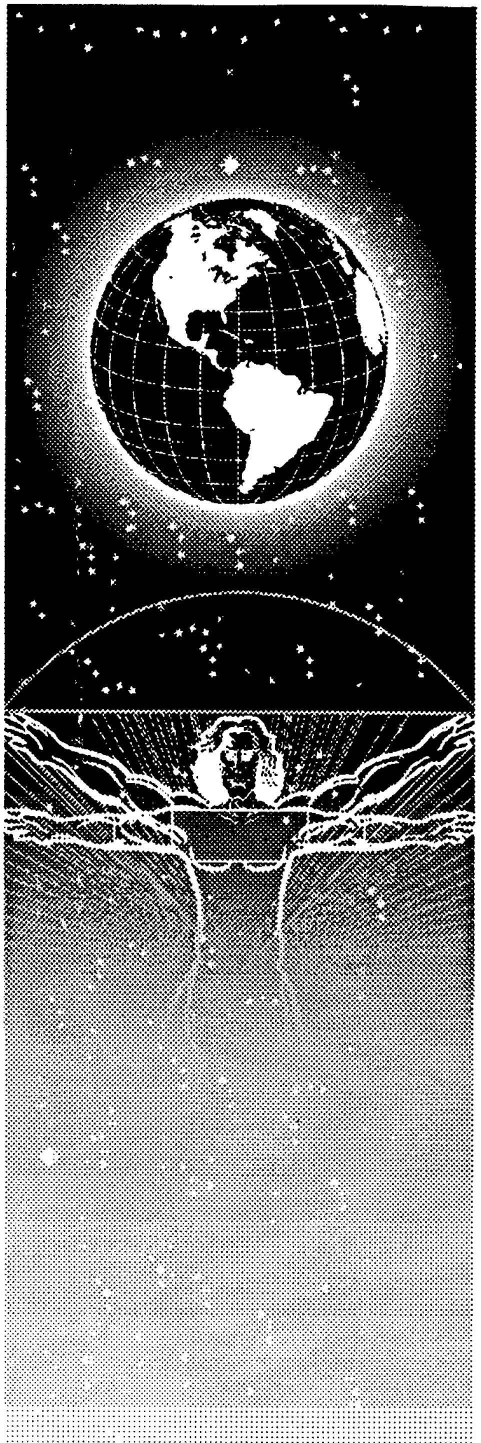
AL/HR-TP-1997-0013



UNITED STATES AIR FORCE
ARMSTRONG LABORATORY

EFFICIENCY OF CLASSIFICATION : A REVISION OF THE BROGDEN TABLE

Cheng Cheng

Department of Mathematical Sciences
Johns Hopkins University
Baltimore, MD

Melody Darby

HUMAN RESOURCES DIRECTORATE
MANPOWER AND PERSONNEL RESEARCH DIVISION
7909 Lindbergh Drive
Brooks AFB, TX 78235-5352

July 1997

Approved for public release; distribution is unlimited.

AIR FORCE MATERIEL COMMAND
ARMSTRONG LABORATORY
HUMAN RESOURCES DIRECTORATE
7909 Lindbergh Drive
Brooks Air Force Base, TX 78235-5352

NOTICES

MELODY M. DARBY
Project Scientist
Manpower & Personnel Research Division

PATRICK C. KYLLONEN
Technical Director
Manpower & Personnel Research Division

MARTIN L. FRACKER, Lt Col, USAF
Acting Chief, Manpower & Personnel Research Division

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
July 1997

3. REPORT TYPE AND DATES COVERED
Interim - June 1996 - September 1996

4. TITLE AND SUBTITLE
Efficiency of Classification: A Revision of the Brogden Table

5. FUNDING NUMBERS
PE - 62202F
PR - 1123
TA - A5
WU - 01

6. AUTHOR(S)
Cheng Cheng
Melody Darby

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Armstrong Laboratory
Human Resources Directorate
Manpower and Personnel Research Division
7909 Lindbergh Drive
Brooks, AFB, TX 78235-5352

8. PERFORMING ORGANIZATION REPORT NUMBER
AL/HR-TP-1997-0013

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
The first author was at the Armstrong Laboratory under the Air Force Office of Scientific Research's Summer Faculty Research Program in 1996
Armstrong Laboratory Technical Monitor: Ms Melody Darby (210) 536-3552

12a. DISTRIBUTION AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The open problem of establishing an accurate adjustment formula for the Brogden (1959) table used in estimating assignment benefits is solved by taking a distribution-theoretic approach in the case of equally correlated jobs. This approach establishes a more sound adjustment formula which provides more accurate results than the formula provided by Alley & Darby (1994). Implications for the use of the adjusted table in the case of correlated jobs is discussed.

14. SUBJECT TERMS
Benefit Estimation, Effectiveness, Job Performance, Selection, Classification, Efficiency, Mean Predicted Performance, Personnel

15. NUMBER OF PAGES
22

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

CONTENTS

TABLES

PREFACE

Special thanks go to Dr. William Alley, Dr. Patrick Kyllonen, Dr. Malcolm Ree, and Dr. Mary J. Skinner, from whom we received valuable inputs on this summer research project. Thanks are also due the Human Resources Directorate of the Armstrong Laboratory at Brooks AFB, TX, where an enjoyable research environment is provided.

