Contents

Note: Comments/suggestions on these lecture notes are welcome on the e-mail: sukuyd@gmail.com to Dr. Suresh Kumar.

Chapter 4

Continuous Probability Distributions

4.1 Definitions

4.1.1 Continuous Random Variable

A continuous random variable is a variable X that takes all values x in an interval or intervals of real numbers.

For example, if X denotes the lifetime of a person, then it is a continuous random variable because lifetime happens to be continuous, no matter how small or big it is.

4.1.2 Probability Density Function (PDF)

The probability of a continuous random variable taking any single specific value is zero. Instead, probabilities are assigned over intervals using a probability density function (PDF). A function f is called PDF of a continuous random variable X if it satisfies the following conditions:

(i) $f(x) \geq 0$ for all x.

(ii) $P(a \leq X \leq b) = \int_a^b f(x)dx$, i.e., $f(x)$ gives probability of X lying in any given interval $[a, b]$,

(iii) $\int_{-\infty}^{\infty} f(x)dx = 1$, since X will certainly will take one value in its entire sample space $(-\infty, \infty)$. It is also called normalization property of f.

We immediately notice the following points.

1. The PDF $f(x)$ provides probability of X over some interval. $f(x)$ at any value of X is simply a value of the PDF and not the probability. By definition, probability for any particular value of X is 0. It has nothing to do with the PDF $f(x)$.

2. The condition $f(x) \geq 0$ implies that the plot of $y = f(x)$, also called the probability curve, lies on or above x-axis.

3. The condition $\int_{-\infty}^{\infty} f(x)dx = 1$ graphically implies that the total area under the probability curve $y = f(x)$ is 1.

4. $P(a \leq X \leq b) = \int_a^b f(x)dx$ gives the area under the probability curve $y = f(x)$ from $x = a$ to $x = b$ as shown in Figure 4.1

5. $P(a \leq X \leq b) = P(a < X \leq b) = P(a < X < b)$ since $P(X = a) = 0$, $P(X = b) = 0$.

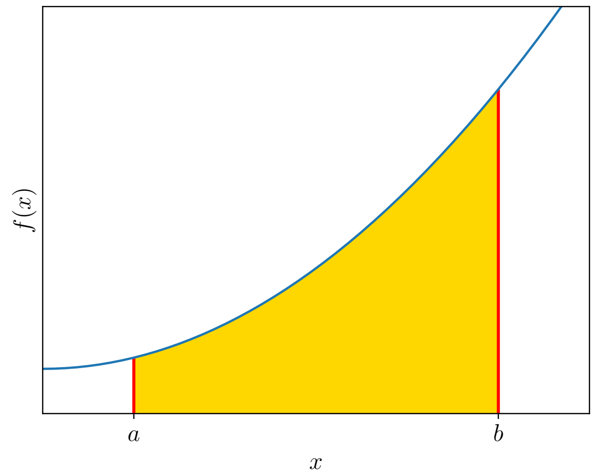

Figure 4.1: The shaded golden region gives the probability $P(a \leq X \leq b) = \int_a^b f(x)dx$.

Ex. Verify whether the function is a PDF of X.

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

Also, evaluate $P(0.2 \leq X \leq 0.6)$.
Sol. We have

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{0} 0\,dt + \int_0^1 2x\,dx + \int_1^{\infty} 0\,dt = \left[x^2\right]_0^1 = (1)^2 - (0)^2 = 1.$$

So given function is PDF of X. Next,

$$P(0.2 \leq X \leq 0.6) = \int_{0.2}^{0.6} 2x\,dx = \left[x^2\right]_{0.2}^{0.6} = (0.6)^2 - (0.2)^2 = 0.36 - 0.04 = 0.32.$$

4.1.3 Cumulative Distribution Function (CDF)

A function F defined by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

is called **cumulative distribution function** (CDF) of X. See Figure 4.2, where the shaded golden region gives the value of $P(X \leq b) = F(b)$.

We notice that
(i) $P(a \leq X \leq b) = \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx = F(b) - F(a)$.
(ii) $F'(x) = f(x)$, provided the differentiation is permissible.

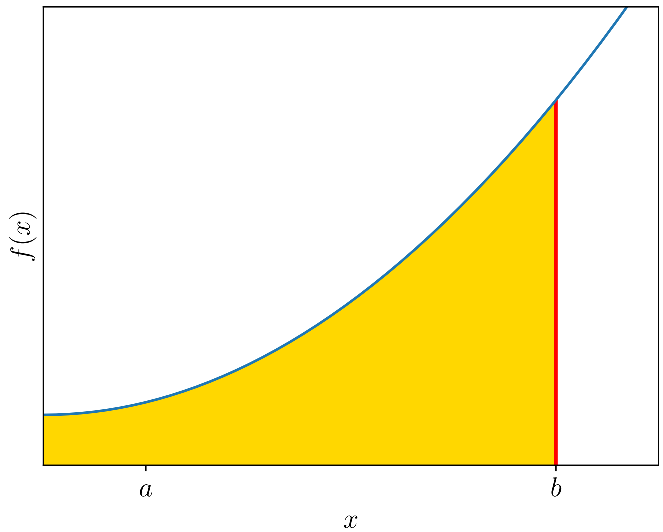

Figure 4.2: The shaded golden region gives the probability $P(X \leq b) = F(b)$.

Ex. The PDF of a continuous random variable X is:

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

Find CDF of X.
Sol. The CDF is given by

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \begin{cases} \int_{-\infty}^{x} 0\,dt & \text{if } x \in (-\infty, 0) \\ \int_{-\infty}^{0} 0\,dt + \int_{0}^{x} 2t\,dt & \text{if } x \in [0,1] \\ \int_{-\infty}^{0} 0\,dt + \int_{0}^{1} 2t\,dt + \int_{1}^{x} 0\,dt & \text{if } x \in (0,\infty) \end{cases} = \begin{cases} 0 & \text{if } x \in (-\infty, 0) \\ x^2 & \text{if } x \in [0,1] \\ 1 & \text{if } x \in (1,\infty) \end{cases}$$

4.1.4 Expectation, MGF, Moments and Variance

The **expectation** of a random variable X having density f is defined as

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

In general, the expectation of $H(X)$, a function of X, is defined as

$$E(H(X)) = \int_{-\infty}^{\infty} H(x) f(x) dx.$$

The **moment generating function** (MGF) of X is defined as

$$m_X(t) = E(e^{Xt}) = \int_{-\infty}^{\infty} e^{xt} f(x) dx.$$

The **k-th moment** about 0, **mean** and **variance** of X are respectively, given by

$$\mu_k' = E(X^k) = \int_{-\infty}^{\infty} x^k f(x) dx = \left[\frac{d^k}{dt^k}[m_X(t)]\right]_{t=0},$$

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

$$\sigma^2 = E(X^2) - E(X)^2 = \int_{-\infty}^{\infty} x^2 f(x) dx - \left(\int_{-\infty}^{\infty} x f(x) dx\right)^2.$$

Ex. The PDF of a continuous random variable X is:

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

Find $E(X)$, $E(X^2)$, σ^2 and MGF. Also find $E(X)$ and $E(X^2)$ using MGF.

Sol. The expected value $E(X)$ is given by:

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

Since $f(x) = 0$ outside the interval $[0, 1]$, the integral simplifies to:

$$E(X) = \int_0^1 x \cdot 2x\, dx = 2\int_0^1 x^2\, dx = 2\left[\frac{x^3}{3}\right]_0^1 = 2\left(\frac{1^3}{3} - \frac{0^3}{3}\right) = 2 \cdot \frac{1}{3} = \frac{2}{3}$$

The expected value of X^2 is given by:

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\, dx$$

Again, since $f(x) = 0$ outside $[0, 1]$, the integral becomes:

$$E(X^2) = \int_0^1 x^2 \cdot 2x\, dx = 2\int_0^1 x^3\, dx = 2\left[\frac{x^4}{4}\right]_0^1 = 2\left(\frac{1^4}{4} - \frac{0^4}{4}\right) = 2 \cdot \frac{1}{4} = \frac{1}{2}$$

The variance σ^2 is given by:

$$\sigma^2 = E(X^2) - [E(X)]^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}$$

MGF

For the given PDF, the MGF is give by:

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx = \int_0^1 e^{tx} \cdot 2x\, dx.$$

Using the integration by parts and simplifying, we get:

$$M_X(t) = \left[x \cdot \frac{2}{t} e^{tx}\right]_0^1 - \int_0^1 \frac{2}{t} e^{tx}\, dx. = \frac{2e^t}{t} - \frac{2e^t}{t^2} + \frac{2}{t^2}.$$

$E(X)$ using MGF

We have

$$M_X(t) = \left[x \cdot \frac{2}{t}e^{tx}\right]_0^1 - \int_0^1 \frac{2}{t}e^{tx}\,dx. = \frac{2e^t}{t} - \frac{2e^t}{t^2} + \frac{2}{t^2}.$$

First, compute the derivative of $M_X(t)$:

$$M_X'(t) = \frac{d}{dt}\left(\frac{2e^t}{t} - \frac{2e^t}{t^2} + \frac{2}{t^2}\right).$$

Using the quotient rule and simplifying, we get:

$$M_X'(t) = \frac{2e^t(t-1)}{t^2} + \frac{4e^t}{t^3} - \frac{4}{t^3}.$$

Now, evaluate $M_X'(t)$ at $t = 0$. This requires taking the limit as $t \to 0$:

$$E(X) = \lim_{t\to 0} M_X'(t).$$

Using L'Hôpital's rule, we find:

$$E(X) = \frac{2}{3}.$$

$E(X^2)$ using MGF

First, compute the second derivative of $M_X(t)$:

$$M_X''(t) = \frac{d}{dt}\left(\frac{2e^t(t-1)}{t^2} + \frac{4e^t}{t^3} - \frac{4}{t^3}\right).$$

Simplify and evaluate at $t = 0$:

$$E(X^2) = \lim_{t\to 0} M_X''(t).$$

Using L'Hôpital's rule, we find:

$$E(X^2) = \frac{1}{2}.$$

4.2 Uniform or Rectangular Distribution

A random variable X is said to have uniform distribution if its density function $f(x)$ is constant for all values of x, in the form:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b] \\ 0 & \text{if } x \notin [a, b] \end{cases}$$

The plot of f is shown in Figure 4.3.

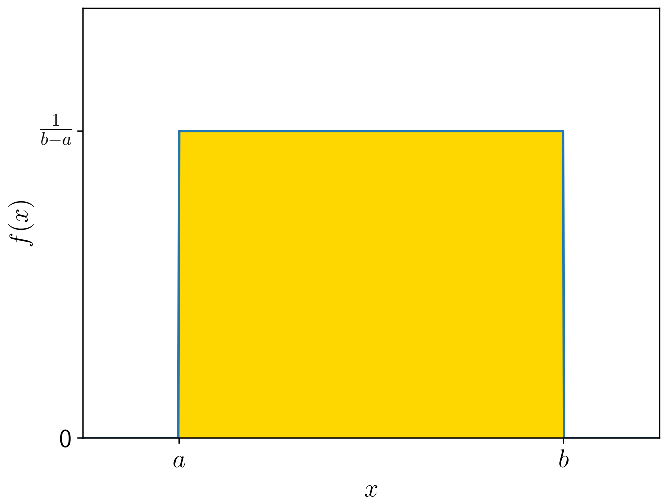

Figure 4.3: The area of the shaded golden rectangular region gives the total probability 1.

You may easily derive the following for the uniform distribution.

$$\mu = E(X) = \frac{1}{b-a}\int_a^b x dx = \frac{b+a}{2},$$

$$\sigma^2 = E(X^2) - E(X)^2 = \frac{1}{b-a}\int_a^b x^2 dx - \left(\frac{1}{b-a}\int_a^b x dx\right)^2 = \frac{(b-a)^2}{12}.$$

$$m_X(t) = E(e^{tX}) = \frac{1}{b-a}\int_a^b e^{tx} dx = \frac{e^{bt} - e^{at}}{(b-a)t}..$$

$$F(x) = \begin{cases} 0, & x \in (-\infty, a) \\ \dfrac{1}{b-a}\int_a^x dt, & x \in [a, b] \\ 1, & x \in (b, \infty) \end{cases} = \begin{cases} 0, & x \in (-\infty, a) \\ \frac{x-a}{b-a}, & x \in [a, b] \\ 1, & x \in (b, \infty) \end{cases}$$

Key points of Uniform Distribution

- $X = x$: Any value in a finite interval $[a, b]$.
- $f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b] \\ 0 & \text{if } x \notin [a, b] \end{cases}$
- $M_X(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}$
- $X \sim \text{Uniform}\left(\frac{a+b}{2}, \frac{(b-a)^2}{12}\right)$

Ex. Suppose that a large conference room at a certain company can be reserved for no more than 4 hours. Both long and short conferences occur quite often. Assume that the length X of a conference has a uniform distribution on the interval [0, 4].
(a) What is the probability density function?
(b) What is the probability that any given conference lasts at least 3 hours?

