Asymptotic Geometry and Concentration

Aukosh Jagannath

April 4, 2013

Abstract

In this document, I attempt to convince the reader that there are interesting questions in asymptotic geometry and try to demonstrate the important role that probabilistic techniques play in this field. I will then turn this on its head and show that using ideas from geometry you can get nontrivial statements in probability.

1 Introduction

The purpose of this document is to demonstrate an interesting connection between (metric) geometry and probability. I will, hopefully, convince you that you can probe interesting geometric questions using probabilistic techniques and interesting probabilistic questions using geometry.

I’ll first demonstrate using a few simple questions that tackling high dimensional problems using your intuition from dimension two and three can be difficult. Many of the results in this talk will be truly high dimensional problems and will, hopefully, help you believe the maxim: A function that depends on many variables in a reasonable way stays very close to its mean.¹

Remark 1. Notation. Hence forth, $\gamma^n$ denotes the $n$–dimensional standard multivariate gaussian, and $\gamma^n(\mu, \Sigma)$ denotes a gaussian with mean $\mu$ and covariance $\Sigma$, thus $\gamma^n = \gamma^n(0, Id)$. $\sigma^n$ denotes the uniform measure over the sphere $S^n$. $\lesssim$ means up to a universal constant etc.

2 The standard examples

Here I’ll discuss a few standard examples that are often used to convince people that their intuition in low dimensions does not follow through to larger dimensions.

Exercise 2. The $n$-sphere is spiky. Consider the unit cube sitting in $\mathbb{R}^n$ centered at the origin. Go to each vertex of the cube and blow a bubble all at the same time and same speed. Whenever any two bubbles touch, they stop growing. When this process has terminated, go to the center of the cube and blow one final bubble. When this bubble touches any of the others, have it stop growing. What happens to the radius of the inner bubble as $n \to \infty$?

Exercise 3. $\ell_\infty^n$ is strictly bigger than in more ways than one $\ell_2^n$. A simple question: what is the radius of the unit volume $\ell_\infty$ ball? how about the $\ell_2$ ball?

Exercise 4. The bulk is very small. Consider the $n$ dimensional gaussian measure. Where is the bulk of the measure?

To do the least problem the following argument will be helpful

Proposition 5. (Gaussian concentration) Let $F$ be $1$–Lipschitz and mean zero. Then

$$\gamma^n(|F(x)| \geq \epsilon) \leq Ce^{-c\epsilon^2/2}$$
Proof. Let $X$ and $Y$ be independent copies of $\gamma^n$, and let $X(t) = sin(t)X + cos(t)Y$ and $X'(t) = cos(t)X + sin(t)Y$. Notice that these are mutually independent copies of $X$. Then

$$\hat{\sigma}_N F(x) = e^{-\lambda \int F(y) d\gamma} \int e^{\lambda F(x)} d\gamma$$

Jensen's inequality gives

$$\leq \int e^{\lambda \int F(x) d\gamma} \int e^{-\lambda \int F(y) d\gamma} d\gamma$$

$$= \int \int e^{\lambda \frac{\pi}{2} \nabla F(X(t)) \cdot X'(t) dt} d\gamma d\gamma$$

Jensen's inequality again

$$\leq \frac{2}{\pi} \int_0^{\pi/2} \int e^{\lambda \frac{\pi}{2} \nabla F(X(t)) \cdot X'(t)} d\gamma d\gamma dt$$

Rot. Inv.

$$= \int e^{\lambda \frac{\pi}{2} \nabla F(X(t)) \cdot Y} d\gamma d\gamma$$

$$\leq e^{\frac{\lambda^2}{4} \nabla F(x)^2} d\gamma$$

$$\leq e^{\frac{\lambda^2}{4} \nabla F(x)^2}$$

Using exponential Chebyshev,

$$\mathbb{P}(F \geq \epsilon) \leq e^{-\lambda \epsilon + \lambda^2 \pi^2 / 4}$$

Minimizing in $\lambda$, we see that

$$\epsilon^2 / \pi^2 = \lambda$$

Minimizes the r.h.s., so

$$\mathbb{P}(F \geq \epsilon) \leq e^{-\epsilon^2 / \pi^2}$$

Remark 6. The $c$ in this argument is suboptimal. The only way I know how to the optimal choice ($c = 1/2$) you have to fight for it using much more advanced techniques.

