

Statistical Tests

oooooooo

Distributions

oooooooooooo
oooooooooooooo
o

Descriptive Stats

ooooooo
oooooooooooo
o

Inferential Stats

oooooooo
oooooooooooo
o

Bayesian Stats

ooooo

Monte Carlo

ooooooo

Statistical Data Modeling

Saumya Bhatnagar

April 3, 2020

Statistical Tests

oooooo

Distributions

oooooooooooo
oooooooooooooo
o

Descriptive Stats

ooooooo
oooooooooooo
o

Inferential Stats

oooooooooooo
oooooooooooo
o

Bayesian Stats

ooooo

Monte Carlo

oooooooo

Table of contents I

Statistical Tests

Tests

Distributions

Distributions

Types of Dist

Max-Likelihood of Distributions

Descriptive Stats

Stats Flow

Descriptive Stats

Some more Terminologies

Inferential Stats

AB Testing

Hypothesis testing

Statistical Tests

oooooo

Distributions

oooooooooooo
oooooooooooooo
o

Descriptive Stats

ooooooo
oooooooooooo
o

Inferential Stats

oooooooooooo
oooooooooooo
o

Bayesian Stats

ooooo

Monte Carlo

oooooooo

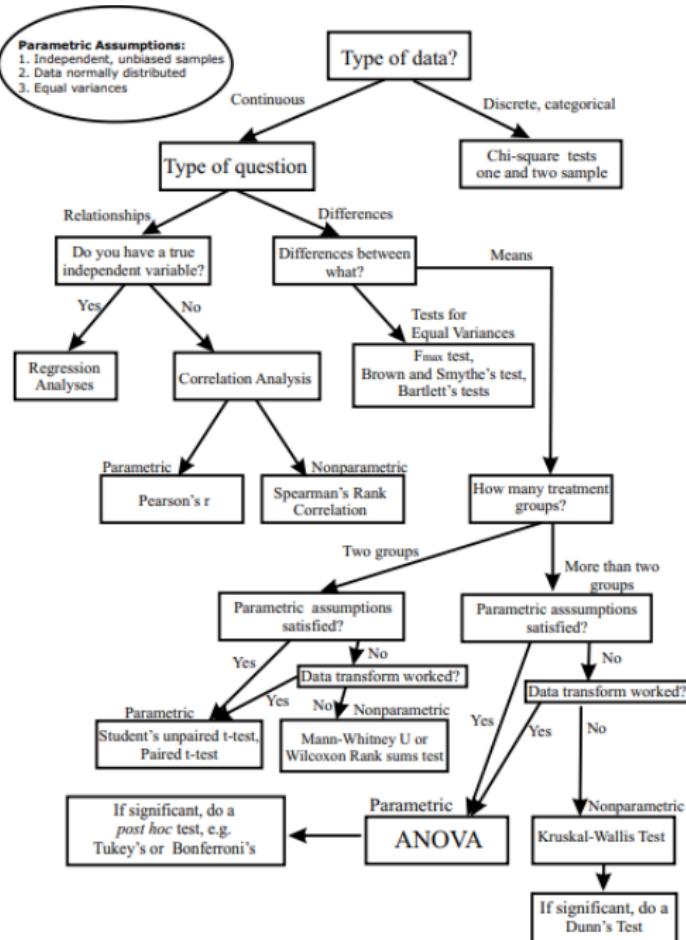
Table of contents II

something

Bayesian Stats

Bayesian Reasoning

Monte Carlo



1. $H_0 : \mu = 100$;
 $H_1 : \mu \neq 100$
2. if H_0 is true,
how extreme is
our sample?
- 3.
- 4.

Statistical Tests

○●○○○

Distributions

○○○○○○○○
○○○○○○○○○○
○

Descriptive Stats

○○○○○
○○○○○○○○

Inferential Stats

○○○○○○○
○○○○○○○○

Bayesian Stats

○○○○

Monte Carlo

○○○○○○

Tests

Precision Recall tradeoff

content...

Statistical Tests
○○●○○○

Distributions
○○○○○○○○
○○○○○○○○○○○○
○

Descriptive Stats
○○○○○
○○○○○○○○
○

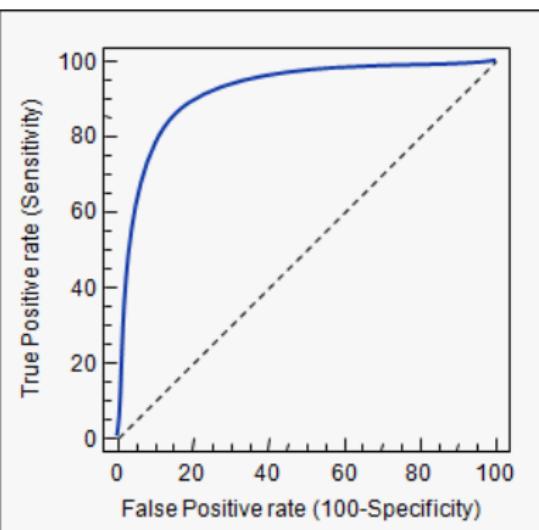
Inferential Stats
○○○○○○○
○○○○○○○○
○

Bayesian Stats
○○○○○

Monte Carlo
○○○○○○○

Tests

Receiver Operating Characteristic (ROC) curve and AUC



AUC values

AUC values	Test quality
0.9-1.0	Excellent
0.8-0.9	V good
0.7-0.8	Good
0.6-0.7	Satisfactory
0.5-0.6	Unsatisfactory

		CONDITION determined by "Gold Standard"			
TOTAL POPULATION		CONDITION POS	CONDITION NEG	PREVALENCE $\frac{\text{CONDITION POS}}{\text{TOTAL POPULATION}}$	
TEST OUT-COME	TEST POS	True Pos TP	Type I Error False Pos FP	Precision Pos Predictive Value PPV = $\frac{\text{TP}}{\text{TEST P}}$	False Discovery Rate FDR = $\frac{\text{FP}}{\text{TEST P}}$
	TEST NEG	Type II Error False Neg FN	True Neg TN	False Omission Rate FOR = $\frac{\text{FN}}{\text{TEST N}}$	Neg Predictive Value NPV = $\frac{\text{TN}}{\text{TEST N}}$
ACCURACY ACC $\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TOT POP}}$	Sensitivity (SN), Recall Total Pos Rate TPR = $\frac{\text{TP}}{\text{CONDITION POS}}$	Fall-Out False Pos Rate FPR = $\frac{\text{FP}}{\text{CONDITION NEG}}$	Pos Likelihood Ratio LR+ $\text{LR}+ = \frac{\text{TPR}}{\text{FPR}}$	Diagnostic Odds Ratio DOR $\text{DOR} = \frac{\text{LR}+}{\text{LR}-}$	
	Miss Rate False Neg Rate FNR = $\frac{\text{FN}}{\text{CONDITION POS}}$	Specificity (SPC) True Neg Rate TNR = $\frac{\text{TN}}{\text{CONDITION NEG}}$	Neg Likelihood Ratio LR- $\text{LR}- = \frac{\text{TNR}}{\text{FNR}}$		

Type I Error: error of first kind, Reject H_0 when H_0 is true. H_0 : usko bimari nahi hai, **incorrect rejection** of true null hyp

Prob of Type I Error = level of significance = $\alpha = 5\%$ (generally)

Type II error: error of second kind, Not reject H_0 when H_0 is false, **failure to reject** false null hyp

Prob of Type II Error = β and power of hypothesis test = $1-\beta$

$1-\beta$ = prob of rejecting a H_0 when H_0 is false

Statistical Tests

oooo●○

Distributions

oooooooo
oooooooooooo
○

Descriptive Stats

oooooo
oooooooooooo
○

Inferential Stats

ooooooo
oooooooooooo
○

Bayesian Stats

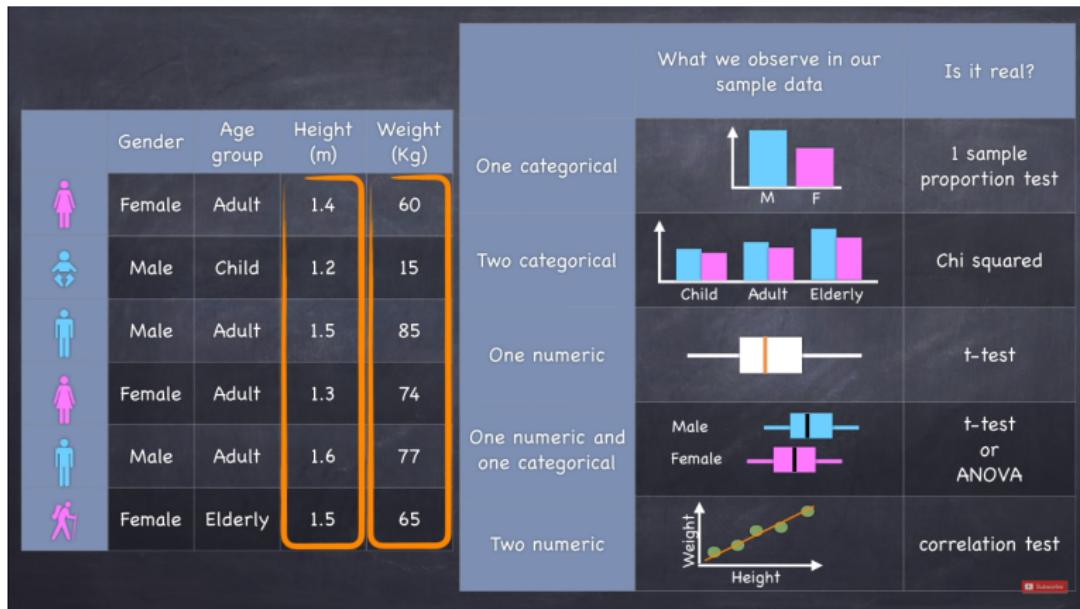
ooooo

Monte Carlo

ooooooo

Tests

test statistics



t-test, anova, chi-square, correlation test

Statistical Tests

oooooo●

Distributions

oooooooooooo
oooooooooooooo
o

Descriptive Stats

oooooooo
o

Inferential Stats

oooooooo
o

Bayesian Stats

ooooo

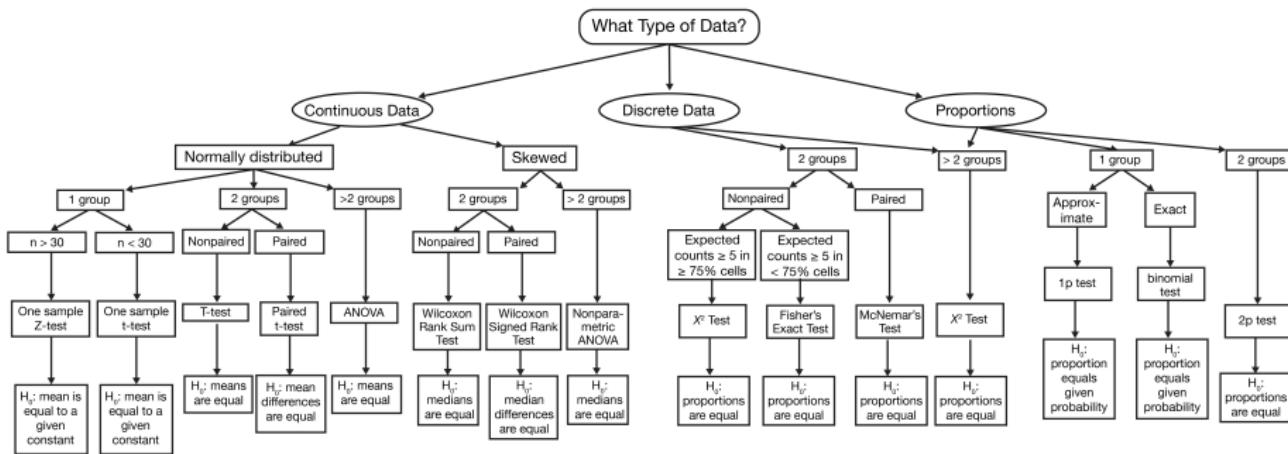
Monte Carlo

ooooooo

Tests

Type of tests

Flow chart: which test statistic should you use?



