

# Package ‘BayesGOF’

October 12, 2022

**Type** Package

**Title** Bayesian Modeling via Frequentist Goodness-of-Fit

**Version** 5.2

**Date** 2018-10-09

**Author** Subhadeep Mukhopadhyay, Douglas Fletcher

**Maintainer** Doug Fletcher <tug25070@temple.edu>

**Description** A Bayesian data modeling scheme that performs four interconnected tasks: (i) characterizes the uncertainty of the elicited parametric prior; (ii) provides exploratory diagnostic for checking prior-data conflict; (iii) computes the final statistical prior density estimate; and (iv) executes macro- and micro-inference. Primary reference is Mukhopadhyay, S. and Fletcher, D. 2018 paper “Generalized Empirical Bayes via Frequentist Goodness of Fit” (<<https://www.nature.com/articles/s41598-018-28130-5>>).

**Depends** orthopolynom, VGAM, Bolstad2, nleqslv

**Suggests** knitr, rmarkdown

**VignetteBuilder** knitr

**License** GPL-2

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2018-10-09 21:50:09 UTC

## R topics documented:

BayesGOF-package . . . . .	2
arsenic . . . . .	2
AutoIns . . . . .	3
ChildIll . . . . .	3
CorbBfly . . . . .	4
DS.entropy . . . . .	5
DS.Finite.Bayes . . . . .	6
DS.macro.inf . . . . .	7
DS.micro.inf . . . . .	8

DS.posterior.reduce . . . . .	9
DS.prior . . . . .	10
DS.sampler . . . . .	12
galaxy . . . . .	14
gLP.basis . . . . .	14
gMLE.bb . . . . .	15
gMLE.nn . . . . .	16
gMLE.pg . . . . .	17
NorbergIns . . . . .	18
rat . . . . .	19
ship . . . . .	20
steroid . . . . .	20
surg . . . . .	21
tacks . . . . .	21
terb . . . . .	22
ulcer . . . . .	22

<b>Index</b>	<b>24</b>
--------------	-----------

---

BayesGOF-package	<i>Bayesian Modeling via Frequentist Goodness-of-Fit</i>
------------------	--

---

### Description

A Bayesian data modeling scheme that performs four interconnected tasks: (i) characterizes the uncertainty of the elicited parametric prior; (ii) provides exploratory diagnostic for checking prior-data conflict; (iii) computes the final statistical prior density estimate; and (iv) executes macro- and micro-inference.

### References

Mukhopadhyay, S. and Fletcher, D., 2018. "Generalized Empirical Bayes via Frequentist Goodness of Fit," *Nature Scientific Reports*, 8(1), p.9983, <https://www.nature.com/articles/s41598-018-28130-5>.

---

arsenic	<i>Arsenic levels in oyster tissue</i>
---------	--

---

### Description

Results from an inter-laboratory study involving  $k = 28$  measurements for the level of arsenic in oyster tissue.  $y$  is the mean level of arsenic from a lab and  $se$  is the standard error of the measurement.

### Usage

```
data("arsenic")
```

**Format**

A data frame of  $(y_i, se_i)$  for  $i = 1, \dots, 28$ .

$y$  mean level of arsenic in the tissue measured by the  $i^{th}$  lab

$se$  the standard error of the measurement by  $i^{th}$  lab

**Source**

Wille, S. and Berman, S., 1995. "Ninth round intercomparison for trace metals in marine sediments and biological tissues," *NRC/NOAA*.

---

AutoIns	<i>Number of claims on an insurance policy</i>
---------	--

---

**Description**

The number of claims on an automobile insurance policy made by  $k = 9461$  individuals during a single year.

**Usage**

```
data("AutoIns")
```

**Format**

A vector of length 9461.

value number of auto insurance claims by the  $i^{th}$  person

**Source**

Efron, B. and Hastie, T., 2016. *Computer Age Statistical Inference* (Vol. 5). Cambridge University Press.

---

ChildIll	<i>Frequency of child illness</i>
----------	-----------------------------------

---

**Description**

Results of a study that followed  $k = 602$  pre-school children in north-east Thailand from June 1982 through September 1985. Researchers recorded the number of times a child became ill during every 2-week period.

**Usage**

```
data("ChildIll")
```

**Format**

A vector of length  $k = 602$ .

value number of times the  $i^{\text{th}}$  child became ill during the study

**Source**

Bohning, D., 2000. *Computer-assisted Analysis of Mixtures and Applications: Meta-analysis, Disease Mapping, and Others* (Vol. 81). CRC press.

---

CorbBfly

*Corbet's Butterfly data*

---

**Description**

The number of times Alexander Corbet captured a species of butterfly during a two-year period in Malaysia.

