Example: Mixture of Normals

$$P(y) \geq E_q[\log P(\theta, y)] - E_q \log q$$

Mean-Field: $q(\theta_1, \ldots \theta_d) = \prod_i q(\theta_i)$

while (ELBO not converged) {

update $q(\theta_1)$ given $q(\theta_{-1})$

update $q(\theta_2)$ given ...

θ_1

$$\log q(\theta_i) = E_{q(\theta_{-i})}[\log P(\theta, y)] + \text{const}$$

$$\theta \sim Dir_K(\alpha) \quad (\text{uniform})$$

$$z_i \sim Mult_K(\theta)$$

$$y_i | z_i \sim N(\mu_{z_i}, \sigma^2_{z_i})$$

$$P(\theta, z, \mu, \sigma^2 | y) \propto$$

$$L(\theta, z, \mu, \sigma^2) P(\cdot,\cdot) \propto$$

$$\left(\prod_{i=1}^{n} \left[\prod_{k=1}^{K} \left(\frac{1}{\sqrt{2\pi\sigma_k^2}} e^{\frac{-(y_i - \mu_k)^2}{2\sigma_k^2}} \right)^{z_{ik}} \theta_k^{z_{ik}} \right] \right)$$

$$\prod^{K} \frac{1}{\sqrt{2\pi\tau^2}} e^{-\frac{\mu_k^2}{2\tau^2}}$$

$$\log P(\theta, y) =$$

$$\sum_{i=1}^{n}\sum_{k=1}^{K}\left(-\log\sigma_k^2 - \frac{(y_i-\mu_k)^2}{2\sigma_k^2} + \log\theta_k\right)z_{ik} + \log \text{Prior}$$

$$\log q(z_i) = \sum_k z_{ik}\left(\underset{q(-z_i)}{E}\frac{(y_i-\mu_k)^2}{2\sigma_k^2} + \underset{q(-z_i)}{E}[\log\theta_k]\right)$$

$$\Rightarrow P(z_{ik}=1) \propto \exp\left[\left(\underset{q(-z_i)}{E}\left[\frac{(y_i-\mu_k)^2}{2\sigma_k^2}\right] + \underset{q(-z_i)}{E}[\log\theta_k]\right)\right]$$

Mult

$$\propto \exp\left(E_\mu\frac{(y_i-\mu_k)^2}{2} \times \underset{1/\sigma^2}{E}\left[\frac{1}{\sigma_k^2}\right] + \underset{\theta_k}{E}[\log\theta_k]\right)$$

$$\log q(\mu_k) = \sum_{i=1}^{n} -\frac{(y_i-\mu_k)^2}{2}\overbrace{\underset{1/\sigma^2}{E}\left[\frac{1}{\sigma_k^2}\right]E[z_{ik}]}^{w_i} + \text{Prior} + \text{const}$$

$$\Rightarrow \mu_k \sim N\left(\frac{\sum w_i y_i}{\sum w_i}, \frac{1}{\sum w_i}\right)$$

