Biostatistics 602 - Statistical Inference
Lecture 25
Bayesian Test & Practice Problems

Hyun Min Kang

April 18th, 2013

Last Lecture

- What is an E-M algorithm?
- When would the E-M algorithm be useful?
- Is MLE via E-M algorithm always guaranteed to converge?
- What are the practical limitations of the E-M algorithm?

Overview of E-M Algorithm (cont'd)

Objective

- Maximize $L(\theta|\mathbf{y})$ or $l(\theta|\mathbf{y})$.
- Let $f(\mathbf{y}, \mathbf{z}|\theta)$ denotes the pdf of complete data. In E-M algorithm, rather than working with $l(\theta|\mathbf{y})$ directly, we work with the surrogate function

$$Q(\theta|\theta^{(r)}) = \mathrm{E}\left[\log f(\mathbf{y}, \mathbf{Z}|\theta)|\mathbf{y}, \theta^{(r)}\right]$$

  where $\theta^{(r)}$ is the estimation of θ in r-th iteration.
- $Q(\theta|\theta^{(r)})$ is the *expected log-likelihood of complete data*, conditioning on the observed data and $\theta^{(r)}$.

Key Steps of E-M algorithm

Expectation Step

- Compute $Q(\theta|\theta^{(r)})$.
- This typically involves in estimating the conditional distribution $\mathbf{Z}|\mathbf{Y}$, assuming $\theta = \theta^{(r)}$.
- After computing $Q(\theta|\theta^{(r)})$, move to the M-step

Maximization Step

- Maximize $Q(\theta|\theta^{(r)})$ with respect to θ.
- The $\arg\max_\theta Q(\theta|\theta^{(r)})$ will be the $(r+1)$-th θ to be fed into the E-step.
- Repeat E-step until convergence

Does E-M iteration converge to MLE?

Theorem 7.2.20 - Monotonic EM sequence

The sequence $\{\hat{\theta}^{(r)}\}$ defined by the E-M procedure satisfies

$$L\left(\hat{\theta}^{(r+1)}|\mathbf{y}\right) \geq L\left(\hat{\theta}^{(r)}|\mathbf{y}\right)$$

with equality holding if and only if successive iterations yield the same value of the maximized expected complete-data log likelihood, that is

$$E\left[\log L\left(\hat{\theta}^{(r+1)}|\mathbf{y},\mathbf{Z}\right)|\hat{\theta}^{(r)},\mathbf{y}\right] = E\left[\log L\left(\hat{\theta}^{(r)}|\mathbf{y},\mathbf{Z}\right)|\hat{\theta}^{(r)},\mathbf{y}\right]$$

Theorem 7.5.2 further guarantees that $L(\hat{\theta}^{(r)}|\mathbf{y})$ converges monotonically to $L(\hat{\theta}|\mathbf{y})$ for some stationary point $\hat{\theta}$.

Bayesian Tests

- Hypothesis testing problems can be formulated in a Bayesian model
- Bayesian model includes
  - Sampling distribution $f(\mathbf{x}|\theta)$
  - Prior distribution $\pi(\theta)$
- Bayesian hypothesis testing is based on the posterior probability
  - In Frequentist's framework, posterior probability cannot be calculated.
  - In Bayesian framework, the probability of H_0 and H_1 can be calculated
    - $\Pr(\theta \in \Omega_0|\mathbf{x}) = \Pr(H_0 \text{ is true})$
    - $\Pr(\theta \in \Omega_0^c|\mathbf{x}) = \Pr(H_1 \text{ is true})$
  - Rejection region can be determined directly based on the posterior probability

Bayesian vs Frequentist Framework

Frequentist's Framework

- θ is considered to be a fixed number
- Consequently, a hypothesis is *either true of false*
  - If $\theta \in \Omega_0$, $\Pr(H_0 \text{ is true}|\mathbf{x}) = 1$ and $\Pr(H_1 \text{ is true}|\mathbf{x}) = 0$
  - If $\theta \in \Omega_0^c$, $\Pr(H_0 \text{ is true}|\mathbf{x}) = 0$ and $\Pr(H_1 \text{ is true}|\mathbf{x}) = 1$

Bayesian Framework

- $\Pr(H_0 \text{ is true}|\mathbf{x})$ and $\Pr(H_1 \text{ is true}|\mathbf{x})$ are function of $\mathbf{x}$, between 0 and 1.
- These probabilities give useful information about the veracity of H_0 and H_1.

