Bayesian Statistics
Part I: History, Philosophy, and Motivation
Part II: Introduction to Probability

Michael Anderson, PhD

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

March 23, 2021

Who was Thomas Bayes?

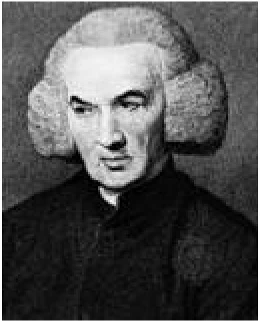
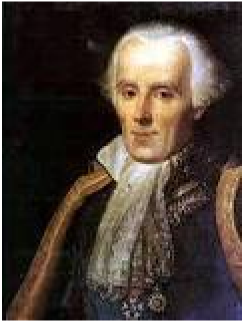
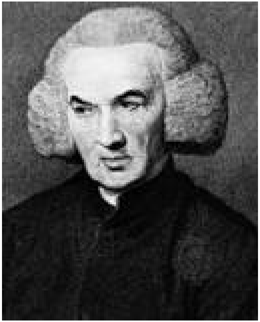
  

Figure: Major players in creation of Bayes' rule

Who was Thomas Bayes?



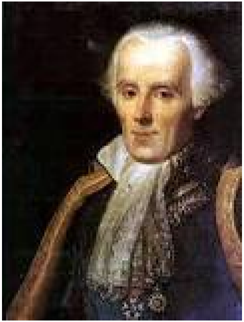


Figure: Major players in creation of Bayes' rule

Sharon Bertsch McGrayne †

- **McGrayne Bios (5:41-9:55)**

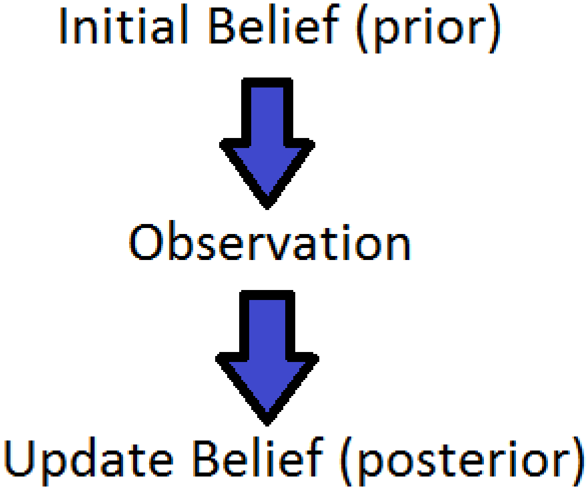
† "The Theory That Would Not Die" How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy"

Bayesian Paradigm

Initial Belief (prior)

Bayesian Paradigm

Bayesian Paradigm



Motivation and History

The allegory of our statistical lives

- Most of us are born Bayesians.

Motivation and History

The allegory of our statistical lives

- Most of us are born Bayesians.

> *"It is remarkable that this science (probability), which originated in the consideration of games of chance, should have become the most important object of human knowledge."*
> $\sim$ Pierre-Simon de Laplace (1749-1827)

Motivation and History

The allegory of our statistical lives

- Most of us are born Bayesians.

"It is remarkable that this science (probability), which originated in the consideration of games of chance, should have become the most important object of human knowledge."
$\sim$ Pierre-Simon de Laplace (1749-1827)

- **McGrayne WWII (16:02-30:23)**
- **McGrayne Air France Flight 447 (2:30-4:00)**

Motivation and History

- But by adolescence...

Motivation and History

- But by adolescence...

  - Fisher's work at Rothamstad advanced experimental design.
    McGrayne Fisher (12:44-16:00)
    McGrayne Obscurity (29:57-36:31)

Motivation and History

- But by adolescence...

  - Fisher's work at Rothamstad advanced experimental design.
    McGrayne Fisher (12:44-16:00)
    McGrayne Obscurity (29:57-36:31)
  - Are these methods appropriate for observational studies?

Motivation and History

- But by adolescence...

