ENVIRONMETRICS
Environmetrics 2007; **18**: 189–203
Published online 18 September 2006 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/env.811

An info-gap approach to power and sample size calculations

David R. Fox[1*,†], Yakov Ben-Haim[2], Keith R. Hayes[3], Michael A. McCarthy[4], Brendan Wintle[4] and Piers Dunstan[3]

[1] *Australian Centre for Environmetrics, University of Melbourne, Melbourne 3010, Australia*
[2] *Technion—Israel Institute of Technology, Haifa, Israel*
[3] *CSIRO Division of Marine Research, GPO Box 1538, Hobart, Tasmania 7001, Australia*
[4] *The School of Botany, University of Melbourne, Melbourne, Australia*

SUMMARY

Power and sample size calculations are an important but underutilised component of many ecological investigations. A key problem with these calculations is the need to estimate or guess the effect size and error variance (the design parameters) prior to the actual data collection. Furthermore, calculations associated with statistical power and sample size are invariably predicated on normal distribution theory. While the central limit theorem ensures the applicability of normal-based inference for reasonably large sample sizes, the impact of violations of this assumed distributional form in the context of power and sample size determinations is rarely considered. This paper uses information-gap theory to provide sample size guidelines that are robust to uncertainties associated with both the design parameters and distributional form. A simple information-gap approach is developed for one- and two-sided hypothesis tests. The model results quantify the extent to which minimum power demands can be protected from uncertainty by taking additional samples, and demonstrate the importance of the combined effects of standard deviation/effect size ratio and assumed distribution in these considerations. Info-gap theory does not eliminate the need for an initial estimate or best guess of the design parameters or the specification of a parametric distribution from which to compute power. It does, however, measure the degree of insurance provided by additional samples in the face of uncertainties in each of these. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS: hypothesis testing; experimental design; environmental monitoring; robustness; parameter uncertainty

1. INTRODUCTION

Hypothesis testing has proved to be a popular and valuable way of assessing statements about environmental condition. A focus on 'compliance' is generally accompanied by a commensurate increase in attention to the related issues of power and sample-size (Lyles *et al*., 1997). These techniques are well known to researchers in many diverse disciplines including, but not limited to, biology, ecology, psychology, medicine, epidemiology, pharmacology and botany (Green, 1989, Fairweather, 1991, Cohen, 1992, Underwood and Chapman, 2003). The utility of power analysis

*Correspondence to: David R. Fox, Australian Centre for Environmetrics, University of Melbourne, Parkville Victoria 3010, Australia.
[†]E-mail: david.fox@unimelb.edu.au

Received 17 January 2006
Accepted 11 July 2006

(PA) is that it enables the researcher to assess the efficacy of a proposed monitoring and analysis strategy. Specifically, it quantifies the probability that a statistical test procedure will detect an effect of some prescribed magnitude when it in fact exists. The uptake and use of PA has been patchy. As noted by Hoenig (http://www.esi-topics.com/nhp/comments/september-02-JohnHoenig.html) PA has been widely advocated in journals and texts as a way of interpreting the results of statistical tests. However, operational and conceptual difficulties coupled with flawed advice on the use of post-hoc or retrospective PA (Reed and Blaunstein, 1995; Thomas and Juanes, 1996) as a way of resolving the 'dilemma of the non-rejected null'[1] (Hoenig and Heisey, 2001) continues to undermine the utility of statistical power calculations (Tversky and Kahenman, 1971; Oakes, 1986).

A particular operational difficulty is the circularity that inevitably arises when assigning important parameter values such as effect size and population error variance *prior* to the data collection effort. Most texts suggest that this is resolved by providing a best guess, utilising results from previous studies, or in some other way, estimating the unknown parameter. In environmental studies, it is particularly difficult to quantify an ecologically/biologically/environmentally meaningful or important 'effect size' due to the paucity of relevant data. Given the uncertainties in these critical inputs it is perhaps not surprising that the researcher is sometimes cautioned against attaching too much credence to the outputs (Fox, 2001). We agree with Lenth (http://www.stat.uiowa.edu/∼rlenth/Power/) that the specification of standardized, or 'T-shirt effect sizes' of 'small', 'medium' and 'large' is not the way to resolve this problem.

Another potential difficulty arises when conventional tools are used for undertaking a PA. Many standard formulae and software utilities for computing power and/or sample sizes are predicated on normal distribution theory. While the central limit theorem (CLT) often provides a 'safety-net' by its asymptotic guarantee of normality in the distribution of certain test statistics, reliance on it may be unfounded in many environmental studies (see, for example Watson and Downing, 1976). Violations of normality usually are most serious in preliminary assessments where effect sizes are 'large' (hence relatively small sample sizes are involved) and the variable of interest has large third and fourth moments (skewness and kurtosis). The impact of violations of the normality assumption on a PA is rarely considered.

In this paper we restrict our attention to mean-based inference. In particular, we have a sample of size n assumed to have been randomly selected from $X \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ for some random variable, X. Our interest centres on using the sample mean $\overline{X}$ to test hypotheses concerning the (unknown) population mean, μ. Either asymptotically or exactly, $\overline{X} \sim N(\mu, \sigma^2/n)$.

Our primary interest is in developing a simple information-gap analysis (Ben-Haim, 2006) for power calculations. The aim is to provide sample size guidelines that are robust to the uncertainty associated with the specification of design parameters (population error variance and ecologically significant effect size) as well as uncertainties in the distributional form of the test statistic. These guidelines can serve as an initial starting point for environmental scientists concerned about the statistical and ecological value of a proposed monitoring design.

In Section 2 we formulate the one- and two-tailed tests of interest. In Section 3 we develop the info-gap analysis for parameter uncertainty and in Section 4 we consider uncertainty in the sampling distribution.

[1]This refers to the situation where the failure of a statistical test to identify a 'significant' result is attributed to the test's low power.

DOI: 10.1002/env

2. PRELIMINARIES

2.1. Single-mean, one-tailed test

We commence with a simple, one-tailed hypothesis testing situation for a single population mean, μ defined by the pair of hypotheses in Equation (1).

$$\begin{aligned} H_0 &: \mu = \mu_0 \\ H_1 &: \mu > \mu_0 \end{aligned} \tag{1}$$

A size-α test rejects H_0 in favour of H_1 for values of the test statistic $\overline{X} > C^*(\alpha)$ where $\overline{X}$ is the mean of a sample of n observations and the 'critical-value' $C^*(\alpha)$ is chosen such that $P[\overline{X} > C^*(\alpha) | \mu = \mu_0] = \alpha$. It is readily verified that $C^*(\alpha)$ is given by Equation (2).

$$C^*(\alpha) = \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} \tag{2}$$

where z_α is the $(1-\alpha)$ 100 percentile of the standard normal distribution. Furthermore, it is evident from Figure 1a that the proportion of times a true null hypothesis would be rejected by this decision rule (Type I error) is precisely α—the so-called level of significance.

