R. Dudley

Some notes relating to J. Beck's paper
"Lower bounds on the approximation of the multivariate empirical process"

1 Introduction

This is not a self-contained exposition. It attempts to fill in details of the proof of Theorem 1 of J. Beck (1985), on slow convergence of the empirical process, for the uniform distribution on the unit cube, to its limiting Gaussian process uniformly over balls in $\mathbb{R}^d$ as d becomes large. This was the topic of the Stochastic Seminar at MIT and University of Connecticut, Storrs, in the fall of 2006. Beside myself, Richard Nickl was a main contributor to the seminar. Dmitry Panchenko and Wen Dong also made helpful suggestions. This exposition will only be readable while also reading Beck's paper.

The theorem is as follows. Let $d \geq 2$ and let P be the uniform distribution on the unit cube I^d in $\mathbb{R}^d$. For $0 < \eta \leq 1$ let $B(d, \eta)$ be the collection of all sets $B \cap I^d$ where B is a ball in $\mathbb{R}^d$ of radius r with $\eta/2 \leq r \leq \eta$. Let G_P be the Gaussian limit process of empirical processes $\nu_n := \sqrt{n}(P_n - P)$.

Theorem 1 (Beck, 1985) *Let $X_1, ..., X_n$ be i.i.d. (P), and take the corresponding P_n and ν_n. Then for any choice of G_P, and a constant $c_1 = c_1(d)$ depending only on d, the probability that for all η with $n^{-1/d} \leq \eta \leq 1$,*

$$\sup_{A \in B(d,\eta)\sqrt{n}} |(\nu_n - G_P)(A)| > c_1(d)(n\eta^d)^{1/2 - 1/(2d)}$$

is larger than $1 - e^{-n}$.

Here's a try at some overview, which may be clear only after going through some of the rest of this writeup and rereading the paper (Beck, 1985). In Beck's paper, Lemma 3 is a corollary of Sauer's Lemma, showing that the class of all balls in d-space is not too large in a Vapnik-Červonenkis sense. Lemmas 4 and 5 give regularity properties of related Gaussian processes. The hypothesis of Lemma 1 is that some signed measure (which in the application will be obtained from a discretization of a Gaussian process) has some such regularity. The conclusion of Theorem 1 is that the empirical process $\sqrt{n}(P_n - P)$ for P_n obtained from X_j i.i.d. uniform on the unit cube can be approximated uniformly over balls by its limiting Gaussian process G_P only within $O(n^{-1/(2d)})$, except with small probability. Lemma 1 shows that, with high probability (pertaining only to the Gaussian process), there is *no way* to choose P_n, however "uniformly" one might distribute the X_j over the cube, to get any better approximation. This illustrates the "irregularity of distribution" situation that Richard Nickl spoke about, cf. Beck and Chen (1987).

In Beck's paper, for each dimension $d = 1, 2, ...$, where only $d \geq 2$ is of interest, we have a fixed probability measure P on the sample space $\mathbb{R}^d$, namely the uniform probability on the unit cube I^d.

Beck uses a "Σ" notation for the union of sets. This seems harmless when the sets are disjoint cubes, although now unusual.

What Beck calls the Wiener 'measure' W is what we call the isonormal process W_P indexed by measurable sets, namely a Gaussian process with mean 0 and covariance $EW_P(A)W_P(B) =$

$P(A \cap B)$. What Beck calls the Brownian 'measure' is what we call the G_P process, also a Gaussian process with mean 0 indexed by measurable subsets of $\mathbb{R}^d$, and with covariance $EG_P(A)G_P(B) = P(A \cap B) - P(A)P(B)$. It's unnecessary to use ad hoc methods, such as "inclusion-exclusion" to define such processes on cubes, then to approximate other sets from inside and outside by unions of small cubes, as Beck does, just to define these processes. Still, the degree of approximation via unions of cubes seems to be crucial to the paper.

2 About the proof of Theorem 1

The definitions of C and l on pp. 292-293 of the paper are not very clear. The notation C seems unfortunate because C is not a constant, it depends on n and d. Actually, 5 to 4 lines from below on p. 296 Beck says C must be "sufficiently large," depending on d, let's say $C \geq M(d)$ for an $M(d)$ to be defined when we reach the point it's needed. (A first restriction is that $M(d) \geq 4$ for all d.) The other defining property of C is that $Cn = 2^{ld}$ for some integer $l \geq 1$. So let's define in terms of $M(d)$

$$l := l(d,n) := \left\lceil \frac{\log_2(M(d)n)}{d} \right\rceil \tag{1}$$

and then set

$$C := C(d,n) := 2^{dl(d,n)}/n. \tag{2}$$

Here the argument d is given first because it influences the order of magnitude of C, whereas n comes in through the requirement that l be an integer. With these definitions, one can check that we do have $2^{ld} = Cn$ and $C \geq M(d)$.

These talks are concerned only with Theorem 1 of Beck's paper, pertaining to balls in $\mathbb{R}^d$. Lemmas 1, 3, 4, and 5 of his paper all are part of the proof of Theorem 1. Thus, Beck's Theorem 2 and Lemma 2 won't be considered.

3 Definitions for Lemma 1

Since $c^*(d)$ occurs in the hypothesis $(d, n, C, *)$ of Lemma 1 (first display on p. 293 of Beck), it seems logically preferable to define $c^*(d)$ before stating Lemma 1. And, since $c^*(d)$ is defined in terms of $c_{21}(d)$, we also need to define $c_{21}(d)$ first.

The constant $c_{21}(d)$ is defined by Beck's last equation in his (30) by

$$c_{21}(d) = Cnv_d(\rho_0/2)^d = v_d 2^{ld-(l+2)d} = v_d/4^d,$$

where v_d is the volume of the unit ball in d dimensions, $v_d = \pi^{d/2}/\Gamma\left(\frac{d}{2}+1\right)$. So $c_{21}(d)$ is a well-defined, explicit constant depending only on d. Now, let's define

$$c^*(d) := \left(\frac{2^{d+3}\sqrt{2}}{\sqrt{2}-1} + 4^{d+1}\right) \cdot \frac{1}{c_{21}(d)}. \tag{3}$$

This also is explicit and depends only on d. The definition of $c^*(d)$ above differs from Beck's given just after his (31) by the insertion of 4^{d+1}, but it will be shown to give (31) as stated, in subsection 7.3.

Sometimes Beck refers to C as if it depends only on d, which is not correct, cf. (1) and (2). Thus to clarify the statement of Lemma 1, I propose replacing "property $(d, n, C, *)$" by "property $(d, n, M(d), *)$" where l is defined by (1) in terms of d, $M(d)$, and n. This hypothesis makes sense for any $M(d) > 4$. The particular $M(d)$ to be used in proving the theorem will be defined after finishing with Lemmas 3, 4, and 5. The last line of Lemma 1 I propose should say that $c_5 = c_5(d, M(d))$ depends only on d and $M(d)$, to avoid the relatively harmless, but annoying, dependence on n. Once $M(d)$ is specified, c_5 will then actually depend only on d.

