Sparse Bayesian Inference on Positive-valued Data using Global-local Shrinkage Priors

Yasuyuki Hamura¹, Takahiro Onizuka²,
Shintaro Hashimoto² and Shonosuke Sugasawa³

¹Graduate School of Economics, The University of Tokyo. JSPS Research Fellow.
²Department of Mathematics, Hiroshima University
³Center for Spatial Information Science, The University of Tokyo

Abstract

In various applications, we deal with high-dimensional positive-valued data that often exhibits sparsity. This paper develops a new class of continuous global-local shrinkage priors tailored to analyzing positive-valued data where most of the underlying means are concentrated around a certain value. Unlike existing shrinkage priors, our new prior is a shape-scale mixture of inverse-gamma distributions, which has a desirable interpretation of the form of the posterior mean. We show that the proposed prior has two desirable theoretical properties; Kullback-Leibler super-efficiency under sparsity and robust shrinkage rules for large observations. We propose an efficient sampling algorithm to carry out posterior inference. The performance of the proposed method is illustrated through simulation and two real data examples, the average length of hospital stay for COVID-19 in South Korea and adaptive variance estimation of gene expression data.

Key words: gamma distribution; Kullback-Leibler super-efficiency; Markov chain Monte Carlo; tail-robustness
1 Introduction

In various statistical applications, we often face a sequence of positive-valued observations such as machine failure time, store waiting time, survival time under a certain disease, an income of a certain group, and so on. A common feature of the data is “sparsity” in the sense that most of the underlying means of observations are concentrated around a certain value (grand mean) while a small part of means is significantly large compared with the other means. To reflect the sparsity structure, a useful Bayesian technique is a class of “global-local shrinkage priors” (e.g. Polson and Scott, 2012) that provides adaptive and flexible shrinkage estimation of underlying means; when the observations are around the ground mean, the posterior mean strongly shrinks the observation toward the grand mean, but the observations that are away from the ground mean remain unshrunk.

This paper proposes a framework for Bayesian inference on a sequence of positive-valued observations by using gamma sampling distributions for observations and new global-local shrinkage priors for mean parameters based on a shape-scale mixture of inverse-gamma distributions. Specifically, we introduce the scaled beta (SB) distribution, and its extension is called the inverse rescaled beta (IRB) distribution as mixing distributions in the shape-scale mixture. Although we found that the proposed hierarchical model provides an interpretable form of posterior mean, it causes a strange phenomenon that is unlikely to appear under Gaussian and Poisson models. We provide concrete guidance for choosing the hyperparameters of the proposed priors. Moreover, we reveal two theoretical properties of the proposed prior, tail-robustness for large means and Kullback-Leibler supper-efficiency under sparsity.

There are several works on shrinkage inference of a sequence of positive-valued data. Using the gamma sampling model (as in our proposal), simultaneous estimation for rate/scale parameters was considered by several authors decades ago (e.g., Berger, 1980; Ghosh and Parsian, 1980; DasGupta, 1986; Dey et al., 1987). Typically, in their classical framework, one benchmark method was first chosen in terms of unbiasedness or equivariance, and then interest focused on constructing another estimator which improves
the benchmark under some loss function. To address the sparsity in positive-valued data, Donoho and Jin (2006) proposed a threshold-type estimator with the false discovery rate control, but the assumed model is an exponential distribution (a special case of gamma distribution), so its applicability is quite limited. More recently, Lu and Stephens (2016) proposed an empirical Bayes shrinkage method customized for variance estimation using a χ²-distribution (a special case of gamma distribution) for the observed sampling variance and a finite mixture of inverse-gamma priors for the true variance. The existing approaches for the gamma model do not take account of sparsity as often confirmed in applications, and they tend to provide a universal shrinkage regardless of the observed values. More importantly, the existing methods only produce point estimates of the underlying means. In contrast, the proposed method can obtain full information of posterior distributions, which enables us to compute posterior means (point estimates) and credible intervals for uncertainty quantification.

In Bayesian analysis, the methodology and application of global-local shrinkage priors were developed last decades. Under Gaussian sequence or normal linear regression models, there have been proposed a variety of shrinkage priors including the most famous horseshoe (Carvalho et al., 2010) prior and related priors (e.g. Armagan et al., 2013; Bhadra, et al., 2017; Bhattacharya et al., 2015; Hamura et al., 2020; Zhang et al., 2020). This prior has an attractive shrinkage property, making it possible to strongly shrink small observations toward zero while keeping large observations unshrunk. Recently, techniques of global-local shrinkage priors for Gaussian data are extended to the (quasi-)sparse count data (e.g. Datta and Dunson, 2016; Hamura et al., 2021a). Under the two types of distributional frameworks, several desirable theoretical properties have been revealed. On the other hand, the theoretical argument, Kullback-Leibler super-efficiency, and tail-robustness of the proposed prior are completely different from ones in the existing works since the proposed prior has a form of shape-scale mixtures that are quite different from ones adopted in the Gaussian and count cases. Therefore, our work contributes to the development of not only the analysis of positive-valued data but also the theory of global-local shrinkage priors.

The remainder of the paper is structured as follows. In Section 2, we introduce
settings and our hierarchical model, and we propose a global-local shrinkage prior based on a kind of beta distribution. Furthermore, we illustrate the properties of the marginal prior and posterior distributions for $\lambda_i$, and also discuss the selection of hyper-parameters of the proposed priors. An efficient posterior computation algorithm is constructed via the Markov chain Monte Carlo (MCMC) method. In Section 3 we show two theoretical properties on the proposed priors. The performance of the proposed method is demonstrated through numerical studies in Section 4 and we apply the method to two real datasets related to the average length of hospital stay for COVID-19 in South Korea and variance estimation of gene expression data in Section 5. Proofs and technical details are given in the Supplementary Material.

2 Bayesian Inference on Positive-Valued Data

2.1 Settings and models

Suppose we observe a sequence of positive-valued data, denoted by $y_1, \ldots, y_n$. For each $i = 1, \ldots, n$, we assume the following gamma model $y_i$:

$$y_i | \lambda_i \sim \text{Ga} \left( \delta_i, \frac{\delta_i}{\lambda_i \eta_i} \right),$$

where $\text{Ga}(\alpha, \beta)$ denotes a gamma distribution with shape parameter $\alpha$ and rate parameter $\beta$, $\delta_i$ is a fixed constant, and $\lambda_i$ is a parameter of interest. Under the model, $E[y_i] = \lambda_i \eta_i$ and $\eta_i$ is a structural component that may be modeled to incorporate covariates and other external information (e.g. spatial information). In what follows, we assume $\eta_i = 1$ for simplicity, under which $\lambda_i$ is interpreted as the mean of $y_i$, but all the computation algorithms and analytical results are valid for the general form of $\eta_i$ as long as $\eta_i$ is conditioned on. As considered in Lu and Stephens (2016), if $y_i$ and $\lambda_i$ are sampling and true variances, respectively, the choice is $\delta_i = n_i/2$, where $n_i$ is a sample size used to compute $y_i$. Moreover, if $y_i$ is a sample mean based on $n_i$ samples generated from an exponential distribution $\text{Exp}(1/\lambda_i)$, it holds that $\delta_i = n_i$, and it reduces the framework of a sequence of exponential data when $n_i = 1$, considered in Donoho and Jin (2006). In the present framework, our interest lies in the simultaneous estimation
of the sequence of positive-valued means $\lambda = (\lambda_1, \ldots, \lambda_n)$ by combining information of a given set of data $y = (y_1, \ldots, y_n)$. In particular, we focus on the structure that most of the observations are located around the grand, whereas some observations are very large. To carry out flexible Bayesian inference even under this situation, we employ an idea of global-local shrinkage that can provide customized shrinkage estimation of $\lambda_i$ depending on the location of the observed value $y_i$.

Specifically, we consider the following prior distribution for $\lambda_i$:

$$\lambda_i | u_i \sim IG(1 + \tau u_i, \beta \tau u_i), \quad i = 1, \ldots, n, \quad (2)$$

where $\beta$ and $\tau$ are unknown global parameters and $u_i$ is a local parameter related to customized shrinkage rule. The prior mean of $\lambda_i$ is $E[\lambda_i] = \beta$, so that $\beta$ is interpreted as a grand mean of underlying heterogeneous means. On the other hand, since $\text{Var}(\lambda_i) = \beta^2/(\tau u_i - 1)$ as long as $\tau u_i > 1$, $\tau$ and $u_i$ control the scale of the prior. Unusual parametrization of (2) is the dependence of both shape and scale parameters on the local parameter $u_i$, so that setting a mixing distribution for $u_i$ leads to a class of shape-scale mixtures of inverse-gamma distributions. However, this parametrization is essential to interpret the form of posterior means of $\lambda_i$.

Under the inverse-gamma prior (2), the conditional posterior distributing of $\lambda_i$ given $u_i$ is $IG(1 + \delta_i + \tau u_i, \delta_i y_i + \beta \tau u_i)$, so that the posterior mean of $\lambda_i$ is given by

$$E[\lambda_i | y_i] = E\left[\frac{\delta_i y_i + \beta \tau u_i}{\delta_i + \tau u_i} \bigg| y_i \right] = y_i - E[\kappa_i | y_i](y_i - \beta),$$

where $\kappa_i = \tau u_i/(\delta_i + \tau u_i) \in (0, 1)$ is known as shrinkage factor that determines the amount of shrinkage of $y_i$ toward the grand mean $\beta$. As desirable properties of $\kappa_i$, $E[\kappa_i | y_i]$ should be close to 1 when $y_i$ is close to the grand mean, leading to strong shrinkage toward $\beta$, while $E[\kappa_i | y_i]$ should be sufficiently small for $y_i$ having large $y_i - \beta$ to prevent bias caused by over-shrinkage. We also note that the global parameter $\tau$ determines the overall shrinkage effect, whereas the local parameter $u_i$ allows $\kappa_i$ to vary over different observations.
2.2 Global-local shrinkage priors

Our hierarchical model can be expressed as

\[ y_i | \lambda_i \sim \text{Ga} \left( \delta_i, \frac{\delta_i}{\lambda_i \eta_i} \right), \quad \lambda_i | u_i \sim \text{IG}(1 + \tau u_i, \beta \tau u_i), \quad u_i \sim \pi(\cdot), \]

where \( \beta \) and \( \tau \) are unknown hyperparameters, for which we assign conjugate gamma priors, as discussed in Section 2.5. Note that we mainly consider the case of \( \eta_i = 1 \). For the local parameter \( u_i \), we suggest two prior distributions. The first one is the scaled beta (SB) prior

\[ \pi_{\text{SB}}(u_i) = \frac{1}{B(a, b)} \frac{u_i^{a-1}}{(1 + u_i)^{a+b}}, \]

where \( a, b > 0 \) are hyperparameters and \( B(a, b) \) is the beta function. The SB distribution is also known as the beta prime distribution (e.g., Johnson et al., 1995), and the related family of distributions has been often used in Bayesian statistics (e.g., Pérez et al., 2017; Hamura et al., 2021b), especially in the context of shrinkage priors. As an alternative prior, we newly propose the inverse rescaled beta (IRB) prior

\[ \pi_{\text{IRB}}(u_i) = \frac{1}{B(b, a)} \frac{1}{u_i (1 + u_i)} \frac{\{\log(1 + 1/u_i)\}^{b-1}}{\{1 + \log(1 + 1/u_i)\}^{b+a}}. \]

Note that the IRB prior for \( u_i \) is equivalent to using the rescaled beta prior (Hamura et al., 2021b) for \( 1/u_i \).

Here, we summarize basic tail properties of \( \pi_{\text{SB}}(u_i) \) and \( \pi_{\text{IRB}}(u_i) \). As is well known, the SB prior has the following properties corresponding to robustness and shrinkage:

- **Robustness.** As \( u_i \to 0 \), we have \( \pi_{\text{SB}}(u_i) \propto u_i^{a-1} \). In particular, \( \pi_{\text{SB}}(\kappa_i) \to \infty \) as \( \kappa_i \to 0 \) if and only if \( a < 1 \).

- **Shrinkage.** As \( u_i \to \infty \), we have \( \pi_{\text{SB}}(u_i) \propto u_i^{-b-1} \). In particular, \( \pi_{\text{SB}}(\kappa_i) \to \infty \) as \( \kappa_i \to 1 \) if and only if \( b < 1 \).

Meanwhile, ignoring log factors, we see that \( \pi_{\text{IRB}}(u_i) \) has the following tail properties:

- **Robustness.** As \( u_i \to 0 \), we have \( \pi_{\text{IRB}}(u_i) \approx u_i^{-1} \). In particular, \( \pi_{\text{IRB}}(\kappa_i) \approx
$\kappa_i \rightarrow \infty$ as $\kappa_i \rightarrow 0$ whatever the value of $a > 0$ is. This is in contrast to the case of the SB prior.

- **Shrinkage.** As $u_i \rightarrow \infty$, we have $\pi_{IRB}(u_i) \propto u_i^{-1-b}$. In particular, $\pi_{IRB}(\kappa_i) \rightarrow \infty$ as $\kappa_i \rightarrow 1$ if and only if $b < 1$. This is exactly as in the case of the SB prior.

### 2.3 Marginal prior for $\lambda_i$

In this section, we consider the behavior of the marginal prior of $\lambda_i$. We assume that the grand mean $\beta$ and global shrinkage parameter $\tau$ are fixed at 1 for simplicity. We first discuss the roles of the hyper-parameters, $a$ and $b$, of the proposed priors, and then we propose particular choices of the hyperparameters. Analytical results are summarized in the following proposition.

**Proposition 1.** Suppose that either $\pi(u_i) = \pi_{SB}(u_i) \propto u_i^{-1} / (1 + u_i)^{a+b}$ or $\pi(u_i) = \pi_{IRB}(u_i) \propto [1/(u_i(1+u_i))] \log(1+1/u_i)^{b-1}/\{1+\log(1+1/u_i)\}^{b+a}$. Then the marginal prior $p(\lambda_i)$ of $\lambda_i$ has the following properties:

(i) As $\lambda_i \rightarrow 0$,

$$p(\lambda_i) \approx \begin{cases} 
\lambda_i^{a-1}, & \text{if } \pi(u_i) = \pi_{SB}(u_i), \\
\lambda_i^{-1}, & \text{if } \pi(u_i) = \pi_{IRB}(u_i).
\end{cases}$$

(ii) As $\lambda_i \rightarrow \infty$,

$$p(\lambda_i) \approx \lambda_i^{-2}.$$

(iii) As $\lambda_i \rightarrow 1$,

$$p(\lambda_i) \rightarrow \begin{cases} 
\infty, & \text{if } b \leq 1/2, \\
C_1 < \infty, & \text{if } b > 1/2,
\end{cases}$$

for some finite positive constant $0 < C_1 < \infty$. 


