Summary of General Hypothesis Test Procedure:

1. Define the **null hypothesis**, which is the uninteresting or default explanation.
2. Assume that the null hypothesis is true, and determine the probability rules for the possible outcomes of the experiment.
3. After collecting data, compute the probability of the final outcome or even more extreme outcomes.

Null Hypothesis: Hy that easily allows you to calculate the probability of a particular statistic

Fair Coin Experiment

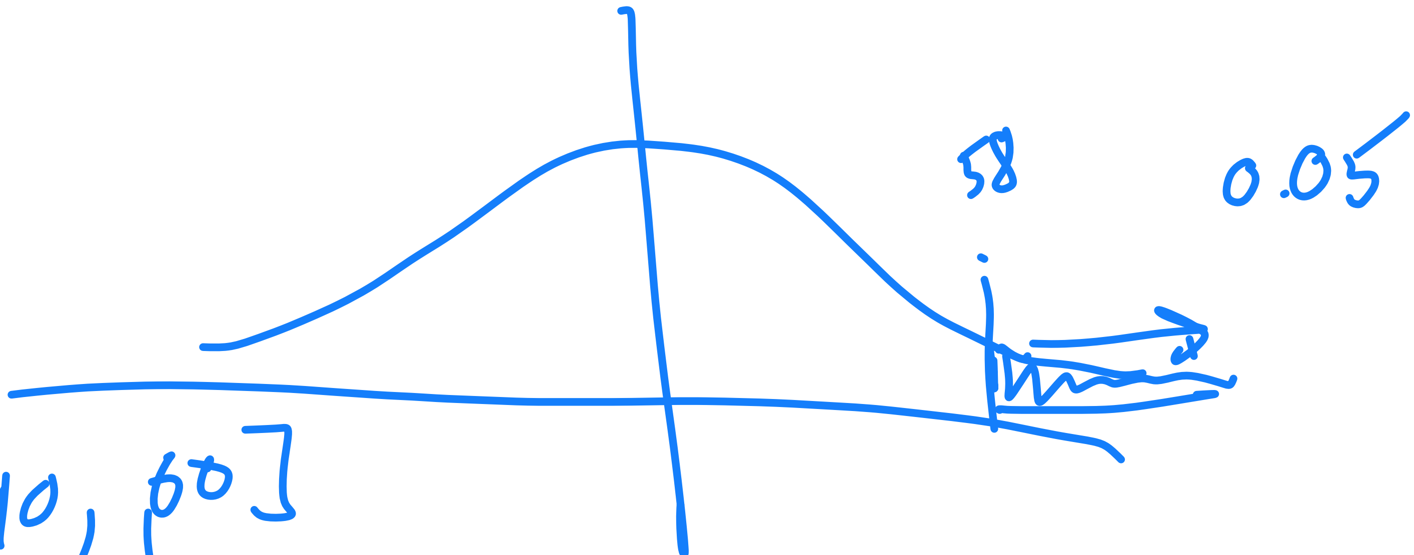
Exp - flips, $T = T(X_1, \dots, X_n) = \sum_{i=1}^{n} X_i$

H_0 - coin fair, $p = 0.5$

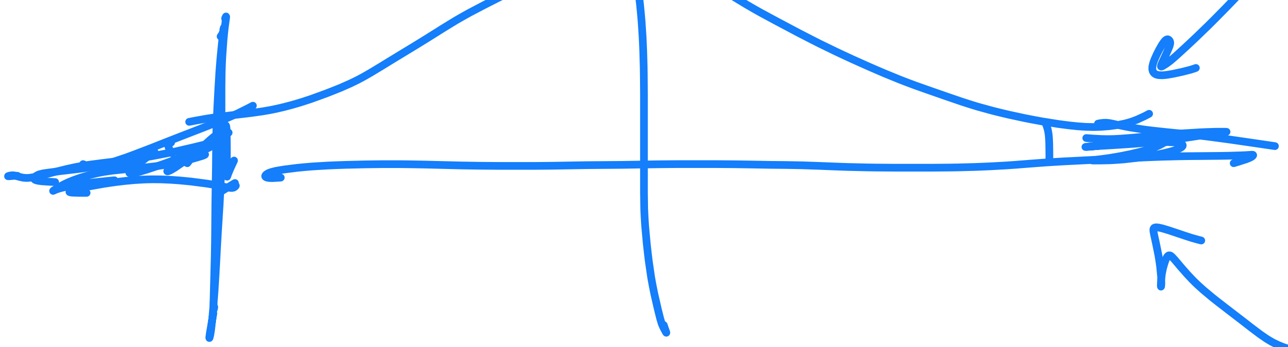
H_1 - $\underbrace{p \neq 0.5}_{\text{double sided}}$, $\underbrace{p > 0.5, \quad p < 0.5}_{\text{single sided}}$

$T \sim Bin(n, p) = Bin(n, \tfrac{1}{2})$ $\quad n = 100$

$t = \sum_i x_i$ $\quad$ qbinom$(.95, 100, \tfrac{1}{2}) = 58$



[40, 60]



Low ~

qbinom(0.025, 100, 1/2)

qbinom(0.975, 100, 1/2)

Approx w/ Normal

$T \sim N(50, 100 \times \frac{1}{4}) = N(50, 25)$

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

$\Updownarrow$

$T = Z\sigma + \mu$

$Z_{1-\alpha} = Z_{0.95} = 1.64$

$(1.64 \times 5) + 50 = 58.2$

Error

Table of error types		Null hypothesis (H_0) is	
		True	**False**
Decision about null hypothesis (H_0)	**Fail to reject**	Correct inference (true negative) (probability = $1-\alpha$)	Type II error (false negative) (probability = β)
	Reject	Type I error (false positive) (probability = α)	Correct inference (true positive) (probability = $1-\beta$)

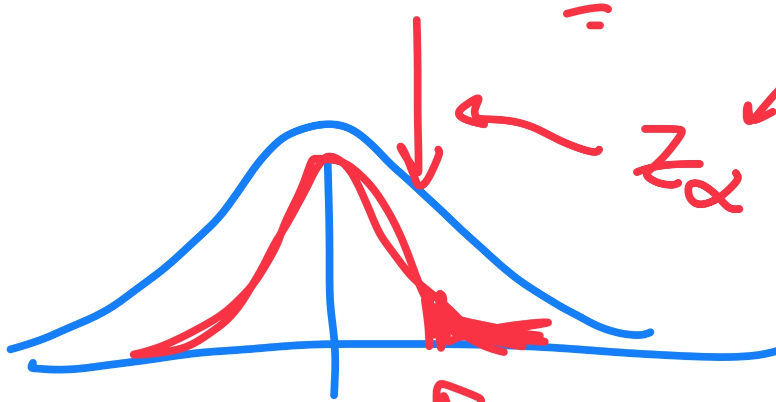
$$P(\text{Reject} \mid H_0 \text{ true}) = \alpha$$

① Hypothesis Testing of Means:

H_0

$\sigma =$

$\bar{X}$ - sample mean.



