Ok, so marginal mean consistency of f isn't hard (or costly in terms of MSE).

What other things about $(x,y) \sim D$ can we find models to satisfy/learn?

What about quantiles? τ is a q-quantile of D if

$$\Pr_y [y \le \tau] = q$$

"q fraction of D is below τ"

It'd be rad to find f that was conditionally (on x) a q-quantile, generally hard/imp. But can ask for it marginally:

Def f has marginal quantile consistency error α for target quantile q if

$$\left| \Pr_{(x,y) \sim D} [y \le f(x)] - q \right| \le \alpha$$

Squared error : marginal mean consistency
pinball loss : quantile

Def: The pinball loss function for quantile q:

$$L_q(\tau, y) = \begin{cases} (y-\tau)q & y > \tau \\ (\tau - y)(1-q) & y \le \tau \end{cases}$$

For D & f,

$$PB_q(f, D) = \mathbb{E}_{(x,y) \sim D} [L_q(f(x), y)]$$

Lemma: For any continuous distribution every $0 \le q \le 1$

$$\tau_q = \underset{\tau \in [0,1]}{\arg\min} \; \mathbb{E}_y [L_q(\tau, y)]$$

is a q-quantile.

Pf Since y is continuous & actually convex in τ so is min. at a pt where it's subderivative is 0

$\int \mathbb{I}(\,)\,[(y-\tau)\,q]\cdot dy$
$\int \mathbb{I}(\,)\,(\tau-y)(1-q)\,dy$
is continuous

$$\frac{d\, \mathbb{E}_y [L_q(\tau, y)]}{d\tau} = \mathbb{E}_y [(1-q)\, \mathbb{1}[y \le \tau] - q\, \mathbb{1}[y > \tau]]$$

$$= \mathbb{E}_y [\mathbb{1}[y \le \tau] - q]$$

$$= \Pr[y \le \tau] - q$$

$= 0$ where τ is a q-quantile.

Can we "patch"/fix f to be q-quantile consistent like we did w. means?

Say f is α-violating q-quantile consisteng

Define Δ : $\Pr[y \le f(x) + \Delta] = q$ [$\exists$ since y is continuous]

$\hat{f}(x) = f(x) + \Delta$

$\hookrightarrow$ now MQC

{And only has lower pinball loss.}

But its improvement in pinball loss depends on density btwn f & $\hat{f}$

Recall ρ-Lipschitzness applied to cond label distribution $D(x)$ says $\forall 0 \le \tau \le \tau' \le 1$

$$\Pr_{y \sim D(x)}[y \le \tau'] - \Pr_{y \sim D(x)}[y \le \tau] \le \rho[\tau - \tau']$$

The whole of the label dist D is ρ-Lipschitz if $\forall x$, $D(x)$ is ρ-Lipschitz. (we just need marginal Lipschitzness not 1 ont

Lemma: Fix any D which is continuous & ρ-Lipschitz. If f has marginal consistency err α wrt q, Δ as above, $\hat{f} = f + \Delta$ then $PB_q(\hat{f}) \le PB_q(f) - \frac{\alpha^2}{2\rho}$

and $PB(\hat{f}) \le PB(f) + |\Delta|\alpha - \frac{\alpha^2}{2\rho}$

This argument is messier (more calculus / geometry) but morally similar (see p14-15 in Uncertainty notes).

Some overview of argument

$$\frac{d\, PB_q(f(x)+\tau)}{d\tau} = \mathbb{E}_x\left[\frac{d\, \mathbb{E}_{y\sim D_x} L_q(f(x)+\tau, y)}{d\tau}\right]$$

