Math 833 De Finetti's theorem Week 3

$\Omega = \{0,1\}^{\mathbb{Z}_+}$ – infinite sequences of 0's and 1's

$\xi_1, \xi_2, \xi_3, \dots$

It has a natural σ-algebra spanned by cylinders of the form $\{\xi \mid \xi_1 = i_1, \dots, \xi_n = i_n\}$

[This is the minimal σ-algebra, which makes all coordinate maps $\xi \to \xi_i$ measurable]

We are studying probability measures μ on Ω with respect to the above σ-algebra. In other words, we are interesting in various possible distributions of random $\xi = (\xi_1, \xi_2, \dots)$

Definition: $S(n)$ – symmetric group of rank n – all permutations of n objects, i.e. bijections $\sigma: \{1,\dots n\} \to \{1,\dots n\}$

We embed $S(n) \subset S(n+1)$ as the subgroup fixing $n+1$, i.e. $\sigma(n+1) = n+1$, and define infinite symmetric group $S(\infty)$ as the union $S(\infty) = \bigcup_{n=1}^{\infty} S(n)$. Elements of $S(\infty)$ are bijections $\sigma: \{1,2,\dots\} \longrightarrow \{1,2,\dots\}$ which fix all but finitely many elements.

$S(\infty)$ acts on Ω permuting the coordinates:

$$\sigma(z_1, z_2, \dots) = (z_{\sigma^{-1}(1)}, z_{\sigma^{-1}(2)}, \dots)$$

note inversion!

$S(\infty)$ also acts on measures on Ω:

$$\sigma(\mu)[A] = \mu[\sigma^{-1}(A)]$$

Question: What are all $S(\infty)$-invariant probability measures on Ω?

corresponding random sequences are called „exchangeable"

Lemma: $S(\infty)$-invariant probability measures form a convex set.

Proof: It means that for any two $S(\infty)$-invariant probability measures μ and ν and for each $0<t<1$, $t\mu + (1-t)\,\nu$ is also a $S(\infty)$-invariant probability measure.

$(t\mu + (1-t)\,\nu)[M] = t\mu[M] + (1-t)\,\nu[M]$ Which is straightforward $\blacksquare$

Notation: $\Omega_{ex}^{S(\infty)}$ – all extreme points of the set of $S(\infty)$-invariant measures

Proposition : The following definitions of being extreme are equivalent:

1) μ is such that if $\mu = \alpha \nu_1 + (1-\alpha)\nu_2$ for $0 \le \alpha \le 1$ and two $S(\infty)$-invariant probability measures ν_1 and ν_2, then either $\alpha = 0$ or $\alpha = 1$. (indecomposable)
(we assume $\nu_1 \neq \nu_2$)

2) μ is such that for each $S(\infty)$-invariant measurable subset $A \subset \Omega$ either $\mu(A) = 0$ or $\mu(A) = 1$
(ergodic)

Proof: Suppose μ is indecomposable and take $S(\infty)$-invariant $A \subset \Omega$.

$$\mu = \mu|_A + \mu|_{A^c} = \mu(A) \cdot \frac{\mu|_A}{\mu(A)} + \mu(A^c) \cdot \frac{\mu|_{A^c}}{\mu(A^c)}$$

restriction to A

complement

$\Downarrow$

Either $\mu(A)$ or $\mu(A^c)$ should be $= 0$.

$\mu|_A(B) = \mu(A \cap B)$

Next, suppose that μ is ergodic and decompose it as $\mu = \alpha \nu_1 + (1-\alpha)\nu_2$. If $\nu_1 \neq \nu_2$, then we can choose a $S(\infty)$-invariant set $A \subset \Omega$, such that

$\nu_1(A) \neq \nu_2(A)$. But either $\mu(A)=0$ or $\mu(A)=1$.

In the first case $0 = \mu(A) = \alpha\,\nu_1(A) + (1-\alpha)\,\nu_2(A)$, which can be then possible only if $\alpha = 0$ or $(1-\alpha)=0$.

