On uncertainty calibration for equivariant functions

Edward Berman, Jacob Ginesin, and Robin Walters

Northeastern University

Email: berman.ed@northeastern.edu

Abstract—**Uncertainty estimation is often important in data-sparse settings where equivariant models tend to excel, including pick-and-place robotics tasks, galaxy morphology classification, and chemical physics. We study the relationship between equivariance, model calibration, and model confidence. We define a generalization of calibration error for regression tasks and prove an upper bound for $G-$invariant and $G-$equivariant models. This work marks a first step toward an understanding for how equivariance relates to uncertainty estimation.**

I. INTRODUCTION

A surprising result of Wang et al. [16] is that equivariant models can still be effective even in cases of mismatch between the model and the data symmetry. This finding motivated the work of [17], which explored how equivariance can affect model *accuracy*, both positively and negatively. However, it is not yet understood how equivariance impacts model *calibration*, loosely defined as the disagreement between a model's accuracy and predicted *confidence*. Understanding both model calibration and confidence is particularly useful in data-sparse settings where equivariant neural networks tend to thrive, such as pick-and-place robotics tasks [10, 15, 14, 4, 7, 9, 8] and molecular physics [19, 13]. While equivariance has proved invaluable in these scenarios, equivariance is not a golden bullet: equivariance comes with well-defined tradeoffs, including fairly limited benefits at scale [18, 11, 6, 2, 1], provable model performance degradation in cases of symmetry mismatch [17], more complex architectures, and the need for more compute. To help quantify these tradeoffs, we seek to study how the inductive bias of equivariance affects different metrics such as calibration error. This in turn allows us to address several unanswered questions on the subject of calibration and confidence of equivariant models. Namely, when does equivariance help a model predict its own confidence? How do different types of symmetry such as correct, incorrect, and extrinsic equivariance [17] affect model calibration?

The purpose of this work is to extend the error bounds given by [17] to a broader class of calibration losses. In this way, we can quantify the effect of equivariance not just on accuracy, but also on calibration.

II. THEORETICAL RESULTS

A. Problem Statements and Notation

In this section, we lay out the problem statements and preliminary notation necessary for our main results. A function is *equivariant* if it commutes with the group action, $gf(x) = f(gx)$. A special case is *invariance*, where the representation of g on the output is trivial, so $f(x) = f(gx)$.

Throughout the problem statements, we introduce additional notation to make certain things explicit. We denote sets, vector spaces, or groups in boldface, such as $\mathbf{E}$. However, we will drop these in the proofs as they become clear. Where we use gx, we are really considering the representation $\rho(g)$, an $n \times n$ matrix, for $g \in \mathbf{G}$. This means that gx, the result of the group action acting on x, is a vector.

a) Invariant Regression Problem Setup: Consider a function $f : \mathbf{X} \to \mathbf{Y}$ where $\mathbf{Y} = \mathbb{R}^n$. Define a model class as a set $\{h : \mathbf{X} \to \mathcal{M} \times \mathcal{S}\}$ where $\mathcal{M} = \mathbb{R}^n$ represents the space of all mean predictions and $\mathcal{S} = \mathbb{R}^n_+$ represents the space of all variance-vectors. One may intuit these vectors as representing diagonal covariance matrices. $\mathcal{M}$ will have elements m and $\mathcal{S}$ will have elements s. Denote the two outputs by h_μ and h_{σ^2}. Let $p : \mathbf{X} \to \mathbb{R}$ be the probability density over the domain $\mathbf{X}$. Denote the subdomain of $\mathbf{X}$ given by the constraint $h_{\sigma^2}(x) = s$, $\mathbf{X_s} = \{x \in \mathbf{X}|h_{\sigma^2}(x) = s\}$. It will be implicit hereafter that $\mathbf{X_s}$ is defined with respect to a specific s. We denote the elements of $\mathbf{X_s}$ as x_s. We will denote the set of orbit representatives of $\mathbf{G}$ in $\mathbf{X_s}$ as $\mathbf{F_s}$. This is the same as the fundamental domain in [17]. The group $\mathbf{G}$ acts on $\mathbf{X}$ and $\mathbf{Y}$. We assume that each set of orbit representatives $\mathbf{F_s}$ satisfies the condition that the union of all pairwise intersections $\cup_{g_1 \neq g_2}(\mathbf{g_1 F_s} \cap \mathbf{g_2 F_s})$ have measure 0. We also assume that $\mathbf{F_s}$ and $\mathbf{G}x_s$ are differentiable manifolds for any $\mathbf{X_s}$ and for all $x_s \in \mathbf{X_s}$.

