Additional Exercises for Introduction to Probability

mj201,tms41

last updated: 20.5.2025, 23:42

Lecture 1 (Introduction to Probability)

1. We have a hash table that has 100 buckets. We add to the table two arbitrary strings that are independently hashed. How many possible ways are there for the strings to be stored in the table?

2. A license plate has 6 places, where the first three are upper case letters A-Z, and the last three places are numeric 0-9. How many such 6-place license plates are possible?

3. Consider a hash table with 100 buckets. 950 strings are hashed and added to the table. a) Is it possible that a bucket in the table contains no entries? b) Is it guaranteed that at least one bucket in the table contains at least two entries? c) Is it guaranteed that at least one bucket in the table contains at least 10 entries? d) Is it guaranteed that at least one bucket in the table contains at least 11 entries?

4. How many ways are there to select 3 books from a set of 6?

5. Old iPhone passcodes were 4-digit. If we can see a fingerprint on the screen of 3 digits (so 1 digit must use used twice), how many distinct passcodes are possible? What is there were only fingerprints only on 2 digits?

6. How many distinct bit strings can be formed from three 0's and two 1's?

7. A company X has 13 new servers that they would like to assign to 3 datacentres, where datacentre A, B and C have 6, 4 and 3 empty server racks, respectively. How many different divisions of the serves are possible?

8. An urn contains 8 red balls and 4 white balls. We draw 2 balls from the urn without replacement.
   a) Assume each draw is equally likely. What is the probability that both balls drawn are red?
   b) Assume red balls have weight r and white balls have weight w, and the probability that a given ball is the next one selected from the urn is its weight divided by the sum of weights of all balls currently in the urn. What is the probability now that both balls drawn are red?

9. There are only two factories that produce light bulbs. The light bulbs produced in factory A work for over 5000 hours in 99% of cases, whereas the ones from the other factory B work for over 5000 hours in only 95% of the cases. Factory A supplies 60% of the market, whereas

factory B supplies the remaining 40% of the market. What is the chance that a purchased light bulb will work for longer than 5000 hours?

10. **Monty Hall Problem**
You are in a game show where there are 3 closed doors in front of you. There is a prize behind one of the doors, all equally likely. You point to one of the doors. The game show host opens either of the other two doors and shows that it does not have a prize. You now get a choice: do you stick with your original door and open it, or should you switch to the other non-open door?
    - You may think you now have 50/50 chance since one door has and one doesn't have the prize behind it. But you would be wrong: let's use Bayes' rule to calculate.
    - Let F_1: prize behind door 1, F_2: prize behind door 2, F_3: prize behind door 3, so $\mathbf{P}[F_i] = \frac{1}{3}$.
    - Without loss of generality, say you pick door 1, and the host opens door 2 **without** the prize behind it (E: open door 2 without a prize).
    - So we want to compare:
        - if we should stick to our choice of door 1: compute probability of a prize being behind door 1 given that door 2 was opened without a prize: $\mathbf{P}[F_1|E]$.
        - if we should switch to door 3: compute probability of a prize being behind door 3 given that door 2 was opened without a prize: $\mathbf{P}[F_3|E]$.

    Given Bayes' rule we have $\mathbf{P}[F_1|E] = \frac{\mathbf{P}[E|F_1]\mathbf{P}[F_1]}{\mathbf{P}[E]}$.
    Compute the probability that the host opens door 2 (when we picked door 1).

Lectures 2, 3 (Random variables, probability mass function, expectation, expectation properties, variance, discrete distributions)

1. You are playing a card game that uses four standard decks of cards. There are 208 card in total. Each deck has 52 cards (13 values with 4 suits each). Cards are only distinguishable based in their suit and value, not which deck they came from.
a) In how many distinct ways can the cards be ordered?
b) You will be dealt the first two cards from the four decks of cards. Cards with values 10, Jack, Queen, King and Ace are considered "good" cards. What is the probability of getting two "good" cards?
c) Over the course of several rounds you observe 100 cards played. Out of the cards played only 15 were "good" cards. You are dealt the next two cards. What is the probability of getting two "good" cards now? You may assume that previously seen cards are not re-dealt.

