**Probability**:

$$\text{Set } \Omega \text{ (finite or countable)}$$
$$P(\omega) > 0, \sum_{\omega} P(\omega) = 1$$
$$A \subseteq \Omega, P(A) = \sum_{\omega \in A} P(\omega)$$

**Example**. Birthday Problem:
"How many people do we need to have chance 50% that 2 or more have same birthday date?"
Say have n people and C categories.

- What is $\Omega$?

$$\Omega = \{\omega : \omega = (\omega_1, \omega_2, \cdots, \omega_n), \omega_i \in \{1, 2, \cdots, C\}\}$$

- What is $P(\omega)$?

$$\text{Try } P(\omega) = \frac{1}{C^n}$$

- What is A?

$$A = \{\omega : \omega_i \neq \omega_j, \forall i, j\}$$
$$P(A) = \sum_{\omega \in A} P(\omega) = \frac{1}{C^n}|A|$$
$$\text{where}|A| = C(C-1)(C-2)\dots(C-n+1)$$

**Answer** 1:

$$P(A) = (1 - \frac{1}{C})(1 - \frac{2}{C})\cdots(1 - \frac{n-1}{C})$$

**Answer** 2: (humans)

$$\text{Use } \log(1-x) \sim -X$$
$$P(A) = \exp \sum_{i=1}^{n} \log(1 - i/C)$$
$$\approx \exp - \sum_{i=1}^{n} i/C$$
$$= \exp - \frac{\binom{n}{2}}{C}$$

Now set

$$e^{-\frac{\binom{n}{2}}{C}} = \frac{1}{2} \Rightarrow n = 1.2\sqrt{C} \approx 23$$

**Answer** 3:

$$\log(1-x) = -x + O(x^2)$$

$$-x - x^2 \leqslant \log(1-x) \leqslant -x, 0 \leqslant x \leqslant \frac{1}{2}$$

**Theorem**: if $n, C$ tend to $\infty$ so that

$$\frac{n^3}{C^2} \to 0, \frac{\binom{n}{2}}{C} \to E$$

$$P(A) \approx e^{-E}$$

**Problem** How many people:
Do we need to have even odds to have triple match?

**Example.**
Put N points down at random in $[0,1]^2$, put $\varepsilon$–Ball around each, what's chance cover?
We put probabilities on $\tau[0,1]$
**Manifolds**

$$\Omega = \{x_1, x_2, \cdots, x_{35} \in \mathbb{R}_+^{35}, \sum x_i = s, \prod x_i = \rho\}$$

**Half Way House:**

$$\Omega = (0,1]$$

Work with intervals $(a_n, b_n] = I_n$,

$$A = \cup_{i=1}^{n} I_n, I_n \text{ disjoint intervals}$$

$$P(A) \triangleq \sum_{I_i \in A} (b_i - a_i)$$

**Model for fair coin tossing:**

$$\omega = .\omega_1\omega_2\omega_3\cdots$$

(using nonterminated: $\frac{1}{2} = 0.011\cdots$)

$$\omega = \sum \frac{d_n(\omega)}{2^n}$$

$$A = \{\omega : d_i(\omega) = 1\}$$

has $P(A) = \frac{1}{2}$ for all $i$, similarly,

$$P(A_1 = E_1, \cdots, A_n = E_n) = \frac{1}{2^n}, \forall E_1, \cdots, E_n \in \{0, 1\}$$

**Theorem** (Borel's Weak Law of Large Numbers):

$$\forall \varepsilon > 0, P\left\{\left|\frac{d_1 + \cdots + d_n}{n} - \frac{1}{2}\right| > \varepsilon\right\} \to 0$$

*Proof*: Define $\Omega_n(\omega) = 2 \cdot d_n(\omega) - 1$
Same to prove

$$\forall \varepsilon, P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\Omega_i\right| > 2\varepsilon\right\} \to 0$$

Note

$$\int_0^1 \Omega_i(\omega)\mathrm{d}\omega = 0$$

$$\int_0^1 \Omega_i(\omega)\Omega_j(\omega) = \delta_{ij}$$

$$\text{So } \int_0^1 (\sum_{i=1}^{n}\Omega_i)\mathrm{d}\omega = 0, \int_0^1 (\sum_{i=1}^{n}\Omega_i)^2\mathrm{d}\omega = n$$

Applying Markov's inequality.

$$\begin{aligned} P\{\left|\frac{1}{n}\sum_{i=1}^{n}\Omega_i\right| > 2\varepsilon\} &= P\left((\sum_{i=1}^{n}\Omega_i)^2 > 4\varepsilon^2 n^2\right) \\ &\leqslant \frac{\int(\sum\Omega_i)^2}{4\varepsilon^2 n^2} \\ &= \frac{1}{4\varepsilon^2 n} \\ &\to 0 \end{aligned}$$

**Lemma:** Markov's inequality: if $f : (0,1) \to \mathbb{R}$,

$$f(\omega) \geqslant 0$$

Then $\forall a > 0$

$$P\{\omega : f(\omega) \geqslant a\} \leqslant \frac{\int_0^1 f(\omega)\mathrm{d}\omega}{a}$$

*Proof*

$$\int_0^1 f(\omega)\mathrm{d}\omega \geqslant \int_A f(\omega) \geqslant aP(A)$$

We want strong law: Borel's normal number theorem:

$$\lim \frac{1}{n}\sum_{i=1}^{n} \Omega_n(\omega) = 0$$

**Problems**: not true
$w = .1111\cdots, w = 0110000111110\cdots$, limit doesn't exist

**Definition**. $A \subseteq \Omega$ to be negligable if

$$\forall \varepsilon > 0, \exists \text{ finite or countable many intervals } I_1, \cdots s.t.$$
$$A \subset \cup_k I_k \text{ and } \sum_k |I_k| < \varepsilon$$

**Example**. rationals in $(0,1]$ negligable.
Note: a finite or countable union of negligable sets is negligable.

**Theorem** (Borel's Normal Number)

$$\Omega = \{0,1\}, I = (a,b], P(I) = b - a$$
$$\omega = \sum_{i=1}^{\infty} \frac{d_i(\omega)}{2^i}, \Omega_i(\omega) = 2d_i(\omega) - 1$$
$$S_n(\omega) = \sum_{i=1}^{n} \Omega_i(\omega)$$

**Theorem**

$$\lim_{n\to\infty} \frac{S_n(\omega)}{n} = 0$$

Except for $\omega$ in a negligible set.

*Proof.* Use subsquence argument.
**Lemma**: Say $x_n$ real numbers, $1 \leqslant i < \infty$, $y_n = x_1 + \cdots + x_n$, then $\frac{y_n}{n} \to 0$, provided

$$\frac{y_{n^2}}{n^2} \to 0$$
$$\sup |x_n| < \infty$$

**Corollary**

$$\lim \frac{S_n(\omega)}{n} = 0 \Leftrightarrow \lim \frac{S_{n^2}(\omega)}{n^2} = 0$$

*Proof.* $h_n^2 \leqslant n \leqslant (h_n + 1)^2$

$$\begin{aligned}
\left|\frac{y_n}{n}\right| &\leqslant \frac{|y_n|}{h_n^2} \\
&= \frac{|y_{h_n^2} - (y_{h_n^2} - y_n)|}{h_n^2} \\
&\leqslant \frac{|y_{h_n^2}|}{h_n^2} + \frac{|y_{h_n^2} - y_n|}{h_n^2}
\end{aligned}$$

if $|x_i| \leqslant C$ all $i$

$$\begin{aligned}
\frac{|y_n - y_{h_n^2}|}{h_n^2} &= \frac{|\sum_{i=h_n^2+1}^{n} x_i|}{h_n^2} \\
&\leqslant \frac{C(n - h_n^2)}{h_n^2} \\
&\leqslant \frac{2C}{h_n} \\
&\to 0
\end{aligned}$$

□

$$B = \{\omega : \lim_n \frac{S_{n^2}(\omega)}{n^2} = 0\}$$

Show $B^c$ is negligable, choose $\delta_n \downarrow 0$

$$\{\omega : \left|\frac{S_{n^2}(\omega)}{n^2}\right| < \delta_n,\ \text{any large n}\} \subseteq B$$

$$\begin{aligned} B^c &\subseteq \{\omega : \left|\frac{S_{n^2}}{n^2}\right| \geqslant \delta_n,\ \text{i.o.}\} \\ &\subseteq \cup_{n=j}^{\infty}\{\omega : \left|\frac{S_{n^2}}{n^2}\right| \geqslant \delta_n\} \\ &\triangleq \cup_{n=j}^{\infty} B_n (\forall j) \end{aligned}$$

$B_n$ is a disjoint union of intervals.

$$P(B_n) \leqslant \frac{1}{n^2\delta_n^2}(\text{Chebyshev's inequality})$$

Done if choose $\delta_n$ s.t.

$$\sum_{n=1}^{\infty} \frac{1}{n^2\delta_n^2} < \infty$$

Choose $\delta_n = n^{-\frac{1}{4}}$

Remarks:

- The set of normal numbers has negligible complement.
  Notice $N^c$ is not countable.

  **Example**. $(d_1(\omega), d_2(\omega), \cdots) = (1, 1, u_3, 1, 1, u_6, \cdots), d_i(\omega) = 1$ unless $i = 3k$.
  Since $\frac{1}{n}\sum_{i=1}^{n} d_i(\omega) \geqslant \frac{2}{3}$. Thus $\omega$ is not normal.
  but $(u_3, u_6, \cdots)$ is a $0-1$ sequence, then such $\omega$ is uncountable.

- Either need better bounds on $P\{\frac{S_n}{n} > a\}$ or use subsequence

- WLLN: $\lim_{n\to\infty} P\{|\frac{S_n}{n}| > \varepsilon\} = 0$
  SLLN: $P(\lim_{n\to\infty} \frac{S_n}{n} = 0) = 1$.
  **Example**. weak law holds but not strong law

$$P(j) = \frac{c}{j^2 \log|j|}, |j| \geqslant 2$$

Pick $X_1, \cdots, X_n$ independently from $P(j)$

$$S_n = \sum_{i=1}^{n} X_i$$

Then $P\{|\frac{S_n}{n}| > \varepsilon\} \to 0$
but $|\frac{S_n}{n}| > \varepsilon$ i.o. with probability 1
Strong law says
$\frac{S_n}{n}$ gets small and stays small.

- Note weak law

$$P\left\{\left|\frac{S_n}{n}\right| > \varepsilon\right\} \leqslant \frac{1}{n\varepsilon^2}$$

But strong law has no finite content.

$$B = \{\omega : \text{ Strong law holds}\} = \cap_{k=1}^{\infty} \cup_{m=1}^{\infty} \cap_{n=m}^{\infty} \{\omega : \left|\frac{S_n}{n}\right| < \frac{1}{k}\}$$

Algebras, $\sigma$-algebras, and construction of probability.
Let $\Omega$ be a set.
**Definition**. A field is a subsets of $\Omega : \mathcal{F}$ is a collection $F \subseteq \Omega$, s.t.

- $\varnothing \in \mathcal{F}$
- $F \in \mathcal{F} \Rightarrow F^c \in \mathcal{F}$
- $A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$

**Example**.

- $$\mathcal{F} = \{\varnothing, \Omega\}$$
- $$\Omega = (0, 1], \mathcal{F} \text{ is all finite disjoint unions of intervals } (a, b]$$

**Definition**. A $\sigma$-field is a field so

$$F_n \in \mathcal{F}, 1 \leqslant n < \infty$$
$$\cup_{i=1}^{\infty} F_i \in \mathcal{F}$$

- All subsets of $\Omega$ is a $\sigma$-field.

- If $\{F_n\}$ are $\sigma$-fields, the $\cap F_n$ is a $\sigma$-field.

- So take any collection Q of subsets of $\Omega$,

$$F(Q) = \text{Smallest } \sigma - \text{field containing Q} = \cap F$$

- Take $(0,1]$, $\mathrm{Q} = \text{All } (a,b]$,

$$F(Q) = \text{Borel sets}$$

**Definition**. A probability space $(\Omega, \mathcal{F}, P)$ is a set $\Omega$,
A $\sigma$-algebra of subsets $\mathcal{F}$ and a function $P : \mathcal{F} \to [0,1]$,

- $$P(\varnothing) = 0$$

- $$P(A^c) = 1 - P(A), A \in \mathcal{F}$$

- $$A_i, 1 \leqslant i < \infty \in \mathcal{F}, A_i \cap A_j = \varnothing,$$

  Then
$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

**What have we learned since Greeks?**
1) Can't define length or area of general sets
2) Wonderful approximate theory.

Constructing measures
$\Omega, \mathcal{F}$ field of subsets, $P$ probability on $\mathcal{F}$,
$\forall E \subset \Omega$, define
$$P^*(E) = \inf \sum_{i=1}^{\infty} P(F_i)$$
Inf over countable collections $\{F_i\}, F_i \in \mathcal{F}, E \subseteq \cup_{i=1}^{\infty} F_i$
**Idea**: (Caratheodory)

$$M = \{A \subset \Omega : \forall E \subset \Omega, P^*(E) = P^*(A \cap E) + P^*(E \cap A^c)\}$$

**Theorem**. M is a $\sigma$-algebra containing F,

- $P^*$ is probability on M,
- $P^*(F) = P(F), F \in F$,
- $P^*$ is unique such.

First step.

- $$P^*(\varnothing) = 0$$
- $$A \subseteq B, P^*(A) \leqslant P^*(B)$$
- $$A \subseteq \cup_{i=1}^{\infty} A_i \Rightarrow P^*(A) \leqslant \sum_i P^*(A_i)$$
$$\textit{Proof } \forall \varepsilon, \exists F_{i,h} \in F, \cup_h F_{i,h} \supseteq A_i$$
$$\sum_{h=1}^{\infty} P(F_{ih}) \leqslant P^*(A_i) + \frac{\varepsilon}{2^i}$$

Note says

$$P^*(E) = P^*(A \cap E \cup A^c \cap E) \leqslant P^*(A \cap E) + P^*(A^c \cap E)$$

So $A \in M \Leftrightarrow P^*(E) \geqslant P^*(A \cap E) + P^*(A^c \cap E)$

**Fact 1**. M is a field

- $\varnothing$
- complements
- suppose $A, B \in M$, show $A \cap B \in M$

$$\begin{aligned}
P^*(E) &= P^*(B \cap E) + P^*(B^c \cap E) \\
&= P^*(A \cap B \cap E) + P^*(A^c \cap B \cap E) + P^*(A \cap B^c \cap E) + P^*(B^c \cap A^c \cap E) \\
\text{(subadditivity)} &\geqslant P^*(A \cap B \cap E) + P^*(A^c \cap B \cap E \cup A \cap B^c \cap E \cup B^c \cap A^c \cap E) \\
&= P^*(A \cap B \cap E) + P^*((A \cap B)^c \cap E)
\end{aligned}$$

**Fact 2**. countable additivity
$\{A_n\}_{n=1}^{\infty} \in M, A_m \cap A_s = \varnothing$, everything.

$$P^*(E \cap (\cup A_i)) = \sum_n P^*(E \cap A_i)$$

*Proof.* Say $\{A_i\}_{i=1}^n$, induction:
(trivial for $n = 1$)
$n = 2$, if $A_1 \cup A_2 = \Omega$, correct!
If not,

$$\begin{aligned}
P^*(E \cap (A_1 \cup A_2)) &= P^*(E \cap (A_1 \cup A_2) \cap A_2) + P^*(E \cap (A_1 \cup A_2) \cap A_1^c) \\
&= P^*(E \cap A_1) + P^*(E \cap A_2)
\end{aligned}$$

Say OK for $n - 1$, given $\{A_i\}^n$,

$$\begin{aligned}
P^*(E \cap \cup_1^n A_i) &= P^*(E \cap (\cup^{n-1} A_i)) + P^*(E \cap A_n) \\
&= \sum_1^n P^*(E \cap A_i)
\end{aligned}$$

$$\begin{aligned}
P^*(E \cap \cup_1^n A_i) &\geqslant P^*(E \cap \cup_{i=1}^n A_i) \\
&= \sum_{i=1}^n P^*(A_i \cap E)
\end{aligned}$$

Let $n \to \infty$.

**Fact 3**. M is a $\sigma$-algebra and $P^*$ is countably additive on M.
Say $\{A_i\}_1^\infty \in M$, show $\cup_1^\infty A_i \in M$
"disjointify":

$$\begin{cases}
A_1' = A_1 \\
A_2' = A_2 \cap (A_1')^c \\
\cdots \\
A_n' = A_n \cap (\cup_{i=1}^{n-1} A_i')^c
\end{cases}$$

then $\cup_1^\infty A_i = \cup_1^\infty A_i'$
Let $A = \cup_1^\infty A_n', F_n = \cup_{i=1}^n A_i' \in M$

$$
\begin{aligned}
P^*(E) &= P^*(E \cap F_n) + P^*(E \cap F_n^c) \\
&\geqslant P^*(E \cap F_n) + P^*(E \cap F_n^c) \\
&= \sum_{i=1}^{n} P^*(E \cap A_i') + P^*(E \cap A^c)
\end{aligned}
$$

So

$$
\begin{aligned}
P^*(E) &\geqslant \sum_{1}^{\infty} P^*(E \cap A_i') + P^*(E \cap A^c) \\
&= P^*(E \cap \cup_1^\infty A_i') + P^*(E \cap A^c) \\
&= P^*(E \cap A) + P^*(E \cap A^c)
\end{aligned}
$$

and $P^*$ add on M by fact 2.

