Gaussian processes

Sunday, November 26, 2017 11:19 PM

A collection of random variables $\{X_i\}_{i \in A}$ is jointly Gaussian/Normal or a Gaussian process if

$\forall S \subseteq A$, S finite, $\forall \{a_i\}_{i \in S}$,

$$\sum_{i \in S} a_i X_i \text{ is normal.}$$

If A is finite $\underline{X}$ is a Gaussian vector.

Lemma: If X is a Gaussian vector and X_i are uncorrelated (i.e. $Cov(X_i, X_j) = 1$) then X_i are independent.

Proof: Let $\mu_i = \mathbb{E} X_i$, $\sigma_i^2 = Var(X_i)$.
Let Y_i be independent, $\mathbb{E} Y_i = \mu_i$, $Var(Y_i) = \sigma_i^2$.

Multi-dimensional Characteristic Function

$$\varphi_X(\theta) = \mathbb{E} e^{i \langle \theta, X \rangle} \quad \text{for } \theta \in \mathbb{R}^d$$

Inversion formula so enough to prove

$$\forall \theta, \ \mathbb{E} e^{i \langle \theta, X \rangle} = \mathbb{E} e^{i \langle \theta, Y \rangle}.$$

$$\langle \theta, Y \rangle = \sum_i \theta_i Y_i \sim N\left(\sum \theta_i \mu_i, \sum \theta_i^2 \sigma_i^2\right)$$

$\mathbb{E} \langle \theta, X \rangle = \sum_i \mathbb{E}\, \theta_i X_i = \sum \theta_i \mu_i$

$Var \langle \theta, X \rangle = Var \left(\sum \theta_i X_i \right) = \sum_{i,j} Cov \left(\theta_i X_i, \theta_j Y_j \right)$

$= \sum_i \theta_i^2 Cov(X_i, X_i) = \sum_i \theta_i^2 \sigma_i^2$

$\Rightarrow \langle \theta, X \rangle \sim N \left(\sum \theta_i \mu_i, \sum \theta_i^2 \sigma_i^2 \right) \stackrel{d}{=} \langle \theta, Y \rangle$

$\Rightarrow \varphi_X(\theta) = \varphi_Y(\theta).$

Hence $X \stackrel{d}{=} Y$ so X_i are independent.

A $d \times d$ matrix V is <u>non-negative semi-definite</u> if $\forall \theta \in \mathbb{R}^d$, $\theta^T V \theta \geq 0$.

If $V = (V_{ij})$ $V_{ij} = Cov(X_i, X_j)$ then V is symmetric non-negative semi-definite.

Let $Y = \sum \theta_i X_i$. Then

$0 \leq Var(Y) = \sum_{i,j} Cov(\theta_i X_i, \theta_j X_j) = \sum_{i,j} \theta_i V_{ij} \theta_j$

$= \theta^T V \theta.$

<u>Theorem</u>

For each $\mu \in \mathbb{R}^d$, V $d \times d$ NSD matrix there is a unique Gaussian Vector distribution distribution $N_d(\mu, V)$ such that for $Y \sim N_d(\mu, V)$, $\mathbb{E} Y = \mu$, $Cov(Y_i, Y_j) = V_{ij}$

In particular a Gaussian vector is determined by its mean and co-variance

If V has full rank it has a density

$$f_y(y) = \frac{1}{\sqrt{(2\pi)^n \operatorname{Det} V}} \exp\left(-\tfrac{1}{2}(y-\mu)^T V^{-1}(y-\mu)\right).$$

Proof:

If $Y, \tilde{Y}$ have the same mean, covariance then

$$\langle \theta, Y \rangle, \langle \theta, \tilde{Y} \rangle \overset{d}{=} N\left(\sum \theta_i \mu_i, \theta^T V \theta\right)$$

So $\varphi_Y(\theta) = \varphi_{\tilde{Y}}(\theta)$.

Construction of Y given μ, V.

Let $Z = (Z_1, \ldots, Z_d)$ be an IID vector of $N(0,1)$.