2. n people go to a party and drop off their hats to a hat-check person. When the party is over, a different hat-check person is on duty, and returns the n hats randomly back to each person. Let X be the random variable representing the number of people who get their own hat back.
a) For $n = 3$, find $E[X]$ by first computing the probability mass function p_X, and then applying the definition of expectation.
b) Find a general formula for $E[X]$, for any positive integer n.

3. Four 6-sided dice are rolled. The dice are fair, so each one has equal probability of producing a value in $\{1, 2, 3, 4, 5, 6\}$. Let X = the minimum of the four values rolled. (It is fine if more

than one of the dice has the minimal value.)
a) What is $P[X \geqslant k]$ as a function of k?
b) What is $E[X]$?
c) Let T be the sum of the values rolled on the four dice. Let S be the sum of the largest three values on the four dice. In other words, $S = T - X$. What is $E[S]$?

4. The lottery works like this: 100 balls numbered 0-99 are placed in an urn, and 5 balls are withdrawn, giving an ordered sequence of 5 numbers. Once balls are drawn they are not replaced before the next are drawn. Citizens buy tickets with 5 numbers of their choosing. The jackpot is awarded for matching all 5 numbers in the right order.
a) When drawing a lottery, how many possible outcomes are there?
b) You decide to play 5 different numbers. You buy several tickets, one for each of permutations of these 5 numbers. How many tickets do you need to buy? What is your chance of winning the jackpot?
c) You can also win a prize if you guess all 5 numbers, but in incorrect order. What is the probability that any ticket could win this prize? What is your probability of winning this prize?

5. A four sided die has sides numbered $1 - 4$. You roll two such dice. Let X be the sum of the two dice.
a) What are the possible values of X?
b) What is the probability mass function of X?
c) What is the expectation of X, $E[X]$?

6. Let X be a randon variable with possible values of 0 and 1. Let $P[X = 0] = 2 \cdot P[X = 1]$ and $E[X] = \frac{1}{3}$.
a) What is the probability mass function of X?
b) What is the variance of of X, $V[X]$?

Lecture 4 (More discrete distributions: Poisson, Geometric, Negative)

1. Cambridge Tigers Korfball (CTK) team has a probability 0.7 to win in a home game, and probability 0.5 to win in an away game. All games are independent. In a season, there are 35 home and 35 away games.
a) What is the probability that CTK win exactly 20 home games?
b) What is the probability that the first home win for CTK is on the fourth home game?
c) CTK plays in a 3-game series in a pattern home-away-home. What is the probability that CTK wins two out of these three games?
d) A new korfball team joins the league and only has probability of 0.05 of winning each game regardless of whether it is home or away. What is the approximation of the probability that this new team wins w season games?

2. A take-away has a footfall of 10 customers per hour. Let us model with by a Poisson process.
a) What is the probability that at least 3 customers come to a take-away over the course of an hour?
b) For how many hours must the take-away be open for the expected number of customers visiting in that time period to be 100? What is the probability that 0 customer come to the take-away during that time period?
c) Suppose that starting at any particular moment in time, the amount of time we must

wait for the next customer to come is a continuous random variable with probability density function $f(x)$ where x is measured in hours:

$$f(x) = \begin{cases} 10e^{-10x} & \text{if } x \geqslant 0, \\ 0 & \text{if } x < 0. \end{cases}$$

What is the probability that we must wait more than 6 minutes (which is 0.1 hours) for the next customer to come to take-away?

3. Consider requests to a web server in 1 second. In the past, server load averages at 2 hits/second. What is the probability of getting less than 5 hits in a second?

4. There are 10 balls in an urn. One is white, 9 are black. Balls are drawn, then replaced. What is the probability that n or more tries are needed to get the white ball?

Lecture 5 (Continuous random variables)

1. Let us toss a fair coin 100 times. Let X be the number of heads. What is the probability that $43 \leqslant X \leqslant 57$? Using continuity correction, approximate the binomial distribution by normal distribution.

