OPTIMAL BAYESIAN CONTROL OF A NONLINEAR REGRESSION PROCESS WITH UNKNOWN PARAMETERS

by

Nicholas M. Kiefer
Department of Economics
Cornell University
Ithaca, New York 14853

and

Yaw Nyarko
Department of Economics
Brown University
Providence, RI 02912

1. Introduction

Economic Agents operating in uncertain, stochastic environments can face a tradeoff between current period expected reward and accumulation of information of uncertain value. For example, a firm producing to meet uncertain demand might produce at the expected current reward maximizing output, based on his current beliefs about the form of the demand curve, or it might choose to experiment by varying output, thus taking short term losses in order to sharpen beliefs about the form of the demand curve. A parametric representation of the agent's problem is made by considering the utility function $u(x,y)$ and the conditional density $f(y|x,\theta)$. Here the random variable y is what the agent is trying to control (e.g., current period profits) and x is the control variable. The parameters θ of the conditional density of y given x are unknown, but the agent has opinions about θ given by a distribution μ. The agent attempts to minimize the present discounted value of the stream of expected losses, $E\Sigma\delta^t u(x_t,y_t)$, where the expectation is taken with respect to current beliefs. The problem is complicated by the fact that beliefs are updated from period to period using Bayes Rule; consequently current period actions can be expected to influence future period beliefs. This introduces stochastic dynamics into the model.

This paper considers the problem in the case in which the density $f(y|x,\theta)$ is a location family. In this case the model can be written $y = g(x,\beta) + \varepsilon$, where ε is an i.i.d. random variable whose distribution may involve unknown parameters. When $g(x,\beta) = x'\beta$ the problem is one of controlling a linear regression process with unknown parameters over an infinite horizon. Many approximate control rules for this problem have been proposed, for example sequential least-squares estimation combined with one-period optimization conditioning on the current estimates. The analogous policy for the nonlinear model is clear. In practice several policies can work "well," though it is possible to compose examples in which the policy men-

Lecture Notes in Control and Information Sciences 121
A. Blaquière (Editor)
Modeling and Control of Systems
in Engineering, Quantum Mechanics, Economics and Biosciences

tioned, for example, is easily improved. From an economic modelling point of view, however, we are interested in the optimal policy, and in the consequences for convergence of beliefs and policies of following the optimal policy. Will it be optimal for an agent to learn the parameters (and thus converge to "rational expectations")?

This paper gives general conditions under which the sequence of beliefs converges to a limit and the sequence of optimal policies converges to a limit. Under further conditions the limit policy is the optimal one-period policy for limit beliefs. Conditions under which the limit belief is point mass at true parameter values, corresponding to consistent parameter estimates are more stringent and are still under investigation.

Least-squares control rules in the linear regression model have been widely discussed and studied analytically by Taylor (1974) and Jordan (1985) and experimentally by Anderson and Taylor (1976). Improvements using a Bayesian approach were suggested by Zellner (1971) and studied by Harkema (1975). The optimal policy in the linear regression case has been studied by Kiefer and Nyarko (1987), who obtain results on convergence of beliefs and policies. convergence in a different class of models has been studied by Easley and Kiefer (1986). Results on optimal learning while controlling a stochastic process are collected along with an example in Kiefer (1988).

2. The Decision Problem: Uncertainty, Policies and Rewards

In this section we sketch the general framework we wish to study.

Let Ω' be a complete and separable metric space, let $\mathcal{F}'$ be its Borel field, and $(\Omega', \mathcal{F}', P')$ a probability space. Define the stochastic process $(\varepsilon_t)^{\infty}$ on $(\Omega', \mathcal{F}', P')$. The ε_t are assumed to be independent and identically distributed, with the common marginal distribution $p(\varepsilon_t|\xi)$ depending on some parameter, ξ in R^h, which is unknown to the agent. We assume that the set of probability measures, $(p(\cdot|\xi)$, is continuous in the parameter ξ (in the weak topology of measures); and that for any ξ, $\int \varepsilon\, p(d\varepsilon|\xi) = 0$. Let $\bar{X}$, the action space, be a compact subset of R^k. Define $\Theta = R^m \times R^h$ to be the parameter space. If the "true parameter" is $\theta = (\beta,\xi) \in \Theta$, and the agent chooses an action $x_t \in \bar{X}$ at date t, then the agent observes y_t, where,

