Stochastic process — sequence of stochastic variables. (discrete - time).

Stochastic variable: ξ is a function defined on a probability space $\{\Omega, \mathcal{A}, P\}$.

Ω is a set of "outcomes", $\mathcal{A}$ set of "events", P - probability measure.

Ex: $A_x = \{\omega \in \Omega \mid \xi(\omega) \leq x\} \in \mathcal{A}$

Expected value: $E\{\xi\} \triangleq \int_\Omega \xi \, dP$

"almost"

Hilbert space: $L^2(\Omega, \mathcal{A}, P) \triangleq \{ \xi \mid E\{|\xi|^2\} < \infty, \ \xi \text{ def. on } \{\Omega, \mathcal{A}, P\} \}$

second-order random variables.

inner product $\langle \xi, \eta \rangle = E\{\xi \bar{\eta}\}$

norm $\|\xi\| = \sqrt{\langle \xi, \xi \rangle} = \sqrt{E\{|\xi|^2\}}$

Equivalence class: $\|\xi - \eta\| = 0 \iff \xi = \eta$ P- almost everywhere ($\int |\xi - \eta|^2 dP = 0$)

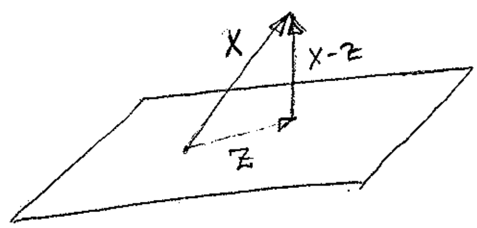
Vector generalizations: $y = [y_1 \ y_2] \Rightarrow \mathcal{Y} = \text{span}\{y_1, y_2\}$

Estimation: Let x and y be s.v. and an observation of y determine the best estimate $z = f(y)$ of x

i.e $\min_{z \in L^2(\Omega, \mathcal{Y}, P)} \|x - z\|$

$x \in \mathbb{R}^n$
$y \in \mathbb{R}^m$

Solution is given by the Orthogonal Projection Lemma 2.2.1



$L^2(\Omega, \sigma(y), P)$

The minimizing z is characterized by

$x - z \perp L^2(\sigma(y))$ i.e. $\langle x - z, y_\alpha \rangle = 0 \quad \forall y_\alpha \in \sigma(y)$

$\Rightarrow z = E[x \mid y]$ the conditional expectation.

Defined by $E\{x I_A\} = E\{z I_A\} \quad \forall A \in \mathcal{Y}$

i.e $\int_A x \, dP = \int_A z \, dP \quad \forall A \in \mathcal{Y}$

where $\mathcal{Y}$ is the σ-algebra generated by y

i.e. $\mathcal{Y} = \{\{y \in B\} \mid B \in \mathcal{A}\} = \sigma(y)$.

$z = f(y)$ is typically a non-linear fcn of y that is difficult to compute and requires knowledge of x and y's joint distribution.

If x and y have a joint Gaussian distribution, then f is linear.

Linear estimation

Let x, y be s.v. with joint covariance matrix

$$\Sigma = E\left\{ \begin{bmatrix} X - E(X) \\ Y - E(Y) \end{bmatrix} \begin{bmatrix} X - E(X) \\ Y - E(Y) \end{bmatrix}' \right\} = \begin{bmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_y \end{bmatrix}$$

Assume: $E(X) = 0$, $E(Y) = 0$ (we will assume this in general too!)

Let $H(y) \triangleq \operatorname{span}\{ y_k \mid k = 1 \dots m \}$ be the subspace linearly generated by the components of y.

Consider: $\left[\min_{z_k \in H(y)} \| x_k - z_k \| \right]$ componentwise best linear estimator of x_k

The optimal $z_k = \hat{x}_k$ is the orthogonal projection of x_k onto $H(y)$ and will be denoted by $\hat{x}_k = E[x_k \mid H(y)] = E^{H(y)} x_k$.

Prop: 2.2.3: x and y zero-mean second-order randomvectors

then $E[x \mid y] = E^{H(y)} x = \Sigma_{xy} \Sigma_y^{+} y$ Moore-Penrose pseudoinverse

Note: for the linear estimation we need only the second-order information.

Discrete-time stochastic processes

Let $y \triangleq \{ y(t) \}_{t \in \mathbb{Z}}$ where $y(t) \in L^2(\Omega, \mathcal{A}, P)$ t - usually time

first moment: mean $m(t) \triangleq E\, y(t)$

second moment: covariance $\Lambda(t,s) = E\{ (y(t) - m(t))(y(s) - m(s))^* \}$

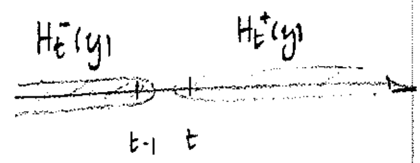
centered: If $m(t)$ known, then consider $y(t) - m(t)$ $\Rightarrow$ assume $m(t) = 0$ $\forall t$.

