



## Skewed probit regression — identifiability, contraction and reformulation

Item Type	Article
Authors	Van Niekerk, Janet;Rue, Haavard
Eprint version	Post-print
Publisher	National Statistical Institute
Journal	Revstat Statistical Journal
Rights	Archived with thanks to Revstat Statistical Journal
Download date	2023-12-09 01:24:43
Link to Item	<a href="http://hdl.handle.net/10754/666049">http://hdl.handle.net/10754/666049</a>

---

---

## SKEWED PROBIT REGRESSION - IDENTIFIABILITY, CONTRACTION AND REFORMULATION\*

---

---

Authors: JANET VAN NIEKERK   
– CEMSE Division, King Abdullah University of Science and Technology,  
Kingdom of Saudi Arabia (Janet.vanNiekerk@kaust.edu.sa)  
HÅVARD RUE  
– CEMSE Division, King Abdullah University of Science and Technology,  
Kingdom of Saudi Arabia (Haavard.Rue@kaust.edu.sa)

1 Received: Month 0000      Revised: Month 0000      Accepted: Month 0000

2 Abstract:

3 • Skewed probit regression is but one example of a statistical model that generalizes  
4 a simpler model, like probit regression. All skew-symmetric distributions and link  
5 functions arise from symmetric distributions by incorporating a skewness parameter  
6 through some skewing mechanism. In this work we address some fundamental issues  
7 in skewed probit regression, and more generally skew-symmetric distributions or skew-  
8 symmetric link functions.

9 We address the issue of identifiability of the skewed probit model parameters by  
10 reformulating the intercept from first principles. A new standardization of the skew  
11 link function is given to provide an anchored interpretation of the inference. Possible  
12 skewness parameters are investigated and the penalizing complexity priors of these are  
13 derived. This prior is invariant under reparameterization of the skewness parameter  
14 and quantifies the contraction of the skewed probit model to the probit model.

15 The proposed results are available in the *R-INLA* package and we illustrate the use  
16 and effects of this work using simulated data, and well-known datasets using the link  
17 as well as the likelihood.

18 Key-Words:

19 • *Skew symmetric, Probit, Binary regression, Penalizing complexity, INLA.*

20 AMS Subject Classification:

21 • 49A05, 78B26.

---

\*The opinions expressed in this text are those of the authors and do not necessarily reflect the views of any organization.

---

## 1. INTRODUCTION

---

1 Skew-symmetric distributions have acclaimed fame due to their ability to  
 2 model skewed data, by introducing a skewness parameter to a symmetric distribu-  
 3 tion, through some skewing mechanism. In the preceding decades, an abundance  
 4 of skewed distributions has been proposed from the basis of symmetric distribu-  
 5 tions, like the skew-normal [30, 3], skew-t [4] and more generally skew-elliptical  
 6 distributions [21]. In each of these skew distributions, an additional parameter is  
 7 introduced that indicates the direction of skewness or alternatively, symmetry.

8 With the introduction of the additional parameter, the inferential problem  
 9 can become more challenging. The identifiability of the parameters and the exis-  
 10 tence of the maximum likelihood estimators (MLEs) are issues to keep in mind. In  
 11 the Bayesian paradigm, the choice of a prior for the skewness parameter emerges.  
 12 Either way, the inference of the skewness parameter is crucial in evaluating the  
 13 appropriateness of the underlying (skewed) model.

14 A continuous random variable  $X$ , follows a skew-normal (SN) distribution  
 15 with location, scale and skewness(shape) parameters  $\xi, \omega$  and  $\alpha$ , respectively, if  
 16 the probability density function (pdf) is as follows:

$$(1.1) \quad g(x) = \frac{2}{\omega} \phi\left(\frac{x - \xi}{\omega}\right) \Phi\left[\alpha \left(\frac{x - \xi}{\omega}\right)\right],$$

17 where  $\alpha \in \mathfrak{R}$ ,  $\omega > 0$ ,  $\xi \in \mathfrak{R}$ , and  $\phi(\cdot)$  and  $\Phi(\cdot)$  are the density and cumulative  
 18 distribution function (CDF) of the standard Gaussian distribution, respectively.  
 19 Denote by  $G(x)$  the CDF of the skew-normal density.

20 The parameterisation in (1.1) poses difficulties since the mean and variance  
 21 depends on  $\alpha$ , as  $E[X] = \xi + \omega\delta\sqrt{2/\pi}$  and  $V[X] = \omega^2(1 - 2\delta^2/\pi)$ , where  $\delta =$   
 22  $\alpha/\sqrt{1 + \alpha^2}$ . This implies that inference for  $\alpha$  will also influence the inference for  
 23 the mean and variance, since both are functions of  $\alpha$ .

24 A similar challenge arises in the binary regression framework where the  
 25 skew-normal link function is used as a generalization of probit regression, namely  
 26 skewed probit regression. The need for asymmetric link functions have been  
 27 noted by [14]. In binary regression, asymmetric link functions are essential in  
 28 cases where the probability of a particular binary response approaches zero and  
 29 one at different rates. In this case, a symmetric link function will result in sub-  
 30 stantially biased estimators with over(under)estimation of the mean probability  
 31 of the binary response, due to the different rates of approaching zero and one  
 32 (see [16] for more details on this issue). Skewed probit regression is an extension  
 33 of probit regression, where covariates are transformed through the skew-normal  
 34 CDF instead of the standard normal CDF.

35 Here, it might not be intuitive when the skewed link function is more appro-  
 36 priate than the symmetric link function. The estimate of the skewness parameter

1 could provide some insights into this, only if the inference of the skewness pa-  
2 rameter is reliable and interpretable.

3       Regarding the inference of the skewness parameter,  $\alpha$  in (1.1), being it  
4 in the skewed probit regression or the skew-normal distribution as the underly-  
5 ing response model (which are conceptually the same estimation setup), various  
6 works have been contributed, most of them dedicated to the skew-normal response  
7 model framework. The identifiability of the parameters in the skew-normal re-  
8 sponse model was investigated by [22] (and skew-elliptical in general), [31] (for  
9 finite mixtures) and [13] (for extensions of the skew-normal distributions). For  
10 binary regression, identifiability of the parameters was considered by [25] where  
11 some issues concerning identifiability were raised. We address the identifiability  
12 problem from a first principles viewpoint, so that the parameters are identifiable,  
13 even with weak covariates, hence adding to [25].

14       In the skew-normal response model, the bias of the MLEs is a well-known  
15 fact (see [34] for more details). For small and moderate sample sizes, the MLE  
16 of the skewness parameter could be infinite with positive probability and the  
17 profile likelihood function has a singularity as the skewness parameter approaches  
18 zero, as noted early on by [3] (see also [26]). Some approaches to alleviate this  
19 feature of the skew-normal likelihood function have been proposed, including  
20 reparameterization of the model by [3] using the mean and variance (instead of  
21 location and scale parameters), or using a Bayesian framework by [27] (default  
22 priors) and [7] (proper priors). Also, [34] used the work of [19] to propose an  
23 adjusted (penalized) score function for frequentist estimation of the skewness  
24 parameter. A penalized MLE approach for all the parameters, including the  
25 skewness parameter, is presented by [5]. Bias-reduction regimes were proposed  
26 by [28].

27       From a Bayesian viewpoint, various priors for the skewness parameter have  
28 been proposed such as the Jeffrey's prior [27], truncated Gaussian prior [1], Stu-  
29 dent t prior and approximate Jeffery's prior [7], uniform prior [2], probability  
30 matching prior [11], informative Gaussian and unified skew-normal priors [12]  
31 and the beta-total variation prior [17]. All of these Bayesian approaches, with  
32 the exception of the latter, are based on somewhat arbitrary prior choices for  
33 mainly mathematical or computational convenience. These priors (as many oth-  
34 ers) are not invariant under reparameterization of the skewness parameter. The  
35 beta-total variation prior presented by [17] is based on the total variation from  
36 the symmetric Gaussian model to the skew-normal model, viewing the skewness  
37 parameter as a measure of perturbation. This prior is indeed invariant under  
38 one-to-one transformation of the skewness parameter.

