Package: ewens (via r-universe)

May 19, 2026

Title Ewens Distribution

Version 0.1.0

Description Implements the probability mass function of, and random draws from, the Ewens distribution, a probability distribution over partitions of integer, as described in Ewens (1972) <doi:10.1016/0040-5809(72)90035-4>.

License MIT + file LICENSE

Encoding UTF-8

URL https://github.com/chrishanretty/ewens

BugReports https://github.com/chrishanretty/ewens/issues

RoxygenNote 7.3.2

Imports copula (>= 1.0)

Suggests knitr, quarto

VignetteBuilder quarto

NeedsCompilation yes

Author Chris Hanretty [aut, cre] (ORCID: <https://orcid.org/0000-0002-8932-9405>)

Maintainer Chris Hanretty <chris.hanretty@gmail.com>

Repository https://cran.r-universe.dev

Date/Publication 2026-05-19 09:30:14 UTC

RemoteUrl https://github.com/cran/ewens

RemoteRef HEAD

RemoteSha 063587773f415af6ff20c3f8a34fc57ca0af5610

Contents

`dewens` *Probability mass function for the Ewens distribution*

Description

Gives the probability mass function for the Ewens distribution, as described in Ewens, Warren (1972). "The sampling theory of selectively neutral alleles". *Theoretical Population Biology*. 3: 87–112. doi:10.1016/0040-5809(72)90035-4.

Usage

```
dewens(x, theta = 1, log = FALSE)
```

Arguments

`x`	A vector giving class memberships of each observation in the sample
`theta`	A non-negative parameter governing the expected sample diversity.
`log`	if TRUE, probabilities are given as log(p). Default is FALSE.

Details

The probability of a vector of counts $m_1, ..., m_n$ is given by the expression

$$\frac{n!}{\theta(\theta+1)...(\theta+n-1)} \prod_{j=1}^{n} \frac{\theta^{m_j}}{j^{m_j} m_j!}$$

Value

A numeric vector giving a probability (or if `log = TRUE`, a log probability)

Examples

```
x <- sample(LETTERS, 120, replace = TRUE)
dewens(x, theta = 1)
dewens(x, theta = 0) ## returns NaN since vector incompatible with zero diversity
```

dewens_k *Probability mass function for the number of classes from a Ewens distribution*

Description

Probability mass function for the number of classes from a Ewens distribution

Usage

```
dewens_k(k, n, theta)
```

Arguments

k	An integer number of classes at which to evaluate the PMF
n	A sample size not less than k
theta	A non-negative parameter governing the expected sample diversity.

Details

The number of classes from a Ewens distribution with parameter θ is given by the expression

$$Pr(K = k) = |S_n^k| \frac{\theta^k}{\theta(\theta+1)...(\theta+n-1)}$$

, where $|S_n^k|$ is the absolute value of a Stirling number of the first kind.

Value

The probability of observing k classes

Examples

```
x <- sample(LETTERS, 120, replace = TRUE)
dewens_k(1, 20, theta = 1) ## Pretty unlikely we just see one class
```

ewens_k_exact — *Calculate expected number of classes in a sample of size n given theta*

Description

The expected number of classes from the Ewens distribution is given by $\theta \sum_{j=1}^{n} \frac{1}{\theta+j-1}$. This is often more convenient than integrating across the PMF given by dewens_k

Usage

```
ewens_k_exact(n, theta)
```

Arguments

n	The sample size
theta	The non-negative parameter governing expected sample diversity

ewens_mle — *Maximum likelihood estimate of theta given sample vector with class memberships*

Description

Maximum likelihood estimate of theta given sample vector with class memberships

Usage

```
ewens_mle(x)
```

Arguments

x	A vector containing class memberships; sample size n and number of classes k are calculated from this

Value

A scalar giving the estimate of theta

gcrp *Draw from a generalized Chinese Restaurant Process*

Description

Draw from a generalized Chinese Restaurant Process

Usage

```
gcrp(n, alpha = 0, theta = 1)
```

Arguments

n	The sample size.
alpha	A parameter between zero and one inclusive governing the expected sample diversity
theta	A non-negative parameter governing the expected sample diversity.

Value

A vector of length n consisting of numeric class labels.

Examples

```
rewens(100, 1)
rewens(120, 0.5)
rewens(10, 0)
```

rewens *Draw from the Ewens distribution*

Description

Returns a vector with class membership

Usage

```
rewens(n, theta = 1)
```

Arguments

n	The sample size.
theta	A non-negative parameter governing the expected sample diversity.

Details

Although this command is described as sampling from the Ewens distribution, it is easier to think of it as a particular instantiation of the Chinese Restaurant Process, run for n "customers". The jth customer

- sits at a new table with probability $\frac{\theta}{j-1+\theta}$, or
- sits at an occupied table with probability $\frac{c}{j-1+\theta}$ where c is the number of customers already at each table.

Value

A vector of length n consisting of numeric class labels.

Examples

```
rewens(100, 1)
rewens(120, 0.5)
rewens(10, 0) ## equal to rep(1, 10)
```

`rgem` *Draw from the Griffiths-Engen-McCloskey distribution*

Description

Draw from the Griffiths-Engen-McCloskey distribution

Usage

```
rgem(alpha = 0, theta = 1, trunc_at = 500)
```

Arguments

`alpha`	A parameter between zero and one
`theta`	A parameter which must be greater than -alpha
`trunc_at`	An integer which specifies the maximum number of components to return

Details

The Griffiths-Engen-McCloskey distribution is the infinite dimensional counterpart to the Ewens sampling distribution. This function does not return an infinite dimensional vector(!), but returns a vector of shares creating by a "stick-breaking" construction. The vector of shares is returned after `trunc_at` sticks are broken; this can mean that there is still a non-negligible residual amount.

Value

A vector of shares of length `trunc_at` which may sum to less than one

Index