NAMRL-1405

MONITORING OPERATIONAL SELECTION SYSTEMS THROUGH FREQUENCY COUNTS: AN APPLICATION OF BAYESIAN PREDICTIVE INFERENCE

D. J. Blower
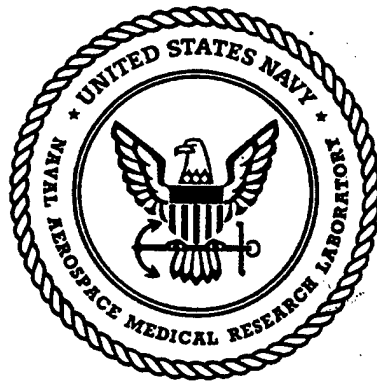
DTIC QUALITY INSPECTED 4

Reviewed and approved 2/4/99

C. H. Armstrong

C. G. ARMSTRONG, CAPT, MSC USN
Commanding Officer

NAVAL AEROSPACE MEDICAL RESEARCH LABORATORY

51 HOVEY ROAD, PENSACOLA, FL 32508-1046

NAMRL-1405

MONITORING OPERATIONAL SELECTION SYSTEMS THROUGH FREQUENCY COUNTS: AN APPLICATION OF BAYESIAN PREDICTIVE INFERENCE

D. J. Blower

ABSTRACT

One of the reasons military selection systems are put into place is to reduce high attrition rates in training. Typically though, after the selection system has experienced some operational use, the rejection and attrition rates are observed to fluctuate over time. The question then naturally arises, "Do these fluctuations represent some substantial change in the real world that should be investigated, or do they merely reflect the statistical vagaries seen in any small sample size?"

In this technical memorandum, the question just posed is addressed quantitatively by standard Bayesian statistical techniques. A predictive inference can be made about future frequency counts based on the empirical data of the past frequency counts. The uncertainty surrounding the theoretical rates mentioned above is handled in the Bayesian approach by an averaging procedure. The likelihood for any new frequencies conditioned on a given rate parameter is averaged over the probability for an acceptable range of these rate parameters. This probability for the rate parameters is in turn conditioned on the past frequency data and theoretical knowledge about the rate parameters.

This kind of analysis is helpful whenever there is concern that something fundamental might have caused the rejection and/or attrition rates of the selection system to change. For example, the rejection rates of a given selection system are suspected of having changed dramatically over the past few years. Before we can attempt to track down the cause of this alleged rate increase, we must first establish that the increase in frequency counts is not simply due to statistical fluctuations inherent in small sample sizes. It is quite easy to be misled into thinking that a "trend" based on relatively small numbers portends a significant change in the underlying rate parameter.

The techniques detailed in this report will help researchers disentangle sample size fluctuations from external perturbations to the rate parameter. In the case of any selection system, these techniques can be employed to determine whether there is a justification for investigating such fundamental changes as a shift in the hardware configuration, a change in the ability levels of the candidate population, changes to the training regime, or changes in the validity of the current prediction algorithm.

Acknowledgments

INTRODUCTION

Military training is a very expensive proposition. Therefore, a concern for keeping costs down motivates attempts to reduce the failures in training due to poor performance. In military jargon, these training failures are called *attritions*. One way to reduce the attrition rate and lower costs is through selection testing. The hope is that a relatively inexpensive test battery can identify relevant traits that underly performance in training and thus weed out some of the personnel who would ultimately fail in training.

The trade-off in using selection tests to reduce the attrition rate is that some candidates are rejected by the test who otherwise would have succeeded in training. The question of whether it is economically worthwhile to implement such a selection test can be answered by statistical decision theory. For a cost-benefit example concerned with the training of Navy pilots, see Blower [1].

In selection research, we typically estimate rejection rates and attrition rates for some targeted population for which the test battery was designed. These rates are derived from some statistical model. The particular statistical model examined here happens to be discriminant analysis. A primary output from discriminant analysis is a set of two normal curves for some linear combination of test scores. Henceforth, such a linear combination of scores from the test battery will be called a *composite score*. One curve is constructed for those who eventually fail training with a second curve for those who eventually pass training. Once a threshold score is established, areas of these normal curves can be used to estimate the rejection rate and the attrition rate. The details of how these numbers are derived are presented in the next section.

It must be emphasized that these rates are *inherently unobservable* parameters. Once a particular statistical model has been chosen, it is not a particular problem to calculate the one number that is the estimate of the rejection rate or the attrition rate. We shall show exactly how this can be accomplished in the next section. But rejection rates cannot be measured; only the actual frequency of rejections can be collected as data.

There is naturally some uncertainty attached to such rate parameters because of the more fundamental uncertainties that underly their derivation. Uncertainties might conceivably arise due to any of the following conditions that are assumed to impact the mean composite scores of the two groups. The first is statistical in nature. The mean composite score for the two groups is based on a small sample size. It would not be surprising to find that the estimate of the mean composite score for either or both of the two groups changed with a larger sample. The other conditions are changes external to the selection system. The means of the composite scores could be drifting upwards or downwards over time due to such causes. For example, the hardware components of the test battery may deteriorate with lack of maintenance. New A/D cards, monitors, and software operating systems introduced to upgrade the system may have an impact on the composite scores. The skill level of the applicants themselves could be shifting over time as the population from which the applicants are drawn changes. Uncertainty may exist concerning changes in the training curriculum that may make it easier or harder for an applicant with a given skill level to pass training.

All of these influences, and others not yet thought of, could conceivably, acting separately or in concert, affect the theoretical rejection and attrition rates. It is our lack of information about these sources that motivates us to attach some measure of uncertainty to the one number we calculated above. The only way to optimally process information in the presence of uncertainty is through probability theory, and the best embodiment of probability theory is the Bayesian approach.

In stark contrast to these myriad uncertainties listed above, there is no uncertainty in the actual frequency counts (assuming we have been attentive enough to carefully record the data). These frequency counts consist of the actual number of those rejected by the test battery and the number failing in training. The question then becomes whether the currently believed theoretical rates are in consonance with the empirical frequency counts.

In this technical memorandum, the question just posed is addressed quantitatively by standard Bayesian statistical techniques. A predictive inference can be made about future frequency counts based on the empirical

data of the past frequency counts. The uncertainty surrounding the theoretical rates mentioned above are handled in the Bayesian approach by an averaging procedure. The likelihood for any new frequencies conditioned on a given rate parameter is averaged over the probability for an acceptable range of these rate parameters. This probability for the rate parameters is in turn conditioned on the past frequency data and theoretical knowledge about the rate parameters. These statements are made precise in upcoming sections of the report.

Those who are in charge of training programs are understandably concerned about managing the flow of personnel. Fluctuations in the number of rejections and attrition from a selection system often raise a red flag of concern. The question then naturally arises, "Do these fluctuations represent some substantial change in the real world that should be investigated, or do they merely reflect the statistical vagaries seen in any small sample size?"

The kind of analysis outlined here is helpful whenever there is concern that something fundamental might have caused the rejection rates and/or attrition rates of the selection system to change. Before we can attempt to track down the cause of this alleged rate increase, we must first establish that the increase or decrease in frequency counts cannot be explained by statistical fluctuations inherent in small sample sizes. It is quite easy to be misled into thinking that a "trend" based on relatively small numbers portends a significant change in the underlying rate parameter. The quantitative ideas developed here are applied to an actual military selection system; the Landing Craft Air Cushion Vehicle (LCAC) Selection System for Operators and Engineers. The results of this analysis are presented in a companion paper [2].

