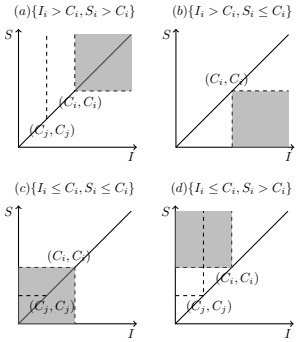
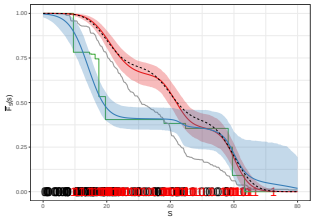
Slide 1

Bayesian Nonparametric Bivariate Survival Regression for Current Status Data

Peter Müller, UT Giorgio Paulon, UT,
V. Giancarlo Sal y Rosas, PUCP,



biv current status data



estimation with (univ) CS data

www.math.utexas.edu/users/pmueller/isbabnp.pdf

1 Partners Notification study

Slide 2

Partners Notification study

Golden et al (2005 NEJM)

- Heterosexual men and women that were treated for gonorrhea and/or chlamydia up to 14 days prior to enrollment Randomization 1:1 to

  Intervention: Vouchers for medication to give to their sex partners, or if they preferred, staff members contacted their partners and provided them with medication without a clinical examination

  Control: Standard treatment

- *Outcome:* Persistent or recurrent gonorrhea and/or chlamydial infection in the original participant.

- *Visits:* Only *one* visit between 3 and 19 weeks after enrollment

Slide 3

Partners Notification study

Table: Descriptive statistics of the cohort

	Control	Intervention	Total
N (%)	921 (50.3)	911 (49.7)	1832
Female	721 (78.3)	720 (79.0)	1441 (78.7)
Age (years)[1]	21 [19-25]	22 [19-26]	21 [19-25]
Initial diagnosis			
Gonorrhea	131 (14.2)	132 (14.5)	263 (14.4)
Genital chlamydia	742 (80.6)	732 (80.4)	1474 (80.5)
Both	48 (5.2)	47 (5.1)	95 (5.1)
Events	118 (12.8)	86 (9.4)	204 (11.1)
Visit time C_i	86 [77-103]	87 [76-103]	87 [77-103]

Slide 4

Data

Disease free $\longrightarrow$ infected I_i $\longrightarrow$ symptoms S_i
or $\rightarrow$ symptoms S_i (due to other causes)

Notation: patient i, $i = 1, \ldots, n$

- I_i time to infection
- S_i time to symptoms
- $L_i = S_i - I_i$ lag time
- C_i visit time (observation time)
- $\Delta_{Ii} = I(I_i \le C_i)$ and $\Delta_{Si} = I(S_i \le C_i)$
- x_i covariates

Observed data: at time of visit record

$$\boldsymbol{Y}_i = (C_i, \Delta_{Ii}, \Delta_{Si}, x_i)$$

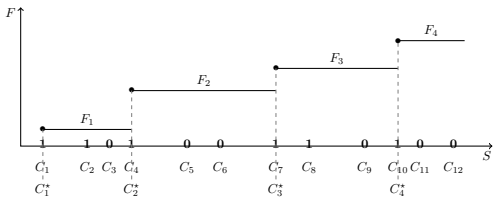
(bivariate) current status data (Jewell & van der Laan, 2003 Handbook of Stat)

2 Current status data

Slide 5

Current status data

Univariate CS data, $\Delta_i = I(S_i \le C_i)$, wlog. ordered by C_i, and $\Delta_1 = 1, \Delta_n = 0$



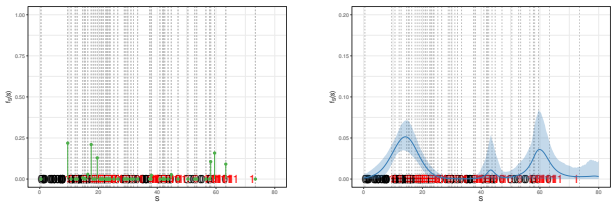
Let $A = \{i > 1 : (\Delta_{i-1}, \Delta_i) = (0, 1)$ (i.e. all left censored following a right censored observations); $A \equiv A \cup \{1\}$.

