Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space

Random variable $X(\omega) : (\Omega, \mathcal{F}) \to (\mathbb{X}, \mathcal{X})$

X is Measurable (X is $(\Omega, \mathcal{F})$ to $(\mathbb{X}, \mathcal{X})$ measurable)

if for any $B \in \mathcal{X}$

$$X^{-1}(B) = \{\omega : X(\omega) \in B\}$$
$$\in \mathcal{F}$$

Define probability measure $\mathbb{P}_X$ on $(\mathbb{X}, \mathcal{X})$

as $\mathbb{P}_X(B) = \mathbb{P}\{X(\omega) \in B\} = \mathbb{P}\, X^{-1}(B)$

$\mathbb{P}_X$ is the distribution of X

pushforward measure

<u>Expectation</u> / mean / average

for each $\mathbb{P}$ on $(\Omega, \mathcal{F})$ there is a map on the space of random variables $\mathbb{E}$

$\mathbb{1}\{\omega \in A\} = 1$ if $\omega \in A$
0 if $\omega \notin A$

Properties / axioms

- $\mathbb{E}\, \mathbb{1}\{\omega \in A\} = \mathbb{P} A \quad \forall A \in \mathcal{F}$

- $\mathbb{E}\, 0 = 0$
- Linearity $\mathbb{E}[aX + bY] = a\mathbb{E}X + b\mathbb{E}Y$

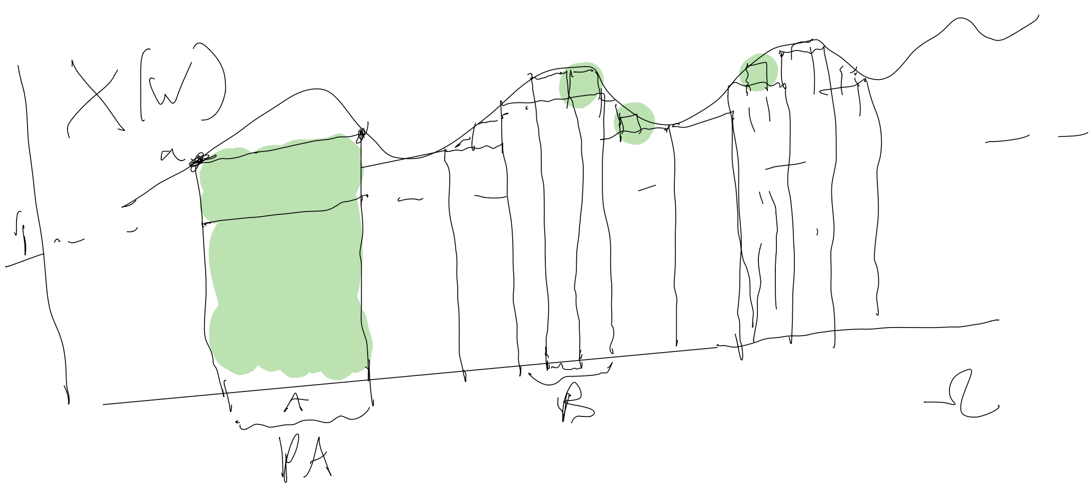
  for any $a, b \in \mathbb{R}$, X, Y random variables
- (monotone convergence). For a sequence $X_1(\omega)$, $X_2(\omega)$, ... such that

  $0 \le X_1(\omega) \le X_2(\omega) \le \dots \uparrow X(\omega)$

  for each $\omega \in \Omega$

  then $\lim_{i \to \infty} \mathbb{E} X_i(\omega) \to \mathbb{E} X(\omega)$

  $\lim \mathbb{E} X_i = \mathbb{E} \lim X_i$



- Take two measures μ and ν

  we say $\mu << \nu$ μ "is dominated by" ν

  or μ "is absolutely continuous" w.r.t ν

if $\nu A = 0$ implies $\mu A = 0$

- A measure μ on $(\Omega, \mathcal{F})$ is σ-finite if $\Omega = \bigcup_{i=1}^{\infty} A_i$ such that $\mu A_i < \infty$ for all i
- Radon Nikodym theorem
  Let μ, ν be σ-finite
  Let $\mu << \nu$.
  Then there is a function $\frac{d\mu}{d\nu} \geq 0$
  such that $\mathbb{E}_{\mu} X = \mathbb{E}_{\nu} X \frac{d\mu}{d\nu}$

  for any random variable X

- $\frac{d\mu}{d\nu} \geq 0$ and $\mathbb{E}_{\nu} \frac{d\mu}{d\nu} = 1$

Call $\frac{d\mu}{d\nu}$ the density of μ with respect to ν

Radon Nikodym derivative

Consider case $\Omega = \mathbb{R}$, can use the uniform measure
also called the Lebesgue measure μ
satisfies $\mu \mathbb{1}\{a < \omega < b\} = b - a$
for any $a, b \in \mathbb{R}$
Defined over a Σ-field containing all such intervals
contains
$(-\infty, a] \cup [b, \infty)$, $[b, c)$, $[b, c]$, $\{b\}$, $\{a\}$

