1 The frequentist interpretation of probability

1.1 Probability of event A as it relates to relative frequency

The probability of an event A, say, $\mathbb{P}(A)=p$, is directly related to the **relative frequency** of the occurrence of event A, as defined in the context of $(S, \Im, \mathbb{P}(.))$.

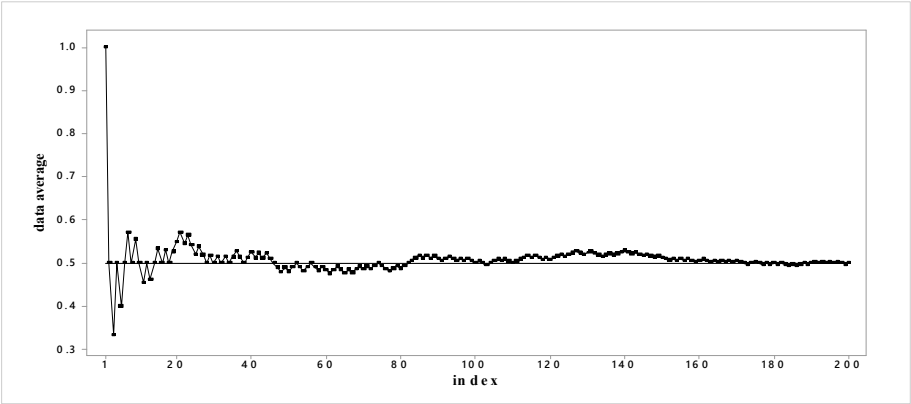
The traditional frequentist interpretation is articulated in terms of the **"long-run" metaphor** which begins with the event of interest, say A, and places this event in the context of a **simple Bernoulli model**:

$$\mathcal{M}_\theta(\mathbf{z})\text{: } Z_k \backsim \mathsf{BerIID}(\theta, \theta(1-\theta)),\ z_k=0,1,\ \theta\in[0,1],\ k\in\mathbb{N}:=(1,2,\ldots),$$

by associating event A with the random variable X via $A=(s\text{: } Z(s)=1)$ and $\overline{A}=(s\text{: } Z(s)=0)$ where the relevant event space is $\Im=\{S, \varnothing, A, \overline{A}\}$. Under the IID assumptions one can invoke the SLLN to claim that in a sequence of n trials the relative frequency $\overline{z}_n=\frac{1}{n}\sum_{i=1}^n z_i$ of occurrence of event A will approximate (oscillate around) the value of its true probability $p=\mathbb{P}(A)$. NOTE that the data $\mathbf{z}_0$ come in the form of a sequence zeros and ones, i.e. $\mathbf{z}_0=(1,0,0,1,1,1,0,....,0,1)$, and thus:

$$\tfrac{1}{n}\textstyle\sum_{i=1}^n z_i=\frac{m}{n} \text{ represents the relative frequency of event } A \text{ occurring,}$$

since it is equal to the number of ones in n data points in $\mathbf{z}_0$.



$\overline{x}_n$ for Bernoulli IID data with n=200

This metaphor, however, carries the seeds of a potential confusion between the stochastic process $\{X_k,\ k\in\mathbb{N}\}$ itself and one its finite realizations $\{x_k\}_{k=1}^n$.

How is such an interpretation formally justified? The simple answer is that it is justified by invoking the SLLN. Under certain probabilistic assumptions (restrictions) on the stochastic process $\{X_k,\ k\in\mathbb{N}\}$, the most restrictive being that it is IID, one can prove *mathematically* that:

$$\mathbb{P}(\lim_{n\to\infty}(\tfrac{1}{n}\textstyle\sum_{k=1}^n X_k)=p)=1 \tag{1}$$

known as **convergence almost surely** (a.s). The SLLN asserts that for any IID process $\{X_k,\ k\in\mathbb{N}\}$ and any event $A\in\Im$, the relative frequency of occurrence of A converges to $\mathbb{P}(A)$ **with probability one** as $n\to\infty$.

First, the result in (1) does *not* involve any claims that the **sequence of numbers** $\{\overline{x}_n\}_{n=1}^{\infty}$, where $\overline{x}_n=\frac{1}{n}\sum_{k=1}^{n}x_k$, converges to p in a purely mathematical sense; the convergence is probabilistic.

Second, the result in (1) holds **irrespective of the particular realization** $\{x_k\}_{k=1}^{n}=(x_1,x_2,...,x_n)$ of the process $\{X_k,\ k\in\mathbb{N}\}$. Indeed, it holds for any realization of $\{X_k,\ k\in\mathbb{N}\}$, as long as the latter satisfies certain **probabilistic assumptions,** such as IID.

Third, the result in (1) refers only to **what happens at the limit** $n=\infty$ and says nothing about the behavior of $\frac{1}{n}\sum_{k=1}^{n}X_k$ for a given n. When one refers to the relative frequency $\frac{1}{n}\sum_{k=1}^{n}x_k$ fluctuating around $\mathbb{P}(A)$ as n changes, one is invoking common sense intuition to conceptualize the convergence associated with the SLLN at the level of the observed frequency.