More generally, one would like to monitor the rejection rates and attrition rates from a given selection system to detect if any changes have occurred that should prompt a review of the selection system. This is the same philosophy behind industrial quality control procedures that seek to identify the causes of glitches in a manufacturing process.

**Initial information about the rate parameters**

This section provides an illustration of how information about the theoretical rate parameters can be extracted from a given statistical model. The two rate parameters we are concerned with here are the rejection rate and the attrition rate that follow from operational use of the selection system. Please refer to Fig. 1 during the ensuing discussion.

At the top of the figure two normal curves are drawn. The curve to the left represents the distribution of test battery scores for those candidates who fail training, and the second curve to its right represents the distribution of test battery scores for those candidates who pass training. The x-axis is actually a composite score based on a linear weighting of some number of individual tests from the overall test battery. The y-axis shows the usual probability density function values for a normal distribution with the stated mean and standard deviation.

The means and standard deviations (σ) for the composite score are shown for the two groups called FAIL and PASS. These means, because they are computed from a small sample size, may differ from the true value of the mean and this is indicated by placing a "hat" above μ. The discriminant analysis procedure tries to construct a composite score such that the means of the two groups are as far apart as possible. As part of this procedure, the σ of the composite score is made equal to 1.00 for both groups. The threshold score is established through statistical decision theory and, in this example, is placed at a value of $+0.20$. A candidate who scores below this threshold, or cut-off score, is predicted to be a failure in training, while any candidate who scores above the threshold score is predicted to be a success in training. For further details of this model see Blower [3].

Given the two normal curves and the placement of the cut-off score, four areas can be defined. These four areas represent the four possible combinations of the two true states of the world, 1) candidate actually passes training and 2) candidate actually fails training, with the two decisions that are made, 1) predict candidate passes and 2) predict candidate fails. There are thus two ways to be right and two ways to be wrong.

From the normal curves and the placement of the threshold score, we can assign percentages to each one of these four areas. Since the estimated mean of the FAIL distribution, $\hat{\mu} = -0.10$, is $.30\sigma$ from the threshold score,
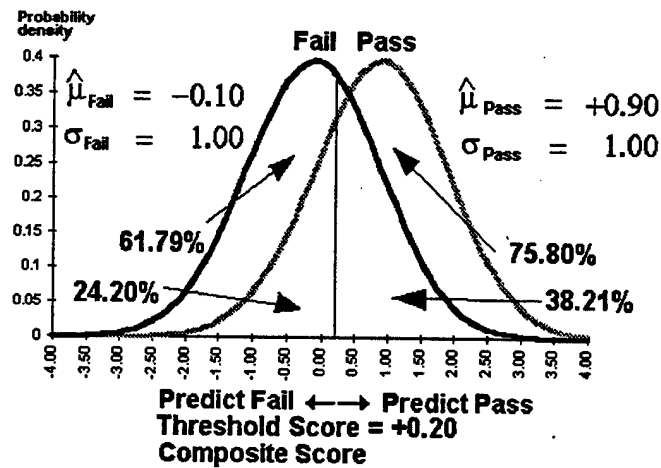
Figure 1: The initial rejection rate and attrition rate estimates for the selection system as derived from a statistical model and information from the R&D phase.

the area in the larger portion of the FAIL distribution is 61.79%. The area in the smaller portion of the FAIL distribution must therefore be 38.21%. Likewise, since the mean of the PASS distribution, $\hat{\mu} = +0.90$, is .70σ from the threshold score, the area in the larger portion of the PASS distribution is 75.80%. The area in the smaller portion of the PASS distribution is 24.20%. These numbers can be found in any table of the normal distribution.

The 2×2 matrix found in the center of the figure contains the actual frequencies for the four areas of the normal curve as obtained from the research and development (R&D) phase prior to the operational implementation of the selection system. It shows the number of candidates who fell into the two categories of a correct prediction, that is, 45 and 19, and the number of candidates who fell into the two categories for a wrong prediction, that is, 15 and 11. Of the total of 90 subjects who participated in the R&D experiments, 60 subjects eventually passed training and 30 subjects eventually failed training.

Given this context, we are now ready to estimate the rejection rate and the attrition rate. The estimated rejection rate is the number of candidates who score below the threshold and are therefore predicted to fail, divided by the total number of candidates tested. The attrition rate is the number of candidates that actually fail training from the number that went into training because their scores were above the cut-off score and were therefore predicted passes. This calculation is shown at the bottom of the figure. The rejection rate is estimated at 37.78% and the attrition rate is estimated at 19.64%.

The composite mean scores of +0.90 and −0.10 shown in Fig. 1 for the PASS and FAIL groups were constructed from a relatively small sample during the R&D phase of the selection system development. From sample size considerations alone, we are uncertain about these mean values and would expect them to fluctuate. For the same reason, the counts in the 2×2 matrix are subject to fluctuation as well. So, all in all, our estimates of the rejection rate and the attrition rate are fraught with some level of uncertainty. Probability theory, and more specifically the Bayesian formalism, are employed to deal with this uncertainty in an optimal fashion.
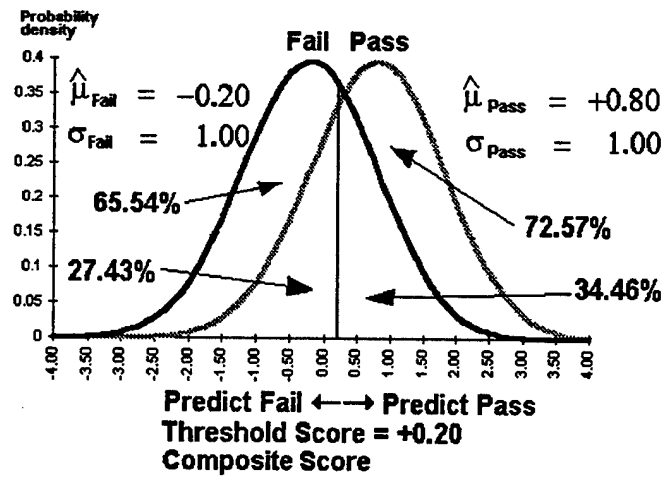
The uncertainty in rejection rates and attrition rates due to small sample sizes

As shown in Fig. 1 above, we formed an estimate of the rejection rate and the attrition rate based on the relatively small sample that was available to us during R&D. The rejection rate was a little less than 38%, and the attrition rate was just below 20%. Most likely, if we were to take another sample, the estimates of μ_{Pass} and μ_{Fail} would change. In addition, the number of candidates actually passing and failing in training would fluctuate as well with another sample.

As a numerical example, consider Figs. 2 and 3 which are just like Fig. 1 except that we consider plausible changes to the estimates of μ_{Pass} and μ_{Fail} by the discriminant analysis program on different samples. In Fig. 2 we show a new sample that gives $\hat{\mu}_{Fail} = -0.20$ and $\hat{\mu}_{Pass} = +0.80$. Notice that the discriminability of the test battery remains the same because the separation between the PASS and FAIL means is still 1.00. Also, the number of candidates actually passing training is lowered to 55 and the number failing raised to 35 to exhibit sampling fluctuations here as well. Now our estimate of the rejection rate is raised to above 42% and the attrition rate to above 23% when compared to the previous estimates.

