STATS 100A: Two or More Random Variables

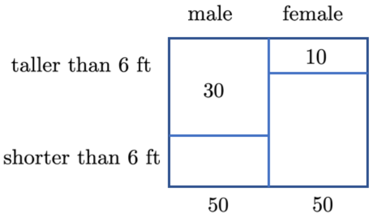
Ying Nian Wu

Department of Statistics
University of California, Los Angeles

Some pictures are taken from the internet.
Credits belong to original authors.

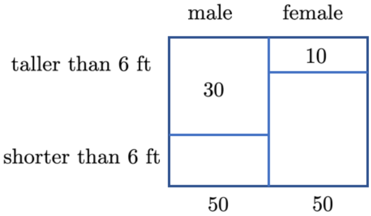
Population

Recall Example 2 in Part 1: Sample a random person from a population of 100 people, 50 males and 50 females. 30 males are taller than 6 ft, 10 females are taller than 6 ft.



Population proportion

Example 2: A male, B tall.



$$P(A) = \frac{|A|}{|\Omega|} = \frac{50}{100} = 50\%.$$

$$P(B) = \frac{|B|}{|\Omega|} = \frac{30+10}{100} = 40\%.$$

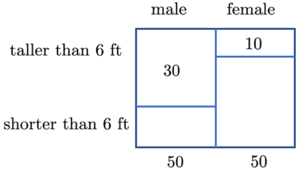
$$P(A \cap B) = \frac{|A \cap B|}{|\Omega|} = \frac{30}{100} = 30\%.$$

Probability = population proportion.

Population proportion

Experiment → outcome → number

Example 2: A male, B tall.



$$P(A|B) = \frac{|A \cap B|}{|B|} = \frac{30}{40} = 75\%.$$

Among tall people, what is the proportion of males?

$$P(B|A) = \frac{|A \cap B|}{|A|} = \frac{30}{50} = 60\%.$$

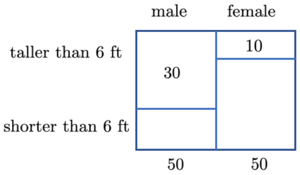
Among males, what is the proportion of tall people?

Conditional probability = proportion within sub-population.

Joint distribution

Example 2: $X \in \{$ male, female$\}$, $Y \in \{$ tall, short$\}$.

$$p(x, y) = P(X = x, Y = y).$$

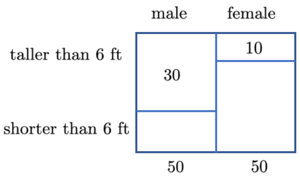


$$p(m, t) = .3; p(m, s) = .2; p(f, t) = .1; p(f, s) = .4.$$

Marginal distribution

Example 2: $X \in \{$ male, female$\}$, $Y \in \{$ tall, short$\}$.

$$p(x, y) = P(X = x, Y = y).$$



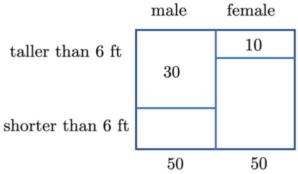
$$P(X = x) = p(x) = \sum_y p(x, y).$$

$$P(Y = y) = p(y) = \sum_x p(x, y).$$

Conditional distribution

Example 2: $X \in \{$ male, female$\}$, $Y \in \{$ tall, short$\}$.

$$p(x, y) = P(X = x, Y = y).$$



$$P(X = x|Y = y) = p(x|y) = p(x, y)/p(y).$$

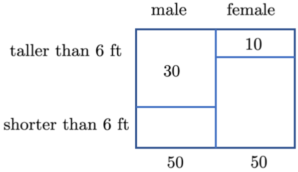
$$P(Y = y|X = x) = p(y|x) = p(x, y)/p(x).$$

Chain rule: $p(x, y) = p(x)p(y|x) = p(y)p(x|y)$.

Rule of total probability

Example 2: $X \in \{$ male, female$\}$, $Y \in \{$ tall, short$\}$.

$$p(x, y) = P(X = x, Y = y).$$

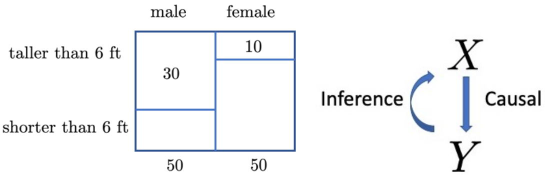


$$p(y) = \sum_x p(x, y) = \sum_x p(x)p(y|x).$$

Bayes rule

Example 2: $X \in \{$ male, female$\}$, $Y \in \{$ tall, short$\}$.

$$p(x, y) = P(X = x, Y = y).$$



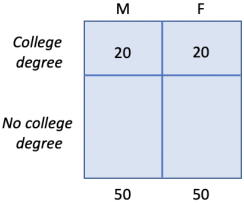
$$p(x|y) = \frac{p(x, y)}{p(y)} = \frac{p(x)p(y|x)}{\sum_{x'} p(x')p(y|x')}.$$

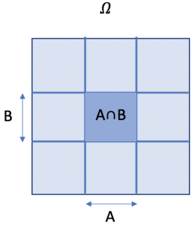
Independence

$$P(A|B) = P(A).$$

$$P(A \cap B) = P(A)P(B).$$

$X \in \{$ male, female$\}$, $Y \in \{$ college, not$\}$





$p(y|x) = p(y).$
$p(x, y) = p(x)p(y|x) = p(x)p(y).$

Reasoning

Recall Example 6: Rare disease example

1% of population has a rare disease.

A random person goes through a test.

If the person has disease, 90% chance test positive.

If the person does not have disease, 90% chance test negative.

If tested positive, what is the chance he or she has disease?

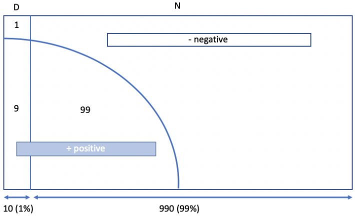
$P(D) = 1\%$.

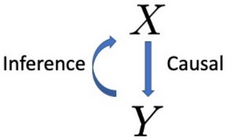
$P(+|D) = 90\%$, $P(-|N) = 90\%$.

$P(D|+) =?$

$X \in \{D, N\}$. $Y \in \{+, -\}$.

