Schur properties of convolutions of gamma random variables

Farbod Roosta-Khorasani*   Gábor J. Székely†

January 20, 2016

Abstract

Sufficient conditions for comparing the convolutions of heterogeneous gamma random variables in terms of the usual stochastic order are established. Such comparisons are characterized by the Schur convexity properties of the cumulative distribution function of the convolutions. Some examples of the practical applications of our results are given.

1 Introduction and Main Result

Linear combinations (i.e., convolutions) of independent gamma random variables (r.v's) often naturally arise in many applications in statistics, engineering, insurance, actuarial science and reliability. As such there has been extensive study of their stochastic properties in the literature. For examples of such theoretical studies as well as applications see [16, 17, 18, 12, 32, 19, 8, 34, 33, 21, 1, 7, 29, 25] and references therein.

Bock et al., [6], and Diaconis and Perlman, [9], in their seminal works, first studied the Schur convexity properties of the cumulative distribution function of the linear combinations of independent gamma r.v's. Ever since, this topic and its variants have been studied by many researchers; see the references mentioned above. However

---
*Dept. of Computer Science, University of British Columbia, Vancouver, Canada. farbod@cs.ubc.ca.

†National Science Foundation, Arlington, Virginia. gszekely@nsf.gov and Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences, Budapest, Hungary.

despite all the efforts, the results in [6] remained the best available, yet far from the best possible. Bakirov, [5] provided a tighter bound for the special case of convolutions of chi-squared r.v's of degree one, which often arise from quadratic forms. Here, we prove results regarding the Schur properties of the tails of convolutions of gamma r.v's with respect to the mixing weights and in terms of the usual stochastic order, and indeed sharpen some results given in [6]. As a consequence, the result in [5] is also generalized.

More specifically, let X_i, $i = 1, 2, \ldots, n$, be n independent and identically distributed (i.i.d) gamma distributed r.v's, parametrized by shape $\alpha > 0$ and rate $\beta > 0$ parameters with the probability density function (PDF)

$$f(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & x \geq 0 \\ 0 & x \leq 0 \end{cases}.$$

Consider the following non-negative linear combinations of such r.v's

$$\sum_{i=1}^{n} \lambda_i X_i,$$

where $\lambda_i \geq 0$, $i = 1, 2, \ldots, n$, are real numbers. For a given $\alpha > 0$, $\beta > 0$ and $x > 0$, define

$$P(\boldsymbol{\lambda}; \alpha, \beta, x) := \Pr\left(\sum_{i=1}^{n} \lambda_i X_i < x\right), \tag{1}$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{R}^n$. The aim of the present paper is to find the conditions allowing one to compare tail probabilities of the form (1) with respect to the mixing weights $\boldsymbol{\lambda}$. In [25, Theorem 2.2], results regarding the extremal values of (1) (i.e., maximal and minimal values with respect to $\boldsymbol{\lambda}$ and for given α, β and x) are proved. Here, we extend those results to be able to compare (1) for any pair of weight vectors.

To that end, let us recall that the vector $\boldsymbol{\lambda}$ is said to *majorize* the vector $\boldsymbol{\mu}$, denoted by $\boldsymbol{\mu} \prec \boldsymbol{\lambda}$, if

$$0 \leq \lambda_n \leq \ldots \leq \lambda_2 \leq \lambda_1, \tag{2a}$$

$$0 < \mu_n \leq \ldots \leq \mu_2 \leq \mu_1, \tag{2b}$$

$$\sum_{i=1}^{k} \mu_i \leq \sum_{i=1}^{k} \lambda_i, \quad \forall k < n, \tag{2c}$$

$$\sum_{i=1}^{n} \mu_i = \sum_{i=1}^{n} \lambda_i. \tag{2d}$$

Note that there is "strict positivity" assumption on $\boldsymbol{\mu}$, but not on $\boldsymbol{\lambda}$. Of course, it is clear that padding $\boldsymbol{\mu}$ (and as a result $\boldsymbol{\lambda}$) with 0's would not change the majorization order and simply add redundant components to both vectors. Given $\boldsymbol{\mu} \prec \boldsymbol{\lambda}$,

$P(.; \alpha, \beta, x)$ is said to be Schur convex if $P(\boldsymbol{\mu}; \alpha, \beta, x) \le P(\boldsymbol{\lambda}; \alpha, \beta, x)$, and it is said to be Schur concave if $P(\boldsymbol{\mu}; \alpha, \beta, x) \ge P(\boldsymbol{\lambda}; \alpha, \beta, x)$. For comprehensive details on the theory of majorization and its applications, refer to the classic book of Marshall and Olkin [20].

It is not hard to show that the variance of $\sum_{i=1}^n \lambda_i X_i$ is a Schur-convex function of $\boldsymbol{\lambda}$ and, indeed, it would be useful to know when $P(.; \alpha, \beta, x)$ exhibit similar properties. For symmetric distributions a fairly general result is known. If $X_1, X_2 \ldots, X_n$ are independent random variables with a common symmetric and log-concave PDF, in [23], it was shown that $\Pr\left(\sum_{i=1}^n \lambda_i X_i < x\right)$ is Schur-convex in $\boldsymbol{\lambda}$ for any $x > 0$. However, for positive random variables with non-symmetric distributions (such as gamma r.v's), to the best of our knowledge, no such general results, except for those in [6] and [5], exist.

For the case of $n \ge 3$, Bock et al., [6, Theorem 3], showed that if $\boldsymbol{\mu} \prec \boldsymbol{\lambda}$ and $\lambda_i > 0 \; \forall i$, then

$$
\begin{aligned}
P(\boldsymbol{\mu}; \alpha, \beta, x) &\ge P(\boldsymbol{\lambda}; \alpha, \beta, x), \quad \forall x > \frac{(n\alpha + 1) \max_i \lambda_i}{\beta}, \\
P(\boldsymbol{\mu}; \alpha, \beta, x) &\le P(\boldsymbol{\lambda}; \alpha, \beta, x), \quad \forall x < \frac{(n\alpha + 1) \min_i \lambda_i}{\beta}.
\end{aligned}
$$

For the special case of chi-squared r.v's of degree one (i.e., $\alpha = \beta = 1/2$), Bakirov, [5], provided the following tighter bound (for $\forall n \ge 1$) for the Schur concavity of $P(.; 1/2, 1/2, x)$:

$$
P(\boldsymbol{\mu}; \frac{1}{2}, \frac{1}{2}, x) \ge P(\boldsymbol{\lambda}; \frac{1}{2}, \frac{1}{2}, x), \quad \forall x > 2s,
$$

where $s = \sum_{i=1}^n \lambda_i = \sum_{i=1}^n \mu_i$. No results concerning the Schur convexity of $P(.; 1/2, 1/2, x)$ was given in [5].

Our main result is stated as follows (the details of its proof are given in the Appendix). The discussions regarding the relative improvements compared to [6] and [5], as well as further extensions are deferred to Section 3.

Theorem 1 *Let $X_i \sim Gamma(\alpha, \beta)$, $i = 1, 2, \ldots, n$, be n i.i.d gamma r.v's, where $\alpha > 0$ and $\beta > 0$. If $\boldsymbol{\mu} \prec \boldsymbol{\lambda}$, then*

$$
\begin{aligned}
P(\boldsymbol{\mu}; \alpha, \beta, x) &\ge P(\boldsymbol{\lambda}; \alpha, \beta, x), \quad \forall x > \frac{(2\alpha + 1)s}{2\beta}, \\
P(\boldsymbol{\mu}; \alpha, \beta, x) &\le P(\boldsymbol{\lambda}; \alpha, \beta, x), \quad \forall x <
\begin{cases}
\dfrac{\alpha s}{\beta}, & n = 2 \\
\dfrac{(\alpha - 1)s}{\beta}, & n \ge 3, \; \alpha > 1
\end{cases},
\end{aligned}
$$

where $s = \sum_{i=1}^n \lambda_i = \sum_{i=1}^n \mu_i$.

