Two for the Price of One: Info-Gap Robustness of the 1-Test Algorithm

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

Abstract

Analysts in many domains must choose a design, a strategy, or an intervention without being able to test all relevant alternatives. We consider a situation in which one of two alternatives must be chosen, while only one alternative can be tested prior to decision. The probability of success from blind choice is 1/2. The probability of success if the distribution of the system attributes is known is 3/4. The 1-test algorithm assures probability greater than 1/2 of choosing the better system based on a single test, even without knowing the probability distribution of the system attributes. If the distribution is poorly known, then info-gap theory can robustify the 1-test algorithm. Using the info-gap robustness function we show that robust-satisficing algorithms may differ from the nominally optimal algorithm when the attribute distribution is uncertain.

Keywords. Testing, design, info-gap.

1 The 1-Test Algorithm

Consider a choice between two design concepts for a technological system. We would like to choose the system with higher reliability (or longer life or lower mean time between failure, etc.). It may be very expensive to construct and test both physical systems. It would be useful if the better system could be reliably chosen based on testing only one system.

Consider the choice between two medical interventions for a specific patient (or macro-economic interventions for a specific economy, or biological interventions in an ecosystem). We can do one or the other, but not both. Given all available information, we are epistemically indifferent between the interventions: we have no reason to believe that one intervention is better than the other, though they are different. We choose one intervention by flipping a fair coin, and we observe the result (reduction in fever, or increase in blood count, etc.). For future reference we would like to know which of the two would have been better.

Decisions such as these can be thought about generically as follows.

Two systems each have a real-valued attribute (e.g. lifetime, reliability, etc.). We would like to choose the system with the larger—better—value, but we are able to measure the attribute of only one system. We must decide if the measured attribute is the smaller or the larger of the two, where we have chosen the system to test by a throw of a fair coin. We know nothing about the distribution of the attribute values, other than that they can take any value in a specified interval.

The 1-test algorithm is stated without proof by Cover [2] and proven by Snapp [6]. The idea is also discussed in a blog [7]. We can formalize it as follows.

Two different real numbers, x_1 and x_2, are chosen by an algorithm unknown to you. One of these numbers, call it x_{r}, is revealed to you, where you know that the probability that $x_{\mathrm{r}} = x_1$ is 0.5. You must decide if x_{r} is the smaller or the larger of the two numbers.

The **1-test algorithm** for deciding whether x_{r} is the smaller or larger of the two values is as follows. Let $q(y)$ be a non-atomic probability density function (pdf) which is positive on an interval containing x_1 and x_2. The interval may be finite, half-finite, or infinite. We will refer to $q(y)$ as the "decision pdf". Decide according to the following decision rule:

1. Draw a random number, y, distributed according to $q(y)$.
2. If $y \geq x_{\mathrm{r}}$ then decide that x_{r} is the smaller of the two x_i.
3. If $y < x_{\mathrm{r}}$ then decide that x_{r} is the larger of the two x_i.

The 1-test algorithm succeeds if the number chosen by the algorithm is in fact the larger of the two numbers.

Let $P_\mathrm{s}(x_1, x_2, q)$ denote the probability of success of the 1-test algorithm using a pdf $q(y)$ applied to real numbers x_1 and x_2. We will prove the following theorem. See Cover [2] and Snapp [6].

Theorem 1 *The probability of success of the 1-test algorithm exceeds 1/2.*

Given:

- *The two numbers, x_1 and x_2, are different.*

- *$q(y)$ is a non-atomic probability density function which is non-zero on an interval containing x_1 and x_2. $q(y)$ is zero outside this interval.*

Then:

$$P_\mathrm{s}(x_1, x_2, q) > \frac{1}{2} \tag{1}$$

Proof of theorem 1. The two numbers are different, so one is larger. Denote the larger of the two numbers by x_1, where x_r is the number which has been revealed. Our information is:

$$\mathrm{Prob}(x_\mathrm{r} = x_1) = \mathrm{Prob}(x_\mathrm{r} = x_2) = 0.5 \tag{2}$$

If x_r is the larger of the two numbers, then the probability of success equals the probability that $y < x_\mathrm{r}$:

$$P_\mathrm{s}(x_\mathrm{r} = x_1) = \int_{-\infty}^{x_1} q(y)\,\mathrm{d}y = Q(x_1) \tag{3}$$

where $Q(y)$ is the cumulative distribution function of $q(y)$. Similarly, if x_r is the smaller of the two numbers, then the probability of success equals the probability that $y \geq x_\mathrm{r}$:

$$P_\mathrm{s}(x_\mathrm{r} = x_2) = \int_{x_2}^{\infty} q(y)\,\mathrm{d}y = 1 - Q(x_2) \tag{4}$$

$P_\mathrm{s}(x_\mathrm{r} = x_1)$ and $P_\mathrm{s}(x_\mathrm{r} = x_2)$ are illustrated in fig. 1.

