Sudakov's typical marginals, random linear functionals and a conditional central limit theorem

Heinrich v. Weizsäcker

February 6, 1997

Summary. V.N. Sudakov [Sud78] proved that the one-dimensional marginals of a high-dimensional second order measure are close to each other in most directions. Extending this and a related result in the context of projection pursuit of P. Diaconis and D. Freedman [Dia84], we give for a probability measure P and a random (a.s.) linear functional F on a Hilbert space simple sufficient conditions under which most of the one-dimensional images of P under F are close to their canonical mixture which turns out to be almost a mixed normal distribution. Using the concept of approximate conditioning we deduce a conditional central limit theorem (theorem 3) for random averages of triangular arrays of random variables which satisfy only fairly weak asymptotic orthogonality conditions.

AMS-Classification: 60B11, 60F05, 28C20, 60G12

1 Notations and basic assumptions

If X is a topological space then $Prob\ (X)$ denotes the set of tight probability measures on the Borel σ-algebra $\mathcal{B}(X)$ of X. The set $Prob\ (X)$ will always be equipped with the topology of convergence in law, i.e. the coarsest topology for which $P \longmapsto P(U)$ is lower semicontinuous for all open set $U \subset X$. Let $\mathcal{H}$ be a real Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$. For $P \in Prob\ (\mathcal{H})$ the symbol $\check{P}$ denotes the measure obtained by reflection at the origin: $\check{P}(B) = P(-B)$. With this we form the symmetrized convolution $P * \check{P} = \mathcal{L}(X - Z)$ where X, Z are independent with distribution P.

We are interested in random marginals of P, i.e. its image under random linear functionals. For the infinite dimensional case we also include a.s. linear functionals: Let $(\Omega, \mathcal{A}, \nu)$ be a probability space and assume that there is a $\mathcal{B}(\mathcal{H}) \otimes \mathcal{A}$-measurable map $F : \mathcal{H} \times \Omega \to I\!R$ such that for all $x, y \in \mathcal{H}$

$$(1) \qquad F(x + y,\ \omega) = F(x, \omega) + F(y, \omega) \qquad \nu -\ a.s.\ .$$

It is essential to allow the exceptional sets in (1) since a Borel measurable map $F(\cdot, \omega)$ on $\mathcal{H}$ which is strictly additive is known to be continuous. If $\Omega = \mathcal{H} = I\!R^n$ we assume that $F(x, \omega) = \langle x, \omega\rangle$. We say that the pair (ν, F) induces the canonical Gaussian cylindrical measure ([GV61]) on $\mathcal{H}$ if the family $(F(x, \cdot\ ))_{x\in\mathcal{H}}$ is a centered Gaussian process with $cov(F(x, \cdot\ ), F(y, \cdot\ )) = \langle x, y\rangle$ for all $x, y \in \mathcal{H}$. For $\omega \in \Omega$ we are interested in the 'marginal' measure

$$(2) \qquad P^\omega = \mathcal{L}_P(F(\cdot, \omega)) \in Prob(I\!R),$$

and in the joint distribution of $\|\cdot\|^2$ and $F(\cdot,\omega)$:

$$P_2^\omega = \mathcal{L}_P(\|\cdot\|^2, F(\cdot,\omega)) \in Prob(I\!R_+ \times I\!R). \tag{3}$$

Note that the map $\omega \longmapsto P^\omega \in Prob(I\!R)$ is $\mathcal{A}-\mathcal{B}(Prob(I\!R))$-measurable since for $\varphi \in C_b(I\!R)$ the map $\omega \longmapsto \int \varphi(u)P^\omega(du) = \int \varphi(F(x,\omega))P(dx)$ is $\mathcal{A}$-measurable by Fubini. Similarly $\omega \longmapsto P_2^\omega$ is $\mathcal{A} - \mathcal{B}(Prob(I\!R_+ \times I\!R))-$measurable.
The n-dimensional identity operator is denoted by I_n.
For every $P \in Prob(\mathcal{H})$ with $\int \| x \|^2 P(dx) < \infty$ let $\mathcal{C}(P)$ be its covariance operator, i.e. the trace class operator given by $\langle \mathcal{C}(P)x, y\rangle = \int_{\mathcal{H}} \langle x, y\rangle\ P(dx)$.
Conversely, given the operator $\mathcal{C}$, the symbol $\rho(\mathcal{C})$ denotes its spectral radius and $\mathcal{N}(\mathcal{C})$ denotes the centered Gaussian measure on $\mathcal{H}$ with corresponding operator $\mathcal{C}(\mathcal{N}(\mathcal{C})) = \mathcal{C}$.
If p is a probability measure on $I\!R_+$ then $p \times \mathcal{N}$ denotes the probability measure on $I\!R_+ \times I\!R$ given by

$$\int_{I\!R_+ \times I\!R} f(u,v)\ (p \times \mathcal{N})(dudv) = \int_{I\!R_+} [\int_{I\!R} f(u,v)\mathcal{N}(u)(dv)]\ p(du).$$

2 The symmetric, unconditioned second order case

The following is a dimension free version of Sudakov's theorem and of the corresponding result of Diaconis and Freedman. It shows that the fluctuation of the one-dimensional marginals of a second order measure P essentially depends only on the spectral radius $\rho\ \mathcal{C}(P)$ of its covariance operator. In this paper we give only applications to asymptotics of finite dimensional situations. But the proof in the infinite dimensional version requires almost no extra work. Of course, the result can also be applied if (Ω, ν) is an abstract Wiener space ([Gro65]) with Cameron-Martin subspace $\mathcal{H}$ and the map which sends x to the ν-equivalence class of $F(x,\cdot)$ is the canonical embedding of $\mathcal{H}$ into $\mathbf{L}^2(\nu)$. In the next section we deduce Theorem 1 from a more general result.

