Quantifying Elephant Social Structure: Using a Bilinear Mixed Effects Model to Elicit Qualities of Elephant Behavior

Eric Vance

Institute of Statistics & Decision Sciences

Duke University

Preliminary Exam

March 29, 2005

Elephant Social Structure

- Only females form families. Males just run around looking to mate.
- Oldest female is the leader since she's the largest and wisest.
- Elephants within a family tend to be related.

Scientific Questions

- Why do elephants stay in large groups even when food is scarce?
- What role does genetics play in elephant social structure?
- How does one quantify social structure in order to assess whether or not groups are larger in the Wet Season vs. the Dry Season?

Data Collection

- Biologists in Kenya ride into the National Park looking for herds of elephants.
- When a herd is spotted, they write down the names of the elephants present.
- The biologists either stay to observe the family or move on to the next herd.

Binomial Data Example

The total number of times data were recorded:

$$\text{Dry Season:} \quad N_{\text{Dry}} = 331$$
$$\text{Wet Season:} \quad N_{\text{Wet}} = 171$$

If Amy, Ang, and Ali are observed together, and the others are missing, then:

$$\boxed{\begin{aligned} y_{\text{AmyAng}} &= y_{\text{AmyAli}} = y_{\text{AngAli}} = 1 \\ n_{\text{AmyAng}} &= n_{\text{AmyAli}} = n_{\text{AngAli}} = 1 \end{aligned}}$$

While for one missing elephant: $\boxed{\begin{aligned} y_{\text{AmyAme}} &= 0 \\ n_{\text{AmyAme}} &= 1 \end{aligned}}$

Whereas for two missing elephants: $\boxed{\begin{aligned} y_{\text{AstAme}} &= 0 \\ n_{\text{AstAme}} &= 0 \end{aligned}}$

Binomial Data, Cont.

In this example of five elephants Amy, Angelina, Alison, Astrid, and Amelia at time $= t$, the $\mathbf{y}$ matrix of successful observations would be:

$\mathbf{y_t} =$

	Amy	Ang	Ali	Ast	Ame
Amy	$\ddots$	1	1	0	0
Ang	1	$\ddots$	1	0	0
Ali	1	1	$\ddots$	0	0
Ast	0	0	0	$\ddots$	0
Ame	0	0	0	0	$\ddots$

The $\mathbf{n_t}$ matrix of potential observations =

	Amy	Ang	Ali	Ast	Ame
Amy	$\ddots$	1	1	1	1
Ang	1	$\ddots$	1	1	1
Ali	1	1	$\ddots$	1	1
Ast	1	1	1	$\ddots$	0
Ame	1	1	1	0	$\ddots$

The Model

- Data is binomial
  - $y_{ij} \sim \text{Bin}(n_{ij}, p_{ij})$
  - y_{ij} is the number of times elephants i and j observed together.
  - n_{ij} is the number of times either i or j observed.

- Use a GLM
  - $E(y_{ij} | \theta_{ij}) = g(\theta_{ij})$.
  - g is the inverse logit link function.
  - So $p_{ij} = \frac{\exp \theta_{ij}}{1+\exp \theta_{ij}}$.

- θ_{ij} ***is the linear predictor.***

Linear Predictor θ_{ij}

How often are elephants together?

- Intrinsic sociability a_i.
  - Sociable elephants will be observed together with other elephants (in groups) more often than unsociable elephants.
- Common intercept β_0.
- Genetic relatedness $\beta_g \, g_{ij}$.
  - DNA samples lead to a measure g_{ij} of how closely elephant i and j are related.
- Normal error γ_{ij}.
- **Pairwise effect** $z_i' z_j$ between elephants i and j.

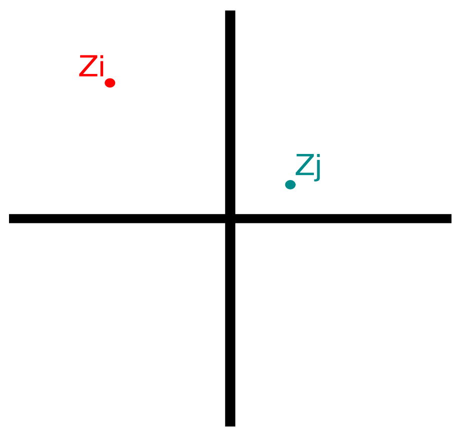
$$\boldsymbol{\theta_{ij} = (\tfrac{1}{2}\beta_0 + a_i) + (\tfrac{1}{2}\beta_0 + a_j) + \beta_g \, g_{ij} + \gamma_{ij} + z_i' z_j}$$

Pairwise Effect

$z_i' z_j$ is the inner product of the positions of elephants i and j in "Social Space".

- For visual interpretability I choose the dimension of social space $k = 2$.

Elephants i and j have positions $\mathbf{z}_i$ and $\mathbf{z}_j$ in 2D social space.



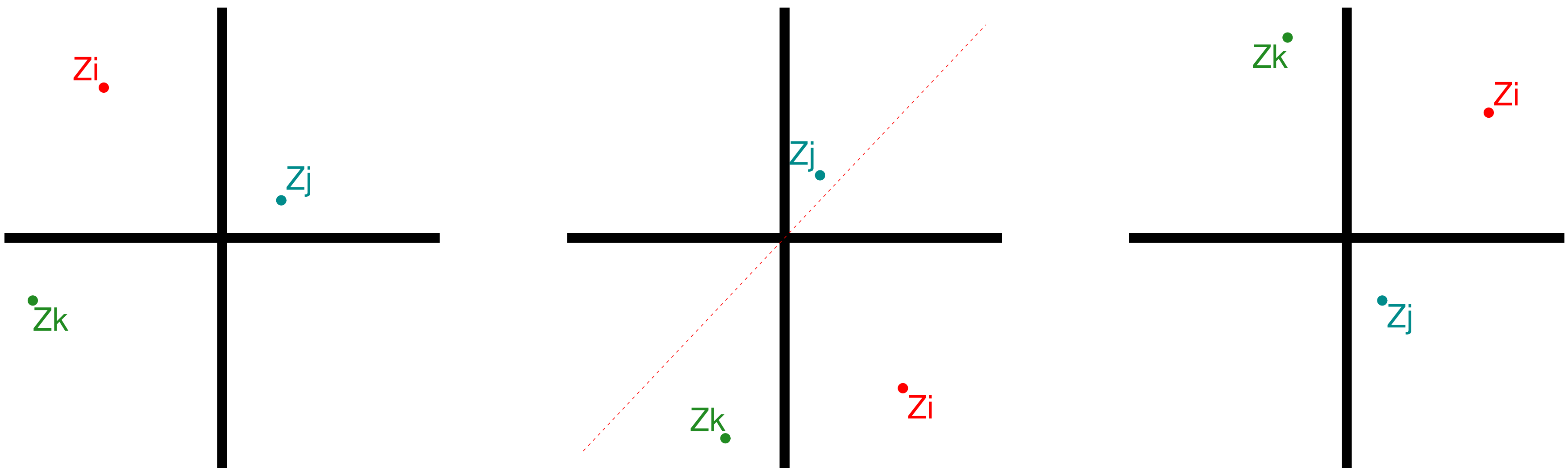
$$\mathbf{z}_i \sim N(\mathbf{0}, \sigma_z^2\,\mathrm{I})$$
$$\mathbf{z}_j \sim N(\mathbf{0}, \sigma_z^2\,\mathrm{I})$$

- If $\mathbf{z_i' z_j} = \mathbf{0}$ then elephants i and j interact as often as their sociabilities a_i ,a_j and their genetics g_{ij} would predict.
- If $\mathbf{z_i' z_j} > \mathbf{0}$ then i and j like each other and are observed together more often than the model would otherwise predict.
- If $\mathbf{z_i' z_j} < \mathbf{0}$ then i and j dislike each other.

Pairwise Effect, Cont.

- Only the inner products of the vectors $z_i'z_j, z_i'z_k, z_j'z_k$ matter.

Reflections or rotations of Social Space do not change inner products.

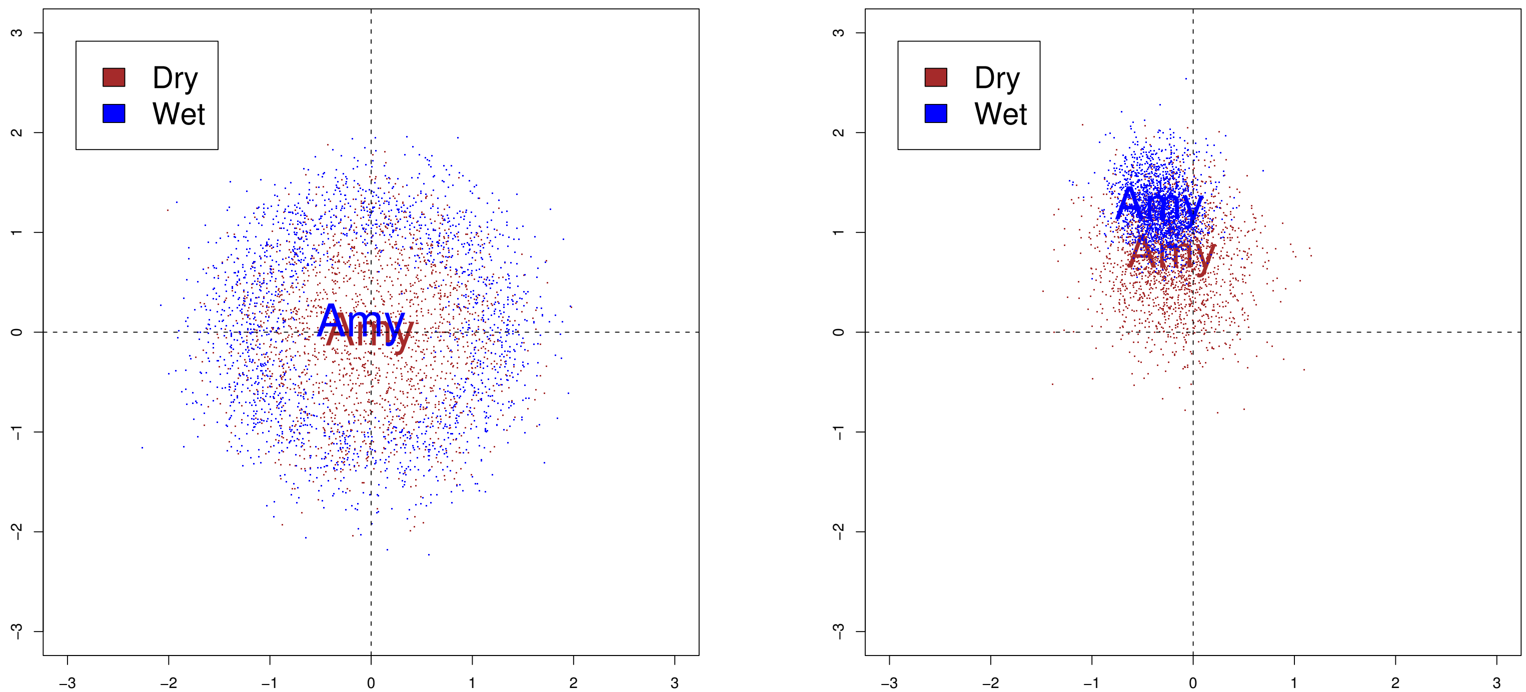


Social Space Reflection Rotation

- All 3 social spaces are equivalent.

