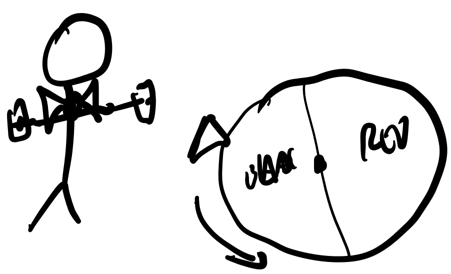
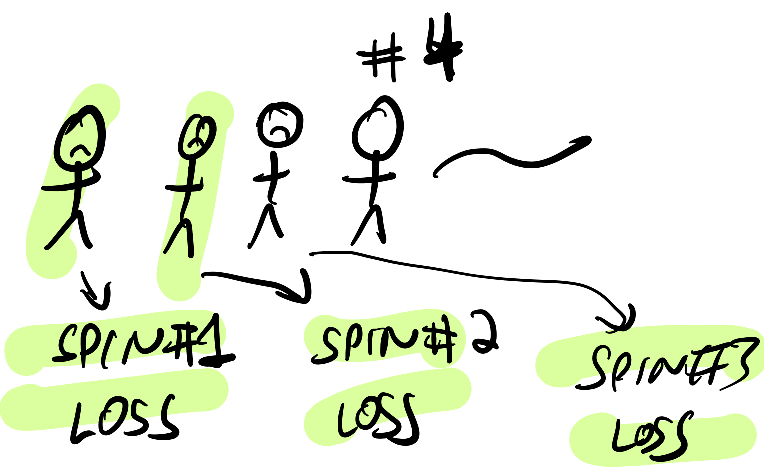
OPEN SEATING TODAY

PLAN:

- TODAY - LIKELIHOOD
- WED - PS 7 (R, JUPYTER)
- THU - OUT-OF-PLACE ZOOM LECTURE 11:00am-12:30pm "WK 10.1"
- FRI - PS 7

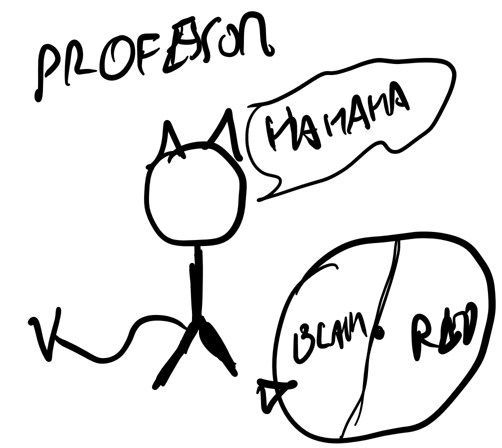
PART II



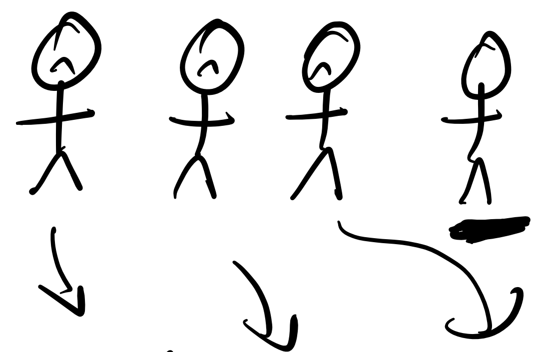


$$\mathbb{P}(\text{SPIN \#4} = \text{LOSS}) = \underline{0.5}$$

PROFESSOR



CHEATING?



