Common Types of distribution

Normal (Gaussian) distn

$f(x) = \frac{1}{\sqrt{2\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma}}$

(1σ) : 68%
(2σ) : 95%
(3σ) : 99.7%

QQ Plots

to check if two datasets are from similar distribr (diagonal plot ⇒ similar distr)

to check Normality (normal distr and)

Quantile of the data

Z-score of data

$Z = \frac{x_i - \mu}{\sigma}$

Black Swan Theory :- Rare Events are more likely to occur Vs as predicted by normal distribution

Student's t distribution

tails size controlled by addn parameter ; υ : degrees of freedom

$\upsilon = 1$ Cauchy distribution (Lorentz) (longer tail)

$\upsilon = \infty$ (Normal distr) (short tail)

95% Confidence interval around Sample mean $= \bar{x} \pm t_{n-1}(0.05) \frac{S}{\sqrt{n}}$

Binomial distribution

$f_{n,p}(k) = \binom{n}{k} p^k q^{n-k}$

$(q = 1-p)$

prob of k success in n trials

$p(\text{success}) = p$

mean = np

std = npq

$(p \approx 1/2, n \to \infty)$ Binomial → Normal

Chi-square statistics

A statistics to measure how far different is the observed distribution from given distn

to check the null hypothesis of independence

Pearson χ^2 statistic : $\sum_{\text{data}} \frac{(O-E)^2}{E}$

O : observed frequency || E : expected frequency

for categorical variables

the χ^2 statistic follows the χ^2 distribution

Goodness of fit

It's actually distribution of $X_1^2 + X_2^2 + \dots X_k^2$ where X : Standard Normal distr

F-distribution

Similar to chi-square statistics but instead of dealing with counts dealing with Continuous variables

F-statistics ~ $\frac{\text{Variability among group mean}}{\text{Variability within each group}}$

Residual Variability

Poisson distribution

given an event occuring λ times per interval of time/space on avg

$f(k) = \frac{\lambda^k e^{-\lambda}}{k!}$ → prob of k events occurence in unit time.

mean → λ
Std → λ

* Can be used to measure request capacity measure

λ might not be const at smaller time frame but over sufficiently larger time frame

Exponential distr

time b/w two consecutive events

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

Weibull Distribution

time to failure when λ is increasing/decreasing

$f(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} \exp\left(-\left(\frac{t}{\eta}\right)^{\beta}\right)$

prob of failure at time t

$\eta : \frac{1}{\lambda}$: characteristic life

β → Shape parameter

> 1 : prob of failure increase over time

< 1 : prob of failure decrease over t.