This paper is organized as follows. In Section 2, we give some examples of the practical applications of our results. In Section 3, we discuss the relative improvements of our results compared to those in [6] and [5]. We also extend Theorem 1 by weakening the majorization requirement. In addition, we give similar results for the case where $n = \infty$. The proofs of our results are given in the Appendix.

2   Examples

In this section, we give examples to demonstrate some practical applications of our results.

2.1   Experimental Design in Signal Detection

Consider the additive model of observations

$$D(t) = \tau s(t) + \eta(t), \quad t \in [0, T]$$

where $D(t)$ is the measured data, $s(t)$ is the signal of interest and $\eta(t)$ is the additive noise. Suppose we have N measurements, taken at discrete time intervals, $0 = t_1 \leq t_1 \leq \ldots \leq t_N = T$. In addition, suppose that $\{\eta(t_i) : i = 1, 2, \ldots, N\}$ is a collection of i.i.d Laplace r.v's with mean zero and variance σ_i^2. In addition, let $\tau = 1$ if there is a signal, and $\tau = 0$ otherwise (i.e., the measured data is in fact entirely the noise). To detect if the signal is present, we can use the signal to noise ratio (SNR)

$$Q(N) := \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sigma_i} |D(t_i)|.$$

Note that $\sigma_i^{-1} D(t)$ is a Laplace r.v, with mean $\tau s(t)$ and variance 1.

A design question to answer is that of , at least how many measurements, a priori, is needed to make sure the probability of Type I error (i.e., when we conclude that the signal is present, when in fact it is missing) is below a desired tolerance $0 < \delta \ll 1$.

If $\tau = 0$, then $\sigma_i^{-1} |D(t)| \sim Gamma(1, \sqrt{2})$, so by Theorem 1 we get

$$P(Q(1) \geq x) \geq P(Q(2) \geq x) \geq \ldots \geq P(Q(N) \geq x) \geq \ldots, \quad \forall x > \frac{3}{2\sqrt{2}}.$$

For a given N, we can easily compute $P(Q(N) \geq x)$. Hence, in order to find the minimum N required, we can increase N until $P(Q(N) \geq x) \leq \delta$.

2.2   Matrix Trace Estimation

The need to estimate the trace of an implicit symmetric positive semi-definite (SPSD) matrix is of fundamental importance (see [28]) and arises in many applications; see

for instance [15, 4, 3, 14, 10, 31, 27, 26, 30, 13, 2] and references therein. The standard approach for estimating the trace of such a matrix A, denoted here by $tr(A)$, is based on a Monte-Carlo method, where one generates N random vector realizations $\mathbf{w}_i$ from a suitable probability distribution D and computes

$$tr_D^N(A) := \frac{1}{N} \sum_{i=1}^{N} \mathbf{w}_i^t A \mathbf{w}_i.$$

One such suitable probability distribution is the standard normal, $\mathcal{N}(0, \mathbb{I})$. This estimator is known as the *Gaussian estimator*, denoted here by $tr_G^N(A)$.

Now, given a pair of small positive real numbers (ε, δ), consider finding an appropriate sample size N such that

$$\Pr\left(tr_G^N(A) \geq (1 - \varepsilon)tr(A)\right) \geq 1 - \delta, \tag{3a}$$

$$\Pr\left(tr_G^N(A) \leq (1 + \varepsilon)tr(A)\right) \geq 1 - \delta. \tag{3b}$$

Such question was first studied in [3] and further improved in [24]. In particular, in [24] it was proved that the inequalities (3) hold if

$$N > \frac{\|A\|}{tr(A)} \frac{8}{\varepsilon^2} \ln(\frac{1}{\delta}), \tag{4}$$

where $\|A\|$ denotes the L_2 norm of the matrix A. The ratio $tr(A)/\|A\|$ is known as the *effective rank* of the matrix (see [11]) and it is a stable quantity compared with the usual rank. The appearance of effective rank in the bound (4) is an indication of a possible relation between the "*skewness*" of the eigenvalues of A and the efficiency of the Gaussian estimator. In other words, the more skewed the eigenvalue distribution is, the worse we expect the Gaussian estimator to perform (i.e., the larger the true sample size required would be). However effective rank is not a consistent measure for skewness and, as such, in [24], this relationship was demonstrated only numerically and no consistent definition for how the relative skewness could be measured was given. Now using the majorization order among eigenvalue vectors as a consistent measure of skewness, the new theoretical results in the present paper fully describe the observations from the numerical examples in [24]. As in the proof of [24, Theorem 1], we see that

$$Pr\left(tr_H^N(A) \leq (1 - \varepsilon)tr(A)\right) = P\left(\boldsymbol{\lambda}; \frac{N}{2}, \frac{N}{2}, (1 - \varepsilon)tr(A)\right), \tag{5}$$

where $\boldsymbol{\lambda}$ the vector of eigenvalues of A sorted in the decreasing order. Consider two SPSD matrices, A_1 and A_2, such that $tr(A_1) = tr(A_2)$ and whose respective eigenvalue vectors, $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$, are sorted in the decreasing order. If $\boldsymbol{\lambda}_2 \prec \boldsymbol{\lambda}_1$, then we say that

the eigenvalue distribution of A_1 is more skewed that that of A_2. If $N > 2/\varepsilon$, then from Theorem 1 we obtain

$$P\left(\boldsymbol{\lambda}_2; \frac{N}{2}, \frac{N}{2}, (1-\varepsilon)tr(A_2)\right) \leq P\left(\boldsymbol{\lambda}_1; \frac{N}{2}, \frac{N}{2}, (1-\varepsilon)tr(A_1)\right).$$

In other words, using the same sample size N, our estimate with A_1 is more likely to be located further away to the left of the true value than that with A_2. Hence in order to make the former estimate better, we need to increase the sample size which, in turn, in some algorithms translates into more computational costs; see [14, 10, 30, 27, 26, 25]. Similar comparisons can be made for $Pr\left(tr_H^N(A) \leq (1+\varepsilon)tr(A)\right)$.

3   Discussions and Further Extensions

The comparison between the results in Theorem 1 and those in [6] and [5] can be summarized as follows:

- For the special case of chi-squared distribution, i.e., $\alpha = \beta = 1/2$, our Schur concavity result is the same as that in [5].

- For $n = 2$, the results of Theorem 1 coincide with [6, Theorem 1]. However, for $n \geq 3$, our results are more uniform than [6, Theorem 3]. Namely, the sufficient conditions given in Theorem 1 are not dependent on the dimensions of the vectors. Additionally, our bounds for x are independent of the particular values of the vector components, such as "$\min_i \lambda_i$" or "$\max_i \lambda_i$". In other words, for all vectors whose sums are equal, we give a fixed bound for x.

