Statistical Methods for Data Analysis

Probability and PDF’s

Luca Lista

INFN Napoli
Definition of probability

• There are two main different definitions of the concept of probability

• Frequentist
  – **Probability** is the ratio of the number of occurrences of an event to the total number of experiments, in the *limit* of very large number of **repeatable** experiments.
  – Can only be applied to a specific classes of events (repeatable experiments)
  – Meaningless to state: “**probability that the lightest SuSy particle’s mass is less than 1 TeV**”

• Bayesian
  – **Probability** measures someone’s the degree of belief that something is or will be true: **would you bet?**
  – Can be applied to most of unknown events (past, present, future):
    • “**Probability that Velociraptors hunted in groups**”
    • “**Probability that S.S.C Naples will win next championship**”
Classical probability

“The theory of chance consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining the number of cases favorable to the event whose probability is sought. The ratio of this number to that of all the cases possible is the measure of this probability, which is thus simply a fraction whose numerator is the number of favorable cases and whose denominator is the number of all the cases possible.”

Pierre-Simon Laplace, A Philosophical Essay on Probabilities
Classical Probability

Probability = \frac{\text{Number of favorable cases}}{\text{Number of total cases}}

- Assumes all accessible cases are \textit{equally probable}
- This analysis is rigorously \textit{valid on discrete cases only}
  - Problems in continuous cases (→ Bertrand’s paradox)

\begin{align*}
P &= \frac{1}{2} \\
P &= \frac{1}{6} \\
&\text{(each dice)} \\
P &= \frac{1}{10} \\
P &= \frac{1}{4}
\end{align*}
What about something like this?

We should move a bit further…
Probability and combinatorial

- Complex cases are managed via combinatorial analysis
- Reduce the event of interest into elementary equiprobable events
- Sample space ↔ Set algebra
  - and/or/not ↔ intersection/union/complement

E.g:

2 = {(1,1)}
3 = {(1,2), (2,1)}
4 = {(1,3), (2,2), (3,1)}
5 = {(1,4), (2,3), (3,2), (4,1)}

etc. …
Random extractions

• Success is the extraction of a red ball in a container of mixed white and red ball

  • Red: $p = 3/10$

  • White: $1 - p = 7/10$

• Success could also be: track reconstructed by a detector, or event selected by a set of cuts

• Classic probability applies only to integer cases, so strictly speaking, $p$ should be a rational number
Multiple random extractions

Extraction path leads to Pascal’s / Tartaglia’s triangle like $(a + b)^n$

\[
\begin{pmatrix} N \\ n \end{pmatrix} = \frac{N!}{n! (N - n)!}
\]
Binomial distribution

• Distribution of number of “successes” on \( N \) trials, each trial with “success” probability \( p \)

\[
P(n; N, p) = \frac{N!}{n!(N - n)!} p^n (1 - p)^{N-n}
\]

• Average: \( \langle n \rangle = Np \)
• Variance: \( \langle n^2 \rangle - \langle n \rangle^2 = Np(1-p) \)
• Frequently used for efficiency estimate
  – Efficiency error \( \sigma = \text{Var}(n)^{\frac{1}{2}} \):
    \[
    \sigma_\varepsilon = \sqrt{\frac{\varepsilon(1 - \varepsilon)}{N}}
    \]
    Note: \( \sigma_\varepsilon = 0 \) for \( \varepsilon = 0, 1 \)
Tartaglia or Pascal?

- **India**: 10th century commentaries of *Chandas Shastra Pingala*, dating 5th-2nd century BC

- **Persia**: Al-Karaji (953–1029), Omar Khayyám (1048-1131): "Khayyam triangle"

- **China**: Yang Hui (楊輝, 1238-1298): "Yang Hui triangle"

- **Germany**: Petrus Apianus (1495-1552)

- **Italy**: Niccolò Fontana Tartaglia *(Ars Magna*, by Gerolamo Cardano, 1545): “Triangolo di Tartaglia”

- **France**: Blaise Pascal *(Traité du triangle arithmétique*, 1655)
Bertrand’s paradox

• Given a randomly chosen chord on a circle, what is the probability that the chord’s length is larger than the side of the inscribed triangle?

