Chapter 9

Tuesday, November 7, 2023 5:28 PM

Chapter 9

Continuous Distributions

Review

So far, we have considered discrete random variables.

- **Binomial:** number of successes in n independent trials, each with probability of success π. $Binom(n, \pi)$
- **Geometric:** number of failures before first success in independent trials each with probability of success π. $Geom(\pi)$
- **Poisson:** number of occurrences of a random event in a fixed time period or space. $Poisson(\lambda)$

Review

We have seen the following definitions in this context.

- **PMF:** f
$$f(x) = P(X = x)$$
- **Expected value:** μ
$$E[X] = \sum_x x \cdot f(x).$$
- **Variance:** σ^2
$$Var[X] = \sum_x (x - \mu)^2 \cdot f(x)$$
- **Standard deviation:** $\sigma = \sqrt{Var[X]}$

Review

We have proved the following rules for expectations for discrete random variables.

- Law of the unconscious probabilist

$$E[t(X)] = \sum_x t(x)f(x).$$

- Linearity of expectation: $E[aX + b] = aE[X] + b$.
- Non-linearity of variance: $Var[aX + b] = a^2 Var[X]$.

Continuous distributions

A continuous random variable is a random variable which (in theory) can take on any real value in some interval.

The fact that a continuous random variable takes uncountably many values eliminates the possibility of assigning a probability to each possible value, as we did in the discrete case.

We return to the key idea behind probability histograms:

$$area = probability.$$

From discrete to continuous

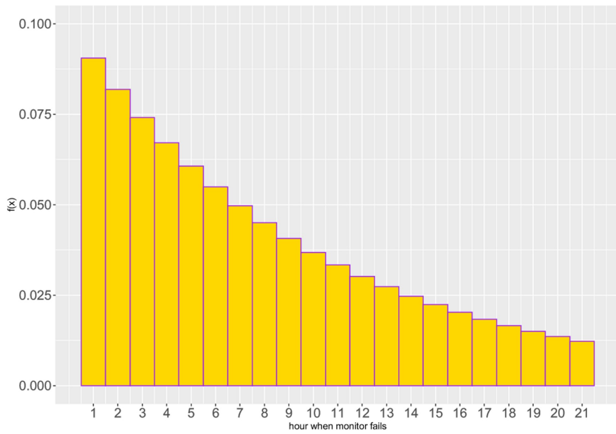
Suppose an electronic surveillance monitor is turned on briefly at the beginning of every hour and has a 0.905 probability of working properly, regardless of how long it has remained in service.

If we let the random variable X denote the hour at which the monitor first fails, then:

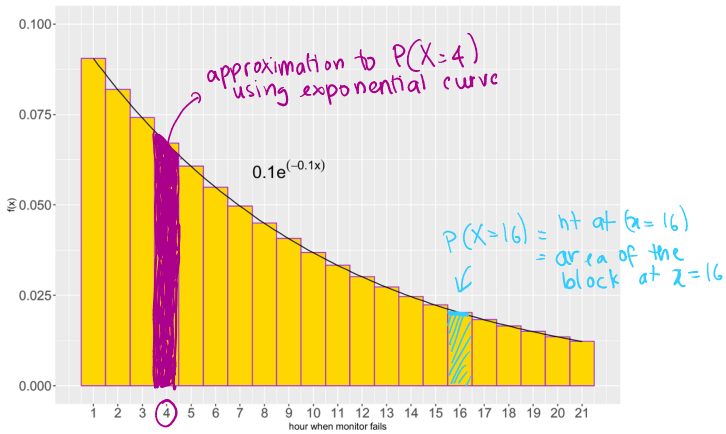
$$\begin{aligned} f(x) = P(X = x) &= \text{P(Monitor works for x-1 hours and then fails)} \\ &= (0.905)^{x-1} \cdot 0.095, \qquad x = 1, 2, 3, \ldots \end{aligned}$$

note: we are assuming that whether or not it fails is independent from hour to hour to write the PMF

Probability histogram of X



Now take a look at the graph below, where the exponential curve $0.1e^{-0.1x}$ is superimposed on the graph.



