Chapter 6

Dependent Tailfree Process and Dependent Multivariate PT

6.1. Linear Dependent Tailfree Process (LDTP)

The popular DDP models for families of random probability measures $\mathcal{G} = \{G_x,\ x \in X\}$ inherits a limitation from the underlying DP prior. The probability measures G_x are a.s. discrete. We earlier discussed a simple fix by convolution with continuous kernels. Alternatively, Jara and Hanson (2011) define a nonparametric Bayesian prior model $p(\mathcal{G})$ that builds on the PT construction and allows to generate absolutely continuous distributions G_x. Recall the construction of the PT prior by defining

$$G(B_{\epsilon 0} \mid B_\epsilon) \equiv Y_{\epsilon 0} \sim \mathsf{Be}(a_{\epsilon 0}, a_{\epsilon 1}) \tag{6.1}$$

for B_ϵ and $B_{\epsilon 0}$ in two adjacent levels of the nested partition. By definition of the PT prior, $Y_{\epsilon 0}$ are independent *across* ϵ. Jara and Hanson (2011) build on (6.1) to define a prior for $\mathcal{G}$. Similar to (6.1) they define

$$Y_{x,\epsilon 0} = G_x(B_{\epsilon 0} \mid B_\epsilon),$$

and introduce dependence *across* x by a simple logistic regression

$$Y_{x\epsilon 0} = \frac{\exp(x'\beta_{\epsilon 0})}{1 + \exp(x'\beta_{\epsilon 0})}, \tag{6.2}$$

leaving independence across ϵ intact. They recommend a g-prior $\beta_{\epsilon 0} \sim N[0, g(X'X)^{-1}]$, with $g = 2n/c$.

Figure 6.1 shows inference in an example. Model (6.2) is a natural equivalent of the PT model for $G \sim \mathsf{PT}$ to families of random probability measures $\{G_x,\ x \in X\}$. However, we should note that, because the logistic model (6.2) for a fixed x does not reduce a beta prior, the implied marginal for G_x is not a PT prior and the model is not, strictly speaking, an extension of the PT model discussed in Chapter 4.

6.2. Dependent PTs

Trippa *et al.* (2011) develop a generalization of the PT prior to a model for related RPMs $\mathcal{G} = \{G_x,\ x \in X\}$ for applications similar to the DDP model. Trippa *et al.* (2011) define a dependent multivariate PT prior (MPT) for $\mathcal{G}$. The advantage of the MPT is the possibility to restrict the model to continuous random distributions and an elegant construction to introduce the dependence. In contrast to the DDP there is no need to track point masses across covariates, and covariates can be of any data format, continuous, categorical or count variables. To avoid confusion we

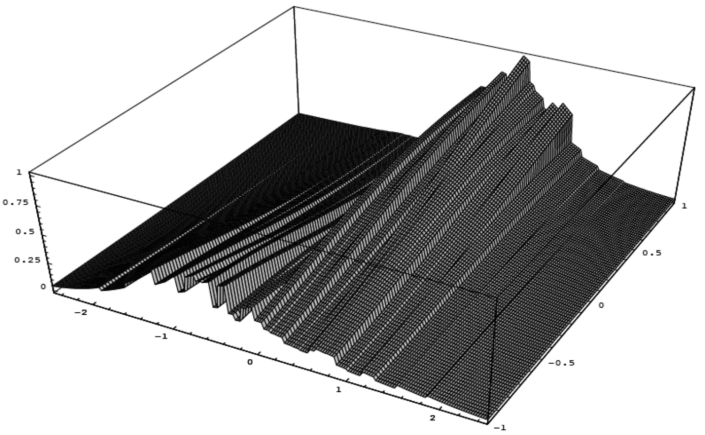

FIG 6.1. *Posterior estimates* $\overline{G}_x = E(G_x \mid data)$. *The figure plots* $\overline{G}_x(t)$ *against* $x \in (-1, 1)$ *and* $-2.5 \leq t \leq 2.5$.

note that this construction is different from the multivariate PT of §4.5. The latter defines a single random probability measure G on a multivariate sample space, while the former constructs a BNP prior for a family $\mathcal{G}$ of random probability measures indexed by x.