**Fact 4**. $F \subset M$
Take $A \in F$, and (given $\varepsilon > 0$), $A_i \in F$
So $E \subset \cup_1^\infty A_i$ and

$$
\sum_{1}^{\infty} P(A_i) \leqslant P^*(E) + \varepsilon
$$

Set $B_n = A \cap A_n$, $C_n = A^c \cap A_n$

$$
\begin{aligned}
E \cap A \subset \cup_1^\infty B_n, E \cap A^c \subset \cup_1^\infty C_n \\
P^*(A \cap E) + P^*(A^c \cap E) &\leqslant \sum_n P(B_n) + P(C_n) \\
&= \sum_n P(A_n) \leqslant P^*(E) + \varepsilon
\end{aligned}
$$

**Fact 5**. $P^*(F) = P(F)$, $F \in F$,
*Proof.* $P^*(F) \leqslant P(F)$
If $F \subset \cup_1^\infty F_n$, $F_n \in F$

$$
P(F) \leqslant \sum_i P(F \cap F_i) \leqslant \sum_1^\infty P(F_i)
$$

$P^*(F) \geqslant P(F)$:

Special case:

$$
\Omega = (0, 1]
$$

$\mathcal{F}$ = disjoint unions of finite intervals.

$$P(\cup_{i=1}^{\infty}(a_i, b_i]) = \sum_{i=1}^{\infty}(b_i - a_i)$$

Need unhappily to show that if
$I = \cup_{i=1}^{\infty}(a_i, b_i]$, then

$$P(I) = \sum_{i=1}^{\infty}(b_i - a_i)$$

We show

$$a) \cup_1^{\infty} I_i \subset I \Rightarrow \sum(b_i - a_i) \leqslant (b - a)$$

$$b) I \subset \cup_1^{\infty} I_i \Rightarrow (b - a) \leqslant \sum_1^{\infty}(b_i - a_i)$$

*Proof* of a): Say $(a_i, b_i), 1 \leqslant i \leqslant n$, disjoint.
$\cup_1^n(a_i, b_i] \subset (a, b]$, show $\sum_1^n(b_i - a_i) \leqslant (b - a)$
Induction on n.
Say ok for $n - 1$,
$(a(a_n, b_n]b]$ has $a_n \geqslant a_i$
Therefore, $\cup_1^{n-1}(a_i, b_k) \subset (a, a_n]$

$$\begin{aligned}\sum_1^n(b_i - a_i) &\leqslant (a_n - a) + b_n - a_n \\ &\leqslant b_n - a \\ &\leqslant b - a\end{aligned}$$

let $n \to \infty$

*Proof* of b): $(a, b] \subset \cup_{i=1}^{\infty}(a_i, b_i]$.
Pick $\varepsilon > 0, \varepsilon < b - a$

$$[a + \varepsilon, b] \subset \cup_{i=1}^{\infty}(a_i, b_i + \frac{\varepsilon}{2^i})$$

By **Heine-Borel Theorem**, $\exists n$,

$$[a + \varepsilon, b) \subset \cup_1^n(a_i, b_i + \frac{\varepsilon}{2^i})$$

$$(b - a - \varepsilon) \leqslant \sum_1^{\infty}(b_i - a_i) + \varepsilon$$

$$\varepsilon \to 0, b - a \leqslant \sum_{i=1}^{\infty}(b_i - a_i)$$

So
$\exists$ probability on M extending length on $\mathcal{F}$
$\mathcal{F} \subseteq \sigma(\mathcal{F}) \subseteq M$.
$M$ is a class of Lebesgue measure sets.

**Definition**. A collection of sets (in $\Omega$) is a $\pi$-system if closed under finite intersections.
**Definition** A collection of sets is a $\lambda$-system,
if

- $\varnothing \in L$
- Closed under complements
- $\{A_i\}_{i=1}^{\infty} \in L_1$ and $A_i \cap A_j = \varnothing$ All $i \neq j$
  $\cup_1^{\infty} A_i \in L$

**Example**. $\Omega = \{1, 2, 3, 4\}$

$$L = \{\varnothing, \Omega, \{1, 2\}, \{3, 4\}, \{1, 3\}, \{2, 4\}\}$$

a) L $\lambda$-system. $A, B \in L, A \subseteq B$ then, $B\ A = B \cap A^c \in L$
*Proof* $B \in L$, so $B^c \in L, A \in L, A \cup B^c \in L$. So, $(A \cup B^c)^c = B \cap A^c \in L$
b) L is both $\lambda$-system and $\pi$-system, then L is $\sigma$-algebra.
*Proof* $A, B \in L, (A\ (A \cap B)) \cup (A \cap B) \cup (B\ (A \cap B)) = A \cup B \in L$
Given $\{A_n\}_1^{\infty} \in L$, make $A'_n$ disjoint.

$$\cup_1^{\infty} A_i = \cup_1^{\infty} A'_i \in L$$

**Theorem** $(\pi - \lambda)$ Let $P$ be $\pi$ system, $L$ be $\lambda$ system, $P \subseteq L$, then $\sigma(P) \subseteq L$.

*Proof* Let $L_0$ = smallest $\lambda$ system containing $P$.
We show $L_0$ is $\pi$-system, so $\sigma(P) \subseteq L_0 \subseteq L$.
$A \in P$, let $L_A = \{B \subseteq \Omega : A \cap B \in L_0\}$.
**Claim** $L_A$ is $\lambda$-system

- $\Omega \in L_A$

- say $B \in L_A, A \cap B \in L_0, A \backslash (A \cap B) \in L_0$
  e.g. $A \cap (A^c \cup B^c) = A \cap B^c \in L_0$, so $B^c \in L_A$.

- $B_i$ disjoint in $L_A$, $B_i \cap A \in L_0 (\cup B_i) \cap A = \cup_1^\infty (B_i \cap A) \in L_0$
  so $\cup B_i \in L_A$.

**Claim**. $A \in P, L_0 \subseteq L_A$
Take $B \in P, A \cap B \in P$

$$P \subseteq L_A \to L_0 \subseteq L_A$$

Take $B \in L_0$ claim $L_0 \subseteq L_B$
*Proof* If $A \in P$, $B \in L_A$, so $A \cap B \in L_0$, so $A \in L_B$.
$P \subseteq L_B \to L_0 \subseteq L_B$
Finally, $B, C \in L_0, B \in L_0, C \in L_B, e.g. B \cap C \in L_0$
So $L_0$ is a $\pi$-system, $L_0$ is $\sigma$-algebra in L.
$\sigma(P) \subseteq L_0 \subseteq L$

**Corollary**.
If two probability measures P, Q agree on $\pi$-system P, then they agree on $\sigma(P)$
*Proof* Let $L = \{A : P(A) = Q(A)\}$
This is a $\lambda$-system if $P(A_i) = Q(A_i), 1 \leqslant i \leqslant 2$
$A_i \cap A_j = \varnothing$, then

$$P(\cup_1^\infty A_i) = Q(\cup_1^\infty A_i)$$

$\Rightarrow \sigma(P) \subset L$
So unique in Caratheodory,
If $P(F) = Q(F), F \in \mathcal{F}, P^*(A) = P^*(A), A \in M$
*Proof* $\mathcal{F}$ is a $\pi$-system.

**Independence**.
$(\Omega, \mathcal{F}, P)$ probability space,

**Definition**. $A, B \in \mathcal{F}$ independent
$P(A \cap B) = P(A)P(B)$.

**Definition**. $\{A_i\}_{i\in I}$ independent if any finite number of intersection:

$$P(A_{n_1} \cap \cdots \cap A_{n_m}) = \prod_{j=1}^{m} P(A_{n_j})$$

**Example**. $\Omega = (0,1], \omega = \sum_{i=1}^{\infty} \frac{d_i(\omega)}{2^i}$,
$A_i = \{d_i = 0\}, \{A_i\}_{i=1}^{\infty}$ are independent,
let $Q_i = \sigma(A_i) = \{\varnothing, \Omega, Q_i = 0, Q_i = 1\}$

$$Q_{odd} = \sigma(Q_{2i+1}, 0 \leqslant i < \infty)$$
$$Q_{even} = \sigma(Q_{2i}, 1 \leqslant i < \infty)$$

$Q_{odd}$ and $Q_{even}$ are independent, how to prove?

**Proposition**. $(\Omega, F, P)$ probability space.
$Q_i, 1 \leqslant 1 \leqslant I$ be independent $\pi$-system, then
$\sigma(Q_i), 1 \leqslant i \leqslant I$ are independent.

*Proof.* Let $\mathcal{B}_i = Q_i \cup \Omega$,
still $\pi$-system, still $\mathcal{B}_i$ independent.
Let $L_1 = \{B_1 \in \Omega : \forall B_j \subseteq \mathcal{B}_j, 0 \leqslant j \leqslant I, P(B_1 \cap B_2 \cdots \cap B_I) = \prod_{i=1}^{\infty} P(B_i)\}$
$\mathcal{B}_1 \subseteq L_1$.
$L_1$ is a $\lambda$-system.

- $\Omega$ in.
- $B \in L$, then $B^c \in L$
- closed under disjoint unions.

So $\sigma(B_1) \subseteq Ł_1$.
Now have independence
$\sigma(B_1), B_2, \cdots, B_I$ so $\sigma(B_1), \sigma(B_2), \cdots, \sigma(B_I)$ independent, done.

**Theorem** (Borel-Cantelli)
$(\Omega, \mathcal{F}, P)$ and $\{A_i\}_{i=1}^{\infty} \in \mathcal{F}$
1) if $\sum_{i=1}^{\infty} P(A_i) < \infty$
then $P\{A_i \text{ i.o.}\} = 0$

2) if $\sum_{i=1}^{\infty} P(A_i) = \infty$, and $A_i$ independent, then $P(A_i \text{ i.o.}) = 1$.

**Definition**. $A_n$ i.o. $= \cap_{n=1}^{\infty} \cup_{m=n}^{\infty} A_m = \{\omega \text{ with } A_i \text{ occuring infinitely often}\}$
*Proof* 1) Given $\varepsilon > 0$, choose $N$,

$$\sum_{i=N}^{\infty} P(A_i) < \varepsilon$$

$$P(A_i \text{ i.o.}) \leqslant P(\cup_{i=N}^{\infty} A_i) \leqslant \sum_{N}^{\infty} P(A_i) < \varepsilon$$

2) Use $1 - x \leqslant e^{-x}$, all $x > 0$
study $P((A_n \text{ i.o.})^c) = P(\cup_{n=1}^{\infty} \cap_{i=n}^{\infty} A_i^c)$.
Now $P(\cap_{i=n}^{\infty}) \leqslant P(\cap_{i=n}^{N} A_i^c) = \prod_{i=n}^{N} P(A_i^c) = \prod_{i=n}^{N}(1 - P(A_i)) \leqslant e^{-\sum_{i=n}^{N} P(A_i)}$
Let $N \to \infty$, R.H.S. $= 0$
so $P(\cap_{i=n}^{\infty} A_i^c) = 0$
so $P((A_n \text{ i.o.})^c) = 0 \Rightarrow P(A_i \text{ i.o.}) = 1$

Back to $\Omega = (0,1], \omega = \sum \frac{d_n}{2^n}, \Omega_i(\omega) = 2d_i(\omega) - 1$
Strong law showed $S_n = \sum_{i=1}^{n} \Omega_i$,

$$\lim \frac{S_n}{n} = 0, a.s.$$

Recall $x_n \in \mathbb{R}$,

$$\overline{\lim} x_n = \lim_{n\to\infty} \sup_{i \geqslant n} x_i$$

$$\underline{\lim} x_n = \lim_{n\to\infty} \inf_{i \geqslant n} x_i$$

$x_n$ has a limit $\Leftrightarrow \overline{\lim} x_n = \underline{\lim} x_n$

**Fact**. with probability 1, $\overline{\lim} S_n = +\infty, \underline{\lim} S_n = -\infty$.
and also $\overline{\lim} \frac{S_n}{\sqrt{n}} = \infty, \underline{\lim} \frac{S_n}{\sqrt{n}} = -\infty$.
What about $\frac{S_n}{n^{\frac{3}{4}}}$?
Hardy showed $\lim \frac{S_n}{\sqrt{n}\log n} = 0$

**Theorem** (Kinchine's Law of Iterated Logarithm) $\overline{\lim} \frac{S_n}{\sqrt{2n \log\log n}} = 1, \underline{\lim} \frac{S_n}{\sqrt{2n \log\log n}} = -1$

*Proof* Setup. $L = \{\omega : \overline{\lim} \frac{S_n}{g(n)} = 1\}$

$$g(n) = \sqrt{2n \log\log n}$$

$$L = (\cap_{c\in(0,1)\cap\mathbb{Q}} L_c) \cap \cap_{c\in(1,\infty)\cap\mathbb{Q}} L_c^c$$
$$L_c = \{\omega : S_n > cg(n) \text{ i.o.}\}.$$

We show

- $$\forall c < 1, P(L_c) = 1, (*)$$
- $$\forall c > 1, P(L_c) = 0. (**)$$

We know $P\{\frac{S_n}{\sqrt{n}} > a_n\} \leqslant e^{-\frac{a_n^2}{2}(1+o(1))}$
Large Deviations.
We prove (*), (**) by Borel-Cantelli.

For us, $a_n = (1+\varepsilon)\sqrt{2\log\log n}$
We choose $n = [\theta^k], \theta > 1$.
Get

$$\sum \frac{1}{k^{1+\varepsilon}} = \begin{cases} < \infty & \varepsilon > 0 \\ = \infty & \varepsilon < 0 \end{cases}$$

Problems:
a) Only have on subsequence (Interpolation)
b) Fix non-independence.

Why now?
a) Illustrate Borel-Cantelli in non-trivial way.
b) We meet

- Large deviations
- Max inequality
- Interpolation

c) Great theorem.

**Step** 1. maximum inequality
$S_0 = 0, m_n = \max_{0\leqslant k\leqslant n} S_k$

**Proposition**. $\forall$ integer $c > 0$,

$$
\begin{aligned}
P\{m_n \geqslant c\} &= P\{S_n \geqslant c\} + P\{S_n > c\} \\
&\leqslant 2P\{S_n \geqslant c\} (\text{Kolmogorov's inequality})
\end{aligned}
$$

*Proof*

$$
\begin{aligned}
P\{m_n \geqslant c\} &= P\{m_n \geqslant c, S_n \geqslant c\} + P\{m_n \geqslant c, S_n < c\} \\
&= I + II
\end{aligned}
$$

$$I = P\{S_n \geqslant c\}$$

For II, break space
At first time j, that $S_j = c$.

$$
\begin{aligned}
F_j &= \{\omega : S_1 < C, S_2 < c, \cdots, S_j = c\} \\
P\{m_n \geqslant c, S_n < c\} &= \sum_j P\{m_n \geqslant c \text{ and } F_j\} \\
&= \sum_j P\{F_j, S_n - S_j < 0\} \\
&= \sum_j P\{F_j\} P\{S_n - S_j < 0\} \\
&= \sum_j P\{F_j\} P\{S_n - S_j > 0\} \\
&= \sum_j P\{F_j \text{ and } S_n - S_j > 0\} \\
&= P\{S_n > c\}
\end{aligned}
$$

**Step** 2. $P\{L_c\} = 0, c > 1, L_c = \{\omega : S_n > cg(n) i.o.\}$
Note $\sum P(S_n > cg(n)) \leqslant \sum \frac{1}{(\log n)^2} = \infty$
So Borel-Cantelli doesn't work, use subsequence:

$$3 < n_1 < n_2 < \cdots, g(n) \uparrow \infty$$

$$\{\omega : S_n(\omega) \geqslant cg(n), \text{ some n}; n_{k-1} < n \leqslant n_k\} \subseteq \{\omega : m_{n_k} > cg(n_{k-1})\} \triangleq A_k$$

$$L_c \subseteq A_k, \text{ i.o.}$$

O.K.

$$\begin{aligned} P\{A_k\} &= P\{\omega : m_{n_k} > cg(n_{k-1})\} \\ &\leqslant 2P\{\omega : S_{n_k} \geqslant cg(n_{k-1})\} \\ &\leqslant 2e^{-c^2\frac{g^2(n_{k-1})}{2n_k}(1+o(1))}(L.D.) \\ &= 2e^{-c^2\frac{n_{k-1}}{n_k}\log\log(n_{k-1})(1+o(1))} \\ &= 2\frac{1}{(\log n_{k-1})^{c^2\frac{n_{k-1}}{n_k}}} \end{aligned}$$

Choose $\theta \in (1, c^2)$
Set $n_k = [\theta^k] \uparrow \infty$

$$P\{A_k\} \leqslant 2\frac{1}{(k\log\theta)^{(1+o(1))\frac{c^2}{\theta}}}$$

$$\sum P(A_k) < \infty, \frac{c^2}{\theta} > 1.$$

Done.

**Step** 3. $P(L_c) = 1, 0 < c < 1$
Given $c < 1$, let $\gamma = \frac{c+1}{2}, \varepsilon = \frac{c+1}{2} - c$
**Claim**. $L_c \supseteq \{\omega : S_{n_k} \geqslant cg(n_k),\ \text{i.o.}\} \supseteq \{A_k \cap B_k,\ \text{i.o.}\}$

$$A_k = \{\omega : |S_{n_{k-1}}| \leqslant \varepsilon g(n_k)\}$$

$$B_k = \{\omega : S_{n_k} - S_{n_{k-1}} \geqslant \gamma g(n_k)\}$$

*Proof* $\omega \in A_n \cap B_n$, $S_{n_k}(\omega) = S_{n_k} - S_{n_{k-1}} + S_{n_{k-1}} \geqslant ((c+\varepsilon) - \varepsilon)g(n_k) = cg(n_k)$

Next. $\{A_n \cap B_n,\ \text{i.o.}\} \supset \{A_n \text{ all large n}\} \cap \{B_n \text{ i.o.}\}$
We show both pieces have probability 1.
For $A_k$,

$$\frac{S_{n_{k-1}}}{g(n_k)} = \frac{S_{n_{k-1}}}{g(n_{k-1})}\frac{g(n_{k-1})}{g(n_k)} \xrightarrow{p} 0$$

From step 2. $|S_{n_{k-1}}| \leqslant 2g(n_{k-1})$ has probability 1.
For all large n, we will choose $n_k \uparrow \infty$, $\frac{n_k}{n_{k-1}} \to \infty$
For $B_k$, they are independent.
Show $\sum P(B_k) = \infty$, so $B_k i.o.$ with probability 1.

