Continuous Probability, Continuous Distributions, Joint Densities 10

Computer Science Mentors 70

April 9 – 11, 2018

1 Continuous Probability

1.1 Questions

1. Given the following density functions, identify if they are valid random variables. If yes, find the expectation and variance. If not, what rules does the variable violate?

   (a) $f(x) = \begin{cases} \frac{1}{4} & \text{if } x \in [\frac{1}{2}, \frac{9}{2}] \\ 0 & \text{otherwise} \end{cases}$

   Solution: Yes. Is non-negative and area sums to 1. $\mathrm{E}[X] = \frac{5}{2}$ $\mathrm{Var}[X] = \frac{4}{3}$

   (b) $f(x) = \begin{cases} x - \frac{1}{2} & x \in [0, \infty) \\ 0 & \text{otherwise} \end{cases}$

   Solution: No. Has negative values on $(0, \frac{1}{2})$

2. For a discrete random variable X we have $\Pr[X \in [a, b]]$ that we can calculate directly by finding how many points in the probability space fall in the interval and how many total points are in the probability space. How do we find $\Pr[X \in [a, b]]$ for a continuous random variable?

Solution: For a continuous RV with probability density function $f(x)$, the probability that X takes on a value between a and b is the area under the pdf from a to b, which is the integral from a to b of $f(x)$.

3. Are there any values of a, b for the following functions which gives a valid pdf? If not, why? If yes, what values?

   (a) $f(x) = -1,\ a < x < b$

   Solution: No. $f(x) \geq 0$ must be true.

   (b) $f(x) = 0,\ a < x < b$

   Solution: No. $\forall a, b \quad \int_a^b 0dx = 0.$

   (c) $f(x) = 10000,\ a < x < b$

   Solution: Yes, $\int_0^a 10000dx = 1 = 10000a - 0 = 1 \implies a = \frac{1}{10000}$

4. For what values of the parameters are the following functions probability density functions? What is the expectation and variance of the random variable that the function represents?

   (a) $f(x) = \begin{cases} ax & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$

   Solution: For a function to represent a probability density function, we need to have that the integral of the function from negative infinity to positive infinity to equal 1 and for f(x) to be greater than or equal to 0. So we need integral over $(-\infty, \infty)$ $\int_{-\infty}^{\infty} f(x) = 1 = \int_0^1 axdx = \frac{ax^2}{2} \mid_0^1 = 1 \iff \frac{a}{2} - 0 = 1 \iff a = 2$
   For RV Y with pdf = $f(x)$, E$(Y) = \int_{-\infty}^{\infty} x \times f(x)dx = \int_0^1 x \times 2xdx = \frac{2x^3}{3} \mid_0^1 = \frac{2}{3} - 0 = \frac{2}{3}$
   Var$(Y) = \int_{-\infty}^{\infty} x^2 f(x)dx - \text{E}[Y]^2 = \int_0^1 x^2 2xdx - \frac{4}{9} = \int_0^1 2x^3dx - \frac{4}{9} = \frac{x^4}{2} \mid_0^1 = \frac{1}{2} - 0 - \frac{4}{9} = \frac{1}{18}$

   (b) $f(x) = \begin{cases} -2x & \text{if } a < x < b\ (a = 0 \vee b = 0) \\ 0 & \text{otherwise} \end{cases}$

Solution: Again we need $f(x) \geq 0$, so here $a, b \leq 0$, so $b = 0$. Then $\int_a^0 f(x)dx = 1 = \int_a^0 -2xdx = \frac{-2x^2}{2} |_a^0$
$= 0 - (\frac{-2a^2}{2}) = \frac{2a^2}{2} = 1 \iff a^2 = 1 \iff a$
$= \pm 1 \implies a = -1$.

