Homework 1

Mark Wilber

2024-08-30

Question 1: Practicing with Bayes Rule

Suppose there are two globes, one of Earth and one of Mars. The Earth globe is 70% water and the Mars globe is 100% land. Someone randomly chooses a globe and tosses it high in the air and catches it and their thumb is on land. Use Bayes Rule to show that the $P(\text{Earth tossed}|\text{Land}) = 0.23$.

Hint: To solve this problem it is useful to remember the law of total probability: $P(\text{Land}) = P(\text{Land}|\text{Earth tossed})P(\text{Earth tossed}) + P(\text{Land}|\text{Mars tossed})P(\text{Mars tossed})$.

$$\begin{aligned}
P(\text{Land}) =& P(\text{Land}|\text{Earth tossed})P(\text{Earth tossed}) + P(\text{Land}|\text{Mars tossed})P(\text{Mars tossed}) \\
=& (1 - 0.7) * 0.5 + 1 * 0.5 \\
=& 0.65
\end{aligned}$$

$$\begin{aligned}
P(\text{Earth tossed}|\text{Land}) =& \frac{P(\text{Land}|\text{Earth tossed}) * P(\text{Earth tossed})}{P(Land)} \\
=& \frac{0.3 * 0.5}{0.65} \\
=& 0.2307692
\end{aligned}$$

(0.3*0.5)/0.65 # Question 2: Sandhill cranes and food supplementation

Wildlife managers are interested in asking the following question related to wildlife viewing opportunities: Does food supplementation in fields along sandhill crane (*Grus canadensis*) migration routes increase the likelihood that cranes will stop over and spend "non-neglible" time in food-supplemented fields (thus increasing wildlife viewing opportunities of these majestic animals)? Some researchers expect that food supplementation will not increase crane stopovers along migration routes as habitat factors such as aerially discernible vegetation structure are more important for determining stopover locations.

To answer this question, researchers identified 30 fallow agriculture fields in relatively close proximity in Alberta, Canada. They randomly assigned 15 of these fields to a "No food supplement" treatment and 15 of the fields to the a " Food supplement" treatment. On the first year of the study, researchers did not supplement any of the fields and used wildlife cameras and viewer observations to determine if fields experienced greater than 2 crane nights of use during the migration season (e.g., 1 crane in the field for 2 days, 2 cranes in the field for 1 day, etc.). If a field experienced greater than 2 crane nights it was given a 1, otherwise it was marked as 0. In the second year, researchers added food supplement to the 15 fields in the "Food supplement" group once a week during the migration season. They again used wildlife cameras and viewer observations to record whether fields experienced greater than (1) or less than (0) 2 crane nights during the migration season.

The results from this management experiment are given in `crane_data.csv`. Load this data into R and use a Bayesian analysis to answer the researcher's question: **Does food supplementation in fields along sandhill crane migration routes increase the probability of that cranes will use the fields as stopover locations?**.

```r
# Reading in the crane data
data = read.csv("crane_data.csv")
print(data)
```

```
##      treatment year1_fields_used year2_fields_used year1_total_fields
## 1   supplement                 3                14                 15
## 2 no_supplement                2                 7                 15
##   year2_total_fields
## 1                 15
## 2                 15
```

When answering this question, be sure to clearly address the following subquestions 1. What are the parameters you are estimating to answer this question? 2. What is the likelihood you are using for the data? 3. What prior distributions are you using? 4. What does the model you are fitting look like when written in model syntax (see below) and in R code? 4. How are you choosing to generate the posterior distribution (e.g., grid approximation vs. quadratic approximation)? 5. How are you comparing posterior distributions to answer the question of interest?

When you provide an answer to the researcher's question, support your statement with a plot of the posterior distribution and statement about how certain you are in your conclusions based on your posterior

Note: You can write a Bayesian model in R markdown using the following syntax.

$$
\begin{aligned}
y &\sim \text{Binomial}(N, p) \\
p &\sim \text{Uniform}(0, 1)
\end{aligned}
$$

the `\\` specifies a line break and the `&` specifies where you want the equations aligned.

Hint: Use the northern harrier and red-tailed hawk example from class as a template. We will learn a much more efficient way to answer this question later in class, but for now think about fitting four separate Bayesian models and comparing the posterior distributions of the parameters. You will need to think about exactly how you want to compare these posterior distributions to answer the question of interest.

Description of crane data set

- `treatment`: Either `supplement` (field had food supplement in year 2) or `no_supplement` (field did not have a food supplement in year 2)
- `year1_fields_used`: The number (out of 15) fields that were used by cranes in year 1
- `year2_fields_used`: The number (out of 15) fields that were used by cranes in year 2
- `year1_total_fields`: Total number of fields in the group in year 1 (15)
- `year2_total_fields`: Total number of fields in the group in year 2 (15)

$$
\begin{aligned}
P(\text{field is used in year k}|\text{field is supplemented}) &= \frac{P(\text{field is supplemented}|\text{field is used in year k}) * P(\text{field is used in year}}{P(\text{field is supplemented})} \\
P(\text{field is used in year k}|\text{field is not supplemented}) &= \frac{P(\text{field is not supplemented}|\text{field is used in year k}) * P(\text{field is used in y}}{P(\text{field is not supplemented})} \\
\text{where } k &= 1, 2
\end{aligned}
$$

1. What are the parameters you are estimating to answer this question?

A1: The proportion of fields used in a given year of the study given the field is supplemented or not.

