PoCS Assignment 2

P. Alexander Burnham

September 11, 2018

PoCS Assignment github repo: https://github.com/alexburn17/BurnhamPoCS

Worked with: Yu, Kewang, Ryan, Sam, Luke and Chris

Problem 1:

a)

Step 1: Set up the problem in terms of probability theory

$$Pr(X \leq x) = \int_a^x P(x)dx$$

$$P(x) \propto x^{-\gamma a}$$

$$P(x) = cx^{-\gamma}$$

Step 3: set up the integration

$$\int_a^b P(x)dx = 1$$

Step 4: simplify

$$c\int_a^b x^{-\gamma}dx = 1$$

Step 5: find antiderivative and substitute in a and be into the upper and lower limits

$$c\frac{x^{-\gamma+1}}{-\gamma+1}|_b^a = 1$$

Step 6:

$$c\frac{b^{-\gamma+1}}{-\gamma+1} - c\frac{a^{-\gamma+1}}{-\gamma+1} = 1$$

Step 6: Simplify

$$c\frac{-\gamma+1}{b^{1-\gamma} - a^{1-\gamma}} = 1$$

Step 6: solve for C

$$c = \frac{\gamma - 1}{b^{1-\gamma} - a^{1-\gamma}}$$

b)

We assume $\gamma > 1$ as $\gamma < 1$ yields an infiniate area under the curve and as this is a probability distribution, it must sum to 1.

Problem 2:

Step 1: Integrate p(x) and x^n to find the n^{th} moment

$$P(x) = x^n c x^{-\gamma}$$

$$\int_a^b x^n P(x) dx$$

Step 2:

$$\int_a^b x^n c x^{-\gamma} dx$$

Step 3: Simplify and integrate

$$c \int_a^b x^{n-\gamma} dx$$

Step 4: substitute in a and b

$$c \frac{x^{n-\gamma+1}}{n-\gamma+1} |_a^b$$

Step 5: Simplify

$$c \frac{b^{n-\gamma+1}}{n-\gamma+1} - c \frac{a^{n-\gamma+1}}{n-\gamma+1}$$

Step 5: simplify n^{th} moment and add in expression for c

$$c \frac{b^{1-\gamma+n} - a^{1-\gamma+n}}{1-\gamma+n} * \frac{\gamma-1}{b^{1-\gamma} - a^{1-\gamma}}$$

Problem 3:

Step 1: take the limit as b goes to infinity

$$\lim_{b \to \infty} \frac{b^{1-\gamma+n} - a^{1-\gamma+n}}{1-\gamma+n} * \frac{\gamma-1}{b^{1-\gamma} - a^{1-\gamma}}$$

Step 2: the b terms go to infinity and are removed

$$\frac{a^{1-\gamma+n}}{1-\gamma+n} * \frac{\gamma-1}{-a^{1-\gamma}}$$

Step 3: a terms cancel

$$\frac{(\gamma-1)a^n}{\gamma-n-1}$$

This means that when b goes to infinity, $\gamma > n+1$ and a dominates

Problem 4:

$$\frac{b^{1-\gamma+n}-a^{1-\gamma+n}}{1-\gamma+n} * \frac{\gamma-1}{b^{1-\gamma}-a^{1-\gamma}}$$

When $a \ll b$, b dominates and $\gamma \leq n+1$

Problem 5:

a)

Step 1: (E=expectation)

$$var(x) = E[(x)-(E(x))^2]$$

$$\sigma^2 = E(x^2)-(E(x))^2$$

Step 2: substitute in the n^{th} moment in for x and substitute in the appropriate exponent for n and simplify

$$\sigma^2 = \frac{b^{3-\gamma}-a^{3-\gamma}}{3-\gamma} * \frac{1-\gamma}{b^{1-\gamma}-a^{1-\gamma}} - \left(\frac{b^{2-\gamma}-a^{2-\gamma}}{2-\gamma} * \frac{1-\gamma}{b^{1-\gamma}-a^{1-\gamma}}\right)^2$$

Step 3: Take the limit as b approaches infinity

$$\lim_{b\to\infty} \frac{b^{3-\gamma}-a^{3-\gamma}}{3-\gamma} * \frac{1-\gamma}{b^{1-\gamma}-a^{1-\gamma}} - \left(\frac{b^{2-\gamma}-a^{2-\gamma}}{2-\gamma} * \frac{1-\gamma}{b^{1-\gamma}-a^{1-\gamma}}\right)^2$$

Step 4: b terms are removed and a terms simplify

$$\sigma^2 = \frac{-a^{3-\gamma}}{3-\gamma} * \frac{1-\gamma}{-a^{1-\gamma}} - \left(\frac{-a^{2-\gamma}}{2-\gamma} * \frac{1-\gamma}{-a^{1-\gamma}}\right)^2$$

$$\sigma^2 = a^2\frac{1-\gamma}{3-\gamma} - a^2\frac{(1-\gamma)^2}{(2-\gamma)^2}$$

