Correlated bivariate Normal competing risks

– simulation findings in an ill-posed problem

Malcolm Hudson[1,2] Valerie Gares [3] Maurizio Manuguerra[1] Val Gebski[2]

ISCB ASC 2018, Melbourne

[1]Macquarie University

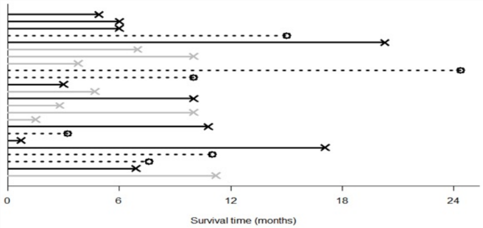
[2]NHMRC Clinical Trials Centre
University of Sydney

[3]INSA Rennes

Outline i

Competitive risks

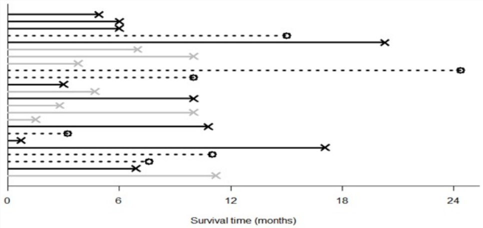
- Study time to several events in competition (24 patients):
  - Myocardial infarction (MI) ———× (10)
  - Death from other causes ———× (8)
  - Censoring $\cdots\circ$ (6)



- Patient status:
  Myocardial infarction
  Withdraw from study
  Death from other causes

Competitive risks

- Study time to several events in competition (24 patients):
  - Myocardial infarction (MI) ———× (10)
  - Death from other causes ———× (8)
  - Censoring $\cdots \circ$ (6)



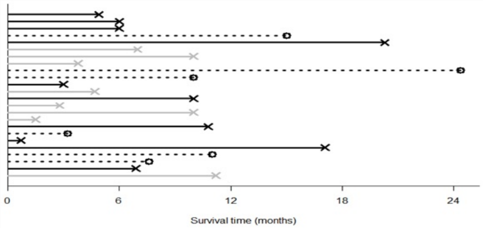
- Patient status:

Myocardial infarction	$\Rightarrow$	Event of interest
Withdraw from study	$\Rightarrow$	Censoring
Death from other causes	$\Rightarrow$	Censoring

Introduces dependence between censoring and event

Competitive risks

- Study time to several events in competition (24 patients):
  - Myocardial infraction (MI) ——— × (10)
  - Death from other causes ——— × (8)
  - Censoring $\cdots\circ$ (6)



- Patient status:

Myocardial infarction	$\Rightarrow$	Event of interest
Withdraw from study	$\Rightarrow$	Censoring
Death from other causes	$\Rightarrow$	**Competing event**

Respects independent censoring assumption in SVA

Bivariate Normal Censored Linear Model

Correlated competing risks model

- Bivariate Normal Censored Linear Model (bnc lm)
  - Log-Time to *first occurring* event
  - Two events: $Y \sim \mathrm{BVN}(\mu, \Sigma)$
  - lm: $\mu = X\beta$
  - *bnc* lm for observed data: $y = \min(Y_1, Y_2, C)$ and $D = 1, 2, 0$
  - observations (y, D) comprise *first-event data*
- Estimation
  - ML solution
  - ML $\leftarrow$ MPL
  - EM algorithm uses imputed times to non-occurring event
- Imputation
  - Conditional moments $\mathbb{E}(Y_1 | Y_1 > y, Y_2 = y,\ D = 2)$
  - $\mathbb{E}(Y_1 | Y_1 > y, Y_2 > y,\ D = 0)$
  - $\mathbb{E}(Y_1^2 | Y_1 > y, Y_2 = y,\ D = 2)$
  - ... *etc*