$S_{n_k} - S_{n_{k-1}} \overset{L}{=} S_{n_k - n_{k-1}}$

$$
\begin{aligned}
P\{B_{n_k}\} &= P\{\omega : S_{n_k - n_{k-1}} \geqslant \gamma g(n_k)\} \\
&\leqslant e^{-(1+o(1))\gamma^2 g^2(n_k)/2(n_k - n_{k-1})} \\
&= e^{-(1+o(1))\gamma^2 (\log\log n_k)\frac{n_k}{n_k - n_{k-1}}}, \ \frac{n_k}{n_{k-1}} \uparrow \infty \\
&= \frac{1}{(\log n_k)(1+o(1))\gamma^2}
\end{aligned}
$$

Choose $\theta \in (1, \frac{1}{\gamma^2})$, set $n_k = e^{k^\theta}$

$$
\sum_k \frac{1}{k^{(1+o(1))\gamma^2\theta}} = \infty
$$

$\gamma^2\theta < 1$

Remarks.
i) We proved for $\pm1$ probability $\frac{1}{2}$.
Same theorem true for any mean 0, variance $\sigma^2$, independent

$$
\overline{\lim} \frac{S_n}{\sqrt{2\sigma^2 n \log\log n}} = 1
$$

$$
\text{e.g. if } P(X_n = 1) = p, P(X_n = -1) = q
$$

$$
\overline{\lim} \frac{S_n - n(p-q)}{\sqrt{2pqn \log\log n}} = 1
$$

ii) Erdǿs proved coin tossing

$$
P\{\frac{S_n}{\sigma\sqrt{n}} > d(n) \text{ i.o.}\} = \begin{cases} 0 & \sum \frac{d(n)}{n} e^{-\frac{1}{2}d(n)^2} < \infty \\ 1 & \sum \frac{d(n)}{n} e^{-\frac{1}{2}d(n)^2} = \infty \end{cases}
$$

$$
\Rightarrow \overline{\lim} \left( \frac{S_n - n(p-q)}{\sqrt{npq}} - \sqrt{2\log\log n} \right) \frac{\sqrt{2\log\log n}}{\log\log\log n} = \frac{3}{2}.
$$

iii) Ordinary LIL.
If $\varepsilon > 0$,

$$
\frac{S_n}{\sqrt{2n\log\log n}} > (1+\varepsilon)
$$

Only finitely often.
Can look at distribution of last time it's $> 1+\varepsilon$
iv) In statistical testing problems

people look at $\{\frac{S_n}{\sqrt{n}} > .96\}$
If so LIL shows wait long enough sure of success.

$$:)\ \log\log 10^{100} = 5.43$$

**Distribution functions on $\mathbb{R}^d$**
$d = 1$, say $\mu$ is probability on $\mathbb{R}$.
Define $F(x) = \mu(-\infty, x]$
Note $(-\infty, x]$ is $\pi$-system.
So $F(x)$ determines $\mu$.

- $$F(-\infty) = \lim_{x\to\infty} F(x) = 0$$
- $$F(\infty) = \lim_{x\to\infty} F(x) = 1$$
- $$F(x) \leqslant F(y), x \leqslant y$$
- $$x_n \downarrow x, F(x_n) = \mu(-\infty, x_n] \to \mu(-\infty, x] = F(x).$$

**Example**.

$$\mu(A) = \delta_0(A) = \begin{cases} 1 & 0 \in A \\ 0 & \text{else} \end{cases}$$

Converse if $F(x)$ satisfies the four conditions,then exists unique

$$\mu(x, y] = F(y) - F(x)$$

Right continuity

$$d = 2, x \in \mathbb{R}^2$$

$\mu$ probability on $\mathbb{R}^2$

$$F(x) = \mu(A_x)$$

$A_x = \{y_1, y_2 : y_i \leqslant x_i, i = 1, 2\}$
This satisfies the four conditions, but not enough!

BOX: $w, x, \xi, y$

$$F(\text{BOX}) = F(A_x) - F(A_w) - F(A_y) + F(A_\xi)$$

Need this $\geqslant 0, \forall$ Box.

H.W.Problem. Let $F(x), G(y)$ be distribution functions on $\mathbb{R}$,

$$U(x,y) = F(x) \wedge G(y)$$
$$L(x,y) = (F(x) + G(y) - 1)_+$$

show these are bivariate distribution functions with margins:

$$U(x,\infty) = F(x)$$
$$U(\infty,y) = G(y)$$
$$L(x,\infty) = F(x)$$
$$L(\infty,y) = G(y)$$

and $\forall H(x,y)$ distribution function with margins $F(x), G(y)$
$L(x,y) \leqslant H(x,y) \leqslant U(x,y)$.
Correlations are extremal.
Application make multivariate distributions without needing normalized constraint:

$$d = 2, F_1(x)F_2(y)F_3(x+y) = A(x,y)$$

General d.

$$A = \{x \in \mathbb{R}^d : a_i < x \leqslant b_i\}$$

Vertices are $V, V_i$ coordinates are $a_i, b_i$
$sgn(V) = (-1)^i, i = \#$ of times $a_i$ appears.
e.g. $n = 2, sgn(a_1, b_2) = -1$
$F(x)$ is real function on $\mathbb{R}^d$

$$\Delta_A F = \sum_V sgn(V) F(V)$$

$$d = 1 : \Delta_A F = F(b) - F(a)$$

**Theorem**. Suppose $F(x)$ is continuous from above,

$\Delta_A F \geqslant 0, \forall$ bounded rectangle:

$$\lim_{x \to -\infty} F(x) = 0, \lim_{x \to \infty} F(x) = 1$$

Then $\exists$ unique probability $\mu$ on Borel sets of $\mathbb{R}^d$,

$$\mu(A) = \Delta_A(F)$$

$\infty$ measures
say $\Omega$ set, $\mathcal{F} : \sigma$-algebra of subsets.
A measure is

$$\mu : \mathcal{F} \to [0, \infty]$$

- $$\mu(\varnothing) = 0$$
- $$A \subseteq B, \mu(A) \leqslant \mu(B)(\text{monotonic})$$
- $$\{A_i\} \text{ disjoint in } \mathcal{F}, \mu(\cup_1^\infty A_i) = \sum_1^\infty \mu(A_i)(\text{countable additivity})$$

Why?

- When we write
  $$\mu(0, x] = \int_0^x e^{-y} \mathrm{dy}$$
  "$\mu$y" is lebesgue on $\mathbb{R}$
- Free (did all work)
- Calderon

Many of theorems same.

**Example**. If $\{A_i\}_{i=1}^{\infty}, A_i \uparrow A$

$$\mu(A) = \lim_i \mu(A_i)$$

*Proof.* $B_i = A_i \backslash A_{i-1}$

$$\mu(A) = \mu(\cup_1^\infty B_i) = \sum_1^\infty \mu(B_i) = \lim \sum_1^n \mu(B_i) = \lim \mu(A_n)$$

Watch it. For probabilities,

$$A \subseteq B, P(B \backslash A) = P(B) - P(A)$$

and $A_n \downarrow A, P(A) = \lim P(A_i)$
But for example length on $\mathbb{R}$,

$$A_n = (-\infty, -n], A_n \downarrow \varnothing$$

$$\lambda(A_n) = \infty, \forall n, \text{ not } \lambda(\phi) = 0$$

**Definition**. $\mu$ on $(\Omega, \mathcal{F})$ is $\sigma$-finite if

$$\exists B_i \in \mathcal{F}, \cup_1^\infty B_i = \Omega, \text{ and } \mu(B_i) < \infty \text{ all i.}$$

**Example**. $\lambda$ on $\mathbb{R}$ is $\sigma$ finite.
**Example**. On $(0,1]$, $\mu(A) = \#\{\text{points in A}\}$ not $\sigma$-finite
Uniqueness $(\Omega, \mathcal{F})$, $\mu_1$, $\mu_2$ measures on $\mathcal{F}$, say P is a $\pi$-system in $\mathcal{F}$, and

$$\exists B_i \in P, \cup_1^\infty B_i = \Omega, \mu_i(B_j) < \infty, \text{ all } i, j$$

**Theorem**. If $\mu_1(B) = \mu_2(B), B \in P$, then

$$\mu_1 = \mu_2 \text{ on } \sigma(P)$$

*Proof.* Pick $F \in P$, $\mu_1(F) = \mu_2(F) < \infty$, $\nu_i(A) = \mu_i(A \cap F), i = 1, 2$
$\nu_1 = \nu_2$ on $P$, and if

$$L = \{A \in \mathcal{F} : \nu_1(A) = \nu_2(A)\}$$

$\lambda$-system, $\pi - \lambda$ says: $\nu_1 = \nu_2$ on $\sigma(P)$
If $\Omega = \cup_{i=1}^\infty B_i, B_i \in P$

$$\mu_1(B_j) = \mu_2(B_j) < \infty$$

Let $A_1 = B_1, A_2 = B_2 \backslash A_1, \cdots$

$$\begin{aligned}\mu_1(A) &= \mu_1(A \cap \cup A_i)\\ &= \sum \mu_1(A \cap A_i)\\ &= \sum \mu_2(A \cap A_i)\\ &= \mu_2(A)\end{aligned}$$

**Example**. On $(0,1]$, $P = \{\varnothing\}$
$\mu_1(\varnothing) = 0, \mu_1(\Omega) = \infty$
$\mu_2(\varnothing) = 0, \mu_2(\Omega) = 1$
agree on $P$, not on $\sigma(P)$

**Outer measure**
$\Omega$ any set
Q any collection of subsets.
$\mu : Q \to [0,\infty]$ any set function. ($\mu(\varnothing) = 0$)
**Definition**. Any $F \subseteq \Omega$,

$$\mu^*(F) = \inf \sum_{i=1}^{\infty} \mu(A_i)$$

Inf over converings of F by sets in Q.
$F \subseteq \cup A_i$, and $\infty$ if no such cover.
**Definition**. An outer measure on $\Omega$ is a function $\mu^*$ all subsets to $[0,\infty]$

- $$\mu^*(\varnothing) = 0$$
- $$A \subseteq B, \mu^*(A) \leqslant \mu^*(B)$$
- $$A \subset \cup_1^\infty A_i, \mu^*(A) \leqslant \sum_1^\infty \mu^*(A_i)$$

Check. Any example $\mu^*$ is an outer measure.
For 3) If some $A_i$ has $\mu(A_i) = \infty$, then done
and if all $\mu^*(A_i) < \infty$, then $\forall \varepsilon > 0, i, \exists \{A_j\}_{j=1}^\infty A_{i,j} \in Q$

$$\sum_j \mu(A_{i,j}) \leqslant \mu^*(A_i) + \frac{\varepsilon}{2^i}$$

$\cup A_{i,j}$ cover $A$,

$$\mu^*(\cup A_i) \leqslant_{i,j} \mu(A_{i,j}) + \frac{\varepsilon}{2^i} \leqslant \sum \mu^*(A_i) + \varepsilon.$$

**Example**. Hausdorff measure.
$\Omega, d$ be a metric space,
$\forall r, \varepsilon > 0$, define $A \subseteq \Omega$,

$$h^*_{r,\varepsilon}(A) = \inf c(r) \sum_{i=1}^{\infty} \left( \frac{\operatorname{diam}(B_i)}{2} \right)^r$$

Inf over coverings of A by balls of diameter $\leqslant \varepsilon$

$$c(r) = \frac{\pi^{\frac{r}{2}}}{\Gamma(\frac{r}{2} + 1)}$$
$$h^*_r(F) = \inf_{\varepsilon > 0} h^*_{r,\varepsilon}(F)$$

(on $\mathbb{R}^d$, $r = d$, this is $\lambda$)
**Measurable functions**.

$$(\Omega, \mathcal{F}), (\Omega', \mathcal{F}')$$

**Definition**. $T : \Omega \to \Omega'$ is measurable if $\forall A' \in \mathcal{F}', T^{-1}(A') \in \mathcal{F}$
Recall. $T^{-1}(A') = \{\omega : T(\omega) \in A'\}$
Check. $T^{-1}(\cup_i A'_i) = \cup_i T^{-1}(A'_i)$

$$T^{-1}(A^c) = T^{-1}(A)^c$$
$$T^{-1}(\cap A'_i) = \cap T(A'_i)$$
$$A' \subseteq B' \Rightarrow T^{-1}(A') \subseteq T^{-1}(B')$$

**Proposition**.
i) If $\mathcal{F}' = \sigma(\mathcal{A}')$
then $T$ is measurable $\Leftrightarrow T^{-1}(A') \in \mathcal{F}, \forall A' \in \mathcal{A}'$
ii) $(\Omega, \mathcal{F}) \xrightarrow{T_1} (\Omega', \mathcal{F}') \xrightarrow{T_2} (\Omega'', \mathcal{F}'')$

$$T_1 \text{ and } T_2 \text{ are measurable } \Rightarrow T_2 \circ T_1 : \Omega \to \mathcal{F}'' \text{ is measurable.}$$

*Proof* If conditions holds,
1) Let $\mathcal{B}' = \{B' \in \Omega' : T^{-1}(B') \in \mathcal{F}\}$. This contains $\mathcal{A}'$.
$Q$ is closed under unions and complement, countable additivity.
2) $(T_2 \circ T_1)^{-1}(\mathcal{F}'') = \{\omega : T_2(T_1(\omega)) \in \mathcal{F}''\}\} = T_1^{-1}(T_2^{-1}(\mathcal{F}'')) \in \mathcal{F}$.

**Definition**. A random variable is a real valued measurable function.

$$\forall \Omega, \mathcal{F}, \text{ and } X : \Omega \to (\mathbb{R}, \text{ Borel}), X \text{ is measurable}$$

**Definition**. A random vector is a measurable function in $\mathbb{R}^n$(Borel)

**Theorem**. $T : (\Omega, F) \to \mathbb{R}^n$ is a random vector $\Leftrightarrow$ $T_i(\omega)$ is a random variable.
*Proof* Say, each $T_i$ is measurable.

$$\{\omega : T(\omega) \leqslant (x_1, \cdots, x_n)\} = \cap_{i=1}^n \{\omega : T_i(\omega) \leqslant x\}$$

If $T$ is measurable,

$$\{\omega : T_i(\omega \leqslant x\} = \cup_{n\in\mathbb{Z}} \{\omega : T(\omega) \leqslant (n, \cdots, x, n, \cdots, n)\}$$

**Theorem**. If $T : \mathbb{R}^a \to \mathbb{R}^b$ is continuous, then $T$ is measurable.

Proof. T continuous $\Leftrightarrow$ $\forall$ closed set $B \in \mathbb{R}^b$, $T^{-1}(B)$ is closed.
So $\mathbb{R}^2 \to \mathbb{R}, (x, y) \to x + y, (x, y) \to xy, (x, y) \to \min(\max)(x, y)$ are measurable.
So if $X_1, X_2$ are random variables, $X_1 + X_2, X_1/X_2, X_1 \cap X_2, X_1 \cup X_2$ are random variables.

$$\Omega \to \mathbb{R} \times \mathbb{R} \to \mathbb{R}$$

So if $X_i, 1 \leqslant i < \infty$ are random variables, then

$$\{\sup(X_i) \leqslant x\} = \cap_{i=1}^\infty \{\omega : X_i \leqslant x\}$$

So sup measurable, so $\overline{\lim} X_i$ is measurable.
So $\underline{\lim} X_n$ is measurable, so $\{\omega : \lim X_n \exists\}$ is measurable.

**Example**. Not everything is measurable.
"Lebesgue mistake"
Let $A$ be a Borel set in $\mathbb{R}$.
$\pi : \mathbb{R}^2 \to \mathbb{R}$, then $\pi(A)$ need not be measurable.
New measures from old.
$(\Omega, F, \mu)$ measure space. $T : (\Omega, F) \to (\Omega', F')$ measurable.
Define $\mu^{T^{-1}}(B') = \mu(T^{-1}(B'))$

Check $\mu^{T^{-1}}$ is a measure.

$$\mu^{T^{-1}}(\varnothing) = \mu\{\omega : T(\omega) \in \varnothing\} = \mu(\varnothing) = 0$$
$$\mu^{T^{-1}}(\cup A_i) = \mu(T^{-1}(\cup A_i)) = \mu(\cup T^{-1}(A_i))$$

**Example**. Haar measure on $O_n$

$$O_n = \{M : n \times n \text{ real } MM^T = I\}$$

**Fact** $\exists$ a probability $\lambda$ on $O_n$
$\lambda(MA) = \lambda(A), \forall M \in O_n, A \subseteq O_n$
How to pick $M \in O_n$ from haar measure?
Let $Z_{i,j}, 1 \leqslant i,j \leqslant n$ be independent standard normal random variables.
Let $A_{i,j} = Z_{i,j}, A \xrightarrow{C.S.} M$ this is Haar new $\sigma-$algebra from old.
If $X_i, 1 \leqslant i < I$ are random variables.
$\sigma(\{X_i\}_{i \leqslant I})$ is a $\sigma$ algebra generated by $\{\omega : X_i \leqslant r_j\}$.