2. What is the likelihood you are using for the data?

A2:

$$
\begin{aligned}
P(\text{field is supplemented}|\text{field is used in year k}) \\
P(\text{field is not supplemented}|\text{field is used in year k})
\end{aligned}
$$

3. What prior distributions are you using?

A3: Uniform$(0, 1)$

4. What does the model you are fitting look like when written in model syntax (see below) and in R code?

A4: Let U_k be a random variable representing the total number of fields in a trial that experienced greater than two "crane nights" in year k, where $k = 1, 2$. Let p_k^s denote the proportion of either supplemented, p_k^1, or not supplemented, p_k^0, fields.

$$\text{Using a uniform prior:} \begin{aligned} U_k &\sim \text{Binomial}(N = 15, p_k^s) \\ p_k^s &\sim \text{Uniform}(0, 1) \end{aligned}$$

4. How are you choosing to generate the posterior distribution (e.g., grid approximation vs. quadratic approximation)?

A4: Grid approximation

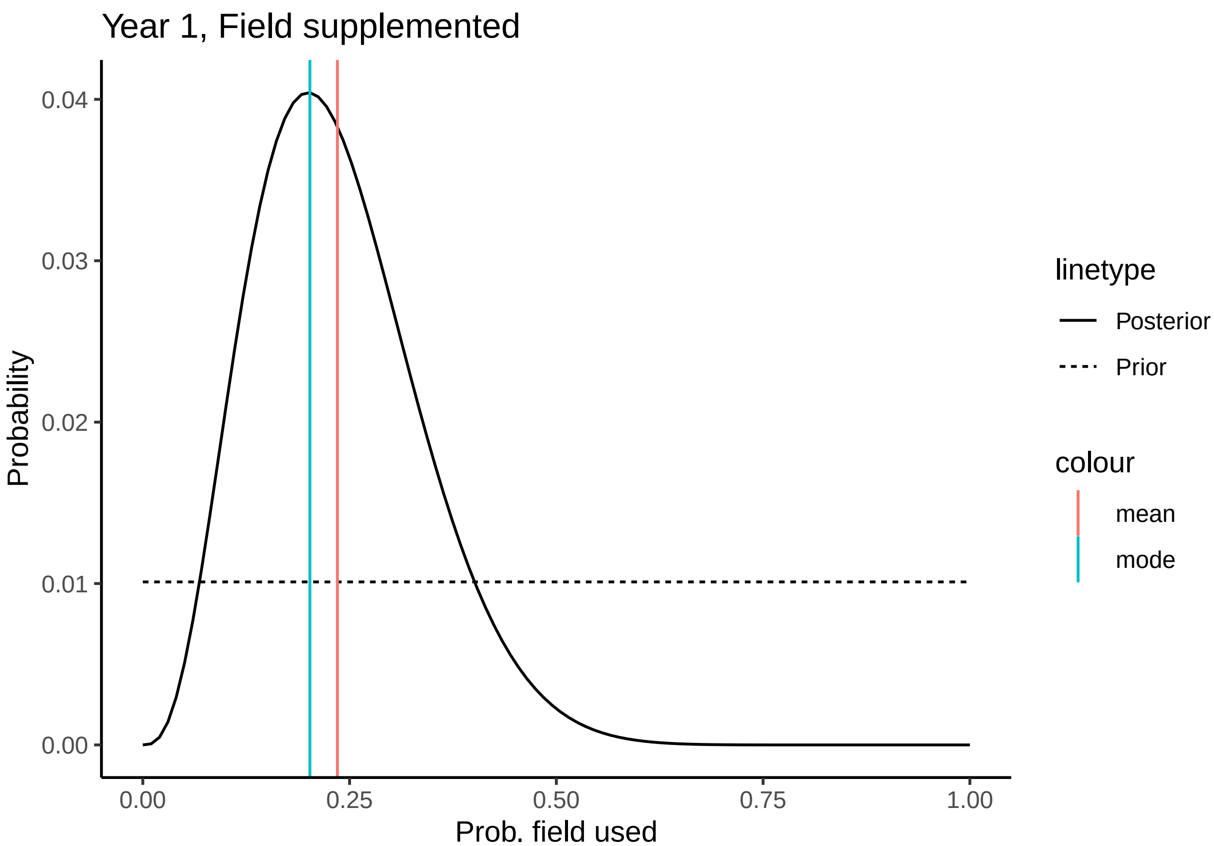
```r
## Extracting information from CSV

y1_supp = data[1,2]
#y1_supp
y1_no_supp = data[2,2]
#y1_no_supp
y2_supp = data[1,3]
#y2_supp
y2_no_supp = data[2,3]
#y2_no_supp

N = 15 #number or trials (fields)


extracted_data = list(y1_supp, y1_no_supp, y2_supp, y2_no_supp)
```

```r
## Grid approximation (with uniform(0,1) as the prior)
library(ggplot2)
# 1. Define grid
p_grid = seq(0, 1, len=100)
dp = p_grid[2] - p_grid[1]

# 2. Compute prior
prior = dunif(p_grid, 0, 1)


posteriors = list()
mean_p = list()
mode_p = list()

for (i in seq_along(extracted_data)) {

  # 3. Compute likelihood
  element <- dbinom(extracted_data[[i]], size=N, prob=p_grid)

  # 4. Compute unnormalized posterior
  unnorm_posterior <- prior * element

  # 5. Normalize posterior to "integrate" to one
```

