10. 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[3] 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.

For example, two systems have an attribute (e.g. lifetime, reliability, etc.) with values x_1 and x_2, but we are able to test and estimate the attribute of only one system. We must decide if the revealed 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.

(a) Let $q(y)$ be a pdf which is positive on all real numbers. Consider the following decision rule:[4]

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

What is an intuitive explanation of why this algorithm works? Show that the probability of successful decision with this rule is greater than 1/2.

(b) Suppose that we have a rough guess of the pdf by which the x_i are chosen. Specifically, suppose we think they are drawn from a joint pdf which is something like $\widetilde{p}(x_1, x_2)$. How should we choose the distribution $q(y)$ which will be used to decide according to the algorithm in part (a)? Formulate the robustness for the probability of successful decision with $q(y)$.

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

$$\mathcal{U}(h, \widetilde{p}) = \left\{ p(x) = \lambda \mathrm{e}^{-\lambda x} : \ \max[0, \ (1-h)\widetilde{\lambda}] \leq \lambda \leq (1+h)\widetilde{\lambda} \right\}, \quad h \geq 0 \tag{24}$$

Furthermore, assume that the pdf used for deciding is also exponential: $q(y) = \gamma \mathrm{e}^{-\gamma y}$. Derive the robustness function (or its inverse) and explore the choice of γ.

(d) Demonstrate the decision algorithm described in part (a) by simulation. Draw N pairs of numbers, (x_1, x_2), independently from a 'generating' pdf $p(x)$ of your choice. Use the decision algorithm described above with a 'deciding' pdf $q(y)$ of your choice. The theoretical probability of success for a pair (x_1, x_2) is $P_\mathrm{s}(x_1, x_2)$ (this was derived in part (a)). The *average theoretical probability of success,* over the N draws, is:

$$\overline{P}_\mathrm{s} = \frac{1}{N} \sum_{i=1}^{N} P_\mathrm{s}(x_1, x_2) \tag{25}$$

The *empirical probability of success,* F, is the fraction of the N draws in which the q-distribution decision algorithm decides correctly: deciding 'smaller' when x_r is the smaller between x_1 and

[3] It is very important that we know the probability is equal. Otherwise, invoking the principle of indifference would lead to a contradiction, as in the typical 2-envelope problem.

[4] The algorithm was proposed by Thomas M. Cover, 1987, Pick the largest number, chapter 5.1 in T. Cover and B. Gopinath, 1987, *Open Problems in Communication and Computation,* Springer-Verlag, Berlin. See also:
◦ Snapp, Robert R., 2005, Tom Covers Number Guessing Game, http://www.cems.uvm.edu/˜snapp/teaching/ coversproblem.pdf.
◦ http://blog.xkcd.com/2010/02/09/math-puzzle.
◦ Yakov Ben-Haim, 2011, Two for the Price of One: Info-Gap Robustness of the 1-Test Algorithm, 7th Intl Symp on Imprecise Probabilities and their Applications, Innsbruck, Austria. Available at: http://info-gap.com/content.php?id=22

x_2 and deciding 'larger' when x_r is the larger between x_1 and x_2. Show that the empirical and the average theoretical probabilities of success agree.

(e) Consider n systems with values $x_1, \ldots, x_n$, which are all different. Suppose that $m < n$ of these values are revealed, with equal probabilities for each system to be revealed. Can the algorithm of part (a) be generalized? Formulate and study the robustness of the choice of the decision pdf.

Solution to problem 10.

(a) The two numbers are different, so one is larger. Denote the larger of the two by x_1, where $x_{\rm r}$ is the number which has been revealed. Our information is:

$$\text{Prob}(x_{\rm r} = x_1) = \text{Prob}(x_{\rm r} = x_2) = 0.5 \tag{79}$$

A "naive" decision algorithm — even chance that $x_{\rm r}$ is the larger or the smaller — has probability of success of 0.5.

Now consider the decision algorithm based on drawing y from $q(x)$.

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

$$P_{\rm s}(x_{\rm r} = x_1) = \int_{-\infty}^{x_1} q(y)\,{\rm d}y = Q(x_1) \tag{80}$$

Similarly, if $x_{\rm r}$ is the smaller of the two numbers, then the probability of success equals the probability that $y \ge x_{\rm r}$:

$$P_{\rm s}(x_{\rm r} = x_2) = \int_{x_2}^{\infty} q(y)\,{\rm d}y = 1 - Q(x_2) \tag{81}$$

$P_{\rm s}(x_{\rm r} = x_1)$ and $P_{\rm s}(x_{\rm r} = x_2)$ are illustrated in fig. 2.

