

THE STATA JOURNAL

Editor

H. Joseph Newton
Department of Statistics
Texas A & M University
College Station, Texas 77843
979-845-3142; FAX 979-845-3144
jnewton@stata-journal.com

Editor

Nicholas J. Cox
Geography Department
Durham University
South Road
Durham City DH1 3LE UK
n.j.cox@stata-journal.com

Associate Editors

Christopher Baum
Boston College

Rino Bellocco
Karolinska Institutet

David Clayton
Cambridge Inst. for Medical Research

Mario A. Cleves
Univ. of Arkansas for Medical Sciences

William D. Dupont
Vanderbilt University

Charles Franklin
University of Wisconsin, Madison

Joanne M. Garrett
University of North Carolina

Allan Gregory
Queen's University

James Hardin
University of South Carolina

Stephen Jenkins
University of Essex

Ulrich Kohler
WZB, Berlin

Jens Lauritsen
Odense University Hospital

Stanley Lemeshow
Ohio State University

J. Scott Long
Indiana University

Thomas Lumley
University of Washington, Seattle

Roger Newson
King's College, London

Marcello Pagano
Harvard School of Public Health

Sophia Rabe-Hesketh
University of California, Berkeley

J. Patrick Royston
MRC Clinical Trials Unit, London

Philip Ryan
University of Adelaide

Mark E. Schaffer
Heriot-Watt University, Edinburgh

Jeroen Weesie
Utrecht University

Nicholas J. G. Winter
Cornell University

Jeffrey Wooldridge
Michigan State University

Stata Press Production Manager

Lisa Gilmore

Copyright Statement: The Stata Journal and the contents of the supporting files (programs, datasets, and help files) are copyright © by StataCorp LP. The contents of the supporting files (programs, datasets, and help files) may be copied or reproduced by any means whatsoever, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the Stata Journal.

The articles appearing in the Stata Journal may be copied or reproduced as printed copies, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the Stata Journal.

Written permission must be obtained from StataCorp if you wish to make electronic copies of the insertions. This precludes placing electronic copies of the Stata Journal, in whole or in part, on publicly accessible web sites, file servers, or other locations where the copy may be accessed by anyone other than the subscriber.

Users of any of the software, ideas, data, or other materials published in the Stata Journal or the supporting files understand that such use is made without warranty of any kind, by either the Stata Journal, the author, or StataCorp. In particular, there is no warranty of fitness of purpose or merchantability, nor for special, incidental, or consequential damages such as loss of profits. The purpose of the Stata Journal is to promote free communication among Stata users.

The *Stata Journal*, electronic version (ISSN 1536-8734) is a publication of Stata Press, and Stata is a registered trademark of StataCorp LP.

Speaking Stata: Density probability plots

Nicholas J. Cox
Durham University, UK
n.j.cox@durham.ac.uk

Abstract. Density probability plots show two guesses at the density function of a continuous variable, given a data sample. The first guess is the density function of a specified distribution (e.g., normal, exponential, gamma, etc.) with appropriate parameter values plugged in. The second guess is the same density function evaluated at quantiles corresponding to plotting positions associated with the sample's order statistics. If the specified distribution fits well, the two guesses will be close. Such plots, suggested by Jones and Daly in 1995, are explained and discussed with examples from simulated and real data. Comparisons are made with histograms, kernel density estimation, and quantile–quantile plots.

Keywords: gr0012, density probability plots, distributions, histograms, kernel density estimation, quantile–quantile plots, statistical graphics

1 Introduction: how do methods become standard?

Many more methods are invented for analyzing data than ever become part of anybody's standard toolkit. What selection processes determine which methods become popular?

Optimists suppose that all really good ideas will sooner or later become noticed and accepted as sound and useful. Indeed, a good idea may be reinvented several times over: smart people will follow the same logical paths, albeit at different times and in varying circumstances. If Gauss had never got around to publishing least-squares, Legendre did anyway. For optimists, the selection of good methods is virtually Darwinian, but without the messy details of sex and death.

Pessimists suppose that the conservatism inherent in many aspects of teaching and research means that few data analysts look far beyond what they had learned in graduate school or what their peer group currently favors. Quite simply, many people are just too busy to get around to much extra reading about possible new methods. Alternatively, many disciplines show extraordinary willingness to experiment with esoteric new techniques, particularly this season's hottest (or coolest) modeling approaches. At the same time, they show extraordinary inertia over more mundane details of data analysis, especially descriptive or presentational methods.

Historians tell us that the details of selection have often been curious and protracted. Broadly speaking, their stories imply that the optimists and pessimists are both largely correct. Every really good idea in statistical science experienced a glorious take-off when it became widely appreciated. But in retrospect, we can also identify failed precursors, similar or even identical, that did not really get off the ground. The bootstrap, which in a strong sense sprang from the head of Bradley [Efron \(1979\)](#), was anticipated by quite a few other ideas ([Hall 2003](#)). The box plot, which is often attributed to John W. Tukey

(1972, 1977), has a longer history stretching over at least a few decades (e.g., Crowe 1933).

