MH1820 Week 4

1. Discrete distribution: Poisson discrete random variable
2. Continuous random variables, PDF and CDF
3. The Exponential and Gamma distribution

continuous.

The Poisson distribution

The Poisson distribution

Some experiments result in counting the **number of times** particular events occur/arrive during a given **time interval**.

Examples:

- Number of phone calls between 9AM and 10AM; the number of customers that arrive at a ticket window between 12noon and 2pm.
- Number of typos on a 10-page report. (here: 10-page is like the "time interval")

Usually, need a parameter λ that measures the average particular events occur per unit time.

(generic) (generic)

X = # arrivals per unit interval.

λ = average # arrivals per unit interval

$= \mathbb{E}[X]$.

A discrete random variable X has a **Poisson distribution**, denoted by $X \sim Poisson(\lambda)$, if its PMF is of the form

$$p(x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad \underline{x = 0, 1, 2, \ldots,}$$

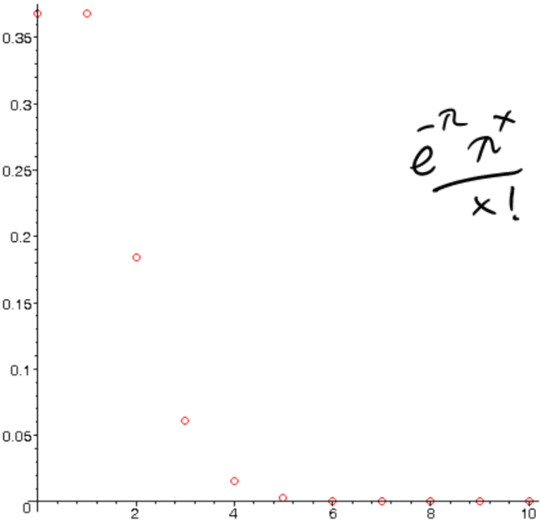
where $\lambda > 0$.

Theorem 1 (Poisson)

If $X \sim Poisson(\lambda)$, then

$$\mathbb{E}[X] = \lambda, \quad \mathrm{Var}[X] = \lambda. \quad \checkmark$$

verify in Tut 6 later.



$$\frac{e^{-\lambda}\lambda^{x}}{x!}$$

PMF of Poisson(1) $\lambda = 1$

Example 2

On average, there are 2 supernovae in the milky way per century. Assuming Possion distribution, what is the probability that there are 2 supernovae in the milky way within one decade?

X = # supernovae per **decade**. ✓

Want calculate $\mathbb{P}(X=2)$

$X \sim \text{Poisson}(\lambda)$

On average, 2 supernovae per century. (100 years)

$\underline{\underline{\quad "\quad}}$ λ $\underline{\underline{\quad "\quad}}$ " decade (10 years)

$\frac{2}{10}$

$X \sim \text{Poisson}(\lambda = 0.2)$

$$\mathbb{P}(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \quad P(X=2) = \frac{e^{-0.2}\,0.2^{2}}{2!}$$

$$\approx 0.016 \ \#.$$

Example 3

In a city, telephone calls to 911 come on the average of two every 3 minutes. If one assumes a Poisson distribution, what is the probability of five or more calls arriving in a 9-minute period?

X = # calls per 9-minute period.

Want $\mathbb{P}(X \geq 5)$.

$X \sim \text{Poisson}(\lambda)$

On average, 2 calls per 3-minute.

$\times 3$ $\times 3$

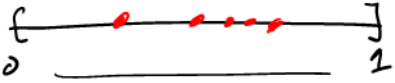
On average, 6 calls per 9-minute

So $\lambda = 6$.

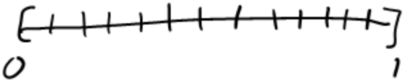
$$\mathbb{P}(X \geq 5) = 1 - \mathbb{P}(X \leq 4)$$
$$= 1 - \mathbb{P}(0 \leq X \leq 4)$$
$$= 1 - \left(\sum_{x=0}^{4} \frac{e^{-6} 6^x}{x!} \right) \approx 1 - 0.285$$
$$\approx 0.715 \quad \#.$$

Not tested:

λ = average # arrivals.



unit interval.



n subintervals.

