Lecture 24 01 Dec

(1.) STATE/DEFINE ε_i

(2.) STATE Model

	est	...
β_1		
β_2		

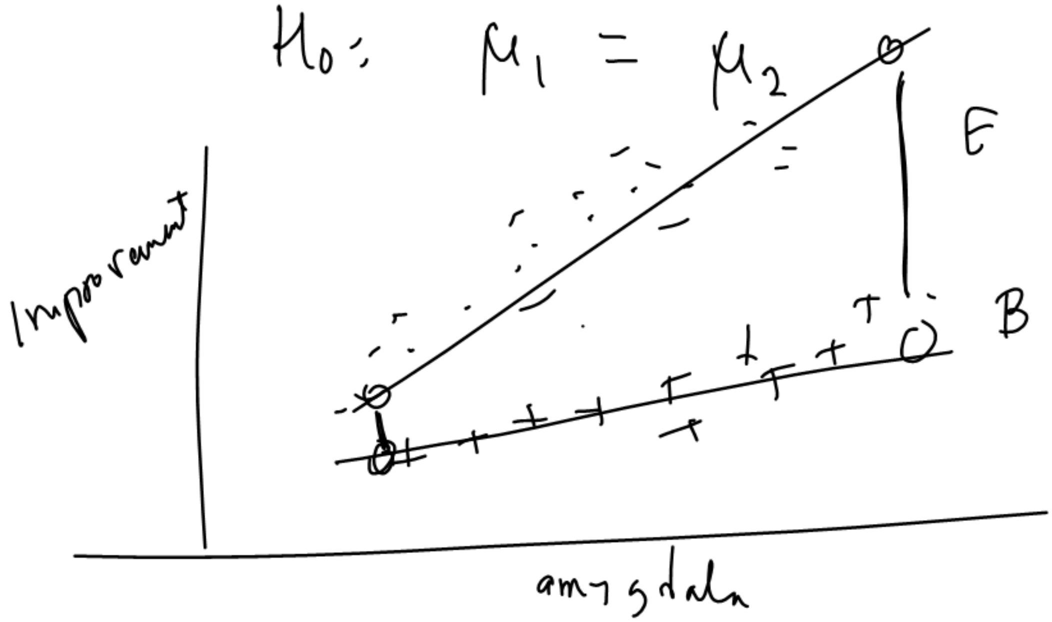
(3.) $Y^1_1, \dots, Y^1_{n_1}$ iid $N(\mu_1, \sigma^2)$

$Y^2_1 \dots Y^2_{n_2}$ iid $N(\mu_2, \sigma^2)$

$H_0: Y^1_i = Y^2_i$ X

$H_0: \mu_1 = \mu_2$

(4.)



Likelihood Function

Dataset: $\{ (x_i, Y_i),\ i = 1, \dots, n \}$

Model: $Y_i = \mu(x_i) + \varepsilon_i,$

ε_i iid $N(0, \sigma^2)$

$\mu(x_i) = \beta_0 + \beta_1 x_i$

$$\Longleftrightarrow \quad Y_i \mid x_i \overset{\text{indep}}{\sim} N\left(\mu(x_i) = \beta_0 + \beta_1 x_i,\ \sigma^2\right)$$

The likelihood function (of the parameters $\beta_0, \beta_1, \sigma^2$):

$$L(\beta_0, \beta_1, \sigma^2) = f(y_1, \dots, y_n \mid x_1 \dots x_n, \beta_0, \beta_1, \sigma^2)$$

$$= \prod_{i=1}^{n} f(Y_i \mid x_i, \beta_0, \beta_1, \sigma^2)$$

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

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{ -\frac{1}{2} \frac{\sum_{i=1}^{n} \left(Y_i - (\beta_0 + \beta_1 x_i)\right)^2}{\sigma^2} \right\}$$

The max likelihood estimator of $\beta_0, \beta_1, \sigma^2$:

$$(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2) = \arg\max L(\beta_0, \beta_1, \sigma^2)$$

$$= \arg\max \log L(\beta_0, \beta_1, \sigma^2)$$

$$= \arg\max \; -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2}\frac{\sum_{i=1}^{n}\left(Y_i - (\beta_0 + \beta_1 x_i)\right)^2}{\sigma^2}$$

Consider σ^2 fixed.

