# Notes on Bayesian Prevalence

June 30, 2019

## Introduction

We consider a population of units (participants or spike-sorted single neuron spike trains) which are of two types. Within the population, a proportion $\gamma$ possess some definable effect, while the proportion of units in the population who do not possess this effect is $1-\gamma$. The *prevalence* of the defined effect within the population is $\gamma$, $(0 < \gamma < 1)$. A random sample of $n$ units is selected from the population and each unit undergoes a test procedure, in which the presence of the defined effect is investigated using a significance test. It is assumed that for each unit the significance level of the test is $a$ (*1 - specificity*) and the power of the test (*sensitivity*) is $b$. Thus, the probability that a randomly selected unit from the population who does not possess the defined effect will produce a significant result is $a$, while the probability that a randomly selected unit from the population who does possess the defined effect will produce a significant result is $b$.

A binary variable – *shows a significant effect* or *does not show a significant effect* is recorded for each unit in the sample, and we suppose that the total number of units who show a significant effect, out of the $n$ tested, is $k$. Let $\theta$ be the probability that a randomly selected unit from the population will show a significant effect. Then

$$\theta = (1-\gamma)a + \gamma b = a + (b-a)\gamma. \tag{1}$$

We will develop the modelling in terms of the parameter $\theta$, and later use (1) to find appropriate results in terms of the prevalence, $\gamma$.

## Modelling

Assuming that the test results on the performance of the units are independent and that the parameter $\theta$ is the same for all units in the population. Let the random variable $X$ denote the number of units out of the $n$ tested which show a significant effect at significance level $a$. Then $X$ follows a binomial distribution and

$$\Pr(X = k|\theta) = \binom{n}{k}\theta^k(1-\theta)^{n-k}, \quad k = 0, 1, \ldots, n, \quad (0 < \theta < 1). \tag{2}$$

We now define a prior distribution to characterise the prior uncertainty about $\theta$. First, we note that under the uncontroversial assumption that $b > a$, we find from (1) that $\theta > a$. Also, since $\gamma < 1$, we find that $\theta < b$. The claim regarding the assumption that $b > a$ is perfectly reasonable

since it would make no sense to employ a test procedure for which the power is less than the significance level. It follows that $a < \theta < b$.

The conjugate prior for $\theta$ is the beta distribution so, bearing in mind the constraint on $\theta$, we assume that the prior distribution for $\theta$ is the following truncated beta distribution with probability density function

$$\begin{aligned} p(\theta|a,b,r,s) &= \frac{1}{B(r,s)} \frac{\theta^{r-1}(1-\theta)^{s-1}}{[F(b;r,s) - F(a;r,s)]}, \quad a < \theta < b, \quad (r > 0, s > 0), \\ &\equiv \frac{\text{Beta}(r,s)}{[F(b;r,s) - F(a;r,s)]} \end{aligned} \tag{3}$$

where $F(x;r,s)$ is the cumulative distribution function (cdf) of $\theta$ given by the following beta cdf,

$$F(x;r,s) = \frac{1}{B(r,s)} \int_0^x \theta^{r-1}(1-\theta)^{s-1}\, dt \tag{4}$$

$\text{Beta}(r,s)$ is the pdf of the beta distribution and $B(r,s)$ is the beta function, both having parameters $r, s$. The selection of values for the parameters $r, s$ depends on prior information about $\theta$. In the absence of any prior information about $\theta$ we will use the choice $r = 1, s = 1$ in practical applications, while keeping the notation general in the formulation. This corresponds to the *a priori* assumption that the prior uncertainty regarding $\theta$ can be represented by a uniform distribution on the interval $(a,b)$.

