Random Process

Jianbin Tan

1 Gaussian Processes

Definition. T is a index set and $(\Omega, \mathcal{F}, P)$ is a probability space, $(X_t)_{t\in T}$ is called a random process if $X_t \in \mathcal{F}$,$\forall\ t \in T$.

Note. (1) If X_t is measurable on $(\Omega_t, \mathcal{F}_t, P_t)$, we can assume that exist a new probability space $(\Omega, \mathcal{F}, P)$ and $X_t' \in \mathcal{F}$ s.t. $X_t' \overset{d}{=} X_t$ by Kolmogorov extension theorem.

(2)The distribution of $(X_t)_{t\in T}$ is uniquely determined by all the distributions of $(X_t)_{t\in T_0}$. T_0 is a finite subset of T.

(3) Normally, the index set T is some compact metric space with order. One may define a measurable structure on E: $(T, \mathcal{B}(T), \mu)$, $\mu(T) < \infty$, if you are interesting in the stochastic integral $\int_T X_t f(t)\mu(dt)$.

Example. Random vector, random matrix and random field.

Example. Random walk: Z_i are independent, mean zero and $X_n = \sum_{k\leq n} Z_k$.

Example. Brownian motion: $T = [0, \infty)$, $X_0 = 0$:

(a) $t_0 < t_1 < ... < t_n$, X_{t_0}, $X_{t_1} - X_{t_0}$,..., $X_{t_n} - X_{t_{n-1}}$ are independent.

(b) $t > s$, $X_t - X_s \sim N(0, t-s)$.

(c) $X_t \in C([0, \infty))$ a.s. P.

Definition. A random process $(X_t)_{t\in T}$ is called a Gaussian process if for $\forall$ finite $T_0 \subset T$, $(X_t)_{t\in T_0}$ has a normal distribution.

Note. Noticed that if $(X_t)_{t\in T}$ is a mean zero Gaussian process, then it's completely determined by $EX_sX_t := K(s,t)$.

Example. If $(X_t)_{t\in T}$ is a Brownian motion, $EX_sX_t = EX_s(X_t - X_s) + EX_s^2 = s$ if $t > s$ $\Rightarrow K(s,t) = \min(s,t)$. We can alternatively define the Brownian motion:

(a) $(X_t)_{t\in T}$ is a mean zero Gaussian process.

(b) $X_0 = 0$, $K(s,t) = \min(s,t)$.

(c) $X_t \in C([0, \infty))$ a.s..

Definition. $(X_t)_{t\in T}$ is a mean zero Gaussian process with $K(s,t)$, then we can define a canonical metric on T: $d_X(s,t) = ||X_t - X_s||_2 = (E(X_t - X_s)^2)^{1/2}$.

Note. (1) The canonical metric may not be a metric since $d_X(s,t) = 0 \nrightarrow s = t$.

(2) Assume that $X_p = 0$, d can determine the covariance function K since $K(s,t) = \frac{d^2(t,p)+d^2(s,p)-d^2(s,t)}{2}$. Then the canonical metric completely determines the mean zero Gaussian process.

Example. Let $T \subset \mathcal{R}^n$ and $g \sim N(0, I_n)$, $X_t = \langle g, t\rangle$, $t \in T$. $(X_t)_{t\in T}$ is called canonical Guassian processes and $d_X(s,t) = ||s-t||_2$.

2 Slepian's Inequality

Lemma 1. Let $X \sim N(0,1)$ and f is a differentiable function on $\mathcal{R}$ s.t. $E|f^{'}(X)| < \infty$ and $E|Xf(X)| < \infty$, we have $Ef^{'}(X) = EXf(X)$.

Proof. Let $f_n = fI_{[-n,n]}$, then $f_n(x) \to f(x)$.

$$Ef_n^{'}(X) = \int_{\mathcal{R}} f_n^{'}(x)p(x)dx = -\int_{\mathcal{R}} f_n(x)p^{'}(x)dx = \int_{\mathcal{R}} f_n(x)p(x)xdx = EXf_n(X)$$

Let $n \to \infty$ and apply dominated convergence theorem. □

Lemma 2. Let $X \sim N(0,\Sigma)$, $f : \mathcal{R}^n \to \mathcal{R}$ and $\triangledown f$ exist, then $\Sigma\ E\triangledown f(X) = EXf(X)$.