(c) $f(x) = \begin{cases} c & -30 < x < -20 \vee -5 < x < 5 \vee 60 < x < 70 \\ 0 & \text{otherwise} \end{cases}$
Dont worry too much about calculations, but you should be able to set up the equations

Solution: We need $\int_{-\infty}^{\infty} f(x)dx = 1$ and $f(x) \geq 0$. So $c \geq 0$
$\int_{\infty}^{\infty} f(x)dx = 1 = \int_{-30}^{-20} cdx + \int_{-5}^{5} cdx + \int_{60}^{70} cdx = cx |_{-30}^{-20} + cx |_{-5}^{5} + cx |_{60}^{70}$
$= 10c + 10c + 10c = 30c = 1 \implies \frac{1}{30}$

$$\begin{aligned}
E(Y) &= \int_{-\infty}^{\infty} x * f(x)dx \\
&= \int_{-30}^{-20} xcdx + \int_{-5}^{5} xcdx + \int_{60}^{70} xcdx \\
&= \frac{x^2c}{2} |_{-30}^{-20} + \frac{x^2c}{2} |_{-5}^{5} + \frac{x^2c}{2} |_{60}^{70} \\
&= \frac{(-30)^2c}{2} - \frac{(-20)^2c}{2} + \frac{5^2c}{2} - \frac{(-5)^2c}{2} + \frac{70^2c}{2} - \frac{60^2c}{2} \\
&= 900c = \frac{900}{30} = 30
\end{aligned}$$

$$\begin{aligned}
E(Y^2) &= \int_{-\infty}^{\infty} x^2 f(x)dx \\
&= \int_{-20}^{30} x^2cdx + \int_{-5}^{5} x^2cdx + \int_{60}^{70} x^2cdx \\
&= \frac{x^3c}{3} |_{-30}^{20} + \frac{x^3c}{3} |_{-5}^{5} + \frac{x^3c}{3} |_{60}^{70} \\
&= \frac{(-30)^3c}{3} - \frac{(-20)^3c}{3} + \frac{5^3c}{3} - \frac{(-5)^3c}{3} + \frac{70^3c}{3} - \frac{60^3c}{3} \\
&= \frac{108250c}{3} = 1202.77\ldots
\end{aligned}$$

5. Define a continuous random variable R as follows: we pick a random point on a disk of radius 1; the value of R is distance of this point from the center of the disk. We will find the probability density function of this random variable.

   (a) What is (should be) the probability that R is between 0 and $\frac{1}{2}$? Why?

   Solution: $\frac{1}{4}$, because the area of the circle with distance between 0 and $\frac{1}{2}$ is $(\pi(\frac{1}{2})^2 = \frac{\pi}{4})$, and the area of the entire circle is π.

   (b) What is (should be) the probability that R is between a and b, for any $0 \leq a \leq b \leq 1$?

   Solution: The area of the region containing these points is the area of the outer circle minus the area of the inner circle, or $\pi b^2 - \pi a^2 = \pi(b^2 - a^2)$. The probability that a point is within this region, rather than the entire circle, is $\frac{\pi(b^2-a^2)}{\pi} = b^2 - a^2$.

   (c) What is a function $f(x)$, for which $\int_a^b f(x)dx$ satisfies these same probabilities?

   Solution: $f(x) = 2x$ because $\int_a^b f(x)dx = [x^2]_a^b = b^2 - a^2$.

2 Continuous Distributions

2.1 Introduction

Uniform Distribution: $U(a, b)$ This is the distribution that represents an event that randomly happens at any time during an interval of time.

- $f(x) = \frac{1}{b-a}$ for $a \leq x \leq b$
- F(x) = 0 for $x < a$, $\frac{x-a}{b-a}$ for $a < x < b$, 1 for $x > b$
- E$(x) = \frac{a+b}{2}$
- Var$(x) = \frac{1}{12}(b-a)^2$

Exponential Distribution: $Expo(\lambda)$ This is the continuous analogue of the geometric distribution, meaning that this is the distribution of how long it takes for something to happen if it has a rate of occurrence of λ.

