Bayesian Sample Size Optimization Method for Integrated Test Design of Missile Hit Accuracy

Guangling Dong[1], Chi He[2], Zhenguo Dai[3], Yanchang Huang[3] and Xiaochu Hang[3]

[1]School of Astronautics, Harbin Institute of Technology, 92 Xidazhijie Street, Harbin, China

[2]School of Mechatronic Engineering, CUST, 7089 Weixing Road, Changchun, China

[3]Department of Test Technology, Baicheng Ordnance Test Center of China, Mailbox 108, Baicheng, China

Keywords: Missile Hit Accuracy, Integrated Test Design, Bayesian Sample Size Determination, Power Prior, Design Prior.

Abstract: Sample size determination (SSD) for integrated test of missile hit accuracy is addressed in this paper. Bayesian approach to SSD gives test designer the possibility of taking into account of prior information and uncertainty on unknown parameters of interest. This fact offers the advantage of removing or mitigating typical drawbacks of classical methods, which might lead to serious miscalculation of the sample size. However, standard power prior based Bayesian SSD method cannot cope with integrated SSD for both simulation test and field test, as large numbers of simulation samples would cause contradiction between design prior and average posterior variance criterion (APVC). In allusion to this problem, we propose a test design effect equivalent method for equivalent sample size (ESS) calculation, which combined simulation credibility, sample size, and power prior exponent to get a rational design prior for subsequent field test. Average posterior variance (APV) of interested parameters is deduced by simulation credibility, sample sizes of two kinds of test, and prior distribution parameters. Thus, we get optimal design equations of integrated test scheme under both test cost constraints and required posterior precision constraint, whose effectiveness are illustrated with two examples.

1 INTRODUCTION

Hit accuracy is a key technical performance index in missile weapon system, which is traditionally testified through live field firing test. As its extremely high expenditure in field test, we can only get very limited test samples as a basis for type approval to reach a low precision assessment result. Thus, in order to solve this problem, integrated test and evaluation with combined use of simulation test and field test was put forward in 1990s (Kraft 1995; Kushman and Briski 1992), and by now, it has become the development direction of test and evaluation (Claxton et al., 2012; Schwartz, 2010; Waters, 2004). In the design stage of integrated test, sample size allocation (SSA) for simulation test and field test is a vital problem. Standard frequentist sample size formulae generally determine a specific sample size through precision requirement of parameter estimation (Adcock, 1997) or statistical power analysis of Hypnosis test (Murphy et al., 2009), as used in design of integrated test, SSA ratio become a problem, and also the equal treatment of simulation test sample and field test sample is widely questioned.

Generally, simulation test samples are strongly correlated with field test samples, but they don't necessarily take on the same distribution parameters. Verification validation and accreditation (VV&A) of simulation system (Balci 1997, 2013; Rebba et al., 2006) are made to get some indexes (such as simulation credibility) for description of this difference. Bayesian sample size determination (SSD) (Clarke and Yuan, 2006; De Santis, 2007; Joseph and Belisle, 1997; Nassar et al., 2011) makes use of prior information to get the minimal sample size through pre-posterior estimated performance criteria of parameter of interest. While it is used in SSD of integrated test, determining a proper weight for prior information (De Santis, 2007) is a problem. Also, the obtained power prior (Ibrahim and Chen, 2000) for design from large size of prior samples is often narrow enough to estimate the unknown parameter of interest, which seems no need for field

Dong G., He C., Dai Z., Huang Y. and Hang X..
Bayesian Sample Size Optimization Method for Integrated Test Design of Missile Hit Accuracy.
DOI: 10.5220/0005510902440253
In *Proceedings of the 5th International Conference on Simulation and Modeling Methodologies, Technologies and Applications* (SIMULTECH-2015), pages 244-253
ISBN: 978-989-758-120-5

test samples. But, this totally inference of interested parameter by simulation test samples is under suspicion, and takes on great risk due to sometimes low credible simulation test samples.

A proper design prior is the basis for Bayesian SSD of integrated test, which is the key content of our research. The rest of this paper is organized as follows: Section 2 introduce the common used SSD methods of both frequentist and Bayesian, as well as the definitions of analysis prior and design prior used in Bayesian SSD. Section 3 analyses the problem of determining exponential factor in getting design prior with power prior, and propose the design effect of test to calculate the equivalent sample size (ESS) of simulation test, which provide the key content of our proposed method for getting a rational design prior. Section 4 deals with optimal SSA for simulation test and field test, which takes average posterior variance of Bayesian estimation of hit accuracy as output, and gets two optimization equations under cost constraints and required posterior precision constraint. Section 5 illustrates the proposed methods with two examples, and proves their effectiveness in practical application. In section 6, the features of our proposed Bayesian SSD for design of integrated test are summarized.

