The Lady Tasting Wine

Joachim Vandekerckhove

The Lady Tasting Wine

The second lady is a Master of Wine

- Claims to be able to discriminate between French Bordeaux wine and a Californian Cabernet Sauvignon, Merlot blend

The Lady Tasting Wine

The second lady is a Master of Wine

- Claims to be able to discriminate between French Bordeaux wine and a Californian Cabernet Sauvignon, Merlot blend
- She is put to a similar test

The Lady Tasting Wine (RRRRRW)

- The same data

The Lady Tasting Wine (RRRRRW)

- The same data
  - RRRRRW, so the likelihood is again $P_R^5(1 - P_R) \times C$

The Lady Tasting Wine (RRRRRW)

- The same data
  - RRRRRW, so the likelihood is again $P_R^5(1 - P_R) \times C$
- But different prior notions

The Lady Tasting Wine (RRRRRW)

- The same data
  - RRRRRW, so the likelihood is again $P_R^5(1 - P_R) \times C$
- But different prior notions
  - One might believe that Masters of Wine can distinguish the Californian imitation from the French original, so that $P_R \geq 0.5$

The Lady Tasting Wine (RRRRRW)

- The same data
  - RRRRRW, so the likelihood is again $P_R^5(1 - P_R) \times C$
- But different prior notions
  - One might believe that Masters of Wine can distinguish the Californian imitation from the French original, so that $P_R \geq 0.5$
  - ... while at the same time doubting that ladies can distinguish the two methods of teamaking, so that $P_R = 0.5$ is most likely (but $P_R > 0.5$ is possible)

The Lady Tasting Wine (RRRRRW)

- The same data
  - RRRRRW, so the likelihood is again $P_R^5(1 - P_R) \times C$
- But different prior notions
  - One might believe that Masters of Wine can distinguish the Californian imitation from the French original, so that $P_R \geq 0.5$
  - ... while at the same time doubting that ladies can distinguish the two methods of teamaking, so that $P_R = 0.5$ is most likely (but $P_R > 0.5$ is possible)
- It is possible for you to disagree and still be sensible

The Lady Tasting Wine (RRRRRW)

- Tea is different from wine

The Lady Tasting Wine (RRRRRW)

- Tea is different from wine
    - This is relevant and useful prior information

The Lady Tasting Wine (RRRRRW)

- Tea is different from wine
  - This is relevant and useful prior information
  - We can assume one prior for the case of wine

The Lady Tasting Wine (RRRRRW)

- Tea is different from wine
  - This is relevant and useful prior information
  - We can assume one prior for the case of wine
    - $p(P_R) = K_w(1.01 - P_R)(P_R - 0.49)$, for $0.5 \leq P_R \leq 1$

The Lady Tasting Wine (RRRRRW)

- Tea is different from wine
  - This is relevant and useful prior information
  - We can assume one prior for the case of wine
    - $p(P_R) = K_w(1.01 - P_R)(P_R - 0.49)$, for $0.5 \leq P_R \leq 1$
  - ... and another for the case of tea:

The Lady Tasting Wine (RRRRRW)

- Tea is different from wine
  - This is relevant and useful prior information
  - We can assume one prior for the case of wine
    - $p(P_R) = K_w(1.01 - P_R)(P_R - 0.49)$, for $0.5 \leq P_R \leq 1$
  - ... and another for the case of tea:
    - $p(P_R) = 0.8$ for $P_R = 0.5$

The Lady Tasting Wine (RRRRRW)

- Tea is different from wine
  - This is relevant and useful prior information
  - We can assume one prior for the case of wine
    - $p(P_R) = K_w(1.01 - P_R)(P_R - 0.49)$, for $0.5 \leq P_R \leq 1$
  - ... and another for the case of tea:
    - $p(P_R) = 0.8$ for $P_R = 0.5$
    - $p(P_R) = K_t(1 - P_R)$ for $P_R > 0.5$

The Lady Tasting Wine (RRRRRW)

- Tea is different from wine
  - This is relevant and useful prior information
  - We can assume one prior for the case of wine
    - $p(P_R) = K_w(1.01 - P_R)(P_R - 0.49)$, for $0.5 \leq P_R \leq 1$
  - ... and another for the case of tea:
    - $p(P_R) = 0.8$ for $P_R = 0.5$
    - $p(P_R) = K_t(1 - P_R)$ for $P_R > 0.5$
- K_t and K_w are chosen such that the sum (or integral) over all possibilities is 1. This is always possible if the distribution is proper. The solution for wine here is easy enough (it is the sum of $(1 - P_R)(P_R - 0.5)$ for all values of P_R), but it isn't in general