$$y_t = g(x_t,\beta) + \varepsilon_t \tag{2.1}$$

and ε is chosen according to $p(\cdot|\xi)$. The function g is assumed measurable;

further restrictions are introduced implicitly through assumptions on the updating equation (2.2) and the reward function (2.3).

One example is the simple linear regression model with unknown slope and intercept and with the ε_t independent draws from the normal distribution with mean zero and variance σ^2. In that example Ω' is R^∞, $\mathcal{F}'$ is the collection of Borel sets on R^∞, and P' is the infinite product of independent univariate normal distributions with means zero and common variance σ^2. The parameter ξ is the variance of ε, σ^2. The action space $\check{X}$ is a closed interval in R^1. The parameter $\beta \in R^2$ consists of the slope and intercept of the regression. The space Θ is $R^2 \times R^1_+$.

Let $\mathcal{T}$ be the Borel field of Θ, and let $P(\Theta)$ be the set of all probability measures on $(\Theta, \mathcal{T})$. Endow $P(\Theta)$ with its weak topology, and note that $P(\Theta)$ is then a complete and separable metric space (see e.g., Parthasarathy (1967, Ch. II, Theorems 6.2 and 6.5)). Let $\mu_0 \in P(\Theta)$ be the prior probability on the parameter space, with finite first moment.

The agent is assumed to use Bayes rules to update the prior probability at each date after any observation of (x_t, y_t). For example, in the initial period, date 1, the prior distribution is updated after the agent chooses an action x_1, and observes the value of y_1. The updated prior, i.e., the posterior, is then $\mu_1 = \Gamma(x_1, y_1, \mu_0)$, where $\Gamma: \check{X} \times R^1 \times P(\Theta) \to P(\Theta)$ represents the Bayes rule operator. If the prior, μ_0, has a density function, then the posterior may be easily computed. In general, the Bayes rule operator may be defined by appealing to the existence of certain conditional probabilities, although some care is needed (see Diaconis and Freedman (1986)). Under some conditions the operator Γ is continuous in its arguments, and we assume this throughout. Any (x_t, y_t) process will therefore result in a posterior process, (μ_t), where for all $t = 1,2,\ldots,$

$$\mu_t = \Gamma(x_t, y_t, \mu_{t-1}) \tag{2.2}$$

Let $\check{H}_n = P(\Theta) \times \prod_{i=1}^{n-1} [\check{X} \times R^1 \times P(\Theta)]$. A <u>partial history</u>, h_n, at date n is any element $h_n = (\mu_0, (x_1, y_1, \mu_1), \ldots, (x_{n-1}, y_{n-1}, \mu_{n-1})) \in \check{H}$; h_n is said to be admissible if (2.2) holds for all $t = 1,2,\ldots, n-1$. Let $\hat{H}_n$ be the subset of $\check{H}_n$ consisting of all admissible partial histories at date n. A <u>policy</u> is a sequence $\pi = (\pi_t)_{t=1}^\infty$, where for each $t \geq 1$, the policy function $\pi_t: H_t \to \check{X}$ specifies the date t action $x_t = x_t(h_t)$, as a Borel function of the partial history, h_t in

H_t, at that date. A policy function is stationary if $\pi_t(h_t) = g(\mu_t)$ for each t, where the function $g(\cdot)$ maps $P(\theta)$ into $\tilde{X}$.

Define $(\Omega, \mathcal{F}, P) = (\theta, \mathcal{F}, \mu_0) \times (\Omega', \mathcal{F}', P')$. Any policy, π, then generates a sequence of random variables $\{(x_t(\omega), y_t(\omega), \mu_t(\omega))\}_{t-1}^{\infty}$ on $(\Omega, \mathcal{F}, P)$ as described above, using (2.1) and (2.2). See Kiefer and Nyarko (1987) for technical details.

