Hypothesis testing

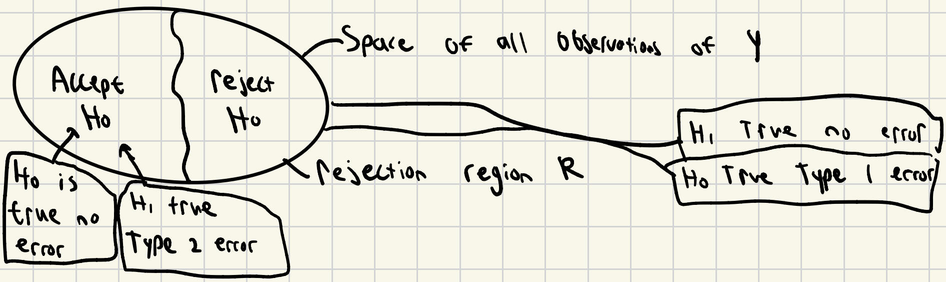
$\theta = 0$
$\theta = 1$ } diff hypothesis

H_0 null hyp, H_1 alt hyp

$P_Y(Y; H_0)$ ← likelihood under the null

1) Type 1 error α, false rejection of H_0
2) Type 2 error β, false acceptance of H_0

Goal: find a decision rate $g(Y)$ that allows to decide if H_0 is true



MAP: Declare $\theta = 1$ if $\underbrace{\underset{\text{Prior}}{P_\Theta(\theta=0)}\,\underset{\text{Conditional}}{P_{Y|\Theta}(Y|\theta=0)}}_{\text{posterior}} < P(\theta=1)\cdot P(Y|\theta=1)$

Can rewrite as

$$L(Y) = \frac{P_{Y;\theta}(Y;\theta=1)}{P_{Y;\theta}(Y;\theta=0)}$$

MAP: $L(Y) > \dfrac{P_\Theta(\theta=0)}{P_\Theta(\theta=1)} \Rightarrow \theta = 1$

MLE: $L(Y) > 1 \Rightarrow \theta = 1$

Example 6 sided die being tested

$H_0 : P_Y(Y_i; H_0) = 1/6$

$H_1 : P_Y(Y; H_1) = 1/4$ for $Y = 1,2$ and $P_Y(Y; H_1) = 1/8$ for $Y = 3,4,5,6$

Observe data:

$$L(Y) = \frac{P_Y(Y; H_1)}{P_Y(Y; H_0)} = \frac{1/4}{1/6} = 1.5 \quad \text{if } Y = 1,2$$

$$= \frac{1/8}{1/6} = .75 \quad \text{if } Y = 3,4,5,6$$

$\xi \leftarrow$ boundary

for $\xi < .75 \rightarrow$ reject H_0 for all y

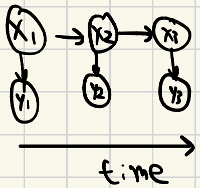
for $.75 < \xi < 1.5$ if $y = 1,2$ reject H_0

$\xi > 1.5$ accept H_0 $\forall y$

$\alpha(\xi) \rightarrow$ type I error

$\alpha($

Lecture 24



"State" follows markov property

"Observations"

$$\mathbb{P}(X_t \mid X_{t-1})$$

$$X_t = A x_t + \underbrace{w_t}_{\text{Noise}}$$

Gaussian

likelihood, prior

$$\mathbb{P}(X_t \mid X_{t-1}, Y_0, \dots, Y_t) \propto \underbrace{\mathbb{P}(Y_0, \dots, Y_t \mid X_t)}_{\text{likelihood}} \cdot \underbrace{\mathbb{P}(X_t \mid X_{t-1})}_{\text{prior}}$$

1) Jointly gaussian
2) Linear transform of indep gaussians is gaussian
3) corr = 0 $\Rightarrow$ indep
4) MMSE = LLSE

