

Constrained Noninformative Priors

Corwin L. Atwood

Published October 1994

Idaho National Engineering Laboratory
Idaho Falls, Idaho 83415

MASTER

Prepared for the
U.S. Nuclear Regulatory Commission
Office for Analysis and Evaluation of Operational Data
Reliability and Risk Analysis Branch
Under DOE Idaho Operations Office
Contract DE-AC07-94ID13223

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED
DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

etc

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, make any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

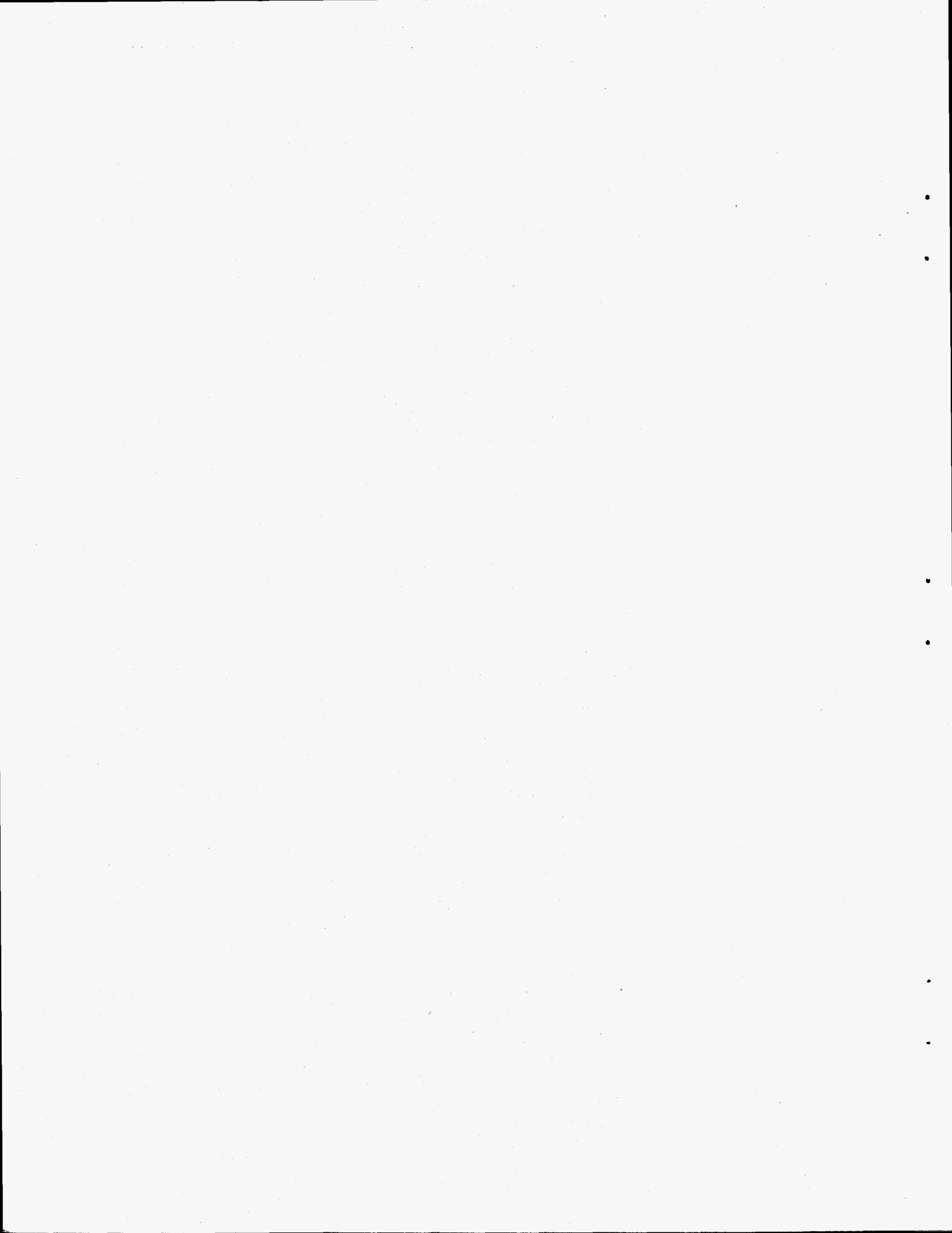
DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

ABSTRACT

The Jeffreys noninformative prior distribution for a single unknown parameter is the distribution corresponding to a uniform distribution in the transformed model where the unknown parameter is approximately a location parameter. To obtain a prior distribution with a specified mean but with diffusion reflecting great uncertainty, a natural generalization of the noninformative prior is the distribution corresponding to the constrained maximum entropy distribution in the transformed model. Examples are given.

