Estimating the Quantile Function

J. O. Ramsay, G. Hooker and J. Nielsen

March 13, 2010

Abstract

The quantile function Q provides direct answers to questions about the size of an observation associated with a pre-specified risk, and as such is often of more value to users of statistics than either the distribution or density function. Nonparametric methods for estimating the quantile function are explored, using a representation of Q as the solution of a differential equation. It is shown that this framework is useful for computation and estimation, and also offers an elegant expression of the relationship between various functional descriptions of univariate variation. Moreover, the transformation of probability u to surprisal $S(u) = -\log_2(1-u)$ bring further simplicity at both the computational and formal levels. The use of these methods is illustrated with some challenging rainfall data.

1 Introduction

The quantile function $x = Q(u)$ specifies the random variable value x that will not be exceeded with probability u, and is the functional inverse of the probability distribution function $u = F(x)$ that specifies u that x will not exceed. Its derivative $q = DQ$ was called the *sparsity* function (Tukey, 1962), and the function $f \circ Q$ was referred to by Parsen (1979) as the *density-quantile* function in order to contrast it with q, which he called the *quantile-density* function. The reciprocal relationship $q(u) = 1/(f \circ Q)(u)$ between the *sparsity* and *density-quantile* functions is easily seen by differentiating $(F \circ Q)(u) = u$. We will use superscripts as mnemonics for textbook distributions, as in $(F^E, Q^E), (F^N, Q^N), (F^W, Q^W)$, and (F^G, Q^G) for the exponential, normal, Weibull and gamma distributions, respectively.

Q's and q's focus our attention on data extremes; F's and f's highlight the usual. Consequently, someone interested in risk will prefer to look at Q or q; a tourist might want to know, "What's the size of a storm that I would only see once is a century ($Q(0.01)$) when I visit Florida this October?" The quantile function in Figure 1 displays the distribution of 355 days of measurable June rainfall in the Canadian prairie city of Regina over 34 years of data, and clearly shows that two rainfalls were exceptional; the larger of these flooded 20,000 basements in 1975. The popularity of the $q-q$ plot is due to its ability to show how data extremes fail to conform to a model distribution, and the false discovery rate for revealing significant outcomes in large numbers of small-sample experiments may be viewed as an application of the $q-q$ plot for the uniform distribution. Finally, many real-life distributions are compound in nature, with tail behavior being due to the occasional intervention of a process quite different from that which generates the bulk of the data. Large convective thunderstorms can now and then coalesce and yield unforgettable weather; such an event in January of 1998 closed downtown Montreal for a week, and is tragically familiar in cities like New Orleans.

The quantile function was championed by Tukey (1962, 1977) as a functional summary of risk. Parzen (1979, 1997, 2004) explored theoretical and applied properties of Q in the context of testing goodness of fit, and many other contributions to the literature deal with random variate generation (Devroye, 1986). Gilchrist (2000) offered an especially readable and application-oriented account of data modelling using quantile functions associated with common parametric distributions. There is also a large literature on the estimation of pre-assigned quantiles and quantile regression, summarized in

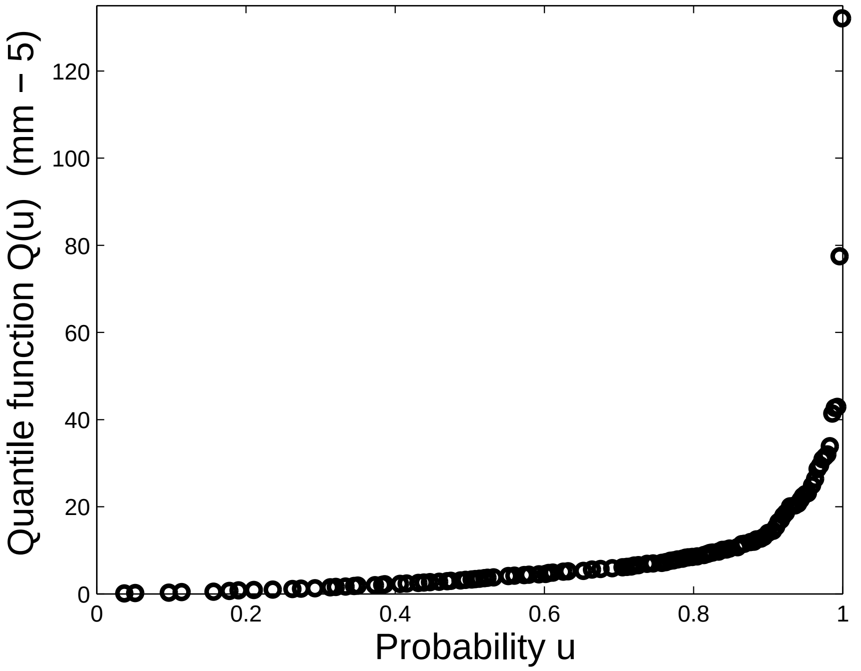

Figure 1: The empirical quantile function $\tilde{Q}_u(u)$ for rainfall exceeding 0.5 mm in June in Regina, Saskatchewan, Canada over the years 1960-1994.

(Koenker, 2005).

Our focus is on parametric and nonparametric methods, or fixed- and open-dimensional techniques, for the estimation of the entire quantile function. We will assume for simplicity that the quantile function is both differentiable and strictly monotonic, but the methods that we propose are easily adaptable to quantile functions with flat spots, corresponding to distributions without connected support.

Estimating Q involves three formidable challenges. The first of these is the monotonicity of Q, and we deal with his by representing Q as the solution of a differential equation. This, it turns out, reveals a surprising functional anti-symmetry connecting Q and F and their derived functions. This defines some effective computing strategies, and also suggests some new families of low-dimensional parametric distributions.

Interesting Q's tend also to display extreme and localized curvature. Because of its monotonicity, too much of Q can be packed against the vertical boundaries of a plot, as in Figure 1, to be considered effective graphical methodology. Moreover, common nonparametric curve fitting methods have a great deal of trouble fitting these near–vertical portions of Q while preserving its roughly linear behavior elsewhere. We cope with this problem by transforming probability u to *surprisal*, $S(u) = -\log_2(1-u)$ and $U(s) = 1 - 2^{-s}$, we use the notation Q_u and Q_s to refer to quantile values plotted against u and s, respectively. Switching to Q_s not only tends to straighten the relation $Q_s(s) = (Q_u \circ U)(s)$; it also simplifies a number of parametric expressions of Q_u's corresponding to pivotal distributions, such as the exponential. We see in Figure 2 that plotting ordered rainfalls against surprisal reveals that, in addition to the two extreme rainfalls, there is a surge in rainfall intensities for $s \in [4, 6]$. This is due to large convective precipitations in the form of thunderstorms or blizzards that represent a different process than the more usual drizzle. The two extreme rainfalls seem not to be consistent with the rest of the data in the sense of being more than two standard errors away from the dashed line, a reasonable parametric quantile function described in Section 2.

The third estimation challenge is the severe change in variation over u or s of the order statistics $X_{(i)}, i = 1, ..., N$ from which a quantile function is to be computed. The asymptotic variance of $X_{(i)}$ is $U_{(i)}(1 - U_{(i)})q(U_{(i)})$ and $(1 - U_{(i)})q_s[S(U_{(i)})]$, $U_{(i)}N = i, i < N$, over u and s, respectively. We see in Figure 1 that the value of sparsity q_u can be huge for extreme data values. Consequently, we seek to control the high-variance portions of Q_u or Q_s by

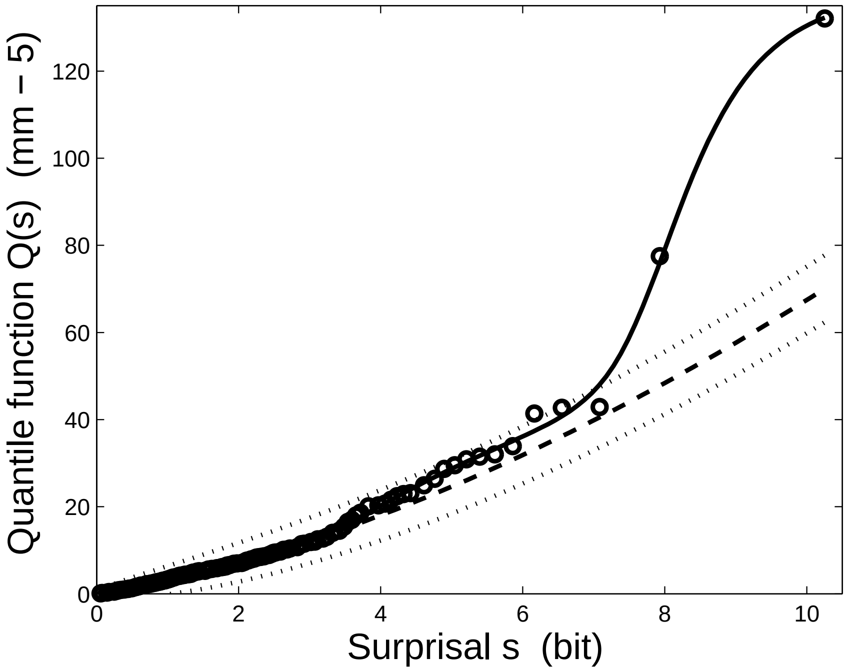

Figure 2: The empirical quantile function $\tilde{Q}_s(s)$ for rain in Regina plotted as a function of surprisal $S(u) = -\log_2(1-u)$, along with a regularized least squares fit (solid line), a fit using a modified Weibull quantile (dashed line), and two standard deviations of order statistics away from this fit (dotted lines).

connecting in a smooth way the information in high-density regions to tail behavior. We also consider parameterizations of the quantile function that effectively separate the modeling of smooth parts of Q from their extremes by connecting quantile estimation to *curve registration* methods that distinguish phase from amplitude variation (Kneip and Ramsay, 2009; Ramsay and Silverman, 2005).