Next, we define a family of probability densities for each of the fibers of h_{σ^2}. Specifically, we define a density over $\mathbf{X_s}$ by $q : \mathbf{X_s} \to \mathbb{R}$ via $q(x_s) = \frac{p(x_s)}{\int_{X_s} p(x)dx}$. For $x \notin \mathbf{X_s}$, $q(x) = 0$. This allows us to define the domain restricted regression error

$$\text{err}_{\text{reg}}(h, s) = \int_{\mathbf{X_s}} q(x_s)\Big\|h_\mu(x_s) - f(x_s)\Big\|_2^2 dx_s. \quad (1)$$

Denote by $q(\mathbf{Gx_s}) = \int_{z \in \mathbf{Gx_s}} q(z)dz$ the probability of the orbit $\mathbf{Gx_s}$ on $\mathbf{X_s}$. Denote by $q_{\text{norm}}(z) = \frac{q(z)}{q(\mathbf{Gx_s})}$ the normalized probability density on the orbit $\mathbf{Gx_s}$ such that $\int_{\mathbf{Gx_s}} q_{\text{norm}}(z)dz = 1$. Let $\mathbb{E}_{\mathbf{Gx_s}}[f]$ be the mean of the function f on the orbit $\mathbf{Gx_s}$ and let $\mathbb{V}_{\mathbf{Gx_s}}[f]$ be the variance of f on the orbit $\mathbf{Gx_s}$,

$$\mathbb{E}_{\mathbf{Gx_s}}[f] = \int_{\mathbf{Gx_s}} q_{\text{norm}}(z)f(z)dz = \frac{\int_{\mathbf{Gx_s}} q(z)f(z)dz}{\int_{\mathbf{Gx_s}} q(z)dz} \quad (2)$$

$$\mathbb{V}_{\mathbf{Gx_s}}[f] = \int_{\mathbf{Gx_s}} q_{\text{norm}}(x)\Big\|\mathbb{E}_{\mathbf{Gx_s}}[f] - f(z)\Big\|_2^2 dz. \quad (3)$$

These definitions arise from the error lower bound on invariant regression in [17]. Intuitively, if a model h is constant on an

orbit where f is varying, then the constant that minimizes the regression error is the average of f on the orbit, and the induced error is the variance of f.

Finally, we define a generalization of the expected normalized calibration error (ENCE), as given by Eqn. 8 in [12]. The main drawbacks of their ENCE metric are two-fold: it is defined in terms of binning approximations and assumes that $h_\mu(x)$ and $h_{\sigma^2}(x)$ are scalar values. Our formalism not only works for vectors and avoids binning, allowing for a discussion of continuous group symmetries. While binning approximations are still necessary to compute ENCE in practice, our theory supports the more generalized continuous case. We take the absolute value function applied to a vector $|\cdot|$ to be applied element-wise. Similarly, if we use $\sqrt{c \cdot s}$ as a vector, that means the square root function was applied element-wise to the vector s after and scaled by a constant c. We will also note that vectors have a partial ordering where $\vec{a} \le \vec{b}$ if $a_i \le b_i$ for all i. Let $\mathbf{D}$ be the region of vectors d bounded by $s_1 \le d \le s_2$. We define a probability density $r : \mathcal{S} \to \mathbb{R}$ such that $\mathbb{P}(h_{\sigma^2}(x) \in \mathbf{D}) = \mathbb{P}(s \in \mathbf{D}) = \int_{\mathbf{D}} r(s) ds$. This is the push-forward of the density of p over h_{σ^2}, equivalently, $r = h_{\sigma^2} \# (p)$.

We now have the necessary machinery to define our generalized ENCE metric. The goal for our learning task is that h_μ fits the function f and h_σ^2 properly predicts confidence by minimizing our generalized ENCE metric (Equation 5):

$$\text{ENCE} = \int_{\mathcal{S}} r(s) \tag{4}$$

$$\cdot \frac{\mathbb{E}_{X_s}\left[\left\|\sqrt{\frac{2}{\pi}}s - |h_\mu(x) - f(x)|\right\|_2^2 \;\middle|\; h_{\sigma^2}(x) = s\right]}{\left\|\sqrt{\frac{2}{\pi}}s\right\|_2^2} ds. \tag{5}$$

We assume that the model class $\{h\}$ is arbitrarily expressive except that it is constrained to be invariant with respect to a group $\mathbf{G}$. That is to say, it is a universal approximator for compactly supported G-invariant functions.

