Lecture Notes on the Optimizer's Curse

Yakov Ben-Haim
Yitzhak Moda'i Chair in Technology and Economics
Faculty of Mechanical Engineering
Technion — Israel Institute of Technology
Haifa 32000 Israel
yakov@technion.ac.il
http://info-gap.com http://www.technion.ac.il/yakov

A Note to the Student: These lecture notes are not a substitute for the thorough study of articles and books. These notes are no more than an aid in following the lectures.

§ **Sources:**

• Smith, James E. and Robert L. Winkler, 2006, The optimizer's curse: Skepticism and postdecision surprise in decision analysis, *Management Science,* Vol. 52, No. 3, pp.311–322.

• Thaler, Richard H., 1992, *The Winner's Curse: Paradoxes and Anomalies of Economic Life,* Princeton University Press.

Contents

1 Probabilistic Analysis **2**
1.1 Formulation . . . 2
1.2 Simple Examples . . . 3
1.2.1 3 Zero-Mean Alternatives . . . 3
1.2.2 n Zero-Mean Alternatives . . . 4
1.2.3 3 Different Alternatives . . . 5
1.3 Distribution of $V_{\hat{i}^\star}$ . . . 6
1.4 Optimizer's Curse Theorem . . . 8

2 Info-Gap Analysis **9**
2.1 Robustness: Formulation . . . 9
2.2 Robustness: Simple Example . . . 10
2.3 Probability of Success and the Proxy Property . . . 11
2.4 Proxy Property: Simple Examples . . . 12
2.4.1 Normal Distribution . . . 12
2.4.2 Uniform Distribution . . . 12
2.5 Standardization and the Proxy Property . . . 14
2.6 Coherence . . . 17
2.7 Coherence and the Proxy Property . . . 19

[0]lectures\risk\lectures\optimizers-curse03.tex 19.5.2010

1 Probabilistic Analysis

1.1 Formulation

§ N **alternatives:** 1, …, n.

- v_i = **Unknown** true value of ith alternative. $v = (v_1, \ldots, v_n)^T$.
- V_i = **Known** estimated value of ith alternative. $V = (V_1, \ldots, V_n)^T$.

§ **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:**

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

Thus, for any choice i, the expected regret is zero:

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

This is because:

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

§ **Best-model optimization:**

$$i^\star = \arg\max_i V_i \tag{4}$$

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

- Example:
  - Suppose $\mathrm{E}(v_i) = \mu$ for all i.
  - Anticipate $\mathrm{E}(V_{i^\star}) > \mu$ 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 $\mathrm{E}(V_{i^\star} - y_{i^\star}) = \mathrm{E}(V_{i^\star}) - \mu > 0$.
- Meaning: On average, best-estimate optimum:
  - **Is over-estimate.**
  - **Has positive regret.**
- We will explore this more deeply later.

1.2 Simple Examples

§ We consider some simple examples from Smith and Winkler (2006).

1.2.1 3 Zero-Mean Alternatives

§ The true values, v_i, are all precisely equal zero. They are not random variables.

§ The estimates, V_i, are all $\mathcal{N}(0, 1)$. See fig. 1.

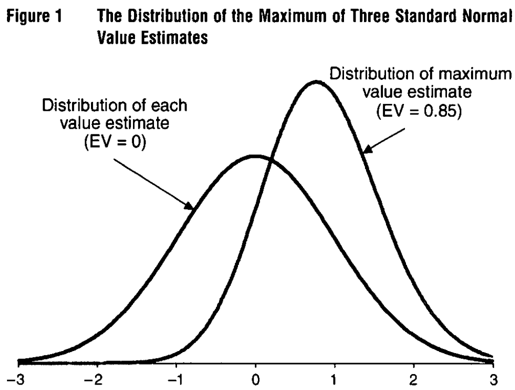


Figure 1: Smith and Winkler (2006), fig. 1.