Key words: Jeffreys prior, risk assessment



SUMMARY

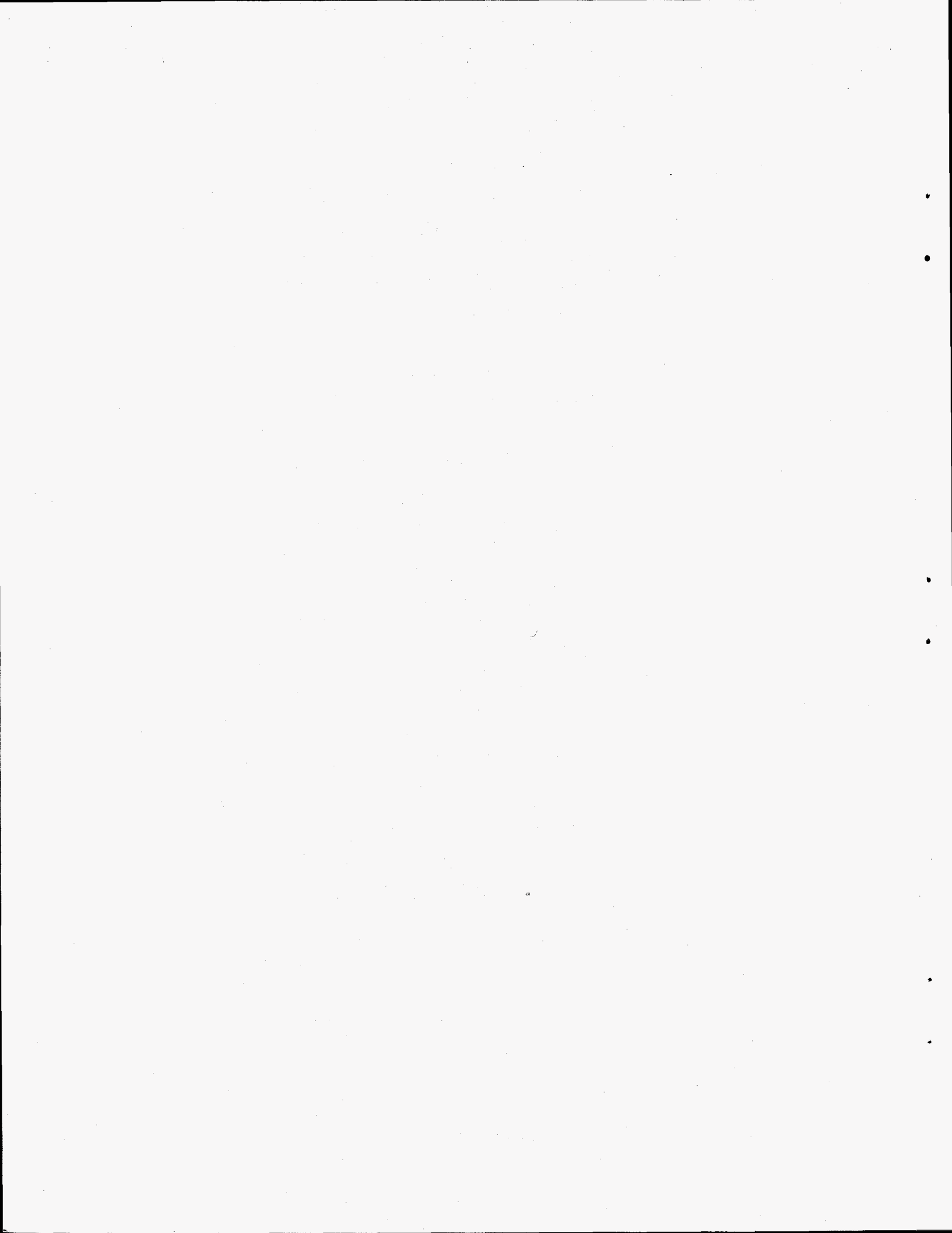
The Jeffreys noninformative prior distribution for a parameter θ can be viewed as the distribution for θ that corresponds to a maximum entropy (i.e. uniform) distribution, in the parameterization where the unknown parameter $\phi(\theta)$ is approximately a location parameter. The constrained noninformative prior, a generalization of the Jeffreys prior, is therefore defined as the distribution for θ corresponding to the maximum entropy distribution on ϕ subject to a constraint. When the constraint is $E(\theta) = \theta_0$, the constrained noninformative prior has the form

$$c \exp(b\theta) J^{1/2}(\theta)$$

where J is the Fisher information.

