Boehringer Ingelheim

Joint meeting of 3 WGs of the IBS / DR

Lübeck, 2009-12-05

G. Nehmiz

Introduction to non-parametric Bayes methods

Overview

- Parametric and nonparametric probability models
- Prior distributions and prior processes
- Overlay of prior information and information from data
- Example: Cox model (counting process formulation)
- Discussion
- References

Parametric and nonparametric probability models

- P: Model class + parameter value → data

  NP: Whole distribution → data

Parametric and nonparametric probability models

- P: Test whether a parameter lies in a given region
  or
  investigation of posterior distribution of the parameter

  NP: Test whether 2 distributions as a whole are equal (reference space necessary)
  or
  Investigation of posterior distribution (continuously indexed family of neighbourhoods) of a distribution

Ref.: Lehmann 1986, 334-337; Brunner/Langer 1999, 32-33

Parametric and nonparametric probability models

- What does the Bayesian synthesis

Prior function → Posterior function ← Likelihood

mean if spaces of whole distributions are investigated instead of a finite-dimensional parameter space?

- In particular, how much “hidden information” is contained in an apparently uninformative prior distribution, selected for convenience or tractability?

Ref.: Berger, J.A.S.A. 2000, 1272 right

Prior distributions and prior processes

Boehringer Ingelheim

- “Definition”: A stochastic process is an indexed family of distributions over a sample space, whereby the indexing has to be “continuous” in a certain sense, or at least “measurable”
- If the sample space has dimension > 1, the process is also called a “random field”

Ref.: Møller/Waagepetersen 2004, 7-11

Prior distributions and prior processes

- A distribution of distributions can be considered as a stochastic process, whereby the index set is itself a distribution and “generates” a set of neighbourhoods around a given distribution
- The given distribution, around which we want to construct the neighbourhoods, is defined on the partitions of the sample space

Ref.: Navarrete et al., Stat. Modelling 2008, 4

Prior distributions and prior processes

Boehringer Ingelheim

- The historically first process of this kind is the Dirichlet process; for each partition, it assigns a Dirichlet distribution to the probabilities of each element of the partition
- We obtain a family of distributions around the given distribution
- The family is conjugate to the given distribution, samples from the given distribution (also if independently censored) can be included
- The distributions in the family are, with probability 1, discrete

Ref.: Ferguson, Ann. Stat. 1973, Gelfand et al. 2007

Prior distributions and prior processes

- The Dirichlet process was applied successfully to the estimation of 1 survival curve with right-censoring
- A sharp prior distribution has to be given first, around which the family of distributions is centered
- The relative weight of the given distribution, relative to the information provided by the data, is described by a non-negative number, c
- The Kaplan-Meier estimator can be seen as a limiting case if c = 0

Ref.: Suzarla/Van Ryzin, J.A.S.A. 1976

Prior distributions and prior processes

The Polya tree is a special case of the Dirichlet process whereby the partitions of the sample space are generated through recursive bisection; degenerate splits are possible. At each branching, the probabilities of the 2 sub-sections are Beta-distributed.

- The Polya tree also needs a given sharp distribution to begin with
- The Polya tree already allows a representation of the Kaplan-Meier curve, in the limiting case that the weight of the prior distribution becomes 0

Ref.: Muliere/Walker, Scand.J.Statist. 1997

Prior distributions and prior processes

The Beta process is defined on $[0,\infty)$. The definition starts with the cumulative hazard function Λ and not with the distribution of the event times

- In the non-continuous case, it is not generally true that $F(t) = \exp(1-\Lambda(t))$
- One has to select a basic hazard function $d\Lambda_0^*(t)$
- It is assumed that the increments $d\Lambda$ are independent and non-negative (i.e. Λ is a Lévy process) and that the $d\Lambda$ are beta-distributed with parameters $c * d\Lambda_0^*(t)$, $c * (1-d\Lambda_0^*(t))$
- The existence is difficult to prove

Ref.: Hjort, Ann.Stat. 1990

Prior distributions and prior processes

- Also the Beta process is conjugated to samples (possibly censored) from the corresponding basic distribution
- In the limit for $c = 0$, the estimated survival function becomes the Kaplan-Meier curve

Ref.: Hjort, Ann.Stat. 1990

Prior distributions and prior processes

- The counting process counts the number of events observed for each interval (details in example below)
- As an associated Lévy process (cumulative intensity process), the Gamma process is often used (see also example below)
- This is problematic as the assumption of independent increments is implausible in particular in neighbouring intervals
- However, an alternative Lévy process is the Beta process (see also example below)

