Chord length at discs with random centre and radius

Martin A.J. van Montfort* and Albert Otten

Abstract

A way of collecting information on the size of objects in the 2-dimensional space consists of studying what is going on on a randomly chosen line (e.g. line transect sampling). Here the purpose is: inference on the distribution of the radii of discs randomly located in a 2-dimensional space; the inference has to be based on measured lengths of intersections (chord lengths) with a fixed chosen line. The paper presents results of some modelling, simulation and estimation to link the distribution of chord lengths at output with the distribution of radii at input.

1 Problem

The real life problem, which gave rise to this paper, deals with quantifying eddies in a river, where the eddies are modelled by cylinders each with its own radius. The size of eddies is important in the study of turbulence, river bank protection, sediment entrainment and flow resistance. It is also of fundamental interest to explain natural phenomena. The distribution of the radii is the main point of interest. The measuring technique, however, yields chord lengths in stead of radii. More precisely, at a fixed point in the middle of the river, the passage of an eddy is detected by flow meters and the passage time is recorded. Multiplication by speed of river flow gives the chord length of the intersection of eddy and central line in the river. If a picture from above were available one could measure the radii directly, but here we only have chord length available of eddies intersecting the central line and passing the flow meters within the measurement period. Eddies disjoint with the central line won't be observed at all.

A variant of this problem in a higher dimension deals with inference on the distribution of the radii of the spherical holes in a cheese to be based on the observed circles on an intersecting plane. Note that large holes have a higher probability to be hit by a plane than small ones.

In Fig. 1 the cross section of the river is scaled from -1 to 1. An eddy is just half way passing the cross section. In practise radii are small compared to river width, and eddies seem to be randomly distributed (both along and across the river). Hence border effects are neglected. The notation and statistical assumptions are as follows (with random variables underlined).

*Department of Mathematics, Wageningen Agricultural University, Dreijenlaan 4, 6703 HA Wageningen, the Netherlands. Tel.: +31 317 484085. Fax: +31 317 483554. E-mail: montfort@rcl.wau.nl

Notation: let

- m be the position of the centre of circular disc,
- r be the radius of the disc,
- h be halve of the chord length, if any.

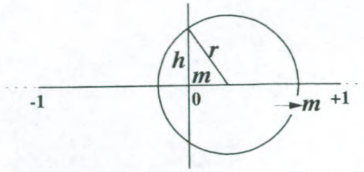

Figure 1: Graphical representation of m, r, h (see text).

Assumptions:

- $\underline{m}$ has a Uniform distribution on $(0,1)$ (this restriction gives the same results as Uniform on $(-1,1)$ but simplifies integrals),
- $\underline{r}$ is independent of $\underline{m}$, with a distribution on $(0,1)$ with cdf $\Phi(r)$ and pdf $\phi(r)$,
- different discs behave independently.

Then the (deficient) distribution of $\underline{h}$ could be derived, which distribution depends on the type and parameters of Φ.

The point of interest is inference on $\underline{r}$ based on h-data. The general solution is presented, and as a special case the solution for truncated Exponential input is given.

2 Deficient distribution of chord halves

From Fig. 1 we see that in case of intersection

$$0 \leq m, r, h \leq 1, \quad m \leq r \quad \text{and} \quad h^2 = r^2 - m^2 .$$

From the definition $\Phi(r) = P[\underline{r} \leq r]$ follows $\Phi(0) = 0$ and $\Phi(1) = 1$. For given h, a halve chord length of at least h will be produced by discs for which $m \leq \sqrt{1-h^2}$ ($= m(h)$ from now on) and $r \geq \sqrt{h^2+m^2}$. Figure 2 shows the region of integration over radius and distance to obtain the probability $P[\underline{h} \geq h]$. After integrating over r at given m (bold line segment) this probability is obtained as integral over m:

$$\begin{aligned} P[\underline{h} \geq h] &= \int_0^{m(h)} P[\sqrt{h^2+m^2} \leq \underline{r} \leq 1] \mathrm{d}m \\ &= \int_0^{m(h)} \{1 - \Phi(\sqrt{h^2+m^2})\} \mathrm{d}m \end{aligned} \tag{1}$$

and the probability that a chord can be observed on a passing eddy is

$$P[\text{obs}] = P[\underline{h} \geq 0] = \int_0^1 \{1 - \Phi(m)\} \mathrm{d}m . \tag{2}$$

Figure 3 shows a possible shape of the (deficient) distribution of $\underline{h}$.