Proof. Let $\Sigma = I$, then

$$E\partial f(X)/\partial X_i = EX_i f(X) \Rightarrow E\triangledown f(X) = EXf(X)$$

Let $Z = \Sigma^{-1/2}X$, then

$$EXf(X) = E\ \Sigma^{1/2}\ Zf(X) = \Sigma^{1/2}\ E\ \Sigma^{1/2}\ \triangledown f(X) = \Sigma\ E\triangledown f(X)$$

□

Note. $EX\triangledown f(X)^T = \Sigma\ E\triangledown^2 f(X)$ since $EX\frac{\partial f(X)}{\partial X_i} = \Sigma\ E(X\frac{\partial^2 f(X)}{\partial X_i \partial X_j})_j$.

Definition. X, Y are two Gaussian random vector valued in $\mathcal{R}^n$, we define a continuous interpolates: $Z(u) = \sqrt{u}\ X + \sqrt{1-u}\ Y$, $u \in [0,1]$.

Lemma 3. Let $X \sim N(0,\Sigma)$, $Y \sim N(0,\Lambda)$ are independent, then for all twice-differentiable $f : \mathcal{R}^n \to \mathcal{R}$, we have $\frac{d}{du}Ef(Z(u)) = \frac{1}{2}tr((\Sigma - \Lambda)\ E\triangledown^2 f(Z))$.

Proof.

$$\frac{d}{du}Ef(Z(u)) = E\triangledown f(Z)^T\ \frac{dZ}{du}(u) = \frac{1}{2}E\triangledown f(Z)^T\ (\frac{X}{\sqrt{u}} - \frac{Y}{\sqrt{1-u}})$$

And $E\triangledown f(Z)^T X$ and $E\triangledown f(Z)^T Y$ can be represented as:

$$E\triangledown f(Z)^T X = E\ tr(X\triangledown f(Z)^T) = tr(E_Y E_X X\triangledown f(Z)^T) = \sqrt{u}\ tr(\Sigma\ E\triangledown^2 f(Z))$$
$$E\triangledown f(Z)^T Y = E\ tr(Y\triangledown f(Z)^T) = tr(E_X E_Y Y\triangledown f(Z)^T) = \sqrt{1-u}\ tr(\Lambda\ E\triangledown^2 f(Z))$$

□

Theorem 1. (Slepian's Inequality, functional form) Consider two mean zero Gaussian random vectors X and Y in $\mathcal{R}^n$, $X \sim N(0, \Sigma)$, $Y \sim N(0, \Lambda)$. Then $Ef(X) \geq Ef(Y)$ if:
(a) $||X_i||_2 = ||Y_i||_2$ and $||X_i - X_j||_2 \leq ||Y_i - Y_j||_2$;
(b) $f : \mathbb{R}^n \to \mathcal{R}$ s.t. $\frac{\partial^2 f}{\partial x_i \partial x_j} \geq 0$ for all $i \neq j$.

Proof. We can supposed that X and Y are independent, if not, we can create another independent X', Y' s.t. $X \stackrel{d}{=} X'$ and $Y \stackrel{d}{=} Y'$. Let $Z(u) = \sqrt{u}\, X + \sqrt{1-u}\, Y$, noticed that:

$$\frac{d}{du} Ef(Z(u)) = \frac{1}{2} tr((\Sigma - \Lambda)\ E\nabla^2 f(Z)) \geq 0 \Rightarrow Ef(X) \geq Ef(Y)$$

□

Corollary 1. (Slepian's Inequality) $P(\max_i X_i \geq \tau) \leq P(\max_i Y_i \geq \tau)$.

