

Prior for spline coefficients  $\beta \sim \sqrt{\tau^{-1}}AR2(2\rho, -\rho)$ ,  $\rho \in (0, 1)$ , where  $AR2$  is standardised with marginal variances = 1.

### PC priors for $\rho$

Assuming  $\beta$  follows an AR2 process with  $\rho = (2\rho, -\rho)$ , with  $\rho \in (0, 1)$  and let  $\Sigma$  be the corresponding correlation matrix, i.e. marginal variances of  $\beta = 1$ . Then the KLD for the construction of the PC prior for  $\rho$ , assuming base model is  $\rho = 0$  (i.e. an i.i.d Normal), is:

$$KLD(f_1||f_0) = \frac{1}{2} \left( \text{tr}(\Sigma_0^{-1}\Sigma_1) - n - \log \left( \frac{|\Sigma_1|}{|\Sigma_0|} \right) \right)$$

Since the base model is  $\rho = 0$ ,  $\Sigma_0 = \mathbf{I}$  and hence  $\text{tr}(\Sigma_0^{-1}\Sigma_1) = \text{tr}(\Sigma_1) = n$ , since  $\Sigma_1$  is a correlation matrix.

$$\begin{aligned} \Rightarrow KLD(f_1||f_0) &= -\frac{1}{2} \log(|\Sigma_1|) \\ &= -\frac{1}{2} \log \left[ (1 - \psi_1^2)^{n-1} (1 - \psi_2^2)^{n-2} \right] \\ &= -\frac{1}{2} \log \left[ \left( 1 - \left( \frac{2\rho}{1+\rho} \right)^2 \right)^{n-1} (1 - \rho^2)^{n-2} \right] \\ &= -\frac{1}{2} \log \left[ (1 + 3\rho)^{n-1} (1 - \rho)^{2n-3} (1 + \rho)^{-n} \right] \end{aligned}$$

The PC prior is defined as an exponential prior on  $d(\rho) = \sqrt{2KLD(f_1||f_0)}$  with rate  $\lambda$ . Consider

$$\begin{aligned} d(\rho) &= \sqrt{2KLD(f_1||f_0)} \\ &= \sqrt{-\log \left[ \left( 1 - \left( \frac{2\rho}{1+\rho} \right)^2 \right)^{n-1} (1 - \rho^2)^{n-2} \right]} \\ &= \sqrt{(1-n) \log(1+3\rho) + (3-2n) \log(1-\rho) + n \log(1+\rho)} \\ &= \sqrt{f(\rho)}, \end{aligned} \tag{1}$$

then

$$\begin{aligned} \frac{\partial d(\rho)}{\partial \rho} &= \frac{1}{2\sqrt{f(\rho)}} \left[ \frac{3-3n}{1+3\rho} + \frac{2n-3}{1-\rho} + \frac{n}{1+\rho} \right] \\ &= \frac{1}{2\sqrt{f(\rho)}} \left[ \frac{(3-3n)(1-\rho)(1+\rho) + (2n-3)(1+3\rho)(1+\rho) + n(1+3\rho)(1-\rho)}{(1+3\rho)(1-\rho)(1+\rho)} \right] \\ &= \frac{1}{2\sqrt{f(\rho)}} \left[ \frac{\rho^2(6n-12) + \rho(10n-12)}{(1+3\rho)(1-\rho)(1+\rho)} \right] > 0 \quad \forall \rho \in (0, 1), \end{aligned}$$

i.e.  $d(\rho)$  is a monotonically increasing positive function in  $\rho$ , which is intuitive as  $\rho \rightarrow 1$  means higher deviation from the base model  $\rho = 0$  and hence higher KLD.

Thus, PC prior for  $\rho$  is

$$\begin{aligned} \pi(\rho) &= \pi_d(d(\rho)) \left| \frac{\partial d(\rho)}{\partial \rho} \right| \\ &= \lambda_\rho e^{-\lambda_\rho d(\rho)} \frac{\partial d(\rho)}{\partial \rho} \end{aligned}$$

To determine the decay rate  $\lambda_\rho$ , the authors suggested inferring from an interpretable probability statement  $P(Q(\rho) > U) = \alpha$  for some  $\alpha$  as the tail probability. Here I plan to directly work on the  $\rho$  scale, e.g. specifying  $P(\rho > 0.999) = 0.01$  as a loose PC prior. Since  $d(\rho)$  is a monotonically increasing one-to-one mapping, the upper bound specified on  $\rho$  can be translated into an upper for  $d(\rho)$  as  $P(d(\rho) > d(U)) = 0.01$ . Knowing the tail probability of an exponential R.V  $P(X > x) = 1 - \text{c.d.f.}(x) = e^{-\lambda x}$ ,  $\lambda_\rho$  can be derived from  $e^{-\lambda_\rho d(U)} = 0.01 \implies \lambda_\rho = -\log(0.01)/d(U)$ .

