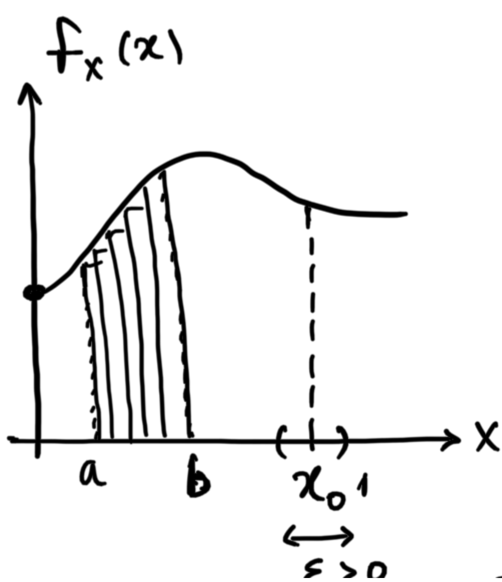
STATISTICS CLASS 1

- Agenda for today's class
  - Take home Q ✓
  - Recap of Poisson dist + Py code
  - Normal distribution
  - sample v. Population
  - Variance and covariance
  - Correlation and causation

- PDF

Density $= \dfrac{\text{mass}}{\text{vol.}}$, PDF $= \dfrac{\text{Probability}}{\text{length}} = \underbrace{f_X(x)}$



Going from PDF $\longleftrightarrow$ Prob.

$$P(a < X < b) = \int_a^b \underbrace{f_X(x)}_{\text{PDF}} \times \underbrace{dx}_{\text{length}}$$

Approximating probability at x_0

$$P\left(x_0 - \frac{\varepsilon}{2} < x_0 < x_0 + \frac{\varepsilon}{2}\right) = \int_{x_0 - \frac{\varepsilon}{2}}^{x_0 + \varepsilon/2} f_X(x)\,dx = \underbrace{f_X(x_0)}\cdot\underbrace{\varepsilon}$$

for small enough ε, $f_X(x)$ is a const $f_X(x_0)$

- Poisson Distribution

Say we are interested in no: of people visiting a website. Lets call the RV

$$X = \text{\# of visits per hour}$$

Prior data reveals $E[X] = 20/\text{hr.} = \lambda$

If X was a binomial, $E[X] = \underbrace{n \cdot p} = 20$

$$\underbrace{\lambda_{HOUR}} = \underbrace{60}_{n} \times \underbrace{\frac{\lambda_{HOUR}}{60}}_{p} = 60 \times \lambda_{MIN}$$

Assuming no: of clicks each minute are independent of each other

$60 \times \lambda_{min}$ } 60 independent bern.

$n \cdot p =$ ... } trials, with $p = \lambda_{min}$

$$\underbrace{P(X=k)}= {}^{60}C_k \cdot \left(\underbrace{\frac{\lambda}{60}}\right)^k \cdot \left(1-\frac{\lambda}{60}\right)^{60-k}$$

$$|||ly \quad P(X=k) = {}^{3600}C_k \left(\frac{\lambda}{3600}\right)^k \left(1-\frac{\lambda}{3600}\right)^{3600-k}$$

$$P(X=k) = \lim_{n\to\infty} {}^{n}C_k \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} = \underbrace{e^{-\lambda} \cdot \frac{\lambda^k}{k!}}$$

$$X \sim Pois(\lambda), \quad P(X=k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}, k = \underbrace{\{0, 1, 2, \ldots\}}$$

- Exponential Distribution

$$\text{Poisson} = \frac{\text{No: of events}}{\text{1 unit of time}}, \quad \text{Exponential} = \frac{\text{Time}}{\text{1 event}}$$

