Example

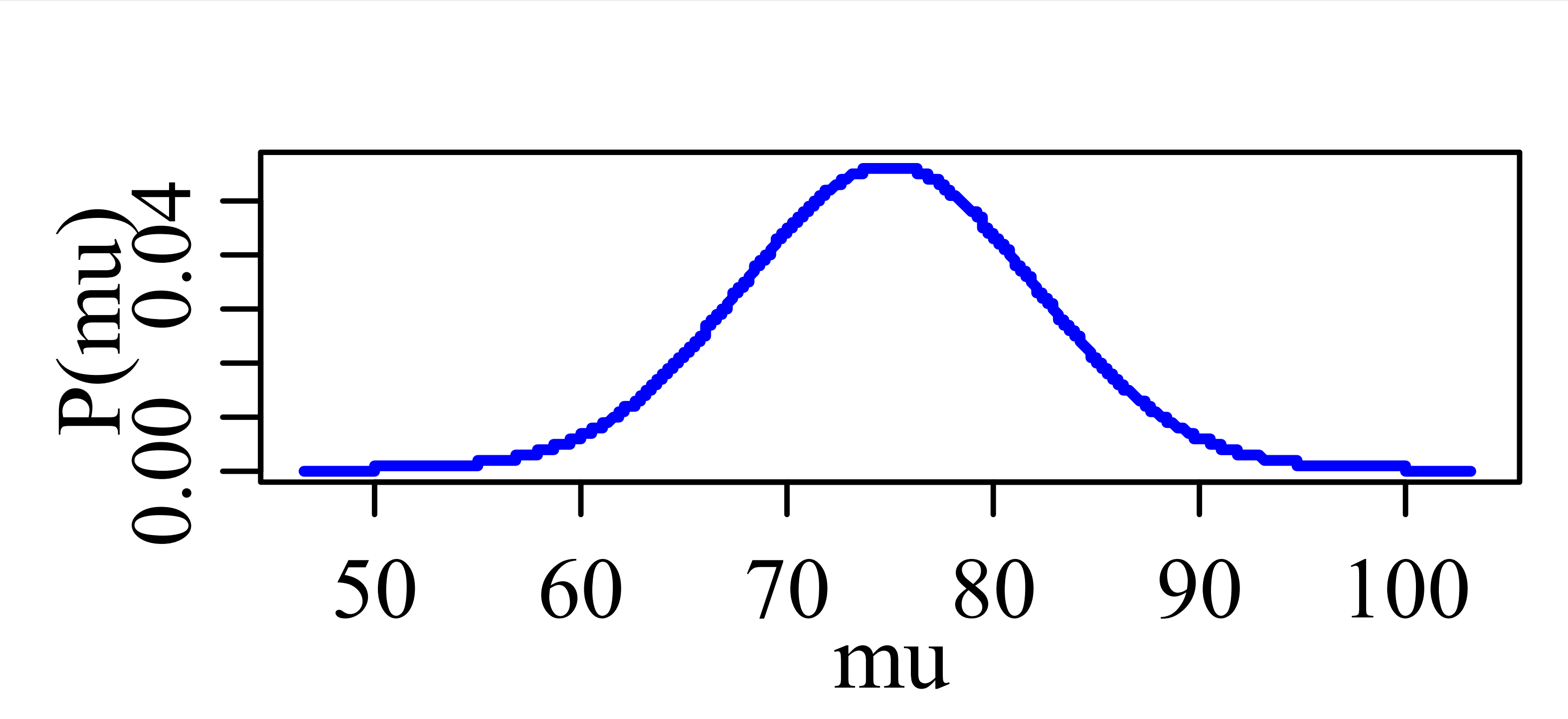
- Suppose we have a test scored from 0-100
- Interested in the distribution of test scores for students
- Have test scores from 10 students
- Let x_i denote the score on the test for student i
- Postulate a distribution for the students test scores
- $x_i \mid \mu, \sigma^2 \sim N(\mu, \sigma^2)$
  - μ: the mean of the population of students' scores
  - σ^2: the variability in the distribution of the students' scores
- $\mathbf{x} = (91, 85, 72, 87, 71, 77, 88, 94, 84, 92)$

Prior for Mean

- $\mu \sim N(\mu_\mu, \sigma_\mu^2)$

Does this make sense?
If not may need to alter prior.

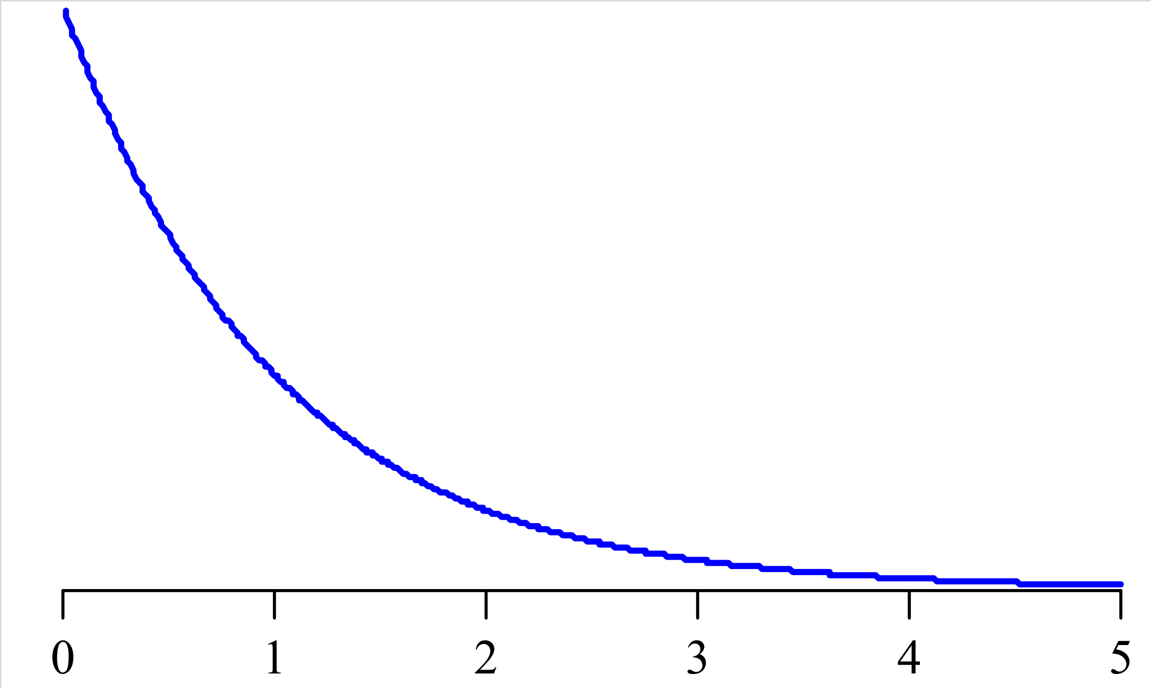
- $\mu_\mu = 75, \sigma_\mu^2 = 50 \rightarrow \mu \sim N(\mu_\mu, \sigma_\mu^2) = N(75, 50)$



