- Probability is subjective for this framework. Therefore, we can define probability for non-repeatable experiment.

Review of Probability

- Let Ω be set collection and subset $\mathcal{F}$ is a σ-algebra if

✓ $\mathcal{F}$ contains Ω

✓ $\mathcal{F}$ is closed under union:
If A and $B \in \mathcal{F}$, then $A \cup B \in \mathcal{F}$

✓ $\mathcal{F}$ is closed under complement.
If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$

- Definition of measure

✓ consider $(\Omega, \mathcal{F})$
The function $\mu : \mathcal{F} \to \mathbb{R}_{\geq 0}$ is called a measure if $\mu(A) \geq 0 \ \forall A \in \mathcal{F}$

✓ Countably additive

If $\{A_n\}$ disjoint. then $\mu(\bigsqcup_n A_n)$

$$= \sum_n \mu(A_n)$$

✓ $\mu(\phi) = 0$

- Definition of probability space $(\Omega, \mathcal{F}, \mu)$

: Ω contains all the sets

: $\mathcal{F}$ a collection of subsets properly defined so everything inside is measurable

: μ = the measure

$(\Omega, \mathcal{F}, \mu)$ is a probability space if

$\mu(\Omega) = 1$

Example 1:

Let $\Omega \equiv \mathbb{R}$, $\mathcal{F}$ is the σ-algebra contains all open set.

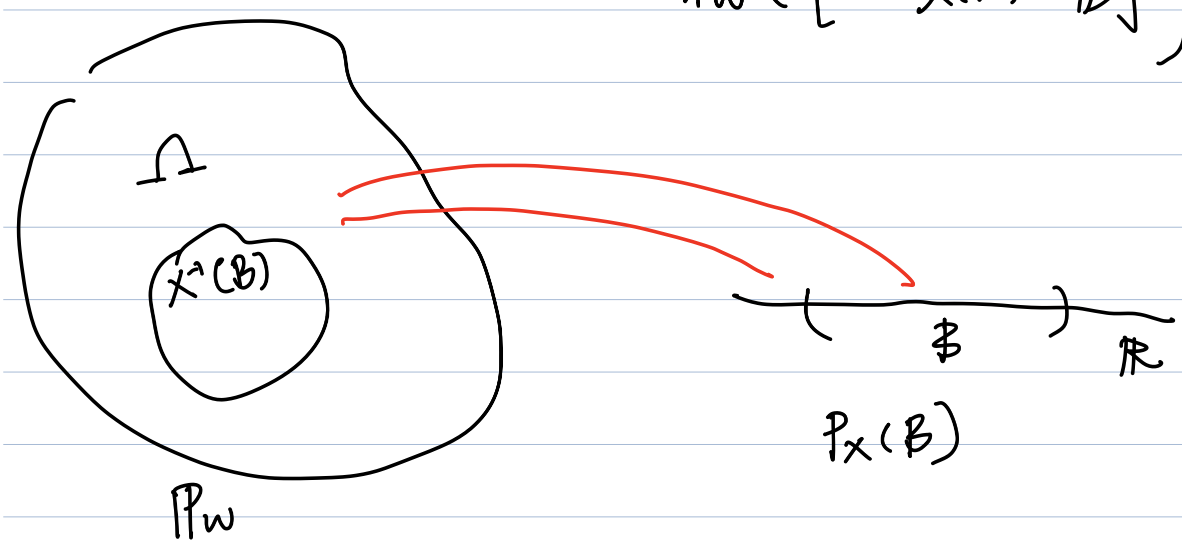
and there is a unique measure μ
$\mu((a,b]) = b-a$ for $a<b \in \mathbb{R}$.
This is called Lebesgue measure.

■ Random variable

function X maps Ω to $\mathbb{R}$ is a real-valued random variable if for all Borel sets $\mathbb{B} \in \mathbb{R}$
$X^{-1}(B) = \{w : X(w) \in \mathbb{B}\}$ is in $\mathcal{F}$.

