1.1 $N \cdot \hat{p}_N = \cancel{N} \cdot \frac{1}{\cancel{N}} (x_1 + x_2 + \cdots x_n)$

$= (x_1 + x_2 + \cdots x_n)$ Binomial Distribution

1.2 $E[\hat{p}_n]$: we know $\hat{p}_N = \frac{1}{N}(x_1 + x_2 + \cdots x_n)$, ~~where N is the total number, $(x_1 + x_2 \cdots x_n)$ is the number who vote~~

By Expect value of binomial distribution,

$E[\hat{p}_N] = E[\frac{x_1 + x_2 \cdots + x_n}{N}] = \cancel{N} \cdot p$

$= p$

1.3 ~~When N ... big~~

$\hat{p}_N = (\frac{x_1 + x_2 + \cdots x_n}{N})$

Binomial distribution : $X \sim \text{Binomial}(N, p) \rightarrow var(X) = N \cdot p(1-p)$

E error : $E((\hat{p}_N - E(\hat{p}_N))^2) < E$ $\because E(\hat{p}_N) = p$

$\therefore = E((\hat{p}_N - p)^2)$

$= (\frac{1}{N})^2 \cdot Var(x_1 + x_2 \cdots + x_N)$

$= (\frac{1}{N})^2 \cdot N \cdot p(1-p)$

$= \frac{1}{N} \cdot p(1-p) < E$

$= \frac{p(1-p)}{E} < N$

1.4 we know $N > \frac{p(1-p)}{E}$

Also, Max $p \cdot (1-p) = 0.25$, which means biggest Var Bernomial$(p) = 0.25$

$\therefore N > \frac{0.25}{E}$

$\therefore p(1-p) = 0.25$

$p = 0.5$

$\therefore N > \frac{0.5 \cdot (1-0.5)}{E}$

1.5 : $1 - 0.95 = 0.05$

$$N \geq \frac{p(1-p)}{0.05\,\varepsilon^2} \xleftarrow{\text{because}} P((|\hat{p}_N - p|)^2 \geq \varepsilon^2) \leq \frac{p(1-p)}{N\varepsilon^2}$$

$$0.05 \leq \frac{p(1-p)}{N\varepsilon^2}$$

$$\cancel{\geq \frac{p-p^2}{0.05\,\varepsilon^2}}$$

$$N \geq \frac{1}{0.05 \cdot 4\varepsilon^2}$$

$$N \geq \frac{1}{0.2\varepsilon^2}$$

$$N \geq \frac{5}{\varepsilon^2}, \quad 95\% \text{ confidence}$$

1.7 P(first head) = 0.5

P(second head) = 0.25

P(second tail) = 0.25

p = Expected $\hat{p}_N = E[\hat{p}_N]$

P(second head) + P(first head & truly support)

$$p = 0.25 + 0.5 \cdot q$$

$$2p = \frac{1}{2} + q$$

$$q = 2p - \frac{1}{2}$$

1.8. $q = 2p - 0.5$

$$\hat{q}_N = \frac{1}{N}(2p - 0.5)$$
$$= 2\hat{p}_N - 0.5$$

$$E[\hat{q}_N] = E[2\hat{p}_N - 0.5)]$$
$$= 2 \cdot \cancel{N} \cdot \frac{1}{N} \cdot p - 0.5$$
$$= 2p - 0.5 = q$$

$$Var(X) = E[X - E(X))^2]$$
$$= E[(\hat{q}_N - q)^2]$$
$$= Var[\hat{q}_N)$$
$$= E[(2\hat{p}_N - 1) - q)^2]$$
$$= Var(2\hat{p}_N - 1)$$

$$Var(\bar{x}) = \frac{1}{N^2}(Var(X_1) + Var(X_2) + \cdots Var(X_N))$$
$$= \bar{E}[((2\hat{p}_N - 1) - q)^2_1 + ((2\hat{p}_N - 1) - q)^2_2 + \cdots + ((2\hat{p}_N - 1) - q)^2_N]$$

$$Var(\hat{q}_N] = E[(2\hat{p}_N - 1) - E[(2\hat{p}_N - 1)]^2]$$
$$= E[2\hat{p}_N - 1 - E(2\hat{p}_N) - 1]$$
$$= E[4E[\hat{p}_N - p]]$$
$$= 4\, Var(\hat{p}_N)$$
$$= 4 \cdot \frac{p(1-p)}{N}$$

