Persistence: Alternative proofs of some results of Greenshtein and Ritov

Jon A. Wellner

University of Washington

- Talk at Newton Institute, Cambridge, 10 January, 2008
- *Email: jaw@stat.washington.edu*
  http: //www.stat.washington.edu/jaw/jaw.research.html

Outline

1. Introduction and framework: persistence
2. A theorem of Greenshtein and Ritov
3. First Proof – via Nemirovski's inequality
4. Second Proof – via bracketing entropy bounds
5. Proof of Nemirovski's inequality
6. Summary; problems and open questions

1. Introduction and framework: persistence

Setting:

- Data: n i.i.d. copies $Z_1, \ldots, Z_n$ of $Z = (Y, X_1, \ldots, X_p) \equiv (Y, \underline{X})$; write $Z_i = (Y^i, X_1^i, \ldots, X_p^i)$, $i = 1, \ldots, n$.
- Dimension of $\underline{X}$, $p = p_n$ large, $p_n = n^\alpha$, $\alpha > 1$
- Goal: Predict Y on the basis of the covariates X_j, $j = 1, \ldots, p$
- Predictors $\widehat{Y}$ of Y of the form $\widehat{Y} = \sum_{j=1}^p \beta_j X_j = \underline{\beta}' \underline{X}$ with $\underline{\beta} \in B_n \subset \mathbb{R}^p$ for each n.
- Natural sets B_n to consider are

$$B_{n,k} \equiv \{\beta \in \mathbb{R}^p : \ \#\{j : \ \beta_j \neq 0\} = k\} = \{\beta \in \mathbb{R}^p : \ \|\beta\|_0 = k\},$$
$$B_{n,b} \equiv \{\beta \in \mathbb{R}^p : \ \|\underline{\beta}\|_1 \leq b\}.$$

where $k = k_n \to \infty$ and $b = b_n \to \infty$.

- For $Z = (Y, \underline{X}) \sim P$ on $(\mathbb{R}^{p+1}, \mathcal{B}_{p+1})$, define

$$L_P(\beta) = E_P \left(Y - \sum_{j=1}^{p} \beta_j X_j \right)^2 .$$

- For $Z = (Y, \underline{X}) \sim P$ on $(\mathbb{R}^{p+1}, \mathcal{B}_{p+1})$, define

$$L_P(\beta) = E_P \left(Y - \sum_{j=1}^{p} \beta_j X_j \right)^2 .$$

- For a given sequence of distributions $\{P_n\}$ of Z and sequence of sets $\{B_n\}$ with $B_n \subset \mathbb{R}^p$, define

$$\beta_n^*(P_n) \equiv \beta_n^* \equiv \text{argmin}_{\beta \in B_n} L_{P_n}(\beta).$$

  Thus β_n^* is a deterministic sequence in $\mathbb{R}^p$ determined by P_n and B_n.

- For $Z = (Y, \underline{X}) \sim P$ on $(\mathbb{R}^{p+1}, \mathcal{B}_{p+1})$, define

$$L_P(\beta) = E_P \left(Y - \sum_{j=1}^{p} \beta_j X_j \right)^2 .$$

- For a given sequence of distributions $\{P_n\}$ of Z and sequence of sets $\{B_n\}$ with $B_n \subset \mathbb{R}^p$, define

$$\beta_n^*(P_n) \equiv \beta_n^* \equiv \text{argmin}_{\beta \in B_n} L_{P_n}(\beta).$$

  Thus β_n^* is a deterministic sequence in $\mathbb{R}^p$ determined by P_n and B_n.

- This corresponds to the unknown "ideal predictor" $\widehat{Y}^* = \underline{\beta}_n^* \underline{X}$ which would be available to us if we knew P_n.

- **Definition.** (Greenshtein and Ritov, 2004).
  Given a set of possible predictors B_n, a sequence of procedures $\{\hat{\beta}_n\}$ is persistent (or *persistent relative to* $\{B_n\}$ and $\{\mathcal{P}_n\}$) if, for every sequence $P_n \in \mathcal{P}_n$

$$L_{P_n}(\widehat{\beta}_n) - L_{P_n}(\beta_n^*) \rightarrow_p 0.$$

2. A theorem of Greenshtein and Ritov

Theorem. If $p = p_n = n^\alpha$ and

$$F(Z_i) \equiv \max_{0 \le j,k \le p} |X_j^i X_k^i - E_{P_n}(X_j^i X_j^i)|$$

satisfies $E_{P_n} F^2(Z_1) \le M < \infty$ for all $n \ge 1$, then for $b_n = o((n/\log n)^{1/4})$ the procedures given by

$$\widehat{\beta}_n \equiv \text{argmin}_{\beta \in B_{n,b_n}} L_{\mathbb{P}_n}(\beta) \tag{1}$$

are persistent with respect to

$$B_{n,b_n} \equiv \{\beta \in \mathbb{R}^p : \ \|\beta\|_1 \le b_n\}.$$

- **Comment 1.** The persistent procedures $\widehat{\beta}_n$ in (1) are equivalent to Lasso estimators with a particular range of the penalty parameters.