Remark 1: If errors are normally distributed and a model is well calibrated, then $\sqrt{\frac{2}{\pi}}\sigma = \overline{|\mu - f(x)|}$ is the same as $\sigma^2 = \overline{(\mu - f(x))^2}$, where the factor of $\sqrt{\frac{2}{\pi}}$ comes from [5]. We choose the former as a component in our metric because it allows us to later apply a theorem from [17].

b) Equivariant Regression Problem Setup: Our setup is the same as for invariant regression with a small addition. Let $a_x : G/G_x \to Gx$ be a map from coset space to an orbit. $\alpha(g, x) = \left(\int G_x dh\right)^{-1} \left|\frac{\partial a_x(\bar{g})}{\partial \bar{g}}\right|$ will act as a jacobian term that allows one to lift integrals over an orbit Gx to the group G, see Appendix A in [17] for more details. We define matrices $Q_{\mathbf{Gx}} \in \mathbb{R}^{n \times n}$, $Q^\dagger(gx) \in \mathbb{R}^{n \times n}$ such that $\int_{\mathbf{G}} Q^\dagger(gx)dg = \text{Id}$ via

$$Q_{\mathbf{Gx}} = \int_{\mathbf{G}} q(gx)\rho_Y(g)^T \rho_Y(g)\alpha(x, g)dg \tag{6}$$

$$Q^\dagger(gx) = Q_{\mathbf{Gx}}^{-1} q(gx)\rho_Y(g)^T \rho_Y(g)\alpha(x, g). \tag{7}$$

We also define $\mathcal{E}_{\mathbf{G}}[f, x]$ by

$$\mathcal{E}_{\mathbf{G}}[f, x] = \int_G Q^\dagger(gx)g^{-1}f(gx)dg. \tag{8}$$

These definitions come from the equivariant regression error lower bound in [17]. The intuition is that $\mathcal{E}$ is the minimizer obtained the average of all inversely transformed f on each orbit Gx.

B. Invariant Regression Upper Bounds

Theorem 1: If a model h is G−invariant, then ENCE given by Equation 5 is bounded below by 0 and bounded above by $1 + \mathbb{E}_{\mathcal{S}}\left[\frac{\text{err}_{\text{reg}}(h,s)}{\|\sqrt{\frac{2}{\pi}}s\|_2^2}\right]$. If $\text{err}_{\text{reg}}(h, s)$ is minimized, then the ENCE upper bound becomes $1 + \mathbb{E}_{\mathcal{S}}\left[\frac{\int_{F_s} q(Gx_s)\mathbb{V}_{Gx_s}[f]dx_s}{\|\sqrt{\frac{2}{\pi}}s\|_2^2}\right]$.

Proof: We begin by observing that the entries in $\sqrt{\frac{2}{\pi}}s$ and $|h_\mu(x) - f(x)|$ are nonnegative. Therefore, we have that $0 \le \mathbb{E}_{X_s}\left[\left\|\sqrt{\frac{2}{\pi}}s - |h_\mu(x) - f(x)|\right\|_2^2 \;\middle|\; h_{\sigma^2}(x) = s\right] \le \mathbb{E}_{X_s}\left[\left\|\sqrt{\frac{2}{\pi}}s\right\|_2^2 + \left\||h_\mu(x) - f(x)|\right\|_2^2 \;\middle|\; h_{\sigma^2}(x) = s\right]$.

Consequently,

$$0 \le \text{ENCE}$$

$$\le \int_{\mathcal{S}} r(s)$$

$$\cdot \frac{\mathbb{E}_{X_s}\left[\left\|\sqrt{\frac{2}{\pi}}s\right\|_2^2 + \left\||h_\mu(x) - f(x)|\right\|_2^2 \;\middle|\; h_{\sigma^2}(x) = s\right]}{\left\|\sqrt{\frac{2}{\pi}}s\right\|_2^2} ds$$

$$= \int_{\mathcal{S}} r(s)$$

$$\cdot \frac{\left\|\sqrt{\frac{2}{\pi}}s\right\|_2^2 + \mathbb{E}_{X_s}\left[\left\||h_\mu(x) - f(x)|\right\|_2^2 \;\middle|\; h_{\sigma^2}(x) = s\right]}{\left\|\sqrt{\frac{2}{\pi}}s\right\|_2^2} ds$$

$$= \int_{\mathcal{S}} r(s)ds + \int_{\mathcal{S}} r(s)$$

$$\cdot \left[\frac{\int_{X_s} q(x_s)\left\|h_\mu(x) - f(x)\right\|_2^2 dx}{\left\|\sqrt{\frac{2}{\pi}}s\right\|_2^2}\right] ds.$$