For a Poisson(λt) random count, where t is the observation period and λ is the event rate, the noninformative prior constrained by $E(\lambda) = \lambda_0$ is gamma with shape parameter 1/2 and mean λ_0 . For binomial data, the constrained noninformative prior is not conjugate to the binomial distribution, but an approximation to the conjugate (beta) distribution is tabulated.

An example from risk assessment shows that the constrained noninformative prior can produce much more realistic answers than the Jeffreys prior distribution.



CONTENTS

ABSTRACT	iii
SUMMARY	v
1. INTRODUCTION	1
2. NONINFORMATIVE PRIORS	2
2.1 General Results	2
2.2 Special Cases	3
3. CONSTRAINED NONINFORMATIVE PRIORS	4
3.1 General Results	4
3.2 Special Cases	6
4. EXAMPLE	9
REFERENCES	14
APPENDIX	15

CONSTRAINED NONINFORMATIVE PRIORS

1. INTRODUCTION

In many risk assessments, the probability of a rare event must be estimated. Two examples are:

- A reliable mechanical system serves as a standby safety system in an operating facility. It is occasionally demanded, and it is periodically tested. The unreliability (probability of failure to function on demand) must be estimated, and the uncertainty in the estimate must be quantified. The system could fail to function because it is out of service for testing or maintenance, or because it fails to start, or because it fails to run for the full required time. Different data sets are relevant for the different failure modes. Each contributing set of data may be sparse, but all should be used if possible.
- A radioactive or toxic waste repository is planned. One quantity to estimate is the probability that the release exceeds a regulatory limit in some specified time. Many possible release scenarios must be considered, such as natural migration of the waste, disruption of the containment barriers by earthquakes, etc. The risks from these scenarios must be combined, and the overall uncertainty quantified.

Often, as in the examples, the process being analyzed is divided into submodels, which must be combined to yield a final answer. Therefore, Bayesian methods are frequently used, because the posterior distributions for the submodels can be propagated easily to yield an overall estimate, with a quantified uncertainty. (For example, see Martz and Waller 1990.) In such an analysis, it is desirable to use only diffuse prior distributions, because the risk assessment will be challenged by people with various agendas. From the viewpoint of defensibility, a noninformative prior is ideal.

From the viewpoint of realism, however, a noninformative prior can be defective. The posterior mean is pulled toward the prior mean, and the noninformative prior mean can be quite unrealistic. Sometimes, few of the rare events have ever been observed, in which case the pull toward the prior mean is strong. For example, the noninformative

prior distribution for a binomial probability has mean $1/2$. In the first example above, it is ridiculous to use a prior distribution that says: the equipment will be out of service half the time; when it is in service, each modeled portion of the system will fail to start half the time; and when the system starts, each portion of the system will fail to run half the time.

In some situations the analysts truly have prior information, leading to a well-justified informative prior. In other cases, however, the analysts have only a vague idea of the realistic values. A desirable prior distribution would have a mean specified by such understanding, but would be otherwise uninformative. That is, the dispersion of the distribution would be large enough that, in some objective sense, it corresponds to ignorance. This is the subject of this note.

2. NONINFORMATIVE PRIORS

Some standard facts about noninformative priors are summarized here, to set the stage for constrained noninformative priors.

2.1 General Results

First, note that for a one-dimensional location parameter, the uniform distribution is an (improper) noninformative prior. Box and Tiao (1973, p. 26) argue that this statement is obvious. Lindley (1958) showed the related fact that the fiducial distribution for a parameter (the distribution that formally gives confidence intervals) is the same as the Bayes posterior distribution if and only if, possibly after a model transformation, the parameter is a location parameter with a uniform prior distribution. Therefore, the uniform prior for a location parameter gives posterior intervals that agree numerically with confidence intervals, which by their definition use only the data, not any prior information.

These considerations motivate the Jeffreys noninformative prior. Transform the original model so that the transformed parameter is approximately a location

parameter. Take a uniform distribution on the transformed parameter, and find the corresponding distribution on the untransformed parameter. This distribution is approximately noninformative. The only detail still to be discussed is the transformation.

Box and Tiao (1973) derive the transformation in general, but the derivation is especially simple when there is a one-dimensional sufficient statistic, that is, when the original distribution is in the exponential family. Let θ be the parameter in the original model, and, without loss of generality, let the sufficient statistic be the maximum likelihood estimator, $\hat{\theta}$. For large samples, $\hat{\theta}$ is approximately normal with mean θ and variance proportional to $J^{-1}(\theta)$, where $J(\theta)$ is the Fisher information in a single observation. Now define

$$\phi = \int^{\theta} J^{1/2}(t) dt \quad (1)$$

so that $d\phi/d\theta = J^{1/2}(\theta)$. (This is the standard asymptotic variance-stabilizing transformation.) Therefore, $\hat{\phi}$ is approximately normal with mean ϕ and constant variance, that is, ϕ is approximately a location parameter. Let f denote the density θ , and let g denote the density of ϕ , with the densities related by

$$f(\theta) = g[\phi(\theta)] |d\phi/d\theta| \quad (2)$$

Therefore, f is proportional to $J^{1/2}(\theta)$ when g is a uniform density. This is the Jeffreys noninformative prior.

