Probability Theory

"A random variable is neither random nor variable."

Gian-Carlo Rota, M.I.T..

Florian Herzog

2013

Probability space

Probability space

A probability space W is a unique triple $W = \{\Omega, \mathcal{F}, P\}$:

- Ω is its sample space
- $\mathcal{F}$ its σ-algebra of events
- P its probability measure

Remarks: (1) The sample space Ω is the set of all possible samples or elementary events ω: $\Omega = \{\omega \mid \omega \in \Omega\}$.
(2)The σ-algebra $\mathcal{F}$ is the set of all of the considered events A, i.e., subsets of Ω: $\mathcal{F} = \{A \mid A \subseteq \Omega, A \in \mathcal{F}\}$.
(3) The probability measure P assigns a probability $P(A)$ to every event $A \in \mathcal{F}$: $P : \mathcal{F} \rightarrow [0, 1]$.

Sample space

The sample space Ω is sometimes called the *universe* of all samples or possible outcomes ω.

Example 1. Sample space

- *Toss of a coin (with head and tail):* $\Omega = \{H, T\}$.
- *Two tosses of a coin:* $\Omega = \{HH, HT, TH, TT\}$.
- *A cubic die:* $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$.
- *The positive integers:* $\Omega = \{1, 2, 3, \dots\}$.
- *The reals:* $\Omega = \{\omega \mid \omega \in \mathbb{R}\}$.

Note that the ωs are a mathematical construct and have per se no real or scientific meaning. The ωs in the die example refer to the numbers of dots observed when the die is thrown.

Event

An event A is a subset of Ω. If the outcome ω of the experiment is in the subset A, then the event A is said to have occurred. The set of all subsets of the sample space are denoted by 2^Ω.

Example 2. Events

- *Head in the coin toss:* $A = \{H\}$.
- *Odd number in the roll of a die:* $A = \{\omega_1, \omega_3, \omega_5\}$.
- *An integer smaller than 5:* $A = \{1, 2, 3, 4\}$, *where* $\Omega = \{1, 2, 3, \dots\}$.
- *A real number between 0 and 1:* $A = [0, 1]$, *where* $\Omega = \{\omega \mid \omega \in \mathbb{R}\}$.

We denote the complementary event of A by $A^c = \Omega \backslash A$. When it is possible to determine whether an event A has occurred or not, we must also be able to determine whether A^c has occurred or not.

Probability Measure I

Definition 1. Probability measure
A probability measure P on the countable sample space Ω *is a set function*

$$P : \mathcal{F} \to [0, 1],$$

satisfying the following conditions

- $P(\Omega) = 1.$
- $P(\omega_i) = p_i.$
- *If* $A_1, A_2, A_3, \ldots \in \mathcal{F}$ *are mutually disjoint, then*

$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} P(A_i).$$

Probability

The story so far:

- Sample space: $\Omega = \{\omega_1, \ldots, \omega_n\}$, finite!
- Events: $\mathcal{F} = 2^\Omega$: All subsets of Ω
- Probability: $P(\omega_i) = p_i \quad \Rightarrow \quad P(A \in \Omega) = \sum_{w_i \in A} p_i$

Probability axioms of Kolmogorov (1931) for elementary probability:

- $\mathrm{P}(\Omega) = 1.$
- If $A \in \Omega$ then $P(A) \geq 0.$
- If $A_1, A_2, A_3, \ldots \in \Omega$ are mutually disjoint, then

$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} P(A_i).$$

Uncountable sample spaces

Most important uncountable sample space for engineering: $\mathbb{R}$, resp. $\mathbb{R}^n$.

Consider the example $\Omega = [0, 1]$, every ω is equally "likely".

- Obviously, $\mathsf{P}(\omega) = 0$.
- Intuitively, $\mathsf{P}([0, a]) = a$, basic concept: **length**!

Question: Has every subset of $[0, 1]$ a determinable length?
Answer: No! (e.g. Vitali sets, Banach-Tarski paradox)

Question: Is this of importance in practice?
Answer: No!

Question: Does it matter for the underlying theory?
Answer: A lot!

Fundamental mathematical tools

Not every subset of $[0, 1]$ has a determinable length $\Rightarrow$ collect the ones with a determinable length in $\mathcal{F}$. Such a mathematical construct, which has additional, desirable properties, is called σ-algebra.

