Introduction to Bayesian non-parametric survival analysis

G. Nehmiz

Joint meeting of 3 WGs of the IBS / DR
Freising 2013-11-07

Boehringer Ingelheim

Overview

Boehringer Ingelheim

Parametric and nonparametric probability models
Parametric and nonparametric inference
Prior distributions and prior processes
Estimation of 1 survival curve
Comparison of 2 survival curves: Cox model (counting process formulation)
Discussion
References

Parametric and nonparametric probability models

Boehringer Ingelheim

Model class + parameter value → data

Whole distribution → data

Parametric and nonparametric inference

Boehringer Ingelheim

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 of a distribution

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

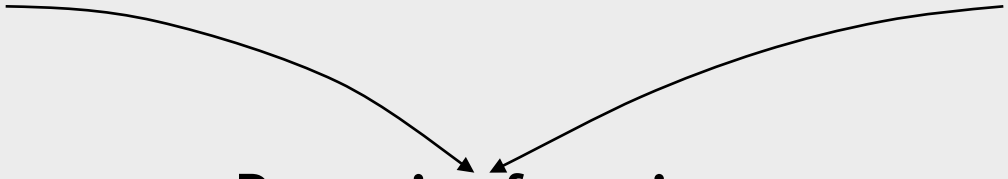
Parametric and nonparametric inference

Boehringer Ingelheim

What does the Bayesian synthesis

Prior function

Likelihood



Posterior function

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"
– details are too difficult now

If the sample space has dimension > 1, the process is also called a "random field"

In our case, the sample space is the time axis $[0,\infty)$

Ref.: Doob 1953, 60+625; Møller/Waagepetersen 2004, 7-11

Prior distributions and prior processes

Boehringer Ingelheim

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 on the sample space.

To construct the neighbourhoods of the given distribution, we look at 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 interval of the partition.
We obtain a family of distributions around the given distribution on the sample space

The following properties can be proven:
- The family is conjugate to samples from the given distribution (also if independently censored)
- The distributions in the family are, with probability 1, discrete

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

Estimation of 1 survival curve

Boehringer Ingelheim

The Dirichlet process was applied successfully to the estimation of 1 survival curve with right-censoring.
The prior family of distributions is constructed to be centered around one exponential distribution

The relative weight of the prior family, relative to the information provided by the data, is described by a non-negative number, c
The Kaplan-Meier estimator can be seen as the limiting case for c = 0

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

Estimation of 1 survival curve

Boehringer Ingelheim

We follow up the example of Suzarla/Van Ryzin 1976:

Time [months]	Event / censoring
0.8	e
1.0	c
2.7	c
3.1	e
5.4	e
7.0	c
9.2	e
12.1	c

Estimation of 1 survival curve

Boehringer Ingelheim

The starting point is an exponential function

$\alpha(u,\infty) := c \bullet e^{-\theta u}$

whereby (u,∞) is the last element of a partition of $[0,\infty)$, c is a weighting parameter, and θ was selected to „fit the median to that of the K-M curve“ (in so far it is a double use of the data like in Empirical Bayes).

The Dirichlet process is constructed around this function – again, very difficult, we illustrate the result only

