Analysis of multivariate survival data

Klaus Holst & Thomas Scheike

March 3, 2020

Overview

When looking at multivariate survival data with the aim of learning about the dependence that is present, possibly after correcting for some covariates different approaches are available in the mets package

- Binary models and adjust for censoring with inverse probabilty of censoring weighting

  - biprobit model

- Bivariate surival models of Clayton-Oakes type

  - With regression structure on dependence parameter
  - With additive gamma distributed random effects
  - Special functionality for polygenic random effects modelling such as ACE, ADE ,AE and so forth.

- Plackett OR model model

  - With regression structure on OR dependence parameter

- Cluster stratified Cox

  Typically it can be hard or impossible to specify random effects models with special structure among the parameters of the random effects. This is possible for our specification of the random effects models.

  To be concrete about the model structure assume that we have paired binomial data $T_1, \delta_1, T_2, \delta_2, X_1, X_2$ where the censored survival responses are $T_1, \delta_1, T_2, \delta_2$ and we have covariates X_1, X_2.

  The basic models assumes that each subject has a marginal on Cox-form

$$\lambda_{s(k,i)}(t) \exp(X_{ki}^T \beta)$$

where $s(k, i)$ is a strata variable.

Gamma distributed frailties

The focus of this vignette is describe how to work on bivariate survival data using the addtive gamma-random effects models. We present two different ways of specifying different dependence structures.

- Univariate models with a single random effect for each cluster and with a regression design on the variance.

- Multivariate models with multiple random effects for each cluster.

The univariate models are then given a given cluster random effects Z_k with parameter θ the joint survival function is given by the Clayton copula and on the form

$$\psi(\theta, \psi^{-1}(\theta, S_1(t, X_{k1})) + \psi^{-1}(\theta, S_1(t, X_{k1}))$$

where ψ is the Laplace transform of a gamma distributed random variable with mean 1 and variance θ.

We then model the variance within clusters by a cluster specific regression design such that

$$\theta = z_j^T \alpha$$

where z is the regression design (specified by theta.des in the software).

This model can be fitted using a pairwise likelihood or the pseudo-likelihood using either

- twostage

- twostageMLE

To make the twostage approach possible we need a model with specific structure for the marginals. Therefore given the random effect of the clusters the survival distributions within a cluster are independent and on the form

$$P(T_j > t | X_j, Z) = exp(-Z \cdot \Psi^{-1}(\nu^{-1}, S(t | X_j)))$$

with Ψ the laplace of the gamma distribution with mean 1 and variance $1/\nu$.

Additive Gamma frailties

For the multivariate models we are given a multivarite random effect each cluster $Z = (Z_1, ..., Z_d)$ with d random effects. The total random effect for each subject j in a cluster is then specified using a regression design on these random effects, with a regression vector V_j such that the total random effect is $V_j^T(Z_1, ..., Z_d)$. The elements of V_J are 1/0. The random effects $(Z_1, ..., Z_d)$ has associated parameters $(\lambda_1, ..., \lambda_d)$ and Z_j is Gamma distributed with

- mean $\lambda_j / V_1^T \lambda$

- variance $\lambda_j / (V_1^T \lambda)^2$

The key assumption to make the two-stage fitting possible is that

$$\nu = V_j^T \lambda$$

is constant within clusters. The consequence of this is that the total random effect for each subject within a cluster, $V_j^T(Z_1, ..., Z_d)$, is gamma distributed with variance $1/\nu$.

The DEFAULT parametrization (var.par=1) uses the variances of the random effecs

$$\theta_j = \lambda_j / v^2$$

For alternative parametrizations one can specify that the parameters are $\theta_j = \lambda_j$ with the argument var.par=0.

Finally the parameters $(\theta_1, ..., \theta_d)$ are related to the parameters of the model by a regression construction M (d x k), that links the d θ parameters with the k underlying α parameters

$$\theta = M\alpha.$$

The default is a diagonal matrix for M. This can be used to make structural assumptions about the variances of the random-effects as is needed for the ACE model for example. In the software M is called theta.des

Assume that the marginal survival distribution for subject i within cluster k is given by $S_{X_{k,i}}(t)$ given covariates $X_{k,i}$.

Now given the random effects of the cluster Z_k and the covariates $X_{k,i}$ $i = 1, \ldots, n_k$ we assume that subjects within the cluster are independent with survival distributions

$$\exp(-(V_{k,i}Z_k)\Psi^{-1}(v, S_{X_{k,i}}(t))).$$

A consequence of this is that the hazards given the covariates $X_{k,i}$ and the random effects Z_k are given by

$$\lambda_{k,i}(t; X_{k,i}, Z_{k,i}) = (V_{k,i}V_k)D_3\Psi^{-1}(v, S_{X_{k,i}}(t))D_t S_{X_{k,i}}(t) \qquad (1)$$

where D_t and D_3 denotes the partial derivatives with respect to t and the third argument, respectively.

Further, we can express the multivariate survival distribution as

$$S(t_1, \ldots, t_m) = \exp(-\sum_{i=1}^{m}(V_i Z)\Psi^{-1}(\eta_l, v_l, S_{X_{k,i}}(t_i)))$$

$$= \prod_{l=1}^{p}\Psi(\eta_l, \eta, \sum_{i=1}^{m}Q_{k,i}\Psi^{-1}(\eta, \eta, S_{X_{k,i}}(t_i))). \qquad (2)$$

In the case of considering just pairs, we write this function as $C(S_{k,i}(t), S_{k,j}(t))$.

