Read me file for the method used in ESS_RegressionCalculator.R

This file provides documentation for an R program that performs the computations needed to obtain the effective sample size (ESS) of a parametric prior, as described in the paper "Determining the Effective Sample Size of a Parametric Prior" by Morita, Thall and Muller (Biometrics 64, 595-602, 2008). Please read this paper carefully before using these computer programs. For questions or request for a reprint of the paper, please contact Satoshi Morita at smorita@urahp.yokohama-cu.ac.jp or Peter Thall at rex@mdanderson.org

What is a prior ESS?

Understanding the strength of a prior distribution relative to the likelihood is a fundamental issue when applying Bayesian methods. This issue may be addressed directly by quantifying the prior information in terms of a number of hypothetical observations (prior effective sample size, ESS) that, starting with a very non-informative prior, would give a posterior very close to the original prior. A prior ESS allows one to judge the relative contributions of the prior and the data to the final conclusions.

Computation of a prior ESS

For several commonly used Bayesian models, a computational method is not needed to obtain the numerical prior ESS values. Some important examples are given in the table presented below. For example, as summarized in the first row of the table, a beta(α, β) prior distribution has ESS = α+β for a Bayesian model having a beta prior and a binomial likelihood.

Prior, likelihood, and traditionally reported prior effective sample size, ESS, for some common models.

Prior	Likelihood	ESS
Be (α, β)	Bin (n, θ)	$\alpha+\beta$
Ga (α, β)	Exp (θ)	α
Inv-χ^2 (ν, σ^2)	N (μ, σ^2)	ν
Dir ($\alpha_1,\ldots,\alpha_d$)	Mn (n, $\theta_1,\ldots,\theta_d$)	$\alpha_1+\cdots+\alpha_d$

For many parametric Bayesian models, however, the prior ESS cannot be determined analytically.

ESS_RegressionCalculator.R:

This is a computer program written in R that computes the prior ESS for either a normal linear regression model or a logistic regression model. For both models, denote the covariate vector by $\boldsymbol{X} = (X_1, \cdots, X_d)$ and the linear term by $\eta(\boldsymbol{X}, \boldsymbol{\theta}) = \theta_0 + X_1\theta_1 + \cdots + X_d\,\theta_d$, for parameter vector $\boldsymbol{\theta} = (\theta_0,\ldots,\theta_d)$. In this program, you may include $d \le 10$ covariates. Denote the outcome variable by Y. For the normal linear model, we assume that $Y \mid \boldsymbol{X}, \boldsymbol{\theta}, \tau \sim N\{\eta(\boldsymbol{X}, \boldsymbol{\theta}), 1/\tau\}$, so that $\eta(\boldsymbol{X}, \boldsymbol{\theta})$ is the mean τ is the precision (inverse variance) parameter. For the logistic regression model, the outcome variable Y is binary and the assumed model is

$Pr(Y = 1 | \boldsymbol{X}, \boldsymbol{\theta}) = \pi(\boldsymbol{X}, \boldsymbol{\theta}) = \exp\{\eta(\boldsymbol{X}, \boldsymbol{\theta})\}/[1+ \exp\{\eta(\boldsymbol{X}, \boldsymbol{\theta})\}]$. In both models, either a normal or a gamma prior may be assumed for each entry of $\boldsymbol{\theta}$. In the normal model, a gamma prior must be assumed for τ. All priors are assumed to be mutually independent.

<u>To run the program, you must specify the following input information:</u>

1. The regression model (normal linear or logistic)
2. The number of covariates (at most 10).
3. The prior for each $\theta_0, \cdots, \theta_d$ using either a normal N (μ, σ^2) or a gamma Ga (α, β), including the numerical values of the hyperparameters for each specified prior.
4. A positive integer '*M*' that is an initial value assumed to be larger than the prior ESS.
5. The number of simulations, '*T*'. The default number is T = 5,000. For example, you may use T = 10,000 to obtain very accurate ESS values, or a value of T as low at 1,000 to reduce runtime.
6. If you would like to compute the ESS of a subvector of $\boldsymbol{\theta}$, specify the parameters to be included in the subvector.