4 About Beck's lemmas 3, 4, and 5

Let's recall some basic definitions and facts about VC classes of sets (whereas Beck seems to treat balls and half-spaces in $\mathbb{R}^d$ as special classes).

Let $\mathcal{C}$ be a class of subsets of a set X and F a finite subset of X. For any set T let $|T|$ be the number of elements of T if T is finite, or $+\infty$ otherwise. Let $\mathcal{C} \sqcap F := \{A \cap F : A \in \mathcal{C}\}$. Let $\Delta^{\mathcal{C}}(F) := |\mathcal{C} \sqcap F|$. For $N = 1, 2, \ldots$, let $m^{\mathcal{C}}(N) := \max\{\Delta^{\mathcal{C}}(F) : |F| = N\} \le 2^N$. Let $S(\mathcal{C}) := \sup\{N : m^{\mathcal{C}}(N) = 2^N\}$ Then $\mathcal{C}$ is called a VC (Vapnik-Chervonenkis) class if and only if $S(\mathcal{C}) < +\infty$.

For nonnegative integers N and r let $\binom{N}{\le r} := \sum_{j=0}^{r} \binom{N}{j}$, which equals 2^N if $r \ge N$. It's a known fact that for any $\mathcal{C}$ and N,

$$m^{\mathcal{C}}(N) \le \binom{N}{\le S(\mathcal{C})},$$

called Sauer's lemma, e.g. UCLT = Dudley (1999), Theorem 4.1.2. The statement has content only for $S(\mathcal{C}) < N < \infty$. The lemma is sharp, e.g. when $\mathcal{C}$ is the class of all sets having S or fewer members, so that $S(\mathcal{C}) = S$. Based on Sauer's lemma (which they seem to have rediscovered independently after earlier giving a less sharp inequality), Vapnik and Chervonenkis showed that: if $S = S(\mathcal{C})$, then

$$m^{\mathcal{C}}(N) \le 1.5 N^S / S!, \quad N \ge S + 2$$

(UCLT, Proposition 4.1.5). If $N = S + 1$, it's easy to check that $m^{\mathcal{C}}(N) \le 2^N \le 2N^S$. If $N \le S$ then the same holds (with the factor of 2 being needed only for $N = 1$), so we get in all cases

$$m^{\mathcal{C}}(N) \le 2N^{S(\mathcal{C})}. \tag{4}$$

For the class $\mathcal{B}(d)$ of all balls in $\mathbb{R}^d$ we have $S(\mathcal{B}(d)) = d + 1$ for all d. I actually published a separate paper proving this (Dudley, 1979). It soon turned out to be a special case of a more general fact. Consider the $(d+1)$-dimensional vector space A_d of all affine functions $x = (x_1, \ldots, x_d) \mapsto a_1 x_1 + \cdots + a_d x_d + c$ for constants a_j and c. Take the fixed function $f(x) := -\|x\|^2 = -x_1^2 - \cdots - x_d^2$. Let H_1 be the set of all functions $f + h$ for $h \in A(d)$. Consider the class $\mathcal{C}$ of all positivity sets $\{x : g(x) > 0\}$ for $g \in H_1$. Then $S(\mathcal{C}) = d + 1$ by a general fact about classes formed in such a way, depending only on the fact that we have a class of functions $A(d)$ forming a real vector space of a given dimension, and on f only in that it's a fixed function. It was first proved by Wenocur and Dudley (1981) and is given in UCLT, Theorem 4.2.1. Special cases were known earlier, e.g. for the class of half-spaces in $\mathbb{R}^d$, Radon (1921, p. 114). The class of sets we get in the present case is exactly $\mathcal{B}(d)$ (including the degenerate case of the empty set when the constant c is too small).

Now, looking at Beck's Lemma 3, $2^{kd} = N$ is the number of points in the kth integer lattice in I^d he considers, and $d+1 = S(\mathcal{B}(d))$, so we get the lemma from the above general considerations (basically Sauer's lemma), and with an improved constant 2 rather than $4d$ in front.

Given Lemma 3, the next step is Lemma 4. Just before it, m is defined as the least integer (≥ 1, as we need but Beck doesn't mention) with

$$2^m/m \geq c_0(d)n^2 \tag{5}$$

where $c_0(d)$ is a constant depending on d to be defined later. One can see from this that as $n \to \infty$, m is asymptotic to $2\log_2 n$.

To clarify Beck's statement of Lemma 4, I'd suggest instead of "0 and l" in the first line it should say "0 or l."

In my old copy of Beck's paper, I missed the bottom line of p. 294. It should say (that the probability of the given event) "is greater than $1 - e^{-n-2}$". (Although the issue of ZW containing the paper was missing when I last looked for it in the MIT library, I have a copy of an original MS of the paper Beck had sent me.)

5 Counting cubes

Dmitry Panchenko and Wen Dong suggested the main ideas for this section. For $t = 0, 1, ...$, recall Beck's notation for cubes of side 2^{-t} included in the unit cube I^d, for each $\underline{i} = (i_1, ..., i_d)$ where each i_j is a nonnegative integer less than 2^t,

$$I(t;\underline{i}) \ := \ \left\{(x_1, ..., x_d) \in \mathbb{R}^d : \ \frac{i_j}{2^t} \leq x_j < \frac{i_j+1}{2^t}, \ j = 1, \ldots, d\right\}. \tag{6}$$

Each such cube will be called a *t-cube*. There are exactly 2^{td} of them, and they are disjoint. Each t-cube is a disjoint union of 2^d $(t+1)$-cubes. The cube $I(t, \underline{i})$ contains a vertex

$$p(t;\underline{i}) := (i_1, ..., i_d)/2^t.$$

In fact this is the only one of its vertices that it contains. The one 0-cube $I(0; (0, 0, \ldots, 0))$ equals the unit cube I^d except for the d faces $x_j = 1$ which have 0 volume.

We are considering the class $\mathcal{B}(d)$ of all balls $B(x, R) := \{y : \ \|x - y\| < R\}$ which can have any center $x \in \mathbb{R}^d$ and any radius $R > 0$. For any given d, ball $G = B(x, R)$, and $t = 0, 1, ...$, let $N_7(G, t)$ be the number of $(t+1)$-cubes included in G but not included in any t-cube included in G (the left side of Beck's inequality (7) in the special case $k = 0$). Let $N_6(G, t)$ be the number of t-cubes $I(t, \underline{j})$ such that the vertex $p(t, \underline{j}) \in G$ but the cube $I(t, \underline{j})$ is not included in G (the left side of Beck's inequality (6) in the special case $k = 0$ and where t replaces m).

Claim. For each $d = 2, 3, ...$ there is a constant $C_{67}(d)$ depending only on d such that

$$\max\left(\sup_{G \in \mathcal{B}(d)} N_6(G, t), \sup_{G \in \mathcal{B}(d)} N_7(G, t)\right) \ \leq C_{67}(d)2^{t(d-1)}$$

for all $t = 0, 1, \ldots$.