Sol. (a) The PDF is given by

$$f(x) = \begin{cases} \frac{1}{4}, & x \in [0,4] \\ 0, & x \notin [0,4] \end{cases}$$

(b) $P(X \geq 3) = \int_3^4 \frac{1}{4} dx = \frac{1}{4}$.

Ex. The Hubble constant H_0 measures the present expansion rate of the universe in the units of km/s/Mpc. Observations suggest that its value could be anywhere between 50 and 100. If H_0 follows uniform distribution in the interval [50, 100], find (a) PDF of H_0 (b) probability that H_0 lies in the interval $[60, 80]$ (c) the expected value of H_0.
Sol. (a) Here, $a = 50$ and $b = 100$, so the PDF of H_0 is given by

$$f(H_0) = \begin{cases} \frac{1}{50}, & H_0 \in [50,100] \\ 0, & H_0 \notin [50,100] \end{cases}$$

(b) $P(60 < H_0 < 80) = \int_{60}^{80} \frac{1}{50} dx = \frac{1}{50}(80 - 60) = 0.4$

(c) $E(H_0) = \frac{50+100}{2} = 75$

Ex. (Estimating Data Access Time) A cloud storage company allocates random access times for clients' data requests. The access time, X, is uniformly distributed between $a = 0.5$ seconds and $b = 1.5$ seconds. Calculate the probability that the access time is less than 1 second. Find the expected access time.
Sol. Here, $a = 0.5$ and $b = 1.5$, so:

$$f(x) = \frac{1}{1.5 - 0.5} = 1.$$

Probability of $X < 1$

The cumulative probability up to $x = 1$ is:

$$P(X < 1) = \int_{0.5}^{1} f(x)\, dx = \int_{0.5}^{1} 1\, dx = [x]_{0.5}^{1} = 1 - 0.5 = 0.5.$$

Thus, the probability is $\mathbf{0.5}\,(\mathbf{50}\%)$.

Expected Access Time

The expected value $E(X)$ for a uniform distribution is given by:

$$E(X) = \frac{a+b}{2}.$$

Substituting $a = 0.5$ and $b = 1.5$:

$$E(X) = \frac{0.5 + 1.5}{2} = 1.$$

Thus, the expected access time is **1** second.

Relevance to Computer Science and AI

1. **Fair Resource Allocation:** This example demonstrates how uniform distributions model randomness in resource allocation, ensuring fairness. In computer systems, access times, load balancing, and random number generation often follow uniform distributions.

2. **Training AI Models:** In artificial intelligence, uniform distributions are used to initialize weights in neural networks (e.g., Xavier initialization) to ensure unbiased training.

3. **Simulations and Testing:** AI systems often rely on simulations where inputs are uniformly distributed across a range to test robustness and optimize algorithms.

4. **Random Sampling:** Continuous uniform distributions are critical in Monte Carlo methods, which are fundamental to probabilistic inference, reinforcement learning, and optimization in AI.

4.3 Exponential Distribution

The waiting time for a Poisson event in a Poisson process follows a distribution called exponential distribution. In the following, we derive its PDF.

Let λ be number of Poisson events per unit time, and Y denote the number of Poisson events in time x. Suppose X represent the time until the first Poisson event occurs. Then the probability that no events occur in time x is given by the Poisson distribution with $Y = 0$:

$$P(X > x) = P(Y = 0) = \frac{(\lambda x)^0 e^{-\lambda x}}{0!} = e^{-\lambda x}$$

It follows that

$$P(X \leq x) = 1 - P(Y = 0) = 1 - e^{-\lambda x}.$$

This is the CDF of X

$$F(x) = 1 - e^{-\lambda x}, \quad x \geq 0$$

So $f(x) = F'(x) = \lambda e^{-\lambda x}$ is the PDF of X.

In general, the PDF of the exponential distribution is given by:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

where $\lambda > 0$ is the rate parameter (or equivalently, the scale parameter $\beta = 1/\lambda$). See Figure 4.4.

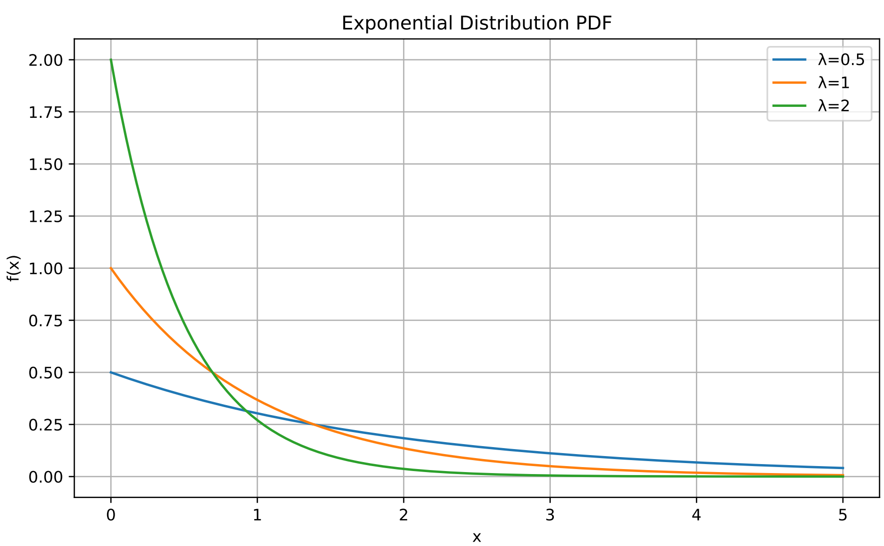

Figure 4.4: Exponential distribution for different values of λ.

The MGF of the exponential distribution is:

$$M_X(t) = E[e^{tX}] = \int_0^\infty e^{tx} f(x)\, dx = \int_{-\infty}^\infty e^{tx} \lambda e^{-\lambda x}\, dx = \lambda \int_0^\infty e^{-(\lambda - t)x}\, dx$$

This integral converges only if $t < \lambda$. Solving the integral:

$$M_X(t) = \lambda \left[\frac{e^{-(\lambda - t)x}}{-(\lambda - t)} \right]_0^\infty = \lambda \left(\frac{1}{\lambda - t} \right) = \frac{\lambda}{\lambda - t}$$

$$M'_X(t) = \frac{d}{dt}\left(\frac{\lambda}{\lambda - t}\right) = \frac{\lambda}{(\lambda - t)^2}$$
$$M''_X(t) = \frac{d}{dt}\left(\frac{\lambda}{(\lambda - t)^2}\right) = \frac{2\lambda}{(\lambda - t)^3}$$

Therefore, mean of X is given by

$$E[X] = M'_X(0) = \frac{\lambda}{(\lambda - 0)^2} = \frac{1}{\lambda}$$

Also,

$$E[X^2] = M''_X(0) = \frac{2\lambda}{(\lambda - 0)^3} = \frac{2}{\lambda^2}$$

So variance is

$$V(X) = E[X^2] - (E[X])^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

Key points of Exponential Distribution

- $X = x$: Waiting time for a Poisson event in a Poisson process with rate parameter λ
- $f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \in [0, \infty) \\ 0 & \text{if } x \notin [0, \infty) \end{cases}$
- $F(x) = 1 - e^{-\lambda x}$
- $M_X(t) = \frac{\lambda}{\lambda - t}$, $t < \lambda$
- $X \sim \text{Exponential}\left(\frac{1}{\lambda}, \frac{1}{\lambda^2}\right)$

Ex. Suppose the time between arrivals at a service desk follows an exponential distribution with a rate parameter $\lambda = 0.5$ (per minute). Find:

1. The probability that the time between arrivals is less than 2 minutes.
2. The mean and variance of the time between arrivals.

Sol.

1. To find probability that $X < 2$ minutes, we use the CDF $F(x) = 1 - e^{-\lambda x}$ of the exponential distribution, that is,

$$P(X < 2) = F(2) = 1 - e^{-\lambda \cdot 2} = 1 - e^{-0.5 \cdot 2} = 1 - e^{-1} \approx 0.6321$$

2. Mean and Variance are given by
   - $E[X] = \frac{1}{\lambda} = \frac{1}{0.5} = 2$ minutes.
   - $V(X) = \frac{1}{\lambda^2} = \frac{1}{(0.5)^2} = 4$ minutes^2.

4.3.1 Memoryless Property of the Exponential Distribution

The memoryless property of the Exponential distribution is one of its most important and distinctive features. This property states that the probability of an event occurring in the future is independent of what has already happened. In other words, the probability of waiting for an additional amount of time is not affected by how much time has already passed.

Formally, for an exponential random variable X with rate parameter λ, the memoryless property is given by:

$$P(X > x + t \mid X > x) = P(X > t), \quad \text{for all } x, t \geq 0.$$

This means that, given that the event hasn't occurred until time x, the distribution of the remaining waiting time is the same as the original distribution, unaffected by the elapsed time.

Explanation of the Memoryless Property

Let's consider the probability that X, the waiting time for an event, is greater than $x + t$ given that X is already greater than x. This is written as the conditional probability:

$$P(X > x + t \mid X > x).$$

Using the definition of conditional probability:

$$P(X > x + t \mid X > x) = \frac{P(X > x + t)}{P(X > x)}.$$

Substituting the survival function $P(X > x) = e^{-\lambda x}$, we get:

$$P(X > x + t \mid X > x) = \frac{e^{-\lambda(x+t)}}{e^{-\lambda x}} = e^{-\lambda t}.$$

This is precisely the survival function $P(X > t)$, which shows that the probability of waiting for t more time is the same as the probability of waiting for t from the beginning, confirming the memoryless property.

Interpretation: The memoryless property means that no matter how much time has already passed without the event occurring, the remaining time until the event happens still follows the same exponential distribution. It "forgets" the past. For example, if a lightbulb lasts an average of 1000 hours and is still working after 500 hours, the expected remaining lifetime of the lightbulb is still exponentially distributed with the same average lifetime (1000 hours), regardless of the 500 hours it has already lasted.

This property is unique to the exponential distribution and is a key characteristic that makes it suitable for modeling random processes such as the time between arrivals of events in a Poisson process or the time until failure of a component in reliability theory.

Ex. Based on extensive testing, it is determined that the time Y in years before a major repair is required for a certain washing machine is characterized by the density function

$$f(y) = \frac{1}{4}e^{-y/4}, \quad y \geq 0.$$

Note that Y is an exponential random variable with $\mu = 4$ years. The machine is considered a bargain if it is unlikely to require a major repair before the sixth year.

(a) What is the probability $P(Y > 6)$?

(b) What is the probability that a major repair is required in the first year?

Sol. (a) $P(Y > 6) = 0.2231$.

Thus, the probability that the washing machine will require major repair after year six is 0.223. Of course, it will require repair before year six with probability 0.777. Thus, one might conclude the machine is not really a bargain.

(b) The probability that a major repair is necessary in the first year is

$$P(Y < 1) = 1 - e^{-1/4} = 1 - 0.779 = 0.221.$$

Ex. (Time Between Successive Failures in a Biological System) In Biological Systems Engineering, the exponential distribution is commonly used to model the time between successive events in systems where events happen independently at a constant rate. A typical example can be found in the *failure rate of cells in a culture system* or the *decay of a radioactive substance* in biological organisms.

Assume that the failure rate of a biological system is constant. On average, one failure occurs every 2 hours. This failure rate can be modeled using the exponential distribution.

The exponential distribution is given by the probability density function (PDF):

$$f(t; \lambda) = \lambda e^{-\lambda t}, \quad t \geq 0$$

Where:

- t is the time until the next failure occurs,
- λ is the rate parameter (the rate of failure per unit time). In this case, $\lambda = \frac{1}{2}$ failures per hour, because we expect 1 failure every 2 hours.

Calculate the probability that the time until the next failure is less than or equal to 3 hours.

Sol. We need to find $P(T \leq 3)$, where T is the time until the next failure. This can be calculated using the cumulative distribution function (CDF) of the exponential distribution:

$$F(t) = 1 - e^{-\lambda t}$$

Substituting the given values:

$$F(3) = 1 - e^{-(\frac{1}{2})(3)} = 1 - e^{-1.5} \approx 1 - 0.2231 = 0.7769$$

Thus, the probability that the time until the next failure is less than or equal to 3 hours is approximately **77.69**%.

Importance in Biological Systems Engineering:

Understanding the time between successive failures or events is crucial for optimizing biological processes. Some of the applications include:

- **Enzyme Kinetics:** The time between reaction steps can be modeled using the exponential distribution to optimize reaction rates in biochemical engineering.
- **Cell Culture Systems:** The failure rate of cells in bioreactor systems can be modeled to predict when maintenance is required or when cells are likely to die.
- **Drug Release:** In drug delivery systems, the time between the release of doses can follow an exponential distribution, helping to model when subsequent doses are likely to be needed.

By applying the exponential distribution, engineers and biologists can make informed decisions about system design, maintenance scheduling, and optimization of biological processes in fields like biomanufacturing and biotechnology. The exponential distribution is widely used in reliability analysis, queuing theory, and survival analysis due to its memoryless property.

4.4 Gamma Distribution

The waiting time for a fixed number of Poisson events in a Poisson process follows a distribution called Gamma distribution. In the following, we derive its PDF.

