Evidence for long-tailed distributions in the Internet

Allen B. Downey
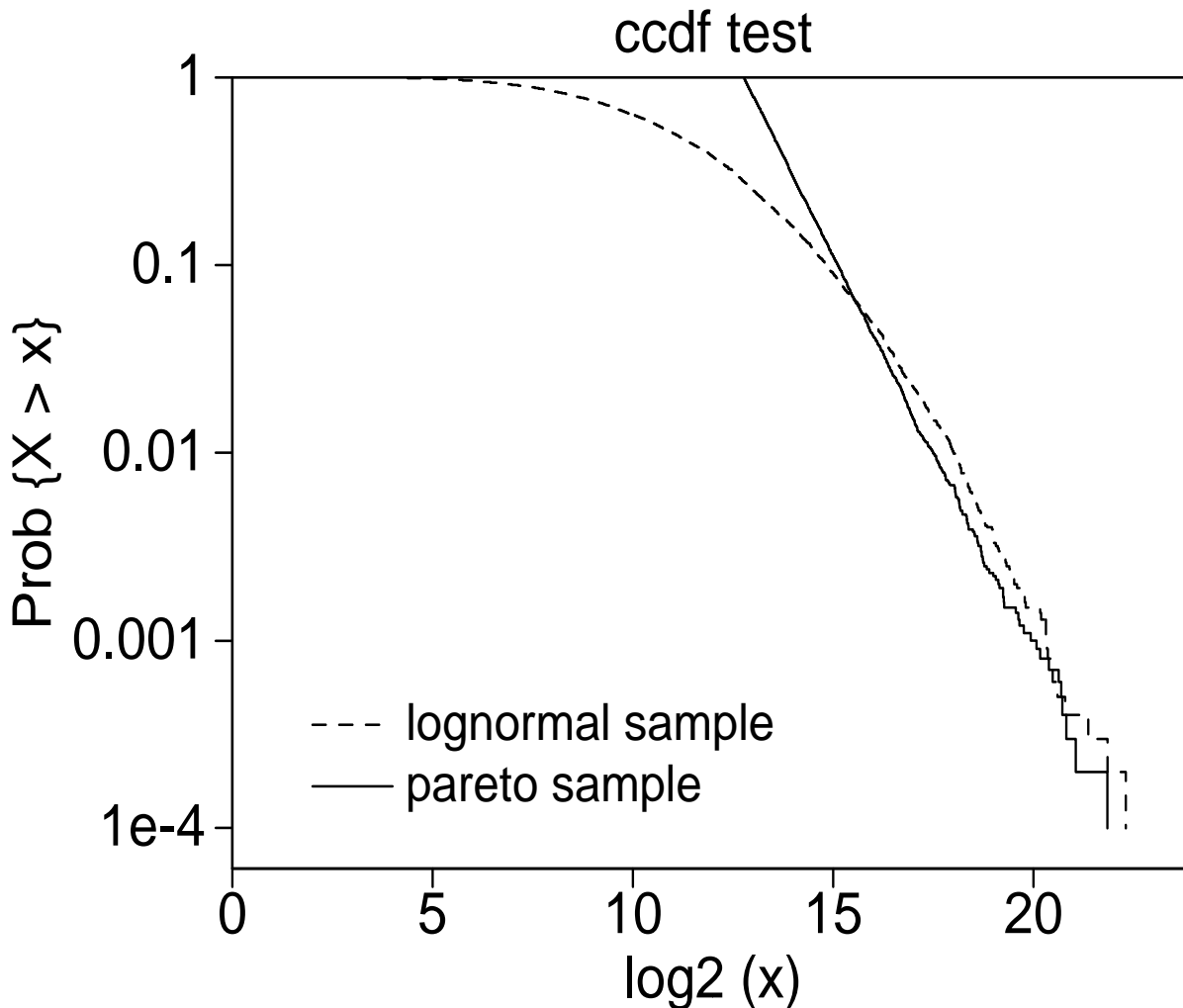
Wellesley College

Self-Similarity

No shortage of explanations...
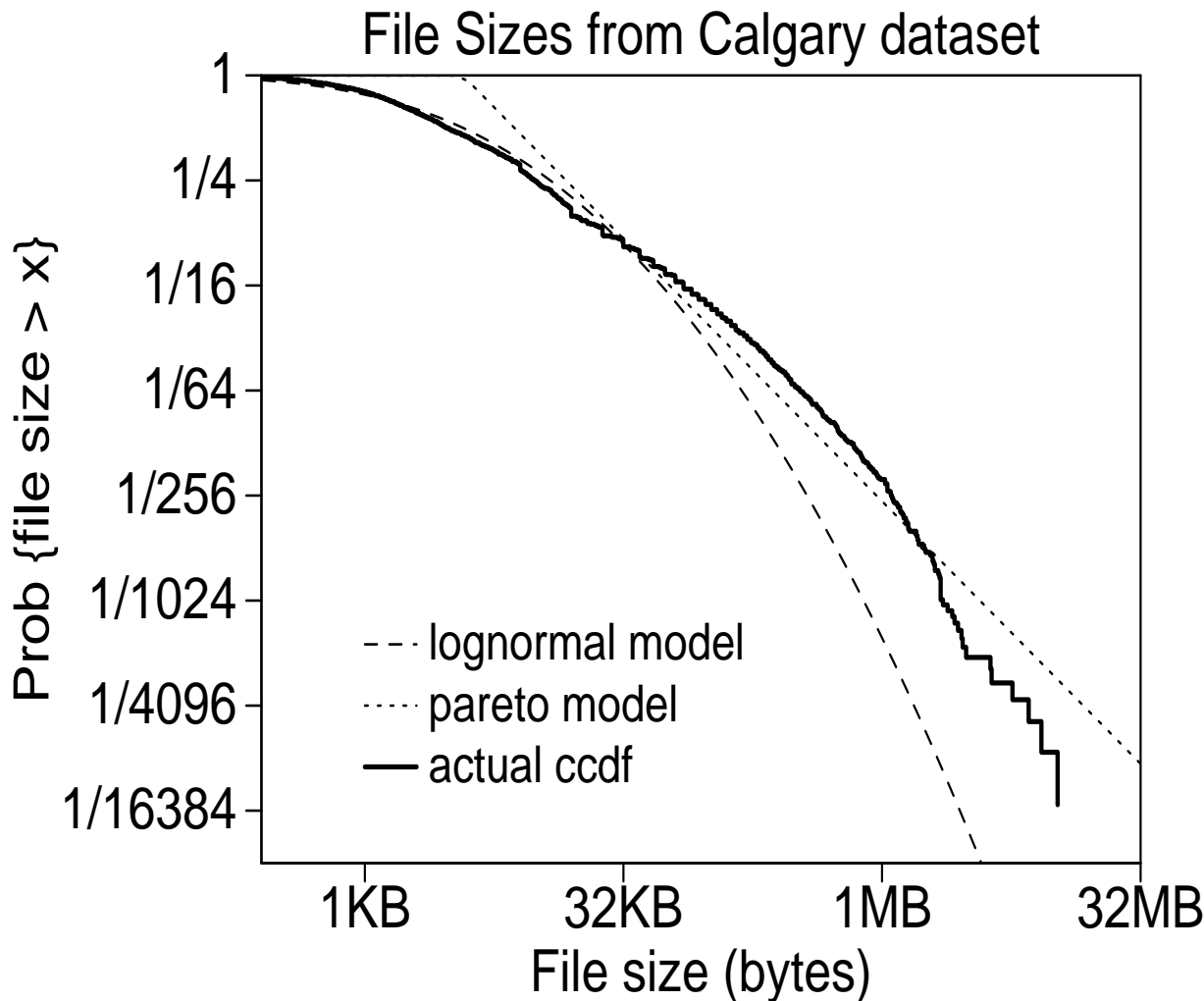
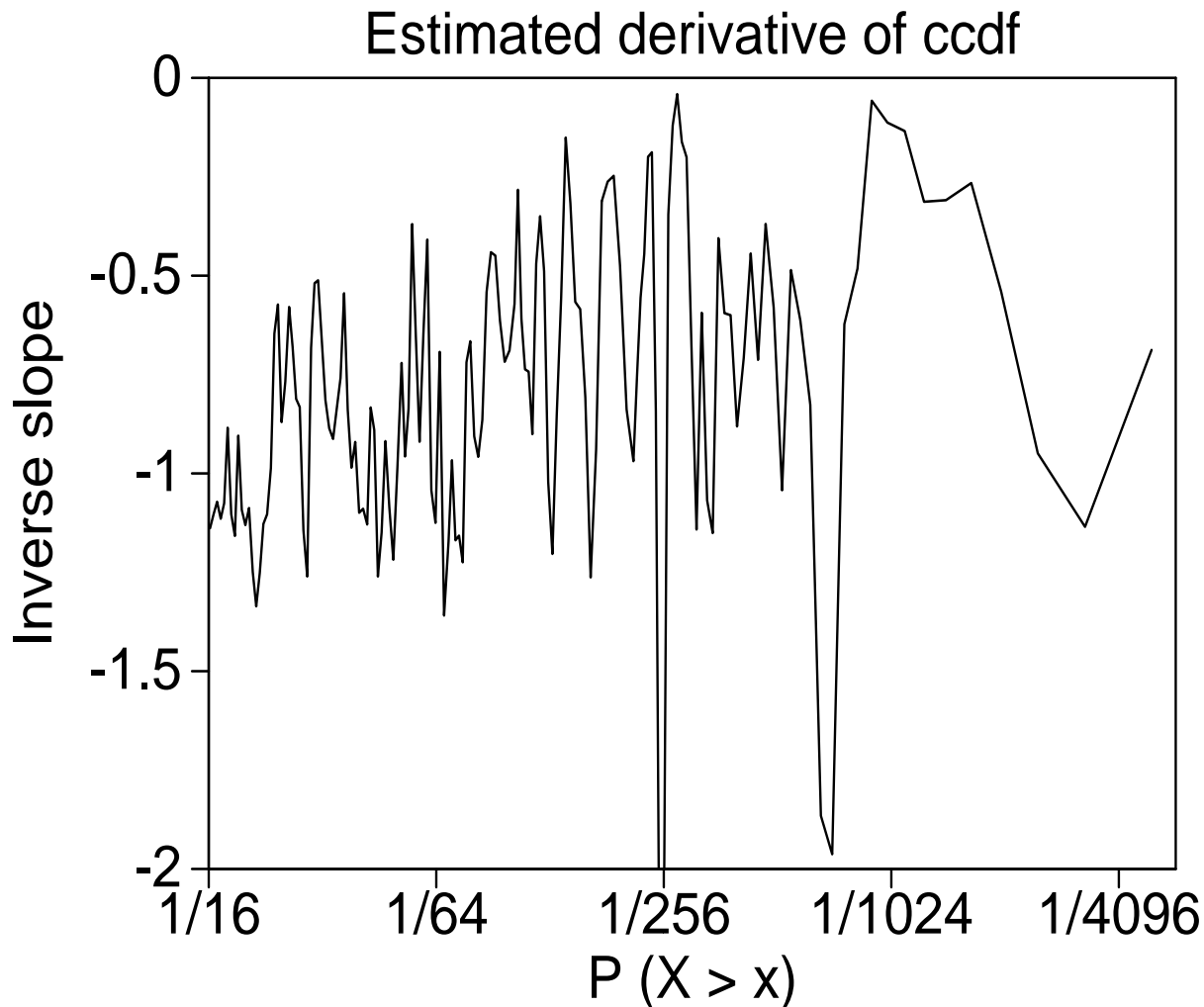- ■ ON/OFF model

- ■ M/G/∞ model

