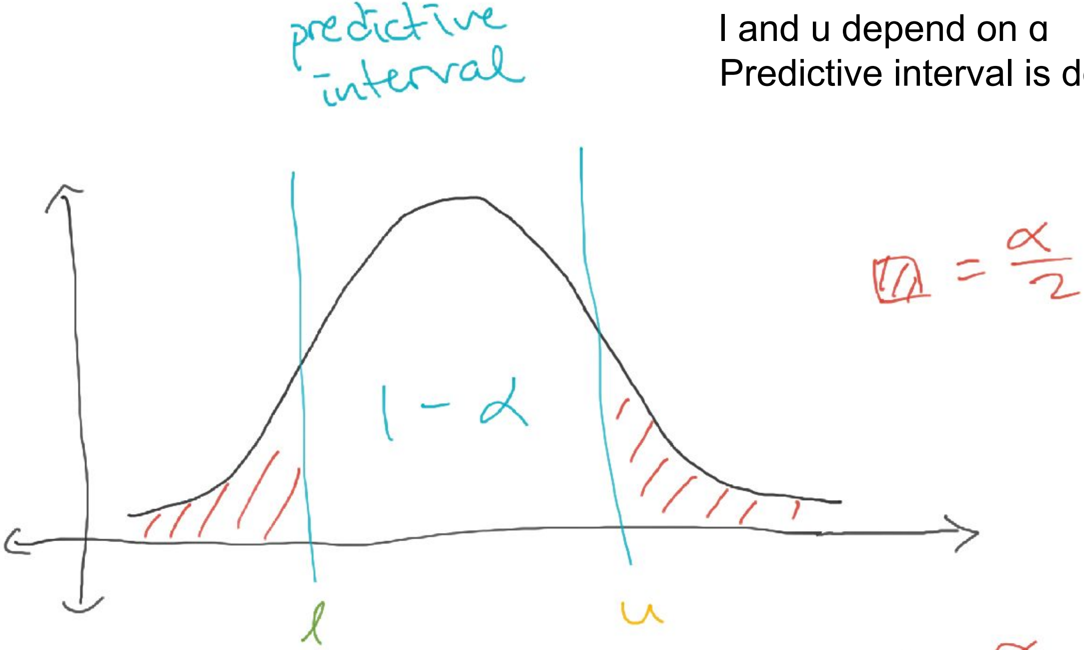
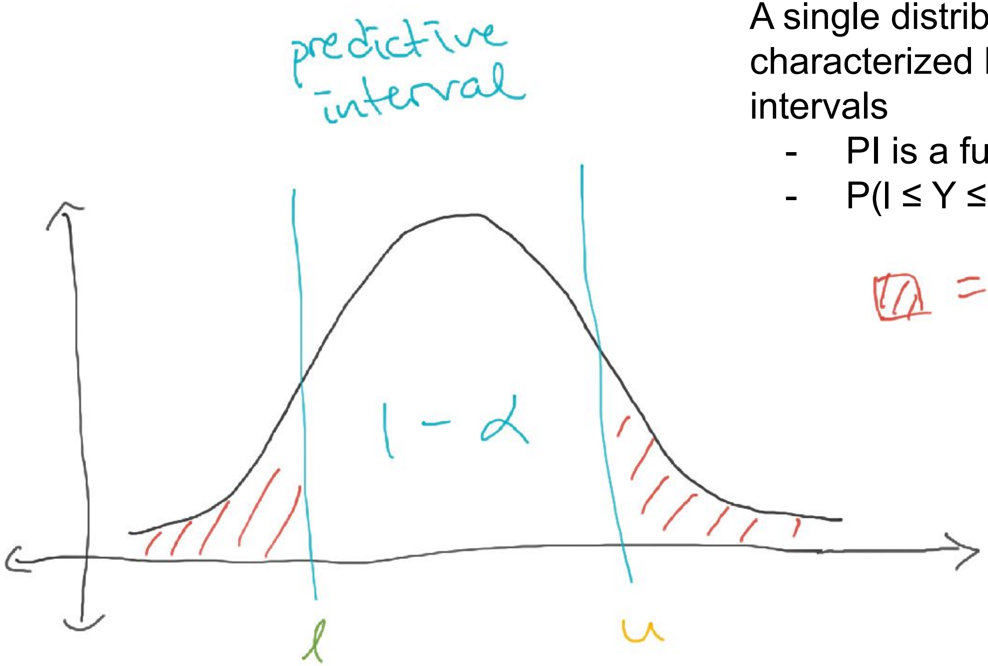
l and u depend on α
Predictive interval is defined by [l, u]



