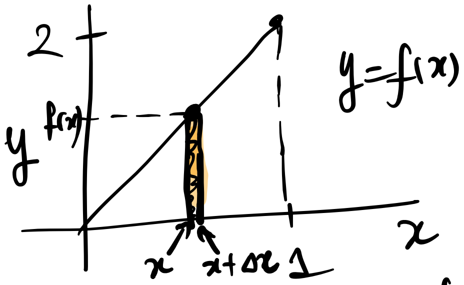
Lecture 35 4/19/2024

X is a continuous r.v. with density $f(x)$

$\boxed{\begin{array}{l} ① f(x) \geq 0 \\ ② \int_{-\infty}^{\infty} f(x)dx = 1 \end{array}}$



$$f(x) = \begin{cases} 2x, & 0<x<1 \\ 0 & o/w \end{cases}$$

Area of shaded rectangle $\approx f(x)\Delta x$

$P(x < X < x+\Delta x)$ = shaded area $\approx f(x)\Delta x$

$$f(x) \approx \frac{P(x < X < x+\Delta x)}{\Delta x}$$

histograms from Data8

$$\text{height of a bin} = \frac{\text{\% of data in bin}}{\text{width of bin}}$$

for density histograms y-axis is density.

$$f(x) \to F(x) = P(X \leq x)$$

area under $y=f(x)$ over the specified interval: $(-\infty, x]$

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

$$F'(x) = f(x)$$

Discrete r.v.

X, p.m.f $f(x) = P(X=x)$

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

$$E(X) = \sum_x x \cdot P(X=x)$$

$$E(X) = \sum_x x \cdot f(x)$$

$$E(g(X)) = \sum_x g(x) P(X=x)$$

$$= \sum_x g(x) f(x)$$

ex: $g(X) = X^2$

$$E(g(X)) = \sum_x x^2 f(x)$$

$$Var(X) = E[(X-\mu)^2]$$

$$g(X) = (X-\mu)^2$$

$$Var(X) = \sum_x (x-\mu)^2 f(x)$$

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx = \mu$$

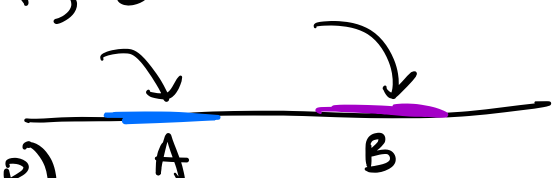
Show that $Var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$

$$Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$$

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

All the properties of expectation & variance carry over. a, b etc are always constants

1. $\mathbb{E}(aX+b) = a\mathbb{E}(X) + b$
2. $Var(aX+b) = a^2 Var(X)$
3. $SD(aX+b) = |a|\, SD(X)$
4. $\mathbb{E}(aX+bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$
5. We say that X, Y are independent if for every interval A, B



$$P(X\in A, Y\in B) = P(X\in A)P(Y\in B)$$

If X, Y are independent, then

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

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

TWO SPECIAL DISTRIBUTIONS

① exponential dsn

② normal dsn

---

① Exponential dsn.

We say that X has the exponential dsn if its density function $f(x) = \begin{cases} e^{-x}, & x>0 \\ 0, & o/w. \end{cases}$

This particular dsn $\left(f(x) = e^{-x}, x>0 \right)$ is called the exp. dsn with rate 1.

X is called an exponential r.v. with rate 1

$$X \sim exp(1)$$

In general, we say that $X \sim exp(\lambda)$ ("X has the exponential dsn with rate λ")

if $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x>0, \lambda>0 \\ 0, & o/w. \end{cases}$

λ is some positive constant.

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

$$= \cancel{\lambda}\left[-\frac{1}{\cancel{\lambda}} e^{-\lambda t}\Big|_{0}^{x}\right] = 1 - e^{-\lambda x}$$

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

We usually use T to denote an exponential r.v.

$$T \sim \exp(\lambda) \Rightarrow P(T \leq x) = 1 - e^{-\lambda x},\ x, \lambda > 0$$

$$\boxed{\boxed{P(T > x)} = \cancel{1} - (\cancel{1} + e^{-\lambda x}) = e^{-\lambda x}}$$

MEMORYLESS PROPERTY OF $T \sim \exp(\lambda)$

$$P(T > s + t \mid T > t) = P(T > s)$$

$$\mathbb{E}(T) = \frac{1}{\lambda} \leftarrow \text{To obtain this, integrate}$$

$$\int_{-\infty}^{\infty} x \cdot f(x)\,dx = \int_{0}^{\infty} x \lambda e^{-\lambda x}\,dx$$

$$\mathrm{Var}(T) = \frac{1}{\lambda^2},\quad SD(T) = \mathbb{E}(T) = \frac{1}{\lambda}$$

$P(T>t)$ is the prob. that T "survives" past time t.

