

Fuzzy Error Bars

Charles J. Geyer

May 20, 2005

1 Introduction

This document tries out “fuzzy error bars” using the R package `ump`. The idea was mentioned in Geyer and Meeden (2005), but was not adequately explored due to lack of space.

2 Example 1: Keeping It Simple

For our first example, we make up some data.

```
> set.seed(42)
> fred <- c("red", "orange", "yellow", "green", "blue", "indigo",
+          "violet", "cyan", "magenta", "white", "black")
> n <- seq(15, by = 5, length = length(fred))
> x <- rbinom(length(n), n, 0.5)
```

and calculate the fuzzy confidence intervals

```
> library(ump)
> p <- seq(0, 1, 1e-04)
> alpha <- 0.05
> phi <- matrix(NA, length(p), length(fred))
> for (i in 1:ncol(phi)) phi[, i] <- umpu.binom(x[i], n[i], p,
+       alpha)
> dimnames(phi) <- list(as.character(p), fred)
```

While we are at it, we calculate the core

```
> make.core <- function(phi) {
+   nam <- dimnames(phi)[[2]]
+   p <- as.numeric(dimnames(phi)[[1]])
```

```

+   stopifnot(all(0 <= p & p <= 1))
+   foo <- matrix(NaN, ncol(phi), 2)
+   for (i in 1:ncol(phi)) {
+     phii <- phi[, i]
+     foo[i, 1] <- min(p[1 - phii == 1])
+     foo[i, 2] <- max(p[1 - phii == 1])
+   }
+   dimnames(foo) <- list(nam, c("low", "high"))
+   return(foo)
+ }
> make.core(phi)

```

	low	high
red	0.4466	0.8472
orange	0.4558	0.8152
yellow	0.2744	0.6154
green	0.4375	0.7484
blue	0.3923	0.6883
indigo	0.3600	0.6400
violet	0.4209	0.6848
cyan	0.2984	0.5486
magenta	0.4049	0.6474
white	0.3994	0.6328
black	0.5020	0.7207

and support

```

> make.support <- function(phi) {
+   nam <- dimnames(phi)[[2]]
+   p <- as.numeric(dimnames(phi)[[1]])
+   stopifnot(all(0 <= p & p <= 1))
+   foo <- matrix(NaN, ncol(phi), 2)
+   for (i in 1:ncol(phi)) {
+     phii <- phi[, i]
+     foo[i, 1] <- min(p[1 - phii > 0])
+     foo[i, 2] <- max(p[1 - phii > 0])
+   }
+   dimnames(foo) <- list(nam, c("low", "high"))
+   return(foo)
+ }
> make.support(phi)

```

	low	high
red	0.3802	0.8959
orange	0.4054	0.8525
yellow	0.2400	0.6531
green	0.4046	0.7767
blue	0.3651	0.7141
indigo	0.3364	0.6636
violet	0.3991	0.7051
cyan	0.2803	0.5685
magenta	0.3874	0.6646
white	0.3832	0.6486
black	0.4863	0.7347

Figure 1 is produced by the following code

```

> par(mar = c(5, 6, 1, 1) + 0.1)
> delta <- 0.3
> foo <- make.support(phi)
> plot(range(foo), c(0, length(fred) - delta), type = "n", axes = FALSE,
+       ylab = "", xlab = "")
> axis(side = 1)
> box()
> for (i in 1:ncol(phi)) {
+   phii <- phi[, i]
+   ppp <- p[1 - phii > 0]
+   xxx <- c(ppp, rev(range(ppp)))
+   yyy <- (1 - phii)[1 - phii > 0]
+   yyy <- c(yyy, 0, 0)
+   yyy <- i - 1 + yyy * (1 - delta)
+   polygon(xxx, yyy)
+   lines(rep(x[i]/n[i], 2), c(i - 1, max(yyy)))
+ }
> axis(side = 2, at = seq(along = fred) - 1 + (1 - delta)/2, lab = fred,
+       las = 1, padj = 1)
> title(xlab = expression(p))

```

and appears on p. 4.