Example 6: Rare disease example





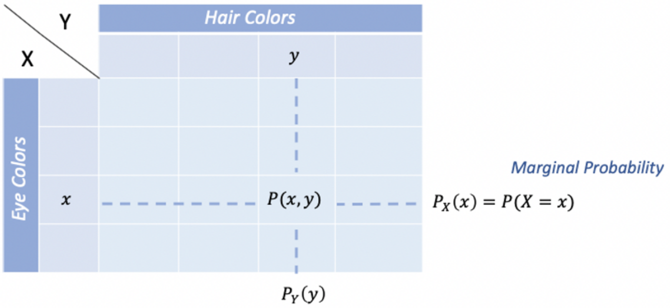
$P(D|+) = \frac{9}{9+99} = \frac{1}{12}$.

$$p(x|y) = \frac{p(x,y)}{p(y)} = \frac{p(x)p(y|x)}{\sum_{x'} p(x')p(y|x')}.$$

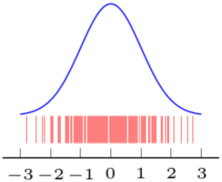
$p(x)$: prior belief. $p(x|y)$: posterior belief.

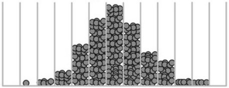
Discrete joint, marginal, conditional

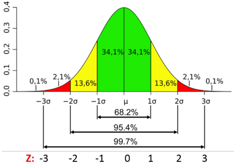
X = eye color, Y = hair color.



Continuous



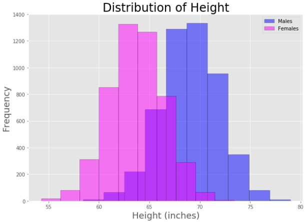




$$P(X \in (x, x+\Delta x)) = f(x)\Delta x.$$
$$f(x) = P(X \in (x, x+\Delta x))/\Delta x.$$

Two continuous random variables

$X =$ height, $Y =$ weight.



STATS 100A HW2

Problem 1 Consider a random walk on integers. We start from $X_0 = 0$, and at each step, we flip a fair coin. If it is head, we move forward by 1, and if it is tail, we move backward by 1. In math notation, $X_{t+1} = X_t + \epsilon_t$, where $\epsilon_t = 1$ with probability 1/2, and $\epsilon_t = -1$ with probability 1/2.

(1) At time $t = 5$, what are the possible values of X_t?

(2) What is the probability of each possible value in (1)?

(3) If 1 million people do the random walk independently, all starting from 0 at $t = 0$. At $t = 5$, what is the distribution of these 1 million people?

Problem 2 Suppose a person does random walk over 2 states, 1 and 2, starting from state 1. At each step, the person will stay with probability 1/3, or move to the other state with probability 2/3. Let X_t be the state of the person at step t, with $X_0 = 1$.

(1) Calculate $P(X_t = 1)$ and $P(X_t = 2)$ for $t = 1, 2, 3, 4$.

(2) Let $K_{ij} = P(X_{t+1} = j | X_t = i)$. Let K be the 2×2 matrix. Write down K.

(3) Let $p^{(t)} = (P(X_t = 1), P(X_t = 2))$. Prove $p^{(t+1)} = p^{(t)}K$, and $p^{(t)} = p^{(0)}K^t$.

(4) If 1 million people do the random walk independently, all starting from 1 at $t = 0$. At $t = 4$, what is the distribution of these 1 million people?

Problem 3 Suppose at any moment, the probability of fire in a classroom is α. Suppose the conditional probability of alarm given fire is β, and the conditional probability of alarm given no fire is γ.

(1) Calculate the marginal probability of alarm using the rule of total probability.

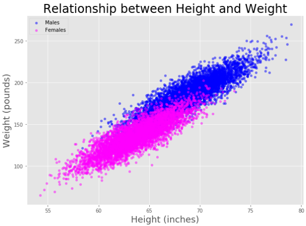
(2) Calculate the conditional probability of fire given alarm using the Bayes rule.

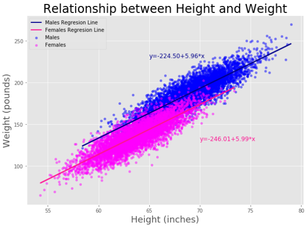
(3) Suppose we repeat the experiment 10,000 times. Suppose $\alpha = 1\%$, $\beta = 99\%$, and $\gamma = 2\%$. Then on average, how may times are there fire? how many times are there alarm? how many times are there false alarm? When there is alarm, how often is it true alarm?

1

Two continuous random variables

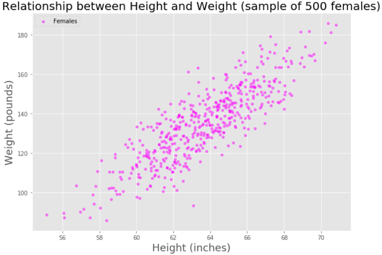
$X = \text{height}$, $Y = \text{weight}$.





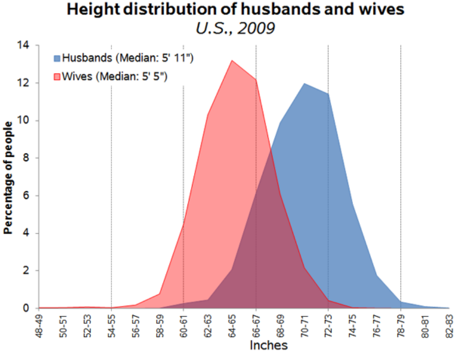
Two continuous random variables

$X =$ height, $Y =$ weight.



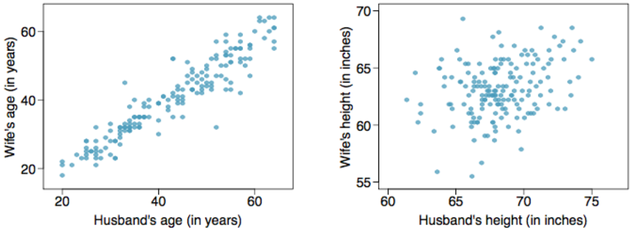
Two continuous random variables

X = husband, Y = wife.



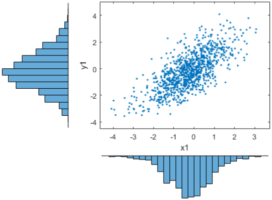
Two continuous random variables

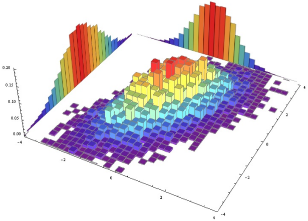
X = husband, Y = wife.