Definition 2. σ-algebra

A collection $\mathcal{F}$ of subsets of Ω is called a σ-algebra on Ω if

- $\Omega \in \mathcal{F}$ *and* $\emptyset \in \mathcal{F}$ *($\emptyset$ denotes the empty set)*
- *If* $A \in \mathcal{F}$ *then* $\Omega \backslash A = A^c \in \mathcal{F}$*: The complementary subset of A is also in* Ω
- *For all* $A_i \in \mathcal{F}$*:* $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

The intuition behind it: collect all events in the σ-algebra $\mathcal{F}$, make sure that by performing countably many elementary set operation $(\cup, \cap, ^c)$ on elements of $\mathcal{F}$ yields again an element in $\mathcal{F}$ (closeness).

The pair $\{\Omega, \mathcal{F}\}$ is called $measure\ space$.

Example of σ-algebra

Example 3. σ-algebra of two coin-tosses

- $\Omega = \{HH, HT, TH, TT\} = \{\omega_1, \omega_2, \omega_3, \omega_4\}$
- $\mathcal{F}_{min} = \{\emptyset, \Omega\} = \{\emptyset, \{\omega_1, \omega_2, \omega_3, \omega_4\}\}$.
- $\mathcal{F}_1 = \{\emptyset, \{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}, \{\omega_1, \omega_2, \omega_3, \omega_4\}\}$.
- $\mathcal{F}_{max} = \{\emptyset, \{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_4\}, \{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}, \{\omega_1, \omega_4\}, \{\omega_2, \omega_3\},$
  $\{\omega_2, \omega_4\}, \{\omega_3, \omega_4\}, \{\omega_1, \omega_2, \omega_3\}, \{\omega_1, \omega_2, \omega_4\}, \{\omega_1, \omega_3, \omega_4\},$
  $\{\omega_2, \omega_3, \omega_4\}, \Omega\}$.

Generated σ-algebras

The concept of generated σ-algebras is important in probability theory.

Definition 3. $\sigma(\mathcal{C})$: σ-algebra generated by a class $\mathcal{C}$ of subsets
Let $\mathcal{C}$ be a class of subsets of Ω. The σ-algebra generated by $\mathcal{C}$, denoted by $\sigma(\mathcal{C})$, is the smallest σ-algebra $\mathcal{F}$ which includes all elements of $\mathcal{C}$, i.e., $\mathcal{C} \in \mathcal{F}$.

Identify the different events we can measure of an experiment (denoted by $\mathcal{A}$), we then just work with the σ-algebra generated by $\mathcal{A}$ and have avoided all the measure theoretic technicalities.

Borel σ-algebra

The Borel σ-algebra includes all subsets of $\mathbb{R}$ which are of interest in practical applications (scientific or engineering).

Definition 4. Borel σ-algebra $\mathcal{B}(\mathbb{R})$)

The Borel σ-algebra $\mathcal{B}(\mathbb{R})$ is the smallest σ-algebra containing all open intervals in $\mathbb{R}$. The sets in $\mathcal{B}(\mathbb{R})$ are called Borel sets. The extension to the multi-dimensional case, $\mathcal{B}(\mathbb{R}^n)$, is straightforward.

- $(-\infty, a), \quad (b, \infty), \quad (-\infty, a) \cup (b, \infty)$
- $[a, b] = \overline{(-\infty, a) \cup (b, \infty)}$,
- $(-\infty, a] = \bigcup_{n=1}^{\infty}[a - n, a]$ and $[b, \infty) = \bigcup_{n=1}^{\infty}[b, b + n]$,
- $(a, b] = (-\infty, b] \cap (a, \infty)$,
- $\{a\} = \bigcap_{n=1}^{\infty}(a - \frac{1}{n}, a + \frac{1}{n})$,
- $\{a_1, \cdots, a_n\} = \bigcup_{k=1}^{n} a_k$.

Measure

Definition 5. Measure
Let $\mathcal{F}$ be the σ-algebra of Ω and therefore $(\Omega, \mathcal{F})$ be a measurable space. The map

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

is called a measure on $(\Omega, \mathcal{F})$ if μ is countably additive. The measure μ is countably additive (or σ-additive) if $\mu(\emptyset) = 0$ and for every sequence of disjoint sets $(F_i : i \in \mathbb{N})$ in $\mathcal{F}$ with $F = \bigcup_{i \in \mathbb{N}} F_i$ we have

$$\mu(F) = \sum_{i \in \mathbb{N}} \mu(F_i).$$

If μ is countably additive, it is also additive, meaning for every $F, G \in \mathcal{F}$ we have

$$\mu(F \cup G) = \mu(F) + \mu(G) \quad \text{if and only if} \quad F \cap G = \emptyset$$

The triple $(\Omega, \mathcal{F}, \mu)$ is called a *measure space.*

Lebesgue Measure

The measure of length on the straight line is known as the Lebesgue measure.