$P(X \leq l) = \frac{\alpha}{2}$ $\quad$ $P(X \leq u) = 1 - \frac{\alpha}{2}$

A single distribution can be characterized by multiple predictive intervals

- PI is a function of F and α → [l, u]
- P(l ≤ Y ≤ u) = 1 - α



$\boxed{//} = \frac{\alpha}{2}$

$P(X \leq l) = \frac{\alpha}{2}$

$P(X \leq u) = 1 - \frac{\alpha}{2}$

$$IS_\alpha(F,y) = (u-l) + \frac{2}{\alpha}(l-y)\boxed{\mathbb{1}(y<l)} + \frac{2}{\alpha}(y-u)\mathbb{1}(y>u)$$

The indicator function of a subset A of a set X is a function

$$\mathbf{1}_A \colon X \to \{0,1\}$$

defined as

$$\mathbf{1}_A(x) := \begin{cases} 1 & \text{if } x \in A\,, \\ 0 & \text{if } x \notin A\,. \end{cases}$$

$$IS_\alpha(F, y) = (u - l) + \boxed{\frac{2}{\alpha}(l - y)\mathbb{1}(y < l)} + \boxed{\frac{2}{\alpha}(y - u)\mathbb{1}(y > u)}$$

The last two terms penalize the model if the prediction is wrong

If the observed value (ground truth) is outside of the predicted interval [l, u], then the last two terms will be > 0

If the observed value is inside of the predicted value, $l \leq y \leq u$, then the last two terms are 0

The penalty is proportional to the distance between y and the lower end l of the interval, with the strength of the penalty depending on the level α

$$IS_\alpha(F, y) = \boxed{(u - l)} + \frac{2}{\alpha}(l - y)\mathbb{1}(y < l) + \frac{2}{\alpha}(y - u)\mathbb{1}(y > u)$$

Predictive interval [l, u]

- u = upper bound
- l = lower bound

This term penalizes prediction intervals that are broad. In other words, the forecaster is rewarded for narrow prediction intervals. For example, a technically correct but useless prediction would be [-∞, ∞]

Since by definition l ≤ u and the last two terms are ≥ 0, interval score (IS) is always ≥ 0

$$WIS_{\alpha_{0:K}}(F,y) = \frac{1}{K+\frac{1}{2}}\left(w_0 \cdot |y-m| + \boxed{\sum_{k=1}^{K} w_k \cdot IS_{\alpha_k}(F,y)}\right)$$

- α_K = Uncertainty level used to define the predictive interval
- K = Number of intervals to be evaluated
- m = Median of the predictive distribution
- $w_k = \alpha_k/2$
- $w_0 = 1/2$

$$WIS_{\alpha_{0:K}}(F,y) = \frac{1}{K+\frac{1}{2}}\left(w_0 \cdot |y-m| + \boxed{\sum_{k=1}^{K} w_k \cdot IS_{\alpha_k}(F,y)}\right)$$

It's not very informative to describe a distribution with a single interval, so it's common to describe a distribution by reporting multiple central predictive intervals at different levels:

- $(1 - \alpha_1) < (1 - \alpha_2) < \cdots < (1 - \alpha_K)$, along with the predict median m (which can be thought of as the central prediction interval at level $(1 - \alpha_0) \to 0$)

The values for the weights (w_0 and w_k) were chosen to show equivalency with a well established score for evaluating full predictive distributions (not shown here)

