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

Metrology Course Structure

Measurement Science & Technology

Optical Metrology *(Weeks 2-4)*	Nanoscale *(Weeks 5-7)*	Probability and Statistics *(Weeks 8-9)*	Electrical Metrology *(Weeks 10-12)*	Quantum *(Weeks 13-14)*
↓	↓	↓	↓	↓
Optical Image Sensors SPADs Microscopy	AFM SEM Nanomag.	Theory & Practical Examples	Noise ADCs (amplitude) TDC (time)	Quantum SI Quantum Applications Supercond.

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

Outline

8.1 Introduction to Probability

8.2 Random Variables

8.3 Moments

8.4 Covariance and Correlation

9.1 Random Processes

9.2 Central Limit Theorem

9.3 Estimation Theory

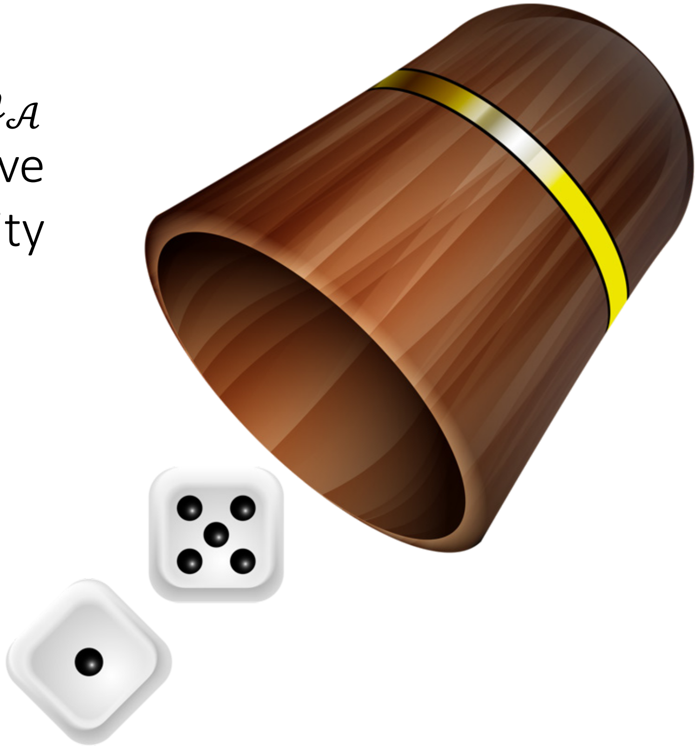
9.4 Accuracy, Precision and Resolution

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

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

Outline

8.1 Introduction to Probability

8.2 Random Variables

8.3 Moments

8.4 Covariance and Correlation

9.1 Random Processes

9.2 Central Limit Theorem

9.3 Estimation Theory

9.4 Accuracy, Precision and Resolution

8.2 Random Variables

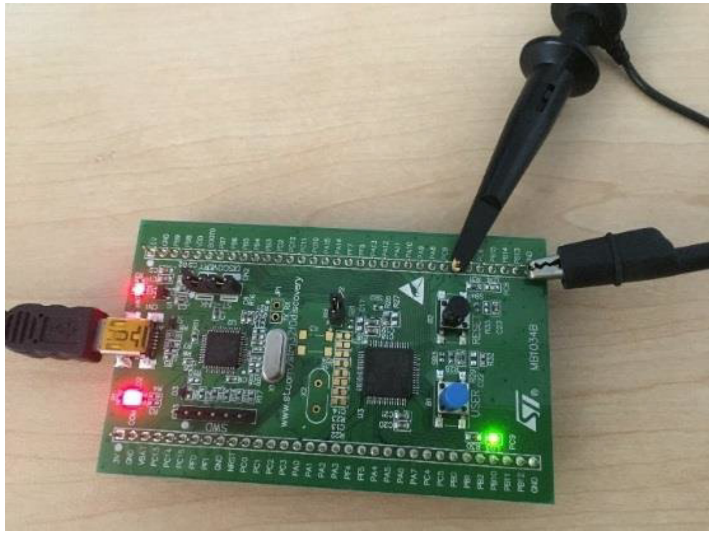
- A Random Variable (RV) is a number $X(s)$ assigned to every outcome s of an experiment.

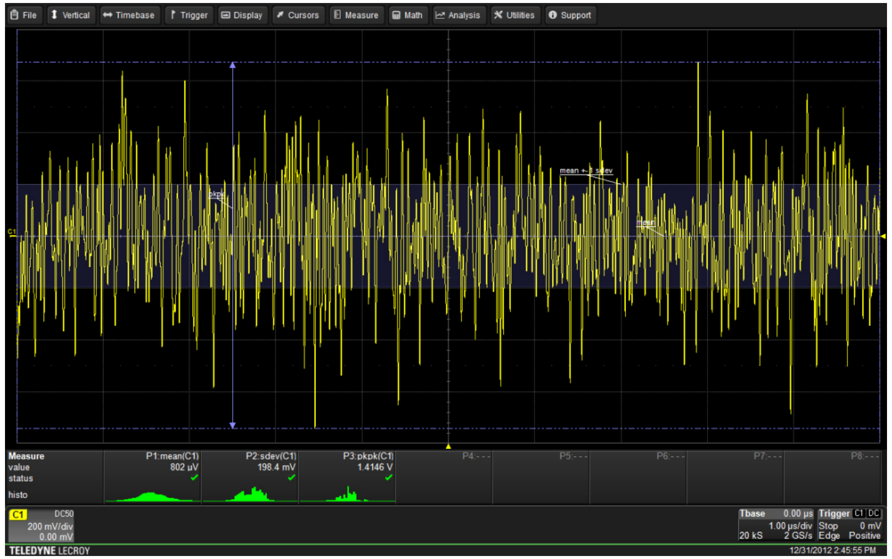
  Examples: the voltage of a random source, etc..

- 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:

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



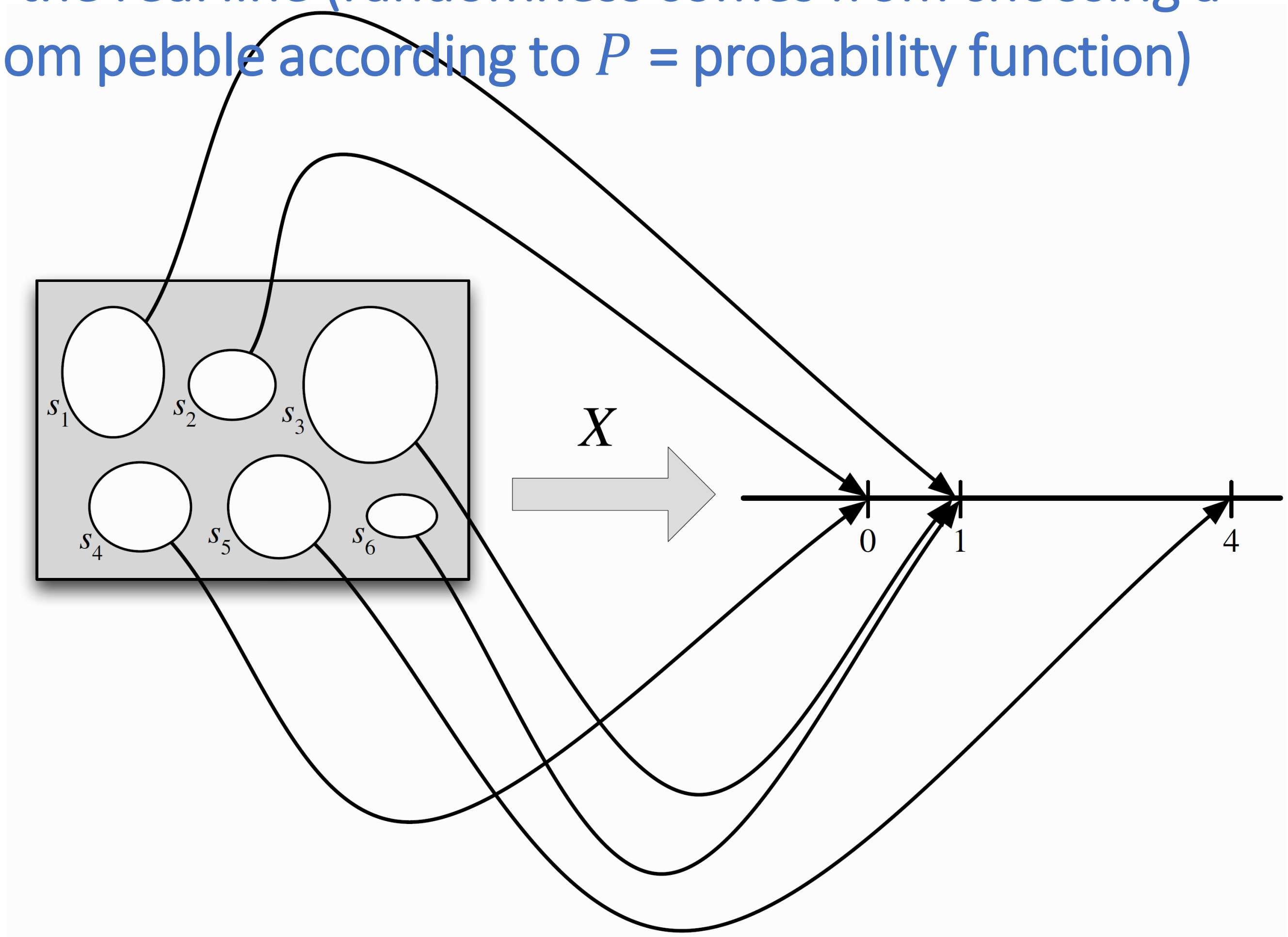


A. Papoulis, *Probability, Random Variables and Stochastic Processes*, 3rd ed., 1991, Chap. 4.1

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

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:

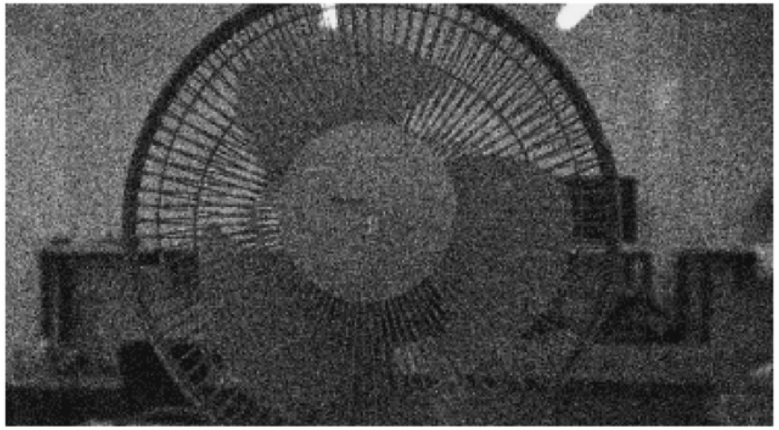
$$\mathcal{S} = \{a_1, a_2, \dots, a_n\}$$

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

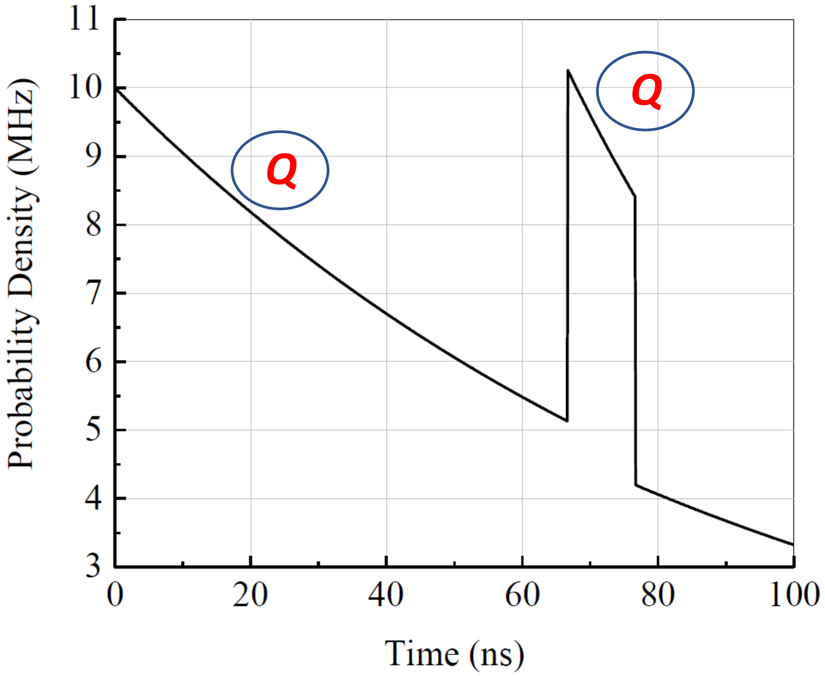
- A Random Variable X is instead said to be continuous if it can take on any value in a given interval, possibly of infinite length. For example its support can be:

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

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.*

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:

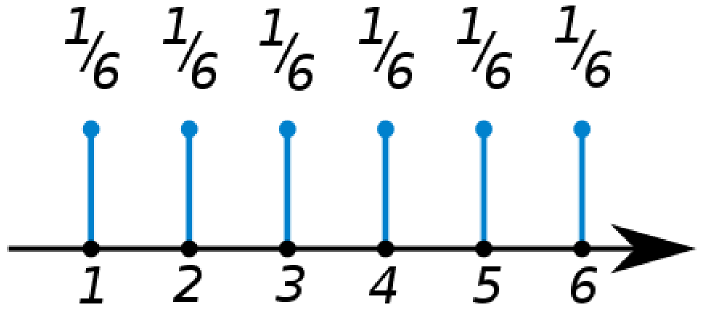
$$\text{PMF: } p_X(x) = P\{X = x\}$$

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

- The PMF needs to satisfy two criteria:

1. Nonnegative:

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



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

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

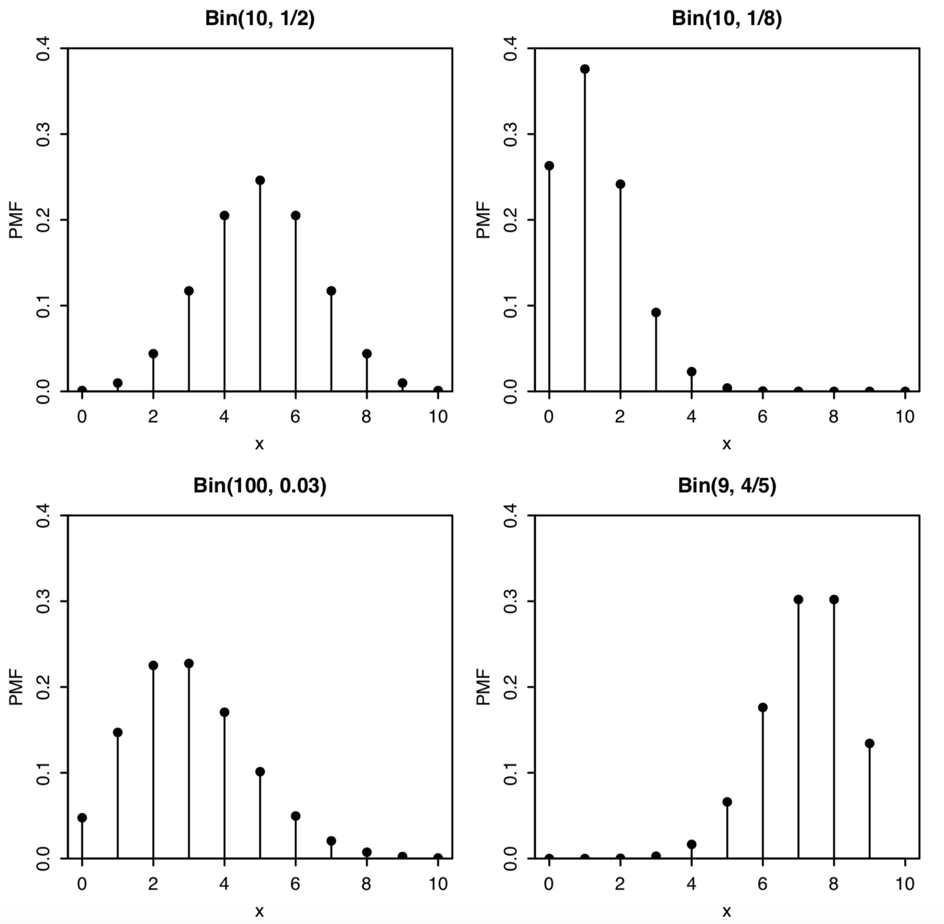
- A discrete RV X is said to have the Bernoulli distribution with parameter p if $P\{X = 1\} = p$ and $P\{X = 0\} = 1 - p$, where $0 < p < 1$.