Prior for Standard Deviation

- $\sigma \sim \text{Exp}(\lambda)$

- $\lambda = 1 \rightarrow \sigma \sim \text{Exp}(\lambda) = \text{Exp}(1)$



Prior Predictive Distribution

Prior Predictive Distribution

The joint distribution for all unknowns $\boldsymbol{\theta}$ and all data $\mathbf{x}$

$$p(\mathbf{x}, \boldsymbol{\theta}) = p(\mathbf{x} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})$$

Marginal distribution for the data, according to the model

$$p(\mathbf{x}^{prior} \cancel{\mid \mathbf{x}}) = \int p(\mathbf{x}, \boldsymbol{\theta})\, d\boldsymbol{\theta} = \int p(\mathbf{x} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})\, d\boldsymbol{\theta}$$

Distribution of data given model parameters

Prior distribution for the parameters

Prior Predictive Distribution

Marginal distribution for the data, according to the model

$$p(\mathbf{x}^{prior} \mid \mathbf{x}) = \int p(\mathbf{x}, \boldsymbol{\theta})\, d\boldsymbol{\theta} = \int p(\mathbf{x} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})\, d\boldsymbol{\theta}$$

Distribution of data given model parameters

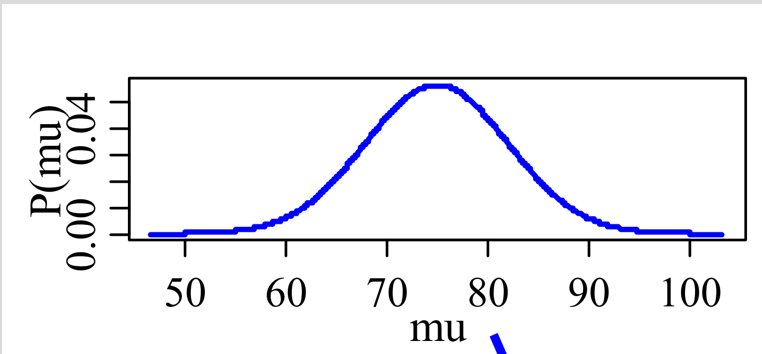
Prior distribution for the parameters

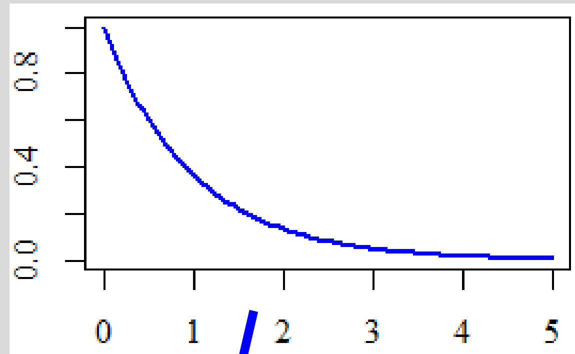
Simulate values by

1. Simulate values for the parameters from the prior $p(\theta)$
2. Using those values for the parameters, simulate from the conditional distribution of data $p(x \mid \theta)$
3. Repeat

Prior Predictive Distribution

1. Simulate values for the parameters from the prior $p(\boldsymbol{\theta})$





μ_{sim} σ_{sim}

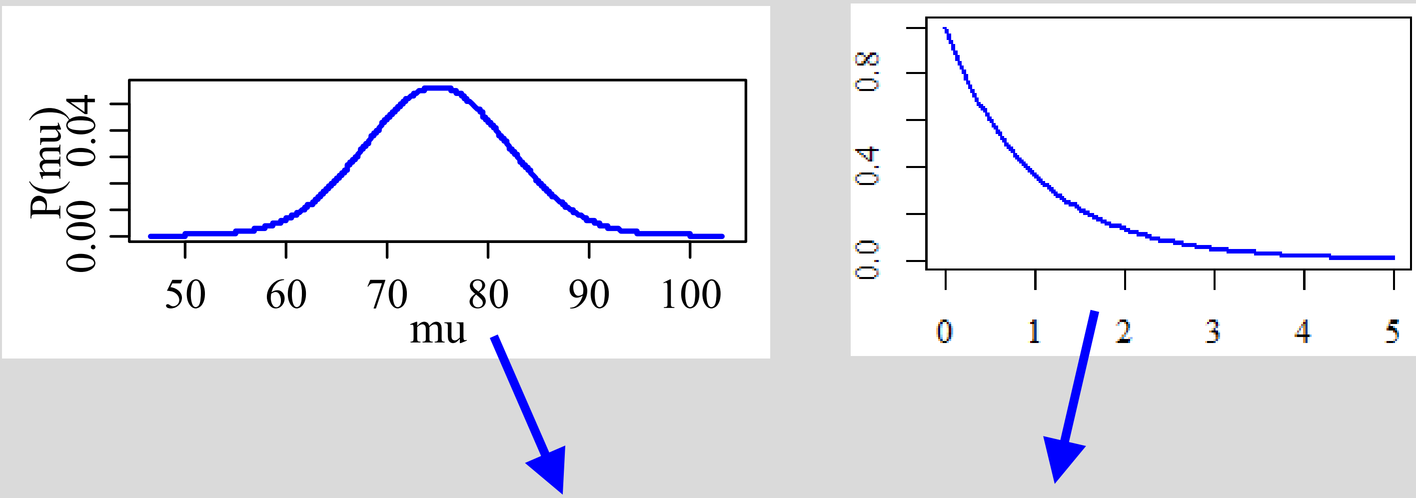
2. Using those values for the parameters, simulate from the conditional distribution of data $p(\mathrm{x} \mid \theta)$