- Let $C^\star_j$ denote the C_i in A, plus $C^\star_{J+1} > \max\{C_i\}$;
- Easy to show, $f_S(s) = \sum_j p_j \delta_{C^\star_j}$

Slide 6

Estimating $f_S(\cdot)$

Easy EM algorithm to estimate $\hat{f}_S(\cdot)$ (Groeneboom & Wellner, 1992).



(a) NP mle (b) mix of normal mle

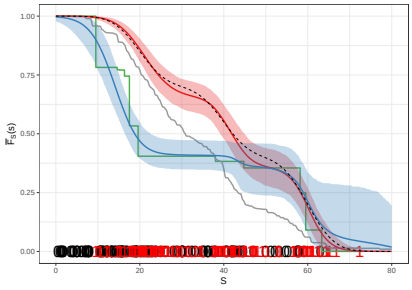
- NP mle tends to shrink f_S towards the extremes;
- Par model, e.g., mix of normals, smooths the mle, but shrinkage persists

Slide 7

BNP to the rescue

To regularize inference we

1. Dependent censoring: regression of S_i on C_i (visit time), $C_i = \min\{S_i + \text{Exp}(\lambda),\ \text{Unif}\}$;
2. BNP prior on f_S



Truth: black dashed

NP mle: green step fct

Parametric mix of N: blue curve

NP mle & dep censoring: gray step function

BNP & dep censoring: red curve

3 Bivariate current status data

Slide 8

Bivariate current status data

$\boldsymbol{Y}_i = (\Delta_{Ii}, \Delta_{Si}, C_i)$, note:

- $(\Delta_I, \Delta_S) = (1,0)$, i.e., $(I < C < S)$: symptoms due to disease of interest *or other*
- $(\Delta_I, \Delta_S) = (0,1)$, i.e., $(I > C > S)$: symptoms due to *other causes*

And two more cases. Let $F_I = p(I \le C)$, $F_S = p(S \le C)$, $F_{IS} = p(I < C, S < C)$.[.25cm] Only F_I, F_S, F_{IS} are identifiable (Wang & Ding, 2000 Bka).

Slide 9

Likelihood fct: Recall
$F_I = p(I \le C)$, $F_S = p(S \le C)$, $F_{IS} = p(I < C, S < C)$.

$$\prod_i F_{IS}^{\Delta_I \Delta_S} \times (F_I - F_{IS})^{\Delta_I(1-\Delta_S)}$$
$$\times (F_S - F_{IS})^{(1-\Delta_I)\Delta_S} \times (1 - F_I - F_S + F_{IS})^{(1-\Delta_I)(1-\Delta_S)}$$

$\longrightarrow$ use prior regularlization and known structure to allow for inference on $F(I,S)$.

Copula models: Ma et al. (2015 Bka); Li et al., (2017, CSDS); Wang & Ding (2000 Bka)
copula $(I,S) \sim C(\xi)$, with sensitivity analysis for ξ.

Dependent probit: Dunson & Dinse (2002, Bmcs) use a probit model with frailities to induce dependence.

Structure: alternatively use structural assumptions to build $F(I,S) \to$ next

Slide 10

Building a bivariate CS data model

Assume

$$F(I,S) = w\, F^\star(I,S) + (1-w)\, F'(I,S)$$

with

Symptom due to disease: $F'(I,S) \implies S > I$;

Symptom due to other cause: $F^\star(I,S)$, no constraint

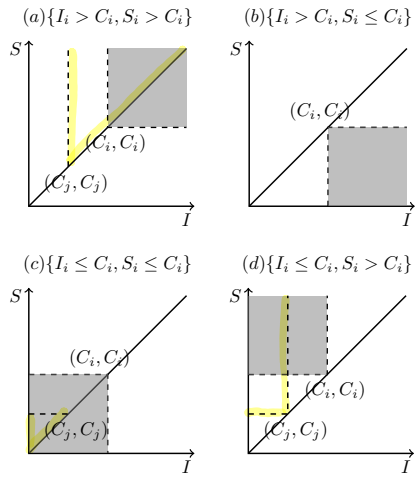
Slide 11

Likelihood factors: Let $F_{11}(C) = p(I < C, S < C)$, $F_{10}(C) = p(I < C, S \ge C)$, $F_{01}(C) = p(I \ge C, S < C)$, $F_{00}(C) = p(I \ge C, S \ge C)$.

