(1) ابتدا تعداد بازی ها را بدست می آوریم که 63 تا می شود و به شرح زیر است:

مرحله اول → 32 بازی → 2^5 بازی
مرحله دوم → 16 بازی → 2^4 بازی
⋮
مرحله ششم → 1 بازی → 2^0 بازی

$$\oplus = 2^5 + 2^4 + \dots + 2^0 = 2^6 - 1 = 63$$

بنابراین 63 آزمایش تصادفی داریم که هر کدام یک متغیر تصادفی x_i نسبت می دهیم. مشخصا متغیر تصادفی های هر مرحله امید ریاضی یکسانی دارند و متغیر تصادفی x_1 تا x_{32} مربوط به لول 1 و x_{33} تا x_{48} مربوط به لول 2 و ... نهایتا هم متغیر تصادفی x_{63} مربوط به لول 6 است.

$$x_i = \begin{cases} 2^{l_i - 1} & \text{پیش بینی درست} \\ 0 & \text{پیش بینی اشتباه} \end{cases}$$

جدول امتیاز متغیر تصادفی ها به شرح رو به رو است و اگر اشتباه پیش بینی کند امتیاز صفر می گیرد و اگر درست پیش بینی کند امتیاز آن 2^{l-1} است. حال امید ریاضی جمع این متغیر تصادفی ها را می خواهیم.

$\leq i \leq 32, \; E[x_i] = \sum_{x=0}^{1} x P(X=x) = \frac{1}{2}$ ، $\underset{33 \leq i \leq 48}{E[x_i]} = \sum_{x=0}^{2} x P(X=x) = 2 \times \frac{1}{4} = \frac{1}{2}$

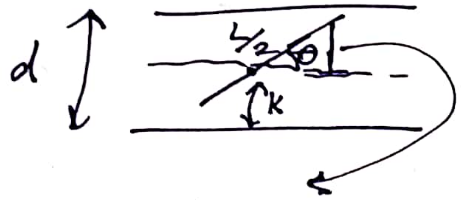
در حالت کلی می‌توان نوشت :

$$E[x_i] = \sum_{x=0}^{2^{l_i}-1} x P(X=x) = 2^{l_i-1} \times \frac{1}{2^{l_i}} = \frac{1}{2}$$

پس امید ریاضی هر بازی $\frac{1}{2}$ است با توجه به این که ۶۳ بازی در مجموع داریم و امید ریاضی خاصیت خطی بودن دارد داریم :

$$E[x_1 + \cdots + x_{63}] = E[x_1] + E[x_2] + \cdots + E[x_{63}] = 63 \times \frac{1}{2} = \cancel{31} \; \boxed{31,5}.$$

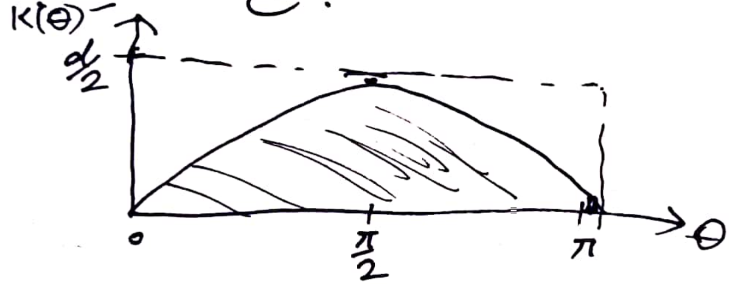
۲) فرض می‌کنیم $L<d$ است. مرکز سوزن را O در نظر می‌گیریم.



$\frac{L}{2}\sin\theta$

زاویه‌ای که سوزن با خط افقی می‌سازد را θ در نظر می‌گیریم و
کمترین فاصله مرکز سوزن از خطوط افقی را با K نشان می‌دهیم.

با تعریف $\sin$ می‌دانیم ضلع روبه‌رو به θ طول $\frac{L}{2}\sin\theta$ دارد و واضح است برای اینکه سوزن ضلع‌ها را قطع کند باید $K\leq\frac{L}{2}\sin\theta$ باشد. در حالت مساوی $K=\frac{L}{2}\sin\theta$ است که رسم K برحسب تابعی از θ باید جالب باشد.