2.2 Special Cases

Box and Tiao work out two examples of interest here.

Binomial Data

When the data are binomial(n, p), the transformation is $\phi = \sin^{-1}\sqrt{p}$, and the noninformative prior density is proportional to $p^{-1/2}(1-p)^{-1/2}$. This is analytically convenient because this distribution is beta(1/2, 1/2), and the beta distributions are conjugate priors to the binomial distribution. When p is the probability of a failure on

demand, and f failures and s successes are observed in $f + s$ demands, the posterior distribution of p is $\text{beta}(f + 1/2, s + 1/2)$. The noninformative prior can be thought of as equivalent to observing half a failure on one trial.

Poisson Data

When the data are $\text{Poisson}(\mu)$, the transformation is $\phi = \sqrt{\mu}$, and the noninformative prior density is proportional to $\mu^{-1/2}$. This is an improper distribution, but except for the constant multiplier its form is a limit of a $\text{gamma}(\alpha, \beta)$ density with $\alpha = 1/2$:

$$\lim_{\beta \rightarrow 0} \beta^\alpha \mu^{\alpha-1} e^{-\beta\mu} / \Gamma(\alpha) .$$

This is analytically convenient because the gamma distributions are conjugate priors to the Poisson distribution. When μ is the expected count of events in some time period t , and μ has a $\text{gamma}(\alpha_{\text{prior}}, \beta_{\text{prior}})$ prior distribution, then the posterior distribution based on n events in the time period is $\text{gamma}(\alpha_{\text{post}}, \beta_{\text{post}})$ with $\alpha_{\text{post}} = \alpha_{\text{prior}} + n$, and $\beta_{\text{post}} = \beta_{\text{prior}} + 1$. Formally, if the prior parameters are $\alpha_{\text{prior}} = 1/2$ and $\beta_{\text{prior}} = 0$, the posterior distribution is $\text{gamma}(n + 1/2, 1)$.

If λ is the *rate* of events per unit time, set $\mu = \lambda t$. Then a $\text{gamma}(\alpha, \beta)$ distribution for μ corresponds to a $\text{gamma}(\alpha, \beta t)$ distribution for λ . The noninformative prior for λ is proportional to $\lambda^{-1/2}$, and the corresponding posterior distribution based on n events in time t is $\text{gamma}(n + 1/2, t)$. The mean of the posterior distribution is $(n + 1/2)/t$, and the noninformative prior may be thought of as equivalent to half an event in no time.

3. CONSTRAINED NONINFORMATIVE PRIORS

This section extends the above results to the case where the prior distribution of θ satisfies a constraint, specifying the mean.

3.1 General Results

Basic Definition

Let θ be the parameter in the original model, and let ϕ , defined by Equation (1), be the approximate location parameter in the transformed model. The one-to-one relation between θ and ϕ will sometimes be denoted by $\phi = \phi(\theta)$ and $\theta = \theta(\phi)$. A constrained noninformative prior for θ will be a distribution corresponding to a density for ϕ that is as nearly uniform as possible, subject to any specified constraints. The constraint considered in this note is on the mean of θ : $E(\theta) = \theta_0$, for some specified θ_0 . The term "nearly uniform distribution" is defined precisely as the constrained maximum entropy distribution. Harris (1982) states that for a multinomial distribution with unordered cells, "entropy is best interpreted as a measure of heterogeneity." It follows, by taking a continuous distribution as a limit of a multinomial distribution with the number of cells becoming infinite, that the entropy of a continuous distribution can be interpreted as a measure of flatness of the density, with large entropy corresponding to near flatness. Entropy is not the only such measure, but it is a widely used one, and is used here.

The constrained noninformative prior on θ is obtained by substituting the correct g into Equation (2). For a noninformative prior, g is constant. For a constrained noninformative prior, it is shown in the appendix that $g[\phi(\theta)]$ is of the form $c \exp(b\theta)$. The constrained noninformative prior on θ is therefore of the form

$$c \exp(b\theta) J^{1/2}(\theta). \quad (3)$$

The constants b and c are such that the density integrates to 1.0 and has the specified mean θ_0 .