Step 5: factor a^2 out and cross multiply

$$\sigma^2 = a^2\left(\frac{(1-\gamma)(2-\gamma)^2}{(3-\gamma)(2-\gamma)^2} - \frac{(3-\gamma)(1-\gamma)^2}{(3-\gamma)(2-\gamma)^2}\right)$$

Step 6: square, distribute, subtract and cancel the numerator

$$(1-\gamma)(4-4\gamma+\gamma^2)-(3-\gamma)(1-2\gamma+\gamma^2)$$

$$4-4\gamma+\gamma^2-4\gamma+4\gamma^2-\gamma^3-(3-6\gamma+3\gamma^2-\gamma+2\gamma^2-\gamma^3)$$

$$numerator = 1 - \gamma$$

$$Yields : \sigma^2 = \frac{1-\gamma}{(3-\gamma)(2-\gamma)^2} a^2$$

Step 6: multiply both sides of equation by negative 1 to get in the form described in the instructions:

$$\sigma^2 = \frac{\gamma - 1}{(\gamma-3)(\gamma-2)^2} a^2$$

- $c_1 = 1$
- $c_2 = 3$
- $c_3 = 2$

Step 6: Take the square root of both sides to get the standard deviation σ

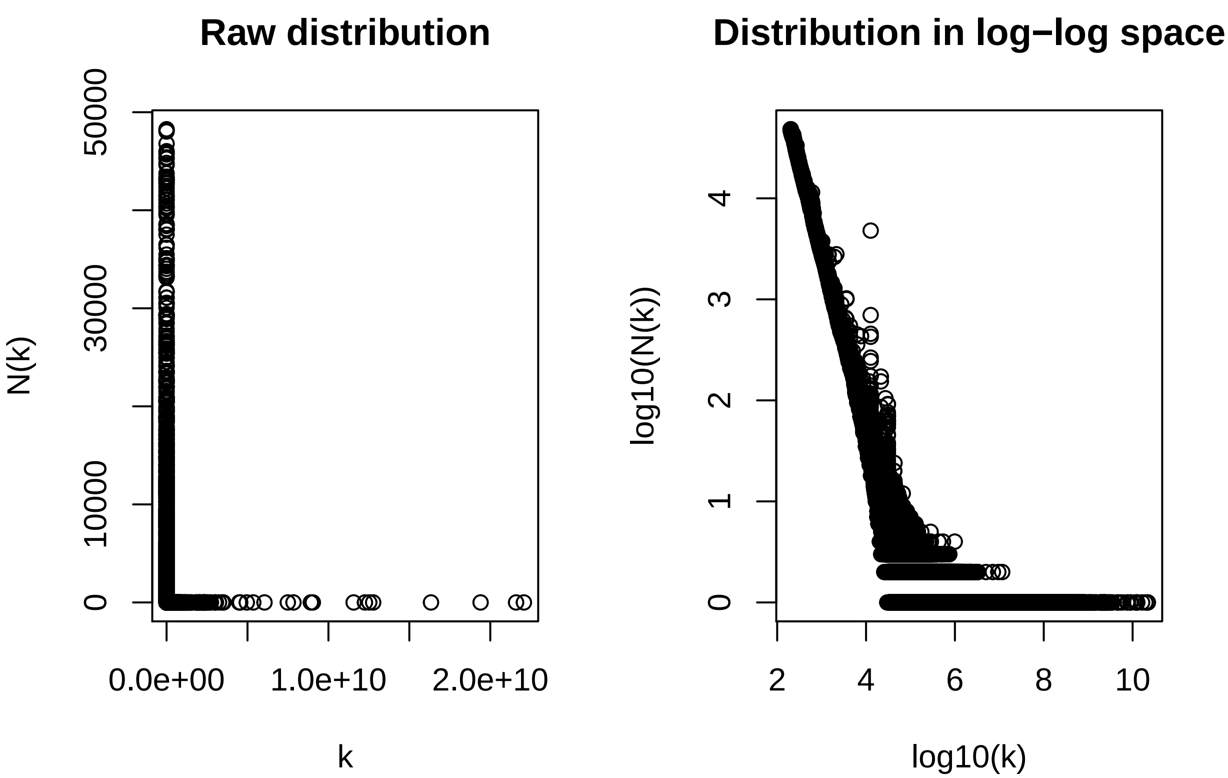
$$\sigma = \frac{a\sqrt{(\gamma-1)}}{\sqrt{(\gamma-3)}(\gamma-2)} a^2$$

Constaints: $\gamma > 3$

b)