Statistical Tests
oooooo

Distributions
●oooooooo
○oooooooooooo
○

Descriptive Stats
○ooooo
○oooooooooooo
○

Inferential Stats
○oooooo
○oooooooooooo
○

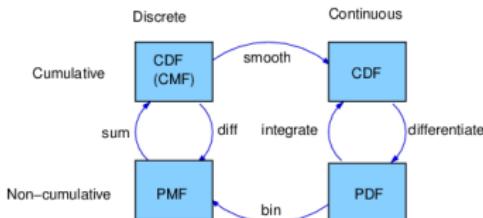
Bayesian Stats
ooooo

Monte Carlo
ooooooo

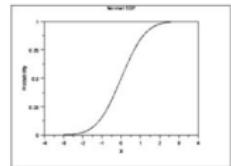
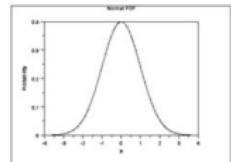
Distributions

PMF, CMF, PDF, CDF

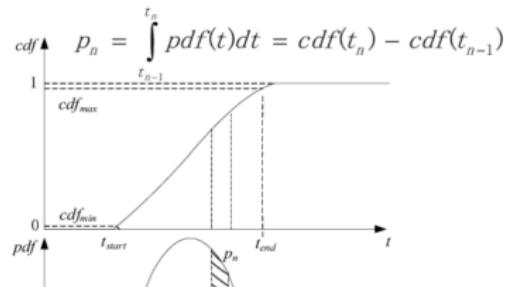
- ▶ A PMF, “f” returns the probability of an outcome:
 $f(x) = P(X = x)$



- Probability density function (p.d.f.) denote f
- Probability mass function (p.m.f.) denote f
 $f(x)=P(X=x)$
- Cumulative distribution function (c.d.f.) denote F
 $F(x)=P(X\leq x)$



- ▶ Reliability function & Hazard Function



Statistical Tests

oooooo

Distributions

●oooooooo
○oooooooooooo
○

Descriptive Stats

○ooooo
○oooooooooooo
○

Inferential Stats

○ooooooo
○oooooooo
○

Bayesian Stats

ooooo

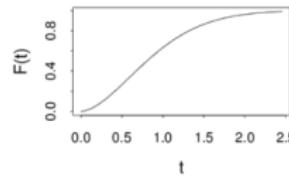
Monte Carlo

oooooooo

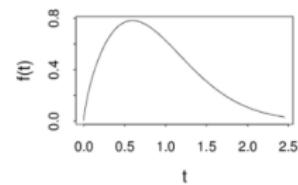
Distributions

PMF, CMF, PDF, CDF

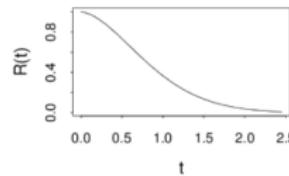
Cumulative Distribution Function



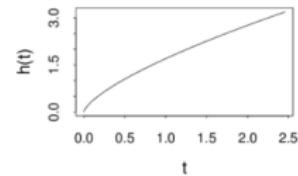
Probability Density Function



Reliability Function



Hazard Function



- Reliability function & Hazard Function

Statistical Tests
oooooo

Distributions
o●oooooooo
oooooooooooo
o

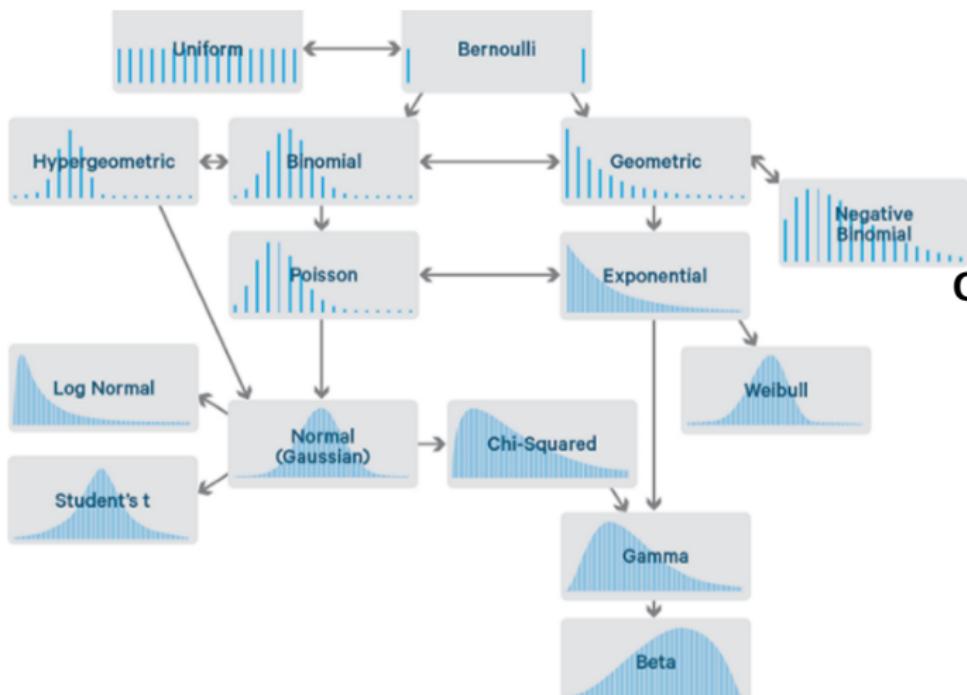
Descriptive Stats
oooooooo
o
o

Inferential Stats
oooooooo
oooooooo
o

Bayesian Stats
ooooo

Monte Carlo
ooooooo

Distributions



Discrete

1. Uniform Discrete or Rectangular
2. Binomial
3. Hypergeometric
4. Poisson
5. Geometric

Continuous

1. Uniform
2. Normal/Gaussian
3. Student's T
4. chi-squared
5. Exponential
6. Beta
7. Triangular
8. Gamma

Statistical Tests
○○○○○

Distributions
○○●○○○○○○
○○○○○○○○○○○○
○

Descriptive Stats
○○○○○○
○○○○○○○○○○
○

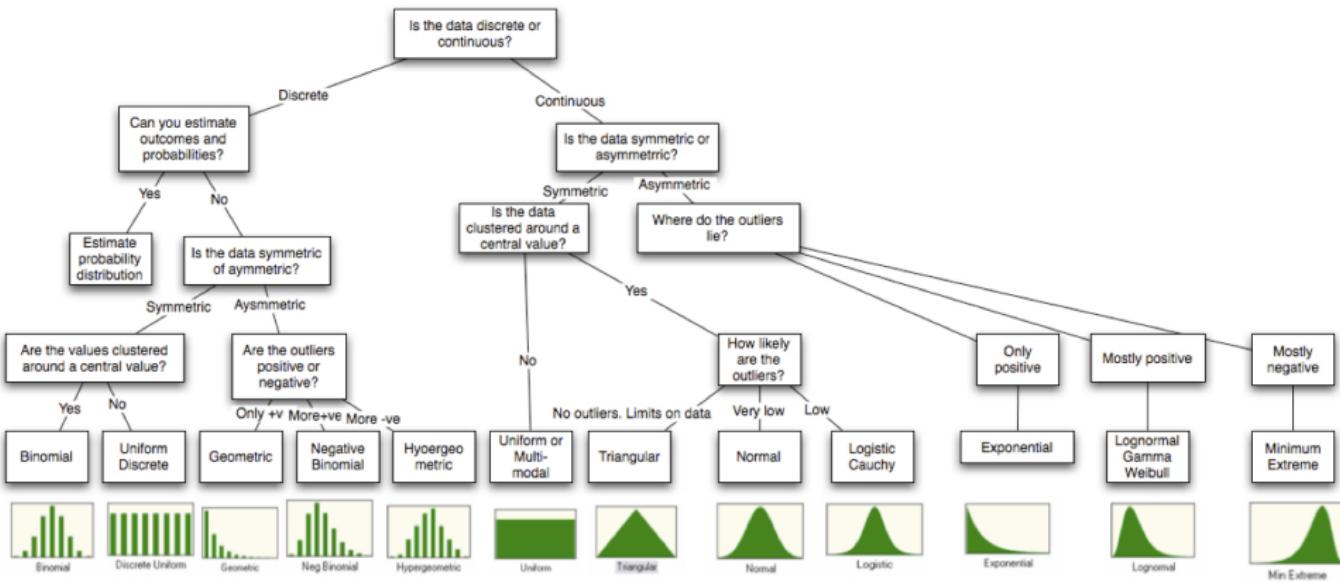
Inferential Stats
○○○○○○○○
○○○○○○○○
○

Bayesian Stats
○○○○

Monte Carlo
○○○○○○

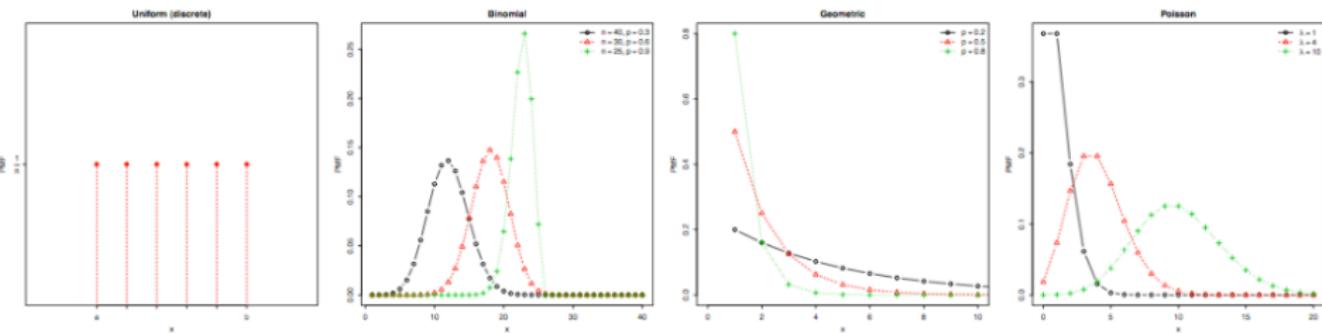
Distributions

Figure 6A.15: Distributional Choices



1.1 Discrete Distributions

		CDF/CMF	PMF	Expected Val of RV	Var of RV	
	Notation ¹	$F_X(x)$	$f_X(x)$	$\mathbb{E}[X]$	$\text{Var}[X]$	
Uniform	Unif $\{a, \dots, b\}$	$\begin{cases} 0 & x < a \\ \frac{ x - a + 1}{b - a} & a \leq x \leq b \\ 1 & x > b \end{cases}$	$\frac{I(a < x \leq b)}{b - a + 1}$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2 - 1}{12}$	$\frac{e^{ax} - e^{-(b+1)x}}{s(b-a)}$
Bernoulli	Bern (p)	$(1-p)^{1-x}$	$p^x (1-p)^{1-x}$	p	$p(1-p)$	$1-p + pe^s$
Binomial	Bin (n, p)	$I_{1-p}(n-x, x+1)$	$\binom{n}{x} p^x (1-p)^{n-x}$	np	$np(1-p)$	$(1-p + pe^s)^n$
Multinomial	Mult (n, p)		$\frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k} \quad \sum_{i=1}^k x_i = n$	np_i	$np_i(1-p_i)$	$\left(\sum_{i=0}^k p_i e^{s_i}\right)^n$
Hypergeometric	Hyp (N, m, n)	$\approx \Phi \left(\frac{x - np}{\sqrt{np(1-p)}} \right)$	$\frac{\binom{m}{x} \binom{m-x}{n-x}}{\binom{N}{x}}$	$\frac{nm}{N}$	$\frac{nm(N-n)(N-m)}{N^2(N-1)}$	
Negative Binomial	NBin (n, p)	$I_p(r, x+1)$	$\binom{x+r-1}{r-1} p^r (1-p)^x$	$r \frac{1-p}{p}$	$r \frac{1-p}{p^2}$	$\left(\frac{p}{1-(1-p)e^s} \right)^r$
Geometric	Geo (p)	$1 - (1-p)^x \quad x \in \mathbb{N}^+$	$p(1-p)^{x-1} \quad x \in \mathbb{N}^+$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{p}{1-(1-p)e^s}$
Poisson	Po (λ)	$e^{-\lambda} \sum_{i=0}^x \frac{\lambda^i}{i!}$	$\frac{\lambda^x e^{-\lambda}}{x!}$	λ	λ	$e^{\lambda(e^s-1)}$