Maximizing the log likelihood wrt β_0, β_1 is identical to minimizing

$$C(\beta_0,\beta_1) = \sum_{i=1}^{n}\left(Y_i - (\beta_0 + \beta_1 x_i)\right)^2$$

Generalizing ...

$$Y_i \mid x_i \begin{cases} 1 & \text{YES} \quad \text{crack} \\ 0 & \text{NO} \quad \text{o/w} \end{cases}$$

Space Challenger Dataset

$$Y_i \mid x_i \sim \text{Bernoulli}\left(\pi(x_i) \right.$$

$$= \Pr(Y_i = 1 \mid x_i)$$

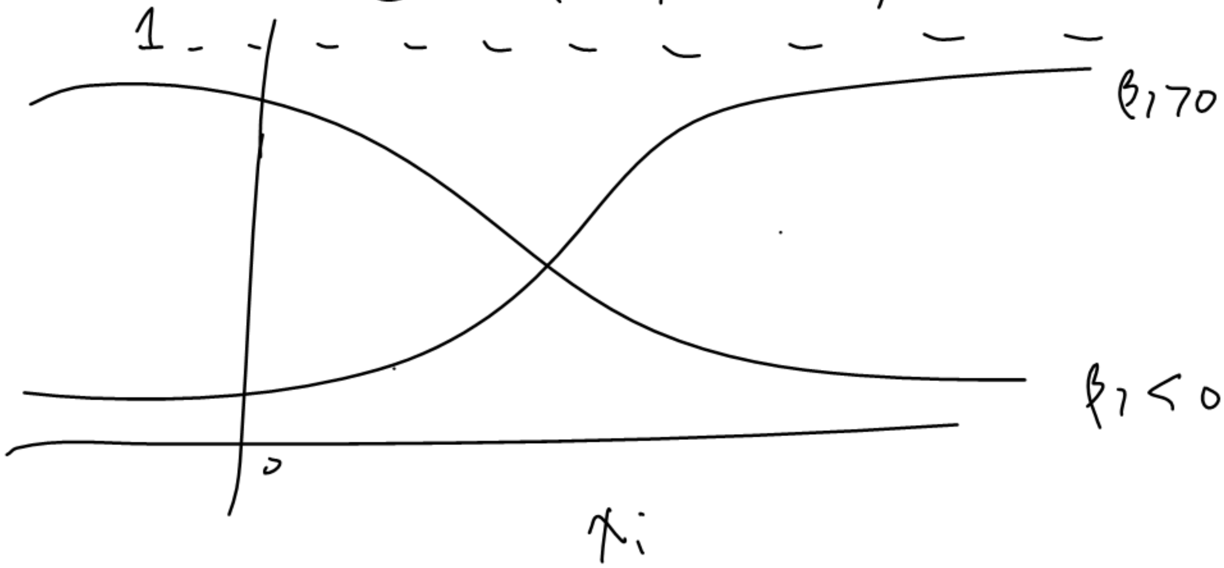
$$\underbrace{\pi(x_i)} = E(Y_i \mid x_i)$$

$$\pi(x_i) = \Pr(Y_i = 1 \mid x_i)$$

$$\underset{=}{\text{e.g.}} \quad \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}$$

Logistic curve

$\in (0, 1)$



The likelihood function of β_0, β_1:

$$L(\beta_0, \beta_1) = \overset{\text{pmf}}{f}(Y_1, \dots Y_n \mid x_1 \dots x_n, \beta_0, \beta_1)$$

$$= \prod_{i=1}^{n} \underbrace{f(Y_i \mid x_i, \beta_0, \beta_1)}$$

$$= \prod_{i=1}^{n} \pi(x_i)^{Y_i} [1 - \pi(x_i)]^{1-Y_i}$$

$$= \prod_{i=1}^{n} \left(\frac{\pi(x_i)}{1 - \pi(x_i)} \right)^{Y_i} (1 - \pi(x_i))$$

