Lecture Notes on Statistical Processes in Physics

Zihang Wang

September 2023

Bernoulli process

A single trial in a Bernoulli process is called a Bernoulli trial, which has two outcomes: success, denoted as 1, with probability p, and failure, denoted as 0, with probability $1-p$. Tossing a coin is a simple example of a Bernoulli process since each toss of the coin can result in one of two possible outcomes: "Heads" ($X = 1$) or "Tails" ($X = 0$) with a probability mass function of p and $1 - p$,

$$P_\delta(p, X) = \begin{cases} p, & X = 1 \\ 1-p, & X = 0 \end{cases}. \tag{1}$$

In quantum mechanics, the Stern-Gerlach experiment serves as a classical illustration of a Bernoulli process. An electron passes through an inhomogeneous magnetic field, yielding two distinct outcomes: the particles are deflected either "up" or "down", based on their intrinsic angular momentum, known as spin. In particular, the initial spin quantum state of a single electron is given by,

$$|\Psi\rangle_e = \sqrt{p}\,|\uparrow\rangle + e^{i\phi}\sqrt{1-p}\,|\downarrow\rangle\,, \tag{2}$$

and ϕ is a general phase factor. The probability is then given by the overlap, e.g. $|\langle\uparrow|\Psi\rangle_e|^2 = p$ and $|\langle\downarrow|\Psi\rangle_e|^2 = 1-p$. We emphasize that the trials in a Bernoulli process are independent.

Binomial Process

A binomial process consists of a sequence of N Bernoulli trials, and the probability mass function for having M successes is determined by the binomial distribution,

$$P_B(p, M) = \binom{N}{M} p^M (1-p)^{N-M}. \tag{3}$$

Instead of having a single electron in the Stern-Gerlach experiment, a beam (or a pack) of electrons (with fixed number N) is utilized, and the outcome intensity (snapshot) is characterized by the binomial distribution. The many-body quantum state can be written as,

$$|\Psi\rangle_{\text{e,beam}} = \bigotimes_{j}^{N} \left(\sqrt{p}\,|\uparrow_j\rangle + e^{i\phi_j}\sqrt{1-p}\,|\downarrow_j\rangle \right), \tag{4}$$

and the probability of observing a particular N-body quantum state, e.g.

$|\Psi\rangle_0 = |\uparrow\uparrow\downarrow\uparrow\ldots\downarrow\rangle$, with M spin up electrons is given by the binomial distribution. The key difference is that a Bernoulli process involves independent trials with two outcomes, while a binomial process counts successes in a fixed number of these trials.

Poisson Process

The Poisson process assumes that events are rare, independent, and isolated*. The probability of M events occurring in a time interval T is described by the probability mass function $P(M,T) = (\gamma T)^M/M!e^{-\gamma T}$, where γ represents a general rate function. The Poisson distribution can be derived from the binomial distribution in the limit $dt \to 0$,

$$P(M,T) = \lim_{dt\to 0} \frac{N!}{M!(N-M)!}(\gamma dt)^M(1-\gamma dt)^{N-M} = \frac{(\gamma T)^M}{M!}e^{-\gamma T}, \tag{5}$$

where $N = T/dt$, and the limits are,

$$\lim_{dt\to 0} \frac{N!}{M!(N-M)!}(\gamma dt)^M = \frac{(\gamma N dt)^M}{M!} = \frac{(\gamma T)^M}{M!}, \quad \lim_{dt\to 0}(1-\gamma dt)^{N-M} = e^{-\gamma T}. \tag{6}$$

In the second equality, we use the series expansion when $N \to \infty$,

$$\lim_{N\to\infty} \frac{N!}{M!(N-M)!} \approx N^M \left(1 - \frac{M(M-1)}{N} + \mathcal{O}(\frac{1}{N^2})\right), \tag{7}$$

and only kept the zero-th order.

The Poisson distribution is particularly useful in various applications, including astronomy, quantum optics, and telecommunications, where the detection of individual photons/particles is essential, and their arrival can be modeled as a Poisson process. For example, Poisson process is used to study atomic decay, photon detection, and quantum tunneling. Let's recall the Stern-Gerlach experiment of a beam of electrons. When $p \to 0$, the binomial distribution becomes the Poisson distribution. We should note that the number of total events goes to infinity.

Let's consider an example, assume stars randomly distributed around us with density n, what is probability that the nearest star is at distance R ?

