Introduction to Applied Bayesian Statistics
Part III: Building Bayes Theorem
Part IV: Prior Specification

Michael Anderson, PhD

Department of Biostatistics and Epidemiology
The University of Oklahoma Health Sciences Center

Sept 14, 2016

Outline

Example

In 2012, the prevalence of HIV among the general U.S. population was estimated to be 0.38% (www.cdc.gov). Rapid diagnostic tests have been developed to test for the presence of HIV infection in as little as 10 minutes. Suppose one such diagnostic test was evaluated in a case-control study and it was observed that out of 1000 subjects with HIV, 967 obtained positive test results and out of 1000 subjects with out HIV, 985 obtained negative test results.

In summary

- $P(D+) = 0.0038 \Rightarrow P(D-) = 1 - 0.0038 = 0.9962$
- $P(T + |D+) = 0.967 \Rightarrow P(T - |D+) = 1 - 0.967 = 0.033$
- $P(T - |D-) = 0.985 \Rightarrow P(T + |D-) = 1 - 0.0985 = 0.015$

Suppose a person in the U.S. is selected at random, given the test, and the result is positive. Should this person begin treatment?

- $P(D + |T+)$?

Venn Diagram of the Example

Venn Diagram of the Example

Note

- $P(D + \cap T+) = P(D+)P(T + | D+)$
- $P(D - \cap T+) = P(D-)P(T + | D-)$
- $P(T+) = P(D+)P(T + | D+) + P(D-)P(T + | D-)$

$$P(D + | T+) = \frac{P(D+\cap T+)}{P(T+)} = \frac{P(D+)P(T+|D+)}{P(D+)P(T+|D+)+P(D-)P(T+|D-)}.$$

Example (continued)

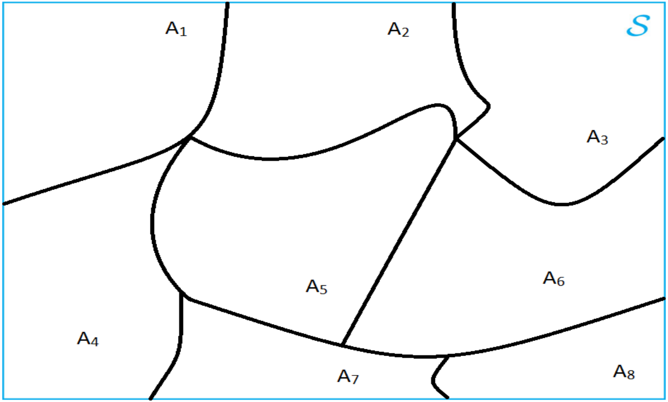
Recall

- $P(D+) = 0.0038 \Rightarrow P(D-) = 1 - 0.0038 = 0.9962$
- $P(T + |D+) = 0.967 \Rightarrow P(T - |D+) = 1 - 0.967 = 0.033$
- $P(T - |D-) = 0.985 \Rightarrow P(T + |D-) = 1 - 0.985 = 0.015$

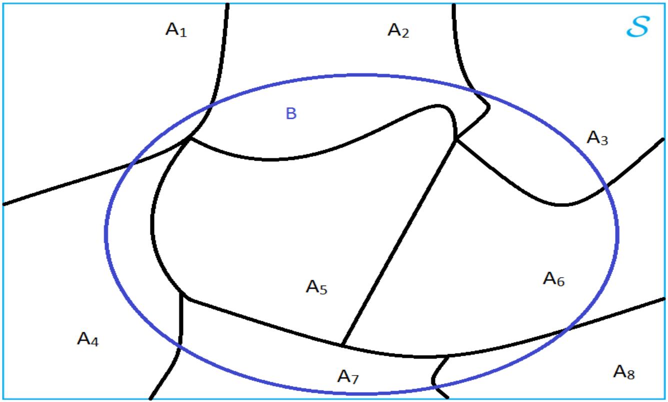
$$P(D + |T+) = \frac{P(D+)P(T+|D+)}{P(D+)P(T+|D+)+P(D-)P(T+|D-)} = \frac{(0.0038)(0.967)}{(0.0038)(0.967)+(0.9962)(0.015)} = 0.1974$$

What does this mean?