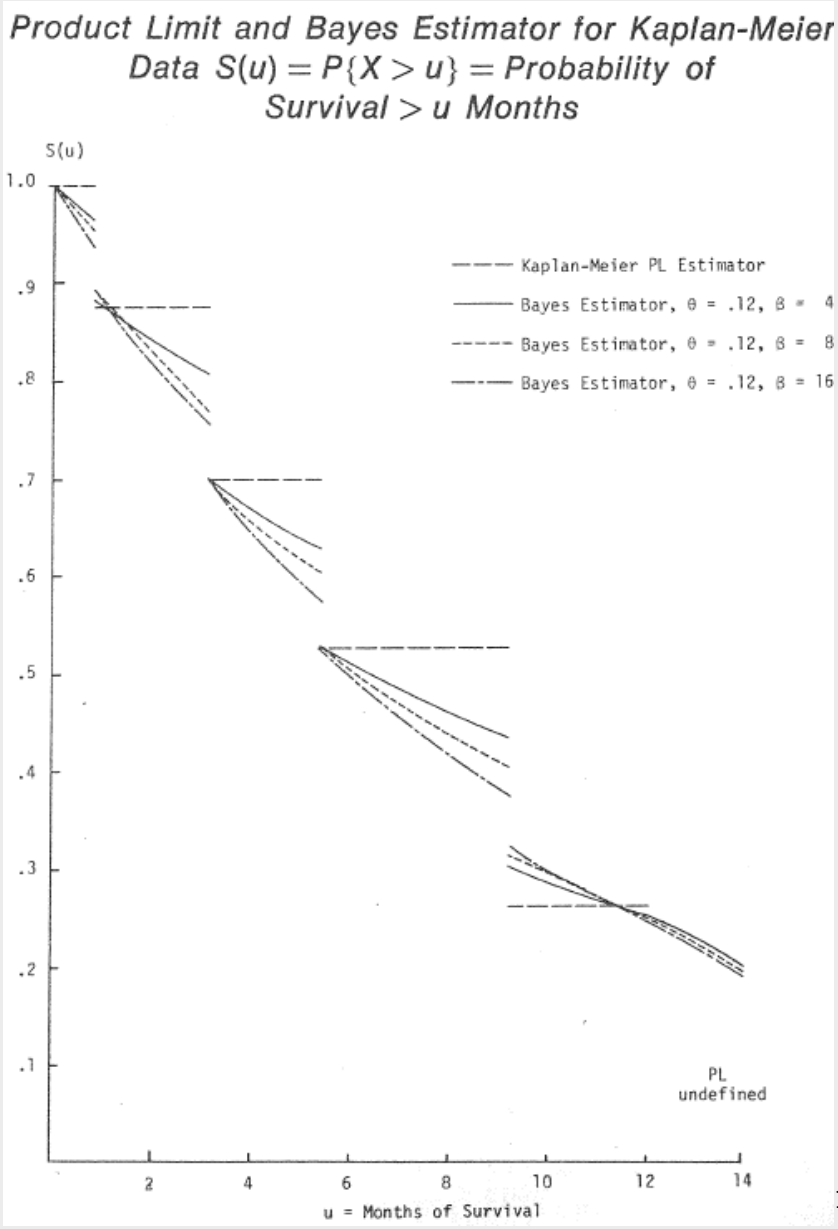
Estimation of 1 survival curve

Boehringer Ingelheim

Values of 0, 4, 8 and 16 for c (in the original called β) give the following posterior curves:

Jumps at event times become smaller with increasing c

Curves have bend points (hardly visible) at censoring times



Estimation of 1 survival curve

Boehringer Ingelheim

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 $1-F(t) = \exp(-\Lambda(t))$.

One has to select a basic hazard function $d\Lambda_0^*(t)$. It is then 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

$c \cdot d\Lambda_0^*(t)$, $c \cdot (1-d\Lambda_0^*(t))$

The existence is difficult to prove.

Ref.: Hjort, Ann.Stat. 1990

Estimation of 1 survival curve

Boehringer Ingelheim

Also the Beta process is conjugated to samples (possibly censored) from the corresponding event-time distribution

In the limit for c = 0, the estimated survival function becomes the Kaplan-Meier curve also here

Ref.: Hjort, Ann.Stat. 1990

Estimation of 1 survival curve

Boehringer Ingelheim

Summing up:

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 apparently “reasonably”

Ref.: Bernardo/Smith 1994, 177-181

Comparison of 2 survival curves: Cox model (with counting process)

Boehringer Ingelheim

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 $t_{n+1} > t_n$.

Comparison of 2 survival curves: Cox model (with counting process)

Boehringer Ingelheim

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

Comparison of 2 survival curves: Cox model (with counting process)

Boehringer Ingelheim

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 patient 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:

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

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

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

Comparison of 2 survival curves: Cox model (with counting process)

Boehringer Ingelheim

Parameters in the PH model $I_j(t) = Y_j(t) \cdot \lambda_0(t) \cdot e^{z_j\beta}$ 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 the double product
Product (j=1,...,n) Product (all i with $t_i \leq T_j$)

$$h_j(t_i)^{dN_j(t_i)} \cdot e^{-h_j(t_i)(t_i - t_{i-1})}$$

where T_j is the event or censoring time of patient j and h_j is the individual hazard.
This is, as function of the $dN_j(t_i)$, prop. to a Poisson distrib..

