List of distributions in hmmTMB

Théo Michelot

2024-04-17

The hidden Markov models implemented in hmmTMB comprise two stochastic processes: an unobserved state process (S_t), defined as a discrete-valued Markov chain, and an observation process (Z_t). The observation depends on the hidden state as follows,

$$Z_t \mid \{S_t = j\} \sim \mathcal{D}(\boldsymbol{\omega}_j),$$

where $\mathcal{D}$ is a distribution, and $\boldsymbol{\omega}_j$ is the vector of its parameters in state $j \in \{1, 2, \dots\}$.

In practice, the choice of $\mathcal{D}$ depends on the observed variable; for example, a beta variable might be appropriate for a variable defined between 0 and 1, or a gamma distribution for a strictly positive variable. The distributions that are currently included in hmmTMB can be accessed using `hmmTMB:::dist_list`, and a list of distributions together with their parameters is shown below in alphabetical order. The word in quotes is the name of the distribution used in the package, which should be used when defining the observation model.

```
"beta": beta(shape1, shape2)
"binom": binomial(size, prob)
"cat": categorical(p2, p3, ...)
"dir": Dirichlet(alpha1, alpha2)
"exp": exponential(rate)
"foldednorm": folded normal(mean, sd)
"gamma": gamma(shape, scale)
"gamma2": gamma2(mean, sd)
"lnorm": log-normal(meanlog, sdlog)
"mvnorm": multivariate normal(mu1, mu2, ..., sd1, sd2, ..., corr12, ...)
"nbinom": negative binomial(size, prob)
"norm": normal(mean, sd)
"pois": Poisson(rate)
"t": Student's t(mean, scale)
```

"truncnorm": truncated normal(mean, sd, min, max)
"tweedie": Tweedie(mean, p, phi)
"vm": von Mises(mu, kappa)
"weibull": Weibull(shape, scale)
"wrpcauchy": wrapped Cauchy(mu, rho)
"zibinom": zero-inflated binomial(size, prob, z)
"zigamma": zero-inflated gamma(shape, scale, z)
"zigamma2": zero-inflated gamma2(mean, sd, z)
"zinbinom": zero-inflated negative binomial(size, prob, z)
"zipois": zero-inflated Poisson(rate, z)
"ztnbinom": zero-truncated negative binomial(size, prob)
"ztpois": zero-truncated Poisson(rate)

Link functions: Note that, for the purpose of statistical estimation, the parameters are transformed through a link function. This is particularly important when an observation parameter is modelled as a function of covariates, because covariate effects are included on the link-transformed parameter. The link functions used for the parameters are the following:

- log for $\omega > 0$ (e.g., standard deviation of `"norm"` distribution)
- logit for $\omega \in [0, 1]$ (e.g., probability of `"binom"` distribution)
- logit of scaled parameter for $\omega \in [a, b]$ (e.g., angular mean $\in [-\pi, \pi]$ of `"vm"` distribution)
- identity for $\omega \in \mathbb{R}$ (e.g., mean of `"norm"` distribution)

Special distributions: The categorical (`"cat"`) and multivariate normal (`"mvnorm"`) distributions don't have a fixed number of parameters, because they depend on the data. The number of parameters for the categorical distribution depends on the number of unique values in the data, and the number of parameters for the multivariate normal distribution depends on the dimension of the data.