

Manuál programu

bending.py

verze: 1.00

Program pro vážený bending korelační a kovarianční matice

2023

Autoři: Zdeňka Veselá, Luboš Vostrý, Michaela Brzáková, Alexandra Novotná

Výzkumný ústav živočišné výroby, v.v.i.

Přátelství 815

104 00 Praha – Uhřetěves

Česká republika

Vypracováno za podpory MZe – projekt NAZV QK1910059 „Využití genomických údajů a optimalizace šlechtitelských postupů u masného skotu“

Obsah

Úvod	2
Metodický postup.....	3
Systémové požadavky	3
Spuštění programu	3
Podkladové soubory	3
matice	3
vahy	4
Výsledky.....	4
Příklady	5
Použitá literatura	6

Úvod

Program bending.py byl vyvinut pro potřeby bendingu korelační nebo kovarianční matice, tak aby bylo dosaženo matice pozitivně definitní.

Matice genetických variancí a kovariancí je nezbytným předpokladem a vstupním údajem pro předpovědi plemenných hodnot metodou BLUP (best linear unbiased prediction). Takováto matice musí splňovat základní předpoklady – musí být čtvercová, symetrická a pozitivně definitní, tak aby byla možná její inverze. Genetické variance a kovariance je nutné před vlastní předpovědí plemenných hodnot odhadnout. V případě jednoduchých modelů zohledňujících menší počet vlastností probíhá odhad genetických variancí a kovariancí společně a v takových případech není nutné získanou kovarianční matici nijak upravovat. V případě složitějších víceznakových modelů s větším množstvím vzájemně korelovaných vlastností, je však často nezbytné jednotlivé prvky kovarianční matice odhadovat samostatně pro menší skupiny vlastností. Důvodem může být to, že komplikovanější modely nedosahují konvergence nebo jsou výpočetně a časově velmi náročné. V takových případech jsou genetické korelace (kovariance) často odhadovány pouze pro kombinaci dvou vzájemně korelovaných vlastností. Takto získané korelace (kovariance) jsou následně použity k vytvoření kompletní matice pro všechny vzájemně korelované vlastnosti. Tato matice však v naprosté většině případů není pozitivně definitní a je nutné ji převést na pozitivně definitní matici postupem nazývaným bending (Hayes a Hill 1981; Meuwissen a Kanis 1988; Essl 2010).

Klasický bending bez zohlednění významnosti jednotlivých korelací je však ve většině případů nežádoucí, neboť při sestavování korelační matice požadujeme, aby u některých korelací (například těch odhadnutých s významně nižší chybou odhadu nebo na podkladě početnějších souborů) docházelo po bendingu k minimálním změnám. Naopak u méně důvěryhodných korelací (například korelace odhadnuté s velkou chybou odhadu, statisticky neprůkazné nebo korelace neodhadnutelné, a tudíž dosazené intuitivně, či na podkladě literárních zdrojů) máme mnohem méně striktní požadavky na jejich případnou změnu po bendingu. Zohlednění významnosti jednotlivých korelací (případně variancí a kovariancí) umožňuje postup váženého bendingu popsáný v práci (Jorjani et al. 2003).

Program bending.py umožňuje provést bending jak metodou klasického postupu bez vážení, tak váženým bendingem. Dále umožňuje provést bending korelační i kovarianční matice. Bending kovarianční matice vede ke změnám jak variancí (na diagonále), tak kovariancí a tím pádem také

odhadnutých koeficientů dědivosti. Bending korelační matice vede pouze ke změnám korelací (respektive kovariancí) a při následném převedení zpět na kovarianční matici umožňuje ponechat hodnoty variancí a koeficientů dědivosti beze změny. Bending korelační matice je vhodný například pro mezinárodní genetické hodnocení. V takovém případě jsou variance odhadovány uvnitř jednotlivých zemí na podkladě velkého množství užitekostí, zatímco korelace (kovariance) mezi zeměmi jsou odhadovány na podkladě menších souborů jedinců, kteří jsou geneticky provázáni.

Metodický postup

Metodický postup programu bending.py vychází z prací Jorjani et al. (2002) a Jorjani et al. (2003).

