Prof. Dr. E. Cramer, Prof. Dr. M. Kateri
T. v. Bentum, M.Sc., L. Kaufmann, M.Sc.

RWTH Aachen, SS 2021

publication: July 9th — solutions available online: July 16th

Applied Data Analysis

Exercise Sheet 6

Exercise 24

Suppose $Y_1, \dots, Y_n$ are independent counts that satisfy a Poisson GLM with canonical link $\log\big(\mathsf{E}(Y_i)\big) = \log(\mu_i) = \sum_{j=1}^{p} \beta_j x_{ij}$, for a set of explanatory variables x_{ij}, $i = 1, \dots, n$, $j = 1, \dots, p$, and model parameters $\beta_1, \dots, \beta_p \in \mathbb{R}$.

(a) Explain how the Poisson GLM can be used to model Binomial/multinomial GLMs and, in particular, how the log-linear model subsumes the logit model/logistic regression.

(b) Show that if $\boldsymbol{X}$ is of full rank, then the Hessian $\mathcal{H}$ (s. II.2.23) is negative definite, so that the MLE $\hat{\boldsymbol{\beta}}$ uniquely exists and determine its asymptotic covariance matrix.

(c) Suppose that independent random variables

$$Y_{ijk} \sim \mathscr{P}(\mu_{ij}), \quad i = 1, \dots, I, \;\; j = 1, \dots, J, \;\; k = 1, \dots, n,$$

are observed. That means a balanced two-way layout of n independent observations for each level of the two factors with I or J levels, respectively. Formulate a Poisson main-effect GLM with log-link-function for $\mu_{11}, \dots, \mu_{IJ}$. Find the likelihood equations and show that μ_{ij} has fitted value $\hat{\mu}_{ij} = \frac{Y_{i++}Y_{+j+}}{nY_{+++}}$, $i = 1, \dots, I$, $j = 1, \dots, J$, where $+$ denotes a sum over all levels of the index:

$$Y_{i++} = \sum_{j=1}^{J}\sum_{k=1}^{n} Y_{ijk}, \;\; Y_{+j+} = \sum_{i=1}^{I}\sum_{k=1}^{n} Y_{ijk}, \;\; Y_{+++} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{n} Y_{ijk}.$$

Exercise 25

(a) Suppose $Y_1, \dots, Y_n$ are independent counts that satisfy a Poisson GLM for $\log\big(\mathsf{E}(Y_i)\big) = \log(\mu_i) = \sum_{j=1}^{p} \beta_j x_{ij}$, for a set of explanatory variables x_{ij}, $i = 1, \dots, n$, $j = 1, \dots, p$, and model parameters $\beta_1, \dots, \beta_p \in \mathbb{R}$.

   (i) Assume that the observed response merely indicates whether each Y_i is positive. Show that $Z_i := \mathbb{1}(Y_i > 0)$, $i = 1, \dots, n$, satisfy a binary GLM with complementary-log-log link, where $\mathbb{1}$ denotes the indicator function, i.e.,

$$\log\Big(-\log\big(1 - \mathsf{P}(Z_i = 1)\big)\Big) = \sum_{j=1}^{p} \beta_j x_{ij}.$$

(ii) Assume that $x_{i1} = 1$, $i = 1, \ldots, n$, and thus, the model is containing an intercept. Denote the MLE of β_j with $\hat{\beta}_j$, $j = 1, \ldots, p$, and the fitted mean values with $\hat{\mu}_i$, $i = 1, \ldots, n$. Show that the average rate of change in the estimated mean satisfies

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\partial \hat{\mu}_i}{\partial x_{ij}} = \hat{\beta}_j \overline{y}, \quad j = 1, \ldots, p.$$

(b) Let $Y \mid X = x \sim \mathscr{P}(x\mu)$, $\mu > 0$, be a Poisson distributed random variable conditional on the value of X and let X be a positive random variable with $\mathsf{E}(X) = 1$ and $\mathrm{Var}(X) = \tau \geq 0$. Show that $\mathsf{E}(Y) = \mu$ and $\mathrm{Var}(Y) = \mu + \tau\mu^2$.

Hint: Use the property

$$\mathrm{Var}(Y) = \mathsf{E}\big(\mathrm{Var}(Y|X)\big) + \mathrm{Var}\big(\mathsf{E}(Y|X)\big),$$

where $\mathsf{E}(Y|X)$ denotes the conditional expectation and $\mathrm{Var}(Y|X)$ denotes the conditional variance of Y given the random variable X.

