The probability density function and geometric mean of the square/line characteristic distance ratio

Abstract

We consider two distances, one between two random points in a square with sides *a* and *a*, and the other between two random points on a line of length *a*. We analytically determine the probability density function for their ratio. The distribution is near log-normal and has a geometric mean of exp(π/3+1/3*ln2-7/12) or 2.0035…. We verify these calculations with a Monte Carlo simulation.

1. Introduction

We wish to determine the length of the line that accommodates the same degree of horizontal variability in atmospheric constituents as a square does.

We introduce the relevant part of Philip's (2007) work (Sect. 2); define our own problem (Sect. 3); note the similarities between Philip's and our problems (Sect. 4); solve our problem (Sect. 5); calculate the geometric mean of the resulting density function (Sect. 6); and verify the results with a Monte Carlo simulation (Sect. 7).

2. Philip's work

Here we introduce the part of Philip's (2007) work (both problem and solution) relevant to ours.

2.1 Problem

He lets the two random points be (X_1, Y_1, Z_1) and (X_2, Y_2, Z_2). He assumes that $X_1$and X_2 are independent and evenly distributed in the interval (0, *a*). The same is assumed for Y_1 and Y_2 in (0, *b*) and for Z_1 and Z_2 in (0, *c*).

We are only interested in the special case where *a*=*b*=*c*. Therefore, *b* and *c* in his equations are replaced with *a* below; suffix ($_a$ as in f_a, g_a, h_a, k_a) is omitted.

He considers the probability distribution function (*P*) for four events.

$$\begin{aligned}
F(t) &\equiv P((X_1 - X_2)^2 \le t) \\
G(s) &\equiv P((X_1 - X_2)^2 + (Y_1 - Y_2)^2 \le s) \\
H(u) &\equiv P((X_1 - X_2)^2 + (Y_1 - Y_2)^2 + (Z_1 - Z_2)^2 \le u) \\
K(v) &\equiv P((X_1 - X_2)^2 + (Y_1 - Y_2)^2 + (Z_1 - Z_2)^2 \le v^2)
\end{aligned}.$$

He lets the corresponding density functions be

$$
\begin{aligned}
f(t) &\equiv dF(t)/dt \\
g(s) &\equiv dG(s)/ds \\
h(u) &\equiv dH(u)/du \\
k(v) &\equiv dK(v)/dv
\end{aligned}
.
$$

All variables in his formulation above are independent of each other, except

$$u = v^2 .$$

2.2 Solution

See his paper for complete derivation of the density functions. In brief, he derives $F(t)$ from the geometry. $f(t)$ is its derivative by definition. $g(s)$ is the convolution of the probability density function f

$$g(s) = \int f(s-t) f(t) dt .$$

$h(u)$ is the convolution of f and g

$$h(u) = \int f(u-s) g(s) ds .$$

Once $h(u)$ is yielded, $k(v)$ is given by replacing u with v^2. More precisely,

$$k(v) \equiv \frac{dK(v)}{dv} = \frac{du}{dv}\frac{dK(v)}{du} = \frac{dv^2}{dv}\frac{dH(u)}{du} = 2vh(v^2) .$$

This is because $K(v)=H(u)$. $K(v)$ and $H(u)$ refer to an identical event (examine the definitions above, while noting $u=v^2$).

Resulting analytical solutions relevant to our work are

$$F(t) = \begin{cases} 1 - \left(1 - \sqrt{t}/a\right)^2, & 0 < t \le a^2; \\ 1, & a^2 < t. \end{cases}$$

$$f(t) = \frac{1}{a\sqrt{t}} - \frac{1}{a^2}, \qquad 0 < t \le a^2.$$

$$g(s) = \begin{cases} \dfrac{\pi}{a^2} - \dfrac{4\sqrt{s}}{a^3} + \dfrac{s}{a^4}, & 0 < s \le a^2; \\ \dfrac{1}{a^2}\left(-2 - \pi + 4\arcsin\left(\dfrac{a}{\sqrt{s}}\right)\right) + \dfrac{4\sqrt{s - a^2}}{a^3} - \dfrac{s}{a^4}, & a^2 < s \le 2a^2. \end{cases}$$

These are his equations (1), (2) and (6), after replacing b with a. Note $g(s)$ takes either one of the two forms above depending on the value of s.

3. Our problem

We define

$$\begin{aligned} &\tau \equiv \ln t^{-1} \\ &\sigma \equiv \ln s \end{aligned}.$$

It follows that

$$\begin{aligned} &t = e^{-\tau} \\ &\frac{dt}{d\tau} \equiv -t \\ &s = e^{\sigma} \\ &\frac{ds}{d\sigma} \equiv s \end{aligned}.$$

As t goes from 0 to a^2, τ goes from ∞ to $\ln(a^{-2})$. As s goes from 0 to $2a^2$, σ goes from $-\infty$ to $\ln(2a^2)$.

We also define, for a non-negative number R,

$$\upsilon \equiv \ln(R^2), 0 \le R \le \infty .$$

As R goes from 0 to ∞, υ goes from $-\infty$ to ∞.

