STAT 234 Lecture 7
Continuous Random Variables
Section 4.1-4.2

Yibi Huang
Department of Statistics
University of Chicago

Coverage:

Continuous Random Variables (Section 4.1-4.2 in MMSA)

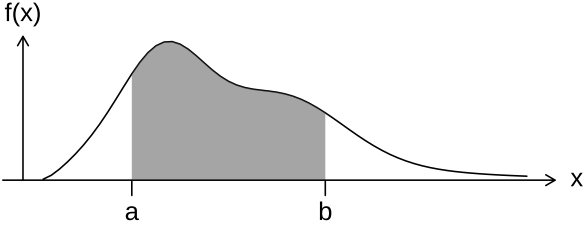
- probability density function (pdf)
- cumulative distribution function (cdf)
- expected values, variance (Section 4.2 in MMSA)
    - Skip [Approximating the Mean Value and Standard Deviation] and [Moment Generating function] on p.174-177

Continuous Random Variables

Continuous Random Variables (Review)

A random variable X is said to have a *continuous distribution* if there exists a non-negative function f such that

$$P(a < X \leq b) = \int_a^b f(x)\,dx, \quad \text{for all } -\infty \leq a < b \leq \infty.$$



Here f is called the *probability density function* (*pdf*), the *density curve*, or the *density* of X.

Conditions of pdf (Review)

A (pdf) $f(x)$ can be of any imaginable shape but must satisfy the following:

- It must be *nonnegative*

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

- The *total area under the pdf must be 1*

$$\int_{-\infty}^{\infty} f(x)\,dx = P(-\infty < X \leq \infty) = 1$$

Interpretation of pdf (Review)

Suppose f is the pdf of X. If f is continuous at a point x, then for small δ

$$P\left(x - \frac{\delta}{2} < X \leq x + \frac{\delta}{2}\right) = \int_{x-\delta/2}^{x+\delta/2} f(u)\,du = \delta f(x).$$

- Is the pdf f of a random variable always ≤ 1?

Interpretation of pdf (Review)

Suppose f is the pdf of X. If f is continuous at a point x, then for small δ

$$P\left(x - \frac{\delta}{2} < X \leq x + \frac{\delta}{2}\right) = \int_{x-\delta/2}^{x+\delta/2} f(u)\, du = \delta f(x).$$

- Is the pdf f of a random variable always ≤ 1?

  No, the pdf $f(x)$ itself is not a probability.

  It's the area underneath $f(x)$ that represents the probability.

$P(X = x) = 0$ If X Is Continuous (Review)

For any continuous random variable X

$$P(X = x) = \int_x^x f(u)du = 0$$

$P(X = x) = 0$ If X Is Continuous (Review)

For any continuous random variable X

$$P(X = x) = \int_x^x f(u)du = 0$$

- What percentage of men are 6-feet tall exactly?
  Those that are 6.00001 or 5.99999 feet tall don't count.

$P(X = x) = 0$ If X Is Continuous (Review)

For any continuous random variable X

$$P(X = x) = \int_x^x f(u)du = 0$$

- What percentage of men are 6-feet tall exactly?
  Those that are 6.00001 or 5.99999 feet tall don't count.
- It doesn't matter whether the end point(s) of an interval is included when calculating the probability of X falling the interval if X is continuous

$$P(a < X < b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b)$$

A pdf $f(x)$ May Not be Continuous

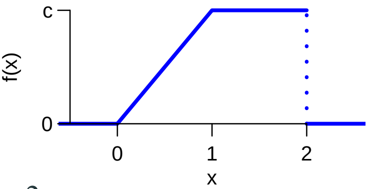
The pdf $f(x)$ of a continuous random variable might not be continuous.

See the example on the next page.

Example 1 (Review)

Consider a continuous random variable X with the pdf

$$f(x) = \begin{cases} cx & \text{if } 0 \leq x \leq 1 \\ c & \text{if } 1 \leq x \leq 2 \\ 0 & \text{elsewhere} \end{cases}$$

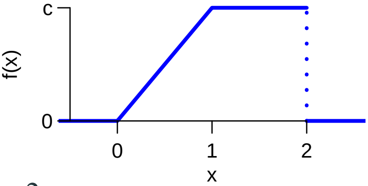


- Note $f(x)$ is not continuous at $x = 2$

Example 1 (Review)

Consider a continuous random variable X with the pdf

$$f(x) = \begin{cases} cx & \text{if } 0 \le x \le 1 \\ c & \text{if } 1 \le x \le 2 \\ 0 & \text{elsewhere} \end{cases}$$

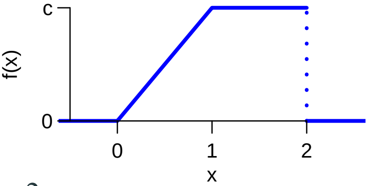


- Note $f(x)$ is not continuous at $x = 2$
- What is the value of c?

