Journal of Informatics and Mathematical Sciences
Volume 3 (2011), Number 3, pp. 201–210

http://www.rgnpublications.com

Using Random Sets to Model Learning in Manufacturing

Paul Speaker and C.R. MacCluer

Abstract. It is widely observed that manufacturing quality metrics improve as experience is gained during production. The traditional empirical *learning curves* modeling such improvements have recently been explained by a predictive model deduced from first principles, namely certain principles imported into artificial intelligence from statistical mechanics. However, this new learning model is limited to a finite lesson pool of paradigm shifts. This paper presents an extension to incremental learning using sampling based on the notion of dynamic random sets.

1. **Introduction to the learning curve**

It is widely observed that manufacturing measurements of quality (*quality metrics*) exhibit improvement as experience is gained during production. Various empirical models for these *learning curves* have taken the form of either power laws

$$C(q) = C_0 q^{-\alpha} \tag{1}$$

or as exponential laws of the form

$$C(q) = C_0 e^{-\lambda q}, \tag{2}$$

where $C(q)$ is the metric associated with the learning curve and q is the accumulated quantity produced [Zangwill and Kantor 1998, 2000]. This metric $C(q)$ is a quantitative measure that a manager would wish to minimize, such as manufacturing cost or production errors. Past models have been little more than a choice between which function is a better fit to the data, rather than a causal explanation for why this might be observed [Speaker 2009].

However, we have recently obtained an explanation for this observed learning-during-production that is based on a small number of assumptions widely held by workers in this field [Speaker and MacCluer 2009]; this result is summarized as

2010 *Mathematics Subject Classification.* Primary, 91B38; Secondary, 60D05, 90B50.
Key words and phrases. Learning curve; Manufacturing costs; Random sets; Management decision making.

Theorem A below. We fix ideas by assuming the quality metric to be production cost per unit manufactured.

Theorem A (Speaker 2009). *As production of a good proceeds, lessons are drawn (with replacement) independently from the pool of n possible lessons. Suppose the i-th lesson, once learned, yields an ongoing cost saving of c_i per unit manufactured. Once this saving is realized, it is unavailable for future improvement. Then the expected cost U per unit is predicted to decrease by the rule*

$$U(k) = U_0 - \sum_{i=1}^{n} [1 - (1 - \alpha e^{\beta c_i})^k] c_i\,, \tag{3a}$$

where U_0 is the initial cost per unit, where k is proportional to accumulated production q, where

$$\frac{1}{\alpha} = \sum_{j=1}^{n} e^{\beta c_j}, \tag{3b}$$

and where β is a measure of the effectiveness of the manager.

All of the model assumptions underlying Theorem A are strongly advocated by many investigators as detailed in [Speaker 2009].

Our goal in this paper is to first remodel cost savings as a result of a mix of *incremental* as well as discrete learning using Stieltjes integrals to replace the sums of (3a) and (3b). An attempt to obtain expected cost savings from arbitrary random sets of lessons via Robbin's theorem is seen to be impossible because of the possible correlation between lessons. We then construct a weakly convergent sequence of dynamic random sets, each of which grows over time to eventually cover all possible lessons, that yields in the weak limit a generalized dynamic random set that is the correct sampling model for dynamic learning. This generalized dynamic sampling preserves the essential refinement of lessons property, yet retains the independence of lesson choice. We conclude by showing that the incremental model devolves to the discrete model (3) when estimated.

2. A model for incremental improvement

Suppose each lesson to be learned has been labeled by exactly one real number x so that the cost savings realized by learning the lessons labeled by points lying in a measurable set X is given by the Stieltjes integral

$$\Delta C(X) = \int_X dC(x), \tag{4}$$

where the *cumulative cost savings function* $C(x)$ is an everywhere defined, nondecreasing, right-continuous function that is continuously differentiable between isolated jumps.

Points of differentiability of the cumulative cost savings function C should be thought of as *incremental improvements,* while jump discontinuities are *paradigm shifts.* For the incremental movements, improvement happens continuously over time. This type of learning is what March called *exploitation* [March 1991] and is considered *supervised learning* in the artificial intelligence context [Russell and Norvig 2002].

On the other hand, the paradigm shift corresponds to March's idea of the *exploration* aspect of learning, which is considered *unsupervised learning* in the artificial intelligence context. Learning under the category of exploration includes, for example, new technology, new worker insight to production, new cost structure, material replacement, etc. Rather than continuous improvement, this type of learning takes the form of sudden leaps forward. While continuous improvement may be modeled with a continuum of learning choices, explorative learning takes the form of a finite number of breakthroughs.