2. Two types of cars, A and B, are being tested for fuel consumption. On test, there are 10 cars of type A and 10 cars of type B, and the consumption of fuel by all 20 vehicles is independent. Let A_x be the number of litres of fuel by car x of type A consumed in the test drive. Let B_y be the number of litres of fuel by car y of type B consumed in the same test drive. Let $A_x \sim Pois(4)$ for $1 \leqslant x \leqslant 10$ and $B_y \sim N(5,3)$ for $1 \leqslant y \leqslant 10$. For every test drive we pick (probabilistically) cars that we will monitor for the amount of fuel consumed and generate the statistics for them. For every test drive a car has (independently) a 0.2 probability of being monitored. Let T be the total amount of fuel (in litres) consumed by the monitored cars.
a) Compute $E[T]$.
b) On a particular test drive, let there be 3 type A cars being monitored (no type B). What is $P[T \geqslant 20]$ on that test drive?
c) Now, on another test drive, let there be 3 type B cars being monitored (no type A). What is $P[T \geqslant 20]$ on that test drive?

Lecture 6,7 (Joint and Marginal Distributions)

1. Let X and Y be a pair of random variables with joint distribution function $F_{X,Y} = F$. Prove that for any $a, b, c, d \in \mathbb{R}$ such that $a < b$ and $c < d$,

$$\mathbb{P}\,[\,a \leqslant X \leqslant b, c \leqslant Y \leqslant d\,] = F(b,d) + F(a,c) - F(a,d) - F(b,c).$$

2. Let X and Y be two random variables with joint distribution function

$$F_{X,Y}(x,y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-x-y} & \text{if } x, y \geqslant 0, \\ 0 & \text{otherwise.} \end{cases}$$

Compute the marginal distribution function of X and Y, $F_X(x)$ and $F_Y(y)$, and their density. What can you conclude about the random variables X and Y?

3. Related to the example about an urn containing balls numbered $1, 2, \ldots, N$ (Slide 7), consider instead the process of drawing $n \leqslant N$ balls without replacement from an urn that contains m red balls and $N - m$ blue balls. Compute the marginal distribution of X_i, where $X_i \in \{0, 1\}$ indicates whether the i-th drawn ball is red. What does the result imply for the expected number of red balls drawn?

4. Prove the alternative formula for the covariance, i.e., $\mathbf{Cov}[X, Y] = \mathbb{E}[X \cdot Y] - \mathbb{E}[X] \cdot \mathbb{E}[Y]$ (Slide 6, Lecture 7).

5. Prove the general form of the Variance of Sum Formula (used in Lecture 11): For any random variables $X_1, X_2, \ldots, X_n$:

$$\mathbf{V}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbf{V}[X_i] + 2 \cdot \sum_{i=1}^{n} \sum_{j=i+1}^{n} \mathbf{Cov}[X_i, X_j].$$

6. Let X and Y be two random variables with covariance $\mathbf{Cov}[X, Y]$. How does the covariance change if we instead take $X' := \alpha \cdot X$ and $Y' := \beta \cdot Y$, and consider $\mathbf{Cov}[X', Y']$? (cf. Slide 21)

7. Proof that the correlation coefficient is scaling-invariant (dimension-less) (Slide 12).

8. Complete the proof that the range of the correlation coefficient is in $[-1, 1]$ (Slide 13).

9. Look up the definition of pairwise independence, and construct three random variables X, Y and Z so that any pair of them is pairwise independent, but the three variables are not independent. (Remark: to emphasise the difference between independence and pairwise independence, some sources use the term "mutual independence".)

Lecture 8,9 (Markov, Chebyshev, Weak Law of Large Numbers, Central Limit Theorem

1. Compute the density function of $X_1 + X_2 + X_3$, where the X_i's are independent random variables with a continuous uniform distribution over $[0, 1]$. (Extension: Can you generalise your result to the sum of $n \geqslant 3$ random variables?)

2. Prove Markov's inequality. (Hint: This follows the lines of the proof of Chebyshev's inequality in the lecture notes.)

