Posterior Probabilities of Hypotheses

Kristyn Pantoja

9/10/2019

Examining Posterior Probability of Hypotheses Calculation

- To find out why the MEDs did not outperform the space-filling design.
- Ideas:
  - Is the calculation for estimating posterior probabilities of hypotheses incorrect?
    - is the evidence, i.e. density of $y|H_\ell$, i.e. $f(y|H_\ell, X)$, incorrectly calculated?
    - are the simulations of $y|H_r$ incorrect?
    - ...
  - Or something else.
- Ways to find out:
  - dissect the code
  - checking that $E[P(H_0|y,X)|X,H_0] = E[P(H_1|y,X)|X,H_1]$. Why is this true? Is it still true if V_0 does not have the same dimensions as V_1, as in our case?

First, Review Posterior Probabilities of Hypotheses

Posterior Probabilities of Hypotheses

- Posterior Probability of model $H_\ell, \ell \in \{0, 1\}$:

$$P(H_\ell|y, X) = \frac{\pi_\ell f(y|H_\ell, X)}{\pi_0 f(y|H_0, X) + \pi_1 f(y|H_1, X)}$$
$$= \frac{f(y|H_\ell, X)}{f(y|H_0, X) + f(y|H_1, X)}$$

  Since we assume $\pi_0, \pi_1 = \frac{1}{2}$.
- $f(y|H_\ell, X) = \phi_{\ell,X}(y)$ is the model evidence, i.e. density of $N_N(X\mu_\ell, \sigma_\varepsilon^2 I + XV_\ell X^T)$ evaluated at a given y and design **D** with N design points.
- $E[P(H_\ell|y, X)|H_r, X]$ may be estimated using MC approximation from simulated responses y under a chosen hypothesis H_r.

Estimate Expected Posterior Probability of a Hypothesis

Estimate the expected posterior probability of hypothesis H_ℓ for J simulations of Y under H_r, given design $\mathbf{D} = \{x_1, ..., x_N\}$:

1. For $j = 1, \ldots, J$:
   1.1 Draw $\beta \sim N(\mu_r, V_r^2)$
   1.2 Draw $y_i^{(j)}|\mathbf{x}_i \sim N(\mathbf{x}_i^T\beta, \sigma_\varepsilon^2)$, $\forall \mathbf{x}_i \in \mathbf{D}$, so $y^{(j)} \in R^N$.
   1.3 $\forall m = \{0, 1\}$, calculate model evidences $f(y|H_m, \mathbf{D})$
   1.4 Calculate the posterior probability of H_ℓ, $P(H_\ell|y^{(j)}, \mathbf{D})$, from simulation j

$$P(H_\ell|y^{(j)}, \mathbf{D}) = \frac{f(y^{(j)}|H_\ell, X)}{f(y^{(j)}|H_0, X) + f(y^{(j)}|H_1, X)}$$

2. Average the estimated posterior probabilities of H_ℓ over $\forall j$ to obtain MC estimate of $E[P(H_\ell|y, \mathbf{D})|H_r, \mathbf{D}]$

Note that $y^{(j)}$ are generated from $N_N(X\beta, \sigma_\varepsilon^2 I)$ and are independent, while the model evidence for H_m marginalizes out β and evaluates $y^{(j)}$ using $f(y|H_m, \mathbf{D})$, the density of $N_N(X\mu_m, \sigma_\varepsilon^2 I + XV_mX^T)$, in which they are no longer assumed to be independent.

An additional design

A center design with all points at (0.5, 0.5)

```
center_design = matrix(rep(0.5, N*2), N, 2)
```

- This design should have the greatest ability to distinguish the hypotheses, and hence should have the highest expected posterior probabilities for the true hypotheses.

Expected Posterior Probabilities

- Evaluating ability to distinguish hypotheses:

	1atT,k=1	1atT,k=4	1atT,k=50	Fast,S=5
E[P(H0|Y,D)|H0,D]	0.816	0.795	0.807	0.844
E[P(H1|Y,D)|H0,D]	0.184	0.205	0.193	0.156
E[BF01 | H0,D]	46.5	63.8	67.2	99.6
E[P(H0|Y,D)|H1,D]	0.191	0.215	0.129	0.158
E[P(H1|Y,D)|H1,D]	0.809	0.785	0.871	0.842
E[BF01|H1,D]	0.722	0.849	0.377	0.477

	Fast,S=5	D-Opt	Space	Random	Center
E[P(H0|Y,D)|H0,D]	0.844	0.541	0.927	0.844	0.633
E[P(H1|Y,D)|H0,D]	0.156	0.459	0.0733	0.156	0.367
E[BF01 | H0,D]	99.6	1.3	1370	1330	3.02
E[P(H0|Y,D)|H1,D]	0.158	0.453	0.0527	0.123	0.369
E[P(H1|Y,D)|H1,D]	0.842	0.547	0.947	0.877	0.631
E[BF01|H1,D]	0.477	1	0.106	0.635	1.07

Why?