  - Fisher's work at Rothamstad advanced experimental design.
    McGrayne Fisher (12:44-16:00)
    McGrayne Obscurity (29:57-36:31)
  - Are these methods appropriate for observational studies?
- Today's message- "It's never too late to have a happy childhood."

Motivation and History

Why go Bayesian?

Motivation and History

Why go Bayesian?

- Another powerful tool for your tool kit.
  - measurement error from misclassification
  - complex dependencies among observations
  - missing data

Motivation and History

Why go Bayesian?

- Another powerful tool for your tool kit.
  - measurement error from misclassification
  - complex dependencies among observations
  - missing data
- Frequentist probabilities don't always make intuitive sense.
  - **McGrayne Frequentist (9:56-12:44)**
  - Example: Probability of an H-bomb accident.

Motivation and History

Why go Bayesian?

- Another powerful tool for your tool kit.
  - measurement error from misclassification
  - complex dependencies among observations
  - missing data
- Frequentist probabilities don't always make intuitive sense.
  - **McGrayne Frequentist (9:56-12:44)**
  - Example: Probability of an H-bomb accident.
- Bayesian "credible intervals" can be correctly interpreted by intro stat students.

Motivation and History

Why go Bayesian?

- Another powerful tool for your tool kit.
  - measurement error from misclassification
  - complex dependencies among observations
  - missing data
- Frequentist probabilities don't always make intuitive sense.
  - **McGrayne Frequentist (9:56-12:44)**
  - Example: Probability of an H-bomb accident.
- Bayesian "credible intervals" can be correctly interpreted by intro stat students.
- P-values and confidence intervals are somewhat ill defined.

Motivation and History

Why go Bayesian?

- Another powerful tool for your tool kit.
  - measurement error from misclassification
  - complex dependencies among observations
  - missing data
- Frequentist probabilities don't always make intuitive sense.
  - **McGrayne Frequentist (9:56-12:44)**
  - Example: Probability of an H-bomb accident.
- Bayesian "credible intervals" can be correctly interpreted by intro stat students.
- P-values and confidence intervals are somewhat ill defined.
  - Example: A study finds that out of 22 subjects with lung cancer 7 are female.

Motivation and History

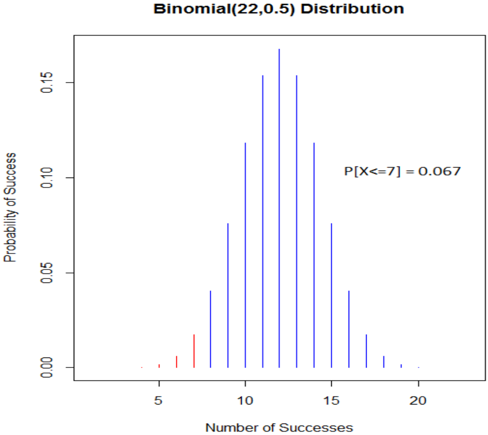
Example: 7 out of 22 subjects with lung cancer are female.

Gender	Gender
0	0
0	0
0	0
1	0
0	0
1	0
1	0
1	0
0	1
0	0
1	1

Motivation and History

Example: 7 out of 22 subjects with lung cancer are female.

Gender	Gender
0	0
0	0
0	0
1	0
0	0
1	0
1	0
1	0
0	1
0	0
1	1



Motivation and History

Example: 7 out of 22 subjects with lung cancer are female.

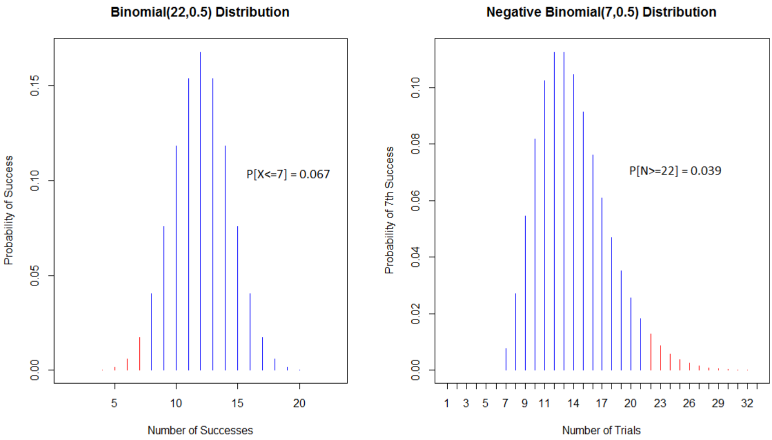


Figure: P-values depend on the researcher's intention.