• “Randomly chosen” is not a well defined concept in this case
• Some classical probability concepts become arbitrary until we move to PDF’s (uniform in which ‘metrics’?)

P = 1/2

P = 1/3

P = 1/4
Probability definition (freqentist)

• A bit more formal definition of probability:
• Law of large numbers:

\[ P(E) = p \quad \text{if} \quad \frac{N(E)}{N} \xrightarrow{P} p \]

\[ \forall \varepsilon \quad \lim_{N \to \infty} P \left( \left| \frac{N(E)}{N} - p \right| < \varepsilon \right) = 1 \]

... isn’t it a circular definition?
In a picture...

**Law of Large Numbers in Average of Die Rolls**

Average converges to expected value of 3.5
Problems with probability definitions

- Frequentist probability is, to some extent, circularly defined
  - A phenomenon can be proven to be random (i.e.: obeying laws of statistics) only if we observe infinite cases
  - F.James et al.: “this definition is not very appealing to a mathematician, since it is based on experimentation, and, in fact, implies unrealizable experiments \( (N \to \infty) \)”. But a physicist can take this with some pragmatism
  - A frequentist model can be justified by details of poorly predictable underlying physical phenomena
    - Deterministic dynamic with instability (chaos theory, …)
    - Quantum Mechanics is intrinsically probabilistic…!
  - A school of statisticians state that Bayesian statistics is a more natural and fundamental concept, and frequentist statistic is just a special sub-case

- On the other hand, Bayesian statistics is subjectivity by definition, which is unpleasant for scientific applications.
  - Bayesian reply that it is actually inter-subjective, i.e.: the real essence of learning and knowing physical laws…

- Frequentist approach is preferred by the large fraction of physicists (probably the majority, but Bayesian statistics is getting more and more popular in many application, also thanks to its easier application in many cases
Axiomatic definition (A. Kolmogorov)

• Axiomatic probability definition applies to both frequentist and Bayesian probability
  – Let $(\Omega, F \subseteq 2^\Omega, P)$ be a measure space that satisfy:
    
    1. $P(E) \geq 0 \ \forall E \in F$
    2. $P(\Omega) = 1$
    3. $\forall (E_1, \ldots, E_n) \in F^n : E_i \cap E_j = 0$
       
       $$P\left( \bigcup_{i=1,\ldots,n} E_i \right) = \sum_{i=1,\ldots,n} P(E_i)$$
    – Terminology: $\Omega =$ sample space, $F =$ event space, $P =$ probability measure

• So we have a formalism to deal with different types of probability
Conditional probability

- Probability of $A$, given $B$: $P(A \mid B)$
- i.e.: probability that an event known to belong to set $B$, is also a member of set $A$:
  - $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$
- Event $A$ is said to be independent on $B$ if the conditional probability of $A$ given $B$ is equal to the probability of $A$:
  - $P(A \mid B) = P(A)$
- Hence, if $A$ is independent on $B$:
  - $P(A \cap B) = P(A) \cdot P(B)$
- $\Rightarrow$ If $A$ is independent on $B$, $B$ is independent on $A$
Prob. Density Functions (PDF)

- Sample space = \{ \vec{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n \}
- Experiment = one point on the sample space
- Event = a subset \( A \) of the sample space
- P.D.F. = \( f(\vec{x}) = f(x_1, \ldots, x_n) \)
- Probability of an event \( A \) =

\[
P(A) = \int_A f(x_1, \ldots, x_n) \, d^n x
\]

- Differential probability:

\[
\frac{d^n P}{dx_1 \cdots dx_n} = f(x_1, \ldots, x_n)
\]

