The tenets of indirect inference in Bayesian models

Dmytro Perepolkin[1,*] Ullrika Sahlin[1] Benjamin Goodrich[2]

First version: 2021-06-29
This version: 2021-08-27

Abstract

This paper extends the application of Bayesian inference to probability distributions defined in terms of quantile function. We introduce the method of indirect likelihood to be used in the Bayesian models with sampling distributions defined by the quantile function. We provide examples and demonstrate the equivalence of this "quantile-based" (indirect) likelihood to the conventional "density-defined" (direct) likelihood. We consider practical aspects of the numerical inversion of quantile function by rootfinding required by the indirect likelihood method. In particular we consider a problem of ensuring the validity of an arbitrary quantile function with the help of Chebyshev polynomials and provide useful tips and implementation of these algorithms in Stan and R. We also extend the same method to propose the definition of an "indirect prior" and discuss the situations where it can be useful. |

[1] Centre for Environmental and Climate Science, Lund University

[2] Applied Statistics Center, Columbia University

[*] Correspondence: Dmytro Perepolkin <dmytro.perepolkin@cec.lu.se>

Contents

You have to start looking at the world in a new way.
– Priya in TENET by Christopher Nolan

1 Introduction

1.1 The quantile function tale

Up until mid-20th century the chart of the cumulative distribution function was always read from "x to y", from observations to probability. It has probably crossed some curious minds before that it could also be read backwards, but the first "documented" case is believed to be Tukey (1965) who "flipped" the CDF sideways or read it "from y to x", from probability to value, calling it the "representing function". Today we know this function as the "inverse CDF" or the "quantile function". Although the research topic related to the properties of quantile function, and the distributions based on it have seen some academic interest, the number of articles published in this area is simply incomparable to the amount of attention enjoyed by the "density-based" methods. Quantile functions have always been viewed as a little awkward to work with, especially given the innovations of Pearson and others who have given us the luxury of the exponential family of probability distributions, and later, Schlaifer and Raiffa (1961) who have introduced the world to the miracle of the conjugate models. Until today, the research related to the quantile distributions is often classified under "empirical, non-parametric and other non-traditional statistics".

Advances in computer science and optimization in the recent decades have made previously intractable problems solvable and previously impossible operations routine and doable. For example, even though the inverse CDF of the normal distribution is still unavailable in closed form, most (if not all) of the modern statistical software packages have it implemented at the core and scientists have learned to never think of it as a huge task to find z-values corresponding to some probabilities. What has been done in the past with the help of a table printed on the back of every statistics textbook, today is done at scale with a click of a button (or even a tick of a processor). This made previously non-invertible distributions, like normal and gamma, "numerically invertible" and the methods using them feasible, accessible and common.

1.2 Aims of the paper

One feature that the quantile distributions lacked until today, is the ability to learn from data. The absence of PDF has been seen as a hurdle to using them in classical Bayesian models. This article attempts to build on the ideas of Parzen (1980), Gilchrist (2000) and Nair et al. (2020) and systematically present and illustrate the Bayesian inference using quantile functions. We call this method of inference "indirect" because instead of dealing with *observations* or *parameters*, it deals with *probabilities* corresponding to these observations or parameters (given the parametrized model). Our aim is to show that the indirect Bayesian inference using intermediate probabilities ("CDF-values") leads to the same posterior beliefs, as the conventional, density-based inference. We highlight and discuss practical challenges related to inverting the non-invertible quantile functions and propose a novel application of the proxy root-finding algorithm for validation of quantile functions.

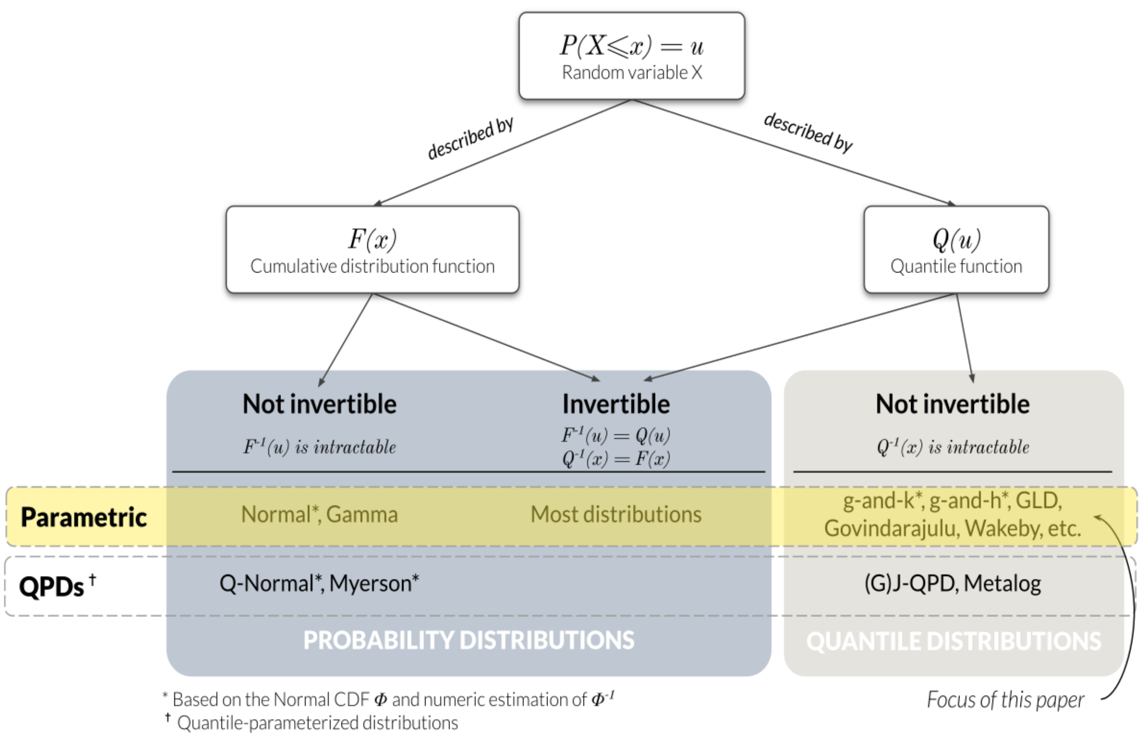

Figure 1: Probability distributions, quantile distributions and parameterization by quantiles.

1.3 Paper structure

Section 2 revisits the functions and identities for characterizing the distribution of a continuous random variable. Then in Section 3 we introduce the terms of "indirect likelihood" and "indirect prior" and show that likelihood (and prior) can be expressed without the probability density function (PDF). Section 4 presents an example of a Bayesian model with indirect likelihood expressed by a non-invertible quantile distribution implemented using Hamiltonian Monte Carlo in Stan (Stan Development Team, 2021).

Section 5 discusses the computational aspects of estimating the intermediate CDF values in indirect likelihood expressed by a quantile distribution. We discuss root-finding algorithms and provide an example model with indirect Tukey *g-and-h* likelihood. We used the Robust Adaptive Metropolis MCMC algorithm by Vihola (2012) interfaced by the `fmcmc` package (Vega Yon and Marjoram, 2019) and a built-in bracketing root-finding algorithm for inverting a quantile function in R.

Section 6 discusses a problem of ensuring that the particular combination of parameters produces a valid quantile function. We propose a proxy-rootfinding algorithm based on Chebyshev polynomials and discuss challenges and computational cost associated with finding the roots of a QDF. We illustrate the use of Chebyshev polynomials for proxy root-finding by approximating the quantile density function of the *g-and-k* distribution and provide custom functions in R implementing the two method of approximation proposed by Boyd (2013).

We conclude the paper by discussion and summary of the results in Section 7. We provide the definitions of the distributions used in this paper in Appendix A.

2 Distribution specification

In this section we briefly review the different ways of specifying a probability distribution to set the scene for the discussion of direct and indirect Bayesian inference. We review the use of the inverse cumulative distribution function (quantile function) to describe statistical distributions and discuss several examples of distributions defined by the quantile function ("quantile distributions") found in the scientific literature (Figure 1).

2.1 Essential functions

Let X be a continuous random variable. It can be expressed via the distribution function, also known as the *cumulative distribution function* (CDF):

$$F_X(x|\theta) = Pr(X \leq x|\theta), \quad \theta \in \mathcal{A} \subset \mathbb{R} \tag{1}$$

Alternative way of describing the random variable X is via the *quantile function* (QF).

$$Q_X(u|\theta) = \inf\{u : F_X(x|\theta) \geq u\}, \quad 0 \leq u \leq 1 \tag{2}$$

If $F_X(x)$ is continuous and non-decreasing over the support of X, then $Q_X(u|\theta)$ is simply an inverse of $F_X(x|\theta)$. Therefore, the quantile function is often referred to as the "inverse CDF", i.e.

$$Q_X(u|\theta) = F_X^{-1}(x|\theta) \tag{3}$$

Unfortunately, not all CDFs are analytically invertible. A distribution defined by a non-invertible quantile function $Q_X(u|\theta)$ is called a *quantile distribution* (Figure 1).