Univariate gaussians

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)/2\sigma^2} \quad \leftarrow \text{RV } X$$

$$Y = aX + b \quad , \quad \mathbb{P}(Y \le y) = \mathbb{P}(aX + b \le y) = \mathbb{P}\left(X \le \frac{y-b}{a}\right)$$

$$= F_X \rightarrow \text{Gaussian}(a\mu + b, a^2\sigma^2)$$

implies linear combos of gaussian are gaussian

$X_1 \sim N(\mu_1, \sigma_1^2)$

$X_2 \sim N(\mu_2, \sigma_2^2)$

$Y = X_1 + X_2$

$$M_X = \mathbb{E}[e^{sX}] = e^{\sigma^2 s^2/2 + \mu s} \quad \text{for gaussian } X$$

$$M_Y = \mathbb{E}[e^{s(X_1+X_2)}] = \mathbb{E}[e^{sX_1} \cdot e^{sX_2}] = \mathbb{E}[e^{sX_1}] \cdot \mathbb{E}[e^{sX_2}]$$

$$= e^{(\sigma_1^2 + \sigma_2^2)s + (\mu_1 + \mu_2)s}$$

Sum of indep gaussians is gaussian

Multivariate Gaussian

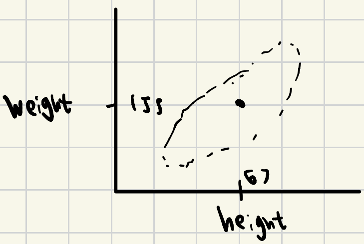
$\vec{X} \in \mathbb{R}^d$, $\vec{\mu} \in \mathbb{R}^d$ mean vector, Cov: $\Sigma \in \mathbb{R}^{d \times d}$

$$f_X(x) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} e^{-1/2 (\vec{X}-\vec{\mu})^T \Sigma^{-1} (\vec{X}-\vec{\mu})}$$

Example

X_1: height in.
X_2: weight lbs.

$$\mu: \begin{bmatrix} 67 \\ 155 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 15.5 & 104 \\ 104 & 1094 \end{bmatrix} \quad \sigma_{ij} = \sigma_{ji}$$



Jointly Gaussian

Def 1: $\vec{X} = (X_1, \dots, X_n)$ $X_i = R.V$, let $Z \in \mathbb{R}^d$, $Z_i \sim N(0,1)$ $\in 1, \dots, d$
$X_1, \dots, X_n$ jointly gaussian if $\exists\, \mu \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times d}$ s.t

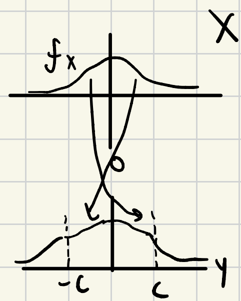
$$\vec{X} = A\vec{Z} + \vec{\mu}$$

Def 2: $X_1, \dots, X_n$ jointly gaussian if any L.C of X_i is gaussian

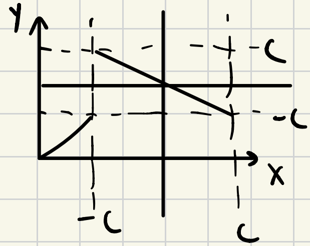
example $X \sim N(0,1)$, $Y = \frac{1}{2}X + \varepsilon$, $\varepsilon \sim N(0, .75)$, $\varepsilon \perp X$

Check: $aX + bY$ gaussian?
$= aX + b/2\, X + b\varepsilon$
$= X(a + b/2) + b\varepsilon \rightarrow$ gaussian

example
$X \sim N(0,1)$ if $|X| > c \rightarrow Y = X$
$|X| < c \rightarrow Y = -X$



Are X, Y jointly gaussian?



L.C: $X + Y = Z$

when $|X| > c$ $= 2X$ ✓

$|X| < c$ $= X + (-X) = 0$

Properties of jointly gaussians

· uncorr ⇒ indep

X_1, X_2 uncorr.