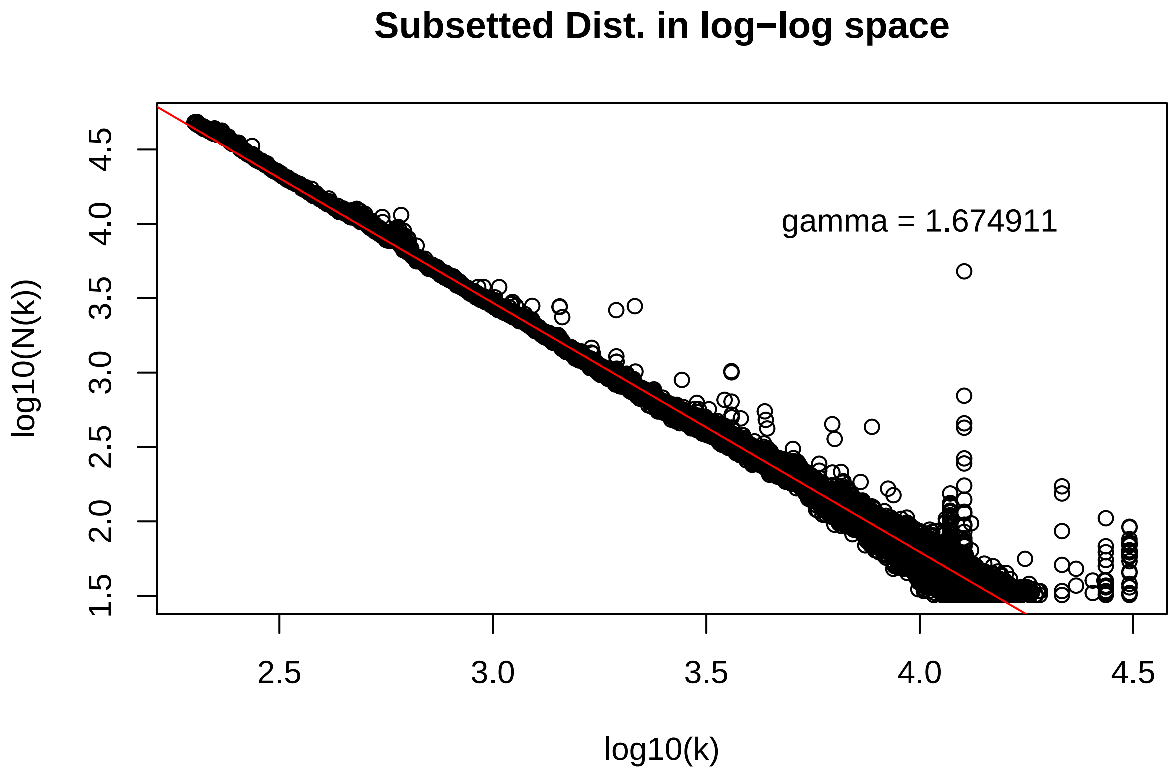
γ must be greater than 3 and when it is below that, sigma goes to infinity. This will be further elucidated in problem 8.

Problem 6



Problem 7:

Data trimed to $\log_{10}\mathrm{N_k} > 1.5$ and γ was estimated to be 1.675




```
##    (Intercept) wordsTrim$logk
##       8.493782      -1.674911
```

Problem 8:

```
## [1] -192921.7
```

mean of k = 3235569 sd of k = 109794501 a estimate = 200 gamma estimate = 1.675

plugging these estimates back into the equation from problem 5 for variance we get a negative number (-192921.7) as our estimate of gamma is less than 3 causing sigma to go to infinity. Also, our computed mean is smaller than our computed sd, which further gives evidence that sigma will go to infiinity, meaning that the moment does not exist. This makes sense given the mean, sd and estimated gamma value as well as the negative variance result by plugging these estimates back in.

Problem 9

a)

This probability can be written as…

$$\frac{P(2\ girls\ at\ least\ 1\ born\ on\ Tuesday)}{P(2\ girls\ at\ least\ 1\ born\ on\ Tuesday)+2(P(at\ least\ 1\ girl\ born\ on\ Tuesday))}$$

Grid can be set up that is 7x7 with a total of 49 cells with T denoting instances were a girl is born on Tuesday.