- 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 $\mathrm{Bin}(n, p)$ 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

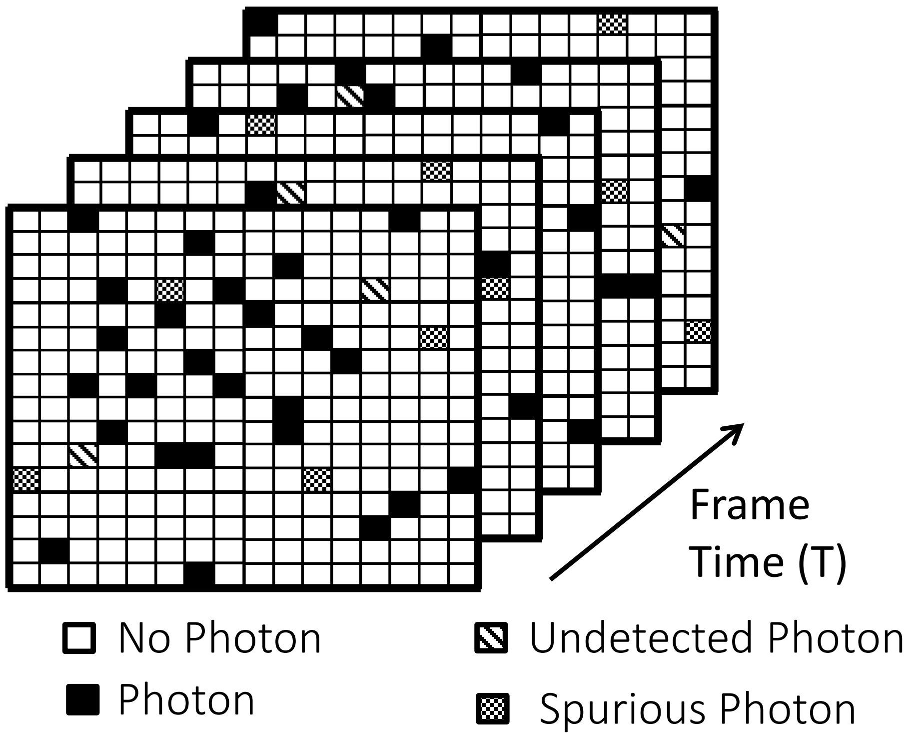
8.2.2 Bernoulli RV – Example

(CMOS) SPAD:
Single-Photon
Avalanche
Photodiode

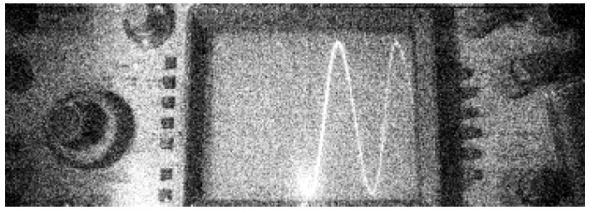
+ time-of-arrival, energy/wavelength, polarization, etc.

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



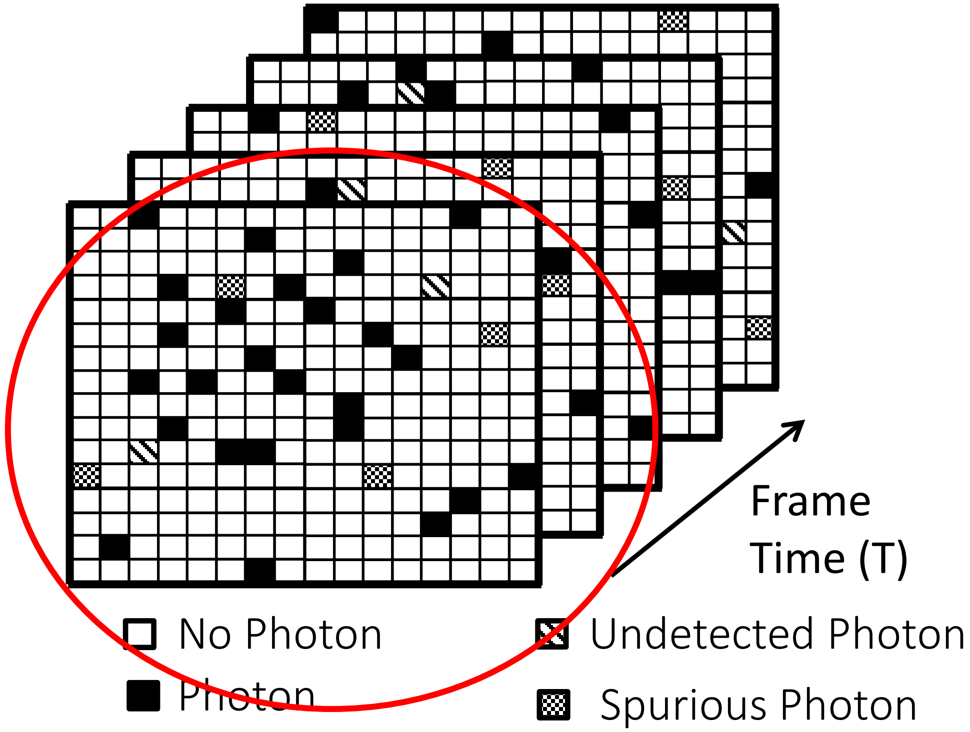
4-bit frame

📖 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

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

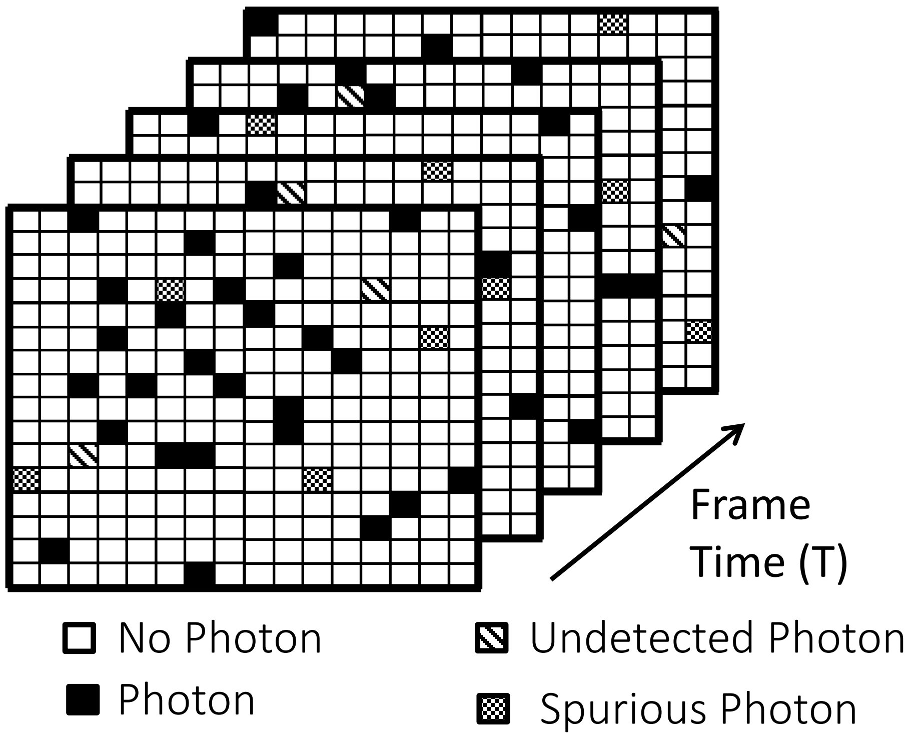
Q # of photons at each pixel: $P\{Z=k\}=\dfrac{e^{-\phi\tau\eta}(\phi\tau\eta)^k}{k!} \Longrightarrow$

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

Ex

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

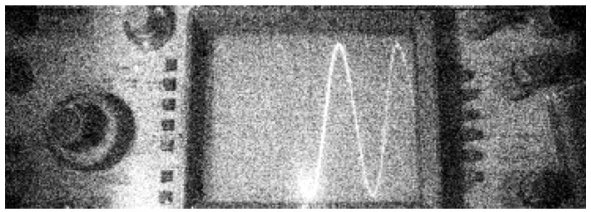
(CMOS) SPAD:
Single-Photon
Avalanche
Photodiode



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

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

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



Y. Hirose, MDPI Sensors(18), 2018

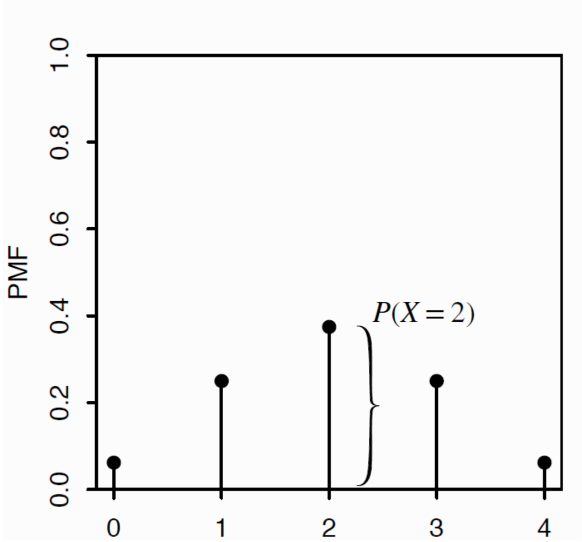
8.2.2 Binomial RV – Example

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

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

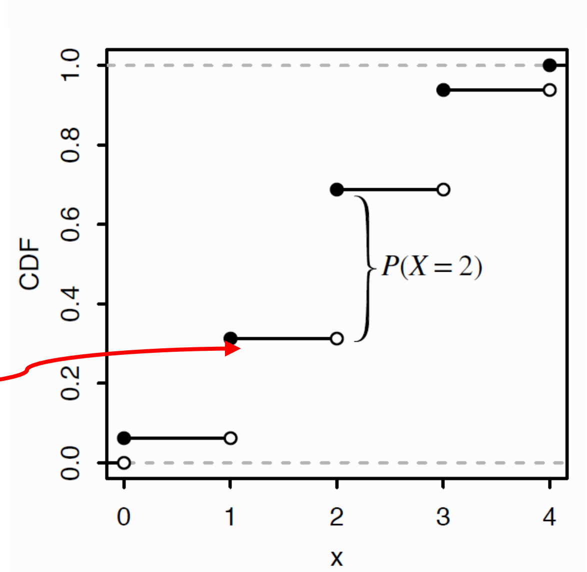


Example: Let X be $\mathrm{Bin}(4, 1/2)$. The cumulative distribution function can be calculated from the probability mass function.

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 $F_X(x_1) \leq F_X(x_2)$

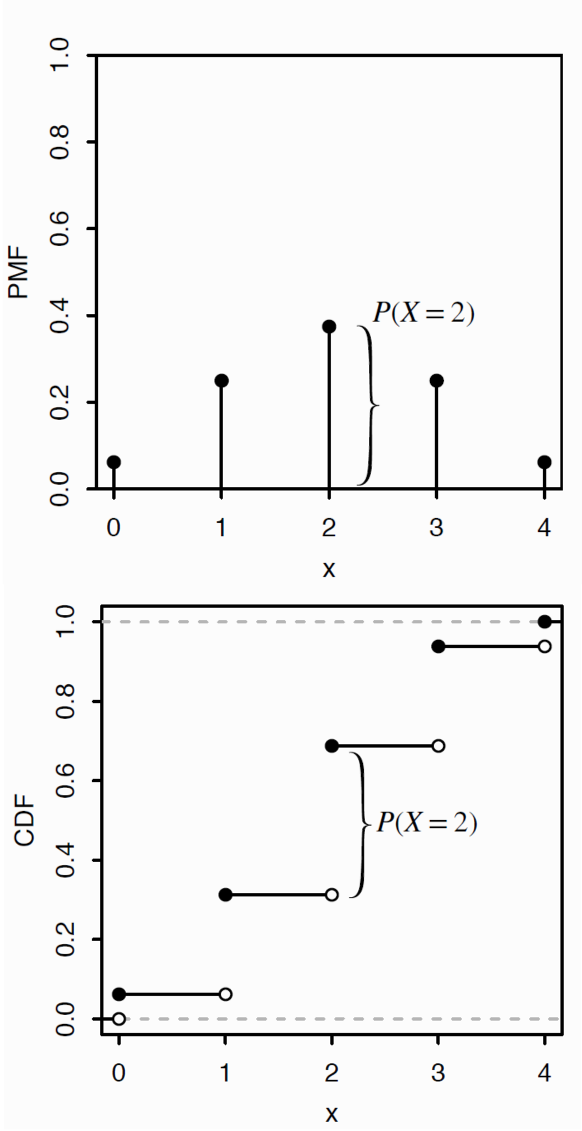
 2. Right-continuous: The CDF is continuous except possibly for some jumps. When there is a jump, the CDF is continuous from the right, i.e. for any a:

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

 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

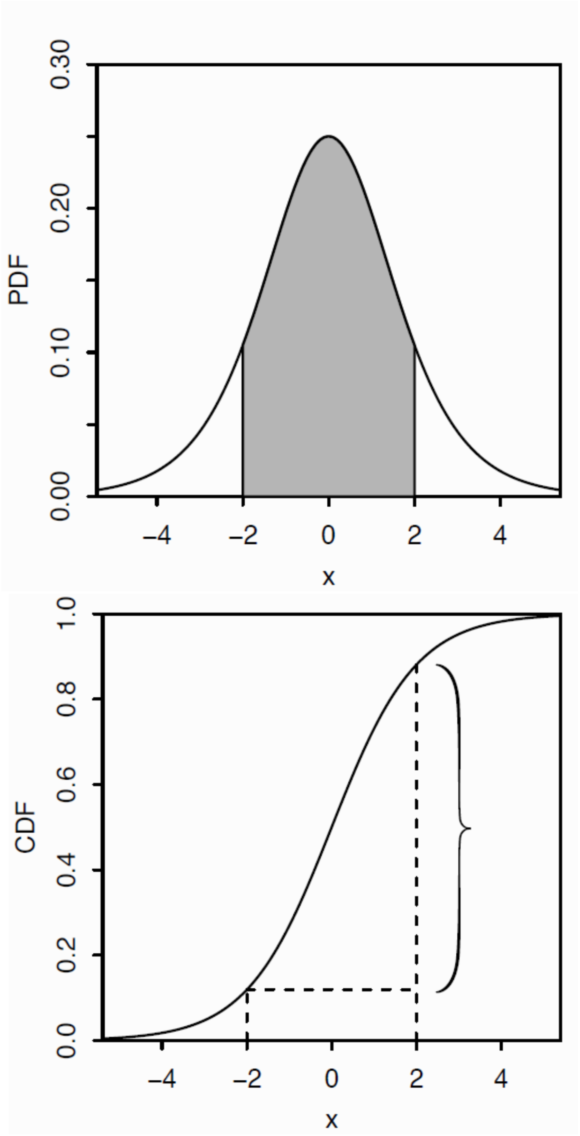
- For a continuous RV X with cumulative distribution function F_X, the probability density function (PDF) f_X is the derivative of the cumulative distribution function (CDF):

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

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

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.

