14 January 2004
revised 18 January 2004

INTRODUCTION TO MATHEMATICAL MODELLING
LECTURE 5: OTHER DISTRIBUTIONS

David A. Meyer

Project in Geometry and Physics, Department of Mathematics
University of California/San Diego, La Jolla, CA 92093-0112
http://math.ucsd.edu/~dmeyer/; dmeyer@math.ucsd.edu

Ubiquity of the normal distribution

In the previous lectures we have discussed the normal distribution (for heights) and the binomial distribution (for the number of heads in a sequence of coin flips). The Central Limit Theorem tells us that the former approximates the latter, and moreover, that it approximates sums of independent, identically distributed (i.i.d.) random variables with finite expectation value and variance. This suggests that the normal distribution should occur in many contexts, and motivates statements like that of Galton quoted in Larsen & Marx [1, p. 276]. We saw, however, that when we tried to approximate some of the data that people in the class collected using normal distributions with the corresponding sample mean and sample variance, the approximations were not always so good.

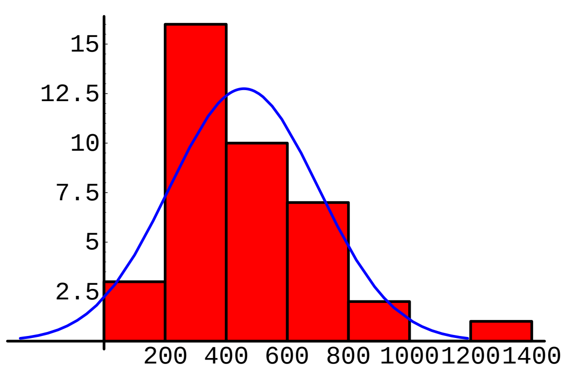

Figure 5.1. Histogram of page lengths of books with the graph of the normal distribution having the same mean and variance.

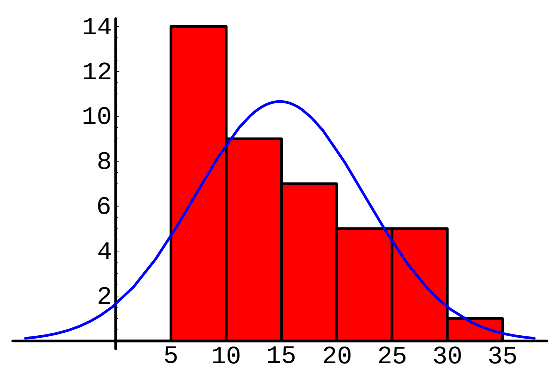

Figure 5.2. Histogram of touchdowns per quarterback with the graph of the normal distribution having the same mean and variance.

One feature that several of the data sets collected have in common is that they contain only positive numbers. This is true for heights, for the sizes of poker pots, for page lengths of books, for ages at which people learn to ride bicycles, for song lengths, for numbers of touchdowns per quarterback, and for ages at death. Although any normally distributed random variable has positive probability of taking any real numbered value, if the mean of the distribution is sufficiently large relative to its variance, the probability of observing a negative value is almost 0. In the case of heights, for example, Figures 2.3 and 2.4 show that the normal distributions approximating the height data do not predict that we will observe any negative heights [2]. When the mean is smaller relative to the variance, however, the normal distribution does not approximate positive valued data so well. Figures 5.1 and 5.2 show data on page lengths of books for a sample of books in the SIO library (collected by Jayme Reynolds) and on touchdowns per quarterback in the 2003 NFL season (collected by Jeremy Greene). In each case the approximating normal distribution implies a substantial probability for negative values which, of course, are not observed.

One possible fix for this problem could be to use the distribution for $|X|$ where X has a normal distribution with mean 0, *i.e.*,

$$f(x) = \begin{cases} \frac{2}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} & \text{if } x \geq 0; \\ 0 & \text{if } x < 0. \end{cases} \tag{5.1}$$

This is a probability density function: it is positive and integrates to 1.

Homework: Compute the mean and variance of the distribution in eq. 5.1.

Figures 5.3 and 5.4 show the approximations to the book length and touchdown data using this distribution with the corresponding variance. There is still, clearly, a problem at values near 0. These plots also illustrate another problem: the data include too many large values to be well approximated by a normal distribution or by the distribution in eq. 5.1. This feature is sometimes discribed as a ‘heavy tail’ or ‘fat tail’. In this course we will try to construct mathematical models of several systems that display this phenomenon.