Refinements in Application

If the constrained noninformative prior has an inconvenient form, the user may wish to approximate it by a similar distribution with an analytically attractive form. This is carried out below when the data are binomial, and is seen to be unnecessary when the data are Poisson.

The constrained noninformative prior is supposed to represent ignorance except for the constraint $E(\theta) = \theta_0$. If θ_0 is a number that is estimated from data, then the prior should also represent the uncertainty in θ_0 . Kass and Steffey (1989) give methods for accomplishing this. They consider empirical Bayes priors, but their ideas can be easily adapted. In particular, their Equation (3.8) adjusts the posterior variance to account for uncertainty in the parameters of the prior distribution. This idea is not considered further here because it goes beyond the main point of this report.

3.2 Special Cases

Poisson Data

Consider Poisson data corresponding to failure rate λ in time t . From Equation (3), the form of the noninformative prior constrained by $E(\lambda) = \lambda_0$ is

$$c \exp(b\lambda) \lambda^{-1/2}.$$

This is a gamma density. Its mean is $(1/2)/b$, so b is forced to equal $1/2\lambda_0$. The posterior distribution is $\text{gamma}(n + 1/2, t + 1/2\lambda_0)$. The prior distribution may be thought of as equivalent to half a failure in time $1/2\lambda_0$.

Binomial Data

Now consider binomial data with parameter p . From Equation (3), the form of the noninformative prior constrained by $E(p) = p_0$ is

$$c \exp(bp) p^{-1/2}(1-p)^{-1/2}. \quad (4)$$

This is not a simple form to use. Its integral is the confluent hypergeometric function, and is familiar to anyone who has worked with the moment generating function of a beta distribution. The integral can be evaluated numerically using the power series expansion of the exponential and the known properties of the beta density (for example, see Section 13.1.2 of Abramowitz and Stegun 1964). By exchanging the roles of p and $1-p$, we see that the density of $1-p$ is the density of p with the sign of b reversed. Therefore, the integral of the density may be calculated using a negative b ,

corresponding to a mean $< 1/2$, or using a positive b , corresponding to a mean $> 1/2$. The case that is not calculated directly can then be found from symmetry. The calculation involves less numerical error for large positive b than for equally large negative b .

Let p be the probability of a failure and $1 - p$ the probability of a success. Assuming that failures and successes are independent events, the posterior distribution based on f failures and s successes is proportional to

$$\exp(bp) p^{f-1/2} (1-p)^{s-1/2},$$

whose moments and quantiles can be found numerically.

A beta density would be much more convenient. Therefore, Table 1 gives the parameters α and β of a beta density that has the same mean and variance as the density (4), for selected values of $p_0 \leq 0.5$. For values of $p_0 \geq 0.5$, reverse the roles of α and β . To interpolate for a value p_0 that is not in the table, use the column for α , and the relation $\beta = \alpha(1 - p_0)/p_0$.

To see how good the beta approximation is, let us compare the constrained noninformative cumulative distribution to the cumulative beta distribution, at selected percentiles of the beta distribution. Table 2 gives this comparison. The beta percentiles were found by the IMSL (1987) function BETIN, and the exact distribution was integrated with the IMSL subroutine QDAWS. The results were the same when the double precision version was used and the allowed relative errors, for calculating b and for integrating the distribution, were decreased. Table 2 shows, for example, that at when $p_0 = 0.10$, the exact noninformative prior assigns probability 0.957 to the interval to the left of the 95th percentile of the beta approximation. The difference between 0.957 and 0.95 is fairly small, and all such differences in the table seem small enough to be adequate for practical use.

Table 1. Parameter b of exact constrained noninformative prior, and beta parameters α and β that match the mean and variance.

p_0	b	α	β
0.50	0.	0.5000	0.5000
0.48	-0.1601	0.4808	0.5208
0.46	-0.3210	0.4630	0.5435
0.44	-0.4835	0.4464	0.5682
0.42	-0.6484	0.4311	0.5953
0.40	-0.8166	0.4168	0.6252
0.38	-0.9891	0.4035	0.6583
0.36	-1.167	0.3911	0.6952
0.34	-1.352	0.3795	0.7368
0.32	-1.545	0.3689	0.7839
0.30	-1.748	0.3590	0.8378
0.29	-1.854	0.3544	0.8678
0.28	-1.964	0.3501	0.9002
0.27	-2.078	0.3459	0.9352
0.26	-2.196	0.3420	0.9733
0.25	-2.319	0.3383	1.015
0.24	-2.447	0.3349	1.060
0.23	-2.582	0.3317	1.111
0.22	-2.723	0.3289	1.166
0.21	-2.873	0.3264	1.228
0.20	-3.031	0.3243	1.297
0.19	-3.201	0.3225	1.375
0.18	-3.383	0.3213	1.464
0.17	-3.579	0.3206	1.565
0.16	-3.793	0.3205	1.683
0.15	-4.027	0.3211	1.820
0.14	-4.287	0.3227	1.982
0.13	-4.579	0.3254	2.177
0.12	-4.910	0.3293	2.415
0.11	-5.292	0.3349	2.709
0.10	-5.743	0.3424	3.081
0.09	-6.285	0.3522	3.561
0.08	-6.958	0.3648	4.195
0.07	-7.821	0.3802	5.051
0.06	-8.978	0.3980	6.235
0.05	-10.61	0.4171	7.925
0.04	-13.08	0.4358	10.46
0.03	-17.22	0.4531	14.65
0.02	-25.53	0.4693	22.99
0.01	-50.52	0.4848	48.00
0.005	-100.5	0.4925	98.00
0.001	-500.5	0.4985	498.0
0.	$-\infty$	0.5000	∞

