OPERATIONS RESEARCH CENTER

working paper

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PRIORS: AN INTERACTIVE COMPUTER PROGRAM FOR FORMULATING AND UPDATING PRIOR DISTRIBUTIONS

by

John VandeVate

OR116-82 JUNE 1982

Prepared under Grant No. 80-IJ-CX-0048 from the National Institute of Justice, U.S. Department of Justice. Points of view or opinions stated in this document are those of the author and do not necessarily represent the official position or policies of the U.S. Department of Justice.

Table of Contents

1. Introduction

PRIORS is an interactive PL/I program written under National Institute of Justice Grant Number 80-IJ-CX-0048. The program is designed to assist evaluators in formulating, modifying and updating prior distributions.

OPT2 is likewise an interactive PL/1 program written under this grant. The products of PRIORS may be useful in formulating Bayesian decision rules with OPT2.

1.1 Why Prior Distributions?

One of the main concerns of evaluations is to collect information. Both the qualitative information of "process evaluators" and the quantitative information of "outcome evaluators" are relevant to evaluations. However, as in many fields, merging these distinct types of information often leads to conflict. We feel that the apparent conflict between "process evaluators" and "outcome evaluators" can in some cases be resolved through Bayesian analysis. The idea is to use the qualitative information of the process evaluator to form a "prior distribution" and the statistical information of the outcome evaluator to update the prior and obtain a "posterior distribution".

More than just a resolution to the conflict between process and outcome evaluators, Bayesian analysis offers the adaptability necessary in the face of such multifaceted and changing problems as crime, drug and alcohol abuse, family counseling, etc. In simple hypothesis tests for example, classical statistics formulates decision rules strongly biased in favor of the null hypothesis.

Bayesian analysis and more specifically conjugate prior distributions offers a tractable, appealing method for overcoming the deficiencies of classical statistics thereby affording a vehicle for resolving the conflict between process and outcome evaluators.

1.2 What is a Prior Distribution?

A prior distribution is as its name suggests, simply a probability distribution for the outcome of some experiment or trial based on information available before the event. Most people for example would set their chances of getting Heads upon tossing a coin at fifty-fifty -- before ever seeing the coin. This simple example captures the essence of prior distributions --

namely prior distributions translate previous and often qualitative knowledge into quantitative information.

Continuing with our coin-tossing example, suppose we wanted to determine whether or not a coin was "fair". First we take the coin and turn it over in our hand, feel its weight and check that one side is Heads and the other Tails. Imagine our chagrin if we had simply begun by tossing the coin a number of times before detecting that both sides were Heads! Then, based on these observations we formulate a prior distribution for the probability that the coin, when tossed, will land Heads. Tossing the coin a number of times we obtain the sequence of observations (0_i) with say 0_1 Heads, 0_2 Tails, etc. With this quantitative information we update our prior to obtain the posterior distribution. The posterior distribution is simply the conditional distribution of p given the sequence of observations (0_i).

One special class of prior distributions, conjugate priors, is mathematically and intuitively appealing in that the prior and posterior distributions come from the same mathematical family. The program PRIORS deals exclusively with these conjugate prior distributions.

2. Hypothesis Testing

Hypothesis testing is no longer simply a laboratory tool. Today it affects the courses of thousands of lives and millions of dollars. FDA regulations are an especially tangible example of the present power of hypothesis testing. Admissions policies to public assistance programs, special education programs, limited medical facilities and psychiatric institutions are, intentionally or not, decision rules for hypothesis tests.

The problems involved in formulating such decision rules, not to mention their consequences, set hypothesis testing in social institutions apart from testing in laboratories. It is neither politically acceptable nor economically feasible to determine which citizens will receive public assistance according to the same formulas used to determine the effectiveness of malathion against Drosophila.

Consider the problem of formulating requirements for admission to the following public assistance program. The law requires that people be admitted solely on the basis of a single summary measure: their present assets. Since a family's economic situation is complex and multifaceted, it is not likely that any single measure will correctly detect all "truly needy" families or all families who are "not truly needy." Yet, we must construct a reasonable decision framework within the structure of the law.

Our problem then is to determine a decision threshold having the property that applicants whose assets exceed the threshold value will not be admitted. We realize that any given threshold value will have dramatic effects on the lives of thousands of people. If for example we set our decision threshold too high, many deserving applicants will be unjustly turned away. On the other hand if we set out decision threshold too low, undeserving applicants may receive money earmarked for the needier. In order to determine the best decision threshold we undertook an extensive retrospective study to determine how the assets of past applicants aligned themselves. Highly trained case workers reviewed the case of each previous applicant. Based on the case history, they decided whether or not the applicant was "truly needy." We then studied the level of assets at the time of application within each group -- "truly needy" and "not truly needy." We found that half of all applicants were, on the basis of this study, considered "truly needy."

Unfortunately, however, there was no level of assets which could unambiguously distinguish between the two groups. In fact the study found the asset distribution shown in Figure 1.

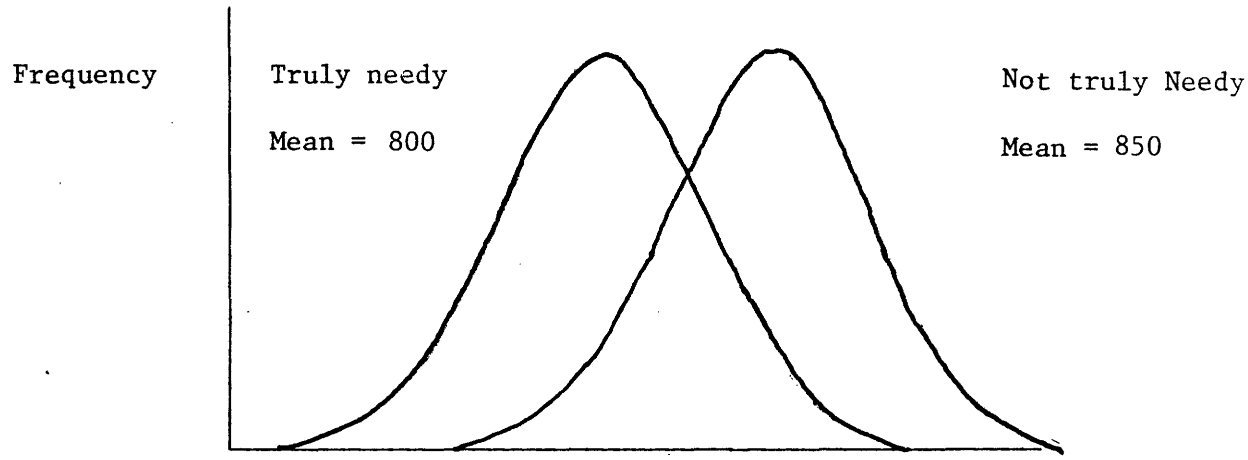

Figure 1

Asset Distributions of "Truly Needy"

and "Not Truly Needy"

It is clear from Figure 1 that regardless of what threshold value we choose we will reject truly needy applicants, accept not truly needy applicants or both. In this situation Classical Statistics would ordinarily prescribe either the .05 alpha-level decision rule or the .05 beta-level decision rule. The .05 alpha-level decision rule is, roughly speaking, designed to ensure that the chances of turning away a truly needy applicant remain below one in twenty. The .05 beta-level decision rule on the other hand ensures that the chances of accepting a not truly needy applicant remain below the same figure.

Straightforward as these rules may seem their consequences may be intolerable to many planners and decision makers. In our case the .05 alpha-level decision rule would admit people with assets not exceeding $840. Anyone else would be rejected. It is clear from Figure 2 that some applicants who are not truly needy would be accepted into our program. In fact 75% of this group would be accepted. If each client in the program costs $1,200.00 then these people alone will cost our program over four million dollars for every ten thousand applicants.

