APPENDIX

A PROOFS

A.1 PROOF OF THEOREM 1

Theorem 1 *For any Model* $\mathsf{M} : \mathcal{X} \to (\mathcal{Y} \to [0,1])$ *and given canonical calibration mapping* $\mathsf{R}(p) = \mathbb{P}\big[[\mathsf{M}(X)](Y) \leq p\big]$, $\mathsf{R} \circ \mathsf{M}$ *is quantile calibrated*

Proof: To show that $\mathsf{R} \circ \mathsf{M}$ is quantile calibrated. we need to show that $\mathbb{P}[(\mathsf{R} \circ \mathsf{M})[X][Y] \leq p] = p$, $\forall p \in [0,1]$ since we are assuming that $\mathsf{R}\,(p)$ is invertible function, which gives us that it is surjective. An equivalent way of showing this is that $\mathbb{P}[\,(\mathsf{R} \circ \mathsf{M})\,[X][Y] \leq \mathsf{R}(p)] = \mathsf{R}\,(p)\ \forall p \in [0,1]$

$$
\begin{aligned}
\mathbb{P}\Big[(\mathsf{R} \circ \mathsf{M})[X][Y] \leq \mathsf{R}(p)\Big] &= \mathbb{P}\Big[\mathsf{R}^{-1}\big((\mathsf{R} \circ \mathsf{M})[X][Y]\big) \leq \mathsf{R}^{-1}\big(\mathsf{R}(p)\big)\Big] && \mathsf{R}^{-1} \text{ is strictly increasing} \\
&= \mathbb{P}\Big[(\mathsf{M}[X])[Y] \leq p\Big] \\
&= \mathsf{R}(p) && \text{By definition}
\end{aligned}
$$

$\square$

A.2 PROOF OF CLAIM 1,2,3

Claim 1 *Let Y be a random variable with CDF F, and let $G = \mathsf{R} \circ F$ be its CDF after composing with mapping R obtained from isotonic regression characterized by $\mathcal{C} = \{c_{(1)}, c_{(2)}, \cdots, c_{(n)}\}$. If there exist $i-1, i, i+1 \in \{0, n\} \wedge c_{(i)} - c_{(i-1)} \neq c_{(i+1)} - c_{(i)}$ then the CDF G is not differentiable and its corresponding probability density function g is not continuous at $F^{-1}(c_{(i)})$*

Proof: First, G can be expressed as follows

$$
G(x) = \begin{cases}
\dfrac{F(x)}{nc_{(1)}}. & -\infty < x \leq F^{-1}(c_{(1)}) \\
\dfrac{F(x) - c_{(1)}}{n(c_{(2)} - c_{(1)})} + \frac{1}{n} & F^{-1}(c_{(1)}) < x \leq F^{-1}(c_{(2)}) \\
\dfrac{F(x) - c_{(2)}}{n(c_{(3)} - c_{(2)})} + \frac{2}{n} & F^{-1}(c_{(2)}) < x \leq F^{-1}(c_{(3)}) \\
\vdots & \vdots \\
\dfrac{F(x) - c_{(n-1)}}{c_{(n)} - c_{(n-1)}} + \frac{n-1}{n} & F^{-1}(c_{(n-1)}) < x \leq F^{-1}(c_{(n)})
\end{cases}
$$

Let $a = F^{-1}(c_{(1)})$. We will show that G not differentiable at a. Similarly we can show that it is not differentiable at the other $n-2$ switching points.

The left derivative is as follows

$$
\lim_{x \to a^-} \frac{G(x) - G(a)}{x - a} = \lim_{x \to a^-} \frac{\dfrac{F(x)}{n.c_{(1)}} - \dfrac{F(a)}{n.c_{(1)}}}{x - a} = \frac{1}{nc_{(1)}} \lim_{x \to a^-} \frac{F(x) - F(a)}{x - a} = \frac{F'(a)}{nc_{(1)}}
$$

The right derivative is as follows

$$\lim_{x\to a^+}\frac{G(x)-G(a)}{x-a}=\lim_{x\to a^+}\frac{\frac{F(x)-c_{(1)}}{n(c_{(2)}-c_{(1)})}+\frac{1}{n}-\frac{1}{n}}{x-a} \tag{1}$$
$$=\frac{1}{n.(c_{(2)}-c_{(1)})}\lim_{x\to a^+}\frac{F(x)-F(a)}{x-a}=\frac{F'(a)}{n.(c_{(2)}-c_{(1)})}$$

Hence G is not differentiable.

