Glossary

See [BAYES] **Glossary** for definitions standard to Bayesian analysis.

always predictor. Predictor that is always included in a model, that is, included in every model in the model space. In `bmaregress`, it is specified with the command within the always group of variables:

```
. bmaregress (varlist, always) ...
```

analytical mean model size. See *mean model size*.

analytical model-size distribution. See *model-size distribution*.

analytical PIP. See *posterior inclusion probability*.

analytical PMP. See *posterior model probability (PMP)*.

Bayesian model averaging (BMA). A special case of model averaging, where the weights correspond to posterior model probabilities. BMA considers a space of plausible models and views a model as a discrete random variable over this space. It uses the Bayes theorem to estimate the posterior model probability based on the model prior and observed data. The posterior distribution of model parameters is then estimated by an average over the posterior distributions conditional on a model weighted by the corresponding posterior model probability.

Bayesian model averaging (BMA) regression. BMA applied to regression analysis, where each model from the model space corresponds to a unique set of predictors. Model uncertainty in the context of Bayesian model averaging regression amounts to the uncertainty of the inclusion of predictors in a model. See [BMA] **bmaregress**.

bivariate inclusion probability. See *joint inclusion probability*.

bivariate jointness. See *jointness*.

BMA. See *Bayesian model averaging (BMA)*.

coefficient sample. See *model parameter sample*.

complements. See *jointness*.

CPMP. See *cumulative posterior model probability (CPMP)*.

cumulative posterior model probability (CPMP). The running sum of the highest to lowest PMP of models. See [BMA] **bmastats models**.

disjointness. A tendency of the two predictors to be included in a model exclusively; that is, if one is included, the other is not. Such predictors are known as *substitutes*, which means that each of them carries the same amount of information about the outcome. There are several measures proposed to estimate disjointness, see [BMA] **bmastats jointness**. Also see *jointness*.

entropy. The negative of the mean of the log density. For a normal distribution with mean μ and variance σ^2, it is defined as $\{1 + \log(2\pi\sigma^2)\}/2$. It is used for comparison with the mean log predictive-score (LPS) when checking model fit using out-of-sample observations. When the variance is known (see `bmastats lps`'s `sigma2()` option), the comparison can also be made for in-sample observations. The closer the mean LPS to the entropy, the better the model fit. See [BMA] **bmastats lps**.

enumeration. See *model enumeration*.

explored model space. See *visited model space*.

fixed g, fixed-g case. A case when a fixed value is used for the g parameter of a Zellner's g-prior. This is the case when one of the following is specified in the `gprior()` option of `bmaregress`: `bench` (the default), `uip`, `ric`, `sqrtn`, `fixed`, or `ebl`. Also see random g.

frequency mean model size. See *mean model size*.

frequency model-size distribution. See *model-size distribution*.

frequency PIP. See *posterior inclusion probability (PIP)*.

frequency PMP. See *posterior model probability (PMP)*.

fully explored model space. See *model enumeration*.

full model space. An entire model space as defined for a Bayesian model averaging model.

g-prior, g-hyperprior. A prior for the g parameter; see *random g* and *fixed g*.

grouped predictors. See *in–out group*.

highest probability model (HPM). Model with the highest posterior model probability. Clarke (2003), among others, states that the HPM is the model closest to the data-generating model whether the latter is in the explored model space or not. See [BMA] **bmastats models**.

HPM. See *highest probability model (HPM)*.

hyperprior. A distribution for a parameter of a prior distribution. In the context of a Bayesian model averaging regression, this is a prior distribution for the g parameter; see g-prior.

important predictor. See *predictor*.

in–out group. A group of predictors that is included in or excluded from a model together during estimation. In `bmaregress`, it is specified with the command in parentheses:

```
. bmaregress (varlist) ...
```

in–out predictor, in–out variable. A predictor (variable) that is included in or excluded from a model during estimation. Also see *always predictor* and [BMA] **bmaregress**.

in–out term. An in–out predictor (*varname*) or an in–out group of predictors (*varlist*); see [BMA] **bmaregress**.

inclusion map. See *variable-inclusion map*.

inclusion probability. In the context of a Bayesian model averaging regression, a probability over the considered model space that a predictor is included in a regression model. This is known as a marginal inclusion probability. Also see *joint inclusion probability* and *posterior inclusion probability (PIP)*.

influential predictor. See *predictor*.

