2

Application of the Negative Binomial/Pascal Distribution in Probability Theory to Electrochemical Processes

Thomas Z. Fahidy
Department of Chemical Engineering, University of Waterloo, Canada

1. Introduction

The subject of interest here is a set of random observations falling into only two categories: success and failure, their single-event probability being constant and independent of any previous occurrence during a sequence of such events. The probability distribution of their number, known generally as Bernouilli trials, is called the negative binomial distribution (NBD) or the Pascal distribution. In this context, success and failure are completely relative concepts, their physical meaning defined by the experimenter or analyst: the appearance of a substandard product, for instance, among acceptable ones may well be considered success by a quality controller whose objective is the identification of substandards. This distribution serves in general for finding the probability of an exact number, or at least, or at most a certain number of failures observed upon so many successes (or vice versa, depending on the chosen definition of success and failure). Its special form, related to the appearance of the *first* success (or failure) is called the geometric distribution.

Although NBD theory has widely been employed in various technical/technological areas, its utility for the analysis of electrochemical phenomena and electron-transfer processes has not yet been demonstrated to the author's knowledge. The purpose of this chapter, in consequence, is to supply such a demonstration via five specific illustrative examples as a means of stimulating further interest in probabilistic methods among electrochemical scientists and engineers.

2. Brief theory

2.1 Basic concepts and definitions

Let N be the random variable denoting the number of failures occurring in successive Bernouilli trials (Appendix A) prior to the occurrence of K successes (also random). Define $M \equiv N + K$. Then, the probability mass function (pmf) of the NBD, defined (e.g., Doherty, 1990) as

$$\text{pmf (NBD)} = P[N = n; K = k] = C(n+k-1;k-1)p^k q^n \quad 0 < p < 1;\ n = 0, 1, 2,\ldots;\ q = 1 - p \qquad (1)$$

or, alternatively [e.g., Walpole et al., 2002] as

$$\text{pmf(NBD)} = P\{M = m; K = k] = C(m-1; k-1)p^k q^{m-k} \quad 0 < p < 1;\ m = k, k+1, k+2,\ldots;\ q = 1 - p \qquad (2)$$

yields the probability that the number of independent Bernouilli trials required to achieve n number of failures prior to k number of successes is m = n + k. The cumulative mass function cmf (Weisstein, n.d.)

$$\text{cmf (NBD)} = P[N \le n'; K = k] = p^k\{C(k-1;k-1) + C(k;k-1)q + C(k+1;k-1)q^2+\ldots+C(n'+k-1;k-1)q^{n'}] = 1 - I_q(n'+1;k) \quad (3)$$

which yields the probability of achieving up to n' (but not more than n') failures prior to achieving k number of successes, is readily computable in terms of the incomplete beta function

$$I_q(n' + 1;k) = \Gamma\,(n' + k + 1)/[\Gamma(\,n' + 1)\Gamma(\,k)]\ \Psi_q\,(k;n') \quad (4)$$

where

$$\Psi_q\,(k;n') \equiv \int_0^q \ [(u^{n'}(1 - u)^{k-1}]\ du$$

requires, in general, numerical integration. Selected values of the Ψ – function are given in Table 1; the k = 5; n' = 0 entry demonstrates insensitivity to small values of success probability, since the value of $\Psi_q\,(5;0) = (1\text{-}p)^5/5$ is essentially 0.2.

Eq.(4) is particularly useful in the case of n'/small k configuration (Appendix B).

k	n'	q = 1 – p			
		0.85	0.90	0.95	0.99
1	2	0.2047	0.2430	0.2858	0.3234
	3	0.1305	0.1640	0.2036	0.2401
	4	0.0887	0.1181	0.1548	0.1902
2	2	0.0742	0.0790	0.0822	0.0833
	3	0.0418	0.0459	0.0489	0.0499
	4	0.0259	0.0259	0.0322	0.0333
3	2	0.0325	0.0330	0.0333	0.0333
	3	0.0159	0.0164	0.0166	0.0167
	4	8.82×10^{-3}	9.28×10^{-3}	9.49×10^{-3}	9.52×10^{-3}
4	2	0.0166	0.0167	0.0167	0.0167
	3	7.06×10^{-3}	7.12×10^{-3}	7.14×10^{-3}	7.14×10^{-3}
	4	3.49×10^{-3}	3.55×10^{-3}	3.57×10^{-3}	3.57×10^{-3}
5	0	0.19998	0.199998	0.1999999	0.2
	1	0.03332	0.03333	0.03333	0.2
	2	9.50×10^{-3}	9.52×10^{-3}	9.52×10^{-3}	0.2

Table 1. A short tabulation of selected values of the Ψ – function in Eq.(4) at small success single-event probabilities and small success numbers

2.2 Important parameters of the NBD

Table 2 contains parameters related to the first four statistical moments, with original notations adjusted to comply with the notation scheme in Doherty (1990) followed in this chapter. The mode (the value of the most frequently occurring random variable) does not exist when k =1 (the case of geometric distribution) and k = 0 (meaningless in NBD context).

Parameter	Expression	Reference
Mean (or expectation)	qk/p	Doherty, 1990
Variance	qk/p^2	Doherty, 1990
Skewness	$(1 + q)/\sqrt{kp}$	Evans, et al., 2000
Kurtosis	$3 + 6/k + p^2/kq$	Evans, et al., 2000
Mode	$q(k – 1)/p; k > 1$	Weisstein, n.d.

Table 2. Fundamental parameters of the negative binomial distribution (NBD)

2.3 The geometric distribution GD: a special case of NBD

Since C(n + k -1; k – 1) reduces to C(n; 0) = 1 when k = 1, the simplified form of Eq.(1) yields the pmf of the geometric distribution as pq: this is the probability of the first success occurring at the n - th Bernouilli trial, (i.e., the probability of (n-1) unsuccessful (“failed”) trials prior to the first success on the n-th trial) with mean q/p, variance q/p^2, skewness $(1 + q)/\sqrt{q}$, and kurtosis $9 + p^2/q$. Similarly, from Eq.(2), when k = 1, C(m – 1; 0) = 1 and it follows that

$$\text{pmf (GD)} = P[N = n\ ;\ K =1] = pq^{m-1} = pq^{n} \tag{5}$$

and

$$\text{cmf (GD)} = P\ [N \le n'\ ;\ K = 1] = p\ (1 + q + q^2 +\ldots+q^{n'}) = 1 – q^{n'+1} \tag{6}$$

2.4 Complementary probabilities

The complement of the cumulative mass function determines the probability that at least (n' + 1) or more failures would occur prior to the appearance of the last success, namely

$$\text{NBD: } 1 – p^k[C(k + 1; k -1) – C(k + 2; k – 1)q – C(k + 3; k – 1)q^2 - \ldots - C(n' + k + 1;q^{n'}] = I_q\ (n' + 1; k) \tag{7}$$

$$\text{GD: } \ 1 – p(1 + q + q^2 +\ldots+q^{n' + 1}) = q^{n' + 1} \tag{8}$$

2.5 Estimation of the single-event success probability from experimental observations

If p is not known a-priori, Evans et al. (2000) recommend the unbiased-method estimator

$$p^*(UB) = (k_0 - 1)/(n_0 + k_0 – 1) \tag{9}$$

and the maximum likelihood-method estimator

$$p^*(ML) = k_0 / (n_0 + k_0) \tag{10}$$

based on available experimental data. The zero subscript refers to anteriority, and the asterisk indicates that Eqs.(9) and (10) are estimators of the unknown (“true”) population parameters. When n_0 and k_0 are small, discrepancy between the two estimates can be very large. Conversely, if $k_0 >> 1$, the two estimates are essentially equal.

