

Additional Exercises for Introduction to Probability (Lectures 7-12)

tms41

last updated: 5.6.2021, 17:55

Lecture 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 $c < d$,

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

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

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

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 $1, 2, \dots, N$ (Slide 7), consider instead the process of drawing $n \leq N$ 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 16).
5. Prove the general form of the Variance of Sum Formula (Slide 19): For any random variables X_1, X_2, \dots, 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 22).
8. Complete the proof that the range of the correlation coefficient is in $[-1, 1]$ (Slide 23).

- 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)

- 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 $n \geq 3$ random variables?)
- Prove Markov's inequality. (Hint: This follows the lines of the proof of Chebyshev's inequality in the lecture notes.)
- Give a proof of Chebyshev's inequality that employs Markov's inequality.
- Updated (Jun 2021. Many thanks to Henry Batchelor and Marcus Handley for some helpful suggestions which prompted this update).** 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 a bit beyond what is covered in this course.
 - What are the differences between the Weak Law of Large Numbers, the Strong Law of Large Numbers and the Central Limit Theorem (CLT)?
 - Can we use the CLT to deduce the Strong Law of Large Numbers?
 - Can we use the CLT to deduce the Weak Law of Large Numbers?
- Let X_1, X_2, \dots, 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 $N(0, 1)$ can you deduce?
- Consider throwing a fair, six-sided die 1000 times, and let $Y \in \{1, \dots, 1000\}$ be the number of times a six occurs. Use the Central Limit Theorem to find values a and b such that

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

- For the example on multiple-choice exam questions (Slide 25), apply the Central Limit Theorem to $\mathbb{P}[\sum_{i=1}^n R_i \geq 5.5]$.
- This question is related to the example loading a container with packets (Slide 26). 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 25 and on Slide 26.
- Argue why the distribution $Cau(2, 1)$ has no expectation and no variance.
- Let X_1, X_2, \dots, X_n be independent samples from the $Cau(2, 1)$ distribution. Give a justification why the average \bar{X}_n does not converge. Hint: Exploit the fact that the sum of independent Cauchy distributions is again a Cauchy distribution.

Lecture 10 (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 ~~(λ)~~ $1/\lambda$ by using an unbiased estimator applied to this data set.
2. Compute the Mean-Squared-Error for the sample mean $\bar{X}_n = \frac{1}{n} \cdot \sum_{i=1}^n X_i$, where X_1, X_2, \dots, 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, \dots , be a sequence of independent and identically distributed samples from the discrete uniform distribution over $\{1, 2, \dots, N\}$. Let $Z := \min \{i \geq 1 : X_i = X_{i+1}\}$. Compute $\mathbb{E}[Z]$ and $\mathbb{E}[(Z - N)^2]$. How can you obtain an unbiased estimator for N ?
6. (related to the Exercise on Slide 23; this is a bit tricky). Prove that for any $n \geq 2$, there is a constant $C > 0$ (independent of n) so that $\mathbb{P}[T_1 < \max\{X_1, X_2, \dots, X_n\}] \geq C$.
7. Prove the Mean Squared Error decomposition formula (Slide 27).
8. Let X_1, X_2, \dots, 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 $Z_1 = X_1$, $Z_2 = \bar{X}_n$ and $Z_3 = (Z_1 + Z_2)/2$ are all unbiased estimators.
 - b) Which of the three estimators would you choose?
9. Consider the following modification of the problem of estimating the population size (Slide 21). 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 ((so that we can to then apply the estimators from the lectures)?
10. Prove the Cauchy-Schwartz Inequality for random variables X and Y :
$$|\mathbb{E}[X \cdot Y]| \leq \sqrt{\mathbb{E}[X^2] \mathbb{E}[Y^2]}.$$
11. Let X be a random variable such that $\mu = \mathbb{E}[X] = 1/2$ and $\mathbf{V}[X] = 1$. What can you deduce about $\mathbb{E}[\ln(2X)]$?
12. (a bit tricky). Let X be a random variable with expectation μ , variance σ^2 and median m . Prove that $|\mu - m| \leq \sigma$.

Lecture 11-12 (Algorithms and Applications)

- (Birthday problem) Let X count the number of collisions among k independent samples from a discrete uniform distribution over $\{1, 2, \dots, n\}$.
 - What is $\mathbb{E}[X]$?
 - Prove that $\mathbb{P}[X > 0] \approx 1 - \exp\left(-\binom{k}{2} \cdot \frac{1}{n}\right)$.
 - Describe how this could be used to obtain an estimator for the population size? (Your estimator does not need to be unbiased) (see Slide 9 of Lecture 11)
- Prove that $\mathbb{E}[Y] = \|p\|_2^2$ (see Slide 9 or 11).
- Prove that the set of random variables $\sigma_{i,j}$ are not pairwise independent (Slide 10/11 of Lecture 11).
- Prove formally that any testing algorithm in this model must have a two-sided error (Slide 13 of Lecture 11).
- 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.
- Prove that if X_1, X_2, \dots, X_n are n independent samples from the continuous uniform distribution $Uni[0, 1]$, then for $Z := \max\{X_1, X_2, \dots, X_n\}$ it holds that $\mathbb{E}[Z] = \frac{n}{n+1}$.
- Assume $n = 4$ in the secretary problem, and for any $1 \leq k \leq 4$ 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.
- (a bit tricky). Consider the secretary problem and let I_1, I_2, \dots, 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.
- (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, \dots indicating whether each parking place $j = 0, 1, 2, \dots$ 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.