Decision tools for breeders

Muhammad Yaseen

2026-11-04

Table of contents

1 Abstract **1**

2 Probabilities of superior performance and stability **2**

3 Posterior ranks **5**

4 Bayesian stability **6**

5 Predictions **7**

6 Which genotype wins where **8**

7 Model choice **9**
7.1 Information criteria 9
7.2 Posterior predictive checks 10

8 Working with posterior, bayesplot and coda **12**

9 Interpretation **13**

10 References **14**

1 Abstract

A posterior sample answers the questions a breeder asks in probabilities: how likely is a genotype to be among the best, how stable is it, and what yield can be expected in a cell that was never grown. This vignette goes through `bam_superiority()`, `bam_ranks()`, `bam_stability()`, `bam_predict()` and `bam_winners()`, and through the model choice tools `bam_compare()` and `bam_ppc()`.

```
fit <-
  bayes_ammi(
```

```
    .data   = Maiz
  , .y      = y
  , .gen    = entry
  , .env    = site
  , .rep    = rep
  , .nIter  = 4000
  , .k      = 2
  , .burnin = 1000
  , .thin   = 1
  , .chains = 2
  , .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   = 9
  , .align  = TRUE
  , .model  = "AMMI"
  , .het    = FALSE
  , .waic   = TRUE
  )
```

2 Probabilities of superior performance and stability

Dias et al. (2022) select the best s fraction of G genotypes, that is $\lceil sG \rceil$ of them, in every draw. Write $\bar{\eta}_{i\cdot}$ for the mean of genotype i over the environments and $\sigma^2_{GE,i} = \frac{1}{E}\sum_j (\eta_{ij} - \bar{\eta}_{i\cdot} - \bar{\eta}_{\cdot j} + \bar{\eta}_{\cdot\cdot})^2$ for its interaction variance. Then

- the probability of superior performance is the share of draws in which $\bar{\eta}_{i\cdot}$ is among the best $\lceil sG \rceil$;
- the probability of superior stability is the share in which $\sigma^2_{GE,i}$ is among the smallest $\lceil sG \rceil$;
- the joint probability is their product;
- the conditional probability is the probability of superior performance within each environment;
- the pairwise probability is $P(\bar{\eta}_{i\cdot} > \bar{\eta}_{i'\cdot})$.

