

- Convergence to the mean

Using Variance.

- Covariance & Correlation.

Rules for expected value:

1. If a, b are constants and X is a random variable then

$$\underline{E(aX + b) = aE(X) + b}$$

2. If X, Y are random variables (dependent or independent)

$$\underline{E(X + Y) = E(X) + E(Y)}$$

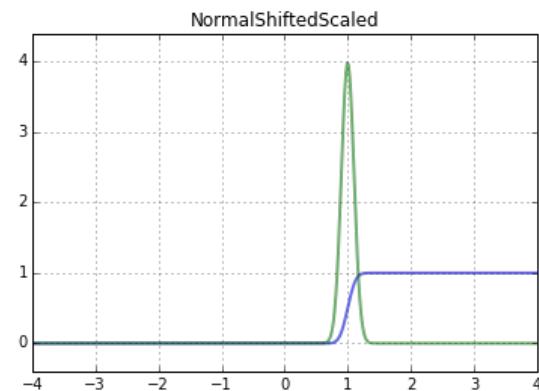
Rules for Variance

1. If a, b are constants and X is a random variable then

$$\underline{\text{Var}(aX + b) = a^2 \text{Var}(X)}$$

2. If X, Y are Independent Random Variables, then

$$\underline{\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)}$$



Shifted and Scaled Normal $N(\mu, \sigma)$

Shift: $\mu = 1$ scale: $\sigma = 0.1$

PDF:

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

CDF:

$$F(x) = \int_{-\infty}^x f(s)ds = 1 - Q\left(\frac{x-\mu}{\sigma}\right)$$

If the RV X is distributed according to $N(\underline{\mu}, \underline{\sigma})$

Then:

$$E(X) = \mu$$

$$\text{Var}(X) = \sigma^2$$

Mean ≠ Average

Mean • $E(X)$ is a property of the distribution, it is not a random variable.

- The average is a random variable:
 - $\text{Average}(x_1, x_2, \dots, x_n) \doteq \frac{1}{n} \sum_{i=1}^n x_i$
- When n is large, the average tends to be close to the mean.

The average

also called the empirical mean

X_i indicates whether there was a click-through on the i th presentation of the Ad.

The X_i are Independent and Identically Distributed Binary Random Variables (**IID Binary RVs**)

$X_i = 1$ indicates there was a click through on the i th presentation of the Ad. Otherwise $X_i = 0$

$$\Pr[X_i = 1] = p, \quad \Pr[X_i = 0] = 1 - p, \quad 0 \leq p \leq 1, \quad p \text{ is the probability of success}$$

$$E[X_i] = 1 \times p + 0 \times (1 - p) = p \quad p \text{ is not a random variable}$$

$$\text{The average is defined to be } S_n \doteq \frac{1}{n} \sum_{i=1}^n X_i \quad S_n \text{ is a random variable}$$

From linearity of expectation we know that

$$E[S_n] = E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \frac{1}{n} \sum_{i=1}^n p = p$$

Next, we consider $\text{Var}(S_n)$

The Binomial Distribution

Exact calculation

Suppose X_1, X_2, \dots, X_n are

independent identically distributed (IID) random variables

$$\Pr[X_i = 1] = p, \quad \Pr[X_i = 0] = 1 - p, \quad 0 \leq p \leq 1$$

We define the average to be another **random variable**

$$S_n \doteq \frac{1}{n} \sum_{i=1}^n X_i \quad \text{What is } \Pr\left(S_n = \frac{m}{n}\right), \quad 0 \leq m \leq n?$$

$$S_n = \frac{m}{n} \quad \text{if and only if for } m \text{ of the } X_i, X_i = 1, \text{ for } n-m \text{ of the } X_i, X_i = 0$$

The probability of each such sequence is: $p^m(1-p)^{n-m}$

The number of such sequences is: $\binom{n}{m}$

$$\Pr\left(S_n = \frac{m}{n}\right) = \binom{n}{m} p^m (1-p)^{n-m} \quad \text{The Binomial distribution.}$$

Alternative derivation for the Binomial distribution

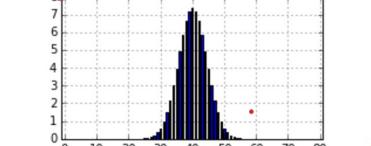
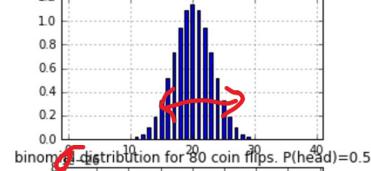
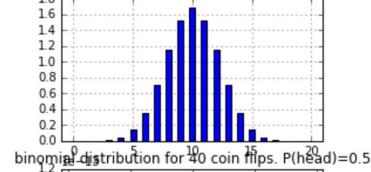
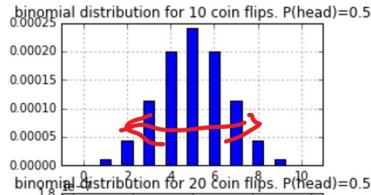
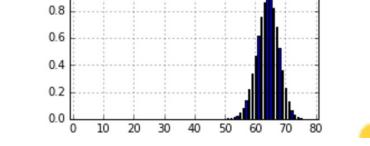
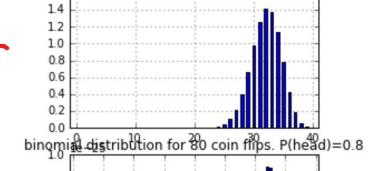
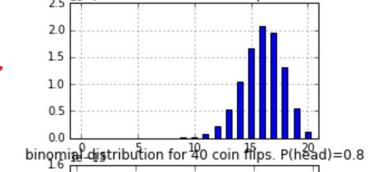
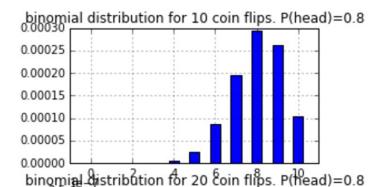
Recall:

$$(a+b)^n = \sum_{i=0}^n \binom{n}{i} a^i b^{n-i}$$

Setting: $a = p, b = (1-p)$

Gives:

$$1 = (p + (1-p))^n = \sum_{i=0}^n \binom{n}{i} p^i (1-p)^{n-i}$$



10

20

40

80

The variance of the average

$$\text{Var}(x) = E((x - \mu)^2)$$
$$E(x^2) - E(x)^2$$
$$p \cdot 1 + (1-p) \cdot 0 > p - p^2 = p(1-p)$$

