ECS315 2014/1 Part IV.1 Dr.Prapun

10 Continuous Random Variables

10.1 From Discrete to Continuous Random Variables

In many practical applications of probability, physical situations are better described by random variables that can take on a *continuum* of possible values rather than a *discrete* number of values. For this type of random variable, the interesting fact is that

- any individual value has probability zero:

$$\text{pmf} \rightarrow P[X = x] = 0 \quad \text{for all } x \tag{18}$$

  and that

- the support is always uncountable.

These random variables are called **continuous random variables**.

First implication:

10.1. We can see from (18) that the pmf is going to be useless for this type of random variable. It turns out that the cdf F_X is still useful and we shall introduce another useful function called *probability density function* (pdf) to replace the role of pmf. However, integral calculus[36] is required to formulate this continuous analog of a pmf. pmf → pdf

10.2. In some cases, the random variable X is actually discrete but, because the range of possible values is so large, it might be more convenient to analyze X as a continuous random variable.

[36]This is always a difficult concept for the beginning student.

Second implication: Talking about $P[X=x]$ is useless ...
So, we usually talk about probability of the form $P[a \le X \le b]$.

Example 10.3. Suppose that current measurements are read from a digital instrument that displays the current to the nearest one-hundredth of a mA. Because the possible measurements are limited, the random variable is discrete. However, it might be a more convenient, simple approximation to assume that the current measurements are values of a continuous random variable.

Example 10.4. If you can measure the heights of people with infinite precision, the height of a randomly chosen person is a continuous random variable. In reality, heights cannot be measured with infinite precision, but the mathematical analysis of the distribution of heights of people is greatly simplified when using a mathematical model in which the height of a randomly chosen person is modeled as a continuous random variable. [17, p 284]

Example 10.5. Continuous random variables are important models for

(a) voltages in communication receivers

(g) noise in communication system

(b) file download times on the Internet

(c) velocity and position of an airliner on radar

(d) lifetime of a battery

(e) decay time of a radioactive particle

(f) time until the occurrence of the next earthquake in a certain region

Example 10.6. The simplest example of a continuous random variable is the "random choice" of a number from the interval $(0, 1)$.

- In `MATLAB`, this can be generated by the command `rand`. In `Excel`, use `rand()`.
- The generation is "unbiased" in the sense that "any number in the range is as likely to occur as another number."
- Histogram is flat over $(0, 1)$.
- Formally, this is called a uniform RV on the interval $(0, 1)$.

Definition 10.7. We say that X is a **continuous random variable**[37] if we can find a (real-valued) function[38] f such that, for any set B, $P[X \in B]$ has the form

$$P[X \in B] = \int_B f(x)dx. \tag{19}$$

$P[\text{some condition(s) on } x] = \int f(x)\, dx$

integrate over all the x that satisfies the condition(s).

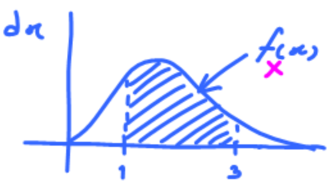
- In particular,

$P[|x| < 2] = P[-2 < x < 2] = \int_{-2}^{2} f(x)\, dx$

$$P[a \le X \le b] = \int_a^b f(x)dx. \tag{20}$$

In other words, the **area under the graph** of $f(x)$ between the points a and b gives the probability $P[a \le X \le b]$.

$P[1 < x < 3] = \int_1^3 f(x)\, dx$



$P[x^2 > 1] = P[x > 1 \text{ or } x < -1] = \int_1^{\infty} f(x)\, dx + \int_{-\infty}^{-1} f(x)\, dx$

$P[x < 3] = \int_{-\infty}^{3} f(x)\, dx$

- The function f is called the ***probability density function*** (pdf) or simply ***density***.

- When we want to emphasize that the function f is a density of a particular random variable X, we write f_X instead of f.