**Usage**

```
data("CorbBfly")
```

**Format**

A vector of length  $k = 501$ .

value number of times Corbet captured the  $i^{\text{th}}$  species

**Source**

Fisher, R.A., Corbet, A.S. and Williams, C.B., 1943. "The relation between the number of species and the number of individuals in a random sample of an animal population." *The Journal of Animal Ecology*, pp.42-58.

**References**

Efron, B. and Hastie, T., 2016. *Computer Age Statistical Inference* (Vol. 5). Cambridge University Press.

---

DS.entropy	<i>Full and Excess Entropy of DS(G,m) prior</i>
------------	---

---

**Description**

A function that calculates the full entropy of a DS(G,m) prior. For DS(G,m) with  $m > 0$ , also returns the excess entropy  $qLP$ .

**Usage**

```
DS.entropy(DS.GF.obj)
```

**Arguments**

DS.GF.obj      Object resulting from running DS.prior function on a data set.

**Value**

ent              The total entropy of the DS(G,m) prior where  $m \geq 0$ .  
qLP              The excess entropy when  $m > 0$ .

**Author(s)**

Doug Fletcher

**References**

Mukhopadhyay, S. and Fletcher, D., 2018. "Generalized Empirical Bayes via Frequentist Goodness of Fit," *Nature Scientific Reports*, 8(1), p.9983, <https://www.nature.com/articles/s41598-018-28130-5>.

**Examples**

```
data(rat)
rat.start <- gMLE.bb(rat$y, rat$n)$estimate
rat.ds <- DS.prior(rat, max.m = 4, rat.start, family = "Binomial")
DS.entropy(rat.ds)
```

---

DS.Finite.Bayes      *Conduct Finite Bayes Inference on a DS object*

---

### Description

A function that generates the finite Bayes prior and posterior distribution, along with the Bayesian credible interval for the posterior mean.

### Usage

```
DS.Finite.Bayes(DS.GF.obj, y.θ, n.θ = NULL,
                cred.interval = 0.9, iters = 25)
```

### Arguments

DS.GF.obj	Object from DS.prior.
y.θ	For Binomial family, number of success $y_i$ for new study. In the Poisson family, it is the number of counts. Represents the study mean for the Normal family.
n.θ	For the Binomial family, the total number of trials for the new study. In the Normal family, n.θ is the standard error of y.θ. Not used for the Poisson family.
cred.interval	The desired probability for the credible interval of the posterior mean; the default is 0.90 (90%).
iters	Integer value of total number of iterations.

### Value

prior.fit	Fitted values for the estimated parametric, DS, and finite Bayes prior distributions.
post.fit	Dataframe with $\theta$ , $\pi_G(\theta y_0)$ , and $\pi_{LP}(\theta y_0)$ .
interval	The $100 \times \text{cred.interval}\%$ Bayesian credible interval for the posterior mean.
post.vec	Vector containing the PEB posterior mean (PEB.mean), DS posterior mean (DS.mean), PEB posterior mode (PEB.mode), and the DS posterior mode (DS.mode).

### Author(s)

Doug Fletcher, Subhadeep Mukhopadhyay

### References

Mukhopadhyay, S. and Fletcher, D., 2018. "Generalized Empirical Bayes via Frequentist Goodness of Fit," *Nature Scientific Reports*, 8(1), p.9983, <https://www.nature.com/articles/s41598-018-28130-5>.

Efron, B., 2018. "Bayes, Oracle Bayes, and Empirical Bayes," Technical Report.

**Examples**

```
## Not run:
### Finite Bayes: Rat with theta_71 (y_71 = 4, n_71 = 14)
data(rat)
rat.start <- gMLE.bb(rat$y, rat$n)$estimate
rat.ds <- DS.prior(rat, max.m = 4, rat.start.family = "Binomial")
rat.FB <- DS.FiniteBayes(rat.ds, y.0 = 4, n.0 = 14)
plot(rat.FB)

## End(Not run)
```

DS.macro.inf

*Execute MacroInference (mean or mode) on a DS object***Description**

A function that generates macro-estimates with their uncertainty (standard error).

**Usage**

```
DS.macro.inf(DS.GF.obj, num.modes = 1,
             method = c("mean", "mode"),
             iters = 25, exposure = NULL)
```

**Arguments**

DS.GF.obj	Object from DS.prior.
num.modes	The number of modes indicated by DS.prior object.
method	Returns mean or mode(s) (based on user choice) along with the associated standard error(s).
iters	Integer value of total number of iterations.
exposure	In the case where DS.GF.obj is from a Poisson family with exposure, exposure is the vector of exposures. Otherwise, the default is NULL.