A second type of error is committed whenever the test procedure fails to reject a false null hypothesis (Type II error). Type II errors are particularly important in environmental sciences since this corresponds to a failure to identify an important environmental impact. The probability of a Type II error (denoted by β) is $P[\overline{X} < C^*(\alpha) | \mu = \mu_1] = \beta$ (Figure 1) where $\mu_1 > \mu_0$. Note that in general, $\alpha + \beta \neq 1$. To make explicit the dependency of β on the exact value of μ_1 we will use the notation $\beta(\mu_1)$.

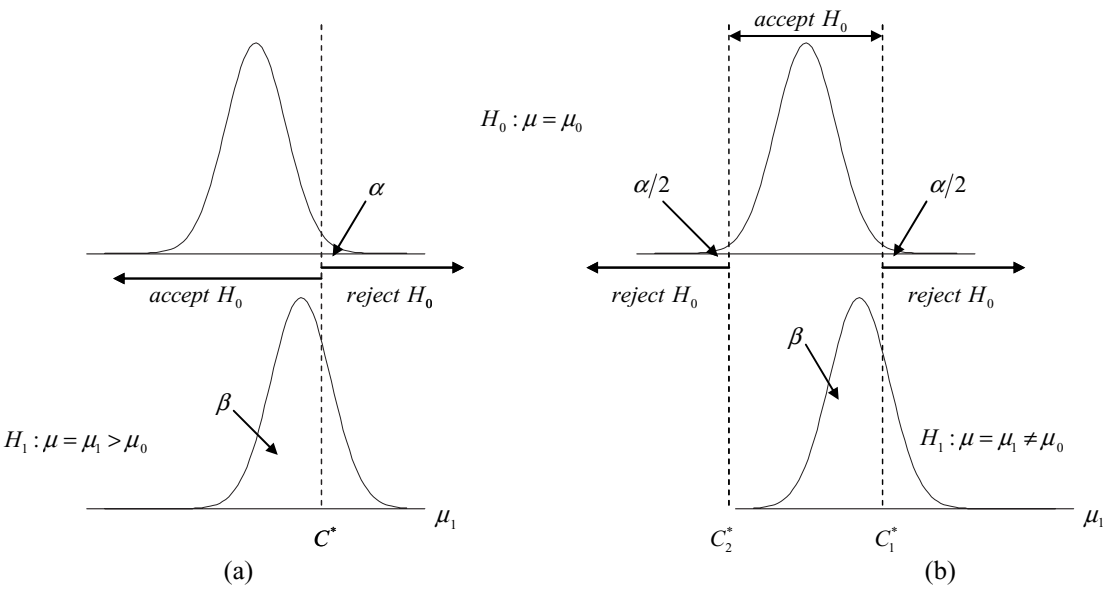

Figure 1. Critical regions and important probabilities for a one-tailed (a) and two-tailed (b) hypothesis test of a single population mean μ

DOI: 10.1002/env

An important quantity used to characterise the efficacy of the test procedure is its statistical power $1-\beta(\mu_1)$ which is the probability that an incorrect null hypothesis is correctly rejected. Power curves for the test defined by Equation (2) are obtained by evaluating the power at specific values of $\mu=\mu_1$ using Equation (3).

$$\text{Power}\,|_{\mu=\mu_1} = 1-\Phi\left[\frac{\mu_0-\mu_1}{\sigma/\sqrt{n}}+z_\alpha\right] \tag{3}$$

where Φ is the cumulative distribution function (cdf) of the standard normal distribution. For fixed sample size, n, the power is a monotonically increasing function of $|\Delta|$ where $\Delta=\mu_0-\mu_1$ is the effect size or minimum detectable difference.

2.2. Single-mean, two-tailed test

The mathematical development for the more general case of a two-sided test is essentially identical. An α-level test of the pair of hypotheses given in Equation (4) is defined by two critical regions $\overline{X}\geq C_1^*(\alpha)$ and $\overline{X}\leq C_2^*(\alpha)$ (Figure 1b).

$$\begin{aligned} H_0 &: \mu=\mu_0 \\ H_1 &: \mu\neq\mu_0 \end{aligned} \tag{4}$$

where

$$C_1^*(\alpha)=\mu_0+z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\quad C_2^*(\alpha)=\mu_0-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{5}$$

For a specified value $\mu=\mu_1$, the power of the test defined by Equation (5) is

$$\text{Power}\,|_{\mu=\mu_1} = 1-\Phi\left[\frac{\mu_0-\mu_1}{\sigma/\sqrt{n}}+z_{\alpha/2}\right]+\Phi\left[\frac{\mu_0-\mu_1}{\sigma/\sqrt{n}}-z_{\alpha/2}\right] \tag{6}$$

Despite advice to the contrary the computation of power for given design parameters $\{n,\alpha\}$, should *not* be used retrospectively to explain why a particular test result was non-significant (Hoenig and Heisey, 2001). The utility of these power considerations is at the *planning* stage of an investigation. As noted above, however, this utility is compromised by the need to specify *a priori* a biologically meaningful effect size ($\Delta=\mu_0-\mu_1$) and population standard deviation (σ). This is invariably difficult to do and many researchers simply opt to compute power for a range of possible effect sizes and/or standard deviation.

3. INFO-GAP FORMULATION FOR PARAMETER UNCERTAINTY

Information-gap (hereafter referred to as info-gap) theory is a recent development designed to assist decision makers faced with severe uncertainty (Carmel and Ben-Haim, 2005; Regan *et al*., 2005; Ben-Haim, 2006). Info-gap theory aims to address the 'robustness' of decision making under uncertainty. It asks the question: how wrong can a model and its parameters be without jeopardising the quality of decisions made on the basis of this model? Put another way it maps the point at which decisions should change in order to be robust to uncertainty in the parameters and functional forms of statistical, economic or bio-physical models.

DOI: 10.1002/env

Info-gap theory derives its robustness functions from three elements: a performance measure, a process model and a non-probabilistic model of uncertainty. The performance measure is a statistical, economic or bio-physical metric of value to the decision maker. The decision maker may wish to increase the performance measure (e.g. dollar value of a share portfolio) or reduce it (e.g. probability of extinction of an endangered species). In each case there is often a critical performance value which defines a change in decision. In our case, the performance measure is the *power* of the statistical test.

