I.

a) Because X follows a discrete probability distribution that describes the probability of X successes (positive cases) in n draws (takes a more extensive PCR test), without replacement, from a finite population (the group that takes basic test) of size N that contains exactly K objects with that feature, where in each draw is either a success (positive) or a failure (negative).

$$P[X=x]=\frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}, \forall \mathrm{x} \in \{\max\{0, \mathrm{n}+\mathrm{K}-\mathrm{N}\}, \ldots, \min\{\mathrm{n}, \mathrm{K}\}\}$$

b) For a population of N, we test them one oy one, when the i -th person is tested positive, i.e. $k_i=1$, we increment K, the count of positive cases within these N people. So $K=\sum_{i=1}^{N} k_i$.
For a population of N, we decide one oy one if choose them into the smaller group of n, when the i-th person is tested positive, i.e., $C_i=1$, we increment n, the size of the smaller group within these N people. So $n=\sum_{i=1}^{N} C_i$.
For a population of N, we decide one oy one if choose them into the smaller group of $n$and test them, when the i-th person is tested positive and he is chosen into smaller group, i.e., $k_iC_i=1$, we increment X, the count of positive cases within the smaller group of size n. So $X=\sum_{i=1}^{N} k_iC_i$.
$\mathbf{E}[n]=\mathbf{E}[\sum_{i=1}^{N} C_i]=\sum_{i=1}^{N}\mathbf{E}[C_i]=N\mathbf{E}[C_i]=n,\ \therefore q=\mathbf{E}[C_i]={}^{n}/_{N}$.
$\mathbf{E}[K]=\mathbf{E}[\sum_{i=1}^{N} k_i]=\sum_{i=1}^{N}\mathbf{E}[k_i]=N\mathbf{E}[k_i]=K,\ \therefore \mathbf{E}[k_i]=\frac{K}{N}$.

Since C_i and k_i are independent, $\mathbf{E}[X]=\mathbf{E}[\sum_{i=1}^{N} C_ik_i]=\sum_{i=1}^{N}\mathbf{E}[C_ik_i]=N\mathbf{E}[C_i]\mathbf{E}[k_i]=K,\ \therefore \mathbf{E}[k_i]=np$.

c) $.K^2=(\sum_{i=1}^{N} k_i)(\sum_{j=1}^{N} k_j)=\sum_{i=1}^{N} {k_i}^2+\sum_{i\neq j} k_ik_j.\ \because \forall i\in\{1,\ldots,N\}, k_i\in\{0,1\}, \therefore\ \forall i\in\{1,\ldots,N\}, k_i={k_i}^2.$
$.\therefore\ K^2=\sum_{i=1}^{N} k_i+\sum_{i\neq j} k_ik_j=K+\sum_{i\neq j} k_ik_j$
$.n^2=(\sum_{i=1}^{N} C_i)(\sum_{j=1}^{N} C_j)=\sum_{i=1}^{N} {C_i}^2+\sum_{i\neq j} C_iC_j.\ \because \forall i\in\{1,\ldots,N\}, C_i\in\{0,1\}, \therefore\ \forall i\in\{1,\ldots,N\}, C_i={C_i}^2.$
$.\therefore\ n^2=\sum_{i=1}^{N} C_i+\sum_{i\neq j} C_iC_j=n+\sum_{i\neq j} C_iC_j$
$X^2=(\sum_{i=1}^{N} k_iC_i)(\sum_{j=1}^{N} k_jC_j)=\sum_{i=1}^{N} {k_i}^2{C_i}^2+\sum_{i\neq j} k_ik_jC_iC_j.\ \because \forall i\in\{1,\ldots,N\}, C_i\in\{0,1\},$
$.\therefore\ \forall i\in\{1,\ldots,N\}, k_iC_i={k_i}^2{C_i}^2. \therefore\ X^2=\sum_{i=1}^{N} k_iC_i+\sum_{i\neq j} k_ik_jC_iC_j=X+\sum_{i\neq j} k_ik_jC_iC_j$
$.\mathbf{E}[X(X-1)]=\sum_{x=0}^{n} x(x-1)\frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}=\sum_{x=2}^{n}\frac{K(K-1)\binom{K-2}{x-2}\binom{N-K}{n-x}}{\frac{N(N-1)}{n(n-1)}\binom{N-2}{n-2}}=\frac{n(n-1)K(K-1)}{N(N-1)}\sum_{x=2}^{n}\frac{\binom{K-2}{x-2}\binom{N-K}{n-x}}{\binom{N-2}{n-2}}=\frac{n(n-1)K(K-1)}{N(N-1)}$
$.=\frac{np(n-1)(pN-1)}{(N-1)}$
$.\ \mathbf{E}[X^2]=\mathbf{E}[X(X-1)]+\mathbf{E}[X]=np+\frac{np(n-1)(pN-1)}{(N-1)}$
$.Var(X)=\mathbf{E}[X^2]-\mathbf{E}[X]^2=np+\frac{np(n-1)(pN-1)}{N-1}-n^2p^2=np\frac{N-1+(n-1)(pN-1)-(N-1)np}{N-1}=\frac{np(1-p)(N-n)}{N-1},$
$.Var(X/n)=\frac{Var(X)}{n^2}=\frac{p(1-p)(N-n)}{n(N-1)}$