Recall the independent random splitting probabilities for a PT random measure $G \sim \mathsf{PT}$,

$$Y_\epsilon = G(B_{\epsilon 0} \mid B_\epsilon) \sim \mathsf{Be}(a_{\epsilon 0}, a_{\epsilon 1}).$$

Note that in anticipation of the upcoming construction we index the random splitting probabilities by ϵ, rather than $(\epsilon, 0)$, as before. By definition of the PT, Y_ϵ are independent across ϵ. The construction of the MPT is conceptually very straightforward. We simply replace the independent beta random variables by random processes $Y_{x,\epsilon}$ with unchanged marginal beta prior, but now with dependence across x.

6.2.1. Multivariate Beta Process

We refer to the desired process $Y_{x,\epsilon}$ as a multivariate beta process (MPT). For clarification we note that the MPT is unrelated to the beta process of Hjort (1990). The MPT uses one realization of the MBP for each ϵ. In the upcoming brief definition of the MBP we simplify notation by dropping the ϵ index in $Y_{x,\epsilon}$.

The construction starts with the representation of a beta random variable Y_x as a ratio of gamma random variables. Let G^o_x, G^1_x denote two independent gamma random variables. Then $Y_x = G^o_x/(G^o_x + G^1_x)$ is a beta random variable. Next we generate the gamma random variables indirectly as illustrated in Figure 6.2 as the random measures $\Gamma(S^o_x)$ and $\Gamma(S^1_x)$ assigned by a gamma process $\Gamma(\cdot)$ to the area circumscribed by kernels centered at x under a gamma process $\Gamma(\cdot)$.

Let X denote the covariate space. The gamma process is defined on $X \times \Re$; if $X = \Re^k$, then the gamma process is indexed by $(k + 1)$ dimensional Borel sets. The trick is that the same construction works just as well to define Y_{x_1} and Y_{x_2} for two covariate values x_1, x_2. The magic is that the overlap of the kernels centered at x_1 and x_2 induces exactly the kind of desired dependence between Y_{x_1} and Y_{x_2}. If the kernels are $N(x, \sigma)$ Gaussian kernels, then the choice of the scale σ determines

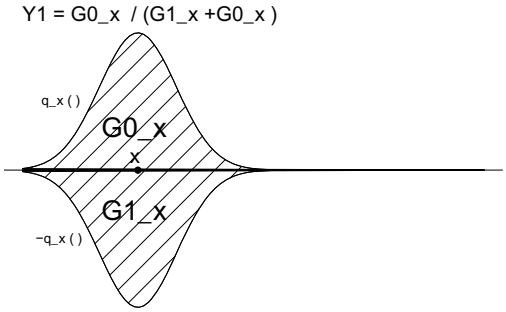

FIG 6.2. *Generating the beta r.v.* Y_x *as ratio of gamma random measures. The gamma random variables are created as random measures* $G_x^0 = \Gamma(S_x^0)$ *and* $G_x^1 = \Gamma(S_x^1)$ *of two the sets* S_x^0 *and* S_x^1 *under a gamma process* $\Gamma(\cdot)$.