**Value**

DS.GF.macro.obj	Object of class DS.GF.macro associated with either mean or mode.
model.modes	For method = "mode", returns mode(s) of estimated DS prior.
mode.sd	For method = "mode", provides the bootstrapped standard error for each mode.
boot.modes	For method = "mode", returns all generated mode(s).
model.mean	For method = "mean", returns mean of estimated DS prior.
mean.sd	For method = "mean", provides the bootstrapped standard error for the mean.
boot.mean	For method = "mean", returns all generated means.
prior.fit	Fitted values of estimated prior imported from the DS.prior object.

**Author(s)**

Doug Fletcher, Subhadeep Mukhopadhyay

**References**

Mukhopadhyay, S. and Fletcher, D., 2018. "Generalized Empirical Bayes via Frequentist Goodness of Fit," *Nature Scientific Reports*, 8(1), p.9983, <https://www.nature.com/articles/s41598-018-28130-5>.

**Examples**

```
## Not run:
### MacroInference: Mode
data(rat)
rat.start <- gMLE.bb(rat$y, rat$n)$estimate
rat.ds <- DS.prior(rat, max.m = 4, rat.start, family = "Binomial")
rat.ds.macro <- DS.macro.inf(rat.ds, num.modes = 2, method = "mode", iters = 5)
rat.ds.macro
plot(rat.ds.macro)
### MacroInference: Mean
data(ulcer)
ulcer.start <- gMLE.nn(ulcer$y, ulcer$se)$estimate
ulcer.ds <- DS.prior(ulcer, max.m = 4, ulcer.start)
ulcer.ds.macro <- DS.macro.inf(ulcer.ds, num.modes = 1, method = "mean", iters = 5)
ulcer.ds.macro
plot(ulcer.ds.macro)
## End(Not run)
```

---

DS.micro.inf

*MicroInference for DS Prior Objects*

---

**Description**

Provides DS nonparametric adaptive Bayes and parametric estimate for a specific observation  $y_0$ .

**Usage**

```
DS.micro.inf(DS.GF.obj, y.0, n.0, e.0 = NULL)
```

**Arguments**

DS.GF.obj	Object resulting from running DS.prior function on a data set.
y.0	For Binomial family, number of success $y_i$ for new study. In the Poisson family, it is the number of counts. Represents the study mean for the Normal family.
n.0	For the Binomial family, the total number of trials for the new study. In the Normal family, n.0 is the standard error of y.0. Not used for the Poisson family.
e.0	In the case of the Poisson family with exposure, represents the exposure value for a given count value y.0.



**Details**

Returns an object of class DS.GF.micro that can be used in conjunction with plot command to display the DS posterior distribution for the new study.

**Value**

DS.mean	Posterior mean for $\pi_{LP}(\theta y_0)$ .
DS.mode	Posterior mode for $\pi_{LP}(\theta y_0)$ .
PEB.mean	Posterior mean for $\pi_G(\theta y_0)$ .
PEB.mode	Posterior mode for $\pi_G(\theta y_0)$ .
post.vec	Vector containing PEB.mean, DS.mean, PEB.mode, and DS.mode.
study	User-provided $y_0$ and $n_0$ .
post.fit	Dataframe with $\theta$ , $\pi_G(\theta y_0)$ , and $\pi_{LP}(\theta y_0)$ .

**Author(s)**

Doug Fletcher, Subhadeep Mukhopadhyay

**References**

Mukhopadhyay, S. and Fletcher, D., 2018. "Generalized Empirical Bayes via Frequentist Goodness of Fit," *Nature Scientific Reports*, 8(1), p.9983, <https://www.nature.com/articles/s41598-018-28130-5>.

**Examples**

```
### MicroInference for Naval Shipyard Data: sample where y = 0 and n = 5
data(ship)
ship.ds <- DS.prior(ship, max.m = 2, c(.5,.5), family = "Binomial")
ship.ds.micro <- DS.micro.inf(ship.ds, y.0 = 0, n.0 = 5)
ship.ds.micro
plot(ship.ds.micro)
```

---

DS.posterior.reduce     *Posterior Expectation and Modes of DS object*

---

**Description**

A function that determines the posterior expectations  $E(\theta_0|y_0)$  and posterior modes for a set of observed data.

**Usage**

```
DS.posterior.reduce(DS.GF.obj, exposure)
```

**Arguments**

DS.GF.obj	Object resulting from running DS.prior function on a data set.
exposure	In the case of the Poisson family with exposure, represents the exposure values for the count data.

**Value**

Returns  $k \times 4$  matrix with the columns indicating PEB mean, DS mean, PEB mode, and DS modes for  $k$  observations in the data set.

