Bibliography

Abecasis, G. R., Cherny, S. S., and Cardon, L. R. (2001). The impact of genotyping error on family-based analysis of quantitative traits. *European Journal of Human Genetics*, 9:130–134.

Adler, R. J. (1981). *The Geometry of Random Fields*. John Wiley & Sons, Chichester.

Akey, J. M., Zhang, K., Xiong, M., Doris, P., and Jin, L. (2001). The effect that genotyping errors have on the robustness of common linkage-disequilibrium measures. *American Journal of Human Genetics*, 68:1447–1456.

Bair, E., Hastie, T., Paul, D., and Tibshirani, R. (2006). Prediction by supervised principal components. *Journal of the American Statistical Association*, 101:119–137.

Bair, E. and Tibshirani, R. (2004). Semi-supervised methods to predict patient survival from gene expression data. *Public Library of Science Biology*, 2:511–522.

Baird, D., Johnstone, P., and Wilson, T. (2004). Normalization of microarray data using a spatial mixed model analysis which includes splines. *Biometrics*, 20:3196–3205.

Balagurunathan, Y., Dougherty, E. R.. Chen, Y., Bittner, M. L., and Trent, J. M. (2002). Simulation of cDNA microarrays via a parameterized random signal model. *Journal of Biomedical Optics*, 7:507–523.

Bellman, R. E. (1961). *Adaptive control processes : a guided tour.* Princeton, NJ. : Princeton University Press.

Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of a genetic-linkage map in man using restricted fragment length polymorphisms. *American Journal of Human Genetics*, 32:314–331.

Burgueño, J., Crossa, J., Grimanelli, D., Leblanc, O., and Autran, D. (2005). Spatial analysis of cDNA microarray experiments. *Crop Science*, 45:748–757.

Chen, Y., Dougherty, E. R., and Bittner, M. L. (1997). Ratio-based decisions and the quantitative analysis of cDNA microarray images. *Journal of Biomedical Optics*, 2:264–374.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74:829–836.

Constantine, A. G. and Hall, P. (1994). Characterising surface smoothness via estimation of effective fractal dimension. *Journal of the Royal Statistical Society, Series B*, 56:97–113.

Cressie, N. A. (1993). *Statistics for Spatial Data.* John Wiley & Sons, New York.

Cui, X., Kerr, M. K., and Churchill, G. A. (2003). Transformations for cDNA microarray data. *Statistical Applications in Genetics and Molecular Biology*, 2:4.

Cullis, B. R. and Gleeson, A. C. (1991). Spatial analysis of field experiments - an extension to two dimensions. *Biometrics*, 47:1449–1460.

Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 41:909–996.

de Pison Ascacibar, J. M., Mere, J. O., Limas, M. C., and de Cos Juez, J. (2000). *fdim: Functions for calculating fractal dimension.* R package version 1.0-3.

DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A., and Trent, J. M. (1996). Use of a cDNA microarray to analyse gene expression patterns in human cancer *Nature Genetics*, 14:457–4600.

Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81:425–455.

Douglas, J. A., Boehnke, M., and Lange, K. (2000). A multipoint method for detecting genotyping errors and mutations in sibling-pair linkage data. *The American Journal of Human Genetics*, 66:1287–1297.

Dudoit, S., Yang, Y. H., Callow, M. J., and Speed, T. P. (2002). Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica Sinica*, 12:111–139.

Efron, B., Tibshirani, R., Storey, J. D., and Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment. *Journal of the American Statistical Association*, 96:1151–1160.

Elston, R., Olsen, J., and Palmer, L., editors (2002). *Biostatistical Genetics and Genetic Epidemiology.* John Wiley and Sons, Chichester, England.

Falconer, K. J. (1990). *Fractal Geometry: Mathematical Foundations and Applications.* John Wiley & Sons, Chichester.

Fan, J., Peng, H., and Huang, T. (2005). Semilinear high-dimensional model for normalization of microarray data: A theoretical analysis and partial consistency. *Journal of the American Statistical Association*, 100:781–796.

Fan, J., Tam, P., Vande Woude, G., and Ren, Y. (2004). Normalization and analysis of cDNA microarrays using within-array replications applied to neuroblastoma ncell response to a cytokine. *Proceedings of the National Academy of Sciences*, 101:1135–1140.

