An essay towards solving a problem in the doctrine of Chances

presented by Alexander Churbanov

University of Nebraska at Omaha, USA

Thomas Bayes (1702–1761)

Reverend Thomas Bayes

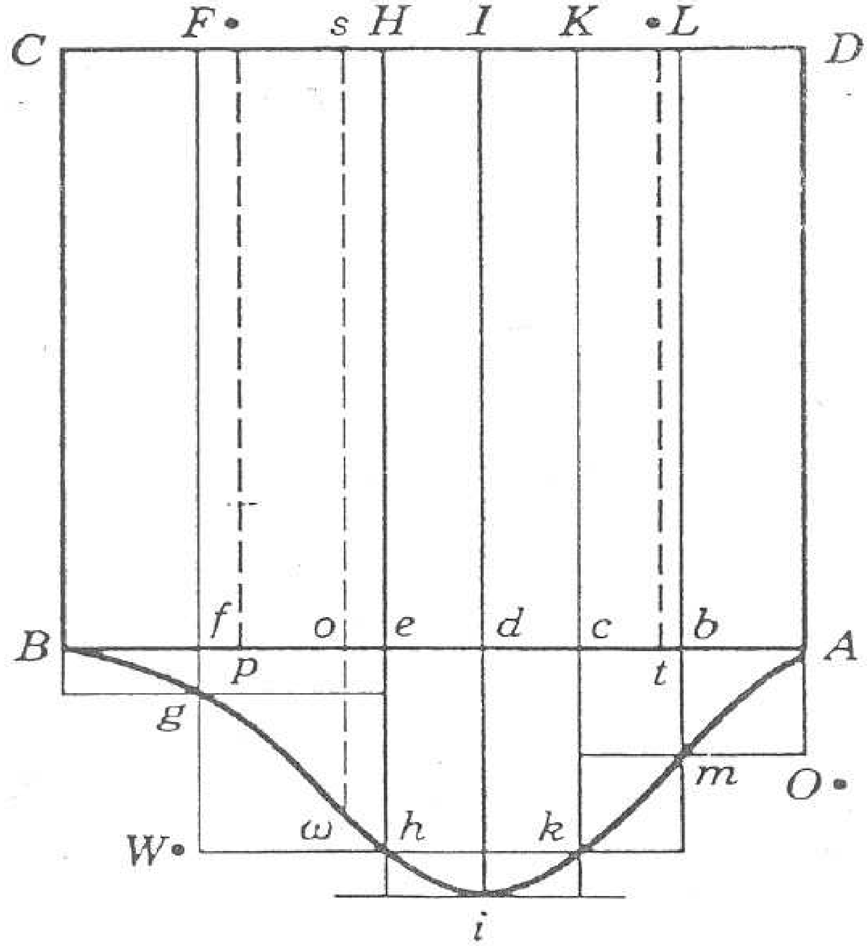
Bayesian paper

- *An essay towards solving a Problem in the doctrine of Chances* was published in Philosophical Transactions of the Royal Society of London in 1763
- For the first time paper treated axioms of probability and derivations of posterior distribution in a rigorous mathematical way

Experiment setting [1]

Suppose the square table with corners $ABCD$ and if the if either of the balls o or W are thrown upon it, there shall be the same probability that it rests upon any one equal part of the table as another, and that it must necessarily rest somewhere upon it.

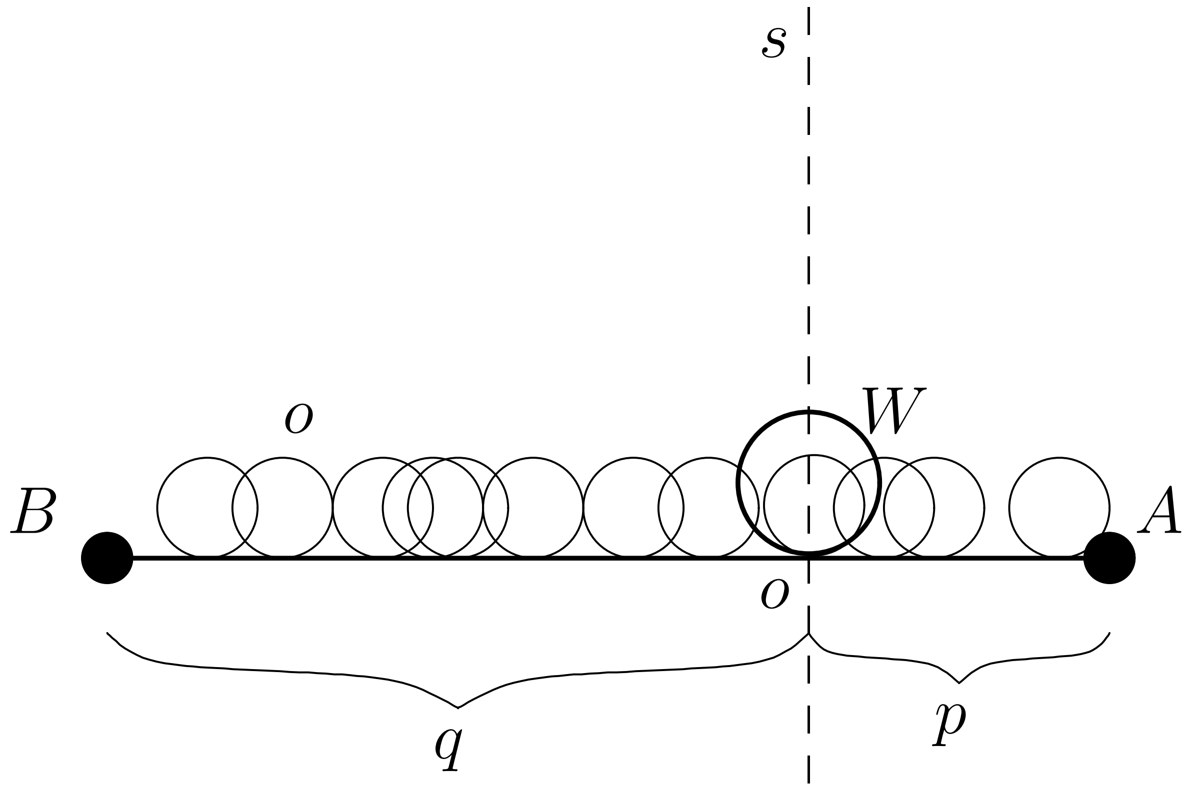
Bayesian table



From Bayes 1763 [1]

Suppose the the ball W is thrown first, and through the point where it rests a line os shall be drawn parallel to AD in s and o; and that afterwards the ball o shall be thrown $p+q$ or (N) times, and that its resting between AD and os after a single throw be called the happening of the event M in a single trial.

Bayesian experiment



In this experiment $p + q = N$

Bayesian experiment

o being uniformly distributed and once o is determined, M has binomial, conditional distribution with o representing success or failure on a single trial, the model is:

$$P(M = p|o) = \binom{N}{p} o^p (1 - o)^{N-p}$$

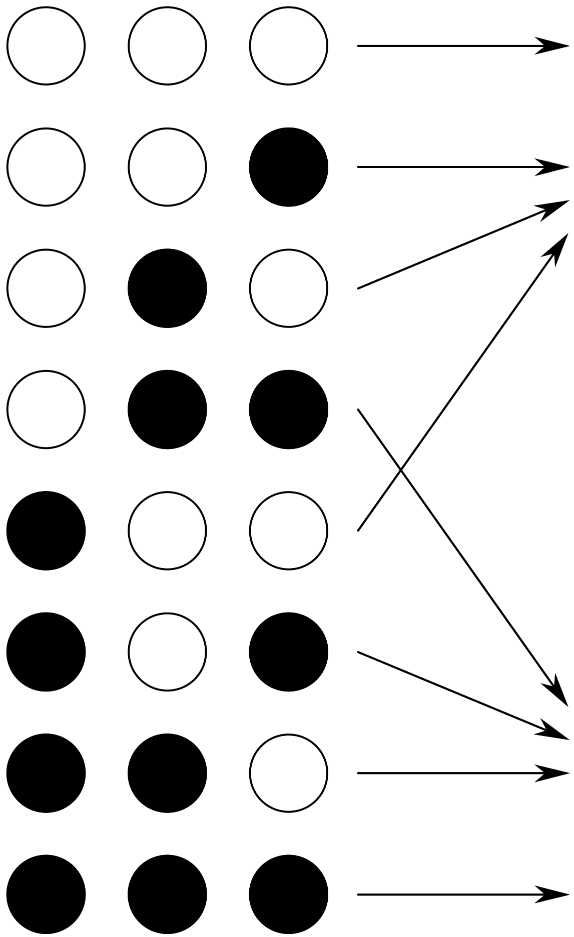
$$\forall p = 0, \ldots, N$$

Binomial experiment example

Given biased coin:

- Probability of heads $\frac{1}{3}$
- Probability of tails $\frac{2}{3}$

Binomial experiment example



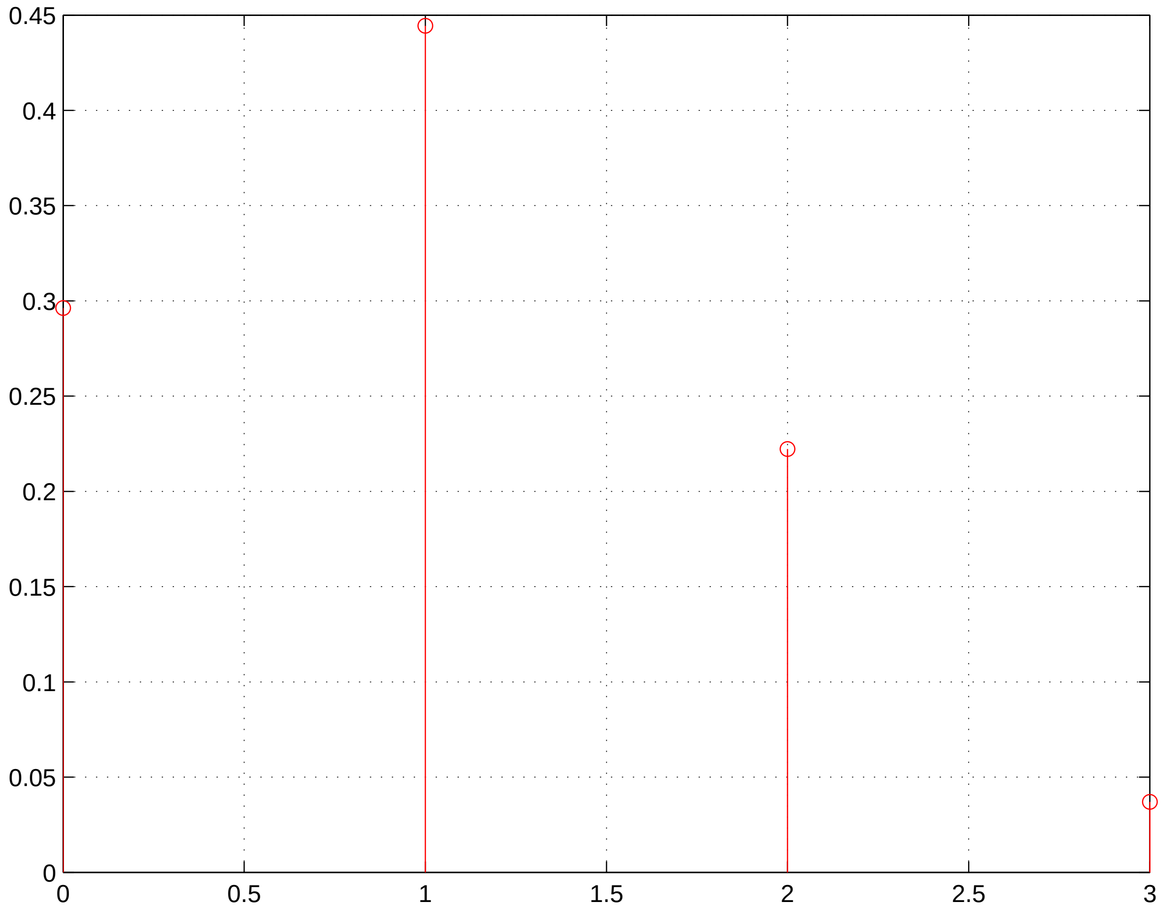
$$\frac{3!}{3!} \times \left[\frac{1}{3}\right]^3 \times \left[\frac{2}{3}\right]^0 \approx 0.037$$

$$\frac{3!}{1!\,2!} \times \left[\frac{1}{3}\right]^2 \times \left[\frac{2}{3}\right]^1 \approx 0.222$$

$$\frac{3!}{1!\,2!} \times \left[\frac{1}{3}\right]^1 \times \left[\frac{2}{3}\right]^2 \approx 0.444$$