If V is NSD there exists U orthonormal $U^T U = I$, and D diagonal such that

$$V = U^T D^2 U.$$

Let $A = DU$, $\quad Y = A^T Z + \mu$, $\quad \mathbb{E} Y = \mu$.

$$\begin{aligned} \operatorname{Cov}(Y_i, Y_j) &= \mathbb{E}\left[\left(\sum_{\ell} A^T_{i\ell} Z_\ell\right)\left(\sum_{\ell'} A^T_{j\ell'} Z_{\ell'}\right)\right] \\ &= \mathbb{E}\left[\sum_{\ell} A^T_{i\ell} A^T_{j\ell'}\right] \\ &= (A^T A)_{ij} = V_{ij}. \end{aligned}$$

If V is full rank, so is D and so is A.

A is invertible transform and has a density.

If V is not full rank then $\exists\, \theta \neq 0$ such that

$V\theta = 0$ so $Var(\theta Y) = \theta^T V \theta = 0 \Rightarrow \theta Y$ constant

If $\theta_j \neq 0$ then $Y_j = \theta_j^{-1} \sum \theta_i Y_i + C$ determined by other co-ordinates. So Y is supported on lower dimensional hyperplane.

Marginals If Y is jointly Gaussian, so is Y_S.

Conditional Distribution:

If S, T disjoint subsets of co-ordinates,

Given $Y_T = y_T$

$Y_S \mid Y_T = y_T$

$\sim N(\mu_S - V_{ST} V_{TT}^{-1}(y_T - \mu_T), V_{SS} - V_{ST} V_{TT}^{-1} V_{TS})$.

Multivariate Central Limit Theorem

If $Y_1, \ldots$ IID random vectors $\mathbb{E}Y = \mu \in \mathbb{R}^d$, $V_{ij} = Cov(Y(i), Y(j))$ then

$$\frac{\sum_{i=1}^{n} (Y_i - \mu)}{\sqrt{n}} \xrightarrow{d} N_d(0, V)$$

Proof a) Lindebergs Method b) Characteristic functions

c) Tightness:

Extract a convergent subsequence, prove that is Gaussian.

On $\mathbb{R}^d$: A sequence $\{\mu_\alpha\}_\alpha$ is tight if $\forall \varepsilon > 0 \ \exists M$ such that

$$\inf_\alpha \mu_\alpha(B_M) \geq 1-\varepsilon.$$

General Space: $\forall \varepsilon \ \exists K_\varepsilon$ compact,

$$\inf_\alpha \mu_\alpha(K_\varepsilon) \geq 1-\varepsilon.$$

Means mass not escaping to infinity.

A family $\Gamma = \{\mu_\alpha\}$ of probability measures is relatively compact if for any $\mu_1, \mu_2, \ldots \in \Gamma$ exists subsequence n_k such that

$\mu_{n_k} \to \nu$ converges weakly

(ν may not be in Γ.).

Prohorov's Theorem A family of prob measures Γ on a separable metric space is tight if and only if it is relatively compact.

We will do case of $\mathbb{R}$.

If Γ not tight, $\exists\, \varepsilon > 0$ and μ_n such that $\mu_n(B_n) \leq 1-\varepsilon$.

If $\mu_n \xrightarrow{w} \mu$, $\exists\, M$ such that $\mu(B_M) > 1-\varepsilon/2$ contradiction.

The does exist a sub-probability measure limit.

Suppose Γ is tight, $\mu_1, \ldots \in \Gamma$ with CDF $F_1, \ldots$

Diagonal Selection Argument:

Choose $q_1, q_2, \ldots$ an enumeration of $\mathbb{Q}$.

For each q_i, $F_n(q_i) \subseteq [0,1]$ must have accumulation points.

Find a subsequence $n_k^{(1)}$ such that

$$F_{n_k^{(1)}}(q_1) \to H(q_1).$$

Find a further subsequence $n_k^{(2)}$ such that

- $F_{n_k^{(2)}}(q_2) \to H(q_2)$.
- $n_k^{(2)} \subseteq n_k^{(1)}$
- $n_1^{(2)} = n_1^{(1)}$

Continue so that $n_k^{(i)}$ satisfies,

Continue so that n_k satisfies,

- $F_{n_k^{(i)}}(q_i) \to H(q_i)$
- $n_k^{(i)} \subseteq n_k^{(i-1)}$
- $n_j^{(i)} = n_j^{(i-1)}$ for $1 \leq j \leq i$.

Let $n_k = n_k^{(k)}$. Then $n_k \subseteq n_k^{(i)}$ for all i

so $F_{n_k}(q_i) \to H(q_i)$

where $H : \mathbb{Q} \to [0,1]$ is non-decreasing.