§ The mean of the distribution of $V_{i^\star}$ is 0.85.
(We will understand this more deeply later.)

§ Thus the average regret, $\mathrm{E}(V_{i^\star} - 0)$, is 0.85.

§ **More generally,** suppose:

- The true values are $v_i = \mu$ for all i. They are not random variables.
- The estimates, V_i, are all $\mathcal{N}(\mu, \sigma^2)$.
- Then $\mathrm{E}(v_{i^\star}) = 0.85\sigma$ which is the average regret.

1.2.2 n Zero-Mean Alternatives

§ If:

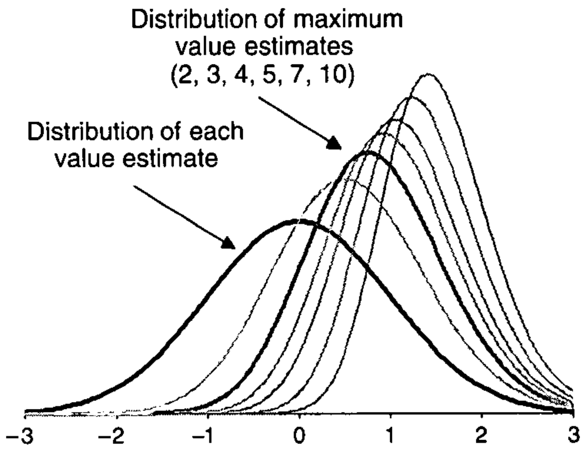
- $v_i = 0$ for all $i = 1, \ldots, n$ (not random variable).
- $V_i \sim \mathcal{N}(0, 1)$ for all $i = 1, \ldots, n$.

§ Then the regret increases as n increases. See fig. 2.

This makes sense:

- ○ $V_{\hat{i}^\star}$ is the maximum of n estimates.
- ○ This maximum tends to increase as n increases.

Figure 2 The Distribution of the Maximum of *n* Standard Normal Value Estimates



Number of alternatives	Expected disappointment
1	0.00
2	0.56
3	0.85
4	1.03
5	1.16
6	1.27
7	1.35
8	1.43
9	1.48
10	1.54

Figure 2: Smith and Winkler (2006), fig. 2.

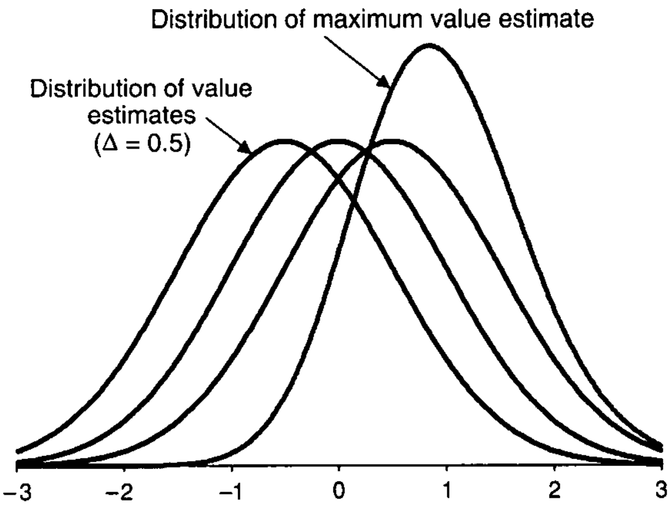
1.2.3 3 Different Alternatives

§ The true values are $v_i = -\Delta,\ 0,\ \Delta$. Not random variables.

§ The estimates are unbiased normal with unit standard deviation: $V_i \sim \mathcal{N}(v_i, 1)$.

§ As the alternatives become more different, we should expect $V_{i^\star}$ to become a better bet. See fig. 3.

