Supplement to “The Fallacy of Placing Confidence in Confidence Intervals”

Richard D. Morey (richarddmorey@gmail.com),
Rink Hoekstra, Jeffrey N. Rouder,
Michael D. Lee, Eric-Jan Wagenmakers

1 The lost submarine: details

We presented a situation where $N = 2$ observations were distributed uniformly:

$$x_i \stackrel{iid}{\sim} \text{Uniform}(\theta - 5, \theta + 5),\ i = 1, \ldots, N$$

and the goal is to estimate θ, the location of the submarine hatch. Without loss of generality we denote x_1 as the smaller of the two observations. In the text, we considered five 50% confidence procedures; in this section, we give the details about the sampling distribution procedure and the Bayes procedure that were omitted from the text.

1.1 Sampling distribution procedure

Consider the sample mean, $\bar{x} = (x_1 + x_2)/2$. As the sum of two uniform deviates, it is a well-known fact that $\bar{x}$ will have a triangular distribution with location θ and minimum and maximum $\theta - 5$ and $\theta + 5$, respectively. This distribution is shown in Figure 1.

It is desired to find the width of the base of the shaded triangle in Figure 1 such that it has an area of .5. To do this we first find the width of the base of the unshaded triangular area marked “a” in Figure 1 such that the area of the triangle is .25. The corresponding unshaded triangle on the left side will also have area .25, which means that since the figure is a density, the shaded region must have the remaining area of .5. Elementary geometry will show that the width of the base of triangle “a” is $5/\sqrt{2}$, meaning that the distance between θ and the altitude of triangle “a” is $5 - 5/\sqrt{2}$ or about 1.46m.

We can thus say that

$$Pr(-(5 - 5/\sqrt{2}) < \bar{x} - \theta < 5 - 5/\sqrt{2}) = .5$$

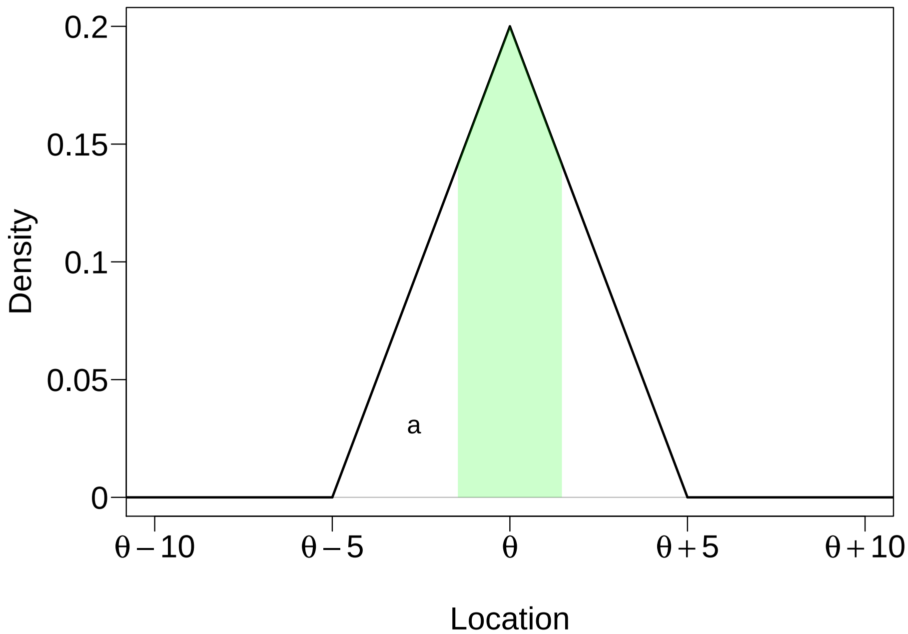

Figure 1: The sampling distribution of the mean $\bar{x}$ in the submarine scenario. The shaded region represents the central 50% of the area. The unshaded triangle marked “a” has area .25.

which implies that, in repeated sampling,

$$Pr(\bar{x} - (5 - 5/\sqrt{2}) < \theta < \bar{x} + (5 - 5/\sqrt{2})) = .5$$

which defines the sampling distribution confidence procedure. This is an example of using $\bar{x} - \theta$ as a pivotal quantity (Casella & Berger, 2002).