$\bar{X} \sim N(\mu_0, \sigma^2/n)$

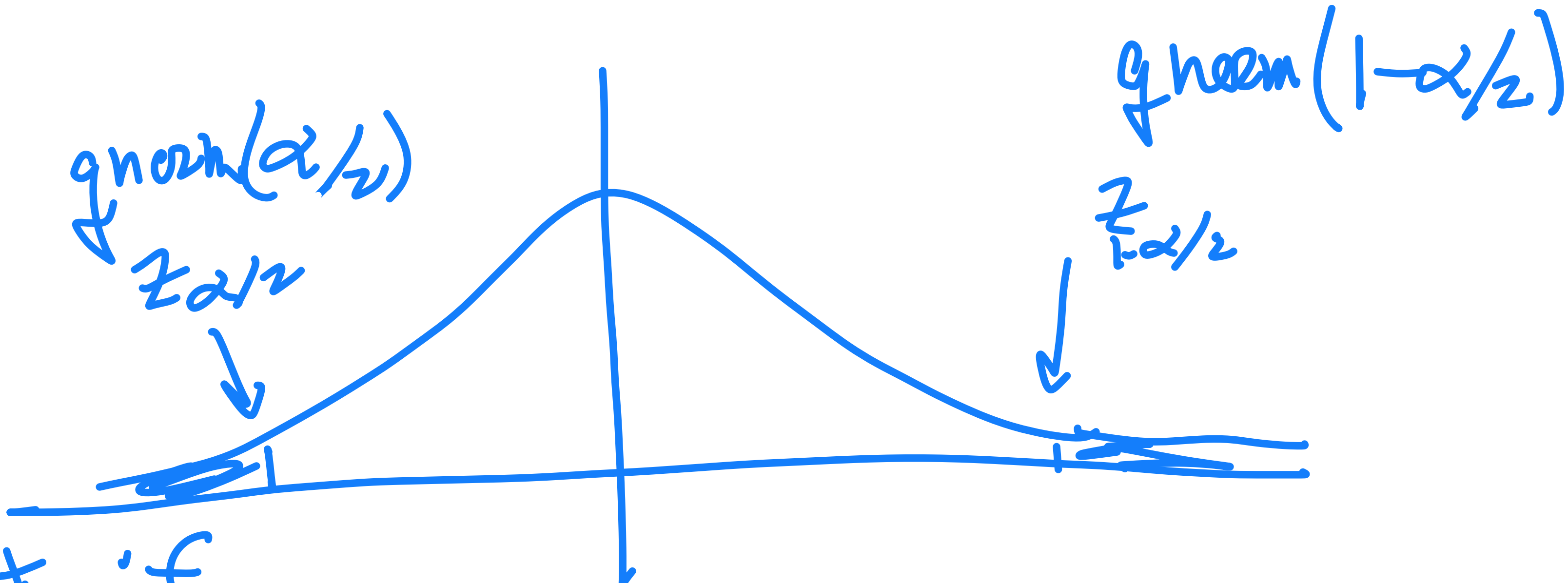
is $Z > Z_\alpha$

critical

$Z = (\bar{X} - \mu_0)/(\sigma/\sqrt{n}) \sim N(0,1)$

Hyp of Mean

1b



Reject if

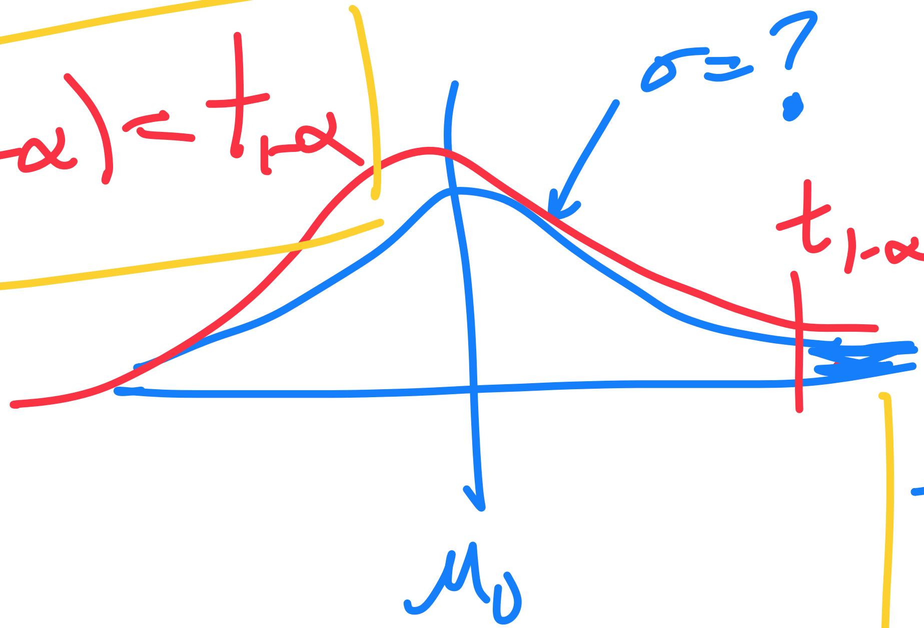
$z < z_{\alpha/2}$

$z > z_{1-\alpha/2}$

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

(2)

$\boxed{qt(1-\alpha) = t_{1-\alpha}}$



$(X_1 \dots X_n)$ - samples

S_n^2 - sample variance

$$\boxed{T = \frac{\bar{X} - \mu_0}{S_n / \sqrt{n}}}$$

$T \sim s\text{-}t(dof = n-1)$

If $T > t_{1-\alpha}$
Reject H_0

Paired Samples - Hypothesis Testing

Same set of specimens two conditions

"before" & "after"

$$\begin{pmatrix} X_1 \\ \\ X_n \end{pmatrix} \quad \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$$