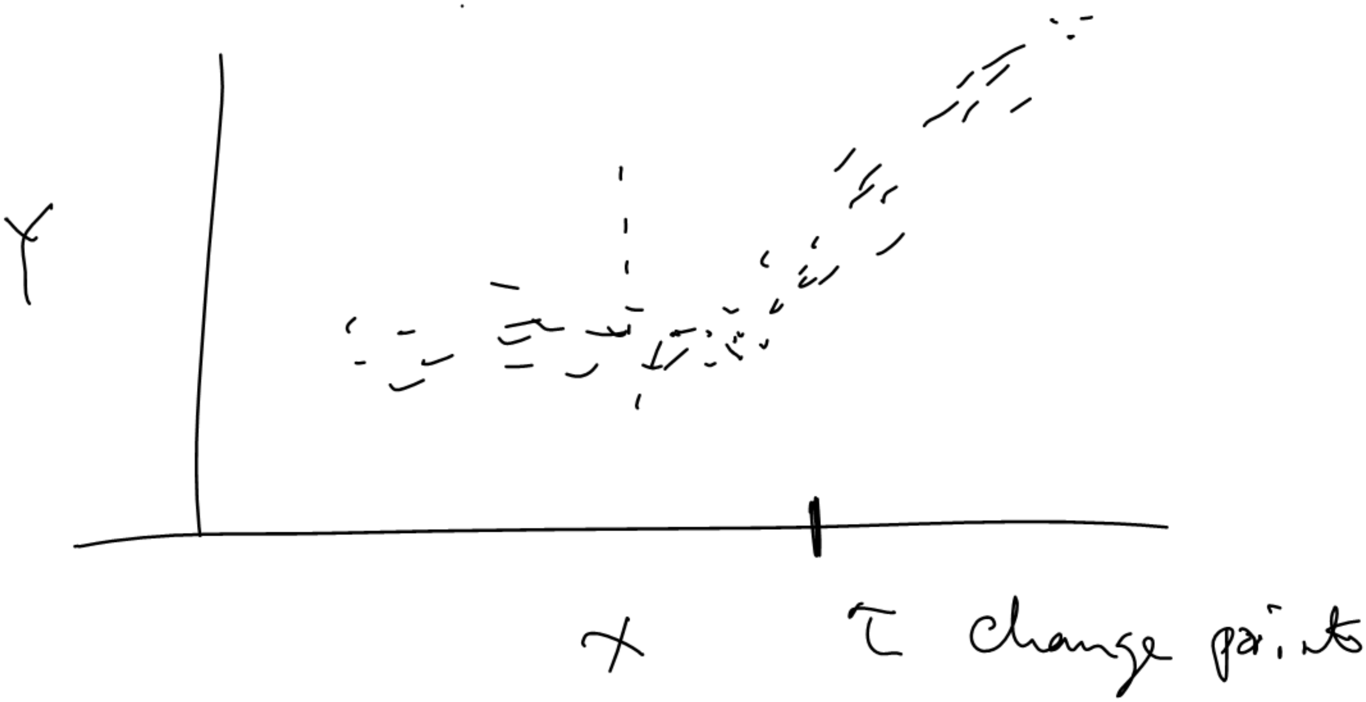
$$= \prod_{i=1}^{n} \left(\exp(\beta_0 + \beta_1 x_i)\right)^{Y_i} \left(\frac{1}{1+\exp(\beta_0+\beta_1 x_i)}\right)$$

$$\frac{\pi(x_i)}{1-\pi(x_i)} = \frac{\dfrac{\exp(\beta_0+\beta_1 x_i)}{\cancel{1+\exp(\beta_0+\beta_1 x_i)}}}{\dfrac{1}{\cancel{1+\exp(\beta_0+\beta_1 x_i)}}}$$

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg\max L(\beta_0, \beta_1).$$

iteratively!

