Simultaneous Inference Under the Vacuous Orientation Assumption

Ruobin Gong

Department of Statistics, Rutgers University

ISIPTA 2019, Ghent, Belgium

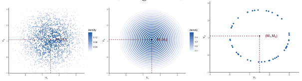
July 3, 2019

Simultaneous Inference Under the Vacuous Orientation Assumption

Ruobin Gong
Department of Statistics, Rutgers University, USA · ruobin.gong@rutgers.edu

I. Motivation

$$\mathbf{E} \sim \text{Normal}(\mathbf{0}, \mathbf{I}_k) \;=\; \underset{\text{(configuration)}}{\mathbf{E}'\mathbf{E} \sim \chi^2_k} \;+\; \underset{\text{(orientation)}}{\mathbf{E} \mid \mathbf{E}'\mathbf{E} \sim \text{isotropic}}$$



Precise simultaneous inference for k unknown quantities must rely on a known correlational structure such as error independence, i.e. $\mathbf{E} \sim \mathbf{Normal}(\mathbf{0}, \mathbf{I}_k)$. We relax this assumption by **keeping the** χ^2_k **configuration** component while **ridding the isotropic orientation** component.

II. Notation & Model

$\mathbf{Y}$ is a k-vector of observable measurements, and corresponding $\mathbf{M}$ its unknown true values. $\mathbf{E}$ is a vector of measurement errors and S^2 an associated variance parameter. Posit $\mathbb{E}$, the following body of **marginal** model evidence:

1. $\mathbf{Y} - \mathbf{M} = \mathbf{E}$: additive measurement error
2. $\mathbf{Y} = \mathbf{y}$: precisely observed measurement
3. Error configuration:
   $$\mathbf{E}'\mathbf{E} = S^2 U, \quad \text{where } U \sim \chi^2_k$$
4. Fixed error variance: $S^2 = s^2$

(4′. Random error variance: $S^2 \sim U_s$)

Auxiliary variables U and U_s are means to express evidence in stochastic form. $\mathbb{E}$ is judged to be *independent* suitable for DS-ECP (see IV). No assumption on error orientation is made.

III. Evidence Projection and Combination

Combination of evidence $\mathbb{E}$ results in a class of subsets of the full model state space

$$\mathsf{R}_{\mathbb{E}} \stackrel{\text{def}}{=\!=} \{(\mathbf{Y}, \mathbf{M}, \mathbf{E}, S^2) \in \Omega : \mathbf{Y} = \mathbf{y},\ \mathbf{Y} - \mathbf{M} = \mathbf{E},\ \mathbf{E}'\mathbf{E} = S^2 U,\ S^2 = s^2\},$$

which is a **multi-valued map** from U to subsets of Ω. Since $U \sim \chi^2_k$, $\mathsf{R}_{\mathbb{E}}$ is a random subset of Ω with distribution inherited from U. The density function of U dictates the mass function of $\mathsf{R}_{\mathbb{E}}$.

Projection of $\mathsf{R}_{\mathbb{E}}$ onto the margin of interest $\mathbf{M}$,

$$\mathsf{R}_{\mathbf{M}|\mathbb{E}} \stackrel{\text{def}}{=\!=} \{\mathbf{M} \in \Omega_{\mathbf{M}} : (\mathbf{M} - \mathbf{y})'(\mathbf{M} - \mathbf{y}) = s^2 U\}$$

where $U \sim \mu_{\mathbb{E}}$, the χ^2_k distribution. $\mathsf{R}_{\mathbf{M}|\mathbb{E}}$ is again a random subset of $\Omega_{\mathbf{M}}$ whose distribution is dictated by U. For every realization $U = u$, $\mathsf{R}_{\mathbf{M}|\mathbb{E}}(u)$ is a k-sphere centered at $\mathbf{y}$ with radius $s\sqrt{u}$. We say that $\mathsf{R}_{\mathbf{M}|\mathbb{E}}$ embodies posterior inference for $\mathbf{M}$ given evidence $\mathbb{E}$.

IV. DS-ECP

Central to Dempster-Shafer Extended Calculus of Probability (**DS-ECP**) is the processing of bodies of independent marginal evidence.

