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.
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.

<table>
<thead>
<tr>
<th></th>
<th>male</th>
<th>female</th>
</tr>
</thead>
<tbody>
<tr>
<td>taller than 6 ft</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>shorter than 6 ft</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>
**Example 2:** A male, B tall.

<table>
<thead>
<tr>
<th></th>
<th>male</th>
<th>female</th>
</tr>
</thead>
<tbody>
<tr>
<td>taller than 6 ft</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>shorter than 6 ft</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>

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

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

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

**Probability = population proportion.**
Population proportion

**Experiment → outcome → number**

**Example 2:** A male, B tall.

<table>
<thead>
<tr>
<th></th>
<th>male</th>
<th>female</th>
</tr>
</thead>
<tbody>
<tr>
<td>taller than 6 ft</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>shorter than 6 ft</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>

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

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

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

<table>
<thead>
<tr>
<th></th>
<th>male</th>
<th>female</th>
</tr>
</thead>
<tbody>
<tr>
<td>taller</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>shorter</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>

$p(m, t) = .3; p(m, s) = .2; p(f, t) = .1; p(f, s) = .4.$
Example 2: $X \in \{ \text{male, female} \}, \ Y \in \{ \text{tall, short} \}$.

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

<table>
<thead>
<tr>
<th></th>
<th>taller than 6 ft</th>
<th>shorter than 6 ft</th>
</tr>
</thead>
<tbody>
<tr>
<td>male</td>
<td>30</td>
<td>50</td>
</tr>
<tr>
<td>female</td>
<td>10</td>
<td>50</td>
</tr>
</tbody>
</table>

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

$$P(Y = y) = p(y) = \sum_x p(x, y).$$
Example 2: \( X \in \{ \text{male, female} \}, \ Y \in \{ \text{tall, short} \}. \)

\[
p(x, y) = P(X = x, Y = y).
\]

\[
P(X = x | Y = y) = p(x | y) = p(x, y) / p(y).
\]

\[
P(Y = y | X = x) = p(y | x) = p(x, y) / p(x).
\]

Chain rule: \( p(x, y) = p(x)p(y | x) = p(y)p(x | y). \)
Rule of total probability

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

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

\[
\begin{array}{c|cc}
 & \text{male} & \text{female} \\
\hline
\text{taller than 6 ft} & 30 & 10 \\
\text{shorter than 6 ft} & 50 & 50 \\
\end{array}
\]

$$ p(y) = \sum_x p(x, y) = \sum_x p(x)p(y|x). $$
Example 2: $X \in \{ \text{male, female} \}$, $Y \in \{ \text{tall, short} \}$.

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

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

\[ P(A|B) = P(A). \]
\[ P(A \cap B) = P(A)P(B). \]

\[ X \in \{ \text{male, female} \}, \ Y \in \{ \text{college, not}\} \]

\[
p(y|x) = p(y).
\]
\[
p(x, y) = p(x)p(y|x) = p(x)p(y).
\]
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?
\[ P(D) = 1\% \]
\[ P(+) | D = 90\%, \quad P(−|N) = 90\% \]
\[ P(D|+) = ? \]
\[ X \in \{D, N\}, \quad Y \in \{+, −\} \].
Example 6: Rare disease example

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

\[
p(x|y) = \frac{p(x, y)}{p(y)} = \frac{p(x)p(y|x)}{\sum_{x'} p(x')p(y|x')}
\]

\[p(x): \text{ prior belief. } p(x|y): \text{ posterior belief.}\]
Discrete joint, marginal, conditional

\[ X = \text{eye color}, \ Y = \text{hair color}. \]
$P(X \in (x, x + \Delta x)) = f(x)\Delta x.$

$f(x) = \frac{P(X \in (x, x + \Delta x))}{\Delta x}.$
Two continuous random variables

\[ X = \text{height}, \ Y = \text{weight}. \]
Two continuous random variables

\[ X = \text{height}, \quad Y = \text{weight}. \]
Two continuous random variables

\[ X = \text{height}, \, Y = \text{weight}. \]
Two continuous random variables

\[ X = \text{husband}, \quad Y = \text{wife}. \]
Two continuous random variables

\[ X = \text{husband}, \ Y = \text{wife}. \]
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

