Pak J. Statist.
2003 Vol. 19(3) pp 301 – 313

THE DISCREPANCY OF P-VALUES AND POSTERIOR PROBABILITY IN POISSON DISTRIBUTION

Rahim Chinipardaz
Department of Statistics, University of Shahid Chamran,
Ahvaz, Iran.

ABSTRACT

The problem of testing a point null hypothesis in Poisson distribution is considered. The Bayesian measure of evidence against the null hypothesis is compared with P-value (or observed significance level). Lindley and Jeffreys approach is followed for Poisson distribution.

Bayesian measure of evidence against the null hypothesis expressed in terms of the infimum of the posterior probability of the null hypothesis when the infimum is taken over the class of the some reasonable priors. The result shows that the P-value is much lower than the lowest probability obtained even if all priors are considered. This irreconcilability between P-value and any Bayesian analysis of null hypothesis is no longer valid for one sided null hypothesis

KEYWORDS

P-values; Posterior Probability; Mixtures; Lower Bounds

1. INTRODUCTION

The irreconcilability of Bayesian evidence and frequentist evidence has been extensively studied by many authors. When the null hypothesis consists of a single point, it has been known that these two measures evidence are different. Lindley in his famous paper, “A Statistical Paradox” considered a Normal distribution with unknown mean, $N(\theta, \sigma^2)$ and showed that for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ with a fixed significance level, the posterior probability of null hypothesis would be increased closely to one as the sample size increased. He used a improper Uniform fixed prior for the parameter mean. Such a result can be obtained by Jeffreys' approach. Jeffreys (1961) suggested reporting the posterior probability of and H_0 and H_1 based on assigning equal prior probability to the two hypotheses and using a conjugate prior. Although he did not refer to it as paradox and did not compare significance level with the posterior probability of null hypothesis again his approach leads to confirming Lindley's result. In an interesting paper by Edwards et al. (1963) it was shown that this discrepancy between two evidences is not dependent on precise fixed prior and its occurrence even as a class of priors is considered. A similar method was used by Dickey (1977) examine the infimum of the Bayes factor of H_0 under classes of priors. The problem has been discussed and commented by some other workers including Bartlett (1957), Pratt (1965), Shafer (1982) Berger (1985), Good (1987), Bas and Sanches (1989) and Lee (1996).

Expanding on these earlier works, Berger and Sellke (1987) considered the Normal distribution with some various class of the priors, including the class of all distributions. The result was even taking all priors the infimum of the posterior probability of H_0 to be true is much larger than the significance level. The similar result was shown by Berger and Delampady (1987) for Binomial distribution (see also, Berger and Mareta, 1994). The distinction has been also studied independently of the sample size. In fact the distinction between the two approaches also exist even without increasing the sample size. An extensive list of references in this area are included in Shafer (1982) and Berger and Mereta (1994). In this paper some attempts have been made to expand this problem to the Poisson distribution. The motivation is that the Poisson distribution has an infinite discrete random variable.

The paper is organized as follows: In section 2 some notations and necessary preliminary results are presented. The main results will be given in the two next sections. Section 3 is devoted to comparison between the posterior probability and the significance level when a fixed prior has been used to obtaining the posterior distribution and the sample size increase. In section 4 some lower bounds on the posterior probability when the prior distribution of H_1 is allowed to vary within some reasonable class of distributions is examined. The comparison is also considered independently of the sample size. Finally, some comments and conclusions are given in section 5.

2. NOTATION AND PRELIMINARY RESULTS

Consider simple situation of observing a random variable X having density $\mathrm{f}(\mathrm{x}\,|\theta)$ where θ is an unknown parameter assuming values in parameter space $\theta \in R$. It is desired to test the null hypothesis $\mathrm{H}_0 : \theta = \theta_0$ versus the alternative $\mathrm{H}_1 : \theta \neq \theta_0$ where θ_0 is a specified value of parameter corresponding to null hypothesis. In classical approach the significance level or P-value is based on the probability of exceeding some test statistic, $T(X)$, from its observed value, i.e.

$$\mathrm{P-value} = \mathrm{P}_{\theta=\theta_0}(\,|\mathrm{T(X)}| > \mathrm{t(X)}) \tag{1}$$