Advice: "Check why the Bayesian probability is not higher for our methods than for the space filling design e.g. check the mean density of the points in the space filling design and our designs (under H_1). If the points in our design have higher density then I think there must be a mistake in the computation of the posterior probability. If the mean density is not higher in our design then there is something wrong with the criterion we are optimizing."

- The density of a given responses y from design $\mathbf{D}$ with corresponding design matrix X under H_1 is the pdf of $N(X\mu_1, \sigma_\varepsilon^2 I + XV_1X^T)$
- The mean densities are averaged over the simulated responses.

Model Evidences

a.k.a. marginal likelihoods a.k.a. densities is calculated as:

```
model_evidence = function(Y, D, N, mean_beta, var_mean,
                          var_e, type){
  # Y is vector of responses
  # X is design matrix
  X = constructDesignX(D, N, type)
  marginaly_mean = X %*% mean_beta
  if(dim(X)[1] > 1){
    marginaly_var = diag(rep(var_e, N)) +
        (X %*% var_mean %*% t(X))
  } else{
    marginaly_var = var_e +
        (X %*% var_mean %*% t(X))
  }
  return(dmvnorm(Y, mean = marginaly_mean,
                 sigma = marginaly_var, log = FALSE))
#}
```

Mean Densities / Evidences for H1 for Y | H1

The density $N(X\mu_1, \sigma_\varepsilon^2 I + XV_1X^T)$ evaluated for each simulated $y^{(j)}$, then averaged over all `numSims` simulations, for each design **D**. (note: scaled by `10e-55`.)

	E[f(Y|H1)|H1]
k1	1665.0220
k4	3377.6071
k50	2425.4645
fast	1021.6549
dopt	574.3946
space	893.9396
random	1154.0077
center	19655.2234

Mean Densities / Evidences for H0, too

(note: scaled by `10e-55`.)

	E[f(Y|H0)|H1]	E[f(Y|H1)|H1]
k1	335.035	1665.022
k4	1372.983	3377.607
k50	153.922	2425.465
fast	8135.418	1021.655
dopt	429.288	574.395
space	0.675	893.940
random	130.429	1154.008
center	17239.127	19655.223

- it seems that there might be some outliers affecting these values, since the fast MED's mean model evidence favors the wrong hypothesis.
- it also doesn't make sense why the center design has such high evidence for H_0 relative to H_1.

What about median?

(scaled by `10e-49`.)

	median(f(Y|H0)|H1)	median(f(Y|H1)|H1)
k1	272.914	2531.599
k4	203.825	3381.728
k50	197.017	3271.117
fast	121.587	2722.603
dopt	452.629	687.078
space	0.648	1245.369
random	11.725	1944.884
center	13226.410	31735.912

Revisiting Expected Posterior Probabilities (y simulated from H1)

Recall,

$$
\begin{aligned}
E[P(H_\ell|Y,D)|H_1,D] &\approx \frac{1}{J}\sum_{j=1}^{N} \frac{f(y^{(j)}|D,H_\ell)\pi_\ell}{f(y^{(j)}|D,H_0)\pi_0 + f(y^{(j)}|D,H_1)\pi_1} \\
&= \frac{1}{J}\sum_{j=1}^{N} \frac{f(y^{(j)}|D,H_\ell)}{f(y^{(j)}|D,H_0) + f(y^{(j)}|D,H_1)} \text{ if } \pi_0 = \pi_1
\end{aligned}
$$

- where $y^{(j)}$ is the jth simulation of responses y under H_1 in this case, and
- $f(y^{(j)}|D,H_\ell)$ is the N-dimensional multivariate normal density of the marginal distribution of $y|H_1$ here, i.e. pdf of $N(X\mu_1, \sigma_\varepsilon^2 I + XV_1X^T)$

Compare Expected (Mean) Posterior Probabilities of H0, H1 (y simulated from H1)

	E[P(H0|Y,D)|H1,D]	E[P(H1|Y,D)|H1,D]
k1	0.2120588	0.7879412
k4	0.1852580	0.8147420
k50	0.1853418	0.8146582
fast	0.1638508	0.8361492
dopt	0.4451813	0.5548187
space	0.0570718	0.9429282
random	0.1140949	0.8859051
center	0.3831320	0.6168680

- why isn't the center design doing best?