In addition to survival times from this model, we assume that we independent right censoring present $U_{k,i}$ such that the given V_k and the covariates $X_{k,i}$ $i = 1, \ldots, n_k$ $(U_{k,1}, \ldots, U_{k,n_k})$ of $(T_{k,1}, \ldots, T_{k,n_k})$, and the conditional censoring distribution do not depend on V_k.

One consequence of the model strucure is that the Kendall's can be computed for two-subjects (i, j) across two clusters "1" and "2" as

$$E\left(\frac{(V_{1i}Z_1 - V_{1j}Z_2)(V_{2i}Z_1 - V_{2j}Z_2)}{(V_{1i}Z_1 + V_{2i}Z_2)(V_{1j}Z_1 + V_{2j}Z_2)}\right) \qquad (3)$$

under the assumption that that we compare pairs with equivalent marginals, $S_{X_{1,i}}(t) = S_{X_{2,i}}(t)$ and $S_{X_{1,j}}(t) = S_{X_{2,j}}(t)$, and that

$S_{X_{1,i}}(\infty) = S_{X_{1,j}}(\infty) = 0$. Here we also use that η is the same across clusters. The Kendall's tau would be the same for (**??**) due to the same additive structure for the frailty terms, and the random effects thus have the same interpretation in terms of Kendall's tau.

Univariate gamma (clayton-oakes) model twostage models

We start by fitting simple Clayton-Oakes models for the data, that is with an overall random effect that is Gamma distrubuted with variance θ. We can fit the model by a pseudo-MLE (twostageMLE) and a pairwise composite likelihood approach (twostage).

The pseudo-liklihood and the composite pairwise likelhood gives quite similar results in this case. In addition the log-parametrization is illustrated with the var.link=1 option. In addition it is specified that we want a "clayton.oakes" model.

```
1  library(mets)
2  data(diabetes)
3
4  # Marginal Cox model with treat as covariate
5  margph <- phreg(Surv(time,status)~treat+cluster(id),data=
       diabetes)
6  # Clayton-Oakes, MLE
7  fitco1<-twostageMLE(margph,data=diabetes,theta=1.0)
8  summary(fitco1)
9
10 # Clayton-Oakes
11 fitco2 <- survival.twostage(margph,data=diabetes,theta=0.0,
       detail=0,
12     clusters=diabetes$id,var.link=1,model="clayton.oakes"
           )
13 summary(fitco2)
14 fitco3 <- survival.twostage(margph,data=diabetes,theta=1.0,
       detail=0,
15     clusters=diabetes$id,var.link=0,model="clayton.oakes"
           )
16 summary(fitco3)
```

```
Loading required package: timereg
Loading required package: survival
Loading required package: lava
mets version 1.2.4
Dependence parameter for Clayton-Oakes model
Variance of Gamma distributed random effects
$estimates
            Coef.        SE        z       P-val Kendall tau         SE
dependence1 0.9526614 0.3543033 2.68883 0.007170289   0.322645 0.08127892


$type
NULL

attr(,"class")
[1] "summary.mets.twostage"
Dependence parameter for Clayton-Oakes model
Variance of Gamma distributed random effects
With log-link
$estimates
            log-Coef.       SE          z     P-val Kendall tau          SE
dependence1 -0.04849523 0.330665 -0.1466597 0.8834006   0.3226451 0.07226526
```

```
$vargam
            Estimate Std.Err   2.5% 97.5%  P-value
dependence1   0.9527    0.315 0.3352  1.57 0.002493


$type
[1] "clayton.oakes"


attr(,"class")
[1] "summary.mets.twostage"
Dependence parameter for Clayton-Oakes model
Variance of Gamma distributed random effects
$estimates
                 Coef.        SE        z      P-val Kendall tau         SE
dependence1 0.9526619 0.3150119 3.024209 0.002492843   0.3226451 0.07226526


$type
[1] "clayton.oakes"


attr(,"class")
[1] "summary.mets.twostage"
```

The marginal models can be either structured Cox model or as here with a baseline for each strata. This gives quite similar results to those before.

```
1  # without covariates but marginal model stratified
2  marg <- phreg(Surv(time,status)~+strata(treat)+cluster(id),
      data=diabetes)
3  fitcoa <- survival.twostage(marg,data=diabetes,theta=1.0,
      clusters=diabetes$id,
4        model="clayton.oakes")
5  summary(fitcoa)
```

```
Dependence parameter for Clayton-Oakes model
Variance of Gamma distributed random effects
With log-link
$estimates
              log-Coef.        SE        z     P-val Kendall tau         SE
dependence1 -0.05683996 0.3322322 -0.171085 0.8641569   0.3208241 0.07239207


$vargam
            Estimate Std.Err   2.5% 97.5%  P-value
dependence1   0.9447   0.3139 0.3296  1.56 0.002613


$type
[1] "clayton.oakes"


attr(,"class")
[1] "summary.mets.twostage"
```

Piecewise constant Clayton-Oakes model

Let the cross-hazard ratio (CHR) be defined as

$$\eta(t_1, t_2) = \frac{\lambda_1(t_1|T_2 = t_2)}{\lambda_1(t_1|T_2 \geq t_2)} = \frac{\lambda_2(t_2|T_1 = t_1)}{\lambda_2(t_2|T_1 \geq t_1)} \tag{4}$$

where λ_1 and λ_2 are the conditional hazard functions of T_1 and T_2 given covariates. For the Clayton-Oakes model this ratio is $\eta(t_1, t_2) = 1 + \theta$, and as a consequence we see that if the co-twin is dead at any time we would increase our risk assessment on the

hazard scale with the constant $\eta(t_1, t_2)$. The Clayton-Oakes model also has the nice property that Kendall's tau is linked directly to the dependence parameter θ and is $1/(1 + 2/\theta)$.

