§5. The Bernoulli Scheme. I. The Law of Large Numbers.

1.

Def A $\boxed{\text{Bernoulli scheme}}$ is a probabilistic model $(\Omega, \mathcal{A}, P)$ w/ $\Omega = \{\omega \mid \omega = (a_1, \cdots, a_n),\ a_i = 0, 1\}$, $\mathcal{A} = \mathcal{P}(\Omega)$, $p(\omega) = p^{\sum a_i} q^{n - \sum a_i}$.

i.e. it is a probabilistic model of n indep. experiments w/ two outcomes.

Define the random variables $\boxed{\xi_1}, \cdots, \boxed{\xi_n}$ by $\xi_i(\omega) = a_i$, where $\omega = (a_1, \cdots, a_n)$.

Define another list of random variables $\boxed{S_0}, \boxed{S_1}, \cdots, \boxed{S_n}$ by

$S_0(\omega) \equiv 0$, and $S_k = \xi_1 + \cdots + \xi_k$, $k = 1, \cdots, n$.

Note $E S_n = E(\xi_1 + \cdots + \xi_n) = \sum_{k=1}^{n} E\xi_k = np. \Rightarrow E\frac{S_n}{n} = p.$ ···· (1).

Moreover, since $\xi_1, \cdots, \xi_n$ are independent,

$$VS_n = V(\xi_1 + \cdots + \xi_n) = \sum_{k=1}^{n} V(\xi_k) = \sum_{k=1}^{n} E(\xi_k - E\xi_k)^2 = \sum_{k=1}^{n}\left[(1-p)^2 p + p^2 q\right] \overset{p+q=1.}{=} npq.$$

Thm. (Chebyshev's inequality)

$(\Omega, \mathcal{A}, P)$: probability space.

ξ : nonnegative random variable.

$\Rightarrow P\{\xi \geq \varepsilon\} \leq E(\xi/\varepsilon)$, $\forall\, \varepsilon > 0$. ···· (3)

[Proof].

$\xi \geq \varepsilon \cdot I_{\{\xi \geq \varepsilon\}}. \Rightarrow E\xi \geq E(\varepsilon \cdot I_{\{\xi \geq \varepsilon\}}) = \varepsilon P\{\xi \geq \varepsilon\}. \Rightarrow E(\xi/\varepsilon) \geq P\{\xi \geq \varepsilon\}.$ #

Cor ···(4)

ξ : random variable.

$\varepsilon > 0$.

$\Rightarrow P(|\xi| \geq \varepsilon) \leq E(|\xi|/\varepsilon)$.

$P(|\xi| \geq \varepsilon) \leq E(\xi^2/\varepsilon^2)$. (using $P(|\xi| \geq \varepsilon) = P(\xi^2 \geq \varepsilon^2)$)

$P(|\xi - E\xi| \geq \varepsilon) \leq (V\xi)/\varepsilon^2$. (replacing $|\xi|$ above by $|\xi - E\xi|$ and using def. of V (i.e. $V\xi = E(\xi - E\xi)^2$).

Some deductions:

Taking $\xi = S_n/n$ in the 3rd ineq. of (4), we have

$$P\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) \leq V\left(\frac{S_n}{n}\right)/\varepsilon^2 = \frac{VS_n}{n^2\varepsilon^2} = \frac{pq}{n\varepsilon^2} \leq 1/4n\varepsilon^2. \cdots (5)$$

For $p+q=1$, $p, q \geq 0$, by calculus, pq maximizes at $p = q = 1/2$.

Thus, for a fixed $\varepsilon > 0$, if n is very large, then the probability of the frequency $\frac{S_n}{n}$ of success deviating from p by ε is rather small.

Notation:

For $n \geq 1$ and $0 \leq k \leq n$, $\boxed{P_n(k)} := C_n^k p^k q^{n-k}$.

Using this notation, we can write $P\{|\frac{S_n}{n}-p|\geq\varepsilon\} = \sum_{\{k:|(k/n)-p|\geq\varepsilon\}} P_n(k)$, and (5) can be rewritten as $\sum_{\{k:|(k/n)-p|\geq\varepsilon\}} P_n(k) \leq \frac{pq}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}$. ... (6)

An immediate consequence is the following: For fixed $\varepsilon>0$,

$$\sum_{\{k:|k/n-p|\geq\varepsilon\}} P_n(k) \to 0 \text{ as } n\to\infty. \quad \cdots (7)$$ or (equivalently)

$$P\{|\frac{S_n}{n}-p|\geq\varepsilon\}\to 0 \text{ as } n\to\infty \quad \cdots (8)$$ or (more precisely)

$$P^{(n)}\{\omega^{(n)} \mid |\frac{S_n^{(n)}(\omega^{(n)})}{n}-p|\geq\varepsilon\} = \sum_{\{k:|k/n-p|\geq\varepsilon\}} P_n(k) \to 0 \text{ as } n\to\infty \quad \cdots (9).$$

Either of (7), (8), or (9) is called $\boxed{\text{James Bernoulli's law of large numbers.}}$

The reason for writing (9) is that P in (8) should vary by n because the probability space are varying w.r.t. n.

i.e. $(\Omega^{(n)}, \mathcal{A}^{(n)}, P^{(n)}), n\geq 1$, $\Omega^{(n)} = \{\omega^{(n)} = (a_1^{(n)}, \cdots, a_n^{(n)}), a_i^{(n)} = 0, 1\}$, $\mathcal{A}^{(n)} = \mathcal{P}(\Omega^{(n)})$, and $P^{(n)}(\omega^{(n)}) = p^{\sum a_i^{(n)}} q^{n-\sum a_i^{(n)}}$.