As was argued for the discrete case in [Speaker and MacCluer 2009], the probability $\mathrm{prob}(X)$ of choosing the lessons forming the measurable set X is a smooth function $p = p(y)$ of the cost savings $y = \Delta C(X)$ resulting from these lessons:

$$\mathrm{prob}(X) = p\left(\int_X dC(x)\right). \tag{5}$$

Because the largest number of distinct learning trials — by Boltzmann's H-theorem [Feynman 1972] — occurs simultaneously with maximum entropy, it follows that $p(y) = \alpha \exp(\beta y)$, where the normalization α will be determined below. See [Speaker and MacCluer 2009].

In particular, if F is the *cumulative probability that all lessons indexed by x or less will be chosen,* then

$$F(x) = \alpha \exp\left(\beta \int_{-\infty}^{x} dC(y)\right) = \alpha \exp(\beta C(x)), \tag{6a}$$

where

$$\frac{1}{\alpha} = \exp\left(\beta \int_{-\infty}^{\infty} dC(x)\right) = \exp\left(\beta C(\infty)\right). \tag{6b}$$

Note that the corresponding probability density function must be

$$f(x) = F'(x) = \alpha\beta \exp(\beta C(x)) c(x) = \beta F(x) c(x), \tag{7a}$$

where

$$c(x) = C'(x) \tag{7b}$$

at points x of differentiability of C. Jumps of C of height c_0 will yield jumps in F of height $\alpha \exp(\beta c_0)$.

But now suppose X is a *random set* of lessons, that is, X is a measurable-set-valued random variable (see Appendix). Each random set X possesses a *hitting probability* $T_X(x)$, which is the probability that x lies in X. This hitting probability is obtained by integrating the characteristic function $\chi_X(x)$ over the probability space of the setvalued random variable X.

For any integrable realvalued function g of a real variable x we have the famous theorem of Robbins [1944] on expected values:

$$E\left[\int_X g(x)dx\right] = \int_{-\infty}^{\infty} g(x)T_X(x)dx = E[g(x) \mid x \in X]. \tag{8}$$

(The first equality of (8) is nothing more than an application of Fubini's theorem — see the Appendix.)

Remark. At first glance, the expected cost savings from learning all the lessons from this random sample of lessons X would from (4) appear to be

$$\Delta\overline{C}(X) = E\left[\int_X dC(x)\right] = \int_{-\infty}^{\infty} T_X(x)\,dC(x). \tag{9}$$

The second equality of (9) is indeed a correct statement of probability. However, lesson choice must be independent for this to be the learning observed in manufacturing — see [Speaker 2009]. Moreover, this independence is necessary to obtain the probability structure (6). Unfortunately, a random set X may possess correlated points, that is, where $x_1 \neq x_2$ yet

$$\text{prob}(x_1 \in X \text{ and } x_2 \in X) \neq T_X(x_1)T_X(x_2).$$

But when correlation is absent, (4) holds for random lesson choice.

Result A. *If lesson choice from X is uncorrelated, then the expected cost savings from learning the random set of lessons of X is indeed given by*

$$\Delta\overline{C}(X) = E\left[\int_X dC(x)\right] = \int_{-\infty}^{\infty} T_X(x)\,dC(x). \tag{10}$$

Our task then is to construct an evolving random lesson sampling technique that grows over time to eventually cover all lessons, a process suggested by Robbins's 1944 modeling of airfield carpet bombing. This dynamic model of learning must allow *uncorrelated sampling of lessons* as well as *additivity of the resulting cost savings* when compound lessons are refined into independent sublessons. Both properties are necessary if we hope to retain the probability structure (6).

3. **Constructing dynamic random sampling**

We present here only the finite incremental learning case, where cumulative cost saving $C = C(x)$ is everywhere continuously differentiable and where $C'(x) = c(x)$ has compact support.

Suppose that learning is taking place at the fixed Poisson rate λ so that the lessons that comprise the measurable set L require $t = \mu(L)/\lambda$ seconds to learn, where μ is Lebesgue measure.

