CONTINUOUS PROBABILITY 11

COMPUTER SCIENCE MENTORS 70

November 12-16, 2018

1 Continuous Probability

1.1 Introduction

When quantities are real-valued or ”continuous”, the number of sample points ω in their probability space Ω is uncountably infinite. So, for any ω, $\mathrm{P}(\omega) = 0$.

As such, when working with continuous random variables, instead of specifying $\mathrm{P}[X = a]$, we specify $\mathrm{P}[a \leq X \leq b]$ for some interval [a, b]:

$$\mathrm{P}[a \leq X \leq b] = \int_a^b f(x)dx$$

The function f is known as the probability density function (PDF). f is a non-negative function over the reals whose total integral is equal to 1 ($\int_{-\infty}^{\infty} f(x) = 1$).

We also define the cumulative distribution function (CDF), the function $\mathrm{P}[X \leq a]$:

$$\mathrm{P}[X \leq a] = \int_{-\infty}^{a} f(x)dx$$

Expectation and variance of continuous R.V:

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} xf(x)dx$$

$$\mathrm{Var}[X] = \int_{-\infty}^{\infty} x^2 f(x)dx - (\int_{-\infty}^{\infty} xf(x)dx)^2$$

1.2 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. 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. $\forall a, b \quad \int_a^b 0 dx = 0$.

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

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

3. 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 infin-

ity 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$(Y) = \int_{-\infty}^{\infty} x^2 f(x)dx - \text{E}[Y]^2 = \int_0^1 x^2 2xdx - \frac{4}{9} = \int_0^1 2x^3 dx - \frac{4}{9} = \frac{x^4}{2} \mid_0^1 = \frac{1}{2} - 0 - \frac{4}{9} = \frac{1}{18}$

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

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

4. 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: $N(\mu, \sigma^2)$

- Mean: μ
- Variance: σ^2
- $f(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{(2\pi)}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$
- The CLT (Central Limit Theorem) states that any unspecified distribution of events will converge to the Gaussian as n increases. For a sequence of iid random variables: $X_1, X_2, ..., X_n$, each with mean μ and variance σ^2,

  $$\frac{X_1 + X_2 + ... + X_n - n\mu}{\sigma\sqrt{n}}$$

  approaches the standard normal distribution $Z \sim N(0, 1)$

2.2 Questions

1. 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

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

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

3. Suppose you have two coins, one that has heads on both sides and another that has tails on both sides. You pick one of the two coins uniformly at random and flip it. You repeat this process 400 times, each time picking one of the two coins uniformly at random and then flipping it, for a total of 400 flips.
Use the CLT to approximate the probability of getting more than 220 heads.

> **Solution:** Let X be the number of heads we get. We have a $\frac{1}{2}$ probability of getting the coin with 2 heads and thus getting a head and we have a $\frac{1}{2}$ probability of getting the coin with 2 tails and not getting heads. Thus $X \sim \mathrm{Bin}(400, \frac{1}{2})$.
> Now, we know that $E[X] = 400 * \frac{1}{2} = 200$ and $\mathrm{Var}(X) = 400 * \frac{1}{2} * (1 - \frac{1}{2}) = 100$, so standard deviation $\sigma = \sqrt{100} = 10$
> Using CLT, we approximate $\frac{X - E[X]}{\sigma} = \frac{X - 200}{10}$ as $Z \sim N(0, 1)$
> Therefore, $P(X > 200) = P(\frac{X-200}{10} > 2) \approx P(Z > 2) = 1 - P(Z \leq 2) \approx 1 - 0.9772$ $= 0.0228$.

3 Joint Densities

3.1 Introduction

A joint density function for two random variables X and Y is a non-negative function f: $\mathbf{R}^2 \to \mathbf{R}$ that has total integral equal to 1 ($\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$).

Joint distribution of X and Y:

$$\text{P}[a \leq X \leq b, c \leq Y \leq d] = \int_c^d \int_a^b f(x,y)dxdy$$

.

3.2 Questions

1. Let X, Y be independent uniform random variables on [0, 2].

   (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 $X^2 + Y^2 < 4$?

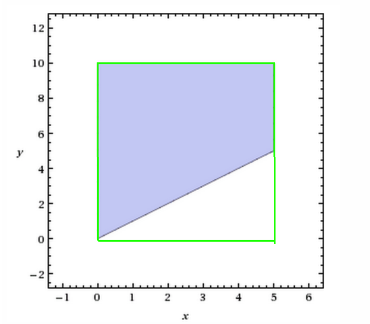
   Solution: The domain for the joint density is the square $[0,2] \times [0,2]$. Notice that $X^2 + Y^2 = 4$ gives us a circle with radius 2. The event $(X^2 + Y^2 < 4)$ is the region bounded by the square and the circle. Thus, $P(X^2 + Y^2 < 4) = \iint f_{X,Y}(x,y)dxdy =$ area $\times f_{X,Y}(x,y) = \frac{1}{4}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

2. Alice and Bob are throwing baseballs and they want to see who can throw a baseball further. The distance Alice throws a baseball is modeled as a uniform distribution between 0 and 5 while the distance Bob throws a baseball is modeled as a uniform distribution between 0 and 10.

