mathfrak{L}} X$, and (2) follows for $f(x) = \|x\|^r$. □

In view of (ii) the notion of d_r-convergence (and the set $\mathfrak{D}_r$) is invariant under equivalent norms on B. Convergence of order r implies convergence of any lower order $0 \le s \le r$, and convergence of order 0 reduces to convergence in distribution. Convergence of *infinite* order $X_n \xrightarrow{d_\infty} X$ resp. *positive* order $X_n \xrightarrow{d_+} X$ is defined as convergence for *all* resp. *some* $r > 0$, provided the distributions P_n, P are in $\mathfrak{D}_\infty := \bigcap_r \mathfrak{D}_r$ resp. $\mathfrak{D}_+ := \bigcup_r \mathfrak{D}_r$.

We now introduce the concept of convergence in distribution of *exponential* order, which is stronger than d_∞-convergence and closely related to convergence of moment generating functions (short: MGF). For $r \ge 0$ let $\mathfrak{E}_r = \mathfrak{E}_r(B)$ be the set of all probability measures on the Borel sets of B, such that $\int \exp(r\|x\|)\, P(dx) < \infty$. Thus the distribution $\mathfrak{L}(X)$ of a random element X in B belongs to $\mathfrak{E}_r$ iff $E\{\exp(r\|X\|)\} < \infty$, i.e. the MGF of $\|X\|$ is finite at the point r.

We say, that a sequence $P_n \in \mathfrak{E}_r$ *converges (weakly) of exponential order r*, abbreviated $P_n \xrightarrow{e_r} P$, iff P_n converges weakly to P and

(3) $$\int \exp(r\|x\|) P_n(dx) \longrightarrow \int \exp(r\|x\|) P(dx) \quad \text{as} \quad n \longrightarrow \infty .$$

As usual, we use the same notation of convergence for random elements X_n, X as for their distributions $P_n = \mathfrak{L}(X_n)$, $P = \mathfrak{L}(X)$, i.e. $X_n \xrightarrow{e_r} X_n$ means convergence in distribution and convergence of the MGF of $\|X_n\|$ to the one of $\|X\|$ at the point r:

(4) $$E(\exp \|rX_n\|) \longrightarrow E(\exp\|rX\|).$$

Equivalent characterizations of convergence of exponential order are given below.

Lemma 2: For random elements X_n, X with distributions in $\mathfrak{E}_r$ the following conditions are equivalent

(i) $X_n \xrightarrow{e_r} X$

(ii) $X_n \xrightarrow{\mathfrak{L}} X$ and $\exp(r\|X_n\|)$ is uniformly integrable for $n \in \mathbb{N}$.

(iii) $E(f(X_n)) \longrightarrow E(f(X))$ for any continuous $f: B \longrightarrow \mathbb{R}$ such that $f(x)/\exp(r\|x\|)$ is bounded.

(iv) $X_n \xrightarrow{\mathfrak{L}} X$ and $\exp(\|X_n\|) \xrightarrow{d_r} \exp(\|X\|)$.

For the real line $B = \mathbb{R}$ and $r > 0$ the following condition is equivalent to (i)-(iv):

(v) The MGF of X_n converges pointwise in $[-r, r]$ to the MGF of X.

Proof: (i) $\Leftrightarrow$ (ii) $\Leftrightarrow$ (iii) follows similarly to Lemma 1, and (i) $\Leftrightarrow$ (iv) is obvious by definition. For (iii) $\Rightarrow$ (v) take $|t| \le r$ and apply (iii) to $f(x) = \exp(tx)$. To prove (v) $\Rightarrow$ (ii), observe first $X_n \xrightarrow{\mathfrak{L}} X$, see e.g. Billingsley (1979), Sec. 30. Hence $\exp(\pm r X_n) \xrightarrow{\mathfrak{L}} \exp(\pm rX)$ and thus $\exp(\pm rX_n)$ is uniformly integrable, see Billingsley (1969), Thm. 5.4. □

Again, convergence of exponential order r implies convergence of any lower exponential order, and convergence of exponential order 0 reduces to convergence in distribution.

Convergence of *infinite* exponential order $X_n \xrightarrow{e_\infty} X$ resp. of *positive* exponential order $X_n \xrightarrow{e_+} X$ is defined as $X_n \xrightarrow{e_r} X$ for *all* resp. *some* $r>0$, provided the distributions are in $\mathfrak{E}_\infty := \bigcap_r \mathfrak{E}_r$ resp. $\mathfrak{E}_+ := \bigcup_r \mathfrak{E}_r$. The notions of e_∞-resp. e_+-convergence are invariant under equivalent norms on B, in contrast to e_r-convergence which depends on the norm.

Convergence of positive exponential order implies convergence of infinite order:

(5) $$X_n \xrightarrow{e_+} X \quad \Rightarrow \quad X_n \xrightarrow{d_\infty} X .$$

Proof: Suppose $X_n \xrightarrow{e_r} X$ for some $r>0$. For any $k \in \mathbb{N}$, $\|X_n\|^k \le k!\, r^{-k} \exp(r\|X_n\|)$ is uniformly integrable by Lemma 2(ii) and hence $X_n \xrightarrow{d_k} X$ follows by Lemma 1(ii). □

2. Basic Properties

We now derive the basic properties for convergence of order r which correspond to well known results on convergence in distribution. Similar results for exponential orders may be derived in view of Lemma 2(iv), some of which are explicitely given below. The results also hold for infinite order, i.e. we allow $0 \le r \le \infty$.

Continuous Functions and d-Convergence

Let X_n, X be random elements on B, and let B′ be another separable Banach space. For a continuous *linear* function A: $B \to B'$ and a constant $b \in B'$ one has

(1) $$X_n \xrightarrow{d_r} X \quad \Rightarrow \quad (AX_n + b) \xrightarrow{d_r} (AX + b) \ .$$

For a general continuous function f: $B \to B'$, such that $\|f(x)\| / (1+\|x\|^s)$ is bounded for some $s>0$, the following holds

(2) $$X_n \xrightarrow{d_{rs}} X \quad \Rightarrow \quad f(X_n) \xrightarrow{d_r} f(X)$$

Note that application of f reduces the order of convergence if $s>1$ (and $r>0$).

Proof: (1) follows from (2) since $\|f(x)\| \le (1+\|x\|)(\|A\|+\|b\|)$. In (2) $X_n \xrightarrow{\mathfrak{L}} X$ implies $f(X_n) \xrightarrow{\mathfrak{L}} f(X)$ by Billingsley (1968), Thm.5.1. By assumption, there is an $M>0$ such that $\|f(x)\| \le M \cdot \mathrm{Max}\{1, \|x\|^s\}$ and hence $\|f(X_n)\|^r \le M^r \cdot \mathrm{Max}\{1, \|X_n\|^{rs}\}$. The uniform integrability of $\|f(X_n)\|^r$ follows from the one of $\|X_n\|^{rs}$, which proves (2). □

Continuous Functions and e-Convergence

For a continuous *linear* function A: $B \to B'$ with norm $s = \|A\|$ and a constant $b \in B'$ one has

(3) $$X_n \xrightarrow{e_{rs}} X \quad \Rightarrow \quad (AX_n + b) \xrightarrow{e_r} (AX + b) \ .$$

For a general continuous function f: $B \to B'$, such that $\|f(x)\| \le s\|x\| + t$ for some $s,t>0$, the following holds

(4) $$X_n \xrightarrow{e_{rs}} X \quad \Rightarrow \quad f(X_n) \xrightarrow{e_r} f(X) \ .$$

The proofs are similar to those of (1)(2).

Constant Random Elements

For $a_n, a \in B$ viewed as constant random elements, usual convergence is equivalent to convergence in distribution of any (exponential) order:

(5) $$\lim_n a_n = a \quad \Longleftrightarrow \quad a_n \xrightarrow{d_r} a \quad \Longleftrightarrow \quad a_n \xrightarrow{e_r} a \ .$$

Proof: $\lim a_n = a$ is equivalent to $a_n \xrightarrow{\mathfrak{L}} a$, and $\|a_n\|^r$, $\exp(r\|a_n\|)$ are uniformly integrable. □

Pairs of Random Elements

Let X_n, X resp. Y_n, Y be random elements on B resp. B′ and consider the pairs (X_n, Y_n), (X,Y) on the product space $B \times B'$ with the maximum-norm $\|(x,y)\| = \max(\|x\|, \|y\|)$.