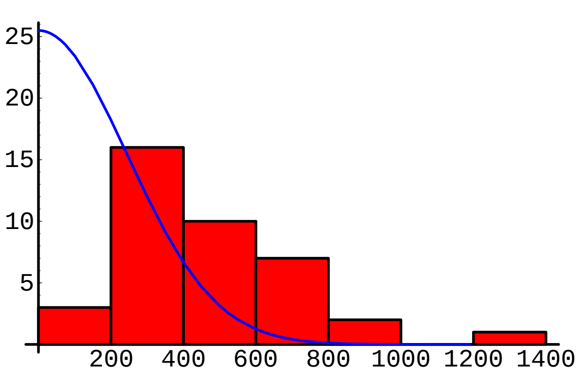

Figure 5.3. Histogram of page lengths of books with the graph of the distribution (5.1) having the same variance.

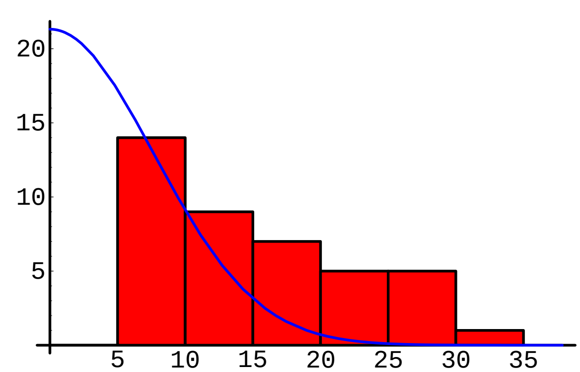

Figure 5.4. Histogram of touchdowns per quarterback with the graph of the distribution (5.1) having the same variance.

The distribution of wealth

Each year *Forbes* publishes a list of the 400 richest individuals in the United States, together with *Forbes'* estimate of their wealths. Figure 5.5 shows the data for 2003 [3]; the horizontal axis is wealth in billions of US\$. For these 400 wealths, the standard deviation is approximately \$4.2B. While it is clear that the distribution (5.1) with the corresponding variance (in blue) does not fit very well near 0 (there are almost twice as many people with wealths between \$600M (the smallest wealth in the top 400) and \$2.5B as this distribution would predict), it fits even worse at large wealths. According to this distribution, the probability that an individual would have a wealth greater than \$30B, is

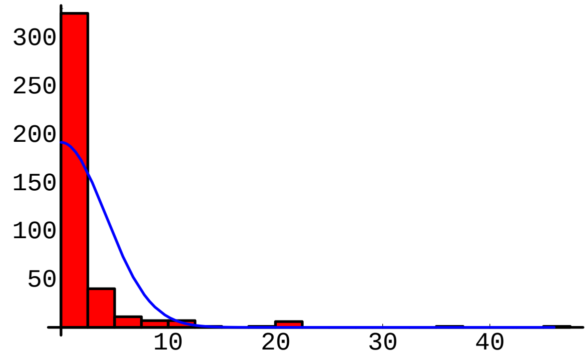

Figure 5.5. Histogram of the wealth of the richest 400 people in the U.S., with the graph of the distribution (5.1) having the same variance.

$$\mathrm{prob}(W > \$30\mathrm{B}) = \int_{30}^{\infty} \frac{2}{\sqrt{2\pi}\sigma} e^{-w^2/2\sigma^2} \mathrm{d}w \approx 6.07253 \times 10^{-13},$$

where $\sigma = 4.16735$. For the U.S. population of about 290M [4], this predicts less than 2×10^{-4} people with wealths above \$30B, which must be contrasted with the existing two: Bill Gates at \$46B and Warren Buffett at \$36B; *i.e.*, it is off by a factor of more than 10^4.

To illustrate this contrast in a way that is perhaps more compelling than Figure 5.5, we plot the data differently. Figure 5.6 shows a plot of wealth as a function of rank on *Forbes'* list: Bill Gates has rank 1, Warren Buffett has rank 2, *etc.* Figure 5.7 shows the same plot for a random sample drawn from the distribution (5.1) with $\sigma = 4.16735$. The two plots are completely different for small ranks, while for large ranks the real wealths decrease more quickly with rank than do the randomly sampled values.