EFFICIENCY OF CLASSIFICATION:
A REVISION OF THE BROGDEN TABLE

I. INTRODUCTION

Classification, and associated measures of effectiveness, are of especial interest to the Military Services because of the nature of large military personnel systems (Johnson & Ziedner, 1990; Campbell & Russell, 1994; Alley & Darby, 1994; and Rosse, Whetzel, & Peterson, 1994). Large numbers of applicants and available military jobs motivate the need to investigate the properties of the assignment of groups of applicants to sets of jobs. Classification efficiency as defined by Brogden (1959) refers to the average level of productivity when a group of applicants are assigned to a set of jobs using full least squares predicted performance scores as the basis for making the assignments. Brogden proposed the mean predicted performance (MPP) as the index of classification efficiency. MPP is the mean of the performance scores of a group of applicants assigned to a set of jobs when that assignment has been made with a methodology which accomplishes an optimal match of people to jobs. Brogden's optimal assignment strategy consisted of computing each applicant's predicted performance scores for each of the jobs in a set and then assigning an individual to the job for which he had the highest performance score. Brogden's work with the largest predicted performance score was based on earlier results from Tippett (1925). Tippett was concerned with the statistical issue of the sampling distribution of extreme values of scores for random samples of varying size and the implications for the properties of range statistics. In order to apply Tippett's findings to the problem of assessing classification efficiency, Brogden made the following simplifying assumptions: 1) all assignments of people to jobs are made according to the highest performance score, 2) there is an infinite supply of applicants, and 3) all jobs have equal quotas and are equally important. Under these three reasonable simplification assumptions enabling theoretical derivation, MPP values vary with the applicant rejection rate and the validity and intercorrelation of the performance scores. Brogden calculated a table (referred to as the Brogden table hereafter) showing the theoretical MPP values in the multiple job system of one to ten jobs for perfectly valid performance estimates and zero correlation of estimates across jobs. For the more general case of non-zero but equally correlated job performance scores, Brogden proposed an adjustment formula. Using Monte Carlo simulation to extend the Brogden table to 500 jobs, Alley and Darby discovered that Brogden's adjustment greatly under estimates the actual MPP in the cases of non-zero rejection rates and provided a remedy. It has been an open problem to establish a sound and more accurate adjustment formula for the Brogden table. Taking a distribution-theoretic approach, this study establishes an adjustment formula that gives more accurate results than the Alley-Darby formula. Mathematical details of the development of the new adjustment formula are relegated to the Appendix.[†]

II. METHODS

The MPP calculation problem studied by Brogden admits a natural formulation in terms of certain probability distributions. Brogden's assumptions enabling the theoretical calculation are equivalent to the following probabilistic model. Mathematical details are in the appendix.

Suppose there are m jobs. The applicant population represented by the predicted performance scores can be treated as an m-component random vector $[C_1, C_2, \cdots, C_m]$, where C_j is the performance score of a (randomly selected) applicant on the jth job.

In the basic situation of unit validity (R=1) and zero correlation between the performance scores (r=0), each C_j is assumed to follow the standard normal distribution, and all C_j's are independent. Because each applicant is assigned to the job corresponding to the highest score, the classification efficiency for the assignment is calculated from the maximum component in the vector, i.e.

[†] With numerous typographical errors, Rosse, Whetzel, and Peterson (1994) discuss the calculation of the theoretical MPP in general terms. Details of a sound development under the Brogden assumptions are given in the Appendix. A correction of the fundamental equation (4) in Rosse, Whetzel, and Peterson is available from the current authors.

$$C^* = \max(C_1, C_2, \cdots, C_m) .$$

In statistical terminology, C^* is the largest order statistic out of the m random variables. The mean predicted performance (MPP) after assignment, when the rejection rate $p=0$, is simply the expected value (or mean) of the largest order statistic C^*. By a known fact of the largest order statistic (See e.g. Larson, 1982, pp. 318-319) and the mathematical definition of mean, the theoretical MPP is given by the following integral.

$$(1) \qquad MPP_0 = \int_{-\infty}^{\infty} xm[\Phi(x)]^{m-1}\phi(x)dx ,$$

where $\phi(x) = (2\pi)^{-1/2} \exp\left(x^{-2}/2\right)$ is the standard normal probability density function (pdf), and $\Phi(x) = \int_{-\infty}^{x} \phi(t)dt$ is the standard normal cumulative distribution function (cdf). The entries on the second column of Table 1 in Brogden and Table A1 in Alley and Darby are the approximated values of the integral (1) for different numbers of jobs. When the rejection rate $p>0$, the theoretical MPP is the following integral:

$$(2) \qquad MPP_p = \frac{1}{1-p} \int_{L}^{\infty} xm[\Phi(x)]^{m-1}\phi(x)dx ,$$

where L is the $\left(p^{1/m}\right)$-th percentile (quantile) of the standard normal distribution.

In the more general situation where the validity $0 < R \leq 1$, and the performance scores $C_1, C_2, \cdots, C_m$ are equally correlated with pairwise correlation $0 < r < 1$, Brogden proposed a model in which all the m scores are related to a common variable K, plus an independent random perturbation following the normal distribution. Symbolically,

$$C_j = K + C'_j , \qquad j = 1,2,\ldots,m,$$

where K follows the normal distribution with mean zero and variance $R^2 r$, and $C'_1, C'_2, \cdots, C'_m$ are independent and identically distributed by the normal distribution with mean zero and variance $R^2(1-r)$, and are independent of K. This model does provide the desired theoretical distribution of the performance scores, but in this situation, a simple integral representation of the theoretical MPP like equation (1) or (2) no longer exists. For practice, a manageable adjustment for the table entries is desirable. The adjustment formula given by Brogden is correct only when the rejection rate is zero. Much improvement is obtained by Alley and Darby based on empirical studies using Monte Carlo simulation. With the known distributional properties of the performance scores and the properties of the quantile function (the inverse function of the cdf), an approximation formula can be established for adjusting the table values to a much better accuracy. This formula is given in the following section together with a revised Brogden table.