As an example to test $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ for Normal distribution

$$T(X) = \sqrt{n}\,|\,\overline{X} - \theta_0\,|\,/\,\sigma$$

where $\overline{X}$ is the sample mean and then

$$\mathrm{P-value} = 2(1-\Phi(t))$$

where Φ is the standard normal cumulative distribution function and

$$t = \sqrt{n}\,|\,\overline{X} - \theta_0\,|\,/\,\sigma\,.$$

In the case of Poisson parameter there is no symmetry of the definition of P-value. Therefore, the P-value is difficult to obtain. Instead of this definition, the ``intrinsic significance level" is suggested. Choosing $T(X) = 1/\,f\,(X\,|\theta_0)$ in (1) leads to defining the P-value as

$$\mathrm{P-value} = \mathrm{P}_{\theta=\theta_0}\left(\{\mathrm{y} : \mathrm{f}\,(\mathrm{y}|\theta_0\,)| \le \mathrm{f}\,(\mathrm{x}|\theta_0)\}\right) \qquad (2)$$

Such a definition has been used by Berger and Delampady, 1987 for the non-symmetric Binomial distribution (see Kemptorne, 1976 for more detail).
In Bayesian approach the probability of $\mathrm{H}_0 : \theta = \theta_0$ is denoted by π_0 and $\pi_1 = 1-\pi_0$ as the probability of $\mathrm{H}_1 : \theta \neq \theta_0$ and it is supposed the mass on H_1 is spread out according to the density $g(\theta)$. The marginal density of X is

$$m(\mathbf{x}) = \pi_0\, f(\mathbf{x}|\theta_0\,) + (1-\pi_0)\, m_g\,(\mathbf{x})$$

where

$$\mathrm{m_g}(\mathrm{x}) = \int_{\theta\neq\theta_0} \mathrm{f}\,(\mathrm{x}|\theta)\mathrm{g}(\theta)\mathrm{d}\theta.$$

The posterior probability of H_0 is given by

$$\mathrm{P}\;(\mathrm{H}_0\,|\,\mathrm{x}) = \left[1+\frac{1-\pi_0}{\pi_0}\times\frac{\mathrm{m_g}(\mathrm{x})}{\mathrm{f}\,(\mathrm{x}\,|\theta_0)}\right]\left[1+\frac{1-\pi_0}{\pi_0}\times\frac{1}{\mathrm{B}(\mathrm{x},\mathrm{g})}\right]^{-1} \qquad (3)$$

where $\mathrm{B}(\mathrm{x},\mathrm{g}) = \mathrm{f}\,(\mathrm{x}\,|\,\theta_0)/\mathrm{m_g}(\mathrm{x})$ is the Bayes factor for H_0 versus H_1.

In Bayesian approach the posterior probabilities of hypotheses are primary Bayesian measures in testing problems. One may calculate the posterior probabilities $\mathrm{P}(\mathrm{H}_0\,|\,\mathrm{x})$

and $P(H_0 \mid x)$ and decides between H_0 and H_1 accordingly. Many people are more comfortable with the Bayes factor. In this article we focuss on $P(H_0 \mid x)$.

3. JEFFREYS-LINDLEY PARADOX IN POISSON DISTRIBUTION

Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. In classical approach and using approximation, the
Normal distribution to Poisson distribution P-value is given

$$P-\text{value} = P_{\theta_0}(|T(X)| > t(X)) = 1 - 2\Phi\left(\frac{t(x) - n\theta_0}{\sqrt{n\theta_0}}\right) \tag{4}$$

where $t(x)$ is $\sum_{i=1}^{n} x_i$. Now consider a Bayesian who gives the probability π_0 to H_0 and preads the mass out on H_1 according to the vague prior density $g(\theta) = 1; \theta \in R^+$. It can be used to obtain the posterior distribution as

$$\begin{aligned}
P(\theta_0 \mid x) &= \frac{\pi_0 \dfrac{e^{-n\theta_0} \theta_0^{t(X)}}{\prod_{i=1}^{n}(x_{i}!)}}{\pi_0 \dfrac{e^{-n\theta_0} \theta_0^{t(X)}}{\prod_{i=1}^{n}(x_{i}!)} + (1-\pi_0)\int_0^\infty \dfrac{e^{-n^{\theta}} \theta^{t(x)}}{\prod_{i=1}^{n}(x_{i}!)} d\theta} \\
&= \left[1 + \frac{1-\pi_0}{\pi_0} t(x)! e^{n\theta_0} (n\theta_0)^{t(x)}\right]^{-1} \\
&\approx \left[1 + \frac{1-\pi_0}{\pi_0} \sqrt{\frac{2\pi}{t(x)}}\, e^{(n^{\theta_0} - t(x))\left(\frac{t(x)}{n\theta_0}\right)^{t(x)+1}}\right]^{-1}
\end{aligned}$$

