Stochastic Calculus. Lect 2

1. Correlation, Independence

If (X,Y) is Gaussian, $EX = EY = 0$

$E(XY) = 0 \iff X, Y$ are independent

This "decorrelation => independence" property is true for Gaussians, but not in general.

Proof: $f(x,y) = \frac{1}{2\pi} (\det \underline{\underline{C}})^{-1/2} e^{-\frac{1}{2}(x,y) \underline{\underline{C}}^{-1} \binom{x}{y}}$

if $E\,xy = 0 \qquad C_{12} = 0.$

$$C = \begin{bmatrix} C_{11} & 0 \\ 0 & C_{22} \end{bmatrix} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$$

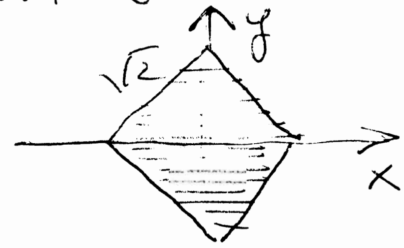
$$f(x,y) = \frac{1}{2\pi} \frac{1}{\sigma_1 \sigma_2} e^{-\frac{1}{2}\frac{x^2}{\sigma_1^2}} e^{-\frac{1}{2}\frac{y^2}{\sigma_2^2}}$$

$$= \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2} x^2/\sigma_1^2} \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2}\frac{y^2}{\sigma_2^2}}$$

$$= f_1(x) \cdot f_2(y).$$

Counterexample if (X, Y) is not Gaussian.

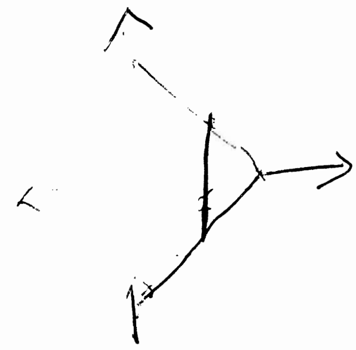
$$f(x,y) = \begin{cases} \frac{1}{2} & |x|+|y| \le 1 \\ 0 & |x|+|y| > 1 \end{cases}$$



Clearly, this is a p.density, and $EXY=0$ by symmetry (notice that $EX=EY=0$).

But X, Y are not independent.

$$f_1(x) = \frac{1}{2}\int_{-(1-|x|)}^{+(1-|x|)} du = 1-|x|.$$



$$f_2(y) = 1-|y|$$

$$f_1(x)\,f_2(y) = (1-|x|)(1-|y|) \neq \frac{1}{2}\,\chi_{(0,1)}(|x|+|y|).$$

In general, independence $\Rightarrow$ correlation $=0$

But correlation $0 \Rightarrow$ independence does not hold. It does hold for Gaussian vectors. [Exercise: find your own counterexample].

2. Conditional Expectation

Assume $E|X| < \infty$ $\quad E X^2 < \infty$.

$\mu = E(X)$ satisfies

$$\mu = \arg\min_a E\left[(X-a)^2\right]$$

The average, or mean, can be interpreted as the "constant that best approximates the r.v. X" in the mean-square sense. So the mean is the "best predictor".

We also have a similar statement for conditional predictors.

Let X, Y be two r.v.'s. Set

$$\boxed{\begin{aligned} &E[X|Y] = f(Y) \quad \text{if} \\ &f(y) = \arg\min_g E\left[(X-g(Y))^2\right] \end{aligned}}$$

This defines the <u>conditional expectation of X given Y</u>

If ~~f_{XY}~~ (X,Y) has a density $f_{12}(x,y)$

$$\min_g \iint f_{12}(x,y)\,(x-g(y))^2\,dx\,dy \qquad g_\varepsilon = g + \varepsilon h$$

$$2\iint f_{12}(x,y)\,(x - g^*(y))\,h(y)\,dx\,dy = 0$$

$$2\int_{-\infty}^{+\infty} f_{12}(x,y)(x - g^*(y))\,dx = 0$$

$$\therefore \quad g^*(y) = \frac{\int f_{12}(x,y)\,x\,dx}{\int f_{12}(x,y)\,dx}$$

