# Estimation the mean street size based on proportion of odds

**11/09/2018, One on Epsilon.**

***Edited by Clara Valtorta and Yoni Nazarathy.***

---

This is a companion to a Curious Epsilon blog post. See One on Epsilon Blog (https://oneonepsilon.com/blog).

In the computation below we assume a specific form of a probability distribution for the number of houses on a street (street size). This distribution is a geometric distribution on $1, 2, 3, \ldots$ with mean $M$. It has a probability mass function:

$$p_M(n) = \frac{1}{M}(1 - \frac{1}{M})^{n-1}.$$

Now for even streets the proportion of odd values is $1/2$ but for odd streets of length n, the probability is

$$\frac{1}{2}\frac{n+1}{n}.$$

Hence if we observe a proprotion $\alpha$ of the streets as odd then it should hold that,

$$\alpha = \sum_{n=1}^{\infty} P(odd \mid n) p_M(n) = \sum_{n=1}^{\infty} \frac{1}{2} p_M(2n) + \sum_{n=1}^{\infty} \frac{1}{2}\frac{2n}{2n-1} p_M(2n-1).$$

We now approximate the right hand sum in Julia (via trunctation up to a large value). We then search for the $M$ that bests fits the observed $\alpha = 0.502$. This gives us an estimate of the mean street size, $M$.

In [1]:

```
N1 = 10000 #upper support of distribution
N2 = 2000 #value up to where to search for best M

geomDist(n,M) = (1/M)*(1-(1/M))^(n-1)
odd(n) = (n+1)/(2n)
alpha(M,dist) = sum([0.5*dist(2k,M) for k in 1:N1]) + sum([odd(2k-1)dist(2k+1,M) for k
in 1:N1])

function findM(observedAlpha)
    errs = [abs(alpha(M,geomDist) - observedAlpha) for M in 1:N2]
    last(findmin(errs))
end

findM(0.502)
```

Out[1]:

644

Hence we estimate that the mean number of houses on a street is $644$. Keep in mind, that the strong assumption made here is the geometric distribution assumption.

---

We can also see how this estimate would vary based on the measurment of $\alpha$. For this lets let $\alpha$ be in the range $(0.501, 0.503)$ and recalculate the matching $M$ for each such $\alpha$. We then obtain a plot showing how different $\alpha$ measurments would yield different estimates of $M$.

In [2]:

```
using PyPlot
ests = [[alpha,findM(alpha)] for alpha in 0.501:0.0001:0.503]
plot(first.(ests),last.(ests))
xlim(0.501,0.503)
xlabel("observed alpha")
ylabel("mean street size");
```