Let λ be number of Poisson events per unit time in a Poisson process, and Y denote the number of Poisson events in time x. Suppose X represent the time until the α Poisson events occur. Then the probability that $\alpha - 1$ events occur in time x is given by:

$$P(X > x) = \sum_{Y=0}^{\alpha-1} P(Y = y) = e^{-\lambda x} \sum_{Y=0}^{\alpha-1} \frac{(\lambda x)^y}{y!}$$

It follows that

$$P(X \leq x) = 1 - \sum_{Y=0}^{\alpha-1} P(Y = y) = 1 - e^{-\lambda x} \sum_{Y=0}^{\alpha-1} \frac{(\lambda x)^y}{y!}$$

This is the CDF of the Gamma distribution:

$$F(x) = 1 - e^{-\lambda x} \sum_{Y=0}^{\alpha-1} \frac{(\lambda x)^y}{y!}, \quad x \geq 0$$

Differentiating the CDF with respect to x, and simplifying, we obtain:

$$f(x) = \frac{\lambda^\alpha}{(\alpha-1)!} x^{\alpha-1} e^{-\lambda x} = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}.$$

In general, the PDF of the Gamma distribution is:

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

where $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt$, $\alpha > 0$ and $\lambda > 0$. See Figure 4.5.

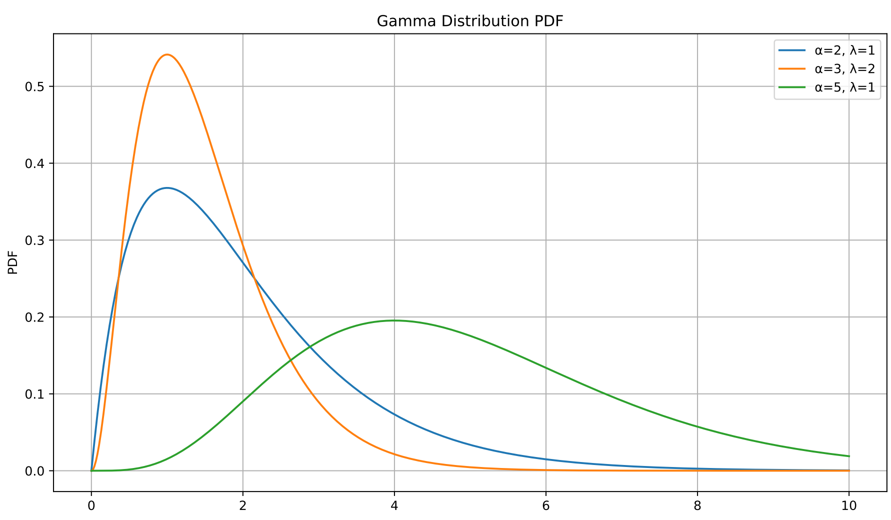

Figure 4.5: Gamma distribution for different values of α and λ.

The parameter α, also called shape parameter, of the Gamma distribution is not restricted to being a positive integer. It can be any positive real number. When the shape parameter α is a positive integer, the Gamma distribution reduces to the **Erlang distribution**, which is commonly used in queuing theory. The Gamma distribution with $\alpha = 1$ is the exponential distribution.

MGF

$$M_X(t) = E[e^{tX}] = \int_0^\infty e^{tx} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}\, dx$$

$$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\lambda-t)x}\, dx$$

Let $u = (\lambda - t)x$, then $du = (\lambda - t)dx$, and the integral becomes:

$$M_X(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty \left(\frac{u}{\lambda - t}\right)^{\alpha-1} e^{-u} \frac{du}{\lambda - t}$$

$$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \frac{1}{(\lambda - t)^\alpha} \int_0^\infty u^{\alpha-1} e^{-u}\, du$$

The integral is the Gamma function $\Gamma(\alpha)$, so:

$$M_X(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha)}{(\lambda - t)^\alpha} = \left(1 - \frac{t}{\lambda}\right)^{-\alpha}$$

Mean

$$M_X'(t) = \alpha \left(1 - \frac{t}{\lambda}\right)^{-\alpha-1} \cdot \frac{1}{\lambda}$$

Evaluate at $t = 0$:

$$E[X] = \alpha \cdot \frac{1}{\lambda} = \frac{\alpha}{\lambda}$$

Variance

$$M_X''(t) = \alpha(\alpha+1) \left(1 - \frac{t}{\lambda}\right)^{-\alpha-2} \cdot \frac{1}{\lambda^2}$$

Evaluate at $t = 0$:

$$E[X^2] = \alpha(\alpha+1) \cdot \frac{1}{\lambda^2}$$

Thus:

$$V(X) = \frac{\alpha(\alpha+1)}{\lambda^2} - \left(\frac{\alpha}{\lambda}\right)^2 = \frac{\alpha}{\lambda^2}$$

Key points of Gamma Distribution

- $X = x$: Waiting time for α Poisson events in a Poisson process with rate parameter λ
- $f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & \text{if } x \in (0, \infty) \\ 0 & \text{if } x \notin (0, \infty) \end{cases}$
- $M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^\alpha$, $t < \lambda$
- $X \sim \text{Gamma}\left(\frac{\alpha}{\lambda}, \frac{\alpha}{\lambda^2}\right)$

```python
# PROGRAM FOR GAMMA DISTRIBUTION PROBABILITIES
from scipy.integrate import quad
from scipy.special import gamma
from math import exp, inf
def gamma_prob():
    pdf = lambda x: (lam**alpha/gamma(alpha))*x**(alpha-1)*exp(-lam*x)
    return quad(pdf, a, b)[0]
# Example usage:
a=0; b = 2; alpha=3; lam=2
print(f"P({a}<X<{b}:alpha={alpha}, lam={lam}): {gamma_prob():.8f}")
a=0; b = 60; alpha=5; lam=0.1
print(f"P({a}<X<{b}:alpha={alpha}, lam={lam}): {gamma_prob():.8f}")

#Output

P(0<X<2:alpha=3, lam=2): 0.76189669
P(0<X<60:alpha=5, lam=0.1): 0.71494350
```

Ex. Suppose the waiting time (in hours) for 3 events in a Poisson process with rate $\lambda = 2$ follows a Gamma distribution. Find:

(a) The probability that the waiting time is less than 2 hours.

(b) The mean and variance of the waiting time.

Sol. (a) **Probability**: The PDF is:

$$f(x; 3, 2) = \frac{2^3}{\Gamma(3)} x^2 e^{-2x}$$

Compute $P(X < 2)$:

$$P(X < 2) = \int_0^2 \frac{8}{2} x^2 e^{-2x}\, dx$$

Using integration by parts or a Gamma distribution table, we find:

$$P(X < 2) \approx 0.7619$$

(b) **Mean and Variance**:

$$E[X] = \frac{\alpha}{\lambda} = \frac{3}{2} = 1.5\,\text{hours}$$

$$V(X) = \frac{\alpha}{\lambda^2} = \frac{3}{4} = 0.75\,\text{hours}^2$$

Ex. (Survival time in a biomedical experiment) In a biomedical study with rats, a dose-response investigation is used to determine the effect of the dose of a toxicant on their survival time. The toxicant is one that is frequently discharged into the atmosphere from jet fuel. For a certain dose of the toxicant, the study determines that the survival time, in weeks, has a gamma distribution with $\alpha = 5$ and $\lambda = \frac{1}{10}$. What is the probability that a rat survives no longer than 60 weeks?

Sol. Let the random variable X be the survival time (time to death). The required probability is

$$P(X \leq 60) = \int_0^{60} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx$$

Using $\alpha = 5$, $\lambda = \frac{1}{10}$ and $-\lambda x = y$, we obtain

$$P(X \leq 60) = \int_{-6}^{0} \frac{1}{\Gamma(5)} y^4 e^y dy = \frac{1}{4!}[e^y(y^4 - 4y^3 + 12y^2 - 24y - 24)]_{y=-6}^{y=0} \approx 0.715$$

Alternatively, we can use the CDF of Erlang distribution

$$F(x; \alpha, \lambda) = 1 - e^{-\lambda x} \sum_{Y=0}^{\alpha-1} \frac{(\lambda x)^y}{y!}$$

to evaluate $P(X \leq 60)$, that is,

$$P(X \leq 60) = F(60; 5, \frac{1}{10}) \approx 0.715$$

So the probability that a rat survives no longer than 60 weeks is 71.5%.

4.5 Normal Distribution (Gaussian distribution)

The normal distribution is the most important continuous probability distribution in the entire field of statistics. It is also known as the **Gaussian distribution**. In fact, the normal distribution was first described by De Moivre in 1733 as the limiting case of Binomial distribution when number of trials is large. This discovery did not get much attention. Around fifty years later, Laplace and Gauss rediscovered normal distribution while dealing with astronomical data. They found that the errors in astronomical measurements are well described by normal distribution. It approximately describes many phenomena that occur in nature, industry, and research. For example, physical measurements in areas such as meteorological experiments, rainfall studies, measurements of manufactured parts, and errors in scientific measurements are often more than adequately explained with a normal distribution.

Remarkably, when n, np and $n(1-p)$ are large, then it can be shown that the binomial distribution is well approximated by a distribution of the form:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \sim \frac{1}{\sqrt{2\pi(np(1-p))}}\, e^{-\frac{1}{2}\left(\frac{x-np}{\sqrt{np(1-p)}}\right)^2},$$

known as the normal distribution.

Formally, a continuous random variable X is said to follow Normal distribution with parameters μ and σ if its density function is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty,\ -\infty < \mu < \infty,\ \sigma > 0.$$

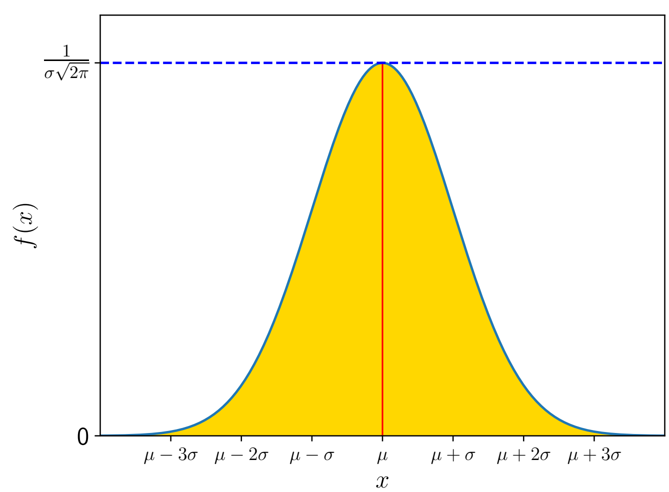

Figure 4.6: The area of the shaded golden region under the normal probability curve gives the total probability 1. The normal probability curve is symmetrical about the vertical red line $x = \mu$. Therefore, $P(X \leq \mu] = 0.5 = P(X \geq \mu)$. Also, the maximum value of $f(x)$ occurs at $x = \mu$, and is given by $f(\mu) = \frac{1}{\sigma\sqrt{2\pi}}$.

For the normal random variable X, we can verify the following:

$$\int_{-\infty}^{\infty} f(x)dx = 1, \quad m_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}, \quad E(X) = \mu, \quad V(X) = \sigma^2.$$

Verify the PDF property: $\int_{-\infty}^{\infty} f(x)dx = 1$

We have

$$\int_{-\infty}^{\infty} f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$
$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2\sigma^2}} dy, \text{ where } y = x - \mu$$
$$= \frac{2}{\sigma\sqrt{2\pi}} \int_{0}^{\infty} e^{-\frac{y^2}{2\sigma^2}} dy$$
$$= \frac{2}{\sigma\sqrt{2\pi}} \int_{0}^{\infty} e^{-r} r^{-\frac{1}{2}} \frac{\sqrt{2}}{2}\sigma dr, \text{ where } \tfrac{y^2}{2\sigma^2} = r$$
$$= \frac{1}{\sqrt{\pi}} \int_{0}^{\infty} e^{-r} r^{-\frac{1}{2}} dr$$
$$= \frac{1}{\sqrt{\pi}} \Gamma(1/2)$$
$$= \frac{1}{\sqrt{\pi}} \sqrt{\pi}$$
$$= 1.$$

MGF

$$m_X(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f(x)dx$$
$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 + tx} dx$$
$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}\left[(x-\mu)^2 - 2\sigma^2 tx\right]} dx$$
$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}\left[x^2+\mu^2-2\mu x - 2\sigma^2 tx\right]} dx$$
$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}\left[x^2-2(\mu+\sigma^2 t)x+\mu^2\right]} dx$$
$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}\left[(x-(\mu+\sigma^2 t))^2 - 2\mu\sigma^2 t - \sigma^4 t^2\right]} dx$$
$$= \frac{1}{\sigma\sqrt{2\pi}} e^{\mu t + \frac{1}{2}\sigma^2 t^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}\left[x-(\mu+\sigma^2 t)\right]^2} dx$$
$$= \frac{1}{\sigma\sqrt{2\pi}} e^{\mu t + \frac{1}{2}\sigma^2 t^2} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2\sigma^2}} dy, \text{ where } y = x - (\mu + \sigma^2 t)$$
$$= \frac{2}{\sigma\sqrt{2\pi}} e^{\mu t + \frac{1}{2}\sigma^2 t^2} \int_{0}^{\infty} e^{-\frac{y^2}{2\sigma^2}} dy$$
$$= \frac{2}{\sigma\sqrt{2\pi}} e^{\mu t + \frac{1}{2}\sigma^2 t^2} \int_{0}^{\infty} e^{-r} r^{-\frac{1}{2}} \frac{\sqrt{2}}{2}\sigma dr, \text{ where } \tfrac{y^2}{2\sigma^2} = r$$
$$= \frac{1}{\sqrt{\pi}} e^{\mu t + \frac{1}{2}\sigma^2 t^2} \int_{0}^{\infty} e^{-r} r^{-\frac{1}{2}} dr$$
$$= \frac{1}{\sqrt{\pi}} e^{\mu t + \frac{1}{2}\sigma^2 t^2} \Gamma(1/2)$$
$$= \frac{1}{\sqrt{\pi}} e^{\mu t + \frac{1}{2}\sigma^2 t^2} \sqrt{\pi}$$
$$= e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

Mean and Variance

$$M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

So it follows that

$$m'_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}(\mu + \sigma^2 t)$$

$$m''_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}[(\mu + \sigma^2 t)^2 + \sigma^2]$$

$$E(X) = m'_X(0) = \mu\ .$$

$$V(X) = E(X^2) - E(X)^2 = m''_X(0) - [m'_X(0)]^2 = \mu^2 + \sigma^2 - \mu^2 = \sigma^2.$$

Thus, the two parameters μ and σ in the density function of normal random variable X are its mean and standard deviation, respectively.