39       Amongst the many works on the skew-normal response model, it seems that  
40 the genesis of the skew-normal model has been neglected. The skew-normal model  
41 was introduced by [3] as an (asymmetric) extension of the Gaussian model. The  
42 motivation for this extension is found in data. When data behaves like the Gaus-  
43 sian model, but the profile of the density is asymmetric, the skew-normal model

1 might be appropriate. Conversely, we need an inferential framework wherein the  
 2 skew-normal model would contract (or reduce) to the Gaussian model, in the ab-  
 3 sence of sufficient evidence of non-trivial skewness. The priors mentioned before  
 4 do not provide a quantification framework with which the modeler can under-  
 5 stand, and subsequently control this contraction. To achieve this, we need to  
 6 consider the model (either skewed probit regression or the skew-normal response  
 7 model) from an information theoretic perspective. Then we can construct a prior  
 8 with which the quantification of contraction (or not) can be done, and used as a  
 9 translation of prior information from the modeler to the model.

10 In this paper we address some issues (identifiability, standardizing, skewness  
 11 parameters) prevalent in skewed-probit regression in Section 2 and construct the  
 12 penalized complexity (PC) prior for the skewness parameter of the link function  
 13 (which is translatable to the skew-normal response model) in Section 4. This PC  
 14 prior is implemented in the *R-INLA* [32] (see also [33], [29]) package for general  
 15 use by others. We use a numerical study to illustrate the solutions proposed  
 16 in Section 2 and apply the PC prior to simulated and real data in Sections 5  
 17 and Section 6. The paper is concluded by a discussion in Section 7 in which  
 18 we sketch the wider applicability of this work and contributions to the wider  
 19 skew-symmetric family.

---

## 2. SKEWED PROBIT REGRESSION AND ISSUES

---

20 We consider skewed probit regression as an extension of probit regression,  
 21 where the link function is the skew-normal CDF instead of the standard normal  
 22 CDF. We formulate skewed probit regression that can include random effects like  
 23 spline functions of the covariates, spatial and/or temporal effects. For this paper,  
 24 we assume the following structure. From a sample of size  $n$ , the responses  $\mathbf{y}_{n \times 1}$   
 25 are counts of successful trials out of  $N_{n \times 1}$  trials and hence we assume a Binomial  
 26 distribution with success probability  $p$ . We gather all  $m$  covariates in  $\mathbf{X}_{n \times m}$  and  
 27 use these to build an additive linear predictor, defined as  $\boldsymbol{\eta}_{n \times 1}$ . So then,

$$(2.1) \quad \begin{aligned} y_i &\sim \text{Binomial}(N_i, p_i) \\ p_i &= G(\eta_i), \quad i = 1, \dots, n \end{aligned}$$

28 where  $G(\cdot)$  is the CDF of the Skew-Normal that depends on  $(\xi, \omega, \alpha)$ . The linear  
 29 predictor  $\eta_i$  is an additive linear predictor defined as follows,

$$(2.2) \quad \eta_i = \beta_0 + \boldsymbol{\beta}' \mathbf{X}_i + \sum_{k=1}^K f^k(\mathbf{Z}_i),$$

30 where  $\mathbf{X}$  and  $\mathbf{Z}$  are the covariates for the fixed and random effects, respectively,  
 31 the functions  $\{f^k(\cdot)\}$  are random effects like spatial, spline, temporal effects with  
 32 hyperparameters  $\boldsymbol{\theta}$ .

---

## 2.1. Issue 1 - Standardizing the link function

---

With the aim of standardizing the link function, [25] assumed  $\xi = 0, \omega = 1$ , similar to [8] and many others. Initially, the idea behind this choice feels intuitive since the skew probit link is an extension of the probit link through the skewness parameter. However, the  $(0, 1)$  parameter values of the probit link should not be naively copied to the skewed probit link. The choice,  $\xi = 0, \omega = 1$  implicitly concedes that a skew-normal density (1.1) with mean

$$E[X] = \alpha \sqrt{\frac{2}{\pi(1 + \alpha^2)}},$$

and variance

$$V[X] = 1 - \frac{2\alpha^2}{\pi(1 + \alpha^2)},$$

1 is used to calculate the probability of success, for all  $\alpha$ . This essentially implies  
 2 that for different skewness parameter values, different means and variances are  
 3 used. This way of standardizing is a parameter-based method, instead of the  
 4 intended property-based method like in the probit link. We do not expect the  
 5 assumption  $\xi = 0, \omega = 1$  to work well since the mean and variance are not  
 6 anchored and can attain many values based on different values of  $\alpha$ .

7 We posit that the mean and the variance (properties of the link) should  
 8 be fixed, like in the probit case, instead of the skew-normal location and scale  
 9 parameters. This is analogous to the idea of the centered parametrization of the  
 10 skew-normal density and mentioned by [9].

11 We propose the link function  $F(y|\alpha)$  that is the CDF of the Skew-Normal  
 12 density (1.1) scaled to have zero mean and unit variance for all values of  $\alpha$ . That  
 13 is,

$$F(y|\alpha) = \int_{-\infty}^y f(x|\alpha) dx$$

14 where

$$(2.3) \quad f(x|\alpha) = \frac{2}{\omega(\alpha)} \phi\left(\frac{x - \xi(\alpha)}{\omega(\alpha)}\right) \Phi\left[\alpha\left(\frac{x - \xi(\alpha)}{\omega(\alpha)}\right)\right],$$

$$\xi(\alpha) = -\omega\alpha \sqrt{\frac{2}{\pi(1 + \alpha^2)}},$$

and

$$\omega(\alpha) = \sqrt{\left(1 - \frac{2\alpha^2}{\pi(1 + \alpha^2)}\right)^{-1}}.$$

15 This provides an anchored link function with zero mean and unit variance, for  
 16 all  $\alpha$ . If this standardization is not used then an arbitrary unknown scale is  
 17 introduced to the model, with no means of recovering it. By fixing the mean and  
 18 variance, we have a better understanding of the properties of the link and we do  
 19 approach the probit case in the neighborhood of  $\alpha = 0$ .

---

## 2.2. Issue 2 - The quantile intercept and identifiability of parameters

---

1        The identifiability of the parameters in skewed probit regression were first  
 2 investigated by [25]. They showed that without the presence of a continuous  
 3 covariate, the intercept  $\beta_0$ , and skewness parameters are not identifiable. This  
 4 is expected due to the traditional definition of the skewed probit model (2.1)  
 5 and (2.2). We rectify the formulation of the skewed probit regression intercept,  
 6 by introducing the quantile intercept, and subsequently solve this issue of non-  
 7 identifiability by returning to first principles.

8        In simple linear regression, the intercept is used to calculate the expected  
 9 value of the linear predictor without any effect from covariates. In probit re-  
 10 gression, the intercept contains information about the probability of the event,  
 11 without the effects from covariates. The value of the intercept should not provide  
 12 any information about the other parameters in the model.

13        However, when we introduce a skewness parameter to a symmetric family to  
 14 formulate a skew-symmetric link then we are fundamentally changing the meaning  
 15 of what is traditionally called the intercept of the linear predictor, i.e.  $\beta_0$  in (2.2).

Consider probit regression with one centered covariate  $X$ ,

$$p = \text{Prob}[Y = 1] = \Phi(\beta_0 + \beta_1 X).$$

Now if  $\beta_1 X = 0$ , then

$$q = \text{Prob}[Y = 1] = \Phi(\beta_0),$$

16 which implies that  $\beta_0$  is the  $q^{\text{th}}$  quantile of the standard Gaussian distribution.  
 17 There is thus a one-to-one relationship between  $q$  and  $\beta_0$ . When  $\beta_1 \neq 0$ , then  
 18  $\text{Prob}[Y = 1]$  changes because of  $\beta_1 X$ , without affecting  $\beta_0$ , because  $\Phi$  remains  
 19 the same function. In this sense,  $\beta_0$  is uninformative for  $\beta_1$ .

Conversely, consider skewed-probit regression from (2.1) and (2.3),

$$p = \text{Prob}[Y = 1] = F(\beta_0 + \beta_1 X | \alpha).$$

Here,  $\beta_0$  should, in the same way, be uninformative for  $\beta_1$ . This does not hold  
 because the dependence of  $\alpha$ . We can ensure this, if

$$q = \text{Prob}[Y = 1] = F(\beta_0 | \alpha)$$

20 is constant for varying  $\alpha$ , which is the case if  $\beta_0$  is defined as the  $q^{\text{th}}$  quantile of  
 21 the distribution with CDF  $F$ . Therefore, we reformulate  $\beta_0$  as

$$(2.4) \quad \beta_0(q, \alpha) = F^{-1}(q | \alpha),$$

22 so  $\beta_0$  is the  $q^{\text{th}}$  quantile of  $F(\cdot | \alpha)$ . The quantile level  $q$  is now the unknown  
 23 intercept-parameter instead of  $\beta_0$ .