By the definition of domain restricted regression error in our problem setup, we conclude

$$\int_{X_s} q(x_s)\left\|h_\mu(x_s) - f(x_s)\right\|_2^2 dx_s = \text{err}_{\text{reg}}(h, s).$$

This gives us

$$\text{ENCE} \le \int_{\mathcal{S}} r(s)\left[1 + \frac{\text{err}_{\text{reg}}(h, s)}{\|\sqrt{\frac{2}{\pi}}s\|_2^2}\right] ds \tag{9}$$

$$= 1 + \mathbb{E}_{\mathcal{S}}\left[\frac{\text{err}_{\text{reg}}(h, s)}{\|\sqrt{\frac{2}{\pi}}s\|_2^2}\right]. \tag{10}$$

Now, if the domain restricted regression error is minimized, then by Theorem 4.8 in [17] we have that

$$\text{err}_{\text{reg}}(h, s) = \int_{F_s} q(Gx_s)\mathbb{V}_{Gx_s}[f]dx_s \quad (11)$$

which completes the proof. ∎

C. Invariant Regression Upper Bound Example

Example 1: In this example, we consider a set X of five chemical compounds and a codomain Y of spectra. We will calculate the ENCE upper bound assuming a minimized regression error. Each molecule $x \in X$ has information related to position and atomic number. We assume that the model class h has $E(2)$-invariance to the atomic positions, where $E(2)$ is the Euclidean group for $\mathbb{R}^2$. The atomic positions are shown in Figure 1. Note that for illustration purposes in this example, we are not giving the exact physically accurate atomic positions as you would find in [13] or a similar dataset. The dataset contains some duplicates of the chemical compounds up to transformations in $E(2)$. That is, the atomic numbers are the same but the atomic positions may be rotated and translated about the origin. Different molecule orientations are notated with $+$ and $\times$.

The input space X contains molecules $\{CH_4(+), CH_4(\times), H_2O, SO_2, NH_3\}$. The map $f : X \to Y$ produces four unique spectra, one for each of the unique molecules. The model class h predicts three unique spectra, as it produces the same spectra for both H_2O and SO_2 due to their equivalence up to $E(2)$ in our simplified example. The model h produces two distinct variance vectors – it predicts variance s_1 for the first two orbits and variance s_2 on the last orbit. The probability of obtaining each of the molecules is $\{0.125, 0.125, 0.125, 0.125, 0.5\}$ for $\{CH_4(+), CH_4(\times), H_2O, SO_2, NH_3\}$ respectively.

Along each row, Figure 1 shows an orbit of the representative set of X in the first two columns, the spectral lines from f and h_μ in the third column, and the associated variance s from $h_{\sigma^2}(x)$ in the last column. Each molecule is titled with the name of the compound and the probability of sampling it in X. Consider the first row. Since h is $E(2)$ invariant, h is able to fully fit the function f on the orbit containing two rotated version of CH_4. Thus, the regression error is zero. Along the second row, h produces the same spectral lines for H_2O and SO_2 since they are the same up to $E(2)$ in our example. Since H_2O and SO_2 are equally probable, the output of h that minimizes the regression error is just the average of the two spectral lines.

We now refer to Eqn. 11 to compute the domain restricted regression error for s_1. The first orbit gave us a regression error of zero. As such, we must have $\mathbb{V}_{Gx_s}[f] = 0$ on this orbit. This is certainly true, as f is invariant to the atomic numbers, which are the same for the various rotations of methane. For

the second orbit, we compute

$$q(Gx_s) = \frac{p(H_2O) + p(SO_2)}{p(CH_4(\times)) + p(CH_4(+)) + p(H_2O) + p(SO_2)}$$
$$= \frac{0.125 + 0.125}{0.125 + 0.125 + 0.125 + 0.125} = 0.5.$$

In other words, when restricted to the domain that outputs a variance of s_1, we have a 50% chance of being on the orbit containing H_2O and SO_2. The mean of the function f on the orbit Gx_s is just the average of the two spectral lines since the probabilities of H_2O and SO_2 are equal. The variance can be computed as the average distance of the real spectral lines from the prediction h, which is itself just the average of the two lines. Obtaining exact values for the spectra from [13] and [3], we can calculate a value of approximately 38.5 for the variance of f on the second orbit. Thus, we can compute $\text{err}_{\text{reg}}(h, s_1) = (0.5 \times 0) + (0.5 \times 38.5) = 19.25$.

