Computing the bias of a coin - Bayesian style

Dorian Goldman

April 25, 2017

Let's imagine that we are flipping a coin, but we don't know how biased it is. Let's call the bias p (ie. if it's fair, then $p = 1/2$). How do we determine, through experiment, what p is? If we are able to flip the coin as many times as we want, then we can use the outcomes of this experiment to determine this. In other words, Bayes theorem allows us to quantify the hypothesis of induction! The major assumption made though, is that there exists a real distribution.
Let's define X_N to be the number of heads obtained after N flips of a coin which has bias p. Then X_N has a distribution given by

$$f(X_N = k|p) = \binom{N}{k} p^k (1-p)^{N-k}.$$

But we really want to know what is $f(p|X_n = k)$, ie. we want to learn the bias from the data. Thus we use Baye's theorem:

$$f(p|X_N = k) = \frac{f(X_N = k|p)f(p)}{f(X_n = k)}.$$

What is $f(X_n = k)$? This is summing up all possibilities over all probabilities. In other words, we have

$$f(X_n = k) = \int_0^1 \binom{N}{k} p^k (1-p)^{N-k} f(p) dp.$$

Thus we have

$$f(p|X_N = k) = \frac{\binom{N}{k} p^k (1-p)^{N-k} f(p)}{\int_0^1 \binom{N}{k} p^k (1-p)^{N-k} f(p) dp}.$$

What is $f(p)$? This is just our prior belief about what p is - since we don't know anything, we set it to be the *uniform distribution*, so $f(p) \equiv 1$. Thus we finally have

$$f(p|X_N = k) = \frac{p^k(1-p)^{N-k}}{\int_0^1 p^k(1-p)^{N-k}dp}.$$

For simplicity in the next sections, are going to write $N = n + m$ so that we have

$$f(p|X_n = m) = \frac{p^m(1-p)^n}{\int_0^1 p^m(1-p)^n dp}.$$

Also recall that the Beta function is defined as

$$B(n+1, m+1) = \int_0^1 p^n(1-p)^m dp.$$

0.1 Case 2: $m = n$ as $n \to +\infty$

$$\mathbb{E}(p|X_n = n) = \frac{B(n+2, n+1)}{B(n+1, n+1)} = \frac{n+1}{2n+3}\frac{B(n+1, n+1)}{B(n+1, n+1)} = \frac{n+1}{2n+3}.$$

$$\mathrm{Var}(p|X_n = n) = \frac{1}{B(n+1, n+1)}\int_0^1 (p - \frac{n+1}{2n+3})^2 p^n(1-p)^n dp \tag{0.1}$$

$$\frac{B(n+3, n+1) - B(n+2, n+1)(n+1)/(2n+3) + (n+1)^2/(2n+3)^2 B(n+1, n+1)}{B(n+1, n+1)}.$$

Using the identity $B(m+1, n) = \frac{m}{m+n}B(m, n)$ repeatedly and using the fact that $n \to +\infty$, we have

$$\mathrm{Var}(p|X_n = n) = \frac{1}{2}\left(1 - \frac{1}{2}\right) + o(n) \text{ as } n \to +\infty. \tag{0.2}$$

$$\mathbb{E}(p|X_n = n) = \frac{1}{2} + o(n) \text{ as } n \to +\infty. \tag{0.3}$$

Next, we need to check if $\{Z_n = p|X_n = n\}_{n\in\mathbb{N}}$ form an independent set of random variables. Indeed, interpreting $\{p|X_n = n\}$ as the bias on a coin for which we have observed n heads out of $2n$ trials, we have

$$\mathbb{E}(Z_m Z_n) = \frac{1}{B(m+1, m+1)B(n+1, n+1)}\int_0^1\int_0^1 p_1 p_2 p_1^m(1-p_1)^m p_2^n(1-p_2)^n dp_1 dp_2 \tag{0.4}$$

$$= \mathbb{E}(Z_m)\mathbb{E}(Z_n) \tag{0.5}$$

ala Fubini's theorem.

Thus $\{Z_n\}_n$ form an indepdent set of random variables with means converging to 1/2 and variances to $\frac{1}{2}(1-\frac{1}{2})$.
Lyapunov CLT: The theorem is named after Russian mathematician Aleksandr Lyapunov. In this variant of the central limit theorem the random variables Xi have to be independent, but not necessarily identically distributed. The theorem also requires that random variables |?Xi?| have moments of some order , and that the rate of growth of these moments is limited by the Lyapunov condition given below.

Theorem 1. *Lyapunov CLT.[6] Suppose* $\{X1, X2, \cdots\}$ *is a sequence of independent random variables, each with finite expected value* μ_i *and* σ_i^2 *i. Define*

$$s_n^2 = \sum_{i=1}^n \sigma_i^2 s_n^2 = \sum_{i=1}^n \sigma_i^2$$

If for some $\delta > 0$, *Lyapunov?s condition*

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n \mathrm{E}\left[|X_i - \mu_i|^{2+\delta}\right] = 0 \lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n \mathrm{E}\left[|X_i - \mu_i|^{2+\delta}\right] = 0$$

is satisfied, then a sum of
converges in distribution to a standard normal random variable, as n goes to infinity:

$$\frac{1}{s_n} \sum_{i=1}^n (X_i - \mu_i) \xrightarrow{d} N(0,1). \frac{1}{s_n} \sum_{i=1}^n (X_i - \mu_i) \xrightarrow{d} N(0,1).$$

Now define $X_i = (p|X_n = n)$ for $i = 1, \cdots, n$ (ie. they're all the same). Then we conclude that

$$\frac{1}{n}\sum_{i=1}^n (p|X_n = n) = p(|X_n = n) \xrightarrow{d} N(1/2, 1/4).$$