Bias in variance component estimation in culled data

Árnason, T.[1], Albertsdóttir, E.[1], Eriksson S.[2], Fikse WF.[2] & Sigurdsson Á.[1]

AUI[1], SLU[2]

Background

- Pre-selection or culling on an unobservable criterion before testing is a common problem in horse breeding
- Culling on criterium correlated with tested traits causes bias in estimation of genetic parameters and genetic evaluations (Robertson,1966,Meyer & Thompson,1984)

Background cont.

- Survival = Racing-status =**Test-status** = 0/1 binary trait
- **If** record available on Gaussian trait **then** test-status=1 **else** test-status=0
- Requires pedigree information on culled animals
- Estimated heritabilty of test-status moderate to high in many horse populations
- EBVs for test-status often important

Background cont.

- Árnason(1999) showed by simulations that bivariate genetic evaluations involving racing-status and a Gaussian trait reduced selection bias, increased accuracy, and increased genetic response when **true** genetic parameters were available
- The procedure was validated by method **R** in Swedish standardbred trotters (SST)
- Racing-status has been included in MT-AM-BLUP for genetic evaluations of SST since 1995

Background cont.

- Genetic and environmental correlations between test-status and the Gaussian performance traits are not readily estimable
- Genetic covariance estimable if environmental covariance is constrained to predifined value
- Expected environmental covariance ?

Objective

- To study the consequences of assuming **zero environmental covariances** between a binary test-status and a Gaussian trait on the estimates of genetic parameters **in culled data**, when the true covariances deviates from zero, using REML and Gibbs sampler methods

Simulated data

- Base population: random 25 males 500 females = 525 animals
- 3 generations random mating 500 females and 25 males per generation = 1575 animals. No phenotypes
- 5 generations: Selection on phenotypic records (mass selection across generations): 500 females and 25 males selected per generation producing 1500 ofspring per generation = 7500 records
- Pedigree list=9600 animals

Simulation procedure

- True genetic parameters: $h^2_1 = h^2_2 = 0.4$; r_A=0.5; r_E=0.0; 0.5; -0.5
- BV=$\mathbf{a}$=0.5$\mathbf{a}_s$ + 0.5$\mathbf{a}_d$ + $\mathbf{m}$ (Mendelian sampling term)
- $\mathbf{m}$=$\sqrt{}$(0.5 c)$\mathbf{C}_G$ $\mathbf{z}$; c=(1-0.5(F_s + F_D)); $\mathbf{V}_G$=$\mathbf{C}_G\mathbf{C}'_G$; $\mathbf{z}$=bivarate random sampled vector ($\mathbf{0}$,$\mathbf{I}$)
- $\mathbf{P}$=$\mathbf{Xb}$ + $\mathbf{a}$ +$\mathbf{C}_e\mathbf{z}$ (underlying phenotypes)
- Fixed effects = generation effects
- Culling frequency 0.5 (σ_{P1}<0) and 0.8 (σ_{P1}<0.84)
- 10 replicates

Model and methods for estimation of genetic parameters

- DMU (Jensen & Madsen, 2008)
- Animal Model Average Information Restriced Maximum Likelihood (AM-AI-REML)
- Multiple Trait Animal Model Linear Threshold Markov Chain Monte Carlo Gibbs Sampler (MT-AM-LT-MCMC-GS)

Results

Estimated variance components for the Gaussian trait (σ^2_A = 0.4; σ^2_E = 0.6; r_A = 0.5). ***ST-AM-AI-REML*** analysis

r_E	Var(A) (s.e.)	Var(E) (s.e.)	$\hat{h}^2$ (s.e)
50% culling			
0.0	**0.383 (0.016)**	**0.588 (0.014)**	**0.394 (0.015)**
0.5	**0.338 (0.012)**	**0.509 (0.010)**	**0.399 (0.013)**
-0.5	**0.498 (0.020)**	**0.520 (0.016)**	**0.489 (0.018)**
80% culling			
0.0	0.403 (0.014)	0.565 (0.011)	0.416 (0.012)
0.5	0.334 (0.022)	0.469 (0.015)	0.414 (0.036)
-0.5	0.534 (0.029)	0.485 (0.018)	0.522 (0.021)

