Probability and Random Variables (ECE313/ECE317)

Fatima Taousser

Departement of Electrical Engineering and Computer Sciences, UTK
ftaousse@utk.edu

Fall 2023

Probability: Random variables

- **Continuous Random variable.**

  - Probability density function

  - Expectation, variance and their properties

  - Cumulative distribution function

  - Uniform distribution

  - Exponential distribution

  - Normal(Gaussian) distribution.

Probability: Random variables

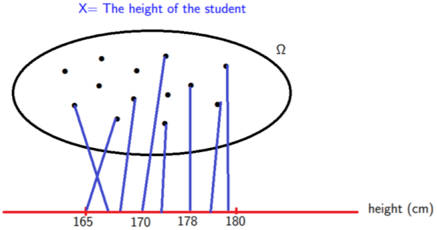
Probability density function (PDF):

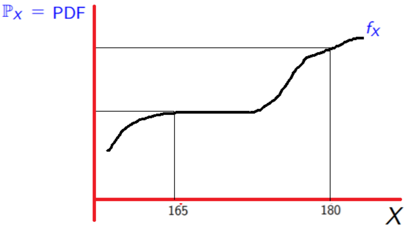
A continuous random variable X takes its values on an interval

Example:

$\Omega = \{$Students of the ECE-313 class$\}$

- Experiment:"Choose a student randomly and measure the height of this student "
- X= The height of the student
- The probability is described by the Probability Density Function (PDF), and denoted by f_X or $\mathbb{P}_X$. The probability can be computed for each range of X in the interval $[160, 180]$.

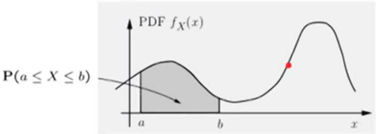




$$\mathbb{P}(160 \leq X \leq 180) = ?$$

Probability: Random variables

▷ Probability density function: Is a function which allows us to represent a probability law in the form of an integral.
$\Rightarrow f_X(.)$ is said a probability density function (PDF) if:
$f_X(.) : \mathbb{R} \to \mathbb{R}$ is positive, inegrable and $\int_{-\infty}^{+\infty} f(x)\, dx = 1$, such that



$$\mathbb{P}(a \le X \le b) = \int_a^b f_X(x)\, dx = \text{The area under the curve from a to b}$$

- A random variable is continuous if it can be described by a PDF

Probability: Random variables

- Properties of the density function:

- $f_X(x) \geq 0$
- $\int_{-\infty}^{+\infty} f_X(x)\ dx = 1$
- $\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)\ dx$
- $\mathbb{P}(X \leq a) = \int_{-\infty}^a f_X(x)\ dx$
- $\mathbb{P}(a \leq X \leq a + \delta) = \int_a^{a+\delta} f_X(x)\ dx \approx f_X(a).\delta,$ where δ is a very small number.
- $\mathbb{P}(X = a) = \int_a^a f_X(x)\ dx = 0$
- $\mathbb{P}(X > a) = 1 - \mathbb{P}(X \leq a)$
- $\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X < b)$. (the probability of a closed interval is equal to the probability of an open interval.)

Probability: Random variables

▷ Cumulative Distribution Function (CDF):

If X is a continuous random variable with PDF $f(x)$ defined on $a \leq x \leq b$, then the cumulative distribution function (CDF), denoted $F(t)$ is given by:

$$F(t) = \mathbb{P}(X \leq t) = \int_a^t f(x)\ dx$$

- The CDF is found by integrating the PDF from the minimum value of X to t.

Probability: Random variables

▷ Expectation and variance of a continuous random variable:

Definition: If X is a continuous random variable with PDF $f(x)$, then the expected value (or the weighted mean) of X is given by

$$\mu = \mathbb{E}(X) = \int_{-\infty}^{+\infty} x\ f(x)\ dx$$

- The formula for the expected value of a continuous random variable is the continuous analog of the expected value of a discrete random variable, where instead of summing over all possible values we integrate.

Definition: The variance of a continuous random variable with PDF $f(x)$, is

$$V(X) = \int_{-\infty}^{+\infty} [x - \mathbb{E}(X)]^2 f(x)\ dx = [\int_{-\infty}^{+\infty} x^2 f(x)\ dx] - \mathbb{E}^2(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X)$$

Definition: The standard deviation of a continuous random variable is

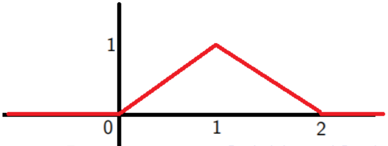
$$\sigma(X) = \sqrt{V(X)}$$

Probability: Random variables

Example: Let the random variable X denote the time that a person waits for an elevator to arrive. Suppose the longest one would need to wait for the elevator is 2 minutes, so that the possible values of X (in minutes) are given by the interval $[0, 2]$. A possible PDF for X is given by

$$f(x) = \begin{cases} x; & 0 \leq x \leq 1 \\ 2 - x; & 1 \leq x \leq 2 \\ 0; & \text{Otherwise} \end{cases}$$