General Bayes' Theorem

Bayes' Theorem

$$p(\theta|x_1,\ldots,x_n) = \frac{p(\theta)p(x_1,\ldots,x_n|\theta)}{\int p(\theta)p(x_1,\ldots,x_n|\theta)d\theta} \tag{1}$$

General Bayes' Theorem

Bayes' Theorem

$$p(\theta|x_1,\ldots,x_n) = \frac{p(\theta)p(x_1,\ldots,x_n|\theta)}{\int p(\theta)p(x_1,\ldots,x_n|\theta)d\theta} \tag{1}$$

- **Prior** distribution of θ.

General Bayes' Theorem

Bayes' Theorem

$$p(\theta|x_1,\ldots,x_n) = \frac{p(\theta)p(x_1,\ldots,x_n|\theta)}{\int p(\theta)p(x_1,\ldots,x_n|\theta)d\theta} \quad (1)$$

- **Prior** distribution of θ.
- **Likelihood** of the data.

General Bayes' Theorem

Bayes' Theorem

$$p(\theta|x_1,\ldots,x_n) = \frac{p(\theta)p(x_1,\ldots,x_n|\theta)}{\int p(\theta)p(x_1,\ldots,x_n|\theta)d\theta} \quad (1)$$

- **Prior** distribution of θ.
- **Likelihood** of the data.
- Normalizing **constant**.

General Bayes' Theorem

Bayes' Theorem

$$p(\theta|x_1,\ldots,x_n) = \frac{p(\theta)p(x_1,\ldots,x_n|\theta)}{\int p(\theta)p(x_1,\ldots,x_n|\theta)d\theta} \quad (1)$$

- **Prior** distribution of θ.
- **Likelihood** of the data.
- Normalizing **constant**.
- **Posterior** distribution of θ.

Bernoulli Distribution

- A Bernoulli experiment consists of a single trial with two possible outcomes (success/failure) with success probability π.

Bernoulli Distribution

- A Bernoulli experiment consists of a single trial with two possible outcomes (success/failure) with success probability π.
- $Bern(\pi)$

Bernoulli Distribution

- A Bernoulli experiment consists of a single trial with two possible outcomes (success/failure) with success probability π.
- $Bern(\pi)$
- $p(x) = \pi^x(1-\pi)^{1-x} \quad x = 0, 1$

Bernoulli Distribution

- A Bernoulli experiment consists of a single trial with two possible outcomes (success/failure) with success probability π.
- $Bern(\pi)$
- $p(x) = \pi^x(1-\pi)^{1-x} \; x = 0, 1$
- $E(x) = \pi$

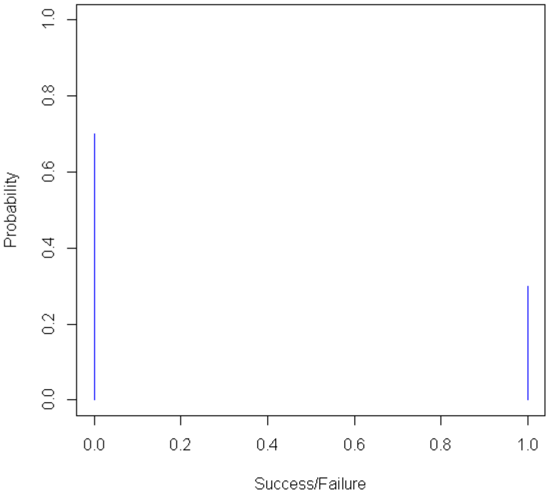
Bernoulli Distribution

- A Bernoulli experiment consists of a single trial with two possible outcomes (success/failure) with success probability π.
- $Bern(\pi)$
- $p(x) = \pi^x(1-\pi)^{1-x} \; x = 0, 1$
- $E(x) = \pi$
- $Var(x) = \pi(1-\pi)$