- **Comment 1.** The persistent procedures $\widehat{\beta}_n$ in (1) are equivalent to Lasso estimators with a particular range of the penalty parameters.
- **Comment 2.** Greenshtein and Ritov (2004) also prove related results for procedures based on the “model selection sets” $B_{n,k}$ under the assumption that $k = k_n = o((n/logn)^{1/2})$.

- **Comment 1.** The persistent procedures $\widehat{\beta}_n$ in (1) are equivalent to Lasso estimators with a particular range of the penalty parameters.
- **Comment 2.** Greenshtein and Ritov (2004) also prove related results for procedures based on the "model selection sets" $B_{n,k}$ under the assumption that $k = k_n = o((n/logn)^{1/2})$.
- **Proof, part 1:** Let $\gamma' = (-1, \beta_1, \ldots, \beta_p)' \equiv (\beta_0, \ldots, \beta_p)' \in \mathbb{R}^{p+1}$, and let $Y \equiv X_0$. Then

$$L_P(\beta) = E_P(Y - \beta' X)^2 = \gamma' \Sigma_P \gamma$$

where $\Sigma_P \equiv (\sigma_{ij}) = (E_P(X_i X_j))_{0 \leq i,j \leq p}$.

- **Proof, part 1, continued:** Let $\mathbb{P}_n$ be the empirical measure of $Z_1, \ldots, Z_n$. Then

$$L_{\mathbb{P}_n}(\beta) = \gamma' \Sigma_{\mathbb{P}_n} \gamma \equiv \gamma' (\widehat{\sigma}_{ij}) \gamma \equiv \gamma' \widehat{\Sigma} \gamma.$$

Define ϵ_{ij}^n and $E = (\epsilon_{ij}^n)$ by

$$\epsilon_{ij}^n \equiv \widehat{\sigma}_{ij} - \sigma_{ij}, \qquad E \equiv (\epsilon_{ij}^n) \equiv \widehat{\Sigma} - \Sigma_P.$$

Then

$$|L_{\mathbb{P}_n}(\beta) - L_{P_n}(\beta)| = |\gamma' (\Sigma_{\mathbb{P}_n} - \Sigma_{P_n}) \gamma| \le \|\Sigma_{\mathbb{P}_n} - \Sigma_{P_n}\|_\infty \|\gamma\|_1^2.$$

- **Proof, part 1, continued:** Thus for $B_{n,b_n} = \{\beta \in \mathbb{R}^p : \ \|\beta\|_1 \le b_n\}$,

$$
\begin{aligned}
&Pr\left(\sup_{\beta \in B_{n,b_n}} |L_{\mathbb{P}_n}(\beta) - L_{P_n}(\beta)| > \epsilon \right) && (2)\\
&\le \quad Pr(\|\Sigma_{\mathbb{P}_n} - \Sigma_{P_n}\|_\infty (1+b_n)^2 > \epsilon) \\
&\le \quad \epsilon^{-1}(b_n+1)^2 E\|\Sigma_{\mathbb{P}_n} - \Sigma_{P_n}\|_\infty . && (3)
\end{aligned}
$$

Thus if we can show that the expectation in the last display satisfies

$$
E\|\Sigma_{\mathbb{P}_n} - \Sigma_{P_n}\|_\infty \le C\sqrt{\frac{\log n}{n}},
$$

then the proof is complete:

