Bayesian Inference via the Blended Paradigm

John Lewis, Yoonkyung Lee, Steven MacEachern

The Ohio State University With Support from:
The Nationwide Center for Advanced Customer Insights
NSF

Tuesday, August 5, 2013

Introduction to the Blended Paradigm

- Traditional Bayesian Framework:
  - $Y \sim [Y|\theta]$ and $\theta \sim [\theta]$
  - Given the observed $Y = y$, pass to the posterior:

$$[\theta|y] \propto [y|\theta][\theta]$$

- Deficiencies:
  - We may not fully believe the likelihood $[Y|\theta]$
  - More modeling may not fix this problem
- Acknowledge inadequacy of our model
- Drive Bayesian update with the 'good' portion of the likelihood

Introduction to the Blended Paradigm

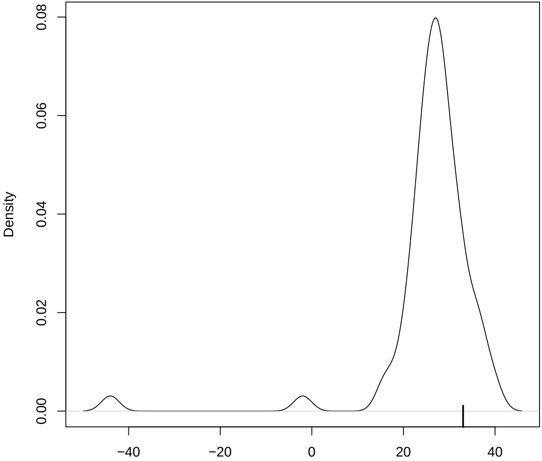
- Blended Paradigm Framework:
  - Here, $T(Y)$ is a 'robust' summary of the data
    - 'Robust' means insensitivity to the deficiencies in the model for inferences of interest
    - Model for $[Y|\theta]$ implies model for $[T(Y)|\theta]$
  - $T(Y) \sim [T(Y)|\theta]$ and $\theta \sim [\theta]$
  - Given the observed $T(y) = t$, pass to the posterior:

$$[\theta|t] \propto [t|\theta][\theta]$$

  - Implementation via MCMC
- Goal : To obtain better inference through a wise choice of $T(Y)$

Proof of Concept-Outliers

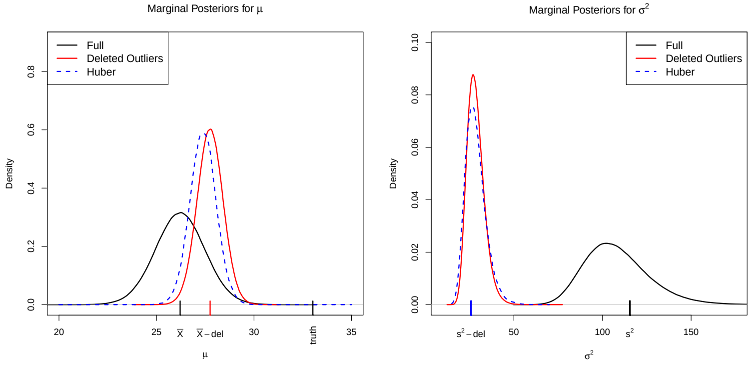
- Simon Newcomb's famous measurements of the speed of light