1 Note that there is (generally) not a one-to-one relationship between  $\beta_0$  and  
 2  $q$  since the  $q^{\text{th}}$  quantile depends on  $\alpha$ . In this new formulation, the intercept  
 3 as defined implicitly by  $q$ , provides no information about  $\beta_1$  and parameters of  
 4  $F(\eta_i|\alpha)$  are identifiable. We return in 5.3 to a numerical study of this issue.

5 This formulation might seem surprising at first sight, but in the case of  
 6 a symmetric link, the intercept is the quantile of a distribution with fixed (no)  
 7 skewness. In the case of the probit or identity links for example, this formulation  
 8 will reduce to the usual intercept parameter since in these cases there is a one-  
 9 to-one relationship between  $\beta_0$  and  $q$ .

10 In terms of implementation in *R-INLA*, the new formulation of the skew  
 11 normal model in terms of  $q$  is available and subsequently, the prior distribution  
 12 for  $q$  can be derived from a corresponding informative  $N(\mu_0, \tau_0)$  prior for  $\beta_0$  in  
 13 the case where  $\alpha = 0$ . This will ensure that the probit and the skewed-probit  
 14 models have comparable priors for their respective "intercept" parameters.

---

### 2.3. Issue 3 - Skewness-related parameters

---

15 It is well-known that the skew-normal likelihood has a (double) singularity  
 16 in the neighbourhood  $\alpha \simeq 0$  [3]. Various adaptations of maximum likelihood  
 17 estimation and some Bayes estimators have been proposed as solutions to this  
 18 singularity. [23] used the Fisher information to propose a reparameterization that  
 19 uses  $\alpha^3$  as the skewness parameter since this solves the double singularity problem  
 20 in the likelihood. In our venture to derive the PC prior for the skewness, we  
 21 derived the Kullback-Leibler divergence (KLD) from the skew-normal link to the  
 22 probit link and noticed the same feature as mentioned in [23]. This resemblance  
 23 is expected since the Fisher information metric is the Hessian of the KLD.

24 From (2.3), the KLD for small  $|\alpha|$  can be found to be

$$\begin{aligned}
 \text{KLD}(\alpha) &= \int f(x|\alpha) \log \frac{f(x|\alpha)}{f(x|\alpha=0)} dx \\
 &= \frac{\pi^2 + 16 - 8\pi}{6\pi^3} \alpha^6 - \frac{144\pi + 3\pi^3 - 38\pi^2 - 168}{6\pi^4} \alpha^8 \\
 &\quad + \frac{-42240\pi - 2560\pi^3 + 16176\pi^2 + 129\pi^4 + 39936}{120\pi^5} \alpha^{10} + \mathcal{O}(\alpha^{12}) \\
 (2.5) \quad &\approx c_1 \alpha^6 + c_2 \alpha^8 + c_3 \alpha^{10}.
 \end{aligned}$$

25 Interestingly, the behavior of  $\alpha$  around  $\alpha = 0$  does not have the usual asymptotics  
 26 (consistency rate of  $\sqrt{n}$ ) since the leading term is  $\alpha^6$ . This implies that the  
 27 estimator of  $\alpha$  in the neighbourhood  $\alpha \simeq 0$ , has a consistency rate  $n^{\frac{1}{6}}$  but a  
 28 skewness parameter  $\gamma = \alpha^3$ , such that  $\alpha = \text{sign}(\gamma) \sqrt[3]{|\gamma|}$ , will have the normal  
 29 asymptotics in the sense that the estimator of  $\gamma$  will be  $\sqrt{n}$  consistent.

30 Even though  $\gamma$  has the usual asymptotic behaviour, the estimate of it is



1 hard to interpret since it does not relate easily to an interpretable property.  
 2 We can instead focus on the more interpretable (standardised) skewness of the  
 3 skew-normal distribution,  $\gamma_1$ , which is a monotone function of  $\gamma$

$$(2.6) \quad \gamma_1 = \frac{(4 - \pi) \left( \sqrt{\frac{2\delta^2}{\pi}} \right)^3}{2 \left( 1 - \frac{2\delta^2}{\pi} \right)^{\frac{3}{2}}},$$

4 where  $\delta = \frac{\alpha}{\sqrt{1+\alpha^2}}$  (and  $\gamma = \alpha^3$ ). The skewness takes values in the interval  
 5  $-0.99527 < \gamma_1 < 0.99527$ , which is correct up to five digits.

6 The question arises if we should formulate a prior for  $\alpha$ ,  $\gamma$  or the skew-  
 7 ness  $\gamma_1$ . If priors are assigned more ad-hoc parameters, this question poses a  
 8 challenge. The PC prior is invariant under reparameterizations [35], implying  
 9 that this framework will produce equivalent priors for  $\alpha$ ,  $\gamma$  and  $\gamma_1$ . They are  
 10 equivalent in the inferential sense, and will produce the same posterior inference.

---

### 3. SKEW-NORMAL MEAN REGRESSION

---

11 In this section we focus on skew-normal regression, although these issues  
 12 also exist in more general skew-symmetric regression models.

13 In the preceding section we mentioned the different parameters that can  
 14 be used to capture the skewness in the skewed probit model, and the proposals  
 15 pertain to the skew-normal regression model as well.

16 Most works on skew-normal regression propose a regression model for the  
 17 location parameter,  $\xi$ , from (1.1). This generalization of Gaussian regression  
 18 seems straightforward but when we keep in mind that the location parameter of  
 19 the Gaussian is equal to the mean, then we can see that regressing through the  
 20 location parameter of the skew-normal is not practical. In the spirit of gener-  
 21 alizing Gaussian regression to skew-normal regression, we should formulate the  
 22 regression model based on the mean. Hence for  $y_i \sim SN(\xi, \omega, \alpha)$  from (1.1),

$$(3.1) \quad E[Y_i] = \eta_i,$$

23 with  $\eta_i$  from (2.2), instead of  $\xi_i = \eta_i$ . Note that here we do not reformulate the  
 24 intercept as in Section 2.2 for skewed probit regression, since the identity link  
 25 function is used. We illustrate the proposed skew-normal regression model in  
 26 Section 6.

---

#### 4. PENALIZING COMPLEXITY PRIOR FOR THE SKEWNESS PARAMETER

---

1        The work of [35] introduced the notion of penalizing complexity priors for  
 2 parameters and provided the framework for deriving priors that quantify the  
 3 contraction from a complex model to a simpler model. These PC priors are  
 4 especially helpful and very needed in cases where priors have traditionally been  
 5 chosen due to mathematical convenience, or convention (see [24] for more details  
 6 on the performance of PC priors). PC priors have been used in various fields of  
 7 research, for example [36] derived the PC priors for autoregressive models while  
 8 [20] derived PC priors for Gaussian random fields.

9        In this section we derive the PC prior for  $\alpha$  due to the invariance of the PC  
 10 prior under reparameterization of the skewness parameter. The derivations of the  
 11 PC prior for  $\gamma$  and  $\gamma_1$  follows then directly from a change-of-variable exercise.

12        Using [35] and (2.5), define the uni-directional distance from the skew-  
 13 normal to the Gaussian density as,

$$(4.1) \quad \begin{aligned} d(\alpha) &= \sqrt{2\text{KLD}(\alpha)} \\ &\approx \sqrt{2(c_1\alpha^6 + c_2\alpha^8 + c_3\alpha^{10})} \end{aligned}$$

14        The penalizing complexity prior for the skewness parameter  $\alpha$  is then formed by  
 15 assigning an exponential prior with parameter  $\theta$  to the distance. The parameter  
 16  $\theta$  incorporates information from the user to control the tail behavior and thus the  
 17 rate of contraction towards the probit link function. The penalizing complexity  
 18 prior follows then directly, as

$$(4.2) \quad \begin{aligned} \pi(\alpha) &= \frac{1}{2}\theta \exp[-\theta d(\alpha)] \left| \frac{\partial d(\alpha)}{\partial \alpha} \right| \\ &\approx \frac{\theta}{2\sqrt{2(c_1\alpha^6 + c_2\alpha^8 + c_3\alpha^{10})}} |2(6c_1\alpha^5 + 8c_2\alpha^7 + 10c_3\alpha^9)| \\ &\times \exp\left[-\theta|\alpha^3|\sqrt{2(c_1 + c_2\alpha^2 + c_3\alpha^4)}\right]. \end{aligned}$$

