Presentation to VicBiostat

Malcolm Hudson

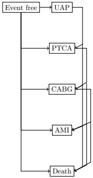
22/06/2017

Moment calculations for correlated BVN variates using Stein's identity

Competing Risks Context of this work

Recent interests in CRs:

- Multi-state modelling of a large RCT
  - LIPID trial, 5 years survival of N=10k women randomized to statin/control
  - intermediate events (interventions, see diagram)



- Factors affecting treatment **recurrence** and death:
  a case study with longitudinal hospital retreatment records
  - recurrent events with observations subject to administrative censoring
  - independence of estimators of cumulative incidence and cumulative mean number of recurrent retreatments from mortality data (when retreatment is highly dependent on mortality)
- Comparz trial: joint analysis of longitudinal patient status and survival
- *Today:* bivariate normal censored data for correlated competing risks

Why study bivariate Normal

Parametric, semi-parametric and non-parametric approaches are univariate

- logNormal survival outcomes with censored outcomes
- Schmee & Hahn, Aitkin (1981), Buckley-James (1979) estimators for censored linear regression

Correlated risks are of interest, but non-identifiability of joint survival time distribution with 2 competing risks

- an "identifiability crisis" (Crowder 1991)
- response was development of many semi-parametric models and Jeong-Fine fully parametric model (Jeong and Fine 2007; see also Tai et al. 2008)

Our goal

- study performance of estimates β, ρ in an ill-posed problem
- sensitivity analysis to:
  - correlation ρ;
  - parametric assumptions.

Three components

1. EM algorithm for censored BVN competing risks
   - includes simulation study of HR estimation sensitivity to ρ in two group censored data design
2. Moment calculations for BVN using Stein's identity
3. R-package **bnc** (BVN competing risks) fitting AFT lm's

EM algorithm for censored data

Simple example (for clarity)

- *known* means of (log-)Normal latent vars, *no censoring*
- $y = \min(Y_1, Y_2)$ with $Y \sim \text{BVN}(0, \Sigma)$, Σ unknown
- Δ identifies which risk is observed (1 or 2)
- two risk times are never *both* observed
- Goal: the ML estimator of Σ from an random sample of y, Δ

Moment calculations for correlated BVN variates using Stein's identity

Malcolm Hudson[1]; Valerie Gares[2]

- *MU Statistics Research Congress 2016*
- *Paper in preparation*

Components

- Correlated competing risks and EM-algorithm application
- lack of identifiability
- Stein's identity for multivariate normal
- Application to Bivariate Normal moments
- examples

[1]Department of Statistics, Macquarie University
[2]Univ. Toulouse, France

Context

In survival analysis the **censored linear regression** model[3] is an EM-algorithm approach to estimating a parametric survival time distribution $P(Y > t)$. Censored data observing $T = \min(Y, C)$, where C is the censor time variable.

1. Complete data by imputation (E-step) of residual survival using $E(Y|Y > \tau, \Delta = 0)$.
2. Follow by ML-estimation based on complete data sufficient statistics

Competing risks survival data when observation of time to event of interest is not possible after the occurrence of a competing event (event *cause* $\Delta \in \{0, 1, 2\}$, if 0, independent censoring).

Observed data is now (T, Δ), where $T = \min(Y_1, Y_2, C)$.

We generalize the (univariate) EM algorithm for *correlated* competing risks.

[3] Schmee and Hahn

Charles Stein

- Charles Stein, mathematician, probabilist and statistician
  - inadmissability of the multivariate normal mean
  - Stein shrinkage
  - Stein Unbiased Risk Estimator

```
r  knitr::include_graphics("Stein5.jpg")
```

Stein's identity

- Identity *characterises* the multivariate Normal distribution
  - e.g. prove limit theorems by showing this identity is satisfied, for arbitrary f
- **Univariate:** $Y \sim N(\mu, \sigma^2)$ **iff**

$$E[(Y-\mu)f(Y)] = \sigma^2 E[f'(Y)]$$

  ($=>$ proof: integration by parts)
- **Multivariate:** $Y \sim MVN(\mu, \Sigma)$ **iff**

$$Cov[Y, f(Y)] = \Sigma E[\nabla f(Y)]$$

Conditional distribution

Let $Z = (Z_1, Z_2) \sim \text{BVN}(0, \Sigma)$, with $\sigma_{11} = \sigma_{22} = 1$, $\sigma_{12} = \rho$.

