Choosing Priors
Probability Intervals

18.05 Spring 2014

Conjugate priors

A prior is conjugate to a likelihood if the posterior is the same type of distribution as the prior.
Updating becomes algebra instead of calculus.

	hypothesis	data	prior	likelihood	posterior
Bernoulli/Beta	$\theta \in [0,1]$	x	$\text{beta}(a,b)$	$\text{Bernoulli}(\theta)$	$\text{beta}(a+1,b)$ or $\text{beta}(a,b+1)$
	θ	$x=1$	$c_1\theta^{a-1}(1-\theta)^{b-1}$	θ	$c_3\theta^{a}(1-\theta)^{b-1}$
	θ	$x=0$	$c_1\theta^{a-1}(1-\theta)^{b-1}$	$1-\theta$	$c_3\theta^{a-1}(1-\theta)^{b}$
Binomial/Beta	$\theta \in [0,1]$	x	$\text{beta}(a,b)$	$\text{binomial}(N,\theta)$	$\text{beta}(a+x,b+N-x)$
(fixed N)	θ	x	$c_1\theta^{a-1}(1-\theta)^{b-1}$	$c_2\theta^{x}(1-\theta)^{N-x}$	$c_3\theta^{a+x-1}(1-\theta)^{b+N-x-1}$
Geometric/Beta	$\theta \in [0,1]$	x	$\text{beta}(a,b)$	$\text{geometric}(\theta)$	$\text{beta}(a+x,b+1)$
	θ	x	$c_1\theta^{a-1}(1-\theta)^{b-1}$	$\theta^{x}(1-\theta)$	$c_3\theta^{a+x-1}(1-\theta)^{b}$
Normal/Normal	$\theta \in (-\infty,\infty)$	x	$\text{N}(\mu_{\text{prior}},\sigma^2_{\text{prior}})$	$\text{N}(\theta,\sigma^2)$	$\text{N}(\mu_{\text{post}},\sigma^2_{\text{post}})$
(fixed σ^2)	θ	x	$c_1\exp\left(\frac{-(\theta-\mu_{\text{prior}})^2}{2\sigma^2_{\text{prior}}}\right)$	$c_2\exp\left(\frac{-(x-\theta)^2}{2\sigma^2}\right)$	$c_3\exp\left(\frac{(\theta-\mu_{\text{post}})^2}{2\sigma^2_{\text{post}}}\right)$

There are many other likelihood/conjugate prior pairs.

Concept question: conjugate priors

Which are conjugate priors?

	hypothesis	data	prior	likelihood
a) Exponential/Normal	$\theta \in [0, \infty)$	x	$\mathrm{N}(\mu_{\text{prior}}, \sigma^2_{\text{prior}})$	$\exp(\theta)$
	θ	x	$c_1 \exp\left(-\frac{(\theta-\mu_{\text{prior}})^2}{2\sigma^2_{\text{prior}}}\right)$	$\theta \mathrm{e}^{-\theta x}$
b) Exponential/Gamma	$\theta \in [0, \infty)$	x	$\mathrm{Gamma}(a, b)$	$\exp(\theta)$
	θ	x	$c_1 \theta^{a-1} \mathrm{e}^{-b\theta}$	$\theta \mathrm{e}^{-\theta x}$
c) Binomial/Normal	$\theta \in [0, 1]$	x	$\mathrm{N}(\mu_{\text{prior}}, \sigma^2_{\text{prior}})$	$\mathrm{binomial}(N, \theta)$
(fixed N)	θ	x	$c_1 \exp\left(-\frac{(\theta-\mu_{\text{prior}})^2}{2\sigma^2_{\text{prior}}}\right)$	$c_2\, \theta^x (1-\theta)^{N-x}$

1. none 2. a 3. b 4. c
5. a,b 6. a,c 7. b,c 8. a,b,c

Answer: 3. b

We have a conjugate prior if the posterior *as a function of* θ has the same form as the prior.

