TAIWANESE JOURNAL OF MATHEMATICS
Vol. 13, No. 6B, pp. 1867-1883, December 2009
This paper is available online at http://www.tjm.nsysu.edu.tw/

POST-DATA ESTIMATION OF BINOMIAL TESTING POWER

Chih-Chien Tsai* and C. Andy Tsao

Abstract. Under a robust Bayesian framework, we study the problem of power estimation in binomial hypothesis testing problems. For testing the binomial mean, the observed powers are maximum likelihood estimates of power functions for one-sided and two-sided problems. With respect to these functions, we show that the observed powers are too large when the data is significant for some beta families of priors. A practical implication about the problem of sample size calculation is discussed.

1. INTRODUCTION

Despite the popularity of hypothesis tests, the associated classical pre-data measures of performance such as significance level or power are not very satisfactory. The practitioners often request supplemental post-data evidential measures once the data is collected. The p-value is one of such measures. It is often taught as a data sensitive evidence against null hypothesis in standard statistical courses. However, whether this is a valid interpretation of the p-value is still subject to study and motivates active researches. See, for example, Berger and Delampady (1987), Berger and Sellke (1987), Casella and Berger (1987), Hwang, et al (1992) and more recent Oh and DasGupta (1999), Sellke, et al (2001), Tsao (2006). The observed power is another emerging measure and has been required in some subject domains as measures of the strength of the experiments. Unfortunately, comparing with the p-values, its theoretical properties are even less understood. See Gillett (1996), Hoenig and Heisey (2001), Tsao and Tseng (2006b) and references therein. Particularly, Gillett (1996) and Tsao and Tseng (2006b) suggest that the usual observed power, as a plug-in estimate of power function, tends to be too large. Furthermore, because the

Received January 22, 2007, accepted January 11, 2008.
Communicated by Yuh-Jia Lee.
2000 *Mathematics Subject Classification*: Primary 62F03, 62F35; Secondary 62F15, 62C10.
Key words and phrases: Bayesian robustness analysis, Beta prior, Observed power, Power function, Bayes estimates, Binomial hypothesis testing, Sample size calculation.
*Corresponding author.

power estimation is usually used in the sample size calculation in the experimental design, this problematic feature in turn leads to unreliable sample size calculation. See, comments and guidances in Lenth (2001).

Motivated by Hoenig and Heisey (2001) and the related works in p-values, for example, Berger and Sellke (1987), Berger and Delampady (1987) and Casella and Berger (1987), Tsao and Tseng (2006b) examines the validity of the observed power under a robust Bayesian framework. They show that the observed power for testing normal means overestimates the power for significant data yet underestimates the power for insignificant data. Some misinterpretations of the observed power are also discussed in their study. In this paper, we investigate the similar problem under binomial setting. The observed powers can be considered as maximum likelihood estimates of power functions. With respect to these functions, we show that the observed powers are too large when the data is significant for some beta families of priors. This conclusion holds both for one-sided and two-sided hypothesis testing problems.

The rest of the paper is organized as follows. In Section 2, we will present the formulation of problem. Some families of priors will be defined. Section 3 contains the main results on the bounds of Bayes estimates. Along with numerical calculations detailed in Section 4, it is shown that the usual observed power tends to be too large. A practical implication about the problem of sample size calculation is discussed in Sections 3 and 4. Finally, the conclusion and discussion are summarized in Section 5.

2. Formulation

Let $X \sim Bin(n, \theta)$ with unknown $0 \leq \theta \leq 1$. Consider the hypothesis testing problems:

(1) $$H_0 : \theta \leq \theta_0 \quad vs. \quad H_1 : \theta > \theta_0,$$

(2) $$H_0 : \theta = \theta_0 \quad vs. \quad H_1 : \theta \neq \theta_0.$$

The hypothesis testing problems (1) and (2), are well-studied, see, for example, Hollander and Wolfe (1973), Conover (1980), and Gibbons (1985). Here we consider the usual nonrandomized exact tests $\Phi_1(X)$, uniformly most power (UMP) level α test for (1), and $\Phi_2(X)$, uniformly most power unbiased (UMPU) level α test for (2). Specifically,

(3) $$\Phi_1(X) = \begin{cases} 1 & \text{if } X \geq c_0 \\ 0 & \text{otherwise} \end{cases}$$

where the smallest $c_0 \in \{0, 1, \cdots, n\}$ satisfies $P_{\theta_0}(X \geq c_0) \leq \alpha$.

$$\Phi_2(X) = \begin{cases} 1 & \text{if } X \leq c_1,\ X \geq c_2 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where the largest c_1 and the smallest $c_2 \in \{0, 1, \cdots, n\}$ satisfy $P_{\theta_0}(X \leq c_1) \leq \alpha/2$ and $P_{\theta_0}(X \geq c_2) \leq \alpha/2$. Their power functions for $\Phi_1(X)$ and $\Phi_2(X)$ are

$$\begin{aligned} \beta_1(\theta) &= P_\theta(X \geq c_0) \\ &= \sum_{i=c_0}^{n} C(n,i)\theta^i(1-\theta)^{n-i} \end{aligned} \tag{5}$$

and

$$\begin{aligned} \beta_2(\theta) &= P_\theta(X \leq c_1) + P_\theta(X \geq c_2) \\ &= \sum_{i=0}^{c_1} C(n,i)\theta^i(1-\theta)^{n-i} + \sum_{i=c_2}^{n} C(n,i)\theta^i(1-\theta)^{n-i}, \end{aligned} \tag{6}$$

respectively. The usual observed powers $\beta_1(\hat{\theta})$ and $\beta_2(\hat{\theta})$ are simply substituting θ in power functions (5) and (6) by its maximum likelihood estimate $\hat{\theta} = \bar{x} = x/n$. Because of invariance property of maximum likelihood estimate, they can be viewed, respectively, as maximum likelihood estimates of power functions $\beta_1(\theta)$ and $\beta_2(\theta)$.