$$\bar{X} \qquad \bar{Y}$$

$$Y_i - X_i = h_i$$

$$(h_1 \ldots h_n)$$

$$\bar{h} = \frac{1}{n} \sum_{i=1}^{n} h_i$$

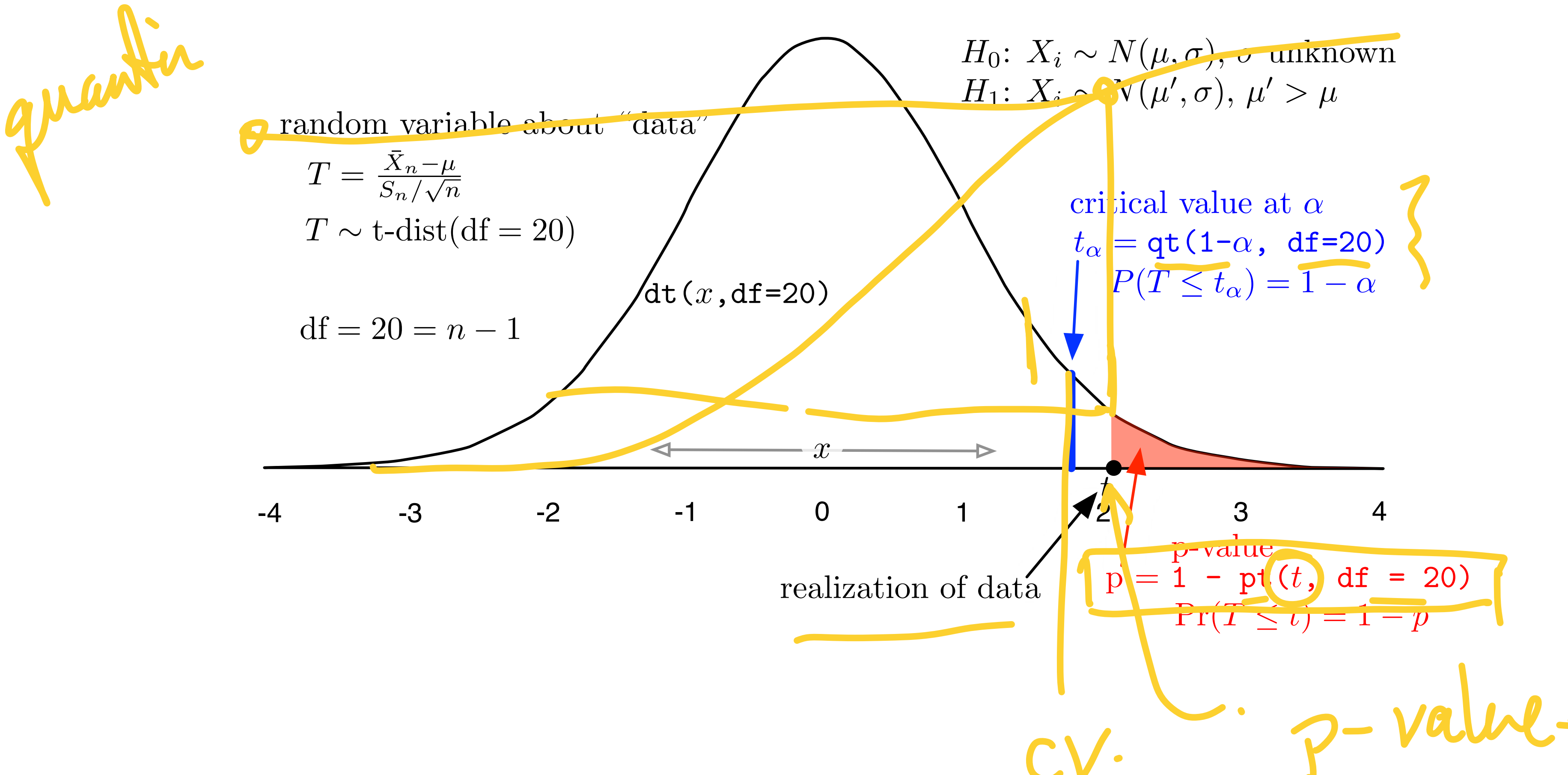
$\bar{h} \sim N\left(0, \sigma^2/n\right) \simeq N\left(0, S_n^2/n\right)$

$$T = \frac{\bar{h} - 0}{S_n/\sqrt{n}} \leftarrow \text{student-t dist.}$$

H_0 = mean of h is zero

$t_{1-\alpha}$ $\quad \alpha = 0.05$

If $T > t_{1-\alpha}$ reject H_0



H_0: $X_i \sim N(\mu, \sigma)$, σ unknown
H_1: $X_i \sim N(\mu', \sigma)$, $\mu' > \mu$
random variable about "data"
$T = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}}$
$T \sim$ t-dist(df $= 20$)
dt(x,df=20)
df $= 20 = n - 1$
critical value at α
$t_\alpha =$ qt(1-α, df=20)
$P(T \le t_\alpha) = 1 - \alpha$
x
-4
-3
-2
-1
0
1
2
3
4
t
realization of data
p-value
p = 1 - pt(t, df = 20)
$\Pr(T \le t) = 1 - p$
quantile
CV.
p-value.

Two sample hypothesis test
equal variances

Scenario: two populations, unknown means, unknown variances. (equal).

$\sigma_x^2 = \sigma_y^2 = \sigma^2$

$X = X_1 \ldots X_n$

$Y = Y_1 \ldots Y_m$

$\bar{X}_n, \bar{Y}_m$

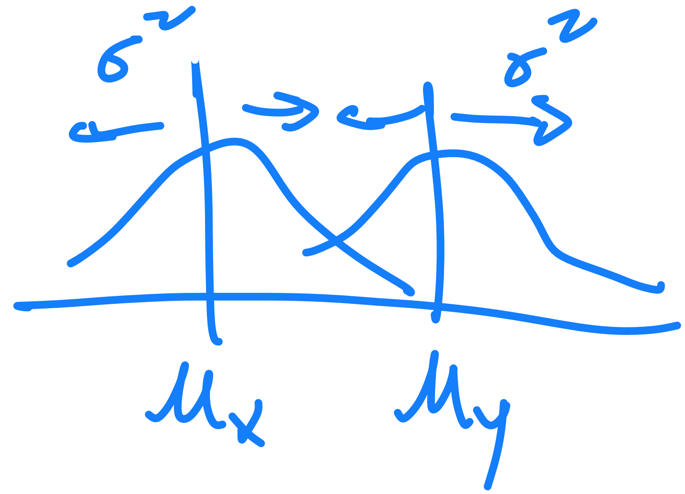
$S_x^2 \quad S_y^2$

$$S_p^2 = \frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}$$

statistic

$$T = \frac{\bar{X}_n - \bar{Y}_m}{S_p\left(\frac{1}{n}+\frac{1}{m}\right)^{\frac{1}{2}}}$$

student with $n+m-2$ degrees of freedom



Statistical Simulation

What is simulation?

Why?

Complex

Predict outcomes. - average

Monte Carlo.

Random #s in Comp.

Pseudo-random #s.

Generate sequences of integers

sequence has memory.

Polynomials → chaotic

uniform dist

Pseudo-Random.

Standard uniform distribution

$u \sim U(0,1)$ ☆

Ex: $\text{Ber}(p)$ $\quad u \leftarrow \text{runif}(\,)$

$$b = \begin{cases} 1 & \text{if } u < p \\ 0 & \text{if } u \geq p \end{cases}$$