Note: If X is a normal random variable with mean μ and variance σ^2, then we write $X \sim N(\mu, \sigma^2)$.

Key points of Normal (Gaussian) Distribution

- Obtained as a limiting case of Binomial distribution: $X \sim \text{Binomial}(np, np(1-p))$ when np and $n(1-p)$ are large enough
- $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- $M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$
- $X \sim N\left(\mu, \sigma^2\right)$

4.5.1 Standard Normal Distribution

Suppose $X \sim N(\mu, \sigma^2)$. Then a variable Z defined by

$$Z = \frac{X - \mu}{\sigma}$$

is called standard normal variable.

Mean and Variance of Z

Noting that $E(X) = \mu$ and $V(X) = \sigma^2$, we have

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

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

Thus, $Z \sim N(0, 1)$.

PDF of Z

Since $Z \sim N(0, 1)$, the PDF of Z reads as

$$\phi(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\dfrac{z^2}{2}}, \quad -\infty < z < \infty.$$

See Figure 4.7.

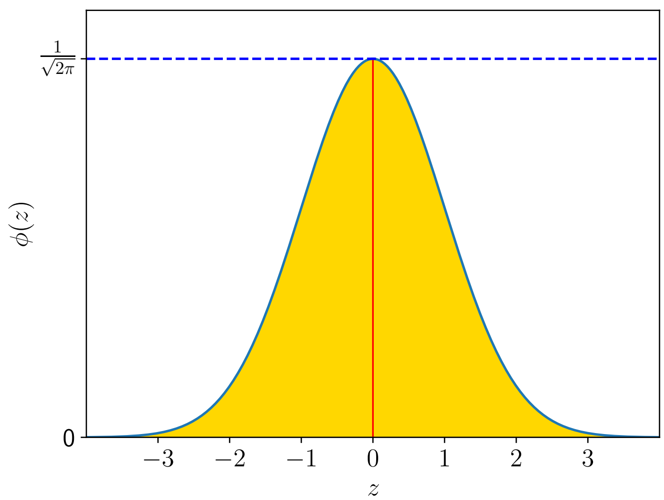

Figure 4.7: The area of the shaded golden region under the standard normal probability curve gives the total probability 1. The normal probability curve is symmetrical about the vertical red line $z = 0$. Therefore, $P(Z \leq 0) = 0.5 = P(Z \geq 0)$. Also, the maximum value of $\phi(z)$ occurs at $z = 0$, and is given by $\phi(0) = \dfrac{1}{\sqrt{2\pi}}$.

CDF of Z

It is given by

$$F(z) = \int_{-\infty}^{z} \phi(z)dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\dfrac{t^2}{2}}\, dt.$$

Probability of normal random variable in an interval

Suppose $X \sim N(\mu, \sigma^2)$. Then $Z = \frac{X-\mu}{\sigma}$ is standard normal variable. Probability of X in an interval (a, b) is calculated as follows:

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) = F\left(\frac{b-\mu}{\sigma}\right) - F\left(\frac{a-\mu}{\sigma}\right).$$

Note that CDF of X or Z cannot be evaluated manually and it requires numerical integration. So CDF values for Z are provided in a table called standard normal distribution table (See Table 4.1). Also, note that because of the symmetry, the CDF values at negative values of Z can be calculated by using $F(-a) = 1 - F(a)$.

```
# PROGRAM FOR NORMAL DISTRIBUTION PROBABILITIES
from scipy.integrate import quad
from math import exp, sqrt, pi, inf
def norm_prob():
    pdf = lambda x: exp(-0.5*((x-mu)/sigma)**2)/(sigma*sqrt(2*pi))
    return quad(pdf, a, b)[0]
# Example usage
a=-2; b=2; mu=0; sigma=1
print(f"P({a}<X<{b}:mu={mu},sigma={sigma})={norm_prob():.8f}")
a=-inf; b=0; mu=0; sigma=1
print(f"P({a}<X<{b}:mu={mu},sigma={sigma})={norm_prob():.8f}")
# Output:
P(-2<X<2:mu=0,sigma=1)=0.95449974
P(-inf<X<0:mu=0,sigma=1)=0.50000000
```

Table 4.1: Standard Normal Distribution Table: It provides CDF values $F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt$ for the standard normal variable Z. Also, note that the CDF values at negative z values can be calculated by using $F(-a) = 1 - F(a)$.

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
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Probabilities in 1σ, 2σ and 3σ Confidence Intervals

From the standard normal distribution table, we see that $F(1) = 0.8413$. It follows that

$$P(|X-\mu| < \sigma) = P(\mu - \sigma < X < \mu + \sigma)$$

$$= P(-1 < Z < 1) = F(1) - F(-1) = F(1) - (1 - F(1)) = 2F(1) - 1 = 2(0.8413) - 1 = 0.6826.$$

This shows that there is approximately 68% probability that the normal variable X lies in the interval $(\mu - \sigma, \mu + \sigma)$, as shown in Figure 4.8. We call this interval as the 1σ confidence interval of X.

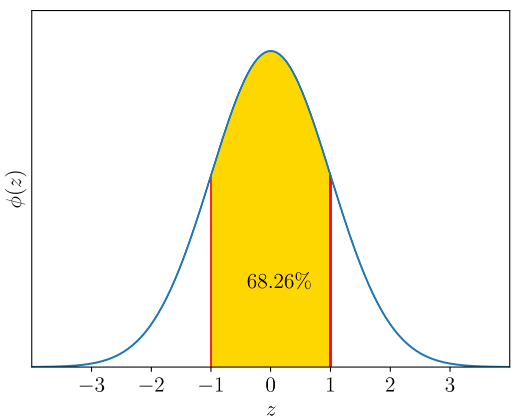

Figure 4.8: The area of the shaded golden region under the standard normal probability curve gives the probability corresponing to the 1σ confidence interval $(\mu-\sigma, \mu+\sigma)$. So $P(\mu-\sigma < X < \mu+\sigma) = P(-1 < Z < 1) = 0.6826$.

Similarly, the probabilities of X in 2σ and 3σ confidence intervals are respectively, are given by

$$P(|X-\mu| < 2\sigma) = P(\mu - 2\sigma < X < \mu + 2\sigma) = P(-2 < Z < 2) = 0.9544,$$

$$P(|X-\mu| < 3\sigma) = P(\mu - 3\sigma < X < \mu + 3\sigma) = P(-3 < Z < 3) = 0.9973.$$

For geometrical clarity, see the left and right panels in Figure 4.9.

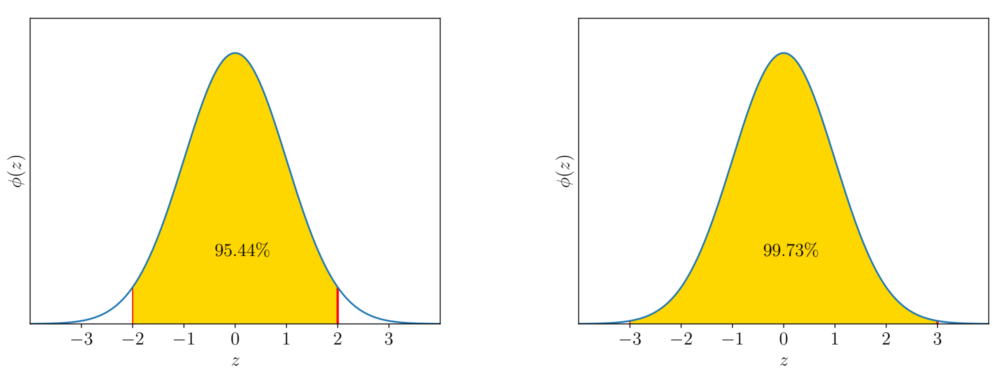

Figure 4.9: **Left panel:** The area of the shaded golden region under the standard normal probability curve gives $P(\mu - 2\sigma < X < \mu + 2\sigma) = P(-2 < Z < 2) = 0.9544$.
Right panel: The area of the shaded golden region under the standard normal probability curve gives $P(\mu - 3\sigma < X < \mu + 3\sigma) = P(-3 < Z < 3) = 0.9973$.

Key points of Standard Normal Distribution

- $Z = \frac{X-\mu}{\sigma}$: Standard Normal Variable where $X \sim (\mu, \sigma^2)$
- $Z \sim N(0,1)$
- $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$
- $M_Z(t) = e^{\frac{1}{2}t^2}$
- $F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt$
- $P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) = F\left(\frac{b-\mu}{\sigma}\right) - F\left(\frac{a-\mu}{\sigma}\right)$
- $F(-a) = 1 - F(a)$
- $F(-a < Z < a) = F(a) - F(-a) = 2F(a) - 1$

Ex. A random variable X is normally distributed with mean 9 and standard deviation 3. Find $P(X \geq 15)$, $P(X \leq 15)$ and $P(0 \leq X \leq 9)$.

Sol. Here X is normal random variable with $\mu = 9$, $\sigma = 3$. So $Z = \frac{X-9}{3}$ is standard normal variable.

$$\therefore\ P(X \geq 15) = P\left(Z \geq \frac{15-9}{3}\right) = P(Z \geq 2) = 1 - F(2) = 1 - 0.9772 = 0.0228$$

$$P(X \leq 15) = 1 - 0.0228 = 0.9772$$

$$P(0 \leq X \leq 9) = P\left(\frac{0-9}{3} \leq Z \leq \frac{9-9}{3}\right)$$

$$= P(-3 \leq Z \leq 0) = F(0) - F(-3) = F(0) - (1 - F(3)) = 0.5 - (1 - 0.9987) = 0.4987$$

Ex. Given a normal distribution with $\mu = 40$ and $\sigma = 6$, find the value of X that has
(a) 45% of the area to the left and
(b) 14% of the area to the right.
Sol. (a) We require a Z value that leaves an area of 0.45 to the left. From the Standard Normal Table 4.1, we find $P(Z < -0.13) = 0.45$, so the desired Z value is -0.13. Hence, using $Z = (X - \mu)/\sigma$, we find

$$X = \sigma Z + \mu = (6)(-0.13) + 40 = 39.22.$$

(b) Here we require a Z value that leaves 0.14 of the area to the right and hence an area of 0.86 to the left. Again, from Standard Normal Table 4.1, we find $P(Z < 1.08) = 0.86$, so the desired Z value is 1.08 and

$$X = \sigma Z + \mu = (6)(1.08) + 40 = 46.48.$$

Ex. In a normal distribution, 12% of the items are under 30 and 85% are under 60. Find the mean and standard deviation of the distribution.
Sol. Let μ be mean and σ be standard deviation of the distribution. Given that $P(X < 30) = 0.12$ and $P(X < 60) = 0.85$. Let z_1 and z_2 be values of the standard normal variable Z corresponding to $X = 30$ and $X = 60$ respectively so that $P(Z < z_1) = 0.12$ and $P(Z < z_2) = 0.85$. From the Standard Normal Table 4.1, we find $z_1 \approx -1.17$ and $z_2 \approx 1.04$ since $F(-1.17) = 0.121$ and $F(1.04) = 0.8508$.

Finally, solving the equations, $\frac{30-\mu}{\sigma} = -1.17$ and $\frac{60-\mu}{\sigma} = 1.04$, we find $\mu = 45.93$ and $\sigma = 13.56$.