The right hand side of (2) is known to be equal to $\mathrm{E}(\underline{r})$, hence $P[\text{obs}] = \mathrm{E}(\underline{r})$. For instance, in case of truncated Exponential radii with $\Phi(r) = (1-\mathrm{e}^{-\alpha r})/(1-\mathrm{e}^{-\alpha})$, we get $P[\text{obs}] = 1/\alpha - 1/(\mathrm{e}^{\alpha} - 1)$.

Based on (1) and (2) we get for the conditional distribution of $\underline{h}$, given an intersection:

$$\begin{aligned} P[\underline{h} \geq h | \text{obs}] &= P[\underline{h} \geq h] / P[\text{obs}] \\ &= \int_0^{m(h)} \{1 - \Phi(\sqrt{h^2+m^2})\} \mathrm{d}m \,/ \int_0^1 \{1 - \Phi(m)\} \mathrm{d}m . \end{aligned} \tag{3}$$

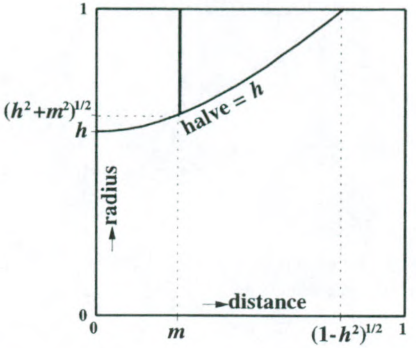

Figure 2: Region of integration over radius and distance.

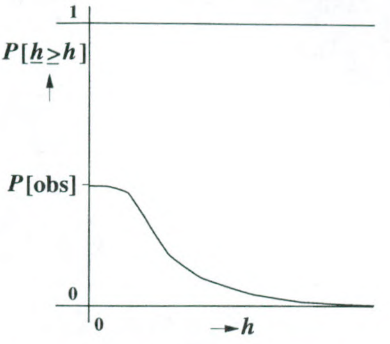

Figure 3: Deficient complementary cdf of the random variable $\underline{h}$.

Let $f(h)$ be the conditional probability density function (pdf) of $\underline{h}$; then

$$f(h) = -\frac{\mathrm{d}}{\mathrm{d}h}P[\underline{h} \geq h|\mathrm{obs}] = -\frac{1}{P[\mathrm{obs}]} \cdot \frac{\mathrm{d}}{\mathrm{d}h}P[\underline{h} \geq h]$$

$$\frac{\mathrm{d}}{\mathrm{d}h}P[\underline{h} \geq h] = m'(h)\{1 - \Phi(1)\} + \int_0^{m(h)} \frac{\mathrm{d}}{\mathrm{d}h}\{1 - \Phi(\sqrt{h^2 + m^2})\}\mathrm{d}m\,.$$

The first term vanishes since $\Phi(1) = 1$. The final result is

$$f(h) = \frac{1}{P[\mathrm{obs}]} \cdot \int_0^{m(h)} \frac{h}{\sqrt{h^2 + m^2}}\phi(\sqrt{h^2 + m^2})\mathrm{d}m\,. \tag{4}$$

Note that P[obs] behaves as a factor only depending on the parameter(s) of $\underline{r}$. Usually numerical integration is needed to give cdf (3) and pdf (4) of $\underline{h}$ for a specified distribution of radii.

In a rough approximation the density could be written proportional to $f(h) = h\phi(h)$. This may be seen by writing

$$f(h) = \frac{1}{P[\mathrm{obs}]} \cdot \int_0^{m(h)/h} \frac{h}{\sqrt{1 + t^2}}\phi(h\sqrt{1 + t^2})\mathrm{d}t\,. \tag{5}$$

and recognising that the integral kernal $1/\sqrt{1 + t^2}$ acts more or less like a delta function at $t = 0$. (Another argument for the approximation is that discs are sampled with probability proportional to size; the conditional radii density given intersection is *exactly* a factor times $r\phi(r)$. This density is diffused when chords in stead of radii are measured.) This approximation to the conditional distribution of $\underline{h}$ turns out to be especially handy in case of Gamma or Beta distributed radii. In fact, this is the main reason to consider this approximation. In that case the approximate distribution of observed chord halves is of the same type, with one of the parameters incremented by one. The quality of the approximation depends on the parameters (shape) of ϕ and could be poor in some cases.