A very useful extension of the model the constant cross-hazard ratio (CHR) model is the piecewise constant cross-hazard ratio (CHR) for bivariate survival data [1], and this model was extended to competing risks in [2].

In the survival setting we let the CHR

$$\eta(t_1, t_2) = \sum \eta_{i,j} I(t_1 \in I_i, t_2 \in I_j) \tag{5}$$

The model lets the CHR by constant in different part of the plane. This can be thought of also as having a separate Clayton-Oakes model for each of the regions specified in the plane here by the cut-points $c(0, 0.5, 2)$ thus defining 9 regions.

This provides a constructive goodness of fit test for the whether the Clayton-Oakes model is valid. Indeed if valid the parameter should be the same in all regions.

First we generate some data from the Clayton-Oakes model with variance 0.5 and 2000 pairs. And fit the related model.

```
d <- simClaytonOakes(2000,2,0.5,0,3)
 margph <- phreg(Surv(time,status)~x+cluster(cluster),data=d)
# Clayton-Oakes, MLE
fitco1<-twostageMLE(margph,data=d)
summary(fitco1)
```

```
Dependence parameter for Clayton-Oakes model
Variance of Gamma distributed random effects
$estimates
            Coef.        SE       z P-val Kendall tau        SE
dependence1 2.01888 0.08970192 22.50654     0   0.5023489 0.01110764


$type
NULL

attr(,"class")
[1] "summary.mets.twostage"
```

Now we cut the region at the cut-points $c(0, 0.5, 2)$ thus defining 9 regions and fit a separate model for each region. We see that the parameter is indeed rather constant over the 9 regions. A formal test can be constructed.

```
 udp <- piecewise.twostage(c(0,0.5,2),data=d,score.method="
     optimize",
             id="cluster",timevar="time",
             status="status",model="clayton.oakes",silent=0)
 summary(udp)
```

```
Data-set  1 out of  4
  Number of joint events: 518 of  2000

Data-set  2 out of  4
  Number of joint events: 274 of  1232

Data-set  3 out of  4
```

```
   Number of joint events: 247 of  1203


Data-set  4 out of  4
  Number of joint events: 594 of  953
[1] 1
Dependence parameter for Clayton-Oakes model
Score of log-likelihood for parameter estimates (too large?)
            0 - 0.5     0.5 - 2
0 - 0.5 0.0019673733 0.001451886
0.5 - 2 0.0007297083 0.003142920


log-coefficient for dependence parameter (SE)
         0 - 0.5         0.5 - 2
0 - 0.5  0.687 (0.069)  0.677 (0.093)
0.5 - 2  0.733 (0.100)  0.718 (0.060)

Kendall's tau (SE)
         0 - 0.5         0.5 - 2
0 - 0.5  0.498 (0.017)  0.496 (0.023)
0.5 - 2  0.51  (0.025)  0.506 (0.015)
```

Multivariate gamma twostage models

To illustrate how the multivariate models can be used, we first set up some twin data with ACE structure. That is two shared random effects, one being the genes σ_g^2 and one the environmental effect σ_e^2. Monozygotic twins share all genes whereas the dizygotic twins only share half the genes. This can be expressed via 5 random effect for each twin pair (for example). We start by setting this up.

The pardes matrix tells how the the parameters of the 5 random effects are related, and the matrix her first has one random effect with parameter θ_1 (here the σ_g^2), then the next 3 random effects have parameters $0.5\theta_1$ (here $0.5\sigma_g^2$), and the last random effect that is given by its own parameter θ_2 (here σ_e^2).

```
1  data <- simClaytonOakes.twin.ace(2000,2,1,0,3)
2
3  out <- twin.polygen.design(data,id="cluster")
4  pardes <- out$pardes
5  pardes
```

```
     [,1] [,2]
[1,]  1.0    0
[2,]  0.5    0
[3,]  0.5    0
[4,]  0.5    0
[5,]  0.0    1
```

The last part of the model structure is to decide how the random effects are shared for the different pairs (MZ and DZ), this is specified by the random effects design (V_1 and V_2) for each pair. This is here specified by an overall designmatrix for each subject (since they enter all pairs with the same random effects design).

