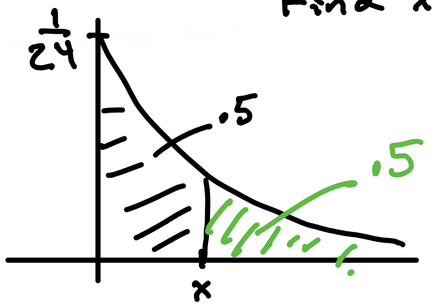
Stat 88 lec 32

warmup : 2:00 - 2:10

3. Let X have the exponential distribution with mean 24 hours. Assume that X is measured in hours.

Find x such that $P(X \le x) = .5$



$f(x) = \lambda e^{-\lambda x}$

$E(X) = \frac{1}{\lambda}$

$F(x) = .5$ find x

$\quad \parallel$

$P(X \le x)$

$F(x) = .5 \Rightarrow S(x) = .5$

$\qquad \parallel$

$P(X > x) = e^{-\lambda x} = e^{-\frac{1}{24}x}$

$e^{-\frac{1}{24}x} = .5$ solve for x

$\log\left(e^{-\frac{1}{24}x}\right) = \log(.5)$

$\quad \parallel$

$-\frac{1}{24}x = \log(.5) \Rightarrow x = \frac{\log(.5)}{-\frac{1}{24}} = \boxed{\frac{\cancel{-}\log(2)}{\cancel{-}\,1/24}}$

log laws

$\log\left(\frac{1}{a}\right) = -\log a$

called half life or median of exponential distribution.

Last time

Sec 10.3 Exponential Distribution

An exponential RV,

$T \sim Exp(\lambda)$, 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 $F(t) = 1 - e^{-\lambda t}$

$P(T \leq t)$

The survival function of T is $S(t) = e^{-\lambda t}$

$P(T > t)$

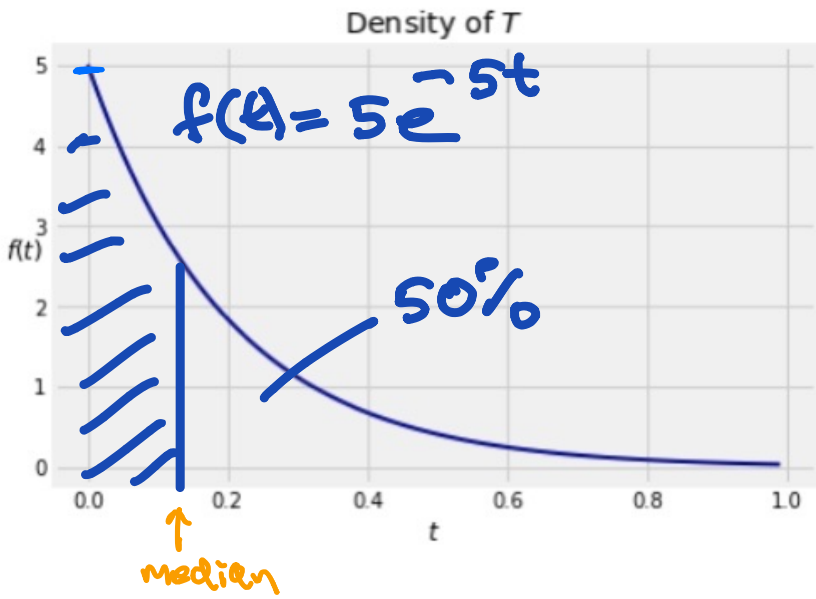
Today ① Sec 10.3 half life of radioactive atoms

② Sec 10.4 The normal distribution

① sec 10.3 half life of radioactive atoms

<u>Median</u>

The median is the 50^{th} percentile



Find t such that,

$$F(t) = .5 \iff S(t) = 1/2$$

$$\frac{1}{2} = S(t)$$

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

$$\Rightarrow \log\left(\frac{1}{2}\right) = -\lambda t$$

$$\Rightarrow -\log(2) = -\lambda t$$

$$\Rightarrow t = \frac{\log(2)}{\lambda}$$

half life = median of exponential lifetime.

$$\boxed{h = \frac{\log(2)}{\lambda}}$$

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

The <u>half life</u>, h, is the median of the exponential lifetime of an atom,

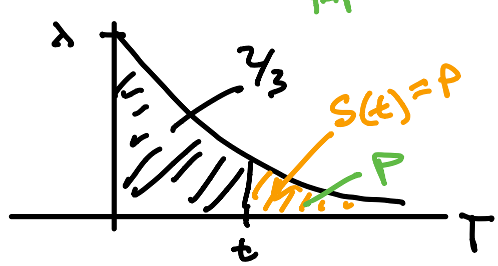
$$h = \frac{\log(2)}{\lambda}$$ defines the rate, λ, of decay.

$$\lambda = \frac{\log(2)}{h}$$

exercise 10.5.4

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?