Here we study the problem of estimating power functions $\beta_1(\theta)$ and $\beta_2(\theta)$ from a robust Bayesian viewpoint. We compare the observed powers with Bayes estimates with respect to conjugate families of priors. The parameter of binomial distribution θ is assumed to have a beta prior, i.e.

$$\pi(\theta) = \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

where $0 < \theta < 1$ and α, $\beta > 0$ and $B(\alpha,\beta) = \int_0^1 \theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta$. The posterior density is

$$\pi(\theta|x) = \frac{1}{B(\alpha+x,\beta+y)}\theta^{\alpha+x-1}(1-\theta)^{\beta+y-1},$$

a $Beta(\alpha+x,\beta+y)$ density, where x is a realization of random variable X and $y = n - x$. Under a squared error loss, the Bayes estimates for power functions $\beta_1(\theta)$ and $\beta_2(\theta)$ are

$$\begin{aligned} E_{\pi(\theta|\bar{x})}\beta_1(\theta) &= \int_0^1 \beta_1(\theta)\pi(\theta|x)d\theta \\ &= \sum_{i=c_0}^{n} C(n,i)\frac{B(\alpha+x+i,\beta+y+n-i)}{B(\alpha+x,\beta+y)} \end{aligned} \tag{7}$$

and

$$
\begin{aligned}
E_{\pi(\theta|\bar{x})}\beta_2(\theta) &= \int_0^1 \beta_2(\theta)\pi(\theta|x)d\theta \\
&= \sum_{i=0}^{c_1} C(n,i)\frac{B(\alpha+x+i,\beta+y+n-i)}{B(\alpha+x,\beta+y)} \\
&\quad + \sum_{i=c_2}^{n} C(n,i)\frac{B(\alpha+x+i,\beta+y+n-i)}{B(\alpha+x,\beta+y)}. \qquad (8)
\end{aligned}
$$

Our problem is to compare the observed power functions $\beta_1(\hat{\theta})$ and $\beta_2(\hat{\theta})$ with

$$\sup_{\pi\in\Gamma} E_{\pi(\theta|\bar{x})}\beta_i(\theta), \quad i = 1, 2,$$

where Γ is the family of beta priors or some of its subsets. Note that if the supremum of Bayes estimates is less than the observed power, it implies that the observed power is too large/optimistic to estimate the power function.

Here we use an alternative parameterization of beta prior: let $p_0 = \frac{\alpha}{\alpha+\beta}$ and some constant $c > 0$ such that $\alpha = cp_0$, $\beta = c(1-p_0)$. Hence

$$E_\pi\theta = p_0 \qquad \text{and} \qquad Var_\pi\theta = \frac{p_0(1-p_0)}{c+1}.$$

Then if π has a $Beta(cp_0, c(1-p_0))$ density, thus $\pi(\theta|x)$ has a $Beta(\alpha_c, \beta_c)$ density where $\alpha_c = cp_0 + x$ and $\beta_c = c(1-p_0) + y$.

The families of beta priors are denoted as

$$\Gamma(p_0) = \{\pi:\ \pi \sim Beta(cp_0, c(1-p_0)),\ c > 0\} \quad \text{for } p_0 \in (0,1);$$

$$\Gamma_c = \{\pi:\ \pi \sim Beta(cp_0, c(1-p_0)),\ p_0 \in (0,1)\} \quad \text{for } c > 0.$$

In this study, we focus on the prior family $\Gamma(p_o)$ for $p_0 = 1/2$ (also as $\Gamma(1/2)$). The choice $p_0 = 1/2$ reflects the noninformativeness or unbiasedness regarding the parameter θ. Note that $\Gamma(1/2)$ includes important priors such as uniform prior $(Beta(1,1),\, c = 2)$ and reference prior $(Beta(1/2, 1/2),\ c = 1)$, etc. The corresponding Bayes estimates are denoted as $E_U\beta_i(\theta)$ and $E_R\beta_i(\theta)$, $i = 1, 2$, respectively. There is an interesting connection between our approach and Walley (1996). If the number of categories is two, Walley's imprecise Dirichlet model approach is similar to deriving bounds of the Bayes estimates with priors belonging to the family Γ_c with $c = 1$ or 2. Restriction of $c = 1$ or 2 might be unreasonable, as commented by O'Hagan (1996), Jennison (1996) and Levi (1996). In contrast, we derive the bounds of Bayes estimate with $\pi \in G(p_0)$ and $p_0 = 1/2$ allowing c to vary. This approach calls for more involved derivation. The interested readers are referred to Tsay and Tsao (2003), Tsao and Tseng (2004) for details.

3. Main Results

We characterize the ordering of $E_{\pi(\theta|x)}\beta_i(\theta), i = 1, 2$ under mild conditions. In turn, the supremum of Bayes estimates with respect to priors in $\Gamma(1/2)$ can be readily obtained.

3.1. One-sided Problems

Firstly, we establish the monotonicity of $\beta_1(\theta)$.

Lemma 1. *The power function $\beta_1(\theta)$ is strictly increasing in θ for $0 < \theta < 1$.*

Proof. The proof follows by noting that

$$\beta_1(\theta) = P(X \geq c_0) = P(Z \leq \theta) \tag{9}$$

where $X \sim Bin(n, \theta)$ and $Z \sim Beta(c_0, n-c_0+1)$, cf. Casella and Berger (1990), p449, for example. Then, it is clear that $\beta_1(\theta)$ is strictly increasing in θ. ■

The following proposition establishes the ordering of the Bayes estimates of power function $E_{\pi(\theta|x)}\beta_1(\theta)$. Recall that x denotes a realization of random variable X and $y = n - x$.

Proposition 1. *If $\pi \in \Gamma(1/2)$ and $x \geq y$, $E_{\pi(\theta|x)}\beta_1(\theta)$ is decreasing in c for $c > 0$.*

Proof. We denote $\pi(\theta|c) = \pi(\theta|c, x)$ where the dependence of x is notationally suppressed. Let $\pi \in \Gamma(1/2)$ with

$$\pi(\theta|c) = \frac{1}{B(\alpha_c, \beta_c)} \exp\left(\frac{c}{2}\ln(\theta(1-\theta))\right) h(\theta)$$

where $h(\theta) = \theta^{x-1}(1-\theta)^{y-1}$. It is easy to see that $\pi(\theta|c)$ has the monotone likelihood ratio property (MLR) in $T(\theta) = \ln(\theta(1-\theta))$ since c is strictly increasing. By Lemma 2 on p. 85 of Lehmann (1986), the proof is completed if we can write $E_{\pi(\theta|x)}\beta_1(\theta) = \int H(t) f_T(t|c)dt$ with $f_T(t|c)$ being the density function of $\ln(\theta(1-\theta))$ when θ has posterior $\pi(\theta|c)$ and $H(t)$ being a decreasing function in t.