Continuous joint density

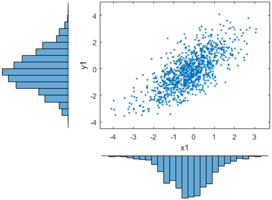
$(X, Y) \sim f(x, y)$.

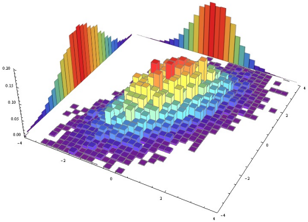




$$P(X \in (x, x + \Delta x), Y \in (y, y + \Delta y) = f(x, y)\Delta x \Delta y.$$
$$f(x, y) = P(X \in (x, x + \Delta x), Y \in (y, y + \Delta y)/\Delta x \Delta y.$$

Marginal density





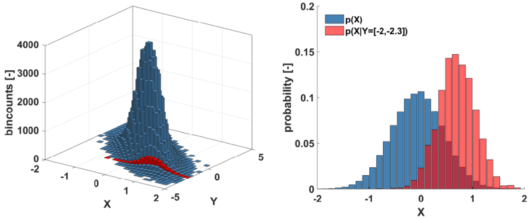
$$P(X \in (x, x + \Delta x)) = f(x)\Delta x.$$

$$P(X \in (x, x+\Delta x)) = \sum_y P(X \in (x, x+\Delta x), Y \in (y, y+\Delta y)).$$

$$f(x)\Delta x = \sum_y f(x, y)\Delta x \Delta y.$$

$$f(x) = \sum_y f(x, y)\Delta y = \int f(x, y) dy.$$

Conditional density



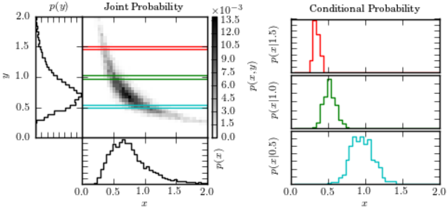
$$P(X \in (x, x+\Delta x) | Y \in (y, y+\Delta y)) = f(x|y)\Delta x.$$

$$P(X \in (x, x+\Delta x) | Y \in (y, y+\Delta y)) = \frac{P(X \in (), Y \in ())}{P(Y \in ())}$$

$$f(x|y)\Delta x = \frac{f(x,y)\Delta x \Delta y}{f(y)\Delta y}.$$

$$f(x|y) = \frac{f(x,y)}{f(y)}.$$

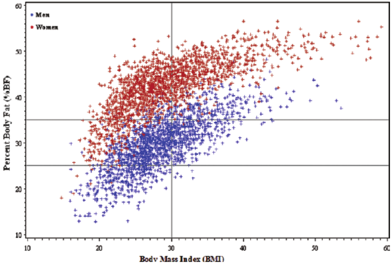
Conditional density



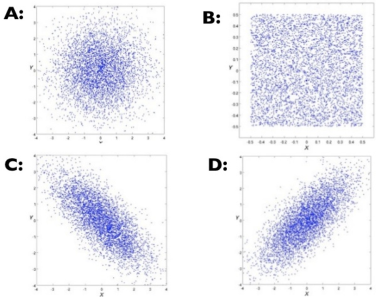
$$f(x|y) = \frac{f(x,y)}{f(y)}.$$

Chain rule: $f(x,y) = f(y)f(x|y)$.

Density

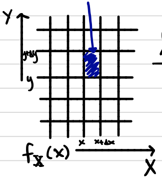

Men
Women
Percent Body Fat (%BF)
Body Mass Index (BMI)
60
50
40
30
20
10
10
20
30
40
50
60

Density



Density

$f(x,y)\cdot\Delta x\cdot\Delta y$



$$\int\cdots dx \longrightarrow f_Y(y) = \int f(x,y)\,dx \xrightarrow{\text{compare}} \boxed{P_X(x) = \sum_y p(x,y)}$$

$$P_Y(y) = \sum_x p(x,y)$$

$$\boxed{f_Y(y)\cdot\Delta y = \sum_{\substack{x\\ \text{bins}}} f(x,y)\cdot\Delta x\cdot\Delta y}$$

$$f_X(x) = \int f(x,y)\,dy$$

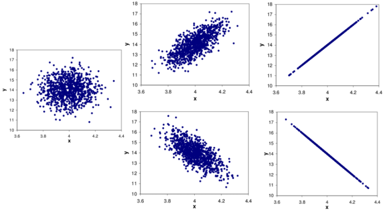
(from before) $\boxed{\dfrac{P(x,y)}{P_Y(y)}}$

form

essence

$$f_{X|Y}(x|y) = \boxed{\frac{f(x|y)}{f_Y(y)}} \quad \text{OR} \quad f(x|y)\cdot\Delta x = \boxed{\frac{f(x,y)\cdot\Delta x\cdot\Delta y}{f(y)\cdot\Delta y}}$$

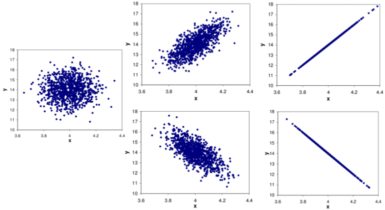
Bivariate Normal



$$X \sim \mathrm{N}(0, 1),$$
$$Y = \rho X + \epsilon;\ \epsilon \sim \mathrm{N}(0, 1 - \rho^2),$$

ϵ is independent of X. Given $X = x$, $Y = \rho x + \epsilon$.

Bivariate Normal



The distribution of points within a vertical slice at x.

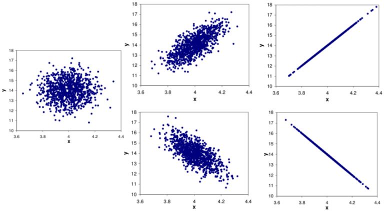
$$\mathbb{E}(Y|X=x) = \mathbb{E}(\rho x + \epsilon) = \rho x.$$

Regression towards the mean, e.g., son's height given father's height.

$$\mathrm{Var}(Y|X=x) = \mathrm{Var}(\rho x + \epsilon) = \mathrm{Var}(\epsilon) = 1 - \rho^2.$$

$$[Y|X=x] \sim \mathrm{N}(\rho x, 1-\rho^2).$$

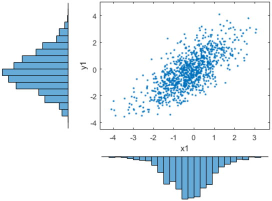
Bivariate Normal

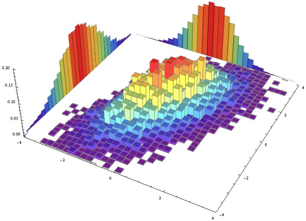


$$
\begin{aligned}
f(x,y) &= f_X(x) f_{Y|X}(y|x) \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left(-\frac{(y-\rho x)^2}{2(1-\rho^2)}\right) \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}(x^2+y^2-2\rho xy)\right].
\end{aligned}
$$

symmetric in (x, y)