$= 1-P$



Find t

$$\lambda = \frac{\log(2)}{h}$$

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

$$\lambda = \frac{\log(2)}{28.8} = .024$$

$$F(t) = 2/3 \Rightarrow S(t) = 1/3$$
$$S(t) = e^{-.024t}$$

$$\Rightarrow -.024t = \log(1/3)$$

$$\boxed{t = \frac{\log(1/3)}{-.024}}$$

For any proportion p of remaining material,

Find t such that $S(t) = P \Rightarrow \boxed{t = \frac{\log(p)}{-\lambda}}$

<u>radiocarbon dating</u>

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 5730 yrs.

If the proportion of ^{14}C to C is p, we can estimate when the animal or plant died,

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

half life formula

$$\Rightarrow t = \frac{-\log(p)}{\log(2)/5730}$$

ex if the ratio of ^{14}C to C is $1/3$ ← p (the proportion of remaining ^{14}C)

the organism died approx

$$\frac{-\log(1/3)}{\log(2)/5730}$$

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

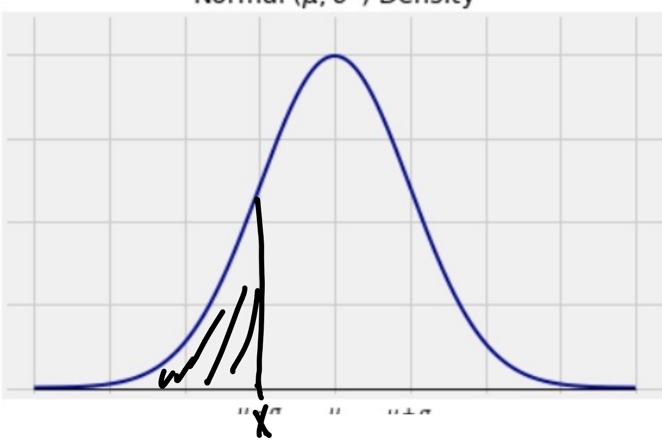
```
9952.812854572363
```

② Sec 10.4 The normal distribution

The random variable X has the *normal* (μ, σ^2) distribution if the density of X is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty$$

Normal (μ, σ^2) Density



To find

$P(X < x) =$

$P\left(\frac{X-\mu}{\sigma} < \frac{x-\mu}{\sigma}\right)$

$\frac{X-\mu}{\sigma} = Z \sim N(0,1)$

Sums of independent normal variables

Fact: the sum of two independent normal RVs is a normal RV.

$X \sim N(\mu_x, \sigma_x^2)$, $Y \sim N(\mu_y, \sigma_y^2)$ } Indep

$X+Y \sim N(\mu_x+\mu_y, \sigma_x^2+\sigma_y^2)$

This result extends to linear combinations of independent normals.

Ex What distribution is $X-2Y+3$?

$\boxed{N(\mu_x-2\mu_y+3, \sigma_x^2+4\sigma_y^2)}$

$E(X-2Y+3) = E(X)-2E(Y)+3 = \mu_x-2\mu_y+3$

$Var(X-2Y+3) = Var(X)+4Var(Y) = \sigma_x^2+4\sigma_y^2$

exercise 10.5.5

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.

$$X_1, X_2, X_3, X_4, X_5 \overset{iid}{\sim} N(150, 20^2)$$

$$W = X_1 + X_2 + X_3 + X_4 + X_5 \sim N(5(150), 5(20^2))$$

$$P(W > w) = .05$$

$$P(W < w) = .95$$

$\overset{\shortparallel}{}$ ← same as $\leq$

$$P\left(Z < \frac{w - 5(150)}{\sqrt{5}(20)}\right)$$

stats.norm.ppf (.95)

$$\Rightarrow \frac{w - 5(150)}{\sqrt{5}(20)} = \Phi^{-1}(.95) = 1.64$$

$$w = 5(150) + 1.64\left(\sqrt{5}(20)\right) = \boxed{823.34}$$

Confidence interval for the difference between means

- $X_1, X_2, \ldots, X_n$ 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

$\bar{X} - \bar{Y}$ is an unbiased estimator for $\mu_X - \mu_Y$

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

$$\operatorname{var}(\bar{X}-\bar{Y}) = \operatorname{var}(\bar{X}) + \operatorname{var}(\bar{Y}) \quad \text{since } \bar{X} \text{ and } \bar{Y} \text{ is indep.}$$

$$= \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}$$

So a 95% CI for $\mu_X - \mu_Y$ is

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

ex

As an 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.

