Description of algorithms implemented in the R package shrinkage

Gwenaël G.R. Leday
Wageningen University and Research

Abstract

This document provides a detailed description of the algorithms implemented in the R package **shrinkage** available at `github.com/gleday/shrinkage`.

Keywords: regression, posterior, R.

1. Linear regression with global shrinkage priors

This section introduces linear regression models with global shrinkage priors, which are the simplest form of shrinkage priors, and yield ridge-type estimators (Hoerl and Kennard 1970).

1.1. Model and priors

Let y denote a n-dimensional response variable and X an n by p observation matrix. Then, the linear regression model with global shrinkage is:

$$y \mid \beta, \sigma^2 \sim \mathrm{N}_n(X\beta, \sigma^2 I_n) \quad (1)$$
$$\beta \mid \sigma^2, \tau^2 \sim \mathrm{N}_p(0, \tau^2\sigma^2 I_p) \quad (2)$$
$$p(\sigma^2) \propto \sigma^{-2} \quad (3)$$

The R package **shrinkage** offers the following choices of priors for τ^2:

1. $\tau^2 \sim \mathrm{InvGamma}(a, b)$
2. $\tau^2 \sim \mathrm{BetaPrime}(a, b)$
3. $\tau^2 \sim \mathrm{InvGaussian}(a, b)$
4. $\tau^2 \sim \mathrm{Gamma}(a, b)$
5. $\tau^2 = \hat{\tau}^2_{\mathrm{ML}}$

1.2. Closed-form inference

When $\tau^2 = \hat{\tau}^2_{\mathrm{ML}}$ the prior variance τ^2 is not endowed with a prior probability distribution but is instead set to the value that maximizes the marginal likelihood (ML) of the model (empirical Bayes). In such case, Bayesian inference is carried out analytically as the marginal posterior distributions of β and σ^2 are available in closed-form:

$$\beta \mid y \sim \mathrm{T}_p\left(\bar{\beta}, \frac{y^T y - \bar{\beta}^T \bar{\Sigma}^{-1} \bar{\beta}}{n} \bar{\Sigma}, n\right) \quad \text{and} \quad \sigma^2 \mid y \sim \mathrm{InvGamma}\left(\frac{n}{2}, \frac{1}{2}\left[y^T y - \bar{\beta}^T \bar{\Sigma}^{-1} \bar{\beta}\right]\right).$$

Here

$$\bar{\Sigma} = \left(X^T X + \tau^{-2} I_p\right)^{-1} \quad \text{and} \quad \bar{\beta} = \bar{\Sigma} X^T y.$$

The marginal likelihood is also available in closed-form:

$$p(y) = \pi^{-\frac{n}{2}} \left(\tau^2\right)^{-\frac{p}{2}} \mid X^T X + \tau^{-2} I_p \mid^{-\frac{1}{2}} \Gamma\left(\frac{n}{2}\right) \left(\frac{1}{2} y^T y - \frac{1}{2} \bar{\beta}^T \bar{\Sigma}^{-1} \bar{\beta}\right)^{-\frac{n}{2}},$$

and its maximizer,

$$\hat{\tau}^2_{\mathrm{ML}} = \arg\max_{\tau^2} \ \log p(y),$$

obtained very efficiently (Karabatsos 2018) when re-writing the log-marginal likelihood

$$\log p(y) = -\frac{n}{2} \log \pi - \frac{p}{2} \log \tau^2 - \frac{1}{2} \sum_{r=1}^{q} \log(d_r^2 + \tau^{-2}) + \log \Gamma\left(\frac{n}{2}\right) - \frac{n}{2} \log\left(\frac{1}{2} y^T y - \frac{1}{2} \sum_{r=1}^{q} \frac{\hat{\theta}_r^2 d_j^4}{d_r^2 + \tau^{-2}}\right),$$

in terms of the singular values d_r, $r = 1, \ldots, q = \min(n, p)$, of $X = UDV^T = FV^T$ and the maximum likelihood estimate $\hat{\theta} = (F^T F)^{-1} F^T y = D^{-1} U^T y$ of $\theta = V^T \beta$.

