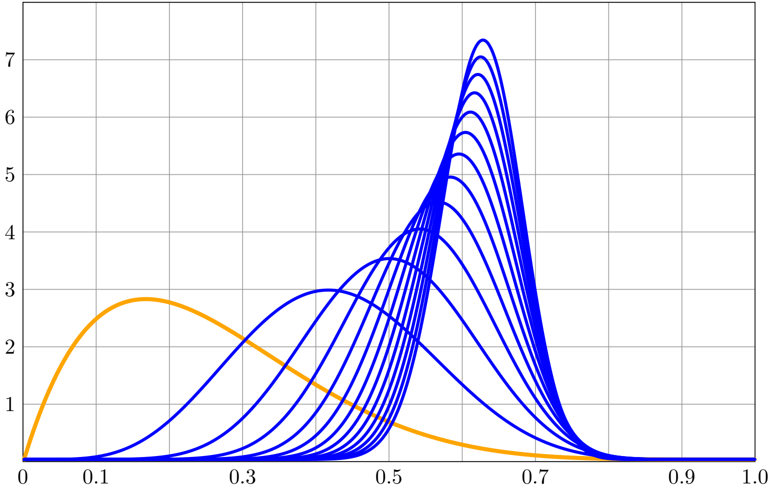
Bayesian Updating: Continuous Priors
18.05 Spring 2022



Announcements/Agenda

Announcements

- Pset due next week (4/4) will be a little longer than usual.

Agenda

- CSI blood types from before break
- Bayesian updating with a continuous range of hypotheses
- Keeping the notation straight
- Using integrals instead of sums

Continuous range of hypotheses

Example. Have a process that can succeed or fail, with unknown probability of success θ. We can hypothesize that θ takes any value in $[0, 1]$.

Formally, label the hypothesis $\mathscr{H}_\theta$, e.g. $\mathscr{H}_{0.5}$, $\mathscr{H}_{0.6}$. More casually, label the hypothesis θ.

Model: 'bent coin' with probability θ of heads.

Example. Waiting for an event to happen. The waiting time $X \sim \text{exp}(\lambda)$ with unknown λ. We can hypothesize that λ takes any value greater than 0.

Example. Have normal random variable with unknown μ and σ. Can hypothesize that (μ, σ) is anywhere in $(-\infty, \infty) \times [0, \infty)$.

Example of Bayesian updating so far

- Three types of coins with probabilities 0.25, 0.5, 0.75 of heads.
- Assume the numbers of each type are in the ratio 1 to 2 to 1.
- Assume we pick a coin at random, toss it twice and get TT.

Compute the posterior probability the coin has probability 0.25 of heads.

(Solution is posted with today's solutions.)

Notation with lots of hypotheses I.

- Now there are 5 types of coins with probabilities 0.1, 0.3, 0.5, 0.7, 0.9 of heads.
- Assume the numbers of each type are in the ratio 1:2:3:2:1 (so fairer coins are more common).
- Again we pick a coin at random, toss it twice and get TT.

Construct the Bayesian update table for the posterior probabilities of each type of coin.

hypotheses $\mathcal{H}$	prior $P(\mathcal{H})$	likelihood $P(\text{data}\mid\mathcal{H})$	Bayes numerator $P(\text{data}\mid\mathcal{H})P(\mathcal{H})$	posterior $P(\mathcal{H}\mid\text{data})$
$\theta_{0.1}$	$1/9$	$(0.9)^2$	0.090	0.297
$\theta_{0.3}$	$2/9$	$(0.7)^2$	0.109	0.359
$\theta_{0.5}$	$3/9$	$(0.5)^2$	0.083	0.275
$\theta_{0.7}$	$2/9$	$(0.3)^2$	0.020	0.066
$\theta_{0.9}$	$1/9$	$(0.1)^2$	0.001	0.004
Total	1		$P(\text{data}) = 0.303$	1

Notation with lots of hypotheses II.

What about 9 coins with probabilities 0.1, 0.2, 0.3, ..., 0.9?

Assume fairer coins are more common with the number of coins of probability θ of heads proportional to $\theta(1-\theta)$

Again the data is TT.

We can do this!