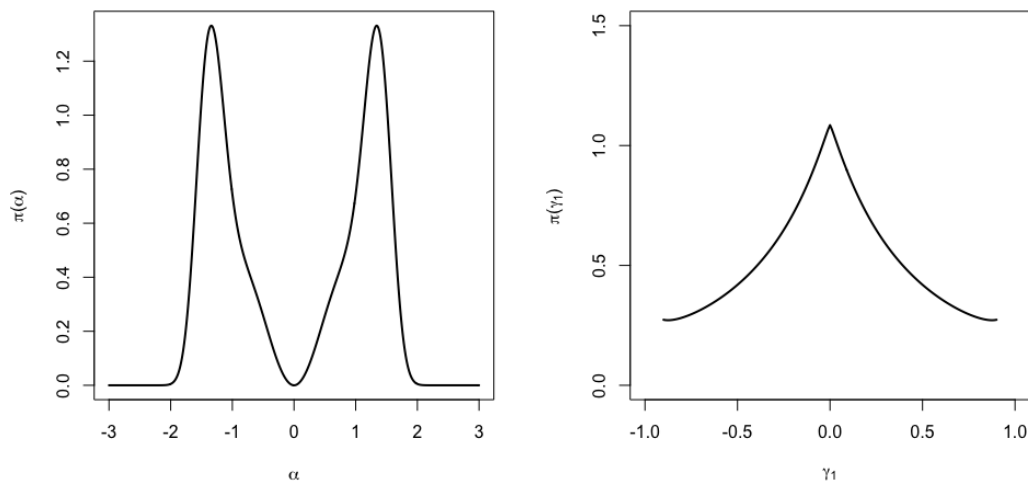
19        for small values of  $|\alpha|$ . The user-defined parameter  $\theta$  is used to govern the con-  
 20 traction towards probit regression, e.g., for small  $p_U > 0$ ,

$$\text{Prob}(d(\alpha) > U) = p_U = \exp(-\theta U)$$

21        which gives  $\theta = -\log p_U/U$ . There is no explicit expression for the penalizing  
 22 complexity prior of  $\alpha$  in general, but the prior can be computed numerically. The  
 23 prior for  $\gamma_1$  is available in the *R-INLA* package[32] with `prior = "pc.sn"` and  
 24 parameter `param=theta`. We use the  $\gamma_1$  reparameterization, since  $\gamma_1$  quantifies the  
 25 skewness as a *property* with good interpretation.

26        The PC priors of  $\alpha$  and  $\gamma_1$  are illustrated in Figure 1 for  $\theta = 5$ , on the  
 27  $\alpha$  and  $\gamma_1$  scales. In Figure 2 various values for  $\theta$  are considered to provide an

1 intuition about the effect of  $\theta$ . From this Figure it is clear that larger values of  
 2  $\theta$  results in higher contraction rates with little mass away from 0. The posterior  
 3 inference of the skewness is not sensitive to the value of  $\theta$  for moderate and large  
 4 samples. In the case of small samples, a very large value of  $\theta$  will contract the  
 5 Bayes estimator towards 0 at a fast rate.



**Figure 1:** PC prior (4.2) for  $\theta = 5$  on the  $\alpha$  scale (left) and the  $\gamma_1$  scale (right)

6

7 From Figure 1 we can see the shape of the PC prior for  $\alpha$  is quite peculiar, but  
 8 has a clear interpretation in terms of a prior on the distance. It just shows that  
 9 if we assign priors to parameters, like  $\alpha$ , instead of to a property, like  $\gamma_1$ , it is  
 10 highly improbable that we could think of a density function for the parameter  
 11 that has good translatable properties. Another interesting note is that from the  
 12 prior density of  $\alpha$  around  $\alpha = 0$ , we can see that most priors of  $\alpha$  proposed in  
 13 literature actually results in underfitting, instead of the usual overfitting, since  
 14 they assign too much density to the neighborhood around  $\alpha = 0$ . Conversely, the  
 15 PC prior of  $\gamma_1$  is as expected with a mode at the value for the probit link.

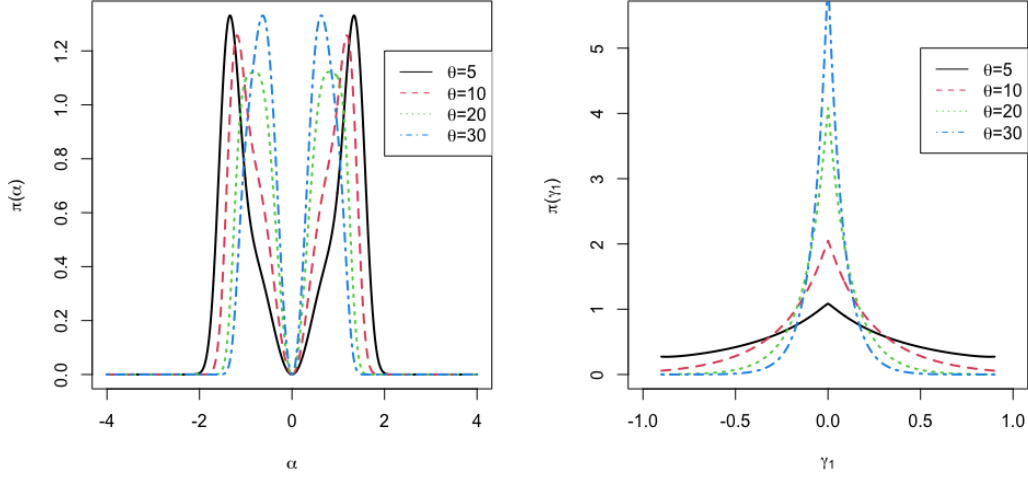
16

---

## 5. SIMULATION STUDY

---

17 In this section we present condensed results from a simulation study with  
 18 the aim to show the results proposed in this work for experiments with a large



**Figure 2:** PC prior (4.2) for various  $\theta$ 's on the  $\alpha$  scale (left) and the  $\gamma_1$  scale (right)

1 and small number of trials. The setup is to simulate linear predictors  $\eta_i =$   
 2  $\beta_0(\alpha, q) + \beta_1 x_i$ , where  $x_i \sim N(0, 0.5)$  for  $i = 1, \dots, n$ . The success probabilities  
 3 are then  $p_i = F(\eta_i | \alpha)$  from (2.1) and subsequently the response variable  $y_i$ ,  
 4 where  $y_i \sim \text{Bin}(N_i, p_i)$ . To investigate the performance of the PC prior for the  
 5 skewness, we consider the PC prior as well as a weak Gaussian prior. Throughout  
 6 this simulation study, we assume  $\theta = 5$  for the PC prior and a weak Gaussian  
 7 prior with parameters  $(0, 10^2)$  for the skewness.

---

### 5.1. Large number of trials

---

8 For an experiment that consists of a large number of trials, we consider  
 9 four simulation scenario's which can be summarized as:

- 10 1.  $q = \frac{1}{3}, \beta_1 = 1, \gamma_1 = 0(\alpha = 0), N_i = 200$
- 11 2.  $q = 0.25, \beta_1 = -1, \gamma_1 = \frac{2}{3}(\alpha = 10), N_i = 200$
- 12 3.  $q = 0.30, \beta_1 = 1, \gamma_1 = \frac{1}{3}(\alpha = 2), N_i = 200$
- 13 4.  $q = 0.10, \beta_1 = -1, \gamma_1 = -\frac{1}{3}(\alpha = -2), N_i = 200$

14 In each case we consider the PC prior as well as the Gaussian prior for the  
 15 skewness  $\gamma_1$ , and weakly informative Gaussian priors for the fixed effects.

---

### 5.1.1. Results

---

1        The fixed effects were recovered well and here we focus on the skewness  
 2  $\gamma_1$ . From Table 1 it is clear that the PC prior (and the Gaussian prior) performs  
 3 as expected since the sample size and number of trials are large. In Figure 3  
 4 the posterior results for the skewness are summarised with coverage probability  
 5 and median length of the credible interval. The results for other scenarios are  
 6 similar and omitted here. From this (and many other) simulation studies, we  
 7 conclude that for a large number of trials the skewed-probit link works well and  
 8 the inference is accurate. It is clear that the PC prior does not contract towards  
 9 the probit model when the data presents strong support for the skewed probit  
 model (scenarios 2,3 and 4).

