Package 'RSizeBiased'

March 29, 2021

Type Package

Title Hypothesis Testing Based on R-Size Biased Samples

Version 0.1.0

Maintainer Dimitrios Bagkavos <dimitrios.bagkavos@gmail.com>

Depends R (>= 3.5.0)

Imports stats, pracma

Description Provides functions and examples for testing hypothesis about the population mean and variance on samples drawn by r-size biased sampling schemes.

License GPL (>= 2)

NeedsCompilation no

RoxygenNote 7.1.1

LazyData true

Author Dimitrios Bagkavos [aut, cre],
Polychronis Economou [aut],
Apostolos Batsidis [aut],
Gorgios Tzavelas [aut]

Repository CRAN

Date/Publication 2021-03-29 08:20:02 UTC

R topics documented:

`Cond.KL.Weib.Gamma` *Kullback-Leibler divergence between the (parametrized with respect to shape and mean or variance) of the Weibull or gamma distribution and its (assumed) maximum likelihood estimates.*

Description

The function returns the Kullback-Leibler divergence (minus a constant) between the (parametrized with respect to shape and mean or variance) underlying Weibull or gamma distribution and its (assumed) maximum likelihood estimates.

Usage

```
Cond.KL.Weib.Gamma(par,nullvalue,hata,hatb,type,dist)
```

Arguments

`par`	The (actual) shape parameter α of the distribution.
`nullvalue`	The (actual) distribution mean or variance.
`hata`	Maximum likelihood estimate of the shape parameter of the distribution.
`hatb`	Maximum likelihood estimate of the scale parameter of the distribution.
`type`	Numeric switch, enables the choice of mean or variance: type: 1 for mean, 2 (or any other value != 1) for variance.
`dist`	Character switch, enables the choice of distribution: type "weib" for the Weibull or "gamma" for the gamma distribution.

Details

The Kullback-Leibler divergence between the Weibull(α, β) or the gamma(α, β) and its maximum likelihood estimate Gamma($\hat{\alpha}, \hat{\beta}$) is given by

$$D_{KL} = (\hat{\alpha}-1)\Psi(\hat{\alpha}) - \log\hat{\beta} - \hat{\alpha} - \log\Gamma(\hat{\alpha}) + \log\Gamma(\alpha) + \alpha\log\beta - (\alpha-1)(\Psi(\hat{\alpha}) + \log\hat{\beta}) + \frac{\hat{\beta}\hat{\alpha}}{\lambda}.$$

Since D_{KL} is used to determine the closest distribution - given its mean or variance - to the estimated gamma p.d.f., the first four terms are omitted from the function outcome, i.e. the function returns the result of the following quantity:

$$\log\Gamma(\alpha) + \alpha\log\beta - (\alpha-1)(\Psi(\hat{\alpha}) + \log\hat{\beta}) + \frac{\hat{\beta}\hat{\alpha}}{\lambda}.$$

For the Weibull distribution the corresponding formulas are

$$D_{KL} = \log\frac{\hat{\alpha}}{\hat{\beta}^{\hat{\alpha}}} - \log\frac{\alpha}{\beta^{\alpha}} + (\hat{\alpha}-\alpha)\left(\log\hat{\beta} - \frac{\gamma}{\hat{\alpha}}\right) + \left(\frac{\hat{\beta}}{\beta}\right)^{\alpha}\Gamma\left(\frac{\alpha}{\hat{\alpha}}+1\right) - 1$$

and since D_{KL} is used to determine the closest distribution - given its mean or variance - to the estimated gamma p.d.f., the first term is omitted from the function outcome, i.e. the function returns the result of the following quantity:

$$-\log\frac{\alpha}{\beta^{\alpha}}+(\hat{\alpha}-\alpha)\left(\log\hat{\beta}-\frac{\gamma}{\hat{\alpha}}\right)+\left(\frac{\hat{\beta}}{\beta}\right)^{\alpha}\Gamma\left(\frac{\alpha}{\hat{\alpha}}+1\right)-1$$

Value

A scalar, the value of the Kullback-Leibler divergence (minus a constant).