Table with 9 hypotheses

hypotheses $\mathcal{H}$	prior $P(\mathcal{H})$	likelihood $P(\text{data}\|\mathcal{H})$	Bayes numerator $P(\text{data}\|\mathcal{H})P(\mathcal{H})$	posterior $P(\mathcal{H}\|\text{data})$
$\theta_{0.1}$	$k(0.1 \cdot 0.9)$	$(0.9)^2$	0.0442	0.1483
$\theta_{0.2}$	$k(0.2 \cdot 0.8)$	$(0.8)^2$	0.0621	0.2083
$\theta_{0.3}$	$k(0.3 \cdot 0.7)$	$(0.7)^2$	0.0624	0.2093
$\theta_{0.4}$	$k(0.4 \cdot 0.6)$	$(0.6)^2$	0.0524	0.1757
$\theta_{0.5}$	$k(0.5 \cdot 0.5)$	$(0.5)^2$	0.0379	0.1271
$\theta_{0.6}$	$k(0.6 \cdot 0.4)$	$(0.4)^2$	0.0233	0.0781
$\theta_{0.7}$	$k(0.7 \cdot 0.3)$	$(0.3)^2$	0.0115	0.0384
$\theta_{0.8}$	$k(0.8 \cdot 0.2)$	$(0.2)^2$	0.0039	0.0130
$\theta_{0.9}$	$k(0.9 \cdot 0.1)$	$(0.1)^2$	0.0005	0.0018
Total	1		$P(\text{data}) = 0.298$	1

$k = 0.606$ was computed so that the total prior probability is 1.

Notation with lots of hypotheses III.

What about 99 coins with probabilities 0.01, 0.02, 0.03, ..., 0.99?

Assume fairer coins are more common with the number of coins of probability θ of heads proportional to $\theta(1-\theta)$

Again the data is TT.

Maybe we can do this ...

