# BMEG 802 – Advanced Biomedical Experimental Design and Analysis

## Bayesian Statistics

Joshua G. A. Cashaback, PhD

# Recap

- Maximum Likelihood Estimation (MLE)
  - Probability Distribution Function
  - Likelihood function
  - 3 Ways to find the Maximum Likelihood Estimation
    - Analytical (Calculus)
    - Brute Force (Grid Search)
    - Optimization (Gradient Descent)

# Today

Bayesian Statistics

- Derivation from Set Theory
- Point Probabilities
  - priors, likelihood, posteriors
- Continuous Probabilites
  - Analytical (Conjugate Priors)
  - Numerical

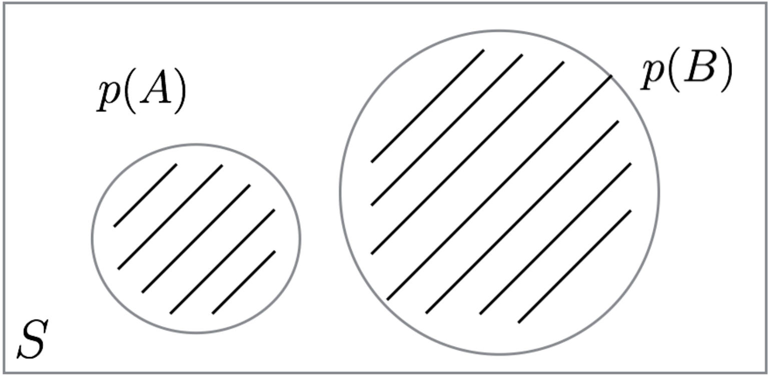
# Bayesian vs. Frequentist

- **Data are treated as fixed observations** vs. data (sample) treated as a random variable
- **Models (parameters) are treated as random variables** vs. models (population parameters) are treated as fixed quantities
- **we compute the probability of all models** vs. we compute the probability of one model ($H_0$)
- **we end up with a richer understanding of relative probability of all models** vs. we make a decision (reject $H_0$ or not)

# Notation

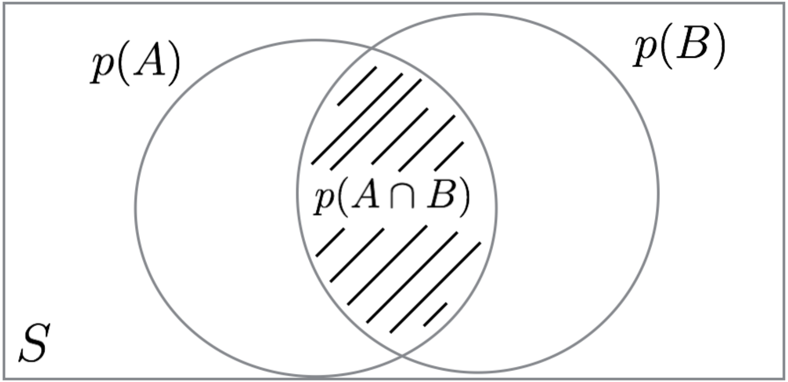
- $S =$ sample space (all possible outcomes)
- $p(A) =$ probability of event A
- $A \cup B =$ union of events A and B
- $A \cap B =$ intersection of events A and B
- $p(B|A) =$ probability of B given A
- $p(A')$ or $p(A^C)$ or $p(\bar{A}) =$ complement probability of $p(A)$

