Continuous Random Variables (Ch. 5.2)

Yifan Zhu

Iowa State University

Outline

Introduction to Continuous Random Variables

Probability Density Functions

Cumulative Distribution Functions

A special case: the exponential distribution

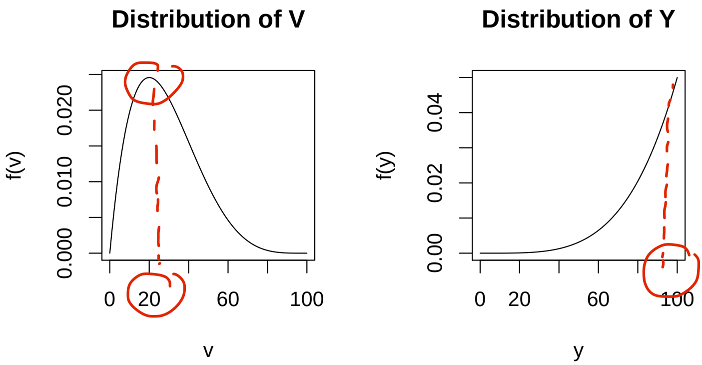
Continuous random variables

- Two types of random variables:
  - **Discrete random variable**: one that can only take on a set of isolated points (X, N, and S).
  - **Continuous random variable**: one that can fall in an interval of real numbers (T and Z).
- Examples of continuous random variables:
  - $Z =$ the amount of torque required to loosen the next bolt (*not* rounded).
  - $T =$ the time you'll have to wait for the next bus home.
  - $C =$ outdoor temperature at 3:17 PM tomorrow.
  - $L =$ length of the next manufactured part.

Continuous random variables

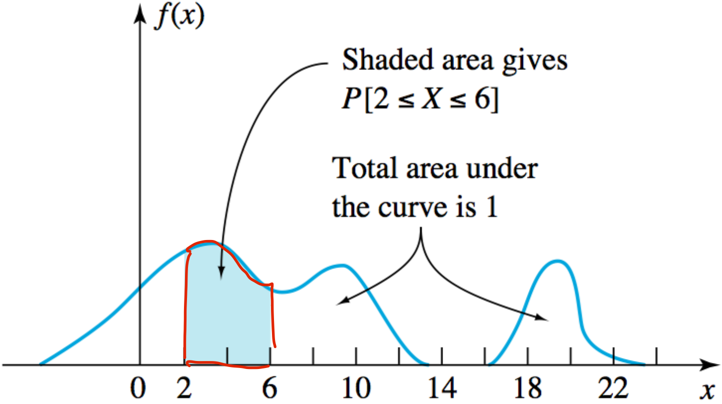
- V: % yield of the next run of a chemical process.
- Y: % yield of a *better* process.
- How do we mathematically distinguish between V and Y, given:
  - Each has the same range: $0\% \leq V, Y \leq 100\%$
  - There are uncountably many possible values in this range.
- We want to show that Y tends to take on higher % yield values than V.

V and Y have *continuous* probability distributions



- The heights of these curves are not themselves probabilities.
- However, the the curves tell us that process Y will yield more product per run on average than process V.

A generic probability density function (pdf)



Outline

Definition: probability density function (pdf)

- A **probability density function (pdf)** of a continuous random variable X is a function $f(x)$ with:

$$f(x) \geq 0 \text{ for all } x.$$

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

The probability of an interval is evaluated by integral:

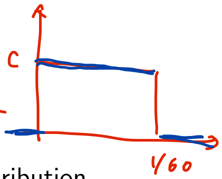
$$\begin{aligned} & P(a \leq X \leq b) \\ = & P(a < X < b) \\ = & P(a \leq X < b) \\ = & P(a < X \leq b) \\ = & \int_a^b f(x)dx, \ a \leq b \end{aligned}$$

- The pdf is the continuous analogue of a discrete random variable's probability mass function.

