

# Package ‘eggCounts’

January 25, 2017

**Imports** boot,coda,utils,testthat,numbers

**Depends** R (>= 3.2.0), Rcpp (>= 0.12.0), rstan (>= 2.11.1), methods

**Suggests** lattice

**Title** Hierarchical Modelling of Faecal Egg Counts

**Version** 1.3

**Date** 2017-01-27

**Description** An implementation of hierarchical models for faecal egg count data to assess anthelmintic efficacy. Bayesian inference is done via MCMC sampling using Stan.

**SystemRequirements** Stan (<http://mc-stan.org/>)

**License** GPL (>= 3)

**LinkingTo** StanHeaders (>= 2.11.0), rstan (>= 2.11.1), BH (>= 1.60), Rcpp (>= 0.12.0), RcppEigen

**LazyLoad** yes

**NeedsCompilation** yes

**URL** [http://www.math.uzh.ch/as/index.php?id=software\\_as](http://www.math.uzh.ch/as/index.php?id=software_as)

**RcppModules** stan\_fit4paired\_mod, stan\_fit4unpaired\_mod, stan\_fit4zipaired\_mod, stan\_fit4ziunpaired\_mod, stan\_fit4nb\_mod, stan\_fit4zinb\_mod

**Author** Craig Wang [cre, aut],  
Michaela Paul [aut],  
Reinhard Furrer [ctb],  
Trustees of Columbia University [cph] (rstanarm)

**Maintainer** Craig Wang <craig.wang@uzh.ch>

**Repository** CRAN

**Date/Publication** 2017-01-25 10:23:13

**R topics documented:**

|                   |    |
|-------------------|----|
| eggCounts-package | 2  |
| echinococcus      | 2  |
| eggs              | 3  |
| fecrtCI           | 3  |
| fecr_probability  | 4  |
| fecr_stan         | 5  |
| fec_stan          | 7  |
| simData1s         | 9  |
| simData2s         | 10 |
| stan2mcmc         | 12 |
| tab1morgan        | 13 |

|              |           |
|--------------|-----------|
| <b>Index</b> | <b>14</b> |
|--------------|-----------|

---

|                   |  |
|-------------------|--|
| eggCounts-package | <i>Hierarchical modelling of faecal egg counts</i> |
|-------------------|--|

---

**Description**

This package implements Bayesian hierarchical models for the analysis of faecal egg count data. Bayesian inference is done via efficient MCMC sampling using Stan.

**Details**

|           |            |
|-----------|------------|
| Package:  | eggCounts  |
| Type:     | Package    |
| Version:  | 1.3        |
| Date:     | 2017-01-27 |
| License:  | GPL (>= 3) |
| LazyLoad: | yes        |

**Author(s)**

Craig Wang <craig.wang@uzh.ch>  
Michaela Paul

---

|              |                                |
|--------------|--------------------------------|
| echinococcus | <i>Faecal egg count sample</i> |
|--------------|--------------------------------|

---

**Description**

This is an example data set containing 24 eggs per gram (epg) values of *Taenia* parasites (*Echinococcus*) in dogs. The correction factor of the diagnostic technique was 50.

**Usage**

```
data(echinococcus)
```

**Examples**

```
data(echinococcus)
table(echinococcus)
```

---

|      |  |
|------|--|
| eggs | <i>Faecal egg count samples (before and after treatment)</i> |
|------|--|

---

**Description**

This is an example data set containing 24 eggs per gram (epg) values before and after anthelmintic treatment. The correction factor of the diagnostic technique was 10.

**Usage**

```
data(eggs)
```

---

|         |   |
|---------|---|
| fecrtCI | <i>Compute standard FECRT according to WAAVP guidelines</i> |
|---------|---|

---

**Description**

Computes the standard Faecal Egg Count Reduction test together with approximate confidence intervals according to the WAAVP guidelines (Coles et al., 1992, 2006). The function also returns bootstrap percentile confidence intervals.