Statistical Tests
○○○○○Distributions
○○○○●○○○○
○○○○○○○○○○○○
○Descriptive Stats
○○○○○○○
○○○○○○○○○○Inferential Stats
○○○○○○○○
○○○○○○○○Bayesian Stats
○○○○○Monte Carlo
○○○○○○○

Distributions

1.2 Continuous Distributions

	Notation	$F_X(x)$	$f_X(x)$	$\mathbb{E}[X]$	$\mathbb{V}[X]$	$M_X(s)$
Uniform	$\text{Unif}(a, b)$	$\begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x > b \end{cases}$	$\frac{I(a < x < b)}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{sb} - e^{sa}}{s(b-a)}$
Normal	$\mathcal{N}(\mu, \sigma^2)$	$\Phi(x) = \int_{-\infty}^x \phi(t) dt$	$\phi(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$	μ	σ^2	$\exp\left\{\mu s + \frac{\sigma^2 s^2}{2}\right\}$
Log-Normal	$\ln \mathcal{N}(\mu, \sigma^2)$	$\frac{1}{2} + \frac{1}{2} \operatorname{erf}\left[\frac{\ln x - \mu}{\sqrt{2\sigma^2}}\right]$	$\frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\ln x - \mu)^2}{2\sigma^2}\right\}$	$e^{\mu+\sigma^2/2}$	$(e^{\sigma^2} - 1)e^{2\mu+\sigma^2}$	
Multivariate Normal	$\text{MVN}(\mu, \Sigma)$		$(2\pi)^{-k/2} \Sigma ^{-1/2} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})}$	μ	Σ	$\exp\left\{\boldsymbol{s}^T \boldsymbol{\mu} + \frac{1}{2} \boldsymbol{s}^T \Sigma \boldsymbol{s}\right\}$
Student's t	$\text{Student}(\nu)$	$I_x\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$	$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}$	0	0	
Chi-square	χ_k^2	$\frac{1}{\Gamma(k/2)} \gamma\left(\frac{k}{2}, \frac{x}{2}\right)$	$\frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2} e^{-x/2}$	k	$2k$	$(1-2s)^{-k/2} s < 1/2$
F	$F(d_1, d_2)$	$I_{\frac{d_1 x}{d_1 x + d_2}}\left(\frac{d_1}{2}, \frac{d_1}{2}\right)$	$\frac{\sqrt{\frac{(d_1 x)^{d_1} d_2^{d_2}}{(d_1 x + d_2)^{d_1+d_2}}}}{x B\left(\frac{d_1}{2}, \frac{d_1}{2}\right)}$	$\frac{d_2}{d_2-2}$	$\frac{2d_2^2(d_1+d_2-2)}{d_1(d_2-2)^2(d_2-4)}$	
Exponential	$\text{Exp}(\beta)$	$1 - e^{-x/\beta}$	$\frac{1}{\beta} e^{-x/\beta}$	β	β^2	$\frac{1}{1-\beta s} (s < 1/\beta)$
Gamma	$\text{Gamma}(\alpha, \beta)$	$\frac{\gamma(\alpha, x/\beta)}{\Gamma(\alpha)}$	$\frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$	$\alpha\beta$	$\alpha\beta^2$	$\left(\frac{1}{1-\beta s}\right)^\alpha (s < 1/\beta)$
Inverse Gamma	$\text{InvGamma}(\alpha, \beta)$	$\frac{\Gamma(\alpha, \frac{x}{\beta})}{\Gamma(\alpha)}$	$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\beta/x}$	$\frac{\beta}{\alpha-1} \alpha > 1$	$\frac{\beta^2}{(\alpha-1)^2(\alpha-2)^2} \alpha > 2$	$\frac{2(-\beta s)^{\alpha/2}}{\Gamma(\alpha)} K_\alpha(\sqrt{-4\beta s})$
Dirichlet	$\text{Dir}(\alpha)$		$\frac{\Gamma\left(\sum_{i=1}^k \alpha_i\right)}{\prod_{i=1}^k \Gamma(\alpha_i)} \prod_{i=1}^k x_i^{\alpha_i-1}$	$\frac{\alpha_i}{\sum_{i=1}^k \alpha_i}$	$\frac{\mathbb{E}[X_i](1-\mathbb{E}[X_i])}{\sum_{i=1}^k \alpha_i + 1}$	
Beta	$\text{Beta}(\alpha, \beta)$	$I_x(\alpha, \beta)$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	$1 + \sum_{k=1}^{\infty} \left(\sum_{r=0}^{k-1} \frac{\alpha+r}{\alpha+\beta+r} \right) \frac{s^k}{k!}$

Statistical Tests
○○○○○

Distributions
○○○○●○○○
○○○○○○○○○○○
○

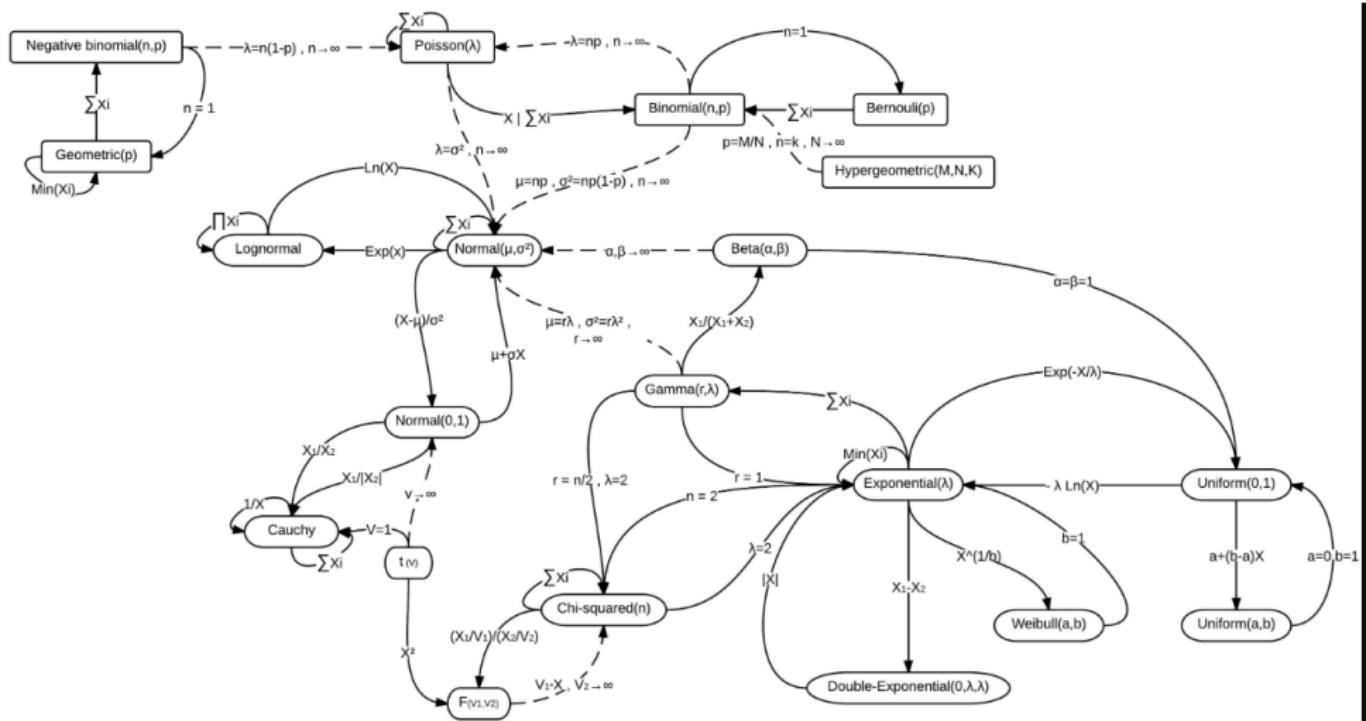
Descriptive Stats
○○○○○○
○
○

Inferential Stats
○○○○○○○
○○○○○○○
○

Bayesian Stats
○○○○

Monte Carlo
○○○○○○

Distributions



Statistical Tests
oooooo

Distributions
oooooooo●ooo
oooooooooooooo
o

Descriptive Stats
oooooooo
o

Inferential Stats
oooooooooooo
o

Bayesian Stats
ooooo

Monte Carlo
oooooooo

Distributions

relation btw various dist

Bernoulli and Binomial: Bernoulli Distribution is a special case of Binomial Distribution with a single trial.

Poisson and Binomial: Poisson Distribution is a limiting case of binomial distribution under the following conditions:

The number of trials is indefinitely large or $\lim_{x \rightarrow \infty}$. The probability of success for each trial is same and indefinitely small or $\lim_{x \rightarrow 0}$. $np = \lambda$, is finite.

Normal and Binomial: Normal distribution is another limiting form of binomial distribution under the following conditions:

The number of trials is indefinitely large, $\lim_{n \rightarrow \infty}$. Both p and q are not indefinitely small. **Normal and Poisson Distribution:** The normal distribution is also a limiting case of Poisson distribution with the parameter $\lim_{\lambda \rightarrow \infty}$.

Statistical Tests
oooooo

Distributions
oooooooo●o
oooooooooooo
o

Descriptive Stats
oooooo
oooooooooooo
o

Inferential Stats
oooooooo
oooooooooooo
o

Bayesian Stats
ooooo

Monte Carlo
ooooooo

Distributions

Beta and Binomial

Prior and Posterior

1. Conjugate prior for binomial

$$x|p \sim \text{Bin}(n, p); p \sim \text{Beta}(a, b) [\text{prior}]$$

$$f(p|X = k) = f(X = k|p)f(p)/f(X = k) \text{ [use bayes]}$$

replace with beta and bin

to learn more

$$p|X \sim \text{Beta}(a + X, b + n - x)$$

Statistical Tests
oooooo

Distributions
oooooooo●
oooooooooooo
o

Descriptive Stats
oooooo
oooooooooooo
o

Inferential Stats
oooooooo
oooooooooooo
o

Bayesian Stats
ooooo

Monte Carlo
ooooooo

Distributions

Poisson event per unit time
how many calls do you get in
a day

The number of printing errors
at each page of the book

Num of metro arrivals in t
time

The number of arrivals
reported in an area on a day.