In Fig. 3, the change in the estimate of μ_{Pass} and μ_{Fail} goes in the opposite direction with yet another hypothesized sample of 90 candidates. $\hat{\mu}_{Fail}$ now equals 0.00 and $\hat{\mu}_{Pass}$ equals +1.00. The sampling variability for the actual number of passes and fails also changes in the opposite direction with 65 passes and 25 fails. In this situation, the estimate of the rejection rate is lowered to about 31%, and the attrition rate is lowered to a little less than 18%. The reader can imagine many other possible combinations arising from repeated sampling of 90 subjects and the resulting distribution of estimated rejection rates and attrition rates.

Reasonable supposition about changes in the values to μ_{Pass} and μ_{Fail} and the actual passes and fails in training from different samples of the same small size leads to different estimates for the rejection rate and attrition rate. Therefore, due to the restrictions of small sample sizes, we are uncertain about the exact value of the rejection and attrition rate.

Probability density

Fail Pass

$\hat{\mu}_{Fail} = -0.20$

$\sigma_{Fail} = 1.00$

$\hat{\mu}_{Pass} = +0.80$

$\sigma_{Pass} = 1.00$

65.54%

72.57%

27.43%

34.46%

Predict Fail ⟵⟶ Predict Pass

Threshold Score = +0.20

Composite Score

		Predicted		
		Pass	**Fail**	
Actual	**Pass**	72.57% 40	27.43% 15	55
	Fail	34.46% 12	65.54% 23	35
		52	38	90

Rejection Rate = 38/90
= 42.22%

Attrition Rate = 12/52
= 23.08%

Figure 2: Plausible changes to the estimates of the mean composite scores as well as to the numbers actually passing and failing training for a new sample size. These kind of changes result in raising the rejection and attrition rate estimates as calculated in Fig. 1.

Fail Pass

$$\hat{\mu}_{Fail} = 0.00 \qquad \hat{\mu}_{Pass} = +1.00$$
$$\sigma_{Fail} = 1.00 \qquad \sigma_{Pass} = 1.00$$

57.93%   78.81%

21.19%   42.07%

Predict Fail ←→ Predict Pass
Threshold Score = +0.20
Composite Score

		Predicted		
		Pass	**Fail**	
	Pass	78.81%	21.19%	
		51	14	65
Actual				
	Fail	42.07%	57.93%	
		11	14	25
		62	28	90

Rejection Rate = 28/90
              = 31.11%

Attrition Rate = 11/62
              = 17.74%

Figure 3: Plausible changes to the estimates of the mean composite scores as well as to the numbers actually passing and failing training for yet a third hypothesized sample size. These kind of changes result in lowering the rejection and attrition rate estimates as calculated in Fig. 1.

The only way to capture uncertainty and deal with it quantitatively is through probability theory. The Bayesian formalism to be outlined in the next few sections uses probability theory to develop its core concepts of prior probability, likelihood, posterior probability, and predictive inferences. In the next section, we begin by discussing the prior probability distribution. It is within this distribution that we are able to express the uncertainty about the rejection and attrition rate parameters just outlined.

Inserting information into the prior distribution

Let θ stand for a rate parameter such as the rejection or attrition rate. The prior distribution for θ, $p(\theta)$, captures our state of knowledge about a rate parameter before any data are gathered, but, nonetheless, must be conditioned on whatever other information may be at our disposal. In this case, this other information is the expected rate of rejection and expected attrition rate stemming from the discriminant analysis model and the prior data gathered during the R&D phase of the selection system. The other information we wish to insert into the prior distribution is the base-level uncertainty about these two expected rates. The prior distribution thus represents our tentative working hypothesis based on all the information in our possession until actual frequency data from the operational selection system force us to modify it through Bayes's Formula.

From a technical mathematical standpoint, the beta distribution is a convenient prior distribution to capture information about parameters like the rate parameter θ that can vary between 0 and 1. To fix the precise shape and location of the beta distribution, we must specify two hyperparameters. These two hyperparameters have traditionally been given the labels of α and β, both of which must be greater than 0.

There are formulas that tell us the mean and standard deviation of any beta distribution as a function of α and β. The mean is especially simple,

$$E(\theta) = \frac{\alpha}{\alpha + \beta} \tag{1}$$

The standard deviation is slightly more complicated, but still easily computed as,

$$SD(\theta) = \sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2 \, (\alpha+\beta+1)}} \tag{2}$$

We want to choose values for α and β that reflect our state of knowledge about the rate parameter θ based on the description given in the second section. For example, we think that the mean of the rejection rate parameter should be about 38%. Likewise, the mean of the attrition rate parameter is about 20%. These estimates are based on the one rather small sample of candidates that we were able to test during R&D. But we are also cognizant of some uncertainty attached to these estimates. Assume that we are willing to let the standard deviation of θ be a value such that $E(\theta) \pm 2SD(\theta)$ captures most of our uncertainty about θ.

Let us first assign a prior probability for the rejection rate parameter called θ_r. If we let $\alpha = 19$ and $\beta = 31$, then using Equations (1) and (2) we find that,

$$E(\theta_r) = \frac{19}{19 + 31}$$

$$= .38$$

$$SD(\theta_r) = \sqrt{\frac{19 \times 31}{50^2 \times 51}}$$

$$= .0680$$

Therefore, if $E(\theta_r) = .38$ and $SD(\theta_r) = .0680$,

$$\text{Rejection rate} = E(\theta_r) \pm 2SD(\theta_r)$$

Prior probability density function for θ_r

$\alpha = 19$
$\beta = 31$
$E(\theta) = .38$
$SD(\theta) = .0680$

θ_r

Rejection Rate Parameter

Figure 4: A sketch of the beta distribution that serves as the repository for a state of knowledge about the true rejection rate parameter based solely on prior R&D data and a discriminant analysis model.

$$= .38 \pm 2(.0680)$$

$$= .2440 \text{ to } .5160$$

and our state of knowledge is that the true rejection rate probably lies in the range from about 25% to almost 52%. This is the price we have to pay for uncertainty about the unobservable rejection rate parameter. Using larger values of α and β than those given above will reduce the uncertainty about $E(\theta)$, and conversely, using smaller values will increase the uncertainty. Fig. 4 contains a sketch of the beta distribution, which serves as the prior probability for the true rejection rate parameter. This curve encapsulates our uncertainty about θ before any frequency data has been collected.