The derivative of the CDF is the *probability density function* (PDF) denoted by

$$f_X(x|\theta) = \frac{dF_X}{dx} \tag{4}$$

Similarly, the derivative of the QF is the *quantile density function* (QDF) denoted by

$$q_X(u|\theta) = \frac{dQ_X}{du}, \quad 0 \leq u \leq 1 \tag{5}$$

The reciprocal of the QDF $[q_X(u|\theta)]^{-1} = f(Q_X(u|\theta))$ is referred to as the *density quantile function* (Parzen, 1980) or *p-pdf* (Gilchrist, 2000).

$$f(Q(u)) = \frac{dF(Q(u))}{dQ(u)} = \frac{dF(Q(u))/du}{dQ(u)/du} = \frac{dF(F^{-1}(u))/du}{q(u)} = \frac{du/du}{q(u)} = [q(u)]^{-1} \tag{6}$$

2.2 Derivatives of the inverses and numerical approximation

Following the inverse function theorem (Price, 1984), for a function to be invertible in the neighborhood of a point it should have a continuous non-zero derivative at that point. If the function is invertible, the derivative of the inverse is reciprocal to the function's derivative. Formally, if dy/dx exists and $dy/dx \neq 0$, then dx/dy also exists and $dx/dy = [dy/dx]^{-1}$. Therefore, for a quantile function $Q(u) = x$, if a QDF $q(u)$ exists and $q(u) \neq 0$, then PDF $f(x)$ also exists and it is equal to $f(x) = [q(u)]^{-1}$ (Figure 2). In Section 3 of this paper, we rely on the density quantile function (DQF) $[q(u|\theta)]^{-1}$, a quantile form of the PDF, to define the likelihood in a Bayesian model based on the quantile distribution.

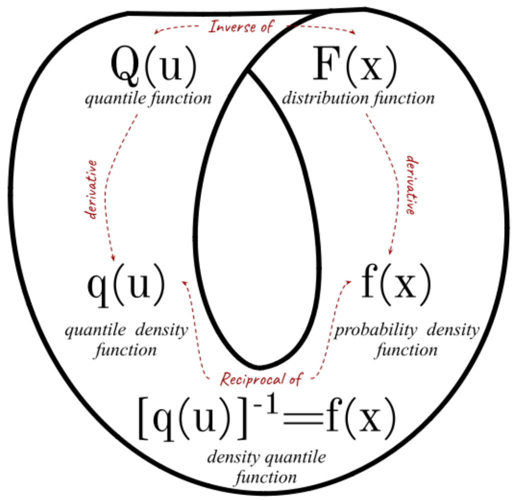

Figure 2: Moebius strip of probability functions.

Even though quantile distributions lack the closed-form CDF $F_X(x|\theta)$, in most cases, the CDF values can be approximated by numerically inverting the $Q_X(u|\theta)$. We denote the numerically inverted quantile function as $\widehat{Q}_X^{-1}(x|\theta)$ or $\widehat{F}_X(x|\theta)$. The inverse of a quantile function $Q(u|\theta)$ at point u, corresponding to the observation x, is obtained by minimizing the difference between the actual observation x and $Q_X(u|\theta)$ by iteratively refining the CDF value u. The details of the numerical inversion algorithms are discussed in Section 4.

2.3 Quantile distributions

Statistical methods utilizing QF and QDF were pioneered by the seminal work of Parzen (1979). Today the research area of quantile distributions is an active field of interest for many scientists. The most popular quantile distributions covered in the literature are Tukey's *g-and-h* and its sibling *g-and-k* distribution (Haynes and Mengersen, 2005; Jacob, 2017; Prangle, 2017; Rayner and MacGillivray, 2002), Tukey's Generalized Lambda Distribution, known as GLD (Aldeni et al., 2017; Chalabi et al., 2012; Dedduwakumara et al., 2021; Fournier et al., 2007; Freimer et al., 1988), Wakeby distribution (Rahman et al., 2015) and Govindarajulu distribution (Nair et al., 2012, 2013). Although in this paper we focus on the parametric quantile distributions, the quantile distributions parameterized by the quantile-probability pairs ("quantile-parametrized" quantile distributions) are also worth a mention. The most prominent examples are the Johnson QPD (J-QPD) and its generalization (Hadlock and

Bickel, 2017, 2019), as well as the metalog distribution (Keelin, 2016; Keelin and Powley, 2011). These distributions play an important role in representing expert beliefs about the variables, parameters or quantities of interest, although they don't lend themselves easily as sampling distributions due to the special nature of their parameterization.

3 Bayesian inference for quantile functions

In this section we adopt the Bayesian updating approach proposed by Nair et al. (2020) and apply it to the classical gamma-exponential model. We introduce the terms of *direct* and *indirect* likelihood using the identities and substitutions introduced in Section 2 and show the equivalence of the two ways of expressing likelihood in Bayesian models.

3.1 Direct and indirect likelihood

Traditional Bayesian inference formula can be equivalently restated using the substitutions involving quantile functions. Assume that the prior information about the scalar parameter θ can be summarized by the prior distribution over the parameter space, Θ. Then, given a random sample of $\underline{x} = \{x_1, x_2, \dots x_n\}$, the posterior distribution of θ can be expressed as:

$$f(\theta|\underline{x}) \propto \mathcal{L}(\theta; \underline{x}) f(\theta) \tag{7}$$

where $f(\theta|\underline{x})$ is the posterior distribution of θ after having observed the sample $\underline{x}$, $f(\theta)$ is the prior distribution of θ, and $\mathcal{L}(\theta; x) = \prod_{i=1}^{n} f(x_i|\theta)$ is the likelihood. We refer to this form of likelihood as "direct", because the observables x are directly used as input to the likelihood function.

Given the random sample of $\underline{x}$, we can use the QF to compute $\underline{Q} = \{Q_1(u_1), Q_2(u_2), \dots Q_n(u_n)|\theta\}$, such that $u_i = F(x_i|\theta), i = 1 \dots n$. The CDF values u_i, which we denote by $\underline{u}|\theta$, can be thought of as degenerate random variables, because they are fully determined given the observables $\underline{x}$ and parameter θ. Since $Q(u_i|\theta) = x_i$ we can substitute $\underline{Q}$ for $\underline{x}$. Then the Bayesian inference formula (7) becomes:

$$f(\theta|\underline{Q}) \propto \mathcal{L}(\theta; \underline{Q}) f(\theta) \tag{8}$$

We refer to the likelihood $\mathcal{L}(\theta; \underline{Q}) = \prod_{i=1}^{n} f(Q(u_i|\theta)) = \prod_{i=1}^{n} [q(u_i|\theta)]^{-1}$ as "indirect", because it relies on computing of intermediate CDF values $u_i = F(x_i|\theta), i = 1 \dots n$. As we have shown, the two forms of likelihood $\mathcal{L}(\theta; \underline{Q})$ and $\mathcal{L}(\theta; \underline{x})$ are proportional to each other.

Therefore, following the likelihood principle, the posterior beliefs about θ are independent of the form of the likelihood function used.

Since the likelihood in the Equation (8) is expressed in terms of $\underline{Q} = Q(\underline{u}|\theta)$, additional transformation is required to arrive at $\underline{u} = F(\underline{x}|\theta)$. In case the closed form of the CDF $F(\underline{x}|\theta)$ is not available, the numeric approximation of $\widehat{Q}^{-1}(\underline{x}|\theta)$ may be used. We discuss the details of the numerical approximation of the inverse quantile function in Section 5 of this paper.

To illustrate the equivalence of the two ways of specifying likelihood in Bayesian models, we use the example from Klugman et al. (2004)[1] regarding the distribution of the claim amounts (referred to hereinafter, as the Claims Example).

Klugman et al. (2004) model the claim amounts using exponential distribution with the mean of $1/\lambda$, where parameter λ is following the gamma distribution with the shape $\alpha = 4$ and the scale $\beta = 0.001$. Given three observations of the claim amounts $\underline{x} = 100$, 950 and 450, we sample from the posterior distribution of the parameter λ using the Hamiltonian Monte Carlo "No U-Turn" sampler implemented in Stan (Stan Development Team, 2021). Since gamma prior is conjugate to the exponential sampling distribution (Pratt et al., 1995) we can verify the distribution of the posterior draws using the analytic solution as $f(\lambda|\underline{x}) = \text{Gamma}(\lambda|\alpha + N, \beta + \sum \underline{x})$, where α and β are the parameters of gamma distribution and $\underline{x} = \{x_1, x_2, \dots x_N\}$ is the sample of observations of size N.

Table 1 and Figure 3 present the posterior distribution of θ from the two gamma-exponential models: one with a direct and another one with an indirect likelihood, along with the posterior from the conjugate model. Stan programs for the corresponding examples are provided in the Supplementary Material.

Table 1: Posterior sample comparison for the models

variable	mean	median	q5	q95	rhat
lambda (direct likelihood)	0.0028324	0.0026905	0.0013352	0.0047949	1.000157
lambda (indirect likelihood)	0.0028291	0.0026787	0.0013695	0.0047937	1.000506

The posterior distributions of parameter λ from the two models are equivalent and match the analytic solution from the conjugate model within a sampling error.