Following Jeffreys, gamma distribution is used as the conjugate prior,

$g(\theta) = \beta^{\alpha} e^{-\beta\theta} \theta^{\alpha-1} \,/\, \Gamma(\alpha), \beta, \alpha > 0 \text{ with } E_g(\theta) = \alpha/\beta = \theta_0$ to obtain the posterior probability of θ_0 given **X**

$$P(\theta_0 \mid x) = \frac{\pi_0 e^{-n\theta_0} \theta_0^{t(x)}}{\pi_0 e^{-n\theta_0} \theta_0^{t(x)} + (1-\pi_0)\int_0^\infty \dfrac{\beta\alpha}{\Gamma(\alpha)} e^{-(B+n)} \theta^{t(x)+\alpha-1} d\theta}$$

$$=\left[1+\frac{1-\pi_0}{\pi_0}\binom{t(x)+\alpha-1}{t(x)}\frac{\alpha^{\alpha}e^{n\theta_0}}{(n\theta_0+\alpha)^{t(x)}}\right]^{-1}.$$

It depends on α. For special case α = 1 and some easy calculation

$$P(\theta_0|x)\approx\left[1+\frac{1-\pi_0}{\pi_0}\sqrt{\frac{2\pi}{t(x)}}\ e^{\,(n\theta_0-t(x))}\left(\frac{t(x)}{n\theta_0+1}\right)^{t(x)+1}\right]^{-1} \qquad (6)$$

Table 1 shows the posterior distribution of $H_0:\theta=1$ versus $H_1:\theta\neq1$ using $\pi_0=1/2$ with $g(\theta)=1$ prior, $P_L\,(H_0x)$ given in (5), and with $g(\theta)=e^{-\frac{\theta}{\theta_0}}\ /\theta_0$ prior, $P_J\,(H_0|\,x)=1$ given in (6), and increasing of $n=n\theta_0$. They are compared with P-value equal to 0.10, 0.05 and 0.01. As can be seen from the Table with a fixed P-value the posterior probability of H_0 be true is increasing to one as *n* increases no matter how small the P-value. Some other value of α was examined and similar results were obtained. For a statistician with Bayesian approach with n = 20, H_0 is supported with t(**x**) = 29, while it is rejected with P-value of 0.05 for a classic statistician.

Table 1: Increasing of $P(H_0\,|\,x)$ with sample size, $\pi_0=1/2$
and P-value= 0.10, 0.05 and 0.01

N= nθ₀	P-value = 0.10		P-value = 0.05		P-value = 0.01	
	$P_L\,(H_0\,|\,x)$	$P_J\,(H_0\,|\,x)$	$P_L\,(H_0|x)$	$P_J\,(H_0|x)$	$P_L\,(H_0|x)$	$P_J\,(H_0|x)$
5	0.183	0.567	0.083	0.402	0.048	0.302
10	0.241	0.599	0.128	0.415	0.061	0.286
20	0.312	0.644	0.200	0.520	0.780	0.293
40	0.392	0.659	0.244	0.550	0.101	0.316
50	0.419	0.714	0.292	0.599	0.110	0.327
100	0.506	0.767	0.358	0.650	0.142	0.370
200	0.592	0.816	0.480	0.718	0.185	0.423
400	0.673	0.859	0.547	0.767	0.237	0.491
1000	0.765	0.903	0.651	0.843	0.324	0.586
5000	0.879	0.953	0.804	0.919	0.511	0.746
10000	0.912	0.966	0.854	0.979	0.681	0.804

The result is simply that small P-value gives a misleading impression as to the validity of H_0. That is because in the case very large sample even minor departure from H_0 leads to rejecting H_0. As Kass and Raftery (1995) also pointed out, frequentist tests tend to reject null hypotheses systematically in very large samples. This is a real problem in sociology where data sets are frequently have thousands of such samples. It should be awarded especially for non-statisticians who dealt with such cases and used common P-values.