The following two subsections review some of the properties of quantile functions and surprisal.

1.1 The quantile function and its properties

Quantile functions are available in parametric form for many textbook distributions, and central to our discussion is that for the exponential distribution, $Q^E(u) = -\tau \log(1-u)$. In Section 2 we will extend the Weibull distribution, for which $Q^W(u) = \tau[-\log(1-u)]^{1/\alpha}$. Measures of the location, scale and shape can be computed directly from the quantile function. The mean and variance of x are

$$\mu = \int_0^1 Q(u)du \quad \text{and} \quad \sigma^2 = \int_0^1 [Q(u) - \mu]^2 du$$

and the median and semi-interquartile range are $Q(0.5)$ and $[Q(0.75) - Q(0.25)]/2$. See Gilchrist (2000) for measures of skewness and other shape descriptors.

New quantile functions and their associated distributions can be generated by various manipulations of $Q(u)$, all of which essentially preserve monotonicity. If X and Y are random variables with quantile functions Q_X and Q_Y, respectively, then

- $a + bQ_X(u)$ is the quantile function for $a + bX$ for a and $b > 0$ fixed constants.

- The quantile function for $aX + bY$ is $aQ_X(u) + bQ_Y(u)$.

- $-Q(1-u)$ is the quantile function for $-X$, and is called the *reflection* of Q. For example, the quantile function for the exponential distribution is $Q_E(u) = -\ln(1-u)$ and its reflection is thus $Q_{-E}(u) = \ln(u)$.

- $1/Q(1-u)$ is the quantile function for $1/X$

- $Q^\alpha(u)$ is the quantile function for $X^\alpha, \alpha > 0$. For example, the quantile function for the uniform distribution is u itself, and u^α is the quantile function for the *power* distribution.

- A convex linear combination of quantile functions is also a quantile function. For example, if we add the exponential and the reflected exponential quantiles

  $$Q_\ell(u) = -\ln(1-u) + \ln(u) = \ln[u/(1-u)],$$

  we obtain the quantile function for the *standard logistic* distribution with density $f(x) = 1/(1+e^{-x})$.

- If two variates are both positive, the product of their quantile functions is a quantile function. For example, the Pareto distribution has $Q(u) = 1/(1-u)^\beta$, and the product of the power and Pareto quantiles, $Q(u) = u^\alpha/(1-u)^\beta$, is the *power–Pareto* distribution.

The first two properties are critical, and imply that quantile functions are apt to be far more convenient in data modeling than either distribution or density functions. For example, we can compare the linear operation $aQ_X(u)+bQ_Y(u)$ with the convolution operation $f_Z(z) = \int f_X(z-y)f_Y(y)dy$ required to find the density of $Z = X+Y$. We will provide a simple process for passing from the quantile function for a sum to the corresponding density.

The quantile function immediately gives us the behavior of the *largest* value $X_{(N)}$ in a random sample of size n, since the quantile function for the extreme value is

$$Q_{(N)}(u) = Q(u^{1/N}).$$

so that the median extreme value from a sample of size n is $Q(0.5^{1/n})$. The quantile function for the smallest value is by symmetry therefore

$$Q_{(1)}(u) = Q(1-(1-u)^{1/n}).$$

Quantile functions for other order statistics can be computed by procedures that are only slightly more complicated.

Quantile functions are especially convenient when generating samples of parameter estimates by bootstrapping, data simulation or other methods. Since we often do this in order to estimate confidence intervals, defined as quantiles of the distribution of parameter estimates, we can estimate the quantile function Q for these estimates, and read the required limits off directly [Quantiles of the pivotal quantity?]. Methods for estimating quantiles functions considered below can be applied to data gained from resampling to estimate quantile functions for parameter estimates.

Finally, of special importance is the fact that, if two quantile functions Q_1 and Q_2 have the same range, then we can transform each quantile function into the other in one or both of two ways:

- A monotone function g of a quantile function $g[Q_u(u)]$ is a quantile function, and
- If h is a monotone function such that $h(0) = 0$ and $h(1) = 1$, then $Q_u[h(u)]$ is also a quantile function.

These properties apply to Q_s as well. The first of these choices changes the vertical scale of Q, and represents what is called *amplitude variation* in signal analysis or functional data analysis. The second changes the location of quantile function features, and therefore represents *phase variation.* Blending these two transformations can be a useful way to transform a relatively simple or low-dimensional Q to be more faithful to the data [maybe something like: "Q to be more adaptable for data fitting"?].

1.2 Surprisal and its properties

The unbounded nonnegative quantity $S(p) = -\log_2(p)$ maps a probability $p \in [0, 1]$ to the positive real line, and is often called by the evocative name *surprisal*, a term first used by Tribus (1961). The more technical name for *surprisal* is *self-information* (Shannon, 1948; Cover and Thomas, 2006). When the logarithm is taken with respect to base 2, the unit of information is the *bit*. For example, $-\log_2(1 - 0.95) = -\log_2(0.05) = 4.32$ bits, the equivalent of getting an average of 4.32 consecutive heads in coin tossing. If $p(x)$ is the value of a probability density function, then $s(x) = -\log_2 p(x)$ is an *infinitesimal surprisal* in the same sense that $p(x)$ is an infinitesimal probability. Considering these infinitesimals as independent leads to the definition of the *entropy* of a distribution $H(X) = E[s] = \int p(x)s(x)dx$, the expected infinitesimal surprisal. The surprisal distribution function is defined by $F_s(x) = S[F_u(x)]$, and the corresponding surprisal density function relates to the probability density function through $f_s(x) = \frac{du}{ds} f_u(x)$.

In this paper we use surprisal to express the information conveyed by an event with right-tail probability $p = 1-u$, so that $S(u) = -\log_2(1-u)$. In this form surprisal increases with u, and quantifies the strangeness or extremity of a large observation in terms of bits. The surprisal of a nonnegative variable x with the value $X = 0$ is 0, since it implies simply that the response will be a non-negative real number, and as such conveys no information about the distribution. Because we have to switch between the natural and the base 2 logarithms, we define here the constant $C_2 = \log(2) \approx 0.69$, so that $\log p = C_2 \log_2 p$.

Surprisal as a measure of information has a natural zero, adds over independent events and therefore has the algebraic structure of magnitudes. Since the human brain has evolved to manipulate the magnitudes of every day life, it is quite likely that surprisal is an easier medium to think in than probability, with its multiplicative algebra that is ill-suited to mental manipulation.

Quantile functions exhibit less curvature when plotted as a function of s, and especially so for the exponential $Q_u(u) = -\tau \log(1-u)$, which becomes $Q_s(s) = \tau s$. As we move through the gamma distribution $\Gamma(\alpha, \beta)$ by increasing α, $\alpha = 1$ being the exponential distribution, the distribution becomes progressively more short-tailed. The top panels in Figure 3 display the quantile function over both u and s; we see that the curvature in Q_s becomes negative for $\alpha = 2$. We also see the close connection between Tukey's box and whisker diagram and the quantile function; the quartile observations correspond to $s = 0.415, 1$ and 1.39 bits, respectively; and the box over s is squeezed to the left to emphasize right tail behavior at the expense of central tendency. For either argument the quantile curve passes through the lower left and upper right corners of the box. The bottom panels of Figure 3 show that the Gaussian quantile function Q_s^N is nearly logarithmic, with an even stronger negative curvature on the right because its tail is shorter than that of a gamma variate.