Examples of Bayesian hypothesis testing procedure

A neutral test between H_0 and H_1

- Accept H_0 is $\Pr(\theta \in \Omega_0|\mathbf{x}) \geq \Pr(\theta \in \Omega_0^c|\mathbf{x})$
- Reject H_0 is $\Pr(\theta \in \Omega_0|\mathbf{x}) < \Pr(\theta \in \Omega_0^c|\mathbf{x})$
- In other words, the rejection region is $\{\mathbf{x} : \Pr(\theta \in \Omega_0^c|\mathbf{x}) > \frac{1}{2}\}$

A more conservative (smaller size) test in rejecting H_0

- Reject H_0 is $\Pr(\theta \in \Omega_0^c|\mathbf{x}) > 0.99$
- Accept H_0 is $\Pr(\theta \in \Omega_0^c|\mathbf{x}) \leq 0.99$

Example: Normal Bayesian Test

Problem

Let $X_1, \cdots, X_n$ be iid samples $\mathcal{N}(\theta, \sigma^2)$ and let the prior distribution of θ be $\mathcal{N}(\mu, \tau^r)$, where σ^2, μ, and τ^2 are known. Construct a Bayesian test rejecting H_0 if and only if $\Pr(\theta \in \Omega_0|\mathbf{x}) < \Pr(\theta \in \Omega_0^c|\mathbf{x})$

Solution

Consider testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$. From previous lectures, the posterior is

$$\pi(\theta|\mathbf{x}) \sim \mathcal{N}\left(\frac{n\tau^2\overline{x} + \sigma^2\mu}{n\tau^2 + \sigma^2}, \frac{\sigma^2\tau^2}{n\tau^2 + \sigma^2}\right)$$

We will reject H_0 if and only if

$$\Pr(\theta \in \Omega_0|\mathbf{x}) = \Pr(\theta \leq \theta_0|\mathbf{x}) < \frac{1}{2}$$

Solution (cont'd)

Because $\pi(\theta|\mathbf{x})$ is symmetric, this is true if and only if the mean for $\pi(\theta|\mathbf{x})$ is less than or equal to θ_0. Therefore, H_0 will be rejected if

$$\begin{aligned} \frac{n\tau^2\overline{x} + \sigma^2\mu}{n\tau^2 + \sigma^2} &< \theta_0 \\ \overline{x} &< \theta_0 + \frac{\sigma^2(\theta_0 - \mu)}{n\tau^2} \end{aligned}$$

Confidence interval and the parameter

Frequentist's view of intervals

- We have carefully said that the interval *covers* the parameter
- not that the parameter is *inside* the interval, on purpose.
- The random quantity is the interval, not the parameter

Example

- A 95% confidence interval for θ is $.262 \leq \theta \leq 1.184$
- "The probability that θ is in the interval [.262,1.184] is 95%" : Incorrect, because the parameter is assumed fixed
- Formally, the interval [.262,1.184] is one of the possible *realized values* of the random intervals (depending on the observed data)

Bayesian interpretation of intervals

- Bayesian setup allows us to say that θ *is inside* [.262, 1.184] with some probability.
- Under Bayesian model, θ is a random variable with a probability distribution.
- All Bayesian claims of coverage are made with respect to the posterior distribution of the parameter.

Credible sets

- To distinguish Bayesian estimates of coverage, we use *credible sets* rather than confidence sets
- If $\pi(\theta|\mathbf{x})$ is a posterior distribution, for any set $A \subset \Omega$
  - The credible probability of A is $\Pr(\theta \in A|\mathbf{x}) = \int_A \pi(\theta|\mathbf{x})d\theta$
  - and A is a *credible set* (or *creditable interval*) for θ.
- Both the interpretation and construction of the Bayes credible set are more straightforward than those of a classical confidence set, but with additional assumptions (for Bayesian framework).