- ■ Protocol models

ccdf test

ccdf test

- Samples (n=10,000) from Pareto and lognormal distributions with similar tail behavior.

File sizes on a Web server

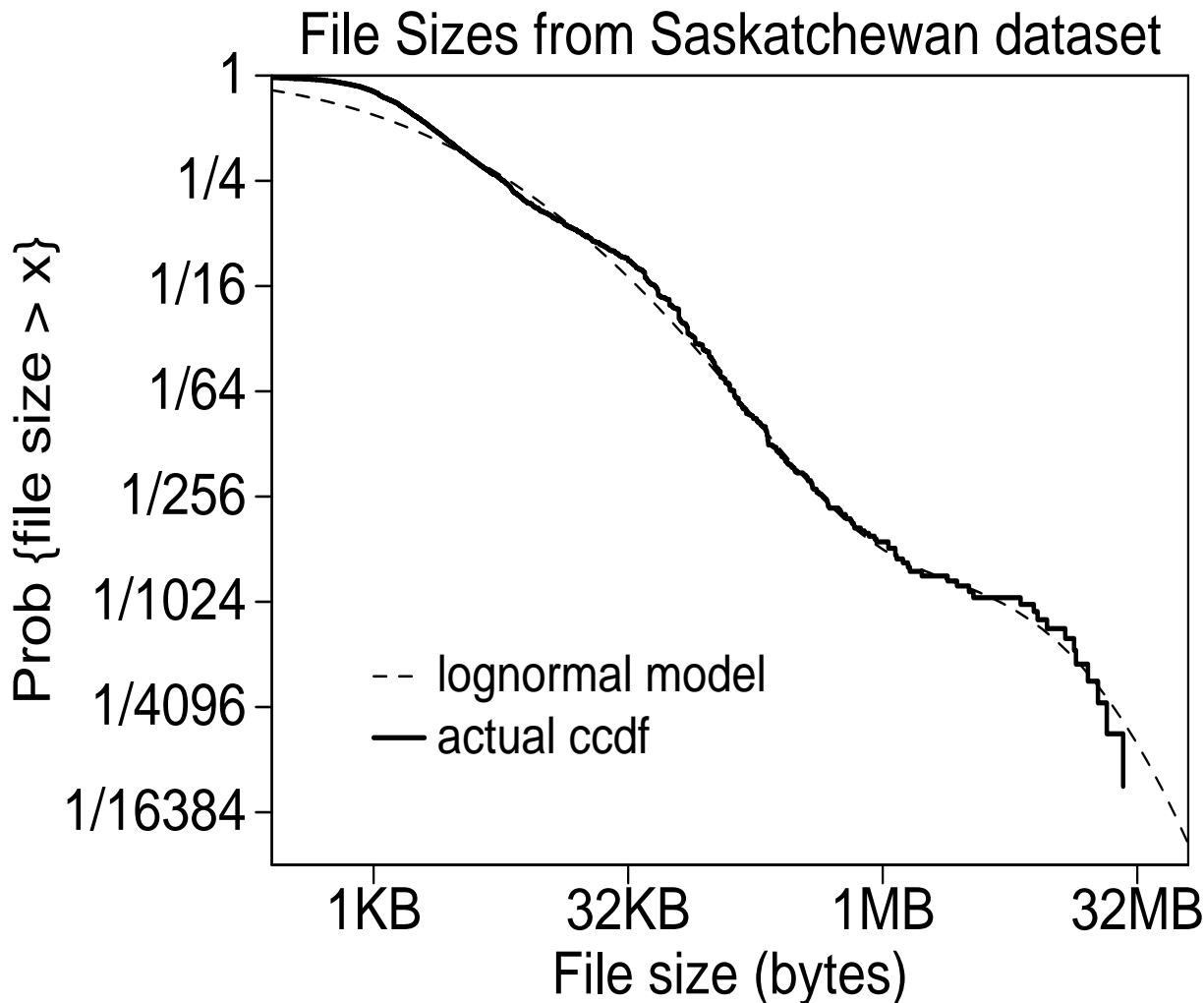File Sizes from Calgary dataset

- Sizes of 15,160 files at the University of Calgary.