- For $n \geq 3$, our Schur convexity result is sharper than that of [6, Theorem 3] when

$$\max_i \lambda_i > \frac{\sum_i \lambda_i}{n} \frac{\alpha + \frac{1}{2}}{\alpha + 1/n}.$$

  Similarly, for $n \geq 3$ and $\alpha > 1$, our Schur concavity result improves that of [6, Theorem 3] when

$$\min_i \lambda_i < \frac{\sum_i \lambda_i}{n} \frac{\alpha - 1}{\alpha + 1/n}.$$

- Most importantly, in [6, Theorem 3], it is required that $\lambda_i > 0$, $i = 1, 2, \ldots, n$. This is a rather strong condition as many interesting comparisons cannot be performed this way. Our results do not make any strict positivity assumption on the components of $\boldsymbol{\lambda}$ and it suffices if they are simply non-negative. As a simple example, consider $\alpha = 2$, $\beta = 1$ and $x < 1$. Then using Theorem 1, we get $P(\boldsymbol{\lambda}_1; 2, 1, x) \geq P(\boldsymbol{\lambda}_2; 2, 1, x) \geq P(\boldsymbol{\lambda}_3; 2, 1, x) \geq P(\boldsymbol{\lambda}_4; 2, 1, x)$ with $\boldsymbol{\lambda}_1 = (1, 0, 0, 0)$, $\boldsymbol{\lambda}_2 = (2/3, 1/3, 0, 0)$, $\boldsymbol{\lambda}_3 = (3/6, 2/6, 1/6, 0)$, and $\boldsymbol{\lambda}_4 = (4/10, 3/10, 2/10, 1/10)$. This comparison is not possible with [6, Theorem 3].

- For the Schur convexity of $P(.; \alpha, \beta, x)$ in the case of $\alpha \leq 1$ and $n \geq 3$, the result in [6, Theorem 3] remains the best available, as we were not able to improve upon it here. In fact, from our method of proof, it seems likely that obtaining a general result for this case is impossible. Indeed, Bock et al. [6, p. 394] give an example which corroborates this observation.

It is possible to weaken the majorization requirement and obtain even more general results. More specifically, relaxing the equality condition in (2) gives the following weak majorization order. Recall that the vector $\boldsymbol{\lambda}$ is said to *weakly majorize* the vector $\boldsymbol{\mu}$, denoted by $\boldsymbol{\mu} \prec_w \boldsymbol{\lambda}$, if

$$0 \leq \lambda_n \leq \ldots \leq \lambda_2 \leq \lambda_1,$$
$$0 < \mu_n \leq \ldots \leq \mu_2 \leq \mu_1,$$
$$\sum_{i=1}^{k} \mu_i \leq \sum_{i=1}^{k} \lambda_i, \quad \forall k \leq n.$$

In the case of weak majorization order, we have the following almost immediate corollary.

Corollary 2 *Let $X_i \sim Gamma(\alpha, \beta)$, $i = 1, 2, \ldots, n$, be n i.i.d gamma r.v's, where $\alpha > 0$ and $\beta > 0$. If $\boldsymbol{\mu} \prec_w \boldsymbol{\lambda}$, then*

$$P(\boldsymbol{\mu}; \alpha, \beta, x) \geq P(\boldsymbol{\lambda}; \alpha, \beta, x), \quad \forall x > \frac{(2\alpha + 1)s_{\boldsymbol{\lambda}}}{2\beta},$$

$$P(\boldsymbol{\mu}; \alpha, \beta, x) \leq P(\boldsymbol{\lambda}; \alpha, \beta, x), \quad \forall x < \begin{cases} \dfrac{\alpha s_{\boldsymbol{\mu}}}{\beta}, & n = 2 \\ \dfrac{(\alpha - 1)s_{\boldsymbol{\mu}}}{\beta}, & n \geq 3, \ \alpha > 1 \end{cases},$$

where $s_{\boldsymbol{\lambda}} = \sum_{i=1}^{n} \lambda_i$ and $s_{\boldsymbol{\mu}} = \sum_{i=1}^{n} \mu_i$.

We can also extend Theorem 1, as well as Corollary 2, for the case where $n = \infty$. For any non-negative ℓ_1 sequence $\boldsymbol{\lambda} = (\lambda_1, \lambda_1, \ldots)$, i.e., $\sum_{i=1}^{\infty} \lambda_i < \infty$, define

$$P_{\infty}(\boldsymbol{\lambda}; \alpha, \beta, x) := \Pr\left(\sum_{i=1}^{\infty} \lambda_i X_i < x\right).$$

The (weak) majorization order is naturally extended to such ℓ_1 sequences.

Corollary 3 *Let $\{X_i \sim Gamma(\alpha, \beta), \ i = 1, 2, \ldots\}$ be a countably infinite collection of i.i.d gamma r.v's, where $\alpha > 0$ and $\beta > 0$. For any two non-negative ℓ_1 sequences, $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$, such that $\boldsymbol{\mu} \prec_w \boldsymbol{\lambda}$, we have*

$$P_{\infty}(\boldsymbol{\mu}; \alpha, \beta, x) \geq P_{\infty}(\boldsymbol{\lambda}; \alpha, \beta, x), \quad \forall x > \frac{(2\alpha + 1)s_{\boldsymbol{\lambda}}}{2\beta},$$

$$P_{\infty}(\boldsymbol{\mu}; \alpha, \beta, x) \leq P_{\infty}(\boldsymbol{\lambda}; \alpha, \beta, x), \quad \forall x < \frac{(\alpha - 1)s_{\boldsymbol{\mu}}}{\beta}, \quad \alpha > 1,$$

where $s_{\boldsymbol{\lambda}} = \sum_{i=1}^{\infty} \lambda_i$ and $s_{\boldsymbol{\mu}} = \sum_{i=1}^{\infty} \mu_i$.

Finally, it might be worth noting that the results such as Theorem 1 show that if $\boldsymbol{\mu} \prec \boldsymbol{\lambda}$, then $P(\boldsymbol{\mu}; \alpha, \beta, x)$ and $P(\boldsymbol{\lambda}; \alpha, \beta, x)$ must have at least one crossing on $x \in (0, \infty)$. Diaconis and Perlman in [9] tried to answer whether this crossing point is unique. However, they only proved this uniqueness for $n = 2$ and for $n \geq 3$, they required to impose further restrictions. Ever since, this has been an open problem which is known as the Unique Crossing Conjecture (UCC) and it is quite remarkable that the UCC has remained open, although all the evidence points towards the direction of it being true.

References

[1] L. Amiri, B. Khaledi, and F. J. Samaniego. On skewness and dispersion among convolutions of independent gamma random variables. *Probability in the Engineering and Informational Sciences*, 25(01):55–69, 2011.

[2] H. Avron. Counting triangles in large graphs using randomized matrix trace estimation. *Workshop on Large-scale Data Mining: Theory and Applications*, 2010.

[3] H. Avron and S. Toledo. Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix. *JACM*, 58(2), 2011. Article 8.

[4] Z. Bai, M. Fahey, and G. Golub. Some large scale matrix computation problems. *J. Comput. Appl. Math.*, 74:71–89, 1996.

[5] N. K. Bakirov. Comparison theorems for distribution functions of quadratic forms of gaussian vectors. *Theory of Probability & Its Applications*, 40(2):340–348, 1995.

[6] M. E. Bock, P. Diaconis, F. W. Huffer, and M. D. Perlman. Inequalities for linear combinations of gamma random variables. *Canadian Journal of Statistic*, 15:387–395, 1987.

[7] P. J. Boland, E. El-Neweihi, and F. Proschan. Schur properties of convolutions of exponential and geometric random variables. *Journal of Multivariate Analysis*, 48(1):157–167, 1994.

[8] J. Bon and E. Pãltãnea. Ordering properties of convolutions of exponential random variables. *Lifetime Data Analysis*, 5(2):185–192, 1999.

