Random Processes, Gaussian Width, Chaining

High Dim Probability & Linear Algebra for ML and Sig Proc

Namrata Vaswani

Iowa State University

Random Process

1. A r. process is a (possibly uncountably infinite) collection of r. variables $X_t, t \in T$. The set T can be a subset of $\Re^n$.
   - In classical examples, the set T is a subset of $\Re$ and often denotes the time index - continuous time or discrete time
     - If T is a finite set, say $T = \{1, 2, \ldots, n\}$, then we get a r. vector in $\Re^n$.
     - If T is set of integers, then we often refer to X_t as a r. sequence
     - Brownian motion: $T = \{t \geq 0\}$, X_t continuous almost surely, and $X_t - X_s \sim \mathcal{N}(0, t - s)$
   - When $T \subseteq \Re^n$: we often use the term random field, e.g., water temperature at different locations on earth
2. Assume zero mean, $\mathbb{E}[X_t] = 0$ for all $t \in T$.
3. Covariance function $\Sigma(t, s) := cov(X_t, X_s)$
4. Canonical pseudo-metric / Increments of X_t: Define a "distance pseudo-metric" on T using the r. process X_t:

$$d_X(t, s) := ||X_t - X_s||_{L^2}, \ where \ ||Z||_{L^2} := \sqrt{E[Z^2]}$$

   $d_X(.,.)$ is a pseudo-metric in general because $d(t, s) = 0$ does not imply $t = s$.
   The book often uses $||X_t - X_s||_2$ at various places but all of it should be $||X_t - X_s||_{L^2}$

Gaussian Random Process (GP)

1. $X_t, t \in T$ is a GP iff for, any finite subset $T_0 \subseteq T$, $(X_t)_{t \in T_0}$ is a Gaussian r. vector

2. Equivalently $X_t, t \in T$ is a GP iff every finite linear combination $\sum_{t \in T_0} a_t X_t$ is a Gaussian r. variable
3. The distribution of a zero-mean GP is completely determined by $\Sigma(t, s)$ and equivalently also by $d(t, s)$ (Ex 7.1.8)
4. Canonical GP: For a set $T \subset \Re^n$,

$$X_t = t^\top g, \ \ t \in T$$

   with $g \sim \mathcal{N}(0, I_n)$
5. For a canonical GP, $d_X(t, s)$ is a metric.

Bounding $\mathbb{E}[\sup_{t \in T} X_t]$ using another GP: Slepian and Sudakov-Fernique

1. *Assuming zero-mean GP everywhere*
2. Slepian's inequality
3. Sudakov-Fernique inequality: If

$$\forall t, s \in T, \ \ E[(X_t - X_s)^2] \leq E[(Y_t - Y_s)^2]$$

   then

$$\mathbb{E}[\sup_{t \in T} X_t] \leq \mathbb{E}[\sup_{t \in T} Y_t]$$

4. Application to get a tight bound on $\mathbb{E}[||A||]$ for A with i.i.d. Gaussian entries.

5. Sudakov-minoration inequality: For any $\epsilon \geq 0$,

$$\mathbb{E}[\sup_{t \in T} X_t] \geq c\epsilon\sqrt{\mathcal{N}(T, d_X, \epsilon)}$$

where $\mathcal{N}(T, d, \epsilon)$ is the covering number of T in metric $d_X(.,.)$ (smallest size of epsilon net that covers T when the eps-balls are defined using $d(.)$).

Proof for compact set T: follows from S-F

- $$\mathbb{E}[\sup_{t \in T} X_t] \geq \mathbb{E}[\sup_{t \in epsNet(T)} X_t]$$
- RHS is sup over a finite set, and hence sup can be replaced by max.
- Define $Y_t = \epsilon g_t/\sqrt{2}$ with $g_t \sim \mathcal{N}(0,1)$.
- For two points $t, s \in epsNet$, the distance is more than epsilon (reason: epsNet is the smallest possible epsNet – maximal eps-separated subset of T)
- Thus, one can show that S-F applies and we get

$$\mathbb{E}[\sup_{t \in epsNet} X_t] \geq \mathbb{E}[\sup_{t \in epsNet} Y_t] = (\epsilon/\sqrt{2})\mathbb{E}[\max_{t \in epsNet} g_t] \geq c\epsilon\sqrt{\log \mathcal{N}(T, d_X, \epsilon)}$$

- Last inequality follows by Ex 2.5.11 that lower bounds max of indep Gaussian r.v.s and the fact that $|epsNet| = \mathcal{N}(T, d_X, \epsilon)$

6. Application to bound covering numbers of the set T in Euclidean distance metric: apply S-m with $X_t = t^\top g, \ t \in T$

Gaussian width

1. G width of set $T \subset \Re^n$

$$w(T) := \mathbb{E}[sup_{t \in T} t^\top g], \quad g \sim \mathcal{N}(0, \mathbf{I}_n)$$

   Note from above that

$$w(T) \geq c\epsilon \sqrt{\log \mathcal{N}(T, \epsilon)}$$

2. Properties : TBD
3. Clearly $w(c_1 T) = c_1 w(T)$
4. Relation to spherical width: roughly $w(T) \approx \sqrt{n} w_s(T)$
5. Define $diam(T)$
6. Examples: computing or upper/lower bounding Gaussian width of unit ball and sphere, of unit ell-1 ball

Chaining: Dudley's inequality (Chap 8.1)

1. TBD
2. Sudakov-minor tells us Dudley is not tight

Applications

1. Application to get a tight bound on $\mathbb{E}[||A||]$ for A with i.i.d. Gaussian entries.
2. By Sudakov-minor,

$$w(T) \geq c\epsilon \sqrt{\log \mathcal{N}(T, \epsilon)}$$

3. Sudakov-minor tells us that Dudley is not tight

4. G width is used to quantify sample complexity of sparse recovery and many similar problems: see Theorem 10.5.1 Proof of this result :
   - This result follows from Escape from Mesh result of Sec 9.4,
   - which, in turn, uses Corollary 8.6.2/8.6.3 (Talagrand's comparison inequality).
   - Proof of Corollary 8.6.2/8.6.3: uses a generic chaining bound (Thm 8.5.3), and lower bound of Theorem 8.6.1
   - Proof of lower bound of Theorem 8.6.1 uses a multi-scale version of Sudakov-minor.
   - Cor 8.6.2, 8.6.3: X_t is a zero-mean subG process, Y_t is zero-mean GP. If

     $$\forall t, s \in T, \ ||X_t - X_s||_{\psi_2} \leq K d_Y(t,s), \quad d_Y(t,s) := ||Y_t - Y_s||_{L^2}$$

     Then,

     $$\mathbb{E}[\sup_{t \in T} X_t] \leq CK \mathbb{E}[\sup_{t \in T} Y_t]$$

     Pick $Y_t = t^\top g$ (canonical GP), then the result becomes: If

     $$\forall t, s \in T, \ ||X_t - X_s||_{\psi_2} \leq K ||t - s||,$$

     Then,

     $$\mathbb{E}[\sup_{t \in T} X_t] \leq CK w(T)$$

   - In summary, the sparse recovery sample complexity guarantee uses
     - ★ Gaussian width and upper bound on it for ell-1 ball

- ★ a difficult multi-scale version of Sudakov-minor (proof of Sudakov-minor uses S-F)
- ★ a generic chaining result (proof of Dudley introduces the chaining idea)

Proof sketches for Slepian and S-F: Gaussian interpolation idea

1. TBD
2. S-F: same overall approach, pick $f_\beta(x) :=??$

bb I

aa