Experimental design with random regressors at the design stage

Luc Pronzato
Laboratoire I3S,
CNRS—Université de Nice–Sophia Antipolis,
Bât. Euclide, "Les Algorithmes",
2000 route des Lucioles, 06410 Biot, France

Abstract: We consider experimental design for estimating the parameters of a regression model when the actual experimental conditions that are used at the experimentation stage may randomly fluctuate around those that are specified at the design stage. However, the values used in the experimentation are observed, and they can be used for estimating the parameters of the model. A direct extension of D-optimal design is to consider the expected value of the determinant of the information matrix with respect to the possible fluctuations of the experimental variables. We show how the analytical expression of such expected determinants can be obtained when the experimental variables are independently distributed. Three examples of experimental design in linear and nonlinear regression are presented.

AMS Subject Classification: Primary 62K05; secondary 15A52, 15A15.

Key words: Experimental design; expected determinant; random matrix; random experiments.

1 Introduction and problem statement

We consider experimental design for parameter estimation in a regression model. There are several circumstances under which the experimental conditions, and thus the regressors, may be considered as random before the experimentation takes place, and expected values of determinants of information matrices are involved.

A first example of such a situation is considered in Pronzato (1999), where experimental conditions ξ_k are independently identically distributed (i.i.d.) random variables, and thus uncontrolled, but are observed sequentially. The decision of observing y_k under experimental conditions ξ_k must be taken on–line. When the criterion for experimental design is D-optimality, the expression of the expected value of the determinant of the information matrix can be used to compute an open–loop–feedback strategy (Bertsekas, 1987).

Another situation is considered in this paper: we assume that the experiment is designed off-line (non-sequentially), but when the experimentation takes place, the actual experiment which is used may differ from the one that was designed. In some cases, the difference may be moderate: for instance, the experimenter may be able to tune the process according to the specifications of the design with a given accuracy. Designing experiments as if no fluctuation would take place is then acceptable. In some other cases, however, the difference may be large. For instance when the time at which observations are made is one of the design variables, and human operation is required, the experimenter may be available only at a random time around that specified by the design. Pharmacokinetic experiments, with observations corresponding to blood samples taken on individuals, provide examples of such a situation, see Section 3. Note that the problem differs from that in Keeler and Reilly (1992): we assume that the actual values of the experimental variables are observed, and they can thus be used for estimating the parameters of the model. However, at the design stage, we consider the experimental variables as random.

We thus consider a regression model with observations y_k, $k = 1, \ldots, N$ given by

$$y_k = \eta(\bar{\boldsymbol{\theta}}, \xi_k) + \epsilon_k$$

with $\eta(\boldsymbol{\theta}, \xi)$ the model response for parameters $\boldsymbol{\theta}$ and experimental conditions ξ, $\bar{\boldsymbol{\theta}}$ the unknown true value of the model parameters and (ϵ_k) a sequence of i.i.d. measurement errors with zero mean and variance σ^2. The parameters are estimated by least–squares. When the experimental conditions can be settled without error, a classical criterion for choosing $\Xi_1^N = (\xi_1, \ldots, \xi_N)$ that yields a precise estimation corresponds to D-optimality, see, e.g., Box and Lucas (1959), Atkinson and Hunter (1968), Fedorov (1972): Ξ_1^N maximizes $\det[\mathbf{M}_N]$, with $\mathbf{M}_N$ the information matrix

$$\mathbf{M}_N = \mathbf{M}(\boldsymbol{\theta}, \Xi_1^N) = \sum_{k=1}^{N} \mathbf{z}_k \mathbf{z}_k^\top ,$$

where

$$\mathbf{z}_k = \mathbf{z}(\boldsymbol{\theta}, \xi_k) = \frac{\partial \eta(\boldsymbol{\theta}, \xi_k)}{\partial \boldsymbol{\theta}} . \tag{1}$$

When $\eta(\boldsymbol{\theta}, \xi)$ is linear in $\boldsymbol{\theta}$, $\eta(\boldsymbol{\theta}, \xi) = \mathbf{z}^\top(\xi)\boldsymbol{\theta}$ and $\mathbf{M}_N = \mathbf{M}(\Xi_1^N)$ does not depends on $\boldsymbol{\theta}$. When $\eta(\boldsymbol{\theta}, \xi)$ is nonlinear in $\boldsymbol{\theta}$, $\mathbf{M}_N = \mathbf{M}(\boldsymbol{\theta}, \Xi_1^N)$ depends on $\boldsymbol{\theta}$. Local D-optimal design then corresponds to maximizing $\det[\mathbf{M}(\boldsymbol{\theta}, \Xi_1^N)]$ for a given nominal value of $\boldsymbol{\theta}$. Average D-optimal design corresponds to maximizing the expected value of $\det[\mathbf{M}(\boldsymbol{\theta}, \Xi_1^N)]$, or $\log\det[\mathbf{M}(\boldsymbol{\theta}, \Xi_1^N)]$, see Fedorov (1980), with respect to $\boldsymbol{\theta}$, for a given prior distribution for $\boldsymbol{\theta}$. Only *local design* will be considered in this paper. We assume that at the design stage the ξ_k's are random variables, with the distribution of ξ_k depending on a design variable ζ_k, and we consider the design criterion $E_{\mathbf{z}_1^N}\{\det[\mathbf{M}_N]\}$, which is a function of $\zeta_1, \ldots, \zeta_N$. For instance, ξ_k may be normally distributed $\mathcal{N}(\zeta_k, \rho^2(\zeta_k))$, see Example 1 in Section 3, or uniformly distributed in $[\zeta_k(1-u), \zeta_k(1+u)]$, see Example 2.