The Lady Tasting Wine (RRRRRW)

- Exercise: Compute and plot these in R

The Lady Tasting Wine (RRRRRW)

- Exercise: Compute and plot these in R
  - Compute these priors for every value of P_R and plot them side-by-side

The Lady Tasting Wine (RRRRRW)

- Exercise: Compute and plot these in R
  - Compute these priors for every value of P_R and plot them side-by-side
    - Use $P_R \in \{0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 1.0\}$

The Lady Tasting Wine (RRRRRW)

- Exercise: Compute and plot these in R
  - Compute these priors for every value of P_R and plot them side-by-side
    - Use $P_R \in \{0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 1.0\}$
  - Wine: $p(P_R) = K_w(1.01 - P_R)(P_R - 0.49)$, for $0.5 \leq P_R \leq 1$

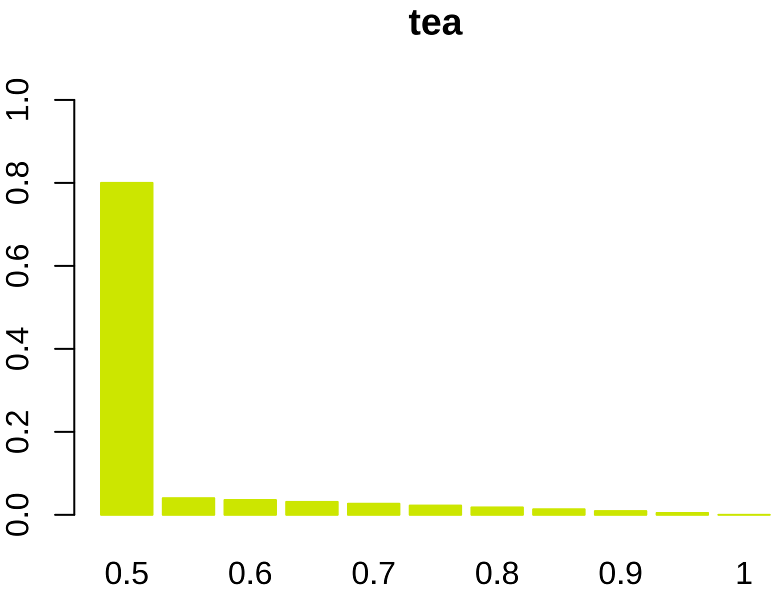
The Lady Tasting Wine (RRRRRW)

- Exercise: Compute and plot these in R
  - Compute these priors for every value of P_R and plot them side-by-side
    - Use $P_R \in \{0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 1.0\}$
  - Wine: $p(P_R) = K_w(1.01 - P_R)(P_R - 0.49)$, for $0.5 \leq P_R \leq 1$
  - Tea: $p(P_R) = 0.8$ for $P_R = 0.5$ and $p(P_R) = K_t(1 - P_R)$ for $P_R > 0.5$

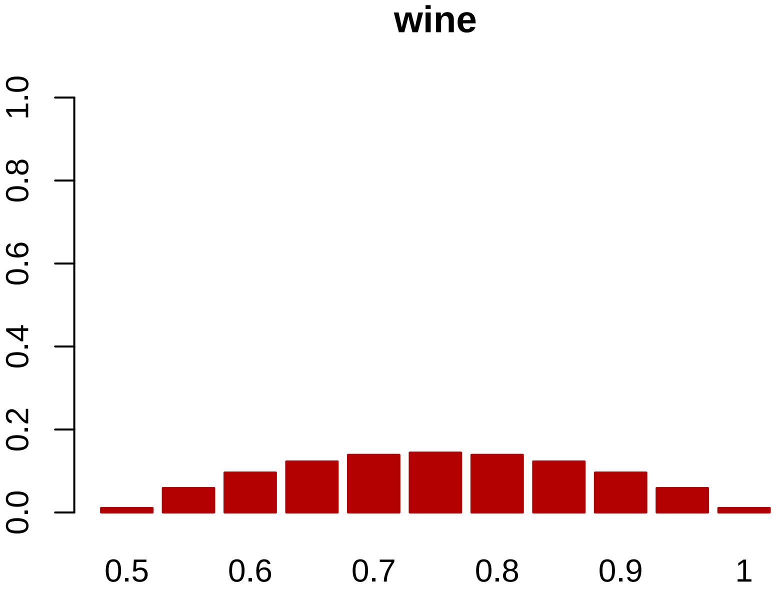
The Lady Tasting Wine (RRRRRW)