f_2

$$\boxed{g^*(y) = \frac{\int f_{12}(x,y)\,x\,dx}{f_2(y)} = f_{X|Y}(y)}$$

3. Tower property:

$$E[E(X|Y)] = E(X) \quad (\text{exercise})$$

4. Conditional expectation for Gaussian variables

$$E[X|Y=y] = \frac{\int f_{12}(x,y)\, x\, dx}{\int f_{12}(x,y)\, dx}$$

$$C = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{bmatrix}$$

Assume centred variables

$$f_{12} = \frac{1}{2\pi} (\det C)^{-1/2} \quad C^{-1} =$$

$$f_{12}(x,y) = \frac{1}{2\pi} \frac{1}{\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2}\left(\frac{x_1^2}{\sigma_1^2} - \frac{2\rho x_1 x_2}{\sigma_1\sigma_2} + \frac{x_2^2}{\sigma_2^2}\right)\right\}$$

Trick: find the best linear predictor

$$\min_b E[(X - bY)^2] \qquad E(X-bY)Y = 0$$

$$\beta^* = \frac{E(xy)}{Ey^2} = \beta.$$

$E(X-\beta y)y = 0 \Rightarrow X-\beta y, y$ are ~~independent~~ independent

$$\Rightarrow E[(X-\beta y)f(y)] = E(X-\beta y)Ef(y)$$
$$= 0 \cdot Ef(y)$$
$$= 0.$$

Thus $X-\beta y$ is ~~independent~~ uncorrelated with all functions of Y. This means that

$$\boxed{E(X|Y) = \beta Y!}$$

For gaussians conditional expectation is equivalent to linear prediction

Multivariate Prediction

$(X_0 X_1, \dots X_N)$ gaussian RV.

$E X_i = 0$.

$C_{ij} = E(X_i X_j) \qquad i = 1, \dots N, \; j = 1, \dots N$

$D = C^{-1}$

$$E(X_0 \mid X_1 \dots X_N) = \sum_{i=1}^{N} \beta_i X_i$$

$$\boxed{\beta_i = \sum_{k=1}^{N} D_{ik} E[X_k, X_0]}$$

Discrete ~~time~~ time auto-regressive models

Define, recursively, $n \geq 1$ (X_0 given)

$$X_{n+1} = a + b X_n + \sigma Y_{n+1}$$

a, b, σ constants $\qquad Y_k \sim N(0,1)$ iid.

$$X_1 = a + bX_0 + \sigma \nu_1$$

$$X_2 = a + b(a + bX_0 + \sigma \nu_1) + \sigma \nu_2$$

$$= a + ab + b^2 X_0 + b\sigma \nu_1 + \sigma \nu_2$$

$$X_3 = a + \cancel{a}b\, X_2 + \sigma \nu_3$$

$$= a + b\left[a + ab + \dots + \sigma \nu_2\right] + \sigma \nu_3$$

$$X_3 = a + ab + ab^2 + b^2\sigma\nu_1 + b\sigma\nu_2 + \sigma\nu_3 + b^3 X_0$$

$$\vdots$$

$$X_n = a\left(\frac{b^n - 1}{b - 1}\right) + \sigma\left[\sum_{j=0}^{n-1} b^j \nu_{n-j}\right] + b^n X_0$$

$$= A_n + B_n$$

$$\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \begin{bmatrix} b \\ \vdots \\ b^n \end{bmatrix} X_0 + \begin{bmatrix} \\ \\ \\ \end{bmatrix} + \sigma \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ b^{n-1} & \cdots\ 1 & 0 \\ b^n\ b^{n-1} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \nu_1 \\ \vdots \\ \nu_n \end{bmatrix}$$

$(X_1, \dots X_n)$ is a <u>Gaussian process</u>

$$X_{n+1} = a + b X_n + \sigma V_{n+1}$$

is called an <u>AR(1) model</u>. It models dependence of a sequence of R.V's saying that the <u>best predictor of X_{n+1}</u> is $a + b X_n$.

$$\boxed{E[X_{n+1} \mid X_1 \ldots X_n] = a + b X_n}$$

<u>Longer tail dependence</u>

$$X_{n+1} = a + \sum_{j=1}^{m} b_j X_{n-j} + \sigma V_n$$

This kind of model is used to study <u>momentum</u> or inertia in time-series signals.