Ref.: Hellmich 2001

Comparison of 2 survival curves: Cox model (with counting process)

Boehringer Ingelheim

A prior distribution for $d\Lambda_0(t)$ (actually for the values of the piecewise constant function I(t)) is

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

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

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

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.

Comparison of 2 survival curves: Cox model (with counting process)

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, Lunn et al. 2013

Comparison of 2 survival curves: Cox model (with counting process)

Boehringer Ingelheim

Prior information:

$c_0 = 1$

$\lambda_0^*(t) = e^{-t/(t_n+1)}$, so that $\lambda_0^*(t_n) \approx 0.3$

Dependence from the prior information still needs to be investigated in greater detail, see e.g. Laud et al. 1998, p. 218-219

Comparison of 2 survival curves: Cox model (with counting process)

Boehringer Ingelheim

WinBUGS results:

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

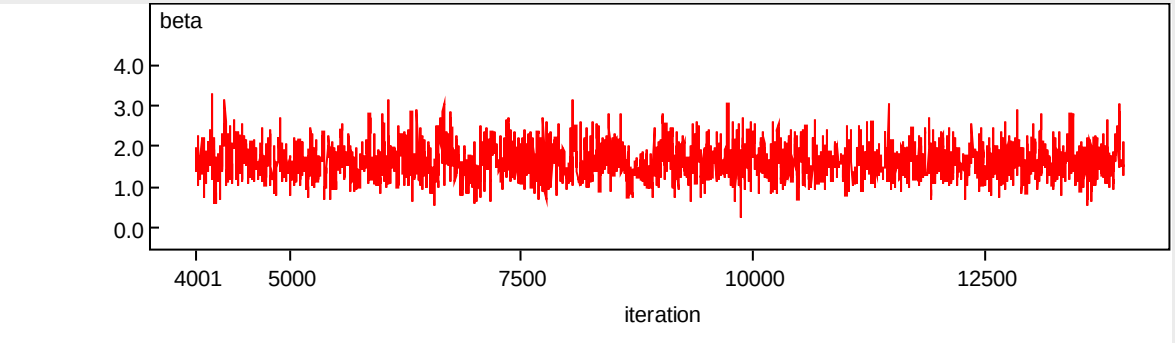
dL0 is the average hazard of both groups

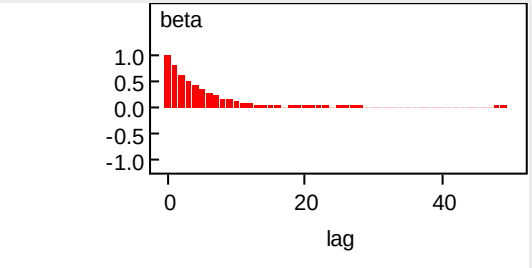
Similar results for the estimated surv. curves of each group

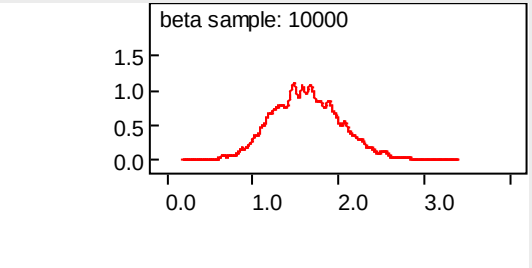
Comparison of 2 survival curves: Cox model (with counting process)

