Lecture 20 Math 340 5/4/16

$$X_1, \ldots, X_n \overset{iid}{\sim} \rho N(\theta_0, \sigma_0^2) + (1-\rho) N(\theta_1, \sigma_1^2) = \sum_{m=1}^{M} \rho_m N(\theta_m, \sigma_m^2)$$

Goal: estimate all $\rho, \theta_0, \sigma_0^2, \theta_1, \sigma_1^2$

MLE approach: no closed form solution

grid search: will be too approx

Multivariate N-R: too difficult

$$P(X \mid \theta_0, \sigma_0^2, \theta_1, \sigma_1^2, \rho) = \prod_{i=1}^{n} \rho \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2\sigma_0^2}(x_i - \theta_0)^2} + (1-\rho) \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2\sigma_1^2}(x_i - \theta_1)^2}$$

What if we knew ...

that X_1 belonged to $N(\theta_1, \sigma_1^2)$, X_2 belonged to $N(\theta_0, \sigma_0^2)$, etc...

this is called a "latent variable"

let $I_i := \mathbb{1}_{\text{the } i^{th} \text{ observation belongs to } N(\theta_0, \sigma_0^2)}$ $\rho^{I_i}(1-\rho)^{1-I_i}$

$I_1, \ldots, I_n \overset{iid}{\sim} \text{Bern}(\rho)$ (no reason to think otherwise) $\Rightarrow I \mid \rho \sim \text{Bin}(n, \rho)$

$$P(X, I \mid \theta_0, \sigma_0^2, \theta_1, \sigma_1^2, \rho) = P(X \mid \theta_0, \sigma_0^2, \theta_1, \sigma_1^2, I, \rho)\, P(I \mid \theta_0, \sigma_0^2, \theta_1, \sigma_1^2, \rho)$$

$$= P(X \mid \theta_0, \sigma_0^2, \theta_1, \sigma_1^2, I)\, P(I \mid \rho)$$

once you know I... no need for ρ

$$= \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2\sigma_0^2}(x_i - \theta_0)^2}\right)^{I_i} \left(\frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2\sigma_1^2}(x_i - \theta_1)^2}\right)^{1-I_i} \rho^{I_i}(1-\rho)^{1-I_i}$$

Is this mathematically valid?

put these inside

$$= \int p(X|\theta, I)\, p(I|\theta)\, dI$$

$$p(X|\theta) = \int_{supp[I]} p(X, I|\theta)\, dI$$

... but does this work?

$$p(X|\theta_0 \ldots \rho) = \int_{supp(I_1)} \cdots \int_{supp(I_n)} \prod_{i=1}^{n} p(X_i I_i|\theta)\, dI_1 \ldots dI_n$$

I_i is discrete so...

$$= \prod_{i=1}^{n} \sum_{I_1 \in \{0,1\}} \cdots \sum_{I_n \in \{0,1\}} \left(\rho \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2\sigma_0^2}(X_i-\theta_0)^2}\right)^{I_i} \left((1-\rho) \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2\sigma_1^2}(X_i-\theta_1)^2}\right)^{1-I_i}$$

each term gets its own sum, and then all are multiplied together

$$= \sum_{I_1 \in \{0,1\}} \left(\rho \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2\sigma_0^2}(X_1-\theta_0)^2}\right)^{I_1} \left((1-\rho) \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2\sigma_1^2}(X_1-\theta_1)^2}\right)^{1-I_i} \cdot \ldots \sum_{I_n \in \{0,1\}} \ldots$$

$$= \left(\rho \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2\sigma_0^2}(X_1-\theta_0)^2} + (1-\rho) \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2\sigma_1^2}(X_1-\theta_1)^2}\right) \cdots (-\; n \;-)$$

$$= \prod_{i=1}^{n} \rho \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2\sigma_0^2}(X_i-\theta_0)^2} + (1-\rho) \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2\sigma_1^2}(X_i-\theta_1)^2}$$ = our original likelihood function

so... $p(X|\theta) = \int p(X|\theta, I)\, p(I|\theta)\, dI$

is called "data augmentation"