Note that log factors are ignored in the above statement. This result is obtained from a more general theorem (Theorem S1) given in the Supplementary Material. Part (i) corresponds to shrinkage and non-shrinkage for small \(\lambda_i\) under the SB prior with \(a > 1\) and the IRB prior, respectively. Part (ii) corresponds to robustness for large \(\lambda_i\) under the proposed priors; if we fix \(u_i\), then we necessarily have \(p(\lambda_i) \propto \lambda_i^{-2-u_i} < \lambda_i^{-2}\) as \(\lambda_i \to \infty\). Part (iii) corresponds to shrinkage for moderate \(\lambda_i\) under the proposed priors with \(b \leq 1/2\); if we fix \(u_i\), then \(p(\lambda_i)\) never diverges at \(\lambda_i = 1\).

In other words, the left tail of \(\pi(u_i)\) can affect the left tail of \(p(\lambda_i)\) if we use the SB prior with \(a \leq 1\) or the IRB prior; the right tail of \(p(\lambda_i)\) is guaranteed to be sufficiently heavy for any values of the hyperparameters; we can expect that a sufficient amount of prior probability mass is put around \(\lambda_i = 1\) if we choose \(b \leq 1/2\) for the SB and IRB priors. Based on these findings, we propose to use \(a > 1\) for the SB prior and \(b \leq 1/2\) for both the SB and IRB priors. In particular, our default choices are \(a = 2\) and \(b = 1/2\) for both the priors.

The marginal prior densities of \(\lambda_i\) under the SB and IRB priors are illustrated in Figure 1. As expected, it can be seen from the right panel that the right tail of \(p(\lambda_i)\) is heavier under the proposed priors than the global shrinkage prior (denoted by GL in Figure 1) when \(u_i\) is fixed, that is, \(\lambda_i \sim IG(2,1)\). Also, it is confirmed that the IRB prior makes the right tail heavier than the SB prior. The left panel shows that the hyper-parameter \(a\) of the SB and IRB priors causes a trade-off between undesirable tail thickness at the origin and desirable tail thickness at infinity. However, for the case of the SB prior, we at least have that \(p(\lambda_i) \to 0\) as \(\lambda_i \to 0\) for \(a = 2\) and for \(a = 3\). The most remarkable point we want to stress here is that under each of the proposed priors, \(p(\lambda_i)\) has a spike at \(\lambda_i = 1\). This means that a large shrinkage effect is expected when we use one of the proposed priors, and this is quite in contrast to the case of fixing \(u_i = 1\), where the mode of \(p(\lambda_i)\) is significantly shifted to the left.

Finally, the choice \(a = 2\) may seem slightly strange in the literature on global-local shrinkage priors. Under \(u_i \sim SB(a,b)\), the shrinkage factor \(\kappa_i = u_i/(1 + u_i)\) follows the beta distribution \(\text{Beta}(a,b)\). The well-known horseshoe prior (Carvalho et al., 2010) corresponds to the case \((a,b) = (1/2,1/2)\), and the resulting prior distribution of \(\kappa_i\) is
Beta(1/2, 1/2), which has the popular U-shaped density. For our model, we do not adopt the choice \((a, b) = (1/2, 1/2)\), since setting \(a = 1/2\) causes unexpected tail-robustness (or lack of desirable shrinkage toward the grand mean) around the origin and since using \(a > 1\) does not affect tail-robustness around infinity much (see also Section 3.1).

Figure 1: Marginal prior densities for \(\lambda_i\). The right panel is an enlarged version of the left panel in log-scaled tail region.

2.4 Marginal posterior of \(\lambda_i\)

Here, we discuss the flexibility of the proposed prior distributions. As an artificial example, we suppose that \(m = 50\) observations (the first 47 observations are 5 and the others are 15, 30, 50) are observed. Furthermore, we set \(\eta_i = \delta_i = 5\), such that the sampling distribution corresponds to the exponential distribution with mean \(\lambda_i\). We show marginal posterior distributions of \(\lambda_i\) given \(y \in \{5, 15, 30, 50\}\) in Figure 2. The marginal posterior under the global shrinkage prior \((u_i = 1)\) over-shrinks the posterior density under a large signal such as \(y = 50\). Furthermore, Figure 2 shows that the IRB prior produces heavier-tailed posterior density functions than the SB prior for large observations. Therefore, we recommend using IRB prior in situations where tail-robustness is more important.
2.5 Posterior computation

We can construct an efficient Metropolis within the Gibbs algorithm for our model by using the approximation method of Miller (2019). Here, we consider the case of the SB prior. The details of posterior computation under the IRB prior are given in the Supplementary Material. In order to simplify sampling of $\tau$, we make the change of variables $\nu_i = \tau u_i$ for $i = 1, \ldots, n$. Then the overall posterior distribution of $(\lambda, \beta, \tau, \nu)$ given $y$ is expressed by

$$p(\lambda, \beta, \tau, \nu|y) \propto \pi(\beta) \pi(\tau) \frac{1}{\tau^n} \prod_{i=1}^{n} \left\{ \pi(\nu_i/\tau) \frac{\beta^{\nu_i+1}\nu_i^\nu_i}{\Gamma(\nu_i)} \frac{1}{\lambda_i^{\nu_i+2}} e^{-\beta \nu_i/\lambda_i} \frac{1}{\lambda_i^{\delta_i}} \exp \left( - \frac{\delta_i y_i}{\lambda_i \eta_i} \right) \right\} ,$$

Figure 2: Marginal posterior densities for $\lambda_i$ under four types of observed values.
where $\nu = (\nu_1, \ldots, \nu_n)$. Since the SB prior density is expressed as

$$
\pi_{SB}(u_i) = \frac{1}{B(a,b)} \frac{u_i^{a-1}}{(1 + u_i)^{a+b}} = \frac{1}{\Gamma(a)\Gamma(b)} \int_0^\infty t_i^{a+b-1} e^{-t_i} u_i^{a-1} e^{-t_i u_i} dt_i
$$

for all $i = 1, \ldots, n$, it follows that

$$
p(\lambda, \beta, \tau, \nu | y) \propto \int_{(0, \infty)^n} \left[ \pi(\beta) \pi(\tau) \frac{1}{\tau^{na}} \prod_{i=1}^n \left\{ t_i^{a+b-1} e^{-t_i} \nu_i^{a-1} e^{-t_i \nu_i / \tau} \frac{\beta^{\nu_i + 1} \nu_i \nu_i^t}{\Gamma(\nu_i)} \frac{1}{\lambda_i} e^{-\beta \nu_i / \lambda_i} \frac{1}{\lambda_i} e^{-(\delta y_i)/(\lambda \eta)} \right\} \right] dt.
$$

We consider $t = (t_1, \ldots, t_n) \in (0, \infty)^n$ as a set of additional latent variables. For the global parameters, we use the conjugate gamma priors $\pi(\beta) = Ga(\beta | a_\beta, b_\beta)$ and $\pi(\tau) = Ga(\tau | a_\tau, b_\tau)$, and we often use $a_\beta = b_\beta = a_\tau = b_\tau = 0.1$ as non-informative priors in simulation.

The variables $\lambda, \beta, \tau, t, \text{and} \nu$ are updated in the following way.

- Sample $\lambda_i \sim IG(\delta_i + \nu_i + 1, \delta_i y_i / \eta_i + \beta \nu_i)$ independently for $i = 1, \ldots, n$.

- Sample $\beta \sim Ga\left( \sum_{i=1}^n \nu_i + n + a_\beta, \sum_{i=1}^n \nu_i / \lambda_i + b_\beta \right)$.

- Sample $\tau \sim GIG\left( -na + a_\tau, 2b_\tau, 2 \sum_{i=1}^n t_i \nu_i \right)$, which has density proportional to $\tau^{-na+a_\tau-1} e^{-b_\tau \tau - \sum_{i=1}^n t_i \nu_i / \tau}$.

- Sample $t_i \sim Ga(a + b, 1 + \nu_i / \tau)$ independently for $i = 1, \ldots, n$.

- The full conditional distribution of $\nu$ is proportional to

$$
\prod_{i=1}^n \{Ga(\nu_i | a, t_i / \tau)Ga(1 / \lambda_i, \nu_i, \beta \nu_i)\},
$$

which can be accurately approximated by using the method of Miller (2019) for each $i = 1, \ldots, n$. The method is based on the gamma approximation of intractable probability density function by matching the first- and second-derivatives of log densities. We use the approximate full conditional distributions as proposal distributions in independent Metropolis-Hastings (MH) steps.
3 Theoretical Properties

In this section, we analytically compare properties of different priors for $u_i$ and, in particular, show two properties of the proposed priors, namely, tail-robustness for large observations (Section 3.1) and desirable Kullback-Leibler risk bound under sparsity (Section 3.2). For simplicity, we fix $\beta = \tau = 1$ in what follows, so that all the theoretical results are conditional on the hyperparameters.

3.1 Tail-robustness for large observations

For a prior $\pi(u_i)$ of local parameter $u_i$, we consider the class given by

$$\sup_{u \geq 1} \{u \pi(u)\} < \infty,$$

(3)

$$\pi(u) \sim C u^{\alpha - 1} \left\{1 + \log(1 + 1/u)\right\}^{1+\gamma} \text{ as } u \to 0 \text{ for some } \alpha \geq 0 \text{ and } \gamma \geq -1,$$

(4)

where $C$ is a positive constant. The notation $f(x) \sim g(x)$ means $\lim_{x \to 0} f(x)/g(x) = 1$. Condition (3) is a technical condition satisfied by most priors. Condition (4) is a condition on the tail of $\pi(u_i)$ at the origin and is satisfied by both the SB and the IRB priors. Because we consider proper distributions only in this paper, the case of $\alpha = 0$ and $\gamma \leq 0$ is excluded.

We consider the tail robustness of the Bayes estimator of $\lambda_i$ given by

$$\hat{\lambda}_i = \hat{\lambda}_i^{\text{ML}} - E[\kappa_i|y_i](\hat{\lambda}_i^{\text{ML}} - 1),$$

where $\hat{\lambda}_i^{\text{ML}} = y_i/\eta_i = y_i$ if $\eta_i = 1$ and $\kappa_i = u_i/(\delta_i + u_i)$. Specifically, we show that the expected shrinkage factor, $E[\kappa_i|y_i]$, converges to zero as $y_i \to \infty$.

Theorem 1. There exists a function $\kappa^*(\cdot) : (0, \infty) \to (0, \infty)$ such that

$$E[\kappa_i|y_i] \sim \frac{1}{\delta_i}(1 + \alpha)\kappa^* \left(\frac{\delta_i}{\eta_i}y_i\right) \to 0$$

as $y_i \to \infty$.

The constant $\alpha \geq 0$ is related to the tail of $\pi(u_i)$ at the origin. The heavier the tail
is, the faster the expected shrinkage factor converges to zero. Finally, we note that if we fix \( u_i = 1 \), then \( E[\kappa_i | y_i] = 1/(\delta_i + 1) \to 0 \) as \( y_i \to \infty \).

### 3.2 Kullback-Leibler super-efficiency under sparsity

We now consider the predictive efficiency for the proposed method (e.g. Polson and Scott, 2010; Carvalho et al., 2010; Datta and Dunson, 2016). In particular, we discuss the Kullback-Leibler divergence between the true sampling density and the Bayes predictive density under the proposed global-local shrinkage prior. We consider the following one-dimensional model

\[
y \sim \text{Ga} \left( \delta, \frac{\delta}{\lambda \eta} \right), \quad \lambda \sim \text{IG}(1 + u, u), \quad u \sim \pi(u).
\]

In the above model, let \( f(y|\lambda) = \text{Ga}(y|\delta, \delta/(\lambda \eta)) \) and let \( \lambda_0 \) be the true value of \( \lambda \). We define the Kullback-Leibler (KL) divergence between \( f(y|\lambda) \) and \( f(y|\lambda_0) \) by

\[
D_{\text{KL}}(\lambda_0, \lambda) = D_{\text{KL}}(f(y|\lambda_0), f(y|\lambda)).
\]

Then we have

\[
D_{\text{KL}}(\lambda_0, \lambda) = \delta \left( 1/\lambda - 1/\lambda_0 - \log 1/\lambda_0 \right) = \delta \left( \lambda_0/\lambda - 1 - \log \lambda_0/\lambda \right).
\]

Furthermore, the KL neighborhood around \( \lambda_0 \) is defined by

\[
A_{\varepsilon}(\lambda_0) = \{ \lambda \in (0, \infty) | D_{\text{KL}}(\lambda_0, \lambda) < \varepsilon \}.
\]

We assume that the prior \( p(\lambda) \) is information dense in the sense of \( p(A_{\varepsilon}(\lambda_0)) > 0 \) for all \( \varepsilon > 0 \). From the Proposition 4 in Barron (1987), we have the Cesa-ro-mean risk \( R_n \) is expressed by

\[
R_n \leq \varepsilon - \frac{1}{n} \log P(\lambda \in A_{\varepsilon}(\lambda_0)), \tag{5}
\]

where \( R_n = n^{-1} \sum_{k=1}^{n} D_{\text{KL}}(f(y|\lambda_0)|\hat{f}_k(\lambda)) \) and \( \hat{f}_k(\lambda) \) is the Bayes predictive density under KL divergence using the posterior density based on \( k \leq n \) observations \( y_1, \ldots, y_k \). We now evaluate the prior probability \( P(\lambda \in A_{\varepsilon}(\lambda_0)) \) in the right-hand side of (5) when \( \lambda_0 = 1 \).
Although we proved the theorem for the univariate case, the convergence in the multivariate case is derived from a component-wise application.

**Theorem 2.** Assume that the true sampling model is $\text{Ga}(\delta, \delta/(\lambda_0 \eta))$. For $\lambda_0 \neq 1$, the Cesáro-mean risk for Bayes predictive density $\hat{f}_n$, which is the posterior mean of the density function $f(\cdot|\lambda)$, satisfies

$$R_n = O\left(n^{-1} \log n\right).$$

If $\lambda_0 = 1$ and if $\pi(u) \propto u^{-1-b}$ as $u \to \infty$ for some $0 < b \leq 1/2$, then

$$R_n = O\left\{n^{-1}(\log n - \log \log n)\right\}.$$

Theorem 2 relates the right tail of $\pi(u_i)$ to the risk given in (5). To achieve Kullback-Leibler super-efficiency, it is sufficient to use $\pi(u_i)$ with a sufficiently heavy tail ($b \leq 1/2$). Thus, $b$ plays a role in controlling sparsity at the grand mean. We remark that fixing $u_i = 1$ corresponds to using a point mass prior for $u_i$ and hence to violation of the sufficient condition that $\pi(u) \propto u^{-1-b}$ as $u \to \infty$.