Definition 1. A body of marginal evidence $\mathbb{E}$ consisting of J pieces is said to be **independent**, if the marginal **auxiliary variables (a.v.s)** associated with each piece are all statistically independent. That is, for $U_j \sim \mu_j$, $j = 1, \cdots, J$,

$$(U_1, \cdots, U_J) \sim \mu_1 \times \cdots \times \mu_J.$$

Notably, deterministic pieces of evidence are associated with degenerate a.v.s, thus always independent of other pieces of evidence.

Dempster's Rule of Combination amounts to 1) taking the *product* of marginal a.v.s, and 2) applying *domain revision* to the joint a.v. to exclude values that result in algebraic incompatibility, i.e. empty intersections of marginal focal sets. Denote μ the prior probability of $\mathbf{U}$, the joint a.v. for $\mathbb{E}$ measurable w.r.t. $\sigma(\Xi)$. *A posteriori* $\mathbb{E}$, revise μ to $\mu_{\mathbb{E}}$ measurable w.r.t. $\sigma(\Xi_{\mathbb{E}}) \subset \sigma(\Xi)$ where $\Xi_{\mathbb{E}} = \{u \in \Xi : \mathsf{R}_{\mathbb{E}}(u) \neq \emptyset\}$, and

$$\mu_{\mathbb{E}} = (\mu \times \mathbf{1}_{\Xi_{\mathbb{E}}}) / \mu(\Xi_{\mathbb{E}}),$$

where $\mathbf{1}_A(S) = 1$ if $S \subseteq A$ and 0 otherwise. For the current model, domain revision of the a.v. is trivial, namely $\mu_{\mathbb{E}} = \mu$.

Stochastic three-valued logic. Posterior inference about assertions concerning the state space is expressed through a probability triple $(\mathsf{p}, \mathsf{q}, \mathsf{r})$, representing weights of evidence "for", "against", and "don't know" about that assertion. Set functions $\mathsf{p}, \mathsf{q}, \mathsf{r} : \Omega_{\mathbf{M}} \to [0, 1]$ are such that for all $H \in \sigma(\Omega_{\mathbf{M}})$,

$$\mathsf{p}(H) = \int_{\{u \in \Xi_{\mathbb{E}} : \mathsf{R}_{\mathbf{M}|\mathbb{E}}(u) \subseteq H\}} d\mu_{\mathbb{E}},$$

The $(\mathsf{p}, \mathsf{q}, \mathsf{r})$ representation is an alternative to a pair of belief and plausibility functions on $\Omega_{\mathbf{M}}$, where p is the belief function and $1 - \mathsf{q}$ (equivalently $\mathsf{p} + \mathsf{r}$) is its conjugate plausibility function.

V. Posterior Inference

Linear forms of hypotheses are expressed by a consistent system of equations $\mathbf{CM} = \mathbf{a}$, where $\mathbf{C}$ is a real-valued p by k matrix with arbitrary p. Define summary statistic

$$t_{\mathbf{y}} = (\mathbf{a} - \mathbf{Cy})'(\mathbf{CC}')^{-1}(\mathbf{a} - \mathbf{Cy}),$$

where in case $p > \text{rank}(\mathbf{C})$, the inverse is taken to be the Moore-Penrose pseudoinverse.

Theorem 3. Posterior probabilities concerning one-sided linear hypothesis $H : \mathbf{CM} \leq \mathbf{a}$ are

$$\{\mathsf{p}(H), \mathsf{q}(H), \mathsf{r}(H)\} = \{F(t_{\mathbf{y}}), 0, 1 - F(t_{\mathbf{y}})\}$$

if $\mathbf{Cy} \leq \mathbf{a}$, and

$$\{\mathsf{p}(H), \mathsf{q}(H), \mathsf{r}(H)\} = \{0, F(t_{\mathbf{y}}), 1 - F(t_{\mathbf{y}})\}$$

otherwise. F is the CDF of scaled χ^2_k with scaling factor s^2 (fixed error variance case).

Posterior $(1 - \alpha)$ **credible regions** of the form

$$A_\alpha = \left\{\mathbf{M} \in \Omega_{\mathbf{M}} : (\mathbf{M} - \mathbf{y})'(\mathbf{M} - \mathbf{y}) \leq F^{-1}_{1-\alpha}\right\},$$

where F^{-1}_α is the α^{th}-quantile of $\mu_{\mathbb{E}}$.