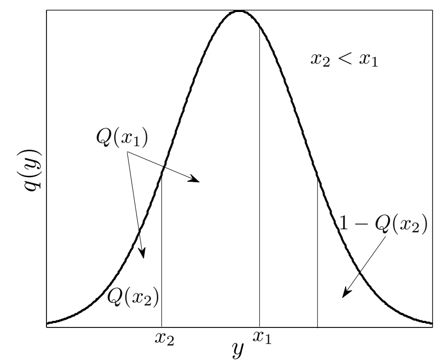

Figure 1: $Q(x_1)$ and $1 - Q(x_2)$ for $x_2 < x_1$; eqs.(3) and (4).

Recall that $x_1 > x_2$ which, since $q(y)$ is non-zero on an interval containing x_1 and x_2, implies:

$$Q(x_1) > Q(x_2) \tag{5}$$

Thus the total probability of success, with the q-based decision algorithm, is:

$$\begin{aligned} & P_\mathrm{s}(x_1, x_2, q) \\ &= \mathrm{Prob}(x_\mathrm{r} = x_1) P_\mathrm{s}(x_\mathrm{r} = x_1) \\ &\quad + \mathrm{Prob}(x_\mathrm{r} = x_2) P_\mathrm{s}(x_\mathrm{r} = x_2) && (6) \\ &= 0.5 Q(x_1) + 0.5[1 - Q(x_2)] && (7) \\ &= 0.5[1 + \underbrace{Q(x_1) - Q(x_2)}_{>0}] > 0.5 && (8) \end{aligned}$$

which completes the proof. ∎

2 Info-Gap Robustness of the 1-Test Algorithm

The system attributes, x_1 and x_2, are random variables. Let $p(x_1, x_2)$ denote their joint pdf. If we knew this distribution we could choose the 1-test decision distribution, $q(y)$, to maximize the probability of success. But suppose we only have a guess or plausible supposition of the joint pdf of x_1 and x_2. That is, we think they are drawn from a joint pdf which is something like $\widetilde{p}(x_1, x_2)$, but the true distribution may have a different shape or different moments. How should we choose $q(y)$?

In this section we introduce info-gap models to represent non-probabilistic uncertainty about the true pdf of x_1 and x_2. We then define the info-gap robustness function and illustrate its use in selecting the decision pdf $q(y)$.

2.1 Info-Gap Uncertainty and Robustness

An info-gap model [1], [4] is a family of nested sets, $\mathcal{U}(h, \widetilde{p})$, $h \geq 0$. The elements of these sets are realizations of the uncertain quantity, which is the joint pdf of x_1 and x_2 in the present case. The set-valued functions, $\mathcal{U}(h, \widetilde{p})$, of an info-gap model, have the following properties:

$$\text{Contraction:}\quad \mathcal{U}(0, \widetilde{p}) = \{\widetilde{p}\} \tag{9}$$

$$\text{Nesting:}\quad h < h' \text{ implies } \mathcal{U}(h, \widetilde{p}) \subseteq \mathcal{U}(h', \widetilde{p}) \tag{10}$$

Contraction states that, in the absence of uncertainty, only a single function—our estimate—applies, so the uncertainty set is a singleton. *Nesting* is the property that the sets become more inclusive as the horizon of uncertainty grows. An info-gap model is a non-probabilistic quantification of uncertainty. It entails no assumptions about probability distributions or about worst cases.

For instance, consider a situation where evidence supports a symmetric pdf $\widetilde{p}(x)$ for $|x| \leq d$, but no evidence is available on the far tails, $|x| > d$, and fat tails are suspected. A simple info-gap model for this situation is:

$$\mathcal{U}(h,\widetilde{p}) = \Big\{ \; p(x) \in \mathcal{P} : \; p(x) = \nu \widetilde{p}(x), \; |x| \le d \qquad p(x) \le \frac{h}{x^2}, \; |x| > d \; \Big\}, \quad h \ge 0 \quad (11)$$

where $\mathcal{P}$ is the set of non-negative and normalized pdfs.

The generic properties of an info-gap model are eqs.(9) and (10), for which eq.(11) is an example. An info-gap model can be a Lévy neighborhood or a contamination neighbor, as treated by Huber [3], but need not be as illustrated by eq.(11) and in [1] and [4].