Proof of claim: We will take $C_{67}(d) \geq 2^d - 1$. This condition is exactly what is needed for the claim to be valid for $t = 0$ (specifically for N_7), so assume from now on that $t \geq 1$.

Let λ^d be d-dimensional Lebesgue measure. This is what will be meant by volume in $\mathbb{R}^d$. Recall that the volume of the unit ball in $\mathbb{R}^d$ is $v_d := \pi^{d/2}/\Gamma(1+\frac{d}{2})$. (This fits with the familiar values for $d = 1, 2, 3$, and can be proved by induction from d to $d+2$ for each $d \geq 2$.) Thus the volume of any ball $B(x, R)$ is $v_d R^d$. The total $(d-1)$-dimensional area of the unit sphere in $\mathbb{R}^d$ is $A_d = (d/dR)v_d R^d|_{R=1} = dv_d$. This also fits with familiar values for $d = 2$ and 3.

A t-cube has diameter $\delta_{d,t} := \sqrt{d}/2^t$ and has a center, such that each point has distance from the center at most $\sqrt{d}/2^{t+1}$.

Let $G = B(x, R)$. For both suprema we want to bound, the cubes are or are included in t-cubes that intersect G but are not entirely included in it. Thus these t-cubes are within $\delta_{d,t}$ of the boundary sphere $\{y : \|y - x\| = R\}$ and so are included in $B(x, R + \delta_{d,t})$ and disjoint from $B(x, R - \delta_{d,t})$ (where the latter is empty if $R \leq \delta_{d,t}$). Also, all the cubes are subsets of the unit cube I^d.

Let's find upper bounds for the d-dimensional volume V of the region

$$I^d \cap B(x, R + \delta_{d,t}) \setminus B(x, R - \delta_{d,t}).$$

If $R \leq \delta_{d,t}$ then we use simply

$$V \leq \lambda^d(B(x, R + \delta_{d,t})) \leq v_d(2\delta_{d,t})^d = 2^d d^{d/2} v_d/2^{td}. \tag{7}$$

If $\delta_{d,t} < R \leq \frac{3}{2}\sqrt{d}$ then we ignore the I^d and use $t \geq 1$, giving the bound

$$V \leq v_d((R+\delta_{d,t})^d - (R-\delta_{d,t})^d) = dv_d \int_{R-\delta_{d,t}}^{R+\delta_{d,t}} r^{d-1}dr \leq dv_d(R+\delta_{d,t})^{d-1} \cdot 2\delta_{d,t} \leq dv_d(2\sqrt{d})^d/2^t. \tag{8}$$

The numbers N_6 and N_7 we need to bound are both 0 unless the boundary sphere of the ball intersects I^d, i.e. for some $q \in I^d$, $\|x - q\| = R$, so assume that to hold. If $R > \frac{3}{2}\sqrt{d}$ then the center x must be outside I^d. The volume V will be bounded by integrating in spherical coordinates with center x. Consider for $G = B(x, R)$ the line L through x and the center $p := p(1, (1, 1, \ldots, 1)) = (1/2, 1/2, \ldots, 1/2)$ of I^d. The distance $\|p - q\| \leq \sqrt{d}/2$ and so by the triangle inequality $\|p - x\| \geq R - \|p - q\| \geq R - \sqrt{d}/2$.

For any $q' \in I^d$ we also have $\|p - q'\| \leq \sqrt{d}/2$. Let M be the line through x and q' and let θ be the maximum value of the acute angle between L and M, namely

$$\theta := \theta_{d,R} = \sin^{-1}\left(\frac{\sqrt{d}/2}{R - \sqrt{d}/2}\right). \tag{9}$$

Thus $\theta \leq \sin^{-1}(1/2) = \pi/6$. In bounding V by an integral in spherical coordinates we have an angular factor A which is the $(d-1)$-dimensional area of a cap on the unit sphere.

Here is an idea of Dmitry Panchenko's that could simplify some of the following but I didn't have time to incorporate it. Instead of only the unit sphere, let's find an upper bound for the $(d-1)$-dimensional area B of the part C of any sphere included in the unit cube. Consider radii from the center of the sphere passing through C and continuing until they encounter the surface S of the unit cube. This gives a continuous one-to-one map of C onto a subset E of S. Let α_{d-1} denote $(d-1)$-dimensional area. Then $B = \alpha_{d-1}(C) \leq \alpha_{d-1}(E) \leq \alpha_{d-1}(S) = 2d$ because a tangent hyperplane to C is perpendicular to the radius and so areas in E, if anything, are larger; the last equation holds since I^d has $2d$ faces, each of unit area. (Any notation in this paragraph is unrelated to notation in the rest of the draft.)

Returning to the older argument, it will be shown that the "area" or $(d-1)$-dimensional volume of a "ball" in the geodesic distance on the unit sphere is bounded above by that of a Euclidean $(d-1)$-dimensional ball of the same radius, which is not at all surprising since the sphere is positively curved. In fact it must be well known, but having happened to write out the following proof, I had no time to search the literature for a reference. For small radii, the bound is asymptotically sharp, which is again not surprising since locally the sphere is approximately flat.)

In dimension $d = 2$, the "area" becomes an arc length on the unit circle, bounded above by

$$A \leq 2\theta = v_1\theta. \tag{10}$$

(Arcs of circles, spanning an angle less than π, with arc length distance, are actually isometric to Euclidean intervals, so there is no curvature effect here.) If $d \geq 3$, then

$$A \leq A_{d-1}\int_{\cos\theta}^{1}(1-x^2)^{(d-3)/2}dx.$$

(The integrand equals 1 for $d = 3$, in accordance with the well-known fact that the area of a zone on a sphere in $\mathbb{R}^3$ is proportional to its height, and for $\theta = \pi/2$ we get $A_2 = 2\pi$, the area of a hemisphere as desired.) With a change of variables $y = 1 - x$ we get for $d \geq 3$

$$A \leq A_{d-1}\int_{0}^{1-\cos\theta}(2y-y^2)^{(d-3)/2}dy.$$

We have $1 - \cos\phi \leq \phi^2/2$ for all real ϕ. For $d = 3$ we get an upper bound

$$A \leq \pi\theta^2 = v_2\theta^2. \tag{11}$$

For $d \geq 4$ we have

$$A \leq A_{d-1}\int_{0}^{1-\cos\theta}(2y)^{(d-3)/2}dy = A_{d-1}2^{(d-3)/2}\frac{2}{d-1}(1-\cos\theta)^{(d-1)/2}$$

and so

$$A \leq \frac{A_{d-1}\theta^{d-1}}{d-1} = v_{d-1}\theta^{d-1}. \tag{12}$$

We have $\sin^{-1}x \leq 2x$ for $x \leq \pi/4$ and so from (9),

$$\theta \leq \frac{\sqrt{d}}{R-\sqrt{d}/2}. \tag{13}$$

By the bounds (10), (11) and (12) we get that for each dimension $d \geq 2$ and $R > \frac{3}{2}\sqrt{d}$,

$$A \leq v_{d-1}\theta^{d-1} \leq v_{d-1}\left(\frac{\sqrt{d}}{R-\sqrt{d}/2}\right)^{d-1}. \tag{14}$$

