King Fahd University of Petroleum & Minerals

DEPARTMENT OF MATHEMATICAL SCIENCES

Technical Report Series

TR 431

Jan 2014

Measuring Acquisition of Statistical Knowledge with a Birth-and-Death Stochastic Process

M. Hafidz Omar and Ibrahim Rahimov

DHAHRAN 31261 ● SAUDI ARABIA ● www.kfupm.edu.sa/math/ ● E-mail: c-math@kfupm.edu.sa

Measuring Acquisition of Statistical Knowledge with a Birth-and-Death Stochastic Process

M. Hafidz Omar[a*], Ibrahim Rahimov[b]

[a]Dept of Mathematics and Statistics, King Fahd Univ of Petroleum &Minerals, PO Box 1161, Dhahran 31261, Saudi Arabia.
[b]Dept of Mathematics and Statistics, Zayed University, PO Box 19282, Dubai, UAE.

Abstract

This paper presents a birth-and-death stochastic model for the process of knowledge acquisition in typical university courses. In particular, we model the acquisition of statistical knowledge in an introductory statistics class. Because concepts in statistics are very much inter-related, students' acquisition of knowledge and their failure to acquire such knowledge is at the heart of their success in Statistics classrooms. We propose the use of the Birth-and-Death Stochastic Process to model the acquisition and non-acquisition rate of statistics knowledge. Based on end-of-chapter quizzes, we were able to fully characterize the students' acquisition level of statistics knowledge. And as a validity measure, we utilize the probabilities of acquisition and of forgetting to predict students' achievement on their comprehensive final exam scores. The acquisition and forgetting probabilities as a group provides a better predictor set than the typical major exams for the course. We found that Birth-and-Death process holds promise for modeling the process of students' learning of statistics. We note that the model may also hold promise in other but similar courses.

Keywords: Birth and Death process; Statistics learning; acquisition; retention; forgetting; regression.

1. Introduction and Literature Review

Success in university courses, in part, depends on students' achievement in various interrelated aspects of knowledge in a course. Instructor's effectiveness in imparting knowledge and skill in these courses depend on not only having the necessary knowledge and skills but also designing the mode and rate at which the subject matter must be imparted to the novice students. However, many university courses offered have knowledge and skill units that are developmentally pre-requisite of other more advanced units and as such are very much inter-related.

In particular, Statistics courses have concepts that are very much inter-related with a few exceptions. Typically students are expected to master not only the material at hand but also previous material. For example, when students learn the estimation chapter, they are expected to have learned the central limit theorem to a satisfactory degree. However, to learn the central limit theorem, students need to have mastered the following main concepts which are:

1) descriptive statistics
2) sampling
3) sampling replications
4) rules for limits and
5) the normal probability distribution.

Although the first three and the last concepts are typically taught in the same course, the fourth

[*] Corresponding author. Tel.: +966-3-860-2471; fax: +966-3-860-2340.
E-mail address: omarmh@kfupm.edu.sa.

concept is taught in a previous course, namely calculus. In addition, students are also expected to apply the following pre-requisite concepts from the same statistics course or from a previous mathematics course:

1) probability rules
2) rules for continuous probability distribution
3) basic mathematical ideas for areas under the curve and
4) mathematical rules for finding inverse of a continuous function.

Students who successfully acquire all concepts have successfully committed these concepts into their long-term memory. However, their success in using these concepts partly relies on their memory system reconsolidation (Sara, 2000). As described by Snow and Lohman (1993), incoming visual and auditorial stimuli such as in classroom lectures must be received and held long enough in a sensory system, segmented or synthesized, and recognized, encoded, or otherwise represented in a memory system so that attention can be directed and further cognitive work can be done. Students who fail to acquire the knowledge concepts may commit their understanding of concepts into short-term memory only, may not consolidate their memory, and thus a forgetting or retention-loss phenomenon may occur as pointed out by many researchers (Dudai, 2004; Maren, 1999; Lee, Everitt, & Thomas, 2004). According to Wixted and Carpenter (2007), the time-to-forgetting can best be modeled by a decaying power function. That is, without effort to reinforce knowledge consolidation, memory of learning units may decay quite rapidly.