Ex. Use MGF of normal distribution to derive 3rd and 4th central moments, and hence show that skewness of normal distribution is 0 and kurtosis is 3.
Sol. The third central moment is:

$$\mu_3 = E[(X-\mu)^3]$$

Using the standardized variable $Z = \frac{X-\mu}{\sigma}$, we can write:

$$\mu_3 = E[(\sigma Z)^3] = \sigma^3 E[Z^3]$$

Likewise, the fourth central moment is

$$\mu_4 = E[(\sigma Z)^4] = \sigma^4 E[Z^4]$$

Now, $E[Z^3]$ and $E[Z^4]$ may be obtained by the third and fourth order derivatives of the MGF of Z at $t = 0$:

$$M_Z(t) = e^{\frac{1}{2}t^2},\ M_Z'(t) = te^{\frac{1}{2}t^2},\ M_Z''(t) = (t^2+1)e^{\frac{1}{2}t^2},\ M_Z'''(t) = (t^3+3t)e^{\frac{1}{2}t^2},\ M_Z^{(4)}(t) = (t^4+6t^2+3)e^{\frac{1}{2}t^2}$$

Therefore,

$$E[Z^3] = M_Z'''(0) = 0,\ E[Z^4] = M_Z^{(4)}(0) = 3$$

It follows that

$$\mu_3 = \sigma^3 \cdot 0 = 0,\ \mu_4 = 3\sigma^4$$

These results confirm that the skewness ($\gamma_1 = \frac{\mu_3}{\sigma^3}$) is **0** and the kurtosis ($\gamma_2 = \frac{\mu_4}{\sigma^4}$) is **3** for a normal distribution.
Ex. (Sensor Noise in Robotics and Cyber-Physical Systems) In robotics and cyber-physical systems (CPS), sensors often provide noisy measurements. This noise can be modeled using the **normal distribution**. A robot uses a LIDAR sensor measuring the distance to an obstacle with a standard deviation of $\sigma = 0.05$ meters. The mean measured distance to an obstacle over 100 readings is $\mu = 2.0$ meters. Estimate the probability that the true distance lies between 1.95 and 2.05 meters.
Sol. The probability that the true distance lies between 1.95 and 2.05 meters is:

$$P(1.95 < X < 2.05) = P\left(\frac{1.95-2.0}{0.05} < Z < \frac{2.05-2.0}{0.05}\right)$$

$$= P(-1 < Z < 1) = 2F(1) - 1 = 2(0.8413) - 1 = 0.6826$$

So there is approximately a **68.26% probability** that the true distance lies within $\pm 1\sigma$ of the mean, which aligns with the empirical rule of the normal distribution.

Importance in Robotics and CPS

- **Sensor Fusion:** Normal distribution is critical in algorithms like Kalman Filters, combining noisy sensor data.
- **Uncertainty Modeling:** It quantifies measurement uncertainty for robust decision-making.
- **Anomaly Detection:** Deviations from the normal distribution indicate sensor faults or environmental anomalies.
- **Control Systems:** Noise modeling ensures stable and reliable control systems.

The normal distribution plays a key role in making robotics and CPS systems intelligent and reliable.

4.5.2 Approximation of Binomial distribution by Normal distribution

If X is a Binomial random variable with parameters n and p, then X approximately follows a normal distribution with mean np and variance $np(1-p)$ provided n is large. For most of the practical purposes, the approximation is acceptable if $np \geq 10$ and $n(1-p) \geq 10$.

The normal distribution also provides a fairly good approximation even when n is small and p is reasonably close to 0.5.

Ex. Let X be the number of heads in a random experiment where a fair coin is tossed 4 times. Find $P(1 \leq X \leq 3)$ using Binomial distribution and approximate the same using normal distribution.

Sol. Here $n = 4$ and $p = 0.5$. So using Binomial distribution, we have

$$P(1 \leq X \leq 3) = P(X = 1) + P(X = 2) + P(X = 3)$$
$$= \binom{4}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^3 + \binom{4}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 + \binom{4}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^1$$
$$= \frac{4}{16} + \frac{6}{16} + \frac{4}{16} = \frac{14}{16} = 0.875.$$

For normal approximation, we first find the mean and variance: $\mu = np = (4)(0.5) = 2$, $\sigma^2 = np(1-p) = (4)(0.5)(0.5) = 1$.

Note that the conditions $np \geq 10$ and $n(1-p) \geq 10$ are not met here, but $p = 0.5$ so we can still proceed for the normal approximation.

Next, we apply half unit or continuity correction in the given range $1 \leq X \leq 3$ of X, so it becomes $0.5 \leq X \leq 3.5$. Then the required normal approximation is given by

$$P(0.5 \leq X \leq 3.5) = P(-1.5 \leq Z \leq 1.5) = 2F(1.5) - 1 = 2(0.9332) - 1 = 0.8664,$$

which is a good approximation to the binomial probability 0.875. See Figure 4.10 for graphical illustration.

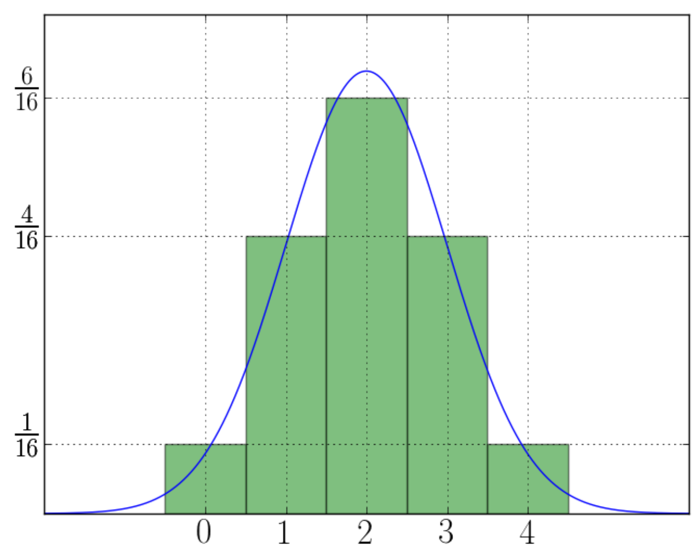

Figure 4.10: Histogram of the binomial distribution with $n = 4$, $p = 0.5$ where X takes the values $0, 1, 2, 3, 4$ with probabilities $1/16$, $4/16$, $6/16$, $4/16$ and $1/16$, respectively. The magenta curve is the normal probability curve with $\mu = np = (4)(0.5) = 2$ and $\sigma^2 = np(1-p) = (4)(0.5)(0.5) = 1$.

Ex. A multiple-choice quiz has 200 questions, each with 4 possible answers of which only 1 is correct. What is the probability that sheer guesswork yields from 25 to 30 correct answers for the 80 of the 200 problems about which the student has no knowledge?

Sol. Here $n = 80$, $p = \frac{1}{4} = 0.25$. So using binomial distribution, we have

$$P(25 \leq X \leq 30) = \sum_{x=25}^{30} \binom{80}{x} (0.25)^x (0.75)^{80-x} \approx 0.11927$$

For normal approximation, we first find the mean and SD: $\mu = np = (80)(0.25) = 20$,
$\sigma = \sqrt{np(1-p)} = \sqrt{(80)(0.25)(0.75)} = 3.873$.
Note that here $np = 20$ and $n(1-p) = 60$, so the conditions $np \geq 10$ and $n(1-p) \geq 10$ are satisfied, and we can proceed for the normal approximation.

Next, we apply half unit or continuity correction in the given range $25 \leq X \leq 30$ of X, so it becomes $24.5 \leq X \leq 30.5$. Then the required normal approximation is given by

$$P(24.5 \leq X \leq 30.5) = P(1.16 < Z < 2.71)$$

$$= P(Z < 2.71) - P(Z < 1.16)$$

$$= 0.9966 - 0.8770 = 0.1196,$$

which is a good approximation to the binomial probability 0.11927.

4.5.3 Approximation of Poisson distribution by Normal distribution

Let X be a Poisson random variable with parameter k. Then for large k, X is approximately normal with mean k and variance k. It follows that the Poisson probabilities can be approximated by normal distribution $N(k, k)$, by using the 0.5 unit or continuity correction on both sides of the given range of X as we did in the binomial case. As a rule of thumb, a Poisson distribution with parameter k can be approximated by a normal distribution when $k \geq 10$.

Ex. Suppose $X \sim \text{Poisson}(k = 20)$. Approximate $P(X \leq 25)$ using the normal approximation.

Sol. Since $k = 20 \geq 10$, the normal approximation is valid. Mean and SD are

- Mean: $\mu = 20$
- Standard deviation: $\sigma = \sqrt{20} \approx 4.472$

For normal approximation, we need to apply continuity correction:

$$P(X \leq 25) \approx P\left(Z \leq \frac{25 + 0.5 - 20}{\sqrt{20}}\right)$$

$$= P\left(Z \leq \frac{5.5}{4.472}\right)$$

$$= P(Z \leq 1.23)$$

From standard normal table, $P(Z \leq 1.23) \approx 0.8907$. Thus, $P(X \leq 25) \approx 0.8907$.

For comparison, the exact probability using the Poisson distribution is:

$$P(X \leq 25) = \sum_{x=0}^{25} \frac{e^{-20} 20^x}{x!} \approx 0.8878$$

The normal approximation (0.8907) is very close to the exact value (0.8878).

4.6 Chi-Squared Distribution

The **chi-squared distribution** is a widely used probability distribution in statistics, particularly in hypothesis testing and confidence interval estimation. It is a special case of the **Gamma distribution** and is closely related to the **normal distribution**.

The chi-squared distribution with k degrees of freedom, denoted as χ_k^2, is the distribution of the sum of the squares of k independent standard normal random variables. If $Z_1, Z_2, \ldots, Z_k$ are independent and identically distributed (i.i.d.) standard normal random variables, then:

$$X = Z_1^2 + Z_2^2 + \cdots + Z_k^2 = \sum_{i=1}^{n} Z_i^2$$

follows a chi-squared distribution with k degrees of freedom with PDF given by

$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2}, \quad x > 0$$

See Figure 4.11.

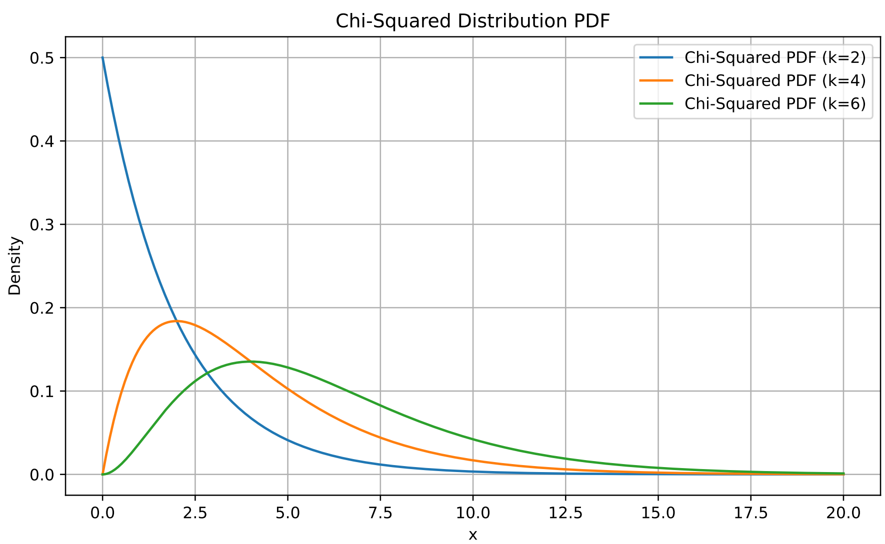

Figure 4.11: Chi-squared distribution for different values of k.

Chi-Squared as a Special Case of Gamma

The chi-squared distribution is a special case of the Gamma distribution. The Gamma distribution has PDF:

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x > 0$$

where α is the shape parameter and β is the rate parameter.

For the chi-squared distribution:

- Shape parameter: $\alpha = \frac{k}{2}$
- Rate parameter: $\lambda = \frac{1}{2}$

Substituting these into the Gamma PDF gives the chi-squared PDF:

$$f(x) = \frac{(1/2)^{k/2}}{\Gamma(k/2)} x^{(k/2)-1} e^{-x/2}, \quad x > 0$$

MGF

The MGF of the chi-squared distribution with k degrees of freedom is:

$$M_X(t) = (1-2t)^{-k/2}, \quad t < \frac{1}{2}$$

Mean

$$M_X'(t) = k(1-2t)^{-(k/2+1)}$$

Evaluate at $t = 0$:

$$E[X] = k(1-0)^{-(k/2+1)} = k$$

Variance

$$M_X''(t) = k(k+2)(1-2t)^{-(k/2+2)}$$

Evaluate at $t = 0$:

$$E[X^2] = k(k+2)(1-0)^{-(k/2+2)} = k(k+2)$$

The variance is:

$$V(X) = E[X^2] - (E[X])^2 = k(k+2) - k^2 = 2k$$

Key points of Chi-squared Distribution

- $X = \sum_{i=1}^{k} Z_i^2$: Sum of squares of k independent standard normal random variables (called Chi-squared (χ_k^2) random variable with k degrees of freedom)
- $f(x) = \begin{cases} \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2} & \text{if } x \in (0,\infty) \\ 0 & \text{if } x \notin (0,\infty) \end{cases}$
- It is also a special case of Gamma distribution with $\lambda = \frac{1}{2}$, $\alpha = 2$.
- $M_X(t) = (1-2t)^{-k/2}, \quad t < \frac{1}{2}$
- $X \sim \chi_k^2(k, 2k)$. Further, $X \sim N(k, 2k)$ provided $k \geq 30$.