1) $f(x) \geq 0$
2) $\int_{-\infty}^{+\infty} f(x)\ dx = \int_0^1 x\ dx + \int_1^2 (2-x)\ dx = [\frac{1}{2}x^2]_0^1 + [2x - \frac{1}{2}x^2]_1^2 =$
$0.5 + (4 - 2 - 2 + 0.5) = 1$



Probability: Random variables

3) The probability that a person waits less than 30 seconds (or 0.5 minutes) for the elevator to arrive is
$\mathbb{P}(0 \leq X \leq 0.5) = \int_0^{0.5} f(x)\ dx = \int_0^{0.5} x\ dx = [\frac{1}{2}x^2]_0^{0.5} = \frac{1}{8} = 0.125$

- $\mathbb{E}(x) = \displaystyle\int_{-\infty}^{+\infty} xf(x)\ dx = \int_1^2 x(2-x)\ dx + \int_0^1 x.x\ dx = \int_1^2 2x - x^2\ dx + \int_0^1 x^2\ dx$

$$= [x^2 - \frac{1}{3}x^3]_1^2 + [\frac{1}{3}x^3]_0^1 = (4 - \frac{8}{3} - 1 + \frac{1}{3}) + \frac{1}{3} = 1$$

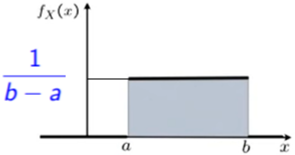
- $V(x) = [\int_{-\infty}^{+\infty} x^2 f(x)\ dx] - \mathbb{E}^2(X) = \int_1^2 x^2(2-x)\ dx + \int_0^1 x^2.x\ dx - 1$

$$= [\int_1^2 2x^2 - x^3\ dx + \int_0^1 x^3\ dx] - 1 = [\frac{2}{3}x^3 - \frac{1}{4}x^4]_1^2 + [\frac{1}{4}x^4]_0^1 - 1 = \frac{7}{6} - 1 = \frac{1}{6}$$

- $\sigma(X) = \sqrt{V(X)} = \dfrac{1}{\sqrt{6}} = 0.408$

Probability: Random variables

- Uniform continuous PDF: with parameters a, b $\rightarrow$ $\mathcal{U}(a, b)$. It is a random variable where all the values are equally likely



$$f_x(x) = \frac{1}{b-a} \Rightarrow \int_a^b f_x(x)\ dx = \int_a^b \frac{1}{b-a}\ dx = \left[\frac{1}{b-a}x\right]_a^b = \frac{b-a}{b-a} = 1$$

$$\mathbb{E}(x) = \int_a^b xf(x)\ dx = \int_a^b x\frac{1}{b-a}\ dx = \left[\frac{1}{2(b-a)}x^2\right]_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}$$

$$V(x) = [\int_a^b x^2 f(x)\ dx] - \mathbb{E}^2(X) = \int_a^b x^2\frac{1}{b-a}\ dx - \left(\frac{b+a}{2}\right)^2 = \left[\frac{1}{3(b-a)}x^3\right]_a^b - \left(\frac{b+a}{2}\right)^2$$

$$= \frac{b^3-a^3}{3(b-a)} - \left(\frac{b+a}{2}\right)^2 = \frac{(b-a)(a^2+ab+b^2)}{3(b-a)} - \frac{(b+a)^2}{4} = \frac{(b-a)^2}{12}$$

Probability: Random variables

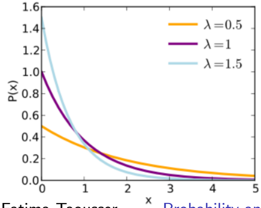
- Exponential Distribution: Is one of the widely used continuous distributions. It is often used to model the time we need to wait before a given event occurs.

- It is often used to answer in probabilistic terms questions such as:

- How long do we need to wait until a customer enters our shop?
- How long will it take before a call center receives the next phone call?
- How long will a piece of machinery work without breaking down?

-If this waiting time is unknown, it is often appropriate to think of it as a random variable having an exponential distribution.

- X:"Is the waiting time for an event to happen" $\rightarrow Exp(\lambda)$, with parameter $\lambda > 0$
- Its PDF is given by: $f_X(x) = \begin{cases} \lambda e^{-\lambda x}; & x > 0 \\ 0; & \text{Otherwise} \end{cases}$



Probability: Random variables

- Exponential Distribution:

- $\int_{-\infty}^{+\infty} f(x)\ dx = \int_0^{+\infty} \lambda e^{-\lambda x}\ dx = [-e^{-\lambda} x]_0^{+\infty} = \underbrace{-e^{-\lambda(+\infty)}}_{=0} + \underbrace{e^{-\lambda(0)}}_{=1} = 1$

- Let us find its CDF. We have

$$F(x) = \int_0^x f(t)\ dt = \int_0^x \lambda e^{-\lambda t}\ dt = [-e^{-\lambda t}]_0^x = -e^{-\lambda x} + e^{-\lambda(0)} = 1 - e^{-\lambda x}$$

