Bocconi University

Academic Year 2025–2026

20592 — Statistics and Probability (Computational Statistics)

Final Project Report

Logistic and Poisson Regression: Fisher Scoring and Bayesian MCMC

Author

Marco Frova (3197241)

1 Introduction

In this project we apply Fisher scoring and Bayesian Markov Chain Monte Carlo methods to two generalized linear models. More specifically, we study credit card ownership with a logistic regression model, and hospital length of stay with a Poisson regression model.

For both problems, we estimate the parameters using Fisher scoring (with the canonical link) and also using Bayesian inference with a random-walk Metropolis–Hastings algorithm. Then we compare the frequentist and Bayesian outputs, we check convergence diagnostics for the MCMC, and we comment on how the covariates seem to affect the response. At the end, we also discuss what happens to Fisher scoring when the link is canonical versus when it is not.

2 Methodology

2.1 Algorithms for GLMs

Fisher scoring is an iterative optimization method for computing the maximum likelihood estimator when no closed-form solution is available. It is closely related to Newton–Raphson, but instead of using the observed Hessian of the log-likelihood, it uses the expected information matrix (the Fisher information). This replacement improves numerical stability and yields updates based on the local curvature of the expected log-likelihood.

Generalized Linear Models belong to the exponential family, for which the score function and Fisher information admit simple expressions depending on the mean–variance relationship and the link function. As a consequence, Fisher scoring leads naturally to the iteratively reweighted least squares (IRLS) algorithm: at each iteration, the log-likelihood is locally approximated by a quadratic form, yielding a weighted least squares problem with weights determined by the variance function and a working response defined by the link. Beyond computation, the Fisher information plays a central inferential role. Under standard regularity conditions, the MLE is asymptotically normal with covariance given by the inverse Fisher information, forming the basis for Wald tests and confidence intervals in GLMs.

The Metropolis-Hastings algorithm is a Markov Chain Monte Carlo (MCMC) method designed to sample from a target distribution known only up to a proportionality constant. It constructs a Markov chain whose stationary distribution coincides with the target density, typically a Bayesian posterior. Under standard conditions of irreducibility and aperiodicity, the chain is ergodic, ensuring convergence of Monte Carlo averages.

In Bayesian generalized linear models, the posterior distribution of the regression coefficients is proportional to the likelihood times the prior and does not admit a closed-form expression. Metropolis–Hastings enables sampling from this posterior without computing the normalizing constant. Posterior means, variances, and credible intervals are approximated by empirical averages of the simulated draws. For large samples, the posterior concentrates around the MLE and becomes approximately normal with covariance related to the inverse Fisher information, linking Bayesian MCMC inference with Fisher scoring–based frequentist inference.

All covariates were standardized prior to estimation, and an intercept was included. Bayesian inference used independent Gaussian priors and symmetric Gaussian random-walk proposals.

2.2 Analytical quantities used in the implementation

Logistic regression (Bernoulli, logit link). Let $\eta = X\beta$ and $p = \sigma(\eta)$ with $\sigma(u) = 1/(1+e^{-u})$. The Bernoulli log-likelihood is $\ell(\beta) = \sum_{i=1}^n (y_i\eta_i - \log(1+e^{\eta_i}))$. The score and Fisher information are $U(\beta) = X^\top(y-p)$ and $\mathcal{I}(\beta) = X^\top W X$, with $W = \operatorname{diag}(p_i(1-p_i))$. Fisher scoring (canonical logit link) updates $\beta^{(t+1)} = \beta^{(t)} + (X^\top W^{(t)} X)^{-1} X^\top (y - p^{(t)})$. In IRLS form, the working response is $z^{(t)} = \eta^{(t)} + (y-p^{(t)})/[p^{(t)}(1-p^{(t)})]$, and convergence is declared when the relative change in β falls below $\varepsilon = 10^{-8}$.

