University of Illinois | **Final Exam** | Department of Economics
Spring 2014 | **Economics 574** | Roger Koenker

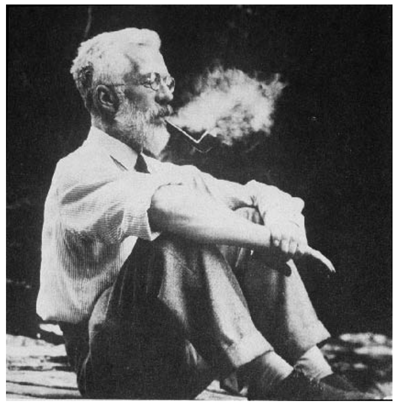

FIGURE 1. R.A. Fisher in a Cloud

The exam is due Friday, May 16 at 5pm. Feel free to email or stop by my office if you feel that there are points that need clarification; I'll post corrections or clarifications on the course web page, if necessary.

(1) In an ancient (2002) 574 final exam I asked to compare several estimators of the location parameter of the Cauchy distribution: the mean, the median, the Cauchy MLE and a one-step estimator starting for the median. The original question and my version of a quick answer is included below. To extend this question, I'd like you to consider three possible generalizations:
 (a) Adding the optimal L-estimator as discussed in L15 as a fifth competitor,
 (b) Introducing a free scale parameter to the Cauchy model,
 (c) Replacing the location parameter formulation by a linear regression.
 These extensions raise a variety of issues from the extremely practical such as, how would we extend the code to implement methods for each case? to more conceptual questions like, how to evaluate simulation evidence and compare to the theoretical Cramer-Rao lower bound? Consider each case and try to briefly sketch your approach.

(2) In the bonus lecture slides on mortality and smoking I've tried to illustrate the main results from the massive study of British doctors in the last half of the 20th century. (An appalling sidelight on this work is provided by Stolley (1991) who recounts R.A. Fisher's role in the debate over smoking and lung cancer.) Suppose that you had access to the raw data from the Doll et al study, and you decided to write a critique of the original analysis arguing that since smoking was presumably a (semi-) voluntary decision some recognition of the endogeneity of the number of years spent smoking should be made. Write a (maximum one page) proposal sketching what you would propose to do.

Stolley, P.D. (1991) When Genius Errs: R.A. Fisher and the Lung Cancer Controversy, *Am. J. Epidemiology*, 133, 416-425.

Appendix A. The 2002 Cauchy Question

Consider the problem of estimating the location parameter, θ, from a random sample of Cauchy random variables with density $f(x, \theta) = \frac{1}{\pi}\left(\frac{1}{1+(x-\theta)^2}\right)$.

(1) Describe briefly how to compute "one-step version" of the mle, assuming that you use the sample median as an initial estimator.
(2) Extend the result in (a) to describe how to compute the mle.
(3) Write an *R* function to compute the estimators in (a) and (b) and conduct a *small*[*] monte-carlo experiment to compare the performance of the mean, median, one step mle, full mle. Hints: (i) see `nlm(` to automate the mle computation, (ii) an n-vector of Cauchy rv's can be generated as `rt(n, 1)`.
(4) Based on the monte-carlo, make a conjecture about the asymptotic variance of the Cauchy mle, and then find the limiting distribution of the mle and compare with your conjecture.

Appendix B. The 2002 Cauchy Answer

```
#Spring 2002 Econ 476 Final Exam Question 2
lik <- function(theta,x){
#log likelihood function for iid cauchy location model scale ==1
sum(-log(1/(1+(x-theta)^2)))
}
onestep <- function(theta,x){
#onestep mle for the iid cauchy location model scale == 1
grad <- sum(2*(x-theta)/(1+(x-theta)^2))
hess <- sum(+4*(x-theta)^2/(1+(x-theta)^2)^2 - 2/(1+(x-theta)^2))
theta - grad/hess
}
set.seed(1917)
R <- 1000
ns <- c(20,40,100)
A <- array(0,c(4,3,R))
for(j in 1:3){
        n <- ns[j]
        for(i in 1:R){
                x <- rt(n,1)
                A[1,j,i] <- mean(x)
                A[2,j,i] <- median(x)
                A[3,j,i] <- onestep(median(x),x)
                A[4,j,i] <- nlm(lik,median(x),x=x)$estimate
                }
        }
a <- apply(A^2,c(1,2),mean)
dimnames(a) <- list(c("Mean","Median","Onestep","MLE"),c("n=20","n=40","n=100"))
caption <- "Mean Squared Errors for 4 estimators of location for iid
Cauchy observations and 3 sample sizes:  1000 replications"
tab <- format(round(a,3))
require(Hmisc)
latex(tab, rowlabel = "Estimator", caption=caption)
```

This runs almost instantaneously and yields the following table.

[*]Small means something like the following design: 3 sample sizes say 20, 50 and 100 with 1000 replications.

Table 1. Mean Squared Errors for 4 estimators of location for iid Cauchy observations and 3 sample sizes: 1000 replications

Estimator	n=20	n=40	n=100
Mean	1057.017	1213.201	667.356
Median	0.137	0.066	0.025
Onestep	0.122	0.053	0.021
MLE	0.119	0.053	0.021