MICRO-428: Metrology

Edoardo Charbon, Claudio Bruschini, Georg Fantner

Institute of Electrical and Micro Engineering (IEM), School of Engineering (STI)

Ecole polytechnique fédérale de Lausanne (EPFL), Neuchâtel, Switzerland

TAs: Samuele Bisi, Alexandre Domenech

EPFL

MICRO-428: Metrology

Week Eight: Elements of Statistics

Claudio Bruschini

TA: Samuele Bisi (2021: Simone Frasca)

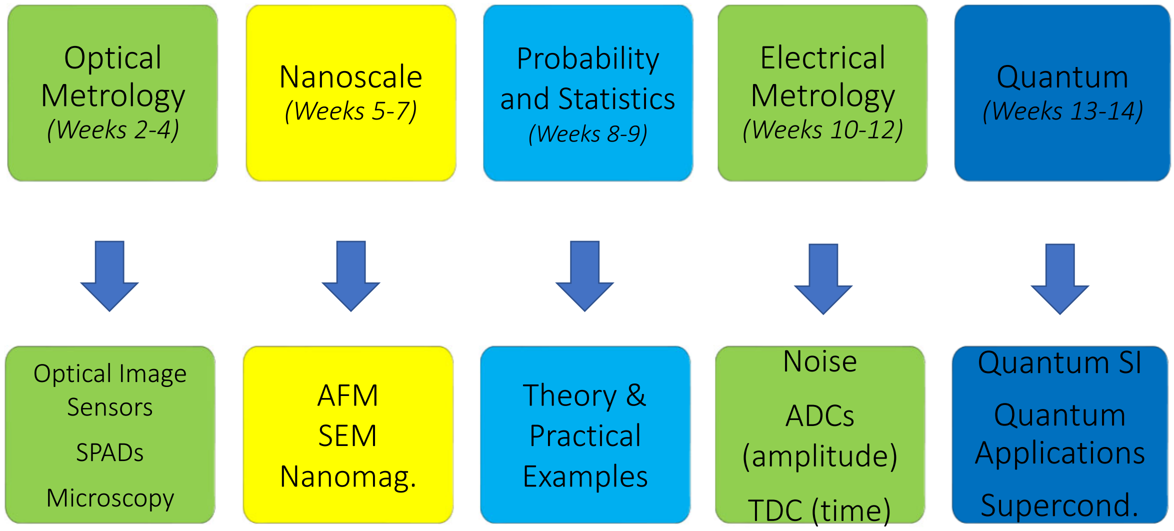
Advanced Quantum Architecture Laboratory (AQUA)

EPFL at Microcity, Neuchâtel, Switzerland

EPFL

Metrology Course Structure

Measurement Science & Technology



Reference Books (Weeks 8&9)

- J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015
- A. Papoulis, Probability, *Random Variables and Stochastic Processes*, 3rd ed., 1991
- S.M. Ross, *Introduction to Probability Models*, 10th ed., 2009
- I.G. Hughes, T.P.A. Hase, *Measurements and their Uncertainties*, 1st ed., 2010
- G.E.P. Box, J.S. Hunter, W.G. Hunter, *Statistics for Experimenters*, 2nd ed., 2005
- J.R. Taylor, *An Introduction to Error Analysis*, 2nd ed., 1997

The first reference, by Blitzstein, was used extensively throughout this lecture as well as the following one. It should still be available from the EPFL library and is a suggested read for these topics.

NB: in general, see also the reference box at the bottom of the slides for notes on the exact chapters, etc.

Outline

The Outline covers both this lecture as well as the next one.

8.1 Introduction to Probability

- The theory of probability deals with averages of mass phenomena occurring sequentially or simultaneously.

- If an experiment is performed n times and the event $\mathcal{A}$ occurs $n_{\mathcal{A}}$ times, and if n is sufficiently large, it is possible to state that the relative frequency $n_{\mathcal{A}}/n$ of occurrence of $\mathcal{A}$ is close to the probability $P\{\mathcal{A}\}$ that the event $\mathcal{A}$ occurs:

$$P\{\mathcal{A}\} \approx n_{\mathcal{A}}/n$$

A. Papoulis, *Probability, Random Variables and Stochastic Processes*, 3rd ed., 1991, Chap. 1

In order to move quickly to practical examples, this part has been substantially abridged and is now available in the Appendix. Please refer to it for the exact definitions.

8.1 Introduction to Probability

- Further formal details in Appendix 8.1 (A8.1)
  - Fair dice example
  - How a probability function maps events to numbers
  - Conditional Probability
  - Bayes' rule & law of total probability (LOTP)
  - Independence of Events

In order to move quickly to practical examples, this part has been substantially abridged and is now available in the Appendix. Please refer to it for the exact definitions.

Outline

8.2 Random Variables

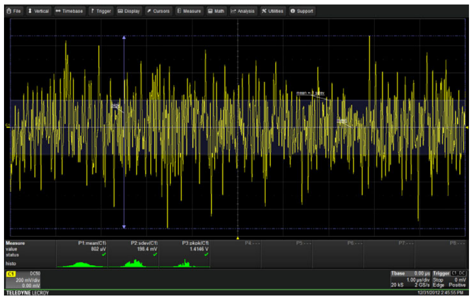
- A Random Variable (RV) is a number $X(s)$ assigned to every outcome s of an experiment.

  Examples: the voltage of a random source, etc..

- The domain of the Random Variable $X(s)$ is $\mathcal{S}$, which is the set of experimental outcomes. It is also called the support of the random variable. Its range is $\mathbb{R}$. Two properties must be satisfied:

  1. The set $\{X(s) \leq x\}$ is an event for every x.
  2. The probabilities of the events $\{X = \infty\}$ and $\{X = -\infty\}$ must be zero:

  $$P\{X = \infty\} = P\{X = -\infty\} = 0.$$



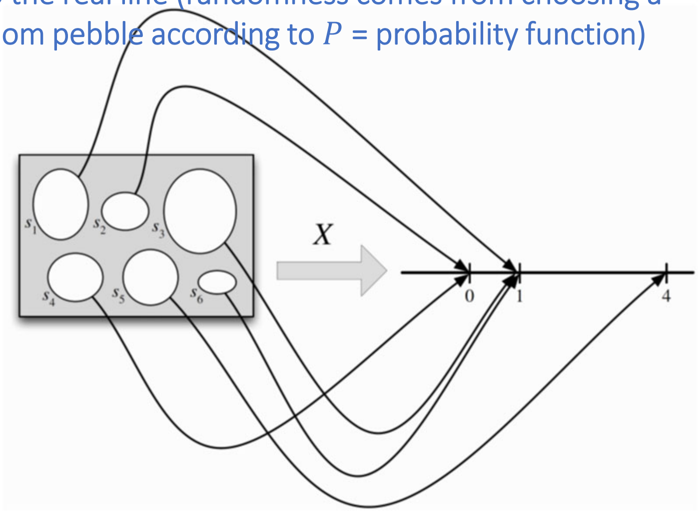
A. Papoulis, *Probability, Random Variables and Stochastic Processes*, 3rd ed., 1991, Chap. 4.1

$\mathcal{S}$ = support is the set of experimental outcomes.

8.2 Random Variables (contd.) – Example

S

Example of random variable mapping X from the sample space S into the real line (randomness comes from choosing a random pebble according to P = probability function)



J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 3.1

The blue “S” on the top right means that this slide is skipped during the lecture, but left in the document to maintain the overall coherence of the material.

8.2 Random Variables (contd.)

- A Random Variable X is said to be discrete if there is a finite list of values $a_1, a_2, \dots, a_n$ or an infinite list of values $a_1, a_2, \dots$ such that $P\{X = a_j \text{ for some } j\} = 1$. In the first case, its support is given by:

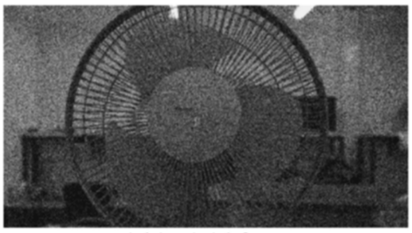
$$\mathcal{S} = \{a_1, a_2, \dots, a_n\}$$

  Example: the outcome from the launch of a dice; the number of photons detected in an image.

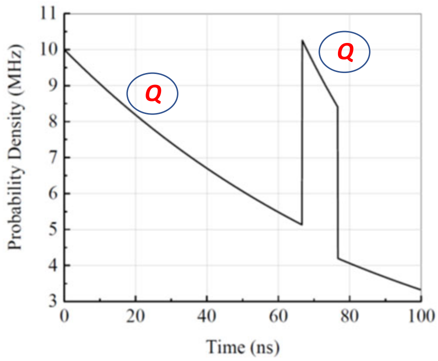
- A Random Variable X is instead said to be continuous if it can take on any value in a given interval, possibly of infinite length. For example its support can be:

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

  Example: time of arrival of a photon in a LiDAR image.



4-bit, 4.4 kfps



SPAD-Based LiDAR first photon PDF

J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 3.2

A. Ulku *et al., A 512x512 SPAD Image Sensor with Integrated Gating for Widefield FLIM.* IEEE JSTQE (2019).
M. Beer *et al., Background Light Rejection..., MDPI Sensors 18, 2018.*

The red “Q” stands for a question asked during class, aimed at going deeper into the subject.

Hints: think of a) outdoors operation (and a first-photon detection set-up, i.e. only the first backscattered photon is detected in a laser period), and b) an illumination with a pulsed laser emitting pulses of finite temporal length…

8.2.1 Probability Mass Functions

How to express the distribution of a (discrete) Random Variable/1

- The probability mass function (PMF) of a discrete RV X is the function:

$$\text{PMF: } p_X(x) = P\{X = x\}$$

Note that this value is positive if $x \in \mathcal{S}$, zero otherwise.

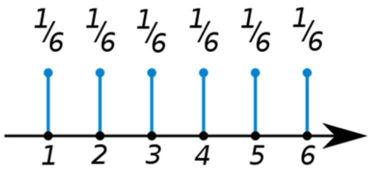
- The PMF needs to satisfy two criteria:

1. Nonnegative:

$p_X(x) > 0$ if $x = x_j$ for some j,
$p_X(x) = 0$ otherwise.

2. Sums to 1:

$$\sum_{j=1}^{\infty} p_X(x_j) = 1$$



J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 3.2

We start to look into the formal definitions for discrete random variables first, before moving to continuous ones.

8.2.1 Probability Mass Functions (contd.) – Example

Example

Imagine to toss two coins at the same time. The possible outcomes are, given that H = head and T = tail, the following: $\mathcal{S} = \{HH, HT, TH, TT\}$. If the Random Variable X is the number of heads, it follows that:

$$p_X(0) = P\{X = 0\} = 1/4$$

$$p_X(1) = P\{X = 1\} = 2/4$$

$$p_X(2) = P\{X = 2\} = 1/4$$

$$p_X(x) = P\{X = x\} = 0 \text{ for all other } x$$

J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 3.2

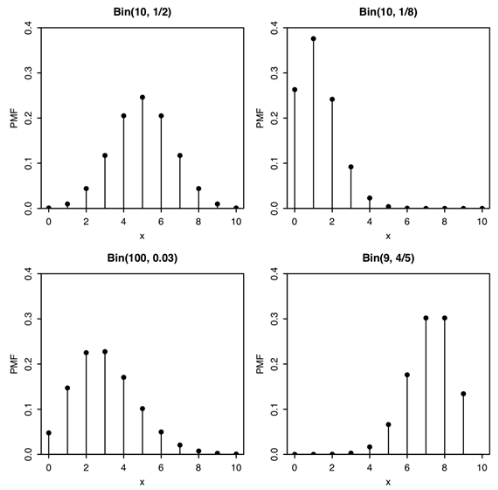
8.2.2 Bernoulli and Binomial RVs

First case: a Random Variable which can only take two values

- A discrete RV X is said to have the Bernoulli distribution with parameter p if $P\{X = 1\} = p$ and $P\{X = 0\} = 1 - p$, where $0 < p < 1$.
- An experiment that can result in either a success or a failure is called a Bernoulli trial.
- Suppose that n *independent* Bernoulli trials are performed. Let p be the probability of success, $1 - p$ the probability of failure, X (RV) the number of successes. The distribution of X is called binomial distribution $\text{Bin}(n, p)$ with parameters n and p:

$$\text{PMF: } P\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k}$$

for $k = 0,1, \dots, n$



J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 3.3

We will now look at some of the most important random variable distributions, illustrated by means of examples from engineering and physics.

NB: the binomial coefficient $\binom{n}{k}$ reads "n choose k".

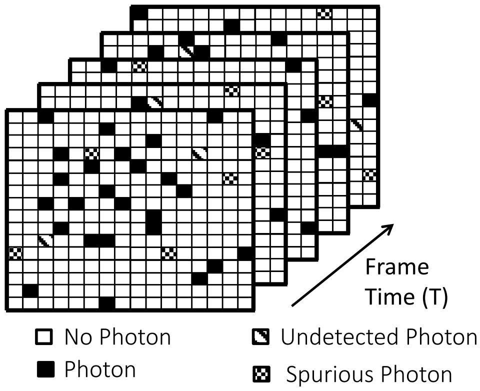
8.2.2 Bernoulli RV – Example

(CMOS) SPAD: Single-Photon Avalanche Photodiode

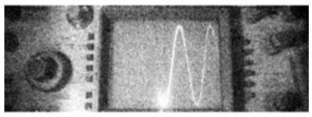
+ time-of-arrival, energy/wavelength, polarization, etc.

Perfect single photon detection limited by

1. Photon detection efficiency (PDE) = QE x FF
2. Temporal Aperture Ratio
3. Dark Count Rate



1-bit frame



4-bit frame

R. Henderson, Edinburgh Univ., ISSCC 2013 – E. Fossum, IISW 2013

A. Chandramouli, *A bit too much? High Speed Imaging from Sparse Photon Counts*, Proc. ICCP 2019

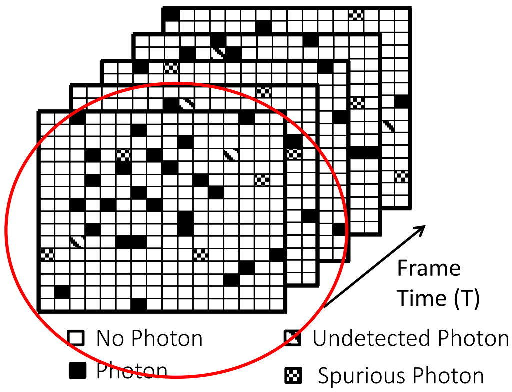
A binary SPAD sensor represent an excellent example to illustrate the properties of Bernoulli and Binomial random variables.

A CMOS SPAD is seen here a source of individual photon detections -> Bernoulli RV. A SPAD array extends this concept to a large number of pixels, whose output is organised in frames as a function of time. Binary SPAD sensors, such as SwissSPAD2, can be read out extremely fast, up to 100 kfps. The detection efficiency is however not perfect, and the sensor is a source of noise (spurious counts) as well.

Individual binary frames can be accumulated (on FPGA or PC) in multi-bit frames → bottom right image.

8.2.2 Bernoulli RV – Example

(CMOS) SPAD:
Single-Photon
Avalanche
Photodiode



ϕ=photon flux $\left({}^{\text{ph}}/_{\text{s}}\right)$, τ=exposure time,
η=quantum efficiency, r=Dark Count Rate (DCR)

Q # of photons at each pixel: $P\{Z = k\} = \dfrac{e^{-\phi\tau\eta}(\phi\tau\eta)^k}{k!} \Longrightarrow$ **Q**

$$P\{B = 0\} = e^{-(\phi\tau\eta + r\tau)}$$
$$P\{B = 1\} = 1 - e^{-(\phi\tau\eta + r\tau)}$$

Ex

R. Henderson, Edinburgh Univ., ISSCC 2013 – E. Fossum, IISW 2013 – S. Ma, *Quanta Burst Photography*, ACM Trans. Graph., Vol. 39, 2020

We look here at the distribution of the number of detected photons in a single frame. $\phi\tau\eta$ Is the average number of detected photons per frame of exposure time τ.

Q: where does this statistical distribution - $P\{Z = k\}$ - come from?

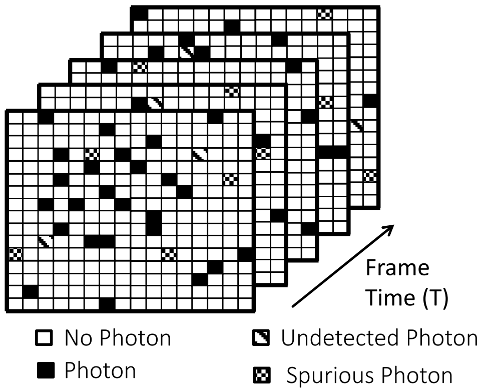
→ *P{Z = k}* is nothing but a Poisson distribution, with mean = lambda = $\phi\tau\eta$ (see 8.2.5). Actually, B = 1 means in this case at least one photon – a purely binary sensor cannot count more than one time. (NB: multi-bit in-pixel architectures are possible.)

The two bottom formulas are more complete ones and take into account the noise contribution as well.

The green "Ex" highlights an exercise or homework which deals with the topic(s) shown in class.

