The tenets of indirect inference in Bayesian models

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

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

> Don't try to understand it. Feel it.
> – Laura in TENET by Christopher Nolan

> The exact shape doesn't matter
> – Dodo in Alice's Adventures in Wonderland by Lewis Carroll

1 Introduction

Introduction paragraph to be added.

1.1   Aims of the paper

- Provide vocabulary
- Show examples and demonstrate the equivalence of the methods
- Clearly explain numerical inversion of quantile function. It has been done poorly before and only using Newton method, whereas we also use Brent)
- Propose novel approach for validating quantile functions, essential for MCMC, because the parameter space is explored randomly and crucial for numerical inversion of quantile functions.
- Introduce indirect prior and discuss its applications.

1.2   Paper structure

Section 2 of this paper reviews the different ways of specifying the probability distributions and their classification depending on the invertibility of their distribution function. It sets the scene for the following sections by reiterating the conditions for the function to be invertible and the relationship between inverse functions and their derivatives. We review the use of the inverse cumulative distribution function (quantile function) to describe statistical distributions and discuss several examples of the distributions defined by the quantile function ("quantile distributions"), found in the scientific literature.

Section 3 leverages the vocabulary and the definitions introduced in the previous section to describe and exemplify the a Bayesian updating approach proposed by Nair et al. (2020) for the distributions which lack an explicit distribution function. We introduce *direct* and *indirect* likelihood using the identities and substitutions introduced in Section 2. We then show the equivalence of two ways of expressing likelihood in Bayesian models and provide an example illustrating the concepts and showcasing their implementation in Stan and R. We then extend the same logic of substitution using the quantile function to define the *direct* and *indirect* prior. 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. We provide several illustrative examples of implementing various combinations of direct and indirect likelihood and prior in Bayesian models using invertible probability distributions and compare the posteriors from these models. All models are using the the Hamiltonian Monte Carlo (HMC) "No U-Turn Sampler" in Stan interfaced by the `rstan` package in R(Stan Development Team, 2021) for updating the parameters in a quantile distribution.

Section 4 discusses practical aspects of using quantile sampling distributions in Bayesian models. Many specialized distributions have been developed for modeling special kinds of

$P(X \leqslant x) = u$
Random variable X

described by   *described by*

$F(x)$
Cumulative distribution function

$Q(u)$
Quantile function

	Not invertible	**Invertible**	**Not invertible**
	$F^{-1}(u)$ is intractable	$F^{-1}(u) = Q(u)$ $Q^{-1}(x) = F(x)$	$Q^{-1}(x)$ is intractable
Parametric	Normal*, Gamma	Most distributions	g-and-k*, g-and-h*, GLD, Govindarajulu, Wakeby, etc.
QPDs †	Q-Normal*, Myerson*		(G)J-QPD, Metalog
	PROBABILITY DISTRIBUTIONS		**QUANTILE DISTRIBUTIONS**

* Based on the Normal CDF Φ and numeric estimation of Φ^{-1}
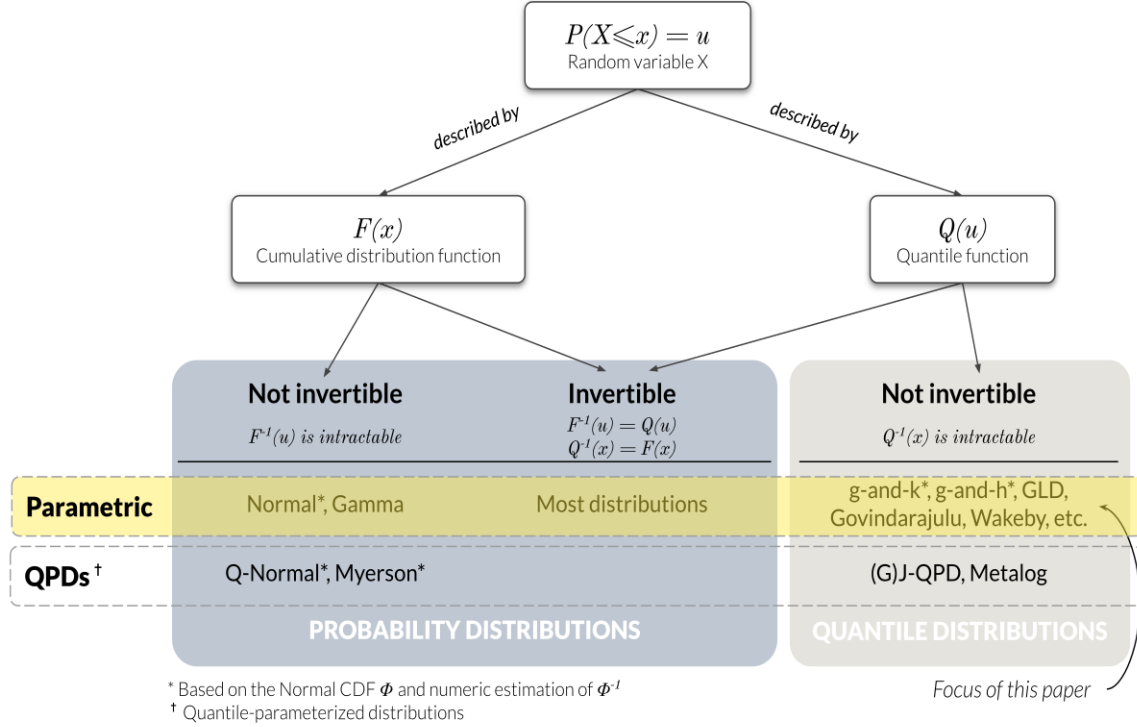† Quantile-parameterized distributions

