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 <u>One on Epsilon Blog</u> (<u>https://oneonepsilon.com/blog</u>).

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:

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

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

$$rac{1}{2}rac{n+1}{n}$$

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

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

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

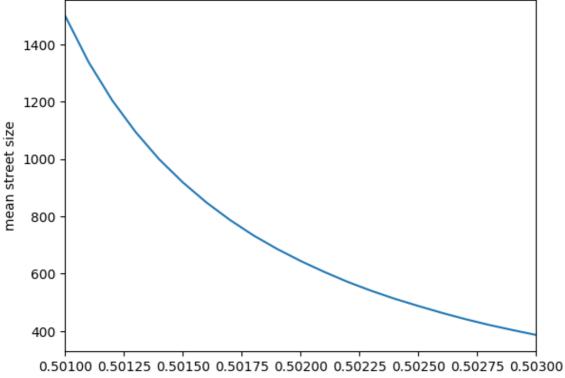
Street Size

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 α . For this lets let α be in the range (0.501, 0.503) and recalculate the matching M for each such α . We then obtain a plot showing how different α 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");
```



observed alpha