I08 - Comparing probabilities

STAT 5870 (Engineering)
Iowa State University

August 28, 2024

One probability

Consider the model $Y \sim Bin(n, \theta)$.

We have discussed a number of statistical procedures to draw inferences about θ:

- Frequentist: based on (asymptotic) distribution of Y/n
  - p-value for test of $H_0 : \theta = \theta_0$,
  - confidence interval for θ,
- Bayesian: based on posterior for θ
  - credible interval for θ,
  - posterior model probability, e.g. $p(H_0|y)$, and
  - posterior probability statements, e.g. $P(\theta < \theta_0|y)$.

Now, we will consider what happens when we have multiple θs.

Two probabilities

Consider the model

$$Y_g \stackrel{ind}{\sim} Bin(n_g, \theta_g)$$

for $g = 1, 2$ and you are interested in the relationship between θ_1 and θ_2.

- Frequentist: based on asymptotic distribution of $\frac{Y_1}{n_1} - \frac{Y_2}{n_2}$:
    - p-value for a hypothesis test, e.g. $H_0 : \theta_1 = \theta_2$,
    - confidence interval for $\theta_1 - \theta_2$,
- Bayesian: based on posterior distribution of $\theta_1 - \theta_2$:
    - credible interval for θ_1, θ_2,
    - posterior model probability, e.g. $p(H_0|y)$, and
    - probability statements, e.g. $P(\theta_1 < \theta_2|y)$.

where $y = (y_1, y_2)$.

Data example

Suppose you have two manufacturing processes and you are interested in which process has the larger probability of being within the specifications.

So you run the two processes and record the number of successful products produced:

- Process 1: 135 successful products out of 140 attempts
- Process 2: 216 successful products out of 230 attempts

In R, you can code this as two vectors:

```
successes = c(135,216)
attempts  = c(140,230)
```

or, better yet, as a data.frame:

```
d = data.frame(process   = factor(1:2),
               successes = successes,
               attempts  = attempts)
```

p-values and confidence intervals

Because there is no indication that you expect one of the two manufacturing processes to have a higher probability, you should perform a two-sided hypothesis test, i.e.

- $H_0 : \theta_1 = \theta_2$
- $H_A : \theta_1 \neq \theta_2$

and calculate a two-sided confidence interval for $\theta_1 - \theta_2$.

```
prop.test(d$successes, d$attempts)


2-sample test for equality of proportions with continuity correction

data:  d$successes out of d$attempts
X-squared = 0.67305, df = 1, p-value = 0.412
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.02417591  0.07448647
sample estimates:
   prop 1    prop 2 
0.9642857 0.9391304 
```

Bayesian analysis

Assume

$$Y_g \stackrel{ind}{\sim} Bin(n_g, \theta_g)$$

and

$$\theta_g \stackrel{ind}{\sim} Be(1, 1).$$

Then the posterior is

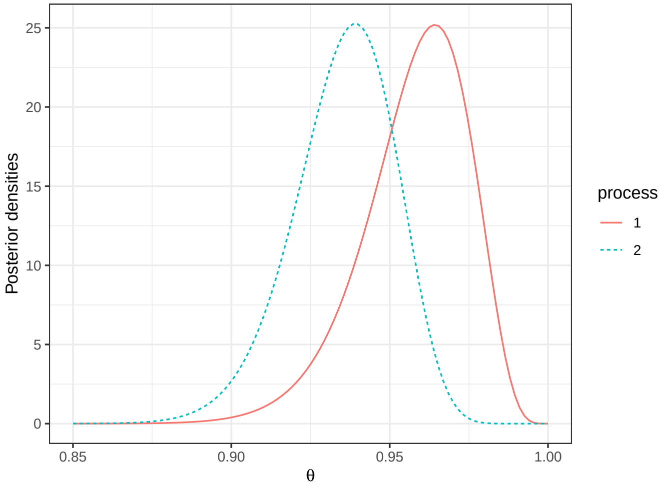
$$\theta_g|y \stackrel{ind}{\sim} Be(1 + y_g, 1 + n_g - y_g).$$

From this we can compute

$$P(\theta_1 < \theta_2|y) = P(\theta_1 - \theta_2 < 0|y)$$

and a credible interval for $\theta_1 - \theta_2$ by simulating values from the posterior and computing $\theta_1 - \theta_2$.

Posteriors


Posterior densities
25
20
15
10
5
0
0.85
0.90
0.95
1.00
θ
process
1
2

Credible interval for the difference

To obtain statistical inference on the difference, we draw samples from the posterior and then calculate the difference:

```
n      <- 1e5
theta1 <- rbeta(n, 1+d$success[1], 1+d$attempts[1] - d$success[1])
theta2 <- rbeta(n, 1+d$success[2], 1+d$attempts[2] - d$success[2])
diff   <- theta1 - theta2

# Bayes estimate for the difference
mean(diff)

[1] 0.02235018

# Estimated 95% equal-tail credible interval
quantile(diff, c(.025,.975))

       2.5%       97.5%
-0.02489203  0.06739588

# Estimate of the probability that theta1 is less than theta2
mean(diff < 0)

[1] 0.16391
```

Multiple probabilities

Now, let's consider the more general problem of

$$Y_g \stackrel{ind}{\sim} Bin(n_g, \theta_g)$$

for $g = 1, 2, \ldots, G$ and you are interested in the relationship amongst the θ_g.