We define $m_1 \equiv k + r, m_2 \equiv n - k + s$. Combination of the likelihood in (2) with the prior in (3) by means of Bayes' theorem gives the posterior probability density function for $\theta$ as

$$p(\theta|k,a,b,r,s) \propto \theta^{m_1-1}(1-\theta)^{m_2-1}, \quad a < \theta < b,$$

and so the posterior p.d.f. is the truncated beta distribution

$$p(\theta|k,a,b,r,s) = \frac{\text{Beta}(m_1,m_2)}{[F(b;m_1,m_2) - F(a;m_1,m_2)]}, \quad a < \theta < b. \tag{5}$$

In the sequel, the cdf and its inverse - the quantile function - for the truncated beta distribution will be required so we now provide expression for these functions: C(x) for the cdf and Q(p) for the quantile function.

$$C(x) = \Pr(\theta < x) = \int_a^x p(\theta|k,a,b,r,s)\, d\theta = \frac{F(x;m_1,m_2) - F(a;m_1,m_2)}{F(b;m_1m_2) - F(a;m_1,m_2)}. \tag{6}$$

Suppose that we wish to find the $p$th quantile, $x$, of the truncated beta distribution in (5). That is: we wish to solve the equation

$$C(x) \equiv \int_a^x p(\theta|k,a,b,r,s)\, d\theta = p. \tag{7}$$

Then using (6), we may write this equation as

$$F(x;m_1,m_2) = (1-p)F(a;m_1,m_2) + pF(b;m_1,m_2) \tag{8}$$

and so

$$x = F^{-1}\left[(1-p)F(a;m_1,m_2) + pF(b;m_1,m_2)\right].$$

Thus, the quantile function for the $p$th quantile of the truncated beta distribution is

$$Q(p) = F^{-1}\left[(1-p)F(a;m_1,m_2) + pF(b;m_1m_2)\right], \tag{9}$$

where as before $F$ is the cdf of the beta distribution given in (4).

## Applications

We now derive some applications of the truncated beta distribution from (5) in relation to the prevalence, $\gamma$.

### Posterior distribution of $\gamma$

Using the standard result for transforming random variables, we find that the posterior p.d.f. for $\gamma$ is

$$p(\gamma|k,a,b,r,s) = c[a+(b-a)\gamma]^{m_1-1}[1-a-(b-a)\gamma]^{m_2-1}, \quad (0<\gamma<1), \tag{10}$$

where the constant $c$ has the form

$$c = \frac{b-a}{B(m_1,m_2)[F(b;m_1,m_2) - F(a;m_1,m_2)]}. \tag{11}$$

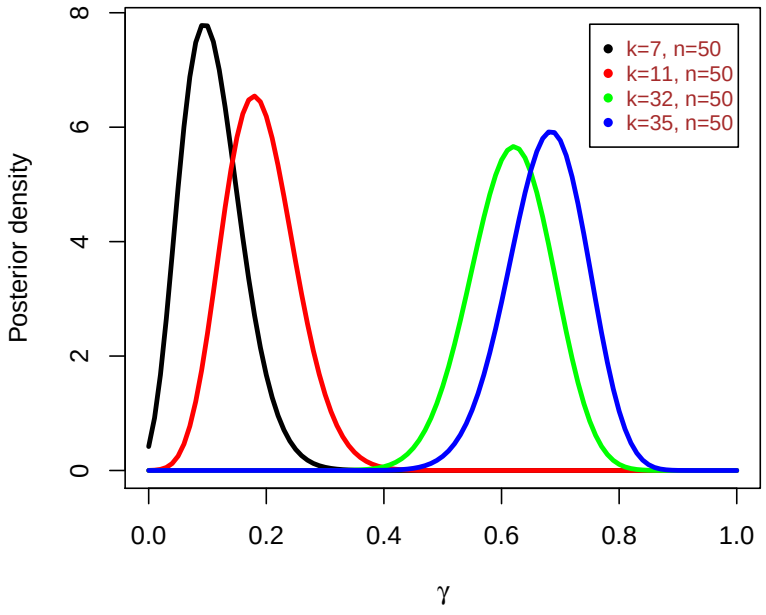


Figure 1: Posterior pdfs of population prevalence for four different choices of the values of $(k,n)$, where $k$ is the number of units, out of $n$ tested, which show a significant result.

