Last lecture

Minimum mean square error estimation (Ch 4.9)

- Recap
  - Constant estimators
  - Unconstrained estimators
  - Linear estimators
- Examples

Joint Gaussian Distribution (Ch 4.11)

- Motivation
- Facts
- Examples

Agenda

- Law of Large Numbers and Central Limit Theorem (Ch 4.10)
  - Equations
  - Examples

- Expectation Maximization (Not tested)
  - Motivation
  - Application
  - Intuition
  - Algorithm
  - Visualization

Equations – Law of Large Numbers

$X_1 \dots X_N$ are uncorrelated, and $\mu_{X_k} = \mu$ and $\sigma_{X_k}^2 \leq C$

- Let $S_n = \sum_1^n X_n$, we have $P\left\{\left|\frac{S_n}{n} - \mu\right| \geq \delta\right\} \leq \frac{C}{n\delta^2}$ for any δ
- Prove by Chebyshev

Example

$X_1 \dots X_{100}$ are RVs with $\mu = 5$ an $\sigma^2 = 1.\ S_n = \sum_1^n X_n$

- Assume $|Cov(X_i, X_j)| \leq 0.1$ if $i = j \pm 1$
- $Cov(X_i, X_j) = 0$ if $|i - j| \geq 2$

- Show $Var(S_{100}) < 120$

- Find upper bound of $P(\left|\frac{S_{100}}{100} - 5\right| \geq 0.5)$

Equations – Central Limit Theorem

$X_1 \ldots X_N$ independent, identically distributed (i.i.d.) RVs

- Let $S_n = \sum_1^n X_n$, we have $\lim_{n\to\infty} P\left\{\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq c\right\} = \Phi(c)$ for any c

- For examples, refer to Gaussian approximation

Final Review Survey

- Open text

- April 30 (Thur)
  - Final Remarks
  - Review on topics before midterm II

- May 5 (Tue)
  - Review on topics after midterm II

- May 7 (Tue)
  - Reading day, no class

#8698199

All the rest of this lecture will not be in the exam

Expectation Maximization

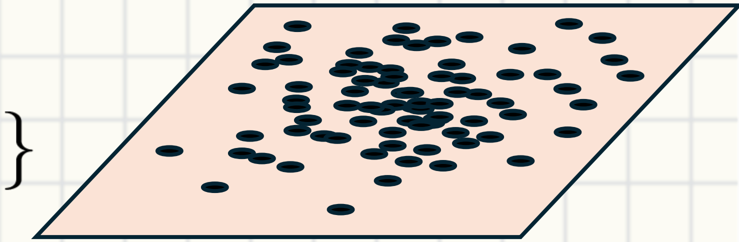
Motivation

- In many cases, we will face a “   ” problem
- Example
    - Given location of many bees x_i, estimate their bee hive location θ
    - If $x_i \sim N(\theta, \sigma^2)$
    - What if they belong to two hives?
- We need to classify/ separate signals to estimate their parameter
- The estimated parameter can help classify the signals

Application

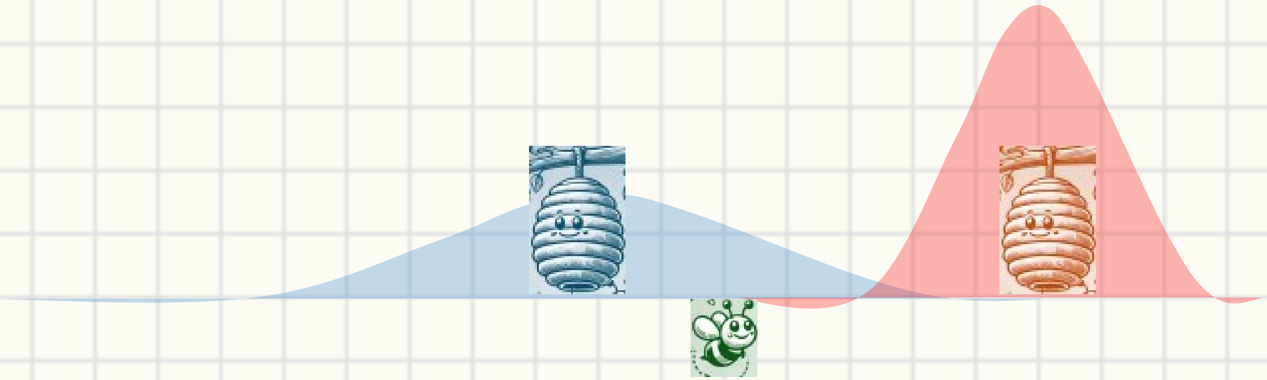
- Unsupervised learning
- Speech recognition
    - Align speech segment to transcriptions
- Image segmentation
- Customer segmentation