Prosaically, even good simple ideas may need a distinct push to get them moving. The push sometimes requires a lot of energy: every field has its enthusiasts who tirelessly promote a method until people start to take notice. Kenneth E. Iverson (1920–2004) was the principal architect of APL and J and a major influence on many mathematical and statistical programming languages, including Stata. He was well advised by the Harvard computer pioneer Howard H. Aiken (1900–1973): “Don’t worry about people stealing your ideas. If they’re any good, you’ll have to ram them down their throats” (Falkoff and Iverson 1981, 682).

The push is sometimes levered by a clever sales pitch, even down to the marketing details of a good name for the method. “Bootstrap” and “box plot” are cases in point, each hinting strongly at something simple and practical, but interestingly different from what you may know about already. Who can summon up much enthusiasm for learning about dispersion diagrams, the term under which climatologists and geographers used precursors of the box plot? The very name sounds dull and dreary.

Appositely for this journal, the push is sometimes eased by the existence of accessible software. Some discussions of software for data analysis seem based on the premise that trapped inside the body of every data analyst there is a programmer longing to get out. It is perhaps closer to the truth to say that most data analysts would much prefer that someone else did the programming, and for many good reasons.

This fairly simple analysis of a general issue provides context for one of my recurrent themes: methods that somehow or other lurk beyond the bounds of what is standard, perhaps for no good reason except the lack of a suitable push. In the next few columns I will look at some personal favorites, writing about them and their Stata implementations.

2 Empirical and theoretical densities

We start with density probability plots, a graphical method for examining how well an empirically derived density function fits a theoretical density function for a specified probability distribution. Here we are focusing on variables measured on scales that are at least approximately continuous. The terminology of density probability plots may suggest to you that, by accident or perversity, a nonstandard name is being used for what will turn out to be a standard kind of graph, but these plots are not standard at all. They appear to have been used very little in the decade since they were invented (Jones and Daly 1995). We will look first at existing methods for the overall problem. Then we will see better where density probability plots fit into the toolkit.

2.1 Histograms and kernel estimates

Suppose that we wish to compare the observed distribution of a variable Y with a specified distribution X with density function $f(x)$. To the density function, there correspond a distribution function $P = F(x)$ and a quantile function $X = Q(p)$. The distribution function and the quantile function are inverses of each other.

The particular density function $f(x|\text{parameters})$ most pertinent to comparison with data will be computed, given values for its parameters, either estimates from data or values chosen for some other good reason. For example, the parameters of a normal (Gaussian) distribution would usually be the mean and the standard deviation. We might estimate the mean and standard deviation, or we might just want to compare a distribution with a standard normal with mean 0 and standard deviation 1. As good courses and texts explain, such density functions are often superimposed on histograms or compared with density estimates produced using a kernel method. Such comparison amounts to comparing a density function fitted or estimated *globally* (say, a normal with particular mean and standard deviation) with a density function fitted or estimated *locally* (say, densities estimated for particular bins or by a moving kernel).

Spelling out the logic, which is simple and usually informal, there are three key ideas:

1. If the distribution model is a good fit, global and local density estimates will be close, but not otherwise.
2. Assessing what is acceptably close may pose a difficult judgment call, but assessing which of two or more candidate distributions is closer to a representation of sample data is often much easier. That is, it will often be easier to say whether a dataset is closer to a normal or to a gamma or a lognormal than it is to say whether a dataset is close enough to a normal to be declared as such.
3. The procedure is often useful for indicating *how* a model fails to fit, for example, in one or both tails or because of outliers. In particular, this is often clearer from a graph than as a by-product of any formalized or automated attempt to attack the problem as one of measuring degree of fit or one of testing a specific hypothesis.

As a matter of history, over most of the last century or so, histograms have been far more commonly used than kernel estimates. This is not really because kernel estimates are a more recent idea. Both ideas can be traced to the 19th century, kernel estimation to at least as far as C. S. Peirce in 1873 (Stigler 1978). The main inhibition has been that, until very recently, histograms were much easier to produce.