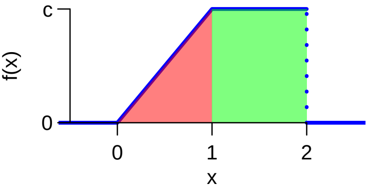
Example 1 (Review)

Consider a continuous random variable X with the pdf

$$f(x) = \begin{cases} cx & \text{if } 0 \le x \le 1 \\ c & \text{if } 1 \le x \le 2 \\ 0 & \text{elsewhere} \end{cases}$$



- Note $f(x)$ is not continuous at $x = 2$
- What is the value of c?



Total Area = Red + Green

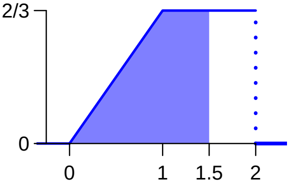
$$= \frac{1 \cdot c}{2} + 1 \cdot c = \frac{3}{2}c = 1$$

$$\Rightarrow \quad c = \frac{2}{3}$$

Example 1 (Cont'd)

What is $P(X \leq 1.5)$?

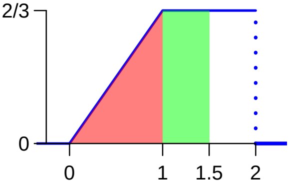
$P(X \leq 1.5) =$



Example 1 (Cont'd)

What is $P(X \leq 1.5)$?

$P(X \leq 1.5) =$

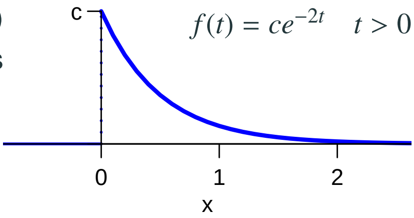


$= \text{Red} + \text{Green}$

$$= \frac{1 \cdot (2/3)}{2} + (0.5)\frac{2}{3} = \frac{2}{3}$$

Example 2

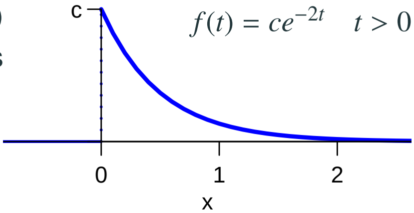
Suppose the lifetime T (in days) of a certain type of batteries has the pdf shown on the right.



- Find the value of c so that $f(t)$ is a legitimate pdf.

Example 2

Suppose the lifetime T (in days) of a certain type of batteries has the pdf shown on the right.



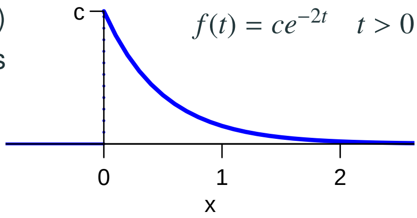
- Find the value of c so that $f(t)$ is a legitimate pdf.

$$\int_{-\infty}^{\infty} f(t)dt = \int_{0}^{\infty} ce^{-2t}dt = -\frac{c}{2}e^{-2t}\Big|_{t=0}^{t=\infty} = \frac{c}{2} - 0 = 1$$

So $c = 2!$

Example 2

Suppose the lifetime T (in days) of a certain type of batteries has the pdf shown on the right.



- Find the value of c so that $f(t)$ is a legitimate pdf.

$$\int_{-\infty}^{\infty} f(t)dt = \int_0^{\infty} ce^{-2t}dt = -\frac{c}{2}e^{-2t}\Big|_{t=0}^{t=\infty} = \frac{c}{2} - 0 = 1$$

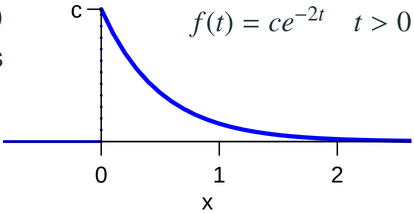
  So $c = 2!$

- Observe that $f(0) = 2e^0 = 2 > 1$!?!