d) $\because 0\le p(1-p)\le\frac{1}{2}*\frac{1}{2}=\frac{1}{4}, \therefore Var\left(\frac{X}{n}\right)=p(1-p)\frac{N-n}{n(N-1)}\le\frac{1}{4}\frac{N-n}{n(N-1)}$.
When $N=1322, n=1103, Var(X/n)=\frac{K(N-K)(N-n)}{N^2n(N-1)}=\frac{219K(1322-K)}{2{,}546{,}485{,}692{,}092}$.

e) According to Chebyshev inequality, $P\left[\left|\frac{X}{n}-p\right|\ge c\right]\le\frac{Var(\frac{X}{n})}{c^2}=0.05, \therefore c=\sqrt{20Var(\frac{X}{n})}$

f) $P\big[p\in[A,B]\big]=P\left[\left|\frac{X}{n}-p\right|\le c\right]\ge 0.95\rightarrow A=\frac{X}{n}-c, B=\frac{X}{n}+c.$
$.c=\sqrt{20Var(\frac{X}{n})}\approx 0.02705. \therefore A\approx 0.55409, B\approx 0.60819.$

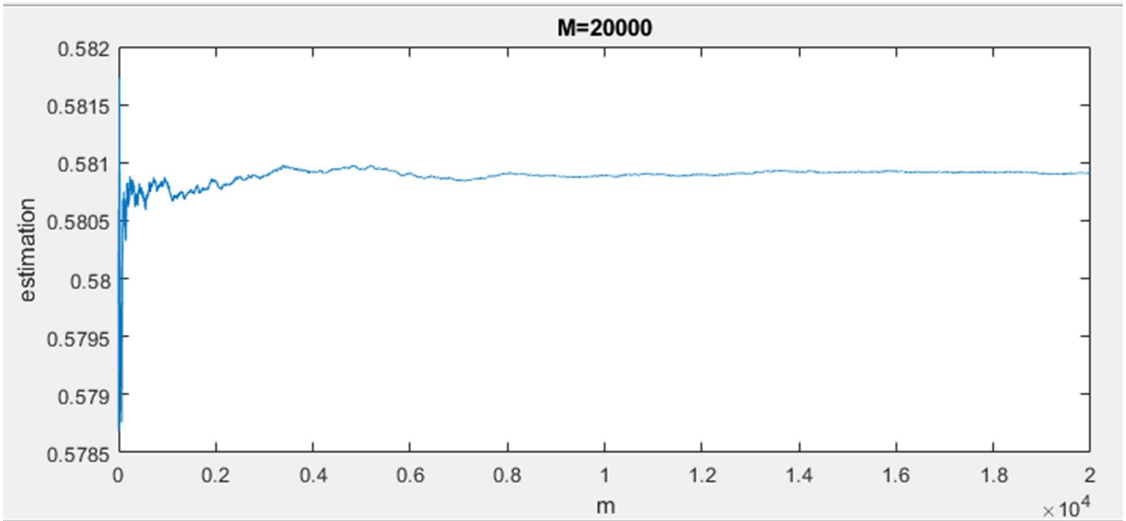
g) We use Markov Inequality on the nonnegative random variable $\left(\frac{X}{n}-p\right)^4$. $\left\{\left|\frac{X}{n}-p\right|\ \ge c\right\}\rightarrow\left\{(\frac{X}{n}-p)^4\ \ge c^4\right\},$
$\therefore P\left[\left|\frac{X}{n}-p\right|\ge c\right]=P\left[(\frac{X}{n}-p)^4\ \ge c^4\right]\le\frac{\mathbf{E}\left[\left(\frac{X}{n}-p\right)^4\right]}{c^4}=\frac{\mathbf{E}\left[\left(\frac{1}{n}\right)^4(X-np)^4\right]}{c^4}=\frac{\mathbf{E}[(X-np)^4]}{n^4c^4}.$

$\frac{\mu_4}{n^4c^4} = 0.05 \rightarrow c = 0.0168266, A \approx 0.56432, B \approx 0.59797$

