Square line picking

Francesco Zanlungo

December 20, 2012

With my co-authors (Z. Yücel, F. Zanlungo, T. Ikeda, T. Miyashita, N. Hagita, *Modeling Indicators of Coherent Motion*, IROS 2012; and *Deciphering the crowd: Modeling and identification of pedestrian group motion*, submitted to Sensors), we faced the problem of finding the probability distribution for the distance of unrelated pedestrians in a square area. Assuming the pedestrian positions are independent and uniformly distributed on the square, the Wolfram web site gives an analytical solution to the problem
(http://mathworld.wolfram.com/SquareLinePicking.html),
but we couldn't find a demonstration, that we provide in this short note (we nevertheless use Mathematica for a couple of troublesome integrals).

1 Line line picking

Let us start with a simpler problem: finding the probability distribution for the distance between two points on a line of length L, assuming both points are given by a uniform distribution,

$$p(x) = \frac{\chi_L(x)}{L}, \tag{1}$$

where

$$\chi_L(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } 0 \leq x \leq L \\ 0 & \text{if } x > L \end{cases}. \tag{2}$$

Assuming

$$0 \leq r \leq 1, \tag{3}$$

the probability distribution for the distance is then

$$\rho_1(r) = \int_0^L \left(\int_0^L p(x)p(y)\delta(|x-y|-r)dy \right) dx, \tag{4}$$

or

$$\rho_1(r) = \int_0^L \left(\int_0^L p(x)p(y)(\delta(x-y-r)+\delta(y-x-r))dy \right) dx. \tag{5}$$

Integrating on y we get

$$\rho_1(r) = \int_0^L p(x)p(x-r)dx + \int_0^L p(x)p(x+r)dx. \tag{6}$$

From

$$\chi_L(x)\chi_L(x-r) = \begin{cases} 0 & \text{if } x < r \\ 1 & \text{if } r \leq x \leq L \\ 0 & \text{if } x > L \end{cases} \tag{7}$$

and

$$\chi_L(x)\chi_L(x+r) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } 0 \leq x \leq L-r \\ 0 & \text{if } x > L-r \end{cases}, \tag{8}$$

we obtain

$$\rho_1(r) = \frac{1}{L^2}\left(\int_r^L dx + \int_0^{L-r} dx\right) \tag{9}$$

and finally

$$\rho_1(r) = \frac{2L-2r}{L^2} = \frac{2}{L} - \frac{2r}{L^2}. \tag{10}$$

The probability decreases linearly with r. Figure 1 compares the solution (10) to the results of a very rough Montecarlo simulation (using only 10^4 points).

