Probability theory 1

17/12/21

Recall: $X_i = 1$ w.p. $\frac{1}{2}$ and $S_0 = 0$

$= -1$ w.p. $\frac{1}{2}$

$S_m = \sum_{i=1}^{m} X_i$ for $1 \le m \le n$

Theorem: Denote $M_n = \max(S_0, S_1, \cdots, S_n)$. Then for $r \geq 0$,

$$P(M_n = r) = P(S_n = r) + P(S_n = r+1).$$

Theorem: (Aresine law for last visit)

For $n \geq 1$, denote $L_n = \frac{1}{n} \max\{0 \le k \le n : S_{2k} = 0\}$

Then $\displaystyle\lim_{n\to\infty} P\left(x < L_n \le y\right) = \frac{2}{\pi}\left(\arcsin\sqrt{y} - \arcsin\sqrt{x}\right) \qquad 0<x<y<1$

- The proof of the above theorem requires two lemmas.

<u>Lemma 1:</u> For $n \geq 1$ and $0 \le k \le n$,

$$P\left(L_n = \frac{k}{n}\right) = P\left(S_{2k}=0\right) P\left(S_{2n-2k}=0\right).$$

<u>Proof of lemma 1:</u> We had shown that

$$P(L_n=0) = \underline{P\left(S_1\neq 0, S_2\neq 0, \dots, S_{2n}\neq 0\right) = P\left(S_{2n}=0\right)} \qquad (1)$$

<u>HW6|8:</u> For a fixed $k \geq 1$, define

$$S_n' = S_{k+n} - S_k, \quad n \geq 0.$$

Show that for $n \geq 0$, $(S_0', S_1', \dots, S_n')$ is independent of S_k and has the distribution as $(S_0, S_1, \dots, S_n)$.

Intuition: $S_0' = S_k - S_k = 0 = S_0$

$S_1' = S_{k+1} - S_k = X_{k+1} \overset{d}{=} X_1 = S_1$

$\vdots$

$(S_0' = x_0, S_1' = x_1, \dots, S_n' = x_n)$

$= (S_0 = x_0, S_1 = x_1, \dots, S_n = x_n)$.

Going back to the proof of lemma 1:

For $0 \leq k \leq n$, $P\left(L_n = \frac{k}{n}\right) = P\left(\max\{0 \leq j \leq n : S_{2j} = 0\} = k\right)$

$S'_n = S_{2k+n} - S_{2k}$

$S'_{2n-2k} = S_{2k+2n-2k} - S_{2k}$
$= S_{2n} - S_{2k}$

$$= P(S_{2k}=0, S_{2k+1}\neq 0, \dots, S_{2n}\neq 0)$$

$$= P(S_{2k}=0, S_{2k+1}-S_{2k}\neq 0, \dots, S_{2n}-S_{2k}\neq 0)$$

$$= P(S_{2k}=0)\, P(S_{2k+1}-S_{2k}\neq 0, \dots, S_{2n}-S_{2k}\neq 0)$$

$$= P(S_{2k}=0)\, P(S'_1\neq 0, S'_2\neq 0, \dots, S'_{2n-2k}\neq 0)$$

by HW6/8 $= P(S_{2k}=0)\, P(S_1\neq 0, S_2\neq 0, \dots, S_{2n-2k}\neq 0)$

by ① $= P(S_{2k}=0)\, P(S_{2n-2k}=0)$. $\square$

Lemma 2: As $n\to\infty$ $\lim_{n\to\infty} \sqrt{\pi n}\, P(S_{2n}=0) = 1$.

$\Rightarrow$ Lemma 2 implies for any $\varepsilon > 0$ $\exists N \in \mathbb{N}$ such that

$$\left| \sqrt{\pi n}\, P(S_{2n}=0) - 1 \right| < \varepsilon \qquad \forall\, n \geq N$$

$$\Rightarrow \quad -\varepsilon < \sqrt{n\pi}\, P(S_{2n}=0) - 1 < \varepsilon$$

$$\Rightarrow \quad \frac{1-\varepsilon}{\sqrt{\pi n}} < P(S_{2n}=0) < \frac{1+\varepsilon}{\sqrt{\pi n}} \qquad \text{——— } (2)$$

<u>Stirling's formula:</u>

As $n \to \infty$, $\quad n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$ i.e.

$$\lim_{n\to\infty} \frac{n!}{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n} = 1 .$$

Fact 1: Suppose that f is a continuous function from a compact interval $[a,b]$ to $\mathbb{R}$. Then,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=\lfloor na \rfloor+1}^{\lfloor nb \rfloor} f\left(\frac{j}{n}\right) = \int_a^b f(x)\,dx$$

where $\lfloor x \rfloor$ is the greatest integer not greater than x.

