Semiparametric inference for survival time models using hierarchical mixture modeling with generalized gamma processes

Alessandra Guglielmi

alessandra.guglielmi@polimi.it

Dipartimento di Matematica, Politecnico di Milano, Milano (Italy)

joint with R. Argiento and A. Pievatolo

CNR-IMATI, Milano (Italy)

Aim of the paper

Survival times data (censored or not)

MODEL: Accelerated failure time (AFT) model (regression model on the $\log$ scale) from a Bayesian semiparametric point of view; the error is distributed nonparametrically as a species sampling mixture model (of densities on the positive reals) where the mixing measure P is a (normalized) generalized gamma measure

AIM: for simulated and real datasets we compute posterior inferences on the regression parameters and the survival times via MCMC (incorporating censoring)

Remark: Density estimation can be performed

The AFT model

Accelerated failure time model

$$\log T = -x'\beta + W$$

T: univariate survival time

$x = (x_1, \ldots, x_p)'$: fixed p-vector of covariates

$\beta = (\beta_1, \ldots, \beta_p)'$: p-vector of regression parameters

W, V errors

Equivalently, if $V := \mathrm{e}^W$:

$$T = \mathrm{e}^{-x'\beta} \cdot V$$

Bayesian semiparametric AFT models

$$\log T = -x'\beta + W \quad \text{or} \quad T = \mathrm{e}^{-x'\beta} \cdot V$$

Christensen & Johnson ('88): $V \sim$ Dirichlet process prior

Walker & Mallick ('99): $W \sim$ 0-median PT prior
Hanson & Johnson ('02): $W \sim$ mixture of 0-median PT
Kottas & Gelfand ('01), Gelfand & Kottas ('03): $W \sim$ semiparametric 0-median family of distributions *(scale mixture of split normals, skewness handled parametrically)*

Kuo & Mallick ('97): $V \sim$ DPM prior *(location mixture of normal kernels)*
Ghosh & Ghoshal ('06): $V \sim$ DPM prior *(scale mixture of Weibull kernels)*
Hanson ('06): $V \sim$ DPM prior *(gamma densities, mixed both over the scale and the shape)*
Argiento, Guglielmi, Pievatolo ('07): $V \sim$ DPM or NIG-mixture prior *(gamma densities, mixed over the scale and the shape)*

Our model

$T_1, \ldots, T_n$ survival times, $x_i = (x_{i1}, \ldots, x_{ip})'$ covariate vector

$$T_i = \mathrm{e}^{-x_i'\beta} \cdot V_i, \; i = 1, \ldots, n,$$

$V_i|\theta_i \stackrel{ind}{\sim} k(\cdot; \theta_i)$ family of densities on $\mathbb{R}^+$ $\quad \Theta \subset \mathbb{R}^s$ parametric space

$\theta_i|G \stackrel{iid}{\sim} G,$

$G \sim q,$ $\quad G$ is a (normalized) generalized gamma measure on Θ

$\beta \perp G, \; \beta \sim \pi(\beta)$

Remark

$V_1, \ldots, V_n|G \stackrel{iid}{\sim} f(v; G) = \int_\Theta k(v; \theta) G(d\theta)$

Generalized gamma measures (Brix, 1999)

μ random measure on $(\Theta, \mathcal{B}(\Theta))$, $\Theta \subset \mathbb{R}^s$
$\sigma \in (0,1]$, $\tau \geq 0$
$\kappa(\cdot)$ non-negative diffuse (finite) measure on Θ (Θ any Polish space),
μ is a *generalized gamma measure* if

- μ is *completely random* (Kingman, 1993), *i.e.* $\mu(B_1), \ldots, \mu(B_k)$ are mutually independent if $B_1, \ldots, B_k$ are disjoint
- for any $B \in \mathcal{B}(\Theta)$, $\mu(B)$ has mgf

$$\mathbb{E}(e^{-s\mu(B)}) = exp\Big(- \frac{\kappa(B)}{\sigma}[(\tau+s)^\sigma - \tau^\sigma]\Big),\ s \geq 0$$
$$(\mu(B) \sim G(\sigma, \kappa(B), \tau))$$

$0 < \sigma < 1$: $G(\sigma, \kappa(B), \tau)$ is the natural exponential family generated by the positive stable law ($\tau = 0$), and μ is the σ-stable process for $\tau = 0$

$\sigma = 1$: μ degenerates on $\kappa(\cdot)$

$\sigma \to 0$: $\mathbb{E}(e^{-s\mu(B)}) = \left(\frac{\tau}{\tau+s}\right)^{\kappa(B)}$ and μ is the gamma random measure

Generalized gamma measures

Theorem (Brix, 1999)

1. $\mu(B) = \int_{[0,+\infty)} yN(dy, B),$

   N a Poisson random measure on $[0, +\infty) \times \Theta$

   ν intensity measure:

$$\nu(A \times B) = \frac{\kappa(B)}{\Gamma(1-\sigma)} \int_A s^{-\sigma-1} \mathrm{e}^{-\tau s} ds = \kappa(B) \int_A \rho(ds),$$

$$A \in \mathcal{B}([0, +\infty)), B \in \mathcal{B}(\Theta)$$

2. μ has no fixed atoms (since κ is diffuse), *i.e.* $\mathbb{P}(\mu(\{x\}) > 0) = 0 \; \forall x$
3. μ is almost surely purely atomic

Generalized gamma random probability measures

G random probability (distribution) built from a generalized gamma random measure μ with parameters $(\sigma, \kappa(\cdot), \tau)$ according to a standard construction via normalization of completely random measures
Kingman (1993), Pitman (2003), James (2005), Antonio Lijoi's talk

Since $\int_{[0,+\infty)\times B} \min(s,1)\nu(ds,dy) = \kappa(B) \int_{[0,+\infty)} \min(s,1)\rho(ds) < +\infty$,

$$\mathbb{P}(\mu(\Theta) =: T < +\infty) = 1,$$

so that

$$G(\cdot) := \frac{\mu(\cdot)}{T} \text{ is a random probability measure on } \Theta$$

$G \sim GG(\sigma, \kappa(\Theta)G_0(\cdot), \tau)$ generalized gamma rpm, $G_0(\cdot) := \kappa(\cdot)/\kappa(\Theta)$

Remark: this parameterization is not unique, *i.e.*
$(\sigma, \kappa(\Theta)G_0(\cdot), \tau)$ and $(\sigma, s^\sigma \kappa(\Theta)G_0(\cdot), \tau/s)$ (for any $s > 0$) yield the same distribution for G (see Pitman, 2003, "scaling property")

Generalized gamma rpm

Pitman (1996, 2003)
$G \sim GG(\sigma, \kappa(\Theta)G_0(\cdot), \tau)$, $0 \le \sigma \le 1$, $\kappa(\Theta) > 0$, $\tau \ge 0$

$$G = \sum_{i=1}^{+\infty} P_i \delta_{X_i}, \qquad (P_i) \perp (X_i), \quad (X_i) \overset{iid}{\sim} G_0,$$

$P_i := \frac{J_i}{T}$, (J_i) jump times of a Poisson process on $[0, +\infty)$ with Levy density

$$\rho(ds) = \frac{1}{\Gamma(1-\sigma)} s^{-\sigma-1} \mathrm{e}^{-\tau s} ds, \quad T = \sum_i J_i$$

$(\tilde{P}_i)$ (ranked) has the Poisson-Kingman distribution with Levy density $\rho(ds)$

- G is a homogeneous Poisson-Kingman rpm with Levy density $\rho(ds)$
- G is a species sampling model $P(\cdot) = \sum_k p_k \delta_{Z_k} + (1 - \sum_k p_k) H(\cdot)$ ((p_k) are positive random weights, $\sum p_k \le 1$, (p_k) and (Z_k) independent, $Z_k \overset{iid}{\sim} H(\cdot)$ non-atomic)

Generalized gamma rpm

Pitman (2003)
Sampling from G induces a random partition $\mathbf{\Pi}$ on the positive integers $\mathbb{N}$
$\theta_1, \ldots, \theta_n | G \overset{iid}{\sim} G$ a.s. discrete: ties among (θ_i)

Generalized gamma rpm

Pitman (2003)
Sampling from G induces a random partition $\mathbf{\Pi}$ on the positive integers $\mathbb{N}$

$\theta_1, \ldots, \theta_n | G \overset{iid}{\sim} G$ a.s. discrete: ties among (θ_i)

If $\mathbf{\Pi_n}$ is the restriction of $\mathbf{\Pi}$ to $\{1, \ldots, n\}$:

$$\mathcal{L}(\theta_1, \ldots, \theta_n) \leftrightarrow \mathcal{L}(\mathbf{\Pi_n}, \theta_1^*, \ldots, \theta_{k(n)}^*)$$

$\theta_1^*, \ldots, \theta_{k(n)}^*$ distinct values in $(\theta_1, \ldots, \theta_n)$

$\pi_n = \{C_1, \ldots, C_k\}$, $C_j = \{i : \theta_i = \theta_j^*\}$, $n_j := \#C_j \geq 1$, $\sum_1^{k(n)} n_j = n$,

$$\mathbb{P}(\mathbf{\Pi_n} = \pi_n, \theta_1^* \in B_1, \ldots, \theta_{k(n)}^* \in B_{k(n)}) = \mathcal{L}(\mathbf{\Pi_n} = \pi_n) \cdot \prod_{j=1}^{k(n)} G_0(B_j)$$

$$(p \text{ symmetric non-negative }) \qquad = p(n_1, \ldots, n_k) \prod_{j=1}^{k(n)} G_0(B_j),$$

p *exchangeable partition probability function* (EPPF) determined by $\mathbf{\Pi}$

Generalized gamma rpm

Pitman (2003)
Sampling from G induces a random partition $\mathbf{\Pi}$ on the positive integers $\mathbb{N}$
$\theta_1, \ldots, \theta_n | G \overset{iid}{\sim} G$ a.s. discrete: ties among (θ_i)

Moreover:

$$\mathbb{P}(\theta_{n+1} \in B | \theta_1, \ldots, \theta_n) = w_{0,n} G_0(B) + \sum_1^k w_{j,n} \delta_{\theta_j^*}(B)$$

$$w_{0,n} = \frac{p(n_1, \ldots, n_k, 1)}{p(n_1, \ldots, n_k)}, \; w_{j,n} = \frac{p(n_1, \ldots, n_{j+1}, \ldots, n_k)}{p(n_1, \ldots, n_k)}.$$

Generalized gamma rpm

$$G \sim GG(\sigma, \kappa(\Theta)G_0, \tau) \quad 0 < \sigma \leq 1, \kappa(\Theta) > 0, \tau \geq 0$$

$\sigma \to 0$: G Dirichlet process with *parameter* $\kappa(\Theta)G_0$

$\sigma = 1/2$: G normalized inverse-gaussian (NIG) process (Lijoi *et al.*, 2005)

$\sigma = 1$: G is degenerate on G_0

$\tau = 0$: if $0 < \sigma < 1$, G Poisson-Dirichlet process with two parameters $(\sigma, 0)$

Generally $\mathcal{L}(G(B_1), \ldots, G(B_k))$ is not available in closed analytic form