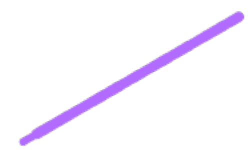
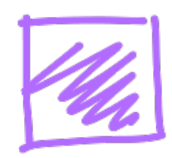
In the second case $0 = \mu(A^c) = \alpha\,\nu_1(A^c) + (1-\alpha)\,\nu_2(A^c)$ which again implies either $\alpha=0$ or $1-\alpha=0$, since necessarily at least one of the numbers $\nu_1(A^c)$, $\nu_2(A^c)$ is positive. ∎

Exercise 1: Consider the measure μ which assigns weight $\frac{1}{2}$ to the singleton $(0,0,0,\dots) \in \Omega$ and weight $\frac{1}{2}$ to $(1,1,\dots) \in \Omega$.

- Is it $S(\infty)$-invariant?
- If yes, then is it extreme?

Theorem 1: The convex set of $S(\infty)$-invariant probability measures is a simplex, which means that each measure has a unique decomposition into linear combination (or integral) of extreme ones.
[with non-negative coefficients of sum = 1]

Analogy: in $\mathbb{R}^2$:

- line  is a simplex (with two exteme points)
- triangle  is a simplex (with three extreme points)
- square  is a convex set with four extreme points, but NOT a simplex

---

Exercise 2: Prove these claims for line, triangle, and square.

Theorem 2: Extreme $S(\infty)$-invariant measures are i.i.d. Bernoulli: $\Omega_{ex}^{S(\infty)} \cong \{0 \le p \le 1\}$ with p corresponding to the i.i.d. sequence $(\xi_1, \xi_2, \dots)$ satisfying $\mathrm{Prob}(\xi_i = 1) = p$.

Remark: $S(\infty)$-invariance of i.i.d. Bernoulli's is obvious, they are also supported on disjoint subsets of Ω due to LLN, which implies that one of them can not be a linear combination of others. The hardest part of the theorem is that there are no other extreme $S(\infty)$-invariant measures.

Corollary 1: Let μ be a $S(\infty)$-invariant probability measure on Ω. There exists a unique probability measure $\nu = \nu(\mu)$ on $[0,1]$, such that $\forall$ measurable $A \subset \Omega$ we have

$$\mu[A] = \int_0^1 B_p[A]\, \nu(dp)$$

Law of i.i.d. Bernoulli-p sequence

Each $S(\infty)$-invariant measure is a mixture of i.i.d. Bernoulli's.

Proof of corollary 1 (modulo Theorems 1 and 2):

By Theorem 1 μ is uniquely decomposed into a convex linear combination of extreme measures.
By Theorem 2 these extreme measures are B_p.
ν is defined as the measure governing the decomposition.

But how do you find ν in practice?

Theorem 3: In notations of Corollary 1 let $(\xi_1, \xi_2, \ldots)$ be μ-distributed. Then the limit

$$\lim_{n\to\infty} \frac{\xi_1 + \ldots + \xi_n}{n}$$

exists almost surely and its distribution is ν.

Compare with Week 2, Slide 5, Theorem 2

Proof of Theorem 3 (modulo Theorems 1 and 2):

By Corollary 1, $(\xi_1, \xi_2, \ldots)$ can be sampled by a two-step procedure:

- First, sample $0 \le p \le 1$ according to ν
- Then sample i.i.d. Bernoulli(p) sequence.

By Strong Law of Large Numbers $\frac{\xi_1 + \ldots + \xi_n}{n} \xrightarrow{n\to\infty} p$ $\blacksquare$

Bruno De Finetti

Born	13 June 1906 Innsbruck, Austria-Hungary
Died	20 July 1985 (aged 79) Rome, Italy
Nationality	Italian
Alma mater	Politecnico di Milano

Corollary 1 and Theorem 3 were brought forward by probabilist/statistician De Finetti.

They play a conceptual role for Bayesian statistics, since a randomly-distributed parameter p of the model appears by itself as a corollary of $S(\infty)$-invariance („exchangeability")

We make steps towards Th 1,2 in the rest of the lecture and finish the proof next week.

Toy problem: Consider $S(n)$-invariant probability measures on $(\xi_1, \dots \xi_n) \in \{0,1\}^n$. What are they?

Proposition: They form a simplex with $n+1$ extreme points parameterized by $k = 0, 1, \dots, n$. The k-th extreme point μ_k^n is the uniform measure on all $(\xi_1, \dots, \xi_n)$ such that $\xi_1 + \dots + \xi_n = k$

Proof: Take any $S(n)$-invariant measure p
Note that any two sequence with the same $\xi_1 + \dots + \xi_n$ can be transformed one into another by some permutation $\sigma \in S(n)$.
Hence, the value of p on them coincides.