- **Proof, part 1, continued:** With $\widehat{\beta}_n \equiv \text{argmin}_{\beta \in B_{n,b_n}} L_{\mathbb{P}_n}(\beta)$ it follows that

$$L_{P_n}(\widehat{\beta}_n) - L_{P_n}(\beta_n^*) \ge 0,$$
$$L_{\mathbb{P}_n}(\widehat{\beta}_n) - L_{\mathbb{P}_n}(\beta_n^*) \le 0,$$

and hence

$$\begin{aligned} 0 &\le L_{P_n}(\widehat{\beta}_n) - L_{P_n}(\beta_n^*) \\ &= L_{P_n}(\widehat{\beta}_n) - L_{\mathbb{P}_n}(\widehat{\beta}_n) + L_{\mathbb{P}_n}(\widehat{\beta}_n) - L_{\mathbb{P}_n}(\beta_n^*) \\ &\qquad + L_{\mathbb{P}_n}(\beta_n^*) - L_{P_n}(\beta_n^*) \\ &\le 2 \sup_{\beta \in B_{n,b_n}} |L_{\mathbb{P}_n}(\beta) - L_{P_n}(\beta)| \to_p 0. \end{aligned}$$

3. First proof (part 2) – via Nemirovski's inequality

Lemma 1. (Nemirovski's inequality)
Let $X_1, \ldots, X_n$ be independent random vectors in $\mathbb{R}^d$, $d \geq 3$, with $EX_i = 0$ and $E\|X_i\|_2^2 < \infty$. Then for every $r \in [2, \infty]$

$$E\Big\| \sum_{i=1}^n X_i \Big\|_r^2 \leq \widetilde{C} \min\{r, \log d\} \sum_{i=1}^n E\|X_i\|_r^2$$

where $\| \cdot \|_r$ is the ℓ_r norm, $\|x\|_r \equiv \{\sum_1^d |x_j|^r\}^{1/r}$ and $\widetilde{C}$ is an absolute constant (i.e. not depending on r or d or n or the distribution of the X_i's).

- **First proof, part 2:** To apply Nemirovski's inequality to bound $E\|\Sigma_{\mathbb{P}_n} - \Sigma_{P_n}\|_\infty$, consider the matrix $\Sigma_{\mathbb{P}_n} - \Sigma_{P_n}$ as a $(p+1)^2-$dimensional vector, and write

$$
\begin{aligned}
\Sigma_{\mathbb{P}_n} - \Sigma_{P_n} &= \sum_{i=1}^n V_i \\
&\equiv \sum_{i=1}^n \frac{1}{n}\Big(X_0^i X_0^i - E(X_0^i X_0^i), X_0^i X_1^i - E(X_0^i X_1^i), \ldots, \\
&\qquad \ldots, X_p^i X_p^i - E(X_p^i X_p^i)\Big) .
\end{aligned}
$$

- **First proof, part 2:** To apply Nemirovski's inequality to bound $E\|\Sigma_{\mathbb{P}_n} - \Sigma_{P_n}\|_\infty$, consider the matrix $\Sigma_{\mathbb{P}_n} - \Sigma_{P_n}$ as a $(p+1)^2-$dimensional vector, and write

$$
\begin{aligned}
\Sigma_{\mathbb{P}_n} - \Sigma_{P_n} &= \sum_{i=1}^n V_i \\
&\equiv \sum_{i=1}^n \frac{1}{n}\Big(X_0^i X_0^i - E(X_0^i X_0^i), X_0^i X_1^i - E(X_0^i X_1^i), \ldots, \\
&\qquad \ldots, X_p^i X_p^i - E(X_p^i X_p^i)\Big) .
\end{aligned}
$$

- By our hypothesis

$$
F(Z_i) \equiv \max_{0 \le j,k \le p} |X_j^i X_k^i - E_{P_n}(X_j^i X_k^i)|
$$

satisfies $E_{P_n} F(Z_i)^2 \le M < \infty$.

- **First proof, part 2, continued:** Then by Jensen's inequality followed by Nemirovski's inequality with $r = \infty$,

$$\begin{aligned}
\{E_{P_n}\|\Sigma_{\mathbb{P}_n} - \Sigma_{P_n}\|_\infty\}^2 &= \left\{E_{P_n}\|\sum_{i=1}^n V_i\|_\infty\right\}^2 \le E_{P_n}\|\sum_{i=1}^n V_i\|_\infty^2 \\
&\le C\log((p_n+1)^2)\sum_{i=1}^n E_{P_n}\|V_i\|_\infty^2 \\
&\le C'\log(4n^{2\alpha})\frac{1}{n^2}\sum_{i=1}^n EF(Z_i)^2 \\
&\le C''\frac{\log n}{n},
\end{aligned}$$

so that

$$E_{P_n}\|\Sigma_{\mathbb{P}_n} - \Sigma_{P_n}\|_\infty \le C''\sqrt{\frac{\log n}{n}}. \qquad \square$$

4. Proof (part 2) – via bracketing entropy bounds

- Let $\mathbb{G}_n \equiv \sqrt{n}(\mathbb{P}_n - P_n)$.