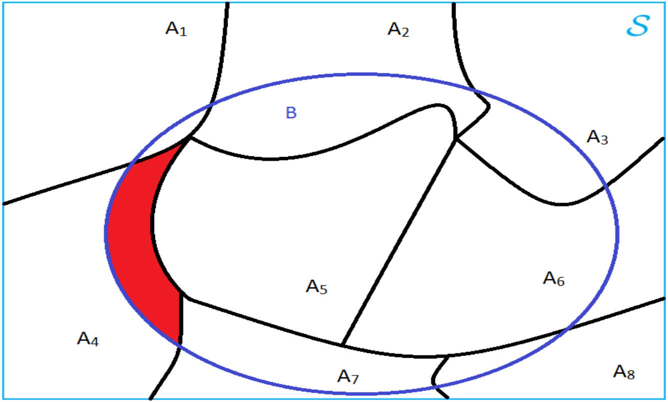
More Generally...



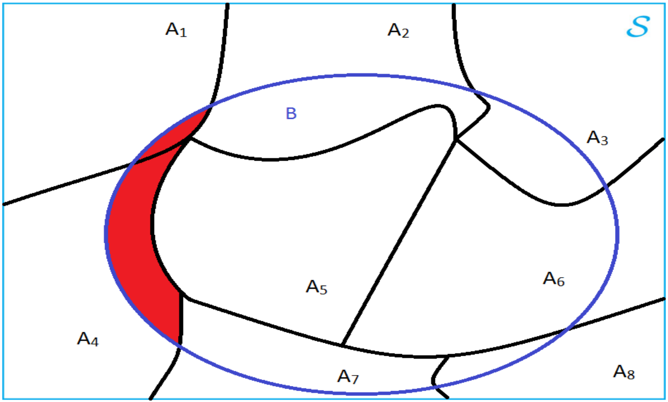
More Generally...



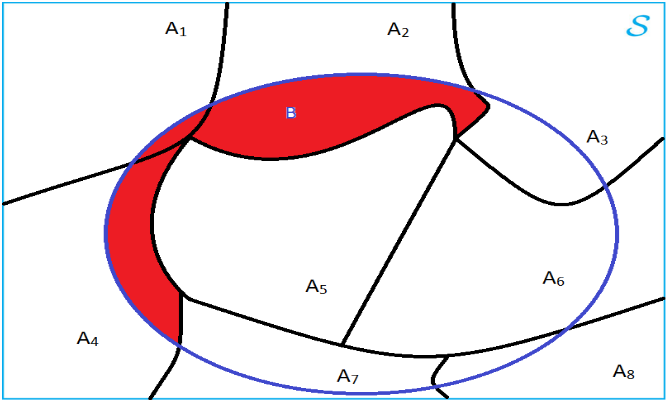
More Generally...



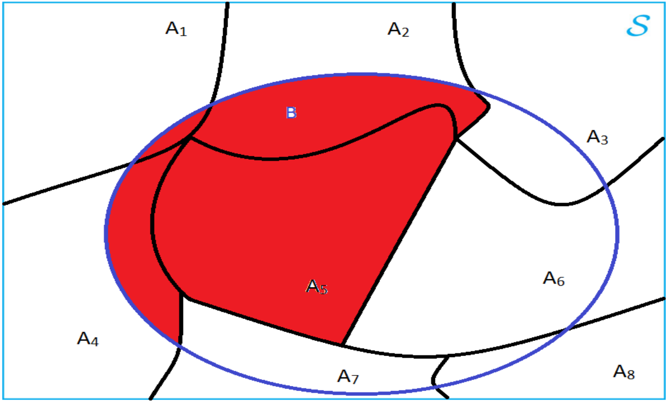
More Generally...