- For continuous cases, the probability of an event made of a single experiment is zero: \( P(\{x_0\}) = 0 \)
- Discrete variables may be treated as “Dirac’s \( \delta \)”
  - uniform treatment of continuous and discrete cases
Variables transformation (discrete)

• 1D case: \( y = Y(x) \)
• \( \{x_1, \ldots, x_n\} \rightarrow \{y_1, \ldots, y_m\} = \{Y(x_1), \ldots, Y(x_n)\} \)
  – Note that different \( Y(x_n) \) may coincide \( (n \neq m) \! \)!
• So:

\[
P(y) = \sum_{i:Y(x_i)=y} P(x_i)
\]

• Generalization to more variables is straightforward:

\[
P(z) = \sum_{i,j:Z(x_i,y_j)=z} P(x_i,y_j)
\]

• Sum on all cases which give the right combination \( (z) \)
• Will see how to generalize to the continuous case and get the error propagation!
Coordinate transformation

- From 2D to 1D:
  \[ f(z) = \int \delta(z - Z(x, y)) f(x, y) \, dx \, dy \]

- Generic change of coordinate (2D):
  \[ f'(x', y') = \int \delta(x' - X'(x, y)) \delta(y' - Y'(x, y)) f(x, y) \, dx \, dy \]

- Generalization of discrete case: sum only on elementary events (experiments) where:
  - \( x'_i = X'_i(x_1, x_2) \) (e.g.: result of sum of two dices)

- Easy to implement with Monte Carlo

- If the relation is invertible, the Jacobian determinant has to be multiplied by the transformed PDF
  - Replace \((x, y)\) in \(f\) with inverse transformation
  - Transform of the \(n\)-D volume with jacobian:
    \[
    \frac{d^n P}{d^n x} = \frac{d^n P}{d^n x'} \det \left| \frac{\partial x'_i}{\partial x_j} \right|
    \]
PDF Examples
Gaussian distribution

\[ g(x; \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left( -\frac{(x - \mu)^2}{2\sigma^2} \right) \]

- Average = \( \langle x \rangle = \mu \)
- Variance = \( \langle (x - \mu)^2 \rangle = \sigma^2 \)
- Widely used mainly because of the central limit theorem

\( \rightarrow \) next slide
Central limit theorem

- The average of $N$ random variables $R_n$ converges to a Gaussian, irrespective to the original distributions.

Adding $n$ flat distributions

Basic regularity conditions must hold (incl. finite variance)
Uniform ("flat") distribution

\[ f(x) = \begin{cases} 
\frac{1}{b-a} & \text{if } a \leq x \leq b \\
0 & \text{if } x < a \text{ or } x > b 
\end{cases} \]

- **RMS** = \[ \sigma = \frac{b - a}{\sqrt{12}} \]

- Model for position of rain drops, time of cosmic ray passage, etc.
- Basic distribution for pseudo-random number generators
Cumulative distribution

• Given a PDF $f(x)$, the cumulative is defined as:

$$F(x) = \int_{-\infty}^{x} f(x') \, dx'$$

• The PDF for $F$ is uniformly distributed in $[0, 1]$:

$$\frac{dP}{dF} = \frac{dP}{dx} \frac{dx}{dF} = \frac{f(x)}{f(x)} = 1$$
Distance from a time $t_0$ to first ‘count’

$t_0$ is an arbitrary time
Also: $t = \text{time between } t_0 \text{ and the first count}$

Notation: $P(n, [t_1, t_2]) = \text{probability of } n \text{ entries in } [t_1, t_2] \text{ given a rate } r$

$P(1, [t, t + \delta t]) = r\delta t$

$P(0, [0, t + \delta t]) = P(0, [0, t])P(0, [t, t + \delta t])$

$P(0, [t, t + \delta t]) \approx 1 - P(1, [t, t + \delta t]) = 1 - r\delta t$

Equivalently:

$P(t_1 > t + \delta t) = P(t_1 > t)(1 - r\delta t)$

Neglect the probability that two or more occur, $O(\delta t^2)$

Independent events
Exponential distribution

\[
P(t_1 > t + \delta t) - P(t_1 > t) = -rP(t_1 > t)
\]

\[
P(t_1 > t) = e^{-rt}
\]

\[
P(t_1 < t) = 1 - P(t_1 > t) = 1 - e^{-rt}
\]

\[
P(t) = \frac{P(t_1 < t + \delta t)}{\delta t} = \frac{dP(t_1 < t)}{dt}
\]

\[
P(t) = \frac{d(P(t_1 < t))}{dt} = \frac{d(1 - e^{-rt})}{dt}
\]

\[
P(t) = re^{-rt}
\]

\[
r = \langle \frac{\Delta n}{\Delta t} \rangle
\]

Cumulative
Poisson distribution

- Probability to have \( n \) entries in \( x \)
  - Expect on average \( \nu = N x / X = r x \)
  - Binomial, where \( p = x / X = \nu / N \ll 1 \)

\[
P(n; \nu, N) = \frac{N!}{n!(N-n)!} \left( \frac{\nu}{N} \right)^n \left( 1 - \frac{\nu}{N} \right)^{N-n}
\]

\[
= \frac{\nu^n}{n!} \frac{N(N-1) \cdots (N-n+1)}{N^n} \left( 1 - \frac{\nu}{N} \right)^N \left( 1 - \frac{\nu}{N} \right)^{-n}
\]

\[
\lim_{N \to \infty} \frac{1}{n!} \frac{N(N-1) \cdots (N-n+1)}{N^n} \left( 1 - \frac{\nu}{N} \right)^N \left( 1 - \frac{\nu}{N} \right)^{-n} = \frac{\nu^n}{n!} e^{-\nu}
\]

Siméon-Denis Poisson (1781-1840)
• For large $\nu$ a Gaussian approximation is sufficiently accurate
Summing Poissonian variables

- Probability distribution of the sum of two Poissonian variables with expected values $\nu_1$ and $\nu_2$:
  \[ P(n) = \sum_{m=0}^{n} \text{Pois}(m; \nu_1) \text{Pois}(n-m; \nu_2) \]

- The result is still a Poissonian:
  \[ P(n) = \text{Pois}(n; \nu_1 + \nu_2) \]

- Useful when combining Poissonian signal plus background:
  \[ P(n; s, b) = \text{Pois}(n; s + b) \]

- The same holds for ‘convolution’ of binomial and Poissonian:
  - Take a fraction of Poissonian events with binomial ‘efficiency’

- No surprise, given how we constructed Poissonian probability!
Demonstration: Poisson $\otimes$ Binomial

$$P(s_0; \mu) = \frac{e^{-\mu} \mu^{s_0}}{s_0!} \otimes B(s; s_0, \varepsilon) = \frac{s_0!}{s!(s_0 - s)!} \varepsilon^s (1 - \varepsilon)^{s_0 - s}$$

$$P(s; \mu, \varepsilon) = \sum_{s_0=s}^{\infty} P(s_0; \mu) B(s; s_0, \varepsilon)$$

$$= \sum_{s_0=s}^{\infty} \frac{e^{-\mu} \mu^{s_0}}{s_0!} \cdot \frac{s_0!}{s!(s_0 - s)!} \varepsilon^s (1 - \varepsilon)^{s_0 - s}$$

$$= \frac{e^{-\mu} (\varepsilon \mu)^s}{s!} \sum_{s_0=s}^{\infty} \frac{\mu^{s_0 - s} (1 - \varepsilon)^{s_0 - s}}{(s_0 - s)!}$$

$$= \frac{e^{-\mu} (\varepsilon \mu)^s e^\mu e^{-\varepsilon \mu}}{s!} = \frac{e^{-\varepsilon \mu} (\varepsilon \mu)^s}{s!} = P(s; \varepsilon \mu)$$
Other frequently used PDFs

- Argus function
- Crystal ball distribution
- Landau distribution
Argus function