If X_n resp. X is independent of Y_n resp. Y (for all n) *or* if X or Y is constant, then

(6) $\quad (X_n, Y_n) \xrightarrow{d_r} (X,Y) \iff X_n \xrightarrow{d_r} X$ and $Y_n \xrightarrow{d_r} Y$.

(7) $\quad (X_n, Y_n) \xrightarrow{e_r} (X,Y) \iff X_n \xrightarrow{e_r} X$ and $Y_n \xrightarrow{e_r} Y$.

Proof : We only prove (6), and (7) follows similarly. Convergence in distribution $(r=0)$ follows from Billingsley (1968), Thm. 3.2 & 4.4 and the continuity of projections. For general $r \geq 0$, note that $\|(X_n, Y_n)\|^r = \mathrm{Max}(\|X_n\|^r, \|Y_n\|^r)$ is uniformly integrable iff $\|X_n\|^r$ and $\|Y_n\|^r$ are. □

Sums and Products (Slutzky's Rules)

Let X_n, X, Y_n, Y be random elements on B and Z_n, Z random variables. Then

(8) $\quad$ X or Y constant, $X_n \xrightarrow{d_r} X$, $Y_n \xrightarrow{d_r} Y \Rightarrow (X_n \pm Y_n) \xrightarrow{d_r} (X \pm Y)$,

(9) $\quad$ X or Y constant, $X_n \xrightarrow{e_{2r}} X$, $Y_n \xrightarrow{e_{2r}} Y \Rightarrow (X_n \pm Y_n) \xrightarrow{e_r} (X \pm Y)$,

(10) $\quad$ X or Z constant, $X_n \xrightarrow{d_{2r}} X$, $Z_n \xrightarrow{d_{2r}} Z \Rightarrow (Z_n \cdot X_n) \xrightarrow{d_r} (Z \cdot X)$.

If B is a Hilbert space with an inner product $\langle -,- \rangle$ then in addition

(11) $\quad$ X or Y constant, $X_n \xrightarrow{d_{2r}} X$, $Y_n \xrightarrow{d_{2r}} Y \Rightarrow \langle X_n, Y_n \rangle \xrightarrow{d_r} \langle X, Y \rangle$.

Note the reduction of the order of convergence from 2r to r in (9-11).

Proof: To prove (8), note that $(X_n, X_n) \xrightarrow{d_r} (X,Y)$ by (6). The function $f(x,y) = x \pm y$ satisfies $\|f(x,y)\| \leq 2\,\|(x,y)\|$ for the maximum norm and (8) follows from (2) with $s=1$. (9) follows in the same way from (4). And (10)(11) are proved using $\|z \cdot x\| \leq \|(x,z)\|^2$, $|\langle x,y \rangle| \leq \|(x,y)\|^2$. □

3. Euclidean Spaces

We now look in particular to finite dimensional euclidean spaces $\mathbb{R}^K$ with $K \in \mathbb{N}$. The Cramér-Wold device to reduce convergence of random vectors to random variables easily extends to convergence of order r. In fact, for K-dimensional random vectors X_n, X one has

(1) $$X_n \xrightarrow{d_r} X \iff \langle t, X_n \rangle \xrightarrow{d_r} \langle t, X \rangle \text{ for any } t \in \mathbb{R}^K .$$

The corresponding result for e_r-convergence does not hold, since the exponential order is not invariant under linear functions. However, we have

(2) $$X_n \xrightarrow{e_+} X \iff \langle t, X_n \rangle \xrightarrow{e_+} \langle t, X \rangle \text{ for any } t \in \mathbb{R}^K .$$

(3) $$X_n \xrightarrow{e_\infty} X \iff \langle t, X_n \rangle \xrightarrow{e_\infty} \langle t, X \rangle \text{ for any } t \in \mathbb{R}^K .$$

Proof : We only prove (1), since (2)(3) follow similarly. "$\Rightarrow$" follows from 2(1).
"$\Rightarrow$": The Cramér-Wold device, e.g. Billingsley (1968) Thm. 7.7, gives $X_n \xrightarrow{\mathcal{L}} X$. If t_k is the k-th unit vector, one has for the maximum norm $\|X_n\|^k = \operatorname{Max}_k |\langle t_k X_n \rangle|^r$ and hence $\|X_n\|^k$ is uniformly integrable, for all k. □

In statistical applications d_1- and d_2-convergence are particulary important, because interest often focuses on expectations and variances of random variables. For r=1 one has [Billingsley(1968), Thm.5.4]

(4) $$X_n \xrightarrow{d_1} X \Rightarrow E(X_n) \longrightarrow E(X),$$

and convergence of order r=2 may be described as follows

(5) $$X_n \xrightarrow{d_2} X \iff$$
$$X_n \xrightarrow{\mathcal{L}} X, \quad E(X_n) \longrightarrow E(X), \quad \operatorname{Cov}(X_n) \longrightarrow \operatorname{Cov}(X),$$

where $\operatorname{Cov}(X)$ denotes the $(K \times K)$-covariance-matrix of X.

Proof of (5): "$\Rightarrow$" In view of (4) it remains to show for any k,l = 1,...,K:

$$\operatorname{Cov}(X_{nk}, X_{nl}) = E(X_{nk}X_{nl}) - E(X_{nk}) \cdot E(X_{nl}) \longrightarrow \operatorname{Cov}(X_k, X_l)$$

resp. $E(X_{nk}X_{nl}) \longrightarrow E(X_k X_l)$. And this follows from $X_{nk}X_{nl} \xrightarrow{\mathcal{L}} X_k X_l$ and the uniformly integrability of $|X_{nk}X_{nl}| \le \|X_n\|^2$, see Billingsley (1968) Thm. 5.4.
"$\Leftarrow$" For any k=1,...,K, we have $X_{nk}^2 \xrightarrow{\mathcal{L}} X_k^2$ and $E(X_{nk}^2) = \operatorname{Var}(X_{nk}) + E(X_{nk})^2$ converges to $\operatorname{Var}(X_k) + E(X_k)^2 = E(X_k^2)$. Hence X_{nk}^2 is uniformly integrable [Billingsley (1968), Th. 5.4] for all k, and so is $\|X_n\|^2 = \operatorname{Max}_k |X_{nk}|^2$. □

Empirical Distributions

Let $X_1,\dots,X_n$ be i.i.d. random variables with distribution function F and denote their empirical distribution function by F_n. For any measurable function g we have by Kchinchin's weak law of large numbers

(1) $$\int g(x)\,F_n(dx) \longrightarrow \int g(x)\,F(dx) = E(g(X))$$

provided $E(g(X))$ exists and is finite. Hence the well known convergence of the empirical distribution F_n to the true distribution F is of any (exponential) order $r>0$, provided the relevant moment of F exists and is finite:

(2) $$E(|X|^r) < \infty \quad \Longrightarrow \quad F_n \xrightarrow{d_r} F.$$

(3) $$E(\exp(r|X|)) < \infty \quad \Longrightarrow \quad F_n \xrightarrow{e_r} F.$$

4. Central Limit Theorems

So far we have mainly given rules to deduce new convergence results from old ones. Now we we will see, that a major tool for establishing convergence in distribution in the first place, namely the central limit theorem (CLT), extends to the stronger concepts of d_r- and e_r-convergence. In both cases, we first prove a general, but rather technical, theorem and state some important special situations as corollaries.

Central Limit Theorem (CLT) for d-Convergence

For each $n \in \mathbb{N}$ let X_{ni} be independent random variables for $i=1,\dots,I(n)$ with $E(X_{ni}) = 0$ and finite variance $Var(X_{ni}) < \infty$. Let $S_n = \sum_i X_{ni}$ and suppose $\sigma_n^2 := Var(S_n) > 0$. Furthermore let $K \geq 4$ be an *even* integer and assume the existence of constants $0 < M_n < \infty$ such that

(i) $$\left| \prod_{i=1}^{I(n)} E(X_{ni}^{k_i}) \right| \leq M_n$$

for any $k_1,\dots,k_{I(n)} \in \{0,\dots,K\}$ such that $k_1+\dots+k_{I(n)} = K$.