2.2 Info-Gap Robustness

The unknown joint pdf of x_1 and x_2 is $p(x_1, x_2)$ where we will assume that the variables x_1 and x_2 are exchangeable: $p(x_1, x_2) = p(x_2, x_1)$. x_1 and x_2 are also exchangeable in the estimated joint pdf, $\widetilde{p}(x_1, x_2)$. $\mathcal{U}(h, \widetilde{p})$ is an info-gap model for uncertainty in $p(x)$.

Let $P_{\rm s}(p, q)$ denote the overall probability of success, regardless of the realizations of x_1 and x_2, based on the 1-test algorithm with decision pdf $q(y)$:

$$P_{\rm s}(p,q) = 2 \int_{-\infty}^{\infty} \int_{x_2}^{\infty} P_{\rm s}(x_1, x_2, q) p(x_1, x_2) \, {\rm d}x_1 \, {\rm d}x_2 \quad (12)$$

In the double integral itself (without the factor 2) we assume that x_1 is greater than x_2. Multiplying by 2 accounts for the other possibility.

We aspire to choose $q(y)$ so that $P_{\rm s}(p, q)$ is no less than a "critical value", $P_{\rm c}$. We know from theorem 1 and eq.(12) that $P_{\rm s}(p, q)$ exceeds 0.5; we might aspire to exceed 0.6 or 0.7. The robustness of any choice of $q(y)$, given aspiration $P_{\rm c}$, is the greatest horizon of uncertainty in the true distribution of x_1 and x_2, up to which all distributions result in probability of success no less than $P_{\rm c}$. Large robustness implies that our estimate, $\widetilde{p}(x_1, x_2)$, can err greatly and the 1-test algorithm with $q(y)$ will still achieve a probability of success no less than $P_{\rm c}$. Small robustness implies high vulnerability to error in the estimate. Clearly, the robustness function $\widehat{h}(q, P_{\rm c})$ establishes preferences on the decision pdfs $q(y)$.

Mathematically, we define the robustness of a decision pdf, $q(y)$, as the greatest horizon of uncertainty, h, up to which the probability of success is no less than the critical value, $P_{\rm c}$, for all possible pdf's at that horizon of uncertainty:

$$\widehat{h}(q, P_{\rm c}) = \max \left\{ h : \; \left(\min_{p \in \mathcal{U}(h, \widetilde{p})} P_{\rm s}(p, q) \right) \ge P_{\rm c} \right\} \quad (13)$$

The robustness, $\widehat{h}(q, P_{\rm c})$, is the least upper bound of the set of h values which satisfice the probability of success at its critical value. We define the robustness to equal zero if the set of h values in eq.(13) is empty.

The info-gap robustness in eq.(13) is different in several respects from the concepts of robustness in robust statistics ([3], section 1.4). First of all, the info-gap model need not represent uncertainty with the neighborhoods usually treated in robust statistics, as mentioned at the end of section 2.1 and illustrated in eq.(11). Eq.(13) does not consider the bias or variance of a statistic, nor the asymptotic (large sample) properties of any statistic, nor does it assume that the statistic is consistent in the sense of converging (in probability) to an asymptotic value. For further discussion of the relation between robust statistics and info-gap robustness see [5].

2.3 Simple Example

We now examine a very simple special case. We know that x_1 and x_2 are chosen independently from an exponential distribution, $p(x) = \lambda {\rm e}^{-\lambda x}$, $x \ge 0$. Our best guess of the coefficient is $\widetilde{\lambda}$ but this guess is very uncertain. We use a fractional-error info-gap model for uncertainty in the exponential coefficient of the pdf by which the x_i are chosen:

$$\mathcal{U}(h, \widetilde{p}) = \left\{ p(x) = \lambda {\rm e}^{-\lambda x} : (1-h)^{+} \widetilde{\lambda} \le \lambda \le (1+h) \widetilde{\lambda} \right\} \quad h \ge 0 \quad (14)$$

where $x^{+} = x$ if $x \ge 0$ and equals zero otherwise. Furthermore, assume that the pdf used for deciding is also exponential: $q(y) = \gamma {\rm e}^{-\gamma y}$. We will derive the robustness function (actually, its inverse) and study the choice of γ.