- By conventional goodness of fit, Pareto wins.

- Tail behavior is not long-tailed.
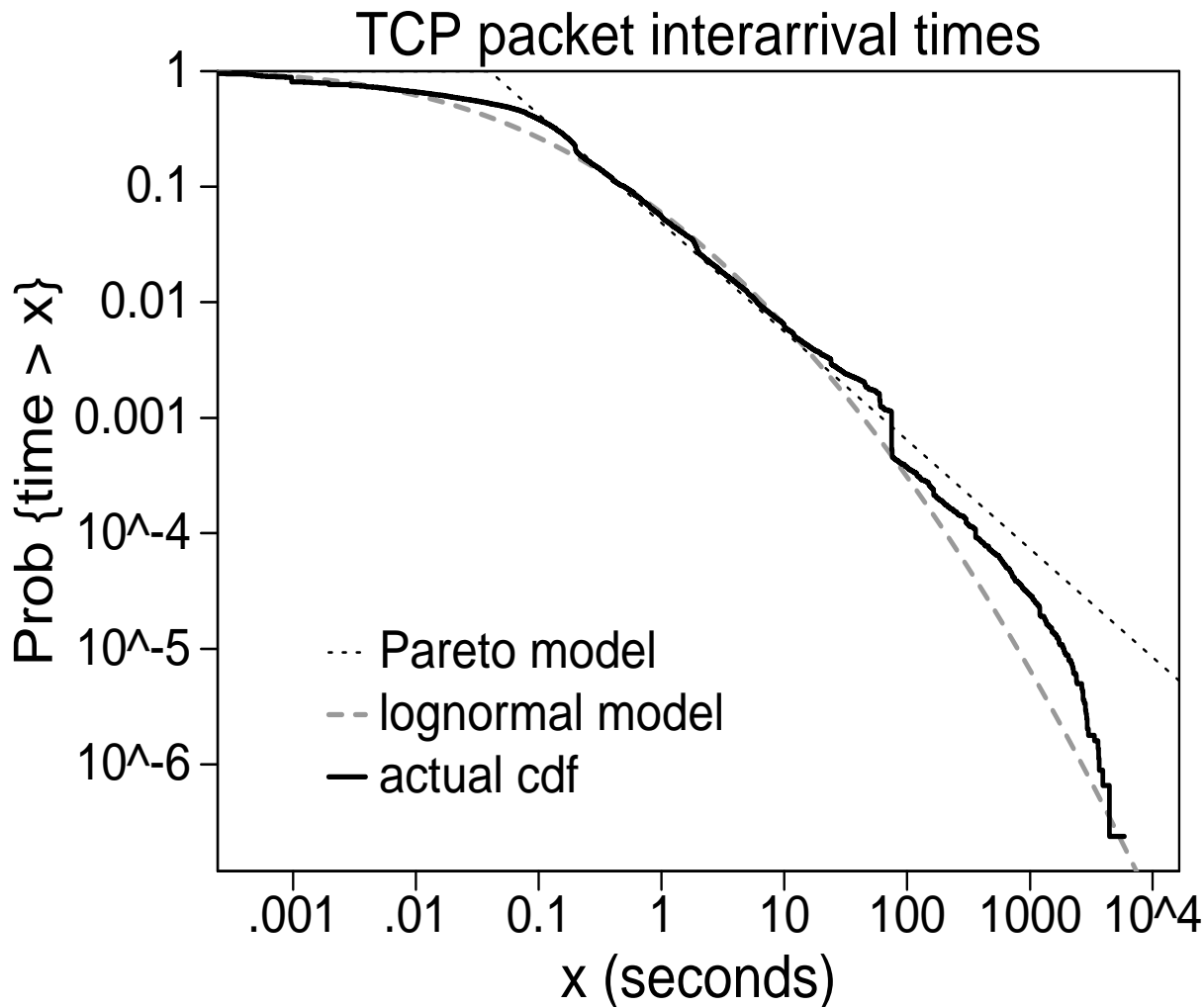
Numerical differentiation

Estimated derivative of ccdf

- Numerical derivatives are noisy.

- Testing for trends is robust.

- Tail curvature = 0.141, p-value $<$ 0.001.