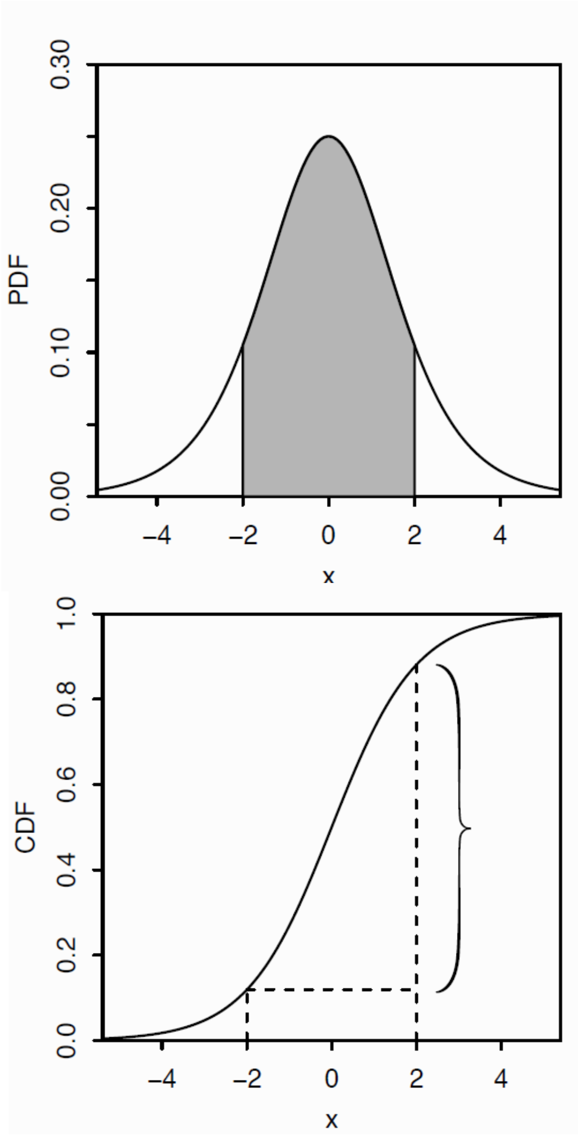
- For a PDF to be valid, two criteria must be met:

1. Nonnegative:

$$f_X(x) \geq 0$$

2. Integrates to 1:

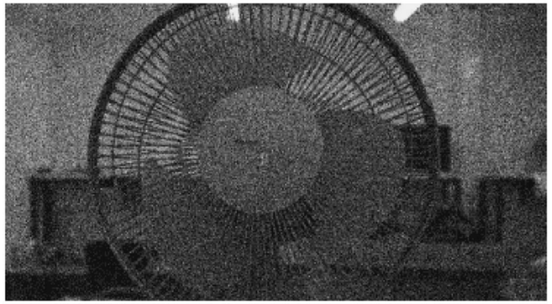
$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1$$



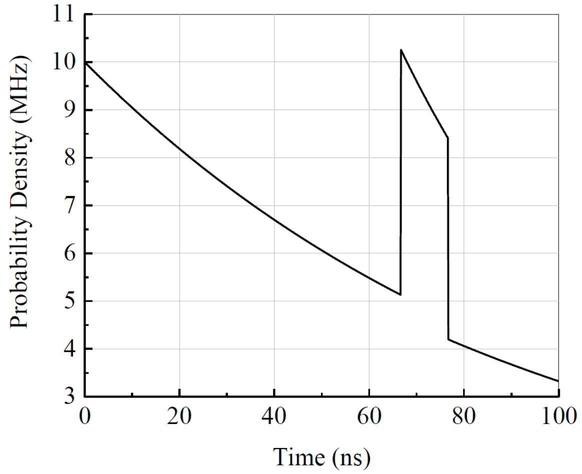
📖 J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 5.1

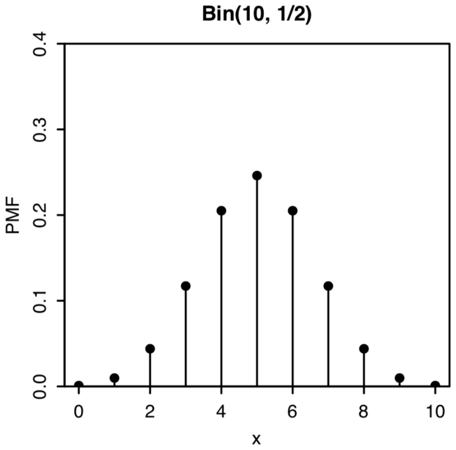
Take-home Messages/W8-1

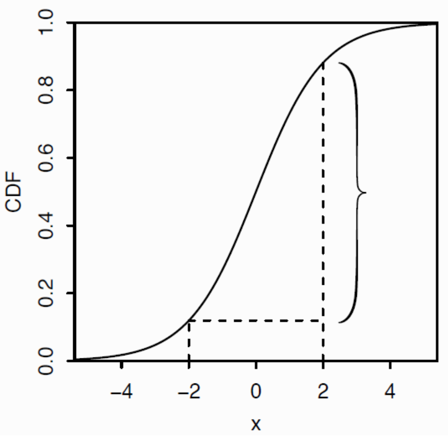
- *Introduction to probability (see also Appendix 8.1):*
  - Basic definitions, conditional probability
  - Bayes' rule, law of total probability, independence of events
- *Random Variables (RVs):*
  - Examples (discrete/continuous)
  - Probability Mass Function (PMF), Cumulative Distribution Function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



4-bit, 4.4 kfps







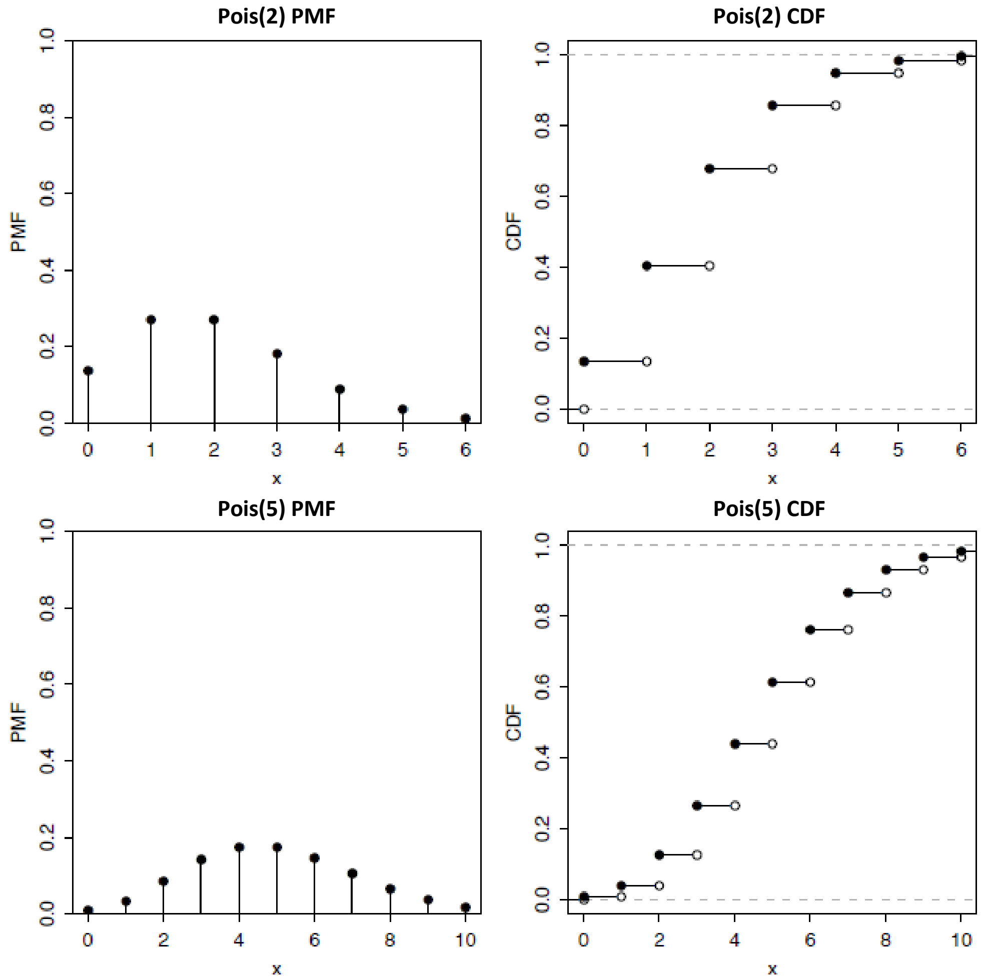
8.2.5 Poisson RV

- 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

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:

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

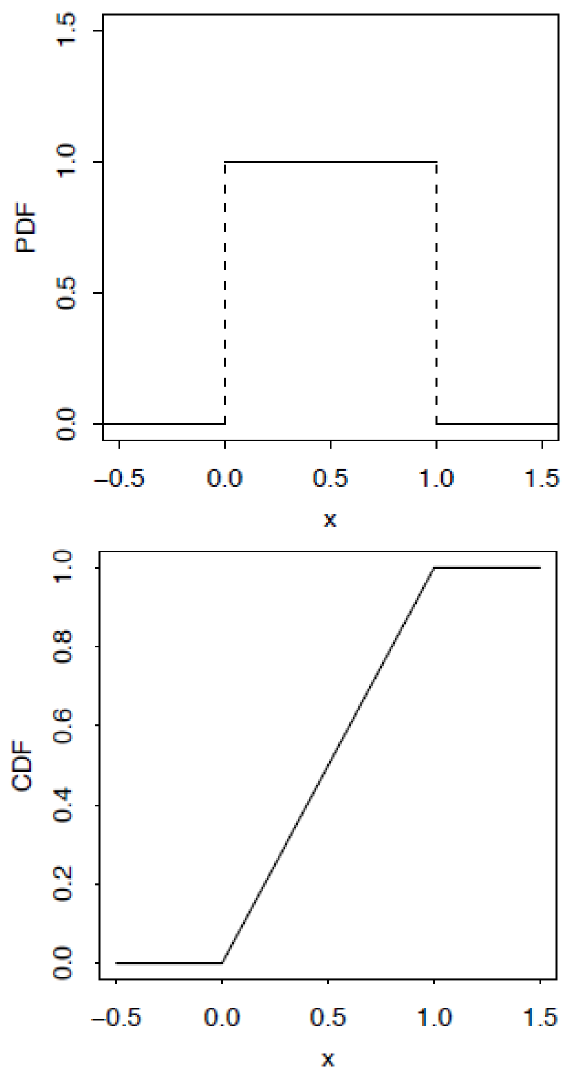
$$U \sim \text{Unif}(a, b)$$

- 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:

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

Unif(0,1) PDF & CDF



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

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:

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

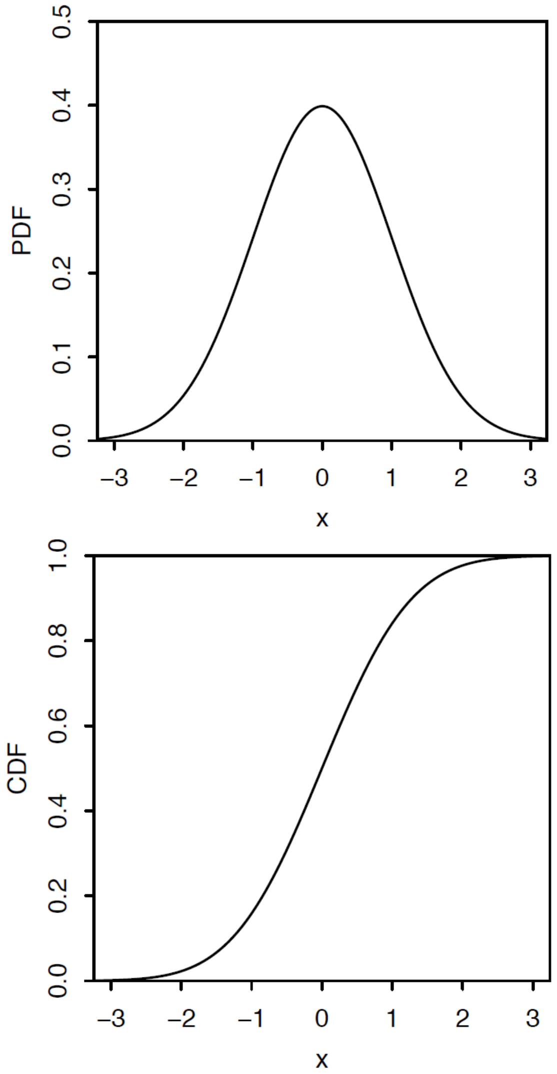
- In the special case of $\mu = 0$ and $\sigma = 1$, the distribution takes the name of standard Normal distribution. We will write it as $Z \sim \mathcal{N}(0,1)$. The standard Normal PDF and CDF are:

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

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

No closed form exists!

Standard Normal PDF/CDF



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

8.2.7 Normal (Gaussian) RV (contd.)

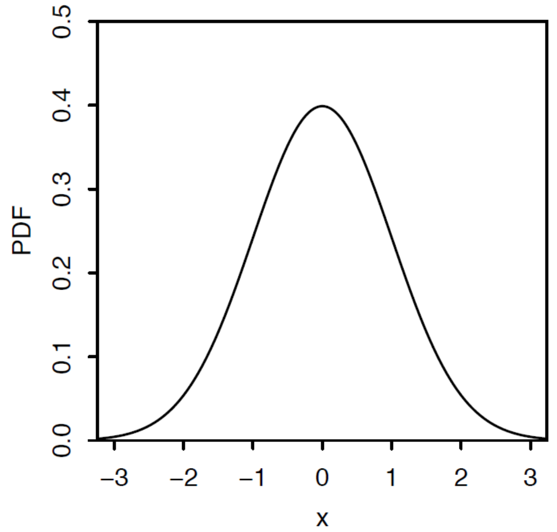
- The standard Normal distribution has the following properties:

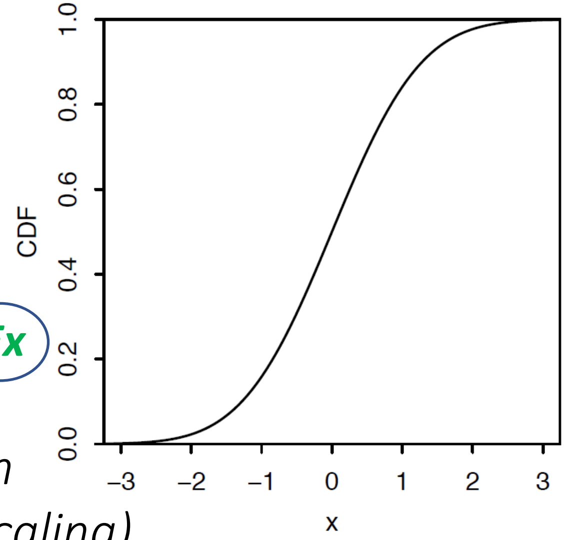
  1. Symmetry of the PDF: φ satisfies $\varphi(z) = \varphi(-z)$
  2. Symmetry of the tail area: the area under the PDF to the left of $-z$ and to the right of z is equal. Using the CDF:
  $$\Phi(z) = 1 - \Phi(-z)$$
  3. Symmetry of Z and $-Z$: If $Z \sim \mathcal{N}(0,1)$, then $-Z \sim \mathcal{N}(0,1)$ 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

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.