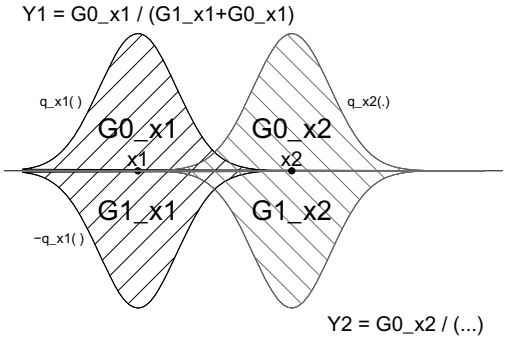

FIG 6.3. *Ratios of gamma random variables define two dependent beta random variables,* $Y_{x_1} = G_{x_1}^o/(G_{x_1}^o + G_{x_2}^1)$ *and* $Y_{x_2} = G_{x_2}^o/(G_{x_2}^o + G_{x_2}^1)$. *The overlap of the kernels determines the correlation. The construction generalizes to families* $\{Y_x;\ x \in X\}$.

the level of dependence. Finally, the construction generalizes to $\{Y_x;\ x \in X\}$, as desired.

In summary, we define the MBP as follows. Let $Q = \{q_x(\cdot)\}$ denote a family of kernels indexed by $x \in X$. For continuous covariates the kernels could be, for example, Gaussian kernels centered at x. Let $S_x^0 = \{(\xi, \zeta),\ \zeta \in X, 0 < \zeta < \alpha_0 q_x(\xi)\}$ denote the area circumscribed by $\alpha_0 q_x(\cdot)$, and similarly for S_x^1 for $-\alpha_1 q_x(\cdot)$. Define $Y_x = \frac{G(S_x^0)}{G(S_x^0)+G(S_x^1)}$ for $x \in X$. In this case we say that $\{Y_x;\ x \in X\} \sim \mathsf{MBP}(\alpha_0, \alpha_1, Q)$.

For use in posterior simulation we note an alternative equivalent construction. For a detailed description of posterior inference see Trippa *et al.* (2011). Here we only introduce the main trick of constructing posterior simulation for the MBP (and thus in the MPT) as standard posterior simulation in DP models. Assume $(Y_{x_1}, \ldots, Y_{x_m}) \sim \mathsf{MBP}(\alpha_0, \alpha_1, Q)$ indexed by x_i, $i = 1, \ldots, m$. In words, we will construct $(Y_{x_1}, \ldots, Y_{x_m})$ as ratios of random measures generated under a DP prior. The construction hinges on the fact that a DP random measure can be written as a normalized gamma process. Thus the ratio of random measures under a DP prior takes the form of a ratio of gamma distributed random variables.

Let $S \equiv \bigcup_{i=1}^m (S_{x_i}^0 \cup S_{x_i}^1)$ denote the region bounded by the $2 \cdot m$ kernels and let $\nu_{x_1 \cdots x_m} \equiv$ denote Lebesgue measure on S. The region S is shown in Figure 6.4.

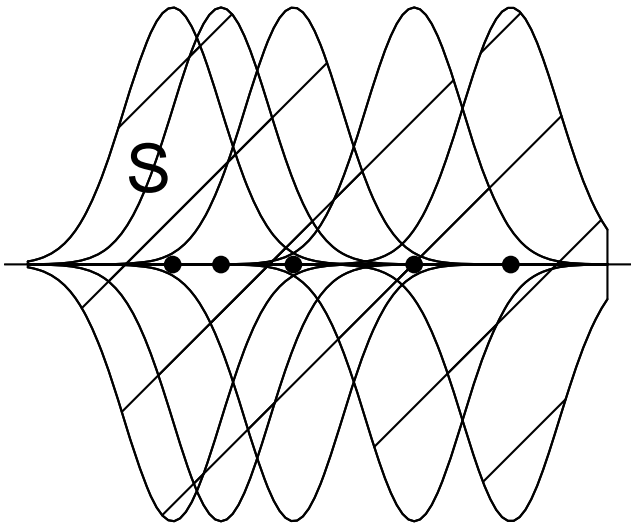

FIG 6.4. *Region S cicumscribed by the union of the* 2 *m kernels.*

Consider a DP random measure

$$D_{x_1 \cdots x_m} \sim \mathsf{DP}(\nu_{x_1 \cdots x_m}).$$

The representation of the DP as a normalized gamma process implies

$$Y_{x_i} = \frac{G(S^0_x)}{G(S^0_x) + G(S^1_x)} \stackrel{d}{=} \frac{D_{x_1 \cdots x_m}(S^0_{x_i})}{D_{x_1 \cdots x_m}(S^0_{x_i} \cup S^1_{x_i})}.$$