θ و K هر دو RV هستند و بازه آن‌ها به صورت زیر تعریف می‌شود:

$$0\leq K\leq\frac{d}{2}$$

$$0\leq\theta\leq\pi$$

$$\text{مساحت هاشور}=\int_0^{\pi}\frac{L}{2}\sin\theta\,d\theta=\frac{L}{2}\left(-\cos\theta\Big]_0^{\pi}\right)=\frac{L}{2}(1-(-1))=L$$

$$\text{احتمال قطع}=\frac{\text{مساحت هاشور}}{\text{مساحت کل}}=\frac{L}{\frac{\pi d}{2}}=\frac{2L}{\pi d}$$

۳) الف) ابتدا تعریف امید ریاضی را می‌نویسیم و البته فرض می‌کنیم که متغیر تصادفی ما، پیوسته است حال داریم:

$$E[x]=\int_{-\infty}^{+\infty} x f_x(x)\,dx \overset{\text{نامنفی}}{=\!=} \int_0^{\infty} x f_x(x)\,dx = \int_0^a x f(x)\,dx + \int_a^{\infty} x f(x)\,dx \geq$$

$$\int_a^{\infty} a f(x)\,dx = aP(x\geq a) \Rightarrow E[x]\geq aP(x\geq a) \Rightarrow P(x\geq a)\leq \frac{E[x]}{a}$$

۳) ب) فرض می‌کنیم $(z-\mu)^2$ برابر k باشد:

$$*\ (z-\mu)^2\geq 0$$

$$P((z-\mu)^2\geq \varepsilon^2) = P(K\geq \varepsilon^2) \overset{\text{مارکوف}}{\leq} \frac{E[k]}{\varepsilon^2} = \frac{E[(k-\mu)^2]}{\varepsilon^2} \overset{\text{تعریف}}{=} \frac{\sigma^2}{\varepsilon^2}$$

* البته برای حدس اولیه عبارت به σ^2 نگاه کردم و سعی کردم به خوبی σ^2 را بسازم. حال داریم

$$\Big\downarrow =$$

$$P(|z-K|\geq \varepsilon)\leq \frac{\sigma^2}{\varepsilon^2}$$

البته یک روش دیگر هم دارد که نامساوی را اثبات کنیم!

۳) ب) اثبات نامساوی چبیشف: این نامساوی هم برای متغیر تصادفی یک bound بدست میاورد.

اگر نامساوی چبیشف را طرفین وسطین کنیم بدست می‌آید

$$\sigma^2 \geq P(|Z-\mu| \geq \varepsilon)\,\varepsilon^2$$

بنابراین باید با انحراف معیار کار کنیم.

$$\sigma^2 \triangleq E[(Z-\mu)^2] = \int_{-\infty}^{+\infty}(z-\mu)^2 f_Z(z)\,dz \geq \int_{-\infty}^{\mu-\varepsilon}(z-\mu)^2 f_Z(z)\,dz + \int_{\mu+\varepsilon}^{+\infty}(z-\mu)^2 f_Z(z)\,dz = *$$

با این کار بازه به طول ۲ε از انتگرال حذف شده.

$$\int_{-\infty}^{\mu-\varepsilon}(z-\mu)^2 f_Z(z)\,dz \geq \int_{-\infty}^{\mu-\varepsilon}\varepsilon^2 f_Z(z)\,dz \quad ②*$$

②* علت ۲ این است که چون توان ۲ داریم کوچکترین مقداری که z می‌تواند بگیرد در بالای انتگرال است

$$\int_{\mu+\varepsilon}^{+\infty}(z-\mu)^2 f_Z(z)\,dz \geq \int_{\mu+\varepsilon}^{+\infty}\varepsilon^2 f_Z(z)\,dz \quad ③$$

$$\overset{2,3}{*} \geq \varepsilon^2\left(\int_{\mu+\varepsilon}^{+\infty} f_Z(z)\,dz + \int_{-\infty}^{\mu-\varepsilon} f_Z(z)\,dz\right) = \varepsilon^2\big(P(Z\geq\mu+\varepsilon)+P(Z\leq\mu+\varepsilon)\big) = \varepsilon^2 P(|Z-\mu|\geq\varepsilon)$$

بنابراین فرهمیدیم.

$$\sigma^2 \geq \varepsilon^2 P(|Z-\mu| \geq \varepsilon) \quad \text{یا} \quad P(|Z-\mu|\geq\varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

۳) ب) فرض می‌کنیم منظور از متغیر تصادفی متغیر تصادفی یکنواخت (uniform) است. بنابراین احتمال داخل بودن $\frac{\pi}{4}$ است.

می‌دانیم نقاط از هم مستقل هستند و احتمال ~~[illegible]~~ موفقیت نیز ثابت است بنابراین داخل یا خارج بودن یک متغیر تصادفی برنولی است.

