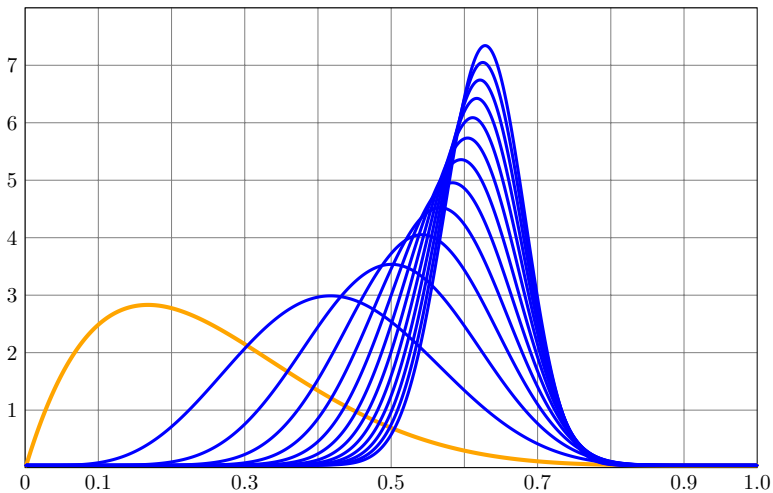


Bayesian Updating: Continuous Priors

18.05 Spring 2017



Continuous range of hypotheses

Example. Bernoulli with unknown probability of success p .
Can hypothesize that p takes any value in $[0, 1]$.
Model: 'bent coin' with probability p of heads.

Example. Waiting time $X \sim \exp(\lambda)$ with unknown λ .
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 (2 times)

Let $C_{0.25}$ stand for the hypothesis (event) that the chosen coin has probability 0.25 of heads. We want to compute $P(C_{0.25}|\text{data})$.

Method 1: Using Bayes' formula and the law of total probability:

$$\begin{aligned}P(C_{.25}|\text{data}) &= \frac{P(\text{data}|C_{.25})P(C_{.25})}{P(\text{data})} \\&= \frac{P(\text{data}|C_{.25})P(C_{.25})}{P(\text{data}|C_{.25})P(C_{.25}) + P(\text{data}|C_{.5})P(C_{.5}) + P(\text{data}|C_{.75})P(C_{.75})} \\&= \frac{(0.75)^2(1/4)}{(0.75)^2(1/4) + (0.5)^2(1/2) + (0.25)^2(1/4)} \\&= 0.5\end{aligned}$$

Method 2: Using a Bayesian update table:

hypotheses \mathcal{H}	prior $P(\mathcal{H})$	likelihood $P(\text{data} \mathcal{H})$	Bayes numerator $P(\text{data} \mathcal{H})P(\mathcal{H})$	posterior $P(\mathcal{H} \text{data})$
$C_{0.25}$	1/4	$(0.75)^2$	0.141	0.500
$C_{0.5}$	1/2	$(0.5)^2$	0.125	0.444
$C_{0.75}$	1/4	$(0.25)^2$	0.016	0.056
Total	1		$P(\text{data}) = 0.281$	1

Solution continued

Please be sure you understand how each of the pieces in method 1 correspond to the entries in the Bayesian update table in method 2.

Note. The total probability $P(\text{data})$ is also called the **prior predictive probability of the data**.

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 $P(\text{data} \mathcal{H})$	Bayes numerator $P(\text{data} \mathcal{H})P(\mathcal{H})$	posterior $P(\mathcal{H} \text{data})$
$C_{0.1}$	1/9	$(0.9)^2$	0.090	0.297
$C_{0.3}$	2/9	$(0.7)^2$	0.109	0.359
$C_{0.5}$	3/9	$(0.5)^2$	0.083	0.275
$C_{0.7}$	2/9	$(0.3)^2$	0.020	0.066
$C_{0.9}$	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 $0.1, 0.2, 0.3, \dots, 0.9$?