**4 Simulation Studies**

We evaluate the performance of Bayesian and frequentist shrinkage methods under gamma response. Let $y_i \sim \text{Ga}(\delta_i, \delta_i/\lambda_i)$ for $i = 1, \ldots, n (= 200)$ and $\delta_i = 5$. We consider the following six scenarios of the true mean $\lambda_i$:

(Scenario 1) $\lambda_i \sim 0.95\delta\mu + 0.05\text{Ga}(20\mu, 2)$,  
(Scenario 2) $\lambda_i \sim 0.9\delta\mu + 0.1\text{Ga}(20\mu, 2)$,
(Scenario 3) $\lambda_i \sim 0.95\delta\mu + 0.05\mu|t_3|$,  
(Scenario 4) $\lambda_i \sim 0.9\text{Ga}(5\mu, 5) + 0.1\mu|t_1|$, 
(Scenario 5) $\lambda_i \sim 0.9\delta\mu + 0.1\text{Ga}(10\mu, 2)$,  
(Scenario 6) $\lambda_i \sim 0.85\delta\mu + 0.15\text{Ga}(10\mu, 2)$,

where $\mu = 5$, $\delta_a$ denotes a point mass at $a$ and $t_c$ denotes a $t$-distribution with $c$ degrees of freedom. In the first three scenarios, most true means $\lambda_i$ are exactly equal to $\mu = 5$, and a small part of true means are very large compared with $\mu$. In scenario 4, most of the true means are concentrated around $\mu$ (not exactly equal to 0). In what follows,
the true means generated from the first and second components will be called null and non-null signals, respectively.

For the simulated data, we apply six methods, the proposed scaled beta (SB) and inverse rescaled beta (IRB) priors, the global shrinkage (GL) prior (setting $u_i = 1$ in the proposed model), shrinkage estimators given by DasGupta (1986), denoted by DG, adaptive variance shrinkage estimators by Lu and Stephens (2016), denoted by VS, and maximum likelihood (ML) estimator $y_i$. We set $a = 2$ and $b = 1/2$ in the SB and IRB priors. We used non-informative gamma priors, $\beta \sim \text{Ga}(0.1, 0.1)$ and $\tau \sim \text{Ga}(0.1, 0.1)$ for SB, IRB and GL priors. For the Bayesian methods, 3000 posterior samples are generated after discarding the first 2000 samples as burn-in. Note that we used the R package “vashr” (https://github.com/mengyin/vashr) to apply the VS method, where the degrees of freedom of $\chi^2$-distribution is set to $2\delta_i (= 10)$.

We first investigate the shrinkage property of the proposed global-local shrinkage priors compared with the other methods. In Figure 3, we show scatter plots of observed values and point estimates (posterior means for the Bayesian methods) produced by five shrinkage methods, under scenario 1. It is observed that the standard shrinkage methods, GL, DG, and VS, linearly shrinks observed value $y_i$, that is, shrinkage factor is constant regardless of $y_i$. On the other hand, the proposed SB and IRB priors more strongly shrink the observed values around $\lambda_i = 5$, showing the adaptive shrinkage property of the global-local shrinkage prior.

We next evaluate mean absolute percentage error (MAPE), defined as $n^{-1}\sum_{i=1}^{n}\lambda_i^{-1}|\lambda_i - \hat{\lambda}_i|$ with a point estimate $\hat{\lambda}_i$. We present boxplots of MAPE for 1000 replications in Figure 4. The results indicate that the proposed SB and IRB provide more accurate point estimates than the other methods in all the scenarios, except for IRB under Scenario 3. The improvement of the proposed methods is remarkable when the null and non-null signals are well-separated, as in Scenarios 1 and 2. Comparing SB and IRB, SB tends to provide smaller overall MAPE than IRB. To compare the two methods more precisely, we computed MAPE only for non-null signals. The averaged values of MAPE for non-null signals are given in Table 1, which shows that IRB performs slightly better than SB for the estimation of non-null signals, and this is consistent with the stronger
Figure 3: Scatter plots of observed values (ML) and point estimates obtained from five shrinkage methods. The right panel is an enlarged version of the left panel. The vertical line in the right panel indicates the location of null signals.

tail-robustness property of IRB than that of SB.

Furthermore, we computed the coverage probability (CP) and average length (AL) of 95% credible/confidence intervals. For the frequentist methods, since DG and VS do not provide interval estimation, we only consider the ML method. The 95% confidence interval of ML can be obtained as \( \left( \frac{y_i}{P_G(0.975; \delta_i, \delta_i)}, \frac{y_i}{P_G(0.025; \delta_i, \delta_i)} \right) \), where \( P_G(\cdot; \alpha, \beta) \) denotes the probability function of \( \text{Ga}(\alpha, \beta) \). The CP and AL averaged over 1000 Monte Carlo replications are given in Table 2. It can be seen that all the three Bayesian methods have empirical CP values larger than the nominal level 0.95 except in Scenario 6, whereas the interval lengths of SB and IRB tend to be smaller than GL and ML in all the scenarios.

Table 1: Mean absolute percentage errors (MAPE) for non-null signals averaged over 1000 Monte Carlo replications.

| Scenario | 1    | 2    | 3    | 4    | 5    | 6    |
|---------|------|------|------|------|------|------|
| SB      | 0.365| 0.359| 0.304| 0.292| 0.388| 0.384|
| IRB     | 0.360| 0.360| 0.284| 0.286| 0.385| 0.386|
Figure 4: Boxplots of mean absolute percentage errors (MAPE) for 1000 Monte Carlo replications.

Table 2: Coverage probabilities and average lengths of 95% credible/confidence intervals, averaged over 1000 Monte Carlo replications.

| Scenario | Coverage probability | Average length |
|----------|----------------------|----------------|
|          | SB      | IRB    | GL    | ML    | SB | IRB | GL | ML |
| 1        | 98.7    | 98.4   | 97.6  | 95.0  | 1.13 | 1.19 | 1.58 | 2.59 |
| 2        | 97.3    | 97.1   | 97.1  | 95.0  | 1.32 | 1.33 | 1.79 | 2.59 |
| 3        | 98.1    | 98.4   | 97.9  | 95.0  | 0.81 | 1.09 | 0.91 | 2.59 |
| 4        | 95.6    | 96.0   | 97.2  | 95.0  | 1.12 | 1.23 | 1.43 | 2.59 |
| 5        | 96.6    | 96.3   | 96.7  | 95.0  | 1.23 | 1.27 | 1.56 | 2.59 |
| 6        | 94.4    | 94.0   | 96.2  | 95.0  | 1.37 | 1.39 | 1.69 | 2.59 |

5 Real Data Example

5.1 Average admission period of COVID-19 in Korea

We first apply the global-local shrinkage techniques to estimate the average length of hospital stay of COVID-19 infected persons. We use the data set available at Kaggle (https://www.kaggle.com/kimjihoo/coronavirusdataset), where the date of admission and discharge is observed for 1587 individuals in Korea. We then group
these individuals regarding 98 cities and three classes of age, young (39 or less), middle (from 40 to 69), and old (70 or more), resulting in \( n = 185 \) groups after omitting empty groups. Assuming exponential distributions with group-specific mean for admission period (days) of each individual, the group-wise sample mean is distributed as \( \text{Ga}(n_i, n_i / \lambda_i) \) for \( i = 1, \ldots, n \), where \( n_i \) is the number of individuals within the \( i \)th group and \( \lambda_i \) is the true mean of admission period specific to the \( i \)th group. Note that \( n_i \) ranges from 1 to 258, and the scatter plot of \( n_i \) and \( y_i \) are given in Figure 6.

We apply the proposed SB prior as well as GL and DG methods. Regarding the prior distributions for grand mean \( \beta \) in the SB and GL models, we assign a non-informative prior, \( \text{Ga}(0,0.1) \). Furthermore, we set \( a = 2 \) and \( b = 1/2 \) in the proposed prior distributions. The posterior means for SB and GL are computed based on 10000 posterior samples (after discarding 3000 samples), whose histograms are shown in Figure 5. The posterior mean of the grand mean \( \beta \) in the SB model was 22.0 (95% credible interval was (21.3, 22.8)), which is consistent with the evidence that the average admission period is around 21 (e.g. Jang et al., 2021). On the other hand, the posterior mean of the grand mean \( \beta \) in the GL model was 24.4, where 95% credible interval was (22.5, 26.6).

We also present the histogram of shrinkage estimates made by DG. It is observed that SB strongly shrinks the observed values toward the grand mean \( \beta \), so that most of the posterior means of the average admission period are concentrated around the grand mean. This is because most groups with large sample means have small sample sizes, and such unreliable information is strongly shrunk. We found that only a single group (old age class of Gyeong-si) has a much larger average admission period, about 35 days. Since the sample size of this group is 107, and the sample mean is about 37, the posterior result seems reasonable. To see more detailed results, we present scatter plots of observed values and posterior means against sample size \( n_i \), in Figure 6. It is observed that the amount of shrinkage (i.e., the difference between observed values and posterior means) decreases as \( n_i \) increases, and observations having small sample sizes strongly shrunk toward the grand mean. From Figure 5 it can also be seen that GL also provides reasonably shrunk estimates of \( \lambda_i \) and DG does not, but the proposed SB prior can provide strongly shrunk point estimates. Moreover, the average length of
95% credible intervals made by SB was 19.0, which was considerably smaller than 22.9 produced by GL.

Figure 5: Histograms of shrinkage estimates (red) and observed values (grey) of average admission period. The vertical line corresponds to the grand mean to which the estimator is shrunk.

Figure 6: Left: Scatter plot of sample size $n_i$ and average admission. Right: Scatter plot of sample size $n_i$ and difference of $y_i$ and posterior mean of $\lambda_i$.

5.2 Variance estimation of gene expression data

We next apply the shrinkage methods to variance estimation of gene expression data. As noted in Lu and Stephens (2016), shrinkage estimation of variance is essential to increase
the estimation accuracy of the mean. We use a popular prostate cancer dataset from Singh et al. (2002). In this dataset, there are gene expression values for $n = 6033$ genes for 50 subjects in control subjects. We compute sampling variances of $n$ gene expressions, distributed as $\text{Ga}(n_i/2, n_i/2\lambda_i)$ for $i = 1, \ldots, n$, where $\lambda_i$ is the true variance of the $i$th gene expression. By assigning non-informative priors, $\text{Ga}(0.1, 0.1)$ for $\beta$ in the SB, IRB, and GL models as well as the VS method. In the Bayesian methods, we computed posterior means using 2000 posterior samples after discarding the first 1000 samples. The histograms of posterior means are shown in Figure 7. As confirmed in the previous example, we can see that the proposed SB and IRB priors can provide more shrunk estimates than the other methods. Furthermore, average lengths of 95% credible intervals made by SB and IRB were 0.640 and 0.639, respectively, which are smaller than 0.663 by GL. This shows the efficiency of the proposed priors.

Figure 7: Histograms of shrinkage estimates (blue) and observed values (grey) of variances of gene expression data.
6 Concluding Remarks

We proposed a new class of continuous global-local shrinkage priors for high-dimensional positive-valued observations. We introduce a shape-scale mixture of inverse-gamma distributions as prior for underlying means, which leads to interpretable forms of posterior means. We showed two theoretical properties of the proposed priors and developed an efficient posterior computation algorithm via MCMC. Simulation studies supported the performance of the proposed methods, and we demonstrated that the proposed priors enable reasonable shrinkage toward the grand mean through two real data examples.

The proposed distributions can also be useful in other models. Since the proposed prior provides a flexible form of the marginal distribution of positive-valued observations, it would be interesting to apply it to accelerated failure time models, which could be an alternative to the Bayesian nonparametric approach (Hanson, 2006; Kuo and Mallick, 1997).

Acknowledgments

Research of the authors was supported in part by JSPS KAKENHI Grant Numbers 20J10427, 21K13835 and 21H00699 from Japan Society for the Promotion of Science.

References

[1] Armagan, A., Dunson, D. and Lee, J. (2013). Generalized double Pareto shrinkage. *Statistica Sinica*, 23, 119–143.

[2] Barron, A.R. (1987). Are Bayes rules consistent in information? *Open problems in communication and computation*, 85–91.

[3] Berger, J. (1980). Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of gamma scale parameters. *The Annals of Statistics*, 8, 545–571.
[4] Bhadra, A., Datta, J., Polson, N.G. and Willard, B. (2017). The horseshoe+ estimator of ultra-sparse signals. *Bayesian Analysis, 12*, 1105–1131.

[5] Bhattacharya, A., Pati, D., Pillai, N.S. and Dunson, D.B. (2015). Dirichlet–Laplace priors for optimal shrinkage. *Journal of the American Statistical Association, 110*, 1479–1490.

[6] Carvalho, C.M., Polson, N.G. and Scott, J.G. (2010). The horseshoe estimator for sparse signals. *Biometrika, 97*, 465–480.

[7] DasGupta, A. (1986). Simultaneous estimation in the multiparameter gamma distribution under weighted quadratic losses. *The Annals of Statistics, 14*, 206–219.

[8] Datta, J. and Dunson, D. (2016). Bayesian inference on quasi-sparse count data. *Biometrika, 103*, 971–983.

[9] Dey D., Ghosh, M. and Srinivasan, C. (1987). Simultaneous estimation of parameters under entropy loss. *Journal of Statistical Planning and Inference, 15*, 347–363.

[10] Donoho, D. and Jin, J. (2006). Asymptotic minimaxity of false discovery rate thresholding for sparse exponential data. *The Annals of Statistics, 34*, 2980–3018.

[11] Ghosh, M. and Parsian, A. (1980). Admissible and minimax multiparameter estimation in exponential families. *Journal of Multivariate Analysis, 10*, 551–564.

[12] Hamura, Y., Irie, K. and Sugasawa, S. (2020). Shrinkage with Robustness: Log-Adjusted Priors for Sparse Signals. *arXiv preprint arXiv:2001.08465*.

[13] Hamura, Y., Irie, K. and Sugasawa, S. (2021a). On global-local shrinkage priors for count data. *Bayesian Analysis, 10*, 1–22.

[14] Hamura, Y., Irie, K. and Sugasawa, S. (2021b). Robust Hierarchical Modeling of Counts under Zero-inflation and Outliers. *arXiv preprint arXiv:2106.10503*.

[15] Hanson, T.E. (2006). Modeling censored lifetime data using a mixture of gammas baseline. *Bayesian Analysis, 1*, 575–594.
[16] Jang, S.Y., Seon, J.-Y., Yoon, S.-J., Park, S.-Y., Lee, S.H. and Oh, I.-H. (2021). Comorbidities and factors determining medical expenses and length of stay for admitted COVID-19 patients in Korea. *Risk Management and Healthcare Policy, 14.*

[17] Johnson, N.L., Kotz, S. and Balakrishnan, N. (1995). Continuous univariate distributions, volume 2. John wiley & sons.

[18] Kuo, L. and Mallick, B. (1997). Bayesian semiparametric inference for the accelerated failure-time model. *Canadian Journal of Statistics, 25,* 457–472.

[19] Lu, M. and Stephens, M. (2016). Variance adaptive shrinkage (vash): flexible empirical Bayes estimation of variances. *Bioinformatics, 32,* 3428–3434.

[20] Miller, J.W. (2019). Fast and accurate approximation of the full conditional for gamma shape parameters. *Journal of Computational and Graphical Statistics, 28,* 476–480.

[21] Pérez, M.-E., Pericchi, L.R. and Ramírez, I.C. (2017). The scaled beta2 distribution as a robust prior for scales. *Bayesian Analysis, 12,* 615–637.

[22] Polson, N.G. and Scott, J.G. (2010). Shrink globally, act locally: Sparse Bayesian regularization and prediction. *Bayesian Statistics, 9,* 501–538.

[23] Polson, N.G. and Scott, J.G. (2012). Local shrinkage rules, Lévy processes and regularized regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 74,* 287–311.