Although the CDF is not differentiable at only a finite number of points, we can still get the PDF by piece-wise differentiation.

$$g(x)=\begin{cases}\frac{f(x)}{nc_{(1)}}. & -\infty < x \leq F^{-1}(c_{(1)})\\ \frac{f(x)}{n(c_{(2)}-c_{(1)})} & F^{-1}(c_{(1)}) < x \leq F^{-1}(c_{(2)})\\ \frac{f(x)}{n(c_{(3)}-c_{(2)})} & F^{-1}(c_{(2)}) < x \leq F^{-1}(c_{(3)})\\ \vdots & \vdots \\ \frac{f(x)}{n.(c_{(n)}-c_{(n-1)})} & F^{-1}(c_{(n-1)}) < x \leq F^{-1}(c_{(n)})\end{cases} \tag{2}$$

Now consider for any $i-1, i, i+1 \in \{0, n\}$ s.t $c_i - c_{i-1} \neq c_{i+1} - c_i$. Let $a = F^{-1}(c_i)$ then

$$\lim_{x\to a^-} g(x)=\lim_{x\to a^-}\frac{f(x)}{n(c_{(i)}-c_{(i-1)})}=\frac{f(a)}{n(c_{(i)}-c_{(i-1)})}$$

$$\lim_{x\to a^+} g(x)=\lim_{x\to a^+}\frac{f(x)}{n(c_{(i+1)}-c_{(i)})}=\frac{f(a)}{n(c_{(i+1)}-c_{(i)})}$$

Since the right limit and left limit do not coincide and by construction of point a, we have that limit does not exist and therefore $g(x)$ is not continuous at a

Note that, most of times, the hypothesis is satisfied, so the smoothness is lost. $\square$

Claim 2 *Let Y_{iso} be transformed random variable after applying isotonic mapping R on random variable Y. Then the expectation of Y_{iso} is as follows*

$$\mathbb{E}[Y_{iso}]=\mu-\frac{\sigma^2}{n}\sum_{i=0}^{n-1}\frac{f(F^{-1}(c_{(i+1)}))-f(F^{-1}(c_{(i)}))}{(c_{(i+1)}-c_{(i)})}$$

Proof: Assume that, before transformation, the random variable is distributed $X\sim\mathcal{N}(\mu,\sigma^2)$ so $f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\frac{-(x-\mu)^2}{2\sigma^2}$

$$
\begin{aligned}
\mathbb{E}[Y_{\text{iso}}] &= \int_{-\infty}^{\infty} x.g(x)dx \\
&= \sum_{i=0}^{n-1} \int_{F^{-1}(c_{(i)})}^{F^{-1}(c_{(i+1)})} x.\frac{f(x)}{n.(c_{(i+1)} - c_{(i)})} dx \\
&= \sum_{i=0}^{n-1} \frac{1}{n(c_{(i+1)} - c_{(i)})} \int_{F^{-1}(c_{(i)})}^{F^{-1}(c_{(i+1)})} x.\frac{1}{\sqrt{2\pi.\sigma^2}} \exp \frac{-(x-\mu)^2}{2.\sigma^2} dx \\
&= \sum_{i=0}^{n-1} \frac{1}{n(c_{(i+1)} - c_{(i)})} \int_{F^{-1}(c_{(i)})}^{F^{-1}(c_{(i+1)})} (x-\mu+\mu).\frac{1}{\sqrt{2\pi.\sigma^2}} \exp \frac{-(x-\mu)^2}{2.\sigma^2} dx \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad x = x - \mu + \mu \\
&= \sum_{i=0}^{n-1} \frac{1}{n(c_{(i+1)} - c_{(i)})} \int_{F^{-1}(c_{(i)})}^{F^{-1}(c_{(i+1)})} (x-\mu)\frac{1}{\sqrt{2\pi.\sigma^2}} \exp \underbrace{\frac{-(x-\mu)^2}{2.\sigma^2}}_{\text{= t and use sub}} dx \\
&\quad + \sum_{i=0}^{n-1} \frac{1}{n(c_{(i+1)} - c_{(i)})} \int_{F^{-1}(c_{(i)})}^{F^{-1}(c_{(i+1)})} \mu\frac{1}{\sqrt{2\pi.\sigma^2}} \exp \frac{-(x-\mu)^2}{2.\sigma^2} dx \qquad \text{using linearity} \\
&= \sum_{i=0}^{n-1} \frac{1}{n(c_{(i+1)} - c_{(i)})} \Big[\frac{-\sigma^2}{\sqrt{2\pi.\sigma^2}} \exp \frac{-(x-\mu)^2}{2.\sigma^2} \Big]_{x=F^{-1}(c_{(i)})}^{x=F^{-1}(c_{(i+1)})} \\
&\quad + \sum_{i=0}^{n-1} \frac{\mu}{n(c_{(i+1)} - c_{(i)})} F(F^{-1}(c_{(i+1)})) - F(F^{-1}(c_i)) \qquad \text{using def of cdf} \\
&= \sum_{i=0}^{n-1} \frac{-\sigma^2}{n} \frac{f(F^{-1}(c_{(i+1)})) - f(F^{-1}(c_{(i)}))}{(c_{(i+1)} - c_{(i)})} + \mu \underbrace{\sum_{i=0}^{n-1} \frac{1}{n} \frac{F(F^{-1}(c_{(i+1)})) - F(F^{-1}(c_i))}{(c_{(i+1)} - c_{(i)})}}_{1} \\
&= \mu - \frac{\sigma^2}{n} \sum_{i=0}^{n-1} \frac{f(F^{-1}(c_{(i+1)})) - f(F^{-1}(c_{(i)}))}{(c_{(i+1)} - c_{(i)})}
\end{aligned}
$$

