

Confidence intervals & credible intervals

Kevin Stadler

Mon Mar 30 10:20:45 2015

Hoekstra et al. (2014) show how nobody (not even statistics professors) know/understand what confidence intervals (CIs) are, how to interpret them or what they are supposed to tell you. Even if you know what they are, in the heat of reading a paper/glimpsing at a graph you are much more likely to interpret them intuitively as something they are not: a 95% confidence interval does *not* tell you that there's a 95% probability of the true mean lying in this interval - confidence intervals are a *frequentist* concept, meaning that there is the assumption of one underlying true mean. The particular CI you're looking at will either contain this (unknown) true mean or it will not. The idea is that, if you were to repeat your sampling procedure on the same population, 95% of the samples' CIs would contain the true mean. But, importantly, for the particular sample you're looking at the CI gives you *no indication of what the true mean is most likely to be, or what region it is likely to lie in!*

If you're interested in knowing the range of possible/likely values of the true mean based on your sample (the intuitive reading that most people go for), you are actually thinking of the [credible interval](#) (aka 'Bayesian confidence interval') for that parameter. The difference between those two types of intervals isn't just a philosophical one, because the Bayesian credible interval will generally be *wider* than the confidence interval, so people's 'intuitive' reading of CI's is an underestimate of the measure that they think they're looking at!

One conclusion from the paper is that this is yet another reason we should all go Bayes, but the obvious next question was: how do you (easily) calculate a credible interval? Let's try:

```
# Make sure results are reproducible
set.seed(101)
# Take a small normally distributed sample
x <- rnorm(10)

# Use a one sample t-test to get the confidence interval for the mean
t.test(x)
```

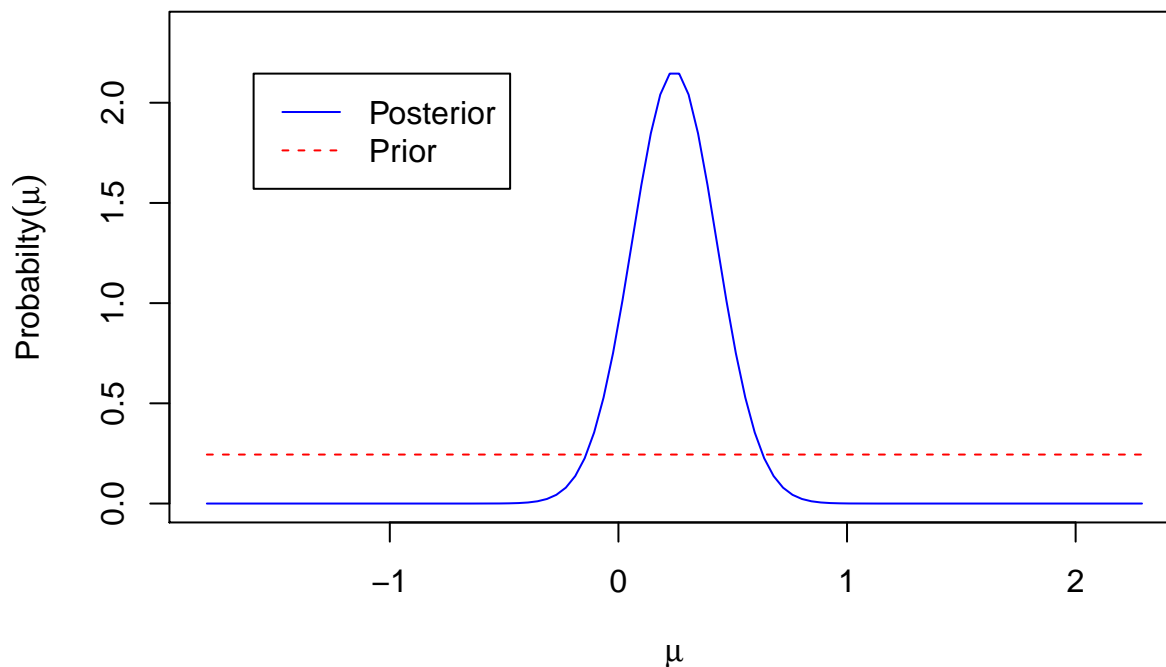
```
##
## One Sample t-test
##
## data: x
## t = 1.3259, df = 9, p-value = 0.2175
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -0.1730163 0.6630965
## sample estimates:
## mean of x
## 0.2450401
```

Determining a credible interval (aka Bayesian confidence interval) for the mean

I first tried the [Bolstad](#) package, which has functions for inferring the most likely mean of a sample that's assumed to be normally distributed. Not making any assumptions about our sample, we use `normgcp` to infer the mean using a flat (uninformative) prior:

```
# the Bayesian inference occurs over a finite number of possible means. n.mu sets
# the number of means considered, which determines the resolution of the posterior
mu.mdl <- Bolstad::normgcp(x, n.mu=100)
```

```
## Standard deviation of the residuals :0.5844
```



