Package ‘ProbBayes’

March 6, 2020

Type Package
Title Probability and Bayesian Modeling
Version 1.1
Author Jim Albert <albert@bgsu.edu>
Maintainer Jim Albert <albert@bgsu.edu>
Depends LearnBayes, ggplot2, gridExtra, shiny
Suggests knitr, rmarkdown
URL https://github.com/bayesball/ProbBayes
License GPL (>= 2)
Encoding UTF-8
LazyData true
NeedsCompilation no
Repository CRAN
Date/Publication 2020-03-06 09:40:07 UTC

R topics documented:

animation_ratings .............................................................. 3
arm_height ........................................................................... 4
bar_plot ................................................................................. 4
batting_2018 ......................................................................... 5
bayesian_crank ................................................................. 5
BBS_survey .......................................................... 6
beta_area ............................................................................. 7
beta_data .............................................................................. 7
beta_draw ................................................................................ 8
beta_interval ................................................................. 9
beta_prior_post .......................................................... 9
beta_quantile ............................................................ 10
book_stats ................................................. 11
buffalo_jan ............................................. 11
career_1978 ............................................ 12
centertitle ............................................. 12
CESample ............................................... 13
ChooseBeta ........................................... 14
ComputerPriceSample ............................... 14
Cowles .................................................. 15
DeathHeartAttackDataNYCfull ................... 15
DeathHeartAttackManhattan ....................... 16
draw_two_p ............................................ 16
dsampling ............................................. 17
dspinner .............................................. 18
electricbills .......................................... 18
federalist_word_study .............................. 19
federer_time_to_serve ............................. 20
fire_calls ............................................. 20
football_field_goals ................................ 21
gas2017 ............................................... 21
gibbs_betabin ......................................... 22
gibbs_discrete ....................................... 22
gibbs_normal .......................................... 23
GradSchoolAdmission ............................. 24
Hamilton_can ......................................... 24
house_prices .......................................... 25
HWhours5schools .................................... 26
increasefont ......................................... 26
JAGS_script .......................................... 27
K DramaData .......................................... 27
LaborParticipation ................................... 28
Madison_can .......................................... 28
many_normal_plots ................................ 29
many_spinner_plots ................................ 29
marriage_counts ..................................... 30
mcdonalds .......................................... 31
metropolis ............................................ 31
movies2017 .......................................... 32
nba_guards ............................................ 33
normal_area ........................................... 33
normal_draw .......................................... 34
normal_interval ..................................... 35
normal_quantile .................................... 35
normal_update ....................................... 36
olympic_butterfly ................................... 37
prior_post_plot ...................................... 37
prob_plot ............................................. 38
ProfessorSalary ................................. 39
pt100price ........................................... 39
**animation_ratings**

Ratings for a set of 2010 animation movies

**Usage**

`animation_ratings`

**Format**

A data frame with 55 observations on the following 6 variables.

- **userId** user ID
- **movieId** movie ID
- **rating** numerical rating
- **timestamp** time when the rating was recorded
- **title** name of the movie
- **Group_Number** numerical ID of movie

**Source**

MovieLens by GroupLens Research
Description

Arm span and height measurements for a sample of students

Usage

arm_height

Format

A data frame with 20 observations on the following 2 variables.

- arm: length of arm span in cm
- height: height in cm

Source

Sample of college students

bar_plot

Bar plot of numeric or character data

Description

Constructs frequency bar plot of a vector of numeric data or a vector of character data

Usage

bar_plot(y, ...)

Arguments

- y: vector of outcomes
- ...: title of the graph

Value

A ggplot2 object containing the bar graph.

Author(s)

Jim Albert
Examples

```r
s <- spinner_data(c(1, 2, 2, 1), nsim=100)
bar_plot(s, "Spinner Data")
y <- c(rep("a", 10), rep("b", 5),
     rep("c", 8), rep("d", 4))
bar_plot(y)
```

---

**batting_2018**  
_Batting Statistics for 2018 Season_

**Description**

Batting statistics collected for all players during the first month and remainder of 2018 baseball season

**Usage**

```
battery_2018
```

**Format**

A data frame with 549 observations on the following 5 variables.

- **Name** name of player
- **AB.x** number of at bats in first month
- **H.x** number of hits in first month
- **AB.y** number of at bats in remainder of season
- **H.y** number of hits in remainder of season

**Source**

Data collected from Retrosheet.org.

---

**bayesian_crank**  
_Computes Posterior Probabilities for Discrete Models_

**Description**

Given a data table with columns Prior and Likelihood, computes posterior probabilities

**Usage**

```
bayesian_crank(d)
```
Arguments
d
data frame with columns Prior and Likelihood

Value
data frame with new columns Product and Posterior

Author(s)
Jim Albert

Examples

df <- data.frame(p=c(.1, .3, .5, .7, .9),
Prior=rep(1/5, 5))
y <- 5
n <- 10
df$Likelihood <- dbinom(y, prob=df$p, size=n)
df <- bayesian_crank(df)

BBS_survey
Trend Estimates of Bird Populations

Description
Trend Estimates for 28 Grassland Bird Species

Usage
BBS_survey

Format
A data frame with 28 observations on the following 4 variables.

Species_Name  name of bird species
Trend  trend estimate
SE  standard error of estimate
N_Site  number of observations at site

Source
North American Breeding Bird Survey
beta_area

Displays Areas Under a Beta Curve

Description
Computes and Displays Areas Under a Beta Curve

Usage
beta_area(lo, hi, shape_par, Color = "orange")

Arguments
- lo: lower bound of interval
- hi: upper bound of interval
- shape_par: vector of shape parameters of the beta curve
- Color: color of shading in the graph

Value
ggplot2 object containing the graphical display.