This representation can be used to construct a Pólya urn scheme to simulate draws from the MBP. Assume that $Z_i \mid Y_{x_i} \sim \mathsf{Ber}(Y_{x_i})$ and $\{Y_{x_i}\} \sim \mathsf{MBP}$. Then $\boldsymbol{Z} = (Z_1, \ldots, Z_m)$ can be generated as follows. First generate a sequence $(\xi_h, \zeta_h) \sim D_{x_1 \cdots x_m}$. These are points in S (area between the kernels). Find the first pair $(\xi_h, \zeta_h) \in S^0_{x_i} \cup S^1_{x_i}$ and record $Z_i = I(\zeta > 0)$ for that pair. Then we repeat the same for Z_2, etc. The key feature is that the (ξ, ζ) sequence can be generated by the Pólya urn scheme for a marginal sample from a DP random measure, marginalizing with respect to $D_{x_1 \cdots x_m}$.

Later, when we use the MBP to define a prior for the random splitting probabilities $Y_{x_i,\epsilon}$ in the MPT, then the Z_i will be the binary digits of observations $y_i \sim G_{x_i}$, with $p(G_x)$ defined by the binary splitting probabilities $Y_{x,\epsilon}$. Details of this construction are described next.

6.2.2. Dependent Multivariate Pólya Tree

The MBP can be used to generate the beta random variables $Y_{x,\epsilon}$ for a PT prior $G_x \sim \mathsf{PT}$ for $x \in X$, with the dependence across x induced by the dependence of the MBP. We first discuss the construction for a uniform centering distribution, i.e., $E(G_x) = \mathsf{Uni}[0,1]$ for all x. We use the dyadic quantiles of the uniform on $[0,1]$ to define the nested partition sequence, i.e., B_ϵ are the sets $[0, \frac{1}{2}), [\frac{1}{2}, 1], [0, \frac{1}{4})$, etc.

Then, we use the MBP to define the random splitting probabilities. For each ϵ define $\{Y_{\epsilon,x}\} \sim \mathsf{MBP}(\alpha_\epsilon, \alpha_\epsilon, Q)$, using one MBP for each ϵ. We define a family of