Figure 3 The Distribution of Maximum Value Estimates with Separation Between Alternatives



Δ	Expected disappointment	Probability of correct choice
0.0	0.85	0.33
0.2	0.66	0.42
0.4	0.51	0.50
0.6	0.39	0.59
0.8	0.30	0.66
1.0	0.22	0.73
1.2	0.17	0.78
1.4	0.12	0.83
1.6	0.10	0.87
1.8	0.07	0.90
2.0	0.05	0.92
2.2	0.03	0.94
2.4	0.02	0.95
2.6	0.01	0.97
2.8	0.01	0.98
3.0	0.00	0.98

Figure 3: Smith and Winkler (2006), fig. 3.

1.3 Distribution of $V_{i^\star}$

§ In this section we derive and study the distribution of $V_{i^\star}$.

- We will understand why its mean exceeds $\mathrm{E}(V_i)$.
- Source: DeGroot, Morris H., 1986, *Probability and Statistics,* 2nd ed., Addison-Wesley, Reading, MA. Section 3.2, pp.182–183.

§ V_i is the estimated value of the ith alternative.

- Its cumulative probability distribution (cpd) is $F_i(y)$.
- All the V_i are statistically independent.

§ $V_{i^\star} = \max_i V_i$.
Its cpd is $G(y)$, derived as follows:

$$\begin{aligned} G(y) &= \mathrm{Prob}(V_{i^\star} \leq y) && (5)\\ &= \mathrm{Prob}(V_1 \leq y, \ldots, V_n \leq y) && (6)\\ &= \prod_{i=1}^{n} F_i(y) && (7)\end{aligned}$$

§ If the V_i are i.i.d. with cpd $F(y)$ and pdf $f(y)$ then:

$$\begin{aligned} G(y) &= [F(y)]^n && (8)\\ g(y) &= n[F(y)]^{n-1} f(y) && (9)\end{aligned}$$

§ Now compare $\mathrm{E}(V_{i^\star})$ and $\mathrm{E}(V_i)$ for i.i.d. case:

$$\begin{aligned} \mathrm{E}(V_{i^\star}) &= \int yn[F(y)]^{n-1} f(y)\,\mathrm{d}y && (10)\\ \mathrm{E}(V_i) &= \int yf(y)\,\mathrm{d}y && (11)\end{aligned}$$

Thus:

$$\mathrm{E}(V_{i^\star}) - \mathrm{E}(V_i) = \int yf(y)\left(n[F(y)]^{n-1} - 1\right)\,\mathrm{d}y \qquad (12)$$

This integral is positive for $n \geq 2$, as we now explain intuitively.

§ Define y_n as the value for which: $n[F(y_n)]^{n-1} = 1$.

- Hence: $F(y_n) = (1/n)^{1/(n-1)}$.
- Some values:

  n: 2, 5, 10, 100.

  $(1/n)^{1/(n-1)}$: 0.5, 0.67, 0.74, 0.95.

  Thus y_n gets large as n gets large.
- Note that $n[F(y)]^{n-1} \leq 1$ iff $y \leq y_n$.
- Hence, from eq.(9), note that $g(y) \leq f(y)$ for $y \leq y_n$.
- Thus, since $g(y)$ is normalized, it is shifted to the right wrt $f(y)$.
- Thus, $\mathrm{E}(V_{i^\star}) \geq \mathrm{E}(V_i)$.

1.4 Optimizer's Curse Theorem

Theorem 1 *The expected regret from the estimated optimal alternative is non-negative.*
Given:

- $V = (V_1, \ldots, V_n)^T$ *are estimated values of* n *alternatives. These are random variables.*
- *The estimates are unbiased:* $\mathrm{E}(V_i|v) = v_i$.
- $v = (v_1, \ldots, v_n)^T$ *are true values of* n *alternatives.*
- $i^\star = \arg\max_i V_i$ *is the index of the most favorable estimate.*

Then:

$$\mathrm{E}(V_{i^\star} - v_{i^\star}|v) \geq 0 \tag{13}$$

where the expectation is with respect to V *conditioned on* v.