Proof. Let $h_n(x) \geq 0$ is some smooth and non-increasing functions s.t. $h_n(x) \to I_{(-\infty,\tau)}(x)$. Let $f_n(x) = \prod_i h_n(x_i) \to \prod_i I_{(-\infty,\tau)}(x_i) = I_{(-\infty,\tau)}(\max_i x_i)$. Noticed that:

$$\frac{\partial^2 f_n}{\partial x_i \partial x_j} \geq 0 \Rightarrow Ef_n(X) \geq Ef_n(Y) \Rightarrow P(\max_i X_i < \tau) \geq P(\max_i Y_i < \tau)$$

□

Note. $EX = \int_0^\infty P(X > t)dt - \int_{-\infty}^0 P(X < t)dt \Rightarrow E\max_i X_i \leq E\max_i Y_i$. For two Gaussian Processes $(X_t)_{t\in T}$ and $(Y_t)_{t\in T}$, $E\sup X_t \leq E\sup Y_t$ also achieves.

Theorem 2. (Sudakov-Fernique Inequality) We drop the condition: $||X_t||_2 = ||Y_t||_2$, which means that if two pseudo-metric d_X and d_Y defined on T s.t. $d_X(s,t) \le d_Y(s,t)$, then $E \sup X_t \le E \sup Y_t$ holds.

Proof. Similarly, we just prove $E \max_i X_i \le E \max_i Y_i$, $X \sim N(0, \Sigma)$, $Y \sim N(0, \Lambda)$. Let $f(z) = \frac{1}{\beta} \log(\sum_i e^{\beta z_i})$, it's easy to show that $f(x) \to \max_i x_i$, if $\beta \to \infty$.

Let $\frac{\partial f}{\partial Z_i}(Z) = \frac{e^{\beta Z_i}}{\sum_m e^{\beta Z_m}} := P_i$ and $\frac{\partial^2 f}{\partial Z_i \partial Z_j}(Z) = \beta(\delta_{ij} P_i - P_i P_j)$. The next step is to show $tr((\Sigma - \Lambda)\ E \nabla^2 f(Z)) \ge 0$:

Since $\sum_{i,j}(EX_i^2 + EX_j^2 - EY_i^2 - EY_j^2)(\delta_{ij}P_i - P_iP_j) = \sum_{i,j}(EX_i^2 - EY_i^2 + EX_j^2 - EY_j^2)\delta_{ij}P_i - \sum_{i,j}(EX_i^2 - EY_i^2 + EX_j^2 - EY_j^2)P_iP_j = 0$

$$\begin{aligned}
\sum_{i,j}(\Sigma_{ij} - \Lambda_{ij})(\delta_{ij}P_i - P_iP_j) &= \sum_{i,j}(EY_iY_j - EX_iX_j)(\delta_{ij}P_i - P_iP_j) \\
&= \frac{1}{2}(\sum_{i,j}(E(X_i - X_j)^2 - E(Y_i - Y_j)^2)(\delta_{ij}P_i - P_iP_j) - \\
&\quad \sum_{i,j}(EX_i^2 + EX_j^2 - EY_i^2 - EY_j^2)(\delta_{ij}P_i - P_iP_j)) \\
&= -\frac{1}{2}(\sum_{i \ne j}(E(X_i - X_j)^2 - E(Y_i - Y_j)^2)P_iP_j \ge 0
\end{aligned}$$

□

Corollary 2. (Norms of Gaussian random matrices) Let $A \in M_{m \times n}(\mathcal{R})$ with iid $N(0,1)$, then $E||A||_2 \le \sqrt{m} + \sqrt{n}$.

Proof. Note that $||A||_2 = \max_{u \in S^{n-1}, v \in S^{m-1}} \langle Au, v \rangle_2$ and $\langle Au, v \rangle_2 = v^T A u = \sum_{i,j} v_i A_{ij} u_j$ is Gaussian distribution, which mean is 0 and variance is 1. We can view $||A||_2$ is a Gaussian process $\{X_{uv}\}$ in $S^{n-1} \times S^{m-1}$ and

$$\begin{aligned}
E(X_{uv} - X_{wz})^2 = E(\sum_{i,j} A_{ij}(v_i u_j - z_i w_i))^2 = \sum_{i,j}(v_i u_j - z_i w_j)^2 &= ||uv^T - wz^T||_F^2 \\
&= ||(u - w)v^T + w(v - z)^T||_F^2
\end{aligned}$$

And $||\alpha\beta^T||_F^2 = tr(\alpha\beta^T\beta\alpha^T) = \beta^T\beta\alpha^T\alpha \Rightarrow ||(u-w)v^T + w(v-z)^T||_F^2 = ||u-w||_2^2 + ||v-z||_2^2 + 2(1 - v^T z)(w^T u - 1) \le ||u-w||_2^2 + ||v-z||_2^2$.