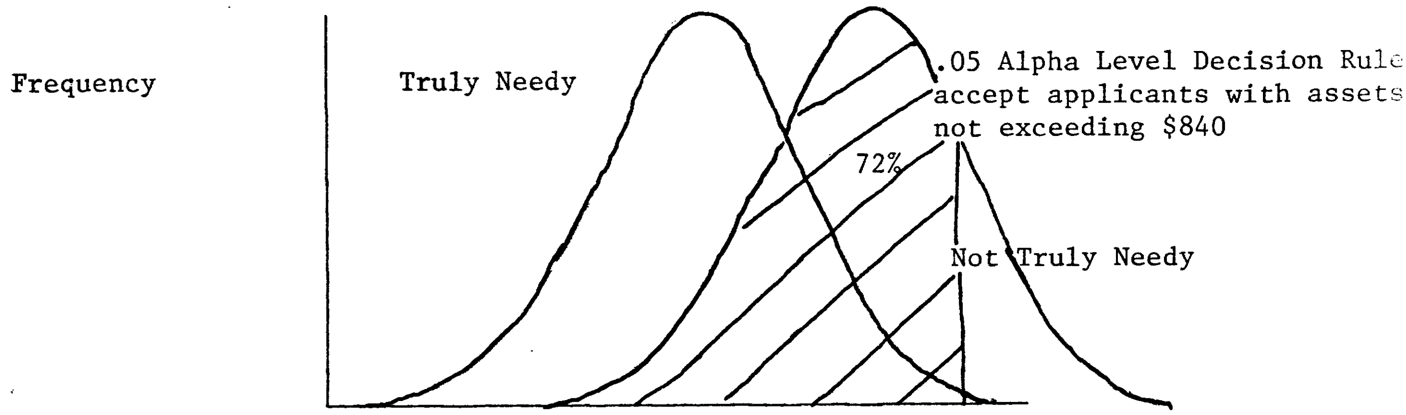

Figure 2

The .05 beta-level (Figure 3) rule will on the other hand prevent this situation. However the consequence of being so parsimonious is that nearly eighty truly needy applicants will be turned out in the cold for every one hundred applying. The costs of this policy when defined broadly, would no doubt be no less than those of the overly generous .05 alpha-level rule.

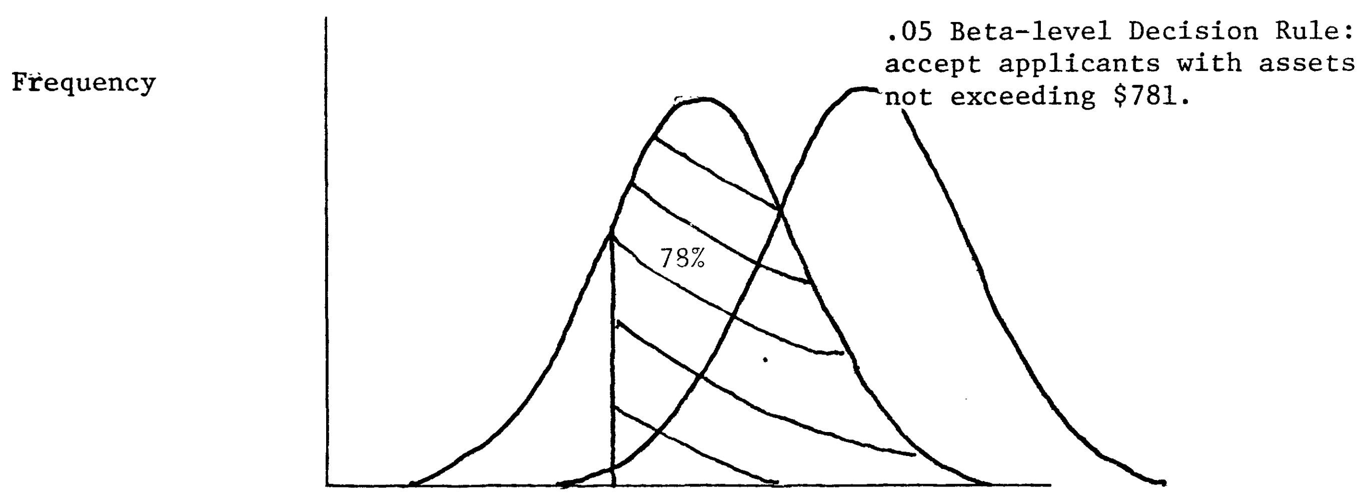

Figure 3

An obvious difficulty with classical statistical decision rules is that they ignore the cost consequences of the various possible outcomes.

Bayesian analysis allows the formulation of decision rules which incorporate the probabilities and costs of the various outcomes of a decision. The interactive program OPT2 assists evaluators in formulating decision rules for hypothesis tests involving Gaussian (normal) distributions. In order to apply

OPT2 it is necessary to have formulated an _a priori_ probability for the null hypothesis or in this case, the hypothesis that an applicant is truly needy.

Since we determined that half of the applicants are truly needy, the _a priori_ probability in this case is 0.5. This probability need however, not always be so objective. It is often necessary and prudent to incorporate more subjective information such as the opinions of experts or previous experience with related situations into one's estimate of the _a priori_ probability. PRIORS will assist a decision maker in this estimation.

In using PRIORS to estimate an _a priori_ probability, simply indicate as in Exhibit I, that you are testing an hypothesis. PRIORS will ask you for your best estimate of the _a priori_ probability and then inform you about some of the consequences of your estimate. If these consequences seem appropriate, you have validated your estimate. Otherwise you should change it.

EXHIBIT I

```
ARE YOU:
1. TESTING AN HYPOTHESIS?
2. ESTIMATING A PARAMETER?
3. UPDATING A PRIOR DISTRIBUTION?
4. NONE OF THE ABOVE
PLEASE TYPE THE NUMBER (1 - 4) OF THE APPROPRIATE OPTION.
:
.1

PLEASE FILL IN THE BLANK.
THE NULL OR NO-EFFECT HYPOTHESIS IS THAT... .an applicant is deserving

WHAT IS YOUR BEST ESTIMATE OF THE PROBABILITY THAT:
AN APPLICANT IS DESERVING
:
..4

THIS ESTIMATE INDICATES THAT YOU FEEL THE PROBABILITY OF NOT OBSERVING
THAT: AN APPLICANT IS DESERVING
EVEN ONCE IN FIVE TRIALS IS:0.07776
WHEREAS THE PROBABILITY OF OBSERVING THAT:
AN APPLICANT IS DESERVING
FIVE CONSECUTIVE TIMES IS:0.01024

WOULD YOU LIKE TO CHANGE YOUR ESTIMATE OF THE PROBABILITY THAT:
AN APPLICANT IS DESERVING?.yes

WHAT IS YOUR BEST ESTIMATE OF THE PROBABILITY THAT:
AN APPLICANT IS DESERVING
:
..5

THIS ESTIMATE INDICATES THAT YOU FEEL THE PROBABILITY OF NOT OBSERVING
THAT: AN APPLICANT IS DESERVING
EVEN ONCE IN FIVE TRIALS IS:0.03125
WHEREAS THE PROBABILITY OF OBSERVING THAT:
AN APPLICANT IS DESERVING
FIVE CONSECUTIVE TIMES IS:0.03125

WOULD YOU LIKE TO CHANGE YOUR ESTIMATE OF THE PROBABILITY THAT:
AN APPLICANT IS DESERVING?.no

THIS INDICATES THAT THE PRIOR PROBABILITY THAT:
AN APPLICANT IS DESERVING
IS:0.50000

WOULD YOU LIKE TO CONTINUE (YES OR NO)? .yes
```

3. Parameter Estimation

Many of the processes studied by evaluators can be accurately represented by underlying probability distributions and described by the parameters characterizing these distributions. Recall for instance the problem of determining the chances of getting Heads upon tossing a certain coin. The outcomes of the tosses can be viewed as a Bernoulli process with p the probability of getting Heads on any toss. Just as the problem of determining the probability of getting Heads on any toss can be reduced to finding the value of p in a Bernoulli process, the problem of describing many processes reduces to determining values for the parameters that describe them. In the following sections (4.1a - 4.1e) we discuss the common distributions addressed by PRIORS, when they arise, their conjugate prior distributions and how to use PRIORS to assess them.