4. LOWER BOUNDS ON POSTERIOR PROBABILITIES

In this section it is attempted to dismiss the conflict between P-value and the posterior probability. One might argue that the difference between $P(H_0 \mid \mathbf{x})$ and P-value given in Table 1 is due to choosing $g(\theta)$ to be specified. This leads to looking at lower bounds on $P(H_0 \mid \mathbf{x})$ when the distribution of θ given that H_1 is true, $g(\theta)$ is allowed to vary within some classes of distributions G where G is a reasonable class of distributions. Therefore, $\inf_{g \in G} P(H_0 \mid x, G)$ is compared with P-value. Letting

$$\underline{P}(H_0 \mid x, G) = \inf_{g \in G} P(H_0 \mid x, G)$$

$$\underline{P}(H_0 \mid x, G) = \left[1 + \frac{1-\pi_0}{\pi_0} \times \frac{\sup_{g \in G} m_g(x)}{f(x \mid \theta_0)}\right]^{-1} = \left[1 + \frac{1-\pi_0}{\pi_0} \times \frac{1}{\underline{B}(x \mid G)}\right]^{-1}.$$

where $\underline{B}(x \mid G)$ is the $\inf_{g \in G} B_g(x) = f(x \mid \theta_0) \;/\; \sup_{g \in G} m_g(x)$ is the Bayes factor for H_0 versus H_1. This approach can highlight discrepancy between posterior probabilities and classical P-values. It is extended by Berger and Sellke (1987) and followed by Berger and Delampady (1987) in normal distribution. Berger and Delampady (1987) then looked at Binomial distribution that is no longer a natural definition of symmetry as normal distribution. They developed a variety of different classes of *G* and showed the similar results as in normal distribution. i.e. for all considered classes the lower bounds were markedly larger than the P-values. The classes are:

G_A ={all possible distributions}
G_{US} ={all uni-model distributions symmetric θ_0 }
G_{Me} = {all distributions with median θ_0 } (7)
G_{Meo} = {all distributions with median and model θ_0 }

These classes are surveyed in this paper.

4.1 Lower Bounds for $G_A = \{$All Distributions$\}$

The simplest results with elementary proof are for G_A

Theorem 1: To test $H_0:\theta=\theta_0$ versus $H_1:\theta\neq\theta_0$ when $g\in G_A$

$$\underline{P}(H_0\,|\,x,G_A)=\left[1+\frac{1-\pi_0}{\pi_0}\times e^{(\theta_0-x)}\left(\frac{x}{\theta_0}\right)^x\right]^{-1} \tag{8}$$

Proof: omitted.

The corresponding P-values and the values of $\underline{P}(H_0\,|\,x,G_A)$ with $\pi_0=1/2$ are given in third and fourth columns of Table 2, respectively. The various values of θ_0 and *x* have been examined. As can be seen lower bounds of posterior probabilities for G_A are much larger than P-values.

The class of G_A is such a large class and includes large unreasonable priors. This leads to suspicion that using G_A biases the conclusion against H_0 and looks at using more reasonable classes of priors.

4.2 Lower Bounds for G_{US} = {All Uni-modal Distributions Symmetric about θ_0 }

A reason required for priors is to be uni-modal symmetry assumption. The class of such priors on $\theta\neq\theta_0$ has been denoted by G_{US}. The following theorem shows minimizing $P(H_0\,|\,x,G)$ over $g\in G_{US}$.

THEOREM 2: To test $H_0:\theta=\theta_0$ versus $H_1:\theta\neq\theta_0$ when $g\in G_{US}$

$$\underline{P}(H_0\,|\,x,G_{US})=\left[1+\frac{1-\pi_0}{\pi_0}\times\frac{\sup\limits_{k>0}\int_{-k}^{k}e^{-\theta(u)}\,\theta(u)^x\,du}{e^{-\theta_0}\theta_0^x}\right]^{-1}$$

where $\theta(u)=\dfrac{u^2+2\theta_0+u\sqrt{u^2+2\theta_0}}{2}$.

Proof: The proof has been given in Berger and Delampady (1987) for Binomial distribution. The similar proof can be used with transforming

$$u(\theta)=\frac{\theta-\theta_0}{\sqrt{\theta}}$$

and inverse function

$$\theta(u)=\frac{u^2+2\theta_0+u\sqrt{u^2+2\theta_0}}{2}.$$

Comparison between P-value and $\underline{P}(H_0|x,G_{US})$ is shown in the third and the fifth columns of Table 2. As was expected, because of $G_A\supset G_{US}$ then $\underline{P}(H_0|x,G_A)\le\underline{P}(H_0|x,G_{US})$ i.e. the greater the difference between P-value, the lower bound on posterior probability.