Intuition

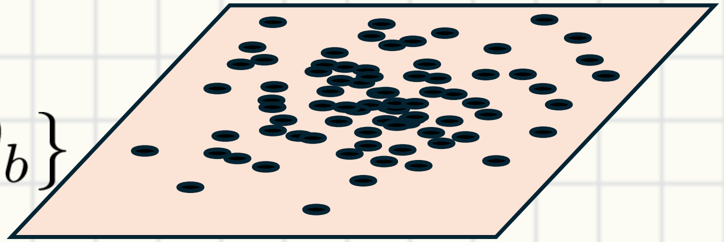


Given: Bee locations X Estimate: bee colors Z and bee hives $\{\theta_r, \theta_b\}$

But we can estimate Z if we know locations of both hives



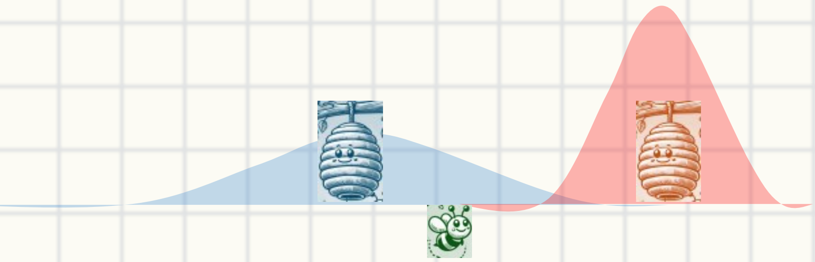
Intuition



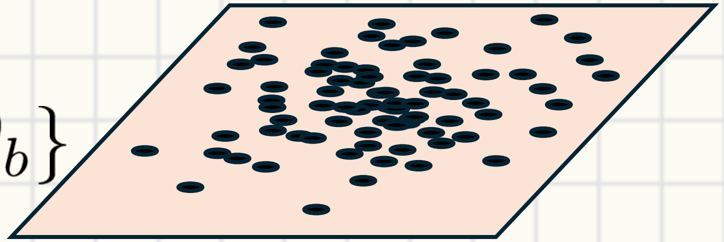
Given: Bee locations X Estimate: bee colors Z and bee hives $\{\theta_r, \theta_b\}$

But we can estimate Z if we know locations of both hives

$$p(z=r|x,\theta_r,\theta_b) = \frac{p(x|z=r,\theta_r)p(z=r|\theta_r)}{p(x|\theta_r,\theta_b)}$$



Intuition



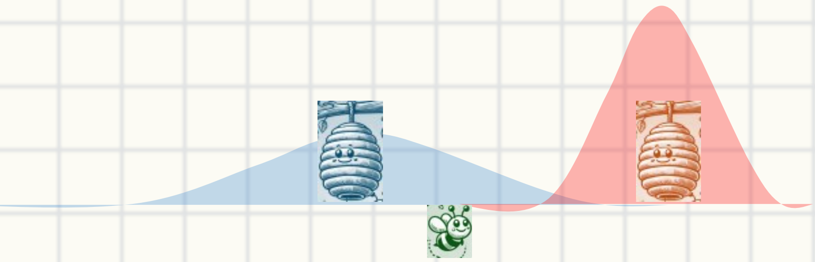
Given: Bee locations X Estimate: bee colors Z and bee hives $\{\theta_r, \theta_b\}$

But we can estimate Z if we know locations of both hives

$$p(z=r|x,\theta_r,\theta_b) = \frac{p(x|z=r,\theta_r)p(z=r|\theta_r)}{p(x|\theta_r,\theta_b)}$$

