Probability

Probability Distribution Function (PDF)

Prof. Dr. Mostafa Elhosseini

https://youtube.com/drmelhosseini

Probability Density Function (PDF)

- The proportion of the population whose values of X lie between a and b is given by

$$\int_{a}^{b} f(x)\,dx$$

Probability Density Function (PDF)

- The proportion of the population whose values of X lie between a and b is given by

$$\int_{a}^{b} f(x)\, dx$$

- Area under the curve does not depend on whether the endpoints a and b are included in the interval.
- Therefore, probabilities involving X do not depend on whether endpoints are included

Example

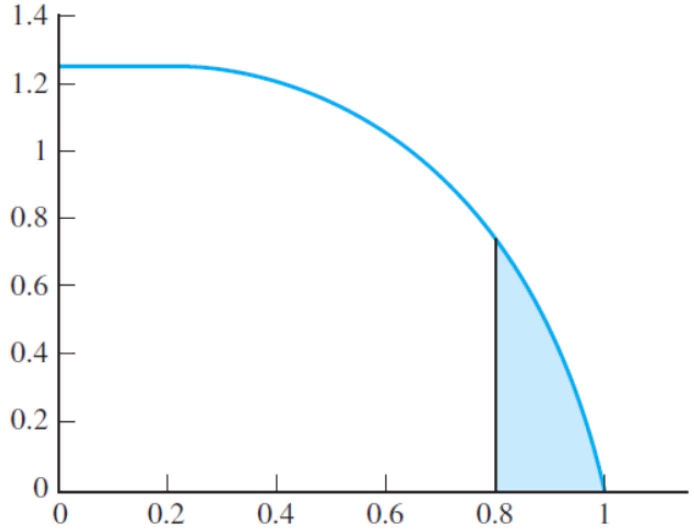
- A **hole is drilled** in a sheet-metal component, and then a **shaft is inserted** through the hole. The **shaft clearance** is equal to the difference between the radius of the hole and the radius of the shaft. Let the random variable **X denote the clearance**, in millimeters. The **probability density function of X** is

$$f(x) = \begin{cases} 1.25(1 - x^4) & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

- Components with clearances larger than **0.8 mm must be scrapped**. **What proportion of components are scrapped**?

Solution

Figure 2.13 (page 104) presents the probability density function of X. Note that the density $f(x)$ is 0 for $x \leq 0$ and for $x \geq 1$. This indicates that the clearances are always between 0 and 1 mm. The proportion of components that must be scrapped is $P(X > 0.8)$, which is equal to the area under the probability density function to the right of 0.8.



This area is given by

$$\begin{aligned}
P(X > 0.8) &= \int_{0.8}^{\infty} f(x)\,dx \\
&= \int_{0.8}^{1} 1.25(1 - x^4)\,dx \\
&= 1.25\left(x - \frac{x^5}{5}\right)\Bigg|_{0.8}^{1} \\
&= 0.0819
\end{aligned}$$

Mean and Variance

Let X be a continuous random variable with probability density function $f(x)$. Then the mean of X is given by

$$\mu_X = \int_{-\infty}^{\infty} x f(x)\, dx \tag{2.35}$$

The mean of X is sometimes called the expectation, or expected value, of X and may also be denoted by $E(X)$ or by μ.

Mean and Variance

Let X be a continuous random variable with probability density function $f(x)$. Then

- The variance of X is given by

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x)\, dx \qquad (2.36)$$

- An alternate formula for the variance is given by

$$\sigma_X^2 = \int_{-\infty}^{\infty} x^2 f(x)\, dx - \mu_X^2 \qquad (2.37)$$

- The variance of X may also be denoted by $V(X)$ or by σ^2.
- The standard deviation is the square root of the variance: $\sigma_X = \sqrt{\sigma_X^2}$.