Stochastic approximation techniques might be used for maximizing $E_{\mathbf{z}_1^N}\{\det[\mathbf{M}_N]\}$ with respect to $\zeta_1, \ldots, \zeta_N$, as done in Pronzato and Walter (1985) for average optimal design. However, the assumption of independence of the ξ_k's allows one to derive the analytical expression of $E_{\mathbf{z}_1^N}\{\det[\mathbf{M}_N]\}$, so that more efficient *deterministic* design algorithms can be used, for instance of the exchange type, see Fedorov (1972), Mitchell (1974). Expressions for expected determinants are derived in Section 2. Section 2.1 concerns the case where the vectors $\mathbf{z}_k$ in $\mathbf{M}_N$ are i.i.d. First, we consider the criterion $E_{\mathbf{z}_1^N}\{\det[\mathbf{Q}+\mathbf{M}_N]\}$, with $\mathbf{Q}$ a given regular matrix, and give a detailed proof of the result. The expression of $E_{\mathbf{z}_1^N}\{\det[\mathbf{M}_N]\}$ is obtained as a corollary and extends a result stated in Pronzato (1998). The case where the $\mathbf{z}_k$'s are independent but have different distributions is considered in Section 2.2. Again, the result is given first for $E_{\mathbf{z}_1^N}\{\det[\mathbf{Q}+\mathbf{M}_N]\}$, with $\mathbf{Q}$ a given regular matrix. The proof is similar to that in Section 2.1, but is more cumbersome and is omitted. Note that this criterion may be of interest in the case of Bayesian estimation, with $\mathbf{Q}+\mathbf{M}_N$ the Bayesian information matrix. The expression of $E_{\mathbf{z}_1^N}\{\det[\mathbf{M}_N]\}$ is given in a Corollary. Finally, three examples are presented in Section 3.

2 The expected value of a determinant

2.1 I.i.d. regressors

The following theorem is stated without proof in Pronzato (1999).

Theorem 1 *Assume that the vectors $\mathbf{z}_l \in \mathcal{Z} \subset \mathbb{R}^p$ are i.i.d., with a probability measure μ such that $\bar{\mathbf{M}} = \int \mathbf{z}\mathbf{z}^\top \mu(d\mathbf{z})$ exists, then for any regular matrix $\mathbf{Q}$,*

$$E_{\mathbf{z}_1^k}\{\det[\mathbf{Q}+\sum_{l=1}^{k}\mathbf{z}_l\mathbf{z}_l^\top]\} = \det[\mathbf{Q}] \times \sum_{l=0}^{\min(k,p)} \mathsf{C}_k^l P_l(\mathbf{t}_1^l)\,, \tag{2}$$

where $\mathbf{z}_1^k = (\mathbf{z}_1,\ldots,\mathbf{z}_k)$, $\mathsf{C}_k^l = k!/[l!(k-l)!]$, $0! = 1$, $\mathbf{t}_1^l = (t_1,\ldots,t_l)$ with $t_l = tr[(\mathbf{Q}^{-1}\bar{\mathbf{M}})^l]$, $l = 1,\ldots,k$, $P_0 = 1$ and

$$P_l(\mathbf{t}_1^l) = l! \sum_{1^{n_1}\ldots l^{n_l}} \frac{(-1)^{l+\sum_{i=1}^l n_i}}{(1^{n_1}\cdots l^{n_l})(n_1!\cdots n_l!)} t_1^{n_1}\cdots t_l^{n_l}\,,\ l \geq 1\,,$$

where the summation $\sum_{1^{n_1}\ldots l^{n_l}}$ is over all non–negative integers $n_1,\ldots,n_l$ such that $\sum_{i=1}^l i n_i = l$, that is, over all partitions of l which have n_i parts of size i.

Proof. The proof is by induction on k.

For $k=1$, direct calculation gives $E_{\mathbf{z}}\{\det[\mathbf{Q}+\mathbf{z}\mathbf{z}^\top]\} = \det[\mathbf{Q}](1+t_1)$.

Assume that (2) is true, and compute

$$\mathcal{E}_{k+1} = E_{\mathbf{z}_1^{k+1}}\{\det[\mathbf{Q}+\sum_{l=1}^{k+1}\mathbf{z}_l\mathbf{z}_l^\top]\} = E_{\mathbf{z}_1^{k+1}}\{\det[(\mathbf{Q}+\mathbf{z}_{k+1}\mathbf{z}_{k+1}^\top)+\sum_{l=1}^{k}\mathbf{z}_l\mathbf{z}_l^\top]\}\,.$$

Using the induction hypothesis and the assumption of independence of the $\mathbf{z}_l$'s, one gets

$$\mathcal{E}_{k+1} = E_{\mathbf{z}}\{\det[\mathbf{Q}'(\mathbf{z})] \times \sum_{l=0}^{\min(k,p)} \mathsf{C}_k^l P_l(\boldsymbol{\tau}_1^l)\}\,,$$

with $\boldsymbol{\tau}_1^k = (\tau_1, \ldots, \tau_k)$, $\tau_l = \mathrm{tr}[(\mathbf{Q}'^{-1}(\mathbf{z})\bar{\mathbf{M}})^l]$ and $\mathbf{Q}'(\mathbf{z}) = \mathbf{Q} + \mathbf{z}\mathbf{z}^\top$. Direct calculation gives

$$\begin{aligned} \det[\mathbf{Q}'(\mathbf{z})] &= q(\mathbf{z}) \det[\mathbf{Q}]\,, \\ \tau_l &= \mathrm{tr}\{[\mathbf{A}(\mathbf{z})]^l\}\,, \end{aligned}$$

where

$$\begin{aligned} q(\mathbf{z}) &= 1 + \mathbf{z}^\top \mathbf{Q}^{-1}\mathbf{z}\,, \\ \mathbf{A}(\mathbf{z}) &= \left(\mathbf{Q}^{-1} - \frac{\mathbf{Q}^{-1}\mathbf{z}\mathbf{z}^\top\mathbf{Q}^{-1}}{q(\mathbf{z})}\right)\bar{\mathbf{M}}\,. \end{aligned}$$

