Single Estimator Overestimation Proof

Hado van Hasselt
Multi-agent and Adaptive Computation Group
Centrum Wiskunde & Informatica

This lemma accompanies the paper "Double Q-learning" that was accepted for publication in the Advances in Neural Information Processing Systems, volume 23 (NIPS 2010). It is a generalization of the result in [1].

Lemma 1. *Let $X = \{X_1, \ldots, X_M\}$ be a set of random variables and let $\mu = \{\mu_1, \ldots, \mu_M\}$ be a set of unbiased estimators such that $E\{\mu_i\} = E\{X_i\}$, for all i. Assume that the set of samples S contains at least one sample for each of the variables in X. Let $\mathcal{M}$ be the set of labels of estimators that maximize the expected values of X:*

$$\mathcal{M} \stackrel{\text{def}}{=} \left\{ j \mid E\{X_j\} = \max_i E\{X_i\} \right\} \quad .$$

Let $Max(S)$ be the set of labels of estimators that yield the maximum estimate for some set of samples S:

$$Max(S) \stackrel{\text{def}}{=} \left\{ j \mid \mu_j(S) = \max_i \mu_i(S) \right\} \quad .$$

Then, for all $j \in \mathcal{M}$

$$E\{\max_i \mu_i\} \geq E\{\mu_j\} = E\{X_j\} \stackrel{\text{def}}{=} \max_i E\{X_i\} \quad . \tag{1}$$

Furthermore, the inequality is strict if and only if $P(j \notin Max) > 0$, for any $j \in \mathcal{M}$.

Proof. Assume $j \in \mathcal{M}$, i.e. μ_j is any estimator whose expected value is maximal. Then

$$\begin{aligned}
E\{\max_i \mu_i\} &= P(j \in Max)E\{\max_i \mu_i\} + P(j \notin Max)E\{\max_i \mu_i\} \\
&= P(j \in Max)E\{\mu_j | j \in Max\} + P(j \notin Max)E\{\max_i \mu_i\} \\
&\geq P(j \in Max)E\{\mu_j | j \in Max\} + P(j \notin Max)E\{\mu_j | j \notin Max\} \\
&= E\{\mu_j\} = E\{X_j\} \stackrel{\text{def}}{=} \max_i E\{X_i\} \quad .
\end{aligned}$$

By definition of Max we have $E\{\max_i \mu_i\} > E\{\mu_j | j \notin Max\}$, for any j. Therefore, the inequality is strict if and only if $P(j \notin Max) > 0$, for some $j \in \mathcal{M}$. If we do not know whether this is the case, we do not know if the inequality in (1) is strict and therefore in general we write $E\{\max_i \mu_i\} \geq \max_i E\{\mu_i\}$. $\square$

References

[1] J. E. Smith and R. L. Winkler. The optimizer's curse: Skepticism and postdecision surprise in decision analysis. *Management Science*, 52(3):311–322, 2006.