Bayesian Estimation of Distributions

March 22, 2021

The following is an approach to finding a parametric estimate of a conditional distribution.

The approach described below is loosely inspired from quantile regression, which we review here briefly.

1 Brief Review of Quantile Regression

If Y is a real-valued random variable with cumulative distribution function $F_Y(y) = P(Y \leq y)$, then the τth quantile of Y is given by the quantile function

$$Q_Y(\tau) = F_Y^{-1}(\tau) = \inf\{y : F_Y(y) \geq \tau\},$$

for some selected quantile $\tau \in (0, 1)$. The loss function used in quantile regression, due to Koenker and Bassett (1978), is defined as

$$\rho_\tau(u) = u(\tau - \mathbb{I}_{(u<0)}), \tag{1}$$

for $\mathbb{I}$ an indicator function. To see that we may obtain the τth quantile of Y by minimising the above, consider an estimate $\hat{y}$ for this quantile. We seek

$$\begin{aligned}
\min_{\hat{y}} \mathbb{E}\big[\rho_\tau(Y - \hat{y})\big] &= \min_{\hat{y}} \int_{-\infty}^{\infty} \rho_\tau(y - \hat{y}) dF_Y(y) \\
&= \min_{\hat{y}} \left\{ (\tau - 1) \int_{-\infty}^{\hat{y}} (y - \hat{y}) dF_Y(y) + \tau \int_{\hat{y}}^{\infty} (y - \hat{y}) dF_Y(y) \right\}.
\end{aligned} \tag{2}$$

Using Leibniz's rule, we can differentiate this expression with respect to $\hat{y}$, which gives

$$\begin{aligned}
\frac{d}{d\hat{y}} \mathbb{E}\big[\rho_\tau(Y - \hat{y})\big] &= \frac{d}{dx}(\tau - 1) \int_{-\infty}^{\hat{y}} (y - \hat{y}) dF_Y(y) + \tau \int_{\hat{y}}^{\infty} (y - \hat{y}) F_Y(y) \\
&= (\tau - 1) \left((\hat{y} - \hat{y}) + \int_{-\infty}^{\hat{y}} \frac{\delta}{\delta \hat{y}} (y - \hat{y}) dF_Y(y) \right) \\
&\quad + \tau \left(-(\hat{y} - \hat{y}) + \int_{\hat{y}}^{\infty} \frac{\delta}{\delta \hat{y}} (y - \hat{y}) dF_Y(y) \right) \\
&= (1 - \tau) F_Y(\hat{y}) - \tau\big(1 - F_Y(\hat{y})\big) \\
&= F_Y(\hat{y}) - \tau.
\end{aligned}$$

Since both of the terms in (2) are positive, the loss function is convex and so setting the above expression equal to zero and solving gives $F_Y(\hat{y}) = \tau$. Hence $\hat{y}$ is indeed the τth quantile of the random variable Y, as required.

In the approach of quantile regression, we seek to approximate the τth conditional quantile function, $Q_Y(\tau|X) = f_{\mathbf{w}_\tau}(X)$, where $f_{\mathbf{w}_\tau}(X)$ is a parametric function approximator with parameters $\mathbf{w}_\tau$. We may obtain the parameters $\mathbf{w}_\tau$ by solving:

$$\mathbf{w}_\tau = \underset{\mathbf{w}\in\mathbb{R}^d}{\arg\min}\, \mathbb{E}\big[\rho_\tau(Y - f_{\mathbf{w}}(X))\big].$$

Given that we usually don't have the distribution function $F_Y(y)$ of Y available and we are instead estimating the parameter values from data, we may instead estimate the parameters from the observations using

$$\hat{\mathbf{w}}_\tau = \underset{\mathbf{w}\in\mathbb{R}^d}{\arg\min} \sum_{i=1}^{n} \rho_\tau\big(y_i - f_{\mathbf{w}}(\mathbf{x}_i)\big).$$