Applications
- → Survival probability
- → Life of a light bulb
- → Time b/w 2 poisson events

$$[\text{Poisson}]: \underline{\lambda} = \frac{\text{3 events}}{\text{Hour}} = \frac{\text{1 event}}{\text{20 min}}$$

$$[\text{Exponential}]: \mu = \frac{1}{\cdot} = \frac{\text{20 min}}{\cdot}$$

λ 1 event

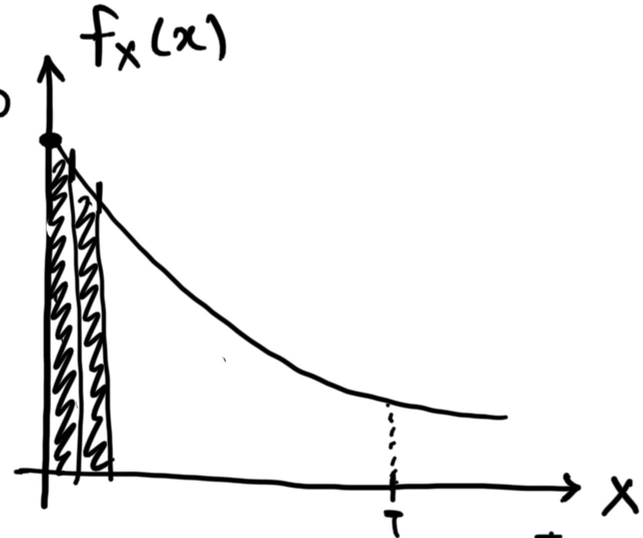
$X \sim \text{Exp}(\lambda)$ → same as poisson λ

$f_X(x) = \lambda e^{-\lambda x}, \; x > 0$

$F_X(x) = 1 - e^{-\lambda x}$

$E[X] = 1/\lambda$

$1 - F(t)$



$P(\text{Lightbulb survives till } T) = F_X(T) = \int_0^T f_X(x)\,dx$

- Memoryless Property

Waiting time does not depend on elapsed time.

Intuitively, given a bulb has survived for 't' minutes, prob. of surviving 's' more minutes is prob. of surviving 's' minutes.

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

$P((X > t+s) \cap (X > t))$ $P(X>t+s)$

$$P((X > t+s) \mid X > t) = \frac{\quad}{P(X > t)} = \frac{\quad}{P(X>t)}$$

$$\Rightarrow \underbrace{P(X > t+s)} = P(X > t) \cdot P(X > s)$$

<u>Ex 1</u>: $P(X > 2) = \underbrace{P(X > 1) \cdot P(X > 1)} = \left(P(X > 1)\right)^2$

<u>Ex 2</u>: $X \sim \underbrace{\mathrm{Exp}(\mu)}$, $P(0 \le X \le 1) = 0.05$

(a) $P(1 \le X \le 2) = (0.05)(0.95)$

(b) $P(2 \le X \le 3) = (0.05)(0.95)^2$

(c) $P(X = 1) = 0$

<u>Ex 3</u>: Say $X \sim \mathrm{Pois}(\lambda)$

λ per 1 unit $\Rightarrow$ λt per t units

$-\lambda t$ ... x

$$P(X=x) = e^{-\lambda t}\frac{(\lambda t)^x}{\underbrace{x!}}, \quad P(X=0) = \underbrace{e^{-\lambda t}}$$

$$\underbrace{F(t)}_{\text{Exp}} = 1-e^{-\lambda t}, \quad 1-F(t) = \overbrace{e^{-\lambda t}}$$

Ex 4: Memorylessness vs. Independence

Memoryless $\Leftrightarrow$ Markov Property

$\underbrace{\text{IId}} \rightarrow$ Poisson $\rightarrow$ Exponential $\rightarrow$ Memoryless

[Independence]: $P(X > A \mid B) = P(X > A)$

[Memoryless]: $P(X > t+s \mid X > s) = \underbrace{P(X > t)}_{\text{Mem.}}$

$$P(X > t+s \mid X > s) = \underbrace{P(X > t+s)}_{\text{Indep.}}$$

- Variance

How far is X from mean $E[X]$.