[1] Example 16.17 on p.544

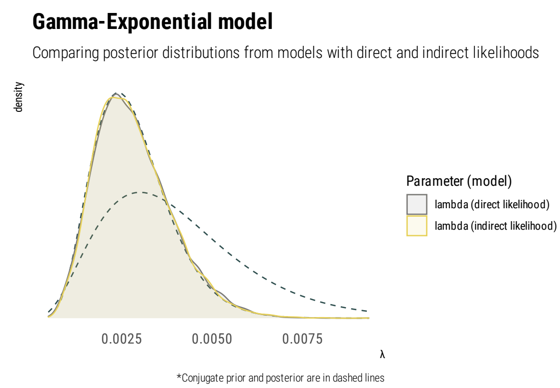

Figure 3: Summary of the posterior samples from the gamma-exponential model with direct and indirect likelihood

3.2 Direct and indirect prior

We can now extend the same logic of substitution using quantile functions to define the *direct* and *indirect* prior. In this section we discuss the transformation of parameters required for implementing the quantile prior and show its connection to the inverse transform used for non-uniform sampling.

Bayesian inference formula can also be restated using the quantile form of the prior. Assume that the prior distribution of θ can be described using the invertible distribution $F_\Theta(\theta) = v$, so that $Q_\Theta(v) = \theta$. Substituting the quantile values $Q_\Theta(v)$ for values of θ, prior beliefs about the parameter(s) of the sampling distribution can be expressed *indirectly* using the distribution of the quantile values corresponding to the CDF value v, given hyperparameter(s) of the prior distribution (9).

$$\begin{aligned} f(Q_\Theta(v)|\underline{x}) &\propto \mathcal{L}(Q_\Theta(v);\underline{x})f(Q_\Theta(v)) \\ [q_\Theta(v|\underline{x})]^{-1} &\propto \mathcal{L}(Q_\Theta(v);\underline{x})[q_\Theta(v)]^{-1} \end{aligned} \tag{9}$$

where $[q_\Theta(v|\underline{x})]^{-1}$ is the indirect form of the quantile posterior, $[q_\Theta(v)]^{-1}$ is the indirect form of quantile prior and $\mathcal{L}(Q_\Theta(v);\underline{x})$ is the direct likelihood, relying on the non-linear parameter transformation $\theta = Q_\Theta(v)$. The likelihood with such parameter transformation requires a Jacobian adjustment(Andrilli and Hecker, 2010), which is equal to the absolute derivative of the transform, i.e. $J(Q_\Theta(v)) = |dQ_\Theta(v)/dv| = |q_\Theta(v)|$. We refer to such formulation of the prior as the *indirect prior* because it describes the prior distribution of probability v corresponding to the parameter θ and not the distribution of the parameter θ directly. Therefore its density is expressed in the quantile form $f(Q_\Theta(v)) = [q_\Theta(v)]^{-1}$.

Of course indirect prior can also be used in combination with an indirect likelihood instead of the direct likelihood shown in (**??**), since, as we showed previously, the two of them lead to the same posterior beliefs about the parameter θ and, consequently, v. In such case, neither prior nor likelihood would require the existence of the closed-form PDF and, therefore, both of them can be represented by quantile distributions.

Provided that the $Q_\Theta(v)$ is a valid (non-decreasing) quantile function, meaning that $q_\Theta(v)$ is non-negative on $v \in [0,1]$, the quantile density term representing the prior and the Jacobian adjustment can be dropped as they are reciprocal to each other.

$$\begin{aligned}
[q_\Theta(v|\underline{x})]^{-1} &\propto \mathcal{L}(Q_\Theta(v);\underline{x})[q_\Theta(v)]^{-1}|q_\Theta(v)| \\
[q_\Theta(v|\underline{x})]^{-1} &\propto \mathcal{L}(Q_\Theta(v);\underline{x})
\end{aligned} \tag{10}$$

where $[q_\Theta(v|\underline{x})]^{-1}$ is the quantile form of the posterior, and the quantile prior $[q_\Theta(v)]^{-1}$ is implied. The quantile function transform $Q_\Theta(v) = \theta, v \in [0,1]$ (given the relevant hyper-parameters) hints the shape of the prior. This formulation represents the quantile function transformation of a variate with a standard uniform prior, as it relies on the QF transformation of the unit parameter v.

In the Claims Example the prior belief about the distribution of the exponential parameter λ was represented by the Gamma distribution, which is, unfortunately not easily invertible (Figure 1). In order to illustrate the indirect specification of the prior, we pick an invertible prior as representation of the expert's beliefs about the exponential parameter λ. The shape of the new prior $Rayleigh(0.003)$ is comparable to the gamma prior $Gamma(4, 0.001)$ used previously (Figure 4).

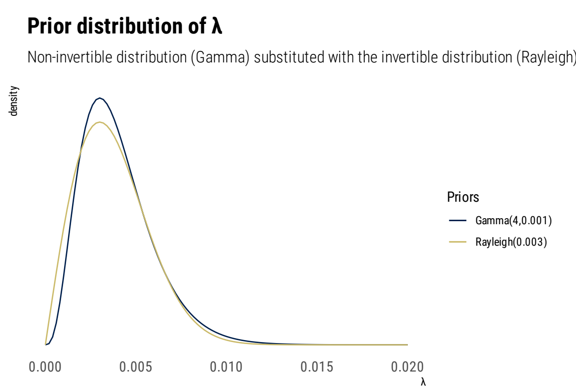

Figure 4: Prior distribution of the rate parameter of Exponentail distribution

Table 2 and Figure 5 compare the posterior distribution of θ from the gamma-exponential

Table 2: Summary of the posterior samples from the Rayleigh-Exponential model with direct and indirect prior and likelihood

variable	mean	median	q5	q95	rhat
lambda (direct prior, direct likelihood)	0.0026969	0.0025534	0.0011192	0.0047638	1.000932
lambda (direct prior, indirect likelihood)	0.0026911	0.0025490	0.0011319	0.0047732	1.001204
lambda (indirect prior, direct likelihood)	0.0027010	0.0025755	0.0011266	0.0047007	1.000218
lambda (indirect prior, indirect likelihood)	0.0027172	0.0025843	0.0011114	0.0047745	1.000886

model with direct and indirect likelihood. Stan programs for corresponding examples are provided in the Supplementary Material.

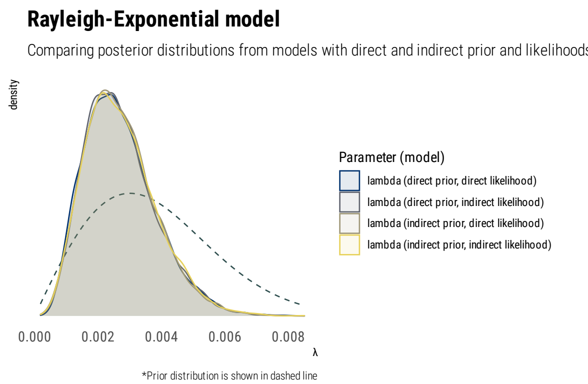

Figure 5: Distribution of the posterior samples from the Rayleigh-Exponential model with direct and indirect prior and likelihood

4 Indirect likelihood with a non-invertible sampling distribution

The models we have used in the previous sections allowed us to choose between *direct* or *indirect* likelihood, because the (exponential) sampling distribution was easily invertible. The main advantage of using the *indirect* form of the likelihood is that it allows us to adopt data generative models based on non-invertible (quantile) distributions. Below, we show an example of updating the parameters of a bathtub-shaped Govindarajulu distribution, which does not have a closed-form CDF or PDF.

We take the dataset provided in Aarset (1987) on time-to-failure of 50 devices (Figure 6). Lifetime reliability data are often modeled using specialized distributions (Nadarajah, 2009) or 2(3)-component mixtures. Nair et al. (2020) used the maximum likelihood to estimate the posterior median of the parameter γ in the Govindarajulu distribution (Nair et al., 2012) given the generalized exponential prior (Gupta and Kundu, 2007). We extend their approach to estimate the full posterior distribution of both parameters in the Govindarajulu likelihood, implementing it in Stan (Stan Development Team, 2021).

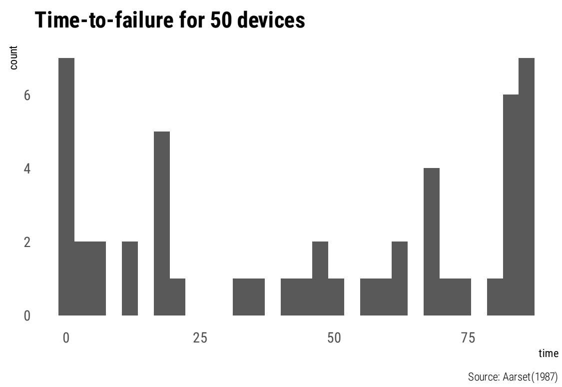

Figure 6: Histogram of the time to failure data

We adopt the generalized exponential prior for the parameter γ of Govindarajulu distribution with hyperparameters $\alpha = 5$ and $\lambda = 1$. The parameter σ of the Govindarajulu distribution can not be lower than the maximum of the observed times-to-failure. Therefore, we defined a shifted exponential prior with the rate $\lambda = 1$ and varying lower bound corresponding to the highest time to failure (in our example equal to 86). The Jacobian for adding a constant to the sampled parameter σ, although required, is equal to one, so there is no impact on the log density, as the shifting transform produces a Jacobian derivative matrix with a unit determinant.