Now, using the very same techniques, we can assign a prior probability to the attrition rate parameter, θ_a. In this case, if we let $\alpha = 20$ and $\beta = 80$, then using Equations (1) and (2) we find that,

$$E(\theta_a) = \frac{20}{20 + 80}$$

$$= .20$$

$$SD(\theta_a) = \sqrt{\frac{20 \times 80}{100^2 \times 101}}$$

$$= .0398$$

Therefore, if $E(\theta_a) = .20$ and $SD(\theta_a) = .0398$,

$$\text{Attrition rate} = E(\theta_a) \pm 2SD(\theta_a)$$

$$= .20 \pm 2(.0398)$$

$$= .1204 \text{ to } .2796$$

8

Prior probability density function for θ_a

Probability Density

$\alpha = 20$
$\beta = 80$
$E(\theta) = .20$
$SD(\theta) = .0398$

95%

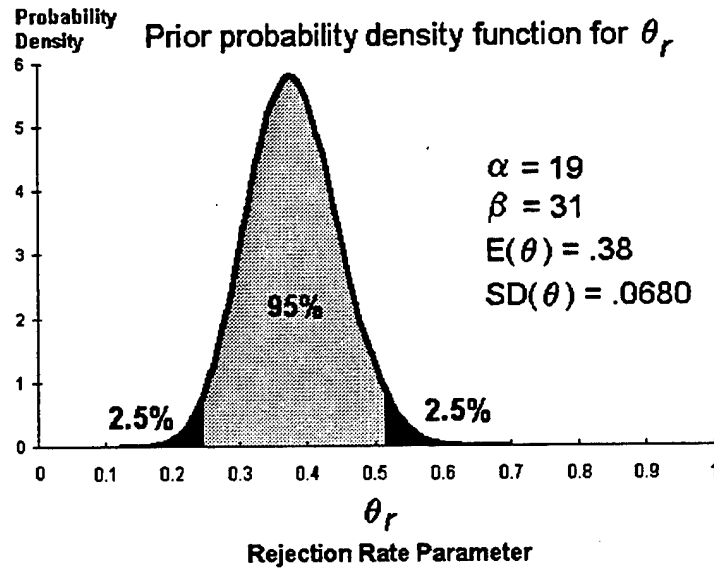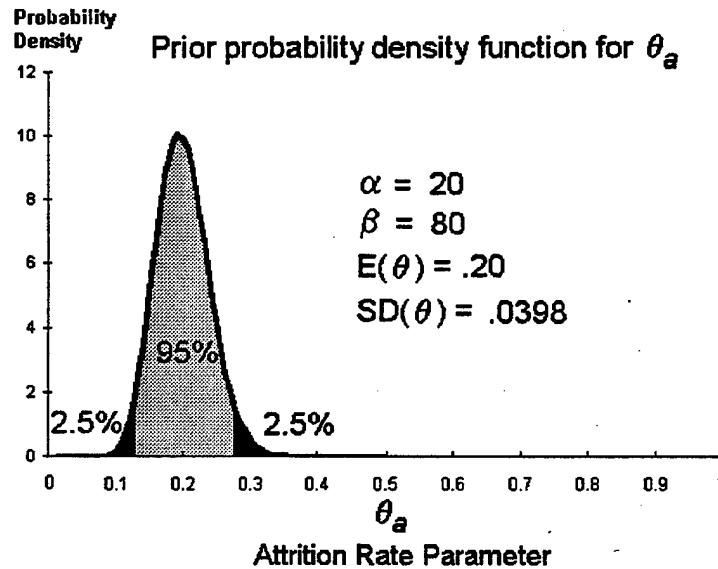2.5%   2.5%

θ_a

Attrition Rate Parameter

Figure 5: A sketch of the beta distribution that serves as the repository for a state of knowledge about the true attrition rate parameter based solely on prior R&D data and a discriminant analysis model.

our state of knowledge is that the true attrition rate probably lies in the range from about 12% to almost 28%. Since it appeared from our numerical examples of the last section that the attrition rate parameter did not vary over as large a range as the rejection rate parameter, we use larger values for α and β. This results in a narrower width for the prior probability for θ_a and, as such, reflects reduced uncertainty about the values θ_a could reasonably assume. Figure 5 shows the prior probability density function assigned to the attrition rate parameter. Notice that the mean is shifted towards the left and the curve possesses a narrower width because we are less uncertain of the range of values that it might take on.

How the state of knowledge about the rate parameters change with data

After the selection system is transitioned from R&D to an operational system, frequency data are collected on the number of rejections and the number of attritions. This additional information can be used to update the state of knowledge about the true rejection and attrition rate parameters. This updating is accomplished through an application of Bayes's Formula. The major result from Bayes's Formula is a revision to the prior distribution. This updated probability distribution is called the posterior distribution because it exists after the data have been incorporated as extra information.

The posterior distribution, when compared to the prior distribution, conditions on two more pieces of additional information that are assumed to be true. The prior distribution conditions on the assumed truth of α and β, while the posterior distribution, say for the rejection rate parameter, conditions not only on α and β, but also on the number of candidates actually tested and the number of candidates rejected because of poor performance on the test battery.

Say that frequency data for rejections and attritions have been collected from the operational selection system for the past 5 years. For brevity, we will discuss only rejection rates, but the same argument applies as well to attrition rates. The notation for these empirical frequency counts that have been gathered in the past is as follows: n is the total number of candidates that have been tested in the past; y is the actual number of candidates that were rejected because their scores fell below the threshold score; and $p(\theta_r | y, n, \alpha, \beta)$ is the notation for the posterior distribution for the rejection rate parameter conditioned on the past frequency counts. The posterior probability is

9

found by an application of Bayes's Formula that relies on the likelihood for the past data and the prior probability of θ_r.

$$p(\theta_r|y, n, \alpha, \beta) = \frac{P(y|\theta_r, n) \times p(\theta_r|\alpha, \beta)}{\int_0^1 d\theta_r \; P(y|\theta_r, n) \times p(\theta_r|\alpha, \beta)} \tag{3}$$

Table 1 presents a simulation of the number of candidates tested and the number rejected by the test battery for each of the five years after the operational implementation of the selection system. The operational selection system tests about 100 candidates per year. The last two columns contain a cumulative record of the number tested, n, and the number rejected, y.

Table 1: The simulated number of candidates rejected by the selection system for the first 5 years after becoming operational.

Year	Number tested	Number rejected	n	y
1	100	32	100	32
2	95	30	195	62
3	106	46	301	108
4	102	35	403	143
5	92	29	495	172

If we wish to incorporate these frequency counts for rejections over the past 5 years into the information already contained in the prior for θ_r, we can set $n = 495$ and $y = 172$ and calculate the posterior distribution. The posterior distribution can be calculated and plotted just like the prior distribution. Again, the appropriate formula for calculating the posterior distribution is contained in Reference [4] as Equation (15). Fig. 6 shows the posterior distribution as conditioned on the empirical data from Table 1 overlaid onto the prior distribution taken from Fig. 5. The additional information in the data has changed the state of knowledge and therefore the uncertainty about the true rejection rate. The posterior curve is shifted to the left with a mean closer to $E(\theta) = .35$, but it is more concentrated around a narrower range of values. We are less uncertain about θ after having collected these data. Before the empirical data were gathered, the state of knowledge about θ was that it probably ranged from .23 to .52 with a mean of .38. After incorporating the data via Bayes's Formula, the state of knowledge about θ_r is updated. The rejection rate parameter has been concentrated into a highly probable range of .32 to .38 with a mean of .35.

The Bayesian formalism for predictive inference

Of great practical interest to the user of the selection system is the question, "How can one *predict* the number of candidates that will be rejected by the test battery in some future time period?" To answer this question, the Bayesian formalism includes a fourth concept, *predictive inference*.

The ingredients for the predictive inference consist of the likelihood for any future number of rejections and the posterior based on empirical data consisting of past frequency counts. The posterior probability, by definition, also contains the information discussed in the second section and which was originally placed into the prior. Before we present the Bayesian-inspired formula for producing a predictive inference, we will spend some time reviewing the notation. The predictive inference formula will then be used to generate graphs that speak to the question posed at the beginning of this section.

