Stochastic Processes

MATH5835, P. Del Moral

UNSW, School of Mathematics & Statistics

Lectures Notes, No. 4

Consultations (RC 5112):

Wednesday 3.30 pm $\rightsquigarrow$ 4.30 pm & Thursday 3.30 pm $\rightsquigarrow$ 4.30 pm

Reminder + Information

References in the slides

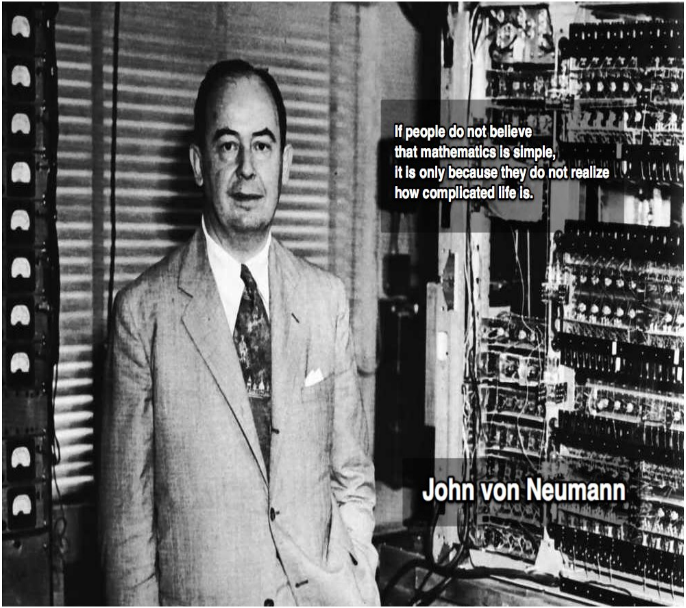
- **Material for research projects** $\rightsquigarrow$ Moodle

  (*Stochastic Processes and Applications* $\ni$ variety of applications)

- **Important results**

  $\subset$ **Assessment/Final exam** $=$ LOGO $=$

Citations of the day



J.v.N. 1903-1957

Gauss/Legendre Regression formula (1800) ⚠

- **Linear-Gaussian estimation problem**

$$\left\{\begin{array}{lcl} X & \sim & \mathrm{N}(m,\sigma^2) \\ Y & = & b\,X + V \quad (X \perp)\ V \ \sim\ \mathrm{N}(0,\tau^2) \end{array}\right.$$

- **Solution ?** $\rightsquigarrow X \mid Y$

Gauss/Legendre Regression formula (1800)

- **Linear-Gaussian estimation problem**

$$\left\{\begin{array}{lcl} X & \sim & \mathrm{N}(m,\sigma^2) \\ Y & = & b\,X+V \quad (X\perp)\ V\ \sim\ \mathrm{N}(0,\tau^2) \end{array}\right.$$

- **Solution ? $\rightsquigarrow$ $X \mid Y$ is a Gaussian r.v. with**

$$\mathbb{E}\,(X|Y)=\mathbb{E}(X)+\frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)}\,(Y-\mathbb{E}(Y))$$

and

$$\mathrm{Var}\,(X\mid Y)\quad=\quad\mathrm{Var}\,(X)\ \left[1-\frac{\mathrm{Cov}(X,Y)^2}{\mathrm{Var}\,(X)\,\mathrm{Var}\,(Y)}\right]$$

Gauss - Legendre formula (1800)

Gauss - Legendre formula (1800)

Gauss - Legendre formula (1800)

$$X \sim \mathrm{N}(m, \sigma^2) \quad \rightsquigarrow \quad Y = b\, X + \mathrm{N}(0, \tau^2) \qquad \Rightarrow \mathbb{E}(Y) = b\ m$$

$$\Downarrow$$

$$Y - \mathbb{E}(Y) = b\, (X - \mathbb{E}(X)) + V \quad \Rightarrow \quad \left\{ \begin{array}{rcl} \mathrm{Var}\,(Y) & = & b^2\sigma^2 + \tau^2 \\ \mathrm{Cov}(X, Y) & = & b\ \sigma^2 \end{array} \right.$$