Table 2. Cumulative distribution of exact constrained noninformative prior, at selected percentiles of the beta approximation.

p_0	Nominal (beta) probability		
	0.90	0.95	0.99
0.50	0.900	0.950	0.990
0.45	0.892	0.943	0.987
0.40	0.888	0.936	0.983
0.35	0.887	0.933	0.979
0.30	0.891	0.932	0.975
0.25	0.898	0.935	0.973
0.20	0.906	0.942	0.975
0.15	0.914	0.951	0.979
0.10	0.916	0.957	0.986
0.05	0.908	0.955	0.989
0.01	0.901	0.951	0.990

4. EXAMPLE

The high pressure coolant injection (HPCI) system in boiling water nuclear power plants was studied (Grant et al. 1994). This is a standby system to inject water into the reactor vessel when water inventory is low and ordinary means of injection are degraded or lost. It is occasionally used and is periodically tested. The data covered seven years, and the customer wanted to see if the unreliability (probability of system failure to function when demanded) had varied during the study period.

Because of the structure of the data and because of engineering considerations, several failure modes were considered. Recovery occurs after a failure if an operator sees that the system is not functioning and takes some simple action to make it function. For example, if the system does not actuate automatically on a start signal, the operator may be able to turn a switch and manually actuate the system. The

failure modes considered were: (1) out of service for testing or maintenance (MOOS), (2) failure to start for reasons other than injection valve failure (FTSO), (3) failure to start because the injection valve failed to open (FTSV), (4) failure to recover from failure to start (FRFTR), (5) failure to run as long as needed (FTR), and (6) failure to recover from failure to run (FRFTR). Failure of the system when demanded equals

$$\text{MOOS or } [(FTSO \text{ or } FTSV) \text{ and } FRFTR] \text{ or } (FTR \text{ and } FRFTR), \quad (5)$$

and the unreliability is the probability of system failure on demand.

The failure and demand counts were extracted from the failure reports and testing frequencies, and are given in Table 3. The numbers account for facts such as the following: The data from unplanned demands were relevant for all failure modes, but the data from tests were relevant only for FTSO and FTR. Some failures to start did not allow the system to attempt to run, while others allowed running, perhaps after a delay. Although at first glance the tabulated demand counts may seem inconsistent, they were established with great care.

Table 3. Data summary (failures/demands) for HPCI reliability study.

<u>Year</u>	<u>MOOS</u>	<u>FTSO</u>	<u>FTSV</u>	<u>FRFTR</u>	<u>FTR</u>	<u>FRFTR</u>
1987	0/16	3/28	0/16	0/1	1/26	0/1
1988	0/10	1/23	1/10	0/2	0/23	0/0
1989	1/7	1/24	0/6	0/0	1/24	0/0
1990	0/13	1/28	0/12	0/1	1/27	1/1
1991	0/9	3/26	0/9	0/1	2/25	1/1
1992	0/6	0/24	0/4	0/0	2/24	0/0
1993	0/2	2/17	0/2	0/0	0/18	0/0
Total	1/63	11/170	1/59	0/5	7/167	2/3

The chi-squared goodness of fit test found no significant difference between years for any failure mode. Confirming this, for each failure mode an attempt to fit a beta-binomial distribution to the data (estimating the parameters of the beta distribution by maximum likelihood) was unable to fit a nondegenerate distribution. Therefore, the most realistic analysis of the data in Table 3 uses the row labeled "Total."