- $\mathbb{E}(x) = \displaystyle\int_{-\infty}^{+\infty} xf(x)\ dx = \int_0^{+\infty} x\lambda e^{-\lambda x}\ dx = \lambda \int_0^{+\infty} xe^{-\lambda x}\ dx \ \rightarrow$ Integrate by part

$$= \lambda\left([\frac{-xe^{-\lambda x}}{\lambda}]_0^{+\infty} - \int_0^{+\infty} \frac{-1}{\lambda} e^{-\lambda x}\ dx\right) = \lambda\left(0 - \frac{1}{\lambda^2}[e^{-\lambda x}]_0^{+\infty}\right) = \lambda\left(\frac{1}{\lambda^2}\right) = \frac{1}{\lambda}$$

- $V(x) = [\int_{-\infty}^{+\infty} x^2 f(x)\ dx] - \mathbb{E}^2(X) = \underbrace{\displaystyle\int_0^{+\infty} x^2\lambda e^{-\lambda x}\ dx}_{\text{integrates by part}} - \frac{1}{\lambda^2}$

$$= \lambda\left([x^2(\tfrac{-1}{\lambda}e^{-\lambda x})]_0^{+\infty} + \frac{2}{\lambda}\int_0^{+\infty} xe^{-\lambda x}\ dx\right) - \frac{1}{\lambda^2} = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Continuous Random Variables

Poisson -Vs- Exponential Distributions:

Poisson	**Exponential**
•Event per unit of time	•Time per single event
•Number of cars passing a tollgate in 1H	•Number of hours between car arrivals
•Number of customers arriving in 1 min	•Number of mins between new arrivals

Exponential Distribution is the time between events in a Poisson process.

Continuous Random Variables

Example:

Visitors arrive at a certain website at an average rate of 3 per/hour $\rightarrow$ Poisson with $\lambda = 3$. Find the probability that the next visitor arrives

a) Within 10 mins

b) After 30 mins

c) In exactely 15 mins

Continuous Random Variables

Solution:

Let X: The waiting time for an arriving visitor $\rightarrow X$ follows and **exponential distribution** with the mean time $\mathbb{E}(X) = \dfrac{1H}{3} = \dfrac{60 \text{ mins}}{3} = 20 \Rightarrow 20 = \dfrac{1}{\lambda} \Rightarrow \lambda = \dfrac{1}{20} \rightarrow Exp(\dfrac{1}{20})$

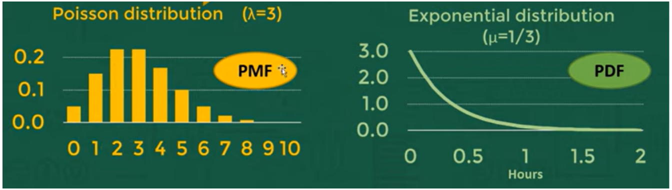
each 60 *mins* $\rightarrow$ 3 person arrive $\rightarrow$ Poisson $\rightarrow \mathbb{E}(X) = \lambda = 3$

each 20 *mins* $\rightarrow$ 1 person arrive $\rightarrow$ Exponential $\rightarrow \mathbb{E}(X) = \dfrac{1}{\lambda} = 20 \; mins = \dfrac{1}{3} \; H$

a) $F_X(x) = \mathbb{P}(X < x) = 1 - e^{-\lambda x} \Rightarrow F_X(10) = \mathbb{P}(X < 10) = 1 - e^{\frac{-10}{20}} = 0.393$

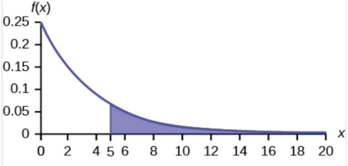
b) $\mathbb{P}(X > 30) = 1 - \mathbb{P}(X < 30) = 1 - F_X(30) = 1 - (1 - e^{\frac{-30}{20}}) = 0.223$

c) $\mathbb{P}(x = 15) = 0$



Continuous Random Variables

Example: Let X = amount of time (in minutes) a postal clerk spends with a customer. The time is known from historical data to have an average amount of time equal to 4 minutes. $\mathbb{E}(X) = 4$ (the average time the clerk spends with a customer) $\Rightarrow \lambda = \frac{1}{4} \Rightarrow X \rightarrow Exp(\frac{1}{4})$.



To calculate the probability that the clerk spend more than 5 min with a customer is given by

$$\mathbb{P}(X > 5) = 1 - \mathbb{P}(X \leq 5) = 1 - F_X(x) = 1 - (1 - e^{-\lambda x}) = 1 - (1 - e^{\frac{-5}{4}}) = 0.286$$

- The probability that a clerk spends 4 to 5 minutes with a randomly selected customer is $\mathbb{P}(4 < X < 5) = F_X(5) - F_X(4) = (1 - e^{\frac{-5}{4}}) - (1 - e^{\frac{-4}{4}}) = 0.7135 - 0.6321 = 0.0814$