For an MZ pair the two share the full gene random effect and the full environmental random effect. In contrast the DZ pairs share the 2nd random effect with half the gene-variance and have both a

non-shared gene-random effect with half the variance, and finally a
fully shared environmental random effect.

```
1  des.rv <- out$des.rv
2  # MZ
3  head(des.rv,2)
4  # DZ
5  tail(des.rv,2)
```

```
  MZ DZ DZns1 DZns2 env
1  1  0     0     0   1
2  1  0     0     0   1
     MZ DZ DZns1 DZns2 env
3999  0  1     1     0   1
4000  0  1     0     1   1
```

Now we call the twostage function. We see that we essentially
recover the true values, and note that the output also compares
the sizes of the genetic and environmental random effect. This
number is sometimes called the heritability. In addition the total
variance for each subject is also computed and is here around 3, as
we indeed constructed.

```
1  aa <- phreg(Surv(time,status)~x+cluster(cluster),data=data)
2  ts <- twostage(aa,data=data,clusters=data$cluster,detail=0,
3         theta=c(2,1),var.link=0,step=0.5,
4         random.design=des.rv,theta.des=pardes)
5  summary(ts)
```

```
Dependence parameter for Clayton-Oakes model
Variance of Gamma distributed random effects
$estimates
              Coef.        SE        z       P-val Kendall tau         SE
dependence1 2.2111163 0.2219525 9.962113 0.000000e+00   0.5250665 0.02503201
dependence2 0.7431872 0.1706430 4.355217 1.329349e-05   0.2709211 0.04535315

$type
[1] "clayton.oakes"

$h
           Estimate Std.Err   2.5%   97.5%   P-value
dependence1   0.7484 0.05898 0.6328 0.8640 6.669e-37
dependence2   0.2516 0.05898 0.1360 0.3672 1.995e-05

$vare
NULL

$vartot
   Estimate Std.Err  2.5% 97.5%   P-value
p1    2.954  0.1426 2.675 3.234 2.514e-95

attr(,"class")
[1] "summary.mets.twostage"
```

The estimates can be transformed into Kendall's tau estimates for
MZ and DZ twins. The Kendall's tau in the above output reflects
how a gamma distributed random effect in the normal Clayton-
Oakes model is related to the Kendall's tau. In this setting the
Kendall's of MZ and DZ, however, should reflect both random
effects.

We do this based on simulations. The Kendall's tau of the MZ is around 0.60, and for DZ around 0.33. Both are quite high and this is due to a large shared environmental effect and large genetic effect.

```
1  kendall.ClaytonOakes.twin.ace(ts$theta[1],ts$theta[2],K
       =10000)
```

```
$mz.kendall
[1] 0.5984888

$dz.kendall
[1] 0.3257834
```

Family data

For family data, things are quite similar since we use only the pair-wise structure. We show how the designs are specified.

First we simulate data from an ACE model. 2000 families with two-parents that share only the environment, and two-children that share genes with their parents.

```
1  library(mets)
2  set.seed(1000)
3  data <- simClaytonOakes.family.ace(2000,2,1,0,3)
4  head(data)
5  data$number <- c(1,2,3,4)
6  data$child <- 1*(data$number==3)
```

```
        time status x cluster   type    mintime lefttime truncated
1 0.26343780      1 1       1 mother 0.26343780        0         0
2 1.14490828      1 1       1 father 0.26343780        0         0
3 0.86649229      1 1       1  child 0.26343780        0         0
4 0.30843425      1 0       1  child 0.26343780        0         0
5 3.00000000      0 0       2 mother 0.07739746        0         0
6 0.07739746      1 0       2 father 0.07739746        0         0
```

To set up the random effects some functions can be used. We here set up the ACE model that has 9 random effects with one shared environmental effect (the last random effect) and 4 genetic random effects for each parent, with variance $\sigma_g^2/4$.

The random effect is again set-up with an overall designmatrix because it is again the same for each subject for all comparisons across family members. We below demonstrate how the model can be specified in various other ways.

Each child share 2 genetic random effects with each parent, and also share 2 genetic random effects with his/her sibling.

```
1  out <- ace.family.design(data,member="type",id="cluster")
2  out$pardes
3  head(out$des.rv,4)
```

```
      [,1] [,2]
[1,] 0.25    0
[2,] 0.25    0
[3,] 0.25    0
```

```
 [4,] 0.25    0
 [5,] 0.25    0
 [6,] 0.25    0
 [7,] 0.25    0
 [8,] 0.25    0
 [9,] 0.00    1
      m1 m2 m3 m4 f1 f2 f3 f4 env
[1,]  1  1  1  1  0  0  0  0   1
[2,]  0  0  0  0  1  1  1  1   1
[3,]  1  1  0  0  1  1  0  0   1
[4,]  1  0  1  0  1  0  1  0   1
```

Then we fit the model

```
1  pa <- phreg(Surv(time,status)~+1+cluster(cluster),data=data)
2  aa <- aalen(Surv(time,status)~+1,data=data,robust=0)
3
4  # make ace random effects design
5  ts <- twostage(pa,data=data,clusters=data$cluster,
6      var.par=1,var.link=0,theta=c(2,1),
7      random.design=out$des.rv,theta.des=out$pardes)
8  summary(ts)
```

```
Dependence parameter for Clayton-Oakes model
Variance of Gamma distributed random effects
$estimates
                Coef.          SE         z P-val Kendall tau          SE
dependence1  2.185967 0.19164670 11.40623      0   0.5222132 0.02187458
dependence2  0.947110 0.07648339 12.38321      0   0.3213691 0.01761183


$type
[1] "clayton.oakes"


$h
            Estimate Std.Err   2.5%   97.5%    P-value
dependence1   0.6977 0.02997 0.6390 0.7564 7.182e-120
dependence2   0.3023 0.02997 0.2436 0.3610  6.359e-24


$vare
NULL


$vartot
   Estimate Std.Err  2.5% 97.5%   P-value
p1    3.133  0.1737 2.793 3.474 1.074e-72


attr(,"class")
[1] "summary.mets.twostage"
```

The model can also be fitted by specifying the pairs that one
wants for the pairwise likelhood. This is done by specifying the
pairs argument. We start by considering all pairs as we also did
before.