Example

- Let Y be the time delay (s) between a 60 Hz AC circuit and the movement of a motor on a different circuit.
- Say Y has a density of the form:

$$f(y) = \begin{cases} c & 0 \le y \le \frac{1}{60} \\ 0 & \text{otherwise} \end{cases}$$



we say that Y has a Uniform(0, 1/60) distribution.

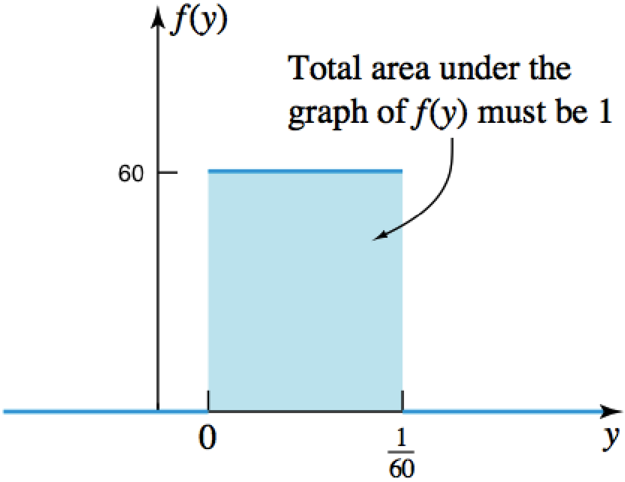
- $f(y)$ must integrate to 1:

$$1 = \int_{-\infty}^{\infty} f(y)dy = \overbrace{\int_{-\infty}^{0} 0dy}^{0} + \int_{0}^{1/60} cdy + \overbrace{\int_{1/60}^{\infty} 0dy}^{0} = \frac{c}{60} = 1$$

- hence, $c = 60$, and:

$$f(y) = \begin{cases} 60 & 0 \le y \le \frac{1}{60} \\ 0 & \text{otherwise} \end{cases}$$

A look at the density



Your turn: calculate the following probabilities.

textbook: $f(y) = \begin{cases} 60, & 0 < y < \frac{1}{60} \\ 0, & \text{otherwise} \end{cases}$

$$f(y) = \begin{cases} 60 & 0 \leq y \leq \frac{1}{60} \\ 0 & \text{otherwise} \end{cases}$$

1. $P(Y \leq \frac{1}{100})$
2. $P(Y > \frac{1}{70})$
3. $P(|Y| < \frac{1}{120})$
4. $P(\left|Y - \frac{1}{200}\right| > \frac{1}{110})$
5. $P(Y = \frac{1}{80})$

1. $P(Y \le \frac{1}{100})$ $\qquad f(y) = 60, \; \underline{0 \le y \le \frac{1}{60}}$

$$= \int_{-\infty}^{\frac{1}{100}} f(y)\, dy$$

$$= \int_{0}^{\frac{1}{100}} f(y)\, dy$$

$$= \int_{0}^{\frac{1}{100}} 60\, dy = \frac{60}{100} = \frac{3}{5}$$

2. $P\left(Y > \frac{1}{70}\right)$

$$= \int_{\frac{1}{70}}^{+\infty} f(y)\,dy$$

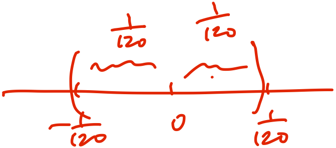
$$= \int_{\frac{1}{70}}^{\frac{1}{60}} 60\,dy = 60\left(\frac{1}{60} - \frac{1}{70}\right) = \frac{1}{7}$$

3. $P\left(|Y| < \frac{1}{120}\right)$

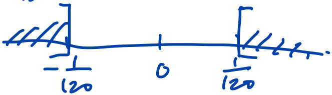
$$= P\left(-\frac{1}{120} < Y < \frac{1}{120}\right)$$

$$= \int_{-\frac{1}{120}}^{\frac{1}{120}} f(y)\,dy$$

$$= \int_{0}^{\frac{1}{120}} 60\,dy = 60 \times \left(\frac{1}{120} - 0\right) = \frac{1}{2}$$