Author(s)
Jim Albert

Examples
lo <- .2
hi <- .4
shape_par <- c(2, 5)
beta_area(lo, hi, shape_par)

beta_data
Simulate random data from a beta curve

Description
Simulate random data from a beta curve

Usage
beta_data(shape_par, nsim=1000)
beta_draw

Arguments

- `shape_par` vector of shape parameters of the beta curve
- `nsim` number of simulations

Value

A vector of random draws from the beta distribution

Author(s)

Jim Albert

Examples

```r
shape_par <- c(12, 8)
beta_data(shape_par, 10)
```

---

**beta_draw**

*Draw a Beta Curve*

Description

Draw a Beta Curve

Usage

```r
beta_draw(shape_pars)
```

Arguments

- `shape_pars` vector of shape parameters of the beta curve

Value

ggplot2 object containing the graphical display.

Author(s)

Jim Albert

Examples

```r
shape_pars <- c(2, 5)
beta_draw(shape_pars)
```
**beta_interval**

*Probability Interval for a Beta Curve*

**Description**

Computes Probability Interval for a Beta Curve

**Usage**

```r
beta_interval(prob, shape_par, Color = "orange")
```

**Arguments**

- `prob`: value of coverage probability
- `shape_par`: vector of shape parameters of the beta curve
- `Color`: color of shading in the graph

**Value**

ggplot2 object containing the graphical display.

**Author(s)**

Jim Albert

**Examples**

```r
shape_par <- c(2, 5)
beta_interval(.5, shape_par)
```

---

**beta_prior_post**

*Plot of Two Beta Curves*

**Description**

Plot of Prior and Posterior Beta Curves

**Usage**

```r
beta_prior_post(prior_shapes, post_shapes)
```

**Arguments**

- `prior_shapes`: vector of shape parameters of the beta prior
- `post_shapes`: vector of shape parameters of the beta posterior
beta_quantile

Displays a Quantile of a Beta Curve

Usage

beta_quantile(prob, shape_par, Color = "orange")

Arguments

- prob: probability value of interest
- shape_par: vector of shape parameters of the beta curve
- Color: color of shading in the graph

Examples

# find the .50 quantile (the median)
prob <- 0.5
shape_par <- c(2, 5)
beta_quantile(prob, shape_par)

# find the .90 quantile (90th percentile)
prob <- 0.9
beta_quantile(prob, shape_par)
**book_stats**

**Text Statistics for Books**

**Description**

Text statistics for a collection of books sold at Amazon.com

**Usage**

book_stats

**Format**

A data frame with 21 observations on the following 3 variables.

- **Book** name of book
- **Complex.Words** percentage of words in the book with three or more syllables
- **Fog.Index** number of years of formal education required to read and understand a passage of text

**Source**

Data collected from Amazon.com website.

---

**buffalo_jan**

**Buffalo snowfall data**

**Description**

Total snowfall in inches for 20 Januarys in Buffalo, New York

**Usage**

buffalo_jan

**Format**

A data frame with 20 observations on the following 2 variables.

- **SEASON** Season
- **JAN** inches of total snowfall

**Source**

career_1978  
*Career Trajectory Data for Baseball Players*

**Description**
Season on-base statistics for collection of MLB baseball players who were born in 1978

**Usage**
career_1978

**Format**
A data frame with 399 observations on the following 6 variables.
- **nameLast**  last name of player
- **Player**  id of player
- **Age**  age of player
- **AgeD**  deviation of age from 30
- **PA**  number of plate appearances
- **OB**  number of on-base events

**Source**
Data collected from Lahman database.

centertitle  
*Centers title in a ggplot2 graphic*

**Description**
Centers and increases font size of a ggplot2 graphic title

**Usage**
centertitle(Color = "blue")

**Arguments**
- **Color**  color of the text in the ggplot2 title

**Value**
ggplot2 theme code to center the title
**Examples**

```r
df <- data.frame(p=c(.1, .3, .5, .7, .9),
                 Prior=rep(1/5, 5))

ggplot(df, aes(p, Prior)) +
  geom_point() +
  ggtitle("My Prior") +
  centertitle()
```

**Description**

Expeditures of U.S. Households

**Usage**

`CEsample`

**Format**

A data frame with 1000 observations on the following 3 variables.

- **UrbanRural** urban/rural status of CU - 1 = urban and 2 = rural
- **TotalIncomeLastYear** amount of CU income before taxes in the last 12 months
- **TotalExpLastQ** CU’s total expenditure in the last quarter

**Source**

U.S. Bureau of Labor Statistics
ChooseBeta

*Shiny App to Choose a Beta Curve*

**Description**

Interactively choose beta curve by selecting the .5 and .9 quantiles

**Usage**

ChooseBeta()

**Value**

None

**Author(s)**

Jim Albert

---

ComputerPriceSample

*Personal Computer Data*

**Description**

Variables on a sample of personal computers

**Usage**

ComputerPriceSample

**Format**

A data frame with 500 observations on the following 5 variables.

- **Price** sales price
- **Speed** clock speed in MHz
- **HardDrive** size of hard drive in MB
- **Ram** size of Ram in MB
- **Premium** premium status of manufacturer

**Source**

Unknown
Description

Data from study to learn about personality determinants of volunteering

Usage

Cowles

Format

A data frame with 1421 observations on the following 5 variables.