Note: AR(m) is equivalent to vector AR(1) [a vector-valued AR(1)]

Gaussian AR(m) models are the simplest time series (stochastic processes) known to scientists, and they are commonly used to model data.
Of course constancy of σ is not often a good assumption, so these models have been generalized by Engle and coll. (ARCH / GARCH models).

Going back to AR(1):

$$X_n = b^n X_0 + a\,\frac{b^n - 1}{b-1} + \sigma \sum_{k=1}^{n} b^{n-k} \nu_k$$

~~If b~~

In particolare,

$$E(X_n) = b^n X_0 + a\,\frac{b^n - 1}{b-1}$$

$$\sigma^2(X_n) = \sigma^2 \sum_{1}^{n} b^{2(n-k)} = \sigma^2\,\frac{b^{2n} - 1}{b^{2n} - 1},$$

$$\cancel{\mathrm{Cov}(X_n, X_m) = \frac{b^{n-m}}{\sigma_n \sigma_m}}$$

$$\cancel{=}$$

$$\cancel{=}$$

$$X_n = b^{n-m} X_m + a \frac{b^{n-m} - 1}{b - 1} + \sum b^j r_j$$

$$E X_n = b^{n-m} E X_m + a \frac{b^{n-m} - 1}{b - 1}$$

$$\mathrm{Cov}(X_n, X_m) = E(X_n - E X_m)(X_m - E(X_m))$$

$$= E\left[b^{n-m}(X_m - E(X_m)) + \sum b^j r_j\right]\left[X_m - E X_m\right]$$

$$= b^{n-m} E(X_m - E X_m)^2$$

$$= b^{n-m} \sigma^2 \cdot \frac{b^{2m} - 1}{b^2 - 1}$$

$$\mathrm{Corr}(X_m, X_n) = \frac{b^{n-m}}{\sqrt{\frac{b^{2n} - 1}{b^2 - 1} \cdot \frac{b^{2m} - 1}{b^2 - 1}}} = \frac{b^{n-m}(b^2 - 1)}{\sqrt{(b^{2n} - 1)(b^{2m} - 1)}}$$

If $|b| < 1$, the correlations decay

$$\rho_{nm} = \frac{b^{|n-m|}(1-b^2)}{\sqrt{(1-b^{2n})(1-b^{2m})}}$$

$$\boxed{\rho_{nm} \sim b^{|n-m|}}$$

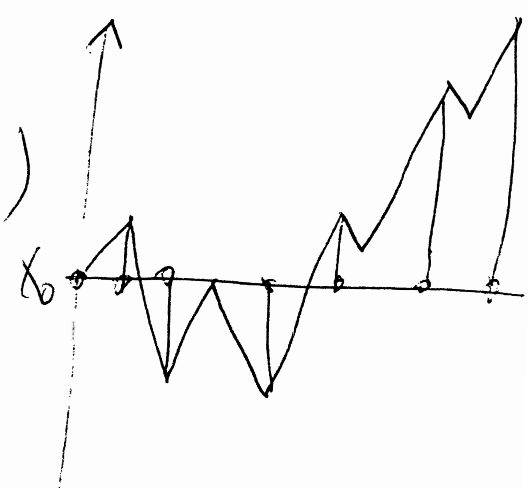
Case $b = 1$

$$X_{n+1} = a + X_n + \sigma \nu_{n+1}$$

This is a RWalk.