Exponential/Normal posterior:

$$f(\theta|x) = c_1 \theta \mathrm{e}^{-\frac{(\theta-\mu_{\text{prior}})^2}{2\sigma_{\text{prior}}^2} - \theta x}$$

The factor of θ before the exponential means this is not the pdf of a normal distribution. Therefore it is not a conjugate prior.

Exponential/Gamma posterior: Note, we have never learned about Gamma distributions, but it doesn't matter. We only have to check if the posterior has the same form:

$$f(\theta|x) = c_1 \theta^a \mathrm{e}^{-(b+x)\theta}$$

The posterior has the form Gamma$(a+1, b+x)$. This is a conjugate prior.

Binomial/Normal: It is clear that the posterior does not have the form of a normal distribution.

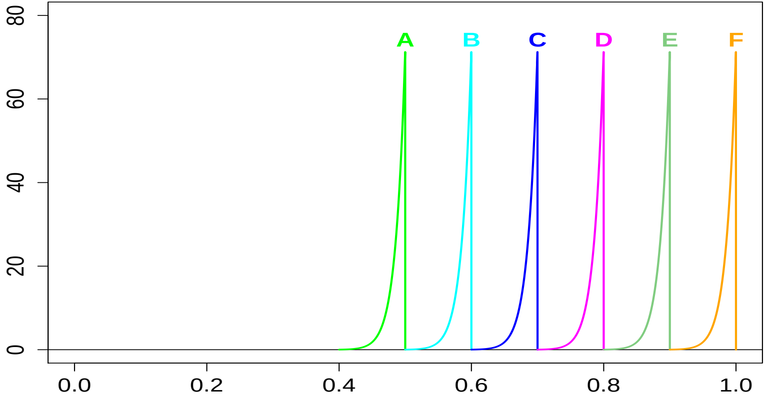
Concept question: strong priors

Say we have a bent coin with unknown probability of heads θ.

We are convinced that $\theta \leq 0.7$.
Our prior is uniform on $[0, 0.7]$ and 0 from 0.7 to 1.

We flip the coin 65 times and get 60 heads.

Which of the graphs below is the posterior pdf for θ?



Solution to concept question

<u>**answer:**</u> Graph C, the blue graph spiking near 0.7.

Sixty heads in 65 tosses indicates the true value of θ is close to 1. Our prior was 0 for $\theta > 0.7$. So no amount of data will make the posterior non-zero in that range. That is, we have forclosed on the possibility of deciding that θ is close to 1. The Bayesian updating puts θ near the top of the allowed range.

Two parameter tables: Malaria

In the 1950's scientists injected 30 African "volunteers" with malaria.

S = carrier of sickle-cell gene

N = non-carrier of sickle-cell gene

$D+$ = developed malaria

$D-$ = did not develop malaria

	$D+$	$D-$	
S	2	13	15
N	14	1	15
	16	14	30

Model

$\theta_S =$ probability an injected S develops malaria.

$\theta_N =$ probabilitiy an injected N develops malaria.

Assume conditional independence between all the experimental subjects.

Likelihood is a function of both θ_S and θ_N:

$$P(\text{data}|\theta_S, \theta_N) = c\,\theta_S^2(1-\theta_S)^{13}\theta_N^{14}(1-\theta_N).$$

Hypotheses: pairs (θ_S, θ_N).

Finite number of hypotheses. θ_S and θ_N are each one of $0, .2, .4, .6, .8, 1$.

Color-coded two-dimensional tables

Hypotheses

$\theta_N \backslash \theta_S$	0	0.2	0.4	0.6	0.8	1
1	(0,1)	(.2,1)	(.4,1)	(.6,1)	(.8,1)	(1,1)
0.8	(0,.8)	(.2,.8)	(.4,.8)	(.6,.8)	(.8,.8)	(1,.8)
0.6	(0,.6)	(.2,.6)	(.4,.6)	(.6,.6)	(.8,.6)	(1,.6)
0.4	(0,.4)	(.2,.4)	(.4,.4)	(.6,.4)	(.8,.4)	(1,.4)
0.2	(0,.2)	(.2,.2)	(.4,.2)	(.6,.2)	(.8,.2)	(1,.2)
0	(0,0)	(.2,0)	(.4,0)	(.6,0)	(.8,0)	(1,0)

Table of hypotheses for (θ_S, θ_N)

Corresponding level of protection due to S:

red = strong, pink = some, orange = none,

white = negative.