- subject: subject number
- neuroticism: measurement of neuroticism
- extraversion: measurement of extraversion
- sex: male or female
- volunteer: no or yes

Source

Unknown.

DeathHeartAttackDataNYCfull

Risk-adjusted mortality outcomes for all NYC hospitals

Description

Reported deaths from heart attack for hospitals in New York City

Usage

DeathHeartAttackDataNYCfull

Format

A data frame with 45 observations on the following 5 variables.

- Hospital: name of hospital
- Borough: borough in New York City
- Type: type of hospital
- Cases: number of heart attack cases
- Deaths: number of deaths
Source

New York State Department of Health

---

DeathHeartAttackManhattan

*Risk-adjusted mortality outcomes for Manhattan hospitals*

Description

Reported deaths from heart attack for hospitals in Manhattan in New York City

Usage

DeathHeartAttackManhattan

Format

A data frame with 13 observations on the following 4 variables.

- **Hospital** name of hospital
- **Type** type of hospital
- **Cases** number of heart attack cases
- **Deaths** number of deaths

Source

New York State Department of Health

---

draw_two_p

*Plot of Distribution of Two Proportions*

Description

Constructs a graph of the probability distribution of two proportions

Usage

draw_two_p(prob_matrix, ...)

Arguments

- **prob_matrix** matrix of probabilities of two proportions with the rows and columns labeled by the values
- **...** other arguments such as the title of the plot
Hypergeometric sampling density

Description

Hypergeometric sampling density

Usage

dsampling(sample_b, pop_N, pop_B, sample_n)

Arguments

- sample_b: number of black balls in sample
- pop_N: number of balls in population
- pop_B: number of black balls in population
- sample_n: number of balls in sample

Value

Value of hypergeometric sampling probability

Author(s)

Jim Albert

Examples

pop_N <- 10
pop_B <- 4
sample_n <- 3
sample_b <- 2
dsampling(sample_b, pop_N, pop_B, sample_n)
**dspinner**

*Computes likelihoods for spinner outcomes*

**Description**

Computes likelihoods for spinner outcomes

**Usage**

```
dspinner(x, Prob)
```

**Arguments**

- `x` vector of spinner observations
- `Prob` matrix of spinner probabilities where each row corresponds to a different spinner

**Value**

column vector consisting of the likelihoods for the different spinners

**Author(s)**

Jim Albert

**Examples**

```r
Prob <- matrix(c(.25, .25, .25, .25,
                 .50, .125, .125, .5,
                 .25, .5, .25, 0), 3, 4, byrow=TRUE)
x <- c(1, 2, 1, 3, 4)
dspinner(x, Prob)
```

**electricbills**

*Electricity Bills*

**Description**

Electricity bills collected for all months for five years

**Usage**

```
electricbills
```
Format

A data frame with 62 observations on the following 3 variables.

- **Year**  year
- **Month**  number of month
- **Amount**  electricity bill in dollars

Source

Data collected for one household in Ohio

---

Federalist_word_study  Frequency use of words for Federalist Papers

Description

Frequency use of words for Federalist Papers written by either Alexander Hamilton or James Madison

Usage

federalist_word_study

Format

A data frame with 56853 observations on the following 7 variables.

- **Name**  name of Federalist paper
- **Total**  total number of words
- **word**  word that is counted
- **N**  frequency of the word
- **Rate**  fraction of words with that word
- **Authorship**  author of paper
- **Disputed**  is authorship disputed?

Source

http://www.gutenberg.org/ebooks/18
federer_time_to_serve  Times to Serve for Roger Federer

Description

Measurements of time to serve for 20 serves of the tennis player Roger Federer

Usage

federer_time_to_serve

Format

A data frame with 20 observations on the following one variable.

time  time to serve in seconds

Source

https://github.com/JeffSackmann

fire_calls  Fire Calls for Zip Code Areas

Description

The number of fire calls and building fires for ten zip codes in Montgomery County, Pennsylvania

Usage

fire_calls

Format

A data frame with 10 observations on the following 3 variables.

Zip_Code  zip code
Fire_Calls  number of fire calls
Building_Fires  number of building fires

Source

kaggle.com
**football_field_goals**  
*Football Field Goals Dataset*

**Description**
Field goal attempt data for three seasons of professional football

**Usage**
football_field_goals

**Format**
A data frame with 3025 observations on the following 5 variables.
- **Team**  name of team
- **Year**  football season
- **Kicker**  last name of kicker
- **Distance**  distance in feet of attempt
- **Success**  attempt was successful (1) or not (0)

**Source**
Data collected by Michael Lopez.

---

**gas2017**  
*Gas bill data*

**Description**
Measurements of average temperature and natural gas bill for each month in 2017

**Usage**
gas2017

**Format**
A data frame with 12 observations on the following 3 variables.
- **Month**  abbreviation of month
- **Temp**  average temperature
- **Bill**  natural gas bill in dollars

**Source**
Personal data collected by a homeowner in Ohio
gibbs_betabin  

_Gibbs sampling of the beta-binomial distribution_

**Description**

Implements Gibbs sampling of the beta-binomial distribution

**Usage**

```r
gibbs_betabin(n, a, b, p = 0.5, iter = 1000)
```

**Arguments**

- `n`: binomial sample size
- `a`: first beta shape parameter
- `b`: second beta shape parameter
- `p`: starting value of proportion in algorithm
- `iter`: number of iterations