The number of soldiers killed
by horse-kick per year

Air conditioners in a lifetime

exponential time per event

What about the interval of
time btw the calls

Num of pages before until x
num of printing errors

Length of time btw metro
arrivals,

Length of time between
arrivals at a gas station

Num of years btw horse-kick
deaths in the army

The life of an Air Conditioner

Statistical Tests
○○○○○

Distributions
○○○○○○○○○○
●○○○○○○○○○○
○

Descriptive Stats
○○○○○
○
○

Inferential Stats
○○○○○○○
○○○○○○○
○

Bayesian Stats
○○○○

Monte Carlo
○○○○○○

Types of Dist

Standard Uniform Density

parameters $a = 0$ and $b = 1$, so the PDF for standard uniform density is given by:

$$f(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Statistical Tests
oooooo

Distributions
oooooooooooo
o●oooooooooooo
o

Descriptive Stats
oooooooo
o
o

Inferential Stats
oooooooooooo
oooooooooooo
o

Bayesian Stats
ooooo

Monte Carlo
oooooooo

Types of Dist

Normal Distribution

Standard Normal Distribution $\mu = 0 ; \sigma = 1$

The 68-95-99.7 rule: Given a normally distributed random variable: $P(\mu - \sigma \leq X \leq \mu + \sigma) \approx .68 \Rightarrow 68\%$ of samples fall within 1 SD of the mean

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx .95$$

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx .997$$

characteristics of Normal distribution:

1. Mean = median = mode
2. The distribution curve is bell-shaped and symmetrical about the line $x=\mu$.
3. The total AUC = 1.
4. Exactly half of the values are to the left of the center and the other half to the right.

Statistical Tests
oooooo

Distributions
oooooooooooo
o●oooooooooooo
o

Descriptive Stats
oooooooooooo
o

Inferential Stats
oooooooooooo
o

Bayesian Stats
oooooo

Monte Carlo
oooooooo

Types of Dist

Student's t distribution

characteristics

1. Underlying dist is Normal
2. Pop dist is unknown
3. sample size is too small for CLT to apply

$$z \sim \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \longleftrightarrow t_{n-1} \sim \frac{\bar{x} - \mu}{s / \sqrt{n}} \text{ t test measures}$$

1. to test hypothesized population mean $\frac{\bar{x} - \mu}{s / \sqrt{n}}$
2. regression $\frac{b - \beta}{SE(b)}$ (uses Std error, as pop std dev is known)
3. 2 sampled t-test: assessing the diff btw 2 pop $\frac{(\bar{x}_1 - \mu_1) - (\bar{x}_2 - \mu_2)}{\sqrt{\frac{s_1^2}{\sqrt{n_1}} + \frac{s_2^2}{\sqrt{n_2}}}}$

Statistical Tests
○○○○○

Distributions
○○○○○○○○○○
○○●○○○○○○○○
○

Descriptive Stats
○○○○○○
○
○

Inferential Stats
○○○○○○○○
○○○○○○○○
○

Bayesian Stats
○○○○

Monte Carlo
○○○○○○○

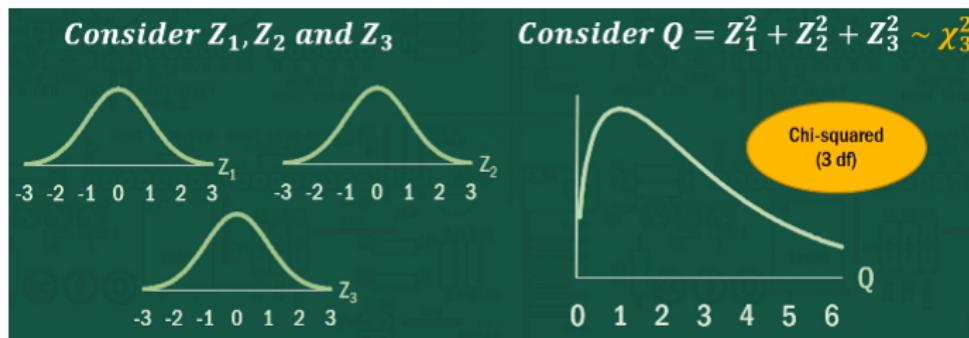
Types of Dist

Chi-squared dist

comes directly from a normal dist (square of selection from standard Normal Distribution) so sample size should be large enough (>5) s.t. CLT applies

for k degrees of freedom: $\chi_k^2 = \sum_{i=1}^k Z_i^2$

$\chi^2 = \sum \frac{(obs-exp)^2}{exp}$ with DF = (row-1)(col-1)



Statistical Tests
oooooo

Distributions
oooooooooooo
oooo●oooooooo
o

Descriptive Stats
ooooooo
oooooooooooo
o

Inferential Stats
oooooooooooo
oooooooooooo
o

Bayesian Stats
ooooo

Monte Carlo
oooooooo

Types of Dist

Binomial, bernoulli, hyper-geometric

Repeat Bernoulli n times and it's Binomial.

Hypergeometric is Binomial without replacement
the properties of a Binomial Distribution are

1. Each trial is independent.
2. There are only two possible outcomes per trial.
3. A total number of n identical trials are conducted.
4. The probability of success and failure is same for all trials.
(Trials are identical.)

Statistical Tests
oooooo

Distributions
oooooooooooo
oooooooo●oooooooo
o

Descriptive Stats
oooooooo
o

Inferential Stats
oooooooooooo
oooooooooooo
o

Bayesian Stats
oooooo

Monte Carlo
oooooooo

Types of Dist

Hypergeometric

1. Discrete
2. equivalent to Binomial, without replacement
3. N = total population
 m =total items of interest in population
 n =sample size
4. region bounded by 0 and m

Characteristics of Poisson distribution:

1. Event are not Independent.
- 2.

Statistical Tests
oooooo

Distributions
oooooooooooo
oooooooo●oooo
o

Descriptive Stats
oooooooo
o

Inferential Stats
oooooooooooo
oooooooooooo
o

Bayesian Stats
ooooo

Monte Carlo
oooooooo

Types of Dist

Poisson

1. Discrete
2. events in fixed region of opportunity (or time interval, t)
3. region bounded by 0 and ∞

Characteristics of Poisson distribution:

1. Independent event.
2. The probability of success over a short interval must equal the probability of success over a longer interval.
3. The probability of success in an interval approaches zero as the interval becomes smaller.
4. The rate at which event occurs is constant (λ)

Poisson RV, X = number of events in t .

mean number of events in t , $\mu = \lambda * t$

The PMF of X : $P(X = x) = e^{-\mu} * \frac{\mu^x}{x!}$

Statistical Tests
oooooo

Distributions
oooooooooooo
oooooooo●oooo
o

Descriptive Stats
oooooooo
o

Inferential Stats
oooooooooooo
oooooooooooo
o

Bayesian Stats
ooooo

Monte Carlo
oooooooo

Types of Dist

Exponential Dist

inverse of Poisson: rate parameter or mean for poisson = λ and
mean for expo = $\beta = 1/\lambda$

Exponential distribution is widely used for survival analysis.

Memoryless-ness:

events must occur at constant rate

events must be independent of each other

probab of event occurring in first min = probab of the event
occurring in $(t+1)$ min

probab of first visitor on website in first min = p

probab of first visitor on website in second min = $(1-p)p$

probab of first visitor on website within third min = $(1-p)^2p$

Each minute graph dropping → exponential decay

Statistical Tests

oooooo

Distributions

Descriptive Stats

Inferential Stats

Bayesian Stats

ooooo

Monte Carlo

ooooooo

Types of Dist

Memorylessness

Memoryless property: $P(X \geq s + t | X \geq s) = P(X \geq t)$

$$P(X \geq s) = 1 - CDF = 1 - P(X \leq s) = e^{-\lambda s}$$

$$P(X \geq s + t | X \geq s) = \frac{P(X \geq s+t, X \geq s)}{P(X \geq s)}$$

$$P(X \geq s + t | X \geq s) = \frac{P(X \geq s+t)}{P(X \geq s)}$$

$$P(X \geq s + t | X \geq s) = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}}$$

$$P(X \geq s + t | X \geq s) = e^{-\lambda t}$$

Memoryless property:

$$E(X | X > a) = a + E(X - a | X > a)$$

$$E(X | X > a) = a + 1/\lambda$$

Statistical Tests

oooooo

Distributions

oooooooooooo
oooooooooooo●o
○

Descriptive Stats

oooooooo
○
○

Inferential Stats

oooooooo○
oooooooo○
○

Bayesian Stats

ooooo

Monte Carlo

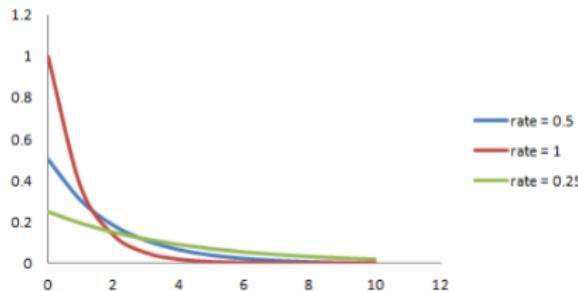
oooooooo

Types of Dist

failure rate of any device at time t , given that it has survived up to t ; $\lambda = \frac{1}{\beta} > 0$

area under the density curve

Exponential Distribution



$$\text{to the left of } x \quad P\{X \leq x\} = 1 - e^{-\lambda x}$$

$$\text{to the right of } x \quad P\{X > x\} = e^{-\lambda x}$$

$$P\{x_1 < X \leq x_2\} = e^{-\lambda x_1} - e^{-\lambda x_2}$$

Statistical Tests
oooooo

Distributions
oooooooooooo
oooooooooooo●
○

Descriptive Stats
oooooooo
○
○

Inferential Stats
oooooooo
○
○

Bayesian Stats
ooooo

Monte Carlo
oooooooo

Types of Dist

to read

<https://www.quora.com/What-is-the-relation-between-standard-normal-and-gamma-distribution>

<https://stats.stackexchange.com/questions/37461/the-relationship-between-the-gamma-distribution-and-the-normal-distribution>

Statistical Tests
○○○○○○

Distributions
○○○○○○○○○○
●

Descriptive Stats
○○○○○
○○○○○○○○
○

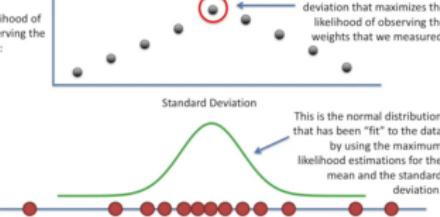
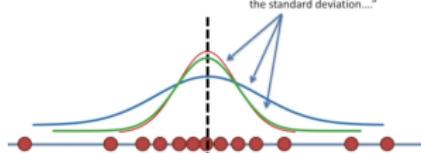
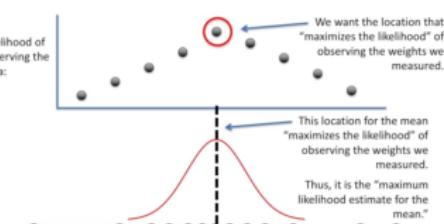
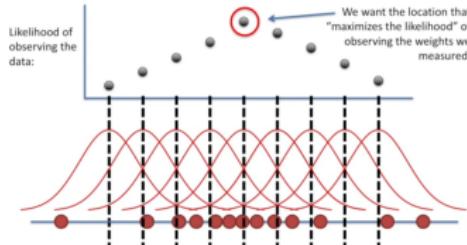
Inferential Stats
○○○○○○
○○○○○○○
○

Bayesian Stats
○○○○○

Monte Carlo
○○○○○○

Max-Likelihood of Distributions

Max Likelihood



Now when someone says that they have the maximum likelihood estimates for the mean or the standard deviation, or for something else...
... you know that they found the value for the mean or the standard deviation (or whatever) that maximizes the likelihood that you observed the things you observed.