Color-coded two-dimensional tables

Likelihoods (scaled to make the table readable)

$\theta_N \backslash \theta_S$	0	0.2	0.4	0.6	0.8	1
1	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.8	0.00000	1.93428	0.18381	0.00213	0.00000	0.00000
0.6	0.00000	0.06893	0.00655	0.00008	0.00000	0.00000
0.4	0.00000	0.00035	0.00003	0.00000	0.00000	0.00000
0.2	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Likelihoods scaled by $100000/c$

$$p(\text{data}|\theta_S, \theta_N) = c\,\theta_S^2(1-\theta_S)^{13}\theta_N^{14}(1-\theta_N).$$

Color-coded two-dimensional tables

Flat prior

$\theta_N \backslash \theta_S$	0	0.2	0.4	0.6	0.8	1	$p(\theta_N)$
1	1/36	1/36	1/36	1/36	1/36	1/36	1/6
0.8	1/36	1/36	1/36	1/36	1/36	1/36	1/6
0.6	1/36	1/36	1/36	1/36	1/36	1/36	1/6
0.4	1/36	1/36	1/36	1/36	1/36	1/36	1/6
0.2	1/36	1/36	1/36	1/36	1/36	1/36	1/6
0	1/36	1/36	1/36	1/36	1/36	1/36	1/6
$p(\theta_S)$	1/6	1/6	1/6	1/6	1/6	1/6	1

Flat prior $p(\theta_S, \theta_N)$: each hypothesis (square) has equal probability

Color-coded two-dimensional tables

Posterior to the flat prior

$\theta_N \backslash \theta_S$	0	0.2	0.4	0.6	0.8	1	$p(\theta_N \mid \text{data})$
1	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.8	0.00000	0.88075	0.08370	0.00097	0.00000	0.00000	0.96542
0.6	0.00000	0.03139	0.00298	0.00003	0.00000	0.00000	0.03440
0.4	0.00000	0.00016	0.00002	0.00000	0.00000	0.00000	0.00018
0.2	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
$p(\theta_S \mid \text{data})$	0.00000	0.91230	0.08670	0.00100	0.00000	0.00000	1.00000

Normalized posterior to the flat prior: $p(\theta_S, \theta_N \mid \text{data})$

Strong protection: $P(\theta_N - \theta_S > .5 \mid \text{data}) = \text{sum of red} = .88075$

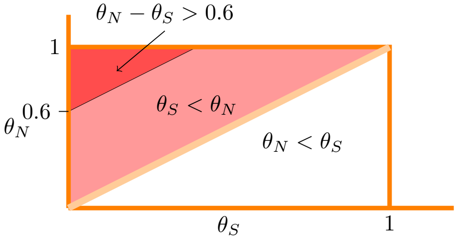
Some protection: $P(\theta_N > \theta_S \mid \text{data}) = \text{sum pink and red} = .99995$

Continuous two-parameter distributions

Sometimes continuous parameters are more natural.

Malaria example (from class notes):
  discrete prior table from the class notes.
  Similarly colored version for the continuous parameters (θ_S, θ_N) over range $[0,1] \times [0,1]$.

$\theta_N \backslash \theta_S$	0	0.2	0.4	0.6	0.8	1
1	(0,1)	(.2,1)	(.4,1)	(.6,1)	(.8,1)	(1,1)
0.8	(0,.8)	(.2,.8)	(.4,.8)	(.6,.8)	(.8,.8)	(1,.8)
0.6	(0,.6)	(.2,.6)	(.4,.6)	(.6,.6)	(.8,.6)	(1,.6)
0.4	(0,.4)	(.2,.4)	(.4,.4)	(.6,.4)	(.8,.4)	(1,.4)
0.2	(0,.2)	(.2,.2)	(.4,.2)	(.6,.2)	(.8,.2)	(1,.2)
0	(0,0)	(.2,0)	(.4,0)	(.6,0)	(.8,0)	(1,0)



The probabilities are given by double integrals over regions.