Hopefully these exercises have convinced you that you have to think at least a little bit to understand high dimensions.

3 Concentration on the sphere

Consider the sphere $S^n$. It’s fairly simple to see that $Vol_{n-1}(S^{n-1}) = \frac{2\pi^{n/2}}{\Gamma(\frac{n}{2})}$. One question you might ask is where does the mass of the sphere live. By playing around with the spheres $S^{n-1}$ sitting in $S^n$ along the $x_1$ direction it is straightforward to find a very important property of spheres.

Theorem 7. (Spherical Concentration) The following bound holds for all $F$ 1-Lipschitz.

$$\sigma^N(\{ F - \int F d\sigma^N | \geq r \}) \leq C e^{-c(n-1)r^2 / 2}$$

Rather than prove this result, I’ll try to demonstrate it, by asking a simple question Where is the mass on the sphere?

Proof. (Demonstration, not proof) The following is an intuitive idea of why you might expect spherical concentration. Consider how much mass on the sphere lives in $< x, e_1 > \in [-1,1] \setminus (-x_1, x_1)$. Note that this
is just
\[
2 \int_{x_1}^{1} Vol_{n-1}(S^{n-1}(\sqrt{1-x^2}))dx = \frac{\Gamma(n+1/2)}{\pi^{n/2} \Gamma(1/2)} \int_{x_1}^{1} (1-x^2)^{-1/2}dx
\]
\[
= C(n) \int_{x_1}^{1} \cos(t)^n dx
\]
\[
\leq C(n) \int_{\arcsin(x_1)}^{\pi/2} e^{-nx^2/2}dx
\]
\[
\leq C(n)e^{-nx^2/2}
\]
since \(C(n)\) is growing like \(O(n^{1/2})\) you see that this bound is (roughly) what I stated above.

Notice that this result implies that the amount of mass in any cap of fixed radius, or even radius growing, but growing slowly is essentially nothing. This leads to a particularly bizarre result as we’ll see.

One really interesting result is

**Corollary 8. (Levy)** Let \(f \in C(S^n)\) \(A = \{f = M_f\}\) where \(M_f\) is the meadian of \(f\). Then

\[
\sigma^n(A_\epsilon) \geq 1 - 2e^{-(n-1)\epsilon^2/2}
\]

**Proof.** \(A_\epsilon = \{f \leq M_f\}_\epsilon \cap \{f \geq M_f\}_\epsilon\) and so \(P(A) \geq 1 - 2e^{-(n-1)\epsilon^2/2}\)

## 4 A not so standard example.

As we saw above the mass of a cap on the sphere is essentially nothing, so a natural question to ask is: Is there a scale of variation on which the cap is not small? The following application of (Not) Poincaré’s limit. \(^1\)

**Theorem 9.** Let \(c > 0\), and \(B(e_1) = \{y : y_1 \geq c/\sqrt{n}\}\). That is \(B(e_1) = B_r(e_1)\) with \(r = \arccos(c/\sqrt{n})\).
Then

\[
\sigma^{n-1}(B(e_1)) \rightarrow \gamma(x > c)
\]

**Proof.** let \(g = (g_1, \ldots, g_{n+1}) \in \mathbb{R}^n\). note that if \(x \in S^{n-1}\) has \(x \sim \sigma^{n-1}\), then if \(g \sim \gamma^n\), we have \(\frac{g}{\|g\|_2} \sim x\). thus

\[
\sigma^{n-1}(\sqrt{n}x_1 \geq c) = \gamma^n\left(\frac{\sqrt{n}}{\|g\|_2} g_1 \geq 1\right)
\]

Now recall that by the strong law of large numbers,

\[
\frac{\sum g_i^2}{n} \overset{a.s.}{\rightarrow} 1.
\]