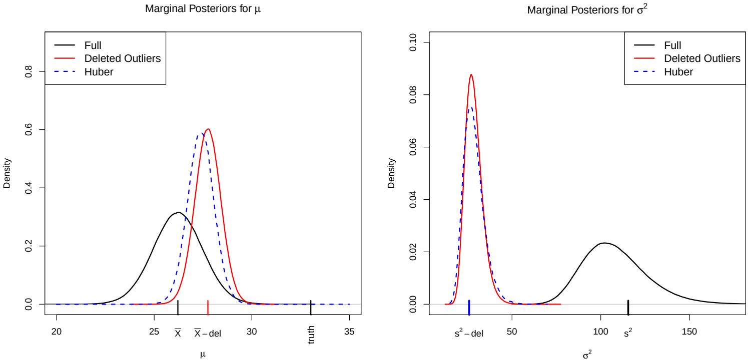
Model

- Model for the data: $Y_i \overset{\text{iid}}{\sim} N(\mu, \sigma^2), i = 1, \cdots, 66$
  - Full Likelihood
  - Restricted Likelihood $T(Y) = (\hat{\mu}, \hat{\sigma})$
  - $(\hat{\mu}, \hat{\sigma})$ robust M-estimators using Huber's loss and 'Proposal 2'
- Fixed priors: $\mu \sim N(\eta, \tau^2)$, $\sigma^2 \sim IG(\alpha, \beta)$
- $\eta = -222$, $\tau = 540$, $\alpha = 5$, $\beta = 200$
- Priors derived from experimental data on the speed of light conducted before Newcomb's study

Posteriors


Marginal Posteriors for μ
Full
Deleted Outliers
Huber
Density
0.0
0.2
0.4
0.6
0.8
20
25
X̄
X̄ – del
30
truth
35
μ
Marginal Posteriors for σ²
Full
Deleted Outliers
Huber
Density
0.00
0.02
0.04
0.06
0.08
0.10
s² – del
50
100
s²
150
σ²

Posteriors



- Overall conclusion: Reduced bias and variance

Model Fitting

- $[T(Y)|\theta]$- typically intractable
- Small dimensions: grid estimation
- Larger dimensions: data augmented MCMC
  - $[\theta, Y | T(y)]$
    - $[\theta|Y, T(y)] = [\theta|Y]$
    - $[Y|\theta, T(y)]$
- First step: full posterior
- Second step: Metropolis-Hastings

MH Step

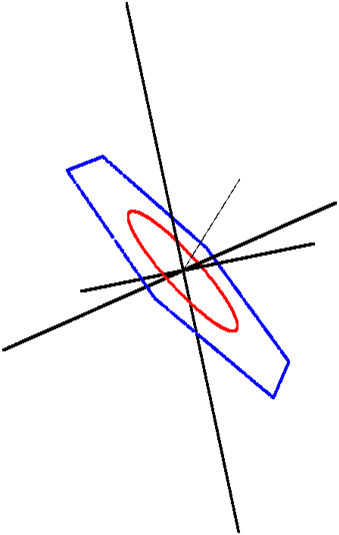
- MH ratio

$$\frac{[y^{**}|\theta, t]}{[y_{curr}|\theta, t]} \frac{p(y_{curr}|y^{**})}{p(y^{**}|y_{curr})}$$

- Ratio of Likelihoods: normalizing constants cancel
- Proposal density:
  - Start $Y^* \sim g$
  - Transform $H : Y^* \rightarrow Y^{**}$, so that $T(Y^{**}) = T(y)$
  - Use this recipe to derive the proposal density (not a standard transformation)

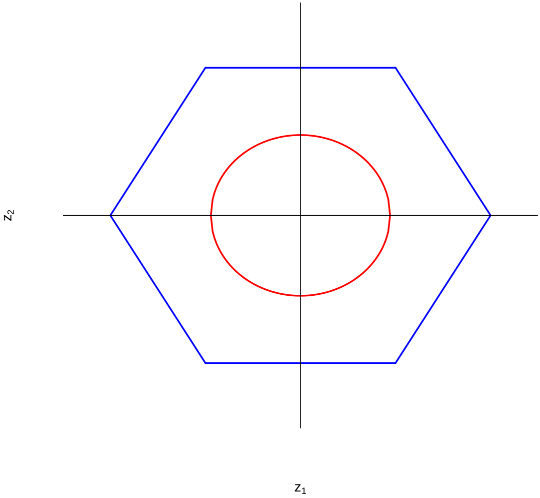
Example 1- $T(y) : \bar{y} = 0, \sum |y_i - \bar{y}| = 1$