**Value**

matrix of simulated draws from the algorithm

**Author(s)**

Jim Albert

**Examples**

```r
sp <- gibbs_betabin(20, 5, 5, 100)
```

---

gibbs_discrete  

_Gibbs sampling of a bivariate discrete distribution_

**Description**

Implements Gibbs sampling for an arbitrary bivariate discrete distribution

**Usage**

```r
gibbs_discrete(p, i = 1, iter = 1000)
```
Arguments

p  matrix defining the probability distribution
i  starting row of the matrix
iter  number of cycles of algorithm

Value

matrix of simulated draws from algorithm

Author(s)

Jim Albert

Examples

p <- matrix(c(4, 3, 2, 1,
  3, 4, 3, 2,
  2, 3, 4, 3,
  1, 2, 3, 4) / 40, 4, 4, byrow = TRUE)
out <- gibbs_discrete(p, 1, 100)

gibbs_normal  
Gibbs sampling of the normal sampling posterior

Description

Implements Gibbs sampling for normal sampling with independent priors on the mean and precision

Usage

gibbs_normal(s, P = 0.002, iter = 1000)

Arguments

s  a list with components y, the observed data, mu0, the prior mean of mu, sigma0, the prior standard deviation of mu, a, the shape parameter of the gamma prior on P, b, the rate parameter of the gamma prior on P
P  starting value of the precision parameter
iter  number of iterations

Value

matrix of simulated draws of (mu, P) from the algorithm

Author(s)

Jim Albert
Examples

```r
s <- list(y = rnorm(20, 5, 2),
    mu0 = 10, sigma0 = 3, a = 1, b = 1)
out <- gibbs_normal(s, P = 0.01, iter=100)
```

GradSchoolAdmission & Graduate School Admission

Description

Study to see what variables are helpful in determining admission to Graduate School

Usage

GradSchoolAdmission

Format

A data frame with 400 observations on the following 3 variables.

- **Admission** student was admitted (1) or not admitted (0)
- **GRE** GRE score
- **GPA** grade point average

Source

Unknown.

Hamilton_can & Frequency use of "can" for Federalist Papers

Description

Frequency use of "can" for Federalist Papers written by Alexander Hamilton

Usage

Hamilton_can
Format
A data frame with 49 observations on the following 6 variables.

Name name of Federalist paper
Total total number of words
word word that is counted
N frequency of the word
Rate fraction of words with that word
Authorship author of paper

Source
http://www.gutenberg.org/ebooks/18

Description
Measurements of house size and selling price for a collection of homes in a city in Ohio

Usage	house_prices

Format
A data frame with 24 observations on the following 2 variables.

price selling price in $1000
size square footage of house

Source
Zillow.com
Homework Hours for Five Schools

Description
Weekly hours spent on homework for students from five schools

Usage
HWhours5schools

Format
A data frame with 116 observations on the following 2 variables.
school  school number of student
hours    weekly hours spent on homework

Source
Unknown.

increasefont

Description
Increases font size on all text in a ggplot2 graphic

Usage
increasefont(Size = 18)

Arguments
Size  font size of all textual elements in a ggplot2 graphic

Value
ggplot2 theme code to increase the font size

Author(s)
Jim Albert
Examples

```r
df <- data.frame(p=c(.1, .3, .5, .7, .9),
                 Prior=rep(1/5, 5))
ggplot(df, aes(p, Prior)) +
geom_point() + increasefont()
```

Description

Model script for JAGS to fit a particular Bayesian model. Currently the possible models are "beta_binomial", "hier_normal", "hier_trajectory", "normal", "regression", "regression_cond_means", and "trajectory".

Usage

```r
JAGS_script(model)
```

Arguments

- `model` name of the model

Value

A character string containing the model script

---

K Drama Data  

Korean Drama Ratings

Description

Ratings of Korean dramas broadcast during different days of the week and different producers

Usage

```r
K Drama Data
```

Format

A data frame with 101 observations on the following 5 variables.

- **Drama** name of drama
- **Schedule** indicator of what day the drama was broadcast
- **Producer** indicator of the producer of the drama
- **Rating** rating of the drama
- **Date** date of rating
Source

AGB Nielsen Media Research Group

LaborParticipation    U.S. Women Labor Participation

Description

U.S. women labor participation and family income

Usage

LaborParticipation

Format

A data frame with 753 observations on the following 2 variables.

Participation  labor participation of the wife
FamilyIncome   family income exclusive of wife’s income in $1000

Source

University of Michigan Panel Study of Income Dynamics

Madison_can    Frequency use of “can” for Federalist Papers

Description

Frequency use of "can" for Federalist Papers written by James Madison

Usage

Madison_can

Format

A data frame with 49 observations on the following 6 variables.

Name  name of Federalist paper
Total total number of words
word word that is counted
N frequency of the word
Rate fraction of words with that word
Authorship author of paper
many_normal_plots

Source

http://www.gutenberg.org/ebooks/18

many_normal_plots  Graph of several normal curves

Description

Graph of several normal curves

Usage

many_normal_plots(list_normal_par)

Arguments

list_normal_par
  list of vectors, where each vector is a mean and standard deviation for a normal distribution

Value

ggplot2 object containing the graphical display.

Author(s)

Jim Albert

Examples

list_normal_par <- list(c(100, 15),
                         c(110, 15), c(120, 15))
many_normal_plots(list_normal_par)

many_spinner_plots  Graphs a collection of spinners

Description

Graphs a collection of spinners

Usage

many_spinner_plots(list_regions)
Arguments

list_regions  list of vectors of integer areas for the spins 1, 2, ...

