

POISSON REGRESSION IN ONE COVARIATE ON MASSIVE DATA

TORSTEN REUTER AND RAINER SCHWABE

ABSTRACT. The goal of subsampling is to select an informative subset of all observations, when using the full data for statistical analysis is not viable. We construct locally D -optimal subsampling designs under a Poisson regression model with a log link in one covariate. A Representation of the support of locally D -optimal subsampling designs is established. We make statements on scale-location transformations of the covariate that require a simultaneous transformation of the regression parameter. The performance of the methods is demonstrated by illustrating examples. To show the advantage of the optimal subsampling designs, we examine the efficiency of uniform random subsampling as well as of two heuristic designs. Further, the efficiency of locally D -optimal subsampling designs is studied when the parameter is misspecified.

1. INTRODUCTION

Progress in technology has lead to the collection of increasingly large data sets. The field of subsampling or subdata selection has gained popularity in recent years, where the aim is to decrease the number of observations in the data set while maintaining as much information as possible. To illuminate fundamental features of the concept, we solely focus on the reduction of observations in massive data for a single covariate, rather than reduction in covariates of high-dimensional data. Subdata selection for massive data can be done via a probabilistic subsampling scheme or through deterministic rules. Earlier works on subsampling for generalized linear models (GLMs) focus on probabilistic methods, in particular on subsampling for logistic regression, see e.g. Wang et al. (2018). More recently there are more works on GLMs, including Poisson regression: For probabilistic subsampling under the A and L -optimality criteria see Ai et al. (2021) and Yu et al. (2022). After Wang et al. (2019) introduced information-based optimal subdata selection (IBOSS) for linear regression, Cheng et al. (2020) proposed IBOSS for logistic regression, a deterministic subsampling technique with a probabilistic initial subsample to estimate the unknown parameter. This is necessary because, as is well known, the optimal design depends on the unknown parameter for GLMs.

In the present paper on Poisson regression we derive locally D -optimal continuous subsampling designs directly bounded by the density of the covariate. Such directly bounded designs were first studied by Wynn (1977) and Fedorov (1989). Recently, Ul Hassan and Miller (2019) derived such bounded optimal subsampling designs for logistic regression in the context of optimal item calibration similarly to our approach. Such subsampling designs can then easily be used for subdata selection by including all observations that lie in the support of the optimal subsampling design and exclude all others. Though an initial step to estimate the parameter is necessary when it is unknown. When there

2020 *Mathematics Subject Classification.* Primary: 62K05. Secondary: 62R07, 62J12, 62D99.
Key words and phrases. Subdata, D -optimality, Massive Data, Poisson Regression.
 Corresponding author: Torsten Reuter. *E-mail address:* torsten.reuter@ovgu.de.

are no constraints on the design, literature on Poisson regression includes Rodríguez-Torreblanca and Rodríguez-Díaz (2007) and Russell et al. (2009).

In Section 2 we introduce the Poisson regression model to be used in this paper. Then, we present a theorem on the support of a locally D -optimal continuous subsampling design as well as a theorem concerning scale-location shifts of the covariate in Section 3. Further, we give examples when the covariate has an exponential or a uniform distribution. In Section 4 we study the efficiency of uniform random subsampling and some heuristic designs in comparison to the optimal subsampling designs. In addition, we consider the loss in efficiency when the regression parameter is misspecified. We add closing remarks in Section 5. Proofs are deferred to an appendix.

2. MODEL SPECIFICATION

We consider pairs $(x_i, y_i), i = 1, \dots, n$, of data, where y_i is the value of the response variable Y_i . x_i is a realization of the random variable X_i . The covariate X_i has probability density function f_X . We suppose that the dependence of the response variable on the covariate X_i is given by a Poisson regression model.

- (A1) Conditionally on the covariate X_i , the response Y_i is Poisson distributed with conditional mean $E(Y_i|X_i) = \exp(\beta_0 + \beta_1 X_i)$.

Model (A1) constitutes a generalized linear model with random covariate and log link. The aim is to estimate the regression parameter $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$. $\mathbf{f}(x) = (1, x)^\top$ denotes the regression function in the linear component $\mathbf{f}(X_i)^\top \boldsymbol{\beta}$ such that $E(Y_i|X_i) = \exp(\mathbf{f}(X_i)^\top \boldsymbol{\beta})$.

We will further assume that the covariate X_i has a continuous distribution satisfying some moment conditions.

- (A2) The covariate X_i has density f_X and $E(X_i^2 \exp(\beta_1 X_i)) < \infty$.

3. SUBSAMPLING DESIGN

We assume that the number of observations n is very large. However, we encounter the challenge of dealing with responses, denoted by Y_i , which are either costly or difficult to observe. Meanwhile, the values x_i of all units X_i of the covariate are readily available. To tackle this problem, we consider a scenario in which the responses Y_i will only be observed for a specific subsampling proportion α of the units, $0 < \alpha < 1$. The selection of these units is based on the knowledge of the covariate values x_i for all units. Our objective is to identify a subsample of pairs (x_i, y_i) that provides the most accurate estimation of the parameter vector $\boldsymbol{\beta}$ by means of the maximum likelihood estimator $\hat{\boldsymbol{\beta}}$. As the covariate X_i has a continuous distribution, we are going to identify a subsample from this distribution that maximizes information, but only covers a percentage α of the distribution. Therefore, we consider continuous designs ξ as measures of mass α on \mathbb{R} with density f_ξ bounded by the density f_X of X_i ensuring $\int f_\xi(x) dx = \alpha$ and $f_\xi(x) \leq f_X(x)$ for all $x \in \mathbb{R}$. A subsample can then be generated according to such a bounded continuous design ξ by accepting units i with probability $f_\xi(x_i)/f_X(x_i)$. To obtain analytical results, we assume that the distribution of the covariate X_i and, hence, its density f_X is known.

The information arising for a single observation at covariate value x is defined by the elemental information $\mathbf{M}(x, \boldsymbol{\beta}) = \exp(\beta_0 + \beta_1 x) \mathbf{f}(x) \mathbf{f}(x)^\top$ (see Russell et al., 2009). For a continuous design ξ ,

the information matrix $\mathbf{M}(\xi, \boldsymbol{\beta})$ is defined by

$$\mathbf{M}(\xi, \boldsymbol{\beta}) = \int \mathbf{M}(x, \boldsymbol{\beta}) \xi(dx) = \exp(\beta_0) \begin{pmatrix} m_0(\xi, \beta_1) & m_1(\xi, \beta_1) \\ m_1(\xi, \beta_1) & m_2(\xi, \beta_1) \end{pmatrix},$$

where $m_k(\xi, \beta_1) = \int x^k \exp(\beta_1 x) f_\xi(x) dx$. The moment condition $E(X_i^2 \exp(\beta_1 X_i)) < \infty$ stated in assumption (A2) for the distribution of the covariates X_i ensures that the entries $m_k(\xi, \beta_1)$ in the information matrix are finite for any bounded continuous design ξ . Otherwise no meaningful optimization would be possible. The moment condition is obviously satisfied when the distribution of X_i has a finite support. It also holds for other not heavy-tailed distributions like the normal distribution. In the case of an exponentially distributed covariate X_i considered below, the additional condition $\beta_1 < \lambda$ on the slope parameter β_1 is required where λ is the rate parameter of the exponential distribution.