The discrete RW $(a=0)$

$$X_n = X_0 + \sigma \sum_{j=1}^{n} \nu_j$$



4. Brownian Motion

We construct a continuous time process which has the following properties

(i) $X(t+\delta t) - X(t)$ indep^t of $X(t)$

(ii) $X(t+\delta t) - X(t) \sim N(0, \delta t)$ $\forall\ \delta t > 0$

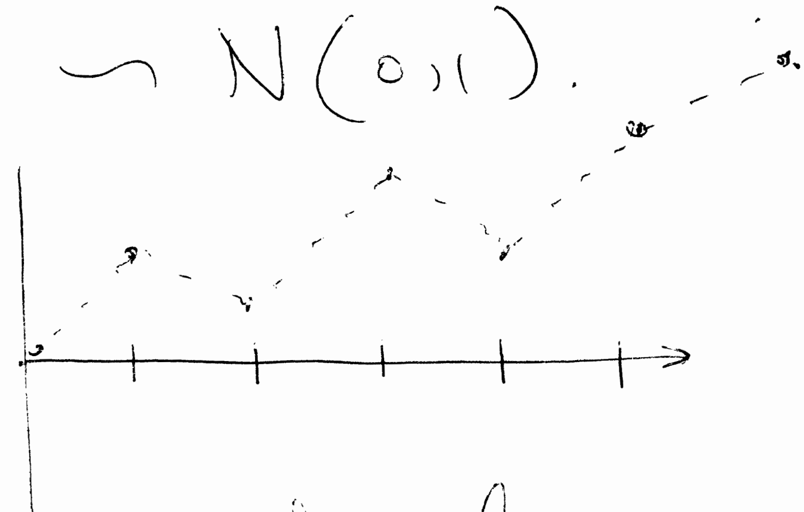
(iii) $X(t)$ is Gaussian ~~tt~~

$(X(t_1) \ldots X(t_n))$ is Gaussian for all $(t_1 \ldots t_m)$.

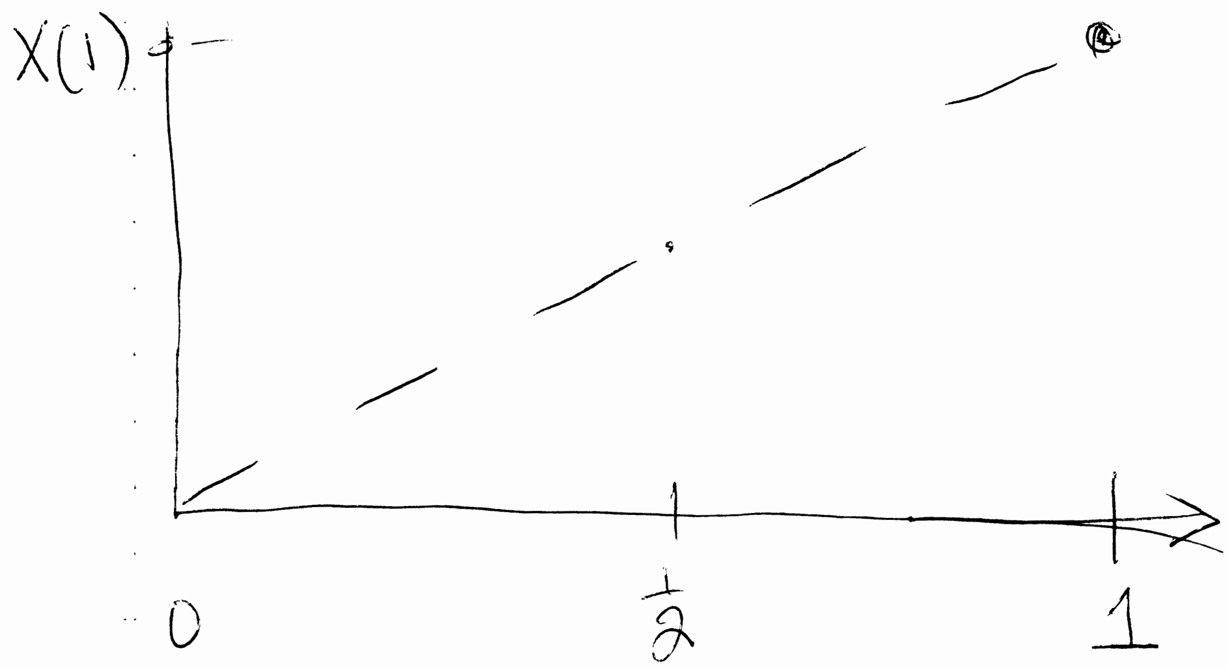
Intuition: Pick any time scale ΔT (eg $\Delta T = 1$). Then, consider

$X(n) = N_1 + N_2 + \cdots + N_n$ where

$N_i \sim N(0,1)$.