$\square$

Lemma 1 *Let $f_{\mu,\sigma}, F_{\mu,\sigma}, F_{\mu,\sigma}^{-1}$ be density, distribution, and quantile functions, respectively, of the normal distribution with mean μ and std σ. Then*

$$
f_{\mu,\sigma}\Big[F_{\mu,\sigma}^{-1}\big(F_{\mu_0,\sigma_0}(y_0)\big)\Big] = \frac{\sigma_0}{\sigma} f_{\mu_0,\sigma_0}(y_0)
$$

Proof: We use the following three properties of normally distributed random variables

1. $F_{\mu,\sigma}(y) = F_{0,1}(\frac{y-\mu}{\sigma})$
2. $f_{\mu,\sigma}(y) = \frac{1}{\sigma} f_{0,1}(\frac{y-\mu}{\sigma})$
3. $F_{\mu,\sigma}^{-1}(p) = \sigma.F_{0,1}^{-1}(p) + \mu$

$$
\begin{aligned}
f_{\mu,\sigma}\Big[F^{-1}_{\mu,\sigma}\big(F_{\mu_0,\sigma_0}(y_0)\big)\Big] &= f_{\mu,\sigma}\Big[F^{-1}_{\mu,\sigma}\big(F_{0,1}(\frac{y_0-\mu_0}{\sigma_0})\big)\Big] && \text{by using (1)}\\
&= f_{\mu,\sigma}\Big[\sigma.F^{-1}_{0,1}(F_{0,1}(\frac{y_0-\mu_0}{\sigma_0}))+\mu\Big] && \text{by using (3)}\\
&= f_{\mu,\sigma}\Big[\sigma.\frac{y_0-\mu_0}{\sigma_0}+\mu\Big] && F^{-1}F(x)=x\\
&= \frac{1}{\sigma}f_{0,1}\Big[\frac{\sigma.\frac{y_0-\mu_0}{\sigma_0}+\mu-\mu}{\sigma}\Big] && \text{by using (2)}\\
&= \frac{1}{\sigma}f_{0,1}\Big[\frac{y_0-\mu_0}{\sigma_0}\Big]\\
&= \frac{\sigma_0}{\sigma}.\frac{1}{\sigma_0}f_{0,1}\Big[\frac{y_0-\mu_0}{\sigma_0}\Big] && \text{Mul and Div by } \sigma_0\\
&= \frac{\sigma_0}{\sigma}f_{\mu_0,\sigma_0}(y_0) && \text{by using (2)}
\end{aligned}
$$

□

Claim 3

$$\mathbb{E}[Y_{iso}] = \mu - \sigma\underbrace{\sum_{i=0}^{n}\frac{1}{n}\frac{\sigma_{(i+1)}p_{(i+1)}-\sigma_{(i)}p_{(i)}}{c_{(i+1)}-c_{(i)}}}_{\delta}$$