**Usage**

```
fecrtCI(egp1, epg2, paired = FALSE, alpha = 0.05, R = 1999)
```

**Arguments**

|        |  |
|--------|--|
| egp1   | faecal egg counts in untreated animals   |
| epg2   | faecal egg counts in treated animals   |
| paired | logical. If true, indicates the samples are paired. Otherwise they are unpaired. |
| alpha  | confidence level of the intervals  |
| R      | number of bootstrap replicates   |

**Value**

A list with

|          |  |
|----------|--|
| estimate | the estimated percentage reduction in mean epg rate    |
| bootCI   | corresponding percentile bootstrap confidence interval |
| approxCI | corresponding approximate confidence interval          |

**Author(s)**

Michaela Paul

**References**

Coles GC, Bauer C, Borgsteede FHM, Geerts S, Klei TR, Taylor MA, Waller, PJ (1992). World Association for the Advancement of Veterinary Parasitology (WAAVP) methods for the detection of anthelmintic resistance in nematodes of veterinary importance, *Veterinary Parasitology*, 44:35-44.

Coles GC, Jackson F, Pomroy WE, Prichard RK, von Samson-Himmelstjerna G, Silvestre A, Taylor MA, Vercruyse J (2006). The detection of anthelmintic resistance in nematodes of veterinary importance, *Veterinary Parasitology*, 136:167-185.

**Examples**

```
## Not run:
data(eggs)
fecrCI(eggs$before, eggs$after, paired=TRUE)

## End(Not run)
```

---

|                  |   |
|------------------|---|
| fecr_probability | <i>Compute the probability of the reduction parameter relative to the given threshold</i> |
|------------------|---|

---

**Description**

Computes the probability of the reduction parameter's marginal posterior density is less than (<), greater or equal to (>=) the threshold.

**Usage**

```
fecr_probability(stanFit, threshold = 0.95,
  lessthan = TRUE)
```

**Arguments**

|           |  |
|-----------|--|
| stanFit   | A stanfit object from the output of fecr_stan()  |
| threshold | numeric. The default threshold is 95% or 0.95.   |
| lessthan  | logical. If true, the probability less than the threshold is computed. Otherwise greater or equal to the threshold is computed. Default is TRUE. |

**Value**

Returns the probability in percentage.

**Author(s)**

Craig Wang <craig.wang@uzh.ch>

**Examples**

```
## Not run:
## load the sample data as a vector
data(eggs)

## apply zero-inflation model to the data vector
model<-fecr_stan(eggs[,1],eggs[,2],rawCounts=FALSE,preCF=10,paired=TRUE,zeroInflation=TRUE)
fecr_probability(model$stan.samples)

## End(Not run)
```

---

fecr\_stan

*Modelling the reduction of faecal egg count data*

---

**Description**

Models the reduction in faecal egg counts data with a (un)paired (zero-inflated) Poisson-gamma model formulation using the Stan modelling language. It is computationally several-fold faster compare to conventional MCMC techniques. For the installation instruction of Stan, please read: [Stan Installation](#).

**Usage**

```
fecr_stan(preFEC, postFEC,rawCounts = FALSE, preCF = 50,
  postCF = preCF, paired = TRUE, zeroInflation=TRUE,
  muPrior, kappaPrior, deltaPrior, phiPrior,
  nsamples = 12000, nburnin = 2000, thinning = 1, nchain = 1,
  ncore = 1, adaptdelta = 0.9, verbose = FALSE)
```