This is a discrete version of $X(\cdot)$ but we still need to define $X(t)$ $t \notin \mathbb{N}$. Consider $t \in (0,1)$



We have

$$X\left(\tfrac{1}{2}\right) = \mathbb{E}\left(X\left(\tfrac{1}{2}\right) \mid X(1)\right) + \xi_1$$

where (i) ξ_1 indept of $X(1)$. But since we expect $X(\cdot)$ to be Gaussian

$$X\left(\tfrac{1}{2}\right) = \cancel{\alpha_1 +}\, \beta_1 \cdot X(1) + \xi_1$$

$$E\left(X\left(\tfrac{1}{2}\right)\right) = 0 \qquad E\, X(1) = 0. \Rightarrow E(\xi_1) = 0.$$

$$\cancel{E\left(X(1)X\left(\tfrac{1}{2}\right)\right) =} \qquad \beta_1 = \frac{\mathrm{Cov}\left(X(1/2)\, X(1)\right)}{\mathrm{Var}(X(1))^2}$$

$$= \frac{1/2}{1} = 1/2$$

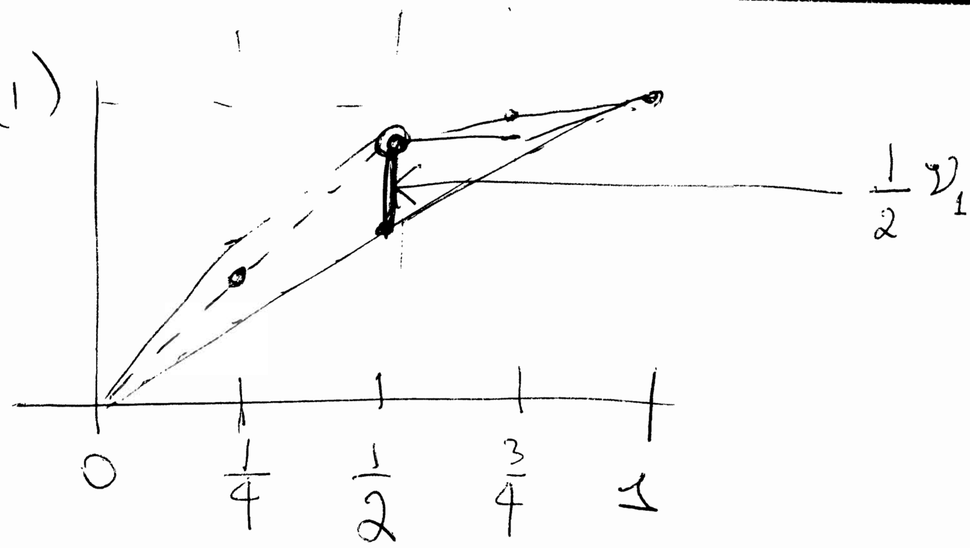
$$X\left(\tfrac{1}{2}\right) = \tfrac{1}{2} X(1) + \xi_1$$

$$\frac{1}{2} = \frac{1}{4} + \sigma^2(\xi_1)$$

$$\boxed{\sigma(\xi_1) = \frac{1}{2}}$$

(17)

$N_1 = X(1)$



$$X\left(\tfrac{1}{4}\right) = \beta_2 X\left(\tfrac{1}{2}\right) + \xi_2$$

$$E\, X\left(\tfrac{1}{4}\right) X\left(\tfrac{1}{2}\right) = \frac{1}{4} = \beta_2 \frac{1}{2} \qquad \beta_2 = \frac{1}{2}$$

~~$E\,X$~~

$$E\left(X\left(\tfrac{1}{4}\right)\right)^2 = \beta_2^2\, E\left(X\left(\tfrac{1}{2}\right)\right)^2 + E\,\xi_2^2$$

$$\frac{1}{4} = \beta_2^2 \frac{1}{2} + \sigma_2^2$$

$$\frac{1}{4} = \frac{1}{8} + \sigma_2^2$$

$$\sigma_2^2 = \frac{1}{8} \qquad \boxed{\sigma_2 = \left(\frac{1}{2}\right)^{3/2}}$$

* This gives an explicit construction of a RW on $\left\{ \frac{m}{2^N} \right\}$ for any N and $m = 0, \ldots, 2^N$ which satisfies the BM statistics.