8.2.2 Binomial RV – Example

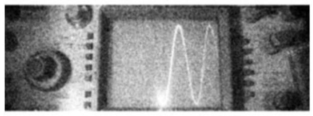
(CMOS) SPAD:
Single-Photon
Avalanche
Photodiode



# of photons k at each pixel for n consecutive (independent) frames:

$$P\{n,k\} = \frac{n!}{(n-k)!\cdot k!}\cdot p_{ph}^{k}\cdot\left(1-p_{ph}\right)^{n-k} \; where$$

$$p_{ph} = 1 - P\{1,0\} = 1 - e^{-(\phi\tau\eta + r\tau)}$$



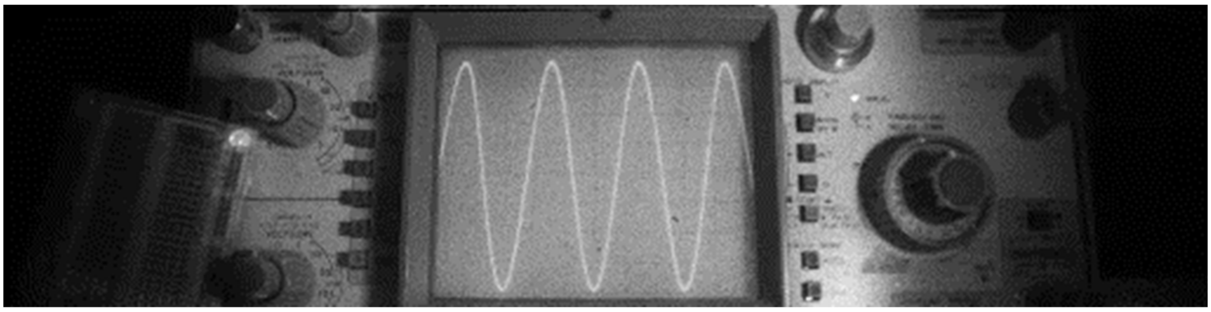
Y. Hirose, MDPI Sensors(18), 2018

The previous results can actually be easily generalised to *n* frames, leading to a practical example of a binomial distributions $P\{n,k\}$.

p_{ph}, the probability of detecting at least one photon, was derived in the previous slide.

Source: Hirose MDPI Sensors 2018.

8.2.2 Binomial RV – Example



This video was taken with the first SwissSPAD camera of 512x128 pixels. Its architecture is quite similar to one of the more advanced SwissSPAD2 sensor. The camera is looking at an analogue oscilloscope, whereby the frames have been added up initially, going back all the way to examples of individual binary frames.

The images are obviously affected by shot noise, but the object being imaged can still be distinguished! And the final images are really binary in nature – each pixel has either recorded at least a photon, or none.

8.2.3 Cumulative Distribution Functions

How to express the distribution of a Random Variable/2

- The cumulative distribution function (CDF) of a discrete RV X is the function F_X given by

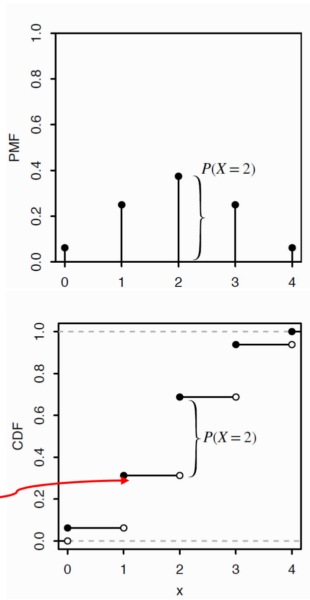
$$\text{CDF: } F_X(x) = P\{X \leq x\}$$

Example: Let X be $\text{Bin}(4, 1/2)$. The cumulative distribution function can be calculated from the probability mass function.

To find, for example, $P\{X \leq 1.5\}$, we sum the PMF over all values of the support that are less than or equal to 1.5:

$$F_X(1.5) = P\{X \leq 1.5\} = P\{X = 0\} + P\{X = 1\} =$$

$$= \binom{4}{0}\left(\frac{1}{2}\right)^4 + \binom{4}{1}\left(\frac{1}{2}\right)^4 = \left(\frac{1}{2}\right)^4 + 4\left(\frac{1}{2}\right)^4 = \frac{5}{16} = 0.3125$$



J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 3.6

8.2.3 Cumulative Distribution Functions (contd.)

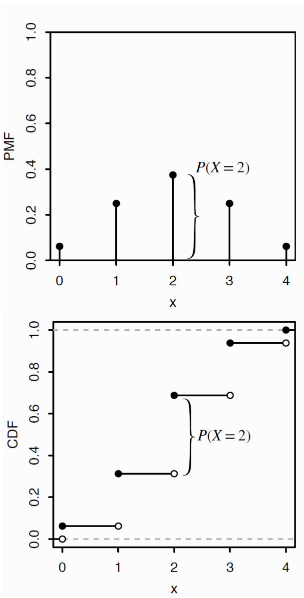
- For a CDF to be valid, the following three criteria must be met:
  1. Increasing: If $x_1 \leq x_2$, then $F_X(x_1) \leq F_X(x_2)$
  2. Right-continuous: The CDF is continuous except possibly for some jumps. When there is a jump, the CDF is continuous from the right, i.e. for any a:

$$F_X(a) = \lim_{x \to a^+} F_X(x)$$

  3. Convergence to 0 and to 1 in the limits:

$$\lim_{x \to -\infty} F_X(x) = 0$$

$$\lim_{x \to \infty} F_X(x) = 1$$



J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 3.6

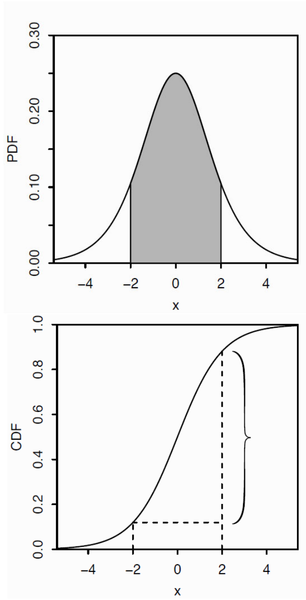
8.2.4 Probability Density Functions

- For a continuous RV X with cumulative distribution function F_X, the probability density function (PDF) f_X is the derivative of the cumulative distribution function (CDF):

$$\text{PDF: } f_X(x) = \frac{d}{dx} F_X(x)$$

hence:

$$\text{CDF: } F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$$



To get a desired probability, integrate the PDF over the appropriate range…

J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 5.1

We now move to the main definitions for a continuous random variable. Note that the knowledge of either PDF or CDF does completely characterise a statistical distribution.

NB: For a PDF f_X, the quantity $f_X(x)$ is not a probability, and in fact it is possible to have $f_X(x) > 1$ for some values of x! In order to obtain a probability, we need to integrate the PDF.

8.2.4 Probability Density Functions (contd.)

- Similarly, by definition of the CDF and the fundamental theorem of calculus:

$$P\{a < X \leq b\} = F_X(b) - F_X(a) = \int_a^b f_X(x)\, dx$$

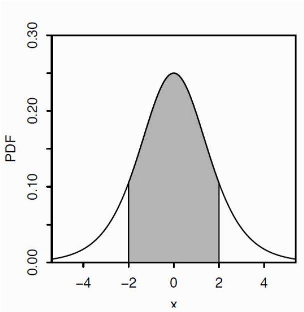
-> Probability = integral of the PDF over a given range.

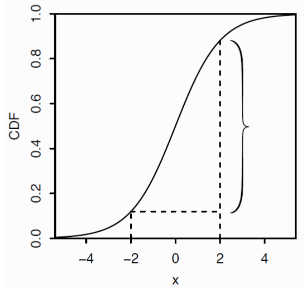
- For a PDF to be valid, two criteria must be met:
  1. Nonnegative:

$$f_X(x) \geq 0$$

  2. Integrates to 1:

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1$$

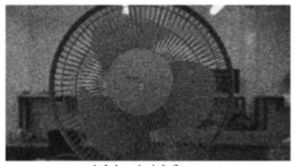




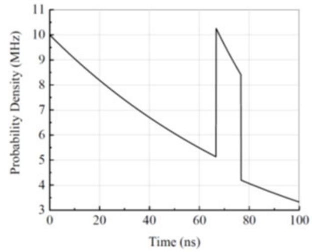
J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 5.1

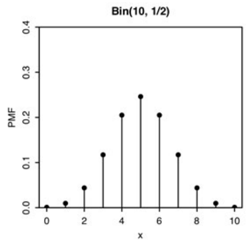
Take-home Messages/W8-1

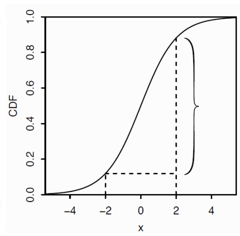
- *Introduction to probability (see also Appendix 8.1):*
  - Basic definitions, conditional probability
  - Bayes' rule, law of total probability, independence of events
- *Random Variables (RVs):*
  - Examples (discrete/continuous)
  - Probability Mass Function (PMF), Cumulative Distribution Function (CDF)
  - Probability Density Function (PDF)
  - Bernoulli, Binomial & related SPAD-based examples



4-bit, 4.4 kfps







First recap section: we summarise here the main definitions, results and examples discussed so far. They should be clear and understood.

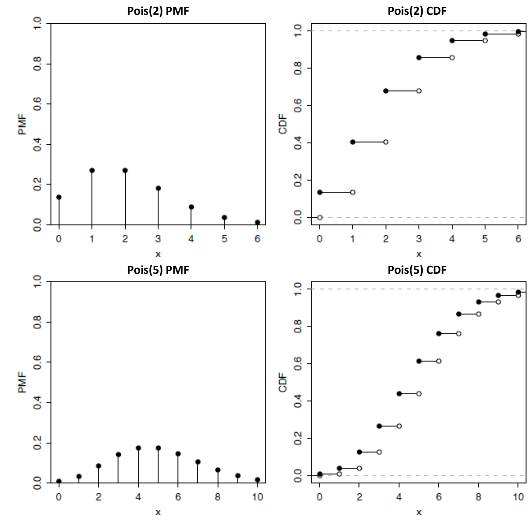
8.2.5 Poisson RV

- A discrete RV X taking on one of the values 0, 1, 2, .. is said to have a Poisson distribution with parameter λ for some $\lambda > 0$ with

$$\text{PMF: } p_X(x) = P\{X = x\} = \frac{e^{-\lambda}\lambda^x}{x!}$$

- It can be demonstrated that the Poisson PMF (we will write $X \sim \text{Pois}(\lambda)$) is a valid PMF since, by Taylor expansion:

$$\sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = e^{\lambda}$$



J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 4.7

We now look at other important statistical distributions and some of their properties. They will be analysed again in the next lecture, in more detail.

Note that a) the Poisson distribution is characterised by a single parameter (λ), and b) the verification/demonstration that the PMF is indeed a valid one. This is a priori not obvious!

8.2.6 Uniform RV

- The continuous uniform RV U on an interval (a, b) is a completely random number between a and b. Its PDF is given by:

$$\text{PDF: } f_U(x) = \begin{cases} \dfrac{1}{b-a} & \text{for } a < x < b, \\ 0 & \text{otherwise.} \end{cases}$$

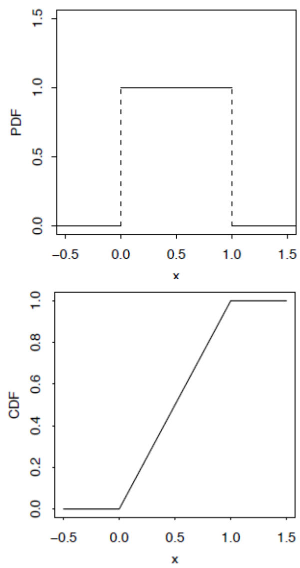
$$U \sim \text{Unif}(a, b)$$

- This is a valid PDF since the area of the PDF is given by the area of a rectangle with width $b - a$ and height $1/(b - a)$.

- Its CDF is given by:

$$F_U(x) = \begin{cases} 0 & \text{if } x \le a, \\ \dfrac{x-a}{b-a} & \text{if } a < x < b, \\ 1 & \text{if } x \ge b. \end{cases}$$

Unif(0,1) PDF & CDF



J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 5.2

Can you think of random variables with this kind of distribution?

Similarly to the previous slide, note here again the verification/demonstration that the PDF is indeed a valid one. This is a priori not obvious!

8.2.7 Normal (Gaussian) RV

- The Normal (Gaussian) distribution (we will write $X \sim \mathcal{N}(\mu, \sigma^2)$) is a famous continuous distribution that is extremely used because of the central limit theorem, which will be explained later. For the continuous Normal RV X, the PDF is:

$$\text{PDF: } f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

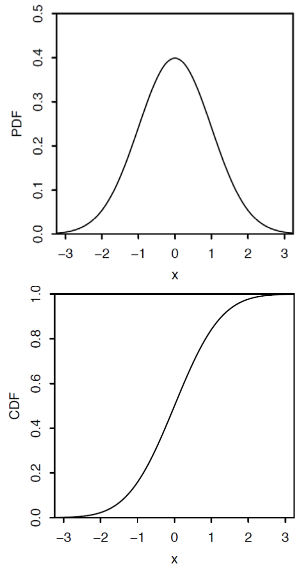
- In the special case of $\mu = 0$ and $\sigma = 1$, the distribution takes the name of standard Normal distribution. We will write it as $Z \sim \mathcal{N}(0,1)$. The standard Normal PDF and CDF are:

$$\text{PDF: } \varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

$$\text{CDF: } \Phi(z) = \int_{-\infty}^{z} \varphi(t)\, dt = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$$

No closed form exists!

Standard Normal PDF/CDF



J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 5.4

The Normal, or Gaussian, distribution is of fundamental importance. It is characterised by two parameters (μ, σ). Note the special case of the *standard* Normal distribution.

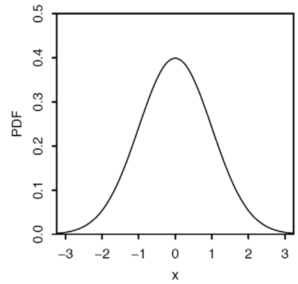
NB: pronunciation: Mu /ˈmjuː/

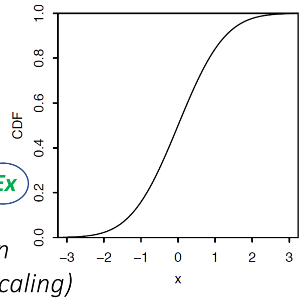
8.2.7 Normal (Gaussian) RV (contd.)

- The standard Normal distribution has the following properties:
  1. Symmetry of the PDF: φ satisfies $\varphi(z) = \varphi(-z)$
  2. Symmetry of the tail area: the area under the PDF to the left of $-z$ and to the right of z is equal. Using the CDF:
  $$\Phi(z) = 1 - \Phi(-z)$$
  3. Symmetry of Z and $-Z$: If $Z \sim \mathcal{N}(0,1)$, then $-Z \sim \mathcal{N}(0,1)$ as well.
- The Normal distribution $X \sim \mathcal{N}(\mu, \sigma^2)$ has PDF and CDF as follows:

$$\text{PDF: } f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2} = \varphi\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}$$

$$\text{CDF: } F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \quad [X = \mu + \sigma Z]$$





Ex

Location-scale transformation (shifting and scaling)

J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 5.4

The second bullet highlights the properties of a location-scale transformation (basically shifting – by μ – and scaling – by σ), which can be very useful to move from one distribution variable to another.

In this case, we can replace the RV Z, distributed according to a standard Normal, by $X = \mu + \sigma Z$, which is thus distributed according to a "general" Normal.

NB: pronunciation: Phi (/faɪ/).

8.2.8 Exponential RV

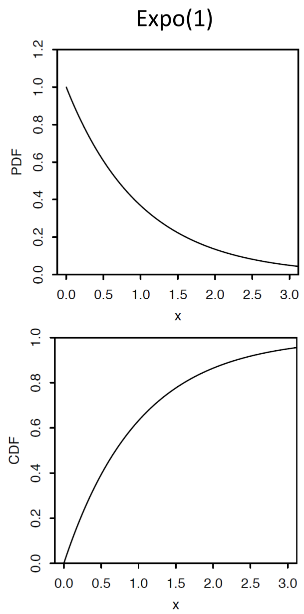
- The exponential is a distribution that represents the amount of failures before the first success (as in time), considering that λ is the success rate per unit time. The average number of successes in the time length t is λt, though the actual number of successes varies randomly.
- A continuous RV X is said to have an exponential distribution (we will write $X \sim \mathrm{Expo}(\lambda)$) with parameter λ if its PDF is:

$$\mathrm{PDF:}\ f_X(x) = \lambda e^{-\lambda x}, \qquad x > 0$$

- The corresponding CDF is

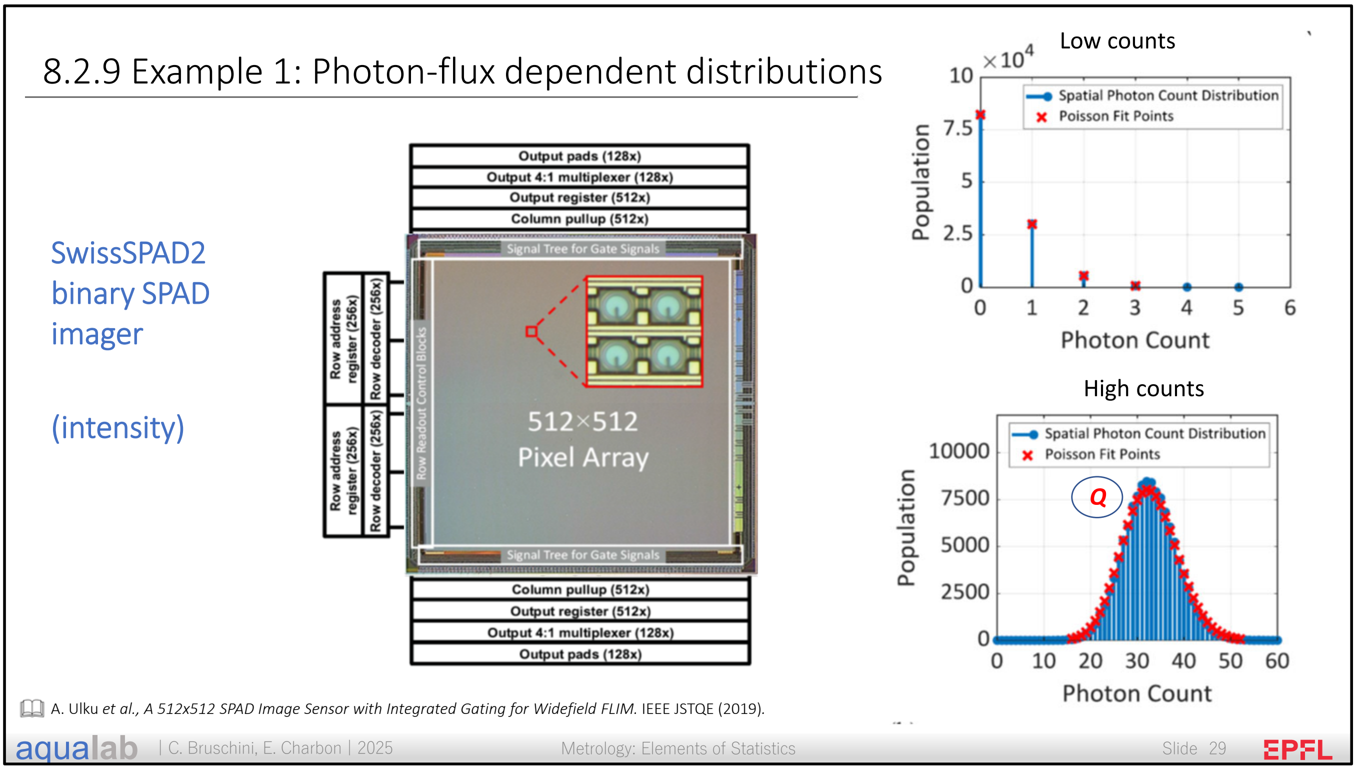
$$\mathrm{CDF:}\ F_X(x) = 1 - e^{-\lambda x}, \qquad x > 0$$

Ex



J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 5.5

Note again the presence of one single variable λ, and the interpretation in terms of success rate (e.g. events/second) and number of successes λt (e.g. events) in a given amount of time t.



