Chaining and Dudley's Inequality

Aarshvi Gajjar

Stochastic Process

- $\{X_t\}_{t \in \mathbb{T}}$ is a random process if it is a collection of random variables on the same probability space, indexed by elements of some set $\mathbb{T}$.
- If $\mathbb{T} = [n]$, then it is a random vector.
- If X_t is an n dimensional normal random variable, where $\mathbb{T} \subset \mathbb{R}^n$ and then it is the canonical Gaussian process.

Expected Supremum

- In many statistical problems, we want to control $\mathbb{E} \sup_{t \in \mathbb{T}} X_t$.
- For example, if $\mathbb{T} = \mathcal{F}$, which is a function class and $z = [z_1, z_2, \ldots, z_n]$ is a data vector.
- If $X_f = \langle f(z), \sigma \rangle$, where σ_i are Rademacher r.v., then $\mathbb{E}_\sigma \sup_{f \in \mathcal{F}} \langle f(z), \sigma \rangle$ is the *Rademacher complexity*, which measures the expressiveness of a function class.
- In many scenarios, like regression, it becomes important to control the expected supremum of the process.

Why is this difficult?

- Note that $X_t \leq \sup_t X_t$ and by monotonicity, $\mathbb{E} X_t \leq \mathbb{E} \sup_t X_t$, which further implies that $\sup_t \mathbb{E} X_t \leq \mathbb{E} \sup_t X_t$.

Background

- $\{X_t\}_{t \in \mathbb{T}}$ is a *subGaussian process* with respect to a metric, d on $\mathbb{T}$ if for any $s, t \in \mathbb{T}$ and $\lambda \in \mathbb{R}$,
$$\mathbb{E} e^{\lambda \cdot (X_t - X_s)} \leq e^{\frac{\lambda^2}{2} \cdot d^2(s,t)}$$

Background

- We will see Dudley's Inequality for Gaussian process, which is $\{X_t\}_{t\in\mathbb{T}}$, where X_t is a mean 0 Gaussian random variable.
- Currently, $\mathbb{T}$ is an arbitrary set, without any geometry on it.
- To add geometry, we induce the Euclidean metric $d : \mathbb{T} \times \mathbb{T} \to \mathbb{R}$, $d(s,t) = \|s - t\|_2$.
- It can be verified that the zero mean Gaussian process is subgaussian w.r.t. the Euclidean metric, i.e. for any $\lambda \in \mathbb{R}$ and any $s,t \in \mathbb{T}$,

$$\mathbb{E}e^{\lambda\cdot(X_t - X_s)} \leq e^{\frac{\lambda^2}{2}\cdot\|s-t\|_2^2}$$

- This means that $(\mathbb{T}, d)$ is a metric space and it satisfies triangle inequality, which is what we need for our main result.

Background: Covering Numbers

- $N(\mathbb{T}, d, \varepsilon)$ is the smallest number of "balls" of radius ε w.r.t. a distance d that can cover $\mathbb{T}$.
- For every $t \in \mathbb{T}$ there exists a point t' in the ε cover such that $d(t, t') \leq \varepsilon$.

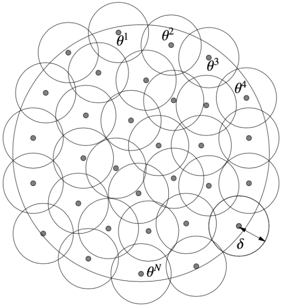


Figure: Geometrically, the union of balls of radius δ centered at $\{\theta^1, \theta^2 \ldots \theta^N\}$ cover $\mathbb{T}$.

Dudley's Inequality

- **Single scale version**: What if we took just 1 net?
- If $\pi(t)$ is the closest point to t on a net,

$$\mathbb{E}\sup_{t\in T}|X_t| \leq \mathbb{E}\sup_{t\in T}\left|X_{\pi(t)}\right| + \mathbb{E}\sup_{t\in T}\left|X_{\pi(t)} - X_t\right|.$$

- The first can be controlled since $\sup_{t\in\mathbb{T}} \to \sup_{\pi(T)\in\text{Net}}$, which is finite. But the second term is uncertain.

Theorem (1967)

Any mean 0 *Gaussian process satisfies the following for some constant* $C > 0$*.*

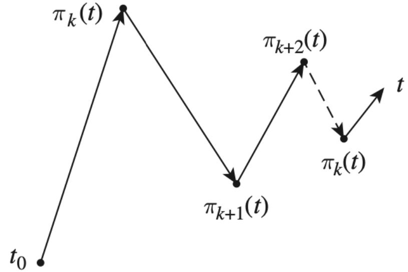
$$\mathbb{E}\sup_{t\in\mathbb{T}} X_t \leq C \cdot \int_0^\infty \sqrt{\log(N(\mathbb{T}, d, \varepsilon))} d\varepsilon$$