Proof. Define $i' = \arg\max_i v_i$. Then:

$$V_{i^\star} - v_{i^\star} \geq V_{i^\star} - v_{i'} \geq V_{i'} - v_{i'} \tag{14}$$

- The left inequality is because $v_{i'} \geq v_{i^\star}$.
- The right inequality is because $V_{i^\star} \geq V_{i'}$.

Now take expectations of eq.(14) w.r.t. V, conditioned on v:

$$\mathrm{E}(V_{i^\star} - v_{i^\star}|v) \geq \mathrm{E}(V_{i^\star} - v_{i'}|v) \geq 0 \tag{15}$$

- The 0 on the right is because the estimates are unbiased: $\mathrm{E}(V_{i'} - v_{i'}|v) = 0$.
- Eq.(15) implies eq.(13). ∎

2 Info-Gap Analysis

§ Related material in "Lecture Notes on Robust-Satisficing Behavior", section 6: Probability of Success. File: lectures\risk\lectures\rsb02.tex.

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

§ **A 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, $\mathrm{E}(v_i|V)$:

$$i^\star = \arg\max_i \mathrm{E}(v_i|V) \tag{16}$$

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

§ **A potential answer. Info-gap robust-satisficing:**

- Satisfice the value: $v_i \geq V_{\mathrm{c}}$. (We will find the regret entering later.)
- Maximize the robustness.

§ **A potential answer. Info-gap opportune-windfalling:**

- Windfall the value: $v_i \geq V_{\mathrm{w}}$ where $V_{\mathrm{w}} \gg V_{\mathrm{c}}$.
- Maximize the opportuneness.

§ **We will explore:**

- Robust-satisficing.
- Proxy theorems.

2.1 Robustness: Formulation

§ **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.:

$$\mathcal{V}(h) = \left\{ v : \left| \frac{v_i - V_i}{s_i} \right| \leq h, \ \forall i \right\}, \quad h \geq 0 \tag{17}$$

Or:

$$\mathcal{V}(h) = \left\{ v : \ (v - V)^T S^{-1} (v - V) \le h^2 \right\}, \quad h \ge 0 \tag{18}$$

§ **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:

$$G(r, v) = r^T v \tag{19}$$

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

$$G(r, v) \ge G_{\mathrm{c}} \tag{20}$$

§ **Robustness:**

$$\widehat{h}(r, G_{\mathrm{c}}) = \max \left\{ h : \ \left(\min_{v \in \mathcal{V}(h)} r^T v \right) \ge G_{\mathrm{c}} \right\} \tag{21}$$

2.2 Robustness: Simple Example

§ We evaluate the robustness, eq.(21), with the info-gap model of eq.(17).

§ Let $\mu(h)$ denote the inner minimum in eq.(21).

- $\mu(h)$ occurs when $r^T v$ is minimal.
- The elements of r are non-negative, so $\mu(h)$ occurs when each v_i is minimal:

$$\mu(h) = \sum_{i=1}^{n} (V_i - s_i h) r_i \tag{22}$$

$$= r^T V - h r^T s \tag{23}$$

Equating this to G_{c} and solving for h yields the robustness:

$$\widehat{h}(r, G_{\mathrm{c}}) = \frac{r^T V - G_{\mathrm{c}}}{r^T s} \tag{24}$$

or zero if this is negative.

§ **Regret.** The numerator in eq.(24) is a regret:

- Regret: outcome lower than expectation.
- $r^T V$: expected outcome.
- Outcome G_{c} would cause regret $r^T V - G_{\mathrm{c}}$.
- **Zero regret** has **zero robustness.**
- **Positive regret** has **positive robustness.**

§ **Preference reversal.** It is evident from eq.(24) that robustness curves of different decisions can cross one another.