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

$$f_X(x) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} e^{-1/2 (x-\mu)^T \Sigma^{-1} (x-\mu)}$$

$$= \frac{1}{(2\pi)(\sigma_1 \sigma_2)} \cdot e^{-\frac{1}{2}[x_1-\mu_1, x_2-\mu_2] \begin{bmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{bmatrix} \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}}$$

$$f_X(x) = \frac{1}{\sqrt{2\pi^2 \sigma_1^2 \sigma_2^2}} e^{-1/2 \left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} \right)}$$

$$= f_{X_1}(x_1) \cdot f_{X_2}(x_2) \quad \therefore \quad \text{cov}(X_1, X_2) = 0 \Rightarrow \text{indep}$$

for jointly gaussian X_1, X_2

Example

$X \sim N(0,1)$

$Y = WX$, $W = \begin{cases} 1 & \text{prob}(1/2) \\ -1 & \text{prob}(1/2) \end{cases}$

$Y | W \sim$ gaussian

X, Y uncorr but not indep

$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

$$= \mathbb{E}[X^2 W] - \mathbb{E}[X]\mathbb{E}[WX]$$

$$= \mathbb{E}[X^2]\mathbb{E}[W] - 0 = 0 \quad \therefore X, Y \text{ uncorr}$$

however $\quad \mathbb{P}(X \le -1 \mid Y = 0) = 0 \quad$ and $\quad \mathbb{P}(X \le -1) \ne 0$

Property 2

L.C of jointly gaussians are jointly gaussian

let X_1, X_2 be jointly gaussian

$$\vec{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = AZ + \mu \quad , \quad A = 2[\;^{d}\;]$$

$$= \begin{bmatrix} A_1 Z + \mu_1 \\ A_2 Z + \mu_2 \end{bmatrix}$$

C, D matrices

$$U = CX_1 + c = C(A_1 Z + \mu) + c$$
$$V = DX_2 + d = D(A_2 Z + \mu) + d$$
$\Rightarrow U, V$ jointly gaussian

Property 3 of J.G

X, Y Want to estimate X w/ Y

where X, Y jointly gaussian

MMSE is LLSE

$$\mathbb{E}[\underbrace{(X - g(Y))^2}_{\text{Cost func.}}]$$

$$\mathbb{E}[(\underbrace{X - \mathbb{E}(X|Y)}_{\text{error}}) \cdot h(Y))] = 0 \rightarrow \text{orthogonality property of MMSE}$$

if you find estimator $\hat{X}$
Show $\mathbb{E}[(X-\hat{X})h(Y)]=0 \rightarrow$ $\hat{X}$ is the MMSE

LLSE: estimator is linear $\hat{X}=aY+b$

$\mathbb{E}[(X-\hat{X})\cdot Y]=0 \rightarrow$ Property of LLSE

LLSE: $\hat{X}=\mathbb{E}[X]+\frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)}(Y-\mathbb{E}(Y))$

Can be shown $\mathbb{E}[X-\hat{X}]=0$ or $\hat{X}$ is unbiased
We can

1) Show $Y, X-\hat{X}$ are uncorr
2) $Y, X-\hat{X}$ linear in X,Y
3) $Y, X-\hat{X}$ indep
4) funcs of indep R.V also indep, $g(Y), X-\hat{X}$ indep
5) All indep R.V uncorr

$\mathbb{E}((X-\hat{X})g(Y))=0$ $\forall g(Y)$ $\therefore$ $\hat{X}$ is the MMSE