Ref.: Sinha/Dey 1998, Laud et al. 1998

Overlay of prior information and information from data

Boehringer Ingelheim

- The data-generating distribution is unknown, all that can be observed is the data (including censoring information)
- In all cases mentioned, the Bayesian synthesis behaves “reasonably” in so far as it depends only from the information that is in the data

Ref.: Bernardo/Smith 1994, 177-181

Example: Cox model (counting process formulation)

- Discretization: For all distinct failure and censoring times t_i (i=1,…,n), consider the risk set R_i. Events / censorings of several patients are possible for a time-point. All censoring is assumed to be non-informative here
- Consider for each patient j (j=1,…,N) the random variable that counts the number of events until t, this is a "counting process" $N_j(t)$
- Indicate by 0/1 whether patient j, while in risk set, has had an event at time $t \in [t_i, t_i+dt)$. Multiple events are possible for a patient but only with different t_is. At the boundaries, define $t_0 := 0$ and an arbitrary $t_{n+1} > t_n$.

Example: Cox model (counting process formulation)

- Risk set (special case: only 1 event / patient):

Patient (j)	Time-point (t_i)				
	t_1	t_2	t_3	. . .	t_n
1	1 (c)	0	0	. . .	0
2	1 (e)	0	0	. . .	0
3	1	1 (c)	0	. . .	0
4	1	1	1 (e)		0
5	1	1	1 (e)		0
.	.	.	.		
:	:	:	:		
N	1	1	1	. . .	1 (e)

(c): Censoring occurs
(e): Event occurs

Example: Cox model (counting process formulation)

- Consider the "intensity process" of patient j:

  $I_j(t)dt := E(dN_j(t) \mid \text{previous events/censorings in } [0,t))$

  whereby $dN_j(t)$ is the increment of $N_j(t)$ in the interval $[t,t+dt)$ and can take the values 0 or 1. $I_j(t)dt$ is the probability that subject j has an event in $[t,t+dt)$, and with $dt \rightarrow 0$, $I_j(t)$ becomes the hazard $h_j(t)$
- While the patient is still in the risk set (as described by a further process $Y_j(t)$), the further assumption is that a covariate vector Z_j influences the hazard multiplicatively:

  $I_j(t) = Y_j(t) * \lambda_0(t) * e^{z_j\beta}$

  with unknown but fixed "baseline hazard" function $\lambda_0(t)$.

Ref.: Clayton 1991, Sinha/Dey 1997, Laud et al. 1998, Hellmich 2001

Example: Cox model (counting process formulation)

- Parameters in the PH model

  $$I_j(t) = Y_j(t) * \lambda_0(t) * e^{z_j\beta}$$

  are β and $\lambda_0(t)$ (or its integral $\Lambda_0(t) := \int_0^t \lambda_0(u)du$, the cumulative hazard function).

  $\lambda_0(t)$ is piecewise constant, in $[t_i,t_{i+1})$ it is =: $\lambda_{0,i}$.

  The likelihood function, given realisations of $N_j(t)$ and $Y_j(t)$, is

  $$L(\beta,\lambda_{0,0},\ldots,\lambda_{0,n}) \sim \text{Product}(i=1,\ldots,n) \text{ of}$$

  $$(1-\lambda_{0,i})^{\text{Sum}(j\in R_i)\, e^{z_j\beta}}$$

  $$* \ \lambda_{0,i}^{\text{Sum(patients with event at } t_i)\, e^{z_j\beta}}$$

Example: Cox model (counting process formulation)

- The prior distributions (considered independent of each other) are:

  Pseudo-constant for β

  and because the $dN_j(t)$ can be considered Poisson-distributed with intensity $I_j(t)dt$ and the Gamma distribution is conjugated to that:

  Gamma $(c^* d\Lambda_0^*(t)\ ,\ c)$ for $d\Lambda_0(t) = \lambda_0(t)dt$
  with a certainty parameter c and an initial guess $\Lambda_0^*(t)$ of the cumulative hazard

  → only true without tied event times

Example: Cox model (counting process formulation)

- Therefore a better prior distribution for $d\Lambda_0(t)$ (actually for the values of the piecewise constant function I(t)) is

  $$\text{Beta}\ (c(t) * d\Lambda_0^*(t)\ ,\ c(t) * (1 - d\Lambda_0^*(t)))$$

  where $d\Lambda_0^*(t)$ is an initial guess, and we assign

  $$c(t) := c_0 * e^{-t/(t_n+1)}$$

  whereby c_0 is one parameter describing the certainty of $d\Lambda_0^*(t)$: Smaller c_0 means less shrinkage and higher weight for the observations t_i.