Let $Y_{uv} = \langle g, u \rangle_2 + \langle h, v \rangle_2$, $g \sim N(0, I_n)$, $h \sim N(0, I_m)$ independent. Then

$$\begin{aligned}
E(Y_{uv} - Y_{wz})^2 = ||u-w||_2^2 + ||v-z||_2^2 &\Rightarrow E||A||_2 \le E \sup\{\langle g, u \rangle_2 + \langle h, v \rangle_2\} \\
&\le E||g||_2 + E||h||_2 = (E||g||_2^2)^{1/2} + (E||h||_2^2)^{1/2} = \sqrt{n} + \sqrt{m}
\end{aligned}$$

□

Lemma 4. $\{X_n\}_{n\leq N}$ iid standard normal distribution, then $E\max_n X_n \asymp \sqrt{\log N}$.

Corollary 3. (Sudakov's Minoration Inequality) $(X_t)_{t\in T}$ is a mean zero Gaussian process and recall $\mathcal{N}(E,d,\varepsilon)$ is the smallest covering number of T by ε-net. Then

$$E\sup X_t \geq c\varepsilon\sqrt{log\ \mathcal{N}(T,d,\varepsilon)}$$

Proof. If (T,d) is totally bounded, recall the ε-separated subset E of T: $d(x,y)>\varepsilon$, $x,y\in E$. And $\mathcal{N}$ denotes the maximal number of $|E|$, then $\mathcal{N}\geq\mathcal{N}(T,d,\varepsilon)$ and $E(X_t-X_s)^2=d_X^2(s,t)\geq\varepsilon^2$, if $s,t\in E$. Let $Y_t=\frac{\varepsilon}{\sqrt{2}}g_t$, $g_t\sim N(0,1)$, then $d_Y(s,t)=\varepsilon^2$:

$$E\sup_{t\in T}X_t\geq E\sup_{t\in E}X_t\geq E\sup_{t\in E}Y_t=\frac{\varepsilon}{\sqrt{2}}E\sup_{t\in E}g_t\geq c\varepsilon\sqrt{\log\mathcal{N}}\geq c\varepsilon\sqrt{\log\mathcal{N}(T,d,\varepsilon)}$$

□

Example. If $(X_t)_{t\in T}$ is a canonical Gaussian process, then $\forall\ \varepsilon>0$, $E\sup_{t\in T}\langle g,t\rangle_2\geq c\varepsilon\sqrt{log\ \mathcal{N}(E,d,\varepsilon)}$, d is Euclidean distance.

3 Gaussian Width

Definition. $E \sup_{t \in T} \langle g, t \rangle_2$ above is defined as the Gaussian width of T: $\omega(T)$, which is a basic geometric quantities associated with subset $T \subset \mathcal{R}^n$.

Property 1.

(a) $\omega(T) < \infty \Leftrightarrow T$ is bounded.
(b) $U \in M_{nn}(\mathcal{R})$ is orthonormal matrix and $y \in \mathcal{R}^n$, then $\omega(UT + y) = \omega(T)$.
(c) $\omega(T) = \omega(conv(T))$.
(d) $\omega(A + B) = \omega(A) + \omega(B)$ and $\omega(aT) = |a|\omega(T)$.
(e) $\omega(T) = \frac{1}{2}\omega(T - T) = \frac{1}{2}E \sup_{x,y \in T} \langle g, x - y \rangle_2$.
(f) $\frac{1}{\sqrt{2\pi}} diam(T) \le \omega(T) \le \frac{\sqrt{n}}{2} diam(T)$.

Proof.

(a) "$\Rightarrow$": $\omega(T) < \infty \Rightarrow T$ is totally bounded $\Rightarrow T$ is bounded. "$\Leftarrow$": $\langle g, t \rangle_2 \le ||g||_2 ||t||_2$ $\Rightarrow \omega(T) \le (\sup_{t \in T} ||t||_2)\ E||g||_2$.

