Priors and the number of bilinear terms

Muhammad Yaseen

2026-11-04

Table of contents

1 Abstract **1**

2 Vague priors **1**

3 Informative priors **2**

4 How many terms? **3**

5 The posterior of the error variance with more terms **5**

6 Information criteria **5**

7 References **5**

1 Abstract

The defaults of `bam_priors()` are vague: the data dominate the posterior. This vignette shows how to set informative priors from earlier trials and how the posterior of the last singular value guides the choice of the number of bilinear terms K.

2 Vague priors

```
f0 <-
  bayes_ammi(
      .data   = Maiz
    , .y      = y
    , .gen    = entry
    , .env    = site
    , .rep    = rep
    , .nIter  = 3000
```

```
    , .k      = 2
    , .burnin = 500
    , .thin   = 1
    , .chains = 1
    , .priors = bam_priors(
          mu_mean     = 0
        , mu_var      = 1e+8
        , tau_mean    = 0
        , tau_var     = 1e+8
        , delta_mean  = 0
        , delta_var   = 1e+8
        , lambda_mean = 0
        , lambda_var  = 1e+8
        , nu          = 1
        , s2          = NULL
        )
    , .level  = 0.95
    , .seed   = 6
    , .align  = TRUE
    , .model  = "AMMI"
    , .het    = FALSE
    , .waic   = TRUE
    )
f0$summary[block %in% c("lambda", "sigma2"), list(parameter, mean, sd,
 ↪  lower, upper)]
```

```
   parameter       mean         sd        lower      upper
1:    sigma2 763107.479 43844.8853 686109.01145 857075.929
2: lambda[1]   5462.493   463.4891   4552.72387   6357.336
3: lambda[2]   1682.008   783.5644     93.71569   2947.843
```

3 Informative priors

Suppose earlier trials in the region give an error variance of about 8×10^5, with the information of 50 plots, and a first singular value near 5000:

```
f1 <-
  bayes_ammi(
      .data   = Maiz
    , .y      = y
    , .gen    = entry
    , .env    = site
    , .rep    = rep
    , .nIter  = 3000
    , .k      = 2
    , .burnin = 500
    , .thin   = 1
    , .chains = 1
```

```
    , .priors = bam_priors(
          mu_mean     = 0
        , mu_var      = 1e+8
        , tau_mean    = 0
        , tau_var     = 1e+8
        , delta_mean  = 0
        , delta_var   = 1e+8
        , lambda_mean = 5000
        , lambda_var  = 250^2
        , nu          = 50
        , s2          = 8e5
        )
    , .level  = 0.95
    , .seed   = 6
    , .align  = TRUE
    , .model  = "AMMI"
    , .het    = FALSE
    , .waic   = TRUE
    )
pm <- rbind(vague       = f0$summary[block %in% c("lambda", "sigma2"),
 ↪  mean],
            informative = f1$summary[block %in% c("lambda", "sigma2"),
             ↪  mean])
colnames(pm) <- f0$summary[block %in% c("lambda", "sigma2"), parameter]
round(pm)
```

```
             sigma2 lambda[1] lambda[2]
vague        763107      5462      1682
informative 767603      5077      4449
```

The prior on λ_k applies to every term. An informative prior suited to the first term therefore also pulls the second towards 5000 (here from 1682 to 4449), a reason to keep `lambda_var` large unless the prior information is about all terms.