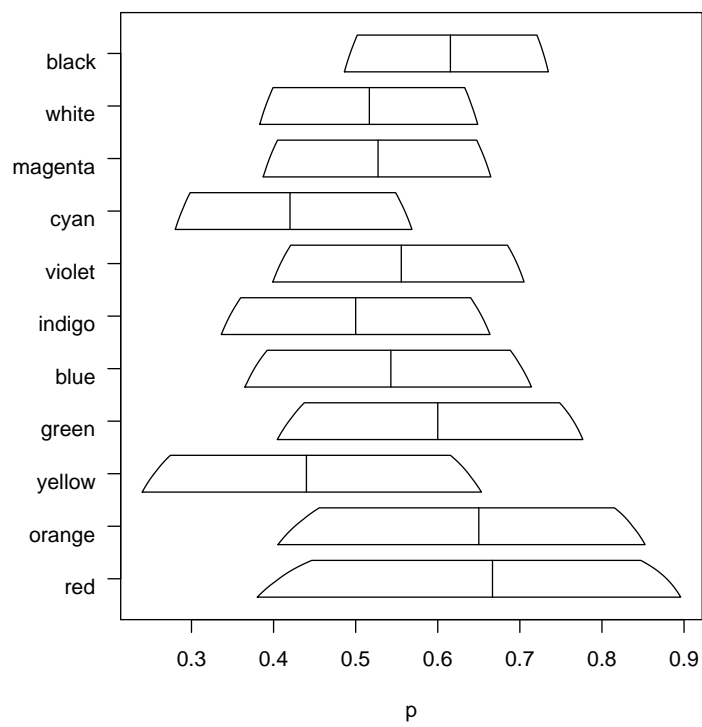


Figure 1: Multiple Fuzzy Error Bars. Fuzzy confidence intervals associated with the UMPU test for the binomial distribution. Sample sizes range between 15 and 65. Height of the error bar denotes the amount of fuzziness (full height, no fuzziness). Line in the middle of each bar denotes the MLE.

3 Example 2: Pointy Haired Bars (PHB)

The associate editor handling Geyer and Meeden (2005) quite rightly pointed out that it is not *always* the case that our fuzzy error bars have full height, so the implied vertical axis for a fuzzy error bar goes from zero at the bottom to one at the top.

However these special cases are self-labeling by their special form. We call them *pointy haired bars* (PHB), a Dilbert allusion. To see how this works, let's do an example of the kind of confidence interval explanation seen in many introductory statistics textbooks where a single plot shows the result of a small simulation study, and (sure enough) about 95% of the 95% confidence intervals for simulated data cover the "simulation truth" parameter value and about 5% don't cover.

Intro stats books usually do this with normal data (or at least quasicontinuous data). We'll use binomial. Also (to be fair to the intro stats books) we obey the "more than five expected in each cell" rule for contingency tables with $n = 100$ and $p_0 = 0.05$ for the simulation truth. Thus we are in "large n " territory according to the time honored (but dumb) rule. Some newer textbooks, such as Wild and Seber (2000) do strongly criticize the time honored rule of thumb.

Let's do it.

```
> n <- 100
> p0 <- 0.05
> nsim <- 50
> x <- rbinom(nsim, n, p0)
> print(x)
```

```
[1] 2 5 5 8 3 11 9 2 5 4 8 5 7 6 7 4 6 0 7 1 3 8 5 4 4
[26] 2 10 4 9 8 6 9 6 4 4 4 7 2 6 6 3 4 5 6 10 6 5 7 3 4
```

O. K. We got a zero, so we'll see the problem which is the ostensible point of this section. Figure 2 is produced by the code very similar to Figure 1, so we will not show the actual code. The figure appears on p. 6.

Hmmmmm. A bit cluttered with 50 intervals, but we'll leave it. The fuzziness can still be seen.

Oh. Almost forgot. How well did we do?

```
> coverage <- 1 - phi[p == 0.05, ]
> sum(coverage == 1)
```

```
[1] 38
```