[37]To be more rigorous, this is the definition for *absolutely* continuous random variable. At this level, we will not distinguish between the continuous random variable and absolutely continuous random variable. When the distinction between them is considered, a random variable X is said to be continuous (not necessarily absolutely continuous) when condition (18) is satisfied. Alternatively, condition (18) is equivalent to requiring the cdf F_X to be continuous. Another fact worth mentioning is that if a random variable is absolutely continuous, then it is continuous. So, absolute continuity is a stronger condition.

[38]Strictly speaking, δ-"function" is not a function; so, can't use δ-function here.

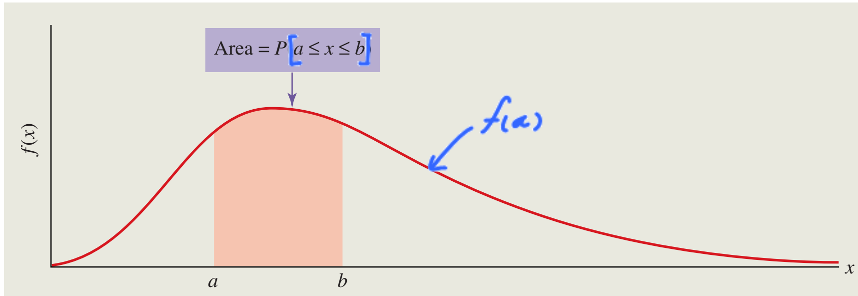
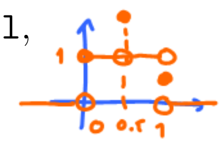
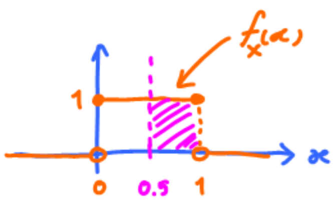


Figure 13: For a continuous random variable, the probability distribution is described by a curve called the probability density function, $f(x)$. The total area beneath the curve is 1.0, and the probability that X will take on some value between a and b is the area beneath the curve between points a and b.

Example 10.8. For the random variable generated by the `rand` command in `MATLAB`[39] or the `rand()` command in `Excel`,





$P[X > 0.5] = 0.5$

$$= \int_{0.5}^{\infty} f_X(x)\,dx = \int_{0.5}^{1} \underbrace{f_X(x)}_{1}\,dx + \cancel{\int_{1}^{\infty} \underbrace{f_X(x)}_{0}\,dx} = x\Big|_{0.5}^{1} = 1 - 0.5 = 0.5$$

Definition 10.9. Recall that the support S_X of a random variable X is any set S such that $P[X \in S] = 1$. For continuous random variable, S_X is usually set to be $\{x : f_X(x) > 0\}$.

Ex. For X in Ex 10.8, support $S_X = [0,1]$.

Ex. For noise in communication system,

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$$

[39]The `rand` command in `MATLAB` is an approximation for two reasons:

(a) It produces pseudorandom numbers; the numbers seem random but are actually the output of a deterministic algorithm.

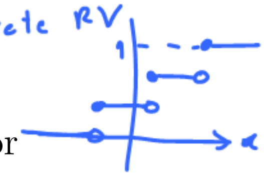
(b) It produces a double precision floating point number, represented in the computer by 64 bits. Thus MATLAB distinguishes no more than 2^{64} unique double precision floating point numbers. By comparison, there are uncountably infinite real numbers in the interval from 0 to 1.

10.2 Properties of PDF and CDF for Continuous Random Variables

10.10. f_X is determined only almost everywhere[40]. That is, given a pdf f for a random variable X, if we construct a function g by changing the function f at a countable number of points[41], then g can also serve as a pdf for X.