2 PRELIMINARIES

2.1 Classical Sample Size Determination

Classical sample size determination method is generally related to interval estimation and hypothesis testing. Estimation of mean value μ for normal distribution with known variance is the simplest condition in SSD, many literatures (Adcock, 1997; Desu and Raghavarao, 1990) have made in-depth study on this problem. Suppose we have a given estimation error d and confidence level 1-α, estimation of μ can be realized by sample mean, thus the SSD formula is

$$\Pr\left[\left|\bar{x}-\mu\right| \le d \,|\mu\right] \ge 1-\alpha \tag{1}$$

As $\sqrt{n}(\bar{x}-\mu)/\sigma \sim N(0,1)$, it follows that inequality (1) is satisfied if sample size n satisfies

$$n \ge \sigma^2 Z_{\alpha/2}^2 / d^2 \tag{2}$$

Statistical power analysis of hypothesis testing extends the rule at inequality(2). Suppose we have a hypothesis of the form

$$H_0 : \mu = \mu_0; \quad H_1 : \mu = \mu_1 \ne \mu_0$$

Significance level α of above hypothesis testing means we reject H_0 while $\left|\bar{x}-\mu_0\right| > \sigma Z_{\alpha/2}/\sqrt{n}$. If the desired power of the alternative when $\left|\mu_1-\mu_0\right| = d$ is 1-β, where d is specified, then the required sample size n is the solution of the following equation

$$\Pr\left[\left|\bar{x}-\mu_0\right| > \sigma Z_{\alpha/2}/\sqrt{n}\,|\mu_1\right] = 1-\beta$$

Define $Z = \sqrt{n}(\bar{x}-\mu_0)/\sigma$, this could be expressed as

$$\Pr\left[\left|Z+\sqrt{n}\cdot d/\sigma\right| > Z_{\alpha/2}\right] = 1-\beta \tag{3}$$

Equation (3) has a unique solution. Desu and Raghavarao (1990) have made detail deduction and gave out its approximate solution

$$n^* = \sigma^2 (Z_{\alpha/2} + Z_\beta)^2 / d^2 \tag{4}$$

Equation (4) is valid as long as α is not too large, and it reduces to inequality (2) when the desired power is 0.5. While the variance σ^2 is unknown, similar formulae can be gotten, for further detail, see (Desu and Raghavarao, 1990) and reference therein.

Samples from simulation test and field test play different roles in performance assessment, so their sample size ratio and weights cannot be determined arbitrarily. In this way, classical SSD method cannot cope with situations in integrated test. As a result, Bayesian SSD method become the inevitable choice in design of integrated missile hit accuracy test.

2.2 Bayesian Sample Size Determination

The key difference in SSD between Bayesian method and classical method is the use of prior information. Suppose that we are interested in choosing the size n of a random sample $\boldsymbol{X}_n$=(X_1, …, X_n), whose joint density function $f_n(\bullet|\theta)$ depends on the unknown parameter vector θ of interest. With Bayesian approach, given sample data $\boldsymbol{x}_n = (x_1, \cdots, x_n)$, the likelihood function $L(\theta; x_n) \propto f_n(x_n \mid \theta)$, and the prior distribution $\pi(\bullet)$ for parameter of interest, we get inference of θ based on elaborations of the posterior distribution:

$$\pi(\theta|\boldsymbol{x}_n) = \frac{f_n(\boldsymbol{x}_n|\theta)\pi(\theta)}{\int_\Theta f_n(\boldsymbol{x}_n|\theta)\pi(\theta)\mathrm{d}\theta}$$

Let $T(\boldsymbol{x}_n)$ denote a generic function of the posterior distribution of θ, which can be controlled by designed test sample size. For instance, $T(\boldsymbol{x}_n)$ could be either the posterior variance, or the width of the highest posterior density (HPD) set or the posterior probability of a certain hypothesis. Before we get actual sample data $\boldsymbol{x}_n$, $T(\boldsymbol{X}_n)$ is a random variable. The idea of Bayesian SSD is to select n so that the observed value $T(\boldsymbol{x}_n)$ is likely to provide accurate information on θ. Pre-computations for SSD are made with the following marginal density function of sample data

$$m_n(x_n;\pi) = \int_\Theta f_n(x_n|\theta)\pi(\theta)\mathrm{d}\theta$$

Above equation shows a mixture of the sampling distribution and the prior distribution of θ. Most Bayesian SSD criteria select the minimal n so that, for chosen values $\varepsilon>0$ and $\varepsilon' \in (0,1)$, one of the two following statements is satisfied:

$$\mathrm{E}\left[T(\boldsymbol{X}_n)\right] \le \varepsilon \tag{5}$$

$$\text{or } \Pr\left[T(\boldsymbol{X}_n) \in A\right] \le \varepsilon' \tag{6}$$

Where, E[•] denotes the expected value computed with respect to marginal density function m_n. Pr[•] express the probability measure corresponding to m_n. A is a subset of the value space that the random variable $T(\boldsymbol{X}_n)$ can assume. Thus, SSD and the subsequent parameter inference is a two-step process: first select n^* by a pre-posterior calculation; and then use $\pi(\theta|\boldsymbol{x}_{n^*})$ to obtain $T(\boldsymbol{x}_{n^*})$.

There are three common used Bayesian SSD criteria for estimation, where criteria (b) and (c) are interval type methods based on the idea of controlling the random length of credible sets for probability distribution, refer to (De Santis, 2007) for further detail.

(a) Average posterior variance criterion (APVC): for a given $\varepsilon>0$, APVC select the minimal sample size n such that

$$\mathrm{E}\left[\mathrm{var}(\theta|\boldsymbol{X}_n)\right] \le \varepsilon \tag{7}$$

Where, $\mathrm{var}(\theta|\boldsymbol{X}_n)$ is the posterior variance of interested unknown parameter θ, and $T(\boldsymbol{x}_n) = \mathrm{var}(\theta|\boldsymbol{x}_n)$, This criterion controls the dispersion of the posterior distribution. Of course, different dispersion measures can be used to derive alternative criteria.