Theorem 1 *Let $\mathcal{H}$ be a real Hilbert space and assume that the pair (ν, F) induces the canonical Gaussian cylindrical measure on $\mathcal{H}$. Let $P \in Prob\ (\mathcal{H})$ and suppose*

$$\int \| x \|^2\ P(dx) \le C < \infty. \tag{4}$$

Let $\overline{P} \in Prob\ (I\!R)$ be the mixture

$$\overline{P} = \int \mathcal{N}(\| x \|^2)\ P(dx). \tag{5}$$

Then for every $\varepsilon > 0$ and every metric κ which induces convergence in law on $Prob\ (I\!R)$ we have

$$\nu\{\ \omega \in \Omega :\ \kappa(P^\omega,\ \overline{P}) > \varepsilon\ \} \le f\,(\rho(\mathcal{C}(P))) \tag{6}$$

where the function f satisfies $f(t) \underset{t\to 0}{\longrightarrow} 0$ and f does not depend on $\mathcal{H}$, ν, P but only on ε, C and κ.

Remark 1 The proof (cf. the end of the next section) will give a particular metric κ^* for which the function f satisfies $f(t) = O\left((Ct)^{\frac{1}{3}}/\varepsilon^2\right)$.

In order to derive from theorem 1 the results of [Sud78] and of [Dia84] let P' be a measure on $I\!R^n$ for large n. The operator $\mathcal{C}(P')$ corresponds to the nonnegative definite matrix C' with entries

$$(\mathcal{C}')_{ij} = \int x_i x_j \; P'(dx). \tag{7}$$

Let $\mathcal{C}'$ have the eigenvalues $\sigma_1^2, \ldots, \sigma_n^2$. Sudakov assumes that $I_n - \mathcal{C}'$ is nonnegative definite which is equivalent to

$$\sigma_i^2 \leq 1 \text{ for all } \; i. \tag{8}$$

Diaconis and Freedman take P' to be the uniform distribution on a finite set such that

$$\sum_{i=1}^{n} (\mathcal{C}')_{ii} = O(n), \tag{9}$$

$$\sum_{i,j=1}^{n} ((\mathcal{C}')_{ij})^2 = o(n^2) \tag{10}$$

and

$$P'\{x \in I\!R^n : \; | \frac{\| x \|}{\sqrt{n}} - 1 \; | > \varepsilon\} = o(1) \tag{11}$$

and both papers consider the marginals $(P')^\omega$ where ω has the uniform distribution Q_n on the unit sphere $\mathbf{S}^{n-1}$, noting that Q_n can be replaced by $\mathcal{N}(\frac{1}{n} I_n)$ since $\mathcal{N}(\frac{1}{n} I_n)$ is rotationally invariant and most points with respect to $\mathcal{N}(\frac{1}{n} I_n)$ are close to the unit sphere.
Now we squeeze this measure: Let $P(B) = P'(\sqrt{n} B)$. Then the $\mathcal{N}(\frac{1}{n} I_n)$-typical behaviour of the marginals of P' and the ν-typical behaviour of the marginals of P coincide for $\nu = \mathcal{N}(I_n)$. Moreover (9) implies (4) with the constant C of the $O(n)$-condition. The left hand side in (10) equals the trace of $\mathcal{C}'^2$. If (10) holds then $\mathrm{tr}\mathcal{C}(P)^2$ converges to zero which is only possible if $\rho(\mathcal{C}(P)) = o(1)$. Thus theorem 1 implies

Corollary 1 *If for $P' \in$ Prob $(I\!R^n)$ conditions (9) and (10) hold as the dimension becomes large then for the uniform distribution Q_n on $\mathbf{S}^{n-1}$ one has*

$$Q_n\{ \; \omega \in \mathbf{S}^{n-1} : \; \kappa \left((P')^\omega, \int \mathcal{N}(\frac{\| x \|^2}{n}) \; P'(dx) \right) > \varepsilon \; \} = o(1). \tag{12}$$

Clearly Sudakov's assumption (8) implies (9) and (10). Sudakov's conclusion was as in the corollary, except that he proved it only for a special metric.
As for the result of Diaconis and Freedman, in the presence of their condition (11), of course the mixed normal in (12) can be replaced by $\mathcal{N}(1)$. Their paper moreover gives sufficient conditions for the empirical measures of large samples chosen from product laws to satisfy (9),(10) and (11). However, in mixed models with nontrivial correlations between the coordinates, one cannot expect (11) to hold. Our proof of theorems 1 and 2 uses an idea similar to the approach of [Dia84].

Here are some instructive elementary examples.