Since $w_k > 0$, $w_0 > 0$ and interval score ≥ 0, WIS ≥ 0

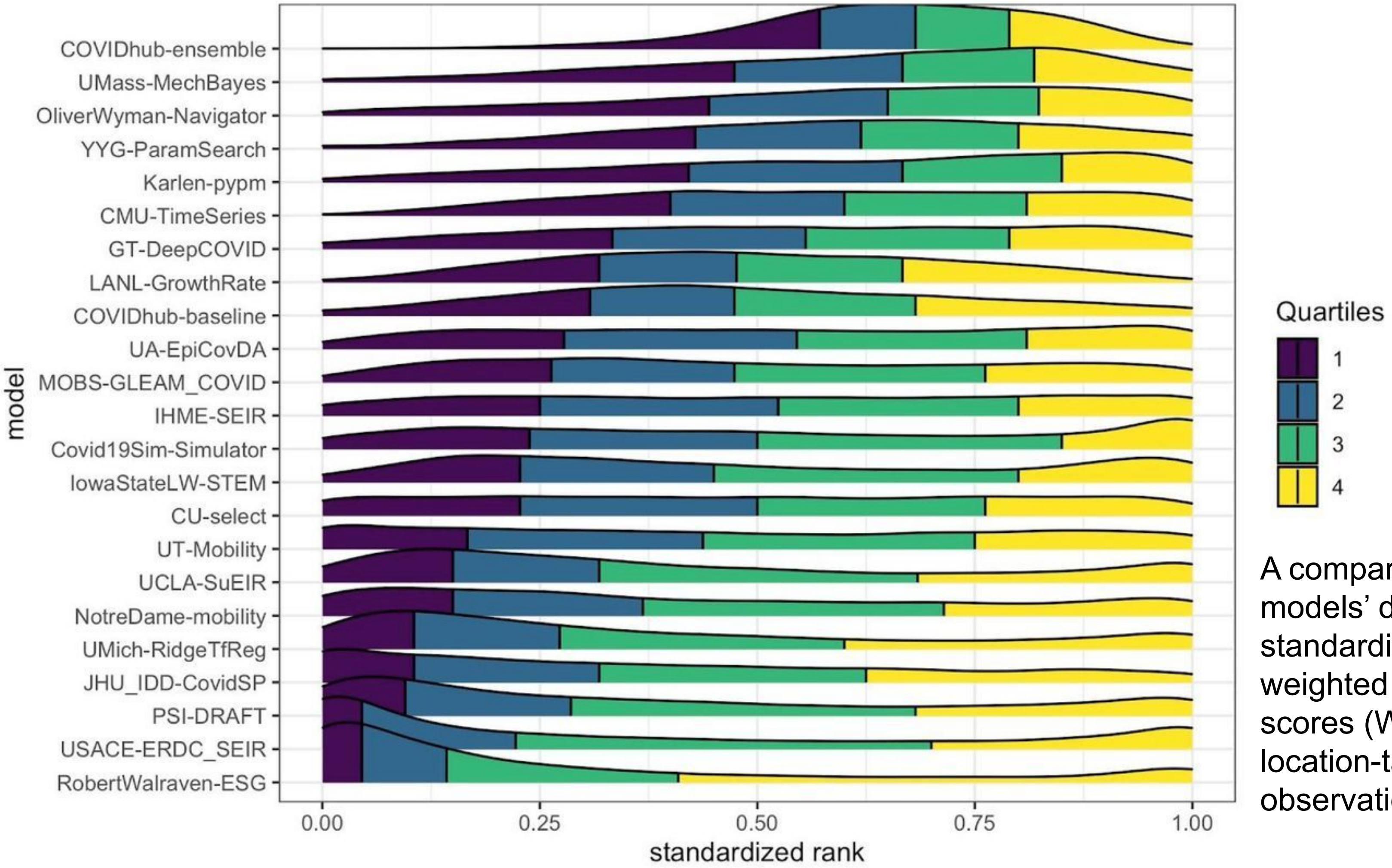
$$sr_{m,i} = 1 - \frac{r_{m,i} - 1}{n_i - 1}$$

This is the WIS standardized rank score that ranks models based on skill from 1 (best) to 0 (worst)

m = current model
i = observation
n_i = total number of models being evaluated (that made the observation for the skill)
$r_{m,i}$ = the rank of model m out of the n_i models, where the model with the best (i.e., lowest) WIS received a rank of 1 and the worst received a rank of n_i

- For example, the best model with $r_{m,i} = 1$ will result in $sr_{m,i} = 1 - 0 = 1$
- Worst performing model with $r_{m,i} = n_i$ will result in $sr_{m,i} = 1 - 1 = 0$

Just a way to standardize ranks, so that models can be compared through multiple observations



A comparison of each models’ distribution of standardized rank of weighted interval scores (WIS) for each location-target-week observation.