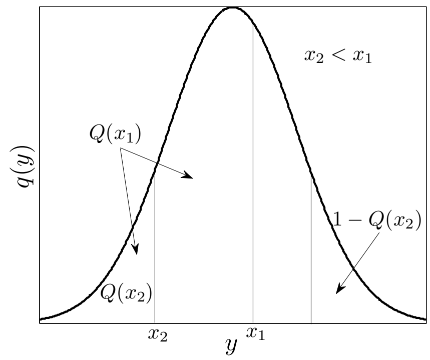

Figure 2: Illustration of $Q(x_1)$ and $1 - Q(x_2)$ where $x_2 < x_1$, defined in eqs.(80) and (81).

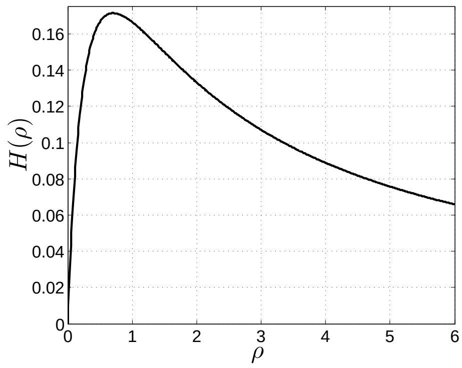

Figure 3: $H(\rho)$ defined in eq.(87).

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

$$\begin{aligned} P_{\rm s}(x_1, x_2, q) &= \text{Prob}(x_{\rm r} = x_1) P_{\rm s}(x_{\rm r} = x_1) + \text{Prob}(x_{\rm r} = x_2) P_{\rm s}(x_{\rm r} = x_2) && (82) \\ &= 0.5 Q(x_1) + 0.5[1 - Q(x_2)] = 0.5[1 + \underbrace{Q(x_1) - Q(x_2)}_{>0}] > 0.5 && (83) \end{aligned}$$

Recall that $x_1 > x_2$ in these relations.

(b) Our rough guess is that the two numbers, $x_1 > x_2$, are drawn from the pdf $\widetilde{p}(x_1, x_2)$, where the variables x_1 and x_2 are exchangeable: $\widetilde{p}(x_1, x_2) = \widetilde{p}(x_2, x_1)$. The actual pdf, $p(x_1, x_2)$, is unknown, though x_1 and x_2 are exchangeable in $p(x_1, x_2)$. Denote an info-gap model for the uncertainty in $\widetilde{p}$ by $\mathcal{U}(h, \widetilde{p})$. For instance, a fractional error info-gap model is:

$$\mathcal{U}(h, \widetilde{p}) = \left\{ p(x_1, x_2) \in \mathcal{P} : \ |p(x_1, x_2) - \widetilde{p}(x_1, x_2)| \le h \widetilde{p}(x_1, x_2) \right\}, \quad h \ge 0 \tag{84}$$

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

If we aspire to a probability of success no less than $P_{\rm c}$, then the robustness to uncertainty in $\widetilde{p}$, of decision based on drawing from $q(x)$, is:[5]

$$\widehat{h}(q, P_{\rm c}) = \max \left\{ h : \ \left(\min_{p \in \mathcal{U}(h, \widetilde{p})} 2 \int_{-\infty}^{\infty} \int_{x_2}^{\infty} P_{\rm s}(x_1, x_2, q) p(x_1, x_2)\,{\rm d}x_1\,{\rm d}x_2 \right) \ge P_{\rm c} \right\} \tag{85}$$

[5]The double integral in eq.(85) is multiplied by 2. In the double integral itself we assume that the first draw is greater than the second draw. Multiplying by 2 takes into account the other possibility.

where $P_s(x_1, x_2, Q) = 0.5[1 + Q(x_1) - Q(x_2)]$.

(c)[6] Let $H(\lambda, \gamma)$ denote the double integral (including the factor 2) in the definition of the robustness in eq.(85). One finds:

$$H(\lambda, \gamma) = \frac{1}{2} + \frac{\lambda\gamma}{(\lambda+\gamma)(2\lambda+\gamma)} \tag{86}$$

$$= \frac{1}{2} + \frac{\rho}{(1+\rho)(1+2\rho)}, \quad \rho = \frac{\lambda}{\gamma} \tag{87}$$

Differentiating we find:

$$\frac{\partial H(\lambda, \gamma)}{\partial \lambda} = \frac{\gamma(\gamma^2 - 2\lambda^2)}{(\lambda+\gamma)^2(2\lambda+\gamma)^2} \tag{88}$$

$H(\lambda, \gamma)$ vs. λ is a unimodal function with a maximum at $\lambda = \gamma/\sqrt{2}$; see fig. 3.