Treating severe respiratory failure*

*Adapted from *Statistics a Bayesian Perspective* by Donald Berry

Two treatments for newborns with severe respiratory failure.

1. CVT: conventional therapy (hyperventilation and drugs)

2. ECMO: extracorporeal membrane oxygenation (invasive procedure)

In 1983 in Michigan:

$19/19$ ECMO babies survived and $0/3$ CVT babies survived.

Later Harvard ran a randomized study:

$28/29$ ECMO babies survived and $6/10$ CVT babies survived.

Board question: updating two parameter priors

Michigan: 19/19 ECMO babies and 0/3 CVT babies survived.

Harvard: 28/29 ECMO babies and 6/10 CVT babies survived.

$\theta_E =$ probability that an ECMO baby survives
$\theta_C =$ probability that a CVT baby survives

Consider the values 0.125, 0.375, 0.625, 0.875 for θ_E and θ_S

1. Make the 4×4 prior table for a flat prior.
2. Based on the Michigan results, create a reasonable informed prior table for analyzing the Harvard results (unnormalized is fine).
3. Make the likelihood table for the Harvard results.
4. Find the posterior table for the informed prior.
5. Using the informed posterior, compute the probability that ECMO is better than CVT.
6. Also compute the posterior probability that $\theta_E - \theta_C \geq 0.6$.
(The posted solutions will also show 4-6 for the flat prior.)

Solution

Flat prior

		θ_E			
		0.125	0.375	0.625	0.875
	0.125	0.0625	0.0625	0.0625	0.0625
θ_C	0.375	0.0625	0.0625	0.0625	0.0625
	0.625	0.0625	0.0625	0.0625	0.0625
	0.875	0.0625	0.0625	0.0625	0.0625

Informed prior (this is unnormalized)

		θ_E			
		0.125	0.375	0.625	0.875
	0.125	18	18	32	32
θ_C	0.375	18	18	32	32
	0.625	18	18	32	32
	0.875	18	18	32	32

(Rationale for the informed prior is on the next slide.)

Solution continued

Since 19/19 ECMO babies survived we believe θ_E is probably near 1.0
That 0/3 CVT babies survived is not enough data to move from a uniform distribution. (Or we might distribute a little more probability to larger θ_C.)
So for θ_E we split 64% of probability in the two higher values and 36% for the lower two. Our prior is the same for each value of θ_C.

Likelihood

Entries in the likelihood table are $\theta_E^{28}(1-\theta_E)\theta_C^6(1-\theta_C)^4$. We don't bother including the binomial coefficients since they are the same for every entry.

		θ_E			
		0.125	0.375	0.625	0.875
	0.125	1.012e-31	1.653e-18	1.615e-12	6.647-09
θ_C	0.375	1.920e-29	3.137e-16	3.065e-10	1.261-06
	0.625	5.332e-29	8.713e-16	8.513e-10	3.504e-06
	0.875	4.95e-30	8.099e-17	7.913e-11	3.257e-07

(Posteriors are on the next slides).

Solution continued

Flat posterior

The posterior table is found by multiplying the prior and likelihood tables and normalizing so that the sum of the entries is 1. We call the posterior derived from the flat prior the flat posterior. (Of course the flat posterior is not itself flat.)

		θ_E			
		0.125	0.375	0.625	0.875
	0.125	.984e-26	3.242e-13	3.167e-07	0.001
θ_c	0.375	.765e-24	6.152e-11	6.011e-05	$\boxed{0.247}$
	0.625	1.046e-23	1.709e-10	1.670e-04	$\boxed{0.687}$
	0.875	9.721e-25	1.588e-11	1.552e-05	0.0639

The boxed entries represent most of the probability where $\theta_E > \theta_C$.