Systémové požadavky

Přestože je k dosažení pozitivně definitní matice pomocí váženého bendingu zapotřebí poměrně velký počet iterací, je celý postup výpočetně a časově nenáročný. Jeho náročnost je přímo úměrná velikosti korelační (kovarianční) matice.

Program bending.py je psaný v programovacím jazyku Python 3, není jej tedy zapotřebí kompilovat. Lze jej spustit v jakémkoliv operačním systému. Pro jeho úspěšné spuštění je nezbytná instalace **Python 3** a dále knihovny (modulu) **Numpy**.

Spuštění programu

Pro spuštění programu jsou zapotřebí dva soubory – „matice“ (soubor s kovarianční, nebo korelační maticí) a „vahy“ (soubor s váhami použitými pro vážený, případně nevážený bending). Oba tyto soubory musí být nahrány do stejného adresáře jako program bending.py. Název souborů není možné měnit bez zásahu do programu bending.py. Program bending.py spustíte běžným příkazem pro spuštění programů v jazyce Python 3 (např. `$ python bending.py` v Linuxu).

Podkladové soubory

matice

Matice korelací nebo kovariancí. Jednotlivé prvky matice musí být odděleny mezerami. Matice musí být čtvercová a symetrická. Hodnoty korelací v korelační matici musí nabývat hodnot od 0 do 1. Na diagonále korelační matice musí být uvedeny hodnoty 1.

Příklad korelační matice:

1	0.95	0.8	0.4	0.4
0.95	1	0.95	0.8	0.4
0.8	0.95	1	0.95	0.8
0.4	0.8	0.95	1	0.95
0.4	0.4	0.8	0.95	1

Příklad kovarianční matice:

100	95	80	40	40
95	100	95	80	40
80	95	100	95	80
40	80	95	100	95
40	40	80	95	100

Program provede kontrolu vstupních matice a automaticky zhodnotí, zda se jedná o korelační, či kovarianční matici. Pokud je průměr mimodiagonálních prvků matice nižší než 1 a zároveň je průměr diagonálních prvků roven 1, program pracuje se zadanou maticí jako s maticí korelační.

vahy

Matice s váhami pro vážený bending. Jednotlivé prvky matice musí být odděleny mezerami. Počet prvků v souboru „vahy“ musí být shodný s počty prvků v souboru „matice“. Matice musí být čtvercová a symetrická. Váhy lze zadávat třemi způsoby:

1. Váhy vyšší než 1: V tomto případě program předpokládá, že čím je zadaná hodnota vyšší, tím je váha dané korelace důležitější. U korelací, kterým je přiřazena vysoká váha, dochází k menším změnám než u korelací, kterým je přiřazena nízká váha. Příkladem takovýchto vah jsou počty jedinců (užitkovosti) – např. počty společných plemenů při odhadu genetických korelací mezi dvěma zeměmi.

1000	500	20	50	200
500	1000	500	5	50
20	500	1000	20	20
50	5	20	1000	200
200	50	20	200	1000

2. Váhy nižší než 1: V tomto případě program předpokládá, že čím je zadaná hodnota nižší, tím je váha dané korelace důležitější. Příkladem použití takovýchto vah jsou standardní chyby odhadu.

0.001	0.002	0.05	0.02	0.005
0.002	0.001	0.002	0.2	0.02
0.05	0.002	0.001	0.05	0.05
0.02	0.2	0.05	0.001	0.005
0.005	0.02	0.05	0.005	0.001

3. Všechny váhy jsou rovné 1: V tomto případě program spustí bending bez použití vah. Respektive všem korelacím (kovariancím) je přiřazena stejná váha.

1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1

Výsledky

Program ukládá výsledné pozitivně definitní matice do stejného adresáře. Pozitivně definitní korelační matice je ukládána do souboru „cor-bend“. V případě bendingu kovarianční matice dále ukládá kromě „cor-bend“ také pozitivně definitní kovarianční matici „cov-bend“. Další výsledky (např. počet iterací potřebných k dosažení pozitivně definitní matice a odchylky mezi původní maticí a maticí po bendingu) jsou vypisovány formou standardního výstupu do konzole. Přesměrovat do samostatného souboru lze klasicky pomocí operátoru „>“ (např. \$ python bending.py > vysledky.log).