As with the first orbit, the minimizing regression error on the third orbit is zero. Since there is only one element in X we can fit, the mapping between the molecule and the spectra is unique and can be made to fit the ground truth. So, $\text{err}_{\text{reg}}(h, s_2) = 0$.

Notice that the probability of obtaining s_1 is the same as the probability of obtaining s_2. Putting it all together, we get

$$\text{ENCE} \leq 1 + \left(0.5 \cdot \frac{0}{||\sqrt{\frac{2}{\pi}}s_2||_2^2}\right) + \left(0.5 \cdot \frac{19.5}{||\sqrt{\frac{2}{\pi}}s_1||_2^2}\right)$$
$$= 1 + \frac{9.625}{||\sqrt{\frac{2}{\pi}}s_1||_2^2}.$$

We conclude that the upper bound for ENCE in our example is $1 + \frac{9.625}{||\sqrt{\frac{2}{\pi}}s_1||_2^2}$. This is plotted as a function of $||\sqrt{\frac{2}{\pi}}s_1||_2^2$ in Figure 2. We can see that in the limit as $||\sqrt{\frac{2}{\pi}}s_1||_2^2$ goes to infinity the upper bound on ENCE becomes 1. Alternatively, in the limit as $||\sqrt{\frac{2}{\pi}}s_1||_2^2$ goes to 0, the upper bound on ENCE diverges. The interpretation of the latter is that if the model is extremely confident, then any deviations from the mean prediction represents similarly extreme miscalibration.

D. Equivariant Regression Upper Bounds

Theorem 2: We assume the model class $\{h\}$ is equivariant. ENCE (Eqn. 5) is bounded below by 0 and bounded above by $1 + \mathbb{E}_{S}\left[\frac{\text{err}_{\text{reg}}(h,s)}{||\sqrt{\frac{2}{\pi}}s||_2^2}\right]$. If $\text{err}_{\text{reg}}(h, s)$ is minimized then the upper bound becomes $1 + \mathbb{E}_{S}\left[\frac{\int_F \int_G q(gx)||f(gx) - g\mathcal{E}_G[f,x]||_2^2 \alpha(x,g)dgdx}{||\sqrt{\frac{2}{\pi}}s||_2^2}\right]$.

Proof: Unlike before, we can not decompose an integral over X_s into an iterated integral over F_s and Gx_s. This is because without the assumption of invariance, the restriction $h_{\sigma^2}(x) = s$ no longer preserves entire orbits. Therefore, we instead note that we can rewrite the domain restricted regression error in terms of the original domain X. Consider the integral in Eqn. 1. Since $q(x) = 0$ for $h_{\sigma^2}(x) \neq s$, we
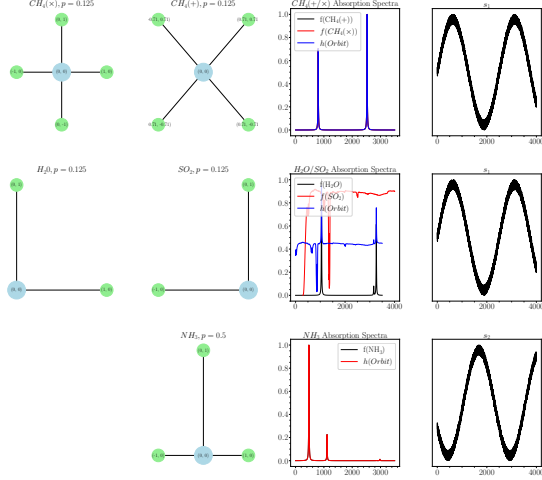
Fig. 1. An example on how the ENCE upper bound behaves for an $E(2)$ invariant model class h on a set of molecules producing absorption spectra. Each row contains an orbit, the absorption spectra specified by f and the error minimizing h, and the predicted variance vector.
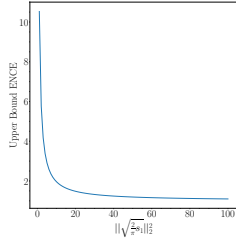