Δ_I	Δ_I	$n_{k\ell}$	likelihood
0	0	1303	$F_{00} = w\, F^\star_{00} + (1-w)\, F'_{00}$
0	1	325	$F_{01} = w\, F^\star_{01} + 0$
1	0	121	$F_{10} = w\, F^\star_{10} + (1-w)\, F'_{10}$
1	1	83	$F_{11} = w\, F^\star_{11} + (1-w)\, F'_{11}$

Slide 12

Likelihood factors



Likelihood factors for the 4 cases under F' and $F^\star$.

Slide 13

Building dependence structure

Using marginal models $F_I(I)$, $F^\star(S)$ and an assumption for $F'(S \mid I)$ we build a dependent model:

- $F^\star$: symptoms due to other causes, $S \perp I$

$$F^\star(I, S) = F_I(I)\, F^\star{}_S(S)$$

- F': symptoms due to disease, $L = S - I$, and $L \perp I$

$$F'(I, S) = F_I(I)\, F_L(S - I)$$

with $L \sim \text{Exp}(\lambda_L)$

Slide 14

Let $F_I = p(I \le C_i) = F_I(C_i)$ and $\bar{F}_I = 1 - F_I$,

$$F_{S|I<C} = p(S \le C_i \mid I \le C_i) = \int_0^{C_i} \frac{f_I(I)}{F_I(C_i)} F_L(C_i - I)\, dI$$

and similarly $\bar{F}_{S|I<C} = p(S > C_i \mid I \le C_i)$. Then

$$\begin{aligned} F_{00} &= \bar{F}_I\,(w\bar{F}^\star_S + 1 - w) = \bar{F}_I\,(1 - w\, F^\star{}_S) \\ F_{01} &= \bar{F}_I\, w\, F^\star{}_S \\ F_{11} &= F_I\,\left(w\, F^\star{}_S + (1-w)\, F'_{S|I<C}\right) \\ F_{10} &= F_I\,\left(w\, \bar{F}^\star{}_S + (1-w)\, \bar{F}'_{S|I<C}\right). \end{aligned}$$

Slide 15

BNP prior

BNP prior: mix of N marginal $F_I(\cdot)$ and $F^\star_S(S)$

$$\begin{aligned} F_I(I) &= \int N(I \mid \overbrace{\mu, \sigma^2}^{\boldsymbol{\theta}})\, dH_I(\boldsymbol{\theta}) \\ F^\star{}_S(S) &= \int N(S \mid \boldsymbol{\theta}) \quad dH_S(\boldsymbol{\theta}) \end{aligned}$$

with your favorite BNP prior $p(H_I)$ and $p(H_S)$.

Regression: letting $\boldsymbol{x}_i \in X$ denote patient-specific covariates, $\boldsymbol{x}_i$ = (gender,arm,age),
extend the BNP prior to $p(H_{I,x};\ \boldsymbol{x} \in X)$.

DDP: We use a dependent DP (DDP) prior with simple linear model for the covariate-dependent atoms.

Slide 16

Parametric sub-models

Two parametric submodels introduce (important) prior knowledge: on

(i) Corr of (C_i, S_i):

$$p(C_i \mid S_i) = \min\{S_i + \text{Exp}(\lambda),\ \text{Unif}\}$$

(ii) Lag times $L_i = S_i - I_i$ under F':

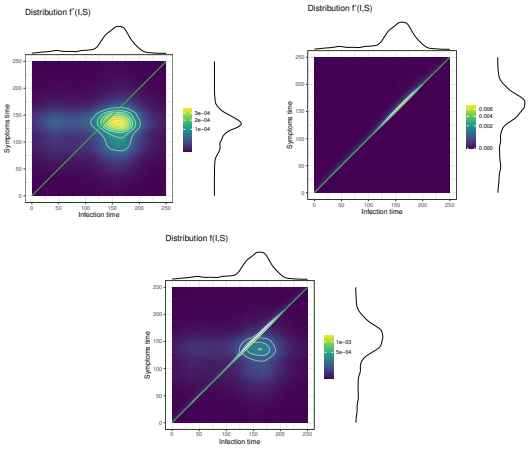
$$L \sim \text{Exp}(\lambda_L),\ \text{with inf prior } p(\lambda_L)$$