1.3. Inference using a Gibbs sampler

The R package **shrinkage** provides approximate inference for priors 1-4 using a Gibbs sampler that samples very efficiently from the posterior conditional distributions of parameters, and provides closed-form inference for prior 5.

Posterior conditional distributions

Regardless of the prior on τ^2, the posterior conditional distributions of β and σ^2 are:

$$\begin{aligned} \beta \mid y, \tau^2, \sigma^2 &\sim \mathrm{N}_p\left(\bar{\beta},\ \sigma^2 \bar{\Sigma}\right), \\ \sigma^2 \mid y, \beta, \tau^2 &\sim \mathrm{InvGamma}\left(\frac{n+p}{2},\ \frac{1}{2}\left[\tau^{-2} \beta^T \beta + (y - X\beta)^T (y - X\beta)\right]\right). \end{aligned}$$

In contrast, the posterior conditional distribution of τ^2 depends on the choice of prior distribution, namely:

1. when $\tau^2 \sim \mathrm{InvGamma}(a, b)$, the posterior conditional distribution of τ^2 is:

$$\tau^2 \mid \beta, \sigma^2 \sim \mathrm{InvGamma}\left(a + \frac{p}{2},\ \frac{1}{2} \sigma^{-2} \beta^T \beta + b\right).$$

2. when $\tau^2 \sim \mathrm{BetaPrime}(a, b)$ two parametrizations can be used:

- the Gamma-Gamma representation of the beta prime distribution (Zhang, Reich, and Bondell 2016):

  $$\tau^2 \sim \text{BetaPrime}(a, b) \quad \Leftrightarrow \quad \tau^2 \mid \gamma^2 \sim \text{Gamma}(a, \gamma^2),\ \gamma^2 \sim \text{Gamma}(b, 1),$$

  yields the posterior conditional distributions:

  $$\begin{aligned} \tau^2 \mid \beta, \sigma^2, \gamma^2 &\sim \text{GIG}\left(\sigma^{-2}\beta^T\beta,\ 2\gamma^2,\ a - \frac{p}{2}\right), \\ \gamma^2 \mid \tau^2 &\sim \text{Gamma}\left(a + b,\ \tau^2 + 1\right). \end{aligned}$$

- the inverse-Gamma-inverse-Gamma representation of the beta prime distribution (Schmidt and Makalic 2019, Proposition 1):

  $$\tau^2 \sim \text{BetaPrime}(a, b) \quad \Leftrightarrow \quad \tau^2 \mid \gamma^2 \sim \text{InvGamma}(b, 1/\gamma^2),\ \gamma^2 \sim \text{InvGamma}(a, 1),$$

  yields the posterior conditional distributions:

  $$\begin{aligned} \tau^2 \mid \beta, \gamma^2, \sigma^2 &\sim \text{InvGamma}\left(b + \frac{p}{2},\ \frac{\beta^T\beta}{2\sigma^2} + \frac{1}{\gamma^2}\right) \\ \gamma^2 \mid \tau^2 &\sim \text{InvGamma}\left(a + b,\ 1 + \frac{1}{\tau^2}\right) \end{aligned}$$

3. when $\tau^2 \sim \text{InvGaussian}(a, b)$ the posterior conditional distribution of τ^2 is:

   $$\tau^2 \mid \beta, \sigma^2 \sim \text{GIG}\left(b + \sigma^{-2}\beta^T\beta,\ \frac{b}{a^2},\ -\frac{1}{2} - \frac{p}{2}\right).$$

4. when $\tau^2 \sim \text{Gamma}(a, b)$ the posterior conditional distribution of τ^2 is:

   $$\tau^2 \mid \beta, \sigma^2 \sim \text{GIG}\left(\sigma^{-2}\beta^T\beta,\ 2b,\ a - \frac{p}{2}\right).$$

