Applied Bayesian Data Analysis — Chapter 9

Kim Albertsson

CERN and Luleå University of Technology

kim.albertsson@ltu.se

October 23, 2019

Chapter 9

Hierarchical models:
When parameters depend upon parameters

Caveat: Minimal coverage of JAGS.

Personal reflections

- Hierarchical models: neat technique to lend statistical strength to model. Like priors, it seems an artform, as opposed to rigorous science. Not saying science cannot be done, just that it's "softer".
- Good examples to give one experience to lean back on. Explains intuitions behind and interpretations of distributions and paramters. E.g. mode of the beta distribution.
- Advice for double checking resutls helpful. E.g.
- Weak discussion on meaningful differences between predictions. So far only "HDI interval covers null hypotesis" therefore no difference between groups. Unsure if I use "null hypothesis" correctly here.

Trick — useful insight applicable once;
Technique — insight with repeated, general, usefulness.

Introduction

Hierarchical models arise in many situations

- E.g. patient recovery after surgery. (Team success rate (θ_s) depends on ω, recovery rate for hospital.)
- E.g. student buying lunch. (Child prob. of buying (θ_s) depends on ω, typical buy rate for school.)

Hierarchical refactoring:

$$
\begin{aligned}
p(\theta, \omega | D) &\propto p(D|\theta, \omega) p(\theta, \omega) \\
&= p(D|\theta) p(\theta|\omega) p(\omega) \qquad (9.1)
\end{aligned}
$$

Compare to Eq. (5.3) applied to the joint parameter probability

$$
\begin{aligned}
p(\theta, \omega | D) &\propto p(D|\theta, \omega) p(\theta, \omega) \\
&= p(D|\theta, \omega) p(\theta|\omega) p(\omega) \qquad \text{(Using 5.3)}
\end{aligned}
$$

I.e. in hierarchical models, data is not dependent on *hidden* parameters.

Introduction

Benefits of hierarchical models:

- Interpretability
- Data efficiency
- Sampling efficiency (e.g. Gibbs sampling depends upon conditionals)

A single coin from a single mint

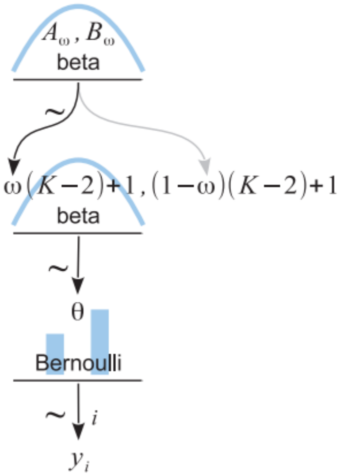
Hierarchical coin model: Factory produces coins with a typical bias.

$$\gamma_i \sim Bernoulli(\theta) \tag{9.2}$$
$$\theta \sim \beta(a_\theta, b_\theta) \tag{9.3}$$
$$\theta \sim \beta(\omega(\kappa - 2) + 1, (1 - \omega)(\kappa - 2) + 1) \tag{9.4}$$
$$\omega \sim \beta(a_\omega, b_\omega) \tag{9.5}$$
$$\kappa = K$$

ω is both mode of β and parameter of model. κ is concentration of β. Figure: Graphical representation of eq. system.

$$\begin{aligned} p(\theta, \omega|\gamma) &= \frac{p(\gamma|\theta, \omega)p(\theta, \omega)}{p(\gamma)} \\ &= \frac{p(\gamma|\theta)p(\theta|\omega)p(\theta)}{p(\gamma)} \end{aligned} \tag{9.6}$$

From def.: γ depends only on θ. From Eq. 5.3: factor joint.



Aside: Hierarchy of data

Given

$$\gamma_i \sim Bernoulli(\theta) \tag{9.2}$$

$$\theta \sim \beta(a, b) \tag{9.3}$$

$$p(\theta|\gamma) = p(\gamma|\theta)p(\theta)$$

I.e. the posterior is factored into a chain of dependencies.

Can we consider the system $\{\gamma, \theta\}$ a hierarchical model? Why; Why not?