We can perform the following statistical procedures:

- Frequentist: based on distribution of $Y_1, \ldots, Y_G$
  - p-value for test of $H_0 : \theta_g = \theta$ for all g,
  - p-value for test of $H_0 : \theta_g = \theta_{g'}$,
  - confidence interval for $\theta_g - \theta_{g'}$,
- Bayesian: based on posterior for $\theta_1, \ldots, \theta_G$:
  - credible interval for $\theta_g - \theta_{g'}$,
  - posterior model probability, e.g. $p(H_0|y)$, and
  - probability statements, e.g. $P(\theta_g < \theta_{g'}|y)$.

where g and $'g$ represent different values.

Data example

Suppose you have three manufacturing processes and you are interested in which process has the larger probability of being within the specifications.

So you run the three processes and record the number of successful products produced:

- Process 1: 135 successful products out of 140 attempts
- Process 2: 216 successful products out of 230 attempts
- Process 3: 10 successful products out of 10 attempts

In R, you can code this as two vectors:

```
successes = c(135,216,10)
attempts  = c(140,230,10)
```

or, better yet, as a data.frame:

```
d = data.frame(process   = factor(1:3),
               successes = successes,
               attempts  = attempts)
```

p-values

The default hypothesis test is

$$H_0 : \theta_g = \theta \quad \text{for all} \quad g \quad \text{versus} \quad H_A : \theta_g \neq \theta_{g'} \text{ for some } g, g'$$

```
prop.test(d$successes, d$attempts)

Warning in prop.test(d$successes, d$attempts):  Chi-squared approximation may be
incorrect


3-sample test for equality of proportions without continuity correction

data:  d$successes out of d$attempts
X-squared = 1.6999, df = 2, p-value = 0.4274
alternative hypothesis: two.sided
sample estimates:
   prop 1    prop 2    prop 3 
0.9642857 0.9391304 1.0000000 
```

Confidence intervals

Confidence interval for $\theta_1 - \theta_3$:

```
# Need to specify a comparison to get confidence intervals of the difference
prop.test(d$successes[c(1,3)], d$attempts[c(1,3)])$conf.int

Warning in prop.test(d$successes[c(1, 3)], d$attempts[c(1, 3)]):  Chi-squared
approximation may be incorrect

[1] -0.10216886  0.03074029
attr(,"conf.level")
[1] 0.95
```

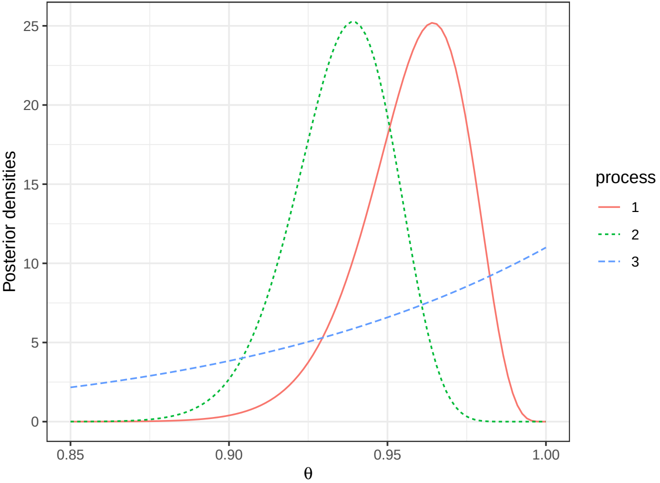
An alternative test

An alternative test for equality amongst the proportions uses `chisq.test()`.

```
d$failures <- d$attempts - d$successes
chisq.test(d[c("successes","failures")])

Warning in chisq.test(d[c("successes", "failures")]):  Chi-squared approximation
may be incorrect

Pearson's Chi-squared test

data:  d[c("successes", "failures")]
X-squared = 1.6999, df = 2, p-value = 0.4274
```

```
chisq.test(d[c("successes","failures")], simulate.p.value = TRUE)

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data:  d[c("successes", "failures")]
X-squared = 1.6999, df = NA, p-value = 0.4158
```

Posteriors



Posterior densities
25
20
15
10
5
0
0.85
0.90
0.95
1.00
θ
process
1
2
3

Credible interval for differences

To compare the probabilities, we draw samples from the posterior and compare them.

```
posterior_samples <- function(d) {
  data.frame(
    rep = 1:1e5,
    name = paste0("theta", d$process),
    theta = rbeta(1e5, 1+d$successes, 1+d$attempts-d$successes),
    stringsAsFactors = FALSE)
}

draws <- d %>% group_by(process) %>% do(posterior_samples(.)) %>% ungroup() %>%
  select(-process) %>% tidyr::spread(name, theta)

# Estimate of the comparison probabilities
draws %>%
  summarize(`P(theta1>theta2|y)` = mean(draws$theta1 > draws$theta2),
            `P(theta1>theta3|y)` = mean(draws$theta1 > draws$theta3),
            `P(theta2>theta3|y)` = mean(draws$theta2 > draws$theta3)) %>%
  gather(comparison, probability)

# A tibble: 3 x 2
  comparison         probability
  <chr>                    <dbl>
1 P(theta1>theta2|y)       0.840
2 P(theta1>theta3|y)       0.632
3 P(theta2>theta3|y)       0.486
```

Summary

Multiple (independent) binomial proportions

- p-values
- confidence intervals
- posterior densities
- credible intervals
- posterior probabilities