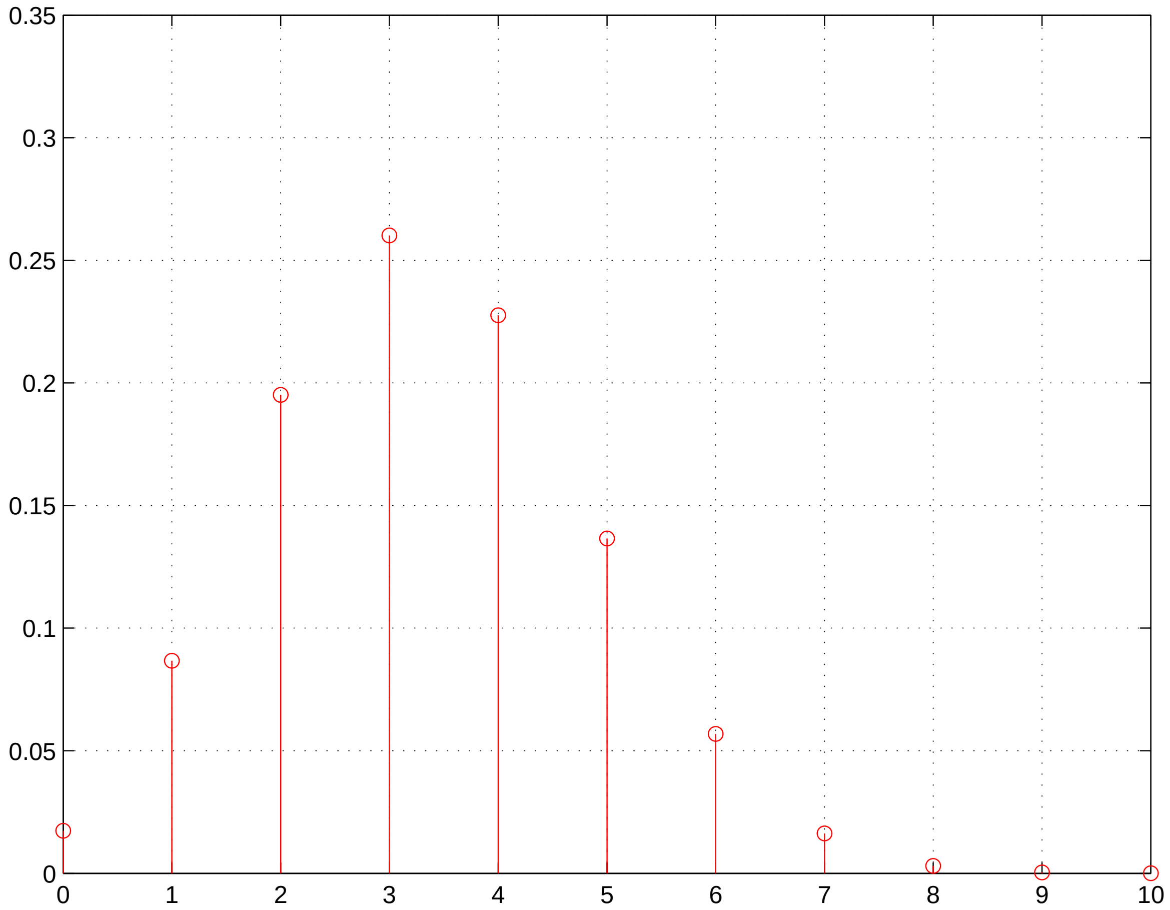
$$\frac{3!}{3!} \times \left[\frac{1}{3}\right]^0 \times \left[\frac{2}{3}\right]^3 \approx 0.296$$

Binomial experiment example



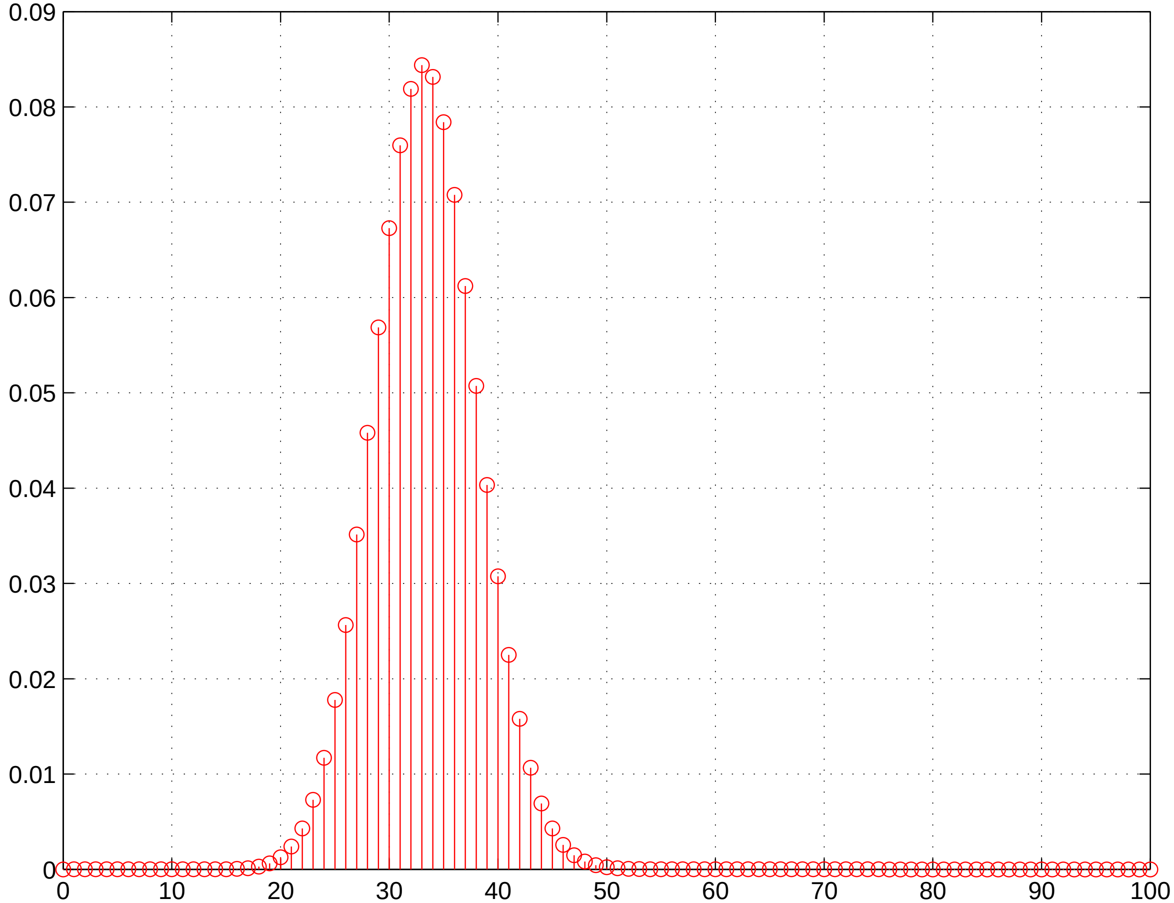
$binomial(0.3, 3)$

Binomial experiment example



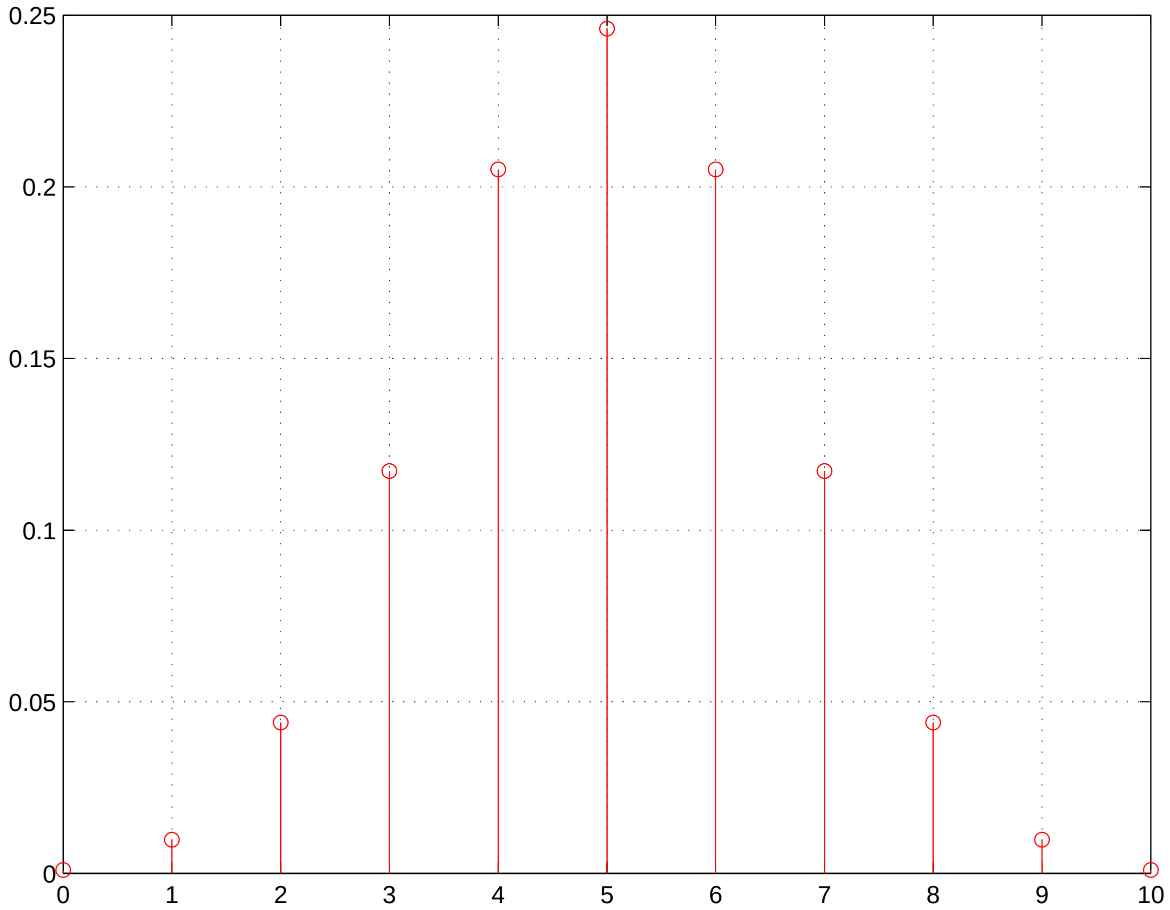
$$binomial(0.3, 10)$$

Binomial experiment example



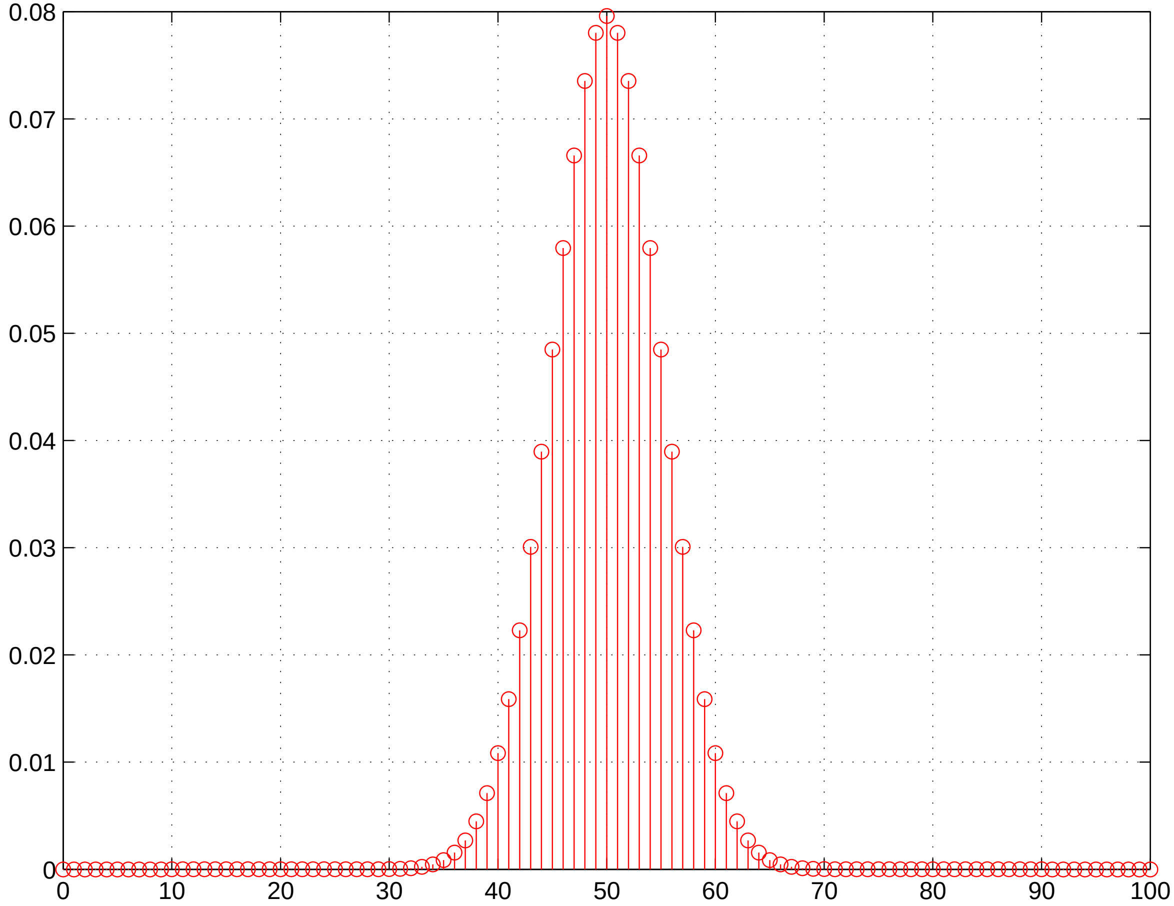
$$binomial(0.3, 100)$$

Binomial experiment example



$$binomial(0.5, 10)$$

Binomial experiment example



$$binomial(0.5, 100)$$

Questions to answer

- Given the event $0 < o < 1$, what is the probability for p balls to end up to the right of the position of o?

Questions to answer

- Given the event $0 < o < 1$, what is the probability for p balls to end up to the right of the position of o?