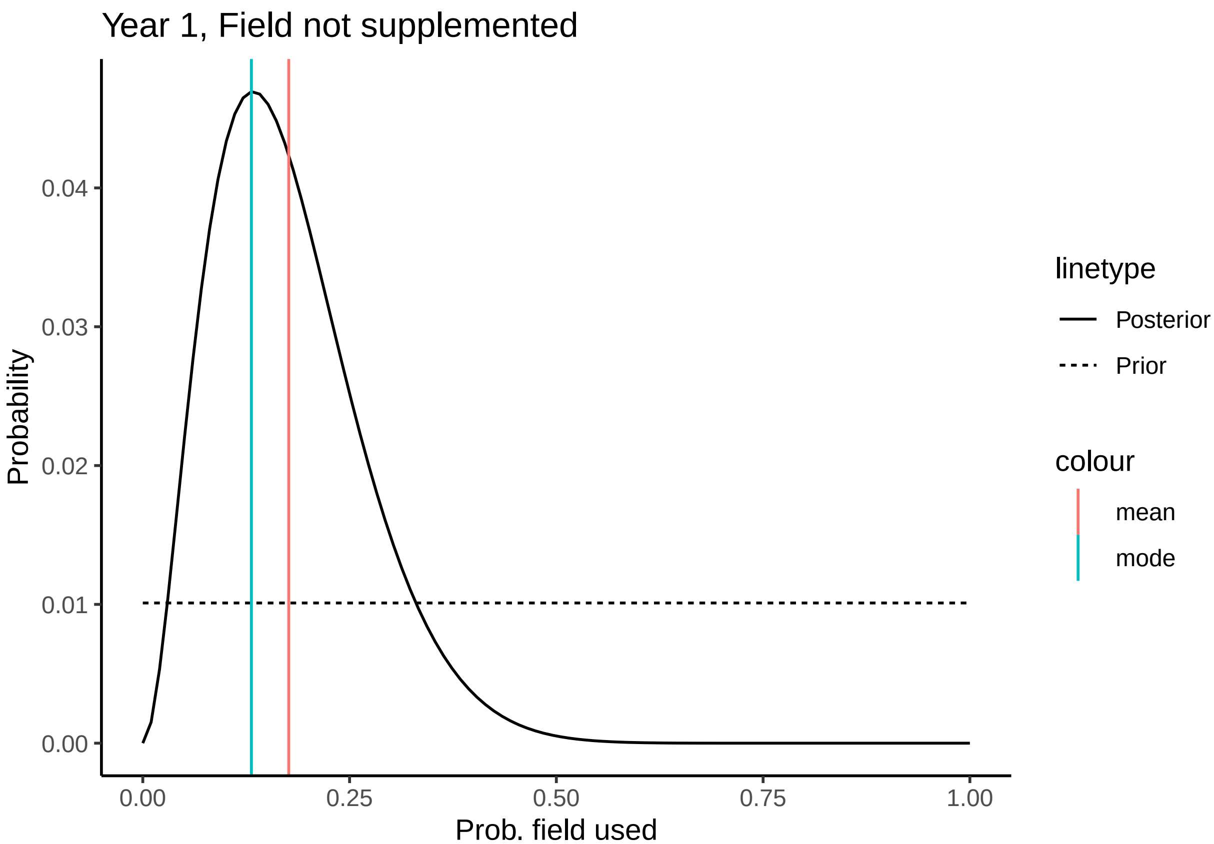
```r
  posterior <- unnorm_posterior / sum(unnorm_posterior)

  # Add prior to list of priors
  posteriors[[i]] <- posterior

  # Computing the means and modes
  mean_p[[i]] <- sum(posterior * p_grid)
  mode_p[[i]] <- p_grid[which.max(posterior)]

}
#posteriors
```

Plotting the posteriors

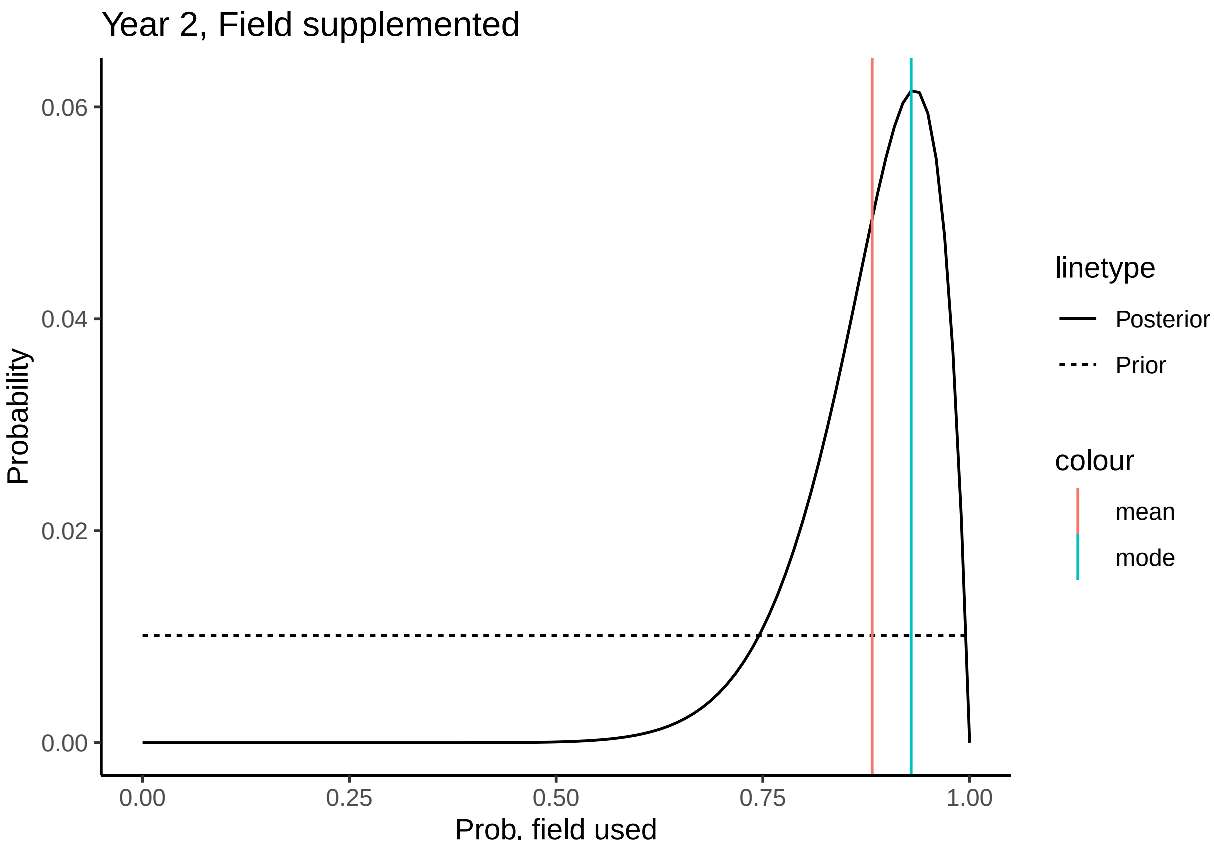
```r
# Year 1, Field supplemented
p1 = ggplot() +
      geom_line(aes(x=p_grid, y=posteriors[[1]], linetype="Posterior")) +
          geom_line(aes(x=p_grid, y=prior*dp, linetype="Prior")) +
          ylab("Probability") + xlab("Prob. field used") +
      ggtitle("Year 1, Field supplemented")+
          theme_classic()

# ggsave("inference1.pdf", width=4, height=3)

p1 + geom_vline(aes(xintercept=mean_p[[1]], color="mean")) +
     geom_vline(aes(xintercept=mode_p[[1]], color="mode"))
```