Value

A ggplot2 object containing the spinner displays

Author(s)

Jim Albert

Examples

regions1 <- c(1, 1, 1)
regions2 <- c(2, 1, 2, 1)
many_spinner_plots(list(regions1, regions2))

marriage_counts  Annual Marriage Counts in Italy

Description

Annual marriage counts per 1000 of the population in Italy from 1936 to 1951

Usage

marriage_counts

Format

A data frame with 16 observations on the following 2 variables.

Year  year
Count  count of marriages per 1000 people

Source

Unknown.
Description

Serving size and calories for a selection of sandwiches from McDonalds

Usage

mcdonalds

Format

A data frame with 11 observations on the following 3 variables.

- **Sandwich**  name of sandwich
- **Size**  serving size in grams
- **Calories**  calories of sandwich

Source

McDonalds restaurant

---

Description

Implements Metropolis sampling for an arbitrary continuous probability distribution

Usage

metropolis(logpost, current, C, iter, ...)

Arguments

- **logpost**  function definition of the log probability function
- **current**  starting value of algorithm
- **C**  half-width of proposal interval
- **iter**  number of iterations
- **...**  other inputs needed in logpost function
movies2017

Value

S vector of simulated values
accept_rate acceptance rate of algorithm

Author(s)

Jim Albert

Examples

```r
lpost <- function(theta, s){
  dnorm(s$ybar, theta, s$se, log = TRUE) +
  dcauchy(theta, s$loc, s$scale, log = TRUE)
}
s <- list(ybar = 20,
  se = 0.4,
  loc = 10,
  scale = 2)
post <- metropolis(lpost, 10, 20, 100, s)
```

movies2017 Movies Sales Data

Description

Weekend and gross sales for a selection of movies released in 2017

Usage

movies2017

Format

A data frame with 10 observations on the following 3 variables.

Movie name of movie
Weekend opening weekend sales in millions of dollars
Gross gross sales in millions of dollars

Source

Internet Movie Database
**nba_guards**  

*Basketball Shooting Data for Point Guards*

**Description**  
Field goal and free throw shooting data for a collection of great NBA point guards

**Usage**  
nba_guards

**Format**  
A data frame with 230 observations on the following 6 variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Player</td>
<td>name of player</td>
</tr>
<tr>
<td>Age</td>
<td>age of player</td>
</tr>
<tr>
<td>FG</td>
<td>field goals</td>
</tr>
<tr>
<td>FGA</td>
<td>field goal attempts</td>
</tr>
<tr>
<td>FT</td>
<td>free throws</td>
</tr>
<tr>
<td>FTA</td>
<td>free throw attempts</td>
</tr>
</tbody>
</table>

**Source**  
Data collected from Basketball-Reference.com.

---

**normal_area**  

*Displays Area Under a Normal Curve*

**Description**  
Computes and Displays Area Under a Normal Curve

**Usage**  

```r
normal_area(lo, hi, normal_pars, Color = "orange")
```

**Arguments**

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>lo</td>
<td>lower bound of interval</td>
</tr>
<tr>
<td>hi</td>
<td>upper bound of interval</td>
</tr>
<tr>
<td>normal_pars</td>
<td>vector of mean and standard deviation of the normal curve</td>
</tr>
<tr>
<td>Color</td>
<td>color of shading in plot</td>
</tr>
</tbody>
</table>
normal_draw

Value

ggplot2 object containing the graphical display.

Author(s)

Jim Albert

Examples

lo <- 10
hi <- 20
normal_pars <- c(25, 10)
normal_area(lo, hi, normal_pars)

normal_draw

Draws a Normal Curve

Description

Draws a Normal Curve

Usage

normal_draw(normal_pars, Color = "red")

Arguments

normal_pars vector of mean and standard deviation of the normal curve
Color color of line in plot

Value

ggplot2 object containing the graphical display.

Author(s)

Jim Albert

Examples

normal_pars <- c(2, 1)
normal_draw(normal_pars)
normal_interval  

**Probability Interval for a Normal Curve**

**Description**
Computes "equal-tails" probability interval for a normal curve

**Usage**

```r
normal_interval(prob, normal_pars, Color = "orange")
```

**Arguments**
- `prob`: value of coverage probability
- `normal_pars`: vector of mean and standard deviation of the normal curve
- `Color`: color of shading in plot

**Value**
ggplot2 object containing the graphical display.

**Author(s)**
Jim Albert

**Examples**

```r
normal_pars <- c(2, 0.5)
prob <- 0.5
normal_interval(prob, normal_pars)
```

---

normal_quantile  

**Displays a Quantile of a Normal Curve**

**Description**
Displays a Quantile of a Normal Curve

**Usage**

```r
normal_quantile(prob, normal_pars, Color = "orange")
```

**Arguments**
- `prob`: probability value of interest
- `normal_pars`: vector of mean and standard deviation of the normal curve
- `Color`: color of shading in plot
Value

ggplot2 object containing the graphical display.

Author(s)

Jim Albert

Examples

```r
normal_pars <- c(100, 10)
prob <- 0.7
normal_quantile(prob, normal_pars)
```

---

**normal_update**

Updates a Normal Prior with Normal Data

Description

Finds the parameters of the normal posterior with normal data and a normal prior

Usage

```r
normal_update(prior, data, teach=FALSE)
```

Arguments

- `prior`: vector with components mean and sd of the normal prior
- `data`: vector with components the sample mean and the standard error of the estimate
- `teach`: logical variable indicating the form of the output

Value

If `teach = TRUE`, returns data frame that displays the mean, precision, and standard deviation for the prior, data, and posterior. If `teach = FALSE`, returns a vector with mean and standard deviation of the posterior.

