# Outlier Analysis with Bayesian inference

Hao Chi Kiang

# Outline

# Normal location-scale model for outlier detection
(Verdinelli and Wasserman 1991)

- Let $y = (y_i)_{i \in I}$ be a family of random quantities with density

$$p(y_i | \mu, \sigma^2, A_i) = (1 - \epsilon)\phi(y_i; \mu, \sigma^2) + \epsilon\phi(y_i; \mu + A_i, \sigma^2)$$

  where
  - $\phi$ is the normal density.
  - $\epsilon$ is the probability of outlying and assumed to be known
  - $\epsilon \in ]0, \frac{1}{2}[$
  - $\mathrm{A} = (\mathrm{A}_\mathrm{i})_{\mathrm{i} \in \mathrm{I}}$ are ‘outliers’ displacement’
- We know $y_i$ and want to infer $\mu$, $\sigma$ and $A$

# Re-expression of location-scale problem

- For each $i \in I$, let $\delta_i$ be independent Bernouli variable, with $E(\delta_i) = \epsilon$. Then the likelihood is:

$$y_i|\mu, \sigma^2, A, \delta \sim N(\mu + \delta_i A_i, \sigma^2)$$

  - $\delta_i = 1$ means $y_i$ is an outlier.
  - Assume $(A_i)_{i\in I}$ are mutually independent
- We need 4 conditional distributions for Gibbs sampling

## Conditional posteriors for Gibbs sampling ($\mu$ and $\sigma^2$)

- Let $y_i^* = y_i - \delta_i A_i$, then $y_i^* \sim N(\mu, \sigma^2)$
- Semi-conjugate prior for $\mu$ and $\sigma$

$$\mu | A, \delta \sim N(\mu_0, \tau_0^2)$$

$$\sigma^2 | A, \delta \ \sim \text{Scaled-inv-}\chi^2(\nu_0, \sigma_0^2)$$

- Posterior

$$\mu | y, \sigma^2, A, \delta \sim N\left(\frac{\tau_0^2 \mu_0 + \sigma^{-2} \sum_{i \in I} y_i^*}{\tau_0^2 + n\sigma^{-2}}, \frac{1}{(\tau_0^2 + n\sigma^{-2})}\right)$$

$$\sigma^2 | \mu, y, A, \delta \sim \text{Scaled-inv-}\chi^2(\nu_0 + n, \frac{\nu_0 \sigma_0^2 + \sum_{i \in I}(y_i^* - \mu)^2}{\nu_0 + n})$$

where

$$n = |I|$$

## Conditional posteriors for Gibbs sampling ($\delta$ and $A$)

- The Bernouli outlier assignments $\delta_i$:

$$E(\delta_i|y,\mu,\sigma^2,A) = \frac{\epsilon\phi(y_i;\mu+A_i,\sigma^2)}{\epsilon\phi(y_i;\mu+A_i,\sigma^2)+(1-\epsilon)\phi(y_i;\mu,\sigma^2)}$$

- The outlier displacement $A_i$
  - Prior $A_i \sim N(0,\xi^{-2})$
  - Posterior

$$\begin{aligned} p(A_i|y,\mu,\sigma^2,\delta) = & \delta_i\phi\left(A_i; y_i-\mu,(\xi^2+\sigma^{-2})^{-1}\right) \\ & + (1-\delta_i)\phi(A_i;0,\xi^{-2}) \end{aligned}$$

- Now we have all the posterior conditional distributions. Gibbs sampling!

# Alternative: assuming unknown $\epsilon$

- In practice, setting $\epsilon = 0.05$ works well
- But we can treat it as random variable if we want
  - Assume $\epsilon$ depends only on $\delta$
  - Prior
$$\epsilon \sim Beta(\alpha, \beta)$$
  - Posterior
$$\epsilon|\delta = Beta\left(\alpha + \sum_{i\in I} \delta_i, \ \ \beta + \sum_{i \in I}(1-\delta_i)\right)$$