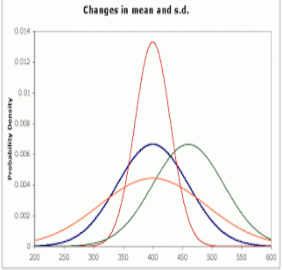
$$\Downarrow$$

$$\mathbb{E}\,(X|Y) \quad = \quad \widehat{m} := \frac{\widehat{\sigma}^2}{\sigma^2}\ m + \left(1 - \frac{\widehat{\sigma}^2}{\sigma^2}\right)\ b^{-1} Y$$

with

$$\mathrm{Var}\,(X \mid Y) \quad := \quad \widehat{\sigma}^2 = \sigma^2\ \left[1 - \frac{b^2\sigma^2}{(b^2\sigma^2 + \tau^2)}\right] = (b^2\tau^{-2} + \sigma^{-2})^{-1}$$

Linear transformations



$$X \sim \mathrm{N}(\widehat{m}, \widehat{\sigma}^2)$$

$$\Longrightarrow \ \mathrm{Law}\left(a\, X \overset{\perp}{+} \mathrm{N}(0, \sigma^2)\right) \ = \ \mathrm{N}\left(a\ \widehat{m},\ a^2\ \widehat{\sigma}^2 + \sigma^2\right)$$

Proof:

$$\left(a\ (X - \widehat{m}) + \mathrm{N}(0, \sigma^2)\right)^2$$

$$= a^2\ \underbrace{(X - \widehat{m})^2}_{\mathbb{E}(\dots)=\widehat{\sigma}^2} + \underbrace{2a\ (X - \widehat{m})\, \mathrm{N}(0, \sigma^2)}_{\mathbb{E}(\dots)=0} + \underbrace{\left(\mathrm{N}(0, \sigma^2)\right)^2}_{\mathbb{E}(\dots)=\sigma^2}$$

The Kalman filter (1960)

Linear-Gaussian $[X_0 \sim \mathrm{N}(m_0, \sigma_0^2) \perp W_n \sim \mathrm{N}(0, \sigma^2) \perp V_n \sim \mathrm{N}(0, \tau^2)]$

$$\left\{ \begin{array}{lcl} X_n & = & a\, X_{n-1} + W_n \\ Y_n & = & b\, X_n + V_n \end{array} \right.$$

Signal

Observation

The Kalman filter (1960)

Linear-Gaussian $[X_0 \sim \mathrm{N}(m_0, \sigma_0^2) \perp W_n \sim \mathrm{N}(0, \sigma^2) \perp V_n \sim \mathrm{N}(0, \tau^2)]$

$$\begin{cases} X_n &= a\, X_{n-1} + W_n \quad \textbf{Signal} \\ Y_n &= b\, X_n + V_n \quad \textbf{Observation} \end{cases}$$

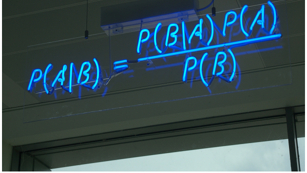
Filtering problem

$$\underbrace{\mathrm{Law}\,(X_n \mid Y_0, \ldots, Y_{n-1})}_{\text{one-step predictor}} \quad \text{and} \quad \underbrace{\mathrm{Law}\,(X_n \mid Y_0, \ldots, Y_{n-1}, Y_n)}_{\text{optimal filter}}$$

$\Updownarrow$ **Bayesian "density" notation**

$$"p(x_n \mid y_0, \ldots, y_{n-1})" \quad \text{and} \quad "p(x_n \mid y_0, \ldots, y_n)"$$

The Kalman filter 1/4



The first observation Y_0:

$$X_0 \sim \mathrm{N}(m_0, \sigma_0^2) \quad \rightsquigarrow \quad Y_0 = b\, X_0 + \mathrm{N}(0, \tau^2)$$

First updating step = the Bayes' rule

$$p(x_0) = \mathrm{N}(m_0, \sigma_0) \quad \to \quad p(x_0 \mid y_0) \propto p(y_0 \mid x_0)\ p(x_0) \quad = \quad \mathrm{N}(\widehat{m}_0, \widehat{\sigma}_0)$$

with the regression formula

$$\begin{aligned} \widehat{m}_0 &:= \frac{\widehat{\sigma}_0^2}{\sigma_0^2}\ m_0 + \left(1 - \frac{\widehat{\sigma}_0^2}{\sigma_0^2}\right)\ b^{-1} Y_0 \\ \widehat{\sigma}_0^2 &= (b^2 \tau^{-2} + \sigma_0^{-2})^{-1} \end{aligned}$$

