Nemirovski’s inequality revisited

Jon A. Wellner

University of Washington
• joint work with Lutz Dümbgen and Sara van de Geer,
• Faculty talk at Graduate Student Orientation, Department of Statistics, University of Washington 22 September, 2008
• Email: jaw@stat.washington.edu
  http://www.stat.washington.edu/jaw/jaw.research.html
Outline

1. A problem from statistics: persistence
2. The 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. Extensions and comparisons
1. A problem from statistics: persistence

Setting:

- **Data:** \( n \) i.i.d. copies \( Z_1, \ldots, Z_n \) of 
  \[ Z = (Y, X_1, \ldots, X_p) \equiv (Y, \underline{X}); \text{ write } Z_i = (Y^i, X_1^i, \ldots, X_p^i), \]
  \( i = 1, \ldots, n. \)

- **Dimension of** \( X \), \( p = p_n \) large, \( p_n = n^\alpha, \alpha > 1 \)

- **Goal:** Predict \( Y \) on the basis of the covariates \( X_j \),  
  \( j = 1, \ldots, p \)

- **Predictors** \( \hat{Y} \) of \( Y \) of the form  
  \[ \hat{Y} = \sum_{j=1}^{p} \beta_j X_j = \underline{\beta}' \underline{X} \]
  with \( \underline{\beta} \in B_n \subset \mathbb{R}^p \) for each \( n \).

- **Natural sets** \( B_n \) to consider are

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

  where \( k = k_n \to \infty \) and \( b = b_n \to \infty \).
For $Z = (Y, X) \sim P$ 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 \( Z = (Y, X) \sim P \) on \((\mathbb{R}^{p+1}, B_{p+1})\), 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

\[
\beta_n^*(P_n) \equiv \beta_n^* \equiv \arg\min_{\beta \in B_n} L_{P_n}(\beta).
\]

Thus \( \beta_n^* \) is a deterministic sequence in \( \mathbb{R}^p \) determined by \( P_n \) and \( B_n \).
• For \( Z = (Y, X) \sim P \) 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 \( \beta^*_n \) is a deterministic sequence in \( \mathbb{R}^p \) determined by \( P_n \) and \( B_n \).

• This corresponds to the unknown “ideal predictor” \( \hat{Y}^* = \beta^*_n X \) 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 $\{P_n\}$) if, for every sequence $P_n \in P_n$

$$L_{P_n}(\hat{\beta}_n) - L_{P_n}(\beta^*_n) \to_p 0.$$
2. The theorem of Greenshtein and Ritov

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

$$F(Z_i) \equiv \max_{0 \leq j, k \leq p} |X^i_j X^i_k - E_{P_n}(X^i_j X^i_k)|$$

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

$$\hat{\beta}_n \equiv \arg\min_{\beta \in B_{n,b_n}} L_{P_n}(\beta)$$

are persistent with respect to

$$B_{n,b_n} \equiv \{\beta \in \mathbb{R}^p : \|\beta\|_1 \leq b_n\}.$$
• **Comment 1.** The persistent procedures $\hat{\beta}_n$ in (1) are equivalent to Lasso estimators with a particular range of the penalty parameters.
• **Comment 1.** The persistent procedures $\hat{\beta}_n$ 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 $k = k_n = o((n/\log n)^{1/2})$. 
• **Comment 1.** The persistent procedures $\hat{\beta}_n$ 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 $k = k_n = o((n/\log n)^{1/2})$.

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

Nemirovski's inequality revisited – p. 10/31
Proof, part 1: With $\hat{\beta}_n \equiv \arg\min_{\beta \in B_n, b_n} L_{P_n}(\beta)$ it follows that

$$L_{P_n}(\hat{\beta}_n) - L_{P_n}(\beta_n^*) \geq 0,$$

$$L_{P_n}(\hat{\beta}_n) - L_{P_n}(\beta_n^*) \leq 0,$$

and hence

$$0 \leq L_{P_n}(\hat{\beta}_n) - L_{P_n}(\beta_n^*)$$

$$= L_{P_n}(\hat{\beta}_n) - L_{P_n}(\hat{\beta}_n) + L_{P_n}(\hat{\beta}_n) - L_{P_n}(\beta_n^*)$$

$$+ L_{P_n}(\beta_n^*) - L_{P_n}(\beta_n^*)$$

$$\leq 2 \sup_{\beta \in B_n, b_n} |L_{P_n}(\beta) - L_{P_n}(\beta)|.$$
Proof, part 1, continued: Let $\mathbb{P}_n$ be the empirical measure of $Z_1, \ldots, Z_n$. Then

