Package ‘ProbBayes’

October 12, 2022

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
<table>
<thead>
<tr>
<th>R topics documented:</th>
<th></th>
</tr>
</thead>
<tbody>
<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_toServe</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>gibbs_betabin</td>
<td>22</td>
</tr>
<tr>
<td>gibbs_discrete</td>
<td>22</td>
</tr>
<tr>
<td>gibbs_normal</td>
<td>23</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>increasefont</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>many_normal_plots</td>
<td>29</td>
</tr>
<tr>
<td>many_spinner_plots</td>
<td>29</td>
</tr>
<tr>
<td>marriage_counts</td>
<td>30</td>
</tr>
<tr>
<td>mcdonalds</td>
<td>31</td>
</tr>
<tr>
<td>metropolis</td>
<td>31</td>
</tr>
<tr>
<td>movies2017</td>
<td>32</td>
</tr>
<tr>
<td>nba Guards</td>
<td>33</td>
</tr>
<tr>
<td>normal_area</td>
<td>33</td>
</tr>
<tr>
<td>normal_draw</td>
<td>34</td>
</tr>
<tr>
<td>normal_interval</td>
<td>35</td>
</tr>
<tr>
<td>normal_quantile</td>
<td>35</td>
</tr>
<tr>
<td>normal_update</td>
<td>36</td>
</tr>
<tr>
<td>olympic_butterfly</td>
<td>37</td>
</tr>
<tr>
<td>prior_post_plot</td>
<td>37</td>
</tr>
<tr>
<td>prob_plot</td>
<td>38</td>
</tr>
<tr>
<td>ProfessorSalary</td>
<td>39</td>
</tr>
<tr>
<td>pt100price</td>
<td>39</td>
</tr>
</tbody>
</table>
### animation_ratings

Ratings for a set of 2010 animation movies

#### Usage

```r
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
arm_height

**Description**

Arm span and height measurements for a sample of students

**Usage**

```r
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

**Description**

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

**Usage**

```r
bar_plot(y, ...)
```

**Arguments**

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

**Value**

A ggplot2 object containing the bar graph.

**Author(s)**

Jim Albert
**batting_2018**

**Description**

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

**Usage**

```
battin_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
```r
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
```r
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**

```r
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)`
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

A `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

Value

ggplot2 object containing the graphical display.

Author(s)

Jim Albert

Examples

```
prior_shapes <- c(4, 6)
post_shapes <- c(19, 16)
beta_prior_post(prior_shapes, post_shapes)
```

---

**beta_quantile**

*Displays a Quantile of a Beta Curve*

Description

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

Value

ggplot2 object containing the graphical display.

Author(s)

Jim Albert

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**

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

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

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()

CEsample

Expeditures of U.S. Households

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
Value

ggplot2 object containing the graphical display.

Author(s)

Jim Albert

Examples

```r
prob_matrix <- testing_prior()
draw_two_p(prob_matrix, title="Testing Prior")
```

---

**Hypergeometric sampling density**

Description

Hypergeometric sampling density

Usage

```r
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

```r
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**  
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
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

---

**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)
```
gibbs_normal

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

```r
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

```r
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

---

**house_prices**  
*House price data*

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.

Increases font size of text

**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
JAGS_script

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

`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 prodcast during different days of the week and different producers

Usage

`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

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**

```r
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**

```r
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**

```r
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.
**mcdonalds**

*Nutritional data for McDonalds Sandwiches*

**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

**metropolis**

*Metropolis sampling of a continuous distribution*

**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
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.

- **Player**  name of player
- **Age**  age of player
- **FG**  field goals
- **FGA**  field goal attempts
- **FT**  free throws
- **FTA**  free throw attempts

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

**Arguments**
- **lo**  lower bound of interval
- **hi**  upper bound of interval
- **normal_pars**  vector of mean and standard deviation of the normal curve
- **Color**  color of shading in plot
normal_draw

Value

ggplot2 object containing the graphical display.

Author(s)

Jim Albert

Examples

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

Description

Draws a Normal Curve

Usage

```r
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

```r
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**

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>prob</code></td>
<td>value of coverage probability</td>
</tr>
<tr>
<td><code>normal_pars</code></td>
<td>vector of mean and standard deviation of the normal curve</td>
</tr>
<tr>
<td><code>Color</code></td>
<td>color of shading in plot</td>
</tr>
</tbody>
</table>

**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**

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>prob</code></td>
<td>probability value of interest</td>
</tr>
<tr>
<td><code>normal_pars</code></td>
<td>vector of mean and standard deviation of the normal curve</td>
</tr>
<tr>
<td><code>Color</code></td>
<td>color of shading in plot</td>
</tr>
</tbody>
</table>
normal_update

**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

Author(s)
Jim Albert

Examples

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

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.
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.