Since almost sure convergence implies convergence in probability, note that \(\forall \epsilon, \delta > 0, \exists N_0: n > N_0\) gives

\[
\gamma^n\left(\frac{\sqrt{n}}{\|g\|_2} |g_1| \geq 1 - \delta < \epsilon\right)
\]

thus

\[
\gamma^n\left(\frac{\sqrt{n}}{\|g\|_2} g_1 \geq c\right) = \gamma^n\left(g_1 \geq \frac{\|g\|_2}{\sqrt{n}} c\right)
\]

\[
\leq \gamma^n\left(g_1 \geq \frac{\|g\|_2}{\sqrt{n}} c, 1 - \delta < \frac{\|g\|_2}{\sqrt{n}} / 1 + \delta + \epsilon\right)
\]

\[
\leq \gamma(g_1 \geq (1 - \delta)c + \epsilon)
\]

\(^1\)Everyone seems to go out of their way to call it poincare’s limit, but at the same time talk at length about how it’s not due to poincare, so I give it this name.
similarly, since for any set \( A \in \sigma(g_1) \)
\[
\gamma(A) - \gamma^n(A, 1 - \delta < \frac{||g||^2}{\sqrt{n}} < 1 + \delta) = \gamma^n(A, \frac{||g||^2}{\sqrt{n}} < 1 > \delta) < \epsilon
\]
combining these results we get
\[
\gamma^n(g_1 \geq \frac{||g||^2}{\sqrt{n}} c) \geq \gamma^n(g_1 \geq \frac{||g||^2}{\sqrt{n}} c, 1 - \delta < \frac{||g||^2}{\sqrt{n}} < 1 + \delta)
\]
\[
\geq \gamma^n(g_1 \geq (1 + \delta)c, 1 - \delta < \frac{||g||^2}{\sqrt{n}} < 1 + \delta))
\]
sandinghing, taking limsup and liminf, and sending epsilon and delta to zero finishes the result.

Now this result might sound uninteresting, but it implies a rather (to me) surprising result. Recall that in convex geometry on euclidean space one generally compares the volume of a convex set to the smallest ball that contains it and the biggest ball that is within it. As you (hopefully) saw above, in the euclidean setting the mass distribution is quite different than you would expect, there is some how massively more mass in a cube than the biggest ball sitting in side it. One natural question might be to ask if such analyses give interesting results with geodesically convex settings. It turns out that this idea gives some surprising results in this setting on the sphere.

To see why, let’s study a very simple geodesically convex set. Consider the simplex on the sphere \( S^{n-1} \subset \mathbb{R}^n \) formed by the \( n \) coordinate vectors, call it \( \Delta \). Note that if you take the point right at the center of this triangle, \( v = \frac{1}{\sqrt{n}} \sum e_i \), it is clear that \( <v, e_i>=\frac{1}{\sqrt{n}} \). Firstly this result itself is a little interesting, it says that in a high dimensional sphere, this vector is getting very far from all of the coordinate vectors but very close to all of the coordinate planes.

If you are already worried at this point, you should be. First notice that by simple arguments \( \sigma^{N-1} \Delta = (\frac{1}{2})^N \) (pretend it’s vectors are a gaussian as above and then it’s the probability that all of the components are positive). This is quite odd though, given the result from the previous section: the probability of the smallest ball containing it is asymptotically a large constant, however, this vanishes asymptotically.

We might try to approximate from the inside by a ball. In particular, the ball we need to consider now is (up to rotation) \( A(e_1) = \{ x : x_1 \geq 1 - \frac{c}{\sqrt{N}} \} \). Note that the correct scale to look at the triangle is logarithmic.
That is,
\[
\frac{1}{N} \log \sigma^{N-1}(\Delta) \rightarrow -\log 2
\]
so it is natural to ask if it is comparable to
\[
\frac{1}{N} \log \sigma^{N-1}(A)
\]
it turns out that it is not.