- Mainly used to model background in mass peak distributions that exhibit a kinematic boundary

\[ A(x; \theta, \xi) = x \sqrt{1 - \left( \frac{x}{\theta} \right)^2} e^{\xi \left[ 1 - \left( \frac{x}{\theta} \right)^2 \right]} \]

- The primitive can be computed in terms of error functions, so the numerical normalization within a given range is feasible
Argus primitive

- For the records:
  - For $\xi<0$:
    \[
    \int A(x; \theta, \xi)dx = \frac{\theta^2}{2} \left\{ \xi^{-1} e^{\xi \left[1 - \left(\frac{x}{\theta}\right)^2\right]} + (-\xi)^{-\frac{3}{2}} \text{erf} \sqrt{-\xi \left[1 - \left(\frac{x}{\theta}\right)^2\right]} \right\}
    \]
  - For $\xi \geq 0$:
    \[
    \int A(x; \theta, \xi)dx = \frac{\theta^2}{2} \left\{ \xi^{-1} e^{\xi \left[1 - \left(\frac{x}{\theta}\right)^2\right]} - \xi^{-\frac{3}{2}} \text{erfi} \sqrt{\xi \left[1 - \left(\frac{x}{\theta}\right)^2\right]} \right\}
    \]
- But please, verify with a symbolic integrator before using my formulae 😊!
Crystal Ball function

- Adds an asymmetric power-law tail to a Gaussian PDF with proper normalization and continuity of PDF and its derivative

\[ CB(x; \alpha, n, \bar{x}, \sigma) = N \cdot \begin{cases} 
\exp\left(-\frac{(x-\bar{x})^2}{2\sigma^2}\right), & \text{for } \frac{x-\bar{x}}{\sigma} > -\alpha \\
A \cdot (B - \frac{x-\bar{x}}{\sigma})^{-n}, & \text{for } \frac{x-\bar{x}}{\sigma} \leq -\alpha 
\end{cases} \]

\[ A = \left( \frac{n}{|\alpha|} \right)^n e^{-\frac{\alpha^2}{2}} \quad B = \frac{n}{|\alpha|} - |\alpha| \]

- Used first by the Crystal Ball collaboration at SLAC

\[ \alpha = 0.1 \]
\[ \alpha = 1 \]
\[ \alpha = 10 \]
\[ \bar{x} = 0, \sigma = 1, n = 1 \]
Landau distribution

- Used to model the fluctuations in the energy loss of particles in thin layers

\[ L(x) = \frac{1}{\pi} \int_0^\infty e^{-t \log t - xt} \sin(\pi t) \, dt. \]

- More frequently, scaled and shifted:

\[ L(x; \mu, \sigma) = L \left( \frac{x - \mu}{\sigma} \right) \]

- Implementation provided by GNU Scientific Library (GSL) and ROOT (`TMath::Landau`)
PDFs in more dimensions
Multi-dimensional PDF

\[ \frac{d^2 P}{dx dy} = f(x, y) \]

- 1D projections (marginal distributions):
  \[ f_x(x) = \int f(x, y) dy \]
  \[ f_y(y) = \int f(x, y) dx \]

- \( x \) and \( y \) are independent if:
  \[ f(x, y) = f_x(x)f_y(y) \]

- We saw that \( A \) and \( B \) are independent events if:
  \[ P(A \cap B) = P(A)P(B) \]
Conditional distributions

- PDF w.r.t. $y$, given $x = x_0$
- PDF should be projected and normalized with the given condition

$$f(y|x_0) = \frac{f(x_0, y)}{\int f(x_0, y) \, dy}$$

- Remind:
  - $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$
Covariance and cov. matrix

• Definitions:
  – Covariance: \( \text{cov}(x, y) = \langle xy \rangle - \langle x \rangle \langle y \rangle \).
  – Correlation: \( \rho_{ij} = \frac{\text{cov}(x_i, x_j)}{\sigma_{x_i} \sigma_{x_j}} \), \(-1 \leq \rho_{ij} \leq 1\).