Statistical Tests
oooooo

Distributions
oooooooooooo
oooooooooooo
o

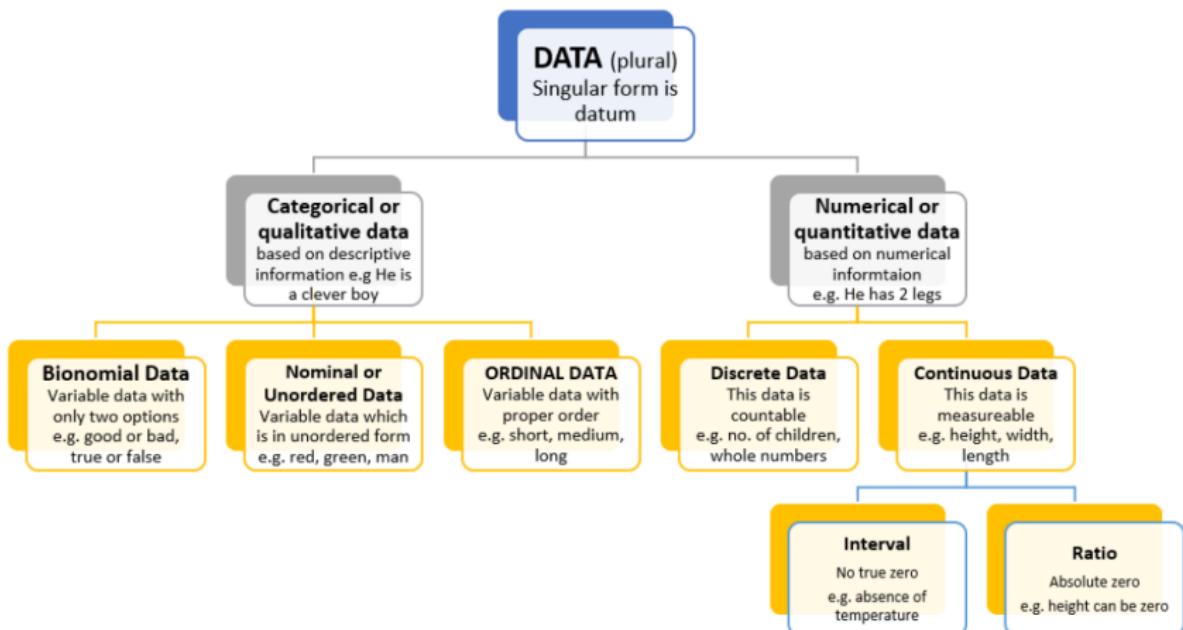
Descriptive Stats
●oooooo
oooooooooooo
o

Inferential Stats
oooooooooooo
oooooooooooo
o

Bayesian Stats
ooooo

Monte Carlo
oooooooo

Stats Flow



Statistical Tests
○○○○○○

Distributions
○○○○○○○○
○○○○○○○○○○
○

Descriptive Stats
○●○○○○
○○○○○○○○

Inferential Stats
○○○○○○○
○○○○○○○

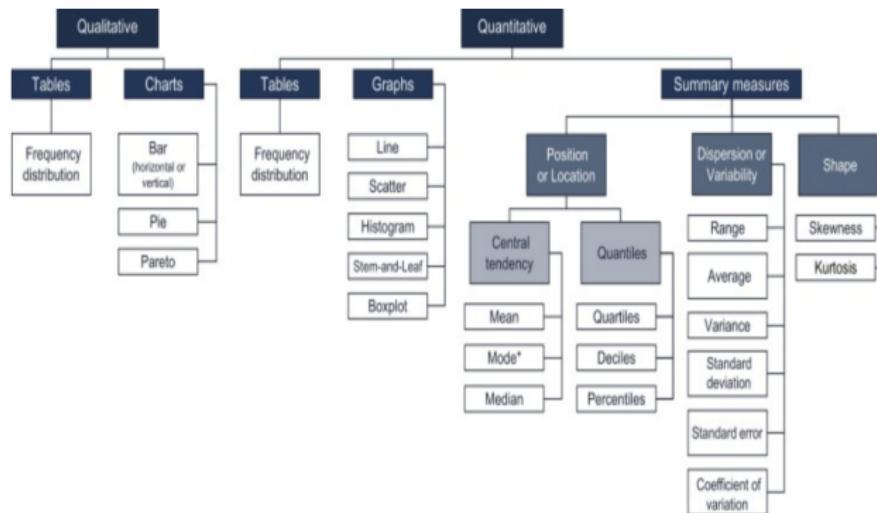
Bayesian Stats
○○○○○

Monte Carlo
○○○○○○

Stats Flow

Types of Analysis

- ▶ Qualitative Analysis/Non-Statistical Analysis gives generic information (uses text, sound and other forms of media).
- ▶ Quantitative Analysis/Statistical Analysis: collecting and interpreting data.



Statistical Tests
○○○○○○

Distributions
○○○○○○○○
○○○○○○○○○○
○

Descriptive Stats
○○●○○○
○○○○○○○○
○

Inferential Stats
○○○○○○
○○○○○○○
○

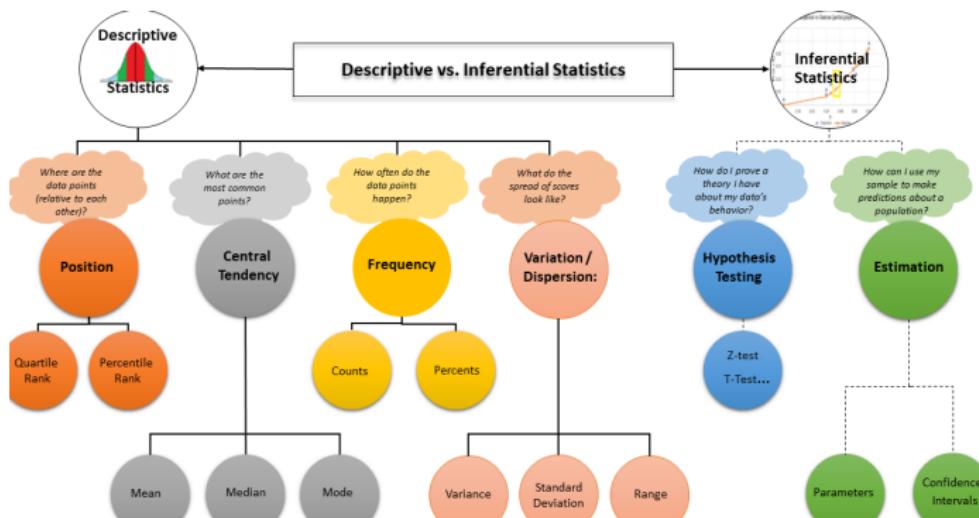
Bayesian Stats
○○○○○

Monte Carlo
○○○○○○

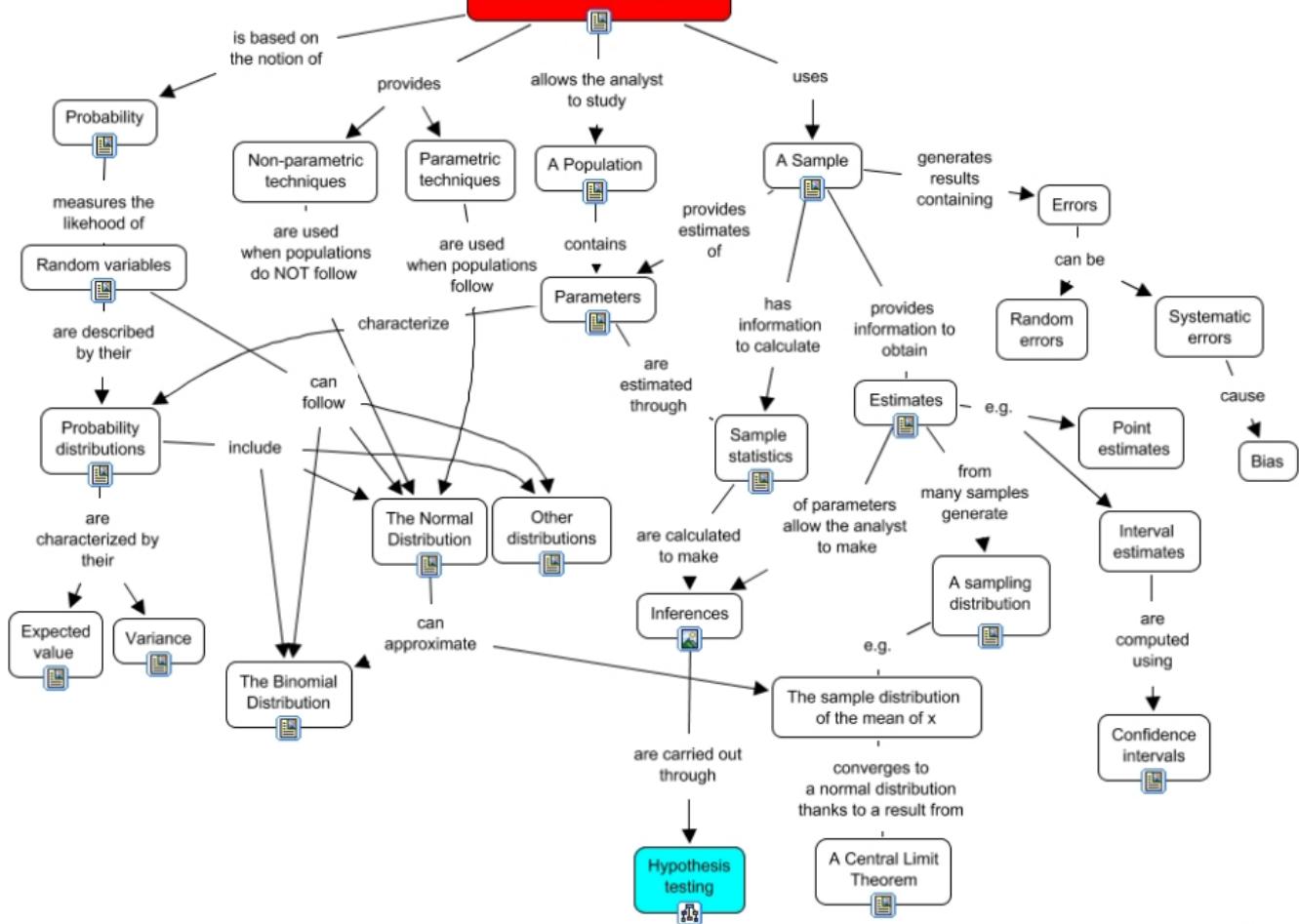
Stats Flow

Types of Statistics

- ▶ Descriptive Statistics: provides descriptions of the population.
- ▶ Inferential Statistics makes inferences and predictions from sample to generalize a population.