**Arguments**

|           |  |
|-----------|--|
| preFEC    | vector of pre-treatment faecal egg counts  |
| postFEC   | vector of post-treatment faecal egg counts   |
| rawCounts | logical. If true, preFEC and postFEC correspond to raw counts (as counted on equipment). Otherwise they correspond to calculated eggs (raw counts times correction factor). Defaults to FALSE. |
| preCF     | correction factor(s) before treatment  |
| postCF    | correction factor(s) after treatment   |

|               |   |
|---------------|---|
| paired        | logical. If true, uses the model for the paired design. Otherwise uses the model for the unpaired design  |
| zeroInflation | logical. If true, uses the model with zero-inflation. Otherwise uses the model without zero-inflation   |
| muPrior       | a list with hyper-prior information for the baseline mean parameter $\mu$ . The default prior is <code>list(priorDist = "gamma", hyperpars=c(1,0.001))</code> , i.e. a gamma distribution with shape 1 and rate 0.001, its 90% probability mass lies between 51 and 2996                |
| kappaPrior    | a list with hyper-prior information for the dispersion parameter $\kappa$ . The default prior is <code>list(priorDist = "gamma", hyperpars=c(1,0.7))</code> , i.e. a gamma distribution with shape 1 and rate 0.7, its 90% probability mass lies between 0.1 and 4.3 with a median of 1 |
| deltaPrior    | a list with hyper-prior information for the reduction $\delta$ . The default prior is <code>list(priorDist = "beta", hyperpars=c(1,1))</code>   |
| phiPrior      | a list with hyper-prior information for the zero-inflation parameter $\phi$ , The default prior is <code>list(priorDist = "beta", hyperpars=c(1,1))</code>  |
| nsamples      | a positive integer specifying how many iterations for each chain (including burn-in samples)  |
| nburnin       | number of burn-in samples   |
| thinning      | thinning parameter, a positive integer specifying the period for saving samples   |
| nchain        | a positive integer specifying the number of chains  |
| ncore         | number of cores to use when executing the chains in parallel  |
| adaptdelta    | the target acceptance rate, a value between 0 and 1   |
| verbose       | logical. If true, prints progress and debugging information   |

### Details

The first time each model with non-default priors is applied, it can take up to 20 seconds for stan to compile the model. Currently the function only support prior distributions with two parameters. For a complete list of supported priors and their parameterization, please consult the list of distributions in [Stan](#).

Sometimes the function outputs informational message from Stan regarding the Metropolis proposal rejections, this is due to the sampler hit the boundary of the parameter space. For some variables, the boundary point is not supported in the distribution. This is not a concern if there are only a few such warnings.

### Value

Prints out the posterior summary of fecr as the reduction, meanEPG.untreated as the mean faecal egg counts before treatment, and meanEPG.treated as the mean faecal egg counts after treatment. The posterior summary contains the mean, standard deviation (sd), 2.5%, 25%, 50%, 75% and 97.5% percentiles, the 95% highest posterior density interval (HPDLow95 and HPDHigh95) and the posterior mode. NOTE: we recommend to use the 95% HPD interval and the mode for further statistical analysis.

The returned value is a list that consists of:

`stan.samples` An object of S4 class `stanfit` representing the fitted results. For more information, please see the [stanfit-class](#) in `rstan` reference manual.

`posterior.summary` A data frame that is the same as the printed posterior summary.

### Author(s)

Craig Wang <craig.wang@uzh.ch>

### See Also

[simData2s](#) for simulating faecal egg counts data with two samples

### Examples

```
## Not run:
## load the sample data as a vector
data(eggs)

## apply zero-inflation model to the data vector
model<-fecr_stan(eggs[,1],eggs[,2],rawCounts=FALSE,preCF=10,paired=TRUE,zeroInflation=TRUE)
samples<-stan2mcmc(model$stan.samples)

## a demonstration
demo("fecm_stan", package = "eggCounts")

## End(Not run)
```

---

fec\_stan

*Modelling of faecal egg count data (one-sample case)*

---

### Description

Models faecal egg counts data in a one-sample case with (zero-inflated) Poisson-gamma model formulation using the Stan modelling language. It is computationally several-fold faster compare to conventional MCMC techniques. For the installation instruction of Stan, please read: [Stan Installation](#).

