Hierarchical Bayes

Peter Lenk
Stephen M Ross School of Business at the University of Michigan
September 2004
Outline

• Bayesian Decision Theory
• Simple Bayes and Shrinkage Estimates
• Hierarchical Bayes
• Numerical Methods
• Batting Averages
• HB Interaction Model
Bayesian Decision Model

**Decision Model**
- Actions
- States
- Consequences

**Inference Model**
- Prior
- Likelihood
- Posterior

**Updating**
- Bayes Theorem

**Integration**

**Result**
- Optimal Decision: Marketing Action & Inference
Bayes Theorem

• Model for the data given parameters
  – $f(y \mid \theta)$ were $\theta = \text{unknown parameters}$
  – E.g. $Y_i = \mu + \varepsilon_i$ and $\theta = (\mu, \sigma)$
  – Likelihood $l(\theta) = f(y_1 \mid \theta) f(y_2 \mid \theta) \ldots f(y_n \mid \theta)$

• Prior distribution of parameters $p(\theta)$

• Update prior
  – $p(\theta \mid \text{Data}) = l(\theta)p(\theta)/f(y)$
  – $f(y) = \text{marginal distribution of data}$
Easy Example:

• Estimate mean from a normal distribution.

• \( Y_i = \mu + \varepsilon_i \)

• Error terms \( \{\varepsilon_i\} \) are iid normal
  – Mean is zero
  – Standard deviation of error terms is \( \sigma \).
  – Assume that \( \sigma \) is known
Conjugate Prior for Mean

- Prior distribution for $\mu$ is normal
  - Prior mean is $m_0$
  - Prior variance is $v_0^2$
  - Precision is $1/v_0^2$
Posterior Distribution

• Observe $n$ data points
• Posterior distribution is normal
  – Mean is $m_n$
  – Variance is $v_n^2$

\[ m_n = w\bar{y} + (1 - w)m_0 \]
\[ w = \frac{n}{\sigma^2} \]
\[ \frac{n}{\sigma^2} + \frac{1}{v_0^2} \]
\[ v_n^2 = \frac{1}{\frac{n}{\sigma^2} + \frac{1}{v_0^2}} \]
Shrinkage Estimators

• Bayes estimators combines prior guesses with sample estimates.
• If the prior precision is larger than sample precision (prior has more information), then put more weight on prior mean.
• If the sample precision is larger the prior precision (sample has more information), then put more weight on sample average.
Example

- Y is normal with mean 10 and Variance 16
- Normal prior for the population mean
  - Mean = 5 & Variance = 2
  - Prior is informative and way off
- Data
  - n = 5, Average = 10.9, Variance = 14.7
- Posterior is normal
  - Mean = 7.4 and variance is 1.2
Prior & Posterior n=5

- Prior
- Posterior
- Likelihood
Prior & Posterior n=50

Prior
Posterior
Likelihood
Use Less Informative Prior

• Y is normal with mean 10 and Variance 16
• Normal prior for the population mean
  – Mean = 5 & Variance = 10 instead of 2
  – Prior is “flatter”
• Data
  – n = 5, Average = 10.9, Variance = 14.7
• Posterior is normal
  – Mean = 9.6 and variance is 2.3
Prior & Posterior n=5

Prior  Posterior  Likelihood

Mean

0  5  10  15
Prior & Posterior n=50

Prior  Posterior  Likelihood
Summary

- Prior has less effect as sample size increases
- Very informative priors give good results with smaller samples if prior information is correct
- If you really don’t know, then use “flatter” or less informative priors
What about Marketing?

- HB revolution in how we think about customers
Henry Ford

All Customers are the Same
Alfred Sloan
Several Common Preferences
Continuous Heterogeneity
Profit Maximization
It Can Get Wild!
HB Model for Weekly Spending

- Within-subject model:
  \[ Y_{i,j} = \mu_i + \epsilon_{i,j} \text{ and } \text{var}(\epsilon_{i,j}) = \sigma_i^2 \]

- Heterogeneity in mean weekly spending or between-subjects
  \[ \mu_i = \theta + \delta_i \text{ and } \text{var}(\delta_i) = \tau^2 \]

- Prior Distribution
  \[ \theta \text{ is } N(u_0,v_0^2) \]