$S_n \doteq \frac{1}{n} \sum_{i=1}^n X_i$ is the number of click-through's in presentations 1, ..., n

$$\begin{aligned} \text{Var}[X_i] &= p \times (1-p)^2 + (1-p) \times (0-p)^2 = \\ &= p \times (1-p) \times (1-p) + p \times p \times (1-p) = \\ &= p \times (1-p) \times (1-p+p) = p(1-p) \end{aligned}$$

As X_i are IID: $\text{Var}[S_n] = \text{Var}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i] = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n}$

$$\sigma(S_n) = \sqrt{\frac{p(1-p)}{n}}$$

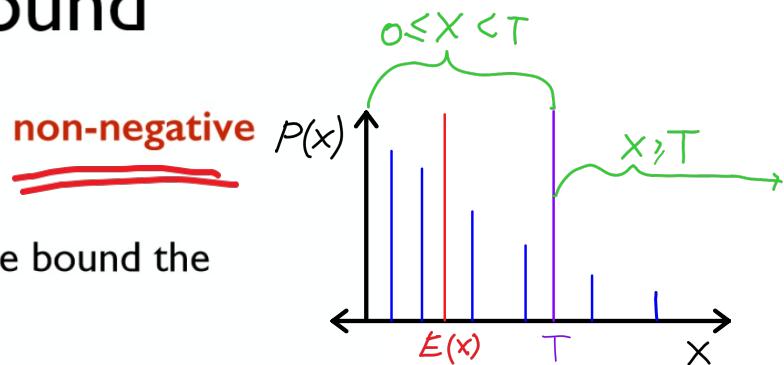
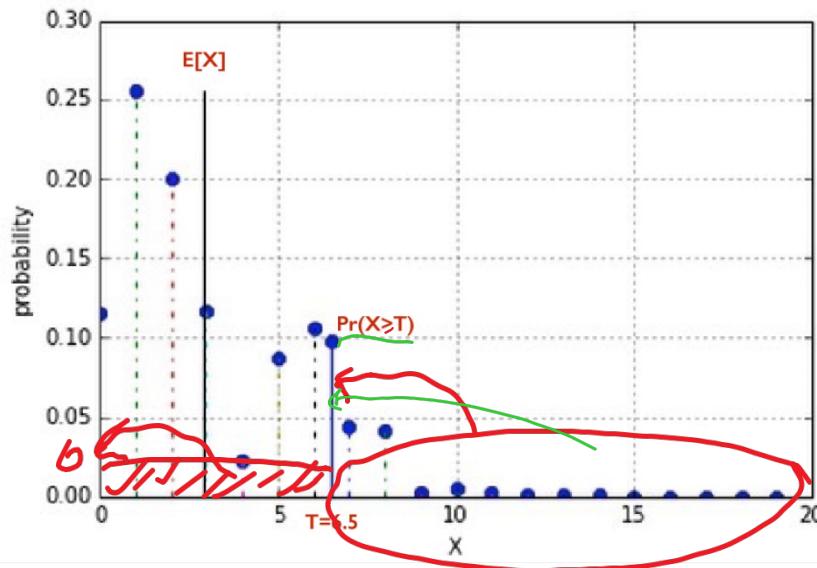
Recall that $\sigma(S_n)$ is proportional to the width of the distribution of S_n

Using the variance to bound the distance from the mean

- Intuition: if the width of the distribution is small then then the probability that the RV is close to the mean is high.
- Proof: to formalize this intuition and make it quantitative we need Chebyshev's bound.
- In order to prove Chebyshev's bound we need Markov bound.
- We'll start with Markov bound.

Markov Bound

- Suppose the RV X is distributed over the **non-negative** integers $0, \dots, 20$
- Suppose we know the mean $E[X]$. Can we bound the probability that $X \geq T$?



$$E[X] = 0 \times \Pr(X < T) + T \times \Pr(X \geq T)$$

$$\Pr(X \geq T) \leq \frac{E(X)}{T}$$

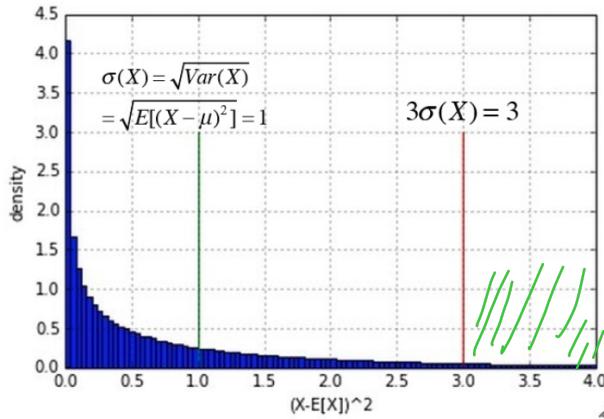
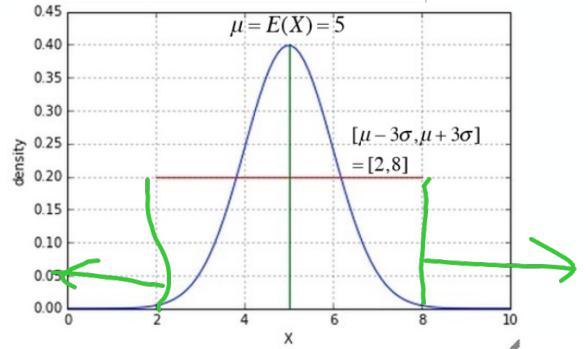
$$E(x), T = K E(x)$$