(b) Average length criterion (ALC): for a given $l>0$, ALC looks for the minimal n to meet the inequality

$$E\left[L_\alpha(\boldsymbol{X}_n)\right] \le l \tag{8}$$

Where, for a fixed $\alpha \in (0,1)$, .$L_\alpha(\boldsymbol{X}_n)$ is the random length for the (1-α) level posterior set of θ. In this case, $T(\boldsymbol{x}_n) = L_\alpha(\boldsymbol{x}_n)$. This criterion was proposed by Joseph (1995) (Joseph et al., 1995), aiming at controlling the average length of the HPD set, but not its variability.

(c) Length probability criterion (LPC): for given $l>0$ and $\varepsilon' \in (0,1)$, LPC choose the smallest n to meet the inequality

$$\Pr\left[L_\alpha(\boldsymbol{X}_n) \ge l\right] \le \varepsilon' \tag{9}$$

Just as for the ALC, $T(\boldsymbol{x}_n)= L_\alpha(\boldsymbol{x}_n)$ and the LPC can be written in the general form (6) with $A = (l\ ,\ U)$, where U denotes the upper bound for the length of the credible interval. Joseph and Belisle (1997) derived the LPC as a special case of the worst outcome criterion (WOC). Refer to (Di Bacco et al., 2003; Joseph et al., 1995) for further details.

SSD criteria for model selection and hypothesis testing are also available in many literatures such as (De Santis, 2004; Reyes and Ghosh, 2013; Wang and Gelfand, 2002; Weiss, 1997). As they are not suitable for sample size determination of integrated test design, we will not make detailed instruction here. For further information, refer to literatures above.

2.3 Two Priors in Bayesian SSD

Two priors approach to SSD, namely analysis prior and design prior, has been already proposed, for instance, by Spiegelhalter and Freedman (1986) and by Joseph et al. (1997). By now, this approach has gotten wide spread usage in many fields. Generally, analysis prior formalizes the pre-test knowledge that we want to take into account, together with test samples, in the final analysis. While design prior describes a scenario which is not necessarily coincident with that of analysis prior, under which we want to choose the sample size. Thus, design prior serves to obtain a marginal distribution m_n that incorporates uncertainty on a guessed distribution for θ. Refer to (De Santis, 2007) for a thorough discussion on this question.

In general, analysis prior can be improper (such as some non-informative priors) as long as the resulting posterior is proper. While the design prior is used for calculating marginal distribution m_n, if it is improper, the resulting marginal distribution may not even exist and the integral that defines m_n being divergent. So, proper design prior should be chosen in Bayesian SSD of integrated test. In order to use historical data for Bayesian SSD, De Santis (2007) went into this problem and gave out a method for getting design prior of location parameter. According to his analysis results, non-informative prior can be used as analysis prior for simulation test samples, and the obtained posterior distribution is proper, which constitutes a suitable prior for subsequent SSD of field test.

However, as simulation test samples are not totally credible, the obtained posterior cannot be used directly as design prior of field test. Power prior (De Santis, 2007; Fryback et al., 2001; Greenhouse and Waserman, 1995) has been taken as a method for incorporating historical information in a design prior and also for deciding the weight that such information has in the SSD process. Therefore, rational design prior for SSD of field test can be obtained based on simulation test samples and a proper use of power prior, which will be discussed in detail in the next section.

3 EQUIVALENT SAMPLE SIZE FOR DESIGN PRIOR

In this section we penetrate into the design prior for Bayesian SSD of field test. The problem and contradiction of using standard power prior from simulation test samples as the design prior of Bayesian SSD in integrated test will be analysed, which lead to our proposed design effect of test for calculating equivalent sample size (ESS) of simulation test. Thus, we get the ESS based design prior for Bayesian SSD.

3.1 Power Prior for Design

Power prior is often used in obtaining design prior of Bayesian SSD, which realizes the weighting fusion of prior information. Ibrahim and Chen (Ibrahim et al., 2003; Ibrahim and Chen, 2000) proposed the application of power prior in information fusion and regression modelling, and analysed its superiority therein. Suppose z_{n_0} is prior data with size n_0 used for providing design prior of field test; $L(\theta; z_{n_0})$ is likelihood function of θ in z_{n_0}, here homogeneity of prior samples and subsequent field test samples is an implicit assumption; $\pi_0(\theta)$ is non-informative prior. Consider posterior distribution $\pi^{\mathrm{P}}(\cdot \mid z_{n_0}, a_0)$ is obtained through multiplication of prior distribution $\pi_0(\bullet)$ and likelihood function $L(\theta; z_{n_0})$, suitably scaled by an exponential factor a_0:

$$\pi^{\mathrm{P}}(\theta \mid z_{n_0}, a_0) \propto \pi_0(\theta) L(\theta; z_{n_0})^{a_0} \qquad a_0 \in (0,1) \tag{10}$$

If π_0 is proper, $\pi^{\mathrm{P}}(\theta \mid z_{n_0}, a_0)$ is also proper; otherwise, z_{n_0} should make $\pi^{\mathrm{P}}(\theta \mid z_{n_0}, a_0)$ a proper distribution function. Exponent a_0 measures the importance of prior data in $\pi^{\mathrm{P}}(\theta \mid z_{n_0}, a_0)$. As $a_0 \to 1$, we get standard posterior distribution of θ with z_{n_0}; while as $a_0 \to 0$, $\pi^{\mathrm{P}}(\theta \mid z_{n_0}, a_0)$ tends to initial non-informative prior π_0. Thus, exponent a_0 determines the weight of z_{n_0} in posterior distribution. This shows we can choose different a_0 for alternative weights on prior information. Spiegelhalter et al. (2004) have gone through this question. Following the definition of power prior, we can get it with a mixture of expression (10) and distribution of a_0. The effect of mixing is to obtain a prior for θ that has a much heavier tails than those which are obtained with fixed a_0. Obviously, exponent a_0 plays an important role in power prior calculation.