With these variables, we consider the following four probability distributions.

$$\begin{aligned} &\Phi(\tau) \equiv P(\ln((X_1 - X_2)^{-2}) \le \tau) \\ &\Gamma(\sigma) \equiv P(\ln((X_1 - X_2)^2 + (Y_1 - Y_2)^2) \le \sigma) \\ &\mathrm{H}(\upsilon) \equiv P(\ln((X_1 - X_2)^2 + (Y_1 - Y_2)^2) + \ln((Z_1 - Z_2)^{-2}) \le \upsilon) \\ &\mathrm{K}(R) \equiv P(\ln((X_1 - X_2)^2 + (Y_1 - Y_2)^2) + \ln((Z_1 - Z_2)^{-2}) \le \ln(R^2)) \end{aligned}.$$

Let the corresponding density functions be

$$\begin{aligned}\varphi(\tau) &\equiv d\Phi(\tau)/d\tau\\ \gamma(\sigma) &\equiv d\Gamma(\sigma)/d\sigma\\ \eta(\upsilon) &\equiv d\mathrm{H}(\upsilon)/d\upsilon\\ \kappa(R) &\equiv d\mathrm{K}(R)/dR\end{aligned}.$$

4. Similarities between Philip's and our problems

In Philip's formulation (Sect. 2.1), the distance between two random points in an *a*x*a*x*a* cube is expressed as $\sqrt{(X_1-X_2)^2+(Y_1-Y_2)^2+(Z_1-Z_2)^2}$. $K(v)$ is the probability that this distance is smaller than v. One can view its density function $k(v)$ as the ultimate goal of his work, and $f(t)$, $g(s)$ and $h(u)$ as tools for deriving $k(v)$.

Likewise, in our formulation (Sect. 3), the square/line ratio of characteristic distances is expressed as $\sqrt{(X_1-X_2)^2+(Y_1-Y_2)^2}/\sqrt{(Z_1-Z_2)^2}$. $\mathrm{K}(R)$ is the probability that this ratio is smaller than R. Our goal is to analytically solve for its density function $\kappa(R)$. We use $\varphi(\tau)$, $\gamma(\sigma)$ and $\eta(\upsilon)$ as tools.

5. Solution to our problem

We take three steps. First, $\gamma(\sigma)$ and $\varphi(\tau)$ are derived based on their analogy to $g(s)$ and $f(t)$, respectively. Second, the convolution of $\gamma(\sigma)$ and $\varphi(\tau)$ yields $\eta(\upsilon)$. Third, replacing υ in $\eta(\upsilon)$ with $\ln R^2$ yields $\kappa(R)$. The first and third steps are straightforward (Sect. 5.1 and 5.3, respectively). The second step is painful and is broken down to three sub-steps (Sect. 5.2.1, 5.2.2 and 5.2.3).

5.1 First step: $\gamma(\sigma)$ and $\varphi(\tau)$

$G(s)$ and $\Gamma(\sigma)$ refer to an identical event (examine their definitions above, while noting $\sigma=\ln s$). Therefore,

$$\Gamma(\sigma)=G(s).$$

Therefore,

$$\gamma(\sigma)\equiv\frac{d\Gamma(\sigma)}{d\sigma}=\frac{ds}{d\sigma}\frac{d\Gamma(\sigma)}{ds}=s\frac{dG(s)}{ds}=sg(s).$$

Because $g(s)$ takes either one of the two forms depending on the value of s (Sect. 2.2), so does $\gamma(\sigma)$.

$$\gamma_1(\sigma) = \frac{s\pi}{a^2} - \frac{4s\sqrt{s}}{a^3} + \frac{s^2}{a^4} = e^{2\sigma}\frac{1}{a^4} - e^{3\sigma/2}\frac{4}{a^3} + e^{\sigma}\frac{\pi}{a^2}, \quad \sigma \le \ln(a^2)$$

$$\gamma_2(\sigma) = \frac{s}{a^2}\left(-2 - \pi + 4\arcsin\left(\frac{a}{\sqrt{s}}\right)\right) + \frac{4s\sqrt{s-a^2}}{a^3} - \frac{s^2}{a^4}$$
$$= \frac{e^{\sigma}}{a^2}\left(-2 - \pi + 4\arcsin\left(\frac{a}{e^{\sigma/2}}\right)\right) + \frac{4e^{\sigma}\sqrt{e^{\sigma} - a^2}}{a^3} - \frac{e^{2\sigma}}{a^4}, \quad \ln(a^2) < \sigma \le \ln(2a^2)$$

$F(t)$ and $\Phi(\tau)$ refer to converse events (examine their definitions above, while noting $\tau = \ln t^{-1}$). Therefore,

$$\Phi(\tau) = -F(t).$$

Therefore,

$$\varphi(\tau) \equiv \frac{d\Phi(\tau)}{d\tau} = \frac{dt}{d\tau}\frac{d\Phi(\tau)}{dt} = -t\frac{d(-F(t))}{dt} = tf(t) = \frac{\sqrt{t}}{a} - \frac{t}{a^2}$$

$$= \frac{\sqrt{e^{-\tau}}}{a} - \frac{e^{-\tau}}{a^2}, \qquad \ln(a^{-2}) \le \tau$$