Example a Let P and P' be the uniform distribution on $\mathbf{S}^{n-1}$ resp. $\sqrt{n}\mathbf{S}^{n-1}$. Then $\mathcal{C}(P) = \frac{1}{n} I_n$ and $\mathcal{C}(P') = I_n$. All marginals $P^\omega, \omega \in \sqrt{n}\mathbf{S}^{n-1}$ and $P'^\omega, \omega \in \mathbf{S}^{n-1}$ are equal and for large n close to $\mathcal{N}(1)$. This is a well known geometric fact.
Example b Let P be the Dirac measure ϵ_e in a unit vector e. The marginal P^ω is the Dirac measure in $\langle e, \omega\rangle$. Then $\mathsf{E}_\nu(P^\omega) = \mathcal{N}(1)$ but the individual marginals are far away from this average. Clearly the spectral radius of $\mathcal{C}(P)$ does not depend on the dimension of the surrounding space.
Example c Let $P' = \frac{1}{n}\sum_{i=1}^n \epsilon_{\sqrt{n}e_i}$ for the standard base $\{ e_1, \cdots, e_n \}$ of $I\!R^n$. Then $\mathcal{C}(P') = I_n$ and hence (9) and (10) are satisfied. For every $\omega = (\omega_1, \cdots, \omega_n) \in I\!R^n$ one has $P'^\omega = \frac{1}{n}\sum \epsilon_{\sqrt{n}\omega_i}$. This gives an interesting necessary condition for those sequences of measures $Q_n \in Prob\ (I\!R^n)$ for which (9) and (10) imply (12): In fact $\frac{1}{n}\sum \epsilon_{\sqrt{n}\omega_i} \approx \mathcal{N}(1)$ asymptotically for Q_n- most ω with respect to Q_n for every such sequence (Q_n).

Note also that the measure $\overline{P}$ is centered whereas this is not assumed for P. One can check directly that the barycenter of P is within a distance $O\left(\rho(\mathcal{C}(P))\right)$ to the origin.

Remark 2 It is not difficult but notationally more involved to extend theorem 1 to k-dimensional marginals (k fixed). The typical k-dimensional marginal is given by the mixed normal

$$\overline{P}^{(k)} = \int_{\mathcal{H}} \ldots \int_{\mathcal{H}} \mathcal{N}((\langle x_i, x_j\rangle)_{1\le i,j\le k})\ P(dx_1)\ldots P(dx_k). \tag{13}$$

In this case one works with the canonical cylindrical measure on the product $\mathcal{H}^k$ or in finite dimensions with the uniform distribution on the Grassmannian manifold of k-dimensional subspaces of $I\!R^n$.

3 General cylindrical measures and approximate conditioning

We want to extend theorem 1 in three ways.

Firstly, instead of proving that the marginals P^ω are close to the mixture $\overline{P}$ we want to show that conditioned on $\| x \|^2 \approx \sigma^2$, the measure P^ω is close to $\mathcal{N}(\sigma^2)$ for ν-most ω. However the following example shows that we cannot work with the usual conditional law $P(\ \cdot \mid\ \| x \|^2)$:

Example d Let $\mathcal{H} = I\!R^n$ and let h be a measurable bijection of S^{n-1} onto the interval $[1, 1+\frac{1}{n}]$. Let P be the image of the uniform distribution Q_n on S^{n-1} under the map $x \mapsto h(x)x$. Then P differs only little from the uniform distribution on S^{n-1} so $\rho(\mathcal{C}(P)) = o(1)$ as n increases. But the vector x with law P can be reconstructed P-almost surely from the value $\| x \|^2$ and hence the conditional distribution of $\langle x, \omega\rangle$ given $\| x \|^2 = \sigma^2$ is the point mass in $\langle x, \omega\rangle$ which is not close to $\mathcal{N}(\sigma^2)$.

As a remedy of this kind of problem, one may speak in general of *'approximate conditional convergence of $\mathcal{L}(Z_n|T_n \approx \vartheta)$ to Q_ϑ'* if the joint law $\mathcal{L}(T_n, Z_n)$ converges to the joint law $\mathcal{L}(T, Z)$

of two random variables with $\mathcal{L}(Z|T = \vartheta) = Q_\vartheta$. Following this idea one has to study the convergence of the joint laws of $\|\cdot\|^2$ and $\langle\cdot,\omega\rangle$ (resp. $F(\cdot,\omega)$) under P, i.e. the measures P_2^ω introduced in (3).

Remark 3 If the pair (v, F) induces the canonical Gaussian cylindrical measure of $\mathcal{H}$ then for every $P \in Prob(\mathcal{H})$

$$\mathsf{E}_\nu P_2^\omega = \mathcal{L}_P(\|\cdot\|^2) \times \mathcal{N} \tag{14}$$

and (for the definition of $\overline{P}$ see (5))

$$\mathsf{E}_\nu P^\omega = \overline{P}. \tag{15}$$

Proof: For every $\varphi \in C_b(I\!R_+ \times I\!R)$ and every $P \in Prob\,(\mathcal{H})$ we have

$$\mathsf{E}_\nu \left(\int \varphi(u,v)\, P_2^\omega(du\,dv) \right) = \int_{\mathcal{H}} \mathsf{E}_\nu\{\varphi(\|\, x\,\|^2, F(x,\omega))\}\; P(dx)$$