As to Bayesian SSD of integrated test, field test samples $\boldsymbol{X}_n$ is still to be observed, we assume π_0 is non-informative prior, and $\pi^{\mathrm{P}}(\theta \mid z_{n_0}, a_0)$ as the design prior obtained from z_{n_0}, a proper marginal distribution of the data $\boldsymbol{X}_n$ can be gotten from

$$m_n(\boldsymbol{x}_n \mid z_{n_0}, a_0) = \int_{\Theta} f_n(\boldsymbol{x}_n \mid \theta) \pi^{\mathrm{P}}(\theta \mid z_{n_0}, a_0) \mathrm{d}\theta \tag{11}$$

Marginal distribution m_n and the resulting sample size depend on the exponent a_0 of power prior. Owing to lack of operational interpretation and without a way to assess the value of a_0, the use of power prior for posterior inference has been criticized, for instance, (Spiegelhalter et al., 2004). De Santis (2007) gave out two interpretations for using power prior in the parameter inference of independent and identically distributed (IID) data:

- When a maximum likelihood estimator for θ exists and is unique, π^{P} is equivalent to a posterior distribution that is obtained by using a sample of size $r = a_0 n_0$, which provides the same maximum likelihood estimator for θ as the entire sample.
- When the model $f_n(\cdot \mid \theta)$ belongs to the exponential family, the prior π^{P} for the natural parameter coincides with the standard

posterior distribution that is obtained by using a sample of size $r = a_0 n_0$ whose arithmetic mean is equal to the historical data mean.

Hence, at least in some standard problems, a power prior can be interpreted as a posterior distribution that is associated with a sample whose informative content on θ is qualitatively the same as that of the historical data set, but quantitatively equivalent to that of a sample of size r. In the design of integrated test for missile hit accuracy, an intuitive idea is to choose directly the simulation system credibility C_0 as exponent a_0 for power prior calculation. However, considering the exponential behaviour of likelihood function for normal random variable, the obtained power prior with extremely large size of prior samples via expression (10) could be a very sharp distribution, which means the design prior of field test already meets with the Bayesian SSD criteria as APVC, ALC and LPC. So, we don't need to carry out any field test for inference of interested parameter. Obviously, this is impractical, as with low simulation credibility, performance inference can takes on high risk even with large size of simulation test samples.

The above contradictory is caused by ignoring the influence of prior sample size n_0 on determining the exponent a_0 of power prior. Taking $C_0 = a_0$, the prior sample size n_0 can still have great influence on the equivalent standard sample size according to $r = a_0 n_0$, which makes the design prior impractical. For this reason, we propose the design effect of test, which can be used to get equivalent sample size n_r of simulation test samples based on the notion of design effect equivalence, and hence get a rational design prior of Bayesian SSD.

3.2 Design Effect of Test

Design effect of test is proposed with a comprehensive consideration on test credibility and posterior estimation performance. Taking average posterior variance of Bayesian SSD as the posterior estimation performance, it is defined as

$$D_E = C \cdot \exp(-L_{\mathrm{APV}}) \qquad (12)$$

Where, C indicates test credibility, L_{APV} denotes the average posterior variance of interested parameter obtained with non-informative prior and test samples. If simulation test has a credibility C_0, and with a sample size n_0, it takes on equal design effect as field test (whose credibility is 1) with a sample size n_r, then we choose the exponential factor $a_0 = n_r / n_0$ for power prior calculation. Thus, according to De Santis's interpretation, we get the equivalent sample size (ESS) n_r for simulation test samples of size n_0 with equal informative content on θ.

From equation (12), while $C_0 = 1$, prior samples is equal to actual field test samples, and $a_0 = 1$; if simulation test credibility C_0 is low, even though sample size n_0 is very large, and $L_{\mathrm{APV}} \rightarrow 0$, as its design effect $D_E \rightarrow C_0$, the ESS n_r would only be a limited value. Taking non-informative prior as analysis prior, and $s_0 = s_1 = 1.2$, exponent a_0 of power prior under different simulation credibility can be worked out based on equivalence of test design effect. Figure 1 shows its relationship with simulation test sample size n_0. In this way, exponent a_0 of power prior decreases with the increment of simulation test sample size n_0, which avoids the above-mentioned assessment risk with only large numbers of simulation test samples.

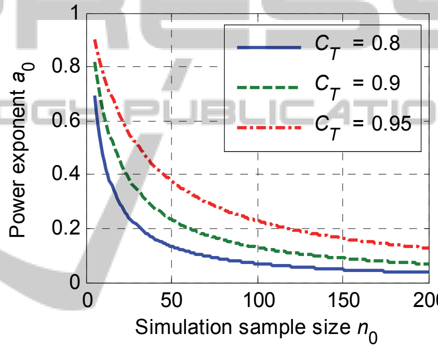

Figure 1: Power exponent a_0 vs. simulation test sample size n_0.