Proof: We first re-substitute $c_{(i)} = F_{\mu_{(i)},\sigma_{(i)}}(y_{(i)})$ and $p_{(i)} = f_{\mu_{(i)},\sigma_{(i)}}(y_{(i)})$, then using the above claim, substituting that $f(F^{-1}(c_{(0)})) = f(F^{-1}(0)) = \lim_{x\to-\infty} f(x) = 0$

$$
\begin{aligned}
&\mu - \frac{\sigma^2}{n}\sum_{i=0}^{n}\frac{f(F^{-1}(c_{(i+1)})) - f(F^{-1}(c_{(i)}))}{(c_{(i+1)}-c_{(i)})}\\
&= \mu - \frac{\sigma^2}{n}\sum_{i=0}^{n}\frac{f(F^{-1}(F_{\mu_{(i+1)},\sigma_{(i+1)}}(y_{(i+1)}))) - f(F^{-1}(F_{\mu_{(i+1)},\sigma_{(i+1)}}(y_{(i+1)})))}{(c_{(i+1)}-c_{(i)})}\\
&= \mu - \frac{\sigma}{n}\Big[\frac{\sigma_{(1)}p_{(1)}}{c_{(1)}} + \sum_{i=1}^{n-1}\frac{\sigma_{(i+1)}p_{(i+1)}-\sigma_{(i)}p_{(i)}}{c_{(i+1)}-c_{(i)}}\Big] && \text{by using } f(F^{-1}(c_{(0)}))=0, c_{(0)}=0\\
&= \mu - \frac{\sigma}{n}\Big[\sum_{i=0}^{n-1}\frac{\sigma_{(i+1)}p_{(i+1)}-\sigma_{(i)}p_{(i)}}{c_{(i+1)}-c_{(i)}}\Big] && \sigma_{(0)}=0, p_{(0)}=0, c_{(0)}=0
\end{aligned}
$$

□

A.3 Proofs of claim 4 and claim 5

Claim 4 *Let* M *be any regression model. Then* M *is perfectly quantile calibrated* iff

$$\text{KL}\Big(\,[\mathsf{M}(X)](Y)||\text{U}\Big) = 0$$

Proof: Since we have that $KL\Big(\,[\mathsf{M}(X)](Y)||U\Big)$ we have that $\mathsf{M}(X)](Y) \triangleq$ Uniform$[0,1]$ from which we have that $\mathbb{P}\big[[\mathsf{M}(X)](Y) \le p\big] = p\ \forall p$ which is the definition of quantile regularization.

□

Claim 5 *Let* $f_{\mu,\sigma}$ *be the marginal distribution of* Y *and* $\Big(F_{\mu_x,\sigma_x} = M[x]\Big)\Big|X = x$ *be the model's predicted cumulative distribution for* $x \in \mathcal{X}$*, then if* $f_{\mu,\sigma}[F^{-1}_{\mu_x,\sigma_x}(p)] = f_{\mu_x,\sigma_x}(F^{-1}_{\mu_x,\sigma_x}(p))\ \forall x, \forall p \in [0,1]$*, we have that* M *is Quantile Calibrated.*

Proof:

$$\begin{aligned}
f_{M[X][Y]}(p) &= \int_{\mathcal{X}} f_{M[x][Y]\Big|X=x}(p)\,.\,f_X(x)dx \\
&= \int_{\mathcal{X}} \frac{d}{dp}\mathbf{P}\Big[\Big(M[x][Y]\Big|X=x\Big) \leq p\Big]\,.\,f_X(x)dx \\
&= \int_{\mathcal{X}} \frac{d}{dp}\mathbf{P}\Big[(F_{\mu_x,\sigma_x}[Y] \leq p\Big]\,.\,f_X(x)dx \\
&= \int_{\mathcal{X}} \frac{d}{dp}\mathbf{P}\Big[(Y \leq F^{-1}_{\mu_x,\sigma_x}(p)\Big]\,.\,f_X(x)dx \\
&= \int_{\mathcal{X}} \frac{d}{dp}\Big[F_{\mu,\sigma}[F^{-1}_{\mu_x,\sigma_x}(p)]\Big]\,.\,f_X(x)dx \\
&= \int_{\mathcal{X}} f_{\mu,\sigma}[F^{-1}_{\mu_x,\sigma_x}(p)]\,.\,\frac{1}{f_{\mu_x,\sigma_x}(F^{-1}_{\mu_x,\sigma_x}(p))}\,.\,f_X(x)dx \\
&= \int_{\mathcal{X}} 1\,.\,f_X(x)dx \\
&= 1
\end{aligned}$$