One has $P_l(\boldsymbol{\tau}_1^l) = l! \sum_{i_1<i_2<\cdots<i_l} \lambda'_{i_1}\lambda'_{i_2}\cdots\lambda'_{i_l}$ for $l \leq p$ and $P_l(\boldsymbol{\tau}_1^l) = 0$ for $l > p$, where the λ'_i's denote the eigenvalues of $\mathbf{A}(\mathbf{z})$. Define $\Delta(\lambda) = \det[\lambda\mathbf{I} - \mathbf{A}(\mathbf{z})]$. One can easily check that $\Delta(\lambda)$ is affine in $1/q(\mathbf{z})$,

$$\Delta(\lambda) = \prod_{l=1}^{p}(\lambda - \lambda'_l) = \sum_{l=0}^{p} \frac{1}{l!} P_l(\boldsymbol{\tau}_1^l)(-1)^l \lambda^{p-l}\,, \tag{3}$$

and the terms $P_l(\boldsymbol{\tau}_1^l)/l!$ coincide with the coefficients of the characteristic polynomial of $\mathbf{A}(\mathbf{z})$ produced by Faddeev'method, see Gantmacher (1966)[9], vol. 1, p. 88. From (3), the terms $P_l(\boldsymbol{\tau}_1^l)$ are affine in $1/q(\mathbf{z})$, and, in the evaluation of τ_l and $P_l(\boldsymbol{\tau}_1^l)$, only the terms linear at most in $1/q(\mathbf{z})$ have to be considered. Therefore, one can replace τ_l by

$$\tau'_l = t_l - l\frac{\mathbf{z}^\top(\mathbf{Q}^{-1}\bar{\mathbf{M}})^l\mathbf{Q}^{-1}\mathbf{z}}{q(\mathbf{z})}\,,$$

which gives

$$\mathcal{E}_{k+1} = \det[\mathbf{Q}] \sum_{l=0}^{\min(k,p)} \mathsf{C}_k^l E_{\mathbf{z}}\{q(\mathbf{z}) P_l(\tau'_1, \ldots, \tau'_l)\}\,.$$

Consider a product term of the form $T(\mathbf{z}) = (\tau'_1)^{n_1} \times \cdots \times (\tau'_k)^{n_m}$ in $P_l(\tau'_1, \ldots, \tau'_l)$. Since only the terms linear at most in $1/q(\mathbf{z})$ need to be kept, $T(\mathbf{z})$ can be replaced by

$$T'(\mathbf{z}) = \prod_{i=1}^{m}(t_i)^{n_i} - \sum_{i=1}^{m}\left(i\frac{\mathbf{z}^\top(\mathbf{Q}^{-1}\bar{\mathbf{M}})^i\mathbf{Q}^{-1}\mathbf{z}}{q(\mathbf{z})} n_i (t_i)^{n_i-1}\prod_{j\neq i}(t_j)^{n_j}\right)\,,$$

and

$$E_{\mathbf{z}}\{q(\mathbf{z})T'(\mathbf{z})\} = (1+t_1)\prod_{i=1}^{m}(t_i)^{n_i} - \sum_{i=1}^{m}\left(i n_i (t_i)^{n_i-1} t_{i+1}\prod_{j\neq i}(t_j)^{n_j}\right)\,.$$

Lengthy but straightforward calculation then gives

$$E_{\mathbf{z}}\{q(\mathbf{z})P_l(\tau_1',\ldots,\tau_l')\} = P_l(\mathbf{t}_1^l) + P_{l+1}(\mathbf{t}_1^{l+1})\,,$$

if $l+1 \leq p$ and $E_{\mathbf{z}}\{q(\mathbf{z})P_l(\tau_1',\ldots,\tau_l')\} = P_l(\mathbf{t}_1^l)$ otherwise, so that

$$\mathcal{E}_{k+1} = \det[\mathbf{Q}] \times \sum_{l=0}^{\min(k+1,p)} \mathsf{C}_{k+1}^l P_l(\mathbf{t}_1^l)\,,$$

which completes the proof.

■

For $l = 1,\ldots,4$, the polynomials P_l are as follows:

$$\begin{aligned} P_0 &= 1\,, \\ P_1(t_1) &= t_1\,, \\ P_2(\mathbf{t}_1^2) &= t_1^2 - t_2\,, \\ P_3(\mathbf{t}_1^3) &= t_1^3 - 3t_1t_2 + 2t_3\,, \\ P_4(\mathbf{t}_1^4) &= t_1^4 - 6t_1^2t_2 + 8t_1t_3 + 3t_2^2 - 6t_4\,. \end{aligned}$$

Up to the sign factor $(-1)^{l+\sum_{i=1}^l n_i}$, the polynomials $P_l(\mathbf{t}_l)/l!$ also coincide with the cycle index of symmetric groups, see, e.g., van Lint and Wilson (1992), p. 465.

From the theorem above, we easily get the property given in Pronzato (1998).

