

# Package ‘PurBayes’

February 19, 2015

**Type** Package

**Title** Bayesian Estimation of Tumor Purity and Clonality

**Version** 1.3

**Date** 2013-05-13

**Author** Nicholas B. Larson

**Maintainer** Nicholas B. Larson <larson.nicholas@mayo.edu>

**Depends** rjags

**Description** PurBayes is an MCMC-based algorithm that uses next-generation sequencing data to estimate tumor purity and clonality for paired tumor-normal data.

**License** GPL-2

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2013-05-16 17:20:31

## R topics documented:

|                                  |   |
|----------------------------------|---|
| PurBayes-package . . . . .       | 2 |
| plot.PurBayes . . . . .          | 2 |
| print.summary.PurBayes . . . . . | 3 |
| PurBayes . . . . .               | 4 |
| summary.PurBayes . . . . .       | 6 |

|              |          |
|--------------|----------|
| <b>Index</b> | <b>7</b> |
|--------------|----------|

---

PurBayes-package      *Bayesian Estimation of Tumor Purity and Clonality*

---

### Description

PurBayes is an MCMC-based mixture-model algorithm that uses next-generation sequencing data to estimate tumor purity and clonality for paired tumor-normal data.

NOTE: Requires JAGS to be installed.

### Details

Package: PurBayes  
Type: Package  
Version: 1.3  
Date: 2013-05-13  
License: GPL-2

### Author(s)

Nicholas B. Larson

Maintainer: Nicholas B. Larson <larson.nicholas@mayo.edu>

---

plot.PurBayes      *Plotting Function for PurBayes*

---

### Description

Generates plot of somatic mutation data and model fit

### Usage

```
## S3 method for class 'PurBayes'  
plot(x, ...)
```

### Arguments

x                      Output object from PurBayes  
...                    Other arguments passed to or from other methods

**Value**

Produces a graphical representation of the somatic mutation data and estimated purity (and sub-clonal cellularities, if detected). These are presented as linear relations against the read data, along with 95% credible intervals.

**Author(s)**

Nicholas B. Larson

**See Also**

[PurBayes](#)

---

print.summary.PurBayes

*Print Summary of PurBayes Analysis*

---

**Description**

S3 method for printing summary of PurBayes output

**Usage**

```
## S3 method for class 'summary.PurBayes'
print(x,...)
```

**Arguments**

|     |   |
|-----|---|
| x   | Object of class <code>summary.PurBayes</code>   |
| ... | Other arguments passed to or from other methods |

**Details**

Prints out summary information from the PurBayes model fit

**Value**

Prints the following information:

Purity Estimate

Posterior distribution summary of tumor purity, given as the posterior median and 2.5% and 97.5% quantiles.

Number of Populations

Value of  $J$ , number of discrete populations of mutations detected by PurBayes

Model Posterior Distributions

Quantiles of all of the parameter posterior distributions in the final model

Penalized Deviance Results

Matrix of penalized deviance results used in determining the final model fit

**Author(s)**

Nicholas B. Larson

**See Also**

[PurBayes,summary.PurBayes](#)

---

PurBayes

*Bayesian Estimation of Tumor Purity and Clonality*

---

**Description**

PurBayes is an iterative Bayesian algorithm which simultaneously estimates tumor purity and clonality using finite mixture models, using the MCMC software JAGS to obtain posterior samples for inference. Using a penalized deviance criterion, PurBayes iteratively fits models increasing in variant population count until an optimal fit is achieved.

**Usage**

```
PurBayes(N, Y, M=NULL, Z=NULL, pop.max=5, prior=NULL, burn.in=50000,
n.post = 10000, fn.jags = "PB.jags", plot = FALSE)
```

**Arguments**

|         |  |
|---------|--|
| N       | numeric vector of total reads for each somatic mutation from the tumor tissue NGS data   |
| Y       | numeric vector of mutant allele supporting read counts for each somatic mutation from the tumor tissue NGS data  |
| M       | optional numeric vector of total reads for germline heterozygous variants. PurBayes uses these to estimate non-reference allele mapping rate to account for mapping bias   |
| Z       | optional numeric vector of alternate allele reads for germline heterozygous variants, corresponding to M   |
| pop.max | Maximum number of variant populations allowed in the iterative modeling procedure. Defaults to 5.  |
| prior   | Optional prior distribution for $\lambda_J$ under the homogenous tumor model. If NULL, defaults to Uniform(0,1). WARNING: This must be provided as a character string written within the JAGS modeling language. |
| burn.in | Number of MCMC draws that are excluded as a burn-in. Defaults to 50000.  |
| n.post  | Number of MCMC draws that are sampled for posterior inference. Defaults to 10000.  |
| fn.jags | File location and name to which write.PB generates the appropriate JAGS model file. Defaults to 'PB.jags' in the current working directory.  |
| plot    | If plot=TRUE, then <a href="#">plot.PurBayes</a> is called to generate a visual representation of the data along with the model fit by PurBayes. Defaults to FALSE.  |