### Usage

```
fec_stan(fec, rawCounts = FALSE, CF = 50, zeroInflation = TRUE,
  muPrior, kappaPrior, phiPrior, nsamples = 12000,
  nburnin = 2000, thinning = 1, nchain = 1,
  ncore = 1, adaptdelta = 0.9, verbose = FALSE)
```

**Arguments**

|               |  |
|---------------|--|
| fec           | vector of faecal egg counts  |
| rawCounts     | logical. If true, preFEC and postFEC correspond to raw counts (as counted on equipment). Otherwise they correspond to calculated eggs (raw counts times correction factor). Defaults to FALSE.   |
| CF            | correction factor or vector of correction factors  |
| zeroInflation | logical. If true, uses the model with zero-inflation. Otherwise uses the model without zero-inflation  |
| muPrior       | a list with hyper-prior information for the baseline mean parameter $\mu$ . The default prior is <code>list(priorDist = "gamma", hyperpars=c(1, 0.001))</code> , i.e. a gamma distribution with shape 1 and rate 0.001, its 90% probability mass lies between 51 and 2996                |
| kappaPrior    | a list with hyper-prior information for the dispersion parameter $\kappa$ . The default prior is <code>list(priorDist = "gamma", hyperpars=c(1, 0.7))</code> , i.e. a gamma distribution with shape 1 and rate 0.7, its 90% probability mass lies between 0.1 and 4.3 with a median of 1 |
| phiPrior      | a list with hyper-prior information for zero-inflation parameter. The default prior is <code>list(priorDist = "beta", hyperpars=c(1, 1))</code>  |
| nsamples      | a positive integer specifying how many iterations for each chain (including burn-in samples)   |
| nburnin       | number of burn-in samples  |
| thinning      | thinning parameter, a positive integer specifying the period for saving samples  |
| nchain        | a positive integer specifying the number of chains   |
| ncore         | number of cores to use when executing the chains in parallel   |
| adaptdelta    | the target acceptance rate, a value between 0 and 1  |
| verbose       | logical. If true, prints progress and debugging information  |

**Details**

The first time each non-default model is applied, it can take up to 20 seconds for stan to compile the model. Currently the function only support prior distributions with two parameters. For a complete list of supported priors and their parameterization, please consult the list of distributions in [Stan](#).

Sometimes the function outputs informational message from Stan regarding the Metropolis proposal rejections, this is due to the sampler hit the boundary of the parameter space. For some variables, the boundary point is not supported in the distribution. This is not a concern if there are only a few such warnings.

**Value**

Prints out summary of meanEPG as the posterior mean egg count. The posterior summary contains the mean, standard deviation (sd), 2.5%, 25%, 50%, 75% and 97.5% percentiles, the 95% highest posterior density interval (HPDLow95 and HPDHigh95) and the posterior mode. NOTE: we recommend to use the 95% HPD interval and the mode for further statistical analysis.

The returned value is a list that consists of:



`stan.samples` An object of S4 class `stanfit` representing the fitted results. For more information, please see the `stanfit-class` in `rstan` reference manual.

`posterior.summary`  
A data frame that is the same as the printed posterior summary.

**Author(s)**

Craig Wang <craig.wang@uzh.ch>

**See Also**

[simData1s](#) for simulating faecal egg count data with one sample

**Examples**

```
## Not run:
## load the sample data as a vector
data(echinococcus)
fec<-echinococcus[[1]]

## apply zero-inflation model to the data vector
model<-fec_stan(fec,rawCounts=FALSE,CF=50,zeroInflation=FALSE)
samples<-stan2mcmc(model$stan.samples)

## a demonstration
demo("fecm_stan", package = "eggCounts")

## End(Not run)
```