The Kalman filter 2/4

The first prediction

$$p(x_0 \mid y_0) \;\rightarrow\; \int_{x_0} \underbrace{p(x_1 \mid x_0)}_{:=\mathrm{Law}(a x_0 + W_1)} \times \underbrace{p(x_0 \mid y_0)}_{:=\mathrm{N}(\widehat{m}_0, \widehat{\sigma}_0^2)} dx_0 = p(x_1 \mid y_0)$$

$$\Downarrow$$

$$a\, \mathrm{N}(\widehat{m}_0, \widehat{\sigma}_0^2) + W_1 \;\sim\; \mathrm{N}\left[\underbrace{a\ \widehat{m}_0}_{:=m_1},\ \underbrace{a^2\ \widehat{\sigma}_0^2 + \sigma^2}_{=\sigma_1^2}\right] = \mathrm{Law}(X_1 \mid \ Y_0)$$

The Kalman filter 3/4

Second updating step = the Bayes' rule

$$p(x_1 \mid y_0) = \mathrm{N}\left(m_1, \sigma_1^2\right)$$

$$\downarrow$$

$$p(x_1 \mid y_0, y_1) \propto p(y_1 \mid x_1)\; p(x_1 \mid y_0) \;=\; \mathrm{N}(\widehat{m}_1, \widehat{\sigma}_1^2)$$

with the regression formula

$$\begin{aligned} \widehat{m}_1 &:= \frac{\widehat{\sigma}_1^2}{\sigma_1^2}\; m_1 + \left(1 - \frac{\widehat{\sigma}_1^2}{\sigma_1^2}\right)\; b^{-1} Y_1 \\ \widehat{\sigma}_1^2 &= (b^2 \tau^{-2} + \sigma_1^{-2})^{-1} \end{aligned}$$

The Kalman filter 4/4

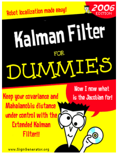
The second prediction

$$p(x_1|y_0,y_1) \;\to\; \int_{x_1} \underbrace{p(x_2 \mid x_1)}_{:=\mathrm{Law}(ax_1+W_2)} \times \underbrace{p(x_1|y_0,y_1)}_{:=\mathrm{N}(\widehat{m}_1,\widehat{\sigma}_1^2)}\, dx_1 = p(x_2|y_0,y_1)$$

$$\Downarrow$$

$$a\;\mathrm{N}(\widehat{m}_1,\widehat{\sigma}_1^2)+W_2 \;\sim\; \mathrm{N}\left[\underbrace{a\;\widehat{m}_1}_{:=m_2},\; \underbrace{a^2\;\widehat{\sigma}_1^2+\sigma^2}_{=\sigma_2^2}\right] = \mathrm{Law}(X_2|Y_0,Y_1)$$

AND SO ON ... = `KALMAN FILTER`



- ⊕ `drone board (altimeter/angle observations)`
- ⊕ `Kalman filters in mathworks`
- ⊕ Section 3.2 + Section 4.5.1 + Section 4.5.4 (& 4.5.3)

Particle Filter ⚠= `Population dynamics model`

Particle Filter ⚠ $=$ Population dynamics model

Initially

$$\left(\xi_0^1,\ldots,\xi_0^N\right) = N \text{ i.i.d. copies of } X_0 \Longrightarrow \frac{1}{N}\sum_{1\le i\le N}\delta_{\xi_0^i} \simeq_{N\uparrow\infty} \mathrm{Law}(X_0)$$

Particle Filter ⚠ = Population dynamics model

Initially

$$(\xi_0^1, \ldots, \xi_0^N) = N \text{ i.i.d. copies of } X_0 \Longrightarrow \frac{1}{N} \sum_{1 \le i \le N} \delta_{\xi_0^i} \simeq_{N \uparrow \infty} \mathrm{Law}(X_0)$$

WHY??

Particle Filter ⚠ $=$ Population dynamics model

Initially

$$(\xi_0^1, \ldots, \xi_0^N) = N \text{ i.i.d. copies of } X_0 \Longrightarrow \frac{1}{N}\sum_{1\leq i\leq N} \delta_{\xi_0^i} \simeq_{N\uparrow\infty} \mathrm{Law}(X_0)$$

WHY??