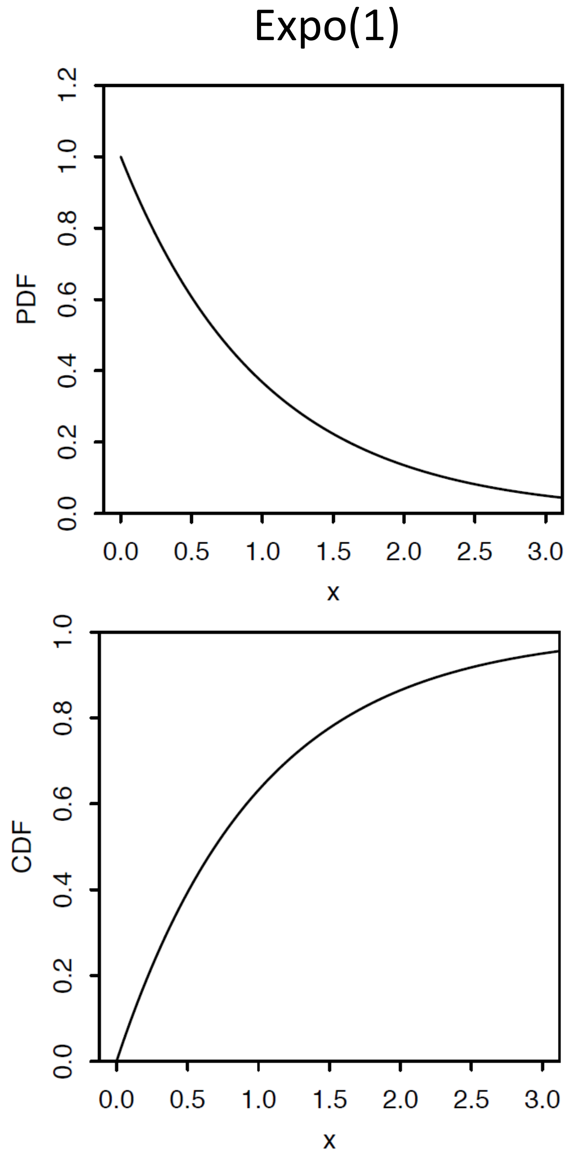
- A continuous RV X is said to have an exponential distribution (we will write $X \sim \mathrm{Expo}(\lambda)$) with parameter λ if its PDF is:

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

- The corresponding CDF is

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

Ex

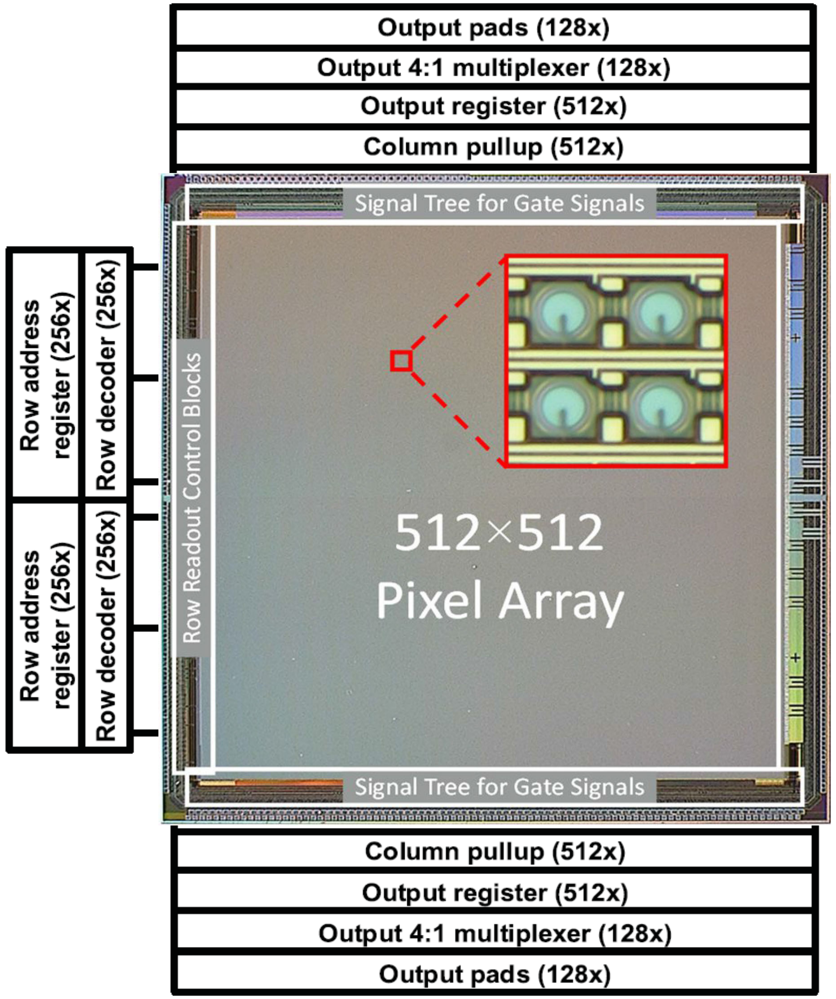


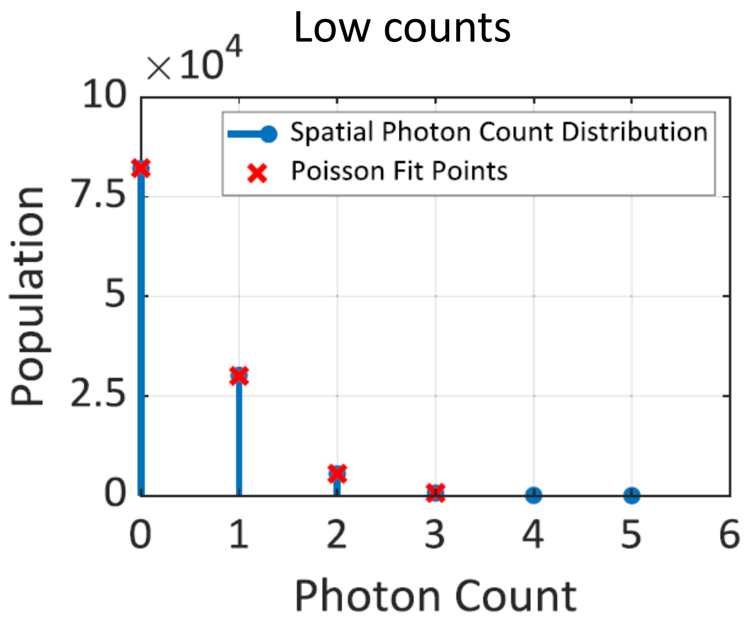
J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 5.5

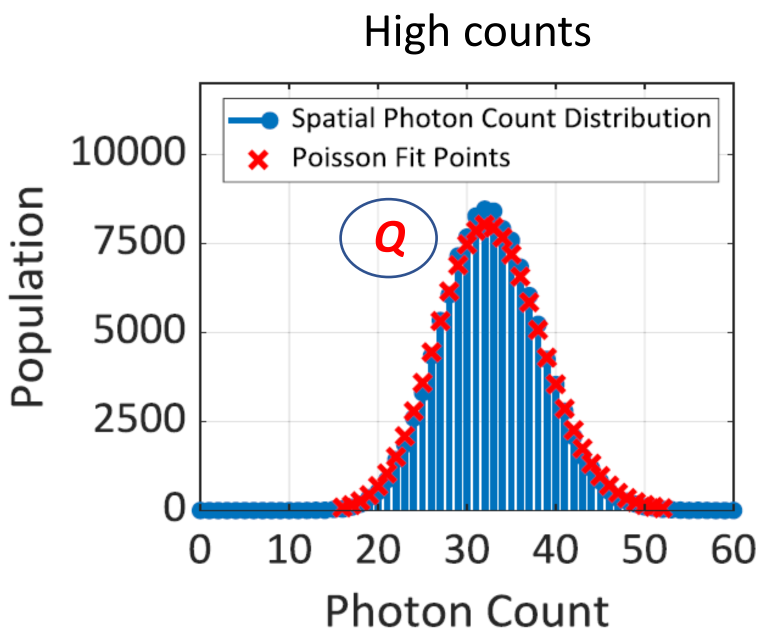
8.2.9 Example 1: Photon-flux dependent distributions

SwissSPAD2 binary SPAD imager

(intensity)