If X is a random variable, it induces a probability measure $\mathbb{P}_X(B) = \mathbb{P}_X(X \in B)$

$$= \mathbb{P}_W(\{w : X(w) \in \mathbb{B}\})$$



and we want to find this random variable because working with real line is easier, allowing us to do addition / multiplication

and anyway the space Ω has the mapping with $\mathbb{R}$

■ for $x \in \mathbb{R}$

$\mathbb{P}_X[(-\infty, x)] = F_X(x)$

is the cumulative distribution function

that you're thinking from any $(-\infty, x)$ look at the pre-image.

properties

✓ $\mathbb{P}_X(\mathbb{R}) = 1$ because

$$\mathbb{P}_w(\{w, X(w) \in \mathbb{R}\}) = \mathbb{P}_w(\Omega) = 1$$

or can say $\lim_{x \to \infty} F(x) = 1$

$\lim F(x) = 0$

$x \to -\infty$

✓ Non-decreasing
that is,
If $x < x'$, then $F_X(x) \leq F_X(x')$

✓ Right-continuous F_X

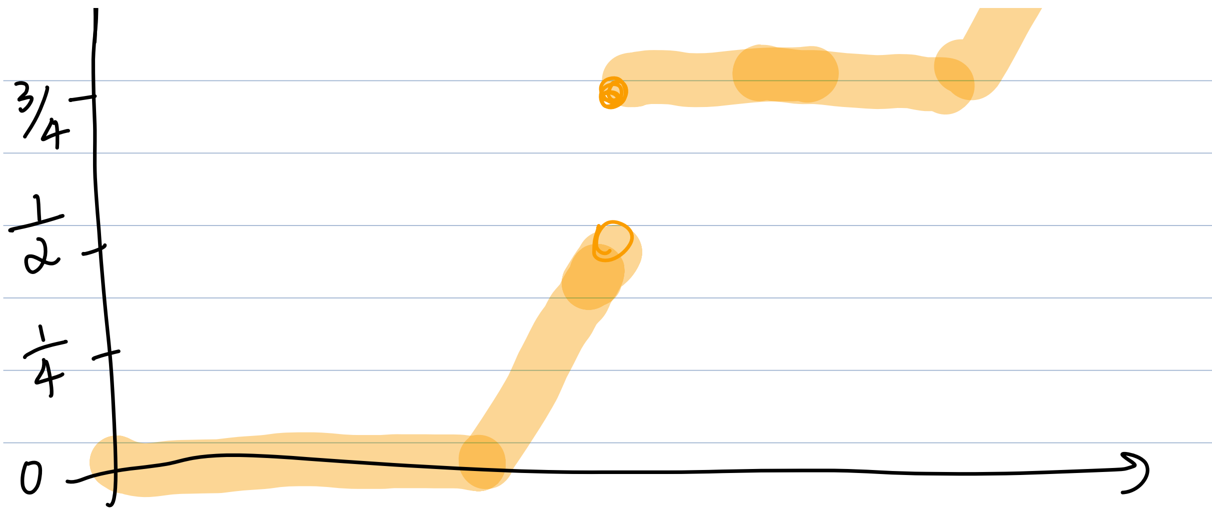
$$\lim_{y \downarrow x} F(y) = F(x)$$

Question:
If you have a CDF, can you construct a random variable accordingly? yes.

theorem:

Consider $\Omega = (0,1]$, and you have the following graph as F_X.

W
17



X ↓ you can construct the cdf by defining

$$X(\omega) = \sup\{y : F(y) < \omega\}$$

↓ that is you pick random point on F, and see where the greatest x is.

Example.

$\Omega = (0,1]$ $\mathcal{F}$ Borel sets. $\mathbb{P}_\omega$ Lebesgue uniform distribution $\mathbb{P}_a((0,t]) = t$ for all $t \in [0,1]$

then

$$\{\omega : \omega \le F(x)\} = \{\omega : X(\omega) \le x\}$$

$$\mathbb{P}_X(X \le x) = \mathbb{P}_\omega(\{\omega : \omega \le F(x)\}) = F(x)$$

theorem:

Suppose $F: \mathbb{R} \to [0,1]$

is 1) non-decreasing

2) right-continuous

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

Then exists a random variable X with CDF F.

Density is constructed upon F_X, CDF.

$F(x) = \int_{-\infty}^{x} f(t)\,dt \quad f \geq 0.$

then we call f is the density function for a random variable.

Theorem:

Let $\mathcal{X}$ be some space

If $f: \mathcal{X} \to \mathbb{R}_{\geq 0}$ such that $\int_{\mathcal{X}} f(x)\,dx < \infty$

then $\exists$ an $\mathcal{X}$-valued random variable with density $\propto f$.

Goal: Construct probability distributions over some set $\mathcal{X}$.

How?

Step 1: Find an function f such that $f: \mathcal{X} \to \mathbb{R}_{\geq 0}$

and $\int_{\mathcal{X}} f(x)\, dx < \infty$

Step 2: Normalization

$$\tilde{f}(t) = \frac{f(t)}{\int_{\mathcal{X}} f(t)\, dt},$$

then $\tilde{f}(t)$ is a valid density.