Fig. 2. Upper bound on ENCE as a function of $||\sqrt{\frac{2}{\pi}s_1}||_2^2$ in Example 1.

may write

$$\text{err}_{\text{reg}}(h,s) = \int_X q(x)\left\|h_\mu(x) - f(x)\right\|_2^2 dx. \qquad (12)$$

Said differently, the integrand takes on a value of 0 outside of X_s, so without loss of generality, we can replace X_s with X and x_s with x everywhere. Using the same argument as we did with invariant regression, we arrive at

$$\text{ENCE} \leq \int_{\mathcal{S}} r(s)\left[1 + \frac{\text{err}_{\text{reg}}(h,s)}{\|\sqrt{\frac{2}{\pi}s}\|_2^2}\right] ds$$

$$= 1 + \mathbb{E}_{\mathcal{S}}\left[\frac{\text{err}_{\text{reg}}(h,s)}{\|\sqrt{\frac{2}{\pi}s}\|_2^2}\right].$$

Applying Theorem 4.9 in [17], if $\text{err}_{\text{reg}}(h,s)$ is a minimizer then we have

$$\text{ENCE} \leq 1+ \qquad (13)$$

$$\mathbb{E}_{\mathcal{S}}\left[\frac{\int_F \int_G q(gx)\|f(gx) - g\mathcal{E}_G[f,x]\|_2^2 \alpha(x,g)dgdx}{\|\sqrt{\frac{2}{\pi}s}\|_2^2}\right]. \qquad (14)$$

This completes the proof. ∎

Remark 2: The upper bounds presented here and in Section II-B are analogous to the maximum χ^2 score one can attain averaged over all of the variances the model predicts.

III. EXPERIMENTAL RESULTS

The goals of this section are to computationally verify our theorem holds on real data, to see how tight our proven upper bound is, and to explore the effect of binning approximations on estimating ENCE. Estimating the true ENCE in practice is difficult because the continuous density $r(\sigma^2)$ must be computed using discrete binning approximations. In this experiment, we show that the theoretical upper bound is satisfied on augmented versions of the QM9 dataset. As with the example, our input is a molecule and we are predicting its absorption spectra. The QM9 dataset is canonicalized such that molecules are unique up to rotation. Therefore, we include 90, 180, and 270 degree rotations of each molecule in the test set. For the true label corresponding to the rotated molecules in the test set, we consider two variations, each corresponding to rows (1) and (2) in Figure 1. *1. Correct Augmentation.* Here, the labels for the rotated molecules are the same as the original. In this circumstance, our $E(3)-$invariance is correct, and we expect our ENCE to be well within the bound. Additionally, we expect that this data augmentation matches the underlying physics of our world. *2. Incorrect Augmentation.* Here, the spectral line has an explicit dependence on the rotation angle of the molecule. Specifically, each 90 rotation corresponds to a reflection over the $y-$axis. This augmentation is not true to real physics, and is done to make our data have incorrect $E(3)-$invariance. In this way, we can illustrate how having incorrect equivariance takes us closer to our theoretical upper bound. Results are presented in Figure 3.
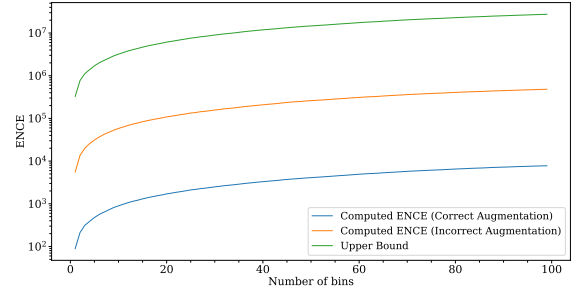

Fig. 3. The Computed ENCE versus the theoretical upper bound. We see that regardless of the number of bins used to estimate the probability density $r(s)$, our computed ENCE is strictly less than the theoretical upper bound, and moreover, when we augment our data set so that we have incorrect equivariance, we get closer to this upper bound.

IV. CONCLUSIONS

In this work, we extended the bounds from [17] to calibration losses, showed an example that demonstrated how the bounds could be realized, and then executed the simplified example on a real dataset. This work marks the first unified theory for relating equivariance to uncertainty.

References

[1] Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, pages 1–3, 2024.