Bayes was interested in the converse problem

- Given the number of p balls to the right of o, what is the probability of o lying in the interval $(f, b) \subset [0, 1]$

$$P(f < o < b | M = p) = ?$$

$$\forall (f, b) \subset [0, 1]; p = 0, 1, \ldots, N$$

Lemma 1

The probability that the point o will fall between any two points f and b in the line AB is the ratio of the distance between the two points to the whole line

$$P(f < o < b) = \frac{fb}{AB}$$

Lemma 2

The ball W having been thrown, and the line os drawn, the probability of the event M in a single trial is the ratio of Ao to AB

$$P(M = 1) = \frac{Ao}{AB}$$

Proposition 8

$$P((f < o < b) \bigcap (M = p)) = \int_f^b \binom{N}{p} o^p (1-o)^{N-p} do$$

- Since there is no reason to favor one value for o over another, all values of o are equally likely
- A prior uniform distribution for o is appropriate since

$$P(M = p) = \int_0^1 \binom{N}{p} o^p (1-o)^{N-p} do = \frac{1}{N+1}$$

Early form of Bayesian theorem

If E_1 and E_2 are any events, ordered in time, then

$$P(E_1|E_2) = \frac{P(E_1 \bigcap E_2)}{P(E_2)}$$

equivalently

$$P(E_1|E_2) = \frac{P(E_2|E_1)\, P(E_1)}{P(E_2)}$$

Bayesian theorem example

To facilitate early detection of breast cancer [3], women are screened using mammography, even if they have no obvious symptoms

- The probability that one of these women has breast cancer is 1%
- If a woman has breast cancer, the probability is 80% that she will have a positive mammography test
- If a woman does not have breast cancer, the probability is 10% of a positive mammography test

Bayesian theorem example

- Imagine a woman with no symptoms having positive mammography test

Bayesian theorem example

- Imagine a woman with no symptoms having positive mammography test
- What is the probability that she actually has breast cancer?

Bayesian theorem example

- Imagine a woman with no symptoms having positive mammography test
- What is the probability that she actually has breast cancer?
- This question was given to 100 physicians [2]

Bayesian theorem example

- Imagine a woman with no symptoms having positive mammography test
- What is the probability that she actually has breast cancer?
- This question was given to 100 physicians [2]
- Ninety five of the physicians gave the answer of approximately 75% instead of the correct answer, which, in this example, is 7.5%.

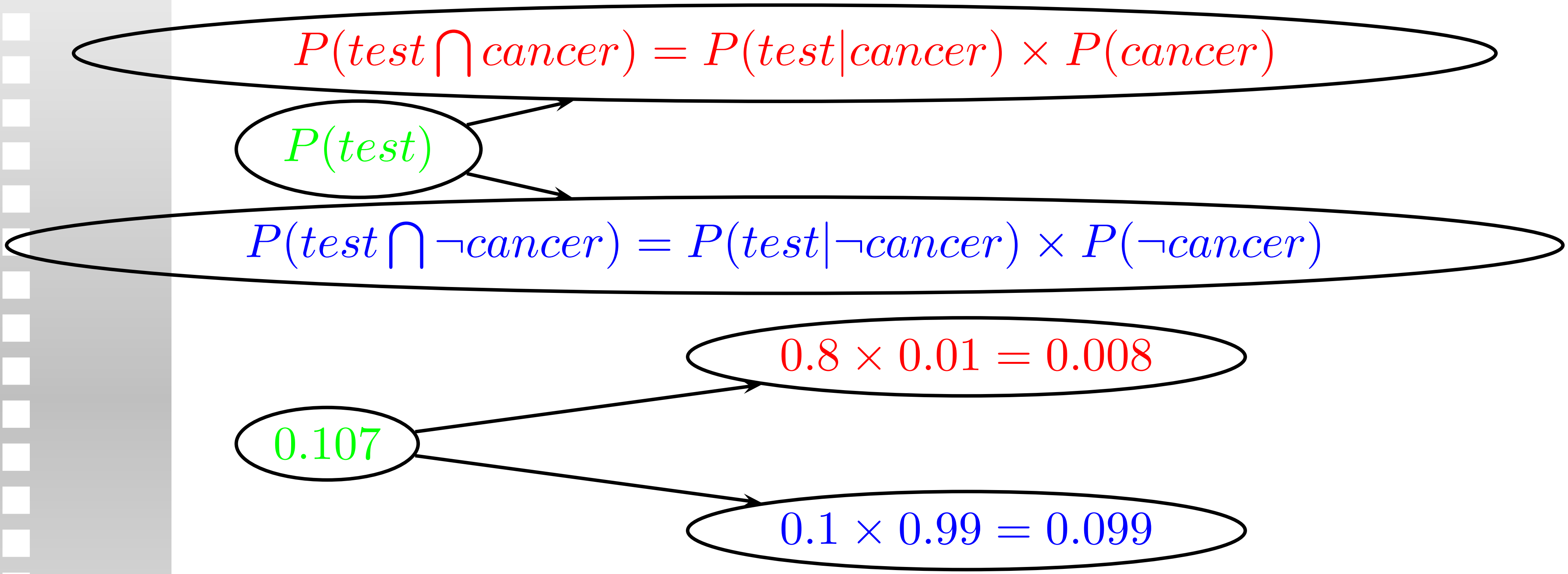
Marginalization of probability

$$P(test \bigcap cancer) = P(test|cancer) \times P(cancer)$$

$$P(test)$$

$$P(test \bigcap \neg cancer) = P(test|\neg cancer) \times P(\neg cancer)$$

Marginalization of probability



Marginalization of probability

$$P(test \bigcap cancer) = P(test|cancer) \times P(cancer)$$

$$P(test)$$

$$P(test \bigcap \neg cancer) = P(test|\neg cancer) \times P(\neg cancer)$$

$$0.8 \times 0.01 = 0.008$$

$$0.107$$

$$0.1 \times 0.99 = 0.099$$

$$P(cancer|test) = \frac{P(cancer) \times P(test|cancer)}{P(cancer) \times P(test|cancer) + P(\neg cancer) \times P(test|\neg cancer)}$$

Marginalization of probability

$$P(test \bigcap cancer) = P(test|cancer) \times P(cancer)$$

$$P(test)$$

$$P(test \bigcap \neg cancer) = P(test|\neg cancer) \times P(\neg cancer)$$

$$0.8 \times 0.01 = 0.008$$

$$0.107$$

$$0.1 \times 0.99 = 0.099$$

$$P(cancer|test) = \frac{P(cancer) \times P(test|cancer)}{P(cancer) \times P(test|cancer) + P(\neg cancer) \times P(test|\neg cancer)}$$

$$\frac{0.01 \times 0.8}{0.01 \times 0.8 + 0.99 \times 0.1} = \frac{0.008}{0.008 + 0.099} = \frac{0.008}{0.107} = 0.075$$

Physicians say...

The article [3] discusses some of the reactions of the physicians to even considering such problems. Here are some quotes:

- *On such a basis one can't make a diagnosis. Statistical information is one big lie.*
- *I never inform my patients about statistical data. I would tell the patient that mammography is not so exact, and I would in any case perform a biopsy.*
- *Oh, what nonsense. I can't do it. You should test my daughter. She studies medicine.*
- *Statistics is alien to everyday concerns and of little use for judging individual persons.*

Proposition 9

$$P((f < o < b)|(M = p)) = \frac{P((f < o < b) \bigcap (M = p))}{P(M = p)}$$

If I give a guess concerning o, the probability I am correct is the ratio of the part of AiB between f and b, to all AiB

$$P((f < o < b)|(M = p)) = \frac{\int_f^b \binom{N}{p} o^p (1-o)^{N-p} do}{\int_0^1 \binom{N}{p} o^p (1-o)^{N-p} do}$$

Why Bayes didn't publish?

Not satisfied with evaluation of

$$\int_f^b o^p (1-o)^{N-p} do$$

Known as the *Incomplete Beta function*, failed to give a general solution to the problem

As we saw

$$\int_0^1 \binom{N}{p} o^p (1-o)^{N-p} do = \frac{1}{N+1}$$

Beta function

$$\int_0^1 o^p(1-o)^{N-p}do = \frac{p!(N-p)!}{(N+1)!}$$

Formula for *Beta function* according to definition

$$\begin{aligned} B(r,s) &= \int_0^1 p^{r-1}(1-p)^{s-1}dp \\ &= \frac{(r-1)!(s-1)!}{(r+s-1)!} = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)} \end{aligned}$$

Here $\Gamma(n) = (n-1)!$ is *Gamma function*

Beta distribution

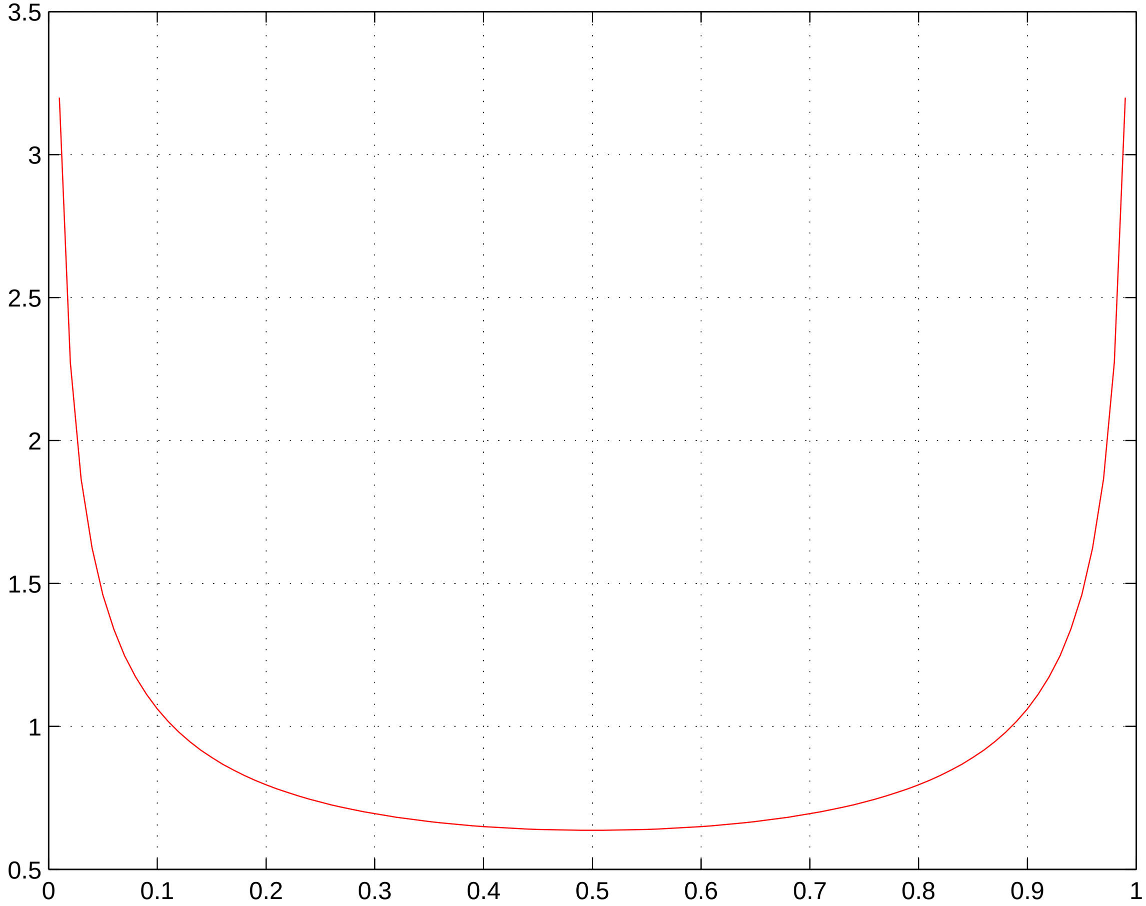
$$P((f < o < b)|(M = p)) = \frac{\int_f^b \binom{N}{p} o^p (1-o)^{N-p} do}{B(p+1, N-p+1)}$$

which is a *Beta distribution* of p with parameters $p+1$ and $N-p+1$

The density function f of the *Beta distribution* is

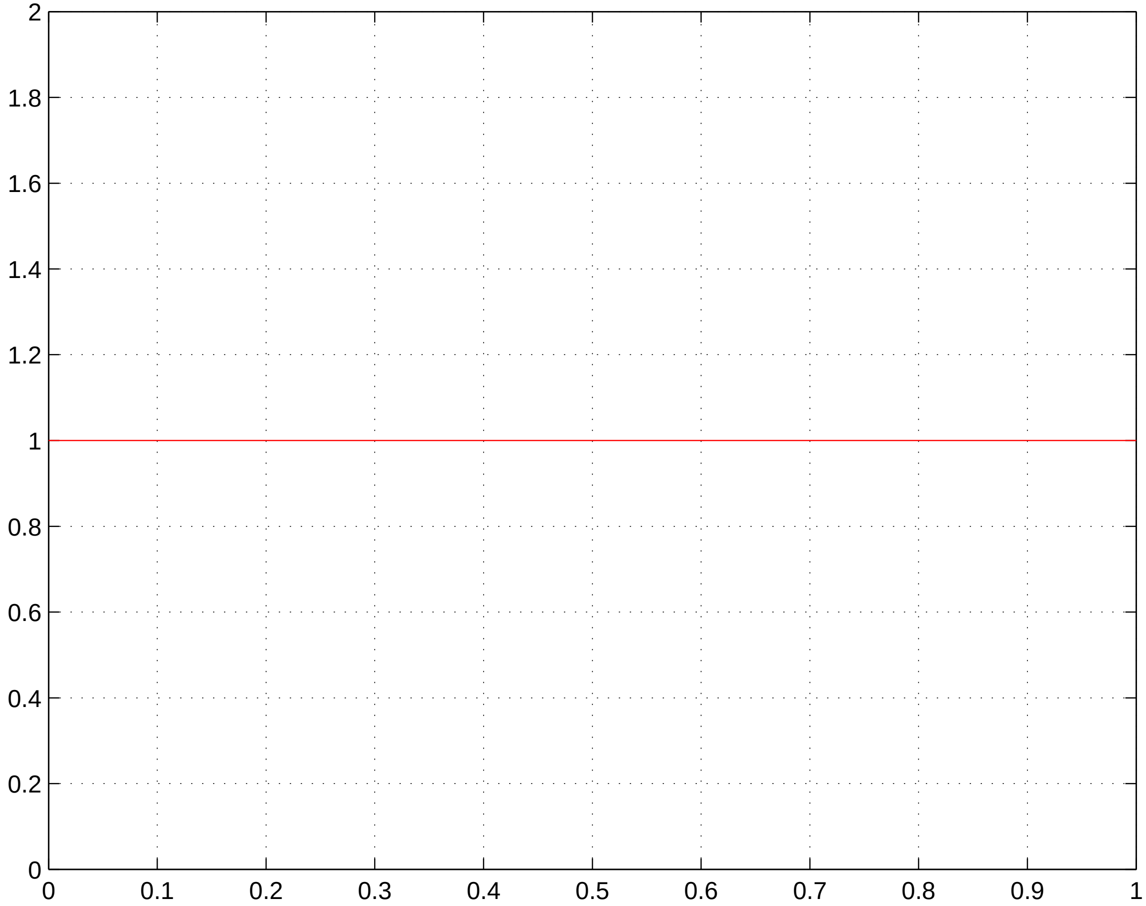
$$f(p) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1}, \qquad 0 \leq p \leq 1$$