$$= \int_{\mathcal{H}} \int_{I\!R} \varphi(\|\, x\,\|^2, v)\; \mathcal{N}(\|\, x\,\|^2)(dv)\; P(dx) = \int_{I\!R_+ \times I\!R} \varphi(u,v)\; \mathcal{L}_P(\|\cdot\|^2) \times \mathcal{N})(dudv)$$

and (15) follows from this or directly by a similar calculation. ■

The point of the second generalization is to relax the assumption on the joint law of the $F(x,\cdot)$. One can choose this joint law according to the properties of P. One needs (14) and (15) only asymptotically and not for all measures on $\mathcal{H}$ but only for two particular measures: for the measure P whose marginals are under consideration, and for the associated measure $P * \check{P}$. Technically this measure $P * \check{P}$ appears because its corresponding expectations in (14) and (15) are related to the 'variance' of the P^ω. In order to show that (15) alone cannot suffice for (6) in the case of arbitrary Gaussian ν consider the following example:

Example e Let $P' = \frac{1}{n}\sum_{i=1}^n \epsilon_{\sqrt{n}e_i}$ be the measure from example c above and consider $P = \frac{1}{n}\sum_{i=1}^n \epsilon_{e_i}$. Let ν be the image of $\mathcal{N}(1)$ under the diagonal map $t \mapsto (t,\cdots,t)$. Then for the ν-typical point $\omega = (t,\cdots,t)$ the marginal P^ω is the Dirac measure in the point t which is far from $\overline{P} = \mathcal{N}(1)$. But $\mathsf{E}_\nu\{P^\omega\} = \mathcal{N}(1)$ i.e. (15) holds for P. On the other hand $P * \check{P}\{\sum_{i=1}^n x_i = 0\} = 1$ and hence $(P * \check{P})^\omega$ is the Dirac measure in the point 0 for ν-a.a. ω, hence (15) fails for $P * \check{P}$ and this measure ν. By the way in this example $\mathsf{E}_\nu\{(P * P)^\omega\} = \mathcal{N}(2)$.

The third extension is that we do no longer assume the existence of second moments. We rather work with the following consequences of the second order assumptions: Condition (4) implies for every $\delta > 0$ by Chebyshev's inequality

$$P\{\ \|\, x\,\| \geq \frac{1}{\sqrt{C\delta}}\ \} \leq \delta \tag{16}$$

and

$$P \otimes P\ \{\ |\langle x,y\rangle| \geq \delta\ \} \leq \frac{C \cdot\ \rho(\mathcal{C}(P))}{\delta^2} \tag{17}$$

where C is the constant in (4). For the proof of (17) choose an ON-basis (e_i) of eigenvectors of $\mathcal{C}(P)$ and note that independent P-distributed vectors X, Y satisfy

$$\mathsf{E}\ \{(\langle X,Y\rangle)^2\} = \mathsf{E}\{\sum_{i=1}^\infty X_i^2 Y_i^2\} \leq \mathsf{E}\{\sum_{i=1}^\infty X_i^2\}(\sup_i \sigma_i^2)\ \leq C \cdot \rho(\mathcal{C}(P))$$

where $X_i = \langle X, e_i\rangle e_i$ and $Y_i = \langle Y, e_i\rangle e_i$. The announced result now reads

Theorem 2 *Let d, d' be two metrics for convergence in law on $Prob(I\!R_+ \times I\!R)$. Let $g, h : (0,\infty) \longrightarrow (0,\infty)$ be two monotone functions such that $g(t) \to \infty$ and $h(t) \to 0$ as $t \to 0$. Then for all $\varepsilon, \eta > 0$ there is some $\delta > 0$ such that the following holds:*
Let $\mathcal{H}$ be a real Hilbert space and let $(\Omega, \mathcal{A}, \nu)$ be a probability space with a jointly measurable map $F : \mathcal{H} \times \Omega \to I\!R$ which is ν-a.s. linear in the first component (cf. (1)). Suppose that $P \in Prob(\mathcal{H})$ satisfies the four conditions

$$P\{\ x \in \mathcal{H} : \|\ x\ \| > g(\delta)\} \le \delta \tag{18}$$

$$P \otimes P\{(x,y) \in \mathcal{H} \times \mathcal{H} :\ |\langle x, y\rangle| > h(\delta)\ \} \le \delta \tag{19}$$

$$d'\left(E_\nu P_2^\omega, \mathcal{L}_P(\|\ \|^2) \times \mathcal{N}\right) \le \delta \tag{20}$$

$$d'\left(E_\nu (P * \check{P})_2^\omega,\ \mathcal{L}_{P*\check{P}}(\|\ \|^2) \times \mathcal{N}\right) \le \delta. \tag{21}$$

Then

$$\nu\{\omega : d\left(P_2^\omega,\ \mathcal{L}_P(\|\ \|^2) \times \mathcal{N}\right) > \varepsilon\} < \eta\,. \tag{22}$$

Corollary 2 *Let the metrics κ, κ' inducing convergence in law in $Prob\ (I\!R)$ and the functions g, h be given. For all $\varepsilon, \eta > 0$ there is some $\delta > 0$ such that*

$$\nu\{\omega : \kappa\left(P^\omega, \overline{P}\right) > \varepsilon\} < \eta \tag{23}$$

holds, whenever (18), (19) and the following conditions are valid:

$$\kappa'(E_\nu P^\omega, \overline{P}) < \delta \text{ and } \kappa'(E_\nu (P * \check{P})^\omega, \overline{P * \check{P}}) < \delta. \tag{24}$$

Let us now turn to the proofs. In the proof of Theorem 2 we first show by a tightness argument that the special choice of the metrics does not matter. We then give in Lemma 1 an example of a special metric for which (18) is not needed and for which there is an explicit estimate for the choice of δ.