10.11. The cdf of any kind of random variable X is defined as

For discrete RV



$$F_X(x) = P[X \le x].$$

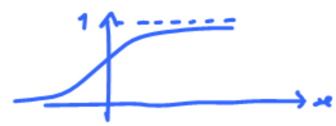
Note that even though there are more than one valid pdfs for any given random variable, the cdf is unique. There is only one cdf for each random variable.

pdf → cdf

10.12. For continuous random variable, given the pdf $f_X(x)$, we can find the cdf of X by

$F_X(5) = P[X \le 5] = \int_{-\infty}^{5} f_X(x)dx$

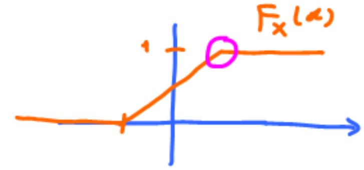
$$F_X(x) = P[X \le x] = \int_{-\infty}^{x} f_X(t)dt.$$



For cont. RV, the cdf is continuous.

cdf → pdf

10.13. Given the cdf $F_X(x)$, we can find the pdf $f_X(x)$ by



- If F_X is differentiable at x, we will set

$$\frac{d}{dx}F_X(x) = f_X(x).$$

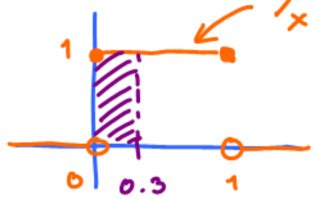
- If F_X is not differentiable at x, we can set the values of $f_X(x)$ to be any value. Usually, the values are selected to give simple expression. (In many cases, they are simply set to 0.)

[40] Lebesgue-a.e, to be exact

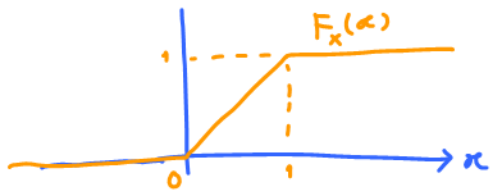
[41] More specifically, if $g = f$ Lebesgue-a.e., then g is also a pdf for X.

pdf → cdf

Example 10.14. For the random variable generated by the `rand` command in `MATLAB` or the `rand()` command in `Excel`,



$F_X(x) = P[X \le x] = \int_{-\infty}^{x} f_X(t)\,dt$

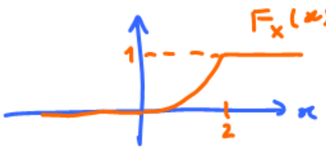


$F_X(-1) = 0$

$F_X(0.3) = 0.3$

cdf → pdf

Example 10.15. Suppose that the lifetime X of a device has the cdf



$$F_X(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{4}x^2, & 0 \le x \le 2 \\ 1, & x > 2 \end{cases} \xrightarrow{\frac{d}{dx}} f_X(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{2}x, & 0 < x < 2 \\ 0, & x > 2 \end{cases}$$

Observe that it is differentiable at each point x except at $x = 2$. The probability density function is obtained by differentiation of the cdf which gives

$$f_X(x) = \begin{cases} \frac{1}{2}x, & 0 < x < 2 \\ 0, & \text{otherwise.} \end{cases}$$

$$f_X(x) = \begin{cases} \frac{1}{2}x, & 0 < x < 2 \\ 0, & \text{otherwise} \end{cases}$$

At $x = 2$ where F_X has no derivative, it does not matter what values we give to f_X. Here, we set it to be 0.

10.16. In many situations when you are asked to find pdf, it may be easier to find cdf first and then differentiate it to get pdf.

Exercise 10.17. A point is "picked at random" in the inside of a circular disk with radius r. Let the random variable X denote the distance from the center of the disk to this point. Find $f_X(x)$.

10.18. Unlike the cdf of a discrete random variable, the cdf of a continuous random variable has no jump and is continuous everywhere.

10.19. $p_X(x) = P[X = x] = P[x \le X \le x] = \int_x^x f_X(t)dt = 0.$

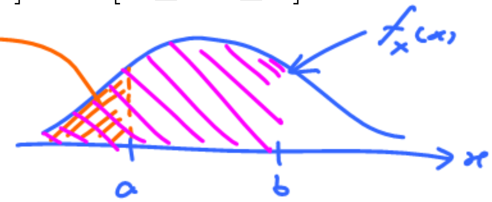
Again, it makes no sense to speak of the probability that X will take on a pre-specified value. This probability is always zero.