Beta priors example



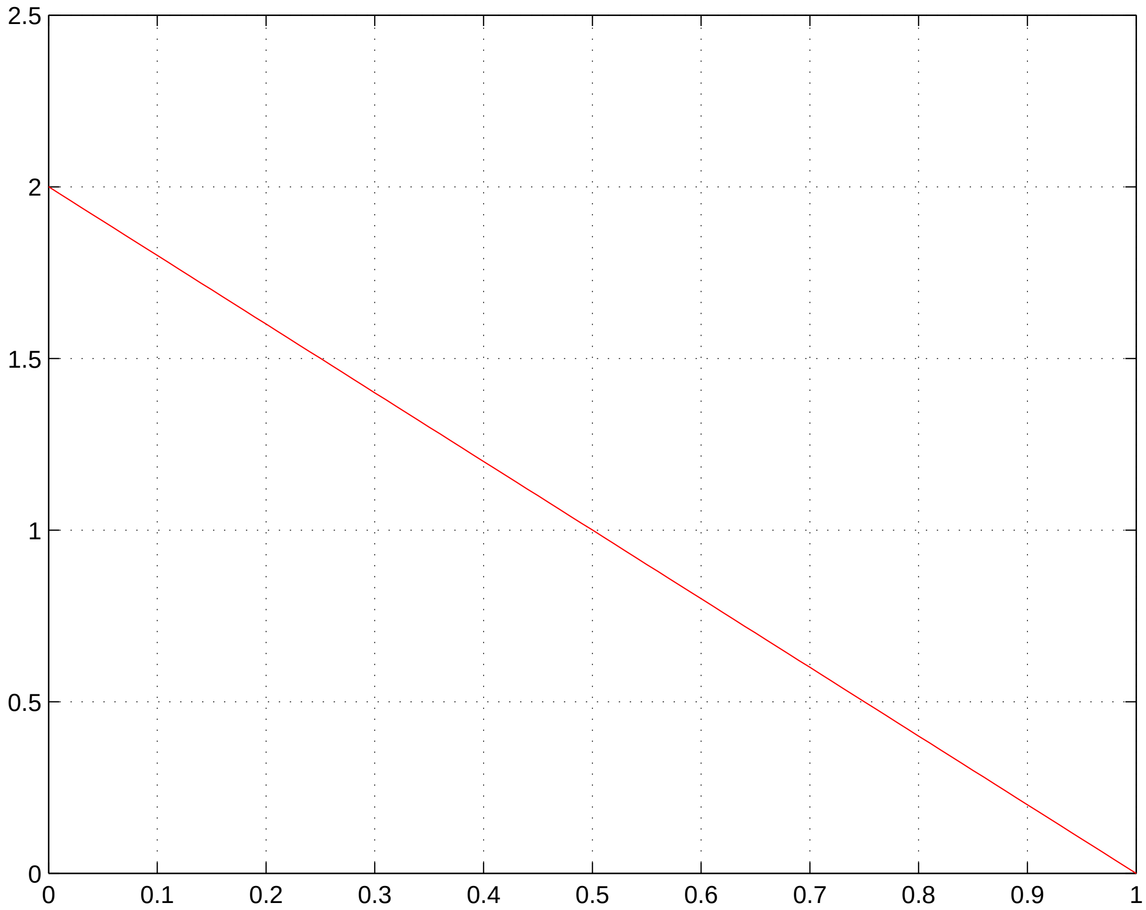
$$Beta(0.5, 0.5)$$

Beta priors example



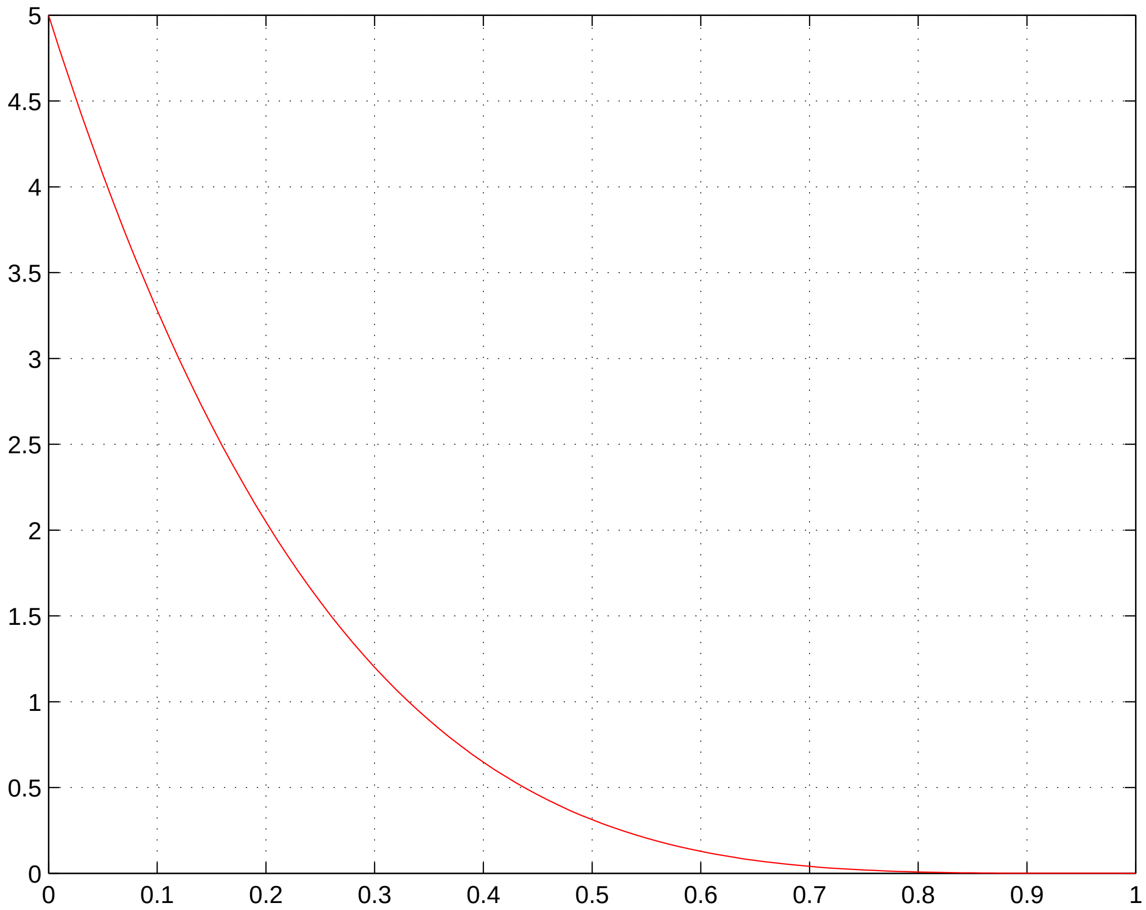
$Beta(1, 1)$

Beta priors example



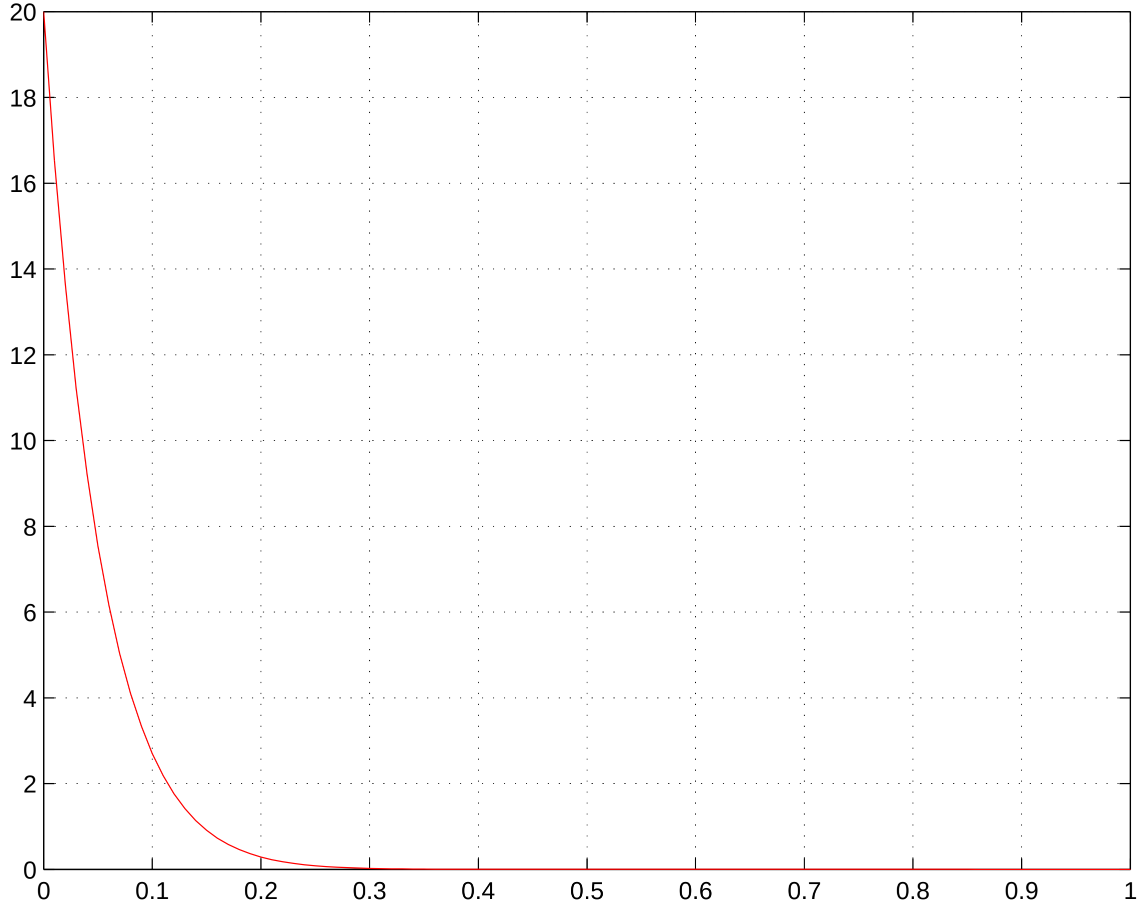
$Beta(1, 2)$

Beta priors example



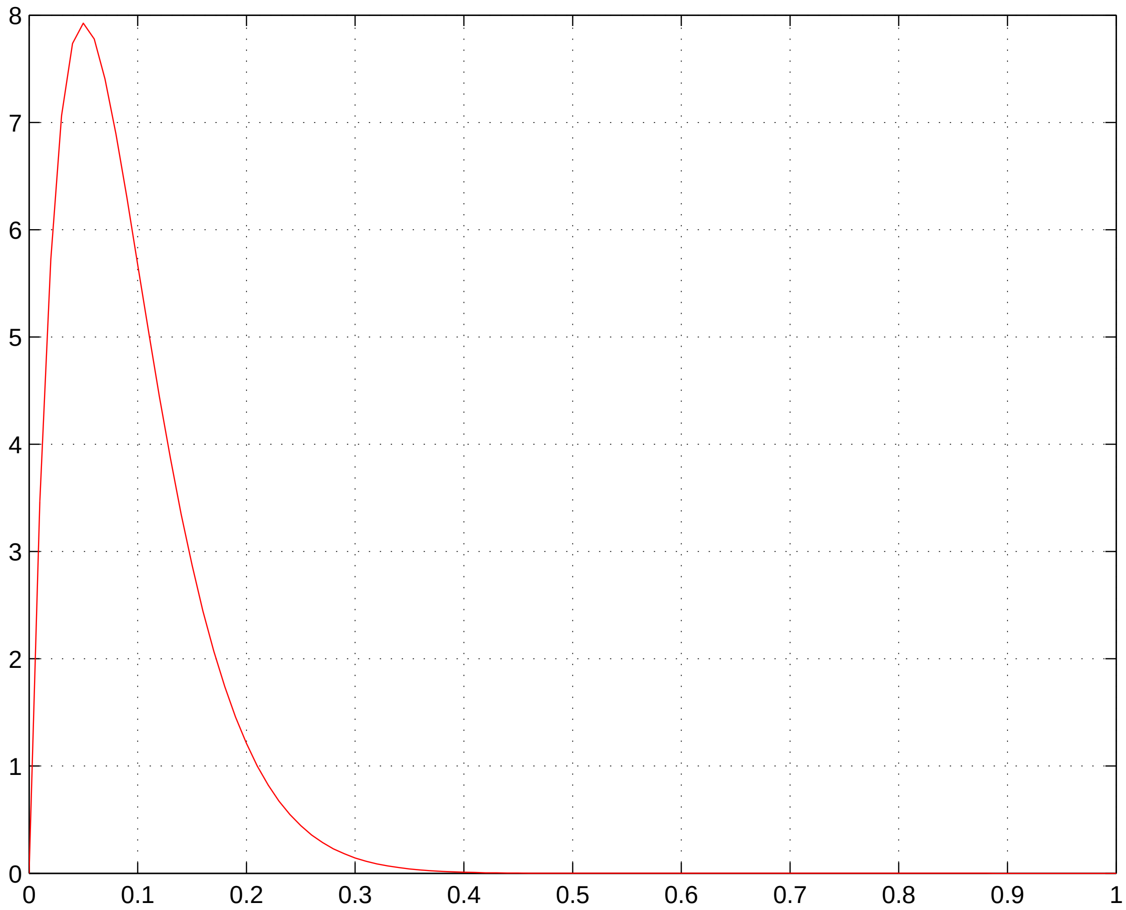
$Beta(1,5)$

Beta priors example



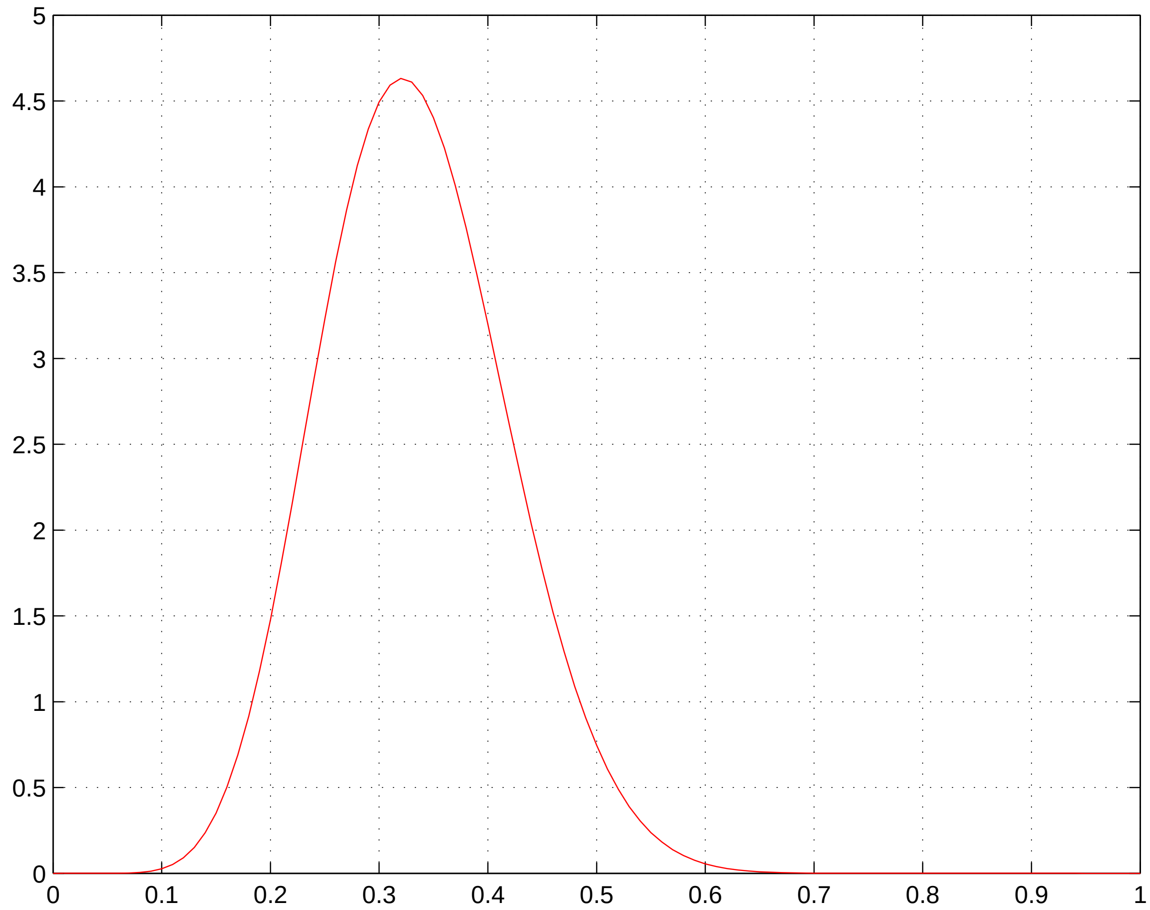
$Beta(1, 20)$

Beta priors example



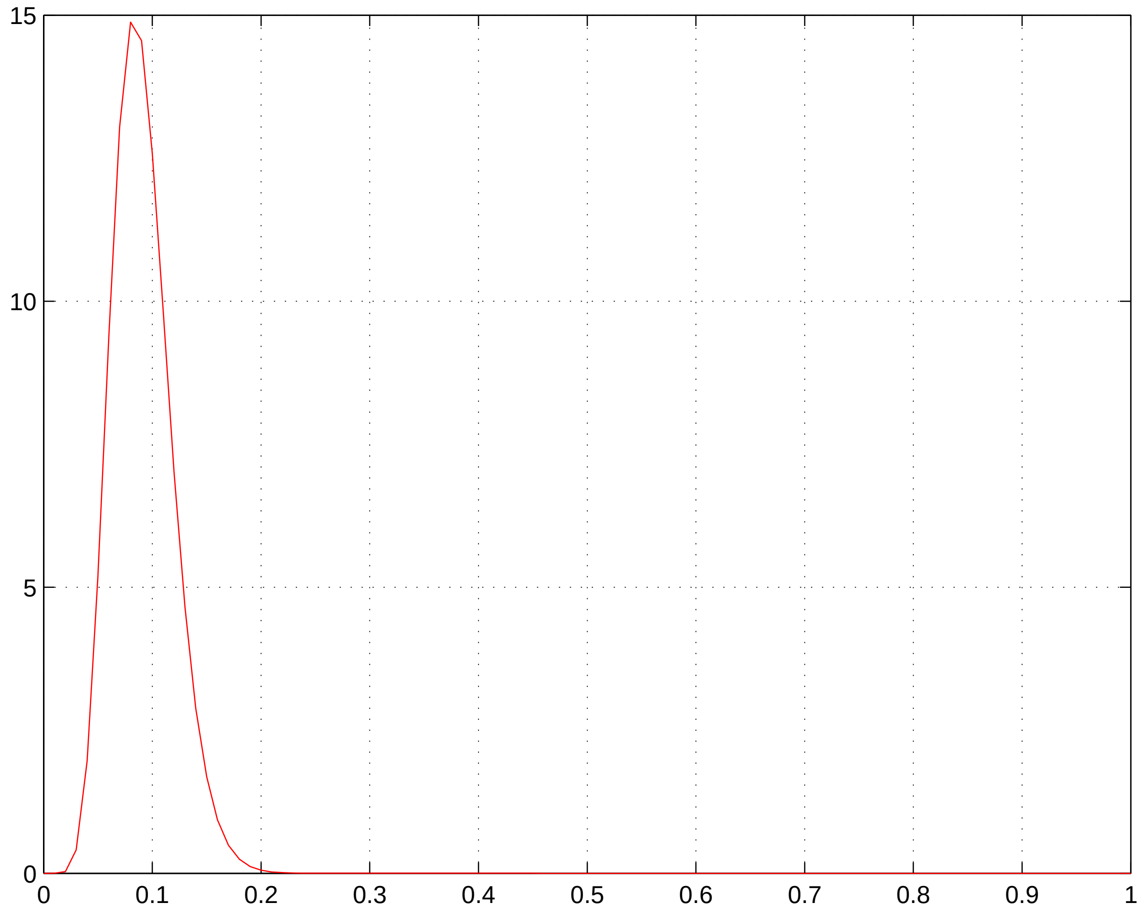
$Beta(2, 20)$

Beta priors example



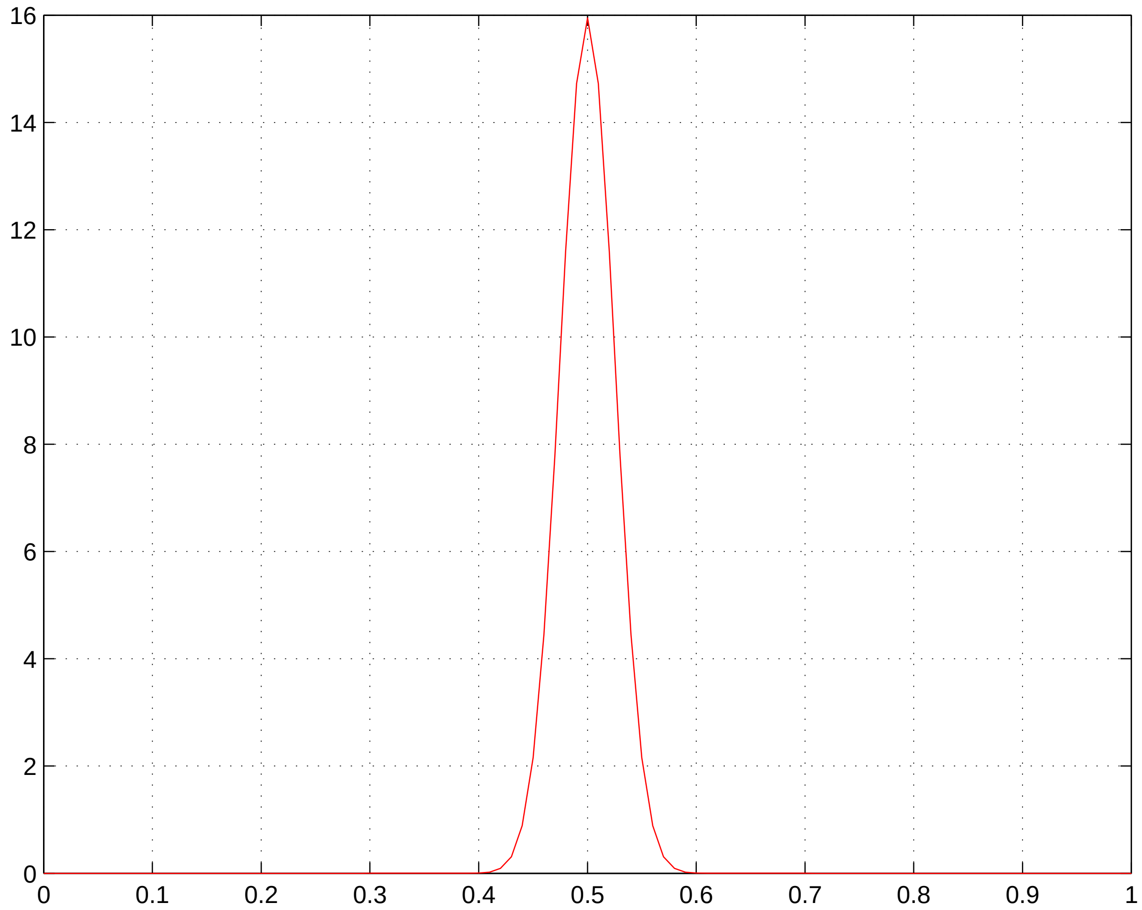
$Beta(10, 20)$

Beta priors example



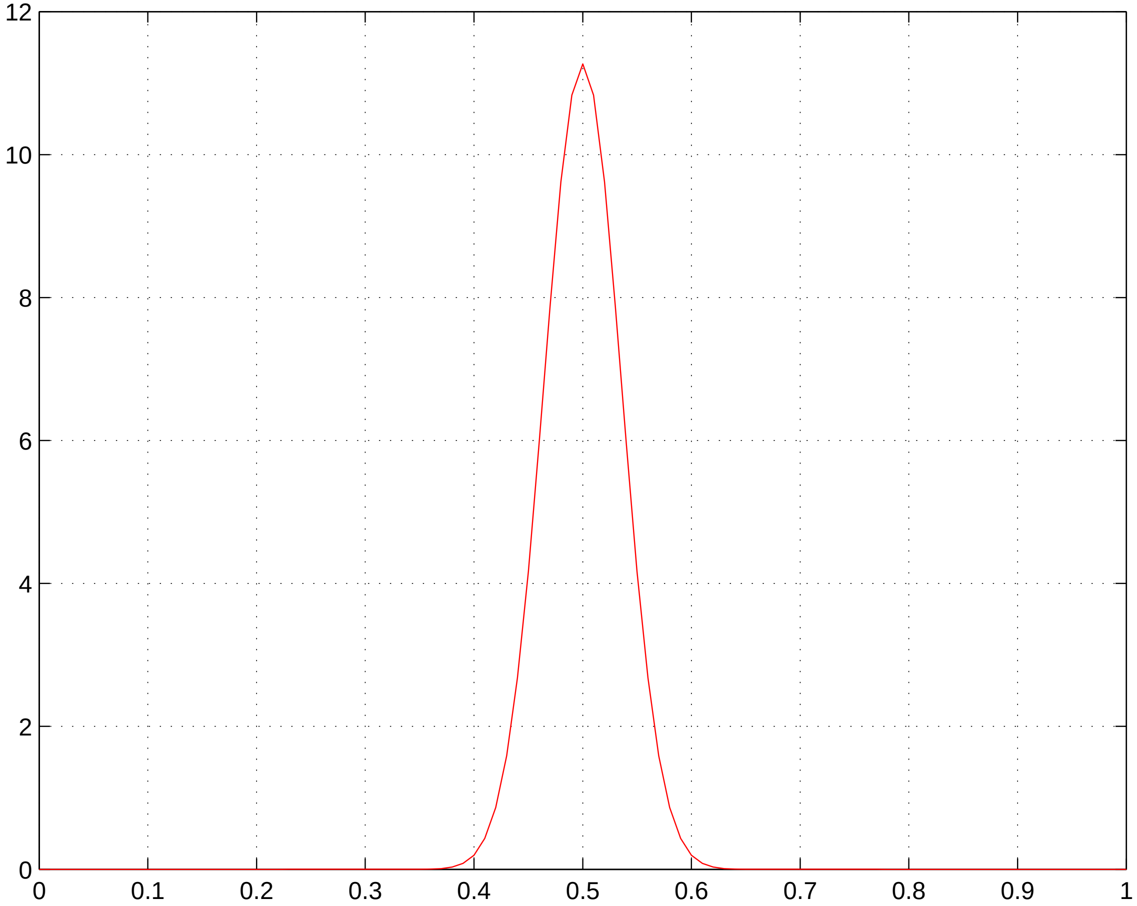
$Beta(10, 100)$

Beta priors example