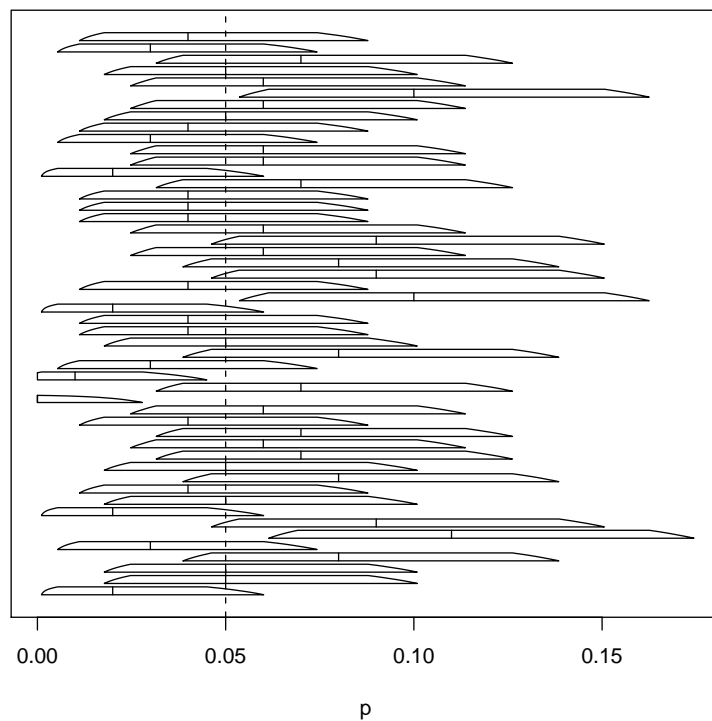


Figure 2: Multiple Fuzzy Error Bars. Fuzzy confidence intervals associated with the UMPU test for the binomial distribution. Height of the error bar denotes the amount of fuzziness (full height, no fuzziness). All bars with flat top have full height. The bar without flat top (number 18 counting from bottom) only goes up to the confidence level, which is 0.9. Line in the middle of each bar denotes the MLE. Vertical dashed line is the “simulation truth” parameter value 0.05. All sample sizes $n = 100$.

```

> sum(coverage == 0)

[1] 5

> sort(coverage[0 < coverage & coverage < 1])

[1] 0.5921725 0.5921725 0.5921725 0.7475198 0.7475198 0.7475198 0.7475198

> mean(coverage)

[1] 0.855332

> sd(coverage)/sqrt(nsim)

[1] 0.04383416

```

So we fully covered 38 times, completely missed 5 times, and partially covered 7 times. The empirical expected coverage was 0.8553 as opposed to the nominal 0.9 level of the intervals. Not bad for simulation sample size 50, only a little more than one standard error (Monte Carlo standard error 0.0438) away from the theoretically correct result.

4 Example 3: Technical Quibbles

It is tempting to describe the exceptional interval in Figure 2 as abutting the end of the data space (zero), but that's not a good description. Because the interval two above it, *also* satisfies that description, but is not problematic.

To see this clearly, here's Figure 2 redone to just show the intervals for $x = 0$ or $x = 1$. This is Figure 3, which appears on p. 8.

Thus we see that the correct description of *full height* fuzzy error bars is *having flat top*, not *abutting the end of the range*. There's another technical reason for this quibble as the following case shows.

```

> n <- 5
> x <- 2
> alpha <- 2/3
> phii <- umpu.binom(x, n, p, alpha)

```

This is Figure 4, which appears on p. 9.

Admittedly a 33.33% confidence interval has no conceivable application. We show it only to make the technical point that very low confidence level

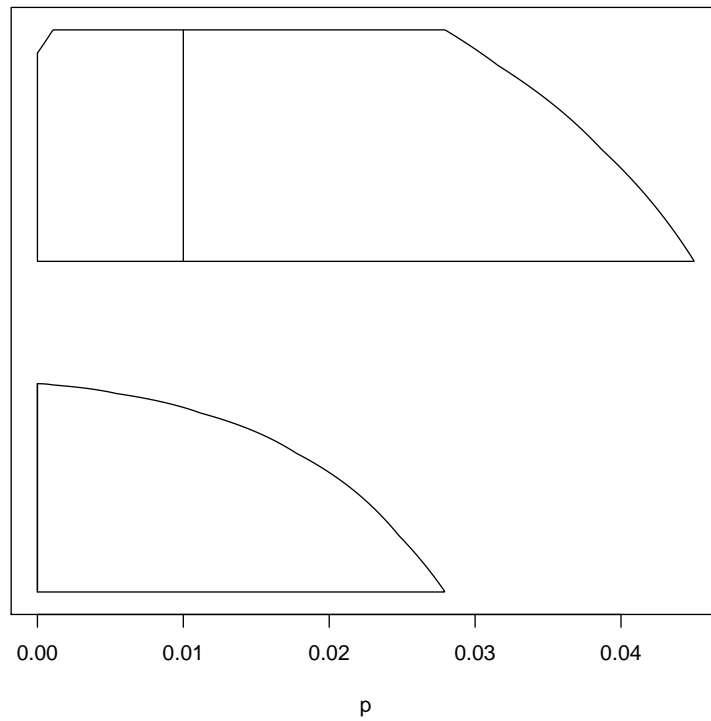


Figure 3: Fuzzy Error Bars that Abut the End. Fuzzy confidence intervals from Figure 2 that have zero for their left endpoint. The bar with flat top has full height. The bar with pointed top only goes up to the confidence level, which is 0.9.

