

Bayesian Basics for Estimating True Effect Size

Nathan (Nat) Goodman

September 15, 2019

This working paper shows an R implementation of Bayes's formula as used in the blog post [What's the Effect Size?](#) to estimate true effect size in two group difference-of-mean studies. The code is adapted from the file `R/stats_effit.R` (function `posterior`) in my [GitHub repo](#).

Bayes's Formula

The line of math below is Bayes's formula for example at hand. An important nuance is that the "probabilities" are *probability densities* - think R's `dnorm`.

$$P(d_{pop} | d_{obs}) = \frac{P(d_{pop}) \times P(d_{obs} | d_{pop})}{P(d_{obs})}$$

1. The term on the left hand side is what we're trying to compute, namely, the posterior probability distribution of d_{pop} for a given value of d_{obs} (0.5 in the blog post's running example).
2. The first term on the right hand side, $P(d_{pop})$, is the prior. For the normal prior in Figures 2 and 3 of the post, this is R's `dnorm` with *mean* = 0.3 and *sd* = 0.1.
3. The next term, $P(d_{obs} | d_{pop})$, is the probability distribution of d_{obs} for a given d_{pop} . For the two group difference-of-mean studies in the post, this is a noncentral t-distribution centered on d_{obs} .
4. The denominator, $P(d_{obs})$, is the probability of a given value of d_{obs} across all values of d_{pop} . This is the integral from $-\infty$ to ∞ of the numerator in Bayes's formula.

R Implementation

Below is R code to compute the posterior for the examples in this post.

```
## Return function for posterior probability of d.pop given d.obs for two group
## difference-of-mean studies with equal sample size and unit standard deviation
##   n is sample size per group
##   d.obs is standardized observed effect size
##   prior is function of d.pop, eg, function(d.pop) dnorm(d.pop,mean=0.3,sd=0.1)
##
posterior=function(n,d.obs,prior) {
  ## probability of d.obs given d.pop for examples at hand. d_d2t defined below
  P_obsGIVENpop=function(d.pop) d_d2t(n,d.pop,d.obs);
  ## numerator in Bayes formula
  numerator=function(d.pop) prior(d.pop)*P_obsGIVENpop(d.pop);
  ## denominator in Bayes formula
  P_obs=integrate(function(d.pop) numerator(d.pop),-Inf,Inf)$value;
  ## final answer
  P_popGIVENobs=function(d.pop) numerator(d.pop)/P_obs;
}

## probability density of noncentral t in terms of n, d.pop, d.obs
##   adapted from https://natgoodman.github.io/repwr/stats.stable.html
d_d2t=function(n,d.pop,d.obs) {
```

```

df=2*(n-1);
t=d.pop*sqrt(n/2);
sqrt(n/2)*suppressWarnings(dt(t,df=df,ncp=sqrt(n/2)*d.obs));
}

## examples with normal prior as in Figures 2 and 3 of the post
## n=10
post=posterior(n=10,d.obs=0.5,prior=function(d.pop) dnorm(d.pop,mean=0.3,sd=0.1));
curve(post,from=-0.2,to=0.8,col='red',lwd=2,xlab='d',ylab='probability density');
curve(dnorm(x,mean=0.3,sd=0.1),col='blue',add=T);
grid();
## n=200;
post=posterior(n=200,d.obs=0.5,prior=function(d.pop) dnorm(d.pop,mean=0.3,sd=0.1));
curve(post,from=-0.2,to=0.8,col='red',lwd=2,xlab='d',ylab='probability density');
curve(dnorm(x,mean=0.3,sd=0.1),col='blue',add=T);
grid();

```

The code exploits one advanced R feature that may be unfamiliar to some readers, namely, the ability of R functions to create other functions on the fly. Each line of `posterior` generates a function that calls functions generated on previous lines or the `prior` function passed as an argument. The final result is likewise a function.

The examples store the function in the variable named `post`, then use `curve` (a base R function) to call `post` and plot the result. The examples also plot the prior (`dnorm(d.pop,mean=0.3,sd=0.1)`) for reference.

Comments Please!

Please post comments on [Twitter](#) or [Facebook](#), or contact me by email natg@shore.net.