# Mutually Exclusive (Disjoint Probability)




$$p(A \cup B) = p(A) + p(B)$$
$$0.7 = 0.4 + 0.3$$

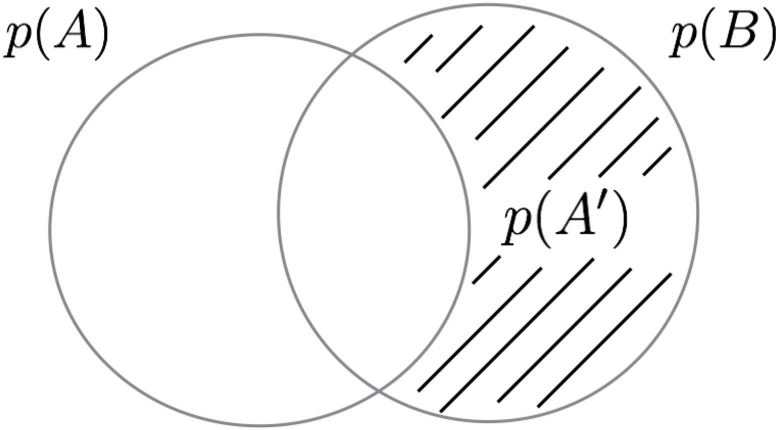
# Joint Probability




$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$
$$0.5 = 0.4 + 0.3 - 0.2$$

- the probability of two events occuring simultaneously

# Complement Probability



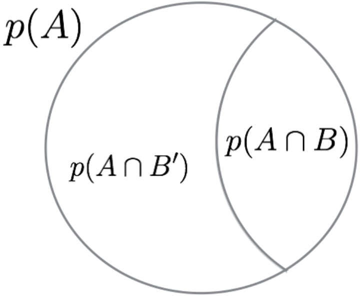

$$p(A') = 1 - p(A)$$

# Marginal Probability

| H \ L | Red | Yellow | Green | Marginal probability P(H) |
|---|---|---|---|---|
| **Not Hit** | 0.198 | 0.09 | 0.14 | **0.428** |
| **Hit** | 0.002 | 0.01 | 0.56 | **0.572** |
| **Total** | **0.2** | **0.1** | **0.7** | **1** |

- Probability of a single event occurring (hit), independent of other events (light)
- e.g., probabilities of getting in an accident at an intersection irrespective of lights
- note: joint probabilities in each cell

# Marginal Probability




$$p(A \cap B') = p(A) - p(A \cap B)$$

The marginal $p(A)$ or $p(B)$ is found by summating their disjoint parts.

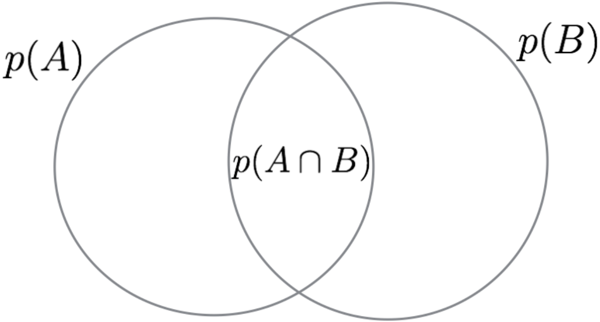
$p(A) = p(A \cap B) + p(A \cap B')$, and similarly

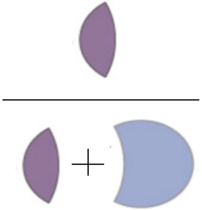
$$p(B) = p(A \cap B) + p(A' \cap B)$$

# Conditional Probability

- $p(accepted) = 0.3$
- $p(funding|accepted) = 0.43$
- $p(funding \cap accepted) = p(funding|accepted) \cdot p(accepted)$
- $p(funding \cap accepted) = 0.43 \cdot 0.3 = 0.13$
- Probability that an event occurs given that another specific event *has already* occurred

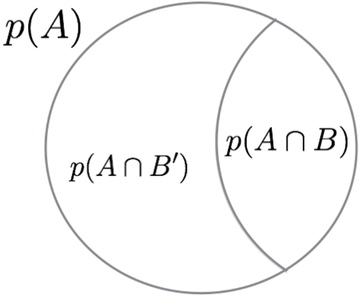
# Conditional Probability



$$p(A \cap B) = p(B|A) \cdot p(A)$$
$$p(B|A) = \frac{p(A \cap B)}{p(A)}$$
$$p(A \cap B) = p(A|B) \cdot p(B) \text{ (in terms of B)}$$
$$p(A|B) = \frac{p(A \cap B)}{p(B)}$$

# Conditional Probability Complements




$$p(A \cap B') = p(B'|A) \cdot p(A)$$

Other friendly complements:

$$p(A' \cap B) = p(B|A') \cdot p(A')$$

$$p(A' \cap B') = p(B'|A') \cdot p(A')$$

# Good News!

Bayes' Theorem is simply a conditional probability!

