1 Sudakov Minoration

Suppose that \((T, \rho)\) is a finite metric space and let \(\{X_t, t \in T\}\) be a stochastic process indexed by \(T\) satisfying

\[
P\{|X_s - X_t| \geq u\} \leq 2 \exp\left(\frac{-u^2}{2\rho^2(s, t)}\right) \quad \text{for all } u > 0.
\]  

(1)

Then for every \(t_0 \in T\), we have

\[
E \sup_{t \in T} |X_t - X_{t_0}| \leq C \int_0^\infty \sqrt{\log M(\epsilon, T)} d\epsilon.
\]  

(2)

There is also a lower bound for the above expectation in terms of packing numbers. This is called Sudakov minoration. Suppose that \(X_t\) has mean zero and that \(E(X_t - X_s)^2 = \rho^2(s, t)\) for all \(s, t \in T\), then

\[
E \sup_{t \in T} (X_t - X_{t_0}) \geq c \sup_{\epsilon > 0} \left(\epsilon \sqrt{\log N(\epsilon, T)}\right)
\]  

(3)

for a positive constant \(c\). Here \(N(\epsilon, T)\) is the \(\epsilon\)-packing number of \(T\).

The proof of Sudakov Minoration goes via Slepian’s lemma:

**Lemma 1.1** (Slepian). Let \((X_1, \ldots, X_n)\) and \((Y_1, \ldots, Y_n)\) denote multivariate normal random vectors with mean 0. Suppose that

\[E(X_i - X_j)^2 \geq E(Y_i - Y_j)^2\]  

for all \(i \neq j\).

Then \(E \max_i X_i \geq E \max_i Y_i\).

**Proof of (3).** Fix \(\epsilon > 0\). Let \(t_1, \ldots, t_n\) be a maximal \(\epsilon\)-packing subset of \(T\) where \(n = N(\epsilon, T)\). Then

\[
E \sup_{t \in T} (X_t - X_{t_0}) \geq E \sup_{1 \leq i \leq n} (X_{t_i} - X_{t_0}).
\]
Let \( g_1, \ldots, g_n \) be independent normal random variables with mean 0 and variance \( \epsilon^2/2 \). We then have
\[
E \left( (X_t - X_{t_0} - (X_{t_j} - X_{t_0}))^2 \right) = \rho^2 (t_i, t_j) \geq \epsilon^2 = E(g_i - g_j)^2
\]
for all \( i, j \). Thus by Slepian’s lemma, we have
\[
E \max_i (X_t - X_{t_0}) \geq E \max_{1 \leq i \leq n} g_i = \frac{\epsilon}{\sqrt{2}} E \max_{1 \leq i \leq n} z_i
\]
where \( z_1, \ldots, z_n \) are independent standard normal random variables. The proof is now complete by noting that the expectation of the maximum of \( n \) i.i.d standard normals is bounded from below by a constant multiple of \( \sqrt{\log n} \).

\[\square\]

2 Dudley’s Entropy Bound Revisited

In the last class, we proved (2) where we assumed that \( T \) is finite and the integral of the square root of the metric entropy is from 0. But, from the proof, it should be clear that we actually proved upper bounds for

\[P \left( \max_{t \in T_N} |X_t - X_{t_0}| \geq a \right) \leq C \exp \left( -\frac{a^2}{CD^2} \right) \quad (4)\]

whenever

\[a \geq C \int_{D/2}^{D/2 - N^{-1}} \sqrt{\log M(\epsilon, T)} \, de.\]

Let us go over the proof of (4) once again. Let us not assume that \( T \) is finite anymore. For each \( i \geq 1 \), let \( T_i \) denote a \( \delta_i = D2^{-i} \)-cover of \( T \) with diameter \( D \). We, in fact, proved

\[P \left( \max_{t \in T_N} |X_t - X_{t_0}| \geq a \right) \leq C \exp \left( -\frac{a^2}{CD^2} \right) \]

whenever

\[a \geq C \int_{D/2}^{D/2 - N^{-1}} \sqrt{\log M(\epsilon, T)} \, de.\]

Let us go over the proof of (4) once again. Let us not assume that \( T \) is finite anymore. For each \( i \geq 1 \), let \( T_i \) denote a \( \delta_i = D2^{-i} \)-cover of \( T \) with cardinality \( M(\delta_i) \). For each \( i = 1, \ldots, N \), let \( \pi : T_i \to T_{i-1} \) be such that for each \( t \in T_i \), we have \( \rho(t, \pi(t)) \leq \delta_{i-1} \). Such a map can be constructed because \( T_{i-1} \) is a \( \delta_{i-1} \)-cover of \( T \).