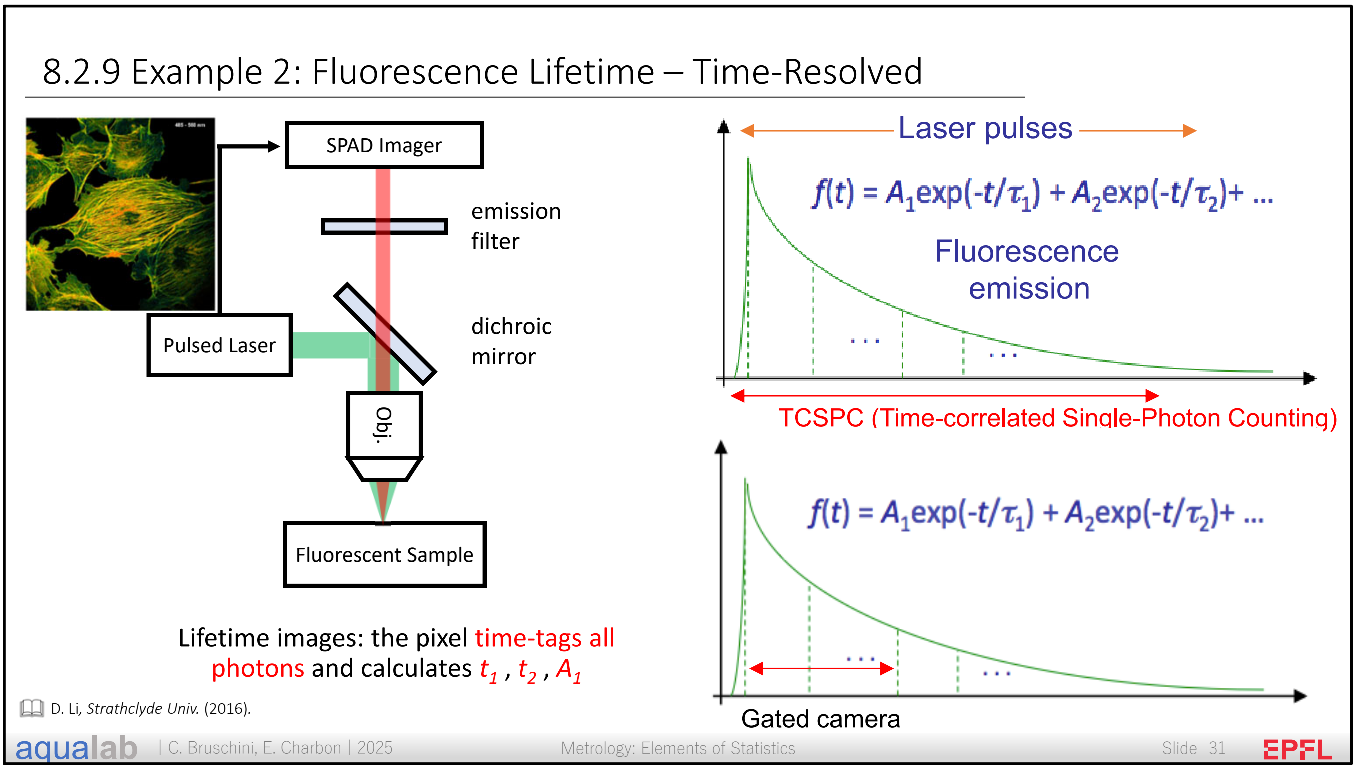
We will now look into four concrete examples from engineering and physics, linked to some of the distributions which we have seen before.

This will also allow to highlight the difference between analytical expressions and real experimental data, the presence of noise or of unaccounted phenomena when modelling a certain process.

The first example is based on binary SPAD imagers used to **measure light intensity and its distribution in time and space**, including the capability of measuring the statistical properties of the impinging photons. This class of sensors allows to output binary frames at very high speed. Further details are available from the companion paper on Moodle.

The top right plot shows a histogramme of the photon count population in eight-bit images, i.e. how many pixels fired how many times, at very low illumination levels. Note the good fit to a Poisson distribution.
Q: how are such low illumination levels obtained? What would you do to check the fit quality at higher photon counts?

The bottom right plot shows the same data at higher illumination levels.
Q: What attracts your attention in the centre of the distribution? What could this be due to?



The second example is from an application in microscopy/life sciences. When fluorophores are excited, they can re-emit light (at a slightly red-shifted wavelength = lower energy / Stokes shift), typically within a very short period, e.g. nanoseconds. The decay is exponential, or a sum of exponentials – see the *f(t)* expression. The corresponding lifetimes provide information on the fluorophore nature and its environment.

Let's suppose that the excitation is periodic, e.g. by means of a pulsed (picosecond) laser. There are fundamentally two ways of sampling the re-emitted light, **measuring its distribution over time**, to determine the fluorophore lifetime:

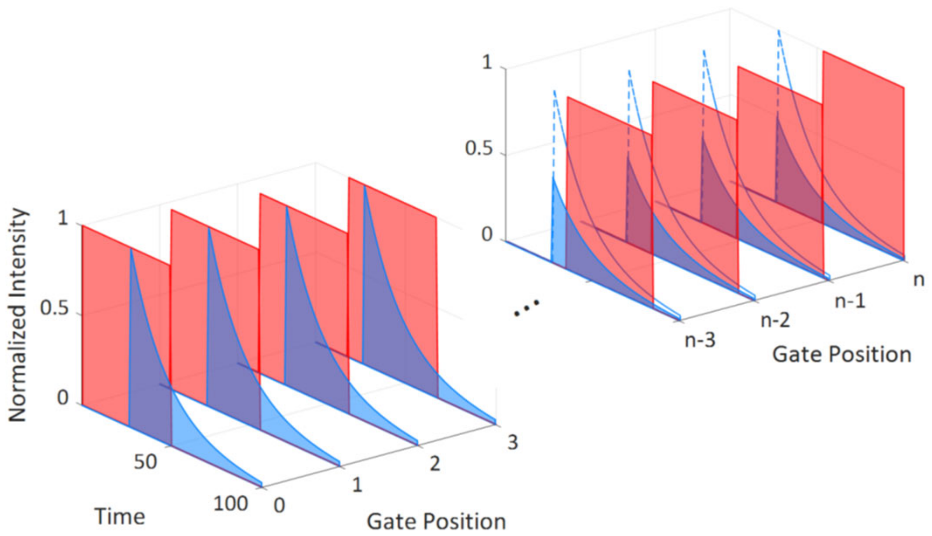
- [TCSPC] either by measuring precisely each time of arrival, using the laser emission time as reference (top), and building a histogramme, from which the lifetime(s) can be extracted by fitting a curve, or using other estimators (see next lecture),

- [Gated] or by accumulating/counting all photons emitted in a given time window (gate) and repeating the operation over several gates, which can be overlapping or not. The lifetime can then be extracted either analytically or computationally.

8.2.9 Example 2: Fluorescence Lifetime – Time-Resolved

SwissSPAD2
binary SPAD
imager

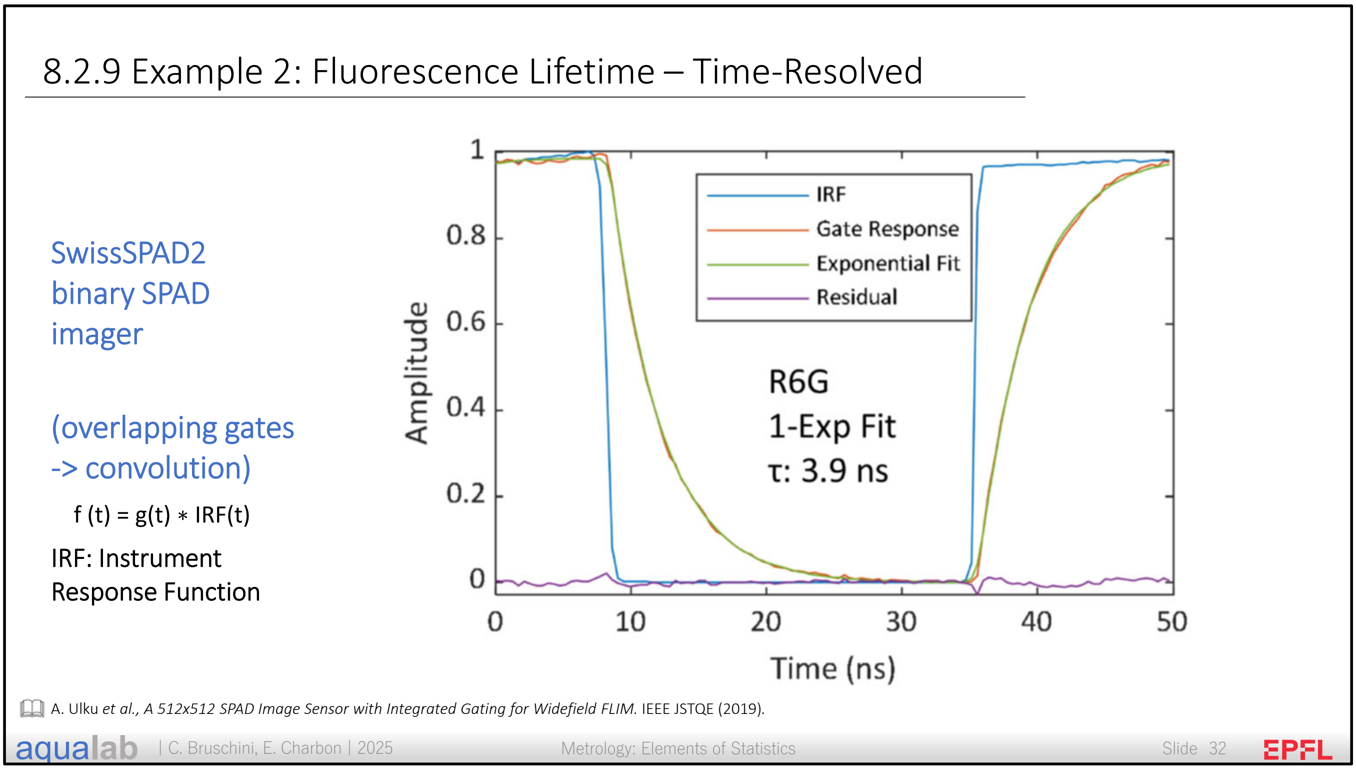
(overlapping gates)



📖 A. Ulku *et al., Large-Format Time-Gated SPAD Cameras for Real-Time Phasor-Based FLIM. EPFL Thèse 8311* (2021).

We see here how to sample the re-emitted light in a Gated approach, when the gates are "long" and overlapping. A sensor like SwissSPAD2 can open and close a gate in correspondence of each laser pulse and count the detected photons per pixel.

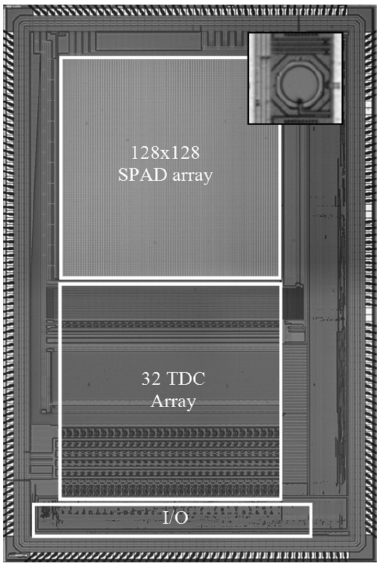
After sufficient statistics has been accumulated at a given gate position, the gate is shifted (by tens or hundreds of picoseconds) to cover another part of the re-emitted light, until the last position (*n* in this plot) is reached.

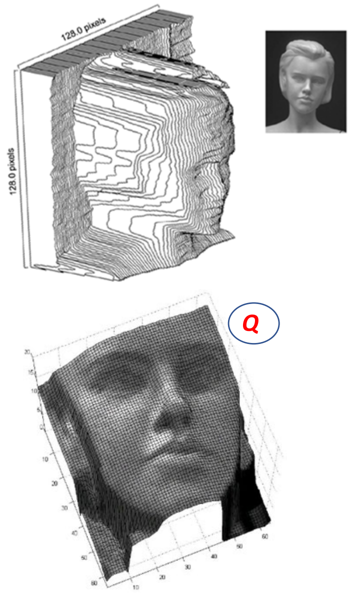


The result is that the measured signal (“Gate Response” in this graph) is not the original exponential any more, but its convolution with the sensor’s response, termed IRF (Instrument Response Function – a square gate for this kind of sensors). This is now the data from which the original lifetime information needs to be extracted.

The corresponding signals are shown here when using a 20 MHz laser -> 50 ns period. The target fluorophore is Rhodamine6G, which has a lifetime of ~3.9 ns. The “Residual” is the difference between the measured data and the fit, and indicates the fit quality.

8.2.9 Example 3: Real Life Truths – LIDAR & Timing Jitter in SPADs





C. Niclass *et al., A 128×128 Single-Photon Image Sensor With Column-Level 10-Bit Time-to-Digital Converter Array.* IEEE JSSC 43 (2008).

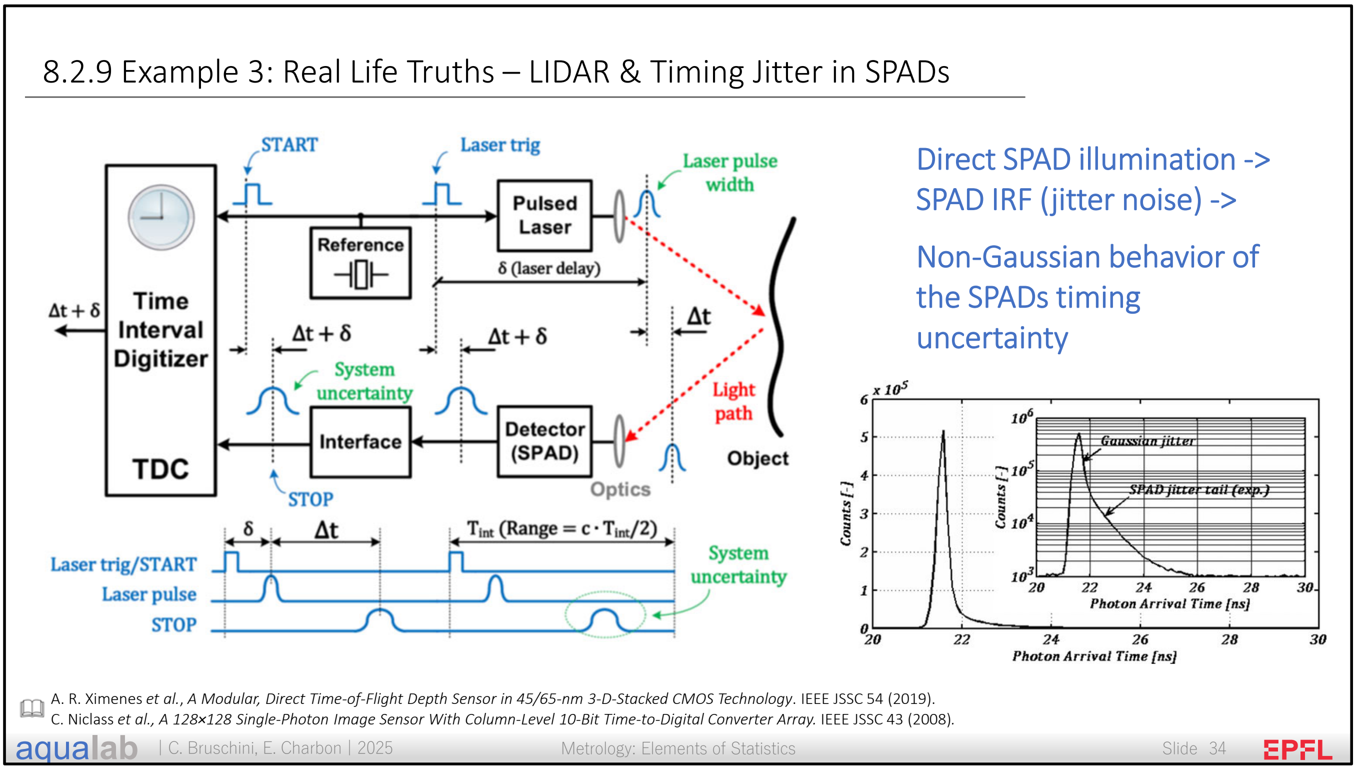
The third example is linked to a LIDAR (Light Detection and Ranging) application, the optical equivalent of RADAR. **A distance is measured by means of timing measurements**. The implementation details are shown in the next slide. There are obviously other 3D measurement techniques such as indirect time of flight, triangulation or interference, which are not discussed here.

Left: photomicrograph of an early EPFL AQUA silicon chip containing 128x128 SPADs coupled to an array of 32 TDCs (Time-to-Digital Converters), to timestamp the arrival of the photons backscattered from a target.

Right: depth-encoded 3D image. Bottom: same target but at much higher precision.

Q: how was this image obtained? Hint: calculate the precision of one single time (= distance) measurement starting from an estimate of the timestamping precision, e.g. 10-100 ps.