The process model is a mathematical summary of the system in question. It describes the relationship between the performance measure and the important characteristics of the system in question. In this example the performance threshold is the power of the statistical test, and the process models are the power equations for a one- and two-sided test, that is Equations (3) and (6). The critical performance measure in power calculations is often taken to be 0.8. This is an arbitrary but well accepted minimum for sample size calculations in biological sciences (Quinn and Keough, 2002).

The info-gap model of uncertainty for the uncertain quantities p in the process model is the unbounded family of nested sets $U(R,\tilde{p})$ of possible realisations p, where R represents the unknown 'horizon of uncertainty' and $\tilde{p}$ our best or initial estimate of p. This model satisfies two axioms:

$$\text{contraction}: \; U(0,\tilde{p}) = \{\tilde{p}\} \tag{7}$$

$$\text{nesting}: \; R < R' \Rightarrow U(R,\tilde{p}) \subset U(R',\tilde{p}) \tag{8}$$

The contraction axiom states that in the absence of uncertainty ($R = 0$), our best estimate $\tilde{p}$ is correct, while the nesting axiom states that the range of uncertain variation increases as the horizon of uncertainty increases. In all cases R is unknown and unbounded, $R \geq 0$. In this example the uncertain quantities are the effect size $\Delta = \mu_0 - \mu_1$ and the population standard deviation σ, such that $p = (\Delta, \sigma)$. Our initial or best estimate of the effect size and standard deviation is denoted $\tilde{p} = (\tilde{\Delta}, \tilde{\sigma})$.

Info-gap theory entertains numerous classes of non-probabilistic uncertainty models (Ben-Haim, 2006). An info-gap model is an unbounded family of nested sets of possible occurrences, such as uncertain parameters or functions. The structure of an info-gap model is chosen according to the available prior information and the nature of the uncertain entities. We will consider several specific info-gap models in this paper.

In this section we consider uncertain parameter values—effect size Δ and standard deviation σ. Let our best estimates of these parameters be $\tilde{\Delta}$ and $\tilde{\sigma}$, respectively. In the one-sided test of Equation (1), $\Delta < 0$ and only negative effect sizes are considered, but the fractional error of the estimate, $(|(\Delta - \tilde{\Delta})/\tilde{\Delta}|)$, is unknown. Likewise, the standard deviation must be positive but the fractional error of the estimate, $(|(\sigma - \tilde{\sigma})/\tilde{\sigma}|)$, is unknown. With this prior information we formulate the following fractional-error info-gap model:

$$U_1(R,\tilde{\Delta},\tilde{\sigma}) = \left\{ (\Delta,\sigma): \quad \begin{array}{l} (1+R)\tilde{\Delta} \leq \Delta \leq \min[0,(1-R)\tilde{\Delta}] \\ \max[0,(1-R)\tilde{\sigma}] \leq \sigma \leq (1+R)\tilde{\sigma} \end{array} \right\}, \quad R \geq 0 \tag{9}$$

This is an unbounded family of nested sets of (Δ, σ) values. The sets become more inclusive as the horizon of uncertainty, R increases. A worst case is not known so R is unbounded.

We will also consider two-sided tests in which the effect size can be either positive or negative. In this case the fractional-error info-gap model becomes:

$$U_2(R,\tilde{\Delta},\tilde{\sigma}) = \left\{ (\Delta,\sigma): \quad \begin{array}{l} \left|\frac{\Delta-\tilde{\Delta}}{\tilde{\Delta}}\right| \leq R \\ \max[0,(1-R)\tilde{\sigma}] \leq \sigma \leq (1+R)\tilde{\sigma} \end{array} \right\}, \quad R \geq 0 \tag{10}$$

DOI: 10.1002/env

The definition of the performance measure, process model and uncertainty model(s) completes the specification of the formulation of the info-gap analysis.

We now turn to the derivation of the robustness function. In info-gap parlance 'robustness' is defined as the greatest horizon of uncertainty, across all uncertain model components, that still meets the pre-defined performance requirement. In our application the robustness of a size α test based on a sample of size n, is the greatest horizon of uncertainty $\hat{R}$ for which all combinations of the uncertain parameters $p = (\Delta, \sigma)$ achieve the minimum required power, that is

$$\hat{R}(n, \beta_c) = \max\left\{R : \left(\min_{(\Delta,\sigma)\in U(R,\tilde{\Delta},\tilde{\sigma})} \text{power}(\Delta, \sigma, n) \geq 1 - \beta_c\right)\right\} \tag{11}$$

where β_c is the critical value of β such that the power of the test is greater than or equal to the minimum required power, for example $1 - \beta_c = 0.8$. Equation (11) is the robustness function for this application of the info-gap model. The strategy of robust-satisficing (Ben-Haim, 2006) is to attempt to guarantee an adequate level of power, by choosing a value of n which is highly robust to uncertainty. Thus, for any given sample size, n, the robustness function indicates the confidence in attaining desired power with that n.

Examination of the process models (Equations (3) or (6)) and the uncertainty models (Equations (9) or (10)) reveals that power decreases as the standard deviation increases, and attains a minimum at uncertainty R when $\sigma = (1 + R)\tilde{\sigma}$. For the one-sided case power decreases as the effect size decreases in absolute value, that is $\Delta = \min[0, (1 - R)\tilde{\Delta}]$. The same effect occurs for the two-sided case, that is power is minimised for $\Delta = (1 - R)\tilde{\Delta}$ but only over the interval $0 \leq R \leq 1$.

Combining the performance measure, process model and uncertainty models allows us to re-write the inner minimum of the robustness function (Equation (11)) as

$$1 - \beta_c \leq 1 - \Phi[h(R, \tilde{\Delta}, \tilde{\sigma}, n) + z_\alpha] \tag{12}$$

for the one-sided case, and

$$1 - \beta_c \leq 1 - \Phi[h(R, \tilde{\Delta}, \tilde{\sigma}, n) + z_{\alpha/2}] + \Phi[h(R, \tilde{\Delta}, \tilde{\sigma}, n) - z_{\alpha/2}] \tag{13}$$

for the two-sided case, where (in both cases)

$$h(R, \tilde{\Delta}, \tilde{\sigma}, n) = \frac{(1 - R)\tilde{\Delta}\sqrt{n}}{(1 + R)\tilde{\sigma}}, \tag{14}$$

and $R \leq 1$. For $R > 1$ the info-gap model in Equations (9) and (10) imply that $h(R, \tilde{\Delta}, \tilde{\sigma}, n) = 0$.