$\{X_i\}_{i \in I}, \{Y_j\}_{j \leqslant J}$ are random variables, they are independent if
$\sigma(\{X_i\}_{i \in I})$ independent of $\sigma(\{Y_j\}_{j \in J})$

**Example**. Let $X_i, 1 \leqslant i < \infty$ be independent uniform random variables.
Find distribution of $M_n = \max(X_i)$
$X_i$ is uniform if $X_i : \Omega \to (0,1]$ has $P^{x_i^{-1}}(A) = \lambda(A)$

$$P(M_n \leqslant x) = P(\max X_n \leqslant x) = P(X_1 \leqslant x)^n = x^n$$

Let's approximate for understanding

$$P(M_n \leqslant 1 - \frac{c}{n}) = (1 - \frac{c}{n})^n \sim e^{-C}$$

**Example**. Let $X_n$ be independent standard normal random variables.

$$P(X_i \leqslant x) = \int_{-\infty}^{x} \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} \mathrm{dt} = \Phi(x)$$
$$\begin{aligned} P(M_n \leqslant x) &= P(X_i \leqslant x, 1 \leqslant i \leqslant n)(\ast) \\ &= \Phi(x)^n \\ &= e^{n \log \Phi(x)} \\ &= e^{n \log(1-(1-\Phi(x)))} \end{aligned}$$

H.W. show that $\forall x > 0$,

$$\frac{xe^{-\frac{x^2}{2}}}{\sqrt{2\pi}(1+x^2} \leqslant \int_x^\infty \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} \mathrm{dt} \leqslant \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}x}$$

So $1 - \Phi(x) \sim (x \text{ large}) \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}x}$

$$* = e^{-n\frac{e^{-x^2}2}{\sqrt{2\pi}x}}$$

take $x = \sqrt{2\log n - \log\log n + C}$

$$* \sim e^{-n\frac{e^{-\frac{1}{2}(2\log n - \log\log n + C}}{\sqrt{2\pi}\sqrt{2\log n}}} = e^{-\frac{e^{-\frac{C}{2}}}{\sqrt{2\pi}}}$$

Double exponential.

**Theorem**. $P\{M_n \leqslant \sqrt{2\log n - \log\log n + C}\} \sim e^{-\frac{e^{-c/2}}{\sqrt{2\pi}}}$

Whats $(\Omega, F)$ and where is?
We build this, recall $\Omega = (0,1]$ Borel $\lambda$
We use this, We have $\omega = \sum_{i=1}^\infty \frac{d_i(\omega)}{2^i}, d_i \in \{0,1\}$
$d_i$ are random variables.
We said $\{d_{2i}\}$ and $\{d_{2i+1}\}$ are independent, therefore,

$$u_0 = \sum_1^\infty \frac{d_{2i}(\omega)}{2^i}$$

and

$$u_1 = \sum_1^\infty \frac{d_{2i+1}(\omega)}{2^i}$$

and both uniform.
Observe. If $F(x)$ is a distribution function on $\mathbb{R}$ and say F continuous strictly increasing. Then $F^{-1}$ exists,
$F(F^{-1}(x)) = x$, U is uniform.

$$P(F^{-1}(U) \leqslant x\} = P(u \leqslant F(x)) = F(x)$$

$$X(\omega) = F^{-1}(\omega) \text{ distributes as } F(x)$$

**Example**. Let $X_1, \cdots$ be i.i.d. standard exponential random variables.

$$P(X_i \geqslant x) = e^{-x}$$

Take $T(y) = -\log y$,

$$\begin{aligned}P(T(U) > x) &= P(-x > \log U)\\ &= P(e^{-x} > U)\\ &= e^{-x}\end{aligned}$$

**Integration(Expectation)**
$(\Omega, \mathcal{F}, M)$ measure space
$f : \Omega \to \mathbb{R} \cup \{\infty\}$ measurable.
Want to define $\int f \mathrm{d}\mu$
Why. In probability $\mu$dx lebesgue.
On $\mathbb{R}, f$ probability density.
Our $\int$ is "better" than Riemann.($L^2$ is complete)
Here $f$ allowed to be $\pm\infty$, convention

$$0 \cdot \infty = 0$$

$\infty - \infty$ not defined.
Use step functions $f : \Omega \to \mathbb{R} \cup \{\pm\infty\}$ taking only finitely many values.

$$f(\omega) = \sum_{i=1}^{N} x_i \delta_{A_i}(\omega), A_i = \{\omega : f(\omega) = x_i\}$$

$\pm\infty$ allowed.

**Fact**. $\forall$ measurable $f, \exists$ step functions $f_n$ s.t.

$$f_n(\omega) \geqslant 0, 0 \leqslant f_n(\omega) \uparrow f(\omega)$$

or

$$f(\omega) < 0, f_n \downarrow f(\omega)$$

*Proof*

$$f_n(\omega) = \begin{cases} -n & \text{If } -\infty \leqslant f(\omega) \leqslant -n \\ -\frac{h-1}{2^n} & \text{If } -\frac{h}{2^n} < f(\omega) \leqslant -\frac{h-1}{2^n}, 1 \leqslant h \leqslant n2^n \\ \frac{h-1}{2^n} & \text{If } \frac{h-1}{2^n} \leqslant f(\omega) < \frac{h}{2^n}, 1 \leqslant h \leqslant n2^n \\ n & \text{If } n \leqslant f(\omega) \leqslant \infty \end{cases}$$

**Definition**. for $f \geqslant 0$ All $\omega$.

$$\int f(\omega)\mu(\mathrm{d}\omega) = \sup \sum_{i=1}^{N} \inf_{\omega \in A_i} f(\omega) \delta_{A_i}(\omega)$$

$A_i$ partition of $\Omega$.
Write any $f$ as $f_+ - f_-$

$$f_+(\omega) = \max(f(\omega), 0), f_-(\omega) = \max(-f(\omega), 0)$$

**Definition** $\int f \mathrm{d}\mu = \int f_+ \mathrm{d}\mu - \int f_- \mathrm{d}\mu$. If not $\infty - \infty$, else $f$ not integrable.
**Theorem**. $f \geqslant 0$ all $\omega$, then
1) If $f$ is a step function,

$$\int f \mathrm{d}\mu = \sum_1^N x_i \mu(A_i)$$

2) $f(\omega) \leqslant g(\omega)$, then

$$\int f \mathrm{d}\mu \leqslant \int g \mathrm{d}\mu$$

3) (Monotone Convergence Theorem)
$0 \leqslant f_n(\omega) \uparrow f(\omega)$ all $\omega$, then

$$\lim_n \int f_n \mathrm{d}\mu = \int \lim_n f_n \mathrm{d}\mu$$

4) $f, g \geqslant 0, \alpha, \beta \geqslant 0, \int (\alpha f + \beta g) \mathrm{d}\mu = \alpha \int f \mathrm{d}\mu + \beta \int g \mathrm{d}\mu$

*Proof*
1) Say $f = \sum_1^n x_i \delta_{A_i}$, Let $\{B_j\}_{j=1}^m$ decomposition.
$\beta_j = \inf_{\omega \in B_j} f(\omega)$
If $A_i \cap B_j \neq \varnothing, \beta_j \leqslant x_i$

$$\begin{aligned}
\sum_j \beta_j \mu(B_j) &= \sum_{i,j} \beta_j \mu(A_i \cap B_j) \\
&\leqslant \sum_{i,j} x_i \mu(A_i \cap B_j) \\
&= \sum_i x_i \mu(A_i)
\end{aligned}$$

$\int f \mathrm{d}\mu \leqslant \sum x_i \mu(A_i)$
Other direction free.
2) Free.
3) By 2),

$$\lim \int f_n \mathrm{d}\mu \leqslant \int f \mathrm{d}\mu$$

Need $\int f \mathrm{d}\mu \leqslant \lim \int f_n \mathrm{d}\mu$,
Only need $S = \sum_{i=1}^n \nu_i \mu(A_i) \leqslant \lim_n \int f_n$
$\forall$ decomposition $\{A_i\}$, $\nu_i = \inf_{\omega \in A_i} f(\omega)$

**Step 1**. Say $S < \infty$, each $\nu_i, \mu(A_i) > 0, < \infty$
Choose $\varepsilon > 0, \varepsilon < \nu_i$ all $i$.
Let $A_{n,i} = \{\omega \in A_i : f_n(\omega) \geqslant \nu_i - \varepsilon\}$
$f_n(\omega) \uparrow f(\omega) \Rightarrow A_{n,i} \uparrow A_i$, so $\mu(A_{n,i}) \uparrow \mu(A_i)$
Now

$$\begin{aligned}\int f_n \mathrm{d}\mu &\geqslant \sum_{i=1}^{N} (\nu_i - \varepsilon)\mu(A_{n,i}) \\ &\to \sum_{i=1}^{N} \nu_i \mu(A_i) - \varepsilon \sum_{i=1}^{N} \mu(A_i)\end{aligned}$$

Let $\varepsilon \downarrow 0, \lim \int f_n \mathrm{d}\mu \geqslant \sum_{i=1}^{N} \nu_i \mu(A_i)$.

**Step 2** $s < \infty$, some $nu_i$ on $\mu(A_i)$ might be 0 or $\infty$,
Say for $1 \leqslant i \leqslant m, \nu_i, \mu(A_i) > 0, < \infty$
Decompose into $A_{m,n}$ as above.
$1 \leqslant i \leqslant m$, and $(\cup^m A_{n,i})^c$

**Step 3**. $S = \infty$, some $i$, $\nu_i \mu(A_i) = \infty$
and $\nu_i \mu(A_i) > 0$
Choose $x, y, 0 < x < \nu_i, 0 < y < \mu(A_i)$
Let $A_{n,i} = \{\omega \in A_i, f_n(\omega) \geqslant x\}$
$f_n \uparrow f$, $A_{n,i} \uparrow A_i = \{\omega : f(\omega) \geqslant x\}$
$\mu(A_n, i) \uparrow \mu(A_i)$
$\int f_n \mathrm{d}\mu \geqslant x\mu(A_{n,i}) \uparrow xy$. Let $x$ or $y \uparrow \infty$

$$\lim_n \int f_n \mathrm{d}\mu \geqslant xy$$

4) Linearity.
Say $f = \sum_{i=1}^{N} x_i \delta_{A_i}$, $g = \sum_{j=1}^{m} y_j \delta_{\beta_j}, \alpha, \beta \geqslant 0$

$$\begin{aligned}\int (\alpha f + \beta g) \mathrm{d}\mu &= \sum_{i,j} (\alpha x_i + \beta y_j) \mu(A_i \cap B_j) \\ &= \alpha \sum_i x_i \mu(A_i) + \beta \sum_j y_j \mu(B_j)\end{aligned}$$

General, $f, g \geqslant 0$, Take $f_n \uparrow f$, $g_n \uparrow g$
Step functions and use 3).

Remarks.
1) This generalized Riemann integral.

$(\Omega, \mathcal{F}, M) = (0,1], \lambda, f(\omega) = \delta_{Rationals}$
Riemann integral.
$\int_0^1 f \mathrm{d}\lambda$ doesn't exists and our $\int f \mathrm{d}\lambda = 0$
2) But $\int_0^\infty \frac{\sin x}{x}$dx is <u>not</u> lebesgue integrable but <u>is</u> Riemann integrable.
Airy function. $A(x) = \frac{1}{\pi}\int_0^\infty \cos(u^3 - xu)$du
3) There is an integral including both kurtzweic-henstock integral.
American Math Monthly 1987 Page 450

**Fatou's lemma**. $(\Omega, F, \mu)$ given,
$f_n(\omega) \geqslant 0$, then

$$\int \underline{\lim} f_n(\omega) \mathrm{d}\mu \leqslant \underline{\lim} \int f_n \mathrm{d}\mu$$

*Proof.* Let $g_n(\omega) = \inf_{k \geqslant n} f_k(\omega) \uparrow \underline{\lim} f_n(\omega)$.
So $\int g_n \uparrow \int \underline{\lim} f_n \mathrm{d}\mu$, also

$$g_n(\omega) \leqslant f_n(\omega)$$

So

$$\int g_n \leqslant \int f_n$$

Take $\underline{\lim}$ both sides.

$$\int \underline{\lim} f_n \leqslant \underline{\lim} \int f_n \mathrm{d}\mu$$

**Example**.
1) $\Omega = (0,1]$,

$$f_n = \begin{cases} I(x > \frac{1}{2}) & n \text{ even} \\ I(x \leqslant \frac{1}{2}) & n \text{ odd} \end{cases}$$

$\underline{\lim} f_n = 0$, $\int f_n = \frac{1}{2}$

$$0 = \int 0 \leqslant \frac{1}{2} = \underline{\lim} \int f_n$$

2) Positivity measures on $(0, \infty)$,

$$f_n = \begin{cases} -1 & \text{on } (n, n+1) \\ 0 & \text{else} \end{cases}$$

$f_n(\omega) \to 0$ all $\omega$
$0 \leqslant -1$? No!

3) On $(0,1]$,

$$f_n = \begin{cases} n^2 & \text{on } (0,\frac{1}{n}) \\ 0 & \text{else} \end{cases}$$

$f_n(\omega) \to 0, \forall\omega$, but $\int f_n = n \uparrow \infty$

$$\int \lim \neq \lim \int$$

Useful to know
If for example $f_n(\omega) \geqslant 0$, a. e. $\mu$, then

$$\int \underline{\lim} f_n \leqslant \underline{\lim} \int f_n$$

**Example**. Fatou. $(0,1]$. Let $\{A_i\}_{i=1}^{\infty}$ be enumeration of rationals.

$$0, 1, \frac{1}{2}, \frac{1}{3}, \frac{2}{3}. \frac{1}{4}, \frac{3}{4}, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}$$

Consider $f(x) = \sum_{i=1}^{\infty} \frac{1}{i^2\sqrt{|x-A_i|}}$
This is $\infty$ all $A_i$
But $f(x) = \lim_n \sum_1^n \frac{1}{i^2}\sqrt{n-A_i}$
and

$$\int_0^{\infty} \leqslant \lim_n \sum \frac{1}{i^2} \int_0^1 \frac{1}{|x-A_i|} \mathrm{dx} < \infty$$

$$\therefore f(x) \text{ is } < \infty \text{ a.e.} x.$$

**Problem** Find single $x$, so $f(x) < \infty$

$$p_n = P\{m_n = a_n\}, q_n = 1 - p_n$$
$$\nu_n(a_n) = e^{-\{\log_2 n\}}$$

**Question**: What's a natural condition for two probabillity measure $\mu_n, \nu_n$ to merge even if don't converge?
One idea:
Pick metric for weak convergences.
$\mu_n, \nu_n$ merge: $d(\mu_n, \nu_n) \to 0$
OR
$\mu_n$ and $\nu_n$ merge $\Leftrightarrow$

$$\forall \text{ bounded continuous function }, \left| \int f \mathrm{d}\mu_\mathrm{n} - \int f \mathrm{d}\mu \right| \to 0$$

**Example**. Say $\mu_n = \delta_n, \nu_n = \delta_{n+\frac{1}{n}}$
Easy to see any standard metric with $d(\mu_n, \nu_n) \to 0$

But $\exists f$ s.t.

$$\int f \mathrm{d}\mu_{\mathrm{n}} \to 1, \int f \mathrm{d}\mu_{\mathrm{n}} \to 0$$

See Persi's paper with freedman. "Uniformities compatible with weak* topology"

Last day, built (on $(\Omega, \mathcal{F}, \mu)$)
$\int f \mathrm{d}\mu$, monotone $\uparrow$, Fatou $\int \underline{\lim} \leqslant \underline{\lim} \int, f \geqslant 0$
**Theorem**. (Dominated convergence)

$$(\Omega, \mathcal{F}, \mu), f_n, g \text{ measurable}$$

- $$|f_n| \leqslant g, \text{ all n}$$
- $$g \text{ integrable}$$
- $$f_n(\omega) \to f(\omega), n \to \infty$$

Then $f$ is integrable and

$$\int f \mathrm{d}\mu = \lim_n \int f_n \mathrm{d}\mu$$

*Proof*
$f_n^+, f_n^- \leqslant g$, so $f^* = \overline{\lim} f_n, f_* = \underline{\lim} f_n \leqslant g$
and $f_n^+, f_n^-$ are integrable.
Cheap trick

$$\begin{aligned} g - f_n, g + f_n &\geqslant 0 \\ \int (g + f_*) &= \int \underline{\lim}(g + f_n) \\ &\leqslant \underline{\lim} \int (g + f_n) \\ &= \int g + \underline{\lim} \int f_n \\ \int (g - f_*) &= \int \underline{\lim}(g - f_n) \\ &\leqslant \underline{\lim} \int (g - f_n) \\ &= \int g - \overline{\lim} \int f_n \end{aligned}$$

So

$$\int f_* \leqslant \underline{\lim} \int f_n \leqslant \overline{\lim} \int f_n \leqslant \int f^*$$

But $f_* = f^* - f$, done.