[24] Singh, D., Febbo, P.G., Ross, K., Jackson, D.G., Manola, J., Ladd, C., Tamayo, P., Renshaw, A.A., D’Amico, A.V. and Richie, J.P. (2002). Gene expression correlates of clinical prostate cancer behavior. *Cancer cell, 1,* 203–209.

[25] Zhang, Y.D., Naughton, B.P., Bondell, H.D. and Reich, B.J. (2020). Bayesian regression using a prior on the model fit: The R2-D2 shrinkage prior. *Journal of the American Statistical Association,* 1–13.
Supplementary Material for “Sparse Bayesian Inference on Positive-valued Data using Global-local Shrinkage Priors”

Yasuyuki Hamura\textsuperscript{1}, Takahiro Onizuka\textsuperscript{2},
Shintaro Hashimoto\textsuperscript{2} and Shonosuke Sugasawa\textsuperscript{3}

\textsuperscript{1}Graduate School of Economics, The University of Tokyo. JSPS Research Fellow.
\textsuperscript{2}Department of Mathematics, Hiroshima University
\textsuperscript{3}Center for Spatial Information Science, The University of Tokyo

This Supplementary Material provides proofs and technical details related to the main text.

S1 Preliminaries

The following facts will be used later in this Supplementary Material.

- As $x \to 0$, $\Gamma(x) \sim 1/x$. As $x \to \infty$, $\Gamma(x) \sim (2\pi)^{1/2}x^{x-1/2}e^{-x}$.

- For any $x > 0$,

\[
\frac{e^{-1/(12x)}}{(2\pi)^{1/2}} < \frac{x^{x-1/2}}{\Gamma(x)e^x} < \frac{1}{(2\pi)^{1/2}}.
\]

\[
\log x - \frac{\Gamma'(x)}{\Gamma(x)} - \frac{1/2}{x} = \log x - \psi(x) - \frac{1/2}{x} = 2 \int_0^{\infty} \frac{1}{t^2 + x^2} e^{2\pi t} - \frac{t}{2} dt.
\]

- Let $c \in \mathbb{R}$. Let $\Psi: (0, \infty) \to [0, \infty)$ be a continuously differentiable function.

Suppose that $x\Psi'(x)/\Psi(x) \to c$ as $x \to 0$ (as $x \to \infty$). Then, for any $v > 0$,

\[
\frac{\Psi(vx)}{\Psi(x)} \to v^c
\]

as $x \to 0$ (as $x \to \infty$).
Here, we investigate properties of the tails of the marginal prior \( p(\lambda_i) \) of \( \lambda_i \) under the proper prior \( u_i \sim \pi(u_i) \).

**Theorem S1.** Let \( \xi_i = 1/\lambda_i - 1 - \log(1/\lambda_i) > 0 \). The marginal prior of \( \lambda_i \) has the following properties.

(i) Suppose that there exist \( \alpha \geq 0 \) and \( \gamma \geq -1 \) such that \( \pi(u) \propto u^{\alpha-1}/\{1 + \log(1 + 1/u)\}^{1+\gamma} \) as \( u \to 0 \). Then, as \( \lambda_i \to 0 \) and \( \lambda_i \to \infty \), we have \( \xi_i \to \infty \) and

\[
p(\lambda_i) \sim \Gamma(\alpha + 1) \frac{1}{\lambda_i^2} \frac{1}{\xi_i^2} \pi \left( \frac{1}{\xi_i} \right) \propto \frac{1}{\lambda_i^2} \frac{1}{\xi_i^{\alpha+1}} \left\{1 + \log(1 + \xi_i)\right\}^{1+\gamma}.
\]

(ii) Suppose that there exists \( b > 0 \) such that \( \pi(u) \propto u^{-1-b} \) as \( u \to \infty \). Then, as \( \lambda_i \to 1 \), we have \( \xi_i \to 0 \) and

\[
p(\lambda_i) \begin{cases} 
\sim \frac{\Gamma(1/2 - b)}{(2\pi)^{1/2}} \left( \frac{1}{\xi_i} \right)^{1+1/2} \pi \left( \frac{1}{\xi_i} \right) \propto \left( \frac{1}{\xi_i} \right)^{1/2-b} \to \infty, & \text{if } b < 1/2, \\
\to \infty, & \text{if } b = 1/2, \\
\sim \int_0^\infty \pi(u) \frac{u^u e^{-u}}{\Gamma(u)} du < \infty, & \text{if } b > 1/2.
\end{cases}
\]

(iii) Suppose that there exists \( a > 0 \) such that \( \pi(u) = \{1/B(a, 1/2)\} u^{a-1}/(1+u)^{a+1/2} \) for all \( u \in (0, \infty) \). Then, as \( \lambda_i \to 1 \), we have \( \xi_i \to 0 \) and

\[
p(\lambda_i) \sim \frac{1}{(2\pi)^{1/2}} \left( \log \frac{1}{\xi_i} \right) \left( \frac{1}{\xi_i} \right)^{3/2} \pi \left( \frac{1}{\xi_i} \right) \propto \log \frac{1}{\xi_i} \to \infty.
\]

(iv) Suppose that there exists \( a > 0 \) such that \( \pi(u) = \{1/B(1/2, a)\} [1/(u(1+u))] [\log(1+1/u)]^{1/2-a}/\{1 + \log(1 + 1/u)\}]^{1/2+a} \) for all \( u \in (0, \infty) \). Then, as \( \lambda_i \to 1 \), we have \( \xi_i \to 0 \) and

\[
p(\lambda_i) \sim \frac{1}{(2\pi)^{1/2}} \left( \log \frac{1}{\xi_i} \right) \left( \frac{1}{\xi_i} \right)^{3/2} \pi \left( \frac{1}{\xi_i} \right) \propto \log \frac{1}{\xi_i} \to \infty,
\]

which is exactly as in part (iii).
For part (i), the SB and IRB priors correspond to setting $\alpha = a$ and $\gamma = -1$ and setting $\alpha = 0$ and $\gamma > 0$, respectively. Part (ii) is directly applicable to both the SB and IRB priors. Although part (ii) does not consider the speed with which $p(\lambda_i)$ tends to infinity as $\lambda_i \to 1$ when $b = 1/2$, this boundary case is treated in parts (iii) and (iv) for the SB and IRB priors with $b = 1/2$, respectively.

In deriving Proposition 1 from Theorem S1, we note that

$$
\xi_i \sim \begin{cases} 
\frac{1}{\lambda_i}, & \text{as } \lambda_i \to 0, \\
\log \lambda_i, & \text{as } \lambda_i \to \infty, \\
(\lambda_i - 1)^2/2, & \text{as } \lambda_i \to 1.
\end{cases}
$$

It follows from parts (iii) and (iv) that

$$
p(\lambda_i) \propto \log \frac{1}{|\lambda_i - 1|} \to \infty
$$

as $\lambda_i \to 1$ for the SB and IRB priors with $b = 1/2$.

**Proof of Theorem S1.** Let $\xi_i = 1/\lambda_i - 1 - \log(1/\lambda_i) \geq 0$ and let

$$f(u) = \frac{u^{u-1}e^{-u}}{\Gamma(u)}$$

for $u \in (0, \infty)$. Then the marginal density can be written as

$$p(\lambda_i) = \frac{1}{\lambda_i^2} \int_0^\infty \pi(u) \frac{u^u}{\Gamma(u)} \frac{1}{\lambda_i^u} e^{-u/\lambda_i} du = \frac{1}{\lambda_i^2} \int_0^\infty u \pi(u) f(u) e^{-u\xi_i} du.$$

For part (i),

$$
\left| \frac{\int_0^\infty u \pi(u) f(u) e^{-u\xi_i} du}{\int_0^1 u \pi(u) f(u) e^{-u\xi_i} du} - 1 \right| \leq \frac{\int_1^{1/2} u \pi(u) f(u) e^{-u\xi_i} du}{\int_0^{1/2} u \pi(u) f(u) e^{-u\xi_i} du} \leq \frac{\int_1^{1/2} u \pi(u) f(u) e^{-u\xi_i} du}{\int_0^{1/2} u \pi(u) f(u) e^{-u\xi_i} du} e^{-(\xi_i - 1)} \to 0.
$$
as $1 < \xi_i \to \infty$. Therefore, as $1/\lambda_i$ or $\lambda_i$ tends to infinity, we have $\xi_i \to \infty$ and

$$\frac{\lambda_i^2 p(\lambda_i)}{\pi(1/\xi_i)/\xi_i^2} \sim \frac{1}{\pi(1/\xi_i)/\xi_i^2} \int_0^1 u \pi(u) f(u) e^{-u \xi_i} du$$

$$= \frac{1}{\pi(1/\xi_i)/\xi_i^2} \int_{\xi_i}^{\xi_i} \frac{1}{\xi_i} \pi(u/\xi_i) f(u/\xi_i) e^{-u} du$$

$$= \int_0^\infty u e^{-u} \frac{1}{\xi_i} \pi(u/\xi_i) f(u/\xi_i) du.$$ 

Note that

$$\sup_{u \in (0,1)} \left( \pi(u/[u^{\alpha-1}/\{1 + \log(1 + 1/u)\}]^{1+\gamma}) \right) \geq \frac{1(u < \xi_i) \pi(u/\xi_i)}{\pi(1/\xi_i)} \to u^{\alpha-1}$$

for all $u \in (0, \infty)$ since $1 + \log(1 + \xi_i)$ is a slowly varying function of $\xi_i \to \infty$ and that

$$\inf_{u \in (0,1)} \left( \pi(u/[u^{\alpha-1}/\{1 + \log(1 + 1/u)\}]^{1+\gamma}) \right) \geq 1(u < \xi_i) f(u/\xi_i) \to 1$$

for all $u \in (0, \infty)$. Then, by the dominated convergence theorem,

$$\frac{\lambda_i^2 p(\lambda_i)}{\pi(1/\xi_i)/\xi_i^2} \sim \int_0^\infty u^\alpha e^{-u} du = \Gamma(\alpha + 1),$$

and this proves part (i).

For part (ii), suppose first that $b < 1/2$. Then

$$p(\lambda_i) \sim \int_0^\infty u \pi(u) f(u) e^{-u \xi_i} du = \int_0^1 u \pi(u) f(u) e^{-u \xi_i} du + \int_1^\infty u \pi(u) f(u) e^{-u \xi_i} du$$

$$= \int_0^1 u \pi(u) f(u) e^{-u \xi_i} du + \int_\xi_i^{\infty} \frac{1}{\xi_i} \pi \left( \frac{u}{\xi_i} \right) f \left( \frac{u}{\xi_i} \right) e^{-u} du.$$ 

By the monotone convergence theorem,

$$\int_0^1 u \pi(u) f(u) e^{-u \xi_i} du \to \int_0^1 u \pi(u) f(u) du \leq \left[ \sup_{u \in (0,1)} \{ u f(u) \} \right] \int_0^1 \pi(u) du < \infty.$$
Meanwhile,

\[
\int_{\xi_i}^{\infty} \frac{1}{\xi_i} \frac{u}{\xi_i} \pi \left( \frac{u}{\xi_i} \right) f \left( \frac{u}{\xi_i} \right) e^{-u \frac{d\xi}{\xi_i}} \left\{ \frac{1}{\xi_i^2} \pi \left( \frac{1}{\xi_i} \right) \right\} = \int_{\xi_i}^{\infty} \frac{1}{u^{1/2+b}} \left( \frac{u}{\xi_i} \right)^{1+b} \pi \left( \frac{u}{\xi_i} \right) \left( \frac{u}{\xi_i} \right)^{1/2} f \left( \frac{u}{\xi_i} \right) e^{-u \frac{d\xi}{\xi_i}}.
\]

For all \( u \in (0, \infty) \), we have that

\[
\infty > \sup_{u \in (1, \infty)} \{ u^{1+b} \pi(u) \} \geq \frac{(u/\xi_i)^{1+b} \pi(u/\xi_i)}{(1/\xi_i)^{1+b} \pi(1/\xi_i)} \to 1
\]

by assumption and also that

\[
\frac{(u/\xi_i)^{1/2} f(u/\xi_i)}{(1/\xi_i)^{1/2} f(1/\xi_i)} = \frac{(u/\xi_i)^{u/\xi_i-1/2} e^{-u/\xi_i}}{\Gamma(u/\xi_i)} \frac{(1/\xi_i)^{1/\xi_i-1/2} e^{-1/\xi_i}}{\Gamma(1/\xi_i)} \left\{ \begin{array}{l} \leq \exp \left( -\frac{1}{12u/\xi_i} \right) \downarrow 1 \\ \geq \exp \left( -\frac{1}{12u/\xi_i} \right) \uparrow 1. \end{array} \right.
\]

Therefore, by the dominated convergence theorem,

\[
\int_{\xi_i}^{\infty} \frac{1}{u^{1/2+b}} \left( \frac{u}{\xi_i} \right)^{1+b} \pi \left( \frac{u}{\xi_i} \right) \left( \frac{u}{\xi_i} \right)^{1/2} f \left( \frac{u}{\xi_i} \right) e^{-u \frac{d\xi}{\xi_i}} \to \int_{0}^{\infty} u^{1/2-b-1} e^{-u \frac{d\xi}{\xi_i}} = \Gamma(1/2-b).
\]

Since

\[
\frac{1}{\xi_i^2} \pi \left( \frac{1}{\xi_i} \right) f \left( \frac{1}{\xi_i} \right) \sim \left( \frac{1}{\xi_i} \right)^{1/2-b} \frac{(1/\xi_i)^{1/\xi_i-1/2} e^{-1/\xi_i}}{\Gamma(1/\xi_i)} \to \infty,
\]

it follows that

\[
p(\lambda_i) \sim \int_{\xi_i}^{\infty} \frac{1}{\xi_i} \frac{u}{\xi_i} \pi \left( \frac{u}{\xi_i} \right) f \left( \frac{u}{\xi_i} \right) e^{-u \frac{d\xi}{\xi_i}} \sim \Gamma(1/2-b) \frac{1}{\xi_i^2} \pi \left( \frac{1}{\xi_i} \right) f \left( \frac{1}{\xi_i} \right) \to \infty.
\]

Next, suppose that \( b > 1/2 \). Then

\[
p(\lambda_i) \sim \int_{0}^{\infty} u \pi(u) f(u) e^{-u \xi_i} \frac{d\xi_i}{\xi_i} \to \int_{0}^{\infty} u \pi(u) f(u) du = \int_{0}^{\infty} u^{1/2} \pi(u) \frac{u^{u-1/2} e^{-u}}{\Gamma(u)} du \leq \frac{1}{(2\pi)^{1/2}} \int_{0}^{\infty} u^{1/2} \pi(u) du < \infty.
\]
by the monotone convergence theorem. Finally, suppose that \( b = \frac{1}{2} \). Then

\[
p(\lambda_i) \sim \int_0^\infty u\pi(u)f(u)e^{-u\xi_i}du \\
\quad \rightarrow \int_0^\infty u\pi(u)f(u)du = \int_0^\infty u^{1/2}\pi(u)\frac{u^{-1/2}e^{-u}}{\Gamma(u)}du \geq \frac{1}{(2\pi)^{1/2}} \int_0^\infty u^{1/2}\pi(u)e^{-1/(12u)}du = \infty.
\]

by the monotone convergence theorem.