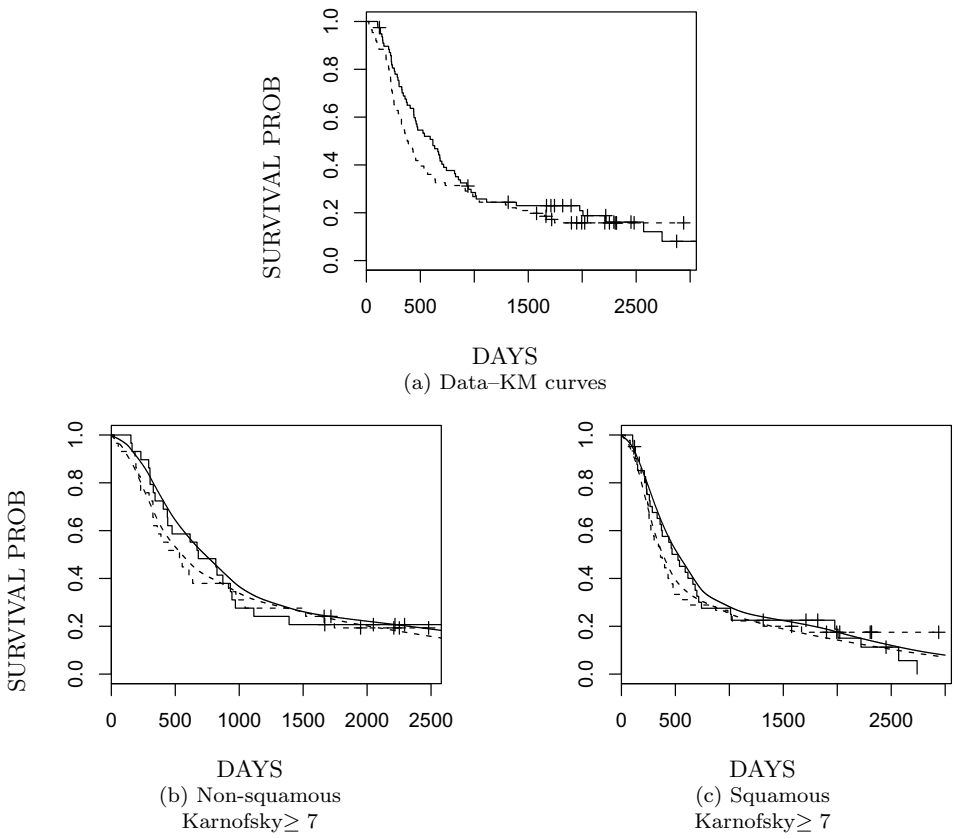

FIG 6.5. *Lung cancer trial. Panel (a) shows the data as Kaplan–Meier curves for treatment (solid line) and control (dashed line). Panels (b) and (c) show the model-based estimates of survival curves arranged by cancer histology, together with the corresponding KM curves.*

RPMs by

$$P_x(B_{\epsilon_1\cdots\epsilon_m}) = \prod_{j=1;\epsilon_j=0}^{m} Y_{\epsilon_1\cdots\epsilon_{j-1}0,x} \prod_{j=1;\epsilon_j=1}^{m} (1 - Y_{\epsilon_1\cdots\epsilon_{j-1}0,x}).$$

We write $\{P_x;\ x \in X\} \sim \mathsf{MPT}(\mathcal{A}, Q, \mathsf{Uni}[0,1])$. The third argument, marks the centering $E(P_x) = \mathsf{Uni}[0,1]$ at the uniform distribution.

Arbitrary centering distributions F_x are easily achieved by modifying the definition to

$$P_x(F_x^{-1}(B_{\epsilon_1\cdots\epsilon_m}) = \prod_{j=1;\epsilon_j=0}^{m} Y_{\epsilon_1\cdots\epsilon_{j-1}0,x} \prod_{j=1;\epsilon_j=1}^{m} (1 - Y_{\epsilon_1\cdots\epsilon_{j-1}1,x}).$$

We write $\{G_x;\ x \in X\} \sim \mathsf{MPT}(\mathcal{A}, Q, F_x)$.

Example 23 *Lung Cancer Trial. Lad* et al. *(1988) report a clinical trial for lung cancer patients. The trial compared radiotherapy versus radiotherapy plus adjuvant chemotherapy. The overall survival data for this study are published in Piantadosi (1997) and is shown in Figure 6.2a. Notice the crossing survival functions The trial enrolled* $n = 164$ *patients, of whom 28 were alive at the end of the followup period.*

The two most important baseline covariates were indicators for squamous versus non-squamous (x_{i1}) histology and performance status at enrollment (x_{i2}). The latter is dichotomized Karnofsky score, with $x_{2i} = 1$ for Karnofsky score ≥ 7. We define a third covariate x_{0i} for treatment assignment with $x_{0i} = 1$ for radiotherapy plus chemotherapy. Trippa et al. *(2011) use the MPT to analyze the data. The MPT model with covariates $\mathbf{x}_i = (x_{i0}, x_{i1}, x_{i2})$ and overall survival times y_i as outcomes implements a fully nonparametric regression for these survival data. Figure 6.2 shows the data and the estimated survival curves.*