The parallels between surprisal $S(u)$ and quantile functions $Q(u)$ seems striking. Both have the same argument, both are closed under addition, both have natural units (at least as a rule for quantiles), both have natural interpretations in terms of real-world experience, and both tend to exhibit accelerating curvature as a function of u and with curvatures that are surprisingly close, as we saw with the exponential distribution. The connection of surprisal with the exponential distribution is further emphasized by the fact that, if variable X has a distribution defined by F, then S has an ex-

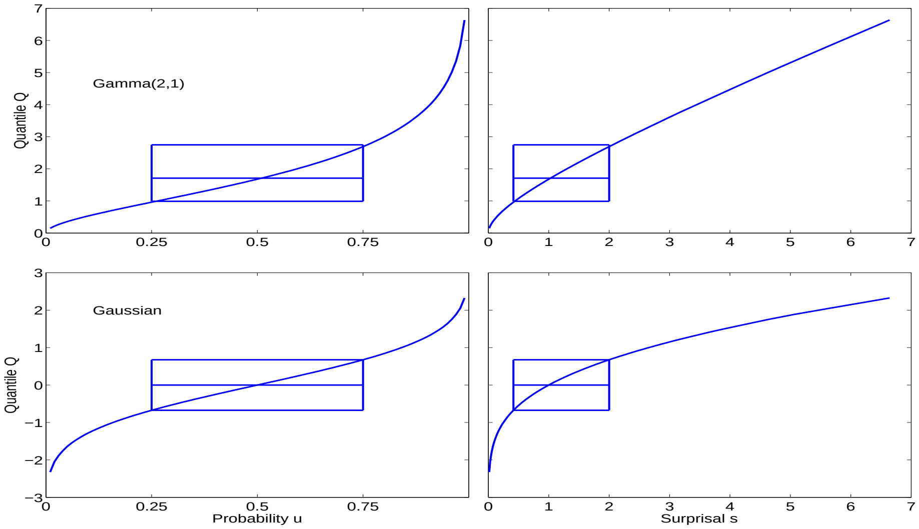

Figure 3: The top left panel displays the quantile function for the gamma distribution $\Gamma(2,1)$ as a function of u, and the top right as a function of s. The bottom panels display the corresponding quantile functions for the standard Gaussian distribution. The boxes are from Tukey's box and whisker diagram, with horizontal lines corresponding to quartile values.

ponential distribution with rate parameter $1/C_2 = 1.4427$, and the values of the composite function $Q_s \circ S$ is also exponentially distributed with rate parameter $\tau/\log(2)$. The exponential distribution, in short, is the functional origin for the functions Q_s. We will see further parallels in the next two sections, where we introduce a functional representation for Q_u and Q_s and a set of differential equations that are useful as both computational tools and as revealers of these links.

2 A representation of Q and a modified Weibull distribution

If, as we assume, Q_u is a strictly differentiable monotone function, then it can be represented as

$$Q_u(u) = \alpha + \beta \int_0^u \exp[W(v)]dv \tag{1}$$

where $W(v)$, the (shifted) *log-sparsity function*, is an unrestricted function (Ramsay, 1996; Ramsay and Silverman, 2005). The representation is even simpler if $x \in [0, \infty)$, since

$$Q_u(u) = \int_0^u \exp[W(v)]dv. \tag{2}$$

A change of variable from u to s yields

$$Q_s(s) = \alpha + \beta \int_0^s \exp[W_s(v)]dv \text{ where } W_s(s) = W[U(s)] - C_2 s + \log C_2. \tag{3}$$

Since the log-sparsity functions W_u and W_s are not constrained in any way, they can be expressed as a linear combination of basis functions:

$$W_u(u) = \sum_{k=1}^{K} c_k \phi_k(u) \text{ and } W_s(s) = \sum_{k=1}^{K} c_k \psi_k(s).$$

This perspective can attract our attention to useful generalizations and modifications of familiar distributions. For example, the Weibull distribution, which contains the exponential distribution, has distribution, quantile and log-sparsity functions

$$\begin{aligned}
F_u(x) &= 1 - \exp[-(x/\tau)^\alpha] \\
Q_u(u) &= \tau[-\ln(1-u)]^{1/\alpha} - 1 \\
W_u(u) &= \log(\tau/\alpha) - \log(1-u) + (1/\alpha - 1)\log[-\log(1-u)] \\
Q_s(s) &= \tau(C_2 s)^{1/\alpha} \\
W_s(s) &= \log(\tau/\alpha) + (1/\alpha - 1)\log(C_2 s) + \log C_2
\end{aligned} \tag{4}$$

and therefore is a linear combination of the basis functions $\phi_1(u) = 1, \phi_2(u) = -\log(1-u)$ and $\phi_3(u) = -\log[-\log(1-u)]$. Note that the second basis function disappears from the s-representation. The first and second basis functions define the exponential log-sparsity W^E, and the third contributes the possibility of sharper than exponential curvature in the quantile on the left when $\alpha < 1$, which accounts for its popularity in the modeling of long-tailed distributions such as survival times.

But the coefficient $\alpha - 1$ implies a negative pole at $u = 0$ for $\alpha < 1$, a zero derivative for $\alpha = 1$, and a positive pole for $\alpha > 1$. The Weibull density is therefore nondifferentiable at $x = 0$ with the singular exception of $\alpha = 0$ and, as a consequence, is more or less incapable of realistically modelling zero-valued observations except in the isolated exponential case. This nasty feature causes no end of computational problems for data values near zero. Ironically, although this distribution is widely used to model long-tailed survival times, Weibull (1951) applied the distribution only to short-tailed data on soil particle sizes, finding α values of the order of 1.5.

A simple fix preserves the Weibull's capacity for long tails but guarantees stable exponential behavior on the right. We replace the third basis function by $\phi_3(u) = -\log[1 - \log(1-u)]$. As $u \to 1$ the unit shift of $-\log(1-u)$ has a vanishing effect, but ϕ_3 is zero at $u = 0$, so that the right behavior of the quantile is essentially that of the exponential for all values of α. The quantile and distribution functions for what we might refer to as the *Exponential-Weibull* or *E-W* distribution are available in closed forms:

$$\begin{aligned}
F_u(x) &= 1 - \exp\{-(x/\tau + 1)^\alpha + 1\} \\
f_u(x) &= (\alpha/\tau)(x/\tau + 1)^{\alpha - 1} \exp[-(x/\tau + 1)^\alpha + 1] \\
Q_u(u) &= \tau\{[1 - \ln(1-u)]^{1/\alpha} - 1\} \\
W_u(u) &= \log(\tau/\alpha) - \log(1-u) + (1/\alpha - 1)\log[1 - \log(1-u)] \\
F_s(s) &= [(x/\tau + 1)^\alpha - 1]/C_2 \\
Q_s(s) &= \tau[(1 + C_2 s)^{1/\alpha} - 1] \\
W_s(s) &= \log(\tau/\alpha) + (1/\alpha - 1)\log(1 + C_2 s) + \log C_2 \qquad (5)
\end{aligned}$$

respectively.

Figure 4 displays the $\log Q^W$ for the Weibull, and its modification $\log Q^{EW}$ for a range of values of α. We see on the right that the quantile shapes are essentially the same, but that on the left the modified log-quantiles coalesce to a common point, whereas their Weibull counterparts diverge. What the Weibull says about the distribution of highly probably events is extremely sensitive to α. In survival terms, long-tailed survival times are associated

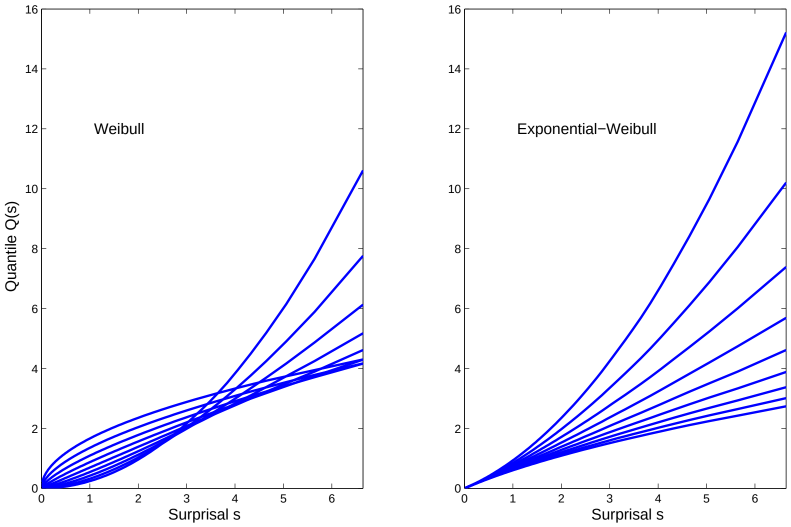

Figure 4: The left panel contains the quantile functions Q_s for the Weibull distribution plotted against surprisal, and the right panel the corresponding quantile functions for the Exponential-Weibull (EW) distribution for $\alpha = 2^{-1:0.25:1}$ and $\tau/\alpha = 1$.

with many very short survival times, but short-tailed survival times are associated with few short times. The Exponential-Weibull distribution disassociates the short survival time distribution from that of the long survival times, which seems more appropriate in many applications.

We see, too, that freeing up the coefficient of ϕ_2 in the E-W distribution implies a multiplication of Q as a function of surprisal by an exponential factor $\exp c_2$ that will ultimately dominate the power function as $s \to \infty$. That is, $c_2 > 0$ implies a positive curvature on the right as the positive exponential exponent takes over. Moreover, $\phi_3 = \log[1+\phi_2]$, and this transformation can be iterated as often as we please, so that, for example, $\phi_4(u) = \log[1+\phi_3(u)]$, $\phi_5(u) = \log[1+\phi_5(u)]$, and so on. Each recursion of the operator $L = \log(1+\cdot)$ adds the capacity for sharper and sharper curvature of the quantile function further and further to the right. Of course the Weibull has this capacity, too, but the recursion is $\log(+\cdot)$. That is, letting $L^0\phi = \phi$, we have the expansions

$$\begin{aligned} W(u|K) &= c_{-1} + \sum_{k=0}^{K} c_k L^k[-\log(1-u)] \\ W_s(s|K) &= c_{-1} + \sum_{k=0}^{K} c_k L^k(s) \end{aligned} \tag{6}$$

which we will refer to as the *E-W basis* system.