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

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

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

Then

$$|L_{\mathbb{P}_n}(\beta) - L_P(\beta)| = |\gamma'(\Sigma_{\mathbb{P}_n} - \Sigma_P) \gamma| \leq \|\Sigma_{\mathbb{P}_n} - \Sigma_P\|_\infty \|\gamma\|_1^2.$$
Proof, part 1, continued: Thus for

\[ B_{n,b_n} = \{ \beta \in \mathbb{R}^p : \| \beta \|_1 \leq b_n \}, \]

\[
P_r \left( \sup_{\beta \in B_{n,b_n}} |L_{P_n}(\beta) - L_{P_n}(\beta)| > \epsilon \right) \leq Pr(\| \Sigma_{P_n} - \Sigma_{P_n} \|_{\infty} (1 + b_n)^2 > \epsilon) \]
\[
\leq \epsilon^{-1} (b_n + 1)^2 E \| \Sigma_{P_n} - \Sigma_{P_n} \|_{\infty} \]. \tag{3}

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

\[
E \| \Sigma_{P_n} - \Sigma_{P_n} \|_{\infty} \leq C \sqrt{\frac{\log n}{n}},
\]

then the proof is complete:
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 $r \in [2, \infty]$

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

where $\| \cdot \|_r$ is the $\ell_r$ norm, $\|x\|_r \equiv \left\{ \sum_{j=1}^{d} |x_j|^r \right\}^{1/r}$ and $\tilde{C}$ 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 $E\|\Sigma_{P_n} - \Sigma_{P_n}\|_\infty$, consider the matrix $\Sigma_{P_n} - \Sigma_{P_n}$ as a $(p + 1)^2$-dimensional vector, and write

$$\Sigma_{P_n} - \Sigma_{P_n} = \sum_{i=1}^{n} V_i$$

$$\equiv \sum_{i=1}^{n} \frac{1}{n} (X_0^i X_0^i - E(X_0^i X_0^i), X_0^i X_1^i - E(X_0^i X_1^i), \ldots, X_p^i X_p^i - E(X_p^i X_p^i)) .$$
• **First proof, part 2:** To apply Nemirovski’s inequality to bound $E\|\Sigma_{P_n} - \Sigma_{P_n}\|_\infty$, consider the matrix $\Sigma_{P_n} - \Sigma_{P_n}$ as a $(p + 1)^2$-dimensional vector, and write

$$ \Sigma_{P_n} - \Sigma_{P_n} = \sum_{i=1}^{n} V_i $$

$$ \equiv \sum_{i=1}^{n} \frac{1}{n} (X_0^i X_0^i - E(X_0^i X_0^i), X_0^i X_1^i - E(X_0^i X_1^i), \ldots, X_p^i X_p^i - E(X_p^i X_p^i)) $$

• By our hypothesis

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

satisfies $E_{P_n} F(Z_i)^2 \leq M < \infty$. 
First proof, part 2, continued: Then by Jensen’s inequality followed by Nemirovski’s inequality with $r = \infty$,

$$\left\{ E_{P_n} \| \Sigma_{P_n} - \Sigma_{ ForP_n} \|_\infty \right\}^2 = \left\{ E_{P_n} \| \sum_{i=1}^{n} V_i \|_\infty \right\}^2 \leq E_{P_n} \| \sum_{i=1}^{n} V_i \|_\infty^2 \leq C \log((p_n + 1)^2) \sum_{i=1}^{n} E_{P_n} \| V_i \|_\infty^2 \leq C'' \log(4n^{2\alpha}) \frac{1}{n^2} \sum_{i=1}^{n} EF(Z_i)^2 \leq C''' \frac{\log n}{n},$$

so that

$$E_{P_n} \| \Sigma_{P_n} - \Sigma_{ ForP_n} \|_\infty \leq C''' \sqrt{\frac{\log n}{n}}. \quad \square$$
4. Proof (part 2) – via bracketing entropy bounds

• Let $G_n \equiv \sqrt{n}(P_n - P_n)$. 
4. Proof (part 2) – via bracketing entropy bounds

- Let $G_n \equiv \sqrt{n}(P_n - P_n)$.