[9] P. Diaconis and M. D. Perlman. Bounds for tail probabilities of weighted sums of independent gamma random variables. *Lecture Notes-Monograph Series*, pages 147–166, 1990.

[10] K. van den Doel and U. Ascher. The chaotic nature of faster gradient descent methods. *J. Scient. Comput.*, 48, 2011. DOI: 10.1007/s10915-011-9521-3.

[11] Yonina C Eldar and Gitta Kutyniok. *Compressed sensing: theory and applications.* Cambridge University Press, 2012.

[12] E. Furman and Z. Landsman. Tail variance premium with applications for elliptical portfolio of risks. *Astin Bulletin*, 36(2):433, 2006.

[13] G. H. Golub, M. Heath, and G. Wahba. Generalized cross validation as a method for choosing a good ridge parameter. *Technometrics*, 21:215–223, 1979.

[14] E. Haber, M. Chung, and F. Herrmann. An effective method for parameter estimation with PDE constraints with multiple right-hand sides. *SIAM J. Optimization*, 22:739–757, 2012.

[15] M.F. Hutchinson. A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines. *J. Comm. Stat. Simul.*, 19:433–450, 1990.

[16] S. Kochar and M. Xu. On the right spread order of convolutions of heterogeneous exponential random variables. *Journal of Multivariate Analysis*, 101(1):165–176, 2010.

[17] S. Kochar and M. Xu. The tail behavior of the convolutions of gamma random variables. *Journal of Statistical Planning and Inference*, 141(1):418–428, 2011.

[18] S. Kochar and M. Xu. Some unified results on comparing linear combinations of independent gamma random variables. *Probability in the Engineering and Informational Sciences*, 26(03):393–404, 2012.

[19] S. Lihong and Z. Xinsheng. Stochastic comparisons of order statistics from gamma distributions. *Journal of Multivariate Analysis*, 93(1):112–121, 2005.

[20] A.W. Marshall and I. Olkin. Inequalities: Theory of majorization and its applications. *Academic, New York*, 1979.

[21] Milan Merkle and Ljiljana Petrović. On schur-convexity of some distribution functions. *Publications de l'Institut Mathématique*, 56(76):111–118, 1994.

[22] J. E. Pečarić, F. Proschan, and Y.L. Tong. *Convex Functions, Partial Orderings, and Statistical Applications*, volume 187 of *Mathematics in Science and Engineering*. Academic Press, 1992.

[23] F. Proschan. Peakedness of distributions of convex combinations. *The Annals of Mathematical Statistics*, 36:1703–1706, 1965.

[24] F. Roosta-Khorasani and U. Ascher. Improved bounds on sample size for implicit matrix trace estimators. *Foundations of Computational Mathematics*, 2014. DOI: 10.1007/s10208-014-9220-1.

[25] F. Roosta-Khorasani, G. Székely, and U. Ascher. Assessing stochastic algorithms for large scale nonlinear least squares problems using extremal probabilities of linear combinations of gamma random variables. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):61–90, 2015. DOI: 10.1137/14096311X.

[26] F. Roosta-Khorasani, K. van den Doel, and U. Ascher. Data completion and stochastic algorithms for PDE inversion problems with many measurements. *Electronic Transactions on Numerical Analysis*, 42:177–196, 2014.

[27] F. Roosta-Khorasani, K. van den Doel, and U. Ascher. Stochastic algorithms for inverse problems involving PDEs and many measurements. *SIAM J. Scientific Computing*, 36(5):S3–S22, 2014.

[28] A. Shapiro, D. Dentcheva, and D. Ruszczynski. *Lectures on Stochastic Programming: Modeling and Theory*. Piladelphia: SIAM, 2009.

[29] G. J. Székely and N. K. Bakirov. Extremal probabilities for gaussian quadratic forms. *Probab. Theory Related Fields*, 126:184–202, 2003.

[30] T. van Leeuwen, S. Aravkin, and F. Herrmann. Seismic waveform inversion by stochastic optimization. *Hindawi Intl. J. Geophysics*, 2011:doi:10.1155/2011/689041, 2012.

[31] J. Young and D. Ridzal. An application of random projection to parameter estimation in partial differential equations. *SIAM J. Scient. Comput.*, 34:A2344–A2365, 2012.

[32] Yaming Yu et al. Some stochastic inequalities for weighted sums. *Bernoulli*, 17(3):1044–1053, 2011.

[33] P. Zhao. Some new results on convolutions of heterogeneous gamma random variables. *Journal of Multivariate Analysis*, 102(5):958–976, 2011.

[34] P. Zhao and N. Balakrishnan. Mean residual life order of convolutions of heterogeneous exponential random variables. *Journal of Multivariate Analysis*, 100(8):1792–1801, 2009.

A Proof

In what follows $X \sim Gamma(\alpha, \beta)$ denotes a gamma r.v parametrized by shape $\alpha > 0$ and rate $\beta > 0$, f_X and F_X stand, respectively, for the probability density function (PDF) and cumulative distribution function (CDF) of a r.v X. Bold face letters denote vectors and the vector components are denoted as $\mathbf{v} = (v_1, v_2, \ldots, v_n)$.

For the proof of Theorem 1, we need to make use of the following additional results. The proof of Theorem 4 is identical to the proof of [29, Theorem 4]. Note that [29, Theorem 4] has been stated in terms of convolution of chi-squared r.v's but the proof, there, has been given for the more general case of arbitrary gamma r.v's. The details of the proofs for Lemmas 5 and 7 can be found in [25]. Lemma 6 has been stated in [25] but the proof is omitted there. We give a detailed proof of Lemma 6 here for completeness.

Theorem 4 is essential in proving our results and it states that an arbitrary convolution of heterogeneous gamma random variables (not necessarily with a common shape or a common rate) has an unique mode. Recall that a PDF, $f(x)$, is called unimodal if there exists a unique $x = a$ such that $f(x)$ is non-decreasing for $x < a$ and $f(x)$ is non-increasing for $x > a$. The point a is called the unique mode of $f(x)$.

Theorem 4 *Let $X_i \sim Gamma(\alpha_i, \beta_i)$, $i = 1, 2, \ldots, n$, be independent r.v's, where $\alpha_i, \beta_i > 0 \; \forall i$. The PDF of $Y_n := \sum_{i=1}^{n} \lambda_i X_i$ is unimodal where $\lambda_i \geq 0 \; \forall i$.*

Lemma 5 ([25, Lemma B.1]) *Let $X_i \sim Gamma(\alpha_i, \beta_i)$, $i = 1, 2, \ldots, n$, be independent r.v's, where $\alpha_i, \beta_i > 0 \; \forall i$. Define $Y_n := \sum_{i=1}^{n} \lambda_i X_i$ for $\lambda_i > 0$, $\forall i$ and $\rho_j := \sum_{i=1}^{j} \alpha_i$. Then for the PDF of Y_n, f_{Y_n}, we have*

(i) $f_{Y_n} > 0$, $\forall x > 0$,

(ii) f_{Y_n} is analytic on $\mathbb{R}^+ = \{x | x > 0\}$,

(iii) $f_{Y_n}^{(k)}(0) = 0$, if $0 \leq k < \rho_n - 1$, where $f_{Y_n}^{(k)}$ denotes the kth derivative of f_{Y_n}.