Properties of generalized gamma rpms

$$\mathbb{E}\left(G(B)\right) = G_0(B) \qquad \mathrm{Var}\left(G(B)\right) = G_0(B)(1 - G_0(B))\mathcal{I}(\sigma, \kappa(\Theta), \tau)$$

$$\mathrm{Cov}(G(B_1), G(B_2)) = \Big(G_0(B_1 \cap B_2) - G_0(B_1)G_0(B_2)\Big)\mathcal{I}(\sigma, \kappa(\Theta), \tau)$$

$$\mathcal{I}(\sigma, \kappa(\Theta), \tau) := (1-\sigma)\Big(1 - \left(\frac{\kappa(\Theta)\tau^\sigma}{\sigma}\right)^{1/\sigma} \exp\left(\frac{\kappa(\Theta)\tau^\sigma}{\sigma}\right)\Big)\Gamma\left(-\frac{1}{\sigma} + 1, \frac{\kappa(\Theta)\tau^\sigma}{\sigma}\right)\Big),$$

where $\Gamma(\alpha, x) := \int_x^{+\infty} \mathrm{e}^{-t} t^{\alpha-1} dt$. If $\eta := \frac{\kappa(\Theta)\tau^\sigma}{\sigma}$, then

$$\mathcal{I} = \mathcal{I}(\sigma, \eta) = \Big(\frac{1}{\sigma} - 1\Big)\eta^{1/\sigma}\mathrm{e}^{\eta}\Gamma\Big(-\frac{1}{\sigma}, \eta\Big)$$

$\forall 0 \leq \sigma < 1 \quad \mathcal{I}(\sigma, \eta) \downarrow$ as η increases and $\forall \eta > 0 \quad \mathcal{I}(\sigma, \eta) \downarrow$ as σ increases

Re-parameterization: $G \sim GG(\sigma, \eta, G_0),\ 0 < \sigma \leq 1, \eta \geq 0$

Properties of generalized gamma rpms

Remark:

$$\theta_1, \ldots, \theta_n | G \overset{iid}{\sim} G, \; G \sim GG(\sigma, \eta, G_0)$$

$\Rightarrow \mathcal{L}(G|\theta_1, \ldots, \theta_n)$ is NOT a GG rpm

Description of the posterior: James (2002, 2005)

Predictive distributions

$\theta_1, \ldots, \theta_n | G \stackrel{iid}{\sim} G$ and $G \sim GG(\sigma, \eta, G_0)$, $0 < \sigma \leq 1$, $\eta \geq 0$

$\theta_1^*, \ldots, \theta_k^*$ distinct observations, $(n_1, \ldots, n_k)$ multiplicities $(\sum_j n_j = n)$,

$$P(\theta_{n+1} \in B | \theta_1, \ldots, \theta_n) = w_0(n, k; \sigma, \eta) G_0(B) + w_1(n, k; \sigma, \eta) \sum_{j=1}^{k} (n_j - \sigma) \delta_{\theta_j^*}(B)$$

Predictive distributions

$\theta_1, \ldots, \theta_n | G \overset{iid}{\sim} G$ and $G \sim GG(\sigma, \eta, G_0)$, $0 < \sigma \leq 1$, $\eta \geq 0$

$\theta_1^*, \ldots, \theta_k^*$ distinct observations, $(n_1, \ldots, n_k)$ multiplicities ($\sum_j n_j = n$),

$$P(\theta_{n+1} \in B | \theta_1, \ldots, \theta_n) = w_0(n, k; \sigma, \eta) G_0(B) + w_1(n, k; \sigma, \eta) \sum_{j=1}^{k} (n_j - \sigma) \delta_{\theta_j^*}(B)$$

$$w_0(n, k; \sigma, \eta) = \frac{2}{n} \sigma \eta \frac{\epsilon_{n+1,k+1}}{\epsilon_{n,k}} = \frac{\sigma}{n} \frac{\sum_{i=0}^{n} \binom{n}{i} (-1)^i \eta^{i/\sigma} \Gamma(k+1-i/\sigma; \eta)}{\sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \eta^{i/\sigma} \Gamma(k-i/\sigma; \eta)}$$

$$w_1(n, k; \sigma, \eta) = \frac{2}{n} \frac{\epsilon_{n+1,k}}{\epsilon_{n,k}} = \frac{1}{n} \frac{\sum_{i=0}^{n} \binom{n}{i} (-1)^i \eta^{i/\sigma} \Gamma(k-i/\sigma; \eta)}{\sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \eta^{i/\sigma} \Gamma(k-i/\sigma; \eta)}$$

Cerquetti (2007) Lijoi et al. (2007)

$$\epsilon_{n,k} := \int_0^{+\infty} \frac{x^{n-1} \mathrm{e}^{-\eta(1+2x)^\sigma}}{(1+2x)^{n-k\sigma}} dx$$

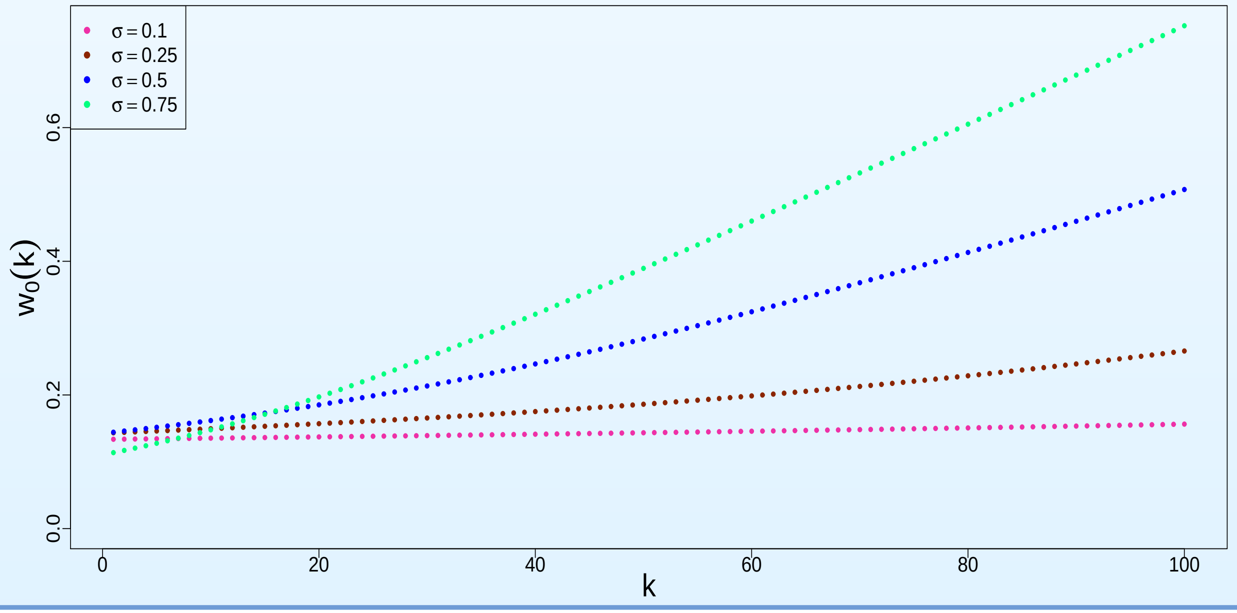
Predictive distributions

$\theta_1, \ldots, \theta_n | G \overset{iid}{\sim} G$ and $G \sim GG(\sigma, \eta, G_0)$, $0 < \sigma \leq 1$, $\eta \geq 0$

$\theta_1^*, \ldots, \theta_k^*$ distinct observations, $(n_1, \ldots, n_k)$ multiplicities ($\sum_j n_j = n$),

$$P(\theta_{n+1} \in B | \theta_1, \ldots, \theta_n) = w_0(n, k; \sigma, \eta) G_0(B) + w_1(n, k; \sigma, \eta) \sum_{j=1}^{k} (n_j - \sigma) \delta_{\theta_j^*}(B)$$

Plot of $w_0(100, k; \sigma, \eta)$ for some values of σ



Prior distributions for the # of distinct observations

K_n number of distinct observations in the sample $(\theta_1, \ldots, \theta_n)$ from $G \sim GG(\sigma, \eta, G_0)$

$$\mathbb{P}(K_n = k) = \mathbf{S}(n,k;\sigma)\frac{2^n e^{\eta}\eta^k}{\Gamma(n)}\epsilon_{n,k} = \mathbf{S}(n,k;\sigma)\frac{e^{\eta}}{\sigma\Gamma(n)}\sum_{i=0}^{n-1}\binom{n-1}{i}(-1)^i\eta^{\frac{i}{\sigma}}\Gamma(k-\frac{i}{\sigma};$$

$\mathbf{S}(n,k;\sigma)$ generalized Stirling numbers of the first kind

The model

$$T_i = \mathrm{e}^{-x_i'\beta} \cdot V_i,\ i = 1, \ldots, n,$$

$$V_i|\theta_i \overset{ind}{\sim} k(\cdot;\theta_i) \qquad \Theta \subset \mathbb{R}^2 \text{ parametric space}$$

$$\theta_i|G \overset{iid}{\sim} G,$$

$$G \sim GG(\sigma, \eta, G_0), \quad G_0(\cdot) := \mathbb{E}(G(\cdot)),\ G_0 \text{ diffuse}$$

$$\beta \perp G,\ \beta \sim \pi(\beta)$$

$$(\sigma, \eta) \sim \pi(\sigma, \eta)$$

$k(\cdot;\vartheta_1,\vartheta_2)$ gamma density with mean ϑ_1/ϑ_2

$G_0 \quad gamma(\omega_1,\gamma_1) \times gamma(\omega_2,\gamma_2)$

Univariate marginal distribution of the error V

If $\omega_1 = \omega_2 = 1$: $f_V(v) = \frac{\gamma_1 \gamma_2}{v(v+\gamma_2)(\gamma_1 + \log(\frac{v+\gamma_2}{v}))^2}$ $(v > 0)$, but

$EV = +\infty$

Univariate marginal distribution of the error V

If $\omega_1 = \omega_2 = 1$: $f_V(v) = \dfrac{\gamma_1 \gamma_2}{v(v+\gamma_2)(\gamma_1 + \log(\frac{v+\gamma_2}{v}))^2}$ $(v > 0)$, but

$EV = +\infty$

median regression model

Prior information: median and IQR as functions of hyperparameters (γ_1, γ_2)

Univariate marginal distribution of the error V

If $\omega_1 = \omega_2 = 1$: $f_V(v) = \frac{\gamma_1\gamma_2}{v(v+\gamma_2)(\gamma_1 + \log(\frac{v+\gamma_2}{v}))^2}$ $(v > 0)$, but

$EV = +\infty$

median regression model

Prior information: median and IQR as functions of hyperparameters (γ_1, γ_2)

If $\omega_2 > j$: $\mathbb{E}(V^j) < +\infty$; in particular

$$\mathbb{E}(V) = \frac{\omega_1\gamma_2}{(\omega_2 - 1)\gamma_1}, \ \omega_2 > 1$$

Univariate marginal distribution of the error V

If $\omega_1 = \omega_2 = 1$: $f_V(v) = \dfrac{\gamma_1\gamma_2}{v(v+\gamma_2)(\gamma_1 + \log(\frac{v+\gamma_2}{v}))^2}$ $(v > 0)$, but

$EV = +\infty$

median regression model

Prior information: median and IQR as functions of hyperparameters (γ_1, γ_2)

If $\omega_2 > j$: $\mathbb{E}(V^j) < +\infty$; in particular

$$\mathbb{E}(V) = \frac{\omega_1\gamma_2}{(\omega_2 - 1)\gamma_1}, \ \omega_2 > 1$$

mean regression model

Prior information: first j moments as functions of hyperparameters (γ_1, γ_2)

Posterior estimates

Since $T_i = \mathrm{e}^{-x'_i\beta} \cdot V_i$, $i = 1, \ldots, n$

$$\hat{f}_{T_{n+1}}(t|T_1, \ldots, T_n) \longleftrightarrow \hat{f}_{V_{n+1}}(v|V_1, \ldots, V_n, \underline{\beta})$$

$$\hat{\underline{\beta}} = \mathbb{E}(\underline{\beta}|T_1, \ldots, T_n) \longleftrightarrow \mathbb{E}(\underline{\beta}|V_1, \ldots, V_n)$$

