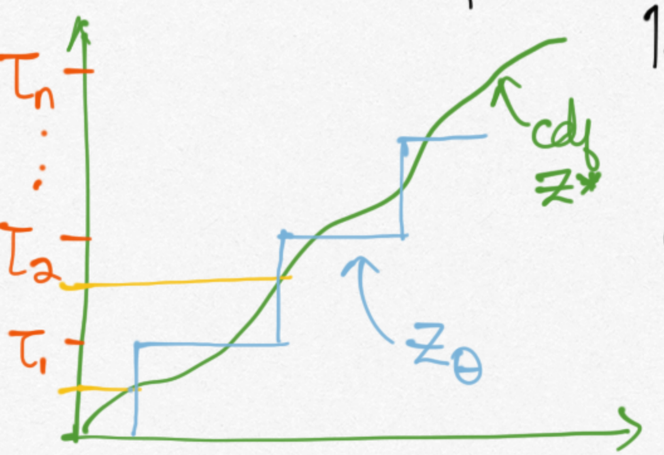
QR-DQN

1. We want to learn distribut^n of return Z
2. We can start with estimate Z_θ (θ: model params) & iteratively improve using bellman operator style updates.
3. It was proved that p-Wass. distance is a γ-contraction mapping & the distributional Bellman operator $T^\pi Z : R + \gamma Z(s',a')$ can converge to a fixed point (using p-Wass in its dual form)
4. $d_p(Z_1, Z_2) \equiv \sup_{s,a} W_p(Z_1(s,a), Z_2(s,a))$
5. Consider some true distribution we want to learn Z^*. If we want to use Wass. dist to learn, Wass dist is the max difference in CDFs' of the 2 distribut^{ns}.



What is the best approx distribut^n?

If we divide cdf into equal prob intervals, say N, then $\tau_i = i/N$. It turns out that best approx Z_θ is at quantile midpoints $\tau' = \frac{\tau_{i-1} + \tau_i}{2}$

So we can learn quantiles using QR.

Quantile Regression

Linear regression estimates "conditional mean" (mean/expectat^n over set of conditions)
If Least absolute deviat^n is used (MAE), we estimate median of target.

QR - we estimate conditional quantiles. (Median is special case, $q = 0.5$)
Given R.V. X: drive time Everett to Seattle
$q \in (0,1)$, $P(X \leq x) = q$
say $P(X \leq 42 \text{ min}) = 0.95$ implies
Prob. of getting to Seattle in ≤ 42 min is 95%

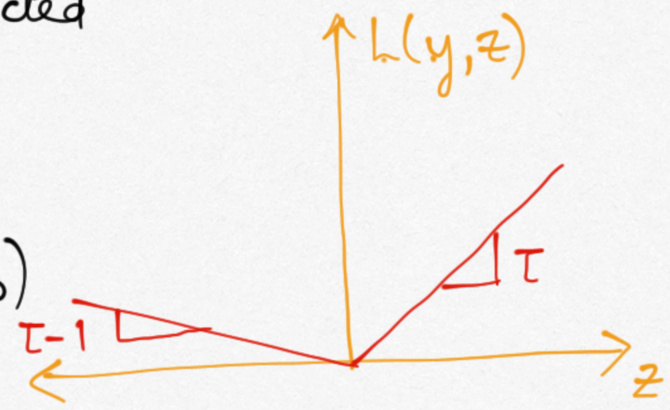
QR Loss (pinball loss)

τ - target quantile, y - actual, z - predicted

$$L_\tau(y-z) = \begin{cases} |(y-z)\tau| & y \geq z \\ |(z-y)(1-\tau)| & y < z \end{cases}$$

or written as $\rho_\tau(u) = u(\tau - \mathbb{I}_{u<0})$

u - residual, $\mathbb{I}$ - indicator fn.



Eg. $\tau = 0.05$, 5%, quantile value = Y of test scores.
Only 5% of students scored below Y
and 95% of students scored above Y.

Let $Y = 10$. If $z = 9$, $L_\tau = (10-9)\cdot(0.05) = 0.05$

$z = 11$, $L_\tau = (11-10)\cdot(1-0.05) = 0.95$

L_τ is higher when we predict on "wrong" side of quantile.

In practice we compute the loss fn between all pairs (θ_i, θ_j) of quantiles of estimated & target distribution.

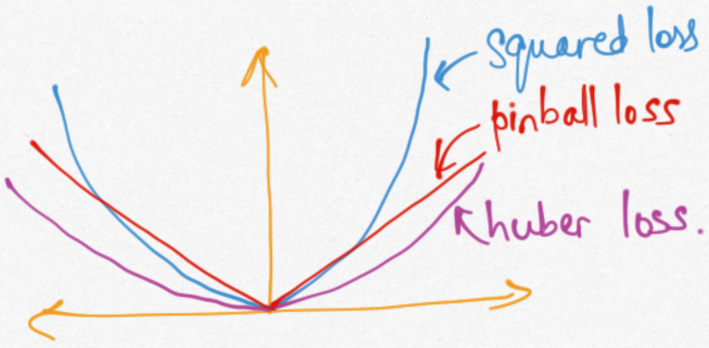
i.e. $\rho^k_{\hat{\tau}_i}(T\theta_j - \theta_i(s,a))$

$T\theta_j$: target (Bellman update)

$\theta_i(s,a)$: current estimate

Huber loss

QR loss is not zero everywhere.
When $u \to 0$, derivative is constant. So authors combine w/ huber loss.



$$L_k(u) = \begin{cases} \frac{1}{2}u^2 & |u| \leq k \\ k(|u| - \frac{1}{2}k) & \text{otherwise} \end{cases}$$

u - residual, $k = 0$ or 1

Quantile huber loss $= |\tau - \mathbb{I}_{u<0}| L_k(u)$