Regularization can be introduced by expressing a basis system as a reproducing kernel Hilbert space (Ramsay and Silverman, 2005). For example, for the Exponential-Weibull distribution expressed in terms of s, we can define the low-dimensional subspace $H_0 = \text{span}\{1, s, \log(1+s)\}$. The linear differential operator

$$\mathcal{L} = \left(\frac{2}{1+C_2 s}\right) \mathrm{D} + \mathrm{D}^2$$

applied to all three basis functions is 0. Consequently, we define the complement space H_1 as a function space containing functions ϕ for which $\mathcal{L}\phi \neq 0$ and with inner product

$$\langle \phi_i, \phi_j \rangle = \int_0^\infty \mathcal{L}\phi_i(s) \mathcal{L}\phi_j(s) ds$$

and which is, therefore, a reproducing kernel Hilbert space.

3 Differential equations for Q and F

Representations (1), (2) imply that $Q(u)$ and $F(x)$ satisfy the differential equations

$$\begin{aligned} \mathrm{D}Q_u(u) &= \exp[W_u(u)] \text{ and } \mathrm{D}F_u(x) = \mathrm{D}u = f_u[Q_u(u)] = \exp[-W_u(u)] \\ \mathrm{D}Q_s(s) &= \exp[W_s(s)] \text{ and } \mathrm{D}F_s(x) = \mathrm{D}s = f_s[Q_s(s)] = \exp[-W_s(s)] \end{aligned} \tag{7}$$

in terms of u and s respectively. Note that the right sides of these equations differ only in terms of the sign of $W_u(u)$ or $W_s(s)$, but as differential equations they are actually quite different. The DQ equations are first order linear differential equations $(\mathrm{D}Q)(u) = 0$ (having a constant as a solution) forced by the function $\exp[W(u)]$, and are therefore *non-autonomous* equations. But the (Du) and (Ds) equations involve u and s on the right hand side, respectively, and are therefore *autonomous* nonlinear first order equations. Steinbrecher and Shaw (2008) also showed that quantile functions can be represent by differential equations for many textbook distributions, and that the numerical approximation to solution of these equations can be a fast and accurate method for estimating quantile functions. The representations (1), (2) and (3) in terms of functions W_u and W_s turn this into a general result.

We often have a closed form expression for the density function $f(x)$, and we can use this to derive an alternative differential equation for Q when W is not analytically available. From (7), we have that

$$W_u(u) = -\ln f_u(x) = -\ln f_u[Q_u(u)] \text{ and } \mathrm{D}Q_u = 1/f_u(Q_u). \tag{8}$$

Taking the logarithm of both sides of $\mathrm{D}Q = 1/f(Q)$ often leads to an alternative basis function formulation. For example, the log-differential equations for the gamma and inverse gamma quantile functions are

$$\begin{aligned} \log \mathrm{D}Q^G_u &= \log[\beta^{-\alpha}\Gamma(\alpha)] + \beta Q^G_u + (1-\alpha)\log Q^G_u \text{ and} \\ \log \mathrm{D}Q^G_u &= \log[\beta^{-\alpha}\Gamma(\alpha)] + \beta Q^G_u + (1+\alpha)\log Q^G_u, \end{aligned} \tag{9}$$

respectively. The right sides are expanded in terms of the basis functions $1, Q^G_u$ and $\log Q^G_u$ and differ only in terms of the coefficient for $\log Q^G_u$. The larger third coefficient gives the inverse gamma function its long positive tail and makes it so appealing as a prior distribution for scale parameters in

Bayesian modeling. More generally, we can consider freeing some or all of the coefficients for these three basis functions, and exploring the possibility of using other bases as well.

4 Evaluating Q_u and Q_s

In most cases $u = F_u(x)$ cannot be analytically solved for $x = Q_u(u)$, and this implies that numerical methods will be required. As noted above, the definition of Q as the indefinite integral of $\exp(W)$ suggests that approximating the solution to the differential equation $\mathrm{D}Q = \exp(W)$ will be useful. The quantile functions in Figure 4 were produced by Matlab commands

```
DQhdl  = @(u, Q, beta) exp(-beta.*log(1-u));
[u, Q] = ode45(DQhdl, 0:0.01:0.99, 0, [], 1.0);
```

The first command defines an anonymous function with function handle `DQhdl`, and the second invokes the Runge-Kutta-Fehlberg solution approximation ode45. The value of β is its final argument. Similar commands in R can be constructed using the differential equation approximation function `ode` in the `deSolve` package.

```
library(deSolve)
DQhd1 <- function(u,Q,beta) {
  out <- exp(-beta*log(1-u))
  list(out) }
beta <- 1
u    <- 0:99/100
out  <- ode(0,u,DQhd1,beta,method="ode45")
```

The differential equation for Q is not stiff, and both sides will typically be differentiable to as high an order as is used for W. Our best experience to date has been with the explicit fourth order Runge-Kutta-Fehlberg Matlab function `ode45`, which we found to be substantially superior to the low-order function `ode23`. This and most such algorithms can be required to provide solutions at a vector of pre-assigned values of u. In our test problems involving moderately long tails, the computation time is dominated by the number of u-values required rather than the accuracy specified for the approximation.

The initial value of Q to be supplied to the function will depend on the distribution. For many distributions over the nonnegative real line, including the Exponential-Weibull, this will be $Q(0) = 0$. But where the slope of the quantile function at zero is steep or infinite, it will be necessary to evaluate the distribution function at a value of x that corresponds to a suitably small value of u or s, and supply x as the initial function value corresponding to the initial argument u or s. The Weibull and the gamma distribution for $\alpha > 1$ are examples.

For distributions over the whole real line, it would be necessary to run the solver forward and backward from some convenient point in the center of [0,1], although in the case of symmetric distributions it is more direct to use the reflection operation $-Q(1-u)$. For example, for the Gaussian and other symmetric distributions, $Q^N(0.5) = 0$ so that $u = 0.5$ would be the right initial argument for solving

$$\mathrm{D}Q^N = \sqrt{2\pi}\exp[(Q^N)^2/2]$$

with an initial function value $Q^N(0.5) = 0$. Using the default precision setting in Matlab's function `ode45`, for example, we get $Q^N(0.975) = 1.9599$, as opposed to 1.9600 returned by function `norminv`. Steinbrecher and Shaw (2008) use this differential equation, as well as its derivative, to approximate quantile functions and asymptotic tail behavior for the normal, t, beta and gamma distributions.

Representations (1) and (2) do have a property that can cause trouble in estimating Q_u for distributions with extremely long tails. Function $\exp(W_u)$ can have a large slope as $u \rightarrow 1$ and this can lead to smaller and smaller step sizes and denser and denser quadrature points for differential equation solution approximations and quadrature methods; resulting in an unacceptable loss of accuracy. For example, $\mathrm{D}\exp[W(0.99)] = 1.3 \times 10^5$ for the modified Weibull Q_u^{EW} with parameters $\tau = 1$ and $\alpha = 0.5$. However, at the same time the inverse quadrature problem associated with integrating over the ordinate to estimate F_u becomes easier and easier. A useful approach, therefore, is to estimate Q_u directly up to some point such as $u = 0.75$, and then approximate the solution to the differential equation for F_u starting at $\hat{Q}(0.75)$ with initial value 0.75 up to a value of x that is associated with a value of u that is considered sufficient.

However, no such problem exists for Q_s, since the curvature of this relationship is always mild. We recommend estimating Q_s, and then transforming to Q_u.

5 Estimating Q_u and Q_s from data $X_1, \ldots, X_N$

We turn to the practical matter of estimating a smooth quantile function from an i.i.d. sample of size N. We take for granted at this point that estimating Q_s as a function of surprisal values is preferable to estimating Q_u, which is in any case easily available from a fine mesh of discrete Q_s values by interpolation. However, the estimation strategies that we propose can also be readily adapted to the estimation of Q_u.

Estimating the quantile function for parametric distribution families is straightforward; maximum likelihood, Bayesian, or other estimation schemes provide parameter estimates, and then the quantile function is either computed directly or approximated as a function of these estimates. For example, maximum likelihood estimation of $\log \alpha$ and $\log \tau$ for the Exponential–Weibull distribution poses no difficulties in our experience, and avoids non-positive parameter estimates. Similarly, we may do the same for distributions defined by the first three basis functions for the E-W basis (6).

Defining the empirical quantile function $\tilde{Q}_s$ is not as straightforward as it is for its counterpart $\tilde{F}_s$. Technically, $\tilde{Q}_s(s) \equiv \inf\{x : \tilde{F}_s(x) \geq s\}$, and is a set-valued function that needs further manipulation to be useful. Parzen (1979) discusses various possibilities, each having their merits for specific objectives. The problem is further complicated when data are heavily discretized, such as in millimeters for rainfall, or even binned, so that large numbers of repeat values are possible. It seems reasonable to define $S_n, n = 1, \ldots, M \leq N$ as $S(U_n)$ where U_n is the center of bin n on the probability scale. Bin width is proportional to the frequency p_n of the bin value $X_{(n)}$, and bin centers are $S_n = S[N^{-1}(\sum_{m=1}^{n} p_m - p_n/2)]$, or $S[(i - 0.5)/N]$ when data values are unique. Either a step function with steps having heights points $X_{(n)}$ over $(S_{n-1}, S_n]$, or a polygon interpolating points $(S_n, X_{(n)})$, can be a useful empirical empirical quantile estimate of Q_s.