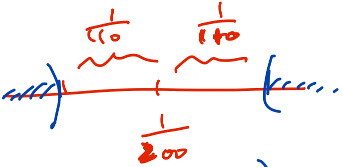


$|Y| \geq \frac{1}{120}$



$$\left(-\infty, -\frac{1}{120}\right] \cup \left[\frac{1}{120}, +\infty\right)$$

4. $P\left(\left|Y - \frac{1}{200}\right| \geq \frac{1}{110}\right)$.



$$= P\left(\left(-\infty, \frac{1}{200} - \frac{1}{110}\right) \cup \left(\frac{1}{200} + \frac{1}{110}, +\infty\right)\right)$$

$$= P\left(Y > \frac{1}{200} + \frac{1}{110}\right) + \underline{P\left(Y < \frac{1}{200} - \frac{1}{110}\right)}.$$

$$= \underline{P\left(Y > \frac{31}{2200}\right)} + \underline{P\left(Y < -\frac{9}{2200}\right)}. = 0$$

$$= \int_{\frac{31}{2200}}^{\frac{1}{60}} 60 \, dy = 60 \times \left(\frac{1}{60} - \frac{31}{2200}\right)$$

$$= \frac{17}{6600}$$

5. $P(Y = \frac{1}{80}) = 0$ (area under the curve).

$$= \int_{\frac{1}{80}}^{\frac{1}{80}} f(y)\, dy$$

$$= \int_{\frac{1}{80}}^{\frac{1}{80}} 60\, dy = 60 \times \left(\frac{1}{80} - \frac{1}{80}\right) = 0$$

$P(X = a) = 0$

true for any continuous random variables.

$P(X = a \text{ or } X = b \text{ or } X = c, \ldots, X = f)$

$= 0$

Let $P = \{p_1, p_2, \ldots, p_n\}$ a finite set.

X a continuous random variable.

$P(X \in P) = 0$

$\Rightarrow$ if two continuous random variable X, Y.

f_X, f_Y are different only on a finite set. these two random variable are essentially the same.

Answers: calculate the following probabilities

1.

$$
\begin{aligned}
P(Y \leq \frac{1}{100}) &= P(-\infty < Y \leq \frac{1}{100}) \\
&= \int_{-\infty}^{1/100} f(y)dy \\
&= \int_{-\infty}^{0} 0dy = \int_{0}^{1/100} 60dy \\
&= \frac{60}{100} = \frac{3}{5}
\end{aligned}
$$

2.

$$
\begin{aligned}
P(Y > \frac{1}{70}) &= P(\frac{1}{70} < Y \leq \infty) \\
&= \int_{1/70}^{\infty} f(y)dy \\
&= \int_{1/70}^{1/60} 60dy + \int_{1/60}^{\infty} 0dy \\
&= 60y \,|_{1/70}^{1/60} + 0 \\
&= 60\left(\frac{1}{60} - \frac{1}{70}\right) \\
&= \frac{1}{7} \approx 0.143
\end{aligned}
$$

3.

$$
\begin{aligned}
P(|Y| < \frac{1}{120}) &= P(-\frac{1}{120} < Y < \frac{1}{120}) \\
&= \int_{-1/120}^{1/120} f(y)dy \\
&= \int_{-1/120}^{0} 0dy + \int_{0}^{1/120} 60dy \\
&= 0 + 60y \mid_0^{1/120} \\
&= 60\left(\frac{1}{120} - 0\right) = \frac{1}{2}
\end{aligned}
$$

4.