In the situation we are examining here, we are concerned with unobserved, or theoretical, rate parameters, *viz.*, the true rejection rate and the true attrition rate. We let θ_r and θ_a stand for such rate parameters. As a rate parameter, θ takes on values between 0 and 1. We have already seen that the prior, $p(\theta)$, encapsulates the
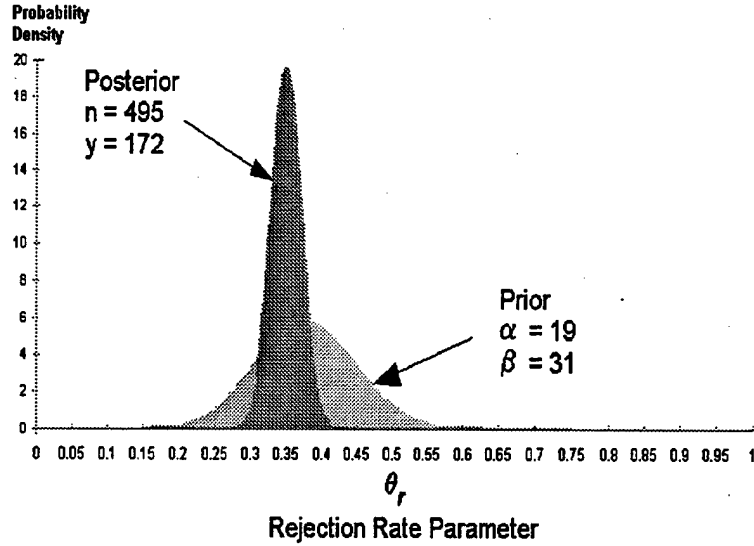
10

Figure 6: A comparison of the posterior distribution for θ_r with the prior. The shift in location and the narrowing of the curve represents the impact of the empirical data.

uncertainty surrounding the true rate parameter before we collect any actual frequencies from the operational system. Using α and β allowed us to specify our prior state of knowledge about the rate parameter. We leveraged α and β to translate what we knew about θ from the particular statistical model we happened to be using. In addition, in what is quite an important feature of this methodology, we included the uncertainty surrounding θ arising from the fluctuations due to small sample size. This uncertainty about θ was reduced when we started collecting data from the operational system. The number of rejections, y, from the total number of candidates tested, n, provided us with additional information that we processed into the posterior distribution.

Finally, we let z stand for the number of candidates that will be rejected by the test battery in some future time period of interest, say the upcoming year. N represents the future total number of candidates that would be tested in the upcoming year. Later in this report, we examine values of N equal to 10, 50 and 100. In projecting the future number of rejections, N can assume any sample size that might be of interest. Values for z can potentially range from $0, 1, 2$, all the way up to N.

$L(z|\theta, N)$ is the likelihood of obtaining $z = 0, 1, 2, \cdots, N$ rejections in some future sample. The likelihood is conditioned on some value of θ, which is taken to be true, and the given value of N. For example, $L(z = 17|\theta_r = .30, N = 50)$ is the likelihood of obtaining 17 rejections from a sample of size 50 when the true underlying rate parameter, in this case the rejection rate, is equal to .30.

Two fundamental axioms of probability theory are used to link the elements just discussed, namely, the likelihood of a future number of rejections and the posterior probability of the rate parameter, into the formula for predictive inference.

$$P(z|N, y, n, \alpha, \beta) = \int_0^1 d\theta_r \ L(z|\theta_r, N) \ p(\theta_r|y, n, \alpha, \beta) \tag{4}$$

For example, $P(z = 17|N = 50, y = 172, n = 495, \alpha = 19, \beta = 31)$ will tell us the probability of obtaining 17 rejections in a future sample of size 50 given that data exist where 172 known rejections occured in a known sample size of 495, and, in addition, where we have specified prior knowledge about the theoretical rate parameter in α and β. Working out the mathematical consequences of Equation (4) is a well-known exercise in Bayesian statistics. All the tedious details are spelled out in a technical report (Blower [4]), which also derives a

11

computational formula that can be inserted into a computer program.

The important point is that, given values for N, z, y, n, α and β, such a program can compute the predictive probability. We shall use a more concise notation for the predictive inference, $P(z)$, that deletes the explicit reference to the fact that N, y and n are known and that α and β are assumed to take on given values. Repeating, $P(z)$ is the Bayesian state of knowledge for obtaining z rejections from a future sample of size N based on past data where y rejections actually occurred in a sample of size n, and where α and β capture any prior knowledge about θ, the rate parameter.

Numerical example of how to calculate z

Before showing the graphs of the predictive probability for the number of rejections for various future sample sizes N, the computational formula will be illustrated for one set of values of N, z, y, n, α and β. First, N is a value that we are free to choose because it happens to be of intrinsic interest in the particular problem that confronts us. Suppose that 100 candidates are projected to be tested in the upcoming year, and we are interested in whether the rejection rate is holding steady at what our theoretical model told us it should be. Halfway through testing, we would like to assess the selection system for any untoward circumstances. We therefore let $N = 50$ and calculate $P(z)$ for all values of z from $z = 0$ to $z = 50$, yielding a discrete probability distribution for 51 values.

Fixed values, n and y, come from previous data. For the present numerical example, we will use the data given in Table 1 to fill in the values for y and n. Therefore, $n = 495$ and $y = 172$ because, as mentioned above, there were 172 rejections from the 495 candidates tested.

All the values needed to perform the calculation are now present. If $z = 17$ and $N = 50$, we want to find the probability of obtaining 17 rejections from the selection system from a future sample of size 50. The inference will be based on data indicating that in the past 5 years 172 candidates were rejected out of the 495 that were tested. In addition, we inserted initial information about the rejection rate through α and β, which indicated that we thought that the mean rejection rate was about .38 with an attached 95% confidence interval stretching from .25 to .52. This range of values reflected our initial uncertainty about the value .38 for the rejection rate parameter because we took notice of the small sample size in the R&D phase and built this into our prior distribution.

The formula to compute the predictive inference contained in Equation (4) is taken from Blower [4].

$$P(z) \quad = \quad f(N, z, n, y, \alpha, \beta) \tag{5}$$

$$= \quad e^x \tag{6}$$

where

$$x \quad = \quad \ln N! - \ln(N - z)! - \ln z! + \ln(\alpha + \beta + n - 1)! - \ln(y + \alpha - 1)!$$

$$- \ln(n - y + \beta - 1)! + \ln(z + y + \alpha - 1)!$$

$$+ \ln(N - z + n - y + \beta - 1)! - \ln(\alpha + \beta + N + n - 1)! \tag{7}$$

Table 2 lists the necessary computations of the log factorials needed to solve Equations (6) and (7). Most computer software designed for numerical calculations will possess built-in functions for determining the natural log transform of factorials. Sometimes, the log transform of the gamma function is provided instead.

Substituting these numbers into Equations (6) and (7) yields,

$$P(z) \quad = \quad f(N = 50, n = 495, y = 172, z = 17, \alpha = 19, \beta = 31)$$

$$= \quad e^x$$

Table 2: A listing of all the numerical values and the log transforms of the factorial functions needed to calculate a predictive probability for the example given in the test.