Based on the data totals, find the unreliability as follows. For each failure mode, use the Jeffreys noninformative prior, the failure count f , and the demand count d to obtain a posterior distribution, $\text{beta}(f + 1/2, d - f + 1/2)$, for the probability of that failure mode. Treat the events in Expression (5) as independent events, and write the probability of Expression (5) in terms of sums and products of the probabilities of the six failure modes. Treat the six Bayes distributions of these probabilities as independent distributions, and find the mean and variance of the overall expression in terms of the means and variances of the six distributions. This results in the exact Bayes mean and variance of the unreliability. To obtain an approximate 90% interval for the unreliability, approximate the distribution as a beta distribution having the mean and variance just found, and use the 5th and 95th percentiles of that beta distribution. This is the approximation advocated by Martz and Waller (1990). All this is easy to accomplish with a computer program that contains very small subprograms or macros, such as a macro to find the probability of the union of two independent events and a macro to find the variance of a product of independent random variables. The calculations for this report were performed with SAS (1988). Based on the last row of Table 3, the mean system unreliability is 0.058, and an approximate 90% interval (5% Bayes probability in each tail) is (0.024, 0.103). This is shown at the right in Figure 1. The dotted horizontal line marks the mean system unreliability, 0.058.

However, the customer wanted to see if the unreliability varied from year to year. Although no significant variation between years was seen for any one failure mode, it was conceivable that a trend or pattern might emerge when the failure modes were combined. Therefore, the unreliability was calculated for each year.

For the first attempt at a calculation, the year-specific unreliability was found in

the same way as the overall unreliability, by updating the Jeffreys prior with the year's data for each failure mode, and combining the posterior distributions for the failure modes. The results are given in Figure 1, with the open circles showing means and the bars showing 90% intervals. This could be called an example of the Lake Wobegon effect—all the years are worse than average. The reason is that, for each failure mode, the so-called noninformative prior pulls the probability toward 0.5.

Therefore, for each failure mode and each year, the distribution for the probability of failure was obtained by updating the beta approximation of the constrained noninformative prior. For example, Table 3 shows that FRFTS had a total of 0 failures in 5 trials. To obtain a mean corresponding to the total counts, even when no failures were observed, the Jeffreys prior was updated with the overall data. The resulting posterior distribution for $P(\text{FRFTS})$ was $\text{beta}(0.5, 5.5)$, with mean $0.5/5.5 = 0.0909$. This mean was used as a constraint when the FRFTS data from an individual year were analyzed. The beta approximation to the constrained noninformative prior with this mean is $\text{beta}(0.3603, 3.963)$; these values could have been obtained accurately enough by interpolating Table 1 for α at $p_0 = 0.0909$, but actually the software that generated Table 1 was used to find α and β . This prior distribution was updated with each year's FRFTS data to find the year-specific posterior distribution for $P(\text{FRFTS})$. The other failure modes were treated similarly, and the year-specific unreliability was found by combining the failure modes.

The results are given in Figure 1, with solid circles showing the means and bars showing 90% intervals. The means are comparable to the overall unreliability, and the uncertainty intervals are larger. These results could be shown to the customer without raising questions about the unreasonably large numbers.

Apart from the general fact that the Jeffreys prior leads to larger posterior unreliabilities than the constrained prior does, the two sets of answers in Figure 1 seem consistent except for 1993. That year has the largest mean, based on the Jeffreys prior, and one of the smaller means, based on the constrained prior. The explanation lies in the fact that only two unplanned demands occurred in 1993. Therefore, the means for $P(\text{MOOS})$ and $P(\text{FTSV})$ are pulled strongly toward the prior means, 0.5 for the Jeffreys prior and much smaller values for the constrained prior.