Moreover

$$\hat{f}_{V_{n+1}}(v|V_1, \ldots, V_n, \underline{\beta}) = \int \Big(\int_\Theta k(v;\theta) G(d\theta) \Big) \mathcal{L}(dG|\theta_1, \ldots, \theta_n, \underline{\beta}) \mathcal{L}(d\underline{\theta}|V_1, \ldots, V_n, \underline{\beta}$$

$$\text{integrating out } G = \int f_{V_{n+1}}(v|\underline{\theta}, \underline{\beta}) \mathcal{L}(d\underline{\theta}|\underline{V}, \underline{\beta}) \qquad (1)$$

$\mathcal{L}(d\underline{\theta}|\underline{V})$ and $\mathcal{L}(d\underline{\beta}|\underline{V})$ in (1): via a Polya urn Gibbs sampler

$$f_{V_{n+1}}(v|\underline{\theta}, \underline{\beta}) = w_0(n,k;\sigma,\eta) f_V(v) + w_1(n,k;\sigma,\eta) \sum_{j=1}^{k} (n_j - \sigma)\, k(v;\theta_j^*)$$

prior marginal distribution of V : $f_V(v) = \int_{\mathbb{R}^+ \times \mathbb{R}^+} k(v;\vartheta_1,\vartheta_2) G_0(d\vartheta_1, d\vartheta_2)$

Algorithms for posterior estimates

$$\hat{f}_{V_{n+1}}(v|V_1,\ldots,V_n,\underline{\beta}) = \int\Big(\int_\Theta k(v;\theta)G(d\theta)\Big)\mathcal{L}(dG|\underline{\theta},\underline{\beta})\mathcal{L}(d\underline{\theta}|\underline{V},\underline{\beta}) \quad (2)$$

(2) is fundamental in the GWCR algorithm (Ishwaran-James, 2003), when independent draws from $\mathcal{L}(dG|\underline{\theta})$ are available: Importance sampler algorithm (iid draws from the importance distribution)

Polya urn sequential importance sampler: sequential draw for $(\theta_1,\ldots\theta_n)$

$$\theta_1 \sim \mathcal{L}(d\theta_1|V_1)$$

$$\theta_{r+1} \sim w_0^*(r,k(r);\sigma,\eta)\mathcal{L}(d\theta_{r+1}|V_{r+1}) + \sum_{j=1}^{k(r)} w_1^*(r,k(r);\sigma,\eta)\delta_{\theta^*_{j,r}}, r = 2,\ldots,n$$

Many w_0 and w_1 must be computed: Polya urn SIS is very *expensive*

Accelerated Polya urn Gibbs sampler: (i) integrating out G (ii) the update rule for $\underline{\beta}$ is simple

Prior for the regression parameters

Reparameterization: $\alpha_j = \mathrm{e}^{\beta_j}, j = 1, \ldots, p$

$\alpha_j \stackrel{ind}{\sim} \text{gamma}(\alpha^*_{1j}, \alpha^*_{2j}), j = 1, \ldots, p$

the full conditional posterior distributions of the α_j's associated to a binary covariate are still gamma

Polya urn Gibbs sampler

Posterior estimates:

$$\hat{f}_{V_{n+1}|\underline{V}}(v \mid data) = \frac{1}{J}\sum_{j=1}^{J} f_{V_{n+1}}(v \mid data, \underline{\theta}^{(j)})$$

$$\hat{S}_{T_{n+1}|\underline{x},\underline{T}}(t \mid x, data) = \frac{1}{J}\sum_{j=1}^{J} S_{V_{n+1}}(\prod_{i=1}^{p}(\alpha_i^{(j)})^{x_i} t \mid data, \underline{\theta}^{(j)})$$

STEP 1 : draw $\underline{\theta}^{(j+1)}$ from $\mathcal{L}(d\underline{\theta}|\underline{\alpha}^{(j)}, data)$ via a

Polya urn scheme with an *acceleration step*:
state space: $(\underline{c}, \underline{\theta})$ where $\underline{\theta} = (\theta_1, \ldots, \theta_n)$ and $\underline{c} = (c_1, \ldots, c_n)$ labels from 1 to $k(n)$ (# distinct value in $\underline{\theta}$) such that $c_i = c_j \Leftrightarrow \theta_i = \theta_j$
$\underline{c} \overset{\text{bijection}}{\longleftrightarrow} \Pi_n = \{C_1, \ldots, C_{k(n)}\}, C_j := \{i : \theta_i = \theta_j^*\} = \{i : c_i = j\}$

MCMC algorithm

Completely observed data

STEP 1 (a): update pairs (c_i, θ_i), for $i = 1, \ldots, n$ from $[(c_i, \theta_i) \mid c_{-i}, \underline{\theta}_{-i}, data]$

$$= [\theta_i \mid c_i, c_{-i}, \underline{\theta}_{-i}, data] \times [c_i \mid c_{-i}, \underline{\theta}_{-i}, data]$$

generalized Polya urn scheme

MCMC algorithm

Completely observed data

STEP 1 (a): update pairs (c_i, θ_i), for $i = 1, \ldots, n$ from $[(c_i, \theta_i) \mid c_{-i}, \underline{\theta}_{-i}, data]$

$$= [\theta_i \mid c_i, c_{-i}, \underline{\theta}_{-i}, data] \times [c_i \mid c_{-i}, \underline{\theta}_{-i}, data]$$

generalized Polya urn scheme

STEP 1 (b): block update the sub-vectors of $\underline{\theta}$ formed by elements with the same value (equivalent to updating the θ_j^* values): acceleration step

Remark: in (a), when $\omega_1 = \omega_2 = 1$, f_V is easy to evaluate (conjugate prior); otherwise: substitute G_0 in f_V with the empirical distribution of a random sample of size m from G_0 (augmentation step)

MCMC algorithm

Completely observed data

STEP 1 (a): update pairs (c_i, θ_i), for $i = 1, \ldots, n$ from $[(c_i, \theta_i) \mid c_{-i}, \underline{\theta}_{-i}, data]$

$$= [\theta_i \mid c_i, c_{-i}, \underline{\theta}_{-i}, data] \times [c_i \mid c_{-i}, \underline{\theta}_{-i}, data]$$

generalized Polya urn scheme

STEP 1 (b): block update the sub-vectors of $\underline{\theta}$ formed by elements with the same value (equivalent to updating the θ_j^* values): acceleration step

Remark: in (a), when $\omega_1 = \omega_2 = 1$, f_V is easy to evaluate (conjugate prior); otherwise: substitute G_0 in f_V with the empirical distribution of a random sample of size m from G_0 (augmentation step)

STEP 2: generate $\underline{\alpha}^{(j+1)}$ from $\mathcal{L}(d\underline{\alpha}|\underline{\theta}^{(j+1)}, data)$: for the α_j's not corresponding to binary covariates a Metropolis step is required

similar to Neal (2000)

MCMC algorithm

STEP 1 (a): update pairs (c_i, θ_i), for $i = 1, \ldots, n$ from $[(c_i, \theta_i) \mid c_{-i}, \underline{\theta}_{-i}, data]$

$$= [\theta_i \mid c_i, c_{-i}, \underline{\theta}_{-i}, data] \times [c_i \mid c_{-i}, \underline{\theta}_{-i}, data]$$

generalized Polya urn scheme

STEP 1 (b): block update the sub-vectors of $\underline{\theta}$ formed by elements with the same value (equivalent to updating the θ_j^* values): acceleration step

Remark: in (a), when $\omega_1 = \omega_2 = 1$, f_V is easy to evaluate (conjugate prior); otherwise: substitute G_0 in f_V with the empirical distribution of a random sample of size m from G_0 (augmentation step)

STEP 2: generate $\underline{\alpha}^{(j+1)}$ from $\mathcal{L}(d\underline{\alpha}|\underline{\theta}^{(j+1)}, data)$: for the α_j's not corresponding to binary covariates a Metropolis step is required

similar to Neal (2000)

STEP 3: augmentation step for censored data - unobserved survival times are sampled one at a time from their full conditional distribution truncated at their censoring points

Choice of (σ, η)

Simulated dataset for density estimation

$$V_i|\theta_i \overset{ind}{\sim} k(\cdot;\theta_i),\ i = 1,\dots,n$$

$$\theta_i|G \overset{iid}{\sim} G,$$

$$G \sim GG(\sigma,\eta,G_0), \quad G_0(\cdot) := \mathbb{E}(G(\cdot)),\ G_0 \text{ diffuse} \tag{3}$$

$$(\sigma,\eta) \sim \pi(\sigma,\eta) \quad (\sigma,\eta) \text{ control the variance of } G \text{ from the mean } G_0$$

(σ, η) fixed or random ?

Choice of (σ, η)

Simulated dataset for density estimation

$$V_i|\theta_i \overset{ind}{\sim} k(\cdot;\theta_i),\ i = 1,\dots,n$$

$$\theta_i|G \overset{iid}{\sim} G,$$

$$G \sim GG(\sigma,\eta,G_0), \quad G_0(\cdot) := \mathbb{E}(G(\cdot)),\ G_0 \text{ diffuse} \qquad (3)$$

$$(\sigma,\eta) \sim \pi(\sigma,\eta) \quad (\sigma,\eta) \text{ control the variance of } G \text{ from the mean } G_0$$

(σ, η) fixed or random ?

test the fit of the $GG(\sigma, \eta, G_0)$-mixture w.r.t. (σ, η): comparison between

$$\mathcal{M}_0 : V_1,\dots,V_n \overset{iid}{\sim} f_V(\cdot) = \int k(\cdot;\theta)G_0(d\theta) \qquad \mathcal{M}_1 : GG(\sigma,\eta,G_0) - \text{mixture } (3)$$

