Lecture 6

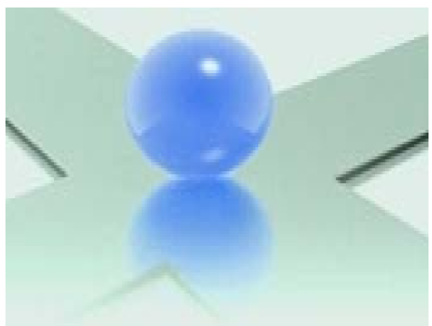
The Optimizer's Curse:

An Info-Gap Response

Yakov Ben-Haim

Technion

Israel Institute of Technology



[0]lectures\talks\lib\tufts2025lec06-001.tex 25.12.2024

Contents

1 Probabilistic Analysis

[0] lectures\talks\lib\optim-curse02tufts.tex 25.12.2024

§ N **alternatives:** $1, \ldots, n$.

- v_i = **Unknown true value of** i**th option.**

  $v = (v_1, \ldots, v_n)^T$.

- V_i = **Known estimate of** i**th option.**

  $V = (V_1, \ldots, V_n)^T$.

§

§ N **alternatives:** $1, \ldots, n$.

- v_i = **Unknown true value of** i**th option.**

  $v = (v_1, \ldots, v_n)^T$.

- V_i = **Known estimate of** i**th option.**

  $V = (V_1, \ldots, V_n)^T$.

§ **Regret:**

- **Choose alternative** i, **expecting** V_i.
- **Obtain realized outcome** y_i.
- **Regret, or disappointment:** $V_i - y_i$.

  Positive regret if $y_i < V_i$.

§ Unbiased estimates:

$$E(V_i|v) = v_i \tag{1}$$

For any i, the expected regret is zero:

$$E(V_i - y_i|v) = 0 \tag{2}$$

This is because:

$$E(V_i|v) = v_i = E(y_i|v) \tag{3}$$

§

§ Unbiased estimates:

$$\mathrm{E}(V_i|v) = v_i \tag{4}$$

For any i, the expected regret is zero:

$$\mathrm{E}(V_i - y_i|v) = 0 \tag{5}$$

This is because:

$$\mathrm{E}(V_i|v) = v_i = \mathrm{E}(y_i|v) \tag{6}$$

§ Questions:

- **Should one optimize on the estimates V_i?**
- **What is the expected regret if you do?**

§ Best-model optimization

(choose most promising alternative):

$$i^\star = \arg\max_i V_i \qquad (7)$$

§

§ Best-model optimization

(choose most promising alternative):

$$i^{\star} = \arg\max_{i} V_i \tag{8}$$

§ Expect positive regret from $V_{i^{\star}}$.

This is the optimizer's curse.

§ Example:

- **Suppose** $\mathrm{E}(v_i) = \mu$ **for all** i**.**
-

§ Example:

- **Suppose** $\mathrm{E}(v_i) = \mu$ **for all** i**.**
- $\mathrm{E}(V_{i^\star}) > \mu$ **since:**
  - $V_{i^\star}$ **is the maximum of** n **estimates.**
  - $V_{i^\star}$ **will tend to be on upper tail.**

    Example: best grade of n **exams.**
  - **Hence** $\mathrm{E}(V_{i^\star} - y_{i^\star}) = \mathrm{E}(V_{i^\star}) - \mu > 0$**.**
-

§ Example:

- **Suppose** $\mathrm{E}(v_i) = \mu$ **for all** i**.**
- $\mathrm{E}(V_{i^\star}) > \mu$ **since:**
  - $V_{i^\star}$ **is the maximum of** n **estimates.**
  - $V_{i^\star}$ **will tend to be on upper tail.**

    Example: best grade of n **exams.**
  - **Hence** $\mathrm{E}(V_{i^\star} - y_{i^\star}) = \mathrm{E}(V_{i^\star}) - \mu > 0$**.**
- **On average, best-estimate optimum:**
  - **Is over-estimate.**
  - **Has positive regret.**
  - **This is the optimizer's curse.**

2 Info-Gap Analysis

§ Related material.[1]

[1] "Lecture Notes on Robust-Satisficing Behavior", section 6: Probability of Success. File: lectures\risk\lectures\rsb02.tex.