3.1 The Bernoulli Process

A Bernoulli process is one in which there are two possible outcomes for any trial: event #1 and event #2. Event #1 occurs on any trial with fixed probability p (generally the quantity of interest), otherwise event #2 occurs. In addition, the outcome of any trial is unaffected by previous trials.

Tossing a coin is for example a Bernoulli process. If the coin is fair, p = 0.5 and Heads or Tails is equally likely to occur on any toss.

Bernoulli processes are common in evaluation settings. Opinion polls for example can often be viewed as Bernoulli processes where p is the fraction of people who would respond favorably. Generally, whenever an independently repeated experiment results in a dichotomy the outcomes can be viewed as a Bernoulli process.

As in Exhibit 2, PRIORS helps you assess your prior distribution to a Bernoulli process by first asking for your best estimate of p. Your response should be some number between zero and one, reflecting your estimate

EXHIBIT 2

ARE YOU:
1. TESTING AN HYPOTHESIS?
2. ESTIMATING A PARAMETER?
3. UPDATING A PRIOR DISTRIBUTION?
4. NONE OF THE ABOVE
PLEASE TYPE THE NUMBER (1 - 4) OF THE APPROPRIATE OPTION.
:
.2

CLASSICAL STATISTICS VIEWS PARAMETERS AS CONSTANTS WITH FIXED YET UNKNOWN VALUES. WE INTEND TO VIEW THEM AS RANDOM VARIABLES WITH PROBABILITY DISTRIBUTIONS. THE PRIOR DISTRIBUTION FOR THE PARAMETER SHOULD DEPEND ON THE DISTRIBUTION IT CHARACTERIZES.

THE PARAMETER YOU ARE TRYING TO ESTIMATE IS FROM:
1. A BERNOULLI PROCESS
2. A POISSON PROCESS
3. A UNIFORM PROCESS
4. AN INDEPENDENT NORMAL PROCESS
5. A NORMAL REGRESSION PROCESS
6. HELP
7. QUIT
PLEASE TYPE THE NUMBER (1 - 7) OF YOUR CHOICE.
:
.1

THE BERNOULLI PROCESS IS ONE IN WHICH THERE ARE TWO POSSIBLE EVENTS: EVENT#1 AND EVENT#2. EVENT#1 OCCURS ON ANY TRIAL WITH FIXED PROBABILITY P (THE PARAMETER WE ARE AFTER) AND EVENT#2 OCCURS WITH PROBABILITY 1-P. TRIALS OCCUR INDEPENDENTLY. THAT IS THE OUTCOME OF ONE TRIAL DOES NOT EFFECT THE OUTCOME OF OTHER TRIALS.

DOES THIS DESCRIBE YOUR PROCESS (YES OR NO)? .yes

PLEASE FILL IN THE BLANK.
EVENT#1 IS THE EVENT THAT... .an applicant is deserving

EVENT#1 IS THE EVENT THAT AN APPLICANT IS DESERVING

WHAT IS YOUR BEST ESTIMATE OF THE FRACTION OF ALL TRIALS FOR WHICH IT IS FOUND THAT AN APPLICANT IS DESERVING
:
..5

IN GENERAL THE MORE TRIALS OF A BERNOULLI PROCESS WE OBSERVE, THE MORE CONFIDENCE WE CAN HAVE IN OUR ESTIMATE OF THE PARAMETER P. WE MUST IN DETERMINING A PRIOR DISTRIBUTION FOR P, DECIDE HOW MUCH CONFIDENCE YOU HAVE IN YOUR EXPERIENCE.

SUPPOSE THAT NONE OF THE NEXT OBSERVATIONS IS THAT:
AN APPLICANT IS DESERVING
HOW MANY SUCH OBSERVATIONS WOULD IT TAKE TO CONVINCE YOU TO CHANGE YOUR ESTIMATE BY MORE THAN .1 ?
:
.50.

~~THIS~~ INDICATES THAT YOUR PRIOR DISTRIBUTION FOR P IS:
A BETA DISTRIBUTION WITH PARAMETERS 100.0000 AND 100.0000
THE MEAN OF THIS DISTRIBUTION IS:.500
THE VARIANCE OF THIS DISTRIBUTION IS: 0.0012
YOUR EQUIVALENT SAMPLE SIZE IS: 200

of the fraction of all trials resulting in Event #1. If your estimate is greater (less) than .5, PRIORS will next ask:

SUPPOSE THAT NONE (ALL) OF THE NEXT TRIALS IS (ARE) THAT:

event #1

HOW MANY SUCH OBSERVATIONS WOULD IT TAKE TO CONVINCE YOU TO CHANGE YOUR ESTIMATE BY MORE THAN .1?

Supposing your estimate is greater than .5.. we hope that with each successive occurrence of event #2 you would reduce your estimate of p. PRIORS is asking you to determine how many successive of occurrences of event #2 it would require to convince you to reduce your estimate of p by .1.

PRIORS will then present the prior distribution:

THIS INDICATES THAT YOUR PRIOR DISTRIBUTION FOR P IS:

A BETA DISTRIBUTION WITH PARAMETERS A AND B

THE MEAN OF THIS DISTRIBUTION IS: Mean

THE VARIANCE OF THIS DISTRIBUTION IS: Variance

YOUR EQUIVALENT SAMPLE SIZE IS: Equivalent sample size

The mean of the distribution represents your best estimate of p, the variance reflects your confidence in that estimate. Your equivalent sample size is a measure of the number of observations you feel your experience is equivalent to. Naturally, the more you know about the process, the larger your equivalent sample size should be.

3.2 The Poisson Process

A Poisson process is an arrival process in which the arrangement and number of arrivals in one time interval do not effect any non-overlapping time interval. Moreover, in a Poisson process arrivals come one at a time and the probability of an arrival in any short interval is proportional to the length of the interval.

Poisson processes arise often in evaluation settings. Crimes, disasters, customer requests, etc. can all be modeled as Poisson processes with the parameter representing the average rate of "arrivals". Consider for instance the problem of estimating the number of husband-wife disputes in a city each year. Since police records do not generally categorize incidents this way, a process evaluator might first ride with police officers, interview those who have previously called the police because of domestic eruptions and undertake other process-related activities. Then, that evaluator would be interviewed carefully to obtain a (personally derived) distribution for the annual rate of husband-wife disputes that require police intervention.

As in Exhibit 3 PRIORS in formulating a prior distribution to this Poisson process will first ask the evaluator to estimate the scope of his/her experience:

YOU JUDGE YOUR EXPERIENCE WITH THIS PROCESS TO BE EQUIVALENT
TO OBSERVING HOW MANY EVENTS OR ARRIVALS?

Obviously the longer and more detailed the process evaluation, the greater the number of observations the evaluators experience will be equivalent to. PRIORS next asks the evaluator for substantive information about the disputes:

WHAT IS YOUR BEST ESTIMATE OF THE AVERAGE TIME BETWEEN ARRIVALS?

It is hoped that during the process evaluation the evaluator developed some insight into the rate at which domestic disputes arise in the city. In answering this question the evaluator should use appropriate units be they minutes, days or years.

After the evaluator has answered all of the appropriate questions PRIORS will present his/her prior distribution as:

YOUR PRIOR DISTRIBUTION FOR THE ARRIVAL RATE IS A GAMMA
DISTRIBUTION WITH PARAMETER r
THIS DISTRIBUTION HAS BEEN MODIFIED BY THE AMOUNT OF
TIME YOU HAVE OBSERVED THIS PROCESS t
THE MEAN OF THE DISTRIBUTION IS: mean
THE VARIANCE IS: variance
YOUR EQUIVALENT SAMPLE SIZE IS: equivalent sample size

The mean represents the evaluators estimate of the arrival rate λ of domestic disputes in the city and the variance reflects his confidence in this estimate.