The following Python program may be used to calculate probability of the chi-squared random variable X in any interval (a, b).

```
# PROGRAM FOR CHI-SQUARED DISTRIBUTION PROBABILITIES
from scipy.integrate import quad
from scipy.special import gamma
def chi_prob():
    pdf = lambda x, k: x**(k/2-1)*np.exp(-x/2)/(2**(k/2)*gamma(k/2))
    return quad(pdf, a, b, args=(k,))[0]
# Example usage:
a=0; b=10; k=5
print(f"P({a}<X<{b}:k={k})={chi_prob():.8f}")
# Output:
P(0<X<10:k=5)=0.92476475
P(0<X<inf:k=5)=1.00000000
```

Ex. Suppose $X \sim \chi_5^2$. Then the probability $P(X < 10) = P(0 < X < 10) = 0.92476475$.
Mean: $E[X] = k = 5$. Variance: $V(X) = 2k = 10$.

4.7 Student's t-Distribution

The **t-distribution**, also known as **Student's t-distribution**[1], is a probability distribution used for estimating the mean of a normally distributed population when the sample size is small and the population standard deviation is unknown. It is widely used in hypothesis testing, confidence intervals, and regression analysis.

If Z is a standard normal random variable and $Y \sim \chi^2_\nu(\nu, 2\nu)$ is a chi-squared random variable with ν degrees of freedom then the random variable

$$T = \frac{Z}{\sqrt{Y/\nu}}$$

follows a t-distribution with ν degrees of freedom, and PDF given by:

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$

See Figure 4.12.

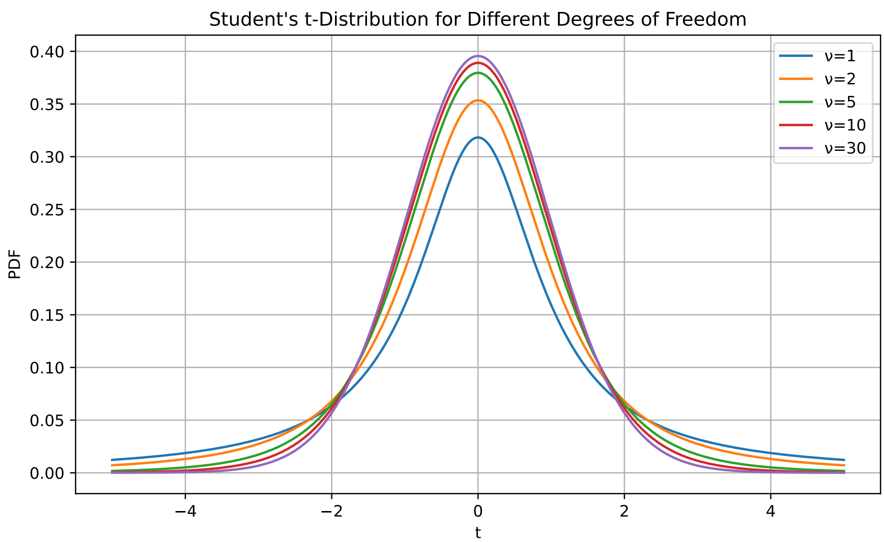

Figure 4.12: Student's t-distribution for different values of ν.

Proof

The t-distribution arises from the ratio of a standard normal random variable Z to the square root of a chi-squared random variable Y divided by its degrees of freedom ν:

$$T = \frac{Z}{\sqrt{Y/\nu}}$$

The joint PDF of Z and Y is the product of their individual PDFs:

$$f(z, y) = f_Z(z) \cdot f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \cdot \frac{y^{\nu/2-1} e^{-y/2}}{2^{\nu/2}\Gamma(\nu/2)}$$

[1]The probability distribution of T was first published in 1908 in a paper written by W. S. Gosset. At the time, Gosset was employed by an Irish brewery that prohibited publication of research by members of its staff. To circumvent this restriction, he published his work secretly under the name "Student". Consequently, the distribution of T is usually called the Student t-distribution or simply the t-distribution.

Define the transformation:

$$T = \frac{Z}{\sqrt{Y/\nu}}, \quad Y = Y$$

The inverse transformation is:

$$Z = T\sqrt{Y/\nu}, \quad Y = Y$$

The Jacobian matrix J of the transformation is:

$$J = \begin{bmatrix} \frac{\partial z}{\partial t} & \frac{\partial z}{\partial y} \\ \frac{\partial y}{\partial t} & \frac{\partial y}{\partial y} \end{bmatrix} = \begin{bmatrix} \sqrt{y/\nu} & \frac{t}{2\sqrt{y\nu}} \\ 0 & 1 \end{bmatrix}$$

The determinant of the Jacobian is:

$$|J| = \sqrt{y/\nu}$$

Using the transformation formula, the joint PDF of T and Y is:

$$f(t, y) = f_{Z,Y}(z(t, y), y) \cdot |J|$$

Substitute $z = t\sqrt{y/\nu}$ and the Jacobian:

$$f(t, y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2 y}{2\nu}} \cdot \frac{y^{\nu/2-1} e^{-y/2}}{2^{\nu/2}\Gamma(\nu/2)} \cdot \sqrt{y/\nu}$$

Simplify the expression:

$$f(t, y) = \frac{y^{(\nu+1)/2-1} e^{-\frac{y}{2}\left(1+\frac{t^2}{\nu}\right)}}{\sqrt{2\pi\nu}\, 2^{\nu/2}\Gamma(\nu/2)}$$

To find the marginal PDF of T, integrate the joint PDF over y:

$$f_T(t) = \int_0^\infty f_{T,Y}(t, y)\, dy = \int_0^\infty \frac{y^{(\nu+1)/2-1} e^{-\frac{y}{2}\left(1+\frac{t^2}{\nu}\right)}}{\sqrt{2\pi\nu}\, 2^{\nu/2}\Gamma(\nu/2)}\, dy$$

Let $u = \frac{y}{2}\left(1 + \frac{t^2}{\nu}\right)$, then $du = \frac{1}{2}\left(1 + \frac{t^2}{\nu}\right) dy$. The integral becomes:

$$f_T(t) = \frac{1}{\sqrt{2\pi\nu}\, 2^{\nu/2}\Gamma(\nu/2)} \int_0^\infty y^{(\nu+1)/2-1} e^{-u} \cdot \frac{2}{1 + \frac{t^2}{\nu}}\, du$$

Simplify and recognize the gamma function:

$$f_T(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

This is the PDF of the t-distribution with ν degrees of freedom.

Mean

For $\nu > 1$, the mean of the t-distribution is:

$$E[T] = \int_{-\infty}^{\infty} t f(t)\, dt = 0,$$

since $tf(t)$ is an odd function over $(-\infty, \infty)$.

Variance

The variance is defined as:

$$V(T) = E[T^2] - (E[T])^2.$$

Since $E[T] = 0$, the variance simplifies to:

$$V(T) = E[T^2].$$

To compute $E[T^2]$, we use the definition of the t-distribution:

$$T = \frac{Z}{\sqrt{Y/\nu}},$$

where:

- $Z \sim \mathcal{N}(0,1)$ (standard normal),
- $Y \sim \chi^2_\nu$ (chi-squared with ν degrees of freedom),
- Z and Y are independent.

1. Compute $E[T^2]$:

$$E[T^2] = E\left[\frac{Z^2}{Y/\nu}\right] = \nu E[Z^2] E\left[\frac{1}{Y}\right].$$

2. Since $Z \sim \mathcal{N}(0,1)$, $E[Z^2] = 1$.
3. For $Y \sim \chi^2_\nu$, the expectation of $\frac{1}{Y}$ is:

$$E\left[\frac{1}{Y}\right] = \frac{1}{\nu - 2}, \quad \text{for } \nu > 2.$$

   This follows from the property of the inverse chi-squared distribution.
4. Substitute these results into $E[T^2]$:

$$E[T^2] = \nu \cdot 1 \cdot \frac{1}{\nu - 2} = \frac{\nu}{\nu - 2}.$$

Thus, the variance of the t-distribution is:

$$V(T) = \frac{\nu}{\nu - 2}, \quad \text{for } \nu > 2.$$

Key points of t-Distribution

- $T = \frac{Z}{\sqrt{Y/\nu}}$ (the random variable (T_ν) of t-distribution with ν degrees of freedom): Z is a standard normal random variable and Y is a chi-squared random variable with ν degrees of freedom.
- $f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$
- $T \sim T_\nu(0, \frac{\nu}{\nu-2})$, $\nu > 2$
- The t-distribution converges to the standard normal, that is, $T \sim N(0,1)$ as the dof increase, and the approximation becomes reasonably good for $\nu \geq 30$).

```
# PROGRAM FOR STUDENT t-DISTRIBUTION PROBABILITIES
from scipy.integrate import quad
from math import pi, sqrt, inf
from scipy.special import gamma
def t_prob():
    pdf = lambda t: gamma((nu+1)/2)/(sqrt(nu*pi)*gamma(nu/2))*(1+t**2/nu)**(-(nu+1)/2)
    return quad(pdf, a, b)[0]
# Example usage
a=-2; b=2; nu=5
print(f"P({a}<T<{b}:nu=5): {t_prob():.8f}")
a=-inf; b=0; nu=5
print(f"P({a}<T<{b}:nu=5): {t_prob():.8f}")

# Output:

P(-2<T<2:nu=5): 0.89806052
P(-inf<T<0:nu=7): 0.50000000
```

4.8 F-Distribution

The **F-distribution** is a continuous probability distribution that arises frequently in statistical analysis, particularly in the context of hypothesis testing (e.g., Analysis of Variance (ANOVA), regression). It is the ratio of two independent chi-squared random variables, each divided by their respective degrees of freedom.

If $U \sim \chi^2_{\nu_1}(\nu_1, 2\nu_1)$ and $V \sim \chi^2_{\nu_2}(\nu_2, 2\nu_2)$ are independent chi-squared random variables, then the ratio $X = \frac{U/\nu_1}{V/\nu_2}$ follows an F-distribution with ν_1 and ν_2 degrees of freedom and PDF:

$$f(x) = \frac{\sqrt{\frac{(\nu_1 x)^{\nu_1} \cdot \nu_2^{\nu_2}}{(\nu_1 x + \nu_2)^{\nu_1+\nu_2}}}}{x \cdot B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)}$$

where$B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1}\, dt$ is the Beta function. See Figure 4.13.

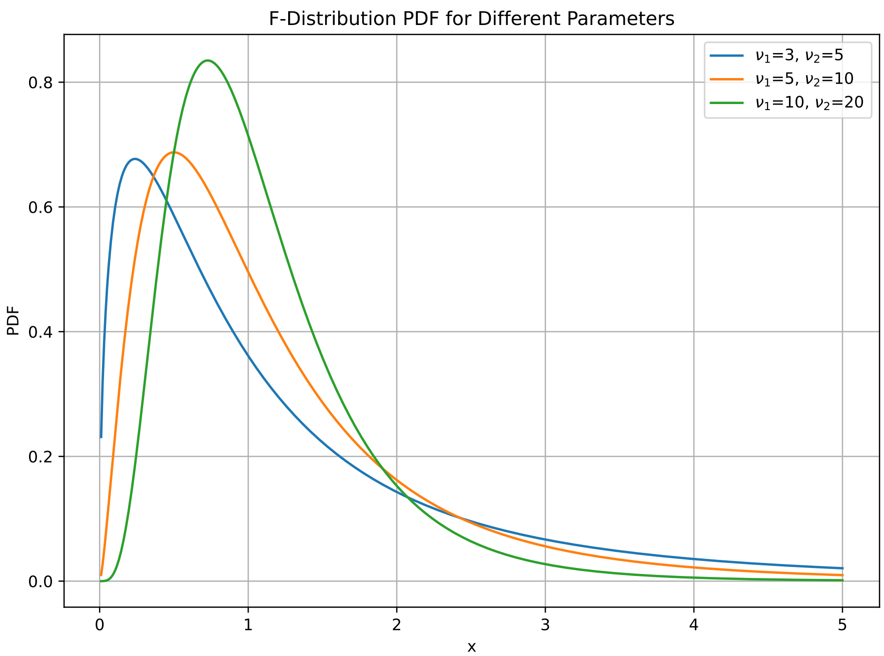

Figure 4.13: F-distribution for different values of ν_1 and ν_2.

The joint PDF of U and V is:

$$f_{U,V}(u, v) = \frac{u^{\nu_1/2-1} e^{-u/2} \cdot v^{\nu_2/2-1} e^{-v/2}}{2^{(\nu_1+\nu_2)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)}$$

Transform (U, V) to (X, Y), where $X = \frac{U/\nu_1}{V/\nu_2}$ and $Y = V$.
Compute the Jacobian of the transformation and integrate out Y to obtain the marginal PDF of X.