Lecture 25

· Jointly gaussian def 1: $X_1, \dots, X_n$ jointly gaussian if

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} = A \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix} + \mu$$

def 2: if $\mu^T X$ is gaussian

Properties

1) linear transformations of gaussian $\rightarrow$ gaussian (MGF)
2) corr $0 \rightarrow$ indep
3) MMSE for $\mathbb{E}(X|Y)$ if X, Y gaussian is LLSE for X from Y

$$\mathbb{E}[X|Y] = aY + b$$

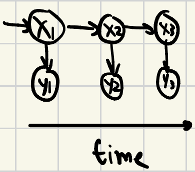
By def: $\begin{bmatrix} X \\ Y \end{bmatrix} = AZ + \mu = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \begin{bmatrix} Z \end{bmatrix} + \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$

$$X = A_1 Z + \mu_1$$
$$Y = A_2 Z + \mu_2$$

Assume A_1, A_2 square and A_2 invertible

$\mathbb{E}[X|Y]$ if $X = A_1 Z + \mu_1 = A_1 A_2^{-1}(Y - \mu_2) + \mu_1 \rightarrow$ LLSE

Kalman Filter



X_i: hidden state obeying markov property
Y_i: observation

Setup

$$X_n = AX_{n-1} + V_n, \quad V_n \sim N(0, V)$$
$$Y_n = CX_n + W_n \quad W_n \sim N(0, W)$$

$X_0 \sim N(\mu, \sigma^2)$ $\therefore X_n$ gaussian
V_n sensor noise, W_n state noise. Indep

Goal

- Estimate X_n from a series of $Y_1, \dots, Y_n$
- Filtering $\mathbb{E}[X_n | Y_1, \dots, Y_n]$
- Prediction $\mathbb{E}[X_{n+1} | Y_1, \dots, Y_n]$
- Smoothing $\mathbb{E}[X_n | Y_1, \dots, Y_n, Y_{n+1} \dots]$

1) X, Y jointly gaussian : $X_n = aX_{n-1} + V_n$ $\quad Y_n = cX_n + W_n$

$$= a(aX_{n-2} + V_{n-1}) + V_n$$

$$\vdots$$

$$X_n = a^n X_0 + \sum_{i=1}^{n} a^{n-i} V_i \qquad \therefore X_n \text{ gaussian}$$

any L.C of X, Y is gaussian

$$fX_n + gY_n \Rightarrow X_n, Y_n \text{ jointly gaussian}$$

Approach

1) Prediction Step : $\mathbb{E}[X_n | Y_1, \dots, Y_{n-1}] = \hat{X}_{n|n-1}$, $Var(X_n - \hat{X}_{n|n-1}) = \hat{\sigma}^2_{n|n-1}$

2) Innovation : $\mathbb{E}[X_n | Y_1, \dots, Y_n] = \hat{X}_{n|n}$, $Var(X_n - \hat{X}_{n|n}) = \hat{\sigma}^2_{n|n}$

Prediction (LLSE = $\mathbb{L}$)

$$\mathbb{E}[X_n | Y_1, \dots, Y_{n-1}] = \mathbb{L}(X_n | Y_1, \dots, Y_{n-1})$$

$$= \mathbb{L}(aX_{n-1} + V_n | Y_1, \dots, Y_n)$$ $V_n \perp X_1, \dots, X_n \wedge Y_1, \dots, Y_n$ so we remove it

$$= a\mathbb{L}(X_{n-1} | Y_1, \dots, Y_{n-1})$$

$$\boxed{\hat{X}_{n|n-1} = a\hat{X}_{n-1|n-1}}$$

$$\mathbb{E}[Y_n | Y_1, \dots, Y_{n-1}) = \mathbb{E}[CX_n + W_n | Y_1, \dots, Y_{n-1}] = c\hat{X}_{n|n-1} = ca\hat{X}_{n-1|n-1}$$