$$
\begin{aligned}
&P(\left|Y - \frac{1}{200}\right| > \frac{1}{110}) \\
&= P(Y - \frac{1}{200} > \frac{1}{110} \text{ or } Y - \frac{1}{200} < -\frac{1}{110}) \\
&= P(Y > \frac{31}{2200} \text{ or } Y < -\frac{9}{2200}) \\
&= P(Y > \frac{31}{2200}) + P(Y < -\frac{9}{2200}) \\
&= \int_{31/2200}^{\infty} f(y)dy + \int_{-\infty}^{-9/2200} f(y)dy \\
&= \int_{31/2200}^{1/60} 60dy + \int_{1/60}^{\infty} 0dy + \int_{-\infty}^{-9/2200} 0dy \\
&= 60y\Big|_{31/2200}^{1/60} + 0 + 0 \\
&= 60\left(\frac{1}{60} - \frac{31}{2200}\right) = \frac{17}{6600} \approx 0.00258
\end{aligned}
$$

5.

$$
\begin{aligned}
P(Y = \frac{1}{80}) &= P(\frac{1}{80} \leq Y \leq \frac{1}{80}) \\
&= \int_{1/80}^{1/80} f(y)dy = \int_{1/80}^{1/80} 60dy \\
&= 60y\Big|_{1/80}^{1/80} = 60\left(\frac{1}{80} - \frac{1}{80}\right) \\
&= 0
\end{aligned}
$$

In fact, for any random variable X and any real number a:

$$
\begin{aligned}
P(X = a) &= P(a \leq X \leq a) \\
&= \int_a^a f(x)dx = 0
\end{aligned}
$$

Outline

Cumulative distribution functions (cdf)

- The **cumulative distribution function** of a random variable X is a function F such that:

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

In other words: Newton-Lebnitz formula

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

- As with discrete random variables, F has the following properties:
  - $F(x) \geq 0$ for all x.
  - F is monotonically increasing.
  - $\lim_{x \to -\infty} F(x) = 0$ $\leftarrow \int_{-\infty}^{-\infty} = 0$
  - $\lim_{x \to \infty} F(x) = 1$ $\leftarrow \int_{-\infty}^{\infty} = 1$

Example: calculating the cdf of Y

- Remember:

$$f_Y(y) = \begin{cases} 60 & 0 \leq y \leq 1/60 \\ 0 & \text{otherwise} \end{cases}$$

- For $y < 0$:

$$F(y) = P(Y \leq y) = \int_{-\infty}^{y} f(t)dt = \int_{-\infty}^{y} 0dt = 0$$

- For $0 \leq y \leq 1/60$:

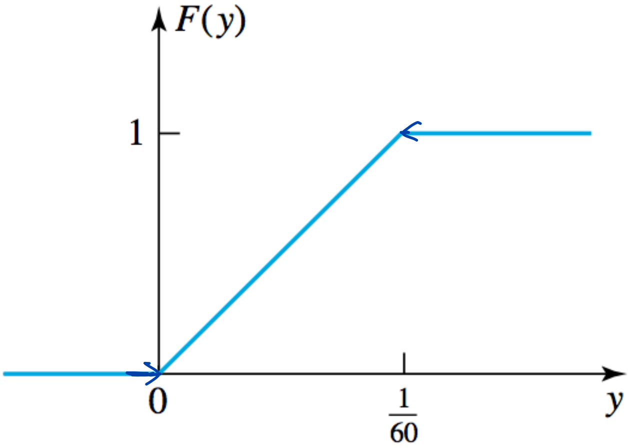
$$F(y) = P(Y \leq y) = \int_{-\infty}^{y} f(t)dt = \int_{-\infty}^{0} 0dt + \int_{0}^{y} 60dt = 60y$$

- For $y > 1/60$:

$$F(y) = P(Y \leq y) = \int_{-\infty}^{y} f(t)dt$$
$$= \int_{-\infty}^{0} 0dt + \int_{0}^{1/60} 60dt + \int_{1/60}^{y} 0dt = 1$$

A look at the cdf

continuous function.