4. Proof (part 2) – via bracketing entropy bounds

- Let $\mathbb{G}_n \equiv \sqrt{n}(\mathbb{P}_n - P_n)$.
- For a class of functions $\mathcal{F} = \{f : \mathcal{Z} \to \mathbb{R}\}$ write $\|\mathbb{G}_n\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} |\mathbb{G}_n(f)|$. For $\mathcal{F}$ with $\#(\mathcal{F}) = d < \infty$, note that $\|\mathbb{G}_n\|_{\mathcal{F}} = \|\mathbb{G}_n(\underline{f})\|_\infty$ where $\mathbb{G}_n(\underline{f}) \equiv (\mathbb{G}_n(f_1), \ldots, \mathbb{G}_n(f_d))$.

4. Proof (part 2) – via bracketing entropy bounds

- Let $\mathbb{G}_n \equiv \sqrt{n}(\mathbb{P}_n - P_n)$.
- For a class of functions $\mathcal{F} = \{f : \mathcal{Z} \to \mathbb{R}\}$ write $\|\mathbb{G}_n\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |\mathbb{G}_n(f)|$. For $\mathcal{F}$ with $\#(\mathcal{F}) = d < \infty$, note that $\|\mathbb{G}_n\|_{\mathcal{F}} = \|\mathbb{G}_n(\underline{f})\|_\infty$ where $\mathbb{G}_n(\underline{f}) \equiv (\mathbb{G}_n(f_1), \ldots, \mathbb{G}_n(f_d))$.
- For each $\epsilon > 0$ let the bracketing number $N_{[\,]}(\epsilon, \mathcal{F}, L_2(P))$ be the minimal number of brackets of $L_2(P)-$size ϵ needed to cover $\mathcal{F}$.

4. Proof (part 2) – via bracketing entropy bounds

- Let $\mathbb{G}_n \equiv \sqrt{n}(\mathbb{P}_n - P_n)$.
- For a class of functions $\mathcal{F} = \{f : \mathcal{Z} \to \mathbb{R}\}$ write $\|\mathbb{G}_n\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |\mathbb{G}_n(f)|$. For $\mathcal{F}$ with $\#(\mathcal{F}) = d < \infty$, note that $\|\mathbb{G}_n\|_{\mathcal{F}} = \|\mathbb{G}_n(\underline{f})\|_\infty$ where $\mathbb{G}_n(\underline{f}) \equiv (\mathbb{G}_n(f_1), \ldots, \mathbb{G}_n(f_d))$.
- For each $\epsilon > 0$ let the bracketing number $N_{[\,]}(\epsilon, \mathcal{F}, L_2(P))$ be the minimal number of brackets of $L_2(P)-$size ϵ needed to cover $\mathcal{F}$.
- For $\delta > 0$, let

$$J_{[\,]}(\delta, \mathcal{F}, L_2(P)) \equiv \int_0^\delta \sqrt{\log(1 + N_{[\,]}(\epsilon, \mathcal{F}, L_2(P))} d\epsilon.$$

Lemma. (Empirical process theory bracketing entropy bound)

$$E^* \|\mathbb{G}_n\|_{\mathcal{F}} \lesssim J_{[\,]}(1, \mathcal{F}, L_2(P_n)) \|F\|_{P_n,2}.$$

Proof. Pollard (1989); see Theorem 2.14.2, van der Vaart and Wellner (1996), page 240. □

- In the current application take $\mathcal{F} = \{f_{j,k}(z) = x_j x_k,\ 0 \le j, k \le p\}$, a finite list of functions of cardinality $\#(\mathcal{F}) = (p_n + 1)^2$.

Lemma. (Empirical process theory bracketing entropy bound)

$$E^*\|\mathbb{G}_n\|_{\mathcal{F}} \lesssim J_{[\,]}(1, \mathcal{F}, L_2(P_n))\|F\|_{P_n,2}.$$

Proof. Pollard (1989); see Theorem 2.14.2, van der Vaart and Wellner (1996), page 240. $\square$

- In the current application take $\mathcal{F} = \{f_{j,k}(z) = x_j x_k,\ 0 \le j, k \le p\}$, a finite list of functions of cardinality $\#(\mathcal{F}) = (p_n + 1)^2$.
- Hence $N_{[\,]}(\epsilon, \mathcal{F}, L_2(P_n)) \le (p_n + 1)^2$ by choosing $\epsilon-$brackets $[l_{j,k}, u_{j,k}]$ given by $l_{j,k}(z) = f_{j,k}(z) - \epsilon/2$ and $u_{j,k}(z) = f_{j,k}(z) + \epsilon/2$.