Příklad výstupu programu:

Zadaná korelační matice:

```
[[1.    0.95 0.8   0.4   0.4 ]
 [0.95 1.    0.95 0.8   0.4 ]
 [0.8   0.95 1.    0.95 0.8 ]
 [0.4   0.8   0.95 1.    0.95]
 [0.4   0.4   0.8   0.95 1.   ]]
```

Zadaná matice vah:

```
[[1000.  500.   20.   50.  200.]
 [ 500. 1000.  500.    5.   50.]
 [  20.  500. 1000.   20.   20.]
 [  50.    5.   20. 1000.  200.]
 [ 200.   50.   20.  200. 1000.]]
```

Průměrné mimodiagonální hodnoty v matici vah: 156.5 > 0

Pro vážení při bendingu budou váhy převedeny na reciproční hodnoty.

Váhy použité při bendingu:

```
[[0.001 0.002 0.05  0.02  0.005]
 [0.002 0.001 0.002 0.2   0.02 ]
 [0.05  0.002 0.001 0.05  0.05 ]
 [0.02  0.2   0.05  0.001 0.005]
 [0.005 0.02  0.05  0.005 0.001]]
```

Počet iterací: 1517

Pozitivně definitní korelační matice:

```
[[1.          0.9357382  0.778766  0.3792007  0.38913593]
 [0.9357382  1.          0.92820954 0.49242246 0.41014093]
 [0.778766   0.92820954 1.          0.7418328  0.6212469 ]
 [0.3792007  0.49242246 0.7418328  1.          0.93374187]
 [0.38913593 0.41014093 0.6212469  0.93374187 1.          ]]
```

Odchyvky korelační matice po bendingu:

```
[[ 0.          0.01426179  0.02123402  0.0207993  0.01086407]
 [ 0.01426179  0.          0.02179046  0.30757754 -0.01014093]
 [ 0.02123402  0.02179046  0.          0.20816721  0.17875313]
 [ 0.0207993  0.30757754  0.20816721  0.          0.01625813]
 [ 0.01086407 -0.01014093  0.17875313  0.01625813  0.          ]]
```

Příklady

Konkrétní příklady použití programu bending.py jsou v adresářích priklad_01 (bending korelační matice vážený počty jedinců), priklad_02 (bending kovarianční matice vážený standardními chybami odhadu) a priklad_03 (bending korelační matice bez vážení). Součástí příkladů jsou vstupní soubory „matice“ a „vahy“ a výsledné soubory „cor-bend“, případně „cov-bend“.

Použitá literatura

ESSL, A., 2010. Choice of an appropriate bending factor using prior knowledge of the parameters. *Journal of Animal Breeding and Genetics* [online]. **108**, 89–101. Dostupné z: doi:10.1111/j.1439-0388.1991.tb00162.x

HAYES, J. F. a W. G. HILL, 1981. Modification of Estimates of Parameters in the Construction of Genetic Selection Indices ('Bending'). *Biometrics* [online]. **37**(3), 483. ISSN 0006341X. Dostupné z: doi:10.2307/2530561

JORJANI, H., L. KLEI a U. EMANUELSON, 2003. A Simple Method for Weighted Bending of Genetic (Co)variance Matrices. *Journal of Dairy Science* [online]. **86**(2), 677–679. ISSN 00220302. Dostupné z: doi:10.3168/jds.S0022-0302(03)73646-7

JORJANI, Hossein, Bert KLEI a Ulf EMANUELSON, 2002. Combining Disparate Estimates of Genetic Correlations. *Interbull Bulletin*. **29**, 1–3.

MEUWISSEN, T. H. E a E KANIS, 1988. Application of bending theory in a pig-breeding situation. *Livestock Production Science* [online]. **18**(1), 85–91. ISSN 0301-6226. Dostupné z: doi:10.1016/0301-6226(88)90069-3