joint inclusion probability. In the context of a Bayesian model averaging regression, a joint probability of inclusion of multiple predictors in a model. This probability is defined over the considered model space. For example, for a pair of predictors X_1 and X_2 and a model M, a bivariate inclusion probability is defined as $P(X_1 \in M, X_2 \in M)$. If the inclusion of X_1 is independent of X_2, then the bivariate joint probability is simply the product of the corresponding marginal inclusion probabilities, $P(X_1 \in M, X_2 \in M) = P(X_1 \in M)P(X_2 \in M)$. Similarly, one can define a joint noninclusion probability as $P(X_1 \notin M, X_2 \notin M)$. The other two joint probabilities of interest are $P(X_1 \in M, X_2 \notin M)$ and $P(X_1 \notin M, X_2 \in M)$. You can view these four probabilities as probabilities corresponding to the cross-tabulation of two random binary indicators $I(X_1 \in M)$ and $I(X_2 \in M)$, where $I(\cdot)$ is an indicator function. The joint probabilities are used to define various jointness measures. See [BMA] **bmastats jointness**.

joint inference. In the context of Bayesian model averaging, joint inference considers joint inclusion probabilities and, more specifically, various jointness measures. Also see *marginal inference*.

joint noninclusion probability. See *joint inclusion probability*.

jointness. A tendency of the two predictors to be included in a model together; that is, if one is included, the other is included too. Such predictors are known as *complements*, which means that, when included together, they provide additional information about the outcome. There are several measures proposed to estimate jointness; see [BMA] **bmastats jointness**. Also see *disjointness*.

jointness measures. Jointness measures assess the degree of inclusion dependency between predictors across the considered model space. They are typically defined for a pair of predictors and are based on the joint inclusion probabilities. See *jointness* and *disjointness*. Also see [BMA] **bmastats jointness**.

log predictive-score (LPS). The negative of the log of the posterior predictive density evaluated at an observation. It is often used to evaluate the model's predictive performance. The smaller the LPS, the better the predictive performance. Also see *entropy* and [BMA] **bmastats lps**.

LPS. See *log predictive-score (LPS)*.

marginal inference. In the context of Bayesian model averaging, marginal inference explores individual characteristics of predictors such as posterior inclusion probability. Also see *joint inference*.

MC3. See *Markov chain Monte Carlo model composition (MC3) algorithm*.

Markov chain Monte Carlo model composition (MC3) algorithm, **MC3 sampling**. MC3 (Madigan and York 1995) is a stochastic algorithm used in Bayesian model averaging to sample models from their posterior distribution over the model space. It explores discrete model spaces formed by subsets of potential predictors by changing one variable, or group of variables, at a time. See [BMA] **bmaregress**.

MCMC model sample. See *model sample*.

MCMC model parameter sample. See *model parameter sample*.

mean model size. The mean of the model-size distribution. In the Bayesian model averaging context, there are prior and posterior mean model sizes. The prior mean model size is the mean of the prior model-size distribution. The posterior mean model size is the mean of the posterior model-size distribution. Posterior mean model size can be computed by using analytical or frequency posterior model probability, in which case it is referred to as, respectively, the analytical or frequency posterior mean model size. The prior mean model size is always computed analytically, but in cases when the model space is not explored fully, it is conditional on the visited models. It is often compared with the posterior mean model size to evaluate the impact of the data on the prior assumption about the model size. See [BMA] **bmastats msize**.

median probability model (MPM). The MPM is a model that includes only influential predictors, predictors with posterior inclusion probability of 0.5 or above. See [BMA] **bmastats models**.

model averaging. In statistics, model averaging is an inferential technique that estimates a quantity of interest by a weighted average of individual model estimates over a space of candidate models. The weights are chosen according to various criteria with the aim of achieving certain asymptotic properties or improving prediction performance. Also see *Bayesian model averaging (BMA)*.

model distribution. In the Bayesian model averaging context, a model is viewed as a random quantity and thus has a distribution. This distribution is defined by model probabilities. There are prior and posterior model distributions. A prior model distribution is a (discrete) distribution defined over the model space and assumed for a model before observing the data. A posterior model distribution is a distribution of the model after observing the data given the assumed prior model distribution.

A prior model distribution represents an assumption about a model before observing the data. Common choices include a uniform prior on the model size and a uniform prior on the model space. The former assigns an equal probability to a model of any size. The latter assigns the same probability $1/|\mathcal{M}|$ to each model M, where $|\mathcal{M}|$ is the total number of models in the model space $\mathcal{M}$.