Scenario	PC prior		Gaussian prior	
	CP	MLCI	CP	MLCI
<b>1</b>	95	0.28	94	0.35
<b>2</b>	96	0.28	97	0.34
<b>3</b>	95	0.31	95	0.34
<b>4</b>	95	0.32	95	0.35

**Table 1:** Coverage probability (CP) and median length of the credible interval (MLCI) for the skewness  $\gamma_1$  under the PC and Gaussian (G) priors, for large  $N_i$

10

---

### 5.2. Small number of trials

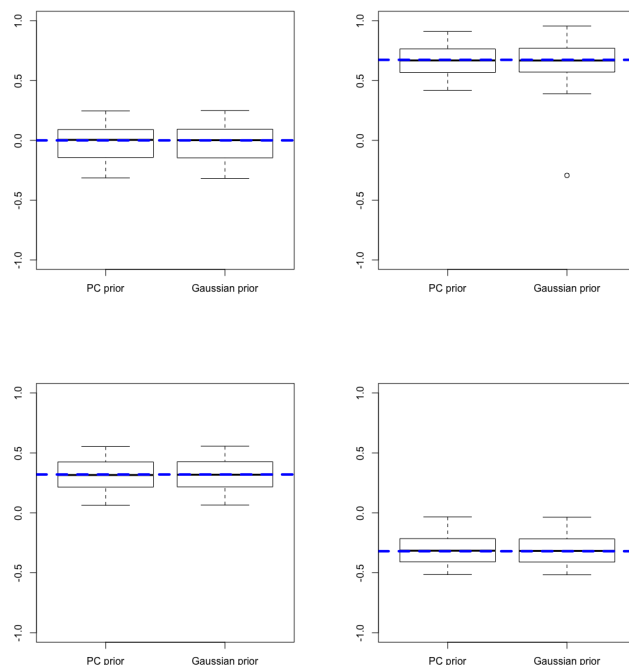
---

11        Here we focus our attention on samples of size 200 of binary trials, and the  
 12 scenario's we consider are:

13 1.  $q = \frac{1}{2}, \beta_1 = 1, \gamma_1 = -\frac{2}{3}(\alpha = -10), N_i = 1$

14 2.  $q = \frac{1}{2}, \beta_1 = 1, \gamma_1 = 0(\alpha = 0), N_i = 1$

15 We consider the PC prior as well as the Gaussian prior for the skewness param-  
 16 eter, and weakly informative Gaussian priors for the fixed effects.



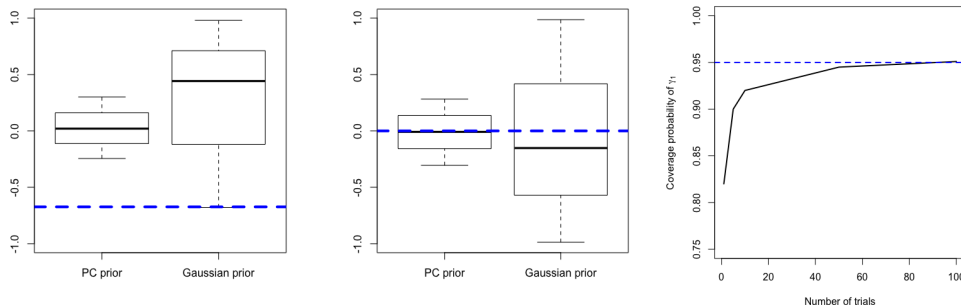
**Figure 3:** Median of 95% credible intervals for the different scenario's with the true skewness (dashed line): Scenario 1, 2 (top left to tight), 3 and 4 (bottom left to right)

---

### 5.2.1. Results

---

1 From Table 2 it is clear that the skewness is not recovered well for a small  
 2 number of trials. In the case of the PC prior, the coverage is poor but the  
 3 credible intervals are still relatively narrow. For the Gaussian prior, the coverage  
 4 is high mainly due to the very wide credible intervals. For a small number of  
 5 trials or binary trials, the skewness is hard to capture. Even though the nominal  
 6 coverage for the Gaussian prior is still high from Table 2, the median length of  
 7 the credible interval implies that the credible intervals span most of the support  
 8 of  $\gamma_1$ . However, the PC prior contracts to zero with relatively narrow credible  
 9 intervals and exhibits poor coverage for  $\gamma_1 \neq 0$ . It is evident that the skewness is  
 10 hard to estimate with a small number of trials. This is not unexpected since in  
 11 binary data, we only observe a success or failure for each subject and subsequently  
 12 the data does not provide sufficient information about the skewness. We need  
 13 repetitions in the data to learn more about the skewness. We can see in Figure 4  
 14 that the PC prior contracts to zero if there is not enough evidence for the skewed  
 15 link, but the Gaussian prior proposes an arbitrary value for the skewness from  
 16 most of the range of  $\gamma_1$  (possibly with the wrong sign as in Figure 4)]. In this



**Figure 4:** 95% credible intervals for  $\gamma_1$  with  $n_i = 1$  and  $\gamma_1 = -\frac{2}{3}$  (left) or  $\gamma_1 = 0$  (middle). Coverage probabilities for  $\gamma_1$  under scenario 1 as  $N_i$  increases (right)

1 case, using the skewed-probit link for binary data might not be useful.

Scenario	PC prior		Gaussian prior	
	CP	MLCI	CP	MLCI
<b>1</b>	65	0.41	90	1.24
<b>2</b>	95	0.33	90	1.45

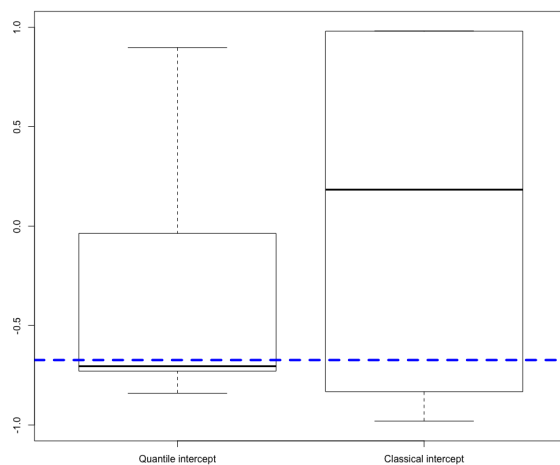
**Table 2:** Coverage probability (CP) and median length of the credible interval (MLCI) for the skewness  $\gamma_1$  under the PC and Gaussian (G) priors, for small  $N_i$

---

### 5.3. Confounding and the effect of the quantile intercept

---

2 In this section we look at the effect of not using the new quantile in-  
 3 tercept. We used a simulated dataset, similar to the preceding section, with  
 4  $q = 0.4, \beta_1 = 0.1, \gamma_1 = -\frac{2}{3}$ . In this setup the linear predictor is close to zero,  
 5 for a centered covariate, the confounding between the classical intercept and the  
 6 skewness parameter is clear. In Figure 5 the median of the 95% credible intervals  
 7 of the skewness (for 500 repetitions) as well as the true value of the skewness  
 8 are presented. On the left we have the case of the quantile intercept and on the  
 9 right, the classical intercept. By using the classical intercept, as in the case of  
 10 GLM, the skewness is not estimated correctly in the sense that the direction is  
 11 not even recovered. It is clear that the quantile intercept solves the confounding  
 12 of the intercept of the linear predictor, with the skewness of the link.



**Figure 5:** Median credible intervals for the skewness  $\gamma_1$  using the quantile intercept vs the classical intercept

---

## 6. APPLICATIONS

---

1        In this section we illustrate the use of skewed probit regression with the  
 2 PC prior using two well-known datasets, the beetle mortality data [10] (binomial  
 3 response with multiple trials) and the UCI Cleveland heart disease data [18]  
 4 (Bernoulli response). We also present the analysis of the Wines data to illustrate  
 5 the use of this work in the skew-normal likelihood.

---

### 6.1. Beetle mortality data

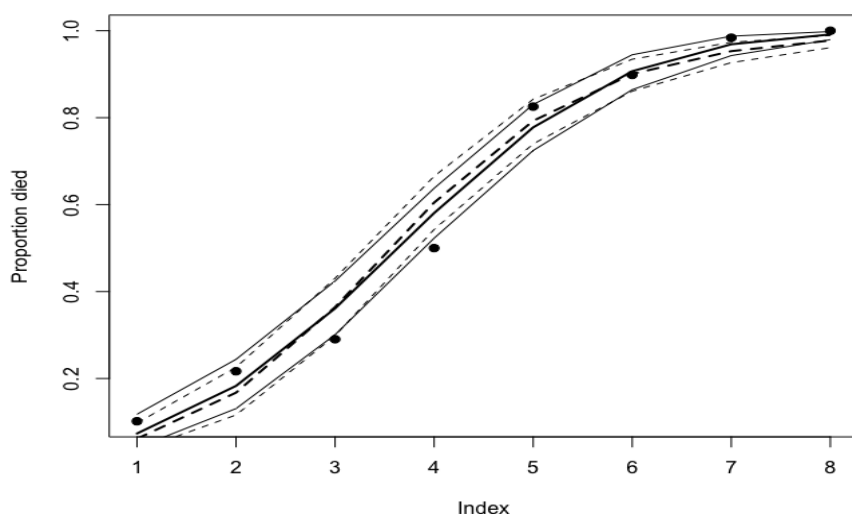
---

6        In this well-known dataset from [15] the number of adult flour beetles killed  
 7 by differing dosages of poison is modelled based on the centered dosage value.  
 8 We use the proposed skewed probit model with the PC prior and the quantile  
 9 intercept. We also fit a probit model and compare the fitted values of both with  
 10 the observed data. These, together with the 95% credible intervals are presented  
 11 in Figure 6. We note that the skewed probit model seem to fit the observed data  
 12 better than the probit model, and the 95% credible interval for the skewness of  
 13 the skewed probit model from Table 3 does not include 0. The marginal log-  
 14 likelihood for the skewed probit model is  $-21.75$  versus  $-23.93$  from the probit  
 15 model. The difference between the marginal log-likelihoods does not provide a  
 16 convincing argument in favor of the skewed probit model, as opposed to the probit  
 17 model.