EXHIBIT 3

```
ARE YOU:
1. TESTING AN HYPOTHESIS?
2. ESTIMATING A PARAMETER?
3. UPDATING A PRIOR DISTRIBUTION?
4. NONE OF THE ABOVE
PLEASE TYPE THE NUMBER (1 - 4) OF THE APPROPRIATE OPTION.
:
.2


CLASSICAL STATISTICS VIEWS PARAMETERS AS CONSTANTS WITH FIXED YET UN-
KNOWN VALUES. WE INTEND TO VIEW THEM AS RANDOM VARIABLES WITH PROBABIL-
ITY DISTRIBUTIONS. THE PRIOR DISTRIBUTION FOR THE PARAMETER SHOULD
DEPEND ON THE DISTRIBUTION IT CHARACTERIZES.

THE PARAMETER YOU ARE TRYING TO ESTIMATE IS FROM:
1. A BERNOULLI PROCESS
2. A POISSON PROCESS
3. A UNIFORM PROCESS
4. AN INDEPENDENT NORMAL PROCESS
5. A NORMAL REGRESSION PROCESS
6. HELP
7. QUIT
PLEASE TYPE THE NUMBER (1 - 7) OF YOUR CHOICE.
:
.2


THE POISSON PROCESS CAN BE VIEWED AS AN ARRIVAL PROCESS IN WHICH:
1. THE ARRIVALS IN ONE PERIOD OF TIME DO NOT EFFECT THE ARRIVALS IN ANY
   NON-OVERLAPPING PERIOD OF TIME.
2. ARRIVALS COME ONE AT A TIME.
3. THE PROBABILITY OF AN ARRIVAL IN A SHORT INTERVAL IS PROPORTIONAL TO
   THE LENGTH OF THE INTERVAL.

DOES THIS DESCRIBE YOUR PROCESS (YES OR NO)? .yes


YOU JUDGE YOUR EXPERIENCE WITH THIS PROCESS TO BE EQUIVALENT TO
OBSERVING HOW MANY EVENTS OR ARRIVALS?
:
.75.


WHAT IS YOUR BEST ESTIMATE OF THE AVERAGE TIME BETWEEN ARRIVALS?
:
.5.0


YOUR PRIOR DISTRIBUTION FOR THE ARRIVAL RATE IS A GAMMA
DISTRIBUTION
WITH PARAMETER:   74.000
THIS DISTRIBUTION IS MODIFIED BY THE AMOUNT OF TIME YOU HAVE
OBSERVED THIS PROCESS:  375.000

THE MEAN OF THE DISTRIBUTION IS: 0.200000
THE VARIANCE IS: 0.000533
YOUR EQUIVALENT SAMPLE SIZE IS:   75.000
```

3.3. The Uniform Distribution

A Uniform or Rectangular process is one in which the value obtained on any trial is evenly distributed between a lower limit and an upper limit. We assume that the value of the lower limit is known and that we are trying to determine the value of the upper limit.

Suppose it was suspected that the time among parolees in a special parole program until recidivism is uniformly distributed between say one day after release and some unknown upper limit. Namely, if someone were released today on this parole program it is believed equally likely that he/she will be arrested tomorrow or any other day before the upper limit ie., given the value of the upper limit is U, the conditional probability that a parolee will recidivate at time t after release is uniformly distributed between L and U where L is known to be the earliest any parolee will recidivate.

In formulating a prior distribution to this uniform process PRIORS will (as in Exhibit 4) ask the evaluator to assess the extent of his/her knowledge:

```
YOU JUDGE YOUR EXPERIENCE WITH THIS PROCESS TO BE EQUIVALENT
TO OBSERVING HOW MANY EVENTS?
```

In this case it is clear that an event is a recidivation and the more the evaluator knows about the program and parolees in general, the larger his/her answer should be. PRIORS will then ask the evaluator to provide a best lower bound to the upper limit of the uniform process:

```
TO YOUR KNOWLEDGE THE LARGEST POSSIBLE VALUE OF ANY TRIAL
FROM THIS PROCESS IS CERTAINLY NO SMALLER THAN WHAT NUMBER?
```

After supplying PRIORS with an upper and lower bound to the possible values of trials from the process his/her prior distribution will appear as:

```
YOUR PRIOR DISTRIBUTION FOR THE UPPER LIMIT OF THIS RECTANGULAR
PROCESS IS A HYPERBOLIC DISTRIBUTION WITH PARAMETER  n
THIS DISTRIBUTION IS DEFINED FOR VALUES GREATER THAN   u
THE MEAN OF THIS DISTRIBUTION IS:   mean
THE VARIANCE IS:   variance
```

Here n represents the number of outcomes observed and u the largest among these. The mean reflects the expected value of the upper limit and the variance indicates our confidence in this estimate. Notice that unlike

THE UNIFORM OR RECTANGULAR PROCESS IS ONE IN WHICH THE VALUE OBTAINED ON ANY TRIAL IS EVENLY DISTRIBUTED BETWEEN A LOWER AND AN UPPER LIMIT. WE ASSUME THAT THE VALUE OF THE LOWER LIMIT IS KNOWN AND THAT WE ARE TRYING TO DETERMINE THE VALUE OF THE UPPER LIMIT. IF YOUR CASE IS JUST THE OPPOSITE THEN SIMPLY REVERSE THE AXIS AGAINST WHICH YOU ARE MEASURING.

EXHIBIT 4

DOES THIS DESCRIBE YOUR PROCESS (YES OR NO)? .yes

YOU JUDGE YOUR EXPERIENCE WITH THIS PROCESS TO BE EQUIVALENT TO OBSERVING HOW MANY EVENTS?

:

.26.

TO YOUR KNOWLEDGE THE LARGEST POSSIBLE VALUE OF ANY TRIAL FROM THIS PROCESS IS CERTAINLY NO SMALLER THAN WHAT NUMBER?

:

.10.

WHAT IS THE SMALLEST VALUE OBSERVATIONS FROM THIS PROCESS CAN EXHIBIT?

:

.0.0

YOUR PRIOR DISTRIBUTION FOR THE UPPER LIMIT OF THIS RECTANGULAR PROCESS IS A HYPERBOLIC DISTRIBUTION WITH PARAMETER: 26
THIS DISTRIBUTION IS DEFINED FOR VALUES GREATER THAN: 10.0000
THE MEAN OF THIS DISTRIBUTION IS: 10.4167
THE VARIANCE IS: 0.1887

other prior distributions the mean is not the evaluators estimate of the upper limit. This is due to the fact that we do not want to over-estimate the upper limit. If our initial estimate is too large, no amount of additional information will correct this. For this reason the evaluator is asked to give a lower bound to the upper limit and not to give an estimate thereof.

3.4 The Normal Process With Independent Samples

An independent normal process is one in which the value of each outcome is selected from a normal or Gaussian distribution. We say the process is independent if the value of each outcome has no effect on any other outcome. PRIORS assumes that the evaluator is trying to formulate prior distributions for the mean and the variance of the underlying normal distribution.

The independent normal is in many areas of evaluation the most common process. Many traits are distributed approximately normally in populations. Height, reading ability and foot-size are for example often approximately normally distributed in human populations. The size of errors in many measurements is also often normally distributed. Moreover, it is often found that if a trait is not normally distributed in a population, stratifying the population leads to normal distributions within each stratum. However it is unfortunately tempting to classify processes rashly as normal. Generally for example such traits as age, income, etc., are not normally distributed within heterogeneous populations.