**Change of measure** and 1, 2, 3 argument.

$$(\Omega, \mathcal{F}, \mu), (\Omega', \mathcal{F}')$$

$T : \Omega \to \Omega'$ is measurable.
Define $\mu^{T^{-1}}$ on $\mathcal{F}'$

$$\mu^{T^{-1}}(B) = \mu(T^{-1}(B))$$

**Theorem** (change of measure)

$$f : \Omega' \to \mathbb{R} \text{ measurable, } f \geqslant 0$$

Then

$$\int f(\omega')\mu^{T^{-1}}(\mathrm{d}\omega') = \int f(T(\omega))\mu(\mathrm{d}\omega)$$

*Proof* 1, 2, 3 argument:
1) Take

$$f(\omega') = \delta_B(\omega') = \begin{cases} 1 & \omega' \in B \\ 0 & \omega' \notin B \end{cases}$$

$$\begin{aligned}\int \delta_B(\omega')\mu^{T^{-1}}(\mathrm{d}\omega') &= \mu(T^{-1}(B)) \\ &= \int \delta_B(T(\omega))\mu(\mathrm{d}\omega)\end{aligned}$$

$$\big(\delta_B(T(\omega)) = \delta_{T^{-1}(B)}(\omega)\big)$$

2) If Ok for $f_1, f_2$, then OK for $f_1 + f_2$.
3) $\forall f \geqslant 0, \exists f_n \uparrow f, f_n$ step functions.

$$\begin{aligned}\int f\mu^{T^{-1}} &= \int \lim f_n \mu^{T^{-1}} \\ &= \lim \int f_n \mu^{T^{-1}} \\ &= \lim \int f_n(T(\omega))\mu(\mathrm{d}\omega) \\ &= \int f(T(\omega))\mu(\mathrm{d}\omega)\end{aligned}$$

**Example**. $\Omega = \{\{0,1\}^n\}, \mu(\omega) = \frac{1}{2^n}$

$$\Omega' = \{0,1,2,\cdots,n\}$$

$$T(\omega) = \sum_1^n \omega_i$$

$$\mu^{T^{-1}}(j) = \frac{\begin{pmatrix} n \\ j \end{pmatrix}}{2^n}$$

Take $f(j) = j^2$.
Theorem says

$$\sum_{j=0}^{n} \frac{j^2 \begin{pmatrix} n \\ j \end{pmatrix}}{2^n} = \frac{1}{2^n} \sum_{\omega_1,\cdots,\omega_n} (\omega_1 + \cdots + \omega_n)^2$$

**Product spaces and Fubini**,
Let $\Omega_1, \Omega_2$ be sets.

$$(\Omega_1 \times \Omega_2) = \{(\omega_1, \omega_2) : \omega_i \in \Omega_i\}$$

If $A \subseteq \Omega_1 \times \Omega_2$
Section $A_{\omega_1} = \{\omega_2 : (\omega_1, \omega_2) \in A\}$
$f(\omega_1, \omega_2)$ any function,

$$f_{\omega_1}(\omega_2) = f(\omega_1, \omega_2)$$

:) sections are great

- $$(\cup_{i \in I} A_i)_{\omega_1} = \cup_i A_{i\omega_1}$$
- $$(\cap A_i)_{\omega_1} = \cap A_{i\omega_1}$$
- $$(A_{\omega_1})^c = (A^c)_{\omega_1}$$

**Definition**. Say $(\Omega_1, \mathcal{F}_1), (\Omega_2, \mathcal{F}_2)$
A rectangle $A = A_1 \times A_2 = \{(\omega_1, \omega_2), \omega_i \in A_i\}$
measurable rectangle if $A_i \in \mathcal{F}_i$
**Definition**. $\mathcal{F}_1 \times \mathcal{F}_2 = \sigma-$algebra containing measurable rectangles.

**Sectioning Lemma**

- $$A \in \mathcal{F}_1 \times \mathcal{F}_2, \text{ then } A_{\omega_1} \text{ and } A_{\omega_2} \text{ are } \mathcal{F}_2, \mathcal{F}_1 \text{ measurable } \forall \omega_1, \omega_2$$

- $$\text{If } f : \Omega_1 \times \Omega_2 \to \mathbb{R} \text{ is measurable , then } f_{\omega_1}, f_{\omega_2} \text{ are } \mathcal{F}_2, \mathcal{F}_1 \text{ measurable.}$$

*Proof*
1) True for $A = A_1 \times A_2$

$$A_{\omega_1} = \begin{cases} A_2 & \text{if } \omega_1 \in A_1 \\ \varnothing & \text{else} \end{cases}$$

Let $\mathcal{S} = \{A \text{ so true}\}$, $\mathcal{S}$ contains rectangles.
$\mathcal{S}$ closed under unions and complements and so $\mathcal{S} \supseteq \sigma\{\text{rectangle}\}$
2) Look at $\{f : \Omega_1 \times \Omega_2 \to \mathbb{R}, \text{ so that 2) holds }\}$,
1. This contains $\delta_{A_1 \times A_2}$
2. $f_1, f_2$ OK, $f_1 + f_2$ OK.
3. Monotone limits.

**Warning** converse fails:
If $A_{\omega_1}, A_{\omega_2}$ measurable, doesn't $\Rightarrow$ A is measurable.
**Example**. $\Omega_1 = \Omega_2 = (0,1], \mathcal{F}_i = \{A : A \text{ is countable or cocountable }\}$ in $(0,1]^2$, w.r.t. $\mathcal{F}_1 \times \mathcal{F}_2$.
Diagonal: $A_{\omega_1}, A_{\omega_2}$ a point sets.
$A \notin \mathcal{F}_1 \times \mathcal{F}_2$

$(\Omega_1, \mathcal{F}_1), (\Omega_2, \mathcal{F}_2)$
**Definition** A kernel $K(\omega_1, \mathrm{d}\omega_2)$ is a function $K : \Omega_1 \times \mathcal{F}_2 \to [0,1]$, such that

- $$\forall A \in \mathcal{F}_2, \omega \to K(\omega, A) \text{ is } \mathcal{F}_1 \text{ measurable}$$

- $$\forall \omega_1, K(\omega_1, \cdot) \text{ is probability on } \mathcal{F}_2$$

**Example**

- $\Omega_1$ is a "parameter space"
  $\{P_\theta(\mathrm{dx})\}_{\theta\in\Theta}$ is a kernel.
  Let $X \sim N(\mu, \sigma^2)$
  $$\Omega_1 = \mathbb{R} \times (0, \infty]$$
  $$K(\nu, \sigma, A) = \int_A \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} \mathrm{dx}$$
- $$K(\omega_1, A) = \mu(A) \text{ some fixed } \mu$$
- $$\Omega_1 = \Omega_2, K(\omega_1, \mathrm{d}\omega_2) \text{ is a Markov chain}$$

**Aim** For $\pi$ probability on $(\Omega, \mathcal{F})$, want to define $\pi K$ on $\mathcal{F}_1 \times \mathcal{F}_2$. Pick $\omega_1$ from $\pi$ and then $\omega_2$ from $K(\omega_1, \cdot)$

**Lemma** Given $K$, and $A \in \mathcal{F}_1 \times \mathcal{F}_2$ measurable, the map $\omega_1 \mapsto K(\omega_1, A_{\omega_1})$ is $\mathcal{F}_1$ measurable.
*Proof* Let $\mathcal{S}$ be $\{A \subseteq \Omega_1 \times \Omega_2 \text{ such that true}\}$

- $$\mathcal{S} \text{ contains } A_1 \times A_2$$
  $$\text{Because } K(\omega_1, (A_1 \times A_2)_{\omega_1}) = \delta_{A_1^c}(\omega_1) K(\omega_1, A_2)$$
  (Recall rectangles are $\pi$−system)

- $\mathcal{S}$ is a $\lambda$−system
  $$K(\omega_1, (\Omega_1 \times \Omega_2)_{\omega_1}) = K(\omega_1, \Omega_2) \text{ is measurable}$$
  If $A \in \mathcal{S}$
  $$K(\omega_1, (A_{\omega_1})^c) = K(\omega_1, (A^c)_{\omega_1}) = 1 - K(\omega_1, A_{\omega_1}) \text{ is measurable}$$
  If $\{A_i\}$ in $\mathcal{S}$ disjoint
  $$K(\omega_1, (\cup_1^\infty A_i)_{\omega_1}) = K(\omega_1, \cup A_{i\omega_1}) = \sum K(\omega_1, (A_i|_{\omega_1}))$$

$\mathcal{S}$ is a $\lambda$–system containing $\pi$–system of rectangles, $\mathcal{S} \supseteq \mathcal{F}_1 \times \mathcal{F}_2$

**Definition** Given $\pi$ on $(\Omega_1, \mathcal{F}_1), K : \Omega_1 \times \mathcal{F}_2 \to [0,1]$,

$$\pi K(A) = \int_{\Omega_1} K(\omega_1, A_{\omega_1})\pi(\mathrm{d}\omega_1)$$

**Theorem** This is a probability measure on $\mathcal{F}_1 \times \mathcal{F}_2$

- $$A = \Omega_1 \times \Omega_2, \int K(\omega_1, \Omega_2)\pi(\omega_1) = 1$$
- $$\pi K(A^c) = 1 - \pi K(A)$$
- $$\begin{aligned} &\{A_i\}_{i=1}^{\infty} \text{ disjoint in } \mathcal{F}_1 \times \mathcal{F}_2 \\ \pi K(\cup A_i) &= \int K(\omega_1, (\cup A_i)_{\omega_1})\pi(\mathrm{d}\omega_1) \\ &= \int \sum_1^{\infty} K(\omega_1, A_{i\omega_1})\pi(\mathrm{d}\omega_1) \\ &= \sum \pi K(A_i) \end{aligned}$$

**Theorem** (Fubini for kernels) Given $\pi K$ as above:

- $\pi K$ is the unique probability measure on $\mathcal{F}_1 \times \mathcal{F}_2$
  $$\text{such that } \pi K(A_1 \times A_2) = \int_{A_1} K(\omega_1, A_2)\pi(\mathrm{d}\omega_1)$$
- $$\begin{gathered} \forall f : \Omega_1 \times \Omega_2 \to [0, \infty] \text{ measurable} \\ \omega_1 \mapsto \int f_{\omega_1}(\omega_2)K(\omega_1, \omega_2) \text{ is } \mathcal{F}_1 \text{ measurable} \end{gathered}$$
- $$\int f(\omega_1, \omega_2)\pi K(\mathrm{d}\omega_1, \mathrm{d}\omega_2) = \int_{\Omega_1} \left( \int_{\Omega_2} f_{\omega_1}(\omega_2)K(\omega_1, \mathrm{d}\omega_2) \right) \pi(\mathrm{d}\omega_1)$$

*Proof*
1) From $\pi - \lambda$ (measurable rectangles: $\pi$−system generating $\mathcal{F}_1 \times \mathcal{F}_2$)
2, 3) Let $\mathcal{S} = \{f \text{ so true}\}$.

$$\begin{aligned} f(\omega_1, \omega_2) &= \delta_{A_1 \times A_2}(\omega_1, \omega_2) \\ &= \delta_{A_1}(\omega_1) K(\omega_1, A_2) \end{aligned}$$

OK for linear combinations, monotone limits.

Remarks

- If $K(\omega_1, A) = \mu(A)$
  This says
  $$\int \int f(\omega_1, \omega_2) \pi(\mathrm{d}\omega_1) \pi(\mathrm{d}\omega_2) = \int \left[ \int f(\omega_1, \omega_2) \mu(\mathrm{d}\omega_2) \right] \pi(\mathrm{d}\omega_1)$$
- $\pi(A) = \pi K(A \times \Omega_2)$ First margin.
- If $\exists Q(\omega_2, \mathrm{d}\omega_1) : \Omega_2 \times \mathcal{F}_1$ kernel, such that
  $$\pi K(A) = \mu Q(A)$$
  $$\mu(B) = \pi K(\Omega_1 \times B)(?)$$
  $$\int K(\omega_1, A_{\omega_1}) \pi(\mathrm{d}\omega_1) = \int Q(\omega_2, A_{\omega_2}) \mu(\mathrm{d}\omega_2)^*$$
  then Q is called a regular conditional distribution for $\pi K$ given $\omega_2$.

$*$ is Bayes theorem in measure theory language: $K$ was $P_\theta(\mathrm{dx})$.

Various extensions.
**Theorem** Let $\pi, K$ be as above:

$$f : \Omega_1 \times \Omega_2 \to [-\infty, \infty]$$

measurable and $\pi K$ integrable.
Set $\mathcal{G} = \{\omega_1 : f_{\omega_1} \text{ is } K(\omega_1, \cdot) \text{ integrable}\}$.
Define $Kf : \Omega_1 \to [-\infty, \infty], Kf(\omega_1) = \begin{cases} \int f_{\omega_1} K(\omega_1, \mathrm{d}\omega_1) & \text{if } \omega_1 \in ? \\ 0 & \text{else} \end{cases}$
Then $\mathcal{G} \in \mathcal{F}, \pi(\mathcal{G}) = 1$

$$\int f \mathrm{d}\omega \mathrm{K} = \int (Kf) \mathrm{d}\pi$$

**Example** Take $\Omega_1 = (0,1], \pi = \lambda$ Lebesgue,

$$\Omega_2 = (0,1], K_{\omega_1}(0) = K_{\omega_1}(1) = \frac{1}{2}$$

$$f(\omega_1, \omega_2) = \frac{(-1)^{\omega_2}}{\omega_1}$$

$$f_{\omega_1}(\omega_2) = \begin{cases} \frac{1}{\omega_1} & \omega_2 = 0 \\ -\frac{1}{\omega_1} & \omega_2 = 1 \end{cases}$$

$$Kf(\omega_1) = \frac{1}{2}\frac{1}{\omega_1} - \frac{1}{2}\frac{1}{\omega_1} = 0$$

But $K(f^+)(\omega_1) = K(f^-)(\omega_1) = \frac{1}{\omega_1}$ and

$$\int f^+ \pi K = \int_0^1 \frac{1}{\omega_1} = \infty$$

IOUs

- **Question** We proved $f_n \geqslant 0, \int \lim f_n = \lim \int f_n$, what about $f_n \downarrow$
  **Example** $f_n$ on $[0,\infty) = \frac{I(x>n)}{n}$ $f_n \downarrow$ i.o.$0 = \lim f_n$ but

  $$0 = \int_0^\infty 0 \mathrm{d}\mu \neq \lim \int f_n = \infty$$

  But if

  $$f_n(\omega) \geqslant 0, f_n(\omega) \downarrow$$

  and

  $$\int f_n \mathrm{d}\mu < \infty, \text{ some} n$$

  Then

  $$\int \lim f_n \mathrm{d}\mu = \lim \int f_n \mathrm{d}\mu$$

  (By dominated convergence)

- Independence
  **Lemma** If

  $$\begin{matrix} A_{11} & A_{12} & A_{13} & \cdots \\ A_{21} & A_{22} & A_{23} & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{matrix}$$

rows finite on $\infty$, columns finite on $\infty$ and all $A_{i,j}$ are independent.
Then
If $\mathcal{F}_n = \sigma(A_{n,j}, 1 \leqslant j < \infty)$ have

$$\mathcal{F}_1, \mathcal{F}_2, \cdots \text{ are independent}$$

*Proof* Let $\mathcal{C}_i = \{\text{All finite intersections of sets } A_{i,j}\}$ These are $\pi-$systems, and

$$C_i = \cap_{j \in J_i} A_{i,j}$$

$$\begin{aligned} P(\cap_{i=1}^n C_i) &= P(\cap_i \cap j A_{i,j}) \\ &= \prod_{i,j} P(A_{i,j}) \\ &= \prod_i P(C_i) \end{aligned}$$

So $\sigma(A_{i,j} 1 \leqslant j < \infty) = \mathcal{F}_i$ are independent.

**Strong law**

**Theorem** Suppose $X_1, X_2, \cdots$ i.i.d. with mean $\mu = \mathbb{E}(X_1) < \infty$, then

$$\frac{S_n}{n} \to \mu, \text{ a.s.}$$

*Proof* w.l.o.g. $X_i \geqslant 0$

$$X = X^+ - X^-, \sum X_i = \sum X_i^+ - \sum X_i^-$$

Truncation

$$Y_i = X_i \delta_{\{X_i \leqslant i\}}$$

$$T_n = \sum_{i=1}^n Y_i$$

Subsequences

$$\alpha > 1, u_n = [\alpha^n]$$

We prove, $\forall \varepsilon > 0$

$$* \sum_{i=1}^\infty P\left\{\left|\frac{T_{u_i} - \mathbb{E}(T_{u_i})}{u_i}\right| > \varepsilon\right\} < \infty$$

$$\begin{aligned} \operatorname{var}(T_n) &\leqslant \sum_{i=1}^n \mathbb{E}(Y_i^2) \\ &= \sum_{i=1}^n \mathbb{E}(x_i^2 \delta_{\{x_i \leqslant i\}}) \\ &= \mathbb{E}\{X_1^2 \sum \delta_{\{X_i \leqslant i\}}\} \\ &\leqslant n\mathbb{E}\{X_1^2 \delta_{\{X_i \leqslant n\}}\} \end{aligned}$$

Chebyshev applied to $*$

$$\begin{aligned} * &\leqslant \sum_{i=1}^\infty \frac{1}{\varepsilon^2 u_i^2} u) i \mathbb{E}(X_1^2 \delta_{\{X_1 \leqslant a_i\}}) \\ &= \frac{1}{\varepsilon^2} \mathbb{E}(X_1^2 \sum_{i=1}^\infty \frac{1}{u_n} \delta_{\{X_i \leqslant u_i\}}) \end{aligned}$$

Study

$$** \sum_{i=1}^\infty \frac{1}{u_i} \delta_{\{x \leqslant u_i\}}$$

Smallest $N_1$ call it $N_x$
$u_{N_x} > x_1$ so $\alpha^{N_x} > x$

$$\alpha^n \leqslant 2u_n (?)$$

$$** \leqslant \sum_{n \geqslant N_x} \frac{2}{\alpha^n} = K\alpha^{-N_x}$$

$K = \frac{2\alpha}{\alpha - 1}$ and since $\alpha^{N_x} > x$

$$** \leqslant \frac{K}{x}$$

$$* \leqslant \frac{k}{\varepsilon^2}\mathbb{E}(X_1) < \infty$$

So Borel-cantelli

$$\frac{T_{u_n} - \mathbb{E}(T_{u_n})}{u_n} \xrightarrow{a.s.} 0$$

We have it for truncated variances on a subsequence
Fight on a way back
Elementary fact
$x_i$ real numbers, $x_i \to x$

$$\frac{1}{n}\sum_{i=1}^{n} x_i \to x$$

Well, $\mathbb{E}(Y_i) = \mathbb{E}(X_i \delta_{\{x_i \leqslant i\}})$

$$\uparrow \mu = \mathbb{E}(X_i)$$

$$\frac{1}{n}\mathbb{E}(T_n) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(Y_i) \to \mu$$

Have

$$\frac{T_{u_n}}{u_n} \to \mu, \text{ a.s.}$$

Next

$$\begin{aligned}
\sum P(X_i \neq Y_i) &= \sum_{i=1}^{\infty} P(X_i > i) \\
&\leqslant \int_0^{\infty} P(X_1 > t)\mathrm{dt} \\
&= \mu \\
&< \infty
\end{aligned}$$

So Borel Cantelli, $X_n \neq Y_n$ finitely often

$$\frac{S_n - T_n}{n} \to 0$$

Have

$$\frac{S_{u_n}}{u_n} \to \mu, \text{ a.s.}$$

Interpolation
Fix $h, u_n \leqslant h \leqslant u_{n+1}$

$$\frac{S_{u_n}}{u_{n+1}} \leqslant \frac{S_h}{h} \leqslant \frac{S_{u_{n+1}}}{u_n}$$

So

$$\frac{u_n}{u_{n+1}} \frac{S_{u_n}}{u_n} \leqslant \frac{S_h}{h} \leqslant \frac{u_{n+1}}{u_n} \frac{S_{u_{n+1}}}{u_{n+1}}$$

$$n \to \infty, \frac{u_{n+1}}{u_n} = \frac{[\alpha^{n+1}]}{[\alpha^n]} \to \alpha$$

$$\frac{1}{\alpha}\mu \leqslant \underline{\lim} \frac{S_h}{h} \leqslant \overline{\lim} \frac{S_h}{h} \leqslant \alpha\mu \text{ a.s.}$$

Let $\alpha = 1 + \frac{1}{n}, P(\cap setsof?) = 1$

$$\frac{S_h}{h} \to \mu \text{ a.s.}$$

Etamadi's ?