The law of large numbers: $\forall f : \mathbb{R} \mapsto \mathbb{R}$ (ex.: $f = 1_{[a,b]}$)

$$\frac{1}{N}\sum_{1\leq i\leq N} f(\xi_0^i) \simeq_{N\uparrow\infty} \mathbb{E}(f(X_0)) = \int f(x_0)\, p(x_0)\, dx_0$$

Exercise:

$$\mathbb{E}\left(\frac{1}{N}\sum_{1\leq i\leq N} f(\xi_0^i)\right) = \ldots?? \qquad \mathbb{E}\left(\left[\frac{1}{N}\sum_{1\leq i\leq N} f(\xi_0^i) - \mathbb{E}(f(X_0))\right]^2\right) = \ldots??$$

Integral representation ⚠

Dirac measure δ_a at some point a:

$$\int f(x)\ \delta_a(dx) \ = \ f(a)$$

$$\Downarrow$$

$$\frac{1}{N}\sum_{1\leq i\leq N} f(\xi_0^i) \ = \ \frac{1}{N}\sum_{1\leq i\leq N}\int f(x_0)\ \delta_{\xi_0^i}(dx_0)$$

Integral representation ⚠

Dirac measure δ_a at some point a:

$$\int f(x)\ \delta_a(dx)\ =\ f(a)$$

$$\Downarrow$$

$$\begin{aligned}
\frac{1}{N}\sum_{1\leq i\leq N} f(\xi_0^i) &= \frac{1}{N}\sum_{1\leq i\leq N}\int f(x_0)\ \delta_{\xi_0^i}(dx_0) \\
&= \int f(x_0)\ \left[\frac{1}{N}\sum_{1\leq i\leq N}\delta_{\xi_0^i}\right](dx_0) \\
&\simeq \int f(x_0)\ \ p(x_0)\ dx_0
\end{aligned}$$

A note on likelihood functions ⚠

$$Y_n = h_n(X_n) + \mathrm{N}(0, \tau^2)$$

$$\Downarrow$$

$$\begin{aligned} p(y_n \mid x_n) &= \frac{1}{\sqrt{2\pi\tau^2}}\, e^{-\frac{1}{2\tau^2}(y_n - h_n(x_n))^2} \\ &= \text{Large values iff } h_n(x_n) \text{ close to the observation } y_n \end{aligned}$$

First updating w.r.t. Y_0 = Selection of individuals

$$p(x_0)dx_0 \simeq \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_0^i}(dx_0)$$

$$\Downarrow$$

$$p(x_0 \mid y_0)dx_0 = \frac{p(y_0 \mid x_0)\ p(x_0)dx_0}{\int\ p(y_0 \mid x_0')\ p(x_0')dx_0'}$$

First updating w.r.t. Y_0 = Selection of individuals ⚠

$$p(x_0)dx_0 \simeq \frac{1}{N} \sum_{1\leq i\leq N} \delta_{\xi_0^i}(dx_0)$$

$$\Downarrow$$

$$\begin{aligned} p(x_0 \mid y_0)dx_0 &= \frac{p(y_0 \mid x_0)\ p(x_0)dx_0}{\int\ p(y_0 \mid x_0')\ p(x_0')dx_0'} \\ &\simeq \sum_{1\leq i\leq N} \frac{p(y_0 \mid \xi_0^i)}{\sum_{1\leq j\leq N} p(y_0 \mid \xi_0^j)} \delta_{\xi_0^i}(dx_0) \end{aligned}$$

$$\Downarrow$$

Sampling N r.v. $\left(\widehat{\xi}_0^1, \ldots, \widehat{\xi}_0^N\right)$ with this law:

$$p(x_0 \mid y_0)dx_0 \simeq \frac{1}{N} \sum_{1\leq i\leq N} \delta_{\widehat{\xi}_0^i}(dx_0)$$

The first prediction ⚠

$$p(x_1 \mid y_0) \quad := \quad \int_{x_0} p(x_1 \mid x_0) \;\times\; \underbrace{p(x_0 \mid y_0)}_{\simeq \frac{1}{N}\sum_{1\le i\le N}\delta_{\widehat{\xi}_0^i}} \; dx_0$$

$\simeq$ *Law of the first state of a Markov chain starting at some* $\widehat{\xi}_0^i$ *with probability* $1/N$, *with* $i = 1, \ldots, N$

$$\Downarrow$$

Sample N transitions $X_0 \rightsquigarrow X_1$ starting at $\widehat{\xi}_0^1, \ldots, \widehat{\xi}_0^N$

$$p(x_1 \mid y_0)\; dx_1 \;\simeq\; \frac{1}{N}\sum_{1\le i\le N} \delta_{\xi_1^i}(dx_1) \quad \text{with} \quad \xi_1^i \;\sim\; p(x_1 \mid \widehat{\xi}_0^i)$$