$$x \sim N(\mu_{\text{sim}}, \sigma^2_{\text{sim}})$$

3. Repeat

Code

1. Simulate values for the parameters from the prior $p(\boldsymbol{\theta})$



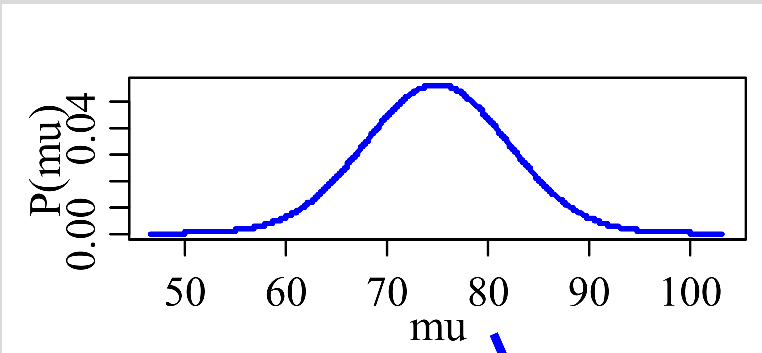
μ_{sim} σ_{sim}

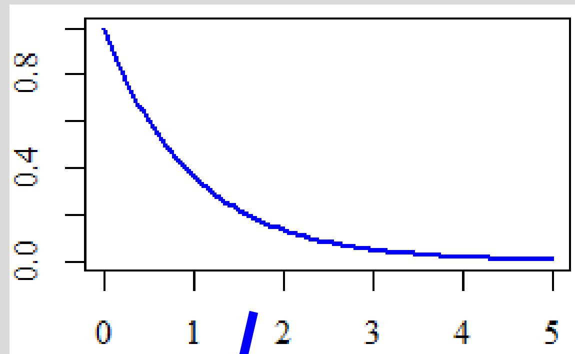
```
n_samples = 10000

mu.mu = 75
sigma.squared.mu = 50
sigma.mu = sqrt(sigma.squared.mu)
sample_mu <- rnorm(n_samples, mu.mu, sigma.mu)
```

Code

1. Simulate values for the parameters from the prior $p(\boldsymbol{\theta})$





μ_{sim} σ^2_{sim}

```
n_samples = 10000

lambda = 1
sample_sigma <- rexp(n_samples, rate=lambda)
```

Code

2. Using those values for the parameters, simulate from the conditional distribution of data $p(x \mid \theta)$

$$x \sim N(\mu_{sim}, \sigma^2_{sim})$$

```
prior_x <- rnorm(n_samples, sample_mu,
sample_sigma)

plot(
  density(prior_x),
  xlab="x", ylab="p(x)",
  font.lab=6,
  font.axis=6,
  lwd=2,
  col="dark green"
)
```

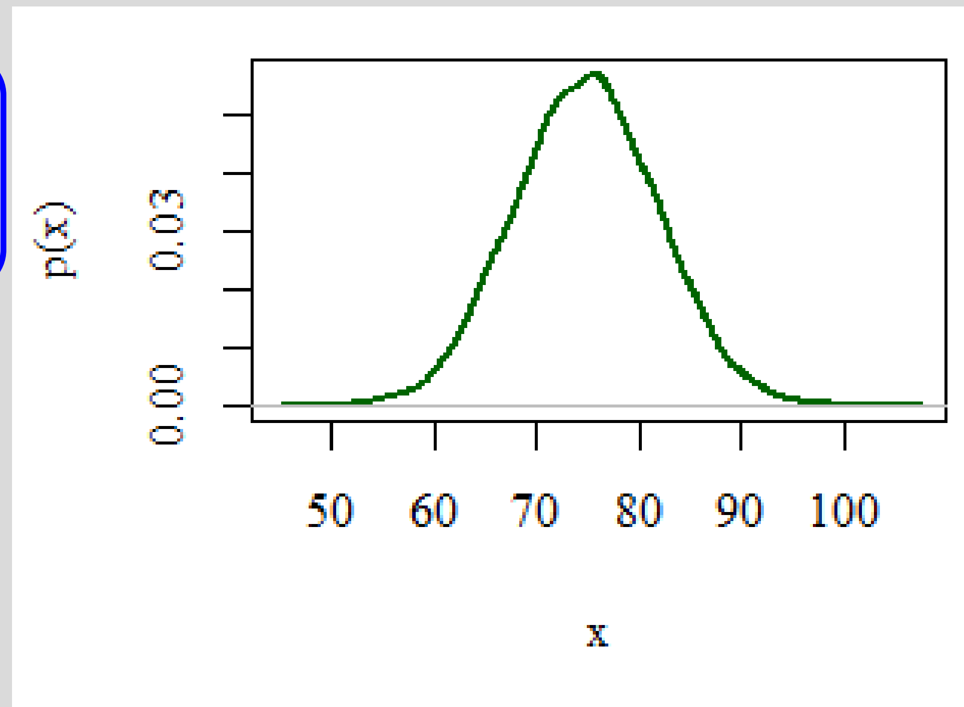
Prior Predictive Distribution

$$p(\mathbf{x}^{prior} \,\cancel{|\, \mathbf{x}}) = \int p(\mathbf{x}, \boldsymbol{\theta})\, d\boldsymbol{\theta} = \int p(\mathbf{x} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})\, d\boldsymbol{\theta}$$

Distribution of data given model parameters

Prior distribution for the parameters

Does this make sense?
If not may need to alter prior.