10.20. $P[X = a] = P[X = b] = 0$. Hence,

$$P[a < X < b] = P[a \le X < b] = P[a < X \le b] = P[a \le X \le b] = F_X(b) - F_X(a)$$

$F_X(a) = P[X \le a] = \int_{-\infty}^{a} f_X(x)\,dx$

$F_X(b) = P[X \le b] = \int_{-\infty}^{b} f_X(x)\,dx$



- The corresponding integrals over an interval are not affected by whether or not the endpoints are included or excluded.
- When we work with continuous random variables, it is usually not necessary to be precise about specifying whether or not a range of numbers includes the endpoints. This is quite different from the situation we encounter with discrete random variables where it is critical to carefully examine the type of inequality.

If X is discrete, we have pmf $p_X(x)$
1) $p_X(x) \geq 0$
2) $\sum_x p_X(x) = 1$

10.21. f_X is nonnegative a.e. and $\int_{-\infty}^{\infty} f_X(x)dx = 1$.

Example 10.22. Random variable X has pdf

$$f_X(x) = \begin{cases} ce^{-2x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

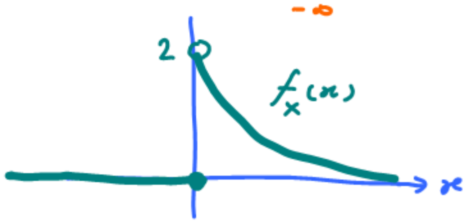
Find the constant c and sketch the pdf.

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

$$\int_{-\infty}^{\infty} f_X(x)dx = \cancel{\int_{-\infty}^{0} \underbrace{f_X(x)}_{0}dx} + \int_{0}^{\infty} \underbrace{f_X(x)}_{ce^{-2x}}dx$$

$$= \int_0^{\infty} ce^{-2x}dx = c\left.\frac{e^{-2x}}{-2}\right|_0^{\infty} = \frac{c}{-2}(0-1) = \frac{c}{2} = 1$$

$$c = 2$$

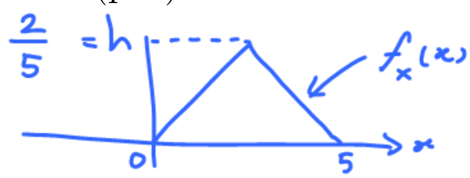


Definition 10.23. A continuous random variable is called ***exponential*** if its pdf is given by

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0 \end{cases}$$

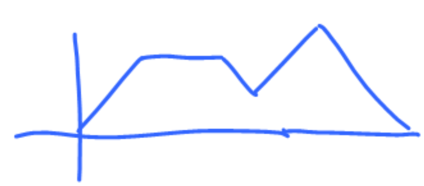
for some $\lambda > 0$

Theorem 10.24. Any nonnegative[42] function that integrates to one is a ***probability density function*** (pdf) of some random variable [8, p.139].



[42]or nonnegative a.e.

$\frac{1}{2} \times h \times 5 = 1$



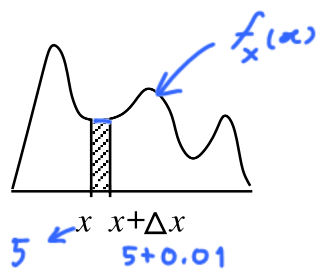
10.25. Intuition/Interpretation:

The use of the word "density" originated with the analogy to the distribution of matter in space. In physics, any finite volume, no matter how small, has a positive mass, but there is no mass at a single point. A similar description applies to continuous random variables.

Approximately, for a small Δx,

$$P\left[X \in [x, x + \Delta x]\right] = \int_x^{x+\Delta x} f_X(t)dt \approx f_X(x)\Delta x.$$

This is why we call f_X the density function.



$$P[5 < X < 5.01] = \int_5^{5.01} f_X(x)\,dx$$
$$\approx f_X(5) \times 0.01$$

Figure 14: $P[x \le X \le x + \Delta x]$ is the area of the shaded vertical strip.