Continuous Random Variables

- The proof of the exponential distribution: **(Geometric + Poisson = Exponential)**
- Let X be an exponential random variable with parameter λ.
- We cut-up each unit of time into n-subintervals of time (n can be a large number)
- The probability that the event occurs during a certain sub-interval is $p = \frac{\lambda}{n}$.
- Let $Y \to \mathcal{G}(p)$ be a geometric distribution with parameter $p = \frac{\lambda}{n}$.
- Let $b > 0$ an integer, and we examine $\mathbb{P}(X \leq b)$.
- We have:

$$\mathbb{P}(X \leq b) \approx \mathbb{P}(Y \leq nb) = \sum_{k=1}^{nb} (1-p)^{k-1} p = \sum_{k=1}^{nb} (1 - \frac{\lambda}{n})^{k-1} (\frac{\lambda}{n})$$

$$= (\frac{\lambda}{n}) \sum_{k=1}^{nb} (1 - \frac{\lambda}{n})^{k-1} = (\frac{\lambda}{n}) \left[\frac{1 - (1 - \frac{\lambda}{n})^{nb}}{1 - (1 - \frac{\lambda}{n})} \right] = 1 - \left((1 - \frac{\lambda}{n})^n \right)^b$$

For $n \to \infty$, we get $\mathbb{P}(X \leq b) = 1 - e^{-\lambda b} = F_X(x) = \int_{-\infty}^{b} f(x)\ dx$

$$\Rightarrow \frac{d}{db}[\mathbb{P}(X \leq b)] = \frac{d}{db}[\int_{-\infty}^{b} f(x)\ dx] \Rightarrow \frac{d}{db}[1 - e^{-\lambda b}] = f(b) \Rightarrow \lambda e^{-\lambda b} = f(b).$$

Continuous Random Variables

- Normal Distribution:

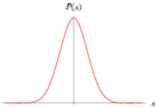
A normal distribution (also called Gaussian distribution) in a variate X with mean μ and variance σ^2: $X \to \mathcal{N}(\mu, \sigma^2)$, is a statistic distribution with probability density function

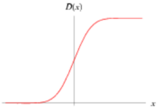
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \quad \mathbb{E}(X) = \mu, \quad V(X) = \sigma^2.$$

- The formula for the cumulative distribution function of the normal distribution is

$$F_X(x) = \int_{-\infty}^{x} f(x)\ dx = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}\ dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-(x-\mu)^2}{2\sigma^2}}\ dx$$

Note that this integral does not exist in a simple closed formula. It is computed numerically.





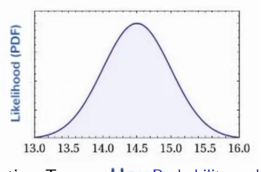
- When the data tends to be around a central value with no bias left or right (symmetry), it is said that it is "Normally Distributed".

Continuous Random Variables

- Example:

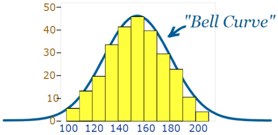
A flight from Sidney to Los Angeles take about 14.5 hours, but it can take from 13 hours to 16 hours $\rightarrow$ Most likely the flight take 14.5 hours.

$$\mu = \mathbb{E} = 14.5, \qquad \sigma = 0.5$$



Continuous Random Variables

- **The Derivation of the Normal Distribution:** Data are said to be normally distributed if their frequency histogram is apporximated by a bell shaped curve.



- The formula of this curve is: $\dfrac{df}{dx} = -k(x-\mu)f(x), \quad k > 0 \;\Rightarrow\; \dfrac{df}{f} = -k(x-\mu)dx$

$$\Rightarrow \int \frac{df}{f} = \int -k(x-\mu)dx \;\Rightarrow \log(f) = \frac{-k}{2}(x-\mu)^2 + C_1 \;\Rightarrow\; f = Ce^{\frac{-k}{2}(x-\mu)^2}$$

- f is a density function if $\int_{-\infty}^{+\infty} f(x)\,dx = 1 \Rightarrow C\int_{-\infty}^{+\infty} e^{\frac{-k}{2}(x-\mu)^2}\,dx = 1.$

Let $v = \sqrt{\frac{k}{2}}(x-\mu) \Rightarrow \begin{cases} x = \sqrt{\frac{2}{k}}v + \mu \\ dx = \sqrt{\frac{2}{k}}\,dv \end{cases} \Rightarrow \int_{-\infty}^{+\infty} f(x)\,dx = \sqrt{\frac{2}{k}}C\int_{-\infty}^{+\infty} e^{-v^2}\,dx$

$= \sqrt{\frac{2\pi}{k}}C = 1 \Rightarrow C = \sqrt{\frac{k}{2\pi}} \Rightarrow f(x) = \sqrt{\frac{k}{2\pi}}e^{\frac{-k}{2}(x-\mu)^2}$ with $k = \sigma^2$

Continuous Random Variables

- **Properties of a normal distribution**

- The curve is symmetric about the center (i.e. around the mean, μ)
$\Rightarrow$ Exactly half of the values are to the left of center and exactly half of the values are to the right.

- The total area under the curve is 1.

Continuous Random Variables

- The standard normal probability distribution $\rightarrow$ $\mathcal{N}(0,1)$

Let the random variable $X \rightarrow \mathcal{N}(\mu, \sigma^2)$ with the normal distribution density $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$.