Proof of Theorem 2: Let d, d' be arbitrary metrics for convergence in law on $Prob(I\!R_+ \times I\!R)$ and suppose that there are special metrics d^*, d'^* for which the assertion holds. Assume that the theorem does not hold for d, d'. Then there are some $\varepsilon, \eta > 0$, two functions g and h and a sequence $(\mathcal{H}_n, P_n, \Omega_n, \nu_n, B_n)$ such that for each n (18), (19), (20) and (21) hold with $\delta_n = \frac{1}{n}$ but (22) fails. Because of (18) the sequence $(\mathcal{L}_{P_n}(\|\ \|_n^2))$ is tight. This implies that the sequences $(\mathcal{L}_{P_n}(\|\ \|_n^2) \times \mathcal{N})$ and (because of (20)) $(E_{\nu_n}(P_n)_2^\omega)$ are also tight. According to Lemma 2 below the sequence of laws $(\mathcal{L}_{\nu_n}(P_n)_2^\omega)$ in $Prob\ (Prob(I\!R_+ \times I\!R))$ is tight. Thus there is a compact subset K of $Prob(I\!R_+ \times I\!R)$ such that for all n one has $\mathcal{L}_{P_n}(\|\ \|^2) \times \mathcal{N} \in K$ and

$$\nu_n\{\omega : (P_n)_2^\omega \notin K\} < \frac{\eta}{2}.$$

On this set K all metrics d and d' which induce the convergence in law are equivalent which contradicts the fact that the theorem holds for d^*, d'^*.
Thus it suffices to prove the theorem for some choice of d^*, d'^*. This is done in the following Lemma. ■

Lemma 1 *Define a metric d^* on $Prob(\mathbb{R}_+ \times \mathbb{R})$ by*

$$d^*(p,q) = \| \mathcal{F}p - \mathcal{F}q \|_{2,\lambda} \tag{25}$$

where $\mathcal{F}$ denotes the Laplace-Fourier transform $\mathcal{F}p(s,t) = \int_{\mathbb{R}_+\times\mathbb{R}} \exp[-su + i\ t\ v]\, p(du\, dv)$ and where λ is a probability measure with a positive Lebesgue density on $\mathbb{R}_+ \times \mathbb{R}$ such that

$$\int_{\mathbb{R}_+\times\mathbb{R}} 4s + 2t^2\, \lambda(dsdt) \leq 1. \tag{26}$$

Then for $0 < \delta < 1$ the conditions (19), (20) and (21) for d^* in the place of d' imply the estimate

$$E_\nu\{[d^*(P_2^\omega, \mathcal{L}_P(\|\cdot\|^2) \times \mathcal{N})]^2\} \leq h(\delta) + 7\delta. \tag{27}$$

In particular the assertion of the theorem holds in the case $d = d' = d^*$. In this case for every δ with

$$h(\delta) + 7\delta \leq \varepsilon^2 \eta \tag{28}$$

the conditions (19) , (20) and (21) imply (22).

Proof: Suppose that (19), (20) and (21) are satisfied for d^* instead of d'. Fix $s, t \in \mathbb{R}_+ \times \mathbb{R}$. The idea of the proof is to estimate the variance of the complex valued random variable $\mathcal{F}P_2^\omega(s,t)$ with respect to the probability measure $\nu \otimes \lambda$ on $\Omega \times \mathbb{R}_+ \times \mathbb{R}$. We use several times the fact that for every $\tau > 0$ and $b \in \mathbb{R}$ the map $\mathbb{R}_+ \ni a \longmapsto \exp[-\tau a + ib]$ has the Lipschitz constant τ. We get for all $s \in \mathbb{R}_+,\ t \in \mathbb{R}$ and ν-a.a. $\omega \in \Omega$ by (1), Fubini's theorem and (19)

$$\begin{aligned}
& |\mathcal{F}P_2^\omega(s,t)|^2 \\
= \ & \Big(\int \exp[-s \| x \|^2 + i\ t\ F(x,\omega)]\ P(dx)\Big)\overline{\Big(\int \exp[-s \| y \|^2 + itF(y,w)]\ P(dy)\Big)} \\
= \ & \int\int \exp[-s(\| x \|^2 + \| y \|^2) + i\ t\ F(x-y,\omega)]\ P(dx)P(dy) \\
\leq \ & \int\int \exp[-s(\| x-y \|^2) + i\ t\ F(x-y,\omega)]\ P(dx)P(dy) + 2h(\delta)s + P\otimes P\{|\langle x,y\rangle| > h(\delta)\} \\
\leq \ & \mathcal{F}(P * \check{P})^\omega(s,t) + 2s\ h(\delta) + \delta
\end{aligned}$$