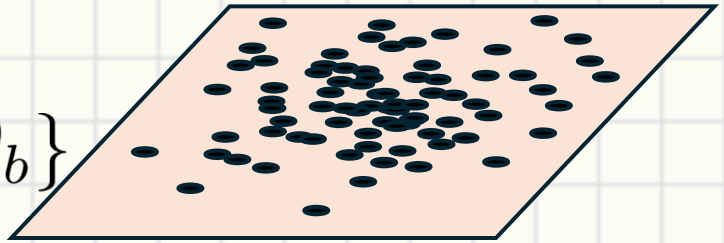
How likely a red bee is here

How likely a bee is here



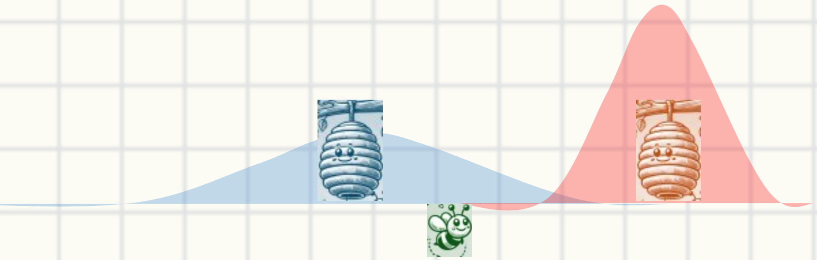
Intuition

Given: Bee locations X Estimate: bee colors Z and bee hives $\{\theta_r, \theta_b\}$



But we can estimate Z if we know locations of both hives

$$p(z = r|x, \theta_r, \theta_b) = \frac{p(x|z = r, \theta_r)p(z = r|\theta_r)}{p(x|\theta_r, \theta_b)}$$



Red bee distribution

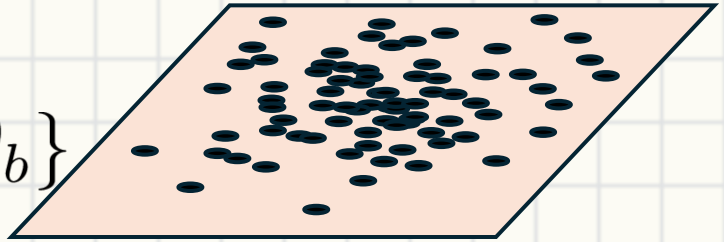
Size of the red hive

$$= \frac{p(x|z = r, \theta_r)p(z = r|\theta_r)}{p(x|z = r, \theta_r)p(z = r|\theta_r) + p(x|z = b, \theta_b)p(z = b|\theta_b)}$$

Red bee likelihood

Blue bee likelihood

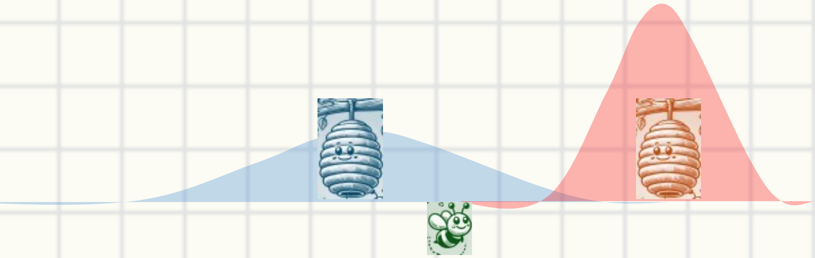
Intuition



Given: Bee locations X Estimate: bee colors Z and bee hives $\{\theta_r, \theta_b\}$

But we can estimate Z if we know locations of both hives

$$p(z=r|x,\theta_r,\theta_b) = \frac{p(x|z=r,\theta_r)p(z=r|\theta_r)}{p(x|\theta_r,\theta_b)}$$



Red bee distribution Size of the red hive

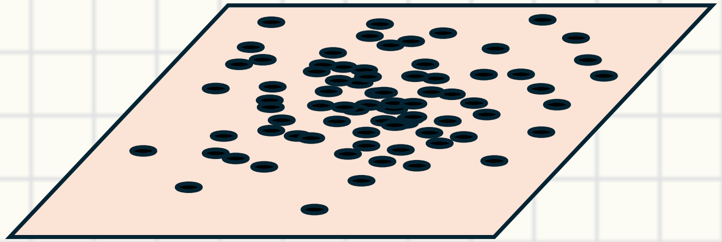
$$= \frac{p(x|z=r,\theta_r)p(z=r|\theta_r)}{p(x|z=r,\theta_r)p(z=r|\theta_r)+p(x|z=b,\theta_b)p(z=b|\theta_b)}$$

Red bee likelihood Blue bee likelihood

How do we know hive locations? We start with a guess $\theta = \{\ \theta_r^{(0)}\ ,\ \theta_b^{(0)}\ \}$

That's our first "chicken"!