Estimated variance components for the Gaussian trait (σ^2_A = 0.4; σ^2_E = 0.6; r_A = 0.5). **ST-AM-AI-REML** analysis

r_E	Var(A) (s.e.)	Var(E) (s.e.)	$\hat{h}^2$ (s.e)
50% culling			
0.0	0.383 (0.016)	0.588 (0.014)	0.394 (0.015)
0.5	0.338 (0.012)	0.509 (0.010)	0.399 (0.013)
-0.5	0.498 (0.020)	0.520 (0.016)	0.489 (0.018)
80% culling			
0.0	**0.403 (0.014)**	**0.565 (0.011)**	**0.416 (0.012)**
0.5	**0.334 (0.022)**	**0.469 (0.015)**	**0.414 (0.036)**
-0.5	**0.534 (0.029)**	**0.485 (0.018)**	**0.522 (0.021)**

Estimated variance components for the Gaussian trait (σ^2_A = 0.4; σ^2_E = 0.6; r_A = 0.5). **MT-AM-LT-GS** analysis

r_E	Var(A) (s.e.)	Var(E) (s.e.)	$\hat{h}^2$ (s.e)	$\hat{r}_A$ (s.e)
50% culling				
0.0	**0.397 (0.016)**	**0.606 (0.013)**	**0.395 (0.014)**	**0.478 (0.021)**
0.5	**0.349 (0.012)**	**0.519 (0.009)**	**0.402 (0.013)**	**0.278 (0.040)**
-0.5	**0.534 (0.017)**	**0.540 (0.014)**	**0.497 (0.018)**	**0.621 (0.018)**
80% culling				
0.0	0.402 (0.019)	0.598 (0.010)	0.401 (0.014)	0.466 (0.048)
0.5	0.325 (0.019)	0.491 (0.013)	0.396 (0.020)	0.100 (0.041)
-0.5	0.610 (0.035)	0.505 (0.019)	0.544 (0.022)	0.715 (0.027)

Estimated variance components for the Gaussian trait (σ^2_A = 0.4; σ^2_E = 0.6; r_A = 0.5). **MT-AM-LT-GS** analysis

r_E	Var(A) (s.e.)	Var(E) (s.e.)	$\hat{h}^2$ (s.e)	$\hat{r}_A$ (s.e)
50% culling				
0.0	0.397 (0.016)	0.606 (0.013)	0.395 (0.014)	0.478 (0.021)
0.5	0.349 (0.012)	0.519 (0.009)	0.402 (0.013)	0.278 (0.040)
-0.5	0.534 (0.017)	0.540 (0.014)	0.497 (0.018)	0.621 (0.018)
80% culling				
0.0	**0.402 (0.019)**	**0.598 (0.010)**	**0.401 (0.014)**	**0.466 (0.048)**
0.5	**0.325 (0.019)**	**0.491 (0.013)**	**0.396 (0.020)**	**0.100 (0.041)**
-0.5	**0.610 (0.035)**	**0.505 (0.019)**	**0.544 (0.022)**	**0.715 (0.027)**

Jump to the Conclusions

Conclusions

- For the parameter combination tested, heritability estimates for the Gaussian trait were significantly biased (upwards) when r_E was negative
- Genetic correlations between test-status and the Gaussian trait were systematically underestimated when r_E was positive and overestimated when r_E was negative

Conclusions - in a nutshell

- The assumption of zero environmental correlation between the culling criterion test-status and a Gaussian trait may sometimes lead to bad estimates of genetic parameters
- Works well if the true r_E is zero (or known)
- Informative prior values for r_E would be valuable

Future directions

- The next step in this study is to investigate the consequences of including test-status in the genetic evaluations of a Gaussian trait in bivariate analysis when the genetic parameters used are wrong
- Preliminary results show that inclusion of test status generally enhances genetic progress even if the genetic parameters used deviate considerably from the true ones!
- Validation important in each case!

Ευχαριστώ - Thank you