The information matrix $\mathbf{M}(\xi, \boldsymbol{\beta})$ serves as a measure for evaluating the performance of the design ξ . Note that $\mathbf{M}(\xi, \boldsymbol{\beta})$ has full rank for any continuous design ξ . This ensures the existence of the inverse

$$\mathbf{M}(\xi, \boldsymbol{\beta})^{-1} = \frac{1}{\exp(\beta_0) d(\xi, \beta_1)} \begin{pmatrix} m_2(\xi, \beta_1) & -m_1(\xi, \beta_1) \\ -m_1(\xi, \beta_1) & m_0(\xi, \beta_1) \end{pmatrix}.$$

where $d(\xi, \beta_1) = m_0(\xi, \beta_1)m_2(\xi, \beta_1) - m_1(\xi, \beta_1)^2$ is the standardized determinant of $\mathbf{M}(\xi, \boldsymbol{\beta})$, $d(\xi, \beta_1) = \exp(-2\beta_0) \det(\mathbf{M}(\xi, \boldsymbol{\beta}))$. Then, $\sqrt{\alpha n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is asymptotically normal with mean zero and covariance matrix $\mathbf{M}(\xi, \boldsymbol{\beta})^{-1}$ for the maximum likelihood estimator $\hat{\boldsymbol{\beta}}$.

Maximization of the information matrix in the Loewner sense of nonnegative definiteness will not be possible, in general. Therefore, we have to consider some one-dimensional information functional. We will focus here on the most popular design criterion, the D -criterion, in its widely used form, $\log(\det(\mathbf{M}(\xi, \boldsymbol{\beta})))$, to be maximized. A subsampling design ξ^* with density f_{ξ^*} that maximizes the D -criterion for a given parameter value $\boldsymbol{\beta}$ will be called locally D -optimal at $\boldsymbol{\beta}$. Maximization of the D -criterion can be interpreted in terms of the covariance matrix as minimization of the volume of the asymptotic confidence ellipsoid for the parameter vector $\boldsymbol{\beta}$.

Remark 3.1. Note that β_0 comes in into the information matrix only by the multiplicative factor $\exp(\beta_0)$. Thus, a locally D -optimal subsampling design ξ^* only depends on the slope β_1 .

For the characterization of a locally D -optimal design, we will make use of an equivalence theorem based on constrained convex optimization (see e.g. Sahn and Schwabe, 2001). For this, we have to distinguish between cases related to the sign of the slope β_1 . In applications, the slope will often be negative ($\beta_1 < 0$). We will focus on that case and establish a representation of the locally D -optimal subsampling designs for $\beta_1 < 0$ first.

Denote by F_X and q_α the cumulative distribution function and the α -quantile of X_i . Let $\mathbb{1}_A$ the indicator function of a set A , i.e. $\mathbb{1}_A(x) = 1$, if $x \in A$ and $\mathbb{1}_A(x) = 0$ otherwise. Further, denote by

$$\psi(x, \xi, \beta_1) = \frac{1}{d(\xi, \beta_1)} \exp(\beta_1 x) (m_0(\xi, \beta_1)x^2 - 2m_1(\xi, \beta_1)x + m_2(\xi, \beta_1))$$

the sensitivity function of a design ξ (see Theorem A.1). Note that the sensitivity function $\psi(x, \xi, \beta_1)$ does not depend on β_0 .

Theorem 3.1. *Let assumptions (A1) and (A2) be satisfied and let $\beta_1 < 0$. Then the subsampling design ξ^* is locally D -optimal at $\boldsymbol{\beta}$ if and only if ξ^* has density $f_{\xi^*}(x) = f_X(x)\mathbb{1}_{\mathcal{X}^*}(x)$ and either*

(i) *there exist $a_1 < a_2 < a_3$ such that*

$$\begin{aligned} \mathcal{X}^* &= (-\infty, a_1] \cup [a_2, a_3], \\ F_X(a_1) + F_X(a_3) - F_X(a_2) &= \alpha, \text{ and} \end{aligned} \tag{3.1a}$$

$$\psi(a_1, \xi^*, \beta_1) = \psi(a_2, \xi^*, \beta_1) = \psi(a_3, \xi^*, \beta_1), \tag{3.1b}$$

or

$$(ii) \mathcal{X}^* = (-\infty, q_\alpha], \tag{3.1a'}$$

$$\psi(x, \xi^*, \beta_1) > \psi(q_\alpha, \xi^*, \beta_1) \text{ for } x < q_\alpha, \text{ and } \psi(x, \xi^*, \beta_1) < \psi(q_\alpha, \xi^*, \beta_1) \text{ for } x > q_\alpha. \tag{3.1b'}$$

Conditions (3.1a) and (3.1a') correspond to the subsampling percentage α while (3.1b) and (3.1b') are related to the conditions on the sensitivity function in the general equivalence theorem for bounded designs (Theorem A.1) reproduced in the Appendix.

In view of the shape $f_{\xi^*}(x) = f_X(x)\mathbb{1}_{\mathcal{X}^*}(x)$ of the density of the continuous optimal subsampling designs ξ^* in Theorem 3.1, the subsampling mechanism becomes deterministic for the optimal design: The subsample can be generated by accepting all units i for which $x_i \in \mathcal{X}^*$ and by rejecting all others.

According to Theorem 3.1, there are two different scenarios for the locally D -optimal design ξ^* . Either the supporting set \mathcal{X}^* consists of two separate intervals $(-\infty, a_1]$ and $[a_2, a_3]$ (scenario (i)) or these intervals will be merged into a single one (scenario (ii)).

Remark 3.2. The optimal subsampling design ξ^* is unique because of the strict concavity of the D -criterion and the shape of the sensitivity function.

For the construction of a locally D -optimal subsampling design by Theorem 3.1, first the conditions of scenario (ii) for an optimal design supported on a single interval can be checked. If scenario (ii) does not apply, the boundary points $a_1 < a_2 < a_3$ for the support \mathcal{X}^* have to be calculated by solving the system of (nonlinear) equations (3.1a) and (3.1b). In the latter case, the rightmost boundary point a_3 of \mathcal{X}^* may lie outside the support of X_i , i. e. $a_3 > x_{\max}$, when the support of the covariate X_i is bounded from above, i. e. $x_{\max} = \text{ess sup}(X_i) < \infty$, where ess sup denotes the essential supremum (see, e. g., Example 3.2 for the uniform distribution below). Then, in scenario (ii), explicit calculation of the rightmost boundary point c is not necessary. Instead, it is sufficient for (3.1b) to verify that $\psi(x_{\max}, \xi^*, \beta_1) \geq \psi(a_1, \xi^*, \beta_1) = \psi(a_2, \xi^*, \beta_1)$.

Remark 3.3. The leftmost boundary point a_1 of a D -optimal subsampling design ξ^* cannot lie outside the range of X_i , i. e. $a_1 > x_{\min}$, where $x_{\min} = \text{ess inf}(X_i)$ the essential infimum of the distribution of X_i .

Remark 3.4. When $\beta_1 = 0$, the information matrix $\mathbf{M}(\xi, \beta)$ is, up to the multiplicative constant $\exp(\beta_0)$, equal to the information matrix $\mathbf{M}(\xi) = \int \mathbf{f}(x)\mathbf{f}(x)^\top \xi(dx)$ in the linear model (treated in Reuter and Schwabe, 2023). Therefore, the D -optimal subsampling design for ordinary linear regression is also locally D -optimal in the Poisson regression model. Hence, according to (Reuter and Schwabe, 2023, Section 4), the subsampling design ξ^* is locally D -optimal for $\beta_1 = 0$ if and only if there exist $a_1 < a_2$ such that

$$\begin{aligned} f_{\xi^*}(x) &= f_X(x)\mathbb{1}_{(-\infty, a_1] \cup [a_2, \infty)}(x), \\ F_X(a_2) - F_X(a_1) &= 1 - \alpha, \text{ and} \\ \psi(a_1, \xi^*, \beta_1) &= \psi(a_2, \xi^*, \beta_1). \end{aligned}$$

By means of equivariance considerations, we may transfer a locally D -optimal subsampling design ξ^* for a covariate X_i to a location-scale transformed covariate $Z_i = aX_i + b$, $a \neq 0$. However, the transformation of a locally D -optimal subsampling design is not as straightforward as in polynomial regression (see Reuter and Schwabe, 2023), but requires a simultaneous transformation of the slope parameter β_1 . This kind of simultaneous transformation typically has to be used in generalizes linear models where the elemental information depends on β_1 by the linear component $\mathbf{f}(x^\top)\beta_1$, see e. g. Radloff and Schwabe (2016).