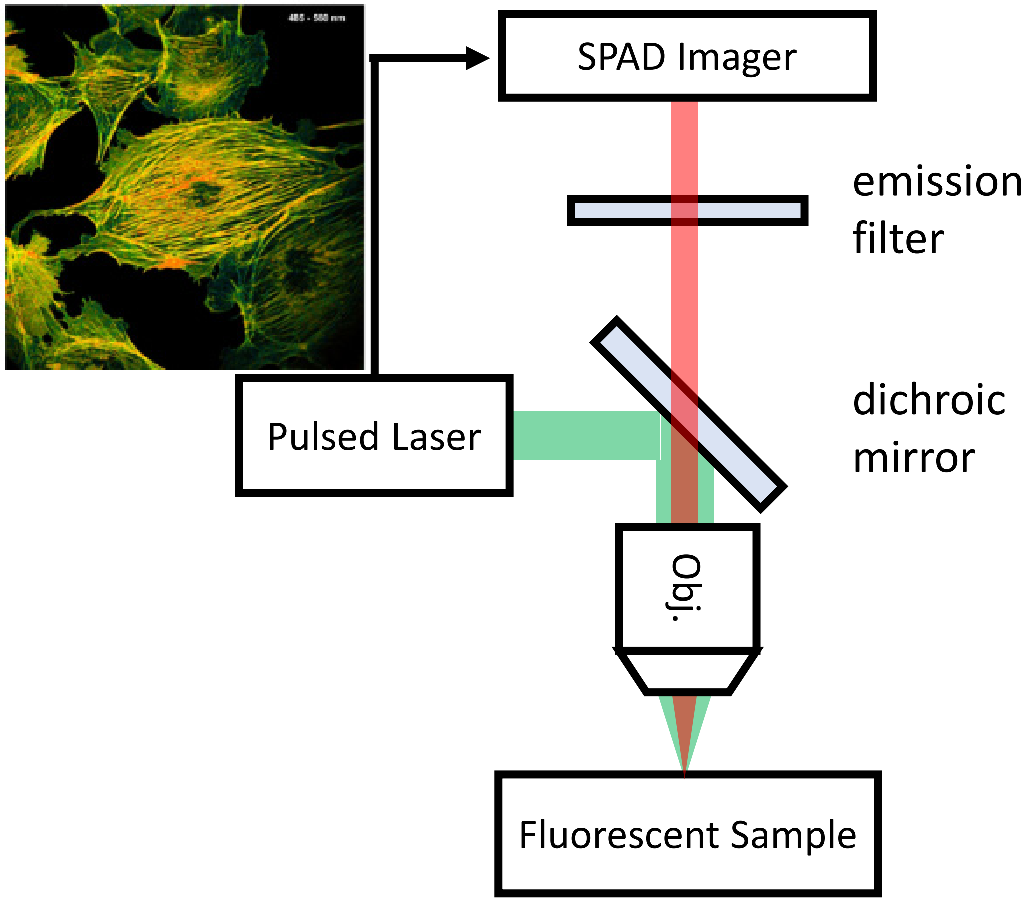




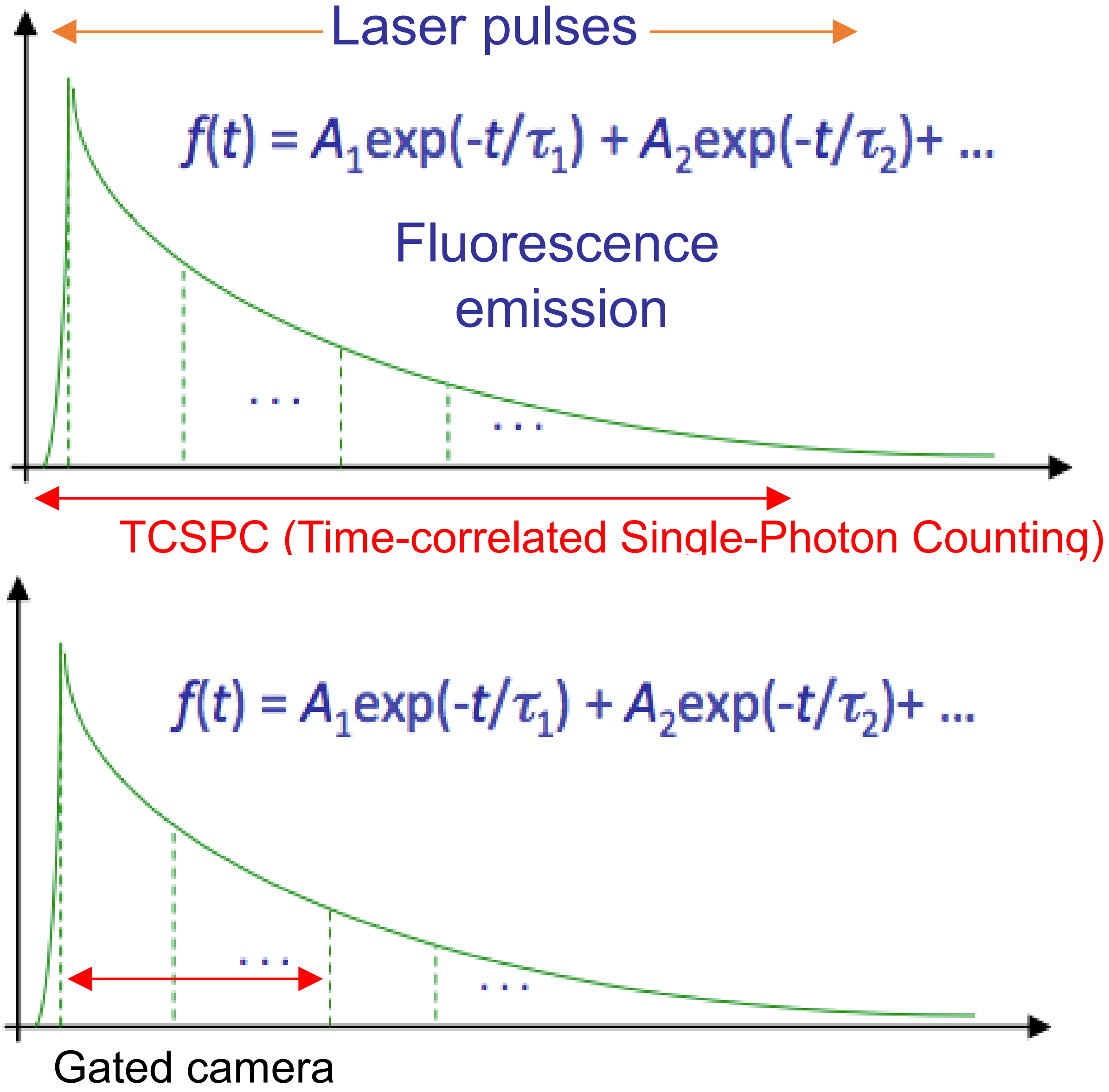


A. Ulku *et al., A 512x512 SPAD Image Sensor with Integrated Gating for Widefield FLIM.* IEEE JSTQE (2019).

8.2.9 Example 2: Fluorescence Lifetime – Time-Resolved



Lifetime images: the pixel time-tags all photons and calculates t_1 , t_2 , A_1

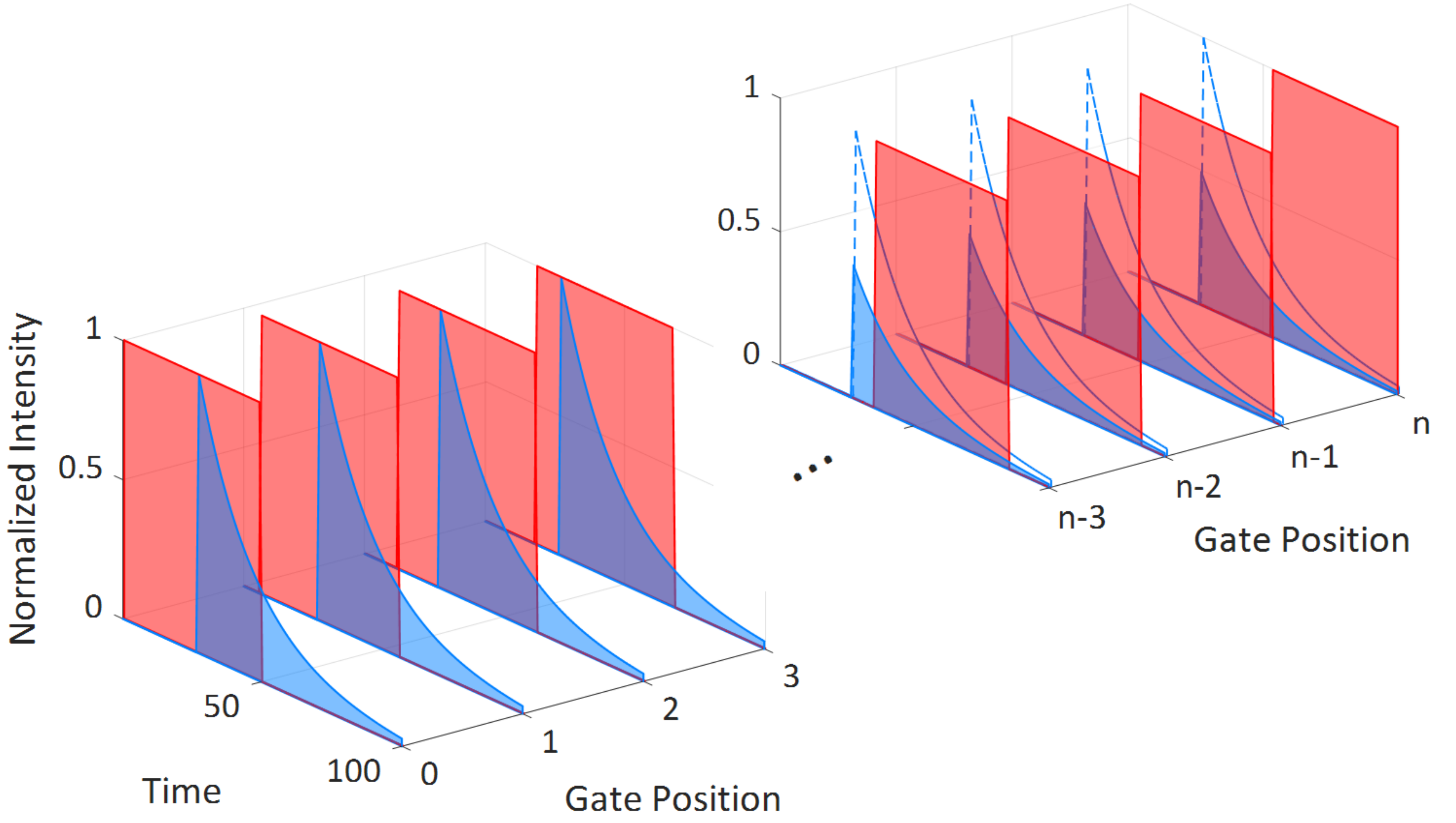


D. Li, *Strathclyde Univ.* (2016).

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).

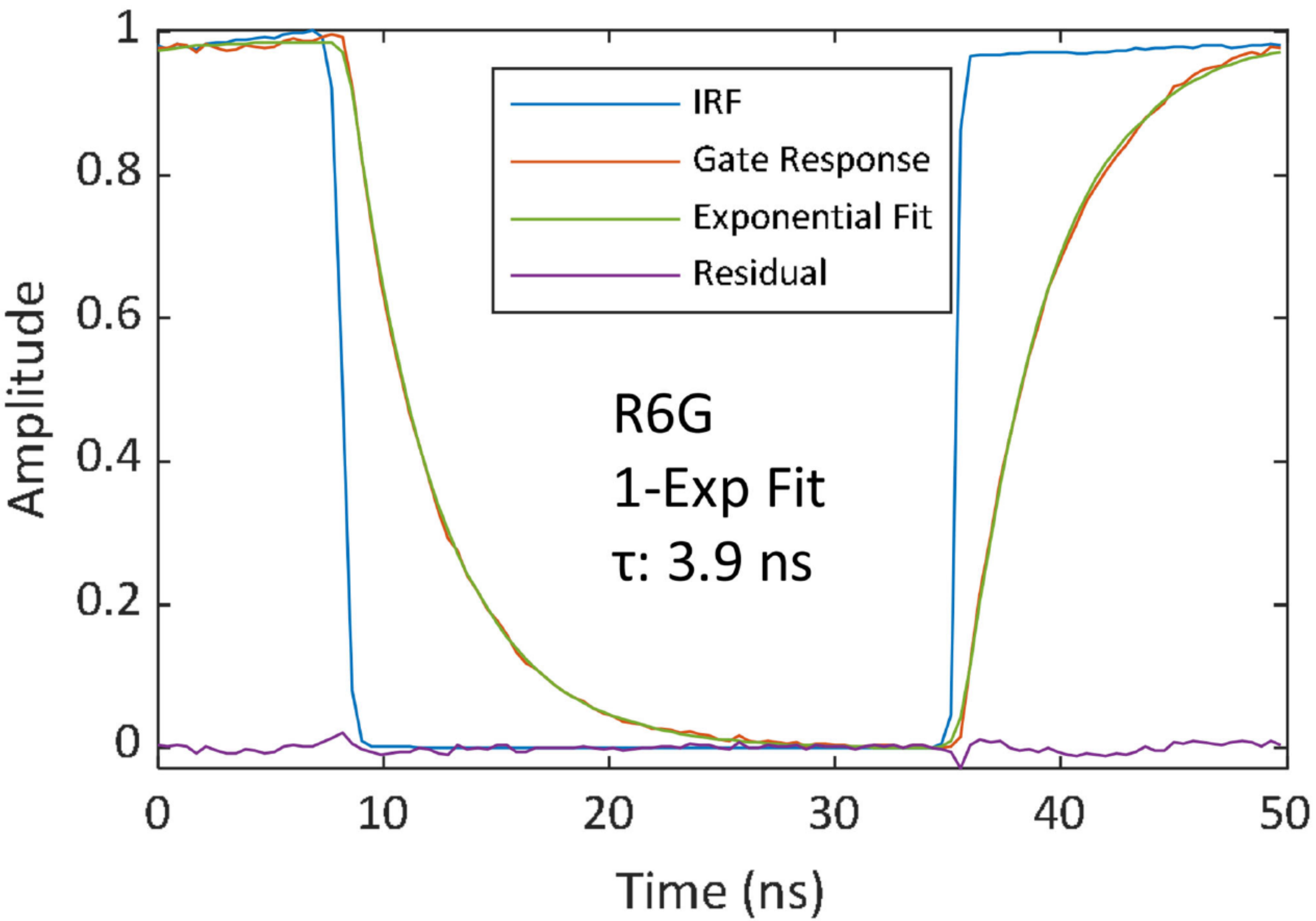
8.2.9 Example 2: Fluorescence Lifetime – Time-Resolved

SwissSPAD2 binary SPAD imager

(overlapping gates -> convolution)