1.9. $P(|\hat{q}_N - E[\hat{q}_N]| < \varepsilon) \geq 90\%$

$P(|\hat{q}_N - q|^2 \geq \varepsilon^2)$

$$\leq \frac{E[(\hat{q}_N - E(\hat{q}_N))^2]}{\varepsilon^2} \leq \frac{4p\cdot(1-p)}{N\varepsilon^2}$$

$$\frac{4\cdot p(1-p)}{N\varepsilon^2} \leq 0.1$$

$$\frac{0.1}{4\cdot p(1-p)} \geq \frac{1}{N\varepsilon^2}$$

$$N \geq \frac{4p\cdot(1-p)}{0.1\,\varepsilon^2}$$

if we don't know p, the Max $p\cdot(1-p) = 0.25$

where $p = 0.5$

$$\therefore N \geq \frac{4\cdot 0.25}{0.1\,\varepsilon^2}$$

$$N \geq \frac{10}{\varepsilon^2}$$

1.10

P_m^i : probability of a person who supports i out of ~~[illegible]~~ m

$$P_i = q_i \cdot P(\text{head}) + P_m^i \cdot P(\text{tail})$$

$$= q_i \cdot \frac{1}{2} + P_m^i \cdot \frac{1}{2} = \frac{1}{2} q_i + \frac{1}{2} \cdot \frac{1}{m}$$

$$\hat{P}_N^i = \frac{1}{N} \cdot P_i$$

$$= \frac{1}{N} \cdot (x_1^1 + x_2^1 + \cdots x_N^1) + (x_1^2 + x_2^2 + \cdots x_N^2) + \cdots + (x_1^i + x_2^i + \cdots x_N^i)$$

$$P_i = \frac{1}{2} q_i + \frac{1}{2m}$$

$$2P_i = q_i + \frac{1}{m}$$

$$q_i = 2P_i - \frac{1}{m}$$

$$E[\hat{q}_N^i] = q_i$$

$$= 2(E[\hat{P}_N^i]) - \frac{1}{m}$$

$$= 2\left(\frac{q_i}{2} + \frac{1}{2m}\right) - \frac{1}{m}$$

$$= q_i + \frac{1}{m} - \frac{1}{m}$$

$$= q_i$$

1.11: $E\left[\sum_{i=1}^{M} (\hat{q}_N^i - q_i)^2\right]$

$= \sum_{i=1}^{M} E[(\hat{q}_N^i - q_i)^2]$

$= \sum_{i=1}^{M} E[(2\hat{p}_N^i - (\frac{1}{M}) - (2p_i - (\frac{1}{M})))^2]$

$= \sum_{i=1}^{M} E[4(\hat{p}_N^i - p_i)^2]$

$E[(\hat{p}_N^i - p_i)^2] = Var(\hat{p}_N)$

from 1.8, $Var(\hat{p}_N) = \frac{p(1-p)}{N}$

$= 4 \sum_{i=1}^{M} Var(\hat{p}_N^i)$

$= 4 \sum_{i=1}^{M} \frac{p_i(1-p_i)}{N}$

2.1 $P(\text{successful defend}) = (p_1 \cdot q_1) + (p_2 \cdot q_2) + \cdots + (p_i \cdot q_i)$
$= \sum_{1}^{i} p_i \cdot q_i$

2.2 If we know $\{q_1 \cdots q_N\}$, I would choose q_i with the highest probability of being attack.

When q is the most likely being attack from 1 to N, we should pick p_i that ~~ ~~ corespond to Max q.

The value of ~~this~~ this p_i should be 1, and the rest of p_i should be 0, since we can ~~ ~~ only protect one site

$\therefore P(\text{successful defense}) = \sum_{1}^{i} p_i \cdot q_i = q_i$

since other p_i are all zeros, except for p_i that we choose is 1.

since you know q_i is the Max q_i,
$P(\text{successful defense}) = \text{Max } q_i$

2.3. From 2.2, we know that the largest value for $P(\text{successful defend})$ will be the value of the Max q_i

Thus, we should make every q_i ~~ ~~ with the same probability, which means $q_i = \frac{1}{N}$, and there is no Max q_i, since every q_i has the same value

Thus, no matter which site defender is going to defend

$P(\text{successful defend})$ will always $= \frac{1}{N}$

$\therefore P(\text{successful attack})$ will be $= 1 - \frac{1}{N}$

2.4. To redo 2.1 in defender perspective :

in attacker perspective : $p(\text{successful defend}) = \sum_1^i p_i \cdot q_i$

$\therefore$ in defender perspective : $p(\text{successful defend}) = 1 - \sum_1^i p_i \cdot q_i$

To redo 2.2 in defender perspective.

in attacker view : p_i will ~~be~~ choose Max q_i, and value of p_i is 1, the rest of $p_i = 0$

$p(\text{successful defend}) = \sum_1^i p_i \cdot q_i = q_i$ same as Max q_i

in defender view : p_i will choose Minimum p_i

$p(\text{successful defend}) =$ minimum p_i out of $\{p_1 \cdots p_N\}$

To redo 2.3 in defender perspective

in attacker view : all ~~p_i~~ p(attack) should have same value, which is $\frac{1}{N}$, so p(successful defend) will be $\frac{1}{N}$, and $p(\text{successful attack}) = 1 - \frac{1}{N}$

in defender view : all p(defend) should have same value

minimum p out of $\{p_1 \cdots p_N\}$

$\therefore$ p(successful attack) should also be $\frac{1}{N}$

$p(\text{successful defend}) = 1 - \frac{1}{N}$

In conclusion, if attacker & defender both know each other's strategy, then they will set p(attack) & p(defend) to $\frac{1}{N}$