- memoryless
- $f(x) = \lambda * e^{-\lambda * x}$
- $F(x) = 1 - e^{-\lambda x}$
- E$(x) = \frac{1}{\lambda}$

Gaussian (Normal) Distribution: $N(\mu, \sigma^2)$

- The CLT states that any unspecified distribution of events will converge to the Gaussian as n increases
- Mean: μ
- Variance: σ^2
- $f(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{(2\pi)}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$

2.2 Questions

1. There are on average 8 office hours in a day. The scores of an exam followed a normal distribution with an average of 50 and standard deviation of 6. If a student waits until an office hour starts, what is the expected value of the sum of the time they wait in hours and their score on the exam?

Solution: Model the beginnings of office hours as 8 points throughout a 24-hour day. Model the student's arrival as 1 point in a 24-hour day. Then in total there are 9 points distributed throughout a 24-hour day. By symmetry, the expected time interval from the student's arrival to the next office hour is $\frac{24}{9}$.

$$\text{E(waiting time)} = \frac{24}{9}$$

$$\text{E(score)} = \text{E}(\mathcal{N}(50, 36)) = 50$$

By linearity of expectation, the sum is $52\frac{2}{3}$.

2. Every day, 100,000,000,000 cars cross the Bay Bridge, following an exponential distribution.

   (a.) What is the expected amount of time between any two cars crossing the bridge?

   Solution: $\frac{1}{100{,}000{,}000{,}000}$ days

   (b.) Given that you havent seen a car cross the bridge for 5 minutes, how long should you expect to wait before the next car crosses?

   Solution: $\frac{1}{100{,}000{,}000{,}000}$ days

3. There are certain jellyfish that dont age called hydra. The chances of them dying is purely due to environmental factors, which well call λ. On average, 2 hydras die within 1 day.

   (a) What is the probability you have to wait at least 5 days for a hydra dies?

   Solution: $\lambda = 2, X \sim Exp(2)$
   $P(X \geq 5) = \int_5^\infty \lambda e^{-\lambda x} dx = \int_5^\infty 2e^{-2x} dx = -e^{-2x}|_5^\infty = e^{-10} = \frac{1}{e^{10}}$

   (b) Let X and Y be two independent discrete random variables. Derive a formula for expressing the distribution of the sum S = X + Y in terms of the distributions of X and of Y.

   Solution: $P(S = m) = \sum_{i=-\infty}^{\infty} P(X = i)P(Y = m - i)$

   (c) Suppose now X and Y are continuous random variables with densities f and g respectively (X,Y still independent). Based on part (a) and your understanding of continuous random variables, give an educated guess for the formula of the density of S = X +Y in terms of f and g.

Solution: $h(t) = \int_{-\infty}^{\infty} f(s)g(t-s)ds$

(d) Use your formula in part (c) to compute the density of S if X and Y have both uniform densities on [0, a].

Solution: Since $f(s)$ is $\frac{1}{a}$ only when $s \in [0, a]$, and 0 everywhere else, we can simplify it to $h(t) = \int_0^a \frac{1}{a} g(t-s)ds$. Consider the case where $t \in [0, a]$. Then $g(t-s)$ will be nonzero (and equal to $\frac{1}{a}$ only when $s \leq t$), so we can further simplify $h(t) = \int_0^t \frac{1}{a}\frac{1}{a} ds = \frac{t}{a^2}$.
Now consider the case where $t \in (a, 2a]$. If so, then $g(t-s)$ is always $\frac{1}{a}$ if $t-s \geq 0$ and $t-s \leq a$ and 0 otherwise. Equivalently, we make sure that $s \leq t$ and $s \geq t-a$. However, recall that we already assumed that $s \leq a$ (or else $f(s) = 0$), so we must restrict ourselves further. Thus, we get $h(t) = \in_{t-a}^{a} \frac{1}{a^2} ds = \frac{1}{a^2}(2a-t)$. So overall, $h(t) = \frac{t}{a^2}$ if $t \in [0, a]$, and $h(t) = 2a-t$ if $t \in (a, 2a]$, and $h(t) = 0$ everywhere else

3 Joint Densities

3.1 Questions

1. Let X, Y be independent uniform (0, 2) random variables.

(a) What is the joint density $f_{X,Y}(x, y)$?