3. Application to selected electrochemical processes

3.1 The electrolytic reduction of acrylonitrile (ACN) to adiponitrile (ADN)

In one of the major electroorganic technologies the 82-90% selectivity of ADN production was ascribed primarily to efficient control of major byproducts: propionitrile and a $C_9H_{11}N_3$-trimer, as well as to the non-electrochemical formation of biscyanoethylether (Danly & Campbell, 1982). It is assumed here that in a hypothetical pilot - plant scale operation of a modified ADN process, batches produced within a set time period will be tested for ADN selectivity S immediately upon production. Denoting the random number of batches exhibiting $S > 85\%$ as N, and the single-event probability of $S \leq 85\%$ as p, the probability that exactly N = n number of batches will exhibit $S > 85\%$ while k number of batches will exhibit the opposite in a sequential sampling is given by Eq.(1). Similarly, the probability that not more than N = n' batches will exhibit $S > 85\%$ is given by Eq.(3). The finding of a "substandard" batch is considered to be success in this context, since the elimination of such batches would be the ultimate goal. Table 3 contains selected numbers of exact failure-probabilities, and cumulative probabilities that no more than two failures will be observed at the indicated values of k and p.

K	p	N					P[N ≤ 2]
		0	1	2	5	10	
1 (*)	0.2	0.20	0.160	0.128	0.0655	0.0215	0.488
	0.5	0.50	0.250	0.125	0.0156	0.0005	0.875
	0.8	0.80	0.160	0.032	0.0003	$1.9\text{x}10^{-8}$	0.992
2	0.2	0.24	0.032	0.077	0.0786	0.0473	0.149
	0.5	0.25	0.125	0.188	0.0469	0.0027	0.563
	0.8	0.64	0.128	0.077	0.0012	$7.2\text{x}10^{-7}$	0.845
3	0.2	0.008	0.006	0.037	0.0551	0.0137	0.045
	0.5	0.125	0.063	0.187	0.0820	0.0019	0.038
	0.8	0.512	0.102	0.123	0.0034	$2.0\text{x}10^{-7}$	0.733
5	0.2	$3.2\text{x}10^{-4}$	0.003	0.003	0.0132	0.0344	0.035
	0.5	0.031	0.156	0.117	0.1231	0.0306	0.164
	0.8	0.328	0.066	0.197	0.0132	$3.4\text{x}10^{-5}$	0.590

Table 3. Failure/success probabilities in Section 3.1 at selected values of the single-event success probability p, success number k (an exact value of the random variable K of the number of substandard batches), and failure number n (an exact value of the random variable N of batches of acceptable quality). (*): geometric distribution

3.2 A nickel-iron alloy plating process

A novel NiFe alloy plating process is envisaged to have been carried out in several sequential experiments adhering to a tightly controlled Ni^{2+}/Fe^{2+} ionic ratio in the cell electrolyte kept within a narrow experimental temperature range. Defining alloy deposits of poor quality as a success (by the same reasoning as in Section 3.1), the experiments are assumed to indicate eight failures prior to three successes. In the absence of any knowledge of single-event success probabilities, the latter can be estimated to be p* (UB) = (3-1)/(8+3-1) = 0.2 [Eq.(9)] and p* (ML) = 3/(8+3) = 0.2727 [Eq.(10)]. Table 4 contains selected values of individual and cumulative probabilities for this process.

n	$P^*_{UB}[N = n]$	$P^*_{ML}[N = n]$	$\sum_n (P^*_{UB})$	$\sum_n (P^*_{ML})$
0	0.0080	0.0203	0.0080	0.0203
1	0.0192	0.0442	0.0272	0.0645
2	0.0307	0.0644	0.0579	0.1289
3	0.0410	0.0780	0.0989	0.2069
4	0.0492	0.0851	0.1480	0.2920
5	0.0551	0.0867	0.2031	0.3787
6	0.0587	0.0840	0.2618	0.4627
7	0.0604	0.0786	0.2922	0.5413
8	0.0604	0.0714	0.3526	0.6128
9	0.0590	0.0635	0.4116	0.6727
10	0.0567	0.0554	0.4683	0.7317
Mean	12		8	
Variance	60		29.34	
Mode	8		5	

Table 4. Failure/success probabilities in Section 3.2 at unbiased and maximum-likelihood estimator values of the single-event success probability based on earlier observations $n_0 = 8$; $k_0 = 3$. N is the random number of good quality alloy specimens in the presence of k = 3 alloy specimens of poor quality.
$P^*_{UB} = C(n+2;2)(0.2)^3(0.8)^n$; $P^*_{ML} = C(n+2;2)(0.2727)^3(0.7273)^n$

3.3 An electrolytic nanotechnological-size cadmium plating process with tagged Cd^{2+} ions

In a hypothetical study of its mechanism, a cadmium plating process is assumed to proceed until a monolayer of about 100 discharged Cd^{2+} ions (ionic radius = 0.097 nm; Dean, 1985) has fully been formed on an approximately 3 $(nm)^2$ deposition area by a 300 nA current pulse of 10 ms duration (corresponding to a current density of about 100 mA/cm^2 in conventional plating technology). Ten percent of the ions in the electrolyte are tagged (e.g., radioactively) for monitoring purposes. Their arrival to the surface with respect to untagged ions may be considered a sequence of Bernouilli trials with probability mass function

$$P[N = m ; K = k] = C(n + k - 1; k - 1)0.1^k 0.9^n \quad (11)$$

and cumulative mass function

$$\begin{aligned} P[N \le n' ; K = k] &= (0.1)^k[C(k-1;k-1) + C(k;k-1)(0.9) + C(k+1;k-1)(0.9)^2 + \ldots + C(n' + k - 1;k-1)(0.9)^{n'} \\ &= 1 - [\Gamma (n' + k + 1)/\{ \Gamma(n' + 1)\, \Gamma(k)]\, \Psi_{0.9} (k;n') \quad (12) \\ &= 1 - (n' + k)!/[(n'!)(k!)]\, \Psi_{0.9} (k;n') \end{aligned}$$

in view of the fundamental relationship between the gamma function and factorials: $\Gamma (x + 1) = x!$

Selected individual and cumulative probabilities pertaining to the arrival of the first five tagged ions to the electrode surface are shown in Table 5.

N	P[N=n] [Eq.(11)]	P[N ≤ n] [Eq.(12)]
0	10^{-5}	10^{-5}
1	4.5×10^{-5}	5.5×10^{-5}
5	7.4×10^{-4}	0.0016
10	0.0035	0.0127
20	0.0129	0.0980
30	0.0196	0.2693
Mode = 36	0.0206	0.3916
40	0.0201	0.4729
Mean = 45	0.0185	0.5688
50	0.0163	0.6548
60	0.0114	0.7909
95	0.0017	0.9998
Variance	450	

Table 5. Failure/success probabilities and principal distribution parameters in Section 3.3, concerning the arrival of the first five tagged Cd^{2+} ions to the approximately 3 $(nm)^2$ deposition area. N is the random variable denoting the number of untagged ions that have arrived at the surface along with the arrival of the K = 5 tagged ions.

3.4 An aluminum anodizing process

In a typical conventional anodizing process (Pletcher & Walsh, 1990) using a 50 g/dm^3 chromium acid electrolyte, high corrosion-resistance opaque grayish white (possibly enamel) films are produced at an electrolyte temperature of 50 ^{0} C. To maintain a deposition rate of 13 – 20 µm/h the cell potential is gradually raised from zero to 30 V while the current density reaches an asymptote within the 10 – 15 mA/cm^2 range.

In a hypothetical research project with the objective of maintaining higher deposition rates, modified chromic acid concentration, and lower temperatures (i.e. lower energy input), a tolerance domain consisting of a combined range of acceptable operating variables in each bath (cell) is to be established, in order to meet new acceptable performance criteria. Considering bath – to – bath performance levels as Bernouilli trials, q is defined as the fraction of acceptable and p = 1 – q as the fraction of unacceptable quality (i.e. "success" in NBD parlance). Consequently, if the random variable N is the number of acceptable baths obtained prior to finding K unacceptable baths, the probability of finding exactly N = n acceptable baths on the k-th unacceptable bath is given by Eq.(1) rewritten as

$$P[N = n; K = k] = (n + k - 1)!/[(k - 1)!\, n!]p^k q^n \quad (13)$$

and specifically, when k = 1,

$$P[N = n; K = 1] = n!/[0!(n-0)!]pq^n = pq^n \quad (14)$$

Selected probability values are displayed in Table 6.