* It is a well-known theorem that this process converges to a continuous function with probability 1. (Kolmogorov).
~~The~~ intuition is that

$$\sigma_n = \frac{1}{2^{\frac{n+1}{2}}} \sim \frac{1}{2^{n/2}} << 1$$

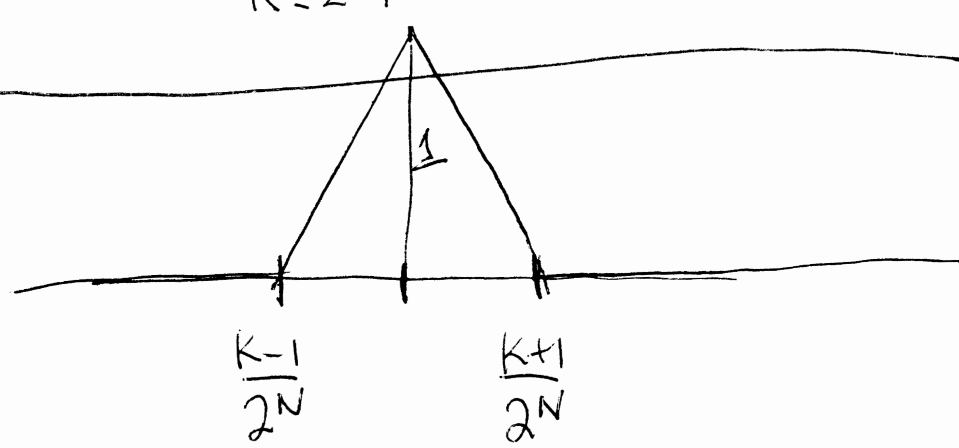
so the amounts that are added to the polygonal curve at the

n^{th} stage is $O\left(\frac{1}{2^{n/2}}\right)$.

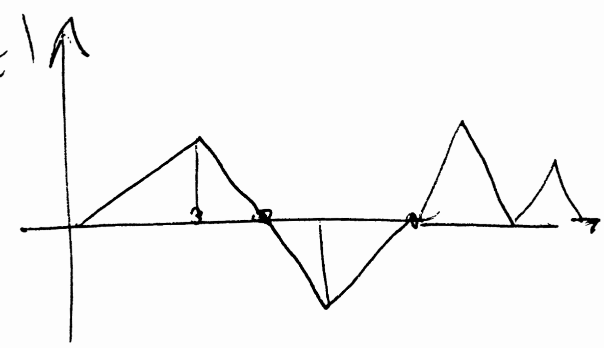
A formal proof is as follows. Let $X_N(t)$ $t \in \left\{\frac{m}{2^N};\ m = 0, \ 2^N\right\}$ be the result of the N^{th} construction. Interpolate linearly to $t \in (0,1)$.

$$\boxed{X_N(t) = \nu_0 + \sum_{j=1}^{N} \sum_{\substack{k \text{ odd} \\ k \le 2^j - 1}} \nu_{k,j} \frac{1}{2^{\frac{j+1}{2}}} T_{k,j}(t)}$$

$T_{k,j}(t) =$



$$X_N(t) = Y_0 + \sum_{j=1}^{N} Y_j(t)$$



$$P\left\{ \max_{t\le 1} |Y_j(t)| > a \right\} \le$$

$$\le \cancel{2^N}\ 2^j\, P\left[\left|Y_j\left(\tfrac{1}{2^j}\right)\right| > a\right]$$

$$\le 2^j \cdot e^{-\frac{1}{2}(2^j)\cdot a^2} \quad 2^{j/2}$$

$$2^{3j/2}\, e^{-\frac{1}{2}(2^j)a^2}$$

$$a = \frac{1}{2^{j\frac{\varepsilon}{2}}} \qquad P\left[\|Y_j\| > \frac{1}{2^{\varepsilon j}}\right] \le$$

$$2^{\frac{3j}{2}}\, e^{-\frac{1}{2} 2^j 2^{-\varepsilon j}}$$

$$P\left[\|Y_j\|_\infty > \frac{1}{2^{\varepsilon j}} \right] \leq 2^{3j/2} e^{-\frac{1}{2} 2^{j(1-\varepsilon)}}$$

$$\Rightarrow \sum_j P\left[\|Y_j\| > \frac{1}{2^{\varepsilon j}} \right] < \infty$$

