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

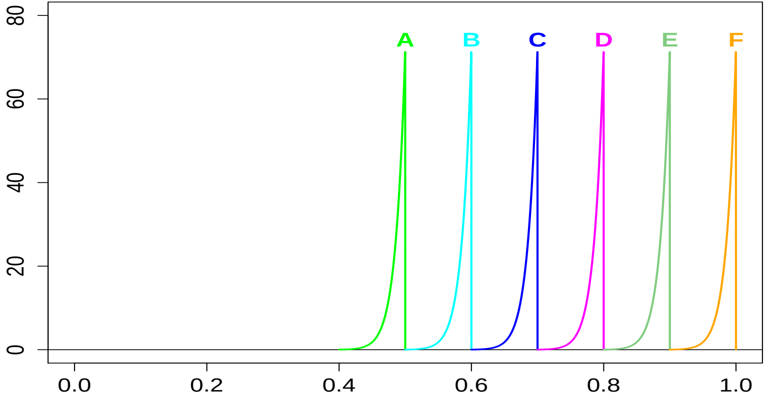
Concept question: strong priors

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?



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	$p(\theta_N\|\text{data})$
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
$p(\theta_S\|\text{data})$	0.00000	0.91230	0.08670	0.00100	0.00000	0.00000	1.00000

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

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

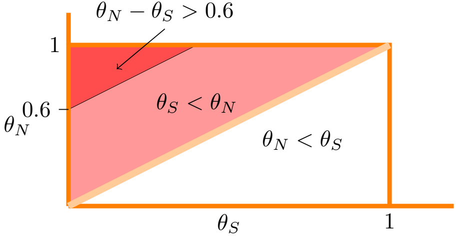
Some protection: $P(\theta_N > \theta_S \,|\, \text{data}) = \text{sum pink and red} = .99995$

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.

θ_E = probability that an ECMO baby survives
θ_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.)

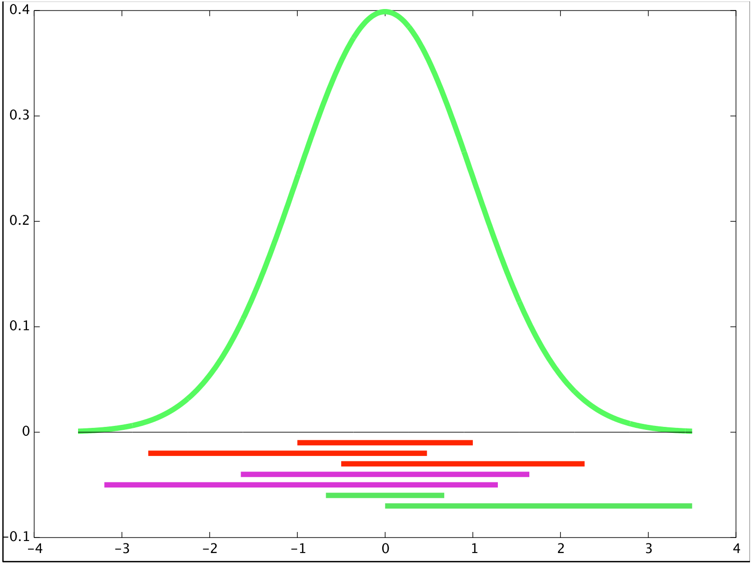
Probability intervals

- **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

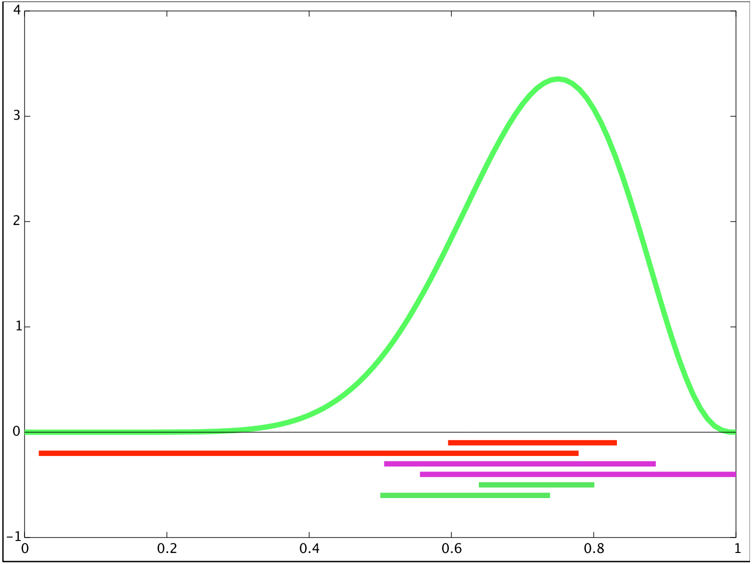
- We have p-probability intervals for the prior $f(\theta)$.
- We have p-probability intervals for the posterior $f(\theta|x)$.
- The latter tends to be smaller than the former. Thanks data!
- Probability intervals are good, concise statements about our current belief/understanding of the parameter of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