Procrustean Transformation

- The posterior draws of the social space vectors must be reflected or rotated to give a coherent picture of the posterior distribution.



Z social space $\mathbf{Z}^*$ transformed social space

- Fix an arbitrary matrix Z_0 of positions in social space, then apply the Procrustean transformation:

$$Z^* = Z_0 Z' \left(Z Z_0' Z_0 Z' \right)^{-\frac{1}{2}} Z$$

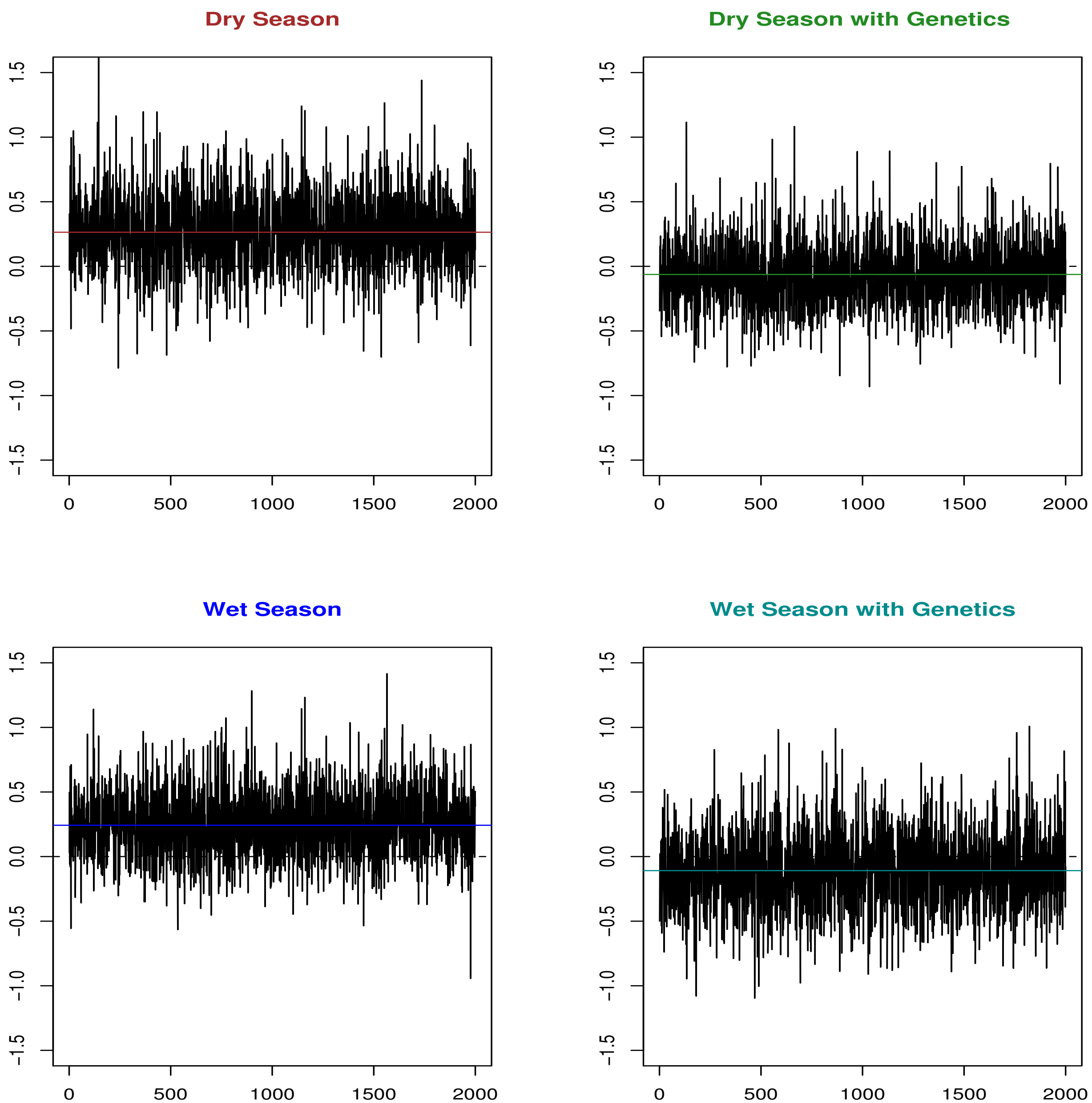
Vague Priors

$$\theta_{\mathbf{ij}} = (\tfrac{\mathbf{1}}{\mathbf{2}}\beta_{\mathbf{0}} + \mathbf{a_i}) + (\tfrac{\mathbf{1}}{\mathbf{2}}\beta_{\mathbf{0}} + \mathbf{a_j}) + \beta_{\mathbf{g}}\,\mathbf{g_{ij}} + \gamma_{\mathbf{ij}} + \mathbf{z_i'z_j}$$

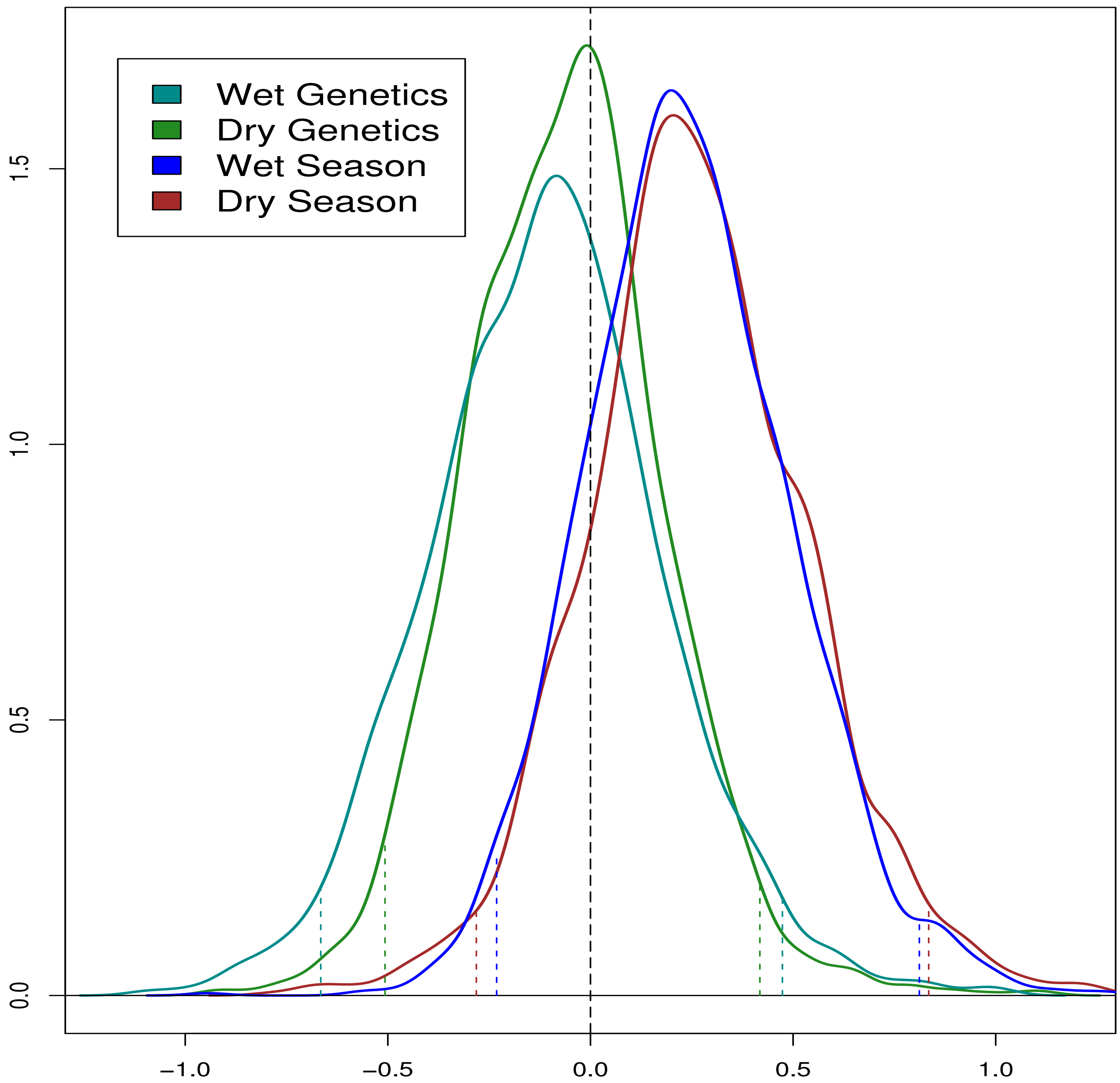
Intercept:	$\beta_0 \sim N(0\,,\ 100)$	
Sociabilities:	$a_i, a_j \sim N(0,\ \sigma^2_{\text{soc}}),$	$\sigma^2_{\text{soc}} \sim IG(\frac{1}{2}\,,\ \frac{1}{2})$
Genetic Coefficient:	$\beta_g \sim N(0\,,\ 100)$	
Pairwise error:	$\gamma_{ij} \sim N(0,\ \sigma^2_{\gamma}),$	$\sigma^2_{\gamma} \sim IG(\frac{1}{2}\,,\ \frac{1}{2})$
Social space:	$\mathbf{z}_i, \mathbf{z}_j \sim N(\mathbf{0},\ \begin{bmatrix}\sigma^2_z & 0\\ 0 & \sigma^2_z\end{bmatrix}),$	$\sigma^2_z \sim IG(\frac{1}{2}\,,\ \frac{1}{2})$