Let $\mu(h, \gamma)$ denote the inner minimum in eq.(85), which is the minimum of $H(\lambda, \gamma)$ as λ varies up to horizon of uncertainty h. $\mu(h, \gamma)$ is the inverse of $\widehat{h}(Q, P_c)$. That is:

$$\mu(h, \gamma) = P_c \quad \text{implies} \quad \widehat{h}(Q, P_c) = h \tag{89}$$

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

$$\lambda_1(h) = (1+h)\widetilde{\lambda} \tag{90}$$

$$\lambda_2(h) = (1-h)^+\widetilde{\lambda} \tag{91}$$

where $x^+ = x$ if $x \geq 0$ and equals zero otherwise.

Let us define the following two functions:

$$\mu_1(h) = H[(1+h)\widetilde{\lambda}, \gamma] \tag{92}$$

$$\mu_2(h) = H[(1-h)^+\widetilde{\lambda}, \gamma] \tag{93}$$

The inner minimum in the definition of the robustness, eq.(85), which is the inverse of the robustness function, is the lesser of the two functions defined in eqs.(92) and (93):

$$\mu(h) = \min_i \mu_i(h) \tag{94}$$

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

$$\gamma^\star = \arg\max_\gamma H(\widetilde{\lambda}, \gamma) \tag{95}$$

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

$$\frac{\partial H(\widetilde{\lambda}, \gamma)}{\partial \gamma} = \frac{\widetilde{\lambda}(2\widetilde{\lambda}^2 - \gamma^2)}{(\widetilde{\lambda}+\gamma)^2(2\widetilde{\lambda}+\gamma)^2} \tag{96}$$

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

$$\gamma^\star = \widetilde{\lambda}\sqrt{2} \tag{97}$$

Figs. 4–6 show robustness curves for different choices of γ, which determines the distribution of 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.(86), where $H = 1/2$ when $\lambda = 0$.

[6] The calculations are made with GapZapper. Domain: Homework-Problems, Application: PS1_RK#10.

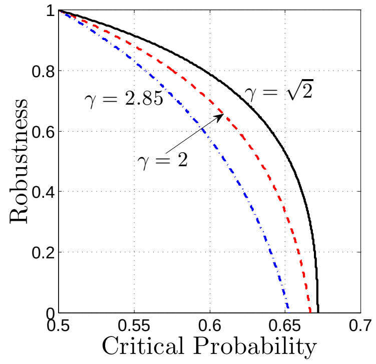

Figure 4: Robustness curves with $\widetilde{\lambda} = 1$.

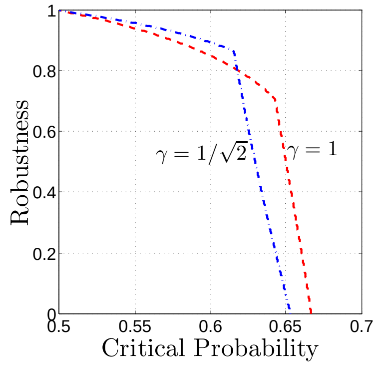

Figure 5: Robustness curves with $\widetilde{\lambda} = 1$.

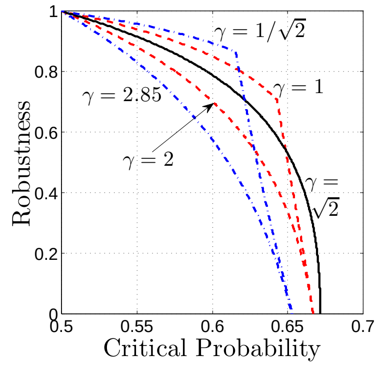

Figure 6: Robustness curves with $\widetilde{\lambda} = 1$.

In fig. 4 we examine values of γ for which $\mu(h)$ in eq.(94) takes only one functional form—$\mu_1(h)$ as we will see—for all horizons of uncertainty, so no kink occurs in the curve. The peak of $H(\lambda, \gamma)$ vs. λ (see fig. 3) occurs when $\lambda = \gamma/\sqrt{2}$, as seen from eq.(88). When $\gamma = \sqrt{2}$ (solid black curve) then, since $\widetilde{\lambda} = 1$, the value of $\mu(0)$ occurs at the peak of $H(\lambda, \gamma)$ vs. λ. As h increases, the value of $\mu(h)$ moves left, down the steep positive slope illustrated in fig. 3. In the other curves of fig. 4, $\widetilde{\lambda} < \gamma/\sqrt{2}$ so the value of $\mu(0)$ occurs on the steep positive slope of H vs λ, and, as h increases, the value of $\mu(h)$ moves left, down the steep positive slope.