We also have a radial one-dimensional factor, which is bounded above by a factor in an intermediate term in (8), namely $2\delta_{d,t}(R+\delta_{d,t})^{d-1}$. Multiplying this by the right side of (14) gives

$$V \leq 2\delta_{d,t}v_{d-1}\left(\frac{\sqrt{d}(R+\delta_{d,t})}{R-\sqrt{d}/2}\right)^{d-1} \leq\ 2v_{d-1}d^{d/2}2^{d-1}/2^t \tag{15}$$

since $R \geq \frac{3}{2}\sqrt{d}$ and $t \geq 1$. Taking the maximum of three constants for each dimension, from (7), (8), and (15) we find that for each dimension $d \geq 2$ there is a constant $\beta(d)$ such that $V \leq \beta(d)/2^t$ for all $t = 1, 2, \ldots$ and all values of $R > 0$. Now, the number of t-cubes that can be included in a region of volume $\leq V$ is at most $2^{td}V \leq \beta(d)2^{t(d-1)}$, whereas the number of $(t+1)$-cubes is at most $2^{(t+1)d}V \leq 2^d\beta(d)2^{t(d-1)}$. We thus obtain

$$\sup_{G\in\mathcal{B}(d)} N_6(G,t) \leq 2^{td}V \leq \beta(d)2^{t(d-1)}$$

and

$$\sup_{G\in\mathcal{B}(d)} N_7(G,t)) \leq 2^d\beta(d)2^{t(d-1)},$$

giving the claim with $C_{67}(d) = 2^d\beta(d)$. □

In Beck's bound (6) it will be helpful if $m \geq l$, in fact the inequality could fail if $m < l$. So let's check that $m \geq l$ for n large enough. From (1) we have since $d \geq 2$ that as $n \to \infty$ $l \sim \log_2 n/d \leq \log_2 n/2$. From (5), recalling $m \geq 1$ we have $m \sim 2\log_2 n$ as $n \to \infty$. So $m > l$ for n large enough, say for $n \geq n_1(d)$ where $n_1(d)$ depends on $M(d)$ and $c_0(d)$, neither of which has been chosen yet.

Suppose that $n \geq n_1(d)$. From the claim we obtain instead of Beck's bounds (6) and (7) of his paper, first for $k = 0$, some weaker bounds with some constants depending only on d in place of 2^d in each. With respect to t, the factor $2^{t(d-1)}$ is the same. Moreover, decomposing a k-cube into m-cubes for $m \geq k$ is the same as decomposing the 0-cube (unit cube) into $(m-k)$-cubes, rescaling by a factor of 2^k, which preserves the family of balls. Likewise, decomposing a k-cube into t-cubes and then $(t+1)$-cubes for $t \geq k$ is the same as decomposing the unit cube into $(t-k)$-cubes, then $(t+1-k)$-cubes. So we obtain the bounds of the form of Beck's (6) and (7), again with other constants in place of 2^d in each, but for general k ($k = l$, in the statement of Beck's Lemma 4) in place of $k = 0$.

The Claim thus gives an inequality in place of Beck's (6),

(6′)

$$\operatorname{card} J(G,m,k,\underline{i})\ \leq\ C_{67}(d)2^{(m-k)(d-1)}, \tag{16}$$

and an inequality in place of Beck's (7),

(7′)

$$\operatorname{card}\{\underline{j}:\ I(t+1;\underline{j}) \subset G_{t+1}(k;\underline{i}) \setminus G_t(k;\underline{i})\}\ \leq\ C_{67}(d)2^{(t+1-k)(d-1)}. \tag{17}$$

Each of the cubes $I(m;\underline{j})$ counted in (16) has volume $1/2^{md}$, and so the volume of their union V for a fixed G is at most $C_{67}(d)2^{-m-k(d-1)}$. So the standard deviation σ of $B(V)$ satisfies

$$\sigma = \sqrt{\lambda^d(V)(1-\lambda^d(V))} \leq \sqrt{\lambda^d(V)} \leq \sqrt{C_{67}(d)}2^{-[m+k(d-1)]/2}.$$

We thus get for a given V and any $q > 0$

$$\Pr(|B(V)| \geq q) \leq 2\left[1 - \Phi\left(qC_{67}(d)^{-1/2}2^{[m+k(d-1)]/2}\right)\right].$$

Then by Beck's Lemma 3 (as slightly improved, replacing $4d$ by 2) we get a bound for the probability that the above event occurs for any V of the given form

$$\Pr(\exists V:\ |B(V)| \geq q) \leq 2 \cdot 2^{(m-k)d(d+1)}\left[1 - \Phi\left(qC_{67}(d)^{-1/2}2^{[m+k(d-1)]/2}\right)\right]. \tag{18}$$

Likewise, each $(t+1)$-cube counted in (17) for a given G has volume $1/2^{t(d+1)}$, so a union U_t of such cubes has volume at most $C_{67}(d)2^{-k(d-1)-t-1}$, and the standard deviation τ_t of $B(U_t)$ satisfies $\tau_t \leq \sqrt{C_{67}(d)}2^{-[k(d-1)+t+1]/2}$. We thus get, by (improved) Lemma 3 again, the probability that $|B(U_t)| \geq q_t$ for any such U_t and $q_t > 0$ bounded above by

$$\Pr\{\exists U_t:\ |B(U_t)| \geq q_t\} \leq 2^{(t+1-k)d(d+1)+2}\left(1 - \Phi\left(q_t(C_{67}(d))^{-1/2}2^{[t+1+k(d-1)]/2}\right)\right). \tag{19}$$

Thus we get, in place of Beck's (8), an inequality, say (8′), with initial factors $4d$ replaced by 2 and $2^{-d/2}$ inside Φ's replaced by $C_{67}(d)^{-1/2}$.

For $c_8'(d)$ not yet chosen, which is not the same as Beck's $c_8(d)$ but plays essentially the same role, set

$$q' \ :=\ c_8'(d)\sqrt{nm}2^{-(m+k(d-1))/2}$$

and for $t = m, m+1, \ldots$,

$$q_t' \ :=\ c_8'(d)\sqrt{n(t+1)}2^{-(t+1+k(d-1))/2}.$$

Then

$$q' + \sum_{t=m}^{\infty} q_t' = c_8'(d)\sqrt{n}2^{-k(d-1)}\sum_{t=m}^{\infty}\frac{\sqrt{t}}{2^{t/2}}$$

where the sum equals $\sqrt{m}2^{-m/2}S$ with, for $m \geq 2$,

$$S = \sum_{r=0}^{\infty}\sqrt{1+\frac{r}{m}}2^{-r/2} \leq \sum_{r=0}^{\infty}\left(1+\frac{1}{m}\right)^{r/2}2^{-r/2} \leq 1/(1-\sqrt{3/4}) < 7.5.$$

We can now define $c_7(d)$, in the statement of Beck's Lemma 4, as $7.5c_8'(d)$, and the two will fit together as they should. We still need to choose $c_8'(d)$.