Inferential statistics



Statistical Tests
oooooo

Distributions
oooooooooooo
oooooooooooo
o

Descriptive Stats
oooo●o
oooooooooooo
o

Inferential Stats
oooooooooooo
oooooooooooo
o

Bayesian Stats
ooooo

Monte Carlo
oooooooo

Stats Flow

Contingency Table and Probabilities

Joint, Marginal and Conditional

- ▶ Joint probabilities for rain and wind:

	no wind	some wind	strong wind	storm
no rain	0.1	0.2	0.05	0.01
light rain	0.05	0.1	0.15	0.04
heavy rain	0.05	0.1	0.1	0.05

- ▶ Marginalize to get simple probabilities:
 - ▶ $P(\text{no wind}) = 0.1 + 0.05 + 0.05 = 0.2$
 - ▶ $P(\text{light rain}) = 0.05 + 0.1 + 0.15 + 0.04 = 0.34$
- ▶ Combine to get conditional probabilities:
 - ▶ $P(\text{no wind}|\text{light rain}) = \frac{0.05}{0.34} = 0.147$
 - ▶ $P(\text{light rain}|\text{no wind}) = \frac{0.05}{0.2} = 0.25$

Statistical Tests
oooooo

Distributions
oooooooooooo
oooooooooooo
o

Descriptive Stats
oooooo●
oooooooooooo
o

Inferential Stats
oooooooooooo
oooooooooooo
o

Bayesian Stats
ooooo

Monte Carlo
oooooooo

Stats Flow

Central Limit Theorem(CLT)

CLT: as $n \uparrow$, the distribution of sample mean or sum approaches a Normal Dist

Law of large num: as a sample size grows, its mean gets closer to the average of the whole population.

(LLN) is a theorem that describes the result of performing the same experiment a large number of times

In a financial context, the law of large numbers indicates that a large entity which is growing rapidly cannot maintain that growth pace forever.

law of averages: the supposed principle that future events are likely to turn out so that they balance any past deviation from a presumed average.

The law of averages is a lay term used to express a belief that outcomes of a random event will “even out” within a small sample

Statistical Tests
oooooo

Distributions
oooooooooooo
oooooooooooo
o

Descriptive Stats
ooooooo
●ooooooo
o

Inferential Stats
ooooooo
ooooooo
o

Bayesian Stats
ooooo

Monte Carlo
ooooooo

Descriptive Stats

Variance, Standard Deviation

- ▶ Variable and Random Variable (RV)
- ▶ Parameter and Hyper-parameter
- ▶ Mean, Median, Mode
- ▶ mode sucks for small samples
- ▶ Range, IQR
- ▶ Standard Deviation (σ): Measure of how spread out the data is from its mean.
- ▶ Variance (σ^2): It describes how much a random variable differs from its expected value. It entails computing squares of deviations. The average of the squared differences from the Mean.
 1. Deviation is the difference bw each element from the mean.
 2. Population Variance = avg of squared deviations
 3. Sample Variance = avg of squared differences from the mean

Statistical Tests
oooooo

Distributions
oooooooooooo
oooooooooooo
o

Descriptive Stats
ooooooo
ooooooo
o

Inferential Stats
ooooooo
ooooooo
o

Bayesian Stats
ooooo

Monte Carlo
ooooooo

Descriptive Stats

EXPECTED VALUE

Discrete random variable $E(X) = \sum_x x p_x(x)$

- ▶ Provided $\sum_x |x| p_x(x) < \infty$. If the sum diverges, the expected value does not exist. **For the jar full of numbered balls**
- ▶ A ball is selected at random; all balls are equally likely to be chosen $P(X = x_i) = \frac{1}{N}$.
- ▶ Say n_1 balls have value v_1 , and n_2 balls have value v_2 , and ... n_n balls have value v_n . Unique values are v_i , for $i = 1, \dots, n$. Note $n_1 + \dots + n_n = N$, and $P(X = v_j) = \frac{n_j}{N}$.
$$E(X) = \frac{\sum_{i=1}^N x_i}{N}$$

Continuous random variable $E(X) = \int_{-\infty}^{\infty} x f_x(x) dx$

- ▶ Provided $\int_{-\infty}^{\infty} |x| f_x(x) dx < \infty$. If the integral diverges, the expected value does not exist.

Statistical Tests

Distributions



Descriptive Stats
○○○○○○
○○●○○○○○○
○

Inferential Stats



Letter	Count
o	1
n	1
t	1
r	1
s	1
e	1
l	1
d	1
c	1
m	1
h	1
g	1
f	1
b	1
v	1
w	1
y	1
x	1
z	1
i	0

Bayesian Stats
00000

Monte Carlo

Descriptive Stats

Sometimes the expected value does not exist

Need $\int_{-\infty}^{\infty} |x| f_x(x) dx < \infty$

For the Cauchy distribution, $f(x) = \frac{1}{\pi(1+x^2)}$.

$$\begin{aligned}
 E(|X|) &= \int_{-\infty}^{\infty} |x| \frac{1}{\pi(1+x^2)} dx \\
 &= 2 \int_0^{\infty} \frac{x}{\pi(1+x^2)} dx \\
 u = 1 + x^2, \quad du = 2x \, dx \\
 &= \frac{1}{\pi} \int_1^{\infty} \frac{1}{u} du \\
 &= \ln u \Big|_1^{\infty} \\
 &= \infty - 0 = \infty
 \end{aligned}$$

\Rightarrow an integral “equals” infinity, it is unbounded above.

Statistical Tests

oooooo

Distributions







Descriptive Stats

oooooo





Inferential Stats

oooooo





Bayesian Stats

ooooo

Monte Carlo

ooooooo

Descriptive Stats

For a RV X with PDF $\rho(x)$. The variance(\mathbb{V}) and the standard deviation(σ_X) of X , are defined by

$$\text{Variance } \sigma^2 = (1/n) \sum_{i=1}^n (x_i - \mu)^2$$

$$\mathbb{V} = \mathbb{E} [(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \int_D (x - \mathbb{E})^2 dP.$$

$$\mathbb{V} = \int_D x^2 dP - \mathbb{E}^2.$$

$$\sigma_X = \sqrt{\mathbb{V}[X]} \quad = \sqrt{\mathbb{E}[X^2] - \mathbb{E}[X]^2}$$

$$\mathbb{V} = \sqrt{\int_D x^2 \rho(x) dx - \left(\int_D x \rho(x) dx \right)^2}.$$

Statistical Tests

oooooo

Distributions

oooooooooooo
oooooooooooooo
o

Descriptive Stats

ooooooo
oooo●oooo
o

Inferential Stats

ooooooo
ooooooo
o

Bayesian Stats

ooooo

Monte Carlo

ooooooo

Descriptive Stats

If one interprets the PDF ($\rho(x)$) as the density of a rod at location (x), then:

The mean, ($\mu = \int x\rho(x) dx$), gives the center of mass of the rod.

The variance, ($V = \int (x - \mu)^2 \rho(x) dx$), gives the moment of inertia about the line ($x = \mu$).

The standard deviation, ($\sigma = \sqrt{V}$), gives the radius of gyration about the line ($x = \mu$).

Statistical Tests
oooooo

Distributions
oooooooooooo
oooooooooooo
o

Descriptive Stats
ooooooo
ooooooo●ooo
o

Inferential Stats
ooooooo
ooooooo
o

Bayesian Stats
ooooo

Monte Carlo
ooooooo

Descriptive Stats

Std error vs std deviation

std error = std dev of the means

std dev quantifies the variation within a set of measurements

std error quantifies variation in the means from multiple sets of measurements

take a sample and get the mean and std dev

take multiple sets of samples and get their means and std dev

plot the means of the various samples

std dev of this plot of means is std error

Statistical Tests

Distributions



Descriptive Stats



Inferential Stats



The figure shows a scatter plot with 10 data points arranged in a bell-shaped curve. The points are represented by small circles. The distribution is roughly symmetric, with most points clustered in the center and fewer points at the extremes.

Bayesian Stats
ooooo

Monte Carlo
oooooooo

Descriptive Stats

coeff of variation

$CV = \frac{sd}{\bar{x}}$, where \bar{x} = sample mean

$$x = [1, 2, 3] \Rightarrow \bar{x} = 2 \text{ and } S_x = 1 \Rightarrow CV(x) = 1/2$$

$$y = [101, 102, 103] \Rightarrow \bar{y} = 102 \text{ and } S_y = 1 \Rightarrow CV(y) = 1/102$$

Higher the CV means higher fluctuations in the dataset

Statistical Tests
oooooo

Distributions
oooooooooooo
oooooooooooo
o

Descriptive Stats
ooooooo
ooooooo●o
o

Inferential Stats
ooooooo
ooooooo
o

Bayesian Stats
ooooo

Monte Carlo
ooooooo

Descriptive Stats

skewness and kurtosis

<https://www.thoughtco.com/what-is-kurtosis-3126241>

skewness

$$\text{mode skewness} = \frac{\text{mean} - \text{mode}}{\text{stddev}}$$

for small dataset, use below:

in skewed data: mode = 3(median) - 2(mean)

$$\text{median skewness} = \frac{3(\text{mean} - \text{median})}{\text{stddev}}$$

$$\text{skewness} = \begin{cases} \text{approx_symmetric}, & -0.5 \leq x \leq 0.5 \\ \text{moderately_skewed}, & 0.5 < |x| < 1 \\ \text{highly_skewed}, & |x| > 1 \end{cases} \quad (2)$$

kurtosis: same mean or sd but diff peakedness

higher peaked => higher kurtosis

Statistical Tests
oooooo

Distributions
oooooooooooo
oooooooooooo
o

Descriptive Stats
ooooooo
ooooooo●
o

Inferential Stats
ooooooo
ooooooo
o

Bayesian Stats
ooooo

Monte Carlo
ooooooo

Descriptive Stats

Moments

I moment: $\frac{\sum x}{n} \Rightarrow \text{mean}$ \Rightarrow considered as values from 0

second moment: $\frac{\sum x^2}{n} \Rightarrow$ values further from 0 will be higher,
so instead we take centralized

second (centralized) moment: $\frac{\sum(x-\mu)^2}{n} \Rightarrow$ variance

third (centralized) moment: $\frac{1}{n} \frac{\sum(x-\mu)^3}{\sigma^3} \Rightarrow$ skew

but since we don't have population mean, we have sample mean,
we adjust the above value with degrees of freedom

II (centralized) moment: $\frac{\sum(x-\bar{x})^2}{n-1} \Rightarrow$ variance

III (centralized) moment: $\frac{n}{(n-1)(n-2)} \frac{\sum(x-\bar{x})^3}{s^3} \Rightarrow$ skew

IV moment: $\frac{n(n+1)}{(n-1)(n-2)(n-3)} \frac{\sum(x-\bar{x})^4}{s^4} - \frac{3(n-1)^2}{(n-2)(n-3)} \Rightarrow$ kurtosis

Statistical Tests
oooooo

Distributions
oooooooooooo
oooooooooooo
o

Descriptive Stats
oooooooooooo
oooooooooooo
●

Inferential Stats
oooooooooooo
oooooooooooo
o

Bayesian Stats
oooooo

Monte Carlo
oooooooooooo

Some more Terminologies

confounding variables

↑ variance, can introduce bias

control variable vs control group

control group → if difficult to identify or control all potential confounding variables

positive confounding: overestimate the effect

negative confounding:

- ▶ underestimate the effect
- ▶ observed association is biased towards NULL

To reduce impact of confounding variables, bias can be eliminated with random samples

Statistical Tests
oooooo

Distributions
oooooooooooo
oooooooooooo
o

Descriptive Stats
ooooooo
oooooooooooo
o

Inferential Stats
●oooooooo
oooooooooooo
o

Bayesian Stats
ooooo

Monte Carlo
oooooooo

AB Testing

Real vs Empirical Conversion

$$1/2 =? 10/20 =? 100/200 =? 1000/2000$$

same empirical conversion rate but different real conv rate because of uncertainty

A = Control group; B = Test group

Statistical Tests
oooooo

Distributions
oooooooooooo
oooooooooooo
o

Descriptive Stats
ooooooo
oooooooooooo
o

Inferential Stats
o●oooooo
oooooooooooo
o

Bayesian Stats
ooooo

Monte Carlo
ooooooo

AB Testing

Types of AB Testing I

Split Testing: divide the traffic to the two new versions

A/A testing (dummy experiment): split traffic to same version, still if difference, that would imply the way users are assigned is impacting the results

Hypothesis testing - Frequentist Approach: Uses χ^2 test?