Equations (12) and (13) can be easily solved numerically which is the approach adopted here. Specifically, a plot of R on the horizontal axis versus the right-hand side of Equations (12) or (13) on the vertical axis is precisely a plot of the robustness $\hat{R}(n, \beta_c)$ versus the power. We now proceed to an example.

3.1. Illustrative example

Figures 2 and 3 plot the robustness function of the one- and two-sided tests for various values of the standardized effect size Δ/σ and for various sample sizes with $\alpha = 0.05$. The vertical axis is the demanded power, $(1 - \beta_c)$, and the horizontal axis is the robustness $\hat{R}(n, \beta_c)$ for sample size n and level of significance α (which is specified by z_α or $z_{\alpha/2}$ in Equations (12) and (13)). That is, the estimated

DOI: 10.1002/env

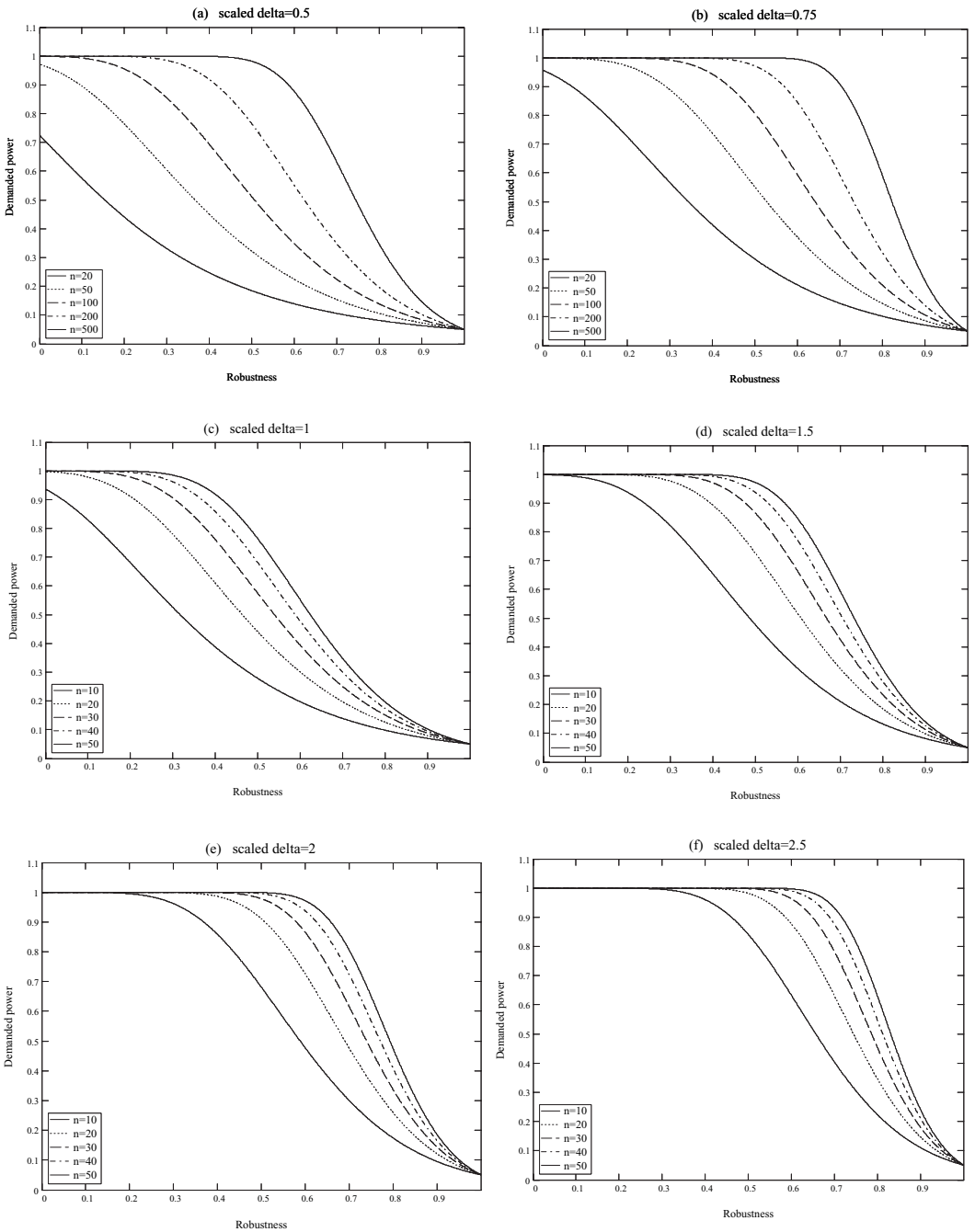

Figure 2. Robustness of power calculations for one-sided hypothesis tests

DOI: 10.1002/env

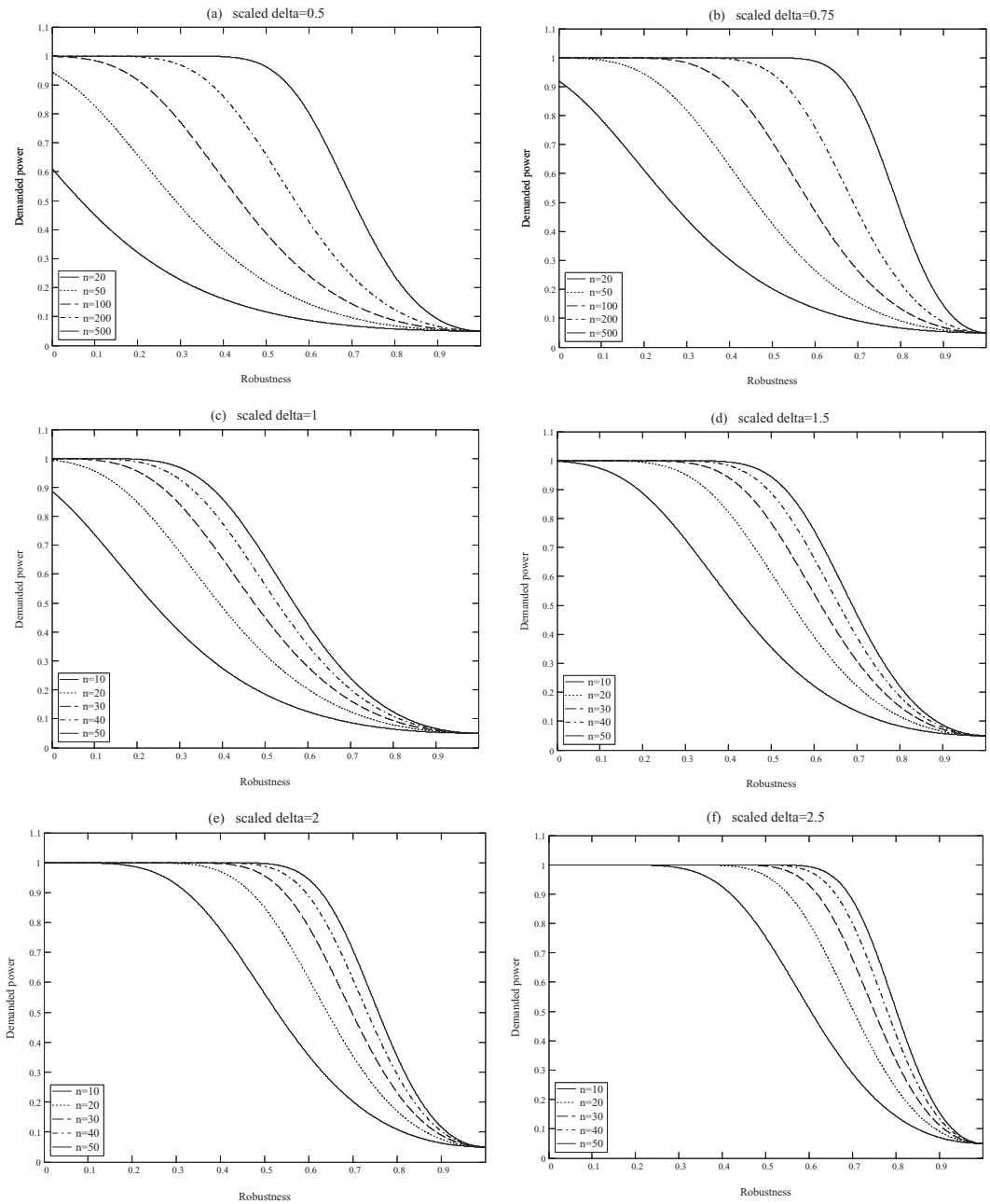

Figure 3. Robustness of power calculations for two-sided hypothesis tests

DOI: 10.1002/env

effect size and standard deviation, $\tilde{\Delta}$ and $\tilde{\sigma}$, can err by a fraction as large as $\hat{R}(n, \beta_c)$ without jeopardizing the requirement that the power exceed $(1 - \beta_c)$ at level of significance α.

These curves accord with intuition. For example, for any required level of power, immunity to parameter and knowledge uncertainty (robustness) is obtained by taking a larger sample size. Conversely, for any fixed level of robustness, the power of the statistical test increases as more samples are taken. The curves are plotted for $0 \leq R \leq 1$. For $R > 1$ we have $h = 0$ and the power which can be achieved is constant at the value reached with $R = 1$.

Furthermore, as robustness approaches unity, the statistical power approaches the level of significance (which is $\alpha = 0.05$ in these figures).

More importantly, these results measure how much robustness to uncertainty can be bought by additional samples, and demonstrate the opposing influence of effect size and standard deviation on power and robustness. Figure 2c, for example shows that for the one-sided case, a minimum power demand of 0.8 can be met with 12% robustness with only 10 samples so long as the standard deviation is approximately the same as the effect size. For a further 40 samples the design parameters can be up to 48% wrong and still achieve the same minimum power demand (a 36% increase in robustness). The same is true for the two-sided case, except in this instance the additional uncertainty associated with the direction of change in effect reduces the robustness to uncertainty in the initial estimates for sample sizes.