Theorem 6. A_α is a *sharp* posterior credible region in the sense that $\mathsf{r}(A_\alpha) = 0$ for all α.

Theorem 7. A_α is *calibrated* to the i.i.d. error model, P^*, in the sense that for all $\mathbf{M}^*$ and all α, $\mathsf{p}(A) = P^*(\mathbf{M}^* \in A) = 1 - \alpha$ and $\mathsf{q}(A) = P^*(\mathbf{M}^* \in A^c) = \alpha$.

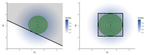


Figure 1: Focal sets that constitute $\mathsf{p}(H)$ for one-sided linear (left) and rectangular (right) hypotheses.

Rectangular regions of the form

$$C_\alpha = \left\{\mathbf{M} \in \Omega_{\mathbf{M}} : \mathbf{M} \in \otimes_{i=1}^k (y_i \pm c_\alpha \cdot s)\right\}$$

parallels Bonferroni simultaneous confidence regions. Probabilities associated with C_α are functions of the standardized half width c_α.

Example 3 *(test for all pairwise contrasts).* The simultaneous test for all pairwise means are identical has null hypothesis

$$H = \cap_{1 \leq i < j \leq k} H_{i,j}, \quad H_{i,j} : M_i = M_j.$$

The number of pairwise contrasts tested is on *quadratic* order of k, but the compound hypothesis H always spans a 1-dimensional subspace of $\Omega_{\mathbf{M}}$. As k increases, the distribution of $\mathsf{r}(H)$ (Figure 2 left) approaches uniform, which is that of a correctly calibrated p-value under the null model, whereas the Bonferroni procedure (Figure 2 right) becomes increasingly conservative for larger k. The vacuous orientation model captures the logical connection among the large number of hypotheses (collinearity), and delivers posterior inference reflective of the geometry of the hypothesis space.

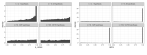


Figure 2: Distribution of $\mathsf{r}(H)$ (left) and Bonferroni p-value (right) for all pairwise contrasts under the null sampling model. For larger k, $\mathsf{r}(H)$ resembles a correctly calibrated p-value, whereas the Bonferroni p-value becomes more conservative.

VI. Future Directions

The vacuous orientation model may extend to

- Elliptical distributions;
- Multivariate and multiple regression;
- Partially vacuous orientation models based on finer variance decomposition.

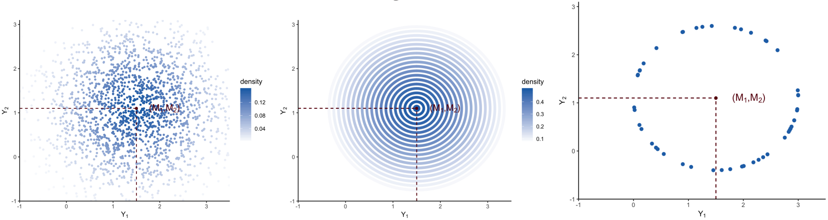
Motivation: simultaneous inference/meta analysis

- $\mathbf{M} = (M_1, \ldots, M_k)$: vector of unknown parameters
- $\mathbf{Y} = (Y_1, \ldots, Y_k)$ a vector of observable data aimed at measuring $\mathbf{M}$
- Each Y_i is a statistic from an experiment which we understand well, but we do not how they relate to one another.

Let $\mathbf{E} = \mathbf{Y} - \mathbf{M}$ denote the vector of measurement errors.

I. Motivation

$$\mathbf{E} \sim \text{Normal}(\mathbf{0}, \mathbf{I}_k) \;=\; \underset{\text{(configuration)}}{\mathbf{E}'\mathbf{E} \sim \chi^2_k} \;+\; \underset{\text{(orientation)}}{\mathbf{E} \mid \mathbf{E}'\mathbf{E} \sim \text{isotropic}}$$

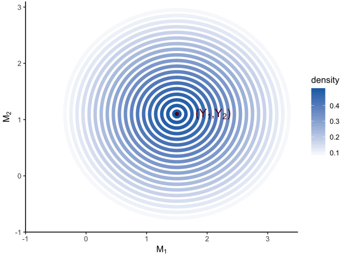


Precise simultaneous inference for k unknown quantities must rely on a known correlational structure such as error independence, i.e. $\mathbf{E} \sim \mathbf{Normal}(\mathbf{0}, \mathbf{I}_k)$. We relax this assumption by **keeping the χ^2_k configuration** component while **ridding the isotropic orientation** component.

