Computational Statistics - STAT 575 - HW #2 - Covariance Estimation

Alex Towell (atowell@siue.edu)

Covariance estimation: bootstrap vs observed fisher information

We use the functions defined by the package *prob.4.2.comp.stats*.

```
if (!require(devtools)) { install.packages('devtools') }
install_github("queelius/prob.4.2.comp.stats")
library("prob.4.2.comp.stats")
options(tinytex.engine_args = '-shell-escape')
```

We invoke the EM algorithm on the observed count data.

```
counts <- c(379,299,222,145,109,95,73,59,45,30,24,12,4,2,0,1,1)

# theta is initial guess for parameter vector,
#   theta = (alpha, beta, mu, lambda)' = (1/3,1/3,2,3).
# the initial value assumes each category is equally probable and mu^(0) and
# lambda^(0) are just arbitrarily chosen to be 2 and 3, with the insight that
# group 3 is more risky than group 2.
theta <- c(1/3,1/3,2,3)
sol <- em.estimator(theta,counts)
theta.em <- sol$estimate
```

We see that $\hat{\theta}_{\text{EM}} = (0.1221648, 0.5625419, 1.4674664, 5.9388786)'$. We use the Bootstrap method to estimate the covariance matrix of $\hat{\theta}_{\text{EM}}$ in the following:

```
cov.bs <- em.cov.bs(theta.em,counts,m=10000)
knitr::kable(cov.bs)
```

	alpha	beta	mu	lamda
alpha	0.0003954	-0.0001803	0.0015914	0.0014345
beta	-0.0001803	0.0004765	0.0001789	0.0016630
mu	0.0015914	0.0001789	0.0126316	0.0139474
lamda	0.0014345	0.0016630	0.0139474	0.0404321

For comparison, we use the observed information matrix, the Hessian of the observed log-likelihood evaluated at $\hat{\theta}_{\text{EM}}$,

$$\hat{\Sigma} = \mathrm{H}(-\ell(\theta))|_{\hat{\theta}_{\text{EM}}}.$$

```
cov.ob <- em.cov.info(theta.em,counts)
knitr::kable(cov.ob)
```

	alpha	beta	mu	lamda
alpha	0	0.0000000	0.0000000	0.0000000
beta	0	0.0005283	0.0000000	0.0000000
mu	0	0.0000000	0.0104683	0.0000000
lamda	0	0.0000000	0.0000000	0.0053377

We see that these covariance matrices appear to be different. It is not clear why. We have separated out the relevant code in the package and put it into the following code block. Dr. Q, do you see anything wrong with this?

```
em.cov.info(theta.em,counts)
```

```
##       alpha         beta          mu       lamda
## alpha     0 0.0000000000 0.00000000 0.000000000
## beta      0 0.0005283146 0.00000000 0.000000000
## mu        0 0.0000000000 0.01046833 0.000000000
## lamda     0 0.0000000000 0.00000000 0.005337692
```