$$\Pr(X \geq 10) \leq \frac{1}{10}.$$

$$E(x) = \sum p_i x_i$$

$$\sum_{x_i < T} p_i x_i + \sum_{x_i \geq T} p_i x_i \geq T p(x_2)$$

Chebyshev's bound



$$\Pr((X - \mu)^2 \geq \lambda^2) \leq \frac{E[(X - \mu)^2]}{\lambda^2} = \frac{Var(X)}{\lambda^2}$$

Plugging in $\lambda = k\sigma(X)$

$$\Pr(|X - \mu| \geq k\sigma(X)) \leq \frac{\sigma(X)^2}{k^2\sigma(X)^2} = \frac{1}{k^2}$$

In the example shown

$$\mu = E(X) = 5$$

$$\sigma = \sqrt{Var(X)} = 1$$

We choose $k = 3$ to get that

$$\Pr(|X - 5| \geq 3) \leq \frac{1}{k^2} = \frac{1}{9}$$

Applying Chebyshev's bound

$$\Pr[|X - \mu| \geq k\sigma(X)] \leq \frac{\sigma(X)^2}{k^2\sigma(X)^2} = \frac{1}{k^2}$$

A few slides ago, we found that

$$\mu(S_n) = p; \quad \sigma(S_n) = \sqrt{\frac{p(1-p)}{n}}$$

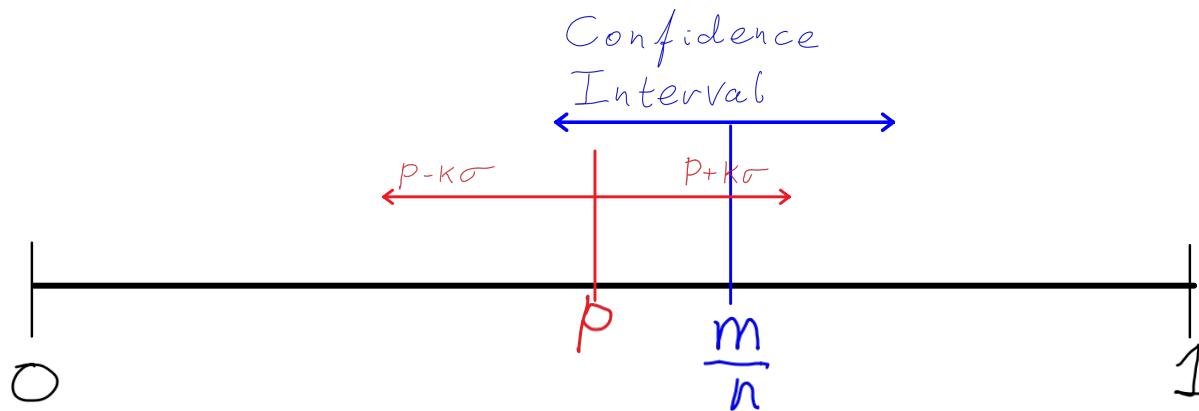
$$\Pr[|S_n - p| \geq k\sqrt{\frac{p(1-p)}{n}}] \leq \frac{1}{k^2}$$

fixing k and letting n increase we see that doubling n decreases
the distance from the mean by $\sqrt{2}$

Confidence Intervals

Suppose we flip a coin n times, and get m heads $n-m$ tails

What can we say about $p = \text{Prob}(\text{heads})$?

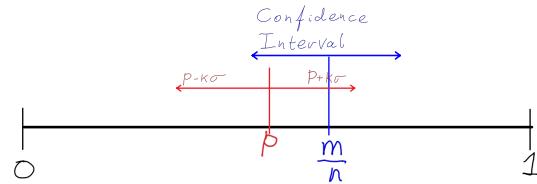


a 99% confidence interval contains all values of p

that assign $\frac{m \text{ heads}}{\text{out of } n \text{ flips}}$ prob. at least 1%

\equiv excludes all values of p

that assign $\frac{m \text{ heads}}{\text{out of } n \text{ flips}}$ prob. less than 1%



chebychev's bound

$$P(|X-\mu| > K\sigma) \leq \frac{1}{K^2}$$

For our random Variable we use:

$$X = \frac{m}{n} ; \mu = p$$

$$\sigma = \sqrt{\frac{p(1-p)}{n}} \leq \sqrt{\frac{\frac{1}{2}(1-\frac{1}{2})}{n}} = \frac{1}{2\sqrt{n}}$$

plugging μ, σ, X into chebychev we get

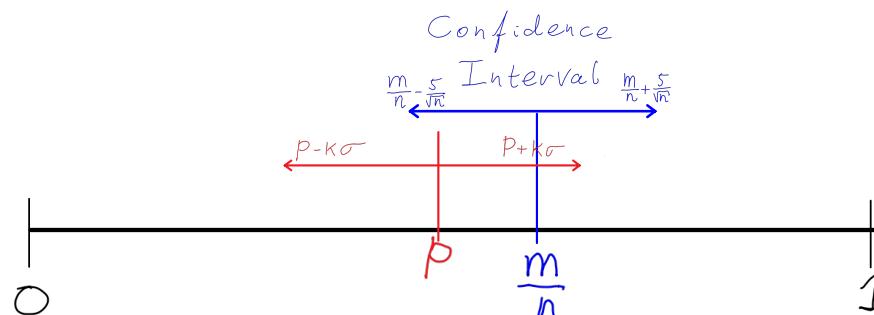
$$P\left(\left|\frac{m}{n} - p\right| \geq \frac{K}{2\sqrt{n}}\right) \leq \frac{1}{K^2}$$

as we want 99% confidence

$$\text{we need } \frac{1}{K^2} \leq 0.01 \Rightarrow K=10$$

and we finally get

$$P\left(\left|\frac{m}{n} - p\right| \geq \frac{5}{\sqrt{n}}\right) \leq \frac{1}{100}$$