Notice how closely the area under the curve approximates the area of the blocks.

It follows that the probability that X lies in some given interval is numerically similar to integral of the exponential curve above that same interval.

Implicit in the similarity here is our sought-after alternative to probability assignments for continuous sample spaces.

Instead of defining probabilities for individual points, we will define probabilities for intervals of points as areas under the graph of some function, where the shape of the function will reflect the desired probability "measure" to be associated with the sample space

Probability Density Function (PDF)

Definition 9.1 A **Probability Density Function** (PDF) is any function f such that

$$f(x) \geq 0, \forall\, x \in \mathbb{R}$$

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

The continuous random variable X defined by the PDF $f(x)$ satisfies

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx. =$$ area of $[a, b]$ under $f(x)$

The set of real numbers where the PDF is strictly positive - the range of the random variable - is called the *support* of the distribution.

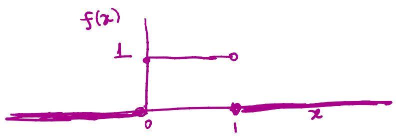
Common Density functions

random variable - is called the *support* of the distribution.

Common Density functions

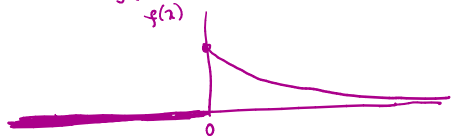
Uniform

$f(x) = 1 \quad 0 \leq x < 1$



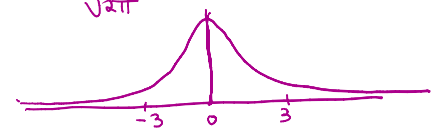
Exponential

$f(x) = e^{-x}, \quad 0 \leq x < \infty$



Std. Normal

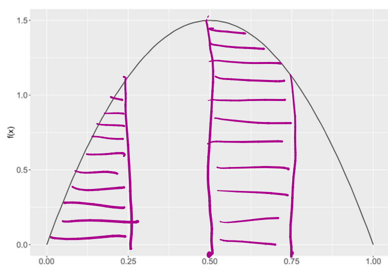
$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad -\infty < x < \infty$



Example 9.1

Consider the function:

$$f(x) = 6x(1-x), \quad 0 \leq x < 1.$$



Verify that this function is a legitimate PDF and use it to calculate $P(0 \leq X \leq \frac{1}{4})$ as well as $P(\frac{1}{2} \leq X \leq \frac{3}{4})$.

$f(x) = 6x(1-x) \quad 0 \leq x < 1$

- Is this a valid PDF? YES
  - Is $f(x) \overset{?}{\geq} 0$ for every x. ✓
  - Is $\int_{-\infty}^{\infty} f(x)\, dx \overset{?}{=} 1$ ✓

$$\int_{-\infty}^{\infty} f(x)\, dx = \int_0^1 6x(1-x)\, dx = 6\int_0^1 (x - x^2)\, dx$$

$$= 6 \cdot \left(\frac{x^2}{2} - \frac{x^3}{3}\right)\Big|_0^1 = 6 \cdot \left(\frac{1}{2} - \frac{1}{3}\right) = 6 \cdot \frac{1}{6} = 1$$

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx = 6\int_a^b (x - x^2)\, dx$$

$$= 6\left(\frac{x^2}{2} - \frac{x^3}{3}\right)\Big|_a^b = 6\left[\frac{b^2}{2} - \frac{b^3}{3} - \left(\frac{a^2}{2} - \frac{a^3}{3}\right)\right]$$

(double check)