- Exercise: Compute and plot these in R
  - Compute these priors for every value of P_R and plot them side-by-side
    - Use $P_R \in \{0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 1.0\}$
  - Wine: $p(P_R) = K_w(1.01 - P_R)(P_R - 0.49)$, for $0.5 \leq P_R \leq 1$
  - Tea: $p(P_R) = 0.8$ for $P_R = 0.5$ and $p(P_R) = K_t(1 - P_R)$ for $P_R > 0.5$
- K_* is the inverse of the sum of everything else over values of P_R

The Lady Tasting Wine (RRRRRW)



The Lady Tasting Wine (RRRRRW)



The Lady Tasting Wine (RRRRRW)

This was an exercise in normalization:

- From a set of values that give us the relative probabilities of possible values of P_R, find the absolute probabilities by enforcing that all probabilities must add up to 1 (the possible values of P_R form a disjoint set)

The Lady Tasting Wine (RRRRRW)

This was an exercise in normalization:

- From a set of values that give us the relative probabilities of possible values of P_R, find the absolute probabilities by enforcing that all probabilities must add up to 1 (the possible values of P_R form a disjoint set)
- Often we can just deal with relative probabilities for convenience

The Lady Tasting Wine (RRRRRW)

This was an exercise in normalization:

- From a set of values that give us the relative probabilities of possible values of P_R, find the absolute probabilities by enforcing that all probabilities must add up to 1 (the possible values of P_R form a disjoint set)
- Often we can just deal with relative probabilities for convenience
- Bayesians will routinely ignore these scaling values and only express equations using the "proportional to" symbol:

$$P(H_0|X) \propto P(H_0)P(X|H_0)$$

The Lady Tasting Wine (RRRRRW)

This was an exercise in normalization:

- From a set of values that give us the relative probabilities of possible values of P_R, find the absolute probabilities by enforcing that all probabilities must add up to 1 (the possible values of P_R form a disjoint set)
- Often we can just deal with relative probabilities for convenience
- Bayesians will routinely ignore these scaling values and only express equations using the "proportional to" symbol:

$$P(H_0|X) \propto P(H_0)P(X|H_0)$$

- $((A \propto B), (B \propto C)) \rightarrow (A \propto C)$

The Lady Tasting Wine (RRRRRW)

This was an exercise in normalization:

- From a set of values that give us the relative probabilities of possible values of P_R, find the absolute probabilities by enforcing that all probabilities must add up to 1 (the possible values of P_R form a disjoint set)
- Often we can just deal with relative probabilities for convenience
- Bayesians will routinely ignore these scaling values and only express equations using the "proportional to" symbol:

$$P(H_0|X) \propto P(H_0)P(X|H_0)$$

- $((A \propto B), (B \propto C)) \rightarrow (A \propto C)$
- I usually make the proportionality explicit to avoid confusion

The Lady Tasting Wine (RRRRRW)

Reshaping our posterior: $p(\theta|x) = Kp(\theta)p(x|\theta)$

The Lady Tasting Wine (RRRRRW)

Reshaping our posterior: $p(\theta|x) = Kp(\theta)p(x|\theta)$

- Wine: $p(P_R|\#R, \#W, wine) =$
  $S_w \times K_w(1.01 - P_R)(P_R - 0.49) \times C(1 - P_R)^{\#W} P_R^{\#R}$

The Lady Tasting Wine (RRRRRW)

Reshaping our posterior: $p(\theta|x) = Kp(\theta)p(x|\theta)$

- Wine: $p(P_R|\#R, \#W, wine) =$
  $S_w \times K_w(1.01 - P_R)(P_R - 0.49) \times C(1 - P_R)^{\#W} P_R^{\#R}$
- Tea:

The Lady Tasting Wine (RRRRRW)

Reshaping our posterior: $p(\theta|x) = Kp(\theta)p(x|\theta)$

- Wine: $p(P_R|\#R, \#W, wine) =$
  $S_w \times K_w(1.01 - P_R)(P_R - 0.49) \times C(1 - P_R)^{\#W} P_R^{\#R}$
- Tea:
  - $p(P_R|\#R, \#W, tea) = S_t \times 0.8 \times C(1 - P_R)^{\#W} P_R^{\#R}$ if $P_R = 0.5$