Files sizes on another Web server

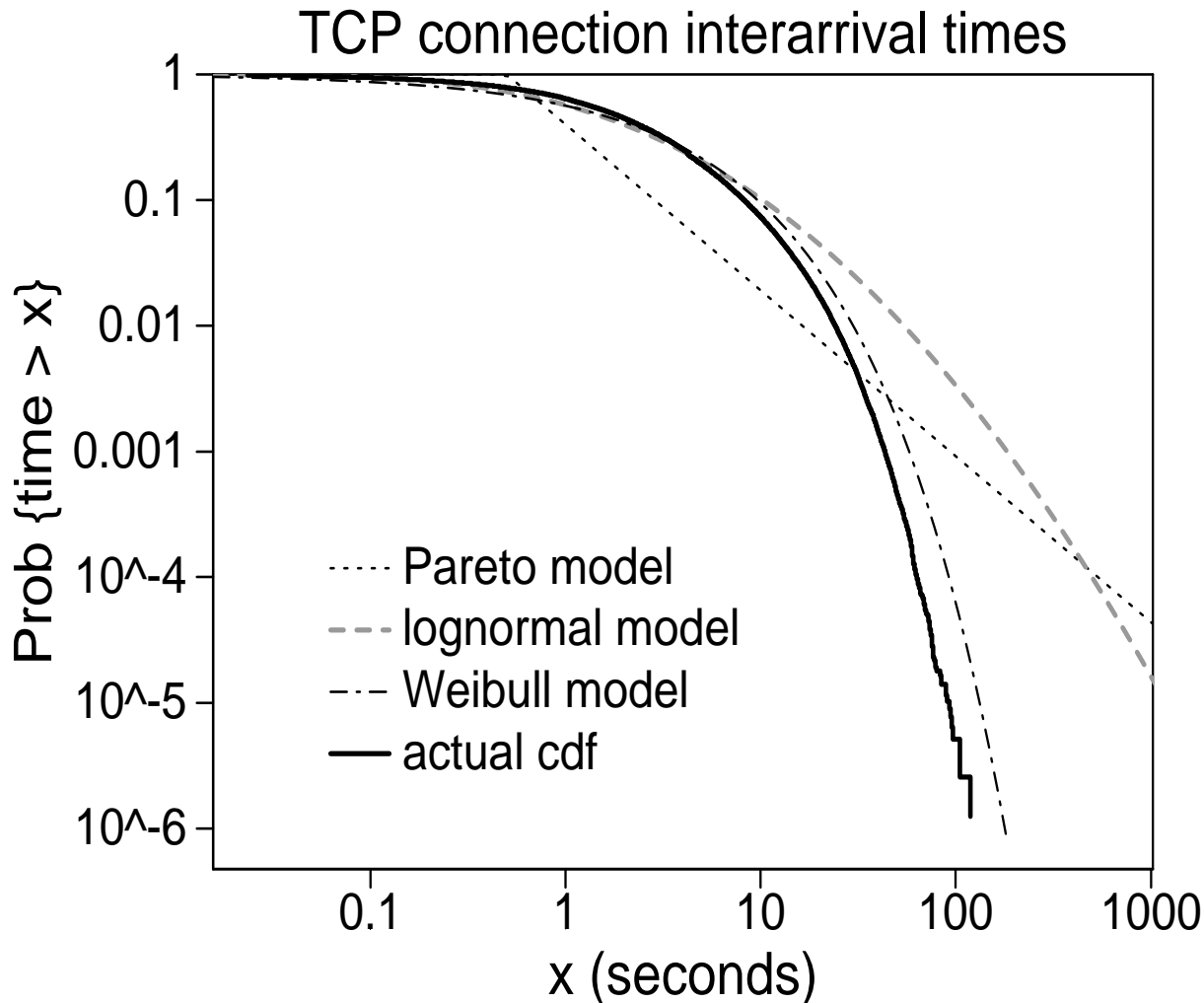File Sizes from Saskatchewan dataset

- Files sizes from University of Saskatchewan.

- Pareto model fits well.

- Two-mode lognormal model fits well.

- Tail curvature test is no help.

Interarrival times, TCP packets
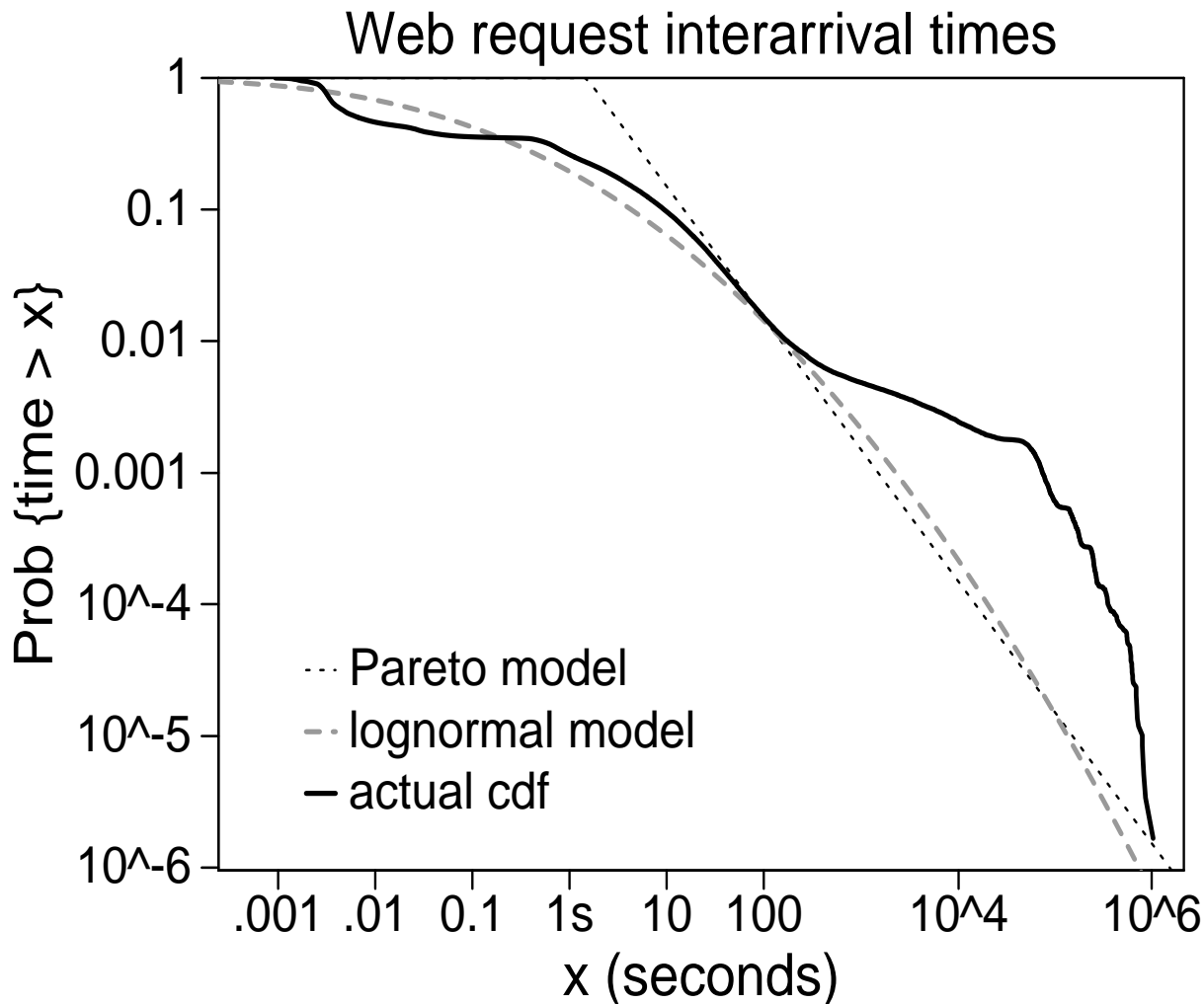

TCP packet interarrival times

- 4 million interarrivals from LBL and DEC datasets.

- Very consistent between datasets.

- Some signs of straightness.

- Extreme tail hard to characterize.
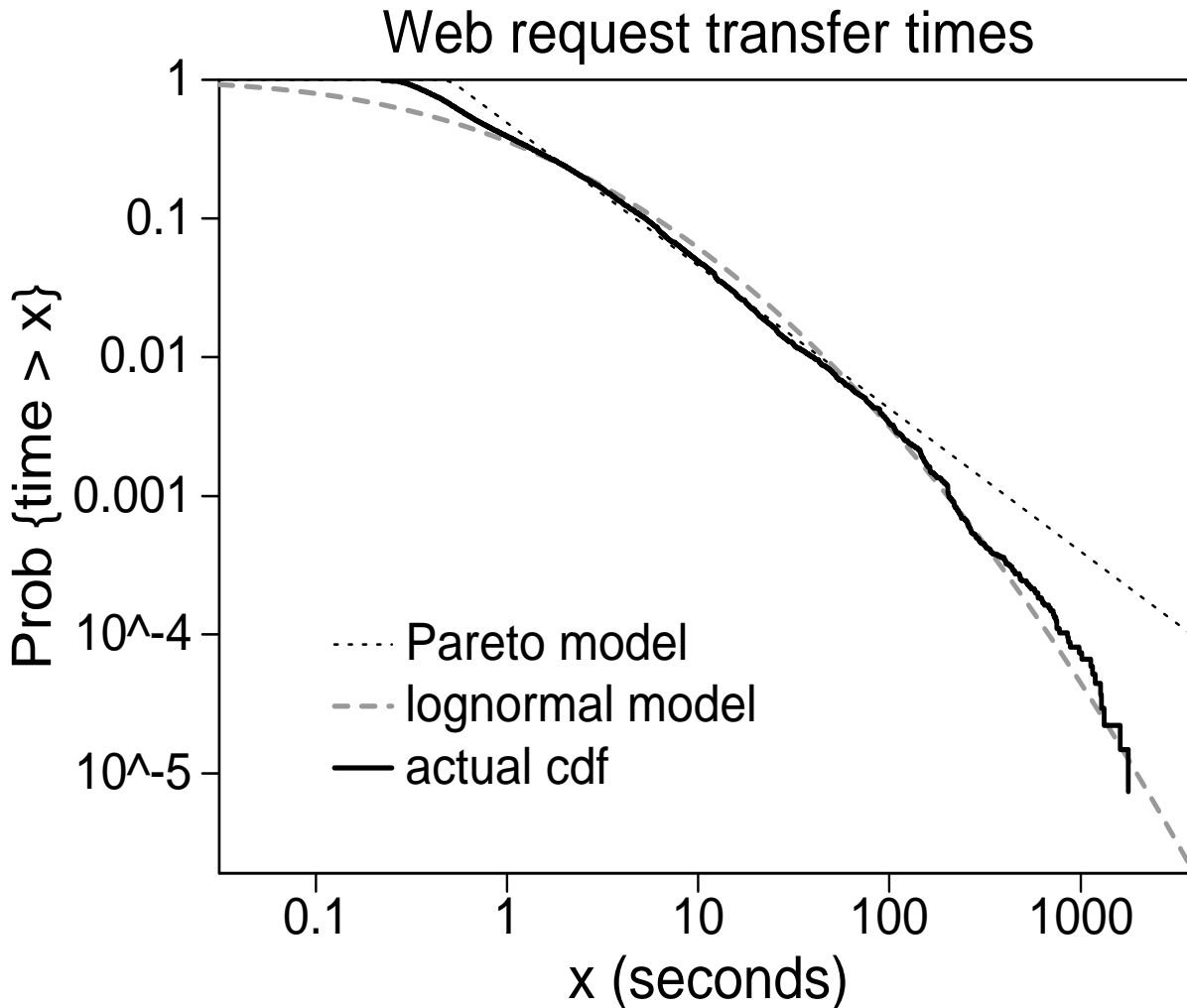
Interarrival times, TCP connections