Lemma. (Empirical process theory bracketing entropy bound)

$$E^* \|\mathbb{G}_n\|_{\mathcal{F}} \lesssim J_{[\,]}(1, \mathcal{F}, L_2(P_n)) \|F\|_{P_n,2}.$$

Proof. Pollard (1989); see Theorem 2.14.2, van der Vaart and Wellner (1996), page 240. $\square$

- In the current application take $\mathcal{F} = \{f_{j,k}(z) = x_j x_k,\ 0 \le j, k \le p\}$, a finite list of functions of cardinality $\#(\mathcal{F}) = (p_n + 1)^2$.
- Hence $N_{[\,]}(\epsilon, \mathcal{F}, L_2(P_n)) \le (p_n + 1)^2$ by choosing $\epsilon-$brackets $[l_{j,k}, u_{j,k}]$ given by $l_{j,k}(z) = f_{j,k}(z) - \epsilon/2$ and $u_{j,k}(z) = f_{j,k}(z) + \epsilon/2$.
- Thus the bound in the lemma becomes

$$E\|\mathbb{G}_n\|_{\mathcal{F}} \lesssim \sqrt{1 + \log\left[(p_n + 1)^2\right]} \|F\|_{P_n,2} \lesssim \sqrt{\log n},$$

- Or, equivalently

$$E\|\Sigma_{\mathbb{P}_n} - \Sigma_{P_n}\|_\infty = E\|\mathbb{P}_n - P_n\|_{\mathcal{F}} \lesssim \sqrt{n^{-1}\log n},$$

in agreement with the bound given by Nemirovski's inequality. $\square$

5. Proof of Nemirovski's inequality

Proof: For given $r \in [2, \infty)$ consider the map V_r from $\mathbb{R}^d$ to $\mathbb{R}$ defined by

$$V_r(x) \equiv \|x\|_r^2.$$

Then V_r is continuously differentiable with Lipschitz continuous derivative ∇V_r. Furthermore

$$V_r(x_+y) \le V_r(x) + y' \nabla V_r(x) + CrV_r(y) \tag{4}$$

for an absolute constant C. Thus, writing $\sum_{i=1}^{n+1} X_i = \sum_{i=1}^{n} X_i + X_{n+1}$, it follows from (4) that

$$V_r(\sum_{i=1}^{n+1} X_i) \le V_r(\sum_{i=1}^{n} X_i) + X'_{n+1} \nabla V_r(\sum_{i=1}^{n} X_i) + CrV_r(X_{n+1}).$$

Taking expectations across this inequality and using independence of X_{n+1} and $\sum_{i=1}^{n} X_i$ together with $E(X_{n+1}) = 0$ yields

$$\begin{aligned} EV_r\left(\sum_{i=1}^{n+1} X_i\right) &\leq E\left\{V_r\left(\sum_{i=1}^{n} X_i\right) + X'_{n+1}\nabla V_r\left(\sum_{i=1}^{n} X_i\right)\right\} \\ &\qquad + CrEV_r(X_{n+1}) \\ &= EV_r\left(\sum_{i=1}^{n} X_i\right) + CrE\|X_{n+1}\|_r^2. \end{aligned}$$

By recursion this yields

$$EV_r\left(\sum_{i=1}^{n+1} X_i\right) \leq Cr\sum_{i=1}^{n+1} EV_r(X_i) \tag{5}$$

and hence the desired result with r rather than $\min\{r, \log d\}$.

To show that we can replace r by $\min\{r, \log d\}$ up to an absolute constant, first note that this follows immediately for $r \leq r(d) \equiv 2 \log d$ with C replaced by $2C$. Now suppose $r > r(d) = 2 \log d$. Recall that for $1 \leq r' \leq r$ we have

$$\|x\|_r \leq \|x\|_{r'} \leq d^{(1/r')-(1/r)} \|x\|_r$$

for all $x \in \mathbb{R}^d$ (by Hölder's inequality).