The Lady Tasting Wine (RRRRRW)

Reshaping our posterior: $p(\theta|x) = Kp(\theta)p(x|\theta)$

- Wine: $p(P_R|\#R, \#W, wine) =$
  $S_w \times K_w(1.01 - P_R)(P_R - 0.49) \times C(1 - P_R)^{\#W} P_R^{\#R}$
- Tea:
  - $p(P_R|\#R, \#W, tea) = S_t \times 0.8 \times C(1 - P_R)^{\#W} P_R^{\#R}$ if $P_R = 0.5$
  - $p(P_R|\#R, \#W, tea) = S_t \times K_t(1 - P_R) \times C(1 - P_R)^{\#W} P_R^{\#R}$ if $P_R > .05$

The Lady Tasting Wine (RRRRRW)

Exercise: plot these in R, using $\#R = 5$, $\#W = 1$

- Wine: $p(P_R|\#R, \#W, wine) =$
  $S_w \times K_w(1.01 - P_R)(P_R - 0.49) \times C(1 - P_R)^{\#W} P_R^{\#R}$

The Lady Tasting Wine (RRRRRW)

Exercise: plot these in R, using $\#R = 5$, $\#W = 1$

- Wine: $p(P_R|\#R, \#W, wine) =$
  $S_w \times K_w(1.01 - P_R)(P_R - 0.49) \times C(1 - P_R)^{\#W} P_R^{\#R}$
- Tea:

The Lady Tasting Wine (RRRRRW)

Exercise: plot these in R, using $\#R = 5$, $\#W = 1$

- Wine: $p(P_R|\#R, \#W, wine) =$
  $S_w \times K_w(1.01 - P_R)(P_R - 0.49) \times C(1 - P_R)^{\#W} P_R^{\#R}$
- Tea:
  - $p(P_R|\#R, \#W, tea) = S_t \times 0.8 \times C(1 - P_R)^{\#W} P_R^{\#R}$ if $P_R = 0.5$

The Lady Tasting Wine (RRRRRW)

Exercise: plot these in R, using $\#R = 5$, $\#W = 1$

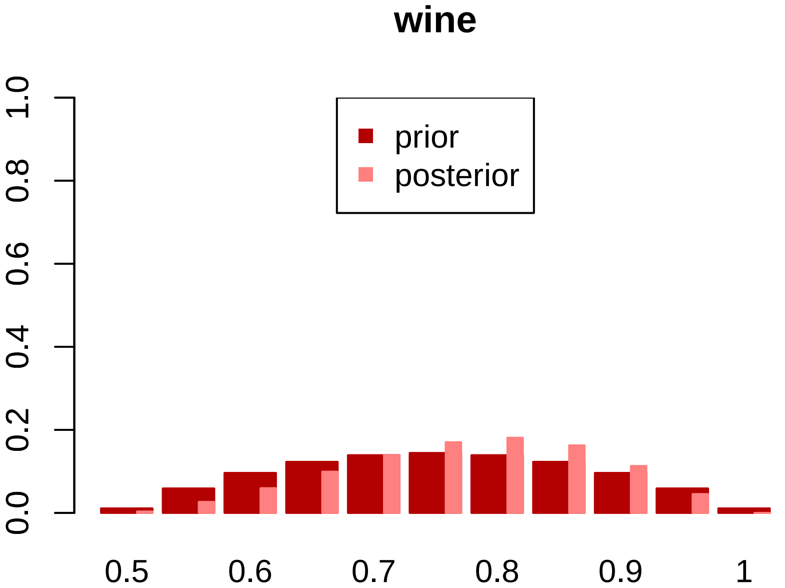
- Wine: $p(P_R|\#R, \#W, wine) = S_w \times K_w(1.01 - P_R)(P_R - 0.49) \times C(1 - P_R)^{\#W} P_R^{\#R}$
- Tea:
  - $p(P_R|\#R, \#W, tea) = S_t \times 0.8 \times C(1 - P_R)^{\#W} P_R^{\#R}$ if $P_R = 0.5$
  - $p(P_R|\#R, \#W, tea) = S_t \times K_t(1 - P_R) \times C(1 - P_R)^{\#W} P_R^{\#R}$ if $P_R > .05$

The Lady Tasting Wine (RRRRRW)