Let the random variable $Z = \dfrac{X-\mu}{\sigma}$. We have

$$\mathbb{E}(Z) = \mathbb{E}\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}(\mathbb{E}(X) - \mu) = \frac{1}{\sigma}(\mu - \mu) = 0$$

$$V(Z) = V\left(\frac{X-\mu}{\sigma}\right) = (\frac{1}{\sigma})^2 V(X) = (\frac{1}{\sigma})^2 \times \sigma^2 = 1$$

$\Rightarrow$ If $X \rightarrow \mathcal{N}(\mu, \sigma^2)$ then $Z = \dfrac{X-\mu}{\sigma} \rightarrow \mathcal{N}(0,1)$, such that

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$$

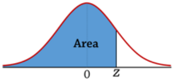
Continuous Random Variables

- Table of the standard normal probability distribution $\rightarrow$ $\mathcal{N}(0,1)$

$Z \rightarrow \mathcal{N}(0,1) \;\Rightarrow\; F_Z(z) = \mathbb{P}(Z \leq z)$ The area under the curve of the density function for all $Z \leq z$

$$F_Z(1.16) = 0.87, \quad F_Z(2.9) = 0.99$$

$$F_Z(-2) = 1 - F_Z(2) = 1 - 0.97$$



- Knowing the standard normal distribution, we can compute the CDF of all normal distribution by considering that $Z = \dfrac{X-\mu}{\sigma}$

Let $X \rightarrow \mathcal{N}(6,4)$ $(\mu = 6, \quad \sigma^2 = 4)$

$$\mathbb{P}(2 \leq X \leq 8) = \mathbb{P}\left(\frac{2-6}{2} \leq \frac{X-6}{2} \leq \frac{8-6}{2}\right)$$

$$= \mathbb{P}\left(-2 \leq \frac{X-6}{2} \leq 1\right) = \mathbb{P}(-2 \leq Z \leq 1)$$

$$\mathbb{P}(Z \leq 1) - \mathbb{P}(Z \geq -2) = \mathbb{P}(Z \leq 1) - [1 - \mathbb{P}(Z \leq 2)]$$

$$= 0.84 - (1 - 0.97) = 0.81$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Continuous Random Variables

- Table of the standard normal probability distribution $\rightarrow$ $\mathcal{N}(0,1)$

$Z \rightarrow \mathcal{N}(0,1) \Rightarrow F_Z(z) = \mathbb{P}(Z \leq z) =$ The area under the curve of the density function for all $Z \leq z$

$$F_Z(1.16) = \mathbb{P}(Z \leq 1.16) = 0.87, \quad F_Z(2.9) = \mathbb{P}(Z \leq 2.9) = 0.99$$

$$F_Z(-2) = \mathbb{P}(Z \leq -2) = 0.0223.$$

- Knowing the standard normal distribution, we can compute the CDF of all normal distribution by considering that $Z = \dfrac{X-\mu}{\sigma}$.

- Let $X \rightarrow \mathcal{N}(6,4) \rightarrow (\mu = 6, \quad \sigma^2 = 4)$

$$\mathbb{P}(2 \leq X \leq 8) = \mathbb{P}(\frac{2-6}{2} \leq \frac{X-6}{2} \leq \frac{8-6}{2})$$

$$= \mathbb{P}(-2 \leq \frac{X-6}{2} \leq 1) = \mathbb{P}(-2 \leq Z \leq 1) = \mathbb{P}(Z \leq 1) - \mathbb{P}(Z \leq -2)$$

$$= 0.84134 - 0.02275 = 0.8186$$

Continuous Random Variables

Example: Most graduate schools of Engineering require applicants for admission to take the Graduate Management Admission Council's GMAT examination. Scores on the GMAT are roughly normally distributed with a mean of 527 and a standard deviation of 112. What is the probability that an individual, scoring above 500 on the GMAT?

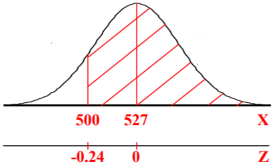
- This is a normal distribution with $\mu = 527$ and $\sigma = 112$, $X \to \mathcal{N}(527, (112)^2)$

$$\Rightarrow f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} = \frac{1}{112\sqrt{2\pi}} e^{\frac{-(x-527)^2}{2(112)^2}}$$

- **Let the standard normal distribution:** $Z = \dfrac{X - 527}{112}$

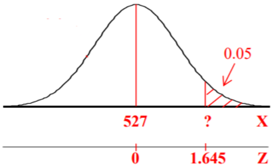
$$\mathbb{P}(X \geq 500) = \mathbb{P}\left(\frac{X - 527}{112} \geq \frac{500 - 527}{112}\right) = \mathbb{P}(Z \geq -0.24)$$

$$= 1 - \mathbb{P}(Z \leq -0.24) = 1 - 0.40517 = 0.5948$$



Continuous Random Variables

- How high must an individual score on the GMAT to score in the highest 5%?