Bernoulli Distribution

- A Bernoulli experiment consists of a single trial with two possible outcomes (success/failure) with success probability π.
- $Bern(\pi)$
- $p(x) = \pi^x(1-\pi)^{1-x} \quad x = 0, 1$
- $E(x) = \pi$
- $Var(x) = \pi(1-\pi)$
- 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.

Binomial Distribution

- A Bernoulli experiment repeated n times.
- Outcome of interest is the number of success in those trials.

Binomial Distribution

- A Bernoulli experiment repeated n times.
- Outcome of interest is the number of success in those trials.
- $Bin(n, \pi)$

Binomial Distribution

- A Bernoulli experiment repeated n times.
- Outcome of interest is the number of success in those trials.
- $Bin(n, \pi)$
- $p(x) = \binom{n}{x}\pi^x(1-\pi)^{n-x}$ $x = 0, 1, \ldots, n$

Binomial Distribution

- A Bernoulli experiment repeated n times.
- Outcome of interest is the number of success in those trials.
- $Bin(n, \pi)$
- $p(x) = \binom{n}{x}\pi^x(1-\pi)^{n-x} \; x = 0, 1, \ldots, n$
- $E(x) = n\pi$

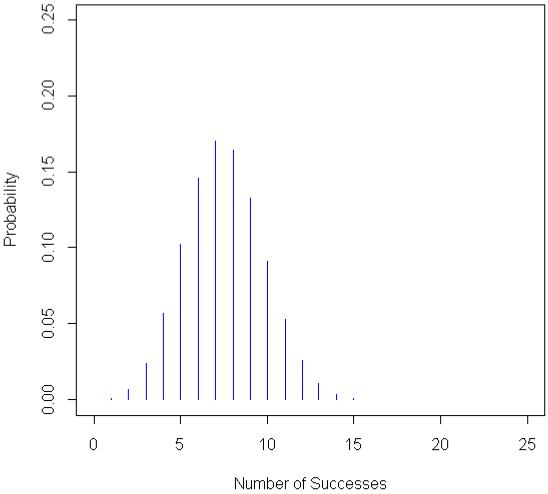
Binomial Distribution

- A Bernoulli experiment repeated n times.
- Outcome of interest is the number of success in those trials.
- $Bin(n, \pi)$
- $p(x) = \binom{n}{x}\pi^x(1-\pi)^{n-x}$ $x = 0, 1, \ldots, n$
- $E(x) = n\pi$
- $Var(x) = n\pi(1-\pi)$

Binomial Distribution

- A Bernoulli experiment repeated n times.
- Outcome of interest is the number of success in those trials.
- $Bin(n, \pi)$
- $p(x) = \binom{n}{x}\pi^x(1-\pi)^{n-x} \; x = 0, 1, \ldots, n$
- $E(x) = n\pi$
- $Var(x) = n\pi(1-\pi)$
- 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



Uniform Distribution

- Puts equal density on every subinterval of the same length between to points $[a, b]$.

Uniform Distribution

- Puts equal density on every subinterval of the same length between to points $[a, b]$.
- $Uniform(a, b)$

Uniform Distribution

- Puts equal density on every subinterval of the same length between to points $[a, b]$.
- $Uniform(a, b)$
- $p(x) = \frac{1}{b-a}\ a \leq x \leq b$

Uniform Distribution

- Puts equal density on every subinterval of the same length between to points $[a, b]$.
- $Uniform(a, b)$
- $p(x) = \frac{1}{b-a}\ a \leq x \leq b$
- $E(x) = \frac{b+a}{2}$

