R. Koenker Spring 2017

Economics 574
Problem Set 2

1.: A famous 18th century problem, and one of the earliest examples of the monte-carlo method, is Buffon's needle problem. It is also a rare, perhaps even unique, example of a case in which we can consider estimating a parameter which describes an aspect of the natural world about which we "know" more than we can ever hope to learn from estimation. This is not the case with the marginal propensity to consume, for example! In Buffon's original formulation a a needle of length l is thrown "at random" onto a floor ruled with parallel lines $d \geq l$ units apart. In n throws let S_n denote the number of times the needle crosses one of the lines.

a.: Show that $\hat{\theta} = dS_n/2ln \to \theta \equiv 1/\pi$. Thus we have a monte carlo strategy to estimate π.

Hint: Interpret "at random" above to mean that that on each throw the center of the needle falls X units from the nearest line with X is uniformly distributed on the interval $[0, d/2]$. The angle, φ, the needle makes with the nearest line is also random and uniformly distributed on $[0, \pi]$. Assume further that X and φ are independent. Try this with a pencil as the needle and the lines formed by the rows of tile on the floor. Figure 1 below may help.

Thus the sample space Ω of the Buffon experiment may be viewed as a rectangle of length π and height $d/2$ on which the rv's (X, φ) have a uniform density. Given this setup the needle will cross a line iff

$$X \leq \frac{l}{2} \sin \varphi$$

Find the probability, p, of this event and note that $S_n =$ binomial (p, n).

b.: Compute the variance of $\hat{\theta}$ suggested in (a).

c.: Find the value of d to minimize $V(\hat{\theta})$. Remember: $l \leq d$.

d.: Use the δ-method to compute the variance of the natural estimator of π, $\hat{\pi} = \hat{\theta}^{-1}$.

e.: In R we can "simulate" Buffon's experiment as follows

```
Sn <- sum(runif(n) < sin (pi * runif(n)))
```

This counts the number of crossings in n trials. Convince yourself that this is correct before abandoning throwing pencils on the floor! Try this with several sample sizes and construct confidence intervals for π.

f.: Since Buffon, there have been many variations on the original problem. Laplace considered a variation in which there is a square grid with vertical and horizontal spacing d. Again the needle is thrown n times but now let V_n denote the number of vertical crossings and H_n the number of horizontal crossings. Based on our previous considerations we know that

$$\hat{\theta}_V = V_n/2n$$

and

$$\hat{\theta}_H = H_n/2n$$

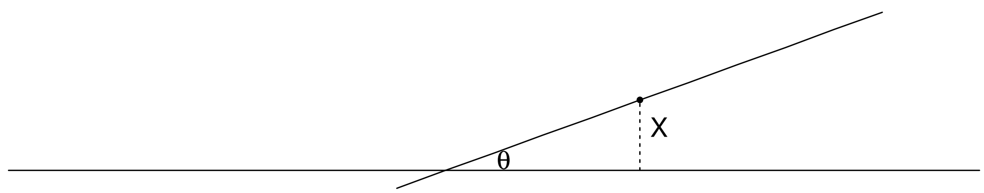

FIGURE 1. Geometry of the Buffon Needle Experiment

where we are now implicitly assuming that $d = l = 1$, are both unbiased estimates of θ. So it is natural to consider combining the estimators to obtain.

$$\tilde{\theta} = (\hat{\theta}_V + \hat{\theta}_H)/2$$

If $\hat{\theta}_V$ and $\hat{\theta}_H$ were independent the variance of the combined estimator would be one-half the variance of either component. Some reflection reveals, however, that the two separate estimates are *negatively* correlated, and therefore the efficiency of the combined estimate will be *greater* than twice the efficiency of $\hat{\theta}_V$ alone. Show that the probability that the needle crosses neither a vertical line nor a horizontal one is[1]

$$P_{\overline{HV}} = 1 - 3\theta$$

so the probability that it crosses both is

$$P_{HV} = P_H + P_V + P_{\overline{HV}} - 1 = \theta$$

[1]Note this implies that $\pi > 3$ contrary to the Iowa legislature which according to an apocryphal story once passed a law that made $\pi = 3$.

since $P_H = 2\theta = P_V$, Show that

$$V(\tilde{\theta}) = \frac{\theta}{n}(\frac{3}{8} - \theta)$$

and compare the efficiency of the corresponding estimator of π with the one obtained from only the "horizontal estimator."

Hint: This example and further improvements are discussed in Perlman and Wichura (1975), "Sharpening Buffon's Needle," *American Statistician*, 157-160. This brief paper is an excellent object lesson in the usefulness of sufficiency and the Rao-Blackwell theory. Even in the simple form given above the example illustrates the utility of combining so-called *antithetic* variates, i.e., the idea that finding negatively correlated estimates of θ can dramatically improve the accuracy of $\hat{\theta}$.