### PC prior for $\tau$

If information is available on the variability on the spline coefficients, the PC prior for  $\tau$  can be inferred directly by a similar procedure of setting an upper bound for  $\tau$ , however the scale of  $\tau$  often depends on several factors, e.g. the spacing of the knots, the scale of the problem, etc. To see this, consider  $y = B\beta$  where  $B$  is the design matrix, then  $\text{VAR}(y) = \tau^{-1} B \Sigma_1(\rho) B'$ . Therefore the scale of  $\tau$  depends on  $\rho$ . The diagonal of the matrix  $B \Sigma_1(\rho) B'$  is relatively uniform, due to the stationarity of the AR2 process assumed and the equal knot spacing and even data. To simplify the construction, we approximate the marginal variance of each  $y_i = \text{VAR}(y_1) = \tau^{-1} B_{1\cdot} \Sigma_1(\rho) B_{1\cdot}'$ , where  $B_{1\cdot}$  is the first row of the design matrix. It is usually easier to give probability statements on the standard deviation of  $y$  instead of  $\beta$ . Hence, the PC prior for  $\tau$  is the type-2 Gumbel prior with decay rate  $\lambda_\tau = -\log(0.01) \sqrt{\text{VAR}(y_1)}/U_y$ , where  $U_y$  is an upper bound of the standard deviation of  $y$ .

Therefore

$$\begin{aligned} f(\tau, \rho) &= f(\tau|\rho)f(\rho) \\ &= \text{type-2 Gumbel}(\tau, \lambda_\tau|\rho) \cdot \pi(\rho) \\ &= \text{type-2 Gumbel}(\tau, \lambda_\tau|\rho) \cdot \lambda_\rho e^{-\lambda_\rho d(\rho)} \frac{\partial d(\rho)}{\partial \rho} \end{aligned}$$

The dependence of  $\lambda_\tau$  on  $\rho$  is through  $\text{VAR}(y_1)$ .

### PC prior for a different variance for the HIV period

It may be more logical to have a different variance, especially for the hump component, during the HIV epidemics. For example, if we have  $\eta \sim \text{AR2}$ , then we can assume  $\beta = \sqrt{W}\eta \sim \sqrt{W}\text{AR2}$ , where  $W$  is a diagonal matrix with elements equal to the marginal variances of each  $\beta$ 's. For simplicity, I set  $W = \text{diag}(w_1, w_1, w_1, \dots, w_2, w_2, w_2)$ , i.e. constant variance before the epidemics and another constant variance after. Obviously, this would destroy the stationarity of the AR2, but it is sensible and hence desirable, as we would not expect the stationarity to hold during the epidemics. Under this formulation, the  $\beta$ 's in the pre-HIV era still enjoys stationarity.

It is slightly more difficult to set up the PC prior for  $w_2$ . To proceed, I re-parameterise the model such that

$$\beta = \sqrt{w_1} \begin{pmatrix} 1 & 0 & 0 & 0 & \dots \\ 0 & 1 & 0 & 0 & \dots \\ \vdots & \ddots & & & \\ & & \sqrt{\gamma} & 0 & 0 \\ & & 0 & \sqrt{\gamma} & 0 \\ & & 0 & 0 & \sqrt{\gamma} \end{pmatrix} \eta,$$

where  $\eta$  follows an  $\text{AR2}(2\rho, -\rho)$  process with marginal variances = 1. Under this parameterisation, we could write the prior  $\pi(w_1, \gamma, \rho) = \pi(w_1|\gamma, \rho)\pi(\gamma|\rho)\pi(\rho)$ , where  $\pi(\rho)$  is the same as above as it does not depend on the variance parameters (since it relates to a correlation matrix), and  $\pi(w_1^{-1}|\gamma, \rho)$  has a gumbel

density as above, because Simpson et al. showed the PC prior of it is independent on the covariance matrix, after conditioning on  $\gamma$  and  $\rho$ .

