

OU_vs_BM

chundra

3/31/2025

Rationale

Phylogenetic regression often assumes that the stochastic process that generates data in related languages is Brownian motion, when in reality there are multiple valid evolutionary processes that could be involved.

This notebook contains preliminary simulation studies designed to investigate whether phylogenetic regression models infer misleading posterior coefficient distributions if the phylogenetic stochastic process is misspecified (specifically, if we assume BM when an Ornstein-Uhlenbeck process was used to generate the data).

For several iterations, we generate synthetic OU data, and fit two phylogenetic regression models (assuming a BM and OU phylogenetic term, respectively) on the simulated data. We compare the posterior distributions of the effect of a synthetic predictor on a numeric response, along with the value used to generate the synthetic data.

Stan models

```
model_BM = "data {  
  int<lower=1> D;  
  vector[D] x;  
  vector[D] y;  
  cov_matrix[D] phy;  
}  
parameters {  
  real theta;  
  real beta;  
  real<lower=0> sigma_phy;  
  real<lower=0> sigma;  
  vector[D] z;  
}  
transformed parameters {  
  vector[D] mu = theta + cholesky_decompose(phy*sigma_phy)*z + beta*x;  
}  
model {  
  theta ~ normal(0,1);  
  beta ~ normal(0,1);  
  sigma_phy ~ normal(0,1);  
  sigma ~ normal(0,1);  
  z ~ normal(0,1);  
  y ~ normal(mu,sigma);  
}"
```

```

model_BM = "data {
  int<lower=1> D;
  vector[D] x;
  vector[D] y;
  cov_matrix[D] phy;
}
parameters {
  real theta;
  real beta;
  real<lower=0> sigma_phy;
  real<lower=0> sigma;
  vector[D] z;
}
transformed parameters {
  vector[D] mu = theta + cholesky_decompose(phy*sigma_phy)*z + beta*x;
}
model {
  theta ~ normal(0,1);
  beta ~ normal(0,1);
  sigma_phy ~ normal(0,1);
  sigma ~ normal(0,1);
  z ~ normal(0,1);
  y ~ normal(mu,sigma);
}"

model_OU = "data {
  int<lower=1> D;
  vector[D] x;
  vector[D] y;
  matrix[D,D] phy;
}
parameters {
  real theta;
  real<lower=0> alpha;
  real beta;
  real<lower=0> sigma_phy;
  real<lower=0> sigma;
  vector[D] z;
}
transformed parameters {
  vector[D] mu = theta + cholesky_decompose(sigma_phy*exp(-.5*phy/alpha) + diag_matrix(rep_vector(.001,
}
model {
  theta ~ normal(0,1);
  beta ~ normal(0,1);
  sigma_phy ~ normal(0,1);
  sigma ~ normal(0,1);
  z ~ normal(0,1);
  y ~ normal(mu,sigma);
}"

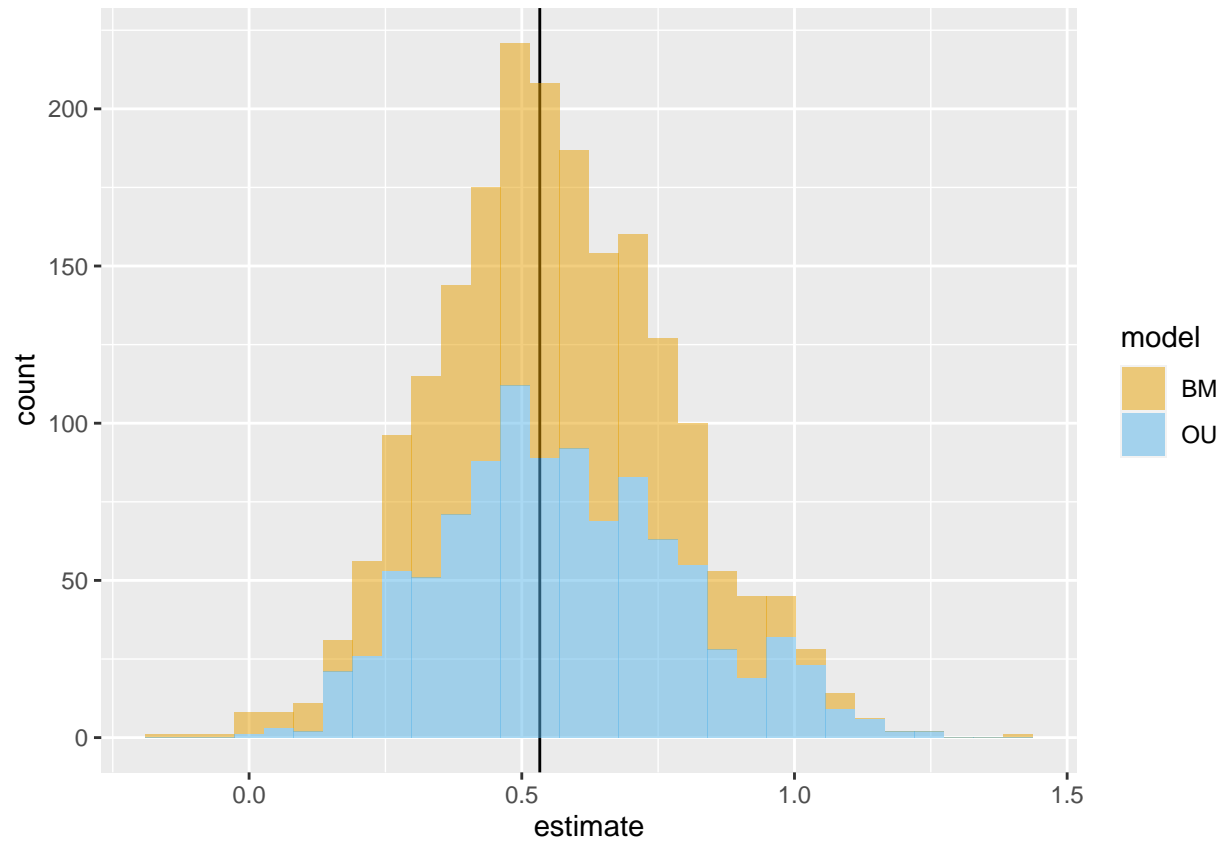
```

Simulations

We generate synthetic data for a phylogenetic regression with an OU 20 times, and fit the above models. Naturally, there are many more dimensions to explore (e.g., propensity of different approaches for false positives/negatives), which we set aside for future work.

```
plots
```

```
## [[1]]
```



```
##
```

```
## [[2]]
```