Fernando, R. L. and Grossman, M. (1989). Marker assisted selection using best linear unbiased prediction. *Genetics Selection Evolution*, 21:467–477.

Fisher, R. A. (1971). *The Design of Experiments, 9th edition.* Hafner Press, New York.

Geldermann, H. (1975). Investigation on inheritance of quantitative characters in animals by gene markers. I. Methods. *Theoretical and Applied Genetics*, 46:319–330.

Gianola, D., Fernando, R. L., and Stella, A. (2006). Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics*, 173:1761–1776.

Gianola, D., Perez-Enciso, M., and Toro, M. A. (2003). On marker-assisted prediction of genetic value: Beyond the ridge. *Genetics*, 163:347–365.

Gilmour, A. R., Cullis, B. R., and Verbyla, A. P. (1997). Accounting for natural and

extraneous variation in the analysis of field experiments. *Journal of Agricultural, Biological and Environmental Statistics*, 2:269–293.

Gilmour, A. R., Cullis, B. R., Welham, S. J., and Thompson, R. (2000). *ASREML Reference Manual, 2nd edition. NSW Agriculture, Orange. Australia.*

Gneiting, T. and Schlather, M. (2004). Stochastic models that separate fractal dimension and the hurst effect. *SIAM Reviews*, 46:269–282.

Göring, H. H. H. and Terwilliger, J. D. (2000). Linkage analysis in the presence of errors II: Marker-locus genotyping errors modeled with hypercomplex recombination fractions. *American Journal of Human Genetics*, 66:1107–1118.

Hahn, L. W., Ritchie, M. D., and Moore, J. H. (2003). Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. *Bioinformatics*, 19:376–382.

Hall, P. and Wood, A. (1993). On the performance of box-counting estimators of fractal dimension. *Biometrika*, 80:246–251.

Hammond, K., Graser, H.-U., and McDonald, C., editors (1992). *Animal Breeding: The modern approach.* Post Graduate Foundation in Veterinary Science, University of Sydney.

Hayes, B. and Goddard, M. E. (2001). The distribution of the effects of genes affecting quantitative traits in livestock. *Genetics Selection Evolution*, 33:209–229.

He, J., Westbrooks, K., and Zelikovsky, A. (2005). Linear reduction methods for tag SNP selection. *International Journal on Bioinformatics Research and Applications*, 1:249–260.

Henderson, C. R. (1973). Sire evaluation and genetic trends. *Proceeding of the Animal Breeding and Genetics Symposium in Honor of Dr. J.L. Lush, American Society of Animal Science and Dairy Science Association, Champaign, IL*, pages 10–41.

Hoeschele, I. and Li, H. (2005). A note on joint versus gene-specific mixed model analysis of microarray gene expression data. *Biostatistics*, 6:183–186.

Horne, B. D. and Camp, N. J. (2004). Principal component analysis for selection of optimal SNP-sets that capture intragenic genetic variation. *Genetic Epidemiology*, 26:11–21.

Huang, H.-C. and Cressie, N. (2000). Deterministic/stochastic wavelet decomposition for recovery of signal from noisy data. *Technometrics*, 42:262–276.

Huang, J., Wang, D., and Zhang, C.-H. (2005). A two-way semilinear model for normalization and analysis of cDNA microarray data. *Journal of the American Statistical Association*, 100:814–829.

Huber, P. J. (1985). Projection pursuit. *The Annals of Statistics*, 13:435–475.

Jin, W., Riley, R. M., Wolfinger, R. D., White, K. P., Passador-Gurgel, G., and Gibson, G. (2001). The contributions of sex, genotype and age to transcriptional variance in drosophila melanogaster. *Nature Genetics*, 29:389–395.

Johnson, R. A. and Wichern, D. W., editors (1988). *Applied multivariate statistical analysis*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Jolliffe, I. T. (1986). *Principal Component Analysis*. Springer, New York.

Kerr, M. K. (2003). Design considerations for efficient and effective microarray studies. *Biometrics*, 59:822–828.

Kerr, M. K. and Churchill, G. A. (2001a). Experimental design for gene expression microarrays. *Biostatistics*, 2:183–201.

Kerr, M. K. and Churchill, G. A. (2001b). Statistical design and the analysis of gene expression microarray data. *Genetical Research*, 77:123–128.