Simulation Aim, Methods

Simulation Goals

- Aim: parametric estimation of first-event data from correlated BVN competing risks
  - Sampling distribution (bias, variance) of MPL estimation of BVN parameters in one- and two- sample datasets of first-event times
  - Sampling distribution of mean difference estimated between Treated and Control subjects

Methods

- One- and two-sample simulations
  - BVN and *non*-Normal copula (with Normal margins)
  - Parameters varying
    i. Sample size (100, 500, 1000)
    ii. Mean difference event-of-interest to competing-event (0, 0.5, 1)
    iii. Event probability (complement *censoring fraction event 1*)
    $\Rightarrow$ prob event-of-interest occurs before end of study (`eof`)
    - Surviving fractions (1, 0.8, 0.6)
    - Competing event censoring *not* included

    iv. Correlation $\rho \in (-0.5, -0.25, 0, 0.25, 0.5)$
    v. Treatment benefit (mean time difference between *Treated and Controls, 2 sample only*)
- Parallel simulation computation
  - **simsalapar** R package Hofert and Mächler (2016)
- Fixed-point accelerated EM iterations Bobb and Varadhan (2014)
- Estimation conducted in our R package **bnc**

Results

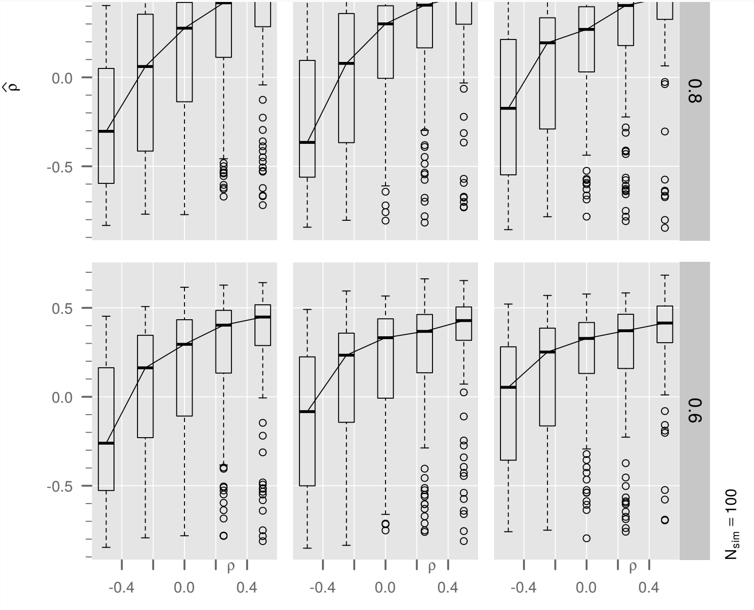
1-sample estimation of correlation

squareEM iterations (n=1000)

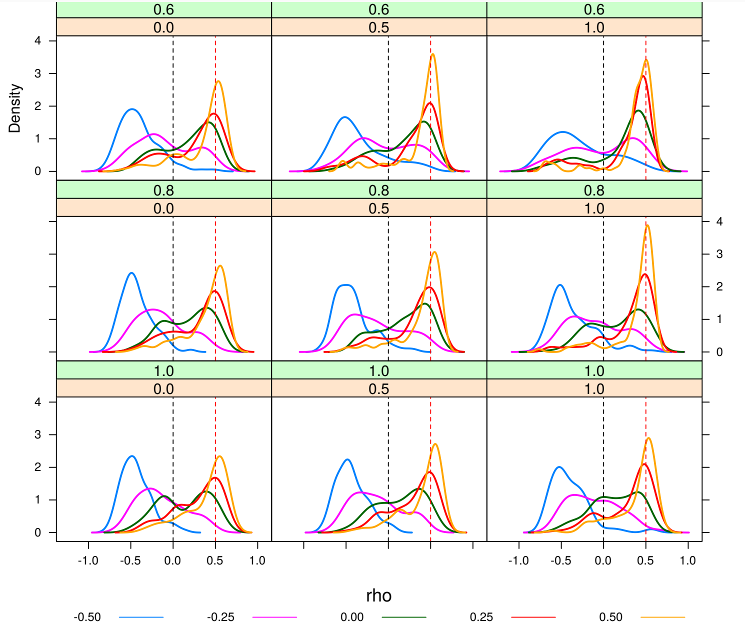
cs	ρ | beta2	0	0.5	1
1	-0.5	36	95	93
	-0.25	83	107	154
	0	107	142	250
	0.25	90	127	250
	0.5	112	227	250
0.8	-0.5	36	90	116
	-0.25	88	108	243
	0	119	163	250
	0.25	146	219	250
	0.5	141	248	250
0.6	-0.5	78	105	250
	-0.25	117	250	235
	0	134	212	250
	0.25	159	240	250
	0.5	184	250	250