u type float

type I - incorrectly reject H_0

prob $= \alpha$

standard
- normal
- t: dist

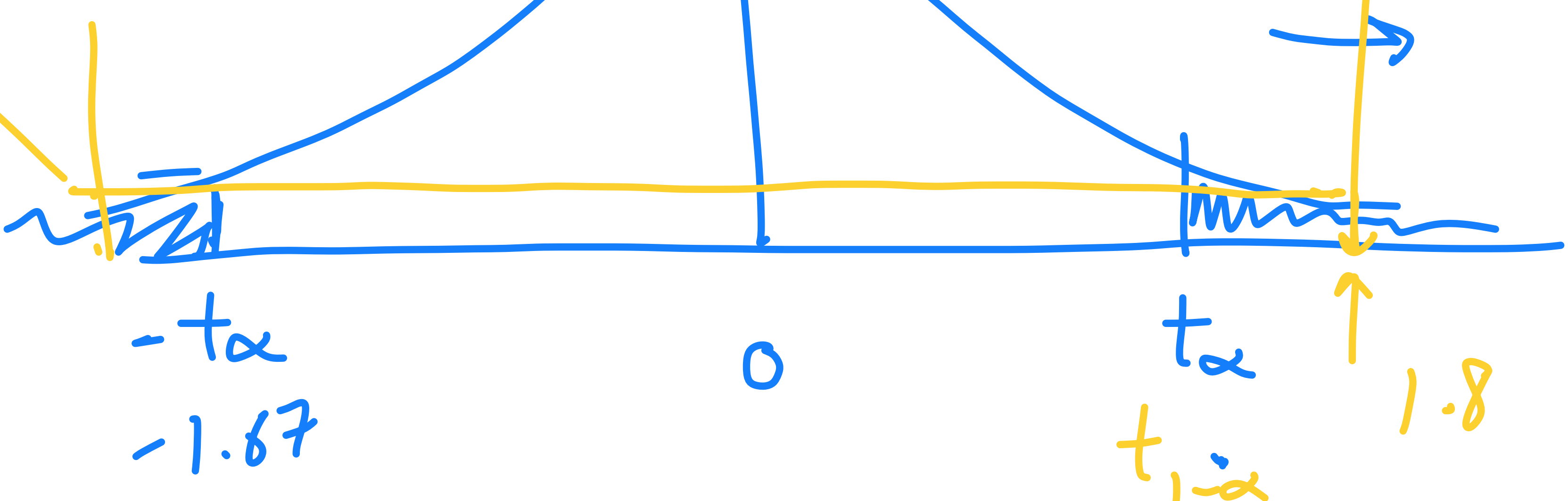
$pt(\#, df)$

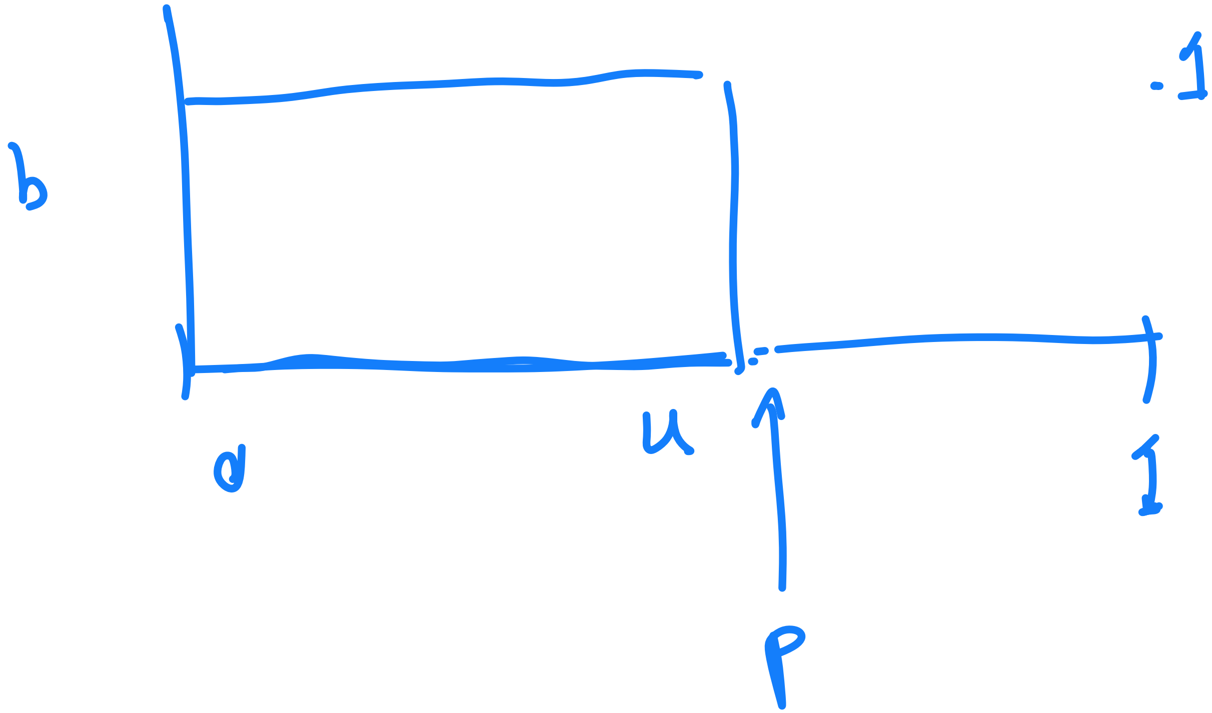
$pnorm(\ )$

p-value $< \alpha$

$\alpha = 0.05$

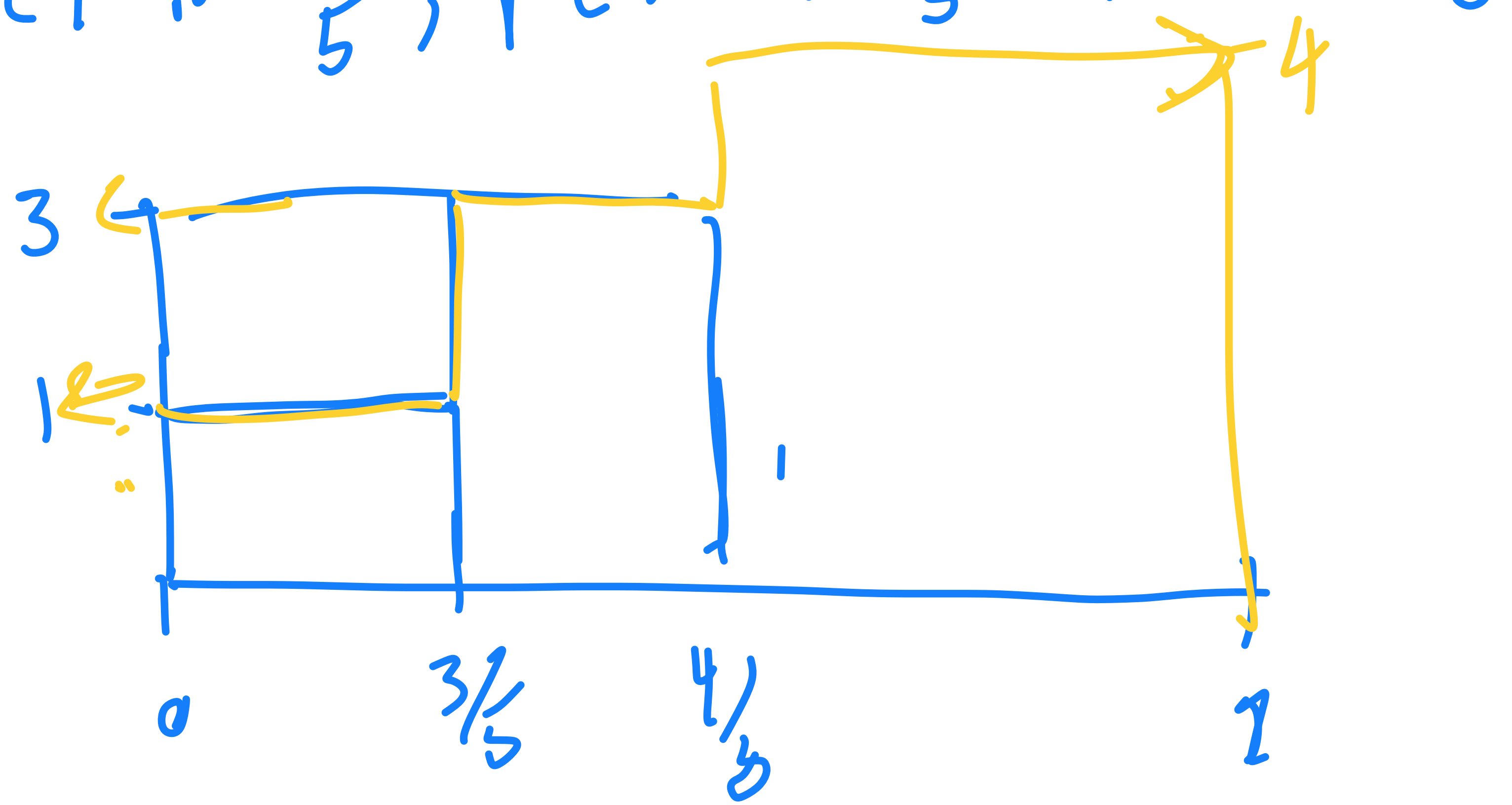
t_n