<u>Rmk</u>:

We can interpret (7), (8), or (8) by
(1) graphic interpretation (see Figure 6) or (2) wandering particle. (See Figure 7).

2.
An interpretation using the language of a large number of experiments.
See P50.

3.
A typical question arising in mathematical statistics:
Given small $\alpha>0$, what is the least number n of observations that guarantees us to have $P\{|\frac{S_n}{n}-p|<\varepsilon\}\geq 1-\alpha$. ⊛

The estimate $P\{|\frac{S_n}{n}-p|\geq\varepsilon\}\leq 1/4n\varepsilon^2$ gives us an upper bound for the answer:
It follows from the inequality that $P\{|\frac{S_n}{n}-p|<\varepsilon\} = 1-P\{|\frac{S_n}{n}-p|\geq\varepsilon\}\geq 1-1/4n\varepsilon^2$.
To achieve ⊛, we can simply let $\alpha = 1/4n\varepsilon^2$. $\Rightarrow n = 1/4\alpha\varepsilon^2$.
As $n\geq 1/4\varepsilon^2\alpha$, ⊛ would be achieved.
However, since Chebyshev's ineq. is really crude, $1/4\varepsilon^2\alpha$ is just an "upper bound" to the answer. #

4.

For notational convenience, denote $\boxed{C(n,\varepsilon)} = \{\omega : |\frac{S_n(\omega)}{n} - p| \le \varepsilon\}$.

$\omega \in C(n,\varepsilon)$ are called typical (or (n,ε)-typical).

Q: What is the size of $C(n,\varepsilon)$ and what are the weights $p(\omega)$, $\omega \in C(n,\varepsilon)$?

A: This will be answered by using a function called "entropy".

Def

$\{p_1, \cdots, p_r\}$: a finite probability distribution. (i.e. p_i are nonnegative w/ $p_1 + \cdots + p_r = 1$).

The entropy of this distribution is $\boxed{H} = -\sum_{i=1}^{r} p_i \ln p_i$, where $0 \ln 0 := 0$.

Rmk:

(1) $H \ge 0$ and $H = 0 \Leftrightarrow \exists\, i$ s.t. $p_i = 1$, $p_j = 0$, $\forall j \ne i$.

(2) By convexity of $x \ln x$, $H \le \ln r$ and H attains its max for $p_1 = \cdots = p_r = 1/r$.

(3) Let's think of a concept of "degree of indeterminacy".

① If $p_i = 1$, $p_j = 0$, $\forall j \ne i$, then the distribution $\{p_1, \cdots, p_r\}$ admits NO indeterminacy (namely 0)
(i.e. the result A_i is completely certain.)

② If $p_1 = \cdots = p_r = 1/r$, then $\{p_1, \cdots, p_r\}$ should admit the maximum indeterminacy.
(i.e. all results are equally possible; impossible to discover any preference).

Thus, as (1), (2) above shows, H, the entropy, can be used to represent the concept of degree of indeterminacy.

Now, we turn to a specific probability space. (I don't know why he abandons Bernoulli!!)

$\Omega = \{\omega \mid \omega = (a_1, \cdots, a_n),\ a_i = 1, \cdots, r\}$, $p(\omega) = p_1^{\nu_1(\omega)} \cdots p_r^{\nu_r(\omega)}$, where $\nu_i(\omega)$ = # of occurrence of i in $\omega = (a_1, \cdots, a_n)$, and $(p_1, \cdots, p_r)$: a probability distri.

For $\varepsilon > 0$, $n = 1, 2, \cdots$, denote

$$\boxed{C(n,\varepsilon)} = \{\omega : |\frac{\nu_i(\omega)}{n} - p_i| < \varepsilon,\ i = 1, \cdots, r\}.$$

← WARNING: Different meaning. (NOT even a generalization)

With these, we have:

Thm. (Macmillan)

Let $p_i > 0$, $i = 1, \cdots, r$ and $0 < \varepsilon < 1$. Then $\exists\, n_0 = n_0(\varepsilon; p_1, \cdots, p_r)$ (i.e. n_0 depends on $\varepsilon, p_1, \cdots, p_r$) s.t. for all $n > n_0$,

(a) $e^{n(H-\varepsilon)} \le N(C(n,\varepsilon_1)) \le e^{n(H+\varepsilon)}$;

(b) $e^{-n(H+\varepsilon)} \le p(\omega) \le e^{-n(H-\varepsilon)}$, $\forall \omega \in C(n,\varepsilon_1)$;

(c) $P(C(n,\varepsilon_1)) = \sum_{\omega \in C(n,\varepsilon_1)} p(\omega) \to 1$ as $n \to \infty$,

where $\varepsilon_1 = \min\{\varepsilon,\ \varepsilon / \{-2\sum_{k=1}^{r} \ln p_k\}\}$.