4 Results

Slide 17

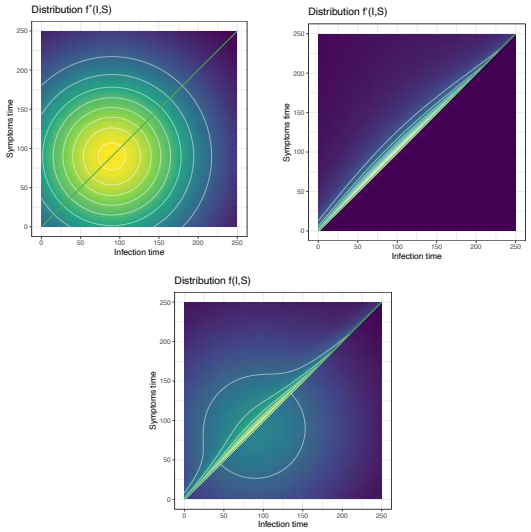
Partner Notification Study - Results

Posterior estimated $F^\star$, F' and F



Slide 18

Compare with prior mean $F^\star$, F' and F



Slide 19

Treatment effect

Recall $F_I(t \mid \boldsymbol{x}) = \int N(t \mid \overbrace{\mu, \sigma^2}^{\boldsymbol{\theta}})\, dH_{\boldsymbol{x}}^{(I)}(\boldsymbol{\theta})$.
Under the linear DDP $\rightarrow$ simple DP mixture:

- Let $\boldsymbol{d}$ be a design vector for covariates $\boldsymbol{x}$. Then

$$\boldsymbol{\theta} \sim H_{\boldsymbol{x}} = \sum_\ell \pi_\ell \delta_{\boldsymbol{d}'\boldsymbol{m}_\ell} \iff \begin{cases} \boldsymbol{m} & \sim H = \sum \pi_\ell \delta_{\boldsymbol{m}_\ell} \\ \boldsymbol{\theta} & = \boldsymbol{d}'\boldsymbol{m} \end{cases}$$

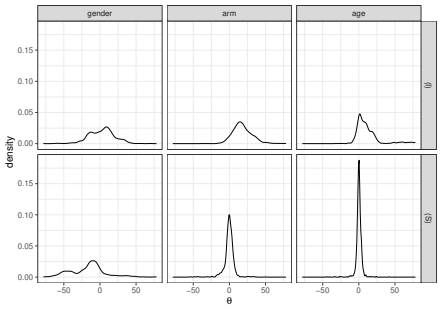
  and

$$F_I(t \mid \boldsymbol{x}) = \int N(t \mid \boldsymbol{d}'\boldsymbol{m}, \sigma^2)\, dH(\boldsymbol{m})$$

- Let $\boldsymbol{m} = (\alpha, \beta, \gamma)$ (for gender, trt, age). Then $H(\boldsymbol{m}) = \sum \pi_\ell \delta_{\boldsymbol{m}_\ell}$ implies $H_\beta(\beta) = \sum \pi_\ell \delta_{\beta_\ell}$.

Slide 20

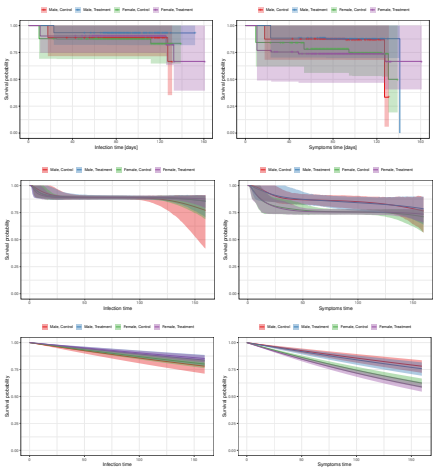
Gender, treatment and age effects in H_I (top) and H_S (bottom) (note that H_S is used under $F^\star$ only)



- Delayed infection times for treated patients
- Earlier time to symptom (due to other causes) for women?
- Some evidence for an age effect, with higher risk for younger patients (more risk taking?)