2.3 Probability of Success and the Proxy Property

§ **Probability of success:**

- Define $q = r^T v$.
- Requirement: $q \geq G_c$.
- $p(q|r)$ = pdf of q given r, which is unknown.
- $P_s(r, G_c)$ = probability of satisfying the requirement with r:

$$P_s(r, G_c) = \text{Prob}(q \geq G_c) = \int_{G_c}^{\infty} p(q|r) \, \mathrm{d}q \tag{25}$$

§ **Probabilistic preferences:**

$$r_1 \succ_p r_2 \quad \text{if} \quad P_s(r_1, G_c) > P_s(r_2, G_c) \tag{26}$$

§ **Robust-satisficing preferences:**

$$r_1 \succ_r r_2 \quad \text{if} \quad \widehat{h}(r_1, G_c) > \widehat{h}(r_2, G_c) \tag{27}$$

§ **Proxy Property:**

Definition 1 *$\mathcal{Q}_r(h)$ and $P(q|r)$ have the **proxy property** at decisions r_1 and r_2 and critical value G_c, with performance function $G(r, q)$, if:*

$$\widehat{h}(r_1, G_c) > \widehat{h}(r_2, G_c) \quad \textit{if and only if} \quad P_s(r_1, G_c) > P_s(r_2, G_c) \tag{28}$$

- The proxy property is symmetric between robustness and probability of success.
- We are particularly interested in the implication from robustness to probability.
- Thus, when the proxy property holds we will sometimes say that robustness is a proxy for probability of success.

§ **Proxy theorem:** The proxy property holds if and only if the info-gap model and the probability distribution are "coherent". We will return to the idea of coherence in section 2.6.

2.4 Proxy Property: Simple Examples

§ Before discussion coherence we examine simple examples of the proxy property, based on the simple example in section 2.2.

2.4.1 Normal Distribution

§ **Let** $q = r^T v$ **be normal**:

$$q \sim \mathcal{N}[r^T V,\ (r^T s)c] \tag{29}$$

where $c > 0$.

§ **The probability of success,** eq.(25), is:

$$\begin{aligned} P_{\mathrm{s}}(r, G_{\mathrm{c}}) &= \mathrm{Prob}(q \geq G_{\mathrm{c}}) & (30)\\ &= \mathrm{Prob}\left(\frac{q - r^T V}{(r^T s)c} \geq \frac{G_{\mathrm{c}} - r^T V}{(r^T s)c}\right) & (31)\\ &= 1 - \Phi\left(\frac{G_{\mathrm{c}} - r^T V}{(r^T s)c}\right) & (32)\\ &= 1 - \Phi\left(-\frac{\widehat{h}(r, G_{\mathrm{c}})}{c}\right) & (33) \end{aligned}$$

where eq.(33) results from eq.(24). $\Phi(\cdot)$ is the cpd of the standard normal variable.

§ **Proxy property holds.**

- From eq.(33) we see that $P_{\mathrm{s}}(r, G_{\mathrm{c}})$ depends on r only through $\widehat{h}(r, G_{\mathrm{c}})$.
- Hence eq.(33) implies eq.(28) and the proxy property holds.

2.4.2 Uniform Distribution

§ **Define uniform distributions as:**

$$p(y|a, b) = \begin{cases} \dfrac{1}{b - a} & \text{if } a \leq y \leq b \\ 0 & \text{else} \end{cases} \tag{34}$$

§ **Suppose** $q = r^T v$ **is uniform,** $p(q|a, b)$, where:

$$\begin{aligned} a &= r^T V - \frac{c}{2} r^T s & (35)\\ b &= r^T V + \frac{c}{2} r^T s & (36) \end{aligned}$$

where $c > 0$.