Your turn: calculate the following using the cdf

$$F(y) = \begin{cases} 0 & y < 0 \\ 60y & 0 \leq y \leq \frac{1}{60} \\ 1 & y > \frac{1}{60} \end{cases}$$

1. $F(1/70)$
2. $P(Y \leq \frac{1}{80})$
3. $P(Y > \frac{1}{150})$
4. $P(\frac{1}{130} \leq Y \leq \frac{1}{120})$

1. $F(1/70) = P(Y \le 1/70)$

$= 60 \times \frac{1}{70} = \frac{6}{7}.$

2. $P(Y \le \frac{1}{80}) = F(\frac{1}{80}) = 60 \times \frac{1}{80} = \frac{3}{4}$

3. $P(Y > \frac{1}{150}) = 1 - P(Y \le \frac{1}{150}).$

$= 1 - F(\frac{1}{150}) = 1 - 60 \times \frac{1}{150} = 1 - \frac{2}{5} = \frac{3}{5}$

$P(X > x) = 1 - P(X \le x)$

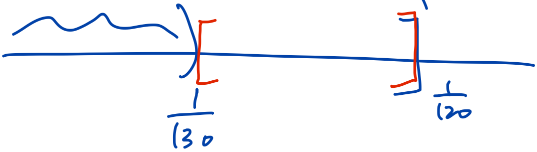
for all kinds random variables

$P(X \ge x) = 1 - P(X < x)$

4. $P\left(\frac{1}{130} \le Y \le \frac{1}{120}\right)$

union

$$= P\left(Y \le \frac{1}{120}\right) - \underbrace{P\left(Y < \frac{1}{130}\right)}_{= P\left(Y \le \frac{1}{130}\right)}.$$



$$= F\left(\frac{1}{120}\right) - F\left(\frac{1}{130}\right)$$

$$= 60 \times \frac{1}{120} - 60 \times \frac{1}{130}$$

$$= \frac{1}{26}$$

Answers: calculate the following using the cdf

1. $F(\frac{1}{70}) = 60\frac{1}{70} = \frac{6}{7}$
2. $P(Y \leq \frac{1}{80}) = F(\frac{1}{80}) = 60\frac{1}{80} = \frac{3}{4}$
3. $$\begin{aligned} P(Y > \frac{1}{150}) &= \int_{1/150}^{\infty} f(y)dy \\ &= \int_{-\infty}^{\infty} f(y)dy - \int_{-\infty}^{1/150} f(y)dy \\ &= 1 - F(1/150) = 1 - \frac{60}{150} \\ &= \frac{3}{5} \end{aligned}$$

   In fact, for any random variable X, discrete or continuous:

   $$P(X \geq x) = 1 - P(X < x)$$

4.

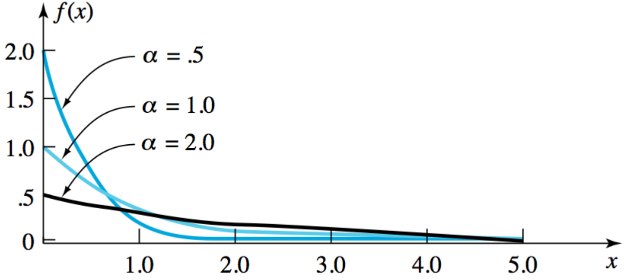
$$
\begin{aligned}
P(\frac{1}{130} \leq Y \leq \frac{1}{120}) &= \int_{1/130}^{1/120} f(y)dy \\
&= \int_{-\infty}^{1/120} f(y)dy - \int_{-\infty}^{1/130} f(y)dy \\
&= F(1/120) - F(1/130) \\
&= 60(1/120) - 60(1/130) \\
&= 1/26 \approx 0.0384
\end{aligned}
$$

Outline

The exponential distribution

- A random variable X has an Exponential(α) distribution if:

$$f(x) = \begin{cases} \frac{1}{\alpha} e^{-x/\alpha} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$



Your turn: for $X \sim \text{Exp}(2)$, calculate the following

$$f(x) = \begin{cases} \frac{1}{2}e^{-x/2} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