- $\hat{\sigma}^2_{n|n-1} = Var(\hat{X}_{n|n-1} - X_n) = Var(a\hat{X}_{n-1|n-1} - X_n) = Var(a\hat{X}_{n-1|n-1} - (aX_{n-1} + V_n))$

$$= a^2 Var(\hat{X}_{n-1|n-1} - X_{n-1}) + Var(V_n)$$

$$\boxed{\sigma^2_{n|n-1} = a^2 \hat{\sigma}^2_{n|n-1} + V}$$

Update Property of LLSE

$\mathbb{L}(X|Y,Z)$ where Y,Z orthogonal or indep if jointly gaussian or $\text{Cov}(Y,Z)=0$

$\longrightarrow \mathbb{L}(X|Y)+\mathbb{L}(X|Z)$

define residual R as

$$R = X - \mathbb{L}(X|Y)$$
$$X = R + \mathbb{L}(X|Y)$$
$$= \mathbb{L}(R|Z) + W + \mathbb{L}(X|Y)$$
$$= \mathbb{L}(X|Y) + \mathbb{L}(X|Z)$$

More generally

$$\mathbb{L}(X|Y,Z) = \mathbb{L}(X|Y) + \mathbb{L}(X|\underbrace{Z - L(Z|Y)}_{\text{innovation of } Z \,:\, \tilde{Z}})$$

where Y,Z not necessarily orthogonal

Innovation updates

$$\hat{X}_{n|n} = \mathbb{L}(X_n|Y_1,\dots,Y_n) + \mathbb{L}(X_n - Y_n - \mathbb{L}(Y_n|Y_1,\dots,Y_{n-1}))$$

$$= \hat{X}_{n|n-1} + \mathbb{L}(X_n|Y_n - C\hat{X}_{n|n-1})$$

Define $\tilde{Y}_n = Y_n - C\hat{X}_{n|n-1}$ "innovation"

$$= CX_n + W_n - C\hat{X}_{n|n-1}$$

$$= C(X_n - \hat{X}_{n|n-1}) + W_n = Ce_n + W_n \quad , \quad e_n = X_n - \hat{X}_{n|n-1}$$

So $\mathbb{L}(X_n|Y_n - C\hat{X}_{n|n-1}) = \dfrac{\text{cov}(X_n,\tilde{Y}_n)}{\text{Var}(\tilde{Y}_n)}(\tilde{Y}_n - \mathbb{E}(\tilde{Y}_n))$

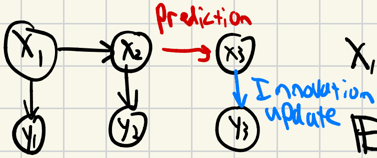
$$\text{Cov}(X_n,\tilde{Y}_n) = C\cdot\text{Cov}(X_n,e_n) + \text{Cov}(X_n,W_n)$$

$$= C\cdot\text{Cov}(e_n + \hat{X}_{n|n-1}, e_n) = C\cdot\text{cov}(e_n,e_n) = C\cdot\text{Var}(e_n) = C\sigma^2_{n|n-1}$$

$$\hat{X}_{n|n} = \underbrace{\hat{X}_{n|n-1}}_{\text{Pred update}} + \overbrace{\boxed{\frac{C\hat{\sigma}^2_{n|n-1}}{C^2\sigma^2_{n|n-1}+W}}}^{\text{Kalman gain : } K_n} \tilde{Y}_n$$

innovation : $Y_n - C\hat{X}_{n|n-1}$

Lecture 26 (Kalmann filter, HMM)



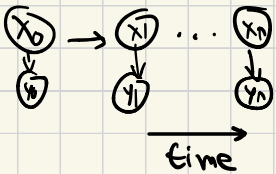
X, Y J.6

$\mathbb{E}(X_i \mid Y, \dots) \to$ LLSE

Filtering

- Pred X_i from X_{i-1}
- Innovation: update X_i from X_{i-1}, Y_0

Hidden markov models



- X_i hidden discrete, form DTMC
- Y_i can be discrete / cts R.V

- for each X_i in state j define $Q(X_i, Y_i)$ as prob of observing Y_i given X_i in a state

- π_0: initial state dist for X_0: $[\_\_\ ]$, $1 \times m$ where m is # of possible state
  ↳ $\pi_0(i) = P(X_0 \text{ in state } i)$

$$P(X_0, \dots, X_n, Y_0, \dots, Y_n) = \pi_0(X_0) \cdot Q(X_0, Y_0) \cdot \left(P(X_0, X_1) \cdot Q(X_1, Y_1) \dots \right)$$

HMM Problems

1) Likelihood: Known Y_is, P, Q, π
   How likely is a sequence of observations?