Theorem 3.2. *Let ξ^* be a locally D -optimal subsampling design at β_1 for a covariate X_i with density f_X . Then, for a covariate Z_i with density $f_Z(z) = \frac{1}{|a|}f_X(\frac{z-b}{a})$, the design ζ^* with density $f_{\zeta^*}(z) = \frac{1}{|a|}f_{\xi^*}(\frac{z-b}{a})$ is locally D -optimal at the transformed parameter β_1/a .*

For $a = -1$, Theorem 3.2 covers sign change. Then we can transfer the characterization of a locally D -optimal subsampling design in the equivalence theorem (Theorem 3.1) to positive values for the slope β_1 .

Corollary 3.3. *Let $\beta_1 > 0$. Then the subsampling design ξ^* is locally D -optimal at β if and only if $f_{\xi^*} = f_X \mathbb{1}_{\mathcal{X}^*}$ and either*

(i) *there exist $a_1 < a_2 < a_3$ such that*

$$\mathcal{X}^* = [a_1, a_2] \cup [a_3, \infty),$$

$$F_X(a_1) + F_X(a_3) - F_X(a_2) = 1 - \alpha, \text{ and}$$

$$\psi(a_1, \xi^*, \beta_1) = \psi(a_2, \xi^*, \beta_1) = \psi(a_3, \xi^*, \beta_1),$$

or

(ii) $\mathcal{X}^* = [q_{1-\alpha}, \infty)$,

$$\psi(x, \xi^*, \beta_1) < \psi(q_{1-\alpha}, \xi^*, \beta_1) \text{ for } x < q_\alpha, \text{ and } \psi(x, \xi^*, \beta_1) > \psi(q_\alpha, \xi^*, \beta_1) \text{ for } x > q_\alpha.$$

To illustrate how the equivalence theorem (Theorem 3.1) can be used to construct locally D -optimal subsampling designs, we consider $\beta_1 < 0$ in the situation of an exponentially and of a uniformly distributed covariate in the following two examples.

Example 3.1 (exponential distribution). We assume the covariate X_i to follow an exponential distribution with rate λ , i. e. X_i has density $f_X(x) = \lambda \exp(-\lambda x)$ for $x \geq 0$. The condition of finite moments $m_k(\xi, \beta_1)$ is satisfied for $\beta_1 < \lambda$ and hence, in particular, for $\beta_1 \leq 0$. For $\beta_1 < 0$, let

$$g_0(t) = \frac{\lambda}{\lambda - \beta_1} \exp(-(\lambda - \beta_1)t), \quad g_1(t) = \left(t + \frac{1}{\lambda - \beta_1}\right) g_0(t) \text{ and } g_2(t) = t^2 g_0(t) + \frac{2}{\lambda - \beta_1} g_1(t)$$

such that $g_k(t) = \int_t^\infty x^k \exp(\beta_1 x) f_X(x) dx$, $t \geq 0$. Then, in scenario (i), the entries in $\mathbf{M}(\xi^*, \beta)$ are

$$m_k(\xi^*, \beta_1) = g_k(0) - g_k(a_1) + g_k(a_2) - g_k(a_3), \quad k = 0, 1, 2,$$

while they reduce to $m_k(\xi^*, \beta_1) = g_k(0) - g_k(q_\alpha)$ in scenario (ii) when there is only one interval, where $q_\alpha = -\log(1 - \alpha)/\lambda$.

In scenario (i), we obtain numerical results for the boundary points a_1 to a_3 solving the system of equations (3.1a) and (3.1b) using the Newton method implemented in the **R** package *nleqslv* by Hasselman (2018). Note that here $a_3 < x_{\max} = \infty$. For the case of a standard exponential distribution ($\lambda = 1$), results are given in Table 1 for selected values of β_1 and α . In addition, we give the values for

the amount $F_X(a_1)$ as well as the percentage of mass the design ξ^* places on the left interval $[0, a_1]$. We also add the result for $\beta_1 = 0$ for reference (see Reuter and Schwabe, 2023).

TABLE 1. Numerical values for the boundary points a_1 , a_2 , a_3 , and q_α , respectively, for selected values of the subsampling proportion α and slope parameter β_1 in the case of a standard exponentially distributed covariate ($\lambda = 1$)

α	β_1	a_1	a_2	a_3, q_α	$F_X(a_1)$	% of mass on $[0, a_1]$
0.01	0.0	0.00579	5.46588	-	0.00577	57.71
	-0.5	0.00501	3.86767	4.14130	0.00500	49.95
	-1.0	0.00500	1.98399	2.02112	0.00499	49.88
	-4.0	0.00496	0.49830	0.50665	0.00495	49.51
0.10	0.0	0.06343	3.25596	-	0.06146	61.46
	-0.5	0.05181	2.92225	5.44835	0.05049	50.49
	-1.0	0.05011	1.83717	2.22435	0.04887	48.87
	-4.0	0.04680	0.47740	0.56896	0.04572	45.72
0.30	0.0	0.21398	2.23153	-	0.19264	64.21
	-0.5	0.17225	1.95006	7.60885	0.15823	52.74
	-1.0	0.15317	1.50902	2.76234	0.14202	47.34
	-4.0	0.12876	0.40855	0.72273	0.12081	40.27
0.75	0.0	0.67278	1.34596	-	0.48971	65.29
	-0.5	0.52804	1.07947	10.89214	0.41024	54.70
	-1.0	0.43176	0.88401	4.28609	0.35063	46.75
	-4.0	-	-	1.38629	-	-

For other values of the rate λ , results can be derived from the case of a standard exponentially distributed covariate via equivariance (Theorem 3.2) by letting $a = 1/\lambda$ and $b = 0$: If we seek a locally D -optimal subsampling design at $\beta_1 < 0$ when the rate is λ , we can first construct a locally D -optimal design at β_1/λ for a standard exponentially distributed covariate and then divide the obtained boundary points by λ . For example, when $\lambda = 2$, $\beta_1 = -1$ and the subsampling proportion is $\alpha = 0.10$, we get the boundary points $0.05181/2$, $2.92225/2$, and $5.44835/2$ from the second line highlighted in the second block of Table 1 such that the locally D -optimal subsampling design wanted is supported on the two intervals $[0, 0.0259]$ and $[1.4611, 2.7242]$.

When the subsampling proportion α goes to zero, the locally D -optimal subsampling design apparently tends to its counterpart in classical optimal design theory which assigns equal weight $1/2$ to two support points $x_1^* = 0$ and $x_2^* = -2/\beta_1$ (see e. g. Rodríguez-Torreblanca and Rodríguez-Díaz, 2007). In particular, we observe $a_2 < x_2^* < a_3$ for all numerically obtained values of a_2 and a_3 .