Thus with $r' = r(d) < r$

$$\begin{aligned} E\|\sum_{i=1}^n X_i\|_r^2 &\leq E\|\sum_{i=1}^n X_i\|_{r(d)}^2 \\ &\leq Cr(d)\sum_{i=1}^n E\|X_i\|_{r(d)}^2 \qquad \text{by (5)} \\ &\leq Cr(d)\sum_{i=1}^n E\left\{d^{\frac{2}{r(d)}-\frac{2}{r}}\|X_i\|_r^2\right\} \\ &\leq Cr(d)d^{2/r(d)}\sum_{i=1}^n E\|X_i\|_r^2 \\ &= 2Ce\log d\sum_{i=1}^n E\|X_i\|_r^2. \end{aligned}$$

Thus Nemirovski's inequality is proved for $r \in [2, \infty)$ with constant $\widetilde{C}$ given by $2eC$ and C the constant of (4).

6. Summary; problems and open questions

Thus it seems that Nemirovski's inequality yields bounds of order comparable to those achieved by bracketing methods from empirical process theory. Since the proofs are very different, it may be worthwhile to explore the exact constants achieved by the two methods in more detail. The following questions are then of particular interest:

- What is the best constant C in the basic inequality (4)?

6. Summary; problems and open questions

Thus it seems that Nemirovski's inequality yields bounds of order comparable to those achieved by bracketing methods from empirical process theory. Since the proofs are very different, it may be worthwhile to explore the exact constants achieved by the two methods in more detail. The following questions are then of particular interest:

- What is the best constant C in the basic inequality (4)?
- What are the best possible bounds of this type obtainable via truncation and Bernstein's inequality as used in traditional empirical process proofs?

6. Summary; problems and open questions

Thus it seems that Nemirovski's inequality yields bounds of order comparable to those achieved by bracketing methods from empirical process theory. Since the proofs are very different, it may be worthwhile to explore the exact constants achieved by the two methods in more detail. The following questions are then of particular interest:

- What is the best constant C in the basic inequality (4)?
- What are the best possible bounds of this type obtainable via truncation and Bernstein's inequality as used in traditional empirical process proofs?
- How do the best possible bounds of the two types mentioned above compare?

6. Summary; problems and open questions

Thus it seems that Nemirovski's inequality yields bounds of order comparable to those achieved by bracketing methods from empirical process theory. Since the proofs are very different, it may be worthwhile to explore the exact constants achieved by the two methods in more detail. The following questions are then of particular interest:

- What is the best constant C in the basic inequality (4)?
- What are the best possible bounds of this type obtainable via truncation and Bernstein's inequality as used in traditional empirical process proofs?
- How do the best possible bounds of the two types mentioned above compare?
- Can Nemirovski's inequality (or the method of proof) be extended to the range $1 \leq r \leq 2$?

Postscript

- During the workshop Lutz Dümbgen showed that Cr in (4) can be replaced by $1 \cdot (r-1)$.

Postscript

- During the workshop Lutz Dümbgen showed that Cr in (4) can be replaced by $1 \cdot (r-1)$.
- Also during the workshop Sara van de Geer obtained a simple inequality of the same type via truncation and Bernstein's inequality. which can be compared to (the refined version of) Nemirovski's inequality with sharp constant.

Postscript

- During the workshop Lutz Dümbgen showed that Cr in (4) can be replaced by $1 \cdot (r-1)$.
- Also during the workshop Sara van de Geer obtained a simple inequality of the same type via truncation and Bernstein's inequality. which can be compared to (the refined version of) Nemirovski's inequality with sharp constant.
- Preliminary comparison of (the refined version of) Nemirovski's inequality with sharp constant and the inequality obtained via Bernstein's inequality and truncation arguments in the case $r = \infty$ and i.i.d. summands X_i show that Nemirovski's inequality is better for $d < 2.33268 \times 10^{18} \equiv d_0$ while van de Geer's inequality is better for $d \geq d_0$. Further checking and comparisons will be reported elsewhere.

Postscript

- During the workshop Lutz Dümbgen showed that Cr in (4) can be replaced by $1 \cdot (r-1)$.
- Also during the workshop Sara van de Geer obtained a simple inequality of the same type via truncation and Bernstein's inequality. which can be compared to (the refined version of) Nemirovski's inequality with sharp constant.
- Preliminary comparison of (the refined version of) Nemirovski's inequality with sharp constant and the inequality obtained via Bernstein's inequality and truncation arguments in the case $r = \infty$ and i.i.d. summands X_i show that Nemirovski's inequality is better for $d < 2.33268 \times 10^{18} \equiv d_0$ while van de Geer's inequality is better for $d \geq d_0$. Further checking and comparisons will be reported elsewhere.