Expectation





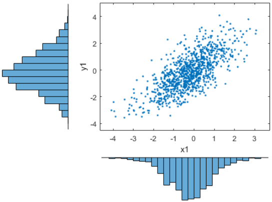
If $(X, Y) \sim p(x, y)$, then

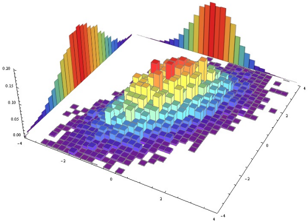
$$\mathbb{E}(h(X, Y)) = \sum_x \sum_y h(x, y) p(x, y).$$

If $(X, Y) \sim f(x, y)$, then

$$\mathbb{E}(h(X, Y)) = \int \int h(x, y) f(x, y) dx dy.$$

Expectation





Population average or long run average of $h(X, Y)$.

$$\frac{1}{n}\sum_{i=1}^{n} h(X_i, Y_i) = \frac{1}{n}\sum_{cells} h(x,y) n f(x,y)\Delta x \Delta y$$

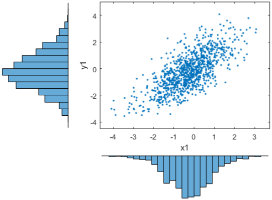
$$\to \int\int h(x,y) f(x,y) dx dy.$$

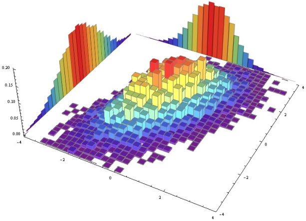
Variance

Let $\mu_h = \mathbb{E}(h(X,Y))$, then

$$\mathrm{Var}(h(X,Y)) = \mathbb{E}[(h(X,Y) - \mu_h)^2].$$

Covariance



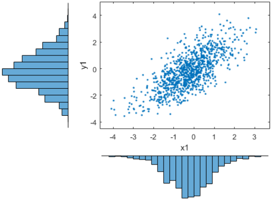


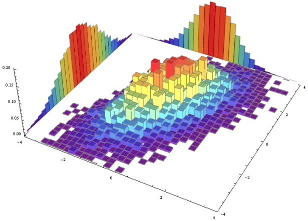
Let $\mu_X = \mathbb{E}(X)$, $\mu_Y = \mathbb{E}(Y)$, we define the covariance

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

It is defined for both discrete and continuous random variables.

Covariance



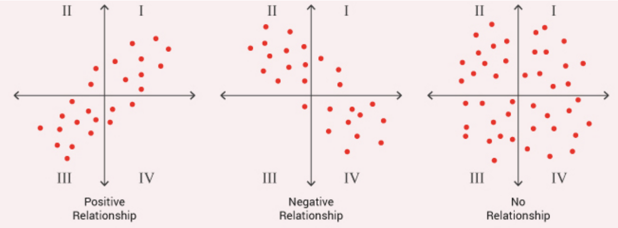


$$(X_i, Y_i) \sim f(x, y),\ i = 1, ..., n.$$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i;\ \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

$$\mathrm{Cov}(X, Y) \doteq \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}).$$

Covariance



$$\mathrm{Cov}(X, Y) \doteq \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}).$$

I, III: $(X_i - \bar{X})(Y_i - \bar{Y}) > 0.$
II, IV: $(X_i - \bar{X})(Y_i - \bar{Y}) < 0.$

Covariance

$$
\begin{aligned}
\mathrm{Cov}(X, Y) &= \mathbb{E}[(X-\mu_X)(Y-\mu_Y)] \\
&= \mathbb{E}[XY - \mu_X Y - X\mu_Y + \mu_X\mu_Y] \\
&= \mathbb{E}(XY) - \mu_X\mathbb{E}(Y) - \mu_Y\mathbb{E}(X) + \mu_X\mu_Y \\
&= \mathbb{E}(XY) - \mu_X\mu_Y \\
&= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).
\end{aligned}
$$

Clearly, $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$ and $\mathrm{Cov}(Y, Y) = \mathrm{Var}(Y)$.

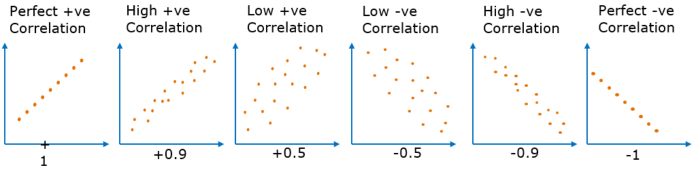
Linearity

$$
\begin{aligned}
& \mathrm{Cov}(aX+b, cY+d) \\
= \;& \mathbb{E}[(aX+b-\mathbb{E}(aX+b))(cY+d-\mathbb{E}(cY+d))] \\
= \;& \mathbb{E}[a(X-\mathbb{E}(X))c(Y-\mathbb{E}(Y))] = ac\mathrm{Cov}(X,Y).
\end{aligned}
$$

Covariance depends on units (meter/foot, kilogram/pound).

$$
\begin{aligned}
& \mathrm{Cov}(X+Y, Z) = \mathbb{E}[(X+Y-\mathbb{E}(X+Y))(Z-\mathbb{E}(Z))] \\
= \;& \mathbb{E}[(X-\mathbb{E}(X)+Y-\mathbb{E}(Y))(Z-\mathbb{E}(Z))] \\
= \;& \mathbb{E}[(X-\mathbb{E}(X))(Z-\mathbb{E}(Z))] + \mathbb{E}[(Y-\mathbb{E}(Y))(Z-\mathbb{E}(Z))] \\
= \;& \mathrm{Cov}(X,Z) + \mathrm{Cov}(Y,Z).
\end{aligned}
$$

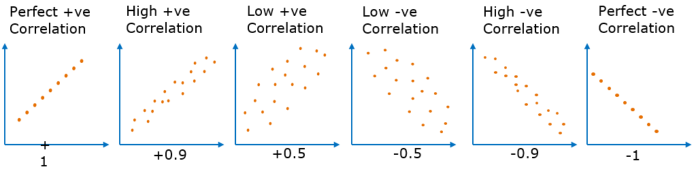
Correlation



Standardize: $X \to (X - \mu_X)/\sigma_X$, $Y \to (Y - \mu_Y)/\sigma_Y$.