2.1 Introduction

§ Question:

Since $V_{i^\star}$ is an unreliable estimate, what should we do?

§ Potential answer. Bayesian analysis

(Smith and Winkler, 2006):

- **Posit prior probabilities for v and conditional probabilities for V given v.**
-

§ Potential answer. Bayesian analysis

(Smith and Winkler, 2006):

- **Posit prior probabilities for v and conditional probabilities for V given v.**
- **Use Bayes' rule to determine posterior probability of v given V.**
-

§ Potential answer. Bayesian analysis

(Smith and Winkler, 2006):

- **Posit prior probabilities for v and conditional probabilities for V given v.**
- **Use Bayes' rule to determine posterior probability of v given V.**
- **Choose alternative based on posterior means, $\mathrm{E}(v_i|V)$:**

$$i^{\star} = \arg\max_{i} \mathrm{E}(v_i|V) \tag{9}$$

-

§ Potential answer. Bayesian analysis

(Smith and Winkler, 2006):

- **Posit prior probabilities for v and conditional probabilities for V given v.**
- **Use Bayes' rule to determine posterior probability of v given V.**
- **Choose alternative based on posterior means, $\mathrm{E}(v_i|V)$:**

$$i^\star = \arg\max_i \mathrm{E}(v_i|V) \qquad (10)$$

- **Smith and Winkler (2006) show that this solution does not have the optimizer's curse!**
-

§ Potential answer. Bayesian analysis

(Smith and Winkler, 2006):

- **Posit prior probabilities for v and conditional probabilities for V given v.**
- **Use Bayes' rule to determine posterior probability of v given V.**
- **Choose alternative based on posterior means, $\mathrm{E}(v_i|V)$:**

$$i^{\star} = \arg\max_{i} \mathrm{E}(v_i|V) \tag{11}$$

- **Smith and Winkler (2006) show that this solution does not have the optimizer's curse!**
- **The problem: where do you get these pdf's?**

§ Potential answer. Info-gap robust-satisficing:

- **Satisfice (don't try to maximize) the value: $v_i \geq V_{\text{c}}$.**

(We will find the regret entering later.)

- **Maximize the robustness.**

§

§ Potential answer. Info-gap robust-satisficing:

- **Satisfice (don't try to maximize) the value:** $v_i \geq V_{\text{c}}$.

(We will find the regret entering later.)

- **Maximize the robustness.**

§ Potential answer. Info-gap opportune-windfalling:

- **Windfall the value:**

$v_i \geq V_{\text{w}}$ **where** $V_{\text{w}} \gg V_{\text{c}}$.

- **Maximize the opportuneness.**

§

§ Potential answer. Info-gap robust-satisficing:

- **Satisfice (don't try to maximize) the value: $v_i \geq V_c$.**

(We will find the regret entering later.)

- **Maximize the robustness.**

§ Potential answer. Info-gap opportune-windfalling:

- **Windfall the value:**

$v_i \geq V_w$ **where** $V_w \gg V_c$.

- **Maximize the opportuneness.**

§ We will explore:

- **Robust-satisficing.**
- **(Proxy theorems.)**

2.2 Robustness: Formulation

§ Observations:

known estimated values of n alternatives: $V = (V_1, \ldots, V_n)^T$**.**

§

§ Observations:

known estimated values of n alternatives: $V = (V_1, \ldots, V_n)^T$.

§ Uncertainty:

- **Unknown true values of n alternatives:** $v = (v_1, \ldots, v_n)^T$.
-

§ Observations:

known estimated values of n alternatives: $V = (V_1, \ldots, V_n)^T$.

§ Uncertainty:

- **Unknown true values of n alternatives:** $v = (v_1, \ldots, v_n)^T$.
- $\mathcal{V}(h)$ = **info-gap model for** v. **E.g.:**

$$\mathcal{V}(h) = \left\{ v : \left| \frac{v_i - V_i}{s_i} \right| \le h, \ \forall i \right\}, \quad h \ge 0 \tag{12}$$

Or:

§ Observations:

known estimated values of n alternatives: $V = (V_1, \ldots, V_n)^T$.