---

simData1s

*Simulate faecal egg count data (1-sample situation)*

---

**Description**

Simulates (zero-inflated) egg count data

**Usage**

```
simData1s(n = 10, mean = 500, kappa = 0.5, phi = 1, f = 50, rounding = TRUE)
```

**Arguments**

|                    |   |
|--------------------|---|
| <code>n</code>     | sample size (number of faeces collected)  |
| <code>mean</code>  | true number of eggs per gram (epg)  |
| <code>kappa</code> | overdispersion parameter, $\kappa \rightarrow \infty$ corresponds to Poisson  |
| <code>phi</code>   | prevalence i.e. proportion of infected animals, between 0 and 1   |
| <code>f</code>     | correction factor of the egg counting technique, either an integer or a vector of integers with length <code>n</code> |

rounding            logical. If true, the Poisson mean for the raw counts is rounded. The rounding applies since the mean epg is frequently reported as an integer value. For more information, please see Details.

### Details

The simulation process does not exactly match the proposed models in [ref:paper], however the simulated data resembles the data observed in real world.

In the simulation of raw (master) counts, it follows a Poisson distribution with some mean. The mean is frequently rounded down if it has a very low value and `rounding = TRUE`, hence there expects to be a negative bias overall when  $\mu < 150$ . Set `rounding = FALSE` if one does not wish to have any bias in the simulated counts.

### Value

A matrix with three columns, namely the observed epg (`obs`), number of eggs counted on the McMaster slide (`master`) and true egg counts (`true`).

### Author(s)

Craig Wang <craig.wang@uzh.ch>  
Michaela Paul

### See Also

[fec\\_stan](#) for analyzing faecal egg count data with one sample

### Examples

```
fec <- simData1s(n=10, mean=500, kappa=0.5, phi=0.7)
```

---

simData2s

*Simulate faecal egg count data (2-sample situation)*

---

### Description

Generates two samples of (zero-inflated) egg count data

### Usage

```
simData2s(n = 10, preMean = 500, delta = 0.1, kappa = 0.5,  
  phiPre = 1, phiPost = phiPre, f = 50, paired = TRUE,  
  rounding = TRUE)
```

**Arguments**

|          |   |
|----------|---|
| n        | sample size (number of faecal samples collected pre- and post-treatment)  |
| preMean  | true number of eggs per gram (epg) (i.e. worm burden) before treatment  |
| delta    | proportion of epg left after treatment, between 0 and 1. $1 - \delta$ is reduction in mean after treatment, $\delta = 0.1$ indicates a 90% reduction  |
| kappa    | overdispersion parameter, $\kappa \rightarrow \infty$ corresponds to Poisson  |
| phiPre   | pre-treatment prevalence (i.e. proportion of infected animals), between 0 and 1   |
| phiPost  | post-treatment prevalence, between 0 and 1  |
| f        | correction factor of the egg counting technique, either an integer or a vector of integers with length n  |
| paired   | logical. If true, paired samples are simulated. Otherwise unpaired samples are simulated.   |
| rounding | logical. If true, the Poisson mean for the raw counts is rounded. The rounding applies since the mean epg is frequently reported as an integer value. For more information, please see Details. |

**Details**

The simulation process does not exactly match the proposed models in [ref:paper], however the simulated data resembles the data observed in real world.

In the simulation of raw (master) counts, it follows a Poisson distribution with some mean. The mean is frequently rounded down if it has a very low value and `rounding = TRUE`, there expects to be a up to 3-10% positive bias in the mean reduction when  $\mu < 150$  and  $\delta < 0.1$ . Set `rounding = FALSE` if one does not wish to have any bias.

**Value**

A matrix with six columns, namely the observed epg (`obs`), number of eggs counted on microscope slide (`master`) and true egg counts (`true`) for both pre- and post- treatment.