Corollary 1 *Assume that the vectors $\mathbf{z}_i \in \mathcal{Z} \subset I\!\!R^p$ are i.i.d., with a probability measure μ such that $\bar{\mathbf{M}} = \int \mathbf{z}\mathbf{z}^\top \mu(d\mathbf{z})$ exists, then, for $k \geq p$,*

$$E_{\mathbf{z}_1^k}\{\det[\frac{1}{k}\sum_{l=1}^k \mathbf{z}_l\mathbf{z}_l^\top]\} = \frac{k!}{k^p(k-p)!}\det[\bar{\mathbf{M}}]\,.$$

Proof. Take $\mathbf{Q} = \alpha\mathbf{I}$ in Theorem 1. This gives $\det[\mathbf{Q}] = \alpha^p$, $t_l = \alpha^{-l}\tau_l$ with $\tau_l = \mathrm{tr}[\bar{\mathbf{M}}^l]$, and $P_l(\mathbf{t}_1^l) = \alpha^{-l}P_l(\boldsymbol{\tau}_1^l)$, so that

$$E_{\mathbf{z}_1^k}\{\det[\alpha\mathbf{I} + \sum_{l=1}^k \mathbf{z}_l\mathbf{z}_l^\top]\} = \sum_{l=0}^{\min(k,p)} \mathsf{C}_k^l \alpha^{p-l} P_l(\boldsymbol{\tau}_1^l)\,.$$

Noticing that $P_p(\boldsymbol{\tau}_1^p) = p!\det[\bar{\mathbf{M}}]$ and taking $\alpha = 0$, one gets the result.

■

When the $\mathbf{z}_i$'s are independent Gaussian vectors with zero mean, Corollary 1 follows from Th. 3.2.15 p. 100 in Muirhead (1982). The case of Gaussian vectors with non-zero mean corresponds to Th. 10.3.7 p. 447 in the same volume. Other particular cases, such as the components of the $\mathbf{z}_i$'s being i.i.d., are considered in Anderson (1989); see also Fortet (1951). Note that the asymptotic version ($k \to \infty$) of Corollary 1 is obvious, since $\lim_{k\to\infty}(1/k)\sum_{l=1}^k \mathbf{z}_l\mathbf{z}_l^\top = \bar{\mathbf{M}}$. Also note that $E\{\det[(1/k)\sum_{l=1}^k \mathbf{z}_l\mathbf{z}_l^\top]\} < \det[\bar{\mathbf{M}}]$ for $p > 1$, so that $\det[(1/k)\sum_{l=1}^k \mathbf{z}_l\mathbf{z}_l^\top]$ underestimates $\det[\bar{\mathbf{M}}]$.

2.2 Independent regressors

An extension of Theorem 1 to the case where the $\mathbf{z}_k$'s are independent but have different probability measures is as follows.

Theorem 2 *Assume that the vectors $\mathbf{z}_l \in \mathcal{Z} \subset I\!\!R^p$, $l = 1, \ldots, k$, are independently distributed, with probability measures μ_l such that $\bar{\mathbf{M}}_l = \int \mathbf{z}_l \mathbf{z}_l^\top \mu_l(d\mathbf{z}_l)$ exists. Then, for any regular matrix $\mathbf{Q}$,*

$$E_{\mathbf{z}_1^k}\{\det[\mathbf{Q} + \sum_{l=1}^{k} \mathbf{z}_l \mathbf{z}_l^\top]\} = \det[\mathbf{Q}] \times \sum_{l=0}^{\min(k,p)} \sum_{\mathcal{S}_l} Q_l(\mathcal{S}_l) \tag{4}$$

with $\mathbf{z}_1^k = (\mathbf{z}_1, \ldots, \mathbf{z}_k)$, $Q_0 = 1$ and

$$Q_l(\mathcal{S}_l) = \sum_{1^{n_1} \cdots l^{n_l}} (-1)^{l + \sum_{i=1}^{l} n_i} \sum_{\sigma_{1^{n_1} \cdots l^{n_l}}(\mathcal{S}_l)} \tau(1^{n_1}) \cdots \tau(l^{n_l}) \,, \; l \geq 1 \,, \tag{5}$$

where:
$\mathcal{S}_l$ is one of the C_k^l subsets of $\{1, \ldots, k\}$ with l elements,
$\sigma_{1^{n_1} \cdots l^{n_l}}(\mathcal{S}_l)$ denotes one of the $\frac{l!}{(1^{n_1} \cdots l^{n_l})(n_1! \cdots n_l!)}$ partitions of a given type $(1^{n_1} \cdots l^{n_l})$ of $\mathcal{S}_l$, with $(i-1)!$ cyclic permutations of each block of size i,
$\tau(i^{n_i}) = t(j_{1_1}, \ldots, j_{1_i}) \times \cdots \times t(j_{n_{i1}}, \ldots, j_{n_{ii}})$, the j_i's being elements of the blocks of size i, and

$$t(i_1, \ldots, i_n) = tr[\mathbf{Q}^{-1} \bar{\mathbf{M}}_{i_1} \times \cdots \times \mathbf{Q}^{-1} \bar{\mathbf{M}}_{i_n}] \,.$$

The summation $\sum_{1^{n_1} \cdots l^{n_l}}$ in (5) is over all non–negative integers $n_1, \ldots, n_l$ such that $\sum_{i=1}^{l} i n_i = l$, that is, over all partitions of l which have n_i parts of size i.