Jaber and Bonney (1997) discussed several models that incorporate forgetting in learning curves to produce industrial products. These learning curves were studied as a function of time from initial production and from momentary stoppages of production lines. Jaber and Bonney's work did not look at the effects of length of breaks on the learning curves.

Globberson et. al (1989) studied the effects of breaks on forgetting when subjects were performing repetitive tasks of data-entering 16 forms. The break length between two data entry sessions varied from 1 to 82 days. Apparently, the longer the production times to produce a product, the more the productivity. However, when the stoppage from producing a product is longer, the forgetting becomes greater.

As acquisition of statistics knowledge concepts also require repetitive tasks to consolidate understanding and consolidate memory systems of concepts, work by Globberson et, al (1989) and Jaber and Bonney (1997) appear to be quite important to shed light on statistics learning. Snow and Lohman (1993) also stressed that an important hypothesis in accumulating information to build a novel train of thought has been the characteristics of the initial perception-memory-attention system, and the skills involved in working this initial system which are fundamentally important in conditioning the success of further cognitive processing. Groen and Parkman (1972) observed that as computational skills are over-learned, response times become quite rapid. They observed that older children and adults for whom computation has become automatic respond rapidly to most types of computational problems.

Recently, some researchers (Janoos, Brown, Morocz, and Wells III, 2013) have used a state-space analysis of working memory. The model used by Janoos et. al (2013) Requires understanding of the neurocognitive processes underlying the working memory of impairment in schizophrenic subjects.

As for our study, we require the understanding of the students' learning process as they acquire new lesson units and as such our proposed model in this paper will concentrate more on how students naturally progress through their statistics course. Teaching Statistics can be quite a challenge when students do not learn the basic pre-requisite knowledge from pre-requisite courses or from previous key building-block chapters. It seems that a good assessment and learning system needs to measure evidence of student mastery of these prerequisite knowledge concepts. But, how and when do we measure these evidences? Also as important is how do we use the evidence?

In this paper, we suggest using a discrete-time homogeneous Birth-and-Death stochastic process model to measure the time series evidence of students' pre-requisite knowledge. The typical structure of a statistics course often introduces concepts sequentially and short quizzes can be given at the end of every complete discussion of concept. This opens the potential to apply the Birth-and-Death stochastic process model.

2. Method

Several important terminologies from the Birth and Death Stochastic Process need to be understood in terms of this Statistics learning process. In this paper, we call the successful **acquisition** of a concept as the *birth* of a memory consolidation of the concept while **forgetting** a concept is termed the *death, decay*, or *disintegration* of the memory of a concept. Meanwhile, failing to acquire a concept is the *non-birth* of a learner's memory for the concept. Since statistical concepts are taught sequentially, a student can accumulate mastery of these concepts in the following ways;

a) He/she can sequentially and successfully acquire all n concepts, which means he/she has passed the course successfully, where n is the total number of concepts in the course as presented in learning units/chapters.
b) He/she can acquire a subset of k concepts (where $1 < \boldsymbol{k} < \boldsymbol{n}$)
c) He/she can acquire some concepts but forget them in which case he/she is left with a subset of *k* concepts
d) He/she does not acquire any concept in which case he/she has $\boldsymbol{k}$=0 understanding of concepts.
e) He/she does not acquire any more concepts after acquiring k concepts.

The following diagram shows these possible transitions for the accumulation of ***n*** concepts in a typical ***n***-unit statistical course.