N	K	p/q		
		0.2/0.8	0.5/0.5	0.8/0.2
5	1 3	0.0656 0.0551	0.0156 0.0820	2.6×10^{-4} 3.4×10^{-3}
10	1 3	0.0214 0.0567	4.9×10^{-4} 0.0081	8.2×10^{-8} 3.5×10^{-6}
15	1 3	0.0070 0.0383	1.5×10^{-5} 5.2×10^{-4}	2.6×10^{-11} 2.3×10^{-9}
Mean	1 3	4 12	1 3	1/4 3/4
Variance	1 3	20 60	1/5 3/5	0.3125 0.9375
Mode	1 3	- 8	- 2	- 1/2

Table 6. Failure/success probabilities and principal distribution parameters in Section 3.4, related to the arrival of anodizing baths of unacceptable quality. N is the number of acceptable baths found at the k-th arrival.

3.5 Non-conducting oxide layer formation on Ti-MnO_2 and Ti-RuO_2 anodes

The possibility of non-conducting oxide layers forming on untreated titanium - manganese oxide and titanium –ruthenium oxide substrate has been known to be a source of failure for dimensionally stable anodes (DSA) over several decades (Smyth, 1966). In this illustration a recently developed experimental DSA is supposed to carry a prohibiting additive embedded in the conducting oxide matrix in order to reduce the presence of nonconductors. The reliability of the additive is estimated to be 98%, i.e. 2% of the DSA are believed to be susceptible to failure. In two independent tests, complying with Bernouilli trial conditions, twelve anode samples were taken randomly for each test from a large ensemble of anodes, finding four (Test 1) and five (Test 2) defective specimens. The defective ratios 4/12 = 0.333 and 5/12 = 0.417 might imply a prima facie rejection of the 2%-defectives-at-most claim by the anode manufacturer, as a questionable ("primitive") means of judgment. A careful NBD – based approach employing cumulative distributions (to account for all possible, not just the actually observed outcome) yields probabilities (via Eq.(2) and Eq.(4), respectively)

$$\text{Test 1: } P[\,N \le 8;\, K = 4] = 0.02^4\{C(3;3) + C(4;3)(0.98) + C(5;3)(0.98^2) + \ldots + C(11;3)(0.98^8)\}$$
$$= 1 - (1980)(5.05\text{x}10^{-4}) = 6.963 \times 10^{-5} \quad (15)$$

and

$$\text{Test 2: } P[\,N \le 7;\, K = 5] = 0.02^3\{C(4;4) + C(5;4)(0.98) + C(6;4)(0.98^2) + \ldots + C(11;4)(0.98^7)\}$$
$$= 1 - (3960)(2.525 \times 10^{-4}) = 2.250 \times 10^{-6} \quad (16)$$

which demonstrate, at a negligible numerical difference, an extremely low likelihood of finding up to eight and up to seven anode specimens, respectively, due to random effects. Hence, the 2% claim appears to be highly questionable.

4. Analysis and discussion

4.1 Computation of intermediate probabilities

The probability of a negative binomial variable N occurring (due to random causes) from a value $N = n_1$ to value $N = n_2$ can be expressed in the simplest form as

$$P[\, n_1 \leq N \leq n_2; K = k \,] = I_q (n_1 ; k) - I_q (n_2 + 1;k) \tag{17}$$

Under specific conditions, tabulations of the incomplete beta function (e.g. Beyer, 1968) may be employed for a quick estimation of probabilities (Appendix C).

4.2 Analysis of the contents of Sections 3.1 – 3.5

The primary role of Tables 2 – 6 resides in reaching a decision whether or not experimental observations indicate the presence of non – random effects causing the observed results. For the ADN – process in Section 3.1, the entries in Table 3 indicate that if, for instance, ten batches with S > 85% selectivity were found along with 1 – 5 batches with S ≤ 85%, this finding would suggest a rather strong promise for the new process, inasmuch as there would be at most an about 0.5% chance for such a result arising from random reasons (at least in the $1 \leq k \leq 5$; $0.2 \leq p \leq 0.8$ ranges). Conversely, if e.g. both selectivity ranges were equally probable ($p = q = 0.5$), the finding of exactly two batches with S > 85% selectivity would be at most slightly promising for the new process, since an about 12 – 19% probability exists for random causes. However, if there were, for instance, at least five S > 85% observations, the process would be judged highly promising.

In a somewhat different manner, the modified process could also be deemed to be acceptable, if the coefficient of variation CV related to the GD describing the appearance of the first S ≤ 85% batch, one of the absolute measures of dispersion, and defined as the ratio of the standard deviation to the mean of the distribution, differed from unity only within a pre-specified (small) fraction. From Table 2, $CV = 1/\sqrt{q}$ when $k = 1$, and it follows that as p becomes progressively smaller, both q and CV approximate unity, the latter from values above. Stipulation of a not more than 10% downward difference from unity (i.e. $CV \leq 1.1$) would specify $p \approx 0.2$, as the highest acceptable single – event probability of S ≤ 85% selectivity; put otherwise, the appearance of the first S ≤ 85% batch after four S > 85% batches would imply promise for the new process. If the stipulation were a more stringent 1% downward difference from unity, $CV \leq 0.1$ would prescribe $p \approx 0.02$, i.e. an only 2% single event probability of S ≤ 85% selectivity.

As shown in Section 3.2, when single-event probabilities are not available from prior sources, their numerical values are highly sensitive to the method of parameter estimation. The largest individual probability: $p^*_{ML} = 0.0867$ occurring at mode = 5 signals at most an approximately 9% maximum chance for randomness-related observation of good quality deposits along with the observation of three low – quality deposits. It follows that the novel process can be considered effective. Cumulative probabilities yield the same qualitative result. If $p^*_{UB} = 0.2$ is accepted for the single-event success (i.e. poor deposit quality), experimental observations of failures (i.e. good deposit quality) with three successes would suggest process reliability in face of an about 6% chance of random occurrence. By contrast, randomness-related probability of up to ten failures along with three successes being almost 50%, the effectiveness of the new process would appear to be rather dubious, if such occurrences were observed experimentally. If $p^*_{ML} \approx 0.27$ is accepted, the almost two-thirds probability of finding up to ten failures due to random effects essentially rules out the new process as a viable alternative to the existing one. Since these estimators come from a single experiment, the conclusions should be accepted only provisionally until further experiments will have established a wider data base.

It is instructive to consider situations with a relatively high number of experimentally observed successes. Assuming, for the sake of argument, $n_0 = 3$ and $k_0 = 10$, the single event success probabilities $p^*(UB) = 9/12 = 0.7500$ and $p^*(ML) = 10/13 = 0.7692$ are expected to deliver near – identical probability estimates. As shown in Table 7, practical discrepancy between UB – based and ML – based probabilities is, indeed, nugatory. Similar conclusions can be reached for first-time success cases with large values of N: if, e.g., K = 1 and N = 10 are set, $P[N = 10] = 7.153 \times 10^{-7}$; $P[N \le 10] = 0.99999976$ (with UB estimators), and $P[N = 10] = 3.299 \times 10^{-7}$; $P[N \le 10] = 0.999999901$ (with ML estimators).