Let $P_{\rm s}(\lambda, \gamma)$ denote the overall probability of success, eq.(12), when the true distribution is exponential with coefficient λ and the decision pdf is exponential with coefficient γ. One finds:

$$P_{\rm s}(\lambda, \gamma) = \frac{1}{2} + \frac{\lambda \gamma}{(\lambda + \gamma)(2\lambda + \gamma)} \quad (15)$$

$$= \frac{1}{2} + \frac{\rho}{(1+\rho)(1+2\rho)}, \quad \rho = \frac{\lambda}{\gamma} \quad (16)$$

Differentiating we find:

$$\frac{\partial P_{\rm s}(\lambda, \gamma)}{\partial \lambda} = \frac{\gamma(\gamma^2 - 2\lambda^2)}{(\lambda + \gamma)^2 (2\lambda + \gamma)^2} \quad (17)$$

$P_s(\lambda,\gamma)$ vs. λ is a unimodal function with a maximum at $\lambda = \gamma/\sqrt{2}$, as illustrated in fig. 2.

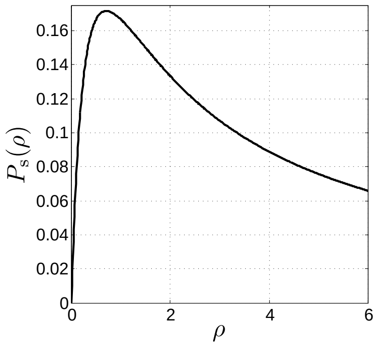

Figure 2: $P_s(\rho)$ defined in eq.(16).

Let $\mu(h,\gamma)$ denote the inner minimum in the definition of the robustness, eq.(13), which is the minimum of $P_s(\lambda,\gamma)$ as λ varies up to horizon of uncertainty h. $\mu(h,\gamma)$ is the inverse of $\widehat{h}(q,P_c)$. That is:

$$\mu(h,\gamma) = P_c \quad \text{implies} \quad \widehat{h}(q,P_c) = h \tag{18}$$

A plot of $\mu(h,\gamma)$ vs. h is the same as a plot of P_c vs. $\widehat{h}(q,P_c)$.

The minimum of $P_s(\lambda,\gamma)$, at horizon of uncertainty h, occurs when λ takes one or the other of its extreme values, which are:

$$\lambda_1(h) = (1+h)\widetilde{\lambda} \tag{19}$$
$$\lambda_2(h) = (1-h)^+\widetilde{\lambda} \tag{20}$$

Let us define the following two functions:

$$\mu_1(h,\gamma) = P_s[(1+h)\widetilde{\lambda},\gamma] \tag{21}$$
$$\mu_2(h,\gamma) = P_s[(1-h)^+\widetilde{\lambda},\gamma] \tag{22}$$

The inner minimum in the definition of the robustness is the lesser of these two functions:

$$\mu(h,\gamma) = \min_i \mu_i(h,\gamma) \tag{23}$$

The nominal optimal choice of γ is the value which maximizes the estimated function $P_s(\widetilde{\lambda},\gamma)$:

$$\gamma^\star = \arg\max_\gamma P_s(\widetilde{\lambda},\gamma) \tag{24}$$

We find $\gamma^\star$ by differentiating $P_s(\widetilde{\lambda},\gamma)$:

$$\frac{\partial P_s(\widetilde{\lambda},\gamma)}{\partial\gamma} = \frac{\widetilde{\lambda}(2\widetilde{\lambda}^2-\gamma^2)}{(\widetilde{\lambda}+\gamma)^2(2\widetilde{\lambda}+\gamma)^2} \tag{25}$$

Thus we see that the nominal optimal choice of γ is:

$$\gamma^\star = \widetilde{\lambda}\sqrt{2} \tag{26}$$

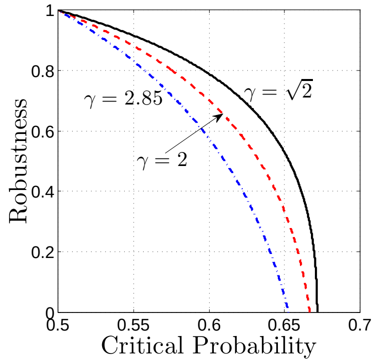

Figure 3: 3 robustness curves.

Note that the nominal optimal decision pdf, $q(y|\gamma^\star)$, differs from the estimated generating pdf, $\widetilde{p}(x|\widetilde{\lambda})$, even if the estimate is correct.

Figs. 3–5 show robustness curves, $\widehat{h}(q,P_c)$ vs P_c, for different choices of γ, which determines the decision pdf, $q(y)$. The estimated value of λ, the coefficient of the estimated distribution of x_i, is $\widetilde{\lambda} = 1$ in all cases.

