Bayes' Theorem and Its Application to Estimating the Bias of a Coin

Carlos Eduardo Gonçalves de Oliveira
LinkedIn: linkedin.com/in/cego669
GitHub: github.com/cego669

Bayes' Theorem

Here, we explore Bayes' theorem by applying it to the problem of estimating the probability density function of the bias of a coin, θ, given a sequence of observed flips. In this experiment, we perform n flips of a coin, k of which result in "heads."

Thus, Bayes' theorem becomes:

$$f_{\Theta|n,k}(\theta|n,k) = \frac{f_\Theta(\theta) P_{n,k|\theta}(n,k|\theta)}{P_{n,k}(n,k)}$$

Note that $f_\Theta(\theta)$ and $P_{n,k}(n,k)$ represent our prior knowledge (*a priori*) about θ and the number of flips.

But what do we know about θ?

We know **nothing** about θ before flipping the coin. Therefore, it is reasonable to assume that all values of θ between 0 and 1 are equally likely. In this sense, we have:

$$f_\Theta(\theta) = \begin{cases} 1, & \text{if } 0 \leq \theta \leq 1, \\ 0, & \text{if } \theta < 0 \text{ or } \theta > 1. \end{cases}$$

Okay, but what about $P_{n,k}(n,k)$?

What is our prior knowledge about the probability function for n and k? It may feel strange to think about this, but the truth is we do not need it. We can apply the law of total probability, which states:

$$P(A) = \sum_n P(A \mid B_n) P(B_n)$$

Since $f_\Theta(\theta)$ is a continuous function, we work with an integral instead of a summation:

$$P_{n,k}(n,k) = \int_0^1 P_{\Theta|n,k}(\theta|n,k) f_\Theta(\theta)\, d\theta$$

Note that $P_{n,k|\theta}(n,k|\theta)$ is simply the probability that the sequence of n flips of the coin with bias θ results in k heads: a binomial distribution. Thus:

$$P_{n,k|\theta}(n,k|\theta) = \binom{n}{k} \theta^k (1-\theta)^{n-k}$$

Substituting the functions, we have to solve this integral:

$$P_{n,k}(n,k) = \int_0^1 \binom{n}{k} \theta^k (1-\theta)^{n-k} \cdot 1 \, d\theta$$

Since $\binom{n}{k}$ is constant with respect to θ, it can be factored out:

$$P_{n,k}(n,k) = \binom{n}{k} \int_0^1 \theta^k (1-\theta)^{n-k} \, d\theta$$

The integral is the normalized form of the beta function, given by:

$$\int_0^1 \theta^k (1-\theta)^{n-k} \, d\theta = \frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)}$$

where $\Gamma(x)$ is the gamma function, which generalizes the factorial for real numbers, with the property $\Gamma(x+1) = x!$ for integers x.

Substituting the integral:

$$P_{n,k}(n,k) = \binom{n}{k} \frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)}$$

Using the relationship between the gamma function and factorials ($\Gamma(k+1) = k!$):

$$P_{n,k}(n,k) = \binom{n}{k} \frac{k!(n-k)!}{(n+1)!}$$

Knowing that $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, we simplify:

$$P_{n,k}(n,k) = \frac{n!}{k!(n-k)!} \cdot \frac{k!(n-k)!}{(n+1)!}$$

$$P_{n,k}(n,k) = \frac{n!}{(n+1)!} = \frac{1}{n+1}$$

Thus:

$$P_{n,k}(n,k) = \frac{1}{n+1}$$

Substituting into Bayes' theorem

With all terms determined, we substitute them into the initial equation:

$$f_{\Theta|n,k}(\theta|n,k) = \frac{1 \cdot \binom{n}{k} \theta^k (1-\theta)^{n-k}}{\frac{1}{n+1}}$$

Simplifying:

$$f_{\Theta|n,k}(\theta|n,k) = (n+1)\binom{n}{k} \theta^k (1-\theta)^{n-k}.$$

Rewriting $\binom{n}{k}$:

$$f_{\Theta|n,k}(\theta|n,k) = (n+1)\frac{n!}{k!(n-k)!} \theta^k (1-\theta)^{n-k}.$$

Interpreting the result

The posterior probability density function $f_{\Theta|n,k}(\theta|n,k)$ gives the probability of the coin's bias being θ, given that we observed k heads in n flips. This result corresponds to a beta distribution, defined as:

$$\mathrm{Beta}(\theta;\alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1},\quad \text{for } 0\leq\theta\leq 1.$$

Here:

$$\alpha=k+1 \quad \text{and} \quad \beta=n-k+1.$$

Thus, $f_{\Theta|n,k}(\theta|n,k)$ follows a beta distribution with parameters $\alpha=k+1$ and $\beta=n-k+1$:

$$f_{\Theta|n,k}(\theta|n,k)=\mathrm{Beta}(\theta;k+1,n-k+1).$$

Conclusion

By applying Bayes' theorem, we demonstrated that the posterior probability density of the coin's bias θ, given k heads in n flips, follows a beta distribution. This result is fundamental and highly interesting in Bayesian inference, as it allows us to update our beliefs about θ based on observed evidence.