Anterior k_0	Anterior n_0	Stipulation k	Stipulation n	Prediction P[N = n]	Prediction Ψ_q (k; n)	Prediction P[N ≤ n]
10	3	1	2	0.0468 0.0410	0.00521 0.00410	0.9843 0.9877
10	3	15	7	0.0948 0.0792	6.81×10^{-8} 4.42×10^{-8}	0.8385 0.8870
15	7	20	10	0.1029 0.1004	6.91×10^{-10} 5.77×10^{-10}	0.9988 0.9990
15	7	10	10	0.0271 0.0194	2.04×10^{-8} 1.45×10^{-8}	0.9_64 0.9_64

Table 7. Selected probabilities based on single event probability estimators in Section 3.2 . In columns 5 – 7 the first entry is UB – based, the second entry is ML – based. In column 7 "9_6" denotes six consecutive nines to indicate the extent of closeness to unity (i.e., certainty).

In the ion-tagging scenario of Section 3.3, and as shown in Table 5, the largest individual probability of untagged ion arrival, predicted by the mode (n = 36) is only about 2%, but there is a nearly 40% cumulative chance that up to 36 untagged ions are in fact at the surface on that particular occasion. At the mean value n = 45, the cumulative probability is about 57%. Given skewness $(1 + 0.9)/\sqrt{(0.9)(5)} \approx 0.90$, and kurtosis $3 = 6/5 + (0.1)^2/[(0.9)(5)] \approx 4.2$, the asymmetric distribution may be considered to be moderately skew (Bulmer, 1979a) and somewhat leptokurtic (Bulmer, 1979b), with respect to the skew-free normal (Gaussian) distribution, serving as reference, whose kurtosis is exactly 3.

Monitoring the arrival of untagged ions in Section 3.3 prior to the presence of the first tagged ion at preset values of their probability may also be an important objective of process analysis. Values obtained via Eq.(5) and Eq.(6) have been rounded to the nearest upward or downward integer in Table 8. The entries in the second column are given by the numerical form of Eq.(5) : $n = - 21.8543 – 9.949122 \ln(P)$, and in the third column by the numerical form of Eq.(6) : $n' = - 9.49122 \ln(P)$. No untagged ions can be expected to arrive prior to the first tagged ion at a probability higher than about 9%, whereas cumulative probabilities of their prior arrival increase with their number. While there is only a 20% probability that one (or no) untagged ion precedes the first tagged ion, it is almost certain for 65 untagged ions to do so. At a higher single - event probability of tagged ion arrival, the number of previously arrived untagged ions is, of course, smaller: when, e.g. p = 0.2, about twenty or thirty such ions may be expected to arrive at a 99% and 99.9% probability, respectively, in contrast with forty three and sixty five, when p = 0.1 .

Probability, P	The rounded number of untagged ions[1]	Up to the number of untagged ions[2]
0.001	44	-
0.005	28	-
0.01	22	-
0.05	7	-
0.09	1	-
0.10	-	0
0.20	-	1
0.50	-	6
0.90	-	21
0.95	-	27
0.99	-	43
0.999	-	65

Table 8. Selected numbers of untagged ions that have arrived at the electrode surface in Section 3.3 prior to the arrival of the first tagged ion at (arbitrary) preset probabilities (p = 0.1; q = 0.9). [1] Eq.(5); [2] Eq.(6)

The effect of the single-event success probability is also illustrated in Table 9 for the aluminizing baths of Section 3.4 . If this probability is low (p = 0.2) for unacceptable baths, the cumulative probabilities show that a relatively large number of acceptable baths can be produced prior to the first unacceptable bath. Conversely, only a relatively small number of acceptable baths can be expected, if this probability is high, before the arrival of the first unacceptable bath. Individual probabilities are much smaller: if, e.g. p = 0.1, it is essentially certain that (at least) up to sixty five acceptable baths would be found before the first unacceptable bath, but the chances of finding *exactly* sixty five baths is only (0.1) $(0.9^{65}) \approx 10^{-4}$.

Cumulative probability	Single event probability, p			
	0.1	0.2	0.5	0.8
0.20	1	0	-	-
0.30	2	1	-	-
0.50	6	2	0	-
0.80	14	6	1	0
0.90	21	9	2	0
0.95	27	12	3	1
0.99	43	20	6	2
0.995	49	23	7	2
0.999	65	30	9	3

Table 9. Selected "not more than" numbers of acceptable baths in Section 3.4 prior to the appearance of the first bath of unacceptable quality at a small, a medium, and a high success probability. Bath numbers are rounded up or down to the nearest integer.

Single-event probabilities pertaining to the event of acceptable (or unacceptable) baths can be readily computed by solving the nonlinear algebraic equation

$$q^n - q^{n+1} = P[q]; \quad n \text{ fixed} \tag{18}$$

The number of observed acceptable baths determines the largest probability obtained from the dP/dq = 0 condition, resulting in

$$P_{max} = n^n/(n+1)^{n+1} \text{ at } q_{max} = n/(n+1) \tag{19}$$

and is illustrated in Table 10. Vacancy in the blocks associated with n = 4, 5 and 6 indicates that the P[q] = 0.10 requirement cannot be satisfied inasmuch as the P_{max} values: 0.0819; 0.0670 and 0.0567 are significantly below 0.10. As shown in the last row of Table 10, the P[q] = 0.1 stipulation can be barely met even at n = 3, but it can be comfortably satisfied at n = 2, (and at n = 1,not shown explicitly; Eq.(18) readily yields q = 1/2 + √0.15 = 0.8873, hence p = 0.1187. Also, from Eq.(19), P_{max} = 1/4 = 0.25, hence $q_{max} = p_{max}$ = 0.5).

P[q] ↓ n→	Single event probability of success, p				
	2	3	4	5	6
0.01	0.0102	0.0103	0.0104	0.0105	0.0107
0.02	0.0209	0.0213	0.0219	0.0224	0.0229
0.03	0.0320	0.0332	0.0345	0.0360	0.0378
0.05	0.0561	0.0602	0.0656	0.0730	0.0853
0.07	0.0833	0.0942	0.1131	0.1667	-
0.10	0.1331	0.1842	-	-	-
q_{max}	0.6667	0.7500	0.8000	0.8333	0.8571
p_{max}	0.3337	0.2500	0.2000	0.1667	0.1429
P_{max}	0.1481	0.1055	0.0819	0.0670	0.0567

Table 10. Single event success probabilities at selected values of P[q] and n in Eq.(18), and the largest attainable P[q] = P_{max} via Eq.(19) in Section 3.4

A more sophisticated (and time consuming) determination of single-event probabilities from a set of experimental observations would require (nonlinear) regression techniques applied to Eq.(18) carrying measurement replicates of the number of acceptable baths that precede the first unacceptable bath.

The situation described in Section 3.5 invites several ramifications with it arising from the sequential sampling plan SSP (Blank, 1980) applied to the statistical procedure. It is instructive to examine the effect of the single-event probabilities on the decision – making process. In compliance with Eq.(4), the cumulative mass function is generalized from Eq.(4) in terms of the incomplete beta function – based method as

$$P[N \le 8; K = 4] = 1 - 1980\,(q^9/9 - 3q^{10}/10 + 3q^{11}/11 - q^{12}/12) \tag{20a}$$

and

$$P[N \le 7; K = 5] = 1 - 3960\,(q^8/8 - 4q^9/9 + 6q^{10}/10 - 4q^{11}/11 + q^{12}/12) \tag{20b}$$

upon algebraic integration, involving the fundamental identities: $(1 - x)^3 = 1 - 3x + 3x^2 - x^3$; $(1 - x)^4 = 1 - 4x + 6x^2 - 4x^3 + x^4$. The variation of these probabilities with the single event probability of failure is depicted in Table 11.