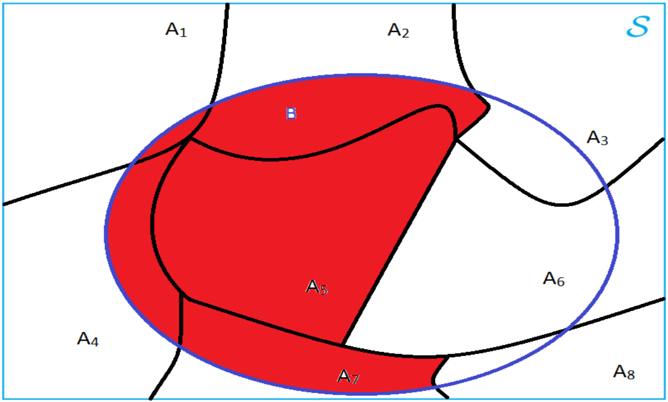
More Generally...



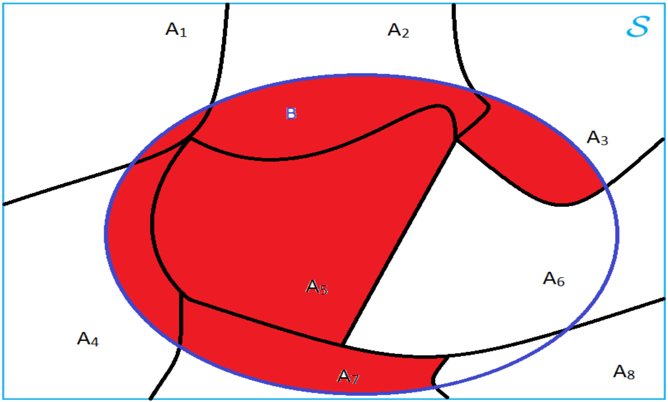
More Generally...



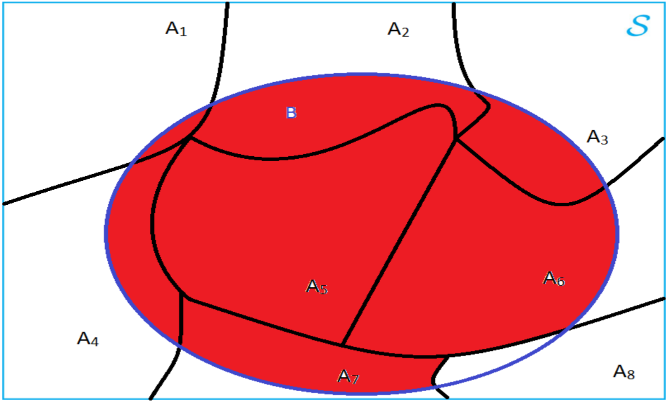
More Generally...



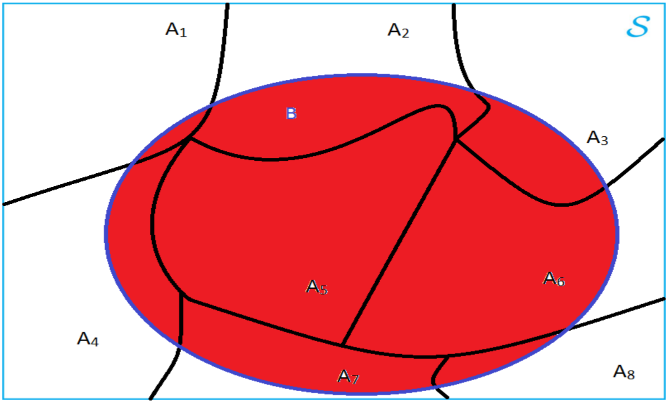
More Generally...



More Generally...



More Generally...



More Formally...