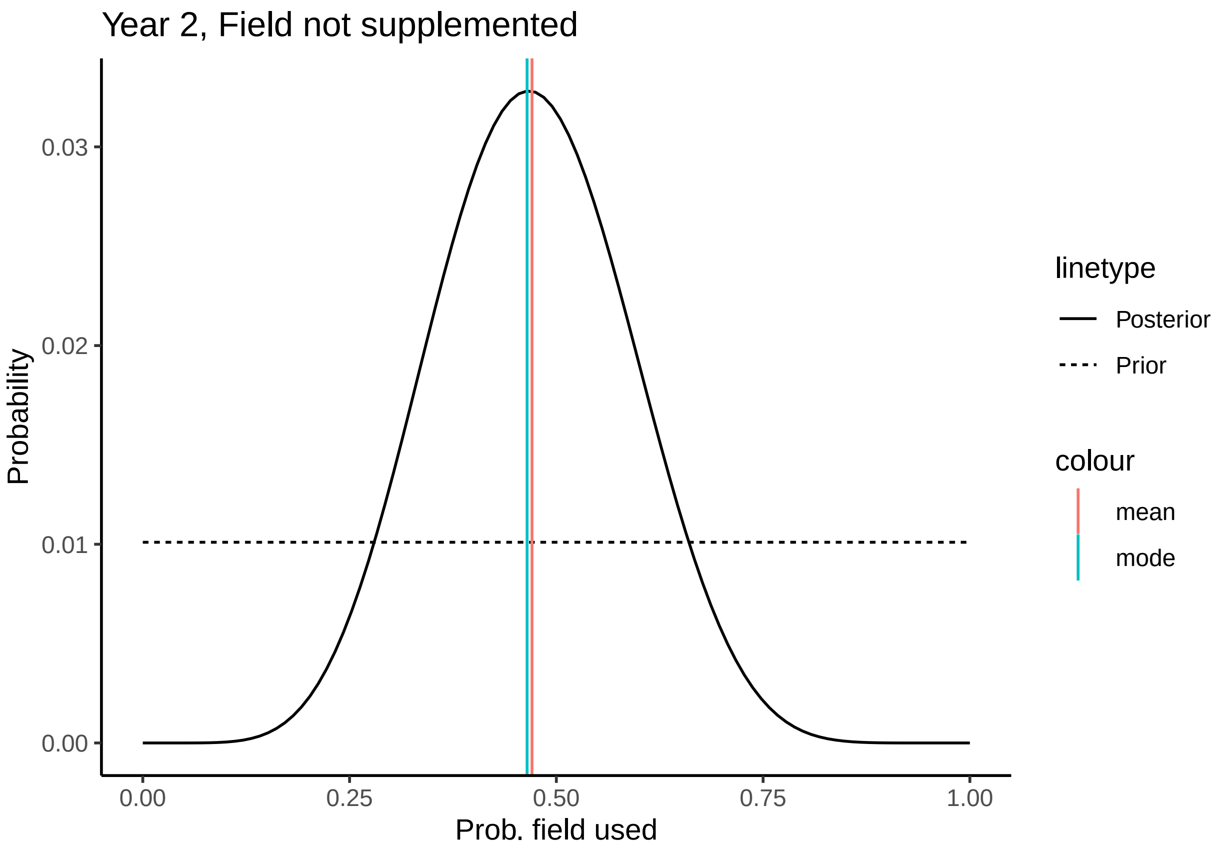
```r
# Year 1, Field not supplemented
p1 = ggplot() +
      geom_line(aes(x=p_grid, y=posteriors[[2]], linetype="Posterior")) +
          geom_line(aes(x=p_grid, y=prior*dp, linetype="Prior")) +
          ylab("Probability") + xlab("Prob. field used") +
      ggtitle("Year 1, Field not supplemented")+
          theme_classic()

# ggsave("inference1.pdf", width=4, height=3)

p1 + geom_vline(aes(xintercept=mean_p[[2]], color="mean")) +
     geom_vline(aes(xintercept=mode_p[[2]], color="mode"))
```




```r
# Year 2, Field supplemented
p1 = ggplot() +
      geom_line(aes(x=p_grid, y=posteriors[[3]], linetype="Posterior")) +
          geom_line(aes(x=p_grid, y=prior*dp, linetype="Prior")) +
          ylab("Probability") + xlab("Prob. field used") +
      ggtitle("Year 2, Field supplemented")+
          theme_classic()

# ggsave("inference1.pdf", width=4, height=3)

p1 + geom_vline(aes(xintercept=mean_p[[3]], color="mean")) +
     geom_vline(aes(xintercept=mode_p[[3]], color="mode"))
```