Example: Cox model (counting process formulation)

Boehringer Ingelheim

- Example data:

18 Leuk: survival analysis using Cox regression

Treatment	Survival time in weeks						
Placebo	1	1	2	2	3	4	4
	5	5	8	8	8	8	11
	11	12	12	15	17	22	23
6-MP	6*	6	6	6	7	9	10*
	10	11*	13	16	17*	19*	20*
	22	23	25*	32*	32*	34*	35*

* indicates censoring

- Matched-pairs structure now ignored

Ref.: Spiegelhalter et al. 1996

Example: Cox model (counting process formulation)

Boehringer Ingelheim

- WinBUGS results, c = 1:

Node statistics

node	mean	sd	MC error	2.5%	median	97.5%	start	sample	
beta	1.629	0.4021	0.01324	0.8882	1.608	2.483	4001	10000	OK
dL0[1]	0.03507	0.02389	3.677E-4	0.004593	0.02981	0.09427	4001	10000	t= 1
dL0[2]	0.03811	0.02574	4.244E-4	0.004999	0.03275	0.1009	4001	10000	t= 2
dL0[3]	0.02114	0.02077	3.988E-4	6.048E-4	0.01488	0.07655	4001	10000	t= 3
dL0[4]	0.04376	0.02971	4.617E-4	0.005707	0.0374	0.1163	4001	10000	t= 4
dL0[5]	0.04806	0.03237	4.493E-4	0.006248	0.04094	0.1294	4001	10000	t= 5
dL0[6]	0.07165	0.03888	5.804E-4	0.01601	0.06458	0.1656	4001	10000	t= 6
dL0[7]	0.02718	0.02615	4.699E-4	7.727E-4	0.01938	0.09738	4001	10000	t= 7
dL0[8]	0.117	0.0522	7.069E-4	0.03554	0.1103	0.2369	4001	10000	t= 8
dL0[9]	0.0371	0.03506	5.769E-4	0.001113	0.02678	0.1301	4001	10000	t=10
dL0[10]	0.08195	0.05177	6.631E-4	0.01088	0.07243	0.206	4001	10000	t=11
dL0[11]	0.1047	0.0644	9.475E-4	0.01471	0.09289	0.2597	4001	10000	t=12
dL0[12]	0.06194	0.05357	8.638E-4	0.002142	0.04721	0.1998	4001	10000	t=13
dL0[13]	0.06817	0.05965	9.734E-4	0.002006	0.0517	0.221	4001	10000	t=15
dL0[14]	0.06937	0.05915	9.341E-4	0.002229	0.05414	0.2193	4001	10000	t=16
dL0[15]	0.09532	0.0753	0.001085	0.004758	0.07646	0.2837	4001	10000	t=17
dL0[16]	0.1985	0.1016	0.001343	0.03303	0.1894	0.4119	4001	10000	t=22
dL0[17]	0.7895	0.2508	0.007882	0.1927	0.9136	1.0	4001	10000	t=23

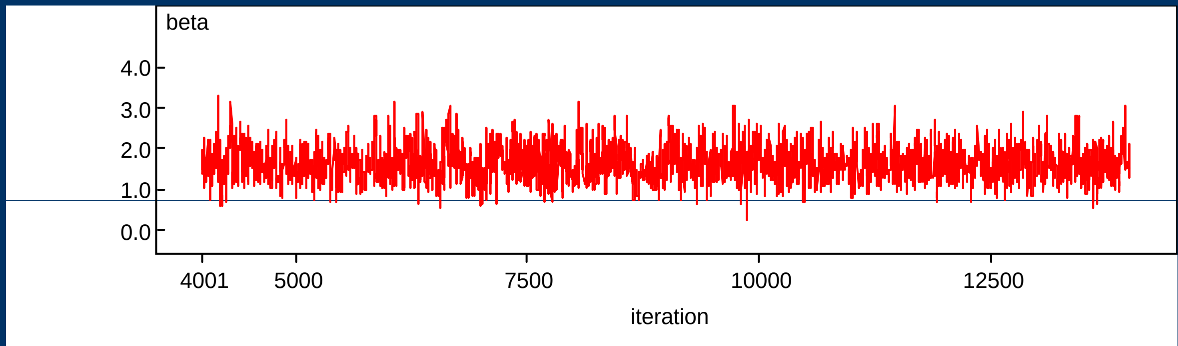
dL0 is the average hazard of both groups