(ii) $I(n)\, M_n^{2/K} \cdot \sigma_n^{-2}$ is bounded as $n \to \infty$

Then, as $n \to \infty$ and $I(n) \to \infty$, the standardized sum convergences to its normal limit of order d_K:

(iii) $$\sigma_n^{-1} S_n \xrightarrow{d_K} N(0,1) \quad .$$

Remark 1: The factors $E(X_{ni}^{k_i})$ in (i) are 1 resp. 0 for $k_i = 0$ resp. 1. Hence the product need only be taken over the set $\{i \mid k_i \geq 2\}$ which has at most $K/2$ elements.

Remark 2: (i) and (ii) are *joint* conditions on the bounds M_n for the moments and the average variance $\sigma_n^2 / I(n)$ of $X_{n1},\dots,X_{nI(n)}$. The following two *separate* conditions imply (i) and (ii):

(i)* $E\{X_{ni}^K\}$ is bounded for all n and all i.

(ii)* $I(n)/\sigma_n^2$ is bounded for all n.

Remark 3: Note, that (iii) for $K=2$ is equivalent to convergence in distribution since $\sigma_n^{-1} S_n$ is standardized.

Corollary 1:

Suppose in addition (in the CLT) for all $n \in \mathbb{N}$, that X_{ni} are i.i.d. variables with variances bounded away from 0 and finite K-th moment $\mu_{nK} = E\{X_{n1}^K\} < \infty$. If the moments μ_{nK} are bounded for $n \to \infty$, then (i)*(iii)* hold and hence (iii) follows.

Corollary 2:

Take $I(n) = n$ in the CLT and let the random variables be of the form $X_{ni} = a_{ni} Y_i$ for real numbers a_{ni} and i.i.d. random variables $Y_1, Y_2, \ldots$ with zero expectation, unit variance and finite K-th moment $\mu_K = E\{Y_1^K\}$. Then the assumptions (i)(ii) and hence (iii) follow from the following conditions on the numbers a_{ni}:

(iv) $\|a_n\|^2 := \sum_i a_{ni}^2 > 0$ (i.e. $a_n \neq 0$)

(v) $\sqrt{n}\ \|a_n\|^{-1} \cdot A_n$ is bounded for $n \to \infty$ with $A_n := \operatorname{Max}_i |a_{ni}|$.

Note, that (iv)(v) hold in the "i.i.d. case" given by $a_{ni} = 1$ for all n and i.

Proof (CLT)

Taking $r = K/2 \geq 2$, Ljapunov's condition follows from

$$\sigma_n^{-K} \sum_i E(X_{ni}^K) \leq \sigma_n^{-K} I(n) M_n = \left[I(n) M_n^{1/r} \sigma_n^{-2}\right]^r \cdot I(n)^{1-r}$$

and (ii), since $I(n) \to \infty$. Hence we have convergence in distribution $\sigma_n^{-1} S_n \xrightarrow{\mathcal{L}} N(0,1)$. This implies [Billingsley (1968), Th. 5.3]:

$$E\left([N(0,1)^2]^r\right) \leq \liminf_n E\left([\sigma_n^{-2}\ S_n^2]^r\right)$$

and hence (iii) holds if we establish

(a) $$\limsup_n E\left(\sigma_n^{-K}\ S_n^K\right) \leq E\left(N(0,1)^K\right) .$$

Now $$S_n^K = K!\ \sum_k \prod_i k_i!^{-1}\ X_{ni}^{k_i} ,$$

where the sum ranges over all $k = (k_1, \ldots, k_{I(n)}) \in \{0, \ldots, K\}^{I(n)}$ such that $k_+ := \sum_i k_i = K$.

Hence $$E\left(\sigma_n^{-K} S_n^K\right) = \sigma_n^{-K}\ K! \sum_k \prod_{i \in J} E\{X_{ni}^{k_i}\} .$$

The sum can be restricted to all k such that $k_i \neq 1$ for all i, since $E(X_{ni}) = 0$. And the product for such k need only be taken over the set $J(k) = \{i \mid k_i \geq 2\}$ which has at most r elements. Since each k is uniquely determined by its values on $J(k)$ we may rearrange the sum above to get

(b) $$E\left(\sigma_n^{-K} S_n^K\right) = K! \sum_{j=1}^{r} E_{nj} \qquad \text{with}$$

(c) $$E_{nj} = \sigma_n^{-K} \sum_{|J|=j} \sum_{k\in \mathfrak{K}(J)} \prod_{i\in J} k_i!^{-1} E\{X_{ni}^{k_i}\} \quad ,$$

where the first sum ranges over all subsets $J\subset\{1,\dots,I(n)\}$ with j elements and the second sum extends over the set $\mathfrak{K}(J) = \{k\in\{2,\dots,K\}^J \mid k_+ = K\}$ which has at most $K^{|J|}$ elements. Now (i) yields

$$\begin{aligned} |E_{nj}| &\le \sigma_n^{-K} \sum_{|J|=j} \sum_{k\in \mathfrak{K}(J)} M_n \\ &\le \sigma_n^{-K} \binom{I(n)}{j} K^j M_n \\ &\le I(n)^{-j}\binom{I(n)}{j} \cdot \left[I(n)\cdot M_n^{1/r}\sigma_n^{-2}\right]^r \cdot I(n)^{j-r} \cdot K^j \ . \end{aligned}$$

By $I(n)^{-j}\binom{I(n)}{j} \longrightarrow j!^{-1}$ as $I(n)\to\infty$ and (ii) we get

(d) $$E_{nj} \longrightarrow 0 \text{ for all } j<r \ .$$

Now looking at j=r in (c), we observe that $\mathfrak{K}(J)$ has for $|J|=r$ exactly one element, given by $k_i=2$ for all i. Hence

$$E_{nr} = \sigma_m^{-K} 2^{-r} \sum_{|J|=r} \prod_{i\in J} E(X_{ni}^2) \ .$$

Now $$\sigma_n^{2r} = \left[\sum_i E(X_{n_i}^2)\right]^r \ge r! \sum_{|J|=r} \prod_{i\in J} E(X_{ni}^2)$$

gives $$K!\, E_{nr} \le K!\, 2^{-r}\, r!^{-1} = E\left(N(0,1)^K\right) \ .$$

Together with (b)(d) we now conclude (a). □

Proof of Remark 2

By (i)* there is a bound $M\ge 1$ such that $E\{|X_{ni}|^k\}<M$ for all $k=0,\dots,K$. To prove (i) in view of remark 1 we observe

$$\left|\prod_i E\left(X_{ni}^{k_i}\right)\right| \le \prod_i E\left(|X_{ni}|^{k_i}\right) \le M^{K/2}$$

So (i) holds for the constant sequence $M_n := M^{K/2}$ and (ii) is then equivalent to (ii)*. □

Proof of Corollary 1

Apply Remark 2 and observe that $\sigma_n^2/I(n) = \mathrm{Var}(X_{n1})$. □

Proof of Corollary 2

To prove (i), let $M := \operatorname*{Max}_{k=0,\dots,K} E\{|Y_1|^k\} < \infty$, $A_n := \operatorname*{Max}_i |a_{ni}|$ and observe

$$\left|\prod_i E\left(X_{ni}^{k_i}\right)\right| \le \prod_i |a_{ni}|^{k_i} E\left(|Y_1|^{k_i}\right)$$
$$\le A_n^K \cdot \prod_i E\left(|Y_1|^{k_i}\right) \le A_n^K \cdot M^{K/2} =: M_n$$

where the last inequality follows from remark 1. Hence (i) holds and (ii) follows from $\sigma_n^2 = \|a_n\|^2$ and (v). □

Central Limit Theorem (CLT) for e-Convergence

For each $n \in \mathbb{N}$ let X_{ni} be independent random variables for $i = 1,\dots,I(n)$ with $E(X_{ni}) = 0$ and finite variance $Var(X_{ni}) < \infty$. Let $S_n := \sum_i X_{ni}$ and suppose $\sigma_n^2 := Var(S_n) > 0$. Furthermore let $r > 0$ and assume

(i) The MGF $M_{ni}(s)$ of X_{ni} is finite for $|s| \le r\sigma_n^{-1}$

(ii) There exists a bound c_n for the third derivative of $\log M_{ni}$

$|(\log M_{ni})'''(t)| \le c_n < \infty$ for all i, $|s| \ge r\sigma_n^{-1}$ and almost all n,

such that $I(n)\sigma_n^{-3} c_n \longrightarrow 0$ as $n \to \infty$.

Then the standardized sum converges to its normal limit of order e_r as $n \to \infty$.