For parts (iii) and (iv),

\[
p(\lambda_i) \sim \int_0^1 u\pi(u)f(u)e^{-u\xi_i}du + \int_1^\infty u\pi(u)f(u)e^{-u\xi_i}du,
\]

where

\[
\int_0^1 u\pi(u)f(u)e^{-u\xi_i}du \rightarrow \int_0^1 u\pi(u)f(u)du < \infty
\]
as in part (ii). By integration by parts,

\[
\int_1^\infty u\pi(u)f(u)e^{-u\xi_i}du = \int_1^\infty \frac{1}{u}u^{2}\pi(u)f(u)e^{-u\xi_i}du \\
= \left[ (\log u)u^{2}\pi(u)f(u)e^{-u\xi_i} \right]_1^\infty - \int_1^\infty (\log u)u^{2}\pi(u)f(u)e^{-u\xi_i}\left\{ \frac{2}{u} + \frac{\pi'(u)}{\pi(u)} + \frac{f'(u)}{f(u)} - \xi_i \right\}du \\
= -\int_1^\infty (\log u)u^{2}\pi(u)f(u)e^{-u\xi_i}\left\{ \frac{2}{u} + \frac{\pi'(u)}{\pi(u)} + \frac{f'(u)}{f(u)} \right\}du + \xi_i \int_1^\infty (\log u)u^{2}\pi(u)f(u)e^{-u\xi_i}du.
\]

Since

\[
\frac{2}{u} + \frac{\pi'(u)}{\pi(u)} + \frac{f'(u)}{f(u)} = \frac{a}{u(1+u)} + \log u - \psi(u) - \frac{1}{u} = \frac{a + 1/2}{u(1+u)} + 2 \int_0^\infty \frac{1}{t^2 + u^2 e^{2\pi t}} - 1 dt \geq 0
\]

for all \( u \in (0, \infty) \) for part (iii) and since

\[
\frac{2}{u} + \frac{\pi'(u)}{\pi(u)} + \frac{f'(u)}{f(u)} = \frac{1/2}{u(1+u)} + \frac{1/2}{u(1+u)} \left\{ \frac{1}{\log(1+1/u)} - u \right\} \\
+ \frac{1/2 + a}{1 + \log(1+1/u) u(1+u)} + 2 \int_0^\infty \frac{1}{t^2 + u^2 e^{2\pi t} - 1} dt \geq 0
\]
for all \( u \in (0, \infty) \) for part (iv), we have, by the monotone convergence theorem,

\[
\int_1^\infty (\log u)u^2\pi(u)f(u)e^{-u\xi_i} \left\{ \frac{2}{u} + \frac{\pi'(u)}{\pi(u)} + \frac{f'(u)}{f(u)} \right\} du
\]

\[
\to \int_1^\infty (\log u)u^2\pi(u)f(u) \left\{ \frac{a + 1/2}{u(1 + u)} + 2 \int_0^\infty \frac{1}{t^2 + u^2 e^{2\pi t} - 1} dt \right\} du
\]

\[
\leq \left[ \sup_{u \in (1, \infty)} \{ u^{3/2}\pi(u)u^{1/2}f(u) \} \right] \int_1^\infty \frac{\log u}{u(1 + u)} \left\{ a + \frac{1}{2} + 2u(1 + u) \int_0^\infty \frac{t}{e^{2\pi t} - 1} dt \right\} du < \infty
\]

for part (iii) and

\[
\int_1^\infty (\log u)u^2\pi(u)f(u)e^{-u\xi_i} \left\{ \frac{2}{u} + \frac{\pi'(u)}{\pi(u)} + \frac{f'(u)}{f(u)} \right\} du
\]

\[
\to \int_1^\infty (\log u)u^2\pi(u)f(u) \left\{ \frac{1/2}{u(1 + u)} + \frac{1/2}{u(1 + u)} \left\{ \log(1 + 1/u) - u \right\} + \frac{1/2 + a}{1 + \log(1 + 1/u)} \frac{1}{u(1 + u)} + 2 \int_0^\infty \frac{1}{t^2 + u^2 e^{2\pi t} - 1} dt \right\} du
\]

\[
\leq \left[ \sup_{u \in (1, \infty)} \{ u^{3/2}\pi(u)u^{1/2}f(u) \} \right] \int_1^\infty \frac{\log u}{u(1 + u)} \left\{ \frac{1}{2} + \frac{1}{2}(1 + u - u) + \frac{1/2 + a}{1 + \log(1 + 1/u)} + 2u(1 + u) \int_0^\infty \frac{t}{e^{2\pi t} - 1} dt \right\} du < \infty
\]

for part (iv). Meanwhile,

\[
\xi_i \int_1^\infty (\log u)u^2\pi(u)f(u)e^{-u\xi_i} du / \left\{ \left( \log \frac{1}{\xi_i} \right) (1/\xi_i)^2 \pi(1/\xi_i)f(1/\xi_i) \right\}
\]

\[
= \int_1^\infty \frac{\log(u/\xi_i)}{\log(1/\xi_i)} u^2\pi(u/\xi_i) f(u/\xi_i) e^{-u} du / \left\{ \left( \log \frac{1}{\xi_i} \right) (1/\xi_i)^2 \pi(1/\xi_i)f(1/\xi_i) \right\}
\]

Note that

\[
\frac{\log(u/\xi_i)}{\log(1/\xi_i)} u^2\pi(u/\xi_i) f(u/\xi_i) e^{-u} \to e^{-u}
\]
for all \( u \in (0, \infty) \). Also, note that

\[
0 \leq \log(u/\xi_i) u^2 \frac{\pi u/\xi_i}{\pi(1/\xi_i)} f(u/\xi_i) e^{-u} \\
\leq \{1 + \log(1 + u)\} u^2 a^{-1} \left( \frac{1}{1 + u/\xi_i} \right)^{a+1/2} \frac{1}{u^{1/2}} e^{\xi_i/12} e^{-u} \\
\leq \{1 + \log(1 + u)\} u^2 a^{-1} \left( 1 + \frac{1}{u^{a+1/2}} \right) \frac{1}{u^{1/2}} e^{1/12} e^{-u} \\
= e^{1/12} \{1 + \log(1 + u)\} (u^{a+1/2} + 1) e^{-u}
\]

for all \( u \in (\xi_i, \infty) \) for part (iii) and that

\[
0 \leq \log(u/\xi_i) u^2 \frac{\pi u/\xi_i}{\pi(1/\xi_i)} f(u/\xi_i) e^{-u} \\
\leq \{1 + \log(1 + u)\} u^2 \left( \frac{1}{1 + u/\xi_i} \right)^{1/2} \left( \frac{1 + \log(1 + \xi_i)}{1 + \log(1 + \xi_i/u)} \right)^{1/2+a} \frac{1}{u^{1/2}} e^{\xi_i/12} e^{-u} \\
\leq \{1 + \log(1 + u)\} u^2 \left( 1 + \frac{1}{u} \right) \left( \frac{\xi_i}{(1 + 1/u)} \right)^{1/2} (1 + u)^{1/2+a} \frac{1}{u^{1/2}} e^{1/12} e^{-u} \\
= e^{1/12} \{1 + \log(1 + u)\} (1 + u)^{3/2+a} \left( 1 + \frac{1}{u} \right)^{1/2} e^{-u}
\]

for all \( u \in (\xi_i, \infty) \) for part (iv). Then, by the dominated convergence theorem,

\[
\int_{\xi_i}^{\infty} \frac{\log(u/\xi_i) u^2 \pi(u/\xi_i) f(u/\xi_i)}{\log(1/\xi_i) \pi(1/\xi_i) f(1/\xi_i)} e^{-u} du \to \int_0^{\infty} e^{-u} du = 1.
\]

Thus, we conclude that

\[
p(\lambda_i) \sim \left( \log \frac{1}{\xi_i} \right) (1/\xi_i)^2 \pi(1/\xi_i) f(1/\xi_i) \to \infty.
\]

This completes the proof. \qed

S3 Lemmas

In this section, we prove four lemmas, which will be used in the next section. Let

\[
\varphi(u) = u \log \left( 1 + \frac{1}{u} \right)
\]

for \( u \in (0, \infty) \).
Lemma S1. The function \( \varphi(\cdot) \) has the following properties.

(i) \( \varphi'(u) = \log(1 + 1/u) - 1/(1 + u) > 0 \) for all \( u \in (0, \infty) \).

(ii) \( \varphi''(u) < 0 \) for all \( u \in (0, \infty) \).

(iii) \( \lim_{u \to 0} \varphi(u) = 0 \) and \( \lim_{u \to \infty} \varphi(u) = 1 \).

(iv) \( y \varphi^{-1}(v/y) = v/\log\{1 + 1/\varphi^{-1}(v/y)\} \to 0 \) as \( v < y \to \infty \) for all \( v \in (0, \infty) \).

(v) \( \varphi'(u) \geq 1/\{2(1 + u)^2\} \) for all \( u \in (0, \infty) \).

(vi) There exists \( 0 < c_1 < 1 \) such that \( \varphi'(\varphi^{-1}(1/y))/\varphi'(\varphi^{-1}(v/y)) \leq (1/v)\{\varphi^{-1}(v/y)/\varphi^{-1}(1/y)\}/(1 - c_1) \) for all \( v, y \in (0, \infty) \) satisfying \( y > \max\{1, v\} \).

(vii) There exists \( c_2 > 0 \) such that \( \varphi^{-1}(v/y)/\varphi^{-1}(1/y) \leq v^{c_2} \) for all \( v, y \in (0, \infty) \) satisfying \( v > 1 \) and \( y > 2v \).

Proof. Parts (i), (ii), and (iii) are trivial. Part (iv) follows since

\[
u/y = \varphi(\varphi^{-1}(u/y)) = \varphi^{-1}(u/y) \log\{1 + 1/\varphi^{-1}(u/y)\}.
\]

Part (v) follows since, by part (i),

\[
\varphi'(u) = \sum_{k=1}^{\infty} \frac{1}{k} \frac{1}{(1 + u)^k} - \frac{1}{1 + u} = \frac{1}{2(1 + u)^2} + \sum_{k=3}^{\infty} \frac{1}{k} \frac{1}{(1 + u)^k}.
\]

For part (vi), let

\[
c_1 = \sup_{0 < t < 1/2} \frac{1/\{1 + \varphi^{-1}(t)\}}{\log\{1 + 1/\varphi^{-1}(t)\}} < \infty.
\]

Then

\[
c_1 = \sup_{0 < u < \varphi^{-1}(1/2)} \frac{1/(1 + u)}{\log(1 + 1/u)} \leq \sup_{0 < u < \varphi^{-1}(1/2)} \frac{1/(1 + u)}{1/(1 + u) + (1/2)/(1 + u)^2} = \sup_{0 < u < \varphi^{-1}(1/2)} \frac{1 + u}{1 + u + 1/2} < 1.
\]
Therefore,

\[
\frac{\varphi'(\varphi^{-1}(1/y))}{\varphi'(\varphi^{-1}(v/y))} = \frac{\log\{1 + 1/\varphi^{-1}(1/y)\} - 1/\{1 + \varphi^{-1}(1/y)\}}{\log\{1 + 1/\varphi^{-1}(v/y)\} - 1/\{1 + \varphi^{-1}(v/y)\}}
\]

\[
\leq \frac{\log\{1 + 1/\varphi^{-1}(1/y)\}}{\log\{1 + 1/\varphi^{-1}(v/y)\} - 1/\{1 + \varphi^{-1}(v/y)\}}
\]

\[
= \frac{\log\{1 + 1/\varphi^{-1}(1/y)\}}{\log\{1 + 1/\varphi^{-1}(v/y)\}}\left[1 - \frac{1/\{1 + \varphi^{-1}(v/y)\}}{\log\{1 + 1/\varphi^{-1}(v/y)\}}\right]
\]

\[
\leq \frac{\log\{1 + 1/\varphi^{-1}(1/y)\} 1}{\log\{1 + 1/\varphi^{-1}(v/y)\} 1 - c_1} = \frac{1}{v} \varphi^{-1}(1/y) 1 - c_1,
\]

where the last equality follows from part (iv). For part (vii), let

\[
c_2 = \sup_{0 < u < \varphi^{-1}(1/2)} \left\{1 + \frac{1}{1 + u} \sum_{k=2}^{\infty} \frac{1}{k (1 + u)^k}\right\} < \infty.
\]

Then, for any \(1 < s < v\), since \(s/y < 1/2\),

\[
\frac{s/y}{\varphi^{-1}(s/y)} \frac{1}{\varphi'(\varphi^{-1}(s/y))} \leq \sup_{0 < t < 1/2} \left\{\frac{s}{\varphi^{-1}(t) \varphi'(\varphi^{-1}(t))}\right\} \leq \sup_{0 < u < \varphi^{-1}(1/2)} \left\{\frac{\varphi(u)}{u} \frac{1}{\varphi'(u)}\right\}
\]

\[
= \sup_{0 < u < \varphi^{-1}(1/2)} \frac{\log(1 + 1/u) - 1/(1 + u)}{\log(1 + 1/u)} = c_2,
\]

where the equality follows from part (i). Thus,

\[
\frac{\varphi^{-1}(v/y)}{\varphi^{-1}(1/y)} = \exp\left\{\left[\log \varphi^{-1}(s/y)\right]_{s=1}^{s=v}\right\} = \exp\left\{\int_{s=1}^{s=v} \frac{1}{s} \frac{s/y}{\varphi^{-1}(s/y)} \frac{1}{\varphi'(\varphi^{-1}(s/y))} ds\right\}
\]

\[
\leq \exp\left(\int_{s=1}^{s=v} \frac{c_2 ds}{s}\right) = v^{c_2}.
\]

This completes the proof. \(\square\)

For \(\varepsilon \geq 0\) and \(j = 0, 1\), let

\[
I_{j,\varepsilon}(y; \delta) = \int_{0}^{\infty} \pi(u) \frac{\Gamma(u + \delta + 1)}{u \Gamma(u)(y + u)^{\delta+1}} \left(\frac{u}{y + u}\right)^u \left(\frac{u}{\delta + u}\right)^j e^{-\varepsilon u} du
\]

for \(y \in (0, \infty)\) for \(\delta \in (0, \infty)\).