Amy, Matriarch of Family AA

Amy Sociability MCMC

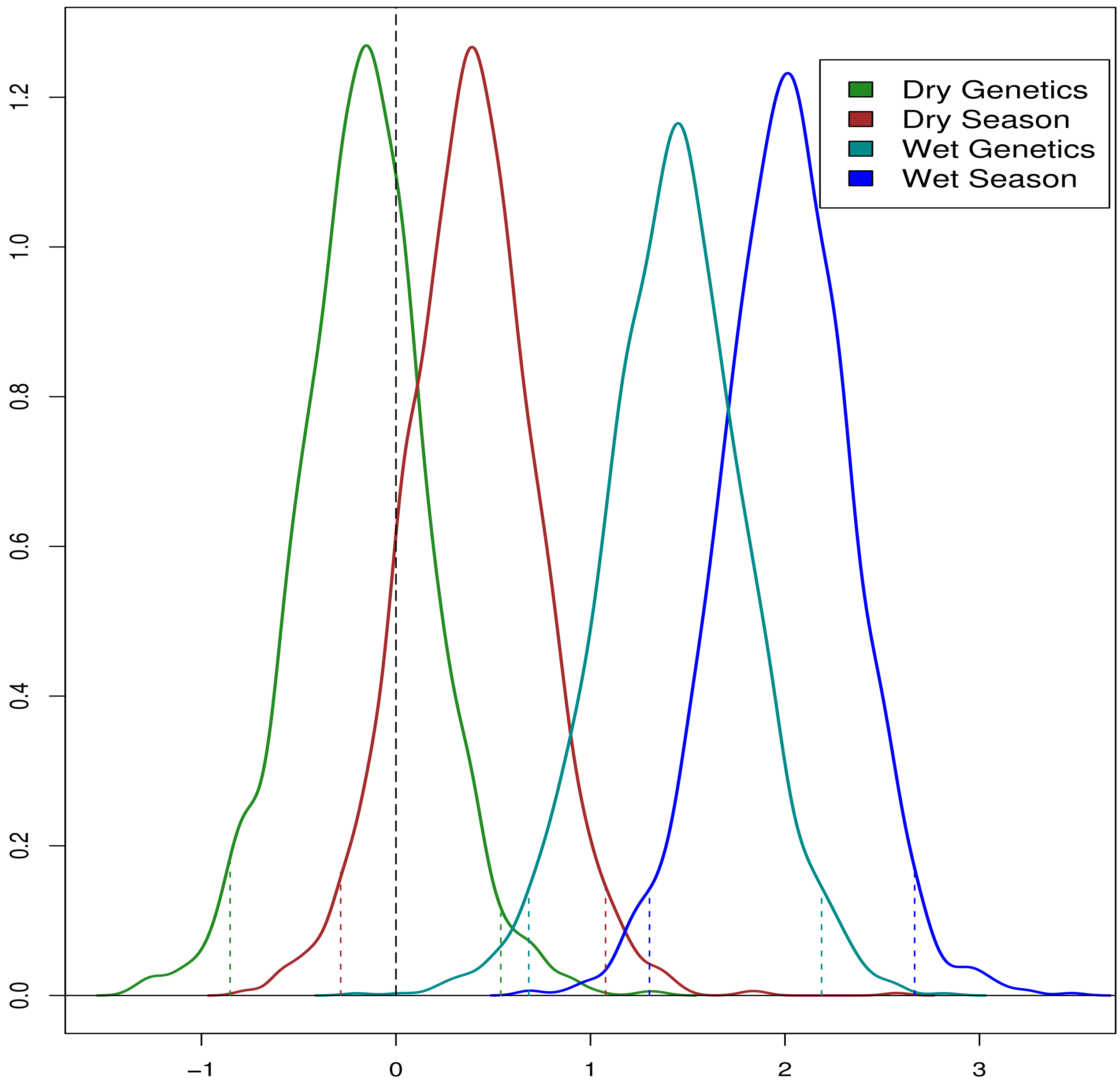


Amy Sociability Posterior Density

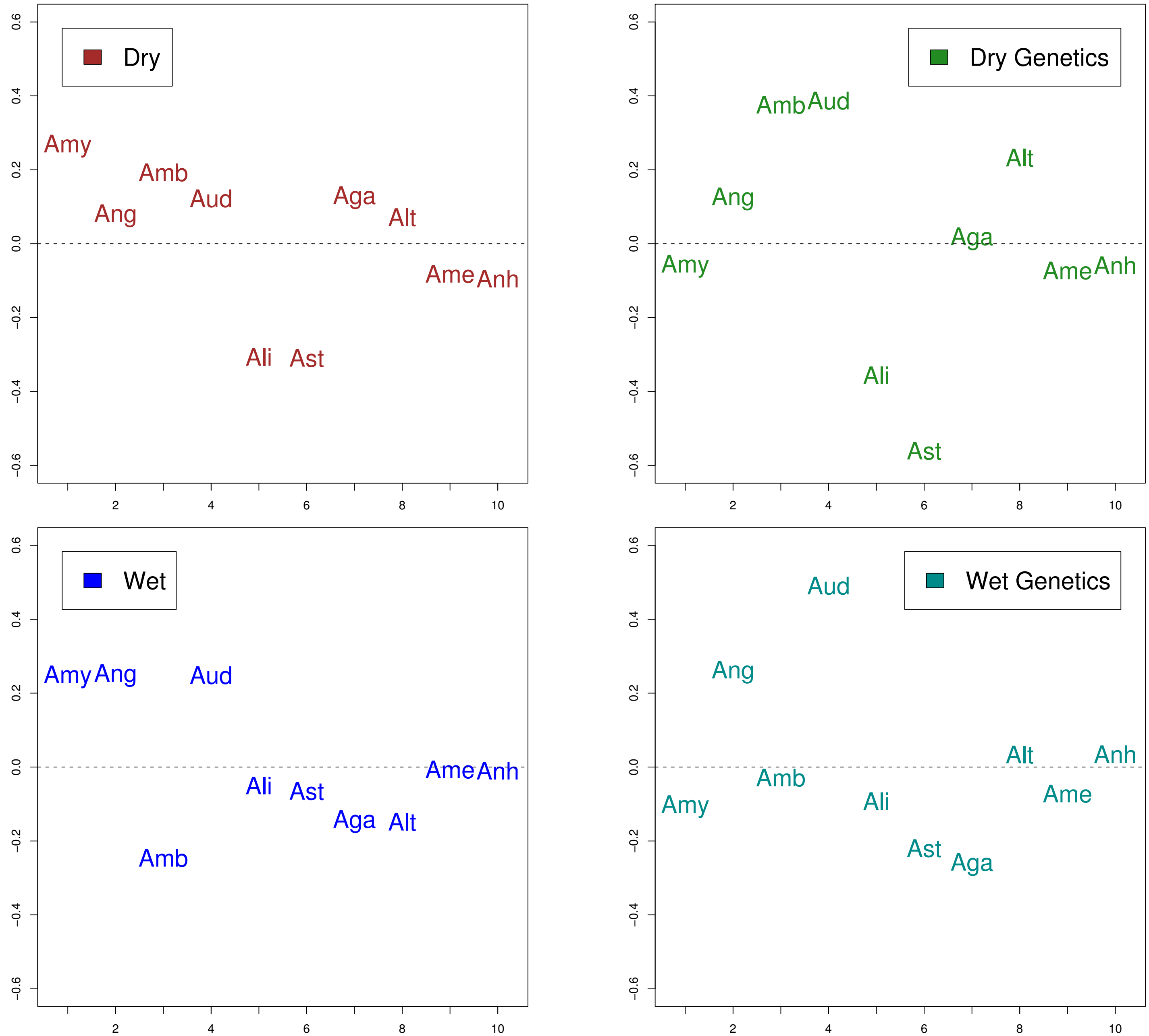


Elephant Family Results

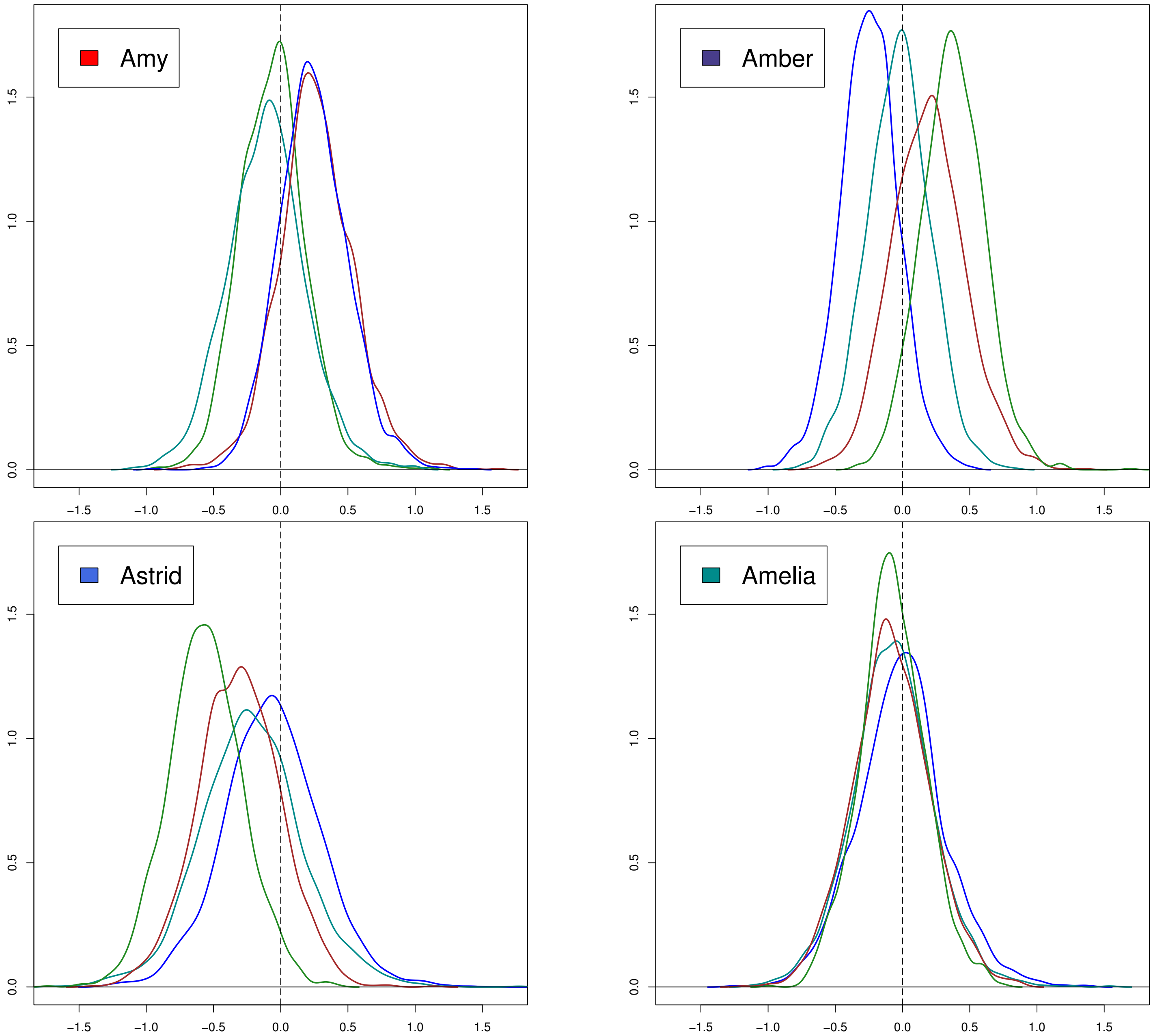
Posterior Intercepts β_0



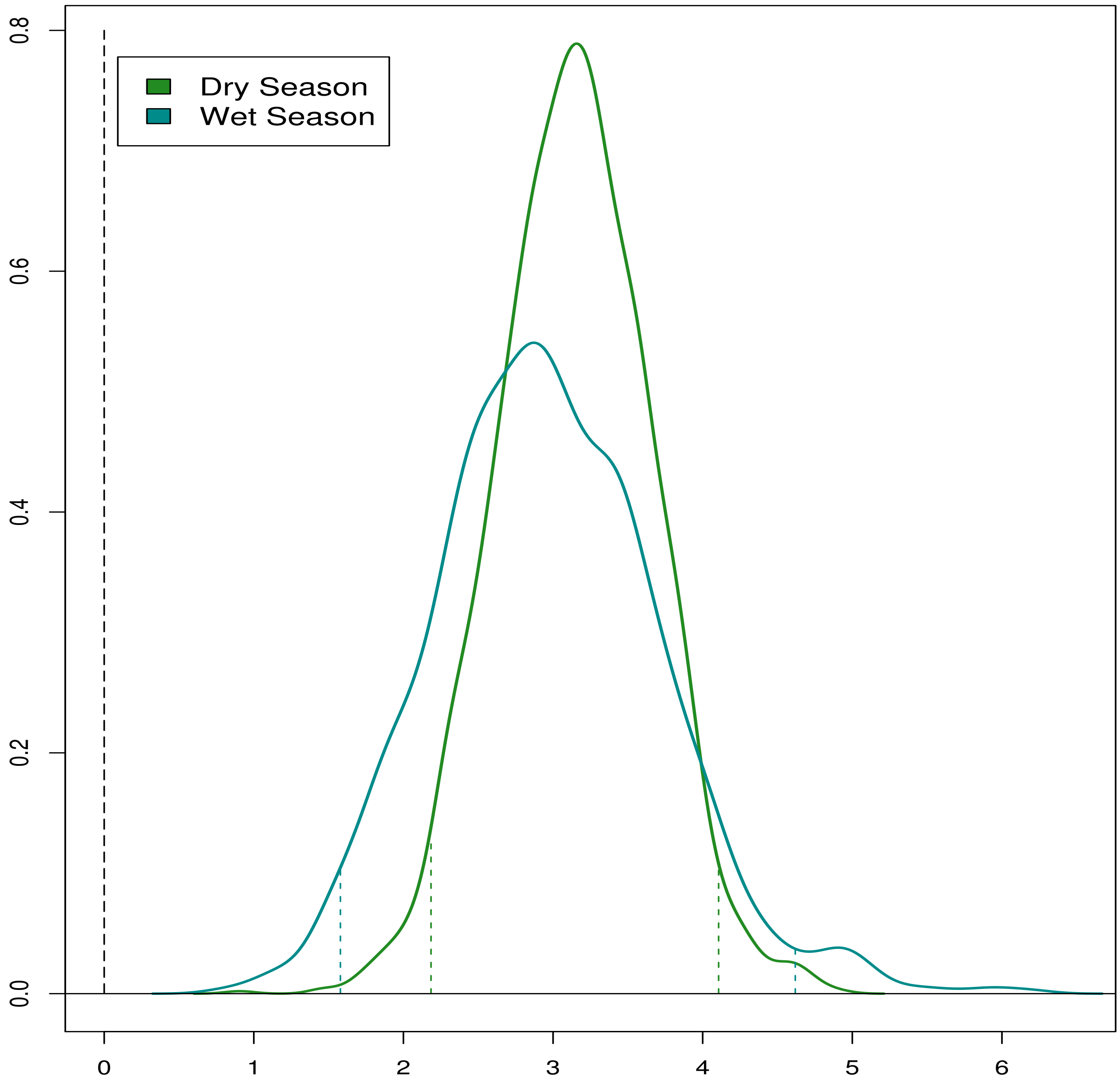
Posterior Sociabilities $\bar{a}$