Focus of this paper

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

data. These distributions lend themselves naturally as sampling distributions in Bayesian models. Unfortunately some of them are quantile distributions, so Bayesian inference with them was complicated. We provide an example of using such distribution in reliability analysis.

In Section 5 discusses the computational aspects of estimating the intermediate CDF values in quantile distributions. We propose a target function and consider several root-finding algorithms which can we applied for numerically inverting the quantile function. The example in this section shows bayesian updating using the 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 algorithm for inverting a quantile function in R.

Section 6 discusses a problem of validating the quantile function, i.e. checking that the combination of parameters defining the distribution is feasible. We present the challenge of checking the quantile function and present the common approaches used for this task. We then propose the use of Chebyshev polynomials (Boyd, 2006) for approximating and root-checking of the quantile density function and discuss the practical aspects of implementing such validation. 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 a custom function 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.

2 Distribution specification

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)(Gilchrist, 2000).

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).

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

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. The derivative of the inverse, then, is equal to the reciprocal of the 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 distribution function $F(x) = u$, if a PDF $f(x)$ exists and $f(x) \neq 0$, then QDF $q(u)$ also exists and it is equal to $q(u) = [f(x)]^{-1}$. In Section [X] of this paper, we rely on the density quantile function (DQF) $[q(u|\theta)]^{-1}$ to define the likelihood in a Bayesian model based on the quantile distribution. We call this method of specifying the likelihood "indirect", as it is using the probability u, representing the observable x (given the parameter θ), and not the observable x directly.

Even though quantile distributions lack the closed-form CDF $F_X(x|\theta)$, in most cases, the it 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 probability u. The details of the numerical inversion algorithms are discussed in Section 4.

A paragraph on quantile distributions interpreting the chart. Although in this paper we focus primarily on the parametric quantile distributions, the quantile distributions parameterized by the quantile-probability pairs (quantile-parametrized quantile distributions) are also briefly mentioned as they represent an important application and extension of quantile distributions.

Statistical methods utilizing QF and QDF were pioneered by the seminal work of Parzen (1979), although a few years prior to that Tukey (1965) discussed these functions referring to them as the "representing function" and the "sparsity index", respectively.

3 Bayesian inference for quantile functions

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 published Bayesian applications of quantile distributions involving computation of the likelihood, rely on the numerical methods for inverting the quantile function and integrating the PDF (Prangle, 2017; Bernton et al., 2019).

In their recent work Nair et al. (2020) showed that the traditional Bayesian inference can be equivalently restated using the substitutions involving the quantile function.

3.1   Direct and indirect likelihood

Assume that the prior information about the parameter θ can be summarized by the prior distribution of $f(\theta)$. 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 $\underline{Q} = \{Q_1(u_1), Q_2(u_2), \dots Q_n(u_n)|\theta\}$, where $u_i = F(x_i|\theta), i = 1 \dots n$ (the probabilities u_i can be also denoted as $\underline{u}|\theta$). 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 $u_i = F(x_i|\theta), i = 1 \dots n$, which are then supplied to the quantile function $\underline{Q}$.

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.

3.1.1 Example

To illustrate the equivalence of the two ways of specifying likelihood in Bayesian models, we use the Example 16.17 from Klugman et al. (2004) on p.544 regarding the distribution of the claim amounts (referred to hereinafter, as the Claims Example).

In this example, the claim amounts are modeled using an exponential distribution with the mean of $1/\lambda$, where λ is following the gamma distribution with the shape $\alpha = 4$ and scale parameter $\beta = 0.001$. Given three observations of the claim amounts $\underline{x} = 100$, 950 and 450, the posterior distribution of the parameter λ is sampled using the HMC algorithm 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.