• Correlated \( n \)-dimensional Gaussian:
  \[
  f(x_1, \cdots, x_n) = \frac{1}{(2\pi)^{n/2} |C|^{1/2}} \exp \left( -\frac{1}{2} (x_i - \mu_i) C_{ij}^{-1} (x_j - \mu_j) \right)
  \]

• where: \( \text{cov}(x_i, x_j) = C_{i,j} \).
Two-dimensional Gaussian

- Product of two independent Gaussians with different $\sigma$
- Rotation in the $(x, y)$ plane

$$g(x', y') = \frac{1}{2\pi\sigma_{x'}\sigma_{y'}} \exp \left( -\frac{1}{2} \left( \frac{x'^2}{\sigma_{x'}^2} + \frac{y'^2}{\sigma_{y'}^2} \right) \right)$$

$$\begin{align*}
x' &= x \cos \varphi - y \sin \varphi \\
y' &= x \sin \varphi + y \cos \varphi
\end{align*}$$

$$\frac{x'^2}{\sigma_{x'}^2} + \frac{y'^2}{\sigma_{y'}^2} = (x, y)^T C^{-1} (x, y)$$

$$C^{-1} = \begin{pmatrix}
\frac{\cos^2 \varphi}{\sigma_{x'}^2} + \frac{\sin^2 \varphi}{\sigma_{y'}^2} & \sin \varphi \cos \varphi \left( \frac{1}{\sigma_{y'}^2} - \frac{1}{\sigma_{x'}^2} \right) \\
\sin \varphi \cos \varphi \left( \frac{1}{\sigma_{y'}^2} - \frac{1}{\sigma_{x'}^2} \right) & \frac{\sin^2 \varphi}{\sigma_{x'}^2} + \frac{\cos^2 \varphi}{\sigma_{y'}^2}
\end{pmatrix}$$
Two-dimensional Gaussian (cont.)

- Rotation preserves the metrics:

\[
|C^{-1}| = \frac{1}{\sigma_{x'}^2 \sigma_{y'}^2} = \frac{1}{\sigma_x^2 \sigma_y^2 (1 - \rho_{xy}^2)}
\]

- Covariance in rotated coordinates:

\[
C = \begin{pmatrix}
\cos^2 \varphi \sigma_{x'}^2 + \sin^2 \varphi \sigma_{y'}^2 & \sin \varphi \cos \varphi (\sigma_{y'}^2 - \sigma_{x'}^2) \\
\sin \varphi \cos \varphi (\sigma_{y'}^2 - \sigma_{x'}^2) & \sin^2 \varphi \sigma_{x'}^2 + \cos^2 \varphi \sigma_{y'}^2
\end{pmatrix}
\]

\[
\begin{align*}
\sigma_x^2 &= \cos^2 \varphi \sigma_{x'}^2 + \sin^2 \varphi \sigma_{y'}^2 \\
\sigma_y^2 &= \sin^2 \varphi \sigma_{x'}^2 + \cos^2 \varphi \sigma_{y'}^2
\end{align*}
\]

\[
\rho_{xy} = \frac{\text{cov}_{xy}}{\sigma_x \sigma_y} = \frac{\sin 2\varphi (\sigma_{y'}^2 - \sigma_{x'}^2)}{\sqrt{\sin 2\varphi (\sigma_{x'}^4 + \sigma_{y'}^4) + 2\sigma_{x'}^2 \sigma_{y'}^2}}
\]
Two-dimensional Gaussian (cont.)

