Confidence Interval for Lift

Data Science, Adobe Target.

I. CONFIDENCE INTERVAL ON THE LIFT

In addition to the lift and confidence, a useful extra statistic to view is the *confidence interval on the lift*. There are nuances to calculating a confidence interval for lift, because the lift of some variant/experience β, over the baseline experience α is a random variable, defined as

$$\Delta_{\beta\alpha} = \frac{X_\beta - X_\alpha}{X_\alpha} \tag{1}$$

where X_α is the performance of experience α. Importantly, the lift is a random variable, which is itself a ***ratio of two random variables***. We must therefore be careful when calculating its variance estimate.

Before discussing this, let us first review the standard theory of confidence intervals. For a samples drawn from some arbitrary distribution, the sample mean $\bar{X}$ has a $C = 1 - 2\alpha$ Confidence interval given by

$$\left(\bar{x} - t^* \frac{s}{\sqrt{N}},\ \bar{x} + t^* \frac{s}{\sqrt{N}}\right) \tag{2}$$

with $t^* = t_\alpha(\nu)$, where ν is the degrees of freedom, and α is the significance level. Here, s is the sample standard deviation, and N is the number of samples. For a 95% confidence interval, and assuming normality of the sampling distribution of the mean, we would have $t^* = 1.96$. The above formula applies for any random variable - but to apply it, we must correctly calculate estimates for the mean and sample standard deviation.

As we mentioned previously in this section, the lift is a ratio of two random variables. To calculate its mean and sample standard deviation, we must therefore use the expressions derived in Appendix A. In particular, we find:

$$E\left[(\text{lift})\right] = E\left[\Delta_{\alpha\beta}\right] = E\left[\frac{X_\beta}{X_\alpha} - 1\right] = E\left[\frac{X_\beta}{X_\alpha}\right] - 1 = \frac{\overline{X}_\beta - \overline{X}_\alpha}{\overline{X}_\alpha} \tag{3}$$

Meanwhile, the Variance of the lift is given by

$$
\begin{aligned}
\text{Var}\left[(\text{lift})\right] &= \text{Var}\left[\frac{X_\beta}{X_\alpha} - 1\right] \\
&= \text{Var}\left[\frac{X_\beta}{X_\alpha}\right] \\
&= \left(\frac{\overline{X}_\beta}{\overline{X}_\alpha}\right)^2 \left[\frac{Var(X_\beta)}{(\overline{X}_\beta)^2} + \frac{Var(X_\alpha)}{(\overline{X}_\alpha)^2}\right]
\end{aligned} \tag{4}
$$

where the covariance term drops out because X_α and X_β are independent random variables (more simply - no user/unit is in both samples). The standard deviation of the lift is then simply the square root of this variance.

Appendix A: Approximations for the Mean and Variance of Ratios of Random Variables

In this section, we derive approximations for the mean and variance of ratios of two random variables. These expressions are useful when calculating a confidence interval for the lift in an experiment. The approach described here is sometimes called the *Delta method*.

Consider two random variables X and Y where Y cannot take values at 0. We wish to derive the expectation and variance of some smooth function of these random variables, $f(X,Y)$. If the probability distribution is sufficiently peaked near to the mean of each variable, then we can do a Taylor expansion around the individual means μ_x and μ_y. This Taylor expansion takes the form:

$$
f(X,Y) = f(\mu_x,\mu_y) + \frac{\partial f(\mu_x,\mu_y)}{\partial x}(X-\mu_x) + \frac{\partial f(\mu_x,\mu_y)}{\partial Y}(Y-\mu_y) + \ldots \tag{A1}
$$

Taking expectations of both sides, and assuming $E[X] = \mu_x$ and $E[Y] = \mu_y$, while $E[(X - \mu_x)^2] = \text{Var}(X)$ and $E[(Y-\mu_y)^2] = \text{Var}(Y)$ while $E[(X-\mu_x)(Y-\mu_y)] = \text{Cov}(X,Y)$, we get to first order:

$$
E[f(X,Y)] \approx f(\mu_x,\mu_y) \tag{A2}
$$

while to second order, we have:

$$
\begin{aligned}
\mathrm{Var}[f(X,Y)] &= E\left\{[f(X,Y) - E(f(X,Y))]^2\right\} \\
&\approx E\left\{\left[f(\mu_x,\mu_y) + \frac{\partial f(\mu_x,\mu_y)}{\partial X}(X-\mu_x) + \frac{\partial f(\mu_x,\mu_y)}{\partial Y}(Y-\mu_y) - f(\mu_x,\mu_y)\right]^2\right\} \\
&= \left(\frac{\partial f(\mu_x,\mu_y)}{\partial X}\right)^2 E\left[(X-\mu_x)^2\right] + 2\frac{\partial f(\mu_x,\mu_y)}{\partial X}\frac{\partial f(\mu_x,\mu_y)}{\partial Y} E\left[(X-\mu_x)(Y-\mu_y)\right] \\
&\quad + \left(\frac{\partial f(\mu_x,\mu_y)}{\partial Y}\right)^2 E\left[(Y-\mu_y)^2\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \left(\frac{\partial f(\mu_x,\mu_y)}{\partial X}\right)^2 \mathrm{Var}(X) + \left(\frac{\partial f(\mu_x,\mu_y)}{\partial Y}\right)^2 \mathrm{Var}(Y) \\
&\quad + 2\frac{\partial f(\mu_x,\mu_y)}{\partial X}\frac{\partial f(\mu_x,\mu_y)}{\partial Y}\mathrm{Cov}(X,Y)
\end{aligned} \tag{A3}
$$

With this general expression available, we can now focus on the ratio of two random variables, i.e. substituting into the above

$$f(X,Y) = \frac{X}{Y}, \tag{A4}$$

with the corresponding partial derivatives:

$$\Longrightarrow \quad \frac{\partial f}{\partial X} = \frac{1}{Y}, \quad \frac{\partial f}{\partial Y} = -\frac{X}{Y^2}, \tag{A5}$$

we finally have the approximations

$$E\left[\frac{X}{Y}\right] \approx \frac{\mu_x}{\mu_y} \tag{A6}$$

while the variance is

$$
\begin{aligned}
\mathrm{Var}\left[\frac{X}{Y}\right] &\approx \frac{1}{\mu_y^2}\mathrm{Var}(X) - 2\frac{\mu_x}{\mu_y^3}\mathrm{Cov}(X,Y) + \frac{\mu_x^2}{\mu_y^4}\mathrm{Var}(Y) \\
&= \frac{\mu_x^2}{\mu_y^2}\left[\frac{\mathrm{Var}(X)}{\mu_x^2} - 2\frac{\mathrm{Cov}(X,Y)}{\mu_x\mu_y} + \frac{\mathrm{Var}(Y)}{\mu_y^2}\right]
\end{aligned} \tag{A7}
$$