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

$$
\begin{aligned}
F(x) &= 1 - e^{-\lambda x} \\
f(x) &= \lambda e^{-\lambda x}
\end{aligned}
\tag{9}
$$

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

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

$$
\begin{aligned}
Q(u) &= -\frac{\ln(1 - u)}{\lambda} \\
q(u) &= \frac{dQ(u)}{du} = \frac{1}{\lambda(1 - u)}
\end{aligned}
\tag{10}
$$

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

Since gamma prior is conjugate to the exponential sampling distribution, the posterior density of λ can be described by

$$
f(\lambda|\underline{x}) = \text{Gamma}(\lambda|\alpha + N, \beta + \sum \underline{x})
\tag{11}
$$

where α and β are the parameters of gamma distribution, $\underline{x} = \{x_1, x_2, ... x_N\}$ is the sample of observations of size N.

Figure 2 and Table 1 compares the posterior distribution of θ from the Gamma-Exponential model with direct and indirect likelihood, along with the posterior from the conjugate model. Stan programs for the corresponding examples are provided in the Supplementary Material.

Gamma-Exponential model

Comparing posterior distributions from models with direct and indirect likelihoods
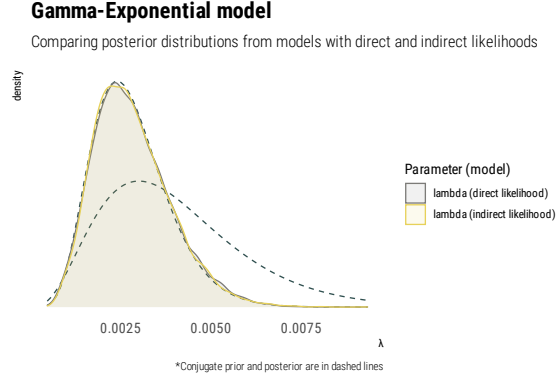
Conjugate prior and posterior are in dashed lines

Figure 2: Summary of the posterior samples from the gamma-exponential model with direct and indirect likelihood

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

3.2 Direct and indirect prior

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 $Q(v) = \theta$. Substituting the quantile values $Q(v)$ 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 probability v, given hyperparameter(s) of the prior distribution (12).

$$f(Q(v)|\underline{x}) \propto \mathcal{L}(Q(v); \underline{x}) f(Q(v))$$
$$[q(v|\underline{x})]^{-1} \propto \mathcal{L}(Q(v); \underline{x})[q(v)]^{-1} \tag{12}$$

where $[q(v|\underline{x})]^{-1}$ is the indirect form of the quantile posterior, $[q(v)]^{-1}$ is the indirect form of quantile prior and $\mathcal{L}(Q(v); \underline{x})$ is the direct likelihood, relying on the non-linear parameter transformation $\theta = Q(v)$. The likelihood with such parameter transformation requires a Jacobian adjustment which is equal to the absolute derivative of the transform, i.e. $J(Q(v)) = |dQ(v)/dv| = |q(v)|$. We refer to such formulation of the prior as the *indirect prior* because relies on the quantile transformation $\theta = Q(v)$ and therefore its density is expressed in the quantile form $[q(v)]^{-1}$.

Of course indirect prior can also be used in combination with indirect likelihood, since, as we argued previously, the two of them lead to the same posterior beliefs about the θ and, consequently, v. In such case, neither prior nor likelihood would require the existence of the closed form PDF and therefore can be represented by quantile distributions.

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

$$[q(v|\underline{x})]^{-1} \propto \mathcal{L}(Q(v); \underline{x})[q(v)]^{-1}|q(v)|$$
$$[q(v|\underline{x})]^{-1} \propto \mathcal{L}(Q(v); \underline{x})$$

(13)

where $[q(v|\underline{x})]^{-1}$ is the quantile form of the posterior, and the prior belief about the parameter θ is represented by the quantile function transform of the parameter $v \in [0,1]$ (given the relevant hyperparameters), such that $Q(v) = \theta$. This formulation is commonly known as the "inverse transform" because it relies on the QF (inverse CDF) transformation of the unit parameter v.

3.2.1 Example

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 can switch to the Rayleigh distribution to represent the expert's beliefs about the exponential parameter λ with the quantile density. The shape of the new prior is comparable to the gamma distribution used previously.

Prior distribution of λ
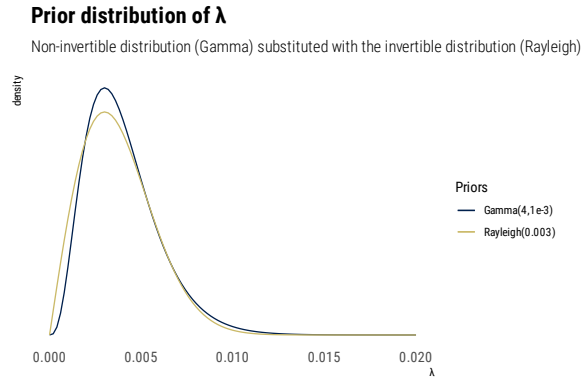Non-invertible distribution (Gamma) substituted with the invertible distribution (Rayleigh)

Figure 3: Prior distribution of parameter

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

$$F(x|\sigma) = 1 - \exp(-x^2/(2\sigma^2)) f(x|\sigma) = \frac{x}{\sigma^2} \exp(-x^2/(2\sigma^2)) \tag{14}$$