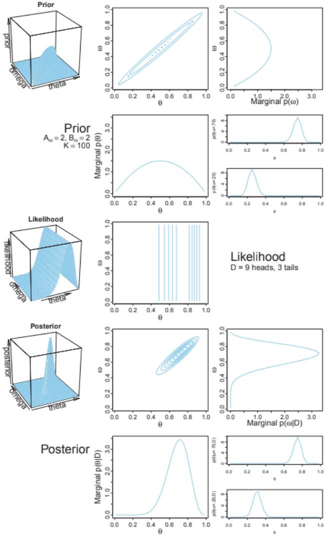
Eq. 9.6 is difficult to solve analytically. Approximate with grid.

Prior uses $b_\omega = 2, b_\omega = 2, K = 100$. I.e. if we know ω we're fairly certain about θ (see $p(\theta|\omega = 0.25)$ plot).

Likelihood contours parallel, data depends only on θ.

Data mostly heads implies ω biased toward upper right quadrant due to tight coupling θ, ω.

Note: Could use second coin to improve estimate of first!

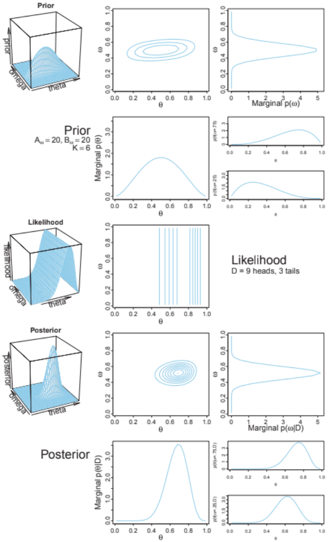


Posterior through grid approximation (II)

Prior uses $b_\omega = 20, b_\omega = 20, K = 6$.
I.e. ω certain but $\theta|\omega$ uncertain.

Posterior mostly refines idea about θ.

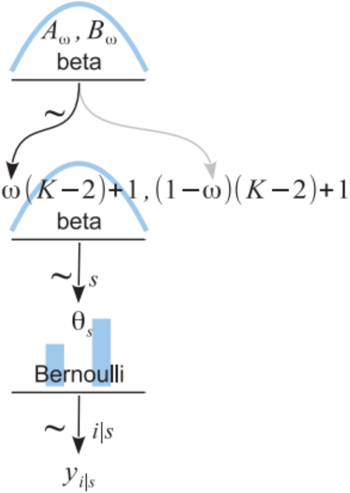
Warning: Mismodelling? See posterior conditional plots. (Model supposes θ close to ω.)



Multiple coins from single mint (I)

Typical case: E.g. testing memory after taking new drug (Note: Acutal variable of interest is the hidden one).

Same as previous figure, with multiple θs.



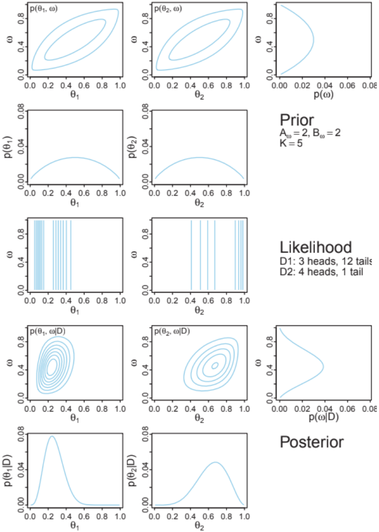
Multiple coins from single mint (II)

Weak coupling between coins ($K = 5$).

$$p(\theta_1, \theta_2, \omega|D) = \frac{p(D|\theta_1, \theta_2, \omega)p(\theta_1, \theta_2, \omega)}{p(D)}$$
$$= \frac{p(D_1|\theta_1)p(D_2|\theta_2)p(\theta_1|\omega)p(\theta_2|\omega)p(\omega)}{p(D)}$$

(Not shown) Conditionals ($p(\theta_s|\omega)$) are spread out.

Datasets imply spread in coin bias. Does little for ω.



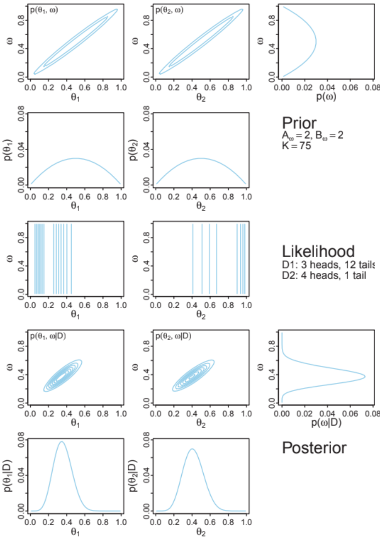
Multiple coins from single mint (III)

Strong coupling between coins ($K = 75$).