**Lemma 10.** Fix \( c > 0 \) and let \( A(e_1) = \{ x : x_1 \geq 1 - \frac{c}{\sqrt{N}} \} \), then
\[
\frac{1}{N + 1} \log \sigma^N(A) \rightarrow -\infty
\]

**Proof.** First recall that for a cap of radius \( r \), we have
\[
\mathbb{P}(Cap_N(r)) = \frac{1}{2} \frac{\beta(\sin(r)^2, \frac{N}{2}, \frac{1}{2})}{\beta(\frac{N}{2}, \frac{1}{2})}
\]
\[ \cos(r) = 1 - \frac{r^2}{N+1} \text{ so } \sin^2(r) = \frac{2r^2}{N+1} - \frac{r^2}{N+1} \text{ so} \]

\[
\beta(\sin(r)^2, \frac{N}{2}, 1/2) = \frac{1}{2} \int_0^{\infty} z^{1/2} (1-z)^{-1/2} dz \leq \frac{1}{2} \int_0^{\frac{2r}{N+1}} z^{n/2} dz = \frac{1}{2(N/2+1)} \left( \frac{2r}{N+1} \right)^{N/2+1}
\]

while

\[
\beta(\frac{N}{2}, 1/2) = \frac{C}{\sqrt{1+1/N}} \left( \frac{1}{1+\frac{1}{N}} \right)^{N/2} \frac{1}{\sqrt{N}}
\]

so that the ratio has logarithm

\[
\frac{N}{2} \left( \log(2c) - \log(N+1) \right) + O(\log(N))
\]

so that when we divide by \(N+1\) and take the limit, we get a quantity that is going to \(-\infty\) like \(-O(\log N)\).

In particular we see that if we set \(c = 1\), we get that the largest ball sitting in \(\Delta\) is dying much faster than the simplex itself.

To summarize

**Proposition.** Let \(\Delta \subset S^{N-1} \subset \mathbb{R}^n = \{x \in S^{n-1} : x_i \geq 0\}\), let \(Out(\Delta)\) be the smallest geodesic ball that contains \(\Delta\) and let \(Inn(\Delta)\) be the largest geodesic ball contained in \(\Delta\). Then

- \(\sigma^{N-1}(Out(\Delta)) \to \int_1^\infty e^{-x^2/2} dx/\sqrt{2\pi} \approx 0.159\)
- \(\frac{1}{N} \log(\sigma^{N-1}(\Delta)) = -\log 2\)
- \(\frac{1}{N} \log(\sigma^{N-1}(Inn(\Delta)) = -O(\log N) \to -\infty\)

### 5 Application: Tomaszewski’s Conjecture

**Conjecture 11.** Let \(a \in S^{n-1} \subset \mathbb{R}^n\) and let \(X \in \{-1, 1\}^n\) be uniformly distributed. Then \(\forall n\)

\[ p_n(a) = \mathbb{P}_X(|a \cdot X| \leq 1) \geq 1/2 \]

Now, let \(Y = \frac{X}{\sqrt{n}}\) so that \(Y \in S^{n-1}\). Then this is equivalent to showing

\[ p(a) = \mathbb{P}_Y(|a \cdot Y| \leq \frac{1}{\sqrt{n}}) \geq 1/2 \]

**Lemma 12.** Let \(a \in S^{n-1}\) be uniformly distributed over the sphere. Then

\[ \mathbb{E}_a[p_n(a)] = \mathbb{P}_a(|a_1| \leq 1/\sqrt{n}) \]

**Proof.** Consider the expected probability on the sphere. That is

\[ \mathbb{E}_a[p_n(a)] \]

Notice that

\[
\mathbb{E}_a[p_n(a)] = \mathbb{E}_a \left[ \mathbb{E}_Y \left[ 1_{|a \cdot Y| \leq \frac{1}{\sqrt{n}}} \right] \right] = \mathbb{E}_Y \left[ \mathbb{E}_a \left[ 1_{|a \cdot Y| \leq \frac{1}{\sqrt{n}}} \right] \right] = \mathbb{E}_Y \left[ \mathbb{P}_a \left( |a_1| \leq \frac{1}{\sqrt{n}} \right) \right] = \mathbb{P}_a(|a_1| \leq 1/\sqrt{n})
\]
where the last inequality follows from the fact that the probability in the third line is actually the measure of a strip and is thus rotation invariant and thus constant in $Y$, and furthermore $Y$ can be taken to be the north pole.