First, $f_T(t|c)$ can be derived easily by noting that for $t < -\ln 4$

$$\begin{aligned} P_c(T \leq t) &= P_c[\ln(\theta(1-\theta)) \leq t] \\ &= P_c(\theta \leq \frac{1}{2}(1 - \sqrt{1-4e^t})) + P_c(\theta \geq \frac{1}{2}(1 + \sqrt{1-4e^t})) \end{aligned}$$

where P_c denotes the probability calculated under the density $\pi(\theta|c)$. Thus the density of T is

$$f_T(t|c) = \frac{e^t}{\sqrt{1-4e^t}}[\pi(\theta_t|c) + \pi(1-\theta_t|c)]$$

with $t < -\ln 4$ and $0 < \theta_t = (1-\sqrt{1-4e^t})/2 < 1/2$. Let $D(\theta_t) = \frac{h(\theta_t)}{h(\theta_t)+h(1-\theta_t)}$ and follow similar steps in Tsao and Tseng (2004), it is left to show that

$$H(\theta_t) = \beta_1(\theta_t)D(\theta_t) + \beta_1(1-\theta_t)D(1-\theta_t)$$

is decreasing in t for all $t \in (-\infty, -\ln 4)$. Thus it suffices to show $\frac{d}{d\theta}H(\theta) \leq 0$ for $0 < \theta < 1/2$ since

$$\frac{d}{dt}H(\theta_t) = \frac{d}{d\theta_t}H(\theta_t)\frac{d}{dt}\theta_t \quad \text{and} \quad \frac{d}{dt}\theta_t = \frac{e^t}{\sqrt{1-4e^t}} > 0.$$

Note that $D(1-\theta) = 1 - D(\theta)$, hence

$$\begin{aligned} \tfrac{d}{d\theta}H(\theta) &= D(\theta)\frac{d}{d\theta}\beta_1(\theta) + \beta_1(\theta)\frac{d}{d\theta}D(\theta) \\ &\quad + \beta_1(1-\theta)\tfrac{d}{d\theta}\left(1-D(\theta)\right) - \beta_1(1-\theta)\tfrac{d}{d\theta}D(\theta) \\ &= \left(\beta_1(\theta) - \beta_1(1-\theta)\right)\frac{d}{d\theta}D(\theta) \\ &\quad + \left[D(\theta)\frac{d}{d\theta}\beta_1(\theta) + D(1-\theta)\frac{d}{d\theta}\beta_1(1-\theta)\right]. \end{aligned} \tag{10}$$

Now, for $0 < \theta < 1/2$, Lemma 1 implies $\beta_1(\theta) - \beta_1(1-\theta) < 0$ and easy calculations show

$$\frac{d}{d\theta}D(\theta) = \frac{\theta^{n-3}(1-\theta)^{n-3}}{[h(\theta)+h(1-\theta)]^2}(1-2\theta)(x-y) \geq 0.$$

The last inequality holds since it is assumed that $x \geq y$. Hence

$$\left(\beta_1(\theta) - \beta_1(1-\theta)\right)\frac{d}{d\theta}D(\theta) \leq 0. \tag{11}$$

Next, recalling (9), then

$$\frac{d}{d\theta}\beta_1(\theta) = \frac{1}{B(c_0, n-c_0+1)}\theta^{c_0-1}(1-\theta)^{n-c_0+1-1}$$

and similarly

$$\frac{d}{d\theta}\beta_1(1-\theta) = \frac{-1}{B(c_0, n-c_0+1)}(1-\theta)^{c_0-1}\theta^{n-c_0+1-1}.$$

Combining these results and definitions of $D(\theta)$ and $h(\theta)$, we have

$$D(\theta)\frac{d}{d\theta}\beta_1(\theta)+D(1-\theta)\frac{d}{d\theta}\beta_1(1-\theta)=\frac{\theta^{-2}(1-\theta)^{-2}}{B(c_0,n-c_0+1)\,[h(\theta)+h(1-\theta)]}G(\theta)$$

where $G(\theta)=\theta^{c_0+x}(1-\theta)^{2n-c_0-x+1}-\theta^{2n-c_0-x+1}(1-\theta)^{c_0+x}$. Since $2(c_0+x)\geq 2n+1$, we have

$$\theta^{2(c_0+x)-2n-1}\leq(1-\theta)^{2(c_0+x)-2n-1}\quad\text{for}\quad 0<\theta<1/2.$$

Cross-multiplying the numerators and denominators then rearranging the terms, we have

$$G(\theta)=\theta^{c_0+x}(1-\theta)^{2n-c_0-x+1}-\theta^{2n-c_0-x+1}(1-\theta)^{c_0+x}\leq 0.$$

Hence,

$$D(\theta)\frac{d}{d\theta}\beta_1(\theta)+D(1-\theta)\frac{d}{d\theta}\beta_1(1-\theta)\leq 0. \tag{12}$$

Combining (10), (11) and (12), it is shown that $\frac{d}{d\theta}H(\theta)\leq 0$ for $0<\theta<1/2$. Thus this completes the proof. ∎

Note that although the power function $\beta_1(\theta)$ seems similar to the probability of winning in the division problem discussed in Tsao and Tseng (2004), Proposition 1 does not easily follow from the similar theorem therein. Particularly, the proof of main theorem in Tsao and Tseng (2004) relies on the crucial condition $m=a+b-1$. In contrast, that condition is no longer needed in our proof. Despite their resemblance, the proof in the above proposition calls for different calculation.

3.2. Two-sided Problems

Lemma 2. *If $c_1+c_2=n$, we have $\beta_2(\theta)=\beta_2(1-\theta)$ for $0<\theta<1$.*

The proof of Lemma 2 is straightforward and skipped.

The technical condition $c_1+c_2=n$ is satisfied in all our calculated examples for the two-sided hypothesis testing problems (2) with $\theta_0=1/2$. Note that the condition imposes a symmetric-like condition on the cutoff points $c_1<c_2$. For the symmetric hypothesis testing problem where $\theta_0=1/2$, it is a natural condition and suitable c_1,c_2 can be found.