- Variances are known
Variances & Covariances

- \( \text{Var}(Y_{i,j} | \mu_i) = \sigma_i^2 \) (known \( \mu_i \))
- \( \text{Var}(Y_{i,j} | \theta) = \tau^2 + \sigma_i^2 \) (unknown \( \mu_i \))
- \( \text{Cov}(Y_{i,j}, Y_{i,k}) = \tau^2 \) for \( j \) not equal to \( k \)
- Observations from different subjects are independent
Precisions = 1/Variance

\[ \text{Pr}(\theta) = \frac{1}{\nu_0^2} \text{ is prior precision} \]

\[ \text{Pr}(\mu_i \mid \theta) = \frac{1}{\tau^2} \]

\[ \text{Pr}(Y_{i,j} \mid \mu_i) = \frac{1}{\sigma_i^2} \text{ and } \text{Pr}(Y_{i,j} \mid \theta) = \frac{1}{\tau^2 + \sigma_i^2} \]

\[ \text{Pr}(\bar{Y}_i \mid \mu_i) = \frac{n}{\sigma_i^2} \text{ and } \text{Pr}(\bar{Y}_i \mid \theta) = \frac{1}{\tau^2 + \frac{\sigma_i^2}{n}} \]
Joint Distribution

\[ P(Y, \mu, \theta) = \]

\[ h(\theta \mid u_0, v_0^2) \prod_{i=1}^{N} g(\mu_i \mid \theta, \tau^2) \prod_{j=1}^{n_i} f(y_{i,j} \mid \mu_i, \sigma_i^2) \]

Prior \hspace{2cm} Between Subjects \hspace{2cm} Within Subjects
Bayes Theorem

\[ P(\mu, \theta | Y) = \frac{P(Y, \mu, \theta)}{\int \int P(Y, \mu, \theta) d\mu d\theta} \]

\[ P(\mu, \theta | Y) = \frac{P(Y, \mu, \theta)}{P(Y)} \]

\[ P(\mu, \theta | Y) \propto P(Y, \mu, \theta) \]

Constant because \( Y \) is fixed & known
Bayes Estimator

• Posterior means are optimal under squared error loss
  \( E(\mu_i|Y) \) and \( E(\theta|Y) \)

• Measure of accuracy is posterior variance
  \( \text{var}(\mu_i|Y) \) and \( \text{var}(\theta|Y) \)
Posterior Distribution of $\theta$

- Normal distribution
- Posterior mean is $u_N$
- Posterior variance is $v_N^2$
- Posterior precision is $Pr(\theta|Y) = 1/v_N^2$
Posterior Precision of $\theta$

“Pr” = Precision = $1$/Variance

$$
Pr(\theta \mid Y) = Pr(\theta) + \sum_{i=1}^{N} Pr(\overline{Y_i} \mid \theta)
$$

$$
Pr(\theta) = \frac{1}{\nu_0^2} \text{ and } Pr(\overline{Y_i} \mid \theta) = \frac{1}{\tau^2 + \frac{\sigma_i^2}{n_i}}
$$
Posterior Mean of $\theta$

\[ u_N = w_0 u_0 + \sum_{i=1}^{N} w_i \bar{Y}_i \]

\[ w_0 = \frac{\Pr(\theta)}{\Pr(\theta | Y)} \quad \text{and} \quad w_i = \frac{\Pr(\bar{Y}_i | \theta)}{\Pr(\theta | Y)} \]
Updating of $\theta$

- Prior Mean $\Rightarrow$ Posterior Mean
  
  $u_0 \Rightarrow u_N$

- Prior Var $\Rightarrow$ Posterior Var
  
  $v_0^2 \Rightarrow v_N^2$
Posterior Mean of $\mu_i$

\[
E[\mu_i \mid Y] = \alpha_i \bar{Y}_i + (1 - \alpha_i) \mu_N
\]

\[
\alpha_i = \frac{\Pr(\bar{Y}_i \mid \mu_i)}{\Pr(\mu_i \mid \theta) + \Pr(\bar{Y}_i \mid \mu_i)} = \frac{n_i}{\sigma_i^2} = \frac{1}{\tau^2} + \frac{n_i}{\sigma_i^2}
\]
Between-Subject Heterogeneity in Mean Household Spending
Between & Within Subjects Distributions