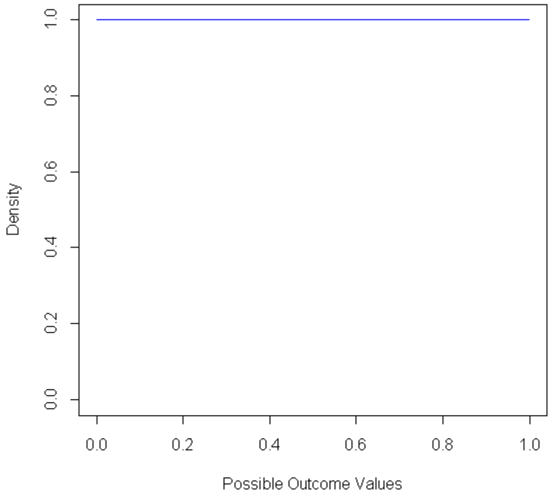
Uniform Distribution

- Puts equal density on every subinterval of the same length between to points $[a, b]$.
- $Uniform(a, b)$
- $p(x) = \frac{1}{b-a}$ $a \leq x \leq b$
- $E(x) = \frac{b+a}{2}$
- $Var(x) = \frac{(b-a)^2}{12}$

Uniform Distribution

- Puts equal density on every subinterval of the same length between to 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.

Normal Distribution

- Symmetric, bell-shaped curve.
- Outcome of interest could have any value in the real numbers.

Normal Distribution

- Symmetric, bell-shaped curve.
- Outcome of interest could have any value in the real numbers.
- $N(\mu, \sigma^2)$

Normal Distribution

- Symmetric, bell-shaped curve.
- Outcome of interest could have any value in the real numbers.
- $N(\mu, \sigma^2)$
- $p(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \; -\infty < x < \infty$

Normal Distribution

- Symmetric, bell-shaped curve.
- Outcome of interest could have any value in the real numbers.
- $N(\mu, \sigma^2)$
- $p(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \; -\infty < x < \infty$
- $E(x) = \mu$

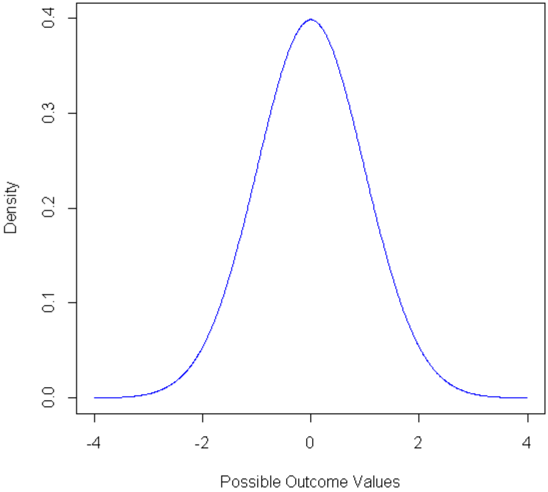
Normal Distribution

- Symmetric, bell-shaped curve.
- Outcome of interest could have any value in the real numbers.
- $N(\mu, \sigma^2)$
- $p(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \; -\infty < x < \infty$
- $E(x) = \mu$
- $Var(x) = \sigma^2$

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



Beta Distribution

- Often used to model probabilities or prevalences.
  Often used as a prior for these quantities.

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 Distribution

- Often used to model probabilities or prevalences.
  Often used as a prior for these quantities.
- Outcome of interest lives on the interval $[0, 1]$.
- $Beta(\alpha, \beta)$

Beta Distribution

- Often used to model probabilities or prevalences.
  Often used as a prior for these quantities.
- Outcome of interest lives on the interval $[0, 1]$.
- $Beta(\alpha, \beta)$
- $p(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \; 0 < x < 1$

Beta Distribution

- Often used to model probabilities or prevalences. Often used as a prior for these quantities.
- Outcome of interest lives on the interval $[0, 1]$.
- $Beta(\alpha, \beta)$
- $p(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \; 0 < x < 1$
- $E(x) = \frac{\alpha}{\alpha+\beta}$