Fourth, the result in (1) provides no information pertaining to the accuracy of $(\frac{1}{n}\sum_{k=1}^{n}x_k)$ as an approximation of $\mathbb{P}(A)$ for a given n; an upper bound for its accuracy is provided by a different limit theorem known as the Law of the Iterated Logarithm (LIL).

The proposed frequentist interpretation has **five key features**:

(a) it revolves around the notion of a statistical model $\mathcal{M}_\theta(\mathbf{x})$, broadly viewed to accommodate non-random samples,

(b) it is firmly anchored on the SLLN,

(c) it is justified on empirical, not a priori, grounds,

(d) the 'long-run' metaphor can be rendered operational, and

(e) the link between the measure-theoretic results and real-world phenomena is provided by viewing data $\mathbf{x}_0$ as a 'truly typical' realization of the stochastic process $\{X_t,\ t\in\mathbb{N}\}$ underlying $\mathcal{M}_\theta(\mathbf{x})$.

In light of the above comments, the frequentist interpretation of probability, based on the SLLN, does *not* endorse the *straight rule* $P(A)=(\frac{1}{n}\sum_{k=1}^{n}x_k)$ as a basis of inference for a particular realization $\{x_k\}_{k=1}^{n}$.

1.2 The probability for statistics and inference procedures

The above framing of the frequentist interpretation of probability for an event A is general enough to be extended in all kinds of different set-ups within frequentist inference, including **the error probabilities**. In the context of a statistical model $\mathcal{M}_\theta(\mathbf{x})$, $\mathbf{x}\in\mathbb{R}_X^n$, the sequence of data come in the form of N realizations $\mathbf{x}^1,\mathbf{x}^2,...,\mathbf{x}^N$ from the same sample space $\mathbb{R}_X^n$.

Example. Consider the following hypotheses:

$$H_0\text{: }\mu\leq\mu_0,\ \text{vs.}\quad H_1\text{: }\mu>\mu_0, \tag{2}$$

in the context of the *simple* (one parameter) *Normal model*:

$$\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{x})\text{: } X_t \backsim \mathsf{N}(\mu,\sigma^2)\,,\ [\sigma^2 \text{ known}],\ t{=}1,2,...,n,...,$$

for which the optimal test is $\mathcal{T}_\alpha{:=}\{\kappa(\mathbf{X}),\ C_1(\alpha)\}$:

$$\begin{array}{l}\text{test statistic: } \kappa(\mathbf{X}){=}\frac{\sqrt{n}(\overline{X}_n-\mu_0)}{\sigma},\ \overline{X}_n{=}\frac{1}{n}\sum_{k=1}^{n}X_k,\\ \text{rejection region: } C_1(\alpha){=}\{\mathbf{x}\text{: } \kappa(\mathbf{x})>c_\alpha\}.\end{array} \tag{3}$$

To evaluate the error probabilities one needs the distribution of $\kappa(\mathbf{X})$ under H_0 and H_1:

[i] $\kappa(\mathbf{X}){=}\frac{\sqrt{n}(\overline{X}_n-\mu_0)}{\sigma}\overset{\mu=\mu_0}{\backsim}\mathsf{N}(0,1)$,

[ii] $\kappa(\mathbf{X}){=}\frac{\sqrt{n}(\overline{X}_n-\mu_0)}{\sigma}\overset{\mu=\mu_1}{\backsim}\mathsf{N}(\delta_1,1),\ \delta_1{=}\frac{\sqrt{n}(\mu_1-\mu_0)}{\sigma}{>}0$ for all $\mu_1>\mu_0$.

These hypothetical sampling distributions are then used to compare H_0 or H_1 via $\kappa(\mathbf{x}_0)$ to the true value $\mu{=}\mu^*$ represented by data $\mathbf{x}_0$ via $\overline{X}_n$, the best estimator of μ. The evaluation of the type I error probability and the p-value is based on [i] $\kappa(\mathbf{X}){=}\frac{\sqrt{n}(\overline{X}_n-\mu_0)}{\sigma}\overset{\mu=\mu_0}{\backsim}\mathsf{N}(0,1)$:

$$\begin{aligned}\alpha&{=}\mathbb{P}(\kappa(\mathbf{X})>c_\alpha;\mu{=}\mu_0),\\ p(\mathbf{x}_0)&{=}\mathbb{P}(\kappa(\mathbf{X})>\kappa(\mathbf{x}_0);\mu{=}\mu_0),\end{aligned}$$

and the evaluation of type II error probabilities and power is based on:

$$\text{[ii] } \kappa(\mathbf{X}){=}\tfrac{\sqrt{n}(\overline{X}_n-\mu_0)}{\sigma}\overset{\mu=\mu_1}{\backsim}\mathsf{N}(\delta_1,1),\text{for } \mu_1>\mu_0.$$

$$\begin{aligned}\beta(\mu_1)&{=}\mathbb{P}(\kappa(\mathbf{X})\leq c_\alpha;\mu{=}\mu_1)\text{ for all }\mu_1>\mu_0.\\ \pi(\mu_1)&{=}\mathbb{P}(\kappa(\mathbf{X})>c_\alpha;\mu{=}\mu_1)\text{ for all }\mu_1>\mu_0.\end{aligned}$$

How do these error probabilities fit into the above frequentist interpretation of probability that revolves around the long-run metaphor?