[2] Johann Brehmer, Sönke Behrends, Pim de Haan, and Taco Cohen. Does equivariance matter at scale? *arXiv preprint arXiv:2410.23179*, 2024.

[3] Coblentz Society. Sulfur dioxide (so_2) infrared spectrum. https://webbook.nist.gov/cgi/cbook.cgi?ID=C7446095&Index=0&Type=IR-SPEC, 1977. URL https://webbook.nist.gov/cgi/cbook.cgi?ID=C7446095&Index=0&Type=IR-SPEC. Spectrum recorded by Dow Chemical Company, digitized by NIST.

[4] Jiahui Fu, Yilun Du, Kurran Singh, Joshua B Tenenbaum, and John J Leonard. Neuse: Neural se (3)-equivariant embedding for consistent spatial understanding with objects. *arXiv preprint arXiv:2303.07308*, 2023.

[5] Roy C Geary. The ratio of the mean deviation to the standard deviation as a test of normality. *Biometrika*, 27 (3/4):310–332, 1935.

[6] Nate Gruver, Marc Anton Finzi, Micah Goldblum, and Andrew Gordon Wilson. The lie derivative for measuring learned equivariance. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=JL7Va5Vy15J.

[7] Haojie Huang, Dian Wang, Xupeng Zhu, Robin Walters, and Robert Platt. Edge grasp network: A graph-based se (3)-invariant approach to grasp detection. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3882–3888. IEEE, 2023.

[8] Haojie Huang, Owen Howell, Dian Wang, Xupeng Zhu, Robin Walters, and Robert Platt. Fourier transporter: Bi-equivariant robotic manipulation in 3d. *arXiv preprint arXiv:2401.12046*, 2024.

[9] Haojie Huang, Dian Wang, Arsh Tangri, Robin Walters, and Robert Platt. Leveraging symmetries in pick and place. *The International Journal of Robotics Research*, 43(4):550–571, 2024.

[10] Dmitry Kalashnikov, Alex Irpan, Peter Pastor, Julian Ibarz, Alexander Herzog, Eric Jang, Deirdre Quillen, Ethan Holly, Mrinal Kalakrishnan, Vincent Vanhoucke, et al. Qt-opt: scalable deep reinforcement learning for vision-based robotic manipulation. corr abs/1806.10293 (2018). *arXiv preprint arXiv:1806.10293*, 2018.

[11] David Klee, Jung Yeon Park, Robert Platt, and Robin Walters. A comparison of equivariant vision models with imagenet pre-training. In *NeurIPS 2023 Workshop on Symmetry and Geometry in Neural Representations*, 2023.

[12] Dan Levi, Liran Gispan, Niv Giladi, and Ethan Fetaya. Evaluating and calibrating uncertainty prediction in regression tasks. *Sensors*, 22(15):5540, 2022.

[13] Raghunathan Ramakrishnan, Pavlo O Dral, Matthias Rupp, and O Anatole Von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific data*, 1(1):1–7, 2014.

[14] Dian Wang, Mingxi Jia, Xupeng Zhu, Robin Walters, and Robert Platt. On-robot learning with equivariant models. *arXiv preprint arXiv:2203.04923*, 2022.

[15] Dian Wang, Robin Walters, Xupeng Zhu, and Robert Platt. Equivariant q learning in spatial action spaces. In *Conference on Robot Learning*, pages 1713–1723. PMLR, 2022.

[16] Dian Wang, Jung Yeon Park, Neel Sortur, Lawson L.S. Wong, Robin Walters, and Robert Platt. The surprising effectiveness of equivariant models in domains with latent symmetry. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=P4MUGRM4Acu.

[17] Dian Wang, Xupeng Zhu, Jung Yeon Park, Mingxi Jia, Guanang Su, Robert Platt, and Robin Walters. A general theory of correct, incorrect, and extrinsic equivariance. *Advances in Neural Information Processing Systems*, 36, 2024.

[18] Yuyang Wang, Ahmed AA Elhag, Navdeep Jaitly, Joshua M Susskind, and Miguel Ángel Bautista. Swallowing the bitter pill: Simplified scalable conformer generation. In *Forty-first International Conference on Machine Learning*, 2023.

[19] Zihan Zou, Yujin Zhang, Lijun Liang, Mingzhi Wei, Jiancai Leng, Jun Jiang, Yi Luo, and Wei Hu. A deep learning model for predicting selected organic molecular spectra. *Nature Computational Science*, 3(11):957–964, 2023.