3. Give a proof of Chebyshev's inequality that employs Markov's inequality.

4. We consider the differences between the Law of Large Numbers and the Central Limit Theorem. (Note that some parts of this question are not easy and go beyond what is covered in this course.)

   a) What are the differences between the Weak Law of Large Numbers, the Strong Law of Large Numbers and the Central Limit Theorem (CLT)?

   b) Why cannot we use the CLT to deduce the Strong Law of Large Numbers?

   c) Demonstrate how the CLT can be used to deduce the Weak Law of Large Numbers.

5. Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. random variables with mean μ and variance σ^2. Applying the CLT to $\sum_{i=1}^{n/2} X_i$, $\sum_{i=n/2+1}^{n} X_i$ and $\sum_{i=1}^{n} X_i$ (after proper scaling and shifting), which property of the standard normal distribution $\mathcal{N}(0,1)$ can you deduce?

6. Consider throwing a fair, six-sided die 1000 times, and let $Y \in \{1, \ldots, 1000\}$ be the number of times a six occurs. Use the Central Limit Theorem to find values a and b such that

$$\mathbb{P}\left[\, 100 \leqslant Y \leqslant 200 \,\right] \approx \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx.$$

7. For the example on multiple-choice exam questions (Slide 25), apply the Central Limit Theorem to $\mathbb{P}\left[\, \sum_{i=1}^{n} R_i \geqslant 5.5 \,\right]$.

8. This question is related to the example loading a container with packets (Slide 19). Also here, we assume that the packets have weights drawn independently from a $Exp(1/2)$ distribution.
   - How large must the capacity of the container be so that we can at least store 40 packets with .99 probability?
   - Optional: Try to explain how this type of application of the CLT differs from the one on Slide 18 and on Slide 19.

9. Argue why the distribution $Cau(2,1)$ has no expectation and no variance.

10. Let $X_1, X_2, \ldots, X_n$ be independent samples from the $Cau(2,1)$ distribution. Give a justification why the average $\overline{X}_n$ does not converge. Hint: Exploit the fact that the sum of independent Cauchy distributions is again a Cauchy distribution.

11. (Exercise on Slide 18) Let $\widetilde{X}_n := \sum_{i=1}^{n} X_i$, where the X_i's are i.i.d. uniform random variables in $\{-1, +1\}$. Using Stirling's approximation for $n!$, conclude that $\mathbf{P}\left[\, \widetilde{X}_n = 0 \,\right] = \Theta(1/\sqrt{n})$ for even integers n.

12. (optional, related to the difference between the Weak Law of Large Numbers and Strong Law of Large Numbers, and the difference between types of convergence) We define an infinite sequence of random variables, $X_1, X_2, \ldots$ as follows. We first generate auxiliary random variables, $Y_0, Y_1, \ldots$ which are independent and $Y_n \sim \mathsf{Uni}\{0, 1, \ldots, 2^n - 1\}$ (i.e., a discrete uniform distribution). Then, we divide the integers into intervals $I_j := [2^j, 2^{j+1} - 1]$ for all $j = 0, 1, \ldots$ For each interval $I_j = [2^j, 2^{j+1} - 1]$, for every $k \in I_j$,

$$X_k = \begin{cases} 1 & \text{if } k = 2^j + Y_j, \\ 0 & \text{otherwise.} \end{cases}$$

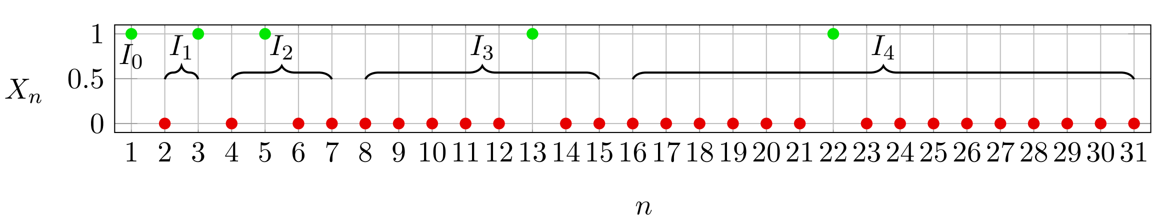

Figure 1: Illustration of the sequence X_n until $n = 31$. In each interval there is exactly one sample equal to 1, while the others are all equal to 0.

a) Prove that for any $\varepsilon > 0$, $\lim_{n\to\infty} \mathbf{P}[\,|X_n| > \varepsilon\,] = 0$.

b) Prove that $\mathbf{P}[\,\lim_{n\to\infty} X_n = 0\,]$ is not equal 1.