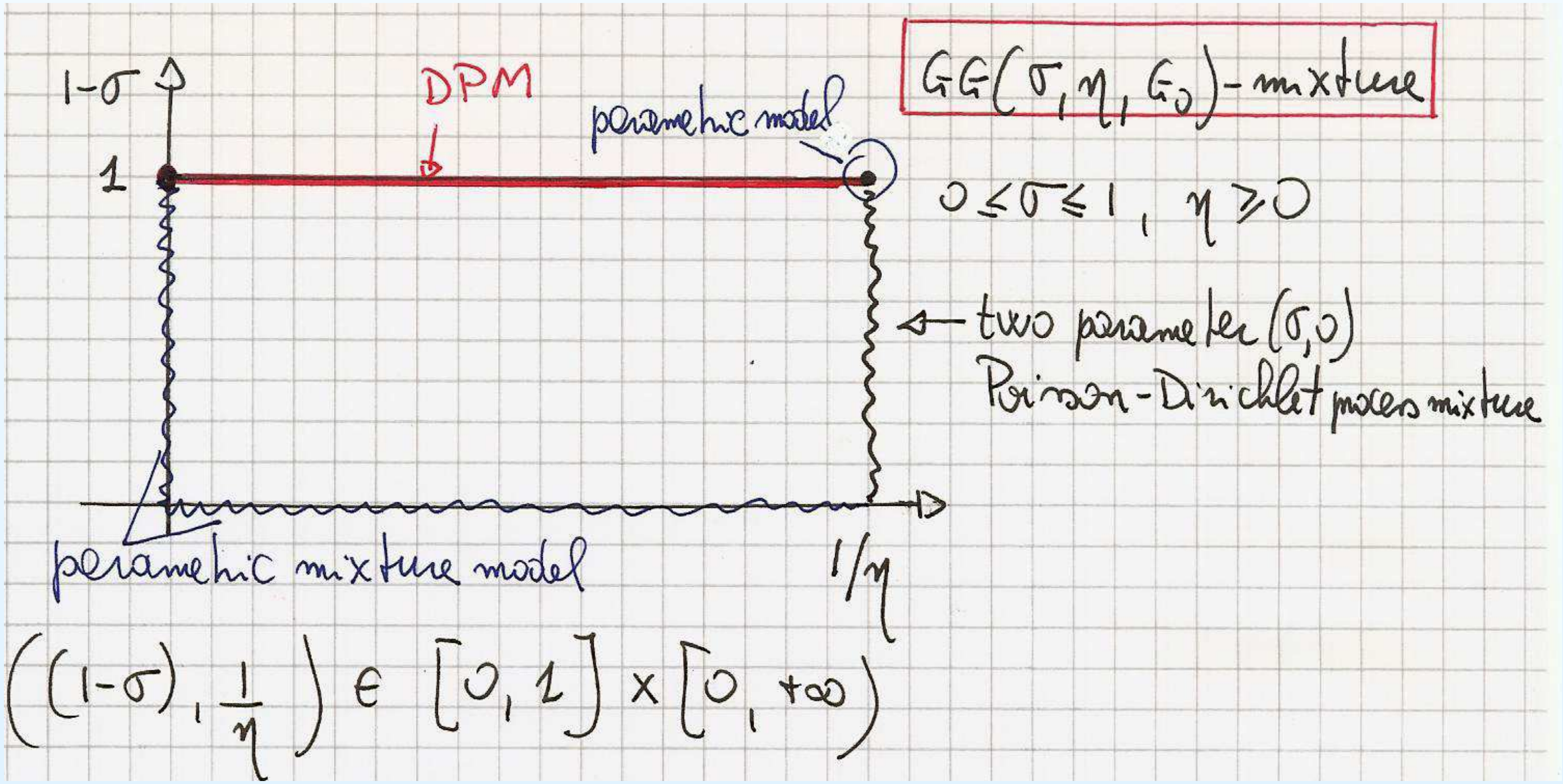
((3) when $\sigma = 1$; parametric mixture)

$\mathcal{M}_1$ is *centered* on $\mathcal{M}_0$, *enlarging* it by adding extra parameters (σ, η) which control the variance from the centering distribution G_0

Bayes factors for $GG(\sigma, \eta, G_0)$-mixtures

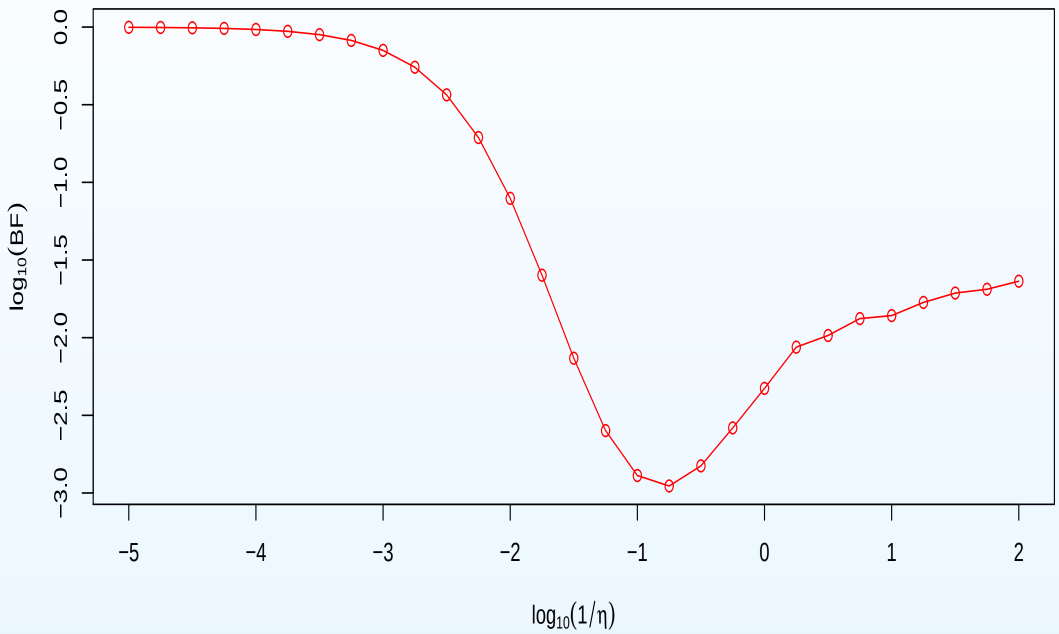
$$BF(\underline{v};\sigma,\eta) := \frac{m_0(\underline{v})}{m_1(\underline{v};\sigma,\eta)}$$

$\min_{(\sigma,\eta)} BF(\underline{v};\sigma,\eta) =: BF(\underline{v};\hat{\sigma},\hat{\eta})$ indicates the parameters (σ,η) most favorable to the alternative, a smaller minimum indicating a better fit for the corresponding nonparametric mixture



Bayes factors for $GG(\sigma, \eta, G_0)$-mixtures

Plot of $\log_{10}(BF(\underline{v}; 0, \eta))$ **as a function of** $log_{10}(\frac{1}{\eta})$



For any $0 \leq \sigma < 1$:

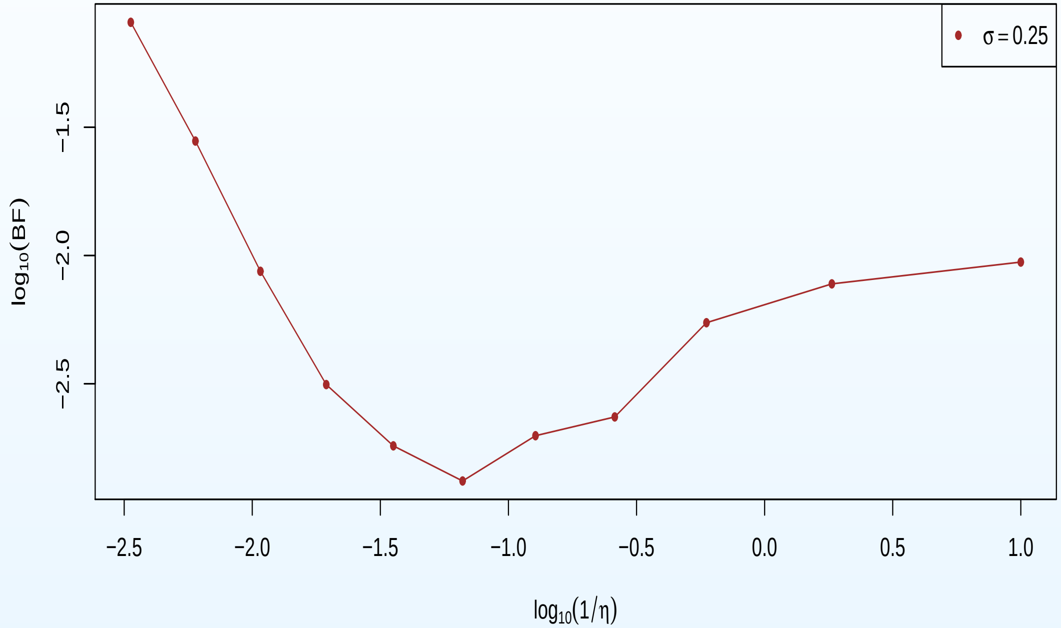
$\frac{1}{\eta} \rightarrow 0$: $\mathrm{Var}(G(B)) \rightarrow 0$ and G"$\rightarrow$"G_0 $\Rightarrow m_1(\underline{v}) \simeq m_0(\underline{v})$ and $BF(\underline{v}) \simeq 1$

$\frac{1}{\eta} \rightarrow +\infty$: $(\tau \rightarrow 0)$ G"$\rightarrow$" Pois-Dir process 2 parameters $(\sigma, 0)$, $BF(\underline{v}) \rightarrow h(\sigma)$

BFs between these extremes: (i) increasing functions of $\frac{1}{\eta}$ or (ii) first decreasing and then increasing

Bayes factors for $GG(\sigma, \eta, G_0)$-mixtures

Plot of $\log_{10}(BF(\underline{v}; 0.25, \eta))$ **as a function of** $log_{10}(\frac{1}{\eta})$



For any $0 \leq \sigma < 1$:

$\frac{1}{\eta} \to 0$: $\text{Var}(G(B)) \to 0$ and G"$\to$"$G_0 \Rightarrow m_1(\underline{v}) \simeq m_0(\underline{v})$ and $BF(\underline{v}) \simeq 1$

$\frac{1}{\eta} \to +\infty$: ($\tau \to 0$) G"$\to$" Pois-Dir process 2 parameters $(\sigma, 0)$, $BF(\underline{v}) \to h(\sigma)$

BFs between these extremes: (i) increasing functions of $\frac{1}{\eta}$ or (ii) first decreasing and then increasing

Bayes factors for $GG(\sigma, \eta, G_0)$-mixtures

Computation of BFs: via a SIS algorithm (Basu-Chib 2003, for DPM)
Computation of many weights $(w_0(r, k(r); \sigma, \eta), w_1(r, k(r); \sigma, \eta))$, $k(r) = 1, \ldots, r$, $r = 2, \ldots, n$ via multiple precision arithmetics (sums of several incomplete gamma functions)
Computations and MCMC simulations via R calling PARI/C library (as suggested by R. Mena)
When η is *small*, computation of BFs via SIS algorithm is tremendously slow

As an *exploratory* example we consider a simulated sample of size $n = 25$

σ	0	0.1	0.25	0.5	0.75
$\min_\eta \log_{10}(BF(\underline{y}; \sigma, \eta))$	-2.955	-2.943	-2.879	-2.618	-2.159
$\hat{\eta}$	$\widehat{\kappa(\Theta)} = 5.623$	48.790	15.101	3.991	0.550

Bayes factors for $GG(\sigma, \eta, G_0)$-mixtures

Computation of BFs: via a SIS algorithm (Basu-Chib 2003, for DPM)
Computation of many weights $(w_0(r, k(r); \sigma, \eta), w_1(r, k(r); \sigma, \eta))$, $k(r) = 1, \ldots, r$, $r = 2, \ldots, n$ via multiple precision arithmetics (sums of several incomplete gamma functions)
Computations and MCMC simulations via R calling PARI/C library (as suggested by R. Mena)
When η is *small*, computation of BFs via SIS algorithm is tremendously slow

As an *exploratory* example we consider a simulated sample of size $n = 25$

σ	0	0.1	0.25	0.5	0.75
$\min_\eta \log_{10}(BF(\underline{v}; \sigma, \eta))$	-2.955	-2.943	-2.879	-2.618	-2.159
$\hat{\eta}$	$\widehat{\kappa(\Theta)} = 5.623$	48.790	15.101	3.991	0.550

$(\sigma, \eta) \sim \pi$?