Algorithm

Algorithm 1 describes a very fast Gibbs sampler, obtained from the conditional distributions described above, that is used in the R package **shrinkage** to fit linear regression models with global shrinkage priors 1-4. Note that the algorithm is different than standard Gibbs samplers on the following aspect: instead of sampling β at iteration i from

$$\beta^{(i)} \sim \text{N}_p\left(\left(X^TX + \frac{I_p}{\tau^{2(i-1)}}\right)^{-1} X^Ty, \sigma^2\left(X^TX + \frac{I_p}{\tau^{2(i-1)}}\right)^{-1}\right),$$

using the procedures of Rue (2001) (when $n > p$) or Bhattacharya, Chakraborty, and Mallick (2016) (when $n \leq p$), Algorithm 1 uses the singular value decomposition of X to sample

$$\theta_j^{(i)} \sim \text{N}\left(\frac{d_j u_j^T y}{d_j^2 + \tau^{2(i-1)}}, \frac{\sigma^2}{d_j^2 + \tau^{2(i-1)}}\right), \quad \text{for } j = 1, \ldots, q,$$

and substitute ${\beta^{(i)}}^T\beta^{(i)}$ and $X\beta^{(i)}$ with ${\theta^{(i)}}^T\theta^{(i)}$ and $F\theta^{(i)}$, respectively. Samples of β are then obtained upon convergence (i.e. when $i = n_{\text{iter}}$) using $\beta^{(i)} = V\theta^{(i)}$, for $i = 1, \ldots, n_{\text{iter}}$. In our experience, the SVD decomposition provides considerable computational speed-ups while giving similar results to standard samplers.

Initialize:
$a = b = 0.5$, $\tau^{2(0)} = \gamma^{2(0)} = 1$, $n_{\text{mcmc}} = 1000$, $n_{\text{burnin}} = 1000$, $n_{\text{iter}} = n_{\text{mcmc}} + n_{\text{burnin}}$
for $i = 1$ **to** n_{iter} **do**
    for $j = 1$ **to** q **do**
        sample $\theta_j^{(i)} \sim \mathrm{N}\left(\frac{d_j u_j^T y}{d_j^2 + \tau^{2(i-1)}}, \frac{\sigma^2}{d_j^2 + \tau^{2(i-1)}}\right)$
    end
    if $\tau^2 \sim \mathrm{InvGamma}(a, b)$ **then**
        sample $\tau^{2(i)} \sim \mathrm{InvGamma}\left(a + \frac{p}{2}, \frac{\theta^{(i)T}\theta^{(i)}}{2\sigma^{2(i-1)}} + b\right)$
    end
    if $\tau^2 \sim \mathrm{BetaPrime}(a, b)$ **then**
        sample $\tau^{2(i)} \sim \mathrm{GIG}\left(\frac{\theta^{(i)T}\theta^{(i)}}{\sigma^{2(i-1)}}, 2\gamma^{2(i-1)}, a - \frac{p}{2}\right)$
        sample $\gamma^{2(i)} \sim \mathrm{Gamma}\left(a + b, \tau^{2(i)} + 1\right)$
    end
    if $\tau^2 \sim \mathrm{InvGaussian}(a, b)$ **then**
        sample $\tau^{2(i)} \sim \mathrm{GIG}\left(b + \frac{\theta^{(i)T}\theta^{(i)}}{\sigma^{2(i-1)}}, \frac{b}{a^2}, -\frac{1}{2} - \frac{p}{2}\right)$
    end
    if $\tau^2 \sim \mathrm{Gamma}(a, b)$ **then**
        sample $\tau^{2(i)} \sim \mathrm{GIG}\left(\frac{\theta^{(i)T}\theta^{(i)}}{\sigma^{2(i-1)}}, 2b, a - \frac{p}{2}\right)$
    end
    sample $\sigma^{2(i)} \sim \mathrm{InvGamma}\left(\frac{n+p}{2}, \frac{1}{2}\left[\frac{\theta^{(i)T}\theta^{(i)}}{\tau^{2(i)}} + \left(y - F\theta^{(i)}\right)^T\left(y - F\theta^{(i)}\right)\right]\right)$
end

Algorithm 1: Gibbs algorithm for linear regression models with global shrinkage priors.

