# Lab 4

# Binomial test

Alice gave Bob a coin and said it is unbiased. How should Bob verify this?

# Binomial test

Alice gave Bob a coin and said it is unbiased. How should Bob verify this?

He flipped the coin 10 times and got 2 heads. Is this coin really unbiased?

# Binomial test

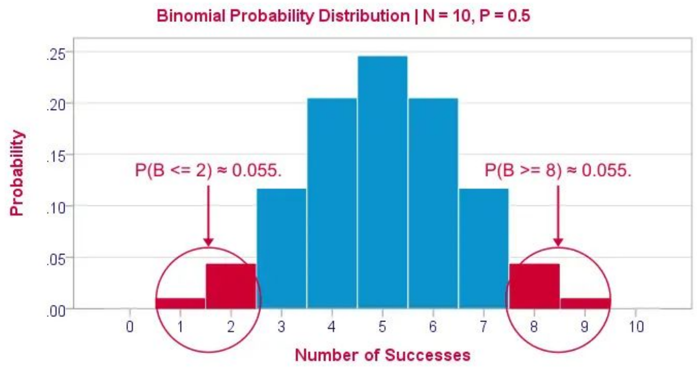
Alice gave Bob a coin and said it is unbiased. How should Bob verify this?

He flipped the coin 10 times and got 2 heads. Is this coin really unbiased?

Test the (null) hypothesis: The coin is unbiased. (p=0.5)

p-value: is the probability of obtaining test results at least as extreme as the result actually observed.

Confidence interval



# Binomial test

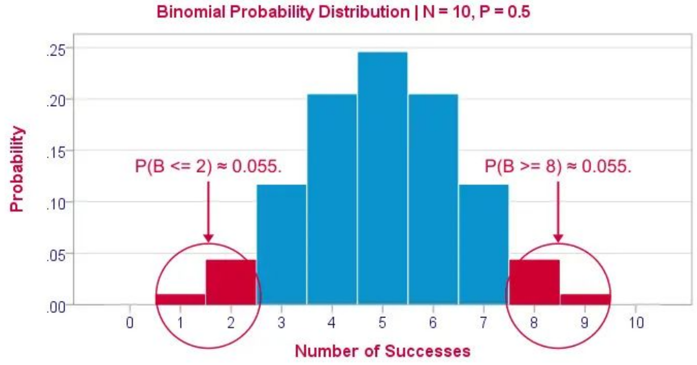
Alice gave Bob a coin and said it is unbiased. How should Bob verify this?

He flipped the coin 10 times and got 2 heads. Is this coin really unbiased?

Test the (null) hypothesis: The coin is unbiased. ($p=0.5$)

p-value: is the probability of obtaining test results at least as extreme as the result actually observed.

5% for hypotesis bound



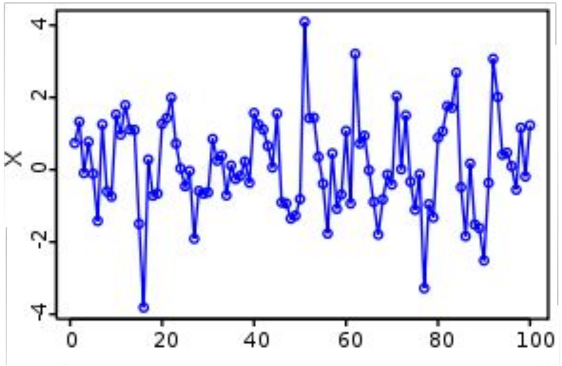

```
>>> scipy.stats.binomtest(2, n=10, p=0.5, alternative='less').pvalue
0.0546875
```

# Autocorrelation

Is there a periodic pattern in X?

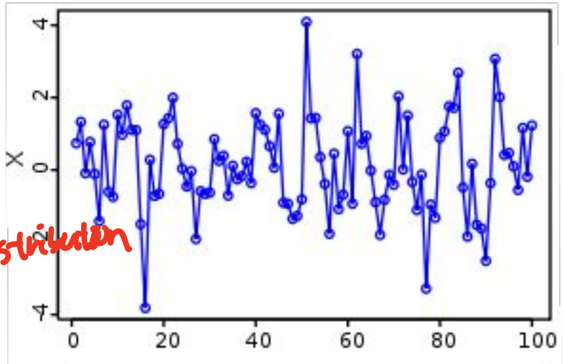
1) Relation between two time series

Correlation



# Autocorrelation

Is there a periodic pattern in X?