Exchangeability

Simplifying Priors Via Exchangeability

- Consider a joint prior distribution $p(\theta_1, \theta_2, \ldots, \theta_n)$

  "In considering the prior knowledge of the θ_i it may often be reasonable to assume their distribution *exchangeable*. That is, that it would be unaltered by any permutation of the suffixes: so that, in particular, the prior opinion of θ_7 is the same as that of θ_4, or any other θ_i; and similarly for pairs, triplets and so on."

  -- Lindley & Smith (1972, p. 2)

- Believe the same things about each parameter
- Supports the use of a ***common*** univariate prior
- $p(\theta_1, \theta_2, \ldots, \theta_n) = \Pi_i \, p(\theta_i \mid \,)$
  - Possibly conditional on other terms

Simplifying Priors Via Exchangeability

- Prior for regression coefficients for *J* predictors

$$p(\beta_1, \beta_2, \ldots, \beta_J) = \prod_{j=1}^{J} p(\beta_j \mid \mu_\beta, \sigma_\beta^2)$$

$$p(\beta_j \mid \mu_\beta, \sigma_\beta^2) = N(\mu_\beta, \sigma_\beta^2)$$

- Popular mechanism for specifying prior distributions, particularly in conjunction with hierarchical model specifications

Bayesian Analyses of the [Traditional Mo]del in JAGS and Stan

Does the difference in prior matter?

	Stan: Bayesian Analysis of Traditional Model			JAGS: Bayesian Analysis of Traditional Model		
	Post. Mean	Post. SD	95% Cred. Interval	Post. Mean	Post. SD	95% Cred. Interval
β_0	-2.52	1.97	(-6.24, 1.46)	-2.53	1.94	(-6.32, 1.29)
β_1	0.66	0.17	(0.31, 0.97)	0.66	0.17	(0.33, 0.99)
β_2	0.38	0.10	(0.18, 0.59)	0.38	0.10	(0.18, 0.58)
σ_ε	1.92	0.20	(1.55, 2.32)	1.91	0.20	(1.57, 2.35)
R^2	0.59	0.08	(0.43, 0.73)	0.59	0.06	(0.45, 0.68)

$\beta_0 \sim N(0, 900)$

$\beta_j \sim N(0, 900)$

$\sigma_\varepsilon \sim \text{Exp}(1)$

$\beta_0 \sim N(0,1000)$ $\quad \tau_\varepsilon \sim \text{Gamma}(1,1)$

$\beta_j \sim N(0,1000)$ $\quad \sigma_\varepsilon = (\tau_\varepsilon)^{-1/2}$

$\beta_0 \sim N(0,1000) \quad \tau_\varepsilon \sim \text{Gamma}(1,1)$

$\beta_j \sim N(0,1000) \quad \sigma_\varepsilon = (\tau_\varepsilon)^{-1/2}$

$\beta_0 \sim U(0,15) \quad \tau_\varepsilon \sim \text{Gamma}(1,1)$

$\beta_j \sim N(0,1000) \quad \sigma_\varepsilon = (\tau_\varepsilon)^{-1/2}$

	Frequentist Analysis of Traditional Model		JAGS: Bayesian Analysis of Traditional Model		BUGS: Bayesian Analysis of Modified Model	
	Est.	95% Conf. Int.	Post. Mean	95% Cred. Interval	Post. Mean	95% Cred. Interval
β_0	-2.54	(-6.41, 1.34)	-2.53	(-6.32, 1.29)	1.07	(0.03, 3.47)
β_1	0.66	(0.33, 0.99)	0.66	(0.33, 0.99)	0.40	(0.15, 0.63)
β_2	0.38	(0.18, 0.59)	0.38	(0.18, 0.58)	0.39	(0.18, 0.59)
σ_ε	1.95	(1.60, 2.37)	1.91	(1.57, 2.35)	1.98	(1.63, 2.43)
R^2	0.60		0.59	(0.45, 0.68)	0.48	(0.35, 0.56)

Additional Examples

Example: Sensitivity of a Test Based on Past Research

Suppose we are interested in the sensitivity of a test, which is the probability that the test will indicate something is there when it actually is (i.e., a true positive)

$$p(\text{test is } + \mid \text{disease is present, status is true, etc.})$$

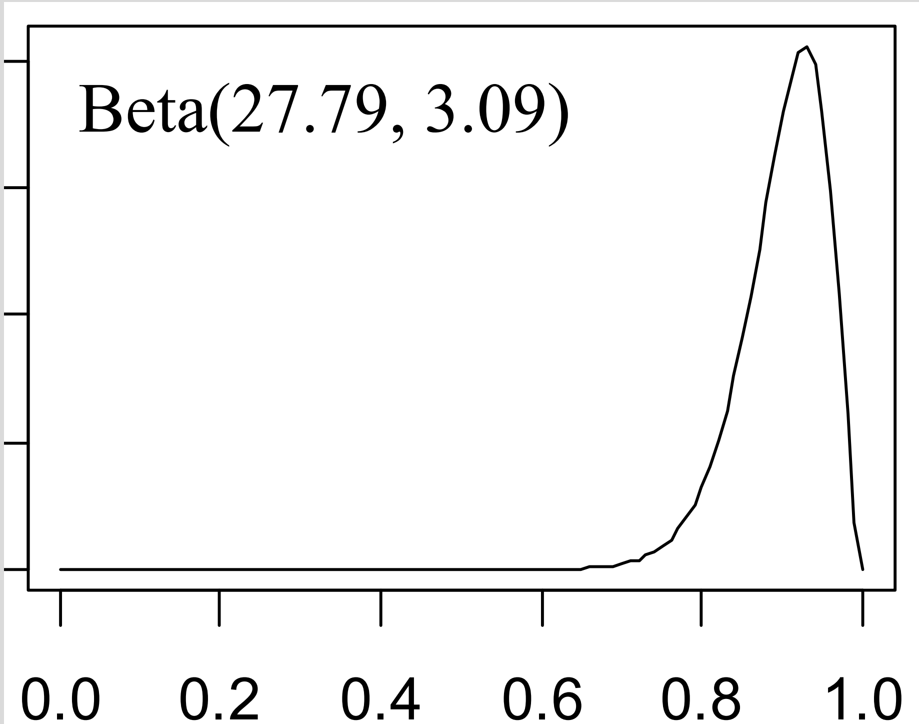
Let θ denote this unknown parameter, which is a probability we could model with a beta prior distribution