**Author(s)**

Doug Fletcher

**References**

Mukhopadhyay, S. and Fletcher, D., 2018. "Generalized Empirical Bayes via Frequentist Goodness of Fit," *Nature Scientific Reports*, 8(1), p.9983, <https://www.nature.com/articles/s41598-018-28130-5>.

**Examples**

```
data(rat)
rat.start <- gMLE.bb(rat$y, rat$n)$estimate
rat.ds <- DS.prior(rat, max.m = 4, rat.start, family = "Binomial")
DS.posterior.reduce(rat.ds)
```

---

DS.prior

*Prior Diagnostics and Estimation*

---

**Description**

A function that generates the uncertainty diagnostic function (U-function) and estimates  $DS(G, m)$  prior model.

**Usage**

```
DS.prior(input, max.m = 8, g.par,
         family = c("Normal", "Binomial", "Poisson"),
         LP.type = c("L2", "MaxEnt"),
         smooth.crit = "BIC", iters = 200, B = 1000,
         max.theta = NULL)
```

**Arguments**

input	For "Binomial", a dataframe that contains the $k$ pairs of successes $y$ and the corresponding total number of trials $n$ . For "Normal", a dataframe that has the $k$ means $y_i$ in the first column and their respective standard errors $s_i$ in the second. For the "Poisson", a vector of that includes the untabled count data.
max.m	The truncation point $m$ reflects the concentration of true unknown $\pi$ around known $g$ .
g.par	Vector with estimated parameters for specified conjugate prior distribution $g$ (i.e beta prior: $\alpha$ and $\beta$ ; normal prior: $\mu$ and $\tau^2$ ; gamma prior: $\alpha$ and $\beta$ ).
family	The distribution of $y_i$ . Currently accommodates three families: Normal, Binomial, and Poisson.
LP.type	User selects either "L2" for LP-orthogonal series representation of U-function or "MaxEnt" for the maximum entropy representation. Default is L2.
smooth.crit	User selects either "BIC" or "AIC" as criteria to both determine optimal $m$ and smooth final LP parameters; default is "BIC".
iters	Integer value that gives the maximum number of iterations allowed for convergence; default is 200.
B	Integer value for number of grid points used for distribution output; default is 1000.
max.theta	For "Poisson", user can provide a maximum theta value for prior; default is the maximum count value in input.

**Details**

Function can take  $m = 0$  and will return the Bayes estimate with given starting parameters. Returns an object of class DS.GF.obj; this object can be used with plot command to plot the U-function (Ufunc), Deviance Plots (mDev), and DS-G comparison (DS\_G).

**Value**

LP.par	$m$ smoothed LP-Fourier coefficients, where $m$ is determined by maximum deviance.
g.par	Parameters for $g$ .
LP.max.uns	Vector of all LP-Fourier coefficients prior to smoothing, where the length is the same as max.m.
LP.max.smt	Vector of all smoothed LP-Fourier coefficients, where the length is the same as max.m.
prior.fit	Fitted values for the estimated prior.
UF.data	Dataframe that contains values required for plotting the U-function.
dev.df	Dataframe that contains deviance values for values of $m$ up to max.m.
m.val	The value of $m$ (less than or equal to the maximum $m$ from user) that has the maximum deviance and represents the appropriate number of LP-Fourier coefficients.
sm.crit	Smoothing criteria; either "BIC" or "AIC".

fam	The user-selected family.
LP.type	User-selected representation of U-function.
obs.data	Observed data provided by user for input.

**Author(s)**

Doug Fletcher, Subhadeep Mukhopadhyay

**References**

- Mukhopadhyay, S. and Fletcher, D., 2018. "Generalized Empirical Bayes via Frequentist Goodness of Fit," *Nature Scientific Reports*, 8(1), p.9983, <https://www.nature.com/articles/s41598-018-28130-5>.
- Mukhopadhyay, S., 2017. "Large-Scale Mode Identification and Data-Driven Sciences," *Electronic Journal of Statistics*, 11(1), pp.215-240.

**Examples**

```
data(rat)
rat.start <- gMLE.bb(rat$y, rat$n)$estimate
rat.ds <- DS.prior(rat, max.m = 4, rat.start, family = "Binomial")
rat.ds
plot(rat.ds, plot.type = "Ufunc")
plot(rat.ds, plot.type = "DSg")
plot(rat.ds, plot.type = "mDev")
```

---

DS.sampler

*Samples data from DS(G,m) distribution.*


---

**Description**

Generates samples of size  $k$  from  $DS(G, m)$  prior distribution.