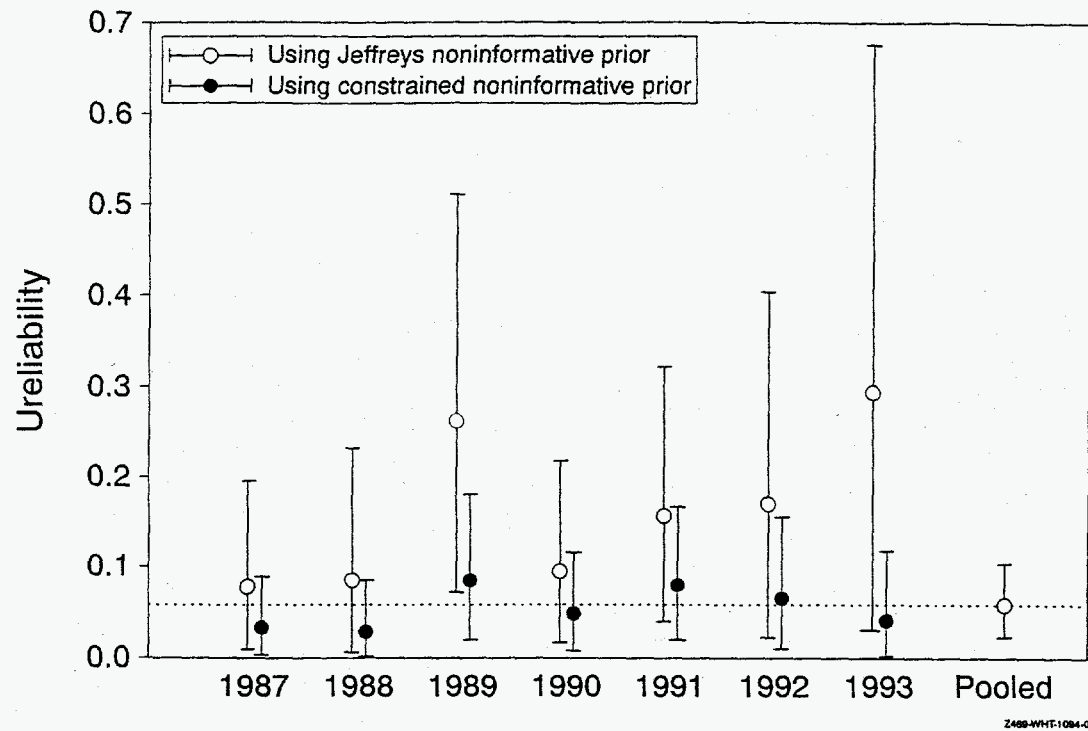


Figure 1. Posterior means and 90% intervals for unreliability, using the Jeffreys noninformative prior and the constrained noninformative priors.

REFERENCES

Abramowitz, Milton, and Irene A. Stegun (1964), *Handbook of Mathematical Functions*, New York: Dover Publications, Inc.

Box, George E. P., and George C. Tiao (1973), *Bayesian Inference in Statistical Analysis*, Reading, MA: Addison-Wesley.

Grant, G. M., C. L. Atwood, C. D. Gentillon, and D. G. Hall (1994), *High Pressure Coolant Injection (HPCI) System Performance, 1987-1993 DRAFT*, EGG-ERTP-10929, Idaho National Engineering Laboratory report.

Harris, B. (1982), "Entropy," in *Encyclopedia of Statistical Sciences, Vol. 2*, S. Kotz and N. L. Johnson, eds., New York: John Wiley & Sons.

IMSL (1987), *User's Manual: Stat/Library* and *User's Manual: Math/Library*, Houston, TX: IMSL, Inc.

Lindley, D. V. (1958), "Fiducial Distributions and Bayes' Theorem," *Journal of the Royal Statistical Society, Series B*, **20**, pp. 102-107.

Martz, H. F. and R. A. Waller (1990), "Bayesian Reliability Analysis of Complex Series/Parallel Subsystems and Components," *Technometrics*, **32**, pp. 407-416.

SAS (1988), *SAS Language Guide for Personal Computers*, Cary, NC: SAS Institute, Inc.

APPENDIX

This section finds the density $g(\phi)$ that maximizes the entropy

$$- \int \log[g(\phi)] g(\phi) d\phi$$

subject to the constraint $E[\theta(\phi)] = \theta_0$. First, convert the problem to one of finite dimension. Choose a finite interval (a, b) on the ϕ -axis, and divide it into subintervals at values ϕ_i . Let Δ_i denote $\phi_i - \phi_{i-1}$, and let g_i denote $g(\phi_i)$. As the interval (a, b) grows large and the Δ_i 's grow small, the integral is approximated by a finite sum. Now the task is to maximize

$$- \sum_i [\log(g_i) g_i \Delta_i]$$

subject to the constraints

$$\sum_i [g_i \Delta_i] = 1$$

and

$$\sum_i [\theta(\phi_i) g_i \Delta_i] = \theta_0 .$$

Using Lagrange multipliers, this is equivalent to maximizing

$$- \sum_i \log(g_i) g_i \Delta_i + \lambda_1 [\sum_i g_i \Delta_i - 1] + \lambda_2 [\sum_i \theta(\phi_i) g_i \Delta_i - \theta_0]$$

with respect to the g_i 's, λ_1 , and λ_2 . The derivative with respect to any g_i is

$$\{-1 - \log(g_i) + \lambda_1 + \lambda_2 \theta(\phi_i)\} \Delta_i .$$

Set this to zero, to see that $\log(g_i)$ must be of the form $a + b\theta(\phi_i)$. Therefore, we have

$$g_i = c \exp[b\theta(\phi_i)] .$$

Therefore, in the limit $g(\phi)$ must equal $c \exp[b\theta(\phi)]$, for all ϕ . When ϕ is written as a function of θ , this expression is $g[\phi(\theta)] = c \exp(b\theta)$.