**Proposition 13.** This satisfies the limit
\[
\lim_{n \to \infty} E_n(p_n(a)) = P(|Z| \leq 1)
\]
where $Z \sim N(0,1)$

**Proof.** This follows immediately from the above lemmas.

**Remark 14.** This lemma is asymptotically sharper than the standard concentration inequality. To see why, note that the standard concentration equality would suggest that
\[
P(a_i > 1/\sqrt{n}) \leq e^{-\frac{n-1}{2} \left(\frac{1}{\sqrt{n}}\right)^2} \to \frac{1}{\sqrt{e}} \approx .61
\]
however as we can see from the above calculation this is a very large over estimate.

### 6 Almost Orthogonality

Let $X_1, \ldots, X_m \sim \sigma^N$. One question we might as is, where do these points live with respect to each other. It turns out they’re quite far from each other, in fact they’re all almost orthogonal from each other with high probability provided $m$ grows polynomially with the dimension. To be precise

**Theorem 15.** Let $X_i$ as above, then
\[
\sigma^N(|\langle X_i, X_j \rangle| \leq \epsilon, \forall i \neq j) \geq 1 - \left(\frac{m}{2}\right) e^{-(N-1)\epsilon^2/2}
\]

The result follows fairly quickly from a few lemmas

**Lemma 16.** For $X_i$ as above, then for any $p \in S^{N-1}$
\[
\sigma^N(|\langle X_1, X_2 \rangle| \leq \epsilon) = \sigma^N(|\langle e_1, X_2 \rangle| \leq \epsilon)
\]

**Proof.** Firstly,
\[
\sigma^N(|\langle X_1, X_2 \rangle| \leq \epsilon) = \int \int 1_{|\langle x, y \rangle| \leq \epsilon} d\sigma(y) d\sigma(x)
\]
now, fix an $x$ and consider the inner integral. Let $O_x$ denote an orthogonal transformation that takes $x$ to $e_1$. Then if $O_x^T z = y$
\[
\int 1_{|\langle x, y \rangle| \leq \epsilon} d\sigma(y) = \int 1_{|\langle x, O_x^T z \rangle| \leq \epsilon} d\sigma \circ O_x(z) = \int 1_{|\langle O_x x, z \rangle| \leq \epsilon} d\sigma(z) = \int 1_{|\langle e_1, z \rangle| \leq \epsilon} d\sigma(z)
\]
thus the inner integral is constant. Pulling it out of the integral gives us the result.

**Lemma 17.** The map $f(x) = x_1$ is $1 - \text{Lip}$ on $S^N$ and mean zero.
Proof. Just note that \( d(x, y) \geq |d(x, e_1) - d(y, e_1)| \) by triangle inequality so that
\[
\frac{|f(x) - f(y)|}{d(x, y)} = \frac{\cos(d(x, e_1)) - \cos(d(y, e_1))}{d(x, y)} \leq \|\cos(x)\|_{Lip([0,1])} = 1
\]
that its mean zero is clear. Again just use rotation invariance.

Proof. (Of Thm) The result now follows by an application of the concentration inequality. Note that by the above
\[
\sigma^N(|<X_1, X_2| \leq \epsilon) = \sigma^N(|e_1, X_2| \leq \epsilon) \geq 1 - \epsilon^{(n-1)\epsilon^2/2}
\]
since the events of the above type are the same probability. We see that by the above lemma we get
\[
\sigma^N(|(X_i, X_j) \leq \epsilon, \forall i \neq j) \geq 1 - \left(\frac{m}{2}\right) e^{-(N-1)\epsilon^2/2}
\]
This seems very surprising at first glance. We're taking more points than the dimension of the space yet somehow we're concluding that they're almost orthogonal. This seems odd. The almost is what is saving us. To see why this is not surprising, note that the event of being of inner product \( \epsilon \) from any point corresponds to a cap of fixed radius, but as discussed above, the probability of any cap of fixed radius less than \( \pi/2 \) is going to vanish in probability. In particular, you might as well take the covering number \( N(\epsilon, S^{n-1}) \), as you can see from above its bounded below by \( e^{(n-1)\epsilon^2}/2 \).