- ▶ p-value
- ▶ confidence intervals
- ▶ needs a fixed sample size in advance
 1. statistical power (how often will you recognize a successful effect): typically 80%
 2. significance level (how often will you observe positive result, however, there's none): typically 5%

Statistical Tests
oooooo

Distributions
oooooooooooo
oooooooooooo
o

Descriptive Stats
ooooooo
oooooooooooo
o

Inferential Stats
●●oooooo
oooooooooooo
o

Bayesian Stats
ooooo

Monte Carlo
ooooooo

AB Testing

Types of AB Testing II

Bayesian AB testing

	Hypothesis Testing	Bayesian A/B Testing
Knowledge of Baseline Performance	Required	Not Required
Intuitiveness	Less, as p-value is a convoluted term	More, as we directly calculate the probability of A being better than B
Sample size	Pre-defined	No need to pre-define
Peeking at the data while the test runs	Not allowed	Allowed (with caution)
Quick to make decisions	Less, as it has more restrictive assumptions on distributions	More, as it has less restrictive assumptions
Representing uncertainty	Confidence Interval (again, a convoluted interpretation which is often misunderstood)	Highest Posterior Density Region – highly intuitive interpretation
Declaring a winner	When sample size is reached and p-value is below a certain threshold	When either "probability to be best" goes above a threshold or the expected loss is below a threshold (in which case a "tie" can be declared between multiple variations)

Statistical Tests
oooooo

Distributions
oooooooooooo
oooooooooooo
o

Descriptive Stats
ooooooo
oooooooooooo
o

Inferential Stats
oo●oooo
oooooooooooo
o

Bayesian Stats
ooooo

Monte Carlo
oooooooo

AB Testing

When can AB test fail

1. in the case of a **referral program**, The referrer and Referee could be split across test and control groups causing spillover on the control or variant group
2. Novelty effects: Prompts and CTA tend to exhibit novelty effects, if not measuring their performance over the long term using a holdout a wrong attribution and/or customer fatigue can happen.
3. What-if scenarios: If you are looking to understand the impact of **not having launched a product**, for instance a subscription offering on a website. A/B test wouldn't be the right fit.

Statistical Tests

oooooo

Distributions

oooooooooooo
oooooooooooo
o

Descriptive Stats

ooooooo
oooooooooooo
o

Inferential Stats

oooo●oooo
oooooooooooo
o

Bayesian Stats

ooooo

Monte Carlo

ooooooo

AB Testing

How long to run the experiment

A p-value measures the probability of observing a difference between the two variants at least as extreme as what we actually observed, given that there is no difference between the variants. Once the p-value achieves statistical significance or we've seen enough data, the experiment is over.

Statistical Tests
oooooo

Distributions
oooooooooooo
oooooooooooo
o

Descriptive Stats
oooooo
oooooooooooo
o

Inferential Stats
oooo●oo
oooooooooooo
o

Bayesian Stats
ooooo

Monte Carlo
oooooooo

AB Testing

Assumptions in Frequentist Approach for AB tests:

- ▶ after obtaining infinite samples, empirical conversion approaches real conversion
- ▶ data is iid (independent and identically distributed)
- ▶ difference btw empirical conv rates for A and B is due to randomness (H_0 : A and B are same)
- ▶ conv of both versions come from the same distributions (same parameters)

Statistical Tests
oooooo

Distributions
oooooooooooo
oooooooooooo
o

Descriptive Stats
ooooooo
oooooooooooo
o

Inferential Stats
oooooo●o
oooooooooooo
o

Bayesian Stats
ooooo

Monte Carlo
oooooooo

AB Testing

When can Frequentist approach fail

1. We need to know how much of the data we need to collect for the test before starting the test.
2. We can't evaluate the result in real-time as we go, instead we need to wait to make any decision until we collect a full of the planned data size.
3. The test result is not intuitively understandable especially for those without a statistical background. (What is P-value again?)
4. The test result can be read as black and white, either it is statistically significant or not. This makes it hard to figure out what to do especially when not statistically significant

Statistical Tests
oooooo

Distributions
oooooooooooo
oooooooooooo
o

Descriptive Stats
oooooo
oooooooooooo
o

Inferential Stats
oooooooo●
oooooooooooo
o

Bayesian Stats
ooooo

Monte Carlo
oooooooo

AB Testing

In case of small improvements of variants I

- ▶ there are scenarios where we want to stick with the null hypothesis when the treatment variant is marginally better than the control. If the treatment requires a lot of engineering maintenance or causes a disruption to the user experience, the costs of implementing the new variant might outweigh the small benefits
- ▶ In scenarios similar to the one of the slightly better model, Bayesian methodology is appealing because it is more willing to accept variants that provide small improvements. Over the next few years, as we perform hundreds of experiments on the same handful of key business metrics, these marginal gains will accumulate on top of each other. Crucially, since we

Statistical Tests
oooooo

Distributions
oooooooooooo
oooooooooooo
o

Descriptive Stats
ooooooo
o

Inferential Stats
oooooo●
ooooooo
o

Bayesian Stats
ooooo

Monte Carlo
ooooooo

AB Testing

In case of small improvements of variants II

conclude an experiment once we are confident it will provide at least a small improvement, we can iterate more quickly and run more experiments over all.

- ▶ By accepting variants that offer a small improvement, Bayesian A/B testing asserts that the false positive rate — the proportion of times we accept the treatment when the treatment is not actually better — is not very important. While this may be shocking to some statisticians, we agree with this sentiment because not all false positives are created equal. Choosing variant B when its conversion rate is 10% and the conversion rate for variant A is 10.1% is a very different mistake than choosing variant B when the conversion

Statistical Tests
oooooo

Distributions
oooooooooooo
oooooooooooo
o

Descriptive Stats
ooooooo
o

Inferential Stats
oooooo●
ooooooo
o

Bayesian Stats
ooooo

Monte Carlo
ooooooo

AB Testing

In case of small improvements of variants III

rates are 10% for B and 15% for A. Yet, under frequentist methodology, these would both count as a single false positive.

- ▶ Instead, Bayesian A/B testing focuses on the average magnitude of wrong decisions over the course of many experiments. It limits the average amount by which your decisions actually make the product worse, thereby providing guarantees about the long run improvement of a metric. We believe that these types of guarantees are much more relevant to Convoy's use case than the false positive guarantees made by frequentist procedures.

Statistical Tests
oooooo

Distributions
oooooooooooo
oooooooooooo
o

Descriptive Stats
ooooooo
oooooooooooo
o

Inferential Stats
oooooooo●
oooooooooooo
o

Bayesian Stats
ooooo

Monte Carlo
oooooooo

AB Testing

In case of small improvements of variants IV

Bayesian

In experiments where the improvement of the new variant is small, Bayesian methodology is more willing to accept the new variant. By using Bayesian A/B testing over the course of many experiments, we can accumulate the gains from many incremental improvements. Bayesian A/B testing accomplishes this without sacrificing reliability by controlling the magnitude of our bad decisions instead of the false positive rate.

Statistical Tests
oooooo

Distributions
oooooooooooo
oooooooooooo
o

Descriptive Stats
oooooo
o

Inferential Stats
ooooooo
●ooooooo
o

Bayesian Stats
ooooo

Monte Carlo
ooooooo

Hypothesis testing

Type I (false positive) and Type II (false negative) Errors

1. FP (REJECT a TRUE hypothesis): the sample population and AB test results → the challenger will increase conversion rates (so you reject the null hypothesis of both rates)
In reality (for the whole population of visitors), the challenger will NOT increase conversion rates for the overall population
2. FN(NOT REJECT a FALSE hypo): the sample population and AB test results → the challenger will not increase conversion rates (so you do not have enough evidence to reject the null hypothesis about the equality of both rates)
In reality, the challenger will increase conversion rates.

Construct the test: Avoid FP: Assume significance level @0.05

Avoid FN: The more sample, the more probability of rejecting of the false hypothesis (more power).

Statistical Tests
oooooo

Distributions
oooooooooooo
oooooooooooo
o

Descriptive Stats
oooooo
oooooooooooo
o

Inferential Stats
oooooooo
o●oooooooo
o

Bayesian Stats
ooooo

Monte Carlo
oooooooo

Hypothesis testing

p_value

probab of data given hypothesis

p-value is the probab of getting a sample as extreme as ours, given H_0 is true

if H_0 is true, how extreme is our sample?

either reject H_0 or do not reject H_0 because there's less evidence to reject it (given α level of significance)

Hypothesis in AB testing: $H_0 = B$ is better

Better performance of B is statistically significant: B is better

Better performance of B is **not** statistically significant: Need more data

importance of results?

calculate confidence intervals to capture uncertainty of experiment

<http://www.statisticshowto.com/what-is-statistical-significance/>

Statistical Tests
oooooo

Distributions
oooooooooooo
o

Descriptive Stats
oooooo
o

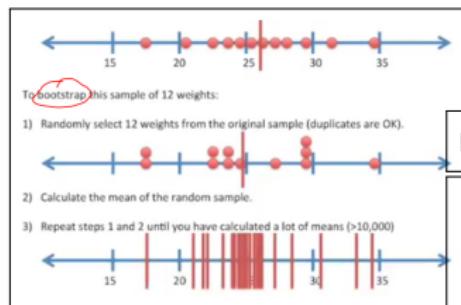
Inferential Stats
ooooooo
o●○oooo

Bayesian Stats
ooooo

Monte Carlo
ooooooo

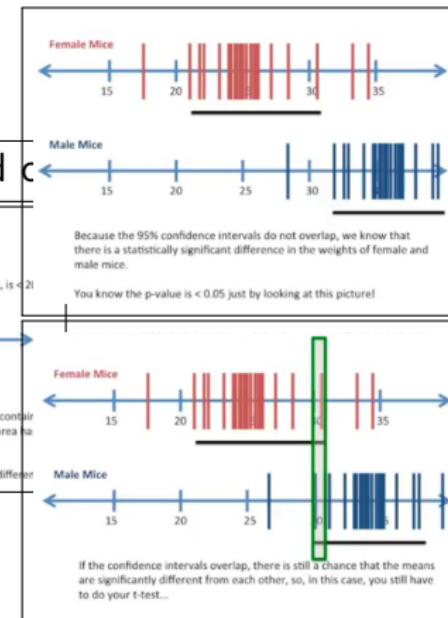
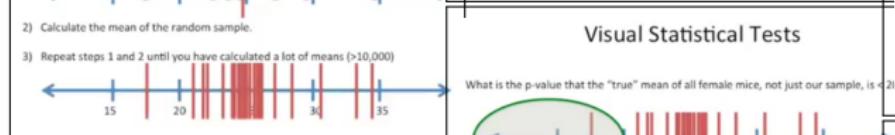
Hypothesis testing

confidence interval



That is to say, the p-value of anything outside of the confidence interval is < 0.05 (and thus, significantly different).

nvjfjnfvnkevnkervnmkd c



Statistical Tests

oooooo

Distributions

oooooooooooo
oooooooooooooo
o

Descriptive Stats

ooooooo
oooooooooooo
o

Inferential Stats

oooooooooooo
oooo●ooooo
o

Bayesian Stats

ooooo

Monte Carlo

ooooooo

Hypothesis testing

Statistical power vs Significance level

content...