Single event probability q of finding an anode of unacceptable quality	Probability of finding up to eight acceptable anodes out of twelve (Test 1)	Probability of finding up to seven acceptable anodes out of twelve (Test 2)
0.50	0.9270	0.8061
0.80	0.2054	0.0726
0.83	0.1324	0.0393
0.84	0.1114	0.0310
0.87	0.0645	0.0148
0.875	0.0528	0.0113
0.88	0.0464	9.50×10^{-3}
0.90	0.0256	4.33×10^{-3}
0.93	7.53×10^{-3}	8.76×10^{-4}
0.95	2.24×10^{-3}	1.84×10^{-5}
0.965	5.92×10^{-4}	5.46×10^{-5}
0.98	6.96×10^{-5}	2.25×10^{-6}
0.99	4.64×10^{-6}	3.78×10^{-7}

Table 11. The cumulative probability that, in Section 3.5, finding up to eight (Test 1, K = 4), and up to seven (Test 2, K = 5) anodes of acceptable quality is due to random effects.

Using Table 11 as a guide, a process analyst setting a cumulative probability of about two percent as the acceptance threshold for the 2% - defectives claim, would be inclined to accept it when (unknown to the analyst) $q \geq 0.93$ in Test 1, and when (again unknown to the analyst) $q \geq 0.88$ in Test 2. A different (and more exacting) analyst setting the claim – acceptance threshold to about 0.05% would accept the claim when (again unknown to the analyst) $q \geq 0.965$.

The point probabilities $P[N = 8] = C(11;3)(0.02^4)(0.98^8) = 2.24 \times 10^{-5}$ (Test 1), and $P[N = 7] = C(11;4)(0.02^5)(0.98^7) = 9.17 \times 10^{-7}$ (Test 2) would favour claim rejection only somewhat more strongly at this low level of single event success probability. This bias toward rejection is much more pronounced at higher p – values, as attested by the p = 0.5 (Test 1: 0.0403; Test 2: 0.0806) and p = 0.2 (Test 1: 0.0443; Test 2: 0.0221) cases, vis – à – vis the top two entries in Table 11. Cumulative probabilities put claim rejections on a firmer ground than point probabilities by increasing their statistical reliability.

In the alternative fixed sample plan FSP approach (Blank, 1980) the size of the anode ensemble is fixed prior to testing. While it is, in principle, a matter of arbitrary choice, it should not be smaller than the mean of the failure occurrences plus the number of successes (i.e. the mean number of acceptable anodes plus the number of defective anodes). Consequently, the ensemble size should be at least $8(0.98)/0.2 + 4 = 396$ in Test 1, and $7(0.98)/0.02 + 5 = 348$ in Test 2, indicating that SSP would be less time (and material) consuming than FSP.

Another decision scheme might be based on the probability that, in face of the 2% - defectives claim the number of acceptable anodes is between two pre-specified values. If, for the sake of argument, these values are 22 and 25 inclusive, and assuming that when a sufficiently large number of anodes were tested three defectives were found, the probability calculated via Eq.(17):

$$[22 \leq N \leq 25; K = 3] = I_{0.98}(22;3) - I_{0.98}(26.3) = 0.0062 \quad (21a)$$

or, equivalently

$$P[22 \leq N \leq 25; K = 3] = (0.02^3)[C(24;2)(0.98^{22}) + C(25;2)(0.98^{23}) + C(26;2)(0.98^{24}) + C(27;2)(0.98^{25})] = 0.0062 \quad (21b)$$

would permit inference of effectiveness for the new process. Similarly, if the decision criterion is based on the incidence of the first defective anode upon the appearance of N anodes of acceptable quality, and it is stipulated that this event occur from the 21st to the 24th test inclusively, the probability

$$P[21 \leq N \leq 24; K = 1] = (0.02)(0.98^{21} + 0.98^{22} + 0.98^{23} + 0.98^{24}) = 0.98^{21} - 0.98^{25} = 0.0508 \quad (22)$$

would lead to the same conclusion. If the somewhat more stringent condition for the appearance of the first defective anode between the 14th and the 17th test inclusive were set, Eq.(22) would be rewritten as

$$P[14 \leq N \leq 17; K = 1] = (0.02)(0.98^{14} + 0.98^{15} + 0.98^{16} + 0.98^{17}) = 0.98^{14} - 0.98^{18} = 0.0585 \quad (23)$$

and a process analyst, inclined to accept the 2% - defectives claim only up to a 4% cumulative probability due to random effects would most likely question the effectiveness of the process. The level of acceptance chosen by the process analyst is an admittedly subjective element in the decision process, but not even quantitative sciences can be fully objective at all times. This statement is all the more valid for applied probability methods as probabilities are prone to be influenced by individual experiences.

5. Caveats related to negative binomial distributions

One inviting pitfall in dealing with NBD – related probability calculations would be the attempt to apply combination strings where they do *not* apply. A case in point can be the 2% - defectives claim in Section 3.5 if, in computing e.g., the probability of finding eight acceptable anodes in the presence of four defective ones, the erroneous path: C(u;8)C(v;4)/C(u + v;12) were chosen. The latter provides the probability of selecting eight items out of u identical items simultaneously with selecting four items out of v identical items, the u and v items being of a different kind, with replacement. Apart from the conceptual error in this scheme, the arbitrary choice of u and v predicts widely different numerical values(if, e.g. u = 60 and v = 20, C(60;8) C(20;4)/C(120; 12) = 0.00117 is considerably different from, e.g., C(45;8) C(6;4)/C(51;12) = 0.0204).

Equally important is the stipulation of mutual independence of the Bernouilli trials. In the context of Section 3.5, e.g., the anode – producing process should have no “memory” of the quality of any previously produced specimen. Otherwise, the NBD – based probability calculations would produce biased, i.e. statistically unreliable results. Similar considerations apply as well to the other illustrative examples. Event interdependence would necessitate working with conditional probabilities; if E_1 were the event of a success and E_2 the event of failure, $P(E_1/E_2)$ would be the probability of a success occurring when a failure has occurred. This topic is further explored in the sequel.

6. Interdependence effects: the role of conditional probabilities and Bayes' theorem

6.1 Fundamental concepts

In terms of conventional set theory, the intersection $(A \cap B)$ denotes the simultaneous existence ("coexistence") of two events, and (A/B) the existence of event A on the condition that event B has happened (is known to have happened). The probability expression $P(A \cap B) = P(A/B)P(B) = P(B/A)P(A)$ serves as the underpinning of Bayes' theorem (also known as Bayes' rule) involving conditional probabilities, written in a general form as

$$P(A_k/B) = P(B/A_k)P(A_k)/[P(B/A_1)P(A_1) + P(B/A_2)P(A_2) + \ldots + P(B/A_k)P(A_k) + \ldots + P(B/A_n)P(A_n)] \quad (24)$$

provided that the A_k, $k = 1,\ldots,n$ events are exclusive (i.e., independent of one another) and exhaustive (i.e., at least one of the events occurs). In the simpler instance of dual event sets (A, and A': not A), (B, and B': not B) it follows from Eq.(24) that

$$P(A/B) = P(B/A)P(A)/[P(B/A)P(A) + P(B/A')P(A')] \quad (25)$$

The denominator in Eq.(24) and Eq.(25) yields the overall probability of event B occurring. Independence of events A_k and B may consequently be defined as $P(B/A_k) = P(B)$ and $P(B/A) = P(B)$, etc. Similarly, the probability of intersecting events is simply the product of their individual probabilities: $P(A \cap B) = P(A)P(B)$, or equivalently, $P(B \cap A) = P(B)P(A)$, in the case of independence.