Suppose that an evaluator is studying a reading program and knows that the reading ability among enrolled students is approximately normally distributed. This knowledge alone clearly reflects relevant prior information. Moreover, the evaluator has some knowledge about the enrolled students' backgrounds as well as knowing how similar programs have performed in the past. This fundamental expertise combined with such process-related activities as sitting in on classes, interviewing students, teachers and administrators, etc. should provide the evaluator with valuable information about the reading ability of students in the program. PRIORS will help assess this prior distribution by first asking the evaluator to estimate the scope of his/her experience:

> YOU JUDGE YOUR EXPERIENCE EQUIVALENT TO OBSERVING HOW MANY OUTCOMES FROM THIS PROCESS?

In this case it is clear that the evaluator should equate his/her experience with knowing the reading ability of some number of enrolled students. The The more he/she knows about the program, the greater this number should be. PRIORS will then ask the evaluator to simulate a normal sample:

> PLEASE TYPE THE VALUES OF OUTCOMES YOU WOULD EXPECT TO OBSERVE FROM THIS PROCESS ONE PER LINE. THERE SHOULD BE AS MANY VALUES AS YOUR ANSWER TO THE LAST QUESTION. TYPE 'DONE' WHEN YOU ARE THROUGH.

The evaluator's response should reflect not only his/her knowledge about the average reading ability, but also about the variation among students. Suppose for example the evaluator estimated his/her experience equivalent to five observations. His/her response to the question about expected observations should consist of five values reflecting both the average reading ability and the degree of difference among students. An answer for example like:

```
.75
.75
.75
.75
.75
'done'
```

is highly unlikely -- not everyone has the same reading ability.

Something like:

```
.75
.60
.75
.80
.85
'done'
```

is more likely. This sample suggests, as exhibit 5 shows, that the evaluator believes the average reading level to be .75 and the variance to be small --

THE INDEPENDENT NORMAL PROCESS IS ONE IN WHICH THE VALUE OF EACH OUTCOME IS SELECTED FROM A NORMAL DISTRIBUTION. THE VALUE OF ONE OUTCOME HAS NO EFFECT ON THE VALUE OF ANY OTHER OUTCOME THE MEAN AND VARIANCE ARE THE UNKNOWN PARAMETERS WE ARE TRYING TO ESTIMATE

EXHIBIT 5

DOES THIS DESCRIBE YOUR PROCESS (YES OR NO)? .yes

YOU JUDGE YOUR EXPERIENCE EQUIVALENT TO OBSERVING HOW MANY OUTCOMES FROM THIS PROCESS?

:
.5.

PLEASE TYPE THE VALUES OF OUTCOMES YOU WOULD EXPECT TO OBSERVE FORM THIS PROCESS, ONE PER LINE.
BE SURE TO USE DECIMALS!
TYPE 'DONE' WHEN YOU ARE THROUGH.

:
.0.75

:
.0.60

:
.0.75

:
.0.80

:
.0.85

:
.done

YOUR MARGINAL PRIOR DISTRIBUTION FOR THE MEAN OF THIS INDEPENDENT NORMAL PROCESS IS A STUDENT'S DISTRIBUTION WITH 4.0000 DEGREES OF FREEDOM.
THIS DISTRIBUTION HAS BEEN MODIFIED TO HAVE MEAN: 0.7500
AND VARIANCE: 0.0035

YOUR MARGINAL PRIOR DISTRIBUTION FOR THE VARIANCE OF THIS INDEPENDENT NORMAL PROCESS IS A GAMMA DISTRIBUTION WITH PARAMETER:
1.0000
THIS DISTRIBUTION HAS BEEN MODIFIED TO HAVE MEAN: 114.2746
THE VARIANCE IS:6529.3477

around .01. We can expect on the basis of this information that the evaluator knows most of the students perform in the 0.45 to 1.0 range.

PRIORS will present prior distributions for the mean or average and the variance as:

YOUR MARGINAL PRIOR DISTRIBUTION FOR THE MEAN OF THIS INDEPENDENT
NORMAL PROCESS IS A STUDENT'S DISTRIBUTION WITH r DEGREES OF FREEDOM.
THIS DISTRIBUTION HAS BEEN MODIFIED TO HAVE MEAN: mean
AND VARIANCE: variance

YOUR MARGINAL PRIOR DISTRIBUTION FOR THE VARIANCE OF THIS INDEPENDENT
NORMAL PROCESS IS A GAMMA DISTRIBUTION WITH PARAMETER p
THIS DISTRIBUTION HAS BEEN MODIFIED TO HAVE MEAN: mean
THE VARIANCE IS: variance

Again the mean of the student's distribution reflects the evaluators estimate of the average reading level and the variance, his confidence in that estimate. The mean of the gamma distribution represents the inverse of the evaluators estimate of the variance for the underlying normal distribution.

3.5 Normal Regression

In normal regression we are trying to predict or estimate the values of some dependent random variable Y as a function of the variables X. In this model we assume that the Y-values are normally distributed with unknown variance and mean equal to some linear function of the X's. We are trying to estimate the variance of Y and the function defining its mean.

Normal regression is common in evaluations since determining the value of the mean of a parameter as a function of other parameters tells us how they effect each other. The rate at which substances cause cancer can for example be modeled as a regression problem. Suppose we are trying to determine the relationship between the heights of parents and that of their children. We might suspect that the height of children, Y, is a linear function of the height of their fathers, X, and the height of their mothers, Z, ie that:

$$Y = AX + BZ + C$$

We are assuming here that height is normally distributed. The problem now reduces to estimating A,B,C and the variance of Y.

As usual we assume some prior knowledge about the relation among heights. In formulating prior distributions for the vector (A,B,C) and the variance of Y, PRIORS will, as in Exhibit 6, first ask how many components are in the vector:

```
YOU ARE TRYING TO ESTIMATE THE MEAN OF Y AS A LINEAR FUNCTION OF
HOW MANY INDEPENDENT VARIABLES?
```

In our case this will be three; father's height, mother's height and other factors or (A,B,C). If however we had included say grandparents height this would be correspondingly larger. Next PRIORS asks us to assess the extent of our experience with the relationship:

```
YOU JUDGE YOUR EXPERIENCE EQUIVALENT TO MAKING HOW MANY OBSERVATIONS?
```

Clearly the more closely we have studied it the larger our answer should be.

Finally, as in the normal process we must simulate observations:

```
PLEASE TYPE IN THE VALUES OF OBSERVATIONS YOU WOULD EXPECT FROM THIS
PROCESS. FOR THE ITH OBSERVATION THE VALUE OF Y(I) IS THE FIRST
ENTRY FOLLOWED BY THE X(I,J)-VALUES. LEAVE A SPACE BETWEEN EACH
ENTRY. EACH OBSERVATION SHOULD START A NEW LINE. THERE SHOULD BE
AS MANY OBSERVATIONS AS YOUR ANSWER TO THE LAST QUESTION.
```

Here too a response like:

```
 Y(J)   X(I,J)
5.7 6.0 5.5 1
5.7 6.0 5.5 1
5.7 6.0 5.5 1
```

for three observations is highly unlikely -- not everyone is the same height. Supposed we assessed our experience equal to five observations and responded with the observations:

```
 Y(I)  X(I,J)
5.7 6.0 5.5 1
6.0 6.1 5.2 1
5.2 6.1 5.5 1
5.9 5.8 5.4 1
5.0 5.2 5.5 1
```

This would reflect more accurately our experience in that for example a man 6.1 ft is likely to have a wife 5.2 ft and a son 6.0 ft or a wife 5.5 ft and a son 5.2 ft. Your answer should reflect your knowledge of the variation within the populations as well as the relations among them. Should you