↑ pseudo-data — "add new" ↑

we "ignore" the [illegible]. But who cares! Here's why...

$$\nabla \ln \mathcal{L}(\theta_0, \sigma_0^2, \sigma_1^2, \rho \mid X) = 0$$ not possible... but if we add in I...

$$\Rightarrow \nabla \ln \prod_{i=1}^{n} \cdots\cdots$$

$$\Rightarrow \nabla \ln \prod_{i=1}^{n} \rho^{I_i} \left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\right)^{I_i} e^{-\frac{1}{2\sigma_0^2}(X_i-\theta_0)^2 I_i} (1-\rho)^{1-I_i} \left(\frac{1}{\sqrt{2\pi\sigma_1^2}}\right)^{1-I_i} e^{-\frac{1}{2\sigma_1^2}(X_i-\theta_1)^2(1-I_i)}$$

$$\Rightarrow \nabla \ln \rho^{\Sigma I_i} \left(\checkmark\right)^{\Sigma I_i} e^{-\Sigma \cdots\cdots} (1-\rho)^{\Sigma 1-I_i} \left(\checkmark\right)^{\Sigma 1-I_i} e^{-\Sigma \cdots\cdots}$$

$(\ln(2\pi) + \ln(\sigma_0^2))$

$$\Rightarrow \nabla \quad \Sigma I_i \ln \rho - \frac{1}{2}\Sigma I_i \ln(2\pi\sigma_0^2) - \frac{1}{2\sigma_0^2}\Sigma_i (X_i-\theta_0)^2 I_i$$

$$+ \left(\Sigma 1-I_i\right)\ln(1-\rho) - \frac{1}{2}\left(\Sigma 1-I_i\right)\ln\left(2\pi\sigma_1^2\right) - \frac{1}{2\sigma_1^2}\Sigma_i (X_i-\theta_1)^2(1-I_i)$$

let $n_1 = \Sigma I_i$, $n_0 = \Sigma 1-I_i$

$$\frac{\partial}{\partial\rho}[\checkmark] = \frac{n_1}{\rho} - \frac{n_0}{1-\rho} = 0 \Rightarrow \frac{n_1}{\rho} = \frac{n_0}{1-\rho} \Rightarrow \frac{1}{\rho} - 1 = \frac{n_0}{n_1}$$

$$\Rightarrow \frac{1}{\rho} = \frac{n}{n_1} \Rightarrow \hat{\rho}_{MLE} = \frac{n_1}{n} = \frac{1}{n}\Sigma I_i$$ obviously... just $\bar{X}$

$$\frac{\partial}{\partial\theta_0}[\,] = -\frac{1}{2\sigma_0^2}\frac{\partial}{\partial\theta_0}\Sigma(X_i-\theta_0)^2 I_i = 0 \Rightarrow \frac{\partial}{\partial\theta_0}\left[\Sigma X_i^2 - 2\theta_0\Sigma X_i + \Sigma(\theta_0^2) I_i\right] = 0$$

$$\Rightarrow -\cancel{2}\Sigma X_i I_i + \cancel{2}\theta_0 n_1 = 0 \Rightarrow \hat{\theta}_{0,MLE} = \frac{\Sigma X_i I_i}{n_1} \quad \& \quad \hat{\theta}_1 = \frac{\Sigma X_i(1-I_i)}{n_0}$$

obviously just $\bar{X}$

$$\frac{\partial}{\partial\sigma_0^2}[\,] = -\frac{n_1}{\cancel{2}\sigma_0^2} + \frac{1}{\cancel{2}(\sigma_0^2)^2}\Sigma(X_i-\hat{\theta}_0)^2 I_i = 0$$

$$\Rightarrow \hat{\sigma}^2_{0,MLE} = \frac{1}{n_1}\Sigma(X_i-\hat{\theta}_0)^2 I_i \quad \& \quad \hat{\sigma}^2_{1,MLE} = \frac{1}{n_0}\Sigma(X_i-\hat{\theta}_1)^2(1-I_i)$$

But this doesn't help us since we don't know $I_1, \dots, I_n$!