$$\mathrm{Cov}\left(\frac{X-\mu_X}{\sigma_X}, \frac{Y-\mu_Y}{\sigma_Y}\right) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}} = \mathrm{Corr}(X,Y).$$

Correlation



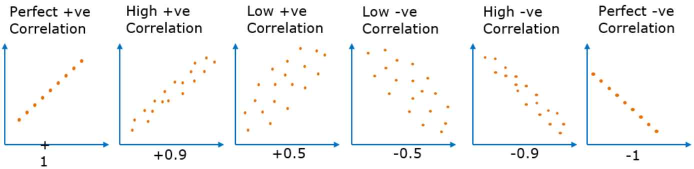
$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}.$$

$$\mathrm{Cov}(X,Y) \doteq \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}).$$

$$\mathrm{Var}(X) \doteq \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2;\ \mathrm{Var}(Y) \doteq \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2.$$

$$\mathrm{Corr}(X,Y) \doteq \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}.$$

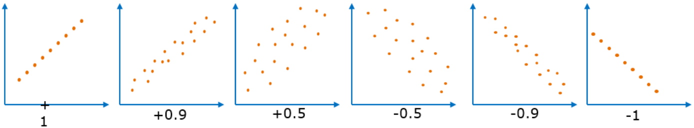
Correlation



Centralize: $\tilde{X}_i = X_i - \bar{X}$; $\tilde{Y}_i = Y_i - \bar{Y}$.

$$\mathrm{Corr}(X, Y) \doteq \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}\sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} = \frac{\sum_{i=1}^n \tilde{X}_i \tilde{Y}_i}{\sqrt{\sum_{i=1}^n \tilde{X}_i^2}\sqrt{\sum_{i=1}^n \tilde{Y}_i^2}}.$$

Correlation

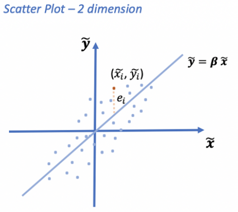


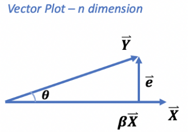
Centralize: $\tilde{X}_i = X_i - \bar{X}$; $\tilde{Y}_i = Y_i - \bar{Y}$.

$$\text{Corr}(X,Y) = \frac{\sum_{i=1}^n \tilde{X}_i \tilde{Y}_i}{\sqrt{\sum_{i=1}^n \tilde{X}_i^2}\sqrt{\sum_{i=1}^n \tilde{Y}_i^2}} = \frac{\langle \mathbf{X}, \mathbf{Y} \rangle}{|\mathbf{X}||\mathbf{Y}|} = \cos\theta.$$

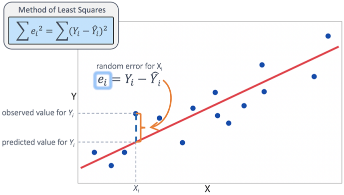
Correlation and regression

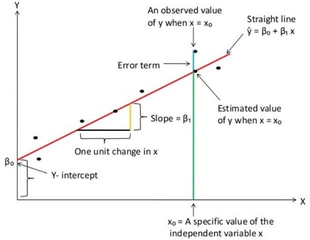
	$\vec{X}$	$\vec{Y}$	$\vec{e}$
1	$\tilde{x}_1$	$\tilde{y}_1$	e_1
⋮	⋮	⋮	⋮
2	$\tilde{x}_i$	$\tilde{y}_i$	e_i
⋮	⋮	⋮	⋮
n	$\tilde{x}_n$	$\tilde{y}_n$	e_n





Correlation and regression



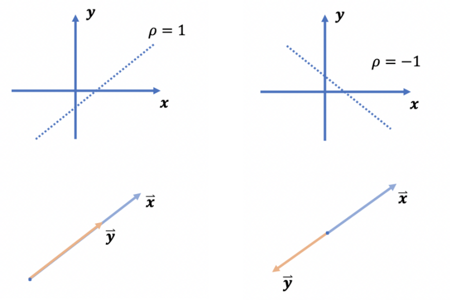


Regression line:

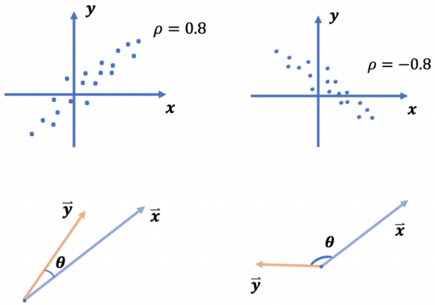
$$\hat{Y} - \bar{Y} = \beta_1 (X - \bar{X}).$$

$$\hat{Y} = \beta_1 X + (\bar{Y} - \beta_1 \bar{X}) = \beta_1 X + \beta_0.$$

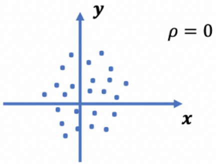
Correlation and regression

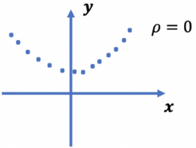


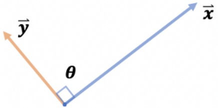
Correlation and regression



Correlation and regression





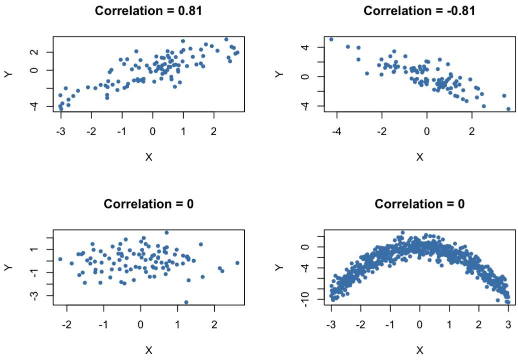


Independence

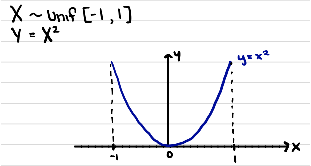
$P(A \cap B) = P(A)P(B)$.
$p(x, y) = p_X(x)p_Y(y)$
$f(x, y) = f_X(x)f_Y(y)$

$$\begin{aligned}
\mathrm{Cov}(X, Y) &= \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] \\
&= \sum_x \sum_y (x - \mu_X)(y - \mu_Y)p(x, y) \\
&= \sum_x \sum_y (x - \mu_X)(y - \mu_Y)p_X(x)p_Y(y) \\
&= \sum_x (x - \mu_X)p_X(x) \sum_y (y - \mu_Y)p_Y(y) \\
&= \left(\sum_x xp_X(x) - \mu_X\right)\left(\sum_y yp_Y(y) - \mu_Y\right) = 0.
\end{aligned}$$

Correlation

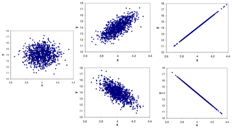


Correlation



Let X be a uniform distribution over $[-1, 1]$. Let $Y = X^2$. Then X and Y are not independent.
However, $\mathbb{E}(XY) = \mathbb{E}(X^3) = 0$, and $\mathbb{E}(X) = 0$. Thus $\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0$.