Lemma 3. *If $c_1+c_2=n$, $\beta_2(\theta)$ is decreasing in θ for $0<\theta<1/2$ and increasing in θ for $1/2<\theta<1$.*

Proof. Note that

$$\begin{aligned}\beta_2(\theta) &= P(X \le c_1) + P(X \ge c_2)\\ &= 1 - P(X \ge c_1+1) + P(X \ge c_2)\\ &= 1 - P(Z_1 \le \theta) + P(Z_2 \le \theta)\end{aligned}$$

where $X \sim Bin(n,\theta)$ and $Z_1 \sim Beta(c_1+1, n-(c_1+1)+1)$, $Z_2 \sim Beta(c_2, n-c_2+1)$. Since $c_1+c_2 = n$, $Z_1 \sim Beta(c_1+1, c_2)$, $Z_2 \sim Beta(c_2, c_1+1)$. Hence,

$$\begin{aligned}&\frac{d}{d\theta}\beta_2(\theta)\\ &= -\frac{d}{d\theta}P(Z_1 \le \theta) + \frac{d}{d\theta}P(Z_2 \le \theta)\\ &= \frac{1}{B(c_2, c_1+1)}\theta^{c_2-1}(1-\theta)^{c_1+1-1} - \frac{1}{B(c_1+1, c_2)}\theta^{c_1+1-1}(1-\theta)^{c_2-1}\\ &= \frac{1}{B(c_2, c_1+1)}\left[\theta^{c_2-1}(1-\theta)^{c_1} - \theta^{c_1}(1-\theta)^{c_2-1}\right].\end{aligned} \tag{13}$$

Since $c_1 < c_2$ and for $0 < \theta < 1/2$ we have

$$\left(\frac{\theta}{1-\theta}\right)^{c_2-1} \le \left(\frac{\theta}{1-\theta}\right)^{c_1}.$$

Cross-multiplying the numerators and denominators then rearranging the terms, we have

$$\theta^{c_2-1}(1-\theta)^{c_1} - \theta^{c_1}(1-\theta)^{c_2-1} \le 0.$$

That is $\frac{d}{d\theta}\beta_2(\theta) \le 0$ for $\theta \in (0, 1/2)$ and $\frac{d}{d\theta}\beta_2(\theta) \ge 0$ for $\theta \in (1/2, 1)$. This completes the proof. ∎

This proposition establishes the ordering of the Bayes estimate of power function $E_{\pi(\theta|x)}\beta_2(\theta)$.

Proposition 2. *If $\pi \in \Gamma(1/2)$ and $c_1 + c_2 = n$, $E_{\pi(\theta|x)}\beta_2(\theta)$ is decreasing in c for $c > 0$.*

Proof. The proof is similar to that of Proposition 1. It suffices to show that $\frac{d}{d\theta}H(\theta) \le 0$ for $0 < \theta < 1/2$. Using $\beta_2(\theta)$ in (10) for two-sided problem instead of $\beta_1(\theta)$, we have

$$\begin{aligned}\frac{d}{d\theta}H(\theta) &= (\beta_2(\theta) - \beta_2(1-\theta))\frac{d}{d\theta}D(\theta)\\ &\quad + \left[D(\theta)\frac{d}{d\theta}\beta_2(\theta) + D(1-\theta)\frac{d}{d\theta}\beta_2(1-\theta)\right].\end{aligned} \tag{14}$$

Now, for $0 < \theta < 1$, Lemma 2 implies $\beta_2(\theta) - \beta_2(1-\theta) = 0$ if $c_1 + c_2 = n$ and thus

$$(\beta_2(\theta) - \beta_2(1-\theta))\, \frac{d}{d\theta} D(\theta) = 0. \tag{15}$$

Next, recalling (13) and similarly

$$\begin{aligned} \frac{d}{d\theta}\beta_2(1-\theta) &= \frac{-1}{B(c_2, c_1+1)} \left[(1-\theta)^{c_2-1}\theta^{c_1} - (1-\theta)^{c_1}\theta^{c_2-1}\right] \\ &= \frac{d}{d\theta}\beta_2(\theta). \end{aligned}$$

Combining the result $\frac{d}{d\theta}\beta_2(1-\theta) = \frac{d}{d\theta}\beta_2(\theta)$ and definitions of $D(\theta)$ and $h(\theta)$, we have

$$D(\theta)\frac{d}{d\theta}\beta_2(\theta) + D(1-\theta)\frac{d}{d\theta}\beta_2(1-\theta) = \frac{d}{d\theta}\beta_2(\theta) \le 0. \tag{16}$$

The last inequality holds since Lemma 3 implies $\beta_2(\theta)$ is decreasing for $0 < \theta < 1/2$.
Combining (14), (15) and (16), we show $\frac{d}{d\theta}H(\theta) \le 0$ for $0 < \theta < 1/2$. This completes the proof. ■

Applying Proposition 2 in Tsay and Tsao (2003), we immediately have Proposition 3 that characterizes the convergence of the Bayes estimate of power functions $E_{\pi(\theta|\bar{x})}\beta_1(\theta)$ and $E_{\pi(\theta|\bar{x})}\beta_2(\theta)$ as c goes to infinity. Its proof is omitted.

Proposition 3. *For $\pi \in \Gamma(p_0)$,*

$$\mid E_{\pi(\theta|\bar{x})}\beta_i(\theta) - \beta_i(\hat{\theta}_\pi) \mid \le K\ Var_{\pi(\theta|x)}\theta, \quad i = 1, 2,$$

where $K > 0$ is a generic constant and

$$\begin{aligned} \hat{\theta}_\pi &= E_{\pi(\theta|x)}\theta = \frac{cp_0 + x}{c+n}, \\ Var_{\pi(\theta|x)}\theta &= \frac{\hat{\theta}_\pi(1-\hat{\theta}_\pi)}{c+n+1} \\ &= \frac{(cp_0+x)(c(1-p_0)+n-x)}{(c+n)^2(c+n+1)}. \end{aligned}$$

As c goes to infinity, the limit of the Bayes estimate of power function, $E_{\pi(\theta|\bar{x})}\beta_1(\theta)(E_{\pi(\theta|\bar{x})}\beta_2(\theta))$, is $\beta_1(p_0)$ $(\beta_2(p_0))$. Besides, when the sample size n is large enough, $E_{\pi(\theta|x)}\theta$ goes to sample mean $\hat{\theta}$ and $Var_{\pi(\theta|x)}\theta$ goes to 0. Thus this proposition also suggests that the Bayes estimate of power function, $E_{\pi(\theta|\bar{x})}\beta_1(\theta)$ ($E_{\pi(\theta|\bar{x})}\beta_2(\theta)$) is close to the corresponding observed power $\beta_1(\hat{\theta})$ ($\beta_2(\hat{\theta})$) for large n.