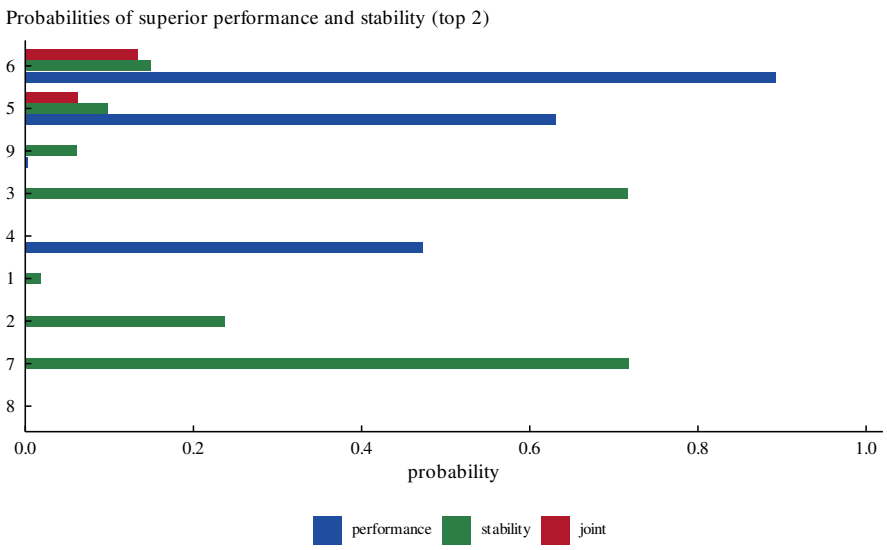
```
sp <-
  bam_superiority(
      fit           = fit
    , .s            = 0.2
    , higher_better = TRUE
    )
sp
```

```
Probabilities of superior performance and stability (selection of 2 of 9 genotypes)

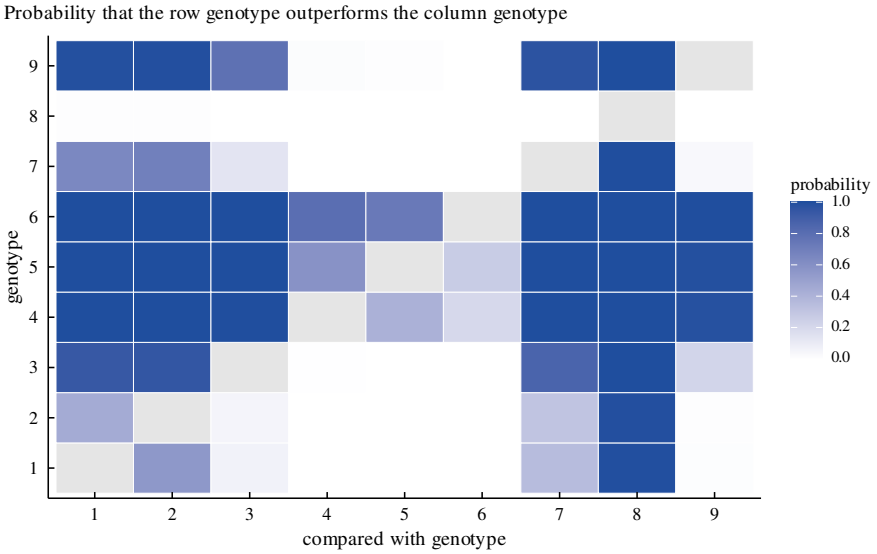
   Gen mean prob_performance prob_stability prob_joint
1:   6 5330         0.892250       0.149750  1.336e-01
2:   5 5246         0.631250       0.098375  6.210e-02
3:   9 4929         0.003125       0.061625  1.926e-04
4:   3 4822         0.000125       0.716375  8.955e-05
5:   4 5217         0.473250       0.000125  5.916e-05
6:   1 4619         0.000000       0.018000  0.000e+00
7:   2 4603         0.000000       0.237375  0.000e+00
8:   7 4672         0.000000       0.718125  0.000e+00
9:   8 4285         0.000000       0.000250  0.000e+00

Sources: Dias et al. (2022). Full references: bam_cite(x).
```

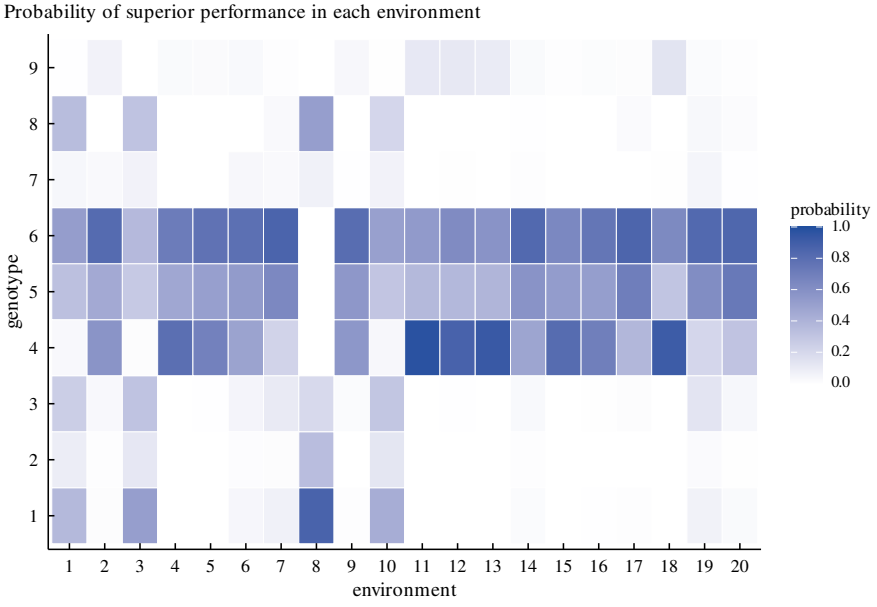
```
plot(x = sp, which = "marginal")
```



```
plot(x = sp, which = "pairwise")
```




```
plot(x = sp, which = "conditional")
```