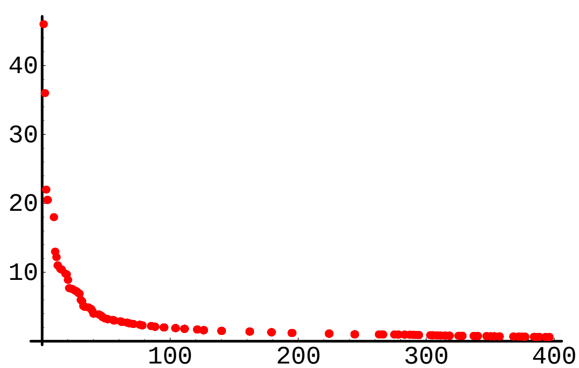

Figure 5.6. The wealths of the richest 400 individuals plotted as a function of their rank.

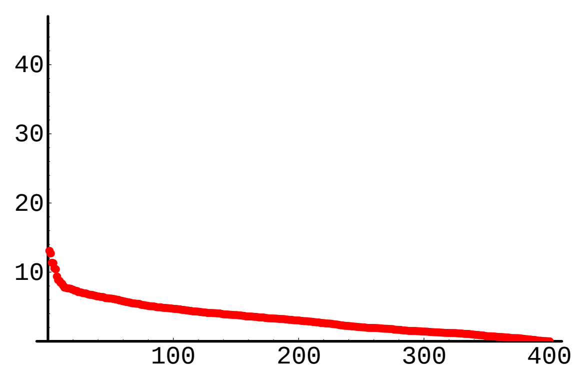

Figure 5.7. 400 values sampled from distribution (5.1) plotted as a function of their rank.

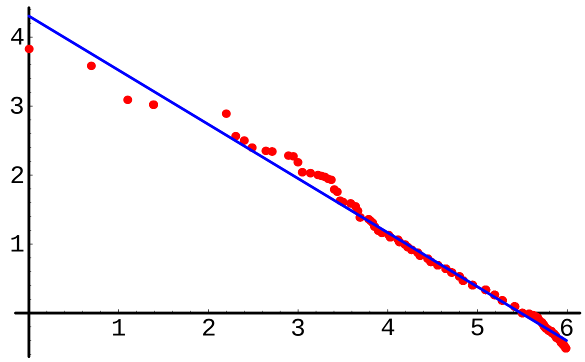

Figure 5.8. Logarithm of the wealth of the richest 400 individuals plotted as a function of logarithm of rank.

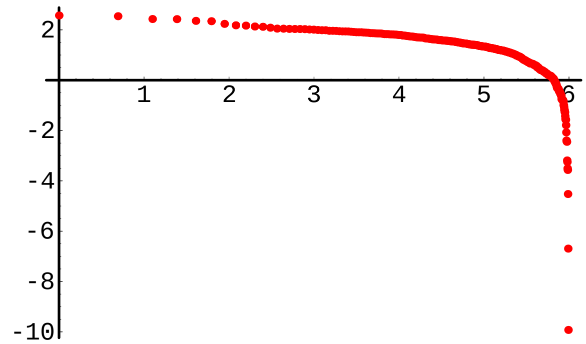

Figure 5.9. Logarithms of 400 values sampled from distribution (5.1) plotted as a function of logarithm of rank.

The contrast is even more compelling visually if these two sets of points are plotted logarithmically. Figure 5.8 shows the logarithm of the wealth plotted as a function of the logarithm of the rank. Similarly, Figure 5.9 shows the logarithm of each sampled value plotted as a function of the logarithm of its rank. The points in Figure 5.8 are well approximated by the line shown (in blue), which has equation

$$\log w = 4.30542 - 0.78502 \log r, \tag{5.2}$$

where w is wealth and r is rank. There is no such approximating line for the points in Figure 5.9.

Homework: Suppose that the distribution of individual wealth in the United States is well approximated by eq. 5.2 across the whole population, not just the richest 400 people. Use this to compute the total individual wealth in the U.S. What fraction is owned by Bill Gates? By the 400 richest people? By the richest 1% of the population? By the richest 10% of the population?

Other distributions

We should be convinced by now that wealths are not distributed according to a normal distribution, or distribution (5.1). There are, of course, many other possible distributions over non-negative values. Next we briefly describe some of the most commonly studied.

Homework: Read Larsen & Marx [1], Chap. 4, for a more complete discussion of these distributions.

Recall from Lecture 3 the example of a gambler flipping a sequence of biased coins, each with probability p of landing head up [5]. The total number of heads in n flips was described by the binomial distribution. There are other random variables that describe other aspects of this example. The first two distributions we consider arise in this context.