Table 3: Summary of the posterior samples from the GenExp-Govindarajulu model using indirect likelihood

variable	mean	median	q5	q95	rhat
gamma	2.041895	2.012847	1.589812	2.588873	1.001060
sigma	86.021648	86.000386	86.000002	86.112457	1.000589

Table 3 and Figure 7 summarize the posterior distribution of the parameters σ and γ of Govindarajulu distribution. We used 2500 post-warmup iterations and 4 chains (see Figure 8 for NUTS parameter plots).

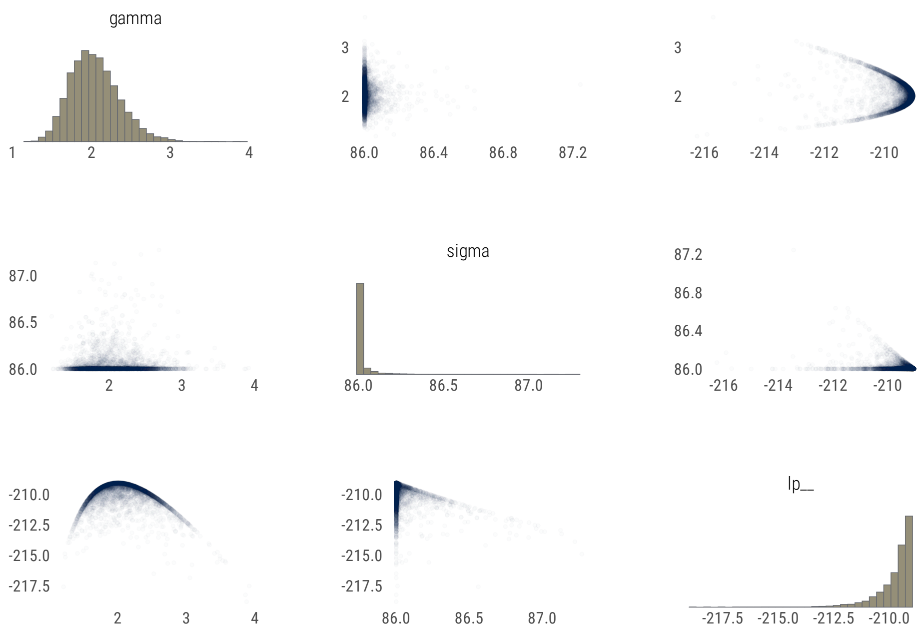

Figure 7: Posterior distribution of parameters in GenExp-Govindarajulu model

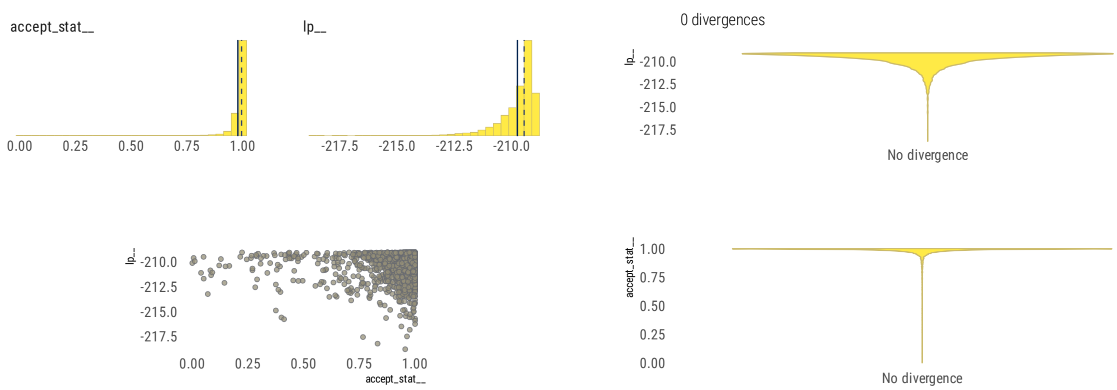

Figure 8: NUTS parameters for GenExp-Govindarajulu model

5 Numerical inverting of quantile function

The core element of *indirect* likelihood method is the use of the intermediate CDF values $\underline{u}$, corresponding to the observables $\underline{x}$ given the parameter θ. These values can either be found analytically, as $Q^{-1}(x) = F(x)$ for invertible probability distributions, or, numerically via root-finding algorithm, as $\widehat{Q}_X^{-1}(x) = \widehat{F}_X(x)$ e.g. in case of quantile distributions (Figure 1).

5.1 Target function

The problem of inverting a quantile function is tantamount to finding the root of a target function

$$\Omega(x; u|\theta) = [x - Q_X(u|\theta)] \tag{11}$$

where x is a known observation, θ is the parameter value, and $u|\theta$ is the CDF value u, such that $\Omega(x; u|\theta) = 0$. Provided that the $Q(u|\theta)$ is a non-decreasing function and x is a fixed observable value, the target function $\Omega(x; u|\theta)$ is non-increasing. The root-finding algorithm uses the target function to take an observable x and "pull in" its inverted equivalent $Q(u|\theta)$ until the two values exactly meet by iteratively adjusting $u|\theta$.

5.2 Root-finding algorithms

In principle, the choice of the algorithms for finding the zeroes of a target function Ω includes two broad groups of methods: *bracketing* and *non-bracketing* (Atkinson, 2008; Burden and Faires, 2011).

The bracketing methods, such as bisection, secant, Lagrange polynomial and Brent method, require a pair of values around the root, i.e. two values of u^+ and u^-, such that $\Omega(x; u^+|\theta) > 0$ and $\Omega(x; u^-|\theta) < 0$. In order for the algorithm to converge on a true value of probability $u|\theta$ faster, the interval $[u^+, u^-]$ needs to be relatively narrow. One way of ensuring that the starting interval is small is via the method, which we call "grid-matching". Because $Q^{-1}(x) = F(x)$ is a non-decreasing function, the interval of CDF values $[u^+, u^-]$ enclosing the true value $u|\theta$ corresponding to the observable x, can be found by matching the observable x to the sorted grid of K quantile values $Q_k^{grd} = Q_X(u_k^{grd}|\theta), \quad k \in 1..K$, where $\forall u_k^{grd}, k \in (1 \dots K) : 0 < u_k^{grd} < 1, \; u_k^{grd} < u_{k+1}^{grd}$ come from a sorted grid of CDF values. Once x is matched to the grid of quantiles Q_k^{grd}, the quantile value immediately preceding the observable x and immediately following it, such that $Q_m^{grd} \leq x \leq Q_{m+1}^{grd}$ can be determined and the corresponding CDF values u_m^{grd} and u_{m+1}^{grd} can be returned. The interval formed by these CDF values can be adopted as $[u^+, u^-]$, as it is guaranteed to contain the value of u corresponding to the root of the target function $\Omega(x; u|\theta)$.

The non-bracketing methods, such as Newton-Rhapson, rely on a single "initial guess" value $u_{(i)}$ and the gradient represented by the derivative of a target function $\Omega'(x; u_{(i)}|\theta)$. Following this method the "improved" value u^* can found as:

$$u^* \approx u_{(i)} - \frac{\Omega(x; u_{(i)})}{\Omega'(x; u_{(i)})} \tag{12}$$

Substituting the target function Z, the Newton-Raphson formula for finding the inverse of the quantile function becomes:

$$u^* \approx u_{(i)} - \frac{x - Q_X(u_{(i)}|\theta)}{d[x - Q_X(u_{(i)}|\theta)]/du_{(i)}} = u_{(i)} + \frac{x - Q_X(u_{(i)}|\theta)}{q_X(u_{(i)}|\theta)} \tag{13}$$

where $u_{(i)}$ is the initial value of the probability (the "CDF value") u, corresponding to the observation x given θ, u^* is the new (improved) value of $u_{(i)}$ after an iteration and $q_X(u_{(i)}|\theta)$ is the QDF of X, which is the first derivative of the QF $Q_X(u_{(i)}|\theta)$ with respect to the CDF value $u_{(i)}$. The procedure can be repeated by taking the approximated u^* as a new initial value $u_{(i)}$ and recomputing the value of u^*, until $|\Omega(x; u_{(i)})|\theta)| < \epsilon$, where ϵ is some small value. For faster convergence it is desirable that the initial guess value $u_{(i)}$ is as close to the true root as possible. The method has been used in the literature for approximating CDFs of several known quantile distributions (see p.99 in Gilchrist, 2000; p.345 in Nair et al., 2013).

There are other methods which can be adapted for finding the zeroes of a function, even though they may be developed for a different purpose. In the absence of other built-in root-finders, we used Stan's `algebra_solver()`, intended for finding the roots of linear systems, to numerically determine the CDF values corresponding to the observable times-to-failure (see example in Section 4 above). Stan's solver is based on the Powell hybrid method, initially developed for finding the local minimum of functions. Powell's hybrid method combines of advantages of the Newton's method with the steepest descent method, which guarantees stable convergence (Powell, 1970).