For a probabilistic analysis of the above, since the quantile regression loss function is always positive, we may re-frame it as a member of the exponential family of distributions by simply taking its negative exponent. This gives an asymmetric Laplace density of the following form (Yu and Moyeed, 2001)

$$e^{-\mathbb{E}\left[\rho_\tau(Y - f_{\mathbf{w}_\tau}(X))\right]} = \frac{\tau^n(1-\tau)^n}{\sigma} \exp\left\{ -\sum_{i=1}^{N} \rho_\tau\left(\frac{y_i - f_{\mathbf{w}_\tau}(\mathbf{x}_i)}{\sigma}\right)\right\},$$

for some scale parameter σ, and τ acting as the asymmetry parameter. For modelling the median, $\tau = 0.5$ and this becomes equivalent to the Laplace distribution. This can be treated as a likelihood function, and maximising this is an effective approach to finding the parameters $\mathbf{w}_\tau$ which give the τth conditional quantile.

However a single τ fit doesn't do justice to the potential of this framework which is to model all quantiles simultaneously; that is, $Q_Y(\tau|X)$ for all $\tau \in (0, 1)$. For this, we need a different approach.

2 A Different Approach

We propose a new approach to modelling distributions which is inspired by the quantile regression framework.

Suppose that the probability of y is found as the following density:

$$P(y) = \frac{1}{Z} \prod_{i=1}^{N} P_i(y),$$

for Z some normalising term and each $P_i(y)$ of the form $\exp\{-g(y)\}$ for $g(\cdot)$ a strictly positive real function. Taking logarithms, we may express the log-probability of y, $\mathcal{L}(y) = \log P(y)$ as the following sum of terms:

$$\mathcal{L}(y) = \sum_{i=1}^{N} \mathcal{L}_i(y) - \log Z.$$

Drawing inspiration from (2), where the residuals are weighted differently for positive and negative residuals, we define each term to be of the form:

$$\mathcal{L}_i(y) = -(y - f_i)\Big(\alpha_i \Theta(y - f_i) - \beta_i \Theta(f_i - y)\Big),$$

for Θ a Heaviside step function, and α_i, β_i and f_i some trainable parameters. We assume $\alpha_i, \beta_i > 0$ for all $i \in [1, N]$, and that the f_i are ordered, such that $f_i \leq f_{i+1}$.

To calculate the value of the normalising term, first consider an expression for the segment where $y \in (f_j, f_{j+1}]$

$$\begin{aligned}
\mathcal{L}(f_j < y \leq f_{j+1}) &= -\sum_{i=1}^{j} \alpha_i (y - f_i) + \sum_{i=j+1}^{N} \beta_i (y - f_i) - \log Z \\
&= \left(\sum_{i=j+1}^{N} \beta_i - \sum_{i=1}^{j} \alpha_i \right) y + \left(\sum_{i=1}^{j} \alpha_i f_i - \sum_{i=j+1}^{N} \beta_i f_i \right) - \log Z \\
&\doteq a_j y + b_j - \log Z
\end{aligned}$$

Where we have defined

$$a_j = \sum_{i=j+1}^{N} \beta_i - \sum_{i=1}^{j} \alpha_i \tag{3}$$

and

$$b_j = \sum_{i=1}^{j} \alpha_i f_i - \sum_{i=j+1}^{N} \beta_i f_i. \tag{4}$$

We can also consider the segment where $y \in (-\infty, f_1]$:

$$\begin{aligned}
\mathcal{L}(-\infty < y \leq f_1) &= \sum_{i=1}^{N} \beta_i (y - f_i) - \log Z \\
&= \underbrace{\sum_{i=1}^{N} \beta_i}_{a_0} y - \underbrace{\sum_{j=1}^{N} \beta_i f_i}_{b_0} - \log Z
\end{aligned}$$

as well as the segment for $y \in (f_N, \infty)$:

$$\begin{aligned}
\mathcal{L}(f_N < y < \infty) &= -\left(\sum_{i=1}^{N} \alpha_i (y - f_i) \right) - \log Z \\
&= \underbrace{-\sum_{i=1}^{N} \alpha_i}_{a_N} y + \underbrace{\sum_{i=1}^{N} \alpha_i f_i}_{b_N} - \log Z,
\end{aligned}$$

where we can see that a_0, b_0, a_N and b_N are consistent with the definition of a_j in Eq. 3 and b_j in Eq. 4.