By definition of d^* and (20), (21) and (26) we conclude

$$\begin{aligned}
& \int E_\nu(|\mathcal{F}P_2^\omega|^2)\ d\lambda \\
\leq \ & \int E_\nu \mathcal{F}((P * \check{P})^\omega)\ d\lambda + \int 2s\ h(\delta)\ d\lambda + \delta \\
\leq \ & \int \mathcal{F}(\mathcal{L}_{P*\check{P}}(\| \,.\, \|^2) \times \mathcal{N})\ d\lambda + h(\delta)/2 + 2\delta
\end{aligned}$$

$$\begin{aligned}
&= \int\int\int \exp(-su + i\ t\ v)\mathcal{N}(u)(dv)\mathcal{L}_{P*\check{P}}(\|\ \|^2)(du)d\lambda + h(\delta)/2 + 2\delta \\
&= \int\int\int \exp(-s - \frac{t^2}{2}) \parallel x - y \parallel^2 P(dx)P(dy)d\lambda + h(\delta)/2 + 2\delta \\
&\leq \int\int\int \exp(-s - \frac{t^2}{2})(\parallel x \parallel^2 + \parallel y \parallel^2)\ dP \otimes P d\lambda + \int 2(s + \frac{t^2}{2})\ h(\delta)\ d\lambda + h(\delta)/2 + 2\delta \\
&\leq \int |\mathcal{F}(\mathcal{L}_P(\|\ \|^2) \times \mathcal{N})|^2\ d\lambda + h(\delta) + 2\delta \\
&\leq \int |E_\nu \mathcal{F}(P_2^\omega)|^2\ d\lambda + h(\delta) + 4\delta,
\end{aligned}$$

since by the Cauchy-Schwarz inequality we have for all p and q the estimate

$$\int |\mathcal{F}p|^2 d\lambda - \int |\mathcal{F}q|^2 d\lambda = \langle \mathcal{F}p + \mathcal{F}q,\ \mathcal{F}p - \mathcal{F}q\rangle_\lambda \leq 2d^*(p,q).$$

Hence using (20) once more we get (27):

$$\begin{aligned}
E_\nu\{\ d^*(P_2^\omega, \mathcal{L}_P(\|\ \|^2) \times \mathcal{N})^2\ \} &\leq E_\nu\{\ (d^*(P_2^\omega, E_\nu(P_2^\omega)))^2\ \} + 2\delta + \delta^2 \\
&\leq \int E_\nu |\mathcal{F}P_2^\omega - E_\nu(\mathcal{F}P_2^\omega)|^2\ d\lambda + 3\delta \leq h(\delta) + 7\delta.
\end{aligned}$$

The last statement follows from Chebyshev's inequality. ∎

Lemma 2 *Let X be a topological space and let $Prob(X)$ be equipped with the topology of convergence in law. Let $\mathcal{Q}$ be a subset of $Prob\ (Prob\ (X))$ such that the set $\{r(Q) : Q \in \mathcal{Q}\}$ is uniformly tight over X where $r(Q)(B) = \int_{Prob\ (X)} \nu(B)Q(d\nu)$. Then $\mathcal{Q}$ is uniformly tight over $Prob\ (X)$.*

Proof: Let (K_l) be a sequence of compact subsets of X such that

$$r(Q)(K_l^c) < \frac{1}{l} \text{ for all } \quad Q \in \mathcal{Q} \text{ and } \quad l \in \mathbb{N}.$$

Let $M_k = \{\nu \in Prob\ (X) : \nu(K_{2^{2m}}^c) \leq 2^{-m} \text{ for all } \quad m \geq k\}$. Then M_k is uniformly tight and hence relatively compact in $Prob\ (X)$, *cf.*[Top71]. Moreover one has $Q\{\nu : \nu(K_{2^{2m}}^c) > 2^{-m}\} < 2^{-m}$ for all $Q \in \mathcal{Q}$ and all m because otherwise

$$r(Q)(K_{2^{2m}}^c) = \int \nu(K_{2^{2m}}^c)Q(d\nu) \geq 2^{-m} \cdot 2^{-m} = 2^{-2m}$$

in contradiction to the choice of K_l. Thus

$$Q(M_k^c) \leq \sum_{m=k}^{\infty} 2^{-m} = 2^{-(k+1)}$$

for all $k \in \mathbb{N}$ and all $Q \in \mathcal{Q}$ which proves that $\mathcal{Q}$ is uniformly tight. ∎

Proof of Corollary 2: The corollary can be deduced from theorem 2 by choosing the metrics d, d' in such a way that the marginal map $Prob\ (I\!R_+ \times I\!R) \longrightarrow Prob\ (I\!R)$ is Lipschitz for d', κ' and for d, κ. However, it is also possible to give a simplified direct argument along the lines of the proof for theorem 2. ■

Proof of Theorem 1: Let $C < \infty, \varepsilon > 0$ and κ be given. Choose $g(t) = \frac{1}{\sqrt{Ct}}, h(t) = t$ and $\kappa' = \kappa$. For $\eta > 0$ let $\delta(\eta) > 0$ be maximal such that for all smaller δ the implication of corollary 2 holds. Let $f : (0, \infty) \longrightarrow (0, \infty)$ be a monotone function such that $f(t) \longrightarrow 0$ as $t \to 0$ and $f(\frac{t^3}{C}) \geq \eta$ whenever $t \geq \delta(\eta)$. We claim that Theorem 1 holds with this f.
Indeed, let $\eta = \nu\{\kappa(P^\omega, \overline{P}) > \varepsilon\}$. Then (23) fails. Now (18) follows from (16), and (24) is trivially satisfied by remark 3 for every $\delta > 0$. Thus by definition of $\delta(\eta)$ for all $\delta < \delta(\eta)$ the condition (19) must fail, i.e.