**Author(s)**

Craig Wang <craig.wang@uzh.ch>  
Michaela Paul

**See Also**

[fecr\\_stan](#) for analyzing faecal egg count data with two samples

**Examples**

```
fec <- simData2s(n=10, preMean=500, delta=0.1, kappa=0.5)

## show the positive bias when the true reduction should be 95%
set.seed(1)
fec <- simData2s(n=1e5, preMean=150, delta=0.05, kappa=0.5)
1-mean(fec[,5])/mean(fec[,2])
```

---

`stan2mcmc`*Convert a Stanfit object to MCMC object*

---

### Description

Converts a large `stanfit` object into a MCMC object for easier analysis, it extracts the relevant MCMC samples of the model from `stanFit` including the baseline mean epg, dispersion, pre- and post-treatment zero-inflation parameters and the calculated reduction.

### Usage

```
stan2mcmc(stanFit)
```

### Arguments

`stanFit`            A `stanfit` object from the output of either `fecr_stan()` or `fec_stan()`

### Details

The output can be analyzed as a typical MCMC object with the functions from the `coda` package.  
NOTE: The resulting MCMC object does not contain warm-up samples and is already thinned.

### Value

A MCMC object with a list of relevant parameters depending on the chosen model.

### Author(s)

Craig Wang <craig.wang@uzh.ch>

### Examples

```
## Not run:
data(epgs)

## apply zero-inflation model for the paired design to the data vector
model <- fecr_stan(epgs[,1], epgs[,2], rawCounts=FALSE, preCF=10, paired=TRUE, zeroInflation=TRUE)
samples <- stan2mcmc(model$stan.samples)
summary(samples)

## End(Not run)
```

---

 tab1morgan

*Abundance of trichostrongyloid eggs in sheep faeces*


---

### Description

This data set contains information about the abundance and distribution of trichostrongyloid eggs in the faeces of 14 groups of commercially farmed sheep given in Table 1 in Morgan et al. (2005). The faecal egg counts were assumed to follow a negative binomial distribution with mean  $m$  and overdispersion parameter  $k$ .

### Usage

```
data(tab1morgan)
```

### Format

A data frame with 14 rows and 13 variables

### Details

The data set has columns:

|                      |  |
|----------------------|--|
| group                | ID number for the groups   |
| ageclass             | age class of sheep: "Lambs" or "Ewes"                            |
| month                | month when samples were taken                                    |
| n                    | number of sheep in group   |
| meanFEC              | mean number of eggs per gram (epg) of faeces                     |
| k                    | estimated overdispersion parameter $k$                           |
| k.low                | lower limit of a 95% confidence interval for $k$                 |
| k.up                 | lower limit of a 95% confidence interval for $k$                 |
| maxFEC               | maximal number of eggs per gram of faeces per group              |
| percentageLarger1000 | percentage of samples with more than 1000 epg                    |
| Chi2                 | goodness-of-fit statistic for the negative binomial distribution |
| df                   | corresponding degrees of freedom                                 |
| p                    | corresponding p-value  |

### Source

Morgan ER, Cavill L, Curry GE, Wood RM, Mitchell ESE (2005). Effects of aggregation and sample size on composite faecal egg counts in sheep, *Veterinary Parasitology*, 121:79-87.

### Examples

```
data(tab1morgan)
if (require("lattice"))
  xyplot(k.low+k.up+k ~meanFEC, type="p", pch=19, col=c(8,8,1), data=tab1morgan)
```

# Index

## \*Topic **datasets**

echinococcus, [2](#)

eggs, [3](#)

tab1morgan, [13](#)

## \*Topic **package**

eggCounts-package, [2](#)

echinococcus, [2](#)

eggCounts (eggCounts-package), [2](#)

eggCounts-package, [2](#)

eggs, [3](#)

fec\_stan, [7](#), [10](#)

fecr\_probability, [4](#)

fecr\_stan, [5](#), [11](#)

fecrtCI, [3](#)

simData1s, [9](#), [9](#)

simData2s, [7](#), [10](#)

stan2mcmc, [12](#)

tab1morgan, [13](#)