Author(s)

Polychronis Economou

R implementation and documentation: Polychronis Economou <peconom@upatras.gr>

References

Economou et. al. (2021). Hypothesis testing for the population mean and variance based on r-size biased samples, under review.

Examples

```
#K-L divergence for the Gamma distribution for shape=2
#and variance=3 and their assumed MLE=(1,1):
 Cond.KL.Weib.Gamma(2,3,1,1,2, "gamma")
#K-L divergence for the Weibull distribution for shape=2
#and variance=3 and their assumed MLE=(1,1):
 Cond.KL.Weib.Gamma(2,3,1,1,2, "weib")
```

---

d_rsize_Weibull — *Weibull size biased distribution of order r.*

Description

Calculates the density of the $r-$size biased Weibull distribution.

Usage

```
d_rsize_Weibull(x,TRpar,r)
```

Arguments

x	Grid points where the functional is being calculated.
TRpar	A vector of length 2, containing the shape and scale parameters of the distribution.
r	The size (order) of the distribution. The special cases $r = 1, 2, 3$ correspond to length, area, volume biased samples respectively and are the most frequently encountered in practice. The case $r = 0$ corresponds to random samples from the Weibull distribution.

Details

The $r-$size density of the observed biased sample $X_1, \ldots, X_n$ is defined by

$$f_r(x;\theta) = \frac{x^r f(x;\theta)}{E(X^r)}$$

where $f(x;\theta)$ is the density of the Weibull distribution and θ the vector of the shape and scale parameters of the distribution.

Value

A vector of length equal to the length of x.

Author(s)

Polychronis Economou

R implementation and documentation: Polychronis Economou <peconom@upatras.gr>

References

Economou et. al. (2021). Hypothesis testing for the population mean and variance based on r-size biased samples, under review.

See Also

p_rsize_Weibull, r_rsize_Weibull

Examples

```
# example of r-size Weibull distribution, r=0,1,2
x<- seq(0, 10, length=50)
dens.0.size<-d_rsize_Weibull(x,c(2,3),0)
dens.1.size<-d_rsize_Weibull(x,c(2,3),1)
dens.2.size<-d_rsize_Weibull(x,c(2,3),2)
plot(x, dens.0.size, type="l", ylab="r-denisty")
lines(x, dens.1.size, col=2)
lines(x, dens.2.size, col=3)
legend("topright",legend=c("r= 0","r= 1","r= 2"),
       col=c("black","red","green"),lty=c(1,1,1))
```

`log_Lik_Weib_gamma_weighted`
Log likelihood function for the weighted gamma or Weibull distributions.

Description

Calculates the log-likelihood function of the weighted gamma or Weibull (depends on user input) distribution.

Usage

```
log_Lik_Weib_gamma_weighted(TRpar,datain,r,dist)
```

Arguments

`TRpar`	A vector of length 2, containing the shape and scale parameters of the distribution.
`datain`	The available sample points.
`r`	The size (order) of the distribution. The special cases $r = 1, 2, 3$ correspond to length, area, volume biased samples respectively and are the most frequently encountered in practice. The case $r = 0$ corresponds to random samples from the Gamma distribution.
`dist`	Character switch, enables the choice of distribution: type "weib" for the Weibull or "gamma" for the gamma distribution.

Details

The log likelihood function of the weighted gamma distribution is defined by

$$\log L = \sum_{i=1}^{n} log f_r(X_i;\theta)$$

where $f_r(x;\theta)$ is the density of the $r-$size biased gamma distribution. Setting $r = 0$ corresponds to the log likelihood of the Gamma distribution.

In the case of Weibull, the log likelihood is defined by

$$\log L = \sum_{i=1}^{n} log f_r(X_i;\theta)$$

where $f_r(x;\theta)$ is the density of the $r-$size biased Weibull distribution. Setting $r = 0$ corresponds to the log likelihood of the Weibull distribution.

Value

A scalar, the result of the log likelihood calculation.

Author(s)

Polychronis Economou

R implementation and documentation: Polychronis Economou <peconom@upatras.gr>

References

Economou et. al. (2021). Hypothesis testing for the population mean and variance based on r-size biased samples, under review.

