Probability and Computation 1920: Homework Assessment

Submit by 2pm Monday 27th Jan via moodle or at student reception

Question 1. *Let $X_1, X_2, \ldots,$ be independent Bernoulli random variables of parameter p.*

(i) Prove that $T = \min\{i \geq 1 : X_i = 1\}$ has geometric distribution. Find its mean and variance.

(ii) Prove that $N = \sum_{i=1}^{n} X_i$ has Binomial Distribution of parameters (n, p). Find its mean and variance.

Question 2. *Let A, B be independent uniformly random subsets of $[n] := \{1, \ldots, n\}$.*

(i) Find $\mathbf{P}[\, A \subseteq B \,]$.

(ii) What is the distribution of $|A|$?

(iii) How about the distribution of $|A \backslash B|$?

(iv) How can you solve part (i) using part (iii)?

Question 3. *Given $k \in [n]$, a k-subset is a subset of $[n]$ with k elements. Let A be a uniformly random k-subset of $[n]$. Compute the mean and variance of the random variable $X = \sum_{j \in A} j$.*

Question 4. *Let $X \geq 0$ be an integer valued random variable.*

(i) Show that $\mathbf{E}[\, X \,] = \sum_{i=0}^{\infty} \mathbf{P}[\, X > i \,]$.

(ii) Find a similar expression for $\mathbf{E}\left[\, X^2 \,\right]$.

Question 5. *Suppose you are throwing an unbiased, 6-faced dice sequentially until a 6 turns up followed by a 5.*

(i) What is the expected waiting time?

(ii) What happens if you are waiting for a 6 followed by a 6?

(iii) Explain the difference.

Question 6. *In Secret Santa each person from a group of size n is assigned someone to buy a gift for. Secret Santa is successful if each person receives exactly one gift and they don't know who from.*

(i) Is Secret Santa possible for any number of people n?

We want to assign names in secret santa and we get everyone to write their name on paper and put them into a hat. People then take it in turns to draw a name from the hat.

(ii) What is the probability the hat method succeeds (nobody gets their own name) when $n = 4$?

We adapt the hat method: If somebody gets their own name put all names back in the hat and try again.

(iii) What is the probability this new algorithm never terminates?

(iv) How many times do we expect to reset it before we have a successful run?

(v) What is the probability of success with the hat method for $n = 5$?

Question 7. *Let $X_1, \ldots, X_n$ i.i.d. samples from a distribution of interest. We know that $\mathbf{E}[\, X_i \,] = \mu$ and $\mathbf{Var}\,[\, X_i \,] = \sigma^2$ for all i, but we do not know the exact values of μ nor σ^2. We are given the mission to find an estimate $\hat{\mu}$ of the actual mean μ. We want the estimate $\hat{\mu}$ to satisfy the (δ, ε) condition: given ε, we want that $\hat{\mu} \in [\mu - \varepsilon\sigma, \mu + \varepsilon\sigma]$ with probability at least $1 - \delta$. How many data points X_i do we need to build an estimator satisfying the (δ, ε) condition?*

- *In a first attempt we can just deliver* $\hat{\mu} = \frac{\sum_{i=1}^{n} X_i}{n}$*, nevertheless, we cannot guarantee a good behaviour of such estimator, as we do not have enough information to compute a Chernoff Bound for it.*

(i) Prove that with $m = \lceil \frac{10}{\varepsilon^2} \rceil$ *data points, we have that* $\hat{\mu}_m = \left(\sum_{i=1}^{m} X_i\right)/m$ *satisfies the* $(1/10, \varepsilon)$ *condition.*

(ii) Write an algorithm that uses at most $O\left(\frac{\log(\delta^{-1})}{\varepsilon^2}\right)$ *data points to build an estimate of* μ *satisfying the* (δ, ε) *condition.*

Hint.

Q7: *For (ii) consider batches of size* $m = \lceil \frac{10}{\varepsilon^2} \rceil$*. What can you say about more than half of them?*