Probability and Statistical Inference

Tianqi Zhang
Emory University

Apr 17th 2025

1 Randomness

1.1 Randomness: A Model of Empirical Observations

Definition 1.1.1 (Latent Space Ω).
We denote the latent space Ω to be the set of all possible outcomes.
Random: A model of an empirically observed property of the world.

Definition 1.1.2 (Random Variable). The random variable X maps each $\omega \in \Omega$ to $\mathbb{R}^d$

Example 1.1.3 Coin toss
The random variable $X : \{\text{Head, Tail}\} \to \mathbb{R}$ is defined to be:

$$X(\text{Head}) = 1, \quad X(\text{Tail}) = 0$$

If we repeat the same experiment under the same conditions, an event $A \subseteq \Omega$ will occur in some experiments but not in others.

1.2 Sampling Frequency

If we conduct n experiments, event A occurs exactly n_A times. Then the sampling frequency of A is:

$$f_n(A) = \frac{n_A}{n}.$$

Example 1.2.1 Coin toss

{3 heads, 97 tails} implies in 100 random experiments with a fair coin:

$$f_{100}(3 \text{ heads}, 97 \text{ tails}) = \frac{3}{100}.$$

1.3 Two Philosophies of Randomness

What happens as $n \to \infty$?

1.3.1 Frequentist Inference

- Stability at large scale: volatility of fluctuations of $f_n(A)$ tends to decrease, where $n \to \infty$.

Theorem 1.3.1 Frequentists' idea

Existence of population probability $P(A) \in [0, 1]$, that is not random, where

$$P(A) = \lim_{n\to\infty} f_n(A)$$

.

However, this limit cannot be deterministic. For example, bad events like:

$$B_n = \{|f_n(A) - P(A)| \geq \varepsilon\}$$

may still occur for large n, which paves the way for the laws of large numbers (LLN) to control B_n.

Remark 1.1 *For all $\varepsilon > 0$, $\Pr(B_n$ happens$) \to 0$ as $n \to \infty$, where:*

$$B_{n,\varepsilon} = \{|f_n(A) - P(A)| \geq \varepsilon\}.$$

1.3.2 Probability Theory vs. Statistical Inference

Remark 1.2 *Probability Theory: For given $P(A)$, compute the probability that a future series of n events lies in an interval:*

$$f_n(A) \in [P(A) - \varepsilon, P(A) + \varepsilon].$$

Remark 1.3 *Statistical Inference: For given statistical evidence,*

1. *A point estimate for $P(A)$:*
$$\lim_{n\to\infty} f_n(A) = P(A)$$

2. *A **confidence interval estimate** for $P(A)$:*

$$CI_n = [f_n(A) - Z_n, f_n(A) + Z_n]$$

Such that $\lim_{n\to\infty} Pr\{P(A) \subset CI_n\} \geq 1-\alpha$ *(commonly* $\alpha = 0.05$*)*

So the two remarks are the inverse problems of each other.

Definition 1.3.2 (Estimator). An **estimator** is a computational rule with given data (a function, or later defined as a random variable of data).

Independent experiments help reduce fluctuations by leveraging concentration of measure results.

1.3.3 Bayesian Estimation

Theorem 1.3.3 Bayesian Idea
The unknown $p = P(A)$ is itself random, and Statistical data reduces uncertainty of the parameter (information update).

- **Prior:** What we believe about $P(A)$ before gathering data.
- **Posterior:** Updated beliefs after gathering data.

The Bernstein-von Mises Theorem might offer some insights reconciling the two schools of thoughts when n is sufficiently large.

Theorem 1.3.4 Bernstein-von Mises Theorem
The posterior distribution is independent of the prior distribution (under some conditions) once the amount of information supplied by a sample of data is large enough