Therefore, $\rho = \sum_{k=0}^{n} \alpha^k \cdot \mu_k^n$, where

$$\alpha^k = \binom{n}{k} \cdot \rho(\underbrace{1,\dots,1}_{k}, \underbrace{0,\dots,0}_{n-k})$$

number of sequences such that $z_1 + \dots + z_n = k$

- $\alpha^k \geq 0$ and $\sum_{k=0}^{n} \alpha^k = 1$ (since ρ and each μ_k^n are probability measures.

- There is a bijection between $S(n)$-invariant measures ρ and $(n+1)$-tuples $(\alpha^0, \alpha^1, \dots, \alpha^n)$ ∎

Remark: The numbers $\alpha^0, \dots, \alpha^n$ define a probability measure on $\{0, 1, \dots n\}$. Hence, we get a correspondence

$S(n)$-invariant measures on $\{0,1\}^n$ $\longleftrightarrow$ measures on $\{0,1,\dots n\}$

Conclusion: For finite n extreme invariant measures are parameterized by values of $\xi_1+\dots+\xi_n$ (or $\frac{\xi_1+\dots+\xi_n}{n}$). Theorems 1 and 2 provide $n\to\infty$ limit of this statement

---

We proceed to $n=\infty$ case

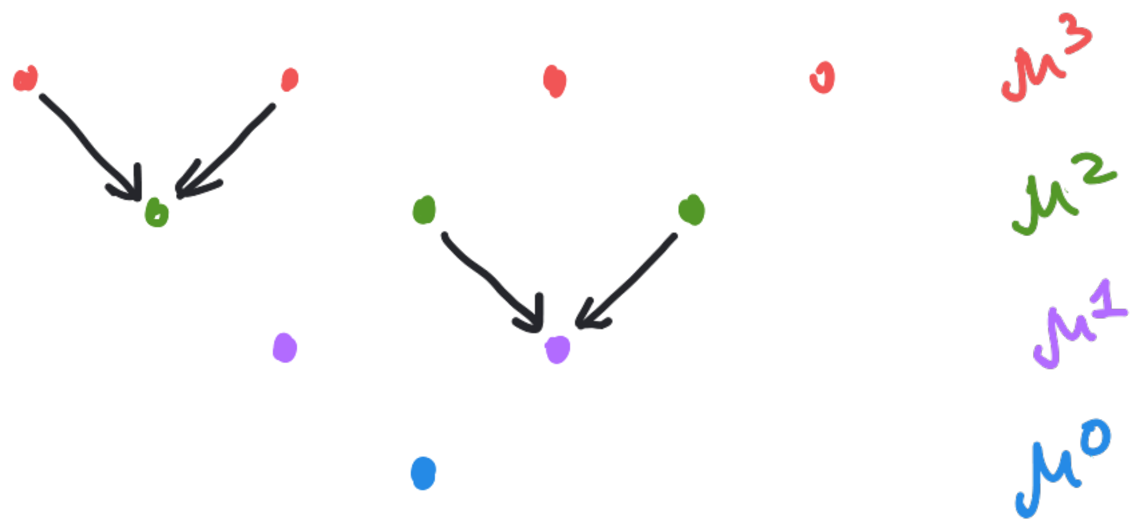
Theorem 4: Given a $S(\infty)$-invariant measure μ, we construct a sequence of measures $(\mu^0, \mu^1, \mu^2, \dots)$, where μ^k is a probability measure on $\{0, 1, \dots k\}$ encoding the distribution of $\xi_1+\dots+\xi_k$ (with μ-distributed $(\xi_1, \xi_2, \dots)$)

Then:

1) We have (for each $0\le i\le k$)
$$\mu^k(i) = \frac{k+1-i}{k+1}\mu^{k+1}(i) + \frac{i+1}{k+1}\mu^{k+1}(i+1)$$
coherency relation

2) $\mu \leftrightarrow (\mu^1, \mu^2, \dots)$ is an isomorphism of convex sets

The way to think about it: we deal with



coherent systems of measures

Each μ^n is an n-dimensional object.
($n+1$ numbers $\mu^n(0), \dots, \mu^n(n)$ subject to $\mu^n(0) + \dots + \mu^n(n) = 1$)

Knowing μ^n we uniquely reconstruct $\mu^{n-1}, \dots, \mu^0$ by iteratively using coherency relations

---

Note a close similarity with setting of Appell sequences from the previous week ?