3 Posterior ranks

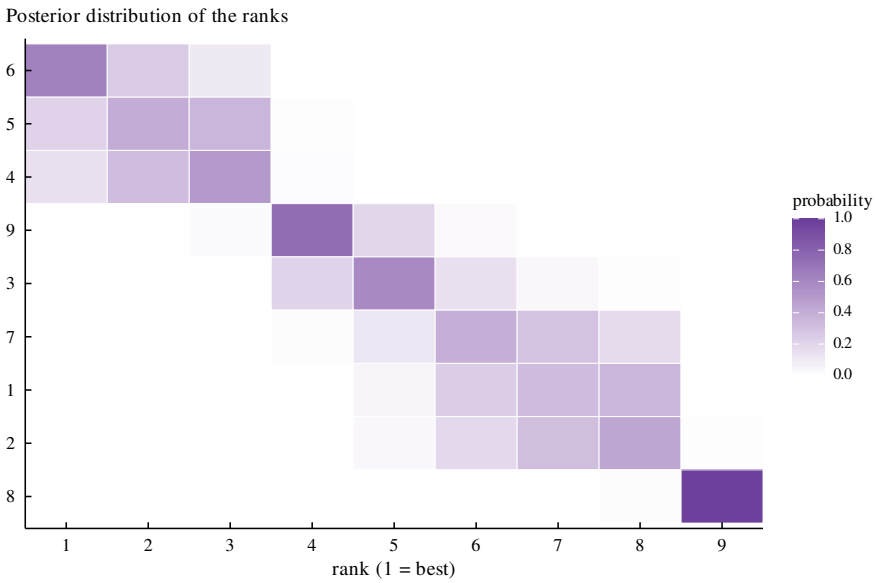
```
rk <-
  bam_ranks(
      fit           = fit
    , higher_better = TRUE
    , .env          = NULL
    )
rk
```

```
Posterior ranks of the genotypes (means over the environments)

   Gen mean_rank median_rank lower upper prob_first
1:   6     1.474           1     1     3   0.634500
2:   5     2.158           2     1     3   0.221000
3:   4     2.402           3     1     3   0.144375
4:   9     4.245           4     3     5   0.000125
5:   3     5.037           5     4     6   0.000000
6:   7     6.492           6     5     8   0.000000
7:   1     7.020           7     5     8   0.000000
8:   2     7.190           7     6     8   0.000000
9:   8     8.981           9     9     9   0.000000

Sources: Dias et al. (2022). Full references: bam_cite(x).
```

```
plot(x = rk)
```



The tile of genotype i and rank r shows $P(\text{rank}_i = r)$. A rank interval that spans several places is the honest answer to "which is second?". Ranks within one environment come from `.env`:

```
bam_ranks(
    fit           = fit
  , higher_better = TRUE
  , .env          = "12"
  )$summary[1:3]
```

```
   Gen mean_rank median_rank lower upper prob_first
1:   4  1.527375           1     1     3   0.624750
2:   6  2.262375           2     1     4   0.229625
3:   5  2.785250           3     1     4   0.115000
```

4 Bayesian stability

`bam_stability()` gives the posterior of two stability measures for every genotype: the interaction variance of Shukla (1972), and the AMMI stability value of Purchase et al. (2000),

$$\text{ASV}_i = \sqrt{\left(\frac{\lambda_1}{\lambda_2}\,\lambda_1^{1/2}\alpha_{i1}\right)^2 + (\lambda_2^{1/2}\alpha_{i2})^2}.$$