With present-day computing power, the integrals (1) and (2), together with the quantity L in (2), can be easily calculated with numerical routines. The integrals can be calculated to a satisfactory accuracy with a trapezoidal rule. Functions for calculating L are readily available in statistical packages such as SAS and S-plus.

III. RESULTS

Table 1 is a revised Brogden table of theoretical MPP values computed using S-plus. The revised Brogden table expresses MPP in standard score metric, mean of zero and standard deviation of one, as does the original table. Note that a new column V is added. In the revised table m denotes the number of jobs and table entries are the mean performance values for different numbers of jobs and rejection rates when the validity $R = 1$ and the between-job correlation $r = 0$. The quantity V is used for adjustment when $R \neq 1$ and/or $r \neq 0$. The adjustment formula is

$$R\sqrt{1-r}\,M_0 + \frac{R}{V}\sqrt{V^2(1-r)+r}\left(M_p - M_0\right),$$

where M_0 is the first column table entry for the given number of jobs and the rejection rate $p = 0$ (pure classification), M_p is the table entry for the given number of jobs and the rejection rate p (selection and classification). For example, for $R = 1$, $r = 0.85$, 16 jobs, and rejection rate $p = 0.6$, the adjusted mean performance is

$$\sqrt{1-0.85}(1.766)+\frac{1}{0.5432}\sqrt{(0.5432)^2(1-0.85)+0.85}(2.295-1.766)=1.604.$$

The table values, other than V, are as in the original Brogden table and the Alley and Darby extension. Inspection of the table indicates that an increase in jobs produces

Table 1
Revised Brogden Table

		Rejection Rate (percentage)									
# Jobs	V	0	10	20	30	40	50	60	70	80	90
1	1.0000	0.000	0.195	0.350	0.497	0.644	0.798	0.966	1.159	1.400	1.755
2	0.8257	0.564	0.721	0.847	0.969	1.092	1.221	1.365	1.531	1.741	2.057
3	0.7480	0.846	0.986	1.100	1.210	1.322	1.441	1.573	1.727	1.923	2.220
4	0.7012	1.029	1.159	1.265	1.368	1.473	1.586	1.711	1.858	2.045	2.330
5	0.6690	1.163	1.285	1.386	1.485	1.586	1.693	1.813	1.955	2.136	2.413
6	0.6449	1.267	1.384	1.481	1.576	1.674	1.778	1.894	2.032	2.208	2.479
7	0.6260	1.352	1.465	1.559	1.651	1.746	1.848	1.961	2.096	2.268	2.534
8	0.6107	1.424	1.533	1.625	1.715	1.807	1.907	2.018	2.149	2.319	2.580
9	0.5978	1.485	1.592	1.682	1.769	1.860	1.957	2.067	2.196	2.363	2.621
10	0.5868	1.539	1.643	1.731	1.817	1.906	2.002	2.110	2.237	2.402	2.657
11	0.5773	1.586	1.689	1.775	1.860	1.948	2.042	2.148	2.274	2.437	2.689
12	0.5690	1.629	1.730	1.815	1.899	1.985	2.078	2.183	2.307	2.469	2.718
13	0.5614	1.668	1.767	1.851	1.933	2.019	2.111	2.214	2.338	2.497	2.745
14	0.5547	1.703	1.801	1.884	1.965	2.050	2.141	2.243	2.365	2.524	2.769
15	0.5487	1.736	1.832	1.914	1.995	2.078	2.168	2.270	2.391	2.548	2.792
16	0.5432	1.766	1.861	1.942	2.022	2.105	2.194	2.295	2.415	2.570	2.813
17	0.5381	1.794	1.888	1.968	2.047	2.129	2.218	2.318	2.437	2.592	2.832
18	0.5334	1.820	1.913	1.993	2.071	2.152	2.240	2.339	2.458	2.611	2.850
19	0.5291	1.844	1.937	2.015	2.093	2.174	2.261	2.360	2.477	2.630	2.868
20	0.5251	1.867	1.959	2.037	2.114	2.194	2.281	2.379	2.495	2.647	2.884
100	0.4294	2.508	2.580	2.642	2.705	2.771	2.843	2.925	3.024	3.154	3.360
200	0.4009	2.746	2.813	2.871	2.929	2.991	3.058	3.136	3.229	3.352	3.547
300	0.3865	2.878	2.941	2.998	3.054	3.113	3.179	3.253	3.344	3.463	3.653
400	0.3772	2.968	3.030	3.085	3.140	3.198	3.261	3.334	3.423	3.540	3.727
500	0.3704	3.037	3.097	3.151	3.205	3.262	3.324	3.396	3.483	3.599	3.783

an increase in expected performance with R and r held constant at 1 and 0 respectively. As noted in Alley and Darby, R and r typically change as the number of jobs increase. This change is primarily attributable to three separate effects: (a) the increase in opportunity for alternative job assignments, (b) a potential increase in validity due to additional performance prediction composites, and (c) a potential decrease in the average among-job intercorrelation, if the additional jobs are sufficiently dissimilar from those already defined.

The performance estimates computed with the revised adjustment procedure are compared with empirical estimates of performance generated from simulation results using SAS (1985) reported in Alley and Darby. In addition to the empirical comparison, a comparison of the revised adjustment estimates with those of Brogden and Alley and Darby yields the results in Table 2. First note that for multiple jobs, $m > 0$, and the pure classification case, $p = 0$, the second term of the revised adjustment formula is zero

Table 2

Comparison of Adjustment Procedures for R=1.