These procedures will be familiar to most readers, but let us remind ourselves how to do it in Stata. Manual references are [R] **histogram**, [G] **twoway histogram**, [R] **kdensity**, and [G] **twoway kdensity**. We will start with problems in which we think we know the answer, taking samples of random numbers from normal and exponential distributions. For reproducibility, set the seed before selecting samples:

```

. set seed 20
. set obs 500
. gen normal_sample = invnormal(uniform())
. label var normal_sample "normal sample"
. gen exp_sample = -ln(uniform())
. label var exp_sample "exponential sample"

```

Note that I used the Stata 9 function name `invnormal()` rather than the name used in previous versions of Stata, `invnorm()`. `invnorm()` continues to work in Stata 9. Readers still using earlier versions will need to use `invnorm()` to replicate this example. Focusing on the sample from the normal, we get the graphs combined in figure 1.

```

. histogram normal_sample, color(none) start(-4) w(0.2)
> plot(function normal = normalden(x,0,1), ra(-4 4))
. kdensity normal_sample, biweight w(0.5) n(500)
> plot(function normal = normalden(x,0,1), ra(-4 4))

```

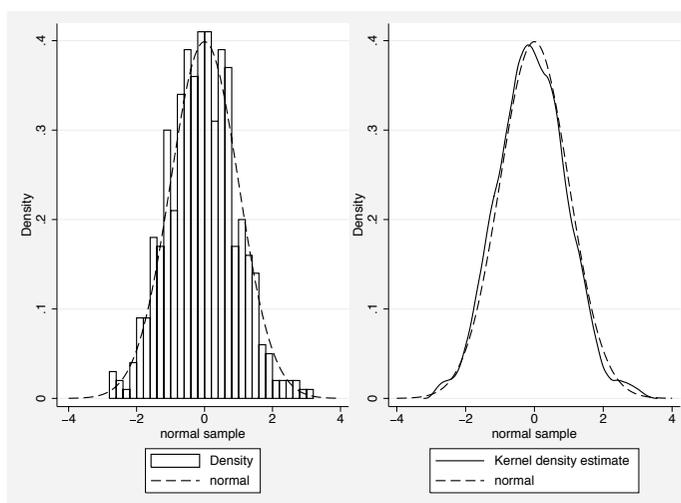


Figure 1: Histogram and kernel estimates of density of a sample of 500 values from a standard normal distribution, shown superimposed

Such an example highlights various familiar points. First, both histograms and kernel estimation require choices, respectively of bin origin and bin width, and of kernel type and kernel width. In particular, if we smooth too much, we throw away detail that might be informative, while if we smooth too little, we might be distracted by detail that is not informative. The choice is often awkward, even with much experience and a well-behaved dataset. Second, we are reminded that even with a moderate sample size—and in this case what should be an excellent fit—empirical distributions can show irregularities that may or may not be diagnostic of some underlying problem.

2.2 Quantile–quantile plots

A common answer to the first point, and to some extent even of the second point, is to think in terms of not densities but quantiles and to produce a probability (meaning, quantile–quantile) plot of observed quantiles versus expected quantiles. The Stata command `qnorm` does this for the Gaussian. See [R] **diagnostic plots**. With our sample, this yields figure 2.

```
. qnorm normal_sample, ms(oh)
```

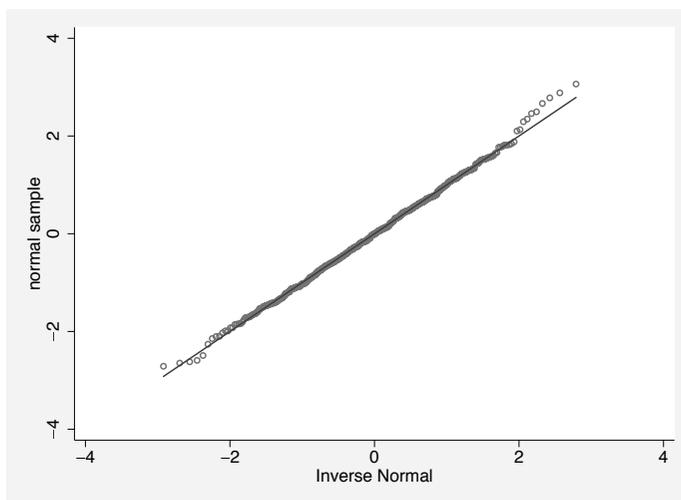


Figure 2: Quantile–normal plot of a sample of 500 values from a standard normal distribution

In this case, other names used are normal quantile plot and probit plot. `qnorm` is not *exactly* what we most need here, as it always uses the mean and standard deviation of the data to fit a normal, rather than externally specified parameters (here mean 0 and standard deviation 1), but that is in practice just a very fine distinction.

Many experienced data analysts would regard the quantile–quantile plot as optimal for this problem. There are almost no awkward choices to be made. The main, and relatively minor, issue is which probability levels or plotting positions are used to evaluate expected quantiles; that is, what p to use in `invnormal(p)`. In addition, plotting observed and expected quantiles on a scatter plot reduces the problem of comparison to one of assessing how far points cluster near the line of equality. This is easier than comparing a curve with a curve. The value of linear reference patterns was emphasized in a previous column (Cox 2004b).