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

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

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

\[ f(x|y) = \frac{f(x, y)}{f(y)}. \]
Conditional density

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

![Graph showing the relationship between Body Mass Index (BMI) and Percent Body Fat (%BF). The graph distinguishes between males and females, with male data points in red and female data points in blue. The distribution of data points suggests a positive correlation between BMI and %BF, with a higher concentration of male data points at lower BMI values and a higher concentration of female data points at higher BMI values.](image-url)
Density

A: 

B: 

C: 

D: 

Distribution
Correlation
Limiting
Density

Distribution

Correlation

Limiting

\[ f(x, y) \cdot \Delta x \cdot \Delta y \]

\[ f_x(x) = \int f(x, y) \, dy \]

\[ f_{X|Y}(x|y) = \frac{f(x, y)}{f_y(y)} \]

\[ \frac{p(x, y)}{r_y(y)} \]

\[ f(x, y) = \frac{f(x, y) \cdot \Delta x \cdot \Delta y}{f_y(y) \cdot \Delta y} \]
\[ X \sim N(0, 1), \]
\[ Y = \rho X + \epsilon; \quad \epsilon \sim N(0, 1 - \rho^2), \]

\(\epsilon\) is independent of \(X\). Given \(X = x\), \(Y = \rho x + \epsilon\).
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.

\[ \text{Var}(Y|X = x) = \text{Var}(\rho x + \epsilon) = \text{Var}(\epsilon) = 1 - \rho^2. \]

\[ [Y|X = x] \sim N(\rho x, 1 - \rho^2). \]
Bivariate Normal

\[ f(x, y) = f_X(x) f_{Y|X}(y|x) \]

\[ = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{x^2}{2} \right) \frac{1}{\sqrt{2\pi(1 - \rho^2)}} \exp \left( -\frac{(y - \rho x)^2}{2(1 - \rho^2)} \right) \]

\[ = \frac{1}{2\pi \sqrt{1 - \rho^2}} \exp \left[ -\frac{1}{2(1 - \rho^2)} (x^2 + y^2 - 2\rho xy) \right]. \]

 symmetric in \((x, y)\)
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

\[
\mathbb{E}(h(X, Y)) = \int \int h(x, y)f(x, y)dxdy.
\]
Population average or long run average of $h(X, Y)$.

$$
\frac{1}{n} \sum_{i=1}^{n} h(X_i, Y_i) = \frac{1}{n} \sum_{\text{cells}} h(x, y)nf(x, y)\Delta x\Delta y
\rightarrow \int \int h(x, y)f(x, y)dxdy.
$$
Let $\mu_h = \mathbb{E}(h(X, Y))$, then

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

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

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

\[
\text{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}).
\]
Covariance

\[
\text{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}).
\]

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

II, IV: \((X_i - \bar{X})(Y_i - \bar{Y}) < 0\).
Covariance

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

Clearly, \(\text{Cov}(X, X) = \text{Var}(X)\) and \(\text{Cov}(Y, Y) = \text{Var}(Y)\).
Covariance depends on units (meter/foot, kilogram/pound).

\[
\text{Cov}(aX + b, cY + d) = \mathbb{E}[(aX + b - \mathbb{E}(aX + b))(cY + d - \mathbb{E}(cY + d))] = a\text{Cov}(X, Y).
\]

\[
\text{Cov}(X + Y, Z) = \mathbb{E}[(X + Y - \mathbb{E}(X + Y))(Z - \mathbb{E}(Z))] = \text{Cov}(X, Z) + \text{Cov}(Y, Z).
\]
Correlation

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

\[
\text{Cov} \left( \frac{X - \mu_X}{\sigma_X}, \frac{Y - \mu_Y}{\sigma_Y} \right) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)}} = \text{Corr}(X, Y).
\]
Correlation

\[
\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)}}.
\]

\[
\text{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}).
\]

\[
\text{Var}(X) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2; \quad \text{Var}(Y) = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y})^2.
\]

\[
\text{Corr}(X, Y) = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}}.
\]
Centralize: $\tilde{X}_i = X_i - \bar{X}; \tilde{Y}_i = Y_i - \bar{Y}$.