Heterogeneity  Subject 1  Subject 2  Subject 3

Spending Distribution

Distribution

0.4
0.3
0.2
0.1
0.0

0 10 20 30 40 50 60

Subject 1  Subject 2  Subject 3

Heterogeneity
2 Observations per Subject
Subject Averages

- Heterogeneity
- Subject 1
- Subject 2
- Subject 3
Pooled Estimate of Mean

Pooled estimate of population average spending
Sample Estimates

• Disaggregate estimate $\bar{Y}_i$ of $\mu_i$ only uses the observations for subject i.
  – Super if 30 or more observations per subject

• Pooled or aggregate estimator $\bar{Y}$ of $\theta$
  – Smaller sampling error
  – Ignores individual difference
HB Shrinkage Estimator

- Take combination of individual average and pooled average

\[ w_i \bar{Y}_i + (1 - w_i) \bar{Y} \]

- What are the correct weights?
- HB automatically gives optimal weights based on
  - Prior variance of \( \mu_i \)
  - Number of observations for subject \( i \)
  - Variance of past spending for subject \( i \)
  - Number of subjects
  - Amount of heterogeneity in household means
Shrinkage Estimates

Spending Distribution

- Heterogeneity
- Subject 1
- Subject 2
- Subject 3
20 Observations per Subject

Distribution

Spending

Heterogeneity
Subject 1
Subject 2
Subject 3
Bayes & Shrinkage Estimates

- Bayes estimators automatically determine the optimal amount of shrinkage to minimize MSE for true parameters and predictions
- Borrows strength from all subjects
- Tradeoff some bias for variance reduction
Good & Bad News

• Only simple models result in equations
• Models we use in marketing require numerical methods to compute posterior mean, posterior standard deviations, predictions and so on.
Numerical Integration

• Compute posterior mean of function $T(\theta)$.

$$E[T(\theta) \mid y] = \int T(\theta) p(\theta \mid y) d\theta$$
Trapezoid Rule

\[ T(x)f(x) \]
Grid Methods

• Very accurate with few functional evaluations
• Need to know where the action is
• Does not scale well to higher dimensions
• You need to be very smart to make it work
Monte Carlo

• Generate random draws $\theta_1, \theta_2, \ldots, \theta_m$ from posterior distribution using a random number generator.

$$E[T(\theta)|y] \approx \frac{1}{m} \sum_{j=1}^{m} T(\theta_j)$$

• What happened to the density of $\theta$?
Good & Bad News

• If your computer has a random number generator for the posterior distribution, Monte Carlo is a snap to do.
• Your computer almost never has the correct random number generator.
Importance Sampling

• Would like to sample from density $f$
• You have a good random number generator for the density $g$
• Importance sampling lets you generate random deviates from $g$ to evaluate expectations with respect to $f$.
• Generate $\phi_1, \ldots, \phi_m$ from $g$
Importance Sampling Approximation

\[ \int T(\theta)f(\theta)d\theta = \int T(\varphi) \frac{f(\varphi)}{g(\varphi)} g(\varphi)d\varphi \]

\[ \approx \frac{\sum_{i=1}^{m} T(\varphi_i)r(\varphi_i)}{\sum_{i=1}^{m} r(\varphi_i)} = \sum_{i=1}^{m} T(\varphi_i)w(\varphi_i) \]

\[ r(\varphi_i) \propto \frac{f(\varphi_i)}{g(\varphi_i)} \quad \text{and} \quad w(\varphi_i) = \frac{r(\varphi_i)}{\sum_{j=1}^{m} r(\varphi_j)} \]
Markov Chain Monte Carlo

- Extension of Monte Carlo
- Random draws are not independent
- Joint distribution $f(\beta, \phi)$ does not have a convenient random number generator.
- Conditional distributions $g(\phi|\beta)$ and $h(\beta|\phi)$ are easy to generate from.
Iterative Generation from Full Conditionals

• Start at $\phi_0$
• Generate $\beta_1$ from $h(\beta|\phi_0)$.
• Generate $\phi_1$ from $g(\phi|\beta_1)$.
• ...
• Generate $\beta_{m+1}$ from $h(\beta|\phi_m)$
• Generate $\phi_{m+1}$ from $g(\phi|\beta_{m+1})$
Baseball Example

• 90 MLB Players in 2000 season.
• Observe at bats (AB) and hits (BA) in May
• Infer distribution of batting averages across players.
• Predict batting averages in October using data from May.
Baseball Batting Averages