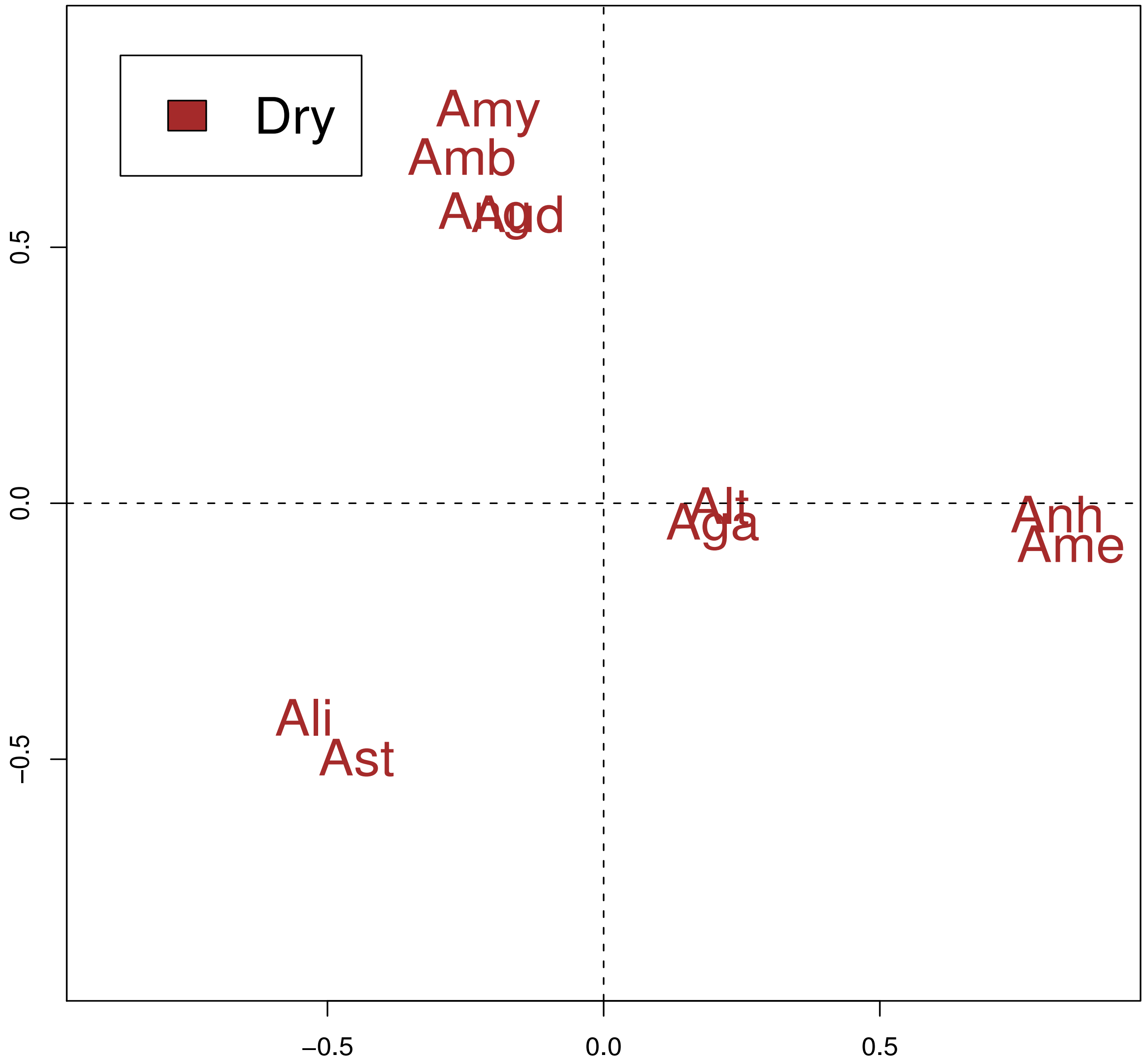
Posterior Sociabilities a_i



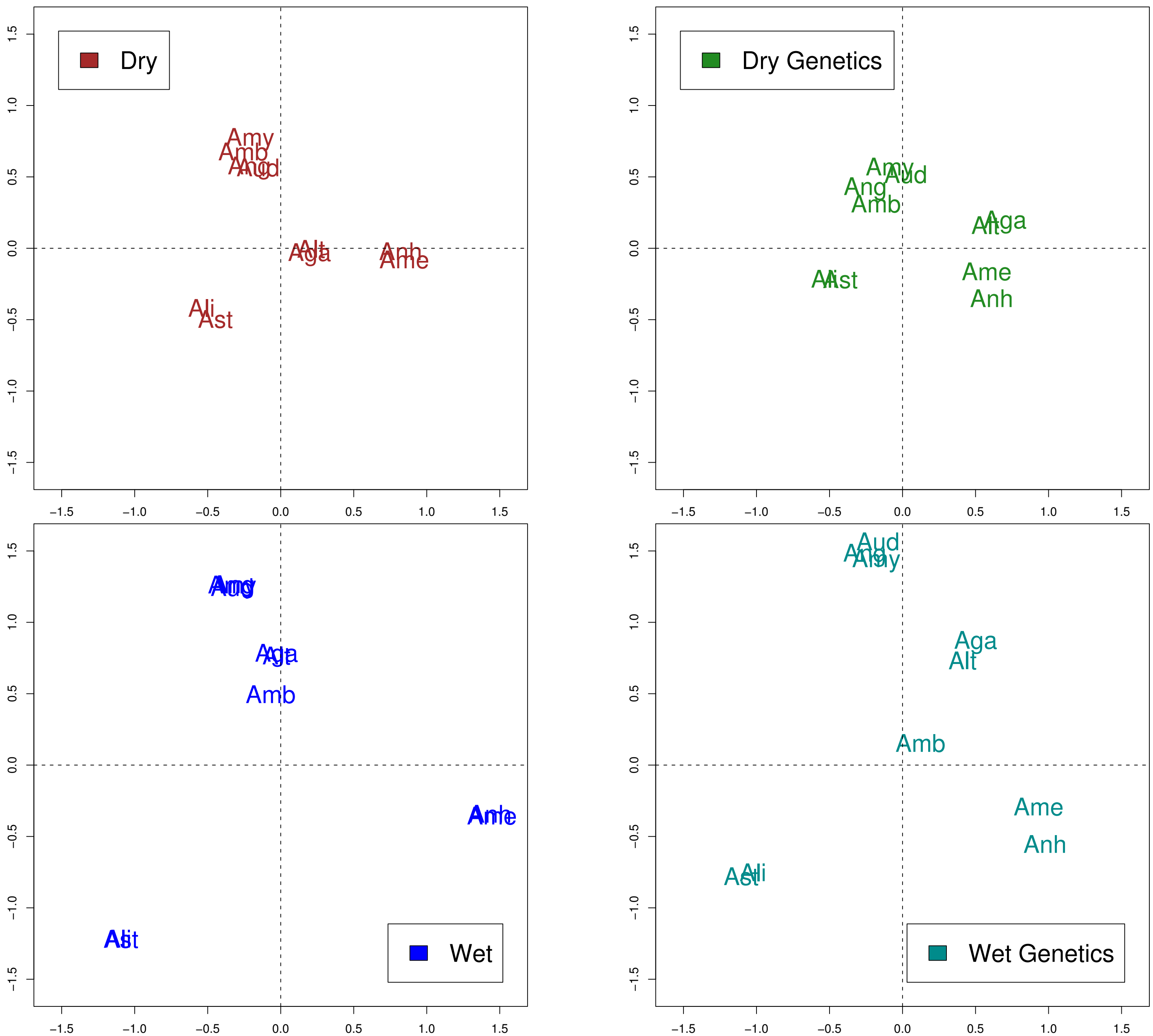
Posteriors for Genetic Coefficients β_g



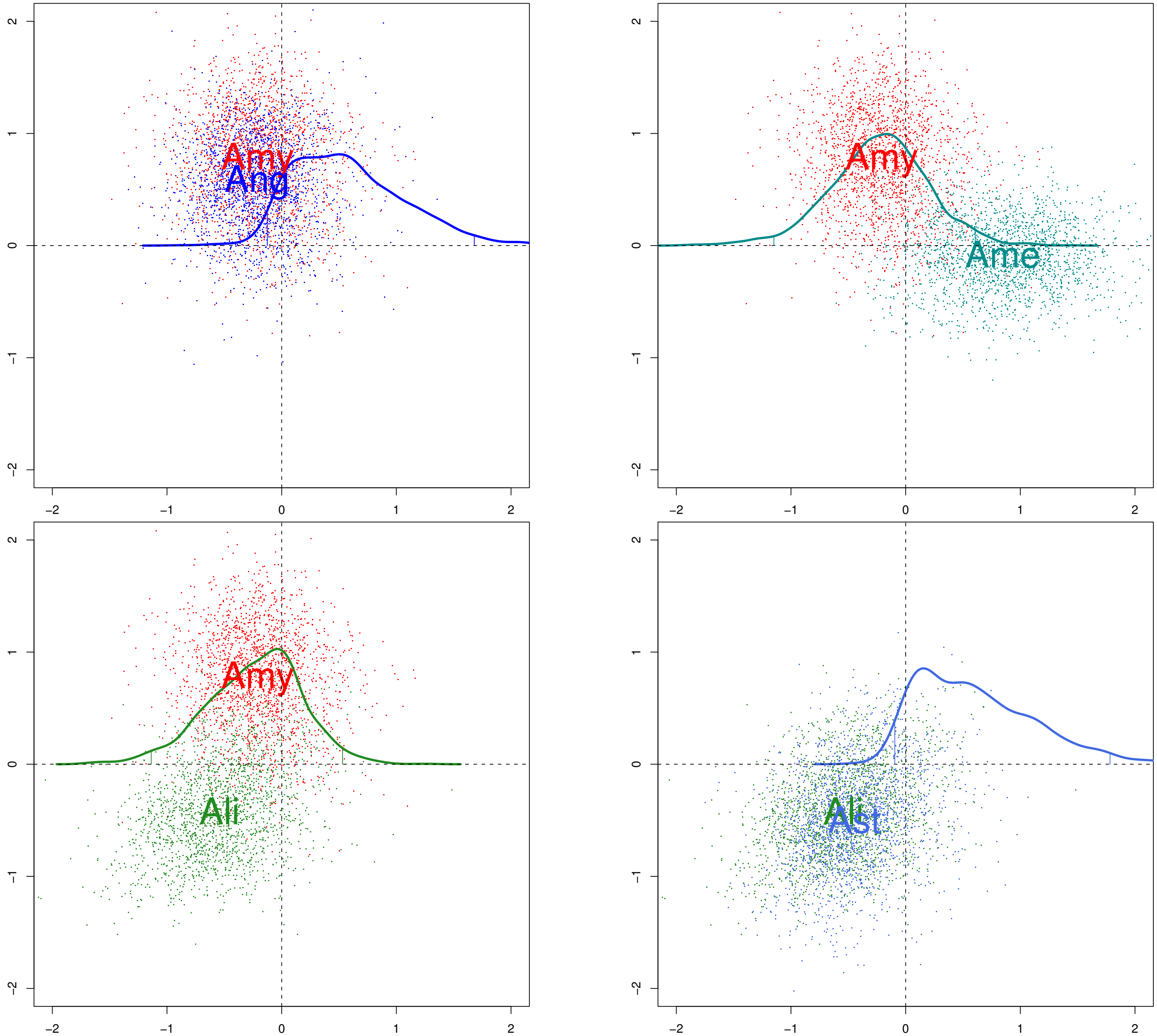
Dry Season Social Space z_i Posteriors



Social Space Posterior Means $\bar{z}_i$



Posterior Innerproducts $z_i' z_j$