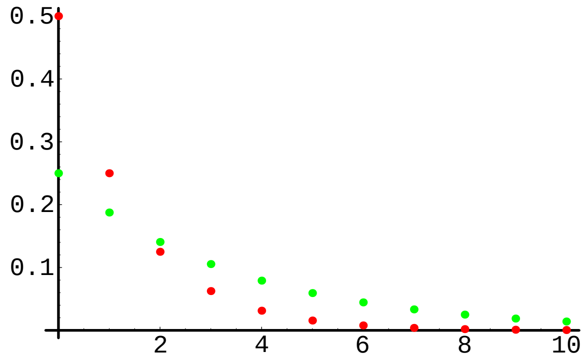

Figure 5.10. Probability functions for geometric distributions with $p = 1/2$ (red), $p = 1/4$ (green).

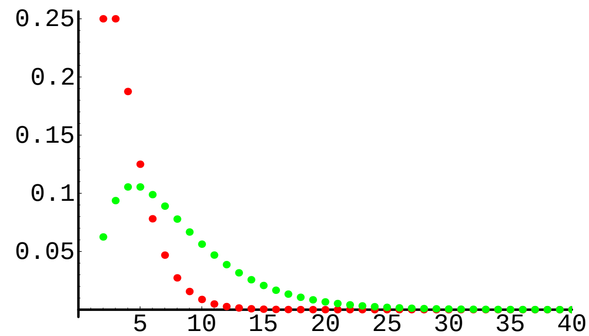

Figure 5.11. Probability functions for negative binomial distributions with $\lambda = 1$ (red), $\lambda = 2$ (green), for $r = 2$ (left) and $r = 4$ (right).

Geometric and negative binomial distribution: For $0 \leq p \leq 1$, let $\text{prob}(T = t) = (1-p)^{t-1}p$, for $t \in \mathbb{Z}_{>0}$. In the sequence of independent biased coin flips, each with probability p of being 1, this is the probability that the first head occurs on the t^{th} flip. Figure 5.10 shows plots of the geometric distribution probability function for $p = 1/2$ (red) and $p = 1/4$ (green). As p decreases, the probability of a short wait decreases, and the probability of a long wait increases. The geometric distribution can be generalized to

$$\text{prob}(T = t) = \binom{t-1}{r-1}(1-p)^{t-r}p^r,$$

for $t \geq 0$. This is the probability that the r^{th} head occurs on the t^{th} flip. For $r = 1$ it is the same as the geometric distribution. Figure 5.11 shows plots of the negative binomial distribution for $r = 2$ with $p = 1/2$ (red) and $p = 1/4$ (green). The probability that $T = 1$ is 0, since there must be at least two coin flips in order to get two heads. For smaller p, shorter waiting times are less likely and longer waiting times are more likely.

Poisson distribution: For $\lambda > 0$, let

$$\text{prob}(W = w) = \frac{e^{-\lambda}\lambda^w}{w!},$$

for $w \in \mathbb{Z}_{\geq 0}$. If events occur randomly at a rate λ per unit of time, the Poisson distribution gives the probability that w occur in a given unit of time. Figure 5.12 shows plots of the probability density functions for Poisson distributions with $\lambda = 1$ (red), $\lambda = 2$ (green) and $\lambda = 4$ (blue). As λ increases, the most likely numbers of occurrences increase.

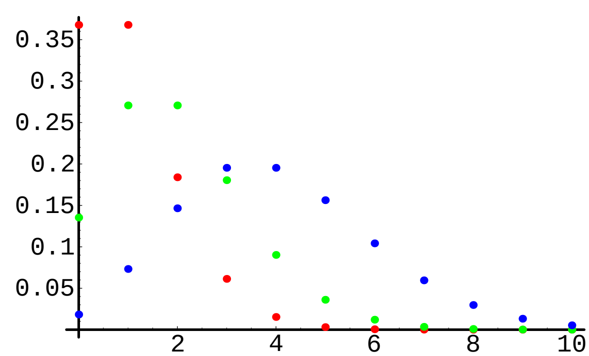

Figure 5.12. Probability density functions for Poisson distributions with $\lambda = 1$ (red), $\lambda = 2$ (green), $\lambda = 4$ (blue).