  Can a pdf $f(x)$ exceed 1?

Example 2

Suppose the lifetime T (in days) of a certain type of batteries has the pdf shown on the right.



- Find the value of c so that $f(t)$ is a legitimate pdf.

$$\int_{-\infty}^{\infty} f(t)dt = \int_{0}^{\infty} ce^{-2t}dt = -\frac{c}{2}e^{-2t}\Big|_{t=0}^{t=\infty} = \frac{c}{2} - 0 = 1$$

  So $c = 2$!

- Observe that $f(0) = 2e^0 = 2 > 1$!?!

  Can a pdf $f(x)$ exceed 1?

  Yes, the pdf $f(x)$ itself is not a probability.

  It's the area underneath $f(x)$ that represents the probability.

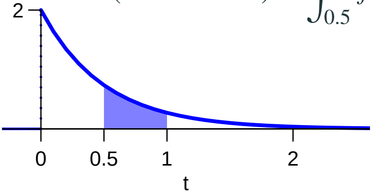
Example 2 (Cont'd)

What is the chance that the battery lasts 0.5 to 1 day?

Example 2 (Cont'd)

What is the chance that the battery lasts 0.5 to 1 day?

$$P(0.5 < T < 1) = \int_{0.5}^{1} f(t)dt = \int_{0.5}^{1} 2e^{-2t}dt$$



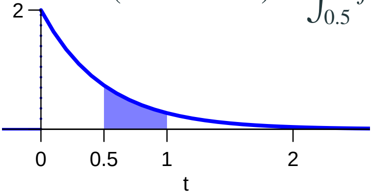
$$= -e^{-2t}\Big|_{t=0.5}^{t=1} = e^{-1} - e^{-2}.$$

Example 2 (Cont'd)

What is the chance that the battery lasts 0.5 to 1 day?

$$P(0.5 < T < 1) = \int_{0.5}^{1} f(t)dt = \int_{0.5}^{1} 2e^{-2t}dt$$

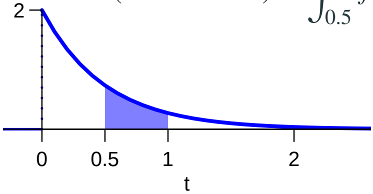
$$= -e^{-2t}\Big|_{t=0.5}^{t=1} = e^{-1} - e^{-2}.$$



What is the chance that the battery last over one day, $P(T > 1)$?

Example 2 (Cont'd)

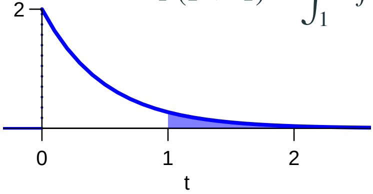
What is the chance that the battery lasts 0.5 to 1 day?



$$P(0.5 < T < 1) = \int_{0.5}^{1} f(t)dt = \int_{0.5}^{1} 2e^{-2t}dt$$

$$= -e^{-2t}\Big|_{t=0.5}^{t=1} = e^{-1} - e^{-2}.$$

What is the chance that the battery last over one day, $P(T > 1)$?



$$P(T > 1) = \int_{1}^{\infty} f(x)dx = \int_{1}^{\infty} 2e^{-2x}dx$$

$$= -e^{-2x}\Big|_{t=1}^{t=\infty} = e^{-2}.$$

Cumulative Distribution Function (cdf)

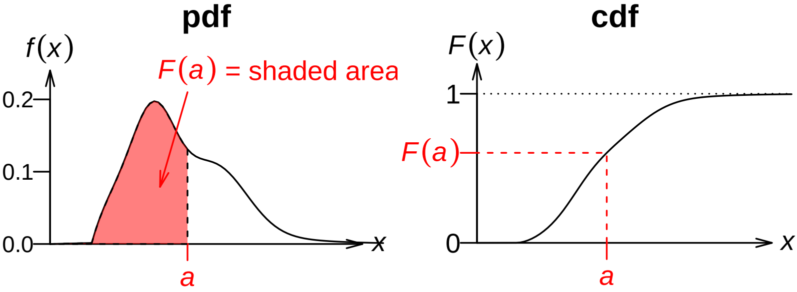
Cumulative Distribution Function (cdf)