Now for each \( i = 0, \ldots, N \), define the map \( \gamma_i : T_N \to T_i \) by

\[\gamma_0(t) \equiv t_0; \quad \gamma_i := \pi_{i+1} \pi_{i+2} \ldots \pi_N, \quad 1 \leq i \leq N - 1 \quad \text{and} \quad \gamma_N := \text{identity}.\]

Now for each \( t \in T_N \), we can write

\[|X_t - X_{t_0}| = |X_{\gamma_N(t)} - X_{\gamma_0(t)}| \leq \sum_{i=1}^{N} |X_{\gamma_i(t)} - X_{\gamma_{i-1}(t)}|\]
It thus follows that
\[ \max_{t \in T_N} |X_t - X_{t_0}| \leq \sum_{i=1}^{N} \max_{t \in T_i} |X_{\gamma_i(t)} - X_{\gamma_{i-1}(t)}| \]

Also because \( \gamma_{i-1}(t) = \pi_i(\gamma_i(t)) \) for \( t \in T_N \), we can write
\[ \max_{t \in T_N} |X_t - X_{t_0}| \leq \sum_{i=1}^{N} \max_{t \in T_i} |X_t - X_{\pi_i(t)}| \]  \quad (5)

Now suppose \( b_1, \ldots, b_N \) be nonnegative numbers with \( \sum_{i=1}^{N} b(i) \leq 1 \). From (5), we obtain the following series of upper bounds
\[
\mathbb{P} \left( \max_{t \in T_N} |X_t - X_{t_0}| \geq a \right) \leq \mathbb{P} \left( \sum_{i=1}^{N} \max_{t \in T_i} |X_t - X_{\pi_i(t)}| \geq a \sum_{i=1}^{N} b_i \right) \\
\leq \sum_{i=1}^{N} \mathbb{P} \left( \max_{t \in T_i} |X_t - X_{\pi_i(t)}| \geq a b_i \right) \\
\leq \sum_{i=1}^{N} \sum_{t \in T_i} \mathbb{P} \left( |X_t - X_{\pi_i(t)}| \geq a b_i \right) \\
\leq \sum_{i=1}^{N} \sum_{t \in T_i} 2 \exp \left( -\frac{a^2 b_i^2}{2 \delta_{i-1}^2 (t, \pi_i(t))} \right) \\
\leq 2 \sum_{i=1}^{N} M(\delta_i) \exp \left( -\frac{a^2 b_i^2}{2 \delta_{i-1}^2} \right) \\
\leq 2 \sum_{i=1}^{N} \exp \left( \log M(\delta_i) - \frac{a^2 b_i^2}{2 \delta_{i-1}^2} \right). \quad (6)
\]

Suppose that \( \beta_1, \ldots, \beta_N \) are fixed nonnegative numbers with \( \beta_1 + \cdots + \beta_N \leq 1/2 \) and let
\[ b_i = \max \left( \beta_i, \frac{2 \delta_{i-1} \sqrt{\log M(\delta_i)}}{a} \right) \]

We can use this sequence \( b_1, \ldots, b_N \) provided \( \sum_{i=1}^{N} b_i \leq 1 \). Because \( \sum_{i=1}^{N} \beta_i \leq 1/2 \), it is enough to ensure that
\[ a \geq 4 \sum_{i=1}^{N} \delta_{i-1} \sqrt{\log M(\delta_i)}. \]
This condition is satisfied if \( a \) is greater than a constant multiple of \( \int \sqrt{\log M(\epsilon)} d\epsilon \) because

\[
\sum_{i=1}^{N} \delta_{i-1} \sqrt{\log M(\delta_i)} = D \sqrt{\log M(D/2)} + \frac{D}{2} \sqrt{\log M(D/4)} + \cdots + \frac{D}{2^{N-1}} \sqrt{\log M(D2^{-N})}
\]

\[
\leq 4 \left[ \int_{D/4}^{D/2} \sqrt{\log M(\epsilon)} d\epsilon + \int_{D/8}^{D/4} \sqrt{\log M(\epsilon)} d\epsilon + \cdots + \int_{D2^{-N-1}}^{D2^{-N}} \sqrt{\log M(\epsilon)} d\epsilon \right]
\]

\[
\leq 4 \int_{D2^{-N-1}}^{D2^{-N}} \sqrt{\log M(\epsilon)} d\epsilon.
\]