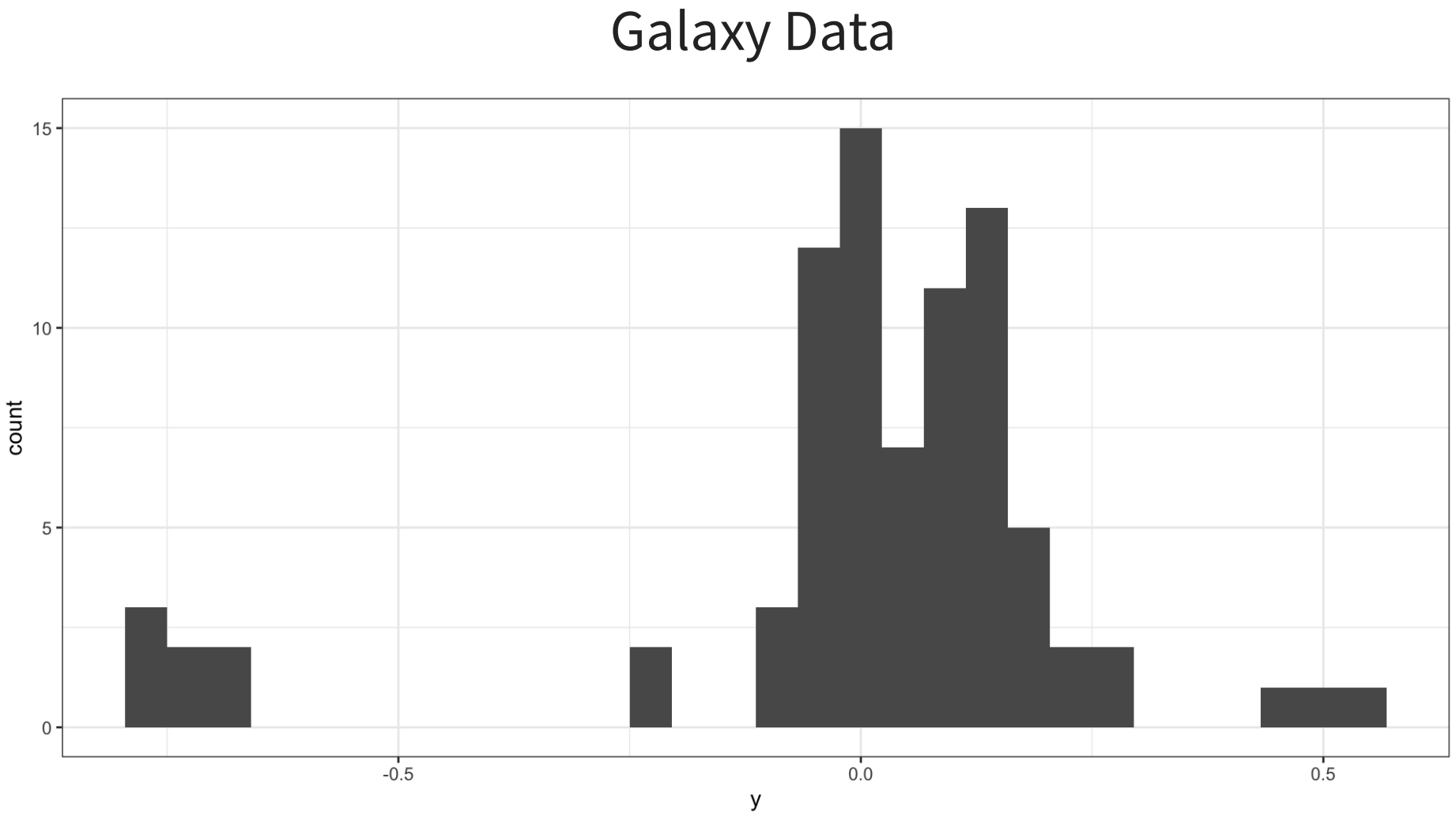
$$\log q(\theta) \overset{+}{=} \sum_{i=1}^{n} \sum_{k} E_{q(\theta_k)}[z_{ik}] \log \theta_k$$

$$\overset{+}{=} \sum_{k} \sum_{i=1}^{n} E[z_{ik}] \log \theta_k$$

$$\Rightarrow q(\theta) \propto \theta_1^{\sum_{i=1}^{n} E[z_{i1}]} \theta_2^{\sum E[z_{i2}]} \cdots \theta_K^{\sum E[z_{iK}]}$$