The rest of the section illustrates how to use the bracketing root-finding algorithm implemented in R for inverting the quantile function of the Tukey's *g-and-h* distribution (Rayner and MacGillivray, 2002). The `stats::uniroot()` function in R implements Brent's method, also known as the Brent-Dekkert algorithm, which combines the features of the bisection and secant methods with the inverse quadratic interpolation. Brent's algorithm performance ranks it only slightly behind more recent Riddler and Zhang methods (Stage, 2013; Zhang, 2011).

5.3 Tukey's g-and-h distribution and the normal QDF

Tukey's *g-and-h* distribution is defined by the quantile funciton:

$$Q_{gnh}(p) = A + Bz[1 + C\tanh(gz/2)]\exp(hz^2/2) \tag{14}$$

where $z = z(p) = \Phi^{-1}(p|0,1)$ is standard normal quantile function, $B > 0$ and $h > 0$. The parameter g is responsible for skeweness, i.e. when $g < 0$ the distribution is skewed left, and when $g > 0$, it is skewed right. The parameter h is responsible for kurtosis. For all values of kurtosis greater than that of the normal distribution, i.e. $h \geq 0$, regardless of value of g, the parameter $C \leq 0.83$ (approximately) results in a proper distribution. In practice the parameter C is often set to 0.8 (Prangle, 2017).

The challenge with *g-and-h* distribution is that it is expressed in terms of $z(p)$ and not in terms of p itself. In order to differentiate the *g-and-h* QF with respect to p we can use the chain rule $q(p) = \frac{dQ}{dp} = \frac{dQ_{gnh}(z)}{dz}\frac{dz}{dp}$. The second part of this expression is the QDF of the standard normal $\frac{dz}{dp} = \frac{d\Phi^{-1}(p)}{dp} = q_{norm}(p)$.

Even though normal distribution does not have a closed-form QF, we can exploit the fact that Stan and R have built-in functions for calculating the $\Phi^{-1}(p)$ and derive QDF (and DQF) of the normal distribution in terms of the normal QF:

$$\frac{d\Phi^{-1}(p)}{dp} = q_{norm}(p) = [f_{norm}(Q_{norm}(p))]^{-1} = [f(\Phi^{-1}(p))]^{-1} \tag{15}$$

Therefore the QDF of the Tukey's *g-and-h* distribution

$$q_{gnh}(p) = \left(0.5B\exp(hz^2/2)[1 + C\tanh(gz/2)](1 + hz^2) + \frac{Cgz}{2\cosh^2(gz/2)}\right)q_{norm}(p) \tag{16}$$

where $z = Q_{norm}(p|0,1)$ is the QF of standard normal and $q_{norm}(p)$ is the QDF of standard normal.

5.4 Root bracketing with grid-matching

The `pgnh()` function in `qpd` package (Perepolkin, 2019) implements $\widehat{Q}^{-1}_{gnh}(x)$ using the grid-matching method to find a pair of values bracketing the root. We use R's built-in `findInterval()` function for locating the index m of the sorted grid of K quantile values $Q_k^{grd} = Q(u_k^{grd}|\theta), \quad k \in 1..K$, such that $Q_m^{grd} \leq x \leq Q_{m+1}^{grd}$. In order to make the search of the root for the target function more numerically stable, we perform the search on the scale of z-values, i.e. standard normal quantile values corresponding to the CDF value p. Once the root is found, the value of z corresponding to the root of the target function can be converted back to the CDF value using the standard normal CDF $\Phi(z) = p$.

The following R code snippet implements the $\widehat{Q}^{-1}_{gnh}(x)$ inverse quantile function (aproximate CDF) for *g-and-h* distribution

```r
pgnh <- function(q, A,B,C=0.8,g,h, n_grid=100L, s_grid=5L, tol=1e-15, maxiter=1e3){
  # target function
  afun <- function(z, q, A,B, C, g, h) {q - qgnh(z,A,B,C,g,h, zscale=TRUE)}
  stopifnot(B>0 && h>=0)
  # service function for producing a grid of values on [0,1]
  p_grd <- make_pgrid(n=n_grid, s=s_grid)
  q_grd <- qgnh(p_grd, A,B,C, g,h, zscale=FALSE)
  idx_lower <- findInterval(q, q_grd, all.inside = TRUE)
  idx_upper <- idx_lower+1L
  int_lower <- stats::qnorm(p_grd[idx_lower]) # convert to z-scale
  int_upper <- stats::qnorm(p_grd[idx_upper]) # convert  to z-scale

  zs <- mapply(function(.q, .il, .iu) { # do this for every element in q
    tmp_zs <- NULL
    tmp_zs <- stats::uniroot(afun, q=.q, A=A, B=B, C=C, g=g, h=h,
                             interval=c(.il,.iu), extendInt="downX",
                             check.conv=TRUE, tol = tol, maxiter = maxiter)
    if(is.null(tmp_zs)) res <- NA_real_ else res <- tmp_zs$root
    res #return the value of the root
  }, q, int_lower, int_upper)

  ps <- stats::pnorm(zs) # convert the z values back to probability scale

  ps[!is.finite(ps)] <- NA_real_
```

```
  ps
}
```

We take a random sample of 100 values from the *g-and-h* distribution with parameters A=5, B=5, C=0.8, g=5 and h=0.25. The distribution of simulated values is shown in Figure 9.

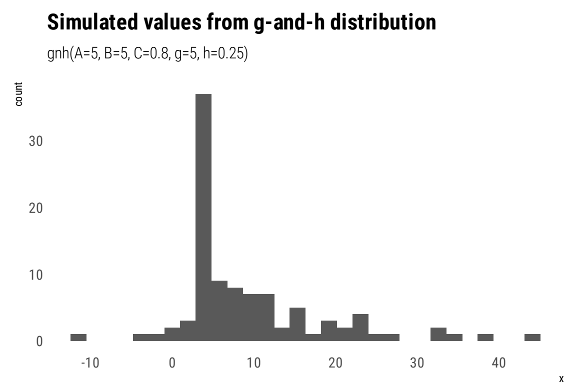

Figure 9: Simulated data from g-and-h distribution

We assumed that the parameters A and B of *g-and-h* distribution are known, and defined a normal prior for parameter g as $Normal(3, 1)$ and Rayleigh prior for parameter h as $Rayleigh(0.3)$. We performed Bayesian inference using Robust Adaptive Metropolis MCMC algorithm by Vihola (2012) interfaced by the `fmcmc` package (Vega Yon and Marjoram, 2019) in R. Table 4 and Figure 10 summarize the posterior distribution of parameters g and h.

Table 4: Posterior summary of parameters in g-and-h distribution

variable	mean	median	q5	q95	rhat
g	5.0062716	5.0136017	4.593153	5.3758618	1.002898
h	0.4223329	0.4165107	0.291208	0.5731965	1.004125

6 Validation of quantile functions

Flexibility of quantile distributions can become their curse, as certain combination of parameters may produce an invalid quantile function. In order for the quantile function to be valid, it needs to be continuous and non-decreasing. Note that it is possible for a non-decreasing quantile function to produce a multi-modal density function and still remain valid. The

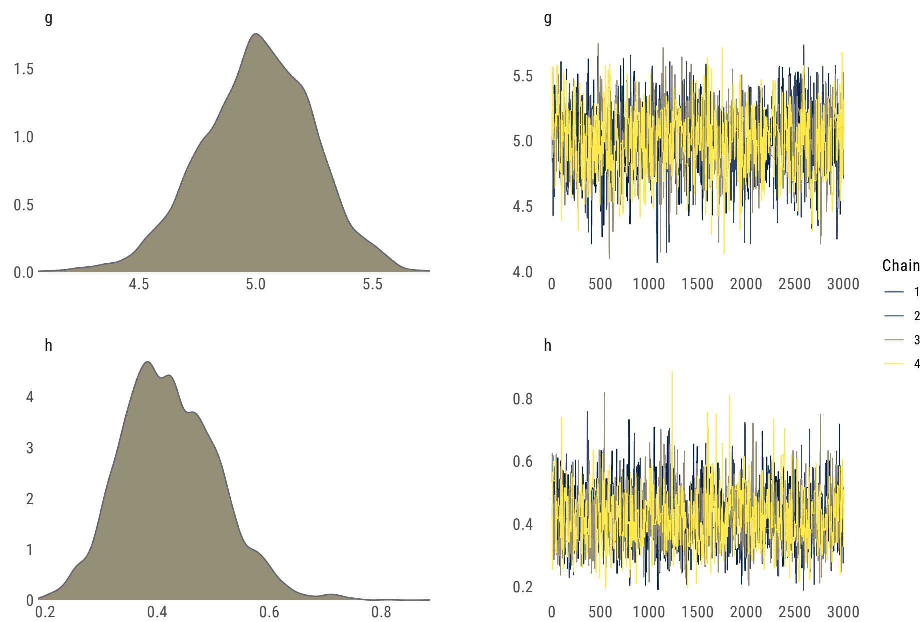

Figure 10: Posterior summary of parameters in g-and-h distribution

violation of the QF feasibility condition manifests itself in the negative values of the QDF, measuring the "rate of change" in the quantile function. An invalid shape of the quantile function can cause the CDF approximation to fail both because the root-finding algorithm, such as Newton-Rhapson, can get stuck in the local minimum and because the QDF can produce an invalid gradient.