The weight λ_1/λ_2 becomes unstable when λ_2 is drawn near zero, so it is computed from the posterior means of λ_1 and λ_2, and the scores from each draw. The yield stability index (YSI) adds the mean rank of the yield to the mean rank of the ASV; small is good.

```
st <-
  bam_stability(
      fit           = fit
    , higher_better = TRUE
    )
st
```

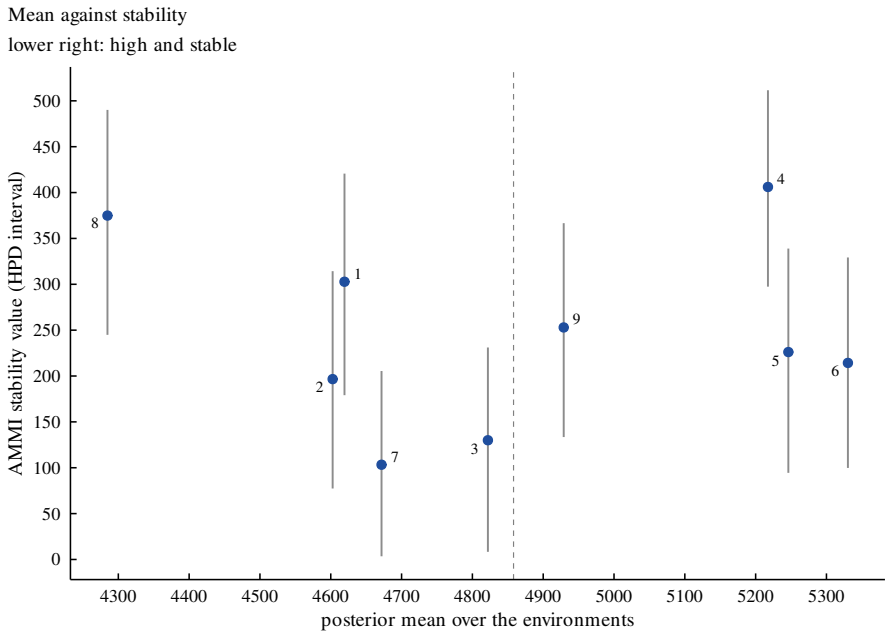
```
Bayesian stability: AMMI stability value and interaction variance
```

```
   Gen mean   ASV ASV_lower ASV_upper variance prob_most_stable rank_mean
1:   6 5330 214.3    99.792     329.2   130842          0.02038     1.474
2:   5 5246 226.1    94.448     338.9   154034          0.01638     2.158
3:   3 4822 130.0     8.340     231.1    58903          0.32462     5.037
4:   7 4672 103.3     3.478     205.4    59410          0.56075     6.492
5:   9 4929 252.9   133.463     366.6   176205          0.00575     4.245
6:   4 5217 406.1   297.425     511.5   414149          0.00000     2.402
7:   2 4603 196.6    77.373     314.3   119325          0.06838     7.190
8:   1 4619 302.8   179.107     420.6   248306          0.00350     7.020
```

```
9:   8 4285 374.9   244.894     490.1   406395          0.00025     8.981
   rank_ASV    YSI
1:    4.350  5.824
2:    4.674  6.832
3:    2.228  7.265
4:    1.725  8.218
5:    5.307  9.553
6:    8.428 10.830
7:    3.815 11.006
8:    6.574 13.594
9:    7.899 16.880
```

Sources: Purchase et al. (2000); Shukla (1972); Farshadfar (2008). Full references: bam_cite(x).

```
plot(x = st)
```