In the next step, φ shows up in the form of $\varphi(\upsilon-\sigma)$

$$\varphi(\upsilon - \sigma) = \frac{\sqrt{e^{-\upsilon+\sigma}}}{a} - \frac{e^{-\upsilon+\sigma}}{a^2}$$

$$= e^{\sigma}\frac{-e^{-\upsilon}}{a^2} + e^{\sigma/2}\frac{e^{-\upsilon/2}}{a}, \qquad \sigma \le \upsilon + \ln(a^2)$$

5.2 Second step: $\eta(\upsilon)$

Just like $h(u)$ is the convolution of f and g, $\eta(\upsilon)$ is the convolution of φ and γ

$$\eta(\upsilon) = \int \varphi(\upsilon - \sigma)\gamma(\sigma)d\sigma$$

Depending on the value of υ, $\eta(\upsilon)$ can be written as $\eta_1(\upsilon)$, $\eta_2(\upsilon)$ or $\eta_3(\upsilon)$

$$\eta_1(\upsilon) = \int_{-\infty}^{\upsilon+\ln(a^2)} \varphi(\upsilon-\sigma)\gamma_1(\sigma)d\sigma, \upsilon \le 0$$
$$\eta_2(\upsilon) = \int_{-\infty}^{\ln(a^2)} \varphi(\upsilon-\sigma)\gamma_1(\sigma)d\sigma + \int_{\ln(a^2)}^{\upsilon+\ln(a^2)} \varphi(\upsilon-\sigma)\gamma_2(\sigma)d\sigma, 0 < \upsilon \le \ln 2$$
$$\eta_3(\upsilon) = \int_{-\infty}^{\ln(a^2)} \varphi(\upsilon-\sigma)\gamma_1(\sigma)d\sigma + \int_{\ln(a^2)}^{\ln(2a^2)} \varphi(\upsilon-\sigma)\gamma_2(\sigma)d\sigma, \ln 2 < \upsilon$$

We pursue $\eta(\upsilon)$ for these three cases individually.

5.2.1 $\eta_1(\upsilon)$

$$\begin{aligned}
\eta_1(\upsilon) &= \int_{-\infty}^{\upsilon+\ln(a^2)} \varphi(\upsilon-\sigma)\gamma_1(\sigma)d\sigma \\
&= \int_{-\infty}^{\upsilon+\ln(a^2)} (e^{\sigma}(-\frac{e^{-\upsilon}}{a^2}) + e^{\sigma/2}(\frac{e^{-\upsilon/2}}{a}))(e^{2\sigma}\frac{1}{a^4} - e^{3\sigma/2}\frac{4}{a^3} + e^{\sigma}\frac{\pi}{a^2})d\sigma \\
&= \int_{-\infty}^{\upsilon+\ln(a^2)} e^{3\sigma}(-\frac{e^{-\upsilon}}{a^6}) + e^{5\sigma/2}(\frac{e^{-\upsilon/2}+4e^{-\upsilon}}{a^5}) + e^{2\sigma}(\frac{-\pi e^{-\upsilon}-4e^{-\upsilon/2}}{a^4}) + e^{3\sigma/2}\frac{\pi e^{-\upsilon/2}}{a^3}d\sigma \\
&= \frac{a^6}{3}e^{3\upsilon}(-\frac{e^{-\upsilon}}{a^6}) + \frac{2a^5}{5}e^{5\upsilon/2}(\frac{e^{-\upsilon/2}+4e^{-\upsilon}}{a^5}) + \frac{a^4}{2}e^{2\upsilon}(\frac{-\pi e^{-\upsilon}-4e^{-\upsilon/2}}{a^4}) + \frac{2a^3}{3}e^{3\upsilon/2}\frac{\pi e^{-\upsilon/2}}{a^3} \\
&= -\frac{1}{3}e^{2\upsilon} + \frac{2}{5}e^{2\upsilon} + \frac{8}{5}e^{3\upsilon/2} + \frac{-\pi}{2}e^{\upsilon} - 2e^{3\upsilon/2} + \frac{2\pi}{3}e^{\upsilon} \\
&= \frac{1}{15}e^{2\upsilon} - \frac{2}{5}e^{3\upsilon/2} + \frac{\pi}{6}e^{\upsilon}, \upsilon \le 0
\end{aligned}$$

Here we have used

$$\int_{-\infty}^{\upsilon+\ln(a^2)} Ce^{p\sigma}d\sigma = \frac{C}{p}e^{p(\upsilon+\ln(a^2))} = \frac{a^{2p}}{p}Ce^{p\upsilon}$$

5.2.2 $\eta_2(\upsilon)$

$\eta_2(\upsilon)$ is the sum of $\eta_{21}(\upsilon)$ and $\eta_{22}(\upsilon)$ below.

$$\eta_{21}(\upsilon) = \int_{-\infty}^{\ln(a^2)} \varphi(\upsilon-\sigma)\gamma_1(\sigma)d\sigma$$
$$\eta_{22}(\upsilon) = \int_{\ln(a^2)}^{\upsilon+\ln(a^2)} \varphi(\upsilon-\sigma)\gamma_2(\sigma)d\sigma$$