On the contrary, we find that scenario (ii) appears for large values of α . This happens when the slope β_1 is strongly negative. More precisely, given α , there is a crossover point β_1^* such that the single interval design with density $f_{\xi^*} = f_X \mathbb{1}_{[0, q_\alpha]}$ is locally D -optimal at β_1 for all $\beta_1 \geq \beta_1^*$. This crossover point becomes stronger negative when α gets smaller and apparently tends to $-\infty$ as $\alpha \rightarrow 0$. On the other hand, when α gets larger, the crossover point apparently tends to zero. In Table 2, we give numerical results for the crossover point β_1^*/λ for selected values of α together with the quantile q_α , the setting x_2^* of the locally D -optimal unbounded design and their ratio. This shows that, for scenario (ii)

to apply, the quantile q_α has to be substantially larger than x_2^* . Vice versa, for given slope $\beta_1 < 0$, there is a critical subsampling proportion α^* such that the single interval design is locally D -optimal for larger subsampling proportions $\alpha \geq \alpha^*$. In particular, when $\beta_1 = 0$, only scenario (i) applies (see Reuter and Schwabe, 2023) and, hence, $\alpha^* = 1$.

We further notice that the percentage of mass on the left interval $[0, a_1]$ is generally larger than 50% for β_1 closer to zero which coincides with what we have seen in Reuter and Schwabe (2023) for the case $\beta_1 = 0$. There, observations from the right tail are more informative and thus more observations are needed on the left tail. Conversely, the percentage of mass on $[0, a_1]$ is smaller than 50% for strongly negative β_1 . Figure 1 depicts the locally D -optimal subsampling designs for $\alpha = 0.5, 0.9$ and $\beta_1 = -1$ along with the corresponding sensitivity functions. The horizontal dotted line represents the threshold s^* from Theorem A.1. The vertical dotted lines depict the boundary points. While smaller subsampling proportions $\alpha \leq 0.1$ are typically of interest in the context of subsampling, our selection of larger subsampling proportions α has been made for the sake of clarity and visibility in the tables and figures.

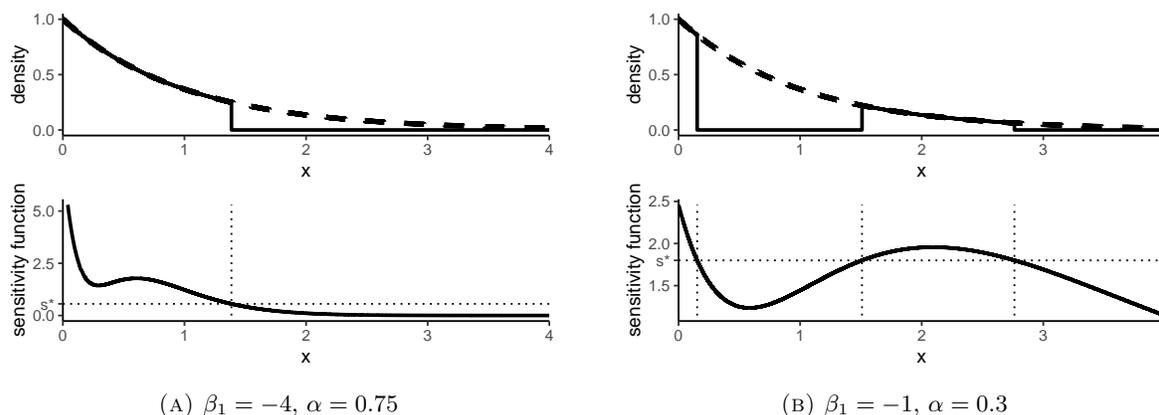


FIGURE 1. Density of the locally optimal design (solid) at β_1 and the standard exponential distribution (dashed, upper panels), and corresponding sensitivity functions (lower panels) for $\beta_1 = -4, \alpha = 0.75$ (left) and $\beta_1 = -1, \alpha = 0.3$ (right)

TABLE 2. Numerical values for the standardized crossover point β_1^*/λ for an exponentially distributed covariate

α	β_1^*/λ	λq_α	λx_2^*	q_α/x_2^*
0.01	-360.34840	0.01005	0.00556	1.81081
0.10	-34.60684	0.10536	0.05779	1.82310
0.30	-10.41165	0.35667	0.19209	1.85679
0.50	-5.49454	0.69314	0.36400	1.90426
0.75	-2.89534	1.38629	0.69077	2.00690
0.90	-1.86128	2.30259	1.07453	2.14288

Example 3.2 (uniform distribution). We assume the covariate to be uniform random on an interval $[x_{\min}, x_{\max}]$ with density $f_X(x) = \frac{1}{x_{\max} - x_{\min}} \mathbb{1}_{[x_{\min}, x_{\max}]}(x)$. The condition of finite moments $m_k(\xi, \beta_1)$ is satisfied for all β_1 .

For $\beta_1 < 0$, let

$$g_0(t) = \frac{\exp(\beta_1 t)}{|\beta_1|(x_{\max} - x_{\min})}, \quad g_1(t) = \left(t + \frac{1}{|\beta_1|}\right) g_0(t) \quad \text{and} \quad g_2(t) = t^2 g_0(t) + \frac{2}{|\beta_1|} g_1(t).$$

In scenario (i), unlike in Example 3.1, the support of the covariate is bounded from above and thus the rightmost boundary point a_3 may be larger than x_{\max} . We denote the essential supremum of ξ^* by $\tilde{a}_3 = \min(a_3, x_{\max})$. Then, in scenario (i), the entries in $\mathbf{M}(\xi^*, \beta)$ are

$$m_k(\xi^*, \beta_1) = g_k(x_{\min}) - g_k(a_1) + g_k(a_2) - g_k(\tilde{a}_3), \quad k = 0, 1, 2,$$

while in scenario (ii), when there is only one interval, they reduce to $m_k(\xi^*, \beta_1) = g_k(x_{\min}) - g_k(q_\alpha)$ where $q_\alpha = (1 - \alpha)x_{\min} + \alpha x_{\max}$.

For the case of a uniform distribution on the unit interval ($x_{\min} = 0$ and $x_{\max} = 1$), optimal boundary points are given in Table 3 for selected values of α and $\beta_1 < 0$. In addition, we give the values for the amount $F_X(a_1)$ as well as the percentage of mass the design ξ^* places on the left interval $[0, a_1]$. We also add formally the result for $\beta_1 = 0$ for reference (see Reuter and Schwabe, 2023).

TABLE 3. Numerical values for the boundary points a_1, a_2, a_3 and q_α , respectively, for selected values of the subsampling proportion α and slope parameter β_1 in the case of a uniformly distributed covariate on $[0, 1]$

α	β_1	a_1	a_2	a_3, q_α	$F_X(a_1)$	% of mass on $[0, a_1]$
0.01	0	0.00500	0.99500	-	0.00500	50.00
	-2	0.00498	0.99498	-	0.00498	49.75
	-4	0.00495	0.49994	0.50499	0.00495	49.51
	-8	0.00490	0.24989	0.25498	0.00490	49.04
0.10	0	0.05000	0.9500	-	0.05000	50.00
	-2	0.04772	0.94772	-	0.04772	47.72
	-4	0.04578	0.49506	0.54928	0.04578	45.78
	-8	0.04269	0.24155	0.29887	0.04269	42.69
0.30	0	0.15000	0.8500	-	0.15000	50.00
	-2	0.13271	0.83271	-	0.13271	44.24
	-4	0.12102	0.46678	0.64577	0.12102	40.34
	-8	0.10847	0.20165	0.39318	0.10847	36.16
0.50	0	0.25000	0.7500	-	0.25000	50.00
	-2	0.20993	0.70993	-	0.20993	41.99
	-4	0.18578	0.42624	0.74046	0.18578	37.16
	-8	-	-	0.50000	-	-

Apart from the situation that $a_3 > x_{\max}$ indicated by a hyphen (-) in the table when $\alpha = 0.5$ and $\beta_1 = -2$, the results are similar to those in Example 3.1: More weight is given to the left interval $[0, a]$ when β_1 is closer to zero. When the subsampling proportion α becomes small, the locally D -optimal subsampling design approaches the locally D -optimal unbounded design equally supported on $x_1^* = 0$ and $x_2^* = -2/\beta_1$. For large values of α , the two intervals are merged into one (e.g. for $\alpha = 0.50$ and $\beta_1 = -8$). Figure 2 depicts the locally D -optimal subsampling designs along the corresponding sensitivity functions in scenario (ii) of a single supporting interval for ξ^* in the left panel. The right

panel exhibits scenario (i) of ξ^* supported on two proper intervals. The horizontal dotted line depicts the threshold s^* . The vertical dotted lines represent the boundary points a_1 , a_2 , and a_3 . The situation when $a_3 > x_{\max}$ is displayed in Figure 3.