All pairs can be written up by calling the familycluster.index
function.

There are 12000 pairs to consider and the last 12 pairs for the last
family is written out here.

```
1  # now specify fitting via specific pairs
2  # first all pairs
3  mm <- familycluster.index(data$cluster)
4  head(mm$familypairindex,n=10)
5  pairs <- matrix(mm$familypairindex,ncol=2,byrow=TRUE)
6  tail(pairs,n=12)
```

```
[1] 1 2 1 3 1 4 2 3 2 4
          [,1] [,2]
[11989,] 7993 7994
[11990,] 7993 7995
[11991,] 7993 7996
[11992,] 7994 7995
[11993,] 7994 7996
[11994,] 7995 7996
[11995,] 7997 7998
[11996,] 7997 7999
[11997,] 7997 8000
[11998,] 7998 7999
[11999,] 7998 8000
[12000,] 7999 8000
```

Then fitting the model using only specified pairs

```
1  ts <- twostage(pa,data=data,clusters=data$cluster,
2          theta=c(2,1),var.link=0,step=1.0,
3          random.design=out$des.rv,
4          theta.des=out$pardes,pairs=pairs)
5  summary(ts)
```

```
Dependence parameter for Clayton-Oakes model
Variance of Gamma distributed random effects
$estimates
               Coef.         SE        z P-val Kendall tau        SE
dependence1 2.185967 0.18986604 11.51321     0   0.5222132 0.02167133
dependence2 0.947110 0.07929082 11.94476     0   0.3213691 0.01825829


$type
[1] "clayton.oakes"


$h
           Estimate Std.Err   2.5%  97.5%    P-value
dependence1   0.6977 0.03044 0.6381 0.7574 2.659e-116
dependence2   0.3023 0.03044 0.2426 0.3619  3.010e-23


$vare
NULL


$vartot
   Estimate Std.Err  2.5% 97.5%   P-value
p1    3.133  0.1713 2.797 3.469 1.057e-74


attr(,"class")
[1] "summary.mets.twostage"
```

Now we only use a random sample of the pairs by sampling these. The pairs picked still refers to the data given in the data argument, and clusters (families) are also specified as before.

```
1  ssid <- sort(sample(1:12000,2000))
2  tsd <- twostage(aa,data=data,clusters=data$cluster,
3           theta=c(2,1)/10,var.link=0,step=1.0,
4           random.design=out$des.rv,iid=1,
5          theta.des=out$pardes,pairs=pairs[ssid,])
6  summary(tsd)
```

```
Dependence parameter for Clayton-Oakes model
Variance of Gamma distributed random effects
$estimates
               Coef.         SE        z        P-val Kendall tau         SE
dependence1 2.001187 0.3400847 5.884378 3.995533e-09   0.5001483 0.04248537
```

```
dependence2 1.054030 0.1277271 8.252205 2.220446e-16    0.3451275 0.02738838

$type
[1] "clayton.oakes"

$h
            Estimate Std.Err    2.5%  97.5%  P-value
dependence1    0.655 0.05345 0.5502 0.7598 1.606e-34
dependence2    0.345 0.05345 0.2402 0.4498 1.089e-10

$vare
NULL

$vartot
   Estimate Std.Err  2.5% 97.5%  P-value
p1    3.055  0.3253 2.418 3.693 5.923e-21

attr(,"class")
[1] "summary.mets.twostage"
```

Sometimes one only has the data from the pairs in addition to for example a cohort estimate of the marginal surival models. We now demonstrate how this is dealt with. Everything is essentially as before but need to organize the design differently compared to before we specified the design for everybody in the cohort.

```
1   ids <- sort(unique(c(pairs[ssid,])))
2
3   pairsids <- c(pairs[ssid,])
4   pair.new <- matrix(fast.approx(ids,c(pairs[ssid,])),ncol=2)
5   head(pair.new)
6
7   # this requires that pair.new refers to id's in dataid
         (survival, status and so forth)
8   # random.design and theta.des are constructed to be the
         array 3 dims via individual specfication from
         ace.family.design
9   dataid <- dsort(data[ids,],"cluster")
10  outid <- ace.family.design(dataid,member="type",id="cluster"
         )
11  outid$pardes
12  head(outid$des.rv)
```

```
     [,1] [,2]
[1,]    1    2
[2,]    3    4
[3,]    3    5
[4,]    4    6
[5,]    7    8
[6,]    9   10
      [,1] [,2]
 [1,] 0.25    0
 [2,] 0.25    0
 [3,] 0.25    0
 [4,] 0.25    0
 [5,] 0.25    0
 [6,] 0.25    0
 [7,] 0.25    0
 [8,] 0.25    0
 [9,] 0.00    1
     m1 m2 m3 m4 f1 f2 f3 f4 env
[1,]  1  1  1  1  0  0  0  0   1
[2,]  0  0  0  0  1  1  1  1   1
[3,]  1  1  1  1  0  0  0  0   1
```

```
[4,]  0  0  0  0  1  1  1  1  1
[5,]  1  1  0  0  1  1  0  0  1
[6,]  1  0  1  0  1  0  1  0  1
```