So we define the survival function $S(t)$ by $S(t) = P(T>t)$

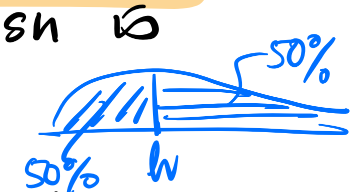
If T models the lifetime of an object $S(t)$ is the chance that it last past time t.

The "memoryless" property of the exp. dsn just means that object "forgets" it has lasted for time t.

$$P(T>s+t \mid T>t) = P(T>s)$$

Median of the exponential distribution.

this is the value s.t. half the dsn is to left of it & half to the right.



If m is the median of the exp. dsn,

$$\underbrace{P(T<m)}_{F(m)} = \underbrace{P(T>m)}_{1-F(m)}$$

$T \sim \exp(\lambda)$

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

$$F(m) = 1 - F(m)$$

$$\underbrace{1-e^{-\lambda h}}_{F(h)} = \underbrace{e^{-\lambda h}}_{S(h) = P(T>h)}$$

h for half-life

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

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

h is the halfway point of the dsn

$$\lambda h = \ln(2) = \log(2)$$

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

log is always ln unless stated otherwise

$$\boxed{h = \underbrace{\log(2)}_{\approx 0.69}\cdot E(T)}$$

$$h < E(T)$$

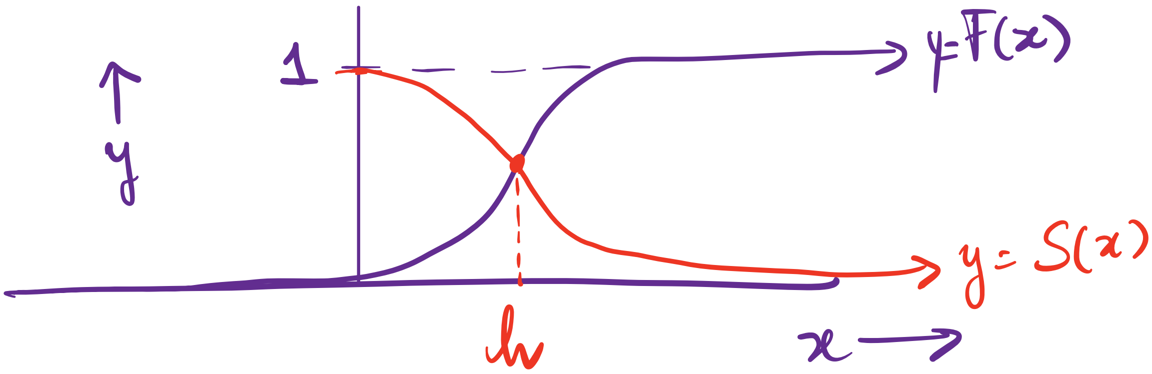
Median < average

Lec 35 Part 1 4/22/2024

h is the value on the axis s.t.

$F(h) = 1 - F(h)$

$P(T < h) = P(T > h) \leftarrow S(h)$



HALF-LIFE

The half-life of a radioactive isotope is the time until half the atoms have decayed.

λ is called the decay rate & we use $T \sim \exp(\lambda)$ to model the lifetime of a radioactive atom.

Ex. 10.5.4 Strontium-90 has a half life of 28.8 years

If we assume exponential decay, how long will it take for $\frac{2}{3}$ of a lump of Strontium-90 to decay?

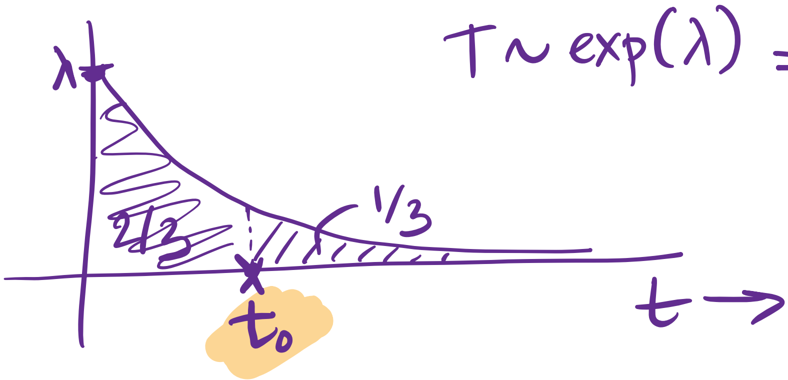
h = half-life which is the value s.t.