§ **Probability of success,** as in eqs.(30) and (31), is:

$$P_{\rm s}(r, G_{\rm c}) = {\rm Prob}(q \geq G_{\rm c}) \tag{37}$$

$$= {\rm Prob}\left(\frac{q - r^T V}{(r^T s)c} \geq \frac{G_{\rm c} - r^T V}{(r^T s)c}\right) \tag{38}$$

§ **Define:**

$$z = \frac{q - r^T V}{(r^T s)c} \tag{39}$$

which is uniform, $p(z|a, b)$, with:

$$a = -\frac{c}{2} \tag{40}$$

$$b = \frac{c}{2} \tag{41}$$

§ **Now probability of success** is analogous to eqs.(32) and (33):

$$P_{\rm s}(r, G_{\rm c}) = {\rm Prob}\left(z \geq \frac{G_{\rm c} - r^T V}{(r^T s)c}\right) \tag{42}$$

$$= 1 - P\left(\frac{G_{\rm c} - r^T V}{(r^T s)c} | a, b\right) \tag{43}$$

$$= 1 - P\left(-\frac{\widehat{h}(r, G_{\rm c})}{c} | a, b\right) \tag{44}$$

where a and b are independent of r, eqs.(40) and (41).

§ **Proxy property holds.**

- From eq.(44) we see that $P_{\rm s}(r, G_{\rm c})$ depends on r only through $\widehat{h}(r, G_{\rm c})$.
- Hence eq.(44) implies eq.(28) and the proxy property holds.

2.5 Standardization and the Proxy Property

§ **Probability of survival.**

• Option i succeeds (survives) if its value is no less than the critical value:

$$v_i \geq v_c \tag{45}$$

• $F_i(\cdot)$ denotes the cumulative probability distribution function of v_i.

• Probability of success for option i is:

$$P_s(i) \;=\; \text{Prob}(v_i \geq v_c) = 1 - F_i(v_c) \tag{46}$$

§ **Standardization class of probability distributions:**

Definition 2 *Let q be a scalar random variable with a pdf which depends on parameters r. The pdf is* ***standardizable*** *and $\theta(q, r)$ is a* ***standardization function*** *if $\theta(q, r)$ is a scalar function which is strictly increasing and continuous in q at any fixed r and whose pdf is the same for all r.*

§ **Example:**

• $f(q|r)$ is a pdf of a random variable q, where r is a vector of parameters of the pdf.

• $f(q|r)$ is a class of pdfs parametrized by r.

• Mean and variance of q are μ_q and σ_q^2. E.g. $r = (\mu_q, \sigma_q^2)$.

• Standardized random variable, with pdf $g(\theta)$, is:

$$\theta = (q - \mu_q)/\sigma_q \tag{47}$$

• If $g(\theta)$ is independent of q then this is a standardization class. That is, if all the standardized random variables in the class have the same pdf, then this is a standardization class.

• Standardization classes are quite common:

◦ the normal, uniform, and exponential distributions all being examples.

◦ The standardized distribution $g(\theta)$ may belong to the standardization class, e.g. normal and uniform, but this is not necessarily true, e.g. the exponential.

¶ **Example: exponential distribution:**

$$f(q|r) = r\mathrm{e}^{-rq}, \quad q \geq 0 \tag{48}$$

Moments:

$$\mathrm{E}(q|r) = \sigma(q|r) = \frac{1}{r} \tag{49}$$

Standardized variable:

$$\theta = \frac{q - \mathrm{E}(q|r)}{\sigma(q|r)} = rq - 1 \tag{50}$$

Standardized density by probability balance:

$$q = \frac{\theta + 1}{r}, \quad \mathrm{d}q = \frac{1}{r}\mathrm{d}\theta \quad \Longrightarrow \quad g(\theta)\mathrm{d}\theta = p(q)\mathrm{d}q = \mathrm{e}^{-rq} r \mathrm{d}q = \mathrm{e}^{-(\theta+1)}\mathrm{d}\theta, \ \theta \ge -1 \tag{51}$$

Standardized density and cumulative distribution:

$$g(\theta) = \mathrm{e}^{-(\theta+1)}, \ \theta \ge -1, \qquad G(\theta) = \int_{-1}^{\theta} g(z)\,\mathrm{d}z = 1 - \mathrm{e}^{-(\theta+1)} \tag{52}$$

$g(\theta)$ is a shifted exponential distribution.