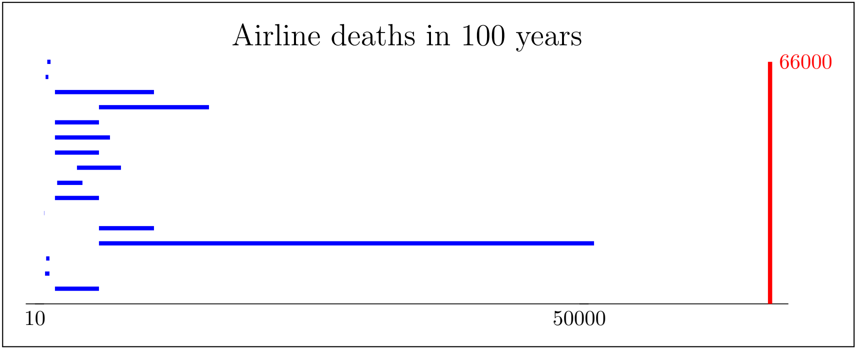
Red $= 0.68$, magenta $= 0.9$, green $= 0.5$

Concept question

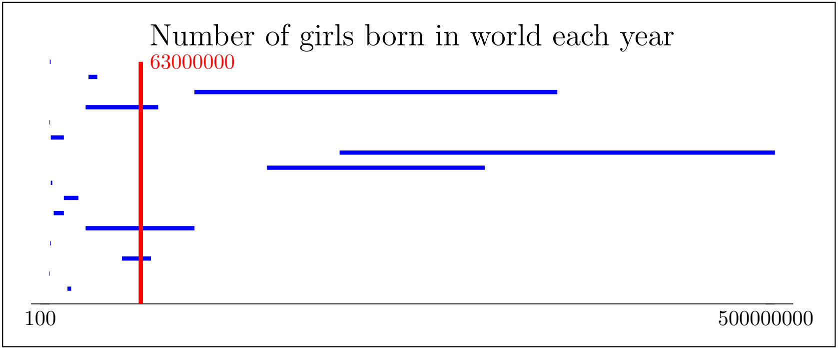
To convert an 80% probability interval to a 90% interval should you shrink it or stretch it?

1. Shrink **2.** Stretch.

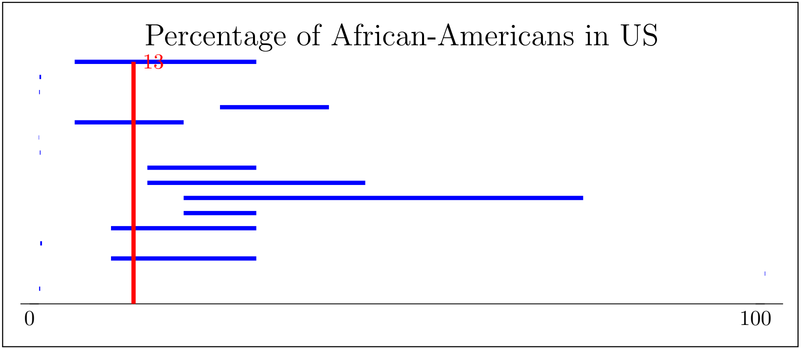
Subjective probability 1 (50% probability interval)



Subjective probability 2 (50% probability interval)

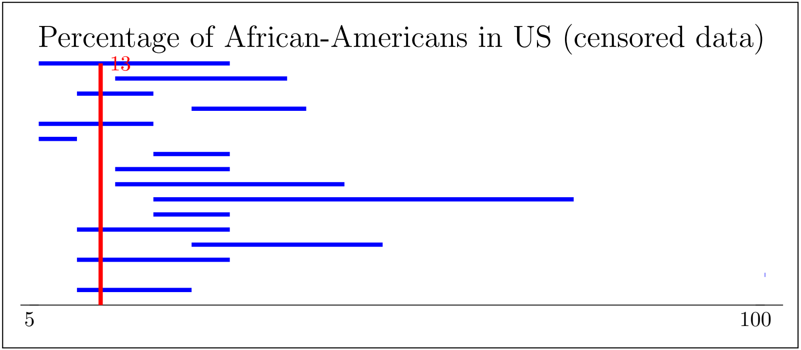


Subjective probability 3 (50% probability interval)

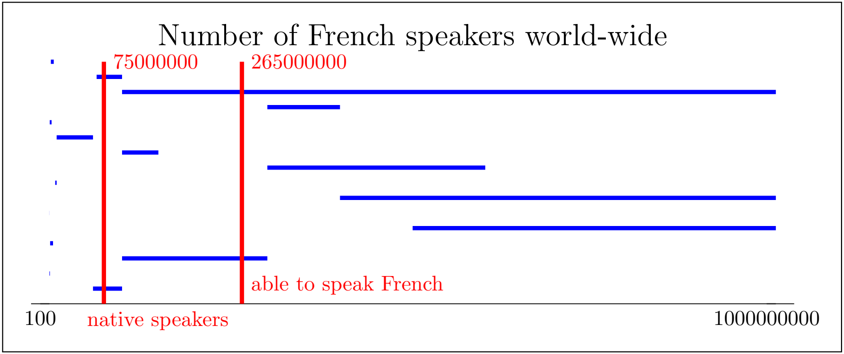


Subjective probability 3 censored (50% probability interval)

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)