$\frac{\lambda}{n}$ = average per subinterval.

$p = \frac{\lambda}{n}$ = "prob" of having an arrival per subinterval.

$\lambda = np$. (success probability p).

$$\mathbb{P}(X = x) = \binom{n}{x}\left(\frac{\lambda}{n}\right)^{x}\left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$= \underbrace{\frac{n!}{(n-x)!\,n^x}}_{\to 1}\ \frac{\lambda^x}{x!}\ \underbrace{\left(1-\frac{\lambda}{n}\right)^{n}}_{\to e^{-\lambda}}\ \underbrace{\left(1-\frac{\lambda}{n}\right)^{-x}}_{\to 1}$$

$$\boxed{\frac{\lambda^x}{x!}e^{-\lambda}}$$

$n \to \infty$

The PMF of a binomial distribution $Binomial(n, p)$ can be approximated by that of $Possion(\lambda)$ with $\lambda = np$.

This works well if $np < 10$ and $n > 50$.

Example 4

Let $X \sim Binomial(100, 0.02)$.

- $\mathbb{P}(X = 2) = \binom{100}{2} 0.02^2 0.98^{98} \approx 0.273$.
- Approximation by $Possion(100 \times 0.02)$,

$$\mathbb{P}(X = 2) = e^{-2}\frac{2^2}{2!} = 0.271.$$

$n = 100 \quad p = 0.02$

$\lambda = 100 \times 0.02.$
$= np.$
$= 2$

Continuous random variables, PDF and CDF

A **continuous random variable** typically is a random variable whose set of possible values is an interval of real numbers or a union of such intervals.

(a, b), $[a, b]$ $(0, 1) \cup [3, 6)$.

Example: An air sample is analyzed and the fraction X of oxygen in the sample is determined (e.g. $X = 0.15$ means that 15% of the volume is taken up by oxygen).

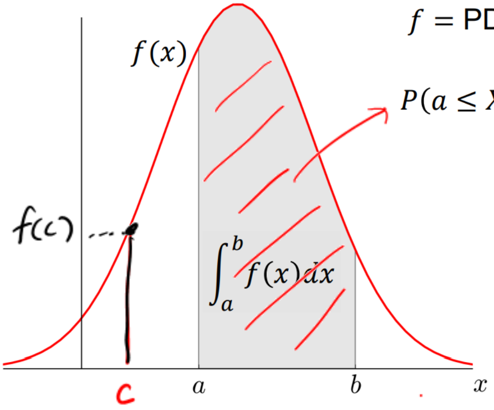
Set of possible values of $X \in [0, 1]$ (interval of real numbers x with $0 \le x \le 1$).

A function f that assigns a nonnegative real number $f(x)$ to each real number x is a **probability density function (PDF)** for a continuous random variable X if

$$\mathbb{P}(\underline{\underline{a \leq X \leq b}}) = \int_a^b f(x)\, dx$$

for all real numbers a, b, $a \leq b$.

- Note $\int_a^b f(x)\, dx$ is the area between the graph of f and the segment of the x-axis between a and b.
- If necessary, we write f_X instead of f to indicate that f belongs to X.



f = PDF of X

$$P(a \le X \le b) = \int_a^b f(x)dx$$

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

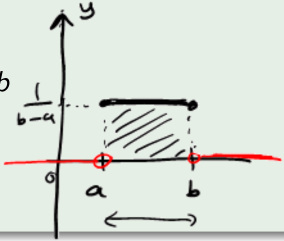
$$P(X=c) \neq f(c).$$

$$P(\underline{X=c}) = 0$$

Example 5 (Uniform distribution)

The random variable X has a **uniform distribution** if its PDF $f(x)$ is equal to a constant on its support. In particular, if the support is the interval $[a, b]$, then