Employing similar arguments, the rest of the posterior probabilities may be written as

$$P(A'/B) = P(B/A')P(A')/[P(B/A')P(A') + P(B/A)P(A)] \quad (26)$$

$$P(A/B') = P(B'/A)P(A)/[P(B'/A)P(A) + P(B'/A')P(A')] \quad (27)$$

$$P(A'/B') = P(B'/A')P(A')/[P(B'/A')P(A') + P(B'/A)P(A)] \quad (28)$$

with the understanding that

$$P(A/B) + P(A'/B) = 1; P(A/B') + P(A'/B') = 1.$$

6.2 Application to the aluminum anodizing process in Section 3.4

In order to illustrate the Bayes' theorem – based approach, the assumption is made that the single event success probability p, i.e., the probability of an anodizing bath performing in an unacceptable manner, does *not* have a strictly defined value; in fact, it varies with a-priori experienced conditional probabilities. In this context A is defined as the event of p being lower or equal to a "believed" or preset value p^*, and the complementary event A' is that $p > p^*$. B is the event of the range of acceptable operating variables falling within a specified tolerance domain, and B' is the event of the range being outside the domain. For the sake of numerical demonstration, the prior probabilities P(A) = 0.85; P(B/A) = 0.97 and P(B/A') = 0.04 are set. They state, on the basis of some prior experience, respectively, that (i) there is an 85% chance for $p \leq p^*$; (ii) a 97% chance that if $p \leq p^*$, the range of operating variables ROV falls within the specified tolerance domain STD, and (iii) a 4% chance that if $p \leq p^*$, the range is outside the domain. The calculations assembled in Table 12 reinforce the $p \leq p^*$ hypothesis inasmuch as P(A/B) is near 100%, whereas P(A'/B) is less than 1%. The probabilities P(A/B') ≈ 0.15 and P(A'/B') ≈ 0.85 strongly imply that if the ROV were outside the STD, the $p \leq p^*$ hypothesis would be essentially untenable.

Probability statement	Bayes' theorem	Probability value, %
$p \leq p^*$ when ROV is within STD	P(A/B) = 0.97x0.85/(0.97x0.85+0.04x0.15)	99.28
$p > p^*$ when ROV is within STD	P(A'/B) = 0.04x0.15/(0.04x0.15+0.97x0.85)	0.72
$p \leq p^*$ when ROV is outside STD	P(A/B') = 0.03x0.85/(0.03x0.85+0.96x0.15)	15.04
$p > p^*$ when ROV is outside STD	P(A'/B') = 0.96x0.15/(0.96x0.15+0.03x0.85)	84.96

Table 12. Application of Bayes' theorem to the anodizing bath in Section 3.4

It is instructive to examine the sensitivity of these results to variations in P(A), P(B/A) and P(B/A') while other pertinent probabilities remain constant. The relationships

$$P(A/B) = 0.97P(A)/[0.04 + 0.93P(A)] \tag{29}$$

$$P(A/B') = 0.03P(A)/[0.96 - 0.93P(A)] \tag{30}$$

$$P(A/B) = 0.85P(B/A)/[0.006 + 0.85P(B/A)] \tag{31}$$

$$P(A/B) = 0.8245/[0.8245 + 0.15P(B/A')] \tag{32}$$

lead to a wealth of useful inferences. An increase in prior probability P(A) produces a general tendency toward unity for posterior probabilities P(A/B) and P(A/B'), although the effect on the latter is less rapid. If P(A) is close to unity, Bayes' theorem predicts little difference with respect to event B or event B'.

Considering specifically the N = 5, K = 3 scenario, and given the finding above that $P[p \leq p^*]$ is very close to unity, it is possible to find the numerical value of p^* that will satisfy the cumulative probability P^* set by the process analyst for accepting the performance of the new anodizing process, expressed as

$$P^* = P[N \leq 5; K = 3] = C(7;2)(p^*)^3[\,1 + q^* + (q^*)^2 + (q^*)^3 + (q^*)^4 + (q^*)^5\,] \tag{33}$$

or alternatively,

$$P^* = P[N \leq 5; K = 3] = 1 - I_q(5;3)\ ;\ q = q^* \tag{34}$$

Since $q^* = 1 - p^*$, and $\Gamma(9)/[\Gamma(6)\Gamma(3)] = 8!/(5!2!) = 168$, it follows from Eq.(34) that

$$P^* = 1 - 168[(q^*)^6/6 - 2(q^*)^7/7 + (q^*)^8/8] \tag{35}$$

If the process analyst sets, for example, $P^* = 1.5\%$, and if the single-event success probability p^* is about 7%, the performance of the new anodizing process can be inferred with great confidence. If five acceptable baths are found experimentally along with three unacceptable baths (the third one being the last one tested), an analyst would likely question positive claims regarding the new anodizing process.

In a somewhat more complicated situation several ranges of single-event probabilities would exist with related conditional probabilities. This case is illustrated by assuming three adjacent success probability ranges: $0.02 \le p < 0.05$; $0.05 \le p < 0.1$; $0.1 \le p < 0.2$, their events denoted as A_1; A_2; A_3; ,respectively. With prior probabilities $P(A_1) = 0.4$; $P(A_2) = 0.5$; $P(A_3) = 0.07$, they represent a firm "belief" in p falling between 0.02 and 0.1, but considerable uncertainty about where it can be expected between these bounds. Similarly, the conditional probabilities, assumed to be $P(B/A_1) = 0.3$; $P(B/A_2) = 0.4$; $P(B/A_3) = 0.15$, reflect this state of affairs. Then, in Eq.(24) the probability of the OVR falling within STD, namely $P(B) = 0.3x0.4 + 0.4x0.5 + 0.15x0.07 = 0.3305$ raises serious doubt about the properness of OVR position, and not surprisingly, the only firm conclusion that can be drawn from the posterior probabilities: $P(A_1/B) = 0.3x0.4/0.3305 = 0.3631$; $P(A_2/B) = 0.4x0.5/0.3305 = 0.6051$ and $P(A_3/B) = 0.15x0.07/0.3305 = 0.03118$ is the rather weak likelihood of p falling between 0.1 and 0.2 due to *random* effects. Let the anodizing process produce ten baths of acceptable, and three baths of unacceptable quality in a subsequent test. Table 13 indicates that inside the domain of single-event success probabilities p must be lower than 0.07 in order to judge the anodizing process acceptable if $P[N = 10;K = 3] \le 0.05$ is stipulated as a reasonable acceptance criterion for the process. Under such circumstances, acceptance would at worst carry with it an about 13% possibility of random effects accompanying the process analyst's decision.

Inherent subjectivity in conditional probabilities arising from reliance on personal experience as well as documented (subjective and/or objective) evidence had been claimed (usually by traditional statisticians) in the past as a weakness of Bayesian methods, but their advantages over traditional statistical approaches have been well recognized (e.g. Arnold, 1990; Manoukian, 1986; Utts & Heckard, 2002). The assertion that "…rational degree of belief is the only valid concept of probability…" (Bulmer, 1979c), represents a perhaps exaggerated, but thought-provoking pro-Bayesian view (Jeffreys, 1983). Section 6.2 portrays (albeit modestly) the usefulness of this segment of modern probability theory with an electrochemical flavour.

P	$P[N=10;K=3] = 66(p^*)^3q^{10}$	$P[N\le10;K=3] = 1- 858(q^{11}/11 - q^{12}/6 + q^{13}/13)$
0.020	$4.314x10^{-4}$	$1.968x10^{-3}$
0.035	$1.982x10^{-3}$	$9.419x10^{-3}$
0.050	$4.940x10^{-3}$	0.0245
0.060	$7.678x10^{-3}$	0.0392
0.070	0.0109	0.0577
0.080	0.0147	0.0799
0.090	0.0187	0.1054
0.100	0.0230	0.1339
0.200	0.0567	0.4983

Table 13. Individual and cumulative probabilities for the three-p-range scenario in Section 6.2

7. Normal distribution – based approximations to NBD with large success occurrences

Rephrasing Theorem 4 – 13 (Arnold, 1990b) concerning large success numbers K, the random variable $(N_k - \mu)/\sigma$ closely approximates the standard random normal variate Z as K increases, and in the limit,

$$(N_k - \mu)/\sigma \rightarrow N(0;1);\ K \rightarrow \infty \tag{36}$$

with mean $\mu \equiv k(1 - p)/p$ and variance $\sigma^2 \equiv k(1 - p)/p^2$. Taking into account the conventional continuity correction required when a discrete distribution is approximated by a continuous distribution, it follows that

$$P[N_k = n_k;\ K = k] \approx \Phi(z'') - \Phi(z') \tag{37}$$

where

$$z' = (n_k - ½ - \mu)/\sigma\ ;\ z'' = (n_k + ½ - \mu)/\sigma \tag{38}$$

and $\Phi(z)$ is the cumulative standard normal distribution function, tabulated extensively in the textbook literature and monographs on probability and statistics.