By the Borel-Cantelli Lemma

$$P\left[\|Y_j\| < \frac{1}{2^{\varepsilon j}} \text{, for } j \text{ sufficiently large} \right] = 1$$

$$\Longrightarrow P\left[\sum_j \|Y_j\|_\infty < \infty \right] = 1$$

$$\Longrightarrow P\left[X(t) \text{ is continuous} \right] = 1$$

$$\boxed{X(t) = \lim_{N \to \infty} X_N(t)}$$

This construction exhibits a Gaussian process $X(t)$ such that

(i) $X(t+\delta t) - X(t)$ and $X(t)$ are independent

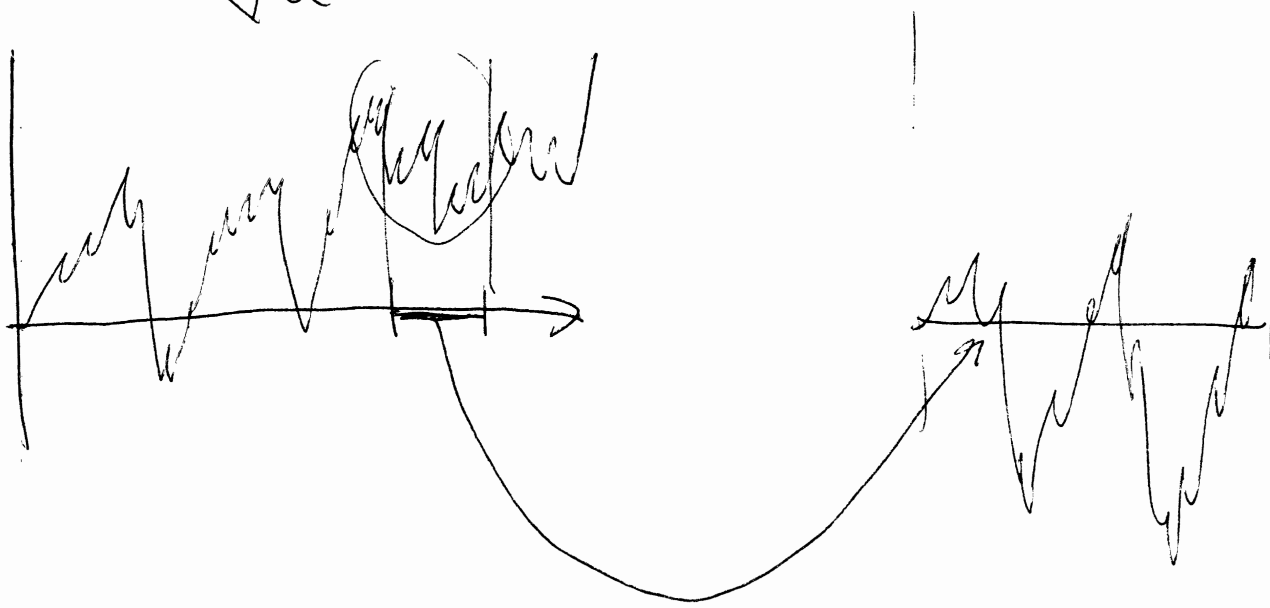
(ii) $X(t) \sim N(0, t)$

(iii) $X(\cdot)$ ~~has~~ is a continuous function of t.

Notice: the construction is the same at <u>all scales</u> (always chop up diadycally and get a sum of independent jumps) therefore $X(t)$ is <u>self-similar</u>

$\forall a > 0$

$$X_a(t) = \frac{X(at)}{\sqrt{a}}$$ is also a BM



Because of this property $X(t)$ is not differentiable (intuitively).
Basically, BM is the standard continuous, homogeneous, random walk.