$F(h) = S(h)$

i.e $P(T \leq h)$ or $P(T < h) = P(T > h)$

$h = 28.8$ years

$$h = \frac{\log(2)}{\lambda} \Rightarrow \lambda = \frac{0.69}{28.8} \approx 0.024$$



$$T \sim \exp(\lambda) \Rightarrow f(t) = \lambda e^{-\lambda t}, \quad t > 0 \quad (\lambda > 0)$$

$$f(0) = \lambda$$

Want to find t_0 such that $\underbrace{P(T > t_0)}_{S(t_0)} = 1/3$

$= 1 - F(t_0)$

$= e^{-\lambda t_0}$

That is $F(t_0) = 2/3$

$\lambda \simeq 0.024$

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

We need to find t_0 such that

$$\underbrace{e^{-\lambda t_0} = \frac{1}{3}}$$

$$e^{\lambda t_0} = 3$$

$$\lambda t_0 = \log(3)$$

$$t_0 = \frac{\log(3)}{\lambda}$$

$$\boxed{t_0 = \frac{\log(3)}{\lambda}}$$

≈ 45.78 years

The Normal distribution §10.4

$f(x)$ for the standard normal distribution:

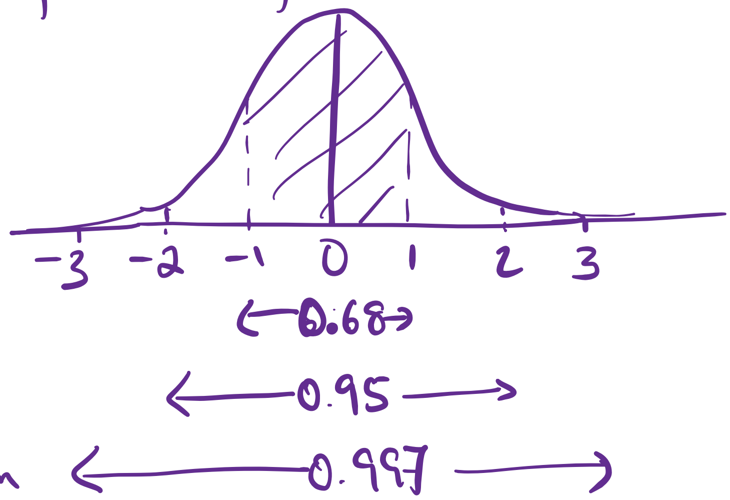
$(Z \sim N(0, 1)$, with $0 = \mu$ and $1 = \sigma^2$

$\mu = E(Z)$

$\sigma^2 = Var(Z)$

$N(\mu, \sigma^2)$ $\qquad f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbb{R}$

The standard normal p.d.f is defined on entire real line.



We can define density for a more general normal distribution centered at μ & with SD σ.

We say that $X \sim N(\mu, \sigma^2)$ distribution

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

If $X \sim N(\mu, \sigma^2)$

$$Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$ ← X in standard units has the $N(0,1)$ distribution

$E(Z) = 0$, $SD(Z) = 1$

$E(X) = \mu \qquad \mathbb{E}\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}\mathbb{E}(X-\mu) = 0$

$$SD\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{|\sigma|} SD(X-\mu) = \frac{1}{\sigma} SD(X) = \frac{\sigma}{\sigma} = 1$$

IMPORTANT FACT If X & Y are independent r.v. & $X \sim N(\mu_X, \sigma_X^2)$ & $Y \sim N(\mu_Y, \sigma_Y^2)$

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

For example If $X \sim N(0,3)$, $Y \sim N(1,4)$ X,Y ind., What is the dsn of $X+Y$?

$X+Y \sim N(0+1, 3+4)$ so $X+Y \sim N(1,7)$

What about $X+3Y-2$?

$X+3Y-2 \sim N(1,39)$

$E(X+3Y-2)$
$= 0+3-2=1$

$Var(X+3Y-2)$
$= Var(X+3Y)$
$= Var(X)+9Var(Y)$
$= 3+9 \cdot 4 = 39$

What about $X-Y$?

$E(X-Y) = -1$

$Var(X-Y) = 7$

$X-Y \sim N(-1,7)$

We use $X-Y$ or rather, dsn of the difference of 2 normal r.v. all the time for Conf. Int & Hyp. tests.

Suppose $X_1, X_2, \ldots X_n$ are iid $N(\mu_X, \sigma_X^2)$

$Y_1, Y_2, Y_3, \ldots Y_m$ are iid $N(\mu_Y, \sigma_Y^2)$

By CLT, $\bar{X} \approx N\left(\mu_X, \frac{\sigma_X^2}{n}\right)$, $\bar{Y} \approx \left(\mu_Y, \frac{\sigma_Y^2}{m}\right)$

$$\underbrace{\bar{X} - \bar{Y}} \approx N\left(\mu_X - \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}\right)$$

What is an approximate 95% CI for $\mu_X - \mu_Y$?

$$\left(\bar{X} - \bar{Y} \pm 2 * SD(\bar{X} - \bar{Y})\right)$$