- **Let the standard normal distribution:** $Z = \dfrac{X - 527}{112}$

$$\mathbb{P}(X \geq \alpha) = 0.05 \Rightarrow \mathbb{P}\left(\frac{X - 527}{112} \geq \frac{\alpha - 527}{112}\right) = 0.05 \Rightarrow \mathbb{P}\left(Z \geq \frac{\alpha - 527}{112}\right) = 0.05$$

$$\Rightarrow 1 - \mathbb{P}\left(Z \leq \frac{\alpha - 527}{112}\right) = 0.05 \Rightarrow \mathbb{P}\left(Z \leq \frac{\alpha - 527}{112}\right) = 1 - 0.05 = 0.95 \Rightarrow \frac{\alpha - 527}{112} = 1.645$$

$$\Rightarrow \alpha = 112(1.645) + 527 = 711.24.$$

Continuous Random Variables

Example: The Edwards Theater chain has studied its movie customers to determine how much money they spend on concessions. The study revealed that the spending distribution is approximately normally distributed with a mean of \$4.11 and a standard deviation of \$1.37. What percentage that customers will spend less than \$3.00 on concessions?

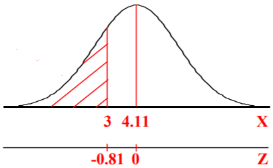
- This is a normal distribution with $\mu = 4.11$ and $\sigma = 1.37$, $X \to \mathcal{N}(4.11, (1.37)^2)$

$$\Rightarrow f_X(x) == \frac{1}{1.37\sqrt{2\pi}} e^{\frac{-(x-4.11)^2}{2(1.37)^2}}$$

- **Let the standard normal distribution:** $Z = \dfrac{X - 4.11}{1.37}$

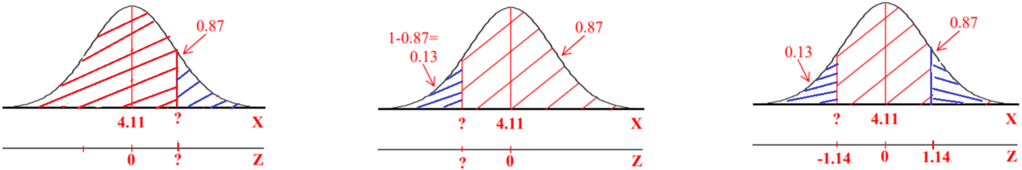
$$\mathbb{P}(X \leq 3) = \mathbb{P}(\frac{X - 4.11}{1.37} \leq \frac{3 - 4.11}{1.37}) = \mathbb{P}(Z \leq -0.81) = 0.20897$$

$$\Rightarrow \mathbb{P}(X < 3) \approx 20.9\%$$



Continuous Random Variables

- What spending amount corresponds to the top 87 % ?



$$\mathbb{P}(X > \alpha) = 0.87 \;\Rightarrow\; \mathbb{P}(X \leq \alpha) = 1 - 0.87 = 0.13$$

$$\Rightarrow \; \mathbb{P}(\frac{X - 4.11}{1.37} \leq \frac{\alpha - 4.11}{1.37}) = 0.13 \;\Rightarrow \mathbb{P}(Z \leq \frac{\alpha - 4.11}{1.37}) = 0.13$$

$$\Rightarrow \; \left(\frac{\alpha - 4.11}{1.37}\right) = -1.13 \;\Rightarrow\; \alpha = (-1.13 \times 1.37) + 4.11 = 2.56$$

$$\Rightarrow \; X \geq \$2.56$$

Continuous Random Variables

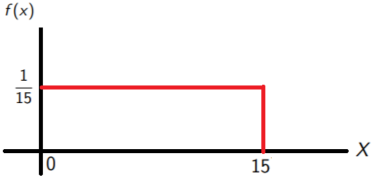
- Table of Continuous Distributions:

Distribution	range X	PDF $f_X(x)$	$\mathbb{E}(X)$	$V(X)$
Uniform $\mathcal{U}([a,b])$	$[a,b]$	$f_X(x) = \frac{1}{b-a}$	$\frac{b+a}{2}$	$\frac{(b-a)^2}{12}$
Exponential $EXP(\lambda)$	$[0,+\infty[$	$f_X(x) = \begin{cases} \lambda e^{-\lambda x}; & x > 0 \\ 0; & \text{Otherwise} \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normal $\mathcal{N}(\mu,\sigma^2)$	$]\text{-}\infty,+\infty[$	$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
Standard Normal $\mathcal{N}(0,1)$	$]\text{-}\infty,+\infty[$	$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$	0	1

Training-Continuous Random Variables

Problem1: Let the uniform distribution $X \to \mathcal{U}(0, 15)$

1) Compute the expectation.
2) Compute the cumulative distribution function (CDF).
3) What is the probability that $X > 7.5$.
4) What is the probability that $X > 5$.
5) What is the probability that $X < 6.5$.
6) What is the probability that $4 < X < 10$?