Boehringer Ingelheim

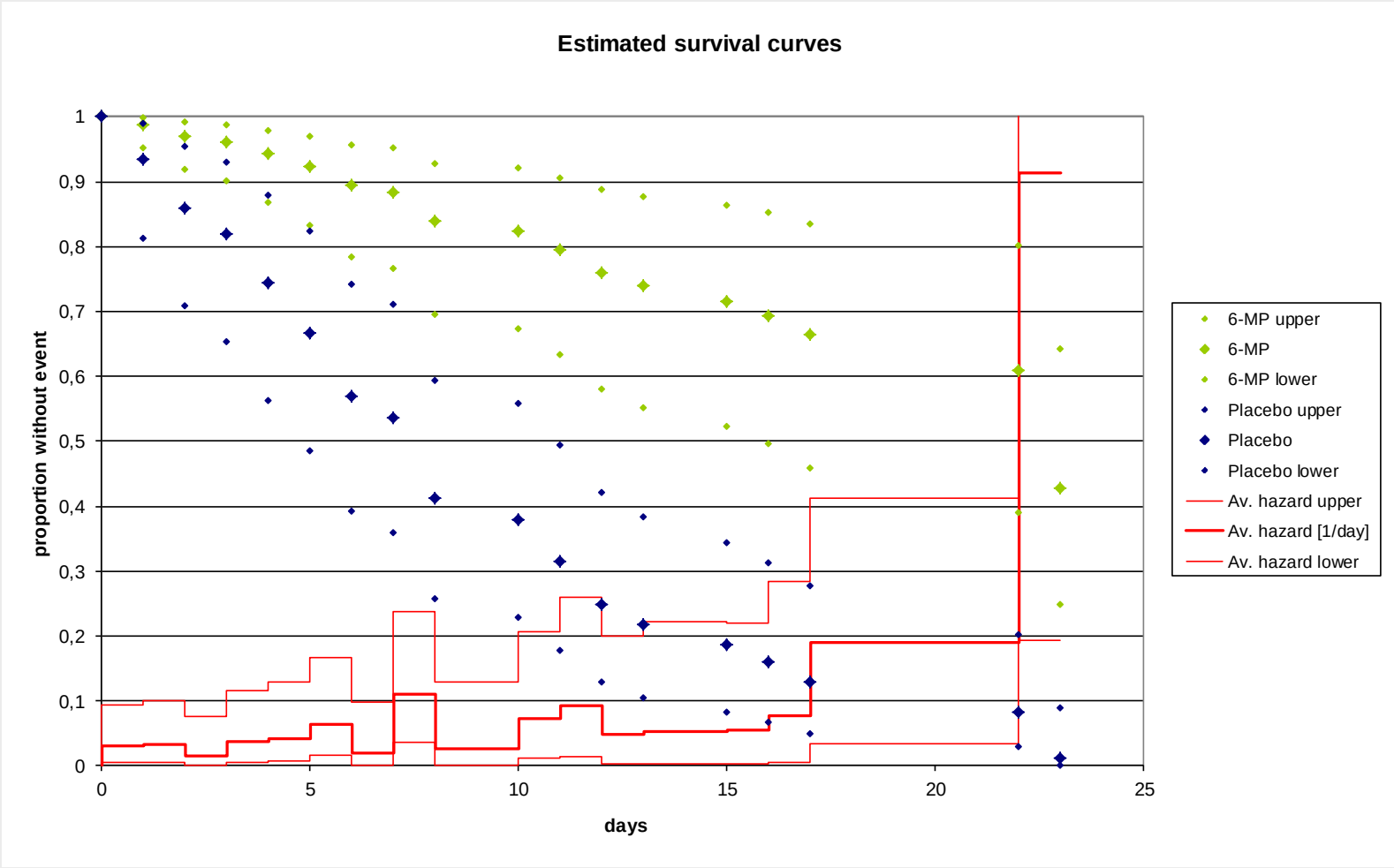
Graphs for the treatment difference parameter “beta”:







Comparison of 2 survival curves: Cox model (with counting process)



All 3 curves have distributions (vertical)

Discussion

Boehringer Ingelheim

In both cases, posterior results depend from the somewhat arbitrary constant c_0.
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.
It is unknown which prior processes could play the role of “reference prior” or of “Jeffreys prior”.

Discussion

Boehringer Ingelheim

What is the result of an investigation of a posterior distribution of a distribution?

We can no longer speak of a posterior plausible region for the parameter, we have to speak e.g. of the “posterior plausible extent of exponentiality”

The biological meaning of such a result is still obscure

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

Doob JL:
Stochastic Processes.
New York/London/Sydney: J. Wiley 1953

References

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

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

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

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

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

References

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

Laud PW, Damien P, Smith AFM:
Bayesian Nonparametric and Covariate Analysis of Failure Time Data.
In:
Dey D, Müller P, Sinha D (eds.):
"Practical Nonparametric and Semiparametric Bayesian Statistics".
New York / Berlin / Heidelberg: Springer-Verlag 1998, 213-225

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

References

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

Lunn D, Jackson C, Best N, Thomas A, Spiegelhalter D:
The BUGS Book.
Boca Raton/FL: CRC Press 2013

Boehringer Ingelheim

Questions?

Thank you