Lemma 6 ([25, Lemma B.2]) *Let $X_i \sim Gamma(\alpha_i, \alpha)$, $i = 1, 2, \ldots, n$, be independent r.v's, where $\alpha_i > 0 \; \forall i$ and $\alpha > 0$. Also let $\psi \sim Gamma(1, \alpha)$ be another r.v independent of all X_i's. If $\sum_{i=1}^{n} \alpha_i > 1$, then the mode, $\bar{x}(\lambda)$, of the r.v $W(\lambda) = Y + \lambda \psi$ is strictly increasing in $\lambda > 0$, where $Y = \sum_{i=1}^{n} \lambda_i X_i$ with $\lambda_i > 0$, $\forall i$.*

Proof By Lemma 5, $\bar{x}(\lambda) > 0$ for $\lambda \geq 0$. By the unimodality of $W(\lambda)$, for any $\lambda > \lambda_0 > 0$, it is enough to show that

$$J\big(\lambda, \bar{x}(\lambda_0)\big) := \left[\frac{d^2}{dx^2} \Pr\left(W(\lambda) \leq x\right) \right]_{x = \bar{x}(\lambda_0)} > 0. \tag{7}$$

Note that $J\big(\lambda_0, \bar{x}(\lambda_0)\big) = 0$ and since $\sum_{i=1}^n \alpha_i > 1$, by Lemma 5(iii), $f_Y(0) = 0$. So we have

$$
\begin{aligned}
J\big(\lambda, \bar{x}(\lambda_0)\big) &= \left[\frac{d}{dx} \int_0^x f_Y(x-z) \frac{\alpha}{\lambda} e^{-\frac{\alpha}{\lambda} z} dz \right]_{x=\bar{x}(\lambda_0)} \\
&= \left[\int_0^x \frac{d}{dx} f_Y(x-z) \frac{\alpha}{\lambda} e^{-\frac{\alpha}{\lambda} z} dz \right]_{x=\bar{x}(\lambda_0)} \\
&= \int_0^{\bar{x}(\lambda_0)} f_Y'(z) \frac{\alpha}{\lambda} e^{-\frac{\alpha}{\lambda}\left(\bar{x}(\lambda_0)-z\right)} dz.
\end{aligned}
$$

Therefore,

$$
\int_0^{\bar{x}(\lambda_0)} f_Y'(z) e^{\frac{\alpha z}{\lambda_0}} dz = \frac{\lambda_0}{\alpha} e^{\frac{\alpha \bar{x}(\lambda_0)}{\lambda_0}} J\big(\lambda_0, \bar{x}(\lambda_0)\big) = 0.
$$

Thus for $\lambda > \lambda_0 > 0$, we have

$$
\begin{aligned}
\frac{\lambda}{\alpha} e^{\frac{\alpha \bar{x}(\lambda)}{\lambda}} J\big(\lambda, \bar{x}(\lambda_0)\big) &= \int_0^{\bar{x}(\lambda_0)} f_Y'(z) e^{\frac{\alpha z}{\lambda}} dz \\
&= \int_0^{\bar{x}(\lambda_0)} f_Y'(z) e^{\frac{\alpha z}{\lambda}} - f_Y'(z) e^{\frac{\alpha z}{\lambda_0}} e^{\alpha \bar{x}(0)\left(\frac{1}{\lambda} - \frac{1}{\lambda_0}\right)} dz \\
&= \int_0^{\bar{x}(\lambda_0)} f_Y'(z) \left(e^{\frac{\alpha z}{\lambda}} - e^{\frac{\alpha z}{\lambda_0} + \alpha \bar{x}(0)\left(\frac{1}{\lambda} - \frac{1}{\lambda_0}\right)} \right) dz \\
&= \int_0^{\bar{x}(\lambda_0)} f_Y'(z) \left(e^{\frac{\alpha z}{\lambda}} - e^{\frac{\alpha z}{\lambda} + \Phi\left(z, \bar{x}(0)\right)} \right) dz,
\end{aligned}
$$

where $\bar{x}(0) > 0$ is the mode of r.v Y and

$$
\Phi\big(z, \bar{x}(0)\big) := \alpha\Big(z - \bar{x}(0)\Big)\left(\frac{1}{\lambda_0} - \frac{1}{\lambda}\right).
$$

Now if $z < \bar{x}(0)$ then $\Phi\big(z, \bar{x}(0)\big) < 0$ and $f_Y'(z) > 0$ so we get $J\big(\lambda, \bar{x}(\lambda_0)\big) > 0$. Similarly if $z > \bar{x}(0)$ then $\Phi\big(z, \bar{x}(0)\big) > 0$ and $f_Y'(z) < 0$ and again we have $J\big(\lambda, \bar{x}(\lambda_0)\big) > 0$.

Lemma 7 ([25, Lemma B.3]) *For some $\alpha_2 \geq \alpha_1 > 0$, let $\xi_1 \sim Gamma(1 + \alpha_1, \alpha_1)$ and $\xi_2 \sim Gamma(1 + \alpha_2, \alpha_2)$ be independent gamma r.v's. Also let $\bar{x} = \bar{x}(\lambda)$ denote the mode of the r.v $\xi(\lambda) = \lambda \xi_1 + (1 - \lambda)\xi_2$ for $0 \leq \lambda \leq 1$. Then $1 \leq \bar{x}(\lambda) \leq \big(2\sqrt{\alpha_1 \alpha_2} + 1\big) / \big(2\sqrt{\alpha_1 \alpha_2}\big)$, $\forall 0 \leq \lambda \leq 1$, with $\bar{x}(0) = \bar{x}(1) = 1$ and, in case of $\alpha_i = \alpha_j = \alpha$, $\bar{x}(1/2) = (2\alpha + 1) / (2\alpha)$, otherwise the inequalities are strict.*

A.1 Proof of Theorem 1

We prove the theorem for the case where $\alpha = \beta$ and $s = 1$. The general case follows from the scaling properties of gamma distribution.

We first consider the case where $n \geq 3$. If $\boldsymbol{\lambda} \succ \boldsymbol{\mu}$, then there exists a finite number, r, of vectors $\boldsymbol{\eta}_i, i = 1, 2, \ldots r$, such that $\boldsymbol{\lambda} = \boldsymbol{\eta}_1 \succ \boldsymbol{\eta}_2 \succ \ldots \succ \boldsymbol{\eta}_{r-1} \succ \boldsymbol{\eta}_r = \boldsymbol{\mu}$, and $\boldsymbol{\eta}_i$ and $\boldsymbol{\eta}_{i+1}$ differ in two coordinates only, $i = 1, 2, \ldots, r - 1$, see [22, 12.5.a]. Thus we may, without loss of generality assume that $\boldsymbol{\lambda}$, and $\boldsymbol{\mu}$ differ only in two coordinates, and in fact, assume that for some $1 \leq j < k \leq n$, we have

$$(\lambda_j, \lambda_k) \succ (\mu_j, \mu_k) \quad \text{and} \quad \lambda_i = \mu_i \text{ for } i \in \{1, \ldots, n\} \backslash \{j, k\}. \tag{8}$$

For $t \in [0, 1]$, define

$$\begin{aligned}
\nu_i(t) &:= t\lambda_i + (1 - t)\mu_i, \quad i = j, k, \\
\nu_i(t) &:= \lambda_i, \quad i \neq j, k, \\
Y(t) &:= \sum_{i=1}^n \nu_i(t)X_i. \tag{9}
\end{aligned}$$