Proof of Theorem 4: First, take a $S(\infty)$-invariant μ.

Our task is to prove the coherency relation

$$\mu^K(i) = \mathrm{Prob}(\xi_1+\dots+\xi_K = i) = \sum_{j=0}^{K+1} \mathrm{Prob}(\xi_1+\dots+\xi_{K+1}=j) \cdot \mathrm{Prob}(\xi_1+\dots+\xi_K=i \mid \xi_1+\dots+\xi_{K+1}=j)$$

Since $\xi_{K+1} \in \{0,1\}$, only $j=i$ and $j=i+1$ give non-zero terms. Hence,

$$\mu^K(i) = \mathrm{Prob}(\xi_1+\dots+\xi_K=i \mid \xi_1+\dots+\xi_{K+1}=i)\, \mu^{K+1}(i) + \mathrm{Prob}(\xi_1+\dots+\xi_K=i \mid \xi_1+\dots+\xi_{K+1}=i) \cdot \mu^{K+1}(i+1)$$

It remains to compute two conditional probabilities

The law of $(\xi_1, \xi_2, \dots, \xi_{K+1})$ given $\xi_1+\dots+\xi_{K+1}=i$ is the uniform measure on $\binom{K+1}{i}$ possible sequences of 0/1.
(by $S(n)$-invariance)

Only those $\binom{K}{i}$ sequences, which have $\xi_{K+1}=0$ give $\xi_1+\dots+\xi_K=i$

Therefore,

$$\text{Prob}(z_1+\dots+z_k=i \mid z_1+\dots+z_{k+1}=i) = \frac{\binom{k}{i}}{\binom{k+1}{i}} =$$

$$= \frac{\frac{k!}{i!\,(k-i)!}}{\frac{(k+1)!}{i!\,(k+1-i)!}} = \frac{k+1-i}{k+1}$$

\# of sequences of (i) 1's and $(k-i)$ 0's

Similarly,

$$\text{Prob}(z_1+\dots+z_k = i \mid z_1+\dots+z_{k+1}=i+1) = \frac{\binom{k}{i}}{\binom{k+1}{i+1}}$$

\# of sequences of $(i+1)$ 1's and $(k-i)$ 0's

$$= \frac{\frac{k!}{i!\,(k-i)!}}{\frac{(k+1)!}{(i+1)!\,(k-i)!}} = \frac{i+1}{k+1}$$

We have shown that each μ gives rise to a coherent system $(\mu^0, \mu^1, \dots)$. It remains to show that each coherent system uniquely determines corresponding $S(\infty)$-invariant μ.

Take $(\mu^0, \mu^1, \dots)$ and define the law of μ-random sequence $(z_1, z_2, \dots) \in \Omega$ through

$$\mathrm{Prob}_\mu\left(z_1 = x_1, \dots z_n = x_n\right) = \frac{\mu^n(x_1 + \dots + x_n)}{\binom{n}{x_1 + \dots + x_n}}$$

- This is a probability measure, since for each $n = 1, 2, \dots$

$$\sum_{x_1, \dots x_n} \frac{\mu^n(x_1 + \dots + x_n)}{\binom{n}{x_1 + \dots + x_n}} = \sum_{i=0}^{n} \mu^n(i) = 1$$

- This is a consistent definition, since by coherency relations, the definition for $n = k$ is obtained from the one for $n = k+1$ by summing over (two) possible values for z_{k+1}

- This is $S(\infty)$-invariant measure, since it is $S(n)$ invariant for each $n=1,2,\ldots$, because $x_1+\ldots+x_n = x_{\sigma(1)}+\ldots+x_{\sigma(n)}$, for each $\sigma\in S(n)$