All our computations were done in R. For the flat posterior:
Probability ECMO is better than CVT is

$$P(\theta_E > \theta_C \,|\, \text{Harvard data}) = 0.936$$
$$P(\theta_E - \theta_C \geq 0.6 \,|\, \text{Harvard data}) = 0.001$$

Solution continued

Informed posterior

		θ_E			
		0.125	0.375	0.625	0.875
	0.125	1.116e-26	1.823e-13	3.167e-07	0.001
θ_C	0.375	2.117e-24	3.460e-11	6.010e-05	0.2473
	0.625	5.882e-24	9.612e-11	1.669e-04	0.6871
	0.875	5.468e-25	8.935e-12	1.552e-05	0.0638

For the informed posterior:

$$P(\theta_E > \theta_C \,|\, \text{Harvard data}) = 0.936$$
$$P(\theta_E - \theta_C \geq 0.6 \,|\, \text{Harvard data}) = 0.001$$

Note: Since both flat and informed prior gave the same answers we gain confidence that these calculations are robust. That is, they are not too sensitive to our exact choice prior.

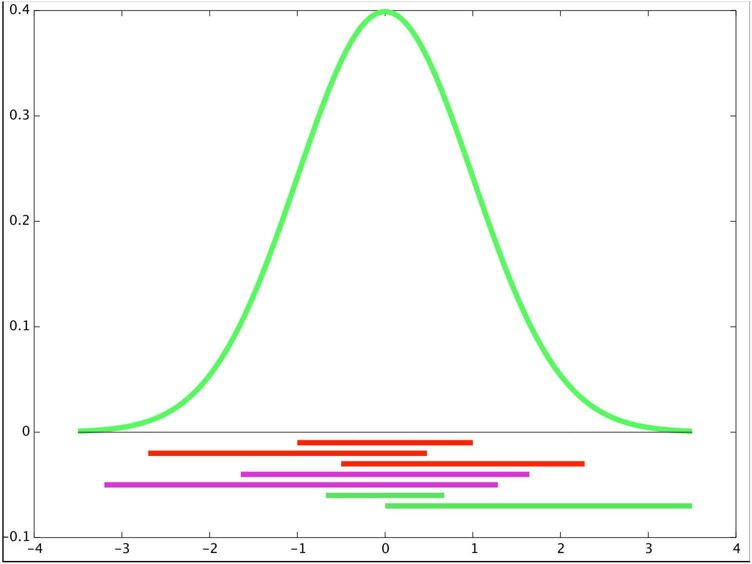
Probability intervals

- **Example.** If $P(a \leq \theta \leq b) = 0.7$ then $[a, b]$ is a 0.7 probability interval for θ. We also call it a 70% probability interval.
- **Example.** Between the 0.05 and 0.55 quantiles is a 0.5 probability interval. Another 50% probability interval goes from the 0.25 to the 0.75 quantiles.
- **Symmetric probability intevals.** A symmetric 90% probability interval goes from the 0.05 to the 0.95 quantile.
- **Q-notation.** Writing q_p for the p quantile we have 0.5 probability intervals $[q_{0.25}, q_{0.75}]$ and $[q_{0.05}, q_{0.55}]$.
- **Uses.** To summarize a distribution; To help build a subjective prior.