η fixed (not too small) and $\sigma \sim Beta(a_0, b_0)$ (discretization on $\{0.01, 0.02, \ldots, 0.99\}$ is computationally convenient) such that $\mathbb{E}(K_n)$ reflects prior opinions

Example 1 - Density estimation

- $n = 100$ data simulated from

$$0.2 \cdot \text{gamma}(40, 20) + 0.6 \cdot \text{gamma}(6, 1) + 0.2 \cdot \text{gamma}(200, 20)$$

- focus on *density estimation* and *detection of the number of cluster* when $V \sim GG(\sigma, \eta, G_0)$-mixtures
- $\omega_1 \in \{1, 3, 10\}$, $\omega_2 = 2$, $\gamma_2 \in \{0.01, 0.1, 1, 10\}$, assigning prior mean $\mathbb{E}(V) = 6$

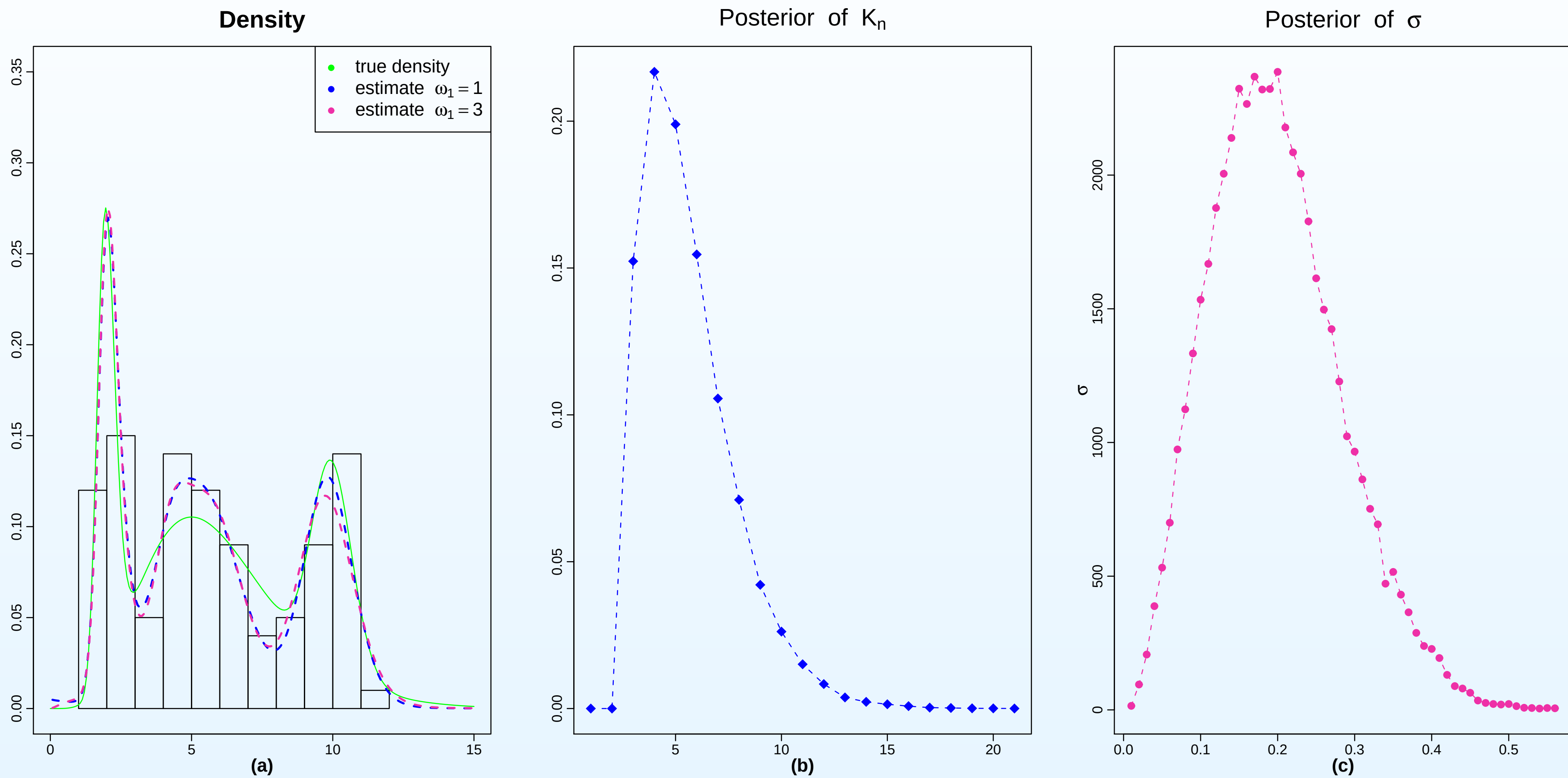
$$\sigma \sim \text{discretized Beta}(1, 1) \quad \eta = 1.5 : \quad \mathbb{E}(K_n) = 30$$

$$\sigma \sim \text{discretized Beta}(1, 7) \quad \eta = 4.5 : \quad \mathbb{E}(K_n) = 6$$

- $\omega_1 = \omega_2 = 1$ (conjugate prior), $\gamma_2 \in \{0.01, 0.1, 1, 10\}$, assigning prior median $med(V) = 5.67$

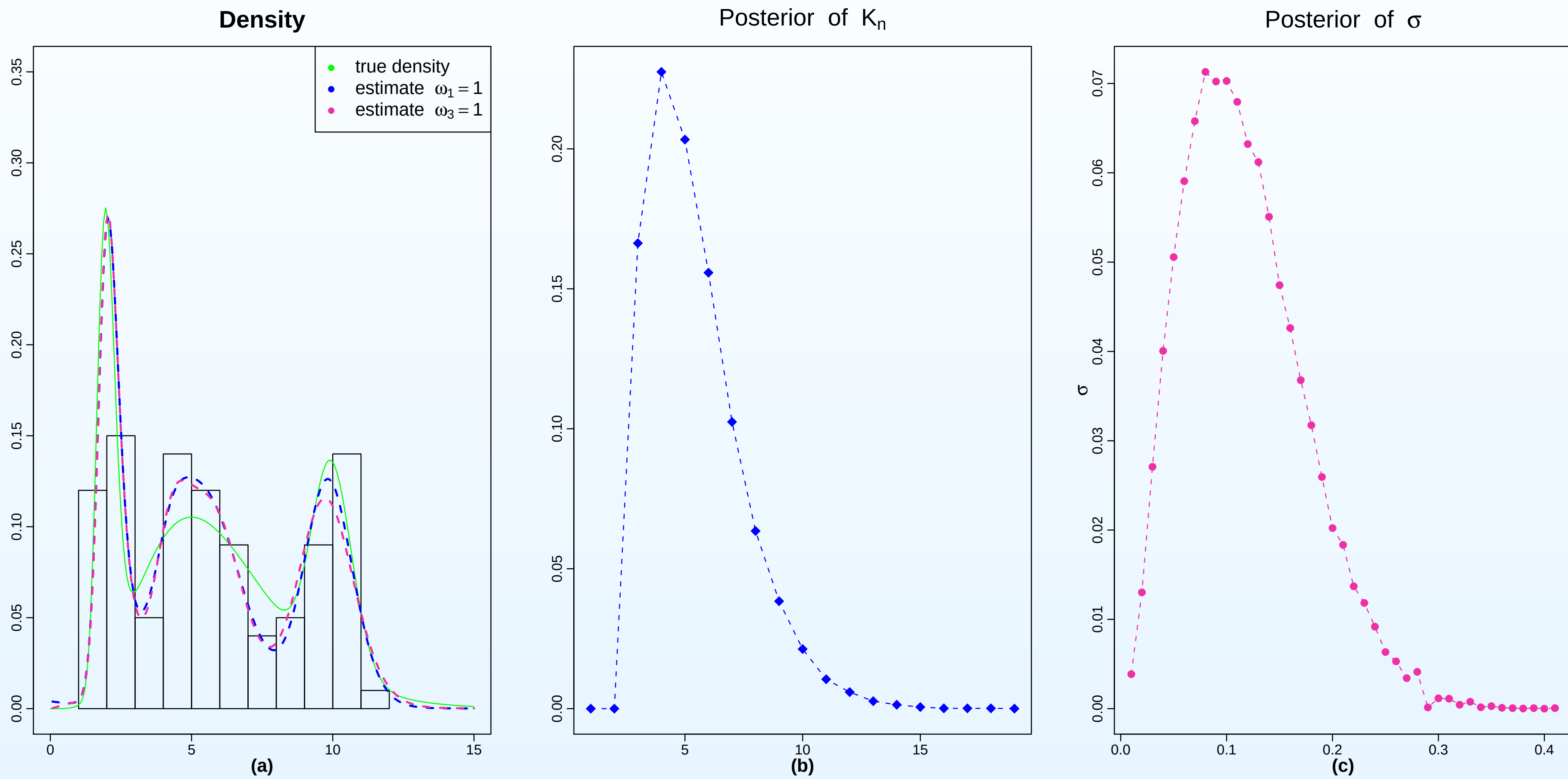
(σ, η) as before

Example 1 - Density Estimation



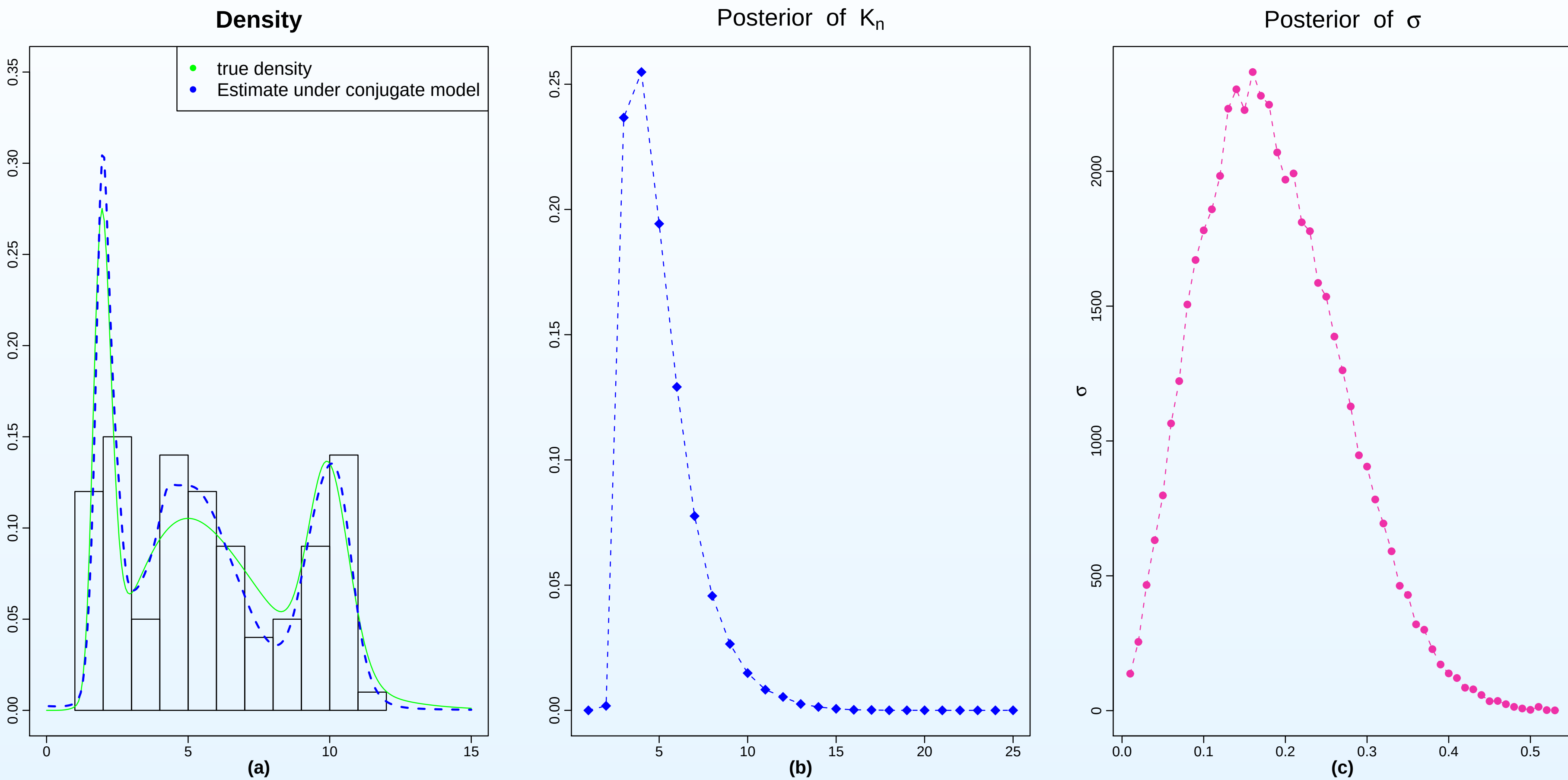
$\omega_2 = 2$ (G_0 non-conjugate), $\gamma_2 = 0.1$, $\sigma \sim Beta(1,1)$ and $\mathbb{E}(K_n) = 30$