Lecture 10,11 (Statistics and Estimators)

1. You model the time that you are spending each week on this course as independent samples from an exponential distribution with unknown parameter λ. After 4 weeks, you record $2, 5, 4, 4$ hours. Estimate $1/\lambda$ by using an unbiased estimator applied to this data set.
2. Compute the Mean-Squared-Error for the sample mean $\overline{X}_n = \frac{1}{n}\cdot\sum_{i=1}^{n} X_i$, where $X_1, X_2, \ldots, X_n$ are i.i.d. samples from some distribution.
3. Let X be a single sample from a Binomial distribution $Bin(n, p)$. In each of the following four cases, decide whether there exists an unbiased estimator and justify your answer.
   a) Assume n is known, but p is unknown and we would like to estimate p.
   b) Assume p is known, but n is unknown and we would like to estimate n.
   c) Assume n and $p \in (0, 1)$ are both unknown, and we would like to estimate $n + p$.
   d) Assume n and p are both unknown, and we would like to estimate $n \cdot p$.
4. Let X be a single sample from a Bernoulli distribution $Ber(p)$, where p is unknown. Can you find an unbiased estimator for p^2? Justify your answer.
5. Let $X_1, X_2, \ldots,$ be a sequence of independent and identically distributed samples from the discrete uniform distribution over $\{1, 2, \ldots, N\}$. Let $Z := \min\{i \geqslant 1 \colon X_i = X_{i+1}\}$. Compute $\mathbb{E}[\,Z\,]$ and $\mathbb{E}\left[\,(Z-N)^2\,\right]$. How can you obtain an unbiased estimator for N?
6. Prove the Mean Squared Error decomposition formula.
7. Let $X_1, X_2, \ldots, X_n$ be n i.i.d. samples from a normal distribution $N(\mu, \sigma^2)$, where μ is unknown but σ is known.
   a) Prove that $Z_1 = X_1$, $Z_2 = \overline{X}_n$ and $Z_3 = (Z_1 + Z_2)/2$ are all unbiased estimators.
   b) Which of the three estimators would you choose?
8. Let $X_1, X_2, \ldots, X_n$ be n i.i.d. random variables with finite μ and finite σ^2. Prove that $\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \leqslant \sum_{i=1}^{n}(X_i - y)^2$, for any $y \in \mathbb{R}$; here, $\overline{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$. What does this imply for the two random variables $\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$ and $\sum_{i=1}^{n}(X_i - \mu)^2$?
9. (see Lecture 11, Slide 6-7; a bit tricky) Consider the estimator $\widetilde{T}_1 := \max(T_1, \max(X_1, X_2, \ldots, X_n))$, where $T_1 := 2\cdot\overline{X}_n - 1$. Using the Bias-Variance Decomposition, what can you conclude about the estimator $\widetilde{T}_1$ in comparison to T_1?
   Hints:
   a) Argue why $\widetilde{T}_1$ is not unbiased.
   b) Try to relate $\mathbf{MSE}[\widetilde{T}_1]$ to $\mathbf{MSE}[T_1]$, differentiating between the cases $T_1 \leqslant \max(X_1, X_2, \ldots, X_n)$ and $T_1 > \max(X_1, X_2, \ldots, X_n)$.
   c) Using the Bias-Variance Decomposition, what does this imply for $\mathbf{V}\left[\,\widetilde{T}_1\,\right]$ versus $\mathbf{V}[\,T_1\,]$?