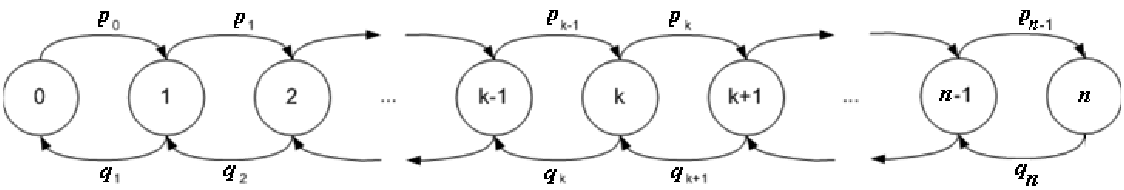

Fig. 1. A Discrete time homogeneous Birth-and-Death Stochastic Process Model of Statistics Learning

Nomenclature

$p_{k,k+1}$	one-step transition probability from state $\boldsymbol{k}$ to the $(\boldsymbol{k}+1)^{th}$ concept acquisition
$p_{k,k-1}$	one-step transition probability from state $\boldsymbol{k}$ to the $(\boldsymbol{k}-1)^{th}$ concept death/decay/disintegration
$p_{k,k}$	transition probability from state $\boldsymbol{k}$ to the $\boldsymbol{k}^{th}$ concept retention
p_k	birth rate in state $\boldsymbol{k}$ to move to the $(\boldsymbol{k}+1)^{th}$ concept acquisition
q_k	knowledge death/decay rate in the $\boldsymbol{k}^{th}$ state to reduce to the $(\boldsymbol{k}-1)^{th}$ state.

2.1. Assumptions of the model and their implications

The model of the memory consolidation system with state space $\mathbf{Z}=\{0,1,2,....,\boldsymbol{n}\}$ where $\boldsymbol{n} < \infty$ is a *Birth-and-Death process* if from any memory consolidation state $\boldsymbol{k}$ in $\mathbf{Z}$ only transition to $\boldsymbol{k}$-1 or $\boldsymbol{k}$+1 is possible, provided that $\boldsymbol{k}$-1 and $\boldsymbol{k}$+1 are also respectively in $\mathbf{Z}$.

The transition probabilities for each unit of time of the Birth-and-Death process have the following properties:

a) transition probability from state $\boldsymbol{k}$ to state $\boldsymbol{k}+1 = \boldsymbol{p_{k,k+1}} = \boldsymbol{p_k}$
b) transition probability from state $\boldsymbol{k}$ to state $\boldsymbol{k}-1 = \boldsymbol{p_{k,k-1}} = \boldsymbol{q_k}$
c) transition probability from state $\boldsymbol{k}$ to state $\boldsymbol{k} = \boldsymbol{p_{k,k}} = 1-(\boldsymbol{p_k}+\boldsymbol{q_k})$
d) $\boldsymbol{j}$-step transition probability from state k to state $\boldsymbol{k}-\boldsymbol{j} = \boldsymbol{p_{k,k-j}} = 0$ (where $\boldsymbol{j}>1$)
e) $\boldsymbol{j}$-step transition probability from state k to state $\boldsymbol{k}+\boldsymbol{j} = \boldsymbol{p_{k,k+j}} = 0$ (where $\boldsymbol{j}>1$)
f) $\boldsymbol{q}_0 = 0$

That is, the implication of condition (f) on the transition probability is that there is no further death/decay beyond those $\boldsymbol{n}$-unit concepts in the typical statistics course. Conditions (a) through (e) shows the sequential nature of the knowledge acquisition process in the typical $\boldsymbol{n}$-unit course. That is, students need to cumulatively acquire each unit successfully in order to successfully acquire the next concept. Condition b above describes the death/decay of the knowledge in the memory system. Condition c describes the rate of non-acquisition and non-decay of knowledge concepts in the memory system. This probability describes the chance for a student to retain a knowledge concept in his/her memory system without advancing to the next memory state or reverting to the previous state. As for conditions d and e, the chance for acquiring more than 1 knowledge concepts in one time and the chance for losing/forgetting more than 1 knowledge concept is zero.

In addition, to ensure that a Birth-and-Death Process is irreducible, we need to supplement the listed assumptions above with the following conditions:
a) $\boldsymbol{p_k} > 0$ for $\boldsymbol{k} = 0,1,\ldots,\boldsymbol{n}-1$ and
b) $\boldsymbol{q}_k > 0$ for $\boldsymbol{k} = 1,\ldots,\boldsymbol{n}$.