Similarly $\lceil x \rceil$ is the smallest integer not smaller that x.

Fact 2: (Fundamental theorem of Calculus)

If $F : [a,b] \to \mathbb{R}$ is continuously differentiable, then

$$\int_a^b F'(x)\,dx = F(b) - F(a).$$

Proof of the theorem: Towards showing

$$\lim_{n\to\infty} P\left(x < L_n \le y\right) = \frac{2}{\pi}\left(\sin^{-1}\sqrt{y} - \sin^{-1}\sqrt{x}\right) \qquad 0 < x, y < 1.$$

$\sup_P U(P,f) = \inf_P L(P,f)$

$U(P,f) \ne L(P,f)$

$\| \qquad \|$

$\sum M_i \delta x_i \qquad \sum m_i \delta x_i$

$\| \qquad \|$

$1 \qquad 0$

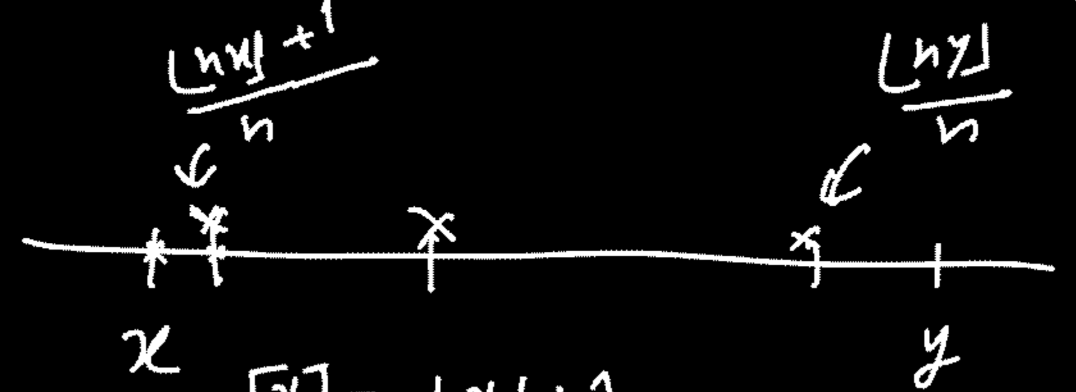
Fix, $0 < \varepsilon < 1$. In view of Lemma 2, There exists a $N \in \mathbb{N}$ such that for all $n \ge N$ we have

$$\frac{1-\varepsilon}{\sqrt{\pi n}} \le P\left(S_{2n} = 0\right) \le \frac{1+\varepsilon}{\sqrt{\pi n}} \qquad \text{(from equation (2))}. \qquad (3)$$

$\left\lfloor \frac{\lfloor x \rfloor}{n} \right\rfloor = \left\lfloor \frac{x}{n} \right\rfloor$

Fix $0 < x < y < 1$ and note that for all $n \ge 1$,

$$P\left(x < L_n \le y\right) = P\left(nx < nL_n \le ny\right)$$



$\lceil x \rceil = \lfloor x \rfloor + 1$

$= \frac{n(\lfloor x \rfloor + 1)}{n} = \frac{n\lfloor x \rfloor + n}{n}$

$$= \sum_{j=\lfloor nx \rfloor+1}^{\lfloor ny \rfloor} P(nL_n = j) = \sum_{j=\lfloor nx \rfloor+1}^{\lfloor ny \rfloor} P\left(L_n = \frac{j}{n}\right)$$

Applying Lemma 1

$$= \sum_{j=\lfloor nx \rfloor+1}^{\lfloor ny \rfloor} P(S_{2j} = 0)\, P(S_{2n-2j} = 0) \quad \text{—— (4)}$$

Now if $n \geq \underbrace{\left\lceil \frac{N}{x} \right\rceil \vee \left\lceil \frac{N}{1-y} \right\rceil}_{= \max\left\{ \left\lceil \frac{N}{x} \right\rceil, \left\lceil \frac{N}{1-y} \right\rceil \right\}}$, then $\forall\, j \in [nx, ny] \cap \mathbb{N}$

