Monotonicity properties of the gamma family of distributions

Iosif Pinelis

Department of Mathematical Sciences
Michigan Technological University
Houghton, Michigan 49931, USA
E-mail: ipinelis@mtu.edu

Abstract

For real $a > 0$, let $X_a$ denote a random variable with the gamma distribution with parameters $a$ and 1. Then $\Pr(X_a - a > c)$ is increasing in $a$ for each real $c \geq 0$; non-increasing in $a$ for each real $c \leq -1/3$; and non-monotonic in $a$ for each $c \in (-1/3, 0)$. This extends and/or refines certain previously established results.

Keywords: stochastic monotonicity, gamma distribution, incomplete gamma function, logarithmic mean

2010 MSC: 26D15, 33B20, 60E15, 62E15

1. Summary and discussion

For any real $a > 0$, let $X_a$ denote a random variable with the gamma distribution with parameters $a$ and 1, so that for any real $c > -a$

$$\Pr(X_a - a > c) = \frac{\Gamma(a, a + c)}{\Gamma(a)},$$

where

$$\Gamma(a, x) = \int_x^{\infty} t^{a-1} e^{-t} \, dt$$

for real $x > 0$; expression (1) defines the incomplete gamma function.

There are quite a few bounds on the incomplete gamma function in the literature; see e.g. [2, 13] and references therein.

The main result of the present paper is

Theorem 1. The probability $\Pr(X_a - a > c)$ is

(I) increasing in real $a > 0$ for each real $c \geq 0$;

(II) decreasing in real $a > -c$ for each real $c \leq -1/3$;

(III) non-monotonic in real $a > -c$ for each $c \in (-1/3, 0)$.
The terms “increasing” and “decreasing” are understood in this note in the strict sense, as “strictly increasing” and “strictly decreasing”.

Remark 2. Since \( P(X_a - a > c) = 1 \) for \( a \in (0, -c) \), in parts (II) and (III) of Theorem 1 one may replace the condition \( a > -c \) by \( a > 0 \), albeit for the price of replacing “decreasing” in part (II) by “non-increasing”.

Corollary 3. For all real \( a > 0 \)

\[ P(X_a - a > 0) < 1/2 < P(X_a - a > -1/3). \]

This immediately follows from parts (I) and (II) of Theorem 1 because, by the central limit theorem, \( P(X_a > a + o(\sqrt{a})) \to 1/2 \) as \( a \to \infty \). In turn, Corollary 3 immediately implies

Corollary 4. For each real \( a > 0 \), the median of \( X_a - a \) is in the interval \((-1/3, 0)\).

Part (I) of Theorem 1 was previously obtained in [8], where it was proved by a quite different method, which does not appear to be working for \( c < 0 \).

Corollary 4 was previously given in [6]. Refinements of this result – but only for the natural values of \( a \) – were obtained in [1, 2, 4, 7].

Corollary 3 improves and generalizes the main result of [14], that \( P(X_n - n > 0) < 1/2 < P(X_n - n > -1) \) for natural \( n \).

As usual for results on stochastic monotonicity (cf. e.g. [3, Section 4] and [11, Section 4]), a straightforward application of Theorem 1 is to statistical testing, as follows. A sample \( Y \) is taken from the centered gamma distribution with shape parameter \( \theta > 0 \) and scale parameter 1. We test the null hypothesis \( H_0: \theta = \theta_0 \) (for some given real \( \theta_0 > 0 \)) versus the alternative hypothesis \( H_1: \theta > \theta_0 \), using the test \( \delta(Y) := I\{Y > c\} \) with a real critical value \( c > 0 \), where \( I(\cdot) \) denotes the indicator function. Then, according to part (I) of Theorem 1 the power function \( \beta_\delta \) of the test, given by the formula \( \beta_\delta(\theta) := E_\theta \delta(Y) = P(Y > c) = P(X_\theta - \theta > c) \) for all real \( \theta > 0 \), will be increasing. In particular, it follows that the test \( \delta \) is unbiased. Part (II) of Theorem 1 can be used similarly.