(Not shown) Conditionals ($p(\theta_s|\omega)$) are concentrated.

Datasets are contradictory (with strong coupling). Data from larger set dominates. Constrains ω.

Grid approximation only feasible in low dimensions.



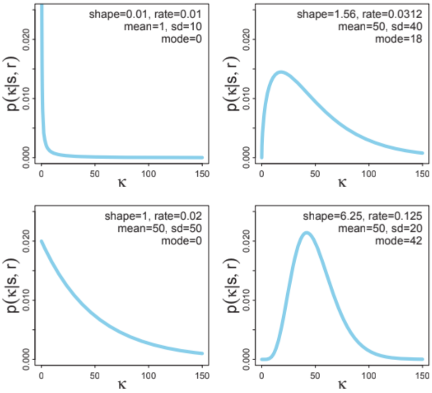
Gamma distribution

Generalised exponential distribution much like β is generalised uniform (among others).

$\mathrm{Gamma}(\kappa|s, r = \theta^{-1})$; where s: shape, r: rate

$$s = \frac{\mu^2}{\sigma^2};\ r = \frac{\mu}{\sigma^2};\ \mu > 0\ (\text{mean}) \tag{9.7}$$

$$s = 1 + \omega r;\ r = \frac{\omega\sqrt{\omega^2 + 4\sigma^2}}{2\sigma^2};\ \omega > 0\ (\text{mode}) \tag{9.8}$$



Related distributions:

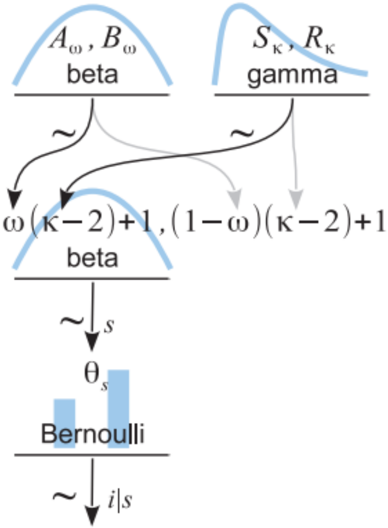
- Exponential distribution $\exp(\lambda) = \mathrm{Gamma}(1, \lambda^{-1})$, models e.g. random decay of particles or electronical components.
- Chi-square $\chi^2(k) = \mathrm{Gamma}(k/2, 2)$, models sum of k independent standard normal random variables.
- ... and many more ...

A realistic model with MCMC

Add realism by

- Relax constraint on κ (draw from Gamma).
- More subjects/coins!

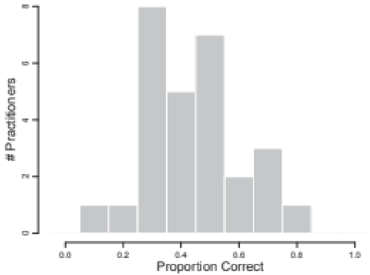
Intuitively, for each coin, κ small: data clustered. If κ large: data spread out.



Therapeutic touch (I)

Recreate investigation of Rosa et al. (1998): Can practitioners feel a body's energy field?

- Flip coin and place experimenter hand near practitioner hand.
- Practitioner reports which hand is closest (binary classification).
- 28 subjects (with caveat)
- Skill of each practitioner: θ_s.
- Effect in group: ω.
- Data in Figure.



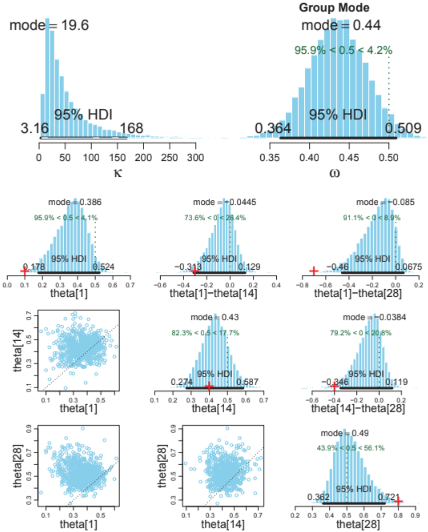
Therapeutic touch (II)

κ shows high auto-correlation (common for chains for high-level parameters). Accuracy of κ not a concern (Why: small changes in κdoes induces only small change in other variables).