Table with 99 coins

Hypothesis $\mathcal{H}$ $\mathcal{H}$	prior $P(\mathcal{H})$ $P(\mathcal{H})$ ($k = 0.04$)	likelihood $P(\text{data} \mid \text{hyp.})$ $P(\text{data} \mid \text{hyp.})$	Bayes numerator $P(\text{data} \mid \text{hyp.})P(\mathcal{H})$	Posterior $P(\text{data} \mid \text{hyp.})P(\mathcal{H})/P(\text{data})$
$\theta_{0.01}$	k(0.01)(1-0.01)	$(1-0.01)^2$	$k(0.01)(1-0.01)^2$	0.001940921
$\theta_{0.02}$	k(0.02)(1-0.02)	$(1-0.02)^2$	$k(0.02)(1-0.02)^2$	0.003765396
$\theta_{0.03}$	k(0.03)(1-0.03)	$(1-0.03)^2$	$k(0.03)(1-0.03)^2$	0.005476951
$\theta_{0.04}$	k(0.04)(1-0.04)	$(1-0.04)^2$	$k(0.04)(1-0.04)^2$	0.007079068
$\theta_{0.05}$	k(0.05)(1-0.05)	$(1-0.05)^2$	$k(0.05)(1-0.05)^2$	0.008575179
$\theta_{0.06}$	k(0.06)(1-0.06)	$(1-0.06)^2$	$k(0.06)(1-0.06)^2$	0.009968669
$\theta_{0.07}$	k(0.07)(1-0.07)	$(1-0.07)^2$	$k(0.07)(1-0.07)^2$	0.01126288
$\theta_{0.08}$	k(0.08)(1-0.08)	$(1-0.08)^2$	$k(0.08)(1-0.08)^2$	0.01246108
$\theta_{0.09}$	k(0.09)(1-0.09)	$(1-0.09)^2$	$k(0.09)(1-0.09)^2$	0.01356654
$\theta_{0.1}$	k(0.1)(1-0.1)	$(1-0.1)^2$	$k(0.1)(1-0.1)^2$	0.01458243
$\theta_{0.11}$	k(0.11)(1-0.11)	$(1-0.11)^2$	$k(0.11)(1-0.11)^2$	0.0155119
$\theta_{0.12}$	k(0.12)(1-0.12)	$(1-0.12)^2$	$k(0.12)(1-0.12)^2$	0.01635805
$\theta_{0.13}$	k(0.13)(1-0.13)	$(1-0.13)^2$	$k(0.13)(1-0.13)^2$	0.01712393
$\theta_{0.14}$	k(0.14)(1-0.14)	$(1-0.14)^2$	$k(0.14)(1-0.14)^2$	0.01781254
$\theta_{0.15}$	k(0.15)(1-0.15)	$(1-0.15)^2$	$k(0.15)(1-0.15)^2$	0.01842682
$\theta_{0.16}$	k(0.16)(1-0.16)	$(1-0.16)^2$	$k(0.16)(1-0.16)^2$	0.01896969
$\theta_{0.17}$	k(0.17)(1-0.17)	$(1-0.17)^2$	$k(0.17)(1-0.17)^2$	0.019444
$\theta_{0.18}$	k(0.18)(1-0.18)	$(1-0.18)^2$	$k(0.18)(1-0.18)^2$	0.01985256
$\theta_{0.19}$	k(0.19)(1-0.19)	$(1-0.19)^2$	$k(0.19)(1-0.19)^2$	0.02019812
$\theta_{0.2}$	k(0.2)(1-0.2)	$(1-0.2)^2$	$k(0.2)(1-0.2)^2$	0.02048341
$\theta_{0.21}$	k(0.21)(1-0.21)	$(1-0.21)^2$	$k(0.21)(1-0.21)^2$	0.02071109
$\theta_{0.22}$	k(0.22)(1-0.22)	$(1-0.22)^2$	$k(0.22)(1-0.22)^2$	0.02088377
$\theta_{0.23}$	k(0.23)(1-0.23)	$(1-0.23)^2$	$k(0.23)(1-0.23)^2$	0.02100402
$\theta_{0.24}$	k(0.24)(1-0.24)	$(1-0.24)^2$	$k(0.24)(1-0.24)^2$	0.02107436
$\theta_{0.25}$	k(0.25)(1-0.25)	$(1-0.25)^2$	$k(0.25)(1-0.25)^2$	0.02109727
$\theta_{0.26}$	k(0.26)(1-0.26)	$(1-0.26)^2$	$k(0.26)(1-0.26)^2$	0.02107516
$\theta_{0.27}$	k(0.27)(1-0.27)	$(1-0.27)^2$	$k(0.27)(1-0.27)^2$	0.02101042
$\theta_{0.28}$	k(0.28)(1-0.28)	$(1-0.28)^2$	$k(0.28)(1-0.28)^2$	0.02090537
$\theta_{0.29}$	k(0.29)(1-0.29)	$(1-0.29)^2$	$k(0.29)(1-0.29)^2$	0.0207623
$\theta_{0.3}$	k(0.3)(1-0.3)	$(1-0.3)^2$	$k(0.3)(1-0.3)^2$	0.02058343
$\theta_{0.31}$	k(0.31)(1-0.31)	$(1-0.31)^2$	$k(0.31)(1-0.31)^2$	0.02037095
$\theta_{0.32}$	k(0.32)(1-0.32)	$(1-0.32)^2$	$k(0.32)(1-0.32)^2$	0.020127
$\theta_{0.33}$	k(0.33)(1-0.33)	$(1-0.33)^2$	$k(0.33)(1-0.33)^2$	0.01985367