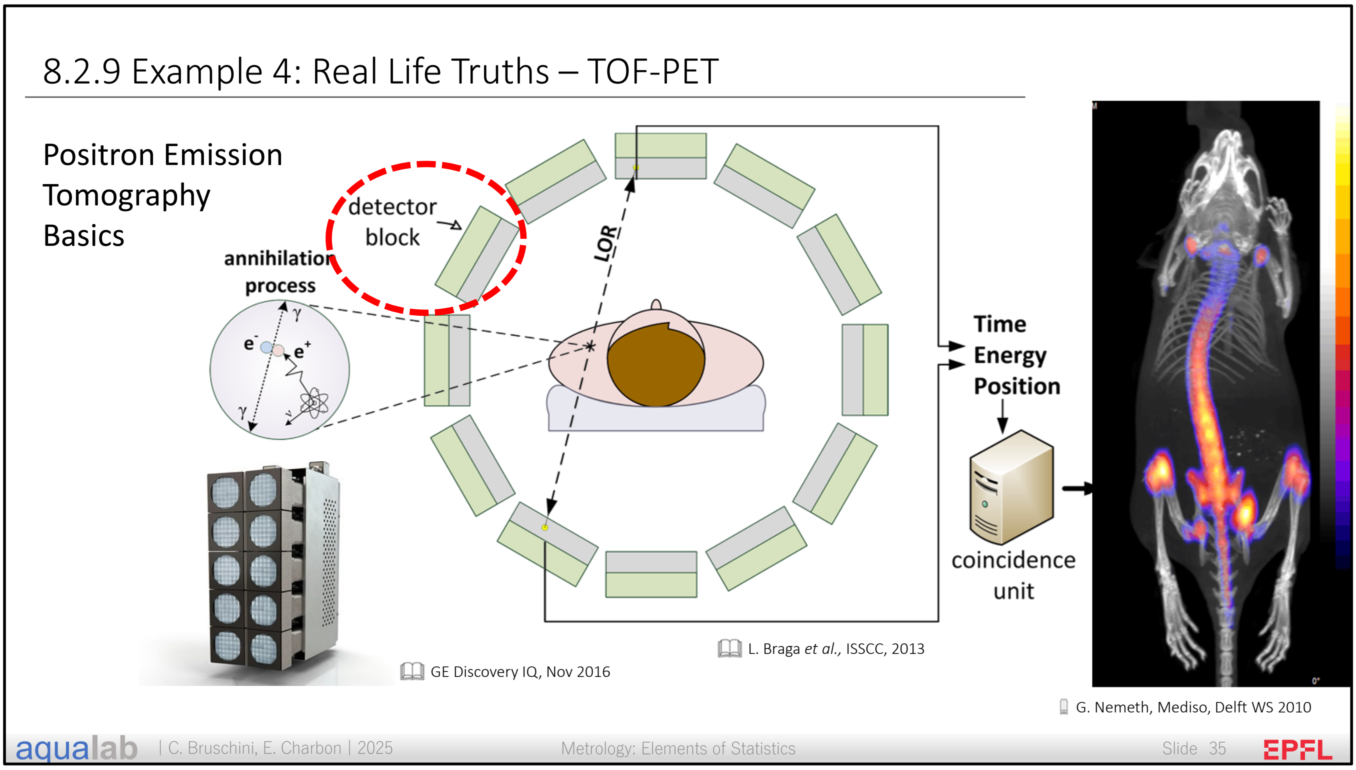
Q: How can this be improved? → next lecture on what happens when averaging repetitive measurements!



Left: basic principles of direct time-of-flight measurements. Note how the start/stop measurements are implemented, the synchronisation with the laser trigger, and the role of the system uncertainty (similar to the IRF in the previous slides) on the final measurement precision.

Right: how does the precision – or timing jitter – of the photodetector come into play? The SPAD response is not infinitely short, but characterised by a Gaussian central section, and an exponential (diffusion) tail on the right. These parts are linked to the device structure, process properties and resulting electric field distributions.

Q: How can the SPAD's IRF be determined? One method consists in illuminating directly the device and timestamping each photon, to then build a histogram similar to the one shown in the plot on the right. Note also the difference between linear and logarithmic scales!



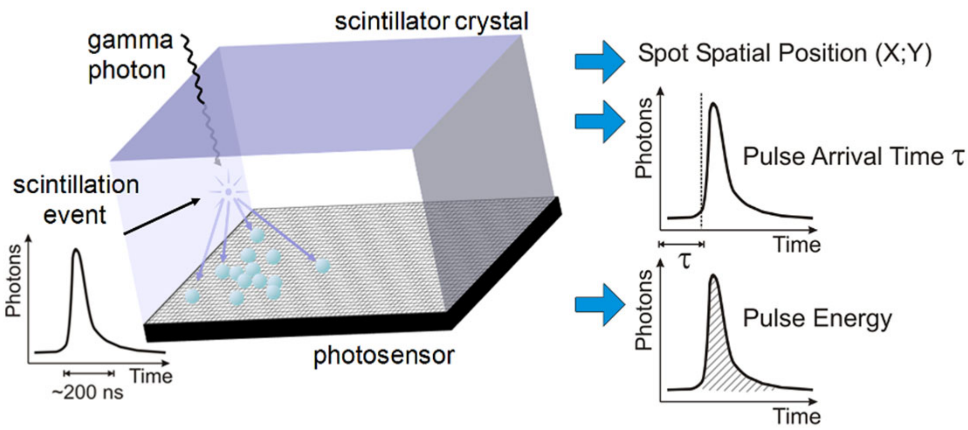
The fourth example is based on a molecular imaging technique which measures tiny concentrations of a radioisotope, by means of (visible) **light intensity and timing measurements**.

[F. Gramuglia, EPFL PhD Thesis #8720, 2022] PET is the most relevant molecular imaging technique available today. The reason for this is its high molecular sensitivity and large tissue penetration. Molecular imaging allows the *in vivo* visualization and quantification of biological processes at the cellular and molecular level by using so-called *molecular probes*. *Molecular probes* are defined as biocompatible image contrast enhancement agents that accumulate and stay in a specific target for a certain time.

In the specific case of PET, the used molecular probes are labeled with a positron-emitting radioisotope. One of the best known molecular probes is 18F-FDG, widely employed in oncology. This sugar-like compound is used to detect cancer cells in high metabolism growing tumor masses.

The molecular probes used are β^+ emitters. When a solution of a molecular probe such as 18F-FDG is injected into a patient undergoing a PET exam, the F-18 emits positrons inside the subject's body. The positron travels for a certain distance, called *positron range*, (~2 mm) before encountering and combining with an electron to forma positronium. This system is unstable, and after a time on the order of ~ 100 ps, the e^-/β^+ annihilation occurs. This causes the emission of two almost collinear γ-photons, each with an energy of E_γ= 511 keV (the equivalent of twice the electron rest masses) and traveling in opposite direction. The detection of these two γ-photons can allow the localization of the annihilation point. For this purpose, the subject under test is placed in a ring scanner.

8.2.9 Example 4: Real Life Truths – TOF-PET



📖 R. Walker *et al.*, IISW, 2013

[ctd] The scanner is fully covered by photodetectors aiming to detect the gammas coming from the patient and extract the so-called *line of response* (LOR). The LOR is defined as the imaginary line connecting the two points in the scanner where a gamma pair is detected and ideally passing through the generation point.

TOF-PET Module → measurement of energy, time-of-arrival and position
The detection of the gammas generated by the e^{-}/β^{+} annihilation can be performed with various types of detector, such as solid-state detectors, ionization chambers, inorganic scintillator-based detectors, etc. In PET, the detection widely relies on the use of inorganic scintillators because they offer the best compromise between energy, timing, and spatial resolution as well as a high detection efficiency for 511 keV γ-photons. These materials have the property of absorbing the incoming radiation and emitting optical photons following a fast exponential timing evolution with a decay time constant τ_d typical of each scintillator.

The emission is isotropic inside the scintillating material, and the amount of light is proportional to the amount of energy deposited in the crystal by the incoming radiation. The light burst (or pulse), generated by a scintillator, reveals when a gamma interacts in the crystal, and it can be detected by instrumenting one or more surfaces of the scintillator with a photodetector. Moreover, if the photodetector can quantify the amount of detected light, it is possible to estimate the energy of the incoming gamma, knowing the light yield.

The image above shows the schematic view of a typical module used in PET applications. In this simplified scheme, the scintillator is optically coupled to a photodetector and a PCB module is used to read out and process the signal.

The curves show the typical profiles in time of a scintillation event (emission of light – thousands of photons in the visible – after conversion of one gamma photon). But eventually, we want to send to the reconstruction software three fundamental pieces of information for each gamma event: **energy, time-of-arrival and position.** We are usually not interested in the full scintillation light waveform and in recording all photons individually!

8.2.9 Example 4: Real Life Truths – Scintillation Light

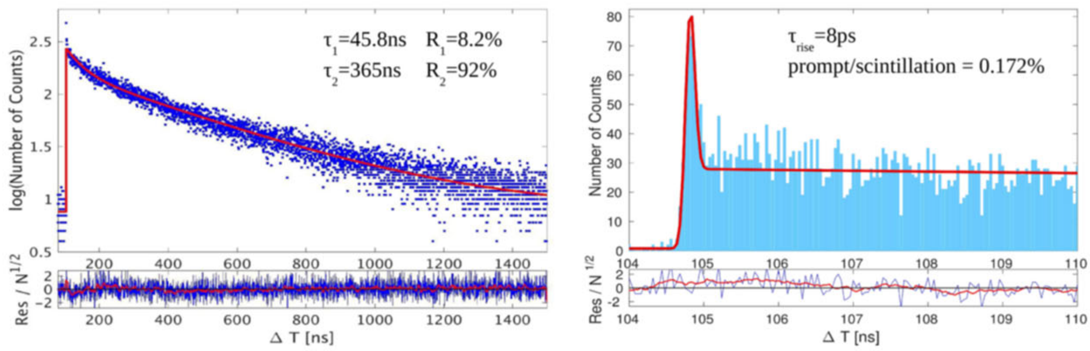


Figure 10. Scintillation decay and rise time of BGO measured with a time correlated single photon counting (TCSPC) setup using 511 keV annihilation gammas (Gundacker *et al* 2016b). The figure on the right hand side shows a pronounced Cherenkov peak at the onset of the scintillation emission with a relative abundance of 0.172% compared to the total amount of photons detected by the stop detector of the TCSPC setup.

Fast vs. “slow” scintillation photons in a heavy scintillating crystal

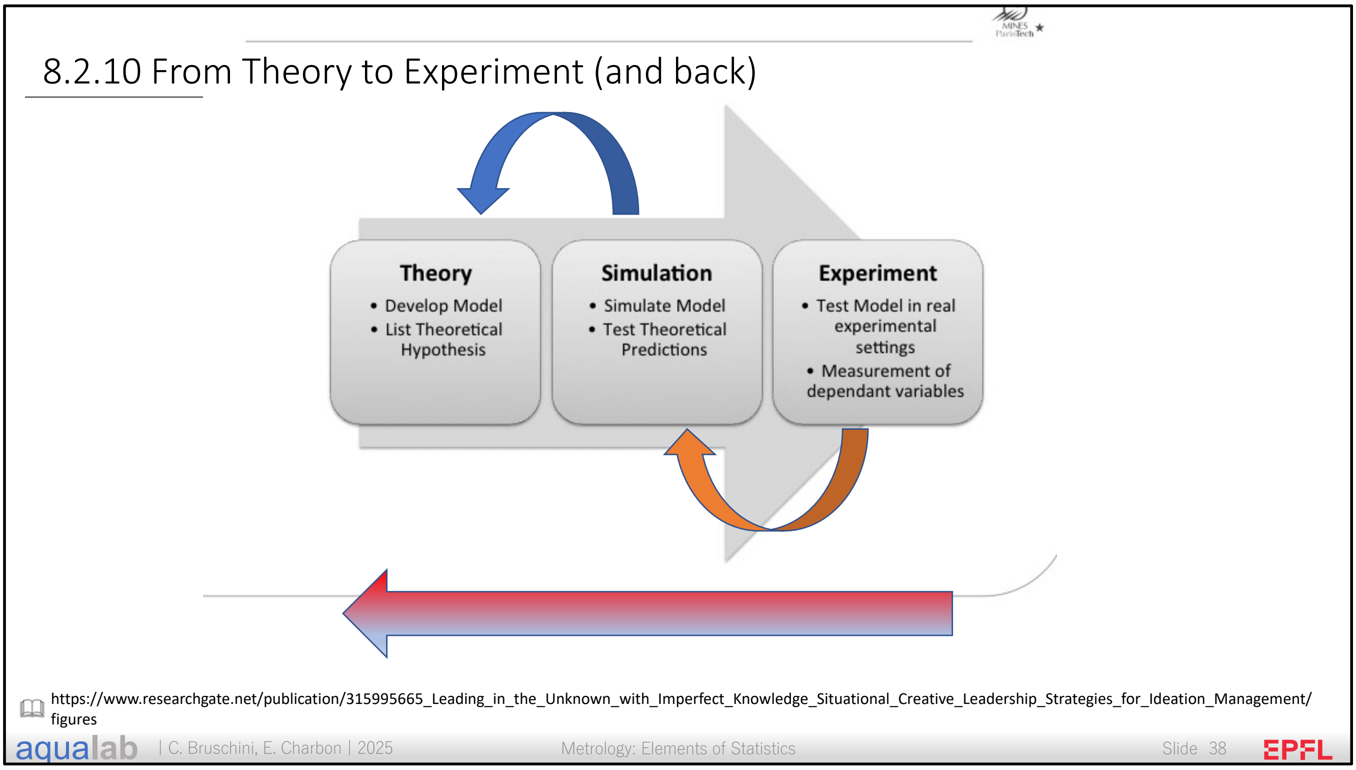
Gundacker S, Auffray E, Pauwels K and Lecoq P *Measurement of intrinsic rise times for various L(Y)SO and LuAG scintillators with a general study of prompt photons to achieve 10 ps in TOF-PET.* IOP Phys. Med. Biol. 61 2802–37

But how do we know that the scintillation light from a scintillator crystal does really have the profile shown in the previous slides? And how can we further exploit the real response of a scintillator?

→ We need to carry out a **precision (timing) measurement**, for example using a radioactive source and detecting as many visible light photons as possible, event after event, similarly to the TCPSC (time-correlated single-photon counting) method shown before.

We can then accumulate all time of arrival data into a histogramme such as the one shown above, which tells us for example that the light intensity decay is bi-exponential rather than monoexponential (left), and that there is actually a small fraction of photons that are emitted right after the gamma conversion (“prompt” events on the right). These could be very useful to improve the timing precision of the PET measurements, and therefore the final image quality!

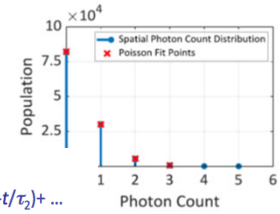
This data does allows allow the material scientists to design improved scintillators, e.g. with better light yield and/or faster decays and/or more prompt photons.

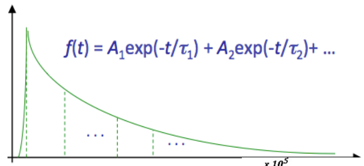


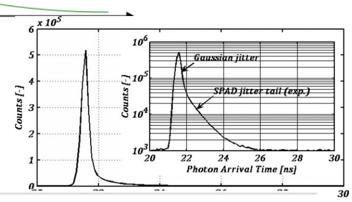
These example highlight the interplay between theory and experiment, possibly going through an intermediate simulation step. Sometimes we move from right to left, using the experimental data to develop/refine a theory.

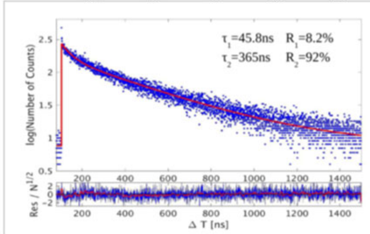
Take-home Messages/W8-2

- *Random Variables:*
  - RV distributions:
    - Poisson ↔ Exponential
    - Uniform, Gaussian
    - …and their main properties (see also W3)
  - Practical examples!
    - Single-photon imager & Poisson light distribution
    - Fluorescence lifetime & exponential decay
    - Timing jitter – combination of distributions ↔ physics
    - Scintillation light – combination of distributions ↔ physics









Second recap section: we summarise here the main definitions, results and examples discussed in this middle section.

Outline

8.3.1 Expected Values

- Given a discrete RV X with support $\mathcal{S} = \{x_1, x_2, \ldots\}$, the expected value (or expectation) of its distribution, which is commonly defined mean, is given by (weighted average):

$$E\{X\} = \sum_{j=1}^{\infty} x_j \, P\{X = x_j\}$$

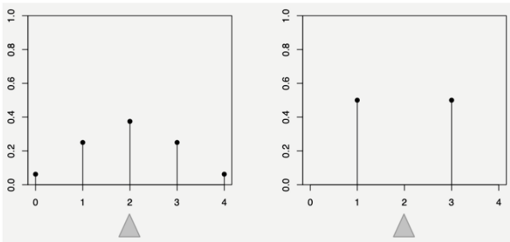
- The expected value is undefined if:

$$\sum_{j=1}^{\infty} |x_j| \, P\{X = x_j\} \to \infty$$

- Similarly, if X is a continuous RV with PDF $f_X(x)$:

$$E\{X\} = \int_{-\infty}^{\infty} x \, f_X(x) \, dx$$

Champions League 25.02.2020 Napoli – Barcelona		
1	X	2
3.26	3.59	2.11



NB: the expected value does not determine the distribution…

J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 4.1

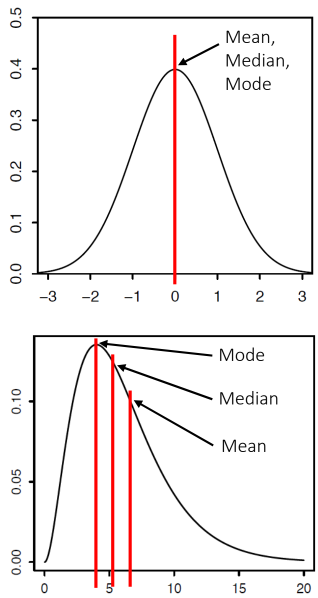
The expected value, or mean, is a very important parameter characterising the distribution of a random variable. Definitions are provided here for a discrete and a random variable. However, knowing the mean is usually far from enough….