Effect	Estimate	95% credible interval
Quantile of the intercept ( $q$ )	0.643	(0.572; 0.703)
Dosage	19.132	(16.074; 22.316)
Skewness ( $\gamma_1$ )	-0.456	(-0.848; -0.053)

**Table 3:** Posterior estimates for the beetle mortality data



**Figure 6:** Fitted and observed proportions (— Skewed Probit, - - Probit) with 95% credible intervals

1

---

## 6.2. Heart disease data

---

2 We will use the Cleveland data obtained by Robert Detrano from the V.A.  
3 Medical Center, Long Beach and Cleveland Clinic Foundation.

4 The response is a binary observation indicating the occurrence of a  $> 50\%$   
5 diameter narrowing in an angiography. Various covariates are available in this  
6 data and we will use a subset of these namely, gender (male/female), type of  
7 chest pain (1 - typical angina, 2 - atypical angina, 3 - non-anginal pain, 4 -  
8 asymptomatic), resting blood pressure, the slope of the peak exercise ST segment  
9 (1 - upsloping, 2 - flat, 3 - down sloping), the number of colored vessels by  
10 fluoroscopy and the results from the thallium heart scan (3 - normal, 6 - fixed  
11 defect, 7 - reversable defect). We centered the two continuous covariates, resting  
12 heart rate and the number of colored vessels by fluoroscopy. Further details can

1 be found in [18].

2 There are 297 subjects with complete information in the dataset of which  
 3 137 experienced the event of  $> 50\%$  diameter narrowing in an angiography. We  
 4 fit a skewed-probit regression model to explain the probability of the event based  
 5 on the values of the covariates similar to [25]. In [25] divergent results were  
 6 obtained based on different estimation frameworks, namely maximum likelihood  
 7 estimation, bootstrap bias correction, Jeffrey’s prior, generalized information ma-  
 8 trix prior and Cauchy prior penalized frameworks. The inconsistent results could  
 9 be attributed to the issues we mentioned in this paper, since all these estima-  
 10 tion methods were developed for the skewed-probit regression model without the  
 11 good standardization, based on the skewness parameter  $\alpha$  and defined using the  
 12 classical intercept.

13 Also, there is a lack of information on the skewness in binary data. The  
 14 consequence is thus that various values of the skewness could be supported. This  
 15 case is a prime example that illustrates the need for the PC prior of the skewness,  
 16 so that we prefer zero skewness a priori (probit regression) and use the data to  
 17 advocate for non-trivial skewness (skewed probit regression).

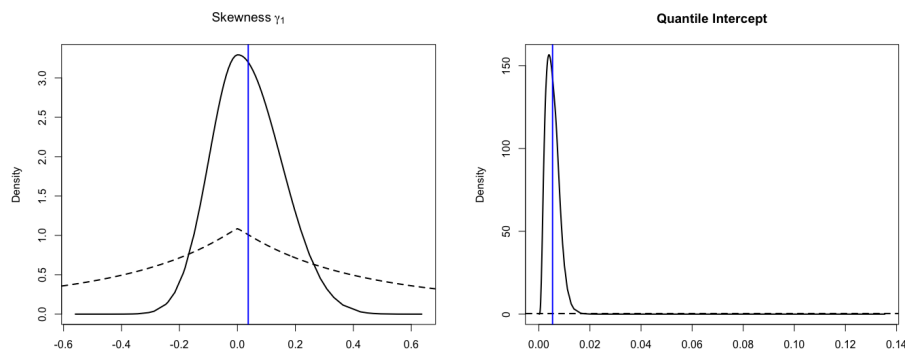
18 Here, we can use the PC prior (4.2) for the skewness and the quantile  
 19 intercept from Section 2.2. All quantitative covariates are centered. The results  
 20 are given in Table 4.

	Posterior mean	95% credible interval
<b>Quantile Intercept (<math>q</math>)</b>	0.045	(0.006; 0.184)
<b>Gender (male)</b>	1.025	(0.605; 1.461)
<b>Type of chest pain (2)</b>	0.198	(-0.538; 0.942)
<b>Type of chest pain (3)</b>	-0.074	(-0.732; 0.590)
<b>Type of chest pain (4)</b>	1.288	(0.673; 1.920)
<b>Resting heart rate</b>	0.016	(0.005; 0.027)
<b>Slope of the peak exercise (2)</b>	1.027	(0.637; 1.452)
<b>Slope of the peak exercise (3)</b>	0.791	(0.059; 1.540)
<b>Number of colored vessels</b>	0.704	(0.477; 0.945)
<b>Skewness (<math>\gamma_1</math>)</b>	0.02	(-0.214; 0.235)

**Table 4:** Results for the Cleveland heart disease data

21 From the estimate of  $\gamma_1$  in Table 4 we deduce that the skewness is not  
 22 supported by the data and a probit regression model could be sufficient. We did  
 23 the analysis using probit regression and the inference is very similar. This result  
 24 of zero skewness coincides with the skewness estimates in [25] using the MLE,  
 25 bootstrap correction, generalized information matrix and cauchy prior penaliza-  
 26 tion approaches. The posterior densities (and prior densities in dashed) of the  
 27 skewness,  $\gamma_1$ , and quantile intercept,  $q$ , are presented in Figure 7.

28 We also see that being a male, having asymptomatic chest pain, higher



**Figure 7:** Posterior (prior - dashed) density of the skewness  $\gamma_1$  (left) and quantile intercept  $q$  (right) with the corresponding point estimates (vertical line)

1 resting heart rate, a flat or downwards slope of the peak exercise ST segment and  
 2 more colored vessels by fluoroscopy, all contribute to a higher probability of the  
 3 event under investigation, i.e.  $> 50\%$  diameter narrowing in an angiography.