The reasonable classes of densities priors may also be obtained by specifying their mode and their median. The class of G_{Me} defined in (7) includes all densities priors with median θ_0. This is an honest decision to devote half of the weight to less than θ_0 and the another half to greater than θ_0. The class of G_{Meo} is more restricted than G_{Me} wherein *g* is also non-decreasing on $[0,\theta_0]$ and non-increasing on $[\theta_0,\infty]$. Because $G_{Me}\supset G_{Meo}$ then $\underline{P}(H_0|x,G_{Me})\le\underline{P}(H_0|x,G_{Meo})$. The next two subsections are devoted to comparison between P-value and $P(H_0|x,G)$ in these two classes.

4.3 Lower Bounds for G_{Me} = { All Distributions with Median θ_0 }

Theorem *3* : $H_0:\theta=\theta_0$ versus $H_1:\theta\ne\theta_0$ when $g\in G_{Me}$

$$\underline{P}(H_0|x,G_A)=\left[\frac{1+\pi_0}{2\pi_0}+\frac{1-\pi_0}{2\pi_0}\times e^{(\theta_0-x)}\left(\frac{x}{\theta_0}\right)^x\right]^{-1} \tag{10}$$

Proof: For $x\ge\theta_0$ we have

$$m_g(x)=\int_0^{\theta_0}\frac{e^{-\theta}\theta^x}{x!}g_1(\theta)d\theta+\int_{\theta_0}^{\infty}\frac{e^{-\theta}\theta^x}{x!}g_2(\theta)d\theta.$$

g_1 and g_2 are both nonnegative and $e^{-\theta}\theta^x$ has its maximum at $\hat{\theta}=x$. Also

$$\int_0^{\theta_0} g_1(\theta)\,d\theta = \int_{\theta_0}^{\infty} g_2(\theta)\,d\theta = \frac{1}{2}$$

$$\sup_{g_1} \int_0^{\theta_0} \frac{e^{-\theta}\theta^x}{x!} g_1(\theta)\,d\theta = \frac{1}{2} \times \frac{e^{-\theta_0}\theta_0^x}{x!} \qquad (11)$$

and

$$\sup_{g_2} \int_{\theta_0}^{\infty} \frac{e^{-\theta}\theta^x}{x!} g_2(\theta)\,d\theta = \frac{1}{2} \times \frac{e^{-x} x^x}{x!} \qquad (12)$$

Combining (11) and (12) with easy calculation complete the proof. For $x < \theta_0$ the similar approach can be applied. The results are shown in the sixth column of Table 2. As can be seen from the Table, the discrepancy between P-values and the posterior probabilities becomes more than

$\underline{P}(H_0 \mid x, G_A)$ and $\underline{P}(H_0 \mid x, G_{US})$.

4.4 Lower Bounds for G_{Meo} = {All Distributions with Median and Mode θ_0 }

Theorem 4*:* $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ when $g \in G_{Meo}$

$$\underline{P}(H_0 \mid x, G_{Meo}) = \begin{cases} \left[\dfrac{1+\pi_0}{2\pi_0} + \dfrac{1-\pi_0}{2\pi_0} \times \dfrac{\sup\limits_{k>0} \frac{1}{k} \int_{\theta_0}^{\theta_0+k} e^{-\theta}\,\theta^x\,d\theta}{e^{-\theta_0}\theta_0^x} \right]^{-1} & x \geq \theta_0 \\ \left[\dfrac{1+\pi_0}{2\pi_0} + \dfrac{1-\pi_0}{2\pi_0} \times \dfrac{\sup\limits_{k>0} \frac{1}{k} \int_{\theta_0-k}^{\theta_0} e^{-\theta}\,\theta^x\,d\theta}{e^{-\theta_0}\theta_0^x} \right]^{-1} & x < \theta_0 \end{cases} \qquad (13)$$

Proof: Suppose $x \leq \theta_0$

$$\sup_{g \in G_{Meo}} m_g(x) = \sup_{g_1} \int_0^{\theta_0} \frac{e^{-\theta}\theta^x}{x!} g_1(\theta)\,d\theta + \sup_{g_2} \int_{\theta_0}^{\infty} \frac{e^{-\theta}\theta^x}{x!} g_2(\theta)\,d\theta\,.$$

g_1 and g_2 are both nonnegative, g_1 is distributed non-decreasingly on $[0, \theta_0]$ and g_2 is distributed non-increasingly on $[\theta_0, \infty]$. Also

$$\int_0^{\theta_0} g_1(\theta)\,d\theta = \int_{\theta_0}^{\infty} g_2(\theta)\,d\theta = \frac{1}{2}.$$

