ICSI/APHY451 Bayesian Data Analysis
December 4, 2020
Dylan VanAllen

Radius of a circle

Abstract

In this project I have determined the radius of a circle using multiple different techniques and comparing them. The first technique was a simple ruler measurement. The second and third techniques used a scattering method combined with Bayesian analysis and a probability density function. It is not always possible to find an analytical solution for the mean of a probability density function, but in this case it is. I have first analytically solved for the mean with associated uncertainty, and then I have used John Skilling's algorithm for nested sampling to determine a mean and uncertainty. While a nested sampling algorithm may be considered overkill for an experiment such as this, it is a great way to implement multiple measuring (and analysis) techniques and compare the precision.

Method

1. Print out (or draw precisely) multiple equally sized circles on a sheet of paper.
    a. It might be fruitful to use a sticky piece of paper, but it is not necessary.
2. Measure the radius (or diameter) of the circles using a ruler or meter stick and keep note of the radius and uncertainty. The uncertainty is one half the smallest increment on the ruler.
3. Lay the paper flat on a surface, and drop grains of rice, one at a time, onto the paper. Rice grains that land off the paper do not count. Only count the total number of grains that landed on the paper and the total number of grains that landed on the paper and inside a circle.

4. Construct a probability density function for $P(R \mid A, N, M, K, I)$, where R is the radius of the circles, A is the area of the paper, N is the number of rice grains that landed on the paper, M is the number of rice grains that landed in a circle, K is the number of circles on the paper, and I is prior information on the problem.
5. Analytically and numerically (using nested sampling) solve for the mean and uncertainty of the radius from the probability density function.

Results

Step 2:

Circle #	Axis of measurement	Boundary	diameter (in)	radius (in)	σ (in)
2	x	outside edge	1.75	0.875	0.0625
2	y	outside edge	1.75	0.875	0.0625
7	x	inside edge	1.68	0.84	0.0625
7	y	inside edge	1.68	0.84	0.0625
11	x	in the middle	1.77	0.885	0.0625
11	y	in the middle	1.78	0.89	0.0625

Applying the equations for the mean and uncertainty:

$$\bar{r} = \sum_{i=1}^{N} \frac{r_i}{N} = 0.87 \text{ in.} \quad \sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (r_i - \bar{r})^2} = 0.020 \text{ in.}$$

Step 3: The results are below. N is the number of grains that landed on the paper.

N	in circle?	N	in circle?	N	in circle?
1	0	18	1	35	0
2	0	19	0	36	1
3	1	20	0	37	0
4	0	21	1	38	1
5	1	22	1	39	1

6	0	23	1	40	0
7	0	24	0	41	0
8	1	25	0	42	0
9	0	26	0	43	0
10	0	27	1	44	1
11	0	28	0	45	0
12	1	29	0	46	1
13	1	30	0	47	0
14	0	31	1	48	0
15	0	32	0	49	1
16	0	33	1	50	0
17	0	34	0	Total in circles (M)	18