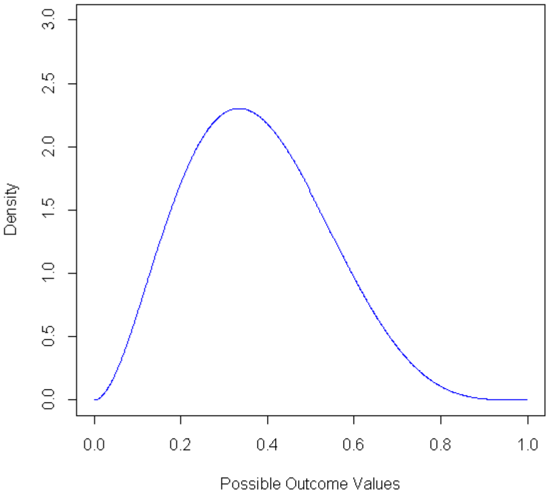
Beta Distribution

- Often used to model probabilities or prevalences. Often used as a prior for these quantities.
- Outcome of interest lives on the interval $[0, 1]$.
- $Beta(\alpha, \beta)$
- $p(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \; 0 < x < 1$
- $E(x) = \frac{\alpha}{\alpha+\beta}$
- $Var(x) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

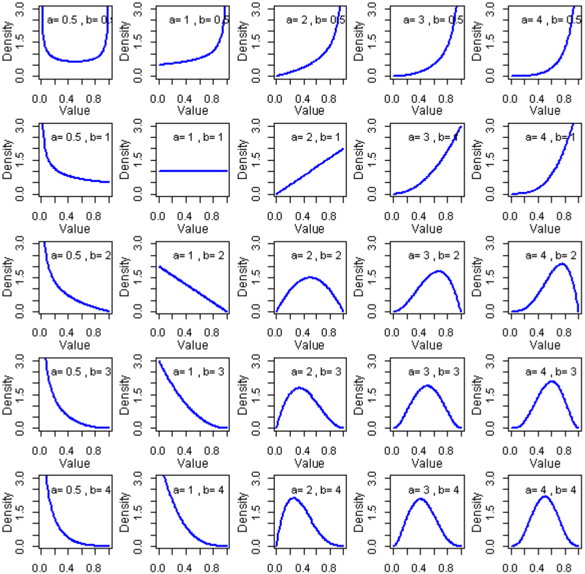
Beta Distribution

- Often used to model probabilities or prevalences.
  Often used as a prior for these quantities.
- Outcome of interest lives on the interval $[0, 1]$.
- $Beta(\alpha, \beta)$
- $p(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \; 0 < x < 1$
- $E(x) = \frac{\alpha}{\alpha+\beta}$
- $Var(x) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
- 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



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) $p(x|\theta)$.

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 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 *sample* $X = (X_1, \ldots, X_n)$ is given by

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

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 *sample* $X = (X_1, \ldots, X_n)$ is given by

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

- Recall that a Bernoulli pmf is defined by $p(x|\theta) = \theta^x(1-\theta)^{1-x}$.

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 *sample* $X = (X_1, \ldots, X_n)$ is given by

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

- Recall that a Bernoulli pmf is defined by $p(x|\theta) = \theta^x(1-\theta)^{1-x}$.
- Then in a sample made up of n independent Bernoulli trials, the joint distribution is given by

$$\prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$$

Likelihood Function

Likelihood Function

Let $\mathrm{p}(x_1, \ldots, x_n|\theta)$ 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

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

is called the *likelihood function*

Likelihood Function

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

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

is called the *likelihood function*

Plotting this function vs θ shows how plausible each θ value is.

Likelihood Function

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

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

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.

Estimating a Single Proportion: Surgery Example

Data is collected from various hospitals in the UK that perform cardiac surgery on babies between 1991 and 1995. The surgery center in Bristol reports 41 deaths and 143 operations during this time. Using vague priors for the proportion of deaths at this surgery center, estimated the posterior prevalence of mortality. See the files

- surgery.data.txt
- surgery.bugs.txt
- surgery.R