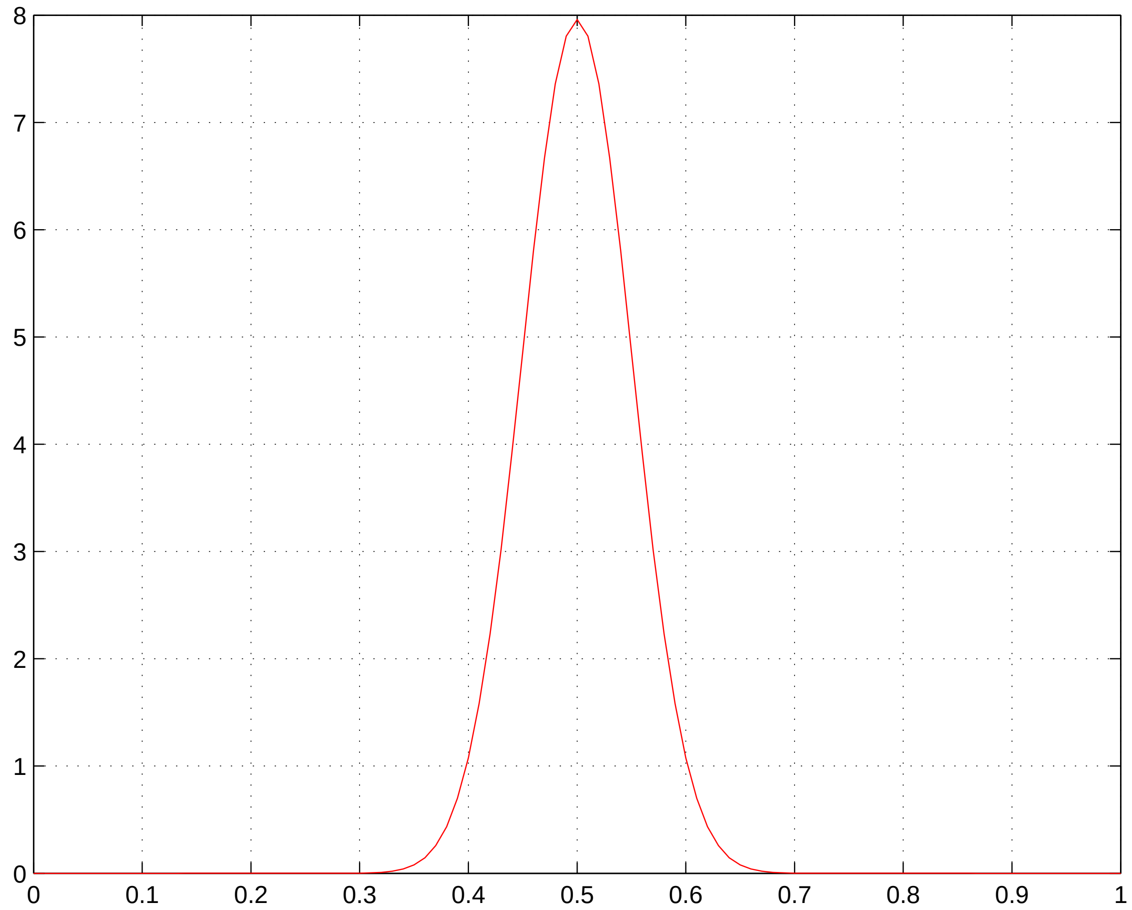
$Beta(200, 200)$

Beta priors example



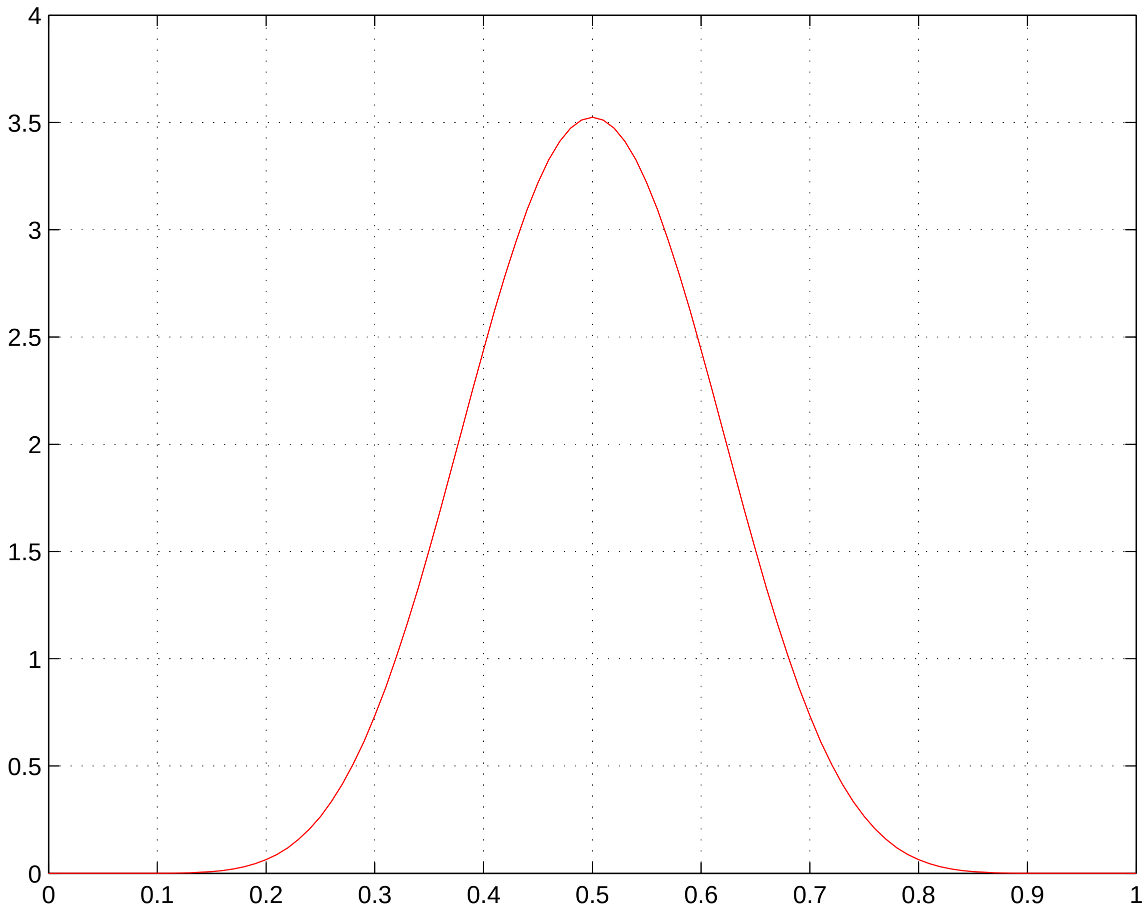
$$Beta(100, 100)$$

Beta priors example



$Beta(50, 50)$

Beta priors example



$Beta(10, 10)$

Beta conjugate prior

Suppose that the prior distribution of p is $Beta(a, b)$, i.e.

$$g(p) = \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)}$$

Likelihood has a binomial distribution

$$f(x|p) = \binom{n}{x} p^x (1-p)^{n-x}$$

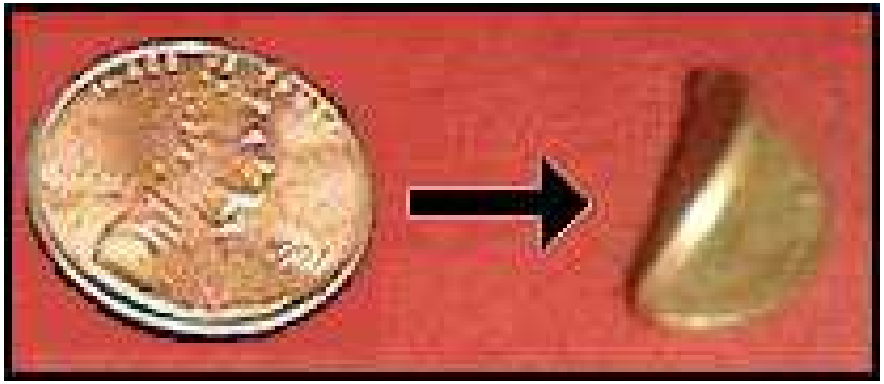
Beta conjugate prior

The posterior distribution of p given x is

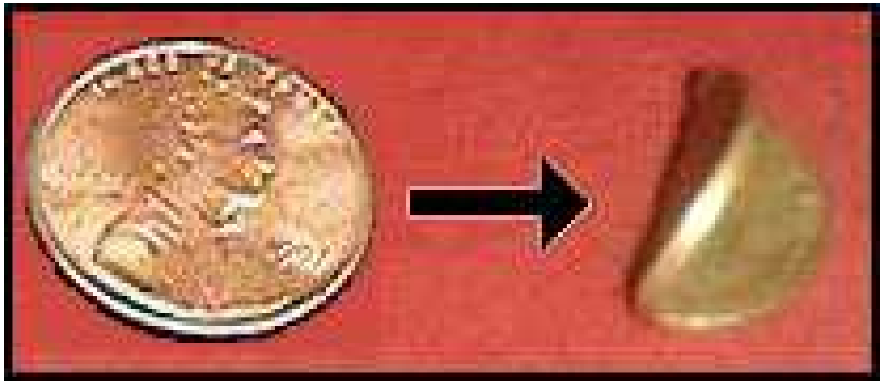