Condition				Estimated M_p		
# Jobs	Rejection Rate	r	Empirical M_p	Brogden (%)	Alley-Darby (%)	Revised (%)
2	0.4	.45	1.00	.81 (81%)	.97 (97%)	.999 (100%)
2	0.4	.85	.85	.42 (49%)	.81 (95%)	.842 (99%)
2	0.6	.45	1.28	1.02 (80%)	1.27 (99%)	1.299 (101%)
2	0.6	.85	1.16	.53 (46%)	1.12 (97%)	1.165 (100%)
4	0.4	.45	1.30	1.10 (84%)	1.25 (97%)	1.301 (100%)
4	0.4	.85	1.02	.57 (56%)	.97 (95%)	1.001 (98%)
4	0.6	.45	1.58	1.27 (80%)	1.52 (96%)	1.588 (100%)
4	0.6	.85	1.32	.66 (50%)	1.26 (95%)	1.333 (101%)
8	0.4	.45	1.57	1.34 (85%)	1.51 (96%)	1.564 (100%)
8	0.4	.85	1.17	.70 (60%)	1.09 (93%)	1.149 (98%)
8	0.6	.45	1.84	1.50 (82%)	1.76 (96%)	1.843 (100%)
8	0.6	.85	1.48	.79 (53%)	1.38 (93%)	1.478 (100%)
16	0.4	.45	1.82	1.56 (86%)	1.73 (95%)	1.798 (99%)
16	0.4	.85	1.29	.82 (63%)	1.21 (93%)	1.274 (99%)
16	0.6	.45	2.06	1.71 (83%)	1.96 (95%)	2.071 (100%)
16	0.6	.85	1.59	.89 (56%)	1.49 (94%)	1.604 (101%)

Note. In the Empirical Column are estimates of the actual MPP obtained by simulation in Alley & Darby. Adjustment formulas should produce values close to these estimates

because $M_p = M_0$; the first term, $R\sqrt{1-r}\,M_0$, is simply Brogden's formula. Moreover, this expression coincides with the Alley and Darby adjustment given in different notation. All three adjustment formulas agree in the pure classification case. For multiple jobs,

$m > 0$, and the selection and classification case, $p > 0$, the appropriate adjustment is not simply $R\sqrt{1-r}\,M_p$ as proposed by Brogden. Alley and Darby suggest, for a given rejection rate, taking the difference between the m job performance estimate (selection and classification) and the single job performance estimate (pure selection), applying the Brogden adjustment to that difference, and then adding the resulting quantity to the single performance estimate. The revised adjustment formula states that the theoretically proper adjustment to the multiple job estimate is the Brogden adjustment in the pure classification case, $p = 0$, plus the properly scaled difference between the m job performance estimate (selection and classification) and the pure classification performance estimate.

Performance estimates in Table 2 for the three adjustment formulas are in the three columns under the label 'Estimated M_p'. In these three columns the numbers in parenthesis are the percentage of the empirical value accounted for by the difference between the empirical value and the revised value. Empirical estimates were calculated for combinations of the following conditions, two rejection rates, .4 and .6; four different multiple job systems, 2, 4, 8, and 10 jobs; and two values of average between job correlation, .45 and .85. Inspection of Table 2 reveals, as shown in Alley and Darby, that Brogden's adjustment underestimates mean performance by varying degrees for the multiple job system. Estimates computed with the Alley and Darby adjustment formula are much closer to, but slightly underestimate, the empirical values. The revised estimates are very close to the empirical estimates with only slight variation due to sampling error.

IV. CONCLUSION

The distributional approach to revising the Brogden table reveals some insights into the theoretical MPP calculation, and consequently a much more accurate adjustment formula is established. It is possible to extend this approach to other theoretical measurement calculations in personnel and education psychology, such as extending the Taylor-Russell tables (Taylor & Russell, 1939; Alley, Darby, & Cheng, 1996) to the situation of multiple equally correlated jobs. Work along this line constitutes further research efforts.

Caution should be exercised when applying the values in Table 1. Real world personnel systems seldom exhibit ideal characteristics such as perfect or even near perfect validities and zero or near zero between job intercorrelations. Table 2 gives an indication of the reduction in performance that can occur with non-zero between job intercorrelations. Applications of Table 1 should be made with a conservative eye toward overstating possible gains in performance.

REFERENCES

Alley, W. E., Darby, M. M., & Cheng, C. (1996, October). *The practical benefits of personnel testing: An extension of the Taylor-Russell tables to multiple job categories*. The 38th Annual Conference of the Military Testing Association, San Antonio, TX.

Alley, W. E., & Darby, M. M. (1994). Estimating the benefits of personnel selection and classification: an extension of the Brogden table. *Educational and Psychological Measurement, 55*, 938-958.

Brogden, H. E. (1959). Efficiency of classification as a function of number of jobs, per cent rejection, and the validity and intercorrelation of job performance estimates. *Educational and Psychological Measurement, 19*, 181-190.

Campbell, J. P., & Russell, T. L. (1994). *Building a joint-service classification research roadmap: methodological issues in selection and classification*. (AL/HR-TR-1994-0013). Brooks AFB, TX: Manpower and Personnel Research Division, Armstrong Laboratory, Human Resources Directorate.

Johnson, C.D., & Ziedner, J. (1990). *Classification Utility: Measuring and improving benefits in matching personnel to jobs*. (IDA Paper-2240) Alexandria, VA: Institute for Defense Analysis.

Larson, H. J. (1982) *Introduction to Probability Theory and Statistical Inference* (3rd ed.) New York: John Wiley and Sons.