## Details

For a given tumor purity level  $\lambda$  PurBayes assumes a binomial-binomial mixture model for the tumor sequence reads which support the alternate allele,  $Y_i^t \sim Bin(N_i, \lambda/2)$ . This model is fit to the data under the assumption of tumor homogeneity. PurBayes also supports the possibility of intra-tumor heterogeneity, whereby the tumor tissue is comprised of additional subclonal variant populations, each with its own 'purity',  $\lambda_j < \lambda$ , for  $j = 1, \dots, J - 1$  and  $\lambda_J \equiv \lambda$ .

The probability that a given variant corresponds to the  $j^{th}$  population is given by  $\kappa_j$ , and  $\boldsymbol{\kappa} = (\kappa_1, \dots, \kappa_J)$  follows a dirichlet prior such that  $\pi(\boldsymbol{\kappa}) \sim Dirichlet(\alpha_1, \dots, \alpha_J)$  for a given variant population quantity  $J$ . PurBayes applies a diffuse prior on  $\boldsymbol{\kappa}$ , such that  $\alpha_1 = \dots = \alpha_J = 1$ . While the user may specify a particular prior for  $\lambda$  under a homogeneous tumor, PurBayes defaults to  $\pi(\lambda_j) \sim Uniform(0, 1)$  for all  $j$ , and uses a sort function to avoid label switching.

The optimality criterion used for model selection with regard to size of  $J$  is based upon the penalized expected deviance (Plummer, 2008) In instances where the optimism cannot be determined, it is approximated by twice the pD value (along with a warning this approximation is being used).

## Value

List object of designated class PurBayes, which includes data inputs N,Y,M,Z, as well as:

|           |   |
|-----------|---|
| n.pop     | Numeric scalar corresponding to number of variant populations detected by PurBayes  |
| PB.post   | mcmc.list object corresponding to posterior samples of PurBayes model parameters. This necessarily includes <i>pur</i> , the tumor purity. If n.pop>1, posterior samples of $\kappa_j$ and $\lambda_j$ for $j = 1, \dots, J$ are also included. |
| dev.mat   | a matrix of the penalized expected deviance results from the model selection procedure. This includes the penalized expected deviance, the difference in PED with the reference model, and the standard error of that difference.               |
| which.ref | indicates which fitted model is the reference model in the penalized expected deviance analysis. This will either be the fitted model with the minimal PED.   |
| jag.fits  | List of learned JAGS models (object class jags) fit in the model selection process  |

## Author(s)

Nicholas B. Larson

## References

Plummer, M. (2008) Penalized loss functions for Bayesian model comparison. *Biostatistics* doi: 10.1093/biostatistics/kxm049

## Examples

```
#Homogeneous tumor example
N.var<-20
N<-round(runif(N.var,20,200))
lambda<-0.75
Y<-rbinom(N.var,N,lambda/2)
```

```
## Not run: PB.hom<-PurBayes(N,Y)

#Heterogeneous tumor example - 1 subclonal population
N.var<-20
N<-round(runif(N.var,20,200))
lambda.1<-0.75
lambda.2<-0.25
lambda<-c(rep(lambda.1,10),rep(lambda.2,10))
Y<-rbinom(N.var,N,lambda/2)
## Not run: PB.het<-PurBayes(N,Y)
```

---

summary.PurBayes      *Analysis summary from PurBayes*

---

## Description

Provides summary analysis of model fitting from PurBayes

## Usage

```
## S3 method for class 'PurBayes'
summary(object,...)
```

## Arguments

|        |   |
|--------|---|
| object | Output from PurBayes                            |
| ...    | Other arguments passed to or from other methods |

## Value

Returns a list with object class summary.PurBayes:

|           |  |
|-----------|--|
| purity    | Posterior distribution summary of tumor purity                     |
| post.dist | Posterior distribution summary of all model parameters             |
| n.pop     | Number of mutation populations detected by PurBayes                |
| dev.out   | Penalized expected deviance results from model selection procedure |

## Author(s)

Nicholas B. Larson

## See Also

[PurBayes](#), [print.summary.PurBayes](#)

# Index

`plot.PurBayes`, [2](#), [4](#)  
`print.summary.PurBayes`, [3](#), [6](#)  
`PurBayes`, [3](#), [4](#), [4](#), [6](#)  
`PurBayes-package`, [2](#)  
`summary.PurBayes`, [4](#), [6](#)