All processes with the same covariance function form an equivalence class.

Let $H(y) \triangleq \overline{\operatorname{span}}\{ y(t) \mid t \in \mathbb{Z} \} = \{ z = \lim_{T \to \infty} \sum_{t=-T}^{T} \sum_{k=1}^{m} a_{kt}\, y_k(t) \}$ (separable Hilbert space)

the past: $H_t^-(y) \triangleq \overline{\operatorname{span}}\{ y(s) \mid s \le t-1 \}$

the future: $H_t^+(y) \triangleq \overline{\operatorname{span}}\{ y(s) \mid s \ge t \}$



Then $H(y) = H_t^+(y) \vee H_t^-(y) = \overline{H_t^+(y) + H_t^-(y)}$

Stationarity

y is wide-sense stationary if $\Lambda(t,s) = \Lambda(t-s) \quad \forall t,s$

$y(t)$ and $y(s)$ are correlated as $y(t-s)$ and $y(0)$

(y is strict-sense stationary if $[y(0), \dots, y(n)]$ and $[y(k) \dots y(k+n)]$ have the same distributions $\forall k,n$)

Ex: Let $\{w(t)\}$ be an independent sequence of s.v. with variance 1, then $\Lambda(t,s) = \delta_{t-s}$ and w is called normalized white noise.

Define: Forward shift operator $U : H(y) \to H(y)$ s.t. $U y_k(s) = y_k(s+1)$
and $U_t : H(y) \to H(y)$ s.t. $U_t y_k(s) = y_k(s+t)$

Properties: $U_t = U^t$, stationarity $\Rightarrow \langle U_t \xi, U_t \eta \rangle = \langle \xi, \eta \rangle$ i.e. U_t isometric

U_t unitary: $U_t^* = U_t^{-1}$ and $U_t^* y_k(s) = y_k(s-t)$

Invariance: Clearly $U H_t^+(y) \subset H_t^+(y)$ (future of tomorrow included in todays future)
$U^* H_t^-(y) \subset H_t^-(y)$ (past of yesterday included in todays past)

Markov property

Let $\{X_t\}_{t \in \mathbb{Z}}$ where $X_t \subset H$ and define $X_t^- \triangleq \overline{\text{span}}\{X_s ; s \leq t\}$
$X_t^+ \triangleq \overline{\text{span}}\{X_s ; s \geq t\}$

Def: $\{X_t\}_{t\in\mathbb{Z}}$ is Markovian if $X_t^- \perp X_t^+ \mid X_t$.

Def: $A \perp B \mid X$ if $\langle \alpha - E^X \alpha, \beta - E^X \beta \rangle = 0 \quad \forall \alpha \in A, \beta \in B$.
"A is conditionally orthogonal to B given X"

$$\Rightarrow E^{X_t^-} \lambda = E^{X_t} \lambda \quad \forall \lambda \in X_t^+ \qquad \left(E^{X_t^+} \mu = E^{X_t} \mu \quad \forall \mu \in X_t^- \right)$$

proof: Follows from $A \perp B \mid X \iff E^{A \vee X} \beta = E^X \beta \quad \forall \beta \in B$. (prop 2.4.2 (iv))

$\beta - E^X\beta \perp X \Rightarrow \langle \alpha, \beta - E^X \beta \rangle = \langle E^X \alpha, \beta - E^X\beta \rangle = 0$ i.e. $\beta - E^X\beta \perp A$

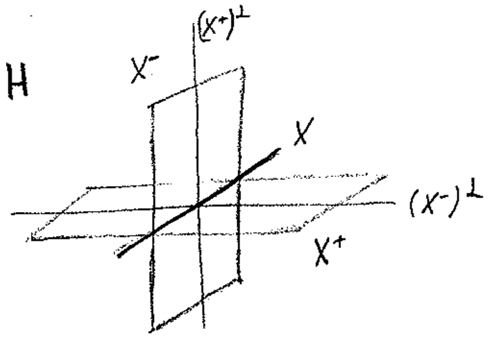
$\Rightarrow \beta - E^X \beta \perp A \vee X \Rightarrow E^{A\vee X}(\beta - E^X\beta) = 0$

$\therefore E^{A \vee X}\beta = E^{A\vee X} E^X \beta = E^X \beta$.

Stationary Markovian family

Consider Markovian $\{X_t\}_{t\in\mathbb{Z}}$, where $X_{t+s} = U_s X_t$ and U_s is a unitary shift.

Let $H \triangleq \bigvee_t X_t$ be the Ambient Hilbert space.