1. $P(X \leq 1)$
2. $P(X > 5)$
3. The cdf F of X

1. $P(X \le 1)$

$$\downarrow \mathrm{Exp}(2)$$

$$= \int_0^1 f(x)\,dx = \int_0^1 \underline{\tfrac{1}{2} e^{-x/2}}\,dx$$

$$= \int_0^1 \left(-e^{-x/2}\right)'\,dx$$

$$= -e^{-x/2}\Big|_0^1 = -e^{-1/2} - (-1)$$

$$= 1 - e^{-1/2}$$

2: $P(X > 5)$

$$= \int_{5}^{+\infty} \frac{1}{2} e^{-x/2} \, dx$$

$$= \left(-e^{-x/2}\right)\Big|_{5}^{+\infty}$$

$$= 0 - \left(-e^{-5/2}\right)$$

$$= e^{-5/2}$$

3. $F(x) = \int_{-\infty}^{x} f(t)\,dt$

$x \leq 0$, $F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{x} 0\,dt = 0$

$x > 0$, $F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{0}^{x} f(t)\,dt$

$= \int_{0}^{x} \frac{1}{2} e^{-\frac{1}{2}t}\,dt$

$= \left(-e^{-\frac{1}{2}t}\right)\Big|_{0}^{x} = -e^{-x/2} - (-1)$

$= 1 - e^{-x/2}$

$$F(x) = \begin{cases} 0, & x \leq 0 \\ 1 - e^{-x/2}, & x > 0. \end{cases}$$

Answers: for $X \sim \text{Exp}(2)$, calculate the following

1.

$$
\begin{aligned}
P(X \le 1) &= \int_{-\infty}^{1} f(x)dx \\
&= \int_{-\infty}^{0} 0dx + \int_{0}^{1} \frac{1}{2} e^{-x/2} dx \\
&= 0 + (-e^{-x/2} |_0^1 \\
&= -e^{-1/2} - (-e^{-0/2}) \\
&= 1 - e^{-1/2} \approx 0.393
\end{aligned}
$$

2.

$$
\begin{aligned}
P(X > 5) &= \int_5^\infty f(x)dx \\
&= \int_5^\infty \frac{1}{2} e^{-x/2} dx \\
&= -e^{-x/2}|_5^\infty \\
&= -e^{-\infty/2} + e^{-5/2} \\
&= e^{-5/2} \approx 0.082
\end{aligned}
$$

3. For $x < 0$:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx)$$
$$= \int_{-\infty}^{x} 0dx = 0$$

For $x \geq 0$:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx$$
$$= \int_{-\infty}^{0} 0dx + \int_{0}^{x} \frac{1}{2}e^{-t/2}dt$$
$$= -e^{-t/2} \mid_{0}^{x} = -e^{-x/2} - (-e^{-0/2})$$
$$= 1 - e^{-x/2}$$

Hence:

$$F(x) = \begin{cases} 1 - e^{-x/2} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

In general, an Exp(α) random variable has cdf:

$$\boxed{F(x) = \begin{cases} 1 - e^{-x/\alpha} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}}$$

Uses of the Exp(α) random variable

- The exponential distribution is the continuous analog of the geometric distribution:
  - A Geometric(p) randomvariable counts the number of trials until a success happens, and the success probability for each trail is the same; An Exp(α) random variable measures the waiting time until a specific event happens, and at any point in time, that event has an equal chance of happening.
  - Memoryless: in Geometric(p), if we know the success has not occured in the first t_0 trails, the additional number of trails (beyond t_0) needed to get a success is still a Geometric(p) random variable; in Exp(α), if we know the event has not happended by time t_0, the additional waiting time for that event to happen is still Exp(α)

Uses of the Exp(α) random variable

Examples:

- Time between your arrival at a bus stop and the moment the bus comes.
- Time until the next person walks inside the library.
- Time until the next car accident on a stretch of highway.