Difference method applied to summary statistics with unequal sample sizes

Jon Minton

Thursday, December 04, 2014

# Introduction

This appendix illustrates how the difference method can be applied in a typical case where the estimates from the lower and the higher states have different sample sizes. For example, imagine the information available was as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| Parameter | Mean | Standard Deviation | Sample Size |
| Lower | 0.35 | 0.20 | 40 |
| Higher | 0.40 | 0.15 | 10 |

We can see through some simple calculations and statistical simulation that some proportion of the paired estimates will violate the monotonicity assumption.

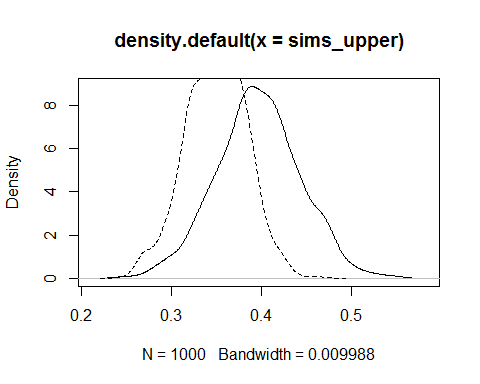
# Upper end of the lower distribution  
0.35 + 2 \* 0.2/(40^0.5)

## [1] 0.4132

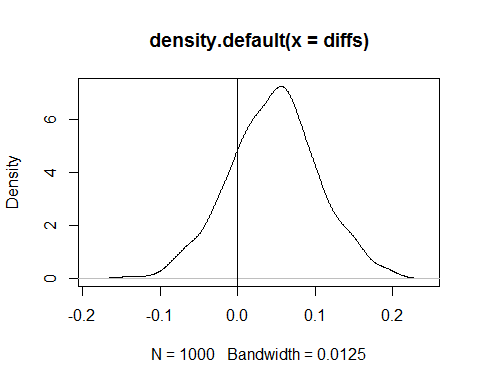
# So, higher than the centre of the Higher distribution  
# Lower end of the higher distribution  
0.40 - 2 \* 0.15/(10^0.5)

## [1] 0.3051

sims\_lower <- rnorm(1000, 0.35, 0.2/(40^0.5))  
sims\_upper <- rnorm(1000, 0.40, 0.15/(10^0.5))  
plot(density(sims\_upper))  
lines(density(sims\_lower), lty="dashed")



diffs <- sims\_upper - sims\_lower  
plot(density(diffs))  
abline(v=0)

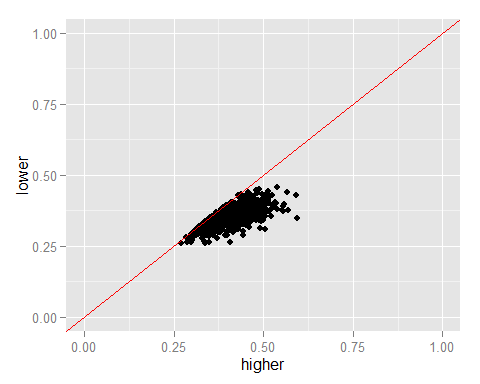


# This shows that some of the density of differences is on the wrong side of the zero line.   
# Let's quantify this:  
length(which(diffs<0))/1000

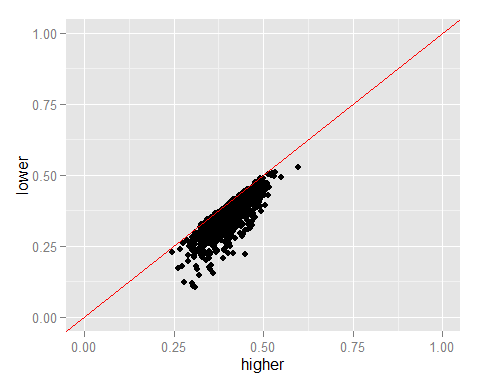
## [1] 0.208

# Estimation using Difference method

library(ggplot2)  
# Specify data  
lower\_mean = 0.35  
lower\_var = (0.20/(40^0.5))^2  
  
higher\_mean = 0.40  
higher\_var = (0.15/(10^0.5))^2  
  
n\_psa=1000  
  
# Load functions  
est\_beta <- function(mu, var) {  
 a <- mu \* ((1 - mu) \* (mu / var) - 1)  
 b <- a \* ((1 - mu) / mu)  
 return(list(a=a, b=b))  
}  
  
get\_dif\_param <- function(  
 u1\_mu, u1\_sd,   
 u2\_mu, u2\_sd,   
 quietly=T  
 ){  
 mu <- u1\_mu - u2\_mu  
   
 sigma2 <- ifelse(u1\_sd > u2\_sd, u1\_sd^2 - u2\_sd^2, u2\_sd^2 - u1\_sd^2)  
 x <- (1 - mu) / mu  
   
 a <- (x/sigma2-1-2\*x-x^2)/(1+3\*x+3\*x^2+x^3)  
 b<-a\*x  
   
 if(quietly==F){  
 print(a/(a+b)) # check mean of delta  
 print(a\*b/(a+b)^2/(a+b+1)) # check variance of delta   
 }  
 return(list(a=a, b=b))  
}  
  
# If upper is simulated as lower + difference  
beta\_lower <- est\_beta(  
 mu=lower\_mean,  
 var=lower\_var  
 )  
  
beta\_dif <-get\_dif\_param(  
 u1\_mu=higher\_mean,  
 u1\_sd=higher\_var^0.5,  
 u2\_mu=lower\_mean,  
 u2\_sd=lower\_var^0.5  
 )  
  
sims\_lower <- rbeta(n\_psa, beta\_lower$a, beta\_lower$b)  
sims\_dif <- rbeta(n\_psa, beta\_dif$a, beta\_dif$b)  
outputs <- data.frame(  
 lower=sims\_lower,  
 higher=sims\_lower + sims\_dif  
 )   
  
qplot(data=outputs, y=lower, x=higher) + geom\_abline(slope=1, intercept=0, colour="red") + xlim(c(0,1)) + ylim(c(0,1))



# If lower is simulated as upper - difference  
  
beta\_higher <- est\_beta(  
 mu=higher\_mean,  
 var=higher\_var  
 )  
  
beta\_dif <-get\_dif\_param(  
 u1\_mu=higher\_mean,  
 u1\_sd=higher\_var^0.5,  
 u2\_mu=lower\_mean,  
 u2\_sd=lower\_var^0.5  
 )  
  
sims\_higher <- rbeta(n\_psa, beta\_higher$a, beta\_higher$b)  
sims\_dif <- rbeta(n\_psa, beta\_dif$a, beta\_dif$b)  
outputs <- data.frame(  
 lower=sims\_higher - sims\_dif,  
 higher=sims\_higher  
 )   
  
qplot(data=outputs, y=lower, x=higher) + geom\_abline(slope=1, intercept=0, colour="red") + xlim(c(0,1)) + ylim(c(0,1))



# Discussion

It is apparent from the scatterplots above that the Difference method does not produce estimates which are on the wrong side of the parity. However, it is also apparent from these images that, in the cases above, the choice of the reference distribution (lower in the first example, and higher in the second example) affects the pattern of scatter. In the absence of additional information is appears neither pattern of scatter is more plausible than the other, but further investigation may be warrented.