An explicit construction of such a "pathological set" is as follows (I believe the buzzword is "Spherical Code"). Take \( p \) prime and consider \( Z_p[x]/\{x^{d+1} = 0\} \) (The polynomials of degree \( d \) with coefficients in \( Z_p \). Number these polynomials \( p_k \) and let \( A_k = \frac{1}{\sqrt{p}} \langle \delta_{p_k(i) = j} \rangle \). View this matrix as a vector in \( \mathbb{R}^{p^2} \), then \( \|A_k\| = 1 \). Taking the inner product between two vectors we see that \( 0 \leq A_k, A_l \leq d/p \) if \( k \neq l \) since the inner product is the number of points at which the polynomials agree, which is at most \( d \) if they're going to be different polynomials. Play with \( p \) and \( d \) to get what you want size wise. Since there are \( p^{d+1} \) polynomials of degree \( d \) we're done.

7 Extensions

This result is far more generic. Talk about Notice that if we replace \( S^N \) with \( \{\pm 1\}^{N+1} \) and \( \sigma \) with \( \mu \) the counting measure (i.e. haar measure) for the hypercube then we have that the same argument follows through with a slightly different constant in the exponential. The proof is the same just note

Lemma 18. The map \( f(x) = (v, x)/n \) is \( 2 - Lip \) on \( \mathbb{F}_2^n \) and mean 0. Where \( v = \sum e_i \)

Proof. Clearly its mean zero. To see the lipschitz-ness note that \( f(x) = \sum x_i/n = (#pos - #neg)/N = 1 - 2#neg/N = 1 - 2d_H(v, x) \) so that
\[
\frac{|f(x) - f(y)|}{d(x, y)} = \frac{2|d_H(v, y) - d_H(v, x)|}{d_H(x, y)} \leq 2
\]
so that since \( \{\mathbb{F}_2^n, \mu_n, d_H\} \) is a normal Levy family, we have the above result.

Result holds in more generic settings. Take riemannian manifolds with lower bounded ricci curvature of increasing dimension. Then by Lichnerowicz’s bound we know that this is a normal levy family. Some examples of such spaces are \( SO(n), W_k^n \) etc. In general, we just need a G-invariant metric probability space.
Proposition 19. The general result is: Let $X, d, \mu$ be a $G$ space, and let $X_i$ be independent copies on $X$ and let $x_0$ be a designated point on $X$ and let $M = \mathbb{E}_\mu d(x_0, y)$

$$\mu(|d(X_i, X_j) - M| \leq \epsilon) \geq 1 - \left(\frac{m}{2}\right) \alpha(\epsilon)$$

Corollary. Consider $W^n_k$ then $\text{diam}(W^n_k) = 4k$ and $\mathbb{E}d(x_0, Y) = 2k$ where $x_0$ is any point and $Y$ is random over the manifold. Thus

$$\mu(|d(X_i, X_j) - 2k| \geq \epsilon) \geq 1 - \left(\frac{m}{2}\right) e^{-C\epsilon^2}$$

8 Geometry gives Probabilistic Results

Consider the following question: Let $(X_t)_{t \in T}$ be some (centered) gaussian process. Can we control the maximum

$$\mathbb{E}\sup_{t \in T} X_t := \sup_{|\mathcal{F}| < \infty} \{\mathbb{E}\sup_{t \in \mathcal{F}} X_t\}$$

The idea of (I believe) Fernique, Sudakov, and others is to induce a metric structure on $T$ and study that.

Let $d(t, s) = ||X_t - X_s||$ and let $N(T, d; \epsilon)$ be the maximum number of points in an $\epsilon$-net.