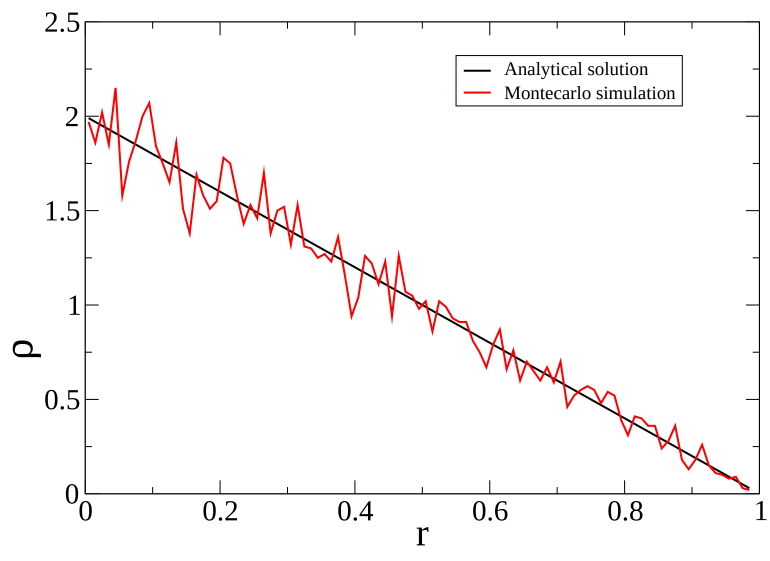

Figure 1

2 Square line picking

Things are harder in two dimensions. Let us assume we pick two points from uniform distributions on a unit square. Let us denote one of our points as (x, y), and the other one as $(x + r\cos(\theta), x + r\sin(\theta))$. The integral on $0 \le \theta < 2\pi$ may be performed by considering the overall length of the portion of the circle of radius r with centre (x, y) that falls inside the square, which will be clearly a function of (x, y, r), and whose integral on x and y will give us the probability distribution we are looking for, $\rho_2(r)$. Actually considering the intersection of the full circle with the four sides would lead us to a lot of troublesome conditions, and for this reason we are going only to consider $\pi \le \theta < 3\pi/2$. Due to the symmetry of the problem, the final result will be obtained multiplying by four. Let us first assume $0 \le r \le 1$. We may divide (figure 2) the square in four areas. In area 1 the quarter of circle under analysis has no intersections with the square, being completely included in it. In areas identified with 2, there is a single intersection, while in 3 we will have two intersections. In area 0 we have no intersections, the circle being completely outside the square. From figure 3, we may see that the length of the circle centred in (x, y) included in the square is $r\phi_2$ were

$$\phi_2 = \frac{\pi}{2} - \phi_1 - \phi_3 = \frac{\pi}{2} - \arccos\left(\frac{x}{r}\right) - \arccos\left(\frac{y}{r}\right). \tag{11}$$

As a result, for $0 \le r \le 1$ we have

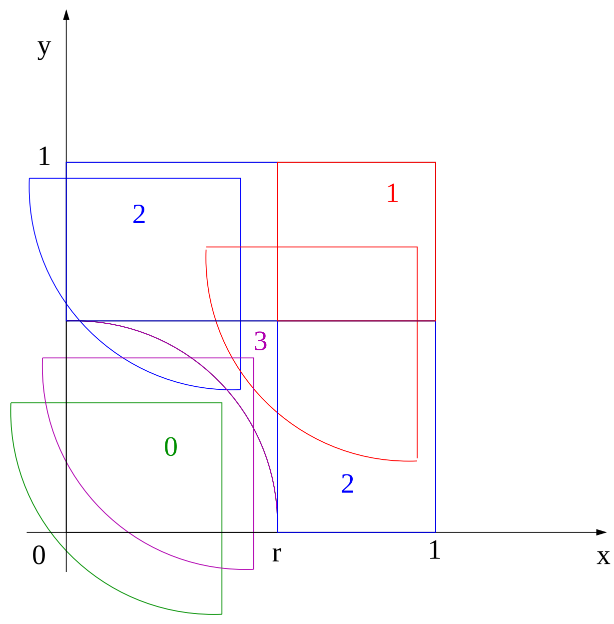

Figure 2

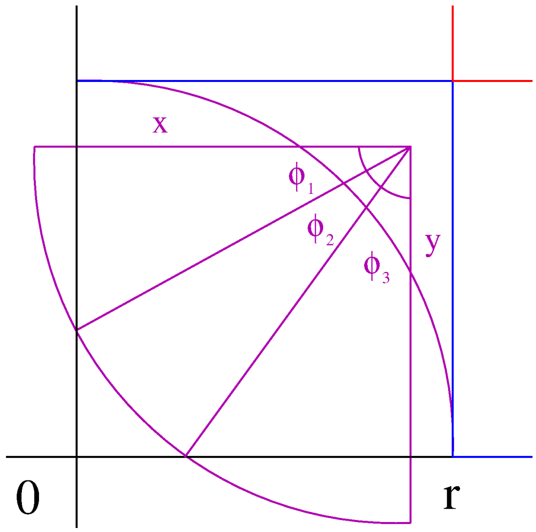

Figure 3

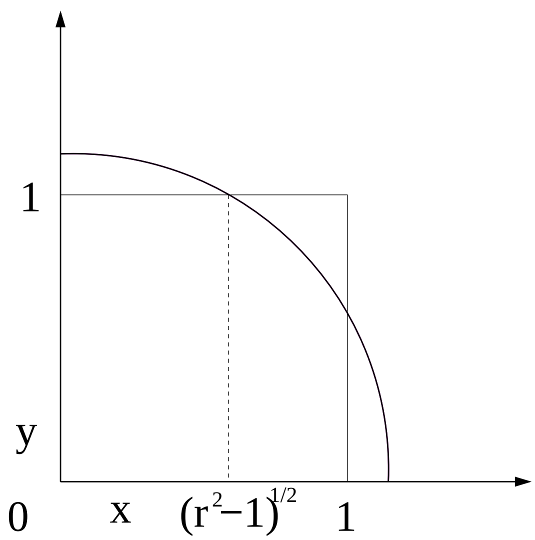

Figure 4

$$\rho_2(r) = I_1(r) - I_2(r) - I_3(r), \tag{12}$$

where

$$I_1(r) \equiv 2r\pi \left(\int_0^r \left(\int_{\sqrt{r^2-x^2}}^1 dy \right) dx + \int_r^1 \left(\int_0^1 dy \right) dx \right), \tag{13}$$