For any $n = 1,2,\ldots,$ let $\mathcal{F}_n$ be the sub-field of $\mathcal{F}$, generated by the random variables (h_n, x_n). Notice that x_n is $\mathcal{F}_n$-measurable but y_n and μ_n are not $\mathcal{F}_n$ - measurable. Next define $\mathcal{F}_\infty = \vee_{n=0}^{\infty} \mathcal{F}_n$.

Let $u:\tilde{X} \times R^1 \to R^1$ be the utility function, so $u(x_t, y_t)$ is the utility to the agent when action x_t is chosen at date t and the observation y_t is made. The reward function $r:\tilde{X} \times P(\theta) \to R^1$, is defined by

$$r(x_t, \mu_{t-1}) = \int_\theta \int_R u(x_t, y_t) p(d\varepsilon_t | \xi) \mu_{t-1}(d\theta) \tag{2.3}$$

The inner integration marginalizes with respect to ε, given the parameter ξ, the outer integration is with respect to parameters. Assume that the reward function is uniformly bounded, continuously, and concave in x for given μ. Note that this assumption restricts $g(\cdot,\cdot)$, $U(\cdot,\cdot)$ and $p(\cdot|\cdot)$.

Let δ in $[0,1)$ be the discount factor. Any policy π generates a sum of expected discounted rewards equal to

$$V_\pi(\mu_0) = \int \sum_{t=1}^{\infty} \delta^{t-1} r(x_t(\omega), \mu_{t-1}(\omega)) P(d\omega) \tag{2.4}$$

where the (x_t, μ_t) processes are those obtained using the policy π. A policy π^* is said to be an optimal policy if for all policies π and all priors μ_0 in in $P(\theta)$, $V_{\pi^*}(\mu_0) \geq V_\pi(\mu_0)$. Even though the optimal policy, π^* (when it exists) may not be unique, the value function $V(\mu_0) = V_{\pi^*}(\mu_0)$ is always well-defined.

3. Existence of a Stationary Optimal Policy

Straightforward dynamic programming arguments can be used to show that stationary optimal policies exist and the value function is continuous.

Theorem 3.1: A stationary optimal policy $g:P(\theta) \to \tilde{X}$ exists. The value function, V, is continuous on $P(\theta)$, and the following functional equation holds:

$$V(\mu) = \max_{x \in \tilde{X}} \{r(x, \mu) + \delta \int V(\tilde{\mu}) p(d\varepsilon | \xi) \mu(d\theta)\} \tag{3.1}$$

where $\tilde{\mu} = \Gamma(x, y, \mu)$ and $y = g(x, \beta) + \varepsilon$, and where the integral is taken over $R^1 \times \Theta$.

Proof: Let $S = \{f : P(\Theta) \to R \mid f$ is continuous and bounded$\}$.

Define $T : S \to S$ by

$$Tw(\mu) = \max_{x \in \tilde{X}} \{r(x, \tilde{\mu}) + \delta \int V(\mu) p(d\varepsilon | \phi) \mu(d\theta)\} \tag{3.2}$$

One can easily show that for $w \in S$, $Tw \in S$; and that T is a contraction mapping. Hence there exists a $v \in S$ such that $v = Tv$. Replacing w with v in (3.2) then results in (3.1); and since $v \in S$, v is continuous. Finally, it is immediate that the solution to the maximization exercise in (3.2) (replacing w with v) results in a stationary optimal policy function (see Blackwell (1965) or Maitra (1968) for the details of the above arguments).

4. Convergence of the Process $\{\mu_t\}$.

In this section we prove that the posterior process converges for P-a.e ω in Ω, to a well-defined probability measure (with the convergence taking place in a weak topology).