Past research studies in the literature have estimated θ and this has averaged out to be .90. And all of the studies have estimated it to be above .80.

Example:
Sensitivity of a Test Based on Past Research

Model prior beliefs that say we expect a value of .90 and we are 95% confident that the value is greater than .80.

Find the beta distribution that has a mean of .90 and 95% of the distribution is above .80:



Example: Prevalence Based on Expert Opinion

Suppose we are interested in how many children have undiagnosed learning disabilities. Let θ denote this unknown parameter, which is a probability we could model with a beta prior distribution.

Analyst: *What do you think is the proportion of children with undiagnosed learning disabilities?*

SME: *My best guess would be around 15%.*

Analyst: *Would it surprise you if it was more than 30%?*

SME: *Eh, not really. I'd buy that.*

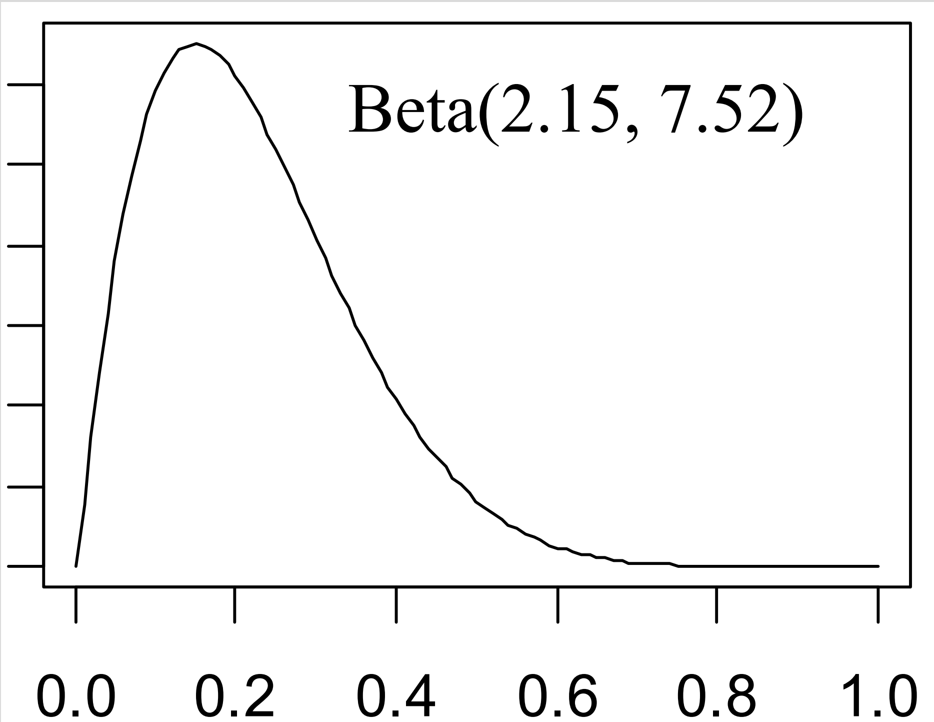
Analyst: *How about 40%?*

SME: *Wow, I'd be really surprised by that. I guess it's possible, but I'd say it's really unlikely.*

Example: Prevalence Based on Expert Opinion

Model prior beliefs that say that our best guess is .15 and we are 90% confident that the value is less than .40.

Find the beta distribution that has a mode of .15 and 90% of the distribution is below .40:



Example: Univ. of Iowa Students with ACT Scores ≥ 19 in the Early 1970s

An administrator reviewing admissions policies at the University of Iowa in the early 1970s needs to estimate the number of student enrolled who received an ACT score of 19 or better.

No summary data is available, and no one wants to look through all that many applications. Will only examine a sample. What should the prior be for θ, the proportion of students at Iowa with scores of 19 or better? Let's use a Beta prior and suppose we have a few pieces of information…

Example: Univ. of Iowa Students with ACT Scores ≥ 19 in the Early 1970s

Pieces of information:

- A 1966 report indicated that the national percentage of college students with ACT scores ≥ 19 was 67% (n = 238,145)
- The report indicated that 77% (n = 77,383) of those enrolled at Midwestern colleges had ACT scores ≥ 19
- This year a 10% national sample indicated the percentage was .66 (n = 55,702)
- In similar kinds of studies, Iowa typically falls near the average for midwestern colleges

This year's data suggest that the story from 1966 still holds up at the national level, with perhaps a slight downtick, so probably at the regional levels too

Example: Univ. of Iowa Students with ACT Scores ≥ 19 in the Early 1970s

Analyst: *What's the most probable value for θ?*

SME: *.75*

Analyst: *Now consider your prior information as if it were the results of an experiment, or previous sample. How many sample observations do you feel that prior information should be worth? How much do you want to weight it?*

SME: *25.*

Example: Univ. of Iowa Students with ACT Scores ≥ 19 in the Early 1970s

Recall our features of the beta distribution:

$$\text{Mode}[\text{Beta}(\alpha,\beta)] = \frac{\alpha - 1}{\alpha + \beta - 2} \qquad \text{Weight}[\text{Beta}(\alpha,\beta)] = \alpha + \beta - 2$$