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 $P(\mathcal{H} \text{data})$
$C_{0.1}$	$k(0.1 \cdot 0.9)$	$(0.9)^2$	0.0442	0.1483
$C_{0.2}$	$k(0.2 \cdot 0.8)$	$(0.8)^2$	0.0621	0.2083
$C_{0.3}$	$k(0.3 \cdot 0.7)$	$(0.7)^2$	0.0624	0.2093
$C_{0.4}$	$k(0.4 \cdot 0.6)$	$(0.6)^2$	0.0524	0.1757
$C_{0.5}$	$k(0.5 \cdot 0.5)$	$(0.5)^2$	0.0379	0.1271
$C_{0.6}$	$k(0.6 \cdot 0.4)$	$(0.4)^2$	0.0233	0.0781
$C_{0.7}$	$k(0.7 \cdot 0.3)$	$(0.3)^2$	0.0115	0.0384
$C_{0.8}$	$k(0.8 \cdot 0.2)$	$(0.2)^2$	0.0039	0.0130
$C_{0.9}$	$k(0.9 \cdot 0.1)$	$(0.1)^2$	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 $0.01, 0.02, 0.03, \dots, 0.99$?

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 .

We could do this ...

Table with 99 coins

Hypothesis H_i	prior $P(H_i)$ $P(H_i) \propto 1$	likelihood $P(\text{data} \text{hyp.})$ $P(\text{data} \text{hyp.})$	Bayes numerator $P(\text{data} \text{hyp.})P(H_i)$	Posterior $P(\text{data} \text{hyp.})P(H_i) / P(\text{data})$
$C_{0.01}$	$k(0.01)(1-0.01)$	$(1-0.01)^2$	$k(0.01)(1-0.01)^2$	0.001940921
$C_{0.02}$	$k(0.02)(1-0.02)$	$(1-0.02)^2$	$k(0.02)(1-0.02)^2$	0.003765396
$C_{0.03}$	$k(0.03)(1-0.03)$	$(1-0.03)^2$	$k(0.03)(1-0.03)^2$	0.005476951
$C_{0.04}$	$k(0.04)(1-0.04)$	$(1-0.04)^2$	$k(0.04)(1-0.04)^2$	0.007079068
$C_{0.05}$	$k(0.05)(1-0.05)$	$(1-0.05)^2$	$k(0.05)(1-0.05)^2$	0.008575179
$C_{0.06}$	$k(0.06)(1-0.06)$	$(1-0.06)^2$	$k(0.06)(1-0.06)^2$	0.009968669
$C_{0.07}$	$k(0.07)(1-0.07)$	$(1-0.07)^2$	$k(0.07)(1-0.07)^2$	0.01126288
$C_{0.08}$	$k(0.08)(1-0.08)$	$(1-0.08)^2$	$k(0.08)(1-0.08)^2$	0.01246108
$C_{0.09}$	$k(0.09)(1-0.09)$	$(1-0.09)^2$	$k(0.09)(1-0.09)^2$	0.01356654
$C_{0.1}$	$k(0.1)(1-0.1)$	$(1-0.1)^2$	$k(0.1)(1-0.1)^2$	0.01458243
$C_{0.11}$	$k(0.11)(1-0.11)$	$(1-0.11)^2$	$k(0.11)(1-0.11)^2$	0.01551119
$C_{0.12}$	$k(0.12)(1-0.12)$	$(1-0.12)^2$	$k(0.12)(1-0.12)^2$	0.01635805
$C_{0.13}$	$k(0.13)(1-0.13)$	$(1-0.13)^2$	$k(0.13)(1-0.13)^2$	0.01712393