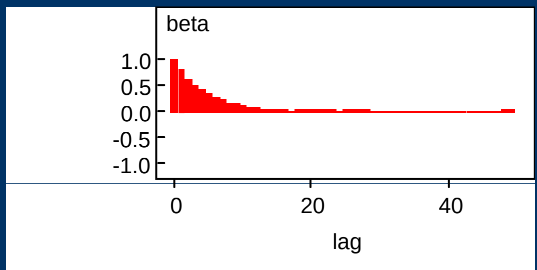
Example: Cox model (counting process formulation)

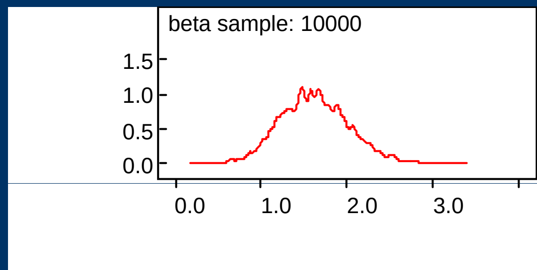
- WinBUGS results:

  Similar results are output for the estimated survival curves of both groups separately

  Graphs of the treatment difference parameter “beta”:

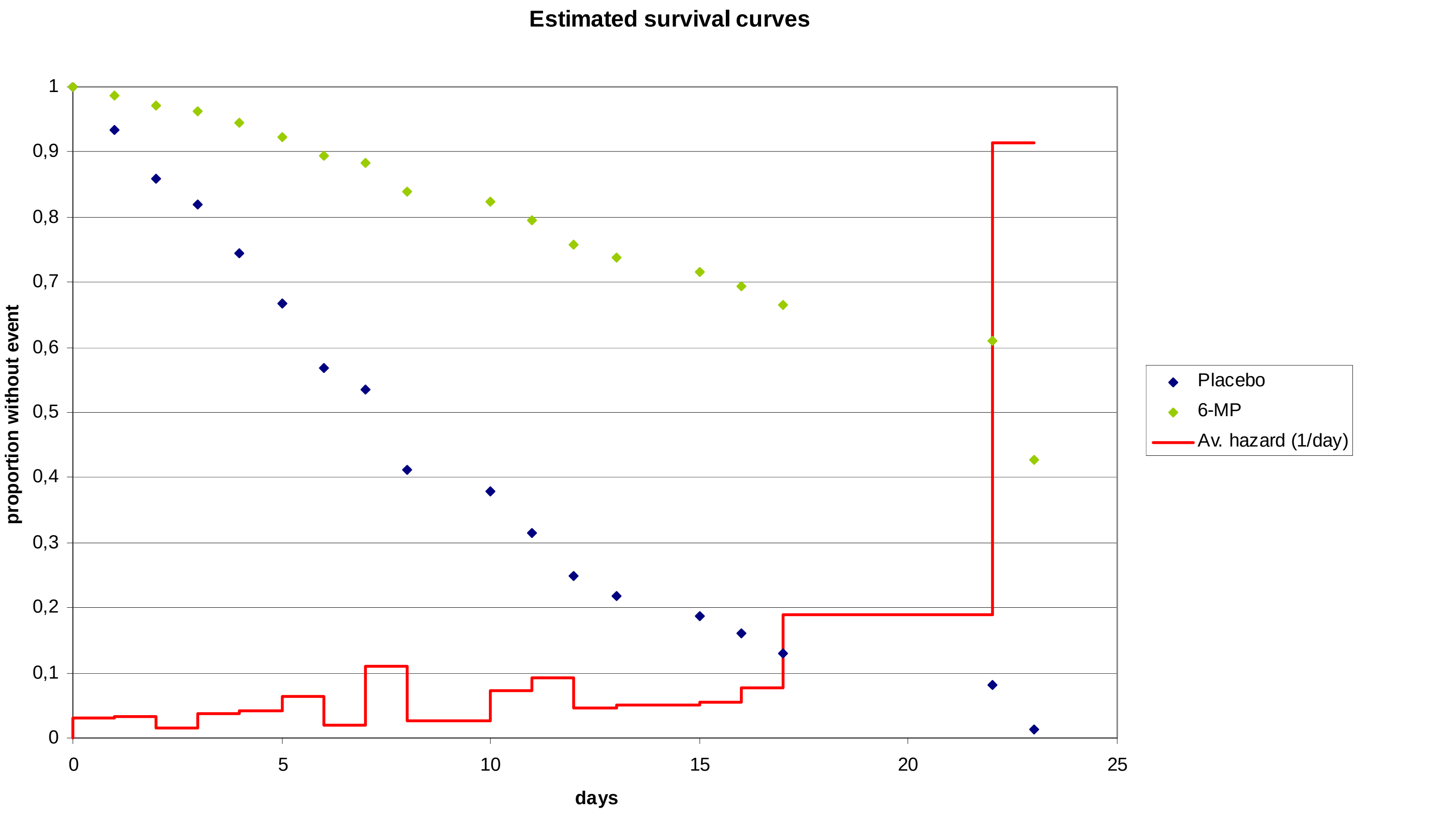






Example: Cox model (counting process formulation)