# Deriving Bayes' Theorem

Remember:

- $p(A|B) = \frac{p(A \cap B)}{p(B)}$,(eq.1)(slide 11)
- $p(A \cap B) = p(B|A) \cdot p(A)$,(eq.2)(slide 11)

Substitute (eq.2) into (eq.1):

$$p(A|B) = \frac{p(B|A) \cdot p(A)}{p(B)}, (eq.3)$$

That's it!

**In terms of statistical models:**

$$p(model|data) = \frac{p(data|model) \cdot p(model)}{p(data)}$$

# Handy Dandy Steps for Point Estimates

$p(A|B) = \frac{p(B|A) \cdot p(A)}{p(B)}, (eq.3)$

*Calculate p(B) by using its marginal probability*

$p(B) = p(A \cap B) + p(A' \cap B)$,$(eq.4)$(slide 9)

# Handy Dandy Steps for Point Estimates

$p(A|B) = \frac{p(B|A) \cdot p(A)}{p(B)}, (eq.3)$

*Calculate p(B) by using its marginal probability*

$p(B) = p(A \cap B) + p(A' \cap B)$,$(eq.4)$(slide 9)

*Substitute (eq.4) into (eq.3)*

$p(A|B) = \frac{p(B|A) \cdot p(A)}{p(A \cap B) + p(A' \cap B)}, (eq.5)$

# Handy Dandy Steps for Point Estimates

$p(A|B) = \frac{p(B|A) \cdot p(A)}{p(B)}, (eq.3)$

**Calculate p(B) by using its marginal probability**

$p(B) = p(A \cap B) + p(A' \cap B)$,$(eq.4)$(slide 9)

**Substitute (eq.4) into (eq.3)**

$p(A|B) = \frac{p(B|A) \cdot p(A)}{p(A \cap B) + p(A' \cap B)}, (eq.5)$

**Since,**

$p(A \cap B) = p(B|A) \cdot p(A)$,$(eq.6)$(slide 11)

$p(A' \cap B) = p(B|A') \cdot p(A')$,$(eq.7)$(slide 12)

*Substitute (eq.6) and (eq.7) into (eq.5)*

$p(A|B) = \frac{p(B|A) \cdot p(A)}{p(B|A) \cdot p(A) + p(B|A') \cdot p(A')}, (eq.8)$

# Why Bayesian???

Powerful way to continually account for new evidence given **prior** beliefs

$$p(A|B) = \frac{p(B|A) \cdot p(A)}{p(B)}$$

x Likelihood
(new evidence)

Prior

Posterior

# POINT PROBABILITES

# POINT PROBABILITES

Powerful way to continually account for new evidence given **prior** beliefs

$$p(A|B) = \frac{p(B|A) \cdot p(A)}{p(B|A) \cdot p(A) + p(B|A') \cdot p(A')}$$

x Likelihood
(new evidence)

Prior

Posterior

p(B) = marginal probability (e.g., true positive & false positive tests)

classic example: A = +covid, A' = -covid, B = + test, B' = - test

# Point Estimate Example

Let's say you take a COVID test and it comes out positive. What is the probability that you have COVID?