¶ Proxy property: example.

- Suppose v_i and v_j both belong to the same standardization class.
- Their info-gap model is eq.(17) and robustness is eq.(24).
- Their standardization functions are:

$$\theta(v_i) = \frac{v_i - \widetilde{v}_i}{c s_i} \tag{53}$$

where $c > 0$.

- $G(\theta)$ = cumulative probability distribution function of the standardized random variables.
- Probability of success for option i is:

$$\begin{aligned} P_{\mathrm{s}}(i) &= \mathrm{Prob}(v_i \ge v_{\mathrm{c}}) = \mathrm{Prob}\left(\frac{v_i - \widetilde{v}_i}{c s_i} \ge \frac{v_{\mathrm{c}} - \widetilde{v}_i}{c s_i}\right) & (54) \\ &= 1 - G\left(\frac{v_{\mathrm{c}} - \widetilde{v}_i}{c s_i}\right) & (55) \\ &= 1 - G\left[-\frac{\widehat{h}(i, v_{\mathrm{c}})}{c}\right] & (56) \end{aligned}$$

where eq.(56) results from eqs.(55) and (24) if $v_{\mathrm{c}} \le \widetilde{v}_i$.

- We see that:

$$P_{\mathrm{s}}(i) > P_{\mathrm{s}}(j) \quad \text{if and only if} \quad \widehat{h}(i, v_{\mathrm{c}}) > \widehat{h}(j, v_{\mathrm{c}}) \tag{57}$$

- This example illustrates a general result:

Standardization implies that the proxy property holds.

- In order to calculate $\widehat{h}(i, v_{\mathrm{c}})$ and hence maximize $P_{\mathrm{s}}(i)$ we must be able to standardize the v_i's, eq.(53).
- This **requires knowing,** for each i:
  - v_i = mean.
  - $c s_i$ proportional to standard deviation.

- This does not **require knowing:**
  - Value of c (actual standard deviations).
  - Identify of pdf.

2.6 Coherence

§ **Coherence:**

- A weak informational-overlap between an info-gap model and a probability distribution.
- Coherence is necessary and sufficient for the proxy property to hold.

§ Scalar uncertainty, q.

- E.g. $q = r^T v$.
- $\mathcal{Q}_r(h)$ is info-gap model for q.
- $P(q|r)$ and $p(q|r)$ are cumulative prob distribution (cpd) and pdf for q.
- $G(r,q)$ is the performance function. Monotonic in q.
- Define:

$$
\begin{aligned}
q^\star(h,r) &\equiv \max_{q\in\mathcal{Q}_r(h)} q && (58)\\
q_\star(h,r) &\equiv \min_{q\in\mathcal{Q}_r(h)} q && (59)\\
\mu(h) &\equiv \min_{q\in\mathcal{Q}_r(h)} G(r,q) && (60)
\end{aligned}
$$

- Define inverse of $G(r,q)$, at fixed r, as follows.

  If $G(r,q)$ *increases* as q increases:

$$G^{-1}(r,G_c) \equiv \max\{q:\ G(r,q)\le G_c\} \quad (61)$$

  If $G(r,q)$ *decreases* as q increases:

$$G^{-1}(r,G_c) \equiv \min\{q:\ G(r,q)\le G_c\} \quad (62)$$

Definition 3 . *$\mathcal{Q}_r(h)$ and $P(q|r)$ are* ***upper coherent*** *at decisions r_1 and r_2 and critical value G_c, with performance function $G(r,q)$, if the following two relations hold for $i=1$ or $i=2$, and $j=3-i$:*

$$P[G^{-1}(r_i,G_c)|r_i] \;>\; P[G^{-1}(r_j,G_c)|r_j] \quad (63)$$

$$G^{-1}(r_i,G_c)-q^\star(h,r_i) \;>\; G^{-1}(r_j,G_c)-q^\star(h,r_j)$$

$$\textit{for } h=\widehat{h}(r_j,G_c) \textit{ and } h=\widehat{h}(r_i,G_c) \quad (64)$$