- A pictorial view of an iso-probability contour

\[
\tan 2\varphi = \frac{2\rho_{xy}\sigma_x\sigma_y}{\sigma_y^2 - \sigma_x^2}
\]

\[
f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1 - \rho_{xy}^2}} \exp \left( -\frac{1}{2(1 - \rho_{xy}^2)} \left( \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} - \frac{2xy\rho_{xy}}{\sigma_x\sigma_y} \right) \right)
\]
1D projections

- PDF projections are (1D) Gaussians:
- Areas of $1\sigma$ and $2\sigma$ contours differ in 1D and 2D!

$$f_x(x) = \int_{-\infty}^{\infty} f(x, y) dy = \frac{1}{\sqrt{2\pi\sigma_x}} e^{-\frac{x^2}{2\sigma_x^2}}$$

$$P_{1D}(n\sigma) = \int_0^n e^{-\frac{x^2}{2}} dx = \text{erf}\left(\frac{n}{\sqrt{2}}\right)$$

$$P_{2D}(n\sigma) = \int_0^n e^{-\frac{r^2}{2}} rdr = 1 - e^{-\frac{n^2}{2}}$$

<table>
<thead>
<tr>
<th></th>
<th>$P_{1D}$</th>
<th>$P_{2D}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1\sigma$</td>
<td>0.6827</td>
<td>0.3934</td>
</tr>
<tr>
<td>$2\sigma$</td>
<td>0.9545</td>
<td>0.8647</td>
</tr>
<tr>
<td>$3\sigma$</td>
<td>0.9973</td>
<td>0.9889</td>
</tr>
<tr>
<td>$1.515\sigma$</td>
<td>0.6827</td>
<td></td>
</tr>
<tr>
<td>$2.486\sigma$</td>
<td>0.9545</td>
<td></td>
</tr>
<tr>
<td>$3.439\sigma$</td>
<td>0.9973</td>
<td></td>
</tr>
</tbody>
</table>
Correlation and independence

- Independent variables are uncorrelated
- But not necessarily vice-versa

\[
f(x, y) = \frac{1}{4} (g(x; r, 1)g(y; 0, 1) + g(x; -r, 1)g(y; 0, 1) + g(x; 0, 1)g(y; r, 1) + g(x; 0, 1)g(y; -r, 1))
\]

\[
\langle x \rangle = \langle y \rangle = 0
\]

\[
\langle x^2 \rangle = \langle y^2 \rangle = 1 + \frac{r^2}{2}
\]

\[
\langle xy \rangle = 0
\]

Uncorrelated, but not independent!
PDF convolution

- Concrete example: add experimental resolution to a known PDF
- The intrinsic PDF of the variable $x_0$ is $f(x_0)$
- Given a true value $x_0$, the probability to measure $x$ is:
  - $r(x, x_0)$
  - May depend on other parameters (e.g.: $\sigma =$ experimental resolution, if $r$ is a Gaussian)
- The probability to measure $x$ considering both the intrinsic fluctuation and experimental resolution is the convolution of $f$ with $r$:
  \[
  g(x) = \int f(x')r(x, x')dx'
  \]
- Often referred to as: $g = f \otimes r$
Reminder of Fourier transform definition:

\[ \hat{f}(k) = \int_{-\infty}^{+\infty} f(x) e^{-ikx} \, dx \quad \leftrightarrow \quad f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(k) e^{ikx} \, dk \]

It can be demonstrated that:

\[ \hat{f} \otimes \hat{g} = \hat{f} \hat{g} \quad \hat{f} g = \hat{f} \otimes \hat{g} \]

In particular, the FT of a Gaussian is still a Gaussian:

\[ \hat{g}(k) = e^{-i k \mu} e^{-\frac{\sigma^2 k^2}{2}} \]

Note: \( \sigma \) goes to the numerator!

Numerically, FFT can be convenient for computation of convolution PDF’s (\( \rightarrow \) RooFit)
A small digression...