$\eta_{21}(\upsilon)$ is a special case of $\eta_1(\upsilon)$, and is obtained by integrating up to $\ln(a^2)$ instead of $\upsilon + \ln(a^2)$.

$$\begin{aligned}
\eta_{21}(\upsilon) &= \int_{-\infty}^{\ln(a^2)} e^{3\sigma}(-\frac{e^{-\upsilon}}{a^6}) + e^{5\sigma/2}(\frac{e^{-\upsilon/2}+4e^{-\upsilon}}{a^5}) + e^{2\sigma}(\frac{-\pi e^{-\upsilon}-4e^{-\upsilon/2}}{a^4}) + e^{3\sigma/2}\frac{\pi e^{-\upsilon/2}}{a^3}d\sigma \\
&= \frac{a^6}{3}(-\frac{e^{-\upsilon}}{a^6}) + \frac{2a^5}{5}(\frac{e^{-\upsilon/2}+4e^{-\upsilon}}{a^5}) + \frac{a^4}{2}(\frac{-\pi e^{-\upsilon}-4e^{-\upsilon/2}}{a^4}) + \frac{2a^3}{3}\frac{\pi e^{-\upsilon/2}}{a^3} \\
&= \frac{1}{30}e^{-\upsilon}((38-15\pi)+(-48+20\pi)e^{\upsilon/2})
\end{aligned}$$

$$\eta_{22}(\upsilon) = \int_{\ln(a^2)}^{\upsilon+\ln(a^2)} \varphi(\upsilon-\sigma)\gamma_2(\sigma)d\sigma$$

$$= \int_{\ln(a^2)}^{\upsilon+\ln(a^2)} (e^{\sigma}(-\frac{e^{-\upsilon}}{a^2}) + e^{\sigma/2}(\frac{e^{-\upsilon/2}}{a}))(\frac{e^{\sigma}}{a^2}(-2-\pi+4\arcsin(\frac{a}{e^{\sigma/2}})) + \frac{4e^{\sigma}\sqrt{e^{\sigma}-a^2}}{a^3} - \frac{e^{2\sigma}}{a^4})d\sigma$$

$$= \Bigg[\frac{1}{30a^6}e^{-\upsilon}(20a^6e^{\upsilon/2}\sigma - 20a^3(\pi+2)e^{\upsilon/2}e^{3\sigma/2} + 15a^2(\pi+2)e^{2\sigma}$$
$$+20a^2e^{3\sigma/2}(4ae^{\upsilon/2}-3e^{\sigma/2})\arcsin(\frac{a}{e^{\sigma/2}})$$
$$+40a^6e^{\upsilon/2}\ln(\sqrt{1-\frac{a^2}{e^{\sigma}}}+1) - 30a^6e^{\upsilon/2}\ln(2(\sqrt{e^{\sigma}-a^2}+e^{\sigma/2}))$$
$$-20a^3e^{\sigma/2}\sqrt{1-\frac{a^2}{e^{\sigma}}}(2a^2-2ae^{\upsilon/2}e^{\sigma/2}+e^{\sigma})$$
$$-2a\sqrt{e^{\sigma}-a^2}(-16a^4+15a^3e^{\upsilon/2}e^{\sigma/2}-8a^2e^{\sigma}-30ae^{\upsilon/2}e^{3\sigma/2}+24e^{2\sigma})$$
$$-12ae^{\upsilon/2}e^{5\sigma/2}+10e^{3\sigma}) \Bigg]_{\ln(a^2)}^{\upsilon+\ln(a^2)}$$

$$= \frac{1}{30a^6}e^{-\upsilon}(20a^6e^{\upsilon/2}\upsilon - 20a^6(\pi+2)e^{\upsilon/2}(e^{3\upsilon/2}-1) + 15a^6(\pi+2)(e^{2\upsilon}-1)$$
$$+20a^6e^{2\upsilon}\arcsin(\frac{1}{e^{\upsilon/2}}) - 10a^6\pi(4e^{\upsilon/2}-3)$$
$$+40a^6e^{\upsilon/2}\ln(\sqrt{1-\frac{1}{e^{\upsilon}}}+1) - 30a^6e^{\upsilon/2}\ln(\sqrt{e^{\upsilon}-1}+e^{\upsilon/2})$$
$$-20a^5\sqrt{e^{\upsilon}-1}(2a-2ae^{\upsilon}+ae^{\upsilon})$$
$$-2a^2\sqrt{e^{\upsilon}-1}(-16a^4+7a^4e^{\upsilon}-6a^4e^{2\upsilon})$$
$$-12a^6e^{\upsilon/2}(e^{5\upsilon/2}-1) + 10a^6(e^{3\upsilon}-1))$$

$$= \frac{1}{30}e^{-\upsilon}(-2e^{3\upsilon} + e^{2\upsilon}(-5\pi-10) + e^{\upsilon/2}(-20\pi+52)$$
$$+\sqrt{e^{\upsilon}-1}(12e^{2\upsilon}+6e^{\upsilon}-8)$$
$$+10e^{\upsilon/2}\ln(\sqrt{e^{\upsilon}-1}+e^{\upsilon/2})$$
$$+20e^{2\upsilon}\arcsin(\frac{1}{e^{\upsilon/2}}) + 15\pi - 40)$$

