Introduction to Probabilistic Reasoning

3. Inferring values – propagating uncertainties.

Giulio D'Agostini

Dipartimento di Fisica

Università di Roma La Sapienza

Summary on probabilistic approach

- Probability means how much we believe something
- Probability depends on available information
  $\longrightarrow$ subjective
- Probability values obey the following basic rules

  1. $0 \leq P(A \mid I) \leq 1$
  2. $P(\Omega \mid I) = 1$
  3. $P(A \cup B \mid I) = P(A \mid I) + P(B \mid I)$ [if $P(A \cap B \mid I) = \emptyset$]
  4. $P(A \cap B \mid I) = P(A \mid B, I) \cdot P(B \mid I) = P(B \mid A, I) \cdot P(A \mid I)$

- All the rest by logic

$\longrightarrow$ And, please, be coherent!

Summary on probabilistic approach

- Probability means how much we believe something
- Probability depends on available information
  $\longrightarrow$ subjective
- Probability values obey the following basic rules

  1. $0 \leq P(A \,|\, I) \leq 1$
  2. $P(\Omega \,|\, I) = 1$
  3. $P(A \cup B \,|\, I) = P(A \,|\, I) + P(B \,|\, I) \quad$ [if $P(A \cap B \,|\, I) = \emptyset$]
  4. $P(A \cap B \,|\, I) = P(A \,|\, B, I) \cdot P(B \,|\, I) = P(B \,|\, A, I) \cdot P(A \,|\, I)$

- All the rest by logic

$\longrightarrow$ And, please, be coherent!

$\Rightarrow$ more comments on $P(A \,|\, B, I)$ $\longrightarrow$

Rule 4 does not <u>define</u> conditional probability

$$4. \quad P(A \cap B \,|\, I) = P(A \,|\, B, I) \cdot P(B \,|\, I) = P(B \,|\, A, I) \cdot P(A \,|\, I)$$

is not a definition of conditional probability (at least in this framework):

- probability <u>is always conditioned</u> by some hypotheses;
- rule nr 4 just relates one belief to the other two ($\Rightarrow$ e.g.: detector efficiency to a 70 GeV Higgs)

Condition is, in general, hypothetical

$A =$ Berlin wins the football match against Athens
$B =$ It rains

$$P(A \,|\, B, I) \;=\; p \tag{1}$$

with I indicating wathever I know about Berlin, Athens, football, etc.

does not means "how much I believe Berlin will win once I know it rains",

but, simpler, "how much I believe Berlin will win under the hypothesis that it rains"

Condition is, in general, hypothetical

$A =$ Berlin wins the football match against Athens
$B =$ It rains

$$P(A \,|\, B, I) \;=\; p \tag{2}$$

with I indicating wathever I know about Berlin, Athens, football, etc.

does not means "how much I believe Berlin will win once I know it rains",

but, simpler, "how much I believe Berlin will win under the hypothesis that it rains"

$$P(A \cap B \,|\, I) = P(A \,|\, B, I) \cdot P(B \,|\, I)$$

$P(A \cap B \,|\, I)$ is then usually different ($\rightarrow$ smaller) than $P(A \,|\, B, I)$

Condition is, in general, hypothetical

$A =$ Berlin wins the football match against Athens
$B =$ It rains

$$P(A\,|\,B, I) \;=\; p \tag{3}$$

with I indicating wathever I know about Berlin, Athens, football, etc.

does not means "how much I believe Berlin will win once I know it rains",

but, simpler, "how much I believe Berlin will win under the hypothesis that it rains"

$$P(A\cap B\,|\,I) = P(A\,|\,B, I)\cdot P(B\,|\,I)$$

$P(A\cap B\,|\,I)$ is then usually different ($\rightarrow$ smaller) than $P(A\,|\,B, I)$
if we know B, then $P(B) = 1$ and $P(A\cap B)$ would be equal to $P(A\,|\,B)$

Conditioned events in terms of bets

$A \mid B$

True: B occurs <u>and</u> A occurs:
$\longrightarrow$ I win the bet

False: B occurs <u>and</u> A does not:
$\longrightarrow$ I lose the bet

Undefined: B does not occur: $\longrightarrow$ bet invalidated

Independence

We remind that two events are called *independent* if

$$P(E \cap H) = P(E)\, P(H)\,.$$

This is equivalent to saying that

- $P(E \,|\, H) = P(E)$ and
- $P(H \,|\, E) = P(H)$,

i.e. the hypothesis that one event has occurs does not change the probability of the other.