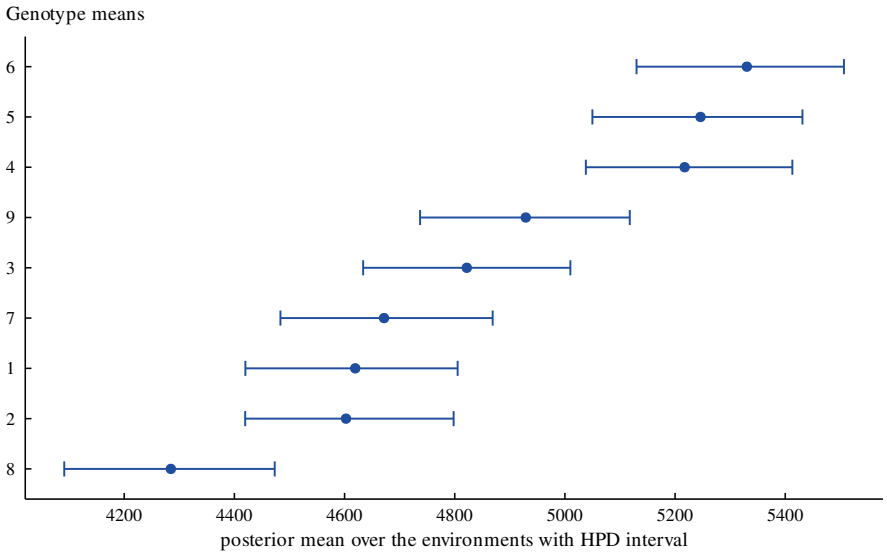
5 Predictions

```
pr <-
  bam_predict(
      fit    = fit
    , .level = 0.95
    )
```

```
pr$genotypes
```

```
   Gen     mean       sd    lower    upper
1:   1 4619.380 97.97855 4419.973 4805.276
2:   2 4602.657 97.04919 4419.602 4798.008
3:   3 4821.905 95.76501 4633.795 5009.925
4:   4 5217.376 96.63632 5038.035 5412.883
5:   5 5246.185 97.27219 5049.912 5431.217
6:   6 5330.296 96.80101 5130.001 5506.541
7:   7 4671.682 98.70976 4483.616 4868.877
8:   8 4284.657 97.54336 4091.166 4473.218
9:   9 4928.959 98.21164 4737.114 5117.745
```

```
plot(x = pr, which = "genotypes")
```



6 Which genotype wins where

```
bam_winners(
    fit           = fit
  , higher_better = TRUE
  )$mega
```

```
   Winner                        Environments  n mean_probability
1:      6 1, 2, 5, 6, 7, 9, 10, 14, 17, 19, 20 11        0.4755682
2:      4            4, 11, 12, 13, 15, 16, 18  7        0.6283929
3:      1                                3, 8  2        0.4708750
```

7 Model choice

7.1 Information criteria

`bayes_ammi()` stores the deviance information criterion (DIC; Spiegelhalter et al. 2002) and the widely applicable information criterion (WAIC; Watanabe 2010) of the observed plots. Both are computed while the chains run, so no draws need to be kept. `bam_compare()` puts several fits side by side:

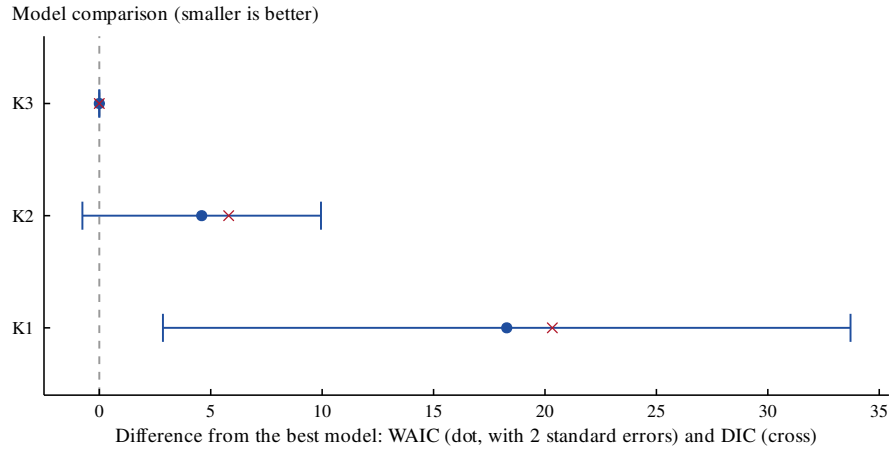
```
fits <-
  lapply(
      X   = 1:3
    , FUN = function(K) {
        bayes_ammi(
            .data   = Maiz
          , .y      = y
          , .gen    = entry
          , .env    = site
          , .rep    = rep
          , .nIter  = 2000
          , .k      = K
          , .burnin = 500
          , .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   = 9
          , .align  = TRUE
          , .model  = "AMMI"
          , .het    = TRUE
          , .waic   = TRUE
          )
      }
    )
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 per environment 11730.38 94.7238 11737.53 90.7011 41.8794
2:    K2     2 per environment 11736.18 86.1939 11742.12 82.5171 41.9031
3:    K1     1 per environment 11750.71 68.4101 11755.81 66.6877 41.7889
   lambda_last_lower lambda_last_mean   dWAIC dWAIC_se
1:            2.7845         1147.040  0.0000   0.0000
2:         1025.6525         2105.971  4.5942   2.6766
3:         4291.9505         5252.469 18.2833   7.7140

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

```
plot(x = cp)
```



WAIC estimates the deviance of new plots of the same genotypes and environments (Gelman et al. 2014). `se` is the standard error of each WAIC; `dWAIC_se` is the standard error of the difference from the best model, computed from the plot-by-plot differences, and is the one to use when comparing (Vehtari et al. 2017). Here K3 has the smallest WAIC; K2 is 4.6 above it, 1.7 standard errors of the difference, and K1 is 18.3 above (2.4 standard errors). The column `lambda_last_lower` adds the posterior view: the HPD interval of the last singular value of K3 reaches down to 2.8, so a term whose interval nearly reaches zero improves prediction little and is weakly supported. Prefer the simplest model among those that the standard errors do not separate.

7.2 Posterior predictive checks

`bam_ppc()` simulates replicated trials from the posterior and compares statistics of the data with their replicated distribution (Gelman et al. 1996). A posterior predictive p-value near 0 or 1 marks a feature of the data the model does not

reproduce.

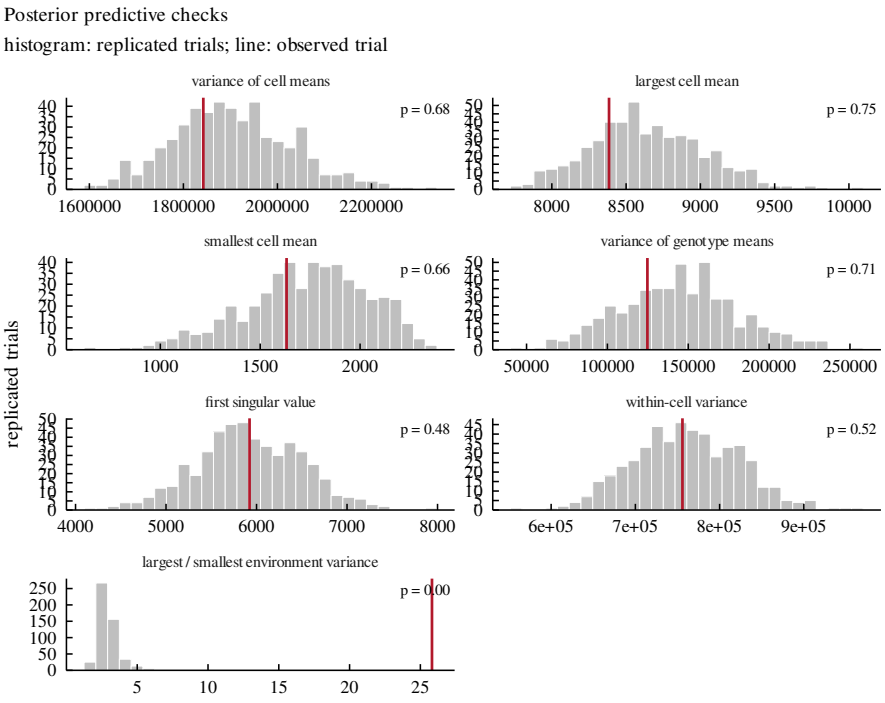
```
pp <-
  bam_ppc(
      fit    = fit
    , .draws = 500
    , .seed  = 2
    )
pp
```