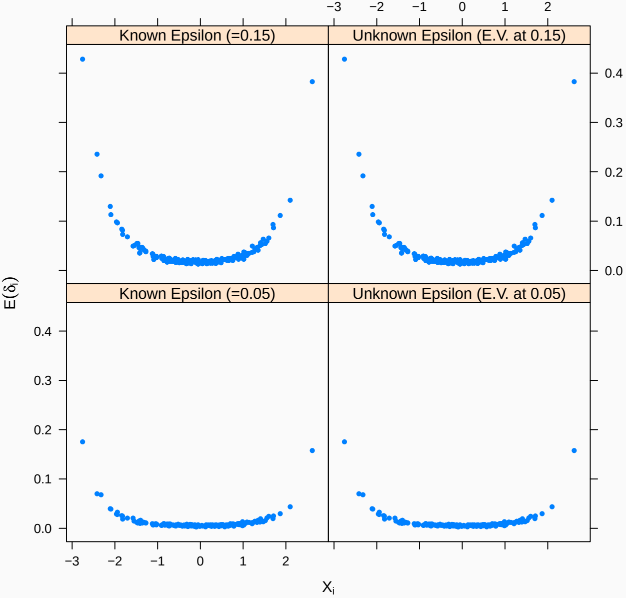
# How to set $\alpha$ and $\beta$?

- User set the $E(\epsilon)$ for the prior
- Additionally, we assume $p(\epsilon < \frac{1}{2}) = 0.99$
- Solve equations r.w.t $\alpha$ and $\beta$

$$\begin{cases} E(\epsilon) = \frac{\alpha}{\alpha+\beta} \\ \frac{\int_0^{\frac{1}{2}} t^{\alpha-1}(1-t)^{\beta-1}dt}{\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt} = 0.99 \end{cases}$$

- Can be done in R with '`uniroot()`' and '`pbeta()`' with three lines of codes

## Probability of being outlier



# Mixture of normals with outlier detection

- Multiple mixtures of normals $\mathrm{N}(\mu_\mathrm{k}, \sigma^2_\mathrm{k})$, $k \in K$
- Mixture Weights: $(w_k)_{k\in K}$
- Augment data with
  - Mixture indicator $(z_i)_{i\in I}$ where $z_i \in \{1, \ldots, k\}$
  - Outliers indicator $(\delta_i)_{i\in I}$
  - Outliers' displacement $(A_i)_{i\in I}$
- Likelihood becomes

$$p(y_i|z, \mu, \sigma, \delta, A, w) = \sum_{k\in K} w_k \phi(y_i; \mu_k + \delta_i A_i, \sigma^2_k)(z_i = k)$$

  where $(E)$ denotes indicator function of proposition $E$
- Note that $A_i$ depends on $z_i$ now

# Mixture of normals with outlier detection ($\mu$ and $\sigma$)

- Let $y_i^* = y_i - \delta_i A_i$. Then

$$p(y_i^*|z, \mu, \sigma, \delta, A, w) = \sum_{k \in K} w_k \phi(y_i^*; \mu_k, \sigma_k^2)(z_i = k)$$

- Semi-conjugate prior for the $\mu$'s and $\sigma$'s

$$\mu_k|A, \delta, w, z_i, \epsilon \sim N(\mu_0, \tau_0^2)$$
$$\sigma_k^2|A, \delta, w, z_i, \epsilon \sim \text{Scaled-inv-}\chi^2(\nu_0, \sigma_0^2)$$

# Mixture of normals with outlier detection ($\mu$ and $\sigma$)

- Conditional Posterior for the $\mu$'s and $\sigma$'s

$$\mu_k|y, \sigma^2, A, \delta, w, z_i, \epsilon \sim N\left(\frac{\tau_0^2\mu_0 + \sigma_k^{-2}\sum_{i\in I_k} y_i^*}{\tau_0^2 + n_k\sigma_k^{-2}}, \frac{1}{(\tau_0^2 + n_k\sigma_k^{-2})}\right)$$

$$\sigma_k^2|y, \mu, A, \delta, w, z_i, \epsilon \sim \text{Scaled-inv-}\chi^2\left(\nu_0 + n_k, \frac{\nu_0\sigma_0^2 + \sum_{i\in I_k}(y_i^* - \mu_k)^2}{\nu_0 + n_k}\right)$$

where

$$I_k = \{i \in I : z_i = k\}$$
$$n_k = |I_k|$$

# Mixture of normals with outlier detection ($z$ and $w$)

- Conditional posterior for $z$

$$p(z_i|y, w, \mu, \sigma, w, A, \delta, \epsilon) = \sum_{k \in K} \frac{w_k \phi(y_i; \mu_k, \sigma_k^2)}{\sum_{m \in K} w_m \phi(y_i; \mu_m, \sigma_m^2)} (z_i = k)$$