- Tchebyshev
- Truncation
- Terpolation
- Tsubsequences

**Remarks**

- This is Nice and strong:

$$\text{If } \frac{S_n}{n} \xrightarrow{?} \mu < \infty, \text{ then } \mathbb{E}(X_1) = \mu$$

- :( No finite content
(S.L.L.N. says $|\frac{S_n}{n} - \mu| < \varepsilon$ and stays there all $n > N$, ? can't be more quantitative without mean.

- Say $X_n$ have $\mathbb{E}(X_n) = 0, \text{var}(1)$.
  Let $m = m(\varepsilon) = \sup_n\{\frac{S_n}{n} \geqslant c\}$

  $$P(m < \infty) = 1$$

  **Theorem** (?)

  $$P\{\varepsilon^2 m(\varepsilon) \leqslant x\} \overset{?}{\to} (2\Phi(\sqrt{x}) - 1)(\chi_1^2 = Z^2)$$

  $$\text{``}m(\varepsilon) = \frac{1}{\varepsilon^2}\text{''}$$

- $X_i$ i.i.d. $F(x), \int_{-\infty}^{\infty} xF(\mathrm{dx}) = 0$

  $$X_n^{(m)} = X_n \delta_{\{|X_n| \leqslant n \vee m\}}, \mu_n^m = \mathbb{E}(X_n^{(m)})$$

  $$S_n^{(m)} = \sum_{i=1}^{n} X_i^{(m)} = \mathbb{E}(S_n^{(m)}), \alpha_m = \sup_i(\mu_i^{(m)})$$

  **Theorem** Let $\varepsilon > 0$, m large enough
  $\alpha_m < \varepsilon$, then

  $$P\{\cup_{n=m}^{\infty} |\frac{S_n}{n}| > \varepsilon\} \leqslant \int_{|X|>m} |X| F(\mathrm{dx}) + ?$$

  "Proof chow-robbins-siegmund great expectations"

- some kind of converse
  Say $X$ has $\mathbb{E}(X^-) < \infty, \mathbb{E}(X^+) = \infty$,
  then $\frac{S_n}{n} \xrightarrow{n\to\infty} \infty$, a.s.
  *Proof* w.l.o.g. $X \geqslant 0$
  Let $X^i = X\delta_{\{X \leqslant i\}} i > 0$
  S.L.L.N.

  $$\frac{S_n}{n} \xrightarrow{n\to\infty} \mathbb{E}(X^i) a.s.$$

  $$\mathbb{E}(X^i) \uparrow \mathbb{E}(X) = \infty$$

- S.L.L.N. is special case of
  a) Ergodic theorem
  b) Martingale convergence

- **Example** $X$ takes values $\{\pm 1, \pm 2, \pm 3, \cdots\}$

$$P\{X = \pm j\} = \frac{c}{(|j|+1)^2 \log(|j|+1)}$$

$$P\{|\frac{S_n}{n}| > \varepsilon\} \xrightarrow{n\to\infty} 0$$

But $|\frac{S_n}{n}| > \varepsilon,$ i.o. all $\varepsilon > 0$

- **Example** Cauchy, $X$ has density $\frac{1}{\pi(1+x^2)} - \infty < x < \infty$

$$P(\frac{S_n}{n} \leqslant x) = \int_{-\infty}^{x} \frac{1}{\pi(1+x^2)} \mathrm{dx}$$

**Definition** on $\mathbb{R}$, let $F_n, F$ be distribution functions, $F_n \Rightarrow F$ "$F_n$ converges to F weakly (or in distribution)".
If $\forall$ continuity x of $F$, $F_n(x) \xrightarrow{n\to\infty} F(x)$.

**Example** $X_n = 1 \pm \frac{1}{n}$ probability $\frac{1}{2}$, $X \equiv 1$.
$F_n(1) \equiv \frac{1}{2}, F(1) = 1$ $F_n \Rightarrow F$ OK.

**Definition** If $X_n, X$ defined on $(\Omega, \mathcal{F}, P)$ say $X_n \xrightarrow{p} X$ "$X_n$ converges to $X$ in probability".
If $\forall \varepsilon > 0, P(|X_n - X| > \varepsilon) \to 0$

$$X_n \xrightarrow{a.s.} X, P(\{\omega : X_n(\omega) \to X(\omega)\}) = 1.$$

**Property** $X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{p} X \Rightarrow F_n \Rightarrow F$, and all implications are strict.
*Proof* $X_n \xrightarrow{a.s.} X \Leftrightarrow \forall \varepsilon > 0, P(|X_n - X| > \varepsilon, \text{ i.o.}) = 0$.
Fix $\varepsilon$, $\{|X_n - X| > \varepsilon \text{ i.o.}\} = \cap_{m=1}^{\infty} \cup_{n=m}^{\infty} \{|X_n - X| > \varepsilon\} \triangleq C_m$.
$C_m \downarrow \varnothing$ with probability 1, so $P(C_m) \downarrow 0$, so $P(C_m) \geqslant P(|X_m - X| > \varepsilon) \to 0$.
Say, $X_n \xrightarrow{p} X$,
then

$$\begin{aligned} P(X \leqslant x - \varepsilon) - P(|X_n - X| > \varepsilon) &\leqslant P(X_n \leqslant x) \\ &\leqslant P(X \leqslant x + \varepsilon) + P(|X_n - X| > \varepsilon) \end{aligned}$$

$$\begin{aligned} \Rightarrow P(X \leqslant x - \varepsilon) &\leqslant \underline{\lim} P(X_n \leqslant x) \\ &\leqslant \overline{\lim} P(X_n \leqslant x) \\ &\leqslant P(X \leqslant x + \varepsilon) \end{aligned}$$

For x continuity point of $F(x)$, let $\varepsilon \to 0$.

Counterexamples
$F_n \Rightarrow F \not\Rightarrow$ convergence in probability.
**Example** Let X and Y be independent standard normals.
Let $X_n = Y, F_n(x) = \Phi(x) = F(x)$

$$P(|X_n - X| > \varepsilon) \not\to 0$$

**Example** $X_n \xrightarrow{p} X \not\Rightarrow X_n \xrightarrow{a.s.} X$

$$(\Omega, \mathcal{F}, P) = (0, 1], \lambda$$

$$\begin{array}{cccccccc} (0, \frac{1}{2}] & (\frac{1}{2}, 1] & (0, \frac{1}{4}] & (\frac{1}{4}, \frac{1}{2}] & (\frac{1}{2}, \frac{3}{4}] & (\frac{3}{4}, 1] & (0, \frac{1}{16}] & \cdots \\ A_1 & A_2 & A_3 & A_4 & A_5 & A_6 & A_7 & \cdots \end{array}$$

$$X_n(\omega) = \delta_{A_i}(\omega), X(\omega) = 0$$
$$P(|X_n - X| > \varepsilon) = P(X_n = 1) \to 0$$
But for $\omega : X_n(\omega) = 1,$ i.o. $\forall\omega$.

**Example** $P(j) = \frac{c}{(|j|+1)^2 \log(5+|j|)}$

$$j = 0, \pm 1, \pm 2, \cdots$$
$$X_i \text{i.i.d. from } P(j)$$
$$S_n = X_1 + \cdots + X_n$$
$$\frac{S_n}{n} \xrightarrow{p} 0 \text{ but } \frac{S_n}{n} \nrightarrow 0, \text{ a.s.}$$

Other notions of convergence

- Setwises $\forall A \in \mathcal{F}$
$$\mu_n(A) \to \mu(A)$$
- Total variation. $\sup_{A \in \mathcal{F}} |\mu_n(A) - \mu(A)| \to 0$

Skorohod's theorem (on $\mathbb{R}$)
**Theorem** Say $F_n \Rightarrow F$, then on $(\Omega, \mathcal{F}, P) = ((0,1], \text{Borel}, \lambda), \exists Y, Y_n : \Omega \to \mathbb{R}$

$$P(Y_n \leqslant x) = F_n(x), P(Y \leqslant x) = F(x)$$

and
$$Y_n(\omega) \xrightarrow{n \to \infty} Y(\omega) \text{ all } \omega$$

*Proof* Say first, $F_n, F$ are continuous for all x (strictly increasing).
Recall, $Y(\omega) = F^{-1}(\omega), Y_n(\omega) = F_n^{-1}(\omega)$ have distribution functions $F_n, F$.

$$\begin{aligned} P(Y(\omega) \leqslant x) &= \lambda(\omega : F^{-1}(\omega) \leqslant x) \\ &= \lambda(\omega : \omega \leqslant F(x)) \\ &= F(x) \end{aligned}$$

And clearly, $Y_n(\omega) \to Y(\omega)$ all $\omega$.
If pick u $(0,1]$, $F^{-1}(u)$ has distribution function F.
Set $Y_n(\omega) = F_n^{-1}(\omega)$ and $Y(\omega) = F^{-1}(\omega)$ for $\omega$ a continuity point of $F$ And

$Y_n(\omega) = Y(\omega) = 0$, rest.

$$P(Y_n \leqslant x) = F(x)$$

Have to argue

$$Y_n(\omega) \to Y(\omega) \text{ all } \omega$$

**Corollary** $F_n \Rightarrow F \Leftrightarrow \forall$ bounded continuous $f : \mathbb{R} \to \mathbb{R}$

$$\int_{-\infty}^{\infty} f(x) F_n(\mathrm{dx}) \to \int f(x) F(\mathrm{dx})$$

*Proof* Say $F_n \Rightarrow F$, choose $Y_n, Y$. Skorohod.

$$Y_n(\omega) \to Y(\omega) \text{ all } \omega$$

$$\int f \mathrm{d}_{\mathrm{F_n}} = \mathbb{E}(f(Y_n)) \to \mathbb{E}(f(Y)) = \int f \mathrm{d}_{\mathrm{F}} (\text{Dominated Convergence Theorem})$$

On the other hand, if $\int f \mathrm{d}_{\mathrm{F_n}} \to \int f \mathrm{d}_{\mathrm{F}}$ all bounded continuous functions. Approximate $f$ as $f_\varepsilon$

$$\begin{aligned} \int f_\varepsilon \mathrm{d}_{\mathrm{F}} &\leqslant \int f_\varepsilon \mathrm{d}_{\mathrm{F_n}} \\ &\leqslant F_n(x) \\ &\leqslant \int f_\varepsilon^+ \mathrm{d}_{\mathrm{F_n}} \\ &\to \int f_\varepsilon^+ \mathrm{d}_{\mathrm{F}} \end{aligned}$$

Let $\varepsilon \to 0$, when x is continuous point, OK.

**Theorem** (Slutsky) Suppose $X_n, Y_n, X$ on $(\Omega, \mathcal{F}, P)$,

$$X_n \Rightarrow X \text{ and } Y_n - X_n \xrightarrow{p} 0$$

Then

$$Y_n \Rightarrow X$$

*Proof* Given any x, choose $y' < x < y''$ and $\varepsilon > 0$ s.t.

$$y' < x - \varepsilon < x < x + \varepsilon < y''$$

$$\mu(y') = \mu(y'') = 0, \mu \Leftrightarrow F$$

$$\begin{aligned} P(X_n \leqslant y'\} - P(|X_n - Y_n| \geqslant \varepsilon) &\leqslant P(Y_n \leqslant x) \\ &\leqslant P(X_n \leqslant y'') + P(|X_n - Y_n| \geqslant \varepsilon) \end{aligned}$$

$$\begin{aligned} F(y') &\leqslant \underline{\lim} P(Y_n \leqslant x) \\ &\leqslant \overline{\lim} P(Y_n \leqslant x) \\ &\leqslant F(y'') \end{aligned}$$

If $F$ is continuous at x, let $y', y'' \to x$.

**Example** Card guessing. $Z_n = \begin{cases} 1 & \text{probability } \frac{1}{n} \\ 0 & \text{probability } 1 - \frac{1}{n} \end{cases}$

$$S_N = \sum_{n=1}^{N} Z_n = \#\text{ correct guesses}$$

$$\mathbb{E}(S_N) = 1 + \frac{1}{2} + \cdots + \frac{1}{N} = \log N - \gamma + o(\frac{1}{N}) \sim \log N$$

$$\mathrm{Var}(S_N) = \sum_{n=1}^{N} \frac{1}{n}(1 - \frac{1}{n}) = \log N - \frac{\pi^2}{6} - \gamma + o(\frac{1}{N}) \sim \log N$$

can show, $P(\frac{S_N - \mu_N}{\sigma_N} \leqslant x) \to \Phi(x) \Leftrightarrow P(\frac{S_N - \log N}{\sqrt{\log N}} \leqslant x\} \to \Phi(x)$

$$X_N = \frac{S_N - \mu_N}{\sigma_N}, Y_n = \frac{S_N - \log N}{\sqrt{\log N}}$$

Ok provided $P(|X_N - Y_N| > \varepsilon) \to 0$
Look at $(\frac{S_N}{\sigma_N} - \frac{S_n}{\sqrt{\log N}}) - (\frac{\mu_N}{\sigma_N} - \frac{\log N}{\sqrt{\log N}})$

**Theorem** (Helly selection)
On $\mathbb{R}$, have any family $F_n(x)$ (Distribution function), then $\exists n_k, 1 \leqslant k < \infty$ and a monotone right continuous F, $0 \leqslant F \leqslant 1$, so that $F_{n_k} \Rightarrow F$.
Note $F$ need not be distribution function.

**Example**

$$F_n \leftrightarrow \delta_n$$

$$F_n \Rightarrow F \equiv 0$$

$$F_n \leftrightarrow \delta_{-n}$$

$$F_n \Rightarrow F \equiv 1$$

*Proof*
Use Cantor

$$\begin{array}{cccc} x_{11} & x_{12} & x_{13} & \cdots \\ x_{22} & x_{22} & x_{23} & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{array}$$

Each row bounded, then $\exists l_r$ and $n_k$ so that

$$x_{r,n_k} \xrightarrow{k\to\infty} l_r$$

All r.
*Proof of Cantor*
By compactness, $\exists n_1$ and $l_1, x_{1,n_1} \to l_1$.
Look at $x_{2,n_1}, \exists n_{1,2}$ subsequence of $n_1$, and $l_2$ so $x_{2,n_{1,2}} \to l_2$.
Keep going.
Now set $n_k = n_{k,k}$.
This is a subsequence of each $n_{i,j}$ and so $x_{i,n_k} \to l_i$ all i.
Back to $F_n$ look at rationals r:

$$F_1(r_i), F_2(r_i), \cdots$$

$\forall n_k$ (by Cantor) for all rationals $r_i$ and $G(r_i), F_{n_k}(r) \xrightarrow{k} G(r)$.
Let $F(x) = \inf_{r>x} G(r)$, this is monotone in x.
**Claim** $F(x)$ is right continuous.
Given $x, \varepsilon > 0, \exists r > x, G(r) < F(x) + \varepsilon, \forall y, x < y < r$ have

$$F(x) \leqslant F(y) \leqslant G(r) \leqslant F(x) + \varepsilon$$

**Claim** If x is continuity point of $F$, then $F_{n_k}(x) \to F(x)$.
Given $\varepsilon$, choose $y < x$ so $F(x) - \varepsilon < F(y)$.
Choose $r, R, y < r < x < R$ and $G(R) < F(x) + \varepsilon$
then

$$F(x) - \varepsilon < F(y) \leqslant G(r) \leqslant F(x) \leqslant G(R) < F(x) + \varepsilon$$

Also, $\forall n, F_n(y) \leqslant F_n(r) \leqslant F_n(x) \leqslant F_n(R)$.
Now take $\underline{\lim}, \overline{\lim} F_{n_k}(x)$ and $\underline{\lim} F_n(x)$ within $\varepsilon$ of $F(x)$ and so $F_{n_k}(x) \to F(x)$.
**Definition** Let $\mu_n$ be probabilities on $\mathbb{R}, \{\mu_n\}$ is tight if $\forall \varepsilon > 0, \exists (a,b]$ so $\mu_n(a,b] > (1-\varepsilon), \forall n$.
$\mu_n$ are "almost compactly supported" mass doesn't drift to $\infty$.
**Theorem** Given $\mu_n$, tightness is a necessary and sufficient condition for: $\forall$ subsequence $n_k, \exists n_{k_i}$ and probability $\mu$ so that $\mu_{n_{k_i}} \Rightarrow \mu$.
*Proof* "$\Rightarrow$" say $\mu_n$ tight, given $n_k, \mu_{n_k}$, by Helly has subsequence $n_{k_i}$ and a $\mu$ (not necessarily probability)

$$\mu_{n_{k_i}} \Rightarrow \mu$$

Choose $(a,b]$ so that $F_n$ is continuous at $a, b$, and $\mu_n(a,b] > (1-\varepsilon)\forall n$.

$$F_{n_k}(b) - F_{n_k}(a) > (1-\varepsilon)$$
$$F_\mu(b) - F_\mu(a) > (1-\varepsilon)$$
$$\Rightarrow \lim_{n\to\infty} F_n(x) = 1$$

Remark:
"Everything" goes over to probability on a complete separable metric space. (See Dudley book)
Let $(\mathcal{X}, d)$ be polish space Borel sets, $\mu_n, \mu$ are probabilities on $(\mathcal{X}, \text{Borel})$.
Define $\mu_n \Rightarrow \mu$, weak*.
If $\forall$ bounded continuous $f : \mathcal{X} \to \mathbb{R}$

- $\int f \mathrm{d}\mu_\mathrm{n} \to \int f \mathrm{d}\mu$

$$\Leftrightarrow \forall \text{ Borel } B, \mu(\partial B) = 0, \mu_n(B) \to \mu(B)$$

- If $\mu_n \Rightarrow \mu, \exists Y_n(\omega), Y(\omega)$ on $(0,1], \lambda$

$$P(Y_n \in B) = \mu_n(B)$$
$$P(Y \in B) = \mu(B)$$

$$Y_n(\omega) \xrightarrow{n\to\infty} Y(\omega) \text{ all } \omega$$

- $\mu_n$ is tight be definition, if $\forall \varepsilon > 0, \exists$ compact $K$, so $\mu_n(K) > (1-\varepsilon)$ all $n$.