In other words, the probability of random variable X taking on a value in a *small* interval around point c is approximately equal to $f(c)\Delta c$ when Δc is the length of the interval.

- In fact, $f_X(x) = \lim\limits_{\Delta x \to 0} \frac{P[x < X \le x + \Delta x]}{\Delta x}$
- The number $f_X(x)$ itself is ***not a probability***. In particular, it does not have to be between 0 and 1.
- $f_X(c)$ is a relative measure for the likelihood that random variable X will take on a value in the immediate neighborhood of point c.

  Stated differently, the pdf $f_X(x)$ expresses how densely the probability mass of random variable X is smeared out in the neighborhood of point x. Hence, the name of density function.

10.26. Histogram and pdf [17, p 143 and 145]:

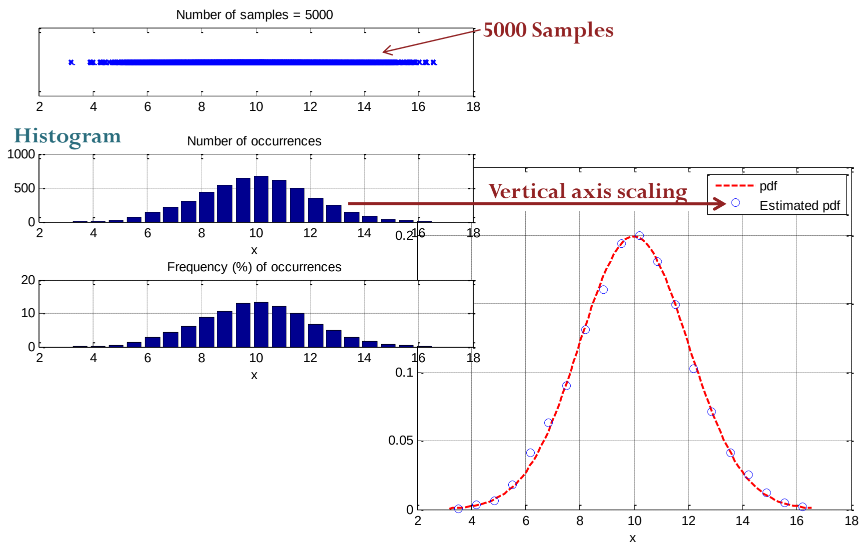


Figure 15: From histogram to pdf.

(a) A probability **histogram** is a bar chart that divides the range of values covered by the samples/measurements into intervals of the same width, and shows the proportion (relative frequency) of the samples in each interval.

- To make a histogram, you break up the range of values covered by the samples into a number of disjoint adjacent intervals each having the same width, say width Δ. The height of the bar on each interval $[j\Delta, (j+1)\Delta)$ is taken such that the area of the bar is equal to the proportion of the measurements falling in that interval (the proportion of measurements within the interval is divided by the width of the interval to obtain the height of the bar).
- The total area under the probability histogram is thus standardized/normalized to one.

(b) If you take sufficiently many independent samples from a continuous random variable and make the width Δ of the base intervals of the probability histogram smaller and smaller, the graph of the probability histogram will begin to look more and more like the pdf.

(c) Conclusion: A probability density function can be seen as a "smoothed out" version of a probability histogram

and normalized

10.3 Expectation and Variance

10.27. ***Expectation***: Suppose X is a continuous random variable with probability density function $f_X(x)$.