Rosse, R. L., Whetzel, D. L., & Peterson, N. G. (1994*). Assessment of the classification efficiency of selection/assignment systems*. (AL/HR-CR-1996-0002). Brooks AFB, TX: Manpower and Personnel Research Division, Armstrong Laboratory,Human Resources Directorate.

SAS Institute Inc. (1985). *SAS user's guide: Basics, version 5 edition.* Cary, NC: SAS Institute Inc.

Taylor, H. C., & Russell, J. T. (1939) The relationship of validity coefficients to the practical effectiveness of tests in selection: discussion and tables. *Journal of Applied Psychology, 23*, 565-578.

Tippett, L. H. C. (1925) On the extreme individuals and the range of samples taken from a normal population. *Biometrika, 17*, 364-387.

APPENDIX

THE DISTRIBUTION THEORY OF THE BROGDEN TABLE

Preliminary

Some concepts and properties of probability distributions on the real line, especially the quantile functions, are given here for future references. Let X be a continuous random variable on the real line $(-\infty, \infty)$. The probability distribution of X is determined by its probability density function (pdf) $f(x)$, which is non-negative and $\int_{-\infty}^{\infty} f(t)dt = 1$. The cumulative distribution function (cdf) is

$$F(x) = \int_{-\infty}^{x} f(t)dt = \Pr\left(X \le x\right) = \{\text{the area under the pdf graph left to } x\}.$$

Note that the cdf $F(x)$ always has values between 0 and 1, and is non-decreasing. For the current purpose it will be technically simple to assume that $F(x)$ is strictly increasing. The *quantile function* (qf) of X, denoted by Q, is the inverse function of the cdf F. That is, for any $0 < u < 1$,

$Q(u) = F^{-1}(u) =$ {*the value* x *such that* $F(x) = u$ } = {*the value* x *such that the area under the pdf graph left to x is u*}. $Q(u)$ is called the *u*-th quantile of the random variable X. Some special quantiles are: the 1st quartile $Q(0.25)$, the median (2nd quartile) $Q(0.5)$, the 3rd quartile $Q(0.75)$, the 40th percentile $Q(0.4)$, etc.

Example A.1 Suppose that X follows the exponential distribution with pdf $f(x) = e^{-x}$, $0 < x < \infty$. Then the cdf is $F(x) = \int_{0}^{x} f(t)dt = \int_{0}^{x} e^{-t}dt = 1 - e^{-x}$. The quantile function $Q(u)$ is obtained by setting $F(x) = u$ and solving for x. We have $F(x) = 1 - e^{-x} = u$, thus $e^{-x} = 1 - u$; taking the natural logarithm on both side gives $x = -\log(1-u)$. Hence the qf $Q(u) = -\log(1-u), 0 < u < 1$.

Example A.2 Suppose that X follows the standard normal distribution $N(0,1)$ with mean 0 and variance 1, and pdf $\phi(x) = \frac{1}{\sqrt{2\pi}} \exp.\left(-x^{-2}\big/2\right)$. Then the cdf is $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp.\left(-t^{2}\big/2\right)dt$.

The qf is

$Q(u) = \Phi^{-1}(u)$. The cdf (and consequently the qf) does not have closed-form expression in this case. They can be calculated numerically in many statistical software packages. In SAS, the cdf Φ is the PROBNORM function, and the qf Φ^{-1} is the PROBIT function.

The quantile function has many useful properties, some of which are given below.

Proposition A.1 *The cdf and the qf are related by* $Q(F(x)) = x$ *for any* x *inside the range of the random variable* X, *and* $F(Q(u)) = u$ *for any* $0 < u < 1$.

Proposition A.2 *As u approaches to 0, the qf $Q(u)$ approaches to the smallest possible value of the random variable X; as u approaches to 1, $Q(u)$ approaches to the largest possible value of X.*

These two properties follow directly from the definitional fact that the qf Q is the inverse function of the cdf F.

Proposition A.3 *The expected value of X, $\mathrm{E}X$, is equal to the integral* $\mathrm{E}X = \int_0^1 Q(u)du.$

Proof: By definition of expected value [Larson 1982, p.118, Definition 3.3.1 (b)],

$$\mathrm{E}X = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{\infty} xdF(x).$$

Now change variable by setting $F(x) = u$. Then by Proposition A.1, $x = Q(u)$, and by Proposition A.2, the range of u is the unit interval (0, 1). Hence the integral becomes

$$\mathrm{E}X = \int_{-\infty}^{\infty} xdF(x) = \int_0^1 Q(u)du,$$

completing the proof.

Example A.3 Consider the exponential random variable in Example A.1. The expected value is

$\mathrm{E}X = \int_0^{\infty} xf(x)dx = \int_0^{\infty} xe^{-x}dx = 1.$ Alternatively, $\mathrm{E}X = \int_0^1 Q(u)du = \int_0^1 -\log(1-u)du = 1.$

Proposition A.4 *Let $0 < p < 1$. The conditional expected value of X given that $X \geq Q(p)$ (i.e. given that X is no less than its p-th quantile), is given by* $\frac{1}{1-p}\int_p^1 Q(u)du.$

The proof is similar to that of Proposition A.3; it is omitted to keep the technicality at a reasonable level. Note that the crucial differences here from the unconditional case in Proposition A.3 are the $1/(1-p)$ factor and the lower limit of integral being p, not 0.