Because $e^{-\theta}\theta^x / x!$ is decreasing on $[x, \infty]$ and $\theta_0 \in [x, \infty]$ then

$$\sup_{g_2} \int_{\theta_0}^{\infty} \frac{e^{-\theta}\theta^x}{x!} g_2(\theta)\,d\theta = \frac{1}{2}\frac{e^{\theta_0}\theta_0^x}{x!} \tag{14}$$

For $\sup_{g_1} \int_0^{\theta_0} \frac{e^{-\theta}\theta^x}{x!} g_1(\theta)\,d\theta$ take

$$u(\theta) = \begin{cases} \theta & \theta \le \theta_0 \\ 2\theta_0 - \theta & \theta > \theta_0 \end{cases}$$

we have

$$\sup_{g_1} \int_0^{\theta_0} \frac{e^{-\theta}\theta^x}{x!} g_1(\theta)\,d\theta = \sup_{g_3} \frac{1}{x!} \int_0^{2\theta_0} h(x,\theta)\, g_3(\theta) d\theta$$

where

$$h(x, \theta_0) = \begin{cases} e^{-\theta}\theta^x & \theta \le \theta_0 \\ e^{-(2\theta_0 - \theta)}(2\theta_0 - \theta)^x & \theta < \theta_0 \end{cases}$$

and $g_3(\theta)$ is nonnegative, uni-modal and symmetric about θ_0 with

$$\int_0^{2\theta_0} g_3(\theta)\, d\theta = \frac{1}{2}.$$

Now because $g_3(\theta)$ is uni-modal and symmetric about θ_0 then it is the mixture of the Uniform distributions $U(\theta_0 - k, \theta_0 + k),\ \ k > 0$ and

$$\sup_{g_3} \int_0^{2\theta_0} \frac{1}{x!} h(x,\theta)\, g_3(\theta)\,d\theta = \sup_{k>0} \frac{1}{4k} \int_{\theta_0 - k}^{\theta_0 + k} \frac{1}{x!} h(x,\theta) d\theta = \sup_{k>0} \frac{1}{2k} \int_{\theta_0 - k}^{\theta_0} \frac{1}{x!} h(x,\theta)\,d\theta \tag{15}$$

Combine (14) and (15) to obtain

$$\sup_{g} m_g(x) = \frac{1}{2x!}\left[e^{\theta_0}\theta_0^x + \sup_{k>0}\frac{1}{k}\int_{\theta_0-k}^{\theta_0} e^{\theta_0}\,\theta_0^x\, d\theta\right]$$

For $x > \theta_0$ the similar way can be followed to obtain (13).

Results of comparison between P-values and lower bounds of posterior probability when *g* belongs to the class of G_{Meo} priors are shown in the last column of Table 2. Because this class is the smallest class between the previous classes distinction between classical measure and Bayesian measure against H_0 hypothesis is much more than before.

Table 2: Lower bounds of $P(H_0|x,G)$ with the various types of prior densities class $\pi_0 = 1/2$

θ_0	X	P-value	$\underline{P}(H_0 \mid x, G_A)$	$\underline{P}(H_0 \mid x, G_{US})$	$\underline{P}(H_0 \mid x, G_{Me})$	$\underline{P}(H_0 \mid x, G_{Meo})$
1	3	0.080	0.215	0.274	0.301	0.339
1	4	0.019	0.073	0.116	0.127	0.167
2	4	0.004	0.017	0.033	0.033	0.049
2	5	0.143	0.316	0.362	0.387	0.414
2	7	0.053	0.171	0.221	0.254	0.300
2	9	0.005	0.023	0.043	0.043	0.068
5	9	0.102	0.216	0.274	0.306	0.347
5	10	0.004	0.172	0.183	0.202	0.257
5	12	0.005	0.029	0.054	0.055	0.089
5	14	0.001	0.004	0.006	0.009	0.016

As can be seen from Table 2 the lower bound for G_{Meo} class is largest. It followed by G_{Me} and G_{US} bounds, respectively. This three classes that attempt to spread mass on both sided about θ_0, have moderately similar lower bounds and much larger than P-values, about four times or even more than P-values. This is expected and supports the notion of discrepancy between significance testing and posterior probabilities in testing precise hypotheses. It should be mentioned that the lower bound for G_A is different than other lower bounds, nearly two or three times as much as P-values. It is because that G_A class is too large and includes many unreasonable priors.