Bivariate normal



$$X \sim \mathrm{N}(0,1),$$
$$Y = \rho X + \epsilon;\ \epsilon \sim \mathrm{N}(0, 1-\rho^2),$$

$$\mathbb{E}(Y) = \mathbb{E}(\rho X + \epsilon) = 0.$$

ϵ and X are independent.

$$\mathrm{Var}(Y) = \mathrm{Var}(\rho X + \epsilon) = \rho^2 \mathrm{Var}(X) + \mathrm{Var}(\epsilon) = 1.$$

$$\mathrm{Cov}(X,Y) = \mathbb{E}(XY) = \mathbb{E}[X(\rho X + \epsilon)] = \rho \mathbb{E}(X^2) + \mathbb{E}(X\epsilon) = \rho.$$
$$\mathbb{E}(X\epsilon) = \mathbb{E}(X)\mathbb{E}(\epsilon) = 0.$$

Variance of sum

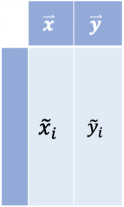
$\mathbb{E}(X+Y) = \sum_x \sum_y (x+y)p(x,y) =$
$\sum_x \sum_y xp(x,y) + \sum_x \sum_y yp(x,y) = \mathbb{E}(X) + \mathbb{E}(Y)$.

$$
\begin{aligned}
\mathrm{Var}(X+Y) &= \mathbb{E}[((X+Y)-\mu_{X+Y})^2] \\
&= \mathbb{E}[((X-\mu_X)+(Y-\mu_Y))^2] \\
&= \mathbb{E}[(X-\mu_X)^2+(Y-\mu_Y)^2+2(X-\mu_X)(Y-\mu_Y)] \\
&= \mathbb{E}[(X-\mu_X)^2]+\mathbb{E}[(Y-\mu_Y)^2]+2\mathbb{E}[(X-\mu_X)(Y-\mu_Y)] \\
&= \mathrm{Var}(X)+\mathrm{Var}(Y)+2\mathrm{Cov}(X,Y).
\end{aligned}
$$

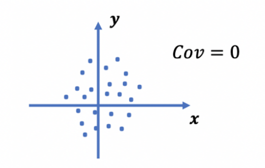
If X and Y are independent, then $\mathrm{Cov}(X,Y)=0$, and

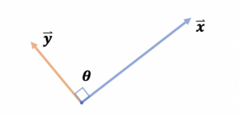
$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

Variance of sum

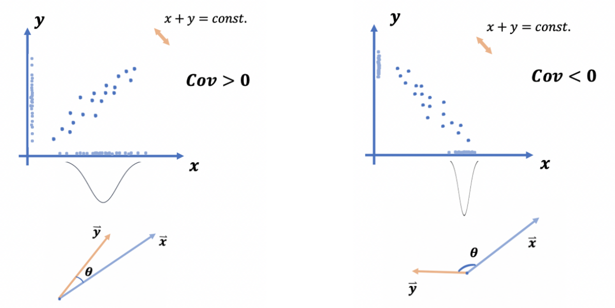


$$\frac{1}{n}\sum_{i=1}^{n} \tilde{x}_i{}^2 = Var(X) = \frac{1}{n}|\vec{x}|^2$$

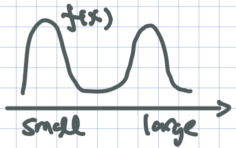




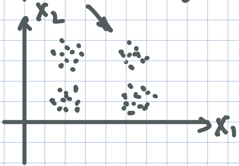
Variance of sum



Average

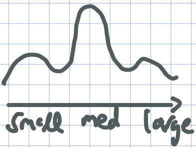


$X_1 \backslash X_2$	small	large
small	small	medium
large	medium	large

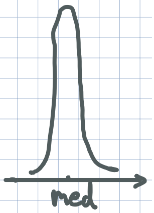


$$\bar{X} = \frac{X_1+X_2}{2}$$

$n=2$



$n=100$



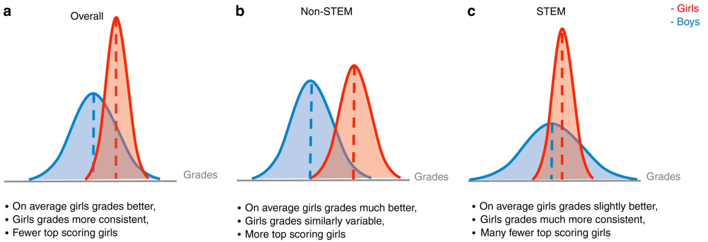
variance

shape smoother

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

CLT LLN

Average


a
Overall
Grades
b
Non-STEM
Grades
c
STEM
- Girls
- Boys
Grades
• On average girls grades better,
• Girls grades more consistent,
• Fewer top scoring girls
• On average girls grades much better,
• Girls grades similarly variable,
• More top scoring girls
• On average girls grades slightly better,
• Girls grades much more consistent,
• Many fewer top scoring girls

Sum and average

$X_i \sim f(x)$, $i = 1, ..., n$, iid: independent and identically distributed.

$$S = \sum_{i=1}^{n} X_i. \ \bar{X} = \frac{S}{n}.$$

$$\mathbb{E}(X_i) = \mu; \ \mathrm{Var}(X_i) = \sigma^2, \ i = 1, ..., n.$$

$$\mathbb{E}(S) = \mathbb{E}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathbb{E}(X_i) = n\mu.$$

$$\mathrm{Var}(S) = \mathrm{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = n\sigma^2.$$

$$\mathbb{E}(\bar{X}) = \frac{\mathbb{E}(S)}{n} = \mu.$$

$$\mathrm{Var}(\bar{X}) = \frac{\mathrm{Var}(S)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Law of large number

$$\mathbb{E}(\bar{X}) = \frac{\mathbb{E}(S)}{n} = \mu.$$

$$\mathrm{Var}(\bar{X}) = \frac{\mathrm{Var}(S)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \to 0.$$

$$\bar{X} \to \mu, \text{ in probability.}$$

$$P(|\bar{X} - \mu| < \epsilon) \to 1, \ \forall \epsilon > 0.$$

Average $\to$ expectation.

