Donsker's Theorem

- Donsker's Theorem as extension of classical CLT
- (Scaled, centered) ECDF $\rightsquigarrow$ Brownian Bridge
- Kolmogorov distribution characterizes largest amplitude of standard Brownian Bridge
- Kolmogorov-Smirnoff test straightforward application of Donsker's Thm & Kolmogorov distribution to check equality of (continuous, 1D) distributions.

I) Probability Spaces and Random Variables

A probability space is a tuple $(\Omega, \mathcal{F}, \mathbb{P})$ where

- Ω is a set of outcomes
- $\mathcal{F}$ is a σ-algebra of subsets of Ω, called events
  - $\Omega \in \mathcal{F}$
  - $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
  - $(A_n)_{n\in\mathbb{N}} \in \mathcal{F} \Rightarrow \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$.
- $\mathbb{P}$ is a (probability) measure:

$$\mathbb{P}: \mathcal{F} \to [0,1] \qquad A \mapsto \mathbb{P}(A)$$

- $\mathbb{P} \geq 0$
- $A \cap B = \phi \Rightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$

Parallel between topology and measure theory

From Williams Prob w/ Mart.

Topology

- Building block → open sets
- Characterize continuity: $f: (S_1, \mathcal{O}_1) \to (S_2, \mathcal{O}_2)$

  is continuous if $f^{-1}(O_2) \in \mathcal{O}_1 \quad \forall O_2 \in \mathcal{O}_2$.

  structure preserving

- Axiomatize notion of 'open set'
  - * arbitrary unions open
  - * finite intersections open

Measure Theory

- Building block → measurable sets
- Characterize measurability: $f: (S_1, \mathcal{F}_1) \to (S_2, \mathcal{F}_2)$

  is measurable if $f^{-1}(A_2) \in \mathcal{F}_1 \quad \forall A_2 \in \mathcal{F}_2$

- Axiomatize notion of 'measurable set'

  (see σ-algebra def. above)

! Need only countable ops requirement; else self-contradicting: Banach-Tarski paradox.

! You cannot avoid measure theory

A random variable on $(\Omega, \mathcal{F}, \cdot)$ is a measurable function $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$

↳ Borel σ-algebra (topology-induced)

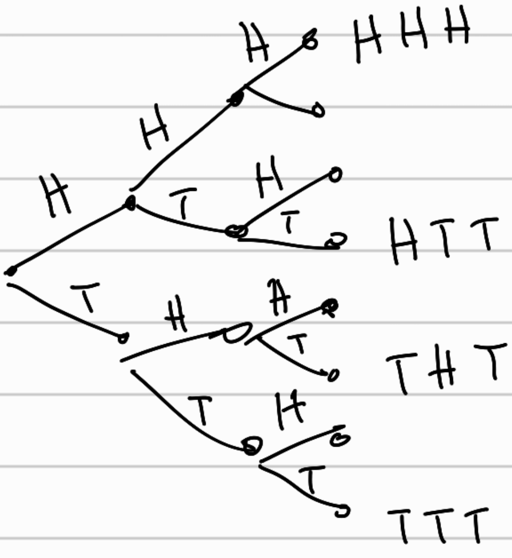
$$\left(\mathcal{B}(\mathbb{R}) = \sigma(\pi(\mathbb{R})) := \sigma\left(\{ (-\infty, x] : x \in \mathbb{R} \} \right) \right)$$

↑ π-system: closed under intersections

! A "random variable" is $\begin{cases} \text{1) not random} \\ \text{2) not a variable} \end{cases}$

Example

$\Omega = \{$



$\}$

$$\mathcal{F} = 2^{\Omega} = \sigma\left(\{\phi, \Omega\} \cup \mathcal{F}_{X_1=H} \cup \mathcal{F}_{X_2=H} \cup \mathcal{F}_{X_3=H}\right)$$

For example $\mathcal{F}_{X_1=H} = \{\phi, \Omega, \text{"}\{Hxx\}\text{"}, \text{"}\{Txx\}\text{"}\}$

! Being told "$X_1=H$" is not sufficient information to "evaluate/carry out measurement" of things like "Second toss is tails" or "There were an odd # of tails"

RV $X(\omega_1 \omega_2 \omega_3) = \begin{cases} 1 & \omega_1 = H \\ 0 & \text{else} \end{cases}$