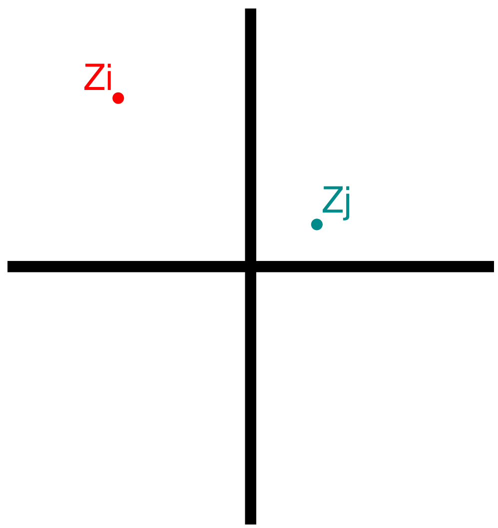


Pickiness in Social Space

- When an elephant is farther from the origin in social space, the inner product term increases and has a larger influence in the model.

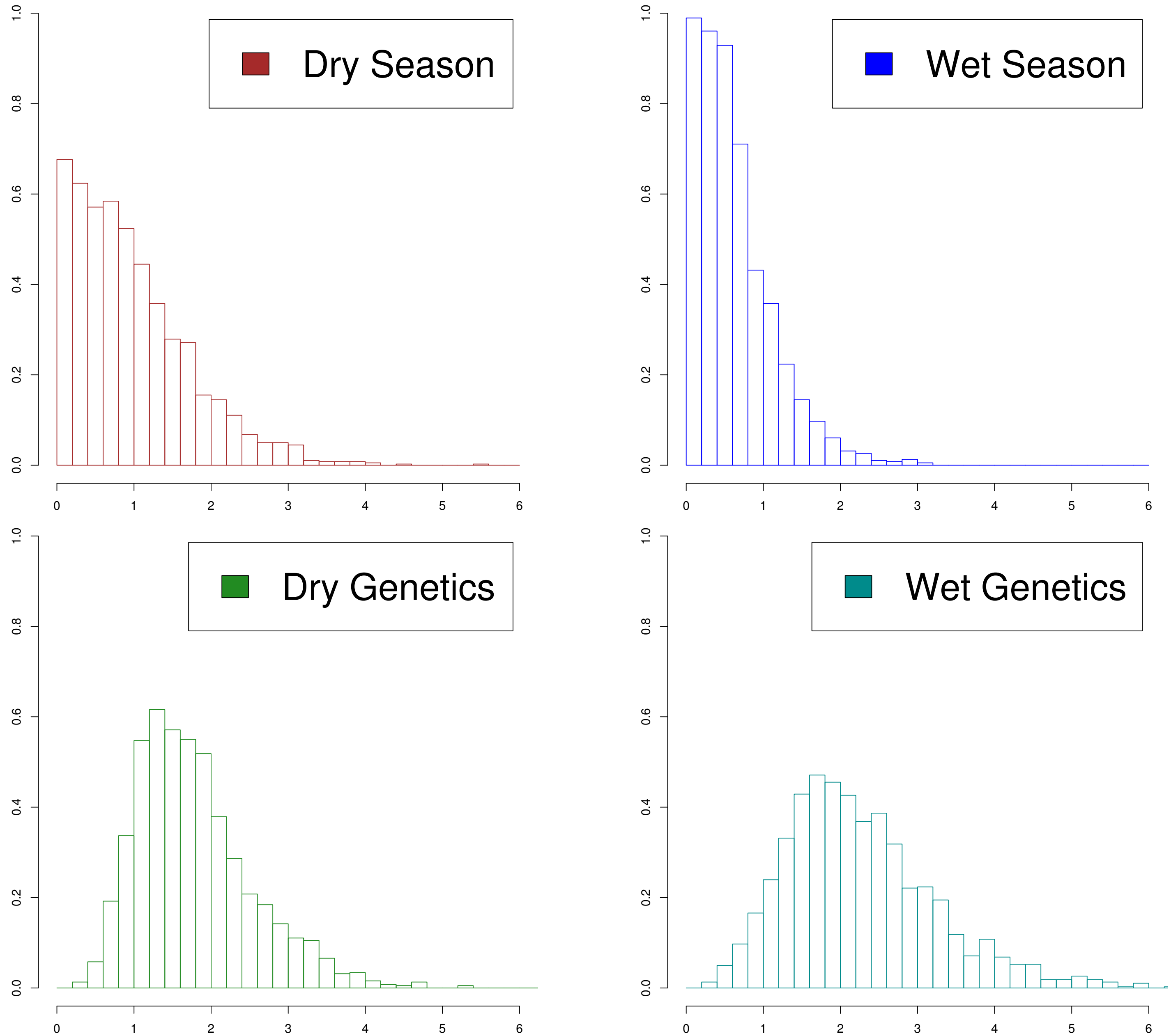
- An elephant close to the origin will have a small pairwise effect in the model.