$n = 3$: sample space lies on the plane defined by $\bar{y} = 0$



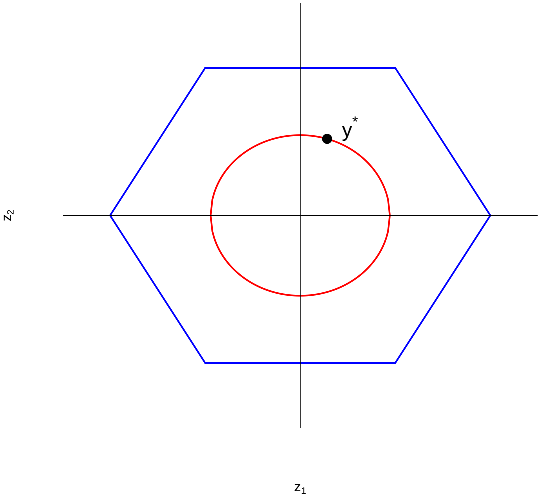
Example 1- $T(y) : \bar{y} = 0, \sum |y_i - \bar{y}| = 1$

Two dimensional view of the plane



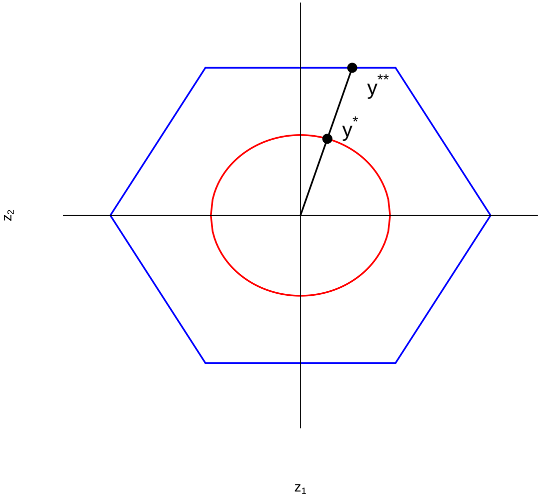
Example 1- $T(y) : \bar{y} = 0, \sum |y_i - \bar{y}| = 1$

Two dimensional view of the plane



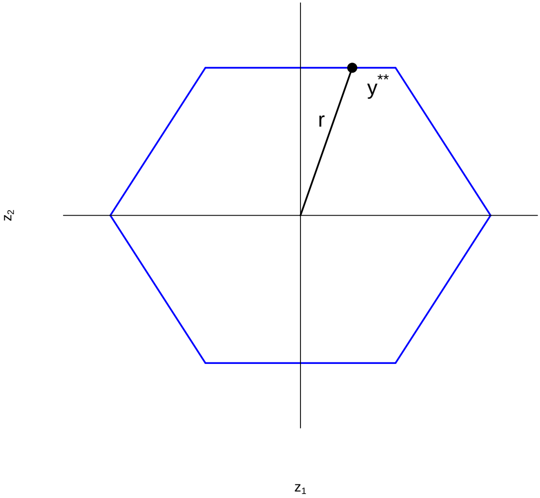
Example 1- $T(y) : \bar{y} = 0, \sum |y_i - \bar{y}| = 1$

Two dimensional view of the plane



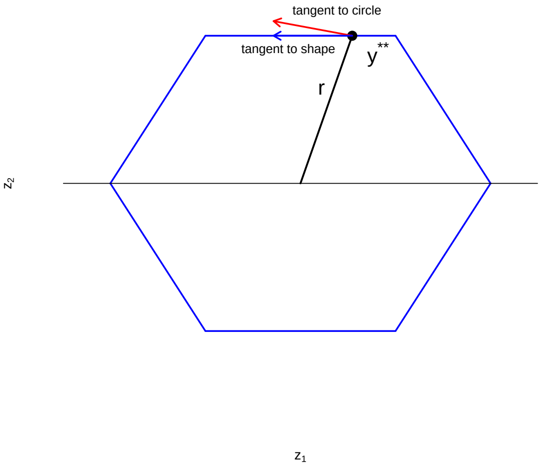
Example 1- $T(y) : \bar{y} = 0, \sum |y_i - \bar{y}| = 1$