4 The posterior densities (and prior densities in dashed) of the fixed effects  
 5 are presented in Figure 8.

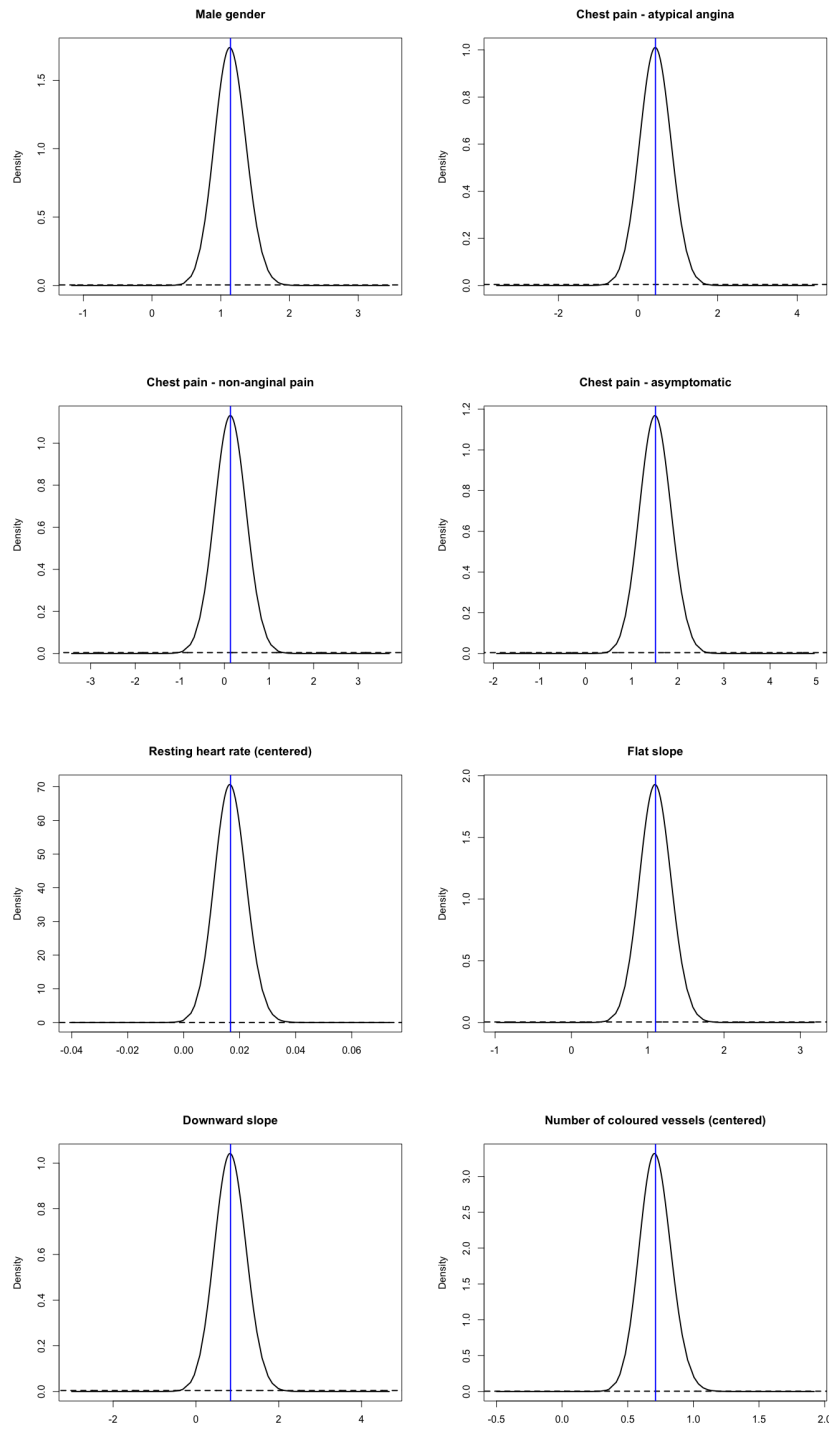
6 We calculated the marginal log-likelihoods for the probit and skewed-probit  
 7 models to be  $-150.62$  and  $-158.41$ , respectively, indicating that the probit model  
 8 is preferred by the data. Both models achieved a correct classification percentage  
 9 of  $84.55\%$ , on a  $50\%$  holdout sample.

---

### 6.3. Wines data

---

10 This section illustrates the new results when the response variable is contin-  
 11 uous and assumed to follow a skew-normal distribution. As mentioned in Section  
 12 3, the results derived in this paper hold for skewed-probit models, as well as skew-  
 13 normal regression models. We use the wines dataset from [6], where the acidity of  
 14 the wine is assumed to follow a skew-normal distribution as illustrated in Figure  
 15 9, where we see the tail behaviour is correctly captured by the fitted Gaussian  
 16 density, but not the skewness. The mean acidity (not the location parameter) is  
 17 modelled using the type of wine, sugar content and pH level as covariates (after  
 18 backwards elimination). We assign PC priors for the precision [35] as well as  
 19 skewness (4.2). The results are given in Table 5. The marginal log-likelihood for  
 20 the skew-normal model is  $-722.21$  and for the Gaussian model it is  $-724.59$ .



**Figure 8:** Posterior (prior - dashed) densities of the fixed effects with the corresponding point estimate (vertical line)

	Posterior mean	95% credible interval
<b>Intercept</b>	77.053	(73.824; 80.252)
<b>Wine (Grignolino)</b>	5.088	(0.478; 9.693)
<b>Wine (Barbera)</b>	23.613	(19.003; 28.280)
<b>Sugar</b>	3.118	(1.150; 5.080)
<b>pH</b>	-8.350	(-10.122; -6.574)
<b>Skewness (<math>\gamma_1</math>)</b>	0.439	(0.128; 0.702)
<b>Precision for the data</b>	0.008	(0.006; 0.009)

Table 5: Results for the wines data

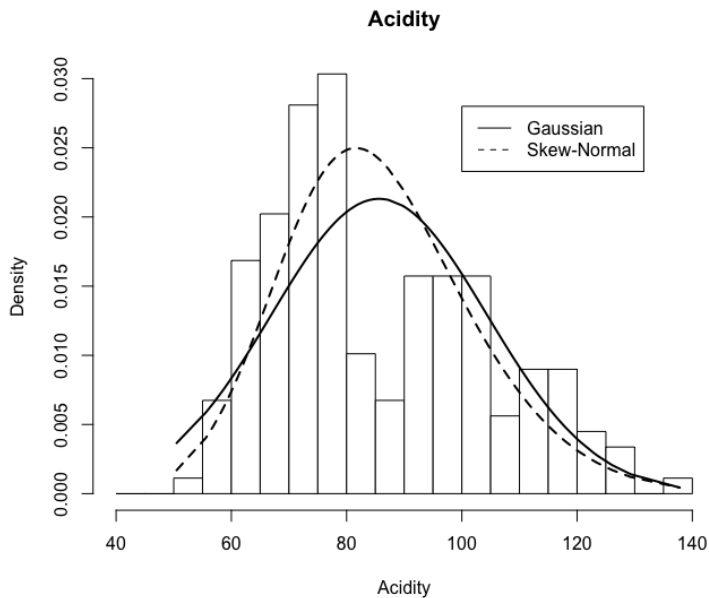


Figure 9: Histogram with model-based Gaussian curve and skew-normal curve

---

## 7. DISCUSSION

---

1        The use of skew-symmetric distributions or links is popular due to the per-  
2        ceived flexibility inherited through the extra parameter that controls the skew-  
3        ness. The skew normal skewness parameter in particular, poses various challenges  
4        in the inference thereof. As we set out with the initial aim to derive the penalizing  
5        complexity prior for the skewness, we realized that there are various other issues  
6        that we could not find addressed in the literature. It is apparent that with  
7        the generalizing to skew-symmetric distributions and links from the symmetric  
1        counterparts, various fundamental concepts have gone amiss.

2 Here we rectify the formulation of the intercept in the linear predictor of  
 3 all skew-symmetric links, firstly to ensure that it behaves as an intercept and  
 4 secondly due to the confounding with the skewness parameter and fixed effects.  
 5 We also show that the popular method of standardizing the skewed link function  
 6 by inheriting the parameter values of the symmetric link, fundamentally changes  
 7 the way the link function maps the data to the linear predictor, and we provide  
 8 an anchored standardization approach. We believe that many of the contradict-  
 9 ing works in this area can be attributed to the inappropriate use of the classical  
 10 intercept and parameter-based standardization, instead of property-based stan-  
 11 dardization. In skew-symmetric regression models, we formulate the regression  
 12 model based on the mean, instead of the location parameter.

13 After the fundamental corrections to the formulation of the skewed-probit  
 14 link, the penalizing complexity prior for the skewness was derived. One particular  
 15 advantage of this prior is that it is invariant to reparameterizations of the skew-  
 16 ness parameter. In light of this, we implemented the PC prior for the skewness in  
 17 *R-INLA* [32] for use by others. We noted, expectedly, that binary data (or with  
 18 few trials) does not provide information about the skewness, and we thus advise  
 19 against the use of the skewed-probit link for data with a small number of trials.  
 20 We advocate the use of the PC prior even more feverently because of this feature,  
 21 since the PC prior will contract to the simpler probit link instead of providing an  
 22 incorrect unreliable estimate of the skewness. Other inferential frameworks might  
 23 not be able to ensure this contraction in the absence of convincing evidence from  
 24 the data about the necessary skewness, and could lead to unfounded complicated  
 25 models.

26 We hope that the issues raised and addressed here will improve the inference  
 27 of the skewed probit model (and more broadly the skew-symmetric links and  
 28 likelihoods) and provide insights into the fundamental considerations necessary  
 29 when distributions or links are generalized.

