Sirindhorn International Institute of Technology

Thammasat University

School of Information, Computer and Communication Technology

ECS315 2017/1 Part IV.1 Dr.Prapun

10 Continuous Random Variables

10.1 From Discrete to Continuous Random Variables

10.1. 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 the random variables to be discussed in this section,

- any individual value has probability zero:

$P[X=5]=0$

$P[X=\pi]=0$

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

$P[X=\sqrt{2}]=0$

and that

- the supports are always uncountable.

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

10.2. Implications:

(a) 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[38] is required to formulate this continuous analog of a pmf.

pmf → pdf

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

(b) Because talking about $P[X = x]$ for continuous RV is useless (always 0), we instead talk about the probability that the RV is in some interval, e.g. $P[a < X < b]$.

10.3. 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.

Example 10.4. 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 10.5. 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. [21, p 284]

Example 10.6. Continuous random variables are important models for

(a) voltages in communication receivers

(b) file download times on the Internet

(c) velocity and position of an airliner on radar

(d) lifetime of a battery

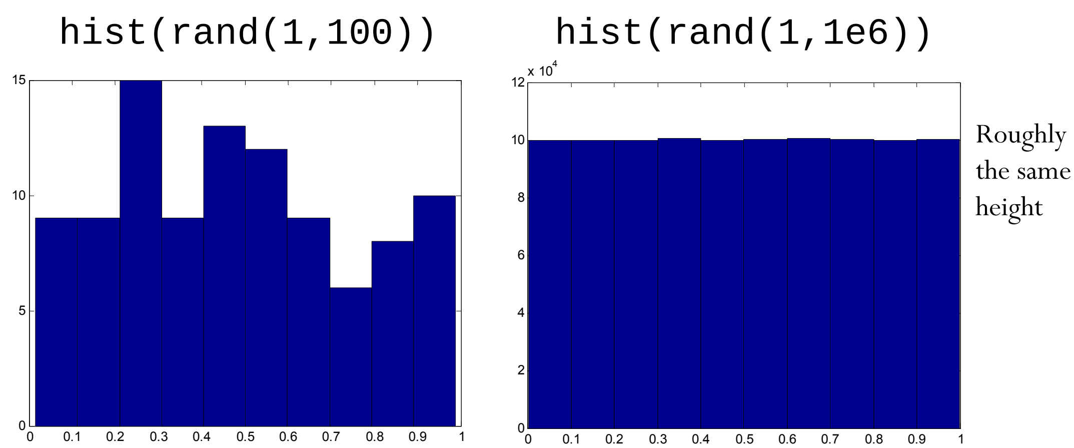
(e) decay time of a radioactive particle

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

(g) noise in communication systems

Example 10.7. 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 (0,1) is as likely to occur as another number."
- Histogram is flat over $(0,1)$ in the limit as the number of samples increases to infinity regardless of the number of bins as long as the bins have the same size. See Figure 19b.
- Formally, this is called a uniform RV on the interval $(0,1)$.



(a) (b)

Figure 19: Histogram of the values generated by MATLAB command `rand`.

Example 10.8. Put a piece of (unit-width and unit-height) paper outdoor. Mark the location of the (center of) first drop of rain on it. Record its horizontal position).

Example 10.9. In MATLAB, there are other commands (such as `randn`) and ways to generate continuous random variables with other shapes of histograms.

Definition 10.10. We say that X is a **continuous random variable**[39] if we can find a (real-valued) function[40] f_X such that, for any set B, $P[X \in B]$ has the form

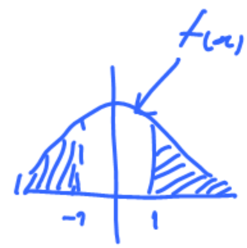
$$P[X \in B] = \int_B f_X(x)dx. \tag{19}$$

Equivalently,

$$P[\text{some condition(s) on } X] = \int_{\substack{\{\text{all the } x \text{ values that} \\ \text{satisfy the condition(s)}\}}} f_X(x)\,dx$$

- In particular,

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



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

$$P[1 < X < 3] = \int_1^3 f_X(x)\,dx \qquad P[X^2 > 1] = \int_{\{x : x^2 > 1\}} f_X(x)\,dx = \int_{(-\infty,-1)\cup(1,\infty)} f_X(x)\,dx$$

$$P[X < 3] = \int_{-\infty}^{3} f_X(x)\,dx \qquad = \int_{-\infty}^{-1} f_X(x)\,dx + \int_1^{\infty} f_X(x)\,dx$$

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

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

[39] 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.

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

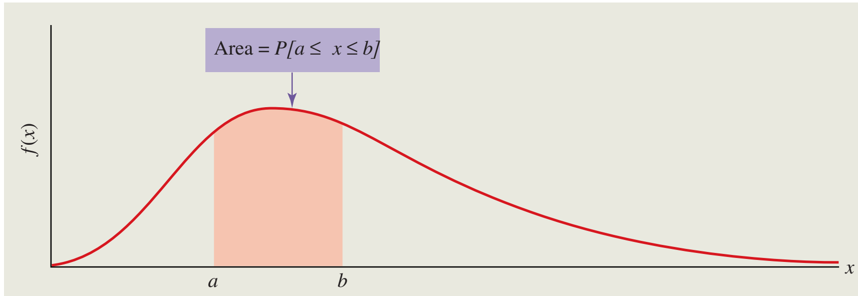
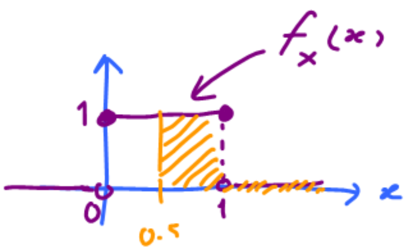


Figure 20: For a continuous random variable, the probability distribution is described by a curve called the probability density function, $f(x)$. The total area beneath the curve is 1.0, and the probability that X will take on some value between a and b is the area beneath the curve between points a and b.