§ Uncertainty:

- **Unknown true values of n alternatives:** $v = (v_1, \ldots, v_n)^T$.
- $\mathcal{V}(h) =$ **info-gap model for** v**. E.g.:**

$$\mathcal{V}(h) = \left\{ v : \ \left| \frac{v_i - V_i}{s_i} \right| \le h, \ \forall i \right\}, \quad h \ge 0 \tag{13}$$

Or:

$$\mathcal{V}(h) = \left\{ v : \ (v - V)^T S^{-1} (v - V) \le h^2 \right\}, \quad h \ge 0 \tag{14}$$

§ **Decision:** r **is the decision vector. E.g.:**

- **A standard unit basis vector, selecting a single alternative.**
-

§ **Decision:** r **is the decision vector. E.g.:**

- **A standard unit basis vector, selecting a single alternative.**
- **An** n**-vector probability distribution selecting a randomized mix of alternatives.**

§

§ **Decision:** r **is the decision vector. E.g.:**

- **A standard unit basis vector,**
  selecting a single alternative.
- **An** n**-vector probability distribution**
  selecting a randomized mix of alternatives.

§ **Performance function. Value:**

$$G(r, v) = r^T v \tag{15}$$

§

§ **Decision:** r **is the decision vector. E.g.:**

- **A standard unit basis vector, selecting a single alternative.**
- **An n-vector probability distribution selecting a randomized mix of alternatives.**

§ **Performance function. Value:**

$$G(r, v) = r^T v \tag{16}$$

§ **Performance requirement. Satisfice the value:**

$$G(r, v) \geq G_{\mathrm{c}} \tag{17}$$

§

§ **Decision:** r **is the decision vector. E.g.:**

- **A standard unit basis vector, selecting a single alternative.**
- **An n-vector probability distribution selecting a randomized mix of alternatives.**

§ **Performance function. Value:**

$$G(r, v) = r^T v \tag{18}$$

§ **Performance requirement. Satisfice the value:**

$$G(r, v) \geq G_{\mathrm{c}} \tag{19}$$

§ **Robustness:**

$$\widehat{h}(r, G_{\mathrm{c}}) = \max \left\{ h : \ \left(\min_{v \in \mathcal{V}(h)} r^T v \right) \geq G_{\mathrm{c}} \right\} \tag{20}$$

2.3 Robustness: Fractional-Error Uncertainty

§ Evaluate the robustness with this info-gap model:

$$\mathcal{V}(h) = \left\{ v : \ \left| \frac{v_i - V_i}{s_i} \right| \le h, \ \forall i \right\}, \quad h \ge 0 \tag{21}$$

§

§ **Evaluate the robustness with this info-gap model:**

$$\mathcal{V}(h) = \left\{ v : \left| \frac{v_i - V_i}{s_i} \right| \le h, \ \forall i \right\}, \quad h \ge 0 \tag{22}$$

§ $\mu(h)$: **inner minimum in definition of robustness:**

- $\mu(h) = \min_{v \in \mathcal{V}(h)} r^T v$.
-

§ **Evaluate the robustness with this info-gap model:**

$$\mathcal{V}(h) = \left\{ v : \ \left| \frac{v_i - V_i}{s_i} \right| \le h, \ \forall i \right\}, \quad h \ge 0 \tag{23}$$

§ $\mu(h)$**: inner minimum in definition of robustness:**

- $\mu(h) = \min_{v \in \mathcal{V}(h)} r^T v$.
- **The elements of** r **are non-negative,**
  so $\mu(h)$ **occurs when each** v_i **is minimal:**

$$\mu(h) = \sum_{i=1}^{n} (V_i - s_i h) r_i \tag{24}$$

$$= r^T V - h r^T s \tag{25}$$

-

§ **Evaluate the robustness with this info-gap model:**

$$\mathcal{V}(h) = \left\{ v : \left| \frac{v_i - V_i}{s_i} \right| \le h, \ \forall i \right\}, \quad h \ge 0 \tag{26}$$

§ $\mu(h)$: **inner minimum in definition of robustness:**

- $\mu(h) = \min_{v \in \mathcal{V}(h)} r^T v$.
- **The elements of r are non-negative,**

  so $\mu(h)$ occurs when each v_i is minimal:

$$\mu(h) = \sum_{i=1}^{n} (V_i - s_i h) r_i \tag{27}$$

$$= r^T V - h r^T s \tag{28}$$

- **Equate to G_c, solve for h yields robustness:**

$$\widehat{h}(r, G_c) = \frac{r^T V - G_c}{r^T s} \tag{29}$$

  or zero if this is negative.