The curves all converge, at the upper left, at $\widehat{h} = 1$ when $P_c = 1/2$. We understand this from eq.(15), where $P_s = 1/2$ when $\lambda = 0$.

In fig. 3 we examine values of γ for which $\mu(h,\gamma)$ in eq.(23) takes only one functional form—$\mu_1(h,\gamma)$—for all horizons of uncertainty, so no kink occurs in the curve. The peak of $P_s(\lambda,\gamma)$ vs. λ (see fig. 2 or eq.(17)) occurs when $\lambda = \gamma/\sqrt{2}$. When $\gamma = \sqrt{2}$ (solid black curve) then, since $\widetilde{\lambda} = 1$, the value of $\mu(0,\gamma)$ occurs at the peak of $P_s(\lambda,\gamma)$ vs. λ. As h increases, the value of $\mu(h,\gamma)$ moves left, down the steep positive slope illustrated in fig. 2. In the other curves of fig. 3, $\widetilde{\lambda} < \gamma/\sqrt{2}$ so the value of $\mu(0,\gamma)$ occurs on the steep positive slope of $P_s(\lambda,\gamma)$ vs λ and, as h increases, the value of $\mu(h,\gamma)$ moves left, down the steep positive slope.

From eq.(26) we see that $\gamma = \sqrt{2}$ is the nominal optimal choice since $\widetilde{\lambda} = 1$. Fig. 3 indicates that this choice is robust-dominant among the values of γ which are shown, and it is clear that this will hold for any value of γ for which $\widetilde{\lambda} \le \gamma/\sqrt{2}$.

Fig. 4 is different from fig. 3: each robustness curve in fig. 4 displays a kink when $\mu(h,\gamma)$ switches from one solution to the other as specified in eq.(23). $\widetilde{\lambda} > \gamma/\sqrt{2}$ in both cases, so $\mu(0,\gamma)$ occurs on the gentle negative-slope portion of $P_s(\lambda,\gamma)$ vs. λ. Thus, for small h, $\mu(h,\gamma)$ moves to the right down the gentle slope. However, at larger h, the value of $(1-h)^+\widetilde{\lambda}$ occurs on the steep positive slope to the left of the peak, and now $\mu(h,\gamma)$ switches and moves left down the steep

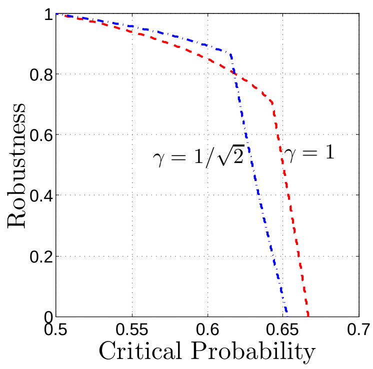

Figure 4: 2 robustness curves.

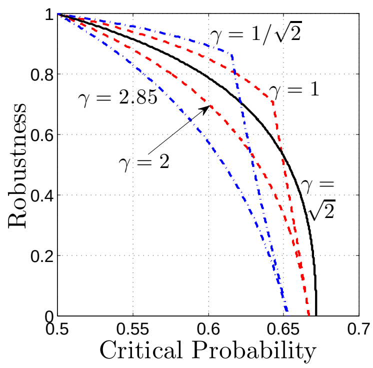

Figure 5: Figs. 3 and 4 combined.

slope. This explains the kink in the robustness curves.

Fig. 5 combines the curves of figs. 3 and 4. What is of particular interest is the intersection between the robustness curves. For instance, the curve for $\gamma = 1$ intersects the curve for $\gamma = \sqrt{2}$ at critical probability $P_c = 0.65$. For greater critical probability, $\gamma = \sqrt{2}$ is more robust (up to $P_c = 0.67$ at which its robustness vanishes). For lower probability, $\gamma = 1$ is more robust. This intersection between robustness curves entails the possibility of reversal of preferences between the corresponding choices of γ (which determines the decision pdf, $q(y)$).

3 Three Properties

We now discuss three generic properties of info-gap robustness curves—trade off, zeroing, and preference reversal—which are illustrated in the example.

Trade off between robustness and performance. Robustness curves, such as in figs. 3–5, are always monotonic, which expresses a trade off between robustness and performance: good performance entails low robustness against uncertainty. In our example, aspiring to high probability of success, P_c, entails low robustness against uncertainty in the generating pdf. This trade off is universal and results from the nesting property of info-gap models, eq.(10). The robustness function, $\widehat{h}(q, P_c)$, quantifies this trade off.