It suffices to show that the CDF of $Y(t)$, in $t \in [0, 1]$, is non-increasing for $x > (2\alpha + 1)/(2\alpha)$ and non-decreasing for $x < (\alpha - 1)/\alpha$. Now we take the Laplace transform of $F_{Y(t)}$ as

$$\begin{aligned}
J(t, z) := \mathcal{L}[F_{Y(t)}](z) &= \int_0^\infty e^{-zx} F_{Y(t)}(x) dx \\
&= \frac{-1}{z} \int_0^\infty F_{Y(t)}(x) d\left(e^{-zx}\right) \\
&= \frac{1}{z} \int_0^\infty e^{-zx} dF_{Y(t)}(x) \\
&= \frac{1}{z} \mathcal{L}[Y(t)](z),
\end{aligned}$$

where $\mathcal{L}[Y(t)](z)$ is the Laplace transform of $Y(t)$ as

$$\mathcal{L}[Y(t)](z) = \prod_{i=1}^n \left(1 + \frac{\nu_i(t)z}{\alpha}\right)^{-\alpha}, \quad \text{for } z \in \mathbb{C}, \ Re(z) > -\min_{1 \leq i \leq n} \frac{\alpha}{\nu_i(t)}.$$

Differentiating with respect to t yields

$$\begin{aligned}
\frac{\partial J}{\partial t}(t, z) &= J(t, z) \frac{\partial}{\partial t}(\ln(J)) \\
&= J(t, z) \sum_{i=j,k} \frac{(\mu_i - \lambda_i)z}{1 + \frac{\nu_i(t)z}{\alpha}}.
\end{aligned}$$

We take the inverse transform to get

$$\begin{aligned}
\frac{\partial}{\partial t} F_{Y(t)}(x) &= \sum_{i=j,k} (\mu_i - \lambda_i) \frac{\partial}{\partial x} \Pr(Y(t) + \nu_i(t)\psi_i \leq x) \\
&= \sum_{i=j,k} (\mu_i - \lambda_i) f_{Y(t)+\nu_i(t)\psi_i}(x) \\
&= (\mu_j - \lambda_j)[f_{Y(t)+\nu_j(t)\psi_j}(x) - f_{Y(t)+\nu_k(t)\psi_k}(x)],
\end{aligned}$$

where $\psi_j, \psi_k \sim Gamma(1, \alpha)$ are i.i.d gamma r.v's which are also independent of all X_i's. By (8), we must have that $\lambda_j \geq \mu_j$, so it suffices to show that $f_{Y(t)+\nu_j(t)\psi_j}(x) \geq f_{Y(t)+\nu_k(t)\psi_k}(x)$ for $x > (2\alpha + 1)/(2\alpha)$ and $f_{Y(t)+\nu_j(t)\psi_j}(x) \leq f_{Y(t)+\nu_k(t)\psi_k}(x)$ for $x < (\alpha - 1)/\alpha$. On the other hand, using the Laplace transform and inverting it again, one can show the following identity (for any integer $k \geq 1$ and reals $a, b \geq 0$)

$$\frac{d^{k-1}}{dx^{k-1}} f_{X+a\psi_1}(x) - \frac{d^{k-1}}{dx^{k-1}} f_{X+b\psi_2}(x) = \frac{1}{\alpha}(b - a)\frac{d^k}{dx^k} f_{X+a\psi_1+b\psi_2}(x), \tag{10}$$

where X is an arbitrary continuous positive r.v and $\psi_i \sim Gamma(1, \alpha)$, $i = 1, 2$, are i.i.d gamma r.v's which are also independent of X. As such we have

$$f_{Y(t)+\nu_j(t)\psi_j}(x) - f_{Y(t)+\nu_k(t)\psi_k}(x) = \frac{1}{\alpha}(\nu_k(t) - \nu_j(t))\frac{\partial}{\partial x} f_{Y(t)+\nu_j(t)\psi_j+\nu_k(t)\psi_k}(x).$$

For given $1 \leq j < k \leq n$, consider the r.v $Y(t) + \nu_j(t)\psi_j + \nu_k(t)\psi_k$. By (9), we get that $\nu_j(t) \geq \nu_k(t)$ for $t \in [0, 1]$, hence, it is only left to show that, for any $t \in [0, 1]$, the mode of this r.v, under the conditions

$$\nu_1(t) \geq \ldots \geq \nu_n(t) \geq 0,$$
$$\sum_{i=1}^n \nu_i(t) = 1,$$

falls between $(\alpha - 1)/\alpha$ and $(2\alpha + 1)/(2\alpha)$. By Lemma 6, the mode of r.v $Y(t) + \nu_1(t)\psi_1 + \nu_2(t)\psi_2$ is greater than that of $Y(t) + \nu_j(t)\psi_j + \nu_k(t)\psi_k$ for $1 < j < k$, and also the mode of $Y(t) + \nu_{n-1}(t)\psi_{n-1} + \nu_n(t)\psi_n$ is smaller than that of $Y(t) + \nu_j(t)\psi_j + \nu_k(t)\psi_k$ for $j < k < n$. Hence, we only need to show that for the mode of the r.v

$$Y_1(t) := Y(t) + \nu_1(t)\psi_1 + \nu_2(t)\psi_2 \tag{11a}$$

denoted by $\bar{x}_1(t)$, we have $\bar{x}_1(t) \leq (2\alpha + 1)/(2\alpha)$ and for the mode of the r.v

$$Y_n(t) := Y(t) + \nu_{n-1}(t)\psi_{n-1} + \nu_n(t)\psi_n \tag{11b}$$

denoted by $\bar{x}_n(t)$, we have $\bar{x}_n(t) \geq (\alpha - 1)/\alpha$. In what follows, we fix any $t \in [0, 1]$ and, for notational simplicity, denote $Y_1(t), \bar{x}_1(t), Y_n(t)$ and $\bar{x}_n(t)$ by $Y_1, \bar{x}_1, Y_n$ and $\bar{x}_n$, respectively.

We first prove the case for Y_1. At any mode of Y_1, we have[1]

$$\left[\frac{\partial}{\partial x} f_{Y_1}(x, \boldsymbol{\nu})\right]_{(\bar{x}_1(\boldsymbol{\nu}), \boldsymbol{\nu})} = 0,$$

$$\left[\frac{\partial^2}{\partial x^2} f_{Y_1}(x, \boldsymbol{\nu})\right]_{(\bar{x}_1(\boldsymbol{\nu}), \boldsymbol{\nu})} < 0,$$

---

[1]Since mode is unique, thus f must be strictly concave at the mode

hence implicit function theorem yields

$$\frac{\partial \bar{x}_1}{\partial \nu_1}(\boldsymbol{\nu}) = -\frac{\frac{\partial^2 f_{Y_1}}{\partial \nu_1 \partial \bar{x}_1}(\bar{x}_1(\boldsymbol{\nu}), \boldsymbol{\nu})}{\frac{\partial^2 f_{Y_1}}{\partial \bar{x}_1^2}(\bar{x}_1(\boldsymbol{\nu}), \boldsymbol{\nu})}. \tag{12}$$

Let $\boldsymbol{\nu}^*$ be where the global maximum of $\bar{x}_1(\boldsymbol{\nu})$, denoted by $\bar{x}_1^*$, occurs. Thus by the necessary condition of maximality, we must have

$$\left[\frac{\partial^2}{\partial \nu_1 \partial x} f_{Y_1}(x, \boldsymbol{\nu})\right]_{(\bar{x}_1^*, \boldsymbol{\nu}^*)} = 0.$$