Poisson regression (log link). Let $\eta = X\beta$ and $\mu = \exp(\eta)$. Up to constants, the log-likelihood is $\ell(\beta) = \sum_{i=1}^n (y_i\eta_i - \mu_i)$. The score and Fisher information are $U(\beta) = X^\top(y-\mu)$ and $\mathcal{I}(\beta) = X^\top W X$, with $W = \operatorname{diag}(\mu_i)$, yielding the Fisher scoring update $\beta^{(t+1)} = \beta^{(t)} + (X^\top W^{(t)} X)^{-1} X^\top (y - \mu^{(t)})$. The corresponding IRLS working response is $z^{(t)} = \eta^{(t)} + (y-\mu^{(t)})/\mu^{(t)}$.

Bayesian posterior and Metropolis–Hastings. For both models, coefficients follow independent Gaussian priors $\beta_j \sim \mathcal{N}(0, \sigma^2)$ with $\sigma^2 = 100$. The log-posterior is $\log \pi(\beta \mid y) = \ell(\beta) - (2\sigma^2)^{-1}\beta^\top\beta + C$. With a symmetric Gaussian random-walk proposal $\beta^\star = \beta + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, s^2 I)$, the Metropolis–Hastings acceptance probability simplifies to $\alpha(\beta, \beta^\star) = \min\{1, \exp[\log \pi(\beta^\star \mid y) - \log \pi(\beta \mid y)]\}$. Posterior summaries and transformed effects (odds ratios and incidence rate ratios) are obtained by applying the transformation to the retained MCMC draws.

2.3 Fisher scoring vs. Newton's method for canonical links

Let $\ell(\beta)$ denote the log-likelihood and $U(\beta) = \nabla \ell(\beta)$ the score. Newton's method for maximizing $\ell(\beta)$ updates $\beta^{(t+1)} = \beta^{(t)} - H(\beta^{(t)})^{-1} U(\beta^{(t)})$, where $H(\beta) = \nabla^2 \ell(\beta)$ is the Hessian. Fisher scoring replaces the observed Hessian with the Fisher information $\mathcal{I}(\beta) = -\mathbb{E}[H(\beta) \mid X]$, yielding $\beta^{(t+1)} = \beta^{(t)} + \mathcal{I}(\beta^{(t)})^{-1} U(\beta^{(t)})$.

For GLMs with canonical link, the log-likelihood has the exponential-family form $\ell(\beta) = \sum_{i=1}^n \{y_i \eta_i - b(\eta_i)\} + c(y)$ with $\eta_i = x_i^\top \beta$ and $\mu_i = b'(\eta_i)$. Differentiating gives $U(\beta) = X^\top (y - \mu)$. Moreover $H(\beta) = -X^\top W X$ with $W = \operatorname{diag}(b''(\eta_i)) = \operatorname{diag}(\operatorname{Var}(Y_i \mid X))$. Since $\mathbb{E}(y \mid X) = \mu$, the expected negative Hessian has the same form, hence $\mathcal{I}(\beta) = X^\top W X = -H(\beta)$. Substituting $H(\beta) = -\mathcal{I}(\beta)$ into Newton's update shows that Newton and Fisher scoring coincide for canonical-link GLMs. This includes logistic regression (logit link, $W_i = \mu_i(1 - \mu_i)$) and Poisson regression (log link, $W_i = \mu_i$).

2.4 What changes under a non-canonical link (e.g. probit)

With a non-canonical link $g(\mu_i) = \eta_i = x_i^\top \beta$, the overall Fisher scoring structure is unchanged (compute $U(\beta^{(t)})$, compute $\mathcal{I}(\beta^{(t)})$, solve $\mathcal{I}(\beta^{(t)})\Delta^{(t)} = U(\beta^{(t)})$, update $\beta^{(t+1)} = \beta^{(t)} + \Delta^{(t)}$), but the explicit expressions for the score and information change because μ_i is no longer $b'(\eta_i)$ and derivatives of the inverse link appear. In particular, the weight matrix is no longer simply $W_i = b''(\eta_i)$; it incorporates link-specific terms involving $d\mu_i/d\eta_i$. A key consequence is that the observed Hessian $H(\beta)$ is generally not equal to $-\mathcal{I}(\beta)$, so Newton's step $\beta^{(t+1)} = \beta^{(t)} - H(\beta^{(t)})^{-1} U(\beta^{(t)})$ and the Fisher scoring step $\beta^{(t+1)} = \beta^{(t)} + \mathcal{I}(\beta^{(t)})^{-1} U(\beta^{(t)})$ need not coincide. For probit regression, for instance, $\mu_i = \Phi(\eta_i)$ and $d\mu_i/d\eta_i = \phi(\eta_i)$ enters the curvature through link-specific weights, leading to different updates than the canonical-logit case.