Terms in Eq. (7)	Value Col. 1	ln of terms in Eq. (7)	Value Col. 3
N	50	$\ln N!$	148.48
$N - z$	33	$\ln(N - z)!$	85.05
z	17	$\ln z!$	33.51
$\alpha + \beta + n - 1$	544	$\ln(\alpha + \beta + n - 1)!$	2866.70
$y + \alpha - 1$	190	$\ln(y + \alpha - 1)!$	810.48
$n - y + \beta - 1$	353	$\ln(n - y + \beta - 1)!$	1721.72
$z + y + \alpha - 1$	207	$\ln(z + y + \alpha - 1)!$	900.46
$N - z + n - y + \beta - 1$	386	$\ln(N - z + n - y + \beta - 1)!$	1916.85
$\alpha + \beta + N + n - 1$	594	$\ln(\alpha + \beta + N + n - 1)!$	3203.92

$$x = -2.19$$

$$P(z) = .1121$$

Therefore the probability of obtaining 17 rejections from a future sample size of 50 is a little over 11%.

Examples of the predictive probability distribution

While the preceeding section showed how one value of z is calculated, a computer program can calculate z for all values of z from 0 to N. The program can then produce a bar graph giving the predictive probability at each one of these z values. The uncertainty about the frequency of future rejections for any given sample size is captured by this predictive probability distribution. The graph can be consulted to obtain an idea of which frequency counts are compatible with this distribution because they fall into an high probability area and which frequency counts seem to be extreme because they fall into an extremely low probability area. The usual criterion will be used that accepts a frequency count that falls within the 95% region of the predictive inference distribution as compatible, while a frequency that falls into the 5% region is viewed as extreme.

This section contains three graphs of the predictive probability distributions for sample sizes of $N = 10$, 50, and 100. These values of N allow one to monitor the frequency counts at an early, mid, or final stage of the upcoming year. These graphs would be used to monitor the performance of the selection system after Year 5 by assessing where the actual number of rejections in a sample of size N fits into the predictive distribution. We begin by showing Figure 7 which contains the graph for a sample of size $N = 10$. As portrayed in this Figure and the next two, the predictive probability distributions are discrete distributions with $z = 0, 1, \cdots N$ along the x-axis. The y-axis indicates the probability for any particular z with the probability calculated by Equations (6) and (7). The distributions are generally symmetrical with the mode of the distribution, that is, the most probable value of z, located in the center of the distribution mass. Some slight positive skewness however is evident for small sample sizes such as $N = 10$.

Looking at Fig. 7, the most probable value for the number of rejections for a future sample size of $N = 10$ is $z = 3$ rejections with a probability just about 25%. If we had to pick one value as our guess for the number of rejections this would be it. Of course, this corresponds nicely to an intuitive guess based on the past empirical data and the theoretical estimate for the rejection rate parameter. All of this was captured in the posterior probability distribution, which peaked at $\theta_r = .35$. The Bayesian formalism never violates common sense in cases where the answer is obvious. Unfortunately, unaided intuition is not so reliable in assessing the probabilities of other possible
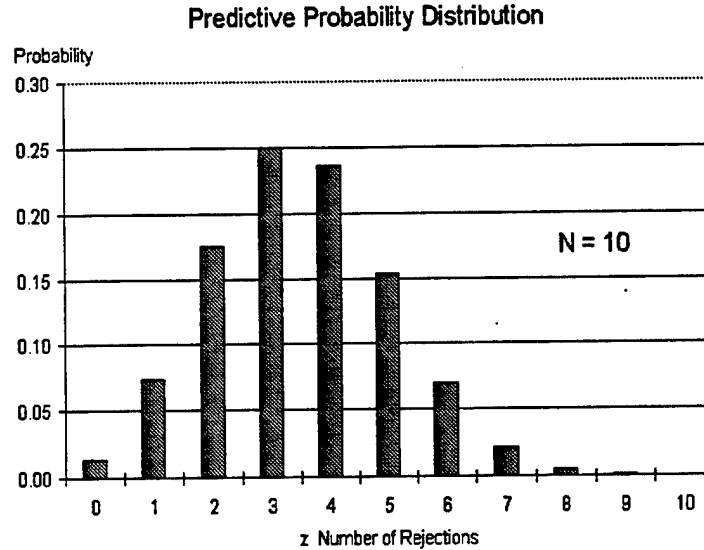
Predictive Probability Distribution

Figure 7: The predictive probability graph showing the probability of the number of rejections for the selection system for a sample size of $N = 10$. This probability distribution is based on past empirical data of 172 rejections out of 495 candidates tested and a theoretical judgment of the rejection rate based on a beta distribution with a mean of .38.

rejection numbers.

It is important to realize that probability distributions reflect the uncertainty about any one value, and it appears that there is a lot of uncertainty for such a small sample size. For example, although 3 rejections is the most probable number of rejections, nearby values cannot be ruled out. 1, 2, 4, 5, and 6 rejections are also not unlikely. The number of rejections that we deem unlikely occurs at the ends of the distribution. Therefore, we would consider no rejections, occurring at the left end of the distribution, and 7 through 10 rejections, occurring at the right end of the distribution, as improbable. The probability of 0 rejections is .0140, and the probability of 7, 8, 9 or 10 rejections is .0221, .0046, .0006, and .00003, respectively. Together these five extreme values make up 4.1% of the total probability, while the other rejection numbers considered as compatible with the distribution make up the remaining 95.9%. A more conservative decision would consider even 7 rejections as falling into the high probability region. This new high probability region now consists of about 98% of the total probability.

Next, consider Fig. 8 where the future sample size has been increased to $N = 50$. This probability distribution is more symmetrical than the previous one which was based on a smaller sample size. The most probable rejection frequency count is 17, again as intuition would suggest, with a probability of 11.21%. This is the value of z that was calculated in the previous section. Nearby values are also quite probable, dropping off somewhat like the normal curve, although it must be remembered that this is a discrete probability distribution. Just a casual visual inspection of the bar graph of probabilities tells us that rejection frequencies lower than about 8 and greater than about 28 are improbable.

Just as before, we would like to identify two values of z, one value at the lower end and one value at the upper end, that together demarcate about 95% of the total probability. For example, in Fig. 8 for the $N = 50$ case, if $z = 10$ at the lower end and $z = 25$ at the upper end, the total probability for all z values in between these two values is just about 95%. On the basis of this analysis, we would say that obtaining $11, 12, \cdots,$ or 24 rejections from a future sample of 50 candidates tested would be compatible with the distribution, but that 0 to 10 rejections or 25 to 50 rejections would be an extreme number.