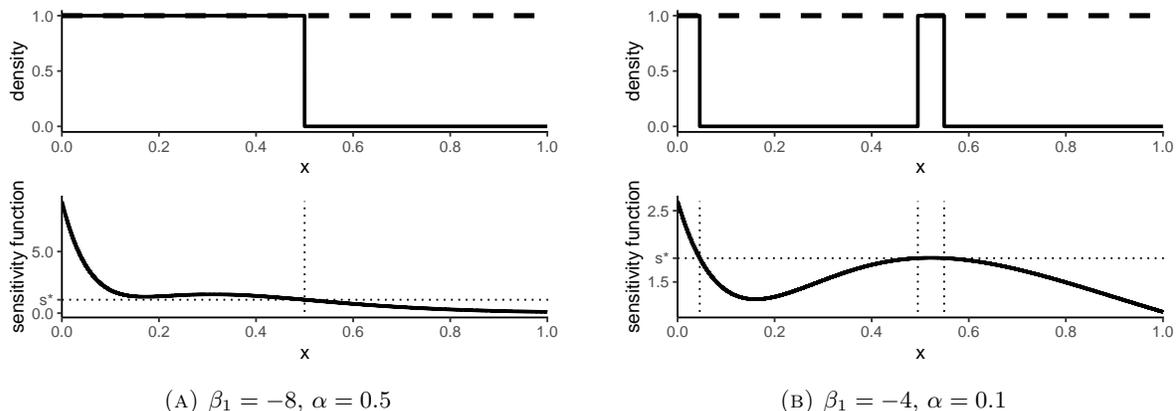


FIGURE 2. Density of the locally optimal design (solid) at β_1 for a uniformly distributed covariate on $[0, 1]$ (dashed, upper panels), and sensitivity functions (lower panels) for $\beta_1 = -8, \alpha = 0.5$ (left) and $\beta_1 = -4, \alpha = 0.1$ (right)

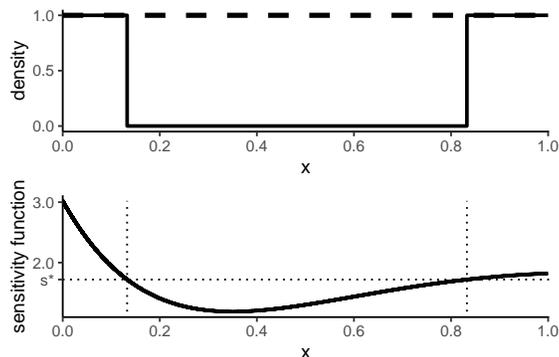


FIGURE 3. Density of the locally optimal design (solid) at β_1 for a uniformly distributed covariate on $[0, 1]$ (dashed, upper panel), and sensitivity functions (lower panel) for $\beta_1 = -2, \alpha = 0.3$

Because of the symmetry of the uniform distribution, locally D -optimal subsampling designs can be derived for positive values of the slope β_1 via equivariance with respect to sign change by letting $a = -1$ and $b = 1$ in Theorem 3.2. For example, when $\beta_1 = 4$ and $\alpha = 0.10$, the optimal boundary points can be obtained from the third line highlighted in the second block of Table 3 as $1 - 0.04578$, $1 - 0.49506$, and $1 - 0.54928$ such that the locally D -optimal subsampling design is then supported on the two intervals $[0.45072, 0.50494]$ and $[0.95422, 1]$.

Further, for other ranges $[x_{\min}, x_{\max}]$ of the uniform covariate, optimal subsampling designs can be obtained by equivariance (Theorem 3.2) as well by letting $a = x_{\max} - x_{\min}$ and $b = x_{\min}$.

4. EFFICIENCY

We want to study the performance of random subsampling as well as some heuristic subsampling designs in the style of IBOSS (see Wang et al., 2019) to quantify the gain in using a locally D -optimal subsampling design. Besides, we are interested in the quality of the heuristic designs and how they compare to random subsampling. Further, we want to investigate the performance of designs when the parameter is misspecified. Specifically, a subsampling design $\xi^*(\beta') = \arg \max \det(\mathbf{M}(\xi, \beta'))$ that is locally D -optimal at β' is studied when the true parameter is β . The performance of a design ξ may be compared to the locally D -optimal subsampling design $\xi^*(\beta)$ using D -efficiency. The D -efficiency of a subsampling design ξ with mass α is defined as

$$\text{eff}_{D,\alpha}(\xi, \beta) = \left(\frac{\det(\mathbf{M}(\xi, \beta))}{\det(\mathbf{M}(\xi^*(\beta), \beta))} \right)^{1/2}.$$

For this definition the homogeneous version $(\det(\mathbf{M}(\xi, \beta)))^{1/2}$ of the D -criterion is used which satisfies the homogeneity condition $(\det(\nu\mathbf{M}))^{1/2} = \nu(\det(\mathbf{M}))^{1/2}$ for all $\nu > 0$ (see Pukelsheim, 1993, Chapter 6.2). Note that by Remark 3.1, the efficiency $\text{eff}_{D,\alpha}(\xi, \beta)$ does not depend on β_0 .

As uniform random subsampling we define the design ξ_α of size α , which has density $f_{\xi_\alpha}(x) = \alpha f_X(x)$. The information matrix of ξ_α is given by $\mathbf{M}(\xi_\alpha, \beta) = \alpha \mathbf{M}(\xi_1, \beta)$. Here, ξ_1 represents the full sample with information matrix $\mathbf{M}(\xi_1, \beta) = \int \exp(\beta_0 + \beta_1 x) \mathbf{f}(x) \mathbf{f}(x)^\top f_X(x) dx$. Thus, the D -efficiency $\text{eff}_{D,\alpha}(\xi_\alpha, \beta)$ of uniform random subsampling can be nicely interpreted as noted in Reuter and Schwabe (2023): for a fixed full sample size n , the required subsample size (mass) $\tilde{\alpha}$ needed to achieve the same precision (measured by the D -criterion), compared to utilizing a locally D -optimal subsampling design ξ^* with mass α , is given by the inverse of the efficiency, $\text{eff}_{D,\alpha}(\xi_\alpha, \beta)^{-1}$, multiplied by α , i. e. $\tilde{\alpha} = \alpha / \text{eff}_{D,\alpha}(\xi_\alpha, \beta)$. For instance, if the efficiency $\text{eff}_{D,\alpha}(\xi_\alpha, \beta)$ equals 0.5, then twice the number of observations would be needed under uniform random sampling compared to a locally D -optimal subsampling design of mass α . Naturally, the full sample has higher information than any proper subsample such that, for uniform random subsampling, $\text{eff}_{D,\alpha}(\xi_\alpha, \beta) \geq \alpha$ holds for all α .