Ex. 1: $E(Z_1 | Z_2 > \tau)$
Conditional distribution of $(Z_1 | Z_2 = b)$ is $N(\rho b, 1 - \rho^2)$ with density

$$p_{1|2}(z|b) = \frac{1}{\sqrt{1-\rho^2}}\, \phi\left(\frac{z - \rho b}{\sqrt{1-\rho^2}}\right). \tag{1}$$

So, with Heavyside function H indicating values greater than 0 (step function 0 or 1)

$$\begin{aligned} E(Z_1 | Z_2 > \tau) &= E\ E^{Z_2}[Z_1 H(Z_2 - \tau)] \\ &= E\ H(Z_2 - \tau) E^{Z_2} Z_1 \\ &= E\ H(Z_2 - \tau)(\rho Z_2) \\ &= \rho E(Z_2 | Z_2 > \tau) \\ &= \rho \frac{\phi(\tau)}{1 - \Phi(\tau)} \end{aligned} \tag{2}$$

Univariate Stein

Ex. 2: $E(Z_1|Z_1 > a, Z_2 = b)$

$$E_{10.01} = E[(Z_1 - \rho b)|Z_1 > a, Z_2 = b] + \rho b \tag{3}$$

Now

$$\begin{aligned} E[(Z_1 - \rho b)|Z_1 > a, Z_2 = b] &= E^{Z_2=b}[(Z1 - \rho b)\, H(Z_1 - a)] \\ &= (1 - \rho^2) E^{Z_2=b}[\delta(Z_1 - a)] \\ &= (1 - \rho^2)\; p_{1|2}(a|b) \end{aligned} \tag{4}$$

Here (4) uses Stein's univariate identity, with the derivative of Heavyside, Dirac's delta. Dirac's delta, in convolution has the *sifting property*[4]

[4] Bracewell 2001; see Mathematica

Bivariate Stein

Ex. 3: $E(Z_1 Z_2 | Z_1 > a, Z_2 > b)$

Consider first $E[(Z_1 Z_2 H(Z_1 - a) H(Z_2 - b)]$.

The 2-d version of Stein's identity is:

$$\begin{aligned} E[\, Z_1\, f(Z_1, Z_2)] &= E[f_1(Z)] + \rho\, E[f_2(Z)] \\ &= (\text{term } 1) + \rho(\text{term } 2) \end{aligned} \tag{5}$$

with f_1 and f_2 defined as the two partial derivatives of

$$f(z_1, z_2) = z_2 H(z_1 - a) H(z_2 - b).$$

Term 1

The expectation term involving partial $f_1(z) = z_2 H(z_2 - b)\delta(z_1 - a)$ then reduces to

$$
\begin{aligned}
&E[Z_2 H(Z_2 - b)\delta(Z_1 - a)] \\
&\qquad = \int_b^\infty dz_2\, z_2 \int_{-\infty}^\infty dz_1\, \delta(z_1 - a)\, p_{12}(z_1, z_2) \\
&\qquad = \int_b^\infty dz_2\, z_2\, p_{12}(a, z_2) \\
&\qquad = \phi(a)\ E^{Z_1 = a}[Z_2 H(Z_2 - b)] \\
&\qquad = \phi(a)\, [\rho a\, E^{Z_1 = a} H(Z_2 - b) + (1 - \rho^2) p_{2|1}(b|a)] \\
&\qquad = \rho\, a\, \phi(a)\, [1 - \Phi\left(\frac{b - \rho a}{\sqrt{1 - \rho^2}}\right)] + (1 - \rho^2) p(a, b)
\end{aligned} \tag{6}
$$

Term 2

Partial $f_2(z) = H(z_1 - a)H(z_2 - b) + z_2 H(z_1 - a)\delta(z_2 - b)$, so we need

$$
\begin{aligned}
E[Z_2 H(Z_1 - a)\delta(Z_2 - b)] &= b \int_a^\infty dz_1\, p_{12}(z_1, b) \\
&= b\phi(b) \int_a^\infty dz_1\, p_{1|2}(z_1|b) \\
&= b\phi(b) \left[1 - \Phi\left(\frac{a - \rho b}{\sqrt{1 - \rho^2}}\right)\right]
\end{aligned}
$$

following a similar argument to that in equation (6).

Bivariate Stein (Ex.3, contd.)