it holds that

$j \geq N$ and $n-j \geq N$, and hence

$$\lceil x \rceil = \lfloor x \rfloor + 1 \qquad x = 2.5$$

$$= \frac{n(\lfloor x \rfloor + 1)}{n}$$

$$\stackrel{?}{=} \frac{\lfloor nx \rfloor + 1}{n}$$

$$P(x < L_n \leq y)$$

$$= P(nx < nL_n \leq ny)$$

$$= \sum_{j=\lfloor nx \rfloor+1}^{\lfloor ny \rfloor} P(nL_n = j)$$

$$P(nL_n = j) = P\left(L_n = \frac{j}{n}\right)$$

$$\left\lceil \frac{N}{x} \right\rceil \leq n$$

$$j \geq nx \geq \left\lceil \frac{N}{x} \right\rceil x \geq \frac{N}{x} x = N$$

$$\frac{1-\varepsilon}{\sqrt{\pi j}}\,\frac{1-\varepsilon}{\sqrt{\pi(n-j)}} \le P(S_{2j}=0)\,P(S_{2n-2j}=0) \le \frac{1+\varepsilon}{\sqrt{\pi j}}\,\frac{1+\varepsilon}{\sqrt{\pi(n-j)}}$$

by 3.

$$\frac{1-\varepsilon}{\sqrt{\pi n}} \le P(S_{2n}=0) \le \frac{1+\varepsilon}{\sqrt{\pi n}} \qquad \text{— 3}$$

That is for large n, summing the above we get,

$$\frac{(1-\varepsilon)^2}{\pi}\sum_{j=\lfloor nx\rfloor+1}^{\lfloor ny\rfloor}\frac{1}{\sqrt{j(n-j)}} \le \sum_{j=\lfloor nx\rfloor+1}^{\lfloor ny\rfloor} P(S_{2j}=0)\,P(S_{2n-2j}=0) \le \frac{(1+\varepsilon)^2}{\pi}\sum_{j=\lfloor nx\rfloor+1}^{\lfloor ny\rfloor}\frac{1}{\sqrt{j(n-j)}} \qquad \text{— 5}$$

|| by 4

$$P(x < L_n \le y)$$

as $n\to\infty$:

$$\frac{(1-\varepsilon)^2}{\pi}\int_x^y \frac{dt}{\sqrt{t(1-t)}} = \frac{(1-\varepsilon)^2\,2}{\pi}\left(\sin^{-1}\sqrt{y}-\sin^{-1}\sqrt{x}\right)$$

applies fundamental thm of calculus.

as $n\to\infty$:

$$\frac{(1+\varepsilon)^2}{\pi}\int_x^y \frac{dt}{\sqrt{t(1-t)}} = \frac{(1+\varepsilon)^2\,2}{\pi}\left(\sin^{-1}\sqrt{y}-\sin^{-1}\sqrt{x}\right)$$

Observe that $\displaystyle\sum_{j=\lfloor nx\rfloor+1}^{\lfloor ny\rfloor} \frac{1}{\sqrt{j(n-j)}} = \frac{1}{n}\sum_{j=\lfloor nx\rfloor+1}^{\lfloor ny\rfloor} \frac{1}{\sqrt{\frac{j}{n}\left(1-\frac{j}{n}\right)}} \xrightarrow{\text{by Fact 1}} \underbrace{\int_x^y \frac{dt}{\sqrt{t(1-t)}}}$

H.W. Check the integration.

From ⑧,

$$\frac{(1-\varepsilon)^2}{\pi}\,2\left(\sin^{-1}\sqrt{y}-\sin^{-1}\sqrt{x}\right) \le \liminf P\left(x < L_n \le y\right) \le \limsup P\left(x < L_n \le y\right) \le \frac{(1+\varepsilon)^2}{\pi}\,2\left(\sin^{-1}\sqrt{y}-\sin^{-1}\sqrt{x}\right)$$

Letting $\varepsilon \downarrow 0$ completes the proof.

Probability generating function:

Definition: Let X be a random variable taking values $0, 1, 2, \cdots$. The probability generating function (PGF) of X is a function $f : [-1, 1] \to \mathbb{R}$ defined by

$$f(s) = \sum_{n=0}^{\infty} P(X=n)\, s^n, \qquad s \in [-1, 1].$$

$a-R \quad a \quad a+R$

$$\sum_{n=0}^{\infty} a_n (x-a)^n$$

$$\sum_{n=0}^{\infty} a_n x^n$$

R

$-R \quad 0 \quad R$

Remark: The PGF f of X can also be represented as

$$f(s) = E(s^X), \qquad |s| \le 1.$$

$$\left| \sum_{n=0}^{\infty} P(X=n)\, s^n \right| < \infty \quad ??$$

$$\leq \sum_{n=0}^{\infty} P(X=n)\, |s|^n \quad . \quad \square$$