From eq.(97) we see that $\gamma = \sqrt{2}$ is the nominal optimal choice since $\widetilde{\lambda} = 1$. Fig. 4 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. 5 is different from fig. 4: each robustness curve in fig. 5 displays a kink when $\mu(h)$ switches from one solution to the other as specified in eq.(94). $\widetilde{\lambda} > \gamma/\sqrt{2}$ in both cases, so $\mu(0)$ occurs on the gentle negative-slope portion of $H(\lambda, \gamma)$ vs. λ. Thus, for small h, $\mu(h)$ 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 $\mu(h)$ switches and moves left down the steep slope. This explains the kink in the robustness curves.

Fig. 6 combines the curves of figs. 4 and 5. 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 $P_{\rm c} = 0.65$. For greater critical probability, $\gamma = \sqrt{2}$ is more robust (up to $P_{\rm c} = 0.67$ 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 pdf of y).

(d)[7]

Figs. 7–9 show empirical and average theoretical probabilities of success, F and $\overline{P}_{\rm s}$. In all cases both the generating and the decision pdf's are normal. The generating pdf has mean and standard deviation of 1 and 0.5, respectively, in all cases. The deciding pdf has standard deviation of 1.5 and mean varying from -2 to 4. The number of (x_1, x_2) pairs, N, is 500, 5,000 and 50,000 in these three figures. Both F and $\overline{P}_{\rm s}$ are functions of the random draws, and these figures show the influence of the sample size. We also see that the highest probability of success occurs when the deciding pdf has a mean at or near the mean of the generating pdf.

Figs. 10–12 show empirical and average theoretical probabilities of success, F and $\overline{P}_{\rm s}$. In all cases the generating pdf is normal with mean and standard deviation of 1 and 0.5 respectively. The deciding pdf is exponential, $q(y) = \gamma e^{-\gamma y}$, with γ between 0.5 and 2. The mean of $q(y)$ is $1/\gamma$. The number of (x_1, x_2) pairs, N, is 500, 5,000 and 50,000 in these three figures. Both F and $\overline{P}_{\rm s}$ are functions of the random draws, and these figures show the influence of the sample size. The highest probability of success appears to occur for γ between 1 and 1.5, which implies a mean of the decision pdf between 0.67 and 1. The mean of the generating pdf is 1.

[7] The calculations are made with GapZapper. Domain: Homework-Problems, Application: PS1_RK#10simul01.

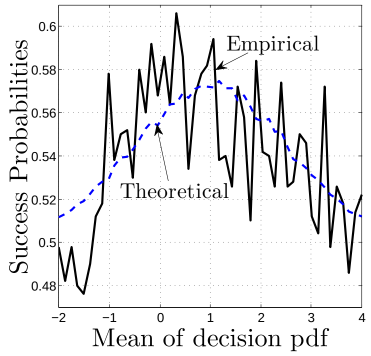

Figure 7: Empirical and average theoretical probabilities of success. $p \sim \mathcal{N}(1, 0.5)$. $q \sim \mathcal{N}(\gamma, 1.5)$, $\gamma \in [-2, 4]$. $N = 500$.

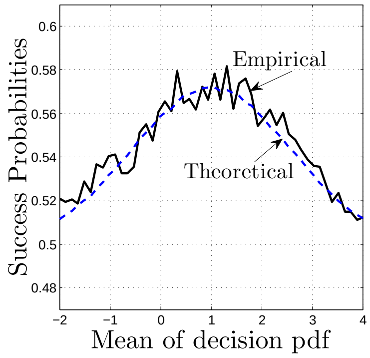

Figure 8: Same as fig. 7 with $N = 5,000$.

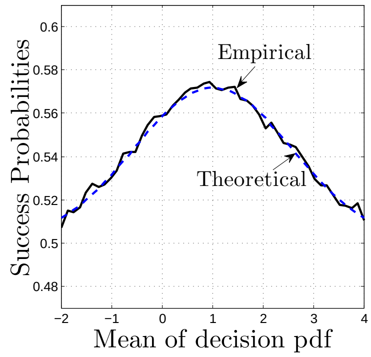

Figure 9: Same as fig. 7 with $N = 50,000$.