2. Proof of Theorem 1

2.1. Proof of part (I) of Theorem 1 (and of part (II) concerning \( c \leq -1 \))

Take any real \( c \) and any real \( a > 0 \lor (-c) \). Then

\[ p(a) := p_c(a) := P(X_a - a > c) = \frac{\Gamma(a, a + c)}{\Gamma(a)} = 1/\left(1 + \frac{\gamma(a, a + c)}{\Gamma(a, a + c)}\right), \quad (2) \]

where \( \gamma(a, a + c) := \Gamma(a) - \Gamma(a, a + c) \). Note that

\[ \Gamma(a, a + c) = \int_{a+c}^\infty t^{a-1}e^{-t} \, dt = (a+c)^a \int_1^\infty x^{a-1}e^{-(a+c)x} \, dx, \]

\[ \gamma(a, a + c) = \int_0^{a+c} t^{a-1}e^{-t} \, dt = (a+c)^a \int_0^1 x^{a-1}e^{-(a+c)x} \, dx. \]
So,
\[ P(X_a - a > c) = \frac{1}{1 + R(a - 1)}, \]
where
\[ R(u) := \frac{I(u)}{J(u)}. \]

\[ I(u) := \int_1^0 f(x)u e^{-(1+c)x} \, dx = \int_0^1 z^u p(z) \, dz, \]
\[ J(u) := \int_1^\infty f(x)u e^{-(1+c)x} \, dx = \int_0^1 z^u q(z) \, dz, \]
\[ f(x) := xe^{1-x}, \]
\[ p(z) := e^{-(1+c)x_1(z)}x_1'(z) > 0, \quad q(z) := -e^{-(1+c)x_2(z)}x_2'(z) > 0, \]
\[ x_1(z) \text{ is the only root } x \in (0, 1) \text{ of the equation } f(x) = z \text{ for } z \in (0, 1), \]
\[ x_2(z) \text{ is the only root } x \in (1, \infty) \text{ of the equation } f(x) = z \text{ for } z \in (0, 1). \]

One might note that for \( z \in (0, 1) \) we have \( x_1(z) = -W(-z/e) \) and \( x_2(z) = -W_1(-z/e) \), where \( W \) denotes the principal branch of Lambert’s function and \( W_{-1} \) denotes its \((-1)\) branch – see e.g. [9, pages 330–331].

It follows that
\[ 2J(u)^2 R'(u) = 2 \int_0^1 \int_0^1 dy \, dx \, (xy)^u \, p(x)q(y)(\ln x - \ln y) \]
\[ = 2 \int_0^1 \int_0^1 dy \, dx \, (xy)^u \, p(x)q(x)(\ln y - \ln x) \]
\[ = \int_0^1 \int_0^1 dy \, dx \, (xy)^u [p(x)q(y) - p(y)q(x)](\ln x - \ln y) \]
\[ = \int_0^1 \int_0^1 dy \, dx \, (xy)^u \, p(y)q(y)[r(x) - r(y)](\ln x - \ln y), \]
where
\[ r := p/q. \]

Differentiating the identities \( f(x_j(z)) \equiv z \) for \( j = 1, 2 \) in \( z \in (0, 1) \), we have
\[ x_j'(z) = \frac{x_j(z)}{(1-x_j(z))z}, \]
which implies
\[ r' = -Am_c, \]
where
\[ A := A(z) := \frac{e^{(c+1)(x_2-x_1)}(x_2-x_1)x_1}{(1-x_1)^3(x_2-1)x_2z} \in (0, \infty) \]
and
\[ m_c := m_c(z) := c(x_1 + x_2 - 2) + (c + 1)(1 - x_1 x_2) \]
\[ = 1 - x_1 x_2 + c(1 - x_1)(x_2 - 1); \] \hfill (5)

here we write \( x_1 \) and \( x_2 \) in place of \( x_1(z) \) and \( x_2(z) \), for brevity.