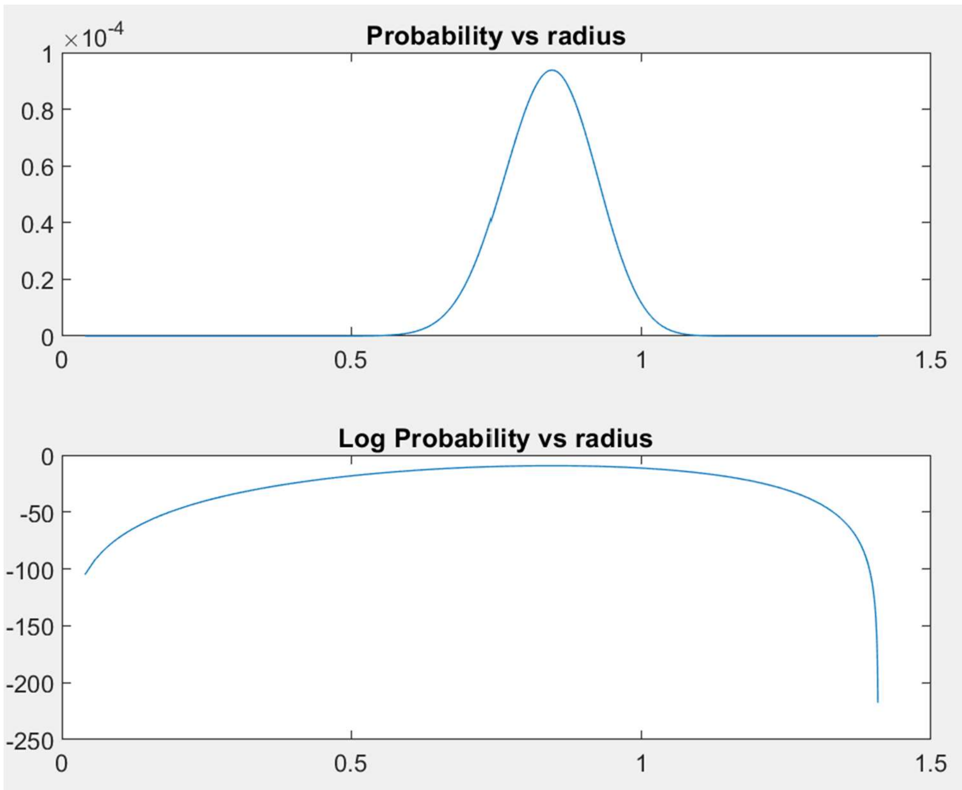
Step 4.

Our goal is to construct a probability density function for $P(R\,|A, N, M, K, I)$. The probability density function for each radius is related to the corresponding area, which is proportional to the probability of M, the number of grains that land inside the circles. This is because the larger the circles, the more space they take up on the sheet. To get a probability function for each M, you count the number of ways to land M rice grains in circles out of M trials and divide by the number of total results. The probability of M rice grains landing in K circles given N trials becomes

$$P(M \mid A, N, R, K, I) = \binom{N}{M} P(grain\ lands\ in\ any\ circle)^{M} P(grain\ lands\ in\ no\ circle)^{N-M}$$

Where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, and the probability of a grain landing in any circle is

$P(C \mid A, N, R, K, I) = \frac{Total\ space\ on\ paper\ inside\ circles}{Total\ area\ of\ paper} = \frac{\pi R^2 K}{A}$ and the probability of it landing outside a circle is $1 - P(C \mid A, N, R, K, I) = \frac{A - \pi R^2 K}{A}$. If choosing a uniform area (as per John Skilling's suggestion), with increments of 0.005, growing no larger than $\frac{Area\ of\ the\ paper}{Number\ of\ circles}$, each radius has a prior probability of $\frac{1}{1246}$, and a posterior probability of $P(R\ |A, N, M, K, I) = \frac{1}{1246}\left(\frac{N!}{M!(N-M)!}\right)\left(\frac{\pi R^2 K}{A}\right)^M \left(\frac{A-\pi R^2 K}{A}\right)^{N-M}$. I have graphed this, as well as the log probability here:



If using nested and stochastic prior sampling, the probability density function is

$P(R \mid A, N, M, K, I) = \frac{1}{number\ of\ prior\ objects}\left(\frac{N!}{M!(N-M)!}\right)\left(\frac{\pi R^2 K}{A}\right)^M \left(\frac{A-\pi R^2 K}{A}\right)^{N-M}$. I made a mistake on this before presentation and have fixed it since. Previously, I was not stochastically setting the prior objects.

Step 5.