TCP connection interarrival times

- **782,000 connections in LBL CONN-7.**

- **Feldmann reports that Weibull fits the bulk.**

- **Fits the tail well, too.**
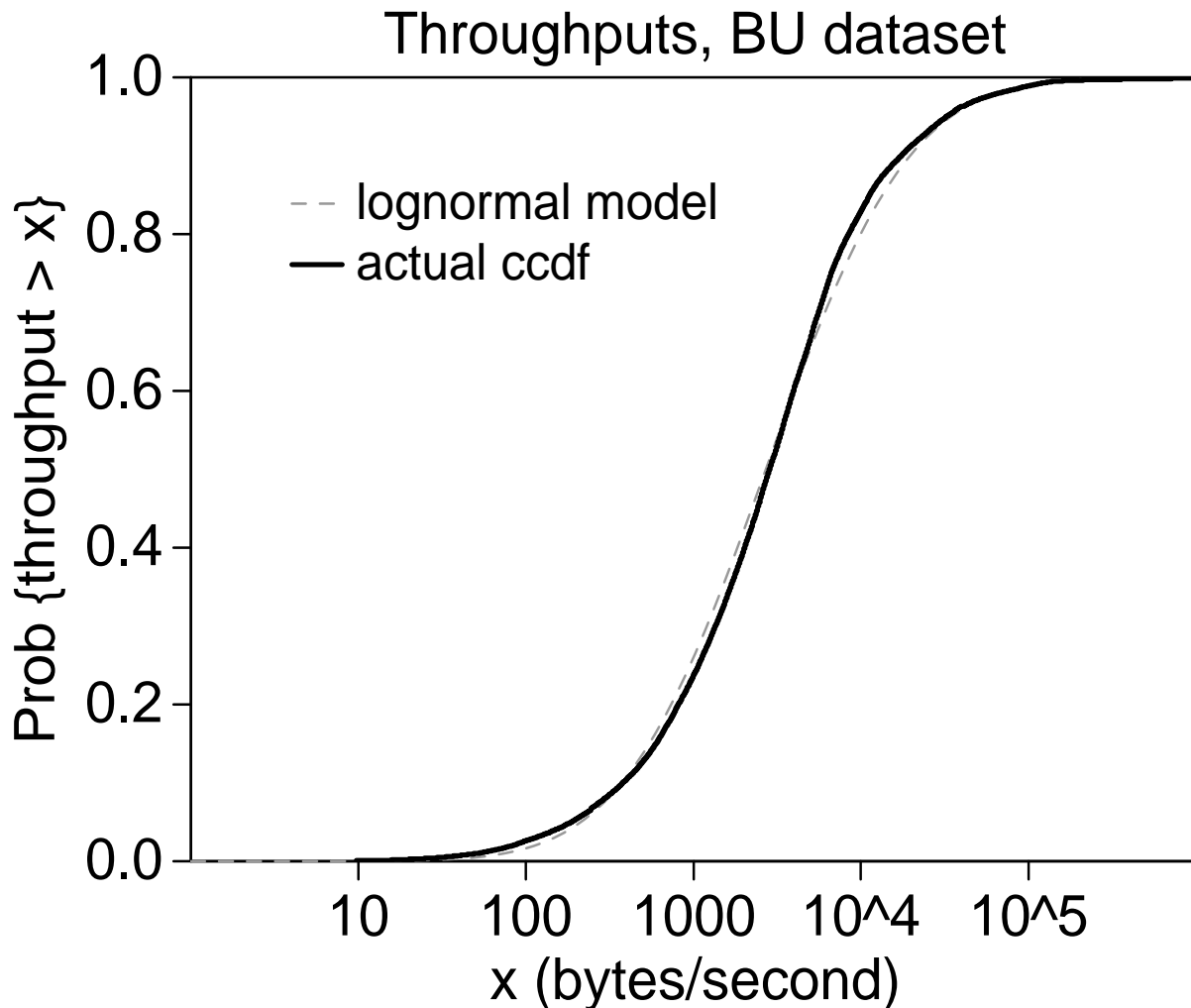
Interarrival times, web requests

Web request interarrival times

- 135,000 requests from instrumented browsers at BU.