The Goal of Statistics

True (or underlying) Distribution

Vs
Empirical Distribution

We have a biased Coin $P(\text{Head}) = \frac{2}{3} = 0.666\dots$

We flip the Coin 100 times and get $\frac{60}{100}$ heads

The empirical distribution is $\hat{P}(\text{Head}) = 0.6$

The empirical distribution is a Random Variable

The true distribution is NOT a Random Variable

Using the empirical distribution We can define:

- Empirical mean = average
- Empirical Variance
- Empirical prob. of events.

As the number of Samples $\rightarrow \infty$

We can estimate the underlying distribution more and more accurately.

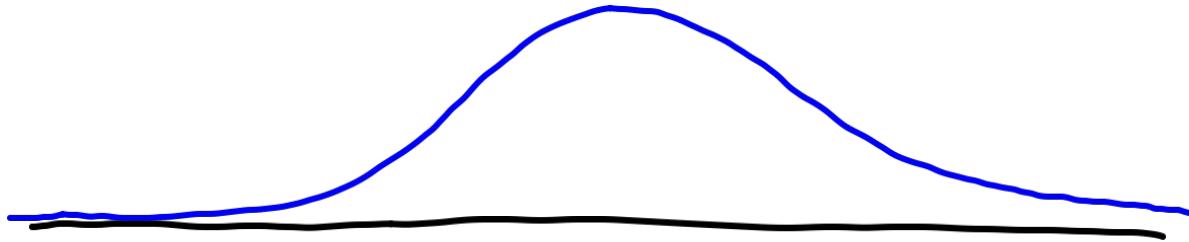
As the number of samples (coin flips) increases

The empirical Converges to the True

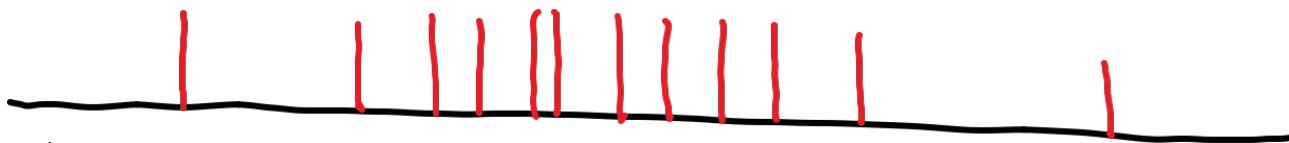
The Question is **How Fast?**

Empirical Dist. Over the Reals

Suppose the true dist. is a density dist over the reals

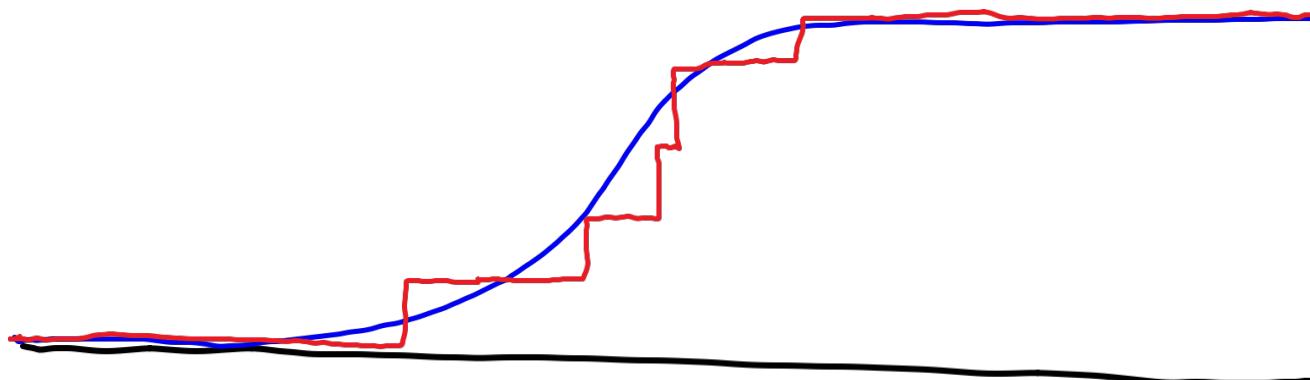


The empirical dist is a point-mass distribution



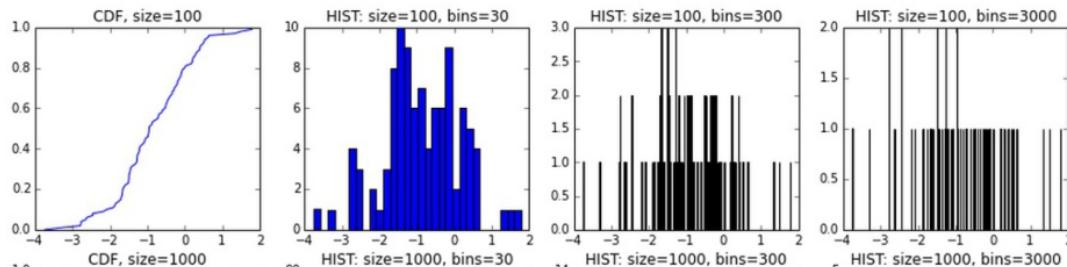
How can the empirical Converge to the density?