To calculate the PC prior for  $\gamma$ , again we consider the  $KLD(f_1||f_0)$ , where the base model  $f_0$  is now at  $\gamma = 1$ , i.e. same variance for the whole period. It can also be seen that under this formulation we can set the prior for  $\gamma$  independent of  $w_1$ , as the  $w_1$  in the base model and the alternative model cancel out.

$$\begin{aligned} KLD(f_1||f_0) &= \frac{1}{2} \left( \text{tr}(\Sigma_0^{-1} \Sigma_1) - n - \log \left( \frac{|\Sigma_1|}{|\Sigma_0|} \right) \right) \\ &= \frac{1}{2} \left( \text{tr}(\Sigma_0^{-1} \sqrt{W} \Sigma_0 \sqrt{W}) - n - \log \left( \frac{|\sqrt{W} \Sigma_0 \sqrt{W}|}{|\Sigma_0|} \right) \right) \\ &= \frac{1}{2} \left( \text{tr}(\Sigma_0^{-1} \sqrt{W} \Sigma_0 \sqrt{W}) - n - (n_\gamma) \log(\gamma) \right), \end{aligned}$$

where  $n_\gamma$  indicates the number of  $\beta$ 's with the different variance.

The first term requires more attention. Using the block matrix inversion formula, write:

$$\begin{aligned} \text{tr}(\Sigma_0^{-1} \sqrt{W} \Sigma_0 \sqrt{W}) &= \text{tr} \left( \begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} \begin{pmatrix} I & 0 \\ 0 & \sqrt{\gamma} I \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & \sqrt{\gamma} I \end{pmatrix} \right) \\ &= \text{tr} \left( \begin{pmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{pmatrix} \begin{pmatrix} A & \sqrt{\gamma}B \\ \sqrt{\gamma}C & \gamma D \end{pmatrix} \right) \\ &= \text{tr}(I + A^{-1}B(D - CA^{-1}B)C - \sqrt{\gamma}A^{-1}B(D - CA^{-1}B)C) + \\ &\quad \text{tr}(-\sqrt{\gamma}(D - CA^{-1}B)CA^{-1}B + \gamma(D - CA^{-1}B)D) \end{aligned}$$

using the cyclic property,

$$\begin{aligned} &= \text{tr}(I) + (1 - 2\sqrt{\gamma}) \text{tr}((D - CA^{-1}B)CA^{-1}B) + \gamma \text{tr}((D - CA^{-1}B)^{-1}D) \\ &= \text{tr}(I) + (1 - 2\sqrt{\gamma}) \text{tr}((D - CA^{-1}B)CA^{-1}B) + \gamma \text{tr}((D - CA^{-1}B)^{-1}D) + \\ &\quad \gamma \text{tr}((D - CA^{-1}B)^{-1}CA^{-1}B) - \gamma \text{tr}((D - CA^{-1}B)^{-1}CA^{-1}B) \\ &= \text{tr}(I) + \gamma \text{tr}(I_{n_\gamma}) + (1 - 2\sqrt{\gamma} + \gamma) \text{tr}((D - CA^{-1}B)CA^{-1}B) \\ &= \text{tr}(W) + (1 - \sqrt{\gamma})^2 \text{tr}((D - CA^{-1}B)CA^{-1}B) \\ &= \text{tr}(W) + (1 - \sqrt{\gamma})^2 \text{tr}(EB) \end{aligned}$$

By comparing term of the block inversion formula, it is easily seen that  $-E = -(D - CA^{-1}B)CA^{-1}$  is the bottom left block matrix of the AR2 precision matrix, and  $B$  is the upper right block matrix of the AR2 covariance matrix. Since the precision matrix of an AR2 process is a band  $2 \cdot (2 + 1) - 1$  matrix,  $E$  is a sparse right triangular matrix with only three non-zero entries:

$$-E = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & -\phi_2\sigma^{-2} & \phi_1(\phi_2 - 1)\sigma^{-2} \\ 0 & 0 & 0 & \dots & 0 & 0 & -\phi_2\sigma^{-2} \\ 0 & 0 & 0 & \dots & & & \end{pmatrix},$$

where  $\sigma^2$  is the ratio of adjustment of the innovations at  $t > 2$  relative to the first two innovations in order to maintain stationarity.  $\sigma^2 = \frac{1+\phi_2}{1-\phi_2}((1-\phi_2)^2 - \phi_1^2)$  for an AR2 process. Given  $\phi_1 = 2\rho$  and  $\phi_2 = -\rho$ ,  $-E$  reduces to

$$-E = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & \frac{\rho(1+\rho)}{(1+3\rho)(1-\rho)^2} & \frac{-2\rho(1+\rho)^2}{(1+3\rho)(1-\rho)^2} \\ 0 & 0 & 0 & \dots & 0 & 0 & \frac{\rho(1+\rho)}{(1+3\rho)(1-\rho)^2} \\ 0 & 0 & 0 & \dots & & & \end{pmatrix}.$$