(c) Let $Y \mid X = x \sim \mathscr{P}(x)$ be a Poisson distributed random variable conditional on the value of X. Further, let $X \sim \Gamma(k, \mu)$ be gamma distributed, $k, \mu > 0$. Show that the marginal distribution of Y is negative binomial. That means, if the pdf $f_{k,\mu}(\cdot)$ of X is given by

$$f_{k,\mu}(x) = \frac{\left(\frac{kx}{\mu}\right)^k}{x\Gamma(k)} \exp\left(\frac{-kx}{\mu}\right), \quad x \geq 0,$$

where $\Gamma(\cdot)$ denotes the gamma-function, it holds that

$$\mathsf{P}(Y = y) = \int_0^\infty \mathsf{P}(Y = y \,|\, X = x) f_{k,\mu}(x)\,\mathrm{d}x = \frac{\Gamma(y+k)}{\Gamma(k)} \cdot \frac{\lambda^k (1-\lambda)^y}{y!},$$

with $\lambda = \frac{k}{\mu+k}$, for each $y \in \mathbb{N}_0$.

Exercise 26

Suppose a clinical study is conducted to assess the effectiveness of some treatment. Assume that the outcome of the treatment is binary, that is either successful or not. The results of the study are summarized in the following table.

	Men		Women	
	Success	Failure	Success	Failure
Treatment	8	5	12	15
Control (no Treatment)	4	3	2	3

Investigate the following questions:

(i) Is the treatment effective for male (female) participants of the study?

(ii) Is the treatment effective if we do not condition on gender?

Explain and justify your answers.

Exercise 27

For a linear model with design matrix $X = (x_{ij})_{i=1,\ldots,n;j=1,\ldots,p}$ consider the object function

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p}\left\{\frac{1}{2}||\boldsymbol{y} - X\boldsymbol{\beta}|| + \lambda||\beta||_q^q\right\} \tag{1}$$

in Lagrangian form for some $\lambda \geq 0$, where $||\cdot||_q$ denotes the L_q norm.

Note that this results in the lasso estimator for $q = 1$ and ridge regression estimator for $q = 2$ (see Slide 216).

(a) Show that the ridge regression estimates (for $q = 2$)

$$\hat{\boldsymbol{\beta}}^{ridge} = (X'X + \lambda I_p)^{-1}X'\boldsymbol{Y}$$

can be obtained as ordinary least squares estimates on an augmented data set (and find the suitable augmentation).

Consider the special case of an orthonormal design matrix X. What do you observe?

(b) Consider the lasso estimator ($q = 1$) for a single predictor $\boldsymbol{x} \in \mathbb{R}^n$, say, and a standardized data set; that is, $\bar{\boldsymbol{y}} = \bar{\boldsymbol{x}} = 0$ and $||\boldsymbol{x}||_2 = 1$. In this case, the object function (1) simplifies to

$$\min_{\beta\in\mathbb{R}}\left\{\frac{1}{2}\sum_{i=1}^{n}(y_i - x_i\beta)^2 + \lambda|\beta|\right\}.$$

Show that the solution is given by

$$\hat{\beta}^{lasso} = \operatorname{sign}(\boldsymbol{x}'\boldsymbol{y})(\boldsymbol{x}'\boldsymbol{y} - \lambda)_+ = \begin{cases}\boldsymbol{x}'\boldsymbol{y} - \lambda, & \boldsymbol{x}'\boldsymbol{y} > \lambda,\\ 0, & |\boldsymbol{x}'\boldsymbol{y}| \leq \lambda\\ \boldsymbol{x}'\boldsymbol{y} + \lambda, & \boldsymbol{x}'\boldsymbol{y} < -\lambda.\end{cases}$$

Hint: First, show that under these assumptions the ordinary least squares estimator is given by $\hat{\beta}^{LS} = \boldsymbol{x}'\boldsymbol{y}$. Then, consider the cases $\hat{\beta}^{LS} >, \leq 0$ separately.

(c) According to (b), the lasso estimator may set a subset of parameters to exactly zero. Why is this, in general, not the case for $q > 1$?

Furthermore, for standardized data (that is, an orthonormal matrix X and a model without intercept β_0) use (b) to find the smallest value for $\lambda > 0$, such that

$$\hat{\boldsymbol{\beta}}^{lasso} = \boldsymbol{0}_p$$

all lasso estimates equal zero.

(d) Show that $||\hat{\boldsymbol{\beta}}^{lasso}||_1$ and $||\hat{\boldsymbol{\beta}}^{ridge}||_2$ increase, respectively, as the tuning parameter $\lambda \to 0$ approaches zero. What do you observe in the limit?