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

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

$$Q(p|\sigma) = \sigma\sqrt{-2\ln(1-p)} q(p|\sigma) = \frac{\sigma}{\sqrt{2}\sqrt{-\ln(1-p)}(1-p)} \tag{15}$$

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

Figure 4 compares the posterior distribution of θ from the gamma-exponential model with direct and indirect likelihood. Stan programs for corresponding examples are provided in the Supplementary Material.
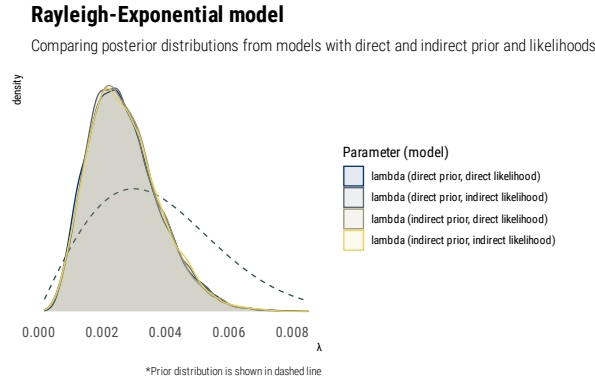

Rayleigh-Exponential model
Comparing posterior distributions from models with direct and indirect prior and likelihoods

Parameter (model)
lambda (direct prior, direct likelihood)
lambda (direct prior, indirect likelihood)
lambda (indirect prior, direct likelihood)
lambda (indirect prior, indirect likelihood)

*Prior distribution is shown in dashed line

Figure 4: Summary of the posterior samples from the Rayleigh-Exponential model with direct and indirect prior and likelihood

4 Specialized quantile sampling distributions

Indirect likelihood would be an curious, yet impractical method of performing Bayesian inference if it did not have an application in models with quantile sampling distributions. Many quantile distributions have been developed with particular application in mind, offering some unique advantages, which are no possible to replicate with conventional density-defined distributions.

Nair et al. (2020) cite an example from Aarset (1987) of time to failure of 50 devices put on the life test at time 0. The data reflects bathtub-shaped hazard rate 5. Lifetime reliability data are often modeled using the three- (or two-) component mixtures, or one of the specialized distributions (Nadarajah, 2009). Nair et al. (2020) used maximum likelihood to estimate the parameters of the Govindarajulu distribution (Nair et al., 2012) described by the generalized exponential prior (Gupta and Kundu, 2007). We reuse the Aarset (1987) example and estimate the full posterior distribution of the parameters in the GenExp-Govindarajulu model.
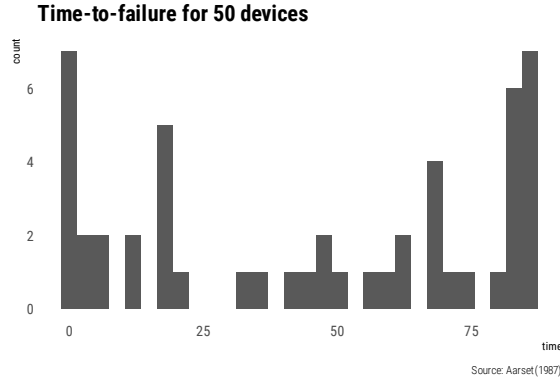

Figure 5: Histogram of the time to failure data

Govindarajulu distribution is defined by QF:

$$Q_X(u) = \sigma\gamma u^\gamma(1 + \gamma^{-1} - u) \tag{16}$$

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

Govindarajulu QDF:

$$q_x(u) = Ku^{\gamma-1}(1 - u) \tag{17}$$

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

Generalized exponential CDF:

$$F(x) = (1 - e^{-\lambda x})^\alpha \tag{18}$$

for $x, \alpha, \lambda > 0$

---

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

Generalized exponential PDF:

$$f(x) = \alpha\lambda(1 - e^{-\lambda x})^{\alpha-1}e^{-\lambda x} \tag{19}$$

Generalized exponential QF

$$Q(u) = -\frac{1}{\lambda}ln(1 - u^{1/\alpha}) \tag{20}$$

We used generalized exponential prior for the parameter γ of Govindarajulu distribution with hyperparameters $ =$5 and $ =$1. The parameter σ of the Govindarajulu distribution can not be lower than the maximum of the observed times-to-failure. We defined exponential prior with $ =$0.5 and varying lower bound corresponding to the highest time to failure equal to 86. The Jacobian for adding a constant, although required, is equal to one, so there will be no impact on the log density, as the shifting transform produces a Jacobian derivative matrix with a unit determinant.
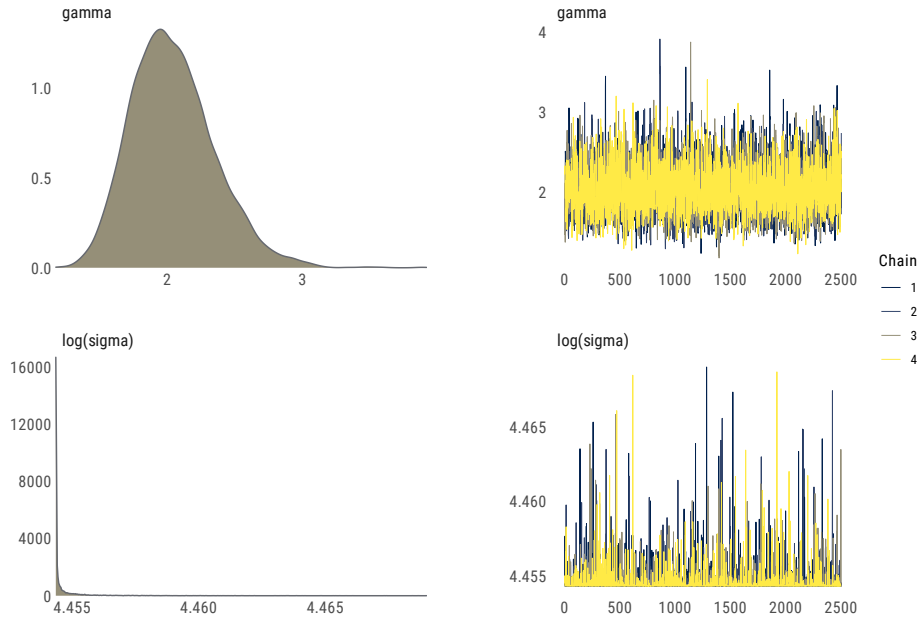

Figure 6: Posterior distribution and traceplot for parameters of GenExp-Govindarajulu model

5   Numerical inverting of quantile function

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

The problem of inverting a quantile function is tantamount to finding the root of a target function $Z(u|\theta) = [x - Q_X(u|\theta)]$ where x is a known observation, θ is the parameter value, and $u|\theta$ is the probability u, such that $Z(u|\theta) = 0$. In principle the choice of the algorithms for finding the zeroes of a target function Z includes two broad groups of methods: *bracketing* and *approximating* (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 $Z(u^-|\theta) < 0$ and $Z(u^+|\theta) > 0$. In order for the algorithm to converge faster, the interval $[u^-, u^+]$ needs to be relatively narrow. As $Q^{-1}(x) = F(x)$ is a non-decreasing function, the interval of probabilities $[u^-, u^+]$ 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 $\forall u_k^{grd}, k \in (1 \ldots K) : 0 < u_k^{grd} < 1$, $u_k^{grd} < u_{k+1}^{grd}$ come from a sorted grid of probability 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 $Q_m^{grd} \le x \le Q_{m+1}^{grd}$ can be determined and the corresponding probabilities u_m^{grd} and u_{m+1}^{grd} can be returned. The interval formed by these probability values is guaranteed to contain the value of u corresponding to the root of the target function $Z(u|\theta)$.

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

$$u^* \approx u_{(i)} - \frac{Z(u_{(i)})}{Z'(u_{(i)})} \tag{21}$$

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

$$u^* \approx u_{(i)} - \frac{x - Q_X(u_{(i)}|\theta)}{d[x - Q_X(u_{(i)}|\theta)]/du_{(i)}} = u_{(i)} + \frac{x - Q_X(u_{(i)}|\theta)}{q_X(u_{(i)}|\theta)} \tag{22}$$

14

where $u_{(i)}$ is the initial value of probability u, corresponding to the observation x given the value of the parameter 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 probability $u_{(i)}$. The procedure can be repeated by taking the approximated u^* as a new initial value $u_{(i)}$ and recomputing the new value of u^*, until $|Z(u_{(i)})|\theta)| < \epsilon$, where ϵ is some small value.

Again, 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).

The Figure 7 illustrates the numerical approximation of probability u corresponding to the observed claim amount equal to 100 in the Claims Example given the sampling model *Exponential*(0.002). We provide the probability guess of 0.5 as pretty ignorant initial value. The algorithm required 4 iterations to arrive at the approximation of the true probability of 0.1812692 given the desired tolerance, which for the purposes of this example we set at 10^{-3}.
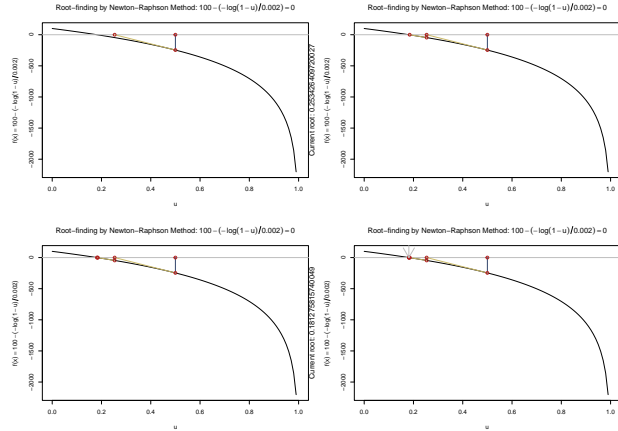

