EM Algorithm

EM stands for Expectation-~~Maximation~~ Maximization.

Originally by DLR, 1977 but ideas go much farther back. DLR united many different ideas and put them in a statistical framework.

meta-algorithm?

EM is not strictly an "algorithm". It is an (abstract) algorithm for creating other algorithms

The basic principle of EM is straightforward.

We observe some data Y ~~[illegible]~~ but there are some data that are unavailable or "missing". Call these data Z.

The observed data Y with the missing data Z are the complete data $X = (Y, Z)$

(1) The complete data have a joint density $g(\underbrace{\cancel{X} z}_{\text{complete data}} | \theta)$

(2) Because of missing Z, we cannot evaluate g. We observe Y with a joint density

$$f(y|\theta) = \int g(y, z|\theta)\, dz$$

$$\ell(\theta|y) = \log f(y|\theta)$$

(3) $\ell(\theta|y)$ is hard to evaluate! (because of $\int$)

(although, maybe not!)

Direct ML maximizes $\ell(\theta | y)$. This may be possible!

Heuristically, the EM algorithm is as follows:

$E \log X$

$E f(Y) \le f(E X)$

① E-step. Given estimate $\theta = \theta_0$

Define $Q(\theta | \theta_0) = \mathbb{E}\left[\log g(y, z | \theta) \,\middle|\, y, \theta_0\right]$

$$= \int p(z | y, \theta_0) \log g(y, z | \theta)\, dz$$

switch $\int$ and log

$\Rightarrow$ add $p(z | y, \theta_0)$

② Maximize $Q(\theta | \theta_0)$ wrt θ, ~~and set~~

~~set $\hat{\theta}_1 = \arg\max_\theta Q(\theta | \theta_0)$~~

~~Set $\theta_0 = \hat{\theta}_1$ and goto ①~~

~~Create a Sequence~~

Set $\theta_{n+1} = \arg\max_\theta Q(\theta | \theta_n)$

Under broad assumptions, $\theta_n \xrightarrow[n \to \infty]{} \hat{\theta}$, MLE

(More later)

For a regular exponential family,

Complete data pdf $g(x|\theta) = h(x)\exp(\theta^T t(x))/a(\theta)$

$$\log g(x|\theta) = \cancel{\log h(x)} + \theta^T t(x) - \log a(\theta)$$

$$Q(\theta|\tilde{\theta})\ \mathbb{E}[\log g(X|\theta)] = \theta^T\,\mathbb{E}[t(x)|\tilde{\theta},y] - \log a(\theta)$$

$$Q'(\theta|\tilde{\theta}) = \mathbb{E}[t(x)|\tilde{\theta},y] - \mathbb{E}_\theta[t(x)] = 0$$

$$\Rightarrow \mathbb{E}_\theta[t(x)] = \mathbb{E}[t(x)|\tilde{\theta},y]$$

Ex: $y_1, y_2, \to y_m, z_{m+1}, \to z_n \sim N(\mu,\sigma^2)$.

$$g(x|\mu,\sigma^2) \propto \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2\right)$$

$$\log g(x|\mu,\sigma^2) = -\frac{n}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum(x_i^2 - 2x_i\mu+\mu^2)$$
$$= -\frac{n}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\left[\sum x_i^2 - 2\mu\sum x_i + \mu^2\right]$$

$$\mathbb{E}\left(\begin{matrix}\sum x_i^2\\ \sum x_i\end{matrix}\middle|\, y\right) = \begin{pmatrix}\sum_{i=1}^{m} y_i^2 + [\tilde{\mu}^2+\tilde{\sigma}^2](n-m+1)\\ \sum_{i=1}^{m} y_i + \tilde{\mu}(n-m+1)\end{pmatrix}$$

$$\mathbb{E}\begin{pmatrix}\sum x_i^2\\ \sum x_i\end{pmatrix} = \begin{pmatrix}(\mu^2+\sigma^2)n\\ \mu n\end{pmatrix}$$