Application to economics
Familiar example: multiple scattering

- Assume the limit of small scattering angles
  - Can add single random scattering angles
    \[ \theta = \sum_i \theta_i \]
- For many steps, the distribution of \( \theta \) can be approximated with a Gaussian
  - \( \langle \theta \rangle = 0 \)
  - \( \sigma_{\theta}^2 = \sum_i \delta^2 = N \delta^2 = \delta^2 x / \Delta x \)
- Hence:
  - \( \sigma_{\theta} \propto \sqrt{x} \)
- This is similar to a Brownian motion, where in general (as a function of time):
  - \( \sigma(t) \propto \sqrt{t} = \sigma \sqrt{t} \)

More precisely (from PDG):

\[ \theta_0 = \frac{13.6 \text{ MeV}}{\beta c p} z \sqrt{x/X_0} \left[ 1 + 0.038 \ln(x/X_0) \right] \]
Stock prices \textit{vs} bond prices

- Stock prices are represented as \textit{geometric Brownian motion}:
  \[ s(t) = s_0 e^{y(t)} \]
  - Where
    \[ y(t) = \mu t + \sigma n(t) \]
- Stock \textit{volatility}: $\sigma$
- Bond growth (risk-free and deterministic):
  \[ b(t) = b_0 e^{\alpha t} \]
- Discount rate: $\alpha$
Average stock option at time $t$

- Average computed as usual:
  \[
  \langle s(t) \rangle = \langle s_0 e^{\mu t} e^{\sigma n(t)} \rangle = s_0 e^{\mu t} \langle e^{\sigma n(t)} \rangle
  \]

- It's easy to demonstrate that, if $w$ is Gaussian:
  \[
  \langle e^w \rangle = e^{\langle w \rangle + \text{Var}(w)/2}
  \]

- The Brownian variance is $\sigma^2 t$, hence:
  \[
  \langle s(t) \rangle = s_0 e^{(\mu + \sigma^2/2) t}
  \]

- The **risk-neutral** price for the stock must be such that it produces no gain w.r.t. bonds:
  \[
  \langle s(t) \rangle = b(t) \implies \mu + \sigma^2/2 = \alpha
  \]
Stock options

- A call-option is the right, buy not obligation, to buy one share at price \( K \) at time \( t \) in the future

- Gain at time \( t \):
  - \( g = s(t) - K \) if \( K < s(t) \)
  - \( g = 0 \) if \( K \geq s(t) \)

- What is the risk-neutral price of the option?
Black-Scholes model

- The **average gain** minus cost $c$ must be equal to bond gain:
  
  $$- c e^{\alpha t} = \langle g \rangle$$

- Equivalently:
  
  $$ce^{\alpha t} = \int_{\log(K/s_0)}^{\infty} (s_0 e^y - K) \frac{1}{\sqrt{2\pi t\sigma^2}} e^{-(y-\mu t)^2/2t\sigma^2} dy$$

- Which gives:
  
  $$c = s_0 \phi(\sigma \sqrt{t} + b) - Ke^{-\alpha t} \phi(b)$$

- Where:
  
  $$b = \frac{\alpha t - \sigma^2 t/2 - \log(K/s_0)}{\sigma \sqrt{t}}$$

- $\phi$ is the cumulative distribution of a normal Gaussian
Limit for $t \rightarrow 0$

- For current time, the price is what you would expect, since there is no fluctuation

$$\lim_{t \rightarrow 0} c(t) = (s_0 - K) \phi\left(-\frac{\ln(K/s_0)}{\sigma \sqrt{t}}\right)$$

$$= \begin{cases} 
  s_0 - K, & \text{if } K < s_0 \\
  0, & \text{if } K \geq s_0 
\end{cases}$$

- At fixed (larger) $t$, the price curve gets ‘smoothed’ by the Gaussian fluctuations
‘Sensitivity’ to stock price and time

\[ \Delta = \frac{\partial c}{\partial s_0} \]

Call Delta versus time to expiration and initial stock price. Strike=25
Black and Scholes

- Black had a PhD in applied Mathematics. Died in 1995
- Scholes won the Nobel price in Economics on 1997
- He was co-funder of the hedge-fund “Long-Term Capital Management”
- After gaining around 40% for the first years, it lost in 1998 $4.6 billion in less than four months and failed after the East Asian financial crisis
The End

Nobody’s perfect!