We have thus shown that for \( a \) satisfying

\[
a \geq 16 \int_{D2^{-N-1}}^{D/2} \sqrt{\log M(\epsilon)} d\epsilon,
\]

we have

\[
P \left( \max_{t \in T} |X_t - X_{t_0}| \geq a \right) \leq 2 \sum_{i=1}^{N} \exp \left( -\frac{a^2 b_i}{2 \delta_{i-1}^2} \right) = 2 \sum_{i=1}^{N} \exp \left( -\frac{2^{2i} a^2 b_i^2}{8D^2} \right)
\]

To convert this into an exponential bound in \( a \), we use the fact that \( b_i \geq \beta_i \) and \( \beta_i \) is an arbitrary nonnegative sequence adding up to \( \leq 1/2 \). We get

\[
P \left( \max_{t \in T} |X_t - X_{t_0}| \geq a \right) \leq 2 \sum_{i=1}^{N} \exp \left( -\frac{2^{2i} a^2 \beta_i^2}{8D^2} \right)
\]

It makes sense to choose \( \beta_i \) in a way that the right hand side above can be summed. Taking \( \beta_i = C 2^{-i} \sqrt{i} \) where \( C \) is chosen so that \( \sum_i \beta_i \leq 1/2 \), we get

\[
P \left( \max_{t \in T} |X_t - X_{t_0}| \geq a \right) \leq 2 \sum_{i=1}^{N} \exp \left( -\frac{2^{2i} a^2 \beta_i^2}{8D^2} \right)
\]

\[
\leq 2 \sum_{i=1}^{\infty} \exp \left( -\frac{C^2 a^2 i}{8D^2} \right)
\]

\[
\leq 2 \exp \left( -\frac{C^2 a^2}{8D^2} \right)
\]

\[
= \frac{2 \exp \left( -\frac{C^2 a^2}{8D^2} \right)}{1 - \exp \left( -\frac{C^2 a^2}{8D^2} \right)}
\]

The denominator is bounded away from 0 because the condition (7) implies that

\[
a \geq 16 \int_{D2^{-N-1}}^{D/2} \sqrt{\log M(\epsilon)} d\epsilon \geq 16 \int_{D/4}^{3D/8} \sqrt{\log M(\epsilon)} d\epsilon \geq 2 \sqrt{\log 2D}.
\]
We have thus shown that for \( a \) satisfying (7), we have the tail bound
\[
P \left( \sup_{t \in T} |X_t - X_{t_0}| \geq a \right) \leq \frac{2}{1 - 2^{-C^2/2}} \exp \left( \frac{-a^2}{CD^2} \right)
\]
for a universal constant \( C \).

3 Special Case

The rest of the stuff in today’s lecture is from Van de Geer’s book.