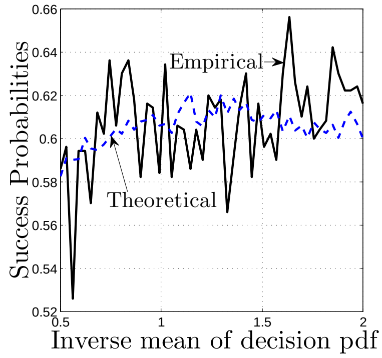

Figure 10: Empirical and average theoretical probabilities of success. $p \sim \mathcal{N}(1, 0.5)$. $q \sim \exp(\gamma)$, $\gamma \in [0.5, 2]$. $N = 500$.

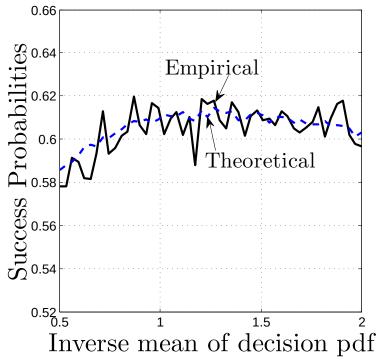

Figure 11: Same as fig. 10 with $N = 5,000$.

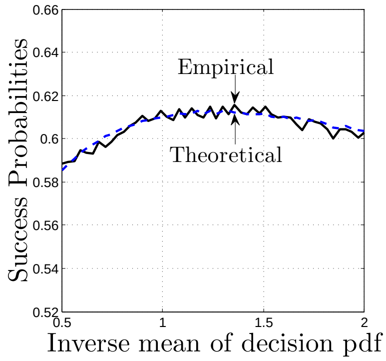

Figure 12: Same as fig. 10 with $N = 50,000$.

Note that both $p(x)$ and $q(y)$ are normal in figs. 7–9, though with different standard deviations. However, in figs. 10–12, $p(x)$ is normal and $q(y)$ is exponential. Nonetheless, P_s is larger in the latter cases than in the former.

Figs. 13–15 show empirical and theoretical success probabilities for various degrees of mis-match of the standard deviation between the generating and the decision pdf's. The generating and the decision pdf's are both normal. In fig. 13 the decision pdf has very small variance compared to the generating pdf, and the maximal success probability is about 0.75. In fig. 14 $p(x)$ and $q(y)$ have the same variance and the maximal success probability is 0.67. In fig. 15 the decision pdf has much larger variance than the generating pdf and the maximum success probability is about 0.56. We also see that the variability of the empirical success probability increases as the decision variance increases. Likewise, the success probability curve becomes broader as the variance of the decision pdf increases.

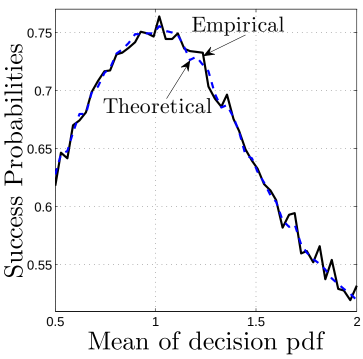

Figure 13: Empirical and average theoretical probabilities of success. $p \sim \mathcal{N}(1, 0.5)$. $q \sim \mathcal{N}(\gamma, s)$, $\gamma \in [0.5, 2]$, $s = 0.005$. $N = 5000$.

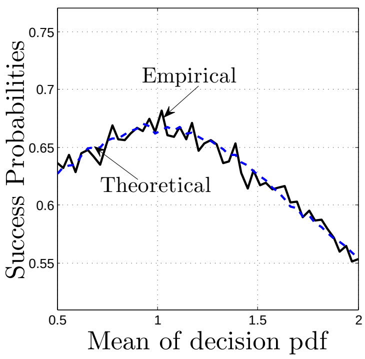

Figure 14: Same as fig. 13 with $s = 0.5$.

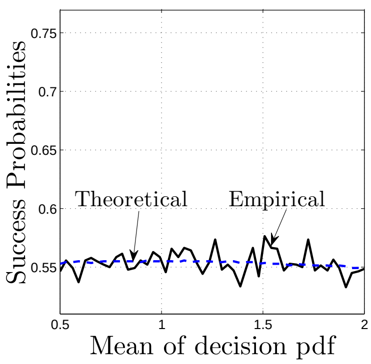

Figure 15: Same as fig. 13 with $s = 2$.