Collectively, Figures 2 and 3 illustrate the result of increasing effect size over the standard deviation, that is the absolute value of the standardised effect size >1. If the effect is twice as large as the standard deviation, then a minimum power demand of 0.8 is satisfied with a robustness of 0.44 for the one-tailed test (Figure 2e) and robustness of 0.39 for the two-tailed test (Figure 3e) using $n = 10$. A further 40 samples only purchases about a 26% increase in robustness for either the one- or two-tailed tests. Clearly the rate of return diminishes rapidly for both power and robustness to uncertainty as the magnitude of the standardised effect size increases—it is easy to spot big 'standardised' changes. Figures 2a,b, and 3a,b illustrate the result of increasing the standard deviation over the effect size, that is the absolute value of the standardised effect size <1. In these cases the manager must work very hard to achieve minimum power. For example, for a one-tailed test for which the standard deviation is twice as large as the effect size (Figure 2a), a sample of $n = 50$ has robustness of only 0.17 for demanded minimum power of 0.8. To increase the manager's immunity to uncertainty to a modest 0.5 would require the purchase of an additional 173 samples! As the standard deviation increases sample sizes must be very large to achieve minimum power with any level of robustness (Figures 2a,b and 3a,b).

4. INFO-GAP FORMULATION FOR DISTRIBUTIONAL UNCERTAINTY

While the preceding formulation and analysis affords useful insights into the info-gap formulation for dealing with uncertainties in *design* parameters $\{\Delta, \sigma\}$, it ignores uncertainties in the assumed distribution that underpins the power calculations. As previously remarked, the CLT asymptotically guarantees the applicability of the normal distribution but in environmental studies, situations arise where assumed normality can be tenuous. This naturally leads us to consider the effects of violations of the normality assumption on statistical power. The next section develops an info-gap approach to this problem for the one-tailed hypothesis test (Equation (1)). The extension to the two-tailed case (Equation (4)) is then straightforward.

DOI: 10.1002/env

4.1. Single mean, one-tailed test

As noted in Section 2.1, the critical region for this α-level test is $\overline{X} > C^*(\alpha)$ where $C^*(\alpha)$ is determined such that $P[\overline{X} > C^*(\alpha)|\mu = \mu_0] = \alpha$. Hereafter $f(x;\ \mu)$ will denote the *pdf* of $\overline{X}$ evaluated at x. We assume this distribution has mean μ. Furthermore, we will assume that $f(x;\ \mu_1) = f(x - \delta;\ \mu_0)$. That is, the distribution under H_1 equals the distribution under H_0 shifted to the right by a distance δ. Thus for any level of significance α, non-negative effect size δ, and *pdf* $f(x;\ \mu)$, we have

$$1 - \alpha = \int_{-\infty}^{C^*(\alpha)} f(x; \mu_0)\mathrm{d}x \tag{15}$$

and

$$\beta(f) = \int_{-\infty}^{C^*(\alpha)} f(x - \delta; \mu_1)\mathrm{d}x = \int_{-\infty}^{C^*(\alpha)-\delta} f(x; \mu_0)\mathrm{d}x = 1 - \alpha - \int_{C^*(\alpha)-\delta}^{C^*(\alpha)} f(x; \mu_0)\mathrm{d}x \tag{16}$$