Example:

$\mathcal{X} = \{0, 1\}$

Fix $p \in [0,1]$

$f(1) = p$

$f(0) = 1 - p$

and I claim this is a valid pdf

Because

$\int_{\mathcal{X}} f(x)\, dx = 1$

and $f(x)$ for $x \in \mathcal{X}$ nonnegative

Example 2

Consider $\mathcal{X}$ $\{0, 1, 2, 3, \ldots\}$

For $\sum_{n\geq 0}^{\infty} \frac{\lambda^n}{n!} = e^{\lambda} < \infty$

then we claim $f(n) = \frac{\lambda^n}{n!}$ can

be a construction base for a density because

$$\tilde{f}(n) = \sum_{n=0}^{\infty} f(n) \cdot e^{-\lambda} = 1$$

and $\tilde{f}(n)$ nonnegative.

therefore as long as f is finite

you can turn this into a density

By normalization.

Example 3.

$\mathcal{X} = [a,b]$. $a<b \in \mathbb{R}$

$f(x) = 1$ for $x \in [a,b]$

then we can define

$$\tilde{f}(x) \equiv \frac{f(x)}{b-a}$$

$$f(x) \equiv \frac{1(a \leq x \leq b)}{b-a}$$ becomes the uniform (a,b)

Example 4.

$\mathcal{X} = (0, \infty)$

Fix $\lambda > 0$

$$\int_0^\infty e^{-\lambda t}\, dt$$

$$= \frac{-1}{\lambda} e^{-\lambda t} \Big|_0^\infty = \frac{1}{\lambda}$$

$e^{-\lambda t}$ is always nonnegative,

then normalizing it

to find that

$f(x) = \lambda e^{-\lambda x}$ for $x > 0$

is a density.

and this is exponential (λ)

Example 5.

Fix $x > 0$

gamma function

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} \, dt < \infty$$

property

$$\Gamma(x+1) = x\Gamma(x)$$

Γ : continuous factorial.

Because $\Gamma(x)$ is finite so
we can use it to define density

consider,

$r, s > 0$

$$\int_0^\infty t^{r-1} e^{-st} \, dt$$

change of variables: $u = st$

$du = s\,dt$

thus, the above becomes

$$\int_0^\infty \left(\frac{u}{s}\right)^{r-1} e^{-u} \frac{du}{s} = \frac{1}{s^r} \int_0^\infty u^{r-1} e^{-v} \, dv$$

$$= \frac{\Gamma(r)}{s^r}$$

then you find the normalization constant!

then you could define

density Gamma (r, s)

$$f(x) = \frac{s^r}{\Gamma(r)} x^{r-1} e^{-sx}$$

Example 6.

Beta funvtion

$$B(a,b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt \quad a, b > 0$$

and property: $B(a,b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

thus you could define a Beta density,

$$f(x) = \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1}.$$

Example 7.

$$\int_{-\infty}^{\infty} e^{-x^2/2}\, dx = \sqrt{2\pi},$$

therefore, you can develop density from this.

$N(0,1)$ density $(2\pi)^{-1/2} e^{-x^2/2}$

Let X be a random variable with density $f(x)$

Let $h: \mathcal{X} \to \mathbb{R}$.

define $E[h(X)] = \int_{\mathcal{X}} h(x) f(x)\, dx$.

If $f(x)$ exists.

Takeaway messages:

- σ-algebra is being constructed in a way that every set satisfies properties to be easy to measure.

- Measure is constructed in a way on measurable set.

- Random variable is a map from set to real line, so that measure is easy to manipulate, because real-lined values can be mapped back to set.

- cumulative density makes the use of notion that random variable talks about pre-image.

- Every CDF can be traced back to a random variable, thus a density, if well-defined

* Every density can be traced back to a random variable, if well-defined

therefore, the above two theorems make the construction of this triple

(CDF, PDF, random variable) easy, under specific condition because,

if you find a function, then you can find

$pdf \to CDF$
$\downarrow X$

if you find a cumulative, you can find

$CDF \to pdf$
$\downarrow X$

if you find a random variable, you can find

$X \to pdf$
$\downarrow CDF$ By definition

* If you have $X_1 \times X_2 \cdots \times X_n$ product space, and you try to define measure on it.