In light of Propositions 1-3, Proposition 4 and Proposition 5 immediately follow.

Proposition 4. *If* $x \geq y$,

$$\begin{aligned} &\inf_{\pi\in\Gamma(1/2)} E_{\pi(\theta|\bar{x})}\beta_1(\theta) &= \lim\nolimits_{c\to\infty,\pi\in\Gamma(1/2)} E_{\pi(\theta|\bar{x})}\beta_1(\theta) \\ \text{and} \quad &\sup_{\pi\in\Gamma(1/2)} E_{\pi(\theta|\bar{x})}\beta_1(\theta) &= \lim\nolimits_{c\to 0,\pi\in\Gamma(1/2)} E_{\pi(\theta|\bar{x})}\beta_1(\theta). \end{aligned}$$

Unfortunately, we do not have parallel theoretical results when $x < y$. However, at least for $\theta_0 \geq 1/2$, the test (3) does not reject H_0 when $x < y$ for any $\alpha < 1/2$. From the practical viewpoint, the results for $x \geq y$ are of greater concern.

Proposition 5. *If* $c_1 + c_2 = n$,

$$\begin{aligned} &\inf_{\pi\in\Gamma(1/2)} E_{\pi(\theta|\bar{x})}\beta_2(\theta) &= \lim\nolimits_{c\to\infty,\pi\in\Gamma(1/2)} E_{\pi(\theta|\bar{x})}\beta_2(\theta) \\ \text{and} \quad &\sup_{\pi\in\Gamma(1/2)} E_{\pi(\theta|\bar{x})}\beta_2(\theta) &= \lim\nolimits_{c\to 0,\pi\in\Gamma(1/2)} E_{\pi(\theta|\bar{x})}\beta_2(\theta). \end{aligned}$$

4. Numerical Calculation

For one-sided and two-sided hypothesis testing problems, we show that the observed powers are too large when the data is significant (contrasting with Bayes estimates with respect to priors in $\Gamma(1/2)$). In addition, the sample size determination method based on the observed power fails to meet the claimed level of power for all cases we calculated. More reasonable methods of sample size determination are introduced.

4.1. Comparison with supremum of Bayes estimates

We consider the hypothesis testing problems for (1) with $\theta_0 = 1/2$, $3/4$ and (2) with $\theta_0 = 1/2$. For each case, the sample size n equals 30. Propositions 4, 5 alleviate the task of the numerical computation for finding bounds. Along with these propositions, we have the lower bounds for $E_{\pi(\theta|\bar{x})}\beta_1(\theta)$ and $E_{\pi(\theta|\bar{x})}\beta_2(\theta)$ are $\beta_1(1/2)$ and $\beta_2(1/2)$, respectively, when π belongs to conjugate priors $\Gamma(1/2)$. For one-sided problem, Table 1 and Table 2 give the estimates and bounds to $\theta_0 = 1/2$ and $3/4$, respectively. For two-sided problem with $\theta_0 = 1/2$, see Table 3. Recall that $E_U\beta_i(\theta)$ and $E_R\beta_i(\theta)$ denote the Bayes estimates with respect to uniform prior and reference prior. It can be readily seen that

Table 1. One-sided testing problem with $\theta_0 = 1/2$, $n = 30$, when $x \geq 20$ the test rejects H_0 at level $\alpha = 0.05$

n	x	$\inf_{\pi\in\Gamma(1/2)}$ $E_{\pi(\theta\vert\bar{x})}\beta_1(\theta)$	$E_U\beta_1(\theta)$	$E_R\beta_1(\theta)$	$\sup_{\pi\in\Gamma(1/2)}$ $E_{\pi(\theta\vert\bar{x})}\beta_1(\theta)$	$\beta_1(\hat{\theta})$
30	15	0.0494	0.1180	0.1199	0.1218	0.0494
30	16	0.0494	0.1750	0.1791	0.1834	0.0993
30	17	0.0494	0.2475	0.2545	0.2619	0.1790
30	18	0.0494	0.3344	0.3448	0.3559	0.2915
30	19	0.0494	0.4326	0.4466	0.4613	0.4316
30	20	0.0494	0.5371	0.5540	0.5718	0.5848
30	21	0.0494	0.6413	0.6600	0.6795	0.7304
30	22	0.0494	0.7384	0.7574	0.7766	0.8489
30	23	0.0494	0.8224	0.8398	0.8571	0.9298
30	24	0.0494	0.8893	0.9037	0.9176	0.9744
30	25	0.0494	0.9378	0.9484	0.9582	0.9933
30	26	0.0494	0.9693	0.9761	0.9821	0.9989
30	27	0.0494	0.9871	0.9909	0.9939	0.9999
30	28	0.0494	0.9957	0.9973	0.9985	1.0000
30	29	0.0494	0.9990	0.9995	0.9998	1.0000
30	30	0.0494	0.9999	1.0000	1.0000	1.0000

Table 2. One-sided testing problem with $\theta_0 = 3/4$, $n = 30$, when $x \geq 27$ the test rejects H_0 at level $\alpha = 0.05$

n	x	$\inf_{\pi\in\Gamma(1/2)}$ $E_{\pi(\theta\vert\bar{x})}\beta_1(\theta)$	$E_U\beta_1(\theta)$	$E_R\beta_1(\theta)$	$\sup_{\pi\in\Gamma(1/2)}$ $E_{\pi(\theta\vert\bar{x})}\beta_1(\theta)$	$\beta_1(\hat{\theta})$
30	15	0.0000	0.0005	0.0005	0.0005	0.0000
30	16	0.0000	0.0010	0.0011	0.0013	0.0000
30	17	0.0000	0.0023	0.0025	0.0028	0.0001
30	18	0.0000	0.0047	0.0054	0.0061	0.0003
30	19	0.0000	0.0095	0.0108	0.0124	0.0011
30	20	0.0000	0.0183	0.0210	0.0242	0.0033
30	21	0.0000	0.0337	0.0389	0.0451	0.0093
30	22	0.0000	0.0596	0.0690	0.0801	0.0241
30	23	0.0000	0.1010	0.1169	0.1357	0.0569
30	24	0.0000	0.1636	0.1889	0.2186	0.1227
30	25	0.0000	0.2531	0.2906	0.3337	0.2396
30	26	0.0000	0.3725	0.4234	0.4804	0.4194
30	27	0.0000	0.5189	0.5813	0.6482	0.6474
30	28	0.0000	0.6802	0.7467	0.8128	0.8635
30	29	0.0000	0.8332	0.8904	0.9398	0.9831
30	30	0.0000	0.9475	0.9806	1.0000	1.0000