$Y(\omega_1 \omega_2 \omega_3) = \begin{cases} 1 & \omega_1 = H \text{ and } \omega_2 = \omega_3 \\ 0 & \omega_1 = T \text{ and } \omega_2 = \omega_3 \\ \sqrt{\pi} & \text{else} \end{cases}$

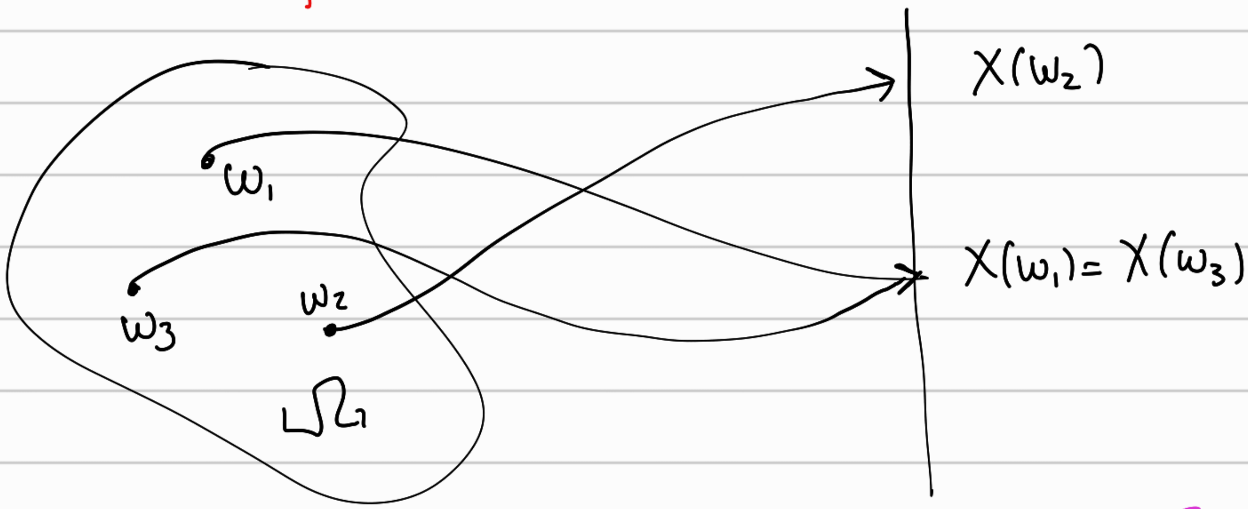
$Y^{-1}[\{1\}] = \{HHH, HTT\}$

$Y^{-1}[\{0\}] = \{THH, TTT\}$

$Y^{-1}[\{\sqrt{\pi}\}] = \Omega \setminus Y^{-1}[\{1, 0\}]$.

! $Y^{-1}[A \text{ op } B] = Y^{-1}[A] \text{ op } Y^{-1}[B]$

! RVs are functions



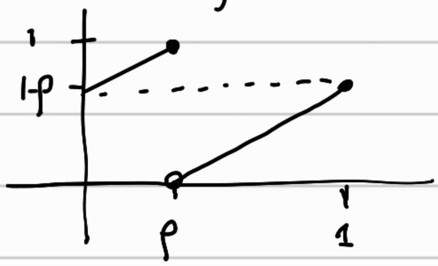
Q. How to approach issue of convergence?

Yes, they are "just functions". But they are fns on a space with more than just a set structure.

Ex $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), \lambda(a,b] \mapsto b-a)$

Thus $X(\omega) = \omega$ is a "uniform 0-1" random variable

so is $1-X$; or



$\{X_p\}_{p \in (0,1)}$

! Uncountably many "$U[0,1]$ rvs"

Q. What do all rvs have in common?

Co-domain!

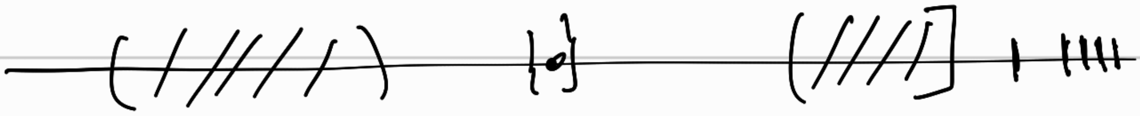
$$(\Omega_1, \mathcal{F}_1, \mathbb{P}_1) \xrightarrow{X_1} (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$
$$(\Omega_2, \mathcal{F}_2, \mathbb{P}_2) \xrightarrow{X_2} (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

Note X ... $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$:

Note X induces a measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

$$X\# \mathbb{P}(A) := \mathbb{P}(\underbrace{X^{-1}[A]}_{\mathcal{F}}) \in [0,1].$$

But Borel sets too flexible, make analysis of $X\#\mathbb{P}$ hard:



The distribution function F of a random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ is defined as:

$$F: \mathbb{R} \to [0,1]$$
$$x \mapsto \mathbb{P}(X \leq x) = \mathbb{P}(X^{-1}(-\infty, x])$$
$$= X\#\mathbb{P}\underbrace{(-\infty, x]}_{\text{focus only on half-spaces}}$$

We say a sequence of random variables $X_1, X_2, \dots$ converges in distribution to a random variable X if

$$\lim_{n\to\infty} F_n(x) = F(x) \; ; \text{ write } X_n \rightsquigarrow X$$

(for all x at which F is continuous)

$\hookrightarrow$ Ex $\mathbb{1}_{\dots} \rightsquigarrow 0$ but $F(0)=1$

$(0, \frac{1}{n})$

while $F_n(0) = 0$.

! Note that $\overset{d}{\rightsquigarrow}$ is a statement about the CDFs, not about the random variables.

! "Weakest type of convergence"

$\overset{d}{\rightsquigarrow}$ Allows us to approximate probability statements on the seq $X_1, X_2, \ldots$ with prob. statements on X:

$$\lim_{n \to \infty} \mathbb{P}(X_n \in B) = \mathbb{P}(X \in B).$$

II) Central Limit Theorem: Old and New

Abraham de Moivre

$\rightsquigarrow$ Pierre-Simon Laplace

$\rightsquigarrow$ Aleksandr Lyapunóv

Let $X_1, X_2, \ldots$ be a sequence of iid r.vs with mean μ and finite variance σ^2. Define the sample average:

$$\overline{X}_n := \frac{X_1 + X_2 + \dots + X_n}{n}$$

! LLN implies $\overline{X}_n \xrightarrow{a.s} \mu$ (and also $\xrightarrow{P}$)

CLT: $$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \rightsquigarrow \mathcal{N}(0,1)$$

Slutsky's Theorem

$X_n \rightsquigarrow X$ and $Y_n \xrightarrow{P} c$, constant. Then

$X_n + Y_n \rightsquigarrow X + c$, $X_n Y_n \rightsquigarrow cX$, $\frac{X_n}{Y_n} \rightsquigarrow \frac{X}{c}$

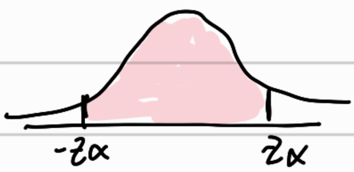
$(c \neq 0)$

Continuous Mapping Theorem

If g is continuous $\mathbb{P}$-a.e, then

$$X_n \xrightarrow{d/p/a.s} X \implies g(X_n) \xrightarrow{d/p/a.s} g(X)$$

Build confidence interval for μ:



$\mathbb{P}\left(-z_\alpha \leq \overline{X}_n - \mu \leq z_\alpha \frac{\sigma}{\ldots}\right) \to \mathbb{P}(|Z| \leq z_\alpha)$

$$\mathbb{P}\left(-z_\alpha \le \frac{\quad}{\sigma/\sqrt{n}} \le z_\alpha \frac{\quad}{\sqrt{n}} \right), \quad \mathbb{P}\left(\quad \le \quad \alpha \right)$$

$\uparrow$ $\mathcal{N}(0,1)$

$$\|$$

$$\mathbb{P}\left(\bar{X}_n + z_\alpha \frac{\sigma}{\sqrt{n}} \ge \mu \ge \bar{X}_n - z_\alpha \frac{\sigma}{\sqrt{n}} \right)$$

$$\|$$

$$\mathbb{P}\left(\bar{X}_n \pm z_\alpha \frac{\sigma}{\sqrt{n}} \ni \mu \right)$$

! This is a prob. statement about random interval, not about μ.

! Note that sample variance $S_n^2 \xrightarrow{P} \sigma^2$ (i.e consistent)
Thus can apply Slutsky's Theorem (and CLT) to replace σ by $\sqrt{S_n^2}$ in CI.