By following recipe

1.
Find a function

$$f: X_1 \times X_2 \cdots X_n \Rightarrow R_{\geq 0}$$

2. check $$\int_{X_1 \cdot X_2 \cdots X_n} f \, dx_1 \cdots dx_n < \infty$$

Namely: Non-negative and finite integral then normalizing it just becomes a density

$f(x_1 \dots x_n)$ is a multivariate density.

$(X_1 \dots X_n)$ is a random vector, and

Each X_i is a random variable with

marginal density

$$f(x_i) = \int f(x_1 \dots x_n)\, dX_{-i}$$

$$\equiv \int f(x_1 \dots x_n)\, dx_1 dx_2 \dots dx_{i-1}, dx_{i+1} \dots dx_n$$

Define Expectation over this random vector $X = (X_1 \dots X_n)$

Being

$$\mathbb{E}(X) = (E(X_1) \dots E(X_n))$$

Covariance Matrix

$$\mathrm{cov}(X) = \mathbb{E}\left[(X - E(X)) (X - E(X))^T \right]$$

$$= \left(\mathrm{cov}(X_i, X_j) \right)_{i,j=1}^{n}$$

which is $n \times n$ matrix

Let (X, Y) be a $\mathcal{X} \cdot \mathcal{Y}$ random variables with density f_{XY}.

$$\int_{\mathcal{X} \times \mathcal{Y}} f_{XY}(x, y)\, dx\, dy = 1.$$

and we would be interested in the relationship of the two.

We define:

definition:

Suppose $\forall A \in \mathcal{X}$ and $\forall B \in \mathcal{Y}$, that

$$\mathbb{P}(X \in A, Y \in B] = P(X \in A) \cdot P(Y \in B)$$

We then say X and Y are independent.

and

$$P(X \in A, Y \in B) = \int_A \int_B f_{XY}(x, y)\, dy\, dx$$

and

$$P(X \in A) = \int_A \int_y f(x,y)\,dy\,dx$$

But from the above definition, we need to check every subsets?

So here comes another definition.

X and Y are independent if and only if

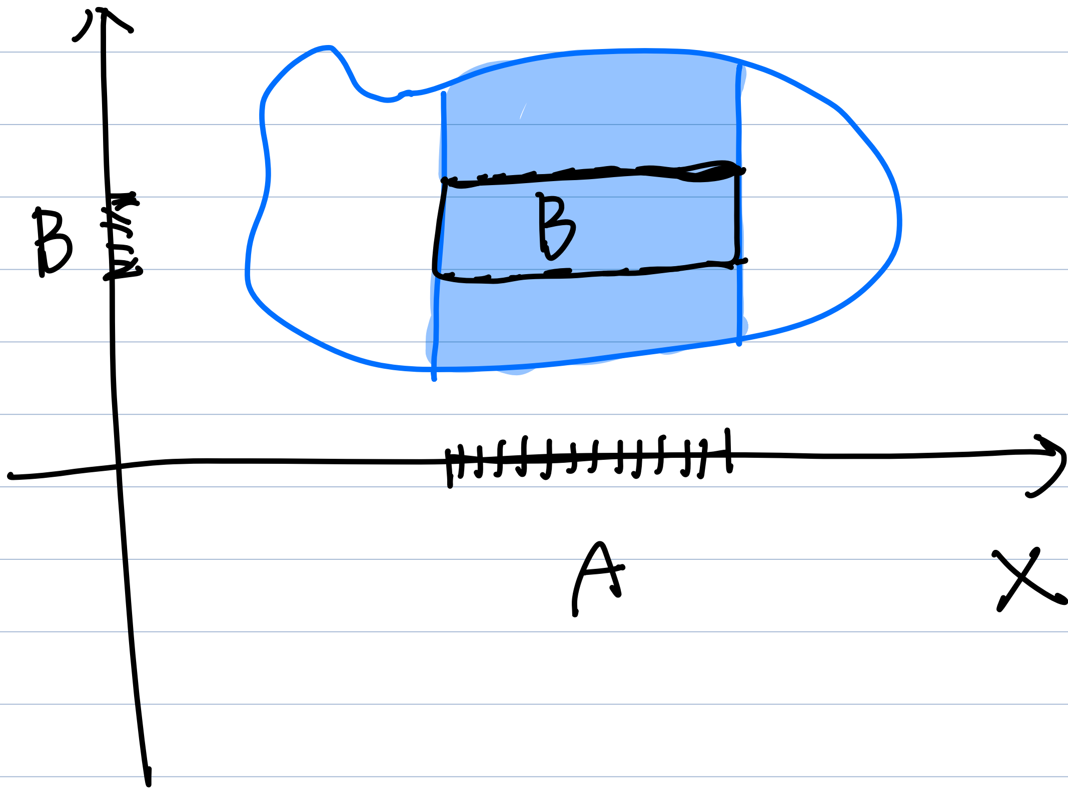
$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$$

where

$$f_X(x) = \int_y f_{XY}(x,y)\,dy$$

$$f_Y(y) = \int_x f_{XY}(x,y)\,dx$$

conditional probability: define a new measure on a subset



define conditional probability by slicing

that

$$p(y \in B \mid X \in A) = \frac{p(X \in A, y \in B)}{p(X \in A)}$$

definition:

For all $x \in \mathcal{X}$, we can define conditional density

$$f(y \mid X=x) = \frac{f_{xy}(x,y)}{f_x(x)}.$$

if you shrink your X into just a slice.

For all $y \in \mathcal{Y}$,

$$f(X \mid Y=y) = \frac{f_{x,y}(x, Y=y)}{f_y(y)}$$

and simply by manipulating definitions, Reaching Bayes theorem

$$f(x|y) \cdot f(y) = f(x,y) = f(y|x) \cdot f(x)$$

then $\quad \ldots = f(y|x) f(x)$

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

To elaborate to more than two variables

$$f(x_1, x_2, x_3)$$

$$\Rightarrow$$

$$f(x_1, x_2) = \int f(x_1, x_2, x_3)\, dx_3$$

$$f(x_1) = \int f(x_1, x_2)\, dx_2$$

$$= \iint f(x_1, x_2, x_3)\, dx_3\, dx_2$$

then

$$f(x_1 \mid x_2, x_3) = \frac{f(x_1, x_2, x_3)}{f(x_2, x_3)}$$

thus

$$f(x_1 \ldots x_n)$$

$$= f(x_2 \dots x_n \mid x_1) \cdot f(x_1)$$

$$= f(x_3 \dots x_n \mid x_1, x_2)\, f(x_2 \mid x_1) \cdot f(x_1)$$

$$= f(x_4 \dots x_n \mid x_1, x_2, x_3)\, f(x_3 \mid x_1, x_2) \cdot f(x_2 \mid x_1) \cdot f(x_1)$$

$\vdots$

$$= f(x_n \mid x_1 \dots x_{n-1}) \cdot f(x_{n-1} \mid x_1 \dots x_{n-2})$$

$\vdots$

$$f(x_2 \mid x_1) \cdot f(x_1)$$

$\downarrow$

the joint density means the joint information is by adding the first information, and the second information depend on the first ... the final piece from previous ones.

this is the basis of ancestral sampling.

Monte-Carlo Simulation

Example:

$x \sim N(0,1)$. then

$$f(x) = (2\pi)^{-1/2} e^{-x^2/2}$$

$$\mathbb{E}X = \int_{-\infty}^{\infty} x f(x) dx$$

$$= \int_{-\infty}^{\infty} \frac{x e^{-x^2/2}}{\sqrt{2\pi}} dx = 0$$

change of variable
Recognizing x odd funtion and symmetry argument

$$E(X^2) \int \frac{x^2 e^{-x^2/2}}{\sqrt{2\pi}} dx = 1$$

By integration by parts

But our ability to compute integral is limited

for example:

$Z \sim Beta(a, b)$

then

$$E(Z) = \int_0^1 z \frac{1}{B(a,b)} z^{a-1} (1-z)^{b-1} dz$$

$$= \frac{1}{B(a,b)} \int_0^1 z^{(a+1)-1} (1-z)^{b-1} \, dz$$

you do not integrate but identifying

$\int_0^1 z^{(a+1)-1} (1-z)^{b-1} \, dz$ is a kernel.

How about

$E\left[\log\left(\frac{z}{1-z}\right)\right]$?

General idea:

Input → 

assume there exists some mechanism

that the output follows distribution p

so the proportion of output in $A \approx P(A)$

Namely, I can always create a machine and this machine can always generate numbers from a speified distribution.

Let $\mathbb{P}$ be a distribution and suppose $X_1 \cdots X_N$ are independent samples from $\mathbb{P}$.

then to know

$$E[h(x)] \approx \frac{1}{N}\sum_{i=1}^{N} h(X_i)$$, similar to law of large number

so.
get X_1,
compute $h(X_1)$
store it

get x_2
compute $h(x_2)$
store it ...
and finally compute $\frac{1}{N}\sum_{i=1}^{N} h(x_i)$

and we can get arbitrary distribution by inverse CDF method, and generate uniform distribution ...
some difficulties?

How to generate any distribution?