Answer: Consider the CDF

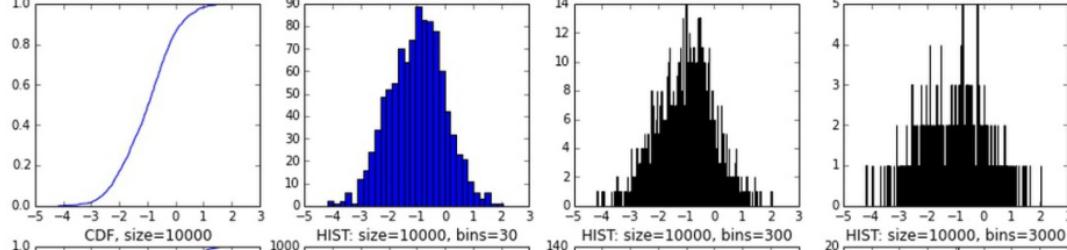


$N(-1, 1) = \text{A normal distribution centered at } -1, \text{ with width 1}$

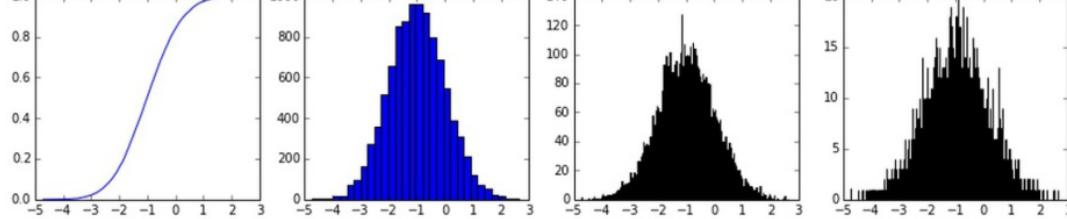
(60)



(100)



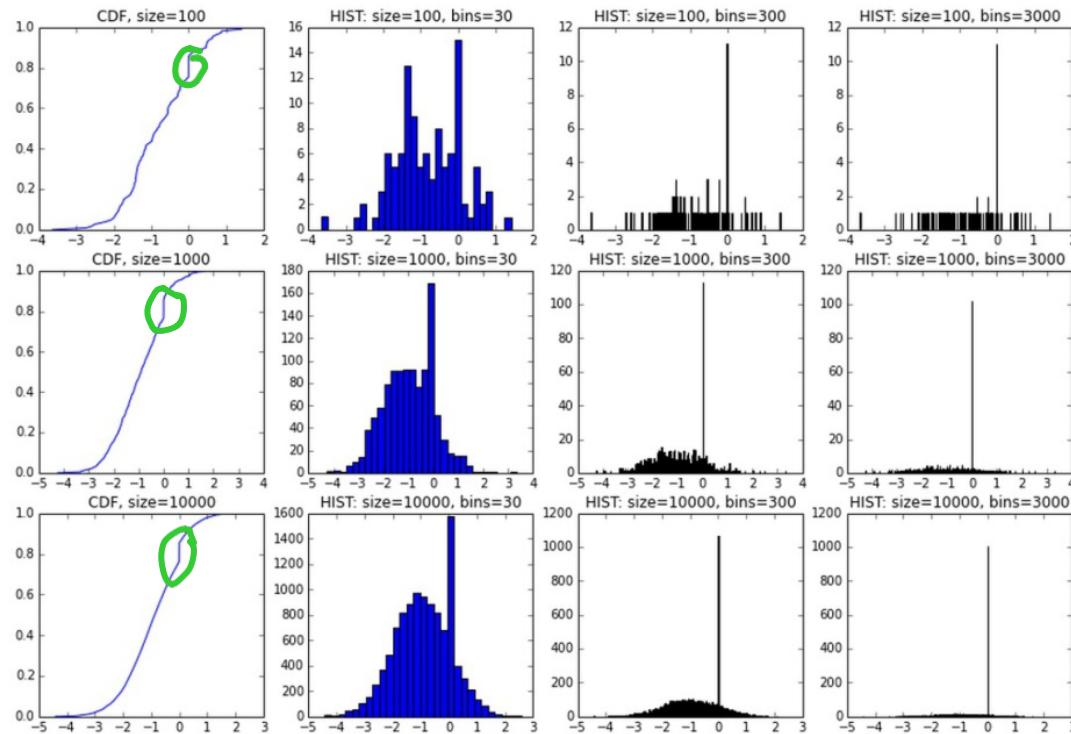
10,000



300

3000

**A mixture of the normal and a point-mass
($10^*N(-1,1) + PM(0)$)**



Testing dependence

of RVs using

Empirical distribution

= DATA

Independent random variables

$$\forall x, y \quad P(X = x \wedge Y = y) = P(X = x)P(Y = y)$$

Independent random variables

$$\forall x, y \quad P(X = x \wedge Y = y) = P(X = x)P(Y = y)$$

	X=0 0.2	X=1 0.1	X=2 0.7
Y=0 0.4	0.08	0.04	0.28
Y=1 0.6	0.12	0.06	0.42

Independent random variables

$$\forall x, y \quad P(X = x \wedge Y = y) = P(X = x)P(Y = y)$$

marginal distributions		X=0	X=1	X=2
Y=0	0.4	0.08	0.04	0.28
	0.6	0.12	0.06	0.42

Do grades depend on gender?