Zeroing of the robustness curve. The robustness, $\widehat{h}(q, P_c)$, will equal zero for some critical value P_c. This value is precisely the estimated performance. Using the notation of our example, the zeroing property is:

$$\widehat{h}(q, P_c) = 0 \quad \text{if} \quad P_c = P_s(\widetilde{p}, q) \tag{27}$$

This means that the robustness is zero when aspiring to a probability of success which equals the estimated probability of success. Estimated outcomes have no robustness against errors in the models that underlie the estimate. Combining this with the trade off property we conclude that only outcomes which are worse than the estimated outcome have positive robustness. This has an important implication for decision under uncertainty. Estimated outcomes are not a good basis for choosing between options because estimated outcomes have no robustness to error in the models and data underlying the estimates.

Preference reversal between options. This paper is based on the idea that more robustness against uncertainty is better than less robustness. This provides a prioritization of options—decision pdf's $q(y)$ in our example—as explained following eq.(12). Figs. 4 and 5 show several examples of intersection between robustness curves for different choices of $q(y)$. For instance, fig. 5 shows crossing between the robustness curves for the nominal optimum ($\gamma = \sqrt{2}$) and a different option ($\gamma = 1$). The former option is preferred if one requires $P_c > 0.65$, while the latter option is preferred if lower probability of success is acceptable. In short, the crossing of robustness curves entails the possibility of reversal of preference between the corresponding options.

4 Extensions of the 1-Test Algorithm

Theorem 1 and the associated decision algorithm relate to selecting a single system from two candidates based on testing only one system. We now consider three candidate systems, where either one or two systems are tested. When testing one system our aim is to select the best of the three systems. When testing two systems our aim is to select the two best systems. We prove two extensions of theorem 1, relating to these two cases, and we propose an hypothesis for more than 3 systems.

4.1 Two Tests, 3 Systems

Consider three systems, each characterized by a single real number, x_i, and assume these numbers are different. Without loss of generality we denote these numbers:

$$x_1 < x_2 < x_3 \tag{28}$$

Two of the systems are tested to reveal their attributes, x_i, where each system has the same probability of being tested. The revealed attributes are:

$$r_1 < r_2 \tag{29}$$

Let s denote the third, unrevealed, number.

Our goal is to select the two best systems, whose attributes are larger than of the third system. We do not need to identify the better of the two best; only to exclude the worst system.

The **2-test 3-system algorithm** is as follows. Let $q(y)$ be a non-atomic pdf which is positive on an interval containing x_1, x_2 and x_3. The interval may be finite, half-finite, or infinite. Select two systems according to the following decision rule:

1. Draw a random number, y, distributed according to $q(y)$.
2. If $y < r_1$, choose the two tested systems.
3. If $r_1 \le y \le r_2$, choose the systems corresponding to r_2 and s.
4. If $r_2 < y$, choose the systems corresponding to r_2 and s.

The probability of blindly choosing the two best systems is 1/3. The following theorem asserts that the above decision algorithm successfully chooses the two best systems with probability strictly exceeding 1/3.

Theorem 2 *The probability of success of the 2-test 3-system algorithm exceeds 1/3.*

Given:

- *The three numbers, x_1, x_2 and x_3, are different.*
- *$q(y)$ is a non-atomic pdf which is positive on an interval containing x_1, x_2 and x_3.*
- *Each system has equal probability of being selected for testing.*

Then:

$$P_{\rm s}(x_1, x_2, x_3, q) > \frac{1}{3} \tag{30}$$

Proof of theorem 2. The three numbers are different, so they can be denoted as in eq.(28). The two revealed numbers are therefore also different and denoted as in eq.(29). Let $R = \{r_1, r_2\}$ denote the set of revealed values. Let s denote the third, unrevealed, number. Let $Q(\cdot)$ denote the cumulative probability distribution of $q(\cdot)$. Since the tested systems are selected with equal probability we can assert:

$$\text{Prob}(s = x_1) = \text{Prob}(s = x_2) = \text{Prob}(s = x_3) = \frac{1}{3} \tag{31}$$

The decision algorithm succeeds at step 2 if $R = \{x_2, x_3\}$ whose probability is 1/3.

The decision algorithm succeeds at step 3 if $R = \{x_1, x_2\}$ or if $R = \{x_1, x_3\}$, each of whose probabilities is 1/3.

The decision algorithm succeeds at step 4 if $R = \{x_1, x_2\}$ or if $R = \{x_1, x_3\}$, each of whose probabilities is 1/3.