2.: A nice 19th century example of the usefulness of the monte-carlo method as an aid to thought about matters probabilistic is Galton's quincunx. The quincunx is a mechanical device developed by Francis Galton and others at the end of the 19th century to illustrate the DeMoivre-LaPlace theorem. The DeMoivre-LaPlace theorem was the first CLT and deals with the special case in which the sequence of iid random variables $X_1, X_2, ...$ are Bernoulli, *i.e.* $X_i = \pm 1$ with probability one half. Most of the major science museums have one, but necessarily the physical ones are somewhat limited in their scope and it would be better (not to speak of cheaper) from a pedagogical point of view to have a computer simulation version.

See Figure 2 for an illustration. Balls are poured into a chute at the top; they fall through the pins represented by the " $\cdot$ "'s -in the simplest case falling to the - left and right to the next level with equal probability – and collecting in the chutes at the bottom in a pattern which looks roughly "Gaussian" if there are a sufficient number of levels of pins and number of balls.

a.: A rather primitive "virtual quincunx" written by your estimable instructor in R is available on the class webpage. Watching it "run" is the econometric equivalent of watching the waves at the seashore. Spend some time doing this to get a feel for the DeMoivre-LaPlace theorem.

b.: Suggest some modification of the existing R code to improve its current implementation or to illustrate some new aspect of the DeMoivre-Laplace problem.

c.: Run a moderately large (say, $n = 1000$) example and evaluate the adequacy of the normal approximation for the resulting allocation of the balls.

3.: Generate several random samples of size 500 from the density

$$f(x) = \phi(x)(1 + \frac{1}{2}\sin(2\pi x))$$

by the rejection method discussed in class and compare several methods of density estimation for these samples.

4.: A common data analytic task is to compare a hypothetical distribution or density with some nonparametric fit. Histograms are sometimes useful for this purpose since if they are constructed correctly you can superimpose the hypothetical density over the bars of the hostogram. This usual way of plotting has some difficulties, however, as a visualization device. First, it is difficult to judge the discrepancy relative to the curve of the hypothetical curve. A suggestion of John Tukey helps in this respect: instead

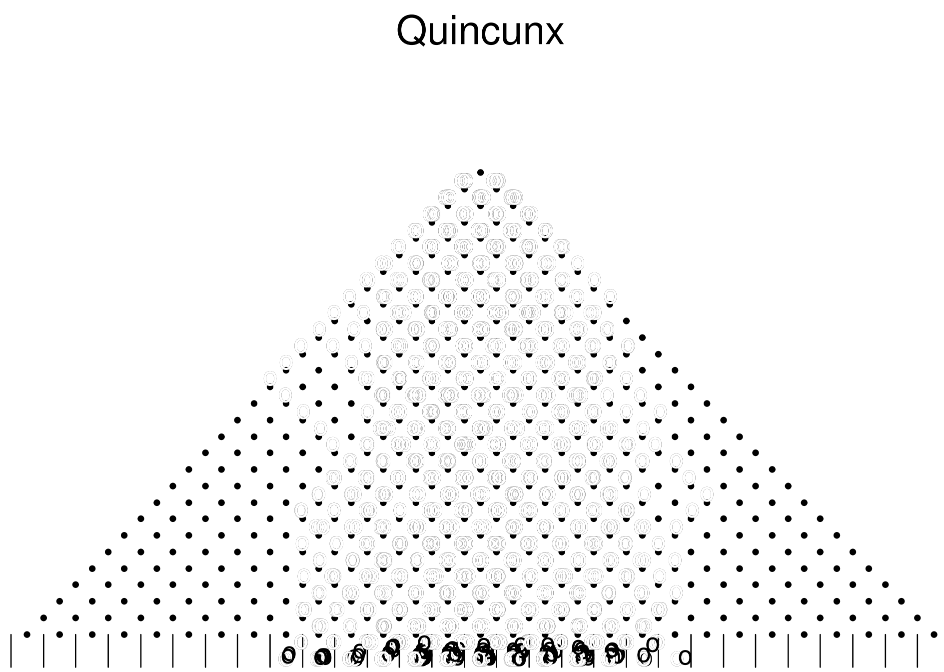

FIGURE 2. Galton's Quincunx

of plotting the bars as emanating from the horizontal axis, plot them so that their upper limits coincide with the hypothetical values and their lower limits can be viewed relative to the straight line of the x-axis. A second difficulty is that the density scale isn't really the right scale to judge these descrepancies. Another suggestion of Tukey's helps to resolve this problem too: rather than plotting bars versus $f(x)$, he proposed plotting square roots of these values. In the inimitable Tukey style, he called these plots – hanging rootograms.