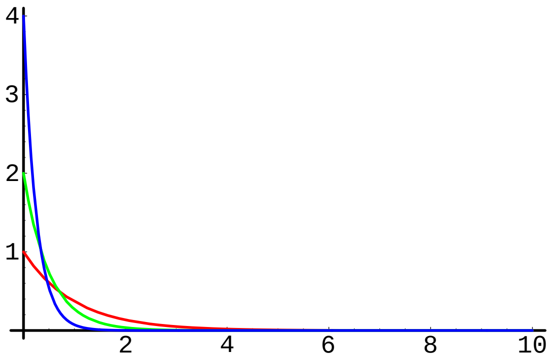

Figure 5.13. Probability density functions for exponential distributions with $\lambda = 1$ (red), $\lambda = 2$ (green), $\lambda = 4$ (blue).

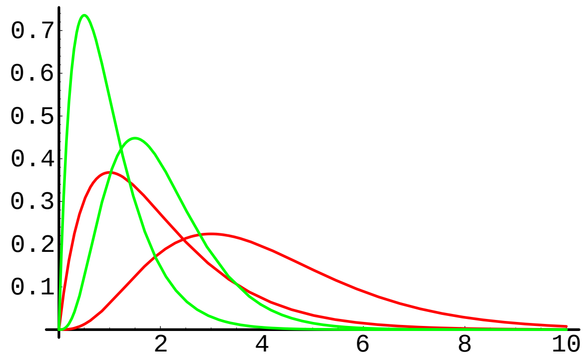

Figure 5.14. Probability density functions for gamma distributions with $\lambda = 1$ (red), $\lambda = 2$ (green), for $r = 2$ (left) and $r = 4$ (right).

Exponential and gamma distribution: For $\lambda > 0$, let $f(t) = \lambda e^{-\lambda t}$, for $t \geq 0$. If events occur randomly at a rate λ per unit of time, this is the probability density function for the waiting time between events. Figure 5.13 shows plots of the exponential distribution probability density function for $\lambda = 1$ (red), $\lambda = 2$ (green) and $\lambda = 4$ (blue). As λ increases, the probability of a short waiting time increases, and the probability of a long waiting time decreases. The exponential distribution can be generalized to

$$g(t) = \frac{\lambda^r}{(r-1)!} t^{r-1} e^{-\lambda t},$$

for $t \geq 0$. This is the probability density function for the waiting time until $r \in \mathbb{Z}_{>0}$ events occur; when $r = 1$ it is just the exponential distribution. Figure 5.14 shows plots for gamma distributions with $\lambda = 1$ (red) and $\lambda = 2$ (green), for $r = 2$ and $r = 4$. As λ increases, there are larger probabilites for smaller waiting times, and as r increases, the probability of a longer waiting time increases.

Homework: Sample 400 points from each of these distributions and plot them as in Figures 5.7 and 5.9. Can you find parameters for any of these distributions so that the sample is similar to the *Forbes* 400 data?

Asymptotic behavior

Despite the possibly apparent similarity between the exponential distribution in Figure 5.13 and the wealth distribution in Figure 5.5, your answer to the homework question above should be "no". None of these distributions have sufficiently 'fat' tails. To be more precise about this statement, we need a little bit of notation.

DEFINITION. Let $f(x)$ and $g(x)$ be two functions defined on some subset of the real numbers that has no upper bound. Then we write $f(x) \sim g(x)$ if and only if

$$\lim_{x \to \infty} \frac{f(x)}{g(x)} = 1.$$

We write $f(x) = \Theta\big(g(x)\big)$ if and only if there is some $0 \neq c \in \mathbb{R}$ such that

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = c.$$

This notation is used to characterize the *asymptotic* behavior of functions. For example, $2x^2 + 3x \sim 2x^2 = \Theta(x^2)$, since as $x \to \infty$, $3x$ becomes insignificantly small compared to x^2.

Homework: Verify this statement using l'Hôpital's rule.

We can compute the asymptotic behavior of the distributions we have introduced in this lecture. For the negative binomial distribution:

$$\binom{t-1}{r-1}(1-p)^{t-r}p^r \sim \frac{t^{r-1}}{(r-1)!}(1-p)^{t-r}p^r = \Theta\big(t^{r-1}(1-p)^{t-r}\big) \quad \text{as} \quad t \to \infty.$$

For the gamma distribution:

$$\frac{\lambda^r}{(r-1)!}t^{r-1}e^{-\lambda t} \sim \frac{\lambda^r}{(r-1)!}t^{r-1}e^{-\lambda t} = \Theta\big(t^{r-1}e^{-\lambda t}\big) \quad \text{as} \quad t \to \infty.$$

For the Poisson distribution:

$$\frac{e^{-\lambda}\lambda^w}{w!} \sim \frac{\lambda^w e^{w-\lambda}}{\sqrt{2\pi w}w^w} = \Theta\bigg(\frac{1}{\sqrt{w}}\Big(\frac{\lambda e}{w}\Big)^w\bigg) \quad \text{as} \quad w \to \infty,$$

where we have used Stirling's formula:

$$w! \sim \sqrt{2\pi w}\Big(\frac{w}{e}\Big)^w \quad \text{as} \quad w \to \infty,$$

the derivation of which would be too great a digression. Each of these functions goes to 0 extremely fast as the argument goes to infinity; none has a 'fat' tail.

