Default and informative priors for 2-level designs

Víctor Peña
Department of Statistical Science
Duke University
vp58@duke.edu

1. Introduction

- **Related work:** Literature on Bayesian design and analysis of 2-level experiments (e.g. [3], [2], and [5], [4]), priors with heredity constraints [7], Bayesian model uncertainty with singular models [1].
- **Main contributions:** **(1)** Inclusion of models with aliased columns with priors that satisfy predictive matching, **(2)** incorporation of (potentially imprecise) prior information, **(3)** in fractional designs, preliminary evidence that priors without heredity constraints select median probability models with abysmal false positive rates.

2. Prior on model space

2.1 Main effects

Suppose we have F factors $x_i \in \{-1, 1\}$. The prior on the powerset of main effects is specified as follows:

1. **Number of active main effects a:** Define the odds a_{fb} that f main effects are active relative to a baseline $b \in \{0, 1, \ldots, F\}$. If $a_{fb} = 1$, the prior on a is uniform.
2. **Odds of being active wrt a baseline:** Let o_{fb} be the odds that predictor x_f is active relative to a baseline x_b. A default choice is $o_{fb} = 1$.

If the prior information is compatible, **multiple priors may satisfy the constraints**. To work around this indeterminacy, we propose finding the **maximum entropy distribution** (that satifies the prior constraints). For our proposed **default values**, the **maximum entropy distribution** is equivalent to putting a hierarchical uniform prior (as in [6]) on the powerset of main effects.

Assuming that a_{fb} and o_{ib} can be elicited **precisely** is often unreasonable. A weaker assumption is that **intervals** on $a_{fb} \in [\ell_{a_{fb}}, u_{a_{fb}}]$ and $o_{fb} \in [\ell_{o_{fb}}, u_{o_{fb}}]$ can be elicited. These bounds can be used to find **imprecise inclusion probabilities**

$$\frac{\ell_{o_{ib}} \min \mathbb{E}(a)}{\sum_{j \neq i} u_{o_{jb}} + \ell_{o_{ib}}} \leq \theta_i \leq \frac{u_{o_{ib}} \max \mathbb{E}(a)}{\sum_{j \neq i} \ell_{o_{jb}} + u_{o_{ib}}},$$

where $\min \mathbb{E}(a)$ and $\max \mathbb{E}(a)$ are the minimum and maximum values of $\mathbb{E}(a)$ over the class of priors on a that are consistent with the imprecise elicitations. For any factor x_f, we can find the **"least"** and **"most favorable"** maximum entropy priors and **find posterior inclusion probabilities under both**.

2.2 Interactions

We recommend priors that satisfy either **strong heredity** (SH) or **weak heredity** (WH). For 2-way interactions, SH implies that $x_i x_j$ can only be included if the main effects x_i and x_j are both active, whereas WH implies that $x_i x_j$ can be included if x_i or x_j is active. If the prior information on the main effects is informative, **heredity constraints will make some interactions more likely to be active than others**. We propose setting the prior on ℓ–way ($\ell \in \{2, \ldots, \ell_{\max}\}$) interactions sequentially as follows:

1. Given the number of $(\ell - 1)$-way interactions in the model, put a **prior on the number of active ℓ-way interactions**. A natural default is a uniform prior.
2. Given the number of active ℓ-way interactions, make all *allowable* ℓ–way interactions (according to heredity constraints) equally likely.

3. Prior on regression parameters

Models are indexed by an indicator $\gamma \in \{0, 1\}^p$ such that $\gamma_i = 1$ if the i-th variable is in the model and $\gamma_i = 0$ if it is not. Given the model index γ, the likelihood is

$$Y \mid \gamma, \alpha, \beta_\gamma, \sigma^2 \sim N_n(1_n \alpha + X_\gamma \beta_\gamma, \sigma^2 I_n).$$

The prior on α, σ^2 is $\pi(\alpha, \sigma^2) \propto 1/\sigma^2$ and $\beta_\gamma \mid \gamma, \alpha, \sigma^2$ is a regular conventional prior [1]:

$$\pi(\beta_\gamma \mid \gamma, \alpha, \sigma^2) = \int_0^{+\infty} N_{|\gamma|}(0_{|\gamma|}, \sigma^2 g (X_\gamma' X_\gamma)^-) \, \pi(\mathrm{d}g)$$

where $(X_\gamma' X_\gamma)^-$ is a generalized inverse of $X_\gamma' X_\gamma$.