Examples

```
#Log-likelihood for the gamma distribution for true parms=(2,3), r=0:
log_Lik_Weib_gamma_weighted(c(2,3), rgamma(100, shape=2, scale=3), 0, "gamma")
#Log-likelihood for the Weibull distribution for true parms=(2,3), r=0:
log_Lik_Weib_gamma_weighted(c(2,3), rweibull(100, shape=2, scale=3), 0, "weib")
```

---

p_rsize_Weibull   *Weibull size biased c.d.f. of order r.*

Description

Calculates the cumulative distribution of the $r-$size biased Weibull distribution.

Usage

```
p_rsize_Weibull(q,TRpar,r)
```

Arguments

q	Points where the functional is being calculated.
TRpar	A vector of length 2, containing the shape and scale parameters of the distribution.
r	The size (order) of the distribution. The special cases $r = 1, 2, 3$ correspond to length, area, volume biased samples respectively and are the most frequently encountered in practice. The case $r = 0$ corresponds to random samples from the Weibull distribution.

Details

The $r-$size c.d.f. of the Weibull density is defined by

$$F_r(y;\theta) = \int_0^y \frac{x^r f(x;\theta)}{E(X^r)}\, dx$$

where θ is a bivariate vector with the the shape and scale of the Weibull distribution.

Value

A vector of length equal to the lemgth of x.

Author(s)

Polychronis Economou

R implementation and documentation: Polychronis Economou <peconom@upatras.gr>

References

Economou et. al. (2021). Hypothesis testing for the population mean and variance based on r-size biased samples, under review.

See Also

d_rsize_Weibull, r_rsize_Weibull

Examples

```
# c.d.f of the r-size Weibull distribution, r=0,1,2 evalutated at a specific point x.
x<- 2
dist.0.size<-p_rsize_Weibull(x,c(2,3),0)
dist.1.size<-p_rsize_Weibull(x,c(2,3),1)
dist.2.size<-p_rsize_Weibull(x,c(2,3),2)
```

---

r_moment_gamma_Weib   *$r-$th moment of the gamma or the Weibull distribution.*

Description

Calculates the $r-$th moment of the gamma or Weibull distribution.

Usage

```
r_moment_gamma_Weib(TRpar,r,dist)
```

Arguments

TRpar	A vector of length 2, containing the shape and scale parameters of the distribution.
r	The size (order) of the distribution. The special cases $r = 1, 2, 3$ correspond to length, area, volume biased samples respectively and are the most frequently encountered in practice. The case $r = 0$ corresponds to random samples from the Gamma distribution.
dist	Character switch, enables the choice of distribution: type "weib" for the Weibull or "gamma" for the gamma distribution.

Details

In the case of the $\Gamma(\alpha, \beta)$ distribution the r-th moment is given by

$$\mu_r = \int_0^\infty x^r f(x;\alpha,\beta)\,dx = \beta^r \frac{\Gamma(\alpha+r)}{\Gamma(\alpha)}, \alpha > -r$$

while for the $W(\alpha, \beta)$ distribution the r-th moment is given by

$$\mu_r = \int_0^\infty x^r f(x;\alpha,\beta)\,dx = \beta^r \Gamma\left(1+\frac{\alpha}{r}\right), \alpha > -r$$

Value

A scalar, the value of the moment.

Author(s)

Polychronis Economou

R implementation and documentation: Polychronis Economou <peconom@upatras.gr>

References

Economou et. al. (2021). Hypothesis testing for the population mean and variance based on r-size biased samples, under review.

Examples

```
#r-moment for the Gamma distribution for true parms=(2,3), r=1:
r_moment_gamma_Weib(c(2,3),1, "gamma")
#r-moment for for the Weibull distribution for true parms=(2,3), r=1:
r_moment_gamma_Weib(c(2,3),1, "weib")
```

---

r_rsize_Weibull — *Weibull size biased random number generation of order r (modified).*

Description

Provides a random sample of size n from the $r-$size biased Weibull distribution (modified).