Kerr, M. K., Martin, M., and Churchill, G. A. (2000). Analysis of variance for gene expression microarray data. *Journal of Computational Biology*, 7:819–837.

Lande, R. and Thompson, R. (1990). Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics*, 124:743–756.

Lin, Z. and Altman, R. B. (2004). Finding haplotype tagging SNPs by use of principal components analysis. *American Journal of Human Genetics*, 75:850–861.

Liò, P. (2003). Wavelets in bioinformatics and computational biology: state of the art perspectives. *Bioinformatics*, 19:2–9.

Mallat, S. (1989). A theory for multiresolutional signal decomposition. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 11:674–693.

Mandelbrot, B. B. and Van Wess, J. W. (1968). Fractional brownian motions, fractional noises and applications. *SIAM Review*, 10:422–437.

McIntyre, G. A. (1955). Design and analysis of two phase experiments. *Biometrics*, 11:324–334.

McLachlan, G. J., Bean, R. W., Jones, L. B.-T., and Zhu, J. (2005). Using mixture models to detect differentially expressed genes. *Australian Journal of Experimental Agriculture*, 45:859–866.

Meuwissen, T. H. E., , and Goddard, M. E. (1996). The use of marker haplotypes in animal breeding schemes. *Genetics Selection Evolution*, 28:161–176.

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157:1819–1829.

Meyer, Y. (1993). *Wavelets: Algorithms and Applications*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and its Applications*, 9:141–142.

Nason, G. P. (2004). *wavethresh: Software to perform wavelet statistics and transforms*. R package version 2.2-8.

Nguyen, D. V., Arpat, A. B., Wang, N., and Carroll, R. J. (2002). DNA microarray experiments: Biological and technological aspects. *Biometrics*, 58:701–717.

Pan, W. (2002). A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments. *Bioinformatics*, 18:546–554.

Quaas, R. L. (1988). Additive genetic model with groups and relationships. *Journal of Dairy Science*, 71:1338–1345.

Quaas, R. L., Everett, E. W., and McClintock, A. C. (1979). Maternal grandsire model for dairy sire evaluation. *Journal of Dairy Science*, 62:1648–1654.

Quaas, R. L. and Pollak, E. J. (1980). Mixed model methodology for farm ranch beef cattle testing programs. *Journal of Animal Science*, 51:1277–1287.

R Development Core Team (2004). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reverter, A., Byrne, K. A., Bruce, H. L., Wang, Y. H., Dalrymple, B. P., and Lehnert, S. (2003). A mixture model-based cluster analysis of DNA microarray gene expression data on brahman and brahman composite steers fed high-,medium-, and low-quality diets. *Journal of Animal Science*, 81:1900–1910.

Reverter, A., Wang, Y. H., Byrne, K. A., Tan, S. H., Harper, G. S., and Lehnert, S. A. (2004). Joint analysis of multiple cDNA microarray studies via multivariate mixed models applied to genetic improvement of beef cattle. *Journal of Animal Science*, 82:3430–3439.

Ritchie, M. D., Hahn, L. W., and Moore, J. H. (2003). Power of multifactor dimensionality reduction for detecting gene-gene interactions in the presence of genotyping error, missing data, phenocopy, and genetic heterogeneity. *Genetic Epidemiology*, 24:150–157.

Ritchie, M. D., Hahn, L. W., Roodi, N., Bailey, L. R., Dupont, W. D., Par, F. F., and Moore, J. H. (2001). Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *American Journal of Human Genetics*, 69:138–147.

Robinson, G. K. (1991). That BLUP is a good thing: The estimation of random effects. *Statistical Science*, 6:15–32.

Roweis, S. (1997). EM algorithms for PCA and SPCA. *Neural Information Processing Systems*, (10):626–632.

Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M.and Stein, L. D., and Sherry, S. e. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, 409:928–933.

Sapir, M. and Churchill, G. A. (2000). Estimating the posterior probability of differential gene expression from microarray data. Technical report, The Jackson Laboratory.

Scaccia, L. and Martin, R. J. (2005). Testing axial symmetry and separability of lattice processes. *Journal of Statistical Planning and Inference*, 131:19–39.

Schadt, E. E., Li, C., Ellis, B., and Wong, W. H. (2001). Feature extraction and normalization algorithm for high-density oligonucleotide gene expression array data. *Journal of Cellular Biochemistry Supplement*, 37:120–125.