Fix \( n \geq 1 \) and let \( W_1, \ldots, W_n \) be independent standard normal variables. Let \( \mathcal{X} \) denote a non-empty set and let \( z_1, \ldots, z_n \) be fixed points in \( \mathcal{X} \). Suppose \( \mathcal{G} \) denote a class of real-valued functions on \( \mathcal{X} \). Define the (pseudo)metric \( \rho \) on \( \mathcal{G} \) by
\[
\rho(g, h) := \left( \frac{1}{n} \sum_{i=1}^{n} (g(z_i) - h(z_i))^2 \right)^{1/2}
\]
and let \( D \) denote the diameter of \( \mathcal{G} \) under \( \rho \). For each \( g \in \mathcal{G} \), define the random variable
\[
X_g := \frac{\sum_{i=1}^{n} W_i g(z_i)}{\sqrt{n}}.
\]
For \( g, h \in \mathcal{G} \), it is clear that \( X_g - X_h \) is normally distributed with mean 0 and variance \( \rho^2(g, h) \). We therefore have
\[
P \{ |X_g - X_h| \geq x \} \leq 2 \exp \left( \frac{-x^2}{2\rho^2(g, h)} \right) \quad \text{for } x > 0
\]
We want an upper bound for
\[
P \left( \sup_{g \in \mathcal{G}} |X_g - X_{g_0}| \geq a \right).
\]
Note that we are not assuming that \( \mathcal{G} \) is finite. Let \( \mathcal{G}_i \) be a \( \delta_i = D2^{-i} \) be a \( \delta_i \)-cover for \( \mathcal{G} \). From (4), we know how to bound
\[
P \left( \sup_{g \in \mathcal{G}_N} |X_g - X_{g_0}| \geq a \right).
\]
We can use Cauchy-Schwarz inequality to bound \( \sup_{g \in \mathcal{G}} |X_g - X_{g_0}| \) in terms of \( \sup_{g \in \mathcal{G}_N} |X_g - X_{g_0}| \). Indeed, fix \( g \in \mathcal{G} \) and let \( h \in \mathcal{G}_N \) be such that \( \rho(g, h) \leq \delta_N \).
Clearly
\[ |X_g - X_{g_0}| \leq |X_h - X_{g_0}| + |X_g - X_h| \]
\[ \leq |X_h - X_{g_0}| + \frac{1}{\sqrt{n}} \left| \sum_{i=1}^{n} W_i(g(z_i) - h(z_i)) \right| \]
\[ \leq |X_h - X_{g_0}| + \rho(g, h) \sqrt{\sum_{i=1}^{n} W_i^2} \]
\[ \leq |X_h - X_{g_0}| + \delta N \sqrt{\sum_{i=1}^{n} W_i^2}. \]

It therefore follows that
\[ \sup_{g \in \mathcal{G}} |X_g - X_{g_0}| \leq \sup_{g \in \mathcal{G}} |X_g - X_{g_0}| + \delta N \sqrt{\sum_{i=1}^{n} W_i^2}. \]

Thus when
\[ \frac{1}{n} \sum_{i=1}^{n} W_i^2 \leq \tau^2, \]
we have
\[ \sup_{g \in \mathcal{G}} |X_g - X_{g_0}| \leq \sup_{g \in \mathcal{G}} |X_g - X_{g_0}| + \delta N \tau \sqrt{n}. \]

As a result,
\[ \mathbb{P} \left( \sup_{g \in \mathcal{G}} |X_g - X_{g_0}| \geq a, \frac{1}{n} \sum_{i=1}^{n} W_i^2 \leq \tau^2 \right) \leq \mathbb{P} \left( \sup_{g \in \mathcal{G}} |X_g - X_{g_0}| \geq a - \delta N \tau \sqrt{n} \right). \]

Dudley’s bound (4) now gives
\[ \mathbb{P} \left( \sup_{g \in \mathcal{G}} |X_g - X_{g_0}| \geq a, \frac{1}{n} \sum_{i=1}^{n} W_i^2 \leq \tau^2 \right) \leq C \exp \left( -\frac{(a - \delta N \tau \sqrt{n})^2}{C \beta^2} \right) \quad (8) \]
provided
\[ a \geq \delta N \tau \sqrt{n} + C \int_{D/2-N-1}^{D/2} \sqrt{\log M(\epsilon, \mathcal{G})} d\epsilon. \]

We now need to choose \( N \). Because we need an upper bound for the right hand side of (8) that decreases exponentially in \( a \), it is reasonable to take \( a \geq 2\delta \tau \sqrt{n} \). Subject to this constraint, it is sensible to take \( N \) as small as possible because then the entropy integral will be small. We thus take
\[ N := \min \{ i \geq 1 : a \geq 2\delta_i \tau \sqrt{n} \}. \]
Note here that, by the Cauchy-Schwarz inequality, \( \sup_{g \in G} |X_g - X_{g_0}| \leq D \tau \sqrt{n} \) on the set \{ \( \sum W_i^2 / n \leq \tau^2 \) \}.

For this choice of \( N \), we have \( a \geq 2 \delta_N \tau \sqrt{n} \) and
\[
a < 2 \delta_{N-1} \tau \sqrt{n} = 8D^{-N-1} \tau \sqrt{n}.
\]
Thus
\[
\delta_N \tau \sqrt{n} + C \int_{D^{-2N-1}/2}^{D/2} \sqrt{\log M(\epsilon, \mathcal{G})} d\epsilon \leq \frac{a}{2} + C \int_{a/(8\tau \sqrt{n})}^{D/2} \sqrt{\log M(\epsilon, \mathcal{G})}.
\]
We have thus shown that
\[
P \left( \sup_{g \in \mathcal{G}} |X_g - X_{g_0}| \geq a, \frac{1}{n} \sum_{i=1}^{n} W_i^2 \leq \tau^2 \right) \leq C \exp \left( - \frac{a^2}{CD^2} \right)
\] (9)
provided
\[
a \geq C \int_{a/(8\tau \sqrt{n})}^{D/2} \sqrt{\log M(\epsilon, \mathcal{G})}.
\] (10)
Note the presence of \( a \) in the right hand side of the above inequality as well.