Nonparametric estimation of Q_s is considered here in terms of two strategies having complementary advantages. The first involves direct nonparametric estimation based on points $(S_{(i)}, X_{(i)})$ where $S_{(i)} = S(U_{(i)})$ with the $U_{(i)}$'s being uniform order statistics using the approaches in Section 5.1 and 5.2.

The second strategy is maximum likelihood estimation, which inevitably requires estimation of the distribution of the surprisal values S_i solving the implicit equation $X_i = Q_s(S_i)$. That is, direct estimation conditions on the quantiles of the s-distribution, and MLE conditions on the quantiles of the

x-distribution.

In addition to the two estimation strategies, we have two functional parameters that can be employed, either separately, or in combination. The first of these is the shifted log sparsity function W_u or W_s in representations (1) and (2). The second is the strictly monotone transformation or *warping function* $h(s)$ in

$$Q_s(s) = Q^*[h(s)] \tag{10}$$

where Q^* is a quantile function either defined by a parametric family or by representation (1) or (2). We use the h-representation

$$h(s) = \frac{M(s)}{M(S_{max})} \quad \text{where} \quad M(s) = \int_0^s \exp[V(v)]dv, \tag{11}$$

where V has the basis function expansion $V(s) = \sum_\ell d_\ell \phi_\ell(s)$. Since the ratio in (11) is invariant with respect to translations of V and most basis function expansions include the constant term either implicitly or explicitly it is necessary to impose the linear restriction $\sum_\ell d_\ell = 0$ for identifiability. Although strictly speaking there is no upper limit to surprisal, in practice one needs to be set in order to evaluate (11). We set this upper limit be be either the largest surprisal value or some value slightly above it which works well.

5.1 Least squares approximation of empirical quantiles

An easy estimation strategy is to fix the $S(i)$'s as above, or at bin centers S_n for tabulated data, and then fit the corresponding $X_{(i)}$ values by regularized weighted nonlinear least squares. In this approach, one "bets" that the data represent the population, in which case the s-values will have distribution $F_S(s) = 1-2^{-s}$, $s > 0$. This approach therefore conditions on the assumption that the model represents population variation in the same way that MLE conditions on the assumption that the density of the data are described by the $\hat{f}(X_i)$ values that MLE estimates.

The estimation of a strictly monotone function $Q_s(s)$ that fits the empirical quantile values is discussed at some length in Ramsay and Silverman (2005) [Is this true? I just quickly glanced over the book and didn't notice much discussion of estimating quantiles.]. A B-spline basis seems natural for W_s since s varies over a closed interval $[0, S_{max}]$, and we have had good experience with equally-spaced knots since the surprisal transformation tends to straighten the quantile function. A higher order E-W basis (6) may also be used, but it does not have the capacity to capture local features such as the storm surge in Q_s that we noted for Regina's rainfall.

We can regularize $W_s(s|\mathbf{c}) = \sum_{k=1}^{K} c_k \psi_k(s)$ where $\mathbf{c} = [c_1, \ldots, c_K]^T$ by trading off the fit to the data and fidelity to the solution of a differential equation $\mathrm{D}Q = \exp(W^*)$ under an assumed distribution parameterized by $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_p]^T$. As an example consider the *E-W* distribution (5) for which $\boldsymbol{\theta} = [\alpha, \tau]^T$ and

$$\mathrm{d}Q(s) = \exp\left\{\ln\left(\frac{\tau}{\alpha}\right) + \ln(C_2) + \left(\frac{1}{\alpha} - 1\right)\ln[1 + sC_2]\right\} = \exp\left\{W^*(s|\boldsymbol{\theta})\right\}.$$

The regularized estimate for $\mathbf{c}$ conditional on $\boldsymbol{\theta}$ and λ are obtained by solving

$$\mathbf{J}(\mathbf{c}, \boldsymbol{\theta}, \lambda) = \mathbf{D}^T \mathbf{V}^{-1}(\mathbf{x} - \mathbf{Q}) - \lambda \mathbf{z} = \mathbf{0} \tag{12}$$

where $\mathbf{x} = [x_{(1)}, \ldots, x_{(N)}]^T$, $\mathbf{Q} = [Q_s(s_1|\mathbf{c}), \ldots, Q_s(s_N|\mathbf{c})]^T$, $E[S_{(i)}] = s_i = \frac{1}{C_2}\sum_{r=1}^{i} 1/(n - r + 1)$, $\mathbf{D}_{N\times K} = \{d_{ij}\} = \{\partial Q_s(s_i|\mathbf{c})/\partial c_j\}$,

$$\mathbf{z}_{K\times 1} = \{z_i\} = \left\{\int_0^{S_{max}} \psi_i(s)[W_s(s|\mathbf{c}) - W^*(s|\boldsymbol{\theta})]\mathrm{d}s\right\},$$

and $\mathbf{V} = \{v_{ij}\}_{N\times N}$ a first order approximation of the covariance of the $X_{(i)}$'s; the v_{ij}'s are given by

$$v_{ij} = \frac{Q_s'(s_i)Q_s'(s_j)}{C_2^2}\sum_{r=1}^{i}\frac{1}{(n - r + 1)^2},\ i \leq j.$$

Here λ controls this trade-off since as $\lambda \to \infty$ we have that $W_s \to W^*$ which implies that $Q_s \to Q$ the solution to $dQ = \exp(W^*)$.

Numerical quadrature can approximate the integrals in (12), and we have used a composite m-point Gauss-Legendre quadrature rule

$$Q_s(s_i|\mathbf{c}) = \int_0^{s_i} q_s(v; \mathbf{c})\mathrm{d}v \approx \sum_{j=1}^{i}\frac{s_j - s_{j-1}}{2}\sum_{k=1}^{m} b_k\, q_s\left(\frac{s_j - s_{j-1}}{2}a_k + \frac{s_j + s_{j-1}}{2}\,\middle|\,\mathbf{c}\right) \tag{13}$$

with weights b_k and abscissas a_k where $q_s = \exp\{W_s\}$ and using the convention that $s_0 = 0$. For each sub-interval $(s_{j-1}, s_j]$ the composite rule will be accurate provided that q can be well approximated by a $2m - 1$ degree polynomial within the interval since (13) is exact for polynomials of degree $2m-1$.

Denote the solution of (12) as $\tilde{\mathbf{c}}(\boldsymbol{\theta},\lambda)$. For given $\tilde{\mathbf{c}}(\boldsymbol{\theta},\lambda)$ an update of $\boldsymbol{\theta}$ conditional on λ denoted $\tilde{\boldsymbol{\theta}}(\lambda)$ is obtained by solving

$$\mathbf{F}^T\mathbf{V}^{-1}(\mathbf{x}-\tilde{\mathbf{Q}})=\mathbf{0} \tag{14}$$

where $\tilde{\mathbf{Q}}$ is $\mathbf{Q}$ evaluated at $\tilde{\mathbf{c}}(\boldsymbol{\theta},\lambda)$, $\mathbf{F}^T_{s\times n}=[\mathbf{f}_1\ \mathbf{f}_2\cdots\mathbf{f}_n]$ and

$$\mathbf{f}_i=-\frac{\partial\mathbf{J}^T}{\partial\boldsymbol{\theta}}\left[\frac{\partial\mathbf{J}}{\partial\tilde{\mathbf{c}}^T}\right]^{-1}\frac{\partial\tilde{Q}(u_i)}{\partial\tilde{\mathbf{c}}}$$

by the application of the Implicit Function Theorem. Estimates $\hat{\mathbf{c}}(\lambda)$ and $\hat{\boldsymbol{\theta}}(\lambda)$ are calculated by iteratively solving (12) and then (14) until convergence. Finally, a value of the complexity parameter λ which balances the trade-off between fit and fidelity to the differential equation needs to be determined. We use GCV which finds the value of λ that minimizes

$$GCV(\lambda)=\frac{n}{(n-\mathrm{edf}(\lambda))^2}(\mathbf{x}-\hat{\mathbf{Q}})^T\hat{\mathbf{V}}^{-1}(\mathbf{x}-\hat{\mathbf{Q}}) \tag{15}$$

where

$$\mathrm{edf}(\lambda)=\mathrm{tr}\left[\hat{\mathbf{D}}\left(\hat{\mathbf{D}}^T\hat{\mathbf{V}}^{-1}\hat{\mathbf{D}}+\lambda\mathbf{G}\right)^{-1}\hat{\mathbf{D}}^T\hat{\mathbf{V}}^{-1}\right] \tag{16}$$

with $\mathbf{G}_{K\times K}=\left\{\int_0^{s_n}\phi_i(u)\phi_j(u)\mathrm{d}u\right\}$ all evaluated at $\hat{\mathbf{c}}(\lambda)$.

The roughness penalty term used in (12)

$$\int_0^{S_{max}}[W_s(s|\mathbf{c})-W^*(s|\boldsymbol{\theta})]^2\mathrm{d}s$$

can be replaced by

$$\lambda\int_0^{S_{max}}\left[(LW)(s|\mathbf{c})\right]^2\mathrm{d}s,$$

where L is a differential operator (Ramsay and Silverman, 2005). For example, $L=2/(1+C_2s)D+D^2$ annihilates the first two E-W basis functions, and therefore penalizes departure from the E-W quantile function. [I think we need to say something here about how these two approaches differ... it seems like they are doing the same thing (in the E-W case we consider) except that the L-spline approach obviates the need for the inner and outer optimization].