An analytical solution of the peak of this probability distribution is available. Setting $\frac{d}{dR}\left(\ln\left(\frac{1}{1246}\left(\frac{N!}{M!(N-M)!}\right)\left(\frac{\pi R^2 K}{A}\right)^M\left(\frac{A-\pi R^2 K}{A}\right)^{N-M}\right)\right) = 0$, and applying the following steps, the analytical solution for the maximum (peak) is found.

$$0 = \frac{d}{dR} lnP = \frac{d}{dR}\left(ln\left(\frac{N!}{M!(N-M)!}\cdot\left(\frac{\pi R^2 K}{A}\right)^M\cdot\left(\frac{A-\pi R^2 K}{A}\right)^{N-M}\right)\right)$$
$$0 = \frac{d}{dR}\left(ln\left(\frac{N!}{M!(N-M)!}\right) + M\cdot ln\left(\frac{\pi R^2 K}{A}\right) + (N-M)\cdot ln\left(\frac{A-\pi R^2 K}{A}\right)\right)$$
$$0 = \frac{d}{dR}\left(M\cdot ln\left(\frac{\pi R^2 K}{A}\right)\right) + \frac{d}{dR} ln\left((N-M)\cdot ln\left(\frac{A-\pi R^2 K}{A}\right)\right)$$
$$0 = \frac{2M}{R} + \frac{(N-M)K\pi R}{K\pi R^2 - A}$$
$$(M-N)K\pi R^2 = 2M(k\pi R^2 - A)$$
$$NK\pi R^2 = MA$$
$$\bar{R} = \sqrt{\frac{MA}{KN\pi}} = 0.84$$

To solve for the uncertainty analytically the following steps were taken.

$$\sigma = \sqrt{-\frac{1}{\frac{\partial^2 logP}{\partial R^2}|_{\bar{R}}}}$$
$$\sigma = \sqrt{-\frac{1}{\frac{\partial}{\partial R}\left(\frac{2M}{R} + \frac{(N-M)K\pi R}{K\pi R^2 - A}\right)|_{\bar{R}}}}$$
$$\sigma = \sqrt{-\frac{1}{\left(\frac{-2M}{\bar{R}^2} + \frac{2K\pi(M-N)}{A-K\pi\bar{R}^2} + \frac{4K^2\pi^2\bar{R}^2(M-N)}{(A-K\pi\bar{R}^2)^2}\right)}}$$
$$\sigma = \sqrt{-\frac{1}{-123.15}} = 0.090$$

To find a numerical solution, I used John Skilling's nested sampling algorithm. Since I used 100 simulation objects the prior was $\frac{1}{100}$, and I stochastically chose areas with uniform likelihood to be sampled. I chose a step size of 0.005 and 1000 iterations. The results of four successive runs of the program are as follows.

```
# iterates = 1000
Evidence: ln(Z) = -11.0536 +- 0.118971
Information: H = 1.41541 nats = 2.04201 bits
mean(r) = 0.854186, stddev(r) = 0.0844955
```

```
# iterates = 1000
Evidence: ln(Z) = -10.5969 +- 0.105241
Information: H = 1.10756 nats = 1.59787 bits
mean(r) = 0.847024, stddev(r) = 0.0735772
```

```
# iterates = 1000
Evidence: ln(Z) = -10.8948 +- 0.115982
Information: H = 1.34518 nats = 1.94068 bits
mean(r) = 0.831808, stddev(r) = 0.078291
```

```
# iterates = 1000
Evidence: ln(Z) = -11.0517 +- 0.120152
Information: H = 1.44365 nats = 2.08275 bits
mean(r) = 0.857973, stddev(r) = 0.0813785
```

Conclusion

The precision on the experiment from the ruler, as expected, is the highest. With just 6 measurements I had lower uncertainty than the analytical Bayesian solution even using such a high number of possible areas as 1246, and 50 rice grains on the paper. It was also lower than the uncertainty given from 1000 iterations of nested sampling. The numerical nested sampling solution interestingly had lower uncertainty than the analytical solution, which tells me that the algorithm is very powerful. However much better the ruler was, it was interesting to see that both Bayesian solutions with even such a relatively low sample size as 50, was so close in predicting the radius of the circle to the solution as measured by the ruler. I suspect that with a larger N, the values for the mean radius as predicted using Bayesian methods and the scattering experiment would be closer and closer to that of the ruler measurements. It was also very helpful to hear the comments of John Skilling. He mentioned to me that the area is what is uniformly likely, not the radius.

References

"Data Analysis: A Bayesian Tutorial", Sivia and Skilling 2006