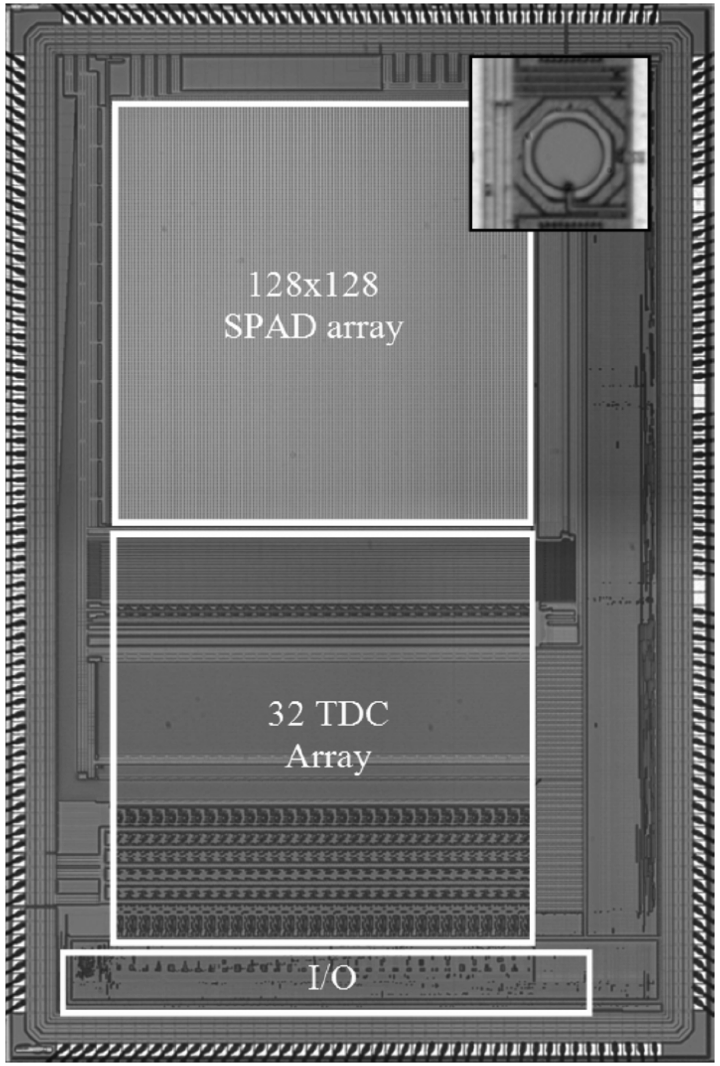
$$f(t) = g(t) * IRF(t)$$

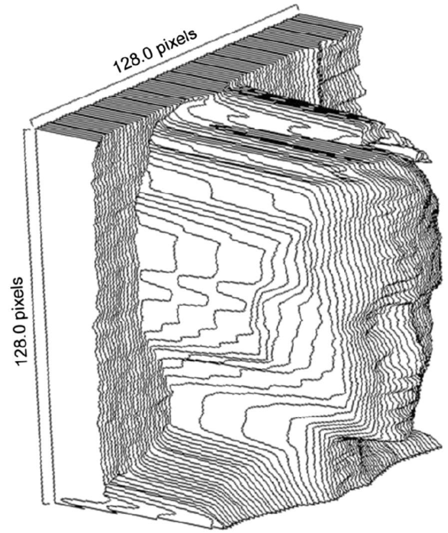
IRF: Instrument Response Function

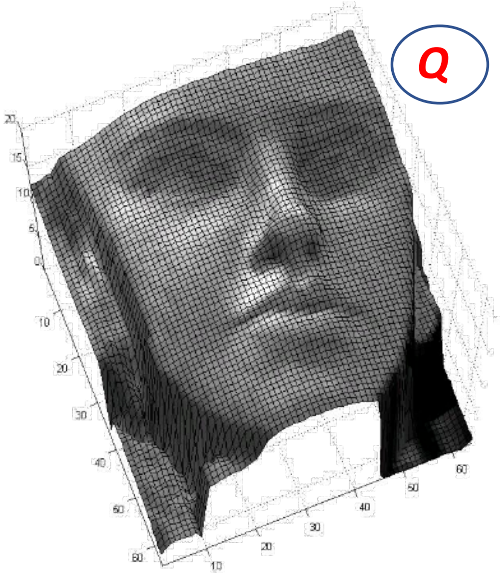


A. Ulku *et al., A 512x512 SPAD Image Sensor with Integrated Gating for Widefield FLIM.* IEEE JSTQE (2019).

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

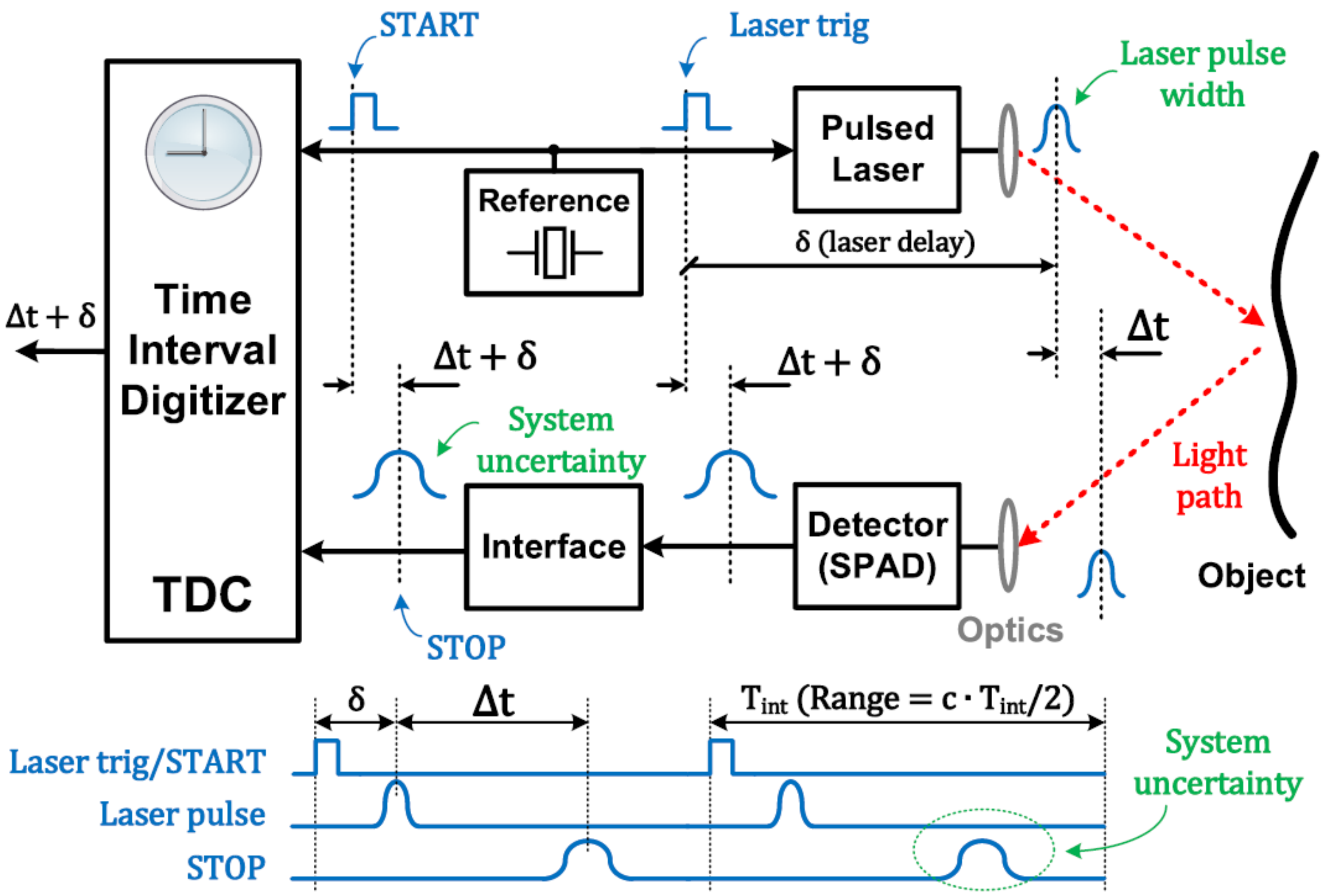






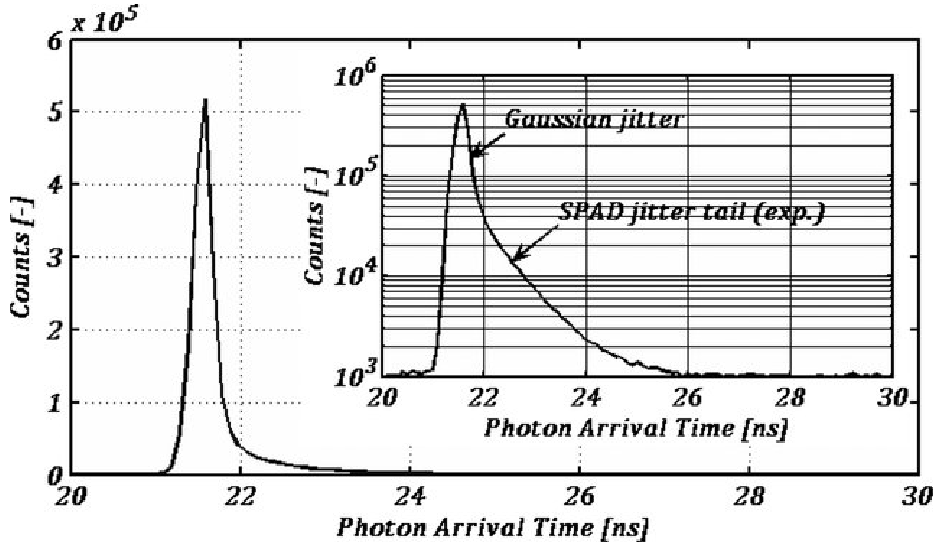
C. Niclass *et al., A 128×128 Single-Photon Image Sensor With Column-Level 10-Bit Time-to-Digital Converter Array.* IEEE JSSC 43 (2008).

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



Direct SPAD illumination -> SPAD IRF (jitter noise) ->

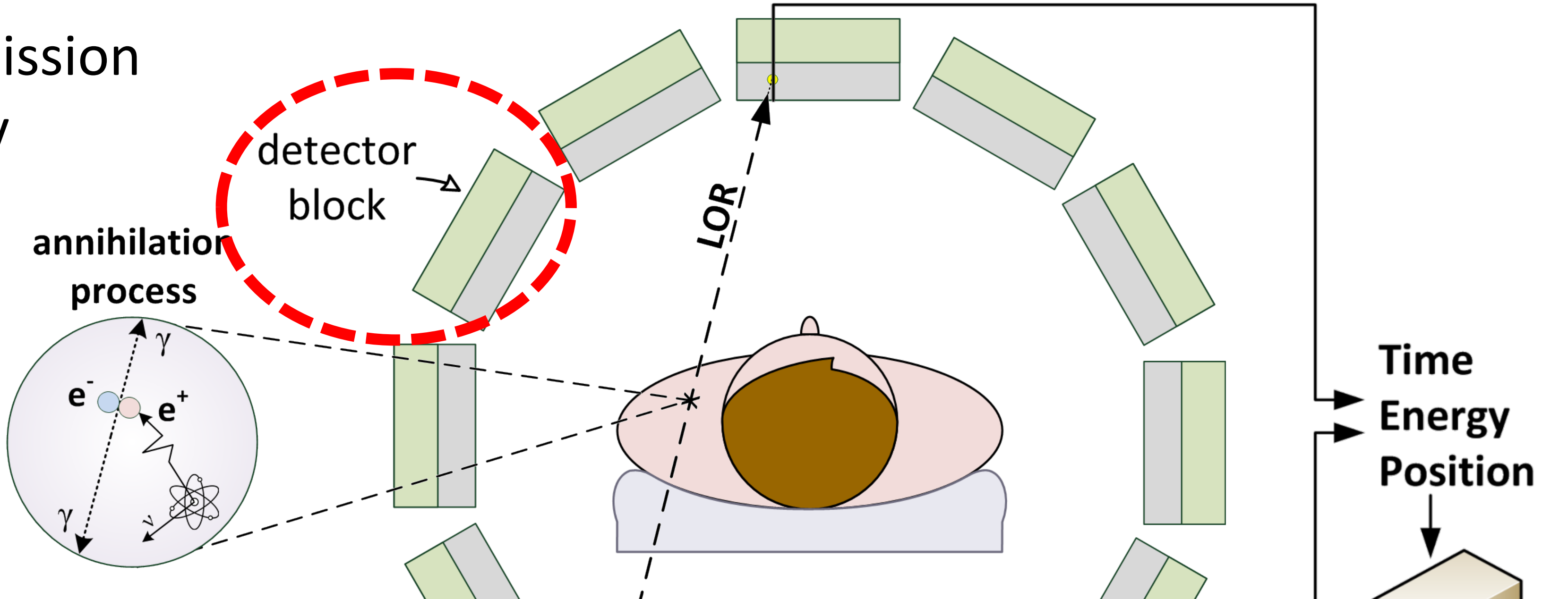
Non-Gaussian behavior of the SPADs timing uncertainty

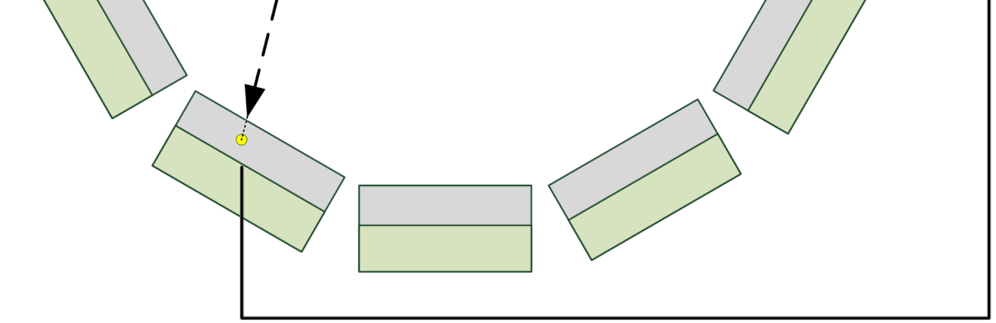


A. R. Ximenes *et al.*, *A Modular, Direct Time-of-Flight Depth Sensor in 45/65-nm 3-D-Stacked CMOS Technology*. IEEE JSSC 54 (2019).
C. Niclass *et al.*, *A 128×128 Single-Photon Image Sensor With Column-Level 10-Bit Time-to-Digital Converter Array*. IEEE JSSC 43 (2008).