Two dimensional view of the plane



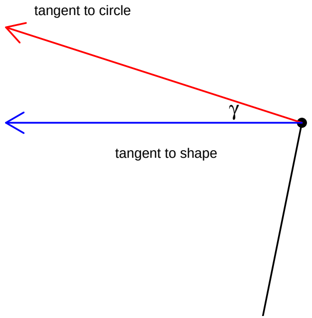
Example 1- $T(y) : \bar{y} = 0, \sum |y_i - \bar{y}| = 1$

Two dimensional view of the plane



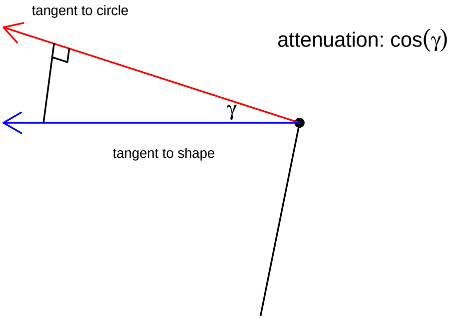
Example 1- $T(y) : \bar{y} = 0, \sum |y_i - \bar{y}| = 1$

Two dimensional view of the plane



Example 1- $T(y) : \bar{y} = 0, \sum |y_i - \bar{y}| = 1$

Two dimensional view of the plane

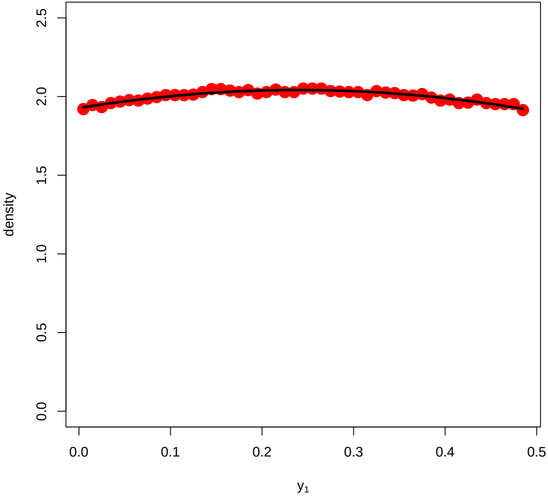


Example 1- $T(y) : \bar{y} = 0, \sum |y_i - \bar{y}| = 1$

- Proposal density
  - Starting distribution g: uniform on the circle
  - Adjustment 1: Stretch out to the target shape
    - r^{-1}
  - Adjustment 2: Attenuation to get density along the target shape
    - $\cos(\gamma)$
- $p(y^{**}) \propto \frac{\cos(\gamma)}{r}$

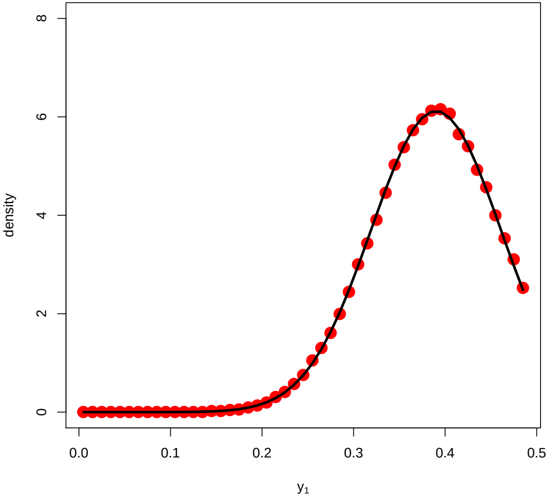
Example 1- $T(y) : \bar{y} = 0, \sum |y_i - \bar{y}| = 1$