Now consider the empirical CDF (also a rv, more precisely a stochastic process)

$\hookrightarrow$ Collection of rvs $(X_t : t \in T)$ on $(\Omega, \mathcal{F}, \mathbb{P}) \mapsto (S, \Sigma)$.

$$F_n(u) := \frac{1}{n} \sum_{i=1}^{n} \underbrace{\mathbb{I}_{\{X_i \le u\}}}$$

This is a Bernoulli$(F(x))$

$$\mathbb{E}\,\mathbb{I}_{\{X_i \le x\}} = \mathbb{P}(X_i \le x) = F(x)$$

Thus, for each x, $F_n(x)$ is itself an average of iid Ber$(F(x))$ r.v.s.

$\hookrightarrow X \sim \mathrm{Ber}(p)$: $\mathbb{E}X = p$; $\mathbb{V}X = \mathbb{E}X^2 - (\mathbb{E}X)^2$
$= \mathbb{E}X - (p)^2$
$= p - p^2$
$= p(1-p)$.

Therefore, by CLT:

$$\frac{F_n(x) - F(x)}{\sqrt{\frac{F(x)(1-F(x))}{n}}} \rightsquigarrow \mathcal{N}(0,1)$$

$$\therefore \underbrace{\sqrt{n}\,\big(F_n(x) - F(x)\big)}_{=:\, G_n(x)} \rightsquigarrow \mathcal{N}\big(0, F(x)(1-F(x))\big)$$

! At each point, the deviation between the empirical and theoretical CDF is (asymptotically) Gaussian

Donsker's Theorem

Q_1: Does $G_n(x)$ converge to a well-defined stochastic process over all of $\mathbb{R}$?

Q_2: If so, can we characterize the limiting process?

$$G_n(x) \rightsquigarrow B(F(x))$$

↳ Brownian Bridge process

II) Kolmogórov-Smirnóv Test

$$\begin{cases} H_0: & F = F^* \leftarrow F^* \text{ unknown, continuous CDF} \\ H_a: & H_0 \text{ is not true}; \quad F \neq F^* \end{cases}$$

Statistic $\quad D_n = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|$

↳ Built from $X_1, \dots, X_n \sim F$

$$= \frac{1}{\sqrt{n}} \sup_{x \in \mathbb{R}} |G_n(x)|$$

(By Glivenko-Cantelli, $D_n \xrightarrow{as} 0$ under H_0)

The Reject H_0 if D_n is large ?

Idea: Reject H_0 if D_n is large.

According to Donsker's Thm:

$$\sqrt{n}\, D_n = \sup_x \sqrt{n}\, |F_n(x) - F(x)|$$

$$\overset{\text{apply CMT}}{\rightsquigarrow} \sup_x |B(F(x))|$$

$$= \sup_{0 \le p \le 1} |B(p)| := K$$

Kolmogorov Distn

To sum up: $\sqrt{n}\, D_n \rightsquigarrow K$.

Thm $\Phi^K(\lambda) := \mathbb{P}\left(\sup_{0 \le p \le 1} |B(p)| \le \lambda \right)$

$$= 1 - 2 \sum_{k=1}^{\infty} (-1)^{k-1} \cdot e^{-2k^2\lambda^2}$$

(Can be proved using Brownian crossing time and Reflection principle)

* Note that limit $\sqrt{n}\, D_n \rightsquigarrow K$ does not

· Note that limit $\sqrt{n}\, D_n \rightsquigarrow K$ ← does not depend on unknown F.

Critical value : K_α s.t $\mathbb{P}(K < K_\alpha) = 1-\alpha$

↳ Reject H_0 if $\sqrt{n}\, D_n > K_\alpha$.

! Test can be extended to check if two samples come from the same distribution

$$D_{nm} := \sup_{x \in \mathbb{R}} \left| F_n(x) - \tilde{F}_m(x) \right|$$

Thm: $\sqrt{\frac{mn}{m+n}}\, D_{nm} \rightsquigarrow K$

! This is not necessarily the most practical test but useful to illustrate principles (and good excuse for some nice math)

IV) Gaussian Processes, Brownian Motion and Brownian Bridge

Recall a stochastic process $X : (X_t : t \in T)$

collection of r.vs $(\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

The FIDIs (finite-dimensional distributions) of X are the collections of multivariate distributions

$$(X_{t_1}, \dots, X_{t_k}) \quad \forall k \in \mathbb{N},\ t_1, \dots, t_k \in T$$

Q? Are the fidis of "$\lim_{n \to \infty} F_n$" multivariate Gaussian?

Not clear (before Donsker)

what we know is that the marginals

$$"X_{t_\ell}" := F_n(x_\ell) \rightsquigarrow \mathcal{N}\big(F(x_\ell), F(x_\ell)(1-F(x_\ell))\big)$$

But we don't know about the joint.