- Conditional posterior for $w$
  - Depends only on $z$
  - Use the Dirichlet-Multinomial conjugate prior

$$w|y, \mu, \sigma, A, \delta, \epsilon \; Dirichlet(\gamma)$$

  - Then posterior is

$$w|y, \mu, \sigma, z, A, \delta, \epsilon \; Dirichlet\left((n_k + \gamma_k)_{k \in K}\right)$$

## Mixture of normals with outlier detection ($\delta$)

- Conditional posterior for $\delta_i$

$$E(\delta_i | y, \mu, \sigma^2, w, z, A, \epsilon) = \frac{\epsilon \phi(y_i; \dot{\mu}_i + A_i, \dot{\sigma}_i^2)}{\epsilon \phi(y_i; \dot{\mu}_i + A_i, \dot{\sigma}_i^2) + (1 - \epsilon)\phi(y_i; \dot{\mu}_i, \dot{\sigma}_i^2)}$$

where

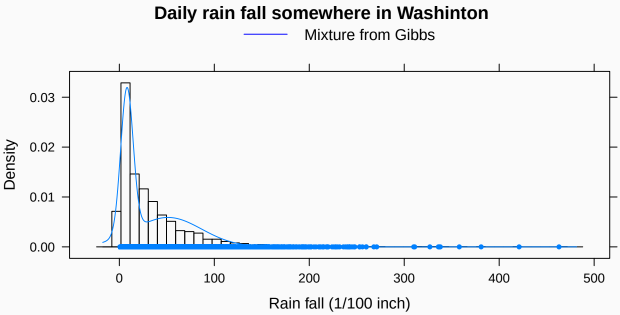
$$\dot{\mu}_i = \sum_{k \in K} \mu_k (z_i = k)$$
$$\dot{\sigma}_i^2 = \sum_{k \in K} \sigma_k^2 (z_i = k)$$

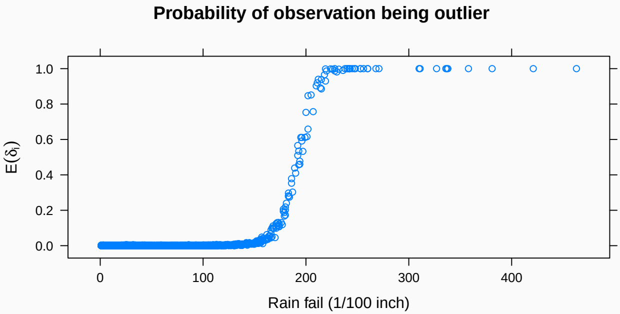
# Mixture of normals with outlier detection ($A$ and $\epsilon$)

- Conditional posterior for $A_i$
  - Prior $A_i \sim N(0, \xi^{-2})$

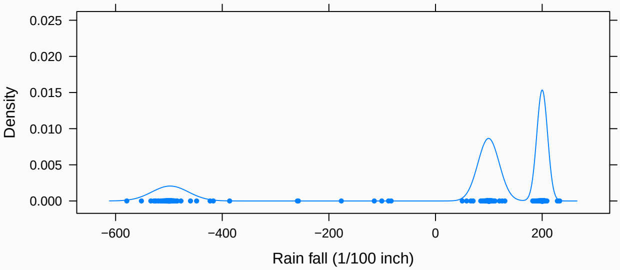
$$
\begin{aligned}
p(A_i|y, \mu, \sigma^2, w, z, \delta, \epsilon) =& \delta_i \phi\left(A_i; y_i - \dot{\mu}_i, (\xi^2 + \dot{\sigma}_i^{-2})^{-1}\right) \\
&+ (1 - \delta_i)\phi(A_i; 0, \xi^{-2})
\end{aligned}
$$

  - Completely the same as non-mixture version, except $\mu$ and $\sigma$ becomes $\dot{\mu}_i$ and $\dot{\sigma}_i$
- Considiton posterior of $\epsilon$ is the same as non-mixture version