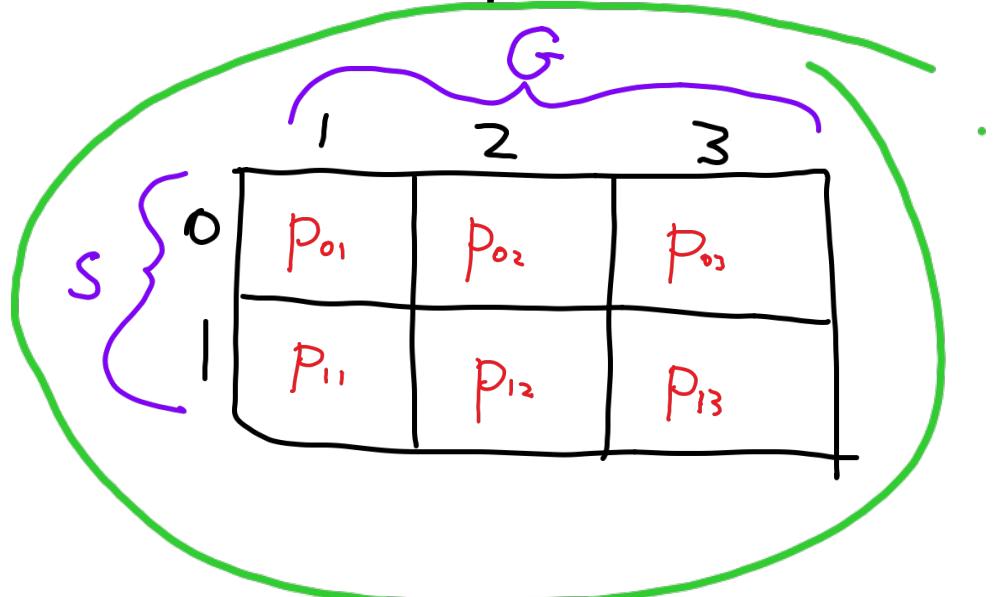
Ω = all students

Define two random Variables

$S: \Omega \rightarrow \{0, 1\}$ (male, female)

$G: \Omega \rightarrow \{1^A, 2^B, 3^C\}$ (grade average)

Are S, G dependent or independent?



n=100
std~1/10

Empirical contingency tables

independent

dependent

true dist

empirical
dist 1

empirical
dist 2

empirical
dist 3

	marginal	A-average	B-average	C-average
marginal	1	0.2500	0.5000	0.2500
Male	0.4000	0.1000	0.2000	0.1000
Female	0.6000	0.1500	0.3000	0.1500

	marginal	A-average	B-average	C-average
marginal	1	0.2600	0.4800	0.2600
Male	0.3900	0.1000	0.2100	0.0800
Female	0.6100	0.1600	0.2700	0.1800

	marginal	A-average	B-average	C-average
marginal	1	0.2200	0.3900	0.3900
Male	0.4400	0.1100	0.1400	0.1900
Female	0.5600	0.1100	0.2500	0.2000

	marginal	A-average	B-average	C-average
marginal	1	0.2200	0.5000	0.2800
Male	0.3500	0.1000	0.1200	0.1300
Female	0.6500	0.1200	0.3800	0.1500

	marginal	A-average	B-average	C-average
marginal	1.0000	0.3000	0.4000	0.3000
Male	0.5000	0.1000	0.2000	0.2000
Female	0.5000	0.2000	0.2000	0.1000

	marginal	A-average	B-average	C-average
marginal	1	0.2700	0.5200	0.2100
Male	0.4800	0.1200	0.2200	0.1400
Female	0.5200	0.1500	0.3000	0.0700

	marginal	A-average	B-average	C-average
marginal	1	0.2600	0.4300	0.3100
Male	0.4700	0.0700	0.2100	0.1900
Female	0.5300	0.1900	0.2200	0.1200

	marginal	A-average	B-average	C-average
marginal	1	0.1800	0.4000	0.4200
Male	0.5500	0.1000	0.1600	0.2900
Female	0.4500	0.0800	0.2400	0.1300

n=10,000
std~1/100

Empirical contingency tables

independent

dependent

true dist

	marginal	A-average	B-average	C-average
marginal	1	0.2500	0.5000	0.2500
Male	0.4000	0.1000	0.2000	0.1000
Female	0.6000	0.1500	0.3000	0.1500

empirical
dist 1

	marginal	A-average	B-average	C-average
marginal	1	0.2473	0.5062	0.2465
Male	0.4021	0.0977	0.2052	0.0992
Female	0.5979	0.1496	0.3010	0.1473

empirical
dist 2

	marginal	A-average	B-average	C-average
marginal	1	0.2530	0.4943	0.2527
Male	0.3925	0.0982	0.1942	0.1001
Female	0.6075	0.1548	0.3001	0.1526

empirical
dist 3

	marginal	A-average	B-average	C-average
marginal	1	0.2457	0.5005	0.2538
Male	0.3893	0.0936	0.1945	0.1012
Female	0.6107	0.1521	0.3060	0.1526

	marginal	A-average	B-average	C-average
marginal	1.0000	0.3000	0.4000	0.3000
Male	0.5000	0.1000	0.2000	0.2000
Female	0.5000	0.2000	0.2000	0.1000

	marginal	A-average	B-average	C-average
marginal	1.0000	0.3000	0.4064	0.2936
Male	0.4991	0.0958	0.2052	0.1981
Female	0.5009	0.2042	0.2012	0.0955

	marginal	A-average	B-average	C-average
marginal	1.0000	0.3058	0.3895	0.3047
Male	0.5044	0.1023	0.1989	0.2032
Female	0.4956	0.2035	0.1906	0.1015

	marginal	A-average	B-average	C-average
marginal	1	0.2984	0.4047	0.2969
Male	0.4970	0.0997	0.1938	0.2035
Female	0.5030	0.1987	0.2109	0.0934