**Usage**

```
DS.sampler(k, g.par, LP.par, con.prior, LP.type, B)
```

```
DS.sampler.post(k, g.par, LP.par, y.0, n.0,
               con.prior, LP.type, B)
```

**Arguments**

k	Total number of samples requested.
g.par	Estimated parameters for specified conjugate prior distribution (i.e beta prior: $\alpha$ and $\beta$ ; normal prior: $\mu$ and $\tau^2$ ; gamma prior: $\alpha$ and $\beta$ ).
LP.par	LP coefficients for DS prior.
con.prior	The distribution type of conjugate prior $g$ ; either "Beta", "Normal", or "Gamma".

LP.type	The type of LP means, either "L2" or "MaxEnt".
y.0	Depending on $g$ , $y_0$ is either (i) the sample mean ("Normal"), (ii) the number of successes ("Beta"), or (iii) the specific count value ("Gamma") for desired posterior distribution(DS.sampler.post only).
n.0	Depending on $g$ , $n_0$ is either (i) the sample standard error ("Normal"), or (ii) the total number of trials in the sample ("Beta"). Not used for "Gamma". (DS.sampler.post only).
B	The number of grid points, default is 250.

### Details

DS.sampler.post uses the same type of sampling as DS.sampler to generate random values from a DS posterior distribution.

### Value

Vector of length  $k$  containing sampled values from DS prior or DS posterior.

### Author(s)

Doug Fletcher, Subhadeep Mukhopadhyay

### References

- Mukhopadhyay, S. and Fletcher, D., 2018. "Generalized Empirical Bayes via Frequentist Goodness of Fit," *Nature Scientific Reports*, 8(1), p.9983, <https://www.nature.com/articles/s41598-018-28130-5>.
- Mukhopadhyay, S., 2017. "Large-Scale Mode Identification and Data-Driven Sciences," *Electronic Journal of Statistics*, 11(1), pp.215-240.

### Examples

```
##Extracted parameters from rat.ds object
rat.g.par <- c(2.3, 14.1)
rat.LP.par <- c(0, 0, -0.5)
samps.prior <- DS.sampler(25, rat.g.par, rat.LP.par, con.prior = "Beta")
hist(samps.prior,15)
##Posterior for rat data
samps.post <- DS.sampler.post(25, rat.g.par, rat.LP.par,
y.0 = 4, n.0 = 14, con.prior = "Beta")
hist(samps.post, 15)
```

galaxy

*Galaxy Data***Description**

The observed rotation velocities and their uncertainties of Low Surface Brightness (LSB) galaxies, along with the physical radius of the galaxy.

**Usage**

```
data("galaxy")
```

**Format**

A data frame of  $(y_i, se_i, X_i)$  for  $i = 1, \dots, 318$ .

y actual observed (smoothed) velocity

se uncertainty of observed velocity

X physical radius of the galaxy

**Source**

De Blok, W.J.G., McGaugh, S.S., and Rubin, V. C., 2001. "High-resolution rotation curves of low surface brightness galaxies. II. Mass models," *The Astronomical Journal*, 122(5), p. 2396.

gLP.basis

*Determine LP basis functions for prior distribution g***Description**

Determines the LP basis for a given parametric prior distribution.

**Usage**

```
gLP.basis(x, g.par, m, con.prior, ind)
```

**Arguments**

x	x values (integer or vector) from 0 to 1.
g.par	Estimated parameters for specified prior distribution (i.e beta prior: $\alpha$ and $\beta$ ; normal prior: $\mu$ and $\tau^2$ ; gamma prior: $\alpha$ and $\beta$ ).
m	Number of LP-Polynomial basis.
con.prior	Specified conjugate prior distribution for basis functions. Options are "Beta", "Normal", and "Gamma".
ind	Default is NULL which returns matrix with $m$ columns that consists of LP-basis functions; user can provide a specific choice through ind.

**Value**

Matrix with  $m$  columns of values for the LP-Basis functions evaluated at  $x$ -values.

**Author(s)**

Subhadeep Mukhopadhyay, Doug Fletcher

**References**

- Mukhopadhyay, S. and Fletcher, D., 2018. "Generalized Empirical Bayes via Frequentist Goodness of Fit," *Nature Scientific Reports*, 8(1), p.9983, <https://www.nature.com/articles/s41598-018-28130-5>.
- Mukhopadhyay, S., 2017. "Large-Scale Mode Identification and Data-Driven Sciences," *Electronic Journal of Statistics*, 11(1), pp.215-240.
- Mukhopadhyay, S. and Parzen, E., 2014. "LP Approach to Statistical Modeling," arXiv: 1405.2601.

---

gMLE.bb

*Beta-Binomial Parameter Estimation*


---

**Description**

Computes type-II Maximum likelihood estimates  $\hat{\alpha}$  and  $\hat{\beta}$  for Beta prior  $g \sim \text{Beta}(\alpha, \beta)$ .