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)
**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.

<table>
<thead>
<tr>
<th>sleeping_times</th>
<th>Sleeping Times</th>
</tr>
</thead>
</table>

**Description**

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

**Usage**

```r
sleeping_times
```

**Format**

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

**hours**  number of hours of sleep

**Source**

Personal collection

<table>
<thead>
<tr>
<th>spinner_bayes</th>
<th>Implements Bayes’ rule for a spinner problem</th>
</tr>
</thead>
</table>

**Description**

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

**Usage**

```r
spinner_bayes(list_regions, prior, data, plot=TRUE)
```
spinner_data

Arguments

- `list_regions`: list of vectors of integer areas for the spins 1, 2, ...
- `prior`: a vector containing the prior probabilities for the spinners
- `data`: a vector containing the spin values where 1, 2, 3, ... are the possible spins
- `plot`: if `plot=TRUE`, a comparative graph of the prior and posterior probabilities is displayed

Value

A data frame with variables `Spinner`, `Prior`, `Likelihood`, `Product`, and `Posterior`

Author(s)

Jim Albert

Examples

```r
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)
```

spinner_data

Simulate random data from a spinner

Description

Simulate random data from a spinner

Usage

`spinner_data(regions, nsim=1000)`

Arguments

- `regions`: vector of integer values for the spins 1, 2, ...
- `nsim`: number of spins

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, ...)
```
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  
*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  
*Tennis Times to 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

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>lo</td>
<td>minimum value of each proportion</td>
</tr>
<tr>
<td>hi</td>
<td>maximum value of each proportion</td>
</tr>
<tr>
<td>n_values</td>
<td>number of values of each proportion</td>
</tr>
<tr>
<td>pequal</td>
<td>probability of the equality of the two proportions</td>
</tr>
<tr>
<td>uniform</td>
<td>indicates if a uniform prior is desired</td>
</tr>
</tbody>
</table>

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)
```
**two_players_time_to_serve**

**Trout20  
*Mike Trout Statcast Data***

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

**Usage**

```r
trot20
```

**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**

```r
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
two_p_summarize  

**Summaries of a probability matrix**

**Description**
Computes posterior of difference $P_2 - P_1$ 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**
```
# use uniform prior over values .2, .3, .4
prob_matrix <- testing_prior(.2, .4, 3, uniform=TRUE)
two_p_summarize(prob_matrix)
```

two_p_update  

**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**

`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

* datasets
  animation_ratings, 3
  arm_height, 4
  batting_2018, 5
  BBS_survey, 6
  book_stats, 11
  buffalo_jan, 11
  career_1978, 12
  CEsample, 13
  ComputerPriceSample, 14
  Cowles, 15
  DeathHeartAttackDataNYCfull, 15
  DeathHeartAttackManhattan, 16
  electricbills, 18
  federalist_word_study, 19
  federer_time_to_serve, 20
  fire_calls, 20
  football_field_goals, 21
  gas2017, 21
  GradSchoolAdmission, 24
  Hamilton_can, 24
  house_prices, 25
  HWhours5schools, 26
  JAGS_script, 27
  KDramaData, 27
  LaborParticipation, 28
  Madison_can, 28
  marriage_counts, 30
  mcdonalds, 31
  movies2017, 32
  nba_guards, 33
  olympic_butterfly, 37
  ProfessorSalary, 39
  pt100price, 39
  pt99price, 40
  pythag2018, 40
  ScoreData, 41
  sleeping_times, 42
  taxi_fares, 46
  tennis_serve, 46
  trout20, 48
  two_players_time_to_serve, 48
  web_visits, 50

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  
KDramaData, 27  
LaborParticipation, 28  
Madison_can, 28  
many_normal_plots, 29  
many_spinner_plots, 29  
marrriage_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  
pt99price, 40  
pythag2018, 40  
random_walk, 41  
ScoreData, 41  
sleeping_times, 42  
spinner_bayes, 42  
spinner_data, 43  
spinner_likelihoods, 44  
spinner_plot, 44  
spinner_probs, 45  
taxi_fares, 46  
tennis_server, 46  
testing_prior, 47  
trout20, 48  
two_p_summarize, 49  
two_p_update, 49  
two_players_time_to_serve, 48  
web_visits, 50