Using $1 - \Phi(x) \leq \exp(-x^2/2)$ for $x \geq 0$, we have

$$1 - \Phi\left(q'C_{67}(d)^{-1/2}2^{[m+k(d-1)]/2}\right) = 1 - \Phi(\kappa(d)\sqrt{nm}) \leq \exp(-\kappa(d)^2mn/2)$$

where $\kappa(d) = c_8'(d)/\sqrt{C_{67}(d)}$. To justify the first term on the last line of Beck's p. 295, a calculation shows it suffices if $\kappa(d) \geq d+1$ if $n \geq 2$ and $m \geq 2$, which will be true for n large enough by (5).

The sum at the bottom of p. 295 starts at $t = m$. To justify the terms in the sum, a similar calculation shows it suffices if $\kappa(d) \geq d+2$ for $n \geq 2$. Thus the last line as a whole will hold if we set

$$c_8'(d) \ :=\ (d+2)\sqrt{C_{67}(d)},$$

which is explicit since $C_{67}(d)$ is given in the Claim and its proof.

The rest of the proof of Lemma 4, in the three-line display at the top of p. 296, then follows easily because: there are 2^{kd} k-cubes $I(k;\underline{i})$; a geometric series is summed; we have $2^{-d}/(1-2^{-d}) <$ for $d \geq 2$; and $k \leq l \leq m$ for n large enough.

6 The rest of the proof of Theorem 1, assuming Lemma 1

Lemma 5 is easy. Beck refers to a paper of Révész. In the Lemma, $q^2/3$ can be replaced by $q^2/(2+\delta)$ for any $\delta > 0$ as Révész does. Since $\mathcal{B}(d)$ is a (uniform) Donsker class, it is in particular a GB-set and GC-set in $L^2(P)$ and Lemma 5 follows from the Landau-Shepp-Marcus-Fernique theorem given e.g. in UCLT, Theorem 2.2.8.

After Lemma 5, Beck's inequality (10) follows directly by rescaling.

In the next display, $c^*(d)$ has, in this version, already been defined in Section 3. In Beck's (11), the first inequality in terms of a sum of binomial upper tail probabilities

$$E(r+1, R, p) \;:=\; \sum_{j=r+1}^{R} \binom{R}{j} p^j (1-p)^{R-j}$$

is immediate. The bound by e^{-n-3}, however, depends on $M(d)$ in (1) which needs to be chosen here large enough. It needs to be shown that can be done.

Of course, binomial tail probabilities are basic in empirical processes, when one deals with empirical measures and processes on a single measurable set. In this case unfortunately there seems to be no nice upper bound by a normal probability as in "Okamoto" cases. Rather we have a small p and large R, a Poisson kind of case. We can apply the Chernoff inequality, due to H. Chernoff in 1952 and given in UCLT, (1.3.9), which gives after the simplification $E(r+1, R, p) \le E(r, R, p)$, for $r \ge Rp$,

$$E(r, R, p) \le \left(\frac{Rp}{r}\right)^r \left(\frac{R(1-p)}{R-r}\right)^{R-r}. \tag{20}$$

The quantities appearing are in this case, where t is any positive integer, $p = c_{10}(d)\exp(-q^2/3)$ where $c_{10}(d)$ is as in Lemma 5, and as Beck states just after his (11), $q = t\sqrt{C(d,n)}/(2c^*(d))$, $R = 2^{l(d,n)d} = C(d,n)n$, $r = n/t^{3/2}$. Here $c^*(d)$ is as defined in Section 3. $M(d)$, needed in the definitions of $l(d,n)$ (1) and thus $C(d,n)$, is not yet chosen. We need to see next how to choose $M(d)$ so that Beck's (11) holds.

To apply the inequality we need to check that $r \ge Rp$, which reduces to $t^{-3/2} \ge Cp$. That isn't immediate since C is not yet chosen and will be large. We'll need to return to this issue later.

Let's first look at the right-hand factor F_2 on the right in (20), where n cancels in the fraction, giving

$$\left(\frac{C(1-p)}{C-t^{-3/2}}\right)^{n(C-t^{-3/2})} \le \left(\frac{C}{C-1}\right)^{nC},$$

so for $D := C-1$

$$F_2 \le \left(1+\frac{1}{D}\right)^{n(D+1)} \le e^{2n} \tag{21}$$

if $C \ge 2$, so that $D \ge 1$.

Now let's look at the left-hand factor on the right in (20), namely $F_1 := (Rp/r)^r$, where again n cancels in the fraction, giving $F_1 = F_{11}F_{12}$ where $F_{11} = (Ct^{3/2})^r$. For any number y such as $t^{3/2}$ with $y \ge 1$, we have $1 \le y^{1/y} \le 2$. It follows that

$$F_{11} \le (2C)^n \le \exp((\log 2 + \log C)n). \tag{22}$$

We have

$$F_{12} = p^r = \left(c_{10}(d) \exp\left(-\frac{Ct^2}{12c^*(d)^2} \right) \right)^{nt^{-3/2}} \leq F_{121}F_{122}$$

where $F_{121} := c_{10}(d)^n$ assuming that $c_{10}(d) \geq 1$, as we can (cf. Lemma 5) and

$$F_{122} = \exp\left(-\frac{Cn\sqrt{t}}{12c^*(d)^2} \right).$$

We then have by (21), (22) and the definition of F_{121}

$$F_2 F_{11} F_{121} \leq e^{n(K+\log C)}$$

where $K = 2 + \log 2 + \log c_{10}(d)$ doesn't depend on t. The entire Chernoff bound applied to just $t = 1$ will give us less than e^{-n-4} if $M(d)$ is large enough so that for $C \geq M(d)$,

$$C \geq 12c^*(d)^2 \left[4 + K + \log C \right], \tag{23}$$

as is possible. Then, because F_{122} becomes smaller as t becomes larger, while F_{11} and F_{121} have bounds not depending on t, we get that $F_1 < 1$ for all t (which is not true for F_2). This implies that $Rp < r$ and so Chernoff's inequality does apply.

We also need to control the sum of the terms for $t \geq 2$. The ratio of the tth term to the $t = 1$ term in F_{122} is

$$\exp\left(-\frac{Cn[\sqrt{t} - 1]}{12c^*(d)^2} \right).$$

The sum of these terms from $t = 2$ to $+\infty$ converges for any $C > 0$. For C large enough, for $n = 1$, by dominated convergence for sums, the sum is less than 1, which then holds also for all n. Thus we need $M(d)$ large enough to make the given sum less than 1 while (23) also holds for $C \geq M(d)$. Then Beck's bound (11) does hold.

A third and (we hope) last requirement on $M(d)$ is that it should be large enough that if Z is a $N(0, 1)$ variable and $C \geq M(d)$, then

$$2\left(1 - \Phi\left(\frac{C\sqrt{n}}{4c^*(d)} \right) \right) \leq 2\exp\left(-\frac{C^2 n}{32c^*(d)^2} \right) < e^{-n-2}.$$

Such a choice is also possible. Then, $M(d)$ is defined and so $C = C(n, d)$ and $l = l(n, d)$ are defined in (2) and (1) respectively. Instead of saying that $C = c_9(d)$ (twice) it seems better to say that $C \geq M(d)$.