Independence

We remind that two events are called *independent* if

$$P(E \cap H) = P(E)\,P(H)\,.$$

This is equivalent to saying that

- $P(E \,|\, H) = P(E)$ and
- $P(H \,|\, E) = P(H)$,

i.e. the hypothesis that one event has occurs does not change the probability of the other. If $P(E \,|\, H) \neq P(E)$, then the events E and H are *correlated*. In particular:

- if $P(E \,|\, H) > P(E)$ then E and H are *positively* correlated;
- if $P(E \,|\, H) < P(E)$ then E and H are *negatively* correlated.

Independence

We remind that two events are called *independent* if

$$P(E \cap H) = P(E)\, P(H)\,.$$

This is equivalent to saying that

- $P(E \,|\, H) = P(E)$ and
- $P(H \,|\, E) = P(H)$,

i.e. the hypothesis that one event has occurs does not change the probability of the other. If $P(E \,|\, H) \neq P(E)$, then the events E and H are *correlated*. In particular:

- if $P(E \,|\, H) > P(E)$ then E and H are *positively* correlated;
- if $P(E \,|\, H) < P(E)$ then E and H are *negatively* correlated.

[By the way, $P(E \cap H \,|\, I) = P(E \,|\, I)\, P(H \,|\, I)$, and so on.]

Independence

We remind that two events are called *independent* if

$$P(E \cap H) = P(E)\, P(H)\,.$$

This is equivalent to saying that

- $P(E \,|\, H) = P(E)$ and
- $P(H \,|\, E) = P(H)$,

i.e. the hypothesis that one event has occurs does not change the probability of the other. If $P(E \,|\, H) \neq P(E)$, then the events E and H are *correlated*. In particular:

- if $P(E \,|\, H) > P(E)$ then E and H are *positively* correlated;
- if $P(E \,|\, H) < P(E)$ then E and H are *negatively* correlated.

[By the way, $P(E \cap H \,|\, I) = P(E \,|\, I)\, P(H \,|\, I)$, and so on.]
$\longrightarrow$ comment on definition Vs use of ‘independence’

Example from six box problem

$$
\begin{aligned}
P(W \mid I) &= 1/2 \\
P(B \mid I) &= 1/2
\end{aligned}
$$

$$
\begin{aligned}
P(W \mid H_j, I) &= j/5 \\
P(W \mid H_j, I) &= (1-j)/5
\end{aligned}
$$

W is positively correlated with H_4,
negatively correlated with H_2, etc

Stochastical dependence vs physical dependence

See slides on logical independence Vs stochastical (probabilistic) independence

. . . not always intuitive

Events and sets

Convenient event $\leftrightarrow$ set analogy:

		Symbol
event	set	E
certain	sample space	Ω
impossible	empty	$\emptyset$
implication	inclusion	$E_1 \subseteq E_2$
opposite (complementary)	complementary	$\overline{E} \quad (E \cup \overline{E} = \Omega)$
logical product	intersection	$E_1 \cap E_2$
logical sum	union	$E_1 \cup E_2$
incompatible	disjoint	$E_1 \cap E_2 = \emptyset$
complete class	finite partition	$\begin{cases} E_i \cap E_j = \emptyset \;\; \forall\, i \neq j \\ \cup_i E_i = \Omega \end{cases}$

Rules of probability

- $P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)$ if $E_i \cap E_j = \emptyset \quad \forall i \neq j$ (just an extension of the basic rule 3).
- $P(E) = 1 - P(\overline{E})$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (generalization of '3')
- $P(E) = P(E \cap H) + P(E \cap \overline{H})$