2) decoding: Know Y_is, P, Q, π
   Best sequence of X_is that explain data?

3) Param estimate: given X_is, Y_is, learn P, Q, π

Decoding problem / Viterbi algo

X^n : Sequence of Xos: $X_0, \dots, X_n$

$$X_n^* = \underset{x^n}{\operatorname{argmax}} \underline{\mathbb{P}(X^n = x^n \mid Y^n = y^n)}$$ MAP estimator

$$\underbrace{\mathbb{P}(Y^n = y^n \mid X^n = x^n)}_{\text{emissions}} \cdot \underbrace{\mathbb{P}(X^n)}_{\text{transition Prob}}$$

$\pi_0(x_0) Q(x_0, y_0) \cdots \mathbb{P}(x_{n-1}, x_n) Q(x_n, y_n)$, taking log $\rightarrow$

$$\log(\pi_0(x_0) \cdot Q(x_0, y_0)) + \dots + \log(\mathbb{P}(x_{n-1}, x_n) \cdot Q(x_n, y_n))$$

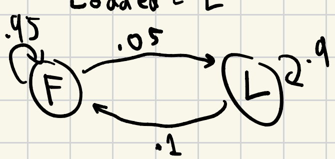
denote $d_0(x_0) = -\log(\pi_0(x_0) \cdot Q(x_0, y_0))$

$$d_m(x_{m-1}, x_m) = -\log(\mathbb{P}(x_{m-1}, x_m) Q(x_m, y_m))$$

$$X^{n*} = \underset{x^n}{\operatorname{argmin}} \left(d_0(x_0) + \sum_{m=1}^{n} d_m(x_{m-1}, x_m) \right) \qquad (1)$$

example: y_i are discrete, mostly fair die but occassionally loaded

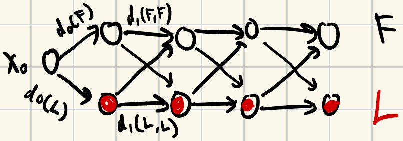
Fair = "F"
Loaded = "L"



$\pi_0 \rightarrow$ always start in fair state

Q: if F, $\mathbb{P}(i) = 1/6$
if L, $\mathbb{P}(6) = 1/2$ and $\mathbb{P}(i \neq 6) = 1/10$

We observe 66162



$y_0 = 6,\ y_1 = 6,\ y_2 = 1,\ y_3 = 6$

$d_0(F) = \pi_0(F) \cdot Q(F,6) = 1 \cdot 1/6\ ,\ -\log(1/6) = .77$

$d_0(L) = \pi_0(L)\, Q(L,6) = 0 \quad = -\log(0) = \inf$

$d_1(F,F) = -\log(.95 \cdot 1/6) = .8$

$d_1(F,L) = -\log(.05 \cdot 1/6) = 1.6$

$\vdots$

$\text{Best}(L_0) = \inf \quad , \text{Best}(L_1) \to \min\begin{pmatrix} F_0 \to L_1 \\ L_0 \to L_1 \end{pmatrix} = \left(\begin{matrix} \text{Best}(F_0) + d(F_0, L_1) = .77 + 1.6 \\ \text{Best}(L_0) + d(L_0, L_1) = \inf \end{matrix}\right)$

$\text{Best}(F_0) = .77 \quad \Rightarrow \text{Best}(L_1) = 2.37$ from F_0