$$P \otimes P\{|\langle x, y\rangle| > \delta\} > \delta.$$

Because of (17) this implies $\rho(\mathcal{C}(P)) \geq \frac{\delta^3(\eta)}{C}$, and hence by the choice of f we get $\nu\{\kappa(P^\omega, \overline{P}) > \varepsilon\} = \eta \leq f(\rho(\mathcal{C}(P)))$.

In order to prove remark 1 let $\kappa^*(p, q) = \| \hat{p} - \hat{q} \|_{2,\lambda_1}$ where the $\hat{}$ indicates the characteristic function and λ_1 is the marginal of λ in lemma 1. Then in (27) one can replace d^* by κ^*. Thus $\delta(\eta) \geq 8\varepsilon^2\eta$ and hence the function f given by $f(t) = \frac{(Ct)^{\frac{1}{3}}}{8\varepsilon^2}$ has the desired properties. ■

4 A conditional central limit theorem

In probabilistic language theorem 2 implies results of the following type. As it turns out, for a triangular array of dependent random variables, even if it does not satisfy the CLT, one still gets a kind of CLT for weighted averages of the array if the weights are chosen in advance independently at random. The idea of the proof is to verify (20) and (21) in theorem 2 with the help of classical central limit arguments. We only consider the most simple situation of *iid* weights.

Theorem 3 *Let the triangular array $(X_{nk})_{n\in\mathbb{N}, 1\leq k\leq n}$ of random variables on some probability space $(\mathcal{X}, \mathcal{B}, \mathbb{P})$ satisfy*

$$\sum_{k=1}^{n} \mathbb{E}(X_{nk}^2) = O(n) \tag{29}$$

$$\sum_{k,j=1}^{n} (\mathbb{E}(X_{nk}X_{nj}))^2 = o(n^2) \tag{30}$$

$$\mathbb{P}\{\max_k |X_{nk}| \geq n\varepsilon\} \underset{n\to\infty}{\longrightarrow} 0 \quad \textit{for every } \varepsilon > 0. \tag{31}$$

Write S_n^2 for $\frac{1}{n}\sum_{k=1}^n X_{nk}^2$. Let $Y_1, Y_2, \ldots$ be iid. variables with mean 0 and variance 1, independent of all X_{nk}. Then the following statements hold.

a) If d is a metric describing convergence in law on $Prob(I\!R_+ \times I\!R)$ then for every $\varepsilon > 0$

$$\mathbb{P}\left\{d\left(\mathcal{L}\Big((S_n^2, \frac{1}{\sqrt{n}}\sum_{k=1}^n Y_k X_{nk}) \mid Y\Big),\ \mathcal{L}(S_n^2)\times\mathcal{N}\right) > \varepsilon\right\} \underset{n\to\infty}{\longrightarrow} 0. \tag{32}$$

b) If moreover

$$\lim_{\zeta\to 0}\limsup_{n\to\infty} \mathbb{P}\{\ S_n^2 > \zeta\ \} = 0 \tag{33}$$

(i.e. the limit points ρ of the (tight) sequence $(\mathcal{L}(S_n^2))$ do not charge the origin) then for every metric κ describing convergence in law on $Prob(I\!R)$ and every $\varepsilon > 0$

$$\mathbb{P}\left\{\kappa\left(\mathcal{L}(\frac{\sum_{k=1}^n Y_k X_{nk}}{\sqrt{n}S_n} \mid Y),\ \mathcal{N}(1)\right) > \varepsilon\right\} \underset{n\to\infty}{\longrightarrow} 0. \tag{34}$$

c) The assertions a) and b) hold even without the assumption (31) if and only if the Y_k are $\mathcal{N}(1)$-distributed.

Proof: a) Let $\varepsilon, \eta > 0$ be given. Let C be the constant in (29). Choose δ according to theorem 2 for the functions $g(t) = \frac{1}{\sqrt{Ct}}$ and $h(t) = t$. Let P_n' be the common distribution of the $X_{nk}, 1 \le k \le n$ on $I\!R^n$. Because of (29)and (30) the conditions (9) and (10) are satisfied. If P_n is the law of the $\frac{X_{nk}}{\sqrt{n}}, 1 \le k \le n$ then $\rho(\mathcal{C}(P_n)) \underset{n\to\infty}{\longrightarrow} 0$ as was shown in the last paragraph preceding corollary 1. The estimates (16) and (17) then show that (18) and (19) hold for every $\delta > 0$ for sufficiently large n.
Now let ν_n be the joint law of $(Y_1, \cdots, Y_n)$. If we can verify (20) and (21) then theorem 2 implies that the left hand side in (32) is $< \eta$ for sufficiently large n and we are done.
Let μ be the one dimensional law of the Y_k and assume (31). Let $\hat\mu$ be the characteristic function of μ. Then $\hat\mu(t) = \exp(-t^2/2) + r(t)t^2$ where $r(t) \underset{t\to 0}{\longrightarrow} 0$. Now for every $(s,t) \in I\!R_+ \times I\!R$ we have by independence