- Pickiness is defined as the length of the vector in social space $|\mathbf{z_i}|$.

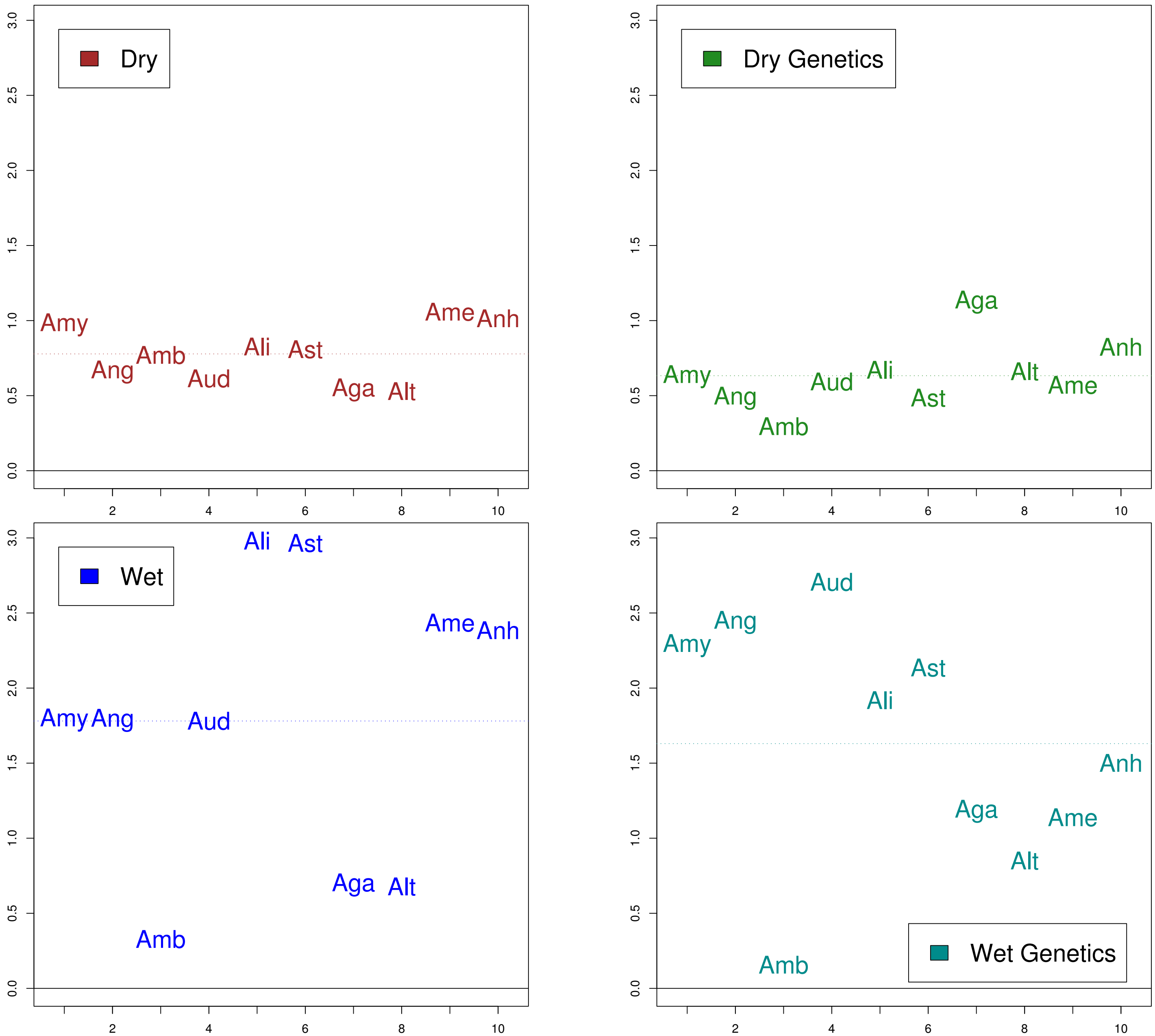


- Elephant i is "pickier" than elephant j.

Amy's Posterior Pickiness



Posterior Pickiness $|z_i|$



Conclusions

- As the matriarch of the family much of Amy's sociability is due to her genetics.
- Some of the posterior sociabilities in Family AA change depending on the season.
- Posterior intercepts β_0 for the Wet seasons are greater than in the Dry seasons, indicating that the elephants are more gregarious during the Wet season.
- Genetic coefficients $\beta_g > 0$ for both Wet and Dry seasons.
- There are clusters of elephants in social space.
- Some elephants are pickier than others, especially in the Wet season.

Future Research

- Include a seasonal intercept (indicator variable) in order to run Wet and Dry season data sets together.

- Use the real, (partially missing) observation matrices instead of the binomial summarized data.

$$\begin{bmatrix} \ddots & 1 & 1 & 1 & \vdots & 0 & 0 \\ 1 & \ddots & 1 & 1 & \vdots & 0 & 0 \\ 1 & 1 & \ddots & 1 & \vdots & 0 & 0 \\ 1 & 1 & 1 & \ddots & \vdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \vdots & \ddots & ? \\ 0 & 0 & 0 & 0 & \vdots & ? & \ddots \end{bmatrix}$$

-

Amy, Matriarch of Family AA

Gibbs Sampling

$$\theta_{ij} = \underbrace{\left(\tfrac{1}{2}\beta_0 + a_i\right)}_{\mathbf{s}} + \underbrace{\left(\tfrac{1}{2}\beta_0 + a_j\right)}_{\mathbf{r}} + \beta_g \, \mathrm{g_{ij}} + \gamma_{ij} + z_i' z_j$$

1. Sample linear effects:
   a) Sample $\beta_d, s, r \,|\, \beta_0, \sigma_a^2, \sigma_\gamma^2, \theta, Z$.
   b) Sample $\beta_0 \,|\, s, r, \sigma_a^2$.
   c) Sample $\sigma_a^2, \sigma_\gamma^2 \,|\, \beta_0, s, r$.

2. Sample bilinear effects:
   a) Sample $z_i \,|\, Z_{-i}, \theta, \beta, s, r, \sigma_z^2, \sigma_\gamma^2$.
   b) Sample $\sigma_z^2 \,|\, Z \sim \mathrm{IG}\left(\frac{1}{2} + \frac{nk}{2}, \frac{1}{2} + \frac{\mathrm{tr}(Z'Z)}{2}\right)$.

3. Update θ_{ij} with a Metropolis step:
   a) Propose $\theta_{ij}^* \sim \mathrm{N}\left(\left(\frac{1}{2}\beta_0 + a_i\right) + \left(\frac{1}{2}\beta_0 + a_j\right) + \beta_g \, \mathrm{g_{ij}} + z_i' z_j \,,\, \sigma_\gamma^2\right)$.
   b) Accept θ_{ij}^* with probability $\left(\frac{p(y_{ij}|\theta_{ij}^*)}{p(y_{ij}|\theta_{ij})} \wedge 1\right)$.

Selecting Dimension k of Social Space

The method of choosing the dimension k of social space depends on the goal of the analysis.

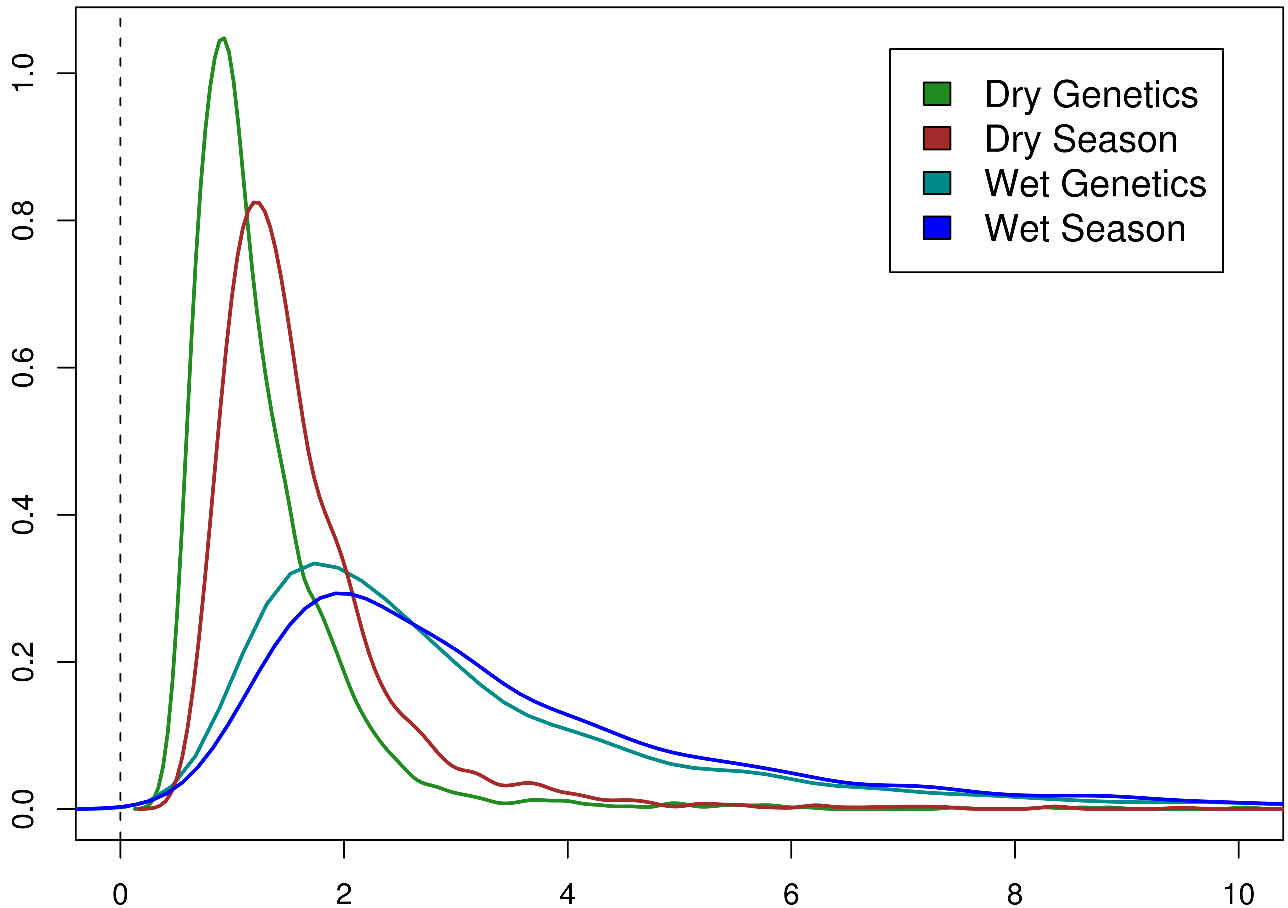
1. Descriptive:
   - Choose $k = 2$ to give easily interpretable results.