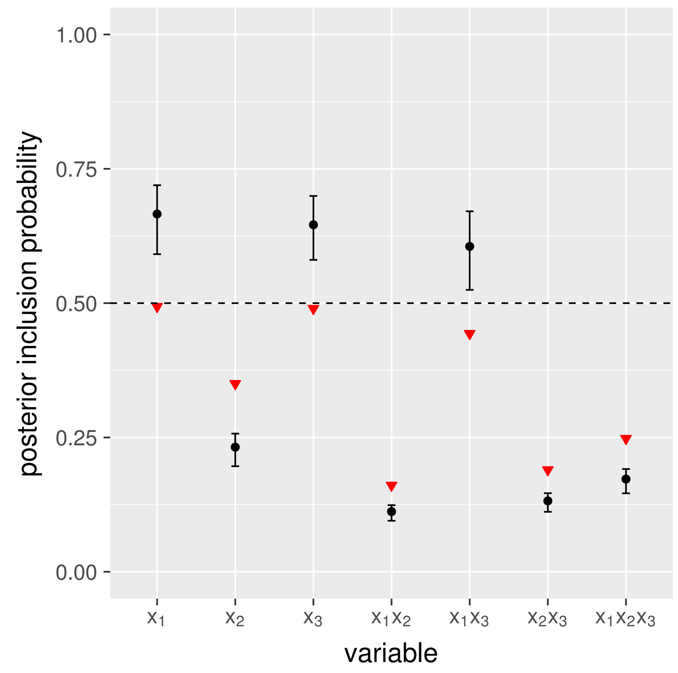
Properties:

- Marginal likelihoods don't depend of the choice of generalized inverse.
- **Predictive matching under aliasing:** If M_η is a model that nests M_γ and adds some extra variables that are **aliased** with variables (main effects or interactions) already in M_γ, then $\mathrm{BF}_{\gamma\eta} = 1$.
- **Predictive matching for saturated models:** If $\mathrm{rank}(X_\gamma' X_\gamma) = n - 1$, the null-based Bayes factor is $\mathrm{BF}_{\gamma 0} = 1$. This case includes **saturated** models and models with k nonaliased variables with $k > n$. [NB: This is a slight generalization of Result 4 in [1].]
- These conditions **aren't satisfied** if the generalized inverse $(X_\gamma' X_\gamma)^-$ is replaced by the **identity matrix**; that is, if the components of β_γ are independent a priori, which is commonly assumed in the literature.
- Without effect hierarchy and under either the uniform or the hierarchical uniform prior on the model space [6], the marginal posterior probabilities of aliased effects are equal. Under heredity constraints, aliased effects can have different marginal posterior probabilities.
- In **orthogonal designs**, prior information on **effect sizes** can be incorporated easily through the **prior mean**.

4. Illustrations

4.1 Informative prior

Simulated data from $y = x_1 + x_2 + x_1 x_3 + \epsilon$, where $\epsilon \sim N(0, I_8)$ and the design is a full 2^3 factorial. We compare the posterior inclusion probabilities under our default SH prior with the results with an imprecise informative prior with $0.9 \leq a_{10} \leq 1$, $1.75 \leq a_{20} \leq 2.25$, $0.9 \leq a_{30} \leq 1$ and $0.9 \leq o_{13} \leq 1.1$, $0.4 \leq o_{23} \leq 0.6$.



4.2 Heredity helps!

The design is the 2^{4-1}_{IV} fractional factorial of Section 4.1. in [4]. We consider 3 scenarios: (1) simulations when the truth is the null model $y \sim N_8(0_8, I_8)$, (2) simulations under SH $y \sim N_8(3x_1 + 2x_2 + x_{12}, I_8)$, (3) simulations under WH $y \sim N_8(3x_1 + x_{12}, I_8)$. We compare the performance of a "naive" hierarchical uniform prior on the variables (without heredity constraints) and our proposed default prior under SH. In all cases, we run $B = 1000$ simulations and compute false positive rates (FPR) and false negative rates (FNR) for the highest probability model (HPM) and the median probability model (MPM).

		null		SH		WH	
		naive	SH	naive	SH	naive	SH
HPM	FPR	0.333	**0.043**	0.150	**0.009**	0.203	**0.072**
	FNR	-	-	0.548	**0.138**	0.516	**0.404**
MPM	FPR	0.752	**0.151**	0.313	**0.016**	0.276	**0.102**
	FNR	-	-	**0.063**	0.118	0.406	**0.406**

References

[1] J. O. Berger, G. Garcia-Donato, M. A. Martinez-Beneito, and V. Peña. Bayesian variable selection in high dimensional problems without assumptions on prior model probabilities. *arXiv preprint arXiv:1607.02993*, 2016.

[2] D. R. Bingham and H. A. Chipman. Incorporating prior information in optimal design for model selection. *Technometrics*, 49(2):155–163, 2007.

[3] G. E. Box and R. D. Meyer. An analysis for unreplicated fractional factorials. *Technometrics*, 28(1):11–18, 1986.

[4] H. A. Chipman and M. S. Hamada. Using Bayesian variable selection to analyze regular resolution IV two-level fractional factorial designs. *Quality and Reliability Engineering International*, 33(3):493–502, 2017.

[5] G. Consonni and L. Deldossi. Objective Bayesian model discrimination in follow-up experimental designs. *Test*, 25(3):397–412, 2016.

[6] J. G. Scott, J. O. Berger, et al. Bayes and empirical-bayes multiplicity adjustment in the variable-selection problem. *The Annals of Statistics*, 38(5):2587–2619, 2010.

[7] D. Taylor-Rodríguez, A. Womack, and N. Bliznyuk. Bayesian variable selection on model spaces constrained by heredity conditions. *Journal of Computational and Graphical Statistics*, 25(2):515–535, 2016.