Now the normalising term Z can be found by summing the integrals of each of these line segments. Beginning with the case where $f_j < y \leq f_{j+1}$:

$$\begin{aligned}
\int_{f_j}^{f_{j+1}} e^{a_j y + b_j}\, dy &= \left[\frac{1}{a_j} e^{a_j y + b_j} \right]_{f_j}^{f_{j+1}} \\
&= \frac{1}{a_j} e_j^b \left(e^{a_j f_{j+1}} - e^{a_j f_j} \right).
\end{aligned}$$

When $y \in (-\infty, f_1]$, we have

$$\begin{aligned}
\lim_{\ell \to -\infty} \int_{\ell}^{f_1} e^{a_0 y + b_0}\, dy &= \lim_{\ell \to -\infty} \left[\frac{1}{a_0} e^{a_0 y + b_0} \right]_{\ell}^{f_1} \\
&= \lim_{\ell \to -\infty} \frac{1}{a_0} e^{b_0} \left(e^{a_0 f_1} - e^{a_0 \ell} \right) \\
&= \frac{1}{a_0} e^{a_0 f_1 + b_0},
\end{aligned}$$

since $a_0 > 0$. Similarly for the segment where $y \in (f_N, \infty)$,

$$\begin{aligned}
\lim_{\ell \to \infty} \int_{f_N}^{\ell} e^{a_N y + b_N}\, dy &= \lim_{\ell \to \infty} \left[\frac{1}{a_N} e^{a_N y + b_N} \right]_{f_N}^{\ell} \\
&= \lim_{\ell \to \infty} \frac{1}{a_N} e^{b_N} \left(e^{a_N \ell} - e^{a_N f_N} \right) \\
&= -\frac{1}{a_N} e^{a_N f_N + b_N}
\end{aligned}$$

since $a_N < 0$. We can now explicitly write down the normalising term Z as the sum of these integrals:

$$\begin{aligned}
Z &= \int_{-\infty}^{f_1} e^{a_0 y + b_0}\, dy + \sum_{j=1}^{N} \int_{f_j}^{f_{j+1}} e^{a_j y + b_j}\, dy + \int_{f_N}^{\infty} e^{a_N y + b_N}\, dy \\
&= \frac{1}{a_0} e^{a_0 f_1 + b_0} + \sum_{j=1}^{N} \frac{1}{a_j} e^{b_j} \left(e^{a_j f_{j+1}} - e^{a_j f_j} \right) - \frac{1}{a_N} e^{a_N f_N + b_N}
\end{aligned}$$

2.1 Efficient Implementation

A naïve computation of Z would run in $O(n^2)$ time, however with a very simple dynamic programming approach, where we compute partial sums of α_i and β_i as well as $\alpha_i f_i$ and $\beta_i f_i$ ahead of time, we can find Z in just $O(n)$ time. (See notebook for details).

References

Roger Koenker and Gilbert Bassett. Regression Quantiles. *Econometrica*, 46 (1):33, January 1978. ISSN 00129682. doi: 10.2307/1913643. URL https://www.jstor.org/stable/1913643?origin=crossref.

Keming Yu and Rana A. Moyeed. Bayesian quantile regression. *Statistics & Probability Letters*, 54(4):437–447, October 2001. ISSN 0167-7152. doi: 10.1016/S0167-7152(01)00124-9. URL https://www.sciencedirect.com/science/article/pii/S0167715201001249.