$\theta_{0.34}$	k(0.34)(1-0.34)	$(1-0.34)^2$	$k(0.34)(1-0.34)^2$	0.01955299
$\theta_{0.35}$	k(0.35)(1-0.35)	$(1-0.35)^2$	$k(0.35)(1-0.35)^2$	0.01922695
$\theta_{0.36}$	k(0.36)(1-0.36)	$(1-0.36)^2$	$k(0.36)(1-0.36)^2$	0.01887751
$\theta_{0.37}$	k(0.37)(1-0.37)	$(1-0.37)^2$	$k(0.37)(1-0.37)^2$	0.01850656
$\theta_{0.38}$	k(0.38)(1-0.38)	$(1-0.38)^2$	$k(0.38)(1-0.38)^2$	0.01811595
$\theta_{0.39}$	k(0.39)(1-0.39)	$(1-0.39)^2$	$k(0.39)(1-0.39)^2$	0.01770747
$\theta_{0.4}$	k(0.4)(1-0.4)	$(1-0.4)^2$	$k(0.4)(1-0.4)^2$	0.01728288
$\theta_{0.41}$	k(0.41)(1-0.41)	$(1-0.41)^2$	$k(0.41)(1-0.41)^2$	0.01684389
$\theta_{0.42}$	k(0.42)(1-0.42)	$(1-0.42)^2$	$k(0.42)(1-0.42)^2$	0.01639214
$\theta_{0.43}$	k(0.43)(1-0.43)	$(1-0.43)^2$	$k(0.43)(1-0.43)^2$	0.01592925
$\theta_{0.44}$	k(0.44)(1-0.44)	$(1-0.44)^2$	$k(0.44)(1-0.44)^2$	0.01545678
$\theta_{0.45}$	k(0.45)(1-0.45)	$(1-0.45)^2$	$k(0.45)(1-0.45)^2$	0.01497625
$\theta_{0.46}$	k(0.46)(1-0.46)	$(1-0.46)^2$	$k(0.46)(1-0.46)^2$	0.0144891
$\theta_{0.47}$	k(0.47)(1-0.47)	$(1-0.47)^2$	$k(0.47)(1-0.47)^2$	0.01399677
$\theta_{0.48}$	k(0.48)(1-0.48)	$(1-0.48)^2$	$k(0.48)(1-0.48)^2$	0.01350062
$\theta_{0.49}$	k(0.49)(1-0.49)	$(1-0.49)^2$	$k(0.49)(1-0.49)^2$	0.01300196
$\theta_{0.5}$	k(0.5)(1-0.5)	$(1-0.5)^2$	$k(0.5)(1-0.5)^2$	0.01250208
$\theta_{0.51}$	k(0.51)(1-0.51)	$(1-0.51)^2$	$k(0.51)(1-0.51)^2$	0.0120022
$\theta_{0.52}$	k(0.52)(1-0.52)	$(1-0.52)^2$	$k(0.52)(1-0.52)^2$	0.01150349
$\theta_{0.53}$	k(0.53)(1-0.53)	$(1-0.53)^2$	$k(0.53)(1-0.53)^2$	0.01100707
$\theta_{0.54}$	k(0.54)(1-0.54)	$(1-0.54)^2$	$k(0.54)(1-0.54)^2$	0.01051404
$\theta_{0.55}$	k(0.55)(1-0.55)	$(1-0.55)^2$	$k(0.55)(1-0.55)^2$	0.01002542
$\theta_{0.56}$	k(0.56)(1-0.56)	$(1-0.56)^2$	$k(0.56)(1-0.56)^2$	0.009542198
$\theta_{0.57}$	k(0.57)(1-0.57)	$(1-0.57)^2$	$k(0.57)(1-0.57)^2$	0.009065309
$\theta_{0.58}$	k(0.58)(1-0.58)	$(1-0.58)^2$	$k(0.58)(1-0.58)^2$	0.008595641
$\theta_{0.59}$	k(0.59)(1-0.59)	$(1-0.59)^2$	$k(0.59)(1-0.59)^2$	0.008134034
$\theta_{0.6}$	k(0.6)(1-0.6)	$(1-0.6)^2$	$k(0.6)(1-0.6)^2$	0.00768128
$\theta_{0.61}$	k(0.61)(1-0.61)	$(1-0.61)^2$	$k(0.61)(1-0.61)^2$	0.007238124
$\theta_{0.62}$	k(0.62)(1-0.62)	$(1-0.62)^2$	$k(0.62)(1-0.62)^2$	0.006805262
$\theta_{0.63}$	k(0.63)(1-0.63)	$(1-0.63)^2$	$k(0.63)(1-0.63)^2$	0.006383342
$\theta_{0.64}$	k(0.64)(1-0.64)	$(1-0.64)^2$	$k(0.64)(1-0.64)^2$	0.005972963
$\theta_{0.65}$	k(0.65)(1-0.65)	$(1-0.65)^2$	$k(0.65)(1-0.65)^2$	0.005574679
$\theta_{0.66}$	k(0.66)(1-0.66)	$(1-0.66)^2$	$k(0.66)(1-0.66)^2$	0.005188993
$\theta_{0.67}$	k(0.67)(1-0.67)	$(1-0.67)^2$	$k(0.67)(1-0.67)^2$	0.004816361
$\theta_{0.68}$	k(0.68)(1-0.68)	$(1-0.68)^2$	$k(0.68)(1-0.68)^2$	0.004457191
$\theta_{0.69}$	k(0.69)(1-0.69)	$(1-0.69)^2$	$k(0.69)(1-0.69)^2$	0.004111843
$\theta_{0.7}$	k(0.7)(1-0.7)	$(1-0.7)^2$	$k(0.7)(1-0.7)^2$	0.00378063