Probability intervals in Bayesian updating

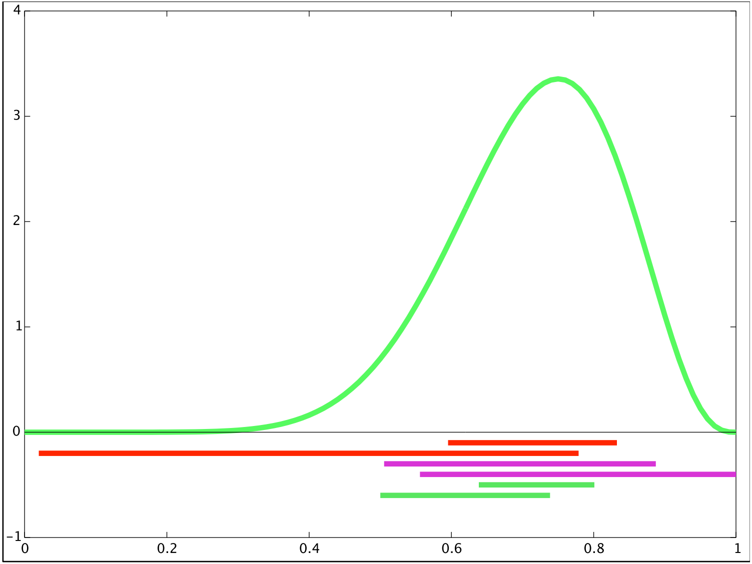
- We have p-probability intervals for the prior $f(\theta)$.
- We have p-probability intervals for the posterior $f(\theta|x)$.
- The latter tends to be smaller than the former. Thanks data!
- Probability intervals are good, concise statements about our current belief/understanding of the parameter of interest.
- We can use them to help choose a good prior.

Probability intervals for normal distributions



Red = 0.68, magenta = 0.9, green = 0.5

Probability intervals for beta distributions



Red $= 0.68$, magenta $= 0.9$, green $= 0.5$

Concept question

To convert an 80% probability interval to a 90% interval should you shrink it or stretch it?

1. Shrink **2.** Stretch.

answer: 2. Stretch. A bigger probability requires a bigger interval.

Reading questions

The following slides contain bar graphs of last year's responses to the reading questions. Each bar represents one student's estimate of their own 50% probability interval (from the 0.25 quantile to the 0.75 quantile).

Here is what we found for answers to the questions:

1. Airline deaths in 100 years: We extracted this data from a government census table at https://www2.census.gov/library/publications/2011/compendia/statab/131ed/2012-statab.pdf page 676 There were 13116 airline fatalities in the 18 years from 1990 to 2008. In the 80 years before that there were fewer people flying, but it was probably more dangerous. Let's assume they balance out and estimate the total number of fatalities in 100 years as $5 \times 13116 \approx 66000$.

2. Number of girls born in the world each year: I had trouble finding a reliable source. Wiki.answers.com gave the number of 130 million births in 2005. If we take what seems to be the accepted ratio of 1.07 boys born for every girl then $130/2.07 = 62.8$ million baby girls.

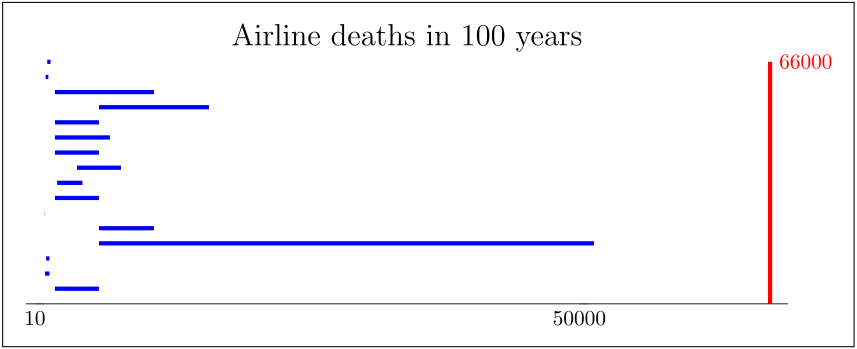
Reading questions continued

3. Percentage of Black or African-Americans in the U.S as of 2015.: 13.3% (https://www.census.gov/quickfacts/)

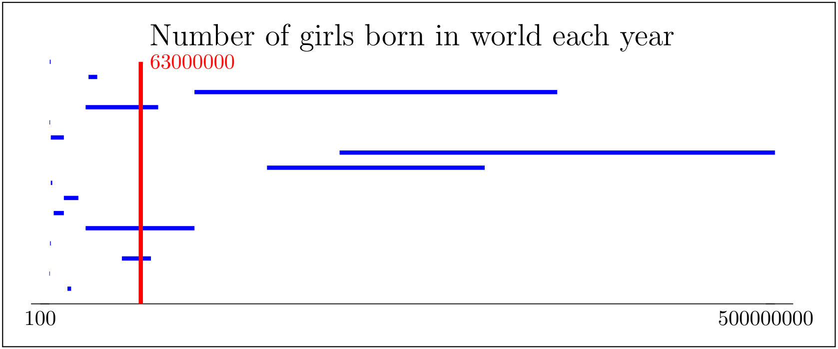
4. Number of French speakers world-wide: 72-79 million native speakers, 265 million primary + secondary speaker (http://www2.ignatius.edu/faculty/turner/languages.htm)