Now fitting the model using only the pair data.

```
1  tsdid <- twostage(aa,data=dataid,clusters=dataid$cluster,
2          theta=c(2,1)/10,var.link=0,step=1.0,
3          random.design=outid$des.rv,iid=1,
4          theta.des=outid$pardes,pairs=pair.new)
5  summary(tsdid)
6  coef(tsdid)
7  coef(tsd)
```

```
Dependence parameter for Clayton-Oakes model
Variance of Gamma distributed random effects
$estimates
              Coef.        SE        z       P-val Kendall tau         SE
dependence1 2.001187 0.3400847 5.884378 3.995533e-09   0.5001483 0.04248537
dependence2 1.054030 0.1277271 8.252205 2.220446e-16   0.3451275 0.02738838

$type
[1] "clayton.oakes"

$h
           Estimate Std.Err   2.5%   97.5%  P-value
dependence1    0.655 0.05345 0.5502 0.7598 1.606e-34
dependence2    0.345 0.05345 0.2402 0.4498 1.089e-10

$vare
NULL

$vartot
   Estimate Std.Err  2.5% 97.5%   P-value
p1    3.055  0.3253 2.418 3.693 5.923e-21

attr(,"class")
[1] "summary.mets.twostage"
              Coef.        SE        z       P-val Kendall tau         SE
dependence1 2.001187 0.3400847 5.884378 3.995533e-09   0.5001483 0.04248537
dependence2 1.054030 0.1277271 8.252205 2.220446e-16   0.3451275 0.02738838
              Coef.        SE        z       P-val Kendall tau         SE
dependence1 2.001187 0.3400847 5.884378 3.995533e-09   0.5001483 0.04248537
dependence2 1.054030 0.1277271 8.252205 2.220446e-16   0.3451275 0.02738838
```

Now we illustrate how one can also directly specify the random.design and theta.design for each pair, rather than taking the rows of the des.rv for the relevant pairs. This can be much simpler in some situations.

```
1  pair.types <- matrix(dataid[c(t(pair.new)),"type"],byrow=T,
       ncol=2)
2  head(pair.new)
3  head(pair.types)
4
5  # here makes pairwise design , simpler random.design og
       pardes, parameters
6  # stil varg, varc
7  # mother, child, share half rvm=c(1,1,0) rvc=c(1,0,1),
8  # thetadesmcf=rbind(c(0.5,0),c(0.5,0),c(0.5,0),c(0,1))
9  #
10 # father, child, share half rvf=c(1,1,0) rvc=c(1,0,1),
11 # thetadescf=rbind(c(0.5,0),c(0.5,0),c(0.5,0),c(0,1))
```

```
12   #
13   # child, child, share half rvc=c(1,1,0) rvc=c(1,0,1),
14   # thetadesmf=rbind(c(0.5,0),c(0.5,0),c(0.5,0),c(0,1))
15   #
16   # mother, father, share 0 rvm=c(1,0) rvf=c(0,1),
17   # thetadesmf=rbind(c(1,0),c(1,0),c(0,1))
18
19   theta.des <- array(0,c(4,2,nrow(pair.new)))
20   random.des <- array(0,c(2,4,nrow(pair.new)))
21   # random variables in each pair
22   rvs <- c()
23   for (i in 1:nrow(pair.new))
24   {
25       if (pair.types[i,1]=="mother" & pair.types[i,2]=="father"
            )
26       {
27       theta.des[,,i] <- rbind(c(1,0),c(1,0),c(0,1),c(0,0))
28       random.des[,,i] <- rbind(c(1,0,1,0),c(0,1,1,0))
29       rvs <- c(rvs,3)
30       } else {
31       theta.des[,,i] <- rbind(c(0.5,0),c(0.5,0),c(0.5,0),c(0,1)
            )
32       random.des[,,i] <- rbind(c(1,1,0,1),c(1,0,1,1))
33       rvs <- c(rvs,4)
34       }
35   }
36   # 3 rvs here
37   random.des[,,7]
38   theta.des[,,7]
39   # 4 rvs here
40   random.des[,,1]
41   theta.des[,,1]
42   head(rvs)
```

```
     [,1] [,2]
[1,]    1    2
[2,]    3    4
[3,]    3    5
[4,]    4    6
[5,]    7    8
[6,]    9   10
     [,1]     [,2]
[1,] "mother" "father"
[2,] "mother" "father"
[3,] "mother" "child"
[4,] "father" "child"
[5,] "child"  "child"
[6,] "child"  "child"
     [,1] [,2] [,3] [,4]
[1,]    1    1    0    1
[2,]    1    0    1    1
     [,1] [,2]
[1,]  0.5    0
[2,]  0.5    0
[3,]  0.5    0
[4,]  0.0    1
     [,1] [,2] [,3] [,4]
[1,]    1    0    1    0
[2,]    0    1    1    0
     [,1] [,2]
[1,]    1    0
```

```
[2,]    1    0
[3,]    0    1
[4,]    0    0
[1] 3 3 4 4 4 4
```

