Stochastic Calculus, Lect 2

1. Correlation, Independence

If \( (X,Y) \) is Gaussian, \( \text{EX} = EY = 0 \)
\( \text{E}(XY) = 0 \iff X, Y \) are independent.

Thus the "correlation \( \Rightarrow \) independence" property
is true for Gaussians, but not in general.

Proof: \( f(x,y) = \frac{1}{2\pi} |\text{det} \ Sigma |^{-1/2} e^{-1/2 (x-y)(\Sigma^{-1})(x-y)} \)

If \( \text{EX} = \text{EY} = 0 \)
\( \Sigma = \begin{bmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \)

\[ f(x,y) = \frac{1}{2\pi} \cdot \frac{1}{\sqrt{\sigma_1^2 \sigma_2^2}} \cdot e^{-\frac{1}{2} \frac{x^2}{\sigma_1^2} - \frac{1}{2} \frac{y^2}{\sigma_2^2}} \]

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

\[ = f_1(x) \cdot f_2(y) \]
Counterexample: If $(X, Y)$ is not Gaussian,
\[ f(x, y) = \begin{cases} \frac{1}{2} & |x + y| \leq 1 \\ 0 & |x + y| > 1 \end{cases} 
\]

Clearly, this is a probability, and $E(XY) = 0$ by symmetry (notice that $EX = EY = 0$).
But $X$ and $Y$ are not independent.

\[ f_1(x) = \frac{1}{2} \int_{-1-x}^{1-x} du = 1 - |x|, \]

\[ f_2(y) = 1 - |y| \]

\[ p(x, y) = (-x)(1-|y|) \neq \frac{1}{2} \, \delta((x, y)) \]

In general, independence $\Rightarrow$ correlation

But correlation $0$ $\Rightarrow$ independence
does not hold. It does hold for Gaussian variables. [Exercise: Find your own counterexample.]
2. Conditional Expectation

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

$\mu = E(X)$ satisfies

$\mu = \arg \min \ E[(X-\mu)^2]$.

The average, or mean, can be interpreted as the "center 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

$E[X|Y] = f(Y)$ if

$f(y) = \arg \min \ E[(X-g(Y))^2]$
This defines the conditional expectation of $X$ given $Y$

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

$\min \int \int f_{X,Y}(x-g(y))^{2}dx\,dy \quad g_{x}=g+\beta h$

$2 \int \int f_{in}(x,y)(x-g(y))h(y)\,dx\,dy = 0$

$2 \int f_{in}(x,y)(x-g(y))\,dx = 0$

$g(y) = \frac{\int f_{in}(x,y)\,dx}{\int f_{in}(x,y)\,dx}$

$g(y) = \frac{\int f_{in}(x,y)\,dx}{\int f_{in}(x,y)\,dx}$

$g(y) = \frac{\int f_{in}(x,y)\,dx}{f_{X}(y)}$
3. Tower property:
\[ E[E(X|Y)] = E(X) \] (exercise)

4. Conditional expectation for Gaussian variables:
\[ E[X|Y=y] = \frac{\int f_{12}(x,y) \times dx}{\int f_{12}(x,y) \times dx} \]

Assume constant variables
\[ f_{12} = \frac{1}{2\pi} \left( \det(C) \right)^{-\frac{1}{2}} \exp \left( \frac{-1}{2} \right) \]
\[ C = \begin{bmatrix} \sigma^2_x & \sigma_{xy} \\ \sigma_{xy} & \sigma^2_y \end{bmatrix} \]

\[ f_{12}(x,y) = \frac{1}{2\pi} \frac{1}{\sqrt{\sigma_x \sigma_y}} \exp \left\{ -\frac{1}{2} \left( \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) \right\} \]

Trick: Find the best linear predictor
\[ \min_E \left[ E[(X - \hat{y})^2] \right] \quad E(X - \hat{y}) \quad \hat{y} = \alpha \]
\[ b^* = \frac{E(xy)}{E(y^2)} = \beta. \]

\[ E[(x - \beta y)^2] = 0 \Rightarrow x - \beta y, y \text{ are independent.} \]

\[ \Rightarrow E[(x - \beta y)f(y)] = E(x - \beta y)E[f(y)] = 0 \cdot E[f(y)] = 0. \]

Thus, \( x - \beta y \) is independent uncorrelated with all functions of \( y \). This means that

\[ E(x|y) = \beta y! \]

For Gaussian random variables, conditional expectation is equivalent to linear prediction.
Multivariate Product

\[(X_0, X_1, \ldots, X_N) \text{ is a Gaussian RV.}\]

\[\mathbb{E} X_i = 0.\]

\[C_{ij} = \mathbb{E}(X_i X_j), \quad i, j = 1, \ldots, N.\]

\[D = C^{-1}\]

\[\mathbb{E}(X_0 | X_1, \ldots, X_N) = \sum_{i=1}^{N} \beta_i X_i\]