$$\widehat{h}(r, G_{\rm c}) = \frac{r^T V - G_{\rm c}}{r^T s} \qquad (30)$$

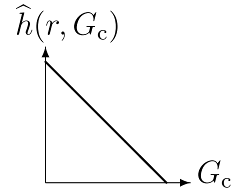
§ The numerator is a regret:

- **Regret: outcome lower than expectation.**
- $r^T V$**: expected outcome.**
- **Outcome** $G_{\rm c}$ **would cause regret** $r^T V - G_{\rm c}$**.**
-

$$\widehat{h}(r, G_c) = \frac{r^T V - G_c}{r^T s} \tag{31}$$

§ **The numerator is a regret:**

- **Regret: outcome lower than expectation.**
- $r^T V$**: expected outcome.**
- **Outcome** G_c **would cause regret** $r^T V - G_c$**.**
- **Zero regret has zero robustness.**
- **Positive regret has positive robustness.**



§ Robustness curve:

- **Trade-off:** robustness vs loss.
- **Zeroing:** no robustness of anticipated regret.
- **Cost of robustness:** slope.

$$\widehat{h}(r, G_{\rm c}) = \frac{r^T V - G_{\rm c}}{r^T s} \tag{32}$$

§ Preference reversal:

Robustness curves of different decisions can cross one another.

$$\widehat{h}(r, G_{\rm c}) = \frac{r^T V - G_{\rm c}}{r^T s} \qquad (33)$$

§ **Dilemma: Choose between 2 decisions,** r_1**,** r_2**:**

$$r_1^T V < r_2^T V \implies r_1: \ \textbf{less estimated return} \qquad (34)$$

$$\implies r_1: \ \textbf{(less estimated regret)} \qquad (35)$$

$$r_1^T s < r_2^T s \implies r_1: \ \textbf{less uncertain} \qquad (36)$$

$$\widehat{h}(r, G_c) = \frac{r^T V - G_c}{r^T s} \tag{37}$$

§ **Dilemma: Choose between 2 decisions,** r_1**,** r_2**:**

$$r_1^T V < r_2^T V \implies r_1 : \textbf{ less estimated return} \tag{38}$$

$$\implies r_1 : \textbf{ (less estimated regret)} \tag{39}$$

$$r_1^T s < r_2^T s \implies r_1 : \textbf{ less uncertain} \tag{40}$$

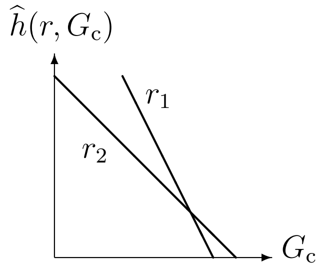


Figure 1: **ROBUSTNESS CURVES**.

§ **Nominal optimum:** r_2**.**

§

$$\widehat{h}(r, G_c) = \frac{r^T V - G_c}{r^T s} \tag{41}$$

§ **Dilemma: Choose between 2 decisions,** r_1, r_2**:**

$$r_1^T V < r_2^T V \quad \Longrightarrow \quad r_1 : \ \textbf{less estimated return} \tag{42}$$

$$\Longrightarrow \quad r_1 : \ \textbf{(less estimated regret)} \tag{43}$$

$$r_1^T s < r_2^T s \quad \Longrightarrow \quad r_1 : \ \textbf{less uncertain} \tag{44}$$

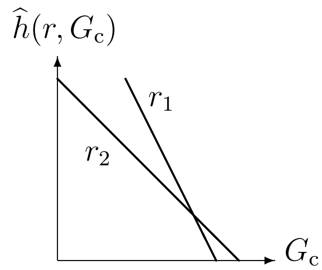


Figure 2: **ROBUSTNESS CURVES**.