Schadt, E. E., Monks, S. A., Drake, T. A., Lusis, A. J., Che, N., Chen, N., Colinayo, V., Ruff, T. G., Milligan, S. B., Lamb, J. R., Cavet, G., Linsley, P. S., Mao, M., Stoughton, R. B., and Friend, S. H. (2003). Genetics of gene expression surveyed in maize, mouse and man. *Nature*, 422:297–302.

Schaeffer, L. R. (1984). Sire and cow evaluation under multiple trait models. *Journal of Dairy Science*, 67:1567–1580.

Schaeffer, L. R. (2006). Strategy for applying genome-wide selection in dairy cattle. *Journal of Animal Breeding and Genetics*, 123:218–223.

Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 270:467–470.

Shai, O., Morris, Q., and Frey, B. (2003). Spatial bias removal in microarray images. Technical report, University of Toronto PSI-2003-21.

Sharma, S. (1996). *Applied multivariate techniques.* John Wiley & Sons, Inc., New York.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman & Hall, London.

Smyth, G. K. and Speed, T. P. (2003). Normalization of cDNA microarray data. *Methods*, 31:265–273.

Soller, M. and Beckmann, J. S. (1983). Genetic polymorphism in varietal identification and genetic improvement. *Theoretical and Applied Genetics*, 67:25–33.

Sorenson, D. A. and Kennedy, B. W. (1984). Estimation of genetic variances from unselected and selected populations. *Journal of Animal Science*, 59:1213–1233.

Speed, T. P., editor (2003). *Statistical analysis of gene expression microarray data.* Chapman & Hall/CRC Press, Boca Raton, FL.

Thomas, J. G., Olson, J. M., Tapscott, S. J., and Zhao, L. P. (2001). An efficient

and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles. *Genome Research*, 11:1227–1236.

Tseng, G. C., Oh, M.-K., Rohlin, L., Liao, J. C., and Wong, W. H. (2001). Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. *Nucleic Acids Research*, 29:2549–2557.

Van Arendonk, J. A. M., Tier, B., and Kinghorn, B. P. (1994). Use of multiple genetic markers in prediction of breeding values. *Genetics*, 137:319–329.

Vinciotti, V., Khanin, R., D'Alimonte, D., Liu, X., Cattini, N., Hotchkiss, G., Bucca, G., de Jesus, O., Rasaiyaah, J., Smith, C. P., Kellam, P., and Wit, E. (2005). An experimental evaluation of a loop versus a reference design for two-channel microarrays. *Bioinformatics*, 21:492–501.

Watson, G. S. (1964). Smooth regression analysis. *Sankhya: The Indian Jounal of Statistics, Series A*, 26:359–372.

Westell, R. A., Quaas, R. L., and Van Vleck, L. D. (1988). Genetic groups in an animal model. *Journal of Dairy Science*, 71:1310–1318.

Westell, R. A. and Van Vleck, L. D. (1987). Simultaneous genetic evaluation of sires and cows for a large population of dairy cows. *Journal of Dairy Science*, 70:1006–1017.

Wierling, C. K., Steinfath, M., Elge, T., Schulze-Kremer, S., Aanstad, P., Clark, M., Lehrach, H., and Herwig, R. (2002). Simulation of DNA array hybridization experiments and evaluation of critical parameters during subsequent image and data analysis. *BMC Bioinformatics*, 3:29.

Wolfinger, R. D., Gibson, G., Wolfinger, E. D., Bennet, L., Hamedah, H., Bushel, P., Afshari, C., and Paules, R. S. (2001). Assessing gene significance from cDNA microarray expression data via mixed models. *Journal of Computational Biology*, 8:625–637.

Wong, G. K., Liu, B., Wang, J., Zhang, Y., Yang, X., Zhang, Z., Meng, Q., Zhou, J., Li, D., and Zhang, J., e. (2004). A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature*, 432:717–722.

Woolaston, A. F., Murison, R. D., and Tier, B. (2005). Analysis of microarrays incorporating adjustments for spatial effects. *Proceedings of the Association for the Advancement of Animal Breeding and Genetics*, (16):262–265.

Woolaston, A. F., Tier, B., and Murison, R. D. (2007). Principal components analysis of SNP data to predict breeding value. *Genetics Selection Evolution*. (Submitted).