b
1
a
u
1
p

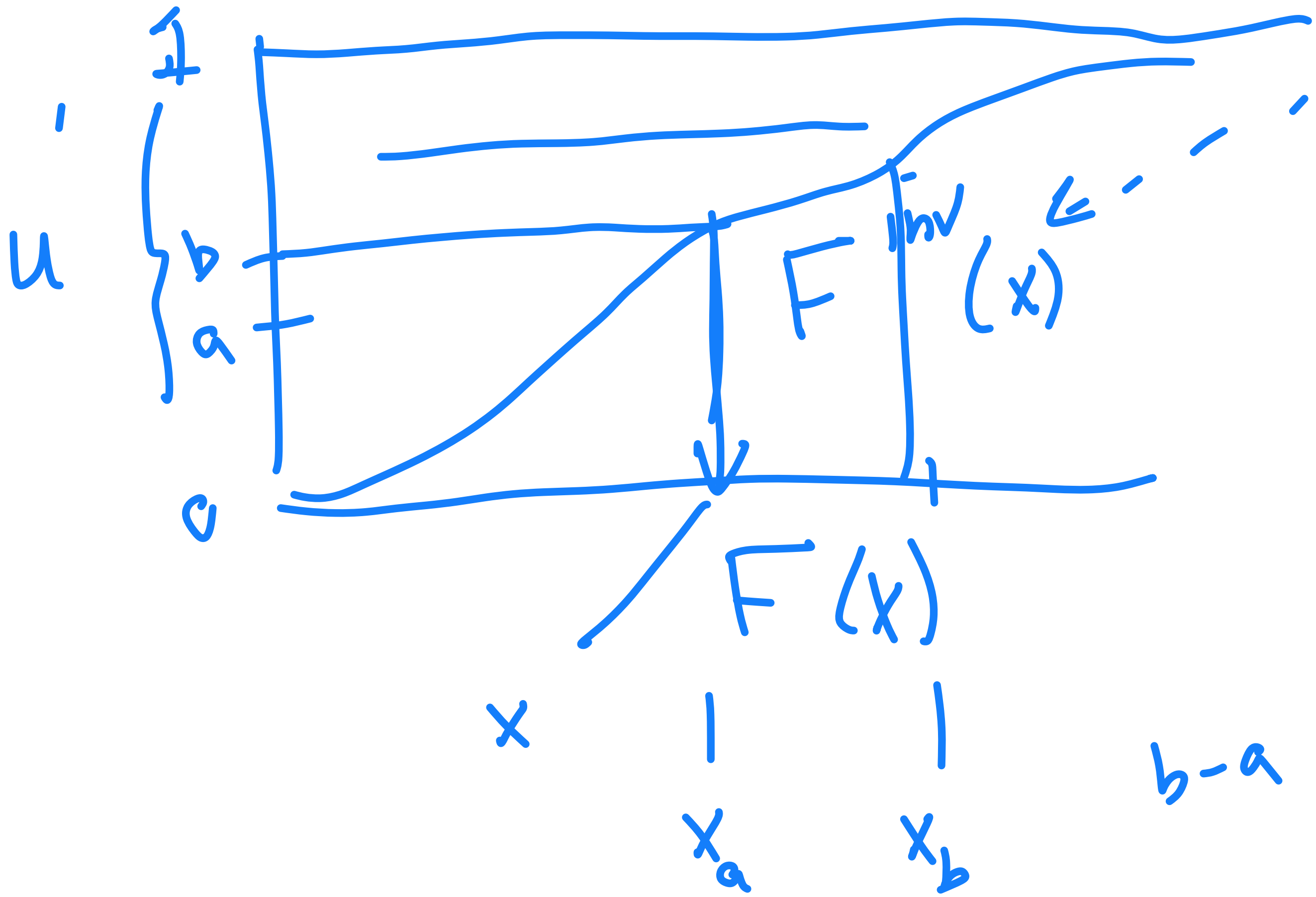
Y RV.

$P(Y=1) = \frac{3}{5}$, $P(Y=3) = \frac{1}{5}$ $P(Y=4) = \frac{1}{5}$



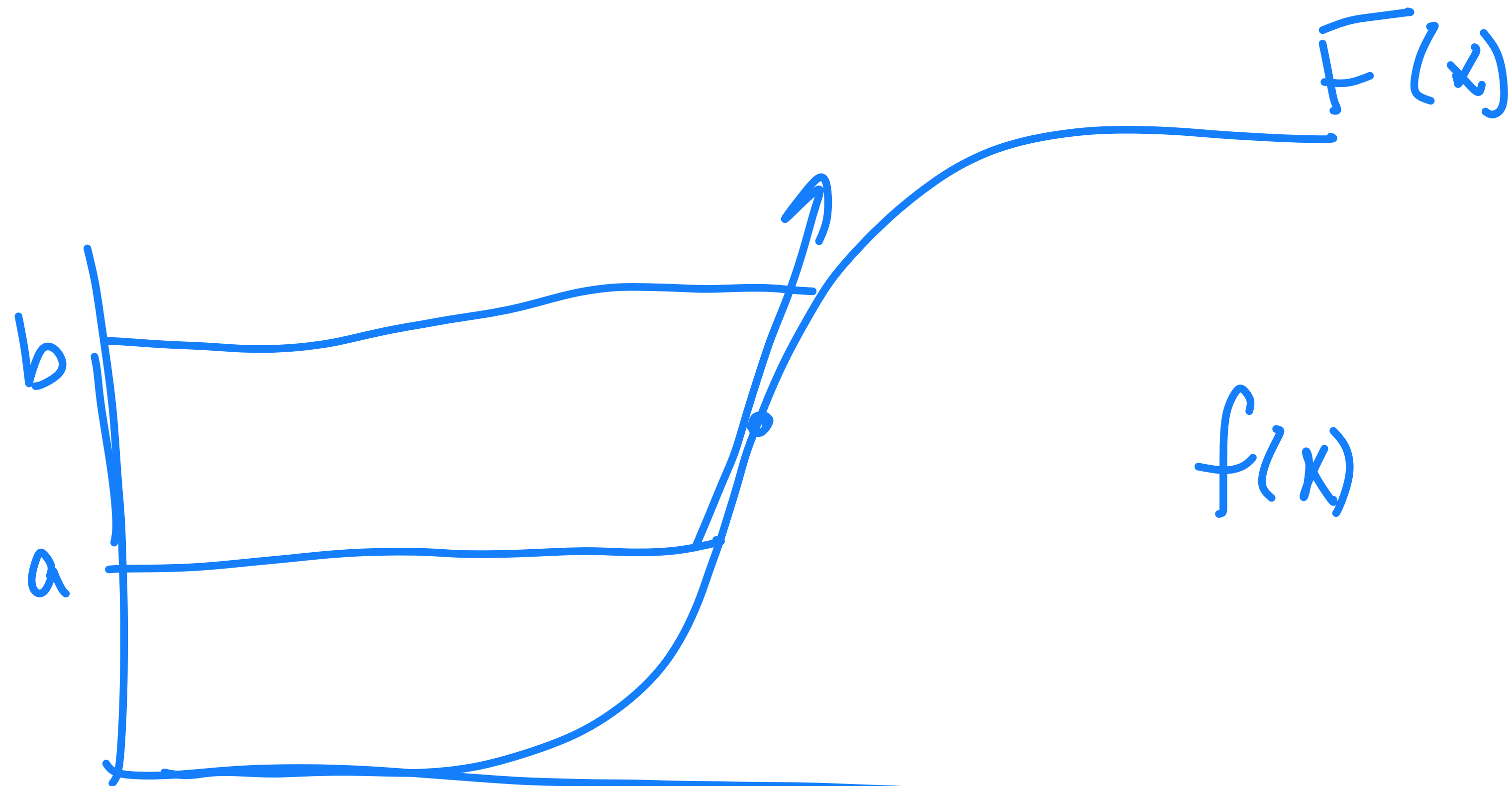
$$y = \begin{cases} 1 & 0 \le u \le 3/5 \\ 3 & \frac{3}{5} < u \le 4/5 \\ 4 & \frac{4}{5} < u \le 1 \end{cases}$$