More on likelihood



DATA: $\{(x_i, Y_i),\ i = 1, \dots, n\}$

Model: $Y_i = \mu(x_i) + \varepsilon_i,$

$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

$$\mu(x_i) = \begin{cases} \beta_0^1 + \beta_1^1 x_i, & x_i \le \tau \\ \beta_0^2 + \beta_1^2 x_i, & x_i > \tau \end{cases}$$

Unknown parameters are:

$\tau, \underbrace{\beta_0^1, \beta_1^1}_{\text{PRE}}, \underbrace{\beta_0^2, \beta_1^2}_{\text{POST}}, \sigma^2$

$$L(\tau, \beta_0^1, \beta_1^1, \beta_0^2, \beta_1^2, \sigma^2)$$

$$= \prod_{i=1}^{n} f(y_i \mid x_i, \ldots)$$

WLOG: $x_1 < x_2 < \ldots < x_{n_1} \leq \tau <$

$x_{n_1+1} < \ldots < x_{n_1+n_2}$

$$= \prod_{i=1}^{n_1} f(y_i \mid \ldots) \cdot \prod_{i=n_1+1}^{n_1+n_2} f(y_i \mid \ldots)$$

$$= \prod_{i=1}^{n_1} \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) \exp\left\{-\frac{1}{2}\frac{1}{\sigma^2}\left(Y_i - (\beta_0^1 + \beta_1^1 x_i)\right)^2\right\}$$

$$\cdot \prod_{i=n_1+1}^{n_1+n_2} \cdots$$

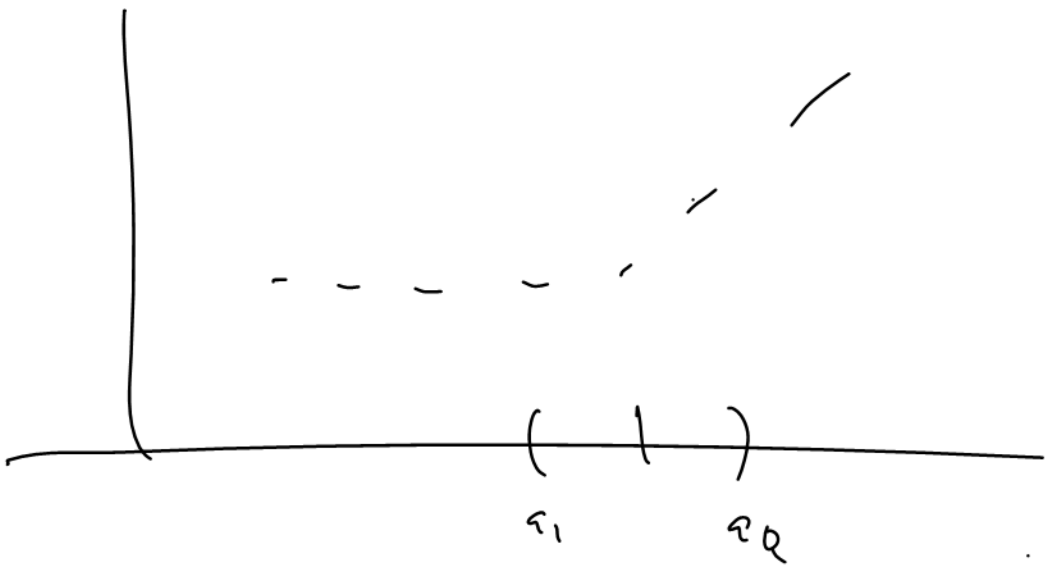
The log likelihood function of $\tau, \beta_0^1, \beta_1^1, \beta_0^2, \beta_1^2, \sigma^2$:

$$l(\quad) = -\frac{n_1+n_2}{2}\ln(2\pi\sigma^2) +$$

$$-\frac{1}{2}\frac{1}{\sigma^2}\sum_{i=1}^{n_1}\left(Y_i - (\beta_0^1 + \beta_1^1 x_i)\right)^2$$

$$-\frac{1}{2}\frac{1}{\sigma^2}\sum_{i=n_1+1}^{n_1+n_2}\left(Y_i - (\beta_0^2 + \beta_1^2 x_i)\right)^2$$