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise.} \end{cases}$$

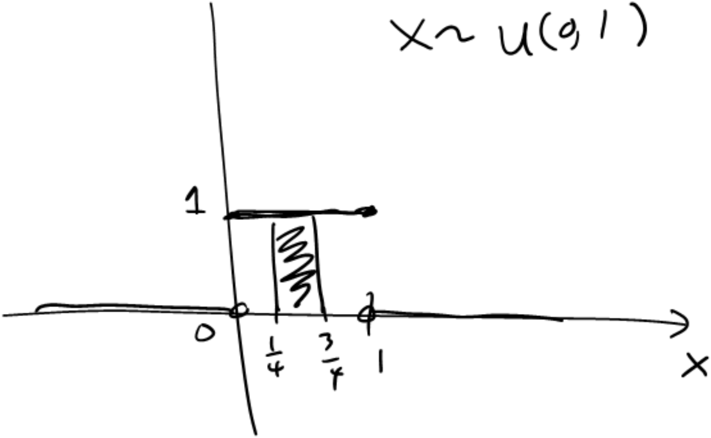


We shall denote it by $X \sim U(a, b)$.

Example: If $X \sim U(0, 1)$, then

$$\mathbb{P}\left(\frac{1}{4} \leq X \leq \frac{3}{4}\right) = \int_{1/4}^{3/4} f(x)\, dx = \int_{1/4}^{3/4} 1\, dx = \frac{3}{4} - \frac{1}{4} = \frac{2}{4} = \frac{1}{2}.$$

$X \sim U(0,1)$



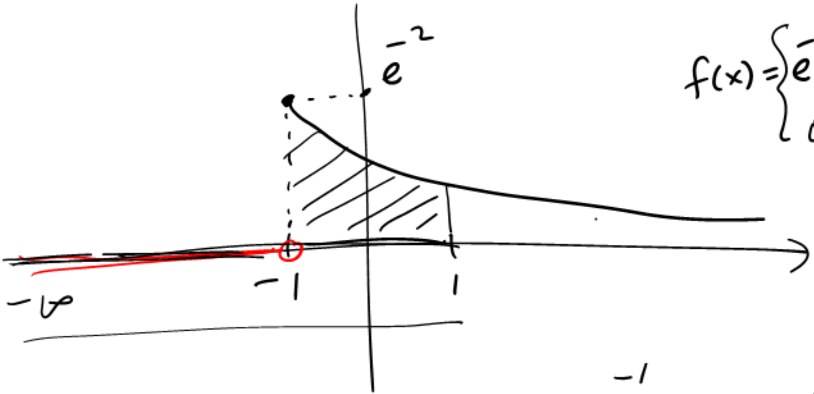
Example 6

Suppose X has PDF

$$f(x) = e^{-x-1}, \quad -1 \leq x < \infty.$$

Compute $\mathbb{P}(X \leq 1)$ and $\mathbb{P}(X \geq 1)$.

Give it a try!



$$f(x) = \begin{cases} e^{-x-1} & x \geq -1 \\ 0 & x < 0 \end{cases}$$

$$\mathbb{P}(X \leq 1) = \int_{-\infty}^{1} f(x)\, dx = \int_{-\infty}^{-1} 0 \, dx + \int_{-1}^{1} e^{-x-1}\, dx$$

$$= \int_{-1}^{1} e^{-x-1}\, dx$$

$$= e^{-1}\left[\int_{-1}^{1} e^{-x}\, dx\right]$$

$$= e^{-1}\left[\left[-e^{-x}\right]_{-1}^{1}\right]$$

$$= e^{-1}\left[-e^{-1} - \left(-e^{1}\right)\right]$$

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

$$\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X < 1)$$

$$= 1 - \mathbb{P}(X \leq 1)$$

$$= 1 - (1 - e^{-2})$$

$$= e^{-2}$$

#.

If X is a continuous random variable with PDF $f(x)$, then the **Cumulative Density Function (CDF)** of X is defined by

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

Note:

- $F(x)$ is nondecreasing
- $0 \leq F(x) \leq 1$. ✓
- $F'(x) = \frac{dF}{dx} = f(x)$ (PDF) ✓

$$\frac{dF}{dx} = \underline{\underline{\frac{d}{dx}\int_{-\infty}^{x} f(t)\, dt}} = \underline{\underline{f(x)}}$$

Fundamental Therem of Calculus.