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

Discrete time auto-regressive models

Define recursively, \( n \geq 1 \), \( (X_0, \text{given} X_{-1}) \)

\[X_{n+1} = a + b \overline{X}_n + \sigma Y_{n+1}\]

\(a, b, \sigma \) constants \( Y_k \sim N(0, \sigma^2) \) iid.
\[ X_1 = a + bX_0 + \sigma V_1 \]
\[ X_2 = a + b(a + bX_0 + \sigma V_1) + \sigma V_2 \]
\[ = a + ab + b^2X_0 + b\sigma V_1 + \sigma V_2 \]
\[ X_3 = a + abX_1 + \sigma V_3 \]
\[ = a + b[a + ab + b\sigma V_1 + \sigma V_2 + \sigma V_3] + \sigma V_3 \]
\[ X_4 = a + abX_2 + b^2\sigma V_1 + b\sigma V_2 + b\sigma V_3 + \sigma^2V_4 \]
\[ X_n = a\left(\frac{b^{n-1}}{b-1}\right) + \sigma \sum_{j=0}^{n-1} b^j \sigma V_{n-j} + b^nX_0 \]

\[ \Rightarrow A_n = B_n \]

\[ \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} b \\ b^2 \\ \vdots \\ b^n \end{pmatrix} \begin{pmatrix} X_0 \\ \vdots \\ X_n \end{pmatrix} + \sigma \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} V_1 \\ \vdots \\ V_n \end{pmatrix} \]

\[ (X_1, \ldots, X_n) \text{ is a Gaussian process} \]
This kind of model is used to study non-linear or nonlinear in low-order differences.

\[ E[X_m | X_{m-1}, X_{m-2}, \ldots, X_1] = a + bX_m + \sigma^2 \]

If \( a + b \) is called an AR(1) model.

\[ X_m = a + bX_{m-1} + \sigma^2 \]

dependence of a sequence of RV's by using the last products of RV's.
Note: AR(m) is equivalent to vector AR(1) (a vector-valued AR(1)).

Covariates AR(m) models are the simplest finite ones (stochastically focused known to scientists), and they are commonly used to model data.

Of course, consistency of $\omega$ is not often a good assumption, so these models have been generalized by Engle and all. (ARCH / GARCH models).

Going back to AR(1).
\[ X_n = \theta^n X_0 + a \frac{\theta^n - 1}{\theta - 1} + \sigma \sum_{k=1}^{n} \theta^{n-k} \epsilon_k \]

**In tabular form,**

\[ E(X_n) = \theta^n X_0 + a \frac{\theta^n - 1}{\theta - 1} \]

\[ \sigma^2(X_n) = \sigma^2 \sum_{k=1}^{n} \theta^{2(n-k)} = \sigma^2 \frac{\theta^{2n} - 1}{\theta^2 - 1} \]

\[ \text{Cov}(X_n, X_m) = \frac{\theta^{n-m}}{\sigma^2} \]

\[ \frac{n}{m} \]
\[ x_n = c^{n-m} x_m + a \frac{c^{n-m} - 1}{c - 1} + \sum c^j y_j \]

\[ E x_n = b E x_m + a \frac{c^{n-m} - 1}{c - 1} \]

\[ \text{Cov} (x_n, x_m) = E (x_n - E x_n)(x_m - E x_m) \]

\[ = E [c^{n-m} (x_m - E x_m) + \sum c^j y_j] [x_m - E x_m] \]

\[ = b^{n-m} E (x_m - E x_m)^2 \]

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

\[ \text{Corr} (x_n, x_m) = \frac{b^{n-m}}{\sqrt{b^{2m} - 1} \sqrt{b^2 - 1}} \]

\[ \frac{b^{2m} - 1}{(b^2 - 1)(b^2 - 1)} \]

\[ \frac{b^{2m} - 1}{(b^2 - 1)(b^2 - 1)} \]
If $|k| < 1$, the correlation decay

$$p_{nm} = \frac{e^{(n-m)(1-k^2)}}{\sqrt{(1-k^2)(1-k^2)}}$$

$$p_{nm} \sim e^{(n-m)}$$

Case $k = 1$

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

This is a Random Walk.

The discrete RW $(a=0)$

$$X_n = X_0 + \sigma \sum_{j=1}^{n} Y_j$$
4. Brownian Motion

We construct a continuous-time process which has the following properties:

(i) \( X(t + \Delta t) - X(t) \) is independent of \( X(t) \)

(ii) \( X(t + \Delta t) - X(t) \sim N(0, \Delta t) \) for all \( \Delta t > 0 \)

(iii) \( X(t) \) is Gaussian, and \( (X(t_1), \ldots, X(t_n)) \) is Gaussian for all \( (t_1, \ldots, t_n) \).