14

Predictive Probability Distribution
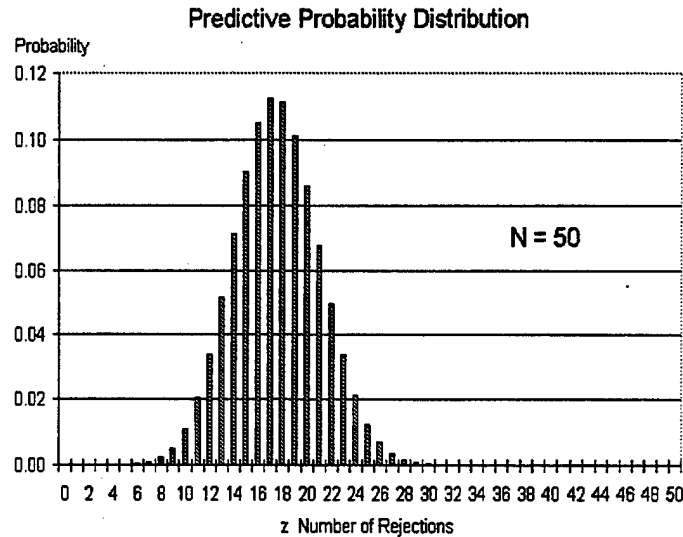
N = 50

z Number of Rejections

Figure 8: The predictive probability graph showing the probability of the number of rejections for the selection system for a sample size of $N = 50$. This probability distribution is based on past empirical data of 172 rejections out of 495 candidates tested and a theoretical judgment of the rejection rate based on a beta distribution with a mean of .38.

Figure 9 can be interpreted in a similar manner. Figure 9 shows the bar graph of the predictive probability for the largest sample size treated, $N = 100$. The graph points to frequency counts in the range of 31 to 39 as the most probable. Together these rejection numbers contribute over 61% to the total distribution. As the sample size N becomes larger, the influence of the posterior distribution shown in Fig. 6 becomes more obvious. Even a quick glance at this distribution tells us that frequency counts below 20 and above 50 are rare. For the more refined analysis, we can pick out $z = 24$ and $z = 46$ as the demarcation frequency counts that together chop off about 5% of the distribution. We would be surprised, for example, to see 20 rejections or 50 rejections from the next sample of 100 candidates, but not surprised to observe 25 rejections or 44 rejections from the same sample size.

Using the predictive distribution as a signal of an external change

If μ_{Pass} and μ_{Fail} were completely stable, then we would expect the posterior distribution to converge upon a single rejection rate and a single attrition rate over time. As the actual frequency data for the number of rejections and the number of attritions accumulates, the posterior probability distribution becomes more and more peaked around one value of θ.

For example, imagine that $n = 10,000$ candidates had been tested over long period of time and that $y = 3800$ of these candidates had been rejected as predicted failures by the test battery. The posterior distribution would look like the curve depicted in Fig. 10. Here we see a very narrow spike centered on $E(\theta_r|y, n, \alpha, \beta) = .38$. Then we would be more confident that Fig. 1 portrays the true state of affairs with μ_{Fail} located at -0.10 and μ_{Pass} located at $+0.90$.

Likewise, using the same numerical example as above, after the 3800 candidates had been rejected by the test battery, 6200 candidates would have been predicted passes and therefore entered training. Let us assume that of these 6200 who entered training, 1240 attrited for an estimated attrition rate of 20%. Technically, we would have a posterior probability distribution for θ_a very closely centered around .20. These frequency counts for the number of attritions add to the weight of evidence provided by the rejection counts concerning the location of μ_{Pass} and μ_{Fail} as shown in Fig. 1.
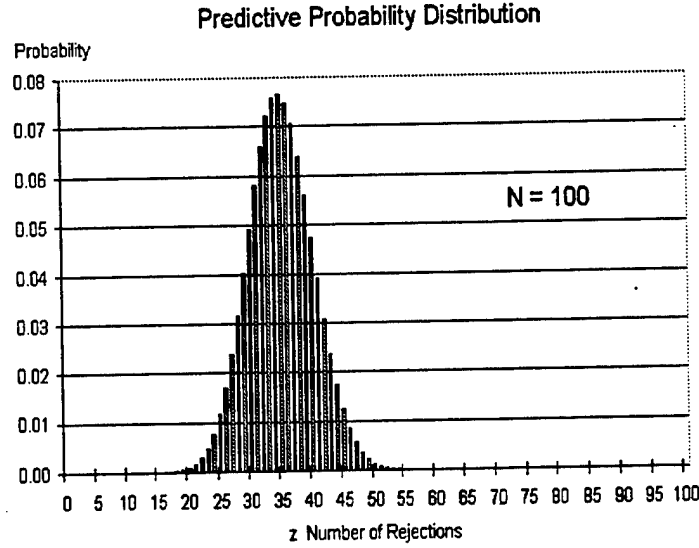
15

Predictive Probability Distribution

Figure 9: The predictive probability graph showing the probability of the number of rejections for the selection system for a sample size of $N = 100$. This probability distribution is based on past empirical data of 172 rejections out of 495 candidates tested and a theoretical judgment of the rejection rate based on a beta distribution with a mean of .38.

As the amount of data increases, the posterior distributions for θ_r and θ_a become even more narrowly peaked around a given value. In the example above, θ_r becomes more peaked around the value of .38. In the limit of an infinite amount of data, the posterior distribution becomes a delta function,

$$p(\theta_r | y, n, \alpha, \beta) = \delta(\theta_r - .38) \tag{8}$$

and the predictive distribution collapses to,

$$P(z | N, y, n, \alpha, \beta) = \int_0^1 d\theta_r \; L(z | \theta_r, N) \; \delta(\theta_r - .38) \tag{9}$$

$$= L(z | \theta_r = .38, N) \tag{10}$$

For example, if one is interested in a future sample size of 100 and wants to know the probability of obtaining 30 rejections, then

$$L(z = 30 | \theta_r = .38, N = 100) = \binom{100}{30} .38^{30} .62^{70}$$

$$= \frac{100!}{30! \; 70!} .38^{30} .62^{70}$$

$$= .0213$$

We have completely eliminated the uncertainty about the location of μ_{Pass} and μ_{Fail} due to the small sample size that we were forced to contend with during R&D. The only concern with sample size now is the size of the future sample N and that uncertainty is accurately reflected in the predictive distribution. This rests on the supposition that μ_{Pass} and μ_{Fail} remain static. On the contrary, events in the real world might cause dynamic changes to
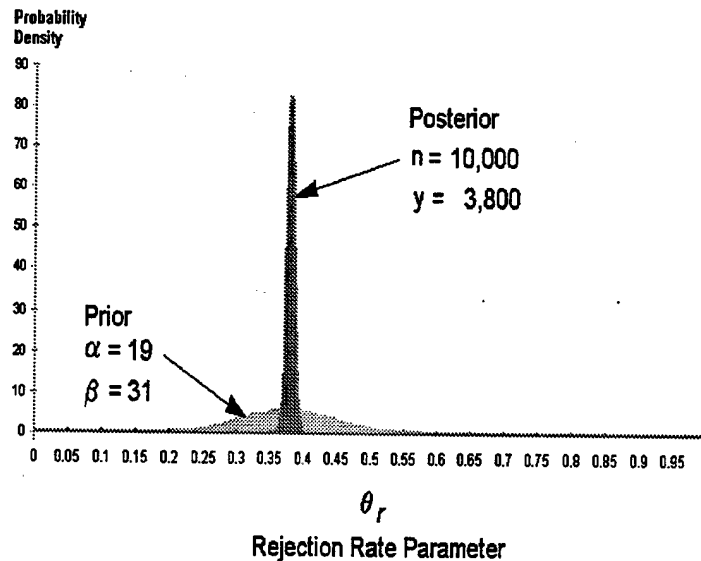
16