But how do we go from the posterior distribution to a credible interval? A first (naive) attempt at getting at a credible interval for the mean by cutting 2.5% off either side:

```
limit <- sum(mu.mdl$posterior)*0.025
lower <- which(cumsum(mu.mdl$posterior)>=limit)[1]
upper <- length(mu.mdl$posterior) - which(cumsum(rev(mu.mdl$posterior))>=limit)[1]
# naive credible interval assuming fixed (known) sigma:
mu.mdl$mu[c(lower,upper)]
```

```
## [1] -0.1061917 0.5549505
```

This interval is actually *smaller* than the confidence interval determined by the t-test above. What's going on? The issue can be seen in the output of `normgcp` above: the function only infers the *mean* of the distribution without considering the *standard deviation* which is unknown, but simply assumed to be identical to the standard deviation of the sample. Known standard deviations can be passed as arguments to the function, but there is normally no reason to assume that we know the standard deviation. Really we will have to make inferences about the mean and standard deviation in tandem (and thus infer a credibility *region* within the distribution's multi-dimensional parameter space).

Bayesian inference over a (ostensibly) normally distributed sample

Here comes *BEST*: *Bayesian estimation supersedes the t test* (Kruschke 2013):

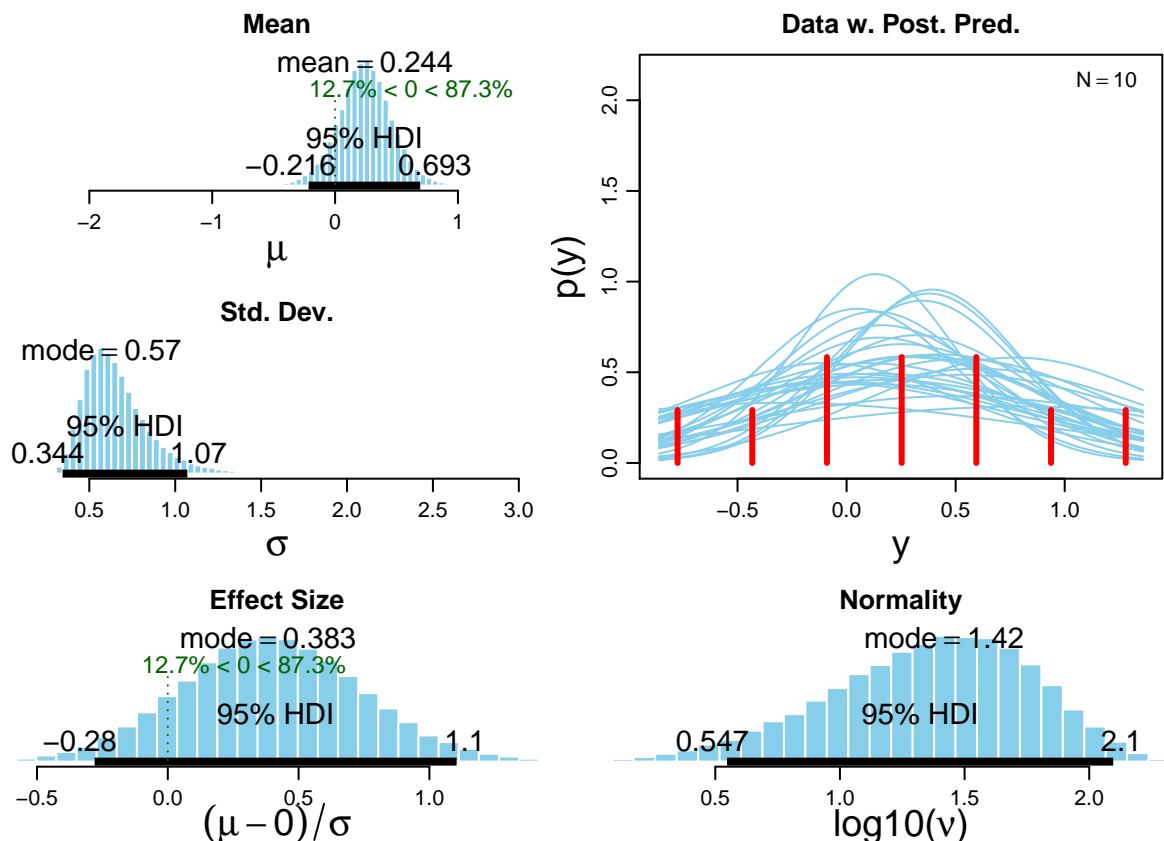
```
library(BEST)
best.mdl <- BESTmcmc(x, verbose=FALSE)
as.matrix(hdi(best.mdl))
```

```
##           mu           nu       sigma
## lower -0.2159010  1.002321 0.3444416
## upper  0.6930611 91.909870 1.0705047
## attr(,"credMass")
## [1] 0.95
```

Not only does this procedure give us a wider credible interval on μ (as we expected), it also gives estimates for standard deviation and ν , a measure of the sample distribution's normality.