$$= \mathbb{E}_x[\Pr(y \le f(x)+\tau] - q]$$

$$= \Pr_{x,y}[y \le f(x)+\tau] - q$$

$$PB_q(f) - PB_q(f') = PB_q(f(x)+\Delta) - PB_q(f(x))$$

$$= \int_0^{\Delta} \frac{d\, PB_q(f(x)+\tau)}{d\tau}\, d\tau$$

$$= \int_0^{\Delta} \Pr_{x,y}[(y \le f(x)+\tau] - q)\, d\tau$$

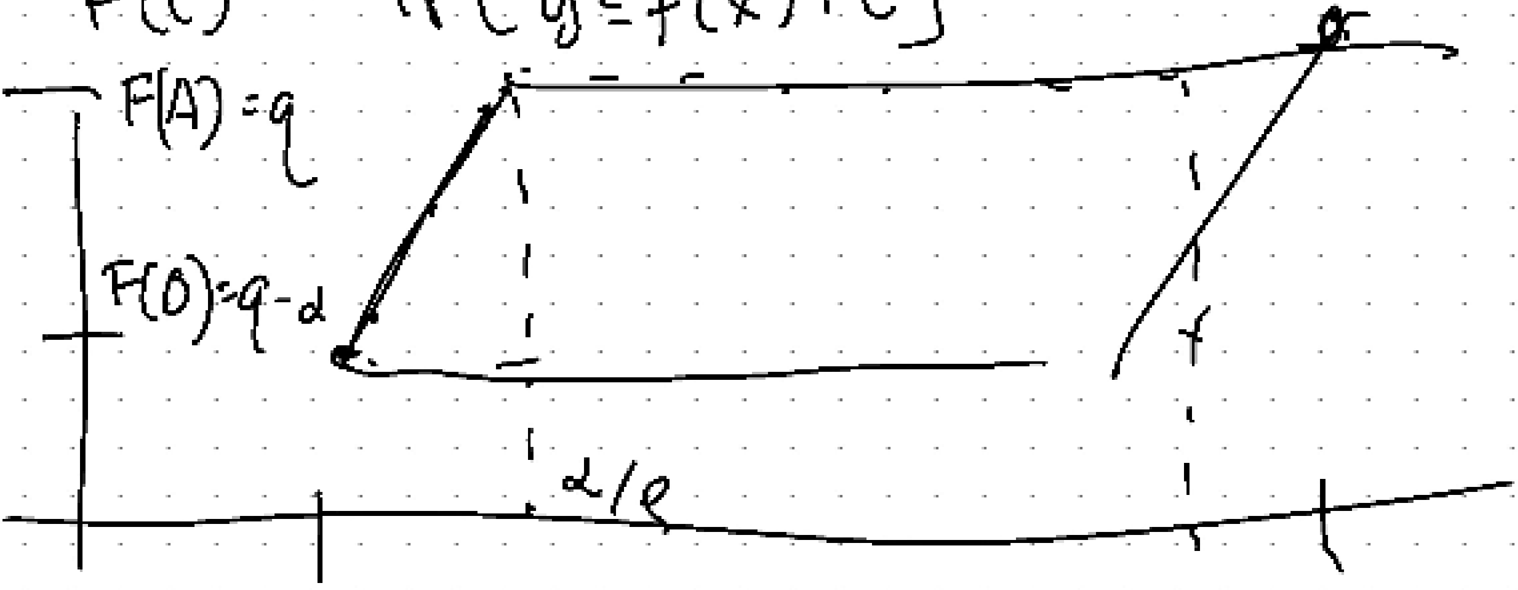
$$\begin{cases} \left[\int_0^{\Delta} \underline{\qquad\qquad} \right]_{d\tau} - |\Delta| q & \Delta \ge 0 \\ \left[\int_0^{\Delta} \Pr(y \le f(x)+\tau] \, d\tau\right] + |\Delta| q & \Delta < 0 \end{cases}$$

If we want to bound T It'll be useful to understand that integral.

lets consider $\Delta \geq 0$ (area under curve between $f(x), f(x)$)

f had quantile $q-\alpha$ $\left(\begin{array}{l} \text{or } q+\alpha \\ q \end{array}\right)$
$\hat{f}$ has quantile q

$F(\tau) = \Pr[y \leq f(x)+\tau]$



0 This area is biggest/smallest Δ if the slopes are as steep as possible

[slope $\leq \rho$!]

Ok, so the Lipschitzness of the CDF of y was critical, cuz it let us bound the AUC of $|F - \hat{F}|$

Eventually we'll be operating on sample distributions which aren't Lipschitz @ small scales (eg spaces btwn samples)

But we really only needed it for $|x-y| \geq \frac{\alpha}{\rho}$

[(ρ, r)-Lipschitzness: $D(x)$ is (ρ, r)-Lipschitz if $\forall\ 0 \leq \tau \leq \tau' \leq 1$ where $\tau' \geq \tau - r$:

$$\Pr_{y \sim D(x)}[y \leq \tau'] - \Pr_{y \sim D(x)}[y \leq \tau] \leq \rho(\tau' - \tau)$$

As alluded ↑, everything we'll do in practice will be wrt $S \sim D^n$ an n-sample from D.