SPIN #1 LOSS

SPIN #2 LOSS

SPIN #3 LOSS

$$\mathbb{P}(\text{SPIN \#4 LOSS}) = 99\%$$

PROBABILITY — AXIOMATIC, RIGOROUS

MODEL —WK 1-5→ CONSEQUENCES

MODEL ←WK 6-10— OBSERVATIONS

STATISTICS — METHODS & APPROACHES

- LEARNING
- INFERENCE
- CLASSIFICATION

PREDICTION

RECALL

PROBABILITY DENSITY EX $p_T(t) = \lambda e^{-\lambda t}$

$p_T(t;\lambda) = \lambda e^{-\lambda t}$

PROBABILITY FUNCTION EX $p_N(k) = \frac{\lambda^k e^{-\lambda}}{k!}$

$\mathbb{P}(N=k) =$

$p_N(k;\lambda) =$

THE LIKELIHOOD FUNCTION IS THE PROBABILITY DENSITY OR PROBABILITY FUNCTION, VIEWED AS A FUNCTION OF THE PARAMETER

EX $L(\lambda) = \lambda e^{-\lambda t}$

NOTE $\int L(\lambda)\, d\lambda \neq 1$

STRATEGY: TO FIND A PARAMETER θ

FROM AN OBSERVATION X, TAKE

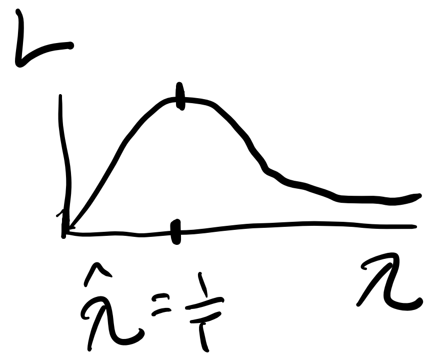
$\theta = \hat{\theta}$ THAT MAXIMIZES $L(\theta)$

$L(\theta; X)$

↳ MAXIMUM LIKELIHOOD

EX T - TIME

$$L(\lambda; T) = \lambda e^{-\lambda T}$$



$$\frac{\partial L}{\partial \lambda} = 0 \ldots$$

EX N OBSERVATIONS IDENTICAL DISTRIBUTION & INDEPENDENT (i.i.d.)

$$L(\theta) = \prod_{i=1}^{N} p_x(x_i; \theta)$$

EX N NORMAL RANDOM VARIABLES iid

$$L(\mu, \sigma; x_i) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$$

$$\frac{\partial L}{\partial \mu} = 0, \quad \frac{\partial L}{\partial \sigma} = 0 \quad \ldots$$

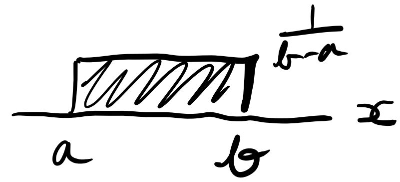
$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

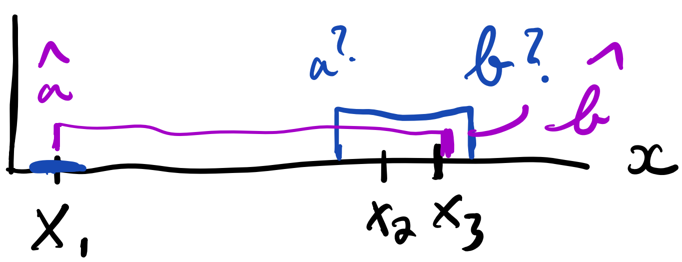
$$\hat{\sigma} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (X_i - \hat{\mu})^2}$$

EX UNIFORM RANDOM VARIABLES

3 iid OBSERVATIONS

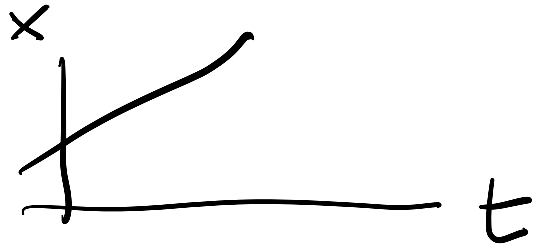
$$p_X(x; a, b) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{ELSE} \end{cases}$$



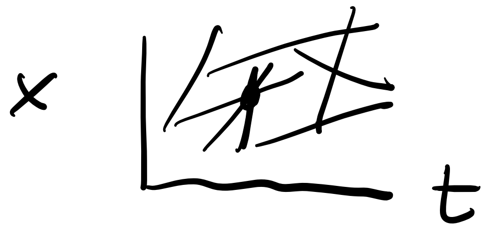


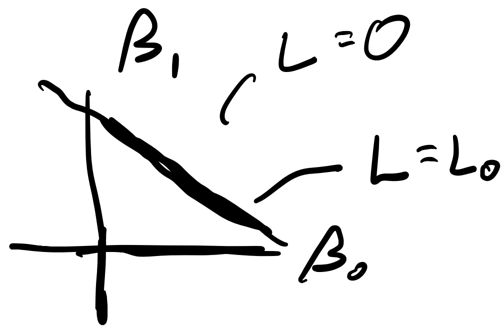
EX $\qquad \dfrac{dx}{dt} = \beta_1 \qquad x(0) = \beta_0$