Posterior Inference

$$\mathsf{R}_{\mathbf{M}|\mathbb{E}} \stackrel{\text{def}}{=\!=} \{\mathbf{M} \in \Omega_{\mathbf{M}} : (\mathbf{M} - \mathbf{y})' (\mathbf{M} - \mathbf{y}) = s^2 U\},$$

is a **random subset** of $\Omega_{\mathbf{M}}$ (concentric hyperspheres), whose distribution is dictated by the *auxiliary variable* $U \sim \chi^2_k$.



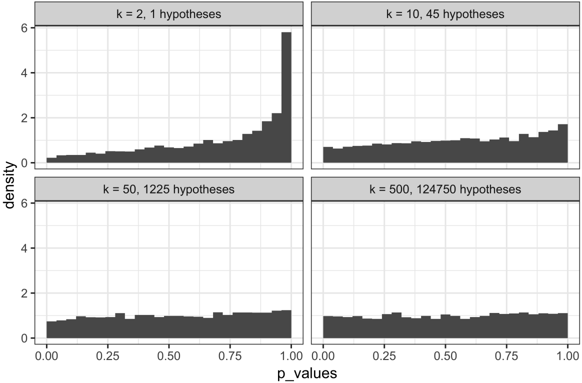
$\mathsf{R}_{\mathbf{M}|\mathbb{E}}$ embodies posterior inference for $\mathbf{M}$ given $\mathbb{E}$.

Testing many collinear hypotheses

Example 3. The simultaneous test for all pairwise means being identical:

$$H = \cap_{1 \leq i < j \leq k} H_{i,j}, \quad H_{i,j} : M_i = M_j.$$

For larger k, $\overline{P}\left(H \mid \mathbb{E}\right)$ approaches uniformity as if a well-calibrated p-value.

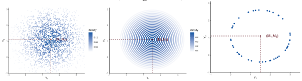


Simultaneous Inference Under the Vacuous Orientation Assumption

Ruobin Gong
Department of Statistics, Rutgers University, USA · ruobin.gong@rutgers.edu

I. Motivation

$$\mathbf{E} \sim \text{Normal}(\mathbf{0}, \mathbf{I}_k) \;=\; \underset{\text{(configuration)}}{\mathbf{E}'\mathbf{E} \sim \chi^2_k} \;+\; \underset{\text{(orientation)}}{\mathbf{E} \mid \mathbf{E}'\mathbf{E} \sim \text{isotropic}}$$



Precise simultaneous inference for k unknown quantities must rely on a known correlational structure such as error independence, i.e. $\mathbf{E} \sim \textbf{Normal}(\mathbf{0}, \mathbf{I}_k)$. We relax this assumption by **keeping the χ^2_k configuration** component while **ridding the isotropic orientation** component.

II. Notation & Model

$\mathbf{Y}$ is a k-vector of observable measurements, and corresponding $\mathbf{M}$ its unknown true values. $\mathbf{E}$ is a vector of measurement errors and S^2 an associated variance parameter. Posit $\mathbb{E}$, the following body of **marginal** model evidence:

1. $\mathbf{Y} - \mathbf{M} = \mathbf{E}$: additive measurement error
2. $\mathbf{Y} = \mathbf{y}$: precisely observed measurement
3. Error configuration:
$$\mathbf{E}'\mathbf{E} = S^2 U, \quad \text{where } U \sim \chi^2_k$$
4. Fixed error variance: $S^2 = s^2$

(4′. Random error variance: $S^2 \sim U_s$)

Auxiliary variables U and U_s are means to express evidence in stochastic form. $\mathbb{E}$ is judged to be *independent* suitable for DS-ECP (see IV). No assumption on error orientation is made.