Author(s)

Jim Albert

Examples

```r
prior <- c(100, 10)
data <- c(110, 15)
normal_update(prior, data)
normal_update(prior, data, teach=TRUE)
```
olympic_butterfly Winning Times in the 100 Meter Butterfly Race

Description

Winning times in seconds for the men’s and women’s 100m butterfly race for the Olympics from 1964 through 2016.

Usage

  olympic_butterfly

Format

A data frame with 28 observations on the following 3 variables.

Year  year of Olympics
Gender  gender
Time  winning time in seconds

Source

https://www.olympic.org/swimming/

prior_post_plot Graphs prior and posterior probabilities

Description

Graphs prior and posterior probabilities from a discrete Bayesian model

Usage

  prior_post_plot(d, Color = "orange")

Arguments

  d  data frame where the first column are the model values, and columns named Prior and Posterior
  Color  fill color for the bars

Value

  ggplot2 object containing the graphical display.
prob_plot

Constructs a graph of a probability distribution

Description

Constructs a graph of a discrete probability distribution

Usage

prob_plot(d, Color = "red", Size = 1.5)

Arguments

d data frame where the first two columns are the variable and associated probabilities
Color color of line in plot
Size width of line in plot

Value

A ggplot2 object containing the plot display

Author(s)

Jim Albert

Examples

d <- data.frame(x=1:5,
    Probability=c(.1, .2, .3, .3, .1))
prob_plot(d)
**ProfessorSalary**  
*Professor Salary Study*

**Description**  
Study on inputs that impact a salary of a professor

**Usage**

`ProfessorSalary`

**Format**

A data frame with 397 observations on the following 7 variables.

- `subject`  subject id
- `rank`  professor rank
- `discipline`  A is theoretical and B is applied
- `yrs.since.phd`  number of years since receipt of doctorate
- `yrs.service`  number of years of service
- `sex`  Female or Male
- `salary`  nine-month salary in dollars

**Source**

Unknown.

---

**pt100price**  
*Prices of One Carat Diamonds*

**Description**

Prices of a sample of one carat diamonds

**Usage**

`pt100price`

**Format**

A data frame with 25 observations on the following 2 variables.

- `diamond`  index of diamond
- `price`  price divided by 100

**Source**

Unknown.
pt99price  
**Prices of 0.99 Carat Diamonds**

**Description**
Prices of a sample of 0.99 carat diamonds

**Usage**
pt99price

**Format**
A data frame with 23 observations on the following 2 variables.
- **diamond**  index of diamond
- **price**  price divided by 100

**Source**
Unknown.

pythag2018  
**Baseball Win-Loss Records**

**Description**
Final standings of the MLB baseball teams in the 2018 season

**Usage**
pythag2018

**Format**
A data frame with 30 observations on the following 7 variables.
- **Team**  team abbreviation
- **League**  league abbreviation
- **W**  number of wins
- **L**  number of losses
- **Pct**  proportion of wins
- **R**  average runs scored
- **RA**  average runs allowed

**Source**
Lahman database
random_walk  

Metropolis sampling of a discrete distribution

Description

Implements Metropolis sampling for an arbitrary discrete probability distribution

Usage

random_walk(pd, start, num_steps)

Arguments

pd  
function containing discrete probability function on the integers 1, 2, ...

start  
starting value of algorithm

num_steps  
number of iterations of algorithm

Value

A vector of simulated values

Author(s)

Jim Albert

Examples

# random walk through a binomial distribution
pd <- function(x){
  dbinom(x, size = 10, prob = 0.5)
}
start <- 4
num_steps <- 50
out <- random_walk(pd, start, num_steps)

ScoreData  

Scores on Achievement Exam

Description

Scores on a 20-question T/F exam

Usage

ScoreData
Format
A data frame with 30 observations on the following 2 variables.

Person subject id
Score number correct in 20-question exam

Source
Data randomly generated.

Description
Sample of sleeping times for a single night for a sample of college students

Usage

sleeping_times

Format
A data frame with 14 observations on the following single variable.

hours number of hours of sleep

Source
Personal collection

Description
Computes and plots the posterior distribution of spinners given a sequence of spins

Usage

spinner_bayes(list_regions,
prior,
data,
plot=TRUE)
spin\textunderscore data

Arguments

\begin{itemize}
  \item \texttt{list\_regions} \hspace{1cm} list of vectors of integer areas for the spins 1, 2, ...
  \item \texttt{prior} \hspace{1cm} a vector containing the prior probabilities for the spinners
  \item \texttt{data} \hspace{1cm} a vector containing the spin values where 1, 2, 3, ... are the possible spins
  \item \texttt{plot} \hspace{1cm} if \texttt{plot=}TRUE, a comparative graph of the prior and posterior probabilities is displayed
\end{itemize}

Value

A data frame with variables \texttt{Spinner}, \texttt{Prior}, \texttt{Likelihood}, \texttt{Product}, and \texttt{Posterior}

Author(s)

Jim Albert

Examples

\begin{verbatim}
regions1 <- c(1, 1, 1)
regions2 <- c(2, 1, 2, 1)
data <- c(1, 1, 1, 2)
spinner\_bayes(list(regions1, regions2),
               prior=c(0.5, 0.5),
               data)
\end{verbatim}

\begin{verbatim}
spinner\_data
\end{verbatim} \hspace{1cm} \textit{Simulate random data from a spinner}

Description

Simulate random data from a spinner

Usage

\begin{verbatim}
spinner\_data(regions, nsim=1000)
\end{verbatim}

Arguments

\begin{itemize}
  \item \texttt{regions} \hspace{1cm} vector of integer values for the spins 1, 2, ...
  \item \texttt{nsim} \hspace{1cm} number of spins
\end{itemize}

Value

A vector of random spins from the spinner

Author(s)

Jim Albert
**Examples**

```r
regions <- c(2, 1, 1, 2)
spinner_data(regions, nsim=20)
```

---

**spinner_likelihoods**  
*Computes likelihood matrix for many spinners*

**Description**

Computes likelihood matrix for many spinners

**Usage**

```r
spinner_likelihoods(regions)
```

**Arguments**

- `regions`  
  list of vectors of integer areas for the spins 1, 2, ...