$$I_2(r) \equiv 8r \int_r^1 \left(\int_0^r \operatorname{arcos}\left(\frac{y}{r}\right) dy \right) dx, \tag{14}$$

and

$$I_3(r) \equiv 4r \int_0^r \left(\int_{\sqrt{r^2-x^2}}^r \left(\operatorname{arcos}\left(\frac{y}{r}\right) + \operatorname{arcos}\left(\frac{x}{r}\right) \right) dy \right) dx. \tag{15}$$

From trivial geometric considerations we have

$$I_1(r) = 2\pi r \left(1 - \frac{\pi r^2}{4}\right). \tag{16}$$

Also

$$I_2(r) = 8r \int_r^1 \left[y \operatorname{arcos}\left(\frac{y}{r}\right) - r\sqrt{1 - \frac{y^2}{r^2}} \right]_{y=0}^{y=r} dx, \tag{17}$$

and thus

$$I_2(r) = 8r^2\pi \int_r^1 dx = 8r^2 - 8r^3. \tag{18}$$

We used Mathematica to compute

$$I_3(r) = \left(6 - \frac{\pi^2}{2}\right) r^3 \tag{19}$$

and finally we get

$$\rho_2(r) = 2r(\pi + r^2 - 4r) \quad \text{for } 0 \le r \le 1. \tag{20}$$

Extending the approach for zone 3 to $1 < r \le \sqrt{2}$ (see figure 4) we have

$$\rho_2(r) = I_4(r) = 4r \int_{\sqrt{r^2-1}}^1 \left(\int_{\sqrt{r^2-x^2}}^1 \frac{\pi}{2} - \operatorname{arcos}\left(\frac{y}{r}\right) - \operatorname{arcos}\left(\frac{x}{r}\right) dy \right) dx. \tag{21}$$

Using again Mathematica we have

$$I_4(r) \equiv 2r \left(4\sqrt{r^2-1} - (r^2 + 2 - \pi) - 4 \operatorname{atan}\left(\sqrt{r^2-1}\right) \right) \tag{22}$$

or

$$\rho_2(r) = \begin{cases} 2r(\pi + r^2 - 4r), & \text{if } 0 \le r \le 1, \\ 2r(4\sqrt{r^2-1} - (r^2 + 2 - \pi) - & \\ 4 \operatorname{atan}\left(\sqrt{r^2-1}\right)) & \text{if } 1 < r \le \sqrt{2} \end{cases} \tag{23}$$

Finally we may check this solution numerically. Let us call $M(r)$ the solution obtained using a Montecarlo method, i.e. extracting x_1, y_1, x_2 and y_2 from pseudo-random uniform distributions and computing the distribution of $r = \sqrt{(x_1-x_2)^2 + (y_1-y_2)^2}$; $N(r)$ the solution obtained using a 2D integrator on a grid for integrals $I_1 - I_2 - I_3$ and I_4; and $\rho_2(r)$ the analytical solution (23). To compute $M(r)$ we use 10^9 pairs of points (overall), and to compute $N(r)$ we use 10^6 grid points (for each point evaluation). All functions are evaluated on

$0 \leq r \leq \sqrt{2}$ with a step $\Delta r = 10^{-3}$. The results are shown in figures 5 (absolute value) and 6 (absolute error). For a Montecarlo method using n points, we expect a relative error of order

$$\varepsilon \approx \frac{1}{\sqrt{n}} \tag{24}$$

In figure 7 we plot the relative error of $M(r)$ versus $1/\sqrt{n(r)}$, where $n(r)$ is the number of points used to evaluate $M(r)$ (i.e the number of points falling between r and $r + \Delta r$), finding a very good agreement.