2. Linear regression with local shrinkage priors

Local shrinkage priors provide more flexibility than global shrinkage priors by allowing the prior variance of the regression parameters to differ between (groups of) variables. Models with local shrinkage priors yield generalized ridge estimators and some sparse estimators.

2.1. Model and priors

Given a partition of the set of p variables into K groups, denote by G_k the set of indexes of the variables that belong to group k and $p_k = \mathrm{card}(G_k)$. (Thus, $\cup_{k=1}^K G_k = \{1, \ldots, p\}$ and $\sum_{k=1}^K p_k = p$.) Then, the linear regression model with local shrinkage priors is

$$\begin{aligned} y \mid \beta, \sigma^2 &\sim N_n(X\beta, \sigma^2 I_n), && (4)\\ \beta \mid \tau^2, \sigma^2 &\sim \mathrm{N}_p(0, \sigma^2 D_\tau) && (5)\\ p(\sigma^2) &\propto \sigma^{-2}. && (6) \end{aligned}$$

where $\tau^2 = (\tau_1^2, \ldots, \tau_K^2)$ and the diagonal matrix D_τ is such that $(D_\tau)_{jj} = \tau_k^2$ if variable

$j = 1, \ldots, p$ belongs to group $k = 1, \ldots, K$.

The R package **shrinkage** offers the following choices of priors for τ^2:

1. $\tau_k^2 \sim \text{InvGamma}(a, b)$
2. $\tau_k^2 \sim \text{BetaPrime}(a, b)$
3. $\tau_k^2 \sim \text{InvGaussian}(a, b)$
4. $\tau_k^2 \sim \text{Gamma}(a, b)$

Several models proposed in the literature can be seen as models with local shrinkage priors, e.g.:

- Bai and Ghosh (2018), when $K = p$ and $\tau_k^2 \sim \text{BetaPrime}(a, b)$
- Caron and Doucet (2008), when $K = p$ and $\tau_k^2 \sim \text{InvGaussian}(a, b)$
- Brown and Griffin (2010) and Caron and Doucet (2008), when $K = p$ and $\tau_k^2 \sim \text{Gamma}(a, b)$

2.2. Closed-form inference

When $\tau^2 = \hat{\tau}^2_{\text{ML}}$ the prior variance τ^2 is not endowed with a prior probability distribution but is instead set to the value that maximizes the marginal likelihood (ML) of the model (empirical Bayes). In such case, Bayesian inference is carried out analytically as the marginal posterior distributions of β and σ^2 are available in closed-form:

$$\beta \mid y \sim \text{T}_p\left(\tilde{\beta}, \frac{y^T y - \tilde{\beta}^T \tilde{\Sigma}^{-1}\tilde{\beta}}{n}\tilde{\Sigma}, n\right) \quad \text{and} \quad \sigma^2 \mid y \sim \text{InvGamma}\left(\frac{n}{2}, \frac{1}{2}\left[y^T y - \tilde{\beta}^T \tilde{\Sigma}^{-1}\tilde{\beta}\right]\right).$$

Here

$$\tilde{\Sigma} = \left(X^T X + D_\tau^{-1}\right)^{-1} \quad \text{and} \quad \tilde{\beta} = \tilde{\Sigma} X^T y.$$