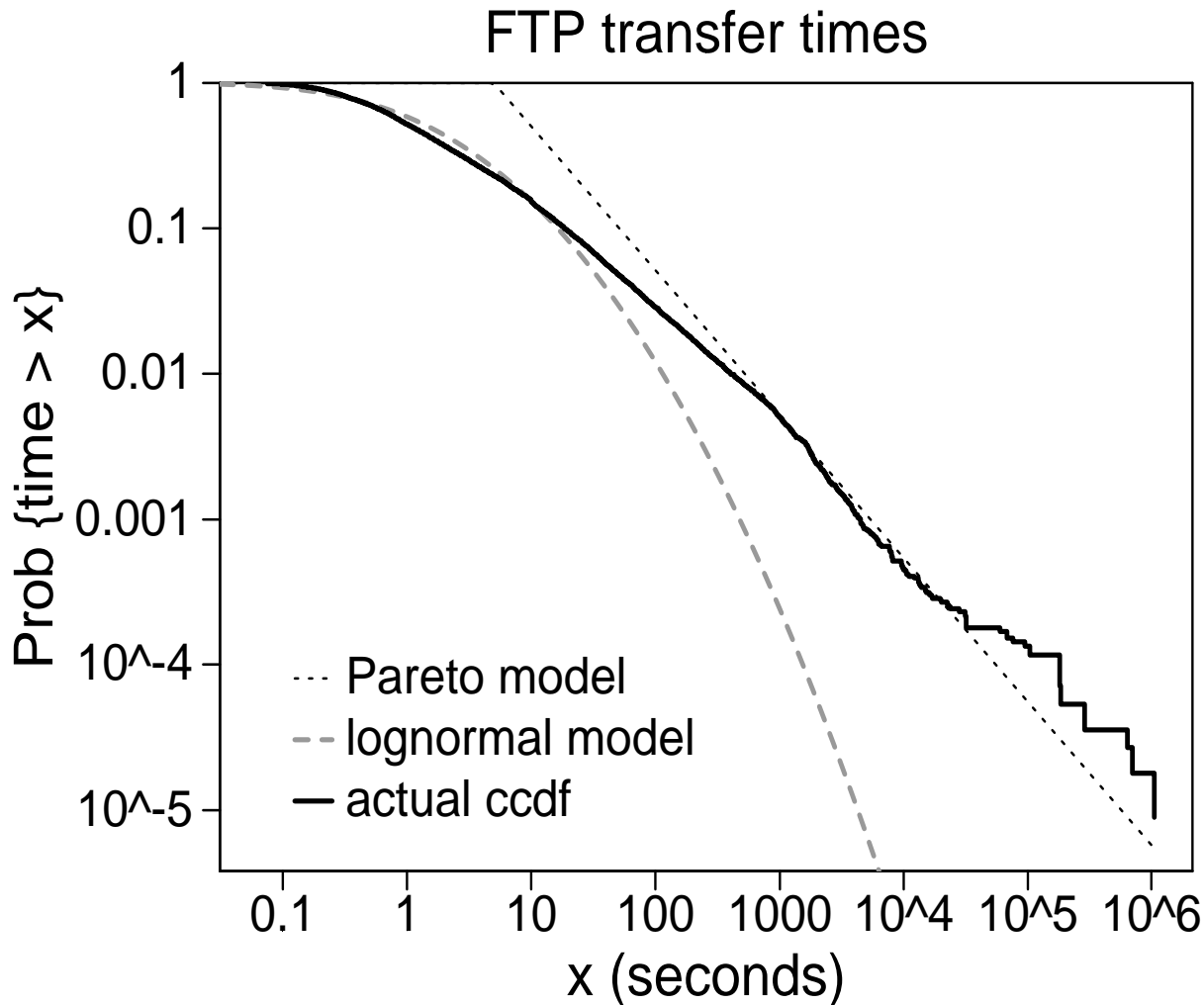- Hard to characterize tail behavior.

http transfer times

Web request transfer times

- **135,000 transfers.**

- **Lognormal model fits the extreme tail.**
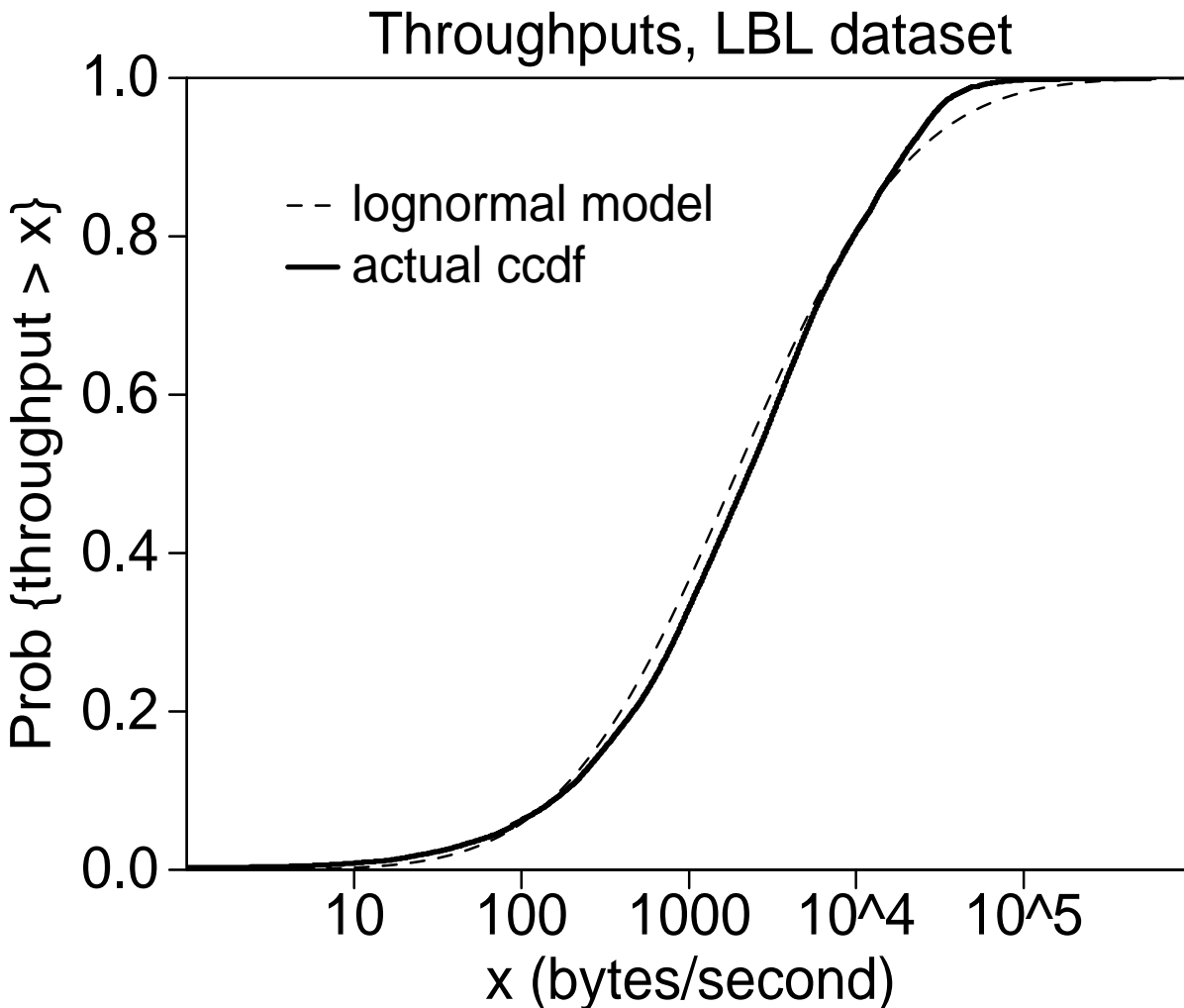
Throughput

Throughputs, BU dataset

- For each transfer, divide size by transfer time.

- Across paths and time, throughput is roughly lognormal.

