RISK ANALYSIS AND SENSITIVITY ANALYSIS:

antithesis or synthesis?

Dr. Jack P.C. Kleijnen,
Professor of Simulation and
Management Informatics
Department of Business and Economics
Tilburg University (Katholieke Hogeschool Tilburg)
5000 LE Tilburg
Netherlands

Abstract

If a dynamic model assumes parameters constant over time, then the posterior mean (i.e. the mean conditional on specific values of these parameters) is relevant. Since parameters are unknown, they must be estimated. Sensitivity analysis quantifies the effects of incorrectly specified values of the parameters. If these effects are important then additional information on the parameters might be collected; otherwise robust solutions are to be sought. If these options do not work then risk analysis can quantify the probability of specific outputs, incorporating the probability distribution of the estimated parameters. Sensitivity analysis changes the values of parameters systematically, whereas risk analysis samples the parameter values. Simple queuing and econometric examples illustrate the two approaches.

Introduction

The present note was prompted by a discussion I had with a colleague involved in econometric modeling. He was contemplating what in the Management Sience jargon would be called "risk analysis", whereas I proposed to him to perform a sensitivity analysis instead. In the statistical jargon the two approaches concentrate on the estimation of the "prior" mean and the "posterior" mean respectively (see below).

Example: A Queuing Simulation

Let $\underline{X}_1$ and $\underline{X}_2$ denote stochastic interarrival time and service time respectively, and let $\underline{Y}$ denote waiting time, where I underline stochastic variables because the distinction between a stochastic variable and its realization or its distribution characteristics is crucial in this note. Then:

$$\underline{Y}_t = \max(\underline{Y}_{t-1} + \underline{X}_{2t} - \underline{X}_{1t}, 0) \tag{1}$$

Assume further that $\underline{X}_1$ and $\underline{X}_2$ are mutually independent and that $\underline{X}_1$ (respectively $\underline{X}_2$) is independently sampled from an exponential distribution with parameter λ (respectively μ). The simulation's purpose is the estimation of ω , the expected value (E) of the response ($\underline{Y}$) in the steady-state:

$$\omega = E(\underline{Y}_t) \text{ for } t \to \infty \tag{2}$$

If the simulation program is succinctly denoted by g and the vector of random numbers by $\underset{\sim}{\underline{r}}$, then ω is estimated by

$$\underline{\hat{\omega}} = g(\lambda, \mu, \underset{\sim}{\underline{r}}) \tag{3}$$

Within a specific simulation run λ is a <u>constant</u> (and so is μ but for simplicity μ is not discussed explicitly below).

Actually, the parameter λ is not known with certainty. Suppose λ is estimated from a sample of n independent interarrival times $\underline{X}_1$:

$$\underline{\hat{\lambda}} = \sum_{i=1}^{n} \underline{X}_{1i}/n \tag{4}$$

so that

$$\mathrm{var}(\underline{\hat{\lambda}}) = \mathrm{var}(\underline{X}_1)/n \tag{5}$$

Because of the Central Limit Theorem I assume normality (N):

$$\underline{\hat{\lambda}} \sim N(\lambda, \sigma_1^2/n) \qquad \text{where } \sigma_1^2 \equiv \mathrm{var}(\underline{X}_1) \tag{6}$$

Observe that if no data on $\underline{X}_1$ were available, then eq. (6) would be replaced by a <u>subjective</u> probability density.

In eq. (2) ω is now interpreted as the <u>posterior</u> expectation of $\underline{Y}_t$ $(t \to \infty)$, i.e., the expectation of $\underline{Y}$ conditional on λ, and the <u>prior</u> expectation of $\underline{Y}$ becomes, say, ω_p. The prior mean ω_p is estimated by sampling $\underline{\hat{\lambda}}$ from a distribution like eq. (6) with estimated parameters $\hat{\lambda}$ and $\hat{\sigma}_1$; see Table I. In Table I the standard error of $\underline{\hat{\omega}}$ is denoted by $\underline{s}_{\hat{\omega}}$; for the computation of $\underline{s}_{\hat{\omega}}$ in autocorrelated situations see Fishman (1978) or Kleijnen (1975, pp. 453-468). Obviously, the prior mean is estimated by

$$\underline{\bar{\hat{\omega}}} = \sum_{j=1}^{m} \underline{\hat{\omega}}_j/m \tag{7}$$

and the standard error of $\underline{\bar{\hat{\omega}}}$ is the square root of

$$\underline{s}^2(\hat{\bar{\underline{\omega}}}) = \sum_{j=1}^{m} (\hat{\underline{\omega}}_j - \hat{\bar{\underline{\omega}}})^2/\{(m-1)m\} \tag{8}$$