- Lets assume our initial, prior guess on whether we have covid based on an exposure is 40%, p(+covid) = 0.4.
- The probability of having a positive test given you have covid is 75%, $p(+test \mid +covid) = 0.75$. (i.e., test sensitivity)
- The probability of having a positive test given you do NOT have covid is 25%: $p(+test \mid -covid) = 0.25$. (i.e., 1 - test specificity)
- We observe a + test. What is the probability that you have COVID, $p(+covid \mid +test)$?
- note: these are fictitious numbers

# Point Estimate Example

$$p(+covid \mid +\, test) = \frac{p(+test \mid +\, covid) \cdot p(+covid)}{p(+test \mid +\, covid) \cdot p(+covid) + p(+test \mid -\, covid) \cdot p(-covid)}$$

# Point Estimate Example

**Knowns:**

$p(+covid) = 0.4$

$p(+test \mid +covid) = 0.75$

$p(+test \mid -covid) = 0.25$

**Unknowns:**

$p(-covid) =?$

$p(+covid \mid +test) =?$

# Point Estimate Example

**Knowns:**

$p(+covid) = 0.4$

$p(+test \mid + covid) = 0.75$

$p(+test \mid - covid) = 0.25$

**Unknowns:**

$p(-covid) = 0.6$; (1 - 0.4)

$p(+covid \mid + test) =?$

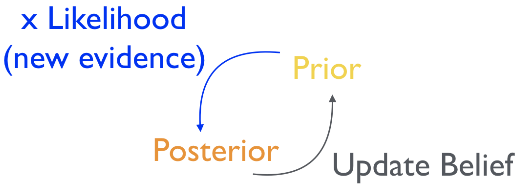
# Point Estimate Example

$$p(+covid \mid +test) = \frac{p(+test \mid +covid) \cdot p(+covid)}{p(+test \mid +covid) \cdot p(+covid) + p(+test \mid -covid) \cdot p(-covid)}$$

$$p(+covid \mid +test) = 0.67 = \frac{0.75 \cdot 0.4}{0.75 \cdot 0.4 + 0.25 \cdot 0.6}$$

# Why Bayesian?

Powerful way to **continually** account for new evidence given prior beliefs




$p(A|B)$ **becomes** $p(A)$ **on the next iteration!**

# Updating

- updating the model (i.e., take another test). Seems like an appropriate thing to do in science
- when new data are gathered, we can re-evaluate a hypothesis
- we do not begin anew (ignorant) each time we ask a question
- previous research provides us information about the merits of the hypothesis
- **the posterior from the previous model becomes the prior for the new model**

# Point Estimate Example - Updating

Let's continue from our previous example. We take another test and it comes out positive. What is our probability of having covid given another positive test?

# Point Estimate Example - Updating

**Knowns:**

$p(+covid) = 0.67$

$p(+test \mid +covid) = 0.75$

$p(+test \mid -covid) = 0.25$

**Unknowns:**

$p(-covid) = ?$

$p(+covid \mid +test) = ?$

# Point Estimate Example - Updating

**Knowns:**

$p(+covid) = 0.67$

$p(+test \mid +covid) = 0.75$

$p(+test \mid -covid) = 0.25$

**Unknowns:**

$p(-covid) = 0.33$; (1 - 0.67)

$p(+covid \mid +test) = ?$

# Point Estimate Example - Updating

$$p(+covid \mid + test) = \frac{p(+test \mid + covid) \cdot p(+covid)}{p(+test \mid + covid) \cdot p(+covid) + p(+test \mid - covid) \cdot p(-covid)}$$

$$p(+covid \mid + test) = 0.86 = \frac{0.75 \cdot 0.67}{0.75 \cdot 0.67 + 0.25 \cdot 0.33}$$

# Point Estimate Example - Updating

Let's keep going and pretend we observed 5 positive tests in row from our initial belief of 40%. Calculating p(+covid|+test) for each iteration leads to:

0.67

0.86

0.95

0.98

0.99

# Influence of PRIOR beliefs

How much of our prediction is influenced by our prior belief that the hand is open or not?

| Prior | Posterior |
|---|---|
| 0.1 | 0.25 |
| 0.2 | 0.43 |
| 0.3 | 0.56 |
| **0.4** | **0.67** |
| 0.5 | 0.75 |
| 0.6 | 0.82 |
| 0.7 | 0.88 |
| 0.8 | 0.92 |
| 0.9 | 0.96 |