Recall, for discrete RV,

$\mathbb{E}X = \sum_x x\, p_X(x)$

$\mathbb{E}[g(X)] = \sum_x g(x) p_X(x)$

$$\mathbb{E}X = \int_{-\infty}^{\infty} x f_X(x) dx \tag{21}$$

$$\mathbb{E}\left[g(X)\right] = \int_{-\infty}^{\infty} g(x) f_X(x) dx \tag{22}$$

In particular,

$\mathbb{E}[X^2] = \sum_x x^2 p_X(x)$

$$\mathbb{E}\left[X^2\right] = \int_{-\infty}^{\infty} x^2 f_X(x) dx$$

$$\text{Var}\, X = \int_{-\infty}^{\infty} (x - \mathbb{E}X)^2 f_X(x) dx = \mathbb{E}\left[X^2\right] - (\mathbb{E}X)^2.$$

$\sigma_X = \sqrt{\text{Var}\, X}$

X_0

Example 10.28. For the random variable generated by the `rand` command in `MATLAB` or the `rand()` command in `Excel`,

$f_X(x) =$

$\mathbb{E}X = \int_{-\infty}^{\infty} x f_X(x)\, dx = \int_0^1 x \times 1\, dx$

$= \int_0^1 x\, dx = \left.\frac{x^2}{2}\right|_0^1 = \frac{1^2}{2} - \cancel{\frac{0^2}{2}} = \frac{1}{2}$

$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx = \int_0^1 x^2 \times 1\, dx = \int_0^1 x^2 dx = \left.\frac{x^3}{3}\right|_0^1 = \frac{1}{3}$

$\text{Var}\, X = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$ $\quad \sigma_X = \frac{1}{\sqrt{12}} = \frac{1}{2\sqrt{3}}$

Example 10.29. For the exponential random variable introduced in Definition 10.23,

$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & \text{otherwise} \end{cases}$

$\mathbb{E}X = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^{\infty} x \lambda e^{-\lambda x} dx = \lambda \int_0^{\infty} x e^{-\lambda x} dx = \lambda \left(-\frac{x}{\lambda} e^{-\lambda x} - \frac{1}{\lambda^2} e^{-\lambda x} \right)\Big|_0^{\infty} = ?$

$= \lambda\left(0 - \left(-\frac{1}{\lambda^2}\right)\right)$

$= \frac{1}{\lambda}$

$$\begin{array}{c|cc|c} & x & e^{-\lambda x} & \\ \frac{d}{dx} & 1 & \frac{1}{-\lambda} e^{-\lambda x} & \int \\ & 0 & \frac{1}{(-\lambda)^2} e^{-\lambda x} & \end{array}$$

$+$ $\quad -$

$$\lim_{x\to\infty} x e^{-\lambda x} = \lim_{x\to\infty} \frac{x}{e^{\lambda x}} \underset{\uparrow}{=} \lim_{x\to\infty} \frac{1}{\lambda e^{\lambda x}} = 0$$

L'Hospital

10.30. If we compare other characteristics of discrete and continuous random variables, we find that with discrete random variables, many facts are expressed as sums. With continuous random variables, the corresponding facts are expressed as integrals.

10.31. All of the properties for the expectation and variance of discrete random variables also work for continuous random variables as well:

(a) Intuition/interpretation of the expected value: As $n \to \infty$, the average of n independent samples of X will approach $\mathbb{E}X$. This observation is known as the "Law of Large Numbers".

(b) For $c \in \mathbb{R}$, $\mathbb{E}[c] = c$

(c) For constants a, b, we have $\mathbb{E}[aX + b] = a\mathbb{E}X + b$.

(d) $\mathbb{E}\left[\sum_{i=1}^{n} c_i g_i(X)\right] = \sum_{i=1}^{n} c_i \mathbb{E}[g_i(X)]$.

(e) $\operatorname{Var} X = \mathbb{E}\left[X^2\right] - (\mathbb{E}X)^2$

(f) $\operatorname{Var} X \geq 0$.

(g) $\operatorname{Var} X \leq \mathbb{E}\left[X^2\right]$.

(h) $\operatorname{Var}[aX + b] = a^2 \operatorname{Var} X$.

(i) $\sigma_{aX+b} = |a|\,\sigma_X$.