The proof is similar to that of Theorem 1 and is omitted. The expressions of Q_l for $l = 1, \ldots, 4$ are as follows:

$$\begin{aligned}
Q_0 &= 1 \,, \\
Q_1(\{1\}) &= t(1) \,, \\
Q_2(\{1,2\}) &= t(1)t(2) - t(1,2) \,, && (6) \\
Q_3(\{1,2,3\}) &= t(1)t(2)t(3) + t(1,2,3) + t(2,1,3) \\
&\quad - t(1)t(2,3) - t(2)t(1,3) - t(3)t(1,2) \,, && (7) \\
Q_4(\{1,2,3,4\}) &= t(1)t(2)t(3)t(4) \\
&\quad - t(1)t(2)t(3,4) - t(1)t(3)t(2,4) - t(1)t(4)t(2,3) \\
&\quad - t(2)t(3)t(1,4) - t(2)t(4)t(1,3) - t(3)t(4)t(1,2) \\
&\quad + t(1)t(2,3,4) + t(1)t(3,2,4) + t(2)t(1,3,4) + t(2)t(3,1,4) \\
&\quad + t(3)t(1,2,4) + t(3)t(2,1,4) + t(4)t(1,2,3) + t(4)t(2,1,3) \\
&\quad + t(1,2)t(3,4) + t(1,3)t(2,4) + t(1,4)t(2,3) \\
&\quad - t(1,2,3,4) - t(2,1,3,4) - t(1,3,2,4) \\
&\quad - t(3,1,2,4) - t(1,4,2,3) - t(1,4,3,2) \,.
\end{aligned}$$

Using the same technique as in the proof of Corollary 1, we get the following corollary of Theorem 2.

Corollary 2 *Assume that the vectors* $\mathbf{z}_l \in \mathscr{Z} \subset I\!R^p$, $l = 1, \ldots, k$, *are independently distributed, with probability measures* μ_l *such that* $\bar{\mathbf{M}}_l = \int \mathbf{z}_l \mathbf{z}_l^\top \mu_l(d\mathbf{z}_l)$ *exists, then for* $k \geq p$

$$E_{\mathbf{z}_1^k}\{\det[\sum_{l=1}^{k} \mathbf{z}_l \mathbf{z}_l^\top]\} = \sum_{\mathcal{S}_p} Q_p(\mathcal{S}_p)\,, \tag{8}$$

where $\mathcal{S}_p$ *and* $Q_p(\mathcal{S}_p)$ *are defined as in Theorem 2, with*

$$t(i_1, \ldots, i_n) = tr[\bar{\mathbf{M}}_{i_1} \times \cdots \times \bar{\mathbf{M}}_{i_n}]\,. \tag{9}$$

The closed–form expression (8) shows that the usual property of multiplication of determinants for replicated experiments does not hold for expected determinants. Consider an experiment Ξ defined by $\mathbf{z}_1^k = (\mathbf{z}_1, \ldots, \mathbf{z}_k)$, and an experiment Ξ' consisting of q replications of Ξ in the following sense: Ξ' is defined by $\mathbf{z}'_{li}$, $l = 1, \ldots, k$, $i = 1, \ldots, q$, which satisfy $E\{\mathbf{z}'_{li}\} = E\{\mathbf{z}_l\}$. Then, in general,

$$E_{\Xi'}\{\det[\sum_{i=1}^{q}\sum_{l=1}^{k} \mathbf{z}'_{li}\mathbf{z}'^{\top}_{li}]\} > q^p E_{\Xi}\{\det[\sum_{l=1}^{k} \mathbf{z}_l \mathbf{z}_l^\top]\}\,, \tag{10}$$

as illustrated in Example 1 below. The reason is that replications yield sets of indices $\mathcal{S}_p$ in (8) that correspond to vectors $\mathbf{z}'_{li}$, $\mathbf{z}'_{lj}$ with the same mean $E\{\mathbf{z}_l\}$. The contribution of such sets to the expected determinant is generally positive, due to random fluctuations around $E\{\mathbf{z}_l\}$. Having an inequality in (10) makes the extension to approximate design theory difficult, since design measures on the product space $\mathscr{Z}^{\otimes p}$ must be considered. Also, we shall see in Example 2 that duplicating one of the p observations in an optimal experiment with p points may be optimal or not depending which one is duplicated, whereas all possible replications give the same result for D-optimality (Atkinson and Hunter, 1968).

3 Examples

Example 1. Consider a linear regression model, with quadratic effect:

$$y_i = \mathbf{z}_i^\top \bar{\boldsymbol{\theta}} + \epsilon_i\,,$$

where $\mathbf{z}_i = [1,\ \zeta_i + \omega_i,\ (\zeta_i + \omega_i)^2]^\top$ is the regressor used in the experiment, (ϵ_i) is an i.i.d. sequence of measurement errors, with zero mean and variance σ^2, and (ω_i) is an independent sequence of disturbances which corrupts the design variables ζ_i. The sequences (ϵ_i)

and (ω_i) are independent, and ω_i is normally distributed $\mathcal{N}(0, \rho^2(\zeta_i))$, with $\rho^2(\zeta) = 0.03\zeta^2$. Corollary 2, with (7,9) and

$$\bar{\mathbf{M}}_i = \begin{pmatrix} 1 & \zeta_i & \zeta_i^2 + \rho^2(\zeta_i) \\ \zeta_i & \zeta_i^2 + \rho^2(\zeta_i) & \zeta_i^3 + 3\zeta_i\rho^2(\zeta_i) \\ \zeta_i^2 + \rho^2(\zeta_i) & \zeta_i^3 + 3\zeta_i\rho^2(\zeta_i) & \zeta_i^4 + 6\zeta_i^2\rho^2(\zeta_i) + 3\rho^4(\zeta_i) \end{pmatrix},$$

gives the expression of $E_{\omega_1^N}\{\det[\mathbf{M}_N]\}$.

