* Announcement

① HW 10 due today (11:59 PM PT)

② HW 11 ~ 11/16

③ No class / OH on Wed (11/11)

④ Quiz 9: Ch 9 - Ch 10.2

After today's lecture

→ HW 11: Q1, Q2, Q3 (Sec 10.1, 10.2)

Q4, Q5 (Sec 10.3)

STAT 88: Lecture 32

Contents

Section 10.3: Exponential Distribution
Section 10.4: Normal Distribution

Warm up: Let $T \sim \text{Exp}(\lambda)$, so it has density $f(t) = \lambda e^{-\lambda t}$.

(a) Find $E(T)$.

(b) Find $\text{Var}(T)$.

Integration by parts

(a) $E(T) = \int_0^\infty t \lambda e^{-\lambda t}\, dt = \int_0^\infty u(t) \cdot v'(t)\, dt = u(t)v(t) \Big|_{t=0}^{\infty} - \int_0^\infty u'(t) v(t)\, dt$

$u = t$, $v = -e^{-\lambda t}$ $\rightsquigarrow$ $u'(t) = 1$, $v'(t) = \lambda e^{-\lambda t}$

$= \underbrace{-t e^{-\lambda t} \Big|_{t=0}^{\infty}}_{=0} + \int_0^\infty e^{-\lambda t}\, dt$

$= \int_0^\infty e^{-\lambda t}\, dt$

$= -\frac{1}{\lambda} e^{-\lambda t} \Big|_{t=0}^{\infty} = \frac{1}{\lambda}$

(b) $E(T^2) = \int_0^\infty t^2 \lambda e^{-\lambda t}\, dt$

$= \frac{2}{\lambda^2}$ (Exercise)

$\text{Var}(T) = E(T^2) - (ET)^2 = \frac{1}{\lambda^2}$ $\rightsquigarrow$ $SD(T) = \sqrt{\frac{1}{\lambda^2}} = \frac{1}{\lambda}$

Last time

Exponential distribution:

An exponential random variable, $T \sim \text{Exp}(\lambda)$, has density $f(t) = \lambda e^{-\lambda t}$. It models the lifetime of objects that don't age, such as a lightbulb.

The cdf of T is

$$F(t) = P(T \le t) = 1 - e^{-\lambda t}. \quad = \int_0^t f(s)\,ds$$

The survival function of T is

$$S(t) = P(T > t) = \underset{\substack{\shortparallel \\ 1-F(t)}}{e^{-\lambda t}}.$$

Memoryless property:

The chance that T survives past time $t + s$, given that it survives past time t is

$$P(T > t + s \mid T > t) = \frac{P(T > t + s)}{P(T > t)} = e^{-\lambda s} = P(T > s).$$

The memoryless property is an excellent reason not to use the exponential distribution to model the lifetimes of people or of anything that ages.

10.3. The Exponential Distribution

Let $\lambda = 2$ bulbs / decade

$\Rightarrow E(T) = \frac{1}{2}$

Expected life time = 5 years

Mean and SD Let $T \sim \text{Exp}(\lambda)$. Then

$$E(T) = \frac{1}{\lambda}.$$

The bigger λ is the sooner the bulb dies.

The variance and SD are given by

$$\text{Var}(T) = \frac{1}{\lambda^2}, \quad \text{SD}(T) = \sqrt{\text{Var}(T)} = \frac{1}{\lambda}.$$

The SD of the exponential distribution is the same as the mean.

Example: Let $X \sim \text{Exp}(\lambda)$ with $E(X) = 10$. Find $P(X > 25 \,|X > 10)$.

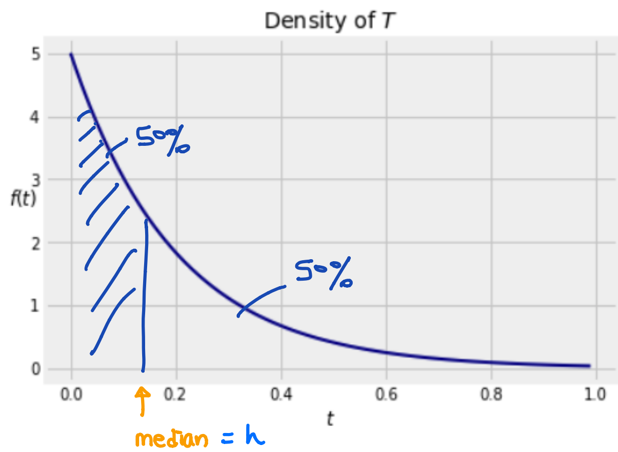
$S(t) = P(T > t) = e^{-\lambda t}$

$E(X) = \frac{1}{\lambda} = 10 \quad \Rightarrow \quad \lambda = 0.1$

$P(X > 25 \mid X > 10) = P(T > 15) = e^{-1.5}$

Memoryless prop.

Median The median is the 50th percentile.