And a 95% CI for $\mu_X - \mu_Y$

$$70{,}000 - 80{,}000 \pm 2\sqrt{\frac{40{,}000^2}{400} + \frac{50{,}000^2}{600}}$$

$$= -10{,}000 \pm 2(2858) = (-15{,}716, -4824)$$

Can you reject the null $\mu_X - \mu_Y = 0$ (for alternative $\mu_X - \mu_Y < 0$ and level 5%) — yes since 0 not in 95% CI.

Test for the equity of two means (A/B test)

We wish to determine if two independent populations have the same mean (i.e. $\mu_Y - \mu_X = 0$)

ex

- 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.

H_0: the mean income in City X is the same as the mean income in City Y ($\mu_x = \mu_y$)

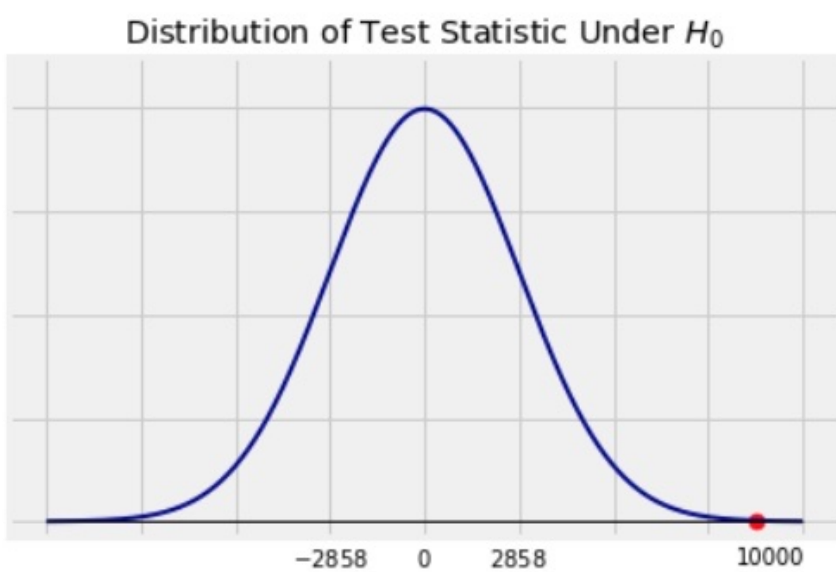
H_1: $\mu_y > \mu_x$

Our T.S. is $\bar{Y} - \bar{X}$ $= 10{,}000$ we reject the null if $\bar{Y} - \bar{X}$ is large.

under Ho:

$\mu_y - \mu_x = 0$ $\quad 2858^2$

$$\bar{Y} - \bar{X} \sim N\left(0, \frac{\sigma_x^2}{400} + \frac{\sigma_y^2}{600}\right)$$



$$\text{P-val} = 1 - \Phi\left(\frac{10{,}000 - 0}{2858}\right) = 1 - \Phi(3.5) = .0002 < .05$$

so reject null.

exercise 10.5.7

7. A simple random sample of 200 students is taken at University A. Independently, a simple random sample of 300 students is taken at University B.

- The number of football games watched by students in Sample A has an average of 1.5 and an SD of 2. The number of football games watched by students in Sample B has an average of 4 and and SD of 1.5.

$\bar{B} - \bar{A}$

a) Construct an approximate 95% confidence interval for the difference between the average number of football games watched by students in the two universities.

$$(\bar{B} - \bar{A}) \pm 2\sqrt{\frac{\sigma_B^2}{300} + \frac{\sigma_A^2}{200}}$$

$$(4 - 1.5) \pm 2\sqrt{\frac{1.5^2}{300} + \frac{2^2}{200}} = (2.17, 2.83)$$

So we can reject the null that $\mu_B - \mu_A = 0$

since 0 not in our CI.