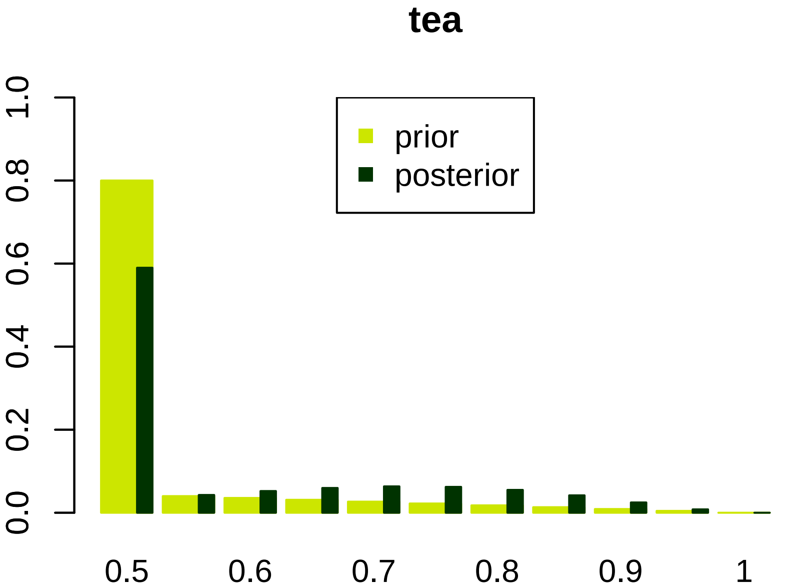
Exercise: plot these in R, using $\#R = 5$, $\#W = 1$

- Wine: $p(P_R|\#R, \#W, wine) =$
  $S_w \times K_w(1.01 - P_R)(P_R - 0.49) \times C(1 - P_R)^{\#W} P_R^{\#R}$
- Tea:
  - $p(P_R|\#R, \#W, tea) = S_t \times 0.8 \times C(1 - P_R)^{\#W} P_R^{\#R}$ if $P_R = 0.5$
  - $p(P_R|\#R, \#W, tea) = S_t \times K_t(1 - P_R) \times C(1 - P_R)^{\#W} P_R^{\#R}$ if $P_R > .05$
- Also, make them pretty.

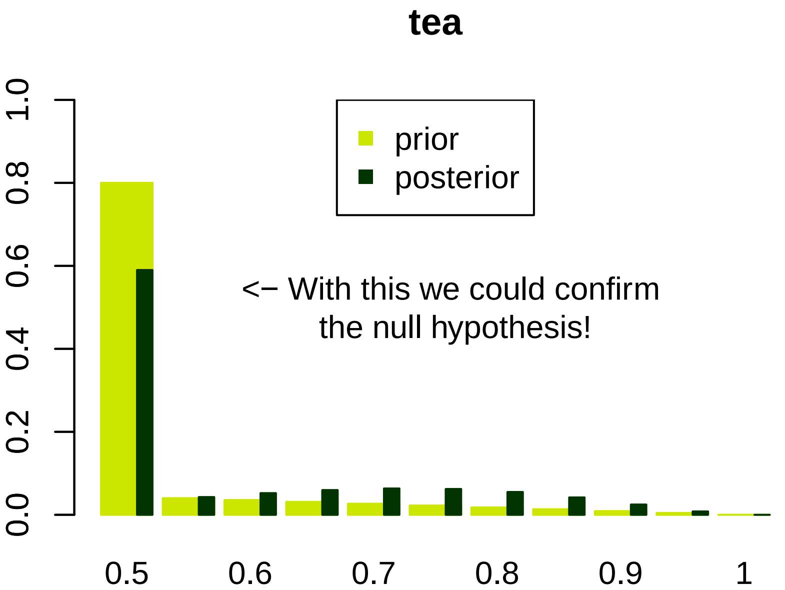
The Lady Tasting Wine (RRRRRW)



The Lady Tasting Wine (RRRRRW)



The Lady Tasting Wine (RRRRRW)



The Lady Tasting Wine

Updating the old posterior with new data:

- Suppose our two ladies return with a renewed thirst.