Further, the condition \( f(x_1) = f(x_2) \) means that the logarithmic mean of \( x_1 \) and \( x_2 \) is 1; recall that the logarithmic mean of two distinct positive real numbers \( x \) and \( y \) is defined by the formula
\[ L(x,y) := \frac{y - x}{\ln y - \ln x}. \]

Then the arithmetic-logarithmic-geometric mean inequality (see e.g. \cite{11}, formula (4)) yields \( \sqrt{x_1 x_2} < 1 < (x_1 + x_2)/2 \), so that \( x_1 + x_2 - 2 > 0 \) and \( 1 - x_1 x_2 > 0 \). So, \( m_c > 0 \) if \( c \geq 0 \) and \( m_c < 0 \) if \( c \leq -1 \). Since the sign of \( r' \) is opposite to that of \( m_c \), we see that the function \( r \) is decreasing (on \((0,1)\)) if \( c \geq 0 \) and increasing if \( c \leq -1 \). Therefore, by \( \text{(i)} \), \( R' < 0 \) and hence \( R \) is decreasing if \( c \geq 0 \) and \( R' > 0 \) and hence \( R \) is increasing if \( c \leq -1 \).

Now part (I) of Theorem \ref{th1} follows by (3) (as well as part (II) concerning \( c \leq -1 \)).

2.2. Proof of part (II) of Theorem \ref{th1}

Now it is also seen that, to complete the proof of part (II) of Theorem \ref{th1}, it suffices to prove Lemma \ref{le1} below; in fact, only the implication \( \text{(v)} \Rightarrow \text{(i)} \) in Lemma \ref{le1} will be needed for this purpose.

**Lemma 1.** Take any real \( c \). The following statements are equivalent to one another:

\( (i) \) \( m_c < 0 \) on \((0,1)\);

\( (ii) \) \( c < \frac{x_1 x_2 - 1}{(1 - x_1)(x_2 - 1)} \) on \((0,1)\);

\( (iii) \) \( c < \frac{xy - L(x,y)^2}{(L(x,y) - x)(y - L(x,y))} \) whenever \( 0 < x < y < \infty \);

\( (iv) \) \( c < \lambda(y) := \frac{y - l(y)^2}{(l(y) - 1)(y - l(y))} \) for all real \( y > 1 \), where \( l(y) := L(1,y) \);

\( (v) \) \( c \leq -1/3 \).

**Proof of Lemma \ref{le1}**. The equivalence \( (i) \iff (ii) \) follows immediately from \( \text{(5)} \).

The implication \( (iii) \Rightarrow (ii) \) holds because \( L(x_1, x_2) = 1 \), as was noted before.

To prove the implication \( (ii) \Rightarrow (iii) \), take any \( x \) and \( y \) such that \( 0 < x < y < \infty \). Let \( b := L(x,y) \). Then \( x/b \in (0,1) \), \( y/b \in (1,\infty) \), and \( f(x/b) = f(y/b) =: z^* \). Then \( x/b = x_1(z^*) \) and \( y/b = x_2(z^*) \). So, (ii) will imply
\[ c < \frac{(x/b)(y/b) - 1}{(1 - (x/b))(y/b - 1)} = \frac{xy - L(x,y)^2}{(L(x,y) - x)(y - L(x,y))}. \]
This proves the implication (ii) \( \Rightarrow \) (iii).

The equivalence (iii) \( \iff \) (iv) follows immediately by homogeneity.

The remaining equivalence (iv) \( \iff \) (v) holds by the following lemma. \(\square\)

**Lemma 2.** The function \( \lambda \) defined in Lemma 1 is increasing on \((1, \infty)\), from \( \lambda(1+) = -1/3 \).

The proof Lemma 2 is based on what was referred to as special l’Hospital-type rule for monotonicity:

**Proposition 5.** [See e.g. \[12, Proposition 4.1.\].] Suppose that \(-\infty \leq A < B \leq \infty \). Let \( f \) and \( g \) be differentiable functions defined on the interval \((A, B)\) such that \( f(A+) = g(A+) = 0 \). Suppose further that \( g \) and \( g' \) do not take on the zero value and do not change their respective signs on \((A, B)\). Finally, suppose that the “derivative ratio” \( f'/g' \) is increasing on \((A, B)\). Then the ratio \( f/g \) is also increasing on \((A, B)\).