$$\begin{aligned} h(p|x) &= \frac{f(x|p)g(p)}{f(x)} \\ &= \frac{B(a,b)\binom{n}{x}p^x(1-p)^{n-x}p^{a-1}(1-p)^{b-1}}{B(a,b)\int_0^1 \binom{n}{x}p^{x+a-1}(1-p)^{n-x+b-1}dp} \\ &= \frac{p^{a+x-1}(1-p)^{n+b-x-1}}{B(a+x,n+b-x)} \\ &= Beta(a+x,n+b-x) \end{aligned}$$

This distribution is thus beta as well with parameters $a' = a + x$ and $b' = b + n - x$.

Bent coin example

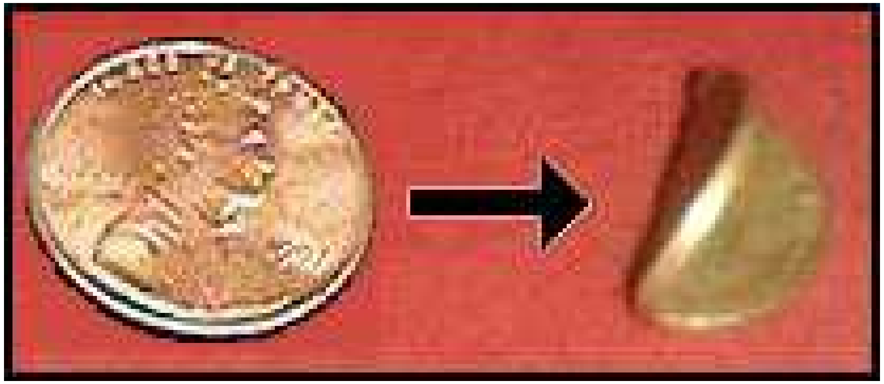


Bent coin example



The coin is severely bent, but on which side is not known

Bent coin example

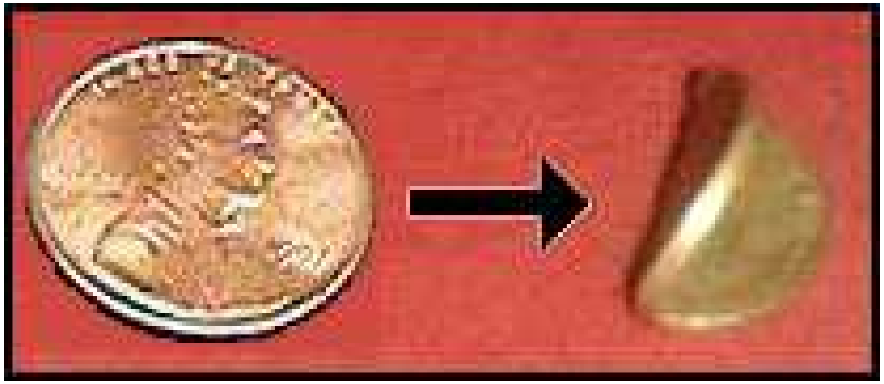


The coin is severely bent, but on which side is not known

In this case we can use mixture of betas to express our prior belief about properties of this coin