Example 1 - Density Estimation



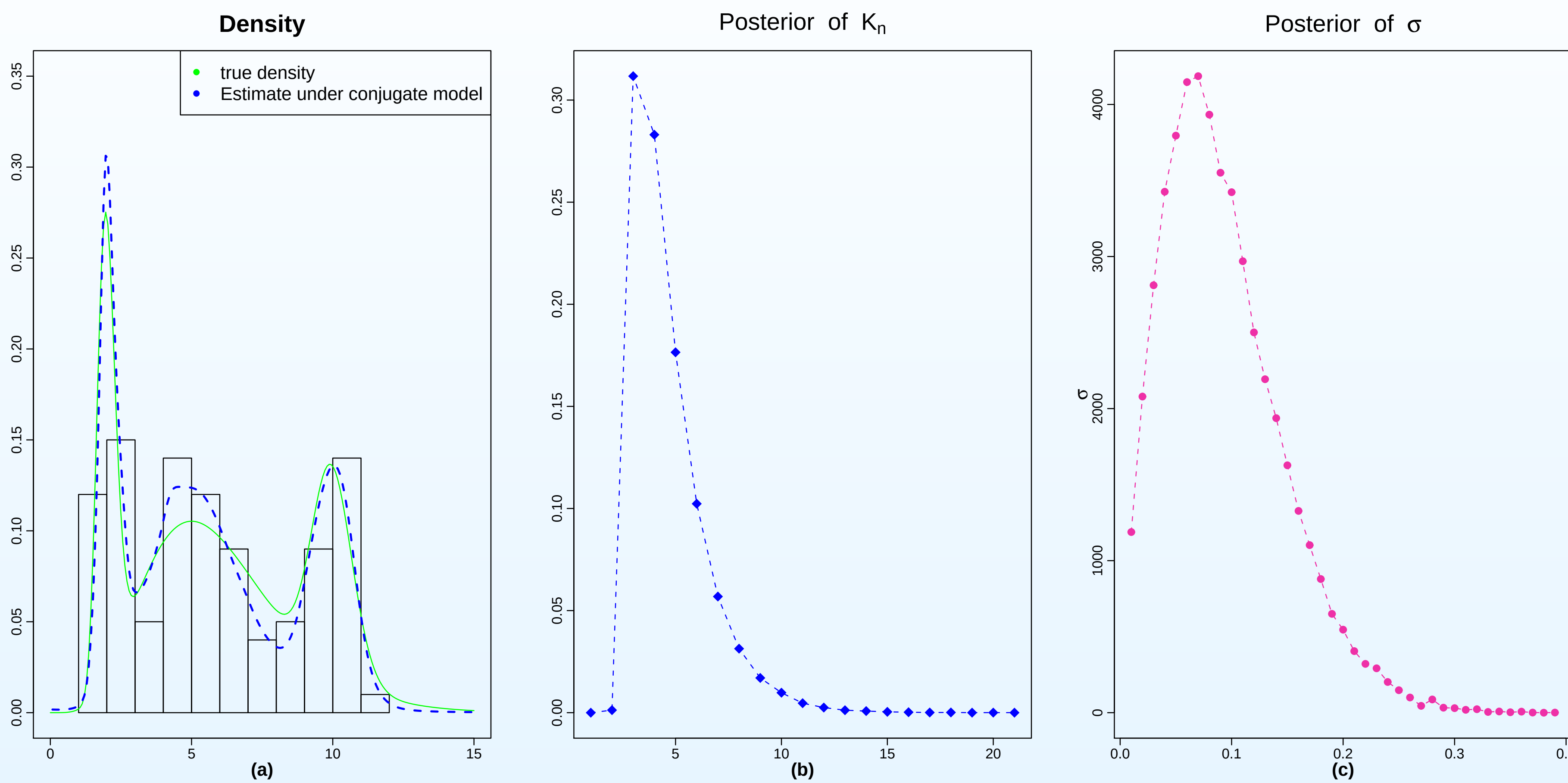
$\omega_2 = 2$ (G_0 non-conjugate), $\gamma_2 = 0.1$, $\sigma \sim Beta(1,7)$ and $\mathbb{E}(K_n) = 6$

Example 1 - Density Estimation



$\omega_1 = \omega_2 = 1$ (G_0 conjugate), $\gamma_2 = 0.01$, $\sigma \sim Beta(1,1)$ and $\mathbb{E}(K_n) = 30$

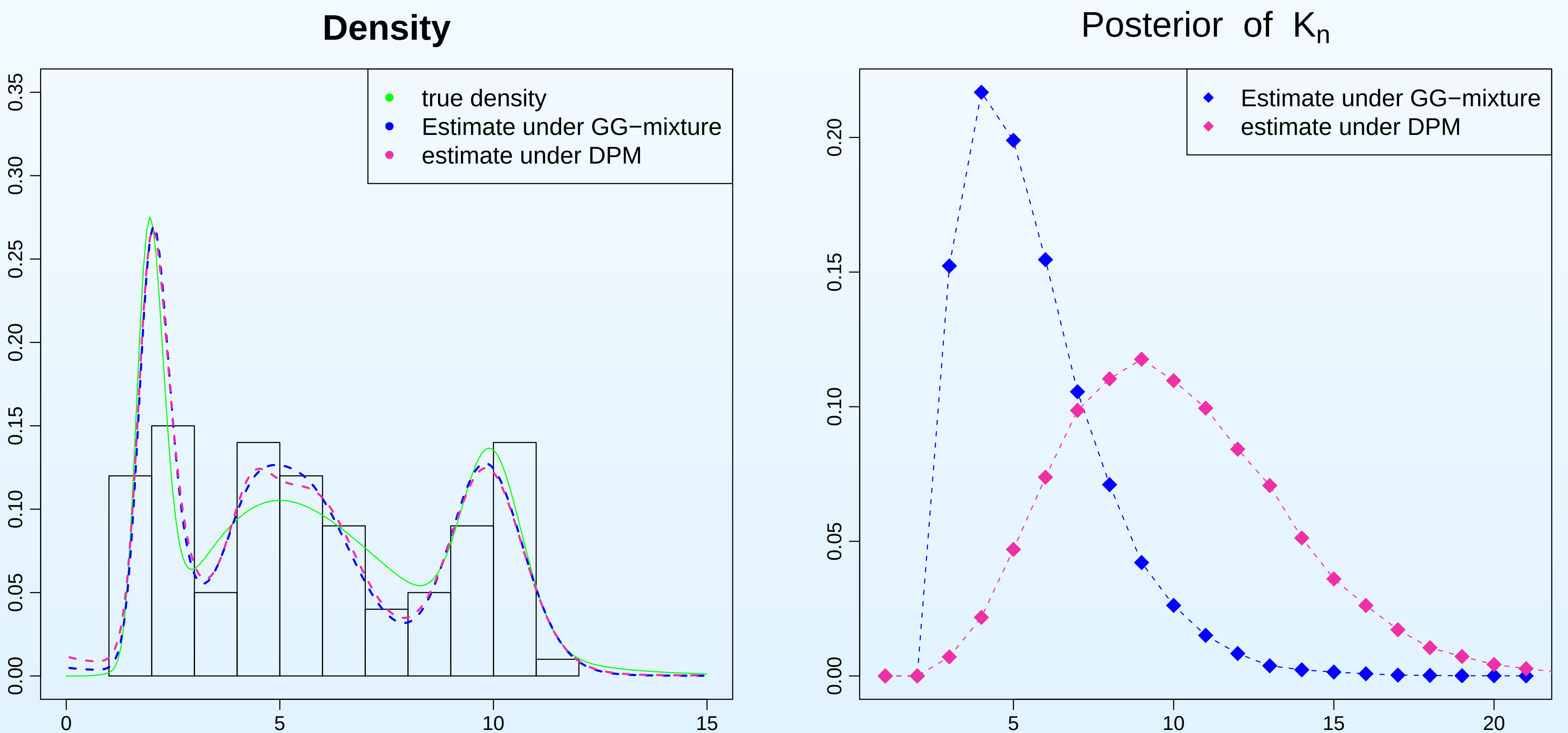
Example 1 - Density Estimation



$\omega_1 = \omega_2 = 1$ (G_0 conjugate), $\gamma_2 = 0.01$, $\sigma \sim Beta(1,7)$ and $\mathbb{E}(K_n) = 6$

Example 1 - Comparison with DPM - random total mass

- $V \sim GG(\sigma, \eta, G_0)$: $\omega_1 = 1, \omega_2 = 2, \gamma_2 = 0.1, \mathbb{E}(V) = 6$
  $\sigma \sim$discretized Beta$(1,1)$, $\mathbb{E}(K_n) = 30$
- $V \sim GG(0, \eta, G_0) = DPM(aG_0)$, same G_0 and $a \sim gamma(3.1, 0.2)$, $\mathbb{E}(K_n) = 30$



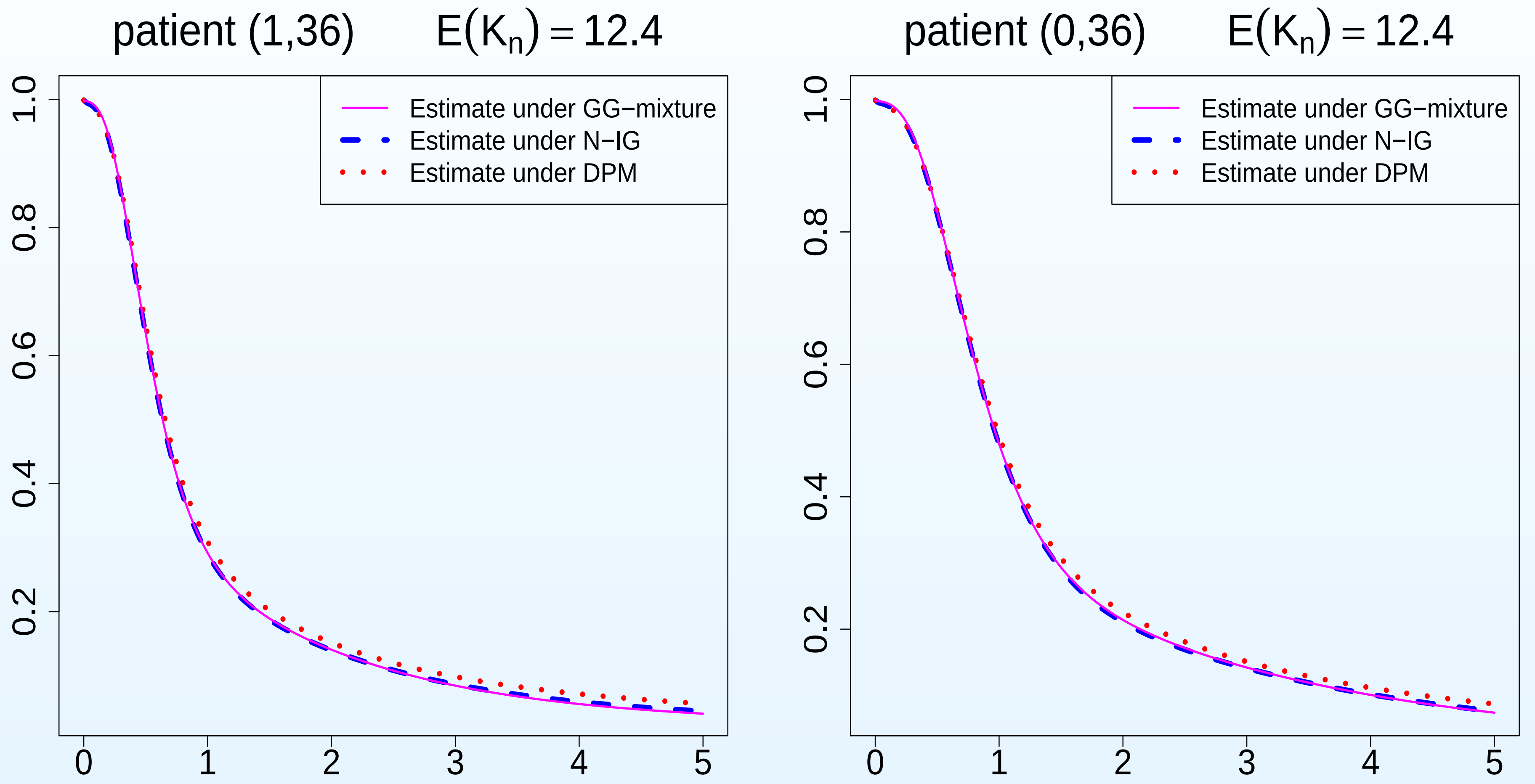
(a) (b)