Rules of probability

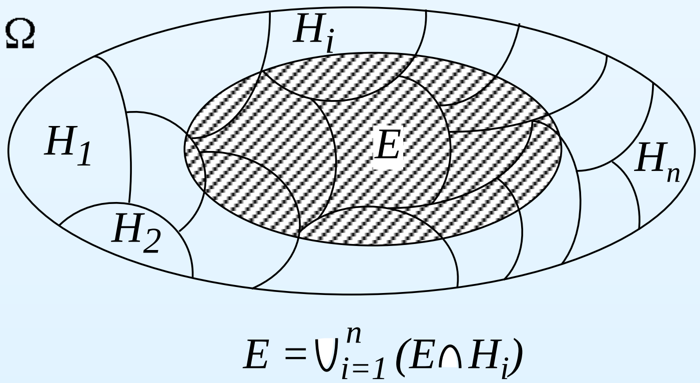
- $P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)$ if $E_i \cap E_j = \emptyset \quad \forall i \neq j$
  (just an extension of the basic rule 3).
- $P(E) = 1 - P(\overline{E})$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (generalization of '3')
- $P(E) = P(E \cap H) + P(E \cap \overline{H})$

$\longrightarrow$ Extension to complete class of events:

$$P(E) = P\left(\bigcup_{i=1}^{n} (E \cap H_i)\right) = \sum_{i=1}^{n} P(E \cap H_i)$$

and, applying '4'

$P(E) \;=\; \sum_i P(H_i) \cdot P(E \,|\, H_i)$
('decomposition law')



$\longrightarrow$ weighted average of $P(E \,|\, H_i)$

Rules of probability

- $P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)$ if $E_i \cap E_j = \emptyset \quad \forall i \neq j$
  (just an extension of the basic rule 3).
- $P(E) = 1 - P(\overline{E})$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (generalization of '3')
- $P(E) = P(E \cap H) + P(E \cap \overline{H})$

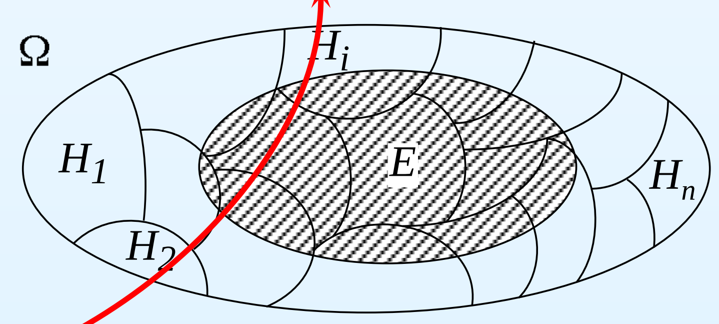
$\longrightarrow$ Extension to complete class of events:

$$P(E) = P\left(\bigcup_{i=1}^{n} (E \cap H_i)\right) = \sum_{i=1}^{n} P(E \cap H_i)$$

and, applying '4'

$$P(E) \;=\; \sum_i P(H_i) \cdot P(E \,|\, H_i)$$

('decomposition law')



$E = \bigcup_{i=1}^{n} (E \cap H_i)$

$\rightarrow$ basis of 'marginalization'

Recovering the combinatorial evaluation formula

$$p = \frac{\#\,\text{favorable cases}}{\#\,\text{possible equiprobable cases}}$$

Given the ‘elementary’, equiprobable n events e_i forming a complete class, i.e. $\cup_i e_i = \Omega$,
we are interested in $P(E)$, where E = “$\cup$ m elementary events”

Recovering the combinatorial evaluation formula

$$p \;=\; \frac{\#\,\text{favorable cases}}{\#\,\text{possible equiprobable cases}}$$

Given the 'elementary', equiprobable n events e_i forming a complete class, i.e. $\cup_i e_i = \Omega$,
we are interested in $P(E)$, where E = "$\cup\; m$ elementary events"

$$\begin{aligned} P(e_i) &= p_0 \\ P(\cup_i e_i) &= \sum_i P(e_i) = n\, p_0 = 1 \\ \rightarrow \quad p_0 &= \frac{1}{n} \\ \rightarrow \quad P(E) &= \sum_{e_i \subset E} P(e_i) = m\, p_0 = \frac{m}{n} \end{aligned}$$

Teoria della probabilità Vs calcolo combinatorio

Capitolo 3 della dispense di Probabilità: da saltare!