Wu, H., Kerr, M. K., Cui, X. Q., and Churchill, G. A. (2003). *MAANOVA: A Software Package for the Analysis of Spotted cDNA Microarray Experiments in the Analysis of Gene Expression Data: An Overview of Methods and Software*. Springer, New York.

Xu, S. (2003). Estimating polygenic effects using markers of the entire genome. *Genetics*, 163:789–801.

Yang, Y. H., Dudoit, S., Luu, P., Lin, D. M., Peng, V., Ngai, J., and Speed, T. P. (2002). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Research*, 30:e15.

Yang, Y. H. and Speed, T. P. (2002). Design issues for cDNA microarray experiments. *Nature Reviews Genetics*, 3:579–588.

Zimdahl, H., Nyakatura, G., Brandt, P., Schulz, H., Hummel, O., Fartmann, B., Campbell, W. D., Droege, M., Monti, J., Lee, Y.-A., Sun, Y., Zhao, S., Winter, E. E., Ponting, C. P., Chen, Y., Kasprzyk, A., Birney, E., Ganten, D., and Hubner, N. (2004). A SNP map of the rat genome generated from cDNA sequences. *Science*, 303:807.

Appendix A

A.1 Algorithm for simulation of population and SNPs

(i) Location of SNPs, l(i), and probability of mutations, p(i), at the SNP loci:

for i = 1, 2, ..., n_s **do**

Sample l(i) from discrete uniform(0,20million)

Sample p(i) from uniform(0,0.5)

end for

(ii) Base population of chromosomes, B(i, j):

for j = 1, 2, ..., 200 **do**

for i = 1, 2, ..., n_s **do**

Sample B(i, j) from Binom(1, p(i))

end for

end for

B(, 1 : 30) = imale(j,1,l) (j = 1, 2, ..., 30 and l = 1, 2, ...,n_s)

B(, 31 : 60) = imale(j,2,l) (j = 1, 2, ..., 30 and l = 1, 2, ...,n_s)

B(, 61 : 130) = ifemale(j,1,l) (j = 1, 2, ..., 70 and l = 1, 2, ...,n_s)

B(, 131 : 200) = ifemale(j,2,l) (j = 1, 2, ..., 70 and l = 1, 2, ...,n_s)

(iii) Base population:

for j = 1, 2, ..., 500 **do**

Randomly choose a sire from 1, 2, ..., 30

Randomly choose starting chromosome for sire, mchr, from 1, 2

Randomly choose a dam from 1, 2 ..., 70

Randomly choose starting chromosome, for dam, fchr, from 1, 2

length=0

while length< 20 million **do**

Sample length to next crossover, lc, from Poisson(1 million)

length=length+lc

for k in 1, 2, .., n_s **do**

if l(k)>length **then**

Switch mchr; BREAK

new(j,1,k)= imale(sire, mchr, k)

end if

end for

end while

length=0

while length<20 million **do**

Sample length to next crossover, lc, from Poisson(1 million)

length=length+lc

for k in 1, 2, .., n_s **do**

if l(k)>length **then**

Switch fchr; BREAK

else

new(j,2,k)= ifemale(dam, fchr, k)

end if

end for

end while

end for

new(1:40, ,) = male(, ,)

new(41:500, ,) = female(, ,)

(iv) Subsequent generations:

for gen=1,2,...10 **do**

for j = 1, 2, ..., 395 **do**

Randomly choose a sire from 1, 2, ..., 40

Randomly choose starting chromosome for sire, mchr, from 1, 2

Randomly choose a dam from 1. 2, ..., 460

Randomly choose starting chromosome for dam, fchr, from 1, 2

length=0

while length< 20 million **do**

Sample length to next crossover, lc, from Poisson(1 million)

length=length+lc

for k in 1, 2, .., ns **do**

if l(k)>length **then**

Switch mchr; BREAK

else

new1(j,1,k)= male(sire, mchr, k)

end if

end for

end while

length=0

while length<20 million **do**

Sample length to next crossover, lc, from Poisson(1 million)

length=length+lc

for k in 1, 2, .., ns **do**

if l(k)>length **then**

Switch fchr; BREAK

new1(j,2,k)= female(dam, fchr, k)

end if

end for

end while

end for

Set:

new1(1:10, ,) = male_{new}(1:10 , ,)

male(1:30, ,) = male_{new}(11:40 , ,)

new1(11:385, ,) = female_{new}(1:385 , ,)

female(1:75, ,) = female$_{new}$(386:460 , ,)

if gen>7 **then**

Keep animals from male$_{new}$ and female$_{new}$ that were not kept in the previous generation to form keep(, ,).