**Lemma S2.** Let \(\delta \in (0, \infty)\).
For any \( y \in (\delta + 1, \infty) \) and any \( \varepsilon > 0 \),

\[
1 \leq \frac{I_{1,0}(y; \delta)/I_{1,\varepsilon}(y; \delta)}{I_{0,0}(y; \delta)/I_{0,\varepsilon}(y; \delta)} \leq \frac{I_{1,0}(\delta + 1; \delta)}{I_{1,\varepsilon}(\delta + 1; \delta)}.
\]

(ii) As \( \varepsilon \to 0 \),

\[
\frac{I_{1,0}(\delta + 1; \delta)}{I_{1,\varepsilon}(\delta + 1; \delta)} = 1.
\]

Proof. Let Part (ii) is trivial. For part (i), we have by the covariance inequality that

\[
\frac{I_{1,0}(y; \delta)/I_{1,\varepsilon}(y; \delta)}{I_{0,0}(y; \delta)/I_{0,\varepsilon}(y; \delta)} = \frac{I_{1,0}(y; \delta)}{I_{0,0}(y; \delta)} \frac{I_{0,\varepsilon}(y; \delta)}{I_{1,\varepsilon}(y; \delta)} \geq 1.
\]

On the other hand,

\[
\frac{I_{1,0}(y; \delta)/I_{1,\varepsilon}(y; \delta)}{I_{0,0}(y; \delta)/I_{0,\varepsilon}(y; \delta)} \leq \frac{I_{1,0}(y; \delta)}{I_{1,\varepsilon}(y; \delta)}.
\]

By the covariance inequality,

\[
\frac{\partial}{\partial y} \log I_{1,0}(y; \delta)/I_{1,\varepsilon}(y; \delta) = -\int_0^\infty \frac{u + \delta + 1}{y + u} d\mu(u)/\int_0^\infty d\mu(u)
+ \int_0^\infty \frac{u + \delta + 1}{y + u} e^{-\varepsilon u} d\mu(u)/\int_0^\infty e^{-\varepsilon u} d\mu(u) \leq 0,
\]

where

\[
d\mu(u) = \pi(u) \frac{\Gamma(u + \delta + 1)}{\Gamma(u)(y + u)^{\delta+1}} \left( \frac{u}{y + u} \right)^{u \delta + 1} du
\]

for \( u \in (0, \infty) \). Therefore,

\[
\frac{I_{1,0}(y; \delta)}{I_{1,\varepsilon}(y; \delta)} \leq \frac{I_{1,0}(\delta + 1; \delta)}{I_{1,\varepsilon}(\delta + 1; \delta)},
\]

and this completes the proof.

In the remainder of this section, we fix \( \delta > 0 \). For \( \varepsilon \geq 0 \) and \( j = 0, 1 \), the integral
$I_j,\varepsilon(y; \delta)$ is denoted by $I_j,\varepsilon(y)$ for $y \in (0, \infty)$ for $\delta \in (0, \infty)$. For $\varepsilon > 0$, let

$$f_j,\varepsilon(v; y) = 1(0 < v < y) \frac{1}{\varphi'(\varphi^{-1}(v/y))} \pi(y\varphi^{-1}(v/y)) \times \frac{\Gamma(y\varphi^{-1}(v/y) + \delta + 1)}{\Gamma(y\varphi^{-1}(v/y)) \gamma^{\delta+1} \{1 + \varphi^{-1}(v/y)\}^{\delta+1} e^{-v} \{\frac{y\varphi^{-1}(v/y)}{\delta + y\varphi^{-1}(v/y)}\}^j e^{-y\varphi^{-1}(v/y)}$$

for $v, y \in (0, \infty)$ for $j = 0, 1$.

**Lemma S3.** For any $y \in (0, \infty)$, any $\varepsilon > 0$, and any $j = 0, 1$, we have

$$I_j,\varepsilon(y) = \int_0^\infty f_j,\varepsilon(v; y)dv.$$

**Proof.** We have

$$I_j,\varepsilon(y) = \int_0^\infty \pi(u) \frac{\Gamma(u + \delta + 1)}{\Gamma(u)(y + u)^{\delta+1}} \left(\frac{u}{y + u}\right)^u \frac{1}{\delta + u} \left(\frac{u}{\delta + u}\right)^j e^{-\varepsilon u} du$$

$$= \int_0^\infty \frac{\Gamma(yu + \delta + 1)}{\Gamma(yu) y^{\delta+1} (1 + u)^{\delta+1}} \left(\frac{yu}{\delta + yu}\right)^j e^{-\varepsilon yu} du$$

$$= \int_0^1 \frac{1}{\varphi'(\varphi^{-1}(t))} \pi(y\varphi^{-1}(t)) \frac{\Gamma(y\varphi^{-1}(t) + \delta + 1)}{\Gamma(y\varphi^{-1}(t)) y^{\delta+1} \{1 + \varphi^{-1}(t)\}^{\delta+1} e^{-yt} \{\frac{y\varphi^{-1}(t)}{\delta + y\varphi^{-1}(t)}\}^j e^{-y\varphi^{-1}(t)} dt$$

$$= \int_0^y \frac{1}{\varphi'(\varphi^{-1}(v/y))} \pi(y\varphi^{-1}(v/y)) \times \frac{\Gamma(y\varphi^{-1}(v/y) + \delta + 1)}{\Gamma(y\varphi^{-1}(v/y)) y^{\delta+1} \{1 + \varphi^{-1}(v/y)\}^{\delta+1} e^{-v} \{\frac{y\varphi^{-1}(v/y)}{\delta + y\varphi^{-1}(v/y)}\}^j e^{-y\varphi^{-1}(v/y)} dv,$$

which is the desired result.

Let

$$g_j(v; y) = \frac{\Gamma(\delta + 1)}{\delta^j} Cy^{j+\alpha-\delta-1} \frac{\{\varphi^{-1}(1/y)\}^{j+\alpha}}{\varphi'((\varphi^{-1}(1/y)) [1 + \log(1 + (1/y) / \varphi^{-1}(1/y))]^{j+\gamma} y^{j+\alpha} e^{-v}}$$

for $v \in (0, \infty)$ and $y \in (1, \infty)$ and $j = 0, 1$.

**Lemma S4.** Let $j = 0, 1$ and $\varepsilon > 0$.

(i) For any $v \in (0, \infty)$, we have $f_j,\varepsilon(v; y) \sim g_j(v; y)$ as $y \to \infty$.

(ii) There exists an integrable function $h_j,\varepsilon(v)$ of $v \in (0, \infty)$ such that $v^{j+\alpha} e^{-v} f_j,\varepsilon(v; y)/g_j(v; y) \leq h_j,\varepsilon(v)$ for all $v \in (0, \infty)$ and $y \in (1, \infty)$.
Proof. For part (i),

\[
f_{j,\epsilon}(v; y) \sim \frac{\pi(y\varphi^{-1}(v/y))}{\varphi'(\varphi^{-1}(v/y))} \Gamma(\delta + 1) \frac{\varphi^{-1}(v/y) e^{-v \{y\varphi^{-1}(v/y)\}^j}}{y^\delta} \\
\sim \frac{1}{\varphi'(\varphi^{-1}(v/y))} \frac{C\{y\varphi^{-1}(v/y)\}^{\alpha-1}}{(1 + \log[1 + 1/\{y\varphi^{-1}(v/y)\}]^{1+\gamma})} \Gamma(\delta + 1) \frac{\varphi^{-1}(v/y) e^{-v \{y\varphi^{-1}(v/y)\}^j}}{y^\delta} \\
\sim \frac{\Gamma(\delta + 1) Cy^{1+\alpha-1}}{\delta^j \varphi'(\varphi^{-1}(v/y))} \frac{\{\varphi^{-1}(v/y)\}^{\gamma+\alpha}}{[1 + \log[1 + (v/y)/\varphi^{-1}(v/y)]]^{1+\gamma}} e^{-v}
\]

as \( y \to \infty \) by part (iv) of Lemma S1, condition (4), and the fact that \( 1 + \log(1 + 1/u) \) is a slowly-varying function of \( u \to 0 \). Note that \( \frac{u\varphi'(u)}{\varphi(u)} \to 1 \) as \( u \to 0 \). Then

\[
\frac{t(\varphi^{-1})'(t)}{\varphi^{-1}(t)} = \frac{\varphi(\varphi^{-1}(t))}{\varphi'(\varphi^{-1}(t))\varphi^{-1}(t)} \to 1
\]

and

\[
\frac{t(\varphi' \circ \varphi^{-1})'(t)}{(\varphi' \circ \varphi^{-1})(t)} = \frac{\varphi^{-1}(t)\varphi''(\varphi^{-1}(t))/\varphi'(\varphi^{-1}(t))}{\varphi^{-1}(t)\varphi'(\varphi^{-1}(t))/\varphi(\varphi^{-1}(t))} \to 0 \tag{S1}
\]

as \( t \to 0 \). Also,

\[
\lim_{t \to 0} \frac{t\varphi'(t)}{\varphi(t)} = \lim_{t \to 0} \frac{[t/\{1 + t/\varphi^{-1}(t)\}]\varphi^{-1}(t) - [t/\{\varphi^{-1}(t)\}^2]/\varphi'(\varphi^{-1}(t))}{1 + \log\{1 + t/\varphi^{-1}(t)\}}
\]

\[
= \lim_{u \to 0} \frac{\varphi(u)/u}{1 + \varphi(u)/u} \frac{1 - 1/[u\varphi'(u)/\varphi(u)]}{1 + \log(1 + \varphi(u)/u)} = 0,
\]

where \( \varphi(t) = 1 + \log(1 + t/\varphi^{-1}(t)) \) for \( t \in (0, 1) \). Therefore,

\[
\varphi^{-1}(v/y) \sim v\varphi^{-1}(1/y),
\]

\[
\varphi'(\varphi^{-1}(v/y)) \sim \varphi'(\varphi^{-1}(1/y)), \quad \text{and}
\]

\[
1 + \log\{1 + (v/y)/\varphi^{-1}(v/y)\} \sim 1 + \log\{1 + (1/y)/\varphi^{-1}(1/y)\}
\]

as \( y \to \infty \) by part (iv) of Lemma S1, condition (4), and the fact that \( 1 + \log(1 + 1/u) \) is a slowly-varying function of \( u \to 0 \).
as \( y \to \infty \). Thus,

\[
\frac{f_{j,\varepsilon}(v; y)}{g_j(v; y)} \sim \frac{\Gamma(\delta + 1) Cy^{j+\alpha - \delta - 1}}{\delta^j \varphi'(-1(1/y)) \Gamma(y\varphi^{-1}(1/y))} \frac{\{v\varphi^{-1}(1/y)\}^{j+\alpha}}{[1 + \log(1 + (1/y)/\varphi^{-1}(1/y))]^{1+\gamma}} e^{-v} = g_j(v; y).
\]

For part (ii), we assume that \( y \) is sufficiently large and we consider the ratio

\[
\frac{f_{j,\varepsilon}(v; y)}{g_j(v; y)} = 1(0 < y < y_j) \frac{\varphi'(-1(1/y))}{\varphi'(-1(1/y))} \frac{\Gamma(y\varphi^{-1}(1/y) + \delta + 1)}{\Gamma(y\varphi^{-1}(1/y))} \frac{y^{j+\alpha} \{1 + \varphi^{-1}(1/y)\}^{\delta+1}}{\{1 + (1/y)/\varphi^{-1}(1/y)\}^{1+\gamma}}
\]

\[
\times \left\{ \frac{y\varphi^{-1}(v/y)}{\delta + y\varphi^{-1}(v/y)} \right\}^j e^{-y\varphi^{-1}(v/y)} \frac{\delta^j}{\Gamma(\delta + 1) Cy^{j+\alpha - \delta - 1}} \frac{\{v\varphi^{-1}(1/y)\}^{j+\alpha}}{[1 + \log(1 + (1/y)/\varphi^{-1}(1/y))]^{1+\gamma}}.
\]

Since

\[
(2\pi)^{1/2}x^{-1/2}e^{-x} \leq \Gamma(x) \leq (2\pi)^{1/2}x^{-1/2}e^{-x}e^1/(12x)
\]

for all \( x \in (0, \infty) \),

\[
\frac{\Gamma(y\varphi^{-1}(1/y) + \delta + 1)}{\Gamma(y\varphi^{-1}(1/y) + 1) y^{\delta+1}} \frac{\{1 + \varphi^{-1}(1/y)\}^{\delta+1}}{\{1 + (1/y)/\varphi^{-1}(1/y)\}^{1+\gamma}}
\]

\[
\leq \frac{\varphi^{-1}(v/y) + \delta + 1}{\varphi^{-1}(v/y) + 1} \frac{\varphi^{-1}(v/y) + 1}{\varphi^{-1}(v/y)} \frac{\varphi^{-1}(v/y) + 1}{\varphi^{-1}(v/y)} \frac{\varphi^{-1}(v/y) + 1}{\varphi^{-1}(v/y)}
\]

\[
\leq \frac{\varphi^{-1}(v/y) + \delta + 1}{\varphi^{-1}(v/y) + 1} \frac{\varphi^{-1}(v/y) + 1}{\varphi^{-1}(v/y)} \frac{\varphi^{-1}(v/y) + 1}{\varphi^{-1}(v/y)} \frac{\varphi^{-1}(v/y) + 1}{\varphi^{-1}(v/y)}
\]

\[
\leq 1 + \frac{\varphi^{-1}(v/y) + \delta + 1}{\varphi^{-1}(v/y) + 1} \frac{\varphi^{-1}(v/y) + 1}{\varphi^{-1}(v/y)} \frac{\varphi^{-1}(v/y) + 1}{\varphi^{-1}(v/y)} \frac{\varphi^{-1}(v/y) + 1}{\varphi^{-1}(v/y)}
\]

where the second inequality follows from the fact that \( 1 \leq (1 + 1/x)^x \leq e \) for all \( x > 0 \).