Further, we analyze the efficiency of two heuristic designs. Again we only consider the case $\beta_1 < 1$. Let the α -quantile of the covariate X_i be denoted by q_α . First, we consider the one-sided design ξ_{os} with density $f_{\xi_{os}}(x) = f_X(x) \mathbb{1}_{(-\infty, q_\alpha]}(x)$ that assigns all of its mass on the left tail of the distribution of the covariate motivated by its optimality for large α . Second, we study the two-sided design ξ_{ts} with density $f_{\xi_{ts}}(x) = f_X(x) \mathbb{1}_{(-\infty, q_{\alpha/2}] \cup [q_{1-\alpha/2}, \infty)}(x)$ that allocates equal mass $\alpha/2$ on both tails of the distribution in the style of the IBOSS method (see Wang et al., 2019).

Example 4.1 (exponential distribution). As in Example 3.1, we assume that the covariate X_i is exponentially distributed with rate λ .

Because uniform random subsampling ξ_α as well as the one- and two-sided designs ξ_{os} and ξ_{ts} are equivariant under location-scale transformations, their efficiency depends only on the slope and the rate by the ratio β_1/λ . In Figure 4, we depict the efficiency of these designs for $\beta_1/\lambda = -1$ and -4 in dependence on the subsampling proportion α . The efficiency of uniform random subsampling is quite low for reasonable proportions $\alpha \leq 0.1$ and, hence, the gain in using the D -optimal subsampling design is substantial. Similarly, the efficiency of the one- and the two-sided design is small for $\alpha \leq 0.1$ and apparently tends to zero for $\alpha \rightarrow 0$ which may be explained by the fact that these designs miss

observations close to the location x_2^* of the locally D -optimal unbounded design. This feature does not apply to uniform random subsampling such that, for very small subsampling proportions, both the one- and the two-sided design is severely less efficient than uniform random subsampling.

As is to be expected, the two-sided IBOSS-like design ξ_{ts} performs much better for β_1 near zero. In particular, for $\beta_1 = 0$, the two-sided design ξ_{ts} only differs slightly from the locally D -optimal subsampling design is ξ^* and has a high efficiency throughout (see Reuter and Schwabe, 2023). Conversely, the one-sided design ξ_{os} performs better for strongly negative β_1 . The vertical dotted line in Figure 4 displays the crossover point α^* . For all $\alpha > \alpha^*$, the one-sided design is the D -optimal subsampling design.

We observe similar behavior in Figure 5. Predictably, the one-sided design performs better for strongly negative β_1 and the two-sided design is better for β_1 closer to zero. Notably, the two-sided design exhibits a nonmonotonic behavior: It performs worst for $\beta_1/\lambda \approx -3.64$ ($\text{eff}_{D,\alpha}(\xi_{ts}, \beta) = 0.07974506$) and attains a local maximum at $\beta_1/\lambda \approx -0.40$ ($\text{eff}_{D,\alpha}(\xi_{ts}, \beta) = 0.9988009$). Further, we again see that uniform subsampling generally performs better for β_1 closer to zero, though it performs best for $\beta_1/\lambda \approx -1.05$ ($\text{eff}_{D,\alpha}(\xi_{ts}, \beta) = 0.6978610$).

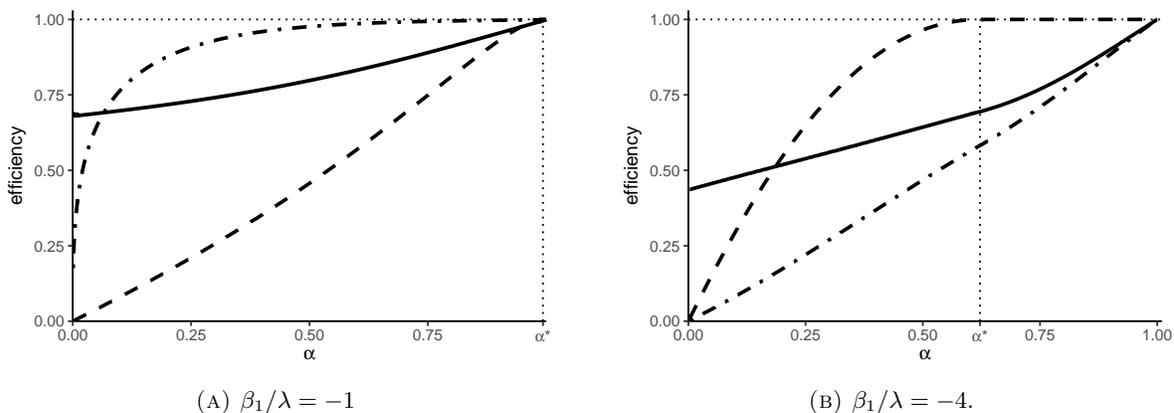


FIGURE 4. D -efficiency of uniform random subsampling (solid), one-sided (dashed), and two-sided (dot-dashed) subsampling design in dependence on the subsampling proportion α for slope-rate ratio $\beta_1/\lambda = -1$ (left) and -4 (right) for an exponentially distributed covariate

For strongly negative β_1 , the behavior of the efficiency of the three designs in Figure 5 gives additional insight. As $\beta_1 \rightarrow -\infty$, the efficiency of uniform random subsampling converges to its lower bound α whereas the efficiency of both one- and two-sided design converge to one. Most of the information is concentrated on the covariate values close to zero. Thus, for strongly negative β_1 the two heuristic designs as well as the D -optimal subsampling design have almost all the information of the full sample. This limiting behavior is not presented in Figure 5 in order to preserve intelligibility for β_1 closer to zero.

Finally, we consider the efficiency of locally D -optimal subsampling designs $\xi^*(\beta')$, when the nominal value β'_1 is misspecified and differs from the true slope β_1 . The left panel of Figure 6 illustrates the efficiency of $\xi^*(\beta')$ in dependence on the subsampling proportion α for selected values of the true ratio β_1/λ , when the nominal value is $\beta'_1/\lambda = -1$. For all values we find that the efficiency of the design

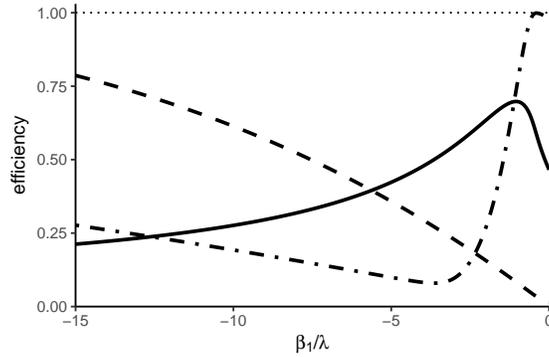
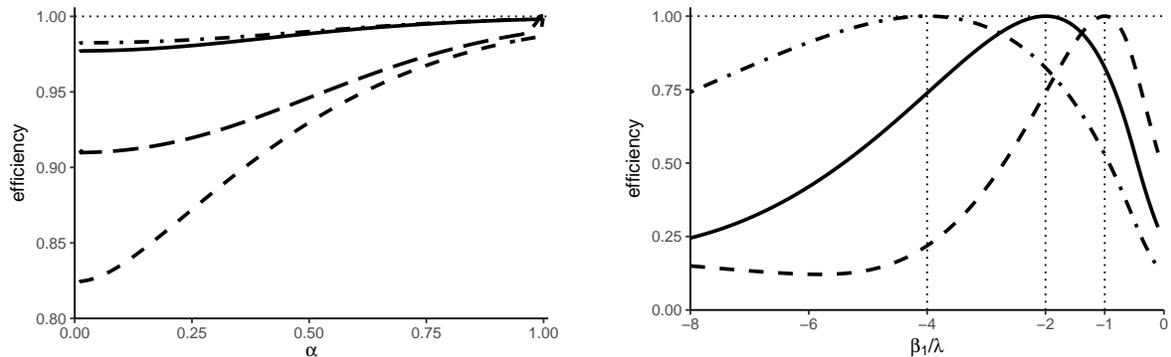


FIGURE 5. D -efficiency of uniform random subsampling (solid), one-sided (dashed), and two-sided (dot-dashed) subsampling design in dependence on the slope-rate ratio β_1/λ for subsampling proportion $\alpha = 0.1$ and an exponentially distributed covariate

$\xi^*(\beta')$ under misspecification declines with decreasing α . When the deviation of the parameter is rather small, $\beta_1/\lambda = -0.8$ and $\beta_1/\lambda = -1.2$, the designs under misspecification are still very efficient, with efficiency above 0.98 for $\alpha = 0.01$. For larger deviations however, the efficiency can drop drastically. In particular, when β_1/λ is closer to 0, the efficiency is more strongly negatively affected than when the deviation of β_1/λ is away from zero. In the right panel of Figure 6, we exhibit the efficiency for various values of the nominal slope-rate ratio in dependence on the true value when the subsampling proportion is $\alpha = 0.1$. The nominal values are indicated by vertical dotted lines.