There is a probability density function, however, that was invented to describe the distribution of wealth. This is Pareto's distribution:

$$f(x) = \alpha k^\alpha x^{-\alpha-1},$$

for $\alpha > 0$, and where $0 < k \leq x$. Figure 5.15 shows the graphs of the Pareto probability density function for $k = 1$ (in red), $k = 2$ (in green) and $k = 4$ (in blue), with $\alpha = 2$. As k increases, the probability of values larger than k increases. Figure 5.16 shows the graphs of the Pareto probability density function for $\alpha = 1$ (red), $\alpha = 2$ (green) and $\alpha = 4$ (blue),

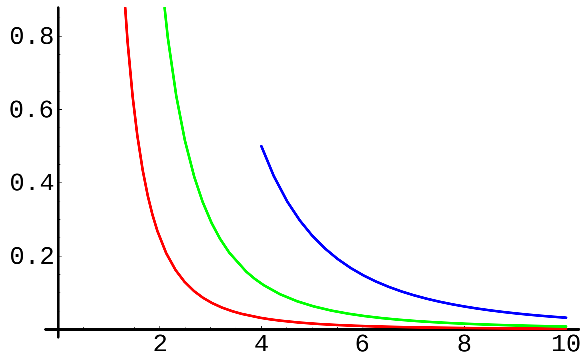

Figure 5.15. Probability density functions for Pareto distributions with $k = 1$ (red), $k = 2$ (green), $k = 4$ (blue) and $\alpha = 2$.

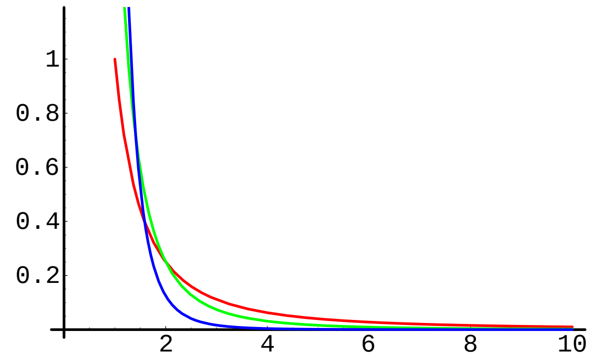

Figure 5.16. Probability density functions for Pareto distributions with $\alpha = 1$ (red), $\alpha = 2$ (green), $\alpha = 4$ and $k = 1$.

with $k = 1$. As α increases, the probability density goes to 0 faster as the argument goes to infinity.

Pareto distributions have 'fat' tails:

$$\alpha k^\alpha x^{-\alpha-1} = \Theta(x^{-\alpha-1}) \quad \text{as} \quad x \to \infty,$$

which goes to 0 much more slowly as $x \to \infty$ than does the normal distribution, or the gamma distribution, or the gamma distribution, or the Poisson distribution.

Homework: Sample 400 points from some Pareto distributions and plot them as in Figures 5.7 and 5.9. Can you find parameters for which the Pareto distribution is similar to the *Forbes* 400 data?
If you collected data that was not well approximated by a normal distribution, plot it as in Figures 5.7 and 5.9. Does one of the other distributions discussed in this lecture appear to approximate your data better than does the normal distribution?

References

[1] R. J. Larsen and M. L. Marx, *An Introduction to Mathematical Statistics and Its Applications* (Upper Saddle River, NJ: Prentice Hall 2001).
[2] D. A. Meyer, "Introduction to Mathematical Modelling: Data collection", http://math.ucsd.edu/~dmeyer/teaching/111winter04/IMM040107.pdf.
[3] "The Forbes 400", *Forbes* (18 September 2003); http://www.forbes.com/richlist2003/rich400land.html.
[4] U.S. Census Bureau, http://www.census.gov/.
[5] D. A. Meyer, "Introduction to Mathematical Modelling: Basic probability theory", http://math.ucsd.edu/~dmeyer/teaching/111winter04/IMM040109.pdf.