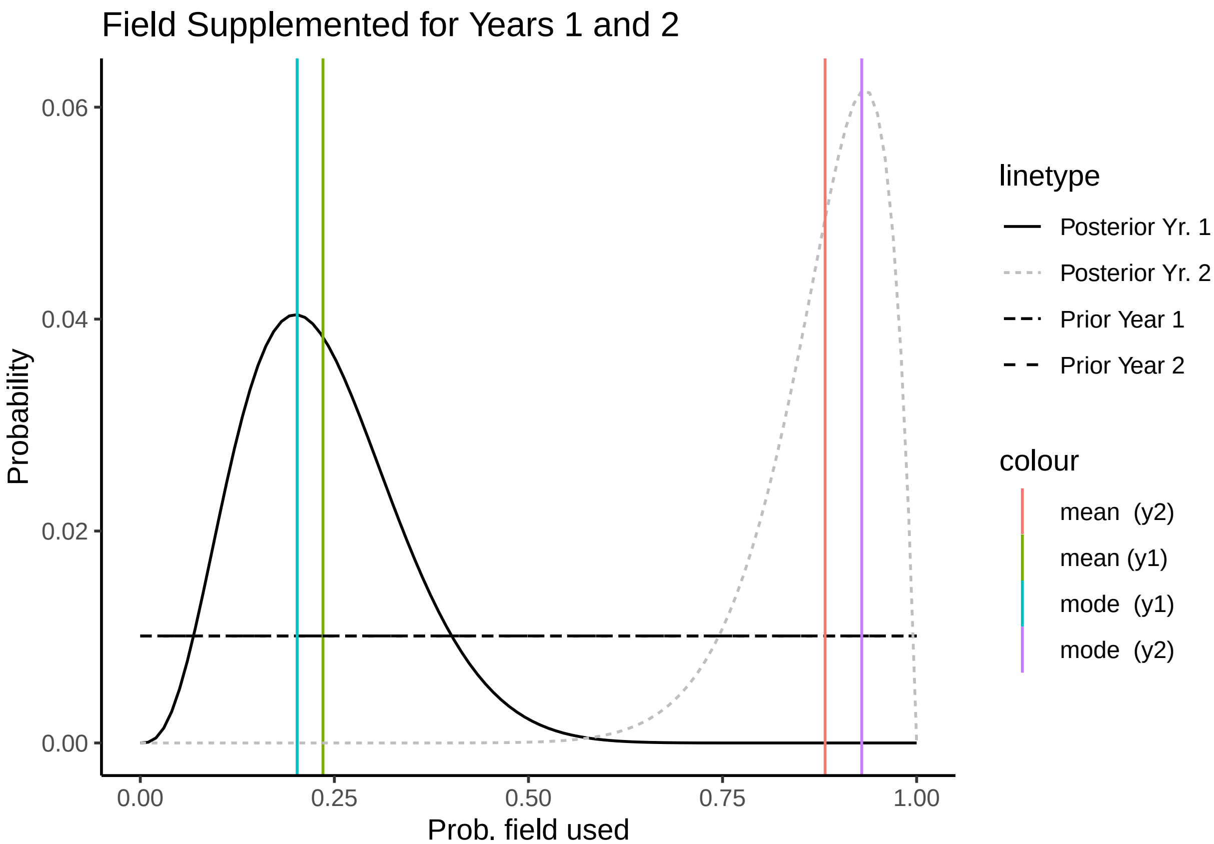

```r
# Year 2, Field not supplemented
p1 = ggplot() +
      geom_line(aes(x=p_grid, y=posteriors[[4]], linetype="Posterior")) +
          geom_line(aes(x=p_grid, y=prior*dp, linetype="Prior")) +
          ylab("Probability") + xlab("Prob. field used") +
      ggtitle("Year 2, Field not supplemented")+
          theme_classic()

# ggsave("inference1.pdf", width=4, height=3)

p1 + geom_vline(aes(xintercept=mean_p[[4]], color="mean")) +
     geom_vline(aes(xintercept=mode_p[[4]], color="mode"))
```