A little algebra shows that

$$\text{Mode}[\text{Beta}(\alpha,\beta)] = \frac{\alpha - 1}{\text{Weight}}$$

$$\text{Weight} \times \text{Mode} = \alpha - 1$$

$$\alpha = \text{Weight} \times \text{Mode} + 1$$

$$\beta = \text{Weight} - \alpha + 2$$

Example: Univ. of Iowa Students with ACT Scores ≥ 19 in the Early 1970s

Recall our features of the beta distribution:

$$\text{Mode}[\text{Beta}(\alpha,\beta)] = \frac{\alpha - 1}{\alpha + \beta - 2} \qquad \text{Weight}[\text{Beta}(\alpha,\beta)] = \alpha + \beta - 2$$

A little algebra shows that

$$\text{Mode}[\text{Beta}(\alpha,\beta)] = \frac{\alpha - 1}{\text{Weight}}$$

$$\text{Weight} \times \text{Mode} = \alpha - 1$$

$$\alpha = \text{Weight} \times \text{Mode} + 1$$

$$\beta = \text{Weight} - \alpha + 2$$

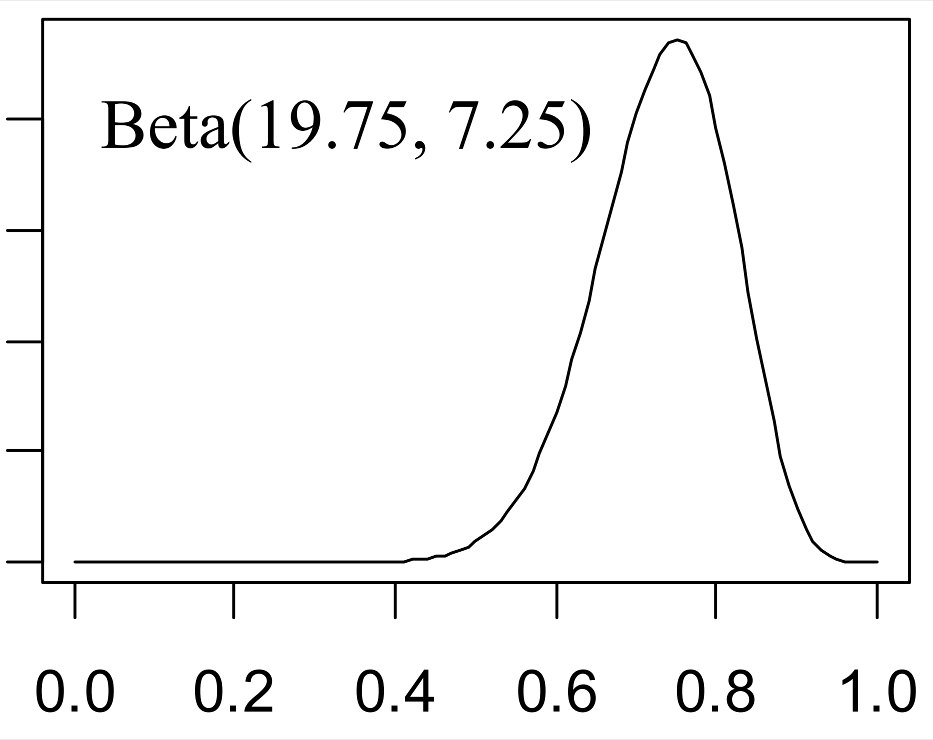
For our example

$$\alpha = 25 \times .75 + 1 = 19.75$$

$$\beta = 25 - 19.75 + 2 = 7.25$$

Example: Univ. of Iowa Students with ACT Scores ≥ 19 in the Early 1970s

The beta distribution that has a mode of .75 and is weighted akin to 25 observations is the Beta(19.75, 7.25) distribution



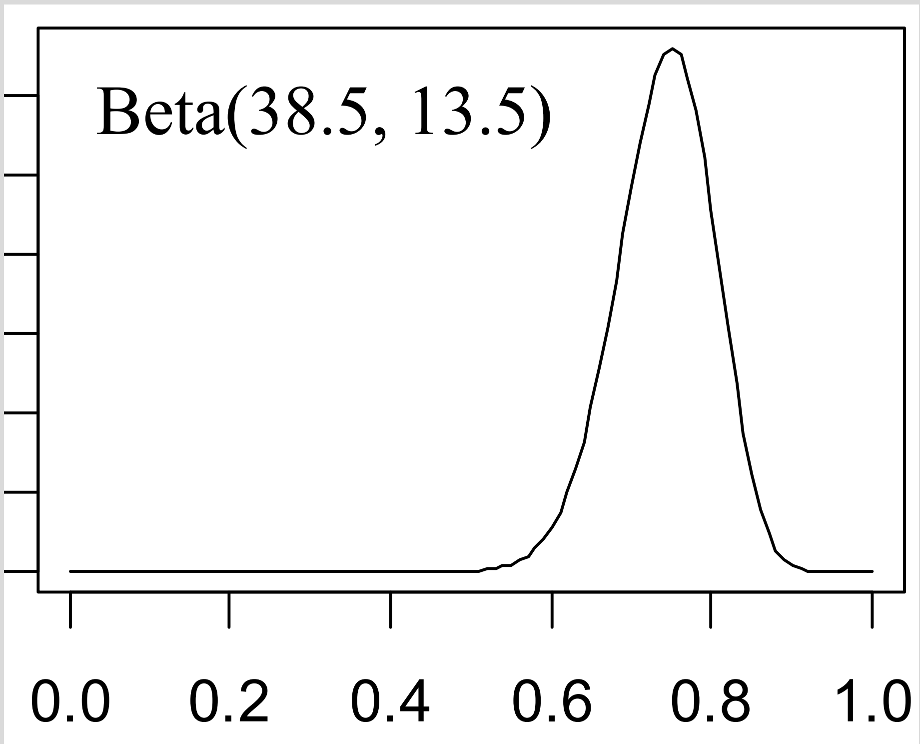
Example: Univ. of Iowa Students with ACT Scores ≥ 19 in the Early 1970s