8.2.9 Example 4: Real Life Truths – TOF-PET

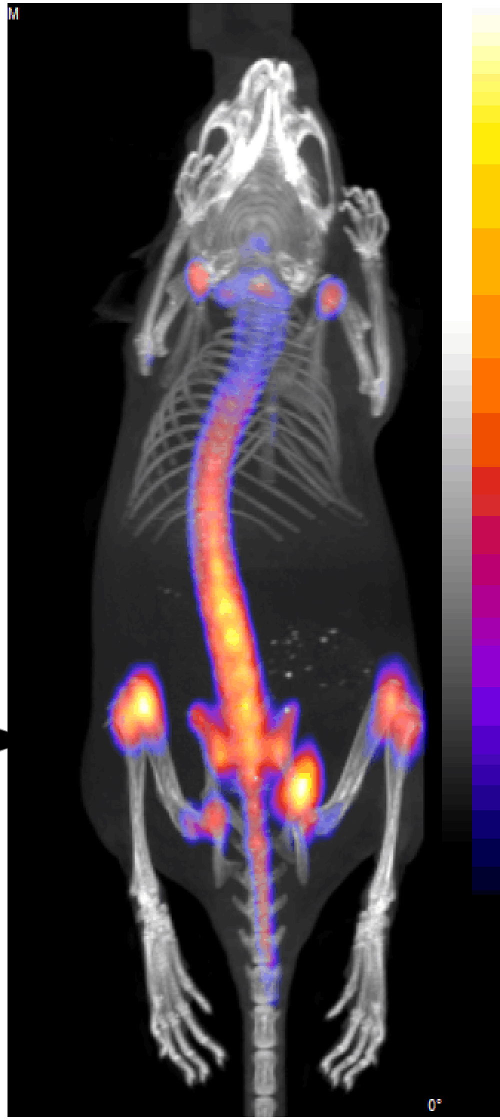
Positron Emission Tomography Basics





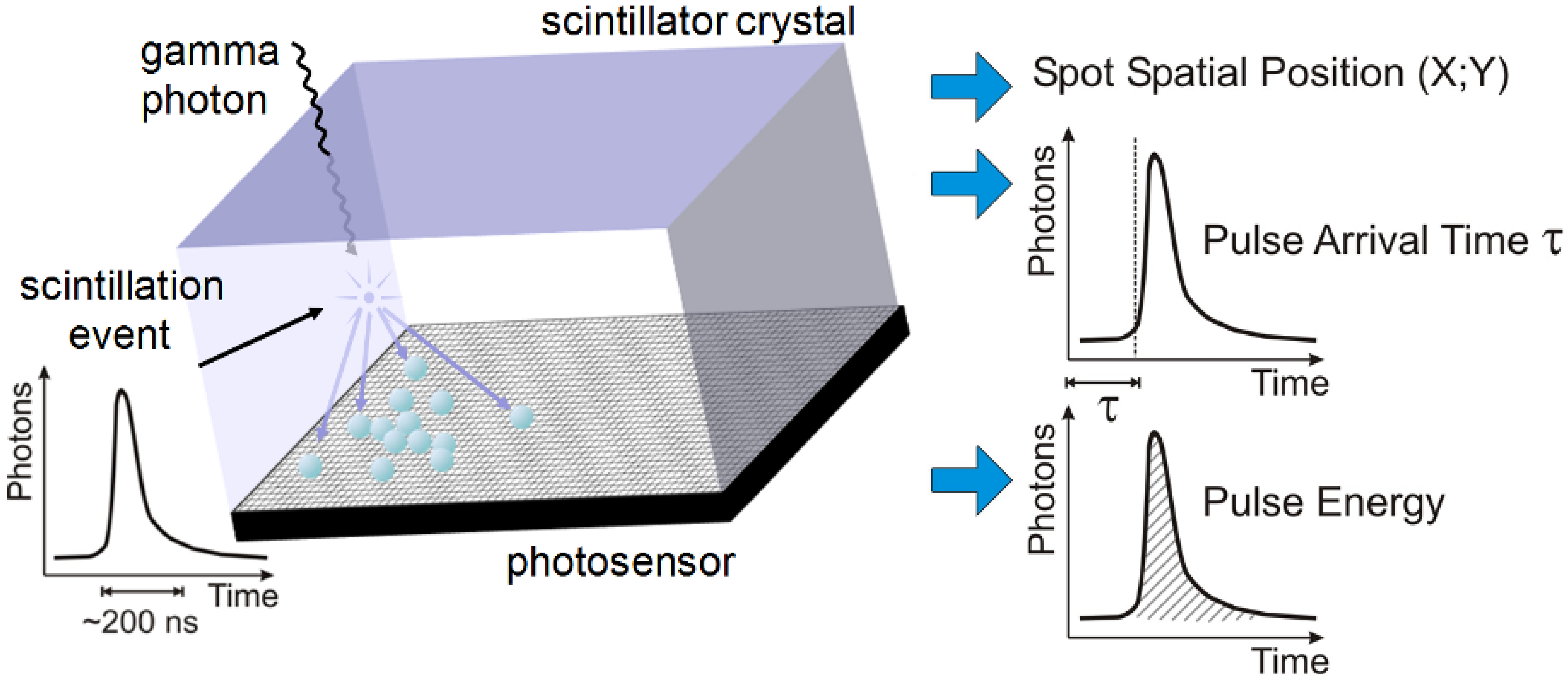
GE Discovery IQ, Nov 2016

L. Braga *et al.*, ISSCC, 2013



G. Nemeth, Mediso, Delft WS 2010

8.2.9 Example 4: Real Life Truths – TOF-PET



R. Walker *et al.*, IISW, 2013

8.2.9 Example 4: Real Life Truths – Scintillation Light

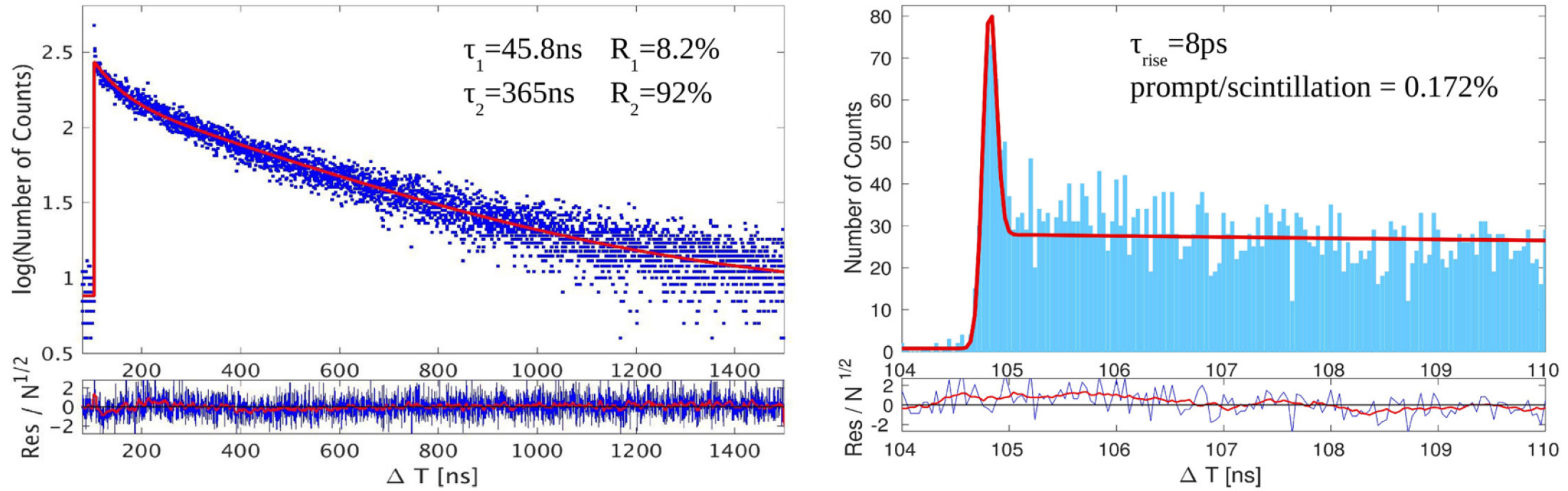
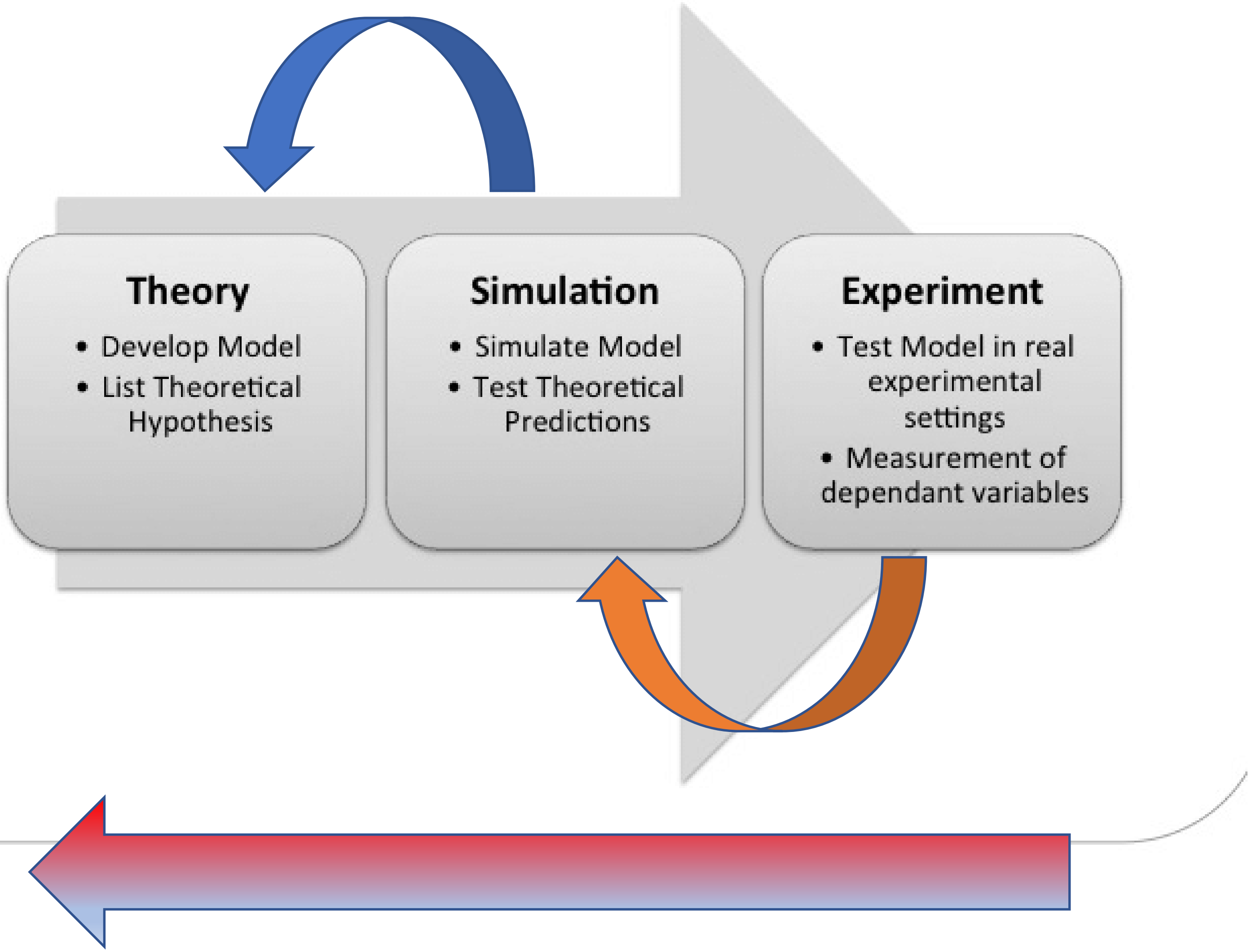


Figure 10. Scintillation decay and rise time of BGO measured with a time correlated single photon counting (TCSPC) setup using 511 keV annihilation gammas (Gundacker *et al* 2016b). The figure on the right hand side shows a pronounced Cherenkov peak at the onset of the scintillation emission with a relative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

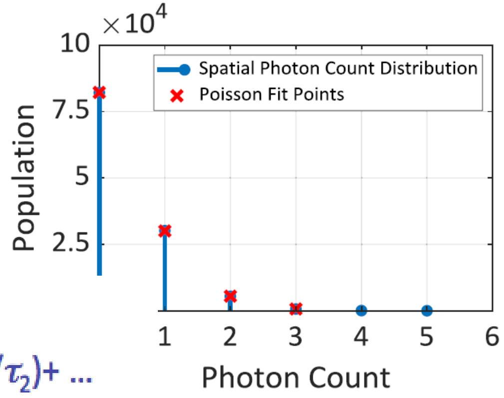
8.2.10 From Theory to Experiment (and back)

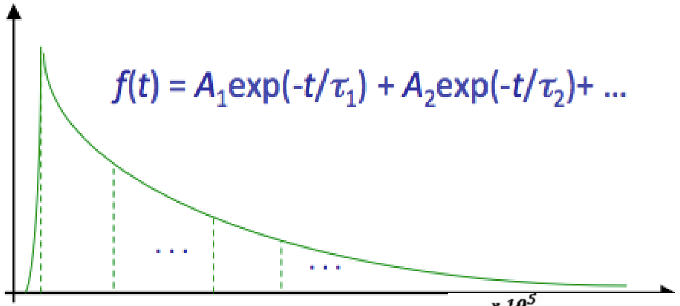


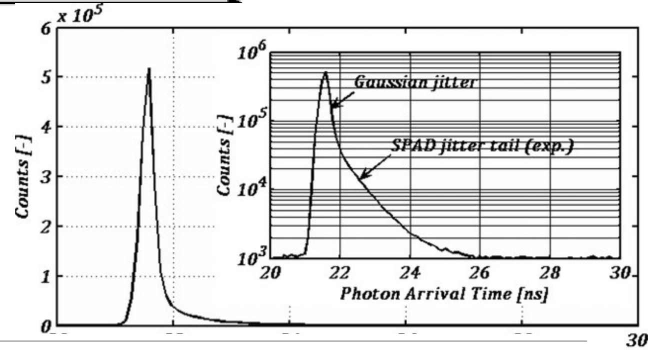
https://www.researchgate.net/publication/315995665_Leading_in_the_Unknown_with_Imperfect_Knowledge_Situational_Creative_Leadership_Strategies_for_Ideation_Management/figures

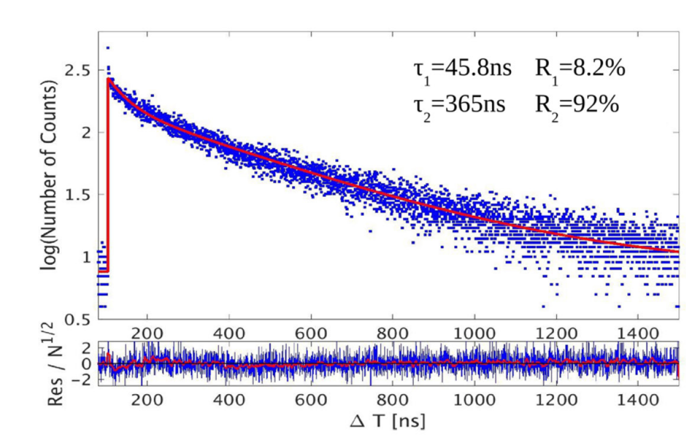
Take-home Messages/W8-2

- *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









Outline

8.1 Introduction to Probability

8.2 Random Variables

8.3 Moments

8.4 Covariance and Correlation

9.1 Random Processes

9.2 Central Limit Theorem

9.3 Estimation Theory

9.4 Accuracy, Precision and Resolution

8.3.1 Expected Values

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

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

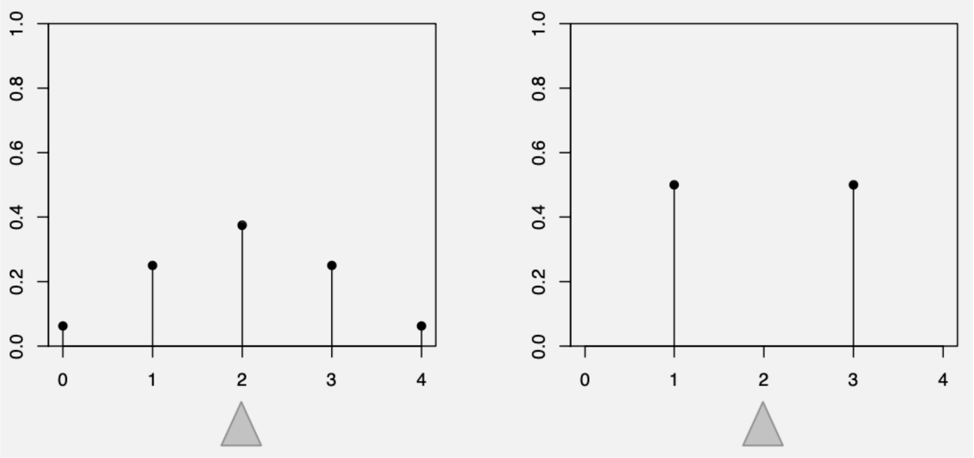
- The expected value is undefined if:

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

- 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…

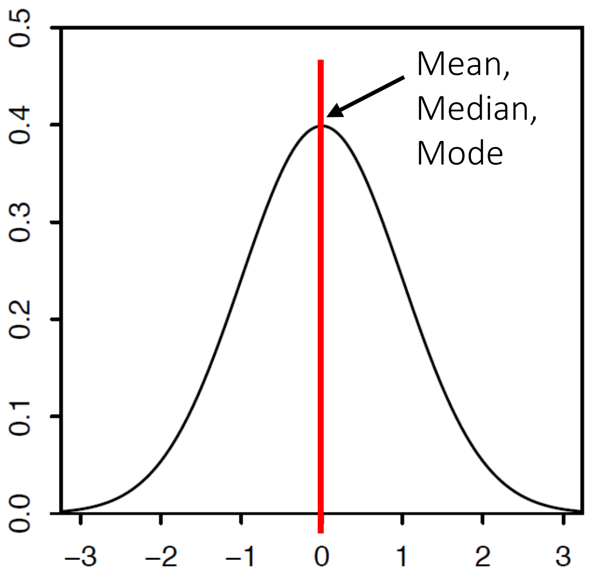
J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 4.1

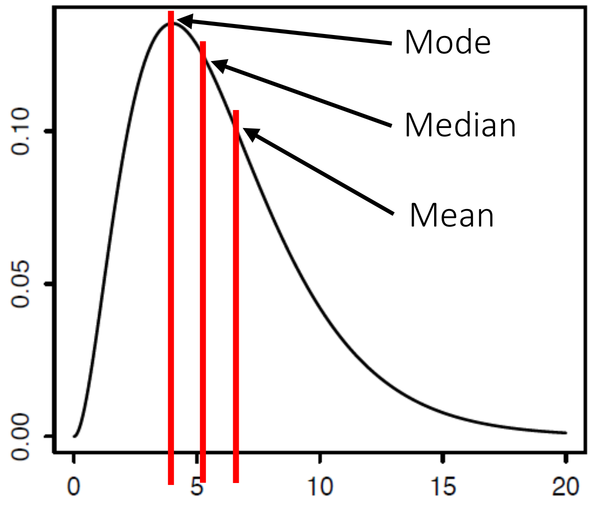
8.3.2 Mean, Median and Mode

- 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):

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

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



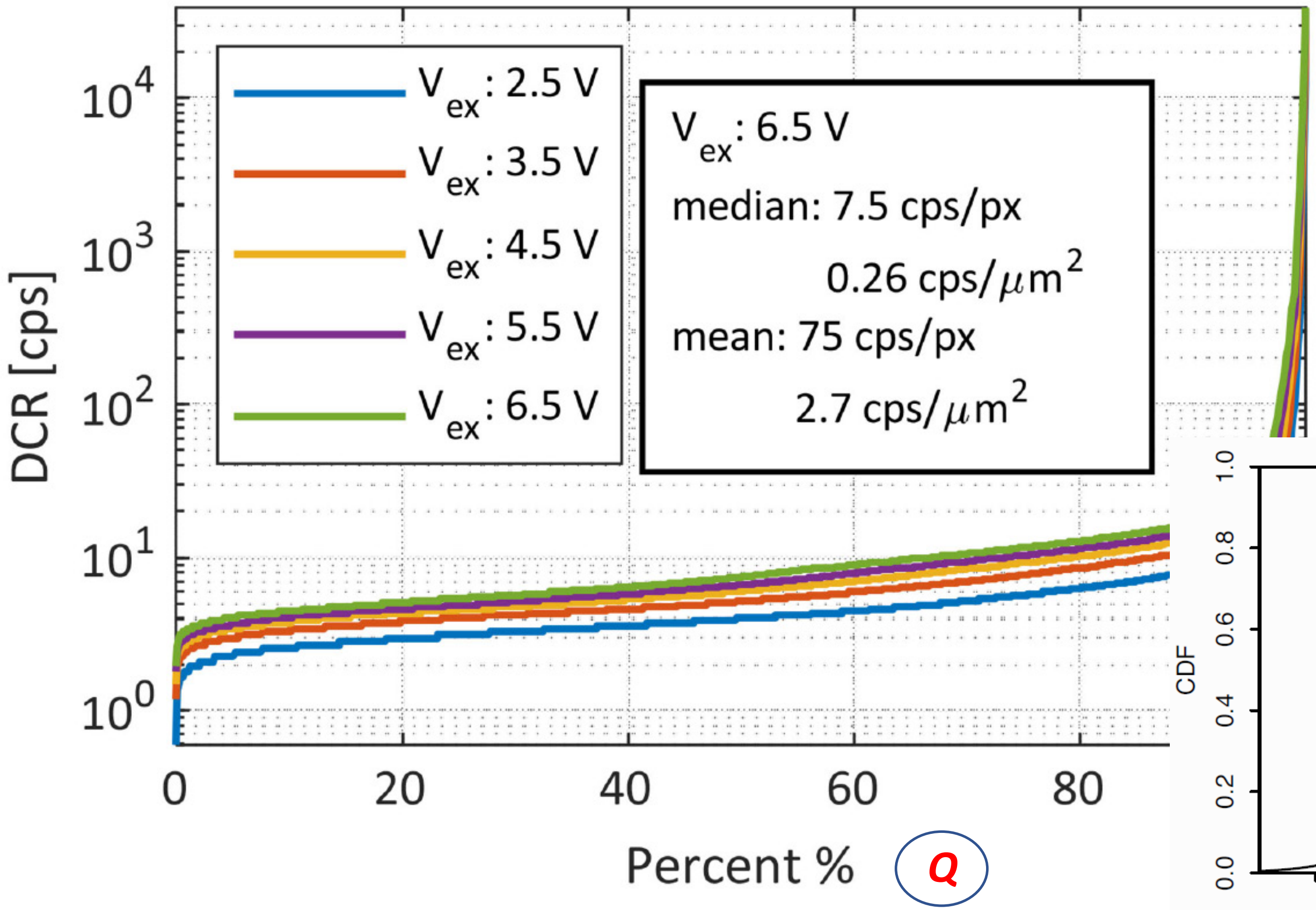


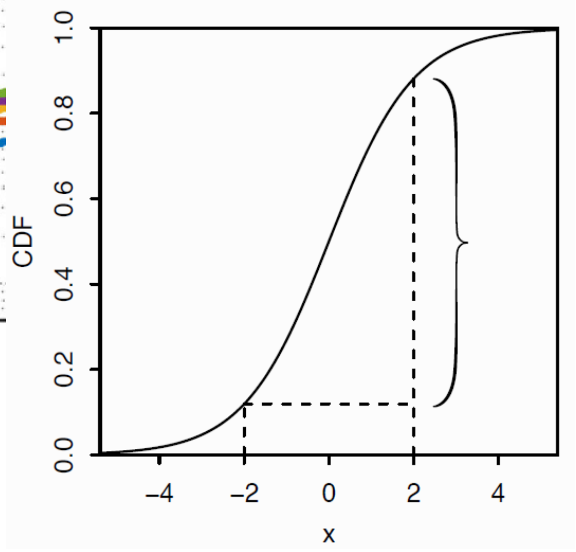
J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 6.1

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).

