Inference of the optimal rate in n of the optimal truncation level δ_n

September 26, 2016

Let $P_1 = Q_1 g_1$ be the limit of P_n^*. Let $\delta_n(P)$ the optimal truncation level under P. Let's assume that $\delta_n(P) \sim C(P) n^{-r(P)}$.

Let $\sigma^2(P)(\delta) = PD_\delta^*(P)^2$ and $b(P)(\delta) = \Psi_\delta(P) - \Psi_0(P)$. Assume that $\sigma(P)^2(\delta) \sim C_0 \delta^{-2\gamma(P)}$, $\gamma(P) \geq 0$, and $b(P)(\delta) = C_1 \delta^{1-\beta(P)}$, $\beta < 1$. Under these assumptions, the optimal truncation level under P is $\delta_n(P) = n^{-\frac{1}{2(\gamma(P)+1-\beta(P))}}$, i.e. $r(P) = \frac{1}{2(\gamma(P)+1-\beta(P))}$.

Asymptotically consistent estimator of the $r(P_1)$: Under the above assumptions, we have proven earlier that we have an estimator $\hat{r}(P_n^*) \xrightarrow{P} r(P_1)$. However simulations have shown that for several reasonable correctly specified parametric models, performance starts being reasonable for $n > 10^7$, which makes it impractical.

Finite sample rule for the choice of the rate: I trained a super-learner on samples of random distributions. This gives a rule that works well in finite samples.

More formally: Consider a random probability distribution P with sample space $\mathcal{O}$. Let π the probability distribution of P.

Let $P_1, ..., P_m \sim^{i.i.d.} \pi$. Let $N_1, ..., N_m \sim \text{Unif}([10^2, 10^5])$.

Let

$$
\begin{aligned}
\mathbf{O}_1 &= (O_{1,1}, ..., O_{1,N_1}), \text{ with } O_{1,1}, ..., O_{1,N_1} \sim^{i.i.d.} P_1, \\
\mathbf{O}_2 &= (O_{2,1}, ..., O_{2,N_2}), \text{ with } O_{2,1}, ..., O_{2,N_2} \sim^{i.i.d.} P_2 \\
&\quad ... \\
\mathbf{O}_m &= (O_{m,1}, ..., O_{m,N_2}), \text{ with } O_{m,1}, ..., O_{m,N_m} \sim^{i.i.d.} P_m.
\end{aligned}
$$

For each P_i, $i \in \{1, ..., m\}$ I know the true optimal rate $r(P_i)$, and for each sample $\mathbf{O}_i$, I compute a bunch of features $\Phi(\mathbf{O}_i) = (\phi_1(\mathbf{O}_i), ..., \phi_p(\mathbf{O}_i))$.

I am using a super learner to regress the true rates onto the features. This yields an estimator $\hat{\rho}_m$ of the conditional expectation of $r(P)$ given $\Psi(\mathcal{O})$.

Assume that for a given sample $\mathbf{O}_i$ I can write my previous asymptotically consistent estimator $\hat{r}$ as a function of the features $\Phi(\mathbf{O}_i)$. Then if I incorporate it in my super learner as a one of the base learners, I hope I'll get, for a given distribution P_0 in the sample space of π and a n iid observations $O'_1, ..., O'_n$ of P_0, $\rho_m(\Phi(O'_1, ..., O'_n)) \xrightarrow{n,m\to\infty} 0$ in probability.

But that doesn't guarantee that a given rule $\rho_m(\Phi(O'_1, ..., O'_n))$ is asymptotically equivalent with $\hat{r}_n(P_n^*)$.

I have another option to provide a rule that would be asymptotically equivalent with $\hat{r}_n(P_n^*)$:

For a given large m, let $\tilde{\rho}_m = \alpha_0 n^{-\lambda_n(\Phi(\mathbf{O'})} \rho_m(\Phi(\mathbf{O'})) + (1 - \alpha_0 n^{-\lambda_n(\Phi(\mathbf{O'})}) \hat{r}_n(P_n^*)$.

Train a super learner to find a good rule $\lambda_n(\Phi(\mathbf{O'})$.