3 Applications and results

Common elements across the two applications. We keep modelling, computation, and reporting choices aligned so the two case studies can be compared directly.

i) Prior choice. All regression coefficients receive independent Gaussian priors, $\beta_j \sim \mathcal{N}(0, 100)$. Since predictors are standardised (mean 0, sd 1), this prior is weakly informative: it stabilises computation and provides gentle shrinkage without forcing the analysis to be prior-driven. The limitation is that a diffuse normal prior is not especially robust when the likelihood is irregular or when a handful of observations have strong leverage. If stronger regularisation were desired, a tighter normal prior would shrink coefficients more decisively. If robustness were the priority, a heavier-tailed prior such as Student-t would still regularise while allowing occasional large effects. To justify this choice, we conducted a prior sensitivity analysis by varying the prior variance over $\sigma^2 \in \{1, 6.25, 25, 100, 400\}$. For both datasets, posterior means, 95% credible intervals, and the ranking of the most influential covariates remained essentially unchanged across this range, and MCMC acceptance rates were stable. This indicates that inference is primarily data-driven rather than prior-driven. We therefore retain $\sigma^2 = 100$ as a reasonable weakly informative default. (See Appendix Tables 4 and 5.)

ii) MCMC diagnostics. Posterior inference is computed with random-walk Metropolis–Hastings. We monitor the sampler using trace plots and autocorrelation. After burn-in, a trace plot should wander around a stable level with no visible drift; this behaviour is consistent with stationarity, meaning the chain is sampling from the target posterior rather than from transient initial conditions. Autocorrelation quantifies dependence between draws k iterations apart. A rapid decay of $\widehat{\rho}(k)$ indicates good mixing and lower Monte Carlo dependence, so fewer iterations are "wasted" on highly correlated draws. In our fits, the monitored coefficients showed stable traces after burn-in and quickly diminishing autocorrelation, supporting the use of posterior means and credible intervals. Trace plots and autocorrelation functions for the monitored coefficients are reported in Appendix A.1 (CreditCard) and Appendix A.2 (azcabgptca).

iii) *Interval reporting and interpretation.* For each dataset we also fit the corresponding GLM by Fisher scoring to obtain the maximum likelihood estimate (MLE). With a weakly informative prior and reasonably informative samples, posterior means are expected to track the MLE closely, which is what we observed. Since our numerical summaries are Bayesian, we report 95% credible intervals. They directly express uncertainty conditional on the observed data and the model: a 95% credible interval contains 95% posterior probability mass for β_j. Confidence intervals from the frequentist fit answer a different question instead (long-run coverage under repeated sampling). Complete posterior summaries for both models are reported in Appendix A.3.

3.1 CreditCard dataset: Bayesian logistic regression

Data, variables, and preprocessing. The response is `card` (acceptance), with $n = 1319$ observations: 1023 accepted (77.56%) and 296 rejected (22.44%). Covariates are `reports`, `age`, `income`, `owner`, `selfemp`, `dependents`, `months`, `majorcards`, and `active`, plus an intercept. All predictors were standardised (mean 0, sd 1). We excluded `share` and `expenditure` to avoid *data leakage*, meaning variables determined by or only observable after the decision outcome; in this context they proxy post-acceptance usage and would inflate apparent performance while muddying interpretation.

Model. For individual i, $Y_i \mid \beta \sim \text{Bernoulli}(p_i)$ with $\text{logit}(p_i) = x_i^\top \beta$. Because predictors are standardised, each coefficient describes the effect of a one-standard-deviation increase in the corresponding covariate on the log-odds of acceptance; exponentiating gives the multiplicative change in the odds.