8.3.2 Mean, Median and Mode

- As previously stated, the mean μ of a RV X is given by its expected value. It is called a measure of the central tendency of the distribution, specifically its center of mass.
- The median m of a RV X is that value such that $P\{X \leq m\} \geq 0.5$ and $P\{X \geq m\} \geq 0.5$. In a continuous RV, it is simply the value at which $F_X(m) = 0.5$.
- The mode c of a RV X is that value that maximizes the PMF (for a discrete RV) or the PDF (for a continuous RV):

$$P\{X = c\} \geq P\{X = x\} \text{ for all } x$$

$$f_X(c) \geq f_X(x) \text{ for all } x$$



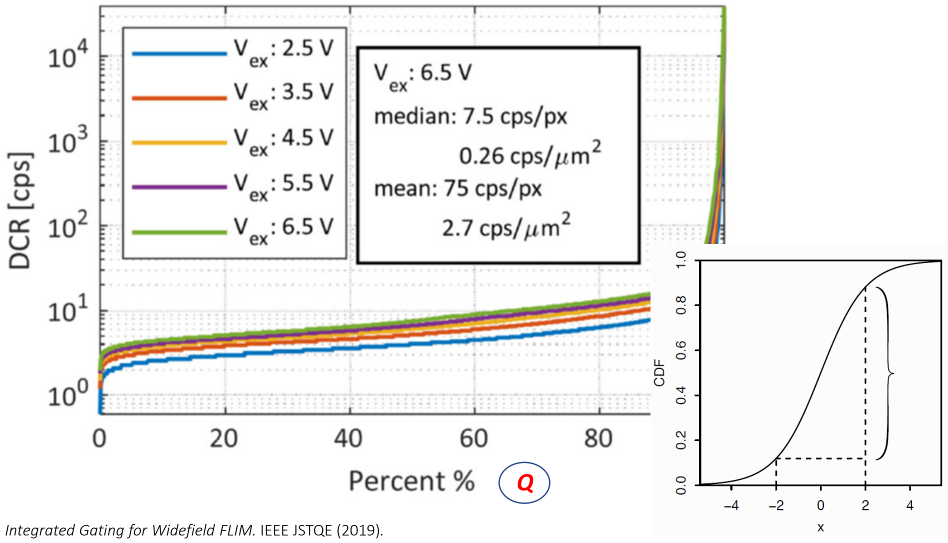
J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 6.1

Sometimes the median is a more useful figure, for example in cases where the target distribution is not symmetric. The mean is indeed quite heavily affected by the distribution tails and outliers – see the *PDF* plots here on the right.

8.3.2 Mean, Median and Mode – Example

SwissSPAD2
binary SPAD imager

noise level (DCR = Dark Count Rate, per pixel)



A. Ulku *et al., A 512x512 SPAD Image Sensor with Integrated Gating for Widefield FLIM.* IEEE JSTQE (2019).

An example of how the mean and median of real distributions can look like is provided here, in the form of the DCR (dark count rate) values, i.e. the noise level in the dark, for the pixels of the previously introduced SwissSPAD2 sensor. Several curves are reproduced for different operating conditions, here the excess bias voltage of the SPADs (i.e. the voltage level beyond breakdown).

Note that in this example the mean and median differ by one order of magnitude!

Q: what is the origin of this graph? Does it represent a PDF?

→ The percent axis is actually the CDF (0-1 interval = 0-100%) and the plot flipped with respect to the "usual" CDF representation (see bottom right image and the CDF definition).

8.3.3 Linearity of expectation and LOTUS

- The most important property of expectation is linearity (actually true for all RV, not only discrete ones). For every given RVs X and Y and any constant c, it follows:

$$E\{X+Y\} = E\{X\} + E\{Y\}$$

$$E\{cX\} = cE\{X\}$$

$E\{\text{dice}\} = 3.5$

- The law of the unconscious statistician (LOTUS) states that, despite $E\{g(X)\}$ does not equal $g(E\{X\})$, there is a way to measure $E\{g(X)\}$ without the need of finding $g(X)$. Given the discrete RV X and the function $g: \mathbb{R} \to \mathbb{R}$, follows:

$$E\{g(X)\} = \sum_x g(x)\, P\{X = x\} \text{ for all } X$$

$E\{2 \text{ dices}\} =$
$E\{\text{dice}\} + E\{\text{dice}\} = 7$

Similarly, if X is a cont. RV with PDF $f_X(x)$: $E\{g(X)\} = \int_{-\infty}^{\infty} g(x) f_X(x) dx$

J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 4.2-4.5, 5.1

Used in Section 8.3.6 (MGF)

8.3.4 Variance

- The variance of a RV X is (*average squared difference -> distribution spread*):

$$Var\{X\} = E\{(X - E\{X\})^2\} = E\{(X - \mu)^2\} = \sigma^2$$

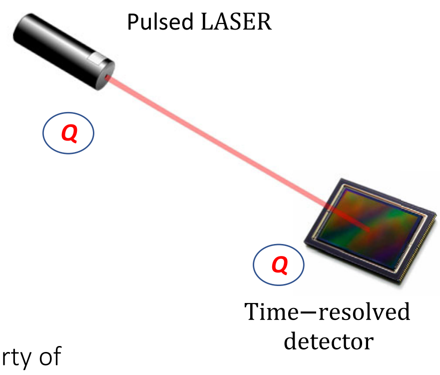
and its square root is called the standard deviation:

$$SD\{X\} = \sqrt{Var\{X\}} = \sigma$$

- For any RV X,

$$Var\{X\} = E\{X^2\} - E\{X\}^2 = E\{X^2\} - \mu^2$$

which can be demonstrated easily using the linearity property of the expected values.



J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 4.6

Another extremely important parameter of the distribution of a random variable is its **variance**, which basically indicates how the distribution deviates from its mean value, and its square root = **standard deviation**.

Let's take as example a time-resolved system composed of a photodetector, which measures the time-of-arrival of impinging photons with a certain timing error σ(detector), and of a laser, which generates a light pulse of width σ(laser).

Q: which is the resulting timing uncertainly of the complete system?

8.3.4 Variance (contd.)

- The Variance has the following properties:

1. $Var\{X + c\} = Var\{X\}$ for any constant c (shift a distribution).

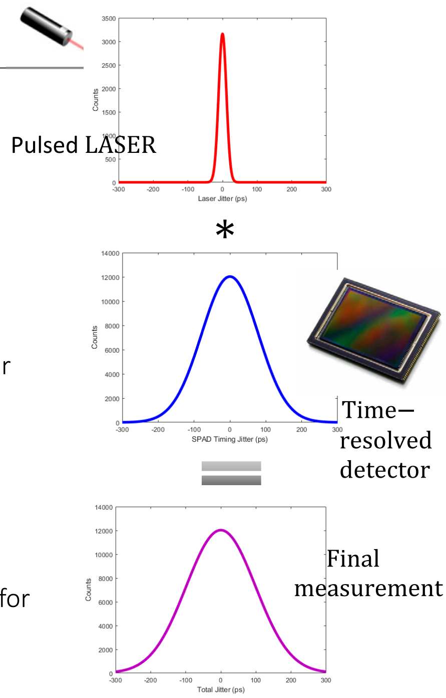
2. $Var\{cX\} = c^2 Var\{X\}$ for any constant c.

3. If X and Y are independent, then $Var\{X + Y\} = Var\{X\} + Var\{Y\}$. This is not true in general if X and Y are dependent. For example, in the case where $X = Y$:

$$Var\{X + Y\} = Var\{2X\} = 4\,Var\{X\} >$$

$$2\,Var\{X\} = Var\{X\} + Var\{Y\}$$

4. All $Var\{X\} \geq 0$, with the equality if and only if $P\{X = a\} = 1$ for some a. [only constants have 0 variance]



J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 4.6

We recall here several important properties of the variance of a random variable. The variances of *independent* RVs add up -> their standard deviations combine quadratically.

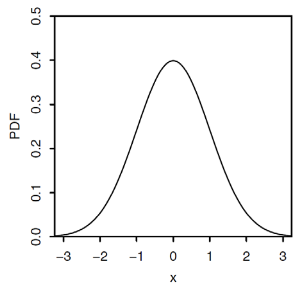
Right: another representation of the time-resolved system shown in the previous slide. Note the convolution ("*") of the distributions, whose variances add up (if they are independent).

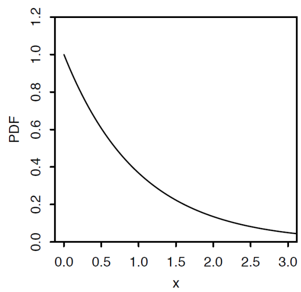
8.3.5 Moments

- Let X be a RV with mean μ and variance σ^2. For any positive n:
  1. the n-th moment of X is $E\{X^n\}$,
  2. the n-th central moment of X is $E\{(X-\mu)^n\}$,
  3. the n-th standardized moment of X is $E\left\{\left(\frac{X-\mu}{\sigma}\right)^n\right\}$.

- As we have seen previously, the first moment of a RV X is its mean value, or, in different words, the center of mass of the distribution:

$$n = 1: \quad \mu = E\{X\}$$





J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 6.2

Generalising the concepts of mean and variance, the distribution of a random variable X can be characterised by its **moments**. The first, second, third and fourth moment are defined in the following.

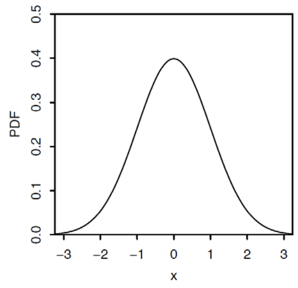
8.3.5 Moments (contd.)

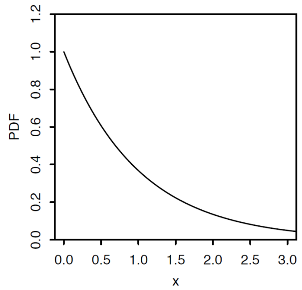
- In the same fashion, the second central moment of a RV X is its variance, or the moment of inertia of the distribution around its center:

$$n = 2: \ \sigma^2 = Var\{X\} = E\{(X - E\{X\})^2\}$$

- The third standardized moment of a RV X is defined as the skewness of the distribution. The skewness is a parameter that measures the asymmetry of the distribution. By standardizing, we make the skewness independent on the position and scale of X (information given by μ and σ):

$$n = 3: \ Skew\{X\} = E\left\{\left(\frac{X - \mu}{\sigma}\right)^3\right\}$$





J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 6.2

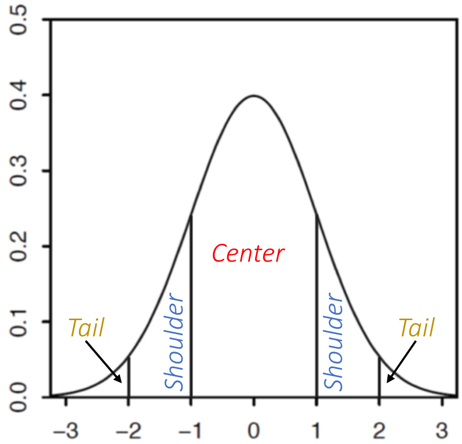
8.3.5 Moments (contd.)

S

- In general, the odd moments give information about the asymmetry of the distribution.
- The fourth standardized moment of a RV X is defined kurtosis of the distribution. If we split the distribution in three main regions, i.e. in the center (1 σ around μ), the shoulders (between 1 and 2 σ's around μ) and the tails (more than 2 σ's from μ), then the kurtosis gives information about the tails.

$$Kurt\{X\} = E\left\{\left(\frac{X-\mu}{\sigma}\right)^4\right\} - 3$$

a classical distribution with large kurtosis is a PDF with a sharp peak at the center, low shoulders and heavy tails.



J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 6.2

8.3.5 Moments (contd.) – Textbook Example

S

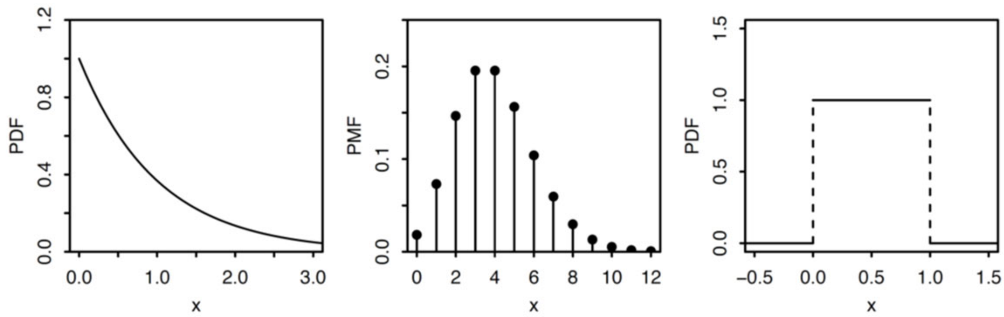
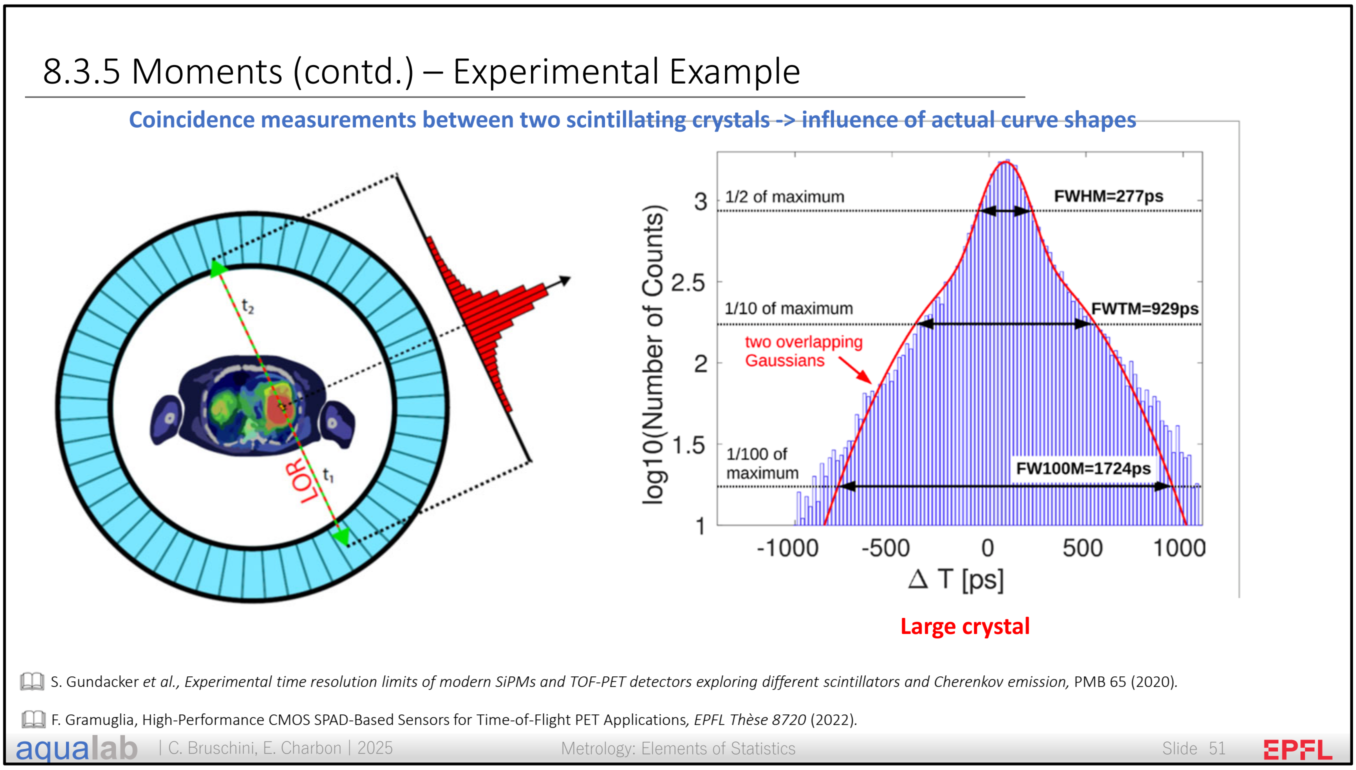


FIGURE 6.6
Skewness and kurtosis of some named distributions. Left: Expo(1) PDF, skewness = 2, kurtosis = 6. Middle: Pois(4) PMF, skewness = 0.5, kurtosis = 0.25. Right: Unif(0, 1) PDF, skewness = 0, kurtosis = −1.2.

J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 6.2

Try to calculate the moments shown in this example!



Moments in a real experimental set-ups: this example shows real experimental distributions of the timing difference of gamma events detected on a given LOR (line of response), such as “opposite” elements of a PET detector (left), or two scintillating crystals placed face-to-face (right).

Left: operating principle of a time-of-flight PET machine. PET detector ring (blue), reconstructed cross-section of a patient (centre), and histogram of the origin of the gamma events along the LOR as calculated from the timing difference t_2-t_1. The time-of-flight information allows a precise determination of the origin of each event and improves the final image quality. The smaller the timing error, the better the resulting SNR. The best detectors can measure the time-of-arrival of gamma events with an error smaller than hundred picoseconds.

Right: timing difference t_2-t_1 as measured in the lab for two experimental crystals placed face-to-face (not shown). Note the vertical logarithmic scale. The resulting distribution is not Gaussian, but well fitted by the sum of two overlapping Gaussians. As a consequence, not only the width of the curve is important (FWHM = Full Width at Half Maximum), but other quantities as well, such as the FWTM (Full Width at 1/10 of Maximum) and FW100M (Full Width at 1/100 of Maximum). These parameters allow to better gauge the importance of the tails of the distributions, and their importance in the final image reconstruction.