3.3 Equivalent Sample Size based Design Prior Elicitation

In the integrated test of missile hit accuracy, mean value μ of impact point is taken as parameter of interest, simulation test samples $\boldsymbol{z}_{n_0} = [z_1, \cdots, z_{n_0}]$ and field test samples $\boldsymbol{y}_{n_r} = [y_1, \cdots, y_{n_r}]$ are independent identically distributed (IID) normal random variables, with both mean value μ and precision λ (reciprocal of variance σ^2) unknown. Then, we can get the posterior distribution of μ from simulation test samples with non-informative analysis prior , which is expressed as

$$\pi(\mu \mid \boldsymbol{z}_{n_0}) = \mathrm{St}(\mu \mid \bar{z}_{n_0}, (n_0 - 1)s_0^{-2}, n_0 - 1), \quad n_0 > 1$$

Where, St express student t distribution; $\bar{z}_{n_0} = \mathrm{mean}(z_i)$ is estimated location parameter; $(n_0 - 1)s_0^{-2}$ is precision, with $s_0^2 = \mathrm{mean}\left((z_i - \bar{z}_{n_0})^2\right)$;

n_0-1 is the degree of freedom for student t distribution. So, the average posterior variance of μ can be derived as

$$L_{\mathrm{APV}}(\mu \mid z_{n_0}) = \mathrm{E}\left(\mathrm{var}(\mu \mid z_{n_0})\right) = \frac{s_0^2}{n_0 - 1}$$

With non-informative prior, suppose field test samples of size n_r take on equal design effect as simulation test samples of size n_0, the computational formula of equivalent sample size is

$$C_0 \exp\left(\frac{-s_0^2}{n_0 - 1}\right) = \exp\left(\frac{-s_r^2}{n_r - 1}\right)$$

And, it follows that

$$n_r = \frac{(n_0 - 1)s_r^2 + s_0^2 - (n_0 - 1)\ln(C_0)}{s_0^2 - (n_0 - 1)\ln(C_0)} \qquad (13)$$

Where, s_r is standard deviation of field test samples. In test design stage, both s_0 and s_r are unknown, we can take $s_r = s_0$ provided a homogeneity of variance assumption. Figure 2 shows the relationship of equivalent sample size n_r and simulation test sample size n_0 under different simulation test credibility, with $s_r = s_0 = 1.2$. It can be seen that the ESS tend to finite value as the increase of simulation sample size, and the higher the simulation credibility, the larger the equivalent sample size, which coincides with practical situation.

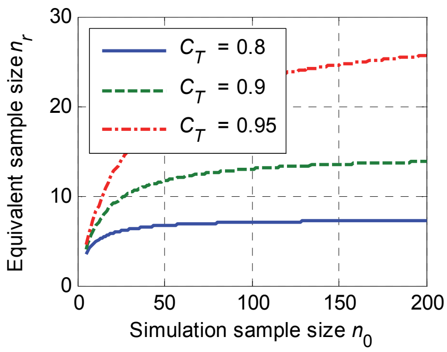

Figure 2: Equivalent sample size of simulation test samples.

Following De Santis's conclusions, as long as we get the equivalent sample size of simulation test samples z_{n_0}, which is IID and normally distributed, the design prior for Bayesian SSD of subsequent field test can be expressed as

$$\begin{aligned} p(\mu, \lambda \mid z_{n_0}) &= \mathrm{Ng}(\mu, \lambda \mid \mu_0, n_r, \alpha_0, \beta_0) \\ &= \mathrm{N}(\mu \mid \mu_0, n_r \lambda) \cdot \mathrm{Ga}(\lambda \mid \alpha_0, \beta_0) \end{aligned} \qquad (14)$$

Where,

$$\mu_0 = \bar{z}_{n_0}$$
$$\alpha_0 = (n_r - 1)/2$$
$$\beta_0 = n_r s_0^2 / 2$$

4 SSD FOR INTEGRATED TEST SCHEME

The purpose of SSD for integrated test of missile hit accuracy is to get a comprehensive test scheme satisfying the requirement of assessment precision with minimal test cost. In this section, we take average posterior variance of missile hit accuracy as assessment precision, and give out the optimal SSD methods for integrated test under both test cost constraint and required assessment precision constraint, according to Bayesian average posterior variance criterion.

4.1 Purpose for Optimal Design of Integrated Test Scheme

Design of integrated test scheme takes on the objective of deciding sample sizes of both simulation test and field test for optimal assessment of missile hit accuracy. So we can avoid resources waste caused by excessive sample size and also assessment risk due to insufficiency of test samples. Therefore, optimal design of integrated test takes on two purposes: first, determining sample size allocation plan for simulation test and field test to get a minimal Bayesian average posterior variance of missile hit accuracy under limited test cost constraint; second, making unified sample size allocation plan for simulation test and field test, to meet the requirement of posterior estimation precision with minimal test cost.

In allusion to integrated test of missile hit accuracy, we want to get a scientific estimation for mean value μ of impact point, whose precision can be expressed with posterior variance. So, optimal design of integrated test scheme will be studied from two aspects as devoted test cost and estimated average posterior variance, with an objective of getting optimal sample size allocation plan for integrated test scheme under two kinds of constraints.