Figure 7: Approximating the probability in the Exponential model using Newton-Rhapson method

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, the previous example utilizing the Givindarajulu likelihood used Stan's solver, intended for finding the roots of linear systems, to numerically find the CDF values corresponding to the observable times to failure. Stan's `algebra_solver()` is based on the Powell hybrid method, initially developed for finding the local minimum of a function (Powell, 1970). Powell's hybrid method combines of advantages of the Newton's method with the steepest descent method, which guarantees stable convergence.

In this section we illustrate the use of the Brent's root-finding method implemented in R as

`stats::uniroot()` for inverting the quantile function of the Tukey's *g-and-h* distribution (Rayner and MacGillivray, 2002).

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

$$Q_{gnh}(p) = A + Bz[1 + C \tanh(gz/2)] \exp(hz^2/2) \tag{23}$$

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).

g-and-h distribution QDF can be obtained by differentiating the QF with respect to the probability p using the chain rule $q(p) = \frac{dQ}{dp} = \frac{dQ_{gnh}(z)}{dz}\frac{dz}{dp}$. The second part of this is the QDF of 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) in terms of QF:

$$\frac{d\Phi^{-1}(p)}{dp} = q_{norm}(p) = [f_{norm}(Q_{norm}(p))]^{-1} = [f(\Phi^{-1}(p))]^{-1} \tag{24}$$

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

$$q_{gnh}(p) = \left(0.5B \exp(hz^2/2)[1 + C\tanh(gz/2)](1 + hz^2) + \frac{Cgz}{2\cosh^2(gz/2)}\right) q_{norm}(p) \tag{25}$$

The `pgnh()` function in `qpd` package (Perepolkin, 2019) implements $\widehat{Q}_{gnh}^{-1}(x)$ 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 $Q_k^{grd} = Q(u_k^{grd}|\theta)$, $k \in 1..K$, 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 probability p. Once the root is found, the value of z corresponding to the root of the target function can be converted back to the probability using the standard normal CDF $\Phi(z) = p$.

Figure 8 shows the distribution of errors of approximating 10000 simulated probability values using the numerical inversion of the g-and-h QF following the described method. After

simulating the values of $p \in (0, 1)$ we calculate the quantile values $Q(p)$ and supply them to the approximation algorithm implemented in *qpd :: pgnh* and compare the result to the original (known) probabilities.

Error of approximating the g-and-h CDF values
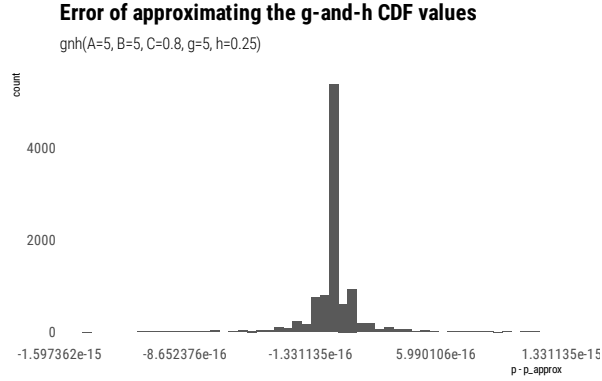
gnh(A=5, B=5, C=0.8, g=5, h=0.25)

Figure 8: Distribution of approximation error for g-and-h CDF values

The absolute error of the approximation does not exceed $1.4432899 \times 10^{-15}$. The desired precision can be set as the argument in `qpd::pgnh()`.

We take a sample of 100 simulated 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.

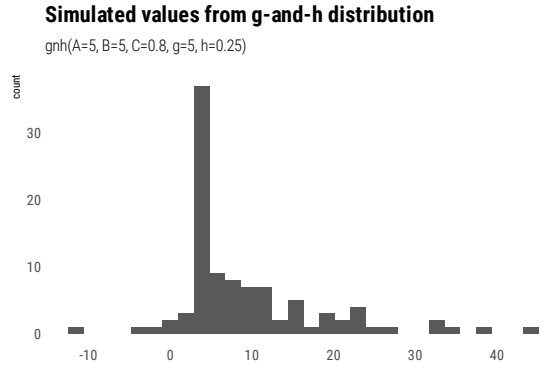**Simulated values from g-and-h distribution**

gnh(A=5, B=5, C=0.8, g=5, h=0.25)

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 $N(3, 1)$ and Rayleigh prior for parameter h as $Rayleigh(0.3)$. We performed Bayesian inference using Robust Adaptive Metropolis MCMC algorithm by

17

Vihola (2012) interfaced by the `fmcmc` package (Vega Yon and Marjoram, 2019) in R. Table 2 and Figure 10 summarize the posterior distribution of parameters *g* and *h*.