The construction. For each natural number k let $X_k(t)$ denote the random set of lessons that is the union of k intervals of type $[a, a + \lambda t/k)$, where the k left end points a are the result of k independent draws with probability given by the density function f of (7). Intuitively, and as the next result shows,

$$E[\mu(X_k(t))] \leq \lambda t, \tag{11}$$

that is, the lessons of $X_k(t)$ require on average at most t seconds to learn. Thus $X_k(t)$ is a random sample of k compound lessons of a certain type which can be learned in time t.

Lemma A. *Let $T_{X_k(t)}(x)$ denote the hitting probability of $X_k(t)$, namely the probability that x will lie in $X_k(t)$. Then*

$$T_{X_k(t)}(x) = 1 - \left[1 - \int_{x-\lambda t/k}^{x} f(y)\,dy\right]^k. \tag{12}$$

Proof. The point x will lie in the random interval $[a, a + \lambda t/k)$ exactly when $x \geq a > x - \lambda t$ with probability

$$\text{prob}\big(a \in (x - \lambda t/k, x]\big) = \int_{x-\lambda t/k}^{x} f(y)\,dy. \tag{13}$$

Thus x lies in none of the k intervals $[a, a + \lambda t)$ forming $X_k(t)$ with probability

$$\left[1 - \int_{x-\lambda t/k}^{x} f(y)\,dy\right]^k,$$

giving (12). □

Lemma B. *Lesson choice is asymptotically independent: Let*

$$T_{X_k(t)}(x_1, x_2) = \text{prob}\big(x_1 \in X_k(t) \textit{ and } x_2 \in X_k(t)\big). \tag{14}$$

If $x_1 \neq x_2$, then for all sufficiently large k,

$$T_{X_k(t)}(x_1, x_2) = T_{X_k(t)}(x_1) \cdot T_{X_k(t)}(x_2). \tag{15}$$

Proof. When $\lambda t/k < |x_2 - x_1|$ it is impossible for both x_1 and x_2 to belong to the same one of the k random intervals $[a, a + \lambda t/k)$ that comprise $X_k(t)$.

Thus as k increases, the sample $X_k(t)$ consisting of k ever-shortening compound lessons begins to reassemble an independent sample of individual lessons, all of which can be learned in time t. □

Lemma C. *The hitting probabilities converge pointwise. In fact, for each x and $t \geq 0$, we have*

$$\lim_{k\to\infty} T_{X_k(t)}(x) = 1 - e^{-\lambda t f(x)}. \tag{16}$$

Proof. Since the probability density f has compact support,

$$\int_{x-\lambda t/k}^{x} f(x)\,dx = \frac{\lambda t f(x)}{k} + O(1/k^2). \tag{17}$$

Therefore

$$\lim_{k\to\infty}\left[1-\int_{x-\lambda t/k}^{x} f(y)\,dy\right]^k = \lim_{k\to\infty}\left[1-\frac{\lambda t f(x)}{k} + O(1/k^2)\right]^k$$

$$= \exp(-\lambda t f(x)). \tag{18}$$

□

4. **Prediction of unit cost**

As seen in Lemma B, our random sample $X_k(t)$ of k compound lessons, learnable in t seconds, approaches for large k an independent sampling of individual lessons. By Lemma C, the corresponding hitting probability $T_{X_k(t)}$ approaches the limiting simple expression (16). Therefore, it is more than plausible that in the limit we have obtained a prediction of the decrease in cost-per-unit during manufacturing.

Theorem B. *Assume cumulative cost saving $C = C(x)$ is continuously differentiable and that its derivative has compact support. If learning is proceeding at the Poisson rate λ, then the expected production cost U per unit is given by the rule*

$$U(t) = U_0 - \int_{-\infty}^{\infty} (1 - e^{\lambda t f(x)})\,dC(x), \tag{19}$$

where U_0 is the initial cost per unit at the onset of production and where the probability density f is given by

$$f(x) = \alpha \exp(\beta C(x)) C'(x). \tag{20}$$

Proof. The unit cost is decreased from its initial cost by the expected cost saving per unit from learning during time t, i.e.,

$$U(t) = U_0 - \Delta\overline{C}(t). \tag{21}$$

But as argued above,

$$\Delta\overline{C}(t) = \lim_{k\to\infty} E\left[\int_{X_k(t)} dC(x)\right].$$

Hence (16) will then yield (19). □

5. **Devolution to the discrete case**

As further evidence that Theorem B gives the correct model for unit cost evolution resulting from incremental learning, let us demonstrate that it cuts back to the discrete case of Theorem A.