Bayes' Theorem

Let $A_1, \ldots, A_k$ be a partition of the sample space, and let B be any event in S. Then, for each $i = 1, \ldots, k$

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{k} P(A_j)P(B|A_j)} \tag{1}$$

- **Prior** probability of A_i.
- **Conditional** of B given A_i.
- **Total Probability**.
- **Posterior** probability of A_i.

Joint Probability Function

Suppose a *sample* is made up of independent observations $X_1, \ldots, X_n$ all assumed to belong to the same identical pdf (or pmf) $p(x|\theta)$. Then $X_1, \ldots, X_n$ are said to be i.i.d. (independent and identically distributed).

Joint pdf (or pmf)

The joint pdf (or pmf) of an i.i.d. *sample* $X = (X_1, \ldots, X_n)$ is given by

$$p(x_1, \ldots, x_n|\theta) = \prod p(x_i|\theta)$$

Likelihood

Likelihood Function

Let $p(x_1, \ldots, x_n|\theta)$ denote the joint pdf or pmf of the *sample* $X = (X_1, \ldots, X_n)$. Then, given $x_1, \ldots, x_n$ is observed, the function of θ defined by

$$L(\theta|x_1, \ldots, x_n) = p(x_1, \ldots, x_n|\theta)$$

is called the *likelihood function*

Plotting this function vs θ shows how plausible each θ value is.
The maximum of the likelihood function is seen as the most plausible value of θ, given the data that was observed.

More Generally...

Bayes' Theorem

$$p(\theta|x_1, \ldots, x_n) = \frac{p(\theta)p(x_1, \ldots, x_n|\theta)}{\int p(\theta)p(x_1, \ldots, x_n|\theta)d\theta} \tag{2}$$

- **Prior** distribution of θ.
- **Likelihood** of the data.
- Normalizing **constant**.
- **Posterior** distribution of θ.
- Not always tractable.
- Conjugate Priors.
- Gibbs Sampling and the Metropolis Algorithm.

Mean (Expected Value) of a Probability Distribution

- Discrete: $E(x) = \sum_x p(x)x$
  - Roll a single fair die infinitely many times. What is the mean of all the rolls?
  - $$E(x) = (1/6)1 + (1/6)2 + (1/6)3 + (1/6)4 + (1/6)5 + (1/6)6 = 3.5$$
- Continuous: $E(x) = \int_x p(x)xdx$
  - For a Normally distributed (continuous) outcome
  - $$E(x) = \int_x \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) xdx = \mu$$

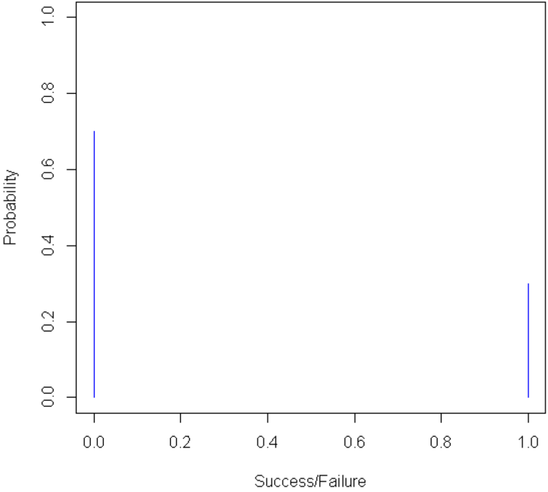
Variance of a Probability Distribution

- Discrete: $Var(x) = \sum_x p(x)(x - E(x))^2$
- Continuous: $Var(x) = \int_x p(x)(x - E(x))^2 dx$
- Note this is just the expected value of $(x - E(x))^2$.
- The positive square root of the variance is the standard deviation.
- **Major point:** The variance can represent our uncertainty about possible beliefs.
  - Less uncertainty about beliefs $\Rightarrow$ smaller variance.
  - Less certain about beliefs $\Rightarrow$ larger variance.