EXHIBIT 6

```
THE NORMAL REGRESSION PROCESS ASSUMES WE ARE TRYING TO PREDICT OR
ESTIMATE THE VALUES OF SOME DEPENDENT RANDOM VARIABLE, Y, AS A LINEAR
FUNCTION OF THE INDEPENDENT VARIABLES, X(.,J). IN THIS MODEL WE
ASSUME THAT THE Y(J)-VALUES ARE NORMALLY DISTRIBUTED WITH UNKNOWN
VARIANCE AND MEAN EQUAL TO SOME LINEAR FUNCTION OF THE X(.,J)-
VALUES. WE ARE TRYING TO ESTIMATE THE VARIANCE OF Y AND THE SLOPE
OF THE LINE.

DOES THIS DESCRIBE YOUR PROCESS (YES OR NO)? .yes


YOU ARE TRYING TO ESTIMATE THE MEAN OF Y AS A LINEAR FUNCTION OF HOW
MANY INDEPENDENT VARIABLES?
:
.3.


YOU JUDGE YOUR EXPERIENCE EQUIVALENT TO MAKING HOW MANY OBSERVATIONS?
:
.5.


PLEASE TYPE IN THE VALUES OF OBSERVATIONS  YOU WOULD EXPECT FROM THIS
PROCESS.
Y(I) AS THE FIRST ENTRY IN ROW I FOLLOWED BY THE X(I,J)-VALUES.
LEAVE A SPACE BETWEEN EACH ENTRY. BE SURE TO USE DECIMAL POINTS.
WAIT FOR THE ':' PROMPT.

   Y(I)    X(I,J)-VALUES
:
.5.7 6.0 5.5 1.0

:
.6.0 6.1 5.2 1.0

:
.5.2 6.1 5.5 1.0

:
.5.9 5.8 5.4 1.0

:
.5.0 5.2 5.5 1.0


THIS DATA HAS BEEN READ AS:
   Y(I)        X(I,J)-VALUES
   5.7000      6.0000      5.5000      1.0000
   6.0000      6.1000      5.2000      1.0000
   5.2000      6.1000      5.5000      1.0000
   5.9000      5.8000      5.4000      1.0000
   5.0000      5.2000      5.5000      1.0000

IS THIS CORRECT? .yes


YOUR MARGINAL PRIOR DISTRIBUTION FOR THE SLOPE OFTHE LINE
IS A  3 DIMENSIONAL STUDENT'S DISTRIBUTION WITH  2
DEGREES OF FREEDOM.
THE MEAN OF THIS DISTRIBUTION IS:
   0.4224      -1.9804      13.8268
IT HAS NO PROPER VARIANCE.
THE CHARACTERISTIC MATRIX OF THIS PRIOR DISTRIBUTION IS:
  171.1000    158.1900    29.2000
  158.1900    146.9500    27.1000
   29.2000     27.1000     5.0000

YOUR MARGINAL PRIOR DISTRIBUTION FOR THE VARIANCE OF THE Y'S IS
A GAMMA DISTRIBUTION WITH PARAMETER:   0.0000
THE MEAN OF THIS DISTRIBUTION IS:   7.1612
THE VARIANCE IS:  51.2835
```

make a mistake here, PRIORS will give you the chance to correct it when you are through.

Given this information PRIORS will present your prior distributions for the vector (A,B,C) and the variance of Y as:

```
YOUR MARGINAL PRIOR DISTRIBUTION FOR THE COEFFICIENTS OF THE
X(I,J)-VALUES IS A  n  DIMENSIONAL STUDENT'S DISTRIBUTION
WITH r DEGREES OF FREEDOM.
THE MEAN OF THIS DISTRIBUTION IS:
    mean vector
THE COVARIANCE MATRIX IS:
    covariance matrix
THE CHARACTERISTIC MATRIX OF THIS PRIOR DISTRIBUTION IS:
    characteristic matric

YOUR MARGINAL PRIOR DISTRIBUTION FOR THE VARIANCE OF THE Y's
IS GAMMA DISTRIBUTION WITH PARAMETER:  p
THE MEAN OF THIS DISTRIBUTION IS:  mean
THE VARIANCE IS:  variance
```

The mean vector of the Student's distribution represents the evaluators estimate in this case of the values (A,B,C) and the variance reflects his/her confidence in that estimate. The characteristic matrix is useful for updating the distribution.

The mean of the gamma distribution is the inverse of the evaluator's estimate of the variance of Y.

5. Posterior Distributions and Updating

The beauty of the prior distributions we formulate with PRIORS is that they readily allow the addition of improved information. We call this process of adding information to a prior distribution "updating". The resulting updated distribution is a "posterior distribution". As we mentioned before the prior distributions formulated with PRIORS are conjugates - that is the posterior is from the same family as the prior. In fact should the evaluator choose to add additional new information, he should treat the posterior exactly as a prior.

To update a prior distribution with PRIORS you must have:

1. formulated a prior distribution with PRIORS and have the description of the distribution on hand.

2. obtained further statistical information about the process.

PRIORS will proceed by asking you about your present prior distribution, then about the additional statistical information. Simply answer the questions and PRIORS will supply you with a description of your posterior distribution. In Exhibit 7 our original prior distribution was:

```
a gamma distribution
with parameter: 74.000.


The distribution has been modified by the amount of time we had
observed the process:  375.000
The mean of the distribution was:  0.2000.
The varience was:  0.000533
Our equivalent sample size was:  75.000
```

Since formulating our prior distribution we observed 25 arrivals with average interarrival time 4.1. Note that after updating a prior we obtain a posterior distribution however should we wish to update again, this posterior would become our present prior distribution.

EXHIBIT 7

```
ARE YOU:
1. TESTING AN HYPOTHESIS?
2. ESTIMATING A PARAMETER?
3. UPDATING A PRIOR DISTRIBUTION?
4. NONE OF THE ABOVE
PLEASE TYPE THE NUMBER (1 - 4) OF THE APPROPRIATE OPTION.
:
.3


THE BEAUTY OF THE PRIOR DISTRIBUTIONS WE FORMULATE WITH THIS PROGRAM
IS THAT THEY READILY ALLOW THE ADDITION OF IMPROVED INFORMATION. WE
CALL THIS PROCESS OF ADDING INFORMATION TO AN ALREADY FORMED PRIOR
'UPDATING'. TO DO THIS WE ASSUME YOU HAVE ALREADY FORMULATED A PRIOR
DISTRIBUTION USING THIS PROGRAM AND THAT SINCE THAT TIME YOU HAVE MADE
ADDITIONAL OBSERVATIONS OF THE PROCESS. IS THIS THE CASE? .yes


WE ASSUME FURTHER THAT YOUR PRIOR DISTRIBUTION IS FOR THE PARAMETER(S)
OF ONE OF THE FOLLOWING PROCESSES:
1. A BERNOULLI PROCESS.
2. A POISSON PROCESS.
3. A UNIFORM PROCESS.
4. AN INDEPENDENT NORMAL PROCESS.
5. A NORMAL REGRESSION PROCESS.
6. NONE OF THE ABOVE
PLEASE TYPE THE NUMBER (1 - 6) OF YOUR PROCESS.
:
.2


WHAT IS THE EQUIVALENT SAMPLE SIZE OF PRESENT PRIOR?
:
.75.0


WHAT IS THE MEAN OF YOUR PRESENT PRIOR DISTRIBUTION?
:
.0.2


SINCE FORMULATING YOUR PRIOR DISTRIBUTION HOW MANY ARRIVALS HAVE YOU
OBSERVED?
:
.25.


WHAT IS YOUR BEST ESTIMATE OF THE AVERAGE INTERARRIVAL TIME FOR THESE
LAST OBSERVATIONS?
:
.4.1


YOUR POSTERIOR DISTRIBUTION FOR THE ARRIVAL RATE IS A GAMMA
DISTRIBUTION
WITH PARAMETER:   99.000
THIS DISTRIBUTION IS MODIFIED BY THE AMOUNT OF TIME YOU HAVE
OBSERVED THIS PROCESS:  477.500

THE MEAN OF THE DISTRIBUTION IS: 0.209424
THE VARIANCE IS: 0.000439
YOUR EQUIVALENT SAMPLE SIZE IS:  100.000

WOULD YOU LIKE TO SEE A PLOT OF YOUR CUMULATIVE POSTERIOR
DISTRIBUTION?.no


WOULD YOU LIKE TO MODIFY THIS DISTRIBUTION (YES OR NO)? .no


THE POSTERIOR DISTRIBUTIONS YOU HAVE FORMULATED ARE NOW YOUR PRESENT
PRIOR DISTRIBUTIONS! TO UPDATE THESE DISTRIBUTIONS SIMPLY TREAT THEM
AS PRIOR DISTRIBUTIONS.
```

5.1 Plots of Cumulative Distributions

After describing your prior distribution(s), PRIORS will ask if you would like to see a plot of your cumulative prior distribution. Should you respond "yes" (or "y") to this question, PRIORS will produce a point plot of the probability the parameter in question will be less than the independent variable. If you do not wish to see this plot type "no" (or "n").