The great merits of quantile–quantile plots are not in dispute here. They are excellent for assessing fit graphically, the task for which they were designed. However, they leave the common desire for a visualization of densities unanswered. This then is the setting for introducing density probability plots.

3 Density probability plots

The density function can also be computed via the quantile function as $f(Q(P))$. Parzen (1979) called this the density-quantile function. (See Parzen [2004] for an overview of his work on quantiles.) For example, if P were 0.5, then $f(Q(0.5))$ would be the density at the median. In practice, P is calculated as the plotting positions p_i attached to values $y_{(i)}$ of a sample of size n , which have rank i ; that is, the $y_{(i)}$ are the order statistics $y_{(1)} \leq \dots \leq y_{(n)}$.

Focusing now on that detail, one simple rule uses $p_i = (i - 0.5)/n$. Most other rules follow one of a family $(i - a)/(n - 2a + 1)$ indexed by a . Although there is literature agonizing about the choice of a , it makes little difference in practice. I tend to use $a = 0.5$ so that $p_i = (i - 0.5)/n$. This can be explained simply as splitting the difference between i/n , which can render the maximum unplotable, as it corresponds to plotting position 1, and $(i - 1)/n$, which can create a similar problem with the minimum, corresponding to plotting position 0.

For our normal example, denote by $\phi()$ and $\Phi()$, as usual, the theoretical density function and distribution function. In Stata 9, `normalden()` is the name for the function previously called `normden()`. The old name continues to work and should be employed by readers still using older versions.

We can thus compute

1. global estimates of the density if normal, given a specified mean and standard deviation, namely, $\phi\{(X - \text{mean})/\text{sd}\}$. In Stata terms, this is `normalden(x, mean, sd)`, and
2. local estimates of the density, namely, $\phi\{\Phi^{-1}(p_i)\}$. In Stata terms, this is `normalden(invnormal(p))`.

As with a quantile–quantile plot, there is only one key graphical decision, the choice of plotting positions. As with previously existing methods, the logic is that if the distribution model is a good fit, global and local density estimates will be close, but not otherwise. The match between the curves allows graphical assessment of goodness of fit and identifying location and scale differences, skewness, tail weight, tied values, gaps, outliers, and so forth.

Such density probability plots were suggested by Jones and Daly (1995). See also Jones (2004). They are best seen as special-purpose plots, like normal quantile plots and their kin, rather than general-purpose plots, like histograms or dot plots. In a program, each distribution fitted (gamma, lognormal, whatever) needs specific code for its quantile and density functions. A Stata implementation is given in the `dpplot` program, published with this column. Fitting a normal is the default. Other distributions that may be fitted include the beta, exponential, gamma, Gumbel, lognormal, and Weibull: some of these require programs to be downloaded from SSC or else a program to be written by the user if parameters are to be estimated from the data, as will usually be desired. In addition, Stata users with some programming expertise should be able to

modify existing programs to fit any favorite distribution that is not supported and also to write extra programs compatible with `dplot` to carry out the necessary calculations.

For our normal problem, the minimum necessary is to specify the parameter values of interest:

```
. dplot normal_sample, param(0 1) ms(oh)
```

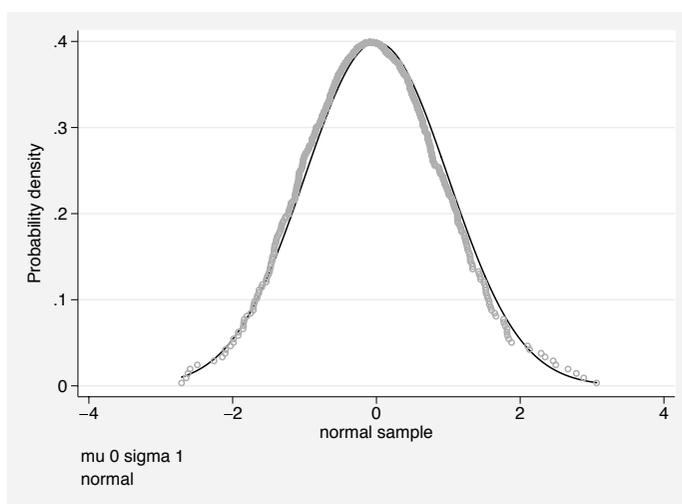


Figure 3: Density probability plot of a sample of 500 values from a standard normal distribution showing fit to that standard normal

Figure 3 gives another graphical take on the problem. The fit would usually be declared good, but it takes some experience, or some theory, to know that the slight irregularities in the tails are not enough to worry about.