Bernoulli Distribution

- A Bernoulli experiment consists of a single trial with two possible outcomes (success/failure) with success probability π.
- $Bern(\pi)$
- $p(x) = \pi^x (1-\pi)^{1-x} \; x = 0, 1$
- $E(x) = \pi$
- $Var(x) = \pi(1-\pi)$
- Example: Plot a Bern(0.3)
  - `x<-c(0,1)`
  - `plot(x,dbinom(x,1,0.3),ylim=range(0,1),type="h", ylab="Probability",xlab="Success/Failure",col="blue")`

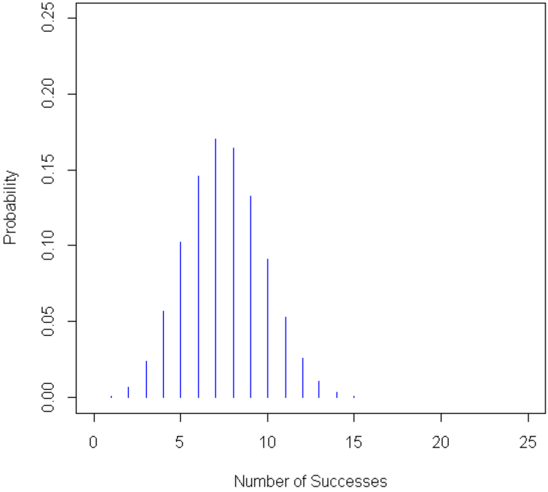
Bernoulli Distribution



Binomial Distribution

- A Bernoulli experiment repeated n times.
- Outcome of interest is the number of successes in those trials.
- $Bin(n, \pi)$
- $p(x) = \binom{n}{x} p^x (1-p)^{n-x} \; x = 0, 1, \ldots, n$
- $E(x) = n\pi$
- $Var(x) = n\pi(1-\pi)$
- Example: Plot a Bin(25,.3)
  - `x<-0:25`
  - `plot(x,dbinom(x,25,.3),ylim=range(0,.25),type="h", ylab="Probability",xlab="Number of Successes",col="blue")`

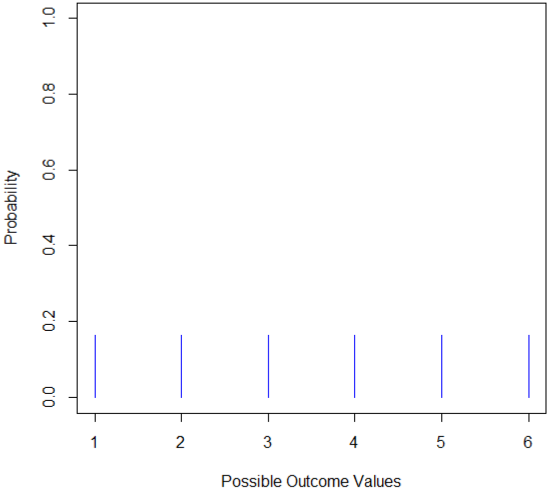
Binomial Distribution



Discrete Uniform Distribution

- Puts equal probability on every observable outcome.
- $DUniform(N)$
- $p(x) = \frac{1}{N}\ x = 1, 2, \ldots, N$
- $E(x) = \frac{N+1}{2}$
- $Var(x) = \frac{(N+1)(N-1)}{12}$
- Example: Plot a discrete uniform for the roll of a fair die.
    - `x<-1:6`
    - ```
      plot(x,rep(1/length(x),length(x)),ylim=range(0,1),type="h",
      ylab="Probability",xlab="Possible Outcome
      Values",col="blue")
      ```
      ```