In Exhibit 8 the independent variable ranges from zero to XMAX = 1.0 and each unit is scale unit = .1 . Whereas the ordinate or y-axis ranges from zero to 1.1 and each unit is .01. The probability that the parameter is less than 0.5 is about .02 and the probability it is less than 0.9 is about 1.0.

Note plots will not be produced for multidimensional distributions.

5.2 Modifying A Distribution

After formulating a prior distribution you may feel it is not exactly what you want. Should this be the case simply respond "yes" (or "y") to the question:

Would you like to modify this distribution (yes or no)?

As in Exhibit 9 PRIORS will ask you whether you would like to change various parameters. Simply answer the questions appropriately and PRIORS will produce a new prior. If you ask to modify a posterior distribution, i.e. if you modify a distribution immediately after updating it, you will be modifying the entire distribution - i.e. not your previous prior nor the additional information, but the updated distribution itself.

26

WOULD YOU LIKE TO SEE A PLOT OF YOUR CUMULATIVE PRIOR DISTRIBUTION?
(YES OR NO)? YES

XMIN	XMAX	YMIN	YMAX
5.0000E-02	1.0000E+00	8.8151E-12	1.0560E+00

ORDINATE (Y AXIS) 0 = 0.0 * $(10)^{0}$,100 ON SCALE $(10)^{0}$

ABCISSA (X-AXIS) 0 = 0.0 * $(10)^{-1}$, SCALE UNIT = $(10)^{-1}$

EXHIBIT 8

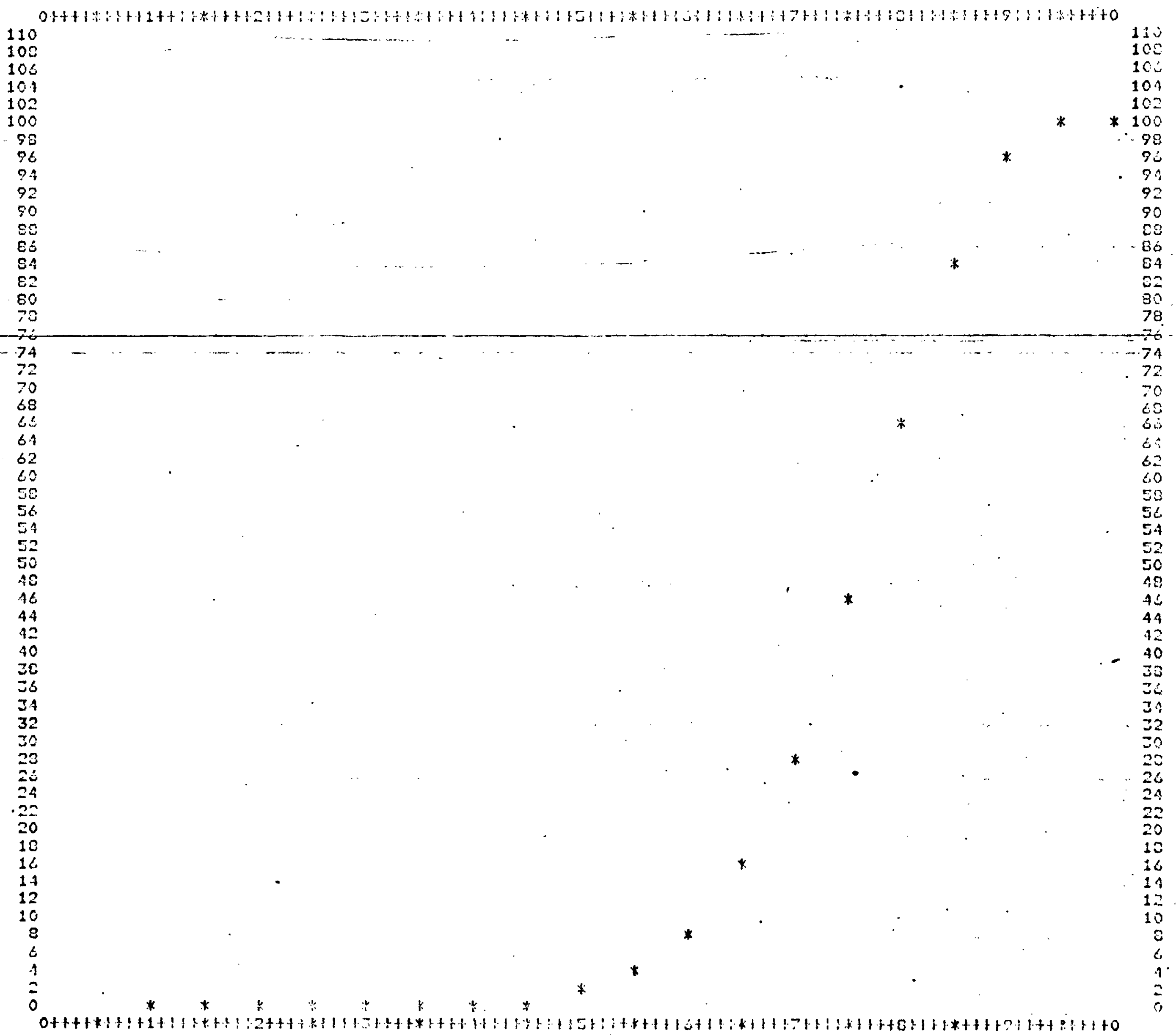

EXHIBIT 9

```
THE POISSON PROCESS CAN BE VIEWED AS AN ARRIVAL PROCESS IN WHICH:
1. THE ARRIVALS IN ONE PERIOD OF TIME DO NOT EFFECT THE ARRIVALS IN ANY
   NON-OVERLAPPING PERIOD OF TIME.
2. ARRIVALS COME ONE AT A TIME.
3. THE PROBABILITY OF AN ARRIVAL IN A SHORT INTERVAL IS PROPORTIONAL TO
   THE LENGTH OF THE INTERVAL.

DOES THIS DESCRIBE YOUR PROCESS (YES OR NO)? .yes

YOU JUDGE YOUR EXPERIENCE WITH THIS PROCESS TO BE EQUIVALENT TO
OBSERVING HOW MANY EVENTS OR ARRIVALS?
:
.100.

WHAT IS YOUR BEST ESTIMATE OF THE AVERAGE TIME BETWEEN ARRIVALS?
:
.4.775

YOUR PRIOR DISTRIBUTION FOR THE ARRIVAL RATE IS A GAMMA
DISTRIBUTION
WITH PARAMETER:   99.000
THIS DISTRIBUTION IS MODIFIED BY THE AMOUNT OF TIME YOU HAVE
OBSERVED THIS PROCESS:  477.500

THE MEAN OF THE DISTRIBUTION IS: 0.209424
THE VARIANCE IS: 0.000439
YOUR EQUIVALENT SAMPLE SIZE IS:  100.000

WOULD YOU LIKE TO SEE A PLOT OF YOUR CUMULATIVE PRIOR
DISTRIBUTION?.no

WOULD YOU LIKE TO MODIFY THIS DISTRIBUTION (YES OR NO)? .yes

WOULD YOU LIKE TO CHANGE THE NUMBER OF TRIALS YOU HAVE SEEN?
YES OR NO)? .yes

HOW MANY ARRIVALS HAVE YOU SEEN?
:
.110.

WOULD YOU LIKE TO CHANGE THE AVERAGE TIME BETWEEN ARRIVALS (YES OR NO)? .yes

WHAT IS THE AVERAGE TIME BETWEEN ARRIVALS?
:
.4.8

YOUR PRIOR DISTRIBUTION FOR THE ARRIVAL RATE IS A GAMMA
DISTRIBUTION
WITH PARAMETER:  109.000
THIS DISTRIBUTION IS MODIFIED BY THE AMOUNT OF TIME YOU HAVE
OBSERVED THIS PROCESS:  528.000

THE MEAN OF THE DISTRIBUTION IS: 0.208333
THE VARIANCE IS: 0.000395
YOUR EQUIVALENT SAMPLE SIZE IS:  110.000

WOULD YOU LIKE TO SEE A PLOT OF YOUR CUMULATIVE PRIOR
DISTRIBUTION?.no
```