Posterior results and interpretation. The intercept represents the baseline log-odds of acceptance for an applicant whose covariates are all at their sample means (since all predictors are standardised). The posterior mean intercept is 1.4702, corresponding to a baseline acceptance probability of

$$p_0 = \text{logit}^{-1}(1.4702) = \frac{1}{1 + e^{-1.4702}} \approx 0.81.$$

Thus, an "average" applicant in the sample has an estimated probability of acceptance of about 81%. The main drivers of acceptance are `reports`, `active`, and `income`.

`reports` has a large negative effect (posterior mean -2.3741, 95% credible interval $[-2.7306, -2.0337]$; odds ratio 0.0931). In words, increasing `reports` by one standard deviation (i.e., moving one SD upward in the sample distribution of the number of major derogatory reports) multiplies the odds of acceptance by about 0.09 (a sharp reduction). Relative to the baseline $p_0 \approx 0.81$, this shift implies a predicted acceptance probability of

$$\text{logit}^{-1}(1.4702 - 2.3741) \approx 0.29$$

for an otherwise average applicant, matching the idea that more adverse credit signals strongly reduce approval chances.

`active` is strongly positive (mean 0.8320, 95% CI $[0.6115, 1.0623]$; odds ratio 2.2979). A one-standard-deviation increase in `active`, so one SD upward in the number of active credit accounts, multiplies the odds of acceptance by roughly 2.3. Starting from $p_0 \approx 0.81$, this corresponds to a predicted acceptance probability of

$$\text{logit}^{-1}(1.4702 + 0.8320) \approx 0.91,$$

consistent with active credit engagement being read as evidence of established access and repayment history.

`income` is positive (mean 0.3904, 95% CI $[0.1786, 0.6062]$; odds ratio 1.4775). A one-standard-deviation increase in `income`, so one SD upward in yearly income, measured in USD 10,000, increases the odds of acceptance by about 48%. Relative to $p_0 \approx 0.81$, this implies a predicted acceptance probability of

$$\text{logit}^{-1}(1.4702 + 0.3904) \approx 0.87,$$

in line with stronger repayment capacity.

Other covariates: direction, significance, and intuition. Among the remaining predictors, `owner` is credibly positive, which is consistent with financial stability or accumulated wealth (after coding the factor and standardising). `majorcards` is also credibly positive, fitting the idea that already holding major cards signals established creditworthiness. On the negative side, `selfemp` is credibly below zero, plausibly reflecting higher income volatility (after coding and standardising), and `dependents` is credibly negative, consistent with larger fixed obligations. By contrast, `age` and `months` are not clearly different from zero at the 95% level once the other covariates are included. Full posterior summaries and odds ratios are reported in Appendix Tables 2 and 6.

3.2 azcabgptca dataset: Bayesian Poisson regression

Data and preprocessing. The response is `los` (length of stay, days), treated as a non-negative count. The dataset has $n = 1959$ observations with covariates `died`, `procedure`, `age`, `gender`, and `type`, plus an intercept. All predictors were standardised.

Model. We fit a Poisson GLM with log link: $Y_i \mid \beta \sim \text{Poisson}(\mu_i)$ and $\log(\mu_i) = x_i^\top \beta$. With standardised predictors, each coefficient corresponds to a one-standard-deviation increase in the covariate. Exponentiating β_j gives an incidence rate ratio (IRR), i.e. the multiplicative change in the expected length of stay.

Posterior results and interpretation. The intercept represents the baseline log expected length of stay for a hypothetical patient whose covariates are all set to their sample means (since all predictors are standardised). The posterior mean intercept is 1.8308, so the baseline expected length of stay is

$$\mu_0 = \exp(1.8308) \approx 6.24 \text{ days.}$$

The dominant driver is `procedure` (mean 0.5645, 95% CI $[0.5465, 0.5829]$; IRR 1.7585). A one-standard-deviation increase in `procedure` (i.e., moving one SD upward in the sample distribution of the procedure indicator/intensity as coded) multiplies the expected length of stay by about 1.76. Relative to the baseline $\mu_0 \approx 6.24$, this corresponds to an expected stay of

$$\mu_0 \times 1.7585 \approx 10.97 \text{ days}$$

for an otherwise average patient (an increase of roughly $10.97 - 6.24 \approx 4.73$ days), consistent with more intensive procedures requiring longer recovery.