# CONTINUOUS PROBABILITIES

# Bayes with Probability Distributions

- in previous example, the likelihood and prior were both single quantities (point probabilities)
- typically Bayesian approaches use full probability distributions
- essentially allows us to evaluate probability of a whole range of possible models, at once

# Back to the Bayesics

$p(A|B) = \frac{p(B|A) \cdot p(A)}{p(B)}$

$p(\theta|y)d\theta = \frac{p(y|\theta) \cdot p(\theta)d\theta}{\int_a^b p(y|\theta) \cdot p(\theta)d\theta}$

$p(\theta|y) = \frac{p(y|\theta) \cdot p(\theta)}{\int_a^b p(y|\theta) \cdot p(\theta)d\theta}$

Marginal probability a normalization constant (sum of prior and likelihood)

$p(\theta|y) \propto p(y|\theta) \cdot p(\theta)$

More details here: https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/readings/MIT18_05S14_Reading13a.pdf

# Bayes' Theorem

$$p(\theta|y) \propto \mathcal{L}(y|\theta) \cdot p(\theta)$$

x Likelihood
(new evidence)

Prior

Posterior

Update Belief

posterior, likelihood, prior can all be defined with probability distributions

# Continuous Probability Example

- Let's revisit the coin flipping example.
- Is the coin fair ($w = 0.5$)?
  - **model**: some proposed process by which the outcome of our coin flip is determined.
    - Binomial Distribution
  - **data**: $k = 2$ heads (# of successes), $n = 3$ flips

# Likelihood Function

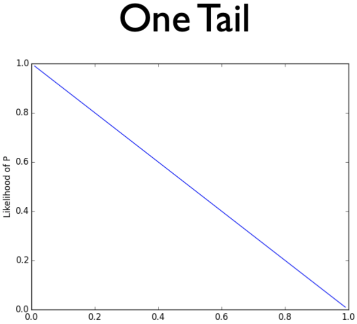
$$\mathcal{L}(w|n, y) = \frac{n!}{y!(n-y)!} w^y (1-w)^{n-y}$$

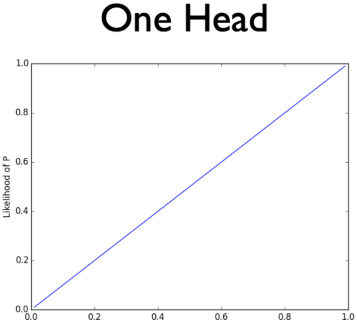
# Likelihood of W for a single toss

$w = 0$ represents the coin is perfectly weighted towards Tails

$w = 1$ represents the coin is perfectly weighted towards Heads

First, lets consider a single toss of Tails and a single toss of Heads



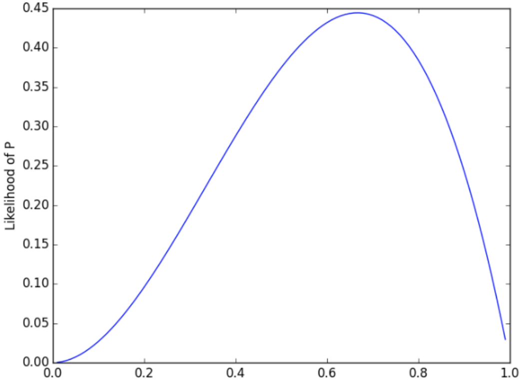

- Note: x-axis is $w$

# Likelihood of P for 3 tosses

$w = 0$ represents the coin is perfectly weighted towards Tails

$w = 1$ represents the coin is perfectly weighted towards Heads

2 Heads, 1 Tails



# The Prior

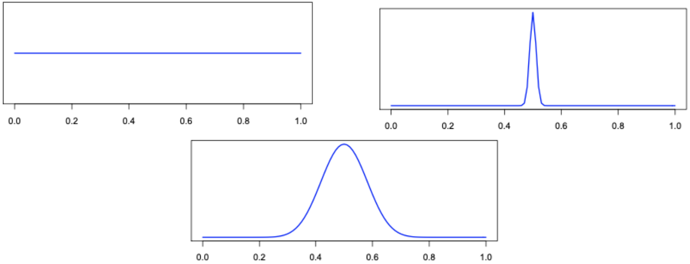
$$p(\theta|y) \propto \mathcal{L}(y|\theta) \cdot p(\theta)$$