$$x(t) = \beta_0 + \beta_1 t$$

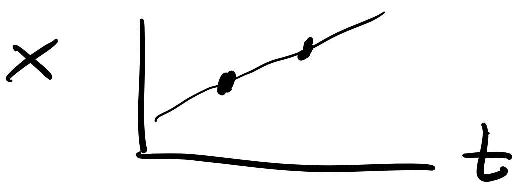


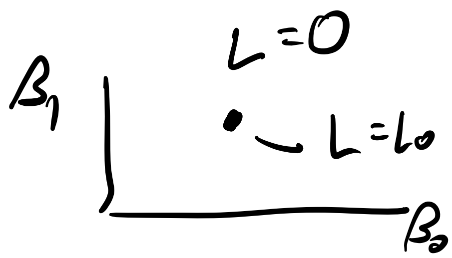
1 DATA POINT



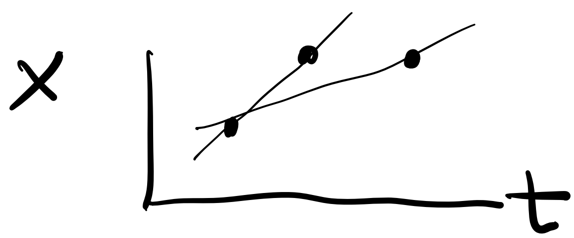


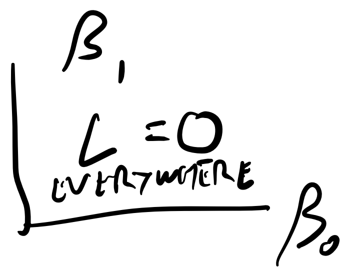
2 DATA POINTS





3 DATA POINTS



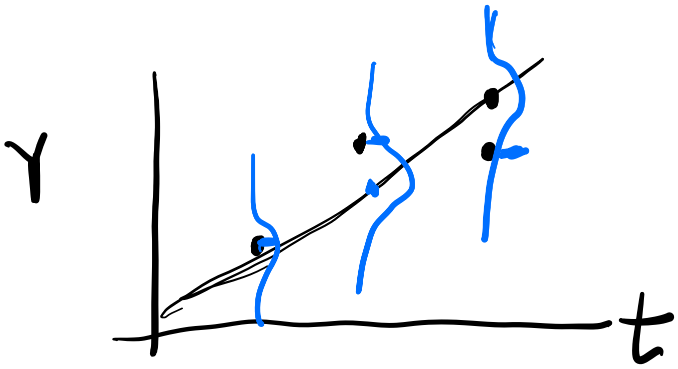


NEW MODEL

$$\frac{dx}{dt} = \beta_1 \quad , \quad x(0) = \beta_0$$

$$Y = X + E_i$$

$\uparrow$ $E_i \sim \text{NORMAL}(0, \sigma)$



$$L(\beta_0, \beta_1, \sigma) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(- \frac{(Y_i - (\beta_0 + \beta_1 t))^2}{2\sigma^2} \right)$$

$$\frac{\partial L}{\partial \beta_0} = 0 \qquad \frac{\partial L}{\partial \beta_1} = 0 \qquad \frac{\partial L}{\partial \sigma} = 0 \qquad \ldots$$

L IS MAXIMIZED WHEN

$$\sum_{i=1}^{N} \left(Y_i - (\beta_0 + \beta_1 t) \right)^2$$ IS

MINIMIZED

(SUM OF SQUARED ERRORS SSE
SUM OF SQUARED RESIDUALS SSR

MAXIMUM LIKELIHOOD FOR NORMAL iid ERRORS IS EQUIVALENT TO LEAST SQUARED ERROR

$$\frac{\partial \, SSR}{\partial \beta} = 0 \Rightarrow \underline{\beta} \text{ IS}$$

EXTREMELY FAST