1.2 Bayesian procedure

The posterior distribution is proportional to the likelihood times the prior. The likelihood is

$$p(x_1, x_2 \mid \theta) \propto \prod_{i=1}^{2} \mathcal{I}(\theta - 5 < x_i < \theta + 5);$$

where $\mathcal{I}$ is an indicator function. Note since this is the product of two indicator functions, it can only be nonzero when both indicator functions' conditions are met; that is, when $x_1 + 5$ and $x_2 + 5$ are both greater than θ, and $x_1 - 5$ and $x_2 - 5$ are both less than θ. If the minimum of $x_1 + 5$ and $x_2 + 5$ is greater than θ, then so to must be the maximum. The likelihood thus can be rewritten

$$p(x_1, x_2 \mid \theta) \propto \mathcal{I}(x_2 - 5 < \theta < x_1 + 5);$$

where x_1 and x_2 are the minimum and maximum observations, respectively. If the prior for θ is proportional to a constant, then the posterior is

$$p(\theta \mid x_1, x_2) \propto \mathcal{I}(x_2 - 5 < \theta < x_1 + 5),$$

This posterior is a uniform distribution over all *a posteriori* possible values of θ (that is, all θ values within 5 meters of all observations), has width

$$10 - (x_2 - x_1),$$

and is centered around $\bar{x}$. Because the posterior comprises all values of θ the data have not ruled out – and is essentially just the classical likelihood – the width of this posterior can be taken as an indicator of the precision of the estimate of θ.

The middle 50% of the likelihood can be taken as a 50% objective Bayesian credible interval. Proof that this Bayesian procedure is also a confidence procedure is trivial and can be found in Welch (1939).

2 Credible interval for ω^2: details

In the manuscript, we compare Steiger's (2004) confidence intervals for ω^2 to Bayesian highest posterior density (HPD) credible intervals. In this section we describe how the Bayesian HPD intervals were computed.

Consider a one-way design with J groups and N observations in each group. Let y_{ij} be the ith observation in the jth group. Also suppose that

$$y_{ij} \stackrel{indep.}{\sim} \text{Normal}(\mu_j, \sigma^2)$$

where μ_j is the population mean of the jth group and σ^2 is the error variance. We assume a "non-informative" prior on parameters $\boldsymbol{\mu}, \sigma^2$:

$$p(\mu_1, \ldots, \mu_J, \sigma^2) \propto (\sigma^2)^{-1}.$$

This prior is flat on $(\mu_1, \ldots, \mu_J, \log \sigma^2)$. In application, it would be wiser to assume an informative prior on these parameters, in particular assuming a population over the μ parameters or even the possibility that $\mu_1 = \ldots = \mu_J = 0$ (Rouder, Morey, Speckman, & Province, 2012). However, for this manuscript we compare against a "non-informative" prior in order to show the differences between the confidence interval and the Bayesian result with "objective" priors.

Assuming the prior above, an elementary Bayesian calculation (Gelman, Carlin, Stern, & Rubin, 2004) reveals that

$$\sigma^2 \mid \boldsymbol{y} \sim \text{Inverse Gamma}(J(N-1)/2, S/2)$$

where S is the error sum-of-squares from the corresponding one-way ANOVA, and

$$\mu_j \mid \sigma^2, \boldsymbol{y} \stackrel{indep.}{\sim} \text{Normal}(\bar{x}_j, \sigma^2/N)$$

where μ_j and $\bar{x}_j$ are the true and observed means for the jth group. Following Steiger (2004) we can define

$$\alpha_j = \mu_j - \frac{1}{J}\sum_{j=1}^{J} \mu_j$$

as the deviation from the grand mean of the jth group, and

$$\begin{aligned} \lambda &= N \sum_{j=1}^{J} \left(\frac{\alpha}{\sigma}\right)^2 \\ \omega^2 &= \frac{\lambda}{\lambda + NJ}. \end{aligned}$$

It is then straightforward to set up an MCMC sampler for ω^2. Let M be the number of MCMC iterations desired. We first sample M samples from the marginal posterior distribution of σ^2, then sample the group means from the conditional posterior distribution for $\mu_1, \ldots, \mu_J$. Using these posterior samples, M posterior samples for λ and ω^2 can be computed.