§ **Nominal optimum:** r_2**.**

§ **Robust satisficing optimum for smaller** G_c**:** r_1**.**

2.4 Robustness: Ellipsoidal Uncertainty

§ **Evaluate the robustness with the info-gap model:**

$$\mathcal{V}(h) = \left\{ v : (v - V)^T S^{-1} (v - V) \le h^2 \right\}, h \ge 0 \tag{45}$$

§

§ **Evaluate the robustness with the info-gap model:**

$$\mathcal{V}(h) = \left\{ v : (v - V)^T S^{-1} (v - V) \leq h^2 \right\}, h \geq 0 \tag{46}$$

§ $\mu(h)$: **inner minimum in definition of the robustness.**

- $\mu(h) = \min_{v \in \mathcal{V}(h)} r^T v$.
-

§ **Evaluate the robustness with the info-gap model:**

$$\mathcal{V}(h) = \left\{ v : (v-V)^T S^{-1} (v-V) \le h^2 \right\}, h \ge 0 \tag{47}$$

§ $\mu(h)$: **inner minimum in definition of the robustness.**

- $\mu(h) = \min_{v \in \mathcal{V}(h)} r^T v$.
- **Lagrange optimization:**

$$H = r^T v + \lambda \left[h^2 - (v-V)^T S^{-1} (v-V) \right] \tag{48}$$

$$0 = \frac{\partial H}{\partial v} = r - 2\lambda S^{-1}(v-V) \tag{49}$$

$$\Longrightarrow \quad v - V = \frac{1}{2\lambda} S r \tag{50}$$

$$h^2 = \frac{1}{4\lambda^2} (Sr)^T S^{-1} (Sr) = \frac{1}{4\lambda^2} r^T S r \tag{51}$$

$$\Longrightarrow \quad \mu(h) = r^T V - h\sqrt{r^T S r} \quad \ge \quad G_{\mathrm{c}} \tag{52}$$

$$\Longrightarrow \quad \widehat{h}(r, G_{\mathrm{c}}) = \frac{r^T V - G_{\mathrm{c}}}{\sqrt{r^T S r}} \tag{53}$$

or zero if this is negative.

§

§ Evaluate the robustness with the info-gap model:

$$\mathcal{V}(h) = \left\{ v : (v - V)^T S^{-1} (v - V) \le h^2 \right\}, h \ge 0 \tag{54}$$

§ $\mu(h)$: inner minimum in definition of the robustness.

- $\mu(h) = \min_{v \in \mathcal{V}(h)} r^T v$.
- **Lagrange optimization:**

$$H = r^T v + \lambda \left[h^2 - (v - V)^T S^{-1} (v - V) \right] \tag{55}$$

$$0 = \frac{\partial H}{\partial v} = r - 2\lambda S^{-1} (v - V) \tag{56}$$

$$\Longrightarrow \quad v - V = \frac{1}{2\lambda} S r \tag{57}$$

$$h^2 = \frac{1}{4\lambda^2} (Sr)^T S^{-1} (Sr) = \frac{1}{4\lambda^2} r^T S r \tag{58}$$

$$\Longrightarrow \quad \mu(h) = r^T V - h \sqrt{r^T S r} \quad \ge \quad G_{\rm c} \tag{59}$$

$$\Longrightarrow \quad \widehat{h}(r, G_{\rm c}) = \frac{r^T V - G_{\rm c}}{\sqrt{r^T S r}} \tag{60}$$

or zero if this is negative.

§ Zeroing and Trade off.

2.5 Probability of Success and the Proxy Property

§ Probability of success:

- **Define** $q = r^T v$**.**
-

§ Probability of success:

- **Define** $q = r^T v$.
- **Requirement:** $q \geq G_c$.
-

§ Probability of success:

- **Define** $q = r^T v$**.**
- **Requirement:** $q \geq G_{\text{c}}$**.**
- $p(q|r) =$ **pdf of** q **given** r**, unknown.**
-

§ Probability of success:

- **Define** $q = r^T v$**.**
- **Requirement:** $q \geq G_{\rm c}$**.**
- $p(q|r) =$ **pdf of** q **given** r**, unknown.**
- $P_{\rm s}(r, G_{\rm c}) =$ **probability of satisfying the requirement with choice-vector** r**:**

$$P_{\rm s}(r, G_{\rm c}) = \text{Prob}(q \geq G_{\rm c}) = \int_{G_{\rm c}}^{\infty} p(q|r)\,{\rm d}q \tag{61}$$