Therefore,

$$
\begin{aligned}
&\eta_2(\upsilon) = \eta_{21}(\upsilon) + \eta_{22}(\upsilon) \\
&= \frac{1}{30} e^{-\upsilon}((38-15\pi) + (-48+20\pi)e^{\upsilon/2}) + \frac{1}{30} e^{-\upsilon}(-2e^{3\upsilon} + e^{2\upsilon}(-5\pi-10) + e^{\upsilon/2}(-20\pi+52) \\
&+ \sqrt{e^{\upsilon}-1}(12e^{2\upsilon} + 6e^{\upsilon} - 8) \\
&+ 10e^{\upsilon/2}\ln(\sqrt{e^{\upsilon}-1} + e^{\upsilon/2}) \\
&+ 20e^{2\upsilon}\arcsin(\frac{1}{e^{\upsilon/2}}) + 15\pi - 40) \\
&= \frac{1}{30} e^{-\upsilon}(-2e^{3\upsilon} + e^{2\upsilon}(-5\pi-10) + 4e^{\upsilon/2} \\
&+ \sqrt{e^{\upsilon}-1}(12e^{2\upsilon} + 6e^{\upsilon} - 8) \\
&+ 10e^{\upsilon/2}\ln(\sqrt{e^{\upsilon}-1} + e^{\upsilon/2}) \\
&+ 20e^{2\upsilon}\arcsin(\frac{1}{e^{\upsilon/2}}) - 2), 0 < \upsilon \le \ln 2
\end{aligned}
$$

5.2.3 $\eta_3(\upsilon)$

$\eta_3(\upsilon)$ is the sum of $\eta_{31}(\upsilon)$ and $\eta_{32}(\upsilon)$ below.

$$
\eta_{31}(\upsilon) = \int_{-\infty}^{\ln(a^2)} \varphi(\upsilon-\sigma)\gamma_1(\sigma)d\sigma
$$

$$
\eta_{32}(\upsilon) = \int_{\ln(a^2)}^{\ln(2a^2)} \varphi(\upsilon-\sigma)\gamma_2(\sigma)d\sigma
$$

$\eta_{31}(\upsilon)$ is the same as $\eta_{21}(\upsilon)$.

$$
\eta_{31}(\upsilon) = \eta_{21}(\upsilon) = \frac{1}{30} e^{-\upsilon}((38-15\pi) + (-48+20\pi)e^{\upsilon/2})
$$

$\eta_{32}(\upsilon)$ is a special case of $\eta_{22}(\upsilon)$, and is obtained by integrating up to $\ln(2a^2)$ instead of $\upsilon + \ln(a^2)$.

$$\eta_{32}(\upsilon) = \int_{\ln(a^2)}^{\ln(2a^2)} \varphi(\upsilon-\sigma)\gamma_2(\sigma)d\sigma$$

$$= \int_{\ln(a^2)}^{\ln(2a^2)} (e^{\sigma}(-\frac{e^{-\upsilon}}{a^2}) + e^{\sigma/2}(\frac{e^{-\upsilon/2}}{a}))(\frac{e^{\sigma}}{a^2}(-2-\pi+4\arcsin(\frac{a}{e^{\sigma/2}})) + \frac{4e^{\sigma}\sqrt{e^{\sigma}-a^2}}{a^3} - \frac{e^{2\sigma}}{a^4})d\sigma$$

$$= \Bigg[\frac{1}{30a^6} e^{-\upsilon}(20a^6 e^{\upsilon/2}\sigma - 20a^3(\pi+2)e^{\upsilon/2}e^{3\sigma/2} + 15a^2(\pi+2)e^{2\sigma}$$
$$+20a^2 e^{3\sigma/2}(4ae^{\upsilon/2} - 3e^{\sigma/2})\arcsin(\frac{a}{e^{\sigma/2}})$$
$$+40a^6 e^{\upsilon/2}\ln(\sqrt{1-\frac{a^2}{e^{\sigma}}}+1) - 30a^6 e^{\upsilon/2}\ln(2(\sqrt{e^{\sigma}-a^2}+e^{\sigma/2}))$$
$$-20a^3 e^{\sigma/2}\sqrt{1-\frac{a^2}{e^{\sigma}}}(2a^2 - 2ae^{\upsilon/2}e^{\sigma/2} + e^{\sigma})$$
$$-2a\sqrt{e^{\sigma}-a^2}(-16a^4 + 15a^3 e^{\upsilon/2}e^{\sigma/2} - 8a^2 e^{\sigma} - 30ae^{\upsilon/2}e^{3\sigma/2} + 24e^{2\sigma})$$
$$-12ae^{\upsilon/2}e^{5\sigma/2} + 10e^{3\sigma}) \Bigg]_{\ln(a^2)}^{\ln(2a^2)}$$

$$= \frac{1}{30}e^{-\upsilon}(e^{\upsilon/2}(10\ln(1+\sqrt{2}) - 20\pi + 2\sqrt{2} + 52) - 48 + 15\pi)$$