max number iterations to converge, nsim=100

Correlation (n=100)



Correlation, n=500 (Density plot)



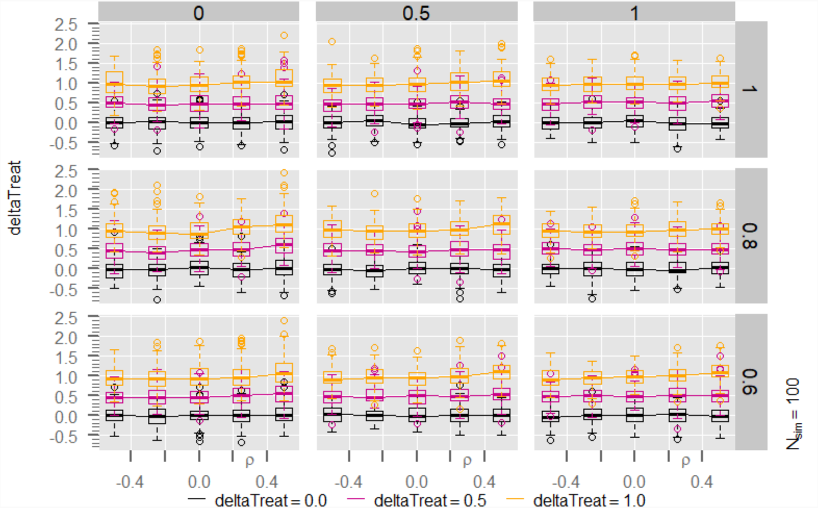
2-sample estimation of treatment benefit

Treatment Benefit $\hat{\Delta} = \hat{\beta}_{21}$ (n=1000)

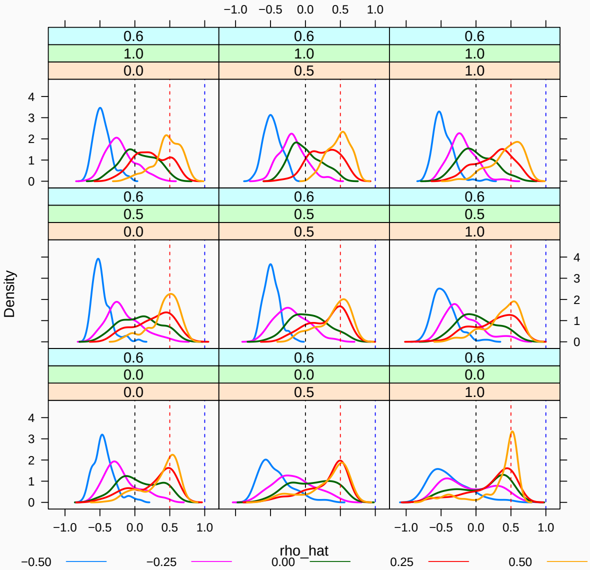
	beta12	0		1	
cs	ρ | deltaTreat	0	0.5	0	0.5
1	-0.5	-0.00	0.50	0.00	0.51
	-0.25	-0.00	0.48	-0.01	0.50
	0	0.00	0.46	0.00	0.50
	0.25	-0.01	0.49	0.00	0.49
	0.5	-0.00	0.52	-0.00	0.49
0.8	-0.5	-0.00	0.50	0.00	0.50
	-0.25	-0.00	0.48	0.01	0.49
	0	0.01	0.49	0.00	0.49
	0.25	-0.00	0.48	-0.01	0.48
	0.5	-0.01	0.51	0.01	0.51
0.6	-0.5	0.02	0.49	-0.01	0.48
	-0.25	-0.02	0.52	-0.00	0.48
	0	-0.01	0.49	0.01	0.48
	0.25	0.02	0.49	0.01	0.50
	0.5	0.00	0.51	-0.01	0.51