ftp transfer times

FTP transfer times
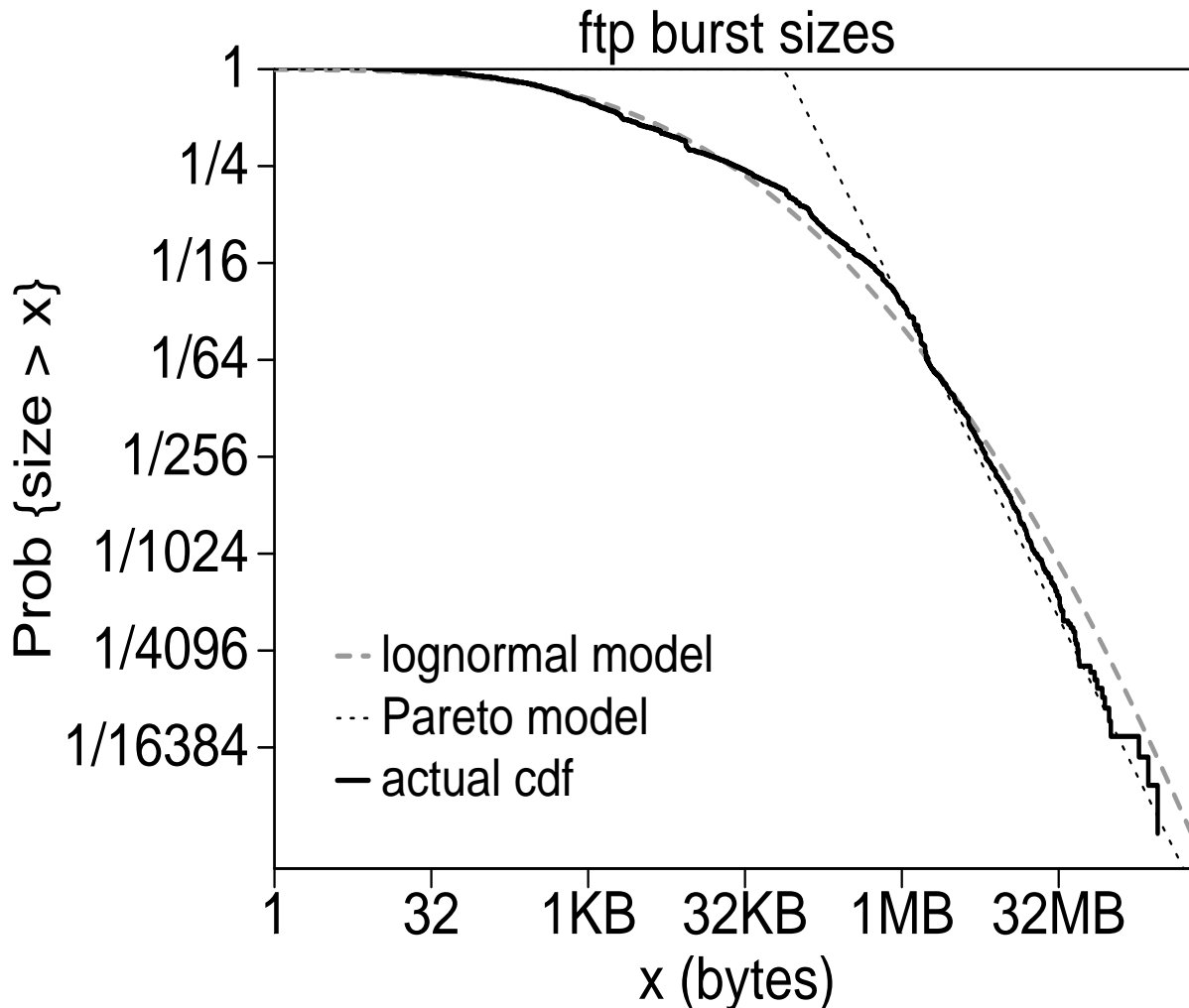

- 105,000 transfers in LBL CONN-7.

- Not so clear that this is lognormal.
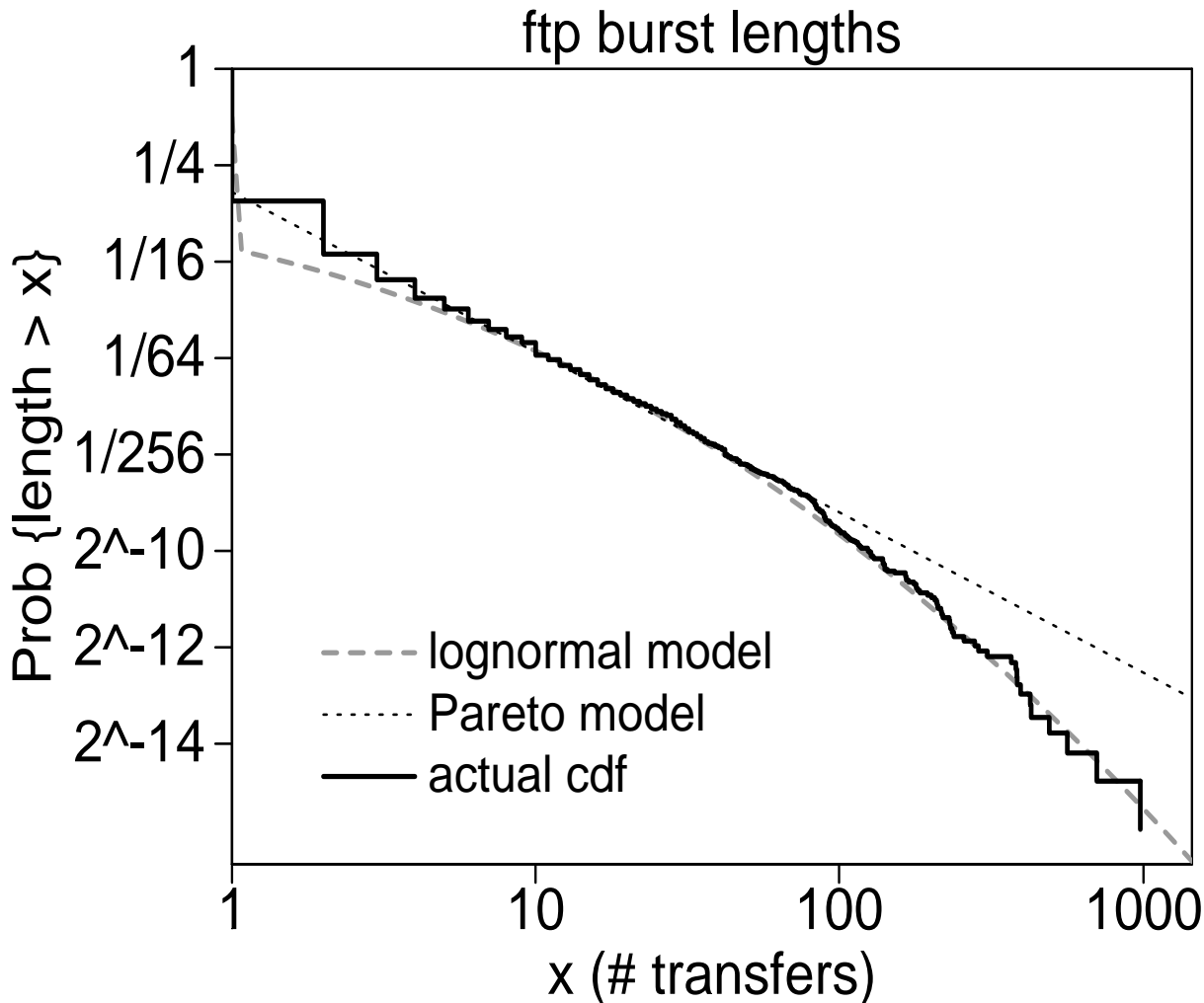
- Paxson used two-stage Pareto model.