4 How many terms?

The paper fitted $K = 2$ terms. A term that the data do not support has a singular value whose posterior reaches down to zero:

```
fits <-
  lapply(
      X   = 1:3
    , FUN = function(K) {
        bayes_ammi(
            .data   = Maiz
          , .y      = y
          , .gen    = entry
          , .env    = site
```

```
          , .rep    = rep
          , .nIter  = 3000
          , .k      = K
          , .burnin = 500
          , .thin   = 1
          , .chains = 1
          , .priors = bam_priors(
                mu_mean     = 0
              , mu_var      = 1e+8
              , tau_mean    = 0
              , tau_var     = 1e+8
              , delta_mean  = 0
              , delta_var   = 1e+8
              , lambda_mean = 0
              , lambda_var  = 1e+8
              , nu          = 1
              , s2          = NULL
              )
          , .level  = 0.95
          , .seed   = 6
          , .align  = TRUE
          , .model  = "AMMI"
          , .het    = FALSE
          , .waic   = TRUE
          )
      }
    )
data.table::rbindlist(l = lapply(X = fits, FUN = function(f) {
  f$summary[block == "lambda", list(K = f$settings$k, parameter, mean,
 ↪  sd, lower, upper)]
}))
```

```
   K parameter      mean       sd        lower    upper
1: 1 lambda[1] 5451.2684 454.4051 4560.3608937 6317.846
2: 2 lambda[1] 5462.4934 463.4891 4552.7238741 6357.336
3: 2 lambda[2] 1682.0078 783.5644   93.7156918 2947.843
4: 3 lambda[1] 5481.9261 465.7028 4533.1825489 6356.072
5: 3 lambda[2] 1951.0831 612.2600  647.9023223 3076.930
6: 3 lambda[3]  901.4993 603.9840    0.1931663 1973.478
```

With $K = 2$ the HPD interval of λ_2 reaches down to 94, and with $K = 3$ that of λ_3 reaches down to 0.19. In both cases the last term is barely supported, whereas the interval of λ_1 stays far from zero. With a third term in the model, λ_2 moves away from zero because the order $\lambda_2 \geq \lambda_3$ gives the third term the weakest pattern. The least-squares singular values (5923, 3070, 2552) do not show any of this, because they are biased upwards. Cornelius et al. (1996) found two terms significant for these data, and Liu (2001, unpublished) suggested two or three.

5 The posterior of the error variance with more terms

```
vapply(
    X         = fits
  , FUN       = function(f) f$summary[parameter == "sigma2", mean]
  , FUN.VALUE = numeric(length = 1)
  )
```

```
[1] 781037.5 763107.5 753154.2
```

Each further term takes $G + E - 1 - 2k$ degrees of freedom from the residual, so the error variance decreases only slowly once the terms are weak.

6 Information criteria

WAIC compares the three fits by their estimated out-of-sample fit (vignette 9):

```
names(fits) <- paste0("K", 1:3)
cp <- do.call(what = bam_compare, args = fits)
cp
```

```
Comparison of Bayesian models (smaller is better)

   model terms variances      DIC      pD     WAIC  p_waic      se
1:    K3     3    common 11860.08 74.9912 11864.18 72.3308 44.4338
2:    K2     2    common 11862.78 68.2585 11866.86 66.4888 44.2687
3:    K1     1    common 11864.96 53.7269 11868.55 53.3171 43.6160
   lambda_last_lower lambda_last_mean  dWAIC dWAIC_se
1:            0.1932         901.4993 0.0000   0.0000
2:           93.7157        1682.0078 2.6782   1.8038
3:         4560.3609        5451.2684 4.3732   5.3261

Sources: Spiegelhalter et al. (2002); Watanabe (2010); Vehtari et al. (2017).
Full references: bam_cite(x).
```

WAIC prefers K3. `dWAIC` is the difference from the best model and `dWAIC_se` its standard error, computed from the plot-by-plot differences between the two models; a difference of less than about two of its standard errors does not separate the models.

7 References

Crossa, J., Perez-Elizalde, S., Jarquin, D., Cotes, J. M., Viele, K., Liu, G. and Cornelius, P. L. (2011). Bayesian estimation of the additive main effects and multiplicative interaction model. *Crop Science*, 51(4), 1458–1469.

doi:10.2135/cropsci2010.06.0343

Cornelius, P. L., Crossa, J. and Seyedsadr, M. S. (1996). Statistical tests and estimators of multiplicative models for genotype-by-environment interaction. In M. S. Kang and H. G. Gauch (eds), *Genotype-by-Environment Interaction*. CRC Press, Boca Raton, 199–234.

Cornelius, P. L. and Crossa, J. (1999). Prediction assessment of shrinkage estimators of multiplicative models for multi-environment cultivar trials. *Crop Science*, 39(4), 998–1009.

Hobert, J. P. and Casella, G. (1996). The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association*, 91(436), 1461–1473.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11, 3571–3594.

Gelman, A., Hwang, J. and Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24(6), 997–1016.