Discrete	Continuous
PMF $\mathbb{P}(X=x)=p(x)$	PDF $\mathbb{P}(a\leq X\leq b) = \int_a^b f(x)\,dx$
CDF $\mathbb{P}(X\leq x)=\sum_{t\leq x} p(t)$	CDF $\mathbb{P}(X\leq x)=\int_{-\infty}^{x} f(t)\,dt$

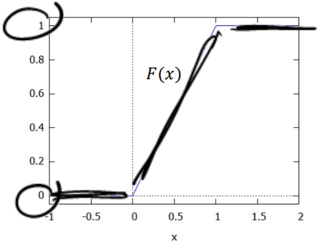
Example 7

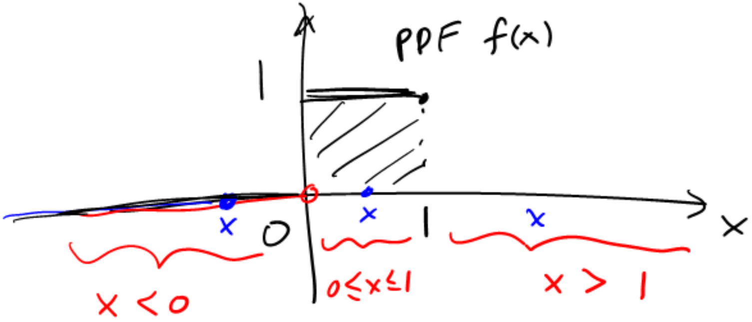
Let $X \sim U(0,1)$ be the uniform distribution on $[0,1]$. Its CDF is given by

$$F(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x \geq 1. \end{cases}$$

Note: For $0 \leq x \leq 1$, we have

$$F(x) = \int_{-\infty}^{x} f(t)\, dt = \int_{0}^{x} 1\, dt = x.$$





$\boxed{x<0}$: $F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{x} 0\,dt = 0$.

$\boxed{0 \le x \le 1}$: $F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{0} 0\,dt + \int_{0}^{x} 1\,dt$

$= x$.

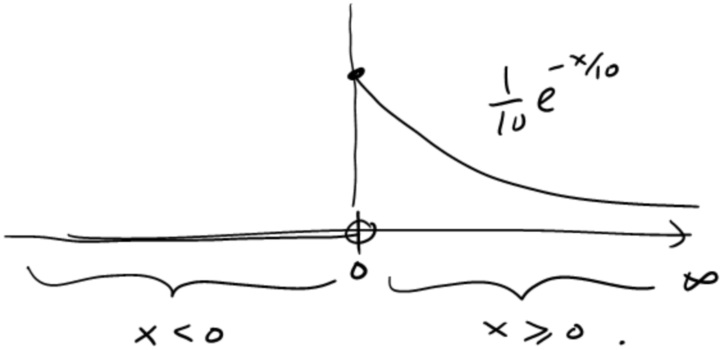
$\boxed{x>1}$: $F(x) = 1$

Example 8

Suppose X has PDF

$$f(x) = \begin{cases} \frac{1}{10} e^{-x/10} & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Compute the CDF of X.



$\boxed{x<0}$: $F(x) = \int_{-\infty}^{x} 0\,dt = 0$

$\boxed{x \geq 0}$: $F(x) = \int_{-\infty}^{x} f(t)\,dt = \underbrace{\int_{-\infty}^{0} 0\,dt}_{0} + \int_{0}^{x} \frac{1}{10}e^{-t/10}\,dt$

✓

$$= \frac{1}{\cancel{10}}\left[\frac{e^{-\frac{t}{10}}}{-\frac{1}{\cancel{10}}}\right]_0^x$$

$$= \left(-e^{-\frac{x}{10}}\right) - (-1)$$

$$= 1 - e^{-\frac{x}{10}}$$

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-\frac{x}{10}} & \text{if } x \geq 0 \end{cases}$$

Let X be a continuous random variable with PDF $f(x)$.

