

CS754 Assignment-4

Saksham Rathi, Ekansh Ravi Shankar, Kshitij Vaidya

Declaration: The work submitted is our own, and we have adhered to the principles of academic honesty while completing and submitting this work. We have not referred to any unauthorized sources, and we have not used generative AI tools for the work submitted here.

Question 2

Part (a)

We are sampling from a set of n coupons, with replacement.

So far, from the previous $j - 1$ trials, we have been getting unique coupons. So, now we have $n - j + 1$ coupons left, out of which we can select any one for uniqueness. Total varieties of coupons = n . So, the probability of picking up a unique coupon is: $\frac{n-j+1}{n}$. (For $j > n$, this probability is 0, because of pigeon-hole principle.)

$q_1 = \frac{n-1+1}{n} = 1$ (This also follows from the fact that the first coupon will be always unique.)

Part (b)

We get our first head in the k^{th} trial, so we get tail in the initial $k - 1$ trials. the probability of getting a head is q , so that of a tail is $1 - q$. Since, each trial is independent of the previous trials, the overall probability is:

$$P = (1 - q)^{k-1}q$$

(where we have also considered the probability of getting a head in the k^{th} trial.)

Part (c)

The probability that $Y = k$ has been calculated in the previous part:

$$P(Y = k) = (1 - q)^{k-1}q$$

To find the expectation of the random variable Y , we need to calculate the following sum:

$$1 \times P(Y = 1) + 2 \times P(Y = 2) + \dots$$

That is, the trial number multiplied by the probability corresponding to that trial.

Writing the above expression in the summation form, we get:

$$\mathbb{E}(Y) = \sum_{k=1}^{\infty} k(1 - q)^{k-1}q$$

Let us call this expression as S , then:

$$S \times (1 - q) = \sum_{k=1}^{\infty} k(1 - q)^k q$$

Subtracting the lower expression from the one above, we get:

$$S - S(1 - q) = (q + 2(1 - q)q + 3(1 - q)^2 q + \dots) - ((1 - q)q + 2(1 - q)^2 q + \dots)$$

We will combine the terms with common powers of $(1 - q)$, so we get:

$$S \times q = q + q(1 - q) + q(1 - q)^2 + \dots + \infty$$

We can use infinite GP expression:

$$S(q) = \frac{q}{1 - (1 - q)}$$

So, we get $\mathbb{E}(Y) = \frac{1}{q}$.

Part (d)

We need to find the variance of Y , which can be written as:

$$\text{Var}(Y) = E(Y^2) - (E(Y))^2$$

The rightmost term, has already been calculated in the previous part. So, let us calculate $E(Y^2)$:

$$\mathbb{E}(Y^2) = \sum_{k=1}^{\infty} k^2 (1 - q)^{k-1} q$$

Let us calculate the following expression:

$$\begin{aligned} & \mathbb{E}(Y^2) - 2\mathbb{E}(Y^2)(1 - q) + \mathbb{E}(Y^2)(1 - q)^2 \\ &= \sum_{k=1}^{\infty} k^2 (1 - q)^{k-1} q - 2 \sum_{k=1}^{\infty} k^2 (1 - q)^k q + \sum_{k=1}^{\infty} k^2 (1 - q)^{k+1} q = \mathbb{E}(Y^2)(1 - (1 - q))^2 = \mathbb{E}(Y^2)(q^2) \end{aligned}$$

Once again, we combine terms with common powers of $1 - q$. So, we get:

$$\mathbb{E}(Y^2)(q^2) = q + 2^2(1 - q)q - 2(1 - q)q + q \sum_{k=1}^{\infty} (1 - q)^{k+1} ((k + 2)^2 - 2(k + 1)^2 + k^2)$$

$$\mathbb{E}(Y^2)(q^2) = q + 2(1 - q)q + 2q \sum_{k=1}^{\infty} (1 - q)^{k+1}$$

$$\mathbb{E}(Y^2)(q^2) = q + 2(1 - q)q + 2q \frac{(1 - q)^2}{1 - (1 - q)} = q + 2q - 2q^2 + 2 - 2q + 2q^2$$

So, we get:

$$\mathbb{E}(Y^2) = \frac{2 - q}{q^2}$$

From this, we get the variance as:

$$\text{Var}(Y) = \frac{2-q}{q^2} - \left(\frac{1}{q}\right)^2 = \frac{1-q}{q^2}$$

Part (e)

Z_n be the random variable denoting the number of trials by which each of the n different coupons were selected at least once. Let t_i be the time to collect the i^{th} coupon after $i-1$ coupons have been collected. Then, $Z_n = t_1 + \dots + t_n$ (we are just summing over the waiting times over every unique coupon obtained).