Statistical Tests

oooooo

Distributions

oooooooooooo
oooooooooooooo
o

Descriptive Stats

ooooooo
oooooooooooo
o

Inferential Stats

oooooooooooo
oooo●oooo
o

Bayesian Stats

ooooo

Monte Carlo

oooooooo

Hypothesis testing

Define sample size in advance to avoid FN

content...

Statistical Tests
oooooo

Distributions
oooooooooooo
oooooooooooo
o

Descriptive Stats
ooooooo
o

Inferential Stats
ooooooo
ooooo●ooo
o

Bayesian Stats
ooooo

Monte Carlo
ooooooo

Hypothesis testing

Problem with significance based testing

- ▶ not intuitive
- ▶ ↑ sample size → experiment run time
- ▶ p-value reaches significance very early on due to novelty effect
- ▶ statistical significance is not valid stopping criteria
- ▶ depends on arbitrary parameters (95% conf, 0.05 p-val)
- ▶ confidence interval: contains true parameter with 95% probability
not 95% probability the true parameter falls within the interval

Statistical Tests

oooooo

Distributions

oooooooooooo
oooooooooooooo
o

Descriptive Stats

ooooooo
oooooooooooo
o

Inferential Stats

oooooooo
oooooooo●○
o

Bayesian Stats

ooooo

Monte Carlo

oooooooo

Hypothesis testing

Multiple Null Hypothesis test

$$1-(1-0.05)^n$$

Statistical Tests

oooooo

Distributions

oooooooooooo
oooooooooooooo
o

Descriptive Stats

ooooooo
o

Inferential Stats

oooooooo
ooooooo●
o

Bayesian Stats

ooooo

Monte Carlo

ooooooo

Hypothesis testing

t-statistics vs z-statistics

when population std dev is known then we use z-statistics, if unknown then we use t-statistics

t-statistics assumes that underlying distribution is normal

t-distribution is bell curved, defined by it's DF (degrees of freedom)

Measure of extremeness, $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

↑ z value, more likely to reject H_0

Statistical Tests

oooooo

Distributions

oooooooooooo
oooooooooooooo
o

Descriptive Stats

ooooooo
oooooooooooo
o

Inferential Stats

oooooooooooo
oooooooooooo

Bayesian Stats

ooooo

Monte Carlo

oooooooo

something

Outlier and Confidence Interval

presented using box and whisker plots

Statistical Tests

oooooo

Distributions

oooooooooooo
oooooooooooo
o

Descriptive Stats

ooooooo
oooooooooooo
o

Inferential Stats

oooooooooooo
oooooooooooo
o

Bayesian Stats

●oooo

Monte Carlo

oooooooo

Bayesian Reasoning

Frequentist vs Bayesian I

Frequentist Approach

treats random events probabilistically and doesn't quantify the uncertainty

usually used for AB Testing when no uncertainty involved such as fair C

Option B is better than A with p-value of 0.03

otherwise need more data

assumes iid

wait till data has been collected

using only data from your current experiment

frequentist methods assume that you repeat your experiment many, many times

take a sample, measure and average their height to produce a point estimate

Statistical Tests

oooooo

Distributions

oooooooooooo
oooooooooooooo
o

Descriptive Stats

ooooooo
o

Inferential Stats

oooooooo
o

Bayesian Stats

o●ooo

Monte Carlo

ooooooo

Bayesian Reasoning

Going Bayesian

Update prior beliefs with new beliefs

Bayes Theorem

$$P(H|E) = \frac{P(E|H)*P(H)}{P(E)}$$

Posterior \propto likelihood x Prior

where

Posterior probab of hypothesis given the evidence = $P(H|E)$

Likelihood of evidence if hypothesis is true = $P(E|H)$

Prior Probab of Hypothesis = $P(H)$

Prior Probab that Evidence is True = $P(E)$

Statistical Tests

oooooo

Distributions







Descriptive Stats

oooooo

o

Inferential Stats

oooooo

o

Bayesian Stats

oo●oo

Monte Carlo

ooooooo

Bayesian Reasoning

Bayesian AB Testing I

Given data:

From sample: CVR_A and CVR_B

From prior population: CVR_A (historical data)
strong prior distribution

Non-overlapping populations

$$P(CVR_A, CVR_B | \text{data}) = P(CVR_A | \text{data}) * P(CVR_B | \text{data})$$

$$P(CVR_A, CVR_B | \text{data}) = \frac{P(\text{data} | CVR_A)P(A)*P(\text{data} | CVR_B)P(B)}{P(\text{data})P(\text{data})}$$

Success/failure output, Likelihood follows Binomial Dist

$$P(\text{views}, \text{clicks} | CVR) = \binom{\text{views}}{\text{clicks}} CVR^{\text{clicks}} (1 - CVR)^{\text{views} - \text{clicks}}$$

Prior probab follows Beta Dist

Statistical Tests

oooooo

Distributions

oooooooooooo
oooooooooooooo
o

Descriptive Stats

ooooooo
o

Inferential Stats

oooooooo
o

Bayesian Stats

oo●oo

Monte Carlo

ooooooo

Bayesian Reasoning

Bayesian AB Testing II

$$P(CVR_A) = \frac{(1-CVR_A)^{b-1} CVR_A^{a-1}}{B(a,b)}$$

Binomial Likelihood implies conjugate Beta

→ Beta Posterior → Prior and Posterior are of same family

$$P(CVR|views, clicks) = \frac{P(views, clicks|CVR)P(CVR)}{P(views, clicks)}$$

where, CVR = Conversion Rate

Statistical Tests

oooooo

Distributions

oooooooooooo
oooooooooooooo
o

Descriptive Stats

ooooooo
oooooooooooo
o

Inferential Stats

oooooooooooo
oooooooooooo
o

Bayesian Stats

oooo●o

Monte Carlo

oooooooo

Bayesian Reasoning

Bayesian AB Testing.. contd...

Posterior Probab

$$\text{Beta}(\text{CVR}_A, a + \text{clicks}_A, b + \text{views}_A - \text{clicks}_A)$$

Choosing a, b for Beta dist

Uninformative prior: Randomly choose a and b

Informative prior: Decide a and b after running the experiment multiple times and already analyzed CVR

Statistical Tests
oooooo

Distributions
oooooooooooo
oooooooooooo
o

Descriptive Stats
ooooooo
oooooooooooo
o

Inferential Stats
oooooooooooo
oooooooooooo
o

Bayesian Stats
oooo●

Monte Carlo
oooooooo

Bayesian Reasoning

get average height bayesian example I

1. Initially, the Bayesian statistician has some basic prior knowledge which is being assumed: for example, that the average height is somewhere between 50cm and 250cm.
2. Then, the Bayesian begins to measure heights of specific citizens, and with each measurement updates the distribution to become a bit more “bell-shaped” around the average height measured so far. As more data is collected, the “bell” becomes sharper and more concentrated around the measured average height.

Statistical Tests
oooooo

Distributions
oooooooooooo
oooooooooooo
o

Descriptive Stats
ooooooo
o

Inferential Stats
ooooooo
o

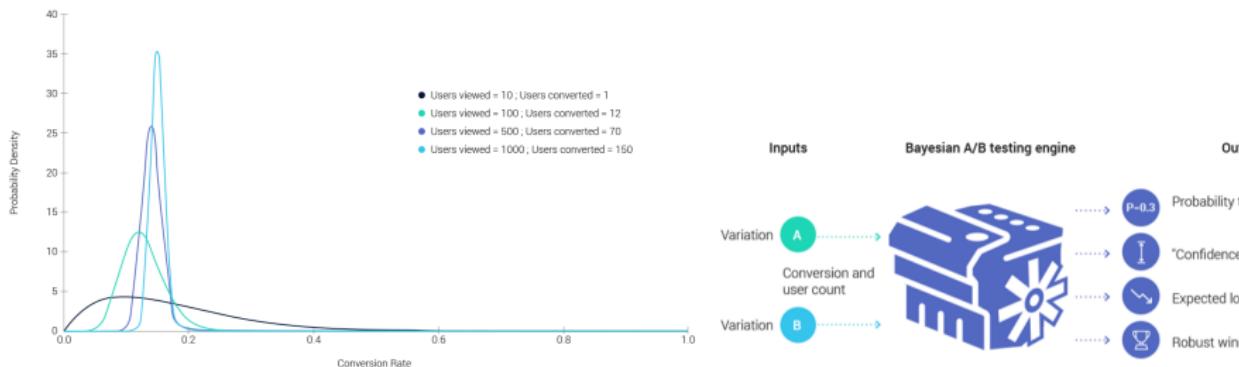
Bayesian Stats
oooo●

Monte Carlo
ooooooo

Bayesian Reasoning

get average height bayesian example II

- For Bayesians, probabilities are fundamentally related to their knowledge about an event. This means, for example, that in a Bayesian view, we can meaningfully talk about the probability that the true conversion rate lies in a given range, and that probability codifies our knowledge of the value based on prior information and/or available data.



Statistical Tests

oooooo

Distributions

oooooooooooo
oooooooooooooo
o

Descriptive Stats

ooooooo
oooooooooooo
o

Inferential Stats

oooooooooooo
oooooooooooo
o

Bayesian Stats

ooooo

Monte Carlo

●oooooooo

Deterministic vs Stochastic

content...

Statistical Tests

oooooooo

Distributions

oooooooooooo
oooooooooooooo
o

Descriptive Stats

ooooooo
oooooooooooo
o

Inferential Stats

oooooooo
oooooooooooo
o

Bayesian Stats

ooooo

Monte Carlo

o●ooooo

Random Walks

content...

Statistical Tests

oooooo

Distributions

oooooooooooo
oooooooooooooo
o

Descriptive Stats

ooooooo
oooooooooooo
o

Inferential Stats

oooooooooooo
oooooooooooo
o

Bayesian Stats

ooooo

Monte Carlo

ooo●oooo

Monte Carlo

content...

Statistical Tests

oooooo

Distributions

oooooooooooo
oooooooooooooo
o

Descriptive Stats

ooooooo
oooooooooooo
o

Inferential Stats

oooooooooooo
oooooooooooo
o

Bayesian Stats

ooooo

Monte Carlo

oooo●oooo

Gibbs Sampler

content...

Statistical Tests

oooooo

Distributions

oooooooooooo
oooooooooooo
o

Descriptive Stats

ooooooo
oooooooooooo
o

Inferential Stats

oooooooooooo
oooooooooooo
o

Bayesian Stats

ooooo

Monte Carlo

oooo●ooo

How to choose the method of predictive modelling. algorithms

<https://www.quora.com/What-is-the-difference-between-supervised-and-unsupervised-learning-algorithms>

<https://blog.udacity.com/2016/04/5-skills-you-need-to-become-a-machine-learning-engineer.html>

<https://towardsdatascience.com/how-to-build-a-data-science-portfolio-5f566517c79c>

<https://www.youtube.com/watch?v=qv6UVOQ0F44>

<https://en.wikipedia.org/wiki/F-test>

numpy matplotlib

<http://slideplayer.com/slide/6260251/>

<http://www.sfu.ca/ ber1/iat802/pdfs/When>

<https://www.youtube.com/watch?v=RlhNbPZC0A>

Statistical Tests

oooooooo

Distributions

oooooooooooo
oooooooooooooo
o

Descriptive Stats

ooooooo
oooooooooooo
o

Inferential Stats

oooooooo
oooooooooooo
o

Bayesian Stats

ooooo

Monte Carlo

oooooo●○

Statistical Tests

oooooooo

Distributions

oooooooooooo
oooooooooooooo
o

Descriptive Stats

ooooooo
oooooooooooo
o

Inferential Stats

oooooooo
oooooooooooo
o

Bayesian Stats

ooooo

Monte Carlo

oooooo●

Thank You!