3 Approximations to the distribution of observed chord halves

3.1 Theoretical remarks

For $\underline{r}$ having a truncated Exponential distribution on $(0, 1)$, so for

$$\Phi(r) = \{1 - \exp(-r/\beta)\}/\{1 - \exp(-1/\beta)\} \quad \text{and} \quad \phi(r) = \text{factor} \times \exp(-r/\beta),$$

we get

$$f(h) \approx \text{factor} \times h \cdot \exp(-h/\beta)$$

with proportionality factors only depending on β. This approximation is in close relation to the density of a gamma-variate with support $(0, \infty)$ and $\alpha \approx 2$:

$$h^{\alpha-1} \exp(-h/\beta)/\beta^{\alpha}\Gamma(\alpha)$$

with a single mode at $\beta(\alpha - 1)$ when $\alpha > 1$.

For $\underline{r}$ having a Beta-distribution on $(0, 1)$, so for

$$\Phi(r) = \int_0^r \phi(t)\mathrm{d}t \quad \text{with} \quad \phi(t) = \frac{t^{p-1}(1-t)^{q-1}}{\Gamma(p)\Gamma(q)/\Gamma(p+q)},$$

we get

$$f(h) \approx \text{factor} \times h^{(p-1)+1}(1-h)^{q-1}$$

indicating a Beta-distribution with an increased parameter p which regulates the behaviour at small h-values and a non changed parameter q regulating the right tail.

In both situations, truncated Exponential and Beta, the behaviour of $\underline{h}$ in the right tail follows the behaviour of $\underline{r}$ in the right tail.

3.2 Support by simulation

A simulation study with always an output of 1000 h-values was performed with two types for the distribution Φ of $\underline{r}$ at input: (1) an at the right side truncated Exponential distribution with expectation β before truncation, and (2) a Beta-distribution with parameters (p, q). The empirical h-distributions at output showed close agreement with the gamma- resp. beta-distributions with ML-fitted parameters.

Table 1: Results at Exponential input of radii with parameter β and assumed gamma output of $\underline{h}$ with estimated parameters $\hat{\beta}, \hat{\alpha}$.

In	$-\log_2(\beta)$	2.5	3	3.5	4	4.5	5
	β	0.177	0.125	0.088	0.063	0.044	0.031
Out	$\hat{\beta}$	0.155	0.117	0.085	0.060	0.041	0.029
	$\hat{\alpha}$	1.71	1.71	1.66	1.64	1.67	1.69

From Table 1 we see that the approximation works well with respect to tailing off in the right tail ($\hat{\beta} \approx \beta$), but α turns out to be somewhat smaller than 2.

From Table 2 we also see that the approximation works well with respect to tailing off in the right tail ($\hat{q}/q \approx 1$); however, the approximation regarding p fails.

Table 2: Result at Beta-input of radii (parameters: p, q) and assumed Beta-output of $\underline{h}$ (estimated parameters $\hat{p}, \hat{q}$). Top of the table: $\hat{p}$; bottom of the table: $\hat{q}/q$.

		q_{In} 2	4	8	16	32	64	Mid mean	
p_{In}	0.5	1.21	1.25	1.41	1.24	1.29	1.35	1.27	$\}\hat{p}$
	1	1.77	1.62	1.52	1.69	1.60	1.64	1.63	
	2	2.16	2.19	2.30	2.06	2.32	2.22	2.21	
	4	2.86	2.69	3.09	2.98	2.94	3.28	2.96	
	0.5	1.15	1.13	1.17	1.05	1.11	1.12	1.12	$\}\hat{q}/q$
	1	1.33	1.13	0.98	1.11	1.05	1.06	1.08	
	2	1.20	1.08	1.06	0.94	1.01	0.94	1.03	
	4	1.11	0.89	0.86	0.81	0.80	0.84	0.85	

3.3 Qualitative remark on estimating the radii distribution

A heavier left tail of the distribution of $\underline{r}$ only highers the fraction of radii, which do not cause an intersection (so no h-value at all), and nearly does not change the distribution of the h-values to show up.