Law of large number

Special case:

$$X = \sum_{i=1}^{n} Z_i, \; Z_i \sim \text{Bernoulli}(p) \text{ iid.}$$

$$\mathbb{E}(X) = np; \; \text{Var}(X) = np(1-p).$$

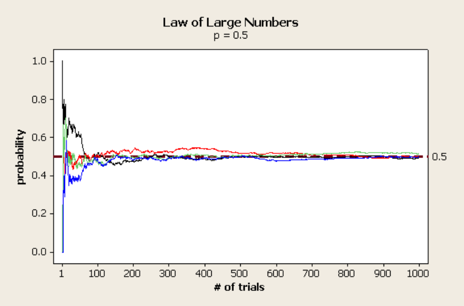
$$\mathbb{E}(X/n) = p; \; \text{Var}(X/n) = p(1-p)/n \to 0.$$

$$X/n \to p, \text{ in probability.}$$

Frequency $\to$ probability.
X/n is average of Z_i. Probability is expectation of Z_i.

Law of large number

Special case:



Keep flipping a fair coin, frequency $\to 1/2$.

Intuition: most of 2^n sequences have frequencies close to 1/2.

Law of large number

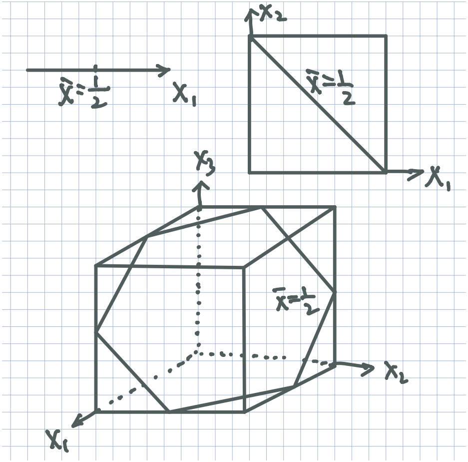
Special case: $X_i \sim \text{Uniform}[0,1]$, iid, $i = 1, ..., n$.

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \to \mathbb{E}(X_i) = 1/2.$$

$$P(|\bar{X} - 1/2| < \epsilon) \to 1, \ \forall \epsilon > 0.$$

Intuition: $(X_1, ..., X_i, ..., X_n)$ is a random point in $\Omega = [0,1]^n$, n-dimensional unit cube.
$A = \{(x_1, ..., x_i, ..., x_n) : |\bar{x} - 1/2| < \epsilon\}$ is the central diagonal piece.
$P(A)$ is the volume of A. $P(A) \to 1$.
No matter how small ϵ is, the volume of the central diagonal piece is almost the same as the volume of the whole n-dimensional unit cube Ω.
Most of the points in Ω belong to A.

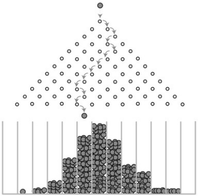
Law of large number



Statistical physics

Most of the points in Ω belong to A.
Suppose $(x_1, ..., x_i, ..., x_n)$ describes a physical system, e.g., $n = 10^{23}$ molecules.
It evolves **deterministically** over time, by traversing with Ω.
Ergodic: it traverses every point in Ω with equal number of visits in the long run.
Then mostly it will be in A, and thus $\bar{x} \doteq 1/2$.
Or at any **random moment**, $(x_i, ..., x_i, ..., x_n) \sim \text{Unif}[0, 1]$ iid, and thus $\bar{x} \to 1/2$.
Law of large number is the reason statistical physics makes sense, even if we assume things move deterministically.

Central limit theorem

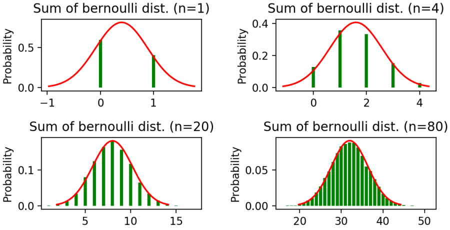


$$X = \sum_{i=1}^{n} Z_i,\ Z_i \sim \text{Bernoulli}(1/2) \text{ iid.}$$

$$X \sim \text{Binomial}(n, 1/2).\ P(X = k) = \frac{\binom{n}{k}}{2^n}.$$

$$P(X = n/2 + z\sqrt{n}/2) = P(X/n = 1/2 + z/(2\sqrt{n}))$$
$$\doteq \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \frac{2}{\sqrt{n}} = f(z)\Delta z.$$

Central limit theorem

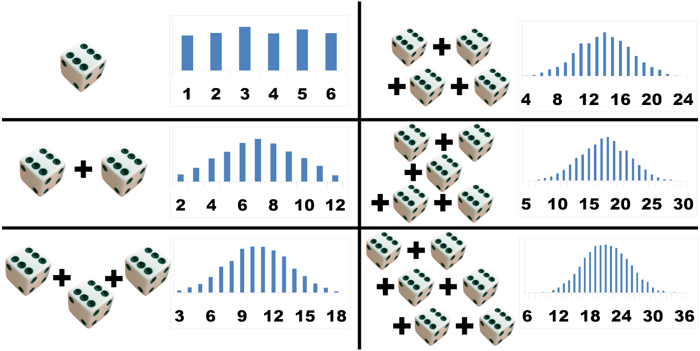


$$X = \sum_{i=1}^{n} Z_i, \;\; Z_i \sim \text{Bernoulli}(p) \text{ iid.}$$

$$\mathbb{E}(X) = np; \;\; \text{Var}(X) = np(1-p).$$

$$X \sim \text{Binomial}(n, p) \sim \text{N}(np, np(1-p))).$$

Central limit theorem

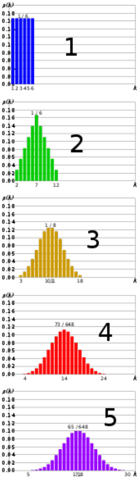


$$S = \sum_{i=1}^{n} X_i.$$

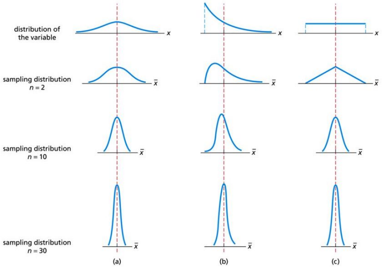
$$\mathbb{E}(X_i) = \mu;\ \mathrm{Var}(X_i) = \sigma^2,\ i = 1, ..., n.$$