4 Convergence Rates of Least Squares Estimators

Consider the regression model: \( Y_i = g_0(z_i) + W_i \) for \( i = 1, \ldots, n \) where \( Y_1, \ldots, Y_n \) are real-valued observations, \( g_0 \in \mathcal{G} \) is an unknown regression function, \( z_1, \ldots, z_n \) are given covariates in a space \( \mathcal{Z} \) and \( W_1, \ldots, W_n \) are independent standard normal random variables.

A least squares estimator for \( g_0 \) satisfies
\[
\hat{g}_n := \arg\min_{g \in \mathcal{G}} \sum_{i=1}^{n} (Y_i - g(z_i))^2.
\]

We would like to obtain convergence rates for \( \hat{g}_n \) as an estimator for \( g_0 \). The important property of \( \hat{g}_n \) is:
\[
\sum_{i=1}^{n} (Y_i - \hat{g}_n(z_i))^2 \leq \sum_{i=1}^{n} (Y_i - g_0(z_i))^2.
\]

Writing \( Y_i = g_0(z_i) + W_i \) and simplifying, we obtain
\[
\sum_{i=1}^{n} (\hat{g}_n(z_i) - g_0(z_i))^2 \leq 2 \sum_{i=1}^{n} W_i (\hat{g}_n(z_i) - g_0(z_i)).
\] (11)
Under the notation:

\[
(W, \hat{g}_n - g_0)_n := \frac{1}{n} \sum_{i=1}^{n} W_i (\hat{g}_n(z_i) - g_0(z_i)),
\]

we can rewrite (11) as

\[
\rho^2(\hat{g}_n, g_0) \leq 2 \langle W, \hat{g}_n - g_0 \rangle.
\]

This inequality is referred to as the Basic Inequality by Van de Geer.

We now prove an upper bound for

\[
\begin{align*}
\star = \mathbb{P} \left\{ \rho(\hat{g}_n, g_0) > \delta, \frac{1}{n} \sum_{i=1}^{n} W_i^2 \leq \tau^2 \right\} \leq \sum_{s \geq 0} \mathbb{P} \left\{ 2^s \delta < \rho(\hat{g}_n, g_0) \leq 2^{s+1} \delta, ||W||_n \leq \tau \right\}. 
\end{align*}
\]

Note that, when \( ||W||_n \leq \tau \), the basic inequality (13) implies (via the Cauchy-Schwarz inequality) that

\[
\rho^2(\hat{g}_n, g_0) \leq 2 \langle W, \hat{g}_n - g_0 \rangle_n \leq 2 \|W\|_n \rho(\hat{g}_n, g_0) \leq 2\tau \rho(\hat{g}_n, g_0)
\]
and hence \( \rho(\hat{g}_n, g_0) \leq 2\tau \) whenever \( ||W||_n \leq \tau \). It follows that there are only a finite number of terms in the right hand side of (14). In fact, the terms in the sum equal zero when \( 2^s \delta > 2\tau \).

By the basic inequality,

\[
\star \leq \sum_{s \geq 0} \mathbb{P} \left\{ (W, \hat{g}_n - g_0)_n > 2^{2s-1} \delta^2, \rho(\hat{g}_n, g_0) \leq 2^{s+1} \delta, ||W||_n \leq \tau \right\}.
\]

For \( \eta > 0 \), if we denote by \( B(g_0, \eta) \) the closed ball \( \{ g \in \mathcal{G} : \rho(g, g_0) \leq \eta \} \), then

\[
\star \leq \sum_{s \geq 0} P_s \text{ where } P_s = \mathbb{P} \left( \sup_{g \in B(g_0, 2^{s+1} \delta)} \sum_{i=1}^{n} W_i (g(z_i) - g_0(z_i)) > n \ 2^{2s-1} \delta^2, ||W||_n \leq \tau \right).
\]