Mean

$$E[X] = E\left[\frac{U/\nu_1}{V/\nu_2}\right] = \frac{\nu_2}{\nu_1} \cdot E\left[\frac{U}{V}\right].$$

Since U and V are independent:

$$E\left[\frac{U}{V}\right] = E[U] \cdot E\left[\frac{1}{V}\right].$$

Substitute: $E[U] = \nu_1$ and $E\left[\frac{1}{V}\right] = \frac{1}{\nu_2 - 2}$ (for $\nu_2 > 2$).

$$E[X] = \frac{\nu_2}{\nu_1} \cdot \nu_1 \cdot \frac{1}{\nu_2 - 2} = \frac{\nu_2}{\nu_2 - 2}, \quad \text{for } \nu_2 > 2.$$

Variance

$$E[X^2] = E\left[\left(\frac{U/\nu_1}{V/\nu_2}\right)^2\right] = \frac{\nu_2^2}{\nu_1^2} \cdot E[U^2] \cdot E\left[\frac{1}{V^2}\right]$$

- For $U \sim \chi^2_{\nu_1}(\nu_1, 2\nu_1)$, $E[U^2] = V(U) + (E[U])^2 = 2\nu_1 + \nu_1^2$.
- For $V \sim \chi^2_{\nu_2}(\nu_2, 2\nu_2)$, $E\left[\frac{1}{V^2}\right] = \frac{1}{(\nu_2-2)(\nu_2-4)}$ (for $\nu_2 > 4$).

Substituting:

$$E[X^2] = \frac{\nu_2^2}{\nu_1^2} \cdot (2\nu_1 + \nu_1^2) \cdot \frac{1}{(\nu_2 - 2)(\nu_2 - 4)}$$

It follows that

$$V(X) = E[X^2] - (E[X])^2 = \frac{\nu_2^2(2\nu_1 + \nu_1^2)}{\nu_1^2(\nu_2 - 2)(\nu_2 - 4)} - \left(\frac{\nu_2}{\nu_2 - 2}\right)^2 = \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}$$

Key points of F-Distribution

- $X = \frac{\chi^2_{\nu_1}/\nu_1}{\chi^2_{\nu_2}/\nu_2}$ (the random variable (F_{ν_1,ν_2}) of F-distribution with ν_1 and ν_2 degrees of freedom): $\chi^2_{\nu_1}$ and $\chi^2_{\nu_2}$ are two independent chi-squared random variables with ν_1 and ν_2 degrees of freedom respectively.
- $f(x) = \frac{\sqrt{\frac{(\nu_1 x)^{\nu_1} \cdot \nu_2^{\nu_2}}{(\nu_1 x + \nu_2)^{\nu_1 + \nu_2}}}}{x \cdot B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)}$
- $X \sim F_{\nu_1,\nu_2}\left(\frac{\nu_2}{\nu_2 - 2}, \frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)}\right)$
- $T = \frac{Z}{\sqrt{\chi^2_\nu/\nu}}$ implies $T^2 = \frac{Z^2}{\chi^2_\nu/\nu} = \frac{\chi^2_1/1}{\chi^2_\nu/\nu}$. It shows that square of a t-distributed random variable follows an F-distribution with 1 numerator degree of freedom and ν denominator degrees of freedom, that is, $T^2 \sim F_{1,\nu}$.

```
# PROGRAM FOR F-DISTRIBUTION PROBABILITIES
from scipy.integrate import quad
from scipy.special import beta
def f_prob():
    pdf = lambda x: ((nu1*x)**nu1 * nu2**nu2 / (nu1*x + nu2)**(nu1+nu2))**0.5 /(x * beta
   (nu1/2, nu2/2))
    return quad(pdf, a, b)[0]
# Example usage:
a=0; b = 2; nu1=2; nu2=7
print(f"P({a}<F<{b}:nu1={nu1}, nu2={nu2}): {f_prob():.8f}")
a=2; b = math.inf; nu1=5; nu2=10
print(f"P({a}<F<{b}:nu1={nu1}, nu2={nu2}): {f_prob():.8f}")

# Output:

P(0<F<2:nu1=2, nu2=7): 0.79442574
P(2<F<inf:nu1=5, nu2=10): 0.16419495
```

4.9 Summary of continuous probability distributions

It would be helpful to know the following key points of continuous probability distributions.

1. Uniform Distribution

- $X = x$: Any value in a finite interval $[a, b]$.
- $f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b] \\ 0 & \text{if } x \notin [a, b] \end{cases}$
- $M_X(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}$
- $X \sim \text{Uniform}\left(\frac{a+b}{2}, \frac{(b-a)^2}{12}\right)$

2. Exponential Distribution

- $X = x$: Waiting time for a Poisson event in a Poisson process with rate parameter λ
- $f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \in [0, \infty) \\ 0 & \text{if } x \notin [0, \infty) \end{cases}$
- $M_X(t) = \frac{\lambda}{\lambda - t}$, $t < \lambda$
- $X \sim \text{Exponential}\left(\frac{1}{\lambda}, \frac{1}{\lambda^2}\right)$

3. Gamma Distribution

- $X = x$: Waiting time for α Poisson events in a Poisson process with rate parameter λ
- $f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & \text{if } x \in (0, \infty) \\ 0 & \text{if } x \notin (0, \infty) \end{cases}$
- $M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^\alpha$, $t < \lambda$
- $X \sim \text{Exponential}\left(\frac{\alpha}{\lambda}, \frac{\alpha}{\lambda^2}\right)$

4. Normal (Gaussian) Distribution

- Obtained as a limiting case of Binomial distribution: $X \sim \text{Binomial}(np, np(1-p))$ when np and $n(1-p)$ are large enough
- $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- $M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$
- $X \sim N\left(\mu, \sigma^2\right)$
- $Z = \frac{X-\mu}{\sigma}$: Standard Normal Variable
- $Z \sim N(0, 1)$
- $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$

- $M_Z(t) = e^{\frac{1}{2}t^2}$
- $F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt$
- $P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) = F\left(\frac{b-\mu}{\sigma}\right) - F\left(\frac{a-\mu}{\sigma}\right)$
- $F(-a) = 1 - F(a)$
- $F(-a < Z < a) = F(a) - F(-a) = 2F(a) - 1$

4. Chi-squared Distribution

- $X = \sum_{i=1}^{k} Z_i^2$: Sum of squares of k independent standard normal random variables (called Chi-squared (χ_k^2) random variable with k degrees of freedom)
- $f(x) = \begin{cases} \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2} & \text{if } x \in (0, \infty) \\ 0 & \text{if } x \notin (0, \infty) \end{cases}$
- It is also a special case of Gamma distribution with $\lambda = \frac{1}{2}$, $\alpha = 2$.
- $M_X(t) = (1 - 2t)^{-k/2}, \quad t < \frac{1}{2}$
- $X \sim \chi_k^2(k, 2k)$

5. Student's t-Distribution

- $T = \frac{Z}{\sqrt{Y/\nu}}$ (the random variable (T_ν) of t-distribution with ν degrees of freedom): Z is a standard normal random variable and Y is a chi-squared random variable with ν degrees of freedom.
- $f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$
- $T \sim T_\nu(0, \frac{\nu}{\nu-2})$, $\nu > 2$

6. F-Distribution

- $X = \frac{U/\nu_1}{V/\nu_2}$ (the random variable (F_{ν_1,ν_2}) of F-distribution with ν_1 and ν_2 degrees of freedom): U and V are two independent chi-squared random variables with ν_1 and ν_2 degrees of freedom respectively.
- $f(x) = \frac{\sqrt{\frac{(\nu_1 x)^{\nu_1} \cdot \nu_2^{\nu_2}}{(\nu_1 x + \nu_2)^{\nu_1+\nu_2}}}}{x \cdot B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)}$
- $X \sim F_{\nu_1,\nu_2}\left(\frac{\nu_2}{\nu_2-2}, \frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)}\right)$

Remark: We have learned the above-mentioned distributions in this chapter as per the need in the later chapters. But note that there exist numerous more continuous distributions applicable in various practical situations.

4.10 Chebyshev's Inequality

If X is normal random variable with mean μ and variance σ^2, then $P(|X-\mu| < k\sigma) = P(|Z| < k) = F(k) - F(-k)$. However, if X is any random variable, then the rule of thumb for the required probability is given by the Chebyshev's inequality as stated below.

If X is a random variable with mean μ and variance σ^2, then

$$P(|X-\mu| < k\sigma) \geq 1 - \frac{1}{k^2}.$$

Proof. By definition of variance, we have

$$\int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \sigma^2$$

$$\implies \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x)dx + \int_{\mu-k\sigma}^{\mu+k\sigma} (x-\mu)^2 f(x)dx + \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x)dx = \sigma^2$$

$$\implies \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x)dx + \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x)dx \leq \sigma^2 \ (\because \ \int_{\mu-k\sigma}^{\mu+k\sigma} (x-\mu)^2 f(x)dx \geq 0)$$

$$\implies \int_{-\infty}^{\mu-k\sigma} k^2\sigma^2 f(x)dx + \int_{\mu+k\sigma}^{\infty} k^2\sigma^2 f(x)dx \leq \sigma^2 \ (\because \ (x-\mu)^2 \geq k^2\sigma^2 \text{ for } x \leq \mu - k\sigma \text{ or } x \geq \mu + k\sigma)$$

$$\implies \int_{-\infty}^{\mu-k\sigma} f(x)dx + \int_{\mu+k\sigma}^{\infty} f(x)dx \leq \frac{1}{k^2}$$

$$\implies 1 - \int_{-\infty}^{\mu-k\sigma} f(x)dx - \int_{\mu+k\sigma}^{\infty} f(x)dx \geq 1 - \frac{1}{k^2}$$

$$\implies \int_{-\infty}^{\infty} f(x)dx - \int_{-\infty}^{\mu-k\sigma} f(x)dx - \int_{\mu+k\sigma}^{\infty} f(x)dx \geq 1 - \frac{1}{k^2} \ (\because \ \int_{-\infty}^{\infty} f(x)dx = 1)$$

$$\implies \int_{\mu-k\sigma}^{\mu+k\sigma} f(x)dx \geq 1 - \frac{1}{k^2}.$$

$$\implies P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}.$$

$$\implies P(|X-\mu| < k\sigma) \geq 1 - \frac{1}{k^2}.$$

Note that the Chebyshev's inequality does not yield the exact probability of X to lie in the interval $(\mu - k\sigma, \mu + k\sigma)$ rather it gives the minimum probability for the same. However, in case of normal random variable, the probability obtained is exact. For example, consider the 2σ interval $(\mu - 2\sigma, \mu + 2\sigma)$ for X. Then, Chebyshev's inequality gives $P(|X-\mu| < 2\sigma) \geq 1 - \frac{1}{4} = 0.75$. In case, X is normal variable, we get the exact probability $P(|X-\mu| < 2\sigma) = 0.9544$. However, the advantage of Chebyshev's inequality is that it applies to any random variable of known mean and variance. Also note that the above proof may be done for discrete random variable as well. So the Chebyshev's inequality is true for discrete random variable as well.

Ex. A random variable X with unknown probability distribution has mean 8 and S.D. 3. Use Chebyshev's inequality to find a lower bound of $P(-7 < X < 23)$.

Sol. Here $\mu = 8$ and $\sigma = 3$. So by Chebyshev's inequality, we have

$$P(8 - 3k < X < 8 + 3k) \geq 1 - \frac{1}{k^2}.$$

In order to get lower bound of $P(-7 < X < 23)$, we choose $k = 5$. We get

$$P(-7 < X < 23) \geq 1 - \frac{1}{25} = 0.96.$$

4.11 Continuous Bivariate Random Variable

Let X and Y be two continuous random variables. Then the ordered pair (X, Y) is called a two dimensional or bivariate continuous random variable.

4.11.1 Joint probability density function

A function f such that

$$f(x,y) \geq 0,\ P[(X,Y) \in R] = \iint_R f(x,y)dxdy,\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1,$$

where R is any region in the domain of f, is called joint probability density function of (X, Y).

4.11.2 Distribution function

The distribution function of (X, Y) is given by

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(x,y)dxdy.$$

4.11.3 Marginal density functions

The marginal density of X, denoted by f_X, is defined as

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy.$$

Similarly, the marginal density of Y, denoted by f_Y, is defined as

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx.$$

4.11.4 Independent random variables

The continuous random variables X and Y are said to be independent if and only if

$$f(x,y) = f_X(x)f_Y(y).$$

4.11.5 Expectation

The expectation or mean of X is defined as

$$E(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)dxdy = \mu_X.$$

In general, the expectation of a function of X and Y, say $H(X,Y)$, is defined as

$$E[H(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(x,y)f(x,y)dxdy.$$

4.11.6 Covariance

If μ_X and μ_Y are the means of X and Y respectively, then covariance of X and Y, denoted by $\text{Cov}(X, Y)$ is defined as

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - E(X)E(Y).$$

Ex. Let X denote a person's blood calcium level and Y, the blood cholesterol level. The joint density function of (X, Y) is

$$f(x, y) = \begin{cases} k, & 8.5 \leq x \leq 10.5, 120 \leq y \leq 240 \\ 0, & \text{elsewhere} \end{cases}$$

(i) Find the value of k.
(ii) Find the marginal densities of X and Y.
(iii) Find the probability that a healthy person has a cholesterol level between 150 to 200.
(iv) Are the variables X and Y independent?
(v) Find $\text{Cov}(X, Y)$.