8.3.6 Moment Generating Functions

- The moment generating function (MGF) of a RV X is defined as:

$$\text{MGF: } \phi(t) = E\{e^{tX}\} = \begin{cases} \sum_x e^{tx} p_X(x), & \text{if } X \text{ is discrete } * \\ \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx\,, & \text{if } X \text{ is continuous } * \end{cases}$$

- We call $\phi(t)$ the moment generating function because all the moments of X can be obtained by successfully differentiating $\phi(t)$. It follows that:

$$\phi'(t) = \frac{d}{dt} E\{e^{tX}\} = E\{Xe^{tX}\} \quad \rightarrow \quad \phi'(0) = E\{X\}$$

$$\phi''(t) = \frac{d}{dt}\phi'(t) = \frac{d}{dt} E\{Xe^{tX}\} = E\{X^2 e^{tX}\} \quad \rightarrow \quad \phi''(0) = E\{X^2\}$$

$$\phi^{(n)}(0) = E\{X^n\}, \qquad \text{for all } n \geq 1$$

The MGF is a "tool" to calculate the moments – by differentiating it – provided that an analytical expression of the random variable is given.

**Using LOTUS (Section 8.3.3)*

S.M. Ross, *Introduction to Probability Models*, 10th ed., 2009, Chap. 2.6

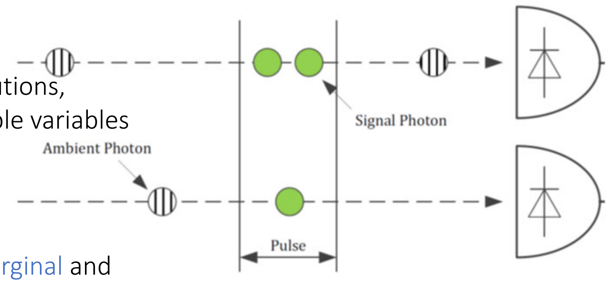
The moment generating function (MGF) $\phi(t)$ is a "tool" to calculate the moments $E\{X^n\}$ of a random variable X by differentiating it – i.e. calculating $\phi^{(n)}(t) \rightarrow \phi^{(n)}(0)$ – provided that an analytical expression of the random variable is given. Its use will be illustrated in the Homeworks.

Outline

8.4 Multivariate Distributions

- During experiments, in real life, we have to deal with multiple RVs. It is very important to know the relationship between different RVs, i.e. if they are independent or dependent on each other.
- The joint distributions, also called multivariate distributions, capture the missing information about how the multiple variables interact.
- The key concepts that will be studied are the joint, marginal and conditional distributions of two variables (see also Appendix A).

Example: LiDAR = detection of backscattered signal photons in presence of background light



$X = signal,$
$Y = noise\ (background, DCR, etc.)$

J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 7

M. Beer *et al., Background Light Rejection…, MDPI Sensors 18, 2018.*

We will now briefly look at more complex distributions of multiple random variables, i.e. *Joint* and *Independent* Distributions.
Marginal and *Conditional* distributions are left for reference in the Appendix, for those who would like to know more.

An example of two random variables – signal and background – is provided on the right, here for the case of the LIDAR application.

8.4.1 Joint Distributions

- The joint distribution of two RVs X and Y provides complete information about the probability of the vector (X, Y) falling into any subset of the plane.
- The joint CDF of two RVs X and Y is a function $F_{X,Y}$ such that:

$$\text{CDF: } F_{X,Y}(x, y) = P\{X \leq x, Y \leq y\}$$

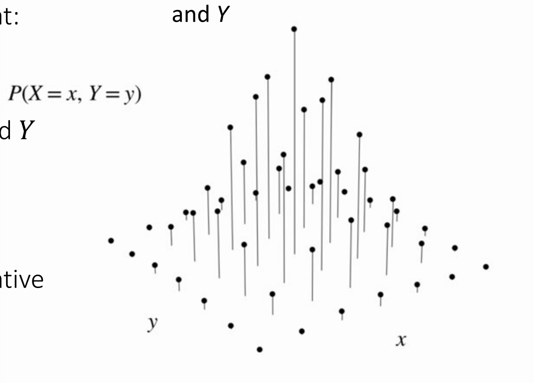
- In the same fashion, the joint PMF of two discrete RVs X and Y is a function $p_{X,Y}$ such that:

$$\text{PMF: } p_{X,Y}(x, y) = P\{X = x, Y = y\}$$

- In the same way of the univariate PMF, it has to be nonnegative and sum up to 1:

$$\sum_x \sum_y P\{X = x, Y = y\} = 1$$

Joint PMF of discrete RVs X and Y



J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 7.1

We generalise here the concepts of CDF and PMF to the joint distributions of two random variables. An example of joint PMF is shown on the right for two discrete RVs.

8.4.1 Joint Distributions (contd.)

- Analogously, the joint PDF of two continuous RVs X and Y is given by:

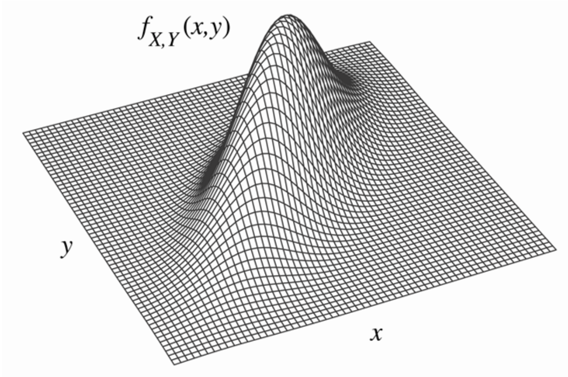
$$\text{PDF: } f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$$

- In order for the joint PDF to be valid, it has to be nonnegative and integrate to 1:

$$f_{X,Y}(x,y) \geq 0 \text{ for all } (x,y)$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx\, dy = 1$$

Joint PDF of continuous RVs X and Y



J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 7.2

Similarly to the previous slide, an example of joint PDF is shown on the right for two continuous RVs.

8.4.2 Independent Distributions

- Two RVs X and Y are independent if:

$$F_{X,Y}(x, y) = F_X(x)\, F_Y(y)$$

which is equivalent to say, for discrete RVs:

$$P\{X = x, Y = y\} = P\{X = x\}P\{Y = y\}$$

and for continuous RVs:

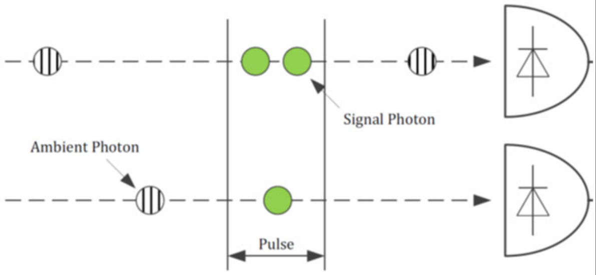
$$f_{X,Y}(x, y) = f_X(x)\, f_Y(y)$$

or

$$f_{Y|X}(y|x) = f_Y(y)$$

for all x and y.

Example: LiDAR employing detection of photon coincidences (within a coincidence window) in presence of background light



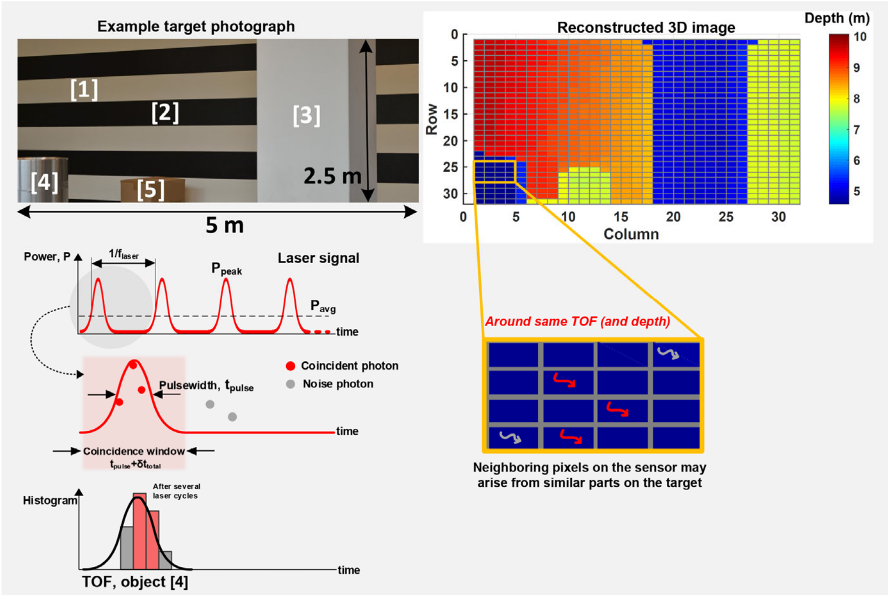
J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 7.1

M. Beer *et al., Background Light Rejection..., MDPI Sensors 18, 2018.*

In the case of independent random variables, the final PDF (probability density function) and CDF (cumulative distribution function) are substantially simplified, reducing to the product of the individual distributions.

NB: the bottom definition is a generalisation to multiple random variables of the *conditional probability* discussed in Appendix A8.1.1 and A8.1.3.

8.4.3 Example: LIDAR & Coincidence Detection



Photon coincidences

Coincidence detection is a well-known technique which utilizes spatio-temporal correlations of photons within a laser pulse to filter out background noise photons which are uniformly distributed in time
-> concept of coincidence window to reduce the likelihood of acquiring noise events (Appendix B)

P. Padmanabhan et al., *Modeling and Analysis of a Direct Time-of-Flight Sensor Architecture for LiDAR Applications*, Sensors 2019

An example of a multivariate distribution as applied to the **measurement of time-of-flight** and the design of a bespoke sensor is provided in Appendix 8.B. We summarise here the basic idea:

Coincidence detection is a well-known technique which utilizes spatio-temporal correlation of photons within a laser pulse to filter out background noise photons which are uniformly distributed in time. The figure above conceptually explains this technique with an example scene and a measured 3D image reconstruction.
The main idea is to exploit the fact that the signal photons reflected from the target are temporally correlated and thus, most likely to be concentrated within a time-window coarsely equal to the total system full width at half maximum, *FWHM,* of the laser pulse....

Instead of letting the sensor integrate events over a long measurement window, imposing this time constraint, referred to as the “coincidence window”, reduces the likelihood of acquiring noise events whose probability of occurrence within that window is very low, thus, electrically enhancing the signal to background noise ratio, SBR. Coincidence may be implemented at the sensor level over clusters/groups of closely-spaced pixels, exploiting a “more-likely” fact that neighboring pixels may belong to similar target depths (and thus, TOFs), as depicted in the figure for the object labelled [4].

8.4.4 Covariance

- The covariance of the joint distribution of two RVs X and Y represents their tendency to go up or down together ("single-number summary"):

$$Cov\{X,Y\} = E\{(X - E\{X\})(Y - E\{Y\})\}$$

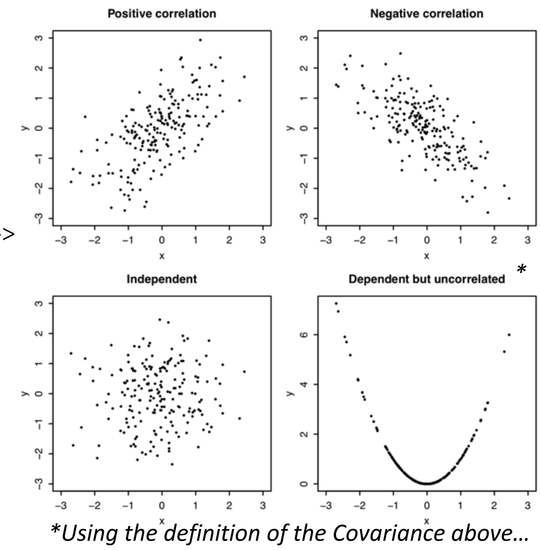
which, using linearity, becomes

$$Cov\{X,Y\} = E\{XY\} - E\{X\}E\{Y\}$$

- If two RVs are independent, then their covariance is zero (-> uncorrelated RVs), because:

$$E\{XY\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f_X(x)\, f_Y(y)\, dx\, dy =$$

$$= \int_{-\infty}^{\infty} x\, f_X(x)\, dx \int_{-\infty}^{\infty} y\, f_Y(y)\, dy = E\{X\}E\{Y\}$$



*Using the definition of the Covariance above…

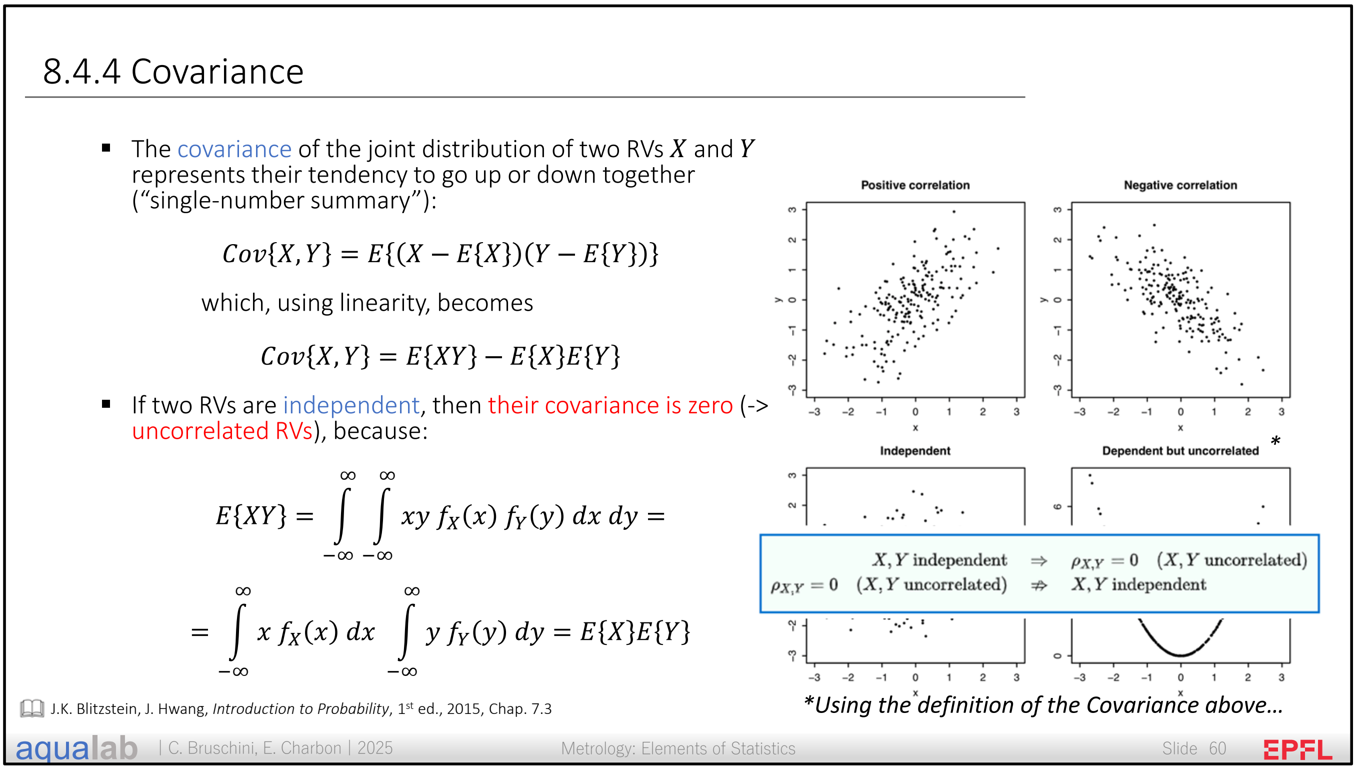
📖 J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 7.3

How can we determine in practice if two random variables are **independent**, which would simplify a lot their joint distribution? We can for example calculate their **covariance**, which is much easier to verify that the statistical independence – if nonzero, it indicates that the RVs are not independent.

Top left: positive correlation ($Cov > 0$), top right: negative correlation ($Cov < 0$), bottom left: independent → uncorrelated ($Cov = Corr = 0$), bottom right: Y is a deterministic function of X ($X \sim \mathcal{N}(0,1)$, $Y = X^2$), but X and Y are uncorrelated (→ $Cov = 0$!) using the definition of the Covariance above*…

*Note 1: The Covariance as defined here is a measure of linear association → RVs can be dependent in nonlinear ways and still have zero covariance.

____



How can we determine in practice if two random variables are **independent**, which would simplify a lot their joint distribution? We can for example calculate their **covariance**, which is much easier to verify that the statistical independence – if nonzero, it indicates that the RVs are not independent.

Top left: positive correlation ($Cov > 0$), top right: negative correlation ($Cov < 0$), bottom left: independent → uncorrelated ($Cov = Corr = 0$), bottom right: Y is a deterministic function of X ($X \sim \mathcal{N}(0,1)$, $Y = X^2$), but X and Y are uncorrelated (→ $Cov = 0$!) using the definition of the Covariance above*…

*Note 1: The Covariance as defined here is a measure of linear association → RVs can be dependent in nonlinear ways and still have zero covariance.

Note 2: the inverse of the theorem above (two RVs independent → uncorrelated RVs) is not true, i.e. just because X and Y are uncorrelated ($Cov = 0$) does not mean that they are independent.

____

8.4.4 Covariance (contd.)

- The covariance , which is much easier to verify that the statistical independence, has the following properties:

1. $Cov\{X, X\} = Var\{X\}$
2. $Cov\{X, Y\} = Cov\{Y, X\}$
3. $Cov\{X, c\} = 0$ for any constant c
4. $Cov\{aX, Y\} = a\, Cov\{X, Y\}$ for any constant a
5. $Cov\{X + Y, Z\} = Cov\{X, Z\} + Cov\{Y, Z\}$
6. $Cov\{X + Y, W + Z\} = Cov\{X, Z\} + Cov\{Y, Z\} + Cov\{X, W\} + Cov\{Y, W\}$
7. $Var\{X + Y\} = Var\{X\} + Var\{Y\} + 2Cov\{X, Y\}$
8. $Var\{X_1 + \cdots + X_n\} = Var\{X_1\} + \cdots + Var\{X_n\} + 2\sum_{i<j} Cov\{X_i, X_j\}$

J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 7.3

Of particular importance are the two last properties, which must be known: simply adding the variances of random variables is not always correct!