TABLE I

Sampling $\hat{\underline{\lambda}}$

Sample	$\hat{\lambda}$	$\hat{\omega}$	$(s_{\hat{\omega}} \equiv s)$
1	$\hat{\lambda}_1$	$\hat{\omega}_1$	(s_1)
2	$\hat{\lambda}_2$	$\hat{\omega}_2$	(s_2)
.			
.			
.			
m	$\hat{\lambda}_m$	$\hat{\omega}_m$	(s_m)

Note that instead of the random sampling scheme of Table I other sampling schemes could be used, e.g., variance reduction can be realized by exploiting the correlation between $\hat{\underline{\lambda}}$ and $\hat{\underline{\omega}}$ using control variates; see Hopmans and Kleijnen (1980).

Other examples

Before proceeding with the discussion of Table I I would like to point out that this discussion also applies to other types of models. In the Appendix I present a simplistic econometric model (where the queuing system's parameters λ and μ are replaced by the regression parameters $\underline{\beta}$). Iman and Conover (1980) present dynamic groundwater-flow models for radioactive waste depositories that can be represented concisely as

$$y = h(\beta_1, \ldots, \beta_k) \tag{9}$$

where output y is a graph y(t) of output as a function of time, and the β's are constants over time; from (simulation) run to run these "constants" vary as specified by their density functions. Rice and Borison (1981) applied similar techniques for the cost estimation of a new plant. And so on.

What to report to the user?

Now I get to the crucial issue: what kind of statements is relevant?

(i) Statisticians like Iman and Conover report the estimated prior mean (eq. 7).

(ii) Management scientists report the estimated posterior mean since their model assumes that in a specific situation a constant λ will hold. More specifically, they obtain the estimate $\hat{\lambda}$ and simulate the corresponding $\hat{\omega}$ through eq. (3); this $\hat{\omega}$ corresponds to a single row in Table I.

(iii) The prudent scientist further performs a sensitivity analysis since he realizes that his assumed value $\hat{\lambda}$ may be completely wrong. So he simulates not for a single value $\hat{\lambda}$ but for a few additional values, the latter values being determined by his experimental design; such designs are discussed in Kleijnen (1975). The experimental design enables him to estimate the response surface, e.g.,

$$\hat{\omega} = \hat{\beta}_0 + \hat{\beta}_1 \lambda + \hat{\beta}_2 \lambda^2 \qquad (10)$$

After having validated this response surface model -see Kleijnen (1981)- the analyst knows whether the response is sensitive to the exact value of λ. If much sensitivity exists then he may try to obtain additional information on λ (increase n in eq. 6). However, getting more accurate information of λ may be impossible. (In econometrics information is restricted to a given time series.) If more accurate estimation is impossible, he may look for queuing systems (including priority rules) that are not very sensitive to the exact value of λ. If such a robust solution is impossible, the scientist has to give "qualified" recommendations; e.g. "if $\lambda = \lambda_0$ then $\hat{\omega} = \hat{\omega}_0$; if λ changes by one unit then $\hat{\omega}$ changes by $\hat{\beta}_1 + 2\hat{\beta}_2$ units".

In risk analysis the probability of occurrence of certain outcomes is quantified, using all rows of Table I. Note that the statistical literature of experimental design specifies which combinations of factors should be simulated. However, the actual levels of the factors are to be specified by the user. Here the variation of $\hat{\underline{\lambda}}$ -see eq. (5)- can be used: consider as extreme values, say, $\hat{\lambda} \pm 1.96\, \hat{\sigma}_1/\sqrt{n}$.

Conclusion

If a dynamic model assumes parameters constant over time, then the posterior mean (i.e. the mean conditional on specific values of these parameters) is relevant. Since parameters are unknown, they must be estimated. Sensitivity analysis quantifies the effects of incorrectly specified values of the parameters. If these effects are important then additional information on the parameters might be collected; otherwise robust solutions are to be sought. If these options do not work then risk analysis can quantify the probability of specific outputs, incorporating the probability distribution of the estimated parameters. Technically speaking, sensitivity analysis changes the values of parameters systematically, whereas risk analysis samples the parameter values. Both techniques are also discussed in Kleijnen (1980, pp. 73-79).