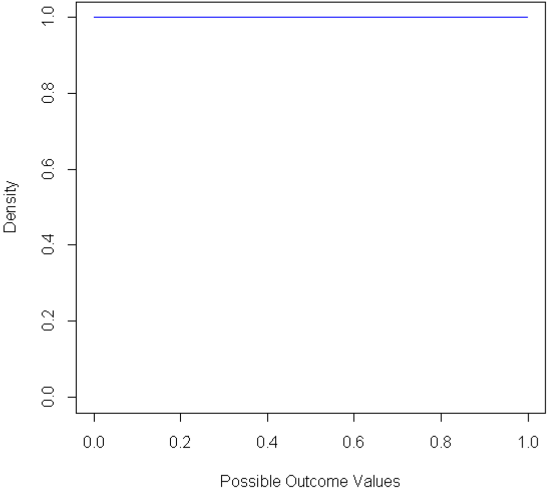
Discrete Uniform Distribution



Uniform Distribution

- Puts equal density on every subinterval of the same length between two points $[a, b]$.
- $Uniform(a, b)$
- $p(x) = \frac{1}{b-a}$ $a \leq x \leq b$
- $E(x) = \frac{b+a}{2}$
- $Var(x) = \frac{(b-a)^2}{12}$
- Example: Plot a Uniform(0,1)
    - `x<-seq(0,1,length=1000)`
    - `plot(x,dunif(x,0,1),ylim=range(0,1),type="l", ylab="Density",xlab="Possible Outcome Values",col="blue")`

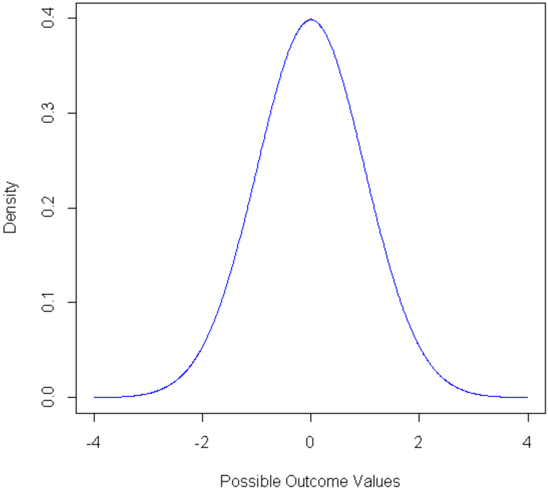
Uniform Distribution



Normal Distribution

- Symmetric, bell-shaped curve.
- Outcome of interest could have any value in the real numbers.
- $N(\mu, \sigma^2)$
- $p(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \ -\infty < x < \infty$
- $E(x) = \mu$
- $Var(x) = \sigma^2$
- Example: Plot a N(0,1)
  - `x<-seq(-4,4,length=1000)`
  - `plot(x,dnorm(x,0,1),ylim=range(0,.25),type="l", ylab="Density",xlab="Possible Outcome Values",col="blue")`

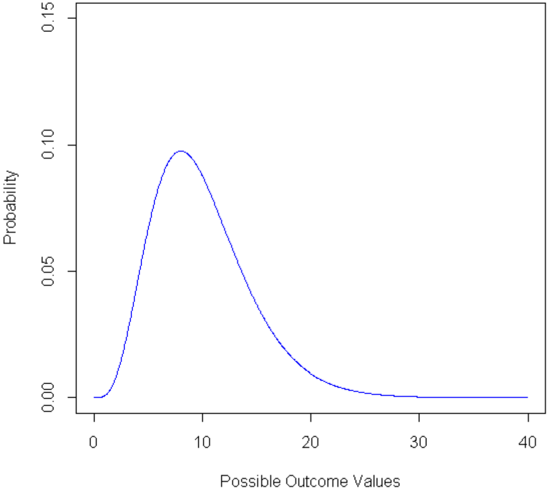
Normal Distribution



Gamma Distribution

- Sometimes used to model lifetimes. Usually right-skewed. Often used as a prior for $1/Var(x)$.
- Outcome of interest must be non negative.
- $Gamma(\alpha, \beta)$
- $p(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}} \; 0 < x < \infty$
- $E(x) = \alpha\beta$
- $Var(x) = \alpha\beta^2$
- Example: Plot a Gamma(5,2)
  - `x<-seq(0,40,length=1000)`
  - `plot(x,dgamma(x,5,.5),ylim=range(0,.15),type="l", ylab="Density",xlab="Possible Outcome Values",col="blue")`