Practitioners, as a class: No effect.

Plus shows score of individual. Deviates from mode due to hierarchical modelling.

Differences: Do subjects perform differently? No!

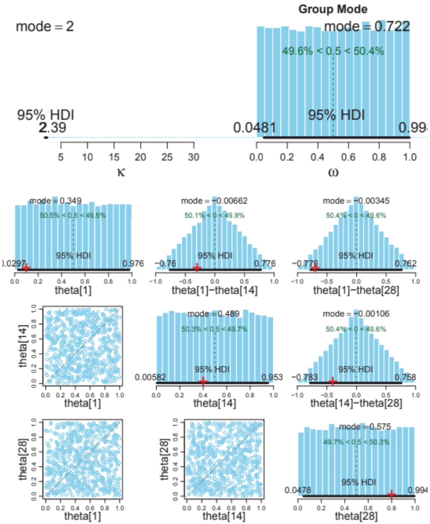


Priors in mid-level paramters

Priors (imputed from top-level priors) are sometimes unintended.

In our case: Group mode prior is uniform. This results in uniform priors in θ.

Priors on differences: Difference between two uniform distributions (triangular).

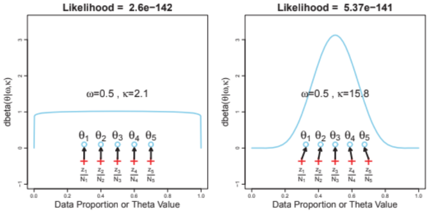


Shrinkage

Low level paramters cluster where high-level paramters are likely, called *shrinkage*.

In previous example: Estimates closer to group mode than sample mean.

HL parameters influenced by all data.
Low level, only by individual data.
Property of hierarchical models.



End!

Aside: Efficiency vs. Effectiveness vs. Efficacy

According to the Cambrigde Dictionary:

Efficiency the good use of time and energy in a way that does not waste any

Effectiveness the ability to be successful and produce the intended results

Efficacy the ability, especially of a medicine or a method of achieving something, to produce the intended result (Am. E.: the quality of being effective; effectiveness)

Efficient working or operating quickly and effectively in an organized way

Effective successful or achieving the results that you want

Efficacious able to produce the intended result

Remaining figures

The rest of the chapter is mainly a more involved example, so figures are only included for discusion if necessary.

Figure 9.13

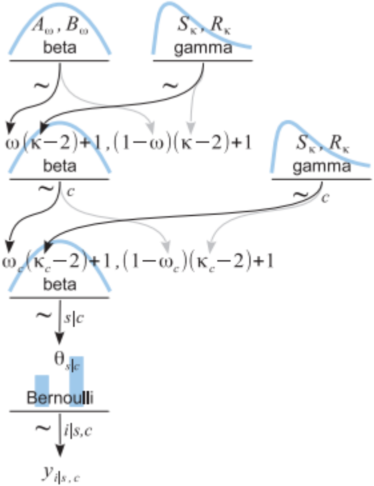

A_ω, B_ω
beta
S_κ, R_κ
gamma
~
~
$\omega(\kappa-2)+1, (1-\omega)(\kappa-2)+1$
beta
S_κ, R_κ
gamma
~
c
~
c
$\omega_c(\kappa_c-2)+1, (1-\omega_c)(\kappa_c-2)+1$
beta
~
$s|c$
$\theta_{s|c}$
Bernoulli
~
$i|s,c$
$y_{i|s,c}$