TOPIC:

DATE:

FILE UNDER:

PAGE:

Ex: $Y_1 \to Y_m, Z_{m+1} \to Z_n \sim \text{Poisson}(\mu)$

$$g(x|\mu) \propto \prod_{i=1}^{n} \mu^{x_i} e^{-\mu}$$

$$\log g(x|\mu) = \sum_{i=1}^{n} x_i \log \mu - \mu$$

$$= \log \mu \sum x_i - n\mu$$

$$\mathbb{E}\left[\sum_{i=1}^{n} x_i \middle| y\right] = \sum_{i=1}^{m} y_i + \sum_{i=m+1}^{n} \tilde{\mu}$$

$$\mathbb{E}_\mu\left[\sum_{i=1}^{n} x_i\right] = \mu n$$

TOPIC

FILE UNDER

DATE

PAGE

Ex: Censored exponential data

$Y_1, \dots, Y_n \sim \text{Exp}(\lambda)$ but some cases are censored on right

Let ~~complete~~ observed data be

$(\min(Y_1, C_1), \delta_1), \dots, (\min(Y_n, C_n), \delta_n)$

where $\delta_i = 1$ if $Y_i \le C_i$ and

$= 0$ if Y_i censored

$$g(x \mid \lambda) \propto \prod_{i=1}^{n} \frac{1}{\lambda} \exp(-x_i/\lambda)$$

$$\log g(x \mid \lambda) = -n \log \lambda - \frac{1}{\lambda} \sum x_i$$

$$= -n \log \lambda - \frac{1}{\lambda}\left\{\sum_{obs} Y_i + \sum_{censored} Z_i\right\}$$

$$\mathbb{E}\left[\log g(x \mid \lambda) \,\middle|\, Y, \tilde{\lambda}\right]$$

$$= -n \log \lambda - \frac{1}{\lambda}\left[\sum_{obs} Y_i + \sum_{mis} c_i + \cancel{1/\tilde{\lambda}}\right]$$

$\mathbb{E}[Z_i \mid Y_i, c_i]$

$$\Rightarrow \hat{\lambda} = \frac{1}{n}\left[\sum_{obs} Y_i + \sum_{mis} c_i + 1/\tilde{\lambda}\right]$$

Roughly,

If ℓ is ~~convex~~ concave then $\{\ell(\theta_n)\}$ is a monotone increasing, bounded, seq. of numbers.

$\Rightarrow$ there is a limit.

~~But does $\theta_n \to \theta$~~

But does this mean $\theta_n \to \hat{\theta}$, MLE?

Not necessarily, but yes for exponential families.

For exp fam, there is always a unique maximizer.

<u>Note</u>: We do not <u>require</u> that

$$\theta_{n+1} = \arg\max_{\theta} Q(\theta \mid \theta_n).$$

We <u>only</u> need $Q(\theta_{n+1} \mid \theta_n) \geq Q(\theta_n \mid \theta_n)$.

↳ This algorithm is Generalized EM (GEM).

Ex. Bivariate Normal w/ missing data

$$Y = (Y_1, Y_2) \sim N(\mu, \Sigma)$$

$$\mu = (\mu_1, \mu_2), \quad \Sigma = \begin{bmatrix} \sigma^2 & V \\ V & \tau^2 \end{bmatrix}$$

$$\begin{bmatrix} Y_1 & Y_2 \\ 3 & 5 \\ 6.1 & 8.3 \\ ? & 7.4 \\ 4 & ? \end{bmatrix} \qquad \theta = (\mu_1, \mu_2, \sigma^2, \tau^2, V)$$

TOPIC: DATE:

FILE UNDER: PAGE:

Ex. Mixture Models

$Y_1, \dots, Y_n$,

$$f(y_i) = \lambda \varphi(y_i \mid \mu_1, \sigma_1^2) + (1-\lambda)\varphi(y_i \mid \mu_2, \sigma_2^2)$$

$$\theta = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \lambda) \quad , \quad \lambda \in (0,1)$$

$$\ell(\theta \mid Y) = \sum_{i=1}^{n} \log\left\{\lambda \varphi(y_i \mid \mu_1, \sigma_1^2) + (1-\lambda)\varphi(y_i \mid \mu_2, \sigma_2^2)\right\}$$

Suppose $Z \sim \text{Bernoulli}(\lambda)$, s.t. if

$z_i = 1$, then $y_i \sim N(\mu_1, \sigma_1^2)$

$z_i = 0$, then $y_i \sim N(\mu_2, \sigma_2^2)$.

CDL:

$$\cancel{L(\theta \mid Y) = \prod_{i=1}^{n} \varphi(y \mid \mu_1, \sigma_1^2)^{z_i}\varphi(y_i}$$

$$f(y_i \mid z_i) = \varphi(y_i \mid \mu_1, \sigma_1^2)^{z_i}\varphi(y_i \mid \mu_2, \sigma_2^2)^{1-z_i}$$

$$p(z_i) = \lambda^{z_i}(1-\lambda)^{1-z_i}$$

$$\cancel{L(\theta) = \prod_{i=1}}$$

$$\ell(\theta) = \sum_{i=1}^{n} \log f(y_i \mid z_i)\, p(z_i)$$

$$= \sum_{i=1}^{n} z_i \log \varphi(y_i \mid \mu_1, \sigma_1^2) + (1-z_i)\varphi(y_i \mid \mu_2, \sigma_2^2)$$

$$+ z_i \log \lambda + (1-z_i)\log(1-\lambda)$$

$$= \sum_{i=1}^{n} z_i \varphi_1 + (1-z_i)\varphi_2 + z_i \log\lambda + (1-z_i)\log(1-\lambda)$$

TOPIC:

DATE:

FILE UNDER:

PAGE:

$\mathcal{Q}_x =$

$\frac{b}{x} = 1 - \frac{a}{x}$

$\frac{b}{x} = \frac{x-a}{x}$

$x = a+b$

$$p(z_i | y_i) \propto p(y_i | z_i)\, p(z_i)$$

$$= \varphi_1^{z_i} \varphi_2^{1-z_i} \lambda^{z_i} (1-\lambda)^{1-z_i}$$

$$= (\lambda \varphi_1)^{z_i} (\varphi_2 (1-\lambda))^{1-z_i}$$

$$= \text{Bernoulli}\left(\frac{\lambda \varphi_1}{\lambda \varphi_1 + (1-\lambda)\varphi_2} \right)$$

$$\parallel$$

$$\pi_i$$

$$\mathbb{E}[z|y] = \pi_i = \frac{\lambda_0 \varphi(y_i | \mu_{10}, \tau_{10}^2)}{\lambda_0 \varphi(y_i | \mu_{10}, \tau_{10}^2) + (1-\lambda_0)\varphi(y_i | \mu_{20}, \tau_{20}^2)}$$

$$Q(\theta|\theta_n) = \mathbb{E}\left[\sum_{i=1}^{n} z_i \log \varphi_1 + (1-z_i) \log \varphi_2 + C \right]$$

$$= \sum_{i=1}^{n} \pi_i \log \varphi_1 + (1-\pi_i) \log \varphi_2 + \pi_i \log \lambda + (1-\pi_i) \log (1-\lambda)$$

$$= \sum_{i=1}^{n} \pi_i \left[-\tfrac{1}{2} \log 2\pi\sigma_1^2 - \frac{1}{2\sigma_1^2}(y_i - \mu_1)^2 \right] + (1-\pi_i) \left[-\tfrac{1}{2} \log 2\pi\sigma_2^2 - \frac{1}{2\sigma_2^2}(y_i - \mu_2)^2 \right]$$