Can be defined on Borel sigma field, smallest Σ-field containing open sets, or sometimes slightly larger one

$$M\{(a,b)\cup(b,c)\} = c-b+b-a = c-a$$

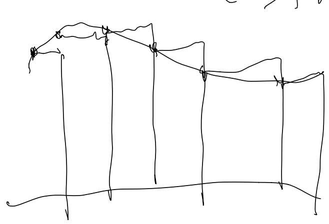
$$M\{(a,c)\} = c-a$$

$$(a,c)/[(a,b)\cup(b,c)] = \{b\}$$

$$M\{b\} = 0$$

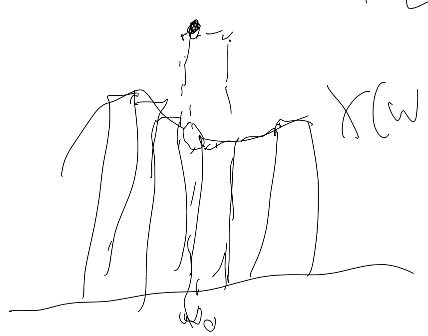
Any countable union of singletons has Lebesgue measure 0

$$M\,X(w) = \int X(w)\,dw$$ where the integral is the standard Riemann integral



for any function that is both Riemann integrable and Lebesgue integrable

Consider



$M(\{w_0\}) = 0$

A continuously distributed random variable X has a density $f_X(w)$ with respect to Lebesgue measure

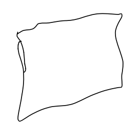
We say $X \sim f_X(\,)$ if f_X is density of X

"is distributed as" $\sim$

Call $f_X(\,)$ the probability density function or pdf of X

$$\mathbb{E}X = \mathbb{E}_\mu X f_X = \int w \, f_X(w)\, dw$$

Multivariate Lebesgue measure on $\mathbb{R}^N$

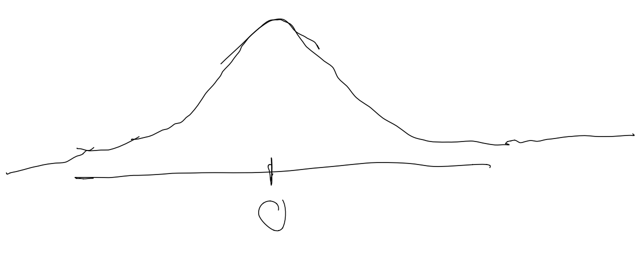


$$\mu \; (a,b)\times(c,d)\times\ldots = (b-a)\times(d-c)\times\ldots$$

$$\mathbb{E}X(w) = \int\ldots\int X(w)\, dw_1\, dw_2 \ldots dw_n$$

Example continuous densities

- Uniform on $[a,b]$, $f_X(w) = \frac{1}{b-a}\mathbb{1}\{a \le w \le b\}$
- (standard) Gaussian on $\mathbb{R}^N$: $f_X(w) = (2\pi)^{-\frac{N}{2}} e^{\frac{w'w}{2}}$



Expectation measures the center of a distribution

Expectations of functions can describe other properties

When it exists, the variance of a random variable X

is $\mathbb{E}[(X(\omega) - \mathbb{E}X(\omega))^2] = \mathbb{E}(X(\omega))^2 - (\mathbb{E}X(\omega))^2$