5.2 L_1 approximation of empirical quantiles

[I haven't edited this section as this approach isn't correct in the dependent case. I guess there are two options: 1) remove it and just make mention of the fact that we have considered the L_1 norm case or 2) work this out again properly. I've started thinking about this but I'd need to work on this some more to get it working]

The use of unweighted error sum of squares is motivated by the assumption of i.i.d. residuals with constant variance, an assumption that surely does not apply to this situation. Section 9.3 in Gilchrist (2000) has a good discussion of the merits of alternative approaches, and further remarks on the limitations of SSE in this context. He tends to prefer the L_1 norm because of its robustness to outliers. Following this logic, it may be preferable to find $\hat{c}(\boldsymbol{\theta}, \lambda)$ by minimizing

$$\sum_{i=1}^{N} |X_{(i)} - Q(S_i)| + \lambda \int_0^1 [W(s|\mathbf{c}) - W^*(s|\boldsymbol{\theta})]^2 \, \mathrm{d}u. \tag{17}$$

In order to minimize (17) an EM algorithm proposed by Phillips (2002) was utilized that modified the L_2-norm optimization strategy by minimizing a weighted version of (12),

$$J(\mathbf{c}, \boldsymbol{\theta}, \lambda) = \sum_{i=1}^{N} v_i (X_{(i)} - Q(S_i|\mathbf{c}))^2 + \lambda \int_0^1 [W(s|\mathbf{c}) - W^*(s|\boldsymbol{\theta})]^2 \, \mathrm{d}u \tag{18}$$

where $v_i = |X_{(i)} - Q(S_i|\mathbf{c})|^{-1}$. In this case $\hat{\mathbf{c}}(\boldsymbol{\theta}, \lambda)$ is obtained by successively solving

$$\mathbf{D}^T \mathbf{V}(\mathbf{x} - \mathbf{Q}) - \lambda \mathbf{z} = \mathbf{0}, \tag{19}$$

the M-step, with diagonal matrix $\mathbf{V}$ containing the v_i's being held fixed at the previous value of $\hat{\mathbf{c}}$, followed by updating the v_i's in the E-step, and repeat until convergence. In practice one must bound the weights v_i since when $|X_{(i)} - Q(s|\mathbf{c})| < \epsilon$ one will encounter numerical difficulties. We used

$$v_i = \begin{cases} 1 & , |X_{(i)} - Q(s|\mathbf{c})| < \epsilon \\ \frac{\epsilon}{\left|X_{(i)} - Q(s|\mathbf{c})\right|} & , \text{otherwise} \end{cases} .$$

To sharpen the approximation one can finally solve (18) once more with $v_i = 1$ for the observations that correspond to the d smallest residuals and 0 for the others where d represents the degrees of freedom. Since we are using a regularized approach and do not have integer valued degrees of freedom we use the effective degrees of freedom

$$\operatorname{edf}(\lambda) = \operatorname{tr}\left[\hat{\mathbf{D}}\left(\hat{\mathbf{D}}^T\hat{\mathbf{V}}\hat{\mathbf{D}} + \lambda\mathbf{G}\right)^{-1}\hat{\mathbf{D}}^T\hat{\mathbf{V}}\right] \tag{20}$$

with rounding to the next largest integer value as a proxy. The value of the smoothing parameter λ is obtained by minimized a weighted version of (15)

$$\frac{n}{(n - \operatorname{edf}(\lambda))^2}(\mathbf{x} - \hat{\mathbf{Q}})^T\hat{\mathbf{V}}(\mathbf{x} - \hat{\mathbf{Q}}). \tag{21}$$

5.3 Maximum likelihood estimation and registration

The negative log likelihood as a function of parameter vector $\mathbf{c}$ is given by

$$-\ln L_u = -\sum_i^N \log(f_u \circ Q_u)(U_i|\mathbf{c}) = \sum_i^N W_u[S_i(X_i|\mathbf{c})]. \tag{22}$$

Although this formulation is elegant as a function of W, it implies that the relation $X_i = Q(U_i)$ must be solved for U_i, and this in turn, from (7), requires approximating the solution to the differential equation $\mathrm{D}u = \exp[-W(u)|\mathbf{c}]$. The corresponding expression involving Q_s and W_s, using $\log f_u(x) = \log f_s(x) + \log(du/ds)$ and neglecting constant terms, is

$$-\ln L_s = \sum_i^N \{W_s[S_i(X_i|\mathbf{c})] + S_i(X_i|\mathbf{c})\} \tag{23}$$

or, in terms of tabulated data,

$$-\ln L_s = \sum_n^M p_n\{W_s[S_n(X_n|\mathbf{c})] + S_n(X_n|\mathbf{c})\}.$$

The *observed quantiles* $S_i(X_i|\mathbf{c})$ and $S_n(X_n|\mathbf{c})$ corresponding to observation X_i and bin center X_n, respectively, are implicit functions of $\mathbf{c}$, and we must use the Implicit Function Theorem to compute their partial derivatives with respect to parameters. From

$$\frac{dQ_s}{d\mathbf{c}} = \frac{\partial Q_s}{\partial S}\frac{dS}{d\mathbf{c}} + \frac{\partial Q_s}{\partial \mathbf{c}} = \frac{dX}{d\mathbf{c}} = 0$$

we obtain

$$\frac{dS}{d\mathbf{c}} = -\left(\frac{\partial Q_s}{\partial S}\right)^{-1}\left(\frac{\partial Q_s}{\partial \mathbf{c}}\right).$$

and, similarly,

$$\frac{d^2S}{d\mathbf{c}^2} = \left(\frac{\partial Q_s}{\partial \mathbf{c}}\left[\frac{\partial^2 Q_s}{\partial \mathbf{c}\partial S} + \frac{\partial^2\ Q_s}{\partial S^2}\frac{dS}{d\mathbf{c}}\right] - \frac{\partial Q_s}{\partial S}\left[\frac{\partial^2 Q_s}{\partial \mathbf{c}\partial S}\frac{dS}{d\mathbf{c}} + \frac{\partial^2\ Q_s}{\partial\ \mathbf{c}^2}\right]\right) / \left[\frac{\partial Q_s}{\partial S}\right]^2.$$

However, we have found direct maximum likelihood estimation in this way to be unstable, to require many iterations to converge, and to often produce estimates of Q_s that are substantially worse than least squares and L_1-based estimates. The estimated values of $S_i(\mathbf{c})$ by necessity have a monotonic relationship to the fixed surprisal values used in direct L_1 and L_2 estimation, and the transformation of these fixed surprisals is itself a functional parameter that, in effect, captures phase discrepancy between the empirical and estimated quantile functions.

This suggests replacing the unrestricted MLE by the estimation of a transformation $h(s)$ of $[0, S_{max}]$ whose smoothness can be controlled, and which is dependent on much less than N parameters. This can be combined with the use of a simple parametric quantile function $Q^*(s|\boldsymbol{\theta})$ or a lower-dimensional basis expansion of $W_s(s|\mathbf{c})$ that fits the data better when evaluated as a function of h, that is as

$$Q_s(s) = Q^*[h(s)] \quad \text{where} \quad h(0) = 0 \text{ and } h(S_{smax}) = S_{max}.$$

In effect, this defines quantile estimation as a *registration problem* (Ramsay and Silverman, 2005) where h is a *warping function* that rescales surprisal in order to better fit the observed data.

Let the warping function h be defined as $h(s) = M(s)/M(S_{max})$ where

$$M(s|\mathbf{c}) = \int_0^s \exp[V(v|\mathbf{c})]\,\mathrm{d}v,$$

$W(s|\mathbf{c}) = \sum_{k=1}^K c_k\phi_k(s) = \mathbf{c}'\boldsymbol{\phi}(s)$, and the K known functions ϕ_k define a basis for representing V. If Q^* offers a reasonable account of the data using a few parameters, than as a rule h will not depart greatly from linearity, so that V will be close to zero everywhere. In this case, using a small number of basis functions works well. Since h is invariant with respect to translations of the coefficient vector $\mathbf{c}$, let $\mathbf{c} = \mathbf{Hd}$ where the K by $K-1$ matrix $\mathbf{H}$ has full column rank, columns summing to zero and satisfies $\mathbf{H}'\mathbf{H} = \mathbf{I}$. Such matrices can be constructed in many ways, including evaluating $K-1$ Fourier series basis functions with period 2π at $2\pi(i-1)/K, i = 1, \ldots, K$.

Assume that $Q^*(s)$ is known, as well as the values $S_i^* = F_s^*(X_i)$, along with its log-sparsity function $W^*(s)$, perhaps after a preliminary fit by least squares or some other convenient method to estimate its parameters. The negative log likelihood is

$$F(\mathbf{d}) = -N \log[M(S_{max}|\mathbf{d})] + \sum_i^N [W(S_i|\mathbf{d}) + C_2 S_i] - \sum_i^N W^*(S_i^*), \qquad (24)$$

where $S_i = h^{-1}(S_i^*)$. The final term does not depend on $\mathbf{d}$ and may therefore be dropped if there are no additional parameters involved in its specification. We may optionally add a roughness penalty $\lambda\mathbf{R}$ to control the roughness of W when using a large K. The values $S_i = M^{-1}[M(S_{max})S_i^*]$ are computed by approximating the solution to the differential equation $\mathrm{D}s = \exp[-W(s|\mathbf{c})]$ for values $M(S_{max}|\mathbf{c})S_i^*$. The gradient and Hessian matrix with respect to $\mathbf{d}$ required to maximize (24) by Newton's method are obtained by approximating the solutions of the differential equations

$$\frac{\partial^2 Q_s}{\partial s \partial \mathbf{c}} = \boldsymbol{\phi} \exp[W_s(s)] \text{ and } \frac{\partial^3 Q_s}{\partial s \partial \mathbf{c}^2} = \boldsymbol{\phi}\boldsymbol{\phi}' \exp[W_s(s)],$$

respectively.