Figure 9.14

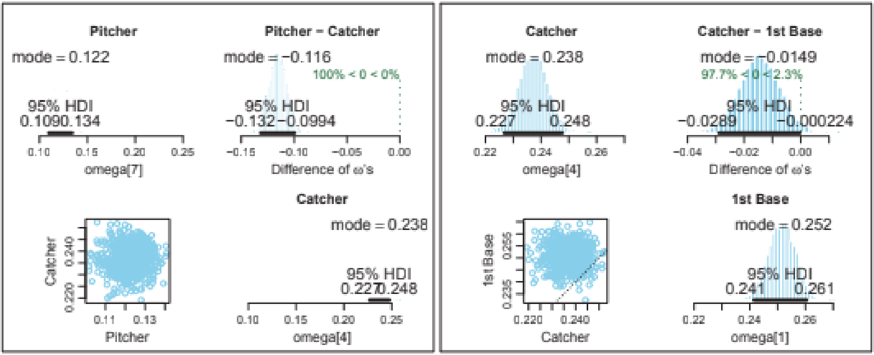
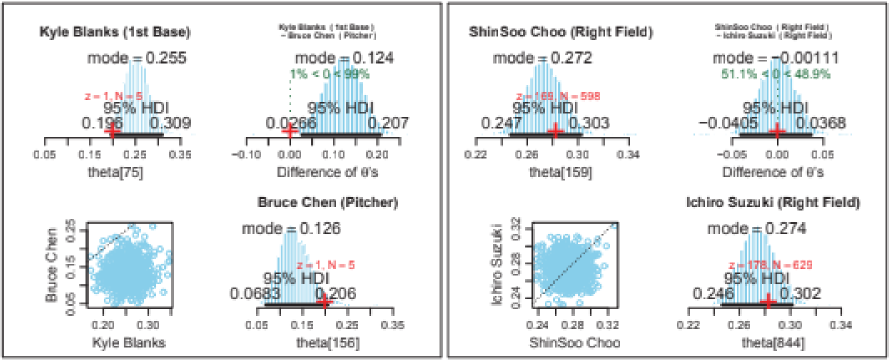
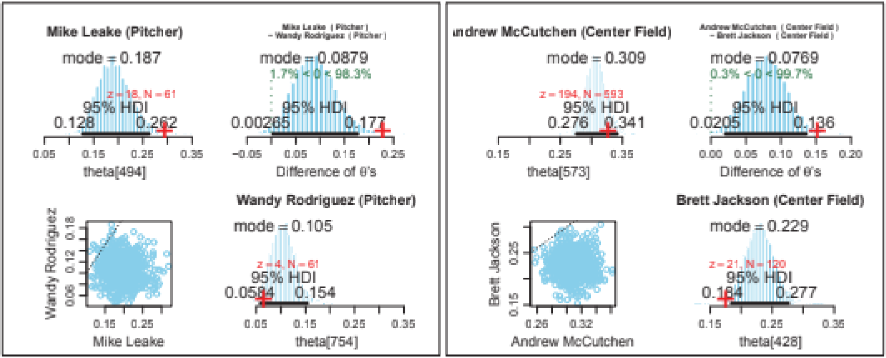

Pitcher
mode = 0.122
95% HDI
0.109 0.134
omega[7]
Pitcher - Catcher
mode = −0.116
100% < 0 < 0%
95% HDI
−0.132 −0.0994
Difference of ω's
Catcher
Pitcher
Catcher
mode = 0.238
95% HDI
0.227 0.248
omega[4]
Catcher
mode = 0.238
95% HDI
0.227 0.248
omega[4]
Catcher - 1st Base
mode = −0.0149
97.7% < 0 < 2.3%
95% HDI
−0.0289 −0.000224
Difference of ω's
1st Base
Catcher
1st Base
mode = 0.252
95% HDI
0.241 0.261
omega[1]

Figure 9.15



Kyle Blanks (1st Base)
mode = 0.255
z = 1, N = 5
95% HDI
0.198
0.309
theta[75]
Kyle Blanks (1st Base) − Bruce Chen (Pitcher)
mode = 0.124
1% < 0 < 99%
95% HDI
0.0266
0.207
Difference of θ's
ShinSoo Choo (Right Field)
mode = 0.272
z = 169, N = 598
95% HDI
0.247
0.303
theta[159]
ShinSoo Choo (Right Field) − Ichiro Suzuki (Right Field)
mode = −0.00111
51.1% < 0 < 48.9%
95% HDI
−0.0405
0.0368
Difference of θ's
Bruce Chen
Kyle Blanks
Bruce Chen (Pitcher)
mode = 0.126
z = 1, N = 5
95% HDI
0.0683
0.206
theta[156]
Ichiro Suzuki
ShinSoo Choo
Ichiro Suzuki (Right Field)
mode = 0.274
z = 178, N = 629
95% HDI
0.246
0.302
theta[844]

