Syntax of model formulas in hmmTMB

Théo Michelot

2025-05-21

The package hmmTMB implements hidden Markov models (HMMs) with flexible covariate dependence on parameters of the observation process or the hidden state process. Here, we describe the syntax required to include covariates. The syntax is borrowed from the R package mgcv, and more information and examples can be found in the documentation of mgcv (Wood (2017)). In particular, see the following links:

- https://stat.ethz.ch/R-manual/R-devel/library/mgcv/html/gam.models.html
- https://stat.ethz.ch/R-manual/R-devel/library/mgcv/html/smooth.terms.html

All formulas presented in this vignette could be passed either to `MarkovChain` (to include as covariate on transition probabilities), or to `Observation` (for covariates on observation parameters). We denote as η_i a generic linear predictor, which is the relevant HMM parameter transformed to its working scale (i.e., through the link function).

1 Base R syntax

The basic syntax used to specify models in R, e.g., for the `lm` function, can be used in hmmTMB.

1.1 Linear effects

First consider the linear effect of a continuous covariate x_1

$$\eta_i = \beta_0 + \beta_1 x_{1i}$$

where $beta_0$ is the intercept and β_1 is the slope parameter for x_1. This model can be specified using the formula

```
f <- ~ x1
```

In the case where there are two continuous covariates x_1 and x_2,

$$\eta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$$

the formula becomes

```
f <- ~ x1 + x2
```

1.2 Interactions

The product of two continuous covariates is called their interaction, and including it in the model allow for the effect of each covariate to depend on the value of the other covariate. The model is

$$\eta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i}$$

and can be specified as

```
f <- ~ x1 * x2
```

1.3 Higher-degree effects

In cases where the effect of a covariate is assumed to be non-linear, it is possible to include powers of the covariate in the model (e.g., covariate squared for a quadratic effect). A model with a quadratic term for x_1 is

$$\eta_i = \beta_0 + \beta_1 x_{1i}^2$$

and can be defined with

```
f <- ~ I(x1^2)
```

More often, we want to include all powers of a variable up to some chosen degree (e.g., terms of degree 1, 2, and 3 for a cubic relationship),

$$\eta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{1i}^2 + \beta_3 x_{1i}^3$$

Such models can succinctly be specified with the special function `poly` (for "polynomial"); for example, for a cubic effect of x_1,

```
f <- ~ poly(x1, 3)
```

1.4 Cyclical effect using trigonometric functions

In some applications, it is useful to include other functions of a covariate. One common case is to use trigonometric functions for covariates with a cyclical effect. For example, if we assume that the continuous variable x_2 is the time of day (between 0 and 24), we might want its effect to be the same for $x_2 = 0$ and $x_2 = 24$. Mathematically, this constraint can be created with the model

$$\eta_i = \beta_0 + \beta_1 \cos\left(\frac{2\pi x_{2i}}{24}\right) + \beta_2 \sin\left(\frac{2\pi x_{2i}}{24}\right)$$

which induces a cyclical pattern with period 24. In R, we use the formula

```
f <- ~ cos(2 * pi * x2 / 24) + sin(2 * pi * x2 / 24)
```

2 Non-linear relationships

In some cases, linear or parametric terms might not be flexible enough to capture the relationship between a model parameter and a covariate. The package hmmTMB implements spline-based models to fit non-parametric smooth functions with automatic smoothness estimation. The heavy lifting of model formulation is done under the hood by mgcv, and hmmTMB uses the same syntax for formulas.

2.1 Smooth terms

Non-parametric smooth functions are implemented as linear combinations of basis functions ϕ_k, of the form

$$f(x) = \sum_{k=1}^{K} \beta^{(k)} \phi^{(k)}(x)$$

where the $\beta^{(k)}$ are unknown parameters (Wood (2017), Section 4.2). These parameters are usually not of direct interest, and only the overall shape of f is the goal of inference. In practice, the $\beta^{(k)}$ are also constrained by a penalty on the wiggliness of f.

To estimate the non-parametric effect of x_1, we could use the model

$$\eta_i = \beta_0 + f_1(x_{1i})$$

There are many ways to specify this model in mgcv, mainly because we have to choose the number K of basis functions, and the family of basis functions. Here is such a model,

```
f <- ~ s(x1, k = 5, bs = "ts")
```

where `k = 5` is the dimension of the basis, and `bs = "ts"` defines a thin plate regression spline basis. Please refer to the mgcv documentation for more information on how to choose these arguments.