Empirical contingency tables

n=1,000,000
std~1/1,000

independent

dependent

true dist

	marginal	A-average	B-average	C-average
marginal	1	0.2500	0.5000	0.2500
Male	0.4000	0.1000	0.2000	0.1000
Female	0.6000	0.1500	0.3000	0.1500

empirical
dist 1

	marginal	A-average	B-average	C-average
marginal	1	0.2500	0.4998	0.2502
Male	0.3998	0.1001	0.2000	0.0997
Female	0.6002	0.1499	0.2998	0.1505

empirical
dist 2

	marginal	A-average	B-average	C-average
marginal	1	0.2504	0.4997	0.2499
Male	0.3991	0.1000	0.1993	0.0999
Female	0.6009	0.1504	0.3004	0.1500

empirical
dist 3

	marginal	A-average	B-average	C-average
marginal	1	0.2504	0.5000	0.2497
Male	0.4005	0.1005	0.2003	0.0996
Female	0.5995	0.1498	0.2997	0.1500

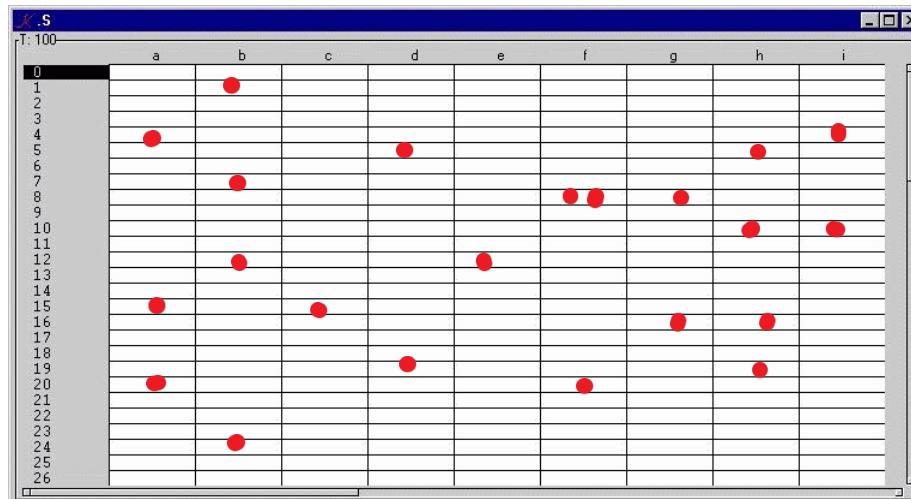
	marginal	A-average	B-average	C-average
marginal	1.0000	0.3000	0.4000	0.3000
Male	0.5000	0.1000	0.2000	0.2000
Female	0.5000	0.2000	0.2000	0.1000

	marginal	A-average	B-average	C-average
marginal	1	0.2997	0.4000	0.3002
Male	0.4997	0.0994	0.2000	0.2003
Female	0.5003	0.2003	0.2000	0.1000

	marginal	A-average	B-average	C-average
marginal	1	0.3006	0.3996	0.2997
Male	0.5002	0.1000	0.2001	0.2001
Female	0.4998	0.2006	0.1995	0.0997

	marginal	A-average	B-average	C-average
marginal	1	0.3001	0.3998	0.3001
Male	0.4995	0.1000	0.1998	0.1997
Female	0.5005	0.2001	0.2000	0.1004

The Data sparsity problem



Suppose $X: \Omega \rightarrow \{1-9\}$ ($a-i$)

$Y: \Omega \rightarrow \{1-26\}$

Are X, Y dependent?

We will need ALOT of data!

If $X: \Omega \rightarrow [0, 1]$

$Y: \Omega \rightarrow [0, 1]$

We will NEVER have enough data

The Covariance approach to finding dependencies

Instead of computing a product for
each entry in the Contingency table

We compute a single number for the whole table.

Cov

If **Cov** far from zero: X and Y are dependent.
close to zero: We can't say anything

What we gain:

1. faster Convergence.
2. less Computation.

What we lose:

* Some dependencies will not be detected

Sum of two random variables

- The sum of two random variables is a random variable:

$$S(\omega) = X(\omega) + Y(\omega)$$

- The expected value is defined to be:

$$E(X) \doteq \sum_{i=1}^n X(\omega_i)P(\omega_i), \quad E(Y) \doteq \sum_{i=1}^n Y(\omega_i)P(\omega_i), \quad E(S) \doteq \sum_{i=1}^n S(\omega_i)P(\omega_i),$$

- We can prove the relation $E(X + Y) = E(X) + E(Y)$ in the following way:

- $E(X + Y) = E(S) =$

$$= \sum_{i=1}^n S(\omega_i)P(\omega_i) = \sum_{i=1}^n (X(\omega_i) + Y(\omega_i))P(\omega_i) =$$

$$= \sum_{i=1}^n X(\omega_i)P(\omega_i) + \sum_{i=1}^n Y(\omega_i)P(\omega_i) = E(X) + E(Y)$$

Product of two random variables

- The product of two random variables is a random variable:

$$M(\omega) = X(\omega) \times Y(\omega)$$

- The expected value is defined to be:

$$E(X) \doteq \sum_{i=1}^n X(\omega_i)P(\omega_i), \quad E(Y) \doteq \sum_{i=1}^n Y(\omega_i)P(\omega_i), \quad E(M) \doteq \sum_{i=1}^n M(\omega_i)P(\omega_i),$$

- Lets analyze $E(XY)$:

$$E(XY) = E(M) = \sum_{i=1}^n M(\omega_i)P(\omega_i) = \sum_{i=1}^n (X(\omega_i)Y(\omega_i))P(\omega_i) = *$$

- $P(\omega_i)$ can be replaced by $P(X(\omega) = x \text{ and } Y(\omega) = y)$. By summing over the possible values of x, y rather than over the outcomes ω_i . Using these 3 observations we can rewrite the last expression as

$$* = \sum_{x,y} xy P(X(\omega) = x \text{ and } Y(\omega) = y) = \#$$

- If $X(\omega), Y(\omega)$ are independent random variables then

$$P(X(\omega) = x \text{ and } Y(\omega) = y) = P(X(\omega) = x) \times P(Y(\omega) = y)$$