Proof

- Note that it suffices to give an upper bound on $\mathbb{E}\sup\left|X_t - X_{t_0}\right|$ for some t_0, since $\mathbb{E}\sup_t|X_t| = \mathbb{E}\sup_t\left|X_t - \mathbb{E}[X_{t_0}\right|] \leq \mathbb{E}\sup_t\left|X_t - X_{t_0}\right|$ by Jensens inequality.
- Now we begin the multiscale approximation of $\mathbb{T}$.
- Let $diam(\mathbb{T}) = 1$ and let t_0 be the center of the ball that covers the entire $\mathbb{T}$.
- Choose $\varepsilon_i = \frac{1}{2^i}$ nets, $\mathbb{T}_k$ of $\mathbb{T}$, for $i \in \mathbb{N}$. This means that for every $t \in \mathbb{T}$ there exists a $t_i \in \mathbb{T}_i$ such that $d(t, t_i) \leq \varepsilon_i$.

Chaining

- Suppose that we are at t. We start from the coarsest approximation, t_0 and find a path to t through a **chain** at points $\pi_1(t), \pi_2(t), \ldots$, where each $\pi_i(t) \in \mathbb{T}_i$.



- $X_{t_0} - X_t = (X_{t_0} - X_{\pi_1(t)}) + (X_{\pi_1(t)} - X_{\pi_2(t)}) \ldots (X_{\pi_\kappa}(t) - X_t)$

Proof

- Bound each link in the chain as $\left|X_{t_0} - X_t\right| \leq \sum_{i=1}^{\infty} \left|X_{\pi_{i-1}(t)} - X_{\pi_i}(t)\right|$
- Since the distance is a metric, we know that $d(\pi_{i-1}(t), \pi_i(t)) \leq d(\pi_{i-1}(t), t) + d(t, \pi_i(t)) \leq \frac{1}{2^{i-1}} + \frac{1}{2^i} = \frac{3}{2^i}$
- Taking supremum, we have that $\mathbb{E} \sup_{t \in \mathbb{T}} \left|X_{t_0} - X_t\right| \leq \sum_{i=1}^{\infty} \mathbb{E} \sup_{\substack{u \in \mathbb{T}_{i-1} \\ v \in \mathbb{T}_i \\ d(u,v) \leq 3\varepsilon_i}} |X_u - X_v|$.

- It can be proved that $X_u - X_v$ is subgaussian under d, and therefore the quantity can be bounded by the sizes of $\mathbb{T}_i$ and $\mathbb{T}_{i-1}$ as follows.

$$
\begin{aligned}
e^{\lambda \cdot \mathbb{E} \sup |X_u - X_v|} \leq \mathbb{E} e^{\lambda \cdot \sup |X_u - X_v|} &= \mathbb{E} \sup e^{\lambda \cdot |X_u - X_v|} \\
&\leq \sum_{v,u} \mathbb{E} e^{\lambda \cdot |X_u - X_v|} \leq |\mathbb{T}_i| |\mathbb{T}_{i-1}| \, e^{\frac{\lambda^2}{2} \cdot d^2(s,t)}
\end{aligned}
$$

- Therefore the second term can be upper bounded as follows.

$$
\mathbb{E} \sup |X_u - X_v| \leq d(u,v) \cdot \sqrt{2 \log(|\mathbb{T}_i| |\mathbb{T}_{i-1}|)} \leq 3\varepsilon_i \cdot \sqrt{2 \log(|\mathbb{T}_i|^2)}
$$

- This gives the following discrete bound

$$
\mathbb{E} \sup_{t \in \mathbb{T}} |X_{t_0} - X_t| \leq C \cdot \sum_{i=1}^{\infty} \frac{1}{2^i} \sqrt{\log(N(\mathbb{T}, d, \varepsilon_i))}
$$

Final touches

- This can be interpreted as a Reimannian sum with $\Delta\varepsilon = \frac{1}{2^{i-1}} - \frac{1}{2^i}$ thus giving us the desired bound of $C \cdot \int_{i=1}^{\infty} \frac{1}{2^i} \sqrt{\log(N, \mathbb{T}, \varepsilon)} d\varepsilon$.
- Q.E.D.

Takeway

- Dudley's Inequality applies to subgaussian process, which include Gaussian and Rademacher.
- If you want to control the expected supremum, you might want to show that your process is subGaussian under some (pseudo) metric.

Example

- Let $\mathbb{T} = B_2^n$. Note that $N \leq (\frac{3}{\varepsilon})^n$, then Dudley's inequality states that $\mathbb{E}\sup_t X_t \leq C \cdot \int_0^1 \sqrt{n \cdot \log(3/\varepsilon)} d\varepsilon \leq C\sqrt{n}$.

References

- High Dimensional Statistics (section 5.3), Martin Wainwright.
- High Dimensional Probability (Chapter 8), Roman Vershynin.
- Notes on Rademacher complexity by Renjie Liao.
- Lecture notes for CMU's Advanced Statistical Learning theory course by Alessandro Rinaldo.