$C_{0.14}$	$k(0.14)(1-0.14)$	$(1-0.14)^2$	$k(0.14)(1-0.14)^2$	0.01781254
$C_{0.15}$	$k(0.15)(1-0.15)$	$(1-0.15)^2$	$k(0.15)(1-0.15)^2$	0.01842682
$C_{0.16}$	$k(0.16)(1-0.16)$	$(1-0.16)^2$	$k(0.16)(1-0.16)^2$	0.01896969
$C_{0.17}$	$k(0.17)(1-0.17)$	$(1-0.17)^2$	$k(0.17)(1-0.17)^2$	0.019444
$C_{0.18}$	$k(0.18)(1-0.18)$	$(1-0.18)^2$	$k(0.18)(1-0.18)^2$	0.01985256
$C_{0.19}$	$k(0.19)(1-0.19)$	$(1-0.19)^2$	$k(0.19)(1-0.19)^2$	0.02019812
$C_{0.2}$	$k(0.2)(1-0.2)$	$(1-0.2)^2$	$k(0.2)(1-0.2)^2$	0.02048341
$C_{0.21}$	$k(0.21)(1-0.21)$	$(1-0.21)^2$	$k(0.21)(1-0.21)^2$	0.02071109
$C_{0.22}$	$k(0.22)(1-0.22)$	$(1-0.22)^2$	$k(0.22)(1-0.22)^2$	0.02088377
$C_{0.23}$	$k(0.23)(1-0.23)$	$(1-0.23)^2$	$k(0.23)(1-0.23)^2$	0.02100402
$C_{0.24}$	$k(0.24)(1-0.24)$	$(1-0.24)^2$	$k(0.24)(1-0.24)^2$	0.02107436
$C_{0.25}$	$k(0.25)(1-0.25)$	$(1-0.25)^2$	$k(0.25)(1-0.25)^2$	0.02109727
$C_{0.26}$	$k(0.26)(1-0.26)$	$(1-0.26)^2$	$k(0.26)(1-0.26)^2$	0.02107516
$C_{0.27}$	$k(0.27)(1-0.27)$	$(1-0.27)^2$	$k(0.27)(1-0.27)^2$	0.02101042
$C_{0.28}$	$k(0.28)(1-0.28)$	$(1-0.28)^2$	$k(0.28)(1-0.28)^2$	0.02090537
$C_{0.29}$	$k(0.29)(1-0.29)$	$(1-0.29)^2$	$k(0.29)(1-0.29)^2$	0.0207623
$C_{0.3}$	$k(0.3)(1-0.3)$	$(1-0.3)^2$	$k(0.3)(1-0.3)^2$	0.02058343
$C_{0.31}$	$k(0.31)(1-0.31)$	$(1-0.31)^2$	$k(0.31)(1-0.31)^2$	0.02037095
$C_{0.32}$	$k(0.32)(1-0.32)$	$(1-0.32)^2$	$k(0.32)(1-0.32)^2$	0.020127
$C_{0.33}$	$k(0.33)(1-0.33)$	$(1-0.33)^2$	$k(0.33)(1-0.33)^2$	0.01985367
$C_{0.34}$	$k(0.34)(1-0.34)$	$(1-0.34)^2$	$k(0.34)(1-0.34)^2$	0.01955289
$C_{0.35}$	$k(0.35)(1-0.35)$	$(1-0.35)^2$	$k(0.35)(1-0.35)^2$	0.01922695
$C_{0.36}$	$k(0.36)(1-0.36)$	$(1-0.36)^2$	$k(0.36)(1-0.36)^2$	0.01887751
$C_{0.37}$	$k(0.37)(1-0.37)$	$(1-0.37)^2$	$k(0.37)(1-0.37)^2$	0.01850656
$C_{0.38}$	$k(0.38)(1-0.38)$	$(1-0.38)^2$	$k(0.38)(1-0.38)^2$	0.01811595
$C_{0.39}$	$k(0.39)(1-0.39)$	$(1-0.39)^2$	$k(0.39)(1-0.39)^2$	0.01770747
$C_{0.4}$	$k(0.4)(1-0.4)$	$(1-0.4)^2$	$k(0.4)(1-0.4)^2$	0.01728288