Example 2 - Dataset with censoring

- $n = 121$ survival times (in thousands of days) of patients suffering from small-cell lung cancer; Walker & Mallick ('99), Kottas & Gelfand ('01), Hanson ('06)
- 2 treatments: A (62 patients) and B (59 patients)
- 23 patients right-censored
- 2 covariates: treatment ($x_1 = 0$ corresponds to A) and entry age; 2 regression parameters (α_1, α_2), no intercept
- estimates (posterior mean) of (α_1, α_2), 90% credible intervals, and estimates of the survival functions
- $\omega_1 = \omega_2 = 1$, $\gamma_2 \in \{0.01, 0.1, 1, 10\}$, $median(V) = 2.44$

  $\sigma \sim$discretized Beta$(1, 4)$ $\quad \eta = 5.77:\quad \mathbb{E}(K_n) = 12.41$ (see EX 3 in Argiento et al., 2007

Example 2 - Survival functions



Estimated survival functions under the $GG(\sigma, \eta, G_0)$-mixture prior for 2 patients for small-cell lung cancer dataset when $m = 2.44$ and $\gamma_2 = 1$.

Example 2 - Estimates of regression parameters

estimates (posterior mean) of (α_1, α_2) and 90% credible interval for hyperparameters as specified before

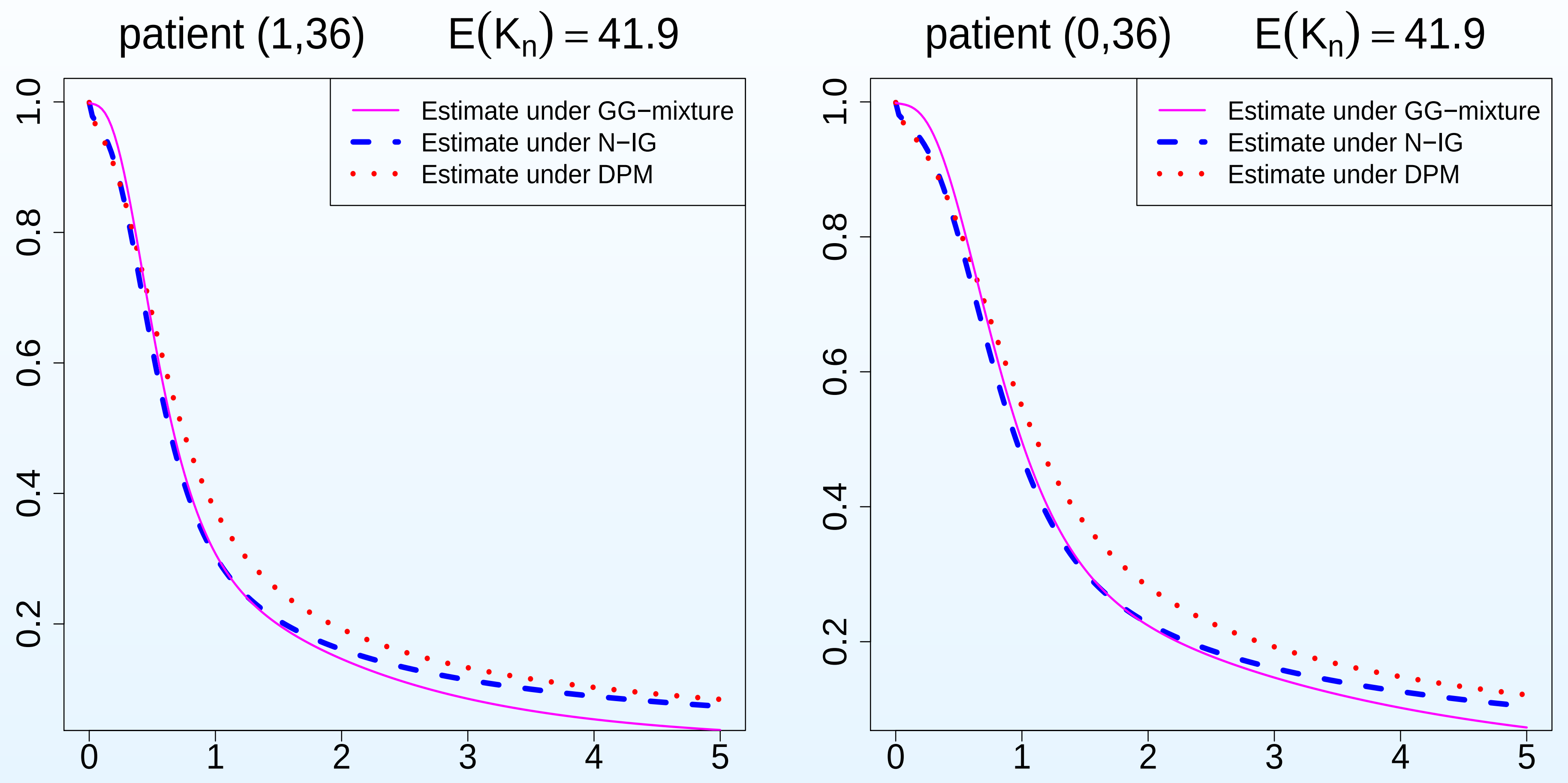
$$1.529(1.229, 1.850); 1.015(1.006, 1.024)$$

$$\mathbb{E}(\sigma | data) = 0.093$$

Robustness analysis with respect to γ_2 ("bandwidth" parameter) and σ

$(\hat{\alpha_1}, \hat{\alpha_2})$	GG (σ random)	N-IG($\sigma = 1/2$)	DPM($\sigma = 0$)
$\gamma_2 = 0.1$	(1.488,1.010)	(1.149,1.011)	(1.437,1.772)
$\gamma_2 = 1$	(1.529,1.015)	(1.533,1.015)	(1.515,1.016)

Example 2 - Survival functions



Estimated survival functions under the $GG(\sigma, \eta, G_0)$-mixture prior for 2 patients for small-cell lung cancer dataset when $m = 2.44$ and $\gamma_2 = 1$, $\mathbb{E}(K_n) = 41.93$, $\sigma \sim$ discretized $Beta(1, 1)$.

Comparison between two $GG(\sigma, \eta, G_0)$-mixture

Predictive fit measure via a cross-validation approach "in the spirit" of Gelfand, Dey, Chang ('92) under the median regression model

$(t_1, \ldots, t_n)$ survival times

$T_j \mid \underline{T}^{(-j)}$, $j \in S^*$= index set of non-censored data

$\tilde{y}_j$, $j \in S^*$, estimated median from the predictive distribution $T_j \mid \underline{T}^{(-j)}$

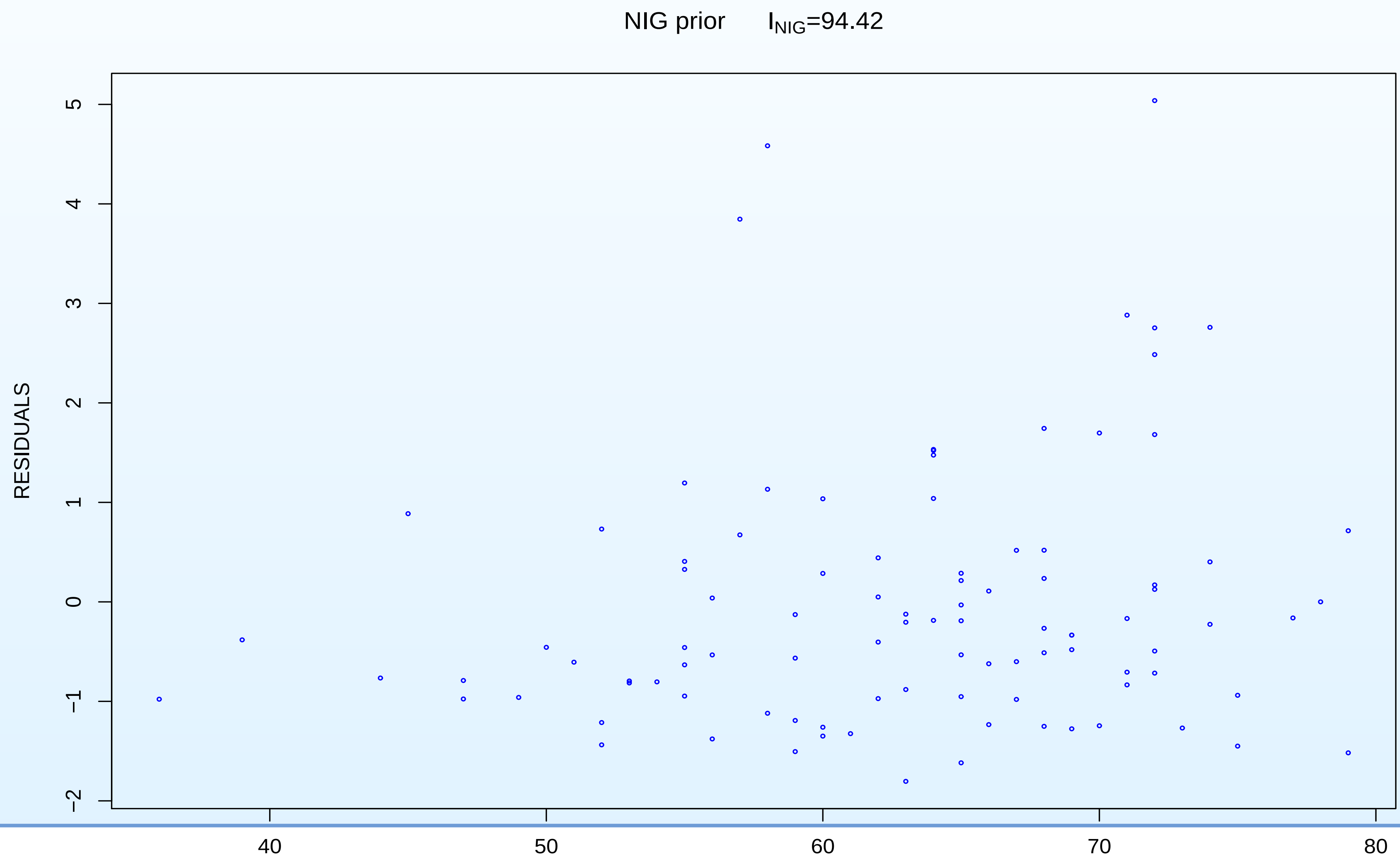
$t_j - \tilde{y}_j$: residual

$$\frac{t_j - \tilde{y}_j}{med_{T_j \mid \underline{T}^{(-j)}} |Y_j - \tilde{y}_j|}$$: standardized residual

$$I := \sum_j \frac{|t_j - \tilde{y}_j|}{med_{T_j \mid \underline{T}^{(-j)}} |Y_j - \tilde{y}_j|}$$: predictive fit index