Turning to the exponential sample, the essential logic is identical. We need to plug in a value for the mean, the single parameter, to compute

1. global estimates of the density if exponential, namely, $(1/\text{mean}) \exp(-X/\text{mean})$. In Stata terms, this is $(1/\text{mean}) * \text{exp}(-X/\text{mean})$, and
2. local estimates of the density. The quantile function is

$$X = -\text{mean} \ln(1 - P)$$

and the density function, as above, is

$$(1/\text{mean}) \exp(-X/\text{mean})$$

Telescoping the two, this is

$$(1/\text{mean}) \exp \left[- \{ -\text{mean} \ln(1 - P) \} / \text{mean} \right]$$

which reduces easily to $(1 - P)/\text{mean}$.

Our mean is tacitly 1, so this is all done for us by one command, yielding figure 4.

```
. dplot exp_sample, dist(exp) param(1) ms(oh)
```

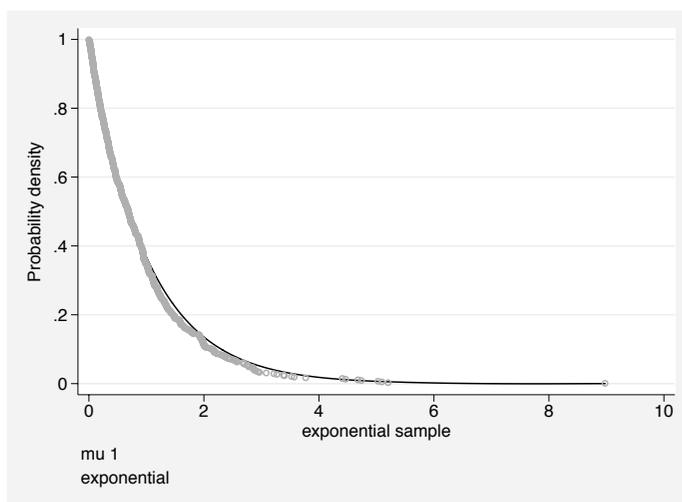


Figure 4: Density probability plot of a sample of 500 values from an exponential distribution with mean 1 showing fit to that distribution.

In this case, the local estimates droop somewhat below the theoretical curve. The fact that the sample mean is rather lower than 1 at 0.942 is pertinent here. Every now and then, our random sample will not match the ideal well, “just by chance”.

4 An example with real data

Let us look at an example with real data, on 158 glacial cirques from the English Lake District ([Evans and Cox 1995](#)). These were examined in a previous column ([Cox 2004a](#)). Glacial cirques are hollows excavated by glaciers that are open downstream, bounded upstream by the crest of a steep slope (wall), and arcuate in plan around a more gently sloping floor. More informally, they are sometimes described as “armchair-shaped”. Glacial cirques are common in mountain areas that have or have had glaciers present. Three among many possible measurements of their size are length, width, and wall height. As a first stab at fitting specific distributions, we consider three two-parameter distributions, namely,

1. The gamma distribution is parameterized using a scale parameter β and a shape parameter α so that the density function is

$$\frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right)$$

Here $\Gamma(u)$ is the gamma function $\int_0^\infty t^{u-1} \exp(-t) dt$.

2. The lognormal distribution is parameterized using a location parameter, the mean of the logs μ , and a scale parameter, the standard deviation of the logs σ , so that the density function is

$$\frac{1}{X\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\ln X - \mu)^2}{2\sigma^2}\right\}$$

3. The Weibull distribution is parameterized using a scale parameter β and a shape parameter γ so that the density function is

$$\frac{\gamma}{\beta} \frac{x^{\gamma-1}}{\beta^\gamma} \exp\left\{-\left(\frac{x}{\beta}\right)^\gamma\right\}$$

Note that these parameterizations all differ from those employed by `streg`. See [ST] `streg`. In each case, the distribution was fitted using maximum likelihood.

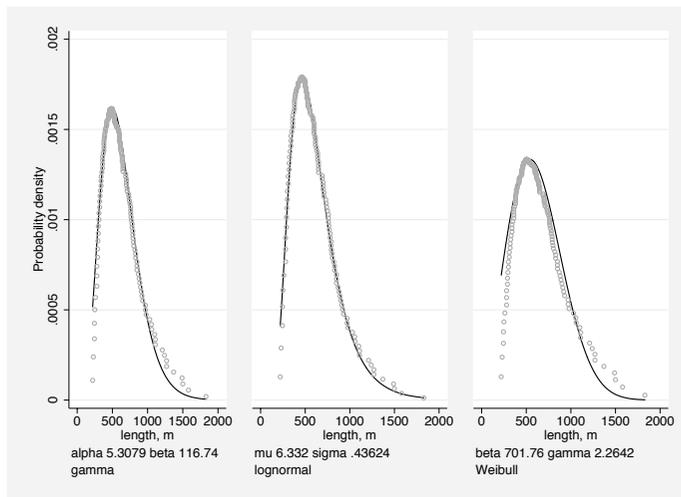


Figure 5: Density probability plots showing fits of cirque length data to gamma, log-normal, and Weibull distributions