$C_{0.41}$	$k(0.41)(1-0.41)$	$(1-0.41)^2$	$k(0.41)(1-0.41)^2$	0.01684389
$C_{0.42}$	$k(0.42)(1-0.42)$	$(1-0.42)^2$	$k(0.42)(1-0.42)^2$	0.01639214
$C_{0.43}$	$k(0.43)(1-0.43)$	$(1-0.43)^2$	$k(0.43)(1-0.43)^2$	0.01592925
$C_{0.44}$	$k(0.44)(1-0.44)$	$(1-0.44)^2$	$k(0.44)(1-0.44)^2$	0.01545678
$C_{0.45}$	$k(0.45)(1-0.45)$	$(1-0.45)^2$	$k(0.45)(1-0.45)^2$	0.01497625
$C_{0.46}$	$k(0.46)(1-0.46)$	$(1-0.46)^2$	$k(0.46)(1-0.46)^2$	0.0144891
$C_{0.47}$	$k(0.47)(1-0.47)$	$(1-0.47)^2$	$k(0.47)(1-0.47)^2$	0.01399677
$C_{0.48}$	$k(0.48)(1-0.48)$	$(1-0.48)^2$	$k(0.48)(1-0.48)^2$	0.01350062
$C_{0.49}$	$k(0.49)(1-0.49)$	$(1-0.49)^2$	$k(0.49)(1-0.49)^2$	0.01300196
$C_{0.5}$	$k(0.5)(1-0.5)$	$(1-0.5)^2$	$k(0.5)(1-0.5)^2$	0.01250208
$C_{0.51}$	$k(0.51)(1-0.51)$	$(1-0.51)^2$	$k(0.51)(1-0.51)^2$	0.0120022
$C_{0.52}$	$k(0.52)(1-0.52)$	$(1-0.52)^2$	$k(0.52)(1-0.52)^2$	0.01150349
$C_{0.53}$	$k(0.53)(1-0.53)$	$(1-0.53)^2$	$k(0.53)(1-0.53)^2$	0.01100707
$C_{0.54}$	$k(0.54)(1-0.54)$	$(1-0.54)^2$	$k(0.54)(1-0.54)^2$	0.01051404
$C_{0.55}$	$k(0.55)(1-0.55)$	$(1-0.55)^2$	$k(0.55)(1-0.55)^2$	0.01002542
$C_{0.56}$	$k(0.56)(1-0.56)$	$(1-0.56)^2$	$k(0.56)(1-0.56)^2$	0.009542198
$C_{0.57}$	$k(0.57)(1-0.57)$	$(1-0.57)^2$	$k(0.57)(1-0.57)^2$	0.009065309
$C_{0.58}$	$k(0.58)(1-0.58)$	$(1-0.58)^2$	$k(0.58)(1-0.58)^2$	0.008595641
$C_{0.59}$	$k(0.59)(1-0.59)$	$(1-0.59)^2$	$k(0.59)(1-0.59)^2$	0.008134034
$C_{0.6}$	$k(0.6)(1-0.6)$	$(1-0.6)^2$	$k(0.6)(1-0.6)^2$	0.00768128
$C_{0.61}$	$k(0.61)(1-0.61)$	$(1-0.61)^2$	$k(0.61)(1-0.61)^2$	0.007238124
$C_{0.62}$	$k(0.62)(1-0.62)$	$(1-0.62)^2$	$k(0.62)(1-0.62)^2$	0.006805262
$C_{0.63}$	$k(0.63)(1-0.63)$	$(1-0.63)^2$	$k(0.63)(1-0.63)^2$	0.006383342
$C_{0.64}$	$k(0.64)(1-0.64)$	$(1-0.64)^2$	$k(0.64)(1-0.64)^2$	0.005972963
$C_{0.65}$	$k(0.65)(1-0.65)$	$(1-0.65)^2$	$k(0.65)(1-0.65)^2$	0.005574679
$C_{0.66}$	$k(0.66)(1-0.66)$	$(1-0.66)^2$	$k(0.66)(1-0.66)^2$	0.005188993