III. Evidence Projection and Combination

Combination of evidence $\mathbb{E}$ results in a class of subsets of the full model state space

$$\mathsf{R}_{\mathbb{E}} \stackrel{\text{def}}{=\!=} \{(\mathbf{Y}, \mathbf{M}, \mathbf{E}, S^2) \in \Omega : \mathbf{Y} = \mathbf{y},\, \mathbf{Y} - \mathbf{M} = \mathbf{E},\, \mathbf{E}'\mathbf{E} = S^2 U,\, S^2 = s^2\},$$

which is a **multi-valued map** from U to subsets of Ω. Since $U \sim \chi^2_k$, $\mathsf{R}_{\mathbb{E}}$ is a random subset of Ω with distribution inherited from U. The density function of U dictates the mass function of $\mathsf{R}_{\mathbb{E}}$.

Projection of $\mathsf{R}_{\mathbb{E}}$ onto the margin of interest $\mathbf{M}$,

$$\mathsf{R}_{\mathbf{M}|\mathbb{E}} \stackrel{\text{def}}{=\!=} \{\mathbf{M} \in \Omega_{\mathbf{M}} : (\mathbf{M} - \mathbf{y})' (\mathbf{M} - \mathbf{y}) = s^2 U\}$$

where $U \sim \mu_{\mathbb{E}}$, the χ^2_k distribution. $\mathsf{R}_{\mathbf{M}|\mathbb{E}}$ is again a random subset of $\Omega_{\mathbf{M}}$ whose distribution is dictated by U. For every realization $U = u$, $\mathsf{R}_{\mathbf{M}|\mathbb{E}}(u)$ is a k-sphere centered at $\mathbf{y}$ with radius $s\sqrt{u}$. We say that $\mathsf{R}_{\mathbf{M}|\mathbb{E}}$ embodies posterior inference for $\mathbf{M}$ given evidence $\mathbb{E}$.

V. Posterior Inference

Linear forms of hypotheses are expressed by a consistent system of equations $\mathbf{CM} = \mathbf{a}$, where $\mathbf{C}$ is a real-valued p by k matrix with arbitrary p. Define summary statistic

$$t_{\mathbf{y}} = (\mathbf{a} - \mathbf{Cy})' (\mathbf{CC}')^{-1} (\mathbf{a} - \mathbf{Cy}),$$

where in case $p > \text{rank}(\mathbf{C})$, the inverse is taken to be the Moore-Penrose pseudoinverse.

Theorem 3. Posterior probabilities concerning one-sided linear hypothesis $H : \mathbf{CM} \leq \mathbf{a}$ are

$$\{\mathsf{p}(H), \mathsf{q}(H), \mathsf{r}(H)\} = \{F(t_{\mathbf{y}}), 0, 1 - F(t_{\mathbf{y}})\}$$

if $\mathbf{Cy} \leq \mathbf{a}$, and

$$\{\mathsf{p}(H), \mathsf{q}(H), \mathsf{r}(H)\} = \{0, F(t_{\mathbf{y}}), 1 - F(t_{\mathbf{y}})\}$$

otherwise. F is the CDF of scaled χ^2_k with scaling factor s^2 (fixed error variance case).

Posterior $(1 - \alpha)$ **credible regions** of the form

$$A_\alpha = \left\{\mathbf{M} \in \Omega_{\mathbf{M}} : (\mathbf{M} - \mathbf{y})' (\mathbf{M} - \mathbf{y}) \leq F^{-1}_{1-\alpha}\right\},$$

where F^{-1}_α is the α^{th}-quantile of $\mu_{\mathbb{E}}$.

Theorem 6. A_α is a *sharp* posterior credible region in the sense that $\mathsf{r}(A_\alpha) = 0$ for all α.

Theorem 7. A_α is *calibrated* to the i.i.d. error model, P^*, in the sense that for all $\mathbf{M}^*$ and all α, $\mathsf{p}(A) = P^*(\mathbf{M}^* \in A) = 1 - \alpha$ and $\mathsf{q}(A) = P^*(\mathbf{M}^* \in A^c) = \alpha$.

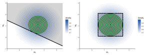


Figure 1: Focal sets that constitute $\mathsf{p}(H)$ for one-sided linear (left) and rectangular (right) hypotheses.

Rectangular regions of the form

$$C_\alpha = \left\{\mathbf{M} \in \Omega_{\mathbf{M}} : \mathbf{M} \in \otimes_{i=1}^{k} (y_i \pm c_\alpha \cdot s)\right\}$$

parallels Bonferroni simultaneous confidence regions. Probabilities associated with C_α are functions of the standardized half width c_α.

