EM Algorithm and ZIP models

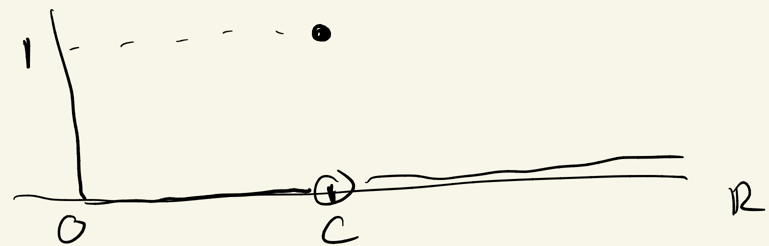
Point mass: Y is a point mass at $c \in \mathbb{R}$

if $P(Y = c) = 1$

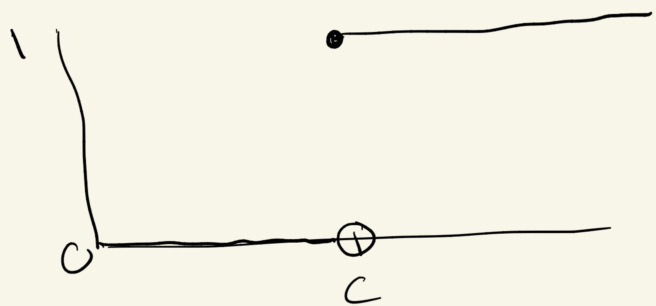
$\Rightarrow P(Y = y) = 0$ for $y \neq c$

Probability function

$$f(y) = \begin{cases} 1 & y = c \\ 0 & y \neq c \end{cases}$$



CDF:



EM algorithm in general

Let θ be an unknown parameter we want to estimate. Let Y be a set of observed data, and Z a set of unobserved latent/missing data.

$$L(\theta) = P(Y \mid \theta) = \int p(Y \mid Z = z, \theta)\, P(Z = z \mid \theta)\, dz$$

Maximizing this likelihood is challenging when we don't observe Z.

<u>EM algorithm</u>:

<u>E step</u>: Let $\theta^{(k)}$ be the current estimate of θ

$$Q(\theta \mid \theta^{(k)}) = \mathbb{E}_{Z \mid Y, \theta^{(k)}}\left[\log L(\theta; Z, Y)\right]$$

<u>M step</u>:

$$\theta^{(k+1)} = \underset{\theta}{\operatorname{argmax}}\; Q(\theta \mid \theta^{(k)})$$

Motivation:

$$\log P(Y|\theta) = \log\left(\int_z P(Y, Z=z|\theta)\,\partial z\right)$$

$$= \log\left(\int_z \frac{P(Y,Z=z|\theta)}{P(Z=z|Y,\theta_{old})} P(Z=z|Y,\theta_{old})\,\partial z\right)$$

$$= \log \mathbb{E}_{Z|Y,\theta_{old}}\left[\frac{P(Y,Z|\theta)}{P(Z|Y,\theta_{old})}\right]$$

$$\geq \mathbb{E}_{Z|Y,\theta_{old}}\left[\log\left(\frac{P(Y,Z|\theta)}{P(Z|Y,\theta_{old})}\right)\right]$$

$$= \underbrace{\mathbb{E}_{Z|Y,\theta_{old}}\left[\log P(Y,Z|\theta)\right]} - \mathbb{E}_{Z|Y,\theta_{old}}\left[\log P(Z|Y,\theta_{old})\right]$$

$$= \mathbb{E}_{Z|Y,\theta_{old}}\left[\log L(\theta;Y,Z)\right] + \overbrace{H(\theta_{old})}^{\geq 0}$$

$$= \underbrace{Q(\theta|\theta_{old})}_{\text{lower bound on log likelihood}} + H(\theta_{old})$$