Note that for any Borel subset, D, of the parameter space Θ, if we suppress the ω's and let, for some fixed ω, $\mu_t(D)$ represent the mass that measure $\mu_t(\omega)$ assigns to the set D, then

$$\mu_t(D) = E[1_{\{\theta \in D\}} | \mathcal{F}_t] \tag{4.1}$$

Define a measure μ_∞ on Θ by setting, for each Borel set D in Θ,

$$\mu_\infty(D) = E[1_{\{\theta \in D\}} | \mathcal{F}_\infty] \tag{4.2}$$

The measure μ_∞ is the limiting posterior distribution and is indeed a well-defined probability measure.

Theorem 4.1. The posterior process $\{\mu_t\}$ converges, for P-a.e. ω in Ω, in the weak topology, to the probability measure μ_∞.

Summary of Proof: Use (4.1) above to show that for any Borel set D in Θ, $\mu_t(D)$ is a Martingale measure, establish that the sequence of probability

measures, $\mu_t(\omega)$, for fixed ω, is tight using the assumption that the first moment of μ_∞ is finite, then apply Prohorov's Theorem (e.g., Billingsley (1968, Theorem 6.1)) to deduce that μ_∞ is a probability measure.

Note that this result on convergence of beliefs is quite different from the standard consistency result looked for in econometrics. The Martingale Convergence Theorem allows us to establish convergence, but the limit measure μ_∞ is a random variable, in the sense that it depends on the particular sequence of shocks realized. In a standard estimation problem, the limit result is that beliefs converge and the limit belief is independent of sample paths, and the limit belief is correct in the sense that μ_∞ assigns point mass to the true parameter value. Standard results do not hold here because along any sample path for which beliefs converge, the sequence of actions $\{x_t\}$ may also be converging. But if actions converge too rapidly, they may not generate enough information to identify all the unknown parameters. One can construct examples in related problems in which this phenomenon occurs (see e.g., Kiefer (1988)).

5. Optimization and Limit Beliefs and Actions

In Theorem 4.1, convergence of beliefs was established for an arbitrary $\{x_t\}$ sequence (i.e., without taking into account the underlying maximization problem). In this section we ask what action (or actions) $\bar{x}$ corresponds to the limiting beliefs μ_∞.

Theorem 5.1 establishes that the limit action is the action which maximizes single period reward for limit beliefs.

Theorem 5.1: The limit action $\bar{x} = \lim_{t\to\infty} x$ exists, is unique for given μ) and maximizes the one-period reward, $r(x,\mu_\infty)$, for limit beliefs μ_∞.

Proof of Theorem 5.1: Recall from Theorem 4.1 that $\lim_{t\to\infty} \mu_t = \mu_\infty$ exists for all sample paths. The sequence $\{x_t\}$ and $\{\mu_t\}$ satisfies for each t (simultaneously, a.e.) the functional equation

$$V(\mu_t) = r(x_t,\mu_t) + \delta\int V(\Gamma(x_t,y_t,\mu_t))p(d\varepsilon|\xi)\mu_t(d\theta). \qquad (5.1)$$

Taking limits along any convergent subsequence gives

$$V(\mu_\infty) = r(\bar{x},\mu_\infty) + \delta\int V(\Gamma(\bar{x},y,\mu_\infty))p(dc|\xi)\mu_\infty(d\theta)$$

where $\bar{x}$ is a limit point of the (x_t) sequence. (In taking the limits one uses the fact that V is bounded and the integral in (5.1) is $E[V(\mu_t)|\mathcal{F}_{t-1}]$ to apply Chung (1974, Theorem 9.4.8).) However, from convergence of beliefs $(\bar{x},y)$ yields no information so $\Gamma(\bar{x},y,\mu_\infty) = \mu_\infty$, and (5.1) becomes $V(\mu_\infty) = r(\bar{x},\mu_\infty) + \delta V(\mu_\infty)$.