Therefore,

\[
\frac{f_{j,\varepsilon}(v; y)}{g_j(v; y)} \leq 1(0 < y < y_j) \frac{\varphi'(-1(1/y))}{\varphi'(-1(1/y))} \frac{\Gamma(y\varphi^{-1}(1/y) + \delta + 1)}{\Gamma(y\varphi^{-1}(1/y))} \frac{y^{j+\alpha} \{1 + \varphi^{-1}(1/y)\}^{\delta+1}}{\{1 + (1/y)/\varphi^{-1}(1/y)\}^{1+\gamma}}
\]

\[
\times \left\{ \frac{y\varphi^{-1}(v/y)}{\delta + y\varphi^{-1}(v/y)} \right\}^j e^{-y\varphi^{-1}(v/y)} \frac{\delta^j}{\Gamma(\delta + 1) Cy^{j+\alpha - \delta - 1}} \frac{\{v\varphi^{-1}(1/y)\}^{j+\alpha}}{[1 + \log(1 + (1/y)/\varphi^{-1}(1/y))]^{1+\gamma}}.
\]
First, suppose that $v > y/2$. Then $v/y > 1/2$ and $y\varphi^{-1}(v/y) \geq y\varphi^{-1}(1/2) \geq \delta + 1 \geq 1$. Therefore, by parts (i) and (v) of Lemma S1 and by condition (3),

\[
\frac{f_{j,\varepsilon}(v; y)}{g_j(v; y)} \leq 1(0 < v < y) \varphi'(\varphi^{-1}(1/y)) \pi(y\varphi^{-1}(v/y)) \times 2^{\delta + 1/2} \left\{ y\varphi^{-1}(v/y) \right\}^{\delta + 1/2} \frac{\varphi^{-1}(v/y) \varphi'(v/y)}{y\varphi^{-1}(v/y)} \frac{\varepsilon^2}{1 + \varphi^{-1}(v/y)} 
\times e^{-\varepsilon y\varphi^{-1}(v/y)} \delta + 1 \frac{1/y^{j+\alpha}}{(\varphi^{-1}(1/y))^{j+\alpha}} \frac{\log\{1 + \varphi^{-1}(1/y)\}}{(\varphi^{-1}(1/y))^{j+\alpha}} \times \frac{1}{y} 
\leq M_1(0 < v < y) \left\{ 1 + \frac{1}{\varphi^{-1}(1/y)} \right\} \left\{ 1 + \varphi^{-1}(v/y) \right\}^2 \frac{1}{y\varphi^{-1}(v/y)} e^{-\varepsilon y\varphi^{-1}(v/y)/3} \times y 
\times 2^{\delta + 1/2} \left\{ y\varphi^{-1}(v/y) \right\}^{\delta + 1/2} \frac{\varphi^{-1}(v/y) \varphi'(v/y)}{y\varphi^{-1}(v/y)} \frac{\varepsilon^2}{1 + \varphi^{-1}(v/y)} 
\times e^{-\varepsilon y\varphi^{-1}(v/y)/3} \frac{1/y^{j+\alpha}}{(\varphi^{-1}(1/y))^{j+\alpha}} \frac{\log\{1 + \varphi^{-1}(1/y)\}}{(\varphi^{-1}(1/y))^{j+\alpha}} \times \frac{1}{y} 
\leq M_2(0 < v < y) \left\{ 1 + \frac{1}{\varphi^{-1}(1/y)} \right\} \left\{ 1 + \varphi^{-1}(v/y) \right\}^2 \frac{1}{y\varphi^{-1}(v/y)} e^{-\varepsilon y\varphi^{-1}(v/y)/3} \times y 
\times 2^{\delta + 1/2} \left\{ y\varphi^{-1}(v/y) \right\}^{\delta + 1/2} \frac{\varphi^{-1}(v/y) \varphi'(v/y)}{y\varphi^{-1}(v/y)} \frac{\varepsilon^2}{1 + \varphi^{-1}(v/y)} 
\times e^{-\varepsilon y\varphi^{-1}(v/y)/3} \frac{1/y^{j+\alpha}}{(\varphi^{-1}(1/y))^{j+\alpha}} \frac{\log\{1 + \varphi^{-1}(1/y)\}}{(\varphi^{-1}(1/y))^{j+\alpha}} \times \frac{1}{y} 
\text{for some } M_1, M_2 > 0. \text{ Thus,} 
\frac{f_{j,\varepsilon}(v; y)}{g_j(v; y)} \leq \frac{M_2 e^{-\varepsilon y\varphi^{-1}(1/2)/3}}{v^{j+\alpha}} \left\{ 1 + \log\{1 + (1/y)/\varphi^{-1}(1/y)\} \right\}^{j+\alpha+1} \frac{1/y^{j+\alpha}}{(\varphi^{-1}(1/y))^{j+\alpha+1}} \frac{\log\{1 + \varphi^{-1}(1/y)\}}{(\varphi^{-1}(1/y))^{j+\alpha+1}} \times \frac{1}{y} 
\leq \frac{M_2}{v^{j+\alpha}} \exp\left( - \left[ \varepsilon' \frac{1}{1/y} - (j + \alpha + 2 + \gamma) \log\{1 + \frac{1}{\varphi^{-1}(1/y)}\} \right] \right) \leq \frac{M_2}{v^{j+\alpha}} \text{ for } \varepsilon' = \varepsilon \varphi^{-1}(1/2)/3, \text{ where the third inequality follows since} 
\frac{1}{\varphi^{-1}(1/y)} \frac{1 + \log\{1 + \frac{1/y}{\varphi^{-1}(1/y)}\}}{1 + \frac{1/y}{\varphi^{-1}(1/y)}} \leq 1 + \frac{1}{\varphi^{-1}(1/y)} 
\text{and the last inequality follows since} 
\varepsilon' = \frac{\varepsilon' \varphi^{-1}(1/y) \log\{1 + 1/\varphi^{-1}(1/y)\}}{(1 + 1/\varphi^{-1}(1/y))} \geq (j + \alpha + 2 + \gamma) \log\{1 + \frac{1}{\varphi^{-1}(1/y)}\} 
\text{for sufficiently large } y > 0. \text{ Hence, } v^{j+\alpha} e^{-v} f_{j,\varepsilon}(v; y)/g_j(v; y) \leq M_2 e^{-v}. \text{ Next, suppose}
that \( 1 < v < y/2 \). Then

\[
\frac{f_{j,v}(v;y)}{g_j(v;y)} \leq 1(0 < v < y) \frac{\pi(y\varphi^{-1}(v/y))}{C(y\varphi^{-1}(v/y))^{\alpha - 1}/(1 + \log[1 + 1/(y\varphi^{-1}(v/y))])^{1+\gamma}} \\
\times (Cy^{\alpha - 1}\{\varphi^{-1}(v/y)\}^{\alpha - 1}/[1 + \log\{1 + (1/y)/\varphi^{-1}(v/y)\}]^{1+\gamma}) \\
\times \varphi'(\varphi^{-1}(v/y)) \{y\varphi^{-1}(v/y) + \delta + 1\}^{\delta + 1/2} \varphi^{-1}(v/y) \\
\times \frac{\{y\varphi^{-1}(v/y)\}^j e^{-\varepsilon y\varphi^{-1}(v/y)}}{\Gamma(\delta + 1)} e^{-\varepsilon y\varphi^{-1}(v/y)} 1/y^{\delta + \alpha} \frac{1}{\Gamma(\delta + 1)} 1_{y^{\delta + \alpha}} \\
\leq 1(0 < v < y) \frac{\pi(y\varphi^{-1}(v/y))}{C(y\varphi^{-1}(v/y))^{\alpha - 1}/(1 + \log[1 + 1/(y\varphi^{-1}(v/y))])^{1+\gamma}} \\
\times \frac{1 + \log\{1 + (1/y)/\varphi^{-1}(v/y)\}}{[1 + \log\{1 + (1/y)/\varphi^{-1}(v/y)\}]^{\delta + 1/2}} e^{-\varepsilon y\varphi^{-1}(v/y)} \frac{1}{\Gamma(\delta + 1)} 1_{y^{\delta + \alpha}} \\
\leq 1(0 < v < y) \frac{1}{C(y\varphi^{-1}(v/y))^{\alpha - 1}/(1 + \log[1 + 1/(y\varphi^{-1}(v/y))])^{1+\gamma}} \\
\times \frac{[1 + \log\{1 + (1/y)/\varphi^{-1}(v/y)\}]^{\delta + 1/2}}{[1 + \log\{1 + (1/y)/\varphi^{-1}(v/y)\}]^{\delta + 1/2}} e^{-\varepsilon y\varphi^{-1}(v/y)} \frac{1}{\Gamma(\delta + 1)} 1_{y^{\delta + \alpha}}.
\]

Here, by conditions (4) and (3),

\[
1(0 < v < y) \frac{\pi(y\varphi^{-1}(v/y))}{C(y\varphi^{-1}(v/y))^{\alpha - 1}/(1 + \log[1 + 1/(y\varphi^{-1}(v/y))])^{1+\gamma}} \\
\leq 1(0 < v < y) \left[ 1(y\varphi^{-1}(v/y) < 1)M_3 + \frac{1}{C(y\varphi^{-1}(v/y))^{\alpha - 1}/(1 + \log[1 + 1/(y\varphi^{-1}(v/y))])^{1+\gamma}} \right] \leq M_5
\]
for some $M_3, M_4, M_5 > 0$. Also, by parts (vi) and (vii) of Lemma $S1$

\[
\frac{1 + \log\{1 + (1/y)/\varphi^{-1}(1/y)\}}{1 + \log\{1 + (1/y)/\varphi^{-1}(v/y)\}} \leq \frac{1 + \gamma_1 \varphi^{-1}(v/y) - 1}{1 - c_1} \leq \frac{1 + \gamma_1 \varphi^{-1}(v/y)}{1 - c_1} \leq \frac{1 + \gamma_1 \varphi^{-1}(v/y)}{1 - c_1} \leq 1 \leq \frac{1 + \gamma_1 \varphi^{-1}(v/y)}{1 - c_1} \leq \frac{1 + \gamma_1 \varphi^{-1}(v/y)}{1 - c_1}
\]

where the third inequality follows since

\[
1 \leq \frac{1 + \log\{1 + (1/y)/\varphi^{-1}(1/y)\}}{1 + \log\{1 + (1/y)/\varphi^{-1}(v/y)\}} \leq 1 + \log\frac{1 + (1/y)/\varphi^{-1}(1/y)}{1 + (1/y)/\varphi^{-1}(v/y)} \leq 1 + \log\frac{\varphi^{-1}(v/y)}{\varphi^{-1}(1/y)}
\]

by the assumption that $v > 1$. Furthermore,

\[
\left\{ y\varphi^{-1}(v/y) + \delta + 1 \right\} \frac{\sqrt{\delta + 1}/e^2}{\Gamma(\delta + 1)} e^{-y\varphi^{-1}(v/y)} \frac{1}{v^{\delta + 1}} \leq \frac{M_6}{v^{\delta + 1}}
\]

for some $M_6 > 0$. Thus,

\[
v^j + \alpha e^{-v} f_{j+\alpha}(v; y) \leq v^j + \alpha e^{-v} M_5 (1 + c_2 \log v) \frac{1}{v^{\delta + 1}} \frac{1}{v^{\delta + 1}} \frac{M_6}{v^{\delta + 1}} \leq M_7 (1 + c_2 \log v)^{\delta + 1} \varphi^{-2}(j+\alpha+1) e^{-v}
\]

for some $M_7 > 0$. Finally, suppose that $0 < v < 1$. Then $\varphi'(\varphi^{-1}(1/y))/\varphi'(\varphi^{-1}(v/y)) \leq 1$ by part (ii) of Lemma $S1$. Also, $y\varphi^{-1}(v/y) \leq y\varphi^{-1}(1/y) \leq 1$ by part (iv) of Lemma
Therefore, by Lemma S2, uniformly in \(y\) and hence \(v\)

$$
\begin{align*}
\frac{f_{j,\varepsilon}(v; y)}{g_j(v; y)} & \leq M_S \frac{\{y\varphi^{-1}(v/y)\}^{\alpha - 1}}{(1 + \log(1 + 1/(y\varphi^{-1}(v/y))))^{1 + \gamma}} \\
& \times \frac{\{y\varphi^{-1}(v/y) + \delta + 1\}^{d + 1/2}}{\{1 + \varphi^{-1}(v/y)\}^{d + 1}} \\
& \times \frac{\{\varphi^{-1}(v/y)\}^j e^{-y\varphi^{-1}(v/y)} 1/v^{j + \alpha} [1 + \log(1 + 1/y)/\varphi^{-1}(1/y))]^{1 + \gamma}}{\{\varphi^{-1}(1/y)\}^{j + \alpha}} \\
& = M_S \frac{[1 + \log(1 + 1/y)/\varphi^{-1}(1/y))]^{1 + \gamma}}{(1 + \log(1 + 1/(y\varphi^{-1}(v/y))))^{1 + \gamma}} \\
& \times \frac{\{y\varphi^{-1}(v/y) + \delta + 1\}^{d + 1/2}}{\{1 + \varphi^{-1}(v/y)\}^{d + 1}} \\
& \times \frac{1}{\{\delta + y\varphi^{-1}(v/y)\}^j} e^{-y\varphi^{-1}(v/y)} \frac{1}{v^{j + \alpha}} \frac{\{\varphi^{-1}(v/y)\}^{j + \alpha}}{\{\varphi^{-1}(1/y)\}^{j + \alpha}} \\
& \leq M_S(d + 2)^{d + 1/2} \frac{1}{\delta^j v^{j + \alpha}} \\
\end{align*}
$$

and hence \(v^{j + \alpha} e^{-v} f_{j,\varepsilon}(v; y)/g_j(v; y) \leq M_S(d + 2)^{d + 1/2} e^{-v}/\delta^j\). This completes the proof.

\(\square\)

**S4 Proof of Theorem 1**

We now prove Theorem 1.

**Proof of Theorem 1.** Note that

$$
p(u_i | y_i) \propto \pi(u_i) \frac{\Gamma(u_i + \delta_i + 1)}{\Gamma(u_i)(\delta_i y_i/\eta_i + u_i)^{d_i + 1}} \left(\frac{u_i}{\delta_i y_i/\eta_i + u_i}\right)^{u_i}.
$$

Then

$$
E[\kappa_1 | y_i] = \frac{I_{1,0}(\delta_i y_i/\eta_i; \delta_i)}{I_{0,0}(\delta_i y_i/\eta_i; \delta_i)} = \frac{I_{1,c}(\delta_i y_i/\eta_i; \delta_i)}{I_{0,c}(\delta_i y_i/\eta_i; \delta_i)}.
$$

Therefore, by Lemma S2

$$
E[\kappa_1 | y_i] \sim \frac{I_{1,c}(\delta_i y_i/\eta_i; \delta_i)}{I_{0,c}(\delta_i y_i/\eta_i; \delta_i)}
$$

uniformly in \(y_i\) as \(\varepsilon \to 0\). Furthermore, by Lemmas S3 and S4 it follows from the
dominated convergence theorem that for each $\varepsilon > 0$,
\[
\frac{I_{1,\varepsilon}(\delta_i y_i/\eta_i; \delta_i)}{I_{0,\varepsilon}(\delta_i y_i/\eta_i; \delta_i)} = \frac{\int_0^\infty f_{1,\varepsilon}(v; \delta_i y_i/\eta_i)dv}{\int_0^\infty f_{0,\varepsilon}(v; \delta_i y_i/\eta_i)dv} \delta_i,
\]
\[
= \frac{\delta_i y_i/\eta_i}{\delta_i} \varphi^{-1}(1/(\delta_i y_i/\eta_i)) \frac{\int_0^\infty v^{1+\alpha}e^{-v}f_{1,\varepsilon}(v; \delta_i y_i/\eta_i)dv}{\int_0^\infty v^\alpha e^{-v}f_{0,\varepsilon}(v; \delta_i y_i/\eta_i)dv} \sim \frac{\delta_i y_i/\eta_i}{\delta_i} \varphi^{-1}(1/(\delta_i y_i/\eta_i))(1 + \alpha)
\]
as $y_i \to \infty$. Thus,
\[
E[\kappa_i|y_i] \sim \frac{\delta_i y_i/\eta_i}{\delta_i} \varphi^{-1}(1/(\delta_i y_i/\eta_i))(1 + \alpha) = o(1)
\]
as $y_i \to \infty$, where the equality follows from part (iv) of Lemma S1.

**S5 Proof of Theorem 2**

Here, we prove Theorem 2. Let $\tilde{\lambda} = 1/\lambda$ and $\tilde{\lambda}_0 = 1/\lambda_0$.