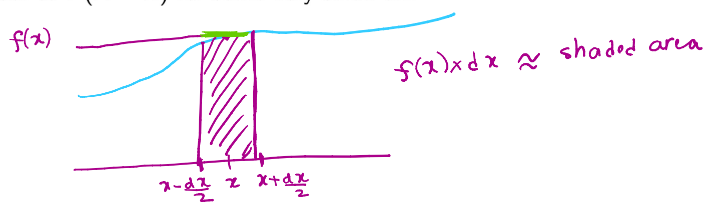
$\therefore P(0 \leq X \leq \frac{1}{4}) = 5/32$

$P(\frac{1}{2} \leq X \leq \frac{3}{4}) = 33/96$

A PDF is not a probability. For instance, it would be incorrect to say

$$f(x) = P(X = x).$$

However, so long as f is continuous at x, you can think of $f(x)\,dx$ as approximately equal to $P(X = x)$ for some very small dx.



The following lemma demonstrates one way in which continuous random variables are very different from discrete random variables.

Lemma 9.1 Let X be a continuous random variable with PDF f. Then for any real number a:

- $P(X = a) = 0$
- $P(X < a) = P(X \leq a)$ → $P(X < a) = \int_{-\infty}^{a} f(x)\,dx$, $P(X \leq a) = \int_{-\infty}^{a} f(x)\,dx$
- $P(X > a) = P(X \geq a)$

I will define range of a continuous random variable as left closed but right open.

ex: $f(x) = 6x(1-x)$ $\quad 0 \leq x < 1$

Cumulative Distribution Function (CDF)

Definition 9.2 Let X be a continuous random variable with PDF f. Then the **Cumulative Distribution Function** CDF for X is

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

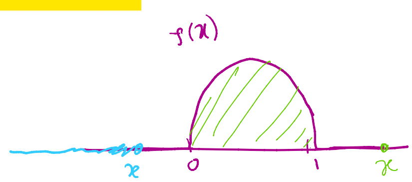
for any $x \in \mathbb{R}$.

Example 9.2

Determine the CDF for the random variable from example 9.1.

$$f(x) = 6x(1-x), \quad 0 \leq x < 1.$$

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

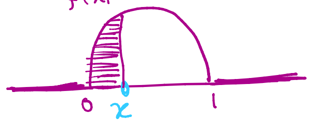


$$= \begin{cases} 0 & x < 0 \\ 3x^2 - 2x^3 & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

$\int_{-\infty}^{x} f(t)\,dt = 1$ for $x \geq 1$

$\boxed{0 \leqslant x < 1}$

$f(x) = 6x(1-x)$



$$F(x) = P(X \leqslant x) = \int_{-\infty}^{x} f(t)\,dt = 6\int_{0}^{x} t(1-t)\,dt = 6\left[\frac{t^2}{2} - \frac{t^3}{3}\Big|_0^x\right] = 6\left[\frac{x^2}{2} - \frac{x^3}{3}\right]$$

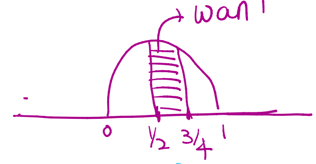
CDF $F(x)$ can also be used to find any probability.

Ex: $P(\frac{1}{2} \leqslant X \leqslant \frac{3}{4}) = P(X \leqslant 3/4) - P(X < \frac{1}{2})$

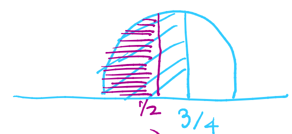
$= P(X \leqslant 3/4) - P(X \leqslant \frac{1}{2})$

$= F(3/4) - F(1/2)$

want



$P(X \leqslant 3/4)$



$P(X < 1/2)$

$F(3/4) = 3 \cdot \left(\frac{9}{16}\right) - 2\left(\frac{3}{4}\right)^3$

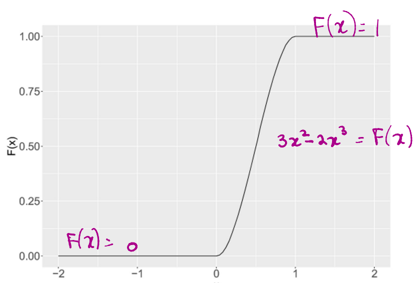
$F(\frac{1}{2}) = 3 \cdot \left(\frac{1}{2}\right)^2 - 2\left(\frac{1}{2}\right)^3$

