MOCU Methodology

Anthony DeGennaro

April 2019

1. **Cost and Optimality**
   - Assume we have a cost function $C(\theta, \psi)$ that quantifies a cost related to our experimental design objective
   - Our experimental design, then, is to seek a $\psi_{IBR}^{\Theta} \in \Psi$ that is optimal on average over all Θ w.r.t. this cost:

     $$\mathbb{E}_\theta[C(\theta, \psi_{IBR}^{\Theta})] \leq \mathbb{E}_\theta[C(\theta, \psi)] \quad \forall \psi \in \Psi$$
   - Of course, if we had perfect knowledge of θ, then we could design a classifier ψ_θ for that specific value that would almost certainly be better than ψ_{IBR}^{Θ}. We can quantify this by computing the cost difference between the two choices, averaged over all of Θ, called the MOCU:

     $$M_\Psi(\Theta) = \mathbb{E}_\theta[C(\theta, \psi_{IBR}^{\Theta}) - C(\theta, \psi_\theta)]$$

     where $\psi_\theta \in \Psi$ denotes the classifier that is optimal for the particular choice of θ. Thus, we should choose experiments in a way that seeks to minimize this MOCU.
2. **Adaptive Experimental Selection**
   - Given a new piece of data $(x, y) \in (x, y)$, we can compute a new MOCU conditioned on that new piece of information. For ease of notation, let $\xi = (x, y)$, and hence $\mathbb{E}_\xi[\cdot]$ refers to the expectation over $\rho(y|x)$ (i.e., the probability that y occurred, given x):

     $$M_\Psi(\Theta|\xi) = \mathbb{E}_{\theta|\xi}[C(\theta, \psi_{IBR}^{\Theta}) - C(\theta, \psi_\theta)]$$
   - Averaging this over many experiments gives the average conditional MOCU:

     $$D_\Psi(\Theta, \xi) = \mathbb{E}_\xi[M_\Psi(\Theta|\xi)]$$
   - The experiment x^* that minimizes this quantity is said to be optimal:

     $$\begin{aligned} x^* &= \text{argmin}_{x \in X} \; D_\Psi(\Theta, \xi) \\ &= \text{argmin}_{x \in X} \; \mathbb{E}_\xi[\mathbb{E}_{\theta|\xi}[C(\theta, \psi_{IBR}^{\Theta}) - C(\theta, \psi_\theta)]] \end{aligned}$$
   - Because x^* also minimizes the quantity $D_\Psi(\Theta, \xi) - M_\Psi(\Theta)$, one can show after some algebra that it also minimizes this quantity:

     $$x^* = \text{argmin}_{x \in X} \; \mathbb{E}_\xi[\mathbb{E}_{\theta|\xi}[C(\theta, \psi_{IBR}^{\Theta|\xi})]] - \mathbb{E}_\theta[C(\theta, \psi_{IBR}^{\Theta})]$$
   - And, because the cost $C(\theta, \psi)$ does not vary with x, we may eliminate it to obtain:

     $$x^* = \text{argmin}_{x \in X} \; \mathbb{E}_\xi[\mathbb{E}_{\theta|\xi}[C(\theta, \psi_{IBR}^{\Theta|\xi})]]$$
3. **MOCU-Specific Calculus**
   - The equation for x^* involves the double-nested expectation $\mathbb{E}_\xi[\mathbb{E}_{\theta|\xi}[\cdot]]$
   - Outer loop: $\mathbb{E}_\xi[F] = \int_\xi F(\xi)\rho(\xi)d\xi$ where:

     $$\rho(\xi) = \mathbb{E}_\theta[\rho(\xi|\theta)] = \int_\theta \rho(\xi|\theta)\rho(\theta)d\theta$$

     Thus, we must know $\rho(\xi|\theta)$ (or be able to sample from it e.g. with a computer model) a priori (assumption)

- Inner loop: $\mathbb{E}_{\theta|\xi}[F] = \int_\xi F(\xi)\rho(\theta|\xi)d\xi$ where, by Bayes' law:

  $$\rho(\theta|\xi) = \frac{\rho(\xi|\theta)\rho(\theta)}{\rho(\xi)}$$

  Thus, we must also know $\rho(\theta)$ (uncertain parameter distribution) a priori (assumption)

4. **Software Algorithmic Implementation**

   inputs :

   (1) enumerable set of discrete values $\Theta = \{\theta_1 \dots \theta_k\}$ with associated probability distribution $\rho(\Theta)$

   (2) enumerable set of discrete actions $\Psi = \{\psi_1 \dots \psi_p\}$

   (3) cost function $C(\theta, \psi)$

   (4) set of experimental choices $X = \{x_1 \dots x_n\}$ with possible outcomes $Y = \{y_1 \dots y_m\}$

   (5) conditional distribution $\rho(Y|X, \Theta)$ (possibly gotten by MC sampling a surrogate)

   algorithm (single step of MOCU experimental choice loop) :

   for all x_i in X:

   for all y_j in Y:

   Compute $\rho(\theta|x_i = y_j)$ via Bayes with priors $\rho(\theta)$, $\rho(y_j|x_i, \theta)$

   Compute $\psi(\Theta|x_i = y_j) = \text{argmin}_\psi \ \mathbb{E}_{\theta|x_i=y_j}[C(\theta, \psi)]$

   Compute $\omega(x_i = y_j) = \mathbb{E}_{\theta|x_i=y_j}[C(\theta, \psi(\Theta|x_i = y_j))]$

   Compute $\mathbb{E}_{y|x_i}[\omega(x_i)]$ (i.e., averaged over the m outcomes in Y)

   Compute $x^* = \text{argmin}_X \ \mathbb{E}_{y|x}[\omega(x)]$ (i.e., minimization over the n experiments in X)

   outputs : x^* , $\psi(\Theta|x^*, Y)$, $\rho(\theta|x^*, Y)$

   (1) x^* (optimal experiment)

   (2) $\psi(\Theta|x^*, Y)$ (robust action for every possible outcome of x^*)

   (3) $\rho(\theta|x^*, Y)$ (conditional posterior of θ for every possible outcome of x^*)