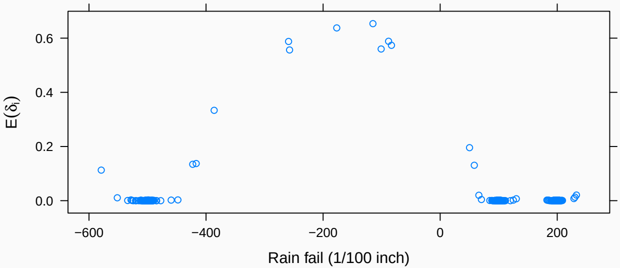

Daily rain fall somewhere in Washinton
Mixture from Gibbs
Density
0.03
0.02
0.01
0.00
0
100
200
300
400
500
Rain fall (1/100 inch)


Probability of observation being outlier
$E(\delta_i)$
1.0
0.8
0.6
0.4
0.2
0.0
0
100
200
300
400
Rain fail (1/100 inch)

**Daily rain fall somewhere in Washinton**




**Probability of observation being outlier**



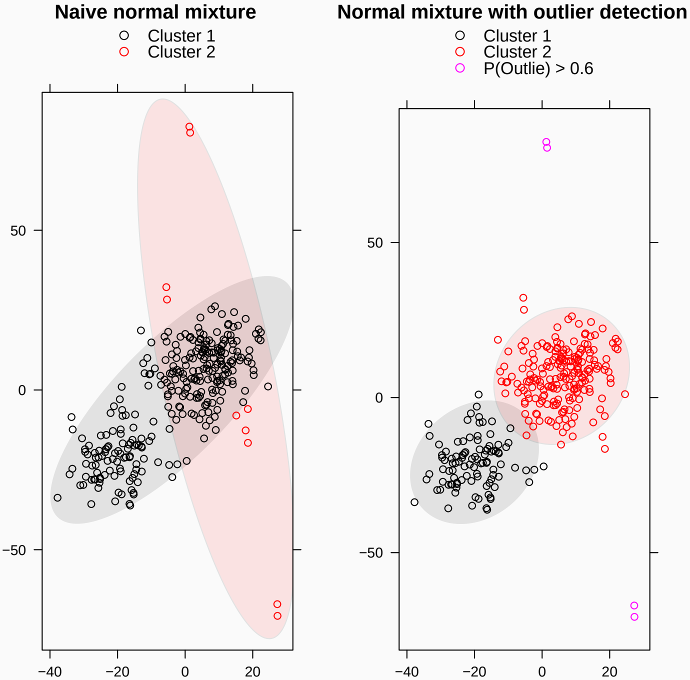
# Mixture of multivariate normals with outlier detection

- Everything is the same for multivariate, except $\mu$ and $\Sigma$
- Use Normal-Inverse-Wishart joint prior instead

$$\Sigma_k | A, \delta, w, z_i, \epsilon \sim IW(\nu_0, \Lambda_0)$$
$$\mu_k | \Sigma, A, \delta, w, z_i, \epsilon \sim N\left(\mu_0, \frac{1}{\kappa_0}\Sigma_k\right)$$
$$\Sigma_k | y, A, \delta, w, z_i, \epsilon \sim IW\left(\nu_0 + n_k, \Lambda_0 + S_k + \frac{\kappa_0 n_k}{\kappa_0 + n_k}(m_k - \mu_0)(m_k - \mu_0)^T\right)$$
$$\mu_k | y, \Sigma, A, \delta, w, z_i, \epsilon \sim N\left(\frac{\kappa_0 \mu_0 + n_k m_k}{\kappa_0 + n_k}, \frac{1}{\kappa_0 + n_k}\Sigma_k\right)$$

where

$$m_k = \frac{1}{n_k}\sum_{i \in I_k} y_i^*$$
$$S_k = \sum_{i \in I_k}(y_i^* - m_k)(y_i^* - m_k)^T$$



Naive normal mixture
Cluster 1
Cluster 2
50
0
−50
−40
−20
0
20
Normal mixture with outlier detection
Cluster 1
Cluster 2
P(Outlie) > 0.6
50
0
−50
−40
−20
0
20