Definition 6. Lebesgue measure on $\mathcal{B}(\mathbb{R})$

The Lebesgue measure on $\mathcal{B}(\mathbb{R})$, denoted by λ, is defined as the measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ which assigns the measure of each interval to be its length.

Examples:

- Lebesgue measure of one point: $\lambda(\{a\}) = 0$.
- Lebesgue measure of countably many points: $\lambda(A) = \sum_{i=1}^{\infty} \lambda(\{a_i\}) = 0$.
- The Lebesgue measure of a set containing uncountably many points:
  - zero
  - positive and finite
  - infinite

6

Probability Measure

Definition 7. Probability measure
A probability measure P on the sample space Ω with σ-algebra $\mathcal{F}$ is a set function

$$P : \mathcal{F} \to [0, 1],$$

satisfying the following conditions

- $P(\Omega) = 1$.
- *If $A \in \mathcal{F}$ then $P(A) \geq 0$.*
- *If $A_1, A_2, A_3, \ldots \in \mathcal{F}$ are mutually disjoint, then*

$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} P(A_i).$$

The triple $(\Omega, \mathcal{F}, P)$ is called a *probability space.*

$\mathcal{F}$-measurable functions

Definition 8. $\mathcal{F}$-measurable function
The function $f : \Omega \to \mathbb{R}$ defined on $(\Omega, \mathcal{F}, P)$ is called $\mathcal{F}$-measurable if

$$f^{-1}(B) = \{\omega \in \Omega : f(\omega) \in B\} \in \mathcal{F} \quad \textit{for all} \quad B \in \mathcal{B}(\mathbb{R}),$$

i.e., the inverse f^{-1} maps all of the Borel sets $B \subset \mathbb{R}$ to $\mathcal{F}$. Sometimes it is easier to work with following equivalent condition:

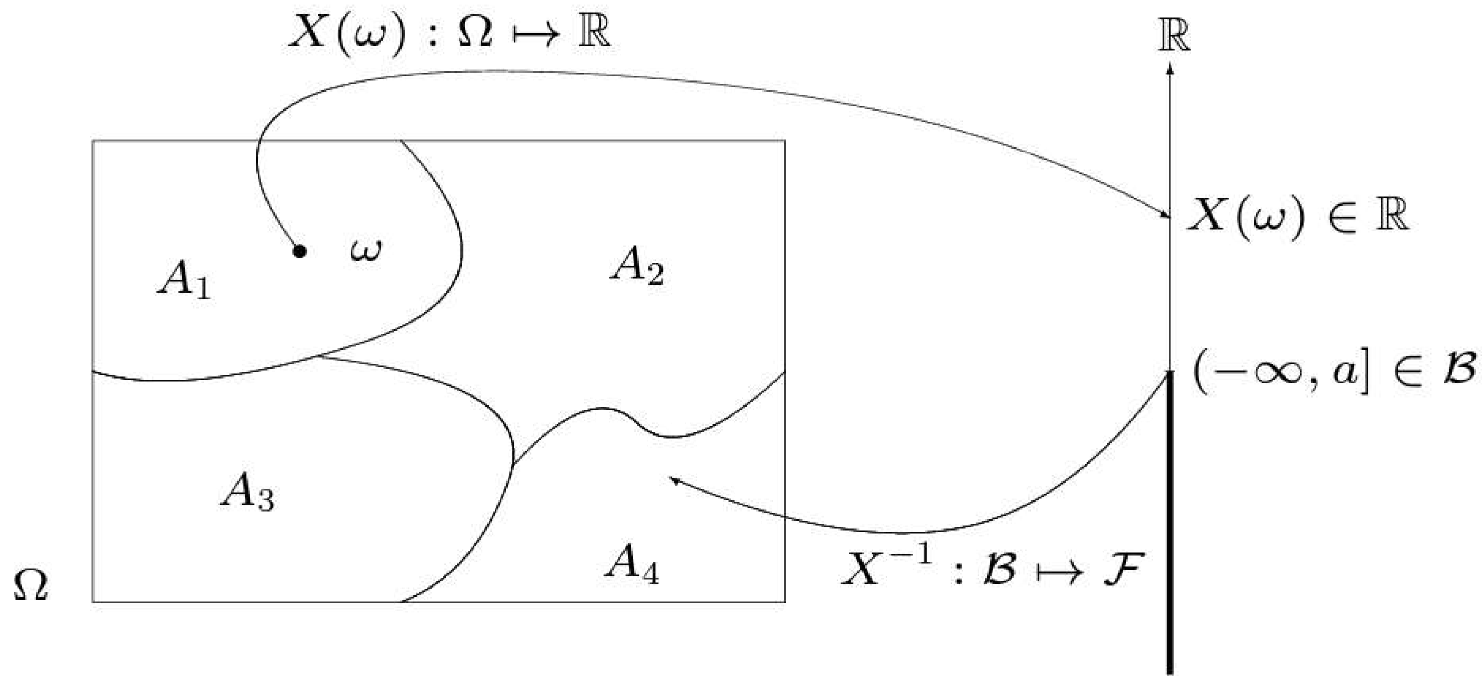
$$y \in \mathbb{R} \Rightarrow \{\omega \in \Omega : f(\omega) \leq y\} \in \mathcal{F}$$