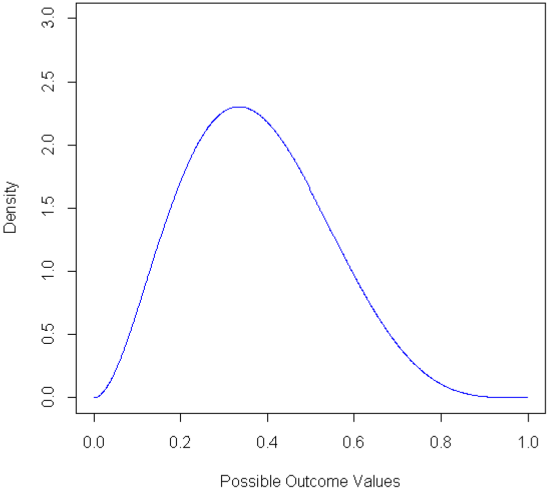
Gamma Distribution



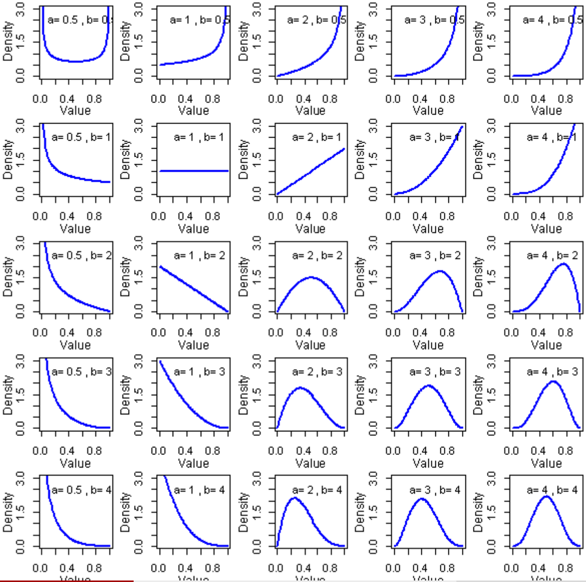
Beta Distribution

- Often used to model probabilities or prevalences.
  Often used as a prior for these quantities.
- Outcome of interest lives on the interval $[0, 1]$.
- $Beta(\alpha, \beta)$
- $p(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \; 0 < x < \infty$
- $E(x) = \frac{\alpha}{\alpha+\beta}$
- $Var(x) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
- Example: Plot a Beta(3,5)
  - `x<-seq(0,1,length=1000)`
  - `plot(x,dbeta(x,3,5),ylim=range(0,3),type="l", ylab="Density",xlab="Possible Outcome Values",col="blue")`

Beta Distribution



Beta Distribution



Conjugate Priors

Posterior has the same distributional form as the prior, then the prior is conjugate for the likelihood.

Table: Common Conjugate Priors and Corresponding Likelihoods.

Likelihood	Prior	Posterior
Bernoulli	Beta	Beta
Binomial	Beta	Beta
Normal	Normal	Normal
Poisson	Gamma	Gamma

Refueling Example Revisited

Table: Results of 24 refueling attempts;
1=Premature Disconnect

Status	Status
0	0
1	0
0	0
0	1
1	1
0	0
0	0
0	1
0	0
0	1
1	0
0	0

- Note that $x_i \in \{0, 1\}$
- Each observation $x_i \sim Bern(\theta)$
- $p(x_i|\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$
- $L(\theta|x_i) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$
- To ease notation let $y = \sum x_i$