To illustrate the scope of the normal distribution it is supposed that in Section 3.4 a preliminary study of the new anodizing process had indicated p = 0.4. It is further stipulated that in a subsequent experimental test twenty acceptable as well as twenty unacceptable baths were found in the usual manner (i.e. the last bath was unacceptable). Given that $\mu = 20x0.6/0.4 = 30$ and $\sigma^2 = 30/0.4 = 75$, Eq.(37) and Eq.(38) yield $z' = (19.5 - 30)/\sqrt{75} = - 1.2124$ and $z'' = (20.5 - 30)/\sqrt{75} = - 1.0970$, respectively. Hence, $\Phi(z') = 0.1131$ and $\Phi(z'') = 0.1357$, resulting in $P[N_k = 20;\ K = 20] \approx 0.1357 - 0.1131 = 0.0226$, versus the rigorous value of $C(39;19)(0.4^{20})(0.60^{20}) = 0.0277$. Since $P[N_k \leq 20;\ K = 20]$ must be larger than $P[N_k = 20;\ K = 20]$, it is not necessary to compute the former, if a lower than 2% cumulative probability were judged sufficient to accept the claim of better performance by the novel process. However, for the sake of completeness, Eq.(4) is shown to corroborate the properness of this reasoning:

$$P[N \leq 20;\ K = 20] = 1 - [\Gamma(41)/\Gamma(21)\Gamma(20)] \int_0^{0.6} u^{20} (1 - u)^{19} du = 0.1298 \tag{39}$$

along with the normal approximation $\Phi(z'') = 0.1357$. These findings do not signal, of course, a better performance.

Table 14 demonstrates that even at a relatively small number of successes the normal approximation to NBD is well within the same order of magnitude, albeit not uniformly so; this is also seen in the instance of cumulative probabilities. As a case in point, from Table 14: $P[0 \leq N \leq 7] = 0.2946$ is obtained employing rigorous NBD theory, whereas the normal approximation via the sum (0.0076 + 0.0017 + 0.0178 + 0.0270 + 0.0358 + 0.0469 + 0.0615 + 0.0700) = 0.2783 differs from the rigorous value only by a relative error of about - 6%. The simpler "shortcut": $P[0 \leq N \leq 7] \approx \Phi\ [(7.5 - 10)/\sqrt{20}] = \Phi\ (- 0.56) = 0.2877$ equally qualifies as a good approximant on account of a relative error of about – 2.3%. The poorer approximation via $\Phi\ [(7 - 10)/\sqrt{20}] \approx \Phi\ (- 0.67) = 0.2514$ with a – 14.7% relative error is the price to pay if the continuity correction is neglected. The last column in Table 14 demonstrates the unevenness of the error magnitudes. Using the relative error, or the error magnitude as a measure of approximation quality is the analyst's decision.

8. Conclusions

The illustrative examples, albeit not exhaustive, demonstrate the potential of NBD theory for analyzing a wide variety of electrochemical scenarios from a probabilistic standpoint. In view of a still rather limited employment of probability – based and statistical methods in the electrochemical research literature, a major intent of the material presented here is a "whetting of appetite" by stimulating cross fertilization between two important disciplines. There remains much more work to be done in this respect.

Success probability, p	Number of failures, n	P[N = n; K = 10]		Magnitude of normal approximation error
		NBD	Standard normal	
0.1	30	8.98×10^{-4}	0.0022	0.0013
0.1	15	2.69×10^{-4}	5.2×10^{-4}	7.0×10^{-4}
0.5	0	9.80×10^{-4}	0.0076	0.0066
0.5	1	0.0049	0.0117	0.0068
0.5	2	0.0134	0.0178	0.0044
0.5	3	0.0269	0.0270	0.0001
0.5	4	0.0436	0.0358	0.0078
0.5	5	0.0611	0.0469	0.0142
0.5	6	0.0764	0.0615	0.0149
0.5	7	0.0673	0.0700	0.0027
0.5	20	0.0093	0.0119	0.0026

Table 14. Comparison of individual probabilities in Section 7 via NBD and standard normal distribution theory

9. Acknowledgment

The manuscript was written on equipment provided by the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Department of Chemical Engineering of the University of Waterloo.

10. Appendix A

The Bernouilli trial is a statistical experiment involving a binomial event with success probability p and failure probability q = 1 – p (success and failure are opposite but arbitrarily defined events). Each trial is independent of any previous trial and p and q do not change from trial to trial. The exponent x in the pmf can take only two values, namely x = 0 (failure with probability q), or x = 1 (success with probability p). This is the Bernouilli distribution with pmf

$$P[X = x] = p^x (1 - p)^{1-x}; \ x = 0, 1 \tag{A.1}$$

If, for instance, there are four anodes in a batch of Section 1 and only one among them is of acceptable quality (bad batch!), 3/4 is the probability of any of the anodes being defective, and 1/4 is the probability of any of the anodes being acceptable. Hence, Eq.(A.1) becomes $P[X=x] = (3/4)^x (1/4)^{1-x}$ with p(0) = 1/4 and p(1) = 3/4. Using the argument (Milton & Arnold, 1990) that in the particular case where x number of trials are needed to obtain three successes, each trial must end with a success and the remaining (x – 1) trials must result in exactly two successes and (x – 3) failures in some order, the probability

$$P[X = x] = C(x - 1;2)\, p^3 (1 - p)^{x-3} \quad x = 3,4,5 \tag{A.2}$$

can be generalized in a straightforward manner to obtain the pmf of the NBD

$$P[X = x] = C(x - 1;k - 1)\, p^k (1 - p)^{x-k} \tag{A.3}$$

and since x = n + k, where n is the number of failures when k successes have been observed, Eq.(1) is finally established. The C(x-1;2) combination in Eq.(A.2) represents the partitioning of (x – 1) elements into 2, and (x – 3) elements in accordance with probability theory, i.e. $C(x – 1;2) = (x – 1)!/([2!(x – 3)!] (=[x^2 -3x + 2]/2)$.

11. Appendix B

When k is sufficiently small, the $\Psi_q(k; n')$ function can be conveniently expressed in terms of polynomials carrying integer powers of its upper limit q, as shown in Table 15 (In the context of the cumulative probabilities, m = n').

Number of successes, k	$\Psi_q(k; m)$
1	$F(q) \equiv q^{m+1}/(m+1)$
2	$F(q) - q^{m+2}/(m+2)$
3	$F(q) - 2q^{m+2}/(m+2) + q^{m+3}/(m+3)$
4	$F(q) - 3q^{m+2}/(m+2) + 3q^{m+3}/(m+3) – q^{m+4}/(m+4)$
5	$F(q) – 4q^{m+2}/(m+2) + 6q^{m+3}/(m+3) - 4q^{m+4}/(m+4)] + f(5)$

Table 15. Polynomial expressions for the Ψ – function at the first five numbers of success.
$f(5) \equiv q^{m+5}/(m + 5)$

12. Appendix C

The utility of incomplete beta function tables is shown by computing the cumulative probability that up to four failures appear prior to the appearance of the eighth (and last) success when p = 0.3, i.e., $1- I_{0.7}(5;8) = I_{0..3}(8;5)$ in terms of the incomplete beta function $I_x(a,b)$ tables (Beyer, 1968). As seen in the excerpt below with entries obtained from the tables rounded to four decimals, $I_{0.3}(8;5) \approx 0.01$; numerical integration of Eq.(A.4) yields 1 – (3960) (2.501 x 10^{-4}) = 0.0095. Similarly, $I_{0.35}(8;5) \approx 0.025$, and 1 – (3960) (2.461 x 10^{-4}) = 0.0255, with x = 1 – 0.35 = 0.65 as upper limit of integration.