Beck then defines $c_0(d)$, which appears in the definition (5) of m, as $(4c_7(d)c^*(d))^2$, recalling again that $c^*(d)$ is as defined in Section 3 above. That brings us to the bottom of p. 296 of Beck's paper.

The rest of the proof is actually straightforward. Starting from the top of p. 297, the inequality in the first three lines is just a triangle inequality. The next line follows by Lemma 4 with $k = l$ for the first term and the definition of B for the second. Also note that $\lambda(I(l; \underline{i})) = 2^{-ld}$ doesn't depend on i. In the fifth line, the second term follows from line 3 from below on p. 296. In the next line, labeled by (12), we use the inequality in the last two lines of p. 296 and the fact that $2^{ld} = Cn$ (2). The probability lower bound stated just after (12) follows from the reasons for the fourth and fifth lines. For (13), recall that $Ball(d, \omega)$ is the set of balls with radius r such that $\omega/2 \le r \le \omega$. Here Lemma 1 is applied straightforwardly. Then Lemma 4 is applied for $k = 0$ (to the entire unit cube instead of small cubes) to give (14), again straightforwardly, and Theorem 1 does follow as Beck states from his (13) and (14) and the same bound at the bottom of p. 296.

7 About the proof of Lemma 1

The hardest part of the proof of Beck's Theorem 1 may be Lemma 1. Its proof is in Section 3 of his paper, pp. 298-302. At any rate that is as long as the rest of the proof combined. This section of course is much indebted to Richard Nickl's three lectures and notes.

7.1 Constants in Fourier transforms

The choice of constants in Fourier transforms turns out to be of no real import to Beck's arguments, as will be seen at the end of this subsection. Still, the paper contains a few equations that are not correct as stated because of such constants, so an attempt is made here to straighten them out.

Let μ be a finite signed measure on the Borel sets of $\mathbb{R}^d$. For a constant F with $0 < F < \infty$, the *Fourier transform* of μ (for the given F) will be defined by

$$(\mathcal{F}\mu)(t) \;:=\; F^d \int e^{-i(x,t)} d\mu(x)$$

where $(x,t) = \sum_{j=1}^d x_j t_j$. Then $\mathcal{F}(\mu)(\cdot)$ is a bounded, continuous, complex-valued function of t. The most often used constants F are $F = 1$ or $F = 1/\sqrt{2\pi}$, each of which allows a particular theorem to be stated without an additional constant. Beck, however, chooses $F = 1/\sqrt{\pi}$ (belatedly, in non-numbered displays after his (19) and (22)).

In defining the characteristic function of a probability measure, $+i$ rather than $-i$ is used in the exponent. The difference is one of convention rather than substance, except that the sign of i in an inverse Fourier transform should be opposite to that in the (direct) transform.

Sometimes, in analysis, other factors are used in the exponent such as $\pm 2\pi i$ instead of $\pm i$. But that is not done in probability theory nor by Beck, so it won't be further considered here.

For a function $f \in \mathcal{L}^1(\mathbb{R}^d, \lambda^d)$, there is a corresponding signed measure μ with $d\mu(x) = f(x)dx$, and we let

$$(\mathcal{F}f)(t) \;:=\; F^d \int e^{-i(x,t)} f(x)dx.$$

Two finite signed measures μ and ν have a convolution, another finite signed measure $\mu * \nu$ defined by

$$(\mu * \nu)(A) \;=\; \textstyle\int \mu(A - y)d\nu(y).$$

If $F = 1$, it's well known (e.g. in probability theory) that $\mathcal{F}(\mu * \nu)(t) = \mathcal{F}(\mu)(t)\mathcal{F}(\nu)(t)$ for all t. For general F we have

$$\mathcal{F}(\mu * \nu)(t) \;=\; F^{-d}(\mathcal{F}\mu)(t)(\mathcal{F}\nu)(t) \tag{24}$$

for all t. Thus Beck's equation (20) needs correction, by a factor of $F^d = \pi^{d/2}$ on the right for his choice of F. It also has a typo, with a hat missing over $f * g$.

If $f \in (\mathcal{L}^1 \cap \mathcal{L}^2)(\mathbb{R}^d, \lambda^d)$ (complex-valued), then by the well-known Plancherel theorem, $\mathcal{F}f \in \mathcal{L}^2(\mathbb{R}^d, \lambda^d)$ also. If $F = 1/\sqrt{2\pi}$ then $\int |\mathcal{F}f(t)|^2 dt = \int |f(x)|^2 dx$, and $f \mapsto \mathcal{F}f$ extends uniquely to an isometry (unitary transformation) of $L^2(\mathbb{R}^d)$ onto itself. For a general F, we get

$$\int |(\mathcal{F}f)(t)|^2 dt = (2\pi F^2)^d \int |f(x)|^2 dx \tag{25}$$

instead of Beck's (21), which needs a correction factor of $(2\pi F^2)^d$ on the left, or 2^d for Beck's F.

There is no choice of F that makes both Beck's (20) and (21) correct as stated. With his F, Beck's (22) also needs correction factors of $(\pi/2)^d$ in the latter two expressions.

This (22) is not used for a while, until the last half page of the proof of Lemma 1. There it is used in his (35), for the definition of $\Delta(\cdot)$ which is unchanged. Then in the first equation of Beck's (36), a factor of $(\pi/2)^d$ should be inserted on the right. Putting that factor also into the term on the right of the next inequality, the next equation becomes true. At any rate the inequality between the first and last expressions in (36) will be true with the proviso that $c_{22})(d, C)$ first appearing and implicitly defined here would preferably be replaced by some constant depending on d and $M(d)$ and thus really only on d, as mentioned in Section 3.

For the Fourier transform of the indicator $1_{G(0,r)}$ of the ball centered at 0 with radius r, I checked Beck's (23) and with his F found $c_{14}(d) = v_{d-1} = \pi^{(d-1)/2}/\Gamma((d+1)/2)$ and in his (25), $c_{15}(d) = (2\pi)^{d/2}$. All that is needed for the proof is that $c_{14}(d)$ and $c_{15}(d)$ are constants > 0 just depending on d.