Analyst: *The values you selected implies that the 50% highest density interval is from .71 to .79. That means you would be indifferent to a bet that says the value is inside this interval vs. outside this interval. If you're not satisfied with this, select another weight.*

SME: (That interval is way too big!) *50*

Example: Univ. of Iowa Students with ACT Scores ≥ 19 in the Early 1970s

The beta distribution that has a mode of .75 and is weighted akin to 50 observations is the Beta(38.5, 13.5) distribution



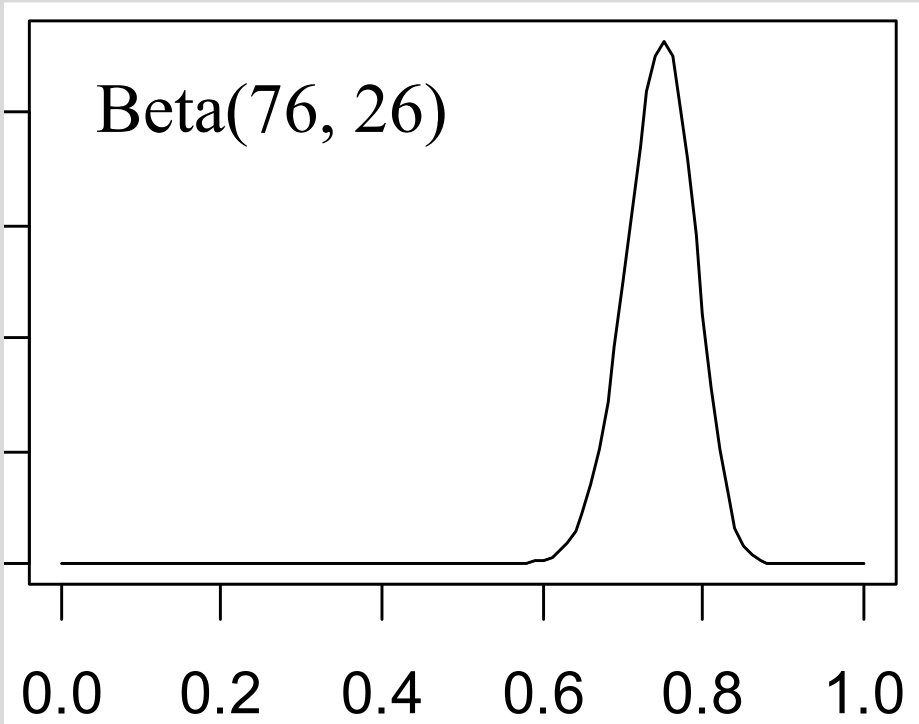
Example: Univ. of Iowa Students with ACT Scores ≥ 19 in the Early 1970s

Analyst: *The values you selected implies that the 50% highest density interval is from .71 to .79. That means you would be indifferent to a bet that says the value is inside this interval vs. outside this interval. If you're not satisfied with this, select another weight.*

SME: (That interval is ***still*** too big!) *100*

Example: Univ. of Iowa Students with ACT Scores ≥ 19 in the Early 1970s

The beta distribution that has a mode of .75 and is weighted akin to 100 observations is the Beta(76, 26) distribution



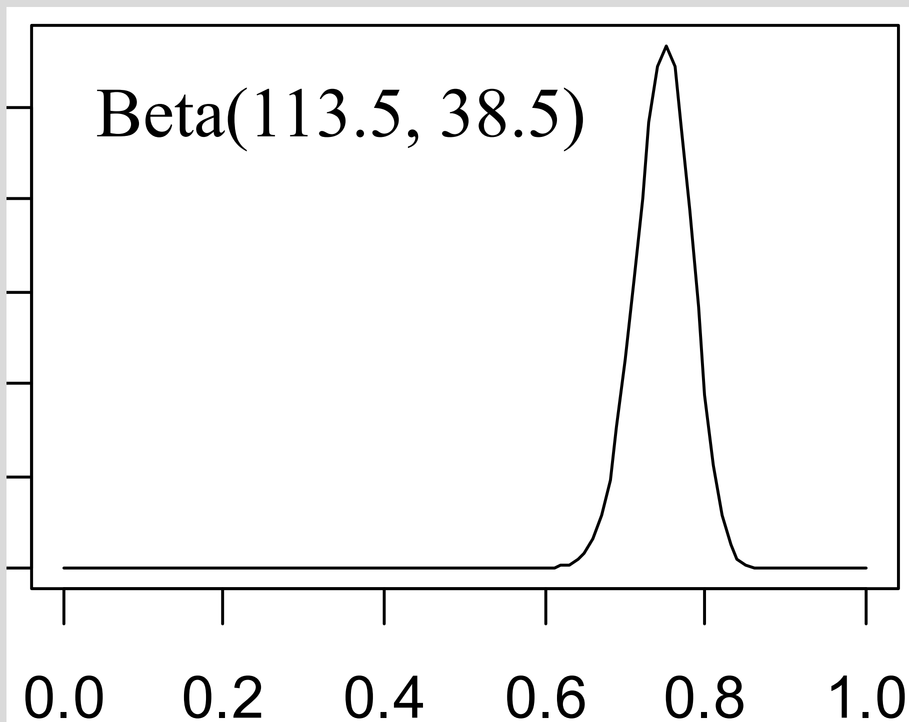
Example: Univ. of Iowa Students with ACT Scores ≥ 19 in the Early 1970s

Analyst: *The values you selected implies that the 50% highest density interval is from .72 to .78. That means you would be indifferent to a bet that says the value is inside this interval vs. outside this interval. If you're not satisfied with this, select another weight.*

SME: (That interval is ***still*** too big!) *150*

Example: Univ. of Iowa Students with ACT Scores ≥ 19 in the Early 1970s

The beta distribution that has a mode of .75 and is weighted akin to 150 observations is the Beta(113.5, 38.5) distribution