For any random variable X, its *cumulative distribution function* (*cdf*) is the function defined by

$$F(x) = F_X(x) = P(X \leq x).$$

One get the cdf of a random variable from its pdf by **integration**:

$$F(x) = \int_{-\infty}^{x} f(u)\,du$$



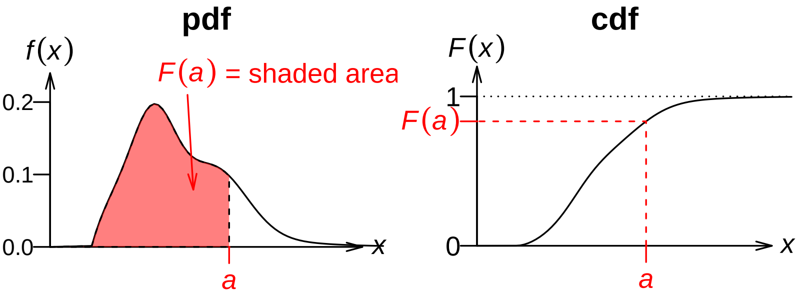
Cumulative Distribution Function (cdf)

For any random variable X, its *cumulative distribution function* (*cdf*) is the function defined by

$$F(x) = F_X(x) = P(X \leq x).$$

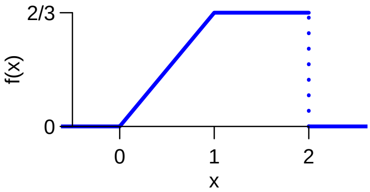
One get the cdf of a random variable from its pdf by **integration**:

$$F(x) = \int_{-\infty}^{x} f(u)\, du$$



Example 1 (cdf)

$$f(x) = \begin{cases} 2x/3 & \text{if } 0 \le x \le 1 \\ 2/3 & \text{if } 1 \le x \le 2 \\ 0 & \text{elsewhere} \end{cases}$$

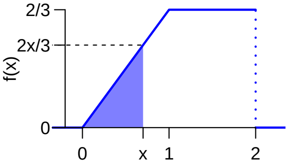


Let's find the cdf $F(x)$ for the density in Example 1 piece by piece.

- For $x < 0$, $F(x) = \int_{-\infty}^{x} f(u)du = 0$ since $f(u) = 0$ for $u < 0$.
- For $0 \le x < 1$,

$$F(x) = P(X \le x) = \int_{\infty}^{x} f(u)du$$

$$= \text{shaded area of}$$

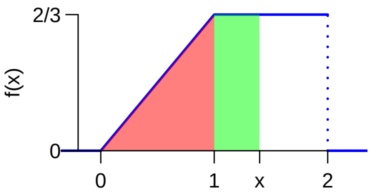


$$= \frac{x \cdot (2x/3)}{2} = \frac{x^2}{3}$$

For $1 \le x \le 2$,

$$F(x) = P(X \le x) = \int_{\infty}^{x} f(u)du$$

$= $ shaded area of



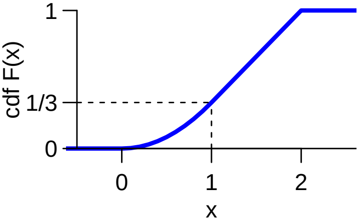
$= $ Red + Green

$$= \frac{1 \cdot (2/3)}{2} + \frac{2}{3} \cdot (x-1) = \frac{1}{3} + \frac{2}{3}(x-1)$$

For $x > 2$, $F(x) = \int_{\infty}^{x} f(u)du = 1$ since the entire area is included.

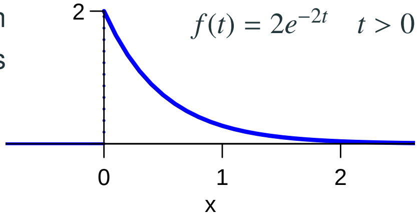
To sum up, the cdf is

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{3}x^2 & \text{if } 0 \le x \le 1 \\ \frac{1}{3} + \frac{2}{3}(x-1) & \text{if } 1 \le x \le 2 \\ 1 & \text{if } x > 2 \end{cases}$$



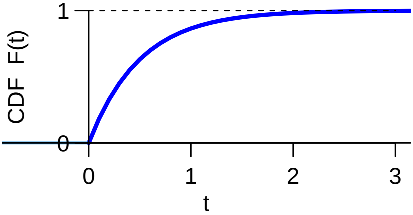
Example 2 (cdf)