- For a class of functions $\mathcal{F} = \{ f : \mathbb{Z} \to \mathbb{R} \}$ write $\|G_n\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |G_n(f)|$. For $\mathcal{F}$ with $\#(\mathcal{F}) = d < \infty$, note that $\|G_n\|_{\mathcal{F}} = \|G_n(\underline{f})\|_{\infty}$ where $G_n(\underline{f}) \equiv (G_n(f_1), \ldots, G_n(f_d))$. 
4. Proof (part 2) – via bracketing entropy bounds

- Let $G_n \equiv \sqrt{n}(P_n - P_n)$. 

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

- 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 $\epsilon$ needed to cover $\mathcal{F}$. 
4. Proof (part 2) – via bracketing entropy bounds

- Let $G_n \equiv \sqrt{n}(P_n - P_n)$.

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

- 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 $\epsilon$ 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)

\[ E^* \| G_n \|_F \lesssim J_1[1, \mathcal{F}, L_2(P_n)] \| F \|_{P_n, 2}. \]


- In the current application take
  \[ \mathcal{F} = \{ f_{j,k}(z) = x_j x_k, \ 0 \leq j, k \leq p \}, \] a finite list of functions of cardinality 
  \[ \#(\mathcal{F}) = (p_n + 1)^2. \]
Lemma. (Empirical process theory bracketing entropy bound)

\[ E^* \|G_n\|_\mathcal{F} \lesssim J[1](1, \mathcal{F}, L_2(P_n))\|F\|_{P_n,2}. \]


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

- Hence \( N[1](\epsilon, \mathcal{F}, L_2(P_n)) \leq (p_n + 1)^2 \) 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)

\[ E^* \|G_n\|_F \lesssim J_{[]} (1, \mathcal{F}, L_2(P_n)) \|F\|_{P_n, 2}. \]


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

• Hence \[ N_{[]} (\epsilon, \mathcal{F}, L_2(P_n)) \leq (p_n + 1)^2 \] 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

\[ E \|G_n\|_F \lesssim \sqrt{1 + \log \left[ (p_n + 1)^2 \right]} \|F\|_{P_n, 2} \lesssim \sqrt{\log n}, \]
• Or, equivalently

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

in agreement with the bound given by Nemirovski’s inequality.
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 $\nabla V_r$. Furthermore

$$V_r(x + y) \leq V_r(x) + y' \nabla V_r(x) + Cr V_r(y) \quad (4)$$

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

$$V_r\left(\sum_{i=1}^{n+1} X_i\right) \leq V_r\left(\sum_{i=1}^n X_i\right) + X_{n+1}' \nabla V_r\left(\sum_{i=1}^n X_i\right) + Cr V_r(X_{n+1}).$$
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

$$EV_r \left( \sum_{i=1}^{n+1} X_i \right) \leq E \left\{ V_r \left( \sum_{i=1}^{n} X_i \right) + X'_{n+1} \nabla V_r \left( \sum_{i=1}^{n} X_i \right) \right\}$$

$$+ Cr EV_r (X_{n+1})$$

$$= EV_r \left( \sum_{i=1}^{n} X_i \right) + Cr E \|X_{n+1}\|_r^2.$$ 

By recursion this yields

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

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

Nemirovski's inequality revisited – p. 21/31
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 \( r' = r(d) < r \)

\[
E \left\| \sum_{i=1}^{n} X_i \right\|_r^{2} \leq E \left\| \sum_{i=1}^{n} X_i \right\|_{r(d)}^{2}
\]

\[
\leq Cr(d) \sum_{i=1}^{n} E \left\| X_i \right\|_{r(d)}^{2}
\]

\[
\leq C r(d) \sum_{i=1}^{n} E \left\{ d^{r(d)/r} \left( \frac{2}{r} \left\| X_i \right\|_r \right)^2 \right\}
\]

\[
\leq C r(d) d^{2/r(d)} \sum_{i=1}^{n} E \left\| X_i \right\|_r^{2}
\]

\[
= 2Ce \log d \sum_{i=1}^{n} E \left\| X_i \right\|_r^{2}
\]

Thus Nemirovski’s inequality is proved for \( r \in [2, \infty) \) with constant \( \tilde{C} \) given by \( 2eC \) and \( C \) the constant of (4).
6. Extensions and comparisons

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.