The Lady Tasting Wine

Updating the old posterior with new data:

- Suppose our two ladies return with a renewed thirst.
- They get $\#R_2 = 44$ corrects and $\#W_2 = 0$ errors

The Lady Tasting Wine

Updating the old posterior with new data:

- Suppose our two ladies return with a renewed thirst.
- They get $\#R_2 = 44$ corrects and $\#W_2 = 0$ errors
- Update the posterior for the lady tasting wine by multiplying the old posterior with the new data:

$$
\begin{aligned}
& p(P_R \mid \#R, \#W, \#R_2, \#W_2, wine) = \\
& \quad S_w \times K_w(1.01 - P_R)(P_R - 0.49) \\
& \qquad \times C(1 - P_R)^{\#W} P_R^{\#R} \times C_2(1 - P_R)^{\#W_2} P_R^{\#R_2}
\end{aligned}
$$

The Lady Tasting Wine

Updating the old posterior with new data:

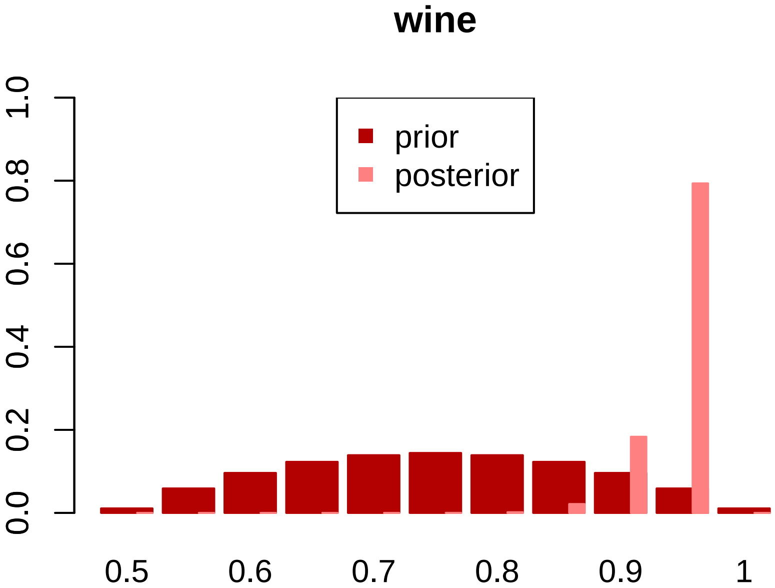
- Suppose our two ladies return with a renewed thirst.
- They get $\#R_2 = 44$ corrects and $\#W_2 = 0$ errors
- Update the posterior for the lady tasting wine by multiplying the old posterior with the new data:

$$
\begin{aligned}
&p(P_R|\#R, \#W, \#R_2, \#W_2, wine) = \\
&\quad S_w \times K_w(1.01 - P_R)(P_R - 0.49) \\
&\qquad \times C(1 - P_R)^{\#W} P_R^{\#R} \times C_2(1 - P_R)^{\#W_2} P_R^{\#R_2}
\end{aligned}
$$

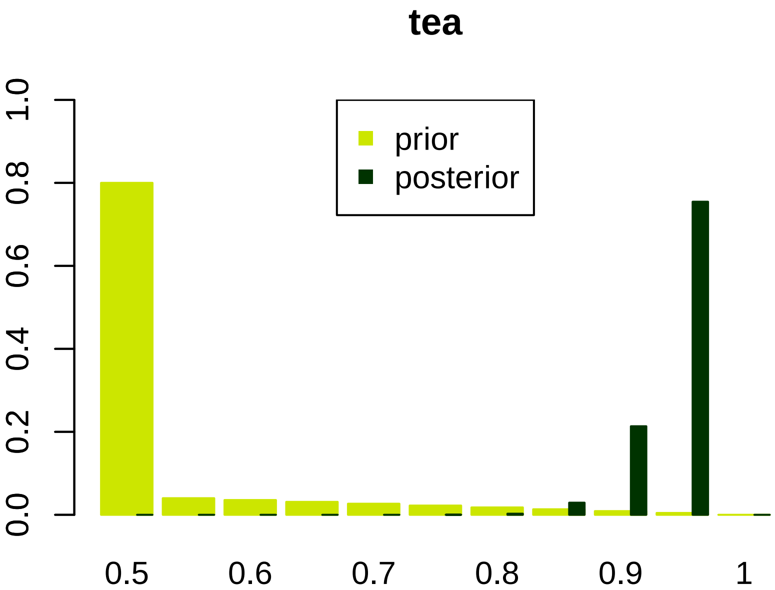
- ... which is equal to:

$$
\begin{aligned}
&p(P_R|\#R + \#R_2, \#W + \#W_2, wine) = \\
&\quad S_w \times K_w(1.01 - P_R)(P_R - 0.49) \\
&\qquad \times C'(1 - P_R)^{\#W + \#W_2} P_R^{\#R + \#R_2}
\end{aligned}
$$

The Lady Tasting Wine



The Lady Tasting Wine



The Lady Tasting Wine

- With enough data, the prior washes out in favor of the data

The Lady Tasting Wine

- With enough data, the prior washes out in favor of the data
  - How many additional correct discriminations will it take before we conclude that $P_R = 1$?