This means that once we know the (random) value $X(\omega)$ we know which of the events in $\mathcal{F}$ have happened.

- $\mathcal{F} = \{\emptyset, \Omega\}$: only constant functions are measurable
- $\mathcal{F} = 2^{\Omega}$: all functions are measurable

$\mathcal{F}$-measurable functions

Ω: Sample space, A_i: Event, $\mathcal{F} = \sigma(A_1, \ldots, A_n)$: σ-algebra of events



X: random variable, $\mathcal{B}$: Borel σ-algebra

$\mathcal{F}$-measurable functions - Example

Roll of a die: $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$
We only know, whether an even or and odd number has shown up:
$\mathcal{F} = \{\emptyset, \{\omega_1, \omega_3, \omega_5\}, \{\omega_2, \omega_4, \omega_6\}, \Omega\} = \sigma(\{\omega_1, \omega_3, \omega_5\})$

Consider the following random variable:

$$f(\omega) = \begin{cases} 1, & \text{if } \omega = \omega_1, \omega_2, \omega_3; \\ -1, & \text{if } \omega = \omega_4, \omega_5, \omega_6. \end{cases}$$

Check measurability with the condition

$$y \in \mathbb{R} \Rightarrow \{\omega \in \Omega : f(\omega) \leq y\} \in \mathcal{F}$$

$\{\omega \in \Omega : f(\omega) \leq 0\} = \{\omega_4, \omega_5, \omega_6\} \notin \mathcal{F} \Rightarrow f$ is not $\mathcal{F}$-measurable.

Lebesgue integral I

Definition 9. Lebesgue Integral
$(\Omega, \mathcal{F})$ *a measure space,* $\mu : \Omega \to \mathbb{R}$ *a measure,* $f : \Omega \to \mathbb{R}$ *is* $\mathcal{F}$*-measurable.*

- *If* f *is a* $\textit{simple function}$*, i.e.,* $f(x) = c_i, \quad$ *for all* $\quad x \in A_i$*,* $c_i \in \mathbb{R}$

$$\int_\Omega f d\mu = \sum_{i=1}^{n} c_i \mu(A_i).$$

- *If* f *is* $\textit{nonnegative}$*, we can always construct a sequence of simple functions* f_n *with* $f_n(x) \leq f_{n+1}(x)$ *which converges to* f*:* $\lim_{n\to\infty} f_n(x) = f(x)$*.*
With this sequence, the Lebesgue integral is defined by

$$\int_\Omega f d\mu = \lim_{n\to\infty} \int_\Omega f_n d\mu.$$

Lebesgue integral II

Definition 10. Lebesgue Integral
$(\Omega, \mathcal{F})$ *a measure space,* $\mu : \Omega \to \mathbb{R}$ *a measure,* $f : \Omega \to \mathbb{R}$ *is* $\mathcal{F}$*-measurable.*