Therefore,

$$\begin{aligned}
\eta_3(\upsilon) &= \eta_{31}(\upsilon) + \eta_{32}(\upsilon) \\
&= \frac{1}{30}e^{-\upsilon}((38-15\pi) + (-48+20\pi)e^{\upsilon/2}) \\
&+ \frac{1}{30}e^{-\upsilon}(e^{\upsilon/2}(10\ln(1+\sqrt{2}) - 20\pi + 2\sqrt{2} + 52) - 48 + 15\pi) \\
&= \frac{1}{30}e^{-\upsilon}(e^{\upsilon/2}(10\ln(1+\sqrt{2}) + 2\sqrt{2} + 4) - 10), \ln 2 < \upsilon
\end{aligned}$$

5.3 Third step: $\kappa(R)$

H(υ) and K(R) refer to an identical event (examine their definitions above, while noting $\upsilon = \ln R^2$). Therefore,

$$\mathrm{K}(R) = \mathrm{H}(\upsilon)$$

Therefore,

$$\kappa(R) \equiv \frac{d\mathrm{K}(R)}{dR} = \frac{d\upsilon}{dR}\frac{d\mathrm{K}(R)}{d\upsilon} = \frac{d\ln(R^2)}{dR}\frac{d\mathrm{H}(\upsilon)}{d\upsilon} = \frac{2}{R}\eta(\ln(R^2))$$

$$\kappa_1(R) = \frac{2}{R}\eta_1(\ln(R^2))$$
$$= \frac{2}{R}(\frac{1}{15}R^4 - \frac{2}{5}R^3 + \frac{\pi}{6}R^2)$$
$$= \frac{1}{15}(2R^3 - 12R^2 + 5\pi R), 0 \le R \le 1$$
$$\kappa_2(R) = \frac{2}{R}\eta_2(\ln(R^2))$$
$$= \frac{1}{15}R^{-3}(-2R^6 + R^4(-5\pi - 10) + 4R$$
$$+ \sqrt{R^2 - 1}(12R^4 + 6R^2 - 8)$$
$$+ 10R\ln(\sqrt{R^2 - 1} + R)$$
$$+ 20R^4 \arcsin(\frac{1}{R}) - 2), 1 < R \le \sqrt{2}$$
$$\kappa_3(R) = \frac{2}{R}\eta_3(\ln(R^2))$$
$$= \frac{1}{15}R^{-3}(R(10\ln(1+\sqrt{2}) + 2\sqrt{2} + 4) - 10), \sqrt{2} < R$$

To express the probability density function on a log scale,

$$d\mathrm{K}(R)/d\ln R$$
$$= R\kappa(R)$$
$$=$$
$$\frac{1}{15}(2R^4 - 12R^3 + 5\pi R^2), 0 \le R \le 1$$
$$\frac{1}{15}R^{-2}(-2R^6 + R^4(-5\pi - 10) + 4R$$
$$+ \sqrt{R^2 - 1}(12R^4 + 6R^2 - 8)$$
$$+ 10R\ln(\sqrt{R^2 - 1} + R)$$
$$+ 20R^4 \arcsin(\frac{1}{R}) - 2), 1 < R \le \sqrt{2}$$
$$\frac{1}{15}R^{-2}(R(10\ln(1+\sqrt{2}) + 2\sqrt{2} + 4) - 10), \sqrt{2} < R$$

6. Geometric mean

Let the geometric mean ratio be *E*. Then

$$\ln E = \int_0^\infty \kappa(R) \ln R dR$$

Let

$$\varepsilon \equiv \ln E^2$$

Then,

$$\varepsilon = 2\int_0^\infty \kappa(R) \ln R dR$$

Because $\upsilon \equiv \ln(R^2)$,

$$\begin{aligned}\varepsilon &= \int \upsilon \eta(\upsilon) d\upsilon \\ &= \int \upsilon \int \varphi(\upsilon - \sigma)\gamma(\sigma) d\sigma d\upsilon\end{aligned}$$

To obtain *E*, we first calculate

$$\mu(\sigma) = \int \upsilon \varphi(\upsilon - \sigma) d\upsilon$$

Then, we get ε as the sum of integrals of the form

$$\varepsilon = \int \mu(\sigma)\gamma(\sigma) d\sigma$$

Finally *E* is given as

$$E = e^{\varepsilon/2}$$

6.1 First step: $\mu(\sigma)$

Because $\varphi(\upsilon - \sigma) = e^{\sigma}(-\frac{e^{-\upsilon}}{a^2}) + e^{\sigma/2}(\frac{e^{-\upsilon/2}}{a}), \sigma \le \upsilon + \ln(a^2)$ (Sect. 5.1),

$$\begin{aligned}&\int \upsilon(e^{\sigma}(-\frac{e^{-\upsilon}}{a^2}) + e^{\sigma/2}(\frac{e^{-\upsilon/2}}{a}))d\upsilon \\ &= \int (\upsilon e^{-\upsilon}(-\frac{e^{\sigma}}{a^2}) + \upsilon e^{-\upsilon/2}(\frac{e^{\sigma/2}}{a}))d\upsilon \\ &= \frac{e^{\sigma}}{a^2}(\upsilon+1)e^{-\upsilon} - \frac{e^{\sigma/2}}{a}(2\upsilon+4)e^{-\upsilon/2}, \sigma - \ln(a^2) \le \upsilon\end{aligned}$$