- $\{\mu_n\}$ tight then $\exists n_k$ and $\mu$ probability so $\mu_{n_k} \Rightarrow \mu$.

Reference

- Billingsley, convergence of probability measures.

- "Just the facts", O. Kallenberg, Foundations of Modern Probability second edition.

**Characteristic functions**.
Let $\mu$ be a probability on $\mathbb{R}$, (Associate $\mathcal{F}$, X) the characteristic function

$$\phi_\mu(t)(= \phi_F(t) = \phi_X(t))$$

is

$$\begin{aligned}\int_{-\infty}^{\infty} e^{itx}\mu(\mathrm{dx}) &= \mathbb{E}(e^{itx}), e^{itx} = \cos(tx) + i\sin(tx)\\ &= \int \cos(tx)\mu(\mathrm{dx}) + i\int \sin(tx)\mu(\mathrm{dx})\end{aligned}$$

Baby Facts

- $\phi_\mu(t)$ exists for all t

- $$\phi_\mu(0) = 1$$

- $\phi_\mu(t)$ is uniformly continuous.
  *Proof*

$$\begin{aligned}|\phi_\mu(t+h) - \phi_\mu(t)| &= \left|\int (e^{i(t+h)x} - e^{itx})\mu(\mathrm{dx})\right|\\ &\leqslant \int \left|e^{i(t+h)x} - e^{itx}\right| \mu(\mathrm{dx})\\ &= \int \left|e^{ihx} - 1\right| \mu(\mathrm{dx})\end{aligned}$$

As $h \to 0$ RHS $\to 0$ independent of t.

Three main theorems.
I) X and Y are independent

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$$

II) If $\mu, \nu$ probabilities and

$$\phi_\mu(t) = \phi_\nu(t) \text{ all } t$$

then $\mu = \nu$.
III) $\mu_n \Rightarrow \mu \Leftrightarrow$

$$\phi_{\mu_n}(t) \xrightarrow{n\to\infty} \phi_\mu(t), \forall t$$

**Example** Card guessing.
Deck $\{1, 2, \cdots, n\}$, guess with complete feedback.

$$\text{Let } X_j = \begin{cases} 0 & 1 - \frac{1}{n} \\ j & \frac{1}{j} \end{cases}, S_n = \sum_{j=1}^n X_j$$

**Claim** $\frac{S_n}{n}$ has nontrivial limit,

$$\phi_j(t) = \mathbb{E}(e^{itX_j}) = 1 - \frac{1}{j} + \frac{e^{itj}}{j}$$

$$\mathbb{E}(e^{it\frac{S_n}{n}}) = \prod_{j=1}^n \left(1 - \frac{1 - e^{it/n}j}{j}\right)$$

$$= e^{\sum_{j=1}^n \log\left(1 - \frac{1-e^{\frac{itj}{n}}}{j}\right)}$$

$$\sim e^{-\frac{1}{n}\sum_{j=1}^n \frac{1-e^{it\frac{j}{n}}}{j/n}}$$

$$\to e^{-\int_0^1 \frac{1-e^{itx}}{x}\mathrm{dx}}$$

Points
a) RHS is a characteristic function for some probability $\mu$ and so

$$P(\frac{S_n}{n} \leqslant x) \to \mu(-\infty, x]$$

b) Not all limits are normal or Poisson.
c) What does $\mu$ look like?
d) This $\mu$ is famous.

F.F.T. (See my paper average random time FFT)
Sino-Soviet Feud
Stable laws, $S_n = \sum X_i, X_i$ i.i.d.

$$P(\frac{S_n - a_n}{b_n} \leqslant x) \to F(x)$$

Infinitly divisible.

$$\begin{matrix} X_{11} & \cdots & X_{1n_1} \\ X_{21} & \cdots & X_{2n_2} \\ \vdots & \vdots & \vdots \end{matrix}$$

$$P\left\{\frac{S_n - a_n}{b_n} \leqslant x\right\} \to F(x) ID$$

Class L. $X_1, X_2, X_3, \cdots$ independent, not i.i.d.
G-K "proved" laws in class L are unimodal.
Chung Proof wrong!
Ibragima published counterexample.
Japanese Ibra wrong
Ibragima publish proof
Sun Ibra wrong
Yamazato OK

On $\mathbb{R}$, last time we proved $\{\mu_n\}$ tight, then $\forall n_k, \exists n_{k(i)}$ and probability $\mu$, such that $\mu_{k(i)} \Rightarrow \mu$

**Corollary**: $\{\mu_n\}$ tight and $\forall \{n_k\}$ s.t. $\mu_{n_k} \Rightarrow \mu$, then $\mu_n \Rightarrow \mu$.

*proof* If not, at some x, $\mu\{x\} = 0, \mu_n(-\infty, x] \nrightarrow \mu(-\infty, x]$.

So $\exists \varepsilon > 0, n_k$ s.t. $|\mu_{n_k}(-\infty, x] - \mu(-\infty, x]| > \varepsilon$.

By tightness, $\exists n_{k(i)}$ s.t. $\mu_{n_{k(i)}} \Rightarrow \mu$.

Three main theorems about random variables.

- If $u$ and $v$ have characteristic functions $\mu_u$ and $\mu_v$, then $u * v$ has characteristic function $\phi_u(t)\phi_v(t)$.
  *proof* If $X$ and $Y$ are random variables distributed as $\mu$ and $\nu$.

$$\begin{aligned}\phi_{\mu*\nu}(t) &= \mathbb{E}e^{it(X+Y)} \\ &= \mathbb{E}e^{itX}e^{itY} \\ &= \mathbb{E}e^{itX}\mathbb{E}e^{itY} \\ &= \phi_\mu(t)\phi_\nu(t)\end{aligned}$$

- Inversion and Uniqueness.
  Say $\mu$ is probability on $\mathbb{R}$, characteristic function $\phi(t)$, then if a, b continuous points of $\mu$,

$$\mu(a,b] = \lim_{T\to\infty} \int_{-T}^{T} \frac{e^{-iat} - e^{-ibt}}{it} \phi(t)\mathrm{dt} = I(t)$$

Consequence: $\mu \mapsto \phi$ is 1-1.

*proof* Need $S(T) = \sin c(T) = \int_0^T \frac{\sin x}{x}\mathrm{dx}, T \geqslant 0$

**Fact** 1 $\lim_{T\to\infty} S(T) = \frac{\pi}{2}$

**Fact** 2 $S(\theta T) = \mathrm{sgn}(\theta) S(T|\theta|), T > 0$

$$\begin{aligned}I(T) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_{-T}^{T}\frac{e^{it(x-a)} - e^{it(x-b)}}{it}\mathrm{dt}\right]\mu(\mathrm{dx}) \\ &= 2\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\mathrm{sgn}(x-a)S(|x-a|t) - \mathrm{sgn}(x-b)S(|x-b|t)}{\phi_T(x)}\mu(\mathrm{dx})\end{aligned}$$

$$\phi_T(x) = \begin{cases} 0 & x < a \\ \frac{1}{2} & x = a \\ 1 & a < x < b \\ \frac{1}{2} & x = b \\ 0 & x > b \end{cases}$$

Then by dominated convergence theorem, $\lim_{T\to\infty} I(T) = \mu(a,b]$
*proof*
If $\phi_\mu(t) = \phi_\nu(t)$ for all t,
know $\mu(a,b] = \nu(a,b], \forall\{\mu,\nu\}$ continuity points of $\mu$ and $\nu$.
Such intervals form $\pi$-system generating Borel field.
**Fact** 1 Can't have $\phi_\mu(t) = \phi_\nu(t)$ except at $t = t_0$. (By continuity of characteristic function)
**Fact** 2 Can have $\phi_\mu(t) = \phi_\nu(t), |t| < A$ not all t.

$$\phi_n(t) = 1 - |t|, |t| \leqslant 1$$

compact support.
If $\phi$ is characteristic function of compact support, then

$$\phi_n(t) = 1 - |s|, |s| \leqslant 1, s = t + 2k, k \in \mathbb{Z}$$

$\phi_\nu$ is also a characteristic function.
**Fact** 3 From 2, if $\mu_1 * \mu_2 = \mu_3 \to \mu_3$, can't conclude $\mu_2 = \mu_3$.
For example, $\mu_\mu^2 = \mu_\mu \mu_\nu$
See Feller Introduction to Probability, Volume II, second edition, chapter 15.

- **Theorem** (Continuity Theorem)

$$\mu_n \Rightarrow\Leftrightarrow \phi_n(t) \to \phi(t) \text{ all } t$$

*proof* "$\Rightarrow$" If $\mu_n \Rightarrow \mu$, make $Y_n, Y, Y_n(\omega) \to Y(\omega)$ all $\omega$

$$\mu_n(t) = \mathbb{E}e^{itY_n} \to \mathbb{E}e^{itY} = \phi(t) \text{ Bounded Convergence Theorem}$$

"$\Leftarrow$" $\phi_n(t) \to \phi(t)$, Claim $\{\mu_n\}$ is tight.

$$u > 0$$

$$\begin{aligned}
\frac{1}{u}\int_{-u}^{u}(1-\phi_n(t))\mathrm{dt} &= \int_{-\infty}^{\infty}\left[\frac{1}{u}\int_{-u}^{u}(1-e^{itx})\mathrm{dt}\right]\nu_n(\mathrm{dx}) \\
&= 2\int_{-\infty}^{\infty}(1-\frac{\sin ux}{ux})\mu_n(\mathrm{dx}) \\
&\geqslant 2\int_{|X|\geqslant\frac{2}{u}}(1-\frac{1}{|ux|})\mu_n(\mathrm{dx}) \\
&\geqslant \mu_n\{x : |x| > \frac{2}{u}\}
\end{aligned}$$

(Notice $|x|\mu \geqslant 2, \frac{1}{|ux|} \leqslant \frac{1}{2}$, then $1 - \frac{1}{|ux|} \geqslant \frac{1}{2}$)
Now $\phi_n(t) \to \phi(t)$ all t, and $\phi(0) = 1$ and $\phi$ continuous.
In neighbor of 0, $\exists u$ small, so

$$\frac{1}{u}\int_{-u}^{u}(1-\phi(t))\mathrm{dt} < \varepsilon$$

Fix u, let $n \to \infty, \frac{1}{u}\int_{-u}^{u}(1-\phi_n(t))\mathrm{dt} < 2\varepsilon$, all $n \geqslant n_0$.
So let $a = \frac{2}{u}, n > n_0, \mu_n(x : |x| > a) < 2\varepsilon$. increase a to deal with $\mu_1, \cdots, \mu_{n_0-1}, \{\mu_n\}$ is tight.
$\Rightarrow \mu_n \Rightarrow \mu$ for $\mu_n$ tight and if $\mu_{n_k} \Rightarrow \mu \Rightarrow \mu_n \Rightarrow \mu$.

Variations

- Inversion theorem says $\int_{-\infty}^{\infty}|\phi(t)|\mathrm{dt} < \infty$
then $\lim_{T\to\infty} I(T) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-iat}-e^{-ibt}}{it}\phi(t)\mathrm{dt}$

$$|\frac{e^{-iat}-e^{-ibt}}{it}| < |b-a|$$

So $\mu(a,b] = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-ita}-e^{-itb}}{it}\phi(t) \leqslant |b-a|$.
So can't have jumps.
So OK for $\forall a, b$.

$$\frac{F(x)-F(x+h)}{h} = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-itx}-e^{-it(x+h)}}{ith}\phi(t)\mathrm{dt}$$

and $h \to 0$ exists $\mu$ has density $f(x)$

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}\phi(t)\mathrm{dt}$$

- What/How can we tell about $\mu$ from $\phi$?
Moments $\mathbb{E}(X^k) < \infty$
$\mathbb{E}(X^k) = (-i)^k\phi^{(k)}(0)$ if $\phi$ is k times differentiable.

What about atoms?
If $\mu$ is supported on $\mathcal{L} = \{0, \pm1, \pm2, \cdots\}$

$$\phi(t) = \sum_{j\in\mathcal{L}}p_j e^{itj}, \phi(t+2\pi) = \phi(t)$$

**Theorem** Have density bounded by M $\Leftrightarrow \int_{-\infty}^{\infty}|\phi(t)| < \infty$
(Markov Moment Problem)

- Example of Theorem.
  P.D. was studying $X = \int_0^1 B_1(\omega)\mathrm{dB}_2(\omega)$

$$Y = \int_0^1 B_2(\omega)\mathrm{dB}_1(\omega)$$

  Are they independent? No!

Central Limit Theorem:

$$X_1, \cdots, X_n \text{ random variables}$$

- "not too different" no one dominates

- "not too dependent", then

$$P\{\frac{S_n - \mu}{\sigma_n} \leqslant x\} = \Phi(x), \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{d}t$$

**Example**

$$X_i = \begin{cases} 1 & \frac{1}{i} \\ 0 & 1 - \frac{1}{i} \end{cases}$$

and independent.

$$\mu_n = \sum_{i=1}^{n} \mathbb{E} X_i = \sum_{i=1}^{n} \frac{1}{i} \sim \log n$$

$$\sigma_n^2 = \sum_{i=1}^{n} \frac{1}{i}(1 - \frac{1}{i}) \sim \log n$$

$$P\{\frac{S_n - \log n}{\sqrt{\log n}} \leqslant x\} \sim \Phi(x)$$

**Example**

$$X_i = \begin{cases} i & \frac{1}{i} \\ 0 & 1 - \frac{1}{i} \end{cases}$$

No CLT.
Lindeberg.

$$\begin{array}{ccccc} x_{11} & x_{12} & \cdots & x_{1r_1} & \cdots \\ x_{21} & x_{22} & \cdots & x_{2r_2} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nr_n} & \cdots \end{array}$$

Triangular array.
All $\mathbb{E}(X_{ij}) = 0, \mathbb{E}(X_{ij}^2) = \sigma_{ij}^2 < \infty$.
$X_{n1}, \cdots, X_{nr_n}$ independent.
Condition (Lindeberg).

$$\forall \varepsilon > 0, \frac{1}{\sigma_n^2} \sum_{k=1}^{r_n} \int_{\{\omega : |X_{nk}(\omega)| > \varepsilon \sigma_n\}} X_{nk}(\omega)^2 P(\mathrm{d}\omega) \to 0$$

$$\sigma_n^2 = \sum_{i=1}^{r_n} \sigma_{ni}^2$$

**Theorem** If Lindeberg, then $P\{\frac{S_n}{\sigma_n} \leqslant x\} \to \Phi(x)$.
We prove $\mathbb{E}e^{it\frac{S_n}{\sigma_n}} \to e^{-\frac{t^2}{2}}$
Need
A) If $\varepsilon_i, \omega_j, 1 \leqslant i, j \leqslant n, |\varepsilon_i| \leqslant 1, |\omega_j| \leqslant 1|$, then $|\prod \varepsilon_i - \prod \omega_i| \leqslant \sum_{i=1}^n |\varepsilon_i - \omega_i|$.
*proof* OK $n = 1, \prod \varepsilon_i - \prod \omega_j = (\varepsilon_1 - \omega_1)(\varepsilon_2 \cdots \varepsilon_n) + \omega_1(\varepsilon_2 \cdots \varepsilon_n - \omega_2 \cdots \omega_n)$

