Stat 201A, Fall 2024: Lab 2

Conceptual review

- When is a sequence of random varibales exchangeable?
- How are the Markov inequality, Chebyshev inequality and the Weak LLN related?

Problem 1

(a) Let X_1, X_2, X_3 be independent $\text{Exp}(\lambda)$ distributed random variables. Find the probability that $\mathbb{P}(X_1 < X_2 < X_3)$.

(b) We deal five cards, one by one, from a standard deck of 52. (Dealing cards from a deck means sampling without replacement.)

  (i) Find the probability that the second card is an ace and the fourth card is a king.

  (ii) Find the probability that the first and the fifth cards are both spades.

  (iii) Find the conditional probability that the second card is a king given that the last two cards are both aces.

Problem 2

Chebyshev's inequality does not always give a better estimate than Markov's inequality. Let X be a random variable with $\mathbb{E}[X] = 2$ and $\mathrm{Var}(X) = 9$. Find the values of t where Markov's inequality gives a better bound for $\mathbb{P}(X > t)$ than Chebyshev's inequality.

Problem 3

A cereal company is performing a promotion, and they have put a toy in each box of cereal they make. There are n different toys altogether and each toy is equally likely to show up in any given box, independently of the other boxes. Let T_n be the number of boxes we need to buy in order to collect the complete set of n toys.

(a) The random variable W_k is the number of boxes we need to open to see a new toy after we have collected k distinct toys. What is the distribution of W_k? Prove that $T_n = 1 + W_1 + W_2 + \cdots + W_{n-1}$.

(b) Calculate the limits $\lim_{n\to\infty} \frac{\mathbb{E}[T_n]}{n \ln(n)}$ and $\lim_{n\to\infty} \frac{\mathrm{Var}[T_n]}{n^2}$.

(c) Use Chebyshev's inequality to estimate $\mathbb{P}(|T_n - \mathbb{E}[T_n]| > \varepsilon n)$.

(d) Show that for any $\varepsilon > 0$ we have

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{T_n}{n \ln(n)} - 1\right| > \varepsilon\right) = 0.$$

This is a weak law of large numbers for the coupon collector's problem.

(e) Using the union bound for the event $E_i^{cn \log(n)}$ that the i-th coupon was not picked in the first $cn \log(n)$ trials, prove that $\mathbb{P}(T_n \geq cn \log(n)) \leq n^{1-c}$.

Problem 4

Cantelli's inequality provides a sharper one-sided bound compared to Chebyshev's inequality. Let X be a random variable with mean μ and variance σ^2, and let $b > 0$.

- Chebyshev's inequality (one-sided):

$$P(X \geq \mu + b) \leq \frac{\sigma^2}{b^2}$$

- Cantelli's inequality:

$$P(X \geq \mu + b) \leq \frac{\sigma^2}{\sigma^2 + b^2}$$

(a) Prove Cantelli's inequality using Markov's inequality.

Hint: Let $Y = X - \mu$. For any $u > 0$,

$$P(Y \geq b) = P(Y + u \geq b + u) \leq P\left((Y + u)^2 \geq (b + u)^2\right).$$

Then use Markov's inequality and find u that minimizes the resulting bound.

(b) Cantelli's inequality implies

$$P(|X - \mu| \geq b) \leq \frac{2\sigma^2}{\sigma^2 + b^2}$$

Comment of the value of this inequality compared to Chebychev's.

Problem 5

The standard Cauchy distribution has the probability density function:

$$f(x) = \frac{1}{\pi\,(1 + x^2)}$$

(a) Does the Weak Law of Large Numbers hold for the Cauchy distribution? Explain why or why not.

(b) Simulate N samples from the standard Cauchy distribution for $N = 10^2, 10^3, 10^4, 10^5$. Calculate the sample averages as N increases. How do the results relate to your explanation in (a)?