# 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 $\omega$ in their probability space $\Omega$ is uncountably infinite. So, for any $\omega$, $\mathrm{P}(\omega) = 0$.

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

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

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

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

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

# 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 $\lambda$.

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

## 2.2 Questions

1. 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) Why is $R$ not $U(0, 1)$?

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

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

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

   (e) Now say that $R \sim U(0, 1)$. Are you more or less likely to hit closer to the center than before?

2. Every day, 10,000,000 cars cross the Bay Bridge, following an exponential distribution.

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

   (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?

3. There are certain jellyfish that dont age called hydra. The chances of them dying is purely due to environmental factors, which well call $\lambda$. On average, 2 hydras die within 1 day. What is the probability you have to wait at least 5 days for a hydra dies?

# 3 Normal Distribution, CLT

## 3.1 Introduction

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

- Mean: $\mu$
- Variance: $\sigma^2$
- $f(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{(2\pi)}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$
- The CDF of the standard normal ($\mathcal{N}(0,1)$) is often written as $\phi(Z)$. If $X \sim \mathcal{N}(0,1), P(X \leq z) = \phi(z)$
- **TIP:** if $X \sim \mathcal{N}(\mu, \sigma^2)$, define $X' = \frac{X-\mu}{\sigma}$. Now, $X' \sim \mathcal{N}(0,1)$. Note the following are equivalent:
$$P(X < c) = P(X' < \frac{c-\mu}{\sigma}) = \phi(\frac{c-\mu}{\sigma})$$
- **The Central Limit Theorem:** If we have a collection of $n$ independent random variables, $X_1, X_2, ..., X_n$ all with mean = $\mu$ and finite variance $\sigma^2$ Define
$$A_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$
or the average of the random variables. As $n \to \infty$, not only does $A_n \to \mu$, but also $A_n \sim \mathcal{N}(\mathrm{E}[A_n], Var(A_n)) \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$.

## 3.2 Questions

1. 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 $n = 400$ times, each time picking one of the two coins uniformly at random and then flipping it, for a total of $400$ flips. Let $X$ be the number heads you get.

   (a) How is $X$ distributed?

   (b) Why doesn't distribution of $X$ look normal, even as $n \to \infty$? Define a new random variable $T$ in terms of $X$ that does.

   (c) What is the expected value and variance of $T$?

Computer Science Mentors CS70 Spring 2018: Suraj Rampure and Yining Liu, with Anindit Gopalakrishnan, Avni Singhal, Hrishi Dharam, Ming Li, Nick Titterton, Nipun Ramakrishnan, Raymond Chong, Ryan Deng and Seiji Sakiyama

(d) Use the CLT to approximate the probability that you get more than 220 heads. You may find it useful to know that $\phi(2) = 0.9772$

# 4 Joint Densities

## 4.1 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:

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

.

## 4.2 Questions

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

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

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

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?

(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?

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%?

   (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)