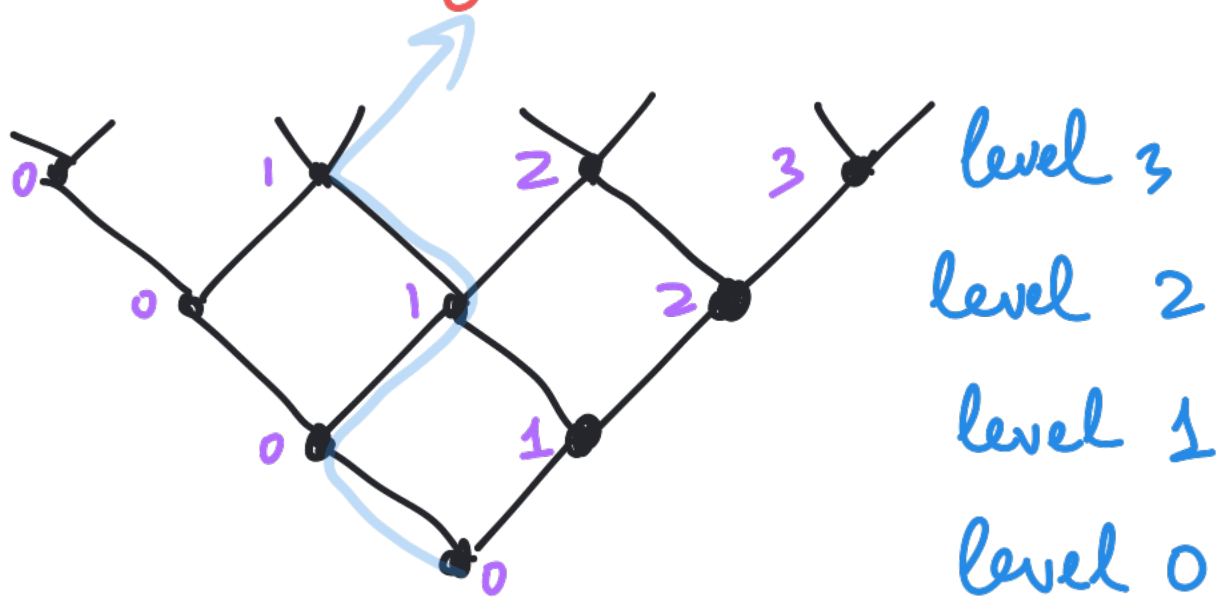
- By Kolmogorov consistency theorem, the probabilities $\mathrm{Prob}(\xi_1=x_1,\ldots,\xi_n=x_n)$ uniquely fix the law of the infinite sequence $(\xi_1,\xi_2,\ldots)$

Finally, it is clear from the construction that whenever $\mu=\alpha\nu+(1-\alpha)\tilde{\nu}$, $0\le\alpha\le1$, then also $\mu^n=\alpha\,\nu^n+(1-\alpha)\,\tilde{\nu}^n$. Hence, the correspondence agrees with the structure of convex sets.

Exercise 3: Let μ be the law of the sequence of i.i.d. Bernoulli random variables. Compute the measures $(\mu^0, \mu^1, \dots)$ and check directly that they satisfy the coherency relations

Here is another point of view on Theorem 4:

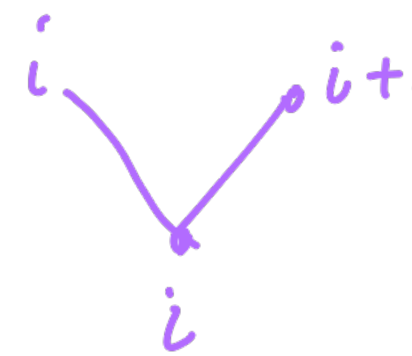
Pascal's graph



Vertices on level K:

$\{0, 1, \dots K\}$

Edges link levels K and $K+1$



Paths in graph

$\updownarrow$

sequences (vertex at level 0, vertex at level 1, ...).
following edges of the graph

$\updownarrow$

$(z_1, z_2, \dots) \in \Omega$ with $z_K = 0 \to$ choose left edge
$z_K = 1 \to$ choose right edge

Under the graph point of view:

$S(\infty)$-invariant random $(\xi_1, \xi_2, \dots)$

$$\Updownarrow$$

Central measure on paths, which means that all initial segments (vertex(0), vertex(1), ..., vertex(k)) are equiprobable, given a fixed value of vertex(k)

$$\Updownarrow$$

Marginals $\mu^k(i) = \mathrm{Prob}(\text{path goes through } i \text{ at level } k)$
$= \mathrm{Prob}(\xi_1 + \dots + \xi_k = i)$

form a coherent system.

Pascal's graph is the simplest example of a branching graph

(generalities coming next week)