And fitting again the same model as before

```
1  tsdid2 <- twostage(aa,data=dataid,clusters=dataid$cluster,
2        theta=c(2,1)/10,var.link=0,step=1.0,
3        random.design=random.des,
4        theta.des=theta.des,pairs=pair.new,pairs.rvs=rvs)
5  summary(tsdid2)
6  tsd$theta
7  tsdid2$theta
8  tsdid$theta
```

```
Dependence parameter for Clayton-Oakes model
Variance of Gamma distributed random effects
$estimates
               Coef.        SE        z       P-val Kendall tau         SE
dependence1 2.001187 0.3400847 5.884378 3.995533e-09   0.5001483 0.04248537
dependence2 1.054030 0.1277271 8.252205 2.220446e-16   0.3451275 0.02738838

$type
[1] "clayton.oakes"

$h
           Estimate Std.Err   2.5%   97.5%   P-value
dependence1    0.655 0.05345 0.5502 0.7598 1.606e-34
dependence2    0.345 0.05345 0.2402 0.4498 1.089e-10

$vare
NULL

$vartot
   Estimate Std.Err  2.5% 97.5%   P-value
p1    3.055  0.3253 2.418 3.693 5.923e-21

attr(,"class")
[1] "summary.mets.twostage"
                [,1]
dependence1 2.001187
dependence2 1.054030
                [,1]
dependence1 2.001187
dependence2 1.054030
                [,1]
dependence1 2.001187
dependence2 1.054030
```

Finally the same model structure can be setup based on a Kinship coefficient.

```
1  # simpler specification via kinship coefficient for each
       pair
2  kinship <- c()
3  for (i in 1:nrow(pair.new))
4  {
5  if (pair.types[i,1]=="mother" & pair.types[i,2]=="father")
       pk1 <- 0 else pk1 <- 0.5
6  kinship <- c(kinship,pk1)
7  }
8  head(kinship,n=10)
9
```

```
10  out <- make.pairwise.design(pair.new,kinship,type="ace")
11  names(out)
12  # 4 rvs here , here independence since shared component has
        variance 0 !
13  out$random.des[,,9]
14  out$theta.des[,,9]
```

```
 [1] 0.0 0.0 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5
[1] "random.design" "theta.des"     "ant.rvs"
    [,1] [,2] [,3] [,4]
[1,]   1    1    0    1
[2,]   1    0    1    1
    [,1] [,2]
[1,]  0.5    0
[2,]  0.5    0
[3,]  0.5    0
[4,]  0.0    1
```

Same same

```
1  tsdid3 <- twostage(aa,data=dataid,clusters=dataid$cluster,
2          theta=c(2,1)/10,var.link=0,step=1.0,
3          random.design=out$random.design,
4          theta.des=out$theta.des,pairs=pair.new,pairs.rvs=
                out$ant.rvs)
5  summary(tsdid3)
6  coef(tsdid3)
```

```
Dependence parameter for Clayton-Oakes model
Variance of Gamma distributed random effects
$estimates
             Coef.        SE        z       P-val Kendall tau        SE
dependence1 2.001187 0.3400847 5.884378 3.995533e-09   0.5001483 0.04248537
dependence2 1.054030 0.1277271 8.252205 2.220446e-16   0.3451275 0.02738838


$type
[1] "clayton.oakes"


$h
          Estimate Std.Err   2.5%   97.5%   P-value
dependence1    0.655 0.05345 0.5502 0.7598 1.606e-34
dependence2    0.345 0.05345 0.2402 0.4498 1.089e-10


$vare
NULL


$vartot
   Estimate Std.Err 2.5% 97.5%   P-value
p1    3.055  0.3253 2.418 3.693 5.923e-21


attr(,"class")
[1] "summary.mets.twostage"
             Coef.        SE        z       P-val Kendall tau        SE
dependence1 2.001187 0.3400847 5.884378 3.995533e-09   0.5001483 0.04248537
dependence2 1.054030 0.1277271 8.252205 2.220446e-16   0.3451275 0.02738838
```

Univariate plackett model twostage models

The copula known as the Plackett distribution, see [3], is on the form   [3] ; ; and

$$
C(u,v;\theta) = \begin{cases} \dfrac{S-(S^2-4uv\theta(\theta-a))}{2(\theta-1)} & \text{if } \theta \neq 1 \\ uv & \text{if } \theta = 1 \end{cases} \tag{6}
$$

with $S = 1 + (\theta - 1)(u + v)$. With marginals S_i we now define the bivariate survival function as $C(u_1, u_2) = H(S_1(t_1), S_2(t_2))$ with $u_i = S_i(t_i)$.

The dependence parameter θ has the nice interpretation that the it is equivalent to the odds-ratio of all 2×2 tables for surviving past any cut of the plane (t_1, t_2), that is

$$\theta = \frac{P(T_1 > t_1 | T_2 > t_2) P(T_1 \le t_1 | T_2 > t_2)}{P(T_1 > t_1 | T_2 \le t_2) P(T_1 \le t_1 | T_2 \le t_2)}.$$

One additional nice feature of the odds-ratio measure it that it is directly linked to the Spearman correlation, ρ, that can be computed as

$$\frac{\theta + 1}{\theta - 1} - \frac{2\theta}{(\theta - 1)^2} \log(\theta) \tag{7}$$

when $\theta \ne 1$, if $\theta = 1$ then $\rho = 0$.