For the info-gap formulation, we consider the sampling distribution of the test statistic to be uncertain (e.g. violations of the CLT due to small sample size; a parent distribution having extreme third and/or fourth moments). Our objective is to determine a sample size that provides an adequate level of robustness to uncertainty in the sampling distribution while guaranteeing a minimum power requirement will be met. Thus, we assume that the actual sampling distribution is unknown, but our best estimate is $\tilde{f}(x)$, which as before depends on the sample size, n. We assume that $\tilde{f}(x)$ is not flat anywhere in its domain. As in previous sections, we will assume a fractional-error info-gap model:

$$U(R,\tilde{f}) = \{f(x) : f \in P, |f(x) - \tilde{f}(x)| \leq R\tilde{f}(x)\}, \quad R \geq 0, \quad x \in \Re \tag{17}$$

where P is the set of all *pdfs* on the domain of x. We assume also that the elements of $U(R,\tilde{f})$ are such that $\beta(f)$ in Equation (16) is continuous in $f(x)$.

Ideally, we would like β to be small. However, since the sampling distribution is uncertain we cannot guarantee that the power $(1 - \beta(\mu_1))$ calculated using Equation (16) is correct. Let $1 - \beta_\mathrm{d}$ be the power which is demanded by the analyst. That is, the analyst requires $\beta(\mu_1) \leq \beta_\mathrm{d}$. The robustness associated with a sample of size n, with the requirement that $\beta(\mu_1)$ be no greater than β_d, is the greatest horizon of uncertainty R up to which all *pdfs* in $U(R,\tilde{f})$ guarantee $\beta(\mu_1) \leq \beta_\mathrm{d}$. That is

$$\hat{R}(n, \beta_\mathrm{d}) = \max\left\{R : \left(\max_{f \in U(R,\tilde{f})} \beta(f)\right) \leq \beta_\mathrm{d}\right\} \tag{18}$$

For convenience, we denote the inner maximum in Equation (18) by $\gamma(R)$. Thus the robustness is the greatest value of R such that $\gamma(R) \leq \beta_\mathrm{d}$. Since the uncertainty sets $U(R,\tilde{f})$ are nested with respect to R, we see that $\gamma(R)$ increases as R increases. Hence the robustness is the greatest value of R at which $\gamma(R) = \beta_\mathrm{d}$. This robustness, obtained from $\gamma(R)$, is evaluated for a specified value of β_d and $C^*(\alpha)$. We use $\gamma(R)$ to find the critical value, $C^*(\alpha)$ whose robustness is maximal at the specified β_d. Thus, the inner maximum in Equation (18) requires us to find the maximum of $\beta(f)$ on $U(R,\tilde{f})$ at fixed R. That is, we must solve the following optimisation problem:

$$\hat{f} = \arg\max_f \beta(f) \quad \text{such that} \quad \hat{f} \in P, \quad (1 - R)\tilde{f}(x) \leq \hat{f} \leq (1 + R)\tilde{f}(x) \tag{19}$$

We will develop an explicit algorithmic solution to this sub-problem. Denote by $\beta(\hat{f}|C^*(\alpha))$ that β obtained using the *pdf* identified in Equation (19) and denote this *pdf* by $\hat{f}(x|C^*(\alpha))$. Note that by

DOI: 10.1002/env

Equation (15), for fixed R, each $C^*(\alpha)$ and $\hat{f}$ corresponds to a unique α, denoted $\alpha(C^*)$ which can be inverted to obtain $C^*(\alpha)$ and for which

$$\gamma(R) = \beta(\hat{f}(x)|\hat{C}^*(\alpha)) \tag{20}$$

Solution to the sub-problem in Equation (19).

Note that $\beta(f)$ in Equation (16) is a linear and continuous function of $f(x)$, so its extreme values, at uncertainty R, occur on the boundary of $U(R,\tilde{f})$ (Kelly and Weiss, 1979, theorem 13, p. 209). Thus, the choice of $f(x)$ which maximizes $\beta(f)$ will switch between the upper envelope $(1+R)\tilde{f}(x)$ and the lower envelope $(1-R)\tilde{f}(x)$.[2] In order to maximize $\beta(f)$ on $U(R,\tilde{f})$ it is necessary to choose f as large as possible when $\overline{x} \leq C^* - \delta$ and as small as possible elsewhere, recalling that f is a proper *pdf*. It is therefore necessary to determine where the *pdf* switches between the lower and upper envelopes. To this end, define the following 'upper cut set' for the nominal *pdf*:

$$X(y) = \{x : \tilde{f}(x) \geq y \text{ and } x \leq C^*(\alpha) - \delta\} \tag{21}$$

In order to maximize $\beta(f)$, $\hat{f}$ needs to be 'large' in $X(y)$ for some y, and contained in $U(R,\tilde{f})$. $X(y)$ is the set of x values in $(-\infty, C^*(\alpha) - \delta]$ for which $\tilde{f}(x)$ is larger than y.

Next define the following partial complement of $X(y)$:

$$X_2(y) = (-\infty, C^*(\alpha) - \delta)] - X(y) \tag{22}$$

$X(y)$ is a set of x values in $(-\infty, C^*(\alpha) - \delta]$ on which $\tilde{f}(x)$ is larger than on the partial complement $X_2(y)$. That is:

$$\min_{x \in X(y)} \tilde{f}(x) \geq \max_{x \in X_2(y)} \tilde{f}(x) \tag{23}$$

Choose a value y_s defined implicitly as:

$$\int_{X(y_s)} \tilde{f}(x)\mathrm{d}x = \frac{1}{2} \tag{24}$$

Now y_s is unique since $\tilde{f}$ is nowhere flat. If $\tilde{f}(x)$ is uni-modal then $X(y_s)$ is a single interval. If $\tilde{f}$ is multi-modal the $X(y_s)$ may contain disjoint intervals. We make no assumption about the modality of $\tilde{f}$. We denote integrals such as Equation (24) by $\tilde{F}[X(y_s)]$ where $\tilde{F}$ is the cdf corresponding to $\tilde{f}$. Next, consider the following *pdf*:

$$\hat{f}(x|C^*(\alpha)) = \begin{cases} (1+R)\tilde{f}(x) & \text{if } x \in X(y_s) \\ (1-R)\tilde{f}(x) & \text{if } x \in X_2(y_s) \\ (1-R)\tilde{f}(x) & \text{if } x > C^*(\alpha) - \delta \end{cases} \tag{25}$$

where, from the definitions of $X(y)$ and $X_2(y)$ in Equations (21) and (22) and the choice of y_s in Equation (24), one can show that:

$$\tilde{F}[X_2(y_s)] + [1 - \tilde{F}(C^*(\alpha) - \delta)] = \tilde{F}[X(y_s)] \tag{26}$$

[2]We need only consider values of $R \leq 1$ in order to cover relevant values of demanded power.