7.2 Asymptotic expansions and Beck's (24) through (29)

The Poisson integral for a Bessel function, given as Beck's (24) with a reference, also appears in the book of Watson (1941, §2.3, (3) p. 35), noting that for $z = x$ real, an integral $\int_{-1}^{1} e^{iux} v(u) du$ where $v(\cdot)$ is an even function equals the integral with e^{iux} replaced by $\cos x$. A function $g(r, t)$ is defined by Beck's (25). Let $\cos_d(x) := \cos(x - (d+1)\pi/4)$ and $\sin_d(x) = \sin(x - (d+1)\pi/4)$. Then display (26) says that the Bessel function $J_{d/2}(x)$ which appears in (25) has the asymptotic expansion

$$\left(\frac{2}{\pi x}\right)^{1/2} \left(\cos_d(x) \sum_{j=0}^{\infty} \frac{B_{2j,d}}{x^{2j}} - \sin_d(x) \sum_{j=0}^{\infty} \frac{B_{2j+1,d}}{x^{2j+1}} \right) \tag{26}$$

as $x \to +\infty$, where the coefficients $B(j, d)$ are given explicitly. The meaning of the asymptotic expansion is that for any fixed d and $N = 0, 1, \ldots$, if $S_N(x)$ is the finite partial sum of the terms with $x^{-i-1/2}$ ($i = 2j$ or $2j+1$) having $i \le N$, then $J_{d/2}(x) - S_N(x) = O(x^{-N-3/2})$ as $x \to +\infty$ (note the factor of order $x^{-1/2}$ outside the parentheses; truncating the sums inside the parentheses gives an error of the order of the lowest-order term omitted). It is not asserted that the series actually converge. If d is odd, then just one of the two series in (26) becomes a finite sum, depending on the residue of d mod 4. If d is even, neither series is a finite sum. If the sums have infinitely many terms it seems that the numerators of the coefficients become quite large, for example of order $(2j - 1 - d)!!^2$ if j is even, where for K odd $K!! = K(K-2) \cdots 3 \cdot 1$. Watson (1941, p. 11) states that (in the special case of such a series for J_0) the "series are, however, not convergent but asymptotic." But if one understands the meaning of asymptotic expansion, the divergence is of no concern. Beck just takes the leading term and applies the meaning for $N = 0$, giving his (27).

After (28) Beck defines the function

$$h(\rho, t) = \frac{2}{\rho} \int_{\rho/2}^{\rho} g^2(r, t) dr.$$

By "uniformly large" Beck appears to mean bounded away from 0, specifically as in inequality (29). On the left side of (29) we need an upper bound for the denominator $h(\rho, t)$. We have by (27) for $\varepsilon = 2^{-d-2}$ that for $\gamma_d(r,t) := c_{16}(d) r^{(d-1)/2} t^{-(d+1)/2} \cos_d(rt)$,

$$|g(r, t) - \gamma_d(r, t)| \ < \ \frac{c_{16}(d) r^{(d-1)/2}}{2^{d+2} t^{(d+1)/2}}$$

for $rt > c_{17}(d, 2^{-d-2})$. Thus taking L^2 norms for the uniform distribution $U[\rho/2, \rho]$ having density $\frac{2}{\rho} 1_{[\rho/2,\rho]}$ we get for $\rho t/2 > c_{17}(d, 2^{-d-2})$ that

$$\|g(\cdot, t) - \gamma_d(\cdot, t)\|_2 \ < \ \frac{c_{16}(d)}{2^{d+2} t^{(d+1)/2}} \left[\frac{2}{\rho} \int_{\rho/2}^{\rho} r^{d-1} dr \right]^{1/2} \ < \ \frac{c_{16}(d) \rho^{(d-1)/2}}{2^{d+2} t^{(d+1)/2}}.$$

It follows that

$$\|g(\cdot, t)\|_2 \ \leq \ \frac{9}{8} c_{16}(d) \rho^{(d-1)/2} / t^{(d+1)/2}.$$

Thus for $c_{166}(d) := 2 c_{16}(d)^2$ and $\rho t > 2 c_{17}(d, 2^{-d-2})$,

$$h(\rho, t) \ = \ \|g(\cdot, t)\|_2^2 \ \leq \ c_{166}(d) \frac{\rho^{d-1}}{t^{d+1}}.$$

For a lower bound, we have for $\rho t/2 > c_{17}(d, 2^{-d-2})$ that

$$\|g(\cdot, t)\|_2 \ \geq \ \frac{c_{16}(d)}{t^{(d+1)/2}} \left[\|r^{(d-1)/2} \cos_d(tr)\|_2 - \frac{\rho^{(d-1)/2}}{2^{d+2}} \right].$$

Now

$$\frac{2}{\rho} \int_{\rho/2}^{\rho} r^{d-1} \cos_d^2(tr) dr \geq \frac{2}{\rho} \left(\frac{\rho}{2} \right)^{d-1} \int_{\rho/2}^{\rho} \cos_d^2(tr) dr \ = \ \frac{\rho^{d-2}}{2^{d-2}} \int_{t\rho/2}^{t\rho} \cos_d^2(x) dx / t.$$

As ρt becomes large, we're integrating $\cos^2$ over a large number of complete cycles, on each of which its average value is 1/2, and the two possibly incomplete cycles near the endpoints become negligible, so that the integral of $\cos^2$ is asymptotic to $t\rho/4$. Thus

$$\|r^{(d-1)/2} \cos_d(tr)\|_2 \geq \left(\frac{\rho}{2} \right)^{(d-1)/2} \cdot \frac{1}{2}$$

for $t\rho \geq c_{177}(d)$ if $c_{177}(d)$ is large enough, in particular $c_{177}(d) \geq 2 c_{17}(d, 2^{-d-2})$. Then

$$\|g(\cdot, t)\|_2 \ \geq \ \frac{c_{16}(d) \rho^{(d-1)/2}}{(2t)^{(d+1)/2}}.$$

Thus $h(\omega, t) \geq c_{167}(d) \omega^{d-1} t^{-d-1}$ for $\omega t \geq c_{177}(d)$ where $c_{167}(d) := c_{16}(d)^2 / 2^{d+3}$.

Let $F \asymp G$ mean that positive functions F and G are of the same order of magnitude under some condition (to be specified in each case), namely that F/G is bounded above and below by strictly positive, finite constants depending only on d.

We have (taking $\omega = \rho$ above) that $h(\rho, t) \asymp \rho^{d-1}/t^{d+1}$ for $\xi := \rho t \in I_2 := [c_{177}(d), +\infty)$. It follows that Beck's (29) holds with some $c'_{20}(d)$ for $0 < \rho \leq \omega \leq 1$ such that $t\rho \in I_2$.

Similarly, from Beck's (28) we get that $h(\rho, t) \asymp \rho^{2d}$ for $0 < \rho \leq 1$ and $\xi = \rho t \in I_0 := (0, c_{199}(d)]$ where $c_{199}(d)$ is a small enough constant depending only on d, specifically smaller than Beck's $c_{19}(d, 1/2)$. For $0 < \rho \leq \omega \leq 1$ such that $t\omega \in I_0$ we have

$$H(\omega, \rho, t) := \frac{h(\omega, t)/\omega^{d-1}}{h(\rho, t)/\rho^{d-1}} \asymp \frac{\omega^{d+1}}{\rho^{d+1}} \ \geq \ 1 \tag{27}$$

and then (29) holds for some $c''_{20}(d)$.