There's a better way: use symbolic notation!

- Let θ be the probability of heads: $\theta = 0.01, 0.02, ..., 0.99$.
- Use θ to also stand for the hypothesis that the coin is of a type with probability of heads $= \theta$.
- We are given a formula for the prior: $p(\theta) = k\theta(1-\theta)$

  k computed so $\sum_{\theta=0.01}^{0.99} p(\theta) = 1$. Using R: $k = 0.06$.
- The likelihood $P(\text{data}|\theta) = P(TT|\theta) = (1-\theta)^2$.

  Total probability $= p(\text{data}) = \sum_{\theta=0.01}^{0.99} p(\text{data}|\theta)p(\theta) = 0.30$

Our 99 row table becomes:

hyp. $\mathcal{H}$	prior $P(\mathcal{H})$	likelihood $P(\text{data}\|\mathcal{H})$	Bayes numerator $P(\text{data}\|\mathcal{H})P(\mathcal{H})$	posterior $P(\mathcal{H}\|\text{data})$
θ	$k\theta(1-\theta)$	$(1-\theta)^2$	$k\theta(1-\theta)^3$	$0.200 \cdot \theta(1-\theta)^3$
Total	1		$P(\text{data}) = 0.300$	1

Notation: big and little letters

1. (Big letters) Event A, probability function $P(A)$.

2. (Little letters) Value x, pmf $p(x)$ or pdf $f(x)$ or $\phi(x)$.

$$\text{`}X = x\text{' is an event:} \quad P(X = x) = p(x).$$

Bayesian updating

3. (Big letters) For hypotheses $\mathcal{H}$ and data $\mathcal{D}$:

$$P(\mathcal{H}),\ P(\mathcal{D}),\ P(\mathcal{H}|\mathcal{D}),\ P(\mathcal{D}|\mathcal{H}).$$

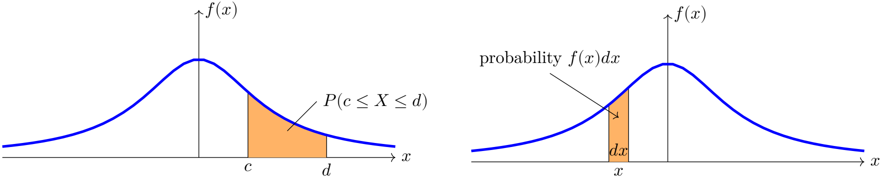
4. (Small letters) Hypothesis values θ and data values x:

$$\begin{array}{cccc} p(\theta) & p(x) & p(\theta|x) & p(x|\theta) \\ f(\theta)\,d\theta & \phi(x)\,dx & f(\theta|x)\,d\theta & \phi(x|\theta)\,dx \end{array}$$

Example. See the coin example in the reading

Review of pdf and probability

X random variable with pdf $f(x)$.
$f(x)$ is a density with units: probability/units of x.



$$P(c \leq X \leq d) = \int_c^d f(x)\, dx.$$

Probability X is in an infinitesimal range dx around x is

$$f(x)\, dx$$

Example of continuous hypotheses

Example. Suppose that we have a coin with probability of heads θ, where θ is unknown. We can hypothesize that θ takes any value in $[0, 1]$.

- Since θ is continuous we need a prior pdf $f(\theta)$, e.g. $f(\theta) = k\theta(1-\theta)$.

- Use $f(\theta)\, d\theta$ to work with probabilities instead of densities.

  For example, the prior probability that θ is in an infinitesimal range $d\theta$ around 0.5 is $f(0.5)\, d\theta$.

- To avoid cumbersome language we will simply say

  'The hypothesis θ has prior probability $f(\theta)\, d\theta$.'

Law of total probability for continuous distributions

Discrete set of hypotheses $\mathcal{H}_1, \mathcal{H}_2, \ldots \mathcal{H}_n$; data $\mathcal{D}$:

$$P(\mathcal{D}) = \sum_{i=1}^{n} P(\mathcal{D}|\mathcal{H}_i)P(\mathcal{H}_i).$$

In little letters: Hypothesis $\theta_1, \theta_2, \ldots, \theta_n$; data x

$$p(x) = \sum_{i=1}^{n} p(x|\theta_i)p(\theta_i).$$

Continuous range of hypothesis θ on $[a, b]$; discrete data x:

$$p(x) = \int_a^b p(x|\theta)f(\theta)\,d\theta$$

Always: "sum" of Bayes numerator column.
Also called the prior predictive probability of the outcome x.

Table question: total probability

(a) A coin has unknown probability of heads θ with prior pdf $f(\theta) = 3\theta^2$. Find the probability of throwing tails on the first toss.

(b) Describe an experiment with success and failure that this models. Include the reason for the prior in your description.

Concept question: discrete or continuous?