Putting this together we can write the total probability of success of the decision algorithm as:

$$\begin{aligned} P_{\rm s}(x_1, x_2, x_3, q) &= \underbrace{\frac{1}{3}\int_{-\infty}^{x_2} q(y)\,{\rm d}y}_{\text{step 2}} \\ &\quad + \underbrace{\frac{1}{3}\int_{x_1}^{x_2} q(y)\,{\rm d}y + \frac{1}{3}\int_{x_1}^{x_3} q(y)\,{\rm d}y}_{\text{step 3}} \\ &\quad + \underbrace{\frac{1}{3}\int_{x_2}^{\infty} q(y)\,{\rm d}y + \frac{1}{3}\int_{x_3}^{\infty} q(y)\,{\rm d}y}_{\text{step 4}} \end{aligned} \tag{32}$$

$$\begin{aligned} &= \underbrace{\frac{1}{3}Q(x_2)}_{\text{step 2}} \\ &\quad + \underbrace{\frac{1}{3}[Q(x_2) - Q(x_1)] + \frac{1}{3}[Q(x_3) - Q(x_1)]}_{\text{step 3}} \\ &\quad + \underbrace{\frac{1}{3}[1 - Q(x_2)] + \frac{1}{3}[1 - Q(x_3)]}_{\text{step 4}} \\ &= \frac{2}{3} - \frac{1}{3}\underbrace{Q(x_1)}_{<1} + \frac{1}{3}\underbrace{[Q(x_2) - Q(x_1)]}_{>0} > \frac{1}{3} \end{aligned} \tag{33}$$

$Q(x_1) < 1$ and $Q(x_2) > Q(x_1)$ because $x_1 < x_2$ and $q(y)$ is positive on an interval containing x_1, x_2 and x_3. This completes the proof. ∎

4.2 One Test, 3 Systems

Consider three systems, each characterized by a single real number, x_i, and assume these numbers are different. Without loss of generality we denote these numbers as in eq.(28). One of the systems is tested to reveal its attribute, r, where each system has the same probability of being tested. t denote the unrevealed numbers.

Our goal is to select the best system, whose attribute is larger than of the other two systems.

The **1-test 3-system algorithm** is as follows. Let $q(y)$ be a non-atomic pdf which is positive on an interval containing x_1, x_2 and x_3. The interval may be finite, half-finite, or infinite. Select a system according to the following decision rule:

1. Draw a random number, y, distributed according to $q(y)$.
2. If $y \le r$, choose the tested system.
3. If $y > r$, choose between the untested systems with equal probability.

The probability of blindly choosing the best system is 1/3. The following theorem asserts that the above decision algorithm successfully chooses the best system with probability strictly exceeding 1/3.

Theorem 3 *The probability of success of the 3-system 1-test algorithm exceeds 1/3.*

Given:

• *The three numbers, x_1, x_2 and x_3, are different.*

• *$q(y)$ is a non-atomic pdf which is positive on an interval containing x_1, x_2 and x_3.*

• *Each system has equal probability of being selected for testing.*

Then:

$$P_{\rm s}(x_1, x_2, x_3, q) > \frac{1}{3} \tag{34}$$

Proof of theorem 3. The three numbers are different, so they can be denoted as in eq.(28). Let r denote the revealed value. Let s and t denote the unrevealed numbers. Let $Q(\cdot)$ denote the cumulative probability distribution of $q(\cdot)$. We can assert:

$$\text{Prob}(r = x_1) = \text{Prob}(r = x_2) = \text{Prob}(r = x_3) = 1/3 \tag{35}$$

The decision algorithm succeeds at step 2 if $r = x_3$ (with probability 1/3).

The decision algorithm succeeds at step 3 if the choice between s and t is correct (with probability 0.5), and if either $r = x_1$ or $r = x_2$ (each with probability is 1/3).

Putting this together we can write the total probability of success of the decision algorithm as:

$$P_{\rm s}(x_1, x_2, x_3, q) = \underbrace{\frac{1}{3}\int_{-\infty}^{x_3} q(y)\,{\rm d}y}_{\text{step 2}} \tag{36}$$

$$+ \underbrace{\frac{1}{2}\frac{1}{3}\left[\int_{x_1}^{\infty} q(y)\,{\rm d}y + \int_{x_2}^{\infty} q(y)\,{\rm d}y\right]}_{\text{step 3}}$$

$$= \underbrace{\frac{1}{3}Q(x_3)}_{\text{step 2}} + \underbrace{\frac{1}{6}\left[(1 - Q(x_1)) + (1 - Q(x_2))\right]}_{\text{step 3}} \tag{37}$$