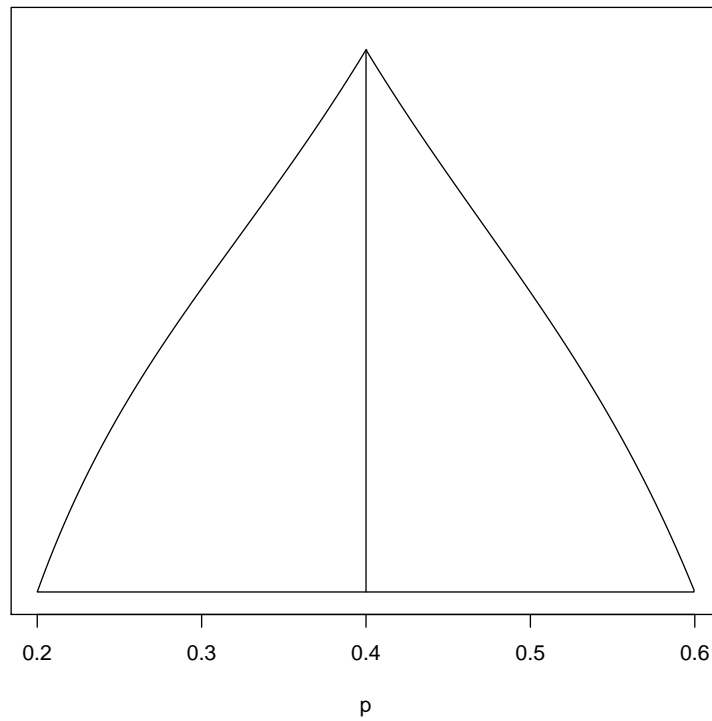


Figure 4: Another Fuzzy Error Bar without Full Height. Fuzzy confidence interval associated with the UMPU test for the binomial distribution. Data $x = 2$, $n = 5$, confidence level 33.33%. The value of the membership function of the fuzzy interval goes from zero at the bottom to 0.965 at the apex.

combined with small n can give another case where the fuzzy error bar does not have full height. And it is also another case where *full height* goes with *flat top* and *non-full height* or *empty core* goes with *pointed top*.

For the curious, the reason for the pointed top is that in order to reject at such a high α we must reject, at least partially, at *all* points, even the point x , even when $p = \hat{p}$, where the fuzzy confidence interval membership function always achieves its maximum (possibly one, possibly less than one). This latter claim is a theorem from unpublished notes of the author: the membership function of a fuzzy confidence interval associated with the UMPU test for a one-parameter discrete exponential family achieves its maximum when the parameter value equals the MLE.

After that bit of theory, we see that abstractly the behavior in the bottom interval of Figure 3 is just the same as the behavior of the interval in Figure 4. The only difference is that in Figure 3, the MLE of the mean value parameter is on the boundary of the sample space, and when that happens, the MLE distribution concentrates all probability at the point on the boundary, and thus we see the irregularity for all α levels. But when the phenomenon occurs in the interior of the sample space, we see the irregularity only for high α levels, hence low confidence levels

$$1 - \alpha < f_{\hat{\mu}}(x)$$

where $\hat{\mu} = x$ is the MLE of the mean value parameter. See equations (3.7a) and (3.7b) and the following discussion in Geyer and Meeden (2005) for explanation.

References

- GEYER, C. J. and MEEDEN, G. D. (2005). Fuzzy and randomized confidence intervals and p-values. Revised and resubmitted (thrice) to *Statistical Science*, <http://www.stat.umn.edu/geyer/fuzz>.
- WILD, C. J. and SEBER, G. A. F. (2000). *Chance Encounters: A First Course in Data Analysis and Inference*. Wiley.