2.2. Measurement of the transition probabilities

After acquiring the $\boldsymbol{k}^{\text{th}}$ concept, acquisition of new concept, retention of the $\boldsymbol{k}^{\text{th}}$ material and forgetting the $\boldsymbol{k}^{\text{th}}$ material are mutually exclusive categories in our model. For the purpose of measuring the transition probabilities, we define the following:
a) $\boldsymbol{p_k} = \boldsymbol{p_{k,k+1}}$ = proportion of correct scores on the $\boldsymbol{k}+1^{\text{th}}$ quiz and
b) $1-\boldsymbol{p_k} = Pr(\boldsymbol{m}=\boldsymbol{k}$ or $\boldsymbol{k}-1 \mid \boldsymbol{j}=\boldsymbol{k})$ = proportion of incorrect scores on the $\boldsymbol{k}+1^{\text{th}}$ quiz.

In particular, $1-\boldsymbol{p_k}$ is the probability of non-acquisition of the $\boldsymbol{k}+1^{\text{th}}$ concept. At this point, this probability is a combination of both the retention probability $Pr(\boldsymbol{m}=\boldsymbol{k} \mid \boldsymbol{j}=\boldsymbol{k})$, and the forgetting/decay probability $Pr(\boldsymbol{m}=\boldsymbol{k}-1 \mid \boldsymbol{j}=\boldsymbol{k})$.

To separately measure which part of this total probability is related to retention of the $\boldsymbol{k}^{\text{th}}$ concept and which part is due to forgetting /decay of this $\boldsymbol{k}^{\text{th}}$ concept, we define the following:
a) $\boldsymbol{w_k}$ = proportion of a linear combination of absence and complement of the homework score on the $\boldsymbol{k}^{\text{th}}$ unit as below

$$\frac{1}{3}\left(1-\frac{\text{attendance on the } k^{\text{th}} \text{ unit}}{\text{Total classes on the } k^{\text{th}} \text{ unit}}\right)+\frac{2}{3}\left(1-\frac{\text{Student score on the } k^{\text{th}} \text{ unit homework}}{\text{Max possible score on the } k^{\text{th}} \text{ unit homework}}\right).$$

The rationale is that since forgetting means the student have obtained the material but then forget due to not reinforcing their knowledge, the first part is related to their forgetting after acquiring knowledge from not attending the classes while the second part is related to their forgetting due to not personally investing time in doing the homework. Also, although students may come to class, the more they don't personally do their work, the more they'll forget. Thus the second part, which deals with forgetting due to not doing homework, is weighted 2 times more than not keeping regular attendance.

b) $Pr(\boldsymbol{m}=\boldsymbol{k}-1 \mid \boldsymbol{j}=\boldsymbol{k}) = \boldsymbol{q_k} = \boldsymbol{w_k}(1-\boldsymbol{p_k})$ = probability of **forgetting** the $\boldsymbol{k}^{\text{th}}$ concept after acquiring these concepts and
c) $Pr(\boldsymbol{m}=\boldsymbol{k} \mid \boldsymbol{j}=\boldsymbol{k}) = 1-\boldsymbol{q_k}-\boldsymbol{p_k}$ = probability of **retaining** the $\boldsymbol{k}^{\text{th}}$ statistics concept after acquiring these concepts.

2.3. Data and Validation of model

The data for the estimation of the model transition probabilities come from a STAT211 Business Statistics I course at King Fahd University of Petroleum and Minerals in the Fall semester of 2011. There were 90 students who took 7 end-of-chapter quizzes for the course.