According to l'Hôpital's rule, both $\upsilon e^{-\upsilon}$ and $\upsilon e^{-\upsilon/2}$ approach 0 as υ approaches ∞. Therefore,

$$
\begin{aligned}
\mu(\sigma) &= \left.\frac{e^{\sigma}}{a^2}(\upsilon+1)e^{-\upsilon} - \frac{e^{\sigma/2}}{a}(2\upsilon+4)e^{-\upsilon/2}\right|_{\upsilon=\sigma-\ln(a^2)}^{\infty} \\
&= -\frac{e^{\sigma}}{a^2}(\sigma-\ln(a^2)+1)e^{-\sigma}a^2 + \frac{e^{\sigma/2}}{a}(2\sigma-2\ln(a^2)+4)e^{-\sigma/2}a \\
&= \sigma-\ln(a^2)+3
\end{aligned}
$$

6.2 Second step: *ε*

ε can be written as

$$
\begin{aligned}
\varepsilon &= \int \mu(\sigma)\gamma(\sigma)d\sigma = \varepsilon_1 + \varepsilon_2 \\
\varepsilon_1 &= \int_{-\infty}^{\ln a^2} \mu(\sigma)\gamma_1(\sigma)d\sigma \\
\varepsilon_2 &= \int_{\ln a^2}^{\ln 2a^2} \mu(\sigma)\gamma_2(\sigma)d\sigma
\end{aligned}
$$

We pursue *ε* for these two terms individually.

6.2.1 ε_1

$$
\begin{aligned}
\varepsilon_1 &= \int_{-\infty}^{\ln a^2} \mu(\sigma)\gamma_1(\sigma)d\sigma \\
&= \int_{-\infty}^{\ln a^2} (\sigma-\ln(a^2)+3)(e^{2\sigma}\frac{1}{a^4} - e^{3\sigma/2}\frac{4}{a^3} + e^{\sigma}\frac{\pi}{a^2})d\sigma \\
&= \frac{1}{4}(2\ln a^2 - 1) - \frac{16}{9}(\frac{3}{2}\ln a^2 - 1) + \pi(\ln a^2 - 1) + (-\ln a^2 + 3)(\frac{1}{2} - \frac{8}{3} + \pi) \\
&= 2\pi - \frac{179}{36}
\end{aligned}
$$

6.2.2 ε_2

$$\varepsilon_2 = \int_{\ln a^2}^{\ln 2a^2} \mu(\sigma)\gamma_2(\sigma)d\sigma$$

$$= \int_{\ln a^2}^{\ln 2a^2} (\sigma - \ln(a^2) + 3)\left(\frac{e^\sigma}{a^2}\left(-2 - \pi + 4\arcsin\left(\frac{a}{e^{\sigma/2}}\right)\right) + \frac{4e^\sigma\sqrt{e^\sigma - a^2}}{a^3} - \frac{e^{2\sigma}}{a^4}\right)d\sigma$$

$$= \Bigg[-\frac{1}{a^4}\Bigg(8a^4\arcsin\left(\frac{a}{e^{\sigma/2}}\right) + e^\sigma(\sigma - \ln(a^2) + 2)a^2(\pi + 2)$$
$$-\frac{8a}{9}\sqrt{e^\sigma - a^2}\left(a^2(-3\sigma + 3\ln(a^2) - 1) + e^\sigma(3\sigma - 3\ln(a^2) + 7)\right)$$
$$-4a^2e^\sigma(\sigma - \ln(a^2) + 2)\arcsin\left(\frac{a}{e^{\sigma/2}}\right)$$
$$-\frac{16a^4}{3}\arctan\left(\frac{a}{\sqrt{e^\sigma - a^2}}\right)$$
$$-4a^3\sqrt{e^\sigma - a^2}(\sigma - \ln(a^2))$$
$$+\frac{1}{4}e^{2\sigma}(2\sigma - 2\ln(a^2) + 5)\Bigg)\Bigg]_{\ln(a^2)}^{\ln(2a^2)}$$

$$= -\frac{1}{a^4}\left(-2\pi a^4 + a^4(\pi + 2)(2\ln 2 + 2) + a^4\left(-\frac{8}{9}(3\ln 2 + 13)\right) - a^4\pi(2\ln 2) + a^4\frac{4}{3}\pi - a^4\ln 2 + a^4\left(2\ln 2 + \frac{15}{4}\right)\right)$$

$$= -\frac{4}{3}\pi + \frac{2}{3}\ln 2 + \frac{137}{36}$$

Therefore,

$$\varepsilon = \varepsilon_1 + \varepsilon_2 = 2\pi - \frac{179}{36} - \frac{4}{3}\pi + \frac{2}{3}\ln 2 + \frac{137}{36} = \frac{2}{3}\pi + \frac{2}{3}\ln 2 - \frac{7}{6}$$

6.3 Third step: *E*

Finally, E is given as

$$E = e^{\varepsilon/2} = \exp\left(\frac{1}{3}\pi + \frac{1}{3}\ln 2 - \frac{7}{12}\right) = 2.00353531591730...$$

Thus, the geometric mean is an irrational number, approximately 2.00.