$$g(p) = 0.5 \times Beta(10, 30) + 0.5 \times Beta(30, 10)$$

Bent coin example



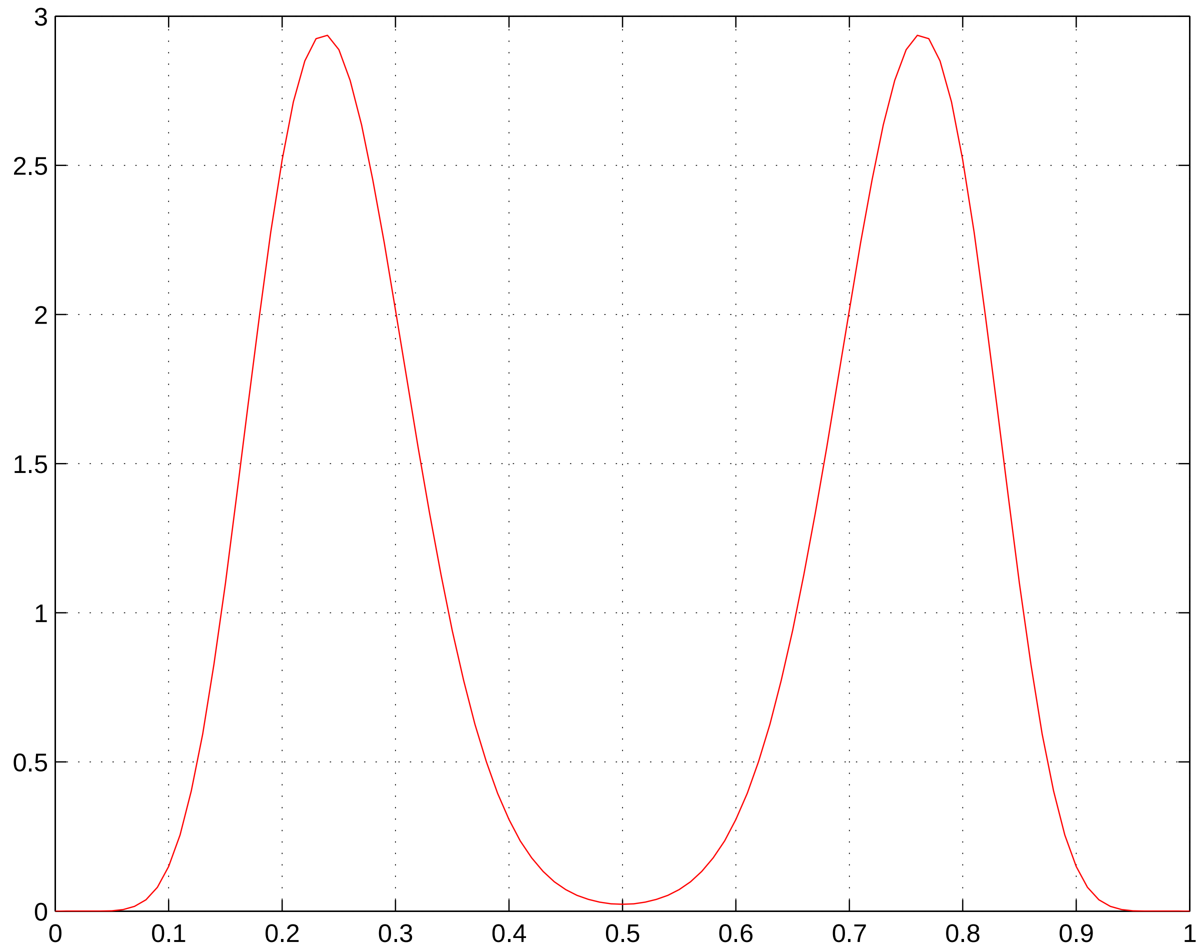
The coin is severely bent, but on which side is not known

In this case we can use mixture of betas to express our prior belief about properties of this coin

$$g(p) = 0.5 \times Beta(10, 30) + 0.5 \times Beta(30, 10)$$

The coin tossing experiment return 33 tails and 102 heads

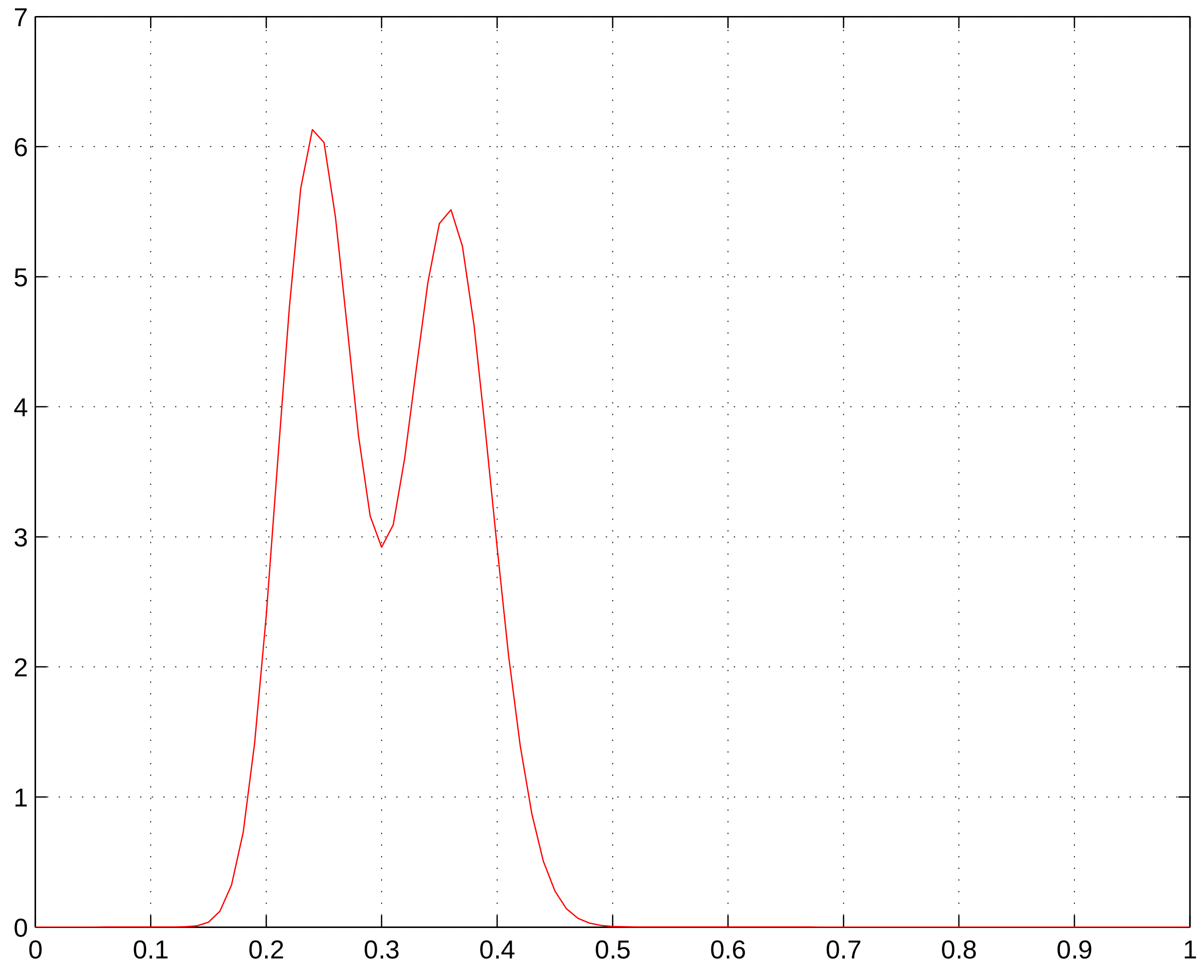
Beta mixture prior



Beta mixture prior

$$g(p) = 0.5 \times Beta(10, 30) + 0.5 \times Beta(30, 10)$$

Beta mixture prior



Beta mixture posterior

$$g(p) = 0.5 \times Beta(43, 132) + 0.5 \times Beta(63, 112)$$

Dirichlet prior

Dirichlet prior Dirichlet prior is conjugate to multinomial distribution. The Dirichlet distribution can be written as

$$D(\Theta|\alpha) = \frac{\Gamma(\sum_{i=1}^{K} \alpha_i)}{\prod_{i=1}^{K} \Gamma(\alpha_i)} \prod_{i=1}^{K} \Theta_i^{\alpha_i - 1}$$

Θ_i satisfy $0 \leq \Theta_i \leq 1$ and $\sum_{i=1}^{K} \Theta_i = 1$

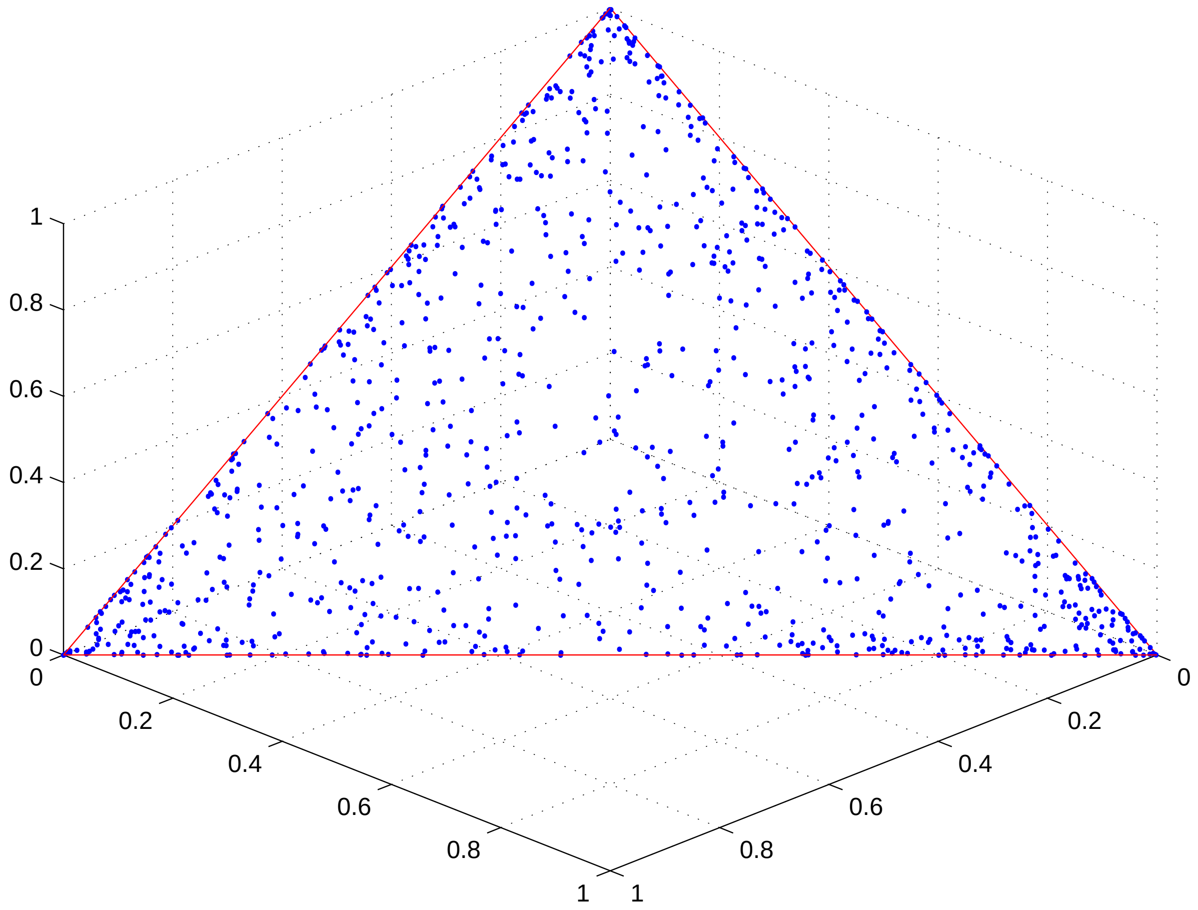
Multinomial distribution

The multinomial distribution corresponding to k balls dropped into n boxes with fixed probability $(p_1, \ldots, p_n)$, (with i^{th} box containing k_i balls) is

$$\binom{k}{k_1, \ldots, k_n} p_1^{k_1} \cdots p_n^{k_n} = \frac{k!}{k_1! \cdots k_n!} p_1^{k_1} \cdots p_n^{k_n}$$