`type` is clearly positive (mean 0.0942, 95% CI $[0.0777, 0.1104]$; IRR 1.0988). A one-standard-deviation increase in `type` (i.e., one SD upward in the coding of the patient/type category) corresponds to roughly a 10% increase in expected stay. Starting from $\mu_0 \approx 6.24$, this implies

$$\mu_0 \times 1.0988 \approx 6.86 \text{ days}$$

(an increase of about 0.62 days), capturing systematic differences across categories.

`gender` is small and negative (mean -0.0489, 95% CI $[-0.0656, -0.0323]$; IRR 0.9523). Interpreted on the IRR scale, a one-standard-deviation increase in `gender` as coded is associated with about a 5% reduction in expected stay. Relative to $\mu_0 \approx 6.24$, this corresponds to

$$\mu_0 \times 0.9523 \approx 5.94 \text{ days}$$

(a decrease of about 0.30 days). Because `gender` is a coded (typically binary) covariate, this "+1 SD" change should be interpreted as a shift across the sample distribution of the coding rather than as a literal one-unit change in gender status.

Other covariates: direction, significance, and intuition. Two additional effects complete the picture. age is credibly positive, consistent with slower recovery and higher complication risk at older ages (under the Poisson log-link, this implies a multiplicative increase in expected days for a one-SD increase in age). died is credibly negative, which is coherent with truncation: conditional on the model and covariates, when death occurs the observed stay can end earlier than it would under a full recovery trajectory. Complete coefficient summaries and IRRs are provided in Appendix Tables 3 and 7.

A Appendix

A.1 MCMC diagnostics: CreditCard dataset

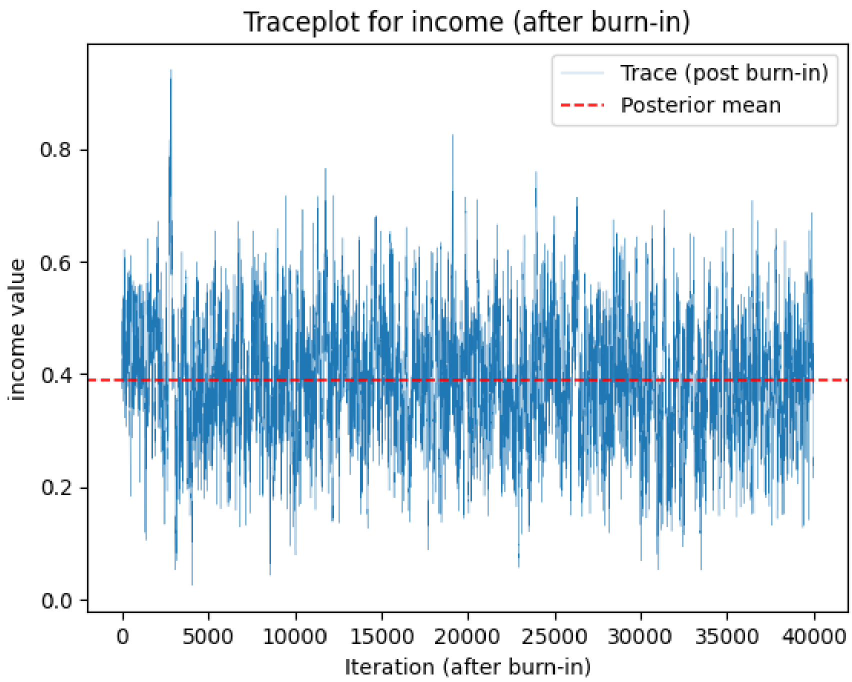

Figure 1: Trace plot for the coefficient associated with income in the CreditCard logistic regression (post burn-in).