What if we can estimate them? How would I estimate I_1?

Math 281 exercise

$X = \mathbb{1}_{\omega \in A}$

$E(X) = ? = P(\omega \in A)$

$X = \mathbb{1}_{\text{Trump will be president}}$

$E(X) = P(\checkmark)$

$I_1 = \mathbb{1}\ X_1$ belongs to $N(\theta_0, \sigma_0^2)$

let $\hat{I}_1 = \underset{P(I_1 = 1 \mid \ldots)}{E[I_1 \mid \theta_0, \sigma_0^2, \theta_1, \sigma_1^2, \rho]} = \dfrac{\rho P(X_1 \mid \theta_0, \sigma_0^2)}{\rho P(X_1 \mid \theta_0, \sigma_0^2) + (1-\rho) P(X \mid \theta_1, \sigma_1^2)}$

$$= \frac{\rho \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2\sigma_0^2}(X_1 - \theta_0)^2}}{\rho \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2\sigma_0^2}(X_1 - \theta_0)^2} + (1-\rho) \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2\sigma_1^2}(X_1 - \theta_1)^2}}$$

let $\hat{I}_2 = P(I_2 = 1 \mid \ldots) = \ldots$

$\vdots$

$\hat{I}_n = P(I_n = 1 \mid \ldots) = \ldots$

We can do a quick calculation for each $I_1, \dots, I_n$

But this doesn't help us since we don't know $\theta_0, \sigma_0^2, \theta_1, \sigma_1^2$!

What if we guess $\theta_{0,0} = 10$, $\sigma^2_{0,0} = 1$, $\theta_{1,0} = 15$, $\sigma^2_{1,0} = 1$, $\rho = 30\%$

or reasonable strong guesses looking at the data

then use these guesses to estimate the augmented data $\hat{I}_1, \dots, \hat{I}_n$

then calculate the MLE's. Repeat. Stop at a certain tolerance(s).

Step 1: guess values for the parameters.

Step 2: Impute the augmented data using expectation (the E-step)

Step 3: Compute the MLE's for the params using the imputed augmented data (the M-step)

Step 4: Repeat 2 & 3 until $|\theta_{t+1} - \theta_t| < \varepsilon$ (prespecified).

It's NEW!

1977 Dempster, Laird, Rubin

But 1974 Rolf Sundberg found it first

1977 - generalized residual

Wu, 1983: proved convergence for a wide variety of parametric models

---

Newton Raphson: useful for solving $f(x) = c$. We used it for finding $\hat{\theta}_{MAP} = \arg\max\{k(\theta|x)\}$ by setting $k'(\theta|x) = 0$ when it was not solvable in closed form.

E-M: useful for cases where MLE's break down due to not in closed form but if you knew some latent data it would be easy.

Can this help with our semi-conjugate model?

$$X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, \sigma^2), \quad \theta \sim N(\mu_0, \tau^2), \quad \sigma^2 \sim \text{InvGamma}\left(\frac{n_0}{2}, \frac{n_0\sigma_0^2}{2}\right)$$

$$P(\theta, \sigma^2 | X) \propto P(X|\theta, \sigma^2) P(\theta) P(\sigma^2)$$

$$\propto \left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i - \theta)^2}\right) \left(e^{-\frac{1}{2\tau^2}(\theta - \mu_0)^2}\right) \left((\sigma^2)^{-\frac{n_0}{2} - 1} e^{-\frac{n_0\sigma_0^2}{2\sigma^2}}\right)$$

$$\propto (\sigma^2)^{-\frac{n_0 + n}{2} - 1} e^{-\frac{\sum x_i^2 + n_0\sigma_0^2}{2\sigma^2}} \underbrace{e^{\frac{\theta n \bar{x}}{\sigma^2}} e^{-\frac{n\theta^2}{2\sigma^2}} e^{-\frac{\theta^2}{2\tau^2}} e^{\frac{\theta\mu_0}{\tau^2}}}$$