Figure 9.16



Mike Leake (Pitcher)
mode = 0.187
z = 18, N = 61
95% HDI
0.128 0.262
theta[494]
Mike Leake (Pitcher) - Wandy Rodriguez (Pitcher)
mode = 0.0879
1.7% < 0 < 98.3%
95% HDI
0.00265 0.177
Difference of θ's
Wandy Rodriguez
Mike Leake
Wandy Rodriguez (Pitcher)
mode = 0.105
z = 4, N = 61
95% HDI
0.0584 0.154
theta[754]
Andrew McCutchen (Center Field)
mode = 0.309
z = 194, N = 593
95% HDI
0.276 0.341
theta[573]
Andrew McCutchen (Center Field) - Brett Jackson (Center Field)
mode = 0.0769
0.3% < 0 < 99.7%
95% HDI
0.0205 0.136
Difference of θ's
Brett Jackson
Andrew McCutchen
Brett Jackson (Center Field)
mode = 0.229
z = 21, N = 120
95% HDI
0.184 0.277
theta[428]

Figure 9.17

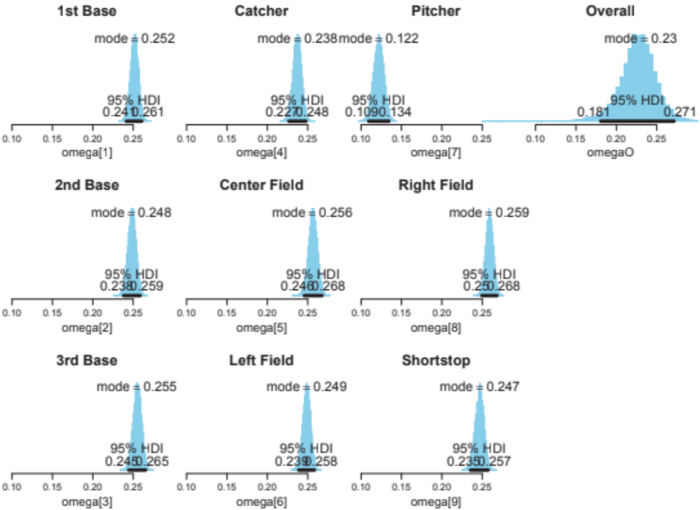
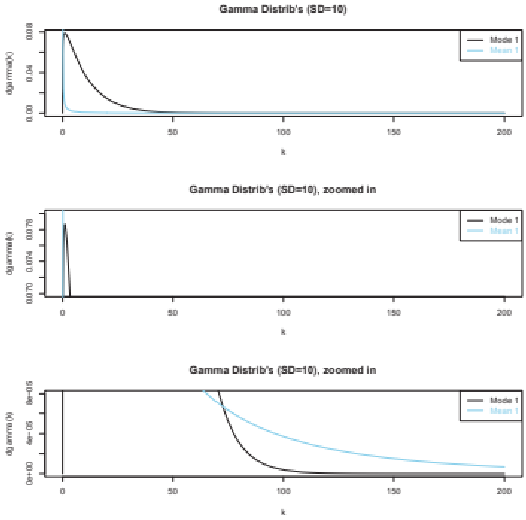

1st Base
mode = 0.252
95% HDI
0.241 0.261
omega[1]
Catcher
mode = 0.238
95% HDI
0.227 0.248
omega[4]
Pitcher
mode = 0.122
95% HDI
0.109 0.134
omega[7]
Overall
mode = 0.23
95% HDI
0.181 0.271
omegaO
2nd Base
mode = 0.248
95% HDI
0.238 0.259
omega[2]
Center Field
mode = 0.256
95% HDI
0.246 0.268
omega[5]
Right Field
mode = 0.259
95% HDI
0.25 0.268
omega[8]
3rd Base
mode = 0.255
95% HDI
0.245 0.265
omega[3]
Left Field
mode = 0.249
95% HDI
0.239 0.258
omega[6]
Shortstop
mode = 0.247
95% HDI
0.235 0.257
omega[9]
0.10 0.15 0.20 0.25

Figure 9.18



Gamma Distrib's (SD=10)
dgamma(k)
0.00
0.04
0.08
Mode 1
Mean 1
0
50
100
150
200
k
Gamma Distrib's (SD=10), zoomed in
dgamma(k)
0.070
0.074
0.078
Mode 1
Mean 1
0
50
100
150
200
k
Gamma Distrib's (SD=10), zoomed in
dgamma(k)
0e+00
4e-06
8e-06
Mode 1
Mean 1
0
50
100
150
200
k

Figure 9.19