2. Assessing model fit: how well does the model explain the data?
   - Various model selection techniques
     - My choice would be to use stochastic search variable selection with point-mass 0 mixture priors on $\sigma^2_{z_1}, \sigma^2_{z_2}, \sigma^2_{z_3}, \ldots$.
     - Requires proper priors.

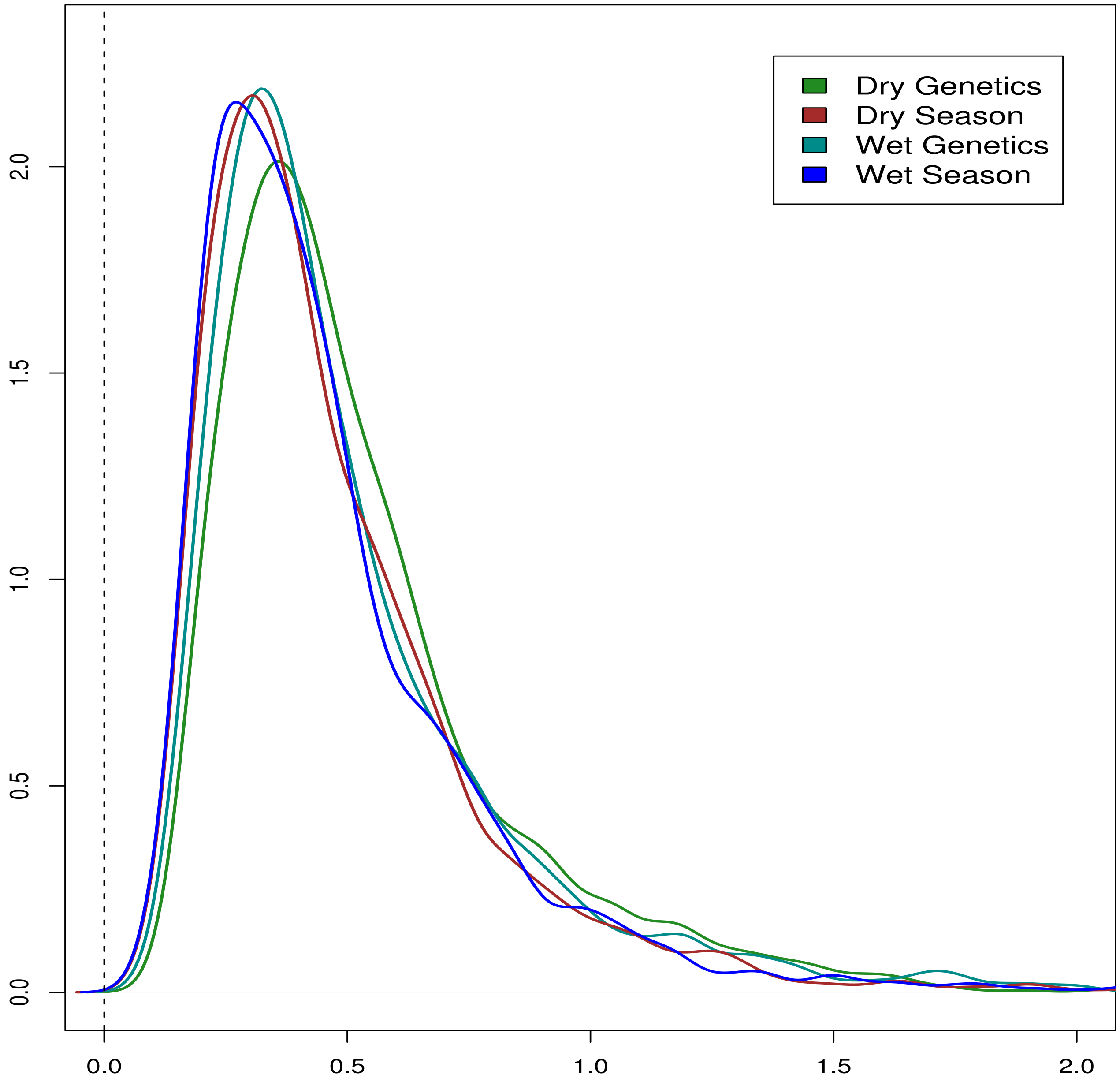
3. Make predictions of unobserved data.
   - Cross-validation techniques.
     - Select k based on the predictive performance $= \sum_{l=1}^{4} \sum_{\{i,j\} \in \mathrm{A}_l} \log p(y_{ij} | \hat{\theta}_{ij})$, where $\hat{\theta}_{ij}$ is the posterior mean excluding pairs in A_l.

Decomposing Error ϵ_{ij}^2

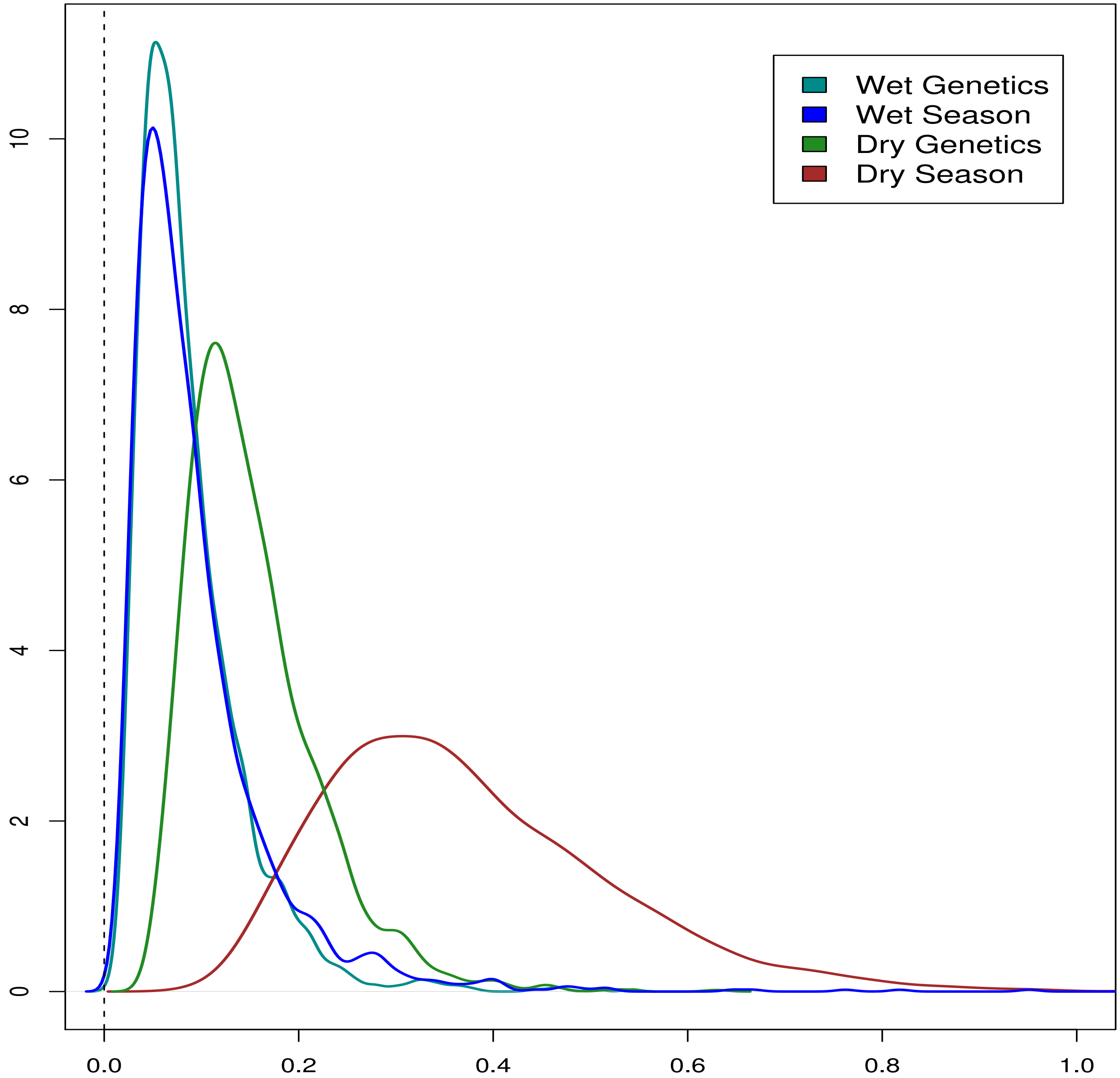
$$\begin{aligned}
\theta_{ij} &= (\tfrac{1}{2}\beta_0 + a_i) + (\tfrac{1}{2}\beta_0 + a_j) + \beta_g\, \mathrm{g}_{ij} + \gamma_{ij} + z_i' z_j \\
&= \beta_0 + \beta_g\, \mathrm{g}_{ij} + \epsilon_{ij} \\
\epsilon_{ij} &= a_i + a_j + \gamma_{ij} + z_i' z_j \\
\mathrm{E}(\epsilon_{ij}^2) &= 2\sigma_a^2 + \sigma_\gamma^2 + \sigma_{z_1}^4 + \sigma_{z_2}^4
\end{aligned}$$