یک تقریب خوب برای احتمال $\frac{\pi}{4}$ تعداد نقاط درست به کل نقاط است (چرا؟)

شهودی اینجوری است و لی عملی ثابت می‌کنیم.

$$x = \begin{cases} 1 & \text{داخل} = \text{Success} \\ 0 & \text{خارج} \end{cases}$$

$$\frac{\text{تعداد درست}}{\text{تعداد کل}} = \frac{\sum_{i=1}^{n} x_i}{n} = p\left(\frac{\pi}{4}\right)$$

بنابراین $\frac{\sum x_i}{n} \simeq \frac{\pi}{4}$. حال یک متغیر تصادفی جدید π' تعریف می‌کنیم که معادل زیر است.

$$x_i \sim \text{Bernoli}(p) = \text{Bernoli}\left(\frac{\pi}{4}\right) \Longrightarrow \begin{cases} E[x_i] = p = \frac{\pi}{4} \\ var[x_i] = pq = \frac{\pi}{4}\left(1 - \frac{\pi}{4}\right) = \frac{\pi(4-\pi)}{16} \end{cases}$$

$$\pi' = \frac{4}{n} \sum_{i=1}^{n} x_i$$

$$\pi' \longrightarrow \quad E[\pi'] = E\left[\frac{4}{n}\sum x_i\right] = \frac{4}{n}\sum_{i=1}^{n} E[x_i] = \frac{4}{n} \times n \times \frac{\pi}{4} = \pi \rightarrow$$

بنابراین π' تخمین خوبی از π است.

$$var[\pi'] = var\left[\frac{4}{n}\sum x_i\right] = \frac{4^2}{n^2}\sum_{i=1}^{n} var[x_i] = \frac{16}{n^2} \times n \times \frac{\pi(4-\pi)}{16} = \frac{\pi(4-\pi)}{n}$$

حال باید یک bound برای π' بدست آوریم.

$$P(|\pi' - E[\pi']| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} \Rightarrow$$

$$P(|\pi' - \pi| \geq 0.01) \leq \frac{\sigma^2}{\varepsilon^2} \leq 0.05 \longrightarrow \frac{\pi(4-\pi)}{(10^{-2})^2 n} \leq 0.05 \Rightarrow n \geq \frac{\pi(4-\pi)}{10^{-4} \times 5 \times 10^{-2}} = \frac{\pi(4-\pi)}{5} \times 10^6$$

می‌توان π را با ۳٫۱۴ جایگذاری کرد و جواب زیر را بدست آورد.

$$n \geq 539353.24 \geq \underline{539354}$$

اما اگر π را نمی‌دانیم و نباید جایگذاری کنیم. کمی بازی با ریاضی :)

$$\pi(4-\pi) = 4\pi - \pi^2 = \underbrace{-(\pi-2)^2}_{\text{همیشه نامثبت}} + 4 \leq 4$$

حال دوباره محاسبه می‌کنیم

$$n \geq \frac{4}{5} \times 10^6 = 0.8 \times 10^6 = 8 \times 10^5$$

3) پ، ب)

$$P(|E_{in} - E_{out}| > \varepsilon) \leq 2\ \text{exponent}(-2\varepsilon^2 N)$$

(دقت کنیم در فرمول هافدینگ نباید متغیر اصلی ضریب داشته باشد)

$$E_{in} = \frac{\varepsilon}{n}\sum_{i=1}^{n} x_i = \varepsilon\,\overline{x_i}, \quad E_{out} = \pi$$

$$P(|\varepsilon\overline{x_i} - \pi| \geq 0.01) = P\left(\left|\overline{x_i} - \frac{\pi}{\varepsilon}\right| \geq \frac{0.01}{4}\right) \leq 2\exp\left(-2\times 10^{-4}\times\left(\frac{1}{4}\right)^2\times N\right) \leq 0.05$$

$$-2\times 2^{-4}\times 10^{-4}\times N \leq \ln 5\times 10^{-2}\times\frac{1}{2} \rightarrow$$

$$N \geq \ln\left(\frac{5}{2}\times 10^{-2}\right)\cdot -\frac{1}{2}\times 2^{4}\times 10^{4} = +3,688\times 8\times 10^{4} \simeq 295110.3$$

$$\underline{N \geq 295111}$$

۴) الف

General Perceptron Model

مدل کلی پرسپترون با ۲ ورودی معادل زیر است:

		AND	sgn
0	0	0	−
0	1	0	−
1	0	0	−
1	1	1	+

$$f(x_1, x_2) = w_1 x_1 + w_2 x_2 + b$$

$$\left.\begin{array}{l} f(0,0) = b < 0 \xrightarrow{\text{example}} b = -1 \\ f(0,1) = w_2 + b < 0 \xrightarrow{\text{ex}} w_2 = 0.7 \\ f(1,0) = w_1 + b < 0 \longrightarrow w_1 = 0.7 \end{array}\right\} \xrightarrow{\text{Model}} f(x_1, x_2) = f(\vec{x}) = 0.7x_1 + 0.7x_2 - 1$$