Increasing $Q(\theta|\theta_{old})$ should increase $\log L(\theta)$

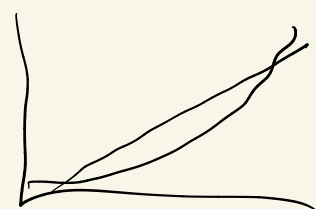
Jensen's inequality:

If φ is a convex function,

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]$$

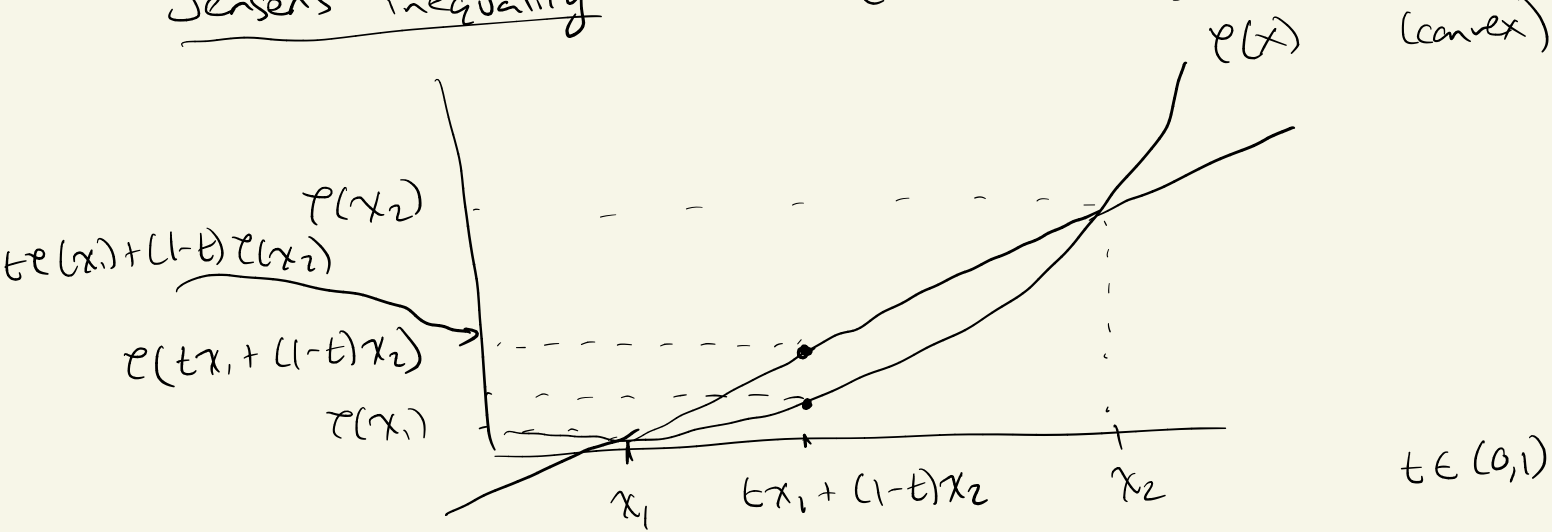
$\Rightarrow$ If φ is concave, then

$$\varphi(\mathbb{E}[X]) \geq \mathbb{E}[\varphi(X)]$$



Jensen's inequality

(visual intuition)



$\varphi(tx_1 + (1-t)x_2)$ representing $\varphi(\mathbb{E}[X])$

$\leq t\varphi(x_1) + (1-t)\varphi(x_2)$ (represents $\mathbb{E}[\varphi(X)]$)