- What is our prior belief? Is the coin weighted or not?

- "uninformative"
  - flat prior - all values of $w$ are equally likely (Bayes / Laplace):
  - Others - Jeffrey's prior, reference priors, maximum entropy
- Informative
  - we have some previous experience / evidence

# The Prior Cont'd

1. I have no clue what W is (flat prior)
2. Every coin we have seen in the past has been fair
3. Most coins have been relatively fair



# Calculating the Posterior

- Analytical
- Numerical

# Analytical

Find a Conjugate Prior

1. IF, the posterior and prior are the same type of distribution, they are conjugate distributions
2. THEN, the prior is a conjugate prior to the likelihood function
3. If we have a conjugate prior we can use hyperparameters to solve the posterior
4. Hyperparameters solved for many distributions: Conjugate Priors - Wikipedia

- The Binomical distribution conjugate prior is the Beta distribution
- The Normal distribution conjugate prior is the Normal distribution

# Analytical

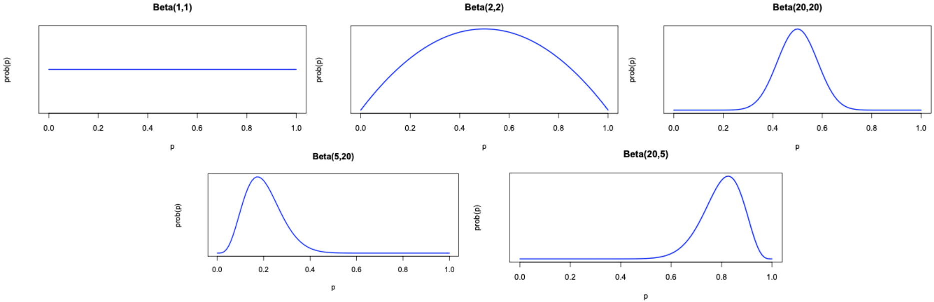
- Back to our coin flipping example.
- Our Likelihood function is the Binomial Distribution
- Binomial distribution's conjugate prior = Beta distribution

$$p(w|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} w^{\alpha-1}(1-w)^{\beta-1}$$

# Beta Distribution Conjugate Prior

$$p(w|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} w^{\alpha-1}(1-w)^{\beta-1}$$

Beta distribution's range [0,1] convenient for our prior $w$ somewhere between 0 and 1




*note: x-axis is $w$

# Analytically Calculating the Posterior

- If we have a conjugate prior
- Then, the posterior is calculated from parameters used in the likelihood and prior (called hyperparameters)

# Analytically Calculating the Posterior Cont'd

Prior: $p(w|\alpha, \beta) = \frac{1}{B(\alpha,\beta)} w^{\alpha-1}(1-w)^{\beta-1}$

- parameters are: $\alpha, \beta$

Likelihood: $\mathcal{L}(w|n, y) = \frac{n!}{y!(n-y)!} w^{y}(1-w)^{n-y}$

- parameters are: $k, n$

Posterior = Likelihood * Prior => after some heavy calculus / algebra

Posterior: $p(w|\alpha_h, \beta_h) = \frac{1}{B(\alpha_h,\beta_h)} w^{\alpha_h-1}(1-w)^{\beta_h-1}$

- hyperparameters are: $\alpha_h, \beta_h$
- $\alpha_h = k + \alpha$
- $\beta_h = n - k + \beta$