DOI: 10.1002/env

Equation (26) guarantees that $\hat{f}$ is a proper *pdf* because the $\pm R\tilde{f}(x)$ terms in Equation (25) cancel out upon integration. Hence $\hat{f}$ belongs to the info-gap model at uncertainty R.

The *pdf* in Equation (25) attains the greatest possible values in $(-\infty, C^*(\alpha) - \delta]$ which are allowed by $U(R,\tilde{f})$, and it does so on the set of largest possible measure.

It is evident that $\hat{f}(x|C^*(\alpha))$ in Equation (25) is the solution of the optimisation problem in Equation (19). To understand this we note the following points. First, no *pdf* in $U(R,\tilde{f})$ can equal $(1+R)\tilde{f}(x)$ on a set whose integral on $\tilde{f}$ is greater than ½ since it would not integrate to unity. In order for $\tilde{f}(x)$ to be a proper *pdf*, it is necessary that the $+R$ term cancel out the $-R$ terms. That is:

$$\int_{X(y_s)} \tilde{f}(x)\mathrm{d}x = \int_{X_2(y_s)} \tilde{f}(x)\mathrm{d}x + \int_{C^*(\alpha)-\delta}^{\infty} \tilde{f}(x)\mathrm{d}x \tag{27}$$

However, from the definition of the sets $X(y)$ and $X_2(y)$, and since $\tilde{f}(x)$ is a *pdf*:

$$1 = \int_{X(y_s)} \tilde{f}(x)\mathrm{d}x + \int_{X_2(y_s)} \tilde{f}(x)\mathrm{d}x + \int_{C^*(\alpha)-\delta}^{\infty} \tilde{f}(x)\mathrm{d}x \tag{28}$$

Thus, if the integral on the left-hand side of Equation (27) exceeds ½, then Equation (28) is violated. Second, y_s has been chosen so that $\hat{f} = (1+R)\tilde{f}$ on a set whose integral on $\tilde{f}$ is equal to ½. Third, this set has the largest possible values of density as shown in Equation (23). Thus, $(x|C^*(\alpha))$ maximizes $\beta(f)$ subject to Equation (19). The value of $\beta(f)$ for the *pdf* in Equation (25), for this value of $C^*(\alpha)$, is:

$$\beta(\hat{f}|C^*(\alpha)) = \tilde{F}(C^*(\alpha)-\delta) + R\tilde{F}[X(y_s)] - R\tilde{F}[X_2(y_s)] = \tilde{F}(C^*(\alpha)-\delta) + R\left(\frac{1}{2} - \tilde{F}[X_2(y_s)]\right) \tag{29}$$

where we have used the fact that $\tilde{F}[X(y_s)] = 1/2$. Furthermore, we see that $\beta(f)$ increases as R increases. This expression for $\beta(\hat{f}|C^*(\alpha))$ is precisely $\gamma(R)$ in Equation (20). As explained in connection with Equation (20), the robustness is the greatest value of R for which $\gamma(R)$ in Equation (20) does not exceed β_d.

4.2. Illustrative example

We consider the one-tailed hypothesis testing situation given by Equation (1) and without loss of generality, set $\mu_0 = 0$. We initially assume known design parameter values of $\sigma = 1$ and $\delta = 0.5$. The effect of uncertainty in the nominal *pdf* (assumed to be normal) for small sample sizes is investigated by the application of Equations (25) and (29) for $n = 3$, 4 and 5 (Figure 4). It is evident from Figure 4 that the power to detect an increase of 0.5 in μ with these small sample sizes is low. Nevertheless, this situation is not atypical in ecological applications such as monitoring rare or threatened species. If the attribute (random variable) being observed is non-normally distributed in the population, then the assumption of normality for the sampling distribution of the test statistic $(\overline{X})$ will be erroneous, resulting in potentially large discrepancies between the *nominal* and *actual* type I and type II errors. The utility of Figure 4 in such situations is that the researcher can assess the immunity to violation of the normality assumption gained by the 'purchase' of extra samples. For example, at a nominal power of 0.25, the robustness to severe uncertainty in the normality assumption increases from 0.1 with $n = 4$ to approximately 0.4 with $n = 5$.

DOI: 10.1002/env

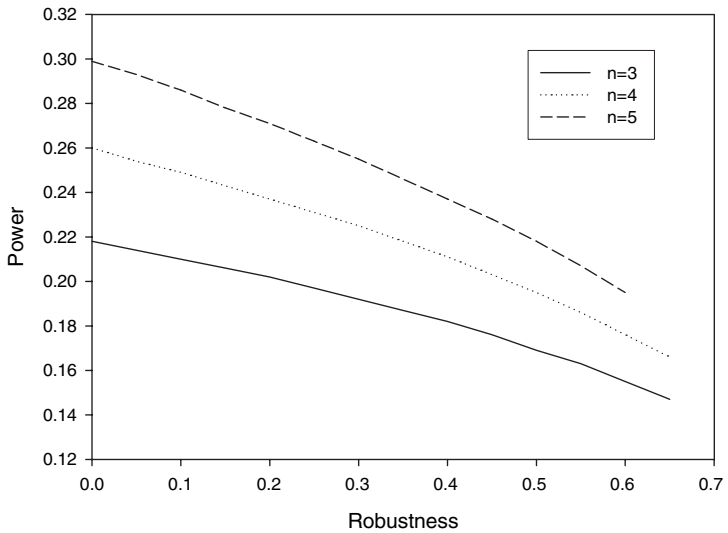

Figure 4. Robustness of power calculations for uncertainty in *pdf* for one-sided test