10. Consider the following modification of the problem of estimating the population size. Instead of sampling without replacement, we sample with replacement. What is the expected number of items we need to sample until we have seen k different IDs?

11. Prove the Cauchy-Schwartz Inequality for random variables X and Y:

$$|\mathbb{E}[X \cdot Y]| \leqslant \sqrt{\mathbb{E}[X^2]\,\mathbb{E}[Y^2]}.$$

12. Let X be a random variable such that $\mu = \mathbb{E}[X] = 1/2$ and $\mathbf{V}[X] = 1$. What can you deduce about $\mathbb{E}[\ln(2X)]$? Hint: Apply Jensen.

13. (a bit tricky). Let X be a random variable with expectation μ, variance σ^2 and median m. Prove that $|\mu - m| \leqslant \sigma$.

14. (Birthday problem) Let X count the number of collisions among k independent samples from a discrete uniform distribution over $\{1, 2, \ldots, n\}$.

    a) What is $\mathbb{E}[X]$?

    b) Prove that $\mathbb{P}[X > 0] \approx 1 - \exp\left(-\binom{k}{2} \cdot \frac{1}{n}\right)$.

    c) Describe how this could be used to obtain an estimator for the population size? (Your estimator does not need to be unbiased) (see also slides of Lecture 11)

15. Consider the algorithm that stops until the first collision is found (Lecture 11). Prove that the expected time until it stops is at most $2 \cdot \sqrt{n}$, where $n = |S|$ is the size of the unknown set. Hint: Consider the realisation of the first $\sqrt{n}$ samples $x_1, x_2, \ldots, x_{\sqrt{n}}$. Assuming these values are all different, consider the first time until the k-th samples satisfies $X_k \in \{x_1, \ldots, x_{\sqrt{n}}\}$ for any $k > \sqrt{n}$.

Lecture 12

1. What is the expected number of local maxima in the secretary problem for n candidates (see Slide 6 of Lecture 12)? (a bit trickier:) Based on this result, suggest an algorithm that outperforms the primitive approach on Slide 7.

2. Prove that if $X_1, X_2, \ldots, X_n$ are n independent samples from the continuous uniform distribution $Uni[0, 1]$, then for $Z := \max\{X_1, X_2, \ldots, X_n\}$ it holds that $\mathbb{E}[Z] = \frac{n}{n+1}$.

3. Assume $n = 4$ in the secretary problem, and for any $1 \leqslant k \leqslant 4$ consider the strategy that accepts the first candidate that is better than the previous $k-1$ candidates. For each possible value of k, compute the probability of hiring the best candidate.

4. (a bit tricky). Consider the secretary problem and let $I_1, I_2, \ldots, I_n$ be the n random variables where $I_j = 1$ if and only the j-th candidate is the best among the first j candidates. Prove that these n random variables are independent.

5. (challenging.) The Parking Problem. You are driving along an infinite street toward your destination, the theatre. There are parking places along the street but most of them are taken. You want to park as close to the theatre as possible but you are not allowed to turn

around. If you see an empty parking place at a distance d before the theatre, should you take it or not?

More specifically, assume you start at point 0 and we have a sequence $X_0, X_1, X_2, \ldots$ indicating whether each parking place $j = 0, 1, 2, \ldots$ is filled or not. Each X_j is an independent Bernoulli random variable with parameter p. By T we denote the (known) place of the theatre. The goal is to minimise $|T - \tau|$, where τ is place where you have parked your car.

6. Recall that in the secretary problem studied in Lecture 12, we assumed that the order of candidates is random (i.e., the ranks of the secretary form a random permutation). Consider now the case where we do not know anything about the order of the candidates (in other words, there might be an adversary who is choosing the order so as to make our strategy perform as poorly as possible). Prove that in this setting, the best possible success probability one can achieve is $\Theta(1/n)$.

Acknowledgements: Thanks to Henry Batchelor and Marcus Handley for spotting typos and errors. Also many thanks to Dimitrios Los for several corrections and additions to the model answers. Lastly, thanks to Justin Lam for sharing his answer to the Parking Problem.