Tutorial - 6

Team Members

111701002 - Ahmed Zaheer Dadarkar
112002004 - Keerthana
111701011 - Devansh Singh Rathore
111801033 - Pellakuru Vamsi

Problem 1: Claim: $P(|R_1 - R_2| \leq 20) \geq 0.45$

Proof:

Define the Random Variables,

$P_1 = R_1 / 100$

$P_2 = R_2 / 100$

Consider the events A and B,

$A = \{|(P_1 - P_2)| > 0.2\}$

$B = \{|P_1 - \mu| > 0.1 \vee |P_2 - \mu| > 0.1\}$

where $\mu = E[P_1] = E[P_2]$

The expected value of both events are the same since both events are the same event, and performed independently of each other.

We claim that $A \rightarrow B$,

for this we assume A to have occurred. Then we have that,

$|P_1 - P_2| > 0.2$

Now, triangle inequality gives us the following,

$|P_1 - \mu| + |P_2 - \mu| \geq |P_1 - P_2| > 0.2$

Hence, we have the following,
$|P_1 - \mu| + |P_2 - \mu| > 0.2$

So at least one of the terms in the sum must be greater than 10,
Hence, we have the following,
$|P_1 - \mu| > 10 \lor |P_2 - \mu| > 10$

Hence, we have shown that $A \rightarrow B$,
So $P(A) \leq P(B)$
By the definition of event B,
$P(B) = P(|P_1 - \mu| > 0.1 \lor |P_2 - \mu| > 0.1)$

Applying the union bound gives us the following,
$\leq P(|R_1 - \mu| > 0.1) + P(|R_2 - \mu| > 0.1)$

And since P_1 and P_2 represent the same event but independently performed, we can write both the terms as the same,

$\leq 2P(|P_1 - \mu| > 0.1)$

We briefly mention the Statement of Hoeffding bounds,
If Y is a sum of n independent Bernoulli Random Variables Y_i's, then we have the following bounds,

$$P(Y > E[Y] + \epsilon) \leq e^{-2\epsilon^2 n}$$

and

$$P(Y < E[Y] - \epsilon) \leq e^{-2\epsilon^2 n}$$

and by applying the union bound to these, we have,

$$P(|Y - E[Y]| > \epsilon) \leq 2e^{-2\epsilon^2 n}$$

We now show that R_1 is a sum of independent 100 Bernoulli Random Variables,

$$P_1 = (X_1 + ... + X_{100}) / 100$$

where

$$X_i = 1 \; if \; i^{th} \; ball \; was \; a \; red \; ball$$

$$= 0 \; if \; i^{th} \; ball \; was \; a \; blue \; ball$$

Hence, P_1 is a sum of independent Bernoulli Random Variables.

Hence, we have ϵ = 0.1, n = 100 in the Hoeffding Bound,

$$P(|P_1 - \mu| > 10) \leq 2e^{-2 * 0.1^2 * 100} = 2e^{-2}$$

Hence,

$$P(A) \leq P(B) \leq 2P(|P_1 - \mu| > 0.1) \leq 4e^{-2}$$

Hence, we have that

$$P(|(P_1 - P_2)| > 0.2) \leq 4e^{-2}$$

Which is the same as,

$$P(|R_1 - R_2)| > 20) \leq 4e^{-2}$$

We consider the complement event,

$$P(|R_1 - R_2)| \leq 20) > 1 - 4e^{-2} = 0.4586 \geq 0.45$$

Hence,

$$P(|R_1 - R_2)| \leq 20) \geq 0.45$$

Hence Proved !

Problem 2:
solution:

Let $X = R_1 / 100$ (This can be written as an average of 100 independent Bernoulli Random Variables, as we had done in Problem 1)

Since we have R red balls and 1000 - R blue balls, the probability of choosing a red ball is as follows,
$p = R / 1000$

We have the following from Hoeffding Bounds,
$P(|X - p| \leq \epsilon) \geq 1 - 2e^{-2\epsilon^2 n}$

We place the value of X and p to get the following,
$P(|R_1 / 100 - R / 1000| \leq \epsilon) \geq 1 - 2e^{-2\epsilon^2 n}$

Hence,
$P(|10R_1 - R| \leq 1000\epsilon) \geq 1 - 2e^{-2\epsilon^2 n}$

Further,
$P(10R_1 - 1000\epsilon \leq R \leq 10R_1 + 1000\epsilon) \geq 1 - 2e^{-2\epsilon^2 n}$

$R_1 = 30$

$P(300 - 1000\epsilon \leq R \leq 300 + 1000\epsilon) \geq 1 - 2e^{-2\epsilon^2 100}$

We want,

$$1 - 2e^{-2\epsilon^2 100} = 0.9$$

We then have,

$$e^{-2\epsilon^2 100} = 0.05$$

We take natural log,

$$-2\epsilon^2 100 = ln(0.05)$$

Then,

$$\epsilon^2 = -ln(0.05) / 200$$

Then

$$\epsilon^2 = 0.01497866136$$

Finally,

$$\epsilon = 0.1223873415$$

Then we have,

$$P(177.6126585 \leq R \leq 422.3873415) \geq 0.9$$

Hence the interval is $[177.6126585, 422.3873415]$!