Usage

```
r_rsize_Weibull(n,TRpar,r)
```

Arguments

n	Number of th sample data points to be provided.
TRpar	A vector of length 2, containing the shape and scale parameters of the distribution.
r	The size (order) of the distribution. The special cases $r = 1, 2, 3$ correspond to length, area, volume biased samples respectively and are the most frequently encountered in practice. The case $r = 0$ corresponds to random samples from the Weibull distribution.

Details

The $r-$size random number generator from the Weibull distribution is implemented based on a change-of-variable technique, to the standard gamma distribution as described by Gove and Patil (1998).

Value

A vector of length n with the random sample.

Author(s)

Polychronis Economou

R implementation and documentation: Polychronis Economou <peconom@upatras.gr>

References

Gove J.H. and Patil G.P. (1998). Modeling the Basal Area-size Distribution of Forest Stands: A Compatible Approach. Forest Science, 44(2), 285-297.

See Also

d_rsize_Weibull, p_rsize_Weibull

Examples

```
#Random number geenration for the r-size Weibull distribution.
r_rsize_Weibull(100,c(2,3),1)
```

`s11.s22` *Variance estimates for test statistics $\zeta\_n, r\hat{}i, i = 1, 2$ specifically for the Weibull and gamma distributions.*

Description

Variance estimates for test statistics $\zeta^i_{n,r}, i = 1, 2$ specifically for the Weibull and gamma distributions.

Usage

```
s11.s22(TRpar,r,sgg,dist)
```

Arguments

`TRpar`	A vector of length 2, containing the shape and scale parameters of the Weibull distribution.
`r`	The size (order) of the distribution. The special cases $r = 1, 2, 3$ correspond to length, area, volume biased samples respectively and are the most frequently encountered in practice. The case $r = 0$ corresponds to random samples from the underlying distribution.
`sgg`	Character switch ("s11" or "s22"), enables choosing between the s11 and s22 options
`dist`	Character switch, enables the choice of distribution: type "weib" for the Weibull or "gamma" for the gamma distribution.

Details

Provided that $\mu_r, r = 1, 2, \ldots$ is the rth moment of the Weibull or the Gamma distribution, then

$$\sigma^2_{1,r} = \mu_r(\mu_{2-r}) - 2\mu_1\mu_{1-r} + \mu_1^2\mu_{-r}$$

and

$$\sigma^2_{2,r} = -4\mu_r\big(2\mu_1^2 - \mu_2) - 2)\mu_1\mu_{1-r} + (2\mu_1^2 - \mu_2)^2 + (8\mu_1^2 - 2\mu_2)\mu_{2-r} - 4\mu_1\mu_{3-r} + \mu_{4-r}\big)$$

Value

A scalar with the value of the variance estimate for the test statistic.

Author(s)

Polychronis Economou

R implementation and documentation: Polychronis Economou <peconom@upatras.gr>

References

Economou et. al. (2021). Hypothesis testing for the population mean and variance based on r-size biased samples, under review.

See Also

zeta_plug_in

Examples

```
#s11 for the Gamma distribution for true parms=(2,3), r=1:
s11.s22(c(2,3),1, "s11", "gamma")
#s22 for for the Weibull distribution for true parms=(2,3), r=1:
s11.s22(c(2,3),1, "s22",  "weib")
```

Size.BiasedMV.Tests *Test statistics.*

Description

The function returns the test statistics for testing a null hypothesis for the mean and a null hypothesis for the varaince.

Usage

```
Size.BiasedMV.Tests(datain_r,r,nullMEAN,nullVAR,start_par,nboot,alpha,prior_sel,distr)
```

Arguments

datain_r	The available sample points.
r	The size (order) of the distribution. The special cases $r = 1, 2, 3$ correspond to length, area, volume biased samples respectively and are the most frequently encountered in practice. The case $r = 0$ corresponds to random samples from the gamma or theWeibull distribution.
nullMEAN	The null value of the distribution mean.
nullVAR	The null value of the distribution variance.
start_par	Vector with two values, containing the starting values for the MLE for the two parameter distribution (Weibull or gamma) .
nboot	Defines the number of bootstrap replications.
alpha	Significance level.
prior_sel	"normal" for the normal distribution or "gamma" for the gamma.
distr	Character switch, enables the choice of distribution: type "weib" for the Weibull or "gamma" for the gamma distribution.