B) $\forall x, t, |e^{itx} - (1 + itx - \frac{t^2x^2}{2})| \leqslant \min(|tx|^2, |tx|^3)$
*proof* w.l.o.g. $\sigma_n^2 = 1$
From B)

$$\begin{aligned}
|\phi_{nk}(t) - (1 - \frac{t^2\sigma_{nk}^2}{2})| &\leqslant \mathbb{E}\min(|tX_{nk}|^2, |tX_{nk}|^3) \\
&\leqslant \varepsilon t^3 \int_{|X_{nk}|<\varepsilon} |X_n k|^2 + t^2 \int_{|X_{nk}|\geqslant\varepsilon} |X_{nk}|^2 \\
&\leqslant \varepsilon t^3 \sigma_{nk}^2 + t^2 \int_{|X_{nk}|\geqslant\varepsilon} |X_{nk}|^2
\end{aligned}$$

By Lindeberg:

$$\sum_{k=1}^{r_n} |\phi_{nk}(t) - (1 - \frac{t^2}{2}\sigma_{nk}^2)| \to 0$$

We show

$$\begin{aligned}
\prod_{k=1}^{r_n} \phi_{nk}(t) &\overset{1}{=} \prod_{k=1}^{r_n} (1 - \frac{t^2}{2}\sigma_{nk}^2) + o(1) \\
&\overset{2}{=} \prod_{k=1}^{r_n} e^{-\frac{t^2}{2}\sigma_{nk}^2} + o(1) \\
&= e^{-\frac{t^2}{2}} + o(1)
\end{aligned}$$

1 From Lindeberg:

$$\sigma_{nk}^2 = \int_{|X_{nk}|<\varepsilon} |X_{nk}|^2 + \int_{|X_{nk}|\geqslant\varepsilon} |X_{nk}|^2 \leqslant \varepsilon^2 + \int_{|X_{nk}|\geqslant\varepsilon} |X_{nk}|^2$$

So $\max \sigma_{nk}^2 \to 0$ So $\forall t, 1 - \frac{t^2}{2}\sigma_{nk}^2$ is in $(0, 1)$ when $n$ sufficient large.
Using A):

$$|\prod_{k=1}^{r_n} \phi_{nk}(t) - \prod_{k=1}^{r_n} (1 - \frac{t^2}{2}\sigma_{nk}^2)| \leqslant \sum |\phi_{nk}(t) - (1 - \frac{t^2}{2}\sigma_{nk}^2| \to 0$$

2)

$$|\prod (1 - \frac{t^2}{2}\sigma_{nk}^2) - \prod e^{-\frac{t^2}{2}\sigma_{nk}^2}| \leqslant \sum_{k=1}^{r_n} \sum_{i=1}^{r_n} |e^{-\frac{t^2}{2}\sigma_{nk}^2} - (1 - \frac{t^2}{2}\sigma_{nk}^2)| *$$

$$\forall z \in \mathbb{C}, |e^z - (1+z)| \leqslant |z|^2 \sum \frac{|z|^{k-2}}{k!} \leqslant |z|^2 e^{|z|}$$

So $* \leqslant \frac{t^4}{4} e^{\frac{t^2}{2}} \sum_{k=1}^{r_n} \sigma_{nk}^4 \leqslant \varepsilon^2 e^{\frac{t^2}{2}} \frac{t^2}{4} \sum_{k=1}^{r_n} \sigma_{nk}^2 = \varepsilon^2 e^{\frac{t^2}{2}} \frac{t^2}{4}$

**Example** $X_i = \begin{cases} 1 & \frac{1}{i} \\ 0 & 1 - \frac{1}{i} \end{cases}$

$$\begin{array}{llll} x_1 & & & \\ x_1 & x_2 & & \\ x_1 & x_2 & x_3 & \cdots \end{array}$$

$$\sigma_n^2 \sim \log n, \forall \varepsilon > 0$$

$$\frac{1}{\log n} \sum \int_{|X_{nk} > \varepsilon\sqrt{\log n}} |X_{nk}(\omega)|^2 \mathrm{dP} = 0$$

for all large n.

**Example** Lyapunov Theorem

Suppose have triangular array and $\mathbb{E}|X_{nk}|^{2+\delta} < \infty$ some $\delta > 0$.

If $\frac{1}{\sigma_n^{2+\delta}} \sum_{k=1}^{r_n} \mathbb{E}(|X_{nk}|^{2+\delta}) \to 0$, then CLT holds.

*proof* Look at Lindeberg:

$$\begin{aligned} \frac{1}{\sigma_n^2} \sum_{k=1}^{n} \int |X_{nk}|^2 \mathrm{dP} &\leqslant \frac{1}{\sigma_n^2} \sum_{k=1}^{n} \int \frac{|X_{nk}|^{2+\delta}}{\varepsilon^\delta \sigma_n^\delta} \mathrm{dP} \\ &\leqslant \frac{1}{\varepsilon^\delta \sigma_n^{2+\delta}} \sum \mathbb{E}|X_{nk}|^{2+\delta} \\ &\to 0 \end{aligned}$$

History:

- Demoivre 1750. Proved for Binomial(p)
- Laplace 1780 "proved" for i.i.d. with $\mathbb{E}e^{tX_i} < \infty$
- Chebyshev 1880 gave first rigorous proof. Assume $\mathbb{E}|X_i|^h < \infty$ all h.
- Lyapunov 1920 $\mathbb{E}|X_i|^{2+\delta}$ OK

- Lindeberg-Levy: 1940 Lindeberg Condition.

10 Examples of CLT:

- Sampling without replacement:
  From an urn: urn has balls labels $Y_1, Y_2, \cdots, Y_N \in \mathbb{R}$.
  Let $X_1, X_2, \cdots, X_n$ be results of sampling without replacement.
  **Theorem** (Erdos-Renyi 1910)
  If $\{X_i\}_{i=1}^N$ "not to different" and $\frac{n}{N}$ small.

$$P\{\frac{S_n - \mu_n}{\sigma_n} \leqslant\} \to \Phi(x)$$

- Holfding Combinatorial CLT.
  Let $A = (A_{ij})_{n\times n}$ matrix.
  Pick $\omega$ a permutation of $\{1, 2, \cdots, n\}$ at random.

$$S_n \sum_{i=1}^{n} A_{i\omega(i)}$$

  n large, $A_{ij}$ "not too wild" $P\{\frac{S_n-\mu_n}{\sigma_n} \leqslant x\} \to \Phi(x)$.
  Let $A_{ij} = |i - j|, 1 \leqslant i, j \leqslant n$.

$$S_n - \sum |i - \omega(i)|$$

- Let $\Omega = \{1, 2, 3, \cdots, N\}, P(j) = \frac{1}{N}$
  For primes 2, 3, 5, $\cdots$, $X_p(\omega) = \begin{cases} 1 & p|\omega \\ 0 & \text{otherwise} \end{cases}$

$$S_N(\omega) = \sum_{p\leqslant N} X_p(\omega) = \# \text{ distinct primes that divide } \omega$$

$$S_N(12) = 2$$

  **Theorem** (Erdos-Kac)

$$P\{\frac{S_n - \log\log N}{\sqrt{\log\log N}} \leqslant x\} \to \Phi(x)$$

- Pick $m \in O(n)$ orthogonal matrix.
  **Theorem** $P\{tr(m) \leqslant x\} \xrightarrow{n} \Phi(x)$

$$\sup |P\{tr(M) \leqslant x\} - \Phi(x)| \leqslant \frac{1}{n!}$$

- Say have a graph at each vertex put uniform on $(0,1)$

$$S_N = \#\text{ local max } = \sum_{\nu \in G} X_\nu$$

$$P(\frac{S_N - \sum \frac{1}{d_\nu + 1}}{\sqrt{\sigma_N^2}} \leqslant x) \sim \Phi(x)$$

Poisson Approximation

$$N = \{0, 1, 2, \cdots\}$$
$$P_\lambda(j) = e^{-\lambda}\lambda^j/j!, j \in N$$
$$P_\lambda * P_\mu = P_{\lambda+\mu}$$
$$\lambda \text{ large: } P_\lambda\{\frac{x-\lambda}{\sqrt{\lambda}} \leqslant t\} \xrightarrow{\lambda\to\infty} \Phi(t)$$

Poisson Heuristic:
If $|I| < \infty, X_i \in \{0,1\}, i \in I$

$$P(X_i = 1) = p_i, W = \sum_{i\in I} X_i$$

Let $\lambda = \mathbb{E}(W) = \sum_{i\in I} p_i$
If $|I|$ "large" with $p_i$ "small"
$X_i$ "not too dependent"
then "$P(W = j) \approx P_\lambda(j)$" If $\mu$ and $\nu$ probabilities on $(\Omega, \mathcal{F})$

$$\|\mu - \nu\| = \sup_{F\in\mathcal{F}} |\mu(F) - \nu(F)|$$

**Definition** A dependency Graph. for $\{X_i\}$ is a simple undirected graph. vertex set I, and some edges $(i, j)$. If $\exists A, B \subseteq I$ with no edge from $A$ to $B$, then $\{X_i\}_{i\in A}, \{X_j\}_{j\in B}$ are independent.
**Example** If $\{X_i\}$ are independent, take empty graph on $I$.
For $i \in I$, let $N_i = \{i\} \cup \{j : (i,j) \in E\}$ (neighborhood).
**Theorem** $\{X_i\}$ binary, $z \in \text{Poisson}(\lambda), \Lambda = \sum_{i\in I} p_i, \{X_i\}$ have dependency graph $(I, E)$. then $W = \sum_{i\in I} X_i$

$$\|L(\omega) - L(z)\|_{TV} \leqslant \min(3, \lambda^{-1})\{\sum_{i\in I}\sum_{j\notin N_i \ \{i\}} p_{ij} + \sum_{i\in I}\sum_{j\in W_i} p_i p_j\}$$

**Example** (Poisson)
$\{X_i\}_{i=1}^n$ are independent.
$P(X_i = 1) = p$, then we get empty dependency graph.
...

I.O.U. $P_\lambda(A) = \sum_{j\in A} \frac{e^{-\lambda}\lambda^j}{j!}, A \subseteq \mathbb{N}$

**Property** (Solving Stein Equation)

$$\forall A \subseteq \mathbb{N}, \exists! f : \mathbb{N} \to \mathbb{R}, s.t.$$

- $$f(?) = 0$$
- $$f(j+1) - jf(j) = \delta_A(j) - P_\lambda(A), \text{ all } j \in \mathbb{N}$$
- $$|f(j)| \leqslant 1.25, |f(j+1) - f(j)| \leqslant \min(3, \frac{1}{\lambda})$$

*Proof* Clearly, there is a unique solution

$$\text{e.g. } f(0) = 0, f(1) = \delta_A(1) - P_\lambda(A)$$

? Write down $f(j)$, multiply ? by $\frac{\lambda^j}{j!}$

$$\frac{\lambda^{j+1}}{j!} f(j+1) - \frac{\lambda^j}{(j-1)!} f(j) = \frac{\lambda^j}{j!}(\delta_A(j) - P_\lambda(A))$$

Sum this in j up to $k-1$,

$$\frac{\lambda^k}{(k-1)!} f(k) = \sum_{j=1}^{k-1} \frac{\lambda^j}{j!}(\delta_A(j) - P_\lambda(A))$$

- $$f(k) = \frac{(k-1)!}{\lambda^k} \sum_{j=1}^{k-1} \frac{\lambda^j}{j!}(\delta_A(j) - P_\lambda(A))$$
- $$f(k) = -\frac{(k-1)!}{\lambda^k} \sum_{j=k}^{\infty} \frac{\lambda^j}{j!}(\delta_A(j) - P_\lambda(A))$$

(Because $\sum_{j=0}^{\infty} \frac{\lambda^j}{j!}(\delta_A(j) - P_\lambda(A)))$

Step 2. Bound $f(k)$ use $|\delta_A(j) - P_\lambda(A)| \leqslant 1$.

- $k \leqslant \lambda + \frac{1}{5}$ use

$$\begin{aligned} |f(k)| &\leqslant \sum_{j=1}^{k-1} \frac{(k-1)!}{\lambda^k} \frac{\lambda^j}{j!} \\ &\leqslant \frac{1}{\lambda} \sum_{j=0}^{k} (\frac{k-1}{\lambda})^{k-(j+1)} \\ &\leqslant \frac{1}{\lambda} \frac{1}{1-\frac{k-1}{\lambda}} \\ &= \frac{1}{(\lambda-(k-1))} \end{aligned}$$

have $k-1 < \lambda + \frac{1}{5} - 1$ or $1 - \frac{1}{5} < \lambda - (k-1)$
or $\frac{1}{\lambda-(k-1)} < \frac{5}{4} = 1.25$

$$|f(k)| \leqslant 1.25$$

- 

$$\begin{aligned} k &> \lambda + \frac{1}{5} \\ |f(k)| &\leqslant \sum_{j=k}^{\infty} \frac{(k-1)!}{\lambda^k} \frac{\lambda^j}{j!} \\ &\leqslant \sum_{j=0}^{\infty} \frac{\lambda^m}{k(k+1)^m} \\ &= \frac{1}{k} \frac{1}{1-\frac{\lambda}{k+1}} \\ &= \frac{k+1}{k(k+1-\lambda)} * \end{aligned}$$

Here $k > \lambda + \frac{1}{5}, k+1 > \lambda + 1 + \frac{1}{5}, k+1-\lambda > \frac{6}{5}$ or $\frac{1}{k+1-\lambda} < \frac{5}{6}$
So $* < \frac{k+1}{k}\frac{5}{6} \Rightarrow |f(k)| < 1.25$ for $k \geqslant 2$

$$f(1) = \lambda^{-1}(\delta_A(0) - P_\lambda(A)$$

$$|f(1)| \leqslant \frac{1}{\lambda}(1-e^{-2}) < 1$$

Largest $A = \{0\}$, smallest $A = \{1, 2, 3, \cdots\}$.
So $|f(k)| \leqslant 1.25$, and $|f(k+1) - f(k)| \leqslant 3$. We need H.W. $|f(k+1) - f(k)| \leqslant \frac{1}{\lambda}$.
3 Basic problems of elementary problem (and variations)

- Birthday Problem
- Coupon Collection
- Matching

I) Let $(V,E)$ be simple graph, vertex set V, edge set E.
Color vertices with c colors. Let W be number of monochromatic edges.
Choose color i with probability $p_i$
(Classical Birthday: $|V| = 23$, complete graph, $p_i = \frac{1}{c}$)
For Poisson set up, $I = \{(i,j) : (i,j) \in E\}$

$$X_i = \begin{cases} 1 & \text{vertex i and j have same color} \\ 0 & \text{otherwise} \end{cases}$$

$$W = \sum_{i \in I} X_i$$

$$\mathbb{E}(X_{(i,j)}) = \sum \theta_a^2$$

$$\lambda = \mathbb{E}(W) = |E| \sum_{a=1}^{c} \theta_a^2$$

Poisson heuristic
If $|E|$ large, $\theta_a$ small with $|E| \sum_a \theta_a^2$ moderate.

$$P(W = j) = \frac{e^{-\lambda} \lambda^j}{j!}$$

$$P(W = 0) = 1 - e^{-\lambda}$$

II) Coupon Collections
N boxes, (outcomes)
Buy $k$ coupons.

$$P(\text{coupon} = i) = \theta_i, 1 \leqslant i \leqslant N$$

Let $W$ = chance all boxes have 1 or more balls.
**Example** How many people in village to have chance $\geqslant \frac{1}{2}$ somebody has birthday on each of $N = 365$ days. (Answer = 2700)

$$X_i = \begin{cases} 1 & \text{Box i empty} \\ 0 & \text{otherwise} \end{cases}$$

$$W = \sum_{i=1}^{N} X_i = \#\text{empties}$$

We want $P(W = 0)$

$$? = (1 - \theta_i)^k, \lambda = \sum_{i=1}^{N} (1 - \theta_i)^k$$

Poisson heuristic says $N$ and ? large, $theta_i$ small and $\lambda$ "is a number".
Then $P(W = 0) \sim e^{-\lambda}$
**Example** $\theta_i = \frac{1}{N}, \lambda = N(1 - \frac{1}{N})^k$.

$$\begin{aligned} ? &= N \log N + C? \\ \lambda &= N(1 - \frac{1}{N})^k \\ &= e^{k \log(1 - \frac{1}{N}) + \log N} \\ &\sim e^{-\frac{k}{N} + \log N} \\ &= e^{-c} \end{aligned}$$

Variation $\theta$ uniform for $\{\theta_i : \sum \theta_i = 1\} A_N$

$$h = 100,000, W = 365$$

III) Matching (Monmort 1708)
Classical version:
Let $\Omega = S_n =$ All n! permutations, pick $\omega \in \Omega$ uniformly,

$$W(\omega) = \#\{i : \omega(i) = i\}$$

$$X_i(\omega) = \begin{cases} 1 & \omega(i) = i \\ 0 & \text{otherwise} \end{cases}$$

$$p_i = P(X_i = 1) = \frac{1}{n}$$

$$\lambda = n \frac{1}{n} = 1$$

$$P(W = 0) = e^{-1}$$

$$P(W > 0) = 1 - e^{-1} \leqslant \frac{2}{3}$$

Variations: say deck has $c_i$ cards. Type i,

$$1 \leqslant i \leqslant C, \sum_{i=1}^{n} c_i = N$$

Shuffle each of two such decks, $W = P(\text{at least one match})$

$$X_i, 1 \leqslant i \leqslant N, X_i = \begin{cases} 1 & \text{Matching ?} \\ 0 & \text{otherwise} \end{cases}$$

?

Problem 4. Pick $\omega \in S_n$ uniformly. Let

$$W(\omega) = \#\{i : \omega(i+1) = \omega(i) + 1\}$$

Find good approximation for $P(W = j)$ (and prove your answer)
For proof, look in paper by A?-Gorden-Goldstein (Statistical Science). Find a version of Steins method that fits and check condition.