

## **BTRY 7180: Generalized Linear Models**

### **Class project**

#### **Requirements and Suggestions**

##### Guidelines

The project has two components

1. A short (10min) presentation in the final week of classes (November 28 and 30). This should be a conference-style talk. The challenge is to comprehensibly condense a set of ideas and experiments into a concise presentation.
2. A short (10 page maximum) written report providing further details on your talk. This should be thought of as an accompaniment to your talk (ie a reminder to me of what you presented) – an explanation of the work you present without requiring substantial background or detailed experimental results. It should present what problem is considered, why it's interesting along with the main ideas and results.
3. You will be required to indicate the project you intend to take by the start of November, to remove duplication, and to ensure that the topic is neither too easy nor too large to be reasonably undertaken.

##### Possible Projects

It is expected that most of the talks here will present results from a published paper – a list of potential papers is given below. Feel free to nominate other papers if you think them of interest; this will be particularly appreciated since the list below is becoming a little “classical”.

Besides presenting a paper, you two further types of project are possible

1. Data analysis, if you happen to be involved in a project with relevant data. To be reasonable, this analysis should not doable within standard software. Ie, at least some non-standard choices of terms of links must be indicated.
2. Implementation of existing methods in software where they are not readily available.

In particular, an implementation of functional generalized linear models (similar to the Muller and Stadtmuller paper listed below) that is compatible with the ‘fda’ library in R would be appreciated by me.

## Paper List

1. Hauke, J., W. Walter and A. Donner, (1977), "Wald's test as applied to hypotheses in logit analysis", *Journal of the American Statistical Association*, 72, 851-853.  
Another paradox in logistic regression – there are a number of follow-up papers for anyone interested.
2. Muller, Hans-Georg and Ulrich Stadtmuller, "Generalized Functional Linear Models", *Annals of Statistics*, 33, 774-805.
3. Wood, S.N. (2011) "Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models", *Journal of the Royal Statistical Society (B)* 73(1):3-36
4. Wood, S.N. (2004) "Stable and efficient multiple smoothing parameter estimation for generalized additive models", *J. Amer. Statist. Ass.* 99:673-686.
5. Aitkin, Murray (1987) (see also Symth, 1989) "Modelling variance heterogeneity in normal regression using GLIM", *Applied Statistics*, 36, 332-339
6. Aitkin, Murray (1996) "A general maximum likelihood analysis of overdispersion in generalized linear models", *Statistics and Computing*, 6, 251-262
7. Danaher, Peter J. and Hardie, Bruce G.S. (2005), "Bacon with your eggs? Applications of a new bivariate beta-binomial distribution", *American Statistician* 59, pp.282-286
8. Drum, Melinda L. and McCullagh, Peter (1993) "REML estimation with exact covariance in the logistic mixed model", *Biometrics*, 49, 677-689
9. Efromovich, Sam and Thomas, Edward V. (1996), "Application of nonparametric binary regression to evaluate the sensitivity of explosives", *Technometrics*, 38, 50-58
10. Follmann, Dean and Wu, Margaret (1995) "An approximate generalized linear model with random effects for informative missing data (Corr: 97V53 p384)", *Biometrics*, 51, 151-168
11. Glonek, G. F. V. and McCullagh, P. (1995) "Multivariate logistic models", *Journal of the Royal Statistical Society, Series B, Methodological* 57, 533-546
12. Jiang, J. (1998), "Consistent estimators in generalized linear mixed models", *J. Amer. Statist. Assoc.* 93 720 729.
13. Jiang, Jiming (1999), "Conditional inference in generalized linear mixed models", *Ann. Stat.* 27, no. 6, 1974-2007
14. Kuk, A. Y. C. (1995) "Asymptotically unbiased estimation in generalized linear models with random effects", *J. Roy. Statist. Soc. Ser. B* 57 395 407.

15. Lang, J.B., McDonald, J.W. and Smith, P.W. (1999) "Association-marginal modeling of multivariate categorical responses: a maximum likelihood approach", *JASA*, 94, 1161-1171
16. Lin, D.Y., Wei, L.J., and Ying, Z. (2002) "Model-Checking Techniques Based on Cumulative Residuals", *Biometrics*, 58, 1-12.
17. Liu, Qing and Pierce, Donald A. (1993) "Heterogeneity in Mantel-Haenszel-type models", *Biometrika*, 80, 543-556
18. Mallick, Bani K. and Alan E. Gelfand (1994) Generalized Linear Models with Unknown Link Functions *Biometrika*, Vol. 81, No. 2, pp. 237-245.
19. McCullagh, Peter (1986) "The conditional distribution of goodness-of-fit statistics for discrete data", *Journal of the American Statistical Association*, 81, 104-107
20. Pierce, Donald A. and Schafer, Daniel W. (1986) "Residuals in generalized linear models", *Journal of the American Statistical Association*, 81, 977-986 (See also "On continuity-corrected residuals in logistic regression", Duffy (1990), *Biometrika* 77, 287-293)
21. Smyth, Gordon K. (1989) "Generalized linear models with varying dispersion", *Journal of the Royal Statistical Society, Series B, Methodological*, 51, 47-60
22. Waller, L.A. and Zelterman, D. (1997), "Loglinear modeling with the negative multinomial distribution", *Biometrics*, 53 971-982.
23. Yee, T. W. and Wild, C. J. (1996) "Vector generalized additive models" *Journal of the Royal Statistical Society, Series B, Methodological*, 58, 481-493
24. Pawitan, Y. (2001) "Two-staged estimation of variance components in generalized linear mixed models", *Journal of Statistical Computation and Simulation*, 69, 1-17