5. COMMENTS AND CONCLUSIONS

The results given in section 4 confirm those given by Berger and Sellke (1987) in normal distributions and those given by Berger and Delampady (1987) in Binomial distributions.

It should be noted that these results will fail to one sided $H_0 : \theta \le \theta_0 \; (\theta \ge \theta_0)$ versus $H_1 : \theta > \theta_0 \; (\theta < \theta_0)$. Because,

$$\mathrm{P-value} = \mathrm{P}(\mathrm{X} \ge \mathrm{x}) = \sum_{j=x}^{\infty} \frac{e^{-\theta_0} \theta_0^j}{j!}$$

Now using vague distribution $g(\theta) = 1$ on entire R^+ as prior

$$\mathrm{P}(\theta \le \theta_0) = \int_0^{\theta_0} \frac{e^{-\theta} \theta^x}{x!} d\theta = \frac{1}{x!} \int_0^{\theta_0} e^{-\theta} \theta^x \, d\theta = \sum_{j=x}^{\infty} \frac{e^{-\theta_0} \theta_0^j}{j!} = \mathrm{P-value}$$

Also using $\Gamma(\alpha, \beta)$ with $\alpha > 1$ as prior distribution it leads to the result that the P-value is even greater than the posterior probability of $P(\theta \le \theta_0 \mid x)$. (see Casella and Berger, 1987 for similar results in location family distributions). This is recommended using a different strategy between one sided testing and sharp null hypothesis testing.

REFERENCES

1. Bartlett, M. S. (1957). A Comment on D. V. Lindley's Statistical Paradox. *Biometrika,* 44, 533: 534.

2. Bas, E. M. and Sanchez, J. A. C. (1989). Testing a Point Null Hypothesis: Asymptotic Robust Bayesian Analysis with Respect to the Prior Given on a Subsigma Field. *International Statistical Review,* 57, 221: 232.

3. Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis.* {2nd ed.} Berlin: Springer-Verlag.

4. Berger, J. O. and Delampady, M. (1987). Testing Precise Hypotheses. *Statist. Sci.,* 2, 317: 352.

5. Berger, J. O. and Mareta, J (1994). Robust Bayesian hypothesis testing in the presence of nuisance parameters, in *J. Statist. Planning and Inferences,* 40, 357: 373.

6. Berger, J. O. and Sellke, T. (1987). Testing a Point Null hypotheses: The Irreconciliability of P-values and Evidence (with discussion), J. Am. Statist. Assoc., 82, 112: 139.

7. Casella, G. and Berger, R. L. (1987). Reconciling Bayesian and Frequentist Evidence in the One-Sided Testing Problem, *J. Am. Statist. Assoc.,* 82, 106: 111.

8. Dickey, J. M. (1977). Is the tail area useful as an approximate Bayes factor? *J. Amer. Statist. Assoc.*, 72: 138-142.

9. Edwards, W., Lindman, H. and Savage, J. (1963). Bayesian Statistical Inference for Psychological Research, *Psychol. Rev.*, 70-3, 193: 242.

10. Good, I. J (1987). Comment on Testing a Point Null hypotheses: The Irreconcilability of P-values and Evidence, by Berger, J. O. and Sellke, T., *J. Am. Statist. Assoc.*, 82, 125: 128.

11. Jeffreys, H. (1961). *Theory of Probability,* 3rd ed. London: Oxford University Press.

12. Kass, R. E. and Raftery, A. E. (1995). Bayes factors, *J. Amer. Statist. Assoc.*, 90, 773-796.

13. Kemptorne, O. (1976). Of what use are tests of significance and tests of hypothesis? *Communication in Statistics, Theory and Methods,* A5(8), 1976.

14. Lee, P. M. (1996). *Bayesian Statistics: An Introduction*, London: Arnold.

15. Lindley, D. V. (1957). A Statistical Paradox, *Biometrika,* 64, 207: 213.

16. Pratt, J. W. (1965). Bayesian Interpretation of Standard Inference Statements. (with Discussion), *J. Roy. Statist. Soc., Ser. B*, 27, 169: 203.

17. Shafer, G. (1982). Lindley Paradox, *J. Am. Statist. Assoc.,*87, 325: 351.