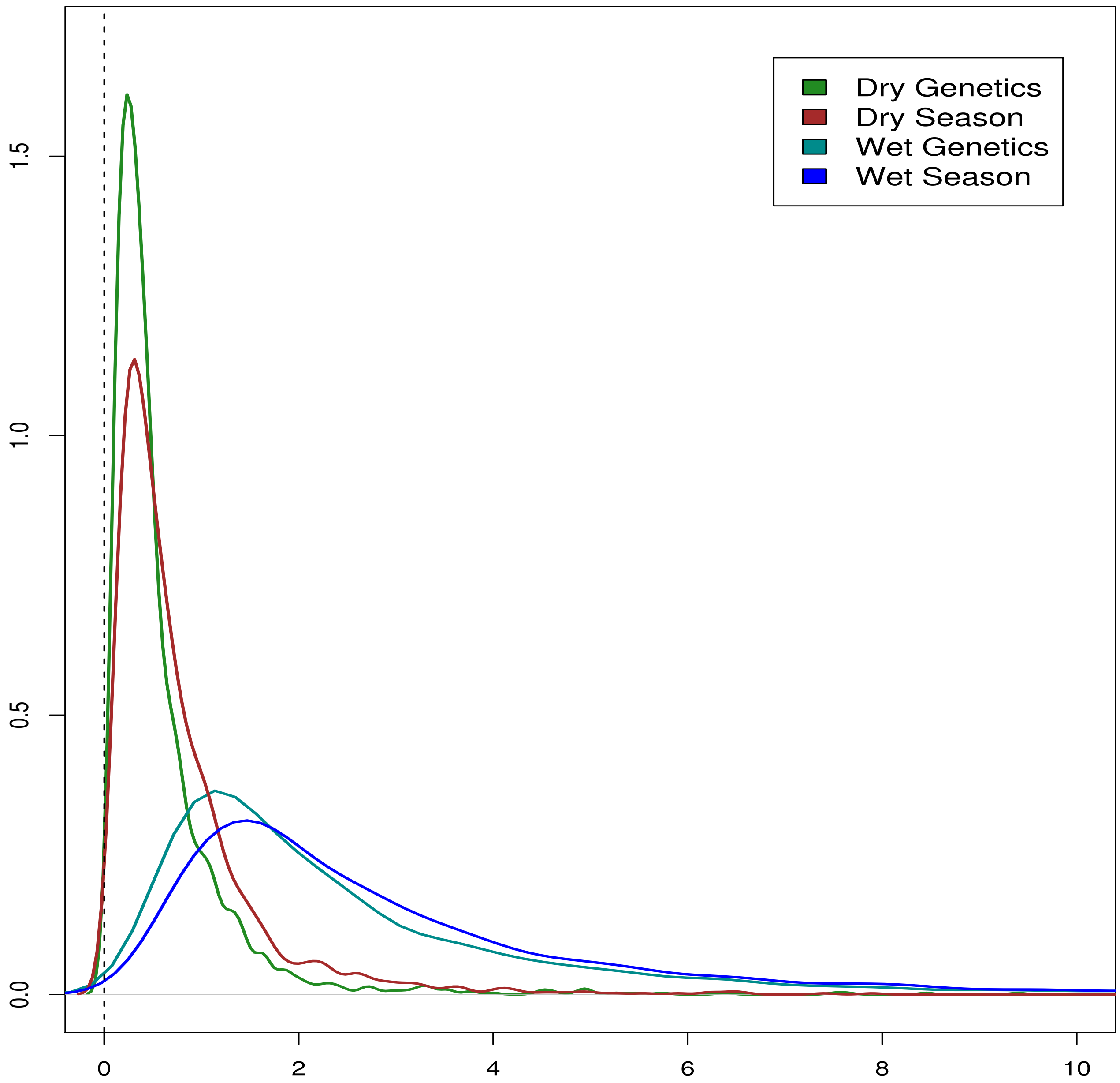
Posterior Sociability Variance $2\sigma_a^2$



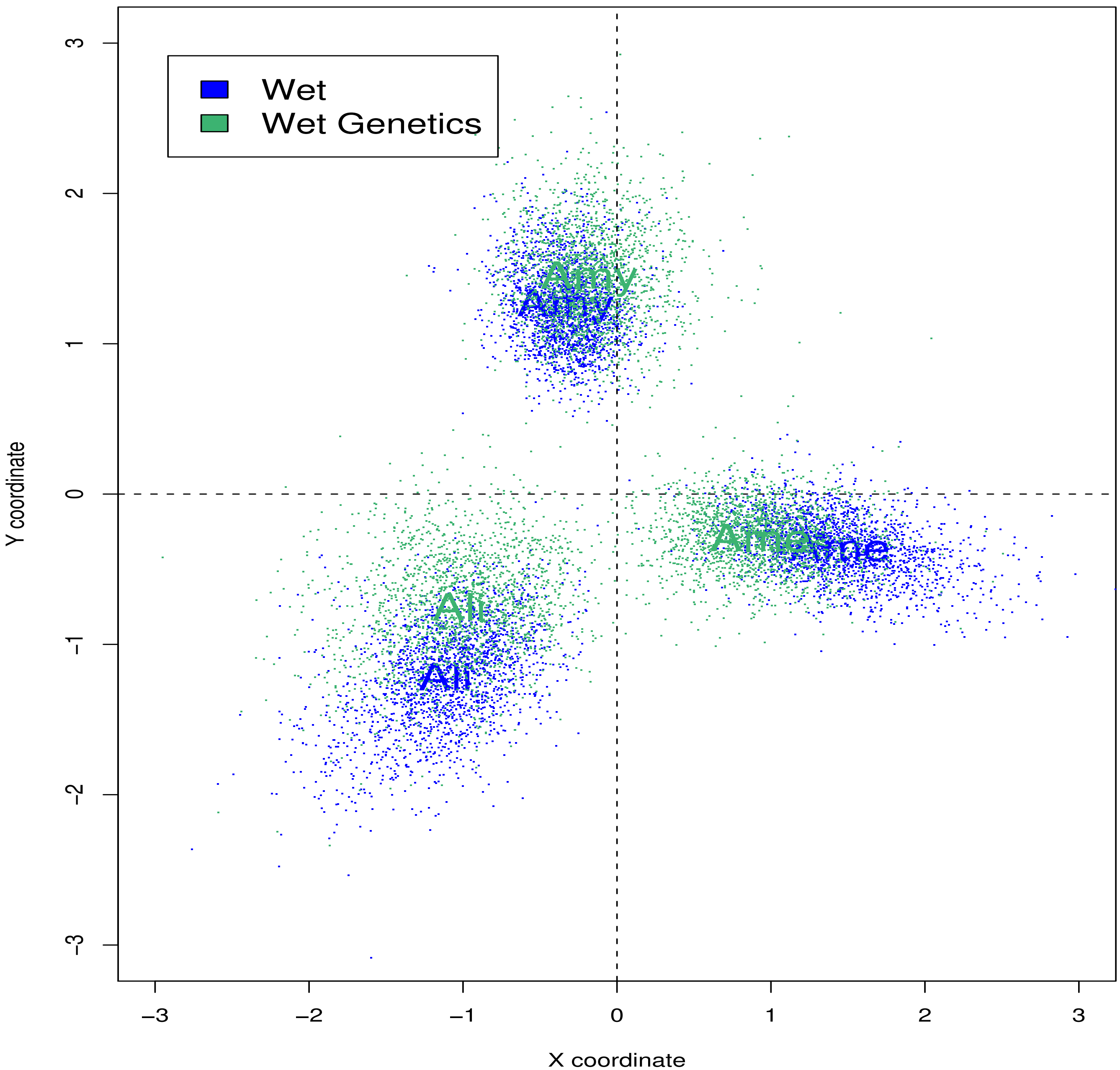
Posterior Normal Error σ_γ^2



Posterior Social Space Variance $(\sigma^4_{z_1} + \sigma^4_{z_2})$



Three Elephants' Social Space Posterior Draws



Raw Data

Date	Time	Spread	Subgroups?	Amy	Ang	Amb	Aud	Ali	Ast	Aga	Alt	Ame	Anh
001024	11:15	45	NO	1	1	1	1					1	1
001024	11:25	70	NO	1	1	1	1					1	1
001024	11:35	80	NO	1	1	1	1					1	1
001024	11:38	10	NO							1	1		
001024	11:48	10	NO							1	1		
001024	11:58	35	NO	1	1	1	1			1	1	1	1
001030	10:50	55	NO					1	1	1	1	1	1
001030	11:00	45	NO					1	1	1	1	1	1
001030	11:10	100	NO					1	1	1	1	1	1
001030	11:20	250	YES					1	1	1	1		
001030	11:20	250	YES									1	1
001030	11:30		YES					1	1	1	1		
001030	11:30		YES									1	1
010618	11:39	120	YES					1	1				
010618	11:39	120	YES	1		1						1	1
010618	11:49	110	NO	1		1		1	1	1	1	1	1
010618	11:59	120	YES					1	1	1	1		
010618	11:59	120	YES	1		1						1	1
010618	12:09	130	NO	1		1	1	1	1	1	1	1	1

Family AA Dry Observations

DRY	Amy	Amy	Ang	Ang	Amb	Amb	Aud	Aud	Ali	Ali	Ast	Ast	Aga	Aga	Alt	Alt	Ame	Ame
	P	A	P	A	P	A	P	A	P	A	P	A	P	A	P	A	P	A
Amy P	272	0	237	7	245	2	245	8	154	64	147	61	215	54	205	54	182	68
Amy A	0	159	35	152	27	157	27	151	118	95	125	98	57	105	67	105	90	91
Ang P	237	7	244	0	224	23	217	36	154	64	142	66	189	80	179	80	162	88
Ang A	35	152	0	187	20	164	27	151	90	123	102	121	55	107	65	107	82	99
Amb P	245	2	224	23	247	0	220	33	153	65	141	67	201	68	194	65	174	76
Amb A	27	157	20	164	0	184	27	151	94	119	106	117	46	116	53	119	73	108
Aud P	245	8	217	36	220	33	253	0	149	69	145	63	205	64	197	62	170	80
Aud A	27	151	27	151	27	151	0	178	104	109	108	115	48	114	56	116	83	98
Ali P	154	64	154	64	153	65	149	69	218	0	191	17	147	122	144	115	115	135
Ali A	118	95	90	123	94	119	104	109	0	213	27	196	71	91	74	98	103	78
Ast P	147	61	142	66	141	67	145	63	191	17	208	0	148	121	145	114	121	129
Ast A	125	98	102	121	106	117	108	115	27	196	0	223	60	102	63	109	87	94
Aga P	215	54	189	80	201	68	205	64	147	122	148	121	269	0	257	2	193	57
Aga A	57	105	55	107	46	116	48	114	71	91	60	102	0	162	12	160	76	105
Alt P	205	54	179	80	194	65	197	62	144	115	145	114	257	2	259	0	185	65
Alt A	67	105	65	107	53	119	56	116	74	98	63	109	12	160	0	172	74	107
Ame P	182	68	162	88	174	76	170	80	115	135	121	129	193	57	185	65	250	0
Ame A	90	91	82	99	73	108	83	98	103	78	87	94	76	105	74	107	0	181

Genetic Relatedness

AA	Amy	Angelina	Amber	Audrey	Alison	Astrid	Agatha	Althea	Amelia
Amy									
Angelina	0.39								
Amber	0.46	0.26							
Audrey	0.31	0.22	0.11						
Alison	0.3	0.02	0.08	0.02					
Astrid	0.25	0.18	0.02	0.14	0.53				
Agatha	0.27	0.08	0.25	-0.15	0.38	0.34			
Althea	0.05	0.05	0.12	-0.01	0.15	0.21	0.46		
Amelia	0.35	0.06	-0.09	0.01	0.09	0.2	0.1	0.02	
Anghared	0.25	0.34	0.02	0.2	-0.06	0.1	0.06	-0.04	0.52