Second updating w.r.t. Y_1 = Selection of individuals ⚠

$$p(x_1 \mid y_0)\ dx_1 \ \simeq \ \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_1^i}(dx_1)$$

$$\Downarrow$$

$$p(x_1|y_0, y_1)dx_1 \quad = \quad \frac{p(y_1 \mid x_1)\ p(x_1 \mid y_0)\ dx_1}{\int\ p(y_1 \mid x_1')\ p(x_1' \mid y_0)\ dx_1'}$$

Second updating w.r.t. Y_1 = Selection of individuals

$$p(x_1 \mid y_0)\ dx_1 \ \simeq\ \frac{1}{N}\sum_{1\leq i\leq N}\delta_{\xi_1^i}(dx_1)$$

$$\Downarrow$$

$$\begin{aligned} p(x_1|y_0,y_1)dx_1 \quad &= \quad \frac{p(y_1 \mid x_1)\ p(x_1 \mid y_0)\ dx_1}{\int\ p(y_1 \mid x_1')\ p(x_1' \mid y_0)\ dx_1'} \\ &\simeq \quad \sum_{1\leq i\leq N}\frac{p(y_1 \mid \xi_1^i)}{\sum_{1\leq j\leq N} p(y_1 \mid \xi_1^j)}\delta_{\xi_1^i}(dx_1) \end{aligned}$$

$$\Downarrow$$

Sampling N r.v. $\left(\widehat{\xi}_1^1,\ldots,\widehat{\xi}_1^N\right)$ with this law:

$$p(x_1|y_0,y_1)dx_1 \simeq \frac{1}{N}\sum_{1\leq i\leq N}\delta_{\widehat{\xi}_1^i}(dx_1)$$

The second prediction ⚠

$$\color{red}{p(x_2|y_0,y_1)} := \int_{x_1} p(x_2 \mid x_1) \times \underbrace{\color{blue}{p(x_1|y_0,y_1)}}_{\simeq \frac{1}{N}\sum_{1\le i\le N}\delta_{\widehat{\xi}_1^i}} \; dx_1$$

$\simeq$ *Law of the second state* X_2 *of a Markov chain given* $X_1 = \widehat{\xi}_1^i$ *with probability* $1/N$.

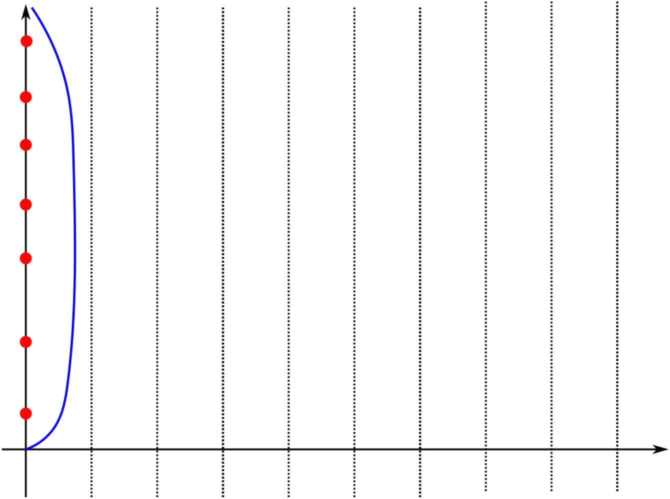
$$\Downarrow$$

Sample N transitions $X_1 \rightsquigarrow X_2$ starting at $\widehat{\xi}_1^1, \ldots, \widehat{\xi}_1^N$

$$p(x_2|y_0,y_1)\; dx_2 \simeq \frac{1}{N}\sum_{1\le i\le N}\delta_{\xi_2^i}(dx_2) \quad \text{with} \quad \xi_2^i \sim p(x_2 \mid \widehat{\xi}_1^i)$$

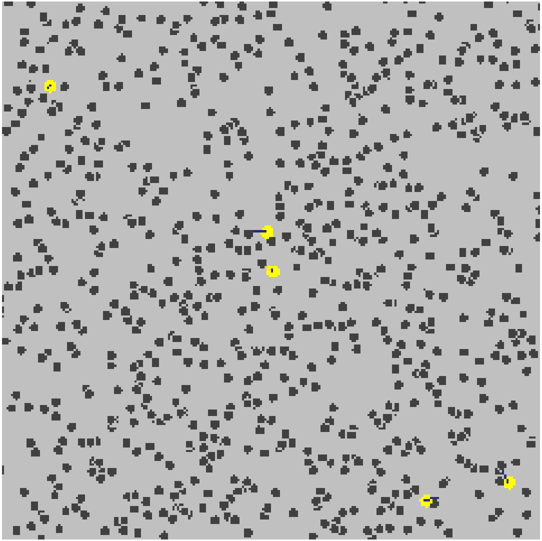
AND SO ON $\ldots =$ PARTICLE FILTER (section 3.2)