- Its **expected value** or **mean** is

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f(x)\,dx.$$

Discrete

$$\mathbb{E}[X] = \sum_{x} x \cdot p(x)$$

- If $g(X)$ is a function of X, then

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\,dx.$$

$$\mathbb{E}[g(x)] = \sum_{x} g(x) p(x)$$

- Similar to discrete random variables, the **variance** and **standard deviation** σ of X can be calculated as follows (where $\mu = \mathbb{E}[X]$):

$$\mathrm{Var}[X] = \mathbb{E}[(X-\mu)^2] = \mathbb{E}[X^2] - \mu^2, \quad \sigma = \sqrt{\mathrm{Var}[X]}.$$

Example 9

The total amount of medical claims (in millions) of the employees of a company has the PDF given by

$$f(x) = \begin{cases} 30x(1-x)^4, & 0 < x < 1. \\ 0 & \text{otherwise.} \end{cases}$$

Find

(i) The mean and the standard deviation of the total in dollars.

(ii) The probability that the total exceeds \$0.2 millions.

Solution.
(i)

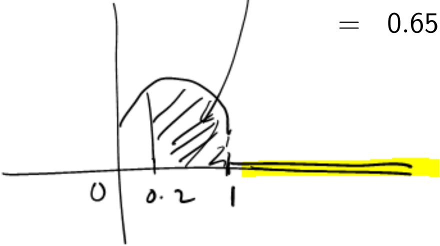
$$\begin{aligned} \text{mean} \quad \mu = \mathbb{E}[X] &= \int_{-\infty}^{\infty} x f(x)\, dx = \int_0^1 x(30x)(1-x)^4\, dx \\ &= 0.286 \text{ millions.} \end{aligned}$$

$$\begin{aligned} \text{variance} = \mathbb{E}[X^2] - \mu^2 &= \int_{-\infty}^{\infty} x^2 f(x)\, dx - (0.286)^2 \\ &= \int_0^1 x^2(30x)(1-x)^4\, dx - (0.286)^2 \\ &= 0.107 - (0.286)^2 = 0.025204 \text{ millions.} \end{aligned}$$

$$\sigma = \sqrt{\mathrm{Var}[X]} = \sqrt{0.025204} = 0.159 \text{ millions.}$$

(ii)

$$\begin{aligned}\mathbb{P}(X > 0.2) &= \int_{0.2}^{\infty} 30x(1-x)^4\,dx \\ &= \int_{0.2}^{1} 30x(1-x)^4\,dx \\ &= 0.6554.\end{aligned}$$



The Exponential and Gamma distribution

We now turn to a continuous random variable that is related to Poisson distribution.

$[0,1]$

- Let λ be the mean/average number of occurrences per unit interval.
- Let $X \sim Poisson(\lambda w)$ be the random variable that counts the number of occurences in an interval of size w.
- Then $\mathbb{P}(\text{no occurences in } [0, w]) = \mathbb{P}(X = 0) = e^{-\lambda w}$.

$$\text{PMF} = \frac{e^{-\lambda}\lambda^{x}}{x!} = P(X=x) \quad \text{if } X \sim \text{Poisson}(\lambda)$$

$$\frac{e^{-\lambda w}(\lambda w)^{0}}{0!} = e^{-\lambda w}$$

λ = # arrivals per unit interval.

$\times w$ $\downarrow$ $\times w$ $\downarrow$

λw = # arrivals per interval of size w.

$X \sim \text{Poisson}(\lambda w)$

X = # arrivals in $[0, w]$

$$\mathbb{P}\left[\text{no occurence in } [0,w]\right]$$

$$= \mathbb{P}(X = 0)$$

Let W = waiting time until the first occurence. Then its CDF $F(w)$ is given by

$$F(w) = \mathbb{P}(W \leq w) = 1 - \mathbb{P}(W > w)$$
$$= 1 - \mathbb{P}(\text{no occurences in } [0, w]) = 1 - e^{-\lambda w}$$

Note that W is nonnegative. For $w \geq 0$, the PDF of W is

$$\frac{dF}{dw} = f(w) = \lambda e^{-\lambda w}.$$

W PDF $f(w) = \lambda e^{-\lambda w}$ $w \geq 0$

$F(w) = 1 - e^{-\lambda w}$

We often let $\lambda = \frac{1}{\theta}$, and say that the random variable X has an **exponential distribution**, denoted by $X \sim Exp(\theta)$, if its PDF is defined by

$$f(x) = \frac{1}{\theta} e^{-x/\theta}, \quad 0 \leq x < \infty.$$

$$= \lambda e^{-\lambda x}$$

Theorem 10 (Exponential distribution)