4.2 Bayesian SSD based on APVC

The problem can be formalized as follows. Take equation (14) as design prior of field test, while we get field test samples $\boldsymbol{x}_n$, mark $\bar{x} = \text{mean}(x_i)$, $s^2 = \text{mean}\left((x_i - \bar{x})^2\right)$, then, the posterior distribution of μ is

$$p(\mu \mid \boldsymbol{x}_n) = \text{St}\left(\mu \middle| \mu_n, (n+n_r)(\alpha_0 + n/2)\beta_n^{-1}, 2\alpha_0 + n\right)$$
$$\mu_n = (n_r + n)^{-1}(n_r\mu_0 + n\bar{x})$$
$$\beta_n = \beta_0 + \frac{ns^2}{2} + \frac{1}{2}(n_r + n)^{-1} n_r n(\mu_0 - \bar{x})^2$$

It follows that the posterior variance of μ is

$$\text{var}\left(\mu \middle| \boldsymbol{x}_n\right) = \frac{\beta_n}{(n+n_r)(\alpha_0 + n/2)} \quad (15)$$

As field test sample $\boldsymbol{x}_n$ is unknown, and can be seen as random variable. So, in order to get the average posterior variance of μ, marginal distribution (11) of field test sample $\boldsymbol{x}_n$ is used in mean value calculation, and it follows that

$$\text{E}\left(\text{var}(\mu \mid \boldsymbol{x}_n)\right) = \frac{\text{E}(\beta_n)}{(n+n_r)(\alpha_0 + n/2)} \quad (16)$$

In this way, average posterior variance of μ is a function of E(ns^2) and $\text{E}\left((\mu_0 - \bar{x})^2\right)$. Where, ns^2 has a Gamma-Gamma distribution, and $\bar{x}$ has a student t distribution, Bernardo (Bernardo and Smith 2000) gave out their probability distribution function as,

$$p(ns^2) = \text{Gg}\left(ns^2 \middle| \alpha_0, 2\beta_0, (n-1)/2\right) \quad (17)$$

$$p(\bar{x}) = \text{St}\left(\bar{x} \middle| \mu_0, n_0 n(n_0 + n)^{-1}\alpha_0\beta_0^{-1}, 2\alpha\right) \quad (18)$$

From equations (13) to (18), we can get the average posterior variance of μ

$$\text{E}[\text{var}(\mu|\boldsymbol{x}_n)] = \frac{n_r s_0^2\left[n_r^2 + (n-4)n_r - n + 1\right]}{(n+n_r-1)(n+n_r)(n_r-1)(n_r-3)} \quad (19)$$

Given the required ε in APVC, we can get the minimal field test sample size n^* according to equation (19), which meets with the requirement of APVC.

4.3 Optimal Design Equations of Integrated Test Scheme

Suppose unit sample costs for simulation test and field test are u_0 and u_1 respectively. Under the constraint of total test cost within T_c, we can get the optimal integrated test scheme with minimal average posterior variance of interested parameter based on APVC of Bayesian SSD, whose design equation is as follows:

$$\min \begin{cases} L_{\text{APV}}(\boldsymbol{N}, C_0, s_0) = \text{E}\left[\text{var}(\mu \mid \boldsymbol{x}_n)\right] \\ \boldsymbol{N} = [n_0, n] \end{cases}$$
$$\text{s.t.} \begin{cases} n_0u_0 + nu_1 \le T_c \\ n_0 > 1; \\ n > 1; \end{cases} \quad (20)$$

Where, $L_{\text{APV}}(\boldsymbol{N}, C_0, s_0)$ represents Bayesian average posterior variance of missile hit accuracy, and works as the objective function of optimization, which is denoted by equation (19). $\boldsymbol{N}$ is the optimizing vector containing sample sizes of simulation test and field test. C_0 is simulation test credibility. s_0 is standard deviation of simulation test samples.

The other condition for optimal design of integrated test scheme is under required precision ε constraint in APVC, seeking the sample size allocation (SSA) plan for minimal total test cost. At this time, the optimal design equation is

$$\min T_c(\boldsymbol{N}) = n_0u_0 + nu_1$$
$$\text{s.t.} \begin{cases} L_{\text{APV}}(\boldsymbol{N}, C_0, s_0) \le \varepsilon; \\ n_0 > 1; \\ n > 1 \end{cases} \quad (21)$$

Equations (20) and (21) provide the optimal design methods for integrated test scheme of missile hit accuracy, whose usage will be demonstrated in detail in Section 5 with two illustrative examples. During their utilization, we get equivalent sample size of simulation test samples for design prior based on credibility C_0 and standard deviation s_0. So, optimal design scheme of integrated test can be obtained in consideration of two kinds of unit sample cost.

5 ILLUSTRATIONS

In the test and evaluation of missile hit accuracy, the key point of integrated test design is to decide the sample size allocation plan for simulation test and field test. So, we can get optimal assessment precision with minimal test consumption. In this section, we demonstrate the optimal design of integrated test scheme under constraints of cost and required assessment precision. In this way, the effectiveness of Bayesian SSD method for design of integrated test is illustrated, where the basic settings for simulation test and field test are shown in Table 1 (not real number, just used for illustration).

Table 1: Basic setting for test schemes.