Algorithms

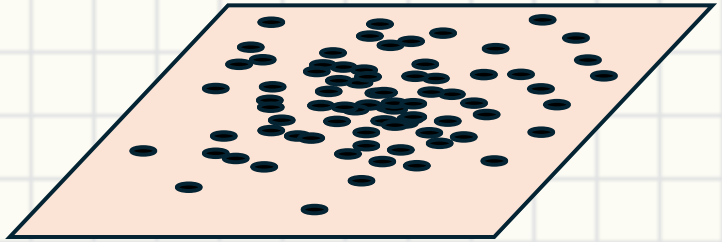


Basic idea:

1. Initialize $\theta = \{ \theta_r^{(0)} , \theta_b^{(0)} \}$ with arbitrary values

2. Write out the likelihood function $p(X, Z|\theta) = p(X|Z, \theta)p(Z|\theta)$ → Somehow need to get Z

Algorithms



Basic idea:

1. Initialize $\theta = \{ \theta_r^{(0)} , \theta_b^{(0)} \}$ with arbitrary values

2. Write out the likelihood function $p(X, Z|\theta) = p(X|Z,\theta)p(Z|\theta)$ → Somehow need to get Z

3. Treat Z as a random variable and compute it's distribution $q(z)$

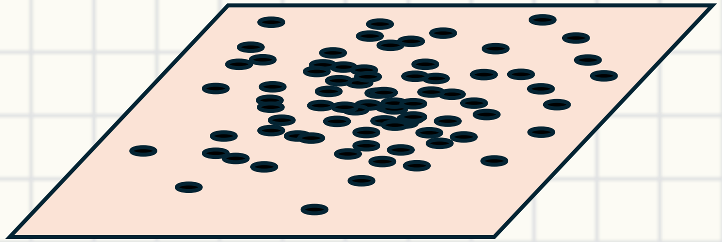
How? Compute the posterior $q(z) = p(Z \mid X, \theta)$
(i.e., probability of bin given ball color, probability of bee type given bee location)

4. Treat likelihood as a function of Z. Then, what is average likelihood?

$$\log p(X, Z \mid \theta)q(z) = logp(X|Z,\theta)p(Z|\theta)q(z)$$

$$= \mathbb{E}_{z\sim q(z)}[logp(X|Z,\theta)p(Z|\theta)]$$

Algorithms



Basic idea:

1. Initialize $\theta = \{ \theta_r^{(0)} , \theta_b^{(0)} \}$ with arbitrary values

2. Write out the likelihood function $p(X, Z|\theta) = p(X|Z, \theta)p(Z|\theta)$ → Somehow need to get Z

3. Treat Z as a random variable and compute it's distribution $q(z)$

   How? Compute the posterior $q(z) = p(Z \mid X, \theta)$
   (i.e., probability of bin given ball color, probability of bee type given bee location)

4. Treat likelihood as a function of Z. Then, what is average likelihood?

$$\log p(X, Z \mid \theta)q(z) = logp(X|Z,\theta)p(Z|\theta)q(z)$$

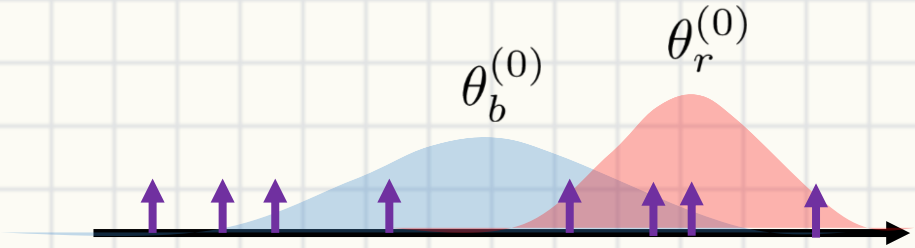
$$= \mathbb{E}_{z\sim q(z)}[logp(X|Z,\theta)p(Z|\theta)]$$