`hdi` stands for *highest density interval* - since the posterior can have any (non-symmetric) shape, there are many 95% “credible intervals”, i.e. many ways to carve out a region that covers 95% of the probability mass. Rather than simply cutting off 2.5% of the probability mass on either end of the posterior, you normally want to select the 95% region which has a greater probability density than any part outside the region. The highest density region of the different parameters is marked in black below¹:

```
plotAll(best.mdl)
```



The relative difference between the confidence intervals' and credible intervals' width depends strongly on the sample size, primarily because we cannot make any strong inferences about the true standard deviation of

¹NB it is not necessarily always the case that the posterior distribution is unimodal, particularly if you're fitting a complex model. In this case the highest density interval might actually be *discontinuous*, i.e. it might actually consist of two disconnected intervals at different locations in the parameter space!

the underlying distribution from a small sample. Having to assume that the standard deviation could be very high means that very different settings of the mean don't actually affect the likelihood of the sample that much, leading to a flatter posterior distribution over possible mean values and a wide credible interval.

```
compare.cis <- function(x) {
  cat("Confidence interval:           ", as.vector(t.test(x)$conf.int), "\n")
  best.mdl <- BESTmcmc(x, verbose=FALSE)
  cat("Highest-density credible interval: ", hdi(best.mdl)[,"mu"], "(sd", hdi(best.mdl)[,"sigma"], ")\n")
  cat("credible:confidence interval ratio: ", unname(diff(hdi(best.mdl)[,"mu"]) / diff(t.test(x)[,"conf
}

# for n=4 the credible interval is on average about 50% wider than the CI
compare.cis(rnorm(4))
```

```
## Confidence interval:           -2.146544 1.992814
## Highest-density credible interval: -3.764133 3.273749 (sd 0.4557651 7.448849 )
## credible:confidence interval ratio: 1.700235
```

```
compare.cis(rnorm(5))
```

```
## Confidence interval:           -0.9106303 0.0950379
## Highest-density credible interval: -1.06393 0.2471275 (sd 0.182476 1.375355 )
## credible:confidence interval ratio: 1.303668
```

```
compare.cis(rnorm(6))
```

```
## Confidence interval:           -1.60889 0.7778873
## Highest-density credible interval: -1.863667 1.008553 (sd 0.5556992 3.029527 )
## credible:confidence interval ratio: 1.203388
```

```
compare.cis(rnorm(8))
```

```
## Confidence interval:           -0.374494 0.8956657
## Highest-density credible interval: -0.3665392 0.946678 (sd 0.2493422 1.50804 )
## credible:confidence interval ratio: 1.033899
```

```
compare.cis(rnorm(10))
```

```
## Confidence interval:           -1.220896 0.6566758
## Highest-density credible interval: -1.273916 0.7712671 (sd 0.7766824 2.427237 )
## credible:confidence interval ratio: 1.08927
```

```
compare.cis(rnorm(20))
```

```
## Confidence interval:           -0.6912493 0.1022275
## Highest-density credible interval: -0.711183 0.1145866 (sd 0.5904682 1.2189 )
## credible:confidence interval ratio: 1.040698
```

```
compare.cis(rnorm(30))
```

```
## Confidence interval:          -0.163708 0.5585145  
## Highest-density credible interval: -0.1665879 0.5765447 (sd 0.7005191 1.274556 )  
## credible:confidence interval ratio: 1.028952
```

```
compare.cis(rnorm(40))
```

```
## Confidence interval:          -0.4084887 0.2142104  
## Highest-density credible interval: -0.4354259 0.1986307 (sd 0.7022036 1.222414 )  
## credible:confidence interval ratio: 1.018239
```

```
compare.cis(rnorm(50))
```

```
## Confidence interval:          -0.390132 0.2065673  
## Highest-density credible interval: -0.3815945 0.2200547 (sd 0.8022392 1.269946 )  
## credible:confidence interval ratio: 1.008295
```

References

- Hoekstra, Rink, Richard D Morey, Jeffrey N Rouder, and Eric-Jan Wagenmakers. 2014. “Robust Misinterpretation of Confidence Intervals.” *Psychonomic Bulletin & Review* 21 (5): 1157–1164. doi:[10.3758/s13423-013-0572-3](https://doi.org/10.3758/s13423-013-0572-3). <http://www.ncbi.nlm.nih.gov/pubmed/24420726>.
- Kruschke, John K. 2013. “Bayesian Estimation Supersedes the T Test.” *Journal of Experimental Psychology: General* 142 (2): 573–603. doi:[10.1037/a0029146](https://doi.org/10.1037/a0029146). <http://www.indiana.edu/~kruschke/BEST/>.