Since we have that $M[X][Y]$Uniform$[0,1]$ we can conclude that M is Perfectly Quantile Calibrated. □

B EXPERIMENTS

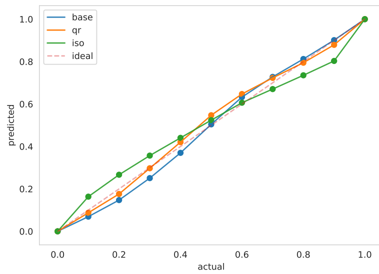
B.1 BAYESIAN LINEAR REGRESSION

For Bayesian Linear Regression we use sklearn's Bayesian Ridge Regression.

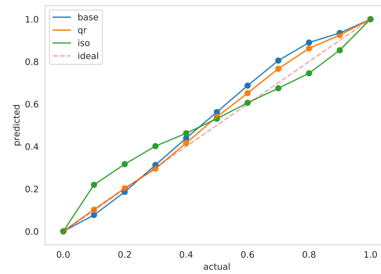
B.2 UCI EXPERIMENTS

We consider following datasets (size-of-data,num-input-features): AirFoil (1503,6) , Bouston Housing (506,13), Concrete Strength (1030,8),Fish Toxicity (908,7),Kin8nm (8192, 9), Protein Structure (45730, 10), Red Wine (1599, 12), White Wine (4898, 12), Yacht Hydrodynamics (308,6), year prediction MSD (515345,91)

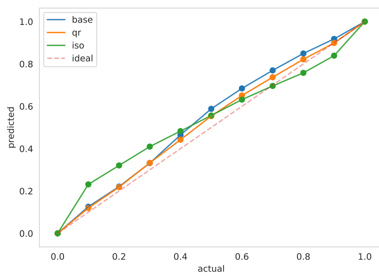
B.3 Calibration plots for Dropout-VI



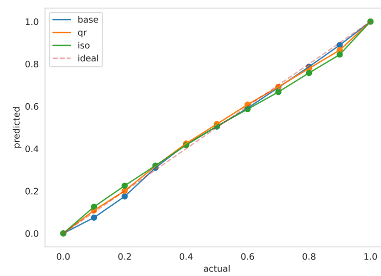
(a) Airfoil



(b) Boston



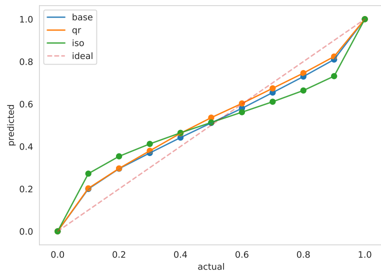
(c) Concrete



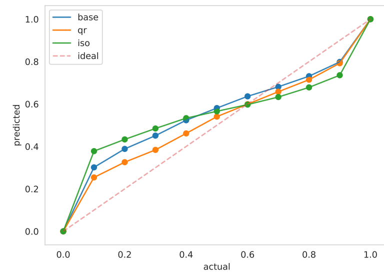
(d) White Wine

Figure 1: Dashed line (y=x) indicates perfect Calibration. The more closer to Dashed line the better

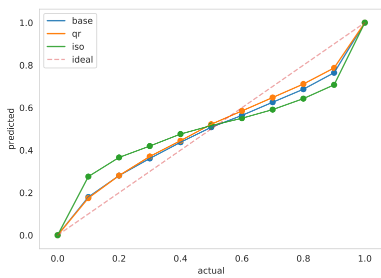
B.4 Calibration plots for Deep Ensembles



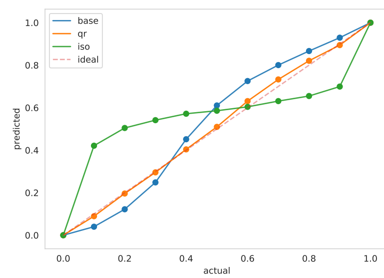
(a) Airfoil



(b) Boston



(c) Red



(d) Yacht Hydrodynamics

Figure 2: Dashed line (y=x) indicates perfect Calibration. The more closer to Dashed line the better