The D-optimal experiment for $N = 9$ observations in $[0, 1]$ corresponds to three observations at each of the design points $0, 1/2$ and 1. Figure 1 presents $\det[E\{\mathbf{M}_9\}] = \det\left[\sum_{i=1}^9 \bar{\mathbf{M}}_i\right]$ (full line) and $E\{\det[\mathbf{M}_9]\}$ (dashed line), as functions of ζ, when the three design points are $0, 1$ and ζ. The maximum of $\det[E\{\mathbf{M}_9\}]$ is obtained for $\zeta = 1$, whereas the maximum of $E\{\det[\mathbf{M}_9]\}$ is obtained at $\zeta \simeq 0.5405$. Figure 1 also gives $3^3 \times E\{\det[\mathbf{M}_3]\}$ when only three observations are made at $0, 1$ and ζ (dash–dotted line). The maximum of this curve is obtained at $\zeta \simeq 0.5442$. Note that $E\{\det[\mathbf{M}_9]\} > 3^3 \times E\{\det[\mathbf{M}_3]\}$. Also note that, as it could be expected, replications reduce the difference between $\det[E\{\mathbf{M}\}]$ and $E\{\det[\mathbf{M}]\}$: the difference between $3^3 \times \det[E\{\mathbf{M}_3\}] = \det[E\{\mathbf{M}_9\}]$ and $3^3 \times E\{\det[\mathbf{M}_3]\}$ is larger than that between $\det[E\{\mathbf{M}_9\}]$ and $E\{\det[\mathbf{M}_9]\}$.

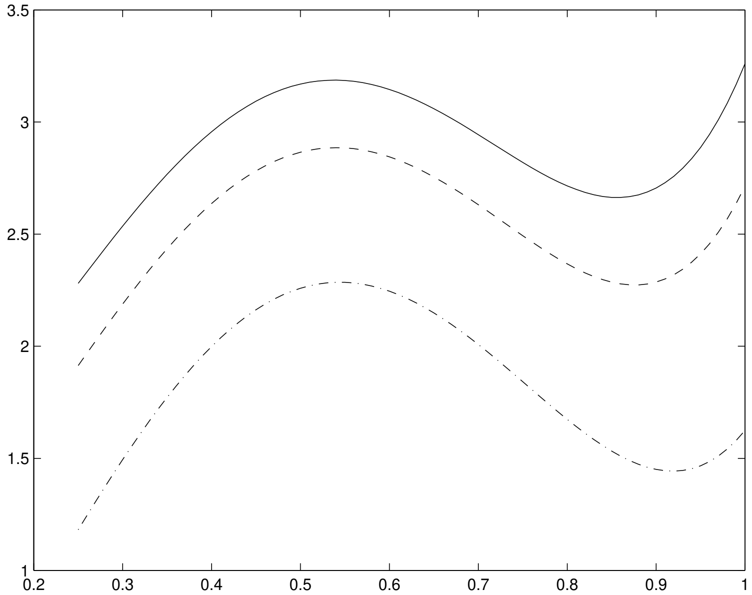

Figure 1: $\det[E\{\mathbf{M}_9\}] = 3^3 \times \det[E\{\mathbf{M}_3\}]$ (full line), $E\{\det[\mathbf{M}_9]\}$ (dashed line) and $3^3 \times E\{\det[\mathbf{M}_3]\}$ (dash–dotted line) as functions of ζ in Example 1. There are three (resp. one) observations at each of the design points $0, 1$ and ζ for $\mathbf{M}_9$ (resp. $\mathbf{M}_3$).

Example 2. Consider the nonlinear regression model

$$\eta(\boldsymbol{\theta}, \xi) = \frac{\theta_1}{\theta_1 - \theta_2}[\exp(-\theta_2\xi) - \exp(-\theta_1\xi)]\,, \tag{11}$$

which describes the evolution of the yield of intermediate product in a consecutive chemical reaction with respect to time ξ, see Box and Lucas (1959), Atkinson and Hunter (1968). The measurement errors ϵ_k are assumed to be i.i.d. with zero mean and variance σ^2. We assume that, at the design stage, the experimental conditions can be considered as independently distributed, ξ_k having the uniform distribution in the interval $[(1-u)\zeta_k, (1+u)\zeta_k]$, with u some given constant in $[0, 1]$. The design problem consists in choosing the ζ_k's, $k = 1, \ldots, N$, that maximize $E_{\Xi_1^N}\{\det[\bar{\mathbf{M}}(\boldsymbol{\theta}, \Xi_1^N)]\}$, with $\Xi_1^N = (\xi_1, \ldots, \xi_N)$. Local optimal design is considered, and we take $\boldsymbol{\theta} = (0.7, 0.2)^\top$, the same value as in Box and Lucas (1959), Atkinson and Hunter (1968). Using (11) and (1), the matrices $\bar{\mathbf{M}}_i$ in (6,9) can easily be calculated. The expression of $E_{\Xi_1^N}\{\det[\mathbf{M}(\boldsymbol{\theta}, \Xi_1^N)]\}$ is then given by (8), which can be maximized for instance by an exchange algorithm (Fedorov, 1972; Mitchell, 1974).

Figure 2 presents the evolution of the optimal design as a function of u when $N = 3$ (dashed lines): there are two observations at the design points on the dashed line on the top and one on the bottom. The evolution of the associated value of $E\{\det[\mathbf{M}_3]\}$ is given on the dashed line in Figure 3. Replicating the smallest value of ζ instead of the largest corresponds to a design which is only *locally* optimal. The corresponding curves are in full line in Figures 2 and 3: in Figure 2, there are two observations at the design points on the full line on the bottom and one on the top. When u tends to zero, the optimal experiment tends to the D-optimal experiment in Box and Lucas (1959), Atkinson and Hunter (1968): one observation at each of the points 1.23 and 6.86 and one replication of either of them. The optimal experiment is rather insensitive to u for values of u reasonably small (say $u < 0.4$), but the expected determinant rapidly decreases, see Figure 3.