The marginal likelihood is also available in closed-form:

$$p(y) = \pi^{-\frac{n}{2}}\left(\prod_{k=1}^{K}(\tau_k^2)^{-\frac{p}{2}}\right) \mid X^T X + D_\tau^{-1} \mid^{-\frac{1}{2}} \Gamma\left(\frac{n}{2}\right)\left(\frac{1}{2}y^T y - \frac{1}{2}\tilde{\beta}^T\tilde{\Sigma}^{-1}\tilde{\beta}\right)^{-\frac{n}{2}},$$

and its maximizer,

$$\hat{\tau}^2_{\text{ML}} = \underset{\tau^2}{\arg\max}\ \log p(y),$$

obtained very efficiently (Karabatsos 2018) when re-writing the log-marginal likelihood

$$\log p(y) = -\frac{n}{2}\log\pi - \frac{p}{2}\log\tau^2 - \frac{1}{2}\sum_{r=1}^{q}\log(d_r^2 + \tau^{-2}) + \log\Gamma\left(\frac{n}{2}\right) - \frac{n}{2}\log\left(\frac{1}{2}y^T y - \frac{1}{2}\sum_{r=1}^{q}\frac{\hat{\theta}_r^2 d_j^4}{d_r^2 + \tau^{-2}}\right),$$

in terms of the singular values d_r, $r = 1, \ldots, q = \min(n, p)$, of $X = UDV^T = FV^T$ and the maximum likelihood estimate $\hat{\theta} = (F^T F)^{-1} F^T y = D^{-1} U^T y$ of $\theta = V^T \beta$.

2.3. Inference using a Gibbs sampler

Inference for linear models with local shrinkage priors is carried out using a Gibbs sampler.

Posterior conditional distributions

Regardless of the prior on τ_k^2, the posterior conditional distributions for β and σ^2 are:

$$\begin{aligned}\beta \mid \tau^2, \sigma^2 &\sim \mathrm{N}_p\left(\tilde{\beta}, \sigma^2\tilde{\Sigma}\right),\\ \sigma^2 \mid \beta, \tau^2 &\sim \mathrm{InvGamma}\left(\frac{n+p}{2}, \frac{1}{2}\left[\beta^T D_\tau^{-1}\beta + (y - X\beta)^T(y - X\beta)\right]\right).\end{aligned}$$

The posterior conditional distribution of τ_k^2, for $k = 1, \ldots, K$, depends on the choice of prior distribution:

1. when $\tau_k^2 \sim \mathrm{InvGamma}(a, b)$, the posterior conditional distribution of τ_k^2 is:

$$\tau_k^2 \mid \beta_k, \sigma^2 \sim \mathrm{InvGamma}\left(a + \frac{p_k}{2}, \frac{1}{2}\sigma^{-2}\beta_k^T\beta_k + b\right),$$

   where β_k denotes the sub-vector of β consisting of the regression parameters of variables in group k.

2. when $\tau_k^2 \sim \mathrm{BetaPrime}(a, b)$ two parametrizations can be used:

   - the Gamma-Gamma representation of the beta prime distribution (Zhang *et al.* 2016):

     $$\tau_k^2 \sim \mathrm{BetaPrime}(a, b) \quad \Leftrightarrow \quad \tau_k^2 \sim \gamma_k^2 \sim \mathrm{Gamma}(a, \gamma_k^2),\ \gamma_k^2 \sim \mathrm{Gamma}(b, 1),$$

     yields the posterior conditional distributions:

     $$\begin{aligned}\tau_k^2 \mid \beta_k, \gamma_k^2, \sigma^2 &\sim \mathrm{GIG}\left(\sigma^{-2}\beta_k^T\beta_k,\ 2\gamma_k^2,\ a - \frac{p_k}{2}\right),\\ \gamma_k^2 \mid \tau_k^2 &\sim \mathrm{Gamma}\left(a + b,\ \tau_k^2 + 1\right).\end{aligned}$$

