Knowledge Gradient (KG) Model.

M: a collection of distinct alternatives.

θ_i: unknown mean $\rightarrow \theta: (\theta_1, \cdots \theta_M)'$.

λ_i: known variance.

$\theta \sim N(\mu^0, \Sigma^0)$ (1)

$x^0, x^1, \cdots, x^{N-1}$: a sequence of N sampling decisions.

x^n selects an alternative from $\{1, \cdots M\}$.

$\varepsilon^{n+1} \sim N(0, \lambda_{x^n}) \rightarrow$ measurement error.

$\hat{y}^{n+1} = \theta_{x^n} + \varepsilon^{n+1}$, $\hat{y}^{n+1} \sim N(\theta_{x^n}, \lambda_{x^n})$.

F^n: σ-algebra generated by samples observed by time n and the identies of their originating alternatives. $(x^0, \hat{y}^1, x^1, \hat{y}^2, \cdots, x^{n-1}, \hat{y}^n)$.

$E_n \sim E[\cdot | F^n]$

$\mu^n := E_n[\theta]$, $\Sigma^n := Cov[\theta | F^n]$

Π: set of experimental designs, satisfying the sequential requirement.

$\Pi := \{(x^0, \cdots, x^{N-1}) : x^n \in F^n\}$.

$\pi = (x^0, \cdots, x^{N-1}) \rightarrow$ a generic element of π.

Target: choose a measurement policy maximizing expected reward.

$$\sup_{\pi \in \Pi} E^\pi [\max_i \mu_i^N]. \quad (2).$$

update equations.

$$\mu^{n+1} = \Sigma^{n+1}((\Sigma^n)^{-1}\mu^n + (\lambda_{x^n})^{-1} \hat{y}^{n+1} e_{x^n}). \quad (3).$$

$$\Sigma^{n+1} = ((\Sigma^n)^{-1} + (\lambda_{x^n})^{-1} e_{x^n}(e_{x^n})')^{-1}, \quad (4).$$

$x = x^n$

$$\mu^{n+1} = \mu^n + \frac{\hat{y}^{n+1} - \mu_x^n}{\lambda_x + \Sigma_{xx}^n} \Sigma^n e_x. \quad (5)$$

$$\Sigma^{n+1} = \Sigma^n - \frac{\Sigma^n e_x e_x' \Sigma^n}{\lambda_x + \Sigma_{xx}^n} \quad (6).$$

$$\tilde{\sigma}(\Sigma, x) := \frac{\Sigma e_x}{\sqrt{\lambda_x + \Sigma_{xx}}}. \quad \tilde{\sigma}: \text{a vector-valued function} \quad (7).$$

$Var[\hat{y}^{n+1} - \mu^n | F^n] = Var[\theta_{x^n} + \varepsilon^{n+1} | F^n] = \lambda_{x^n} + \Sigma^n_{x^n x^n}$.

Define random variables $(Z^n)_{n=1}^N$, $Z^{n+1} := (\hat{y}^{n+1} - \mu^n) / \sqrt{Var[\hat{y}^{n+1} - \mu^n | F^n]}$.

So $\mu^{n+1} = \mu^n + \tilde{\sigma}(\Sigma^n, x^n) Z^{n+1}$ (8)

$\Sigma^{n+1} = \Sigma^n - \tilde{\sigma}(\Sigma^n, x^n)(\tilde{\sigma}(\Sigma^n, x^n))' = \Sigma^n - Cov[\mu^{n+1} | F^n]$.

$(V^n)_n$ ~ a sequence of value functions. one for each time n.

$V^n : S \to \mathbb{R}$.

$V^n(s) := \sup_{\pi \in \Pi} E^\pi [\max_i \mu_i^N | S^n = s]$ for every $s \in S$.

Resulting expression:

$V^N(s) = \max_{x \in [1, \cdots M]} \mu_x$. for every $s = (\mu, \Sigma) \in S$.

$0 \le n < N$

$V^n(s) = \max_{x \in [1, \cdots M]} E[V^{n+1}(S^{n+1}) | S^n = s, x^n = x]$ for every $s \in S$. (9)

Define Q factors. $Q^n : S \times \{1, \cdots, M\} \mapsto \mathbb{R}$, as

$Q^n(s, x) := E[V^{n+1}(S^{n+1}) | S^n = s, x^n = x]$. for every $s \in S$. ~~(9)~~

$V^{\pi, n}(s) := E^\pi [V^N(S^N) | S^n = s]$ for every $s \in S$.

π is said to be optimal if $V^n(s) = V^{\pi, n}(s)$ for every $s \in S$ and $n \le N$. (10)

π^*:

$X^{\pi^*, n}(s) \in \arg\max_{x \in [1, \cdots, M]} Q^n(s, x)$.

for every $s \in S$, $n < N$, and $x \in \{1, \cdots, M\}$. is optimal.

Define KG Policy π^{KG}:

$X^{KG}(S) \in \arg\max_{x} E_n[\max_i \mu_i^{n+1} | S^n = S, x^n = x] - \max_i \mu_i^n$,

①. $N=1$, KG policy meets the requirements of (1).

② KG policy is the only stationary myopically optimal policy.

$$X^{KG}(S^n) \in \arg\max_{x} \tilde{\sigma}_x(\Sigma^n, x) f\left(\frac{-|\mu_x^n - \max_{i \neq x} \mu_i^n|}{\tilde{\sigma}_x(\Sigma^n, x)}\right).$$

$- f(z) := \varphi(z) + z\Phi(z)$

φ: normal probability density function.

Φ : normal cumulative distribution function.

If Σ^n is diagonal, then $\tilde{\sigma}_x(\Sigma^n, x) = \Sigma_{xx}^n / \sqrt{\lambda_x + \Sigma_{xx}^n}$

In general case, Σ^n is not diagonal.

$X^{KG}(S^n)$

$= \arg\max_{x} \mathbb{E}[\max_i \mu_i^n + \tilde{\sigma}_i(\Sigma^n, x^n) Z^{n+1} | S^n, x^n = x] - \max_i \mu_i^n$

$= \arg\max_{x} h(\mu^n, \tilde{\sigma}(\Sigma^n, x))$.

$h: \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}$ defined by $h(a, b) = E[\max_i a_i + b_i Z] - \max_i a_i$.