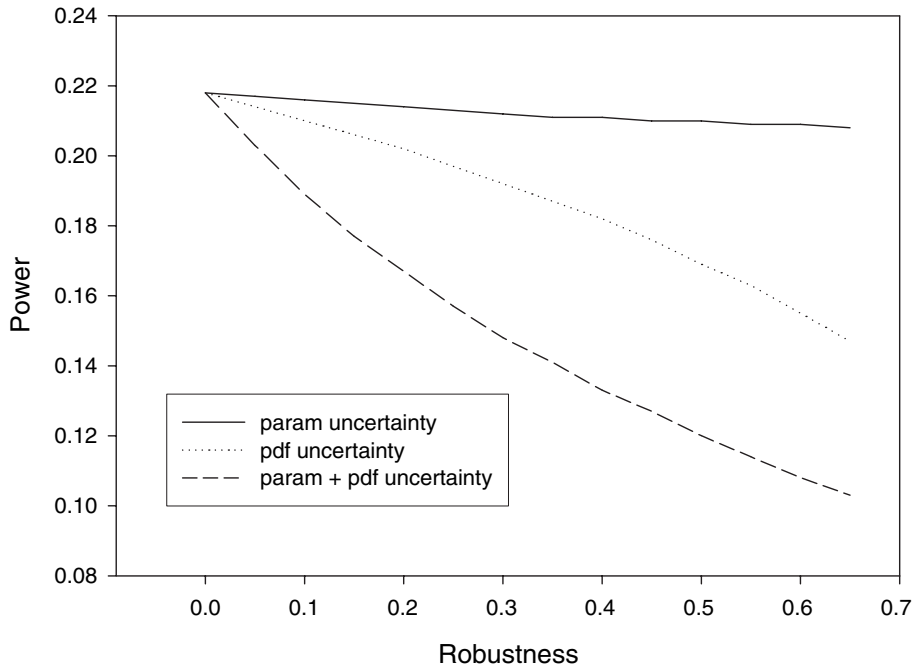

Figure 5. Robustness of power calculations for both parameter uncertainty and *pdf* uncertainty for one-sided test

DOI: 10.1002/env

In concluding this section, the *combined* effect of parameter *and* distributional uncertainty is investigated for the previous example with $n = 3$ (Figure 5). It can be seen from Figure 5 that the impact of distributional uncertainty is uniformly greater than parameter uncertainty and, perhaps more importantly, the combined effect is greater than the sum of the individual effects.

5. CONCLUSIONS

This paper introduces a new dimension to conventional power and sample size calculations. Info-gap theory was developed to provide a logical and consistent approach to quantifying the impact on assumed optimality arising from severe uncertainty in model parameters and functions (Ben-Haim, 2006). Initial info-gap applications focused on engineering design problems. Its application to statistical power and sample size analysis (PSA), particularly in the context of environmental sampling, appears here for the first time. We believe it is a potentially powerful adjunct to conventional practice given the pervasiveness and 'fragility' of PSA. This fragility arises from the fact that the results of PSA are likely to be severely wrong when there is high uncertainty in critical parameters (such as the population variance and the effect size) and/or extreme violations of the assumed normal distribution. The latter situation is characteristic of many ecological studies where small sample sizes are the norm and the variable of interest is decidedly non-normal. The salient feature of the info-gap approach is that it explicitly acknowledges these problems thereby allowing the researcher to assess 'how wrong' he or she can be in the specification of an effect-size, say while still meeting some demanded power criterion.

While our approach may not be perfect and some residual conceptual difficulties remain[3] we nevertheless believe these are outweighed by the additional insights gained by the explicit recognition and treatment of uncertainty at all levels of PSA. If nothing else, we believe the info-gap approach will engender a greater awareness among environmental researchers of the potentially seriously flawed decision-making that can arise from the routine application of conventional PSA.

ACKNOWLEDGEMENTS

This paper was based on the outcomes of the workshop 'Decision making for complex problems in conservation' conducted by the ARC Centre of Excellence for Mathematics and Statistics of Complex Systems (MASCOS) at the University of Melbourne, February 27th–March 4th, 2005. The authors are particularly indebted to Prof. Mark Burgman of the University of Melbourne for his inspiration and facilitation of the workshop. We also thank the anonymous referee whose comments on an earlier draft greatly improved the final version.

REFERENCES

Ben-Haim Y. 2006. *Information-gap decision theory: Decisions under severe uncertainty* (2nd Edition). Academic Press: San Diego.

Carmel Y, Ben-Haim Y. 2005. Info-gap robust-satisficing model of foraging behavior: Do foragers optimize or satisfice? *American Naturalist* **166**: 633–641.

Cohen J. 1992. A power primer. *Psychological Bulletin* **112**(1): 155–159.

Fairweather PG. 1991. Statistical power and design requirements for environmental monitoring. *Australian Journal of Marine and Freshwater Research* **42**: 555–567.

Fox DR. 2001. Environmental power analysis—a new perspective. *Environmetrics* **12**: 437–449.

Green RH. 1989. Power analysis and practical strategies for environmental monitoring. *Environmental Research* **50**: 195–205.

[3]For example, a violation of the distributional assumption would no doubt mean that the prescribed test statistic would relinquish 'optimal' statistical properties such as being uniformly most powerful (UMP).

DOI: 10.1002/env

Hoenig JM, Heisey DM. 2001. The abuse of power: The pervasive fallacy of power calculations for data analysis. *The American Statistician* **55**(1): 19–24.
Kelly PJ, Weiss ML. 1979. *Geometry and convexity: A study in mathematical methods*. Wiley: New York.
Lyles RH, Lawrence LL, Rappaport SM. 1997. Assessing regulatory compliance of occupational exposures via the balanced one-way random effects ANOVA model. *Journal of Agricultural, Biological, and Environmental Statistics* **2**(1): 64–86.
Oakes M. 1986. *Statistical inference: A commentary for the social and behavioural sciences*. Wiley: New York.
Quinn G, Keough M. 2002. *Experimental design and data analysis for biologists*. Cambridge University Press.
Reed JM, Blaunstein AR. 1995. Biologically significant population declines and statistical power. *Conservation Biology* **11**(1): 281–282.
Regan HM, Ben-Haim Y, Langford B, Wilson WG, Lundberg P, Andelman SJ, Burgman MA. 2005. Robust decision making under severe uncertainty for conservation management. *Ecological Applications* **15**(4): 1471–1477.
Thomas L, Juanes F. 1996. The importance of statistical power analysis: An example from Animal Behaviour. *Animal Behaviour* **52**: 856–859.
Tversky A, Kahenman D. 1971. Belief in the law of small numbers. *Psychological Bulletin* **76**: 105–110.
Underwood AJ, Chapman MG. 2003. Problems of measuring biodiversity in coastal habitats: A summary of issues. In *Conserving Marine Environments. Out of Sight Out of Mind*, Hutchings P, Lunney D (eds). Royal Zoological Society of New South Wales: Mosman, NSW; 103–107.
Watson D, Downing PB. 1976. Enforcement of environmental standards and the central limit theorem. *Journal of the American Statistical Association* **71**(355): 567–574.

DOI: 10.1002/env