6.1 Parameter conditions

In some instances it is possible to avoid problematic cases by limiting the allowable range of parameter values. However, in highly parameterized quantile distribution, parameters may be mutually dependent, which makes it impossible to limit the problematic cases by imposing restrictions on each parameter independently. For example, Wakeby distribution has five parameter conditions, four of which are related to more than one parameter (simultaneous conditions), some of which are mutually exlusive, e.g. "either $\beta + \sigma > 0$ or $\beta = \gamma = \delta = 0$". These complex conditions and bounds imposed on parameters may also be difficult to implement in MCMC samplers.

Expressing parameter conditions requires good understanding of the distribution, and even then there might me some combinations of parameters which produce an illegal quantile function. For example, Prangle (2017) and Rayner and MacGillivray (2002) discuss the range of valid parameter values for the *g-and-k* distribution. The parameter k shoud be larger than -0.5, but some values of $-0.5 < k < 0$ can also lead to a quantile function being invalid. The irregularities can be difficult to spot on the quantile function graph, but the effect on the derivative functions (QDF and DQF) can be quite dramatic.

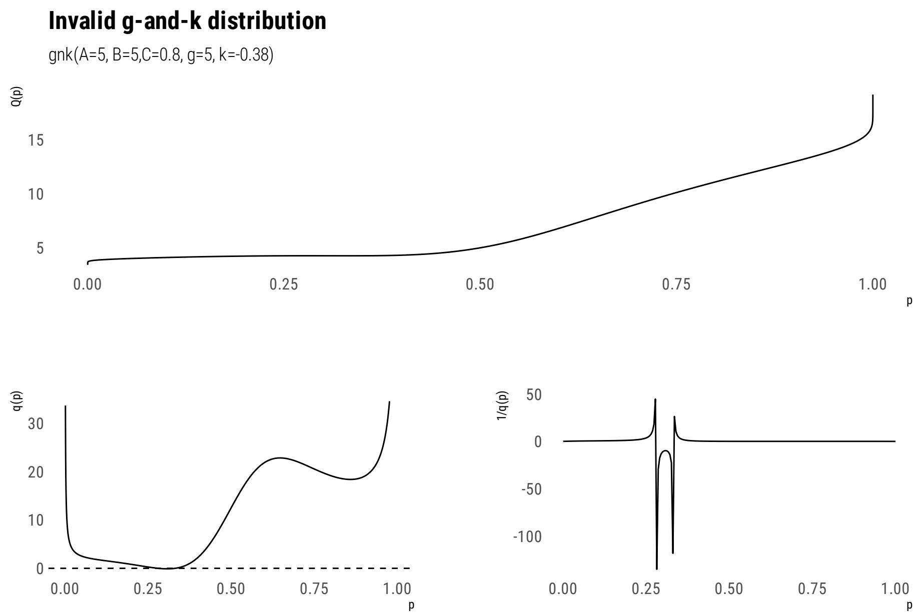

Figure 11: Some combination of parameters of g-and-k distribution may lead to an invalid quantile function

6.2 Grid-checking

The validity of the quantile function can also be checked by computing the QDF values corresponding to the tight grid of probabilities spread across the [0,1] range. Should any one of the QDF values be negative we can reject the quantile function with there parameters as invalid. This requires that the probability grid is tight enough to eliminate the possibility of the QDF "dipping" below zero between the grid values. If the cost of computing the quantile function is relatively low, and the distribution is relatively smooth (no sharp wiggles) this strategy can be reasonable. For a very tight probability grid, though, this method can become computationally expensive.

6.3 Global minimum check

An alternative approach may include finding the global minimum of the QDF and checking if it is positive. However, given the complex shape of the QDF (Figure 11), this may prove to be computationally expensive, as well.

6.4 Direct root-finding

A QDF can also be checked for roots. The values between the roots (or, if only one root is found, on one side of the root) are likely to be negative. Therefore if the roots are found this can indicate that the corresponding QF is invalid. The challenge with finding the QDF roots is that in most cases we can not use any of the bracketing methods, because the presence of roots is not known a-priori. The Newton-Rhapson method can also be problematic, due to difficulties of finding the "good" initial value. Also gradient-based methods, such as Newton-Rhapson, would require a derivative "quantile convexity function" (QCF) which may not be available, should the QDF be invalid. When validating the quantile function we are not interested in finding the roots of the QDF per se, but rather in using them as an indication that a QDF may take negative values on the range [0,1] and thus, that the corresponding QF is invalid.

6.5 Proxy root-finding

Finally, we can represent the QDF with a proxy function, the roots of which can be computed analytically. This method is referred to in the literature as the "proxy root-finding" and has been studied extensively (Boyd, 2013). Chebyshev polynomials are known for their ability to approximate functions of arbitrary complexity (Boyd and Gally, 2007). In order to increase the precision of the approximation either higher degree polynomials should be used or the function range should be partitioned, keeping the degree of the polynomials applied on the subdivisions low(Boyd, 2006). The main computational load of the Chebyshev polynomial method is computing the eigenvalues of the Chebyshev-Frobenius companion matrix (Boyd, 2013). The logic of the method relying on a large number of partitions is that it might be computationally cheaper to find eigenvalues of many small matrices, than finding eigenvalues of a large matrix. The `qpd` package (Perepolkin, 2019) implements several functions for computing the coefficients, finding roots and evaluating the Chebyshev polynomial of arbitrary degree on any interval of a function. For QPD the full range of the function is $(0,1)$, but `qpd::is_qdf_valid()` can check a QDF function passed by the user using arbitrary number of subdivisions fitting the Chebyshev polynomial of a user-defined degree to every partition. Another strategy discussed in Boyd (2013) is "recursive partitioning", where the algorithm start by fitting a single polynomial to the whole function range and comparing the goodness of fit (e.g. the sum of squares) to some small value. If the error exceeds the threshold, the range is split in two and polynomial is fit to each subsegment, repeating the algorithm recursively.

Given the shape of the QDF function (most likely U-shaped) the largest error will be in the tails, so it might be a smart idea to make more splits toward the tails and less splits in the middle of the function range. We implemented S-shaped subdivision scheme in `qpd::is_qdf_valid()` and compared its performance to the linearly partitioned quantile function.

6.6 Dealing with false roots

When using high degree polynomials on complex QPDs false-positives are not uncommon. Boyd (2006) suggests using the roots identified by the proxy-rootfinding method as starting values for the Newton-Rhapson algorithm to refine (or refute) the roots. This adds to computational complexity and requires the presence of a valid QCF. The method we adopted in the `qpd` package is based on the idea of using the proxy roots as subdivisions of a QDF (0,1) range and checking a value from every segment of the function range formed by the proxy roots (e.g. if only one proxy root is found, checking one value on each side of the root). This way the number of evaluations required for assuring non-negativity of QDF can be significantly reduced.

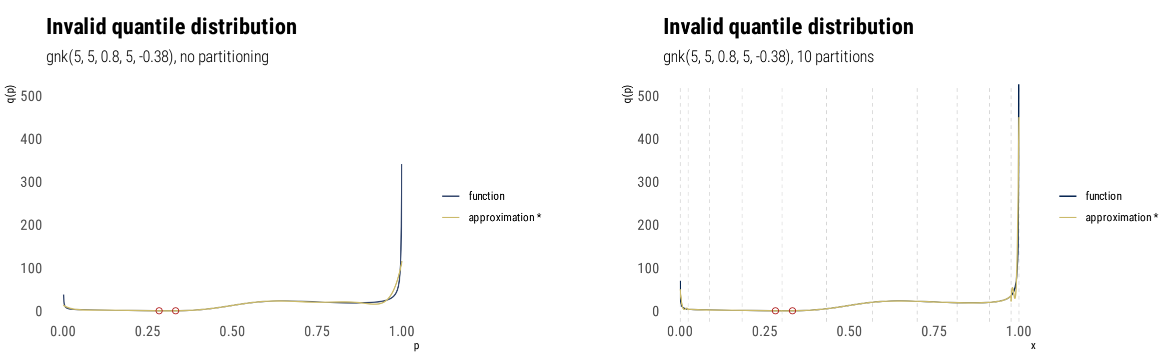

Figure 12: Chebyshev polynomial approximation: valid QDF and true-positive proxy roots

Figure 12 illustrates the ideas presented in this section. The *g-and-k* distribution

$GnK(5, 5, 0.8, 5, -0.38)$ has an invalid QF. Graph on the left shows approximation of QDF with Chebyshev polynomials of degree 10 without paritioning the range. Two roots are identified and they are correctly highlighting the area where QDF takes negative values. The function range can also be partitioned as shown on the graph to the right. Here partitioning is performed using the S-shaped scheme implemented in `qpd::make_pgrid()`. The parameter `s` is passed to the beta inverse transform so that the partitions are made at the points corresponding to $Q_{beta}(\{u_{(M)}\}, s, s)$, where $\{u_{(M)}\}$ is a uniformly distributed grid of M points. Uneven grid facilitates better fit in the tails of the QDF, but most importantly, allows us to use lower degree polynomial. In the paritioned case presented on the right, we used cubic Chebyshev polynomial fit on each of the 10 segments.

