Estimating a relative change using a log-transformation of the outcome

Brice Ozenne

June 15, 2020

1 Result

Let's denote by Y the outcome and by G a group variable G (binary variable). We are interested in the relative change in Y between the groups. We decide to model the group effect on the log scale:

$$\log(Y) = Z = \alpha + \beta G + \varepsilon \text{ where } \varepsilon \sim \mathcal{N}\left(0, \sigma^2\right)$$

We claim that:

$$\frac{\mathbb{E}\left[Y|G=1\right] - \mathbb{E}\left[Y|G=0\right]}{\mathbb{E}\left[Y|G=0\right]} = e^{\beta} - 1$$

2 Proof

2.1 Re-writting the model as a multiplicative model

We can re-write the model as:

$$Y = e^{\alpha+\beta G} e^{\varepsilon} \text{ where } \varepsilon \sim \mathcal{N}\left(0, \sigma^2\right)$$

So for $g \in \{1,2\}$:

$$\mathbb{E}\left[Y|G=g\right] = e^{\alpha+\beta g}\mathbb{E}\left[e^{\varepsilon}\right]$$

Then:

$$\begin{aligned}
\frac{\mathbb{E}\left[Y|G=1\right] - \mathbb{E}\left[Y|G=0\right]}{\mathbb{E}\left[Y|G=0\right]} &= \frac{e^{\alpha+\beta}\mathbb{E}\left[e^{\varepsilon}\right] - e^{\alpha}\mathbb{E}\left[e^{\varepsilon}\right]}{e^{\alpha}\mathbb{E}\left[e^{\varepsilon}\right]} \\
&= \frac{e^{\alpha+\beta} - e^{\alpha}}{e^{\alpha}} = e^{\beta} - 1
\end{aligned}$$

2.2 Using a Taylor expansion

Using a second order Taylor expansion of $\exp(Z)$ around $\mu(G) = \alpha + \beta G$ and assuming that the first moments of Z are finite and the remaining moments are neglectable regarding the factorial of the moment order (i.e. $\forall i \geq 1$, $\frac{1}{i!}\mathbb{E}\left[\varepsilon^i\right] < +\infty$ and $\sum_{i=1}^{\infty} \frac{1}{i!}\mathbb{E}\left[\varepsilon^i\right] < +\infty$), we get:

$$
\begin{aligned}
Y = e^Z &= e^\mu + \sum_{i=1}^{\infty} \frac{1}{i!}(Z-\mu)^i \frac{\partial^i e^\mu}{(\partial \mu)^i} \\
&= e^{\alpha+\beta G} + \sum_{i=1}^{\infty} \frac{1}{i!}(Z-\alpha-\beta G)^i e^{\alpha+\beta G} \\
\mathbb{E}\left[Y|G=g\right] &= e^{\alpha+\beta G} + \sum_{i=1}^{\infty} \frac{1}{i!}\mathbb{E}\left[(Z-\alpha-\beta g)^i\right] e^{\alpha+\beta G} \\
&= e^{\alpha+\beta G}\left(1 + \sum_{i=1}^{\infty} \frac{1}{i!}\mathbb{E}\left[\varepsilon^i\right]\right)
\end{aligned}
$$

where we used that the distribution of ε is independent of g. [Optional] ε follows a zero-mean normal distribution, so the uneven moments are 0:

$$
\mathbb{E}\left[Y|G=g\right] = e^{\alpha+\beta G}\left(1 + \sum_{i=1}^{\infty} \frac{1}{2i!}\mathbb{E}\left[\varepsilon^{2i}\right]\right)
$$

We can now express our parameter of interest:

$$
\begin{aligned}
\Delta_G &= \frac{\mathbb{E}\left[Y|G=1\right] - \mathbb{E}\left[Y|G=0\right]}{\mathbb{E}\left[Y|G=0\right]} = \frac{\mathbb{E}\left[Y|G=1\right]}{\mathbb{E}\left[Y|G=0\right]} - 1 \\
&= \frac{e^{\alpha+\beta}\left(1 + \sum_{i=1}^{\infty} \frac{1}{2i!}\mathbb{E}\left[\varepsilon^{2i}\right]\right)}{e^{\alpha}\left(1 + \sum_{i=1}^{\infty} \frac{1}{2i!}\mathbb{E}\left[\varepsilon^{2i}\right]\right)} - 1 \\
&= e^\beta - 1
\end{aligned}
$$