Suppose that we have prior information that $\tau \in \{a_1, \dots a_Q\}$



Candidate change point : a_1

$$\underline{x_1 < \dots < x_{m_1}} \le a_1 < \underline{x_{m_1+1} \dots < x_n}$$

$$\ell(\tau = a_1, \beta_0^1, \beta_1^1, \beta_0^2, \beta_1^2, \sigma^2)$$

$$= -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2}\frac{1}{\sigma^2}\sum_{i=1}^{m_1}\left(Y_i - (\beta_0^1 + \beta_1^1 x_i)\right)^2$$

$$- \frac{1}{2} \frac{1}{\sigma^2} \sum_{i=m_1+1}^{n} \left(Y_i - \left(\beta_0^2 + \beta_1^2 x_i\right)\right)^2$$

$$\hat{\beta}_0^1(a_1), \hat{\beta}_1^1(a_1), \hat{\beta}_0^2(a_1), \hat{\beta}_1^2(a_1), \hat{\sigma}^2(a_1)$$

$$l\left(a_1, \hat{\beta}_0^1(a_1) \ldots, \hat{\beta}_1^2(a_1), \hat{\sigma}^2(a_1)\right)$$

Continue with other candidate change-points $a_2, \ldots, a_Q$

Finally:

$$\hat{\tau} = \underset{a_j \in \{a_1 \ldots a_Q\}}{\arg\max}\, l(a_j, \ldots)$$

Discussion

$$\underline{Y} = X\beta + \underline{\varepsilon}$$

$$\underline{\varepsilon} \sim N(\underline{0}, I \otimes \sigma^2)$$

$$\hat{\beta} = (X'X)^{-1}X'\underline{Y}$$

$$\hat{\underline{Y}} = X\hat{\beta} = \underbrace{X(X'X)^{-1}X'}_{P}\underline{Y}$$

$$P \quad \underline{Y}$$

$$SSE(1) = (\underline{Y} - X\hat{\beta})'(\underline{Y} - X\hat{\beta})$$

$$\underline{R} = \underline{Y} - X\hat{\beta} = \underline{Y} - P\underline{Y}$$

$$= \underbrace{(I - P)}\underline{Y}$$

Recall: $Y \sim N(X\beta, I \otimes \sigma^2)$

$$\underline{R} = (I - P)\underline{Y}$$

$$\sim N\Big(E\underline{R} = \underbrace{E(I-P)\underline{Y}}, $$

$$\operatorname{cov}(\underline{R}) = (I-P)\operatorname{cov}(\underline{Y})(I-P)' \Big)$$

$$E\underline{R} = (I-P)E\underline{Y} = (I-P)X\underline{\beta}$$

$$= X\underline{\beta} - \underbrace{PX}_{X}\beta$$

$$= \underline{0}$$

$$\operatorname{cov}(\underline{R}) = (I-P)(I \otimes \sigma^2)(I-P)'$$

$$= (I-P)(I-P)' \otimes \sigma^2$$

$$= (I-P) \otimes \sigma^2$$

$$\underline{R} \sim N(\underline{0}, (I - P) \otimes \sigma^2)$$

$$\underline{R} \otimes \frac{1}{\sigma} \sim N(\underline{0}, (I - P))$$

<u>Lemma</u>. $\underline{U} \sim N(\underline{0}, W)$

where W is symmetric & idempotent

$$\Rightarrow \quad \underline{U}'\underline{U} \sim \chi^2(df = \text{trace}(W))$$

$$\underline{R} \otimes \frac{1}{\sigma} \sim N\left(\underline{0}, \ (I-P)\right)$$

$(I-P)$ is symmetric & idempotent

$(I-P)' = (I-P)$ & $(I-P)^2 = I-P$

Lemma $\Rightarrow$

$$\underline{\underline{\left(\underline{R} \otimes \frac{1}{\sigma}\right)'\left(\underline{R} \otimes \frac{1}{\sigma}\right)}} \sim \chi^2$$

with $df = \text{trace}\,(I-P)$

$$\frac{\underline{R}'\underline{R}}{\sigma^2} \sim \chi^2\Big(df = n - \underbrace{\text{trace}\,(P)}_{n-d}\Big)$$

$$\text{trace}\,(P) = \text{trace}\left(X(X'X)^{-1}X'\right)$$

$$= \text{trace}\left((X'X)^{-1}\underline{X'X}\right)$$

$$= \text{trace}\left(I_d\right)$$

d = # columns of X