It can be seen that the efficiency decreases faster for β_1/λ towards zero than for stronger negative values. In particular, the efficiency increases again when β_1/λ goes to $-\infty$.



(A) True parameter $\beta_1/\lambda = -0.5$ (dashed), -0.8 (solid), -1.2 (dot-dashed), and -1.5 (long dashed) (B) Locally D -optimal subsampling designs for $\beta_1'/\lambda = -1$ (dashed), -2 (solid), and -4 (dot-dashed)

FIGURE 6. Efficiency of the locally D -optimal subsampling design for $\beta_1'/\lambda = -1$ and various subsampling proportions α (left) and for subsampling proportions $\alpha = 0.1$ and various values of the nominal slope-rate ratio β_1'/λ (right) in dependence on the true slope-rate ratio β_1/λ for an exponentially distributed covariate

5. CONCLUDING REMARKS

Our investigation centers on a theoretical approach to evaluate subsampling designs under distributional assumptions on the covariate in the case of Poisson regression on a single covariate. We adjust a standard equivalence theorem to Poisson regression, given a general distribution of the covariate and negative slope parameter β_1 . This equivalence theorem also characterizes the support of the locally D -optimal subsampling design and allows us to derive such designs for a given covariate and slope parameter. Then, we establish a theorem to identify locally D -optimal subsampling designs under a scale-location transformation of the covariate and simultaneous rescaling of the slope parameter. We make use of this to give a corollary to the equivalence theorem for $\beta_1 > 0$. It is worthwhile noting that many of the results can be extended from D -optimality to other criteria within Kiefer's class of Φ_q -optimality criteria, including, in particular, linear criteria. The derivation relies mostly on the fact that the sensitivity function can be factorized into the exponential function and a quadratic polynomial, rather than its specific form. Our efficiency analysis shows, among other things, that heuristic one- or two-sided designs can be highly efficient under certain circumstances, however, they display substantial loss in efficiency for the most relevant small subsampling proportions. Addressing uncertainty about the parameter β_1 and the covariate distribution may involve an initial random subsampling step, before deploying the locally D -optimal subsampling design. Lastly, note that the results presented here may be extended to polynomial Poisson regression, where the linear predictor is a polynomial of degree q in the covariate X_i . Then, the equation $\psi(x, \xi, \beta) = s$ has at most $2q + 1$ solutions and the support of ξ^* is the union of at most $q + 1$ intervals.

ACKNOWLEDGMENTS

The work of the first author is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 314838170, GRK 2297 MathCoRe.

APPENDIX A. PROOFS

Before we establish the equivalence theorem (Theorem 3.1), we introduce some technical tools: The directional derivative of the D -criterion at design ξ in the direction of a design η is $\Psi(\xi, \eta, \beta) = \text{trace}(\mathbf{M}(\xi, \beta)^{-1}\mathbf{M}(\eta, \beta)) - 2$. Here, η may be any design of total mass α which is not necessarily required to have a density bounded by f_X . The sensitivity function $\psi(x, \xi, \beta) = \text{trace}(\mathbf{M}(\xi, \beta)^{-1}\mathbf{M}(\xi_x, \beta))$ is the essential part of the directional derivative at ξ in the direction of a single point design ξ_x with all mass α at point x . Then

$$\begin{aligned} \psi(x, \xi, \beta) &= \alpha \exp(\beta_0 + \beta_1 x) \mathbf{f}(x)^\top \mathbf{M}(\xi, \beta)^{-1} \mathbf{f}(x) \\ &= \frac{\alpha}{d(\xi, \beta_1)} \exp(\beta_1 x) (m_0(\xi, \beta_1) x^2 - 2m_1(\xi, \beta_1) x + m_2(\xi, \beta_1)) \end{aligned}$$

does not depend on β_0 and will be denoted by $\psi(x, \xi, \beta_1)$, for short. Note that, for any continuous subsampling design ξ , the information matrix $\mathbf{M}(\xi, \beta)$ is positive definite and, hence, $\psi(x, \xi, \beta_1)$ is well-defined.

For convenience, we reproduce an equivalence theorem for subsampling designs in a general model context which follows from Corollary 1(c) in Sahm and Schwabe (2001).

Theorem A.1. *Let condition*

(A) $P(\psi(X_i, \xi, \beta_1) = s) = 0$ *for any* ξ *and* s

be satisfied. Then the subsampling design ξ^* *is locally* D -*optimal at* β *if and only if there exist a set* \mathcal{X}^* *and a threshold* s^* *such that*

- (i) ξ^* *has density* $f_{\xi^*}(x) = f_X(x)\mathbb{1}_{\mathcal{X}^*}(x)$
- (ii) $\psi(x, \xi^*, \beta_1) \geq s^*$ *for* $x \in \mathcal{X}^*$, *and*
- (iii) $\psi(x, \xi^*, \beta_1) < s^*$ *for* $x \notin \mathcal{X}^*$.

Next we establish that condition (A) holds for the Poisson regression model.

Lemma A.2. *Given* ξ *and* s , *the equation* $\psi(x, \xi, \beta_1) = s$ *has, at most, three different solutions in* x .

Proof. For $\beta_1 = 0$, the sensitivity function is a quadratic polynomial in x . Hence, there are, at most, two solutions.

For $\beta_1 \neq 0$, the sensitivity function $\psi(x, \xi, \beta) = \exp(\beta_1 x)q(x)$ factorizes into the exponential function ($\exp(\beta_1 x)$) and a quadratic polynomial q with positive leading term. Because $\psi(x, \xi, \beta_1)$ is positive, only $s > 0$ has to be considered. Let $v(x) = q(x) - s \exp(-\beta_1 x)$. The third derivative $v^{(3)}(x) = s\beta_1^3 \exp(-\beta_1 x)$ has no roots. By iterative application of the mean value theorem, we see that v has, at most, three roots. Because the solutions of $\psi(x, \xi, \beta_1) = s$ are the roots of v , this completes the proof. \square

Condition (A) follows from the continuous distribution of the covariate X_i .