**Proof of Lemma 2.** Note that \( \lambda = f/g \), where

\[
\begin{align*}
  f(y) &= \frac{y \ln^2 y - (y - 1)^2}{y}, \\
  g(y) &= \frac{(y - \ln y - 1)(y \ln y - y + 1)}{y},
\end{align*}
\]

everywhere in this proof, \( y \) is an arbitrary real number \( > 1 \). Note also that \( f(1+) = g(1+) = 0 \). Next, here the “derivative ratio” is

\[
\frac{f'(y)}{g'(y)} = \frac{f_1(y)}{g_1(y)},
\]

where

\[
\begin{align*}
  f_1(y) &= yf'(y) = \frac{1}{y} - y + 2 \ln y, \\
  g_1(y) &= yg'(y) = \frac{(y - 1)^2 \ln y}{y}.
\end{align*}
\]

We have \( f_1(1+) = g_1(1+) = 0 \). Next, the “derivative ratio” for \( f_1/g_1 \) is

\[
\frac{f'_1(y)}{g'_1(y)} = \frac{f_2(y)}{g_2(y)},
\]

where

\[
\begin{align*}
  f_2(y) &= \frac{y^2}{y^2 - 1} f_1(y) = \frac{1 - y}{1 + y}, \\
  g_2(y) &= \frac{y^2}{y^2 - 1} g_1(y) = \ln y + \frac{y - 1}{1 + y}.
\end{align*}
\]

We have \( f_2(1+) = g_2(1+) = 0 \). Further, the “derivative ratio” for \( f_2/g_2 \) is

\[
\frac{f_3(y)}{g_3(y)} = -\frac{2y}{1 + 4y + y^2},
\]

whose derivative \( 2(y^2 - 1)/(1 + 4y + y^2)^2 \) is \( > 0 \), for real \( y > 1 \). Applying now Proposition 5 three times, we see that \( \lambda = f/g \) is indeed increasing. Moreover, applying the l’Hospital-type rule for limits three times, we see that \( \lambda(1+) = r_3(1+) = -1/3 \). Lemma 2 is now proved. \(\square\)
This completes the proof of parts (I) and (II) of Theorem 1.

Remark 6. It follows from Lemma 1 that

\[ L(x, y) < \tilde{G}(x, y) := \sqrt{xy + \frac{1}{3} (L(x, y) - x)(y - L(x, y))} \]  

whenever \(0 < x < y < \infty\), and the constant factor \(\frac{1}{3}\) here is optimal. This complements the logarithmic-geometric mean inequality \(\sqrt{xy} < L(x, y)\) for distinct positive real \(x, y\). Also, inequality (6) represents an improvement of the arithmetic-logarithmic mean inequality \(L(x, y) < \frac{1}{2} (x + y)\). Indeed, one can show that

\[ \tilde{G}(x, y) < \frac{1}{2} (x + y), \]

again whenever \(0 < x < y < \infty\). This can be done by a method similar to the one used in the proof of Lemma 2, but this time also utilizing the general l'Hospital-type rule for monotonicity given by [12, Corollary 3.1].

2.3. Proof of part (III) of Theorem 1

Take any \(c \in (-1/3, 0)\). Then, by (2), \(p((-c)+) = 1\), whereas \(p(a) < 1\) for real \(a > -c\). So, \(p(a) = P(X_a - a > c)\) is not increasing in \(a\) in any right neighborhood of 0.

To complete the proof of part (III) of Theorem 1 it suffices to show that \(p(a + 1) > p(a)\) for all large enough \(a > 0\). Recalling (2) again and then using integration by parts in the integral expression for \(\Gamma(a + 1, a + 1 + c)\), we have

\[
\Gamma(a + 1)(p(a + 1) - p(a)) = \Gamma(a + 1, a + 1 + c) - a\Gamma(a, a + c)
\]

\[
= (a + c + 1)^a e^{-a-c-1} - a \int_{a+c}^{a+c+1} x^{a-1} e^{-x} \, dx
\]

\[
= (a + c + 1)^a \left( e^{-a-c-1} - a \int_{1/(a+c+1)}^{1} u^{a-1} e^{-(a+c+1)u} \, du \right).
\]

