Papers about inference for adaptive-collected data

September 27, 2021

1 Post-Contextual-Bandit Inference

1.1 Short story

I have some adaptive-collected data (X_t, Y_t, A_t) and the corresponding adaptive policy g_t. Now I try to evaluate some target (a policy, treatment effect) using a function $g^*(a|x)$.

I know I can use importance sampling to get an unbiased estimator like IPW (inverse propensity weighted) and its advanced DR (double robust). But many experiments have shown that they are not asymptotically normal, which means that I cannot easily construct a confidence interval with a specified confidence level.

OK is there any solution in the literature? Yes, some work proposes a so-called stabilized estimator, which divides each term of estimator by its conditional standard deviation based on the history. Such trick has been applied in the non-contextual bandit setting, so ths work deals with the contextual setting.

The main difficulty may be the construction of the estimator of conditional standard deviation.

1.2 Simple formulation

The outcome at time t is $O(t) = (X(t), A(t), Y(t))$ corresponding to context, action and reward.

The action A_t is drawn from a known policy g_t, which is $O(1), O(2), \ldots, O(t-1)$-measurable.

$X(t)$ and $Y(t)$ have their own but unknown time-independent distributions which are $Q_{0,X}, Q_{0,Y}(\cdot|A, X)$.

Evaluation target:

$$\Psi(Q_X, Q_Y) = \int y Q_X(dx) g^*(a|x) d\mu_A(a) Q_Y(dy|a, x), \tag{1}$$

Note that g^* is irrelevant to policy g_t, terrible notation. $|g^*(a|x)| \leq G$ for any x, a.

Of course, we can simplify the calculation about Y:

$$\bar{Q}_0(a,x) = E_{Q_{0,Y}}(\cdot|x,a)[Y] = \int yQ_{0,Y}(dy|a,x) \tag{2}$$

So we can rewrite

$$\Psi(Q_X, \bar{Q}) = \int \bar{Q}(a,x)Q_X(dx)g^*(a|x)d\mu_A(a), \tag{3}$$

Now we can introduce the basic DR estimator:

$$D'(g,\bar{Q})(x,a,y) = \frac{g^*(a|x)}{g(a|x)}(y - \bar{Q}(a,x)) + \int \bar{Q}(a',x)g^*(a'|x)d\mu_A a' \tag{4}$$

What is canonical gradient?

$$D = D' - \Psi \tag{5}$$

For convenience, define an integration operator

$$P_{Q,g}f = \int f(x,a,y)Q_X(dx)g(a|x)d\mu_A(a)Q_Y(dy|a,x) \tag{6}$$

when g is $O(1),\ldots,O(t-1)$-measurable, then just replace the expectation with conditional expectation.

1.3 Proof sketch

2 steps: how to show the asymptotic normality and how to construct a consistent estimator of conditional standard deviation.

Asymptotic normal estimator

Stabilized estimator:

$$\hat{\Psi}_T = (\frac{1}{T}\sum_{t=1}^{T}\hat{\sigma}_t^{-1})^{-1}\frac{1}{T}\sum_{t=1}^{T}\hat{\sigma}_t^{-1}D'(g,\hat{\bar{Q}}_{t-1}) \tag{7}$$

How to show it?

1. construct a martingale difference sequence (MDS):

   $Z_{t,T} = T^{-\frac{1}{2}}[\hat{\sigma}_t^{-1}D'(O(t)) - P_{Q_0,g_t}\hat{\sigma}_t^{-1}D'(O(t))]$

   Note that RHS is the conditional expectation of LHS.

2. For such martingale triangular array, the target is to use Lindeberg condition to prove the CLT, so just need to prove the condition is satisfied.

   - Show that sum of variances of $Z_{t,T}$ is convergent. (Use assumption that $\hat{\sigma}_t$ is consistent and non degenerate efficiency bound)
   - Lindeberg's condition: $\sum_{t=1}^{T} E[Z_{t,T}^2 1(Z_{t,T} \geq \epsilon)] \xrightarrow{p} 0$. $Z_{t,T} = O(\delta_t^{-1}T^{-1/2}\hat{\sigma}_t^{-1})$, $\delta_t \geq Ct^{-1/2}$(assumption),