# Let's update based on each toss

coin flip: $n = 3$ trials, $k = 2$ success

- 1st toss = Heads
- 2nd toss = Tails
- 3rd toss = Heads

Let's assume a flat prior

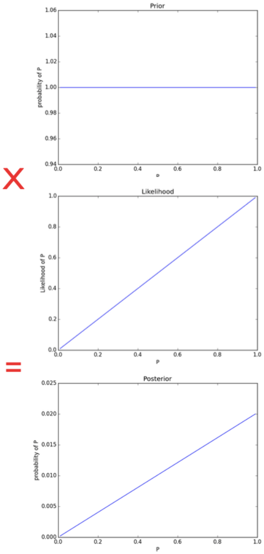
- all $w$ equal

# Toss One - Heads

**Prior:** $p(P|\alpha, \beta); \alpha = 1, \beta = 1$

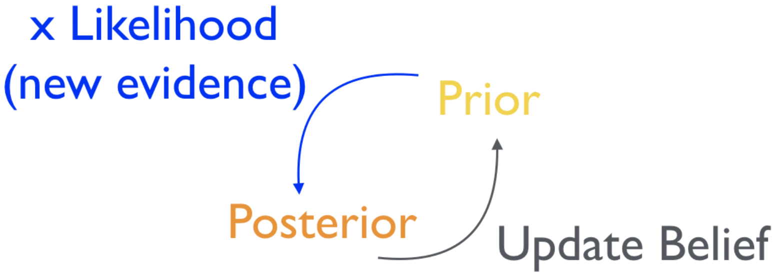
**Likelihood:** $L(P|k, n); k = 1, n = 1$

**Posterior:** $p(P|\alpha_h, \beta_h); \alpha_h = 2, \beta_h = 1$




21

# Same Update Procedure

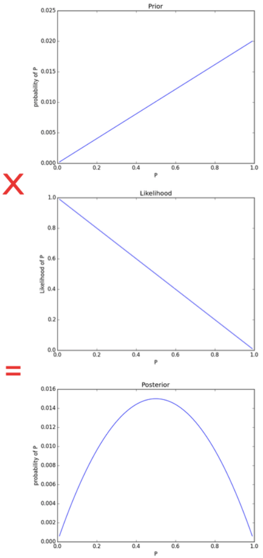


# Toss Two - Tails

**Prior:** $p(P|\alpha,\beta); \alpha = 2, \beta = 1$

**Likelihood:** $L(P|k,n); k = 0, n = 1$

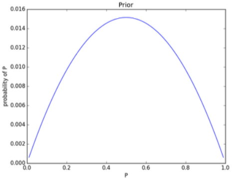
**Posterior:** $p(P|\alpha_h,\beta_h); \alpha_h = 2, \beta_h = 2$