Example: Possible credible set

Problem

Let $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \text{Poisson}(\lambda)$ and assume that $\lambda \sim \text{Gamma}(a, b)$. Find a 90% credible set for λ.

Solution

The posterior pdf of λ becomes

$$\pi(\lambda|\mathbf{x}) = \text{Gamma}\left(a + \sum x_i, [n + (1/b)]^{-1}\right)$$

If we simply split the α equally between the upper and lower endpoints,

$$\frac{2(nb+1)}{b}\lambda \sim \chi^2_{2(a+\sum x_i)} \quad (\text{if } a \text{ is an integer})$$

Therefore, a $1-\alpha$ confidence interval is

$$\left\{\lambda : \frac{b}{2(nb+1)}\chi^2_{2(\sum x_i + a), 1-\alpha/2} \le \lambda \le \frac{b}{2(nb+1)}\chi^2_{2(\sum x_i+a),\alpha/2}\right\}$$

Remark: Credible probability and coverage probability

- It is important not to confuse credible probability with coverage probability
- Credible probabilities are the Bayes posterior probability, which reflects the experimenter's subjective beliefs, as expressed in the prior distribution.
  - A Bayesian assertion of 90% coverage means that the experimenter, upon combining prior knowledge with data, is 90% sure of coverage
- Coverage probability reflects the uncertainty in the sampling procedure, getting its probability from the objective mechanism of repeated experimental trials.
  - A classical assertion of 90% coverage means that in a long sequence of identical trials, 90% of the realized confidence sets will cover the true parameter.

Practice Problem 1 (from last lecture)

Problem

Suppose $X_1, \cdots, X_n$ are iid samples from $f(x|\theta) = \theta\exp(-\theta x)$. Suppose the prior distribution of θ is

$$\pi(\theta) = \frac{1}{\Gamma(\alpha)\beta^\alpha}\theta^{\alpha-1}e^{-\theta/\beta}$$

where α, β are known.

(a) Derive the posterior distribution of θ.

(b) If we use the loss function $L(\theta, a) = (a-\theta)^2$, what is the Bayes rule estimator for θ?

(a) Posterior distribution of θ

$$
\begin{aligned}
f(\mathbf{x},\theta) &= \pi(\theta)f(\mathbf{x}|\theta)\pi(\theta) \\
&= \frac{1}{\Gamma(\alpha)\beta^\alpha}\theta^{\alpha-1}e^{-\theta/\beta}\prod_{i=1}^n\left[\theta\exp\left(-\theta x_i\right)\right] \\
&= \frac{1}{\Gamma(\alpha)\beta^\alpha}\theta^{\alpha-1}e^{-\theta/\beta}\theta^n\exp\left(-\theta\sum_{i=1}^n x_i\right) \\
&= \frac{1}{\Gamma(\alpha)\beta^\alpha}\theta^{\alpha+n-1}\exp\left[-\theta\left(1/\beta+\sum_{i=1}^n x_i\right)\right] \\
&\propto \text{Gamma}\left(\alpha+n-1,\frac{1}{\beta^{-1}+\sum_{i=1}^n x_i}\right) \\
\pi(\theta|\mathbf{x}) &= \text{Gamma}\left(\alpha+n-1,\frac{1}{\beta^{-1}+\sum_{i=1}^n x_i}\right)
\end{aligned}
$$

(b) Bayes' rule estimator with squared error loss

Bayes' rule estimator with squared error loss is posterior mean. Note that the mean of $\text{Gamma}(\alpha,\beta)$ is $\alpha\beta$.

$$
\begin{aligned}
\pi(\theta|\mathbf{x}) &= \text{Gamma}\left(\alpha+n-1,\frac{1}{\beta^{-1}+\sum_{i=1}^n x_i}\right) \\
E[\theta|\mathbf{x}] &= E[\pi(\theta|\mathbf{x})] \\
&= \frac{\alpha+n-1}{\beta^{-1}+\sum_{i=1}^n x_i}
\end{aligned}
$$

Practice Problem 2

Problem

Suppose $X_1,\cdots,X_n$ are iid random samples from Gamma distribution with parameter $(3,\theta)$, which has the pdf

$$f(x|\theta) = \frac{1}{2\theta^3}x^2e^{-x/\theta} \qquad (x>0)$$

You may use the result that $2\sum_{i=1}^n X_i/\theta \sim \chi^2_{6n}$.