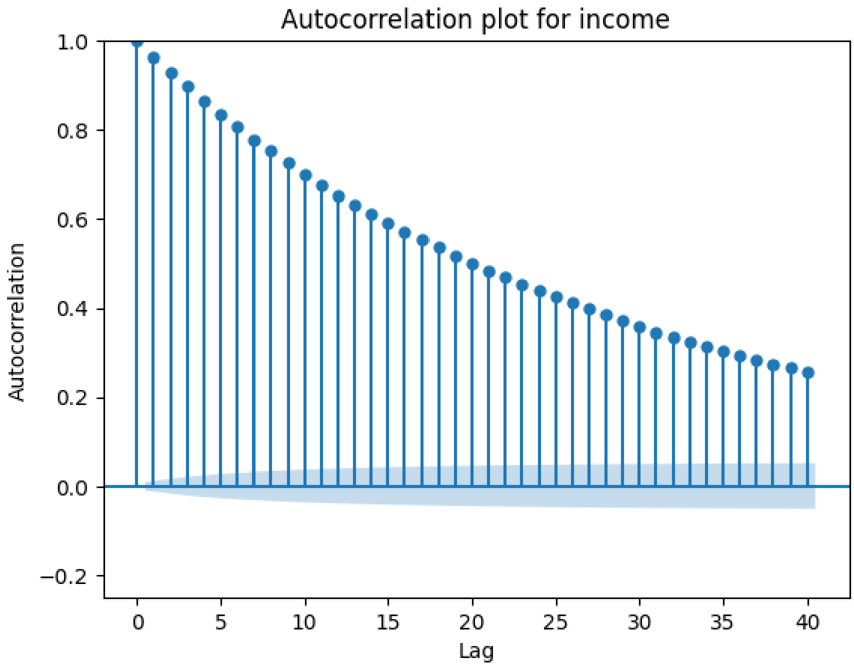

Figure 2: Autocorrelation function for the coefficient associated with `income` in the CreditCard logistic regression.

A.2 MCMC diagnostics: azcabgptca dataset

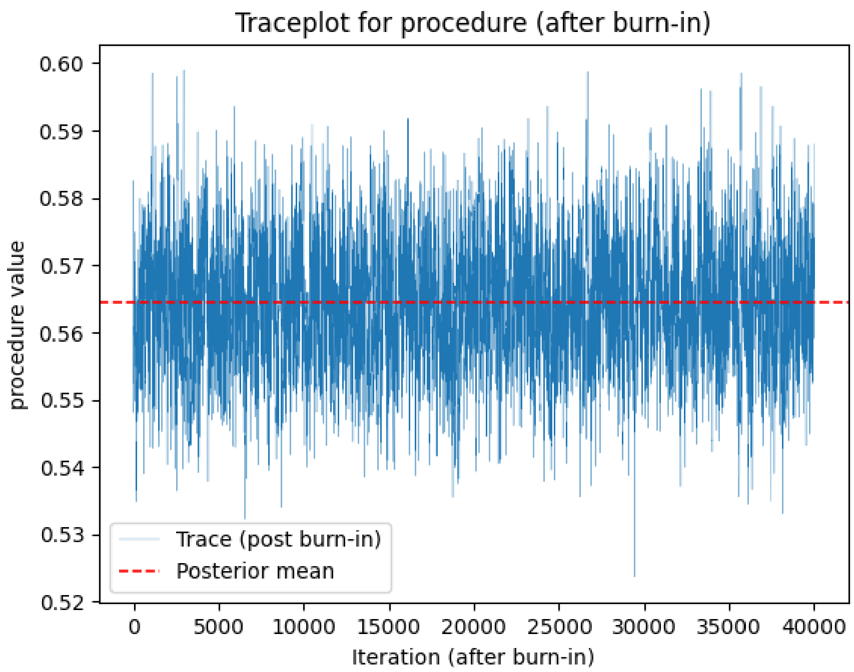

Figure 3: Trace plot for the coefficient associated with `procedure` in the azcabgptca Poisson regression (post burn-in).

