Loss function - Illustration

June 2024

To better understand the proposed losses, they have been applied to distributions with different shapes. For comparison purposes, the quadratic loss, used in standard prediction methods that give point predictions, was calculated:

$$L(Z, p_j) = (Z - \mathbb{E}(\widehat{p}))^2,$$

where $\mathbb{E}(\widehat{p})$ is the mean of $\widehat{p}$.

Figure 1 presents a comparison of the probability density functions for different models. In these examples, each model has a Beta-binomial distribution given by $c + BetaBin(100, a, b)$. The random variable is not important here, as the interest lies in evaluating how well each proposed model's distribution fits the observed value. The parameter c represents the distance of the distribution's mean from the observed value, while the parameters a and b control the shape of the distribution. If a and b are different, the distribution is asymmetric, and these parameters determine whether most of the distribution is located to the right or left of its mean.

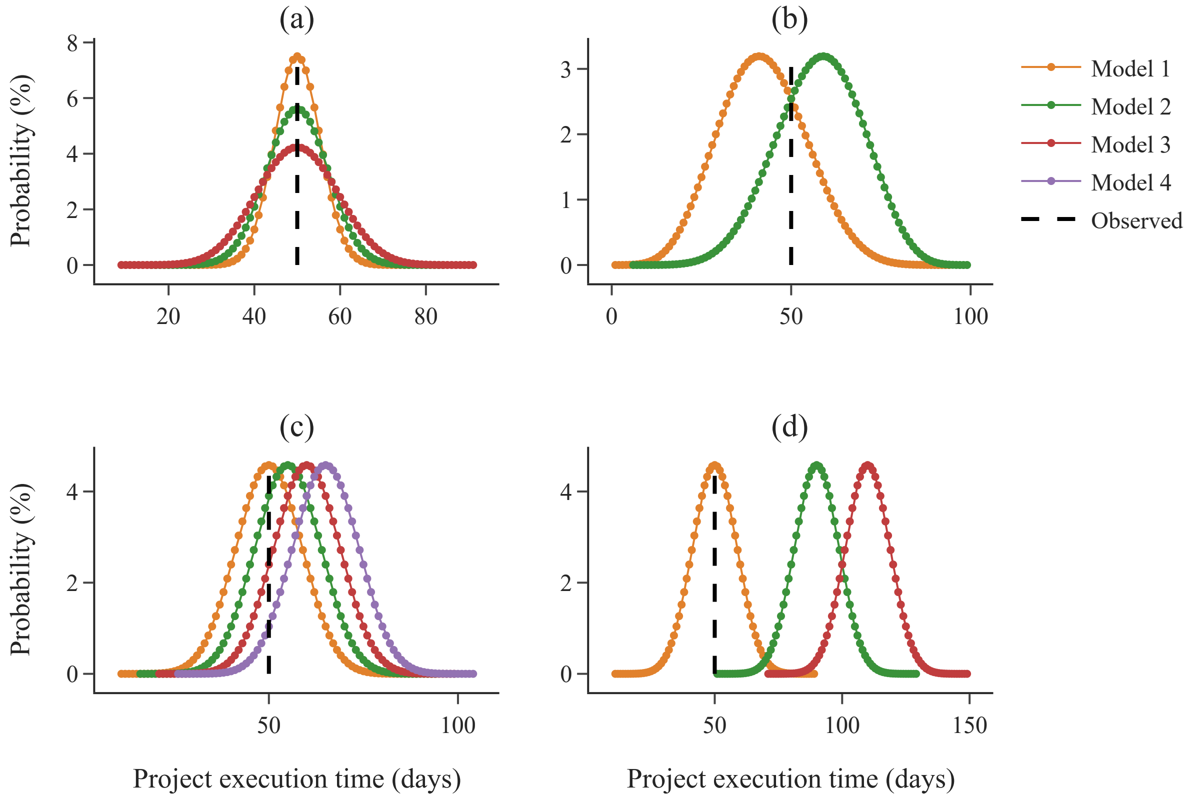

Figure 1: Comparison of probability mass function of different models for (a) distributions with the same mean and different variances, (b) distributions symmetric around the observed value, (c) distributions with the same variance and means farther from the observed value, and (d) distributions with zero probability for the observed value.

Table 1 shows the loss measures for the selected distributions. In Figure 1-(a), the distributions have the same mean but different variances. The quadratic loss is zero for all distributions because the mean equals the observed value. However, other loss measures indicate that the distributions of models 1, 2, and 3 perform progressively worse as the variance increases. This demonstrates that metrics summarizing the entire distribution into a single value, such as the mean, are inadequate for capturing the entire uncertainty of the distribution.

In Figure 1-(b), the distributions are symmetric concerning the observed value and are concentrated on different sides. The quadratic loss, density, unweighted pinball, and tail-weighted pinball give equal results because these measures are invariant to rotation around the observed value and do not consider whether the density is concentrated below or above the observed value. The distribution of model 1 has lower quantiles farther from the observed value, resulting in a higher left-tail-weighted pinball loss. Conversely, the distribution of model 2 has upper quantiles farther from the observed value, resulting in a higher right-tail-weighted pinball loss. Thus, the choice of weights in the pinball loss allows for different weighting of the distribution's quantiles.

In Figure 1-(c), the distributions have equal variances, but the means move away from the observed value. As expected, the more distant distributions perform worse for all metrics because their densities are further from the observed value.

Finally, in Figure 1-(d), the observed value lies outside the distributions of models 2 and 3. Although the distribution of model 3 is farther from the observed value, the density loss is the same for both distributions. This occurs because this loss considers the shape of the distribution and whether the prediction is correct with high probability. If the observed value is not within the density, the metric remains unchanged for translations, as long as the value remains outside the density. Therefore, this loss does not quantify the extent of incorrect predictions, which may be unsuitable in some situations.

Table 1: Quadratic loss, density loss, and pinball loss values for different groups and models. Each group contains models characterized by a specific distribution $c + BetaBin(100, a, b)$ with parameters c, a e b.

		Distributions					Pinball			
Group	Model	c	a	b	MSE	Density	unweighted	two tailed	left tail	rigth tail
3*(a)	1	0	400	400	0	-0.150	0.62	0.05	0.20	0.20
	2	0	50	50	0	-0.112	0.82	0.07	0.27	0.27
	3	0	20	20	0	-0.084	1.09	0.09	0.36	0.36
2*(b)	1	0	8	11	62	-0.051	2.47	0.13	0.99	0.50
	2	0	11	8	62	-0.051	2.47	0.13	0.50	0.99
4*(c)	1	0	25	25	0	-0.091	1.01	0.08	0.34	0.34
	2	5	25	25	25	-0.078	1.57	0.09	0.31	0.65
	3	10	25	25	100	-0.048	3.09	0.14	0.77	1.05
	4	15	25	25	225	-0.021	5.21	0.26	1.67	1.46
3*(d)	1	0	25	25	0	-0.091	1.01	0.08	0.34	0.34
	2	40	25	25	1600	0.000	17.57	1.27	7.79	3.55
	3	60	25	25	3600	0.000	27.57	2.11	12.78	5.21