Appendix: A Simple Econometric Regression Model

(i) Suppose the true (unknown) model is

$$\underline{y}_t = \beta_0 + \beta_1 x_t + \underline{e}_t$$

with Normally Independently Distributed (NID) noise:

$$\underline{e} \sim NID\ (0,\ \sigma_e^2)$$

(ii) The true model is estimated (without specification error) from a time series (y_t, x_t) with $t \quad T_1 = \{1,\ldots,n\}$:

$$y_t = \hat{\beta}_0 + \hat{\beta}_1 x_t + \underline{u}_t \qquad t \quad T_1$$

with estimated noise

$$\underline{u}_t \sim NID\ (0,\ \sigma_u^2)$$

where

$$\hat{\sigma}_u^2 = \sum_1^n (y_t - \hat{y}_t)^2/(n-2)$$

with

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_t$$

(iii) A future time path is simulated over T_2, say $T_2 = \{n+1,\ldots,n+5\}$:

$$y_t = \hat{\beta}_0 + \hat{\beta}_1 x_t + \hat{\underline{u}}_t \qquad t \quad T_2$$

with $\hat{\underline{u}}_t \sim NID\ (0,\ \hat{\sigma}_u^2)$

and x_t (t T_2) being known (exogenous) input.

(iv) If not the posterior expectation of, say, y_{n+5} is desired then (iii) is replaced by the following procedure:

(a) Sample $\hat{\beta}_0$ and $\hat{\beta}_1$ from

$$\beta \sim N(\hat{\underset{\sim}{\beta}}, \hat{\underset{\sim}{\Omega}}_\beta)$$

where the parameters of this multivariate normal distribution follow from the regression analysis of the historical time series (see ii):

$$\hat{\underset{\sim}{\beta}} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1} \underset{\sim}{X}'\underset{\sim}{Y}$$

$$\hat{\underset{\sim}{\Omega}}_\beta = (\underset{\sim}{X}'\underset{\sim}{X})^{-1} \hat{\sigma}_u^2$$

(b) Given the sample (β_0, β_1), simulate over T_2:

$$y_t = \hat{\beta}_0 + \hat{\beta}_1 x_t + \hat{u}_t \qquad t \quad T_2$$

where

$$\hat{u}_t \sim \text{NID}\ (0, \hat{\sigma}_u^2(\hat{\underset{\sim}{\beta}}))$$

$$\text{with } \hat{\sigma}_u^2(\hat{\underset{\sim}{\beta}}) = \sum_1^n (y_t - \hat{y}_t(\hat{\underset{\sim}{\beta}}))^2/(n-2)$$

$$\text{and } \hat{y}_t(\hat{\underset{\sim}{\beta}}) = \hat{\beta}_0 + \hat{\beta}_1 x_t$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ follow from (a), and not from step (ii).

(c) Repeat (b) n_1 times ($n_1 \geqq 1$).

(d) Repeat (a) n_2 times ($n_2 >> 1$).

Note that common random numbers should not be used when sampling $\hat{u}$ since they inflate the estimated prior mean. Antithetic variates can be recommended in step (c) above. See Kleijnen (1975) for details.

References

Fishman, G.S., Principles of Discrete Event Simulation. John Wiley & Sons, Inc., New York, 1978.

Hopmans, A.C.M. and J.P.C. Kleijnen, Regression estimation in simulation. Journal Operational Research Society, 31, no. 11, 1980, pp. 1033-1038.

Iman, R.L. and W.J. Conover, Small sample sensitivity analysis for computer models, with an application to risk assessment. Communications in Statistics, A9, no. 17, 1980, pp. 1749-1842.

Kleijnen, J.P.C., Statistical techniques in simulation (in two volumes). Marcel Dekker, Inc., New York, 1974/1975. (Russian translation: Publishing House "Statistics", Moscow, 1978.)

Kleijnen, J.P.C., Computers and Profits; Quantifying Financial Benefits of Information. Addison-Wesley Publishing Company, Reading (Massachusetts), 1980.

Kleijnen, J.P.C., Regression analysis for simulation practioners. Journal Operational Research Society, 32, 1981, pp. 35-43.

Rice, T.R. and A.B. Borison, Probabilistic cost estimating for the Great Plains coal gasification plant. Interfaces, 11, no. 2, April 1981, pp. 62-68.