8.4.5 Correlation

- The correlation between two RVs X and Y is given by (unitless version of the covariance):

$$Corr\{X,Y\} = \frac{Cov\{X,Y\}}{\sqrt{Var\{X\}\,Var\{Y\}}}$$

- Notice that this formulation is insensitive to scaling. In fact:

$$Corr\{cX,Y\} = \frac{Cov\{cX,Y\}}{\sqrt{Var\{cX\}\,Var\{Y\}}} = \frac{c\,Cov\{X,Y\}}{\sqrt{c^2\,Var\{X\}\,Var\{Y\}}} = Corr\{X,Y\}$$

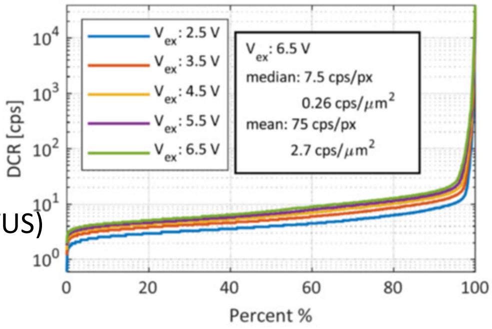
- Moreover, the correlation is bounded:

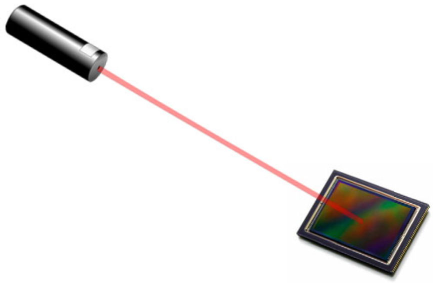
$$-1 \le Corr\{X,Y\} \le 1$$

J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 7.3

Take-home Messages/W8-3

- *Moments:*
  - Expected value (mean), median, mode
  - Linearity and law of the unconscious statistician (LOTUS)
  - Variance/standard deviation and its properties
    - *Example of laser and time-resolved measurement*
  - Moments: general definitions, MGF
- *Covariance and Correlation:*
  - Multivariate, joint and independent distributions
  - Covariance and correlation
    - *Covariance properties(!), e.g.* $Var\{X_1 + \cdots + X_n\}$





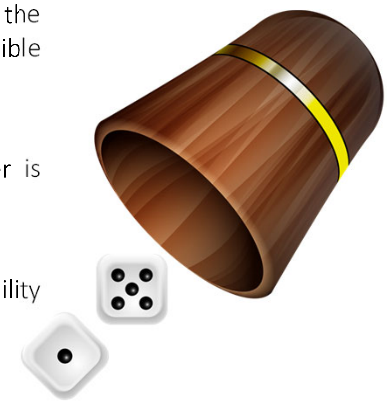
Third recap section: we summarise here the main definitions, results and examples discussed in this third and final section.

Appendix

Appendix 8.1: Introduction to Probability (contd.) – Example

Fair dice

- The classic example to explain the concept of probability is the fair dice. In a fair dice, the probability of obtaining one of the six faces, for example to get the number three, is, as we know, the ratio between the number of positive configurations and the number of total possible configurations: $P\{\text{face is 3}\} = 1/6$.
- In the same fashion, the probability of obtaining an odd number is $P\{\text{face is odd}\} = 3/6$.
- The fair dice represents the classical example of uniform probability distribution, as we will see.

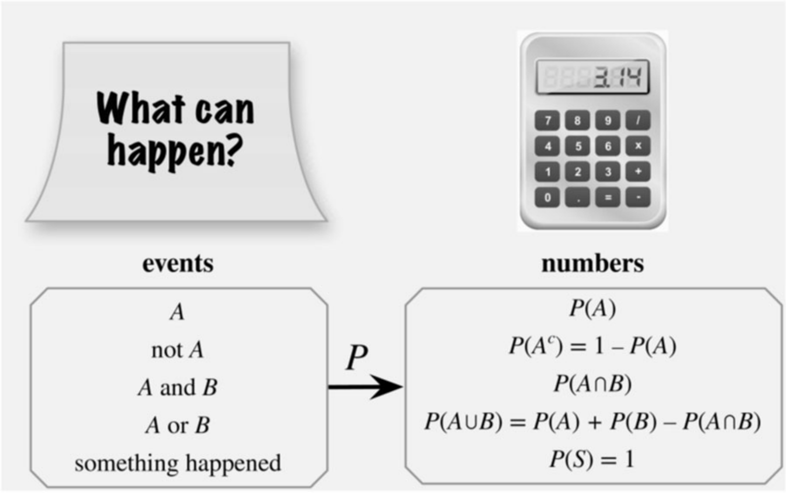


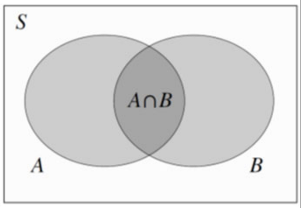
A. Papoulis, *Probability, Random Variables and Stochastic Processes*, 3rd ed., 1991, Chap. 1

This section contains several exact definitions and properties of importance for probability theory.

A8.1 Introduction to Probability (contd.)

How a probability function maps events to numbers



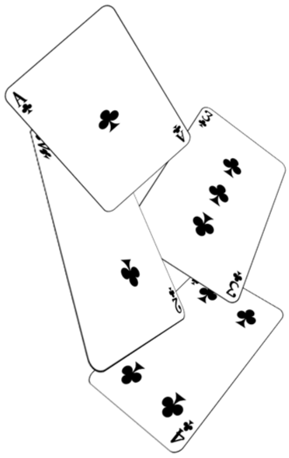


J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 1. 6, 1.7

A8.1.1 Conditional Probability

Thinking conditionally – whenever we observe new evidence (i.e., obtain data), we acquire information that may affect our uncertainties.

Conditional probability answers one simple question: how should we update our beliefs in light of the evidence we observe?



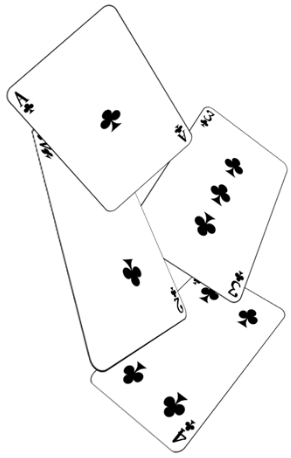
- If $\mathcal{A}$ and $\mathcal{B}$ are events with $P\{\mathcal{B}\} > 0$, then the conditional probability of $\mathcal{A}$ given $\mathcal{B}$ ($\mathcal{B}$ being the evidence which we observe) is *defined* as:

$$P\{\mathcal{A}|\mathcal{B}\} = \frac{P\{\mathcal{A} \cap \mathcal{B}\}}{P\{\mathcal{B}\}}$$

J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 2.2, 2.3

A8.1.1 Conditional Probability (contd.) – Example

Example: Two cards are extracted from a standard deck. Let $\mathcal{A}$ be the event that the first card is a heart, and $\mathcal{B}$ the event that the second card is red. Find $P\{\mathcal{A}|\mathcal{B}\}$ and $P\{\mathcal{B}|\mathcal{A}\}$.



- From naïve definition of probability:

$$P\{\mathcal{A}\cap\mathcal{B}\} = \frac{13}{52}\cdot\frac{25}{51} = \frac{25}{204}\,(= P\{\mathcal{B}\cap\mathcal{A}\})$$

while $P\{\mathcal{A}\} = 1/4$ and $P\{\mathcal{B}\} = 1/2$.

- Follows:

$$P\{\mathcal{A}|\mathcal{B}\} = \frac{P\{\mathcal{A}\cap\mathcal{B}\}}{P\{\mathcal{B}\}} = \frac{25/204}{1/2} = \frac{25}{102}$$

$$P\{\mathcal{B}|\mathcal{A}\} = \frac{P\{\mathcal{B}\cap\mathcal{A}\}}{P\{\mathcal{A}\}} = \frac{25/204}{1/4} = \frac{25}{51}$$

J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 2.2

This example is more involved than what it looks like at first sight…

- Naïve definition = just count the possible outcomes (Blitzstein 1.3).

- P{B}=1/2 is a bit less obvious but can be demonstrated (Blitzstein pp. 42-43 – "there are 26 favorable possibilities for the second card, and for each of those, the first card can be any other card (recall from Chapter 1 that chronological order is not needed in the multiplication rule).").

NB: "… the chronological order in which cards were chosen does not dictate which conditional probabilities we can look at. When we calculate conditional probabilities, we are considering what information observing one event provides about another event, not whether one event causes another."

NB: "P{A given B} and vice versa introduces an evidence which fundamentally changes the outcome (in terms of probability)."

A8.1.1 Conditional Probability (contd.)

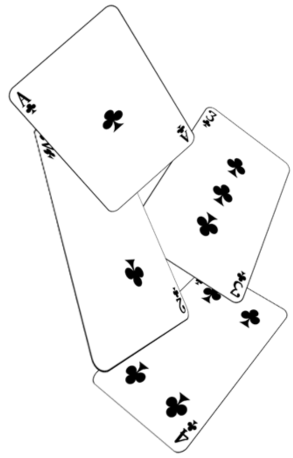
- From the definition of the conditional probability:

$$P\{\mathcal{A}|\mathcal{B}\} = \frac{P\{\mathcal{A} \cap \mathcal{B}\}}{P\{\mathcal{B}\}}$$

we can derive the following theorem:

$$P\{\mathcal{A} \cap \mathcal{B}\} = P\{\mathcal{B}\}P\{\mathcal{A}|\mathcal{B}\} = P\{\mathcal{A}\}P\{\mathcal{B}|\mathcal{A}\} = P\{\mathcal{B} \cap \mathcal{A}\}$$

since $P\{\mathcal{A} \cap \mathcal{B}\} = P\{\mathcal{B} \cap \mathcal{A}\}$



Applying it repeatedly, we can generalize to the intersection of *n* events (*commas = intersections*):

$$P\{\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n\}$$
$$= P\{\mathcal{A}_1\}P\{\mathcal{A}_2|\mathcal{A}_1\}P\{\mathcal{A}_3|\mathcal{A}_1, \mathcal{A}_2\} \ldots P\{\mathcal{A}_n|\mathcal{A}_1, \ldots, \mathcal{A}_{n-1}\}$$

J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 2.2, 2.3

NB: Intersection = AND (the intersection A ∩ B is the event that occurs if and only if both A and B occur).

A8.1.2 Bayes' Rule and Total Probability

- Manipulating the relationship:

$$P\{\mathcal{A} \cap \mathcal{B}\} = P\{\mathcal{B}\}P\{\mathcal{A}|\mathcal{B}\} = P\{\mathcal{A}\}P\{\mathcal{B}|\mathcal{A}\}$$

we can derive the following theorem (Bayes' rule):

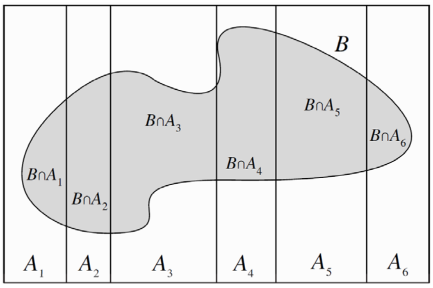
$$P\{\mathcal{A}|\mathcal{B}\} = \frac{P\{\mathcal{B}|\mathcal{A}\}P\{\mathcal{A}\}}{P\{\mathcal{B}\}}$$

which can be extremely useful in case $P\{\mathcal{B}|\mathcal{A}\}$ is much easier to find than $P\{\mathcal{A}|\mathcal{B}\}$, or vice versa.

- Sometimes, it can be extremely convenient to split a complex statistical problem into smaller pieces. In order to do that, one can apply the law of total probability (LOTP)*:

$$P\{\mathcal{B}\} = \sum_{i=1}^{n} P\{\mathcal{B} \cap \mathcal{A}_i\} = \sum_{i=1}^{n} P\{\mathcal{A}_i\}\, P\{\mathcal{B}|\mathcal{A}_i\}$$

*Relates conditional to unconditional probabilities



Ex

J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 2.3

An example of the application of Bayes' rule is provided in the Homeworks.

NB: $\mathcal{A}_i$ are (still) events.

A8.1.3 Independence of Events

- Two events are stochastically independent if:

$$P\{\mathcal{A} \cap \mathcal{B}\} = P\{\mathcal{A}\}P\{\mathcal{B}\}$$

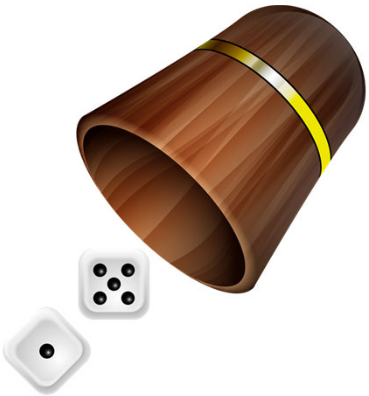
and if $P\{\mathcal{A}\} > 0$ and $P\{\mathcal{B}\} > 0$ then this is equivalent to (from the definition of the conditional probability):

$$P\{\mathcal{A}|\mathcal{B}\} = P\{\mathcal{A}\}, \qquad P\{\mathcal{B}|\mathcal{A}\} = P\{\mathcal{B}\}$$

- In words, two events $\mathcal{A}$ and $\mathcal{B}$ are independent if learning that $\mathcal{B}$ occurred has no influence on the probability of the event $\mathcal{A}$ to happen (and vice versa).
- As consequence, it also has no influence on the probability of the opposite of $\mathcal{A}$, $\mathcal{A}^c$:

$$P\{\mathcal{A}^c|\mathcal{B}\} = 1 - P\{\mathcal{A}|\mathcal{B}\} = 1 - P\{\mathcal{A}\} = P\{\mathcal{A}^c\}$$

- Hence, if $\mathcal{A}$ and $\mathcal{B}$ are independent, then also $\mathcal{A}^c$ and $\mathcal{B}^c$ are. Sometimes this property can be extremely useful.



J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap 2.5

We refer to this definition of independent events when discussing independent random variables.

Appendix 8.A – Multivariate Distributions: Marginal Distributions

- For discrete RVs X and Y, the marginal (or unconditional) PMF of X is given by:

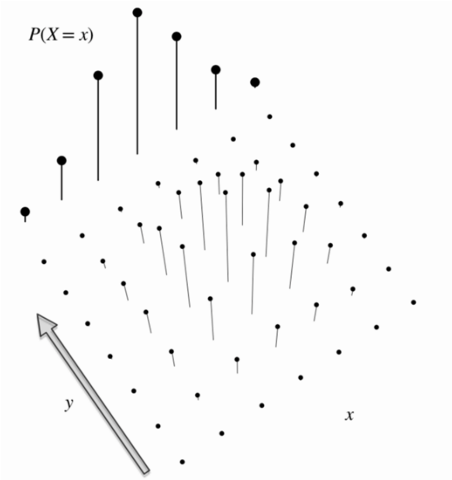
$$P\{X = x\} = \sum_y P\{X = x, Y = y\}$$

(distribution of X alone by summing over all Y)

- In the same way, the marginal CDF of X is obtained by:

$$F_X(x) = P\{X \leq x\} = \lim_{y \to \infty} P\{X \leq x, Y \leq y\} =$$

$$= \lim_{y \to \infty} F_{X,Y}(x, y)$$



Marginal PMF example

J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 7.1

Marginal and *Conditional* distributions are summarised in this Appendix, for those who would like to know more.

Top & image: the marginal PMF is indeed the sum along Y (not the projection!).

Bottom: the marginal CDF can indeed also be obtained as the limit of the joint CDF when $y \to \infty$.

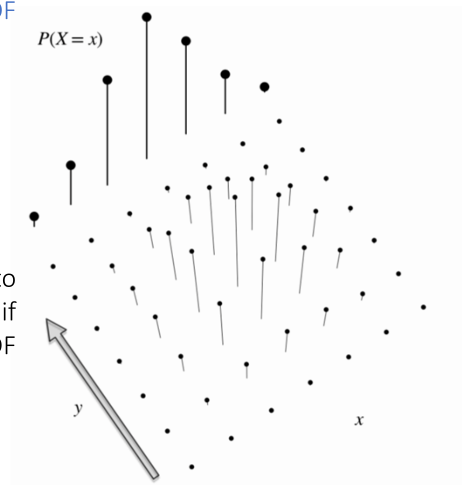
Appendix 8.A – Multivariate Distributions: Marginal Distributions

- For continuous RVs X and Y with joint PDF $f_{X,Y}$, the marginal PDF of X is given by:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy$$

- In the more general case of more than two RVs, all that needs to be done is an integration along the unwanted RVs. For example, if we have the joint PDF of X, Y, W and Z, but we want the joint PDF of the distributions in X and W:

$$f_{X,W}(x,w) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y,W,Z}(x,y,w,z)\, dy\, dz$$



J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 7.2

NB: the image should actually show the marginal distribution for continuous variables.

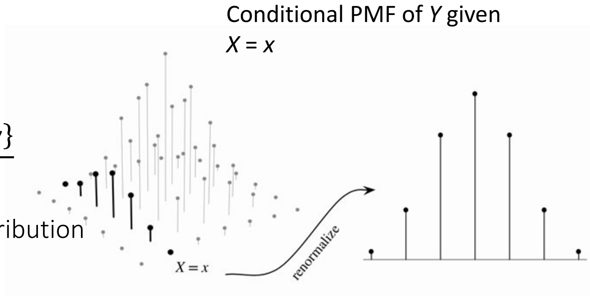
Appendix 8.A – Multivariate Distributions: Conditional Distributions

- For discrete RVs X and Y, the conditional PMF of Y given $X = x$ is given by:

$$P\{Y = y|X = x\} = \frac{P\{X = x, Y = y\}}{P\{X = x\}}$$

(we observe the value of X and want to update our distribution of Y to reflect this information)

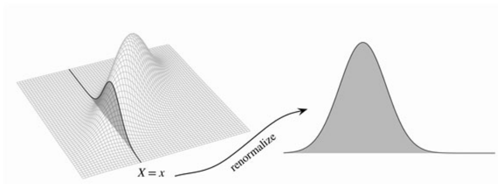
Conditional PMF of Y given $X = x$



- It is possible to obtain the conditional PMF of X given $Y = y$ also using Bayes' rule or the law of total probability (LOTP):

$$P\{Y = y|X = x\} = \frac{P\{X = x|Y = y\}\, P\{Y = y\}}{P\{X = x\}}$$

$$P\{X = x\} = \sum_{y} P\{X = x|Y = y\}\, P\{Y = y\}$$



J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 7.1

Conditional distributions:

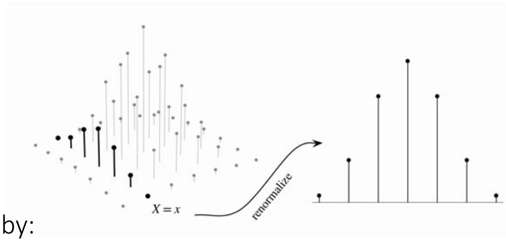
just note that $P\{X = x\}$ with which we renormalize is simply the marginal PMF defined two slides before!