Let $I = [a, b]$ be the smallest closed interval containing the compact support of f. We may, as always, translate and rescale our lesson labeling system at will, so that in this case $I = [0, 1]$. Let us partition I in the usual way into n congruent subintervals $[x_{i-1}, x_i]$, with $0 = x_0 < x_1 < \cdots < x_n = 1$ and $\Delta x_i = x_i - x_{i-1} = 1/n$.

Then as in (4) the cost saving accrued by the compound lesson $[x_{i-1}, x_i)$ is

$$c_i = \int_{x_{i-1}}^{x_i} dC(x), \tag{22a}$$

which will be chosen via (7) with probability

$$p_i = \int_{x_{i-1}}^{x_i} f(x)\, dx = \frac{f(x_i^*)}{n}. \tag{22b}$$

We take snapshots in time t_k under the time scaling where one compound lesson $[x_{i-1}, x_i)$ is learned on average per second. Then because λ is the average rate of learning per lesson length, we have $t_k \lambda = k/n$.

Therefore the Riemann-Stieltjes sum approximation of the integral of (19) takes on the form

$$\int_{-\infty}^{\infty} 1 - e^{-\lambda t_k f(x)}\, dC(x) = \int_0^1 1 - e^{-kf(x)/n}\, dC(x) \tag{23}$$

$$= \sum_{i=1}^{n} \left[1 - e^{-kf(x_i^*)/n}\right] c_i + O(1/n^2)$$

$$= \sum_{i=1}^{n} \left[1 - \left(1 - \frac{f(x_i^*)}{n}\right)^k\right] c_i + O(k/n^2) \tag{24}$$

$$= \sum_{i=1}^{n} [1 - (1 - p_i)^k]\, c_i \;+\; O(k/n^2), \tag{25}$$

which is consistent with the discrete result (3).

6. Summary

We have obtained a prediction of how cost per unit falls during the manufacturing of a good when learning is incremental. When incremental lessons are grouped into discrete compound lessons, the model devolves into a previously obtained prediction of unit cost when learning occurs in discrete jumps.

Acknowledgments

We thank Dr. David Field of General Motors Research for initially proposing this quest for the underlying basis for the learning curve. We thank Professor Peter Magyar of Michigan State University for managing the student industrial project team that first examined this problem. We also thank Professor Roger Calentone of Michigan State University for his guidance on matters of supply chain management.

Appendix

Let us nail down the somewhat slippery notion of a random set.

Definition. Let $\mathscr{M}$ denote the family of all measurable subsets B of the real numbers $\mathbf{R}$. Let $\Sigma(\mathscr{M})$ be a σ-algebra of subsets S of $\mathscr{M}$ which contains at the very least, for each real x, the collection S_x of all B that contain x. Suppose π is a probability measure on $\Sigma(\mathscr{M})$. Then the identity map

$$X : \mathscr{M} \to \mathscr{M} \tag{A1}$$

is called the *random set* of the probability space $\big(\mathscr{M}, \Sigma(\mathscr{M}), \pi\big)$.
(Some authors prefer to restrict the setvalues that X can take on [Goutsias], a practice which conflicts with the traditional usage for realvalued random variables. In our formulation, the probability π often may possess small support so that X may take on most setvalues almost never.)

The function $T_X : \mathbf{R} \to [0,1]$ given by the expected value that x lies in X, in symbols

$$T_X(x) = E[\chi_X(x)] = \int_{\mathscr{M}} \chi_B(x)\, d\pi(B), \tag{A2}$$

is called the *hitting probability* of x in X.

Robbins's Theorem (1944). *For any integrable realvalued function g of a real variable x,*

$$E\left[\int_X g(x)\, dx\right] = \int_{\mathbf{R}} g(x) T_X(x)\, dx. \tag{A3}$$

Proof. By Fubini's theorem,

$$\begin{aligned} E\left[\int_X g(x)\, dx\right] &= \int_{\mathscr{M}} \int_B g(x)\, dx\, d\pi(B) \\ &= \int_{\mathscr{M}} \int_{\mathbf{R}} \chi_B(x) g(x)\, dx\, d\pi(B) \\ &= \int_{\mathbf{R}} g(x) \int_{\mathscr{M}} \chi_B(x)\, d\pi(B)\, dx \\ &= \int_{\mathbf{R}} g(x) T_X(x)\, dx. \end{aligned} \tag{A4}$$

Let us revisit the construction of the random sets $X_k(t)$ of §3. Fix $t > 0$ and set $\lambda t = b$. Suppressing all mention of t, the random set X_k is a category of sets of special type, namely sets that are the union of k half-open intervals $[a, a + b/k)$, each such union determined uniquely (within permutations of subscripts) by the k-tuple $(a_1, a_2, \ldots, a_k)$ of left endpoints, chosen by k successive independent draws with probability density f. Since the outcome of the random variable X_k is declared to be only such unions, these unions form the support K_k of the probability π_k of the probability space $(\mathscr{M}, \Sigma_k(\mathscr{M}), \pi_k)$ belonging to this random set X_k. Let us now deduce $\Sigma_k(\mathscr{M})$ and π_k.