- troppo spesso si confondono corsi di probabilità con corsi di calcolo combinatorio

Teoria della probabilità Vs calcolo combinatorio

Capitolo 3 della dispense di Probabilità: da saltare!

- troppo spesso si confondono corsi di probabilità con corsi di calcolo combinatorio

  a volte lo si confonde con rudimenti di statistica descrittiva

Teoria della probabilità Vs calcolo combinatorio

Capitolo 3 della dispense di Probabilità: da saltare!

- troppo spesso si confondono corsi di probabilità con corsi di calcolo combinatorio

  a volte lo si confonde con rudimenti di statistica descrittiva

  quasi sempre manca la parte inferenziale ("il problema essenziale del metodo sperimentale", secondo Poincaré).

Recovery frequency based "evaluation rule"

Reminder:

$p \longrightarrow f_n$: Bernoulli theorem

$f_n \longrightarrow p$: Bayes' Theorem
(under well defined assumptions – Laplace rule of succession)

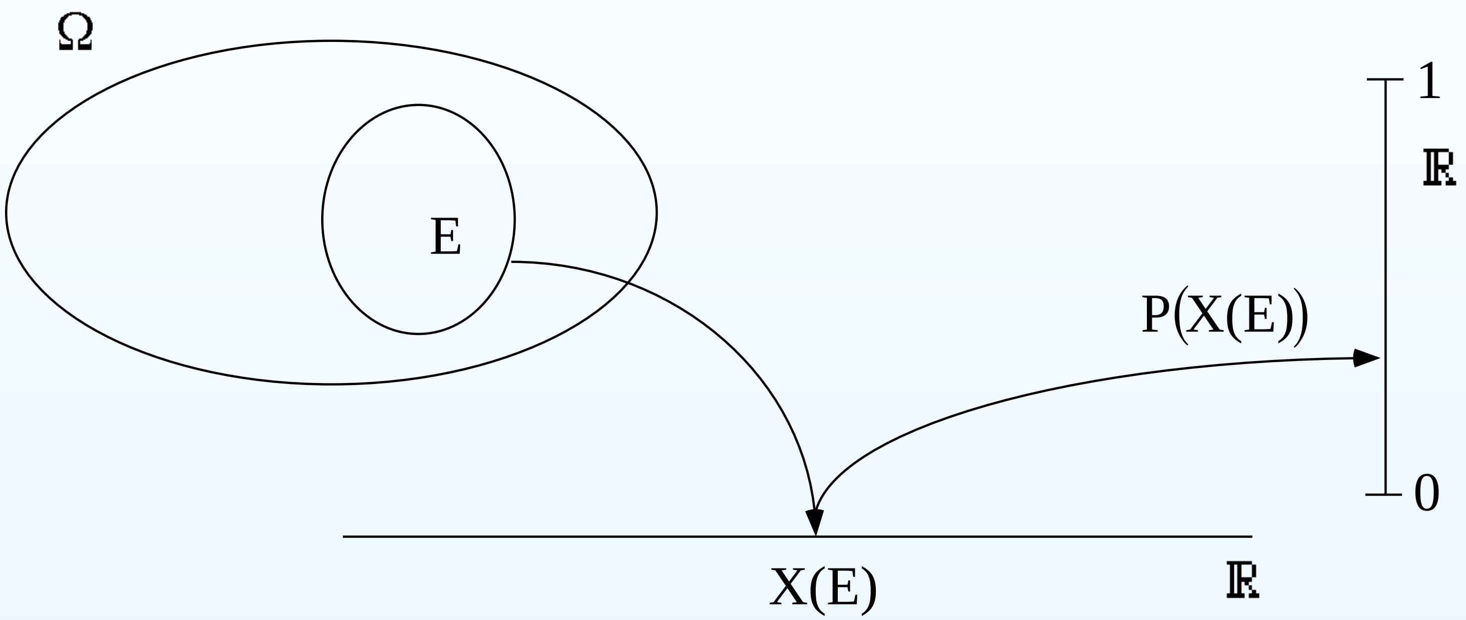
$\longrightarrow$ If we know exactly the box composition there is little to learn about the proportion p

Errori dei giocatori del lotto (e non solo. . .)