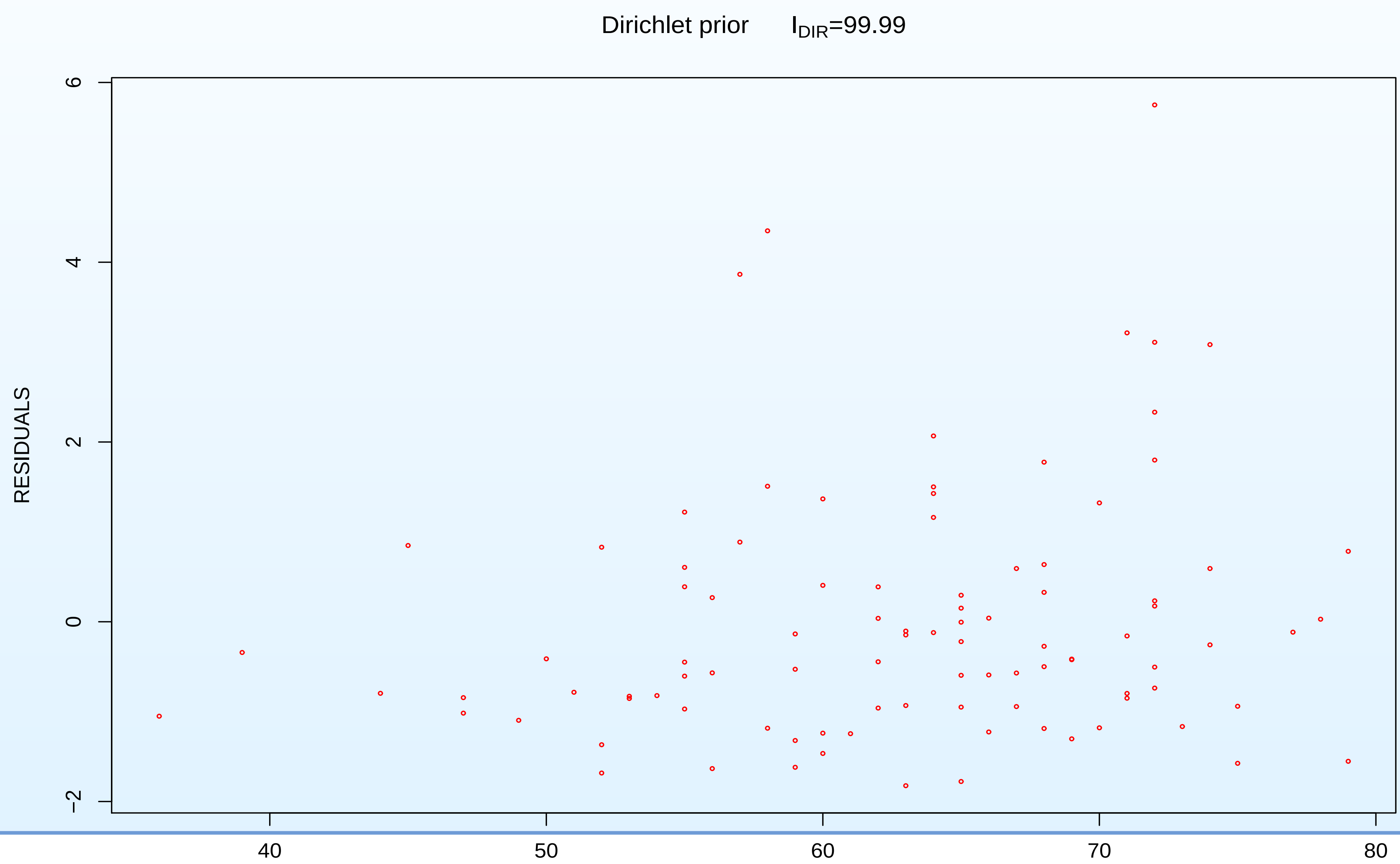
Example 2 - Standardized Residuals

$\sigma = 1/2$: AFT when the error $V \sim$ NIG-mixture

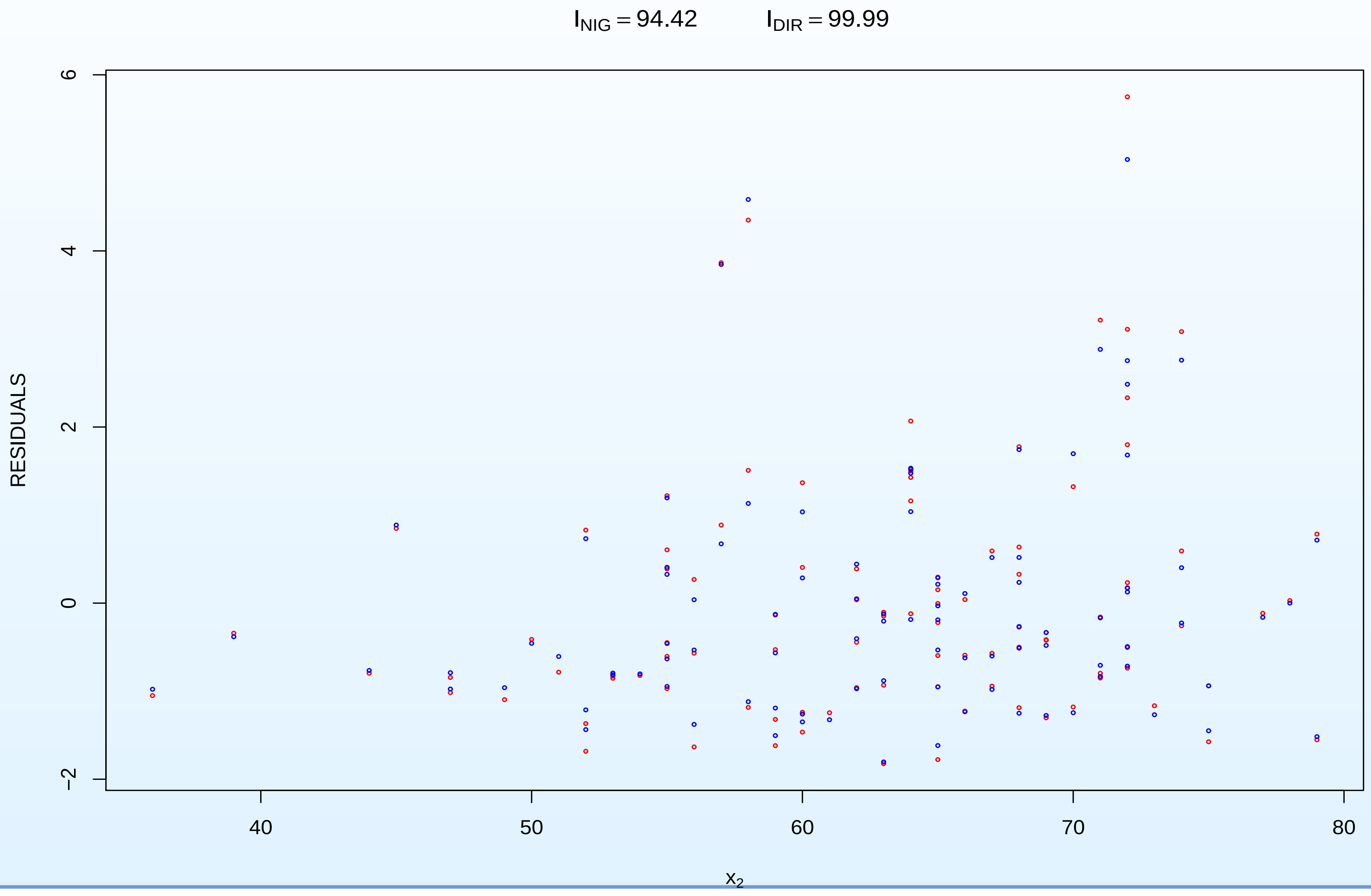


Example 2 - Standardized Residuals

$\sigma \to 0$: AFT when the error $V \sim$ DPM



Example 2 - Standardized Residuals



Comments and future work

- the density estimates (for densities on the positive reals) under the $GG(\sigma, \eta, G_0)$-mixture model with $\sigma \sim \pi(\sigma)$ fundamentally agree with estimates under *similar* nonparametric models ($GG(\sigma, \eta, G_0)$-mixture with σ fixed) but seem more robust with respect to the choice of hyperparameters in G_0
- $GG(\sigma, \eta, G_0)$-mixture model with a random σ seems more effective in detecting the number of clusters in density estimation (as concluded in Lijoi *et al.* 2007 for mixtures of normals)
- $GG(\sigma, \eta, G_0)$-mixture model with $\sigma \sim \pi(\sigma)$ in the AFT setting seems more flexible than $GG(\sigma, \eta, G_0)$-mixture model with σ fixed
- the flexibility of the $GG(\sigma, \eta, G_0)$-mixture models is provided at higher computational cost
- $k(\cdot; \vartheta_1, \vartheta_2)$=Weibull as already suggested in the literature
- simulation of prior/posterior trajectories of G in order to
  - efficiently compute Bayes factors for $GG(\sigma, \eta, G_0)$-mixtures
  - simulate functionals of G

Thanks to Fabrizio Ruggeri

References

- Argiento (2007), PhD Thesis, Bocconi University, Milano.
- Argiento, Guglielmi, Pievatolo (2007), *submitted*.
- Brix (1999), *Adv. Appl. Prob.*, 31, 929-953.
- Cerquetti (2007), *Statist. Probab. Lett.*.
- Christensen, Johnson (1988), *Biometrika*, 75, 693-704.
- Gelfand, Dey, Chang,(1992), In *Bayesian Statistics 4. Proceedings of the Fourth Valencia International Meeting.*
- Gelfand, Kottas (2003), *Scand. J. Statist.*, 30, 651-665.
- Ghosh, Ghosal (2006), in *Bayesian Statistics and its Applications (S. K. Upadhyay et al., eds.), Anamaya Publishers, New Delhi*, 213-229.
- Hanson (2006), *Bayesian Analysis*, 1, 575-594.
- Hanson, Johnson (2002), *JASA*, 97, 1020-1033.
- Ishwaran, James (2003), *Statistica Sinica*, 13, 1211-1235.
- James (2002), arXiv:math.ST/0205093
- James (2005), *Ann. Statist.*, 33, 1771-1799.

References

- Kottas, Gelfand (2001), *JASA*, 96, 1458-1468.
- Kuo, Mallick (1997), *Canad. J. Statist.*, 25, 475-472.
- Lijoi, Mena, Pruenster (2005), *JASA*, 100, 1278-1291.
- Lijoi, Mena, Pruenster (2007), *JRSS, B*.
- Pitman (2003), in *Science and Statistics: a Festschrift for Terry Speed*, volume 40 of *Lectures Notes - Monograph Series*.
- Walker, Mallick (1999), *Biometrics*, 55, 477-483.