2.5 Failure is $(1-p)$, Successful is q, cost is C

$$E(X) = [(1-p_1)\cdot q_1 \cdot c_1] + [(1-p_2)\cdot q_2 \cdot c_2] + \dots + [(1-p_i)\cdot q_i \cdot c_i]$$

$$= \sum_{i=1}^{N} (1-p_i)\cdot q_i \cdot c_i$$

2.6 if we know $\{q_1, \dots q_N\}$, and same as 2.2

when we ~~set~~ find the largest q_i, we set our p_i to 1, and rest of p_i to 0

In this case, we will get rid of the largest cost

since $(1-p_i)\cdot q_i \cdot c_i = (1-1)\cdot q_i \cdot c_i = 0$

After that we will get the smallest cost

$\therefore$ we should set p_i that corespond to Max q_i to 1 and rest of p_i to 0, then we will minimize the expect cost

2-7 Same as 2-3, if we make a cost of one of site greater than cost of other site then defender will likely to defend this Max cost site.

Thus, we should make every site to have same value of cost. Depend on the C, if C is greater, then our q_i should be smaller

So that our $C_N \cdot q_N$ for every site will be the same or as close as possible.

$$q_1 \cdot C_1 \approx q_2 \cdot C_2 \approx \cdots \approx q_i \cdot C_i$$

Otherwise, if we make our expect cost for one of the site to be larger, then defender will more likely to protect that site. And we might get nothing or cost smaller than $q_1 \cdot C_1 = q_2 \cdot C_2 = \cdots = q_i \cdot C_i$

2.8 To redo 2.5. in defender perspective

in attacker view:
$$E(x) = \sum_{i=1}^{N} (1-p_i)\cdot q_i \cdot c_i$$

in defender view:
$$E(x) = \sum_{i=1}^{N} (1-p_i)\cdot q_i \cdot c_i$$

To redo 2.6 in defender perspective

in attacker view: we set our p_i that corespond to Max q_i to 1, and rest of p_i to 0.

in defender view: insted of $c_N \cdot q_N$, it will equal to $(c_N \cdot (1-p_N))$

After we find the largest $(c_N \cdot (1-p_N))$ we will set p_i to 1, and rest to be 0.

To redo 2.7 in defender perspective

in attacker view: the $c_N \cdot q_N$ for every site should be as close as possible

$$q_1 \cdot c_1 \approx q_2 \cdot c_2 \approx \cdots \approx q_i \cdot c_i$$

in defender view: the $c_A \cdot (1-p_A)$ for every site should be as close as possible

$$c_1 \cdot (1-p_1) \approx c_2 \cdot (1-p_2) \approx \cdots \approx c_A (1-p_A)$$

Final Strategy: look at the next page

2.8 To redo 2.5. in defender perspective

in attacker view: $E(x) = \sum_{i=1}^{N} (1 - p_i) \cdot q_i \cdot c_i$

in defender view:

$$E(x) = \sum_{i=1}^{N} (1-p_i) \cdot q_i \cdot c_i$$

To redo 2.6 in defender perspective

in attacker view: we set our pi that corespond to Max qi to 1, and rest of pi to 0.

in defender view: instead of $c_N \cdot q_N$, it will equal to $c_N \cdot (1-p_N)$

After we find the largest $c_N \cdot (1-p_N)$

we will set p_i to 1, and rest to be 0.

To redo 2.7 in defender perspective

in attacker view: the $c_N \cdot q_N$ for every site should be as close as possible

$$q_1 \cdot c_1 \approx q_2 \cdot c_2 \approx \cdots \approx q_i \cdot c_i$$

in defender view: the $c_A \cdot (1-p_A)$ for every site should be as close as possible

$$c_1 \cdot (1-p_1) \approx c_2 \cdot (1-p_2) \approx \cdots \approx c_A (1-p_A)$$

In conclusion: final ~~str~~ strategy will be we should set the average cost for every site to be as close as possible.

Bonus:

1. I would ask people, which four numbers do their PIN has. They can tell me the random sequence of those four numbers. And clearly they are not been ask to give their PIN honestly, since we do not know the sequence of those four numbers. However, based on that we will know how many people use only 1 number PIN, how many people use 2 numbers PIN, how many people use 3 numbers PIN, and how many people use 4 numbers PIN. Also, based on numbers that people tell us, we can list all of the possible outcomes. If people only tell us 1 number, then the possible outcome will be 1. If people only tell us 2 numbers, then the possible outcome will be 2^4. If people only tell us 3 number, then the possible outcome will be 3^4. If people only tell us 4 numbers, then the possible outcome will be 4^4. Based on this data, we can estimate the probabilities people use various PINs with.

Second approach: We knows that on the very old phone, we have the number keyboard that only contain 9 numbers, and each number relate to 3 letters. For example number 2 relate to a, b, c. Thus, we can let people use their PIN (at random sequence) and type a word. Since each number relate to 3 letters, people is not going to give the PIN honestly.

2. Attacker should attack the site with average reward, because if the site has reward greater than average, defender will defend it. if the site has reward lower than average, there is no point of attacking it, since the reward is low
Defender should protect the site with reward greater than average.

If they can negotiate beforehand, they defender should let attacker to attack the site with both relatively high reward, and relatively low cost.