## Bayesian Regression with outlier detection

- Simple linear regression model

$$y_i|\beta = \beta^T x_i + r_i$$

where $(r_i)_{i \in I}$ are mutually independent $N(0, \sigma^2)$

- But it hurts when there are outlying residuals
- Solution: Augment the data and redefine the likelihood

$$y_i|\beta, \sigma^2, A, \epsilon, (\delta_i = 0) = \beta^T x_i + r_i$$
$$y_i|\beta, \sigma^2, A, \epsilon, (\delta_i = 1) = A_i + r_i$$

  - Implication:
    - $E(r_i|\delta, A) = 0 + \delta_i A_i$
    - $\beta$ doesn't depend on outlying data any more

# Bayesian Regression with outlier detection ($\beta$ and $\sigma$)

- Joint prior for $\beta$ and $\sigma$

$$\beta|\sigma^2, \delta, A, \epsilon \sim N(\mu_0, \sigma^2\Omega_0^{-1}) \quad \sigma^2|\delta, A, \epsilon \sim \text{Scaled-inv-}\chi^2(\nu_0, \sigma_0^2)$$

- Posterior

$$\beta|y, \sigma^2, \delta, A, \epsilon \sim N\left(\mu_n, \sigma^2\Omega_n^{-1}\right) \quad \sigma^2|y, \delta, A, \epsilon \sim \text{Scaled-inv-}\chi^2\left(\nu_n, \sigma_n\right)$$

where

$$\mu_n = \Omega_n^{-1}(\dot{X}^T\dot{y} + \Omega_0\mu_0)$$

$$\sigma_n = \frac{1}{\nu_n}(\nu_0\sigma_0^2 + (\dot{y}^T\dot{y} + \mu_0^T\Omega_0\mu_0 - \mu_n^T\Omega_n\mu_n))$$

$$\nu_n = \nu_0 + \sum_{i\in I}(\delta_i = 0)$$

$$\Omega_n = \dot{X}^T\dot{X} + \Omega_0$$

$$\dot{X} = \begin{bmatrix} x_{j_1} & x_{j_2} & \cdots \end{bmatrix}^T, \; j_p \in \{i : \delta_i = 0\}$$

$$\dot{y} = \begin{bmatrix} y_{j_1} & y_{j_2} & \cdots \end{bmatrix}^T$$

## Bayesian Regression with outlier detection ($\delta$)

- Conditional posterior of $\delta$

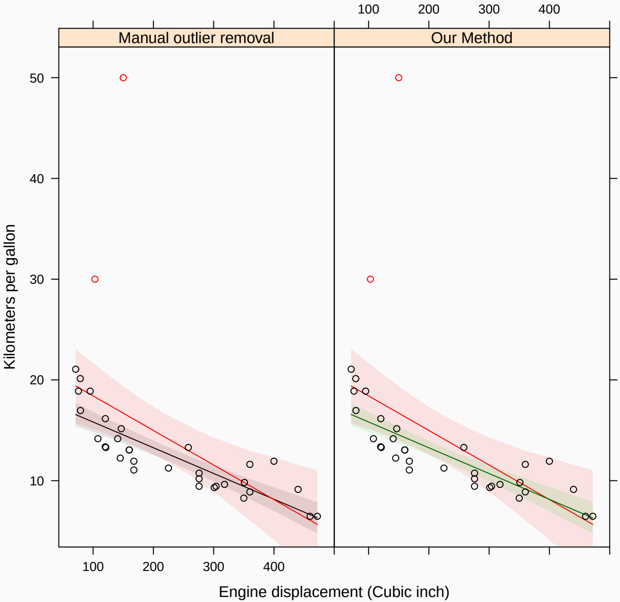
$$E(\delta_i|y,\beta,\sigma^2,A,\epsilon) = \frac{\epsilon\phi(r_i;A_i,\sigma^2)}{\epsilon\phi(r_i;A_i,\sigma^2) + (1-\epsilon)\phi(r_i;0,\sigma^2)}$$

  Note that $r$ is given when both $y$, $\beta$ are given:

$$r_i = y_i - \beta^T x_i$$

- Conditonal posterior of $A$ and $\epsilon$ are just same as before

## Fuel efficiency of some cars vs. Engine displacement (mtcars data set in R)



## Works Cited

Verdinelli, Isabella and Larry Wasserman (1991). "Bayesian analysis of outlier problems using the Gibbs sampler". In: *Statistics and Computing* 1, pp. 105–117.

## Source code

`https://github.com/hckiang/bayesian-outlier-model`