**Proof of Theorem 2** Let $\rho(x) = x - 1 - \log x$ for $x \in (0, \infty)$. Then $A_\varepsilon(\lambda_0) = \{\lambda \in (0, \infty) | \rho(\lambda_0/\lambda) < \varepsilon/\delta\} = \{1/\tilde{\lambda} \in A_\varepsilon(\tilde{\lambda}_0)\}$, where $A_\varepsilon(\tilde{\lambda}_0) = \{\tilde{\lambda} \in (0, \infty) | \rho(\tilde{\lambda}/\tilde{\lambda}_0) < \varepsilon/\delta\}$. Since $\rho'(x) \geq 0$ if and only if $x \geq 1$ for any $x \in (0, \infty)$, there exist $\epsilon^U_{\varepsilon/\delta} \in (0, 1)$ and $c_{\varepsilon/\delta}^U \in (0, \infty)$ such that $\rho(1-c_{\varepsilon/\delta}^U) = \rho(1+c_{\varepsilon/\delta}^U) = \varepsilon/\delta$ and $A_\varepsilon(\tilde{\lambda}_0) = (\tilde{\lambda}_0 - c_{\varepsilon/\delta}^U, \tilde{\lambda}_0 + c_{\varepsilon/\delta}^U\tilde{\lambda}_0)$. Since $c_{\varepsilon/\delta}^U = \varepsilon/\delta + \log(1+c_{\varepsilon/\delta}^U) > \varepsilon/\delta$, we have $\int_{A_\varepsilon(\tilde{\lambda}_0)} d\tilde{\lambda} > \varepsilon\tilde{\lambda}_0/\delta$.

Now, suppose first that $\lambda_0 \neq 1$. Then

\[
P(\lambda \in A_\varepsilon(\lambda_0)) = P(\tilde{\lambda} \in A_\varepsilon(\tilde{\lambda}_0))
\]
\[
= \int_{A_\varepsilon(\tilde{\lambda}_0)} \left\{ \int_0^\infty \pi(u) \frac{u^u}{\Gamma(u)} \tilde{\lambda}^u e^{-\tilde{\lambda} u} du \right\} d\tilde{\lambda}
\]
\[
\geq \int_{A_\varepsilon(\tilde{\lambda}_0)} \left\{ \int_0^1 \pi(u) \frac{u^u}{\Gamma(u)} (1 - c_{\varepsilon/\delta}^U) \tilde{\lambda}_0^u e^{-(\tilde{\lambda}_0 + c_{\varepsilon/\delta}^U\tilde{\lambda}_0) u} du \right\} d\tilde{\lambda}
\]
\[
\geq \frac{\varepsilon\lambda_0}{\delta} (1 - c_{\varepsilon/\delta}^U)e^{-(\tilde{\lambda}_0 + c_{\varepsilon/\delta}^U\tilde{\lambda}_0)} \int_0^1 \pi(u) \frac{u^u}{\Gamma(u)} \tilde{\lambda}_0^u du < \infty.
\]
Therefore, by (5),

\[ R_n \leq \varepsilon + \frac{1}{n} \left[ \log \frac{\delta}{\varepsilon} + \log \left\{ \frac{e^{\lambda_0 + \epsilon/\delta} \lambda_0}{\lambda_0 (1 - cL \epsilon/\delta)} \right\} \right. \int_0^1 \pi(u) \frac{u^\nu}{\Gamma(u)} \lambda_0^\nu \, du \right] \]

which is \( O(n^{-1} \log n) \) when \( \varepsilon = 1/n \).

Next, suppose that \( \lambda_0 = 1 \) and that \( \pi(u) \sim u^{-1-b} \) as \( u \to \infty \) for some \( 0 < b \leq 1/2 \).

Then we have

\[
P(\lambda \in A_\varepsilon(\lambda_0)) = P(\tilde{\lambda} \in \tilde{A}_\varepsilon(\tilde{\lambda}_0)) = \int_{\tilde{A}_\varepsilon(\tilde{\lambda}_0)} \left\{ \int_0^\infty \pi(u) \frac{u^u}{\Gamma(u)} \lambda^u e^{-\lambda u} \, du \right\} \, d\tilde{\lambda}
= \int_{\tilde{A}_\varepsilon(\tilde{\lambda}_0)} \left\{ \int_0^\infty \pi(u) \frac{u^u e^{-u}}{\Gamma(u)} e^{-\rho(\tilde{\lambda}) u} \, du \right\} \, d\tilde{\lambda}
\geq \int_{\tilde{A}_\varepsilon(\tilde{\lambda}_0)} \left\{ \int_0^\infty \pi(u) \frac{u^u e^{-u}}{\Gamma(u)} e^{-\epsilon u/\delta} \, du \right\} \, d\tilde{\lambda}
\geq \frac{\varepsilon}{\delta} \int_0^\infty \pi(u) \frac{u^u e^{-u}}{\Gamma(u)} e^{-\epsilon u/\delta} \, du.
\]

Since the inequality

\[
\frac{u^u e^{-u}}{\Gamma(u)} \geq \frac{1}{(2\pi)^{1/2}} u^{1/2} e^{-1/(12u)}
\]

holds for all \( u \in (0, \infty) \),

\[
\int_0^\infty \pi(u) \frac{u^u e^{-u}}{\Gamma(u)} e^{-\epsilon u/\delta} \, du \geq \int_0^\infty \pi(u) \frac{1}{(2\pi)^{1/2}} u^{1/2} e^{-1/(12u)} e^{-\epsilon u/\delta} \, du \\
\geq \frac{e^{-1/12}}{(2\pi)^{1/2}} \int_1^\infty \pi(u) u^{1/2} e^{-\epsilon u/\delta} \, du.
\]

Since \( \pi(u) \geq C' u^{-1-b} \) for all \( u \geq 1 \) for some \( C' > 0 \) by assumption,

\[
\frac{1}{C'} \int_1^\infty \pi(u) u^{1/2} e^{-\epsilon u/\delta} \, du \geq \int_1^\infty u^{-1/2-b} e^{-\epsilon u/\delta} \, du \geq \int_1^\infty u^{-1-b} e^{-\epsilon u/\delta} \, du \\
= \left[ (\log u) e^{-\epsilon u/\delta} \right]_1^\infty + \frac{\varepsilon}{\delta} \int_1^\infty (\log u) e^{-\epsilon u/\delta} \, du = \int_\varepsilon^{\infty} \left( \log \frac{u}{\varepsilon/\delta} \right) e^{-u} \, du \\
\geq \int_1^\infty \left( \log \frac{u}{\varepsilon/\delta} \right) e^{-u} \, du \left( \log \frac{1}{\varepsilon/\delta} \right) \int_1^\infty \log \{u/(\varepsilon/\delta)\} e^{-u} \, du
\]

when \( \varepsilon < \delta \). Note that \( \log \{u/(\varepsilon/\delta)\}/\log \{1/(\varepsilon/\delta)\} \to 1 \) as \( \varepsilon \to 0 \) and that if \( \varepsilon < \delta/2 \),
then

\[
\log \left\{ \frac{u}{(\varepsilon/\delta)} \right\} = \exp \left( \log \log \left[ \frac{s}{\varepsilon/\delta} \right]_{s=1}^{s=u} \right) = \exp \left( \int_1^u \frac{1}{\log \left\{ \frac{s}{(\varepsilon/\delta)} \right\}} \frac{1}{s} ds \right) \leq \exp \left( \int_1^u \frac{1}{\log 2} ds \right) = u^{1/\log 2}
\]

for all \( u > 1 \). Then, by the dominated convergence theorem,

\[
\int_1^\infty \frac{\log \left\{ \frac{u}{(\varepsilon/\delta)} \right\}}{\log \left\{ 1/(\varepsilon/\delta) \right\}} e^{-u} du \rightarrow \int_1^\infty e^{-u} du = e^{-1}
\]

as \( \varepsilon \to 0 \). Thus,

\[
P(\lambda \in A(\lambda_0)) \geq \frac{\varepsilon e^{-1/12}}{\delta (2\pi)^{1/2} C^\delta} \left( \log \frac{\delta}{\varepsilon} \right) \int_1^\infty \frac{\log \left\{ \frac{u}{(\varepsilon/\delta)} \right\}}{\log \left\{ 1/(\varepsilon/\delta) \right\}} e^{-u} du
\]

and it follows from (5) that

\[
R_n \leq \varepsilon + \frac{1}{n} \left( \log \frac{\delta}{\varepsilon} - \log \log \frac{\delta}{\varepsilon} + \log \left[ \frac{e^{1/12}(2\pi)^{1/2}}{C^\delta} \right] / \int_1^\infty \frac{\log \left\{ \frac{u}{(\varepsilon/\delta)} \right\}}{\log \left\{ 1/(\varepsilon/\delta) \right\}} e^{-u} du \right),
\]

which is \( O(n^{-1}(\log n - \log \log n)) \) when \( \varepsilon = 1/n \). This completes the proof. \( \square \)

S6 Posterior Sampling Under the IRB Prior

By using the integral expression for the IRB density given in the following proposition, we can construct an MCMC algorithm in which \( u_1, \ldots, u_n \) are easily updated.

**Proposition S1.** The IRB prior can be expressed as

\[
B(b, a)_{\text{IRB}}(u_i | b, a) = \int_{(0, \infty)^3} \frac{s_i^{-b} w_i^{b+a-1}}{\Gamma(1 - b) \Gamma(b + a)} e^{-w_i} \frac{z_i^{s_i+w_i}}{\Gamma(s_i + w_i + 1)} e^{-z_i} u_i^{s_i+w_i-1} e^{-z_i u_i} d(s_i, w_i, z_i)
\]

\[
= \int_{(0, \infty)^3} \left[ \text{Ga}(s_i | 1 - b, \log(1 + 1/u_i)) \text{Ga}(w_i | b + a, 1 + \log(1 + 1/u_i)) \text{Ga}(z_i | s_i + w_i + 1, 1 + u_i) \right] \\
\times \frac{1}{u_i(1 + u_i) \{\log(1 + 1/u_i)\}^{1-b} \{1 + \log(1 + 1/u_i)\}^{b+a}} d(s_i, w_i, z_i).
\]
Proof. We have

\[ B(b, a) \text{IRB}(u_i|b, a) \]
\[ = \frac{1}{u_i(1 + u_i)} \int_0^\infty s_i^{-b} e^{-s_i \log(1+1/u_i)} ds_i \int_0^\infty w_i^{b+a-1} e^{-w_i \log(1+1/u_i)} dw_i \]
\[ = \int (0, \infty)^2 \frac{s_i^{-b} w_i^{b+a-1}}{\Gamma(1-b) \Gamma(b+a)} e^{-w_i} u_i(1 + u_i) (1 + 1/u_i)^s_i w_i ds_i, w_i \]
\[ = \int (0, \infty)^3 \frac{s_i^{-b} w_i^{b+a-1}}{\Gamma(1-b) \Gamma(b+a)} e^{-w_i} \frac{1}{\Gamma(s_i + w_i + 1)} u_i^{s_i + w_i - 1} d(s_i, w_i, z_i) \]
\[ = \int (0, \infty)^3 \frac{s_i^{-b} w_i^{b+a-1}}{\Gamma(1-b) \Gamma(b+a)} e^{-w_i} \frac{1}{\Gamma(s_i + w_i + 1)} u_i^{s_i + w_i - 1} d(s_i, w_i, z_i) \]
\[ = \int (0, \infty)^3 [\text{Ga}(s_i|1 - b, \log(1 + 1/u_i)) \text{Ga}(w_i|b + a, 1 + \log(1 + 1/u_i)) \text{Ga}(z_i|s_i + w_i + 1, 1 + u_i)] \]
\[ \times \frac{1}{u_i(1 + u_i)} \{\log(1+1/u_i)\}^{1-b} \{1+\log(1+1/u_i)\}^{b+a} d(s_i, w_i, z_i) \]
and this proves the proposition.  

We make the change of variables \( \nu_i = \tau u_i \) for \( i = 1, \ldots, n \). We consider \( (s, w, z) \in (0, \infty)^3 n \) as a set of additional latent variables. The overall posterior distribution of \((\lambda, \beta, \tau, s, w, z, \nu)\) given \( y \) is

\[
p(\lambda, \beta, \tau, s, w, z, \nu|y) \propto \pi_\beta(\beta) \pi_\tau(\tau) \frac{1}{\tau^n} \prod_{i=1}^n \left\{ \frac{\beta^\nu_i+1 \nu_i^\nu_i}{\Gamma(\nu_i)} \frac{1}{\lambda_i^\nu_i+2} e^{-\beta_i \nu_i / \lambda_i} \frac{1}{\lambda_i^\delta_i} \exp\left(-\frac{\delta_i y_i}{\lambda_i \eta_i}\right) \right\} \]
\[
\times s_i^{-b} w_i^{b+a-1} e^{-w_i} \frac{1}{\Gamma(s_i + w_i + 1)} u_i^{s_i + w_i - 1} e^{-z_i s_i + w_i - 1} e^{-z_i u_i / \tau} \]
\[
= \pi_\beta(\beta) \pi_\tau(\tau) \frac{1}{\tau^n} \prod_{i=1}^n \left\{ \frac{\beta^\nu_i+1 \nu_i^\nu_i}{\Gamma(\nu_i)} \frac{1}{\lambda_i^\nu_i+2} e^{-\beta_i \nu_i / \lambda_i} \frac{1}{\lambda_i^\delta_i} \exp\left(-\frac{\delta_i y_i}{\lambda_i \eta_i}\right) \right\} \]
\[
\times \frac{1}{\nu_i / \tau} \frac{1}{(1 + \nu_i / \tau)} \{\log(1+\nu_i / \tau)\}^{1-b} \{1+\log(1+\nu_i / \tau)\}^{b+a} \] .

The variables \( \lambda, \beta, \tau, (s, w, z), \) and \( \nu \) are updated in the following way.

- Sample \( \lambda_i \sim \text{IG}(\delta_i + \nu_i + 1, \delta_i y_i / \eta_i + \beta \nu_i) \) independently for \( i = 1, \ldots, n. \)
- Sample $\beta \sim \text{Ga}(\sum_{i=1}^{n} \nu_i + n + a_\beta, \sum_{i=1}^{n} \nu_i / \lambda_i + b_\beta)$.

- Sample $\tau \sim \text{GIG}(-\sum_{i=1}^{n} (s_i + w_i) + a_\tau, 2b_\tau, 2\sum_{i=1}^{n} z_i \nu_i)$, which has density proportional to $\tau^{-\sum_{i=1}^{n} (s_i + w_i) + a_\tau - 1} e^{-b_\tau \tau - \sum_{i=1}^{n} z_i \nu_i / \tau}$.

- Independently for $i = 1, \ldots, n$,

  1. sample $s_i \sim \text{Ga}(1 - b, \log(1 + \tau / \nu_i))$ and $w_i \sim \text{Ga}(b + a, 1 + \log(1 + \tau / \nu_i))$ independently and

  2. sample $z_i \sim \text{Ga}(s_i + w_i + 1, 1 + \nu_i / \tau)$.

- The full conditional distribution of $\nu$ is proportional to

  \[
  \prod_{i=1}^{n} \{ \text{Ga}(\nu_i | s_i + w_i, z_i / \tau) \text{Ga}(1/\lambda_i | \nu_i, \beta \nu_i) \},
  \]

  which can be accurately approximated by using the method of Miller (2019) for each $i = 1, \ldots, n$. We use the approximate full conditional distributions as proposal distributions in independent MH steps.