- Which implies:

$$\begin{aligned} \# &= \sum_{x,y} xy P(X(\omega) = x \text{ and } Y(\omega) = y) = \sum_{x,y} xy P(X(\omega) = x) P(Y(\omega) = y) = \\ &= \sum_x x P(X(\omega) = x) \times \sum_y y P(Y(\omega) = y) = E(X)E(Y) \end{aligned}$$

Expected value for a product of independent RVs

$$E(XY) = \sum_x \sum_y xy P(X = x \wedge Y = y) =$$

Expected value for a product of independent RVs

$$\begin{aligned} E(XY) &= \sum_x \sum_y xy P(X = x \wedge Y = y) = \\ &= \sum_x \sum_y xy P(X = x)P(Y = y) = \sum_x x P(X = x) \sum_y y P(Y = y) = \end{aligned}$$

Expected value for a product of independent RVs

$$\begin{aligned} E(XY) &= \sum_x \sum_y xy P(X = x \wedge Y = y) = \\ &= \sum_x \sum_y xy P(X = x)P(Y = y) = \sum_x x P(X = x) \sum_y y P(Y = y) = \\ &= E(X)E(Y) \end{aligned}$$

Covariance

Recall $\mu_X \doteq E(X)$, $\mu_Y \doteq E(Y)$

$$\begin{aligned}\text{Cov}(X, Y) &\doteq E((X - \mu_X)(Y - \mu_Y)) = \\&= E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X \mu_Y = E(XY) - E(X)E(Y)\end{aligned}$$

Recall $\text{Var}(X) = E(X^2) - E(X)^2 = \text{Cov}(X, X)$

Cov(X,Y)≠0 implies that **X** and **Y** are not independent
but **Cov(X,Y)=0** does not imply that **X** and **Y** are independent

Why do we need Corr in addition to Cov ?

- Suppose $\text{Cov}(X,Y)=3$ and $\text{Cov}(X,Z)=1$
- Is X more correlated with Y than with Z?
- Not necessarily, it might be that $Y=3Z$
- We want to have a measure of correlation that is independent of scaling, i.e.

$$\forall a > 0, b > 0: \text{Corr}(aX, bY) = \text{Corr}(X, Y)$$

Correlation coefficient

$$\text{Corr}(X, Y) \doteq \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

- $\text{Corr}(aX+c, bY+d) = \text{Corr}(X, Y)$ if $a, b > 0$
- $\text{Corr}(X, Y)$ varies from -1 to $+1$
- $\text{Corr}(X, Y) > 0 \Leftrightarrow X$ and Y are “correlated”
- $\text{Corr}(X, Y) < 0 \Leftrightarrow X$ and Y are “anti-correlated”
- $\text{Corr}(X, Y) = 1 \Leftrightarrow X = aY, a > 0$
- $\text{Corr}(X, Y) = -1 \Leftrightarrow X = aY, a < 0$

Correlation coefficient

$$\text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}$$

- The covariance depends on scaling and units, correlation coefficient does not

$$\forall a > 0, b > 0 \quad \text{Corr}(aX, bY) = \text{Corr}(X, Y)$$

- The correlation coefficient varies between -1 and 1.

correlation vs. dependence

- If X, Y are independent then $\text{Cov}(X, Y) = 0$ and $\text{Corr}(X, Y) = 0$
- If $\text{Corr}(X, Y) \neq 0$ then X, Y are dependent.
- No implications in the opposite directions
- If X is a random variable that takes n discrete values, and Y is a random variable that takes m discrete values,
Then checking for independence requires checking nm values.
- Checking for zero correlation requires calculation of just one value.
- Correlation is the quick and dirty way to detect strong dependencies, but it cannot find them all.

Examples

	X=1	X=2	X=3	X=4
Y=1	1/4	1/4	0	0
Y=2	0	0	0	0
Y=3	0	0	1/4	1/4

$$\mu(X) = 2.5, \mu(Y) = 2$$

$$\text{cov}(X,Y) = \frac{1}{4}(-1.5 * -1) + \frac{1}{4}(-.5 * -1) + \frac{1}{4}(.5 * 1) + \frac{1}{4}(1.5 * 1) = 1$$

	X=1	X=2	X=3	X=4
Y=1	0	0	0	1/4
Y=2	0	1/4	1/4	0
Y=3	1/4	0	0	0

$$\mu(X) = 2.5, \mu(Y) = 2$$

$$\text{cov}(X,Y) = \frac{1}{4}(-1.5 * 1) + \frac{1}{4}(-.5 * 0) + \frac{1}{4}(.5 * 9) + \frac{1}{4}(1.5 * -1) = -\frac{3}{4}$$

	X=1	X=2	X=3	X=4
Y=1	1/4	0	0	1/4
Y=2	0	0	0	0
Y=3	1/4	0	0	1/4

$$\mu(X) = 2.5, \mu(Y) = 2$$

$$\text{cov}(X,Y) = \frac{1}{4}(-1.5 * 1) + \frac{1}{4}(-1.5 * -1) + \frac{1}{4}(1.5 * 1) + \frac{1}{4}(1.5 * -1) = 0$$

$$P(X=1)=P(X=4)=1/2, P(Y=1)=P(Y=3)=1/2$$

X and Y are independent because all of the joint probabilities are either 0 or 1/4

	X=1	X=2	X=3	X=4
Y=1	1/8	0	0	1/8
Y=2	0	1/4	1/4	0
Y=3	1/8	0	0	1/8

1. $\text{Cov}(X,Y)=0$
2. X and Y are independent

A. 1 and 2 B. 1 and not 2 C. not 1 and 2 D. not 1 and not 2