Then $H = (X^+)^\perp \oplus X \oplus (X^-)^\perp$

$X^- = X \oplus (X^+)^\perp$

$X^+ = X \oplus (X^-)^\perp$

Lemma A.1.6

$TX \subset X$
$\Downarrow$
$T^* X^\perp \subset X^\perp$

$U_t X^+ \subset X^+$
$\Downarrow$

Since $X^- = X \oplus (X^+)^\perp$ $\quad U_t^* X^- \subset X^-$ $\quad U_t^* (X^+)^\perp \subset (X^+)^\perp$

invariant for U_t^*

invariant for the adjoint of U_t^* in X^-

$\Rightarrow$ X is invariant for T_t, $T_t X \subset X$, where T_t is the adjoint of U_t^* in X^-.

Compressed right shift: $\quad T_t = E^{X^-} U_t \; : \; X^- \to X^- \quad t \geq 0.$ not unitary

Prop: 2.6.1.1 The family $\{X_t\}_{t\in\mathbb{Z}}$ generated by $X_t = U_t X$ is Markovian iff X is an invariant subspace for the compressed right shift

i.e $\quad E^{X^-} U|_X = E^X U|_X.$

Let $V_t = U X_t^- \ominus X_t^-$ $\Rightarrow$ $X_{t+1}^- = X_t^- \oplus V_t$ $\qquad V_t$ - "the new info" $\qquad V_s \perp V_t,\ t\neq s$

If $\xi(t) = U_t \xi$, then $\xi(t+1) = E^{X_{t+1}^-} \xi(t+1) = \underbrace{E^{X_t^-} U}_{U(X_t)} \xi(t) + E^{V_t} \xi(t+1)$

Thm 2.6.2: For any $\xi \in X$, $\xi(t) = U_t \xi$ evolves in time as

$$\xi(t+1) = U(X_t)\,\xi(t) + v_\xi(t)$$

where $v_\xi(t)$ is white noise. (stationary sequence of orthogonal s.v.).

Def: 2.6.3: A stochastic system on H is a pair (x, y) of stochastic processes such that the subspaces X_t and $H(y)$ are contained in H and

$$(H_t^-(y) \vee X_t^-) \perp (H_t^+(y) \vee X_t^+) \mid X_t \qquad t \in \mathbb{Z}$$

x is the state process $\;\Rightarrow\;$ $X_t^- \perp X_t^+ \mid X_t$ $\qquad$ Markovian

y is the output process $\;\Rightarrow\;$ $H_t^-(y) \perp H_t^+(y) \mid X_t$ $\qquad$ Splitting property

Thm 2.6.4: All finite-dimensional stochastic systems on $t \in \mathbb{Z}^+$ have a representation of the type

$$\Sigma \begin{cases} x(t+1) = A(t)\,x(t) + B(t)\,w(t) & x(0) = x_0 \\ y(t) = C(t)\,x(t) + D(t)\,w(t) \end{cases}$$

where x_0 is a s.v., $E(x_0) = 0$, w is a norm. white noise $\perp x_0$

Conversely, any pair (x, y) generated this way is a stochastic system

Proof: Read through at home!

Causality

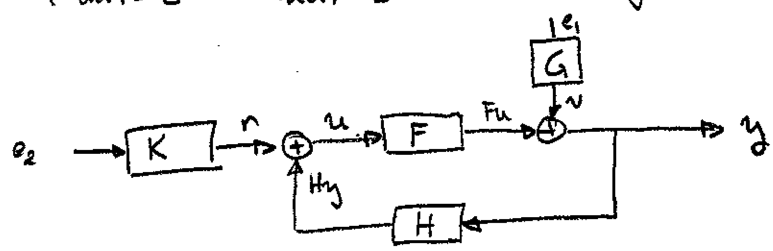
Given two stochastic processes (y, u), jointly stationary, how do we know if u is causing y? (or the opposite).

"There is causality from u to y" if $E[y(t) \mid H(u)] = E[y(t) \mid H^-_{t+1}(u)]$ $\forall t$

Future values of u will not affect the estimate of y

No feedback interpretation

Consider
$$\begin{cases} y(t) = E^{H^-_{t+1}(u)} y(t) + E^{(H^-_{t+1}(u))^\perp} y(t) = Fu + v & \text{where } v(t) \perp H^-_{t+1}(u) \\ u(t) = E^{H^-_{t+1}(y)} u(t) + E^{(H^-_{t+1}(y))^\perp} u(t) = Hy + r & \text{where } r(t) \perp H^-_{t+1}(y) \end{cases}$$



Causality $u \to y$
$\Updownarrow$
Feedback $H = 0$.

Then $y(t) = y_s(t) + y_d(t)$ where $y_s(t) \perp y_d(\tau)$ $\forall t, \tau$.

Stochastic component of y: $y_s(t) = y(t) - E^{H^-_{t+1}(u)} y(t) = y(t) - E^{H(u)} y(t) = E^{H(u)^\perp} y(t)$

Deterministic component of y: $y_d(t) = E^{H(u)} y(t)$.

(forward) innovation process $e_s(t) \triangleq y_s(t) - E^{H^-_t(y_s)} y_s(t)$

Prop 2.6.8: In the feedback-free case $e_s(t) = y(t) - E\{y(t) \mid H^-_t(y) \vee H(u)\}$