3 Note for power calculation

3.1 Recall: delta-method for normally distributed variables

Theory: we recall that for a random variable Y with finite first two moments, the delta method applied around the mean for a transformation f is:

$$f(Y) = f(\mu) + f'(\mu)(Y-\mu) + \frac{1}{2}f''(\mu)(Y-\mu)^2 + \frac{1}{6}f'''(\mu)(Y-\mu)^3 + o\left((Y-\mu)^2\right)$$

where $\mu = \mathbb{E}[Y]$. Introducing $\sigma^2 = \mathbb{V}ar[Y]$, we have:

$$\begin{aligned}\mathbb{E}[f(Y)] =& f(\mu) + f'(\mu)(\mathbb{E}[Y]-\mu) + \frac{1}{2}f''(\mu)\mathbb{E}\left[(Y-\mu)^2\right] + \frac{1}{6}f'''(\mu)\mathbb{E}\left[(Y-\mu)^3\right] + o\left(\mathbb{E}\left[(Y-\mu)^3\right]\right)\\ =& f(\mu) + \frac{\sigma^2}{2}f''(\mu) + o\left(\mathbb{E}\left[(Y-\mu)^3\right]\right)\end{aligned}$$

for a normal distribution since $\mathbb{E}[(Y-\mu)^3] = 0$. Also:

$$\begin{aligned}\mathbb{V}ar[f(Y)] =& \left(f'(\mu)\right)^2 \mathbb{V}ar[\mathbb{E}[Y]-\mu] + f'(\mu)f''(\mu)\mathbb{E}\left[(Y-\mu)^3\right]\\ &+ \left(\frac{f'(\mu)f'''(\mu)}{3} + \frac{\left(f''(\mu)\right)^2}{4}\right)\mathbb{E}\left[(Y-\mu)^4\right] + o\left(\mathbb{E}\left[(Y-\mu)^4\right]\right)\\ =& \left(f'(\mu)\right)^2\sigma^2 + 3\sigma^4\left(\frac{f'(\mu)f'''(\mu)}{3} + \frac{\left(f''(\mu)\right)^2}{4}\right) + o\left(\mathbb{E}\left[(Y-\mu)^4\right]\right)\end{aligned}$$

Application: exponential transformation ($f = \exp$)

$$\begin{aligned}\mathbb{E}[\exp(Y)] &\approx \exp(\mu)\left(1 + \frac{\sigma^2}{2}\right)\\ \mathbb{V}ar[\exp(Y)] &\approx \exp(2\mu)\left(\sigma^2 + \frac{7}{4}\sigma^4\right)\end{aligned}$$

Illustration:

```
n <- 1e4
mu <- 0.1
sigma2 <- 0.1
X <- rnorm(n, mean = mu, sd = sqrt(sigma2))
fX <- exp(X)
```

```
## first order
c(error_mean = mean(fX) - exp(mu),
  errorPC_mean = 100*(mean(fX) - exp(mu))/mean(fX))
```

```
c(error_var = var(fX) - exp(2*mu)*sigma2,
  errorPC_var = 100*(var(fX) - exp(2*mu)*sigma2)/var(fX))
```

```
 error_mean errorPC_mean
 0.05783048   4.97252058
 error_var errorPC_var
0.02343271 16.09687872
```

```
## second order
c(mean = mean(fX),
  error_mean = mean(fX) - exp(mu)*(1+sigma2/2),
  errorPC_mean = 100*(mean(fX) - exp(mu)*(1+sigma2/2))/mean(fX))
c(var = var(fX),
  error_var = var(fX) - exp(2*mu)*(sigma2 + (7/4)*sigma2^2),
  errorPC_var = 100*(var(fX) - exp(2*mu)*(sigma2 + (7/4)*sigma2^2))/var
    (fX))
```

```
       mean   error_mean errorPC_mean
 1.163001402  0.002571938  0.221146614
        var   error_var errorPC_var
0.145572982 0.002058158 1.413832496
```