Numeri ritardatari: si confonde il teorema di Bernoulli ("mal raccontato. . . ") come una Legge e quindi sembra che i numeri in ritardo si precipitino verso il foro di uscita, sgomitando, per mettersi in regola con la "legge dei grandi numeri" (forse temono qualche terribile sanzione. . .).

Numeri caldi : si confondono distribuzioni statistiche (un dato empirico di statistica descrittiva) con distribuzioni di probabilità, tendendo così a credere che ci siano dei numeri particolarmenti vivaci e anche 'sovversivi' che tentano di uscire, irrispettosi della "legge dei grandi numeri".

From events to uncertain numbers

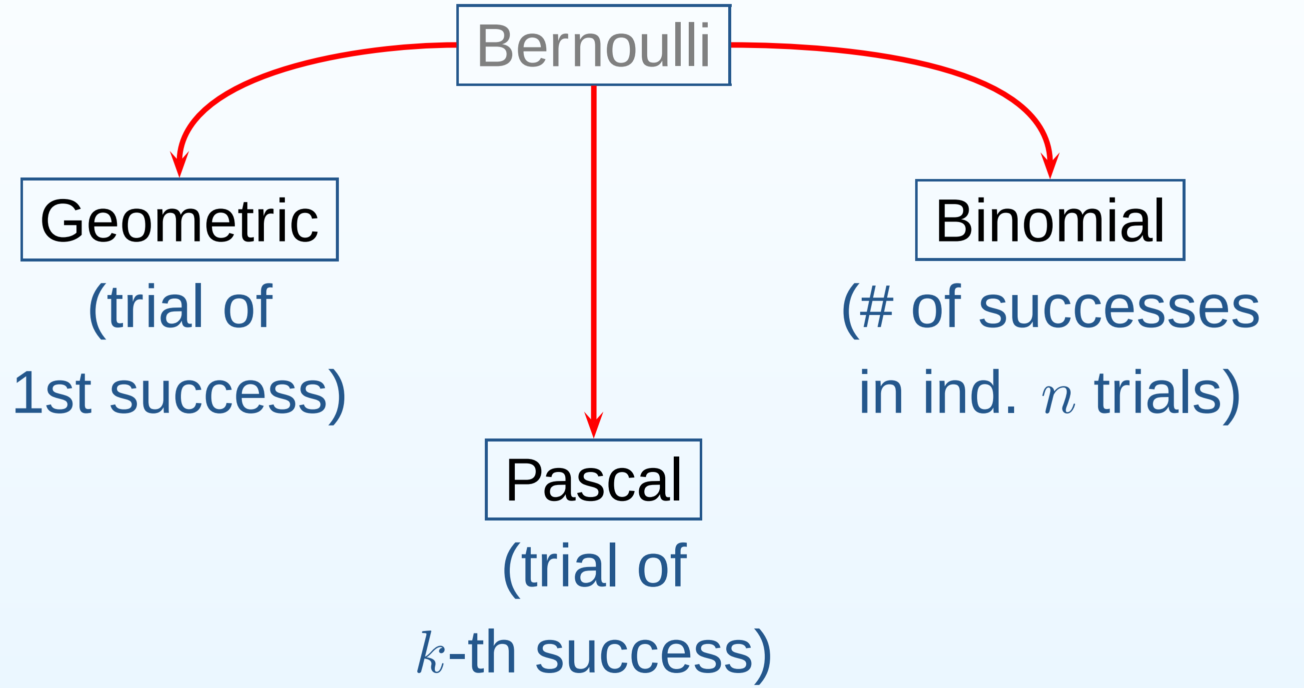


Uncertain numbers are associated to events

- Rolling one die: $X = 4 \leftrightarrow$ ‘face marked with 4’
  (note: no intrinsic order in the numbers associated a die)