(b) $\langle Ut + y, g \rangle_2 = \langle t, U^T g \rangle_2 + \langle y, g \rangle_2 \Rightarrow \omega(UT + y) = E \sup_{t \in T} \langle Ut + y, g \rangle_2 = E \sup_{t \in T} \langle t, U^T g \rangle_2 = E \sup_{t \in T} \langle t, g \rangle_2$.

(c) $\langle \sum_n a_n t_n, g \rangle_2 = \sum_n a_n \langle t_n, g \rangle_2 \le \sup_{t \in T} \langle t, g \rangle_2 \Rightarrow \sup_{h \in conv(T)} \langle h, g \rangle_2 \le \sup_{t \in T} \langle t, g \rangle_2 \Rightarrow \omega(conv(T)) \le \omega(T) \Rightarrow \omega(conv(T)) = \omega(T)$.

(d) $\omega(A + B) = E \sup_{t \in A+B} \langle t, g \rangle_2 = E \sup_{t_1 \in A, t_2 \in B} \langle t_1 + t_2, g \rangle_2 = \omega(A) + \omega(B)$. $\omega(aT) = E \sup_{t \in T} \langle at, g \rangle_2 = |a| E \sup_{t \in T} \langle t, sign(a) g \rangle_2 = |a|\omega(T)$.

(e) $\omega(T - T) = 2\ \omega(T)$.

(f) $\omega(T) \ge \frac{1}{2}E|\langle g, x - y \rangle_2|$, $\forall\ x, y \in T$. Notice that $\langle g, x - y \rangle_2 \sim N(0, ||x - y||_2^2)$. Let $X \sim N(0, \sigma^2)$,

$$E|X| = \int |x| p(x) dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^\infty \exp(-\frac{x^2}{2\sigma^2}) dx^2 = \sqrt{\frac{2\sigma^2}{\pi}}$$

Then $\omega(T) \ge \frac{1}{\sqrt{2\pi}} ||x - y||_2 \Rightarrow \omega(T) \ge \frac{1}{\sqrt{2\pi}} diam(T)$
$\omega(T) \le \frac{1}{2} \sup_{x,y \in T} ||x - y||_2\ E||g||_2 \le \frac{\sqrt{n}}{2} \sup_{x,y \in T} ||x - y||_2$. □

Lemma 5. If $g \sim N(0, I)$, then $r = ||g||_2$ and $\theta := \frac{g}{r}$ are independent.

Note. If $g \sim N(0, I_n)$, $g = r\theta$. Moreover, $r \to \sqrt{n}$ and $\theta \sim Unif(S^{n-1})$. We can approximately think $N(0, I_n) \approx \sqrt{n}\ Unif(S^{n-1})$.

Definition. $\omega_s(T) := E \sup_{t \in T} \langle \theta, t \rangle_2 = \frac{1}{2} E \sup_{x,y \in T} \langle \theta, x - y \rangle_2$, $\theta \sim Unif(S^{n-1})$.

Note. $\omega(T) = E \sup_{t \in T} \langle t, r\theta \rangle_2 = Er\ \omega_s(T) \approx \sqrt{n}\ \omega_s(T)$. $\omega_s(T)$ indeed is the mean width of T.

Example. Let $B_p^n = \{x \in \mathcal{R}^n; ||x||_p \leq 1\}$, $\omega_s(B_p^n) = \int_{S^{n-1}} \frac{1}{||\theta||_p} d\theta$:

$$\omega(B_p^n) = \int_0^\infty r\ P(dr) \int_{S^{n-1}} \frac{1}{||\theta||_p} d\theta = E \frac{||g||_2^2}{||g||_q}$$

If $p = 2$, $\omega(B_p^n) = Er$.

Definition. We define a s squared version of Gaussian width $h(T) := \sqrt{E \sup_t \langle g, t \rangle_2^2}$, $g \sim N(0, I_n)$.

Property 2. $2\omega(T) \leq h(T - T) \leq 2C\ \omega(T)$.

Definition. For a bounded $T \in \mathcal{R}^n$, the statistical dimension of T: $d(T) = \frac{h(T-T)^2}{diam(T)^2} \asymp \frac{\omega(T)^2}{diam(T)^2}$.

Theorem 3. $d(T) \leq \dim(T)$.

Proof.

