

Stochastics
Problem sheet 6 - Concentration theorems
Fall 2022

1. We flip a fair coin 40000 times. Let S be the number of heads. We want to estimate the value y such that $P(S < y)$ is as close to 95% as possible.
 - (a) Estimate y based on the central limit theorem.
 - (b) Using Berry–Esseen, give a bound on the error of the estimate from part (a) and give a lower and upper bound on y .
 - (c) Give an upper bound on y using Hoeffding.
2. We flip a fair coin 40000 times. We want to estimate the probability that we have at least 22000 heads.
 - (a) Try to apply the central limit theorem. What happens?
 - (b) Apply Hoeffding to obtain an upper bound on this probability.
 - (c) Apply Cramér to obtain an upper bound on this probability. (The Cramér rate function of the Bernoulli distribution with parameter p is $I(x) = x \ln \left(\frac{x(1-p)}{(1-x)p} \right) + \ln \left(\frac{1-x}{1-p} \right)$.)
3. A water cleaning facility cleans the waste water (sewage) from n factories. For each factory, the maximum daily output is 200 tons of waste water and the average daily output is 100 tons. The capacity of the water cleaning facility is C tons of water per day. We say that there is *overflow* on a given day if the total waste water produced by the factories exceeds the capacity.
 - (a) $n = 100$ and $C = 14000$. Calculate the probability of overflow.
 - (b) Assuming $n = 100$, determine an upper bound on C such that the probability of overflow is at most 10^{-6} .
 - (c) Given $C = 12000$, calculate the maximum number of factories that can be allowed such that the probability of overflow is at most 10^{-6} .
4. A student is participating in a test which has 100 simple choice questions (with 2 possible answers). For each question, she knows the answer with probability $1/2$. If she doesn't know the answer, she picks one of the answers randomly. Estimate the probability that she gives a correct answer to at least 80 questions.
5. * Compute the Cramér rate function of the distribution $\text{POI}(\lambda)$.
6. We toss a fair coin 1000 times. Give a large deviation estimate on the probability that it comes up heads at least 600 times.
7. E-mails arrive at a server according to a Poisson process with a rate of 1 e-mail per minute. Estimate the probability that at least 1500 e-mails arrive during one day.
8. E-mails arrive at a server according to a Poisson process with a rate of 1 e-mail per minute. We want to estimate the probability that during one day, more than 1800 e-mails arrive. Which of the central limit theorem, Cramér and Hoeffding can be used and why? Give an upper bound on the probability.

(Help: $\text{EXP}(\mu)$ has a Cramér rate function $I(x) = \mu x - 1 - \ln(\mu x)$, and $\text{POI}(\lambda)$ has a Cramér rate function $I(x) = x \ln \frac{x}{\lambda} - x + \lambda$ (for $x > 0$).)
9. A power facility provides 14000 (kW) power to nearby households. Households are put into 2 categories:
 - small households have an average electricity need of 2 kW and maximum need of 15 kW,
 - large households have an average electricity need of 4 kW and maximum need of 25 kW.The facility currently serves 3000 small households. Give an upper bound on the number of large households that can be served from the same facility (in addition to the small households) such that the probability of an outage is smaller than 10^{-7} .

10. Jack plays roulette in the casino. In each round, he bets 10 dollars on red. After 200 rounds, he has lost 150 dollars. Should he be suspicious? Estimate the probability that after 200 rounds, he will lose at least 150 dollars. (In roulette, there are 18 red, 18 black and 1 green slots, one of which is selected uniformly at random in each round.)
11. A truck is carrying three types of packages.
 - Small packages have an average weight of 1 kgs and maximum weight of 3 kgs.
 - Medium packages have an average weight of 2 kgs and maximum weight of 5 kgs.
 - Large packages have an average weight of 4 kgs and maximum weight of 10 kgs.

The truck carries 180 small packages, 160 medium packages and 100 large packages. Its maximal capacity is 1500 kgs. Give an upper bound on the probability that the total weight of the packages exceeds the capacity.

12. A test has 6 problems, each worth a maximum of 10 points. The test is taken by 60 prepared and 40 unprepared students. The result of each student for each problem is random and independent from the others. For each prepared student, the expected value of the score for a single problem is 8 points, while for each unprepared student, the expected value of the score for a single problem is 3 points. Give a large deviation estimate for the probability that the average score of all students is over 45 points.
13. The East Siberian–Pacific Ocean oil pipeline collects the production of 700 oil wells in Siberia and forwards the oil to China. Daily production of the oil wells is independent; for a single well, daily production is never below 490 barrels and never more than 1380 barrels. The total average daily production is 560000 barrels.
 - (a) What should be the capacity of the pipeline if we want the probability of overflow to be at most 10^{-10} ? What about 10^{-8} and 10^{-6} ? At what capacity will this probability be 0?
 - (b) We have more detailed information about the wells: 400 wells have daily production between 490 and 1040 barrels and 300 wells have production between 880 and 1380 barrels. Give a better estimate for the capacity. The probability of overflow should be at most 10^{-10} .

(Remark: the pipeline actually exists; search for ESPO pipeline.)

14. In a large country, there are two political parties. 60% of all voters prefers party A and 40% prefers party B. Give a large deviation estimate on the probability that a survey of 500 people shows party B to be stronger.
15. A postal service delivery truck carries n packages. Each package has maximum weight 5 kg; the average package weight is 2 kg. The capacity of the truck is 1000 kg. Set the value of n so that the probability of n packages being overweight is under 10^{-4} .
16. A city has 40000 households. The garbage produced by a single household during one day is never more than 50 litres; its mean is 20 litres, the deviation 10 litres.
 - (a) What should be the daily capacity C of the garbage processing plant in order that the probability of overflow is under 1%?
 - (b) Why is the central limit theorem not applicable when we want a probability of 10^{-8} instead of 1%? Use Hoeffding in this case to give an upper bound on C .

HW5 (Deadline: Nov. 2, 12:15) A local internet service provider has 12000 clients. Based on their subscription and their habits, they can be put into 3 categories:

- novice: average bandwidth usage is 100 Mbps, and never more than 200 Mbps;
- regular: average bandwidth usage is 160 Mbps, and never more than 300 Mbps;
- master: average bandwidth usage is 250 Mbps, but never more than 400 Mbps.

The number of clients in each category is respectively 3500, 6500 and 2000. What should be the total bandwidth provided in order that the probability of the actual demand exceeding the total bandwidth is at most 10^{-6} ? What should be the total bandwidth provided if the probability criterion is 10^{-7} ? And for 10^{-8} ?