The following function will sample from the marginal posterior distribution of ω^2:

```
## Assumes that data.frame y has two columns:
## $y is the dependent variable
## $grp is the grouping variable, as a factor
Bayes.posterior.omega2

## function (y, conf.level = 0.95, iterations = 10000)
## {
##     J = nlevels(y$grp)
##     N = nrow(y)/J
##     aov.results = summary(aov(y ~ grp, data = y))
##     SSE = aov.results[[1]][2, 2]
##     sig2 = 1/rgamma(iterations, J * (N - 1)/2, SSE/2)
##     lambda = matrix(NA, iterations)
##     group.means = tapply(y$y, y$grp, mean)
##     for (m in 1:iterations) {
##         mu = rnorm(J, group.means, sqrt(sig2[m]/N))
##         lambda[m] = N * sum((mu - mean(mu))^2/sig2[m])
##     }
##     mcmc(lambda/(lambda + N * J))
## }
```

The `Bayes.posterior.omega2` function can be used to compute the posterior and HPD for the first example in the manuscript. The `fake.data.F` function, defined in the file `steiger.utility.R` (available with the manuscript source code at https://github.com/richarddmorey/ConfidenceIntervalsFallacy), generates a data set with a specified F statistic.

```
cl = .683 ## Confidence level corresponding to standard error
J = 3 ## Number of groups
N = 10 ## observations in a group

df1 = J - 1
df2 = J * (N - 1)

## F statistic from manuscript
Fstat = 0.1748638

set.seed(1)
y = fake.data.F(Fstat, df1, df2)

## Steiger confidence interval
steigerCI = steigerCI.omega2(Fstat,df1,df2, conf.level=cl)
samples.omega2 = Bayes.posterior.omega2(y, cl, 100000)
```

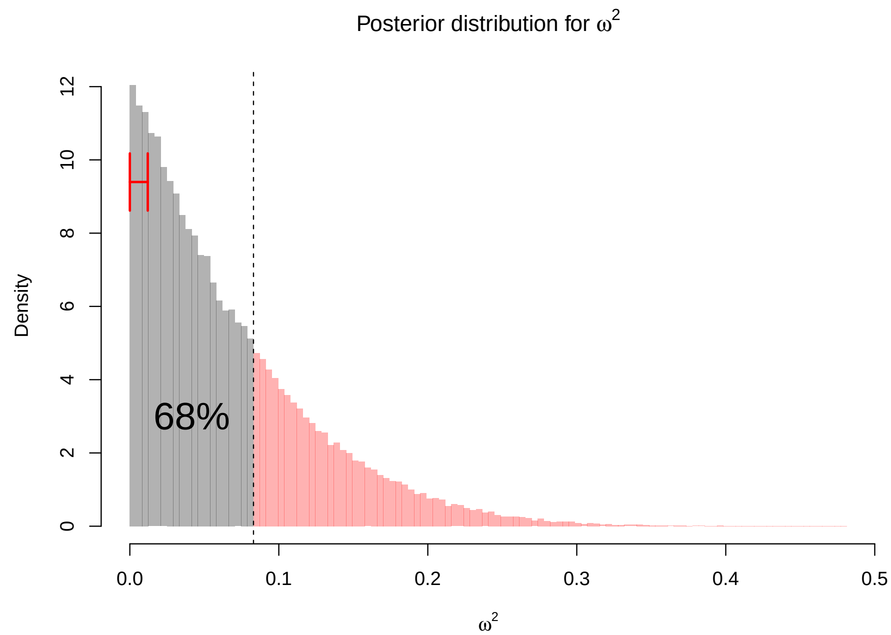

Figure 2: Histogram of the posterior MCMC samples for ω^2. The 68% Bayesian HPD credible interval is highest density region than captures 68% of the posterior density, shown in gray. The vertical dashed line denotes the upper bound of the HPD. The 68% Steiger confidence interval is shown as the interval near the top.

We can compute the Bayesian HPD interval with the 'HPDinterval' function in the package 'coda':

```
library(coda)

HPDinterval( samples.omega2, prob = cl )

##              lower      upper
## var1 5.219606e-06 0.08299102
## attr(,"Probability")
## [1] 0.683
```

References

Casella, G. & Berger, R. L. (2002). *Statistical inference*. Pacific Grove, CA: Duxbury.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis (2nd edition)*. London: Chapman and Hall.

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*, 356–374.

Steiger, J. H. (2004). Beyond the *F* test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods*, *9*(2), 164–182.

Welch, B. L. (1939). On confidence limits and sufficiency, with particular reference to parameters of location. *The Annals of Mathematical Statistics*, *10*(1), 58–69. Retrieved from http://www.jstor.org/stable/2235987