We need to find h such that

$$F(h) = 0.5 \text{ or } S(h) = 0.5.$$

" $\log(\frac{1}{x}) = -\log x$ "

Then

$$F(h) = S(h) \Longrightarrow \quad 1 - e^{-\lambda h} = e^{-\lambda h}$$

$$\Rightarrow \quad 1 = 2e^{-\lambda h}$$

$$\Rightarrow \quad \frac{1}{2} = e^{-\lambda h}$$

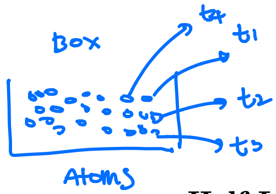
$$\Rightarrow \quad \log(\tfrac{1}{2}) = -\lambda h$$

$$\Rightarrow \quad h = \frac{\log(\frac{1}{2})}{-\lambda} = \frac{-\log 2}{-\lambda} = \frac{\log 2}{\lambda}$$

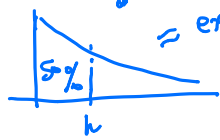
So the median of exponential lifetime is

$$h = \frac{\log 2}{\lambda}. \quad < \quad E(T) = \frac{1}{\lambda}$$

half life
= median of exponential lifetime.



$t_1, t_2, \ldots, t_N$, Histogram



$\approx$ exponential distribution

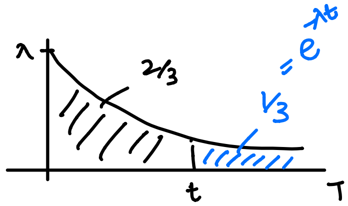
Half Life We often model the time until a radioactive particle is emitted using the exponential distribution.

The half life of a radioactive isotope is defined as the time by which half of the atoms of the isotope will have decayed. This means the half life is the median of the exponential lifetime of an atom.

rate of decay

$$h = \frac{\log 2}{\lambda} \quad \text{or} \quad \lambda = \frac{\log 2}{h}.$$

Example: (Exercise 10.5.4) Strontium 90 has a half-life of 28.8 years. Assuming exponential decay, about how many years will it take for 2/3 of a lump of Strontium 90 to decay?



Find t

$P(T \le t) = F(t) = \frac{2}{3}$

$= 1 - e^{-\lambda t}$

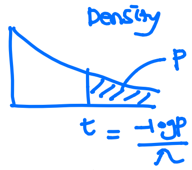
$\leadsto \lambda = \frac{\log 2}{28.8} = 0.024$

$\Rightarrow e^{-\lambda t} = \frac{1}{3}$

$\Rightarrow -\lambda t = \log\left(\frac{1}{3}\right)$

$\Rightarrow t = \frac{\log\left(\frac{1}{3}\right)}{-\lambda} = \frac{-\log 3}{-0.024}$

Radiocarbon Dating We can generalize the calculation above to any proportion. remaining



$$p = e^{-\lambda t} \iff t = \frac{-\log p}{\lambda}.$$

This is used to estimate the age of an object containing animal or plant material. When an animal or plant dies, its radiocarbon ^{14}C decays exponentially with half life h = 5730 years. If the proportion of ^{14}C to C is p, we can estimate when the animal or plant dies:

half life formula

$$t = \frac{-\log p}{\lambda} \text{ where } \lambda = \frac{\log 2}{5730}, \leftarrow h$$

therefore

$$t = \frac{-\log p \cdot 5730}{\log 2}.$$

Example: If the ratio of ^{14}C to C remaining in object is 0.3, the organism died approximately

$$t = \frac{-\log(0.3) \cdot 5730}{\log 2}.$$

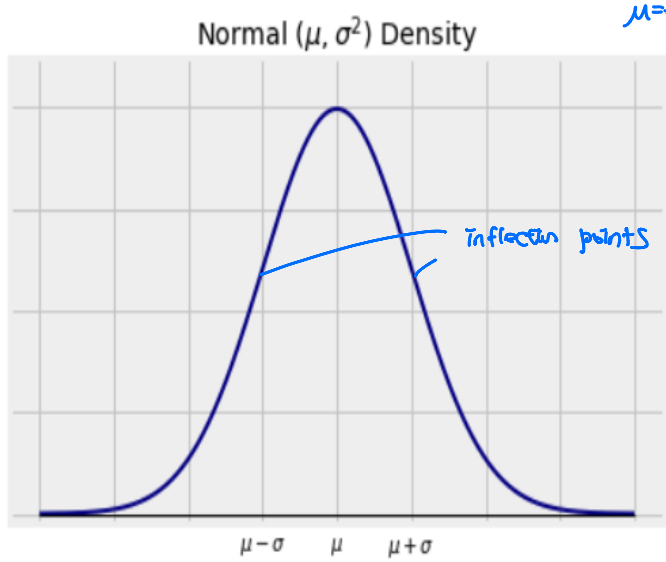
```
-np.log(0.3) / (np.log(2) / 5730)
```

```
9952.812854572363
```

10.4. The Normal Distribution

The random variable X has the normal distribution $\mathcal{N}(\mu, \sigma^2)$ if the density of X is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \text{ for } -\infty < x < \infty.$$



$\mu=0$, $\sigma=1$

→ standard normal curve

The random variable Z has the standard normal distribution if the distribution of Z is $\mathcal{N}(0, 1)$.