This model has a more free parameter than the Clayton-Oakes model.

```
1   library(mets)
2   data(diabetes)
3
4   # Marginal Cox model with treat as covariate
5   margph <- phreg(Surv(time,status)~treat+cluster(id),data=
        diabetes)
6   # Clayton-Oakes, MLE
7   fitco1<-twostageMLE(margph,data=diabetes,theta=1.0)
8   summary(fitco1)
9
10  # Plackett model
11  mph <- phreg(Surv(time,status)~treat+cluster(id),data=
        diabetes)
12  fitp <- survival.twostage(mph,data=diabetes,theta=3.0,Nit
        =40,
13         clusters=diabetes$id,var.link=1,model="plackett")
14  summary(fitp)
15
16  # without covariates but with stratafied
17  marg <- phreg(Surv(time,status)~+strata(treat)+cluster(id),
        data=diabetes)
18  fitpa <- survival.twostage(marg,data=diabetes,theta=1.0,
19        clusters=diabetes$id,score.method="optimize")
20  summary(fitpa)
21
22  fitcoa <- survival.twostage(marg,data=diabetes,theta=1.0,
        clusters=diabetes$id,
23        model="clayton.oakes")
24  summary(fitcoa)
```

```
Dependence parameter for Clayton-Oakes model
Variance of Gamma distributed random effects
$estimates
            Coef.       SE       z      P-val Kendall tau       SE
dependence1 0.9526614 0.3543033 2.68883 0.007170289   0.322645 0.08127892


$type
```

```
NULL

attr(,"class")
[1] "summary.mets.twostage"
Dependence parameter for Odds-Ratio (Plackett) model
With log-link
$estimates
            log-Coef.        SE        z       P-val Spearman Corr.        SE
dependence1   1.14188 0.2784057 4.101497 4.104867e-05      0.3648217 0.08158643


$or
            Estimate Std.Err 2.5% 97.5%   P-value
dependence1    3.133  0.8721 1.423 4.842 0.0003283


$type
[1] "plackett"

attr(,"class")
[1] "summary.mets.twostage"
Dependence parameter for Clayton-Oakes model
Variance of Gamma distributed random effects
With log-link
$estimates
              log-Coef.        SE         z     P-val Kendall tau        SE
dependence1 -0.05683487 0.3239422 -0.1754476 0.8607279   0.3208252 0.07058583


$vargam
            Estimate Std.Err   2.5% 97.5%  P-value
dependence1   0.9448   0.306 0.3449 1.545 0.002022


$type
[1] "clayton.oakes"

attr(,"class")
[1] "summary.mets.twostage"
Dependence parameter for Clayton-Oakes model
Variance of Gamma distributed random effects
With log-link
$estimates
              log-Coef.        SE        z     P-val Kendall tau        SE
dependence1 -0.05683996 0.3322322 -0.171085 0.8641569   0.3208241 0.07239207


$vargam
            Estimate Std.Err   2.5% 97.5%  P-value
dependence1   0.9447  0.3139 0.3296  1.56 0.002613


$type
[1] "clayton.oakes"

attr(,"class")
[1] "summary.mets.twostage"
```

With a regression design

```
1  mm <- model.matrix(~-1+factor(adult),diabetes)
2  fitp <- survival.twostage(mph,data=diabetes,theta=3.0,Nit
       =40,
3         clusters=diabetes$id,var.link=1,model="plackett",
4         theta.des=mm)
5  summary(fitp)
```

```
Dependence parameter for Odds-Ratio (Plackett) model
With log-link
$estimates
                log-Coef.        SE        z       P-val Spearman Corr.
factor(adult)1   1.098333 0.3436264 3.196298 0.001392032      0.3519988
```

```
factor(adult)2  1.231962 0.4938132 2.494794 0.012603018      0.3909505
                    SE
factor(adult)1 0.1016635
factor(adult)2 0.1417283

$or
             Estimate Std.Err  2.5% 97.5%  P-value
factor(adult)1    2.999   1.031 0.9792 5.019 0.003613
factor(adult)2    3.428   1.693 0.1102 6.746 0.042861

$type
[1] "plackett"

attr(,"class")
[1] "summary.mets.twostage"
```

```
1  # Piecewise constant cross hazards ratio modelling
2
3  d <- subset(simClaytonOakes(2000,2,0.5,0,stoptime=2,left=0),
       !truncated)
4  udp <- piecewise.twostage(c(0,0.5,2),data=d,score.method="
       optimize",
5               id="cluster",timevar="time",
6               status="status",model="plackett",silent=0)
7  summary(udp)
```

```
Data-set  1 out of  4
  Number of joint events: 529 of  2000

Data-set  2 out of  4
  Number of joint events: 248 of  1212

Data-set  3 out of  4
  Number of joint events: 254 of  1219

Data-set  4 out of  4
  Number of joint events: 633 of  960
[1] 1
Dependence parameter for Plackett model
log-coefficient for dependence parameter (SE)
        0 - 0.5        0.5 - 2
0 - 0.5  1.761 (0.083)  1.628 (0.128)
0.5 - 2  1.74  (0.128)  2.017 (0.092)

Spearman Correlation (SE)
        0 - 0.5        0.5 - 2
0 - 0.5  0.532 (0.020)  0.499 (0.033)
0.5 - 2  0.527 (0.032)  0.593 (0.021)
```