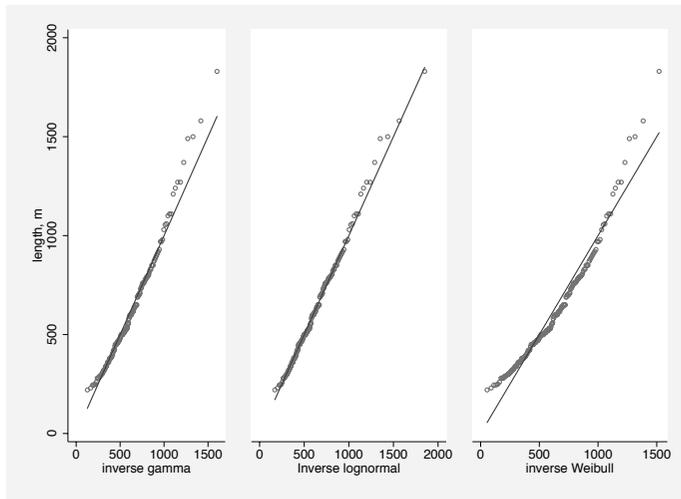


Figure 6: Quantile–quantile plots showing fits of cirque length data to gamma, lognormal, and Weibull distributions

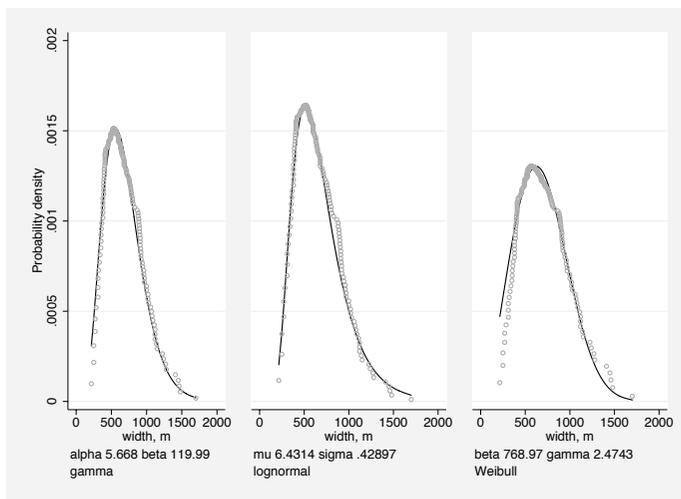


Figure 7: Density probability plots showing fits of cirque width data to gamma, lognormal, and Weibull distributions

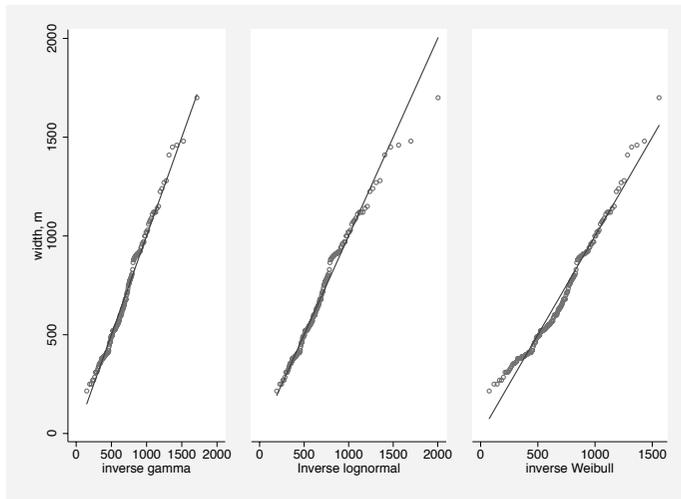


Figure 8: Quantile–quantile plots showing fits of cirque width data to gamma, lognormal, and Weibull distributions

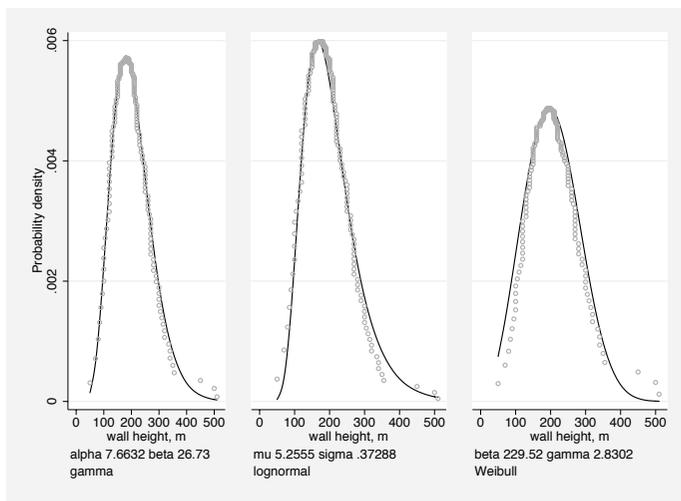


Figure 9: Density probability plots showing fits of cirque wall height data to gamma, lognormal, and Weibull distributions