10.32. ***Chebyshev's Inequality***:

$$P\left[|X - \mathbb{E}X| \geq \alpha\right] \leq \frac{\sigma_X^2}{\alpha^2}$$

or equivalently

$$P\left[|X - \mathbb{E}X| \geq n\sigma_X\right] \leq \frac{1}{n^2}$$

- This inequality use variance to bound the "tail probability" of a random variable.
- Useful only when $\alpha > \sigma_X$

Example 10.33. A circuit is designed to handle a current of 20 mA plus or minus a deviation of less than 5 mA. If the applied current has mean 20 mA and variance 4 mA^2, use the Chebyshev inequality to bound the probability that the applied current violates the design parameters.

Let X denote the applied current. Then X is within the design parameters if and only if $|X - 20| < 5$. To bound the probability that this does not happen, write

$$P\left[|X-20| \geq 5\right] \leq \frac{\operatorname{Var} X}{5^2} = \frac{4}{25} = 0.16.$$

Hence, the probability of violating the design parameters is at most 16%.

10.34. Interesting applications of expectation:

(a) $f_X(x) = \mathbb{E}\left[\delta\left(X - x\right)\right]$

(b) $P\left[X \in B\right] = \mathbb{E}\left[1_B(X)\right]$

Tutorial Example

Discrete RV

Use pmf $p_X(x)$

$$F_X(x) = P[X \le x] = \sum_{t:\,t\le x} p_X(t)$$

Suppose $p_X(x) = \begin{cases} cx^2, & x \in \{1,2\}, \\ 0, & \text{otherwise.} \end{cases}$

To find c, we know that

$$\sum_x p_X(x) = 1.$$

Therefore, $c(1)^2 + c(2)^2 = 1$

$$c = \frac{1}{5}.$$

So, $p_X(x) = \begin{cases} 1/5, & x = 1, \\ 4/5, & x = 2, \\ 0, & \text{otherwise.} \end{cases}$

$$\mathbb{E}X = \sum_x x\, p_X(x) = 1 \times \frac{1}{5} + 2 \times \frac{4}{5} = \frac{9}{5}$$

$$\mathbb{E}[X^2] = \sum_x x^2 p_X(x) = 1^2 \times \frac{1}{5} + 2^2 \times \frac{4}{5} = \frac{17}{5}$$

$$\operatorname{Var} X = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = \frac{17}{5} - \left(\frac{9}{5}\right)^2 = \frac{85-81}{25} = \frac{4}{25}$$

$$\sigma_X = \sqrt{\operatorname{Var} X} = \frac{2}{5}$$

Continuous RV

Use pdf $f_X(x)$

$$F_X(x) = P[X \le x] = \int_{-\infty}^{x} f_X(t)\,dt$$

Suppose $f_X(x) = \begin{cases} cx^2, & x \in (1,2), \\ 0, & \text{otherwise.} \end{cases}$

To find c, we know that

$$\int_{-\infty}^{\infty} f_X(x)\,dx = 1.$$

Therefore,

$$\int_1^2 cx^2\,dx = c\left.\frac{x^3}{3}\right|_1^2 = c\,\frac{7}{3} = 1$$

$$c = \frac{3}{7}.$$

So,

$$f_X(x) = \begin{cases} \frac{3}{7}x^2, & x \in (1,2), \\ 0, & \text{otherwise.} \end{cases}$$

$$\mathbb{E}X = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_1^2 x\,\frac{3}{7}x^2\,dx = \frac{3}{7}\left.\frac{x^4}{4}\right|_1^2$$

$$= \frac{3 \times 15}{7 \times 4} = \frac{45}{28}.$$

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx = \int_1^2 x^2\,\frac{3}{7}x^2\,dx$$

$$= \frac{3}{7}\left.\frac{x^5}{5}\right|_1^2 = \frac{3 \times 31}{7 \times 5} = \frac{93}{35}.$$

$$\operatorname{Var} X = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = \frac{93}{35} - \left(\frac{45}{28}\right)^2 = \frac{291}{3920}$$

$$\approx 0.0742$$

$$\sigma_X = \sqrt{\operatorname{Var} X} \approx 0.2725$$