6 Simulation results

[I haven't edited this section since we are planning to beef up the sim study. I can certainly help out with this but we need to formalize the study design.]

We use simulation results to give an idea of the relative performance of the three types of estimates of the quantile function: the two-parameter Exponential-Weibull estimate, the least squares estimate using eight B-spline basis functions over [0,1], and the registered Exponential-Weibull maximum likelihood estimate.

We chose three population models, each defined by the three parameter basis (6). The first is a short-tailed distribution defined by coefficients 1.0, 0.8 and 0.0, respectively; the second a mildly long tail defined by coefficients 1.0, 1.0 and 0.2, and a the third a severely long-tailed distribution defined by coefficients 1.0, 1.2 and 0.5. The quantiles as a function of surprisal are display in the top three panels of Figure 6. For each design, we simulated 100 random samples, each of size $N = 100$.

We then calculated the bias and root-mean-squared error (RMSE) of the quantile function estimate by averaging across the 100 samples. The bias, viewed as a fraction of the RMSE, was negligible, even for the largest observation, and we do not consider it further. The ratio of the RMSE to the true quantile value is shown for each design in the corresponding bottom panel. For the short tailed design, all three methods have about the same RMSE, and reach about 15% of Q on the right, but with relatively little variation except on the left, where the ratio is taken with respect to small Q-values. For the mildly long-tailed case, least squares estimation does a bit worse than the parametric E-W and MLE methods, reaching about 25% of Q on the right; and we see that MLE does slightly better than the simple E-W case everywhere. In the severely long-tailed case least squares is severely poor relative to the other two methods, and again MLE consistently outperforms the parametric E-W approach, reaching about 30% of Q on the right. The bad performance of least squares on the right is due to over-fitting of the largest few empirical quantiles.

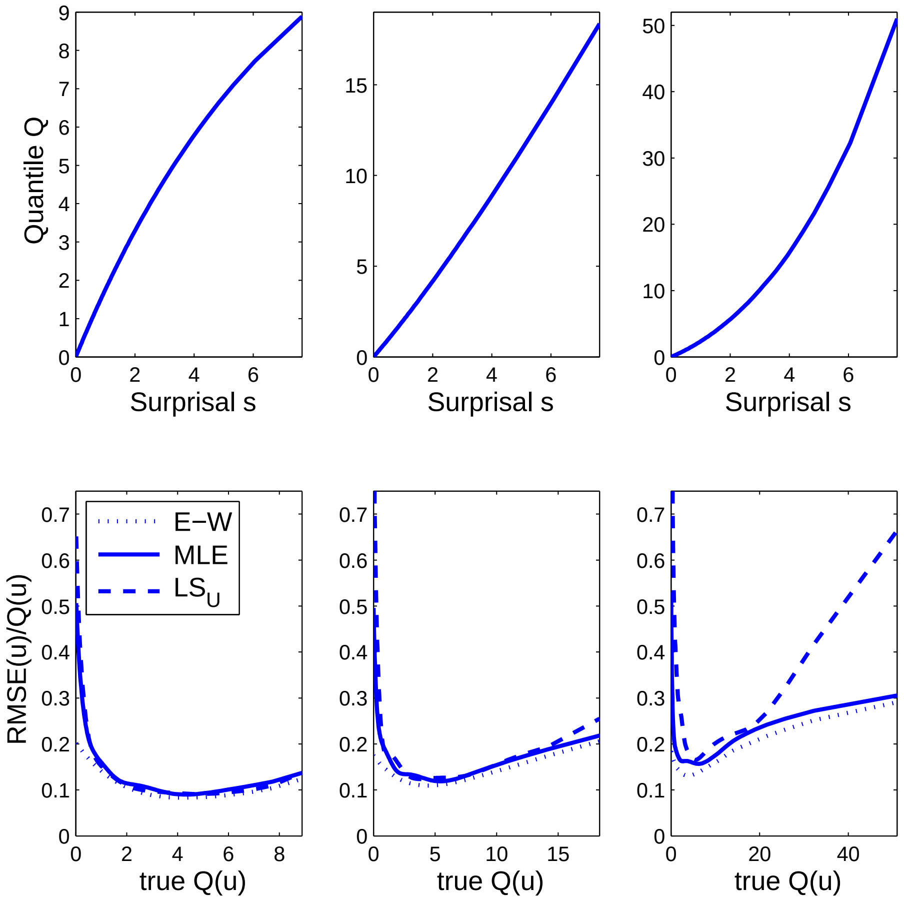

Figure 5: The top three panels display quantile functions for problems having a shorter than exponential tail, approximately exponential, and longer than exponential tail, respectively. The bottom panels show the corresponding root-mean-squared-error function divided by the quantile function averaged over 100 random samples, each of size 100. Results are shown are for the parametric Exponential-Quantile estimate (dotted line), least squares estimate (dashed line) and maximum likelihood registration of the Exponential-Quantile estimate (solid line).

7 Quantile functions for rainfall in Regina

Soil moisture is chronically in deficit over the Canadian prairies, and therefore precipitation is of intense interest to growers, crop insurers, and the scientists who provide the information. Precipitation is best viewed as a bivariate process consisting of the number of days between rain events where no rain is possible, and the amount of rain that falls on days where rain actually is observed or is declared possible by meteorologists. For example, the total rainfall for the critical month of June can be near normal, but consist of two large storms separated by enough dry days to destroy the crop. Hail is a serious crop hazard, and is associated with intense thunderstorms occurring along weather fronts.

The data were daily precipitation amounts over the period 1961 to 1994 recorded at the Regina airport, a growing region that is usually highly productive, but has historically suffered drought conditions about once a decade. Precipitation levels were only used for approximately 35% of the days for which there was measurable precipitation (at least 0.2 mm), and so the values analyzed only consist of the recorded levels larger than 0.2 mm. The data were grouped into months resulting in 355 measurable precipitation values for June, as an example, which when ordered correspond to the sequence $[1/2N, 3/2N, \ldots, 2N-1/2N]$ of u-values extending to a maximum of 0.9986, or 9.47 bits.

We fit the Exponential-Weibull baseline distribution (5) to each month's data by maximum likelihood estimation. The exponent α varied between 0.40 to 0.43 over the May-August growing period, indicating a rather long-tailed distribution; and the scale parameter τ had values 0.78, 1.15, 0.93 and 0.63, reflecting that fact the June has on average somewhat more rain than August. The quantile function Q_s^{EW} for June is seen in the right panel of Figure 2 as extending horizontally over the surprisal values assigned to the data, but substantially under-estimating the 132 mm maximum. Elsewhere, $Q^{'*}$ does a reasonable job, although it does not have the local fit capacity to capture the storm surge quantiles ranging from 4 to 6 bits. [What is $Q^{'*}$?]

The warping function h in the nonparametric maximum likelihood method was defined by 12 order 5 B-spline basis functions with equally-spaced knots. Figure 7 shows the probability warp for June, which shifts probability to the right over the whole unit interval, indicating that more of the data are concentrated in the left small-precipitation zone than the E-W distribution is able to accommodate. The net result is to shift the quantile function to

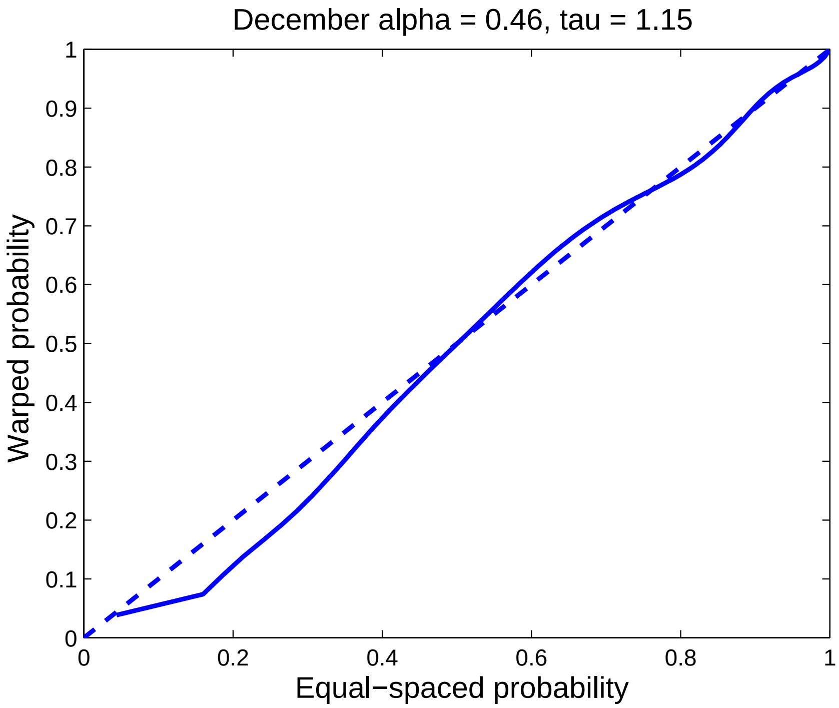

Figure 6: The solid line indicates the value of warping function $h(s)$ that transforms probability values to those that improve the fit of the Exponential-Weibull baseline quantile function values $Q^*[h(s)]$ to the empirical quantiles.