$$\inf_{\pi\in\Gamma(1/2)} E_{\pi(\theta|\bar{x})}\beta_i(\theta) \leq E_U\beta_i(\theta) \leq E_R\beta_i(\theta) \leq \sup_{\pi\in\Gamma(1/2)} E_{\pi(\theta|\bar{x})}\beta_i(\theta),\ i = 1, 2,$$

for all cases we calculated. Note that these are Bayes estimates with respect to priors belonging to $\Gamma(1/2)$ when $c \to \infty$, $c = 2$, $c = 1$ and $c \to 0$, respectively. These numerical calculations are consistent with our theoretical findings that the Bayes estimate increases in c when the prior belonging to $\Gamma(1/2)$.

Furthermore, it is noted that

$$\sup_{\pi \in \Gamma(1/2)} E_{\pi(\theta|\bar{x})}\beta_i(\theta) \le \beta_i(\hat{\theta}),\ i = 1, 2,$$

Table 3. Two-sided testing problem with $\theta_0 = 1/2$, $n = 30$, when $x \le 9$ and $x \ge 21$ the test rejects H_0 at level $\alpha = 0.05$

n	x	$\inf_{\pi\in\Gamma(1/2)}$ $E_{\pi(\theta\|\bar{x})}\beta_2(\theta)$	$E_U\beta_2(\theta)$	$E_R\beta_2(\theta)$	$\sup_{\pi\in\Gamma(1/2)}$ $E_{\pi(\theta\|\bar{x})}\beta_2(\theta)$	$\beta_2(\hat{\theta})$
30	0	0.0428	0.9997	0.9999	1.0000	1.0000
30	1	0.0428	0.9976	0.9988	0.9995	1.0000
30	2	0.0428	0.9910	0.9943	0.9967	1.0000
30	3	0.0428	0.9751	0.9820	0.9876	0.9995
30	4	0.0428	0.9451	0.9566	0.9667	0.9958
30	5	0.0428	0.8969	0.9132	0.9284	0.9803
30	6	0.0428	0.8288	0.8490	0.8688	0.9389
30	7	0.0428	0.7424	0.7649	0.7876	0.8591
30	8	0.0428	0.6424	0.6651	0.6886	0.7384
30	9	0.0428	0.5359	0.5568	0.5787	0.5888
30	10	0.0428	0.4309	0.4484	0.4670	0.4318
30	11	0.0428	0.3351	0.3485	0.3629	0.2895
30	12	0.0428	0.2547	0.2641	0.2742	0.1771
30	13	0.0428	0.1945	0.2006	0.2071	0.1001
30	14	0.0428	0.1573	0.1612	0.1654	0.0566
30	15	0.0428	0.1448	0.1479	0.1513	0.0428
30	16	0.0428	0.1573	0.1612	0.1654	0.0566
30	17	0.0428	0.1945	0.2006	0.2071	0.1001
30	18	0.0428	0.2547	0.2641	0.2742	0.1771
30	19	0.0428	0.3351	0.3485	0.3629	0.2895
30	20	0.0428	0.4309	0.4484	0.4670	0.4318
30	21	0.0428	0.5359	0.5568	0.5787	0.5888
30	22	0.0428	0.6424	0.6651	0.6886	0.7384
30	23	0.0428	0.7424	0.7649	0.7876	0.8591
30	24	0.0428	0.8288	0.8490	0.8688	0.9389
30	25	0.0428	0.8969	0.9132	0.9284	0.9803
30	26	0.0428	0.9451	0.9566	0.9667	0.9958
30	27	0.0428	0.9751	0.9820	0.9876	0.9995
30	28	0.0428	0.9910	0.9943	0.9967	1.0000
30	29	0.0428	0.9976	0.9988	0.9995	1.0000
30	30	0.0428	0.9997	0.9999	1.0000	1.0000

are found in all cases we calculated when the data is significant. These inequalities do not follow from earlier theoretical results in Section 3. Specifically, we establish only the ordering of Bayes estimates therein but not the comparison with the observed powers. For one-sided testing problem (cf. Tables 1 and Table 2 for $\theta_0 = 1/2$ and $3/4$, respectively), when the test rejects null hypothesis at level α, the observed powers are larger than all Bayes estimates with respect to priors in $\Gamma(1/2)$. Take Table 1 for example, when $X \geq 15$ the test rejects null hypothesis at level α, the observed powers are larger than the upper bounds $\sup_{\pi\in\Gamma(1/2)} E_{\pi(\theta|\bar{x})}\beta_1(\theta)$. For two-sided testing problem (cf. Table 3), when the test rejects null hypothesis at level α, the observed powers are larger than all Bayes estimates with respect to priors in $\Gamma(1/2)$. Take Table 3 for example, when $X \leq 5$ or $X \geq 15$ the test rejects null hypothesis at level α, the observed powers are larger than $\sup_{\pi\in\Gamma(1/2)} E_{\pi(\theta|\bar{x})}\beta_2(\theta)$.

4.2. Sample size calculation

The results of extremity of observed powers have important implication on sample size calculation. For testings (1) and (2), we consider the nonrandomized exact tests (3) and (4). The objective, basically, is to determine the sample sizes such that

$$\beta_1(\theta) \geq \beta_0 \quad \text{and} \quad \beta_2(\theta) \geq \beta_0 \tag{17}$$

for one-sided and two-sided testing problems, respectively, where β_0 is a given desired level of power. In practice, there is uncertainty in the parameter θ, the sample size is usually determined by substituting in (17) an estimate of $\hat{\theta} = \bar{x} = x/n$ derived from a pilot study. It is natural to consider a pilot-main experiment setup. Let $X_p \sim Bin(m, \theta)$ be the sample from the pilot study. We assess

$$P_{X_p|\theta=\theta_0+\delta}(\beta_1(\theta) \geq \beta_0) \tag{18}$$

and

$$P_{X_p|\theta=\theta_0+\delta}(\beta_2(\theta) \geq \beta_0) \tag{19}$$

for one-sided and two-sided testing problems, respectively, where δ is effect size and n^* represents a sample size from a determination method. Note that $\beta_1(\theta)$ in (18) relates with the estimated sample size n^* in two aspects: First, n^* is the smallest integer such that $\beta_1(\theta) \geq \beta_0$. Secondly, $\beta_1(\theta)$ is in turn a function of n^* inside the probability in (18). The comments apply similarly to $\beta_2(\theta)$ in (19). For reasonable θ, if (18) and (19) are large enough (close to 1), then that determination method is acceptable. Otherwise, it suggests the calculated sample size might be unreliable. This approach is closely related to β content/$(1-\alpha)$ confidence type evaluation, see, for example, Tsao and Tseng (2006a).