Example 3 *(test for all pairwise contrasts).* The simultaneous test for all pairwise means are identical has null hypothesis

$$H = \cap_{1 \leq i < j \leq k} H_{i,j}, \quad H_{i,j} : M_i = M_j.$$

The number of pairwise contrasts tested is on *quadratic* order of k, but the compound hypothesis H always spans a 1-dimensional subspace of $\Omega_{\mathbf{M}}$. As k increases, the distribution of $\mathsf{r}(H)$ (Figure 2 left) approaches uniform, which is that of a correctly calibrated p-value under the null model, whereas the Bonferroni procedure (Figure 2 right) becomes increasingly conservative for larger k. The vacuous orientation model captures the logical connection among the large number of hypotheses (collinearity), and delivers posterior inference reflective of the geometry of the hypothesis space.

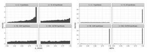


Figure 2: Distribution of $\mathsf{r}(H)$ (left) and Bonferroni p-value (right) for all pairwise contrasts under the null sampling model. For larger k, $\mathsf{r}(H)$ resembles a correctly calibrated p-value, whereas the Bonferroni p-value becomes more conservative.

IV. DS-ECP

Central to Dempster-Shafer Extended Calculus of Probability (**DS-ECP**) is the processing of bodies of independent marginal evidence.

Definition 1. A body of marginal evidence $\mathbb{E}$ consisting of J pieces is said to be **independent**, if the marginal **auxiliary variables (a.v.s)** associated with each piece are all statistically independent. That is, for $U_j \sim \mu_j$, $j = 1, \cdots, J$,

$$(U_1, \cdots, U_J) \sim \mu_1 \times \cdots \times \mu_J.$$

Notably, deterministic pieces of evidence are associated with degenerate a.v.s, thus always independent of other pieces of evidence.

Dempster's Rule of Combination amounts to 1) taking the *product* of marginal a.v.s, and 2) applying *domain revision* to the joint a.v. to exclude values that result in algebraic incompatibility, i.e. empty intersections of marginal focal sets. Denote μ the prior probability of $\mathbf{U}$, the joint a.v. for $\mathbb{E}$ measurable w.r.t. $\sigma(\Xi)$. *A posteriori* $\mathbb{E}$, revise μ to $\mu_{\mathbb{E}}$ measurable w.r.t. $\sigma(\Xi_{\mathbb{E}}) \subset \sigma(\Xi)$ where $\Xi_{\mathbb{E}} = \{u \in \Xi : \mathsf{R}_{\mathbb{E}}(u) \neq \emptyset\}$, and

$$\mu_{\mathbb{E}} = (\mu \times \mathbf{1}_{\Xi_{\mathbb{E}}}) / \mu(\Xi_{\mathbb{E}}),$$

where $\mathbf{1}_A(S) = 1$ if $S \subseteq A$ and 0 otherwise. For the current model, domain revision of the a.v. is trivial, namely $\mu_{\mathbb{E}} = \mu$.

Stochastic three-valued logic. Posterior inference about assertions concerning the state space is expressed through a probability triple $(\mathsf{p}, \mathsf{q}, \mathsf{r})$, representing weights of evidence "for", "against", and "don't know" about that assertion. Set functions $\mathsf{p}, \mathsf{q}, \mathsf{r} : \Omega_{\mathbf{M}} \to [0, 1]$ are such that for all $H \in \sigma(\Omega_{\mathbf{M}})$,

$$\mathsf{p}(H) = \int_{\{u \in \Xi_{\mathbb{E}} : \mathsf{R}_{\mathbf{M}|\mathbb{E}}(u) \subseteq H\}} d\mu_{\mathbb{E}},$$

The $(\mathsf{p}, \mathsf{q}, \mathsf{r})$ representation is an alternative to a pair of belief and plausibility functions on $\Omega_{\mathbf{M}}$, where p is the belief function and $1 - \mathsf{q}$ (equivalently $\mathsf{p}+\mathsf{r}$) is its conjugate plausibility function.

VI. Future Directions

The vacuous orientation model may extend to

- Elliptical distributions;
- Multivariate and multiple regression;
- Partially vacuous orientation models based on finer variance decomposition.