$$\Rightarrow \hat{\mu}_1 = \frac{\sum \pi_i y_i}{\sum \pi_i} \qquad \hat{\sigma}_1^2 = \frac{\sum (y_i - \hat{\mu}_1)^2 \pi_i}{\sum \pi_i}$$

$$\hat{\mu}_2 = \frac{\sum (1-\pi_i) y_i}{\sum \pi_i} \qquad \hat{\sigma}_2^2 = \frac{\sum (1-\pi_i)(y_i - \hat{\mu}_2)^2}{\sum (1-\pi_i)}$$

$$\hat{\lambda} = \frac{1}{n} \sum \pi_i$$

TOPIC

FILE UNDER:

DATE

PAGE

Ex: One-way random effects

$$Y_{ij} \overset{iid}{\sim} N(\mu_i, \sigma^2)$$

$$Y_{ij} = \mu_i + \varepsilon_i \quad \sim N(0, \sigma^2)$$

$$\mu_i \overset{iid}{\sim} N(\alpha, \tau^2)$$

$$\cancel{p(Y_{i1}, Y_{i2}, \ldots Y_{in_i} \mid \alpha, \tau^2) \propto \int \left[\prod_{j=1}^{n_i} \varphi(Y_{ij} \mid \mu_i, \sigma^2)\right] \varphi(\mu_i \mid \alpha, \tau^2)\, d\mu_i}$$

$$p(Y \mid \mu, \sigma, \tau^2) \propto \prod_{i=1}^{G} \left[\prod_{j=1}^{n_i} \varphi(Y_{ij} \mid \mu_i, \sigma^2)\right] \varphi(\mu_i \mid \alpha, \tau^2)$$

$$\log p = \sum_{i=1}^{G} \sum_{j=1}^{n_i} \log \varphi(y_{ij} \mid \mu_i, \sigma^2) + \log \varphi(\mu_i \mid \alpha, \tau^2)$$

$$= \sum_{i=1}^{G} \left[\sum_{j=1}^{n_i} -\frac{1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(Y_{ij} - \mu_i)^2\right] - \frac{1}{2}\log\tau^2 - \frac{1}{2\tau^2}(\mu_i - \alpha)^2$$

$$N = \sum_{i=1}^{G} n_i$$

$$= -\frac{N}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_i \sum_j (Y_{ij} - \mu_i)^2 - \frac{G}{2}\log\tau^2 - \frac{1}{2\tau^2}\sum_i (\mu_i - \alpha)^2$$

$$\frac{\partial}{\partial\sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma^3}\sum_i \sum_j (Y_{ij} - \mu_i)^2$$

$$-\frac{1}{2}\sigma^{-2}$$
$$+\sigma^{-3}$$
$$-\frac{N}{2\sigma^2} 2\sigma$$

$$Y_{ij}^2 - 2\mu_i Y_{ij} + \mu_i^2$$

$$p(\mu_i \mid \sim) \propto \left[\prod_{j=1}^{n_i} \varphi(Y_{ij} \mid \mu_i, \sigma^2)\right] \varphi(\mu_i \mid \alpha, \tau^2)$$

$$N\left(\hat\alpha + \frac{\hat\tau^2}{\hat\sigma^2/n_i + \hat\tau^2}(\bar Y_i - \hat\alpha),\ \frac{\hat\sigma^2/n_i\, \hat\tau^2}{\hat\sigma^2/n_i + \hat\tau^2}\right)$$

$$Y_{ij}^2 - 2\hat\mu_i Y_{ij} + \hat\mu_i^2 + \hat\sigma_{\mu_i}^2$$

$$\bar Y_i^2 - 2\mu_i \bar Y_i + \hat\mu_i^2 + \hat\sigma_{\mu_i}^2$$

$$(Y_{ij} - \bar Y_i + \bar Y_i - \mu_i)^2$$

$$\sum S_i^2 + (\bar Y_i - \mu_i)^2$$
$$+ \text{cross}$$

$f(y|\theta)$ obs.