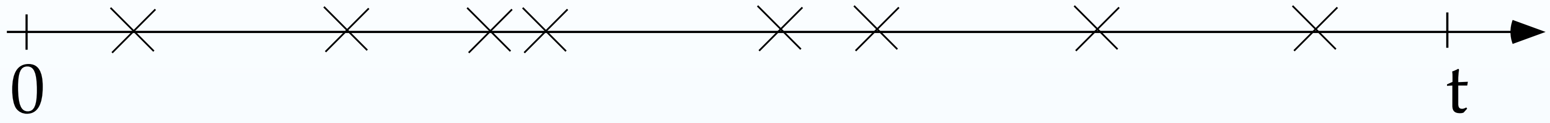
$\rightarrow$ $P(X = 4) = P($‘face marked with 4’$)$

Distributions derived from the Bernoulli process



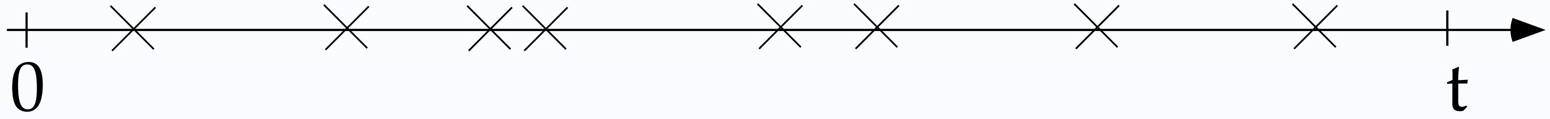
(Binomial well known. We shall not use the Pascal)

Poisson process



Phenomena that might occur or not at a give time (or at a give position), with r the intensity of the process.

Poisson process

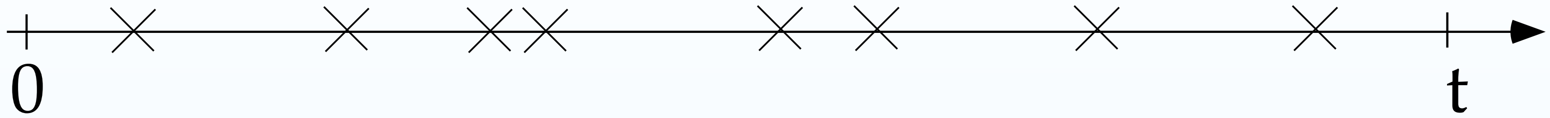


Phenomena that might occur or not at a give time (or at a give position), with r the intensity of the process.

- direct probability problems:
  - given a fixed ‘measuring’ time T, how many counts do we expect?
    $\rightarrow$ Poisson distribution: $\lambda = r\,T$;
  - How much time do we have to wait before the first event occurs? $\rightarrow$ exponential distribution: $\tau = 1/r$;