}

Example: Mixture of Normals



Example: Normal Mixture

Example: Normal Mixture

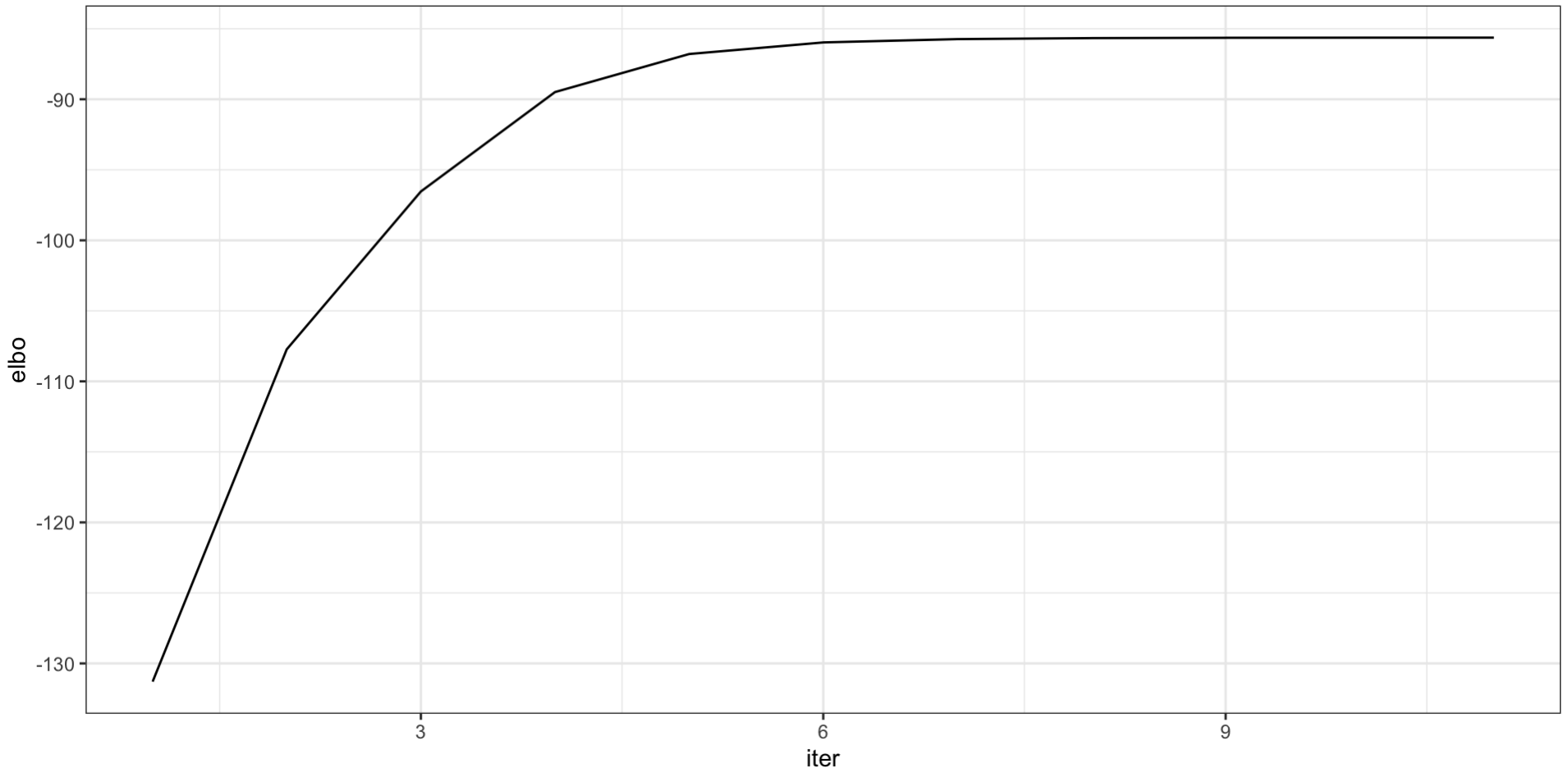
```r
elbo_prev <- -Inf
elbo_cur <- sum(z_probs * (matrix(-y^2/2, nrow=n, ncol=k) +
                           y %o% mu_mean - rep(1, n) %o% (mu_sq_mean)/2)*prec_mean)
elbo <- c()

while((elbo_cur - elbo_prev)^2 > 1e-5) {

  z_probs_raw <- (matrix(-y^2/2, nrow=n, ncol=k) + y %o% mu_mean - rep(1, n) %o% (mu_sq_mean)/2)*prec_mean +
    matrix(digamma(alpha_k) - digamma(sum(alpha_k)), nrow=length(y), ncol=k, byrow=TRUE)
  z_probs_raw <- t(apply(z_probs_raw, 1, function(x) x - max(x)))
  z_probs <- exp(z_probs_raw)
  z_probs <- z_probs/rowSums(z_probs)


  w_i <- t(t(z_probs) * prec_mean)
  mu_mean <- colSums(y * w_i) / (colSums(w_i) + 1)
  mu_sq_mean <- 1/(colSums(w_i) + 1) + (colSums(y * w_i) / (colSums(w_i) + 1))^2

  alpha_k <- colSums(z_probs) + 1

  elbo_prev <- elbo_cur
  elbo_cur <- sum(z_probs * (matrix(-y^2/2, nrow=n, ncol=k) + y %o% mu_mean - rep(1, n) %o% (mu_sq_mean)/2)*p

  elbo <- c(elbo, elbo_cur)
}
```

Conv.

$q(Z_{ik})$

Mult for Z_n

ELBO convergence



Mean and cluster assignments