7. Verification with a Monte Carlo simulation

We compare $\kappa(R)$ with frequency distribution, M, from our Monte Carlo simulation. We calculated the distance between two randomly selected points on a square box with the corners at (0,0), (1,0), (1,1) and (0,1), and the distance between two randomly selected points on a line segment between (0,0) and (1,0). We took their ratio. We did this ten million times.

Logically, for a Monte Carlo simulation with n iterations (n is ten million for our case) and an abscissa interval (the bin width of histogram) w,

$$M = \kappa(R)*n*w.$$

The frequency distribution, M_{log}, given against log-equal intervals w_{log} is

$$M_{log} = R\kappa(R)*n*w_{log}.$$

Our verification strategy is to plot the simulation results (M and M_{log}), plot the analytical solution ($\kappa(R)*n*w$ and $R\kappa(R)*n*w_{log}$), and see if these two agree with each other.

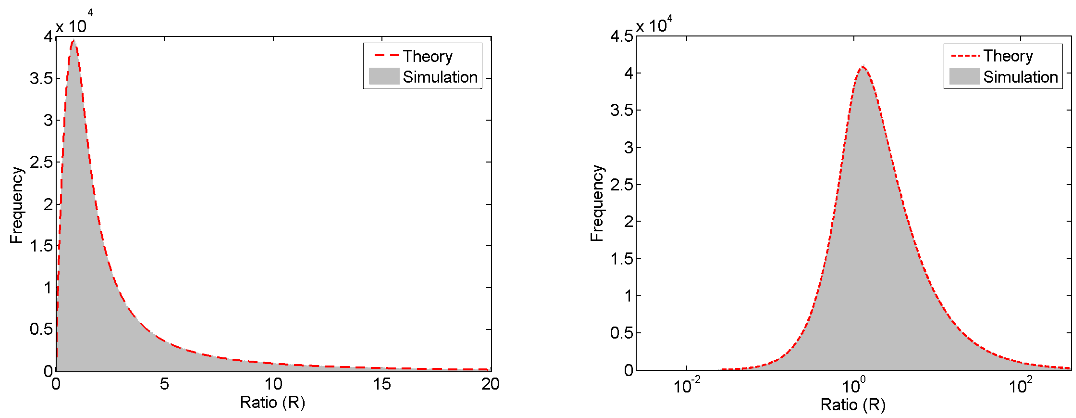

Figure 1. Results of the Monte Carlo simulation and theoretical calculation of the frequency distribution of the distance ratio. The same comparison is shown on a linear (left panel) and logarithmic (right) scale.

They do. These figures also reveal that the distribution is more similar to log-normal (right panel) than to normal (left panel).

As for the geometric mean, our simulation (n=1e7), when repeated 1e5 times, returned a geometric mean of 2.0035…, with a one geometric standard deviation range of 2.0027-2.0043. This is consistent with the analytical solution (2.0035…) and verifies that the solution is indeed different from 2.000.

Taking the arithmetic mean is less preferable because of the evidently skewed distribution of the ratio on a linear scale (left panel). On the arithmetic basis, the mean ratio is ~20 according to our Monte Carlo calculation, high because of a few yet extremely large values when the denominator of the ratio is near 0. Moreover, the ratio of the distance means is significantly different (1.56) from the mean of the ratio (~20), which would raise the question as to which number should be used as the guideline conversion factor if we were to choose the arithmetic method. On the geometric basis, the distribution (on a log scale) looks closer to normal. The mean ratio and the ratio of the distance means are the same (2.0035…).

Reference

J. Philip, The probability distribution of the distance between two random points in a box, unpublished manuscript (2007), TRITA MAT 07 MA 10, http://www.math.kth.se/~johanph/habc.pdf.

Acknowledgements

We thank the anonymous referee #1 for verifying our derivation with a Monte Carlo simulation (independently from our own presented in Section 7). Wolfram Mathematica solved $\eta_{22}(\upsilon)$ and ε_2.

http://integrals.wolfram.com/index.jsp?expr=%28-Exp[x]*Exp[-u]%2Fa^2%2BExp[x%2F2]*Exp[-u%2F2]%2Fa%29%28Exp[x]%2Fa^2*%28-2-p%2B4*arcsin[aExp[-x%2F2]]%29%2B4*Exp[x]*Sqrt[Exp[x]-a^2]%2Fa^3-Exp[2x]%2Fa^4%29

http://integrals.wolfram.com/index.jsp?expr=%28x%2Bq%29%28Exp[x]%2Fa^2*%28-2-p%2B4*arcsin[aExp[-x%2F2]]%29%2B4*Exp[x]*Sqrt[Exp[x]-a^2]%2Fa^3-Exp[2x]%2Fa^4%29