Example 3. We consider the following compartmental model, taken from D'Argenio (1990), describing the plasma concentration of an orally administrated drug:

$$\eta(\boldsymbol{\theta}, \xi) = \frac{DK_a}{V(K_a - K_e)}[\exp(-K_e\xi) - \exp(-K_a\xi)]\,, \tag{12}$$

where D is the bolus dose of the drug, V represents the distribution volume of the sampled (plasma) compartment, K_a is the absorption rate from the input to the sampled compartment and K_e is the elimination rate from the sampled compartment. The design variable ξ corresponds to the sampling time for plasma samples. The dose D is known, and the parameters to be estimated are $\boldsymbol{\theta} = (V, K_a, K_e)^\top$. The experiment is designed for the nominal value $V = 15(l)$, $K_a = 2.0(h^{-1})$ and $K_e = 0.25(h^{-1})$. The (local) D-optimal experiment with $N = n \times p$ then consists of n replications at each of the design points 0.43, 1.80 and 6.16.

Assume that the experimenter who administrates the drug at time 0 is totally available until T, so that plasma samples can be collected at any specified time between 0 and T. Between $\xi = T$ and $\xi_{\max} = 10$, the time when the experiment is terminated, another experimenter collects the remaining plasma samples, but at random times uniformly distributed in $[T, 10]$.

Figure 4 presents $E\{\det[\mathbf{M}_3]\}$ as a function of T for the D-optimal design (dashed line)

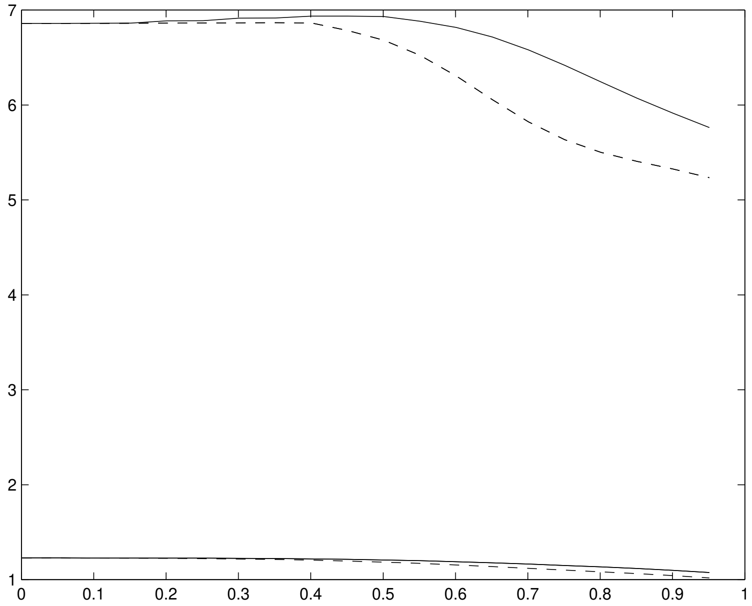

Figure 2: Optimal design for $E\{\det[\mathbf{M}_3]\}$ (dashed lines: two observations on the top) and locally–optimal design (full lines: two observations on the bottom) as functions of u in Example 2.

and for the design that maximizes $E\{\det[\mathbf{M}_3]\}$ for each value of T (full line). The first two sampling times of this optimal design are presented in Figure 5 (the third sampling time is random in $[T, 10]$).

The curve in dashed line in Figure 4 is discontinuous at $T = 1.8$, one of the sampling times of the D-optimal design, since for $T > 1.8$ the second sample is collected precisely at the D-optimal time. The initial decrease of the curve is due to the importance of taking the second sample early enough, and the probability of this event decreases when T increases. For the curve in full line, the second sampling time is always taken smaller than T, close to T for $1 \leq T \leq 1.8$, and then close to the value 1.8 of the second D-optimal sampling time, see Figure 5.

Acknowledgements: The author thanks M. Barbier from the University of Tourves for helpful discussions.

References

[1] T. Anderson. Evaluation of the expected value of a determinant. *Statistics & Probability Letters*, 8:25–26, 1989.

[2] A. Atkinson and W. Hunter. The design of experiments for parameter estimation. *Technometrics*, 10(2):271–289, 1968.

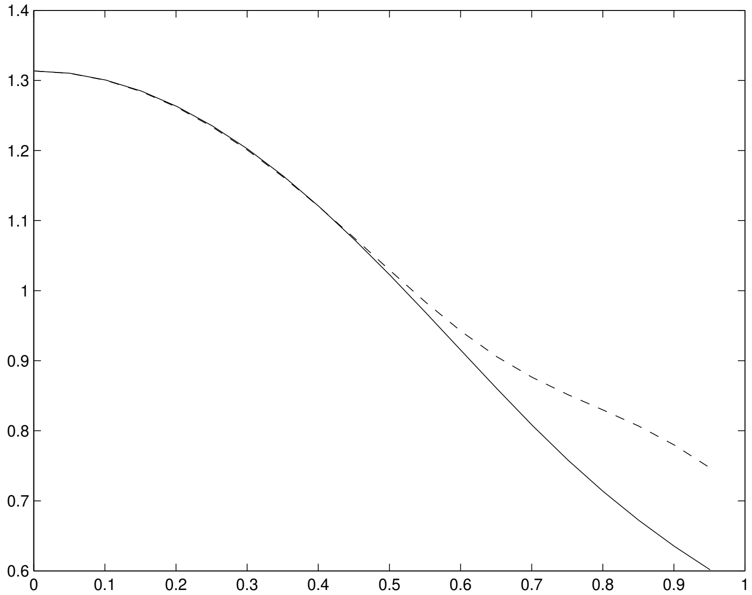

Figure 3: $E\{\det[\mathbf{M}_3]\}$ for the optimal design (dashed line) and locally–optimal design (full line) as functions of u in Example 2.