Figure 1 shows some posterior pdfs for $\gamma$.

### Lower bound for $\gamma$

We can determine a lower bound , $\gamma_c$, for the prevalence by exploiting the relationship between $\theta$ and $\gamma$ from (1) in the form

$$\gamma = \frac{\theta - a}{b - a} \tag{12}$$

and we note that

$$\gamma \geq \gamma_c \iff \theta \geq \theta_c \equiv a + (b - a)\gamma_c. \tag{13}$$

Then from (1) and (7),

$$\Pr(\gamma \geq \gamma_c) = \Pr(\theta \geq \theta_c) = 1 - \Pr(\theta < \theta_c) \equiv 1 - C(\theta_c).$$

Using (5) we first find a posterior interval for $\theta$ of the form $(\theta_c, 1)$ which has posterior probability $p$ by solving

$$\int_{\theta_c}^{b} p(\theta|k, a, b, r, s)\, d\theta = p,$$

which can be written as

$$C(\theta_c) = 1 - p,$$

so that from (9)

$$\theta_c = Q(1 - p).$$

Then from (13) we find the corresponding lower bound for $\gamma$ as

$$\gamma_c = \frac{\theta_c - a}{b - a} \tag{14}$$

### MAP estimate of $\gamma$

The MAP estimate is the posterior mode for $\gamma$. It is given by

$$\begin{cases} 0 & \hat{\theta} \leq a \\ \frac{\hat{\theta}-a}{b-a} & a < \hat{\theta} < b, \quad \text{where} \quad \hat{\theta} = \frac{m_1 - 1}{m_1 + m_2 - 2} \\ 1 & \hat{\theta} \geq b \end{cases}$$

When $r = 1, s = 1, \hat{\theta} = k/n$.

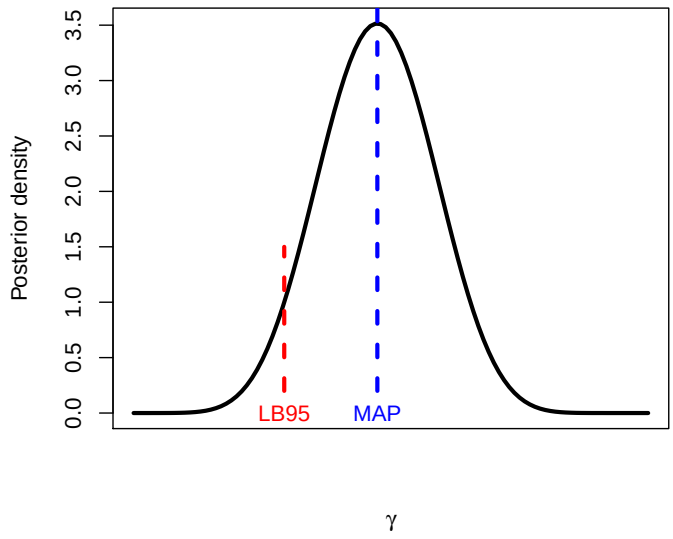


Figure 2: An illustration of the 0.95 lower bound for $\gamma$, denoted by LB95, as well as the MAP estimate of $\gamma$, when $k = 10$ units, out of a total of $n$ units, show a significant result.

An illustration of a 0.95 lower bound as well as a MAP estimate are shown in Figure 2.