```
Posterior predictive checks (500 replicated trials)

                                  statistic  observed  mean_rep p_value
1:                  variance of cell means 1.842e+06 1904569.7   0.676
2:                       largest cell mean 8.385e+03    8639.7   0.748
3:                      smallest cell mean 1.632e+03    1735.1   0.662
4:              variance of genotype means 1.247e+05  144524.3   0.710
5:                    first singular value 5.923e+03    5918.8   0.482
6:                    within-cell variance 7.559e+05  760830.0   0.522
7: largest / smallest environment variance 2.582e+01       2.9   0.000

Sources: Gelman et al. (1996). Full references: bam_cite(x).
```

```
plot(x = pp)
```



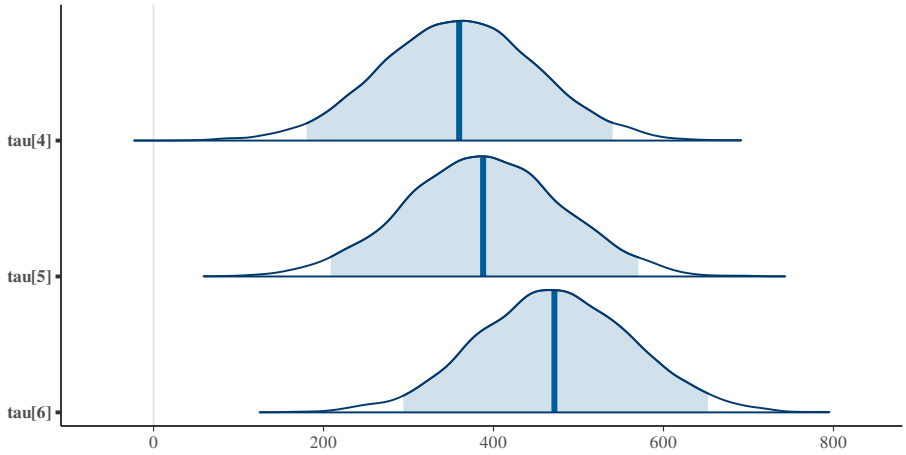
8 Working with posterior, bayesplot and coda

The draws convert to the formats of **posterior** and **coda**, so the plots of **bayesplot** apply:

```
dr <- posterior::as_draws_df(x = fit)
posterior::summarise_draws(.x = posterior::subset_draws(x = dr, variable
 ↪  = c("lambda[1]", "lambda[2]", "sigma2")))
```

```
# A tibble: 3 x 10
  variable     mean  median     sd    mad      q5    q95  rhat ess_bulk ess_tail
  <chr>       <dbl>   <dbl>  <dbl>  <dbl>   <dbl>  <dbl> <dbl>    <dbl>    <dbl>
1 lambda[1]   5480.   5483.   456.   449.   4730. 6.22e3  1.00    6599.    7072.
2 lambda[2]   1657.   1731.   775.   822.    259. 2.83e3  1.00     935.    2044.
3 sigma2    762721. 761141. 43834. 43572. 693279. 8.36e5  1.00    4684.    6929.
```

```
bayesplot::mcmc_areas(x = fit$draws, pars = c("tau[4]", "tau[5]",
 ↪  "tau[6]"), prob = 0.95)
```



9 Interpretation

```
bam_interpret(
    x             = sp
  , higher_better = TRUE
  , top           = 5
  )
```

```
Probabilities of superior performance and stability

Selecting 2 genotype(s): the highest joint probability of superior
performance and stability is that of 6 (0.13).