We have been treating the parameters μ and σ like the mean and the SD of the distribution, and indeed it turns out (after calculations that we will not do) that:

$$E(X) = \mu \text{ and } \mathrm{SD}(X) = \sigma.$$

$$= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\cdot\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx$$

To find $P(X < x)$,

$$P(X < x) = P\left(\underbrace{\frac{X-\mu}{\sigma}}_{= Z} < \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Sums of Independent Normal Variables

Fact: The sum of two independent normal random variables is a normal random variable.

$$X \sim \mathcal{N}(\mu_X, \sigma_X^2) \text{ and } Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2).$$

If X and Y are independent,

$$X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2).$$

This result extends to linear combinations of independent normal random variables.

Example: What is the distribution of $X - 2Y + 3$?

$\sim N(\quad , \quad)$

$E(X-2Y+3) = E(X) - 2E(Y) + 3$

$= \mu_X - 2\mu_Y + 3$

$Var(X-2Y+3) = Var(X-2Y)$

Indep. $= Var(X) + Var(-2Y)$

$= Var(X) + 4Var(Y)$

$= \sigma_X^2 + 4\sigma_Y^2$

Example: (Exercise 10.5.5) The weights of five randomly sampled people are i.i.d. normally distributed random variables with mean 150 pounds and SD 20 pounds. Let W be the total weight of the five people, measured in pounds. If possible, find w such that $P(W > w) = 0.05$. If this is not possible, explain why not.

$$X_1, X_2, \cdots, X_5 \overset{\text{i.i.d.}}{\sim} N(150, \underset{20^2}{\underset{\shortparallel}{400}})$$

$$W = X_1 + X_2 + \cdots + X_5 \sim N(5 \cdot 150, 5 \cdot 400)$$

$$P(W > w) = 0.05$$

$$\Rightarrow P(W \le w) = 0.95$$

$$\shortparallel$$

$$P\Big(\underbrace{\frac{W - 750}{\sqrt{2000}}}_{\substack{\shortparallel \\ Z}} \le \frac{w - 750}{\sqrt{2000}}\Big)$$

$$\shortparallel$$

$$\Phi\left(\frac{w-750}{\sqrt{2000}}\right)$$

$$\mathrm{Var}(X_1 + \cdots + X_5)$$
$$= \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_5)$$
$$= 5 \cdot 400$$

stats.normal.ppf(0.95)

$$\Rightarrow \frac{w - 750}{\sqrt{2000}} = \Phi^{-1}(0.95) = 1.64$$

$$\Rightarrow w = 1.64 \cdot \sqrt{2000} + 750$$

$S_n = X_1 + \cdots + X_n$, $E(X_i) = \mu$, $\mathrm{Var}(X_i) = \sigma^2$

$SD(X_i) = \sigma$

$$E(S_n) = E(X_1) + \cdots + E(X_n)$$
$$= n \cdot \mu$$

$$\mathrm{Var}(S_n) \underset{\text{indep.}}{=} \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) = n \cdot \sigma^2$$

$$SD(S_n) = \sqrt{\mathrm{Var}(S_n)} = \sqrt{n} \cdot \sigma$$

Confidence Interval for the Difference Between Means Suppose you have two independent samples as follows:

- $X_1, X_2, \ldots, X_n$ are i.i.d. with mean μ_X and SD σ_X.
- $Y_1, Y_2, \ldots, Y_m$ are i.i.d. with mean μ_Y and SD σ_Y.

} indep.

You want to estimate the difference $\mu_X - \mu_Y$. Then $\bar{X} - \bar{Y}$ is an unbiased estimator for $\mu_X - \mu_Y$.

$E(\bar{X}-\bar{Y}) = \mu_X - \mu_Y$

By CLT, we know

- $\bar{X}$ is approximately $\mathcal{N}(\mu_X, \frac{\sigma_X^2}{n})$.
- $\bar{Y}$ is approximately $\mathcal{N}(\mu_Y, \frac{\sigma_Y^2}{m})$.

Then $\bar{X} - \bar{Y}$ is approximately $\mathcal{N}(\qquad , \qquad)$ and an approximate 95% CI for $\mu_X - \mu_Y$ is given by

$$\bar{X} - \bar{Y} \pm 2\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}.$$

Example: Suppose you have drawn samples of people independently from two cities, and suppose you have collected the following data:

- The incomes of the 400 sampled people in City X have an average of 70,000 dollars and an SD of 40,000 dollars.
- The incomes of the 600 sampled people in City Y have an average of 80,000 dollars and an SD of 50,000 dollars.

Find a 95% CI for the difference between the mean incomes in the two citie