$h(T - T)^2 = E \sup_{t,s \in T} \langle g, t - s \rangle_2^2 \leq E \sup_{z \in diam(T) B_2^{\dim(T)}} \langle g, z \rangle_2^2 = diam(T)^2 E \sup_{z \in B_2^{\dim(T)}} \langle g, z \rangle_2^2 = diam(T)^2 E \sup_{z \in B_2^{\dim(T)}} ||g||_2^2 \langle \theta, z \rangle_2^2 = diam(T)^2\ \dim(T)$. □

Note. If T is a Ball, $d(T) = dim(T)$.

Example. $A \in M_{mn}(\mathcal{R})$, then

$$h(AB_2^n - AB_2^n) = E \sup_{t,s \in AB_2^n} \langle g, t - s \rangle_2^2 = n\ E \sup_{t,s \in AB_2^n} \langle \theta, t - s \rangle_2^2$$

Noticed that $\partial AB_2^n = \{x \in \mathcal{R}^n; \sum_{i \leq n} \frac{x_i^2}{\lambda_i^2} = 1\}$, λ_i is the singular value of $A \Rightarrow h(AB_2^n - AB_2^n) = 4n \int_{S^{n-1}} \frac{1}{\sum_{i \leq n} \frac{\theta_i^2}{\lambda_i^2}} d\theta = 4||A||_F^2 \Rightarrow d(AB_2^n) = \frac{||A||_F^2}{||A||_2^2}$.

Note. $r(A) = d(AB_2^n)$ is the stable rank of A.

4 Random Projection s of sets

$G_{n,m}$ is the collection of all the m dimensional subspace of $\mathcal{R}^n$ and P is a random projection from $\mathcal{R}^n$ onto $E \sim Unif(G_{n,m})$. And without loss of generality, we can assume that rows of P are orthonormal.

Theorem 4. *$P(diam(PT) \le C(\omega_s(T) + \sqrt{\frac{m}{n}} diam(T))) \ge 1 - 2e^{-m}$.*

Proof.

$diam(PT) = \sup_{x\in T-T} ||Qx||_2 = \sup_{x\in T-T} \max_{z\in S^{m-1}} \langle Qx, z\rangle_2$. Let $\mathcal{N} = \mathcal{N}(S^{m-1}, 1/2)$ $\Rightarrow |\mathcal{N}| \le 5^m$.

$\sup_{x\in T-T} \max_{z\in S^{m-1}} \langle Qx, z\rangle_2 \le 2 \sup_{x\in T-T} \max_{z\in \mathcal{N}} \langle Qx, z\rangle_2 = 2 \max_{z\in\mathcal{N}} \sup_{x\in T-T} \langle x, Q^T z\rangle_2$.

Let $z \sim Unif(S^{m-1})$, $E \sup_{x\in T-T} \langle x, Q^T z\rangle_2 = 2\omega_s(T)$. Note that $f(z) = \sup_{x\in T-T} \langle x, Q^T z\rangle_2$ is Lipschitz function and its norm is $diam(T)$.

$$P(\sup_{x\in T-T} \langle x, Q^T z\rangle_2 \ge 2\omega_s(T) + t)) \le 2\exp(-\tfrac{cnt^2}{diam(T)^2})$$

Then $P(\max_{z\in\mathcal{N}} \sup_{x\in T-T} \langle x, Q^T z\rangle_2 \ge 2\omega_s(T) + t)) \le 2|\mathcal{N}| \exp(-\frac{cnt^2}{diam(T)^2})$. Let $t = C\sqrt{\frac{m}{n}} diam(T)$, if C is large enough, $P(\frac{1}{2} diam(QT) \ge 2\omega_s(T) + C\sqrt{\frac{m}{n}} diam(T))) \le 2e^{-m}$. $\square$

Note. Let m be the phase transition:

$$m = \frac{\left(\sqrt{n} w_s(T)\right)^2}{\text{diam}(T)^2} \asymp \frac{w(T)^2}{\text{diam}(T)^2} \asymp d(T)$$

Then

$$\text{diam}(PT) \le \begin{cases} C\sqrt{\frac{m}{n}}\,\text{diam}(T), & m \ge d(T) \\ C w_s(T), & m \le d(T) \end{cases}$$