Math 775: Homework 6

Alex Dewey

May 11, 2016

1 Notes

2. Input Data 6.2. 2.13.

2 Exercises

We observe independent Bernoulli variables X_i which depend on draws from unobservable normal variables $Z_i \sim N(\zeta, \sigma^2)$– $X_i = 1 \iff Z_i > u$ for some known u. We're interested in MLEs for ζ and σ^2.

Exercise 1.

a. The likelihood is $p^S(1-p)^{n-S}$ pretty much by construction. We draw excchangeable pairs $X_i = 1\ Z_i > u$ with probability equal to a standard normal draw from $p = P(Y(\sim N(0,1)) > \frac{u-\zeta}{\sigma})$ and $S = \sum X_i$ is just the natural sufficient statistic for this process.

b. If we consider z_i to be the complete data, then we're drawing n IID values from $N(\zeta, \sigma^2$.

The form

$$\prod \frac{1}{\sqrt{2\pi}\sigma} \exp - \frac{1}{2\sigma^2}(z_i - \zeta)^2$$

is just the product of those n independent distributions.

To get the log-likelihood we just take the logarithm of the above:

$$\log(p(z_i|x_i, \zeta, \sigma^2) = \log \frac{1}{(\sqrt{2\pi}\sigma)^n} + -\frac{1}{2\sigma^2} \sum (z_i - \zeta)^2 =$$

$$-\frac{n}{2} \log \frac{1}{2\pi\sigma^2} + -\frac{1}{2\sigma^2} \sum (z_i^2 - 2\zeta z_i + \zeta^2)$$

.

Since we're interested in the log-likelihood given only the observed variables x_i, we get the expected log-likelihood by replacing z_i, z_i^2 in the above with $E[z_i|x_i], E[z_i^2|x_i]$

c. We've already done the expectation step, now we need to do the maximization step by taking derivatives and finding the posterior mode for ζ, σ^2. ζ is easier to start with.

$$\frac{d}{d\zeta} E[\log(p(z_i|x_i, \zeta, \sigma^2)] = -\frac{1}{2\sigma^2} \sum (2E[z_i|x, \zeta, \sigma^2] - 2\zeta)$$

.

Setting this equal to zero yields $\hat{\zeta} = E[z_i|X, \zeta, \sigma^2]$.

$$\frac{d}{d\sigma^2} E[\log(p(z_i|x_i, \zeta, \sigma^2)] = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum (E[z_i^2|x, \zeta, \sigma^2] - 2\zeta E[z_i|x, \zeta, \sigma^2] + \zeta^2)$$

.

Now, for σ we use the mode for $\hat{\zeta}$ to simplify the quadratic expression to $E[z_i^2|x, \zeta, \sigma^2] - E[z_i|x, \zeta, \sigma^2]^2$. Setting this equal to zero, and rearranging:

$$\frac{n}{\sigma^2} = \frac{1}{(\sigma^2)^2} \sum (E[z_i^2|x, \zeta, \sigma^2] - E[z_i|x, \zeta, \sigma^2]^2)$$
$$n\sigma^2 = \sum (E[z_i^2|x, \zeta, \sigma^2] - E[z_i|x, \zeta, \sigma^2]^2)$$

.

We get

$$\hat{\sigma^2} = \frac{1}{n} \big[\sum E[z_i^2|x, \zeta, \sigma^2] - \frac{1}{n} \sum [z_i|x, \zeta, \sigma^2]^2) \big]$$

through simple manipulations of the interior. This is the maximization step, so this completes the algorithm.

d. First note that the function
$H_i(t) = \frac{\psi(t)}{1-\phi(t)}$ if $X_i = 1$, and $H_i(t) = -\frac{\psi(t)}{\phi(t)}$ if $X_i = 0$.
is just a censored normal above and below the standardized score $t = \frac{u-\zeta}{\sigma}$. So it represents the two possible expected values that Z_i can take on when we know (because we know the value of x_i) that Z_i is either above or below u, and we have no other information, so all we can do is take the expected value for a given draw from the left- or right-truncated normal.

This is made much more obvious

$$E[z_i|x, \zeta, \sigma^2] = \zeta + \sigma H_i(\frac{u - \zeta}{\sigma})$$

simply if it's written as:

$$\frac{E[z_i|x, \zeta, \sigma^2] - \zeta}{\sigma} = H_i(\frac{u - \zeta}{\sigma})$$

.

As for:

$$E[z_i^2|x, \zeta, \sigma^2] = \zeta^2 + \sigma^2 + \sigma(u + \zeta) H_i(\frac{u - \zeta}{\sigma})$$

We're interested in the expectation of Z_i^2 over the truncated left- or right-normal.

This is a common distribution but normal integrals are notoriously difficult to integrate analytically (it's just integration by parts), so I looked up the variance for a right-truncated normal distribution to demonstrate that the above is right. $t = \frac{u-\zeta}{\sigma}$ for space constraints.

The expectation is

$$E(Z|Z_i < u) = \zeta - \sigma\frac{\psi(t)}{\phi(t)}$$

and the variance is

$$Var(Z|Z_i < u) = \sigma^2(1 - t\frac{\psi(t)}{\phi(t)} - (\frac{\psi(t)}{\phi(t)})^2$$

$E[Z^2] = Var[Z] + E[Z]^2$.

$$E(Z^2|Z_i < u) = \zeta^2 - 2\zeta\sigma\frac{\psi(t)}{\phi(t)} + \sigma^2\frac{\psi(t)}{\phi(t)}^2 + \sigma^2[1 - t\frac{\psi(t)}{\phi(t)} - (\frac{\psi(t)}{\phi(t)})^2] =$$

$$\zeta^2 - 2\zeta\sigma\frac{\psi(t)}{\phi(t)} + \sigma^2 - \sigma^2\frac{u-\zeta}{\sigma}\frac{\psi(t)}{\phi(t)} = \zeta^2 + 2\zeta\sigma h(t) + \sigma^2 + \sigma(u-\zeta)h(t) =$$

$$\zeta^2 + \zeta\sigma h(t) + \sigma^2 + \sigma u h(t) = \zeta^2 + \sigma^2 + \sigma(\zeta + u)h(t)$$

e. By Theorem 5.3.5 in (Robert/Casella, 1999, pp. 215), we know that as long as our expected complete data-likelihood $Q(\theta|\theta_0, x)$ is continuous in our prior $\zeta_0\,\sigma_0$ and our parameters ζ, σ, "every limit point of an EM sequence" is a stationary point of $L(\theta|x)$.

Robert/Casella have a neat proof which analytically computes the iterative steps, but in this case, it suffices to note Theorem 5.3.5 and that every stationary point is an MLE in this model.

Proof: The Z_i are normal and IID so we're effectively computing a linear regression on with spherical errors Z_i. Takeaway: One or more MLEs for $(\hat{\zeta}, \hat{\sigma^2})$ are guaranteed to exist, so it's a well-defined question to ask if they will converge to one of these MLEs.

Furthermore: If a point $(\hat{\zeta}, \hat{\sigma^2})$ is a fixed point after iterating in part (c.), then it also satisfies the ordinary least squares equations for those parameters, and so it is an MLE.

Finally: If a point is not stationary after the iteration in part (c), Theorem 5.3.5 guarantees that it will converge to a fixed point (which by the argument in the above paragraph) is an MLE.