Suppose where the maximum occurs, i.e., $\boldsymbol{\nu}^*$, there is a nonzero coefficient, ν_3, such that $0 < \nu_3 < \nu_2 \leq \nu_1$. The case of $\nu_3 = 0$ will be dealt with at the end of the proof. Fixing all other coefficients, we vary ν_2 and ν_3 under the condition $\nu_2 + \nu_3 = const$. That is, we take the directional derivative in the direction $\delta\boldsymbol{\nu} = (0, 1, -1, 0 \ldots, 0)$ for the "n" dimensional vector $\boldsymbol{\nu} = (\nu_1, \nu_2, \nu_3, \ldots, \nu_n)$. In other words, we consider the change in the direction of $\boldsymbol{\nu} + \gamma\delta\boldsymbol{\nu}$. Using Laplace transform, we get

$$\frac{d}{d\gamma}\mathcal{L}[Y_1](z) = \mathcal{L}[Y_1](z)\frac{d}{d\gamma}\ln\left[\mathcal{L}[Y_1](z)\right]$$

$$= \mathcal{L}[Y_1](z)\frac{d}{d\gamma}\left(-(1+\alpha)\ln(1+\frac{\nu_2 z}{\alpha}) - \alpha\sum_{\substack{i=0 \\ i \neq 2}}^{n}\ln(1+\frac{\nu_i z}{\alpha}) - \ln(1+\frac{\nu_1 z}{\alpha})\right)$$

$$= \mathcal{L}[Y_1](z)\left(-\frac{(1+\alpha)z}{\alpha}\frac{1}{(1+\frac{\nu_2 z}{\alpha})} + \frac{z}{(1+\frac{\nu_3 z}{\alpha})}\right)$$

$$= z\mathcal{L}[Y_1](z)\left(\frac{1}{(1+\frac{\nu_3 z}{\alpha})} - \frac{1}{(1+\frac{\nu_2 z}{\alpha})} - \frac{1}{\alpha}\frac{1}{(1+\frac{\nu_2 z}{\alpha})}\right)$$

$$= z\mathcal{L}[Y_1](z)\left(\frac{\frac{(\nu_2-\nu_3)z}{\alpha}}{(1+\frac{\nu_3 z}{\alpha})(1+\frac{\nu_2 z}{\alpha})} - \frac{1}{\alpha}\frac{1}{(1+\frac{\nu_2 z}{\alpha})}\right).$$

Now inverting the above and differentiating (w.r.t x), we get

$$\frac{\partial^2}{\partial \gamma \partial x} f_{Y_1}(x, \boldsymbol{\nu}) = \frac{(\nu_2 - \nu_3)}{\alpha}\frac{\partial^3}{\partial x^3} f_{\tilde{Y}_{23}}(x, \boldsymbol{\nu}) - \frac{1}{\alpha}\frac{\partial^2}{\partial x^2} f_{Y_1+\nu_2\xi_2}(x, \boldsymbol{\nu}), \tag{13}$$

where $\tilde{Y}_{23} = Y_1 + \nu_2\xi_2 + \nu_3\xi_3$ with $\xi_2, \xi_3 \sim Gamma(1, \alpha)$ being i.i.d r.v.'s, independent of all others appearing before. So at the maximum, $\bar{x}_1^*$, we must have

$$\left[\frac{\partial^2}{\partial \gamma \partial x} f_{Y_1}(x, \boldsymbol{\nu})\right]_{(\bar{x}_1^*, \boldsymbol{\nu}^*)} = 0. \tag{14}$$

Now consider a slight perturbation of $\delta\boldsymbol{\nu} = (0, 1, -1, 0\ldots, 0)$ as $\delta\boldsymbol{\nu}(\varepsilon_0) = (0, 1, -(1 + \varepsilon_0), 0\ldots, 0)$ for some $\varepsilon_0 > 0$. Computations similar as above yields

$$\frac{\partial^2}{\partial\gamma\partial x} f_{Y_1}(x, \boldsymbol{\nu}) = \frac{(\nu_2 - \nu_3)}{\alpha} \frac{\partial^3}{\partial x^3} f_{\tilde{Y}_{23}}(x, \boldsymbol{\nu}) \quad - \quad \frac{1}{\alpha}\frac{\partial^2}{\partial x^2} f_{Y_1 + \nu_2\xi_2}(x, \boldsymbol{\nu})$$

$$+ \quad \varepsilon_0 \frac{\partial^2}{\partial x^2} f_{Y_1 + \nu_3\xi_3}(x, \boldsymbol{\nu}).$$

So at the maximum, $\bar{x}_1^*$, using (14), we must have

$$\left[\frac{\partial^2}{\partial\gamma\partial x} f_{Y_1}(x, \boldsymbol{\nu})\right]_{(\bar{x}_1^*, \boldsymbol{\nu}^*)} = \varepsilon_0 \left[\frac{\partial^2}{\partial x^2} f_{Y_1 + \nu_3\xi_3}(x, \boldsymbol{\nu})\right]_{(\bar{x}_1^*, \boldsymbol{\nu}^*)}$$

$$= -\frac{\alpha\varepsilon_0}{\nu_3}\left[\frac{\partial}{\partial x} f_{Y_1 + \nu_3\xi_3}(x, \boldsymbol{\nu})\right]_{(\bar{x}_1^*, \boldsymbol{\nu}^*)} < 0.$$

This followed from (10) (with $k = 2$, $a = 0$ and $b = \nu_3$) and Lemma 6, as well as noting that $\left[\frac{\partial}{\partial x} f_{Y_1}(x, \boldsymbol{\nu})\right]_{(\bar{x}_1^*, \boldsymbol{\nu}^*)} = 0$. Using (12), the previous relation implies that the mode must increase if a sufficiently small step is taken along the perturbed direction $\delta\boldsymbol{\nu}(\varepsilon_0) = (0, -1, (1 + \varepsilon_0), 0\ldots, 0)$. In other words, for $\varepsilon_1 > 0$ small enough, for the mode of r.v

$$Y_1^{(\varepsilon_0,\varepsilon_1)} := \nu_1 X_1 + (\nu_2 - \varepsilon_1)X_2 + \big(\nu_3 + \varepsilon_1(1 + \varepsilon_0)\big)X_3 + \sum_{i=4}^{n} \nu_i X_i + \nu_1\psi_1 + (\nu_2 - \varepsilon_1)\psi_2, \quad (15)$$

denoted by $\bar{x}_1^{(\varepsilon_0,\varepsilon_1)}$, we must have that

$$\bar{x}_1^* < \bar{x}_1^{(\varepsilon_0,\varepsilon_1)}. \tag{16}$$

It is clear that ε_1 depends on ε_0 and consequently, as ε_0 gets smaller, ε_1 might get smaller as well. However, $\varepsilon_1\varepsilon_0 \in o(\varepsilon_0)$ and $\varepsilon_1\varepsilon_0 \in o(\varepsilon_1)$, where "$o$" denotes the "*little o*". In other words, consider decreasing sequences of $(\varepsilon_0^n)_{n=1}^{\infty}$ and $(\varepsilon_1^n)_{n=1}^{\infty}$. Since $\lim_{n\to\infty}(\varepsilon_1^n\varepsilon_0^n)/\varepsilon_0^n = \lim_{n\to\infty}(\varepsilon_1^n\varepsilon_0^n)/\varepsilon_1^n = 0$, it follows that, as ε_0 gets smaller, $\varepsilon_1\varepsilon_0$ term in (15) vanishes at a faster rate than either of ε_0 or ε_1. Hence using (15) and (16), as well as the continuity of $\bar{x}_1$, we can consider a r.v

$$Y_1^{\varepsilon} := \nu_1 X_1 + (\nu_2 - \varepsilon)X_2 + (\nu_3 + \varepsilon)X_3 + \sum_{i=4}^{n} \nu_i X_i + \nu_1\psi_1 + (\nu_2 - \varepsilon)\psi_2,$$