In brief: Pick any time scale \( \Delta t \) (e.g., \( \Delta t = 1 \)). Then, consider

\[ X(n) = N_1 + N_2 + \ldots + N_n \]

where
This is a discrete version of $X(t)$ but we still need to define $X(t) + \epsilon \in \mathcal{N}$. Consider $\theta \in (0, 1)$.
We have

\[ X(\frac{1}{2}) = E(X(\frac{1}{2}) \mid X(1)) + \frac{\sigma_1}{\sqrt{2}} \]

where \( X(1) \) except of \( X(1) \). But we expect \( X(1) \) to be Gaussian

\[ X(\frac{1}{2}) = \beta_1 \cdot X(1) + \frac{\sigma_1}{\sqrt{2}} \]

\[ E(X(\frac{1}{2})) = 0 \quad E(X(1)) = 0 \quad \implies E(X(\frac{1}{2})) = 0 \]

\[ \beta_1 = \frac{\text{Cov}(X(\frac{1}{2}), X(1))}{\text{Var}(X(1))^2} \]

\[ \beta_1 = \frac{1}{2} \]

\[ X(\frac{1}{2}) = \frac{1}{2} \cdot X(1) + \frac{\sigma_1}{\sqrt{2}} \]

\[ \frac{1}{2} = \frac{1}{2} + \sigma^2(\frac{1}{2}) \]

\[ \sigma(\frac{1}{2}) = \frac{1}{2} \]
\[ N_2 = X(0) \]

\[ X(\frac{1}{4}) = \frac{1}{2} \, X(\frac{1}{2}) + \frac{1}{2} X(1) \]

\[ \mathbb{E} X(\frac{1}{4}) \, X(\frac{1}{2}) = \frac{1}{4} = \frac{1}{2} \frac{1}{2} \quad \Rightarrow \quad \sigma_2 = \frac{1}{2} \]

\[ \mathbb{E} X \]

\[ \mathbb{E} (X(\frac{1}{4}))^2 = \sigma_2^2 \mathbb{E} (X(\frac{1}{2}))^2 + \frac{1}{2} \frac{1}{2} \]

\[ \frac{1}{4} = \frac{1}{2} \sigma_2^2 + \sigma_2^2 \]

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

\[ \sigma_2^2 = \frac{1}{8} \]

\[ \sigma_2 = \left( \frac{1}{2} \right)^{\frac{1}{2}} \]
This gives an explicit construction of a RW on $\left\{ \frac{m^n}{2^N} \right\}$ for any $N$ and $M = 0$, $2^N$ 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

$$S_n = \frac{1}{2^{\frac{n}{2}}} \sim \frac{1}{2^{n/2}} \ll 1$$

so the amounts that are added to the polygonal curve at the
The $n^{th}$ stage is $O \left( \frac{1}{2^{n/2}} \right)$.

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

\[ X_N(t) = Y_0 + \sum_{j=1}^{N} \sum_{k \text{ odd}}^{j} v_{k,j} \frac{1}{2^{k+1}} T_{k,j}(t) \]

\[ T_{k,j}(t) = \begin{cases} 1 & \text{if } \frac{k-1}{2^N} \leq t < \frac{k+1}{2^N} \end{cases} \]
\[X_N(t) = y_0 + \sum_{j=1}^{N} Y_j(t)\]

\[\mathbb{P}\left( \max_{t \leq 1} |Y_j(t)| > a \right) \leq\]

\[\leq \sum_{j \geq 1} 2^j \mathbb{P}\left( |Y_j(\frac{1}{2^j})| > a \right)\]

\[= 2^j \mathbb{P}\left( |Y_j(\frac{1}{2^j})| > a \right)\]

\[= \frac{1}{2^{j/2}} e^{-\frac{1}{2}(2^j) \cdot a^2} \cdot \frac{\sqrt{2}}{2^{j/2}} - \frac{1}{2}(2^j) a^2\]

\[a = \frac{1}{2^{j/2}} \mathbb{P}\left( \|Y_j\| > \frac{1}{2^{j/2}} \right) \leq\]

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

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

By the Borel-Cantelli Lemma

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

\[
\Rightarrow P\left( \sum_j \| Y_j \|_\infty < \infty \right) = 1
\]

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

\[
X(t) = \lim_{N \to \infty} X_N(t)
\]
This contraction exhibits a Gaussian process \( X(t) \) such that

(i) \( X(t_2) - X(t_1) \) and \( X(t_1) \) are independent.

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

(iii) \( X(t) \) is a continuous function of \( t \).

Notice: the contraction is the same at all scales (always chop up diagonally and get a sum of independent jumps).

Therefore \( X(t) \) is self-similar.
If $a > 0$

$$X_a(t) = \frac{X(a + t) - X(a)}{\sqrt{a}}$$

is also a BM.

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