Here we consider $E_U\beta_i(\theta)$, $E_R\beta_i(\theta)$ and $\sup_{\pi\in\Gamma(1/2)} E_{\pi(\theta|\bar{x})}\beta_i(\theta)$, as alternatives to the observed power as estimates of $\beta_i(\theta)$. Let $\widehat{n}_{PI}$, $\widehat{n}_U$, $\widehat{n}_R$ and $\widehat{n}_{sup}$ denote the sample size estimates are calculated by using power estimates $\beta_i(\hat{\theta})$, $E_U\beta_i(\theta)$, $E_R\beta_i(\theta)$ and $\sup_{\pi\in\Gamma(1/2)} E_{\pi(\theta|\bar{x})}\beta_i(\theta)$. Simulations follow the steps.

Step 1. Specify m, θ_0, δ, α and β_0 and generate sample $X_p \sim Bin(m,\theta)$ from the pilot study. Then calculate $\beta_i(\hat{\theta})$, $E_U\beta_i(\theta)$, $E_R\beta_i(\theta)$ and$\sup_{\pi\in\Gamma(1/2)} E_{\pi(\theta|\bar{x})}\beta_i(\theta)$.

Step 2. Calculate the sample sizes $n^* = \widehat{n}_{PI}$, $\widehat{n}_{sup}$, $\widehat{n}_R$ and $\widehat{n}_U$ that guarantee (17) with different estimation methods of $\beta_i(\theta)$.

Step 3. Under true parameter θ, use sample size estimates $n^* = \widehat{n}_{PI}$, $\widehat{n}_{sup}$, $\widehat{n}_R$ and $\widehat{n}_U$ to calculate $\beta_i(\theta)$, $i = 1, 2$.

Step 4. Repeat above steps N times. Approximate (18) and (19) by $\frac{1}{N}\sum_{i=1}^{N} 1\left[\beta_1(\theta) \geq \beta_0\right]$ and $\frac{1}{N}\sum_{i=1}^{N} 1\left[\beta_2(\theta) \geq \beta_0\right]$ where $N = 1000$.

Table 4 and Table 5 give the simulated results for one-sided and two-sided cases respectively with $\theta_0 = 1/2$, $\delta = 1/4$ $(3/8)$, $\alpha = 0.05$, $\beta_0 = 0.8$ and $m = 30, 60, 90, 120$. When δ equals $1/4$, we find that (18), (19) are much smaller then 1. That means all of the estimates ($\widehat{n}_{PI}$, $\widehat{n}_U$, $\widehat{n}_R$ and $\widehat{n}_{sup}$) yield the power that falls below the desired level β_0. However, when δ equals $3/8$, we find that (18) and (19) are close to or equal 1 for estimates $\widehat{n}_U$ and $\widehat{n}_R$ while (18) and (19) are much smaller then 1 for estimates $\widehat{n}_{PI}$ and $\widehat{n}_{sup}$. Overall, it implies that the practice of substituting an estimate $\hat{\theta}$ from pilot study in formula fails to meet the given desired level β_0, yielding a power that falls below the desired level β_0 for all cases we calculated. Therefore, the estimates $\widehat{n}_U$ and $\widehat{n}_R$ are more reasonable choices over the estimates $\widehat{n}_{PI}$ and $\widehat{n}_{sup}$.

Table 4. One-sided testing problem with $\theta_0 = 1/2$, $\delta = 1/4$ $(3/8)$, $\beta_0 = 0.8$ and $\alpha = 0.05$

m	30	60	90	120
$n^* = \widehat{n}_{PI}$, $P_{X_p\|\theta}(\beta_1(\theta)\geq\beta_0)$	0.652 (0.550)	0.664 (0.715)	0.681 (0.735)	0.696 (0.805)
$n^* = \widehat{n}_{sup}$, $P_{X_p\|\theta}(\beta_1(\theta)\geq\beta_0)$	0.652 (0.773)	0.767 (0.833)	0.835 (0.825)	0.828 (0.885)
$n^* = \widehat{n}_R$, $P_{X_p\|\theta}(\beta_1(\theta)\geq\beta_0)$	0.797 (0.904)	0.859 (0.997)	0.891 (1.000)	0.918 (1.000)
$n^* = \widehat{n}_U$, $P_{X_p\|\theta}(\beta_1(\theta)\geq\beta_0)$	0.797 (1.000)	0.859 (1.000)	0.932 (1.000)	0.947 (1.000)

Table 5. Two-sided testing problem with $\theta_0 = 1/2$, $\delta = 1/4$ $(3/8)$, $\beta_0 = 0.8$ and $\alpha = 0.05$

m	30	60	90	120
$n^* = \widehat{n}_{PI}$, $P_{X_p\|\theta}(\beta_2(\theta)\geq\beta_0)$	0.486 (0.718)	0.550 (0.760)	0.589 (0.826)	0.618 (0.854)
$n^* = \widehat{n}_{sup}$, $P_{X_p\|\theta}(\beta_2(\theta)\geq\beta_0)$	0.652 (0.718)	0.769 (0.760)	0.765 (0.900)	0.828 (0.922)
$n^* = \widehat{n}_R$, $P_{X_p\|\theta}(\beta_2(\theta)\geq\beta_0)$	0.652 (0.896)	0.769 (0.982)	0.835 (1.000)	0.878 (1.000)
$n^* = \widehat{n}_U$, $P_{X_p\|\theta}(\beta_2(\theta)\geq\beta_0)$	0.797 (1.000)	0.853 (1.000)	0.891 (1.000)	0.918 (1.000)

5. Conclusion and Discussion

Under a robust Bayesian framework, we show that the observed powers are too large when the data is significant and priors belonging to $\Gamma(1/2)$. This conclusion holds both for one-sided and two-sided hypothesis testing problems. Furthermore, because the observed powers are too large to estimate power in many situations, this undesirable property leads to unreliable sample size calculation. An empirical study about the problem of sample size calculation is carried out. The usual sample size determination method based on the observed power always fails to meet the given desired level of power for all cases we calculated. More reasonable sample sizes can be calculated when the power is better estimated. For example, the Bayes estimates with respect to uniform or reference priors are better alternatives to the observed power on this regard.