Sol. (i) $f(x, y)$ being joint pdf, we have

$$1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = \int_{120}^{240} \int_{8.5}^{10.5} k dx dy = 240k.$$

So $k = 1/240$ and $f(x, y) = 1/240$

(ii) The marginal density of X is

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_{120}^{240} \frac{1}{240} dy = \frac{1}{2}, \quad 8.5 \leq x \leq 10.5.$$

Similarly, the marginal density of Y is

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx = \int_{8.5}^{10.5} \frac{1}{240} dx = \frac{1}{120}, \quad 120 \leq y \leq 240.$$

(iii) The probability that a healthy person has a cholesterol level between 150 to 200, is

$$P(150 \leq Y \leq 200) = \int_{150}^{200} f_Y(y) dy = \frac{5}{12}.$$

(iv) We have

$$f_X(x) f_Y(y) = \frac{1}{2} \times \frac{1}{120} = \frac{1}{240} = f(x, y).$$

This shows that X and Y are independent.

(v) We find

$$E(X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y) dx dy = \int_{120}^{240} \int_{8.5}^{10.5} \frac{x}{240} dx dy = 9.5,$$
$$E(Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x, y) dx dy = \int_{120}^{240} \int_{8.5}^{10.5} \frac{y}{240} dx dy = 180,$$

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy = \int_{120}^{240}\int_{8.5}^{10.5} \frac{xy}{240}dxdy = 1710.$$

Hence, $\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = 1710 - 9.5 \times 180 = 0$.

Ex. The joint density function of (X, Y) is

$$f(x,y) = \begin{cases} c/x, & 27 \le y \le x \le 33 \\ 0, & \text{elsewhere} \end{cases}$$

(i) Find the value of c.
(ii) Find the marginal densities and hence check the independence of X and Y
(iii) Evaluate $P(X \le 32,\ Y \le 30)$.

Sol. (i) Here the given range of (X, Y) is the triangular region common to the three regions given by the inequalities $y \ge 27$, $y \le x$ and $x \le 33$, as shown in Figure 4.14.

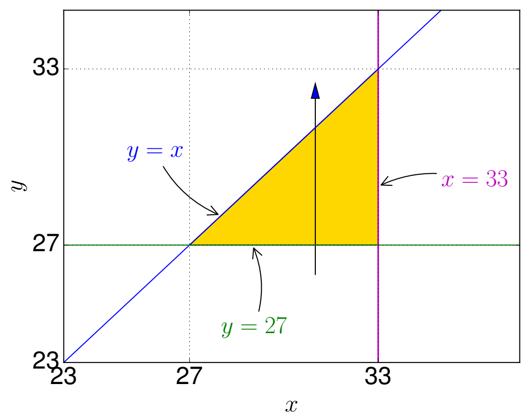

Figure 4.14: The shaded golden region is the triangular region given by the inequalities $y \ge 27$, $y \le x$ and $x \le 33$. The vertical ray enters the given region through the line $y = 27$ and leaves at the line $y = x$. The x-value at the leftmost point $(27, 27)$ of the region is $x = 27$, and at the rightmost points (all points on the line $x = 33$) is $x = 33$.

Considering vertical ray through the given region, we find that the x limits are from $x = 27$ to $x = 33$, and y limits are from $y = 27$ to $y = x$. Therefore, to find c, we use

$$\int_{27}^{33}\int_{27}^{x} f(x,y)dydx = 1$$

and we get

$$c = \frac{1}{6 - 27\ln(33/27)}.$$

(ii) $f_X(x) = \int_{y=27}^{y=x} \frac{c}{x}dy = c(1 - 27/x)$, $27 \le x \le 33$

$f_Y(y) = \int_{x=y}^{x=33} \frac{c}{x}dx = c(\ln 33 - \ln y)$, $27 \le y \le 33$.

We observe that $f(x,y) = c/x \ne f_X(x)f_Y(y)$. So X and Y are not independent.

(iii) To calculate the probability $P[X \leq 32,\ Y \leq 30]$, we need to integrate the joint density over the shaded golden region shown in Figure 4.15. Considering the horizontal ray through this region, we find that the x limits are from $x = y$ to $x = 32$, and y limits are from $y = 27$ to $y = 30$.

$$\therefore P[X \leq 32,\ Y \leq 30] = \int_{27}^{30} \int_{y}^{32} \frac{c}{x} dx dy = c(3 \ln 32 + 3 - 30 \ln 30 + 27 \ln 27).$$

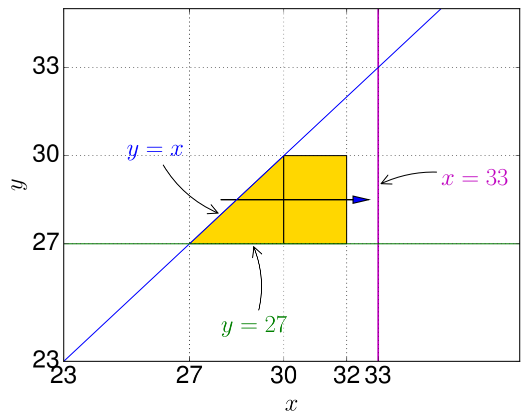

Figure 4.15: The shaded golden region is given by $X \leq 32,\ Y \leq 30$. The horizontal ray enters this region through the line $x = y$ and leaves at the line $x = 32$. The y-value at the bottomost points the region is $y = 27$, and at the uppermost points is $y = 30$.

4.12 Multivariate Gaussian Distribution

The **multivariate Gaussian distribution** (or **multivariate normal distribution**) generalizes the one-dimensional Gaussian to higher dimensions. It describes the joint distribution of a random vector $\boldsymbol{\theta} \in \mathbb{R}^n$ with mean $\bar{\boldsymbol{\theta}}$ and covariance matrix $\mathbf{C}$.

The PDF is given by:

$$N(\boldsymbol{\theta}; \bar{\boldsymbol{\theta}}, \mathbf{C}) = (2\pi)^{-n/2} |\mathbf{C}|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})^T \mathbf{C}^{-1} (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})\right)$$

where:

- $\boldsymbol{\theta}$ is an n-dimensional random vector.
- $\bar{\boldsymbol{\theta}}$ is the **mean vector**.
- $\mathbf{C}$ is the **covariance matrix** (symmetric and positive-definite).
- $|\mathbf{C}|$ is the determinant of $\mathbf{C}$.
- $\mathbf{C}^{-1}$ is the inverse of $\mathbf{C}$.

Example: Bivariate Gaussian ($n = 2$)

Let $\boldsymbol{\theta} = [\theta_1, \theta_2]^T$ follow a bivariate Gaussian distribution.

Parameters

- Mean vector: $\bar{\boldsymbol{\theta}} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$
- Covariance matrix: $\mathbf{C} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$

PDF Computation

1. **Determinant of C:**

$$|\mathbf{C}| = (1)(1) - (0.5)(0.5) = 0.75$$

2. **Inverse of C:**

$$\mathbf{C}^{-1} = \frac{1}{0.75} \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix} = \begin{bmatrix} \frac{4}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{4}{3} \end{bmatrix}$$

3. **Exponent term:**

$$(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})^T \mathbf{C}^{-1} (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}) = \frac{4}{3}\theta_1^2 - \frac{4}{3}\theta_1\theta_2 + \frac{4}{3}\theta_2^2$$

4. **Final PDF:**

$$N(\boldsymbol{\theta}; \bar{\boldsymbol{\theta}}, \mathbf{C}) = \frac{1}{2\pi\sqrt{0.75}} \exp\left(-\frac{1}{2}\left(\frac{4}{3}\theta_1^2 - \frac{4}{3}\theta_1\theta_2 + \frac{4}{3}\theta_2^2\right)\right)$$

Interpretation

- The distribution is centered at $[0, 0]$.
- θ_1 and θ_2 are **positively correlated** (since $C_{12} = 0.5 > 0$).
- Contours of equal probability are **ellipses** (due to correlation).

4.13 Calculating Sigma Difference Between Two Distributions

1. Symmetric Uncertainties with Covariance

For two measurements with means μ_1, μ_2 and covariance matrix $\mathbf{C}$:

$$\mathbf{C} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

The combined uncertainty is:

$$\sigma_{\text{combined}} = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$

The significance of discrepancy is:

$$N_\sigma = \frac{|\mu_1 - \mu_2|}{\sigma_{\text{combined}}}$$

2. Asymmetric Uncertainties

For measurements with asymmetric errors $\mu_1{}^{+\sigma_{1+}}_{-\sigma_{1-}}$ and $\mu_2{}^{+\sigma_{2+}}_{-\sigma_{2-}}$:

- If $\mu_1 > \mu_2$: Use σ_{1-} and σ_{2+}
- If $\mu_1 < \mu_2$: Use σ_{1+} and σ_{2-}

The combined uncertainty becomes:

$$\sigma_{\text{combined}} = \sqrt{\sigma_{\text{rel},1}^2 + \sigma_{\text{rel},2}^2 - 2\rho\sigma_{\text{rel},1}\sigma_{\text{rel},2}}$$

Interpretation

- $< 2\sigma$: Consistent (no significant tension)
- 2σ–3σ: Moderate tension
- $> 3\sigma$: Strong tension
- $> 5\sigma$: Highly significant (discovery threshold in physics)

This interpretation is commonly used in statistical analyses, especially in physics and cosmology, to quantify the level of agreement or discrepancy between theoretical predictions and experimental/observational data. The sigma (σ) levels correspond to standard deviations in a normal distribution:

- $< 2\sigma$ **(Consistent)**: The result is within a reasonable range of expected statistical fluctuations, indicating no significant tension.
- $2\sigma - 3\sigma$ **(Moderate tension)**: Some deviation is observed, but it is not strong enough to warrant serious concern.
- $> 3\sigma$ **(Strong tension)**: The deviation is becoming significant, suggesting a possible inconsistency that may require further investigation.
- $> 5\sigma$ **(Highly significant, discovery threshold)**: This is the gold standard in physics (e.g., in particle physics experiments like those at CERN). A result beyond 5σ is considered strong enough to claim a discovery, as the probability of it being due to random fluctuations is extremely low (~ 1 in 3.5 million).

Example: Hubble Tension

$$\begin{aligned}
&\text{SH0ES: } H_0 = 73.04 \pm 1.04 \text{ km/s/Mpc} \\
&\text{Planck: } H_0 = 67.4 \pm 0.5 \text{ km/s/Mpc} \\
&\rho = 0.1 \\
\Delta\mu &= 5.64 \\
\sigma_{\text{combined}} &= \sqrt{1.04^2 + 0.5^2 - 2 \times 0.1 \times 1.04 \times 0.5} \approx 1.14 \\
N_\sigma &= \frac{5.64}{1.14} \approx 4.95\sigma
\end{aligned}$$

This shows that Hubble tension is a major tension in modern cosmology.

The sigma thresholds mentioned earlier are widely used in statistical analysis, particularly in physics, cosmology, and other sciences dealing with probabilistic inference. However, their exact interpretation as tension criteria varies across fields. Here are some references that discuss these statistical thresholds:

Particle Physics & Discovery Significance

The 5σ threshold for discovery is standard in particle physics, as established in:

- G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *"Asymptotic formulae for likelihood-based tests of new physics,"* Eur. Phys. J. C **71**, 1554 (2011) [arXiv:1007.1727 [physics.data-an]].

The discovery of the Higgs boson at CERN was based on a 5σ criterion:

- CMS Collaboration, *"Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,"* Phys. Lett. B **716**, 30 (2012) [arXiv:1207.7235 [hep-ex]].

Tension in Cosmology

The usage of 2σ, 3σ, and $> 3\sigma$ as measures of tension is common in cosmological parameter estimation:

- Planck Collaboration, *"Planck 2018 results. VI. Cosmological parameters,"* A&A **641**, A6 (2020) [arXiv:1807.06209 [astro-ph.CO]].
- E. Di Valentino, A. Melchiorri, J. Silk, *"Planck evidence for a closed Universe and a possible crisis for cosmology,"* Nat. Astron. **4**, 196–203 (2020) [arXiv:1911.02087 [astro-ph.CO]].

General Statistical Interpretation

The Gaussian probability interpretation of these thresholds is discussed in:

- B. Efron and R. J. Tibshirani, *"An Introduction to the Bootstrap,"* (Chapman & Hall/CRC, 1994).
- A. Stuart, K. Ord, and S. Arnold, *"Kendall's Advanced Theory of Statistics,"* (Oxford University Press, 1999).