Software for Weighted Homogeneity Test Statistic Calculations

Edward Susko

Department of Mathematics and Statistics, Dalhousie University

Introduction

The R routines for weighted homogeneity test statistic calculations are in the file `wthom.q` and documentation for them is given there as well. The functions assume that you have software available that will give you the maximum likelihood estimator of the mixing distribution for any choice of the number of components. The functions become available to a running R (or S-Plus) session through a `source` call:

```
> source("wthom.q")
> source("/home/you/Rcode/wthom.q") # if a different directory than R
                                    # was started in
```

The main function is `wthom.test` which computes the weighted homogeneity test statistic. It requires funtions that will compute the component density for the mixture as well as the homogeneity test function, $t(x, \theta)$. Examples of these that can be used for some common mixtures are available in `wthom.q`; they can also be used as a template for different $t(x, \theta)$ choices and/or for different mixtures.

The main function

The main function is `wthom.test` which computes the weighted homogeneity test statistic for a test of the number of components in a mixture. It can be called with `wthom.test(fx, tf, x, wt, theta)`, where the arguments are

- `fx` :a function that computes the component density for the mixture.
- `tf` :a function that computes the homogeneity test function $t(x, \theta)$.
- `x` :the vector or the matrix of observations. For a matrix, individual observation vectors should be stored in columns.
- `wt` :the vector of weight parameters.

`theta` :a vector or matrix of component parameters. The jth column should give the parameters for the jth component of the mixture.

The function `fx` should be of the form `fx(d,x,theta)` and return the vector of component densities (`d = 0`) for the vector (matrix) `x` of all observations and single parameter `theta`. If `d=1`, it should return the matrix of partial derivatives. Each row should give the vector of partial derivatives for the corresponding x value. Example component densities for some common mixtures are listed below and are present in `wthom.q`.

The function `tx` should be of the form `tx(x,theta)` and return the vector of homogeneity test functions $t(x, \theta)$ for the vector (matrix) `x` of all observations and a single parameter `theta`.

The function call `wthom.test(fx, tf, x, wt, theta)` will return a list with components:

`tstat` : the test statistics for each of the components.

`se` : the corresponding standard error estimates.

An example invocation that obtains the p-value for the test that two components are sufficient when data are generated from the normal mixture density

$$\sum_{j=1}^{m} \pi_j \exp[-(x - \mu_j)^2/(2\sigma_j^2)]/(\sigma_j \sqrt{2\pi})$$

is given as:

```
> mu <- c(-2.752136, -0.416129)
> v <- c(0.540891, 3.361862)
> wt <- c(0.198672, 0.801328)
> theta <- rbind(mu, v)
> ts <- wthom.test(wthom.norm.fx, wthom.norm.tf, x, wt, theta)
> pvalue <- 2*pnorm(-abs(ts$ts/ts$se))
```

Here the values $-2.752136, -0.416129$ give the estimated means (the μ_j) for a two-component mixture. The values $0.540891, 3.361862$ give the corresponding variances (the σ_j^2) and $0.198672, 0.801328$ the weights (the π_j).

Component densities

Component density functions are available for normal, binomial and Poisson distributions. If the mixture density is

$$\sum_{j=1}^{m} \pi_j f(x, \theta_j)$$

the component density is $f(x, \theta)$. The function `fx` should be of the form `fx(d,x,theta)` and return the vector of component densities (`d=0`) for the vector (matrix) `x` of all observations and single parameter `theta`. If `d=1`, it should return the matrix of partial derivatives. Contained in `wthom.q` are the component density functions:

- `wthom.norm.fx`: Computes the normal component density

  $$f(x;\theta) = \exp[-(x-\mu)^2/(2\sigma^2)]/(\sigma\sqrt{2\pi})$$

  Here $\theta = (\mu, \sigma^2)^T$.

- `wthom.pois.fx`: Computes the Poisson component density

  $$f(x;\theta) = \exp(-\theta)\theta^x/x!$$

  ignoring the constant factor $x!$ that need not be included in likelihood inference.

- `wthom.binom.fx`: Computes the binomial component density:

  $$f(x;\theta) = \binom{m}{x}\theta^x(1-\theta)^{m-x}$$

  The constant factor $\binom{m}{x}$ is ignored. In the function call `wthom.binom.fx(d, x, theta)`,

  `x`: the observations are assumed across columns. `x[,i]` gives the ith observation: the number of successes, `x[1,i]`, out of `x[2,i]` trials.

Homogeneity test functions

The function `tx` should be of the form `tx(x,theta)` and return the vector of homogeneity test functions $t(x, \theta)$ for the vector (matrix) `x` of all observations and a single parameter `theta`. Contained in `wthom.q` are the homogeneity test functions:

- `wthom.norm.tf`: $t(x,\theta) = |x-\mu|/\sigma^{2^3} - 4/\sqrt{2\pi}$ where $\theta = (\mu, \sigma^2)$. For the normal component density.
- `wthom.pois.tf`: $t(x,\theta) = (x-\theta)^2 - x$. For the Poisson component density.
- `wthom.binom.tf`: $t(x,\theta) = x(x-1)/\theta + (m-x)(m-x-1)/(1-\theta) - m(m-1)$. For the binomial component density.

Susko, E. (2003). Weighted Tests of Homogeneity for Testing the Number of Components in a Mixture. *Journal of Computational Statistics and Data Analysis: Special Issue on Mixtures*, **41**, 367–378.