**Usage**

```
gMLE.bb(success, trials, start = NULL, optim.method = "default",
        lower = 0, upper = Inf)
```

**Arguments**

success	Vector containing the number of successes.
trials	Vector containing the total number of trials that correspond to the successes.
start	initial parameters; default is NULL which allows function to determine MoM estimates as initial parameters.
optim.method	optimization method in <code>optim()</code> stats.
lower	lower bound for parameters; default is 0.
upper	upper bound for parameters; default is infinity.

**Value**

estimate	MLE estimate for beta parameters.
convergence	Convergence code from <code>optim()</code> ; 0 means convergence.
loglik	Loglikelihood that corresponds with MLE estimated parameters.
initial	Initial parameters, either user-defined or determined from method of moments.
hessian	Estimated Hessian matrix at the given solution.

**Author(s)**

Aleksandar Bradic

**References**

<https://github.com/SupplyFrame/EmpiricalBayesR/blob/master/EmpiricalBayesEstimation.R>

**Examples**

```
data(rat)
### MLE estimate of alpha and beta
rat.mle <- gMLE.bb(rat$y, rat$N)$estimate
rat.mle
### MoM estimate of alpha and beta
rat.mom <- gMLE.bb(rat$y, rat$N)$initial
rat.mom
```

gMLE.nn

*Normal-Normal Parameter Estimation***Description**

Computes type-II Maximum likelihood estimates  $\hat{\mu}$  and  $\hat{\tau}^2$  for Normal prior  $g \sim \text{Normal}(\mu, \tau^2)$ .

**Usage**

```
gMLE.nn(value, se, fixed = FALSE, method = c("DL", "SJ", "REML", "MoM"))
```

**Arguments**

value	Vector of values.
se	Standard error for each value.
fixed	When FALSE, treats the input as if from a random effects model; otherwise, will treat it as if it a fixed effect.
method	Determines the method to find $\tau^2$ : "DL" uses Dersimonian and Lard technique, "SJ" uses Sidik-Jonkman, "REML" uses restricted maximum likelihood, and "MoM" uses a method of moments technique.

**Value**

estimate	Vector with both estimated $\hat{\mu}$ and $\hat{\tau}^2$ .
mu.hat	Estimated $\hat{\mu}$ .
tau.sq	Estimated $\hat{\tau}^2$ .
method	User-selected method.



**Author(s)**

Doug Fletcher

**References**

Marin-Martinez, F. and Sanchez-Meca, J., 2010. "Weighting by inverse variance or by sample size in random-effects meta-analysis," *Educational and Psychological Measurement*, 70(1), pp. 56-73.

Brown, L.D., 2008. "In-season prediction of batting averages: A field test of empirical Bayes and Bayes methodologies," *The Annals of Applied Statistics*, pp. 113-152.

Sidik, K. and Jonkman, J.N., 2005. "Simple heterogeneity variance estimation for meta-analysis," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(2), pp. 367-384.

**Examples**

```
data(ulcer)
### MLE estimate of alpha and beta
ulcer.mle <- gMLE.nn(ulcer$y, ulcer$se, method = "DL")$estimate
ulcer.mle
ulcer.reml <- gMLE.nn(ulcer$y, ulcer$se, method = "REML")$estimate
ulcer.reml
```

gMLE.pg

*Negative-Binomial Parameter Estimation***Description**

Computes Type-II Maximum likelihood estimates  $\hat{\alpha}$  and  $\hat{\beta}$  for gamma prior  $g \sim \text{Gamma}(\alpha, \beta)$ .

**Usage**

```
gMLE.pg(cnt.vec, exposure = NULL, start.par = c(1,1))
```

**Arguments**

cnt.vec	Vector containing Poisson counts.
exposure	Vector containing exposures for each count. The default is no exposure, thus exposure = NULL.
start.par	Initial values that will pass to optim.

**Value**

Returns a vector where the first component is  $\alpha$  and the second component is the scale parameter  $\beta$  for the gamma distribution:  $\frac{1}{\Gamma(\alpha)\beta^\alpha} \theta^{\alpha-1} e^{-\frac{\theta}{\beta}}$ .

**Author(s)**

Doug Fletcher

**References**

Koenker, R. and Gu, J., 2017. "REBayes: An R Package for Empirical Bayes Mixture Methods," *Journal of Statistical Software, Articles*, 82(8), pp. 1-26.

**Examples**

```
### without exposure
data(ChildIll)
ill.start <- gMLE.pg(ChildIll)
ill.start
### with exposure
data(NorbergIns)
X <- NorbergIns$deaths
E <- NorbergIns$exposure/344
norb.start <- gMLE.pg(X, exposure = E)
norb.start
```

---

NorbergIns

*Norberg life insurance data*

---

**Description**

The number of claims  $y_i$  on a life insurance policy for each of  $k = 72$  Norwegian occupational categories and the total number of years the workers in each category were exposed to risk ( $E_i$ ).