Suppose $X \sim \text{Bernoulli}(\theta)$ where the value of θ is unknown. If we use Bayesian methods to make probabilistic statements about θ then which one of the following is true?

1. The random variable is discrete, the space of hypotheses is discrete.

2. The random variable is discrete, the space of hypotheses is continuous.

3. The random variable is continuous, the space of hypotheses is discrete.

4. The random variable is continuous, the space of hypotheses is continuous.

Bayes' theorem for continuous distributions

- θ: continuous parameter with pdf $f(\theta)$ and range $[a, b]$;
- x: random discrete data;
- likelihood: $p(x|\theta)$.

Bayes' Theorem.

$$f(\theta|x)\, d\theta = \frac{p(x|\theta) f(\theta)\, d\theta}{p(x)} = \frac{p(x|\theta) f(\theta)\, d\theta}{\int_a^b p(x|\theta) f(\theta)\, d\theta}.$$

Not everyone uses $d\theta$ (**but they should**):

$$f(\theta|x) = \frac{p(x|\theta) f(\theta)}{p(x)} = \frac{p(x|\theta) f(\theta)}{\int_a^b p(x|\theta) f(\theta)\, d\theta}.$$

Bayesian update tables: continuous priors

$X \sim$ Bernoulli(θ). Unknown θ

Continuous hypotheses θ in [0,1].
Data x.
Prior pdf $f(\theta)$ (prior prob $f(\theta)\,d\theta$).
Likelihood $p(x|\theta)$.

hypothesis	range	prior	likelihood	Bayes numerator	posterior
θ	$[0,1]$	$f(\theta)\,d\theta$	$p(x\|\theta)$	$p(x\|\theta)f(\theta)\,d\theta$	$\dfrac{p(x\|\theta)f(\theta)\,d\theta}{p(x)}$
Total	$[0,1]$	1		$p(x) = \int_0^1 p(x\|\theta)f(\theta)\,d\theta$	1

;

Note $p(x) =$ the prior predictive probability of x.

Board question 1

'Bent' coin: unknown probability θ of heads.

Prior: $f(\theta) = 2\theta$ on $[0, 1]$.

Data: toss and get heads.

(a) Find the posterior pdf to this new data.

(b) Suppose you toss again and get tails. Update your posterior from problem 1 using this data.

(c) On one set of axes graph the prior and the posteriors from parts (a) and (b).

Board Question 2

Same scenario: bent coin $\sim$ Bernoulli(θ).

Flat prior: $f(\theta) = 1$ on $[0, 1]$

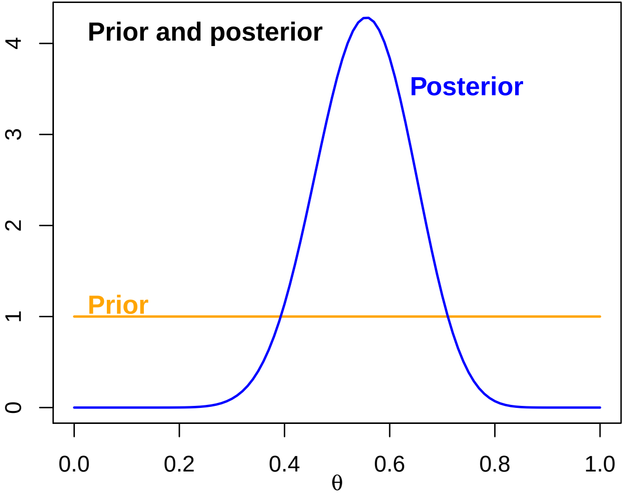
Data: toss 27 times and get 15 heads and 12 tails.

Use this data to find the posterior pdf.

Write an integral formula for the normalizing factor (total probability of the data), but do not compute it. Call its value T and give the posterior pdf in terms of T.

Solution graphs

Here are graphs of the prior and posterior for the previous question.



Beta distribution

$Beta(a, b)$ has density

$$f(\theta) = \frac{(a+b-1)!}{(a-1)!(b-1)!}\theta^{a-1}(1-\theta)^{b-1}$$

https://mathlets.org/mathlets/beta-distribution/

Observation: The coefficient is a normalizing factor so if

$$f(\theta) = c\theta^{a-1}(1-\theta)^{b-1}$$

is a pdf, then

$$c = \frac{(a+b-1)!}{(a-1)!(b-1)!}$$

and $f(\theta)$ is the pdf of a Beta(a, b) distribution.