$$S \sim \mathrm{N}(n\mu, n\sigma^2).$$

Central limit theorem



Central limit theorem

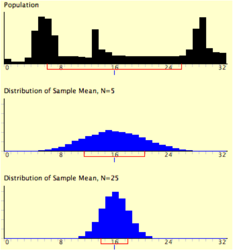


$$S = \sum_{i=1}^{n} X_i. \ \bar{X} = S/n.$$

$$\mathbb{E}(X_i) = \mu; \ \mathrm{Var}(X_i) = \sigma^2, \ i = 1, ..., n.$$

$$S \sim \mathrm{N}(n\mu, n\sigma^2). \ \bar{X} \sim \mathrm{N}(\mu, \sigma^2/n).$$

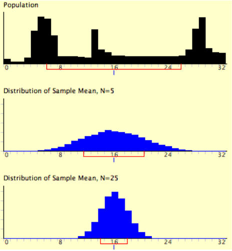
Central limit theorem



Universal, regardless of the distribution of each X_i.

$$S \sim \mathrm{N}(n\mu, n\sigma^2).\ \bar{X} \sim \mathrm{N}(\mu, \sigma^2/n).$$

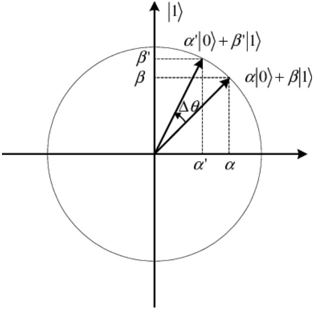
Central limit theorem



Universal, regardless of the distribution of each X_i.

$$S \sim \mathrm{N}(n\mu, n\sigma^2).\ \bar{X} \sim \mathrm{N}(\mu, \sigma^2/n).$$

Quantum coin flipping

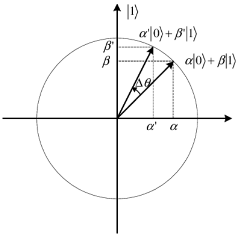


$\Psi = (\Psi(0) = \alpha, \Psi(1) = \beta) = \alpha|0\rangle + \beta|1\rangle$ is a vector rotating over time.
When observed, $P(0) = |\alpha|^2$. $P(1) = |\beta|^2$.
$|\alpha|^2 + |\beta|^2 = 1$. ($\alpha$ and β are complex numbers)
Superposition of $|0\rangle$ and $|1\rangle$.
Qubit for quantum computer

Schrodinger's cat

1 is alive, and 0 is dead.
$\Psi = (\Psi(0) = \alpha, \Psi(1) = \beta) = \alpha|0\rangle + \beta|1\rangle$ is a vector rotating over time.
When observed (measured), $P(0) = |\alpha|^2$. $P(1) = |\beta|^2$.
$|\alpha|^2 + |\beta|^2 = 1$. ($\alpha$ and β are complex numbers)
interpretation: Probability is the **subjective** uncertainty of the **observer** before **measuring** the result. Or frequency that the observer sees a result in long run repetition.
Heisenberg uncertainty principle

Quantum die rolling

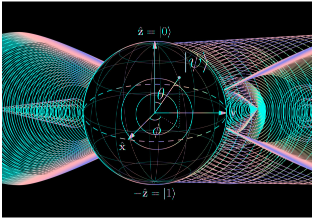


6-dimensional vector rotating over time

$$\begin{aligned}\Psi &= (\Psi(1), \Psi(2), \Psi(3), \Psi(4), \Psi(5), \Psi(6)) \\ &= \alpha_1|1\rangle + \alpha_2|2\rangle + \alpha_3|3\rangle + \alpha_4|4\rangle + \alpha_5|5\rangle + \alpha_6|6\rangle\end{aligned}$$

When observed, $P(1) = |\alpha_1|^2$. $P(2) = |\alpha_2|^2$, ..., $P(6) = |\alpha_6|^2$.
$|\alpha_1|^2 + |\alpha_2|^2 + ... + |\alpha_6|^2 = 1$.
Superposition of $|1\rangle$, $|2\rangle$, ..., $|6\rangle$.

Quantum computer



n qubits: superposition of 2^n states.
2^n-dimensional vector rotating over time.
$\Psi = (\Psi(HHH...H), \Psi(HHH...T), ..., \Psi(TTT...T))$.

Electron position

$\Psi = (\Psi(x), \forall x)$.

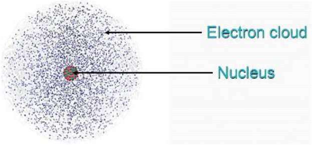
Wave function evolving according to Schrodinger's equation.

Infinite-dimension vector rotating.

$P(X \in (x, x+\Delta x)) = f(x)\Delta x$.

$f(x) = |\Psi(x)|^2$.

In 2D space $x = (x_1, x_2)$ or (x, y)



$$P(X \in (x, x+\Delta x), Y \in (y, y+\Delta y)) = f(x, y)\Delta x \Delta y.$$

Electron position

$\Psi = (\Psi(x), \forall x)$.

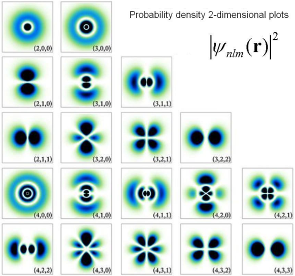
Wave function evolving according to Schrodinger's equation.

Infinite-dimension vector rotating.

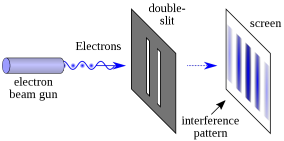
$P(X \in (x, x + \Delta x)) = f(x)\Delta x$.

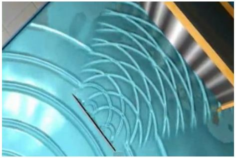
$f(x) = |\Psi(x)|^2$.

Electron cloud, physics and chemistry



Double slit experiment
$\Psi = (\Psi(x), \forall x)$.
Wave function evolving according to Schrodinger's equation.
Infinite-dimension vector rotating.
$P(X \in (x, x + \Delta x)) = f(x)\Delta x$.
$f(x) = |\Psi(x)|^2$.





Particle and wave duality
wave function, subject belief of observer
probability density function of particle position