APPENDIX I

The Distributions

The Bernoulli Process

Has distribution: Probability of event #1 = p
Probability of event #2 = 1-p

Prior distribution: H Beta distribution with parameters a and b

$$\beta_{a,b}(p) = \frac{(a+b-1)!}{(a-1)!(b-1)!} p^{a-1}(1-p)^{b-1} \text{ for } 0 \leq p \leq 1$$

a corresponds to the number of times event #1 was observed
b corresponds to the number of times event #2 was observed
a + b is the equivalent sample size.

Posterior distribution

$$B_{a+a',b+b'}(p) = \frac{(a+a'+b+b'-1)!}{(a+a'-1)!(b+b'-1)!} p^{a+a'-1}(1-p)^{b+b'-1}$$

for $0 \leq p \leq 1$

a' corresponds to the number of times event #1 was observed since formulating prior
b' corresponds to the number of times event #2 was observed since formulating prior
a' + b' is actual sample size since formulating prior

The Poisson Process

Has distribution:

$$P_{\lambda,T}(k) = \frac{(\lambda T)^k e^{-\lambda T}}{k!} \quad \text{for } k = 0,1,\ldots.$$

where λ is average arrival rate
and T is the elapsed time.

Prior distribution: A Gamma distribution with parameter r modified by t

$$G_{r,t}(\lambda) = \frac{e^{-\lambda t}(\lambda t)^r}{r!}$$

where r is the number of events observed
and t is the length of time observing the process.

Posterior distribution:

$$G_{r+r',t+t'}(\lambda) = \frac{e^{-\lambda(t+t')}(\lambda(t+t'))^{r+r'}}{(r+r')!}$$

r' corresponds to the number of additional observations in t' additional time units.

The Uniform Process

Has distribution:

$$P_{U,L}(t) = \frac{1}{U-L} \quad \text{for } L < t < U$$

where L is lower limit and U is upper limit of process

Prior distribution: A Hyperbolic distribution with parameter n defined for values greater than ν

$$H_{n,\nu,L}(t) = (n - 1) \frac{(\nu - L)^{n-1}}{(t - L)^{n}} \quad \text{for } t > \nu$$

where n is the number of observations and ν is the largest value observed.

Posterior distribution:

$$H_{n+n',\nu',L}(t) = (n+n' - 1) \frac{(\nu' - L)^{n+n'-1}}{(t - L)^{n+n'}}$$

where n' is the number of additional observations ν' is the largest value observed in n+n' trials.

The Normal Process

Has distribution:

$$P_{u,\sigma^2}(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(t-u)^2} \quad \text{for } -\infty < t < \infty$$

where u is the mean and σ^2 is the variance

Marginal Prior Distribution for the mean: A Student's distribution with r degrees of freedom.

$$S_r(t) = \frac{\Gamma(\frac{n}{2})}{\sqrt{\pi}\,\Gamma(r/2)} \; r^{\frac{r}{2}} \left(r + \frac{n}{s}(t-u)^2\right)^{-\frac{n}{2}} \sqrt{n/2}$$

where n is the number of observations r+1 u is their mean and s is their variance.

Posterior distribution:

$$S_{r+n'}(t) = \frac{\Gamma(\frac{n+n'}{2})}{\sqrt{\pi}\;\Gamma(\frac{r+n'}{2})} \; (r+n')^{\frac{r+n'}{2}} \left(r+n' + \frac{n+n'}{s''}(t-u'')^2\right)^{-\frac{n+n'}{2}} \sqrt{\frac{n+n'}{2}}$$

where n' is the number of subsequent observations $u'' = (n'u'+nu)/n+n'$ and $s'' = [(n'-1)s' + n'u'^{2} - + rs - nu^2 - (n+n')u''^2]/r+n'$ u' is the mean of the subsequent observations and s' is their variance.

Marginal Prior Distribution for the variance: A Gamma Distribution with parameter p.

$$G_p(t) = e^{-(p+1)st} \; \frac{((p+1)st)^p \; (p+1)s}{p!}$$

p is $\frac{n-3}{2}$ and S is the sample variance

Posterior Distribution

$$G_{p+\frac{n'}{2}}(t) = e^{-(p+\frac{n'}{2}+1)s''t} \; \frac{((p+\frac{n'}{2}+1)s''t)^{p+\frac{n'}{2}}}{(p+\frac{n'}{2})!} (p+\frac{n'}{2}+1)s''$$

the mean of the distribution is $\frac{1}{\hat{\sigma}^2}$ where $\hat{\sigma}^2$ is our estimate of the variance of the normal process.

Normal regression

Has distribution:

$$P_{xi}(t_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-1/2\sigma^2(t_i - \Sigma B_j X_{ij})^2}$$

where $\Sigma B_u X_{ij}$ is the expected value of ti

Prior distribution for B: An n dimensional Student's Distribution with r degrees of freedom.

$$S_{n,r}(t) = \frac{r^{r/2}\,\Gamma(\frac{r+n}{2} - 1)}{\pi^{n/2}\,\Gamma(\frac{r}{2} - 1)} \left(r + (t-u)\left(\frac{r-2}{r}\right) c^{-1} (t-u)^t\right)^{-\left(\frac{r+n}{2}\right)} \left|\frac{r-2}{r} c^{-1}\right|^{1/2}$$

where u is the mean and c is the co-variance matrix.

Posterior distribution:

$$S_{n,r+m'}(t) = \frac{(r+m')^{\frac{r+m'}{2}}\,\Gamma(\frac{r+m'+n}{2} - 1)}{\pi^{n/2}\,\Gamma(\frac{r+m'}{2} - 1)} \left(r+m'+(t-u'')\left(\frac{r+m'-2}{r+m'}\right) c''^{-1} (t-u'')^t\right)^{-\left(\frac{r+m'+n}{2}\right)} \times$$

$$\left|\frac{r+m'-2}{r+m'} c''^{-1}\right|^{1/2}$$

where m' is the number of subsequent observations

$$c''^{-1} = [c'\nu + c\nu]^{-1}/\nu''$$

Marginal Prior Distribution for the variance: A Gamma Distribition with parameter p.

$$G_p(t) = e^{-(p+1)\nu t} \frac{(p+1)\nu t)^p}{p!} (p+1)\nu$$

where p is 1/2 r -1

Posterior Distribution:

$$G_{p + \frac{m'}{2}}(t) = e^{-(p + \frac{m'}{2} + 1)\nu'' t} \frac{((p + \frac{m'}{2} + 1)\nu'' t)^{p + \frac{m'}{2}}}{(p+\frac{m'}{2})!} (p + \frac{m'}{2} + 1)\nu''$$

The mean of this distribution is the inverse of our estimate of the variance.