Standard deviations of posterior probabilities

to see how much the simulated results vary

	sd[P(H0|Y,D)|H1,D]	sd[P(H1|Y,D)|H1,D]
k1	0.2548127	0.2548127
k4	0.2387937	0.2387937
k50	0.2450811	0.2450811
fast	0.2500627	0.2500627
dopt	0.1789188	0.1789188
space	0.1625855	0.1625855
random	0.2170421	0.2170421
center	0.2288063	0.2288063

Standard errors of posterior probabilities

to evaluate the estimates

	se[P(H0|Y,D)|H1,D]	se[P(H1|Y,D)|H1,D]
k1	0.0113956	0.0113956
k4	0.0106792	0.0106792
k50	0.0109604	0.0109604
fast	0.0111831	0.0111831
dopt	0.0080015	0.0080015
space	0.0072710	0.0072710
random	0.0097064	0.0097064
center	0.0102325	0.0102325

Median Posterior Probabilities

	med[P(H0|Y,D)|H1,D]	med[P(H1|Y,D)|H1,D]
k1	0.0997931	0.9002069
k4	0.0671411	0.9328589
k50	0.0624812	0.9375188
fast	0.0400389	0.9599611
dopt	0.4868862	0.5131138
space	0.0012915	0.9987085
random	0.0110149	0.9889851
center	0.3533053	0.6466947

Min Posterior Probabilities

	min[P(H0|Y,D)|H1,D]	min[P(H1|Y,D)|H1,D]
k1	0.0000022	0.0206296
k4	0.0000011	0.0461397
k50	0.0000286	0.0091967
fast	0.0000001	0.0144091
dopt	0.0015599	0.3376954
space	0.0000000	0.0037366
random	0.0000000	0.0065922
center	0.0112449	0.0656253

Max Posterior Probabilities

	max[P(H0|Y,D)|H1,D]	max[P(H1|Y,D)|H1,D]
k1	0.9793704	0.9999978
k4	0.9538603	0.9999989
k50	0.9908033	0.9999714
fast	0.9855909	0.9999999
dopt	0.6623046	0.9984401
space	0.9962634	1.0000000
random	0.9934078	1.0000000
center	0.9343747	0.9887551

Looking at XVX'

Trace as measure of size of matrix: an idea of how influential the XVX^T term is in variance of marginal y

	tr(XVX')|H0	tr(XVX')|H1
k1	0.7935644	0.9268491
k4	0.7919814	0.9207008
k50	0.7931731	0.9237086
fast	0.8012551	0.9460766
dopt	1.0000000	1.5000000
space	0.8518519	1.0855510
random	0.8399868	1.0354603
center	0.7500000	0.8125000

- The order of size of XVX^T for the designs is the same for H_0 and H_1.
- It makes sense that the D-optimal design has the largest traces for H_0 and H_1, since it maximizes X^TX, and $Tr(AA^T) = Tr(A^TA)$ (and a spectral decomposition would show that V_ℓ just scales this value I think)
- Larger XVX^T means flatter marginal density means less evidence? Or could mean higher density, compared to the other more peaked marginal distribution, if the data is not close to the mean (even though it should be).
- **Does the posterior probability somehow consider that data in the same location can only update the prior to a certain extent (like we suspect in the MSE)?**
  - would changing V_0, V_1 prior variance on β solve this problem somehow? Or is it affecting each $XV_\ell X^T$ equally as much?
  - probably X is the largest influencer.

Looking at Designs with a Single Point

What about for single data points?

- Consider the singleton designs, $x_0 = (0, 0)$ and $x_1 = (0.5, 0.5)$ as well as $x_2 = (1, 1)$ (just to make sure its results are similar to x_0)and $x_3 = (0.25, 0.25)$

	x0	x1	x2	x3
E[f(y|H0)|H1]	2.779552	2.048166	1.8356224	2.4007685
E[f(y|H1)|H1]	2.779552	2.482934	1.6691286	2.7114773
XVX'|H0	0.005000	0.007500	0.0150000	0.0056250
XVX'|H1	0.005000	0.008125	0.0250000	0.0056641
E[P(H0|Y,D)|H1,D]	0.500000	0.415153	0.4966713	0.4419608
E[P(H1|Y,D)|H1,D]	0.500000	0.584847	0.5033287	0.5580392

Looking at Designs with Two Points

Two points?

- Consider the singleton designs, $D_0 = \{(0,0),(0,0)\}$, $D_1 = \{(0.5,0.5),(0.5,0.5)\}$, $D_2 = \{(1,1),(0.5,0.5)\}$, and $D_3 = \{(0,1),(0.5,0.5)\}$

	D0	D1	D2	D3
E[f(y|H0)|H1]	9.251962	6.0488940	4.1099478	4.8138863
E[f(y|H1)|H1]	9.251962	7.7712084	5.0994378	5.8255014
tr(XVX')|H0	0.010000	0.0150000	0.0225000	0.0175000
tr(XVX')|H1	0.010000	0.0162500	0.0331250	0.0231250
E[P(H0|Y,D)|H1,D]	0.500000	0.3995208	0.3724031	0.3915504
E[P(H1|Y,D)|H1,D]	0.500000	0.6004792	0.6275969	0.6084496

- It seems that expected posterior probabilities of hypotheses for D_1, D_2, D_3 are pretty close, in spite of D_2, D_3 including less important points
- It is also suspicious that D_2 does better than D_1
- there seems some kind of mechanism for disfavoring repeated points (and maybe does other things?)