(a) Derive the asymptotic size α LRT for testing $H_0:\theta=\theta_0$ vs. $H_1:\theta\neq\theta_0$.

(b) Derive the UMP level α test for $H_0:\theta=\theta_0$ vs. $H_1:\theta=\theta_1$, where $\theta_1>\theta_0$.

(c) Derive the UMP level α test for $H_0:\theta\leq\theta_0$ vs. $H_1:\theta>\theta_0$.

Solution (a) - Obtaining MLEs

$$
\begin{aligned}
L(\theta|\mathbf{x}) &= \prod_{i=1}^n\left[\frac{1}{2\theta^3}x_i^2e^{-x_i/\theta}\right] \\
l(\theta|\mathbf{x}) &= \sum_{i=1}^n\left[-\log 2-3\log\theta+2\log x_i-\frac{x_i}{\theta}\right] \\
&= -n\log 2-3n\log\theta+2\sum_{i=1}^n\log x_i-\frac{1}{\theta}\sum_{i=1}^n x_i \\
l'(\theta|\mathbf{x}) &= -\frac{3n}{\theta}+\frac{1}{\theta^2}\sum_{i=1}^n x_i = 0 \\
\hat{\theta} &= \frac{1}{3n}\sum_{i=1}^n x_i
\end{aligned}
$$

Solution (a) - Obtaining MLEs

$$l''(\theta|\mathbf{x})\big|_{\theta=\hat\theta} = \frac{3n}{\theta^2} - \frac{2}{\theta^3}\sum_{i=1}^n x_i\Big|_{\theta=\hat\theta} = \frac{3n}{\hat\theta^2} - \frac{6n}{\hat\theta^2} < 0$$

Because $L(\theta|\mathbf{x}) \rightarrow 0$ as θ approaches zero or infinity, $\hat\theta = \frac{1}{3n}\sum_{i=1}^n x_i$.

Solution (a) - Constructing asymptotic size α LRT

The rejection region of asymptotic size α LRT is

$$\begin{aligned}
-2\log\lambda(\mathbf{x}) &= -2\left[l(\theta_0|\mathbf{x}) - l(\hat\theta|\mathbf{x})\right] \\
&= 6n\log\theta_0 + \frac{2}{\theta_0}\sum x_i - 6n\log\hat\theta - \frac{2}{\hat\theta}\sum x_i \\
&= 6n\log\theta_0 + \frac{2}{\theta_0}\sum x_i - 6n\log\left(\frac{1}{3n}\sum x_i\right) - 6n > \chi^2_{1,\alpha}
\end{aligned}$$

$$\begin{aligned}
R &= \left\{\mathbf{x} : \frac{2}{\theta_0}\sum x_i - 6n\log\sum x_i > \chi^2_{1,\alpha} + 6n[1-\log(3n\theta_0)]\right\} \\
&= \left\{\mathbf{x} : \sum x_i - 3n\theta_0\log\sum x_i > \frac{\theta_0}{2}\chi^2_{1,\alpha} + 3n\theta_0[1-\log(3n\theta_0)]\right\}
\end{aligned}$$

Solution (b) - UMP level α test for simple hypothesis

For $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$,

$$\frac{L(\theta_1|\mathbf{x})}{L(\theta_0|\mathbf{x})} = \frac{\frac{1}{2^n\theta_1^{3n}}\exp\left[-\frac{\sum x_i}{\theta_1}\right]\prod x_i^2}{\frac{1}{2^n\theta_0^{3n}}\exp\left[-\frac{\sum x_i}{\theta_0}\right]\prod x_i^2} = \left(\frac{\theta_0}{\theta_1}\right)^{3n}\exp\left[\frac{\theta_1-\theta_0}{\theta_0\theta_1}\sum x_i\right]$$