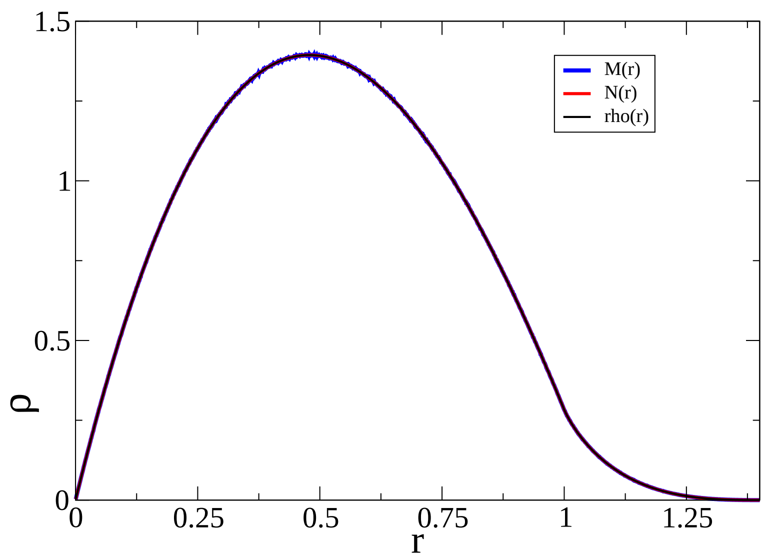

Figure 5

3 Some real pedestrian data

We show also some real world pedestrian data. We tracked *unrelated* pedestrians (i.e. pedestrians not in the same social group) in two square areas of linear size D (see our papers for further details). Scaling the units leads us to the probability distribution

$$\rho_2(r) = \begin{cases} D^{-1} 2r(\pi + r^2 - 4r), & \text{if } 0 \leq r \leq D, \\ D^{-1} 2r(4\sqrt{r^2 - 1} - (r^2 + 2 - \pi) - \\ 4 \text{ atan}(\sqrt{r^2 - 1})) & \text{if } D < r \leq D\sqrt{2} \end{cases} \tag{25}$$

Obviously our formula will be valid for pedestrian positions in the limit $D \gg c$ where c is the linear size of the pedestrian body (the *pedestrian diameter* that we can approximate as $c = 0.4$ meters). Since our tracking areas were of a few