The non-parametric effects of several covariates can be included in an additive model,

$$\eta_i = \beta_0 + f_1(x_{1i}) + f_2(x_{2i})$$

where the different smooth functions are estimated separately, and can have different specifications. For example, to use a cubic regression spline for x_1 and thin place regression spline for x_2,

```
f <- ~ s(x1, k = 5, bs = "cs") + s(x2, k = 10, bs = "ts")
```

2.2 Varying-coefficient model

Interactions between a smooth function and a continuous covariate can be defined as

$$\eta_i = \beta_0 + f(x_{1i}) \times x_{2i}$$

which is implemented through the `by` argument of `s`,

```
f <- ~ s(x1, by = x2, k = 5, bs = "ts")
```

2.3 Factor-smooth interaction

Similarly, the interaction between a smooth function and a categorical (factor) covariate is something like

$$\eta_i = \beta_0 + f_{l1}(x_{1i}) \quad \text{if } \mathrm{ID}_i = l$$

where a separate function f_{l1} is estimated for each level l of the categorical covariate `ID` (which could be the ID of the time series in a data set with several time series). This is also defined using `by`,

```
f <- ~ s(x1, by = ID, k = 5, bs = "ts")
```

2.4 Cyclical splines

Although describing every type of spline implemented in mgcv is beyond the scope of this document, one other relevant option is to have the model

$$\eta_i = \beta_0 + f(x_{2i})$$

with f constrained to be cyclical. This is a more flexible version of the trigonometric formula presented above. In mgcv, such functions can be implemented with the `"cc"` basis, i.e., with a formula like

```
f <- ~ s(x1, k = 5, bs = "cc")
```

2.5 Isotropic multivariate smooth

Isootropic smooth interactions between two or more continuous variables can also be specified. For two covariates x_1 and x_2, we might represent this by the equation

$$\eta_i = \beta_0 + f(x_{1i}, x_{2i})$$

where f is specified in terms of multidimensional basis functions similarly to one-dimensional smooths. There are two main options for this model in mgcv, depending on whether the variables are on the same scale or not.

If x_1 and x_2 are on the same scale (e.g., x and y spatial coordinates), we might expect the interaction to be isotropic, and this can be modelled with

```
f <- ~ s(x1, x2, k = 5, bs = "ts")
```

Note that this only works with thin plate regression splines (and not with cubic splines for example).

If x_1 and x_2 are **not** on the same scale, tensor product splines are needed instead, but this is not currently supported in hmmTMB.

3 Random effects

One great feature of mgcv is that it treats independent normal random effects as a special case of basis-penalty smooth model. This is for example discussed at the following URL: https://stat.ethz.ch/R-manual/R-devel/library/mgcv/html/smooth.construct.re.smooth.spec.html.

3.1 Random intercept

Consider a model with a random effect of the ID (categorical) variable,

$$\eta_i = \beta_0 + \beta_{1l} \quad \text{if } \mathrm{ID}_i = l$$

where the random intercepts are independent and identically distributed as,

$$\beta_{1l} \sim N(0, \sigma^2_{\mathrm{ID}})$$

Here, β_0 is the population-level (fixed) intercept, and β_{1l} is the deviation from the population mean for group/individual l.

In mgcv (and therefore hmmTMB), such a model is specified with the `"re"` basis, i.e.,

```
f <- ~ s(ID, bs = "re")
```

3.2 Random slope

We can also model the (linear) effect of variable x_1 as being ID-specific, i.e., use a random slope model of the form

$$\eta_i = \beta_0 + (\beta_1 + \beta_{2l})x_{1i} \quad \text{if } \mathrm{ID}_i = l$$

where the random slopes are independent and identically distributed as,

$$\beta_{2l} \sim N(0, \sigma^2_{\mathrm{ID}})$$

Like for the random intercept, β_1 is a fixed effect for the population-level slope and β_{2l} is a random deviation from it. This can also be specified with `bs = "re"`,

```
f <- ~ s(ID, by = x1, bs = "re")
```

Note that the group variable `ID` comes first, and the continuous variable `x1` is passed through the `by` argument.

References

Wood, Simon N. 2017. *Generalized Additive Models: An Introduction with R*. CRC press.