What guarantees can we give for algs(S) wrt D?

Thm 1 Fix a model f, dist D, $S \sim D^n$. Let $\tilde{f}(x) = f(x) + \ldots$ satisfy MMC on S. Then, w.p. $1 - \delta$ over S $\tilde{f}$ has α-MMC wrt D, for

$$\alpha \leq \sqrt{\frac{2 \log \frac{1}{\delta}}{n}}$$

Pf [Hoeffding's! $X_1, \dots X_n$ independent RV's $a_i \le X_i \le b_i$ then $\Pr[|\sum X_i - \mathbb{E}[\sum X_i]| \ge t] \le 2\exp\left(\frac{-2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$

$$\Delta = \frac{1}{n}\sum_{(x,y)\in S}(y - f(x)) \qquad -\frac{1}{n} \le \frac{1}{n}(y - f(x)) \le \frac{1}{n}$$

$$\mathbb{E}[\Delta] = \mathbb{E}_{(x,y)\sim D}[y - f(x)] \overset{\text{Hoeffding's}}{\Longrightarrow} \Pr[|\Delta - \mathbb{E}[y-f(x)]| \ge t] \le 2\exp\left(\frac{-nt^2}{2}\right) = \delta$$

$$\Pr\left[|\Delta - \mathbb{E}[y - f(x)]| \ge \sqrt{\frac{2\log\frac{2}{\delta}}{n}}\right] \le \delta$$

$$|\mathbb{E}_{(x,y)\sim D}[\hat{f}(x) - y]| = |\mathbb{E}[f(x) + \Delta - y]|$$

$$\le |\Delta - \mathbb{E}[y - f(x)]|$$

Not quite what we want... $\hat{f}$?

$\le$

Analogously, quantile consistency on a sample will imply apx quantile consistency on D. The "slack" parameter is $\approx$ same (α-QC on sample $\Rightarrow \alpha + \sqrt{\frac{\log\frac{2}{\delta}}{2n}}$ on D)

But the argument requires using Dvoretzky-Kiefer-Wolfowitz (DKW) which says CDFs $\hat{F}$ are well-concentrated w.r.t. F

$$\Pr[|F(c) - \hat{F}_S(c)| \ge t] \le 2\exp(-2nt^2)$$

(See p 14-15 out of Uncertainty notes for pf)

Everything we've done so far has been for iid data

Adversarial sequences (x_i, y_i) are generally way harder.

However, for MMC and MQC, we can get nice adversarial guarantees.

Fix an alg A.

An adversary selects

transcript $\Pi = \{(x_1, p_1, y_1)$ features, prediction, label

....

$(x_T, p_T, y_T)\}$

$[0,1]$

We'll say Π is α-MMC if

$$\left| \frac{1}{T} \sum p_t - \frac{1}{T} \sum y_t \right| = \alpha$$

& α-MQC for quantile q if

$$\left| \frac{1}{T} \sum \mathbb{1}[y_t \le p_t] - q \right| = \alpha$$

---

we can look @ sequential but iid data as a warmup

1. $(\hat{x}_t, \hat{y}_t) \sim D$
2. Give alg $(\hat{x}_t, \hat{y}_t - f(\hat{x}_t)) = (x_t, y_t)$

Fix model f

w. responses p_t

Let $\Delta = \frac{1}{T} \sum p_t$

$\hat{f}(x) = f(x) + \Delta$

Then if Alg is α-MMC for any adversary, wp $1-\delta$ $\hat{f}$ is α'-MMC for $\alpha' \le \alpha + \sqrt{\frac{2 \log 2/\delta}{T}}$

Pf π is α-MMC since Alg is α MMC

so - $\left|\frac{1}{T}\sum p_t - \frac{1}{T}\sum y_t\right| \le \alpha$

Let $\bar{\Delta} = \frac{1}{T}\sum\left[\hat{y}_t - f(\hat{x}_t)\right]$

$$= \left|\Delta - \frac{1}{T}\sum y_t\right| = \left|\Delta - \frac{1}{T}\sum\left[\hat{y}_t - f(\hat{x}_t)\right]\right|$$

$$= \left|\Delta - \bar{\Delta}\right| \le \alpha \quad \bigstar$$

$(\hat{x}_t, \hat{y}_t) \sim_{iid} D \Rightarrow \mathbb{E}[\bar{\Delta}] = \mathbb{E}[y - f(x)]$