$$Var[X] = E[\underbrace{(X_i - E[X])}]$$

We don't look at abs $|\cdot|$ because abs values are not differentiable.

$$Var[X] = E\left[X^2 - 2X \cdot E[X] + E[X]^2 \right]$$

$$= E[X^2] - 2\,E[X]^2 + E[X]^2$$

$$= E[\underbrace{X^2}] - (E[X])^2$$

$$\sigma = SD(X) = (\underbrace{V[X]}_{(\cdot)^2})^{1/2} \Big\} \text{ same units}$$

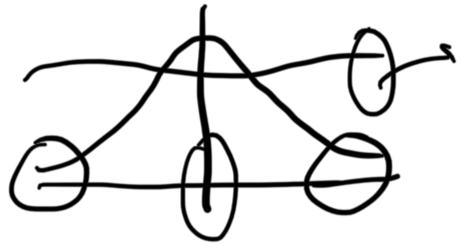
$$Var(c + X) = Var[X]$$

$$Var(c \cdot X) = c^2\, V[X]$$

- Normal Distribution

Applications → Height/wt. distrib
→ Returns stocks

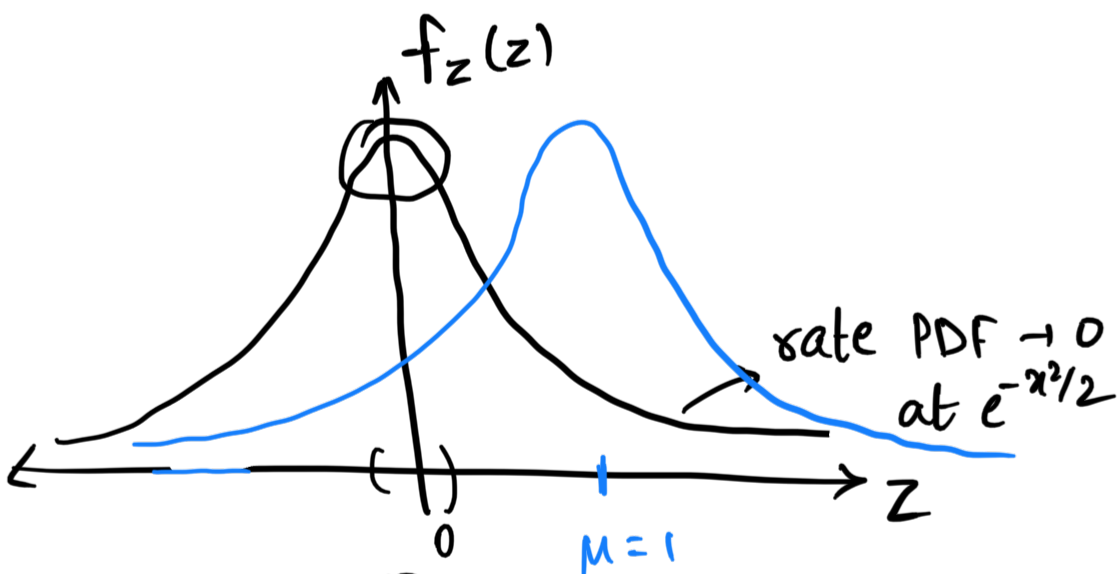
→ income



$Z \sim N(\underset{\text{mean}}{0}, \underset{\text{var}}{1})$ [standard Normal]

$$f_Z(z) = \underbrace{\frac{1}{\sqrt{2\pi}}}_{\text{Normalizing}} \cdot \underbrace{\exp\left(-\frac{z^2}{2}\right)}_{\text{Kernel}}$$

kernel $e^{-x^2/2}$ – symmetric



$$E[Z] = \int_{-\infty}^{\infty} z \cdot f_Z(z) \cdot dz = 0$$

$$V[Z] = E[Z^2] - (\cancel{E[Z]})^2 = E[Z^2] = 1$$

$$-Z \sim N(0,1) \,,\; Z \sim N(0,1)$$

$$\left\{ X = \underbrace{\mu}_{\text{location}} + \underbrace{\sigma}_{\text{scale}} Z \right\} \text{decomposition}$$