Solution (b) - UMP level α test (cont'd)

Let $T = \sum X_i$. Then under H_0, $\frac{2}{\theta_0}T \sim \chi^2_{6n}$.

$$\begin{aligned}
\alpha &= \Pr\left[\left(\frac{\theta_0}{\theta_1}\right)^{3n}\exp\left[\frac{\theta_1-\theta_0}{\theta_0\theta_1}T\right] > k\right] \\
&= \Pr(T > k^*)
\end{aligned}$$

So, the rejection region is

$$R = \left\{\mathbf{x} : T(\mathbf{x}) = \sum x_i > \frac{\theta_0}{2}\chi^2_{6n,\alpha}\right\}$$

Solution (c) - UMP level α test for composite hypothesis

We need to check whether T has MLR. Because $Y = 2T/\theta \sim \chi^2_{6n}$.

$$f_Y(y|\theta) = \frac{1}{2^{3n}\Gamma(3n)} y^{3n-1} e^{-y/2}$$

$$f_T(t|\theta) = \frac{1}{2^{3n-1}\Gamma(3n)\theta}\left(\frac{2t}{\theta}\right)^{3n-1} e^{-t/\theta} = \frac{1}{\Gamma(3n)\theta}\left(\frac{t}{\theta}\right)^{3n-1} e^{-t/\theta}$$

For arbitrary $\theta_1 < \theta_2$,

$$\frac{f_T(t|\theta_2)}{f_T(t|\theta_1)} = \frac{\frac{1}{\Gamma(3n)\theta_2}\left(\frac{t}{\theta_2}\right)^{3n-1} e^{-t/\theta_2}}{\frac{1}{\Gamma(3n)\theta_1}\left(\frac{t}{\theta_1}\right)^{3n-1} e^{-t/\theta_1}} = \left(\frac{\theta_1}{\theta_2}\right)^{3n} \exp\left[\frac{\theta_2-\theta_1}{\theta_1\theta_2} t\right]$$

is an increasing function of t. This T has MLR property.

Solution (c) - Constructing UMP level α test

Because T has MLR property, UMP level α test for $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$ has a rejection region $T > k$, and $\Pr(T > k) = \alpha$. Therefore, the UMP level α test is identical to the answer of part (b), whose rejection is

$$R = \left\{\mathbf{x} : T(\mathbf{x}) = \sum x_i > \frac{\theta_0}{2}\chi^2_{6n,\alpha}\right\}$$

Practice Problem 3

Problem

Let $(X_1, Y_1), \cdots, (X_n, Y_n)$ be a random samples from a bivariate normal

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}\right)$$

We are interested in testing $H_0 : \mu_X = \mu_Y$ vs. $H_1 : \mu_X \neq \mu_Y$.

(a) Show that the random variables $W_i = X_i - Y_i$ are iid $\mathcal{N}(\mu_W, \sigma^2_W)$.

(b) Show that the above hypothesis can be tested with the statistic

$$T_W = \frac{\overline{W}}{\sqrt{S_W^2/n}}$$

where $\overline{W} = \frac{1}{n}\sum_{i=1}^n W_i$ and $S_W^2 = \frac{1}{n-1}\sum_{i=1}^n (W_i - \overline{W})^2$. Furthermore, show that, under H_0, T_W follows the Student's t distribution with $n-1$ degrees of freedom.

Solution (a)

To solve Problem (a), we first need to know that, if $\mathbf{Z} \sim \mathcal{N}(\mathbf{m}, \Sigma)$, then

$$A\mathbf{Z} \sim \mathcal{N}(A\mathbf{m}, A\Sigma A^T)$$

Let $\mathbf{Z} = [X_i \ Y_i]^T$, $\mathbf{m} = [\mu_X \ \mu_Y]^T$, and $A = [1 \ -1]$. Then

$$\begin{aligned} A\mathbf{Z} &= X_i - Y_i = W_i \\ &\sim \mathcal{N}(A\mathbf{m}, A\Sigma A^T) \\ &= \mathcal{N}(\mu_X - \mu_Y, \sigma_X^2 - 2\rho\sigma_X\sigma_Y + \sigma_Y^2) \\ &= \mathcal{N}(\mu_W, \sigma_W^2) \end{aligned}$$