Cont RV's in Sim



quantile

$f(x)$ - pdf



F(x)
b
a
f(x)

Generate samples from $\mathrm{Exp}(\lambda)$

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

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

$$1 - e^{-\lambda x} = u \implies 1 - u = e^{-\lambda x}$$

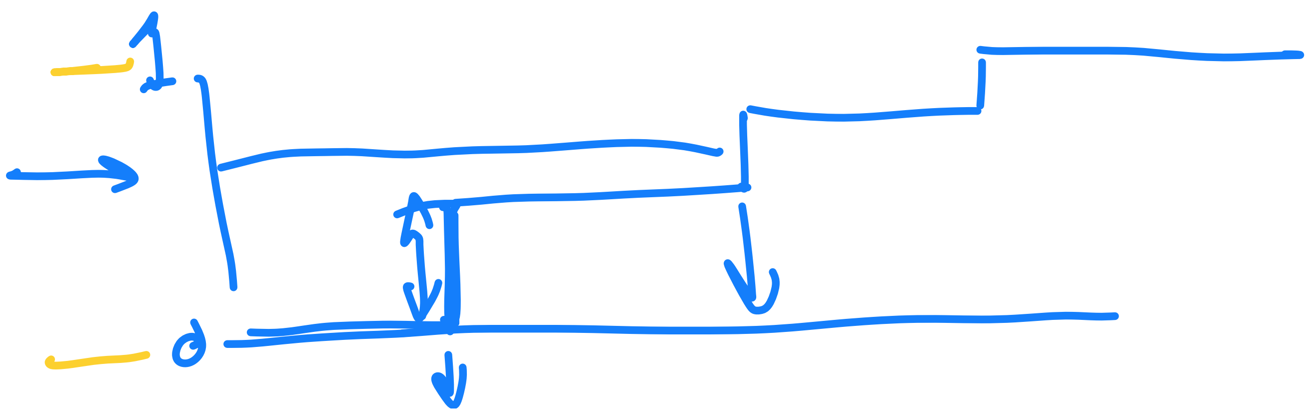
$$\ln(1-u) = -\lambda x \qquad x = -\frac{\ln(1-u)}{\lambda} \doteq -\frac{\ln(u)}{\lambda}$$

Generating Random #s.

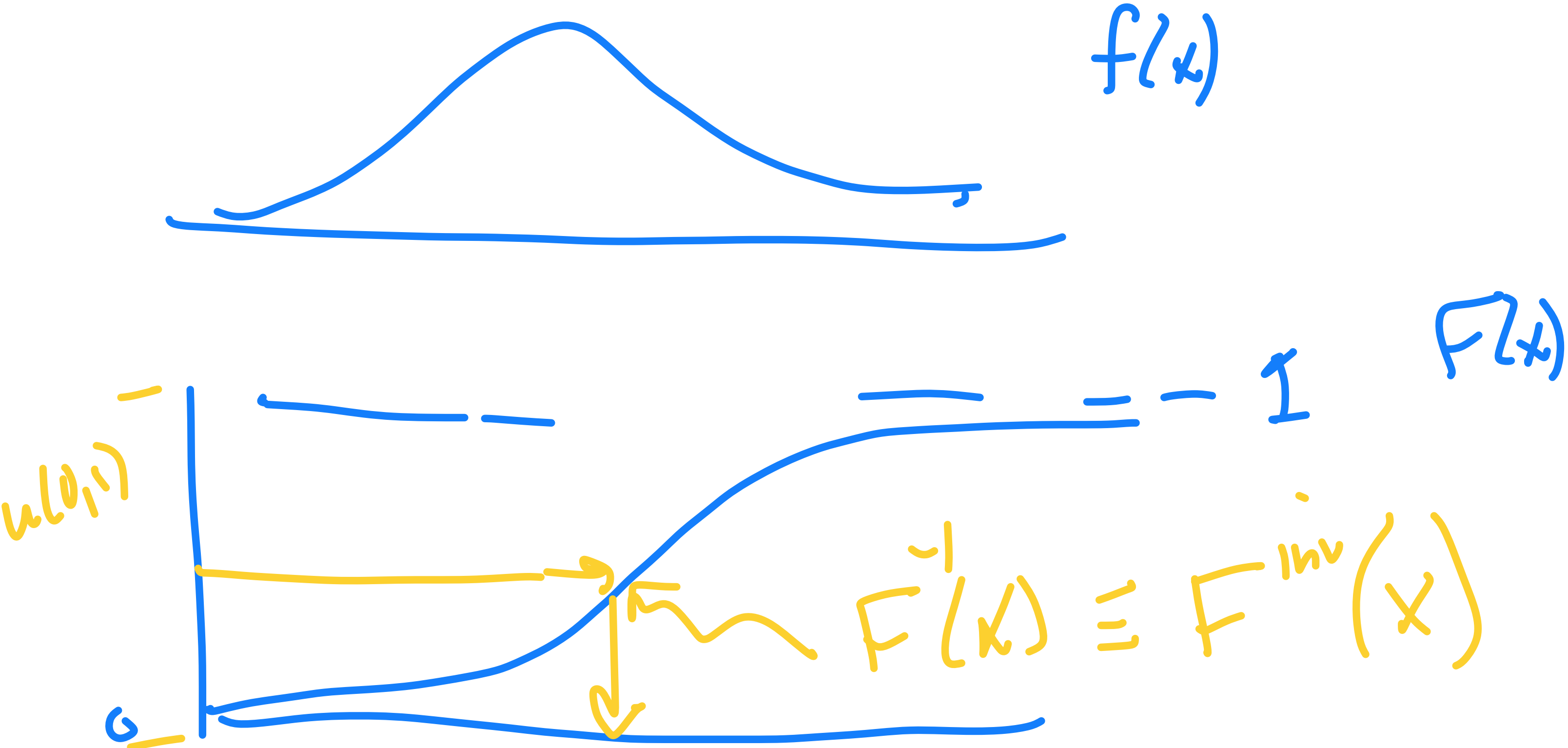
$u \sim U(0,1)$

Ex. Bernoulli

$$b = \begin{cases} 1 & u \le p \\ 0 & > p \end{cases} \qquad b = \begin{cases} 0 & u \le 1-p \\ 1 & u > 1-p \end{cases}$$



Cont. RV's.



$$F^{-1}(x) \equiv F^{inv}(x)$$

Exp RV's.