Table 2: Posterior summary of parameters in g-and-h distribution

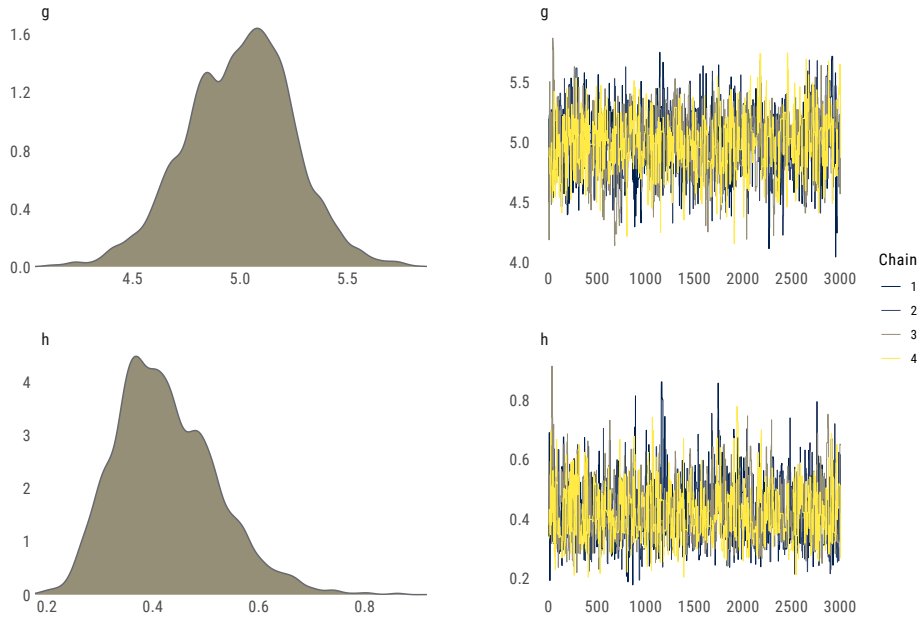| variable	mean	median	q5	q95	rhat
g	4.9965104	5.011251	4.5878278	5.3899243	1.006485
h	0.4248716	0.413202	0.2876798	0.5961615	1.002276

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

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 (DQF) and still remain. The violation of feasibility condition of the QF manifests itself in the negative values of QDF, as it measures the "rate of change" in the quantile function.

In some cases it is possible to avoid problematic cases by limiting the range of values for the parameters. For example, Prangle (2017) and Rayner and MacGillivray (2002) discuss the

range of valid parameter values for *g-and-k* distribution. The parameter k shoud be larger than -0.5, but some values of k, between -0.5 and 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.
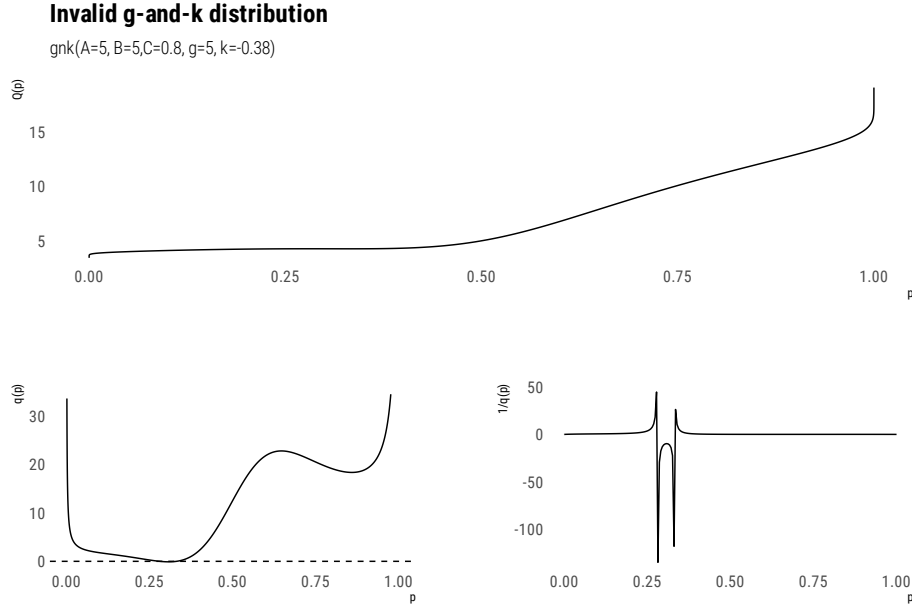
Invalid g-and-k distribution

gnk(A=5, B=5,C=0.8, g=5, k=-0.38)

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

The invalid shape of the quantile function can cause the CDF approximation to fail both because the search (e.g. using the Newton-Rhapson method) can get stuck in the local minima and because the QDF can produce an invalid gradient.