Same thing for $f_X(x)$ on the next slide with respect to the definition of marginal PDF for continuous RVs.

Appendix 8.A – Multivariate Distributions: Conditional Distributions

- For continuous RVs X and Y with joint PDF $f_{X,Y}$, the conditional PDF of Y for $X = x$ is given by:

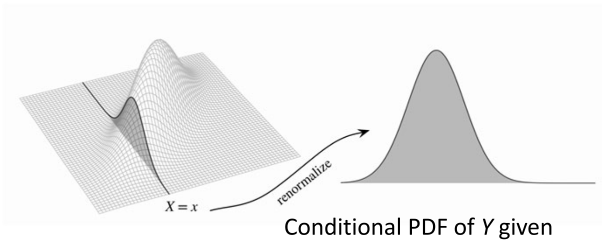
$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$



- The continuous analogs of Bayes' rule or the LOTP are given by:

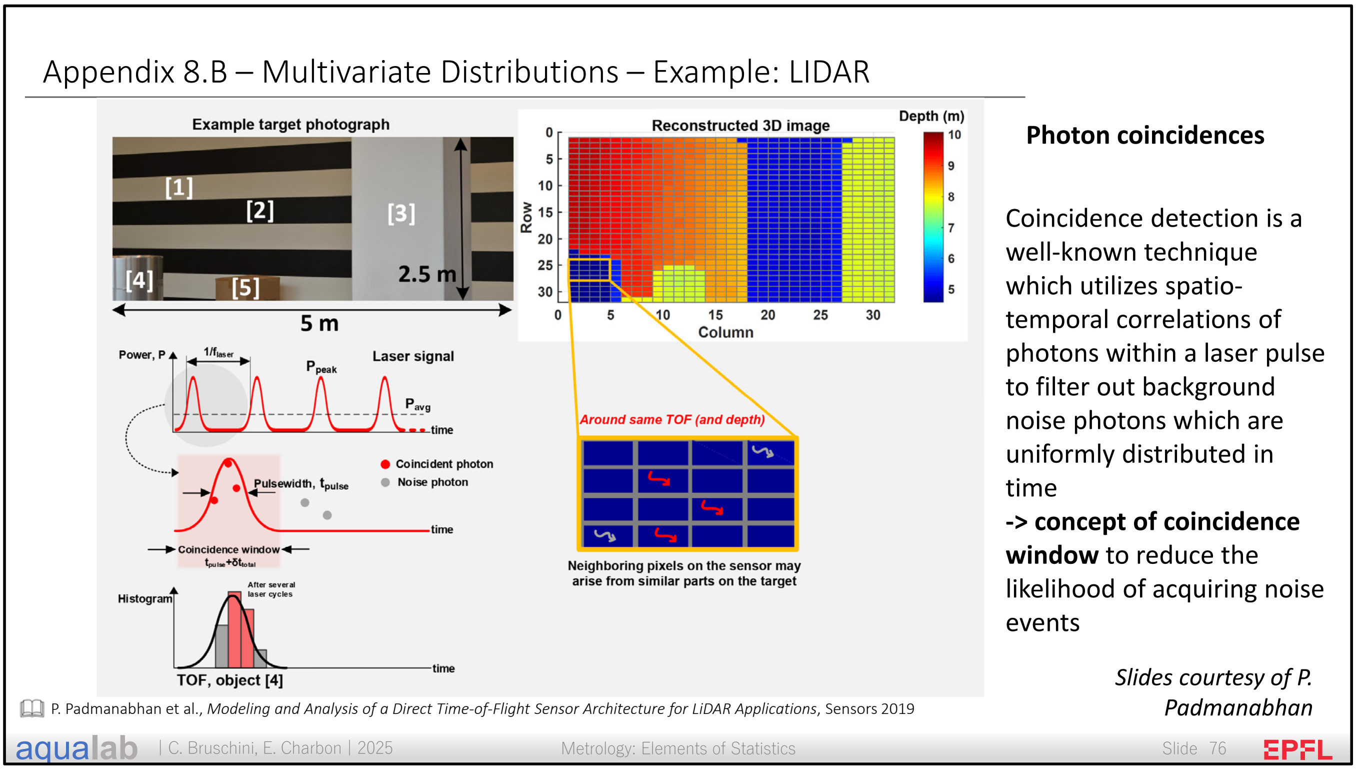
$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y)\, f_Y(y)}{f_X(x)}$$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y)\, f_Y(y)\, dy$$



Conditional PDF of Y given $X = x$

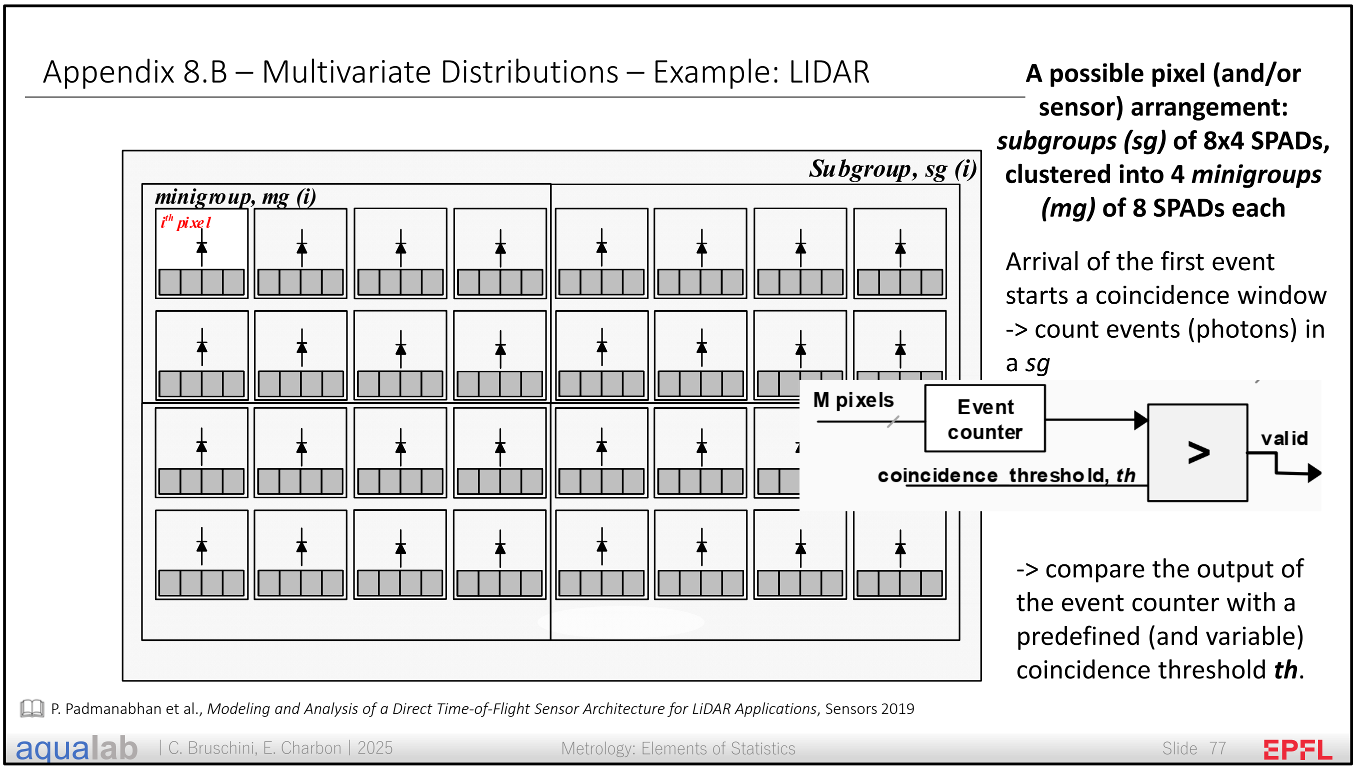
J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 7.2



An example of a multivariate distribution as applied to the **measurement of time-of-flight** and the design of a bespoke sensor is provided here in Appendix 8.B, as a complement to slide 8.4.3, for those who would like to know more.

Coincidence detection is a well-known technique which utilizes spatio-temporal correlation of photons within a laser pulse to filter out background noise photons which are uniformly distributed in time. The figure above conceptually explains this technique with an example scene and a measured 3D image reconstruction.
The main idea is to exploit the fact that the signal photons reflected from the target are temporally correlated and thus, most likely to be concentrated within a time-window coarsely equal to the total system full width at half maximum, *FWHM,* of the laser pulse….

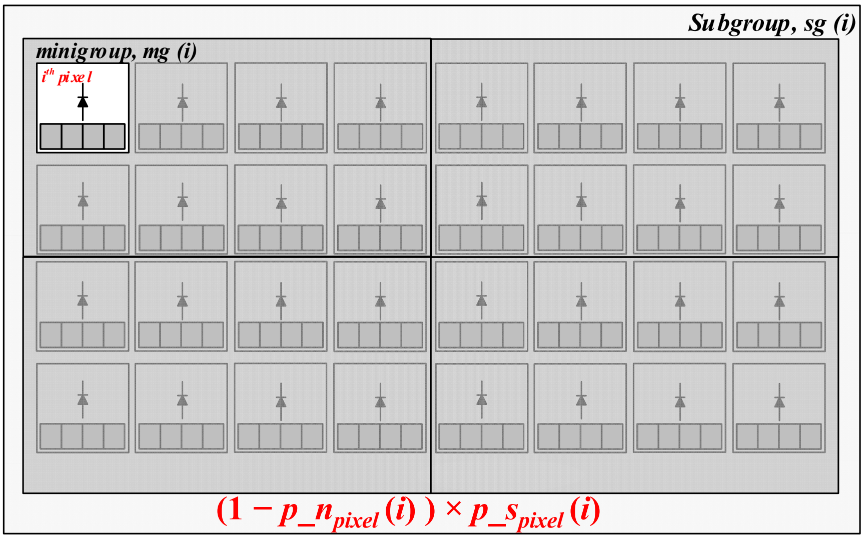
Instead of letting the sensor integrate events over a long measurement window, imposing this time constraint, referred to as the "coincidence window", reduces the likelihood of acquiring noise events whose probability of occurrence within that window is very low, thus, electrically enhancing the signal to background noise ratio, SBR. Coincidence may be implemented at the sensor level over clusters/groups of closely-spaced pixels, exploiting a "more-likely" fact that neighboring pixels may belong to similar target depths (and thus, TOFs), as depicted in the figure for the object labelled [4].



The sensor may be visualized as an array of modules, called, subgroups, where every subgroup is clustered into an array of $M = 4 \times 8$ SPADs in this example capable of performing photon detection. The subgroup, *sg* (M pixels), is further clustered into N minigroups, *mg*, comprising of (M/N) number of pixels each.

Arrival of the first event starts a coincidence window, *twindow.* There is an event counter in every subgroup which tracks the number of photons within a coincidence window. A comparator logic is used to compare the output of the event counter with a predefined (and variable) coincidence threshold, *th*. Whenever the event count exceeds *th*, a signal is considered valid.

Appendix 8.B – Multivariate Distributions – Example: LIDAR



(a) Detect 1st signal photon at *ith* pixel

Mathematically:

$p\_s_{th}(i)$ = P(detecting *th* number of valid signal events within *t_window*) =

Probability of detecting a signal event in a pixel *i*, = $p\_s_{pixel}(i)$,

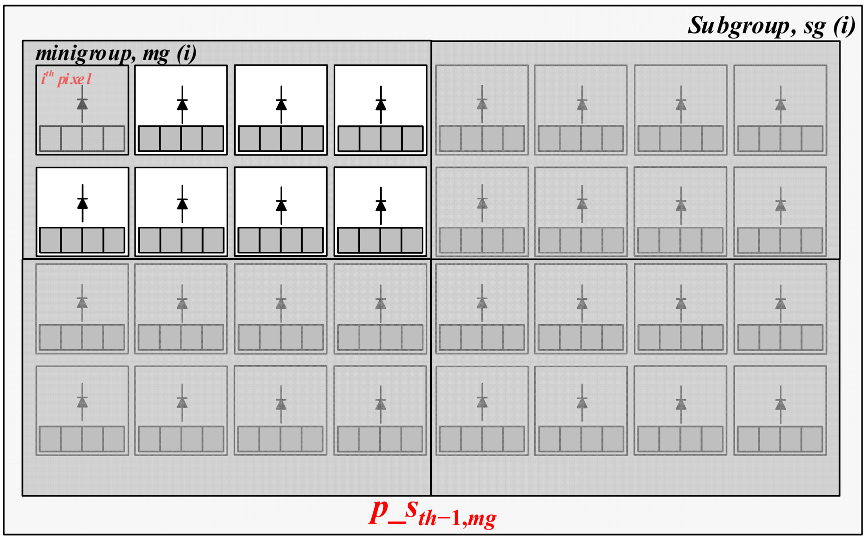
given that no noise photon is detected at pixel *i*, = $(1 - p\_n_{pixel}(i))$,

and…

s = signal, n = noise

P. Padmanabhan et al., *Modeling and Analysis of a Direct Time-of-Flight Sensor Architecture for LiDAR Applications*, Sensors 2019

Appendix 8.B – Multivariate Distributions – Example: LIDAR



(b) Detect (*th-1*) photons in *mg(i)*

… and $p\_s_{th-1}, sg(i)$ = P(detecting *th* - 1 signal events in the rest of the subgroup).

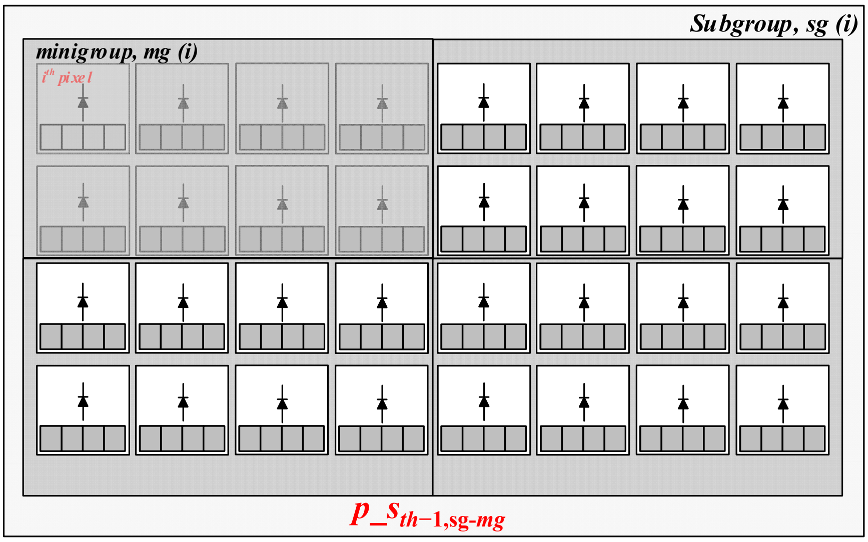
But $p\_s_{th-1}, sg(i)$ =

union operation of individual probabilities of detecting (*th* - 1) signal photons in the minigroup *mg*(*i*) = $p\_s_{th-1}, mg(i)$,

…

P. Padmanabhan et al., *Modeling and Analysis of a Direct Time-of-Flight Sensor Architecture for LiDAR Applications*, Sensors 2019

Appendix 8.B – Multivariate Distributions – Example: LIDAR



(c) Detect (*th-1*) photons in *sg(i) - mg(i)*

… or in the rest of the subgroup, *sg*(*i*) - *mg*(*i*),

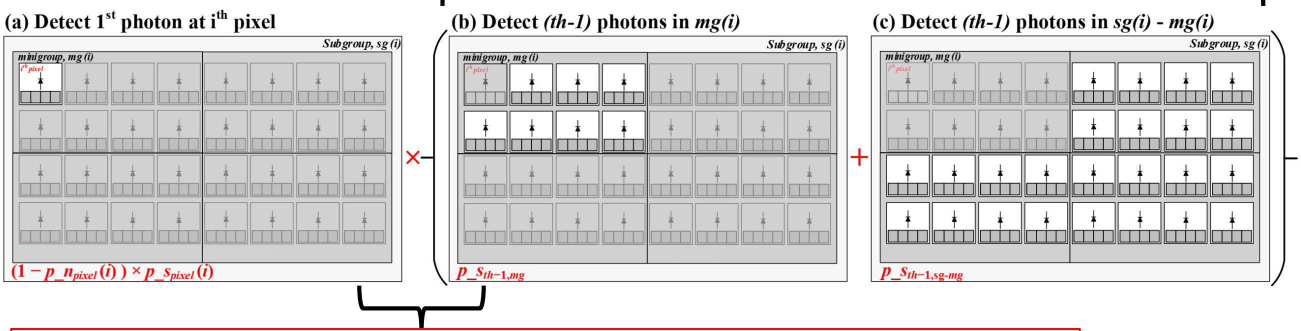
= $p\_s_{th-1}, sg(i)-mg(i)$

P. Padmanabhan et al., *Modeling and Analysis of a Direct Time-of-Flight Sensor Architecture for LiDAR Applications*, Sensors 2019

Summarising:

Union operation over (*th-1*) photons

$$p\_s_{th-1,sg}(i) = p\_s_{th-1,mg} \cup p\_s_{th-1,sg-mg}$$



$$p\_s_{th}(i) = (1 - p\_n_{pixel}(i)) \times p\_s_{pixel}(i) \times p\_s_{th-1,sg}(i)$$

P. Padmanabhan et al., *Modeling and Analysis of a Direct Time-of-Flight Sensor Architecture for LiDAR Applications*, Sensors 2019

The probability of detecting "th" number of valid signal events within *twindow* can be calculated as a conditional probability of detecting a signal event in a pixel, *i*, given that no noise photon is detected at pixel *i* and (*th* - 1) signal events are detected in the rest of the subgroup.

→ The final conditional probability of detecting th-1 signal photons in the subgroup…..