<table>
<thead>
<tr>
<th></th>
<th>The Cleveland Indians - 1995</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>May</td>
</tr>
<tr>
<td></td>
<td>BA</td>
</tr>
<tr>
<td>Murray</td>
<td>.442</td>
</tr>
<tr>
<td>Belle</td>
<td>.400</td>
</tr>
<tr>
<td>Vizquel</td>
<td>.204</td>
</tr>
<tr>
<td>Pena</td>
<td>.148</td>
</tr>
</tbody>
</table>
Estimating a Probability

- $n$ at bats in May
- $X =$ number of hits
- $p =$ batting average for season
- $X$ has a binomial distribution
  - mean $np$
  - variance $np(1-p)$
Binomial Distribution

\[ n = 50 \quad p = 0.3 \quad \text{mean} = 15.00 \quad \text{STD} = 3.24 \]
Need Prior for Batting Average $p$

- $0 < p < 1$
- Beta distribution is popular choice
- It has two parameters: $\alpha$ and $\beta$
- Density is proportional to $p^{\alpha-1}(1-p)^{\beta-1}$
- Prior Mean = $\alpha/(\alpha+\beta)$
Beta Prior for $p$

$$f[p \mid \alpha, \beta] = \text{Beta}(p \mid \alpha, \beta)$$

$$= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1 - p)^{\beta-1} \quad \text{for} \quad 0 \leq p \leq 1$$

$$\Gamma(\alpha) = \int_{0}^{\infty} x^{\alpha-1} e^{-x} \, dx$$
Mean and Variance

\[ E(p) = \frac{\alpha}{\alpha + \beta} \]

\[ V(p) = \frac{E(p)[1 - E(p)]}{\alpha + \beta + 1} \]
Beta Distribution

alpha = 1
beta = 1
mean = 0.50
Beta Distribution

alpha = 5
beta = 1
mean = 0.83
Beta Distribution

$\alpha = 0.5$

$\beta = 0.5$

Mean = 0.50
Bayes Theorem:
Update prior for $p$ after observing $n$ and $x$

$$f [ p \mid x, \alpha, \beta ] = \frac{\Pr[ x \mid p ] f [ p \mid \alpha, \beta ]}{\int_0^1 \Pr[ x \mid q ] f [ q \mid \alpha, \beta ]dq}$$

$$\propto \Pr[ x \mid p ] f [ p \mid \alpha, \beta ]$$

$$\propto p^{\alpha + x - 1} (1 - p)^{\beta + n - x - 1}$$

$$= Beta (p \mid \alpha + x, \beta + n - x)$$
Inference About P: Posterior is also Beta

<table>
<thead>
<tr>
<th>Prior Parameters</th>
<th>Posterior Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>$\alpha + x$</td>
</tr>
<tr>
<td>$\beta$</td>
<td>$\beta + n - x$</td>
</tr>
</tbody>
</table>
Posterior Mean of $p$: It's another shrinkage estimator

$$w_n \hat{p}_n + (1-w_n)(\text{Prior Mean})$$

$$\hat{p}_n = \frac{x}{n} \quad \text{and} \quad w_n = \frac{n}{\alpha + \beta + n}$$
prior mean = 0.25

\[ p = \text{Beta} & \ x = \text{Binomial} \]

\[ \alpha = 5 \quad \beta = 15 \]

\[ N = 50 & \ X = 18 \]

Density

P

Prior

Posterior
Hierarchical Bayes Model

• Variation within batter $i$:
  – $X_i$ given $p_i$ has a binomial distribution

• Variation among batters:
  – $p_i$ is a beta distribution with parameters $\alpha$ and $\beta$.

• Prior distribution for $\alpha$ and $\beta$
  – Gamma (chi-square) distribution
Gamma Distribution

\[ g(y) = G(y \mid r, s) \]

\[
= \frac{s^r}{\Gamma(r)} y^{r-1} e^{-sy} \quad \text{for} \quad y > 0
\]

\[ E(Y) = \frac{r}{s} \quad \text{and} \quad V(Y) = E(Y) \frac{1}{s} \]
Gamma Distribution

alpha = 1  beta = 0.1
mean = 10.00
STD = 10.00
Gamma Distribution

\[
\begin{align*}
\text{Density} & \quad 0.07 \\
0 & \quad 0.01 \\
10 & \quad 0.02 \\
20 & \quad 0.03 \\
30 & \quad 0.04 \\
40 & \quad 0.05 \\
50 & \quad 0.06 \\
\end{align*}
\]

\[\alpha = 10 \quad \beta = 0.5\]

mean = 20.00
STD = 6.32
Specify Prior Parameters: 
\( r, s, u \) \& \( v \)