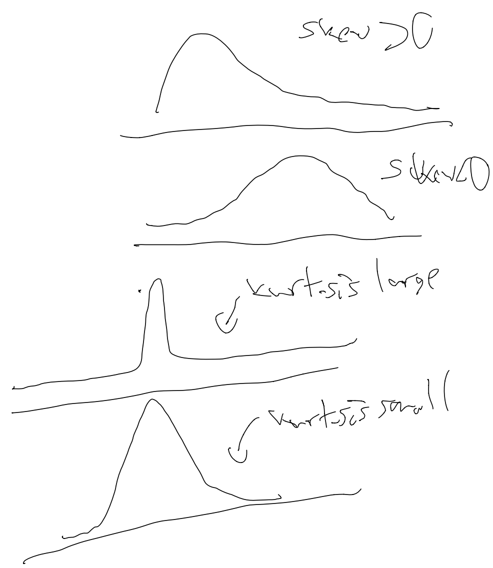
measures dispersion

The k-th (central) moment of X

$$\mathbb{E}[(X(\omega) - \mathbb{E}X(\omega))^k]$$

$k = 3$: skewness

$k = 4$: kurtosis



Another useful measure: Counting measure

μA = # of elements in A

defined on Ω finite or countable

"Discrete random variable" X has density $p_X(\omega)$ w.r.t. counting measure. call $p_X(\omega)$ the "probability distribution" or probability mass function (p.m.f.)

$$\underline{\mathbb{E}X(w)} = E_w[X(w)p_X(w)] = \sum_w X(w) \cdot P_X(w)$$

For $X \in \mathbb{R}^N$ r.v. with distribution $\mathbb{P}_X$

The sets $\{ \cancel{X(w)} \leq x \}$ for $x \in \mathbb{R}^N$ fully determine distribution (if range is Borel σ-field)

$$\mathbb{P}_X \{ X(w) \leq x \} = F(x) \quad ; \mathbb{R}^N \to [0,1]$$

Cumulative distribution function (CDF)

If F is differentiable $\frac{d}{dx}F = f(x)$ is the density w.r.t. Lebesgue measure

If $x_1 > x_2$ then $F(x_1) \geq F(x_2)$

Let $X_1(w)$, $X_2(w)$ defined on $(\Omega, \mathcal{F})$

The distribution of (X_1, X_2) is the "joint distribution"

$\mathbb{P}_{X_1, X_2}$ The distribution of X_1, X_2 individually with marginal distributions.

If (X_1, X_2) has density $f_{X_1 X_2}(x_1, x_2)$ wrt Lebesgue measure

Then
$$f_{X_1}(x_1) = \int f_{(X_1, X_2)}(x_1, x_2)\,dx_2$$
$$f_{X_2}(x_2) = \int f_{(X_1, X_2)}(x_1, x_2)\,dx_1$$

Fubini's theorem $\iint F(x_1, x_2)\,dx_1\,dx_2 = \iint F(x_1, x_2)\,dx_2\,dx_1$ for bounded F

(X_1, X_2) are independent if

$$\mathbb{E} f(X_1) g(X_2) = \mathbb{E}_{X_1} f(X_1)\, \mathbb{E}_{X_2} g(X_2)$$

for any f, g

If densities exist, independence is equivalent to

$$f_{X_1,X_2}(X_1, X_2) = f_{X_1}(X_1) f_{X_2}(X_2)$$

Extends to > 2 variables

Says X_1 "provides no information" about X_2 or vice versa

Let (X_1, X_2) be a vector-valued random variable with a density $f_{X_1,X_2}(\cdot,\cdot)$

The conditional density of X_1 given X_2 is

$$f_{X_1|X_2}(X_1|X_2) = \frac{f_{X_1,X_2}(X_1, X_2)}{f_{X_2}(X_2)}$$ when this exists.

Gives a conditional expectation

$$\mathbb{E}[g(X_1) | X_2] = \int g(x_1) f_{X_1|X_2}(x_1 | X_2)\, dx_1$$

This is a function of X_2 and so is a random variable

Law of iterated expectations. (L.I.E.)

$$\mathbb{E}_{X_2}\left(\mathbb{E}[g(X_1) | X_2]\right) = \mathbb{E}\, g(X_1)$$

for all g

Further it we have $f(X_2)g(X_1)$

$$\mathbb{E}_{X_2}\left(\mathbb{E}\left[f(X_2)g(X_1) \mid X_2\right]\right) = \mathbb{E}_{X_2}\left(f(X_2)\mathbb{E}\left[g(X_1) \mid X_2\right]\right)$$

$$= \mathbb{E} f(X_2)\, g(X_1)$$

If X_1, X_2 are independent, $\mathbb{E}\left[g(X_1) \mid X_2\right]$

$$= \mathbb{E}\left[g(X_1)\right]$$

"almost surely"

meaning $\mathbb{P}\left(\mathbb{E}\left[g(X_1) \mid X_2\right] = \mathbb{E}\left[g(X_1)\right]\right) = 1$