(iii) $$\sigma_n^{-1} S_n \xrightarrow{e_r} N(0,1) .$$

Remark 1: If $I(n) \to \infty$, then (ii) follows from the two conditions

(ii)* $I(n)/\sigma_n^2$ is bounded.

(iii)* c_n is bounded .

Corollary 3

Take $I(n) = n$ in the CLT above and let the random variables be of the form $X_{ni} = a_{ni}Y_i$ for i.i.d. random variables $Y_1, Y_2, \dots$, with mean 0, variance 1, and real numbers a_{ni} satisfying

(iv) $$\|a_n\|^2 := \sum_i a_{ni}^2 > 0 \quad (\text{i.e. } a_n \ne 0)$$

(v) $\sqrt{n}\, \|a_n\|^{-1} \cdot A_n$ is bounded for $n \to \infty$ with $A_n := \operatorname*{Max}_i |a_{ni}|$.

Finally assume

(i)* The MGF of Y_1 is finite on $[-r, +r]$.

Then conditions (i)(ii) hold and hence (iii) follows. Note that (iv)(v) hold in the "i.i.d. case" given by $a_{ni}=1$ for all n, i.

Proof (CLT)

The MGF of $\sigma_n^{-1} S_n$ is $M_n(t) = \prod_i M_{ni}(t\sigma_n^{-1})$ and by Lemma 2(v) is suffices to show

(a) $$\log M_n(t) = \sum_i \log M_{ni}(t\sigma_n^{-1}) \longrightarrow \tfrac{1}{2} t^2 \quad \text{for} \quad |t| \le r .$$

Since $M_{ni}(0) = 1$, $M'_{ni}(0) = E(X_{ni}) = 0$, $\sigma_{ni}^2 := \mathrm{Var}(X_{ni}) = M''_{ni}(0)$, Taylors formula gives $$\log M_{ni}(t\sigma_n^{-1}) = \tfrac{1}{2} \sigma_{ni}^2 t^2 \sigma_n^{-2} + \tfrac{1}{6} c_{ni} t^3 \sigma_n^{-3}$$
with $|c_{ni}| \le c_n$ for almost all n. Hence

$$\log M_n(t) = \tfrac{1}{2} t^2 + \tfrac{1}{6} t^3 \sigma_n^{-3} \sum_i c_{ni}$$

and (a) follows from (ii) since $\left|\sum_i c_{ni}\right| \le I(n) \cdot c_n$ for almost all n. □

Proof of Corollary 3

Denote the MGF of Y_1 by M. By $\sigma_n = \|a_n\|$, $|a_{ni}| \le \sigma_n$ and (i)* we have that $M_{ni}(s) = M(a_{ni}s)$ is finite for $|s| \le r\sigma_n^{-1}$. Furthermore $(\log M_{ni})'''(s) = a_{ni}^3 (\log M)'''(a_{ni} s)$ for $|s| < r\sigma_n^{-1}$ and hence

$$|(\log M_{ni})'''(s)| \le A_n^3 d_n \quad \text{with} \quad d_n = \sup_{|t| < rA_n\sigma_n^{-1}} |(\log M)'''(t)| .$$

Now (v) implies $A_n\sigma_n^{-1} \longrightarrow 0$ and hence $d_n < \infty$ for almost all n. Thus (ii) holds for $c_n = A_n^3 d_n$ (for almost all n) in view of (v). □

For later reference we note two important limit results, being well known in view of Lemma 2(v), which also emerge as special cases of the CLT above.

Example 1: Binomial Limit Theorem

For a sequence X_n of Binomial(n,p_n) variables the standardization of X_n converges to its normal limit of infinite exponential order as $n \to \infty$, provided p_n is bounded away from 0 and 1:

$$(n\,p_n[1-p_n])^{-1/2} \cdot (X_n - n\,p_n) \xrightarrow{e_\infty} N(0,1)\,.$$

To deduce this result from corollary 1 take $n=I(n)$ independent binomial$(1,p_n)$ variables Y_{ni} and put $X_{ni}=Y_{ni}-p_n$. □

Example 2: Poisson Limit Theorem

For a sequence X_n of random variables with Poisson-distribution such that $\lambda_n := E(X_n) \longrightarrow \infty$, the standardized variable converges in distribution of infinite exponential order to its normal limit

$$\lambda_n^{-1/2} \cdot (X_n - \lambda_n) \xrightarrow{e_\infty} N(0,1)\,.$$

To deduce this result from the CLT take $n=I(n)$ independent Poisson variables Y_{ni} with expectation $n^{-1}\lambda_n$ and put $X_{ni}=Y_{ni}-n^{-1}\lambda_n$. □

5. Stochastic Taylor Formula (δ-Method) for d-Convergence

Let Y_n, Z_n be I-dimensional random vectors with values in $V \subset \mathbb{R}$ which converge in probability to an element $a \in \mathbb{R}^I$

(1) $$Y_n \xrightarrow{P} a, \qquad Z_n \xrightarrow{P} a .$$

Suppose further, that the scaled difference converges of order $s \geq 0$ to a random vector U

(2) $$c_n[Y_n - Z_n] \xrightarrow{d_s} U ,$$

with scaling factors $c_n \to \infty$. Now let $G: A \longrightarrow \mathbb{R}^K$ be a continuously differentiable function defined on an open convex neighbourhood $A \subset \mathbb{R}^I$ of a containing V, and consider the random variable

(3) $$S_n = \sup_{0 \leq \xi \leq 1} \left\| DG\big(Y_n + \xi(Z_n - Y_n)\big) \right\| ,$$

where $DG(x)$ denotes the derivative of G at x. Finally suppose for some $t \geq 0$:

(4) $$S_n^t \text{ is uniformly integrable .}$$

Under these assumptions the convergence (2) can be "transformed" via G into

(5) $$c_n[G(Y_n) - G(Z_n)] \xrightarrow{d_r} DG(a) \cdot U$$

with $r = \frac{1}{2}\,\mathrm{Min}\{s,t\}$. Furthermore, the stochastic Taylor formulas hold

(6) $$c_n\big[G(Y_n) - G(Z_n) - DG(a)[Y_n - Z_n]\big] \xrightarrow{d_r} 0 ,$$

(7) $$c_n\big[G(Y_n) - G(Z_n) - DG(Y_n)[Y_n - Z_n]\big] \xrightarrow{d_r} 0 .$$

Remark 1: (2) implies $[Y_n - Z_n] \xrightarrow{P} 0$, so that one condition in (1) follows from the other. If $Z_n = a$ is constant, then (1) follows from (2).

Remark 2: If U has an I-dimensional normal distribution $N(\mu, \Sigma)$, then $D \cdot U$ with $D = DG(a)$ has a K-dimensional normal distribution $N(D\mu, D\Sigma D^T)$.

Remark 3: For compact V, the condition (4) holds for any $t \geq 0$. Taking $t \geq s$ gives $r = \frac{1}{2}s$.

Remark 4: For $s = t = 0$ the condition (4) holds and (2)(5-7) reduce to convergence in distribution, since $r = 0$.

Proof of (5)-(7)

By Taylor's formula, e.g. Dieudonné (1960), (8.14.3) we get

(a) $$c_n[G(Y_n)-G(Z_n)] = c_n H_n[Y_n-Z_n] \quad \text{with}$$
$$H_n = \int_0^1 DG\big(Y_n+\xi(Z_n-Y_n)\big)d\xi \ .$$

For $D = DG(a)$ we have

$$\|H_n-D\| \le \sup_{0\le\xi\le1} \|DG\big(Y_n+\xi(Z_n-Y_n)\big)-D\| .$$

Since the righthand side converges in probability to 0 by (1), we get $[H_n-D] \xrightarrow{P} 0$. Now $\|H_n-D\| \le 2\,\mathrm{Max}\{S_n, \|D\|\}$ implies by (4) that $\|H_n-D\|^t$ is uniformly integrable and thus $[H_n-D] \xrightarrow{d_t} 0$. Since $2r\le s,t$ we conclude by(2) $c_n[H_n-D][Y_n-Z_n] \xrightarrow{d_r} 0$, using Cramér-Wold's device and Slutzky's rule 2(10). Hence (6) follows in view of (a).
By (2) we have $c_n D[Y_n-Z_n] \xrightarrow{d_s} D\cdot U$, so that (6) implies (5). To deduce (7) from (6), observe $DG(Y_n) \xrightarrow{d_t} D$ by (1), since $\|DG(Y_n)\|^t \le S^t$ is uniformly integrable. As above, with $DG(Y_n)$ in place of H_n, we get $c_n[DG(Y_n)-D][Y_n-Z_n] \xrightarrow{d_r} 0$, and (7) follows from (6). □

Example 1 (Log-Transformation)

Let $I=1$, $V=A=(0,\infty)$ and $G=\log$. Then $S_n \le \mathrm{Max}\{Y_n^{-1}, Z_n^{-1}\}$ and (4) holds if Y_n^{-t} and Z_n^{-t} are uniformly integrable.