Recall the pdf for the lifetime T (in days) of a certain type of batteries is $f(t) = 2e^{-2t}$, $t > 0$



The cdf $F(t)$ is

$$F(t) = \begin{cases} 0 & \text{if } t < 0 \\ \int_{-\infty}^{t} f(x)dx = \int_0^t 2e^{-2u}du = -e^{-2u}\Big|_0^t = 1 - e^{-2t} & \text{for } t \geq 0 \end{cases}$$



Obtaining the PDF from the CDF

The PDF can be obtained from the cdf by differentiation.

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

Example 1

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{3}x^2 & \text{if } 0 \le x \le 1 \\ \frac{1}{3} + \frac{2}{3}(x-1) & \text{if } 1 \le x \le 2 \\ 1 & \text{if } x > 2 \end{cases} \Rightarrow \frac{d}{dx}F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{2}{3}x & \text{if } 0 \le x \le 1 \\ \frac{2}{3} & \text{if } 1 \le x \le 2 \\ 0 & \text{if } x > 2 \end{cases}$$

Observe $\frac{d}{dx}F(x)$ is exactly the pdf $f(x)$.

Example 2. For the cdf of the battery life distribution

$$F(t) = \begin{cases} 0 & \text{if } t < 0 \\ 1 - e^{-2t} & \text{for } t \ge 0 \end{cases} \Rightarrow \frac{d}{dx}F(x) = \begin{cases} 0 & \text{if } t < 0 \\ 2e^{-2t} & \text{for } t \ge 0 \end{cases}$$

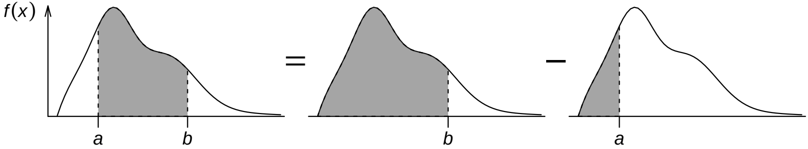
Using the cdf to Compute Probabilities

Let X be a continuous rv with pdf $f(x)$ and cdf $F(x)$. Then for any number a,

$$P(X > a) = 1 - F(a)$$

and for any two numbers a and b with $a < b$,

$$P(a \le X \le b) = F(b) - F(a)$$



Recall in **Example 2**, we computed $P(0.5 < T < 1)$ by integrating the pdf. We can also compute using the cdf, $F(t) = 1 - e^{-2t}$, $t > 0$.

$$P(0.5 < T < 1) = F(1) - F(0.5) = (1 - e^{-2}) - (1 - e^{-1}) = e^{-1} - e^{-2}$$

which agrees with our prior calculation.

Properties of cdfs

- The cdf $F(x) = P(X \leq x)$ is a probability, and hence it must be *between 0 and 1*.

$$0 \leq F(x) \leq 1$$

- cdfs are always *non-decreasing*. For $a < b$

$$F(b) - F(a) = P(X \leq b) - P(X \leq a) = P(a < X \leq b) \geq 0$$

- The cdf of a continuous r.v. must be *continuous*. As $\delta \to 0$

$$F(x + \delta) - F(x) = \int_x^{x+\delta} f(u)du \to 0$$

Expected Values

Expected Values

Let X be a continuous random variable with density f_X, then the **expectation** of X is

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

Suppose $Y = g(X)$ is a function of X. The **expectation** of Y is

$$\mathrm{E}(Y) = \mathrm{E}(g(X)) = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx.$$

Variance and Standard Deviation

The *variance* of a continuous r.v. X, with density $f(x)$, and mean μ, is denoted as Var(X), σ_X^2, or simply σ^2, is defined as

$$\text{Var}(X) = \text{E}(X-\mu)^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx., \quad \text{where } \mu = \text{E}(X).$$

The *standard deviation* (*SD*) is the square root of the variance,

$$\text{SD}(X) = \sigma = \sqrt{\text{Var}(X)}.$$

Properties of the Expected Value and Variance

Property 1. The shortcut formula to find the variance remains valid for continuous random variables.