Test type	Unit sample cost ($)	Pre-estimated standard deviation	Credibility (C_T)
Simulation	50	1.2	0.70~0.95
Field	2000	1.2	1.0

5.1 Optimal Design with Cost Constraint

Suppose we have a total appropriation budget T_C =20, 000 $, and want to get the optimal missile hit accuracy assessment. Optimization equation for SSA of integrated test can be built according to (20), whose solution is straight forward using a nonlinear constrained optimization algorithm. Thus, we get optimal sample size allocation plans under different simulation credibility (SC), where n_0 and n_1 are sample sizes for simulation test and field test respectively, as shown in Table 2. Obviously, the ratio of simulation test samples in optimal integrated test scheme and the corresponding equivalent sample sizes increase with simulation credibility. Meanwhile, average posterior variances decrease with simulation test credibility, which indicates the precision of Bayesian posterior assessment is better under higher simulation test credibility.

Table 2: Optimal integrated test schemes under cost constraint.

SC	SSA		ESS	APV
C_0	n_0	n_1	n_r	L_{APV}
0.70	80	8	4.8410	0.2321
0.80	80	8	6.9659	0.1407
0.90	80	8	12.6516	0.0813
0.95	160	6	24.8609	0.0494

Figure 3 shows the relationship between field test sample size n and average posterior variance under total test cost constraint with a simulation test credibility C_0=0.8, where simulation test sample size is determined by the residual fund of field test. In this way, integrated test with combined use of simulation test and field test takes on better assessment effect than traditional only with field test samples mode.

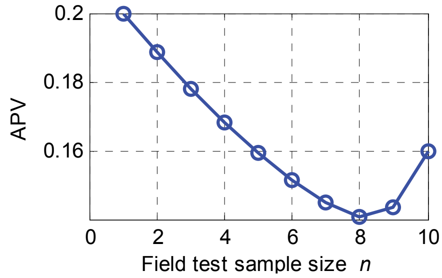

Figure 3: Average posterior variance vs. field test sample size.

5.2 Optimal Design with Required Precision Constraint

Suppose we expect an assessment precision of ε=0.1 (take APVC here) for missile hit accuracy, optimization goal is to get an integrated test scheme with minimal cost. Thus, optimization equation for SSA of integrated test can be built according to (21), and optimal SSA plan under different simulation test credibility can be obtained using a nonlinear constrained optimization algorithm, as shown in Table 3. In this way, under APVC constraint, sample size of field test in optimal test scheme is on the decrease as simulation credibility increases, and meanwhile the total cost for integrated test scheme is decreasing. Exponent of power prior (EoPP) a_0 become larger and larger as the increment of simulation credibility C_0, which indicates simulation test samples with higher credibility take on greater weight in posterior assessment. Thus, less field test samples are needed, and total test cost is reduced.

Table 3: Optimal integrated test schemes under required ε in APVC.

SC	SSA		EoPP	APV	TC
C_0	n_0	n_1	a_0	L_{APV}	T_C
0.70	88	30	0.0552	0.1000	64400
0.80	65	16	0.1056	0.0999	35250
0.90	101	3	0.1290	0.0999	11050
0.95	25	2	0.5576	0.0985	5250

Note that with a given ε=0.1, classical SSD method gives a number of 16. So we can see that with low simulation test credibility (as C_0=0.7, 0.8), integrated test scheme show no predominance, which is due to the negative influence on Bayesian posterior assessment brought by poor prior information. This fact also indicates that if distribution difference of simulation test samples and field test samples is big (namely simulation test takes on low credibility), classical design of experiment and assessment method without consideration on prior information could be more rational. While simulation test credibility is high, adopting integrated test scheme improves test efficiency greatly by saving a great deal of test cost.

In Table 4, we give out several designed test schemes satisfying the required precision ε=0.1 in APVC, where the simulation test credibility is 0.9. It is thus clear that as the increase of simulation test sample size, exponential factor a_0 of power prior decreases, and equivalent sample size tend to finite value. Thus, it would not happen that design prior obtained with large number of simulation test samples has already met with Bayesian SSD criteria, which avoids the credibility risk of posterior assessment for parameter of interest. So, we can see the rationality of equivalent sample size, which is obtained based on equal test design effect.

Table 4: Test schemes for required ε=0.1 in APVC with C_0=0.9.

SSA		ESS	EoPP	APV	TC
n_0	n_1	n_r	a_0	L_{APV}	T_C
187	2	13.7318	0.0734	0.1000	13350
101	3	13.0240	0.1290	0.0999	11050
68	4	12.3517	0.1816	0.0998	11400
50	5	11.6866	0.2337	0.0999	12500
40	6	11.1206	0.2780	0.0995	14000
32	7	10.4854	0.3277	0.0999	15600
27	8	9.9583	0.3688	0.0997	17350

6 CONCLUSIONS

The paper considers SSD problem for integrated test design of missile hit accuracy. Power prior is used in weighted fusion of simulation test samples, and ESS of simulation test based on equal test design effect is proposed to get the design prior of Bayesian SSD for field test. Thus, an advanced Bayesian solution, with systematic and scientific planning concept, for SSA problem in design of integrated test scheme is obtained, which takes on the following features:

(1) SSD is realized by Bayesian pre-posterior calculation. A reasonable estimation for posterior distribution of interested parameter θ can be obtained with comprehensive usage of historical data and prior information before actual field test. Hence, minimal sample size for required assessment precision can be worked out.

(2) Calculation method for ESS of simulation test is given out based on the idea of power prior and test design effect, which get an equivalency of large number of low credible simulation test samples with limited standard filed test samples. Thus, we avoid the discrepancy that taking simulation credibility directly as exponent of power prior, design prior obtained from large numbers of simulation test samples has already met with Bayesian SSD criteria. And in this way, the assessment risk is reduced.