[3] D. Bertsekas. *Dynamic Programming. Deterministic and Stochastic Models.* Prentice-Hall, Englewood Cliffs, 1987.

[4] G. Box and H. Lucas. Design of experiments in nonlinear situations. *Biometrika*, 46:77–90, 1959.

[5] D. D'Argenio. Incorporating prior parameter uncertainty in the design od sampling schedules for pharmacokinetic parameter estimation experiments. *Mathematical Biosciences*, 99:105–118, 1990.

[6] V. Fedorov. *Theory of Optimal Experiments.* Academic Press, New York, 1972.

[7] V. Fedorov. Convex design theory. *Math. Operationsforsch. Statist., Ser. Statistics*, 11(3):403–413, 1980.

[8] R. Fortet. Random determinants. *J. of Research of the National Bureau of Standards*, 47(6):465–470, 1951.

[9] F. Gantmacher. *Théorie des Matrices.* Dunod, Paris, 1966.

[10] S. Keeler and P. Reilly. The design of experiments when there are errors in all the variables. *Can. J. Chem. Eng.*, 70:774–778, 1992.

[11] T. Mitchell. An algorithm for the construction of "D-optimal" experimental designs. *Technometrics*, 16:203–210, 1974.

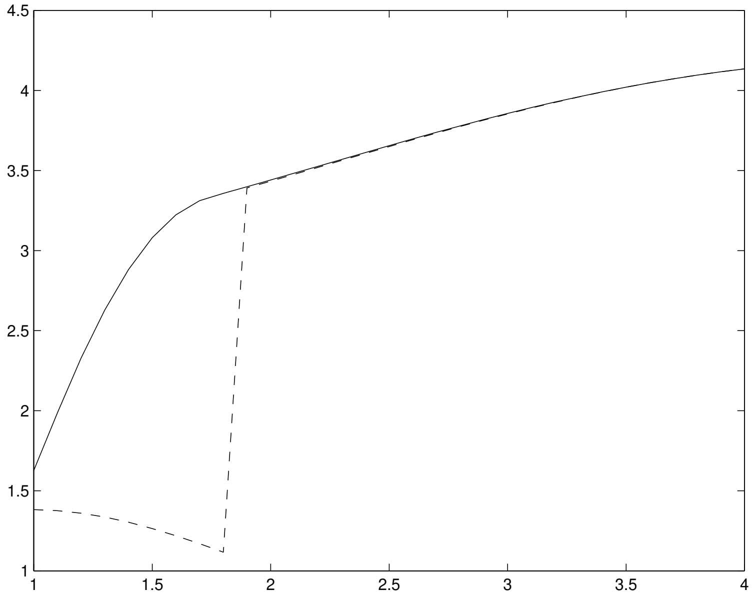

Figure 4: $E\{\det[\mathbf{M}_3]\}$ ($\times 10^{12}$) for the D optimal design (dashed line) and the optimal design (full line) as functions of T in Example 3.

[12] R. Muirhead. *Aspects of Multivariate Statistical Theory.* Wiley, New York, 1982.

[13] L. Pronzato. On a property of the expected value of a determinant. *Statistics & Probability Letters*, 39:161–165, 1998.

[14] L. Pronzato. Sequential selection of observations in randomly generated experiments. *Tatra Mountains Math. Pub.*, 17:167–175, 1999.

[15] L. Pronzato and E. Walter. Robust experiment design via stochastic approximation. *Mathematical Biosciences*, 75:103–120, 1985.

[16] J. van Lint and R. Wilson. *A Course in Combinatorics.* Cambridge University Press, Cambridge, 1992.

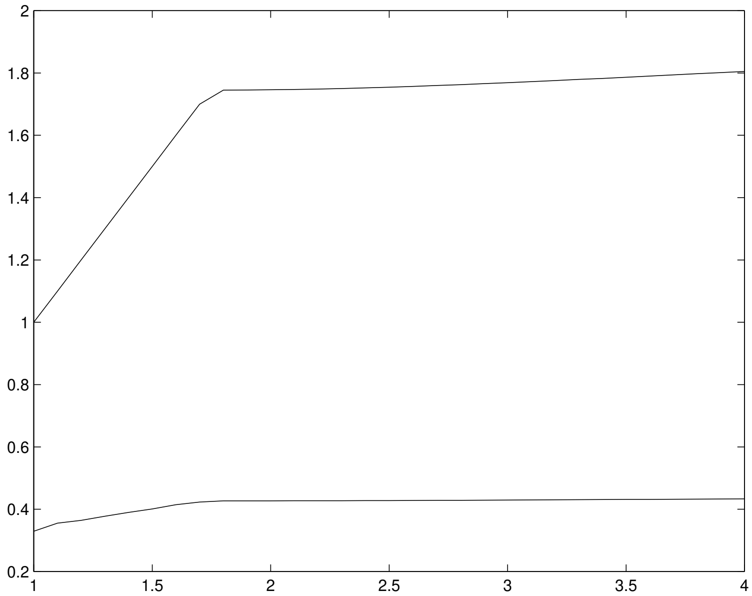

Figure 5: First two sampling times of the optimal design (full line) as functions of T in Example 3.