$$= \frac{1}{3} + \frac{1}{6}\underbrace{[Q(x_3) - Q(x_1)]}_{>0} + \frac{1}{6}\underbrace{[Q(x_3) - Q(x_2)]}_{>0} \tag{38}$$

$$> \frac{1}{3} \tag{39}$$

which completes the proof. ∎

4.3 m Tests, n Systems

Consider n systems, each characterized by a single real number, x_i, and assume these numbers are different. Without loss of generality we denote these numbers:

$$x_1 < x_2 < \cdots < x_n \tag{40}$$

m of the systems are tested to reveal their attributes, x_i, where each system has the same probability of being tested. The revealed attributes are:

$$r_1 < r_2 < \cdots < r_m \tag{41}$$

Let $R = \{r_1, \ldots, r_m\}$ denote the set of revealed values. Let R_j denote the set R after removing the j smallest elements: $R_j = \{r_{j+1}, \ldots, r_m\}$, for $j = 0, \ldots, m$. Thus $R_0 = R$ and $R_m = \emptyset$. Define $r_0 = -\infty$ and $r_{m+1} = \infty$.

Our goal is to select the m best systems, whose attributes are larger than all the remaining systems. We do not need to identify the values of these m best systems; only to exclude the $n - m$ worst systems.

The m-test n-system algorithm takes a slightly different form depending on whether or not the number of tested systems, m, is less than the number of untested systems, $n - m$. If $m \le n - m$ then the best m systems may be entirely in the untested set. If $m > n-m$ then at least some tested systems are among the best m systems. We specify these two realizations of the decision algorithm separately.

Let $q(y)$ be a non-atomic pdf which is positive on an interval containing x_1, x_2 and x_3. The interval may be finite, half-finite, or infinite.

If $m \le n-m$, the m-test n-system algorithm is as follows. Select m systems according to the following decision rule:

1. Draw a random number, y, distributed according to $q(y)$.

2. For $j = 0, \ldots, m$, if $r_j \le y < r_{j+1}$, choose the systems corresponding to R_j and choose j untested systems equi-probably from among all untested systems.

If $m > n-m$, the m-test n-system algorithm is as follows. Select m systems according to the following decision rule:

1. Draw a random number, y, distributed according to $q(y)$.
2. For $j = 0, \ldots, n-m$, if $r_j \le y < r_{j+1}$, choose the systems corresponding to R_j and choose j untested systems equi-probably from among all untested systems.
3. For $j = n-m+1, \ldots, m$, if $r_j \le y < r_{j+1}$, choose the systems corresponding to R_{n-m} and choose all $n-m$ untested systems.

The number of distinct subsets of m from among the n systems is the binomial coefficient $\binom{n}{m}$, which we denote γ_{nm}. Only one of these subsets contains the m best systems. Thus the probability of blindly choosing the m best systems is $1/\gamma_{nm}$. We hypothesize that one could prove, in analogy to theorems 1–3, that the above decision algorithm chooses the m best systems with probability strictly exceeding $1/\gamma_{nm}$.

5 Further Questions

The 1- and 2-test algorithms can probably be further generalized in various ways. Likewise, the info-gap analysis can be realized in many different forms, especially by using different info-gap models to represent different types of prior information about the uncertain generating pdf. Many questions remain to be explored. We mention a few possible extensions of our results.

(1) In some situations the systems are evaluated by multiple criteria, not by only one attribute as we have done. (2) One might consider adaptive testing, wherein intermediate results indicate whether or not to continue testing. (3) One would like to know what is the best possible probability of success.

References

[1] Ben-Haim, Yakov, 2006, *Info-Gap Decision Theory: Decisions Under Severe Uncertainty,* 2nd edition, Academic Press, London.

[2] Cover, Thomas M., 1987, Pick the largest number, chapter 5.1 in T. Cover and B. Gopinath, 1987, *Open Problems in Communication and Computation,* Springer-Verlag, Berlin.

[3] Huber, Peter J., 1981, *Robust Statistics,* Wiley, New York.

[4] Info-gap decision theory, http://info-gap.com.

[5] Keren, Carmit, 2009, *Info Gap Bayesian Classification,* M.Sc. thesis, Technion-Israel Institute of Technology (in English).

[6] Snapp, Robert R., 2005, Tom Covers Number Guessing Game, http://www.cems.uvm.edu/˜ snapp/teaching/coversproblem.pdf

[7] xkcd, http://blog.xkcd.com/2010/02/09/math-puzzle.