

1. In the Chinese Appetizer problem n people are eating n different appetizers arranged on a circular, rotating table. Someone spins the tray so that each person receives a random appetizer. What is the probability that everyone gets the same appetizer as before? How does this compare with the bound obtained using Markov's inequality?

Solution

(i) There are n equally likely orientations of the tray after it stops spinning. Everyone gets the right appetizer in exactly one of these n orientations. Therefore the probability is $1/n$.

(ii) Let X be the number of people that get the right appetizer. In one orientation all n people get the right appetizer and $X=n$. For the other $n-1$ orientations $X=0$. Therefore $E[X] = 1$. By Markov's inequality,

$$P(X \geq n) \leq E[X]/n = 1/n$$

So in this example Markov's inequality is exact.

2. Consider two games. In game A, each time we play we win €2 with probability $2/3$ and lose €1 with probability $1/3$. In game B, each time we play we win €1002 with probability $2/3$ and lose €2001 with probability $1/3$. What is the expected winnings in both games? What is the variance? Using Chebyshev's inequality, compute an upper bound on the probability that after playing 10 rounds of each game the winnings deviate by more than ± 10 from the expected value. Write a Matlab simulation to estimate the probability that make a loss after 10 rounds of each game.

Solution

Let X be the winnings.

Game A.

$$E[X] = 2 \times 2/3 - 1 \times 1/3 = 1.$$

$$\text{Var}(X) = E[X^2] - E[X]^2 = 4 \times 2/3 + (-1)^2 \times 1/3 - 1 = 2$$

Game B.

$$E[X] = 1002 \times 2/3 - 2001 \times 1/3 = 1.$$

$$\text{Var}(X) = E[X^2] - E[X]^2 = 1002^2 \times 2/3 + (-2001)^2 \times 1/3 - 1 = 2004002$$

So both games have the same expected winning, namely €1, but very different variance.

Using Chebyshev, let Y be the winnings after 10 rounds of a game. $E[Y] = 10E[X]$, $\text{Var}(Y) = 10\text{Var}(X)$. So,

$$\text{Game A: } P(|Y - 10| \geq 10) \leq 20/100 = 0.2$$

Game B: $P(|X-10| \geq 10) \leq 20040020/100=200400$. We know the probability can be no more than 1, so this says no more than the trivial fact that $P(|X-10| \geq 10) \leq 1$.

Matlab:

```
P=[]; N=100000;
for i=1:N,
    A=0; B=0;
    for j=1:10,
        r=rand; A = A+2*(r<=2/3)-1*(r>2/3);
        r=rand; B = B+1002*(r<=2/3)-2001*(r>2/3);
    end
    P=[P; A B];
end
sum(P(:,1)<0)/N, sum(P(:,2)<0)/N
```

Probability of a loss after 10 rounds is about 0.02 in game A and about 0.44 in game B.

3. In a poll n turkeys selected independently at random are asked whether they vote for Christmas or not. Let X be the number of yes votes in our sample and use $S=X/n$ as our estimate of the actual fraction of turkeys who like Christmas. Let $X_i=1$ when the i^{th} turkey likes Christmas and 0 otherwise and assume $X_i \sim \text{Ber}(p)$. Using Chebyshev's inequality, how big should n be to ensure that this estimate is within 0.04 of the true fraction at least 95% of the time ?

Solution

Let $X_i=1$ if the i^{th} turkey likes Christmas and 0 otherwise. Then $E[X_i]=1p+0(1-p)=p$ and $\text{Var}(X_i) = E[X_i^2]-E[X_i]^2=1p+0(1-p)-p^2=p(1-p)$.

Therefore $E[X]=np$, $\text{Var}(X)=np(1-p)$ and $E[S]=E[X/n]=E[X]/n=p$, $\text{Var}(S)=\text{Var}(X/n)=\text{Var}(X)/n^2=p(1-p)/n$. Since $0 \leq p \leq 1$ then $0 \leq p(1-p) \leq 1/4$ and so $\text{Var}(S) \leq 1/4n$.

By Chebyshev's inequality,

$$P(|S-p| \geq 0.04) \leq 1/(4 \times 0.04^2 n)$$

and we require the RHS to be less than 0.05 (so that $|S-p| \leq 0.04$ holds 95% of the time). That is we require $1/(4 \times 0.04^2 n) \leq 0.05$ i.e. $n \geq 1/(4 \times 0.04^2 \times 0.05) = 3125$.

4. Suppose you need to design a content delivery network. In a 5 minute interval you get $N=1$ million content requests and each needs to be served from one of your 1000 servers. Suppose each incoming job is assigned uniformly at random to a server. What is the average load on each server ? Use the Chernoff inequality to upper bound the probability that each server has no more than 1250 jobs. Recall that the Chernoff inequality for a Binomial random variable $X \sim \text{Bin}(n,p)$ is $P(X \geq (1+\delta)\mu) \leq \exp(-\mu ((1+\delta)\log(1+\delta) - \delta))$ where $\mu=np$.

Solution

(i) Let's look at one of the servers. Let $X_i=1$ if job i is assigned to the server and 0 otherwise. $P(X_i=1)=1/1000$ since there are 1000 servers and $P(X_i=0)=1- P(X_i=1)$. Therefore $E[X_i]=1/1000$. Let $S=\sum_{i=1}^N X_i$ be the number of jobs assigned to our server. Then $E[S]=N/1000 = 1000$.

(ii) S is a Binomial random variable, $S \sim \text{Bin}(N,p)$. The Chernoff inequality for S is:

$$P(S \geq (1+\delta)\mu) \leq \exp(-\mu ((1+\delta)\log(1+\delta) - \delta))$$

where $\mu=Np=1000$. Our interest is in $(1+\delta)\mu=1250$ i.e. $\delta=1250/\mu-1 = 0.25$ and so

$$P(S \geq 1250) \leq \exp(-1000 (1.25\log 1.25 - 0.25)) \approx \exp(-28) \approx 7 \times 10^{-13}$$

We can therefore view a load of 1250 jobs as effectively being an upper limit on the load a server will experience.