- *If* f *is an arbitrary, measurable function, we have* $f = f^+ - f^-$ *with*

$$f^+(x) = max(f(x), 0) \qquad and \qquad f^-(x) = max(-f(x), 0),$$

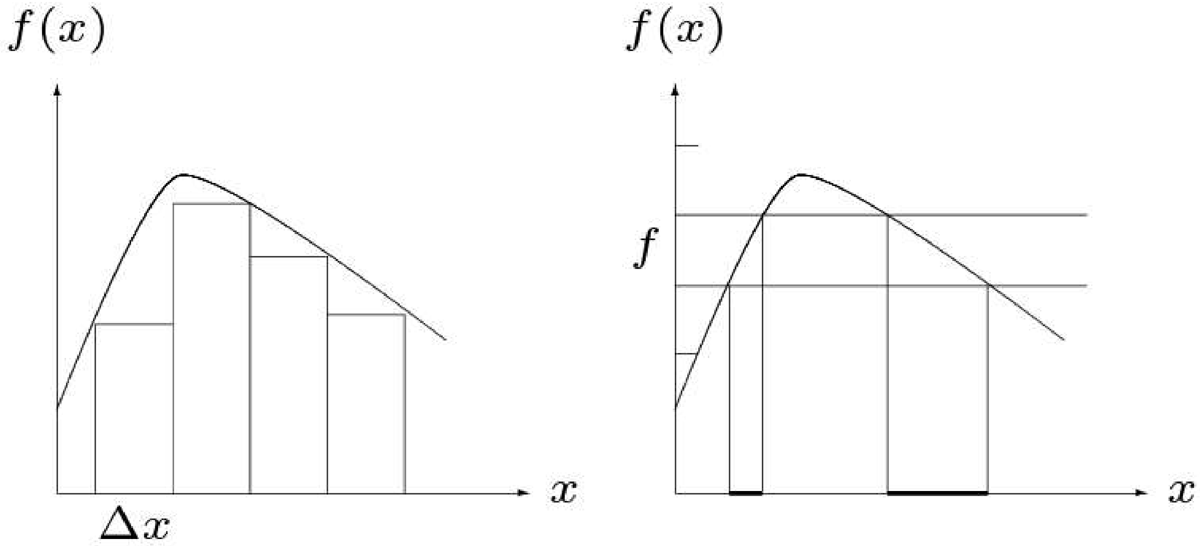
and then define

$$\int_\Omega f d\mu = \int_\Omega f^+ dP - \int_\Omega f^- dP.$$

The integral above may be finite or infinite. It is not defined if $\int_\Omega f^+ dP$ *and* $\int_\Omega f^- dP$ *are both infinite.*

Riemann vs. Lebesgue

The most important concept of the Lebesgue integral is that the **limit of approximate sums** (as the Riemann integral): for $\Omega = \mathbb{R}$:



Riemann vs. Lebesgue integral

Theorem 1. Riemann-Lebesgue integral equivalence
Let f be a bounded and continuous function on $[x_1, x_2]$ *except at a countable number of points in* $[x_1, x_2]$*. Then both the Riemann and the Lebesgue integral with Lebesgue measure* μ *exist and are the same:*

$$\int_{x_1}^{x_2} f(x)\, dx = \int_{[x_1,x_2]} f\, d\mu.$$

There are more functions which are Lebesgue integrable than Riemann integrable.

Popular example for Riemann vs. Lebesgue

Consider the function

$$f(x) = \begin{cases} 0, & x \in \mathbb{Q}; \\ 1, & x \in \mathbb{R} \backslash \mathbb{Q}. \end{cases}$$

The Riemann integral

$$\int_0^1 f(x)dx$$

does not exist, since lower and upper sum do not converge to the same value.
However, the Lebesgue integral

$$\int_{[0,1]} f d\lambda = 1$$

does exist, since f(x) is the indicator function of $x \in \mathbb{R} \backslash \mathbb{Q}$.

Random Variable

Definition 11. Random variable
A real-valued random variable X is a $\mathcal{F}$-measurable function defined on a probability space $(\Omega, \mathcal{F}, P)$ mapping its sample space Ω into the real line $\mathbb{R}$:

$$X : \Omega \rightarrow \mathbb{R}.$$

Since X is $\mathcal{F}$-measurable we have $X^{-1} : \mathcal{B} \rightarrow \mathcal{F}$.