$I_x(8;5)$	0.10	0.05	0.025	0.01	0.005
X	0.4410	0.3910	0.3489	0.3024	0.2725

Intermediate values may be approximated by various methods of interpolation. If, for example, the upper limit of the integral is set to x = 0.67, the value $I_{0.33}(8;5) \approx 0.01 + (0.025 –$

0.01)/(0.3489 – 0.3024) = 0.019 is about 8% higher than 0.0176 obtained via numerical integration. Alternatively, the data may be correlated via a properly selected regression. In the case discussed here, inspection of data suggests the semi-linearized form

$$\ln[I_{0.33}(8;5)] = a + b(x) \tag{A.4}$$

or a fully linearized form

$$\ln[I_{0.33}(8;5)] = c + d\ln(x) \tag{A.5}$$

Conventional least - squares fitting (e.g., Neter et al, 1990) yields a = - 10.00894;b = 17.74349; c = 2.836044; d = 6.231738 with coefficients of determination of 0.991 and 0.999, respectively, and $I_{0.33}(8;5)$ estimates 0.017598 and 0.017597. Disagreement only in the sixth decimal is a fortuitous finding, inasmuch as such closeness is not guaranteed, in general. The fact, that the pre-integral coefficient - carrying factorials and the Ψ – function often involve the multiplication of very large and very small numbers, might be viewed by some analysts as an incentive for preferring interpolation or regression methods.

13. Appendix D

List of symbols

a, b	general variables or arguments
C(a, b)	binomial coefficient (or combination): a!/[b!(a-b)!]
cmf	cumulative mass function
E_i	i-th probabilistic event
GD	geometric distribution
$I_q(\ldots)$	incomplete beta function {Eq.(4)]
K	random variable, denoting the number of successes; k its numerical value
M, N	random variables, denoting the number of failures; m, n their numerical value, respectively
ML	maximum likelihood
N(0;1)	standardized random normal distribution with zero mean and unit variance
NBD	negative binomial distribution
P[…]	probability of a variable or an event
$P[E_2/E_1]$	conditional probability of event E_2 occurring upon the occurrence of event E_1
p	single - event success probability
pmf	probability mass function
q	single-event failure probability
S	selectivity
u	“dummy” integration variable
UB	unbiased
Z	standard normal (Gaussian) variate, z its numerical value
μ	mean of the normal distribution
σ^2	variance of the normal distribution
Γ(…)	gamma function of its argument
Φ	standard normal probability distribution function of variate Z
Ψ(…)	auxiliary function [Eq.(4)]

Special symbols

Subscript 0: earlier observations (earlier data)
Superscript *: set (threshold) value

14. References

Arnold, S. F. (1990a). *Mathematical Statistics*, 535 – 570, Prentice Hall, ISBN 0-13-561051-6, Englewood Cliffs, New Jersey,USA

Arnold, S. F. (1990b). loc. cit., 148 -152.

Beyer, W. H. (Ed.) (1968). *Handbook of tables for Probability and Statistics*, 251– 265, CRC Press, ISBN 0-8493-0692-2, Boca Raton, Florida, USA

Blank, L. (1980). *Statistical Procedures for Engineering, Management and Science*, 208-211, McGraw Hill, ISBN 0-07-005851-2 2, New York, USA

Bulmer, M. G. (1979a). *Principles of Statistics*, 63, Dover, ISBN 0-486-63760-3, New York, USA

Bulmer, M. G. (1979b). *Principles of Statistics*, 65, Dover, ISBN 0-486-63760-3, New York, USA

Bulmer, M. G. (1979c). *Principles of Statistics,* 173, Dover, ISBN 0-486-63760-3, New York, USA

Danly, D. E., & Campbell, C. R. (1982). Experience in the Scale-up of the Monsanto Adiponitrile Process, In: *Technique of Electroorganic Synthesis, Scale-up and Engineering Aspects*, Weinberg, N. L. & Tilak, B. V., 285, Wiley and Sons, ISBN 0-471-06359-2(r.3), New York, USA

Dean, J. A. (Ed.) (1985). *Lange's Handbook of Chemistry*, 13th edn., **3**-121, McGraw Hill, ISBN 0-07-016192-5, New York, USA

Dougherty, E. R. (1990). *Probability and Statistics for the Engineering, Computing and Physical Sciences*, 149-152, Prentice Hall, ISBN 0-13-711995-X, Englewood Cliffs, New J ersey, USA

Evans, M., Hastings, N. & Peacock, B. (2000*). Statistical Distributions*, 3rd edn. 140-144, Wiley and Sons, ISBN 0-471-37124-6(pbk.), New York, USA

Jeffreys, H. (1983). *Theory of Probability*, 3rd edn. Oxford[Oxfordshire]: Clarendon Press, ISBN/ISSN 0198531931(pbk.), UK

Manoukian, E. B. (1986). *Modern Concepts and Theorems of Mathematical Statistics*, 39 – 40, Springer-Verlag, ISBN 0-387-96186-0, New York, Berlin, Heidelberg, Tokyo; 3-540-96186-0, Berlin, Heidelberg, New York, Tokyo

Milton, J.S. & Arnold, J. C. (1990). *Introduction to Probability and Statistics*, 2nd edn., 61, McGraw Hill, ISBN 0-07-042353-9, New York, USA

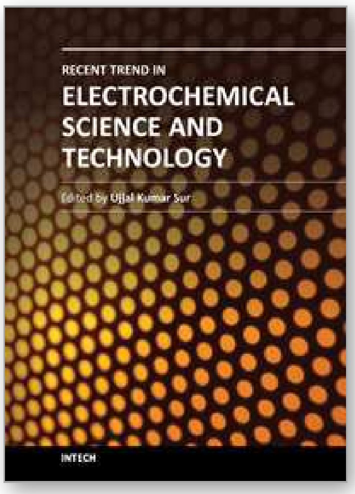
Pletcher, D. & Walsh, F. C. (1990). *Industrial Electrochemistry*, 2nd edn., 439, Chapman and Hall, ISBN 0-412-30410-4, London, UK

Smyth, D. M. (1966). The Heat Treatment of Anodic Oxide Film on Tantalum, In: *Journal of the Electrochemical Society*, Vol. 113, No. 12 (December 1966), pp. 1271-1274, ISN 0013-4651

Utts, J. M., & Heckard, R. F. (2002). *Mind on Statistics*, 319-319, Duxbury, ISBN 0-534-35935-3, Pacific Grove, California, USA

Walpole, R. E., Myers, R. H., Myers, Sh. L., & Ye, K. (2002) *Probability and Statistics for Engineers and Scientists*, 7th edn., 132-135, Prentice Hall, ISBN 0-13-041529-4, Upper Saddle River, New Jersey, USA

Weisstein, E. W. (n.d.) Negative Binomial Distribution, In: http://mathworld.wolfram.com/NegativeBinomialDistribution. *html*, Available from MathWorld-A Wolfram Web Source



Recent Trend in Electrochemical Science and Technology
Edited by Dr. Ujjal Kumar Sur

ISBN 978-953-307-830-4
Hard cover, 306 pages
Publisher InTech
Published online 27, January, 2012
Published in print edition January, 2012

This book titled "Recent Trend in Electrochemical Science and Technology" contains a selection of chapters focused on advanced methods used in the research area of electrochemical science and technologies; descriptions of electrochemical systems; processing of novel materials and mechanisms relevant for their operation. This book provides an overview on some of the recent development in electrochemical science and technology. Particular emphasis is given both to the theoretical and the experimental aspect of modern electrochemistry. Since it was impossible to cover the rich diversity of electrochemical techniques and applications in a single issue, the focus is on the recent trends and achievements related to electrochemical science and technology.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821