medians of 100 replicates

Treatment Benefit, n=100 (estimates)



Treatment Benefit, n=1000, sf=0.6 (density plots)



Follow-up

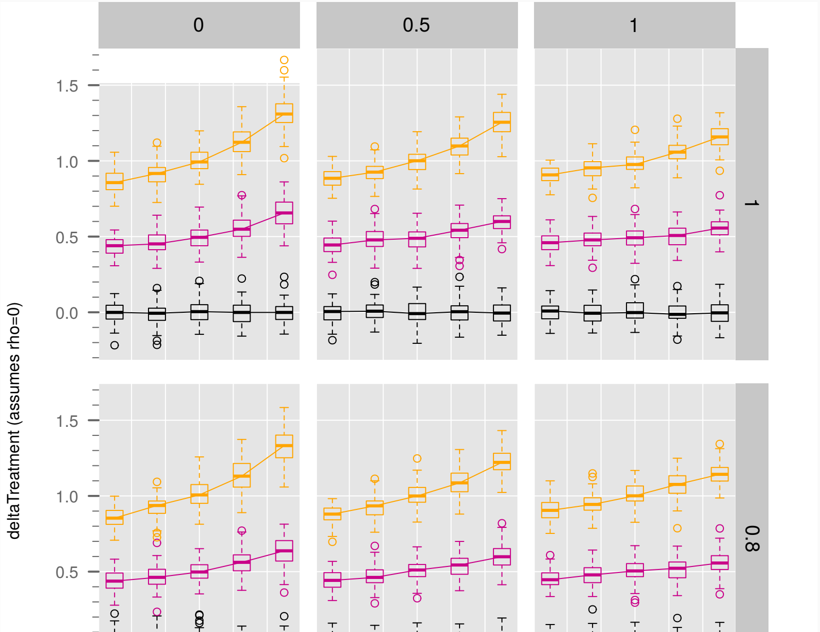
Fixed correlation, a la GEE

- MPL solution with rho fixed
  - [?] (R package **bnc**)

Fixed correlation, rationale

- Estimating ρ is unstable $\Rightarrow$ fix its value a priori (to 0).
- $\rho = 0$ is convenient
  - Competing event $\rightarrow$ (further) independent censoring
  - Fit regression coefficients of event 1 in `survival::survreg`
  - Generalized Estimating Equations (GEEs)
    - Consistency/ unbiasedness to estimating regression coefficients?
    - Likelihood scores with assumed covariance structure