From part(a), we have that the probability q_j of getting a new coupon on the j^{th} trial is $\frac{n-j+1}{n}$. If we do not get a new coupon, then the probability would remain the same for the next trial (as these many coupons are left to be turned up). So, this is essentially same as tossing a coin and waiting for a head, where head has probability q_j . So, basically t_j is the event Y , which we have used in the previous parts, with probability q_j . Expectation is linear so:

$$\mathbb{E}[Z_n] = \sum_{i=1}^n \mathbb{E}[t_i] = \sum_{i=1}^n \frac{1}{q_i} = \frac{n}{n-i+1} = n \left(\frac{1}{n} + \dots + \frac{1}{n} \right)$$

We can use the inequalities mentioned in the question statement:

$$\mathbb{E}[Z_n] = n \left(\log n + \gamma + \mathcal{O}\left(\frac{1}{n}\right) \right)$$

Again, we know that t_j random variables are independent from each other (each trial itself is independent), so we can get the following expression of variance:

$$\begin{aligned} \text{Var}(Z_n) &= \sum_{i=1}^n \text{Var}(t_i) = \sum_{i=1}^n \frac{1-q_i}{q_i^2} = \sum_{i=1}^n \frac{1}{q_i^2} - \frac{1}{q_i} \leq \sum_{i=1}^n \frac{1}{q_i^2} = \sum_{i=1}^n \frac{n^2}{(n-i+1)^2} \\ \text{Var}(Z_n) &\leq n^2 \sum_{i=1}^n \frac{1}{i^2} \end{aligned}$$

We can again use one of the inequalities mentioned in the question:

$$\text{Var}(Z_n) \leq n^2 \sum_{i=1}^{\infty} \frac{1}{i^2} \leq \frac{n^2 \pi^2}{6}$$

Thus, we have got an upper bound on the variance of Z_n . (If we want an even stronger bound, then we can use the expectation approximation in the variance calculation.)

Part (f)

Markov inequality states that:

$$\mathbb{P}(Z_n \geq t) \leq \frac{\mathbb{E}(Z_n)}{t}$$

We can use the expectation approximation calculated in the previous part, and get the following:

$$\mathbb{P}(Z_n \geq t) \leq \frac{n \left(\log n + \gamma + \mathcal{O}\left(\frac{1}{n}\right) \right)}{t}$$

Thus we have found an upper bound using Markov's inequality.

Part (g)

Since, we do not have a mod operator, we need to use one-sided Chebyshev inequality (also famously known as Cantelli's inequality), which states that (σ is the variance of the random variable):

$$\mathbb{P}(X - \mathbb{E}(X) \geq \lambda) \leq \frac{\sigma^2}{\sigma^2 + \lambda^2}$$

Making the expression compatible to our case, we want:

$$\mathbb{P}(Z_n \geq (t - \mathbb{E}(Z_n)) + \mathbb{E}(Z_n))$$

Putting the λ expression so obtained in the right hand side, we get

$$\frac{\sigma^2}{\sigma^2 + \lambda^2} = \frac{1}{1 + \frac{\lambda^2}{\sigma^2}}$$

Now, we also know that $\sigma^2 = \text{Var}(Z_n) \leq \frac{n^2 \pi^2}{6}$ from one of the previous parts. Therefore:

$$\mathbb{P}(Z_n \geq t) \leq \frac{1}{1 + \frac{(t - \mathbb{E}(Z_n))^2}{\frac{n^2 \pi^2}{6}}}$$

We can put the expectation expression in this, and simplify it further. We could have used double sided Chebyshev inequality too, however the bound would have been loose in that case.