A posterior model distribution reflects the effect of the observed data on the prior assumption. The estimation of a posterior model distribution is at the heart of Bayesian model averaging.

model enumeration. A full enumeration of the model space. In the regression framework with p predictors, model enumeration considers all 2^p models formed by including or excluding each of the p predictors. Enumeration is feasible only with a moderate number of predictors. In `bmaregress`, model enumeration is available only when $p \leq 24$. If $p \leq 12$, `bmaregress` uses model enumeration automatically. For $12 < p \leq 24$, you can use the `enumeration` option to perform model enumeration. Also see *model sampling*.

model parameter sample. A Markov chain Monte Carlo sample from the posterior distribution of a model parameter. In a Bayesian model averaging regression, model parameters are the regression coefficients, the intercept, and the error variance. The posterior samples for these parameters are generated by the `bmacoefsample` command; see [BMA] **bmacoefsample**.

model posterior. See *posterior model distribution*.

model prior. See *prior model distribution*.

model probability. In the context of Bayesian model averaging, a model M is viewed as a discrete random variable defined on a model space $\mathcal{M}$ with a probability $0 \leq P(M) \leq 1$, such that $\sum_{M^* \in \mathcal{M}} P(M^*) = 1$. Model probabilities $P(M)$'s define a model distribution. Prior model probabilities $P(M)$'s define a prior model distribution. Posterior model probabilities $P(M|\mathbf{y})$'s, model probabilities given the observed outcome $\mathbf{y}$, define a posterior model distribution; also see *posterior model probability (PMP)*.

model sample. A Markov chain Monte Carlo sample from a posterior model distribution, estimated by the MC3 algorithm. Also see *posterior model probability (PMP)*.

model sampling. Simulation of a Markov chain Monte Carlo model sample. See [BMA] **bmaregress**.

model size. The number of predictors in the model, typically ignoring the constant term. For a model M, the model size is commonly denoted as $|M|$. See [BMA] **bmastats msize** and [BMA] **bmagraph msize**.

model space. A set of models considered for model averaging. In a regression setting, the model space includes 2^p distinct models, which correspond to all possible combinations of inclusions and exclusions of p predictors. In the presence of always predictors p_a, the model space contains 2^{p-p_a} models. In the presence of groups of predictors, an entire group is considered as one predictor in the definition of the model space. Depending on the context, the model space can sometimes imply the visited model space.

model uncertainty. Model uncertainty can be defined in many ways. In a regression setting, model uncertainty is uncertainty associated with the inclusion of predictors in a regression model. With p predictors, there are 2^p possible models. In this case, model uncertainty is defined with respect to these 2^p models.

model-size distribution. In the Bayesian model averaging context, a model is viewed as a random quantity. Its model size is random too and thus has a distribution. There are prior and posterior model-size distributions. The prior model-size distribution represents an assumption about the model size before observing the data. The posterior model-size distribution reflects the effect of the observed data on this prior assumption. It is useful to explore these distributions to evaluate the presumed and observed complexity of a Bayesian model averaging model.

The posterior model-size distribution can be computed by using the analytical or frequency posterior model probability, in which case it is referred to as, respectively, the analytical or frequency posterior model-size distribution. The prior model-size distribution is always computed analytically, but when the model space is not explored fully, it is conditional on the visited models. A beta-binomial distribution with shape parameters of one is commonly used as a noninformative prior for the model size. This is the default used by `bmaregress`. See [BMA] **bmagraph msize** and [BMA] **bmastats msize**.

modified MC3. A Markov chain Monte Carlo sampling algorithm used by `bmaregress` in the case of a random g. It uses the MC3 algorithm to sample from the model space and an adaptive Metropolis–Hastings algorithm to sample the g parameter. See *Methods and formulas* of [BMA] **bmaregress**.

MPM. See *median probability model (MPM)*.

noninclusion probability. The complementary probability of the inclusion probability.

PIP. See *posterior inclusion probability (PIP)*.

PMP. See *posterior model probability (PMP)*.

posterior coefficient sample. See *model parameter sample*.

posterior distribution. In a Bayesian model averaging regression, there is a posterior distribution of models and parameter g, which is estimated or simulated by `bmaregress`. There is also a posterior distribution of model parameters, which is simulated by `bmacoefsample` after `bmaregress` and defined as a mixture of the conditional posterior distributions given a model weighted by the posterior model probability.

posterior inclusion probability (PIP). PIP is the probability that a predictor is included in a model computed over the model space given the observed data and a prior model probability. PIP is used as a measure of the predictor's importance. Often, predictors with PIP of 0.5 or above are considered important predictors. Analytical and frequency PIPs are computed based on the respective analytical or frequency posterior model probabilities. See [BMA] **bmaregress** and [BMA] **bmastats pip**.

posterior mean model size. See *mean model size*.

posterior model distribution. See *model distribution*.

posterior model parameter sample. See *model parameter sample*.