So h-data nearly do not contain information on the left tail of the distribution of $\underline{r}$; or, stated the other way around, h-data are informative on the right tail of the distribution of $\underline{r}$.

Figure 4 illustrates this feature for two input radii densities which differ primarily in the left tail.

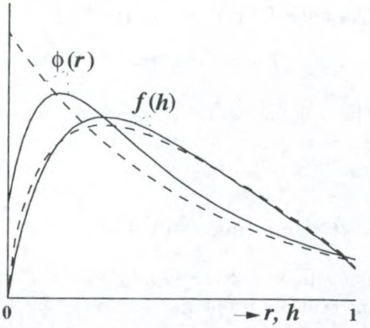

Figure 4: Two situations (solid and dashed) of probability density function $\phi(r)$ of $\underline{r}$ at input and $f(h)$ of $\underline{h}$ at output.

4 ML-estimation of the distribution of radii

The most straightforward technique to estimate the parameters of the radii-distribution Φ given the type of Φ (e.g. truncated Exponential) is the Maximum Likelihood method (ML) in its most crude form. This means: maximise $L = \sum \ln f(h_i)$, see (4), with respect to the parameter(s) of Φ. Because of the intractability of $f(h)$ in (4) numerical integration of $f(h)$ is required to get L at fixed parameter(s), followed by numerical maximisation of L with respect to the parameters. This procedure was applied to the truncated Exponential model for radii. An additional problem is that the upper bound of the support of $\underline{r}$ is not known in the real life data set. Note that the supports of $\underline{r}$ and $\underline{h}$ are the same (with zero being the lower bound). Here the upper bound of the support of $\underline{r}$ was chosen to be a multiple of the maximum h-value (multipliers chosen: 1.1 (0.1) 1.5), and hopefully this choice will nearly not influence the resulting estimate of $\Phi(r)$. The real life data set, Table 3, contains length of chord halves up to a factor. A Q-Q-plot of the data (not

Table 3: Frequency distribution of observed chord halves.

value (h)	1	2	3	4	5	6	7	8	9	10
frequency	150	182	92	40	21	9	5	3	1	3

shown) supports an Exponential right tail. Here $h_{\text{max}} = 10$ and the upperbound was chosen to be $10 \times 1.1, \cdots, 10 \times 1.5$. ML-estimates for α are presented in Table 4.

Table 4: ML-estimates of α and $P[\underline{r} > 5.5]$ for truncated Exponential Φ.

multiplier	1.1	1.2	1.3	1.4	1.5
$\hat{\alpha}$	7.04	7.75	8.44	9.12	9.79
$\hat{P}[\underline{r} > 5.5]$	0.029		0.028		0.027

The radii-distribution in the length unit of h reads

$$P[\underline{r} \leq r] = (1 - e^{-\alpha \tilde{r}})/(1 - e^{-\alpha}) \quad \text{with} \quad \tilde{r} = r/(\text{multiplier} \times h_{\text{max}})\,.$$

The last line of Table 4 presents an estimate of a tail probability at different choices of the multiplier; the arbitrarily chosen multiplier seems to have nearly no influence on the resulting tail probability.

5 Concluding remarks

The equality of the supports for the distribution of radii and of half chords is essential in the real life problem. The fact that the chords are not measured as length complicates the problem and gives rise to additional research. No radii distributions seem to exist which give rise to a simple distribution of the chord length with simple estimation of the parameters. The same problem arises in three dimensions. Think of holes in an infinitely large cheese, where the holes are modelled to be spheres with random radii. The intersection with a randomly chosen plane shows circles with radii to be measured. So the measured radii of the observed circles have to tell the story about the distribution of the radii of the holes (to estimate the fraction non cheese in a cheese). Note that the relative area of the circles is an unbiased estimator for the fraction of contents of the holes, without possibilities for inference on the distribution of the radii of the holes. Theory on what is happening in a 3-dimensional space by measuring in a 2- or 1-dimensional subspace could be found in literature on stereology. For a more general context the literature on PPS (Probabilities Proportional to Size) sampling could be consulted.

Dr. G.M. Smart (National Institute for Water and Atmosphere at Christchurch (NZ)) asked attention for these problems during a sabbatical stay of Van Montfort (August 1996).

Ontvangen: 04-11-1996
Geaccepteerd: 25-07-1997