Training-Continuous Random Variables

- Solution:

$$f(x) = \frac{1}{15}, \quad \text{for} \quad 0 \leq x \leq 15$$

1) $\mathbb{E}(X) = \int_{-\infty}^{+\infty} xf(x)\ dx = \int_0^{15} \frac{1}{15}x\ dx = \frac{1}{15}\int_0^{15} x\ dx = \frac{1}{15}[\frac{1}{2}x^2]_0^{15} = \frac{15}{2} = 7.5$

2) $F(x) = \int_{-\infty}^{x} f(t)\ dt = \int_0^x \frac{1}{15}\ dt = \frac{1}{15}[t]_0^x = \frac{1}{15}x.$

3) $\mathbb{P}(X > 7.5) = 0.5 \ \rightarrow$ We can compute it using the CDF:

$$\mathbb{P}(X > 7.5) = 1 - \mathbb{P}(X \leq 7.5) = 1 - F(7.5) = 1 - \frac{7.5}{15} = 0.5$$

.

4) $\mathbb{P}(X > 5) = 1 - \mathbb{P}(X \leq 5) = 1 - F(5) = 1 - \dfrac{5}{15} = 0.66.$

Or using the integral: $\mathbb{P}(X > 5) = \int_5^{15} f(x)\ dx = \int_5^{15} \frac{1}{15}\ dx = [\frac{1}{15}x]_5^{15} = 1 - \dfrac{5}{15} = 0.66$

5) $\mathbb{P}(X < 6.5) = F(6.5) = \dfrac{6.5}{15} = 0.43 \rightarrow \mathbb{P}(X < 6.5) = \int_0^{6.5} \frac{1}{15}\ dx = [\frac{1}{15}x]_0^{6.5} = 0.43$

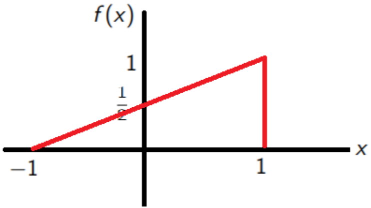
6) $\mathbb{P}(4X < 10) = \mathbb{P}(X < 10) - \mathbb{P}(X \leq 4) = F(10) - F(4) = \dfrac{10}{15} - \dfrac{4}{15} = 0.4.$

Training-Continuous Random Variables

Problem2: Let X be a continuous random variable with the following probability density function:

$$f(x) = \frac{1}{2}(x+1), \qquad -1 < x < 1$$

1) Compute the expectation.
2) Compute the cumulative distribution function (CDF).
3) What is the probability that $X > 0$.
4) What is the 64th percentile of X?



Training-Continuous Random Variables

- Solution:

1) $\mathbb{E}(X) = \int_{-\infty}^{+\infty} xf(x)\; dx = \int_{-1}^{+1} \frac{1}{2}x(x+1)\; dx = \frac{1}{2}\int_{-1}^{+1}(x^2+x)\; dx = \frac{1}{2}[\frac{1}{3}x^3 + \frac{1}{2}x^2]_{-1}^{1} = \frac{1}{3}$

2) $F(x) = \int_{-\infty}^{x} f(t)\; dt = \int_{-1}^{x} \frac{1}{2}(t+1)\; dt = \frac{1}{2}[\frac{1}{2}t^2 + t]_{-1}^{x}$

$$= \frac{1}{2}[(\frac{1}{2}x^2 + x) - (\frac{1}{2}(-1)^2 + (-1))] = \frac{1}{2}(\frac{1}{2}x^2 + x + \frac{1}{2}) = \frac{1}{4}(x+1)^2$$

2) $\mathbb{P}(X > 0) = 1 - \mathbb{P}(X < 0) = 1 - F(0) = 1 - \dfrac{1}{4} = \dfrac{3}{4}$

3) To find the 64th percentile, we just need to set $F(x) = 0.64$

$$F(x) = 0.64 \Rightarrow \frac{1}{4}(x+1)^2 = 0.64 \Rightarrow (x+1)^2 = 4 * 0.64 = 2.56$$

$$\Rightarrow x + 1 = \sqrt{2.56} = \pm 1.6 \Rightarrow x = -2.6 \text{ or } x = 0.6$$

Since $x = -2.6$ is not in the range of $-1 < x < 1$, so $x = 0.6$

$$\Rightarrow \mathbb{P}(X < 0.6) = 0.64$$

Training-Continuous Random Variables

- Problem3:

The time intervals between successive barges passing a certain point on a busy waterway have an exponential distribution with mean 8 minutes.

1) Find the probability that the time interval between two successive barges is less than 5 minutes.
2) Find the probability that the time interval between two successive barges is greater than 3 minutes.
3) Find a time interval t such that we can be 95% sure that the time interval between two successive barges will be greater than t.