$g(y,z|\theta)$ complete

Ascent property of EM algorithm

The sequence $\theta_{n+1} = \arg\max_\theta Q(\theta|\theta_n)$

has the property that for each n,

$$\ell(\theta_{n+1}|y) \geq \ell(\theta_n|y)$$

$$\Rightarrow \log f(y|\theta_{n+1}) \geq \log f(y|\theta_n)$$

w/ strict inequality when $Q(\theta_{n+1}|\theta_n) > Q(\theta_n|\theta_n)$

Jensen's inequality first.

~~Define $D(f\|g) = -\log \mathbb{E}$~~

Define $D(f\|g) = \mathbb{E}_f[\log f/g]$

~~$D(f\|g)$ is zero $\Leftrightarrow$ $f=g$~~

$D(f\|g) \geq 0$ (information inequality)

Pf: $D(f\|g) = \mathbb{E}_f[\log f/g]$

$= \mathbb{E}_f[-\log g/f]$

Jensen's inequality, $\geq -\log \mathbb{E}_f[g/f]$

(−log is a convex function)

$= -\log \int g/f \cdot f$

$= -\log 1$

$= 0$

$D(\cdot\|\cdot)$ is like a distance (but no Δ inequality)

$$h = \frac{g(y,z)}{f(y|\theta)}$$

$$= \log f(y|\theta_0) -$$

$$\log f(y|\theta_0) - \mathbb{E}_h [\log (g(y,z|\theta_0)]$$

$$= \mathbb{E}_h \left[\log \frac{f(y|\theta_0)}{g(y,z|\theta_0)} \right]$$

$$\leq \log \mathbb{E}_h \left[\frac{f(y|\theta_0)}{g(y,z|\theta_0)} \right] = \log 1 = 0$$

$$\log f(y|\theta) \geq Q(\theta|\theta_0) + \boxed{\leq 0}$$

Minorizing function

$$\log f(y|\theta) = \log \int g(y, z|\theta)\, dz$$

MBS class 11/26

$$\log f(y|\theta) - \log f(y|\theta_0)$$

$$= \log \int g(y,z|\theta)\,dz - \log \int g(y,z|\theta_0)\,dz$$

$$= \log \frac{\int g(y,z|\theta)\,dz}{\int g(y,z|\theta_0)\,dz}$$

$$= \log \frac{\int g(y,z|\theta_0)\frac{g(y,z|\theta)}{g(y,z|\theta_0)}\,dz}{\int g(y,z|\theta_0)\,dz}$$

$$= \log \int h(z|y,\theta_0)\frac{g(y,z|\theta)}{g(y,z|\theta_0)}\,dz = \log \mathbb{E}_h\left[\frac{g}{g_0}\right]$$

log is concave

$$\geq \int h(z|y,\theta_0) \log \frac{g(y,z|\theta)}{g(y,z|\theta_0)}\,dz = \mathbb{E}_h\left[\log \frac{g}{g_0}\right]$$

$$= \int h(z|y,\theta_0) \log g(y,z|\theta)\,dz$$

$$- \int h(z|y,\theta_0) \log g(y,z|\theta_0)\,dz$$

$$\log f(y|\theta) \geq \log f(y|\theta_0) + \mathbb{E}_h\left[\log g(y,z|\theta)\right] - \mathbb{E}_h\left[\log g(y,z|\theta_0)\right]$$