Hoenig and Heisey (2001) reviews the practical implementation and interpretation of the observed power. It also points out the observed power is a decreasing function of the p-value of the test. Therefore, the observed power essentially renders no extra information. We subscribe to their points of view. However, as mentioned in Tsao and Tseng (2006b), the observed power, good or bad, is an estimate of the performance of a given α level test. In perspective, our results suggest the observed power might overestimate the power when the data is significant.

Note that the deriving the ordering of Bayes estimates in c for fixed p_0, say 0.5, is not straightforward. Technically, our approach requires detailed analysis of functional form of $\beta_i(\theta)$ (after suitable transformation) and intelligent utilization of the monotone likelihood property. Besides, in the proof of Proposition 2, the relation between binomial and beta distribution is vital and the condition $c_1 + c_2 = n$ is needed. An alternative approach has been recently developed in Tsao and Tseng (2006a). The problem of power estimation of binomial hypothesis testing problems using normal approximate test is investigated in Tsai (2004).

Our results show the extremity of observed power for both one-sided and two-sided hypothesis testing problems. They are similar to Tsao (2006) using a smooth null estimation approach, Tsao and Tseng (2006b) and Gillett (1996) in the context of retrospective power surveys. These results differ from the parallel results for p-values using an accuracy estimation approach in which different conclusions are reached respectively for one-sided and two-sided hypothesis testing problems. Roughly speaking, the p-value and the infimum of Bayes estimate coincide for one-sided hypothesis testing problems but differ for two-sided hypothesis testing problems. See, for example, Berger and Delampady (1987), Berger and Sellke (1987) and Casella and Berger (1987).

Acknowledgment

The research is supported in part by NSC 94-2118-M-259-001 and NSC 96-

2118-M-214-001.

References

1. J. O. Berger and M. Delampady, Testing precise hypothesis (with discussion), *Statistical Science*, **2** (1987), 317-352.
2. J. O. Berger and T. Sellke, Testing a point null hypothesis: the irreconcilability of p-values and evidence, *Journal of American Statistical Association*, **82** (1987), 112-122.
3. G. Casella and R. L. Berger, Reconciling Bayesian and Frequentist evidence in the one-sided testing problem, *Journal of American Statistical Association*, **82** (1987), 106-111.
4. G. Casella and R. L. Berger, *Statistical Inference*, Wadsworth, California, 1990.
5. W. J. Conover, *Practical Nonparametric Statistics*, 2nd Edition, John Wiley and Sons, New York, 1980.
6. J. D. Gibbons, *Nonparametric Methods for Quantitative Analysis*, 2nd Edition, American Sciences Press, Ohio, 1985.
7. R. Gillett, Retrospective power surveys, *The Statistician*, **45** (1996), 231-236.
8. J. M. Hoenig and D. M. Heisey, The abuse of power: the pervasive fallacy of power calculation for data analysis, *The American Statistician*, **55** (2001), 19-24.
9. M. Hollander and D. A. Wolfe, *Nonparametric Statistical Methods*, John Wiley and Sons, New York, 1973.
10. J. T. Hwang, G. Casella, C. Robert, M. Wells and R. Farrel, Estimation of accuracy in testing, *Annals of Statistics*, **20** (1991), 490-509.
11. C. Jennison, Discussion of ”Inferences from multinomial data: Learning from a bag of marbles”, *Journal of Royal Statistical Society B*, **58** (1996), 36.
12. E. L. Lehmann, *Testing Statistical Hypotheses*, 2nd Edition, John Wiley and Sons, New York, 1986.
13. R. V. Lenth, Some practical guidelines for effective sample size determination, *The American Statistician*, **55** (2001), 187-193.
14. I. Levi, Discussion of ”Inferences from multinomial data: Learning from a bag of marbles”, *Journal of Royal Statistical Society B*, **58** (1996), 47.
15. H. S. Oh and A. DasGupta, Comparison of the P-value and posterior probability, *Journal of Statistical Planning and Inference*, **76** (1999), 93-107.
16. A. O’Hagan, Discussion of ”Inferences from multinomial data: Learning from a bag of marbles”, *Journal of Royal Statistical Society B*, **58** (1996), 34.
17. T. Sellke, M. J. Bayarri and J. O. Berger, Calibration of p-values for testing point null hypotheses, *The American Statistician*, **55** (2001), 62-71.

18. C.-C. Tsai, *Post-data estimation of approximate power in binomial mean testing problem*, Technical Report, National Dong Hwa University, 2004.

19. C. A. Tsao, Assessing post-data weight of evidence, *Journal of Statistical Planning and Inference*, **136** (2006), 4012-4025.

20. C. A. Tsao and Y.-L. Tseng, A statistical treatment of the problem division, *Metrika*, **59** (2004), 289-303.

21. C. A. Tsao and Y.-L. Tseng, Confidence estimation for tolerance intervals, *Annals of the Institute of Statistical Mathematics*, **58** (2006a), 441-456.

22. C. A. Tsao and Y.-L. Tseng, *Power estimation for testing normal means*, Technical Report, National Dong Hwa University, 2006b.

23. J.-J. Tsay and C. A. Tsao, Statistical gambler's ruin problem, *Communications in Statistics-Theory and Methods*, **32** (2003), 1337-1359.

24. P. Walley, Inferences from multinomial data: Learning from a bag of marbles (with discussion), *Journal of Royal Statistical Society B*, **58** (1996), 3-57.

Chih-Chien Tsai
Department of Applied Math,
I-Shou University,
Kaohsiung 84001,
Taiwan
E-mail: tjj@isu.edu.tw

C. Andy Tsao
Department of Applied Math,
National Dong Hwa University,
Hualien 97401,
Taiwan
E-mail: chtsao@mail.ndhu.edu.tw