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

$$X = -\frac{\ln(u)}{\lambda}$$

Note:
many distributions - formulas & approx
Normal.

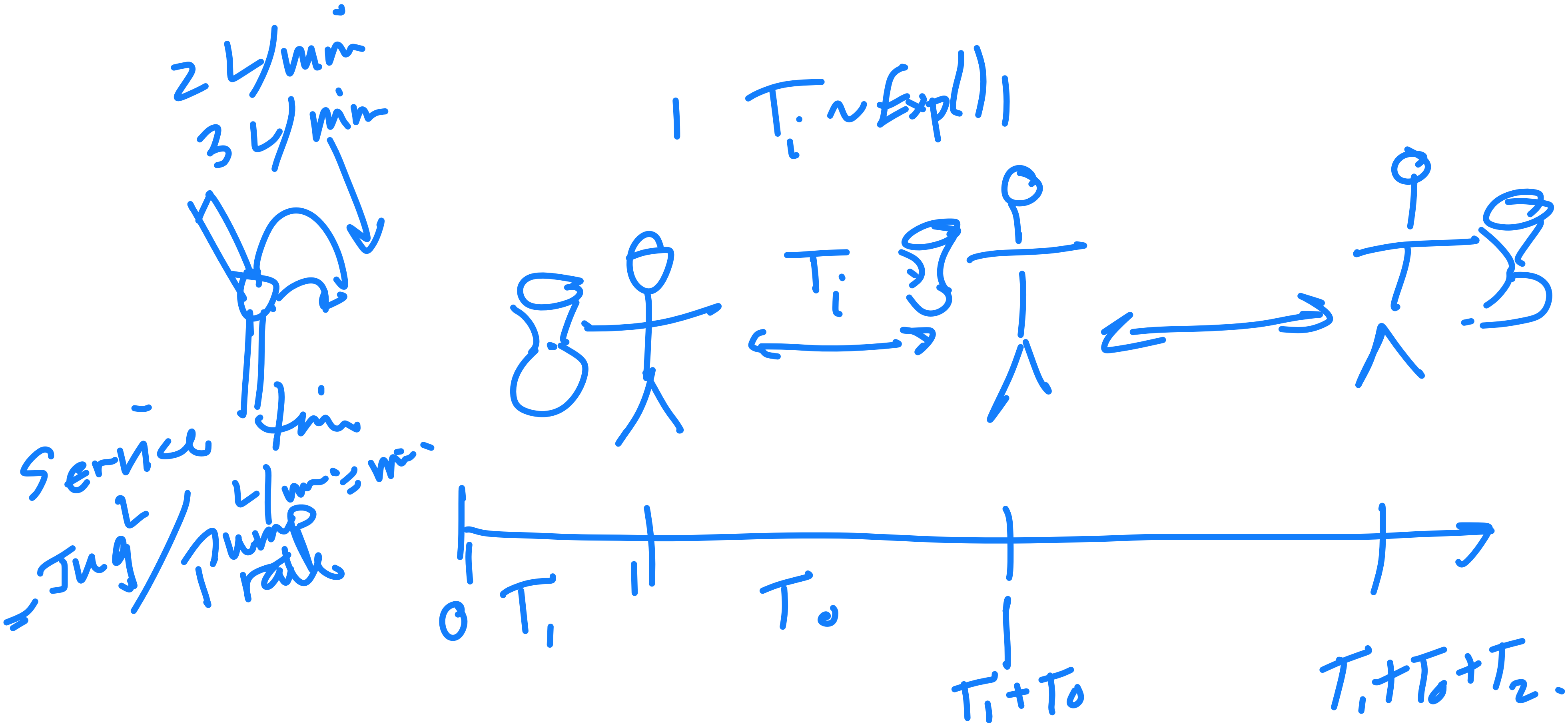
Simulations

1) Identify random variables
   - strategy for generation
2) Generate #s
3) Apply "model"
4) Analyze outcomes.
   descriptive statistics

HW 7 – Q2

$\lambda = \frac{1}{min}$ $n = 120 \times \lambda$.