   - the inverse-Gamma-inverse-Gamma representation of the beta prime distribution (Schmidt and Makalic 2019, Proposition 1):

     $$\tau_k^2 \sim \mathrm{BetaPrime}(a, b) \quad \Leftrightarrow \quad \tau_k^2 \mid \gamma_k^2 \sim \mathrm{InvGamma}(b, 1/\gamma_k^2),\ \gamma_k^2 \sim \mathrm{InvGamma}(a, 1)$$

     yields the posterior conditional distributions:

     $$\begin{aligned}\tau_k^2 \mid \beta_{j\in G_k}, \gamma_k^2, \sigma^2 &\sim \mathrm{InvGamma}\left(b + \frac{p_k}{2}, \frac{\beta_k^T\beta_k}{2\sigma^2} + \frac{1}{\gamma_k^2}\right),\\ \gamma_k^2 \mid \tau_k^2 &\sim \mathrm{InvGamma}\left(a + b,\ 1 + \frac{1}{\tau_k^2}\right).\end{aligned}$$

3. when $\tau_k^2 \sim \mathrm{InvGaussian}(a, b)$ the posterior conditional distribution of τ_k^2 is

$$\tau_k^2 \mid \beta_k, \sigma^2 \sim \mathrm{GIG}\left(b + \sigma^{-2}\beta_k^T\beta_k,\ \frac{b}{a^2},\ -\frac{1}{2} - \frac{p_k}{2}\right)$$

4. when $\tau_k^2 \sim \text{Gamma}(a,b)$ the posterior conditional distribution for τ_k is

$$\tau_k^2 \mid \beta_k, \sigma^2 \quad \sim \quad \text{GIG}\left(\sigma^{-2}\beta_k^T\beta_k,\ 2b,\ a-\frac{p_k}{2}\right).$$

Algorithm

Algorithm 2 describes the Gibbs sampler for linear regression models with local shrinkage priors. Note that to sample β at iteration i the R package **shrinkage** uses the method of Rue (2001) when $n > p$ and the method of Bhattacharya *et al.* (2016) when $n \leq p$.

Initialize:
$a = b = 0.5$, $\tau_1^{2\,(0)} = \ldots = \tau_K^{2\,(0)} = \gamma^{2\,(0)} = 1$, $n_{\text{mcmc}} = 1000$, $n_{\text{burnin}} = 1000$, $n_{\text{iter}} = n_{\text{mcmc}} + n_{\text{burnin}}$
for $i = 1$ **to** n_{iter} **do**
  sample $\beta^{(i)} \sim \text{N}_p\left(\left(X^TX + D_\tau^{(i-1)^{-1}}\right)^{-1} X^Ty, \sigma^2\left(X^TX + D_\tau^{(i-1)^{-1}}\right)^{-1}\right)$
  for $k = 1$ *to* K **do**
    if $\tau_k^2 \sim \text{InvGamma}(a,b)$ **then**
      sample $\tau_k^{2\,(i)} \sim \text{InvGamma}\left(a + \frac{p_k}{2}, \frac{\beta_k^{(i)^T}\beta_k^{(i)}}{2\sigma^{2\,(i-1)}} + b\right)$
    end
    if $\tau_k^2 \sim \text{BetaPrime}(a,b)$ **then**
      sample $\tau_k^{2\,(i)} \sim \text{GIG}\left(\frac{\beta_k^{(i)^T}\beta_k^{(i)}}{\sigma^{2\,(i-1)}}, 2\gamma_k^{2\,(i-1)}, a - \frac{p_k}{2}\right)$
      sample $\gamma_k^{2\,(i)} \sim \text{Gamma}\left(a + b, \tau_k^{2\,(i)} + 1\right)$
    end
    if $\tau_k^2 \sim \text{InvGaussian}(a,b)$ **then**
      sample $\tau_k^{2\,(i)} \sim \text{GIG}\left(b + \frac{\beta_k^{(i)^T}\beta_k^{(i)}}{\sigma^{2\,(i-1)}}, \frac{b}{a^2}, -\frac{1}{2} - \frac{p_k}{2}\right)$
    end
    if $\tau_k^2 \sim \text{Gamma}(a,b)$ **then**
      sample $\tau_k^{2\,(i)} \sim \text{GIG}\left(\frac{\beta_k^{(i)^T}\beta_k^{(i)}}{\sigma^{2\,(i-1)}}, 2b, a - \frac{p_k}{2}\right)$
    end
  end
  sample $\sigma^{2\,(i)} \sim \text{InvGamma}\left(\frac{n+p}{2}, \frac{1}{2}\left[\beta^{(i)^T} D_\tau^{(i)} \beta^{(i)} + \left(y - X\beta^{(i)}\right)^T\left(y - X\beta^{(i)}\right)\right]\right)$
end

Algorithm 2: Gibbs algorithm for linear regression models with local shrinkage priors.