For the purpose of validation of the model, we calculated the individual probabilities of knowledge acquisition, individual probabilities of knowledge failures, joint probability of knowledge acquisition and the joint probability of knowledge decay for each student. To examine the predictive power of the proposed Birth-and Death process model, we then use these probabilities to predict the students' comprehensive final exam scores. In particular, use the following predictor variables to explain the variation in the comprehensive final exam scores:

1) Major Exam 1 score
2) Major Exam 2 score
3) The sum of acquisition probabilities (denoted as "sum *p*")
4) The sum of forgetting probabilities (denoted as "sum *q*")
5) The individual acquisition probabilities
6) The individual forgetting/decay probabilities
7) The product of acquisition probabilities (denoted as "product *p*")
8) The product of forgetting probabilities (denoted as "product *q*")
9) The product of retention probabilities (denoted as "product *r*")

With the above predictors, the general estimated regression model is given below:

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_p X_p .$$

For the purpose of benchmarking, we shall compare the performance of the elements of the Birth-and-Death Stochastic Process model against the performance of the regression model with major 1 and major 2 exam scores. Note that we do not include the retention probabilities in the regression model because by the Birth and Death process model, they are linearly dependent on the sum of the acquisition and forgetting probabilities from state ***k***.

3. Results

Results of the validation by regression model are summarized in this section. Table 1 provides the performance of the predictor variables in explaining the comprehensive final exam scores. Table 2 further investigates which elements of the Birth-and-Death Process model are good predictors of the comprehensive final exam scores.

Table 1 presents the predictors of comprehensive final exam scores and the characteristics of each regression model such as the numerator and denominator degrees of freedom (df), the percentage of variation explained (R^2), the R^2 adjusted for the number of predictors (R^2 adj), the regression standard error of estimate (s), the model F-test ratio, and the observed significance level, p-value (P). The best regression model would have a significant model F-test ratio where the p-value would be less than α = 0.05 significance level, the smallest standard error of estimate (s), the largest percentage of variation explained (R^2) and the largest adjusted R^2.

As pointed out earlier, the model in the first row which contains both the first and second major exam score serves as a benchmark to compare other models. This is mainly because in the STAT211 course, these two exams make up a *summative* performance measure of a number of concepts in the course. Also, the majority of the STAT211 grade, (exactly 80%) comprises mainly of the first, second exam, and the comprehensive final exam in the ratio of 1:1:2. As such, the regression model with major exam scores as predictors appears the best logical reference to compare other competing models against. From the first row of Table 1, major 1 and major 2 exams together explain 29.1% of the variation in the final exam performance (R^2). Although this model is significant, this R^2 is less when compared to the straight sum of the state probabilities on the second row which R^2 implies that it explains 32.1%. One thing to note is that the major exam scores are a *summative* measure of the students' performance over several learning units whereas the probabilities as measured by end-of-chapter quizzes are the measures of success at the end of each unit.

Table 1. Comparison of Predictors in Explaining Final Exam Performance

Predictors	df_1	df_2	R^2	R^2 adj	s	F	P
Exam 1 and Exam 2	2	87	29.1	27.4	5.867	17.83	0.000
Sum p	1	88	32.1	31.4	5.707	41.65	0.000
Sum q	1	88	12.8	11.8	6.469	12.91	0.001
$p_0, p_1, p_2, p_3, p_4, p_5, p_6$	7	82	36.4	31	5.722	6.71	0.000
p's, and $q_1, q_2, q_3, q_4, q_5, q_6, q_7$	14	75	43.5	32.9	5.641	4.12	0.000
p's , q's, product p, product q, product r	17	72	45.5	32.7	5.652	3.54	0.000

When we allow the state probabilities to be least-square weighted such as they are on the fourth line of the table, the R^2 improves to 36.4%. Adding the state probabilities for *forgetting*, as shown on the fifth row, improves this explanatory power to 43.5%. If we use the state probabilities as well as their products, the variation of the final exam explained is increased to 45.5%.

Table 2 further investigates which elements of the Birth-and-Death Process model are good predictors of the comprehensive final exam scores. When the state probabilities and their products are allowed to be inputted together into the model and step-by-step chosen into the regression model, a subset of the state probabilities are significant and quite important to be included.