- MCMC for $[Y_1|\theta, T(y), Y_1 > 0], \theta = (\mu, \sigma) = (0, 1)$



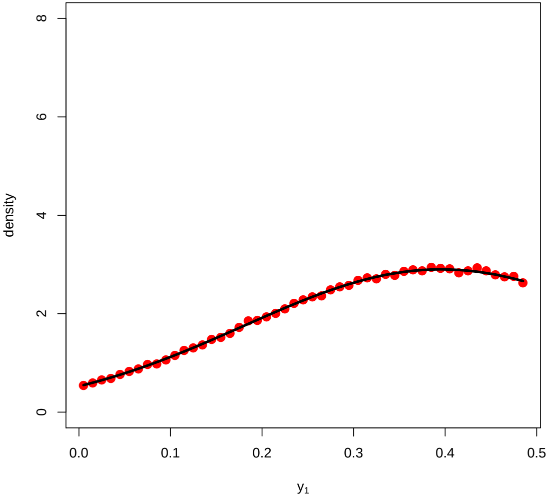
Example 1- $T(y) : \bar{y} = 0, \sum |y_i - \bar{y}| = 1$

- MCMC for $[Y_1|\theta, T(y), Y_1 > 0], \mu = (.4, -.5, .1), \sigma = 0.1$



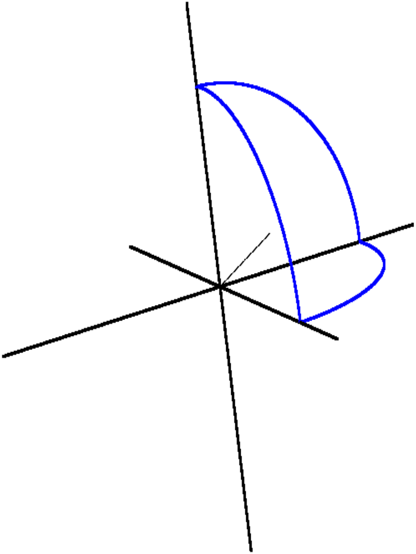
Example 1- $T(y) : \bar{y} = 0, \sum |y_i - \bar{y}| = 1$

- MCMC for $[Y_1|\theta, T(y), Y_1 > 0], \mu = (.4, -.5, .1), \sigma = 0.3$



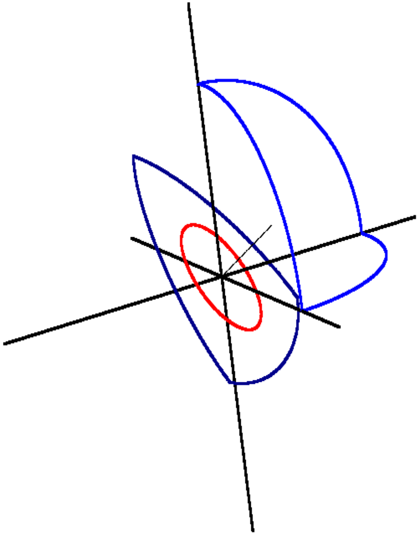
Example 2- $T(y) : \min y = 0, \sum(y_i - \min y)^2 = 1$

$n = 3$: sample space does not lie in a plane



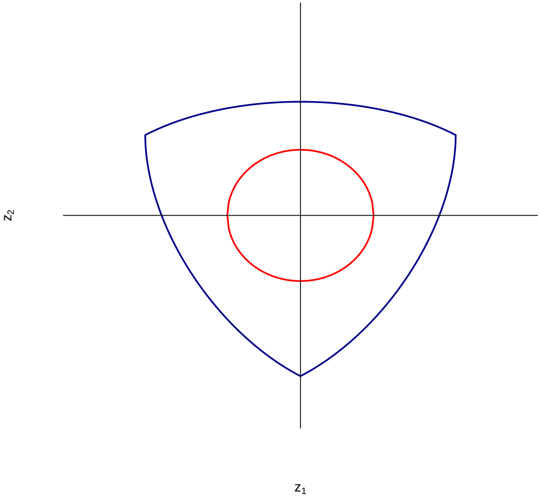
Example 2- $T(y)$: $\min y = 0, \sum (y_i - \min y)^2 = 1$