So, letting \(a \to \infty\),

\[
b := c + 1 \in (2/3, 1), \quad \varepsilon := \frac{1}{a + b}(\frac{1}{2} 0),
\]

and using the substitution \(z = (a + b)v\), we get \(v = \varepsilon z, a = 1/\varepsilon - b\), and

\[
(p(a + 1) - p(a)) \frac{\Gamma(a + 1)}{(a + c + 1)^a e^{-a-c-1}} = 1 - a \int_0^{1/(a+b)} (1 - v)^{a-1} e^{(a+b)v} \, dv
\]

\[
= \int_0^{1} g(\varepsilon, z) \, dz,
\]

where

\[
g(\varepsilon, z) := 1 - (1 - b\varepsilon)(1 - z\varepsilon)^{1/\varepsilon - b-1} e^z
\]

\[
= 1 - \exp\{ (z - b + bz - z^2/2)\varepsilon + O(\varepsilon^2) \}
\]

\[
= -(z - b + bz - z^2/2)\varepsilon + O(\varepsilon^2);
\]
everywhere here, the constant factors in the \(O(\cdot)'s\) are universal. So,

\[
2 \int_0^1 g(\varepsilon, z) \, dz = (b - 2/3)\varepsilon + O(\varepsilon^2) > 0
\]

for all small enough \(\varepsilon > 0\), in view of (7). Thus, indeed \(p(a + 1) > p(a)\) for all large enough \(a > 0\).

This completes the proof of part (III) of Theorem 1, and thereby the entire proof of Theorem 1.

References

[1] Adell, J.A., Jodrás, P., 2005. Sharp estimates for the median of the \(\Gamma(n + 1, 1)\) distribution. Statist. Probab. Lett. 71, 185–191. URL: https://doi.org/10.1016/j.spl.2004.10.025

[2] Alm, S.E., 2003. Monotonicity of the difference between median and mean of gamma distributions and of a related Ramanujan sequence. Bernoulli 9, 351–371. URL: https://doi.org/10.3150/bj/1068128981

[3] Alzer, H., 1997. On some inequalities for the incomplete gamma function. Math. Comp. 66, 771–778. URL: https://doi.org/10.1090/S0025-5718-97-00814-4

[4] Alzer, H., 2005. Proof of the Chen-Rubin conjecture. Proc. Roy. Soc. Edinburgh Sect. A 135, 677–688. URL: https://doi.org/10.1017/S0308210500004066

[5] Anderson, T.W., Samuels, S.M., 1967. Some inequalities among binomial and Poisson probabilities, in: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, University of California Press, Berkeley, Calif., pp. 1–12. URL: https://projecteuclid.org/euclid.bsmsp/1200512976

[6] Chen, J., Rubin, H., 1986. Bounds for the difference between median and mean of gamma and Poisson distributions. Statist. Probab. Lett. 4, 281–283. URL: https://doi.org/10.1016/0167-7152(86)90044-1

[7] Choi, K.P., 1994. On the medians of gamma distributions and an equation of Ramanujan. Proc. Amer. Math. Soc. 121, 245–251. URL: https://doi.org/10.2307/2160389

7
[8] Chojnacki, W., 2008. Some monotonicity and limit results for the regularised incomplete gamma function. Ann. Polon. Math. 94, 283–291. URL: https://doi.org/10.4064/ap94-3-7 doi:10.4064/ap94-3-7

[9] Corless, R.M., Gonnet, G.H., Hare, D.E.G., Jeffrey, D.J., Knuth, D.E., 1996. On the Lambert W function. Adv. Comput. Math. 5, 329–359. URL: http://dx.doi.org/10.1007/BF02124750 doi:10.1007/BF02124750

[10] Hoeffding, W., 1956. On the distribution of the number of successes in independent trials. Ann. Math. Statist. 27, 713–721. URL: https://doi.org/10.1214/aoms/1177728178 doi:10.1214/aoms/1177728178

[11] Lin, T.P., 1974. The power mean and the logarithmic mean. Amer. Math. Monthly 81, 879–883. URL: https://doi.org/10.2307/2319447 doi:10.2307/2319447

[12] Pinelis, I., 2006. On l’Hospital-type rules for monotonicity. JIPAM. J. Inequal. Pure Appl. Math. 7, Article 40, 19 pp. (electronic), www.emis.de/journals/JIPAM/images/157_05_JIPAM/157_05.pdf

[13] Pinelis, I., 2020. Exact lower and upper bounds on the incomplete gamma function. arXiv:2005.06384 [math.CA], to appear in Mathematical Inequalities & Applications.

[14] Vietoris, L., 1983. Dritter Beweis der die unvollständige Gammafunktion betreffenden Lochsschen Ungleichungen. Österreich. Akad. Wiss. Math.-Natur. Kl. Sitzungsber. II 192, 83–91.