Let S be an element of $\Sigma_k(\mathscr{M})$, that is, a π_k-measurable set of measurable sets. Since the subset $S \setminus K_k$ of all measurable sets of S not of type X_k has π_k-measure zero and may be discarded, the remaining subcollection $S \cap K_k$ occurs with relative frequency

$$\begin{aligned}\pi_k(S) &= \pi_k(S \cap K_k)\\ &= \lambda_k(A_S)\\ &= \int_{A_S} f(a_1)f(a_2)\cdots f(a_k)\, da_1 \cdots da_k\,,\end{aligned} \tag{A5}$$

where A_S is the set of all left endpoints $(a_1, a_2, \ldots, a_k)$ (over all permutations of their subscripts) of the sets in S of type X_k. The proper sigma algebra $\Sigma_k(\mathscr{M})$ is thus seen to be the smallest collection of all sets S of measurable sets B such that the left endpoints $(a_1, a_2, \ldots, a_k)$ of members $B \in S$ of type X_k form a measurable set A of $\mathbf{R}^k$.

Result. *The sequence of dynamic random sets X_k of §3 converges weakly in the sense that for each integrable realvalued function g of a real variable and each $t > 0$,*

$$E[g(x)|x \in X_k] = \int_{\mathbf{R}} g(x)T_{X_k}(x)\,dx \to \int_{\mathbf{R}} g(x)\big(1 - e^{-\lambda t f(x)}\big)\,dx. \tag{A6}$$

Proof. The result follows from the Lebesgue dominated convergence theorem. □

Remark. We have been unable to decide whether or not these random sets $X_k(t)$ of §3 converge to a random set. It is impossible for their associated probability measures to converge strongly (in distribution) to a measure associated with some putative limit random set, since such strong convergence leads to statements contradicting Choquet's theorem — that a capacity functional for a random set has to be uppersemicontinuous [Goutsias]. However, it may be possible that the $X_k(t)$ converge in some useful weak sense like the above to an actual random set with the hitting probability given in (16). The weak limit [Billingsley 1968] of the probability measures λ_k of (A5) to a probability product measure on $\mathbf{R}^\infty$ is one possible candidate for the probability associated with this limit random set (once proper identifications are made). But in the end it may surprisingly turn out that learning during the production of a good is given by a *generalized, but not actual* evolving random set with a generalized hitting probability given by (16).

References

[1] P. Billingsley, *Convergence of Probability Measures,* John Wiley and Sons, New York, 1968.

[2] R.P. Feynman, *Statistical Mechanics,* Addison Wesley, 1972.
[3] J. Goutsias, P.S. Mahler and H.T. Nguyen, *Random Sets,* IMA Volumes in Mathematics and its Applications, Springer, New York 1997.
[4] E.E. Robbins, On the measure of a random set, *Annals of Mathematical Statistics* **15** (1944), 70–74.
[5] Paul Speaker, *Mathematical Models of the Manufacturing Learning Curve,* Dissertation, Michigan State University, 2009.
[6] Paul Speaker and C.R. MacCluer, A derivation of the manufacturing learning curve, *European Journal of Operations Research*, 2009 (to appear).
[7] W.I. Zangwill and P.B. Kantor, Toward a theory of continuous improvement and the learning curve, *Management Science* **44**(7) (1998), 910–920.
[8] W.I. Zangwill and P.B. Kantor, The learning curve: a new perspective, *International Transactions in Operations Research* **7**(6) (2000), 595–607.

Paul Speaker, *Department of Mathematics, Michigan State University, East Lansing, MI 48824, USA.*
E-mail: `speakerp@msu.edu`

C.R. MacCluer, *Department of Mathematics, Michigan State University, East Lansing, MI 48824, USA.*
E-mail: `maccluer@msu.edu`

Received April 8, 2011
Accepted July 26, 2011