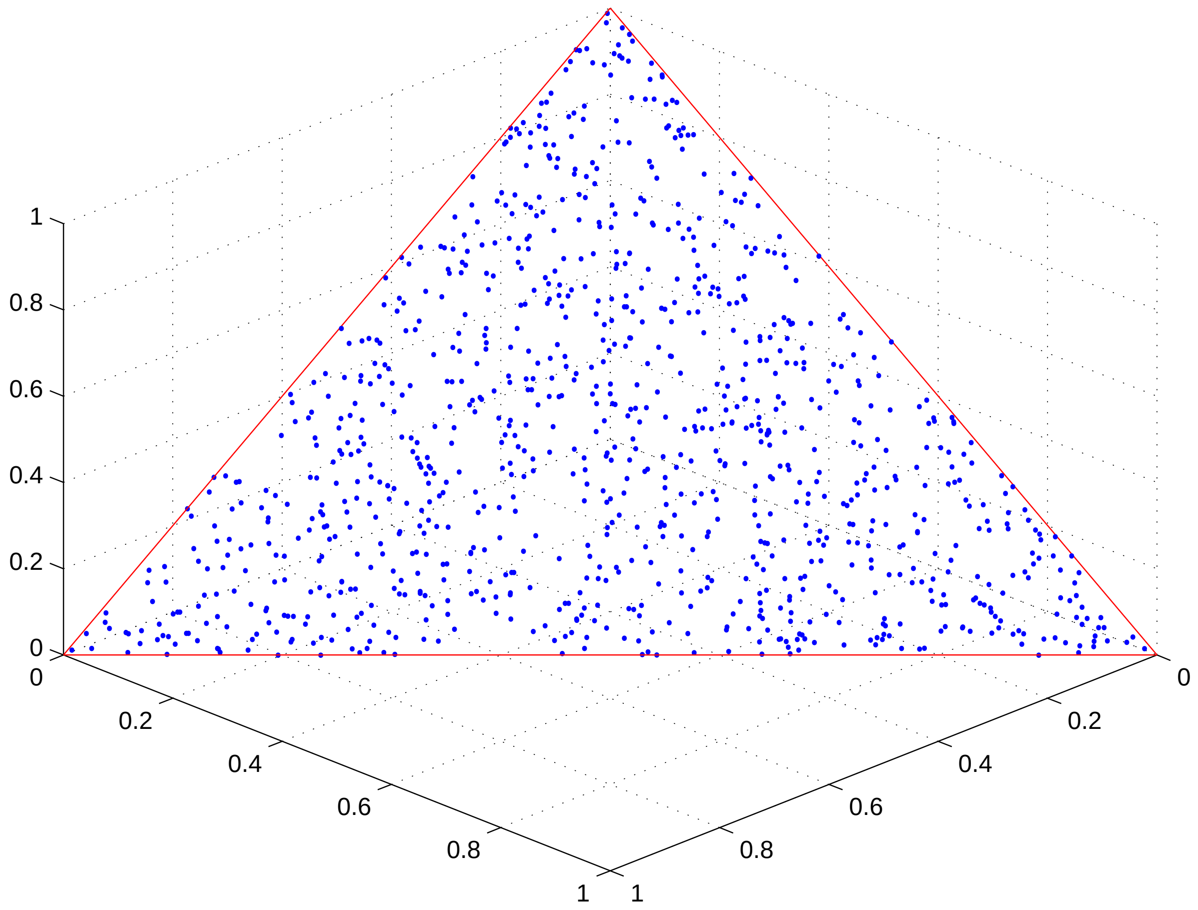
The Dirichlet is a convenient prior because the posterior having observed frequencies $(k_1, \ldots, k_n)$ is Dirichlet with probability $(\alpha_1 + k_1, \ldots, \alpha_n + k_n)$.

Dirichlet priors example



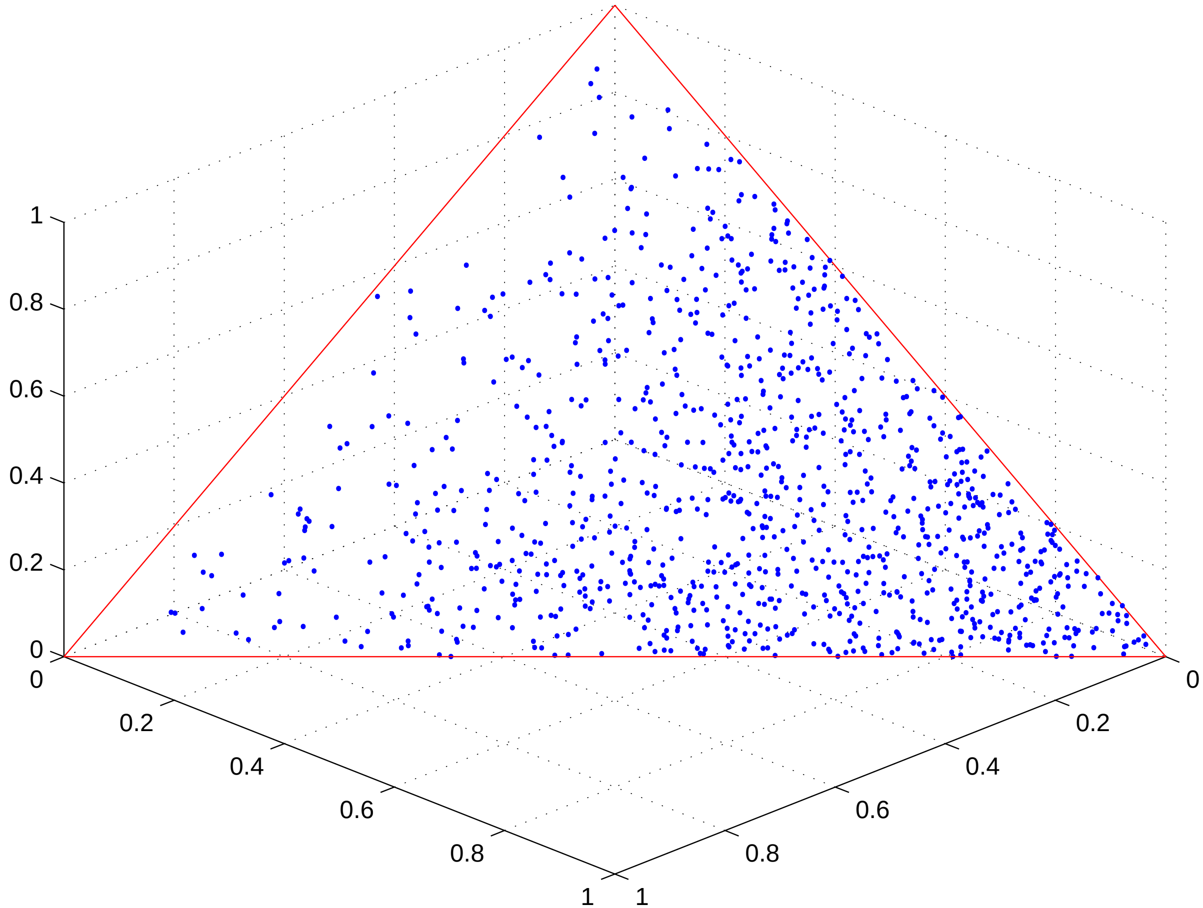
$$Dirichlet(0.5, 0.5, 0.5)$$

Dirichlet priors example



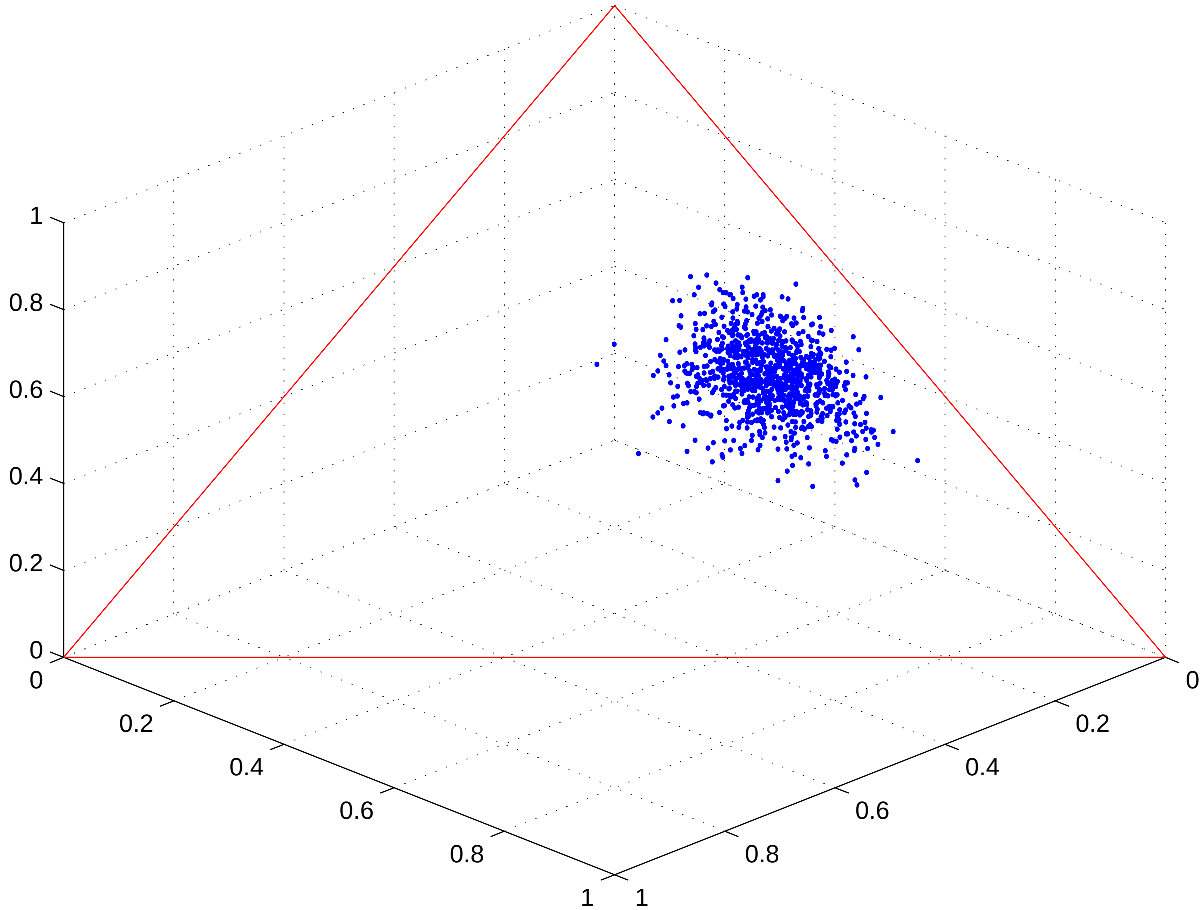
$$Dirichlet(1, 1, 1)$$

Dirichlet priors example



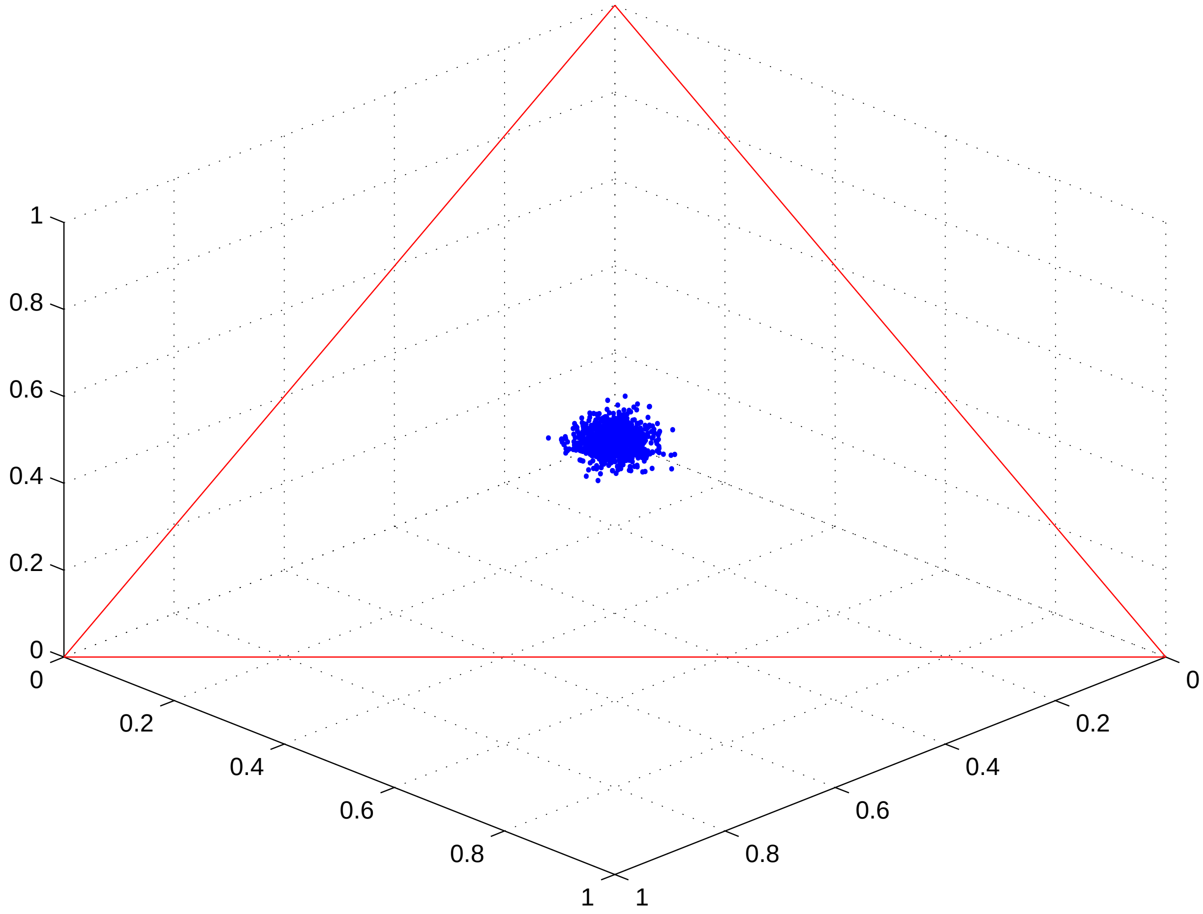
$$Dirichlet(1, 2, 1)$$

Dirichlet priors example



$$Dirichlet(10, 30, 30)$$

Dirichlet priors example



$$Dirichlet(200, 200, 200)$$

References

[1] Rev. Thomas Bayes, *An essay towards solving a problem in the doctrine of chances*, Philosophical Transactions of the Royal Society of London (1763).

[2] D. M. Eddy, *Variations in phsician practice: the role of uncertainty*, Professional Judgment: A Reader in Clinical Decision Making (A. Elstein and J. Dowie, eds.), Cambridge University Press, 1988, pp. 45–590.

[3] Gerd Gigerenzer, *The psychology of good judgment*, Medical Decision Making **16** (1996), no. 3, 273–280.