This is distribution

Autocorrelation: the similarity between observations of a random variable as a function of the time lag between them.

$$\text{Cov}(X,Y) = E(X-E[X])(Y-E[Y])$$

$$= \sum_{x,y} (x-1)(y-1)/n$$

$$= (-1+0+-1)/3 = -\frac{2}{3}$$

bias

$$\sum_{x,y} (x-1)(y-1)/(n-1) = -1$$

**Example**

Given a sequence X: 0,1,2,1,0. Name Y has the sequence with lag 2: NaN,NaN,0,1,2.
E[X] = 1(discard the first two entries in X), E[Y] = 1
Cov = *sum*(-1, 0, -1) / (3-1) = -1

these two are sample (Empirical)

Variance $\sum (x_i - E[x])^2$

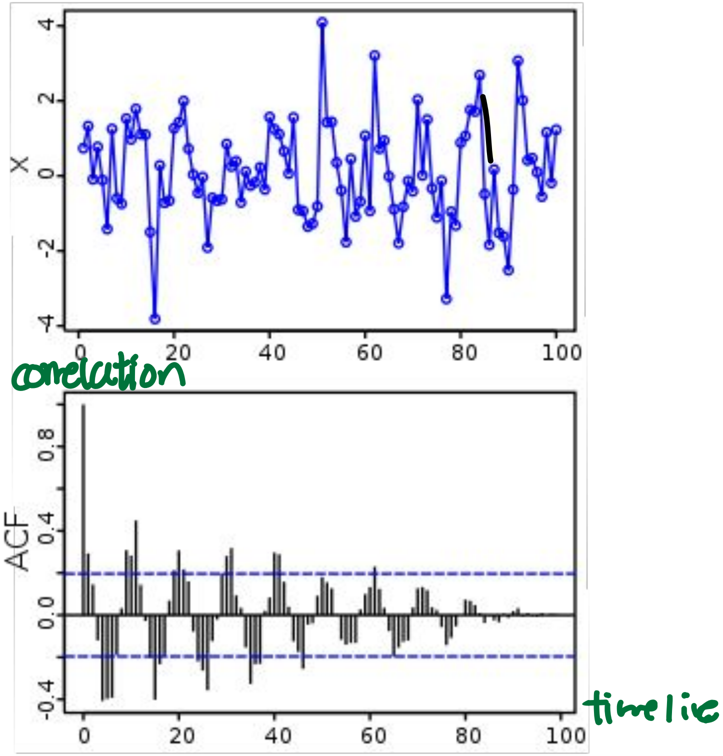
$n-1$

# Autocorrelation

Is there a periodic pattern in X?

Autocorrelation: it is the similarity between observations of a random variable as a function of the time lag between them.

ACF: plot the autocorrelation w.r.t. the time lag.




Every 8 times stamps.

Sending a message

| A | B | C | D |
| --- | --- | --- | --- |
| 00 | 01 | 10 | 11 |

| A | B C D |
| --- | --- |
| 0 | 1 |

# Entropy and Mutual Information

optimal solution

↑ minimize

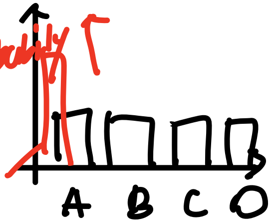
## Entropy

$$H(X) = -\sum_{x \in X} Pr[X = x] \cdot \log_2(Pr[X = x])$$

log(x<1) give you -

probability ↑ Entropy ↓

Probability ↓ Entropy ↑




Interpretation: The minimum expected number of bits needed to transfer the observation.

minimum numbers of bits you use to represent

Example: Compute the entropy of a fair coin toss.

$$H(X) = -(Pr[head] \cdot \log_2(Pr[head]) + Pr[tail] \cdot \log_2(Pr[tail]))$$
$$= -(0.5 \cdot \log_2(0.5) \cdot 2)$$
$$= 1$$

0 or 1.

# Entropy and Mutual Information

A | B C O

0 | 1

$Pr[x, y] \leq Pr_x[x]$

Higher

Mutual information Absolute value

$$I(X;Y) = \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} P_{(X,Y)}(x,y) \log\left(\frac{P_{(X,Y)}(x,y)}{P_X(x)\, P_Y(y)}\right)$$

Interpretation: It measures how much knowing one of these variables reduces uncertainty about the other.

How related between two infos

Properties

A↑ B↓

optimization

maximize connection

- Non-negativity
- Equals 0 if X and Y are independent
- Symmetric, i.e., $I(X;Y) = I(Y;X)$