$$\frac{\overset{(g)}{(g)\begin{matrix} - & T & - & - & - & - & - \\ T & T & T & T & T & T & T \\ - & T & - & - & - & - & - \\ - & T & - & - & - & - & - \\ - & T & - & - & - & - & - \\ - & T & - & - & - & - & - \\ - & T & - & - & - & - & - \end{matrix}}}{\overset{(g)}{(g)\begin{matrix} - & T & - & - & - & - & - \\ T & T & T & T & T & T & T \\ - & T & - & - & - & - & - \\ - & T & - & - & - & - & - \\ - & T & - & - & - & - & - \\ - & T & - & - & - & - & - \\ - & T & - & - & - & - & - \end{matrix}} + \overset{(b)}{(g)\begin{matrix} - & - & - & - & - & - & - \\ T & T & T & T & T & T & T \\ - & - & - & - & - & - & - \\ - & - & - & - & - & - & - \\ - & - & - & - & - & - & - \\ - & - & - & - & - & - & - \\ - & - & - & - & - & - & - \end{matrix}} + \overset{(g)}{(b)\begin{matrix} - & T & - & - & - & - & - \\ - & T & - & - & - & - & - \\ - & T & - & - & - & - & - \\ - & T & - & - & - & - & - \\ - & T & - & - & - & - & - \\ - & T & - & - & - & - & - \\ - & T & - & - & - & - & - \end{matrix}}}$$

This simplifies to...

$$= \frac{\frac{13}{49}}{\frac{13}{49} + \frac{7}{49} + \frac{7}{49}} = \frac{13}{27} = 0.4814815$$

This question has a great deal to do with the language or wording of the question itself. The intuitive answer is 1/2. There is one girl already. Another child has a 50% chance of being another girl. However, the addition of Tuesday alters the probability.

The question states, "A parent has two children, not twins, and one is a girl born on a Tuesday. What's the probability that both children are girls?"

If we were stating, "A parent has two children, one of which is a girl. What is the probability the other is a girl?"" The answer would be 1/3.

$$P(gg) = \frac{gg}{gg + gb + bg}$$

Boy-boy does not enter into the calculation as we already know at least one child is a girl. Instinctively, the addition of the day should not matter. If the question asked something like, "what is the probability of having a girl born on Tuesday and a girl born on any other day of the week?" It would make sense that the day of the week matters. However, the question is rather more subtle in its wording and distinctly ignores order. If order were included, the answer would be 1/2. The addition of information alters the probability space. Now a conditional probability with day of the week as a factor exists. The more information given, the closer to 1/2 we get as shown in part b.

b)

This probability can be written as...

$$\frac{P(2\ girls\ at\ least\ 1\ born\ Dec.\ 31)}{P(2\ girls\ at\ least\ 1\ born\ on\ Dec.\ 31)+2(P(at\ least\ 1\ girl\ born\ on\ Dec.\ 31))}$$

Grid can be set up that is 365x365 with a total of 133225 cells with D31 denoting instances were a girl is born on December 31. Assume no leap year.

$$\frac{\begin{array}{cc|ccccc} & & & & (g) & & \\ & & 1 & 2 & 3 & .. & 365 \\ & 1 & - & - & - & - & D31 \\ (g) & 2 & - & - & - & - & D31 \\ & 3 & - & - & - & - & D31 \\ & : & - & - & - & - & : \\ & 365 & D31 & D31 & D31 & .. & D31 \end{array}}{\begin{array}{cc|ccccc} & & & & (g) & & \\ & & 1 & 2 & 3 & .. & 365 \\ & 1 & - & - & - & - & D31 \\ (g) & 2 & - & - & - & - & D31 \\ & 3 & - & - & - & - & D31 \\ & : & - & - & - & - & : \\ & 365 & D31 & D31 & D31 & .. & D31 \end{array} \begin{array}{cc|ccccc} & & & & (g) & & \\ & & 1 & 2 & 3 & .. & 365 \\ & 1 & - & - & - & - & D31 \\ +\ (b) & 2 & - & - & - & - & D31 \\ & 3 & - & - & - & - & D31 \\ & : & - & - & - & - & : \\ & 365 & - & - & - & - & D31 \end{array} \begin{array}{cc|ccccc} & & & & (b) & & \\ & & 1 & 2 & 3 & .. & 365 \\ & 1 & - & - & - & - & - \\ +\ (g) & 2 & - & - & - & - & - \\ & 3 & - & - & - & - & - \\ & : & - & - & - & - & - \\ & 365 & D31 & D31 & D31 & .. & D31 \end{array}}$$

This simplifies to...

$$= \frac{\frac{729}{133225}}{\frac{729}{133225} + \frac{365}{133225} + \frac{365}{133225}} = \frac{729}{1459} = 0.4996573$$

This additional information has pushed the probability even closer to 1/2. The reason for this asymptotic approach to 1/2 is due to the ratio of overlapping scenarios in the girl-girl case, which is always $1/(2 * \textit{length of } n)$. This ratio decreases as n increases. For 1 day of a week, n=14 and the overlap proportion is 1/14. For a single day of the year, it is much smaller, 1/730. This discrepancy drives the overall promotion towards 1/2 as n increases.