Find $P(X \geqslant 7/8) = 1 - P(X < 7/8)$

$= 1 - P(X \leqslant 7/8) = 1 - F(7/8)$

$= 1 - \left(3 \cdot \left(\frac{7}{8}\right)^2 - 2\left(\frac{7}{8}\right)^3\right)$

A plot of the CDF we just derived is shown below. It is S-shaped which is actually a rather common shape for the CDF of a continuous random variable to have.



- $F(x)$ is a continuous function when X is a continuous random variable.
  - It is also differentiable almost everywhere. (there may be some values for which it is not differentiable though, but this doesn't create any problems as we will see later)
  - For both continuous and discrete random variables, we have the result

$$\lim_{x \to -\infty} F(x) = 0 \quad \text{and} \quad \lim_{x \to \infty} F(x) = 1$$

CDF to PDF

Definition 9.2

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

should remind us of the Fundamental Theorem of Calculus which states when F is defined as the integral of f, then f is the derivative of F.

Lemma 9.2 Suppose $F(x)$ is the CDF of a continuous random variable X with PDF f. Then

$$f(x) = \frac{d}{dx} F(x)$$

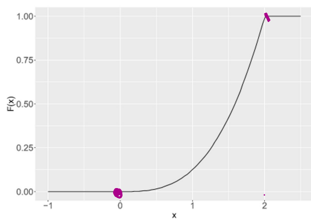
is a PDF for X.

PDF = derivative of CDF

Example 9.3

Let

$$F(x) = \begin{cases} 0 & x < 0 \\ x^3/8 & 0 \le x < 2 \\ 1 & x \ge 2 \end{cases}$$



Find a PDF $f(x)$ for X.

By the Fundamental theorem of calculus

$f(x) = \frac{d}{dx} F(x) = F'(x)$

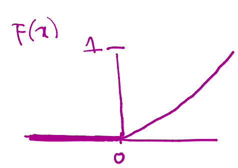
left hand derivative

$$= \begin{cases} 0 & x < 0 \\ 0 & x = 0 \longrightarrow F'(0-) = 0; \quad F'(0+) = 0 \\ 3x^2/8 & 0 < x < 2 \\ \text{undefined} & x = 2 \longrightarrow F'(2-) = 3 \times \frac{2^2}{8} = \frac{3}{2}; \quad F'(2+) = 0 \\ 0 & x > 2 \end{cases}$$

With a continuous random variable, the PDF at a point does not matter, that is, it does not impact any subsequent probability calculations. So we set the "undefined" value to any non-negative number we like.

$$\therefore \quad f(x) = \begin{cases} \frac{3x^2}{8} & 0 \le x < 2 \\ 0 & \text{otherwise} \end{cases}$$

Ex: $F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-x} & 0 \le x < \infty \end{cases}$



Find a PDF $f(x)$.

$$f(x) = F'(x)$$

$$= \begin{cases} e^{-x} & 0 \le x < \infty \\ 0 & \text{otherwise} \end{cases}$$

Note that the CDF F in Example 9.2 is not differentiable at $x = 2$.

However, since the value of $f(x)$ at a point does not affect subsequent probability calculations, we can set $f(2)$ to be any number we like.

For this reason, the CDF is the more important of the two functions for continuous random variables, since probabilities are completely determined by it.

Definition 9.3 Two random variables X and Y are said to be **identical in distribution**, denoted

$$X \stackrel{d}{=} Y$$

if their CDFs are the same. That is,

$$F_X(x) = F_Y(x) \quad \forall x.$$