References

Bai R, Ghosh M (2018). "On the beta prime prior for Scale Parameters in High-Dimensional Bayesian Regression models." *arXiv preprint arXiv:1807.06539*.

Bhattacharya A, Chakraborty A, Mallick BK (2016). "Fast sampling with Gaussian scale mixture priors in high-dimensional regression." *Biometrika*, p. asw042.

Brown PJ, Griffin JE (2010). "Inference with normal-gamma prior distributions in regression problems." *Bayesian analysis*, **5**(1), 171–188.

Caron F, Doucet A (2008). "Sparse Bayesian nonparametric regression." In *Proceedings of the 25th international conference on Machine learning*, pp. 88–95.

Hoerl AE, Kennard RW (1970). "Ridge regression: Biased estimation for nonorthogonal problems." *Technometrics*, **12**(1), 55–67.

Karabatsos G (2018). "Marginal maximum likelihood estimation methods for the tuning parameters of ridge, power ridge, and generalized ridge regression." *Communications in Statistics - Simulation and Computation*, **47**(6), 1632–1651. doi:10.1080/03610918.2017.1321119.

Rue H (2001). "Fast sampling of Gaussian Markov random fields." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **63**(2), 325–338.

Schmidt DF, Makalic E (2019). "Bayesian Generalized Horseshoe Estimation of Generalized Linear Models." In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 598–613. Springer.

Zhang Y, Reich BJ, Bondell HD (2016). "High Dimensional Linear Regression via the R2-D2 Shrinkage Prior." 1609.00046.

A. Distributions

The following table provides notation of distributions used throughout the document.

notation	**domain**	**name**	**density**
$x \sim \mathrm{InvGamma}(a, b)$	$x > 0$	inverse-Gamma	$p(x \mid a, b) \propto x^{-a-1} \exp\left\{-\frac{b}{x}\right\}$
$x \sim \mathrm{Gamma}(a, b)$	$x > 0$	Gamma	$p(x \mid a, b) \propto x^{a-1} \exp\left\{-bx\right\}$
$x \sim \mathrm{BetaPrime}(a, b)$	$x > 0$	beta prime	$p(x \mid a, b) \propto x^{a-1}(1+x)^{-a-b}$
$x \sim \mathrm{GIG}(a, b, c)$	$x > 0$	generalized inverse-Gaussian	$p(x \mid a, b, c) \propto x^{c-1} \exp\left\{-\frac{a/x+bx}{2}\right\}$
$x \sim \mathrm{InvGaussian}(a, b)$	$x > 0$	inverse-Gaussian	$p(x \mid a, b) \propto x^{-\frac{3}{2}} \exp\left\{-\frac{b(x-a)^2}{2a^2x}\right\}$
$x \sim \mathrm{N}_p(m, V)$	$x \in \mathbb{R}^p$	Multivariate Normal	$p(x \mid m, V) \propto \mid V \mid^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x-m)^T V^{-1}(x-m)\right\}$
$x \sim \mathrm{T}_p(m, V, d)$	$x \in \mathbb{R}^p$	Multivariate T	$p(x \mid m, V, d) \propto \left[1 + \frac{1}{d}(x-m)^T V^{-1}(x-m)\right]^{-\frac{d+p}{2}}$

Table 1: List of probability distributions.

Affiliation:

Gwenaël G.R. Leday
Biometris
Wageningen University and Research
Droevendaalsesteeg 1, 6708 PB, Wageningen
The Netherlands
E-mail: gwenael.leday@wur.nl