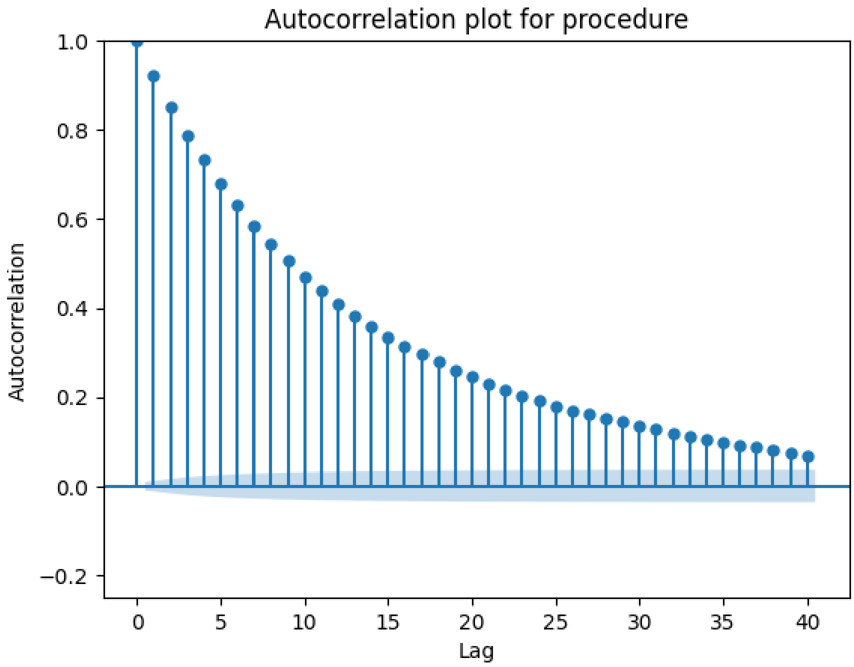

Figure 4: Autocorrelation function for the coefficient associated with `procedure` in the azcabgptca Poisson regression.

A.3 MCMC settings and posterior summaries

Table 1: Random-walk Metropolis–Hastings settings and acceptance rates.

Model	Iterations	Burn-in	Acceptance rate
Logistic (CreditCard)	50,000	10,000	0.333
Poisson (azcabgptca)	50,000	10,000	0.310

Table 2: Posterior summaries for the CreditCard logistic regression (RW-MH).

Coefficient	Mean	SD	95% credible interval
intercept	1.4702	0.0920	[1.2947, 1.6537]
reports	-2.3741	0.1779	[-2.7306, -2.0337]
age	-0.1277	0.0991	[-0.3183, 0.0743]
income	0.3904	0.1111	[0.1786, 0.6062]
owner	0.2445	0.1023	[0.0424, 0.4420]
selfemp	-0.1888	0.0762	[-0.3353, -0.0320]
dependents	-0.3019	0.0872	[-0.4735, -0.1321]
months	0.0373	0.0945	[-0.1459, 0.2230]
majorcards	0.1973	0.0734	[0.0515, 0.3426]
active	0.8320	0.1162	[0.6115, 1.0623]

Table 3: Posterior summaries for the azcabgptca Poisson regression (RW-MH).

Coefficient	Mean	SD	95% credible interval
intercept	1.8308	0.0096	[1.8115, 1.8492]
died	-0.0414	0.0081	[-0.0575, -0.0255]
procedure	0.5645	0.0093	[0.5465, 0.5829]
gender	-0.0489	0.0086	[-0.0656, -0.0323]
age	0.0429	0.0085	[0.0263, 0.0598]
type	0.0942	0.0084	[0.0777, 0.1104]

Table 4: Prior sensitivity analysis for the CreditCard logistic regression. The table reports the effect of varying the Gaussian prior variance on the posterior mean and 95% credible interval of the coefficient associated with `income`, together with the MCMC acceptance rate and the ranking of the three most influential covariates.