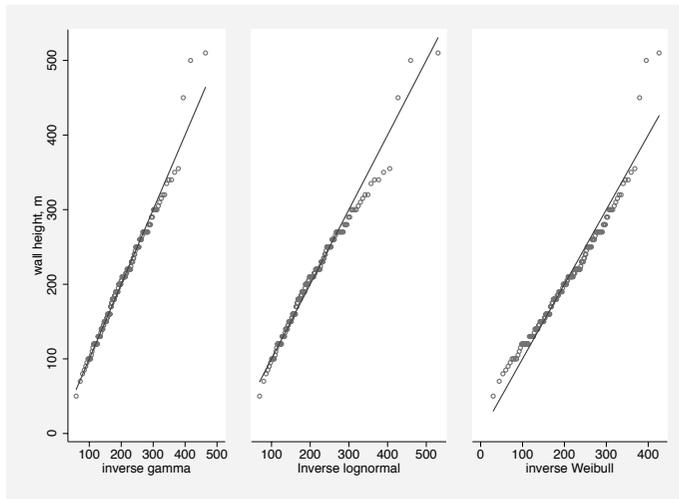


Figure 10: Quantile–quantile plots showing fits of cirque wall height data to gamma, lognormal, and Weibull distributions

Figures 5 through 10 show the results of fitting these distributions to the three size variables, using both density probability plots and quantile–quantile plots. The Weibull is a poor fit in every case, so we will say no more about it. In the fight between the gamma and the lognormal, the latter seems to have the edge for length (figures 5 and 6) and the former for width (figures 7 and 8) and wall height (figures 9 and 10). Three moderate outliers for wall height are enigmatic, however they are viewed.

This is not a complete story. Many would feel it worth exploring three-parameter distributions bringing in nonzero thresholds. Cirques cannot be arbitrarily small (i.e., arbitrarily close to zero in size), as the glaciers that form them cannot be: tiny snow patches do not become glaciers without growing first. But the main message here is one of method, suggesting that density probability plots can nicely complement quantile–quantile plots.

5 The modality of the distribution fitted is always shown

Many users of density probability plots are surprised by one of their features. A little thought shows that this feature is an inevitable consequence of what is being requested, but it is worth explanation.

Suppose that we do something rather silly: we plot the exponential sample to see how far it is normal. To mix silliness and sense, we will use the theoretical principle that the mean and standard deviation should both be 1:

```
. dpplot exp_sample, param(1 1) ms(oh)
```

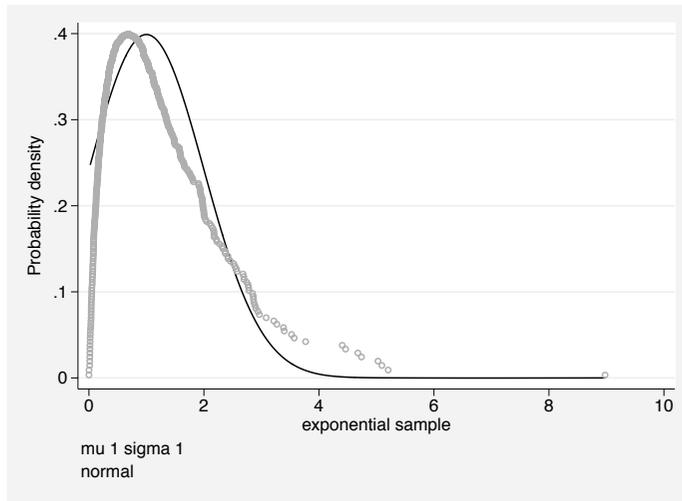


Figure 11: Density probability plot of a sample of 500 values from an exponential distribution with mean 1 showing fit to a normal distribution with mean 1 and standard deviation 1

The plot in figure 11 shows that the fit is poor, but that should be a foregone conclusion. The surprise may be that the empirical or local density estimate is strongly unimodal with definite tails on both left and right. It does not look Gaussian, but even less does it look exponential. How does that shape arise?

The key here is that what you are seeing is the exponential sample trying its very hardest to look as normal (Gaussian!) as possible. That is, you are seeing a normal density, but continuously transformed. When data points are closer together, or further apart, than they would be if the data were really normal, then the normal is also squeezed or stretched accordingly. But those deformations affect horizontal position only, and the density ordinates are exactly as calculated by plugging into `normalden()`. The modality remains that of the normal distribution being fitted.

