11th problem set for Probability and Statistics — 30th April

More on independent vectors

1. (Buffon's Needle) We randomly throw a needle of length ℓ onto an infinite floor. The floor consists of boards, whose edges form parallel lines at a distance $d \geq \ell$. Determine the probability that the needle will cross the edge of any board.

Total probability

2. Determine $P(X < Y)$ for continuous independent random variables $X \sim U(0,2)$ and $Y \sim U(0,1)$. Solve it in several ways:

(a) Directly from a picture.

(b) By considering various possibilities of Y, using the formula (an analogy to the law of total probability)

$$P(X < Y) = \int_0^1 f_Y(y) P(X < Y \mid Y = y) dy.$$

(c) By considering various possibilities of X, using the formula

$$P(X < Y) = \int_0^2 f_X(x) P(X < Y \mid X = x) dx.$$

Convolution

3. Let $X, Y, Z \sim U(0,1)$ be independent random variables.

(a) What is the distribution of $X + Y$? Determine the density in two ways – using the convolution formula and "by the picture".

(b) What is the distribution of $X + Y + Z$? For simplicity, determine the density function only on the interval $[0,1]$.

(c) How to verify the result by sampling?

Applications of Inequalities

4. We roll a die, and if the result is 1 or 2, we get a single point. Let X be the number of points obtained after n (independent) rolls. Estimate the probability that $X \geq n/2$:

(a) Using Markov's inequality.

(b) Using Chebyshev's inequality.

(c) For a specific n, how can this value be determined exactly?

5. A statistician wants to estimate the average height h (in meters) of people in a population using n independent samples $X_1, \ldots, X_n$, selected uniformly at random from all possible people. For the estimate, they use the sample mean $\bar{X}_n = (X_1 + \cdots + X_n)/n$. They estimate that the standard deviation of a single sample is at most 1 meter.

(a) How large should n be chosen so that the standard deviation of $\bar{X}_n$ is at most 1 cm?

(b) For which n does Chebyshev's inequality ensure that $\bar{X}_n$ differs from h by at most 5 cm with a probability of at least 99 %? (That is, $P(|\bar{X}_n - h| \leq 5) \geq 0.99$.)

(c) The statistician notices that all measured people have a height in the interval $(1.4, 2.1)$. How should they adjust the estimate of the standard deviation? How will the answers to the previous questions change?

List of Formulas

- **Convolution:** Let X, Y be continuous independent random variables. Then $S = X + Y$ has density $f_S(s) = \int_{-\infty}^{\infty} f_X(x) f_Y(s-x)dx$.
- **Markov's inequality:** $P(X \geq a\mathbb{E}(X)) \leq \frac{1}{a}$ for $X \geq 0$.
- **Chebyshev's inequality:** $P(|X - \mathbb{E}(X)| \geq t\sigma_X) \leq \frac{1}{t^2}$.

Hints

1: Draw a picture and describe the position of the needle using two random variables (shift and angle). It can be solved as an area problem or using joint density.

3: Use the formula at the end of the first page.

4: What is the distribution of X? Do you remember the formulas for its mean and variance?

5: How to find the variance of the sum of independent random variables? The variance of a multiple of a constant?

Practice Exercises

6. Let $X, Y, Z \sim Exp(\lambda)$ be independent random variables.
(a) What is the distribution of $X + Y$? (b) What is the distribution of $X + Y + Z$?

7. Calculating the area of a circle by random sampling. We generate a random point in a square (both coordinates have a distribution $U(0,1)$). Let X_i be the indicator of the event "the i-th point lies within the inscribed circle".
(a) Determine $\mathbb{E}(X_i)$, $\text{var}(X_i)$.
(b) Let $\bar{X}_n = (X_1 + \cdots + X_n)/n$. Determine $\mathbb{E}(\bar{X}_n)$ and $\text{var}(S_n)$.
(c) For which n do you expect to get the result correct to one decimal place? Two, three, ...?
(d) Another calculation of the area of a circle: $Y_i = \sqrt{1 - U_i^2}$, where $U_i \sim U(0,1)$. Note that $\mathbb{E}(Y_i)$ is the area of a quarter-circle, i.e., $\pi/4$. What is $\text{var}(Y_i)$? What is $\text{var}(_n)$?
(e) Which method is more accurate?

8. We know that the average number of points on a test was 40 (out of 100). Estimate the proportion of students with at least 80 points. Improve the estimate if you know that the standard deviation of the number of points is 10.

9. A meter-long stick is broken into three pieces by one of the methods described below. For each of them, calculate the probability that the three pieces can form a triangle. (Hint: first consider when three positive numbers with sum one are sides of a triangle.)
(a) We randomly select two break points.
(b) We randomly select the first break point. Then we do the same with the piece of stick in the right hand.
(c) We randomly select the first break point. Then we do the same with the larger piece of stick.

10. Randomly select a point from a triangle with vertices at $[0,0]$, $[0,1]$, and $[1,0]$, i.e., the probability of each subset is proportional to its area. Let X, Y be the coordinates of the chosen point.
(a) Find the joint density $f_{X,Y}$.
(b) Find the marginal density f_Y.
(c) Find the conditional density $f_{X|Y}$.
(d) Calculate $\mathbb{E}(X \mid Y = y)$ and use the Law of Total Expectation to calculate $\mathbb{E}(X)$ (using $\mathbb{E}(Y)$).
(e) Calculate $\mathbb{E}(X)$ using the previous part and symmetry.

11th homework, due next week

11. Let $X, Y, Z \sim Exp(\lambda)$ be independent random variables.
(a) What is the distribution of $X + Y$? (b) What is the distribution of $X + Y + Z$?