Scenario

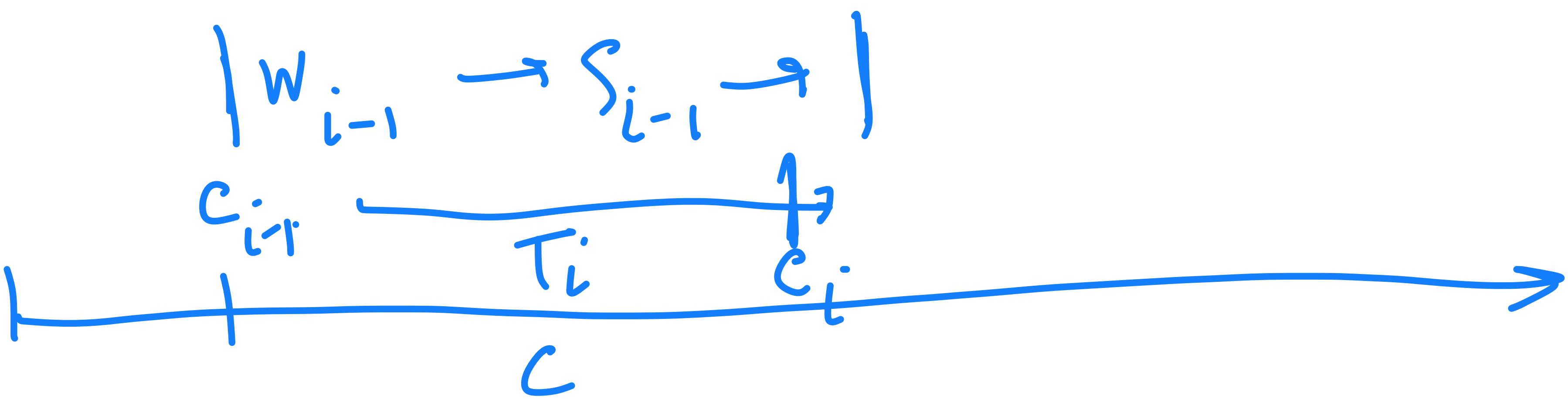


$$W_i = \max\{W_{i-1} + S_{i-1} - T_i, 0\}.$$

waiting time

service time

Time between arrival $i-1$ and T_i

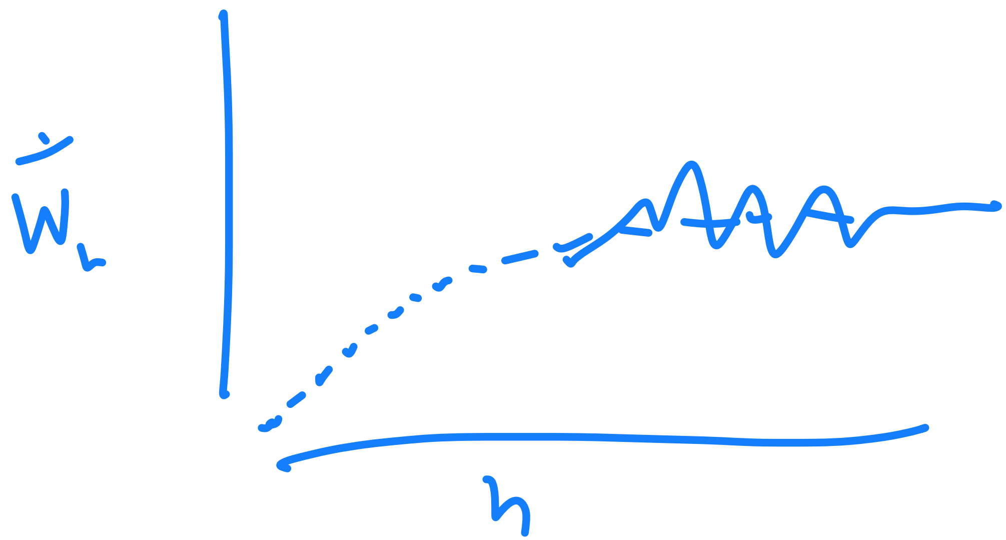


waiting times

mean, median, max, 95% percentile

histogram

$$\bar{W}_n = \frac{\sum_{i=1}^{n} W_i}{n}.$$



Linear Regression

Fitting models

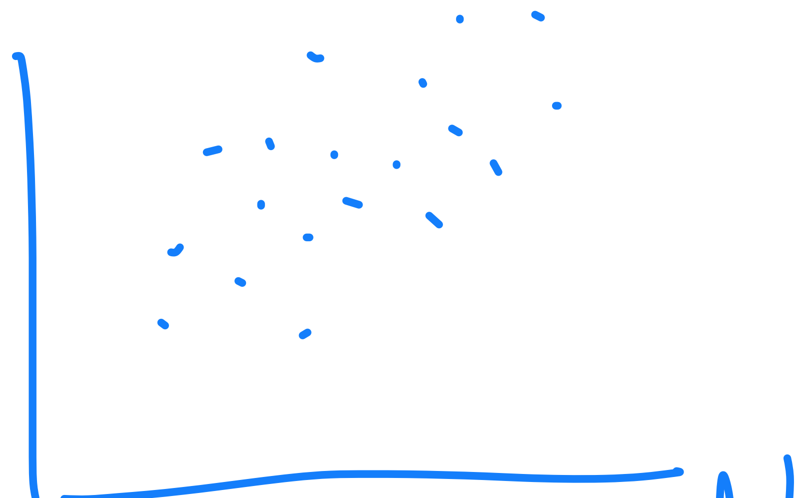
$y_i = \alpha + \beta x_i + \varepsilon_i$

dependent - y_i

ε_i - random noise

$\varepsilon_i \sim N(\ )$

$E[\varepsilon_i] = 0$



x_i - independent

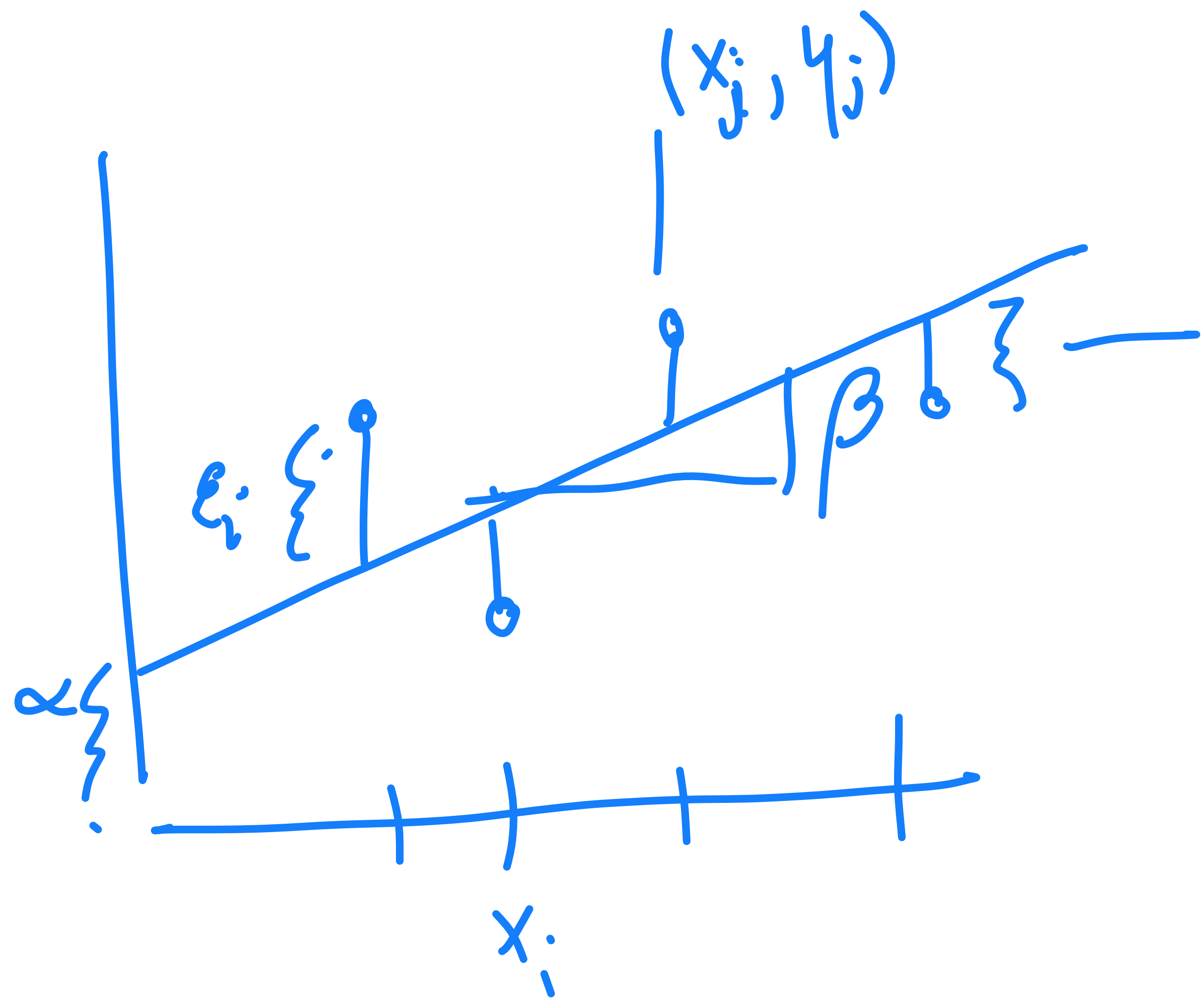
Models.

Generative model.

Find parameters

α, β

ε - parameters of dist



(x_j, y_j)
ε_i
α
β
x_i

Method of least squares

Objective function / penalty.

$$S(\alpha,\beta) = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$$