Let's first consider the total number of stars within a differential volume,

$$N = ndV = 4\pi R^2 ndR. \tag{8}$$

The probability of have no stars within $r < R$ is given by a Poisson process,

$$p(r < R, 0 \text{ star}|n) \sim \frac{(ndV)^0}{0!} e^{-n\frac{4\pi}{3}R^3}, \tag{9}$$

and it is worth noting that it only tells a proportionality without normalization. Similarly, the probability of having one star within the radius is proportional to the following,

$$p(r \approx R, 1 \text{ star}|n) \sim \frac{(ndV)^1}{1!} e^{-dV} = \frac{(n4\pi R^2 dR)^1}{1!} e^{-n4\pi R^2 dR}. \tag{10}$$

Therefore, the joint probability has the product form,

$$\begin{aligned} &p(r < R, 0 \text{ star}, r \approx R, 1 \text{ star}|n) = p(r < R, 0 \text{ star}|n) p(r \approx R, 1 \text{ star}|n) \\ &\sim n4\pi R^2 dR e^{-n4\pi R^2 dR} e^{-n\frac{4\pi}{3}R^3} \approx n4\pi R^2 dR e^{-n\frac{4\pi}{3}R^3}. \end{aligned} \tag{11}$$

Frequentist Inference

The frequentist uses the concept of frequency or repeated sampling. It focuses on estimating statistical parameters, e.g. means and variances and making statistical inferences based on the given data. In frequentist statistics, parameters governing the underlying distribution are treated as fixed, unknown values, and the goal is to estimate these parameters using point estimates (e.g., maximum likelihood) or confidence intervals. For example, in the Stern-Gerlach experiment, the underlying distribution is *assumed* to be the binomial distribution and the controlling parameter is $p = p_0$. The frequentist approach does not incorporate prior beliefs on the parameter p (it is given and fixed), and it relies solely on the data at hand.

Hypothesis testing is a key component of frequentist statistics: we start with two competing hypotheses, null and alternative hypothesis. The former is a declaration of no difference, whereas the later stands in direct opposition to the null, suggesting a significant difference as an alternative. As an example, in the Stern-Gerlach experiment, the frequentist would first propose a binomial parameter $p_0 = 0.5$ with a null hypothesis: "the probability of electron with spin up configuration is equal to 0.5". Using statistical tests and analysis, such as confidence intervals, the frequentist approach would then assess whether the observed outcomes significantly deviate from the null hypothesis's expected probability of 0.5. The alternative hypothesis usually rejects the null, e.g. "the probability of electron with spin up configuration is *not* equal to 0.5".

Bayesian Inference

The Bayesian approach is based on Bayes' theorem,

$$P(\text{Parameter}|\text{Data}) = \frac{P(\text{Data}|\text{Parameter})P(\text{Parameter})}{P(\text{Data})}, \tag{12}$$

where "Data" represents the observed data, and "Parameter" represents the variables that govern the underlying probability distribution function from which the observed data is drawn.

$P(\text{Parameter}|\text{Data})$ is referred to as the **posterior** probability, representing the conditional probability of governing parameters given a set of observed data points. $P(\text{Data}|\text{Parameter})$ is the **likelihood** function that captures the probability of obtaining the observed data with the specified set of controlling parameters. $P(\text{Parameter})$ is referred to as a **prior** distribution, which contains the assumptions regarding the probability distribution of the controlling parameters. $P(\text{Data})$ is known as the **evidence**, representing the overall probability of observing a particular dataset. Since we are mainly interested in the relative probability of various controlling parameters. Therefore, we absorb it into the normalization coefficient.

Bayesian inference treats both observed data and parameters as random variables with probability distributions, in contrast to the frequentist approach. Bayesian inference typically begins with prior beliefs or knowledge about the controlling parameters. For instance, in the Stern-Gerlach experiment, one might assume a prior probability distribution for the parameter p, such as a uniform distribution. The likelihood function quantifies the probability of observing the data with a specific set of controlling parameters and often relies on certain assumptions. For example, it might be described by a binomial distribution with the controlling parameter p in the Stern-Gerlach experiment. The posterior probability represents the updated beliefs about controlling parameters after integrating the observed data. It can then be employed as the new prior for subsequent observations, enabling the continuous refinement of beliefs in the presence of additional data.

Bayesian inference incorporates subjective prior beliefs and provides posterior probability distributions of the controlling parameters, while frequentist inference focuses on objective measures based solely on observed data. The choice between the two approaches often depends on the specific problem and available data.