Proof of Theorem 3.1. If ξ^* is locally D -optimal, then, by Theorem A.1, its density has the shape $f_{\xi} = f_X \mathbb{1}_{\mathcal{X}}$ and $\mathcal{X}^* = \{x; \psi(x, \xi^*, \beta_1) \geq s^*\}$ for some $s^* > 0$. Because $\beta_1 < 0$, the sensitivity function $\psi(x, \xi^*, \beta_1)$ ranges from ∞ for $x \rightarrow -\infty$ to 0 for $x \rightarrow \infty$ with $\psi(x, \xi^*, \beta_1) > 0$ throughout. Thus, the number of sign changes in $\psi(x, \xi^*, \beta_1) - s^*$ is odd and, by Lemma A.2, equal to one or three. Hence, \mathcal{X}^* consists of one or two intervals including a left open interval $(-\infty, a_1]$, say, and potentially a second finite interval $[a_2, a_3]$. Conditions (3.1a) and (3.1a'), respectively, follow from the subsampling percentage α . If there are two intervals, then $\psi(a_k, \xi^*, \beta_1) = s^*$, $k = 1, 2, 3$, by continuity of the sensitivity function and we get condition (3.1b) in scenario (i). If there is only one interval, then condition (3.1b') follows from (ii) and (iii) in Theorem A.1 which completes the proof that the locally D -optimal subsampling design satisfies the properties stated in Theorem 3.1.

Conversely, by the shape of the sensitivity function, the properties stated in Theorem 3.1 imply the equivalence conditions in Theorem A.1 which proves the reverse statement. \square

Proof of Remark 3.3. Assume $a_1 \leq x_{\min}$. Then

$$m_1(\xi^*, \beta_1) = \int_{a_2}^{a_3} x \exp(\beta_1 x) f_X(x) dx > a_2 \int_{a_2}^{a_3} \exp(\beta_1 x) f_X(x) dx = a_2 m_0(\xi^*, \beta_1)$$

and q attains its minimum at $m_1(\xi^*, \beta_1)/m_0(\xi^*, \beta_1) > a_2$. Hence, the sensitivity function $\psi(x, \xi^*, \beta_1) = \exp(\beta_1 x)q(x)$ is strictly decreasing on $(-\infty, a_2]$ such that $\psi(a_1, \xi^*, \beta_1) > \psi(a_2, \xi^*, \beta_1)$ which leads to a contradiction to the optimality condition (3.1b). \square

Proof of Theorem 3.2. The proof goes along the same lines as in Radloff and Schwabe (2016). Denote by g the location-scale transformation $g(x) = ax + b$. Let $Z_i = g(X_i)$. Note that only the distribution of the covariate plays a role, but not the covariate itself. The transformation g is conformable

with the regression function $\mathbf{f}(x)$, i. e. there exists a nonsingular matrix $\mathbf{Q} = \begin{pmatrix} 1 & 0 \\ b & a \end{pmatrix}$ such that $\mathbf{f}(ax + b) = \mathbf{Q}\mathbf{f}(x)$ for all x . For a design ξ bounded by f_X , we define the transformed design $\zeta = \xi^g$ which has density $f_\zeta(z) = \frac{1}{|a|}f_\xi(\frac{z-b}{a})$ and is bounded by the density $f_Z(z) = \frac{1}{|a|}f_X(\frac{z-b}{a})$ of Z_i . Further, let $\tilde{\boldsymbol{\beta}} = (\mathbf{Q}^\top)^{-1}\boldsymbol{\beta} = (\beta_0 - \beta_1 b/a, \beta_1/a)^\top$. By the transformation theorem for measure integrals,

$$\begin{aligned} \mathbf{M}(\zeta, \tilde{\boldsymbol{\beta}}) &= \int \exp(\beta_0 + \beta_1(z - b)/a)\mathbf{f}(z)\mathbf{f}(z)^\top \zeta(dz) \\ &= \int \exp(\beta_0 + \beta_1 x)\mathbf{Q}\mathbf{f}(x)\mathbf{f}(x)^\top \mathbf{Q}^\top \xi(dx) \\ &= \mathbf{Q}\mathbf{M}(\xi, \boldsymbol{\beta})\mathbf{Q}^\top. \end{aligned}$$

Therefore $\det(\mathbf{M}(\zeta, \tilde{\boldsymbol{\beta}})) = \det(\mathbf{Q})^2 \det(\mathbf{M}(\xi, \boldsymbol{\beta}))$. Thus ξ^* maximizes the D -criterion over the set of subsampling designs bounded by f_X for β_1 if and only if ζ^* maximizes the D -criterion over the set of subsampling designs bounded by f_Z for β_1/a . \square

REFERENCES

- Mingyao Ai, Jun Yu, Huiming Zhang, and HaiYing Wang. Optimal subsampling algorithms for big data regressions. *Statistica Sinica*, 31:749–772, 2021.
- Qianshun Cheng, HaiYing Wang, and Min Yang. Information-based optimal subdata selection for big data logistic regression. *Journal of Statistical Planning and Inference*, 209:112–122, 2020. doi: <https://doi.org/10.1016/j.jspi.2020.03.004>.
- Valerii V. Fedorov. Optimal design with bounded density: optimization algorithms of the exchange type. *Journal of Statistical Planning and Inference*, 22:1–13, 1989. doi: [https://doi.org/10.1016/0378-3758\(89\)90060-8](https://doi.org/10.1016/0378-3758(89)90060-8).
- Berend Hasselman. *nleqslv: Solve Systems of Nonlinear Equations*, 2018. URL <https://CRAN.R-project.org/package=nleqslv>. R package version 3.3.2.
- Friedrich Pukelsheim. *Optimal Design of Experiments*. Wiley, New York, 1993.
- Martin Radloff and Rainer Schwabe. Invariance and equivariance in experimental design for nonlinear models. In J. Kunert, C.H. Müller, and A.C. Atkinson, editors, *mODa 11-Advances in Model-Oriented Design and Analysis*, pages 217–224. Springer, 2016.
- Torsten Reuter and Rainer Schwabe. Optimal subsampling design for polynomial regression in one covariate. *Statistical Papers*, 64:1095–1117, 2023. doi: <https://doi.org/10.1007/s00362-023-01425-0>.
- C. Rodríguez-Torreblanca and J.M. Rodríguez-Díaz. Locally D-and c-optimal designs for Poisson and negative binomial regression models. *Metrika*, 66:161–172, 2007.
- Kenneth G. Russell, David C. Woods, Susan M. Lewis, and John A. Eccleston. D-optimal designs for Poisson regression models. *Statistica Sinica*, 19:721–730, 2009.
- Michael Sahm and Rainer Schwabe. A note on optimal bounded designs. In A. Atkinson, B. Bogacka, and A. Zhigljavsky, editors, *Optimum Design 2000*, pages 131–140. Kluwer, Dordrecht, 2001.
- Mahmood Ul Hassan and Frank Miller. Optimal item calibration for computerized achievement tests. *Psychometrika*, 84:1101–1128, 2019.
- HaiYing Wang, Rong Zhu, and Ping Ma. Optimal subsampling for large sample logistic regression. *Journal of the American Statistical Association*, 113:829–844, 2018.

HaiYing Wang, Min Yang, and John Stufken. Information-based optimal subdata selection for big data linear regression. *Journal of the American Statistical Association*, 114:393–405, 2019.

Henry P. Wynn. Optimum designs for finite populations sampling. In S.S. Gupta and D.S. Moore, editors, *Statistical Decision Theory and Related Topics II*, pages 471–478. Academic Press, New York, 1977.

Jun Yu, HaiYing Wang, Mingyao Ai, and Huiming Zhang. Optimal distributed subsampling for maximum quasi-likelihood estimators with massive data. *Journal of the American Statistical Association*, 117:265–276, 2022. doi: <https://doi.org/10.1080/01621459.2020.1773832>.

TORSTEN REUTER. OTTO VON GUERICKE UNIVERSITY MAGDEBURG. UNIVERSITÄTSPLATZ 2, 39106 MAGDEBURG, GERMANY

Email address: `torsten.reuter@ovgu.de`

RAINER SCHWABE. OTTO VON GUERICKE UNIVERSITY MAGDEBURG. UNIVERSITÄTSPLATZ 2, 39106 MAGDEBURG, GERMANY

Email address: `rainer.schwabe@ovgu.de`