Example 2 (Probability Integral Transformation)

Let $I=1$, $V=A=(0,1)$ and suppose for some $\alpha\ge0$

(8) $$M := \sup_{0<x<1} x^\alpha(1-x)^\alpha\,|DG(x)| < \infty .$$

Then $S_n \le 2^\alpha\cdot M\cdot \mathrm{Max}\{Y_n^{-\alpha}, (1-Y_n)^{-\alpha}, Z_n^{-\alpha}, (1-Z_n)^{-\alpha}\}$ and (4) follows from the condition

(9) $$Y_n^{-\alpha t}, (1-Y_n)^{-\alpha t}, Z_n^{-\alpha t}, (1-Z_n)^{-\alpha t} \text{ are uniformly integrable.}$$

The most commonly used transformations G for probabilities satisfy (8): the Probit-, Logit-, and log-log-transformation (all $\alpha=1$), as well as the inverse distribution function of the Cauchy-distribution ($\alpha=2$).

6. Counterexamples

Having established the following hierarchy of convergences

$$X_n \xrightarrow{e_\infty} X \;\Rightarrow\; X_n \xrightarrow{e_+} X \;\Rightarrow\; X_n \xrightarrow{d_\infty} X \;\Rightarrow\; X_n \xrightarrow{d_+} X \;\Rightarrow\; X_n \xrightarrow{\mathfrak{L}} X\,,$$

we now show by means of counterexamples that none of the above implications can be reversed. Furthermore we will see that convergence of any (exponential) order does not imply convergence of any higher (exponential) order.

The basic construction for all counterexamples is as follows. For random variables X and Y_n with densities f and g_n we consider the random variable X_n with a mixed distribution having the density $f_n = c_n g_n + (1-c_n)\cdot f$ for some $0<c_n<1$. Thus

(1) $\quad E(h(X_n)) = c_n E(h(Y_n)) + (1-c_n) E(h(X))$ for any measurable function h.

For the distribution functions F_n and F of X_n and X we have for the supremum norm $\|F_n-F\| \le c_n$. Assuming $c_n \to 0$ we always get $\|F_n-F\| \to 0$ and in particular $X_n \xrightarrow{\mathfrak{L}} X$. Now let Z be a random variable and put

(2) $\quad X = \exp(Z), \qquad Y_n = \exp(b_n Z) \qquad$ with $b_n > 0$.

Using the moment generating function M of Z we get for $r \ge 0$

(3) $\quad E(X^r) = M(r), \qquad E(X_n^r) = c_n M(b_n r) + (1-c_n) M(r).$

Provided these moments are finite, we conclude

(4) $\quad X_n \xrightarrow{d_r} X \quad\Leftrightarrow\quad E(X_n^r) \longrightarrow E(X^r) \quad\Leftrightarrow\quad c_n \cdot M(b_n r) \to 0\,.$

Example 1: d_0-convergence does not imply d_+-convergence.
Let Z have a standard normal distribution $N(0,1)$, so that $M(t) = \exp(\frac{1}{2}t^2)$, and take $c_n = \exp(-n)$. For $b_n^2 = 2n^2$ we conclude from (4) that $X_n \xrightarrow{d_r} X$ does *not* hold for any $r>0$. Note, that $E(X_n^r)$ is finite but converges to ∞. □

Example 2: d_r-convergence for any $0 \le r < s$ does not imply d_s-convergence.
Suppose Z has an exponential distribution with expectation μ.Then $M(t) = (1-t\mu)^{-1}$ for $t<\mu^{-1}$. For fixed $0<s<\mu^{-1}$ we choose $b_n = s^{-1}\mu^{-1}(1-c_n) < s^{-1}\mu^{-1}$ and conclude from (4) that $X_n \xrightarrow{d_r} X$ for any $0 \le r < s$, but *not* $X_n \xrightarrow{d_r} X$. Note, that $E(X_n^s)$ is finite and converges to $1+E(X^s)$. □

Example 3: d_∞-convergence does not imply e_+-convergence.
Taking $Z_n = \log X_n$ in example 1, we get $Z_n \xrightarrow{\mathcal{L}} Z$ but *not* $Z_n \xrightarrow{e_r} Z$ for any $r>0$ by Lemma 2(v). Furthermore from (1) with $h(x) = (\log x)^r$ one obtains $E(Z_n^r) \to E(Z^r)$ for any $r>0$. Thus we have $Z_n \xrightarrow{d_\infty} Z$ but *not* $Z_n \xrightarrow{e_+} Z$, although the MGF of Z_n and Z are finite on $\mathbb{R}$. □

Example 4: e_r-convergence for any $0 \le r < s$ does not imply e_s-convergence.
Taking $Z_n = \log X_n$ in example 2, we get by Lemma 2(v) $Z_n \xrightarrow{e_r} Z$ for any $0 \le r < s$, but *not* $Z_n \xrightarrow{e_s} Z$. Note again, that $E(\exp(s Z_n)) = E(X_n^s)$ is finite and converges to $1+E(\exp(s Z)) = 1 + E(X^s)$. □

The first two examples involve only *positive* random variables X_n, X, because absolute moments for these are simple to calculate. However other examples are easily given, like the following with a standard normal limit.

Example 5: Convergence of any order $0 \le r < 2$ to $N(0,1)$ does not imply d_2-convergence. Suppose $\mathcal{L}(X) = N(0,1)$, $\mathcal{L}(Y_n) = N(0,n)$ and let X_n have a mixed distribution as above with $c_n = n^{-1}$. Then $X_n \xrightarrow{d_r} X$ holds for any $0 \le r < 2$, but $X_n \xrightarrow{d_2} X$ does *not* hold, since $E(X_n^2) \longrightarrow 2$ and $Var(X_n^2) \longrightarrow 2$. □

The last example is of particular interest in statistics, since convergence of a random variable X_n, e.g. a scaled parameter estimate with $E(X_n)=0$, to a normal limit $N(0,\sigma^2)$ are often encountered. If we only have convergence in distribution together with $E(X_n) \longrightarrow 0$, but *not* of convergence of order 2, we can only conclude $\liminf_n Var(X_n) \ge \sigma^2$ [e.g. Billingsley (1968), Thm.5.3]. Hence the "asymptotic" variance σ^2 of the normal limit may be considerably smaller than the actual variance of X_n, unless, of course, we have convergence of order 2 (like e.g. in the central limit theorem above).

7. Statistical Applications

To illustrate the concepts and methods above, we now strengthen some results on asymptotic normality used in statistics to d_r-convergence. Further examples of d_r-convergence in connection with bootstrap methods may be found in Bickel & Freedman (1981) and Freedman (1981).

Transformation of Binomial Rates

Let X be binomial(n,p)-distributed, and let $\overline{X} = \frac{1}{n}X$ be the observed rate. In some applications one is interested in the transformed rate $G(\overline{X})$ for a suitable transformation $G:(0,1) \longrightarrow \mathbb{R}$, e.g. the probit- or logit-transformation. Since $G(x)$ is not defined for $x=0$, $x=1$ we use a modified rate $\tilde{X}$ which coincides with $\overline{X}$ except for the values 0,1 where $\tilde{X}$ takes the values $cn^{-\gamma}$, $1-cn^{-\gamma}$ for some constants $0<c<1$, $\gamma \geq 1$, e.g. $c=\frac{1}{2}$, $\gamma=1$ as Berkson (1955) suggested.
The difference beetween $\overline{X}$ and $\tilde{X}$ is negligable as $n \to \infty$, since $\sqrt{n}(\overline{X}-\tilde{X}) \xrightarrow{e_\infty} 0$ by Lemma 2(v). Hence $\tilde{X}$ has the same asymptotic distribution as $\overline{X}$ (see 4. Example 1)

(1) $$\sqrt{n}(\tilde{X}-p) \xrightarrow{e_\infty} N(0,pq) \quad \text{as } n \longrightarrow \infty \quad \text{and } q = 1-p.$$

Furthermore one can show that

(2) $$\tilde{X}^{-1} \xrightarrow{d_\infty} p^{-1}, \qquad (1-\tilde{X})^{-1} \xrightarrow{d_\infty} (1-p)^{-1}.$$

Application of 5. Example 2 gives the following limit result for the transformed rate, provided the transformation G satisfies condition 5(8)

(3) $$\sqrt{n}\left[G(\tilde{X}) - G(p)\right] \xrightarrow{d_\infty} N(0,\sigma^2) \quad \text{as } n \longrightarrow \infty,$$

with $\sigma^2 = pq[G'(p)]^2$. Note, that 5(9) follows from (2).