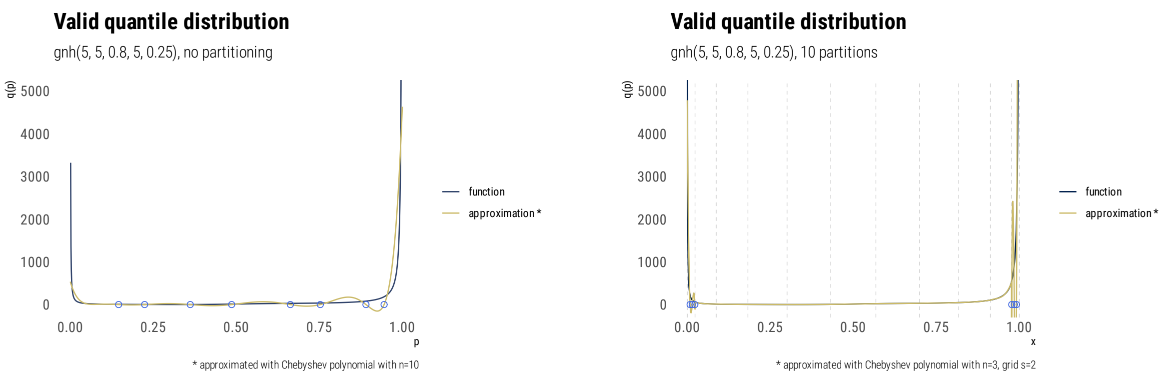

Figure 13: Chebyshev polynomial approximation: valid QDF and false-positive proxy roots

Unfortunately, not all quantile functions are behaving as nicely. Figure 13 presents a valid *g-and-h* distribution which we also approximated with Chebyshev polynomial of degree 10. All roots are false. They appear because of extreme curvature of the QDF in the tails. When the range is partitioned, the false roots move into the outer-most bins, which are in fact "safest" areas of the QDF curve (above the y-axis) even by visual inspection. We need to evaluate the function less than 10 times to refute the roots and assure ourselves that the function is valid. Increasing the degree of the fitted polynomial (or increasing the number and/or density of the QDF range partitions) might reduce the number of false roots and move them even further out towards the tails. Given the low number of roots, and relatively inexpensive evaluation of QDF, these strategies may not be justifiable.

7 Discussion

7.1 Indirect inference

Quantile distributions are expanding the menu of options available to the scientists and offer a wide selection of distributions to be used as likelihoods and/or priors. Moreover new distributions can be created on demand, because the sum of two valid quantile functions is again a quantile function (Gilchrist, 2000). Newly created quantile distribution is guaranteed to be amenable to inverse sampling (by definition). In this paper we presented an *inverse likelihood* method which makes them useable as sampling distributions as well. A class of quantile distributions, which we did not cover in this paper, namely, quantile-parametrized distributions (Keelin and Powley, 2011; Hadlock and Bickel, 2017; Keelin, 2016) can be especially useful as priors, given their highly interpretable parameterization.

The first application of Bayesian inference for quantile distributions was proposed by Parzen (2004), who outlined the method of approximate Bayesian inference using the rejection sampling with the comparison density function (see sections 21-24 in Parzen, 2004). In the recent years, many more applications of the approximate Bayesian computation (ABC) to the models defined by the quantile distributions appeared in the literature (Allingham et al., 2009; Drovandi and Pettitt, 2011; Dunson and Taylor, 2005; McVinish, 2012; and Smithson and Shou, 2017). ABC methods normally do not require computation of the likelihood, which in case of quantile distributions is convenient, as they lack an explicit CDF and PDF. The few Bayesian applications of quantile distributions involving computation of the likelihood found in the literature, rely on the numerical methods for inverting the quantile function and integrating it into the PDF (Prangle, 2017; Bernton et al., 2019). The Bayesian inference proposed in this paper makes numeric integration unnecessary, as the density calculation becomes available in closed form using the density quantile function.

7.2 Practical tips for implementing quantile distributions in Stan and in R.

We implemented several quantile distributions in R (see `qpd` package) and in Stan (see Supplementary Materials) using certain conventions which might simplify their usage in other applications. First of all, we followed the prefix convention in R for naming probability functions "p" for CDF (e.g. `pnorm()`), "q" for QF (e.g. `qnorm()`), "d" for PDF (`dnorm()`) and "r" for random number generators (`rnorm`). We extended the prefix list for quantile distributions: "f" for QDF (e.g. `fnorm()`), "dq" for DQF (e.g. `dqnorm()`) and "ff" for QCF, second derivative of QF (e.g. `ffnorm()`). Moreover, all approximate inverse QF (or CDF) functions have *tolerance* and *maximum number of iterations* arguments exposed (with reasonable default values), which indicate that even though the "p"-prefixed function is available, it is based on the approximation. We used bracketing root-finding algorithm with grid-matching to provide initial values to the target function (exposing the control over the grid to the user).

In Stan, custom functions are implemented in the `functions{}` block. In most cases we had to implement a QF (for various transformations, including inside target function), sometimes both vectorized and scalar version, because user functions in Stan can have only one signature. We also implemented a QDF to be used, for example, in the Newton-Rhapson method, which requires the first derivative of QF and as a service function inside DQF. Because DQF is used in likelihood, Stan requires that the function name ends with `_lpdf` suffix. Should the tilde sampling notation of be used, the suffix can be omitted. We hope that in the future quantile density suffix `_ldqf` will be accepted as well, but until that time we opted for double-suffixing.

As we mentioned Stan allows only one signature for user function per name. Although it is possible to implement vectorized log-DQF function to be used for likelihood, we opted for scalar version (and therefore put it inside the loop over the observations). Implementing scalar version of the log-density-quantile function is unavoidable for specification of the indirect prior, so in order to minimize the code repetition we use scalar version for the likelihood as well. In the process of testing the Stan code, none of our custom probability functions caused problem or divergences. Challenges were more often caused by the lookup function (which performs the grid-matching) or the rootfinding functions (we used built-in `algebra_solver()`). We could not get the recently implemented `algebra_solver_newton()` to work inside the indirect likelihood without divergences, so we used the default Powell hybrid algorithm. When implementing custom functions that will be used outside of `transformed data{}` block, the usage of step-like functions should be minimized, in order to avoid possible issue with divergent transitions (see Stan Functions Reference Section 3.7 for details).

We calculated the CDF-values $\underline{u}$ directly in the model block inside the likelihood loop and not as `transformed parameters{}`. Below is the example of the `model{}` block from our GenExp-Govindarajulu model. Implementing 1D root-finding using the solver for linear algebra is a little awkward. We hope Stan will get its own ("auto-diffable") Brent's root-finder soon.

```
model {
  vector[N] u;
```

```
  // create grid of quantile values
  vector[M] xs_grd = govindarajulu_v_qf(ys_grd, gamma, sigma);
  // look-up u corresponding to closes value from the grid
  vector[N] u_guess = vlookup(x_srt, xs_grd, ys_grd);
  // priors
  target += genexp_s_lpdf(gamma | genexp_alpha, genexp_lambda);
  target += exponential_lpdf(dsigma | exp_lambda);
  // likelihood
  for (i in 1:N){
   // numerical inversion
   u[i] = approx_govindarajulu_cdf_algebra(x_srt[i], u_guess[i], gamma, sigma, rel_tol,
   // likelihood statement
   target += govindarajulu_s_ldqf_lpdf(u[i] | gamma, sigma);
  }
}
```

We found that when using *indirect* inference (both in Stan and in R), good initial values are essential. We tried to cue initial values closer to the mode of the prior to facilitate better mixing of MCMC chains.

Embracing and expanding the use of quantile distributions in Bayesian inference can enable new solutions for old problems and enrich the toolkit available to scientists for performing hard inference tasks. We hope that the *indirect* inference methods presented in this paper can contribute to expanding body of knowledge in Bayesian statistics and fuel further research in the area of quantile distributions.

8 Appendix A. Distributions used in the paper

8.1 Exponential distribution

Exponential distribution function $F(x)$ and the probability density function $f(x)$ are given by

$$\begin{aligned} F(x) &= 1 - e^{-\lambda x} \\ f(x) &= \lambda e^{-\lambda x} \end{aligned} \tag{17}$$

where $\lambda > 0$ and $x \in [0, \infty)$.

Exponential quantile function $Q(u)$ and quantile density function $q(u)$ are

$$
\begin{aligned}
Q(u) &= -\frac{\ln(1-u)}{\lambda} \\
q(u) &= \frac{dQ(u)}{du} = \frac{1}{\lambda(1-u)}
\end{aligned}
\tag{18}
$$

where $\lambda > 0$ and $u = F(x), u \in [0, 1]$

8.2 Rayleigh distribution

Rayleigh distribution function $F(x)$ and probability density function $f(x)$ are:

$$
\begin{aligned}
F(x|\sigma) &= 1 - \exp(-x^2/(2\sigma^2)) \\
f(x|\sigma) &= \frac{x}{\sigma^2}\exp(-x^2/(2\sigma^2))
\end{aligned}
\tag{19}
$$

where $\sigma > 0$ is Rayleigh scale parameter.