posterior model probability (PMP). PMP is a model probability after observing the data with respect to the considered model space and given the assumed prior model probability. Consider a model M from a model space $\mathcal{M}$ and an observed outcome $\mathbf{y}$. Let $P(M)$ be a prior model probability, $f(\mathbf{y}|M)$ be the density of $\mathbf{y}$ given M, and $f(\mathbf{y})$ be the marginal density of $\mathbf{y}$. Then, using the Bayes formula, PMP is defined as

$$P(M|\mathbf{y}) = \frac{f(\mathbf{y}|M)P(M)}{f(\mathbf{y})}$$

Analytical PMP is computed by using the analytical expressions, which are available only with a fixed g. Frequency PMP is computed from a Markov chain Monte Carlo model sample, which is available with a random g or with a fixed g when `bmaregress`'s `sampling` option is specified or implied. See [BMA] **bmastats models** and [BMA] **bmagraph pmp**.

posterior model sample. See *model sample*.

posterior model-size distribution. See *model-size distribution*.

posterior noninclusion probability, **1 - PIP**. The complementary probability of the posterior inclusion probability. This is the probability mass at zero in a mixture posterior distribution of a regression coefficient. See [BMA] **bmagraph coefdensity**.

predictor, **predictor variable**. A variable used to predict an outcome or included in a model for the outcome. In the Stata context, this can be an existing variable in the dataset, or it can be a virtual variable, as described in [U] **11.4.3 Factor variables**, corresponding to a level of a factor variable or to an interaction term. In a regression context, predictor refers to any term in the specification of a regression function. A predictor with a high posterior inclusion probability (PIP), typically 0.5 or above, is considered an important predictor. A predictor with a lower PIP, typically less than 0.5, is considered a weak predictor. By default, `bmaregress` does not report predictors with PIP less than 0.01.

predictor-inclusion map. See *variable-inclusion map*.

prior mean model size. See *mean model size*.

prior model distribution. See *model distribution*.

prior model probability. See *model probability*.

prior model-size distribution. See *model-size distribution*.

random g, **random-g case**. A case when a random value is used for the g parameter of a Zellner's g-prior. This is the case when g is assumed to follow a distribution, hyperprior, and one of the following hyperpriors is specified in the `gprior()` option of `bmaregress`: `betashrink`, `betabench`, `hyperg`, `hypergn`, `zsiow`, or `robust`. Also see *fixed g*.

regression, Bayesian model averaging. See *Bayesian model averaging (BMA) regression*.

regression coefficient sample. See *model parameter sample*.

sampling correlation. A correlation between the analytical posterior model probability (PMP) and the frequency PMP. `bmaregress` reports it in the header when Markov chain Monte Carlo sampling is used. The sampling correlation is used to assess Markov chain Monte Carlo convergence. The closer it is to unity, the better. With a random g, when the analytical PMP is not available, the sampling correlation is computed as a correlation between the harmonic-mean estimator of the analytical PMP and the frequency PMP. See *Convergence of BMA* in *Remarks and examples* of [BMA] **bmaregress**.

shrinkage, **shrinkage factor**, **shrinkage parameter**. Shrinkage is defined as $g/(g+1)$, where g is the parameter of a Zellner's g-prior. The smaller this value, the more the coefficients are shrunk toward zero.

strong predictor. See *predictor*.

substitutes. See *disjointness*.

variable-inclusion summary, **predictor-inclusion summary**. A summary of the inclusion or exclusion of a predictor in a model. `bmastats models` provides this information in a tabular format with a column for each model and with a row for each predictor. Predictors that are included in the model are marked with an `x` in the corresponding row. For brevity, predictors with posterior inclusion probability less than 0.01 are not displayed. Also see *variable-inclusion map* for the graphical representation. See [BMA] **bmastats models**.

variable-inclusion map. A bar graph that plots models and their cumulative posterior model probability on the x axis and the predictors on the y axis. The bars of this graph represent coefficients. The signs of the coefficients are distinguished by color. For instance, `bmagraph varmap` uses blue for positive coefficients, red for negative coefficients, and gray for zero coefficients of predictors that were not included in a model. Also see *variable-inclusion summary* for the tabular representation. See [BMA] **bmagraph varmap**.

visited model space, **visited models**. A subset of models from a model space considered or "visited" by a Markov chain Monte Carlo sampling algorithm. With model enumeration, the visited model space is the same as the full model space.

weak predictor. See *predictor*.

Zellner's g-prior. A prior assumed for regression coefficients in a Bayesian model averaging regression. The g parameter controls the shrinkage of the coefficients toward zero. It can be fixed or random. Large g-values mean less shrinkage. In general, the larger the values of g, the more similar the Bayesian model averaging results are to the standard regression results. See [BMA] **bmaregress**.

References

Clarke, B. 2003. Comparing Bayes model averaging and stacking when model approximation error cannot be ignored. *Journal of Machine Learning Research* 4: 683–712.

Madigan, D., and J. York. 1995. Bayesian graphical models for discrete data. *Journal of Statistical Review* 63: 215–232. https://doi.org/10.2307/1403615.