Distribution function

Definition 12. Distribution function
The distribution function of a random variable X, defined on a probability space $(\Omega, \mathcal{F}, P)$, *is defined by:*

$$F(x) = P(X(\omega) \leq x) = P(\{\omega : X(\omega) \leq x\}).$$

From this the probability measure of the half-open sets in $\mathbb{R}$ is

$$P(a < X \leq b) = P(\{\omega : a < X(\omega) \leq b\}) = F(b) - F(a).$$

Density function

Closely related to the distribution function is the density function. Let $f : \mathbb{R} \mapsto \mathbb{R}$ be a nonnegative function, satisfying $\int_{\mathbb{R}} f d\lambda = 1$. The function f is called a density function (with respect to the Lebesgue measure) and the associated probability measure for a random variable X, defined on $(\Omega, \mathcal{F}, P)$, is

$$P(\{\omega : \omega \in A\}) = \int_A f d\lambda.$$

for all $A \in \mathcal{F}$.

Important Densities I

- Poisson density or probability mass function ($\lambda > 0$):

$$f(x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad , \quad x = 0, 1, 2, \ldots \; .$$

- Univariate Normal density

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)}$$

. The normal variable is abbreviated as $\mathcal{N}(\mu, \sigma)$.

- Multivariate normal density ($x, \mu \in \mathbb{R}^n; \Sigma \in \mathbb{R}^{n \times n}$):

$$f(x) = \frac{1}{\sqrt{(2\pi)^n det(\Sigma)}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)} \; .$$

Important Densities II

- Univariate student t-density ν degrees of freedom $(x, \mu \in \mathbb{R}^1; \sigma \in \mathbb{R}^1$

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu\sigma}}\Big(1 + \frac{1}{\nu}\frac{(x-\mu)^2}{\sigma^2}\Big)^{-\frac{1}{2}(\nu+1)}$$

- Multivariate student t-density with ν degrees of freedom $(x, \mu \in \mathbb{R}^n; \Sigma \in \mathbb{R}^{n\times n})$:

$$f(x) = \frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\frac{\nu}{2})\sqrt{(\pi\nu)^n det(\Sigma)}}\Big(1 + \frac{1}{\nu}(x-\mu)^T\Sigma^{-1}(x-\mu)\Big)^{-\frac{1}{2}(\nu+n)}.$$

Important Densities II

- The chi square distribution with degree-of-freedom (dof) n has the following density

$$f(x) = \frac{e^{-\frac{x}{2}} \left(\frac{x}{2}\right)^{\frac{n-2}{2}}}{2\Gamma(\frac{n}{2})}$$

which is abreviated as $Z \sim \chi^2(n)$ and where Γ denotes the gamma function.

- A chi square distributed random variable Y is created by

$$Y = \sum_{i=1}^{n} X_i^2$$

where X are independent standard normal distributed random variables $\mathcal{N}(0,1)$.

M

Important Densities III

- A standard student-t distributed random variable Y is generated by

$$Y = \frac{X}{\sqrt{\frac{Z}{\nu}}},$$

where $X \sim \mathcal{N}(0,1)$ and $Z \sim \chi^2(\nu)$.

- Another important density is the Laplace distribution:

$$p(x) = \frac{1}{2\sigma} e^{-\frac{|x-\mu|}{\sigma}}$$

with mean μ and diffusion σ. The variance of this distribution is given as $2\sigma^2$.

Expectation & Variance

Definition 13. Expectation of a random variable
The expectation of a random variable X, defined on a probability space $(\Omega, \mathcal{F}, P)$, *is defined by:*

$$E[X] = \int_\Omega X dP = \int_\Omega x f d\lambda.$$

With this definition at hand, it does not matter what the sample Ω is. The calculations for the two familiar cases of a finite Ω and $\Omega \equiv \mathbb{R}$ with continuous random variables remain the same.