$$\lambda \leftrightarrow \tau$$

Poisson process



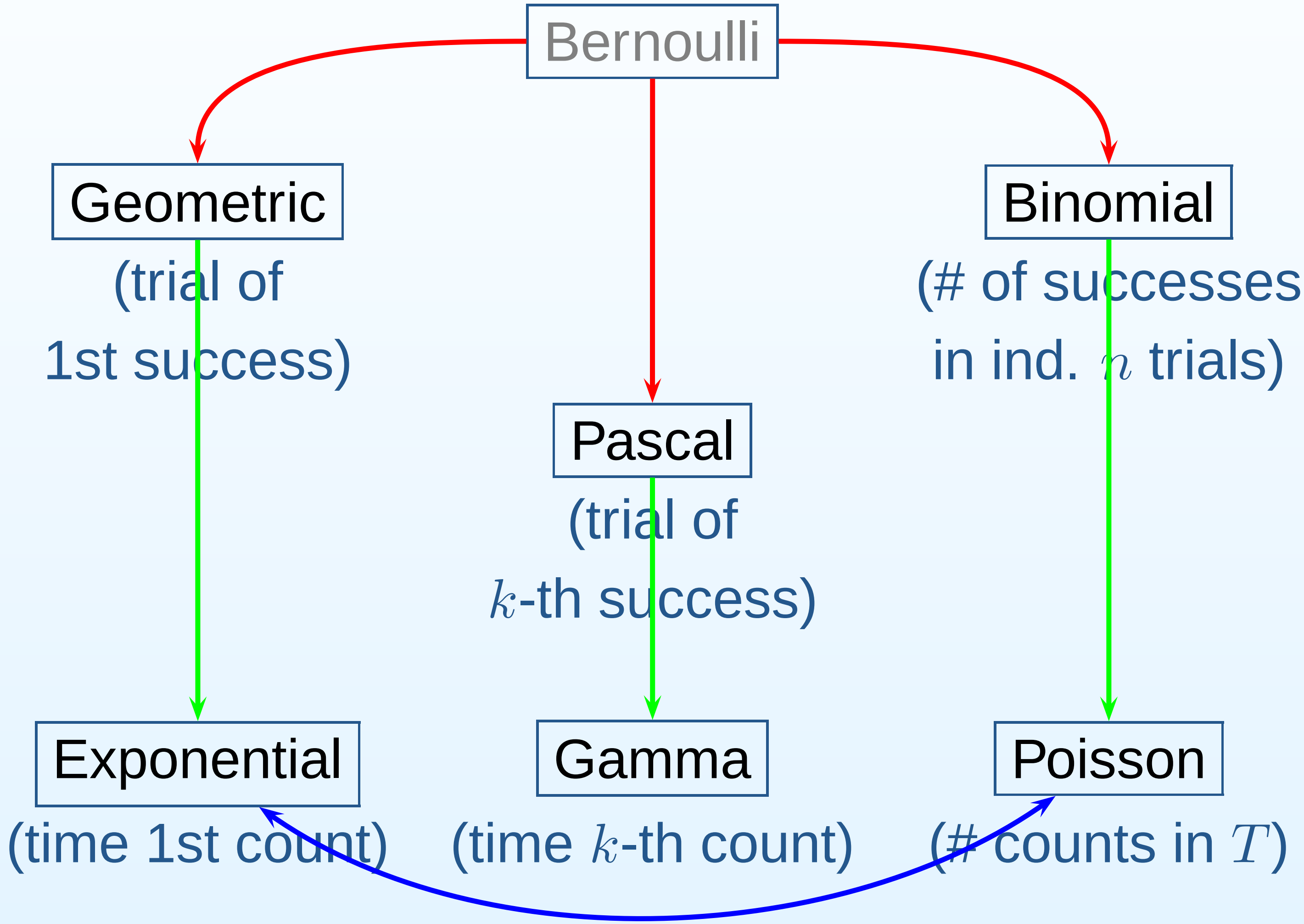
Phenomena that might occur or not at a give time (or at a give position), with r the intensity of the process.

- direct probability problems:
  - given a fixed 'measuring' time T, how many counts do we expect?
    $\rightarrow$ Poisson distribution: $\lambda = r\,T$;
  - How much time do we have to wait before the first event occurs? $\rightarrow$ exponential distribution: $\tau = 1/r$;

$$\lambda \leftrightarrow \tau$$

- inverse probability problems:
  $\Rightarrow$ infer r from nr. of counts, from timing of counts, etc.

Distributions derived from the Bernoulli process



Seguita su lucidi tradizionali. . .

- Distribuzioni di variabili contine: uniforme; triangolari; esponenziale; gaussiana.
- Distribuzioni ‘vettori aleatori’: generalità (in particolare, covarianza, coefficiente di correlazione e distribuzioni condizionate).
- Distribuzione normale bivariata e multivariata; distribuzione condizionata nel caso bivariato.
- Propagazione delle incertezze: dal problema generale al caso particolare di combinazione lineare ed estensione mediante linearizzazione (con caveat sull’uso della linearizzazione); soluzione generale mediante campionamento (Monte Carlo); trasformazione della matrice di covarianza (combinazioni lineari e linearizzazioni).

. . . seguita su lucidi tradizionali. . .

- Teorema centrale del limite e applicazioni; dal processo di Bernoulli al random walk con limite gaussiano ed estensioni varie (dalla rovina del giocatore alla distribuzione di Maxwell delle molecole).
- Inferenza parametrica applicata a mu di una gaussiana, inclusa distribuzione predittiva.
- Quadro generale inferenziale-predittivo in termini di reti bayesiane.

Per riferimenti, link, etc. vedi sul sito.

End

FINE