Table 2. Stepwise Regression Results of Predictors in Explaining Final Exam Performance

Predictors	df_1	df_2	R^2	R^2 *adj*	s	F	P
sum p, q_7	2	87	35.1	33.6	5.612	23.6	0.000
sum p, q_7, q_4	3	86	37.7	35.5	5.530	17.4	0.000
sum p, q_7, q_4 , p_0	4	85	41.1	38.3	5.409	14.8	0.000

From row 1 of Table 2 in particular, the sum of the acquisition probabilities and probability of forgetting for the 7th unit appear as important predictors of the regression model. The percentage of variation in the comprehensive final exam explained by this model is 35.1% which is higher than the benchmark R^2 of 29.1% when the two major exams were the predictors. The standard error of estimate (s) is also lower. The model is improved further if we include the probability of forgetting for the 4th unit where the R^2 is 37.7%, *adjusted* R^2 is 35.5%, and s is 5.530.

When we include further the acquisition probability for the first unit, the R^2 increases to 41.1%. In addition, we obtain the smallest standard error of estimate (s) and obtain the largest *adjusted* R^2 of 38.3% which is far superior to the model with the two major exams.

4. Conclusion and Limitations of the Study

The final regression model, as presented in the last row of Table 2, highlights the importance of the acquisition probabilities for statistics learning unit. The model also highlights the forgetting probabilities of the 7th unit which discusses the sampling distributions and the 4th unit which discusses probability. In addition, the model also puts importance in the acquisition probability of the first unit which discuss the data types, collection, and measurement.

It appears that the Birth-and-Death process model holds some promise in modeling the statistics learning process. When broken into the individual learning units, modeling the mastery of new units, retention of the current units, and forgetting the current units play some role in predicting students' success on the comprehensive course final exam.

Since this is the first research on the application of the Birth-and-Death Stochastic Process to the learning of statistics, more work on applying this model to life data is needed. The validation results obtained in this paper may be different if different data set were obtained.

The approach we took in this paper can be applicable to student's learning process in other courses as well. However, work on mathematics learning or learning in other courses similar to the statistics learning environment still needs to be done.

Acknowledgements

The first author would like to recognize the research facilities at King Fahd University of petroleum and Minerals. The second author is grateful to the University College of Zayed University for encouraging his interest in mathematics and statistics educational research.

References

Beichelt, F. (2006). *Stochastic Processes in Science, Engineering and Finance*. Boca Raton: Chapman & Hall.

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annu. Rev. Psychol.*, 55, pp. 51-86.

Globerson, S., Levin, N. and Shtub, A. (1989). The impact of breaks on forgetting when performing a repetitive task. *IIE Trans.*, 21, 376-381.

Groen , G.J. and Parkman, J.M. (1972). A Chronometric analysis of simple addition. *Psychological Review*, 79, 329-343.

Jaber, M.Y. and Bonney, M. (1997). A comparative study of learning curves with forgetting, *Appl. Math. Modelling*, 21, 523-531.

Janoos, F., Brown, G., Morocz, I., and Wells III, W. (2013). State-space analysis of working memory in schizophrenia: an FBIRN study. *Psychometrika*, 78(2), 279–307.

Lee, J. L., Everitt, B. J. & Thomas, K. L. (2004). Independent cellular processes for hippocampal memory consolidation and reconsolidation. *Science*, 304, 839–843.

Maren, S. (1999). Long-term potentiation in the amygdala: a mechanism for emotional learning and memory. *Trends Neurosci*, 22, 261-267.

Sara, S. J. (2000). Retrieval and reconsolidation: toward a neurobiology of remembering. *Learn. Mem.* 7, 73–84.

Snow, R. E., & Lohman, D. F. (1993). Implications of Cognitive Psychology for Educational Measurement. In R. L. Linn (Ed.), *Educational Measurement.* (3rd ed.). (pp. 263-331). Phoenix: American Council on Education and Oryx Press.

Wixted, J. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology*, 55, 235–269.

Wixted, J.T. and Carpenter, S. K. (2007) The Wickelgren Power Law and the Ebbinghaus Savings Function, *Psychological Science*, 18 (2), 133-134.