We now apply (9) with \( a = \sqrt{n}2^{2s-1} \delta^2, \mathcal{G} = B(g_0, 2^{s+1} \delta) \) so that \( D = 2^{s+2} \delta \) to obtain

\[
P_s \leq C \exp \left( \frac{-n2^{4s-2} \delta^4}{C2^{2s+4} \delta^2} \right) = C \exp \left( \frac{-n2^{2s} \delta^2}{C} \right)
\]
provided

\[
\sqrt{n}2^{2s} \delta^2 \geq C \int_{2^{2s} \delta^2 / (16\tau)}^{2^{s+1} \delta} \sqrt{\log M(\epsilon, B(g_0, 2^{s+1} \delta))} \, d\epsilon.
\]

(15)
If (15) is satisfied for all \( s \) such that \( 2s\delta \leq 2\tau \), then
\[
\star \leq C \sum_s \exp \left( -\frac{n2^{2s}\delta^2}{C} \right). \tag{16}
\]
Let us now give a sufficient condition on \( \delta \) so that (15) is satisfied for all \( s \) such that \( 2s\delta \leq 2\tau \). Let the function \( \Psi \) be such that
\[
\Psi(u) \geq \int_{u^2/(2^\delta \tau)}^u \sqrt{\log M(\epsilon, B(g_0, u), \epsilon)} d\epsilon \tag{17}
\]
for \( 0 < u \leq 4\tau \) (we may assume that \( \Psi \) is only defined for \( 0 < u \leq 4\tau \)). Then, in order to ensure (15), we need \( \delta > 0 \) to be such that \( \sqrt{n}u^2 \geq C\Psi(u) \) for all \( u = 2^{s+1}\delta \leq 4\tau \).

Suppose that we assume that \( \Psi(u)/u^2 \) is decreasing on \( (0, 4\tau] \). Then, if \( \sqrt{n}u^2 \geq C\Psi(u) \) holds for \( u = \delta \), then it holds for all \( u = 2^{s+1}\delta \leq 4\tau \). This is because
\[
\frac{\Psi(2^{s+1}\delta)}{2^{s+2}\delta^2} \leq \frac{\Psi(\delta)}{\delta^2} \leq \frac{\sqrt{n}}{C}.
\]
We have thus shown that
\[
P \{ \rho(\hat{g}_n, g_0) > \delta, ||W||_n \leq \tau \} \leq C \sum_s \exp \left( -\frac{n2^{2s}\delta^2}{C} \right) \tag{18}
\]
provided
\[
\sqrt{n}\delta^2 \geq C\Psi(\delta) \tag{19}
\]
where \( \Psi \) is a function defined on \( (0, 4\tau] \) such that (17) is true and such that \( \Psi(u)/u^2 \) is decreasing on \( (0, 4\tau] \). Finally, suppose that it is also true that \( \sqrt{n}\delta \geq B \) for a constant \( B \), then note that the right hand of the above inequality is bounded from above by a constant times \( \exp(-n\delta^2/C) \).

**Example 4.1 (Finite dimensional class).** Suppose that the class of functions \( \mathcal{G} \) is finite dimensional with dimension \( d \). Then it is possible to show that
\[
M(B(g_0, u), \epsilon) \leq \left( \frac{Cu}{\epsilon} \right)^d.
\]
Then
\[
\int_{u^2/(2^\delta \tau)}^u \sqrt{\log M(\epsilon, B(g_0, u), \epsilon)} d\epsilon \leq \int_0^u \sqrt{d \log \frac{u}{\epsilon}} d\epsilon \leq u\sqrt{d} \int_1^\infty \frac{\sqrt{\log(Ct)}}{t^2} dt = Cuv\sqrt{d}.
\]
We can thus take \( \Psi(u) = Cu\sqrt{d} \) and then \( \delta = C\sqrt{d/n} \) would satisfy \( \sqrt{n}\delta^2 \geq C\psi(\delta) \). We thus have
\[
P \{ \rho(\hat{g}_n, g_0) > \delta, ||W||_n \leq \tau \} \leq \exp \left( -\frac{n\delta^2}{C} \right),
\]
whenever $\delta \geq C\sqrt{d/n}$. The right hand side does not depend on $\tau$ so we may let $\tau \to \infty$ on the left hand. This will lead to the risk bound $\mathbb{E}\rho(\hat{g}_n, g_0) \leq C\sqrt{d/n}$. 