Details

The test statistics implemented are given by the Plug-in and the bootstrap Methods as described in section 3.1 and 3.2 of Economou et al (2021).

Value

An object containing the following components.

- `par` — A vector of the MLE of the distribution parameters.
- `loglik` — A scalar, the maximized log-likelihood.
- `CovMatrix` — The Variance - Covariance matrix of the MLEs.
- `Zeta_i` — A vector of the values of the $\zeta^i_{n,r}, i = 1, 2$ test statistics (if defined)
- `Tivalues` — A vector of the values of the $T^i_{n,r}, i = 1, 2$ test statistics
- `T1_bootstrap_quan` — A vector of the bootstrap quantiles for the $T^1_{n,r}$ test statistic for each one of the significance levels alpha.
- `T2_bootstrap_quan` — A vector of the bootstrap quantiles for the $T^2_{n,r}$ test statistic for each one of the significance levels alpha.
- `NullValues` — A vector of the null values of the distribution mean and variance.
- `distribution` — Character representing the choice of distribution: "weib" for the Weibull or "gamma" for the gamma distribution.
- `alpha` — A vector of significance levels for the test level.
- `bootstrap_p_mean` — A scalar with the bootstrap p-value for testing the mean.
- `bootstrap_p_var` — A scalar with the bootstrap p-value for testing the variance.
- `decision` — A matrix of 0 and 1 of the decisions taken for each one of the significance levels alpha based on the bootstrap method. The first row corresponds to the null hypothesis for the mean and the second to the null hypothesis for the variance.
- `asymptotic_p_mean` — A scalar with the asymptotic p-value for testing the mean (if $\zeta^1_{n,r}$ is defined).
- `asymptotic_p_var` — A scalar with the asymptotic p-value for testing the variance (if $\zeta^2_{n,r}$ is defined).
- `decisionasympt` — A matrix of 0 and 1 of the decisions taken for each one of the significance levels alpha based on the plug-in method and the asymptotic distribution of the test statistics. The first row corresponds to the null hypothesis for the mean and the second to the null hypothesis for the variance.
- `prior_selection` — Character representing the choice of the prior distribution for the bootstrap method: "normal" for the normal distribution or "gamma" for the gamma.

Author(s)

Polychronis Economou

R implementation and documentation: Polychronis Economou <peconom@upatras.gr>

References

Economou et. al. (2021). Hypothesis testing for the population mean and variance based on r-size biased samples, under review.

Examples

```
data(ufc)
datain_r <- ufc[,4]
nullMEAN <- 14 #according to null mean in Sec. 6.3,  Economou et. al. (2021).
nullVAR <- 180 #according to null variance in Sec. 6.3,  Economou et. al. (2021).
Size.BiasedMV.Tests(datain_r, 2, nullMEAN, nullVAR,  c(2,3), 100, 0.05, "normal", "gamma")
```

`T1T2.Mean.Var` *Test statistic $T\_n, r\hat{}1$ or $T\_n, r\hat{}2$ depending on user input.*

Description

The test statistics $T^1_{n,r}$ and $T^2_{n,r}$ are consistent estimators of the mean value $\mathrm{E}(X)$ and variance $\mathrm{Var}(X)$ respectively given an $r-$size biased sample.

Usage

```
T1T2.Mean.Var(datain,r, type)
```

Arguments

`datain`	The available sample points.
`r`	The size (order) of the distribution. The special cases $r = 1, 2, 3$ correspond to length, area, volume biased samples respectively and are the most frequently encountered in practice. The case $r = 0$ corresponds to random samples from the underlying distribution.
`type`	Numeric switch: type =1 corresponds to the T1 statistic while any other numeric value will cause calculation of T2.