Now we show that $\bar{x}$ solves the problem

$$\max_{x\in\bar{X}} r(x,\mu_\infty) \tag{5.2}$$

Suppose on the contrary that there is an $\hat{x}\in\bar{X}$ such that $r(\hat{x},\mu_\infty) > r(\bar{x},\mu_\infty)$. Then by the functional equation

$$V(\mu_\infty) \geq r(\hat{x},\mu_\infty) + \delta\int V(\Gamma(\hat{x},\hat{y},\mu_\infty))p(dc|\theta)\mu_\infty(d\theta). \tag{5.3}$$

But by Blackwell's Theorem (see e.g., Kihlstrom (1984, Lemma 1, p. 18)), since the experiment "observe $(\hat{x},\hat{y})$" is trivially sufficient for the experiment "make no observations," we obtain,

$$\int V(\Gamma(x,y,\mu_\infty))p(dc|\phi)\mu_\infty(d\theta) \geq V(\mu_\infty) \tag{5.4}$$

Hence, from (5.3) and (5.4) $V(\mu_\infty) > r(\bar{x},\mu_\infty) + \delta V(\mu_\infty)$, which is a contradiction. So $\bar{x}$ solves problem (5.2); that is, $\bar{x}$ maximizes the one-period reward $r(x,\mu)$ for limit beliefs, μ_∞. Since $r(\cdot,\mu_\infty)$ is strictly concave in x, $\bar{x}$ must be unique.

6. Conclusion

We have considered the decision problem facing an agent controlling a nonlinear regression process when parameters in the mean function and in the error distribution are unknown. The agent faces a tradeoff between accumulating information by varying the values of the regressors and accumulating one-period reward by following the one-period expected reward maximizing policy. We show that the problem can be brought into the dynamic programming framework and that the value function satisfies the usual functional equation. The sequence of beliefs about the unknown parameters is shown to converge almost surely. Further, the optimal action process converges to the one-period optimal action under limit beliefs.

7. Acknowledgements

This research is supported in part by the National Science Foundation.

REFERENCES

Anderson, T.W. and J. Taylor, (1976), "some Experimental Results on and Statistical Properties of Least Squares Estimates in Control Problems," Econometrica, 44:1289-1302.

Billingsley, P., (1968), Convergence of Probability Measures, Wiley, New York.

Blackwell, D., (1965), "Discounted Dynamic Programming," Annals of Mathematical Statistics, 36, pp. 2226-235.

Chung, K.L., (1974), A Course in Probability Theory, 2nd edition, Academic Press, New York.

Diaconis, P. and D. Freedman, (1986), "On The Consistency Of Bayes Estimates," Annals of Statistics, 14, 1-26 (discussion and rejoinder 26-27).

Easley, D. and N.M. Kiefer, (1986), "Controlling a Stochastic Process with Unknown Parameters," Cornell University working paper, forthcoming in Econometrica.

Harkema, R., (1975), "An Analytical Comparison of Certainty Equivalence and Sequential Updating," JASA, 70, 348-350.

Kiefer, N.M. and Y. Nyarko, "Control of a Linear Regression Process with Unknown Parameters" in W. Barnett, E. Berndt and H. White (eds.), Dynamic Econometric Modelling, New York: Cambridge University Press, 1987.

Kiefer, N.M., "Optimal Collection of Information by Partially Informed Agents," Cornell working paper, 1988.

Kihlstrom, R.E., (1984), "A 'Bayesian' Exposition of Blackwell's Theorem on the Comparison of Experiments," in Bayesian Models in Economic Theory, eds. M. Boyer and R.E. Kihlstrom, Elsevier Science Publishers B.V.

Jordan, J.S., (1985), "The Strong Consistency of the Least Squares Control Rule and Parameter Estimates," manuscript.

Maitra, A., (1968), "Discounted Dynamic Programming in Compact Metric Spaces," Sankhya, Ser A, 30, pp. 211-216.

Parthasarathy, K., (1967), Probability Measures on Metric Spaces, Academic Press, New York.

Taylor, J.B., (1974), "Asymptotic Properties of Multiperiod Control Rules in the Linear Regression Model," International Economic Review, 15, 472-484.

Zellner, A., (1981), An Introduction to Bayesian Inference in Econometrics, Wiley: New York.