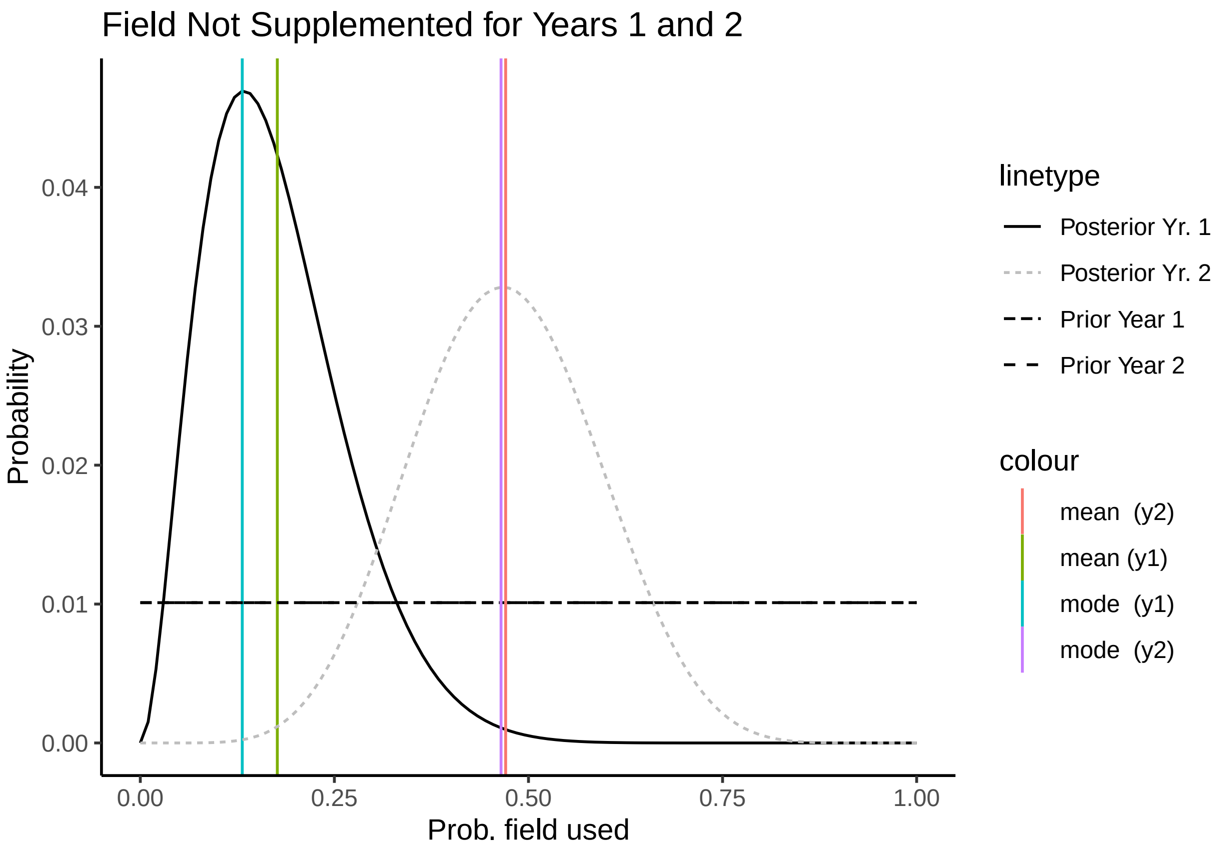
5. How are you comparing posterior distributions to answer the question of interest?

A5:

```
# Field supplemented

p1 = ggplot() +
      geom_line(aes(x=p_grid, y=posteriors[[1]], linetype="Posterior Yr. 1"), color = "black") +
          geom_line(aes(x=p_grid, y=prior*dp, linetype="Prior Year 1")) +
      geom_vline(aes(xintercept=mean_p[[1]], color="mean (y1)")) +
        geom_vline(aes(xintercept=mode_p[[1]], color="mode  (y1)"))+
      geom_line(aes(x=p_grid, y=posteriors[[3]], linetype="Posterior Yr. 2"), color = "gray") +
          geom_line(aes(x=p_grid, y=prior*dp, linetype="Prior Year 2")) +
          ylab("Probability") + xlab("Prob. field used") +
      ggtitle("Field Supplemented for Years 1 and 2")+
      geom_vline(aes(xintercept=mean_p[[3]], color="mean  (y2)")) +
        geom_vline(aes(xintercept=mode_p[[3]], color="mode  (y2)"))+
      theme_classic()

p1
```




```r
# Field not supplemented

p1 = ggplot() +
      geom_line(aes(x=p_grid, y=posteriors[[2]], linetype="Posterior Yr. 1"), color = "black") +
          geom_line(aes(x=p_grid, y=prior*dp, linetype="Prior Year 1")) +
      geom_vline(aes(xintercept=mean_p[[2]], color="mean (y1)")) +
        geom_vline(aes(xintercept=mode_p[[2]], color="mode  (y1)"))+
      geom_line(aes(x=p_grid, y=posteriors[[4]], linetype="Posterior Yr. 2"), color = "gray") +
          geom_line(aes(x=p_grid, y=prior*dp, linetype="Prior Year 2")) +
          ylab("Probability") + xlab("Prob. field used") +
      ggtitle("Field Not Supplemented for Years 1 and 2")+
      geom_vline(aes(xintercept=mean_p[[4]], color="mean  (y2)")) +
        geom_vline(aes(xintercept=mode_p[[4]], color="mode  (y2)"))+
      theme_classic()

p1
```