Training-Continuous Random Variables

-Solution:

1) Let X = The time interval between successive passing barges $\rightarrow$ with a mean $\mu = 8$ $\Rightarrow \lambda = \frac{1}{8} = 0.125 \Rightarrow X \rightarrow Exp(\frac{1}{8})$ such that,

$$f(x) = \begin{cases} \lambda e^{-\lambda x}; & x > 0 \\ 0; & \text{Otherwise} \end{cases} = \begin{cases} (\frac{1}{8})e^{-\frac{1}{8}x}; & x > 0 \\ 0; & \text{Otherwise} \end{cases}$$

with a cumulative distribution $F(x) = 1 - e^{-\lambda x} = 1 - e^{-(\frac{1}{8})x}$

$$\mathbb{P}(X < 5) = F(5) = 1 - e^{-(\frac{1}{8})5} = 0.46$$

2) $\mathbb{P}(X > 3) = 1 - \mathbb{P}(X \leq 3) = 1 - F(3) = 1 - [1 - e^{-(\frac{1}{8})3}] = 0.68$

3) $\mathbb{P}(X > t) = 0.95 \Rightarrow 1 - \mathbb{P}(X < t) = 1 - F(t) = e^{-(\frac{1}{8})t} = 0.95.$

$$-(\frac{1}{8})t = \log(0.95) \;\Rightarrow\; t = -8\log(0.95) = 0.41 \;\rightarrow \text{That is } 24.6s$$

Training-Continuous Random Variables

Problem4:

X is a normally distributed random variable with mean $\mu = 30$ and standard deviation $\sigma = 4$. Find

1) $\mathbb{P}(x < 40)$

2) $\mathbb{P}(x > 21)$

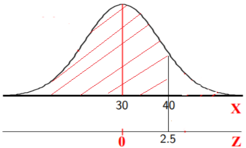
3) $\mathbb{P}(30 < x < 35)$

Solution: $X \to \mathcal{N}(30, (4)^2)$

1) $\mathbb{P}(X < 40) = \mathbb{P}(\frac{X-30}{4} < \frac{40-30}{4}) = \mathbb{P}(Z < \frac{10}{4}) = \mathbb{P}(Z < 2.5)$

- Using the table of the standard normal distribution we get :

$$\mathbb{P}(Z < 2.5) = 0.993 \Rightarrow \mathbb{P}(X < 40) = 0.993$$

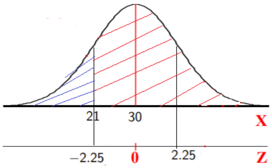


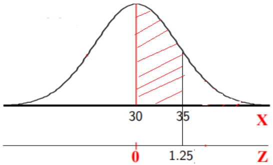
Continuous Random Variables

2) $\mathbb{P}(X > 21) = \mathbb{P}(\frac{X-30}{4} > \frac{21-30}{4}) = \mathbb{P}(Z > \frac{-9}{4}) = \mathbb{P}(Z > -2.25)$

$$= 1 - \mathbb{P}(Z < -2.25) = 0.9878$$

3) $\mathbb{P}(30 < X < 35) = \mathbb{P}(\frac{30-30}{4} < \frac{X-30}{4} < \frac{35-30}{4}) = \mathbb{P}(0 < Z < 1.25)$
$= \mathbb{P}(Z < 1.25) - \mathbb{P}(Z < 0) = 0.8944 - 0.5 = 0.3944 \Rightarrow \mathbb{P}(30 < X < 35) = 0.3944$





Continuous Random Variables

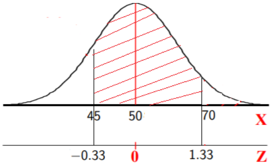
Problem5: For a certain type of computers, the length of time between charges of the battery is normally distributed with a mean of 50 hours and a standard deviation of 15 hours. A person owns one of these computers and wants to know the probability that the length of time will be between 45 and 70 hours.

Solution: X : The length of time charges of the battery $\rightarrow \mathcal{N}(50h, (15h)^2)$

$$\mathbb{P}(45 < X < 70) = \mathbb{P}\left(\frac{45-50}{15} < \frac{X-50}{15} < \frac{70-50}{4}\right) = \mathbb{P}\left(\frac{-5}{15} < Z < \frac{20}{15}\right)$$

$$= \mathbb{P}(-0.33 < Z < 1.33) = \mathbb{P}(Z < 1.33) - \mathbb{P}(Z < -0.33) = 0.9082 - 0.3707 = 0.5375$$

$$\Rightarrow \mathbb{P}(45 < X < 70) = 0.5375$$



Continuous Random Variables

Problem6: The length of life of an instrument produced by a machine has a normal distribution with a mean of 12 months and standard deviation of 2 months. Find the probability that an instrument produced by this machine will last

1) less than 7 months.

2) between 7 and 12 months.

Solution: X : The life time of the instrument $\to \mathcal{N}(12,(2)^2)$

1) $\mathbb{P}(X<7)=\mathbb{P}(\frac{X-12}{2}<\frac{7-12}{2})=\mathbb{P}(Z<\frac{-5}{2})=\mathbb{P}(Z<-2.5)=0.0062$

$$\Rightarrow \mathbb{P}(X<7)=0.0062$$

2) $\mathbb{P}(7<X<12)=\mathbb{P}(\frac{7-12}{2}<\frac{X-12}{2}<\frac{12-12}{2})=\mathbb{P}(-2.5<Z<0)$
$=\mathbb{P}(Z<0)-\mathbb{P}(Z<-2.5)=0.5-0.0062=0.4938$