Log-Odds-Ratio for Poisson Variables

Let X_{11}, X_{12}, X_{21}, X_{22} be independent random variables having Poisson distributions with $\lambda_{ij}:=E(X_{ij}) > 0$. Regarding the X_{ij} as a 2×2-contingency table

X_{11}	X_{12}
X_{21}	X_{22}

one is (among other things) interested in the log-odds-ratio of the observed table

(4) $$U = \log X_{11} + \log X_{22} - \log X_{12} - \log X_{21} .$$

To derive a limit result for U we first consider $\log X$ for a single Poisson random variable X with $E(X) = \lambda$. To avoid infinite values of $\log X$ we replace X by a random variable $\tilde{X}$ which coincides with X except for the value 0 where $\tilde{X}$ takes a fixed value $0<c<1$, e.g. $c=\frac{1}{2}$. For $\lambda\to\infty$ the difference beetween X and $\tilde{X}$ is negligible since $\lambda^{-1/2}(X-\tilde{X}) \xrightarrow{e_\infty} 0$ by Lemma 2(v). Hence $\tilde{X}$ has the same asymptotic distribution as X (see 4. Example 1)

(5) $$\sqrt{\lambda}\,(\tilde{X}\lambda^{-1}-1) \xrightarrow{e_\infty} N(0,1) \quad \text{as} \quad \lambda\to\infty .$$

Application of 5. Example 1 gives

(6) $$\sqrt{\lambda}\left[\log\tilde{X} - \log\lambda\right] \xrightarrow{d_\infty} N(0,1) \quad \text{as} \quad \lambda\to\infty ,$$

since the condition 5(4) follows from

(7) $$\lambda\tilde{X}^{-1} \xrightarrow{d_\infty} 1 .$$

We now return to the table (X_{ij}) and consider a limit process $\lambda_{ij}\to\infty$, such that $\lambda_{ij}/\lambda_{++}$ converges to a constant $c_{ij}>0$, with $\lambda_{++} = \lambda_{11}+\lambda_{12}+\lambda_{21}+\lambda_{22}$. Then the aymptotic distribution of the log-odds-ratio follows easily from (6) applied to each $\tilde{X}_{ij}$:

(8) $$\frac{1}{\sqrt{\gamma}}\left[\log\left(\frac{\tilde{X}_{11}\tilde{X}_{22}}{\tilde{X}_{12}\tilde{X}_{21}}\right) - \log\left(\frac{\lambda_{11}\lambda_{22}}{\lambda_{21}\lambda_{22}}\right)\right] \xrightarrow{d_\infty} N(0,1)$$

with $\gamma = \lambda_{11}^{-1}+\lambda_{12}^{-1}+\lambda_{21}^{-1}+\lambda_{22}^{-1}$.

The Linear Model

Let $Y_1, Y_2, \ldots$ be a sequence of observations satisfying the linear model

(1) $$Y_i = \langle x_i, \beta \rangle + \varepsilon_i \ ,$$

where $x_i \in \mathbb{R}^s$ is a vector of known covariantes, $\beta \in \mathbb{R}^s$ is the unknown vector of parameters, and ε_i is a random error. We assume that the errors $\varepsilon_1, \varepsilon_2, \ldots$ are i.i.d. with

(2) $$E(\varepsilon_1) = 0, \quad 0 < \sigma^2 := \mathrm{Var}(\varepsilon_1) < \infty \ .$$

The vector $Y_{[n]} = (Y_1, \ldots, Y_n)$ of the first n observations may then be written in the usual form of a linear model

(3) $$Y_{[n]} = X_{[n]}\beta + \varepsilon_{[n]}$$

where $X_{[n]}$ is the $(n \times s)$-matrix with rows $x_1^T, \ldots, x_n^T$, and the error $\varepsilon_{[n]} = (\varepsilon_1, \ldots, \varepsilon_n)$ satisfies

(4) $$E(\varepsilon_{[n]}) = 0, \quad \mathrm{Cov}(\varepsilon_{[n]}) = \sigma^2 I_n \ , \qquad (I_n \text{ is the } n \times n \text{ identity matrix}).$$

The least square estimator of Θ based on the first n observations is

(5) $$\hat{\Theta}_{[n]} = (X_{[n]}^T X_{[n]})^{-1} X_{[n]}^T Y_{[n]}$$

with expectation Θ and covariance matrix $\sigma^2 (X_{[n]}^T X_{[n]}^T)^{-1}$. It is well known, e.g. Schach & Schäfer (1978) Sec. 2.6, that the standardized estimator

(6) $$Z_n = \sigma^{-1} (X_{[n]}^T X_{[n]})^{-1/2} [\hat{\Theta}_{[n]} - \Theta]$$

converges to the s-dimensional standard normal distribution $N_s(0, I_s)$, if certain conditions on the covariates hold. Here $A^{1/2}$ denotes the Cholesky-root of a positive definite matrix A, i.e. the unique lower triangular matrix such that $A^{1/2}(A^{1/2})^T = A$.
The convergence of Z_n to its normal limit can be shown to be of order d_r resp. e_r, provided the corresponding moments of the error-distribution exist.

Theorem: Assume the following condition on the covariantes

(i) The sequence $x_1, x_2, \ldots$ is bounded.

(ii) $X_{[n]}$ has full rank s for $n \geq s$.

(iii) The trace of $n(X_{[n]}^T X_{[n]})^{-1}$ is bounded for $n \geq s$.

Then the following convergence results hold as $n \to \infty$:

(iv) $Z_n \xrightarrow{d_2} N_s(0, I_s)$.

(v) If $E(\varepsilon_1^K) < \infty$ for an *even* $K \in \mathbb{N}$, then $Z_n \xrightarrow{d_K} N_s(0, I_s)$.

(iv) If the MGF of ε_1 is finite on some open neighbourhood of 0 resp. on $\mathbb{R}$ then $Z_n \xrightarrow{e_+} N_s(0, I_s)$ resp. $Z_n \xrightarrow{e_\infty} N_s(0, I_s)$.

Remark 1: If $X_{[n]}$ has full rang for some $n \geq s$, then condition (ii) holds after rearranging the first n observations.

Remark 2: Condition (iii) follows from the stronger "standard" assumption

(iii)* $\quad n^{-1} X_{[n]}^T X_{[n]} \longrightarrow V$ positive definite, as $n \to \infty$.

Remark 3: To prove (iv), the condition (i) may be weakened to the following consequence of (iii)*:

(i)* $\quad n^{-1/2} \|x_n\| \longrightarrow 0$ as $n \to \infty$.

Proof (Theorem)

ad (iv): By Cramér-Wold's device it suffices to show for any $t \in \mathbb{R}$, $t \neq 0$

(vii) $$U_n := \langle t, Z_n \rangle \xrightarrow{d_2} N(0, \|t\|^2) .$$

Without loss of generality, we may assume $\|t\| = 1$. From the representation

$$Z_n = \sigma^{-1} C_n^T X_{[n]}^T \varepsilon_{[n]} \quad \text{with} \quad C_n = (X_{[n]}^T X_{[n]})^{-1/2}$$

we get

(viii) $$U_n = \sigma^{-1} \langle a_n, \varepsilon_{[n]} \rangle = \sum_i a_{ni} \sigma^{-1} \varepsilon_i \quad \text{with} \quad a_n = X_{[n]} C_n t, \ \|a_n\|^2 = 1 .$$

We first show for the maximum-norm $\|\cdot\|_{max}$

(ix) $\quad \sqrt{n}\, \|a_n\|_{max}$ is bounded.

This follows from

(x) $$\sqrt{n}\ \|a_n\|_{max} \leq s^2 \cdot \|X_{[n]}\|_{max} \cdot \|\sqrt{n}\, C_n\|_{max} \cdot \|t\|_{max}$$

together with (i) and (iii) in view of

(xi) $\quad \|A\|_{max}^2 \leq \mathrm{trace}(AA^T) \quad$ for any $s \times s$ matrix A.