**Value**

A matrix where each row corresponds to the outcome probabilities for one spinner.

**Author(s)**

Jim Albert

**Examples**

```r
sp1 <- c(2, 1, 1)
sp2 <- c(1, 1, 1, 1)
regions <- list(sp1, sp2)
spinner_likelihoods(regions)
```

---

**spinner_plot**  
*Constructs a spinner*

**Description**

Constructs a spinner with different regions

**Usage**

```r
spinner_plot(probs, ...)
```
spinner_probs

Arguments

probs       vector of probabilities for the spins 1, 2, ...
...         optional vector of values and title

Value

A ggplot2 object containing the spinner display

Author(s)

Jim Albert

Examples

probs <- rep(.2, 5)
spinner_plot(probs,
values=c("A", "B", "C", "D", "E"),
title="My Spinner")
# probs does not need to be normalized
spinner_plot(c(1, 2, 1, 2))
### taxi_fares

**Description**
Sample of taxi fares from a particular city

**Usage**
taxi_fares

**Format**
A data frame with 20 observations on the following single variable.

- **fare** taxi cab fare

**Source**
Personal collection

### tennis_serve

**Description**
Data on time to serve for six professional tennis players

**Usage**
tennis_serve

**Format**
A data frame with 6 observations on the following 3 variables.

- **Player** last name of player
- **n** number of serves
- **ybar** mean time to serve

**Source**
https://github.com/JeffSackmann
testing_prior

Testing prior for two proportions

Description

Constructs a discrete distribution for two proportions under a testing or uniform hypotheses

Usage

```
testing_prior(lo=.1, hi=.9, n_values=9,
              pequal=0.5, uniform=FALSE)
```

Arguments

- `lo`: minimum value of each proportion
- `hi`: maximum value of each proportion
- `n_values`: number of values of each proportion
- `pequal`: probability of the equality of the two proportions
- `uniform`: indicates if a uniform prior is desired

Value

matrix of probabilities where the rows and columns are labeled by the values of the proportions

Author(s)

Jim Albert

Examples

```
# testing prior where each proportion is
# .1, .3, .5, .7, .9
Prob <- testing_prior(.1, .9, 5)
# uniform prior over same proportion values
Prob <- testing_prior(.1, .9, 5, uniform=TRUE)
```
### trout20  
*Mike Trout Statcast Data*

**Description**
Launch speed and distance traveled for a sample of balls hit by the baseball player Mike Trout

**Usage**
`trout20`

**Format**
A data frame with 25 observations on the following 2 variables.

- **launch_speed**  launch speed in mph
- **hit_distance_sc**  distance in feet

**Source**
Major League Baseball Advanced Media

### two_players_time_to SERVE  
*Times to Serve for Two Tennis Players*

**Description**
Measurements of time to serve serves of the tennis players Roger Federer and Rafael Nadal

**Usage**
`two_players_time_to_serve`

**Format**
A data frame with 100 observations on the following 2 variables.

- **Player**  last name of player
- **time**  time to serve in seconds

**Source**
https://github.com/JeffSackmann
Summaries of a probability matrix

**Description**
Computes posterior of difference P2 - P1 of a probability matrix of two proportions

**Usage**
two_p_summarize(prob_matrix)

**Arguments**
- prob_matrix: probability matrix where the rows and columns are labeled with the values of the proportions

**Value**
data frame with variables diff21 and Prob where diff21 = P2 - P1

**Author(s)**
Jim Albert

**Examples**
```r
# use uniform prior over values .2, .3, .4
prob_matrix <- testing_prior(.2, .4, 3, uniform=TRUE)
two_p_summarize(prob_matrix)
```

---

Posterior updating of two proportions

**Description**
Computes posterior distribution of two proportions with a discrete prior

**Usage**
two_p_update(prior, s1f1, s2f2)

**Arguments**
- prior: prior probability matrix where the rows and columns are labeled with the values of the proportions
- s1f1: number of successes and number of failures from first sample
- s2f2: number of successes and number of failures from second sample
value

posterior probability matrix

author(s)

Jim Albert

Examples

```r
prior <- testing_prior()
s1f1 <- c(3, 10)
s2f2 <- c(8, 20)
two_p_update(prior, s1f1, s2f2)
```

---

web_visits  

*Website tracking data*

**Description**

Number of visits to a blog website for different weeks and days of the week

**Usage**

```r
web_visits
```

**Format**

A data frame with 28 observations on the following 3 variables.