The required conditional expectation then is the ratio

$$
\begin{aligned}
E_{11.00} &= \frac{E[(Z_1 Z_2 H(Z_1 - a)H(Z_2 - b)]}{P(a,b)} \\
&= \rho + (1-\rho^2)\,\frac{p_{12}(a,b)}{P(a,b)} \\
&\quad + \rho\frac{a\phi(a)\Psi\left(\frac{b-\rho a}{\sqrt{1-\rho^2}}\right) + b\,\phi(b)\Psi\left(\frac{a-\rho b}{\sqrt{1-\rho^2}}\right)}{P(a,b)}
\end{aligned}
\tag{7}
$$

with $\Psi(z) = 1 - \Phi(z), \quad P(a,b) = P(Z_1 > a, Z_2 > b).$

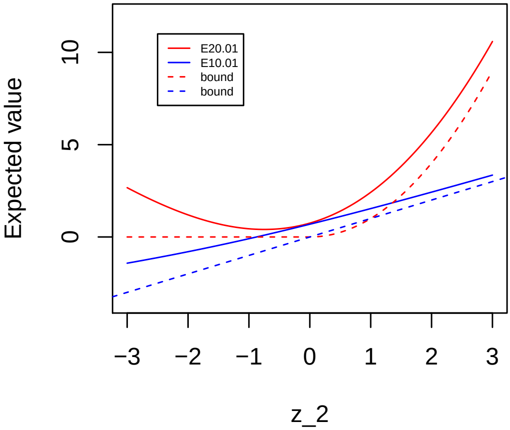
Numerical examples

$$E_{10.01} = E(Z_1 \mid Z_1 > \tau, Z_2 = \tau)$$

```
E10.01 <- function(y) {
    z <- y * (1 - rho)/sdet
    rho * y + sdet * dnorm(z)/pnorm(z, lower = F)
}

E20.01 <- function(y) {
    z <- (y - rho * y)/sdet
    det * (1 + y * dnorm(z)/pnorm(z, lower = F)) + rho * y
}
```

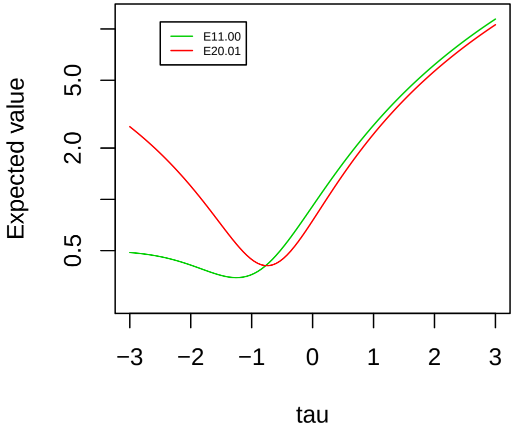
Plots $\rho = 0.5$



$E(Z_1 Z_2 \mid Z_1 > \tau, Z_2 > \tau)$

```
rho <- 0.5
det <- (1 - rho^2)
sdet <- sqrt(det)
p12 <- function(tau) {
    1/(2 * pi)/sdet * exp(-(1 - rho) * tau^2/det)
}
g0 <- function(x, Ycu) {
    dnorm(x) * pnorm((Ycu - rho * x)/sdet, lower = F)
}
```

Plot: covariance and second moment



Censored Linear model and EM algorithm for BVN correlated Competing Risks

Valerie Gares[5];
Malcolm Hudson[6]; Mauritzio Manuguerra [7];
Val Gebski [8]

- *Paper in preparation*
- Aim: parametric competing risk for correlated competing risks

[5]Univ. Toulouse, France
[6]Macquarie University and NHMRC CTC
[7]Macquarie University
[8]NHMRC CTC

R package for BVN correlated Competing Risks

Mauritzio Manuguerra [9];
Valerie Gares[10];
Malcolm Hudson[11]

- package BNC (under development)
- *Paper in preparation*

bnc package

- bivariate normal censored (linear model)
- include code for rho fixed and test code for copula data

[9]Macquarie University
[10]Univ. Toulouse, France
[11]Macquarie University and NHMRC CTC

Tests

Simple:

- Is Likelihood increasing? (nL) when does it not?
- effect of different starting points on convergence
  - perfect start versus start estimate assuming rho=0 (-> independent censoring)
- how many iterations to convergence?
- use package for correlation paper's data example (Table 3) (but, will need SE's)
- with y=min(Y1,Y2), i.e. no censoring:
  - the problem is under-determined/ ill-posed?
- with rho=0, compare results with estimation from censored data
  – time to Event 1 censored by cause 2 or eof, parametric lognormal model
  – R package survreg

Further tests

- Table 2 copula (non-BVN example)
  - write test code generating data, table results

Complex:

- standard errors of EM solution
  - Jamshidian and Jennrich 2000, DCM approach
  - code for package
- ML solution with rho fixed
  - code Anderson & Olkin for package
- together these enable:
  - completion of simulation Tables 1 and 3 using package

Figure 2: Thanks for your attention!

Figure 3: Thanks for your attention!

Figure 4: Charles Stein