Now U_n has expectation 0, variance 1, and $U_n \xrightarrow{\mathcal{L}} N(0,1)$ follows from the central limit theorem, since Lindeberg's condition may be derived from $\|a_n\|_{max} \longrightarrow 0$. And (vi) follows, since U_n is standardized.

ad (v)(vi): The proofs are similar to (iii), using Cramér-Wold's device and the corollaries to the central limit theorems for d- and e-convergence above, which apply in view of (viii)(ix). □

Proof (Remark 3)

Condition (i)* implies $n^{-1/2} \|X_{[n]}\|_{max} \longrightarrow 0$ and using (x) we get instead of (ix) only $\|a_n\|_{max} \longrightarrow 0$, which was sufficient to derive (iv) above. □

Estimation of Quantiles and the Median

Sen (1959) has shown that the (suitably scaled) distribution of sample quantiles (e.g. the median) converges to its normal limit of order d_r, for any $r \in \mathbb{N}$, as the sample size tends to infinity, provided of course the underlying sample distribution has certain properties. We are now going to prove this result under slightly different assumptions, using the asymptotic moments of beta distributions given in Sen (1959), Sec. 2. and the above properties of d_r-convergence.

Let $X_1,\dots,X_n$ be i.i.d. random variables with a distribution concentrated on some open intervall $I \subset \mathbb{R}$. We assume that the distribution function F of X_1 has a continuous densitity $f = F' > 0$ on I. Then for any $0 < p < 1$, the p-th quantile of F is defined by $\theta_p := F^{-1}(p) \in I$. Of course $p = \frac{1}{2}$ gives the median of F. The estimate of θ_p should be the corresponding quantile of the empirical distribution function F_n of $X_1,\dots,X_n$ given by $F_n(x) = \#\{i \mid x_i \le n\}/n$. However F_n being a step function has no unique inverse F_n^{-1} and the quantile is usually defined as $\hat{\theta}_p = F_n^{-1}(p)$ with some definite choice of inverse like the *left* resp. *right-continuous inverse* of a distribution function G given by

(1) $\quad G^-(p) = \inf\{x \mid p \le G(x)\}$ (left-continuous inverse) resp.

(2) $\quad G_-(p) = \sup\{x \mid G(x) \le p\}$ (right-continuous inverse).

To be explizit, we take $\hat{\theta}_p = F_n^-(p)$. Results for the other estimate $F_{n-}(p)$ may be derived from those of $F_n^-(p)$ in view of $F_{n-}(p) = -G_n^-(1-p)$, where G_n is the empirical distribution function of the variables $-X_1,\dots,-X_n$. First we derive the limit result for the uniform distribution on (0,1) and generalize to arbitrary continuous distributions using the δ-method.

Theorem 1 (Uniform case): If F is the uniform distribution on I = (0,1) then as $n \to \infty$:

(i) $\quad \sqrt{n}\left[F_n^-(p) - p\right] \xrightarrow{d_\infty} N(0,pq)$ with $q = 1-p$.

(ii) $\quad E\left[F_n^-(p)^{-s}\right] \longrightarrow p^{-s}$, provided $n > s/p$, $s \in \mathbb{N}$.

(iii) $\quad E\left[(1-F_n^-(p))^{-s}\right] \longrightarrow q^{-s}$, provided $n > s/q$, $s \in \mathbb{N}$.

Remark: The conditions on n in (ii)(iii) garantee the existence of the inverse moments.

Proof: $\hat{\theta}_n = F_n^-(p)$ has a Beta(a_n,b_n)-distribution with $a_n+b_n = n+1$, where a_n is the smallest integer $\geq np$ resp. b_n is the greatest integer $\leq 1+nq$. From Sen (1959), (2.11), (with a_n,b_n interchanged) we get for any $s\in\mathbb{N}$

$$E\{Z_n^s\} \longrightarrow E\{N(0,pq)^s\}, \quad \text{where} \quad Z_n = \sqrt{n}\,[\hat{\theta}_n-\mu_n], \quad \mu_n = E\{\hat{\theta}_n\} = a_n/(n+1).$$

This implies $Z_n \xrightarrow{\mathcal{L}} N(0,pq)$ [Billingsley (1979) Thm.30.2], and hence $Z_n \xrightarrow{d_\infty} N(0,pq)$. Adding $\sqrt{n}\,(\mu_n-p) \longrightarrow 0$ and using 2(3), 2(6) gives (i). Now (ii) follows from

$$E\{\hat{\theta}_n^{-s}\} = B(a_n-s,b_n)/B(a_n,b_n) = n^{(s)}/(a_n-1)^{(s)} \ ,$$

where $B(-,-)$ is the beta funtion, $k^{(r)} = k(k-1)...(k-r+1)$ is the descending factorial, and $s<np\leq a_n$. Finally (iii) follows similarly, since $1-\hat{\theta}_n$ has a Beta(b_n,a_n) distribution. □

Theorem 2 (General Case)

Let F have a continuous density $f = F' > 0$ on its support I and assume for some $\alpha\geq 0$

(i) $F(x)^\alpha\,[1-F(x)]^\alpha\,f(x)^{-1}$ is bounded for $x\in I$.

Then for any $r\in\mathbb{N}$ and $n\longrightarrow\infty$ with $n > 2\,\alpha\cdot r\cdot\max(p^{-1},q^{-1})$, $q=1-p$:

(ii) $\sqrt{n}[\hat{\theta}_p-\theta_p] \xrightarrow{d_r} N(0,\sigma^2)$ with

(iii) $\sigma^2 = (pq)/f(\theta_p)^2$.

Remark 1: (i) is a condition on the tails of the distribution and is equivalent to condition 5(8) for $G=F^{-1}$. Sen (1958) uses another condition, namely the existence of the δ-th moment of F for some $\delta>0$, to derive (ii), using slighly different conditions on F.

Remark 2: The condition on n garantees that the r-th moment of the left-hand side in (ii) exists.

Proof: The variables $Y_i = F(X_i)$ are i.i.d. with a uniform distribution on (0,1). Since F has an inverse F^{-1}: $(0,1) \longrightarrow I$ we get $X_i = F^{-1}(Y_i)$. For the empirical distribution function F_n resp. H_n of X_i resp. Y_i we have $F_n^- = F^{-1}\circ H_n^-$. By Thm.1(i) we have $\sqrt{n}\,[H_n^-(p)-p] \xrightarrow{d_\infty} N(0,pq)$, and application of the δ-method (see 5. Example 2 with $G=F^{-1}$) will give (ii)(iii) provided condition 5(9) holds for $t=2r$, i.e. $H_n^-(p)^{-2\alpha r}$ and $[1-H_n^-(p)]^{-2\alpha r}$ are uniformly integrable. And this follows from Thm.1 (ii-iii) for $n>2\,\alpha\cdot r\cdot\max\{p^{-1},q^{-1}\}$. □

References

Berkson, J.(1953): A Statistically Precise and relatively simple Methods of Estimating the Bioassay with Quantal Response, Based on the Logistic Function. Journal of American Statistical Association,44, 565-599.

Bickel, P.J., Freedman, D.A.(1981): Some Asymptotic Theory for the Bootstrap. The Annals of Statistics, 9, 1196-1217.

Billingsley, P.(1968): Convergence of Probability Measures. Wiley, New York.

Billingsley, P.(1979): Probability and Measure. Wiley, New York.

Bishop, Y.M.M., Fienberg, S.E., Holland, P.W.(1975). Discrete Multvariate Analysis: Theory and Practice. Cambridge (Mass.), MIT-Press.

Dieudonné, J.(1960): Foundations of Modern Analysis. Academic Press, New York.

Freedman, D.A.(1981): Bootstrapping Regressions Models. The Annals of Statistics, 9, 1218-1228,.

Johnson, N.L., Kotz, S.(1970-72): Distribution in Statistics. Vol. 1-4, New York, Wiley.

Schach, S., Schäfer, Th. (1978). Regressions- und Varianzanalyse. Berlin, Springer.

Sen, P.K.(1959): On the Moments of the Sample Quantiles. Calcutta Statistical Association Bulletin, 9, 1-19.