Theorem 20. (Sudakov’s Minoration)

$$\epsilon(\log N(T; \epsilon))^{1/2} \leq \mathbb{E}\sup_{t \in T} X_t$$

Proof. Take $N \leq N(T, \epsilon)$ some net corresponding to it, let $y_t = \frac{\epsilon}{\sqrt{2}} g_t$ where $g_t$ are standard gaussians for $t \in N$. Then

$$||Y_t - Y_s|| = \epsilon \leq d(s, t) = ||X_t - X_s||$$

By Slepian’s lemma then gives

$$\mathbb{E}\sup_{t \in N} Y_t \leq 2\mathbb{E}\sup_{t \in N} X_t \leq 2\mathbb{E}\sup_{t \in T} X_t$$

finally, since for $N$ i.i.d centered gaussians

$$\mathbb{E}\max Y_t \geq c\sqrt{\log(N)}$$

for large $N$, we’re done.

Example 21. Let $g$ be a standard gaussian vector in $\mathbb{R}^n$ and $X_t = (g, t)$ for $t \in B(0, 1)$. Suppose that there are $N$ points in the maximal $\epsilon$ net, then the balls of radius $\epsilon$ around these points cover $B(0, 1)$. Thus $N(T, d; \epsilon) \geq \text{Vol}B(0, 1)/\text{Vol}(B(0, \epsilon)) = \epsilon^{-d}$, we so that $\epsilon(d \log \epsilon)^{1/2} \leq \mathbb{E}\sup_{t \in T} X_t$

The other side of this is given by Fernique and simplified by Talagrand (I believe).

Theorem 22. (Fernique+Talagrand) “The Generic Chaining Argument”. Let $T$ be finite, and let $t_0 \in T$, $T_0 = \{t_0\} \subseteq T_1 \subseteq \ldots \subseteq T$ where we assume that for $n$ large $T_n = T$, and $|T_n| \leq 2^n$. Let $X_t$ be centered. Then

$$\mathbb{E}\sup_{t \in T} X_t \lesssim \sup_{n \geq 0} 2^{n/2} d(t, T_n)$$

2This is to avoid measurability issues
Proof. Let $\pi_n : T \to T_n$ be a “nearest point” map. i.e. $d(\pi_n(t), t) = d(t, T_n)$. Note that

$$X_t - X_{t_0} = \sum_{n \geq 1} X_{\pi_n(t)} - X_{\pi_n-1(t)}$$

Using the gaussian concentration result from above, note that

$$P(|X_{\pi_n(t)} - X_{\pi_n-1(t)}| \geq \epsilon) \leq e^{-\frac{\epsilon^2}{2d(\pi_n(t), \pi_n-1(t))^2}}$$

so that

$$P(|X_{\pi_n(t)} - X_{\pi_n-1(t)}| \geq u \sqrt{d(\pi_n(t), \pi_n-1(t))^2}) \leq e^{-\frac{u^2}{2}}$$

Let

$$E_u = \{ \forall n \geq 1, t \in T, |X_{\pi_n(t)} - X_{\pi_n-1(t)}| \leq u \sqrt{d(\pi_n(t), \pi_n-1(t))^2}) \}$$

note that by the union bound, since there are at most $|T_n| |T_{n-1}|$ pairs $\pi_n(t), \pi_{n-1}(t)$

$$P(E_u^c) \leq \sum_{n \geq 1} 2^{2n+1} e^{-u^2/2} \leq e^{-u^2/2}$$

Let $S = \sup_t \sum_n 2^{n/2} d(\pi_n(t), \pi_{n-1}(t))$ then

$$P(X_t - X_{t_0} \geq uS) \leq P(E_u^c) \leq e^{-u^2/2}$$

so that

$$E \sup_t X_t = E \sup_t X_t - X_{t_0} = \int_0^\infty P(X_t - X_{t_0} \geq s)ds = S \int_0^\infty P(X_t - X_{t_0} \geq uS)ds \approx S$$

but $d(\pi_n(t), \pi_{n-1}(t)) \leq 2d(t, T_{n-1})$ so we’re done. 

\qed