$\mathcal{Q}_r(h)$ and $P(q|r)$ are ***lower coherent*** *if eqs.(63) and (64) hold when $q^\star(h,r)$ is replaced by $q_\star(h,r)$.*

- Coherence implies "information overlap" between $\mathcal{Q}_r(h)$ and $P(q|r)$.
- Eq.(63) depends on $P(q|r)$ but not on h or $\mathcal{Q}_r(h)$.

- Eq.(64) depends on h and $\mathcal{Q}_r(h)$ but not on $P(q|r)$.
- Coherence implies that knowledge of one function reveals something about the other.

§ **Example.** Following are coherent with $G(r, q) = q/r$:

$$P(q|r) = 1 - \mathrm{e}^{-rq} \tag{65}$$

$$\mathcal{Q}_r(h) = \left\{ q: \ 0 \le q \le \frac{h}{r} \right\}, \quad h \ge 0 \tag{66}$$

- As r increases, $P(q|r)$ and $\mathcal{Q}_r(h)$ both become more highly concentrated.
- Each reveals something about the other. There is some "coherence" between them.

§ **Example.** Following are **not** coherent with $G(r, q) = q/r$: Exponential distribution, eq.(65), and:

$$\mathcal{Q}_r(h) = \{ q: \ 0 \le q \le rh \}, \quad h \ge 0 \tag{67}$$

- As r increases, $P(q|r)$ becomes more highly focussed while $\mathcal{Q}_r(h)$ becomes more dispersed.

2.7 Coherence and the Proxy Property

§ We now state and discuss an important theorem:

coherence is necessary and sufficient for the proxy property to hold.

Definition 4 *An info-gap model, $\mathcal{Q}_r(h)$,* ***expands upward continuously*** *at h if, for any $\varepsilon > 0$, there is a $\delta > 0$ such that:*

$$|q^\star(h', r) - q^\star(h, r)| < \varepsilon \quad \textit{if} \quad |h' - h| < \delta \tag{68}$$

Continuous downward expansion *is defined similarly with $q_\star(\cdot)$ instead of $q^\star(\cdot)$.*

We can now state a proposition.[1]

Proposition 1 *Info-gap robustness to an uncertain scalar variable, with a loss function which is monotonic in the uncertain variable, is a proxy for probability of survival if and only if the info-gap model $\mathcal{Q}_r(h)$ and the probability distribution $P(q|r)$ are coherent.*

Given:

• *At any fixed decision r, the performance function, $G(r, q)$, is monotonic (though not necessarily strictly monotonic) in the scalar q.*

• *$\mathcal{Q}_r(h)$ is an info-gap model with the property of nesting.*

• *r_1 and r_2 are decisions with positive, finite robustnesses at critical value $G_{\rm c}$.*

• *$\mathcal{Q}_r(h)$ is continuously upward (downward) expanding at $\widehat{h}(r_1, G_{\rm c})$ and at $\widehat{h}(r_2, G_{\rm c})$ if $G(r, q)$ increases (decreases) with increasing q.*

Then: *The* ***proxy property*** *holds for $\mathcal{Q}_r(h)$ and $P(q|r)$ at r_1, r_2 and $G_{\rm c}$ with performance function $G(r, q)$.*

If and only if: *$\mathcal{Q}_r(h)$ and $P(q|r)$ are* ***upper (lower) coherent*** *at r_1, r_2 and $G_{\rm c}$ with performance function $G(r, q)$ which increases (decreases) in q.*

[1] Yakov Ben-Haim, Robust-satisficing and the probability of survival, working paper, \papers\Proxy Thms\prx22.tex.