Pf

Recall $Q(\theta_{n+1}|\theta_n) \geq Q(\theta_n|\theta_n)$

$$\ell(\theta_{n+1}^{*}) - \cancel{\ell(\theta_n)} = \log f(y|\theta_{n+1}) - \cancel{\log f(y|\theta_n)}$$

$$= \log f(y|\theta_{n+1}) + \underbrace{Q(\theta_{n+1}|\theta_n) - Q(\theta_{n+1}|\theta_n)}_{0}$$

$$= Q(\theta_{n+1}|\theta_n) - \left[Q(\theta_{n+1}|\theta_n) - \log f(y|\theta_{n+1}) \right]$$

$$= Q(\theta_{n+1}|\theta_n) - \left[\mathbb{E}_p\left[\log g(y,z|\theta_{n+1}) \middle| y, \theta_n \right] - \log f(y|\theta_{n+1}) \right]$$

$$= Q(\theta_{n+1}|\theta_n) - \mathbb{E}_p\left[\log \frac{g(y,z|\theta_{n+1})}{f(y|\theta_{n+1})} \middle| y, \theta_n \right]$$

$$= Q(\theta_{n+1}|\theta_n) - \mathbb{E}_p\left[\log p(z|y,\theta_{n+1}) \middle| y, \theta_n \right]$$

$\hookrightarrow p(z|y,\theta_n)$

assumption

$$\geq Q(\theta_n|\theta_n) - \mathbb{E}_p\left[\log p(z|y,\theta_{n+1}) \middle| y, \theta_n \right]$$

by information inequality

$$\geq Q(\theta_n|\theta_n) - \mathbb{E}_p\left[\log p(z|y,\theta_n) \middle| y, \theta_n \right]$$

$$= Q(\theta_n|\theta_n) - \mathbb{E}_p\left[\log \frac{g(y,z|\theta_n)}{f(y|\theta_n)} \middle| y, \theta_n \right]$$

$$= Q(\theta_n|\theta_n) - \mathbb{E}_p\left[\log g(y,z|\theta_n) \middle| y, \theta_n \right] + \log f(y|\theta_n)$$

$$= Q(\theta_n|\theta_n) - Q(\theta_n|\theta_n) + \log f(y|\theta_n)$$

$$= \log f(y|\theta_n)$$

$$= \ell(\theta_n)$$

$\square$

TOPIC: Missing Information

DATE:

FILE UNDER:

PAGE:

$f(y|\theta)$: observed data density

$g(y,z|\theta)$: complete data density

$h(z|y,\theta) \triangleq \dfrac{g(y,z|\theta)}{f(y|\theta)}$ missing data density

$$f(y|\theta) = \frac{g(y,z|\theta)}{h(z|\theta,y)}$$

$$-\log f(y|\theta) = -\log g(y,z|\theta) - \left[-\log h(z|y,\theta)\right]$$

$\dfrac{\partial}{\partial\theta\partial\theta^T}$

$$-\frac{\partial}{\partial\theta\partial\theta^T}\log f(y|\theta) = -\frac{\partial}{\partial\theta\partial\theta^T}\log g(y,z|\theta) - \left[-\frac{\partial}{\partial\theta\partial\theta^T}\log h(z|y,\theta)\right]$$

$\mathbb{E}_{z|y}$

$$\underbrace{I_y(\theta)}_{\text{obs}} = \underbrace{I_{y,z}(\theta)}_{\text{complete}} - \underbrace{I_{z|y}(\theta)}_{\text{missing}}$$

How to compute $I_{y,z}(\theta)$ and $I_{z|y}(\theta)$?

Let $S(y|\theta) = \dfrac{\partial}{\partial\theta}\log f(y|\theta)$

$$S(y,z|\theta) = \frac{\partial}{\partial\theta}\log g(y,z|\theta)$$

TOPIC:

FILE UNDER:

DATE:

PAGE:

Louis 1982 showed:

$$I_{z|y}(\theta) = \mathbb{E}_{z|y}\left[S(y,z|\theta)\, S(y,z|\theta)^T\right] - S(y|\theta)\, S(y|\theta)^T$$

with expectation taken wrt $h(z|y,\theta)$

$$\Rightarrow I_y(\theta) = \underbrace{I_{y,z}(\theta) - \mathbb{E}_{z|y}\left[S(y,z|\theta)S(y,z|\theta)^T\right]}_{\text{Computation only on complete data}} - \underbrace{S(y|\theta)S(y|\theta)^T}_{= 0 \text{ for } \theta = \hat{\theta}}$$