Rayleigh quantile function $Q(p)$ and quantile density function $q(p)$ are:

$$
\begin{aligned}
Q(p|\sigma) &= \sigma\sqrt{-2\ln(1-p)} \\
q(p|\sigma) &= \frac{\sigma}{\sqrt{2}\sqrt{-\ln(1-p)}(1-p)}
\end{aligned}
\tag{20}
$$

where $\sigma > 0$ and $p \in [0, 1]$.

8.3 Govindarajulu distribution

Govindarajulu distribution defined by the quantile function has the following QF and QDF:

$$
\begin{aligned}
Q_X(u) &= \sigma\gamma u^\gamma(1 + \gamma^{-1} - u) \\
q_x(u) &= K u^{\gamma-1}(1-u)
\end{aligned}
\tag{21}
$$

where $K = \sigma\gamma(\gamma + 1)$, and $\gamma > 0$.

The distribution has support on $(Q(0), Q(1)) = (0, \sigma)$[2].

[2] For definition of the Govindarajulu distribution with shifted support see Nair et al. (2012)

8.4 Generalized exponential distribution

Generalized exponential distribution has the following CDF and PDF:

$$\begin{aligned} F(x) &= (1 - e^{-\lambda x})^{\alpha} \\ f(x) &= \alpha\lambda(1 - e^{-\lambda x})^{\alpha-1} e^{-\lambda x} \end{aligned} \tag{22}$$

where $x, \alpha, \lambda > 0$.

References

Aarset, M. V. (1987). How to Identify a Bathtub Hazard Rate. *IEEE Transactions on Reliability*, R-36(1):106–108.

Aldeni, M., Lee, C., and Famoye, F. (2017). Families of distributions arising from the quantile of generalized lambda distribution. *Journal of Statistical Distributions and Applications*, 4(1):25.

Allingham, D., King, R. A. R., and Mengersen, K. L. (2009). Bayesian estimation of quantile distributions. *Statistics and Computing*, 19(2):189–201.

Andrilli, S. F. and Hecker, D. (2010). *Elementary Linear Algebra*. Elsevier Academic Press, Amsterdam ; Boston, 4th ed edition.

Atkinson, K. E. (2008). *An Introduction to Numerical Analysis*. John wiley & sons.

Bernton, E., Jacob, P. E., Gerber, M., and Robert, C. P. (2019). Approximate Bayesian computation with the Wasserstein distance. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 81(2):235–269.

Boyd, J. P. (2006). Computing real roots of a polynomial in Chebyshev series form through subdivision. *Applied Numerical Mathematics*, 56(8):1077–1091.

Boyd, J. P. (2013). Finding the Zeros of a Univariate Equation: Proxy Rootfinders, Chebyshev Interpolation, and the Companion Matrix. *SIAM Review*, 55(2):375–396.

Boyd, J. P. and Gally, D. H. (2007). Numerical experiments on the accuracy of the Chebyshev–Frobenius companion matrix method for finding the zeros of a truncated series of Chebyshev polynomials. *Journal of Computational and Applied Mathematics*, 205(1):281–295.

Burden, R. L. and Faires, J. D. (2011). *Numerical Analysis*. Brooks/Cole, Cengage Learning, Boston, MA, 9th ed edition.

Chalabi, Y., Scott, D. J., and Wuertz, D. (2012). Flexible distribution modeling with the generalized lambda distribution.

Dedduwakumara, D. S., Prendergast, L. A., and Staudte, R. G. (2021). An efficient estimator of the parameters of the generalized lambda distribution. *Journal of Statistical Computation and Simulation*, 91(1):197–215.

Drovandi, C. C. and Pettitt, A. N. (2011). Likelihood-free Bayesian estimation of multivariate quantile distributions. *Computational Statistics & Data Analysis*, 55(9):2541–2556.

Dunson, D. B. and Taylor, J. A. (2005). Approximate Bayesian inference for quantiles. *Journal of Nonparametric Statistics*, 17(3):385–400.

Fournier, B., Rupin, N., Bigerelle, M., Najjar, D., Iost, A., and Wilcox, R. (2007). Estimating the parameters of a generalized lambda distribution. *Computational Statistics & Data Analysis*, 51(6):2813–2835.

Freimer, M., Kollia, G., Mudholkar, G. S., and Lin, C. T. (1988). A study of the generalized Tukey lambda family. *Communications in Statistics-Theory and Methods*, 17(10):3547–3567.

Gilchrist, W. (2000). *Statistical Modelling with Quantile Functions*. Chapman & Hall/CRC, Boca Raton.

Gupta, R. D. and Kundu, D. (2007). Generalized exponential distribution: Existing results and some recent developments. *Journal of Statistical Planning and Inference*, 137(11):3537–3547.

Hadlock, C. C. and Bickel, J. E. (2017). Johnson Quantile-Parameterized Distributions. *Decision Analysis*, 14(1):35–64.

Hadlock, C. C. and Bickel, J. E. (2019). The Generalized Johnson Quantile-Parameterized Distribution System. *Decision Analysis*, 16(1):67–85.

Haynes, M. and Mengersen, K. (2005). Bayesian estimation of g-and-k distributions using MCMC. *Computational Statistics*, 20(1):7–30.

Jacob, P. (2017). Likelihood calculation for the g-and-k distribution.

Keelin, T. W. (2016). The Metalog Distributions. *Decision Analysis*, 13(4):243–277.

Keelin, T. W. and Powley, B. W. (2011). Quantile-Parameterized Distributions. *Decision Analysis*, 8(3):206–219.

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2004). *Loss Models: From Data to Decisions*. Wiley Interscience, Hoboken, N.J, 2nd ed edition.

McVinish, R. (2012). Improving ABC for quantile distributions. *Statistics and Computing*, 22(6):1199–1207.

Nadarajah, S. (2009). Bathtub-shaped failure rate functions. *Quality & Quantity*, 43(5):855–863.

Nair, N. U., Sankaran, P., and Balakrishnan, N. (2013). *Quantile-Based Reliability Analysis*. Springer New York, New York, NY.

Nair, N. U., Sankaran, P. G., and Dileepkumar, M. (2020). Bayesian inference in quantile functions. *Communications in Statistics - Theory and Methods*, 0(0):1–13.

Nair, N. U., Sankaran, P. G., and Vineshkumar, B. (2012). The Govindarajulu Distribution: Some Properties and Applications. *Communications in Statistics - Theory and Methods*.

Parzen, E. (1979). Nonparametric Statistical Data Modeling. *Journal of the American Statistical Association*, 74(365):105–121.

Parzen, E. (1980). Data Modeling Using Quantile and Density-Quantile Functions. Technical report, TEXAS A AND M UNIV COLLEGE STATION INST OF STATISTICS.

Parzen, E. (2004). Quantile Probability and Statistical Data Modeling. *Statistical Science*, 19(4):652–662.

Perepolkin, D. (2019). *Qpd: Tools for Quantile-Parameterized Distribiutions*.

Powell, M. J. (1970). A hybrid method for nonlinear equations. *Numerical methods for nonlinear algebraic equations*.

Prangle, D. (2017). Gk: An R Package for the g-and-k and generalised g-and-h Distributions. *arXiv:1706.06889 [stat]*.

Pratt, J. W., Raiffa, H., and Schlaifer, R. (1995). *Introduction to Statistical Decision Theory*. MIT press.

Price, G. B. (1984). *The Inverse-Function Theorem*, pages 237–262. Springer New York, New York, NY.

Rahman, A., Zaman, M. A., Haddad, K., El Adlouni, S., and Zhang, C. (2015). Applicability of Wakeby distribution in flood frequency analysis: A case study for eastern Australia. *Hydrological Processes*, 29(4):602–614.

Rayner, G. D. and MacGillivray, H. L. (2002). Numerical maximum likelihood estimation for the g-and-k and generalized g-and-h distributions. *Statistics and Computing*, 12(1):57–75.

Schlaifer, R. and Raiffa, H. (1961). *Applied Statistical Decision Theory*.

Smithson, M. and Shou, Y. (2017). CDF-quantile distributions for modelling random variables on the unit interval. *British Journal of Mathematical and Statistical Psychology*, 70(3):412–438.

Stage, S. A. (2013). Comments on an improvement to the Brent's method. *International Journal of Experimental Algorithms*, 4(1):1–16.

Stan Development Team (2021). RStan: The R interface to Stan.

Tukey, J. W. (1965). Which Part of the Sample Contains the Information? *Proceedings of the National Academy of Sciences*, 53(1):127–134.

Vega Yon, G. and Marjoram, P. (2019). Fmcmc: A friendly MCMC framework. *The Journal of Open Source Software*, 4(39).

Vihola, M. (2012). Robust adaptive Metropolis algorithm with coerced acceptance rate. *Statistics and Computing*, 22(5):997–1008.

Zhang, Z. (2011). An improvement to the Brent's method. *International Journal of Experimental Algorithms*, 2(1):21–26.