Example 10.11. For the random variable generated by the `rand` command in `MATLAB`[41] or the `rand()` command in `Excel`,



$$P[X > 0.5] = \int_{0.5}^{\infty} f_X(x)\,dx = \int_{0.5}^{1} \overbrace{f_X(x)}^{1}\,dx + \cancel{\int_{1}^{\infty} \overbrace{f_X(x)}^{0}\,dx}$$

$$= x\Big|_{0.5}^{1} = 1 - 0.5 = 0.5$$

Definition 10.12. 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 $\{x : f_X(x) > 0\}$.

Example 10.13. For the random variable X in Example 10.11,

$$S_X = [0, 1]$$

Example 10.14. For noise in communication systems,

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

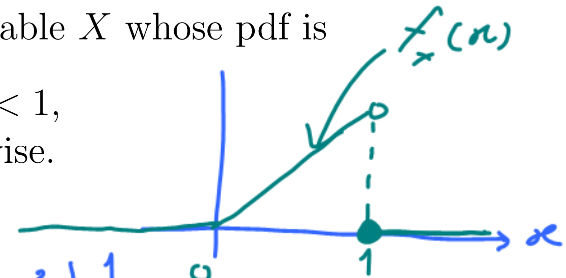
$$S_X = (-\infty, \infty)$$

[41]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.

Example 10.15. Consider a random variable X whose pdf is



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

(a) Find $P[X > 0.5]$.

$$= \int_{0.5}^{\infty} f_X(x)\,dx = \int_{0.5}^{1} 2x\,dx = \frac{\cancel{2}x^2}{\cancel{2}}\Big|_{0.5}^{1} = 1^2 - 0.5^2 = 0.75$$

(b) Find $P[0.2 < X < 0.3]$.

$$= \int_{0.2}^{0.3} f_X(x)\,dx = x^2\Big|_{0.2}^{0.3} = 0.3^2 - 0.2^2 = 0.05$$

(c) Find $P[0.19 < X < 0.21]$. $\approx f_X(0.2)\,\Delta x = (2\times 0.2)\,0.02 = 0.008$

$$= \int_{0.19}^{0.21} f_X(x)\,dx = \int_{0.19}^{0.21} 2x\,dx = x^2\Big|_{0.19}^{0.21} = 0.21^2 - 0.19^2 = 0.008$$

(d) Find $P[0.79 < X < 0.81]$. $\approx f_X(0.8)\,\Delta x = (2\times 0.8)\times 0.02 = 0.032$

$$= \int_{0.79}^{0.81} f_X(x)\,dx = x^2\Big|_{0.79}^{0.81} = 0.81^2 - 0.79^2 = 0.032$$

Observation: From the pdf expression, we know that $f_X(0.8) > f_X(0.2)$.

(a) Does this imply $P[X = 0.8] > P[X = 0.2]$? No! From (18), we know that both probabilities are 0.

(b) $f_X(0.8) > f_X(0.2)$ simply means the RV X is more likely to be in the small interval around 0.8 than in the small interval (of the same length) around 0.2. In fact, the ratio of the two probabilities is approximately the ratio of the pdf values.

10.16. 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,

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

This is why we call f_X the density function.

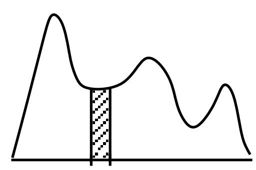


Figure 21: $P[x \le X \le x + \Delta x]$ is the area of the shaded vertical strip.

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

10.17. Histogram and pdf [21, p 143 and 145]:

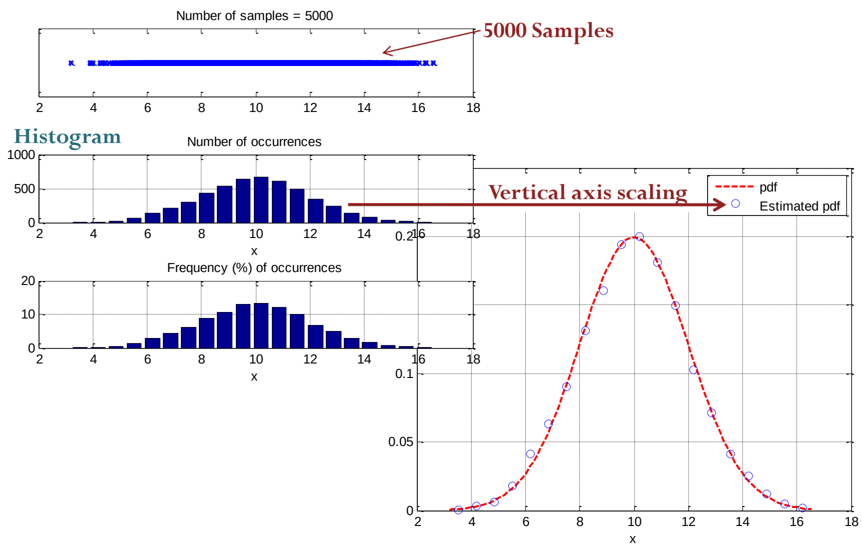


Figure 22: From histogram to pdf.

(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, 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" normalized version of a (probability) histogram