Some illustrations

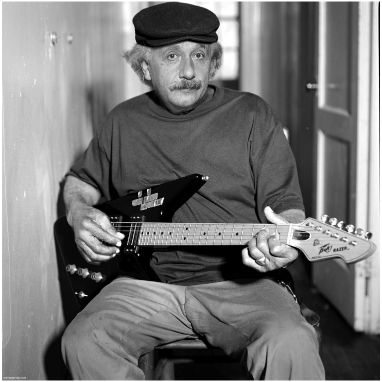


⊕ 2D localization ⊕ robot localization

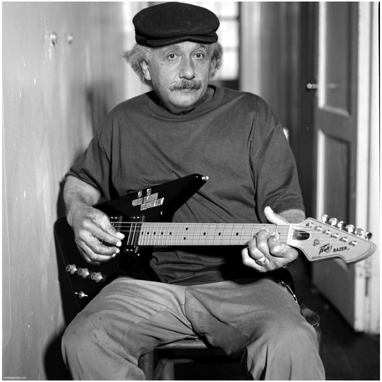
The lost equation - Introduction to Brownian motion



Second citation of the day

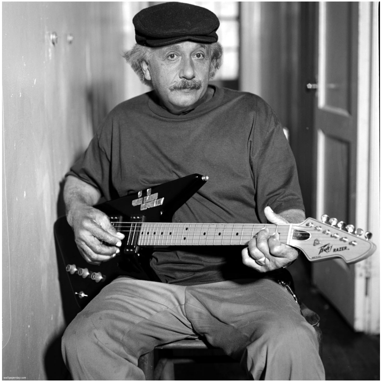


Second citation of the day



Since the mathematicians have invaded the theory of relativity, I do not understand it myself anymore. – *Albert Einstein (1879-1955)*

Second citation of the day



Since the mathematicians have invaded the theory of relativity, I do not understand it myself anymore. – *Albert Einstein (1879-1955)*

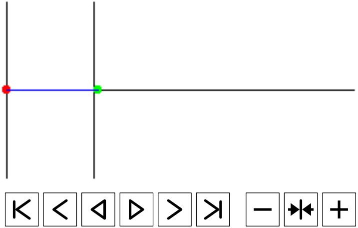
His seventh-grade teacher, exasperated by young Albert's insubordination, famously told Einstein he "would never get anywhere in life".
[A. Calaprice, The New Quotable Einstein (Princeton U.P. 2005)].

Brownian motion B_t or W_t

Discrete time version : "dt" time steps $\oplus$ fair coin tossing

$$W_t := W_{t-dt} + \begin{cases} +\sqrt{dt} & \text{if } \textbf{Heads} \\ -\sqrt{dt} & \text{if } \textbf{Tails} \end{cases}$$

or $$W_t := W_{t-dt} + \sqrt{dt} \times \mathrm{N}(0,1)$$



$\oplus$ 2d BM (video) BM-scale.avi/mp4

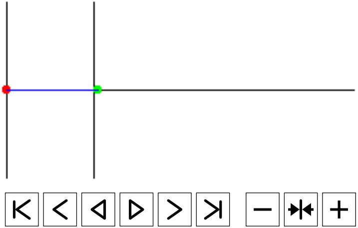
$\Downarrow$

Brownian motion B_t or W_t

Discrete time version : "dt" time steps $\oplus$ fair coin tossing

$$W_t := W_{t-dt} + \begin{cases} +\sqrt{dt} & \text{if} \quad \textbf{Heads} \\ -\sqrt{dt} & \text{if} \quad \textbf{Tails} \end{cases}$$

$$\text{or} \qquad W_t := W_{t-dt} + \sqrt{dt} \times \mathrm{N}(0,1)$$



$\oplus$ 2d BM (video) BM-scale.avi/mp4

$\Downarrow$

$dt \;=\; 10^{-10000000}??$ $\simeq_{dt\sim 0}$ **Continuous time $\oplus$ stochastic calculus**