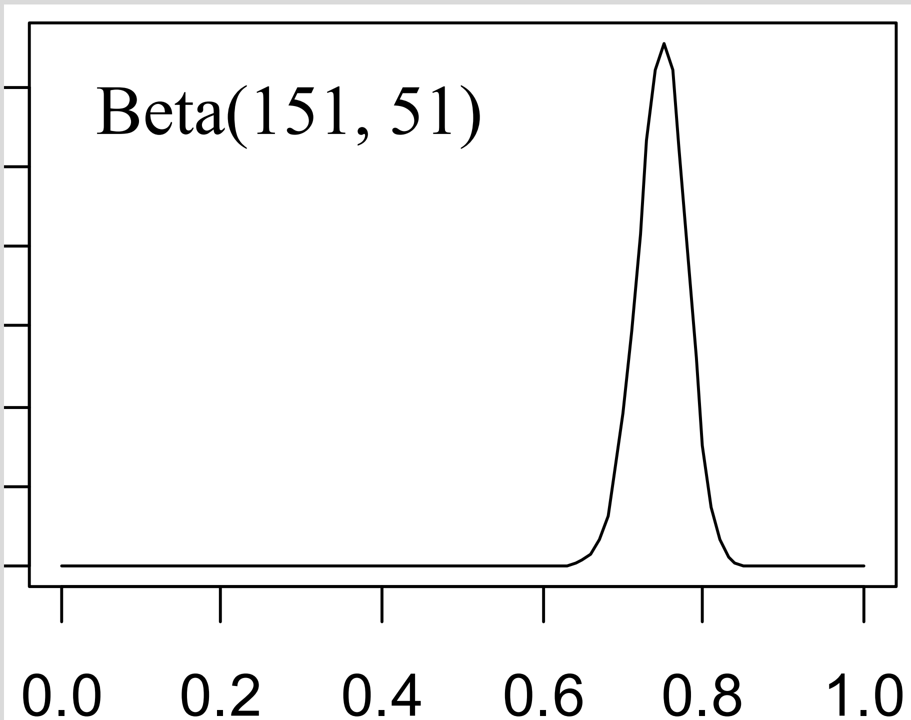
Example: Univ. of Iowa Students with ACT Scores ≥ 19 in the Early 1970s

Analyst: *The values you selected implies that the 50% highest density interval is from .73 to .77. That means you would be indifferent to a bet that says the value is inside this interval vs. outside this interval. If you're not satisfied with this, select another weight.*

SME: (That interval is decent, but what happens if keep going up?) *200*

Example: Univ. of Iowa Students with ACT Scores ≥ 19 in the Early 1970s

The beta distribution that has a mode of .75 and is weighted akin to 200 observations is the Beta(151, 51) distribution



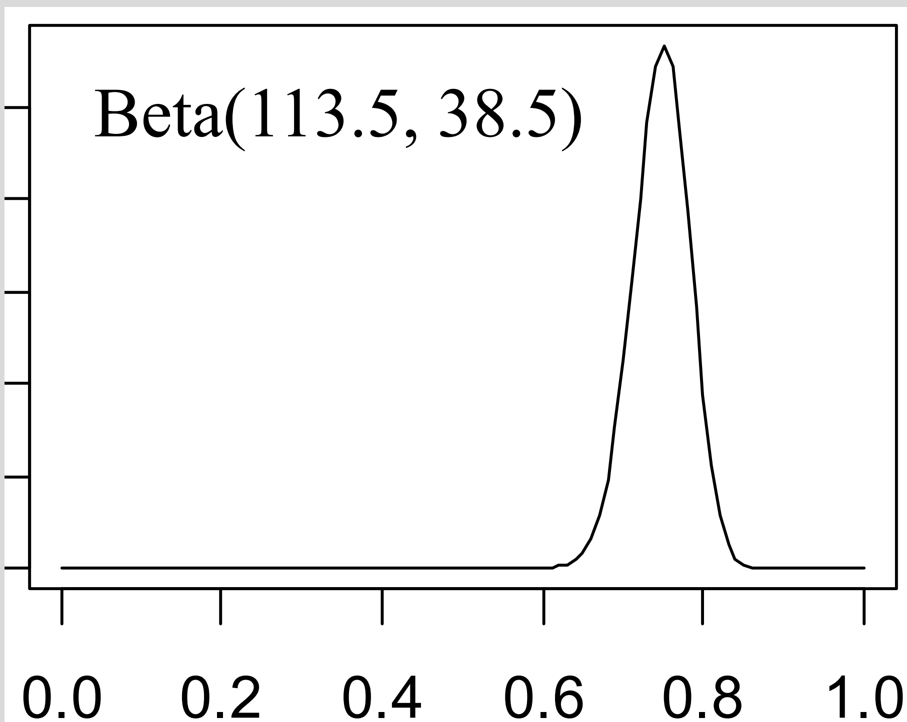
Example: Univ. of Iowa Students with ACT Scores ≥ 19 in the Early 1970s

Analyst: *The values you selected implies that the 50% highest density interval is from .73 to .77. That means you would be indifferent to a bet that says the value is inside this interval vs. outside this interval. If you're not satisfied with this, select another weight.*

SME: (That interval is the same, to two decimal places, so that didn't really move the needle. Let me go back to 150) *150*

Example: Univ. of Iowa Students with ACT Scores ≥ 19 in the Early 1970s

The beta distribution that has a mode of .75 and is weighted akin to 150 observations is the Beta(113.5, 38.5) distribution. It has a 50% highest density interval of (.73, .77)



Example: Univ. of Iowa Students with ACT Scores ≥ 19 in the Early 1970s

The beta distribution that has a mode of .75 and is weighted akin

to 150 … has a

50% h…

Example taken from Novick & Jackson (1974), in their description of Novick's *Bayesian Computer-Assisted Data Analysis* (CADA, 1971)