Definition 14. Variance of a random variable
The variance of a random variable X, defined on a probability space $(\Omega, \mathcal{F}, P)$, *is defined by:*

$$\text{var}(X) = E[(X - E[X])^2] = \int_\Omega (X - E[X])^2 dP = E[X^2] - E[X]^2.$$

Normally distributed random variables

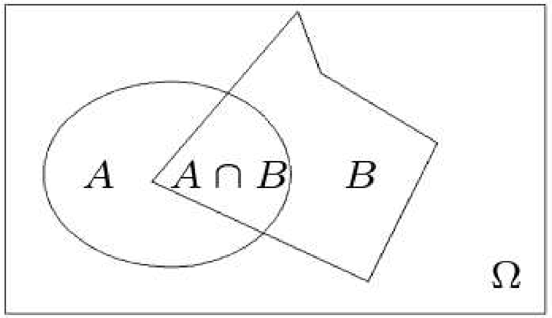
The shorthand notation $X \sim \mathcal{N}(\mu, \sigma^2)$ for normally distributed random variables with parameters μ and σ is often found in the literature. The following properties are useful when dealing with normally distributed random variables:

- If $X \sim \mathcal{N}(\mu, \sigma^2))$ and $Y = aX + b$, then $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2))$.
- If $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2))$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2))$ then $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ (if X_1 and X_2 are independent)

Conditional Expectation I

From elementary probability theory (Bayes rule):

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} \quad , \quad P(B) > 0.$$



$$E(X|B) = \frac{E(XI_B)}{P(B)} \quad , \quad P(B) > 0.$$

Conditional Expectation II

General case: $(\Omega, \mathcal{F}, P)$

Definition 15. Conditional expectation

Let X be a random variable defined on the probability space $(\Omega, \mathcal{F}, P)$ with $E[|X|] < \infty$. Furthermore let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$ ($\mathcal{G} \subseteq \mathcal{F}$). Then there exists a random variable Y with the following properties:

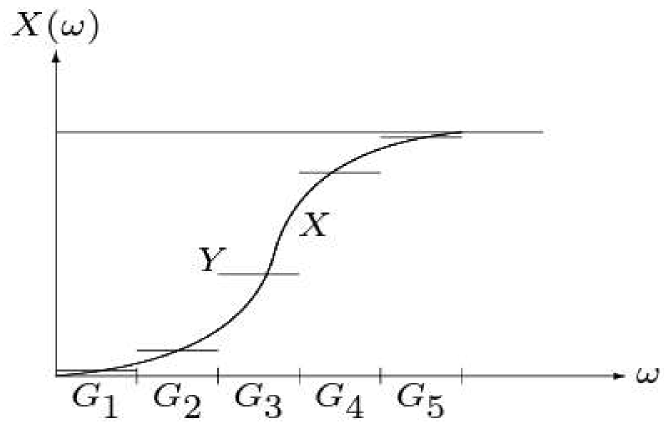
1. *Y is $\mathcal{G}$-measurable.*
2. $E[|Y|] < \infty$.
3. *For all sets G in $\mathcal{G}$ we have*

$$\int_G Y dP = \int_G X dP, \qquad \text{for all} \quad G \in \mathcal{G}.$$

The random variable $Y = E[X|\mathcal{G}]$ is called conditional expectation.

Conditional Expectation: Example

Y is a piecewise linear approximation of X.
$Y = E[X|\mathcal{G}]$



For the trivial σ-algebra $\{\emptyset, \Omega\}$:
$Y = E[X|\{\emptyset, \Omega\}] = \int_\Omega X dP = E[X]$.

Conditional Expectation: Properties

- $E(E(X|\mathcal{F})) = E(X)$.
- If X is $\mathcal{F}$-measurable, then $E(X|\mathcal{F}) = X$.
- Linearity: $E(\alpha X_1 + \beta X_2|\mathcal{F}) = \alpha E(X_1|\mathcal{F}) + \beta E(X_2|\mathcal{F})$.
- Positivity: If $X \geq 0$ almost surely, then $E(X|\mathcal{F}) \geq 0$.
- Tower property: If $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$, then

$$E(E(X|\mathcal{F})|\mathcal{G}) = E(X|\mathcal{G}).$$

- Taking out what is known: If Z is $\mathcal{G}$-measurable, then

$$E(ZX|\mathcal{G}) = Z \cdot E(X|\mathcal{G}).$$

Summary

- σ-algebra: collection of the events of interest, closed under elementary set operations
- Borel σ-algebra: all the events of practical importance in $\mathbb{R}$
- Lebesgue measure: defined as the length of an interval
- Density: transforms Lebesgue measure in a probability measure
- Measurable function: the σ-algebra of the probability space is "rich" enough
- Random variable X: a measurable function $X : \Omega \mapsto \mathbb{R}$
- Expectation, Variance
- Conditional expectation is a piecewise linear approximation of the underlying random variable.