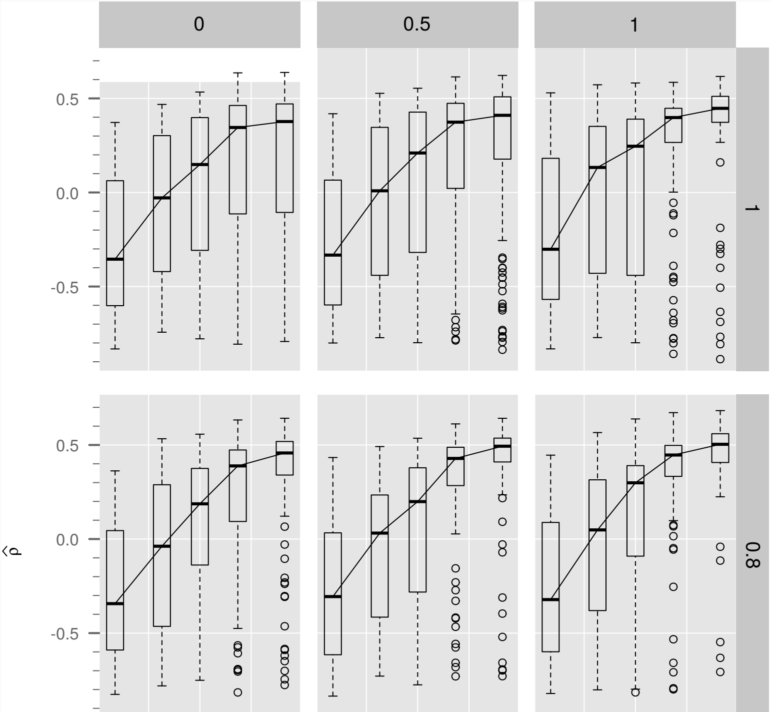
survreg coef of treatment benefit assuming $\rho = 0$, n=100



Robustness to non-Normality

Frank copula

- We set standard Normal *marginals*
- Association parameter, Frank θ
  - $\theta \to \tau \to \rho$
- *Joint* distribution is non- BVN



0
0.5
1
1
0.8
0.5
0.0
-0.5
0.5
0.0
-0.5
$\hat{\rho}$

Summing up

So you want to fit BVN to two competing risks?

- Optimize to maximize Likelihood (*routine*)
- EM algorithm (*safety first*)
- Mildly Penalize Likelihood, adjust EM alg for MPL
- Accelerate EM (Aitken via squareEM)
- Change start-point: initial small number of EM iterations
- R-package **bnc**

What could possibly go wrong?

- Singular Hessian, failure to remain in $-1 < \rho < 1$
- i. Be prepared to wait (forever) for convergence;
  ii. Convergence failure possible when $\hat{\rho} = +1$ or -1
- Still very slow, but sure, convergence
- Very fast, but can fail to converge

So you want to fit BVN to two competing risks?

- Optimize to maximize Likelihood (*routine*)
- EM algorithm (*safety first*)
- Mildly Penalize Likelihood, adjust EM alg for MPL
- Accelerate EM (Aitken via squareEM)
- Change start-point: initial small number of EM iterations
- R-package **bnc**

What could possibly go wrong?

- Singular Hessian, failure to remain in $-1 < \rho < 1$
- i. Be prepared to wait (forever) for convergence;

  ii. Convergence failure possible when $\hat{\rho} = +1$ or -1
- Still very slow, but sure, convergence
- Very fast, but can fail to converge

So you want to fit BVN to two competing risks?

- Optimize to maximize Likelihood (*routine*)
- EM algorithm (*safety first*)
- Mildly Penalize Likelihood, adjust EM alg for MPL
- Accelerate EM (Aitken via squareEM)
- Change start-point: initial small number of EM iterations
- R-package **bnc**

What could possibly go wrong?

- Singular Hessian, failure to remain in $-1 < \rho < 1$
- i. Be prepared to wait (forever) for convergence;
   ii. Convergence failure possible when $\hat{\rho} = +1$ or -1
- Still very slow, but sure, convergence
- Very fast, but can fail to converge

So you want to fit BVN to two competing risks?

- Optimize to maximize Likelihood (*routine*)
- EM algorithm (*safety first*)
- Mildly Penalize Likelihood, adjust EM alg for MPL
- Accelerate EM (Aitken via squareEM)
- Change start-point: initial small number of EM iterations
- R-package **bnc**

What could possibly go wrong?