   (a) Assuming Alice and Bob throw independently, what is the probability that Bob's throw will be further than Alice's?

   Solution: Let $A \sim U[0,5]$ be the distance of Alice's throw and let $B \sim U[0,10]$ be the distance of Bob's throw.
   We know the density of A is $\frac{1}{5}$ and the density of B is $\frac{1}{10}$

We can draw the following graph where x is Alice's throw and y is Bob's throw.



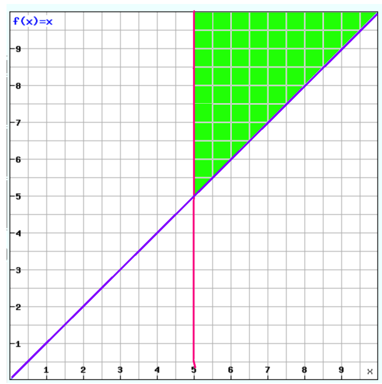
The green outline is the entire sample space while the shaded region is the region of interest. As we can see, the blue region takes up $\frac{3}{4}$ of the total sample space.

A more algebraic approach is to use double integrals.
We fix Alice's throw to be between 0 and 5 and we only consider Bob's throw if it is greater than Alice's.
The joint pdf $f_{A,B} = f_A * f_B = \frac{1}{50}$ as A and B are independent.
Therefore,

$$P(B > A)$$
$$= \int_0^5 \int_a^{10} f_{A,B} db da$$
$$= \int_0^5 \int_a^{10} \frac{1}{50} db da$$
$$= \int_0^5 (\frac{1}{50} b)\Big|_a^{10} da$$
$$= \int_0^5 \frac{1}{5} - \frac{a}{50} da$$
$$= (\frac{1}{5}a - \frac{a^2}{100})\Big|_0^5 = 1 - \frac{25}{100} = \frac{3}{4}$$

(b) Now Alice has improved her throwing abilities and her throwing distance is now also uniform on the interval 0 to 10, which is the same as Bob.
Given that Alice's throw was greater than 5, what is the probability that Bob throws further than her?

Solution: Let $A \sim U[0,10]$ be the distance of Alice's throw and let $B \sim U[0,10]$ be the distance of Bob's throw.
We know the density of A is $\frac{1}{10}$ and the density of B is $\frac{1}{10}$
We want $P(B > A | A > 5)$
We can draw the following graph where x is Alice's throw and y is Bob's throw.



Here let the y-axis be Bob's throw and let the x-axis be Alice's throw.
Here, the purple line represents the region in which Bob throws further than Alice.
The red line down the middle represents the fact that Alice's throw was greater than 5, so we focus our attention on everything that is right of the red line.
The green shaded region represents the probability that Bob's throw is greater than Alice's, conditioned on the fact that Alice's throw was greater than 5.
Therefore, we can see that the green shaded region is $\frac{1}{4}$ of the entire region to the right of the red line.

We can also take an algebraic approach and use double integrals.
Here we fix Alice's throw to be between 5 and 10 as we know she threw for more than 5.
As like the previous problem, we only consider when Bob throws a distance greater than Alice.
The joint pdf $f_{A,B} = f_A * f_B = \frac{1}{100}$ as A and B are independent.
Therefore,
$P(B > A | A > 5) = \frac{P(B>A, A>5)}{P(A>5)} = \frac{P(B>A, A>5)}{\frac{1}{2}}$
$= 2 \int_5^{10} \int_a^{10} f_{A,B} \, db da$
$= 2 \int_5^{10} \int_a^{10} \frac{1}{100} \, db da$

$= 2 \int_5^{10} (\frac{1}{100} b)\Big|_a^{10} da$
$= 2 \int_5^{10} \frac{1}{10} - \frac{a}{100} da$
$= 2(\frac{1}{10}a - \frac{a^2}{200})\Big|_5^{10} = 2((1 - \frac{100}{200}) - (\frac{1}{2} - \frac{25}{200})) = 2(1 - \frac{1}{2} - \frac{1}{2} + \frac{1}{8}) = 2 * \frac{1}{8} = \frac{1}{4}$

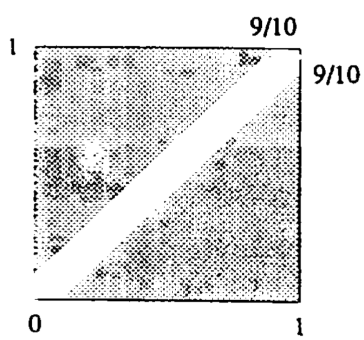
3. A group of students takes CSM mock final. After the exam, each student is told his or her percentile rank among all students taking the exam.

   (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, $P(|X - Y| > 0.1) = 2 \times \frac{1}{2}(\frac{9}{10})^2 = 0.81$.

4. Let X and Y be two independent discrete random variables.

   (a) Derive a formula for expressing the distribution of the sum S = X + Y in terms of the distributions of X and of Y.

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

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

(c) 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