- **Week**  week number
- **Day**  day of the week
- **Count**  number of website visits

**Source**

Personal data collected from Wordpress.com
## Index

*Topic **datasets**

<table>
<thead>
<tr>
<th>Dataset Name</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>animation_ratings</td>
<td>3</td>
</tr>
<tr>
<td>arm_height</td>
<td>4</td>
</tr>
<tr>
<td>batting_2018</td>
<td>5</td>
</tr>
<tr>
<td>BBS_survey</td>
<td>6</td>
</tr>
<tr>
<td>book_stats</td>
<td>11</td>
</tr>
<tr>
<td>buffalo_jan</td>
<td>11</td>
</tr>
<tr>
<td>career_1978</td>
<td>12</td>
</tr>
<tr>
<td>CEsample</td>
<td>13</td>
</tr>
<tr>
<td>ComputerPriceSample</td>
<td>14</td>
</tr>
<tr>
<td>Cowles</td>
<td>15</td>
</tr>
<tr>
<td>DeathHeartAttackDataNYCfull</td>
<td>15</td>
</tr>
<tr>
<td>DeathHeartAttackManhattan</td>
<td>16</td>
</tr>
<tr>
<td>electricbills</td>
<td>18</td>
</tr>
<tr>
<td>federalist_word_study</td>
<td>19</td>
</tr>
<tr>
<td>federer_time_to_serve</td>
<td>20</td>
</tr>
<tr>
<td>fire_calls</td>
<td>20</td>
</tr>
<tr>
<td>football_field_goals</td>
<td>21</td>
</tr>
<tr>
<td>gas2017</td>
<td>21</td>
</tr>
<tr>
<td>GradSchoolAdmission</td>
<td>24</td>
</tr>
<tr>
<td>Hamilton_can</td>
<td>24</td>
</tr>
<tr>
<td>house_prices</td>
<td>25</td>
</tr>
<tr>
<td>HWhours5schools</td>
<td>26</td>
</tr>
<tr>
<td>JAGS_script</td>
<td>27</td>
</tr>
<tr>
<td>KDramaData</td>
<td>27</td>
</tr>
<tr>
<td>LaborParticipation</td>
<td>28</td>
</tr>
<tr>
<td>Madison_can</td>
<td>28</td>
</tr>
<tr>
<td>marriage_counts</td>
<td>30</td>
</tr>
<tr>
<td>mcdonalds</td>
<td>31</td>
</tr>
<tr>
<td>movies2017</td>
<td>32</td>
</tr>
<tr>
<td>nba_guards</td>
<td>33</td>
</tr>
<tr>
<td>olympic_butterfly</td>
<td>37</td>
</tr>
<tr>
<td>ProfessorSalary</td>
<td>39</td>
</tr>
<tr>
<td>pt100price</td>
<td>39</td>
</tr>
<tr>
<td>pt99price</td>
<td>40</td>
</tr>
<tr>
<td>pythag2018</td>
<td>40</td>
</tr>
<tr>
<td>ScoreData</td>
<td>41</td>
</tr>
<tr>
<td>sleeping_times</td>
<td>42</td>
</tr>
<tr>
<td>taxi_fares</td>
<td>46</td>
</tr>
<tr>
<td>tennis_serve</td>
<td>46</td>
</tr>
<tr>
<td>trout20</td>
<td>48</td>
</tr>
<tr>
<td>two_players_time_to_serve</td>
<td>48</td>
</tr>
<tr>
<td>web_visits</td>
<td>50</td>
</tr>
<tr>
<td>animation_ratings</td>
<td>3</td>
</tr>
<tr>
<td>arm_height</td>
<td>4</td>
</tr>
<tr>
<td>bar_plot</td>
<td>4</td>
</tr>
<tr>
<td>batting_2018</td>
<td>5</td>
</tr>
<tr>
<td>bayesian_crank</td>
<td>5</td>
</tr>
<tr>
<td>BBS_survey</td>
<td>6</td>
</tr>
<tr>
<td>beta_area</td>
<td>7</td>
</tr>
<tr>
<td>beta_data</td>
<td>7</td>
</tr>
<tr>
<td>beta_draw</td>
<td>8</td>
</tr>
<tr>
<td>beta_interval</td>
<td>9</td>
</tr>
<tr>
<td>beta_prior_post</td>
<td>9</td>
</tr>
<tr>
<td>beta_quantile</td>
<td>10</td>
</tr>
<tr>
<td>book_stats</td>
<td>11</td>
</tr>
<tr>
<td>buffalo_jan</td>
<td>11</td>
</tr>
<tr>
<td>career_1978</td>
<td>12</td>
</tr>
<tr>
<td>centertitle</td>
<td>12</td>
</tr>
<tr>
<td>CEsample</td>
<td>13</td>
</tr>
<tr>
<td>ChooseBeta</td>
<td>14</td>
</tr>
<tr>
<td>ComputerPriceSample</td>
<td>14</td>
</tr>
<tr>
<td>Cowles</td>
<td>15</td>
</tr>
<tr>
<td>DeathHeartAttackDataNYCfull</td>
<td>15</td>
</tr>
<tr>
<td>DeathHeartAttackManhattan</td>
<td>16</td>
</tr>
<tr>
<td>draw_two_p</td>
<td>16</td>
</tr>
<tr>
<td>dsampling</td>
<td>17</td>
</tr>
<tr>
<td>dspinner</td>
<td>18</td>
</tr>
<tr>
<td>electricbills</td>
<td>18</td>
</tr>
<tr>
<td>federalist_word_study</td>
<td>19</td>
</tr>
<tr>
<td>federer_time_to_serve</td>
<td>20</td>
</tr>
<tr>
<td>fire_calls</td>
<td>20</td>
</tr>
<tr>
<td>football_field_goals</td>
<td>21</td>
</tr>
</tbody>
</table>