---

## Appendix

---

30 We give here a small example for how to do skew probit regression in *R-*  
 31 *INLA*. In the code below, the unusual statement is `remove.names="(Intercept)"`  
 32 which remove the intercept in the formula *after* doing the expansion of factors in  
 33 the model. We need this as we replace the traditional intercept with the quantile  
 34 intercept in the link, and the expansion of factors depends on the presence or  
 35 not, of an intercept in the model.

```
36 library(INLA)
37 n = 200
38 Ntrials = 200
1 x = rnorm(n, sd = 0.5)
```

```

2 eta = x
3 skew <- 0.5
4 prob = inla.link.invsn(eta, skew = skew, intercept = 0.75)
5 y = rbinom(n, size = Ntrials, prob = prob)
6 r = inla(y ~ 1 + x,
7   family = "binomial",
8   data = data.frame(y, x),
9   Ntrials = Ntrials,
10  control.fixed = list(remove.names = "(Intercept)",
11                        prec = 1),
12  control.family = list(
13    control.link = list(model = "sn",
14                        hyper = list(
15                          skew = list(param = 10))))))
16 summary(r)

```

---

## REFERENCES

---

- 17 [1] ARELLANO-VALLE, R.B., BOLFARINE, H. AND LACHOS, V.H. (2007). Bayesian  
18 inference for skew-normal linear mixed models, *Journal of Applied Statistics*, **34**,  
19 6, 663–682.
- 20 [2] AZEVEDO, C.L.N., BOLFARINE, H. AND ANDRADE, D.F. (2011). Bayesian infer-  
21 ence for a skew-normal IRT model under the centered parameterization, *Compu-  
22 tational Statistics & Data Analysis*, **55**, 1, 141–163.
- 23 [3] AZZALINI, A. (1985). A class of distributions which includes the normal ones,  
24 *Scandinavian Journal of Statistics*, 171–178.
- 25 [4] AZZALINI, A. AND CAPITANIO, A. (2003). Distributions generated by perturba-  
26 tion of symmetry with emphasis on a multivariate skew t-distribution, *Journal of  
27 the Royal Statistical Society: Series B (Statistical Methodology)*, **65**, 2, 367–389.
- 28 [5] AZZALINI, A. AND ARELLANO-VALLE, R.B. (2013). Maximum penalized likeli-  
29 hood estimation for skew-normal and skew-t distributions, *Journal of Statistical  
30 Planning and Inference*, **143**, 2, 419–433.
- 31 [6] AZZALINI, A., (2013). *The skew-normal and related families*, 3rd ed, Cambridge  
32 University Press.
- 33 [7] BAYES, C.L. AND BRANCO, M.D. (2007). Bayesian inference for the skewness  
34 parameter of the scalar skew-normal distribution, *Brazilian Journal of Probability  
35 and Statistics*, 141–163.
- 36 [8] BAZÁN, J.L., BRANCO, M.D. AND BOLFARINE, H. AND OTHERS (2006). A skew  
37 item response model, *Bayesian Analysis*, **1**, 4, 861–892.
- 38 [9] BAZÁN, J.L., BOLFARINE, H. AND BRANCO, M.D. (2010). A framework for  
39 skew-probit links in binary regression, *Communications in Statistics-Theory and  
40 Methods*, **39**, 4, 678–6972.
- 41 [10] BLISS, C.I. (1935). The calculation of the dosage-mortality curve, *Annals of  
1 Applied Biology*, **22**, 1, 134–167.

- [11] CABRAS, S., RACUGNO, W., CASTELLANOS, M.E. AND VENTURA, L. (2012). A matching prior for the shape parameter of the skew-normal distribution, *Scandinavian Journal of Statistics*, **39**, 2, 236–247.
- [12] CANALE, A., KENNE, P., EULOGE, C. AND SCARPA, B. (2016). Bayesian modeling of university first-year students' grades after placement test, *Journal of Applied Statistics*, **43**, 16, 3015–3029.
- [13] CASTRO, L.M., MARTÍN, E.S. AND ARELLANO-VALLE, R.B. (2013). A note on the parameterization of multivariate skewed-normal distributions, *Brazilian Journal of Probability and Statistics*, 110–115.
- [14] CHEN, M., DEY, D.K. AND SHAO, Q. (1999). A new skewed link model for dichotomous quantal response data, *Journal of the American Statistical Association*, **94**, 448, 1172–1186.
- [15] COLLET, D., (2003). *Modelling Binary Data. , 2nd ed*, Chapman & Hall/CRC, Boca Raton , FL.
- [16] CZADO, C. AND SANTNER, T.J. (1992). The effect of link misspecification on binary regression inference, *Journal of Statistical Planning and Inference*, **33**, 2, 213–231.
- [17] DETTE, H., LEY, C. AND RUBIO, F. (2018). Natural (Non-) Informative Priors for Skew-symmetric Distributions, *Scandinavian Journal of Statistics*, **45**, 2, 405–420.
- [18] DUA, D. AND GRAFF, C. (2017). UCI Machine Learning Repository, *University of California, Irvine, School of Information and Computer Sciences*, <http://archive.ics.uci.edu/ml>.
- [19] FIRTH, D. (1993). Bias reduction of maximum likelihood estimates, *Biometrika*, **80**, 1, 27–38.
- [20] FUGLSTAD, G.A., SIMPSON, D., LINDGREN, F. AND RUE, H. (2018). Constructing Priors that Penalize the Complexity of Gaussian Random Fields, *Journal of the American Statistical Association*, 1–8.
- [21] GENTON, M.G. (2004). *Skew-elliptical Distributions and their Applications: A Journey beyond Normality*, CRC Press.
- [22] GENTON, M.G. AND ZHANG, H. (2012). Identifiability problems in some non-Gaussian spatial random fields, *Chilean Journal of Statistics*, **3**, 2, 171–179.
- [23] HALLIN, M. AND LEY, C. AND OTHERS (2014). Skew-symmetric distributions and Fisher information: the double sin of the skew-normal, *Bernoulli*, **20**, 3, 1432–1453.
- [24] KLEIN, N. AND KNEIB, T. (2016). Scale-Dependent Priors for Variance Parameters in Structured Additive Distributional Regression, *Bayesian Analysis*, **11**, 4, 1071–1106.
- [25] LEE, D. AND SINHA, S. (2019). Identifiability and bias reduction in the skew-probit model for a binary response, *Journal of Statistical Computation and Simulation*, **89**, 9, 1621–1648.
- [26] LISEO, B. (1990). The skew-normal class of densities: inferential aspects from a Bayesian viewpoint, *Biometrika*, **50**, 59–70.
- [27] LISEO, B. AND LOPERFIDO, N. (2006). A note on reference priors for the scalar skew-normal distribution, *Journal of Statistical Planning and Inference*, **136**, 2, 373–389.



- 3 [28] MAGHAMI, M.M., BAHRAMI, M. AND SAJADI, F.A. (2020). On bias reduction  
4 estimators of skew-normal and skew-t distributions, *Journal of Applied Statistics*,  
5 1–23.
- 6 [29] MARTINS, T.G., SIMPSON, D., LINDGREN, F. AND RUE, H. (2013). Bayesian  
7 computing with INLA: New features, *Computational Statistics and Data Analysis*,  
8 **67**, 68–83.
- 9 [30] O’HAGAN, A. AND LEONARD, T. (1976). Bayes estimation subject to uncertainty  
10 about parameter constraints, *Biometrika*, **63**, 1, 201–203.
- 11 [31] OTINIANO, C.E.G., RATHIE, P.N. AND OZELIM, L.C.S.M. (2015). On the  
12 identifiability of finite mixture of skew-normal and skew-t distributions, *Statistics*  
13 *& Probability Letters*, **106**, 103–108.
- 14 [32] RUE, H., MARTINO, S. AND CHOPIN, N. (2009). Approximate Bayesian inference  
15 for latent Gaussian models by using integrated nested Laplace approximations,  
16 *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **71**, 2,  
17 319–392.
- 18 [33] RUE, H., RIEBLER, A., SØRBYE, S.H., ILLIAN, J.B., SIMPSON, D. AND LIND-  
19 GREN, F. (2017). Bayesian computing with INLA: A review, *Annual Reviews of*  
20 *Statistics and Its Applications*, **4**, 395–421.
- 21 [34] SARTORI, N. (2006). Bias prevention of maximum likelihood estimates for scalar  
22 skew normal and skew t distributions, *Journal of Statistical Planning and Infer-*  
23 *ence*, **136**, 12, 4259–4275.
- 24 [35] SIMPSON, D., RUE, H., RIEBLER, A., MARTINS, T.G. AND SØRBYE, S.H. AND  
25 OTHERS (2017). Penalizing model component complexity: A principled, practical  
26 approach to constructing priors, *Statistical Science*, **32**, 1, 1–28.
- 27 [36] SØRBYE, S.H. AND RUE, H. (2017). Penalised Complexity Priors for Stationary  
28 Autoregressive Processes, *Journal of Time Series Analysis*, **38**, 6, 1467–1492.