The Lady Tasting Wine

- With enough data, the prior washes out in favor of the data
  - How many additional correct discriminations will it take before we conclude that $P_R = 1$?
    - Ha! Trick question! Since they already made an error, P_R can never be 1!

The Lady Tasting Wine

- With enough data, the prior washes out in favor of the data
  - How many additional correct discriminations will it take before we conclude that $P_R = 1$?
    - Ha! Trick question! Since they already made an error, P_R can never be 1!
  - Alright, suppose that the ladies actually made no error in the first 6 trials. How many more will it take before the posterior probability of $P_R = 1$ dominates the alternatives?

The Lady Tasting Wine

- With enough data, the prior washes out in favor of the data
  - How many additional correct discriminations will it take before we conclude that $P_R = 1$?
    - Ha! Trick question! Since they already made an error, P_R can never be 1!
  - Alright, suppose that the ladies actually made no error in the first 6 trials. How many more will it take before the posterior probability of $P_R = 1$ dominates the alternatives?
    - After 34 corrects, $p(P_R|data)$ for wine tasting accrues at $P_R = 1$

The Lady Tasting Wine

- With enough data, the prior washes out in favor of the data
  - How many additional correct discriminations will it take before we conclude that $P_R = 1$?
    - Ha! Trick question! Since they already made an error, P_R can never be 1!
  - Alright, suppose that the ladies actually made no error in the first 6 trials. How many more will it take before the posterior probability of $P_R = 1$ dominates the alternatives?
    - After 34 corrects, $p(P_R|data)$ for wine tasting accrues at $P_R = 1$
    - ... but nothing Dr. Muriel does will convince us that $P_R = 1$, because a priori, $p(P_R = 1) = 0$.

The Lady Tasting Wine

- With enough data, the prior washes out in favor of the data
  - How many additional correct discriminations will it take before we conclude that $P_R = 1$?
    - Ha! Trick question! Since they already made an error, P_R can never be 1!
  - Alright, suppose that the ladies actually made no error in the first 6 trials. How many more will it take before the posterior probability of $P_R = 1$ dominates the alternatives?
    - After 34 corrects, $p(P_R|data)$ for wine tasting accrues at $P_R = 1$
    - ... but nothing Dr. Muriel does will convince us that $P_R = 1$, because a priori, $p(P_R = 1) = 0$.
    - Cromwell's Rule is a general recommendation to give a prior nonzero mass at any point that is not a logical impossibility.

Summary

- Fisher argued for a dichotomy: either (a) an event of small probability under H_0 has occurred, or (b) H_0 is false

Summary

- Fisher argued for a dichotomy: either (a) an event of small probability under H_0 has occurred, or (b) H_0 is false
- This did not work

Summary

- Fisher argued for a dichotomy: either (a) an event of small probability under H_0 has occurred, or (b) H_0 is false
- This did not work
  - the probability needs to include events that did not occur but were as, or more, extreme

Summary

- Fisher argued for a dichotomy: either (a) an event of small probability under H_0 has occurred, or (b) H_0 is false
- This did not work
  - the probability needs to include events that did not occur but were as, or more, extreme
- This did not work either

Summary

- Fisher argued for a dichotomy: either (a) an event of small probability under H_0 has occurred, or (b) H_0 is false
- This did not work
  - the probability needs to include events that did not occur but were as, or more, extreme
- This did not work either
  - it is ambiguous what is ‘more extreme’

Summary

- Fisher argued for a dichotomy: either (a) an event of small probability under H_0 has occurred, or (b) H_0 is false
- This did not work
  - the probability needs to include events that did not occur but were as, or more, extreme
- This did not work either
  - it is ambiguous what is 'more extreme'
- The solution involves Bayes' theorem

Summary

- Fisher argued for a dichotomy: either (a) an event of small probability under H_0 has occurred, or (b) H_0 is false
- This did not work
  - the probability needs to include events that did not occur but were as, or more, extreme
- This did not work either
  - it is ambiguous what is 'more extreme'
- The solution involves Bayes' theorem
  - Compare probabilities of the data under H_0 and alternatives