$n = 3$: sample space does not lie in a plane



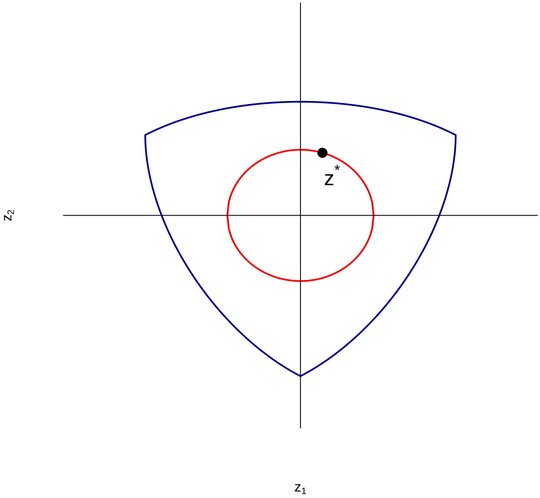
Example 2- $T(y) : \min y = 0, \sum (y_i - \min y)^2 = 1$

Two dimensional view of the plane



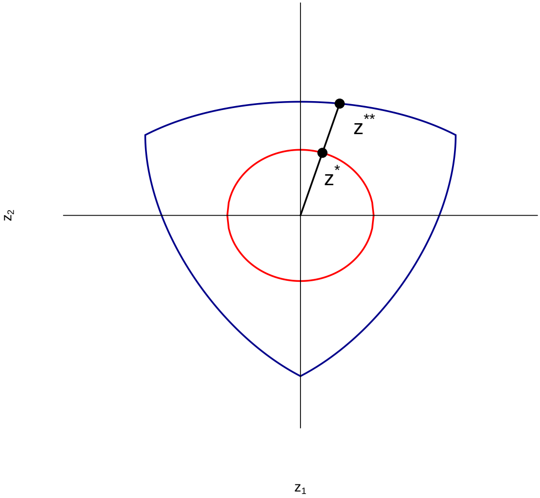
Example 2- $T(y)$: $\min y = 0, \sum (y_i - \min y)^2 = 1$

Two dimensional view of the plane



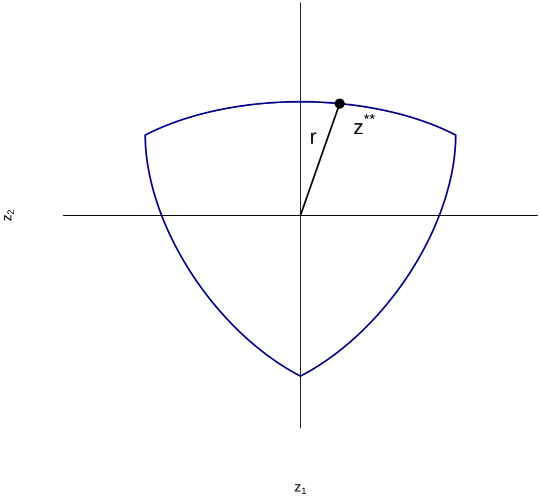
Example 2- $T(y) : \min y = 0, \sum (y_i - \min y)^2 = 1$

Two dimensional view of the plane



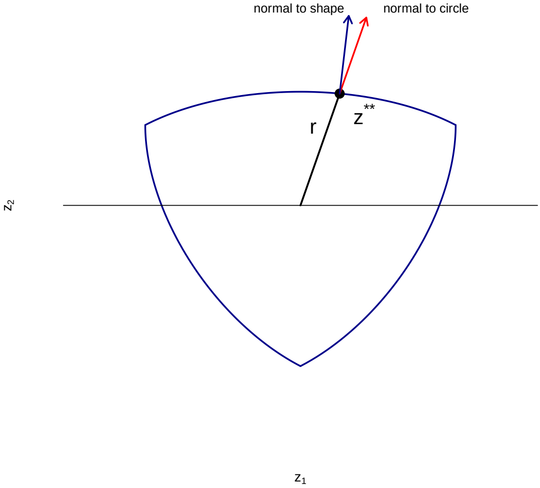
Example 2- $T(y)$: $\min y = 0, \sum(y_i - \min y)^2 = 1$