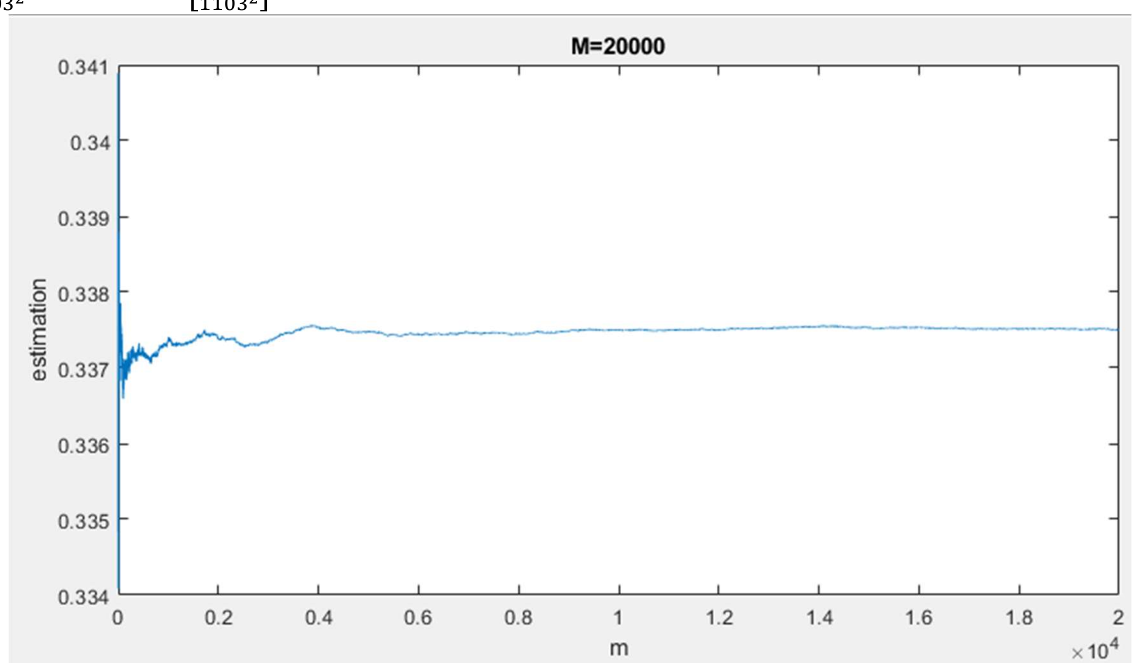
h) Chernoff bound: for any $t \in [0,1-p]$. $\mathbf{P}[X \geq (p+t)n] \leq e^{-D_{KL}(p+t||p)n}$, where $D_{KL}(p+t||p)$gives the relative entropy of $p+t$ and p.
https://arxiv.org/pdf/1507.08298 this paper also provides exponential bounds for hypergeometric distribution with previous studies mentioned.

II

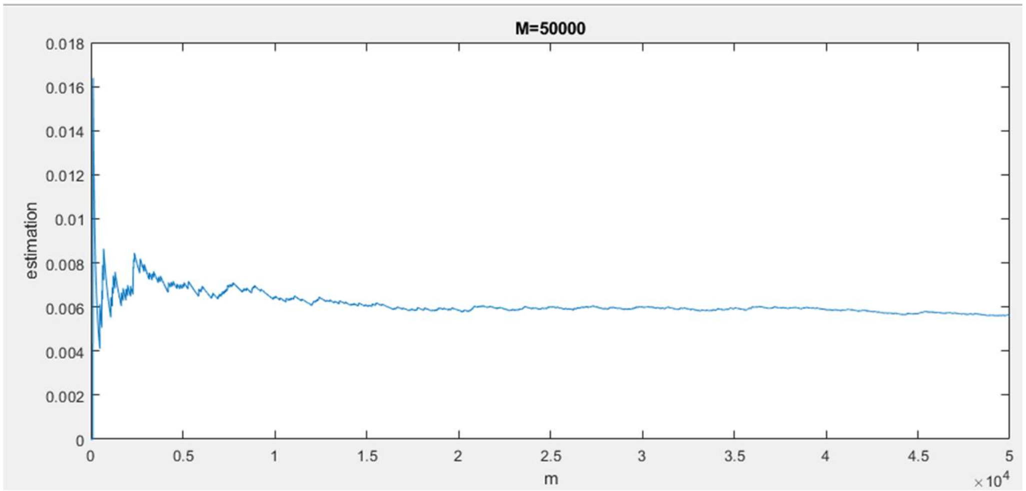
a) Mean of $\frac{1}{j}\sum_{m=1}^{j}\frac{X_m}{1103}$ is 0.5809, variance is 3.1625e-08. $\frac{1}{j}\sum_{m=1}^{j}\frac{X_m}{1103}$ is an unbiased estimator of p. We want large j due to Law of Large Numbers.



b) $\frac{1}{j}\sum_{m=1}^{j}\frac{{X_m}^2}{1103^2} = 0.3375; E\left[\frac{{X_m}^2}{1103^2}\right] = 0.3370$



c) For this problem, the upper bound 0.05 is quite loose. Mean of $\frac{1}{j}\sum_{m=1}^{j}1_{\{|\frac{X_m}{1103}-p|\geq c\}}$ is 0.0061, variance is 3.3888e-07.



M=50000
estimation
m
0.018
0.016
0.014
0.012
0.01
0.008
0.006
0.004
0.002
0
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
$\times 10^4$