If $W \sim Exp(\theta)$, then

$$\mathbb{E}[W] = \theta, \quad \mathrm{Var}[W] = \theta^2.$$

$$\mathbb{E}[W] = \frac{1}{\lambda} \quad \mathrm{Var}[W] = \frac{1}{\lambda^2}.$$

Example 11

Customers arrive in a certain shop according to a Poisson process at mean rate of 20 per hour. What is the probability that the shopkeeper will have to wait more than 5 minutes for the arrival of the first customer?

W = waiting time (minutes) until first customer.

Want $\mathbb{P}(W > 5)$

$W \sim Exp(\theta)$ $\qquad \theta = \frac{1}{\lambda}$

Poisson :

average of 20 per 60 minutes.

" " $\boxed{\frac{20}{60}}$ per minute.

$\lambda = \frac{1}{3}$.

$$W \sim \text{Exp}\left(\theta = \frac{1}{\lambda} = \frac{1}{1/3} = 3\right).$$

$$P(W>5) = \int_5^\infty \frac{1}{\theta} e^{-x/\theta}\, dx$$

$$= \int_5^\infty \frac{1}{3} e^{-x/3}\, dx$$

$$= \frac{1}{3}\left[\frac{e^{-x/3}}{-1/3}\right]_5^\infty = 0 + e^{-5/3}$$

$$\approx 0.1889$$

#.

Let W denote the waiting time until the αth occurrence in a Poisson process with $\lambda = \frac{1}{\theta}$. Then W has a **Gamma distribution** with **shape** parameter α and **scale** parameter θ, denoted by $W \sim Gamma(\alpha, \theta)$, with PDF given by

$$f(w) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} w^{\alpha-1} e^{-w/\theta}, \quad 0 \leq w < \infty.$$

$$\xrightarrow{\alpha=1} \frac{1}{\theta} e^{-w/\theta}.$$

Here:

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y}\, dy, \quad \alpha > 0.$$

For our purpose, α is usually a positive integer, and so

$$\Gamma(\alpha) = (\alpha - 1)!.$$

$$Gamma(\alpha = \underline{1}, \theta) = Exp(\theta)$$

Theorem 12 (Gamma distribution)

If $W \sim Gamma(\alpha, \theta)$, then

$$\mathbb{E}[W] = \alpha\theta, \quad \mathrm{Var}[W] = \alpha\theta^2.$$

averge 30 per 60 mutes

averg. $\frac{30}{60} = \frac{1}{2}$ per minute.

Example 13

Suppose the number of customers per hour arriving at a shop follows a Poisson distribution with mean 30. What is the probability that the shopkeeper will wait for more than 5 minutes until the second customer arrives?

$$\alpha = 2 \qquad \lambda = \frac{1}{2}$$

$$\Rightarrow \theta = \frac{1}{\lambda} = 2.$$

Solution. W = waiting time (in minutes) until the second customer arrives. Then

$$W \sim \mathit{Gamma}\,(\alpha = 2, \theta = 2)\,. \quad (\text{why?})$$

Want to compute $\mathbb{P}(W > 5)$.

$$
\begin{aligned}
\mathbb{P}(W>5) &= \int_5^\infty \frac{1}{\Gamma(2)(2)^2} w^{2-1} e^{-w/2}\, dw \\
&= \frac{1}{4}\int_5^\infty w e^{-w/2}\, dw \\
&= \frac{1}{4}\left(\left[-2we^{-w/2}\right]_5^\infty - \int_5^\infty (-2)e^{-w/2}\, dw\right) \\
&\quad \text{(by integration-by-parts)} \\
&= \frac{1}{4}\left(10e^{-5/2} + 2\left[(-2)e^{-w/2}\right]_5^\infty\right) \\
&= \frac{1}{4}\left(10e^{-5/2} + 4e^{-5/2}\right) \\
&= 0.2873.
\end{aligned}
$$

$\square$