Figure 10: How the posterior distribution can become a very narrow spike after a large amount of empirical data. This posterior distribution represents very little uncertainty about the value of .38 for the rejection rate parameter.
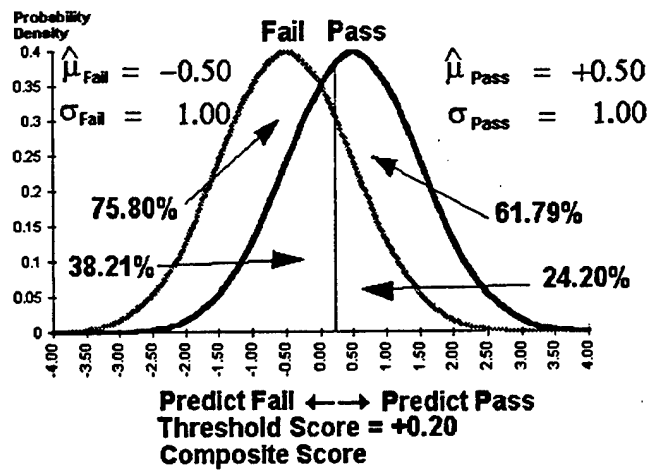
μ_{Pass} and μ_{Fail}. Any disruption to μ_{Pass} and μ_{Fail} would translate into changes for the number of rejections and attritions.

Similarly, μ_{Pass} and μ_{Fail} could be impacted by real-world events that cause them to move further apart or move closer together. They do not necessarily have to move in concert. For example, the mean of the PASS group could be moving down and the mean of the FAIL group could be moving up, and vice versa. These changes are *not* due to the small sample sizes discussed at the beginning of the report. The uncertainty attending small sample size was integrated into the prior probability distribution. The present uncertainty is engendered by external perturbations such as hardware malfunctions, lack of calibration, new computer operating systems, changes in skill level of the candidate pool, or changes in the training criterion, to mention just a few possible causes.

We would like to have some signal that would alert us to these kinds of major perturbations to μ_{Pass} and μ_{Fail}. We suggest that a low probability frequency count from either of the predictive probability distributions serve as that warning signal. For example, in Fig. 8 we constructed the predictive probability distribution for θ_r for $N = 50$ based on the prior distribution and some empirical data. Say that after a half year of testing we found that the selection system had rejected 26 candidates. Twenty six candidates is an extremely high number of rejections and happens to fall into the upper 2 1/2% region of the predictive probability distribution. Of the 24 candidates who did enter training, 4 were attrited.

A plausible scenario explaining this large number of rejections is that μ_{Pass} and μ_{Fail} have been affected by some external cause. If μ_{Pass} is shifted downwards to +0.50 and μ_{Fail} is shifted downwards by the same amount to −0.50, then the estimated rejection numbers and attrition numbers from this model are compatible with the actual observed data. Look at Fig. 11 which shows this downward movement of μ_{Pass} and μ_{Fail}. According to this model, 26 rejections and 4 attritions are now to be expected. It could be that the external perturbation was an overall degradation in the candidates's skill level. Another possibility is that insertion of a new A/D card inadvertently made some tasks harder, causing every candidate to score lower although their skill level remained unchanged.

The main point is that an anomalous number of rejections or attritions, as judged by the predictive probability

17

Probability
Density

Fail Pass

$\hat{\mu}_{Fail} = -0.50$ $\hat{\mu}_{Pass} = +0.50$

$\sigma_{Fail} = 1.00$ $\sigma_{Pass} = 1.00$

75.80%   **61.79%**

38.21%   **24.20%**

Predict Fail ←—→ Predict Pass
Threshold Score = +0.20
Composite Score

	Predicted		
	Pass	**Fail**	
Pass	61.79% 20	38.21% 13	33
Fail	24.20% 4	75.80% 13	17
	24	26	50

Actual (row label for Pass/Fail)

Number of rejections = 26
Number of attritions = 4

Figure 11: A plausible downward movement in the mean composite scores for the PASS and FAIL groups to explain the abnormally high number of rejections. This shift in the means represents the impact of some external perturbation.

18

distribution, should be investigated and, if possible, the cause of the external perturbation should be tracked down. This is the quality control feature of the predictive distribution that alerts us to the fact that the locations of μ_{Pass} and μ_{Fail} may be shifting. On the other hand, if the number of rejections and attritions remains compatible with the predictive distribution, then the uncertainty about the non shifting locations of μ_{Pass} and μ_{Fail} is being reduced.

References

1. Blower, D. J. *A Cost-Benefit Analysis of the Impact of Selection Testing on Advanced Flight Training* NAMRL Special Report 97-1, Naval Aerospace Medical Research Laboratory, Pensacola, FL, 1997.

2. Blower, D. J. *Analysis of Rejection and Attrition Data from the LCAC Selection System.* In Review. Naval Aerospace Medical Research Laboratory, Pensacola, FL, 1998.

3. Blower, D. J. *The Prediction Algorithm for the Landing Craft Air Cushion (LCAC) Selection System.* NAMRL-1395, Naval Aerospace Medical Research Laboratory, Pensacola, FL, 1996.

4. Blower, D. J. *A Quantitative Framework for Conducting Cost Benefit Analyses of Selection Tests.* NAMRL-1397, Naval Aerospace Medical Research Laboratory, Pensacola, FL, 1997.

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE 4 Feb 1999	3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE
Monitoring Operational Selection Systems Through Frequency Counts: An Application of Bayesian Predictive Inference

5. FUNDING NUMBERS

PMS 377
Work Unit CE02

6. AUTHOR(S)
David J. Blower

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Aerospace Medical Research Laboratory
51 Hovey Road
Pensacola Fl 32508-1046

8. PERFORMING ORGANIZATION REPORT NUMBER

NAMRL-1405

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Naval Ocean Systems Center
2531 Jefferson Davis Hwy
Arlington, VA 22242-5160

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Approved for public release; distribution unlimited.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Military selection systems are put into place in an attempt to reduce high attrition rates in training. Typically, though, after the selection system has experienced some operational use, the rejection rates and attrition rates are observed to fluctuate over time. The question then naturally arises, "Do these fluctuations represent some substantial change in the real world that should be investigated, or do they merely reflect the statistical vagaries seen in any small sample size?"

In this technical memorandum the question just posed is addressed quantitatively by standard Bayesian statistical techniques. A predictive inference can be made about future frequency counts based on the empirical data of the past frequency counts. This kind of analysis is helpful whenever there is concern that something fundamental might have caused the rejection rates and/or attrition rates of the selection system to change. For example, the rejection rate of a given selection system is suspected of having dramatically changed over the past few years. Before we can attempt to track down the cause of this alleged rate increase, we must first establish that the increase in frequency counts is not simply due to statistical fluctuations inherent in small sample sizes. It is quite easy to be misled into thinking that a "trend" based on relatively small numbers portends a significant change in the underlying rate parameter. The techniques detailed in this report will help researchers disentangle sample size fluctuations from external perturbations to the rate parameter. In the case of any selection system, these techniques can be employed to determine whether there is a justification for investigating such fundamental changes as a shift in the hardware configuration, a change in the ability levels of the candidate population, changes to the training regime, or changes in the validity of the current prediction algorithm.

14. SUBJECT TERMS

Military selection tests, Attrition rates, Bayesian statistics

15. NUMBER OF PAGES
25

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT SAR

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102