5. Maximize expected likelihood to get new θ
6. Return to step #1 until convergence

7. Return θ^*

Visualization

E-Step : Given estimated hive locations, Calculate bee color



Given θ**, estimate** $q(Z) = p(Z|X, \theta)$

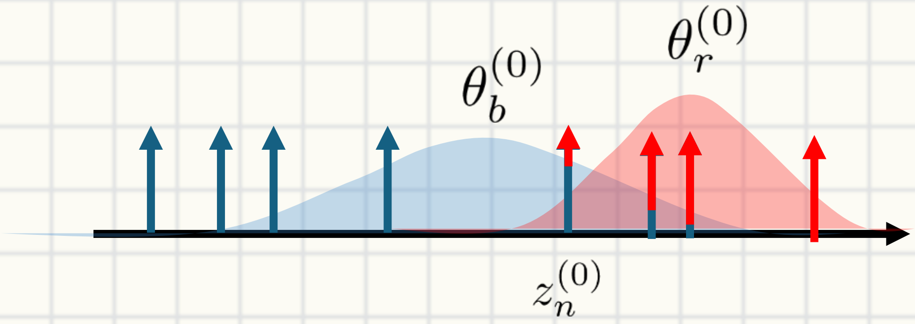
$$p(z_n^{(0)} = r | x_n, \theta_r^{(0)}) = \frac{p(x_n | z_n^{(0)} = r, \theta_r^{(0)} p(z^{(0)} = r | \theta_r^{(0)})}{p(x | \theta_r^{(0)}, \theta_b^{(0)})}$$

Visualization

E-Step : Given estimated hive locations, Calculate bee color

Given θ, estimate $q(Z) = p(Z|X, \theta)$

$$p(z_n^{(0)} = r|x_n, \theta_r^{(0)}) = \frac{p(x_n|z_n^{(0)} = r, \theta_r^{(0)} p(z^{(0)} = r|\theta_r^{(0)})}{p(x|\theta_r^{(0)}, \theta_b^{(0)})}$$



Visualization

E-Step : Given estimated hive locations, Calculate bee color

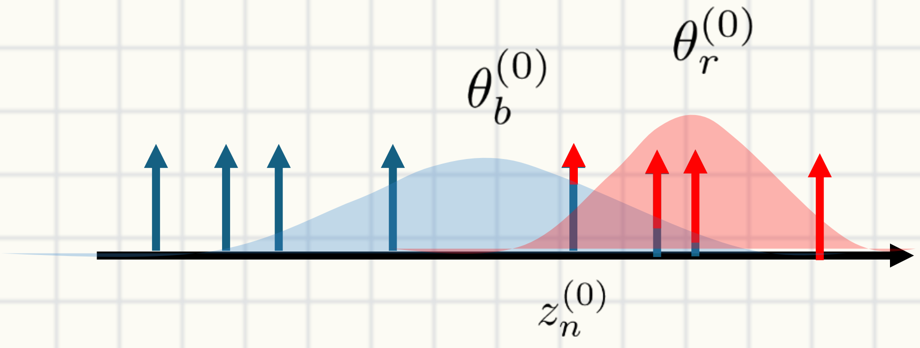
Given θ**, estimate** $q(Z) = p(Z|X, \theta)$

$$p(z_n^{(0)} = r | x_n, \theta_r^{(0)}) = \frac{p(x_n | z_n^{(0)} = r, \theta_r^{(0)}) p(z^{(0)} = r | \theta_r^{(0)})}{p(x | \theta_r^{(0)}, \theta_b^{(0)})}$$

M-Step : Given estimated bee color, Calculate hive locations

Given $q(Z)$**, estimate** θ

$$\theta^{(i+1)} = argmax_{\theta} \mathbb{E}_{q(Z)} [logp(X|Z, \theta) + logp(Z|\theta)]$$



Visualization

E-Step : Given estimated hive locations, Calculate bee color

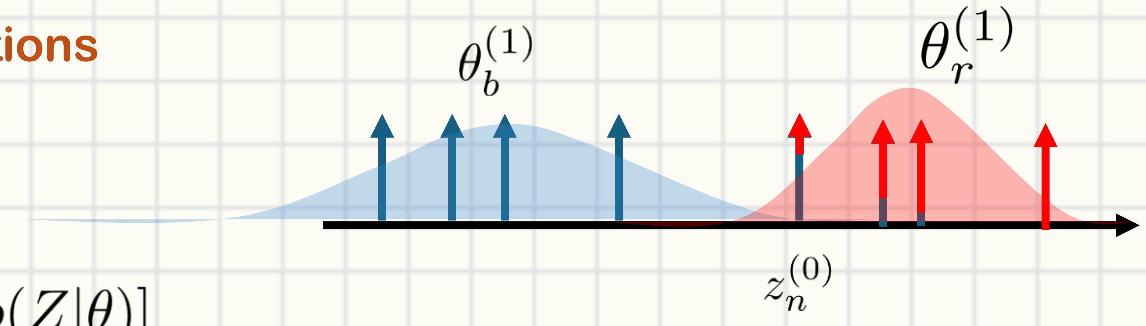
Given θ**, estimate** $q(Z) = p(Z|X,\theta)$

$$p(z_n^{(0)} = r|x_n, \theta_r^{(0)}) = \frac{p(x_n|z_n^{(0)} = r, \theta_r^{(0)} p(z^{(0)} = r|\theta_r^{(0)})}{p(x|\theta_r^{(0)}, \theta_b^{(0)})}$$