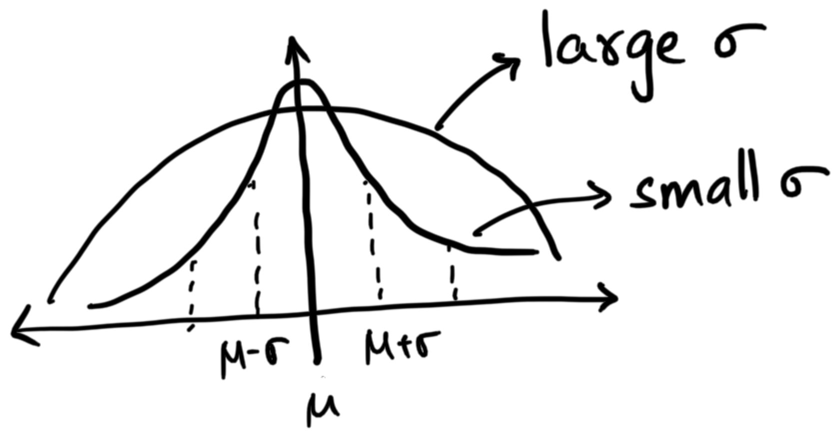
$$E[X] = E[\mu + \sigma Z] = \mu$$

$$V[X] = V[\sigma Z] = \sigma^2 \cdot V[Z] = \sigma^2$$

$$\left. Z = \underbrace{\frac{X-\mu}{\sigma}} \right\} \text{Z-statistic}$$

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \underbrace{\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)}_{\text{kernel}}$$

kernel is big when x is close to μ



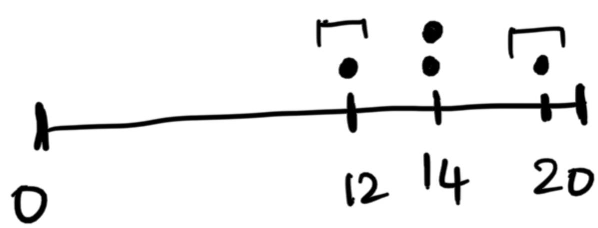
$$P(|X-\mu| < \sigma) \sim 0.68$$

$$P(|X-\mu| < 2\sigma) \sim 0.95$$

$$P(|X-\mu| < 3\sigma) \sim 0.997$$

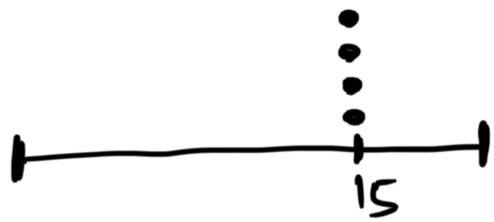
- Moments

  {Mean, median, SD, Quantile, skeness kurtosis}



$$\mu = \frac{12+14+14+20}{} = 15$$

$$\mu_1 = \frac{\quad}{4}$$



sq. dist. from 0

$$\left\{ \mu_2 = \frac{12^2 + 14^2 \times 2 + 20^2}{4} \sim \underbrace{234} \right.$$

$$\mu_2' = 15^2 = \underbrace{225}$$

Var. is sq. dist. from mean

	CRUDE	CENTERED
(1)	$\Sigma x / n$	$\Sigma x / n$
(2)	$\Sigma x^2 / n$	$\Sigma (x-\mu)^2 / n$
(3)	$\Sigma x^3 / n$	$\Sigma (x-\mu)^3 / \sigma^3 \cdot n$

(4) $\sum x^4/n$ $\sum (x-\mu)^4/\sigma^4 \cdot n$