The validity of the quantile function can 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 be tight enough to eliminate the situations where the QDF "dips" into the negative space between the grid values. For a very tight probability grid, this method can be computationally expensive.

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

Finally, the QDF can 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

19

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 in advance. Newton-Rhapson method can also be problematic, due to difficulties of finding the "good" starting value. Also the approximation methods 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 per se, but rather in using them as an indication that a QDF may take negative values on the range [0,1].

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 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 paritioned, keeping the degree of the polynomials applied on the subdivisions low. Boyd (2006) discusses two methods correponding to these strategies. The first of them referred to as the "tredecic method", envisages that regardless of number of partitions the polynomial of fixed degree (M=13) is used. The other method called "megacubes" refers to applying a low-degree polynomial on a very large number of partitions (enumerated in thousands). The tradeoff is computing the low degree polynomial many times or relatively high degree polynomial a few times. 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 "megacubes" method (and any method relying on a large number of paritions) is that it might be computionationally cheaper to find eigenvalues of many small matrices, compared to 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 range of the function is $(0, 1)$, but `qpd::is_qpd_valid()` can check a qpd function passed by the user on arbitrary number of subdivisions (using either uniform or an S-shaped subdivision scheme) automating the process of subdividing the range of a function and fitting the Chebyshev polynomial of a user-defined degree to every partition.

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 debunk the presence of) the roots. This adds to computational complexity and requires the presence of a valid QCF. The method adopted by the `qpd` package is based on the idea of using the proxy roots as subdivisions of a QDF (0,1) range and checking one value from every segment of the function range formed by the proxy roots (e.g. if only one proxy root is found, checking a single value to the left
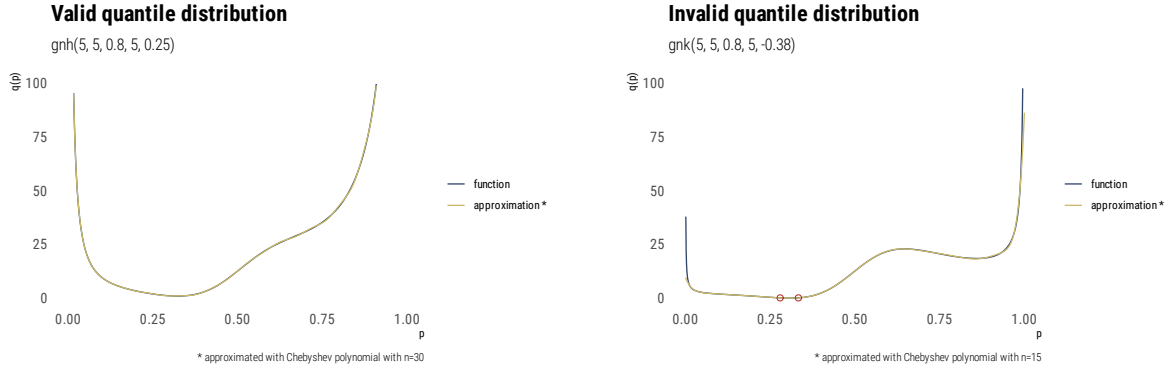
and to the right of the root).

Figure 12: Chebyshev polynomial approximation for validating two quantile distributions

DP Comments:

- For the Specialized distributions I have the flood flow dataset, which has its own distribution - Wakeby, which was developed for modeling of floods or thunderstorms. Wakeby has crazy parametrization (many impossible combinations of parameters) so it would be cool to discuss how it can be done in Stan.

- Especially when we talk about indirect priors, they make sense almost solely in the context of QPDs, ie when we have expert-elicited assessment of parameters. For Wakeby distribution there are some parameters from the published papers and some can just be given by opinion. I am hesitant to introduce any of the QPDs in this paper. Lets keep that Pandora box locked until the second paper.

- I can easily include small snippets of code if that is considered educational. Just need to find a way of not being overwhelming. All models are in the `stan/` folder in the repo.

7 Discussion

> You are inverted, the world is not.
> – Wheeler to the Protagonist in TENET by Christopher Nolan

Here some of the ideas for which I can write a paragraph or two of discussion.

7.1 Moebius loop of the probability functions

maybe this can be moved upfront

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

7.3 Finding good "initial values"

Challenges with grid-matching in Stan (divergences)

7.4 Root-finding algorithms in Stan and R

`nlm()`? numerical derivatives? For Stan I just want to rant that there's still no built-in 1D rootfinder.

7.5 Performance tradeoffs with Chebyshev polynomials

This is worth a separate paper and a lot has been written, so just citing the literature. Reality is that it is expensive and loses to the tight grid for QDFs of moderate complexity. Needs implementation in C++.

7.6 Expert-specified indirect priors

This will hint towards the quantile-parametrized quantile distribution, foreshadowing the next paper.

References

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

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

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.

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.

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.

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.

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.

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.

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.