ftp throughput

Throughputs, LBL dataset

- - lognormal model
— actual ccdf

- Again, roughly lognormal.
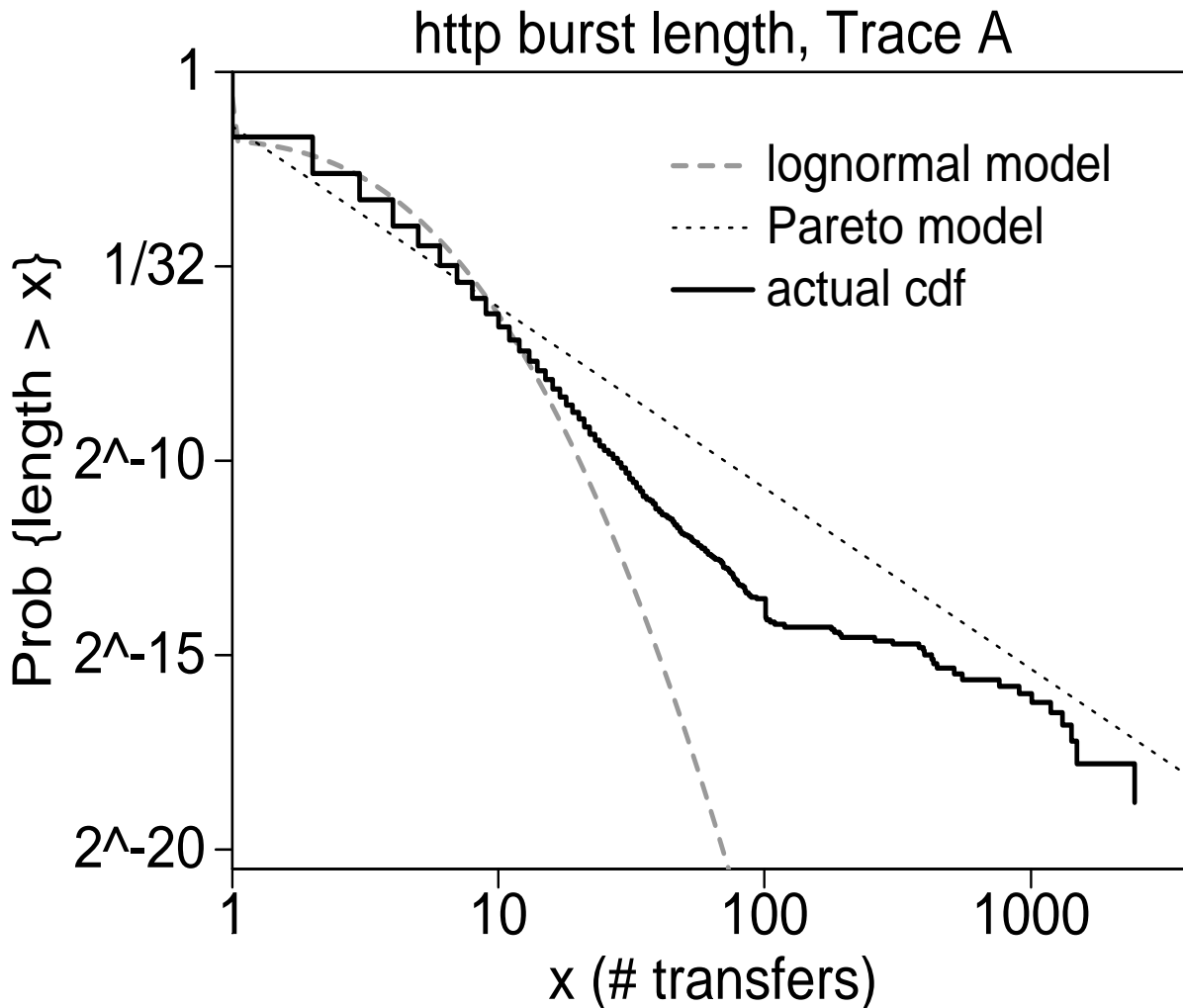- Top end compressed by hw limitations.

ftp burst sizes

ftp burst sizes

- Two or more transfers with <4s between.

- 56,000 bursts.
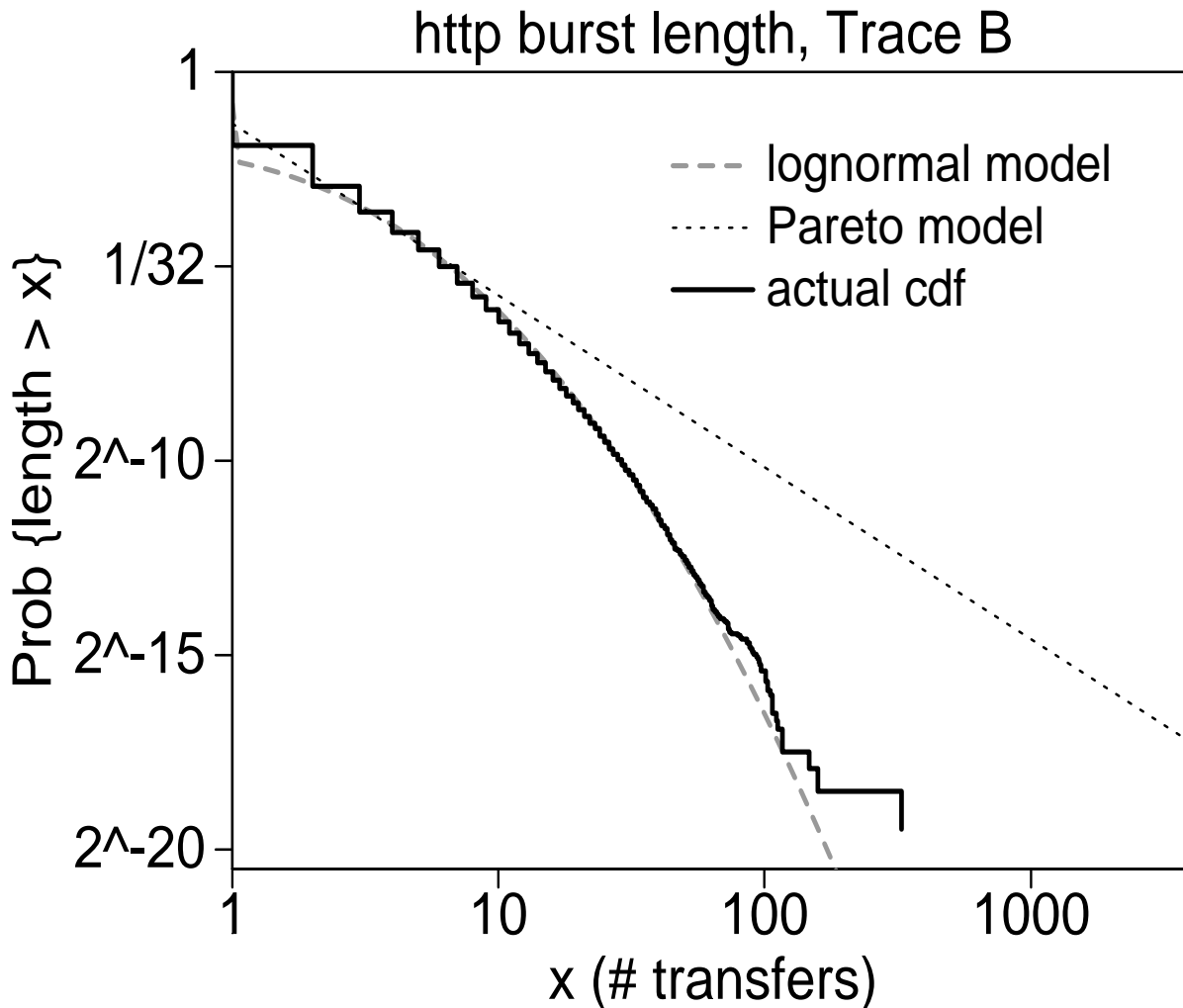
- Fairly convincing straight line.

ftp burst lengths

ftp burst lengths

- How many transfers in a burst?
- 85% are singletons.
- Lognormal? Pareto?

http burst lengths

http burst length, Trace A

- 456,000 http connections from Charzinski trace.

- 70% of connections make a single request.

- Tail behavior hard to characterize.

More http burst lengths

http burst length, Trace B

- 739,000 connections recorded by Charzinski.
- Pretty clearly lognormal.