Example A.4 Consider again the exponential random variable in Example A.1. Let $p = 0.4$. The 40th percentile is then $Q(0.4) = -\log(1-0.4) = -\log(0.6)$. The conditional expected value of X given that X is greater than or equal to its 40th percentile is then

$$\frac{1}{1-p}\int_p^1 Q(u)du = \frac{1}{1-0.4}\int_{0.4}^1 -\log(1-u)du \approx 1.511.$$

For the standard normal variable in Example A.1, the conditional expected value is

$\frac{1}{1-0.4}\int_{0.4}^1 \Phi^{-1}(u)du \approx 0.644.$

Proposition A.5 *Let X be a random variable with quantile function* $Q_X(u)$*, and let* $Y = A + BX$*, with* A *and* B *real numbers. Then the quantile function of Y is* $Q_Y(u) = A + BQ_X(u)$.

The Brogden Table

Brogden's (1959) assumptions are equivalent to treating the applicants' predicted performance scores as certain random variables and the applicant population being represented by the joint probability distribution of the performance scores. In tune with Brogden's notation, let *m* be the number of jobs, and let C_1, C_2, Λ, C_m denote the predicted performance scores. It is assumed that the *m* scores are standardized to have a common mean 0 and a common variance R^2, which is the validity. It is further assumed that the performance scores C_1, C_2, Λ, C_m jointly follow an *m*-variate normal distribution.

Since every applicant is assigned to the job corresponding to the highest score, the allocation efficiency is simply the largest of C_1, C_2, Λ, C_m : $C^* = \max(C_1, C_2, \Lambda, C_m)$, i.e. the largest order statistic of the *m* random variables.

Brodgen (1959) considered calculating the mean allocation efficiency in 4 cases: (1) uncorrelated performance scores and zero rejection rate, (2) uncorrelated performance scores and non-zero rejection rate, (3) equally correlated performance scores (i.e. all (C_i, C_j) pairs have common correlation $0 < r < 1$) and zero rejection rate, and (4) equally correlated performance scores and non-zero rejection rate. Each of cases is discussed below.

Case 1: uncorrelated performance scores and zero rejection rate ($r = 0$, $p = 0$). In this case the scores C_1, C_2, Λ, C_m are independent and identically distributed (i.i.d.) as normal with mean 0 and variance $R^2 - N(0, R^2)$. For convenience first let $R = 1$. Then the C_j 's have the common standard normal probability density function (pdf) $\phi(x) = (2\pi)^{-1/2} \exp\left(x^{-2}/2\right)$ and the cumulative distribution function (cdf) $\Phi(x) = \int_{-\infty}^{x} \phi(t)dt$. The mean allocation efficiency (or MPP) is nothing but the expected value of the maximum score $C^* = \max(C_1, C_2, \Lambda, C_m)$. Because C^* is the largest order statistic out of the *m* scores, by a known fact of the largest order statistic (Larson 1982, pp. 318-319) and the normal distribution assumption of the performance scores, C^* has cdf

$$F(x) = [\Phi(x)]^m, \qquad -\infty < x < \infty, \tag{3}$$

and pdf

(4) $$f(x) = \frac{d}{dx}F(x) = m[\Phi(x)]^{m-1}\phi(x), \qquad -\infty < x < \infty,$$

Application of the definition of expected value [Larson 1982, p.118, Definition 3.3.1 (b)] yields

(5) $$MPP_0 = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{\infty} xm[\Phi(x)]^{m-1}\phi(x)dx \ ,$$

which is exactly equation (1). Expected values and standard deviation of C^* for several values of m are calculated in Table I, Tippett (1925).

Alternatively, the quantile function (qf) $Q(u)$ of C^* can be obtained by setting the cdf (3) to u and solving for x: $F(x) = [\Phi(x)]^m = u$, so $\Phi(x) = u^{1/m}$, and $x = \Phi^{-1}(u^{1/m})$. Hence the qf of C^* is

(5a) $$Q(u) = \Phi^{-1}(u^{1/m}), \qquad 0 < u < 1.$$

Then by Proposition A.3, the alternative expression of the expected value of C^* (the MPP) is

(5b) $$MPP_0 = \int_0^1 \Phi^{-1}(u^{1/m})du.$$

If the validity $R \neq 1$, the MPP is simply $R(MPP_0)$ because C^* is just scaled by R.

Case 2: uncorrelated performance scores and non-zero rejection rate (r = 0, p > 0). The distributions of the performance scores and the maximum score C^* remain the same as in Case 1. But if the rejection rate p>0, i.e. only the upper $(1-p)$ portion of the applicant population is selected, then the allocation efficiency is the conditional expected value of C^* given that C^* is greater than or equal to its pth percentile (quantile). It is enough to only recall at this point that under the given distributional assumptions, the pth quantile of C^* is the number L such that the probability of $C^* \leq L$ is p, i.e. $F(L) = p$ with F the cdf of C^* given in equation (3) (i.e. the cumulative area under the pdf graph left to L is p.) Equivalently, the probability of $C^* \geq L$ is $(1-p)$ (i.e. the cumulative area under the pdf graph right to L is $1-p$.) For convenience, again first let $R = 1$. Following the definition of a conditional probability law, (Larson, 1982, pp. 61-65, 241-243), the conditional cdf of C^* given $C^* \geq L$ is (restricting $x \geq L$)

$$F(x|p) = \frac{\Pr(C^* \leq x \; and \; C^* \geq L)}{\Pr(C^* \geq L)} = \frac{\Pr(L \leq C^* \leq x)}{\Pr(C^* \geq L)} = \frac{F(x) - F(L)}{1-p} = \frac{[\Phi(x)]^m - p}{1-p}, \quad L \leq x < \infty,$$

and $F(x|p) = 0$ if $x < L$. The conditional pdf of C^* given it is greater than or equal to its pth percentile is then

(6) $$f(x|p)=\frac{d}{dx}F(x|p)=\frac{1}{1-p}m\left[\Phi(x)\right]^{m-1}\phi(x),\quad L\le x<\infty.$$

The MPP corresponding to rejection rate p is the conditional expected value

(7a) $$MPP_p=\int_L^\infty xf(x|p)dx=(1-p)^{-1}\int_L^\infty xm\left[\Phi(x)\right]^{m-1}\phi(x)dx,$$

which is exactly equation (2). Note that in the integral the <u>conditional</u> pdf $f(x|p)$, <u>not</u> the unconditional one in (4), is used.