Note $I_{y,z}(\hat{\theta}) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta\partial\theta'} \log g(y,z|\theta) \,\Big|\, \theta_n, y\right]$

$$= -Q''(\hat{\theta} \mid \hat{\theta})$$

Meilijson (89) ~~stated~~ that if y's are iid, then

$\cancel{I_y(\theta) = Var(S(y|\theta))}$

$\cancel{\frac{1}{n}\sum}$

$$S(y|\theta) = \sum_{i=1}^{n} S(y_i|\theta)$$

TOPIC: DATE:

FILE UNDER: PAGE:

$$I_y(\theta) = \mathrm{Var}(s(y|\theta))$$

$$= \frac{1}{n}\sum s(y_i|\theta)s(y_i|\theta)^T - \underbrace{\left[\frac{1}{n}\sum s(y_i|\theta)\right]\left[\frac{1}{n}\sum s(y_i|\theta)\right]^T}_{0 \text{ for } \theta=\hat{\theta}}$$

How?

Louis '82 showed

$$s(y|\theta) = \mathbb{E}_{z|y}\left[s(y,z|\theta)\right]$$

$$I_y(\hat{\theta}) = \frac{1}{n}\sum \mathbb{E}_{z|y}\left[s(y_i,z_i|\hat{\theta})\right]\mathbb{E}_z\left[s(y_i,z_i|\hat{\theta})\right]^T$$

Ex: Mixture models

$$\ell_i(\theta) = z_i \log \varphi(y_i|\mu_1,\sigma_1^2) + (1-z_i)\log(y_i|\mu_2\,\sigma_2) + z_i\log\lambda + (1-z_i)\log(1-\lambda)$$

$$s(y_i z_i|\theta) = \begin{pmatrix} \frac{y_i-\mu_1}{\sigma_1^2} z_i \\ \frac{y-\mu_2}{\sigma_2^2} 1-z_i \\ \left(-\frac{1}{\sigma_1} + \frac{1}{\sigma_1^3}(y_i-\mu_1)\right) z_i \\ \left(-\frac{1}{\sigma_2} + \frac{1}{\sigma_2^3}(y_i-\mu_2)\right)(1-z_i) \\ \frac{z_i}{\lambda} + \frac{1-z_i}{1-\lambda} \end{pmatrix}$$

Aitken Acceleration (Louis 1982)

① Given θ_n, compute θ_{n+1}

② Compute $(1-\hat{J})^{-1} = I_x I_y^{-1}$

③ $\theta^*_{n+1} = \theta_n + (1-\hat{J})^{-1}(\theta_{n+1} - \theta_n)$

Set $\theta_{n+1} = \theta^*_{n+1}$

In neighborhood of MLE

Adaptive Barrier Method

We want to minimize ~~f(x)~~ $f(x)$

subject to $g_i(x) \doteq u_i' x - c_i \geq 0$ (linear constraints)

$u_i \rightarrow$ known

~~the~~ We can use the following surrogate function at iteration K

$$R(x \mid x_n) = f(x) - \lambda \sum_{i=1}^{\ell} \left[g_i(x_n) \log g_i(x) - u_i' x \right]$$

for $\lambda > 0$ $\qquad f(x) - R(x \mid x_n)$ concave + max at x_n

$$f(x_{n+1}) = R(x_{n+1} \mid x_n) + \left[f(x_{n+1}) - R(x_{n+1} \mid x_n) \right]$$

$$\leq R(x_n \mid x_n) + \left[f(x_n) - R(x_n \mid x_n) \right]$$

$$\leq f(x_n)$$