Set $\tilde{H} : \mathbb{R} \to [0,1]$ as

- $\tilde{H}(x) = \inf\{ H(q) : q > x \}$.

Then $\tilde{H}(x)$ is

a) non-decreasing

b) right continuous since if $\tilde{H}(x) < h$ then $\exists\, q > x$ with $H(q) < h$ so $\tilde{H}(y) < h$ for all $x \leq y < q$.

c) If x is a continuity point of $\tilde{H}(x)$ then $F_{n_k}(x) \to \tilde{H}(x)$.

$\forall \varepsilon > 0$, $\exists \delta$ such that $\tilde{H}(y) \in (H(x) - \varepsilon, H(x) + \varepsilon)$ for $y \in [x - \delta, x + \delta]$.

If $q \in \mathbb{Q} \cap (x - \delta, x + \delta)$ then

$F_{n_k}(q) \to H(q) < \tilde{H}(x) + \varepsilon$ ...

$F_{n_k}(q) \to H(q) \leq \tilde{H}(x) + \varepsilon$ otherwise

for $q < y < x + \delta$ $\quad H(y) > \tilde{H}(x) + \varepsilon$ so

$\Rightarrow \limsup F_{n_k}(x) \leq \tilde{H}(x) + \varepsilon$

Similarly $\liminf F_{n_k}(x) \geq \tilde{H}(x) - \varepsilon$.

d) $\lim_{x \to \infty} \tilde{H}(x) = 1$

$\forall \varepsilon > 0, \ \exists M$ such that $\mu_n(B_M) \geq 1 - \varepsilon$

$\Rightarrow F_{n_k}(M) \geq 1 - \varepsilon$

so $\tilde{H}(x) \geq 1 - \varepsilon$ for $x > M$.

e) Similarly $\lim_{x \to -\infty} \tilde{H}(x) = 0$.

Together $\tilde{H}(x)$ is a CDF of a probability measure and

$$F_{n_k} \xrightarrow{d} \tilde{H}$$

Hence μ_n is relatively compact.

Lemma: If every subsequence n_k has a further subsequence $n'_k \subseteq n_k$ such that $\mu_{n'_k} \to \mu$ then $\mu_n \to \mu$.

Proof: If $f \in C_B$, need to show $\int f d\mu_n \to \int f d\mu$.

take n_k such that

$$\limsup \int f d\mu_n = \lim \int f d\mu_{n_k}$$

Also $= \lim \int f d\mu_{n_k'} = \int f d\mu$

so $\limsup \int f d\mu_n = \int f d\mu$

Similarly $\liminf \int f d\mu_n = \int f d\mu$.

Hence $\mu_n \xrightarrow{w} \mu$.

Back to CLT:

- Let μ_n be the law of $\sum_{i=1}^{n} (Y_i - \mu) / \sqrt{n}$.
- By Chebyshev's Inequality μ_n are tight.
- Let n_k be a subsequence
  - By Prohorov $\exists\, n_k' \subseteq n_k$ such that $\mu_{n_k'} \xrightarrow{w} \nu$ for some ν.

  We will show that $\nu \sim N_d(0, V)$.

Let $Z_n \sim \mu_n$ and $Z \sim \nu$.

For any $\theta \in \mathbb{R}^d$, $\langle \theta, Z_n \rangle \xrightarrow{d} N(0, \theta^T V \theta)$ by regular CLT.

If $f \in C_B(\mathbb{R})$ then $g : \mathbb{R}^d \to \mathbb{R}$.

$g(x) = f(<\theta, x>) \in C_B(\mathbb{R}^d)$ so

$\mathbb{E}\, g(Z_{n_k'}) \to \mathbb{E}\, g(Z) = \mathbb{E}\, f(<\theta, Z>)$

$= \mathbb{E}\, f(<\theta, Z_{n_k'}>) \to \mathbb{E}\, f(N(0, \theta^T V \theta))$

Hence $\theta^T Z \stackrel{d}{=} N(0, \theta^T V \theta)$.

So Z is Gaussian (all linear combinations are Gaussian) and must be $N(0, \theta^T V \theta)$.

Hence $\mu_n \xrightarrow{d} N_d(0, V)$.