**Usage**

```
data("NorbergIns")
```

**Format**

A data frame of the occupational group number (group), the number of deaths (deaths), and the years of exposure (exposure) for  $i = 1, \dots, 72$ .

group Occupational group number

deaths The number of deaths in the occupational group resulting in a claim on a life insurance policy.

exposure The total number of years of exposure to risk for those who passed.

**Source**

Norberg, R., 1989. "Experience rating in group life insurance," *Scandinavian Actuarial Journal*, 1989(4), pp. 194-224.

## References

Koenker, R. and Gu, J., 2017. "REBayes: An R Package for Empirical Bayes Mixture Methods," *Journal of Statistical Software, Articles*, 82(8), pp. 1-26.

---

rat

*Rat Tumor Data*

---

## Description

Incidence of endometrial stromal polyps in  $k = 70$  studys of female rats in control group of a 1977 study on the carcinogenic effects of a diabetic drug phenformin. For each of the  $k$  groups,  $y$  represents the number of rats who developed the tumors out of  $n$  total rats in the group.

## Usage

```
data("rat")
```

## Format

A data frame of  $(y_i, n_i)$  for  $i = 1, \dots, 70$ .

$y$  number of female rats in the  $i^{th}$  study who developed polyps/tumors

$n$  total number of rats in the  $i^{th}$  study

## Source

National Cancer Institute (1977), "Bioassay of phenformin for possible carcinogenicity," *Technical Report No. 7*.

## References

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., and Rubin, D.B., 2014. *Bayesian Data Analysis* (Vol. 3). Boca Raton, FL: CRC press.

Tarone, R.E., 1982. "The use of historical control information in testing for a trend in proportions," *Biometrics*, pp. 215-220.

---

ship *Portsmouth Navy Shipyard Data*

---

### Description

Data represents results of quality-control inspections executed by Portsmouth Naval Shipyard on lots of welding materials. The data has  $k = 5$  observations of number of defects  $y$  out of the total number of tested  $n = 5$ .

### Usage

```
data("ship")
```

### Format

A data frame of  $(y_i, n_i)$  for  $i = 1, \dots, 5$ .

$y$  number of defects found in the  $i^{th}$  inspection

$n$  total samples tested in the  $i^{th}$  inspection

### Source

Martz, H.F. and Lian, M.G., 1974. "Empirical Bayes estimation of the binomial parameter," *Biometrika*, 61(3), pp. 517-523.

---

steroid *Nasal Steroid Data*

---

### Description

The standardized mean difference  $y_i$  and standard errors  $se_i$  for seven randomised studies on the use of topical steroids in treatment of chronic rhinosinusitis with nasal polyps.

### Usage

```
data("steroid")
```

### Format

A data frame of  $(y_i, se_i)$  for  $i = 1, \dots, 7$ .

$y$  standard mean difference of clinical trials for topical steroids found in the  $i^{th}$  study

$se$  standard error of the standard mean difference for the  $i^{th}$  study

### Source

IntHout, J., Ioannidis, J. P., Rovers, M. M., & Goeman, J. J., 2016. "Plea for routinely presenting prediction intervals in meta-analysis," *BMJ open*, 6(7), e010247.

---

surg	<i>Intestinal surgery data</i>
------	--------------------------------

---

**Description**

Data involves the number of malignant lymph nodes removed during intestinal surgery for  $k = 844$  cancer patients. For each patient,  $n$  is the total number of satellite nodes removed during surgery from a patient and  $y$  is the number of malignant nodes.

**Usage**

```
data("surg")
```

**Format**

A data frame of  $(y_i, n_i)$  for  $i = 1, \dots, 844$ .

$y$  number of malignant lymph nodes removed from the  $i^{th}$  patient

$n$  total number of lymph nodes removed from the  $i^{th}$  patient

**Source**

Efron, B., 2016. "Empirical Bayes deconvolution estimates," *Biometrika*, 103(1), pp. 1-20.

---

tacks	<i>Rolling Tacks Data</i>
-------	---------------------------

---

**Description**

An experiment that requires a common thumbtack to be "flipped"  $n = 9$  times. Out of these total number of flips,  $y$  is the total number of times that the thumbtack landed point up.

**Usage**

```
data("tacks")
```

**Format**

A data frame of  $(y_i, n_i)$  for  $i = 1, \dots, 320$ .