### Highest posterior density interval for $\gamma$

Depending on the shape of the posterior pdf for $\gamma$, the HPDI can take several forms. It could be (i) a two-sided interval, (ii) a one-sided interval or (iii) a set of disjoint intervals. Case (i) happen when the posterior pdf is unimodal and the mode occurs when $\gamma$ is neither 0 nor 1. Case (ii) occurs when posterior mode occurs when $\gamma = 0$ or when $\gamma = 1$. Case (iii) is not relevant here but it occurs when the posterior pdf is multimodal. We focus on Case (i). Then the HPDI with posterior probability $p$ is the shortest interval of values of $\gamma$ for which the posterior probability that $\gamma$ lies between the endpoints of this interval is equal to $p$. We assume that $r = 1, s = 1$.

We first find the HPDI for $\theta$ which has posterior probability $p$, and then use relation (12) to derive the corresponding interval for $\gamma$. Mathematically, it is required to find endpoints $e_1, e_2$ for $\theta$ such that

$$\begin{aligned} C(e_2) - C(e_1) &= p, \\ p(e_2) - p(e_1) &= 0, \end{aligned}$$

where $C$ is the cdf of the truncated beta distribution defined in (7) and $p(e)$ is the posterior pdf for $\theta$ in (5) evaluated at $\theta = e$, with $r = 1, s = 1$.

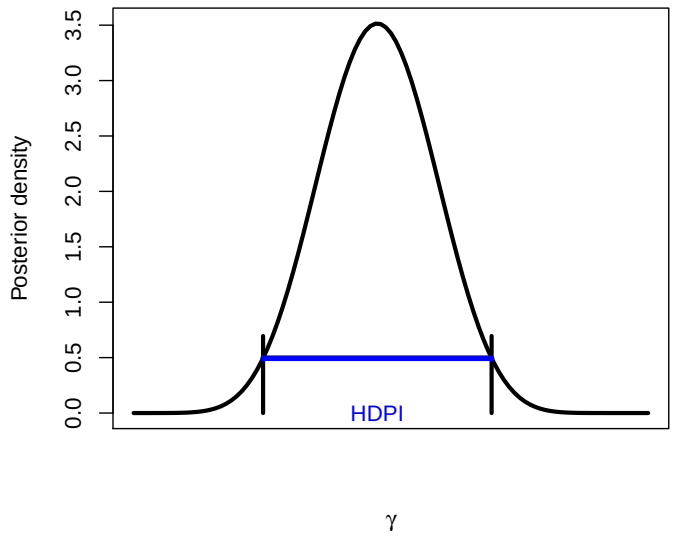


Figure 3: The HPDI when $k = 10$ and $n = 20$.

The HPDI for $\gamma$ is then computed using (12). One-sided intervals occur when $k = 0$ or $k = n$ or if the HPDI for $\theta$ has a left-hand endpoint less than or equal to $a$ or a right-hand endpoint that is greater than or equal to $b$. An illustration is shown in Figure 3.

### Sampling distribution

For a given unknown population prevalence $\gamma$, there will be variation in the number $k$ of significant results obtained from repeated sets of tests in which $n$ units are tested. Various statistics, such as (i) the length of the HDPI for $\gamma$ (ii) the MAP estimate of $\gamma$ and (iii) a lower bound for $\gamma$, are all subject to this sampling variation. It is useful then to consider the sampling distribution of each of these statistics and then compute its mean and standard deviation.

For a given number $n$ of units, the value of $k$ can be anything from 0 to $n$, with probability distribution given in (2). Let $S$ be a statistic of interest which has value $s_k$ when $k$ out of $n$ tests are significant $(k = 0, 1, \ldots, n)$. Then the mean value of $S$ is

$$\mu_n = \sum_{k=0}^{n} \Pr(X = k|\theta) s_k \tag{15}$$

and the standard deviation of $S$ is

$$\sigma_n = \sqrt{\sum_{k=0}^{n} \Pr(X = k|\theta)(s_k - \mu_n)^2}. \tag{16}$$

Formulae (15), (16) are then applied by taking $S$ in turn to be (i) the length of the HDPI for $\gamma$, (ii) the MAP estimate of $\gamma$ and (iii) a lower bound for $\gamma$, or any other relevant statistic.