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]$$

Class activity

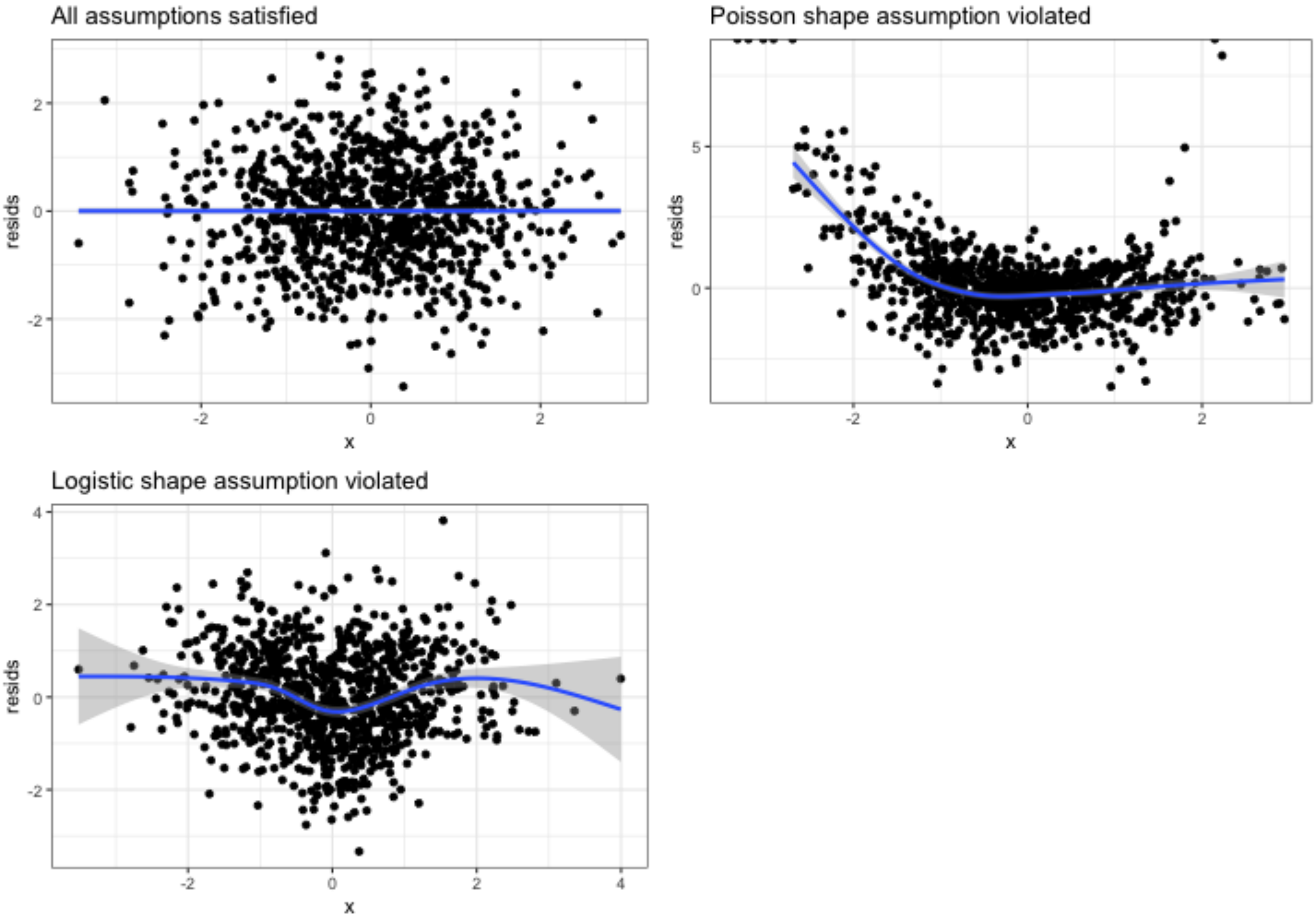
https://sta712-f22.github.io/class_activities/ca_lecture_31.html

$$\log\left(\frac{\alpha_i}{1-\alpha_i}\right) = \gamma_0 + \gamma_1 X_i \qquad \gamma_0 = 0 \quad \gamma_1 = 1$$

$$\log(\lambda_i) = \beta_0 + \beta_1 X_i \qquad \beta_0 = 1 \quad \beta_1 = 1$$

$$\alpha_i = \frac{\exp\{\gamma_0 + \gamma_1 x_i\}}{1 + \exp\{\gamma_0 + \gamma_1 x_i\}}$$

Assessing the shape assumption



Logistic component vs. Poisson component