which has the same form as Y_1 but whose mode, denoted by $\bar{x}_1^{\varepsilon}$, satisfies

$$\bar{x}_1^* < \bar{x}_1^{\varepsilon},$$

which contradicts the maximality of $\bar{x}_1^*$. This means that, if $\nu_3 \neq 0$, then the maximum must occurs at $\boldsymbol{\nu}^*$ with $\nu_3 = \nu_2 \leq \nu_1$. Similar arguments show that at the maximum, we must have $\nu_n = \ldots = \nu_3 = \nu_2 \leq \nu_1$. But by (13) and the fact that

$$\left[\frac{\partial^2}{\partial x^2} f_{Y_1 + \nu_2\xi_2}(x, \boldsymbol{\nu})\right]_{(\bar{x}_1^*, \boldsymbol{\nu}^*)} = -\frac{\alpha}{\nu_2}\left[\frac{\partial}{\partial x} f_{Y_1 + \nu_2\xi_2}(x, \boldsymbol{\nu})\right]_{(\bar{x}_1^*, \boldsymbol{\nu}^*)} < 0,$$

we must have that $\nu_3 \neq \nu_2$. As such, the only possibility is $\nu_n = \ldots = \nu_3 = 0$, i.e., the maximum mod of Y_1 coincides with that of r.v $\zeta = \nu\zeta_1 + (1-\nu)\zeta_2$ where $\zeta_1, \zeta_2 \sim Gamma(1+\alpha, \alpha)$. So now lemma 7 gives

$$\bar{x}_1^* \leq \frac{2\alpha+1}{2\alpha}.$$

Now consider the case for Y_n. Suppose where the minimum occurs, i.e., $\boldsymbol{\nu}^*$, we have, $\nu_{n-1} > 0$ and $\nu_{n-2} > \nu_{n-1}$. Now by the same reasoning as in the case of Y_1, we can consider a r.v

$$Y_n^\varepsilon := \sum_{i=1}^{n-3} \nu_i X_i \nu_1 X_1 + (\nu_{n-2}-\varepsilon)X_{n-2} + (\nu_{n-1}+\varepsilon)X_{n-1} + \nu_n X_n + (\nu_{n-1}+\varepsilon)\psi_{n-1} + \nu_n\psi_n,$$

which has the same form as Y_n but whose mode, denoted by $\bar{x}_n^\varepsilon$, satisfies

$$\bar{x}_n^\varepsilon < \bar{x}_n^*,$$

which contradicts the minimality of $\bar{x}_n^*$. This means that, if $\nu_{n-1} \neq 0$, then the minimum must occurs at $\boldsymbol{\nu}^*$ with $\nu_{n-2} = \nu_{n-1}$. But again, as in the end of the proof for the case of Y_1, we must have that $\nu_{n-2} \neq \nu_{n-1}$. As such the only possibility is $\nu_{n-1} = 0$ (which would also mean that $\nu_n = 0$, although this case can also be independently established with the same reasoning as above). Hence, the minimum mode of Y_n must coincide with that of the r.v $\sum_{i=1}^{n-2} \nu_i X_i$. Again, appealing to the same line of reasoning, at the minimum mode of $\sum_{i=1}^{n-2} \nu_i X_i$, we must either have $\nu_j = 0$ or $\nu_{j-1} = \nu_j$ for any $1 < j \leq n-2$. Hence, for $n \geq 3$, we get

$$\bar{x}_n^* = \min_{1 \leq j \leq n-2} \bar{x}_{j\alpha} = \frac{\alpha-1}{\alpha},$$

where $\bar{x}_{j\alpha}$ denotes the mode of the r.v $Gamma(j\alpha, j\alpha)$.

The case of $n = 2$, amounts to studying the mode of r.v $\zeta = \nu\zeta_1 + (1-\nu)\zeta_2$, denoted by $\bar{x}(\zeta)$, where $\zeta_1, \zeta_2 \sim Gamma(1+\alpha, \alpha)$. Direct application of Lemma 7 yields $1 \leq \bar{x}(\zeta) \leq (2\alpha+1)/(2\alpha)$. Theorem 1 is proved. ∎

A.2 Proof of Corollary 2

The proof goes along the same line as that of Theorem 1. For $t \in [0,1]$, we again define

$$
\begin{aligned}
\nu_i(t) &:= t\lambda_i + (1-t)\mu_i, \quad i = j, k, \\
\nu_i(t) &:= \lambda_i, \quad i \neq j, k, \\
Y(t) &:= \sum_{i=1}^{n} \nu_i(t)X_i.
\end{aligned}
$$

where

$$\nu_1(t) \geq \ldots \geq \nu_n(t) \geq 0, \quad t \in [0,1],$$

$$\sum_{i=1}^{n} \nu_i(t) = s_{\boldsymbol{\mu}} + t(s_{\boldsymbol{\lambda}} - s_{\boldsymbol{\mu}}).$$

Computations identical to the proof of Theorem 1 give

$$\bar{x}_1^*(t) \ \leq \ \left(s_{\boldsymbol{\mu}} + t(s_{\boldsymbol{\lambda}} - s_{\boldsymbol{\mu}})\right)\frac{2\alpha + 1}{2\alpha},$$

$$\bar{x}_n^*(t) \ \geq \ \left(s_{\boldsymbol{\mu}} + t(s_{\boldsymbol{\lambda}} - s_{\boldsymbol{\mu}})\right)\frac{\alpha - 1}{\alpha},$$

where $\bar{x}_1^*(t)$ and $\bar{x}_n^*(t)$ are, respectively, the maximum and minimum mode of r.v's $Y_1(t)$ and $Y_n(t)$ which are defined similarly as in (11). Now for $t \in [0,1]$, we get

$$\bar{x}_1^* \ = \ \max_{t \in [0,1]} \bar{x}_1^*(t) \leq \frac{(2\alpha + 1)s_{\boldsymbol{\lambda}}}{2\alpha},$$

$$\bar{x}_n^* \ = \ \min_{t \in [0,1]} \bar{x}_n^*(t) \geq \frac{(\alpha - 1)s_{\boldsymbol{\mu}}}{\alpha},$$

which give the desired results. ∎

A.3 Proof of Corollary 3

Define

$$P_n(\boldsymbol{\lambda}_n; \alpha, \beta, x) := \Pr\left(\sum_{i=1}^{n} \lambda_i X_i < x\right),$$

where

$$\boldsymbol{\lambda}_n := (\lambda_1, \lambda_2, \ldots, \lambda_n).$$

By the continuity from above, it is clear that

$$P_\infty(\boldsymbol{\lambda}; \alpha, \beta, x) = \lim_{n \to \infty} P_n(\boldsymbol{\lambda}_n; \alpha, \beta, x).$$

Now since $\boldsymbol{\mu}_n \prec_w \boldsymbol{\lambda}_n$, Corollary 2 yields

$$P_n(\boldsymbol{\mu}; \alpha, \beta, x) \ \geq \ P_n(\boldsymbol{\lambda}; \alpha, \beta, x), \quad \forall x > \frac{(2\alpha + 1)\sum_{i=1}^{n} \lambda_i}{2\beta},$$

$$P_n(\boldsymbol{\mu}; \alpha, \beta, x) \ \leq \ P_n(\boldsymbol{\lambda}; \alpha, \beta, x), \quad \forall x < \frac{(\alpha - 1)\sum_{i=1}^{n} \mu_i}{\beta}, \quad \alpha > 1.$$

Taking the limit as $n \to \infty$ gives the desired result. ∎