$y$  number of times a thumbtack landed point up in the  $i^{th}$  trial

$n$  total number of flips for the thumbtack in the  $i^{th}$  trial

**Source**

Beckett, L. and Diaconis, P., 1994. "Spectral analysis for discrete longitudinal data," *Advances in Mathematics*, 103(1), pp. 107-128.

---

terb	<i>Terbinafine trial data</i>
------	-------------------------------

---

**Description**

During several studies of the oral antifungal agent terbinafine, a proportion of the patients in the trial terminated treatment due to some adverse effects. In the data set,  $y_i$  is the number of terminated treatments and  $n_i$  is the total number of patients in the  $i^{th}$  trial.

**Usage**

```
data("terb")
```

**Format**

A data frame of  $(y_i, n_i)$  for  $i = 1, \dots, 41$ .

y number of patients who terminated treatment early in the  $i^{th}$  trial

n total number of patients in the  $i^{th}$  clinical trial

**Source**

Young-Xu, Y. and Chan, K.A., 2008. "Pooling overdispersed binomial data to estimate event rate," *BMC Medical Research Methodology*, 8(1), p. 58.

---

ulcer	<i>Recurrent Bleeding of Ulcers</i>
-------	-------------------------------------

---

**Description**

The data consist of  $k = 40$  randomized trials between 1980 and 1989 of a surgical treatment for stomach ulcers. Each of the trials has an estimated log-odds ratio that measures the rate of occurrence of recurrent bleeding given the surgical treatment.

**Usage**

```
data("ulcer")
```

**Format**

A data frame of  $(y_i, se_i)$  for  $i = 1, \dots, 40$ .

y log-odds of the occurrence of recurrent bleeding in the  $i^{th}$  study

se standard error of the log-odds for the  $i^{th}$  study

**Source**

Sacks, H.S., Chalmers, T.C., Blum, A.L., Berrier, J., and Pagano, D., 1990. "Endoscopic hemostasis: an effective therapy for bleeding peptic ulcers," *Journal of the American Medical Association*, 264(4), pp. 494-499.

**References**

Efron, B., 1996. "Empirical Bayes methods for combining likelihoods," *Journal of the American Statistical Association*, 91(434), pp. 538-550.

# Index

## \* DS Main Functions

- DS.Finite.Bayes, 6
- DS.macro.inf, 7
- DS.micro.inf, 8
- DS.prior, 10
- gLp.basis, 14

## \* datasets

- arsenic, 2
- AutoIns, 3
- ChildIll, 3
- CorbBfly, 4
- galaxy, 14
- NorbergIns, 18
- rat, 19
- ship, 20
- steroid, 20
- surg, 21
- tacks, 21
- terb, 22
- ulcer, 22

## \* package

- BayesGOF-package, 2

- arsenic, 2
- AutoIns, 3

- BayesGOF (BayesGOF-package), 2
- BayesGOF-package, 2

- ChildIll, 3
- CorbBfly, 4

- DS.entropy, 5
- DS.Finite.Bayes, 6
- DS.GF.macro.obj (DS.macro.inf), 7
- DS.GF.micro.obj (DS.micro.inf), 8
- DS.GF.obj (DS.prior), 10
- DS.macro.inf, 7
- DS.micro.inf, 8
- DS.posterior.reduce, 9

- DS.prior, 10
- DS.sampler, 12

- EXP.score (DS.prior), 10

- galaxy, 14
- gLp.basis, 14
- gMLE.bb, 15
- gMLE.nn, 16
- gMLE.pg, 17

- Local.Mode (DS.macro.inf), 7
- LP.smooth (DS.prior), 10

- maxent.LP.par (DS.prior), 10
- maxent.obj.convert (DS.prior), 10
- mom.est (gMLE.bb), 15

- NorbergIns, 18

- plot.DS\_FB\_obj (DS.Finite.Bayes), 6
- plot.DS\_GF (DS.prior), 10
- plot.DS\_GF\_macro\_mean (DS.macro.inf), 7
- plot.DS\_GF\_macro\_mode (DS.macro.inf), 7
- plot.DS\_GF\_micro (DS.micro.inf), 8
- print.DS\_FB\_obj (DS.Finite.Bayes), 6
- print.DS\_GF (DS.prior), 10
- print.DS\_GF\_macro\_mean (DS.macro.inf), 7
- print.DS\_GF\_macro\_mode (DS.macro.inf), 7
- print.DS\_GF\_micro (DS.micro.inf), 8

- rat, 19
- Reduce.LP.coef.univ (DS.prior), 10
- rPPD.ds (DS.sampler), 12

- ship, 20
- steroid, 20
- surg, 21

- tacks, 21
- terb, 22



ulcer, [22](#)

weight.fun.univ (DS.prior), [10](#)