Adress of author: Prof. Dr. Gerhard Osius
Universität Bremen, Fachbereich 3 (Mathematik/Informatik)
28334 Bremen, Postfach 33 04 40
Germany
e-mail: osius@math.uni-bremen.de

Date: 21-Mar-1989 (printed edition)
8-Jan-2001 (PDF-File, with minor corrections)

Universität Bremen Mathematik-Arbeitspapiere

Stand November 2000 ISSN 0173 - 685 X

Vertrieb der Hefte 4, 14, 23, 26 durch Universitätsbuchhandlung, Bibliothekstr. 3, D-28359 Bremen. Vertrieb der übrigen Hefte (soweit nicht vergriffen) durch die Autoren oder FB 3 Mathematik/Informatik Universität Bremen, Postfach 330440, D-28334 Bremen.

1. Ulrich Krause (1976): Strukturen in unendlichdimensionalen konvexen Mengen, 74 S.
2. Fritz Colonius, Diederich Hinrichsen (1976): Optimal control of heriditary differential systems. Part I, 66 S.
3. Günter Matthiessen (1976): Theorie der heterogenen Algebren, 88 S.
4. H. Wolfgang Fischer, Jens Gamst, Klaus Horneffer (1976): Skript zur Analysis, Band 1 (11. Auflage 2000), 286 S.
5. Wolfgang Schröder (1977): Operator-algebraische Ergodentheorie für Quantensysteme, 59 S.
6. Rolf Röhrig, Michael Unterstein (1977): Analyse multivariabler Systeme mit Hilfe komplexer Matrixfunktionen, 216 S.
7. Horst Herrlich, Hans-Eberhard Porst, Rudolf-Eberhard Hoffmann, Manfred Bernd Wischnewsky (1976): Nordwestdeutsches Kategorienseminar, 193 S.
8. Fritz Colonius, Diederich Hinrichsen (1977): Optimal Control of Hereditary Differential Systems. Part II: Differential State Space Description, 36 S.
9. Ludwig Arnold (1977): Differentialgleichungen und Regelungstheorie, 185 S.
10. Rudolf Lorenz (1977): Iterative Verfahren zur Lösung großer, dünnbesetzter symmetrischer Eigenwertprobleme, 104 S.
11. Konrad Behnen, Hans-Peter Kinder, Gerhard Osius, Rüdiger Schäfer, Jürgen Timm (1977): Dose-Response-Analysis, 206 S.
12. Hans-Friedrich Münzner, Dieter Prätzel-Wolters (1978): Minimalbasen polynomialer Moduln, Strukturindizes und BRUNOVSKY-Transformationen, 53 S.
13. Konrad Behnen (1978): Vorzeichen-Rangtests mit Nullen und Bindungen, 53 S.
14. H. Wolfgang Fischer, Jens Gamst, Klaus Horneffer, Eberhard Oeljeklaus (1978): Skript zur Linearen Algebra, Band 1 (13. Auflage 2000), 249 S.
15. Günter Ludyk (1978): Abtastregelung zeitvarianter Einfach- und Mehrfachsysteme, 54 S.
16. Momme Johs Thomsen (1977): Zur Theorie der Fastalgebren, 146 S.
17. Klaus Horneffer, Horst Diehl (1978): Modellrechnungen zur anaeroben Reduktionskinetik des Cytochroms P-450, 34 S.
18. Horst Herrlich, Rudolf-Eberhard Hoffmann, Hans-Eberhard Porst, Manfred Bernd Wischnewsky (1979): Structure of Topological Categories, 252 S.
19. Hans-Friedrich Münzner, Dieter Prätzel-Wolters (1979): Geometric and moduletheoretic approach to linear systems. Part I: Basic categories and functors, 28 S.
20. Hans-Friedrich Münzner, Dieter Prätzel-Wolters (1979): Geometric and moduletheoretic approach to linear systems. Part II: Moduletheoretic characterization and reachability, 28 S.
21. Eckart Beutler, Hans Kaiser, Günter Matthiessen, Jürgen Timm (1979): Biduale Algebren, 165 S.
22. Horst Diehl, Detlef Harbach, Jürgen Timm (1980): Planung und Auswertung von Atomabsorptions-Spektrometrie-Untersuchungen mit der Additionsmethode, 44 S.
23. H. Wolfgang Fischer, Jens Gamst, Klaus Horneffer (1981): Skript zur Analysis, Band 2 (6. Auflage 1996), 299 S.
24. Horst Herrlich (1981): Categorical Topology 1971-1981, 105 S.
25. Horst Herrlich, Rudolf-Eberhard Hoffmann, Hans-Eberhard Porst, Manfred Bernd Wischnewsky (1981): Special Topics in Topology and Category Theory, 108 S.

26. H. Wolfgang Fischer, Jens Gamst, Klaus Horneffer (1984): Skript zur Linearen Algebra, Band 2 (7. Auflage 1999), 257 S.
27. Rudolf-Eberhard Hoffmann (1982): Continuous Lattices and Related Topics, 314 S.
28. Horst Herrlich, Rudolf-Eberhard Hoffmann, Hans-Eberhard Porst (1987): Workshop on Category Theory, 169 S.
29. Harald Boehme (1987): Zur Berufspraxis des Diplommathematikers, 16 S.
30. Jürgen Timm (1986): Mathematische Modelle der Dosis-Wirkungsanalyse bei den experimentellen Untersuchungen der Arbeitsgruppe zur karzinogenen Belastung des Menschen durch Luftverunreinigung, 65 S.
31. Dieter Denneberg (1988): Mathematik für Wirtschaftswissenschaftler. I. Lineare Algebra, 97 S.
32. Peter E. Crouch, Diederich Hinrichsen, Anthony J. Pritchard, Dietmar Salamon (1988, previous edition University of Warwick 1981): Introduction to Mathematical Systems Theory, 244 S.
33. Gerhard Osius (1989): Some Results on Convergence of Moments and Convergence in Distribution with Applications in Statistics, 27 S.
34. Dieter Denneberg (1989): Verzerrte Wahrscheinlichkeiten in der Versicherungsmathematik, Quantilsabhängige Prämienprinzipien, 24 S.
35. Eberhard Oeljeklaus (1989): Birational splitting of homogeneous Albanese bundles, 30 S.
36. Gerhard Osius, Dieter Rojek (1989): Normal Goodness-of-Fit Tests for Parametric Multinomial Models with Large Degrees of Freedom, 38 S.
37. Dieter Denneberg (1990): Mathematik zur Wirtschaftswissenschaft. II. Analysis, 59 S.
38. Ulrich Krause, Cornelia Zahlten (1990): Arithmetik in Krull monoids and the cross number of divisor class groups, 29 S.
39. Dieter Denneberg (1990): Subadditive Measure and Integral, 39 S.
40. Ulrich Krause, Peter Ranft (1991): A limit set trichotomy for monotone nonlinear dynamical systems, 31 S.
41. Angelika van der Linde (1992): Statistical analyses with splines: are they well defined? 22 S.
42. Dieter Denneberg (1992): Lectures on non-additive measure and integral (new edition: Non-additive measure and integral. TDLB 27, Kluwer Academic, Dordrecht (1994)), 114 S.
43. Gerhard Osius (1993): Separating Agreement from Association in Log-linear Models for Square Contingency Tables With Applications, 23 S.
44. Hans-Peter Kinder, Friedrich Liese (1995): Bremen-Rostock Statistik Seminar, 5. - 7. März 1992, 110 S.
45. Dieter Denneberg (1995): Extension of a measurable space and linear representation of the Choquet Integral, 30 S.
46. Dieter Denneberg, Michael Grabisch (1996): Shapley value and interaction index, 20 S.
47. Angelika Bunse-Gerstner, Heike Faßbender (1996): A Jacobi-like method for solving algebraic Riccati equations on parallel computers, 24 S.
48. Hans-Eberhard Porst editor (1997): Categorical methods in algebra and topology - a collection of papers in honour of Horst Herrlich, 498 S.
49. Angelika van der Linde, Gerhard Osius (1997): Estimation of nonparametric risk functions In matched case-control studies, 28 S.
50. Angelika van der Linde (1997): Estimating the smoothing parameter in generalized spline-based regression, 46 S.
51. Ursula Müller, Gerhard Osius (1998): Asymptotic normality of goodness-of-fit statistics for sparse Poisson data, 15 S.
52. Ursula Müller (1999): Nonparametric regression for threshold data, 18 S.
53. Gerhard Osius (2000): The association between two random elements – A complete characterization in terms of odds ratios, 32 S.