$$\begin{aligned}(\mathcal{F}(E_\nu P_2^\omega) - \mathcal{F}\mathcal{L}_p(\|\ \|^2)\times\mathcal{N})(s,t) &= \mathbb{E}\left(\exp[-sS_n^2](\exp\left[\frac{it}{\sqrt{n}}\sum_{k=1}^n X_{nk}Y_k\right] - \exp\left[-\frac{t^2S_n^2}{2}\right])\right)\\ &= \mathbb{E}\left(\exp[-sS_n^2]\left(\prod_{k=1}^n \hat\mu(\frac{tX_{nk}}{\sqrt{n}}) - \exp\left[-\frac{t^2S_n^2}{2}\right]\right)\right)\end{aligned}$$

Using the expansion of $\hat\mu$ and the condition (31) it is easy to see that

$$\prod_{k=1}^n \hat\mu(\frac{tX_{nk}}{\sqrt{n}}) - \exp\left[\frac{-t^2S_n^2}{2}\right] \underset{n\to\infty}{\longrightarrow} 0$$

in $\mathbb{P}$-measure. By dominated convergence we get (20) for sufficiently large n the metric d^* of Lemma 1. From (31) one also concludes that

$$\mathbb{P}(\max|X_{nk} - Z_{nk}| \ge n\varepsilon) \longrightarrow 0$$

where (Z_{nk}) is an independent triangular array with the same distribution as (X_{nk}). Thus the measure $P_n * \check{P}_n$ allows the same argument and hence (21) holds as well for d^*.
As for b) the condition (33) allows to assume that $S_n^2 > \zeta$ holds a.s. for some $\zeta > 0$: Simply replace for each n the underlying probability measure $\mathbb{P}$ by the conditional measure given the event $\{S_n^2 > \zeta\}$. By this change the (conditional) distribution of each of the random variables $\frac{\langle Y,X \rangle}{S}$ is changed in total variation at most by $\mathbb{P}\{S_n^2 \leq \zeta\}$ which is asymptotically arbitrarily small for sufficiently small ζ.
Now let $g \in C_c(I\!R)$ be a continuous function of compact support. Let f be bounded and continuous on $I\!R_+ \times I\!R$ such that $f(s,t) = g(\frac{t}{\sqrt{s}})$ whenever $s \geq \zeta$. Part a) implies for sufficiently large n that with probability close to 1 one has

$$\mathbb{E}\left(g(\frac{\langle Y,X \rangle}{\sqrt{n}S_n})|Y\right) = \mathbb{E}\left(f(S_n^2, \frac{\langle Y,X \rangle}{\sqrt{n}})|Y\right) \approx \mathbb{E}\int f(S_n^2,t)\,\mathcal{N}(S_n^2)(dt)$$

$$= \mathbb{E}\int_{I\!R} f(S_n^2, S_n t)\,\mathcal{N}(1)(dt) = \int g(t)\,\mathcal{N}(1)(dt)$$

which proves (34) modulo a tightness argument as in the proof of theorem 2.
c) If the Y_k have a $\mathcal{N}(1)$-distribution then (20) and (21) follow from (14) without reference to (31). The example c implies together with the Glivenko-Cantelli theorem that without the additional condition (31) the conclusions are only valid for $\mathcal{N}(1)$-distributed Y_k, proving part c) of the theorem. ∎

Let us stop here by asking two natural questions:

1. Suppose in theorem 3, $X_{nk} = X_k$ for some suitable stochastic process $(X_k)_{k\in\mathbb{N}}$. Can one get a.s. convergence of the conditional laws instead of convergence in probability?
2. Is there a common generalization of Dvoretzky's theorem on sections of convex bodies and the k-dimensional version of theorem 1 (cf. Remark 2) ?

Acknowledgement. A first version of this paper was written while the author was guest of the Steklov Institute at St. Petersburg. The author is grateful to V.N. Sudakov for the stimulating introduction to the subject and to J. Pitman and P. Bickel for drawing his attention to the paper [Dia84].

References

[Dia84] P. Diaconis and D. Freedman. Asymptotics of graphical projection pursuit. *Ann. of Stat.*, 12:793–815, 1984.

[Gro65] L. Gross. Abstract Wiener Spaces. *Proc. Fifth Berkeley Symp. on Math. Stat. and Prob.*, 2:31–42, 1965.

[GV61] I.M. Gelfand and N.Ya. Vilenkin. *Generalized functions 4 (English translation)*. Academic Press, New York, 1961.

[Sud78] V.N. Sudakov. Typical distributions of linear functionals on the spaces of high dimension. *Soviet Math. Dokl.*, 19 no.6:1578–1582, 1978.

[Top71] F. Topsøe. *Topology and Measure*. Springer Lecture Notes in Math. 133, Berlin-Heidelberg-New York, 1971.

Heinrich v. Weizsäcker
Fachbereich Mathematik der Universität
67663 Kaiserslautern
Germany
e-mail: weizsaecker@mathematik.uni-kl.de