Summary

- Fisher argued for a dichotomy: either (a) an event of small probability under H_0 has occurred, or (b) H_0 is false
- This did not work
    - the probability needs to include events that did not occur but were as, or more, extreme
- This did not work either
    - it is ambiguous what is 'more extreme'
- The solution involves Bayes' theorem
    - Compare probabilities of the data under H_0 and alternatives
    - Different hypotheses weighted by prior beliefs

Summary

- Fisher argued for a dichotomy: either (a) an event of small probability under H_0 has occurred, or (b) H_0 is false
- This did not work
  - the probability needs to include events that did not occur but were as, or more, extreme
- This did not work either
  - it is ambiguous what is 'more extreme'
- The solution involves Bayes' theorem
  - Compare probabilities of the data under H_0 and alternatives
  - Different hypotheses weighted by prior beliefs
  - Priors are modified by the data to yield posterior beliefs

Summary

- Fisher argued for a dichotomy: either (a) an event of small probability under H_0 has occurred, or (b) H_0 is false
- This did not work
  - the probability needs to include events that did not occur but were as, or more, extreme
- This did not work either
  - it is ambiguous what is 'more extreme'
- The solution involves Bayes' theorem
  - Compare probabilities of the data under H_0 and alternatives
  - Different hypotheses weighted by prior beliefs
  - Priors are modified by the data to yield posterior beliefs
  - Then compare the various possible explanations for what has happened, and compare posterior beliefs with priors

Summary

- Classical analysis is biased against H_0

Summary

- Classical analysis is biased against H_0
  - The classical "significance level" is typically less than the posterior probability of H_0

Summary

- Classical analysis is biased against H_0
    - The classical "significance level" is typically less than the posterior probability of H_0
    - H_0 will be more easily discounted using Fisher's method than with the Bayesian approach

Summary

- Classical analysis is biased against H_0
    - The classical "significance level" is typically less than the posterior probability of H_0
    - H_0 will be more easily discounted using Fisher's method than with the Bayesian approach
        - The vast number of significance tests that are used today will encourage specious beliefs in the efficacy of drugs, treatments, or experimental manipulations

Summary

- Classical analysis is biased against H_0
  - The classical "significance level" is typically less than the posterior probability of H_0
  - H_0 will be more easily discounted using Fisher's method than with the Bayesian approach
    - The vast number of significance tests that are used today will encourage specious beliefs in the efficacy of drugs, treatments, or experimental manipulations
    - Whenever you read some effect having been detected, remember that it probably refers to significance, which too easily suggests an effect when none exists

Summary

- Bayesian analysis gives us everything we want

Summary

- Bayesian analysis gives us everything we want
    - We (usually) either want to know

Summary

- Bayesian analysis gives us everything we want
  - We (usually) either want to know
    - if H_0 is true (as with tea), or

Summary

- Bayesian analysis gives us everything we want
    - We (usually) either want to know
        - if H_0 is true (as with tea), or
        - how big an effect is (as with wine)

Summary

- Bayesian analysis gives us everything we want
  - We (usually) either want to know
    - if H_0 is true (as with tea), or
    - how big an effect is (as with wine)
  - The posterior tells us exactly what we need to know

Summary

- Bayesian analysis gives us everything we want
  - We (usually) either want to know
    - if H_0 is true (as with tea), or
    - how big an effect is (as with wine)
  - The posterior tells us exactly what we need to know
    - In contrast to the p-value, which is a probability for something that did not happen under the assumption of a hypothesis that may not be true

Conclusions

- Bayesian analysis uses prior knowledge

Conclusions

- Bayesian analysis uses prior knowledge
  - Fisher's analysis uses only probabilities assuming guessing and does not handle alternative hypotheses

Conclusions

- Bayesian analysis uses prior knowledge
  - Fisher's analysis uses only probabilities assuming guessing and does not handle alternative hypotheses
  - The Bayesian view recognizes that ones opinion of tasting the two liquids may be different or that the ladies may have different skills

Conclusions

- Bayesian analysis is comparative

Conclusions

- Bayesian analysis is comparative
  - We compare the probabilities of the observed event under H_0 and under the alternatives

Conclusions

- Bayesian analysis is comparative
  - We compare the probabilities of the observed event under H_0 and under the alternatives
  - Contrast with Fisher's approach which involves only the probability of the data under H_0

Conclusions

- Bayesian analysis is comparative
  - We compare the probabilities of the observed event under H_0 and under the alternatives
  - Contrast with Fisher's approach which involves only the probability of the data under H_0
  - If evidence is produced to support some thesis, one must also consider the reasonableness of the evidence were the thesis false