Two dimensional view of the plane



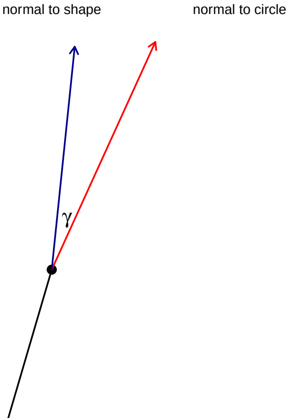
Example 2- $T(y) : \min y = 0, \sum(y_i - \min y)^2 = 1$

Two dimensional view of the plane



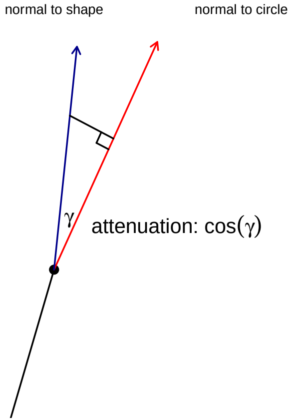
Example 2- $T(y)$: $\min y = 0, \sum (y_i - \min y)^2 = 1$

Two dimensional view of the plane



Example 2- $T(y)$: $\min y = 0, \sum (y_i - \min y)^2 = 1$

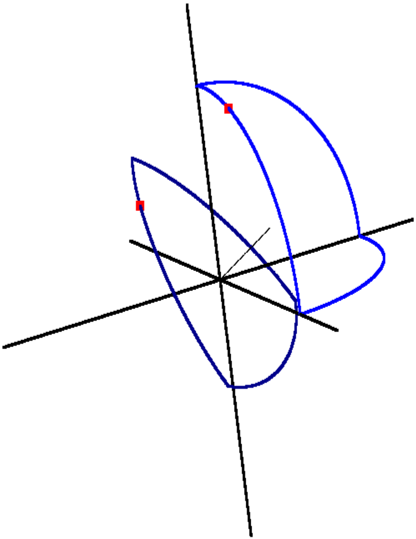
Two dimensional view of the plane



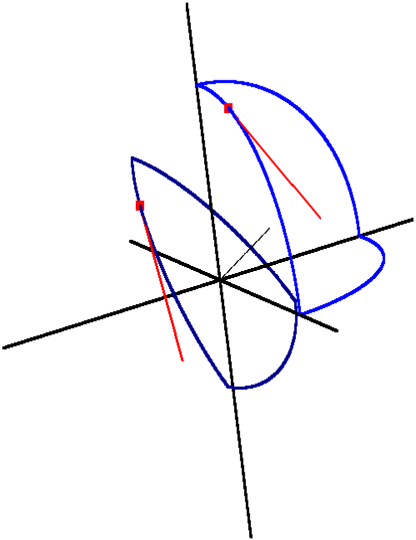
Example 2- $T(y) : \min y = 0, \sum (y_i - \min y)^2 = 1$

- Earlier adjustments result in distribution in the plane
- Need the distribution in the original space
- One final adjustment
  - Comparison of the infinitesimal arc lengths in projected space to full space