M-Step : Given estimated bee color, Calculate hive locations

Given $q(Z)$**, estimate** θ

$$\theta^{(i+1)} = argmax_\theta \mathbb{E}_{q(Z)}[logp(X|Z,\theta) + logp(Z|\theta)]$$

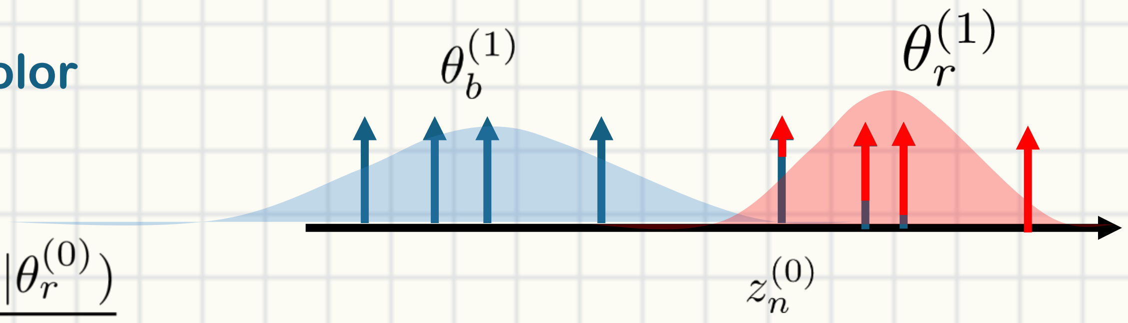


Visualization

E-Step : Given estimated hive locations, Calculate bee color

Given θ**, estimate** $q(Z) = p(Z|X, \theta)$

$$p(z_n^{(0)} = r | x_n, \theta_r^{(0)}) = \frac{p(x_n | z_n^{(0)} = r, \theta_r^{(0)}) p(z^{(0)} = r | \theta_r^{(0)})}{p(x | \theta_r^{(0)}, \theta_b^{(0)})}$$



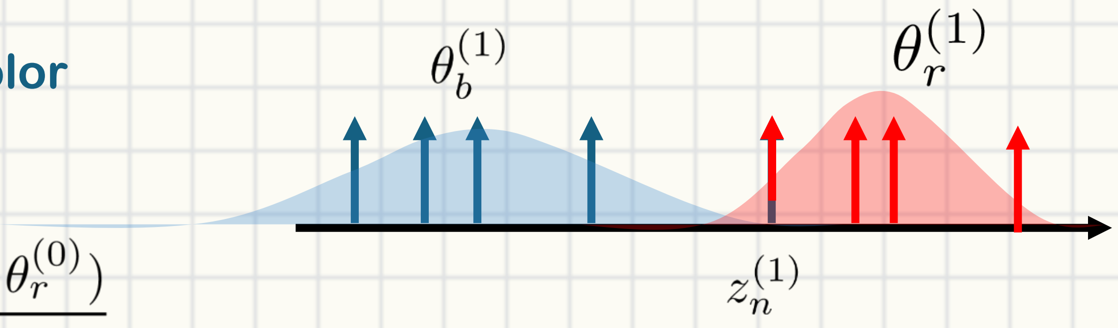
M-Step : Given estimated bee color, Calculate hive locations

Given $q(Z)$**, estimate** θ

$$\theta^{(i+1)} = argmax_\theta \mathbb{E}_{q(Z)}[logp(X|Z, \theta) + logp(Z|\theta)]$$

Visualization

E-Step : Given estimated hive locations, Calculate bee color



Given θ, estimate $q(Z) = p(Z|X,\theta)$

$$p(z_n^{(0)} = r|x_n, \theta_r^{(0)}) = \frac{p(x_n|z_n^{(0)} = r, \theta_r^{(0)}p(z^{(0)} = r|\theta_r^{(0)})}{p(x|\theta_r^{(0)}, \theta_b^{(0)})}$$

M-Step : Given estimated bee color, Calculate hive locations

Given $q(Z)$ **, estimate** θ

$$\theta^{(i+1)} = argmax_\theta \mathbb{E}_{q(Z)}[logp(X|Z,\theta) + logp(Z|\theta)]$$