Solution: Since X and Y are independent, we know that $f_{X,Y}(x, y) = f_X(x)f_Y(y)$. Both $f_X(x)$ and $f_Y(y)$ are uniform (0, 2) random variables, so $f_X(x) = \frac{1}{2}$ and $f_Y(y) = \frac{1}{2}$. Thus, $f_{X,Y}(x, y) = f_X(x)f_Y(y) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$.

(b) What is the probability that $X^2 + Y^2 < 4$?

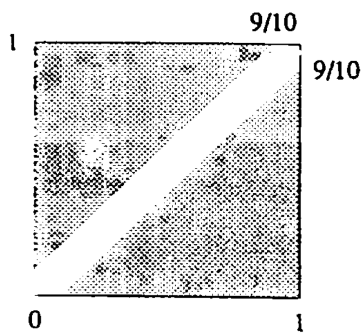
Solution: The domain for the joint density is the square $[0, 2] \times [0, 2]$. Notice that $X^2 + Y^2 = 4$ gives us a circle with radius 2. The event $(X^2 + Y^2 < 4)$ is the region bounded by the square and the circle. Thus, $P(X^2 + Y^2 < 4) = \iint f_{X,Y}(x, y)dxdy =$ area $\times f_{X,Y}(x, y) = \frac{1}{4} \times \pi \times 2^2 \times \frac{1}{4} = \frac{\pi}{4}$

2. A group of students takes CSM mock final. After the exam, each student is told his or her percentile rank among all students taking the exam.

(a) If a student is randomly picked, what is the probability that the student's percentile rank is over 70%?

> **Solution:** Let X be the percentile rank of the student. Then X is a random variable uniformly distributed between 0 and 1. Thus, $P(X > 0.7) = 0.3$.

(b) If two students are picked independently at random, what is the probability that their percentile ranks differ by more than 10%? (Hint: draw a diagram to determine the region of the event)

> **Solution:** Let X be the percentile rank of the first student, and Y be the percentile rank of the second student. Since X and Y are two independent uniform(0, 1) random variable, we know the joint density of X, Y is $f_{X,Y}(x, y) = 1$. The event that {two ranks differ by more than 10%} is equivalent to $|X - Y| > 0.1$, which region is denoted by the shaded area in the following diagram:
>
> 
>
>
> Thus, $P(|X - Y| > 0.1) = 2 \times \frac{1}{2}(\frac{9}{10})^2 = 0.81$.

3. Five students are enrolled in a CSM section which starts at 2pm. They never arrive before 2pm but they all agree to arrive within berkeley time. That is, each person arrives uniformly distributed between 2pm and 2:10pm, independently.

(a) A and B are two students in the section. What is the probability that A arrives at least 2 minutes before B?

> **Solution:** This problem is very similar to the previous problem. The difference is that the region is only one triangle. Thus, P(A arrives at least two minutes before B) = $\frac{1}{2} \times \frac{8}{10} \times \frac{8}{10} = 0.32$.

(b) Find the probability that the event that the first person arrives before 2:02pm, and the last person arrives after 2:08pm.

Solution: *This problem shows that you don't always have to integrate the joint density in order to get the probability.*

Let X be the event that the first person to arrive arrives before 2:02pm and Y be the event that the last person to arrive arrives after 2:08pm. Then,

$$\begin{aligned}
P(X \cap Y) &= 1 - P(\neg(X \cap Y)) \\
&= 1 - P(\neg X \cup \neg Y) \\
&= 1 - P(\neg X) - P(\neg Y) + P(\neg X \cap \neg Y) \\
&= 1 - (\frac{8}{10})^5 - (\frac{8}{10})^5 + (\frac{6}{10})^5
\end{aligned}$$