Sample exam questions
Advanced Econometrics 2: Foundations of Machine Learning
Oxford, 2020

Maximilian Kasy

1. Consider the normal-means estimation problem as discussed in class, where $\boldsymbol{\theta} \in \mathbb{R}^k$ and $\boldsymbol{X} \sim N(\boldsymbol{\theta}, I_k)$, with quadratic error loss, $L(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \sum_i (\widehat{\theta}_i - \theta_i)^2$. Recall that Stein's Unbiased Risk Estimate for an estimator of the form $\widehat{\boldsymbol{\theta}} = \boldsymbol{X} + \boldsymbol{g}(\boldsymbol{X})$ is given by
$$\widehat{R} = k + \sum_i \left(\widehat{\theta}_i - X_i\right)^2 + 2 \cdot \sum_i \partial_{x_i} g_i(\boldsymbol{X}).$$
Suppose that $\widehat{\boldsymbol{\theta}}$ is equal to the so-called Lasso-estimator,
$$\widehat{\boldsymbol{\theta}} = \underset{\tilde{\boldsymbol{\theta}}}{\operatorname{argmin}} \sum_i \left(\tilde{\theta}_i - X_i\right)^2 + \lambda \cdot \sum_i |\tilde{\theta}_i|,$$
for some non-random constant λ.
   (a) What is the function $\boldsymbol{g}$ corresponding to the Lasso estimator?
   (b) Explicitly derive $\widehat{R}$ for the Lasso estimator.
   (c) How could you use this expression to choose λ optimally?

2. Consider data from the multi-armed bandit setting, where we have a sample of observations (Y_t, D_t) for $t = 1, \ldots, T$, with $D_t \in 1, \ldots, k$, and $Y_t = Y_t^{D_t} \in \mathbb{R}$, and D_t was assigned by some adaptive algorithm.

   Assume that the potential outcomes $(Y_t^1, \ldots, Y_t^k)$ are i.i.d. across t, with $E[Y_t^d] = \theta^d$. Let $T_T^d = \sum_t \mathbf{1}(D_t = d)$ and let $\bar{Y}_T^d = \frac{\sum_t \mathbf{1}(D_t=d) \cdot Y_t}{T_T^d}$ be the sample average of outcomes for the observations assigned to treatmend d.

   Derive a bound for the probability

$$P(\max_d \bar{Y}_T^d > M)$$

   for some constant $M > \max_d \theta^d$. Does the bound simplify if T_T^d is much smaller for one of the treatment values d, relative to the others?

   Hints:

   - *Recall the large deviations inequality* $P(\bar{Y}_T^d - \theta^d > \epsilon) < \exp(-T_T^d \cdot \psi^*(\epsilon))$ *where* $\psi^*(\cdot)$ *is the Legendre transform of the log moment generating function of* Y^d.
   - *Recall the inequality, which holds for for any pair of events* A, B, $P(A \cup B) \leq P(A) + P(B)$.