A Gaussian process is such that all its fidis are Multivariate Gaussian

> $Z \sim \mathcal{N}(\mu, \Sigma)$ is an n-dimensional multivariate Gaussian if $\forall v \in \mathbb{R}^n$, $v^T Z \sim \mathcal{N}(v^T \mu, v^T \Sigma v)$

? Determined by collection of expectations and covariance matrices of the fidis

$$\begin{cases} \mu_X(t) = \mathbb{E}\, X_t \\ C_X(t,s) = \mathbb{V}(X_t, X_s) \end{cases}$$

Brownian Motion

Louis Bachelier (1900; "Théorie de la spéculation")

⇝ Albert Einstein (yes, that Einstein)

⇝ Norbert Wiener (proved continuous, but w/w different)

? This is a "dual" definition / existence theorem

A stochastic process $W = (W_t : t \in [0, \infty))$

is known as a Brownian Motion or Wiener Process

- $W_0 = 0$ a.e. ⇝ $\mathbb{P}(\{\omega : W_0(\omega) \neq 0\}) = 0$

- Process has stationary, independent increments

$$\begin{cases} \forall\, 0 < t_1 < t_2 \dots < t_k : \quad W_{t_i} - W_{t_{i-1}} \perp\!\!\!\perp W_{t_j} - W_{t_{j-1}} \\ \quad\quad i \neq j \\ \forall\, t > 0, \quad W_t = W_t - W_0 \sim \mathcal{N}(0, t) \end{cases}$$

- Process has continuous sample paths:

  $\forall \omega \quad W_t(\omega)$ is continuous in t

! BM fidis are Multivariate Gaussian, thus BM is a Gaussian Process (GP) by def.

! Note W_{t-s} and $W_t - W_s \sim \mathcal{N}(0, t-s)$ for $s<t$

Exercise $\quad \mu_W(t) = \mathbb{E} W_t = 0$

$$C_W(t,s) = \mathbb{E} W_t W_s = \min(s,t)$$

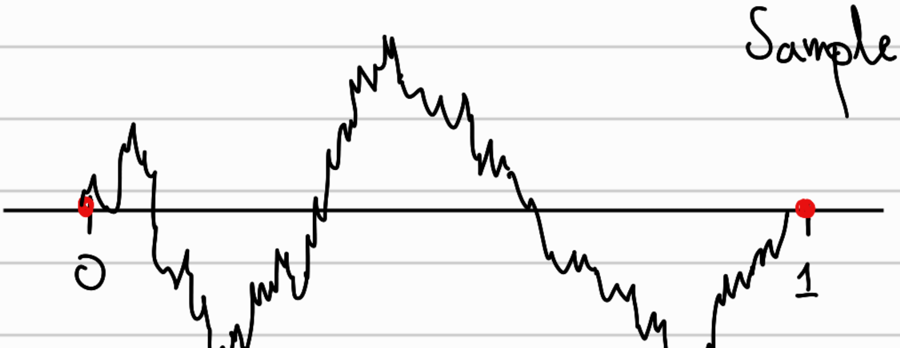
Brownian Bridge

Consider the process

$$B_t := W_t - t \cdot W_1 \qquad 0 \le t \le 1$$

! Note $\begin{cases} B_0 = W_0 - 0 \cdot W_1 = 0 \\ B_1 = W_1 - 1\, W_1 = 0 \end{cases}$

Sample path



Thm
$$\begin{cases} \mu_B(t) = ? \\ c_B(s,t) = ? \end{cases}$$

Proof $\mu_B(t) = \mu_W(t) - t\,\mu_W(1) = 0$

$$C_B(s,t) = \mathbb{V}(B_t, B_s) \qquad s<t$$

$$= \mathbb{E}\, B_t B_s - \cancelto{0}{\mathbb{E} B_t}\ \cancelto{0}{\mathbb{E} B_s}$$

$$= \mathbb{E}\,(W_t - t W_1)(W_s - s W_1)$$

$$= \mathbb{E}\, W_t W_s - t\,\mathbb{E}\, W_s W_1 - s\,\mathbb{E}\, W_t W_1 + st\,\mathbb{E}\, W_1^2$$

$C_W(s,t)=s$

$$= s - st - \cancel{st} + \cancel{st} = s(1-t)$$

In particular, $C_B(s,s) = \mathbb{V}(B_s) = s(1-s)$

Like variance of Ber(p)