#

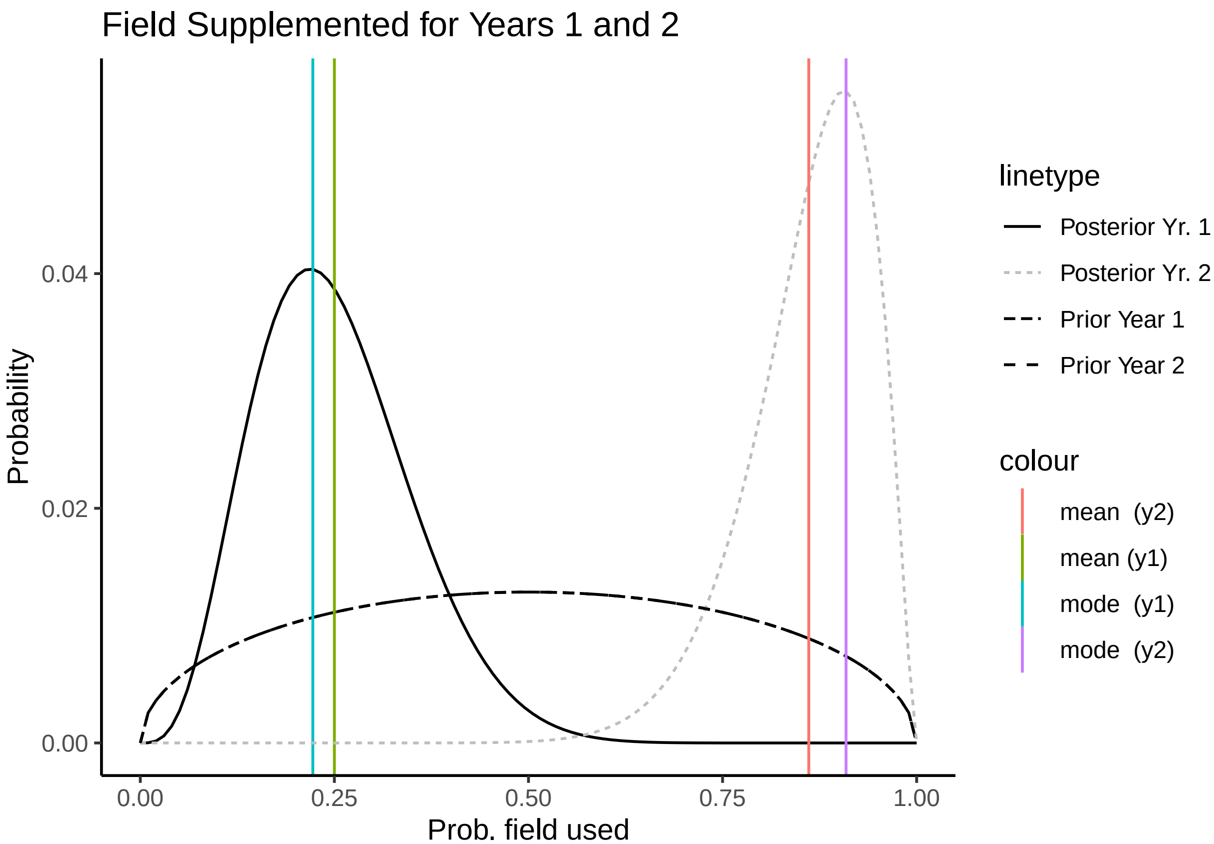
Reconsidering the model using a Beta distribution as the prior:

$$U_k \sim \text{Binomial}(N, p_k^s)$$
$$p_k^s \sim \text{Beta}(0.5 * 3, (1 - 0.5) * 3)$$