- Singular Hessian, failure to remain in $-1 < \rho < 1$
-
  i. Be prepared to wait (forever) for convergence;
  ii. Convergence failure possible when $\hat{\rho} = +1$ or -1
- Still very slow, but sure, convergence
- Very fast, but can fail to converge

So you want to fit BVN to two competing risks?

- Optimize to maximize Likelihood (*routine*)
- EM algorithm (*safety first*)
- Mildly Penalize Likelihood, adjust EM alg for MPL
- Accelerate EM (Aitken via squareEM)
- Change start-point: initial small number of EM iterations
- R-package **bnc**

- Singular Hessian, failure to remain in $-1 < \rho < 1$
- i. Be prepared to wait (forever) for convergence;
  ii. Convergence failure possible when $\hat{\rho} = +1$ or -1
- Still very slow, but sure, convergence
- Very fast, but can fail to converge

So you want to fit BVN to two competing risks?

- Optimize to maximize Likelihood (*routine*)
- EM algorithm (*safety first*)
- Mildly Penalize Likelihood, adjust EM alg for MPL
- Accelerate EM (Aitken via squareEM)
- Change start-point: initial small number of EM iterations
- R-package **bnc**

- Singular Hessian, failure to remain in $-1 < \rho < 1$
- i. Be prepared to wait (forever) for convergence;
  ii. Convergence failure possible when $\hat{\rho} = +1$ or -1
- Still very slow, but sure, convergence
- Very fast, but can fail to converge

Conclusion

- We fit a BVN censored model for competing risks using a novel EM algorithm (package **bnc**).
- Despite the ill-posed model, estimation is feasible.
  Estimates:

ρ:	correlation estimates	highly variable ($n = 1000$)
β:	regression coefficients	precise (even for $n = 100$)

- Regression coefficients provide estimates of treatment benefit in a parametric AFT competing risks survival model.
- Outlook
  - Need further comparison of estimates of B fixing ρ (e.g. assuming $\rho = 0$) with MPL estimates of the BNC model.
  - Asymptotic properties of this GEE-like procedure warrant investigation.

References

Aitkin, M. 1981. "A Note on the Regression Analysis of Censored Data." *Technometrics* 23 (2). American Statistical Association; American Society for Quality: 161–63. http://www.jstor.org/stable/1268032.

Bobb, Jennifer F., and Ravi Varadhan. 2014. *TurboEM: A Suite of Convergence Acceleration Schemes for EM, MM and Other Fixed-Point Algorithms*. https://CRAN.R-project.org/package=turboEM.

Crowder, M. 1991. "On the Identifiability Crisis in Competing Risks Analysis." *Scandinavian Journal of Statistics* 18 (3): 223–33. https://doi.org/10.2307/4616205.

Hofert, Marius, and Martin Mächler. 2016. "Parallel and Other Simulations in R Made Easy: An End-to-End Study." *Journal of*

Statistical Software 69 (4): 1–44.
https://doi.org/10.18637/jss.v069.i04.

Jeong, Jong-Hyeon, and Jason P. Fine. 2007. "Parametric Regression on Cumulative Incidence Function." *Biostatistics* 8 (2): 184.
https://doi.org/10.1093/biostatistics/kxj040.

correlation: $\hat{\rho}$ (n=1000)

cs	ρ | beta2	0	0.5	1
1	-0.5	-0.48	-0.49	-0.49
	-0.25	-0.22	-0.23	-0.21
	0	0.04	0.04	0.05
	0.25	0.32	0.32	0.34
	0.5	0.49	0.47	0.48
0.8	-0.5	-0.48	-0.48	-0.47
	-0.25	-0.22	-0.21	-0.19
	0	0.08	0.09	0.13
	0.25	0.32	0.37	0.35
	0.5	0.48	0.48	0.48
0.6	-0.5	-0.48	-0.47	-0.46
	-0.25	-0.24	-0.20	-0.09
	0	0.10	0.13	0.22
	0.25	0.35	0.38	0.39
	0.5	0.48	0.48	0.47

medians of 100 replicates