5. Number of abortions in the U.S. each year: 1.2 million (http://www.guttmacher.org/in-the-know/characteristics.html)

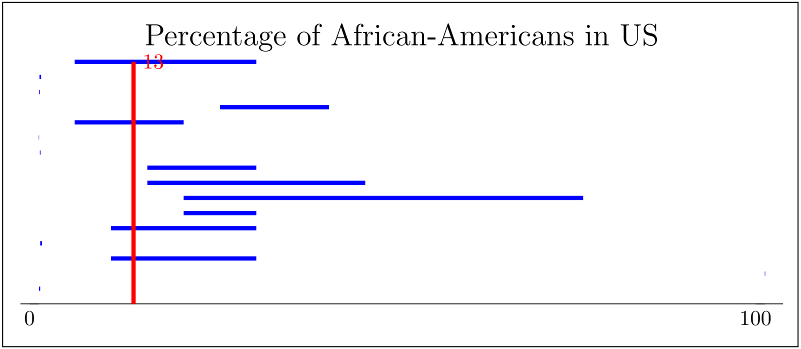
Subjective probability 1 (50% probability interval)



Subjective probability 2 (50% probability interval)

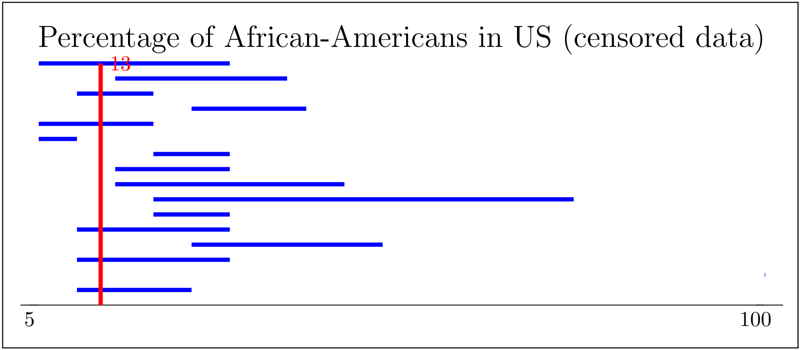


Subjective probability 3 (50% probability interval)

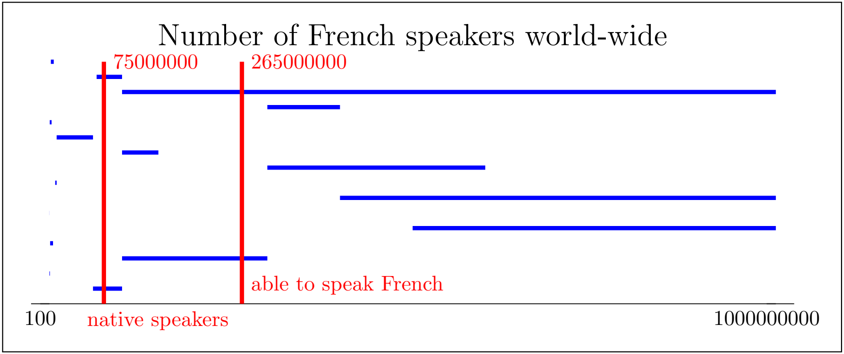


Subjective probability 3 censored (50% probability interval)

Censored by changing numbers less than 1 to percentages and ignoring numbers bigger that 100.



Subjective probability 4 (50% probability interval)



Subjective probability 5 (50% probability interval)