- WinBUGS results:



- All 3 curves have distributions (vertical)

Discussion

- As a first step, robustness w.r.t. selection of c needs to be investigated, see e.g. Laud et al. 1998, p. 218-219
- Interpretation of prior information on cumulative hazard remains difficult
- Interpretation of the limitations that arise from the mathematical properties of the processes still not sufficiently understood.

References

Lehmann EL:
“Testing Statistical Hypotheses”.
New York ...: John Wiley & Sons, 2nd ed. 1986

Brunner E, Langer F:
“Nichtparametrische Analyse longitudinaler Daten”.
München/Wien: R. Oldenbourg Verlag 1999

Berger JO:
Bayesian Analysis: A Look at Today and Thoughts of Tomorrow.
J.A.S.A. 2000; 95 (452): 1269-1276

Navarrete C, Quintana FA, Müller P:
Some issues in nonparametric Bayesian modelling using species sampling models.
Statistical Modelling 2008; 8 (1): 3-21

Boehringer Ingelheim

References

Møller J, Waagepetersen RP:
“Statistical Inference and Simulation for Spatial Point Processes”.
Boca Raton/FL ...: Chapman & Hall / CRC 2004

Ferguson T:
A Bayesian analysis of some nonparametric problems.
Annals of Statistics 1973; 2 (1): 209-230

Gelfand AE, Guindani M, Petrone S:
Bayesian Nonparametric Modelling for Spatial Data Using Dirichlet Processes. In:
Bernardo JM, Bayarri MJ, Berger JO, Dawid AP, Heckerman D, Smith AFM, West M (eds.):
“Bayesian Statistics 8”.
Oxford: Oxford University Press 2007, 175-200

References

Suzarla V, Van Ryzin J:
Nonparametric Bayesian Estimation of Survival Curves from Incomplete Observations.
J.A.S.A. 1976; 71 (356): 897-902

Muliere P, Walker S:
A Bayesian Non-parametric Approach to Survival Analysis Using Polya Trees.
Scandinavian Journal of Statistics 1997; 24: 331-340

Hjort NL:
Nonparametric Bayes estimators based on Beta processes in models for life history data.
Annals of Statistics 1990; 18 (3): 1259-1294

Bernardo JM, Smith AFM:
"Bayesian Theory".
Chichester ...: John Wiley & Sons 1994

References

Sinha D, Dey DK:
Survival Analysis Using Semiparametric Bayesian Methods.
In:
Dey D, Müller P, Sinha D (eds.):
"Practical Nonparametric and Semiparametric Bayesian Statistics".
New York / Berlin / Heidelberg: Springer-Verlag 1998, 195-211

Laud PW, Damien P, Smith AFM:
Bayesian Nonparametric and Covariate Analysis of Failure Time Data.
In:
Dey D, Müller P, Sinha D (eds.): ..., 213-225

Bernardo JM, Smith AFM:
"Bayesian Theory".
Chichester ...: John Wiley & Sons 1994

References

Gilks WR, Best NG, Tan KKC:
Adaptive Rejection Metropolis Sampling within Gibbs Sampling.
Appl. Stat. 1995; 44 (4): 455-472

Gilks WR, Neal RM, Best NG, Tan KKC:
Corrigendum: Adaptive Rejection Metropolis Sampling.
Appl. Stat. 1997; 46 (4): 541-542

http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml

Spiegelhalter D, Thomas A, Best N, Gilks W:
BUGS 0.5 Examples, Volume 1 (version i).
Cambridge: MRC Biostatistics Unit 1996

References

Clayton DG:
A Monte Carlo Method for Bayesian Inference in Frailty Models.
Biometrics 1991; 47 (2): 467-485

Sinha D, Dey DK:
Semiparametric Bayesian Analysis of Survival Data.
J. A. S. A. 1997; 92: 1195-1212

Hellmich M:
Bayes'sche Untersuchung von zensierten Daten.
Presentation, Homburg/Saar 2001,
http://www.imbei.uni-mainz.de/bayes/Documents/baysur.pdf

Questions?

Thank you