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

$$= \frac{\sum_{i=1}^{n} \tilde{X}_i \tilde{Y}_i}{\sqrt{\sum_{i=1}^{n} \tilde{X}_i^2} \sqrt{\sum_{i=1}^{n} \tilde{Y}_i^2}}.$$
Centralize: $\tilde{X}_i = X_i - \bar{X}$; $\tilde{Y}_i = Y_i - \bar{Y}$.

\[
\text{Corr}(X, Y) = \frac{\sum_{i=1}^{n} \tilde{X}_i \tilde{Y}_i}{\sqrt{\sum_{i=1}^{n} \tilde{X}_i^2} \sqrt{\sum_{i=1}^{n} \tilde{Y}_i^2}} = \frac{\langle X, Y \rangle}{|X| |Y|} = \cos \theta.
\]
Correlation and regression

<table>
<thead>
<tr>
<th>1</th>
<th>( \bar{x}_1 )</th>
<th>( \bar{y}_1 )</th>
<th>( e_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>2</td>
<td>( \bar{x}_i )</td>
<td>( \bar{y}_i )</td>
<td>( e_i )</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>n</td>
<td>( \bar{x}_n )</td>
<td>( \bar{y}_n )</td>
<td>( e_n )</td>
</tr>
</tbody>
</table>

Scatter Plot – 2 dimension

\[
\tilde{y} = \beta \tilde{x}
\]

Vector Plot – \( n \) dimension

\[
\tilde{y} = \beta \tilde{x}
\]
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

\[ \rho = 1 \]

\[ \rho = -1 \]
Correlation and regression

\[ \rho = 0.8 \]

\[ \rho = -0.8 \]
Correlation and regression

\[ \rho = 0 \]

\[ \bar{y} \quad \bar{x} \]

\[ \theta \]
Independence

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

\[ \text{Cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] \]
\[ = \sum_x \sum_y (x - \mu_X)(y - \mu_Y)p(x, y) \]
\[ = \sum_x \sum_y (x - \mu_X)(y - \mu_Y)p_X(x)p_Y(y) \]
\[ = \sum_x (x - \mu_X)p_X(x)\sum_y (y - \mu_Y)p_Y(y) \]
\[ = \left( \sum_x xp_X(x) - \mu_X \right) \left( \sum_y yp_Y(y) - \mu_Y \right) = 0. \]
Correlation

- Correlation = 0.81
- Correlation = -0.81
- Correlation = 0
- Correlation = 0
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 $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0$. 

$X \sim \text{unif} [-1, 1]$

$Y = X^2$
Bivariate normal

\[ X \sim N(0, 1), \]
\[ Y = \rho X + \epsilon; \quad \epsilon \sim N(0, 1 - \rho^2), \]

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

\( \epsilon \) and \( X \) are independent.

\[ \text{Var}(Y) = \text{Var}(\rho X + \epsilon) = \rho^2 \text{Var}(X) + \text{Var}(\epsilon) = 1. \]

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

\[ \mathbb{E}(X \epsilon) = \mathbb{E}(X) \mathbb{E}(\epsilon) = 0. \]
### Variance of sum

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

\[
\text{Var}(X + Y) = \mathbb{E}[((X + Y) - \mu_{X+Y})^2] \\
= \mathbb{E}[((X - \mu_X) + (Y - \mu_Y))^2] \\
= \mathbb{E}[(X - \mu_X)^2 + (Y - \mu_Y)^2 + 2(X - \mu_X)(Y - \mu_Y)] \\
= \mathbb{E}[(X - \mu_X)^2] + \mathbb{E}[(Y - \mu_Y)^2] + 2\mathbb{E}[(X - \mu_X)(Y - \mu_Y)] \\
= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y).
\]

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

\[
\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).
\]
Variance of sum

\[ \frac{1}{n} \sum_{i=1}^{n} \bar{x}_i^2 = \text{Var}(X) = \frac{1}{n} |\bar{x}|^2 \]
Variance of sum

$\text{Cov} > 0$

$\text{Cov} < 0$
Average

\[ \bar{x} = \frac{x_1 + x_2}{2} \quad n=2 \]

\[ x \sim N(\mu, \frac{\sigma^2}{n}) \]

CLT

LLN

Distribution

Correlation

Limiting
Average

- **Overall**
  - On average girls grades better,
  - Girls grades more consistent,
  - Fewer top scoring girls

- **Non-STEM**
  - On average girls grades much better,
  - Girls grades similarly variable,
  - More top scoring girls

- **STEM**
  - On average girls grades slightly better,
  - Girls grades much more consistent,
  - Many fewer top scoring girls
Sum and average