Slide 21

Inference under simplified models



data (KM)

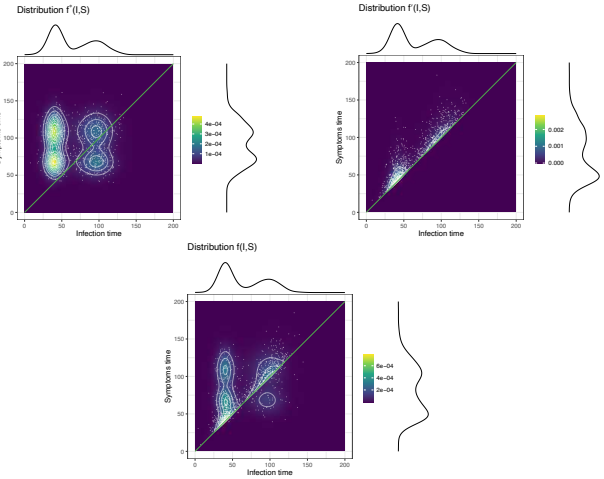
indep linear DDP

Gumbel

5 Simulation

Slide 22

Simulation



White dots are the (simulated & known) true event times. However, inference only conditions on Δ_I, Δ_S.

Slide 23

More simulations

(I) independent C_i & dependent (I, S); (II) dependent C_i & independent (I, S) $(w = 1)$, and (III) dependent censoring & dependent (I, S).

	Sample Size	Distr.	De Iorio et al.	Bivariate Gumbel	Our method
(I)	$n = 250$	Inf.	1.64 (0.92, 3.01)	4.01 (3.14, 5.59)	1.10 (0.09, 2.24)
		Sym.	2.98 (1.11, 5.01)	6.15 (5.31, 8.77)	1.33 (0.18, 3.72)
	$n = 1000$	Inf.	1.32 (0.73, 1.90)	3.76 (3.19, 4.54)	0.50 (0.04, 1.80)
		Sym.	2.32 (1.19, 3.25)	5.99 (5.31, 6.99)	1.30 (0.54, 2.66)
(II)	$n = 250$	Inf.	0.96 (0.74, 1.56)	3.44 (3.08, 4.59)	0.99 (0.13, 2.07)
		Sym.	8.44 (5.21, 12.30)	11.75 (9.18, 18.01)	**0.76** (0.22, 2.16)
	$n = 1000$	Inf.	0.80 (0.50, 1.10)	3.12 (3.03, 3.41)	**0.19** (0.05, 0.50)
		Sym.	8.18 (6.28, 10.32)	10.74 (9.58, 12.49)	**0.12** (0.02, 0.37)
(III)	$n = 250$	Inf.	4.45 (3.00, 6.30)	4.24 (3.09, 5.79)	**0.45** (0.08, 1.14)
		Sym.	9.82 (6.70, 13.20)	8.08 (5.72, 12.15)	**0.24** (0.03, 0.81)
	$n = 1000$	Inf.	4.10 (3.18, 4.96)	3.96 (3.24, 4.81)	**0.13** (0.01, 0.35)
		Sym.	9.94 (8.44, 11.71)	7.98 (6.31, 10.06)	**0.05** (0.01, 0.15)

MSE for $F_I(\cdot)$ ("Inf") and $F_S(\cdot)$ ("Symp"), De Iorio = two indep DDPs

Conclusion

Slide 24

Conclusion

- BNP model for biv current status data;
- Despite limited information in observed data, we get meaningful inference even under moderate n;
- Combining the flexible BNP model with some known structure provides sufficient regularization;
- While "BNP is always right", need to be careful to recognize restricted identifiability;
- Anticipating desired inference in the model construction we get concise inference on covariate effects (beyond visual comparison of survival curves).
- BNP priors for marginal distributions are combined based on some basic assumptions, more specific than generic copula constructions.