the right, and Figure 2 shows that the maximum likelihood estimate $Q_s(s)$ is now able to reach the level of the extreme rainfall, but does so by declaring a 12-bit or 0.0002 tail probability for this event, well beyond the range of values assigned to the data. At about 10 rainfall days on the average in June, this amounts to a rainfall of this magnitude occurring an average of once every 500 years. The Regina residents that we talked to thought that this was about right. [I hope they are all so lucky to have lived long enough to consider this correct ;-)]

The least squares fit to the empirical quantile function, seen in the right panel of Figure **??** as closely fitting the extreme value, was based on six Exponential-Weibull basis functions. Of course, this fit probably deserves the epithet "naive" since it assumes the fixed equi-spaced u-sequence and the corresponding s-sequence to be correct. Which quantile functions seems

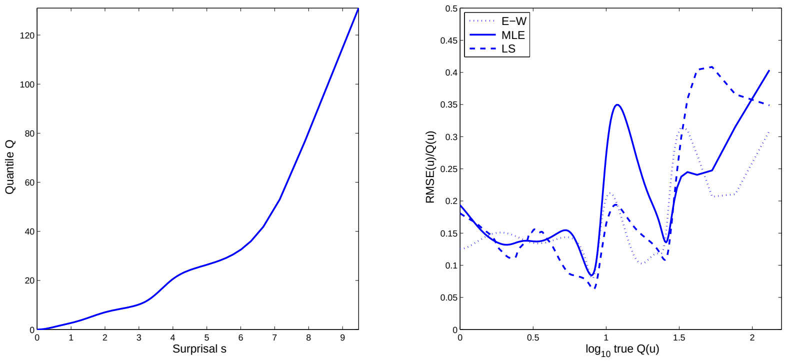

Figure 7: The left panel shows the quantile function as estimated by least squares for the Regina June rainfall data. The storm surge due to severe convective rainfall events is the bulge between 4 and 6 bits. The right panel plots root-mean-square-error over 100 simulated samples divided by the quantile value against the common logarithm of the quantile value. The three methods of estimation are the Exponential-Weibull quantile, maximum likelihood estimation using 12 basis functions, and least squares as a function of u using 8 spline basis functions.

better? An insurer would probably find the maximum likelihood estimate to be too optimistic for comfort, and probably also see the least squares solution as a bit pessimistic. The Exponential-Weibull appears to strike a reasonable compromise at the extremes, and also captures the storm surge quantiles.

In order to investigate the precision of estimation of Q using these methods, we used the simulation approach of Section 6, but this time using as the true quantile function the least squares estimate from the actual data, and the same sample size $N = 355$. This is effectively a form of parametric bootstrapping. The results over 1000 simulated samples are shown in Figures 7 and 7. The maximum likelihood estimate fails to track the storm surge between 15 to 30 mm and does not do well at the extremes. On the whole the preferred method here seems to be fitting the data with the Exponential-Weibull model.

8 Discussion

While this paper aims to develop methods for the parametric and nonparametric estimation of quantiles functions, perhaps the deeper theme is the power of differential equations to capture structural relationships among functions and their functional inverses. The two pairs of equations (7) are based on the reciprocal relation that holds between the derivative of a function and the derivative of its functional inverse, and this relation also connects probability and surprisal through $\mathrm{D}u = C_2 \exp(-C_2 s)$ and $\mathrm{D}s = C_2^{-1} \exp(C_2 s)$. The representations of Q and Q_s in terms of W and W_s, respectively, of (1), (2) and (3) are themselves based on the second order differential equation $\mathrm{D}^2 Q(u) = w(u)\mathrm{D}Q(u)$ and its companion for Q_s (Ramsay, 1996).

Figure 8 summarizes the functions that link the arguments u, s, x and $w = \log q$. It is striking that w and its corresponding functions $W(u)$ and $W_s(s)$ occupy such a central position in this diagram. The differential equations that we have used to both represent and compute F and W_s on the data side and Q_u and Q_s on the probability/surprisal side emphasize this pivotal role, and its capacity for allowing us to move with ease between probability and surprisal as substrates for data analysis.

A second underlying theme is the essential equivalence of probability u and surprisal s as substrates for a functional representation of variation in data. The two are not at all equivalent at the cognitive level, however. Surprisal is a magnitude, and adds over information arriving in independent packets, that is to say experientially, and therefore is much easier to manipulate by mental arithmetic than probability, whose multiplicative group has proved to be a curse for two and a half centuries of gamblers and students, and even on occasion for professional statisticians. In any case, displaying quantile functions over surprisal is a most helpful graphic device.

When we looked at familiar distributions from the perspective of W, we saw that many of them can be modified to give new distributions with helpful features, such as the Exponential-Weibull case, and also generalized in terms of obvious basis function sequences to add additional flexibility. To be sure, W is not always expressible analytically, but where the density function f is available, $\mathrm{D}F = f = \exp W$ and therefore $W = \log(f \circ Q)$ permits its numerical approximation, as well as its estimation on the basis of empirical quantile functions or smoothed versions of them. Moreover, we saw that in (9) that $\log \mathrm{D}Q$ often has an interesting expansion in terms of functions of Q

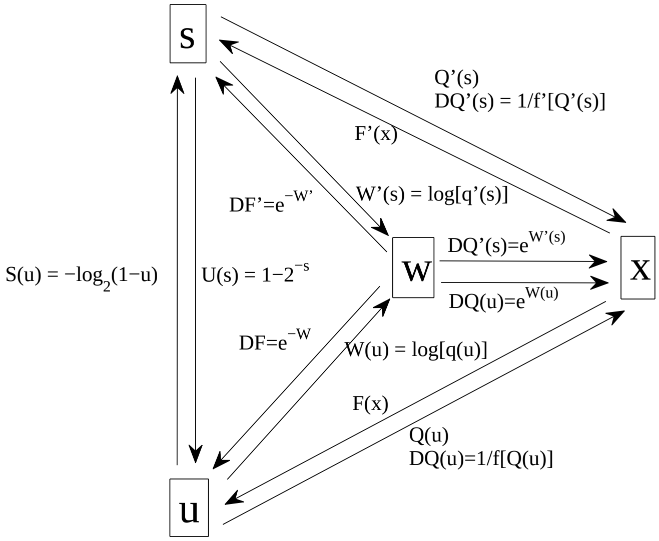

Figure 8: A graph of the relationships between arguments u, s and $w = \log q(u)$.

that may also inspire interesting modifications and generalizations of familiar distributions.

Finally, it seems clear to us why far-sighted pioneers in our field such as Tukey and Parzen pointed to Q as the function that would best convey to our clients concepts such as risk. The insights of Kullback (1959) into the use of information instead of probability as the fundamental continuum underlying statistical theory, along with the fact that the surprisal and quantile functions are essentially identical in the exponential case, seems to call for a fresh visit to the foundations of our field.

References

Cover, T. M. and Thomas, J. A. (2006) Elements of Information Theory, Second Edition. New York: Wiley.

Devroye, L. (1986) *Non-uniform Random Variate Generation.* New York: Springer.

Gilchrist, W. G. (2000) *Statistical Modelling with Quantile Functions.* London: Chapman & Hall/CRC.

Koenker, R. (2005) *Quantile Regression.* Cambridge: Cambridge University Press.

Kullback, S. (1959) *Information Theory and Statistics.* Wiley.

Parzen, E. (1979) Nonparametric statistical data modeling (with discussion). *Journal of the American Statistical Association,* **74,** 105–131.

Parzen, E. (1997) Concrete statistics, in Ghosh, S., Schucany, W.R. and Smith, W. B. (Eds.) *Statistics of Quality*, New York: Marcell Dekker.

Parzen, E. (2004) Quantile probability and statistical data modeling. *Statistical Science,* **19,** 652–662.

Phillips, R. F. (2002) Least absolute deviations estimation via the EM algorithm. *Statistics and Computing*, **12**, 281–285.

Ramsay, J. O. (1996) Estimating smooth monotone functions. *Journal of the Royal Statistical Society, Series B,* **60,** 365–375.

Sheather, S. J. and Marron, J. S. (1990) Kernel quantile estimators, *Journal of the American Statistical Association*, **85**, 410-416.

Steinbrecher, G. and Shaw, W. T. (2008) Quantile mechanics, *European Journal of Applied Mathematics*, **19**, 87-112.

Tribus, M. (1961) *Thermostatistics and Thermodynamics.* Princeton, NJ: D. van Nostrand Company, Inc.

Tukey, J. (1962) The future of data analysis. *Annals of Mathematical Statistics,* **33,** 1–67.

Tukey, J. (1977) *Exploratory Data Analysis.* Reading, Mass.: Addison-Wesley.

Weibull, W. (1951) A statistical distribution function of wide applicability, *Journal of Applied Mechanics - Transactions of the ASME,* **18,** 293-297.