23

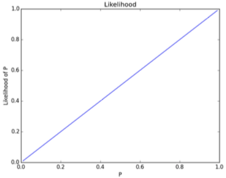
# Toss Three - Heads

**Prior:** $p(P|\alpha, \beta); \alpha = 2, \beta = 2$



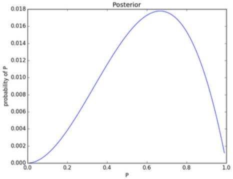

X

**Likelihood:** $L(P|k, n); k = 1, n = 1$




=

**Posterior:** $p(P|\alpha_h, \beta_h); \alpha_h = 3, \beta_h = 2$



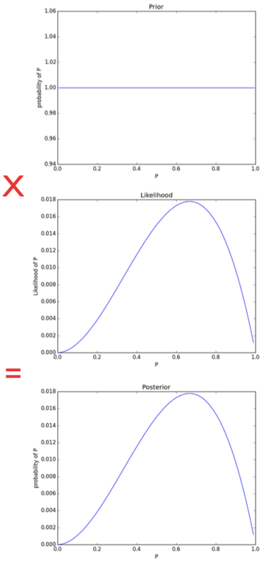

24

# OR, in a Single Step (all tosses already made)

**Prior:** $p(P|\alpha, \beta); \alpha = 1, \beta = 1$

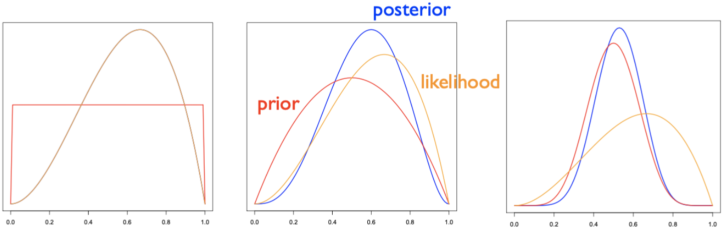
**Likelihood:** $L(P|k, n); k = 2, n = 3$

**Posterior:** $p(P|\alpha_h, \beta_h); \alpha_h = 3, \beta_h = 2$



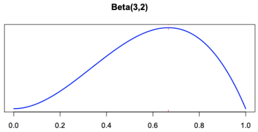

25

# Effect of Prior



# Analytically Describing the Posterior

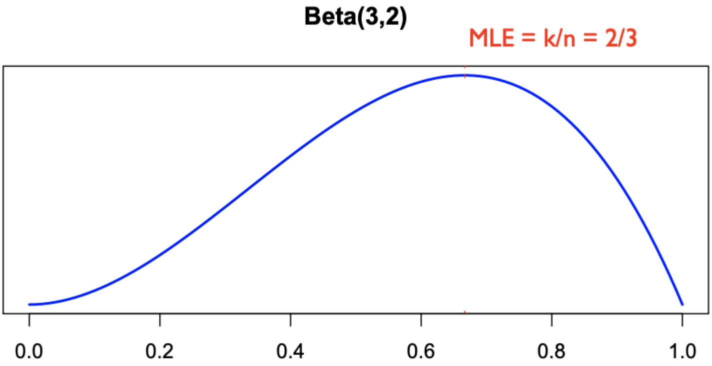
Graphically:




Summary Statistics:

- expression for mean, variance, mode, etc.
- mean $= \frac{\alpha}{\alpha+\beta}$
- variance $= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

# Bayesian vs. MLE




Bayesian: the posterior tells us the probability of all possible w's

MLE (frequentist approach): w is 0.667

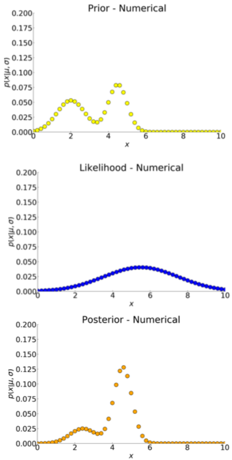
- does not incorporate prior information

# Numerically Calculating the Posterior

- Option 1: Grid approximation by discretizing the prior
- Option 2: Markov Chain Monte Carlo (MCMC)

# Grid Approximation

X




=

Any Shape

Tradeoff between grid coarseness and speed

# Criticisms of Bayesian Approach

- The prior: too much “subjectivity”?
- Data fixed, models (parameters) random
- Often difficult to find analytical solutions

# Advantages of Bayesian Approach

- Bayesian approach allows for incorporating previous findings in a principled way
- frequentist involves testing only one hypothesis (model) : the null hypothesis ... Bayesian estimates probability of all models (parameter values)
- interval estimates (and other such measures of posterior) have a clearer meaning than CIs in frequentist approaches
- in Bayesian approach we get full posterior distribution, a much richer picture than just a mean +/- CI or s.e.

# Applications

- Linear Regression: https://statswithr.github.io/book/introduction-to-bayesian-regression.html
  - pretty heavy stuff...
- Kalman Filters
- Multisensory Integration, Illusions, Sensorimotor Adaptation, etc.
- Bayes Factors
- etc.

# Next Week

Markov Chain Monte Carlo (MCMC)

- sampling the posterior