Explain why the second suggestion is reasonable based on what is known about the variability of histogram bars. For a good example of how this looks in practice see
`http://addictedtor.free.fr/graphiques/graphcode.php?graph=31`.

For an amusing description of a bit more about the rationale for these methods see Wainer, H. "The Suspended Rootogram and Other Visual Displays: An Empirical Validation", *Am. Statistician*, 28, 143-45.

5.: Consider the kernel density estimator

$$\hat{f}_n(x) = n^{-1} \sum K_h(x - X_i) = (nh)^{-1} \sum K((x - X_i)/h)$$

where $K(\cdots)$ is a proper symmetric density, i.e., $K(u) \geq 0$ and $\int K(u)du = 1$, $K(u) = K(-u)$, with mean zero $\mu(K) \equiv \int uK(u) = 0$, and $\mu_2(K) \equiv \int u^2 K(u)du = 1$.

a.: Show that if $h \to 0$, $\hat{f}_n(X)$ is asymptotically unbiased, that is $E\hat{f}_n(x) \to f(x)$.

b.: Write the bias of $\hat{f}_n(x)$ as

$$B(\hat{f}_n(x)) = \int K(s)f(x - hs)ds - f(x)$$

expand f as a quadratic and show that

$$B(\hat{f}_n(x)) = \frac{h^2}{2}\mu_2(K)f''(x) + o(h^2)$$

for $h \to 0$.

c.: Now consider the variance of $\hat{f}_n(x)$ showing that if $nh \to \infty$,

$$V(\hat{f}_n(x)) = (nh)^{-1}||K||_2^2 f(x) + o((nh)^{-1})$$

where $||K||_2^2 = \int K^2(t)dt$.

d.: Conclude from parts b.) and c.) that the optimal bandwidth in the sense of minimizing MISE is

$$h_0 = (n\mu_2^2(K)||f''||_2^2/||K||_2^2)^{-1/5}$$

e.: Show that choosing $h = h_0$ gives

$$MISE = O_p(n^{-4/5})$$

f.: If we normalize K so that $\mu_2(K) = 1$, the optimal K should minimize $||K||_2^2 = \int K^2(t)dt$. Show that this is accomplished by the so-called Epanichnikov kernel $K(u) = \frac{3}{4}\sigma^{-1}(1 - (u/\sigma)^2)I(|u| \leq \sigma)$ for $\sigma = \sqrt{5}$.

6.: A simple simulation experiment: Consider the problem of testing whether data is coming from the standard normal distribution or the McCullagh density of PS 1. There are two obvious approaches: the first is the standard Kolmogorov-Smirnov test based on the statistic:

$$T_n = \sup_x |F_n(x) - \Phi(x)|$$

This is conveniently computed in R with the function `ks.test` and will produce p-values. The other test is the likelihood ratio test, based on the statistic,

$$S_n = 2\sum \log(f(X_i)/\phi(X_i))$$

The statistic S_n doesn't have an obvious limiting distribution but you can easily find empirical critical values and then compute estimated power by simply finding the empirical rejection frequency of the test statistics. In fact it is easy to see that the LR test statistic has a normal limiting distribution so if you wanted, you could recenter S_n by computing the expectation of each of the summands, but this isn't really necessary. For the record this mean is about -0.1387, but now we might also want its variance.... Compare the power of the two tests for a few moderate samples sizes, say 50, 100, 500, 1000. It is perhaps instructive, to get some intuition about the performance of the two tests to plot the distribution function of the McCullagh density. This is easily done in R with the following code:

```
f <- function(x) dnorm(x)*(1 + .5*sin(2*pi*x))
F <- function(x) {
     F <- x
     for(i in 1:length(x))
             F[i] <- integrate(f,-Inf,x[i])
     F
     }
plot(x,F(x),type="l")
```

7.: As a further example of the futility of moments, consider the problem of estimating the raw second moment $S \equiv \mathbb{E}X^2$ for X lognormal, i.e. $\log X \sim \mathcal{N}(\mu, \sigma^2)$. The sample analogue $\hat{S} = n^{-1} \sum X_i^2$ has mean $S = \exp\{2\mu + 2\sigma^2\}$, but note that, for $Z \sim \mathcal{N}(0,1)$,

$$\begin{aligned} \mathbb{P}(X^2 > S) &= \mathbb{P}(\log(X^2) > 2\mu + 2\sigma^2) \\ &= \mathbb{P}(\mu + \sigma Z > \mu + \sigma^2) \\ &= \mathbb{P}(Z > \sigma) \\ &= 1 - \Phi(\sigma) \end{aligned}$$

Thus, for $\sigma = 4$ even when n is quite large we never see X's in the extremely long right tail of the distribution, and consequently the sample analogue estimator is quite poor. Provide some simulation evidence for this, and contemplate the its implications for the law of large numbers. Hint: what would Pafnuty Chebyshev have to say?