Therefore,

$$\text{tr}(\mathbf{EB}) = \frac{-\rho(1+\rho)}{(1+3\rho)(1-\rho)^2} \text{ACF}(2) + \frac{2\rho(1+\rho)^2}{(1+3\rho)(1-\rho)^2} \text{ACF}(1) + \frac{-\rho(1+\rho)}{(1+3\rho)(1-\rho)^2} \text{ACF}(2),$$

where  $\text{ACF}(\cdot)$  is the auto-correlation function. For AR2,

$$\begin{aligned} \text{ACF}(1) &= \frac{\phi_1}{1-\phi_2} = \frac{2\rho}{1+\rho} \\ \text{ACF}(2) &= \phi_1 \text{ACF}(1) + \phi_2 = \frac{\rho(3\rho-1)}{1+\rho} \\ \implies \text{tr}(\mathbf{EB}) &= \frac{2\rho(1+\rho)^2}{(1+3\rho)(1-\rho)^2} \text{ACF}(1) - \frac{2\rho(1+\rho)}{(1+3\rho)(1-\rho)^2} \text{ACF}(2) \\ &= \frac{2\rho(1+\rho)^2}{(1+3\rho)(1-\rho)^2} \left( \frac{2\rho}{1+\rho} \right) - \frac{2\rho(1+\rho)}{(1+3\rho)(1-\rho)^2} \left( \frac{\rho(3\rho-1)}{1+\rho} \right) \\ &= \frac{2\rho^2(1+\rho)}{(1+3\rho)(1-\rho)^2} \left( (1+\rho) \frac{2}{1+\rho} - \frac{3\rho-1}{1+\rho} \right) \\ &= \frac{2\rho^2(1+\rho)}{(1+3\rho)(1-\rho)^2} \left( \frac{3-\rho}{1+\rho} \right) \\ &= \frac{2\rho^2(3-\rho)}{(1+3\rho)(1-\rho)^2} \end{aligned}$$

Hence, given  $\rho$ ,

$$\begin{aligned} KLD(f_1 \| f_0) &= \frac{1}{2} \left( \text{tr}(\Sigma_0^{-1} \sqrt{\mathbf{W}} \Sigma_0 \sqrt{\mathbf{W}}) - n - (n_\gamma) \log(\gamma) \right) \\ &= \frac{1}{2} \left( n - n_\gamma + \gamma n_\gamma + (1 - \sqrt{\gamma})^2 \frac{2\rho^2(3-\rho)}{(1+3\rho)(1-\rho)^2} - n - n_\gamma \log(\gamma) \right) \\ &= \frac{1}{2} \left( (\gamma - \log(\gamma) - 1) n_\gamma + (1 - \sqrt{\gamma})^2 \frac{2\rho^2(3-\rho)}{(1+3\rho)(1-\rho)^2} \right) > 0, \end{aligned}$$

$$\begin{aligned} d(\gamma|\rho) &= \sqrt{2KLD((f_1 \| f_0))} \\ &= \sqrt{(\gamma - \log(\gamma) - 1) n_\gamma + (1 - \sqrt{\gamma})^2 \frac{2\rho^2(3-\rho)}{(1+3\rho)(1-\rho)^2}}, \end{aligned}$$

$$\begin{aligned} \frac{\partial d(\gamma|\rho)}{\partial \gamma} &= \frac{1}{2d(\gamma|\rho)} \left( n_\gamma - \frac{n_\gamma}{\gamma} - (1 - \sqrt{\gamma}) \frac{1}{\sqrt{\gamma}} \frac{2\rho^2(3-\rho)}{(1+3\rho)(1-\rho)^2} \right) \\ &= \frac{1}{2d(\gamma|\rho)} \left( n_\gamma \left( 1 - \frac{1}{\gamma} \right) - (1 - \sqrt{\gamma}) \frac{1}{\sqrt{\gamma}} \frac{2\rho^2(3-\rho)}{(1+3\rho)(1-\rho)^2} \right) > 0 \text{ if } \gamma > 1, \text{ and } < 0 \text{ if } \gamma < 1 \end{aligned}$$

Finally, the PC prior for  $\gamma$  is then

$$\pi(\gamma|\rho) = \pi(d(\gamma|\rho)) \left| \frac{\partial d(\gamma|\rho)}{\partial \gamma} \right|$$

where  $\pi(d(\gamma|\rho)) \sim \text{Exp}(\lambda_\gamma)$  with  $\lambda_\gamma$  set a priori as  $-\log(0.01)/d(U_\gamma)$  and  $U_\gamma$  specifies an upper bound for the ratio of the marginal variance in the HIV period to the pre-HIV era.

When fitting to the model, I expect  $\gamma$  to be  $> 1$  most of the time since the HIV epidemics should introduce more variability.