(3) Taking APV of missile hit accuracy as the output, we get the optimization design equations for integrated test scheme with Bayesian SSD method under both test cost constraint and required precision ε constraint in APVC. Thus, we give out a Bayesian SSD method for integrated test design of missile hit accuracy with optimal efficiency.

REFERENCES

Adcock, C. J., 1997. Sample size determination: a review. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 46(2), 261-283.

Balci, O., 1997. Verification validation and accreditation of simulation models, 135-141., In *Proceedings of the 29th conference on Winter simulation*, Atlanta, GA, USA.

Balci, O., 2013. Verification, Validation, and Testing of Models, 1618-1627. In Gass, S., and Harris, C.M., eds., *Encyclopedia of Operations Research and Management Science*. Springer, New York.

Bernardo, J. M., and Smith, A. F. M., 2000. *Bayesian theory*. John Wiley & Sons, New York.

Clarke, B., and Yuan, A., 2006. Closed form expressions for Bayesian sample size. *The Annals of Statistics,* 34(3), 1293-1330.

Claxton, J. D., Cavoli, C., and Johnson, C., 2012. *Test and Evaluation Management Guide*, The Defense Acquisition University Press. Fort Belvoir, VA. 6th Edition.

De Santis, F., 2004. Statistical evidence and sample size determination for Bayesian hypothesis testing. *Journal of statistical planning and inference*, 124(1), 121-144.

De Santis, F., 2007. Using historical data for Bayesian sample size determination. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(1), 95-113.

Di Bacco, M., Amore, G. D., Scalfari, F., De Santis, F., & Pacifico, M. P., 2004. Two experimental settings in clinical trials: predictive criteria for choosing the sample size in interval estimation. *Applied Bayesian Statistical Studies in Biology and Medicine,* 109-130. Kluwer Academic Publishers, Norwell, MA, USA.

Desu, M. M., and Raghavarao, D., 1990. *Sample Size Methodology*. Academic Press, Boston.

Fryback, D. G., Chinnis, J. O., and Ulvila, J. W., 2001. Bayesian cost-effectiveness analysis. *International Journal of Technology Assessment in Health Care,* 17(1), 83-97.

Greenhouse, J. B., and Waserman, L., 1995. Robust Bayesian methods for monitoring clinical trials. *Statistics in Medicine,* 14(12), 1379-1391.

Ibrahim, J. G., Chen, M. H., and Sinha, D., 2003. On optimality properties of the power prior. *Journal of the American Statistical Association,* 98(461), 204-213.

Ibrahim, J. G., and Chen, M. H., 2000. Power prior distributions for regression models. *Statistical Science,* 15(1), 46-60.

Joseph, L., Du Berger, R., Belisle, P., and Others., 1997. Bayesian and mixed Bayesian/likelihood criteria for sample size determination. *Statistics in Medicine,* 16(7), 769-781.

Joseph, L., Wolfson, D. B., and Du Berger, R., 1995. Sample size calculations for binomial proportions via highest posterior density intervals. *The Statistician*, 44(2), 143-154.

Joseph, L., and Belisle, P. 1997. Bayesian sample size determination for normal means and differences between normal means. *Journal of the Royal Statistical Society: Series D (The Statistician),* 46(2), 209-226.

Kraft, E. M., 1995. Integrated Test and Evaluation- A knowledge-based approach to system development., In *1st AIAA Aircraft Engineering, Technology, and Operations Congress*. AIAA 95-3982, Los Angeles, US.

Kushman, K. L., and Briski, M. S., 1992. Integrated test and evaluation for hypervelocity systems. In *AIAA 17th Aerospace Ground Testing Conference*. American Institute of Aeronautics and Astronautics, Nashville, TN.

Murphy, K. R., Myors, B., and Wolach, A. H., 2009. *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests*. Routledge, New York, NY.

Nassar, M. M., Khamis, S. M., and Radwan, S. S., 2011. On Bayesian sample size determination. *Journal of Applied Statistics,* 38(5), 1045-1054.

Rebba, R., Mahadevan, S., and Huang, S. P., 2006. Validation and error estimation of computational models. *Reliability Engineering & System Safety,* 91(10-11), 1390-1397.

Reyes, E. M., and Ghosh, S. K., 2013. Bayesian Average Error-Based Approach to Sample Size Calculations for Hypothesis Testing. *Journal of Biopharmaceutical Statistics*, 23(3), 569-588.

Schwartz, M., 2010. *Defense acquisitions: How DoD acquires weapon systems and recent efforts to reform the process.* Paper presented at the Congressional Research Service, Washington, DC.

Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P., 2004. *Bayesian approaches to clinical trials and health-care evaluation.* John Wiley & Sons, Ltd, Chichester, West Sussex, England.

Spiegelhalter, D. J., and Freedman, L. S., 1986. A predictive approach to selecting the size of a clinical trial, based on subjective clinical opinion. *Statistics in medicine,* 5(1), 1-13.

Wang, F., and Gelfand, A. E., 2002. A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. *Statistical Science,* 17(2), 193-208.

Waters, D. P., 2004. Integrating Modeling and Simulation with Test and Evaluation Activities. In *USAF Developmental Test and Evaluation Summit*. Woodland Hills, California: AIAA 2004-6800.

Weiss, R. 1997. Bayesian sample size calculations for hypothesis testing. *Journal of the Royal Statistical Society: Series D (The Statistician),* 46(2), 185-191.