By Proposition A.4 and the qf expression of C^* in (5a), the alternative expression of the conditional expected value (the MPP) is

(7b) $$MPP_p=\frac{1}{1-p}\int_p^1 Q(u)du=\frac{1}{1-p}\int_p^1\Phi^{-1}\left(u^{1/m}\right)du.$$

If $R\ne 1$ then the MPP is simply $R(MPP_p)$, for the same reason as before.

The same as establishing (5a), because $F(L)=p$, L can be obtained by setting the cdf (3) to p and solving for x, giving $L=x=\Phi^{-1}(p^{1/m})$, with Φ^{-1} the inverse function of the standard normal cdf (the PROBIT function in SAS). It's interesting to note that the pth quantile of C^*, L, is the $(p^{1/m})$-th quantile of the normal distribution.

Case 3: equally correlated performance scores (i.e. all $\left(C_i,C_j\right)$ pairs have common correlation $0<r<1$) and zero rejection rate. For this more general case Brogden proposed the model that the job performance scores are related to a common latent variable K, plus independent normal random perturbation: $C_j=K+C'_j$, $j=1,2,\ldots,m$. In order that all C_j's follow the $N\left(0,R^2\right)$ distribution and are equally correlated, it is assumed that K follows the $N\left(0,R^2r\right)$, and C'_1,C'_2,Λ,C'_m are i.i.d. $N\left(0,R^2(1-r)\right)$ and are independent of K. This model then provide the desired theoretical distribution of the performance scores. Now the allocation efficiency

(8) $$C^*=\max\left(C_1,C_2,\Lambda,C_m\right)=K+\max\left(C'_1,C'_2,\Lambda,C'_m\right)=K+C^{*\prime},$$

where $C^{*\prime}=\max\left(C'_1,C'_2,\Lambda,C'_m\right)$. Recall that, as in Case 1, when the rejection rate $p=0$, the MPP is the expected value (or mean) of C^*. Because the mean of K is zero and mean is additive, (8) gives that the mean of C^* (the MPP) is simply the mean of K plus the mean of $C^{*\prime}$, which is equal to the expected value (or mean) of $C^{*\prime}$. Furthermore, because of the distributional assumption on the C_j's,

$C^{*\prime}$ differs from the C^* in Case 1 only by a scaling factor $R\sqrt{1-r}$. Therefore the MPP in this case is $R\sqrt{1-r}\,MPP_0$, the Brogden adjustment.

Case 4: equally correlated performance scores and non-zero rejection rate. Brogden attempts to completely reduce the correlated job score cases to the above independent job score cases by observing that the common variable K has zero mean, so that C^* has the same expected value as $C^{*\prime}$. However, this reduction is correct only when the rejection rate p=0. The reason Brogden's adjustment fails in the p>0 case lies in the fact that in the non-zero rejection cases the MPP is the conditional expected value of C^* conditioned on its being greater than or equal to its pth percentile L, and unfortunately for correlated job scores, L is no longer simply $\Phi^{-1}(p^{1/m})$.

The exact probability distribution of C^* in this case is rather complex. Consequently, simple integral expressions for the theoretical MPP like those in equations (1) and (2) no longer exist. In order to establish an adjustment formula in the same spirit as those of Brogden and Alley & Darby (1994), a distributional approximation method is employed. Equation (8) shows that C^* is $C^{*\prime}$ plus an independent normal random variable K with mean zero and standard deviation $R\sqrt{r}<1$. And as mentioned before, $C^{*\prime}$ differs from the C^* in Case 1 only by a scaling factor $R\sqrt{1-r}$, so its exact probability distribution can be readily established. The basic idea then, is to approximate the probability distribution of C^* by properly adjusting the distribution of $C^{*\prime}$, so that the expected value and the variance of the adjusted probability distribution matches with those of C^*. The approximation is accomplished by the following steps.

(*i*) Mean and variance of $C^{*\prime}$: Note that $C^{*\prime}$ differs from the C^* in Case 1 only by a scaling factor $R\sqrt{1-r}$, more precisely, let $C_1{}^*$ be the C^* in Case 1, then $C^{*\prime} = (R\sqrt{1-r})C_1{}^*$. So The expected value of $C^{*\prime}$ is $\mu' = R\sqrt{1-r}\,MPP_0$ (as in Case 3), and The variance of $C^{*\prime}$ is $\sigma'^2 = R^2(1-r)V^2$, where

$$V^2 = \int_{-\infty}^{\infty} (x - MPP_0)^2\, m[\Phi(x)]^{m-1}\phi(x)dx,$$

and MPP_0 is given in equation (5) or (5b). The values of V are given in the first column of the revised Brogden Table in Section III.