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

\[ S = \sum_{i=1}^{n} X_i, \quad \bar{X} = \frac{S}{n}. \]

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

\[ \mathbb{E}(S) = \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{E}(X_i) = n\mu. \]

\[ \text{Var}(S) = \text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i) = n\sigma^2. \]

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

\[ \text{Var}(\bar{X}) = \frac{\text{Var}(S)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}. \]
Law of large number

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

\[ \text{Var}(\bar{X}) = \frac{\text{Var}(S)}{n^2} = \frac{n\sigma^2}{n} = \frac{\sigma^2}{n} \to 0. \]

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

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

Average \to \text{expectation.}
Law of large number

Special case:

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

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

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

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

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 $\rightarrow 1/2$.

**Intuition:** most of $2^n$ sequences have frequencies close to $1/2$. 
Law of large number

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

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

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

**Intuition:** \((X_1, \ldots, X_i, \ldots, X_n)\) is a random point in \(\Omega = [0, 1]^n\), \(n\)-dimensional unit cube.

\(A = \left\{ (x_1, \ldots, x_i, \ldots, x_n) : |\bar{x} - 1/2| < \epsilon \right\}\) is the central diagonal piece.

\(P(A)\) is the volume of \(A\). \(P(A) \to 1.\)

No matter how small \(\epsilon\) is, the volume of the central diagonal piece is almost the same as the volume of the whole \(n\)-dimensional unit cube \(\Omega\).

Most of the points in \(\Omega\) belong to \(A\).
Law of large number
Most of the points in $\Omega$ belong to $A$. Suppose $(x_1, ..., x_i, ..., x_n)$ describes a physical system, e.g., $n = 10^{23}$ molecules. It evolves \textbf{deterministically} over time, by traversing with $\Omega$. \textbf{Ergodic}: it traverses every point in $\Omega$ 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 \textbf{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

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

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

\[ P(X = n/2 + z\sqrt{n}/2) = P(X/n = 1/2 + z/(2\sqrt{n})) \]
\[ = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \frac{2}{\sqrt{n}} = f(z) \Delta z. \]
Central limit theorem

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

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

\[ X \sim \text{Binomial}(n, p) \sim \mathcal{N}(np, np(1 - p)). \]
Central limit theorem

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

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

\[ S \sim \mathcal{N}(n\mu, n\sigma^2). \]
Central limit theorem
Central limit theorem

\[ S = \sum_{i=1}^{n} X_i, \quad \bar{X} = S/n. \]

\[ \mathbb{E}(X_i) = \mu; \quad \text{Var}(X_i) = \sigma^2, \quad i = 1, \ldots, n. \]

\[ S \sim N(n\mu, n\sigma^2). \quad \bar{X} \sim N(\mu, \sigma^2/n). \]
Central limit theorem

Universal, regardless of the distribution of each $X_i$.

$$S \sim N(n\mu, n\sigma^2). \quad \bar{X} \sim N(\mu, \sigma^2/n).$$
Central limit theorem

Universal, regardless of the distribution of each $X_i$.

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

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 \( \beta \) are complex numbers)

Superposition of \( |0\rangle \) and \( |1\rangle \).

Qubit for quantum computer
Quantum mechanics

Schrodinger’s cat

\( \Psi = (\Psi(0) = \alpha, \Psi(1) = \beta) = \alpha|0\rangle + \beta|1\rangle \) is a vector rotating over time.

When observed (measured), \( P(0) = |\alpha|^2 \). \( P(1) = |\beta|^2 \).

\( |\alpha|^2 + |\beta|^2 = 1 \). (\( \alpha \) and \( \beta \) 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 mechanics

Quantum die rolling

6-dimensional vector rotating over time

\[
\Psi = (\Psi(1), \Psi(2), \Psi(3), \Psi(4), \Psi(5), \Psi(6))
\]

\[
= \alpha_1 |1\rangle + \alpha_2 |2\rangle + \alpha_3 |3\rangle + \alpha_4 |4\rangle + \alpha_5 |5\rangle + \alpha_6 |6\rangle
\]

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

\[ n \text{ qubits: superposition of } 2^n \text{ states.} \]

\[ 2^n \text{-dimensional vector rotating over time.} \]

\[ \Psi = (\Psi(HHH\ldots H), \Psi(HHH\ldots T), \ldots, \Psi(TTT\ldots T)). \]
**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
Quantum mechanics

**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