3.2 Two independent groups

Theory: consider two groups $G = 0$ and $G = 1$ for which we want to compare the percentage difference in outcome Y. Our parameter of interest is:

$$\frac{\mathbb{E}\left[Y|G=1\right] - \mathbb{E}\left[Y|G=0\right]}{\mathbb{E}\left[Y|G=0\right]} = \gamma$$

and we assume that on the original scale:

$$\mathbb{V}ar\left[Y\right] = \mathbb{V}ar\left[Y|G=1\right] = \mathbb{V}ar\left[Y|G=0\right] = \sigma_Y^2$$

and

$$\mathbb{E}\left[Y|G=0\right] = \alpha_Y$$

We only assume that the outcome is normally distribution after log transformation, i.e. $\log(Y) \sim \mathcal{N}\left(a_0, s_0^2\right)$ in the first group and $\log(Y) \sim \mathcal{N}\left(a_1, s_1^2\right)$. We can use the delta method to identify these parameters:

$$\begin{aligned}
\alpha_Y &= \exp(a_0)\left(1+\frac{s_0^2}{2}\right)\\
\sigma_Y^2 &= \exp(2a_0)\left(s_0^2+\frac{7}{4}s_0^4\right)\\
\alpha_Y(\gamma+1) &= \exp(a_1)\left(1+\frac{s_1^2}{2}\right)\\
\sigma_Y^2 &= \exp(2a_1)\left(s_1^2+\frac{7}{4}s_1^4\right)
\end{aligned}$$

i.e.

$$\begin{aligned}
\frac{\alpha_Y^2}{\sigma_Y^2} &= \frac{\left(1-\frac{s_0^2}{2}\right)^2}{s_0^2+\frac{7}{4}s_0^4}\\
a_0 &= \frac{1}{2}\log\left(\frac{\sigma_Y^2}{\left(s_0^2+\frac{7}{4}s_0^4\right)}\right)\\
\frac{\alpha_Y^2(\gamma+1)^2}{\sigma_Y^2} &= \frac{\left(1-\frac{s_1^2}{2}\right)^2}{s_1^2+\frac{7}{4}s_1^4}\\
a_1 &= \frac{1}{2}\log\left(\frac{\sigma_Y^2}{\left(s_1^2+\frac{7}{4}s_1^4\right)}\right)
\end{aligned}$$

The first and third equation can be solved numerically.

Illustration:

```
alpha_Y <- 1.15
sigma2_Y <- 0.15

s <- uniroot(function(x){alpha_Y^2/sigma2_Y - (1+x/2)^2/(x+x^2*7/4)},
               interval = c(0,1))$root
a <- log(sigma2_Y/(s+s^2*7/4))/2
c(a = a, s = s)
```

```
         a          s 
0.08802784 0.10608948 
```

We can check that:

```
c(exp(a)*(1+s/2), exp(2*a)*(s+s^2*7/4))
```

```
[1] 1.149944 0.150000
```

i.e.

```
Z <- rnorm(1e4, mean=a, sd = sqrt(s))
mean(exp(Z))
var(exp(Z))
```

```
[1] 1.152237
[1] 0.1496768
```

Note: an alternative approach is to use a log-normal distribution with parameters:

$$s^2 = \log\left(1 + \frac{\sigma^2}{\alpha^2}\right)$$
$$a = \log(\alpha) - \frac{s^2}{2}$$

Here it gives:

```
s <- log(1+sigma2_Y/alpha_Y^2)
a <- log(alpha_Y) - s/2
c(a = a, s = s)
```

```
         a          s
0.08604307 0.10743775
```

We can check that:

```
exp(a + s/2) - alpha_Y
(exp(s)-1)*exp(2*a + s) - sigma2_Y
```

```
[1] 0
[1] -5.551115e-17
```

and

```
Y <- rlnorm(1e4, meanlog=a, sdlog = sqrt(s))
mean(Y)
var(Y)
```

```
[1] 1.146438
[1] 0.1462818
```