For the intermediate range $\xi = t\rho \in I_1 := [c_{199}(d), c_{177}(d)]$, Beck's (26), (27), and (28) are not useful but we can return to (25). With the relationship $x := rt$ (already used in effect in Beck's proof), we get

$$h(\rho, t) = \frac{2}{\xi} \int_{\xi/2}^{\xi} g^2(x/t, t)dx = \frac{c_{15}(d)^2}{t^{2d}} \cdot \frac{2}{\xi} \int_{\xi/2}^{\xi} x^d J_{d/2}(x)^2 dx,$$

where $c_{15}(d)$ is a constant depending only on d, first appearing in Beck's (25), depending on $c_{14}(d)$ in the Fourier transform of the indicator of the unit ball in his (23) and the constants in the Bessel function in the Poisson formula (24) with $k = d/2$. According to my calculations, $c_{15}(d)^2 \equiv (2\pi)^d$.

From Beck's (23), one can see that $g(r, t)$ is an analytic function of the real variables t and r. Thus from Beck's (25) one could see that for each $d = 1, 2, ...$, the Bessel function $J_{d/2}$ is an analytic function except for a possible singularity at 0. As shown in the book of Watson (1941) in the first few pages of Chapter 3, $J_{d/2}$ extends to an entire holomorphic function of a complex variable z if d is even (Watson §3.1, (8)). If d is odd then there is a singularity (branch point, $\sqrt{z}$ times a holomorphic function) at $z = 0$ (the same formula, extended by an argument on the following page). In either case, $J_{d/2}$ on the open half-line $(0, \infty)$ is real analytic. For $x \in I_1$, which is a bounded interval bounded away from 0, $J_{d/2}$ is analytic with at most finitely many zeroes. Thus $\xi \mapsto \int_{\xi/2}^{\xi} x^d J_{d/2}(x)^2 dx$ is a continuous (indeed analytic) function of ξ, bounded and bounded away from 0 for $\xi \in I_1$. It follows that for $\xi \in I_1$, $h(\rho, t) \asymp t^{-2d}$. Now note that for $\xi = t\rho \in I_1$, $t \asymp 1/\rho$. Thus we can also write $h(\rho, t) \asymp \rho^{d-1}/t^{d+1}$ for $\xi \in I_1$, the same as for $\xi \in I_2$. And so, Beck's (29) holds for some constant $c'''_{20}(d)$ whenever $t\rho \in I_1$. (Then $t\omega \in I_1$ or I_2, which one doesn't matter.)

We still need to consider the case $t\rho \in I_0$ and $t\omega \in I_1 \cup I_2$. Then

$$H(\omega, \rho, t) \asymp \frac{t^{-d-1}}{\rho^{d+1}} = \frac{1}{(t\rho)^{d+1}} \geq c_{199}(d)^{-d-1} > 0,$$

so (29) holds also in this last of the possible cases, and so it holds generally as stated, with $c_{20}(d)$ the minimum of the corresponding constants in the different cases.

7.3 A suggested proof of Beck's inequality (31)

This subsection was incorporated in one of Richard Nickl's lectures and the handout for it. The left side of inequality (31) is bounded above by

$$\sum_{\underline{i}} \int_{R^d} |\beta(G(\underline{x}, r) \cap I(l; \underline{i}))| d\underline{x}.$$

For each individual $\underline{i}$, since $I(l; \underline{i})$ is a cube of side 2^{-l} and $r \leq \rho_0 = 2^{-l-1}$, the cube $G(\underline{x}, r) \cap I(l; \underline{i})$ can be non-empty only for $\underline{x}$ in a cube of side 2^{1-l}. Thus we can bound each integral by an upper bound for the integrand times $2^{d(1-l)} = 2^{d-dl} = 2^d/(Cn)$.

Let $r \in [\rho_0/2, \rho_0]$ be fixed. For $k = 0, 1, ...$, let $I_k := \{\underline{i} : \ 2^k \leq \sup_{\underline{x}} |\beta(G(\underline{x}, r) \cap I(l; \underline{i}))| c^*(d)\sqrt{n} < 2^{k+1}\}$. Then by condition $(d, n, C, *)$ for $t = 2^k$, $\text{card}(I_k) \leq n/2^{3k/2}$. Let

$$I_{-1} := \{\underline{i} : \sup_{\underline{x}} |\beta(G(\underline{x}, r) \cap I(l; \underline{i})| < 1/(c^*(d)\sqrt{n})\}.$$

Then $\text{card}(I_{-1}) \leq 2^d$ since each ball $G(\underline{x}, r)$ intersects at most 2^d of the cubes $I(l; \underline{i})$, as Beck has just noted.

For $\underline{i} \in I_k$ with $k \geq 0$, $|\beta(G(\underline{x}, r) \cap I(l; \underline{i}))| \leq 2^{k+1}/(c^*(d)\sqrt{n})$ by definition of I_k. Therefore, considering only the terms with $k \geq 0$, we get a sum

$$2^{d(1-l)} \sum_{k=0}^{\infty} \frac{n}{2^{3k/2}} \cdot \frac{2^{k+1}}{c^*(d)\sqrt{n}}$$

which by summing a geometric series in k gives

$$\frac{2^d}{Cn} \frac{2\sqrt{n}}{c^*(d)} \cdot \frac{\sqrt{2}}{\sqrt{2}-1} = \frac{2^{d+1}\sqrt{2}}{Cc^*(d)(\sqrt{2}-1)} \cdot \frac{1}{\sqrt{n}}.$$

Including the $k = -1$ term gives a further summand $4^d/(c^*(d)Cn^{3/2}) \leq 4^d/(c^*(d)C\sqrt{n})$. Then, inserting the definition of $c^*(d)$ from (3), we get the last term of Beck's (31) as stated.

I don't see why one can omit the $k = -1$ term entirely.

8 Side issues

One interesting possibility is to use an entirely different approach, not approximating by unions of small cubes. Balls can be well approximated by other balls, although of course they don't fit together via disjoint unions unless possibly infinite unions.

Question: How does one exactly verify the inequalities given as (6) and (7) of Beck's paper, with the given constants? I was able to do it for (7) and $d = 2$. One simplifying consideration in these proofs is that decomposing a cube of side 2^{-k} into smaller cubes of side 2^{-t-1} corresponds to decomposing the unit cube into subcubes of side 2^{k-t-1}, as already mentioned in the proof above by way of volumes. So, in proving the given bounds one can assume $k = 0$.

REFERENCES

Beck, József (1985), "Lower bounds on the approximation of the multivariate empirical process," *Z. Wahrscheinlichkeitsth. verw. Geb.* **70**, 289-306.

Beck, József, and Chen, William W. L. (1987), *Irregularities of distribution*, Cambridge University Press.

Dudley, R. M. (1979), Balls in R^k do not cut all subsets of $k+2$ points, *Advances in Math.* **31**, 306-308.

Dudley, R. M. (1999), *Uniform Central Limit Theorems*, Cambridge University Press.

Radon, J. (1921), Mengen konvexer Körper, die einen gemeinsamen Punkt enthalten, *Math. Ann.* **83**, 113-115.

Watson, G. N. (1941, repr. 1952), *A Treatise on the Theory of Bessel Functions*, 2d ed. (1st ed. 1922), Cambridge University Press.

Wenocur, R., and Dudley, R. M. (1981). Some special Vapnik-Červonenkis classes, *Discrete Math.* **33**, 313-318.