- Priors: \( \alpha \) is \( G(r,s) \) \& \( \beta \) is \( G(u,v) \).
- \( E(\alpha) = r/s \) and \( V(\alpha) = E(\alpha)/s \).
- \( s \) determines variance relative to mean.
- I used \( s = 0.25 \) or the variance is four times larger than the mean.
- Same for \( v \).
\[ E(p) = E[E(p | \alpha, \beta)] \]
\[ = \frac{r}{r + u} \]
\[ = p_0 \]

\[ c = V[E(p | \alpha, \beta)] = \frac{p_0(1-p_0)}{r+u+1} \]

\[ r = p_0 \left( \frac{p_0(1-p_0)}{c} - 1 \right) \]

\[ u = (1-p_0) \left( \frac{p_0(1-p_0)}{c} - 1 \right) \]
Specify Prior Parameters

• Guess a mean of all batting averages: $p_0 = 0.25$
• Measure of my uncertainty of that guess: $c = 0.01$
• Parameter $r = 4.4$
• Parameter $u = 13.3$
MCMC for Batting Averages

• Need full conditionals for $p_i$ give $\alpha$ and $\beta$
  – Beta distribution

• Need full conditionals for $\alpha$ and $\beta$ given $p_i$
  – Unknown distribution
  – Use Metropolis algorithm
MCMC: Full Conditionals for Player i Batting Average $p_i$

$$f[p_i \mid x_i, \alpha, \beta] \propto \Pr(x_i \mid p_i) f(p_i \mid \alpha, \beta)$$

$$\propto p_i^{a+x_i-1}(1-p_i)^{\beta+n_i-x_i-1}$$

$$= Beta(p_i \mid \alpha + x_i, \beta + n_i - x_i)$$
MCMC: Full Conditional for $\alpha$ and $\beta$

$$g(\alpha, \beta \mid x_1, \ldots x_n, p_1, \ldots p_n)$$

$$\propto \prod_{i=1}^{n} p_i^{\alpha-1}(1-p_i)^{\beta-1} g(\alpha \mid r, s) g(\beta \mid u, v)$$
Metropolis Algorithm

• Want to generate $\theta$ from $f$
• Instead, generate candidate value $\phi$ from $g(\cdot|\theta)$
  – Density $g$ can depend on $\theta$
  – eg Random walk: $\phi = \theta + \delta$
• With probability $\alpha(\theta,\phi)$ accept $\phi$ as the new value of $\theta$
• With probability $1 - \alpha(\theta,\phi)$ keep $\theta$
Transition Probability

\[ \alpha(\theta, \varphi) = \max \left\{ \frac{f(\varphi)g(\theta | \varphi)}{f(\theta)g(\varphi | \theta)}, 1 \right\} \]

- f is the full conditional density of \( \theta \)
- Ratios: do not need to know constants
- Usually compute \( \alpha \) on log scale.
- Works if densities are not zero
- Works better if g is close to f
Alpha and Beta vs Iteration

Parameter

MCMC Iteration

BETA

ALPHA
Posterior of Alpha
## Parameters Estimates

<table>
<thead>
<tr>
<th></th>
<th>Prior</th>
<th>Posterior</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>17.8</td>
<td>26.2</td>
</tr>
<tr>
<td>(std)</td>
<td>(8.4)</td>
<td>(4.6)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>53.2</td>
<td>68.2</td>
</tr>
<tr>
<td>(std)</td>
<td>(14.6)</td>
<td>(11.7)</td>
</tr>
</tbody>
</table>
Prediction of Season Averages

<table>
<thead>
<tr>
<th></th>
<th>RMSE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>0.060</td>
<td>17.0%</td>
</tr>
<tr>
<td>Bayes</td>
<td>0.032</td>
<td>9.4%</td>
</tr>
</tbody>
</table>
Batting Averages
Bayes Shrinks MLE
HB Conjoint
Lenk, DeSarbo, Green, Young (1996)

• Evaluated computer profiles on a 0 to 10 scale for “likelihood to purchase”
  – 0 = Would not buy
  – 10 = Would definitely buy

• Design
  – 178 subjects
  – 13 attributes with two levels each
  – 20 profiles per subject
Attributes: Effect Coding