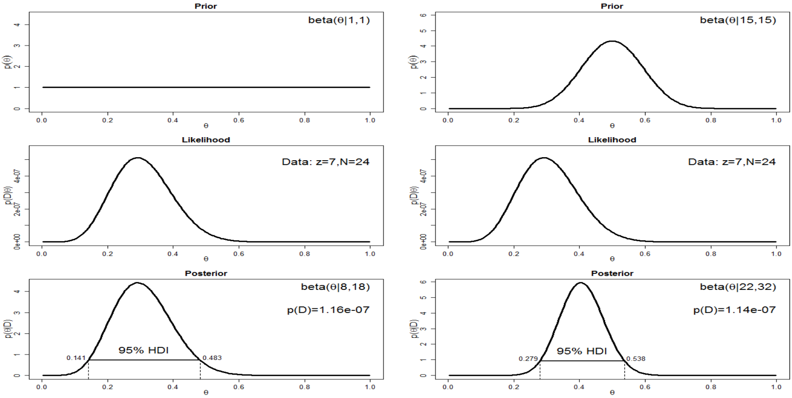
- Prior: $p(\theta) = Beta(\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$
- Mean: $\frac{\alpha}{\alpha+\beta}$
- Variance: $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Refueling Example Revisited

- Posterior: $P(\theta|y) = \frac{p(\theta)p(y|\theta)}{\int p(\theta)p(y|\theta)d\theta} = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}\theta^{y}(1-\theta)^{n-y}}{\int \theta^{\alpha-1}(1-\theta)^{\beta-1}\theta^{y}(1-\theta)^{n-y}d\theta}$
- $= \frac{\theta^{\alpha-1+y}(1-\theta)^{\beta-1+n-y}}{\int \theta^{\alpha-1+y}(1-\theta)^{\beta-1+n-y}d\theta} = \frac{\Gamma(y+\alpha+n-y+\beta)}{\Gamma(y+\alpha)\Gamma(n-y+\beta)}\theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}$
- $= Beta(\alpha + y, n - y + \beta)$
- Mean: $\frac{\alpha+y}{\alpha+\beta+n} = \left(\frac{\alpha+\beta}{\alpha+\beta+n}\right)\left(\frac{\alpha}{\alpha+\beta}\right) + \left(1 - \frac{\alpha+\beta}{\alpha+\beta+n}\right)\left(\frac{y}{n}\right)$
- Variance: $\frac{(\alpha+y)(n-y+\beta)}{(\alpha+\beta+n)^2(\alpha+\beta+2n)}$

Refueling Example Revisited

Consider a $Beta(1, 1)$ and $Beta(15, 15)$ as possible priors



- Mean : 0.308 0.407
- Var : 0.008 0.004

Prior Beliefs

- Knowledge about parameters (or common data biases) BEFORE any data is observed.

We rarely conduct experiments without some degree of prior belief.

- Hypotheses of statistical tests come from prior beliefs.
- Sample size calculations require prior information.
- Identification of outliers depends on what was expected *a priori*.
- Anticipated bias in the data could be handled with prior knowledge of the bias.

Eliciting Prior Beliefs

The goal of eliciting prior beliefs is to turn them into a probability function for use in analysis.

Hierarchy of Prior Knowledge

- Meta analysis of the topic from literature.
- Previous literature of single study.
- Experts' opinion.-Visualizations such as the online MATCH tool are very useful.

In the absence of any of the above sources one may use diffuse (vague) or non-informative priors.

Using the MATCH Uncertainty Elicitation Tool

MATCH Uncertainty Elicitation Tool

- Step 1. From the homepage select the range of possible values using the Upper and Lower limits.
- Step 2. Select Roulette Input Mode. There are other input modes but this one is easy and clever.
- Step 3. Place chips in bins along the specified interval to reflect prior beliefs.
- Step 4. Click "Fitting & Feedback" to see best fitting distributions with accompanying parameter values.

Using the MATCH Uncertainty Elicitation Tool