Details

The test statistic $T^1_{n,r}$ is defined by

$$T^1_{n,r} = \frac{\sum_{i=1}^n X_i^{1-r}}{\sum_{i=1}^n X_i^{-r}}.$$

The test statistic $T^2_{n,r}$ is defined by

$$T^2_{n,r} = \frac{\sum_{i=1}^n X_i^{2-r}}{\sum_{i=1}^n X_i^{-r}} - \left(\frac{\sum_{i=1}^n X_i^{1-r}}{\sum_{i=1}^n X_i^{-r}}\right)^2.$$

Value

A scalar, the value of the test statistic for the given sample.

Author(s)

Polychronis Economou

R implementation and documentation: Polychronis Economou <peconom@upatras.gr>

References

Economou et. al. (2021). Hypothesis testing for the population mean and variance based on r-size biased samples, under review.

Examples

```
#e.g.:
T1T2.Mean.Var(rgamma(100, 2,3),0, 1)
```

---

ufc *Upper Flat Creek forest cruise tree data*

Description

Forest measurement data from the Upper Flat Creek unit of the University of Idaho Experimental Forest, measured in 1991.

Usage

```
ufc
```

Format

A data frame with 336 observations on the following 5 variables; plot (plot label), tree (tree label), species (species kbd with levels DF, GF, WC, WL), dbh.cm (tree diameter at 1.37 m. from the ground, measured in centimetres.), height.m (tree height measured in metres).

Details

The inventory was based on variable radius plots with 6.43 sq. m. per ha. BAF (Basal Area Factor). The forest stand was 121.5 ha. This version of the data omits errors, trees with missing heights, and uncommon species. The four species are Douglas-fir, grand fir, western red cedar, and western larch.

Source

Harold Osborne and Ross Appelgren of the University of Idaho Experimental Forest.

References

Robinson, A.P., and J.D. Hamann. 2010. Forest Analytics with R: an Introduction. Springer.

Examples

```
data(ufc)
```

`zeta_plug_in` *$\zeta\_n,r\hat{}i, i = 1, 2$ test statistic for the Weibull or the gamma distribution (depending on user input.*

Description

Studentized version of the $T^i_{n,r}, i = 1, 2$ test statistic for the Weibull/gamma distribution.

Usage

```
zeta_plug_in(null_value, datain,r,EST_par,type, dist)
```

Arguments

`null_value`	The parameter value in the hypothesis test under the null
`datain`	The available sample points.
`r`	The size (order) of the distribution. The special cases $r = 1, 2, 3$ correspond to length, area, volume biased samples respectively and are the most frequently encountered in practice. The case $r = 0$ corresponds to random samples from the underlying distribution.
`EST_par`	A vector of length 2, containing the shape and scale parameters of the Weibull distribution.
`type`	Numeric switch: type =1 returns the $\zeta^1_{n,r}$ test statistic, any other value returns $\zeta^2_{n,r}$
`dist`	Character switch, enables the choice of distribution: type "weib" for the Weibull or "gamma" for the gamma distribution.

Details

When `type=1` the function returns

$$\sqrt{n}\frac{T_{n,r^1} - \mu^0}{\sigma_{1,r}(\hat{\theta}_n)} \to N(0, 1)$$

after using the fact that under the null we have $\mu_1 = \mu^0$. Any other value for `type` returns

$$\sqrt{n}\frac{T_{n,r^2} - \sigma_0^2}{\sigma_{2,r}(\hat{\theta}_n)} \to N(0, 1)$$

in which case the fact that var(X)= σ_0^2 under the null has been used.

Value

A scalar with the value of the test statistic.

Author(s)

Polychronis Economou

R implementation and documentation: Polychronis Economou <peconom@upatras.gr>

References

Economou et. al. (2021). Hypothesis testing for the population mean and variance based on r-size biased samples, under review.

Examples

```
data(ufc)
datain_r <- ufc[,4]
nullMEAN <- 14
# ml estimates = c(2.6555,8.0376),  taken from section 6.2 in Economou et. al. (2021).
zeta_plug_in(nullMEAN, datain_r, 2, c(2.6555,8.0376),1, "gamma") #corresponds to mean

nullVar <- 180
zeta_plug_in(nullVar, datain_r, 2, c(2.6555,8.0376),2, "gamma") #corresponds to var
```

Index