end if

end for

(v) Calculate MBVs and phenotypes:

for j = 1, 2, ..., n_a **do**

Sample additive SNP effects, a(j) from a $Gamma(0.59, 0.71)$ distribution.

end for

For each animal, count the number of mutations at the jth SNP with an additive effect,

$$q_{ij} = \text{keep}(i, 1, j) + \text{keep}(i, 2, j).$$

then,

$$T = \sum_{j=1}^{j=n_a} q_{ij} a_j,$$

and var(**T**) = σ_a^2. Calculate phenotypic values:

$$y_i = MBV_i + \epsilon,$$

$\epsilon \sim N(0, (\frac{\sigma_a^2}{h^2} - \sigma_a^2))$

A.2 Accuracy of kernel regression estimate

Define the linear function from MBVs to phenotypes:

$$P : \mathbb{R}^n \rightarrow \mathbb{R}^n$$

$$(MBV_1, MBV_2, ..., MBV_n) \rightarrow (y_1, y_2, ..., y_n), P(\mathbf{MBV}) := \mathbf{MBV} + \epsilon, \epsilon \sim N(0, \sigma_e^2),$$

and the linear function from the phenotypes to estimated MBVs:

$$\mathbf{S} : \mathbb{R}^n \rightarrow \mathbb{R}^n$$

$$(y_1, y_2, ..., y_n) \rightarrow (\hat{g}(x_1), \hat{g}(x_2), ..., \hat{g}(x_2)), \mathbf{S}(\mathbf{y}) = \mathbf{Sy},$$

where $\mathbf{S}$ is the matrix $\{w_{ij}\}$ and w_{ij} is defined in equation (5.6).

We require an expression for the accuracy of prediction, where accuracy is defined as below:

$$\text{accuracy} := \rho(MBV_i, \mathbf{S}(P(MBV_i))). \tag{A.1}$$

By definition,

$$\rho(MBV_i, \mathbf{S}(P(MBV_i))) = \frac{\text{cov}(MBV_i, \mathbf{S}(P(MBV_i)))}{\sqrt{\text{var}(MBV_i)\text{var}(\mathbf{S}(P(MBV_i)))}},$$

$$\text{var}(MBV_i) = \sigma_a^2,$$

and

$$h^2 == \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}.$$

Now,

$$\text{var}(P(MBV_i)) = \sigma_a^2 + \sigma_e^2.$$

After substituting $\sigma_e^2 = \frac{\sigma_a^2}{h^2} - \sigma_a^2$, it is found that

$$\text{var}(P(\mathbf{MBV_i})) = \frac{\sigma_a^2}{h^2},$$

so that:

$$\text{var}(\mathbf{S}(P(\mathbf{MBV_i}))) = \frac{\sigma_a^2}{h^2}(\mathbf{SS}^T)_{ii}.$$

Also,

$$\begin{aligned}\text{cov}(MBV_i, \mathbf{S}(P(MBV_i))) &= \text{cov}(MBV_i, \mathbf{S}(MBV_i + \epsilon)) \\ &= \text{cov}(MBV_i, \mathbf{S}(MBV_i) + \mathbf{S}\epsilon) \\ &= \text{cov}(MBV_i, \mathbf{S}(MBV_i)) \\ &= \text{cov}(MBV_i, MBV_i)\mathbf{S}^T\end{aligned}$$

and under the assumption that all animals are unrelated,

$$\text{cov}(MBV_i, \mathbf{S}(P(MBV_i))) = \sigma_a^2 S_{ii}.$$

Now,

$$\rho(MBV_i, \mathbf{S}(P(MBV_i))) = \frac{\sigma_a^2 S_{ii}}{\sqrt{\sigma_a^2 \frac{\sigma_a^2}{h^2}(\mathbf{SS}^T)_{ii}}} = \frac{\sqrt{h^2} S_{ii}}{\sqrt{(\mathbf{SS}^T)_{ii}}}.$$

That is, assuming animals are unrelated, the accuracy of prediction for the Nadaraya-Watson estimator is proportional to the square root of heritability.