1. Hotline support
2. RAM
3. Screen Size
4. CPU
5. Hard Disk
6. Multimedia
7. Cache
8. Color
9. Retail Store
10. Warrantee
11. Software
12. Guarantee
13. Price
Subject-Covariates

- Female: 1 if female and 0 if male
- Years: # years of work experience
- Own: 1 if has computer & 0 else
- Nerd: 1 if technical background & 0 else
- Apply: # of software applications
- Expert: self-report expertise rating
<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std Dev</th>
<th>MIN</th>
<th>MAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEMALE</td>
<td>0.275</td>
<td>0.448</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>YEARS</td>
<td>4.416</td>
<td>2.369</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>OWN</td>
<td>0.876</td>
<td>0.330</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>NERD</td>
<td>0.275</td>
<td>0.448</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>APPLY</td>
<td>4.287</td>
<td>1.574</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>EXPERT</td>
<td>7.618</td>
<td>1.902</td>
<td>3</td>
<td>10</td>
</tr>
</tbody>
</table>
Interaction Model

• Within-Subjects

\[ Y_i = X_i \beta_i + \varepsilon_i \text{ for } i = 1, \ldots, n \]

\[ [\varepsilon_i] = N_m(\varepsilon_i \mid 0, \sigma^2 I_m) \]

• Between-Subjects Heterogeneity

\[ \beta_i = \Theta' z_i + \delta_i \]

\[ [\delta_i] = N_p(\delta_i \mid 0, \Delta) \]
Average over Posterior Means and Std Dev of Partworths Across Subjects

<table>
<thead>
<tr>
<th></th>
<th>PostMean</th>
<th>PostSTD</th>
<th></th>
<th>PostMean</th>
<th>PostSTD</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNST</td>
<td>4.757</td>
<td>1.404</td>
<td>Cache</td>
<td>0.031</td>
<td>0.461</td>
</tr>
<tr>
<td>HotLine</td>
<td>0.095</td>
<td>0.487</td>
<td>Color</td>
<td>0.026</td>
<td>0.371</td>
</tr>
<tr>
<td>RAM</td>
<td>0.347</td>
<td>0.467</td>
<td>Dstrbtn</td>
<td>0.078</td>
<td>0.378</td>
</tr>
<tr>
<td>ScrnSz</td>
<td>0.193</td>
<td>0.405</td>
<td>Wrrnty</td>
<td>0.124</td>
<td>0.392</td>
</tr>
<tr>
<td>CPU</td>
<td>0.392</td>
<td>0.646</td>
<td>Sftwr</td>
<td>0.196</td>
<td>0.399</td>
</tr>
<tr>
<td>HrdDsk</td>
<td>0.171</td>
<td>0.501</td>
<td>Grnttee</td>
<td>0.112</td>
<td>0.427</td>
</tr>
<tr>
<td>MultMd</td>
<td>0.494</td>
<td>0.574</td>
<td>Price</td>
<td>-1.127</td>
<td>0.873</td>
</tr>
</tbody>
</table>
Impact of Covariates on Partworths

<table>
<thead>
<tr>
<th></th>
<th>CNST</th>
<th>RAM</th>
<th>CPU</th>
<th>Dstrbtn</th>
<th>Wrrnty</th>
<th>Grntee</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNST</td>
<td>3.74</td>
<td>0.52</td>
<td>-0.15</td>
<td>0.05</td>
<td>-0.01</td>
<td>0.03</td>
<td>-1.55</td>
</tr>
<tr>
<td>FEMALE</td>
<td>-0.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.40</td>
</tr>
<tr>
<td>YEARS</td>
<td>-0.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OWN</td>
<td>-0.10</td>
<td>0.17</td>
<td>0.17</td>
<td>0.20</td>
<td></td>
<td>-0.12</td>
<td>-0.20</td>
</tr>
<tr>
<td>NERD</td>
<td>-0.27</td>
<td>0.15</td>
<td>0.16</td>
<td></td>
<td>-0.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>APPLY</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXPERT</td>
<td>0.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Summary

• HB allows individual-level coefficients
• Two level model
  – With-in subjects
  – Between subjects (heterogeneity)
• HB shrinks unstable, subject-level estimators to population mean
Summary

• BDT provides integrated framework for making decisions and inference
• Good models consider all sources of uncertainty
• Good methods keep track of all sources of uncertainty
• Bayes does both