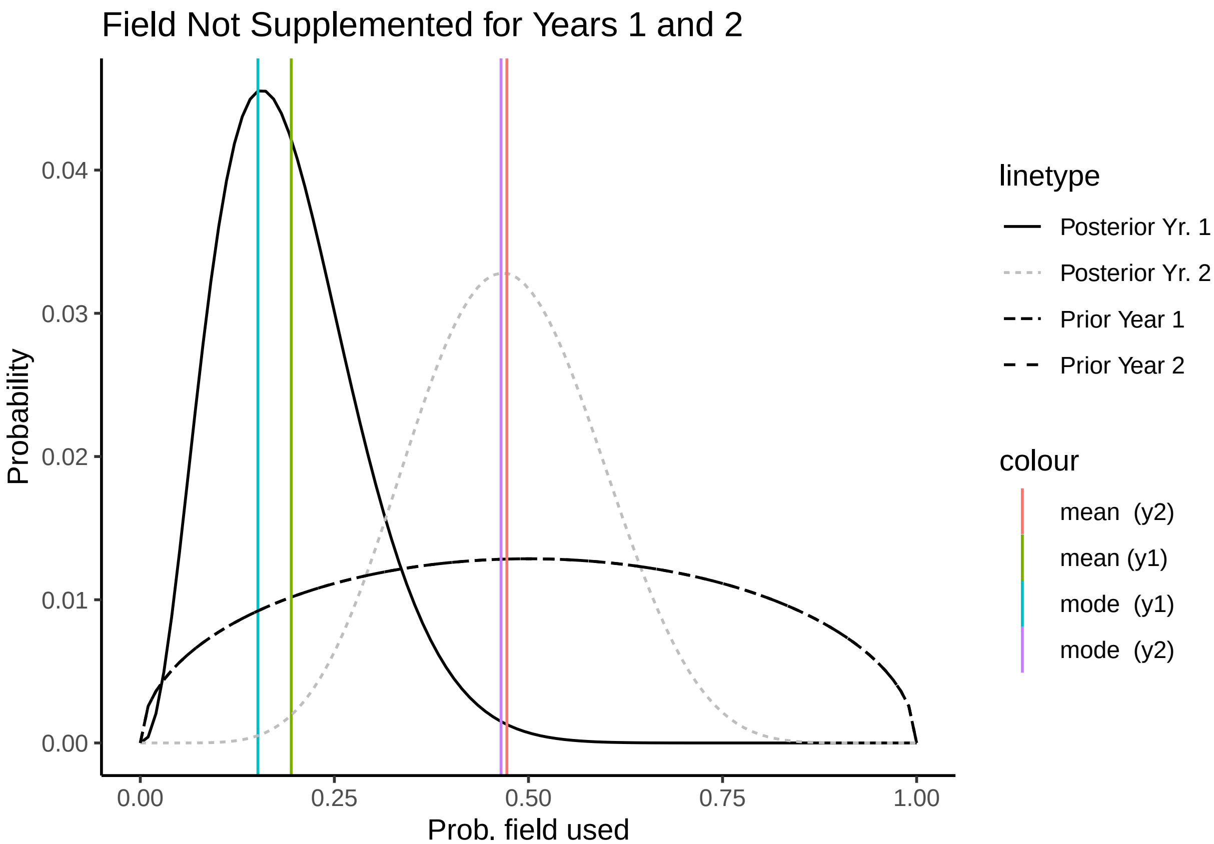
```
## Grid approximation (with Beta(0,1) as the prior)


# 1. Define grid
p_grid = seq(0, 1, len=100)
dp = p_grid[2] - p_grid[1]

# 2. Compute prior
mu = 0.5 # mean
phi = 3 # precision
a = mu*phi
b = (1 - mu)*phi
prior_beta = dbeta(p_grid, a, b)


posteriors_beta = list()
mean_p_beta = list()
mode_p_beta = list()

for (i in seq_along(extracted_data)) {

  # 3. Compute likelihood
```

```
  element <- dbinom(extracted_data[[i]], size=N, prob=p_grid)

  # 4. Compute unnormalized posterior
  unnorm_posterior <- prior_beta * element

  # 5. Normalize posterior to "integrate" to one
  posterior <- unnorm_posterior / sum(unnorm_posterior)

  # Add prior to list of priors
  posteriors_beta[[i]] <- posterior

  # Computing the means and modes
  mean_p_beta[[i]] <- sum(posterior * p_grid)
  mode_p_beta[[i]] <- p_grid[which.max(posterior)]

}
#posteriors
```

```
# Field supplemented; beta prior

p1 = ggplot() +
      geom_line(aes(x=p_grid, y=posteriors_beta[[1]], linetype="Posterior Yr. 1"), color = "black") +
          geom_line(aes(x=p_grid, y=prior_beta*dp, linetype="Prior Year 1")) +
      geom_vline(aes(xintercept=mean_p_beta[[1]], color="mean (y1)")) +
        geom_vline(aes(xintercept=mode_p_beta[[1]], color="mode  (y1)"))+
      geom_line(aes(x=p_grid, y=posteriors_beta[[3]], linetype="Posterior Yr. 2"), color = "gray") +
          geom_line(aes(x=p_grid, y=prior_beta*dp, linetype="Prior Year 2")) +
          ylab("Probability") + xlab("Prob. field used") +
      ggtitle("Field Supplemented for Years 1 and 2")+
      geom_vline(aes(xintercept=mean_p_beta[[3]], color="mean  (y2)")) +
        geom_vline(aes(xintercept=mode_p_beta[[3]], color="mode  (y2)"))+
      theme_classic()

p1
```




```r
# Field not supplemented; Beta prior

p1 = ggplot() +
      geom_line(aes(x=p_grid, y=posteriors_beta[[2]], linetype="Posterior Yr. 1"), color = "black") +
          geom_line(aes(x=p_grid, y=prior_beta*dp, linetype="Prior Year 1")) +
      geom_vline(aes(xintercept=mean_p_beta[[2]], color="mean (y1)")) +
        geom_vline(aes(xintercept=mode_p_beta[[2]], color="mode  (y1)"))+
      geom_line(aes(x=p_grid, y=posteriors[[4]], linetype="Posterior Yr. 2"), color = "gray") +
          geom_line(aes(x=p_grid, y=prior_beta*dp, linetype="Prior Year 2")) +
          ylab("Probability") + xlab("Prob. field used") +
      ggtitle("Field Not Supplemented for Years 1 and 2")+
      geom_vline(aes(xintercept=mean_p_beta[[4]], color="mean  (y2)")) +
        geom_vline(aes(xintercept=mode_p_beta[[4]], color="mode  (y2)"))+
      theme_classic()

p1
```




```r
print(mode_p)
```

```
## [[1]]
## [1] 0.2020202
##
## [[2]]
## [1] 0.1313131
##
## [[3]]
## [1] 0.9292929
##
## [[4]]
## [1] 0.4646465
```

```r
print(mode_p_beta)
```

```
## [[1]]
## [1] 0.2222222
##
## [[2]]
## [1] 0.1515152
##
## [[3]]
## [1] 0.9090909
##
## [[4]]
## [1] 0.4646465
```

```r
print(mode_p[[3]] - mode_p[[1]])
```

```
## [1] 0.7272727
```

```r
print(mode_p_beta[[3]] - mode_p_beta[[1]])
```

```
## [1] 0.6868687
```

```r
print(mode_p[[4]] - mode_p[[2]])
```

```
## [1] 0.3333333
```

```r
print(mode_p_beta[[4]] - mode_p_beta[[2]])
```

```
## [1] 0.3131313
```

Conclusion:

We can see from the plots (with both the uniform and beta priors) that the probability of the cranes using the fields increases whether or not the field is supplemented. If the fields used for this experiment were relatively close to each other, the cranes could have seen the entire area as having a better food supply once they discovered the supplementary food in the first year and revisited the area as a whole in search of food.

With both priors, the MAP increases dramatically between years 1 and 2 in the supplemented fields:

Experiment	Unif. Prior	Beta Prior
Supp. Field	0.7272727	0.6868687
No Supp.	0.3333333	0.3131313