Example 2- $T(y) : \min y = 0, \sum(y_i - \min y)^2 = 1$

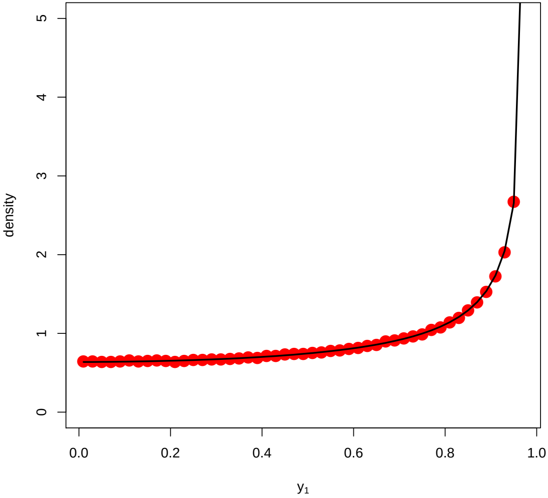


Example 2- $T(y) : \min y = 0, \sum (y_i - \min y)^2 = 1$



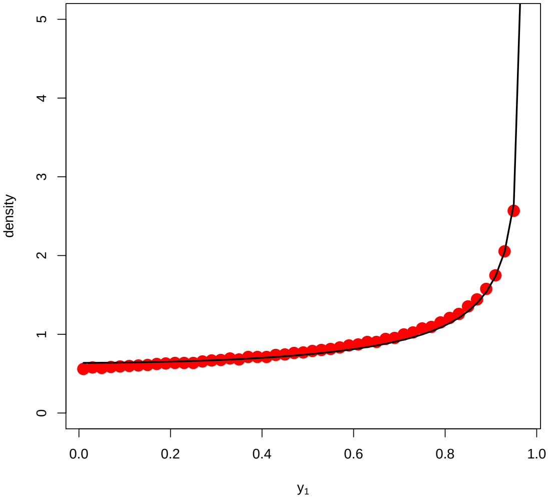
Example 2- $T(y)$: $\min y = 0, \sum (y_i - \min y)^2 = 1$

- MCMC for $[Y_1 | \theta, T(y), Y_2 = 0]$, $\theta = (\mu, \sigma) = (0, 1)$



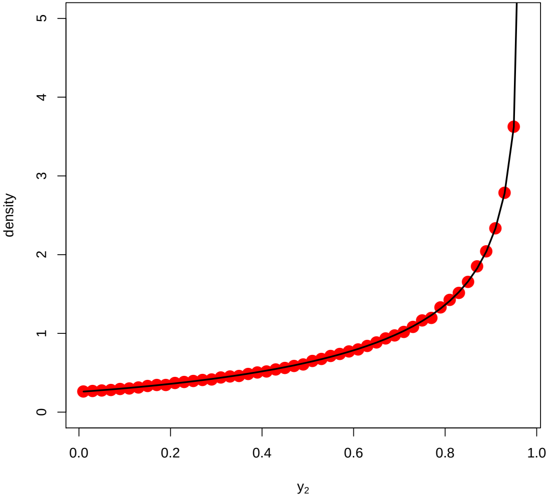
Example 2- $T(y)$: $\min y = 0, \sum(y_i - \min y)^2 = 1$

- Without the second adjustment



Example 2- $T(y)$: $\min y = 0, \sum(y_i - \min y)^2 = 1$

- MCMC for $[Y_2|\theta, T(y), Y_1 = 0]$, $\mu = (-.5, .4, .1)$, $\sigma = .5$



Summary of the MH Step

- The recipe for the proposal
  - Start Y^* uniform on unit sphere
  - Transform $H : Y^* \to Y^{**}$ so that $T(Y^{**}) = T(y)$
- H involves
  - Stretch in deviation space: r^{-1}
  - Attenuation in the deviation space
  - Attenuation for the original space

Summary of the MH Step

- $T(Y) = (L(Y), S(Y))$
- Both attenuations involve ∇L and ∇S
- Extension to n dimensions
  - Stretch in deviation space: $r^{-(n-2)}$
  - Attenuation in deviation space: compare norms of the $n-1$ dimensional tangent spaces for the sphere and the target manifold
  - Attenuation in original space: comparing $n-2$ dimensional volumes in the original space to the deviation space
  - Calculations need ∇L and ∇S

Conclusion

- The illustration today concerned the location and scale problem
  - Concern is outliers
- Modeling and computation extends to more interesting structures
  - Inclusion of covariates
  - Hierarchical models
- Benefits include
  - Reduced bias and smaller posterior variance
  - Ability to pool information
  - Ability to incorporate external information
- Blended paradigm also addresses model misspecification