- In fact Nemirovski’s basic deterministic inequality

\[
V_r(x + y) \leq V_r(x) + y' \nabla V_r(x) + CrV_r(y)
\]  \hspace{1cm} (6)

holds in the following more precise form:

\[
V_r(x + y) \leq V_r(x) + y' \nabla V_r(x) + (r - 1)V_r(y)
\]  \hspace{1cm} (7)

where \( V_r(x) = \|x\|_r^2 \).
Thus Nemirovski’s inequality for sums of independent $X_i$’s holds in the form

$$E\left\| \sum_{i=1}^{n} X_i \right\|_r^2 \leq C(r, d) \sum_{i=1}^{n} E\|X_i\|_r^2$$

where $C(r, d) = \min\{r - 1, e(2 \log(d) - 1)\}$. In particular, when $r = \infty$,

$$E\left\| \sum_{i=1}^{n} X_i \right\|_\infty^2 \leq e(2 \log(d) - 1) \sum_{i=1}^{n} E\|X_i\|_\infty^2.$$
Two alternative methods for deriving similar bounds:

- By “type and co-type” theory together with symmetrization, (8) holds with

\[
C(r, d) = \begin{cases} 
8 \left( \frac{\Gamma((r+1)/2)}{\pi} \right), & 2 \leq r < \infty \\
2\pi c_d^2, & r = \infty.
\end{cases}
\]

where \( c_d^2 = E \max_{1 \leq j \leq d} Z_j^2 \leq 2 \log(d) \) for \( d \geq 3 \).
Two alternative methods for deriving similar bounds:

- By “type and co-type” theory together with symmetrization, \( (8) \) holds with

\[
C(r, d) = \begin{cases} 
8 \left( \frac{\Gamma((r+1)/2)}{\pi} \right), & 2 \leq r < \infty \\
2\pi c_d^2, & r = \infty.
\end{cases}
\]

where \( c_d^2 = \mathbb{E} \max_{1 \leq j \leq d} Z_j^2 \leq 2 \log(d) \) for \( d \geq 3 \).

- By truncation and Bernstein’s inequality \( (8) \) holds with

\[
C(\infty, d) = \left\{ 1 + 3.46 \sqrt{\log(2d)} \right\}^2.
\]
Figure 1: Comparison of $C(\infty, d)$ obtained via the three proof methods: Blue (bottom) = Nemirovski; Red and Black (middle) = type-inequalities / probability in Banach spaces; Green (top) = truncation and Bernstein inequality
- What sort of skill set(s) does a successful researcher in that area usually have?
  strong math background; interest in statistical theory
• What sort of skill set(s) does a successful researcher in that area usually have?
  strong math background; interest in statistical theory
• Why/How did you get involved in this subfield/area?
  Became interested in statistical theory as a grad student; wrote a book or two on empirical process theory.
- What sort of skill set(s) does a successful researcher in that area usually have?  
  strong math background; interest in statistical theory

- Why/How did you get involved in this subfield/area?  
  Became interested in statistical theory as a grad student; wrote a book or two on empirical process theory.

- Who else in the department works on similar things?  
  Galen Shorack and I have worked together on empirical process theory. Most other colleagues work on some aspects of theory / methodology.
• - What sort of skill set(s) does a successful researcher in that area usually have?
  strong math background; interest in statistical theory
• - Why/How did you get involved in this subfield/area?
  Became interested in statistical theory as a grad student; wrote a book or two on empirical process theory.
• - Who else in the department works on similar things?
  Galen Shorack and I have worked together on empirical process theory. Most other colleagues work on some aspects of theory / methodology.
• - If your project is interdisciplinary, how did you connect with people from other departments?
  Interdisciplinary in the sense of interactions between math (probability theory) and statistics. Connections made by attending national and international meetings + (random ?) contacts via internet.