Findings:
- Probability of superior performance, largest first: 6 0.89, 5 0.63, 4 0.47;
  under 0.005: 9, 3, 1, 2, 7, 8.
- Probability of superior stability, largest first: 7 0.72, 3 0.72, 2 0.24, 6
  0.15, 5 0.10, 9 0.06, 1 0.02; under 0.005: 8, 4.
- Most likely superior in performance: 6; in stability: 7.

Recommendations:
- For a wide recommendation consider 6 and 5 (joint probability of at least
  0.05); the probabilities by environment (component conditional) show where
  each genotype is superior.

Cautions:
- The joint probability multiplies the two probabilities as in Dias et al.
  (2022), which treats them as independent.
```

```
Sources:
- Dias, K. O. G., dos Santos, J. P. R., Krause, M. D., Piepho, H.-P.,
  Guimaraes, L. J. M., Pastina, M. M. and Garcia, A. A. F. (2022). Leveraging
  probability concepts for cultivar recommendation in multi-environment
  trials. Theoretical and Applied Genetics. doi:10.1007/s00122-022-04041-y
```

```
bam_interpret(
    x             = st
  , higher_better = TRUE
  , top           = 5
  )
```

```
Bayesian stability

Best yield stability index (expected ranks of mean and ASV): 6, 5, 3, 7 and
9; most probably the most stable: 7 (0.56).

Findings:
- ASV with HPD intervals: 6 214 (99.8 to 329); 5 226 (94.4 to 339); 3 130
  (8.34 to 231); 7 103 (3.48 to 205); 9 253 (133 to 367); 4 406 (297 to 512);
  2 197 (77.4 to 314); 1 303 (179 to 421); 8 375 (245 to 490).

Recommendations:
- Consider 6, 5 and 3: they combine a good mean with a small AMMI stability
  value.

Cautions:
- The ASV uses the first two terms; with one term it is the absolute first
  score.

Sources:
- Purchase, J. L., Hatting, H. and van Deventer, C. S. (2000). Genotype x
  environment interaction of winter wheat (Triticum aestivum L.) in South
  Africa: II. Stability analysis of yield performance. South African Journal
  of Plant and Soil, 17(3), 101-107.
- Shukla, G. K. (1972). Some statistical aspects of partitioning
  genotype-environmental components of variability. Heredity, 29(2), 237-245.
  doi:10.1038/hdy.1972.87
- Farshadfar, E. (2008). Incorporation of AMMI stability value and grain
  yield in a single non-parametric index (GSI) in bread wheat. Pakistan
  Journal of Biological Sciences, 11(14), 1791-1796.
```

10 References

Dias, K. O. G., dos Santos, J. P. R., Krause, M. D., Piepho, H.-P., Guimaraes, L. J. M., Pastina, M. M. and Garcia, A. A. F. (2022). Leveraging probability

concepts for cultivar recommendation in multi-environment trials. *Theoretical and Applied Genetics*. doi:10.1007/s00122-022-04041-y

Shukla, G. K. (1972). Some statistical aspects of partitioning genotype-environmental components of variability. *Heredity*, 29(2), 237–245.

Purchase, J. L., Hatting, H. and van Deventer, C. S. (2000). Genotype x environment interaction of winter wheat in South Africa: II. Stability analysis of yield performance. *South African Journal of Plant and Soil*, 17(3), 101–107.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B*, 64(4), 583–639.

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.

Vehtari, A., Gelman, A. and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5), 1413–1432.

Gelman, A., Meng, X.-L. and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6(4), 733–807.

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M. and Gelman, A. (2019). Visualization in Bayesian workflow. *Journal of the Royal Statistical Society A*, 182(2), 389–402.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B. and Burkner, P.-C. (2021). Rank-normalization, folding, and localization: an improved R-hat for assessing convergence of MCMC. *Bayesian Analysis*, 16(2), 667–718.

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