More generally with density probability plots, results may be difficult to decipher if observed and reference distributions differ in modality or other aspects of gross shape. So if the density function of the distribution being fitted is monotone decreasing or unimodal with two tails, then so too must be $f(Q(P))$, whatever the shape of $f(Y)$.

The practical implication is that when the fit between empirical and theoretical densities is really poor, a density probability plot may well be obscure on quite why and how the fit is poor. The last example was silly on purpose, but there is little to stop even experienced data analysts doing silly things by accident. In particular, density probability plots are not especially suitable for very naive users.

6 Virtues and vices

Extending the discussion in [Jones and Daly \(1995\)](#), the advantages of these plots include

- *Ease of interpretation.* Some people find them easier to interpret than quantile–quantile plots.
- *Fewer awkward choices.* No choices of binning or origin (cf. histograms, dot plots, etc.) or of kernel or of degree of smoothing (cf. density estimation) are required.
- *Flexibility with regard to sample size.* They work well for a wide range of sample sizes. Nevertheless, just as with all other methods, a sample of at least moderate size is preferable (say, ≥ 25).
- *Bounded support is clear.* If X has bounded support in one or both directions, then this should be clear.

The disadvantages include

- *Modality may not match.* Results may be difficult to decipher if observed and reference distributions differ in modality.
- *Tails may be cryptic.* It may be difficult to discern subtle differences in one or both tails of the observed and reference distributions. On the other hand, tails are not always crucial, and it is arguable that quantile–quantile plots may have the opposite weakness of overemphasizing tails.
- *Comparison of curves.* Comparison is of a curve with a curve: some people argue that graphical references should, where possible, be linear (and ideally horizontal). On the other hand, `dpplot` has a `generate()` option to save its results and a `diff` option to show the difference between the two guesses at the density function.
- *Not extensible.* There is no simple extension to comparison of two samples with each other.

7 Conclusions

Density probability plots use sample data to show how close a data sample is to a specified distribution. They portray two guesses at the density function, one produced globally, usually by estimating the parameters of the distribution, and one produced locally by evaluating the density-quantile function at the data points. Although proposed a decade ago, they appear to have been used very little, and this column has provided further publicity, a pointer to a Stata implementation, and a discussion of their advantages and limitations. In particular, they avoid the choices of bin or kernel width imposed by histograms or kernel density estimation, and they usefully complement quantile–quantile plots in conveying what distribution is being fitted, as well as how well that distribution fits.

8 Acknowledgments

Ian S. Evans, Stephen Jenkins, and Tim Sofer aided this project in various ways.

9 References

- Cox, N. J. 2004a. Speaking Stata: Graphing distributions. *Stata Journal* 4(1): 66–88.
- . 2004b. Speaking Stata: Graphing agreement and disagreement. *Stata Journal* 4(3): 329–349.
- Crowe, P. R. 1933. The analysis of rainfall probability. *Scottish Geographical Magazine* 49: 73–91.
- Efron, B. 1979. Bootstrap methods: another look at the jackknife. *Annals of Statistics* 7: 1–26.
- Evans, I. S. and N. J. Cox. 1995. The form of glacial cirques in the English Lake District, Cumbria. *Zeitschrift für Geomorphologie* 39: 175–202.
- Falkoff, A. D. and K. E. Iverson. 1981. The evolution of APL. In *History of Programming Languages*, ed. R. L. Wexelblat, 661–691. New York: Academic Press.
- Hall, P. 2003. A short prehistory of the bootstrap. *Statistical Science* 18: 158–167.
- Jones, M. C. 2004. Hazelton, M. L. 2003, “A graphical tool for assessing normality”, *American Statistician* 57: 285–288: Comment. *American Statistician* 58: 176–177.
- Jones, M. C. and F. Daly. 1995. Density probability plots. *Communications in Statistics, Simulation, and Computation* 24: 911–927.
- Parzen, E. 1979. Nonparametric statistical data modeling. *Journal of the American Statistical Association* 74: 105–131.
- . 2004. Quantile probability and statistical data modeling. *Statistical Science* 19: 652–662.
- Stigler, S. M. 1978. Mathematical statistics in the early states. *Annals of Statistics* 6: 239–265.
- Tukey, J. W. 1972. Some graphic and semigraphic displays. In *Statistical Papers in Honor of George W. Snedecor*, ed. T. A. Bancroft and S. A. Brown, 293–316. Ames, IA: Iowa State University Press.
- . 1977. *Exploratory Data Analysis*. Reading, MA: Addison–Wesley.

About the Author

Nicholas Cox is a statistically minded geographer at Durham University. He contributes talks, postings, FAQs, and programs to the Stata user community. He has also co-authored fifteen commands in official Stata. He was an author of several inserts in the *Stata Technical Bulletin* and is an Editor of the *Stata Journal*.