5. (A proof of Weierstrass theorem using law of large numbers).

Recall: (Weierstrass approximation theorem).

Given $f: [0,1] \to \mathbb{R}$ cont. Then $\exists$ seq. of poly. $B_n : [0,1] \to \mathbb{R}$ s.t. $B_n \to f$ unif. on $[0,1]$.

[Proof].

Define $B_n(p) = \sum_{k=0}^{n} f(\frac{k}{n}) C_n^k p^k q^{n-k}$, where $q = 1-p$.

These are called the $\boxed{\text{Bernstein polynomials}}$ since Bernstein invented this proof.

$\because f$: cont. on $[0,1]$ $\therefore f$: bdd and unif. cont. Given $\varepsilon > 0$.

Choose $\delta > 0$ s.t. $|f(x) - f(y)| \le \varepsilon$, $\forall |x-y| \le \delta$.

Also, choose $M > 0$ s.t. $|f(x)| \le M$, $\forall x \in [0,1]$.

Then $|f(p) - B_n(p)| = \left| \sum_{k=0}^{n} [f(p) - f(\frac{k}{n})] C_n^k p^k q^{n-k} \right|$ (since $\sum_{k=0}^{n} C_n^k p^k q^{n-k} \equiv 1$).

$$\le \sum_{k=0}^{n} |f(p) - f(\tfrac{k}{n})| C_n^k p^k q^{n-k} = \left[\sum_{\{k \mid |k/n - p| \le \delta\}} + \sum_{\{k \mid |k/n - p| > \delta\}} \right] (\cdots)$$

$$\le \underbrace{\sum_{\{k \mid |k/n - p| \le \delta\}} \varepsilon \cdot C_n^k p^k q^{n-k}}_{\text{(using unif. cont.)}} + \underbrace{2M \sum_{\{k \mid |k/n - p| > \delta\}} C_n^k p^k q^{n-k}}_{\text{(using bdd)}} \le \underbrace{\varepsilon}_{\sum_{k=0}^{n} C_n^k p^k q^{n-k} = 1} + \underbrace{2M \cdot \frac{1}{4n\delta^2}}_{(6)}.$$

Thus, $B_n \to f$ unif. on $[0,1]$. #

§6. The Bernoulli Scheme. II. Limit Theorems. (Local, De Moivre-Laplace, Poisson).

1.

SKIP!! I can't understand the meaning of Local Limit Theorem, which has some limit concept concerning the little O-notation.

2.

Recall the notation $P_n(k) = C_n^k p^k q^{n-k}$.

For $-\infty < a \le b < \infty$, denote $\boxed{P_n(a,b]} = \sum_{a < x \le b} P_n(np + x\sqrt{npq})$, where $\sum$ is over those x w/ $np + x\sqrt{npq}$ integer.

$\boxed{\Phi(x)} := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$. (Recall that $\int_{-\infty}^{\infty} e^{-t^2/2} dt = \sqrt{2\pi}$).

Thm. (De Moivre-Lapace Integral Theorem).

Let $0 < p < 1$ and $p + q = 1$.

Then

$$\sup_{-\infty \le a < b \le \infty} \left| P_n(a,b] - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx \right| \to 0, \text{ as } n \to \infty. \quad \cdots (21).$$

Note that $ES_n = np$, $VS_n = npq$ and $P\{S_n = k\} = P_n(k)$.

Thus, rewriting (21), we have

$$\sup_{-\infty\le a<b\le\infty} \left| P\left\{a < \frac{S_n - ES_n}{\sqrt{VS_n}} \le b\right\} - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\,dx \right| \to 0 \quad \text{as } n\to\infty.$$

Hence, for any $-\infty \le A < B \le \infty$, we have

$$P\{A < S_n \le B\} - \left[\Phi\left(\frac{B-np}{\sqrt{npq}}\right) - \Phi\left(\frac{A-np}{\sqrt{npq}}\right)\right] \to 0 \quad \text{as } n\to\infty \quad \cdots (22)$$

Example:

Formula (22) is quite useful:

Let a true die be tossed for 12000 times.

What is the probability P that the number of 6's lie in $(1800, 2100]$?

[Sol].

Let $p = 1/6$, $q = 5/6$, $n = 12000$. by (22)

Then $P = P\{1800 < S_n \le 2100\} \approx \Phi\left(\frac{2100 - 12000\times 1/6}{\sqrt{12000\times 1/6 \times 5/6}}\right) - \Phi\left(\frac{1800 - 12000\times 1/6}{\sqrt{12000\times 1/6\times 5/6}}\right) \approx 0.992$.

#