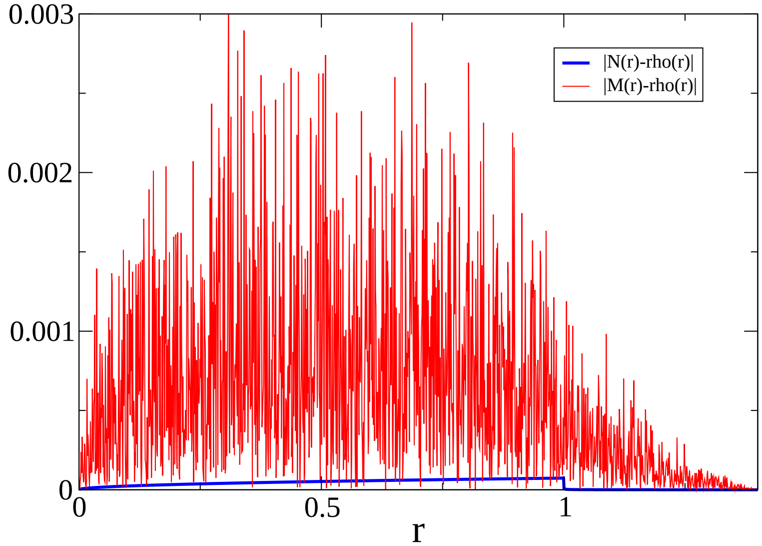

Figure 6

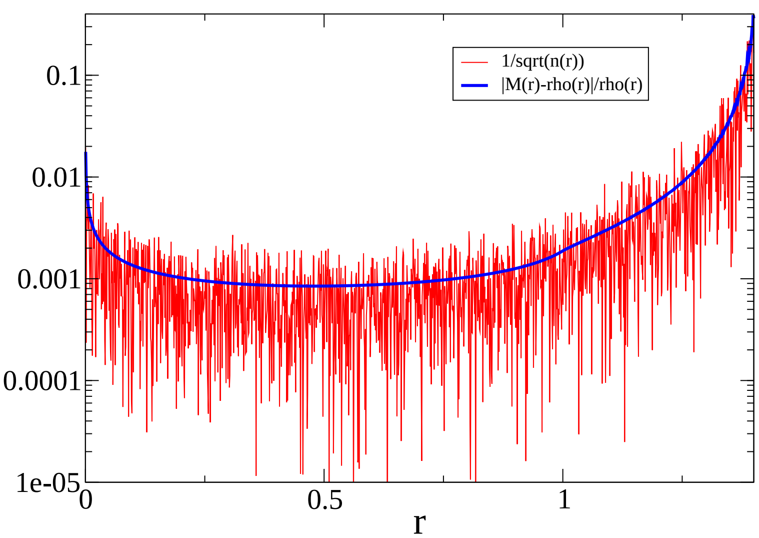

Figure 7

meters, the effects of this cutoff are not big but still visible. For the scope of our papers we just proposed a scaled version of eq. (25) to account for the cutoff, i.e. we substituted r with $r' = r - c$ and D with $D' = D - c/\sqrt{2}$. The results of our tracking are compared to eq. (25), to its scaled model (the *proposed model* of our papers) and to a Gaussian best fit in figures 8 and 9, respectively for the two investigated environments. You can note that the parameter free square line picking formula describes qualitatively better the data than the Gaussian approach. This result is attained even if in the latter approach two parameters (μ and σ) had to be calibrated and thus the oscillations due to the small size samples could be "over-fitted". A rigorous derivation of a formula dealing

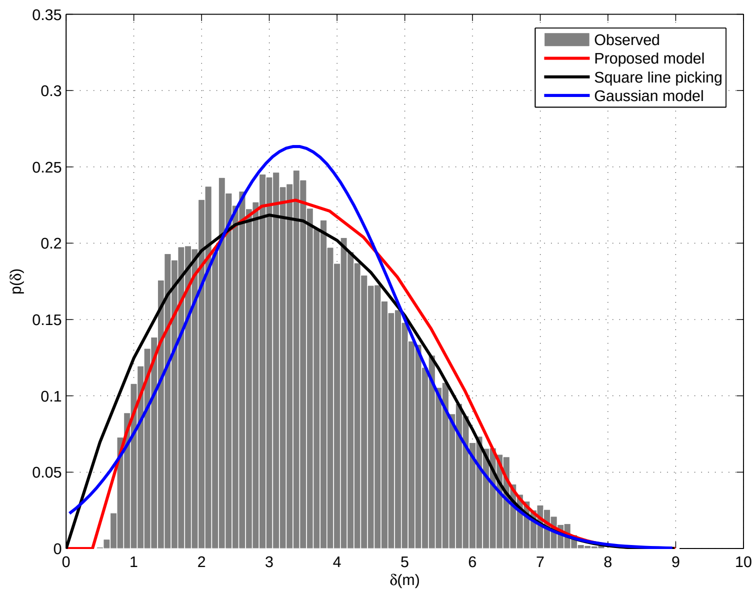

Figure 8

with with the cutoff c will imply coping with the non-trivial collision avoiding behaviour around $r \approx c$, which is more complex than just picking from uniform distributions. For the scope of our paper we needed a formula that could deal with the cutoff in order to avoid overestimation of the likelihood for unrelated pedestrians with respect to pedestrians in groups for $r \approx c$, and the proposed scaling worked fine. We also note that while in our papers we used square tracking areas, Montecarlo integration can be used to obtain the equivalent of (25) for any geometrical shape.

4 Acknowledgements

We thank Thomas Kaczmarek for useful discussion.

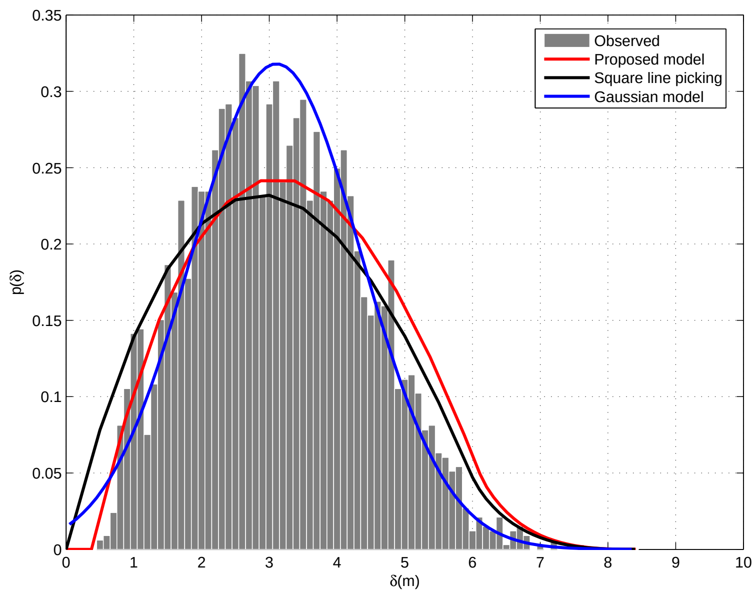
Observed
Proposed model
Square line picking
Gaussian model
p(δ)
δ(m)

Figure 9