§ Probabilistic preferences:

$$r_1 \succ_{\rm p} r_2 \quad \textbf{if} \quad P_{\rm s}(r_1, G_{\rm c}) > P_{\rm s}(r_2, G_{\rm c}) \tag{62}$$

§

§ Probabilistic preferences:

$$r_1 \succ_{\rm p} r_2 \quad \textbf{if} \quad P_{\rm s}(r_1, G_{\rm c}) > P_{\rm s}(r_2, G_{\rm c}) \tag{63}$$

§ Robust-satisficing preferences:

$$r_1 \succ_{\rm r} r_2 \quad \textbf{if} \quad \widehat{h}(r_1, G_{\rm c}) > \widehat{h}(r_2, G_{\rm c}) \tag{64}$$

§

§ Probabilistic preferences:

$$r_1 \succ_{\rm p} r_2 \quad \textbf{if} \quad P_{\rm s}(r_1, G_{\rm c}) > P_{\rm s}(r_2, G_{\rm c}) \tag{65}$$

§ Robust-satisficing preferences:

$$r_1 \succ_{\rm r} r_2 \quad \textbf{if} \quad \widehat{h}(r_1, G_{\rm c}) > \widehat{h}(r_2, G_{\rm c}) \tag{66}$$

§ Proxy property:

$$\widehat{h}(r_1, G_{\rm c}) > \widehat{h}(r_2, G_{\rm c}) \quad \textbf{iff} \quad P_{\rm s}(r_1, G_{\rm c}) > P_{\rm s}(r_2, G_{\rm c}) \tag{67}$$

§

§ **Probabilistic preferences:**

$$r_1 \succ_{\rm p} r_2 \quad \textbf{if} \quad P_{\rm s}(r_1, G_{\rm c}) > P_{\rm s}(r_2, G_{\rm c}) \tag{68}$$

§ **Robust-satisficing preferences:**

$$r_1 \succ_{\rm r} r_2 \quad \textbf{if} \quad \widehat{h}(r_1, G_{\rm c}) > \widehat{h}(r_2, G_{\rm c}) \tag{69}$$

§ **Proxy property:**

$$\widehat{h}(r_1, G_{\rm c}) > \widehat{h}(r_2, G_{\rm c}) \quad \textbf{iff} \quad P_{\rm s}(r_1, G_{\rm c}) > P_{\rm s}(r_2, G_{\rm c}) \tag{70}$$

§ **Theorem: Robust is a proxy for probability of success under fairly general conditions.**

§

§ **Probabilistic preferences:**

$$r_1 \succ_{\rm p} r_2 \quad \textbf{if} \quad P_{\rm s}(r_1, G_{\rm c}) > P_{\rm s}(r_2, G_{\rm c}) \tag{71}$$

§ **Robust-satisficing preferences:**

$$r_1 \succ_{\rm r} r_2 \quad \textbf{if} \quad \widehat{h}(r_1, G_{\rm c}) > \widehat{h}(r_2, G_{\rm c}) \tag{72}$$

§ **Proxy property:**

$$\widehat{h}(r_1, G_{\rm c}) > \widehat{h}(r_2, G_{\rm c}) \quad \textbf{iff} \quad P_{\rm s}(r_1, G_{\rm c}) > P_{\rm s}(r_2, G_{\rm c}) \tag{73}$$

§ **Theorem: Robust is a proxy for probability of success under fairly general conditions.**

§ **Hence:**

- **Choosing r to maximize the robustness, $\widehat{h}(r, G_{\rm c})$, also maximizes the probability of success, $P_{\rm s}(r, G_{\rm c})$.**
- **We can maximize probability of success without knowing probability distributions!**

3 Summary

§ **Optimizer's curse: Choosing the putative optimum:**

- **Causes regret.**
- **Is unrealistic.**

§

§ **Optimizer's curse: Choosing the putative optimum:**

- **Causes regret.**
- **Is unrealistic.**

§ **Info-gap response:**

- **Satisfice the outcome (sub-optimal but good enough).**
- **Maximize the robustness to uncertainty.**

. . .

§ Optimizer's curse: Choosing the putative optimum:

- **Causes regret.**
- **Is unrealistic.**

§ Info-gap response:

- **Satisfice the outcome (sub-optimal but good enough).**
- **Maximize the robustness to uncertainty.**

Any questions?