Example

- A **hole is drilled** in a sheet-metal component, and then a **shaft is inserted** through the hole. The **shaft clearance** is equal to the difference between the radius of the hole and the radius of the shaft. Let the random variable **X denote the clearance**, in millimeters. The **probabilitv densitv function of X** is

$$f(x) = \begin{cases} 1.25(1 - x^4) & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

- Find the mean clearance and the variance of the clearance

Example

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

$$= \int_0^1 x[1.25(1 - x^4)]\, dx$$

$$= 1.25 \left(\frac{x^2}{2} - \frac{x^6}{6} \right) \bigg|_0^1$$

$$= 0.4167$$

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

$$= \int_0^1 x^2[1.25(1 - x^4)]\, dx - (0.4167)^2$$

$$= 1.25 \left(\frac{x^3}{3} - \frac{x^7}{7} \right) \bigg|_0^1 - (0.4167)^2$$

$$= 0.0645$$

Chebyshev's Inequality

- **Chebyshev's inequality** relates the mean and the standard deviation by providing a **bound on the probability** that a random variable takes on a value that **differs from its mean by more than a given multiple of its standard deviation**.
- Specifically, the probability that a random variable differs from its mean by k standard deviations, or more is never greater than $1/k^2$

Chebyshev's Inequality

Chebyshev's Inequality

Let X be a random variable with mean μ_X and standard deviation σ_X. Then

$$P(|X - \mu_X| \geq k\sigma_X) \leq \frac{1}{k^2}$$

Example

- The length of a rivet manufactured by a certain process has mean $\boldsymbol{\mu_x = 50\ mm}$ and standard deviation $\boldsymbol{\sigma_x = 0.45\ mm}$.
- What is the largest possible value for the probability that the length of the rivet is outside the interval 49.1–50.9 mm?

Example

Solution

Let X denote the length of a randomly sampled rivet. We must find $P(X \leq 49.1$ or $X \geq 50.9)$. Now

$$P(X \leq 49.1 \text{ or } X \geq 50.9) = P(|X - 50| \geq 0.9) = P(|X - \mu_X| \geq 2\sigma_X)$$

Applying Chebyshev's inequality with $k = 2$, we conclude that

$$P(X \leq 49.1 \text{ or } X \geq 50.9) \leq \frac{1}{4}$$

Chebyshev's inequality is valid for any random variable and does not require knowledge of the distribution. Because it is so general, the bound given by Chebyshev's inequality is in most cases much greater than the actual probability. Example 2.47 illustrates this.

Example

Assume that the probability density function for X, the length of a rivet in Example 2.46, is

$$f_X(x) = \begin{cases} [477 - 471(x-50)^2]/640 & 49 \le x \le 51 \\ 0 & \text{otherwise} \end{cases}$$

It can be verified that $\mu_X = 50$ and $\sigma_X = 0.45$. Compute the probability that the length of the rivet is outside the interval 49.1–50.9 mm. How close is this probability to the Chebyshev bound of $1/4$?

Example

Solution

$$
\begin{aligned}
P(X \leq 49.1 \text{ or } X \geq 50.9) &= 1 - P(49.1 < X < 50.9) \\
&= 1 - \int_{49.1}^{50.9} \frac{477 - 471(x-50)^2}{640}\, dx \\
&= 1 - \left. \frac{477x - 157(x-50)^3}{640} \right|_{49.1}^{50.9} \\
&= 0.01610
\end{aligned}
$$

The actual probability is much smaller than the Chebyshev bound of $1/4$.

Because the Chebyshev bound is generally much larger than the actual probability, it should only be used when the distribution of the random variable is unknown. When the distribution is known, then the probability density function or probability mass function should be used to compute probabilities.

References

- Statistics for Engineers and Scientists, Fourth Edition, William Navidi Colorado School of Mines
- Agresti, A., Franklin, C. and Klingenberg, B., 2017. Statistics: The art and science of learning from data (Fourth Edi).
- Khan Academy