Prior var	Prior sd	Acc. rate	Top 3 covariates	$\mathbb{E}[\beta_{\text{income}}]$	95% CI
1.00	1.0	0.330	reports, active, income	0.3757	[0.1660, 0.5931]
6.25	2.5	0.336	reports, active, income	0.3901	[0.1754, 0.6118]
25.00	5.0	0.335	reports, active, income	0.3871	[0.1857, 0.6078]
100.00	10.0	0.333	reports, active, income	0.3904	[0.1786, 0.6062]
400.00	20.0	0.336	reports, active, income	0.3888	[0.1770, 0.6203]

Table 5: Prior sensitivity analysis for the azcabgptca Poisson regression. The table reports the effect of varying the Gaussian prior variance on the posterior mean and 95% credible interval of the coefficient associated with `procedure`, together with the MCMC acceptance rate and the ranking of the three most influential covariates.

Prior var	Prior sd	Acc. rate	Top 3 covariates	$\mathbb{E}[\beta_{\text{procedure}}]$	95% CI
1.00	1.0	0.313	procedure, type, gender	0.5647	[0.5461, 0.5835]
6.25	2.5	0.310	procedure, type, gender	0.5644	[0.5459, 0.5830]
25.00	5.0	0.311	procedure, type, gender	0.5647	[0.5466, 0.5833]
100.00	10.0	0.310	procedure, type, gender	0.5645	[0.5465, 0.5829]
400.00	20.0	0.310	procedure, type, gender	0.5645	[0.5465, 0.5829]

A.4 Odds ratios and incidence rate ratios

Table 6: Odds ratios for the CreditCard logistic regression (posterior means and 95% credible intervals).

Parameter	OR mean	CI low	CI high
reports	0.0931	0.0652	0.1308
age	0.8801	0.7274	1.0771
income	1.4775	1.1955	1.8334
owner	1.2770	1.0433	1.5558
selfemp	0.8280	0.7151	0.9685
dependents	0.7394	0.6228	0.8763
months	1.0380	0.8643	1.2498
majorcards	1.2181	1.0529	1.4085
active	2.2979	1.8433	2.8930

Table 7: Incidence rate ratios for the azcabgptca Poisson regression (posterior means and 95% credible intervals).

Parameter	IRR mean	CI low	CI high
died	0.9594	0.9441	0.9749
procedure	1.7585	1.7271	1.7912
gender	0.9523	0.9365	0.9682
age	1.0438	1.0267	1.0616
type	1.0988	1.0808	1.1167

A.5 Fisher scoring (IRLS) estimates

Table 8: Fisher scoring (IRLS) estimates for the CreditCard logistic regression.

Coefficient	$\hat{\beta}_{\mathrm{FS}}$
intercept	1.4582
reports	-2.3556
age	-0.1269
income	0.3832
owner	0.2374
selfemp	-0.1919
dependents	-0.3022
months	0.0338
majorcards	0.1953
active	0.8339

Table 9: Fisher scoring (IRLS) estimates for the azcabgptca Poisson regression.

Coefficient	$\hat{\beta}_{\text{FS}}$
intercept	1.8310
died	-0.0412
procedure	0.5643
gender	-0.0489
age	0.0432
type	0.0943

A.6 Dataset snapshots (first five rows)

Table 10: First five rows of the CreditCard dataset.

card	reports	age	income	share	expenditure	owner	selfemp	dependents	months	majorcards	active
yes	0	37.6667	4.52	0.0332699	124.983	yes	no	3	54	1	12
yes	0	33.25	2.42	0.00521694	9.85417	no	no	3	34	1	13
yes	0	33.6667	4.5	0.00415556	15	yes	no	4	58	1	5
yes	0	30.5	2.54	0.0652138	137.869	no	no	0	25	1	7
yes	0	32.1667	9.7867	0.0670506	546.503	yes	no	2	64	1	5

Table 11: First five rows of the azcabgptca dataset.

died	procedure	age	gender	los	type
0	1	73	0	51	0
0	0	67	0	30	1
0	1	69	0	43	0
0	1	65	0	32	0
0	1	79	0	42	1