Solution (b)

Because $\mu_W = \mu_X - \mu_Y$, testing

$$H_0 : \mu_X = \mu_Y \quad \text{vs.} \quad H_1 : \mu_X \neq \mu_Y$$

is equivalent to testing

$$H_0 : \mu_W = 0 \quad \text{vs.} \quad H_1 : \mu_W \neq 0$$

When $U_i \sim \mathcal{N}(\mu, \sigma^2)$ and both mean and variance are unknown, we know that LRT testing $H_0 : \mu = \mu_0$ vs. $H_0 : \mu \neq \mu_0$ follows that

$$T_U = \frac{\overline{U} - \mu_0}{\sqrt{S_U^2/n}}$$

and T_U follows T_{n-1} under H_0.

Solution (b) (cont'd)

Therefore, the LRT test for the original test, $H_0 : \mu_W = 0$ vs. $H_1 : \mu_W \neq 0$ is

$$T_W = \frac{\overline{W}}{\sqrt{S_W^2/n}}$$

and T_W follows T_{n-1} under H_0.

Practice Problem 4

Problem

Let $f(x|\theta)$ be the logistic location pdf

$$f(x|\theta) = \frac{e^{(x-\theta)}}{(1+e^{(x-\theta)})^2} \qquad -\infty < x < \infty, \ -\infty < \theta < \infty$$

(a) Show that this family has an MLR

(b) Based on one observation X, find the most powerful size α test of $H_0 : \theta = 0$ versus $H_1 : \theta = 1$.

(c) Show that the test in part (b) is UMP size α for testing $H_0 : \theta \leq 0$ vs. $H_1 : \theta > 0$.

Solution for (a)

For $\theta_1 < \theta_2$,

$$\frac{f(x|\theta_2)}{f(x|\theta_1)} = \frac{\frac{e^{(x-\theta_2)}}{(1+e^{(x-\theta_2)})^2}}{\frac{e^{(x-\theta_1)}}{(1+e^{(x-\theta_1)})^2}} = e^{(\theta_1-\theta_2)}\left(\frac{1+e^{(x-\theta_1)}}{1+e^{(x-\theta_2)}}\right)^2$$

Let $r(x) = (1+e^{x-\theta_1})/(1+e^{x-\theta_2})$

$$r'(x) = \frac{e^{(x-\theta_1)}(1+e^{(x-\theta_2)}) - (1+e^{(x-\theta_1)})e^{(x-\theta_2)}}{(1+e^{(x-\theta_2)})^2} = \frac{e^{(x-\theta_1)} - e^{(x-\theta_2)}}{(1+e^{(x-\theta_2)})^2} > 0 \qquad (\because x-\theta_1 > x-\theta_2)$$

Therefore, the family of X has an MLR.

Solution for (b)

The UMP test rejects H_0 if and only if

$$\begin{aligned}
\frac{f(x|1)}{f(x|0)} &= e\left(\frac{1+e^x}{1+e^{(x-1)}}\right)^2 > k \\
\frac{1+e^x}{1+e^{(x-1)}} &> k^* \\
\frac{1+e^x}{e+e^x} &> k^* \\
X &> x_0
\end{aligned}$$

Because under H_0, $F(x|\theta=0) = \frac{e^x}{1+e^x}$, the rejection region of UMP level α test satisfies

$$\begin{aligned}
1 - F(x|\theta=0) &= \frac{1}{1+e^{x_0}} = \alpha \\
x_0 &= \log\left(\frac{1-\alpha}{\alpha}\right)
\end{aligned}$$

Solution for (c)

Because the family of X has an MLR, UMP size α for testing $H_0 : \theta \leq 0$ vs. $H_1 : \theta > 0$ should be a form of

$$\begin{aligned}
X &> x_0 \\
\Pr(X > x_0|\theta = 0) &= \alpha
\end{aligned}$$

Therefore, $x_0 = \log\left(\frac{1-\alpha}{\alpha}\right)$, which is identical to the test defined in (b).