Maybe there's a better way

Use some symbolic notation!

- Let θ be the probability of heads: $\theta = 0.01, 0.02, \dots, 0.99$.
- Use θ to also stand for the hypothesis that the coin is of type with probability of heads = θ .
- We are given a formula for the prior: $p(\theta) = k\theta(1 - \theta)$
- The likelihood $P(\text{data}|\theta) = P(TT|\theta) = (1 - \theta)^2$.

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	$k\theta(1 - \theta)$	$(1 - \theta)^2$	$k\theta(1 - \theta)^3$	$0.2000 \cdot \theta(1 - \theta)^3$
Total	1		$P(\text{data}) = 0.300$	1

(We used R to compute $k = .0600$ so that the total prior probability is 1. Then we used it again to compute $P(\text{data})$ and $k/P(\text{data}) = .2000$.)

Notation: big and little letters

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

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

' $X = x$ ' is an event: $P(X = x) = p(x)$.

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 :

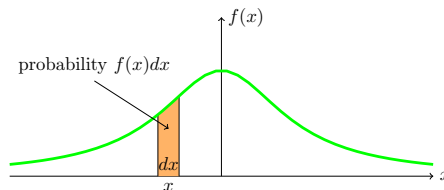
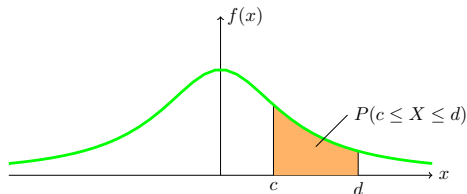
$$\begin{array}{cccc} p(\theta) & p(x) & p(\theta|x) & p(x|\theta) \\ f(\theta) d\theta & \phi(x) dx & f(\theta|x) d\theta & \phi(x|\theta) dx \end{array}$$

Example. Coin example in 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, e.g. 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 $\mathcal{H}_1, \mathcal{H}_2, \dots, \mathcal{H}_n$; data \mathcal{D} :

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

In **little letters**: Hypothesis $\theta_1, \theta_2, \dots, \theta_n$; 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 :

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

Also called **prior predictive probability** of the outcome x .

Board question: total probability

1. 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.
2. Describe a setup with success and failure that this models.

answer: 1. Take $x = 1$ for heads and $x = 0$ for tails. The likelihood $p(x = 0|\theta) = 1 - \theta$.

The law of total probability says

$$p(x = 0) = \int_0^1 p(x = 0|\theta)f(\theta) d\theta = \int_0^1 (1 - \theta)3\theta^2 d\theta = 1/4.$$

2. There are many possible examples. Here's one:

A medical treatment has unknown probability θ of success. A priori we think it's a good treatment so we use a prior of $f(\theta) = 3\theta^2$ which is biased towards success. The first use of it is succesful, so after updating we are even more biased for the treatment.

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**):

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

Concept question

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 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.

answer: 2. A Bernoulli random variable takes values 0 or 1. So X is discrete. The parameter θ can be anywhere in the continuous range $[0,1]$. Therefore the space of hypotheses is continuous.

Bayesian update tables: continuous priors

$X \sim \text{Bernoulli}(\theta)$. Unknown θ

Continuous hypotheses θ in $[0,1]$.

Data x .

Prior pdf $f(\theta)$.

Likelihood $p(x|\theta)$.

hypothesis	prior	likelihood	Bayes numerator	posterior
θ	$f(\theta) d\theta$	$p(x \theta)$	$p(x \theta)f(\theta) d\theta$	$\frac{p(x \theta)f(\theta) d\theta}{p(x)}$
Total	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

'Bent' coin: unknown probability θ of heads.

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

Data: toss and get heads.

1. Find the posterior pdf to this new data.
2. Suppose you toss again and get tails. Update your posterior from problem 1 using this data.
3. On one set of axes graph the prior and the posteriors from problems 1 and 2.

See next slide for solution.

Solution

Problem 1

hypoth.	prior	likelihood	Bayes numerator	posterior
θ	$2\theta d\theta$	θ	$2\theta^2 d\theta$	$3\theta^2 d\theta$
Total	1	$T = \int_0^1 2\theta^2 d\theta = 2/3$		1

Posterior pdf: $f(\theta|x) = 3\theta^2$. (Should graph this.)

Note: We don't really need to compute T . Once we know the posterior density is of the form $c\theta^2$ we only have to find the value of c makes it have total probability 1.

Problem 2

hypoth.	prior	likelihood	Bayes numerator	posterior
θ	$3\theta^2 d\theta$	$1 - \theta$	$3\theta^2(1 - \theta), d\theta$	$12\theta^2(1 - \theta) d\theta$
Total	1	$\int_0^1 3\theta^2(1 - \theta) d\theta = 1/4$		1

Posterior pdf: $f(\theta|x) = 12\theta^2(1 - \theta)$.

Board Question

Same scenario: bent coin $\sim \text{Bernoulli}(\theta)$.

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

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

1. Use this data to find the posterior pdf.

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

answer: $f(\theta|x) = \frac{1}{T} \theta^{15} (1 - \theta)^{12}$. (Called a Beta distribution.)



Beta distribution

$Beta(a, b)$ has density

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

<http://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.