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mu^2.$$

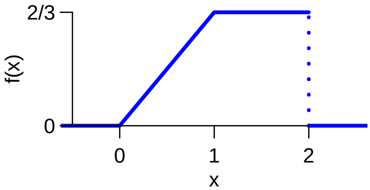
Property 2. For any constants a and b, the following identities are also valid for continuous r.v. X.

- $\mathrm{E}(aX + b) = a\,\mathrm{E}(X) + b$
- $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$
- $\mathrm{SD}(aX + b) = |a|\mathrm{SD}(X)$

The proofs are similar to the ones for the discrete case, just replacing the summation $\sum$ with the integral $\int$, and hence are omitted.

Example 1 (Mean, Variance, SD)

$$f(x) = \begin{cases} 2x/3 & \text{if } 0 \le x \le 1 \\ 2/3 & \text{if } 1 \le x \le 2 \\ 0 & \text{elsewhere} \end{cases}$$



$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^1 x \frac{2x}{3} dx + \int_1^2 x \frac{2}{3} dx$$

$$= \frac{2x^3}{9}\Big|_0^1 + \frac{x^2}{3}\Big|_1^2 = \frac{2}{9} + \frac{4}{3} - \frac{1}{3} = \frac{11}{9}$$

$$\mathrm{E}(X^2) == \int_{-\infty}^{\infty} x^2 f(x) dx = \int_0^1 x^2 \frac{2x}{3} dx + \int_1^2 x^2 \frac{2}{3} dx$$

$$= \frac{x^4}{6}\Big|_0^1 + \frac{2x^3}{9}\Big|_1^2 = \frac{1}{6} + \frac{16}{9} - \frac{2}{9} = \frac{31}{18}$$

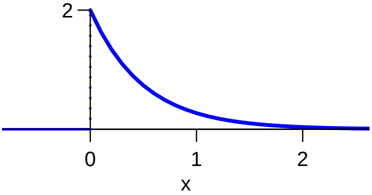
$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = \frac{31}{18} - (\frac{11}{9})^2 = \frac{37}{162}$ by the shortcut formula.

$\mathrm{SD}(X) = \sqrt{37/162} \approx 0.478.$

Example 2 (Expected Value)

For the pdf $f(t) = 2e^{-2t}$, $t > 0$, the expected value is

$$E(T) = \int_{-\infty}^{\infty} t f(t) dt = \int_0^{\infty} 2te^{-2t} dt$$



To find $E(T)$, we need to use *integration by part*.

$$\int_a^b g(t)h'(t)dt = \int_a^b g(t)dh(t) = g(t)h(t)\Big|_a^b - \int_a^b h(t)g'(t)dt.$$

With $h(t) = e^{-2t}$ and $g(t) = -t$, we get

$$E(T) = \int_0^{\infty} 2te^{-2t} dt = \int_0^{\infty} -tde^{-2t} = -te^{-2t}\Big|_0^{\infty} - \int_0^{\infty} e^{-2t} d(-t)$$

$$= 0 + \int_0^{\infty} e^{-2t} dt = -\frac{1}{2}e^{-2t}\Big|_0^{\infty} = \frac{1}{2}.$$

Example 2 (Variance by the Shortcut Formula)

$$\mathrm{E}(T^2) = \int_{-\infty}^{\infty} t^2 f(t)dt = \int_0^{\infty} 2t^2 e^{-2t} dt$$

To find $\mathrm{E}(T^2)$, we need to do *integration by part* again. With $h(t) = e^{-2t}$ and $g(t) = -t^2$, we get

$$\begin{aligned} \mathrm{E}(T^2) = \int_0^{\infty} 2t^2 e^{-2t} dt &= \int_0^{\infty} -t^2 d(e^{-2t}) \\ &= -t^2 e^{-2t}\Big|_0^{\infty} - \int_0^{\infty} e^{-2t} \underbrace{d(-t^2)}_{=-2t} \\ &= 0 + \int_0^{\infty} 2te^{-2t} dt \end{aligned}$$

Observe $\int_0^{\infty} 2te^{-2t} dt$ is exactly $\mathrm{E}(T) = 1/2$ we just calculated, and hence $\mathrm{E}(T^2) = 1/2$.

We can then use $\mathrm{E}(T^2)$ to find the variance.

$$\mathrm{Var}(T) = \mathrm{E}(T^2) - (\mathrm{E}(T))^2 = \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

and $\mathrm{SD}(T) = \sqrt{\mathrm{Var}(T)} = \sqrt{1/4} = 1/2$.