So Hoeffding $\left|\left[\frac{1}{T}\left(\hat{y}_t - f(\hat{x}_t)\right)\right]\right| \le \frac{1}{T}$

$$\Rightarrow \Pr\left[\left|\bar{\Delta} - \mathbb{E}[y - f(x)]\right| \ge \epsilon\right] \le 2\exp\left(\frac{-2T^2\epsilon}{4}\right)$$

$$\le \delta \quad \text{for } \epsilon \ge \sqrt{\frac{2\log\frac{2}{\delta}}{T}}$$

$$\left|\mathbb{E}\left[y - \hat{f}(x)\right]\right| = \left|\mathbb{E}\left[f(x) + \Delta - y\right]\right|$$

$$\le \left|\mathbb{E}\left[f(x) + \bar{\Delta} - y\right]\right| + \alpha \quad [\text{By } \bigstar]$$

$$= \left|\mathbb{E}\left[f(x) + \mathbb{E}[\bar{\Delta}] + (\bar{\Delta} - \mathbb{E}[\bar{\Delta}]) - y\right]\right| + \alpha \quad = 0$$

$$= \left|\mathbb{E}\left[\bar{\Delta} - \mathbb{E}[\bar{\Delta}]\right]\right| + \alpha$$

$$\le \alpha + \sqrt{\frac{2\log\frac{2}{\delta}}{T}}$$

We don't really need the $(x_t, y_t) \sim_{iid} D$ to get α-MMC:

Online MMC

Let $y_0 = 0$
For $t = 1 \ldots T$
get x_t, ignore, predict $p_t = y_{t-1}$
Observe y_t

This is $\frac{1}{T}$-MMC: [Which sucks since $y = (0\,1\,0\,1\,0\,1\,0\,1\,0\,1\,0)$

$\left| \frac{1}{T} \sum p_t - \frac{1}{T} \sum y_t \right|$

$= \left| \frac{1}{T} \sum y_{t-1} - \frac{1}{T} y_t \right| = \frac{1}{T} |y_0 - y_T| \leq \frac{1}{T}$.

So marginal mean guarantees aren't very strong (& are super easy to guarantee).

Quantiles are a bit trickier. We'll argue

Online MQC $\Rightarrow$ Distribution MQC, $\exists$ a deterministic algo for OMQC

We will need to randomize instead of average

Thm Fix an Alg A which, facing any adversary for T rounds is α-MQC for quantile q.
Fix a model $f: \mathcal{X} \to [0,1]$ & feed A the adversary simulation, $x_t = \tilde{x}_t$, $y_t = \tilde{y}_t - f(\tilde{x}_t)$
$\pi = \{(x_t, p_t, y_t)\}$
Let Δ be a RV $\sim_{\text{unif}} [p_1, \ldots, p_T]$
Then $\forall \delta \in [0,1]$, $\hat{f}(x) = f(x) + \Delta$ is α'-MQC for quantile q for

$$\alpha' \leq \alpha + \sqrt{\frac{2 \log \frac{2}{\delta}}{T}}$$

$\left(\Pr[y \leq f(x) + \Delta - q] \leq \alpha' \right]$

Pf

Since Π is α-MQC for q,

$$\left| P \right.$$

$$= \left| \frac{1}{T} \sum_t^t \left[f(\hat{x}_t) - \hat{y}_t \le P_t \right] - q \right|$$

Let D' : , F' its CDF

$\hat{y} = y - f(x)$

$(x,y) \sim D$.

Ideally $F'(x) = \Pr_{y \sim D'}[y \le x]$

would be a good q-quantile,

but... P_t's aren't iid! y_t's ($\Rightarrow \hat{y}_t$'s) are

So Hoeffding isn't applicable.

But, Azuma's $\neq$ is:

Let $X_1, \dots X_n$ be (not nec ind) RVs $|X_i| \le c_i$

$X_{<i} = (X_1, \dots X_{i-1})$. Then $\forall\ t > 0$

$$\Pr\left[\left| \sum_{i=1}^n X_i - \sum \mathbb{E}[X_i \mid X_{<i}] \right| \ge t \right] \le 2\exp\left(\frac{-t^2}{2\sum_{i=1}^n c_i^2} \right)$$