$$f(1,1) = w_1 + w_2 + b > 0 \longrightarrow \checkmark \text{ checked}$$

OR

input		out	sgn
0	0	0	−
0	1	1	+
1	0	1	+
1	1	1	+

$$f(0,0) = b < 0 \longrightarrow b = -1$$

$$f(0,1) = w_2 + b > 0 \longrightarrow w_1 = 1.5 \longrightarrow f(x_1, x_2) = 1.5x_1 + 1.5x_2 - 1$$

$$f(1,0) = w_1 + b > 0 \longrightarrow w_2 = 1.5$$

$$f(1,1) = w_1 + w_2 + b > 0 \longrightarrow \checkmark$$

۴- ب

XOR

in	out	sgn
0 0	0	−
0 1	1	+
1 0	1	+
1 1	0	−

$f(x_1, x_2) = w_1 x_1 + w_2 x_2 + b$

$f(0,0) = b < 0 \longrightarrow$ b منفی است

$f(0,1) = w_2 + b > 0 \Longrightarrow w_2 > 0$

$f(1,0) = w_1 + b > 0 \longrightarrow w_1 > 0$

$f(1,1) = w_1 + w_2 + b < 0$ ✗

نامساوی آخر غیر ممکن است زیرا:

$$\left.\begin{array}{l} (1)\ w_1 + b > 0 \\ (2)\ w_2 > 0 \end{array}\right] \oplus \longrightarrow w_1 + w_2 + b > 0$$!✗

تناقض

۴ ـ ب

ابتدا یک کران بالا برای اندازه وزن در مرحله k ام بدست می‌آوریم. فرض $w^0 = 0$

$$\|w^k\|^2 = \|w^{k-1} + y^i x^i\|^2 = \|w^{k-1}\|^2 + \|y^i x^i\|^2 + \underbrace{2(w^{k-1} \cdot x^i) y^i}_{\text{همیشه نامثبت}}$$

$$\leq \|w^{k-1}\|^2 + \|y^i x^i\|^2 = \|w^{k-1}\|^2 + \|x^i\|^2 \leq \|w^{k-1}\|^2 + R^2$$

پس $\|w^k\|^2 \leq \|w^{k-1}\|^2 + R^2 \xrightarrow{\text{استقرا}} \|w^k\|^2 \leq \|w^0\|^2 + kR^2 \Rightarrow \|w^k\|^2 \leq kR^2$ $\boxed{1}$

حال یک کران پایین برای ضرب داخلی (w^*, w^k) داریم.

$$w^* \cdot w^k = w^* \cdot (w^{k-1} + x^i y^i) = w^{k-1} \cdot w^* + y^i (x^i \cdot w^*) > w^{k-1} \cdot w^* + \lambda$$

دوباره با استقراء داریم.

$$w^* \cdot w^k > w^* \cdot w^{k-1} + \lambda \xrightarrow{\text{استقراء}} w^* \cdot w^k > w^0 + k\lambda \xrightarrow{\text{اندازه}} \|w^*\| \|w^k\| > k\lambda \Rightarrow \|w^k\| > k\lambda \quad \boxed{2}$$

می‌دانیم $a \cdot b \leq \|a\| \|b\| \Rightarrow w^* \cdot w^k \leq \|w^*\| \|w^k\| = \|w^k\|$

$\boxed{1}$ و $\boxed{2}$ $\quad k^2\lambda^2 \leq \|w^k\|^2 \leq kR^2 \Rightarrow k^2\lambda^2 \leq kR^2 \xrightarrow{\div k\lambda^2} k \leq \frac{R^2}{\lambda^2} = \left(\frac{R}{\lambda}\right)^2$

۵) الف) خیر. تضمینی وجود ندارد زیرا تعمیم در واقع براساس اکثریت داده های in-sample انجام می شود و اونی انتخاب می شود که تعداد بیشتری از داده های in-sample به آن نزدیک باشند اما نمی توانیم به قطع در مورد generalization error داده های out-of-sample نظری بدهیم.

ب) بله وقتی این اتفاق می افتد که تمام یا اکثر داده های out-of-sample لیبل ۱- داشته باشند.

پ) الگوریتم A_1 فرضیه h_1 را انتخاب می کند و A_2، h_2 را زیرا تمام داده های in-sample لیبل ۱+ دارند.
چون توزیع y به گونه ای است که $P(y=1)=0.9$ و $P(y=0)=0.1$ است پس در کل خطای h_1، ۹۰ درصد شده و خطای h_2، ۱۰ درصد بنابراین $h_1 \gg h_2$ یعنی h_1 همیشه بهتر از h_2 است پس با احتمال $\boxed{1}$ A_1 بهتر از A_2 خواهد بود.

ت) مشخص است به ازای $p<0.5$ فرضیه ای که A_2 تولید می کند (h_2) بهتر از A_1 خواهد بود زیرا توزیع داده های خروجی به گونه ای است که اگر h_2 پیش بینی کند (۱-) احتمال بیشتری برای درست بودن دارد.