Type I error probability. The event of interest for the evaluation of α is:

$$(Z{=}1){:=}A{=}\{\mathbf{x}\text{: } \kappa(\mathbf{x})>c_\alpha\},\ \forall\mathbf{x}{\in}\mathbb{R}^n,$$

and the distribution where the probabilities come from is [i] $\kappa(\mathbf{X}){=}\frac{\sqrt{n}(\overline{X}_n-\mu_0)}{\sigma}\overset{\mu=\mu_0}{\backsim}\mathsf{N}(0,1)$. One draws N IID samples of size n from $\mathsf{N}(0,1)$, that give rise to the realizations $\mathbf{x}^1$, $\mathbf{x}^2$, ..., $\mathbf{x}^N$.For each sample realization one evaluates $\kappa(\mathbf{x}^i)$ and considers the relative frequency of event A occurring. That relative frequency is the sample equivalent to the significance level α.

The power of the test. The event of interest is:

$$(Z{=}1){:=}A{=}\{\mathbf{x}:\kappa(\mathbf{x})>c_\alpha\},\ \forall\mathbf{x}{\in}\mathbb{R}^n,$$

but the distribution from where the realizations $\mathbf{x}^1$, $\mathbf{x}^2$, ..., $\mathbf{x}^N$ come from is:

$$\text{[ii] } \kappa(\mathbf{X})=\frac{\sqrt{n}(\overline{X}_n-\mu_0)}{\sigma} \overset{\mu=\mu_1}{\backsim} \mathsf{N}(\delta_1,1),\ \mu_1>\mu_0.$$

The same evaluation as that associated with α will now give rise to the relative frequency associated with power of the test at $\pi(\mu_1)$ for a specific μ_1.

The type II error probability. The event of interest is also

$$(Z{=}1){:=}A{=}\{\mathbf{x}:\kappa(\mathbf{x})\leq c_\alpha\},\ \forall\mathbf{x}\in\mathbb{R}^n,$$

and the distribution from where the realizations $\mathbf{x}^1$, $\mathbf{x}^2$, ..., $\mathbf{x}^N$ come from is

$$\text{[ii] } \kappa(\mathbf{X})=\tfrac{\sqrt{n}(\overline{X}_n-\mu_0)}{\sigma} \overset{\mu=\mu_1}{\backsim} \mathsf{N}(\delta_1,1),\ \mu_1>\mu_0.$$

The p-value. The event of interest for the evaluation of the p-value is:

$$(Z{=}1){:=}A{=}\{\mathbf{x}:\ \kappa(\mathbf{x})>\kappa(\mathbf{x}_0)\},\ \forall\mathbf{x}\in\mathbb{R}^n,$$

and the distribution where the probabilities come from is [i] $\kappa(\mathbf{X})=\frac{\sqrt{n}(\overline{X}_n-\mu_0)}{\sigma} \overset{\mu=\mu_1}{\backsim} \mathsf{N}(0,1)$. The data specificity of the p-value does not matter in this case because:

$$p(\mathbf{X}) \overset{\mu=\mu_0}{\backsim} \mathsf{U}(0,1),$$

which implies that $\mathbb{P}(\kappa(\mathbf{X})<c;\mu{=}\mu_0){=}c$.

Post-data severity. The event of interest will be either of the events

$$(Z{=}1){:=}A{=}\{\mathbf{x}:\kappa(\mathbf{x})\gtrless\kappa(\mathbf{x}_0)\},\ \forall\mathbf{x}\in\mathbb{R}^n,$$

depending on the inferential claim $\mu\gtrless\mu_0+\gamma$ evaluated. The distribution from where the realizations $\mathbf{x}^1$, $\mathbf{x}^2$, ..., $\mathbf{x}^N$ come from is [ii] $\kappa(\mathbf{X})=\frac{\sqrt{n}(\overline{X}_n-\mu_0)}{\sigma} \overset{\mu=\mu_1}{\backsim} \mathsf{N}(\delta_1,1)$, with one **caveat**: the legitimate realizations $\mathbf{x}^1$, $\mathbf{x}^2$, ..., $\mathbf{x}^N$ should take values $\kappa(\mathbf{x}_0)\pm\varepsilon$. This is necessary because under $\mu{=}\mu_1$ the distribution associated with events $\{\mathbf{x}$: $\kappa(\mathbf{x})\gtrless\kappa(\mathbf{x}_0)\}$ is *not* Uniformly distributed.