(*ii*) Mean and variance of C^*: From (8), by additivity of expected values, the expected value of C^* is

$$(9)\qquad \mu = \mathrm{E}C^* = \mathrm{E}K + \mathrm{E}C^{*\prime} = \mathrm{E}C^{*\prime} = \mu' = R\sqrt{1-r}\,MPP_0$$

because the expected value of K is 0. By independence of K and $C^{*\prime}$, the variance of C^* is

$$\text{(10)} \qquad \sigma^2 = \operatorname{Var} K + \operatorname{Var} C^{*\prime} = R^2 r + R^2(1-r)V^2 .$$

(*iii*) Standardize $C^{*\prime}$: The random variable $\overline{C}' = \frac{1}{\sigma'}(C^{*\prime} - \mu')$ has mean 0 and variance 1. So the random variable $\overline{C}^* = \mu + \sigma \overline{C}'$ has mean μ and variance σ^2, the same as those of C^*. Hence C^* has the probability distribution approximately the same that of $\overline{C}^*$, in the sense of matching mean and variance. Approximate the distribution of C^* by the distribution of $\overline{C}^*$. We accomplish this by approximating the quantile function of C^* with the quantile function of $\overline{C}^*$.

(*iv*) Quantile function of $\overline{C}^*$: From step (*i*), $C^{*\prime} = (R\sqrt{1-r})C_1{}^*$ and from step (*iii*),

$$\text{(11)} \qquad \begin{aligned} \overline{C}^* &= \mu + \sigma \overline{C}' = \mu + \sigma\left[(C^{*\prime} - \mu')/\sigma'\right] = \mu + \sigma\left[(R\sqrt{1-r}C_1{}^* - \mu')/\sigma'\right] \\ &= \left(\mu - \frac{\sigma}{\sigma'}\mu'\right) + \left(\frac{\sigma}{\sigma'}R\sqrt{1-r}\right)C_1{}^* = A + BC_1{}^*, \end{aligned}$$

where $A = \mu - \sigma\mu'/\sigma'$ and $B = \sigma R\sqrt{1-r}/\sigma'$. By Proposition A.5, the quantile function of $\overline{C}^*$ is $\overline{Q}(u) = A + BQ_1(u)$, where $Q_1(u)$ is the quantile function of $C_1{}^*$. Recall form step (*i*) that $C_1{}^*$ is the C^* in Case 1, whose quantile function is given in equation (5a) -- $Q_1(u) = \Phi^{-1}(u^{1/m})$. Therefor substituting the values of $A, B, Q_1(u)$ gives the quantile function of $\overline{C}^*$ as

$$\overline{Q}(u) = A + BQ_1(u) = \left(\mu - \frac{\sigma}{\sigma'}\mu'\right) + \left(\frac{\sigma}{\sigma'}R\sqrt{1-r}\right)\Phi^{-1}(u^{1/m}), \quad 0 < u < 1.$$

(*v*) Approximate the quantile function $Q(u)$ of C^* by that of $\overline{C}^*$:

$$\text{(12)} \qquad Q(u) \approx \left(\mu - \frac{\sigma}{\sigma'}\mu'\right) + \left(\frac{\sigma}{\sigma'}R\sqrt{1-r}\right)\Phi^{-1}(u^{1/m}), \quad 0 < u < 1.$$

(*vi*) The adjustment formula: The MPP is the conditional expected value of C^* given it is greater than or equal to its pth percentile (i.e. given $C^* \geq Q(p)$). So by Proposition A.4,

$$MPP = \frac{1}{1-p}\int_p^1 Q(u)du.$$

Replacing $Q(u)$ by its approximation given in equation (12). we have

$$MPP = \frac{1}{1-p}\int_p^1 Q(u)du \approx \frac{1}{1-p}\int_p^1 \left[\left(\mu - \frac{\sigma}{\sigma'}\mu'\right) + \left(\frac{\sigma}{\sigma'}R\sqrt{1-r}\right)\Phi^{-1}\left(u^{1/m}\right)\right]du$$

$$= \left(\mu - \frac{\sigma}{\sigma'}\mu'\right) + \left(\frac{\sigma}{\sigma'}R\sqrt{1-r}\right)\left[\frac{1}{1-p}\int_p^1 \Phi^{-1}\left(u^{1/m}\right)du\right].$$

By equation (7b), the integral in the brackets is the MPP_p in Case 2, so

$$MPP \approx \left(\mu - \frac{\sigma}{\sigma'}\mu'\right) + \left(\frac{\sigma}{\sigma'}R\sqrt{1-r}\right)MPP_p \ .$$

Plugging in the values of $\mu, \sigma, \mu', \sigma'$ obtained in steps (*i*) and (*ii*) gives

$$MPP \approx \left(R\sqrt{1-r}MPP_0 - \frac{\sqrt{R^2 r + R^2(1-r)V^2}}{\sqrt{R^2(1-r)V^2}}R\sqrt{1-r}MPP_0\right) + \left(\frac{\sqrt{R^2 r + R^2(1-r)V^2}}{\sqrt{R^2(1-r)V^2}}\right)R\sqrt{1-r}MPP_p$$

$$= R\sqrt{1-r}MPP_0 - \frac{R\sqrt{r+(1-r)V^2}}{R\sqrt{1-r}V}R\sqrt{1-r}MPP_0 + \frac{R\sqrt{r+(1-r)V^2}}{R\sqrt{1-r}V}R\sqrt{1-r}MPP_p$$

$$= R\sqrt{1-r}MPP_0 - \frac{R\sqrt{r+(1-r)V^2}}{V}MPP_0 + \frac{R\sqrt{r+(1-r)V^2}}{V}MPP_p$$

$$= R\sqrt{1-r}MPP_0 + \frac{R}{V}\sqrt{r+(1-r)V^2}\left(MPP_p - MPP_0\right).$$

The last expression is exactly the adjustment formula given in Section III.