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 $g\colon \mathbb{R} \to \mathbb{R}$, follows:

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

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

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

📖 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*):

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

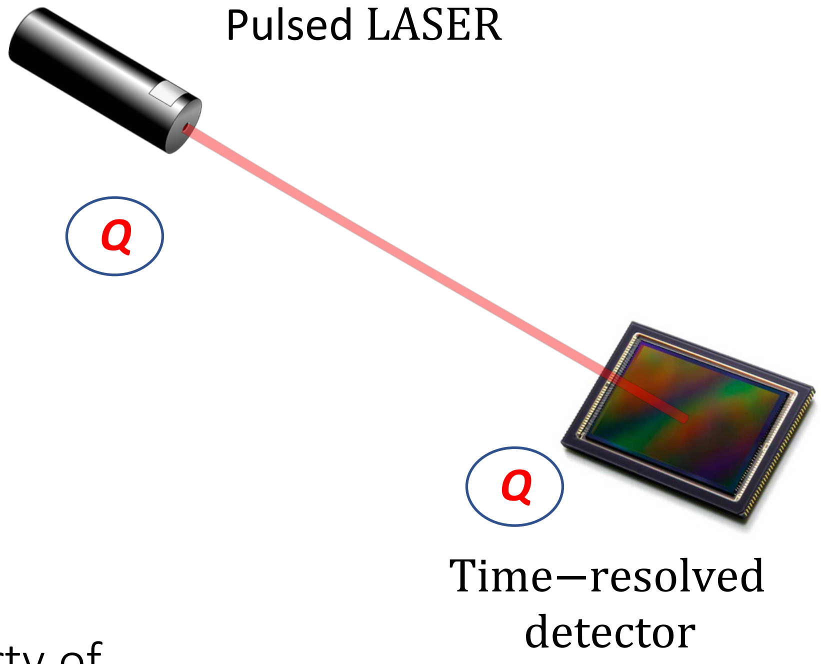
and its square root is called the standard deviation:

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

- 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

8.3.4 Variance (contd.)

- The Variance has the following properties:

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

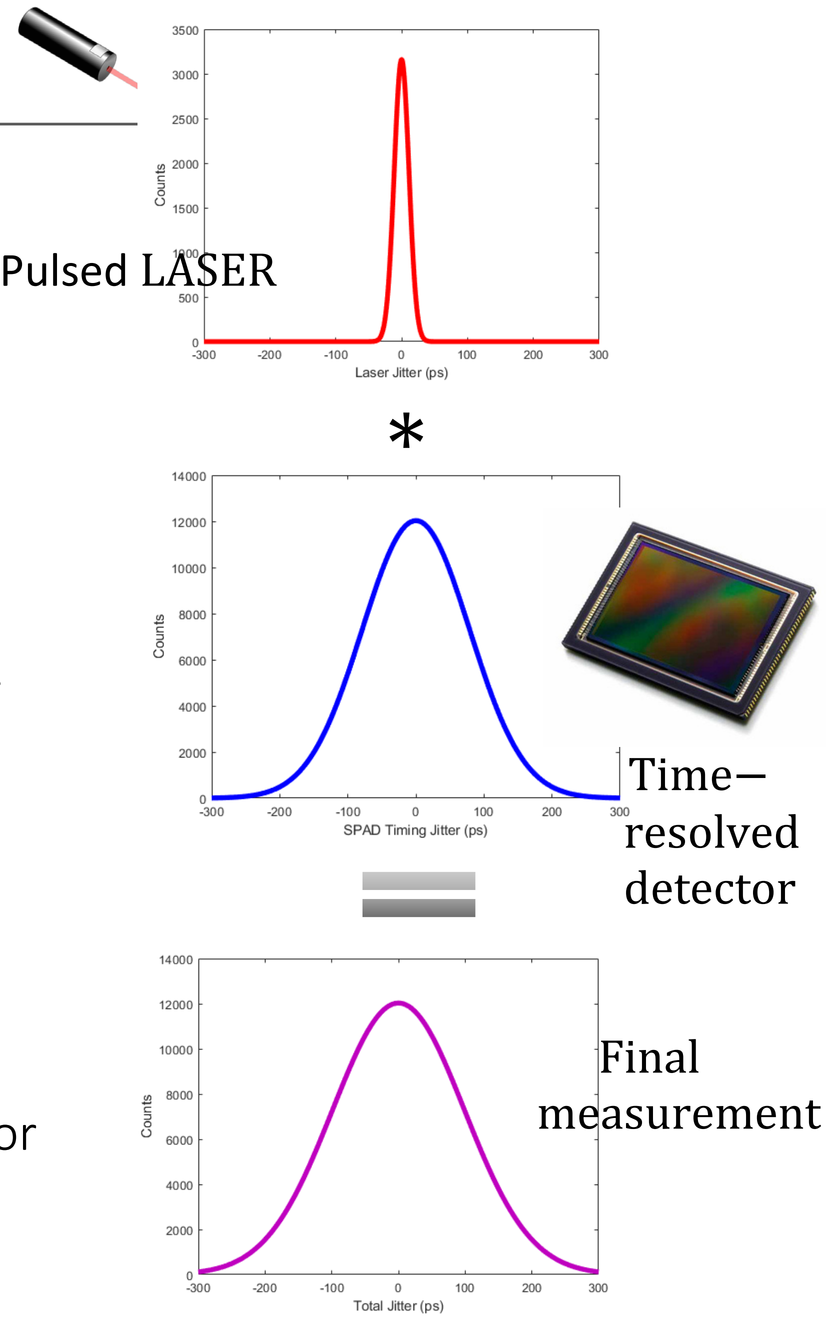
2. $Var\{cX\} = c^2 Var\{X\}$ for any constant c.

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]



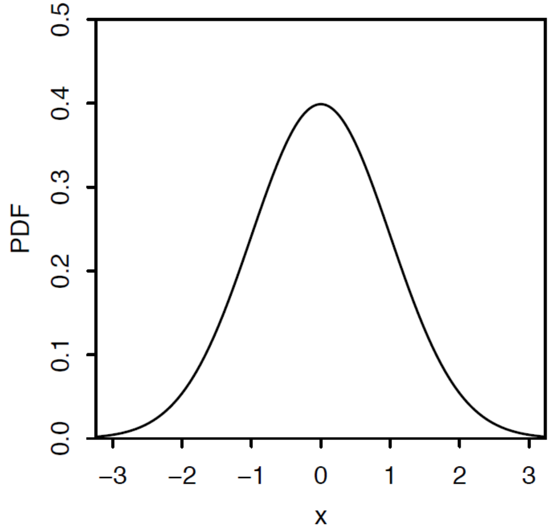
J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 4.6

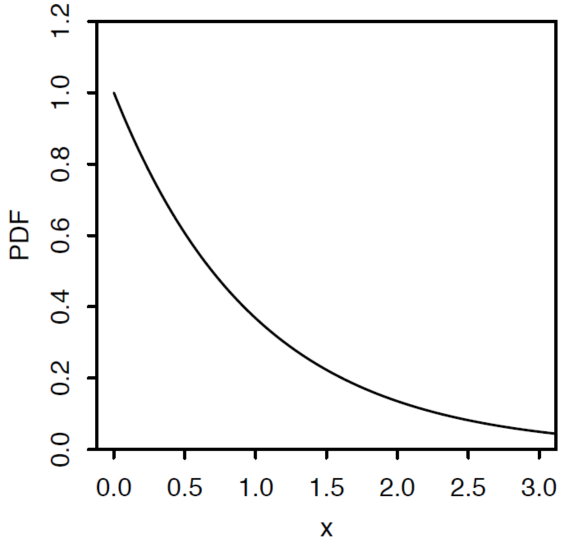
8.3.5 Moments

- 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:

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





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

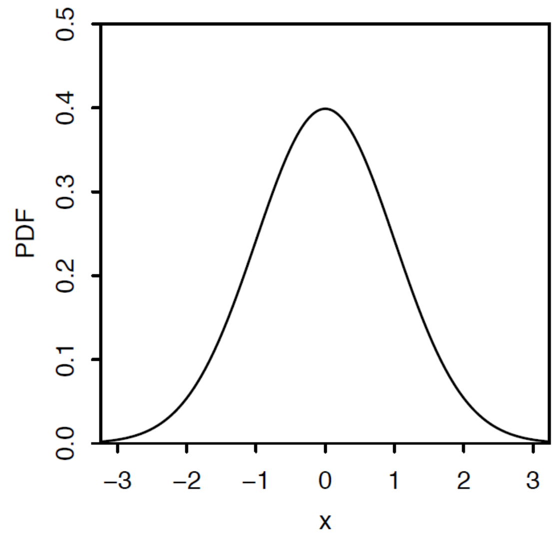
8.3.5 Moments (contd.)

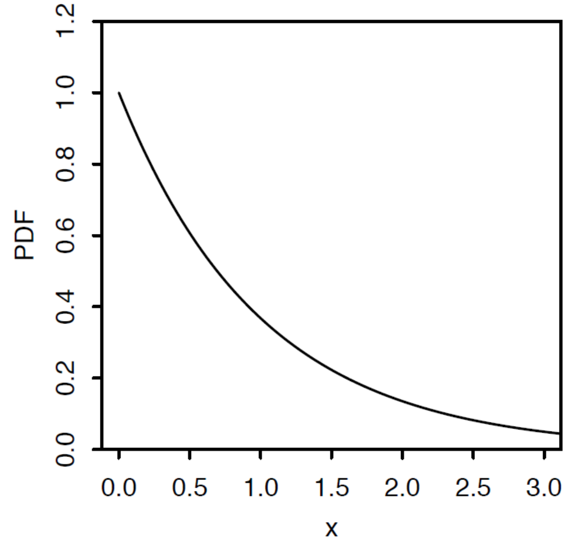
- In the same fashion, the second central moment of a RV X is its variance, or the moment of inertia of the distribution around its center:

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

- 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

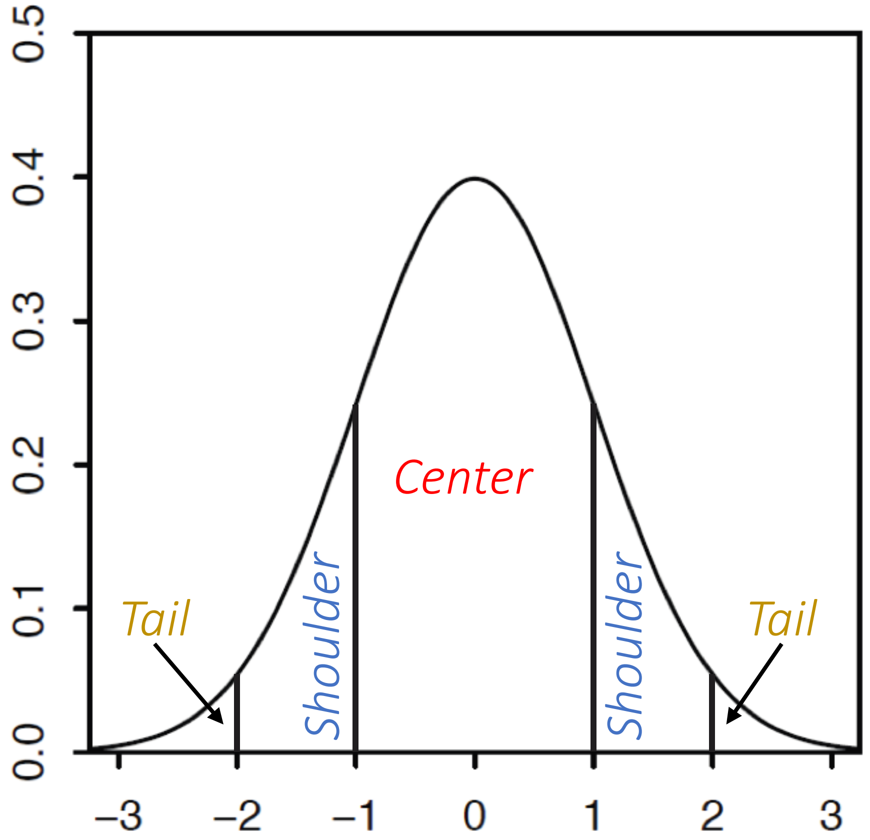
8.3.5 Moments (contd.)

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

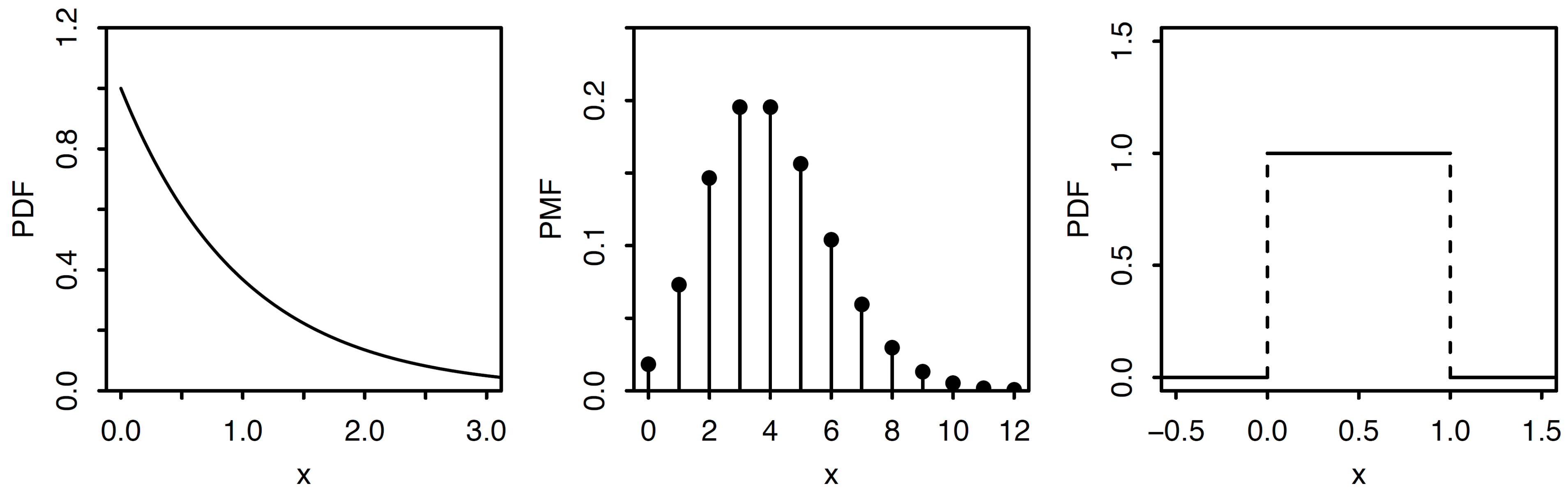
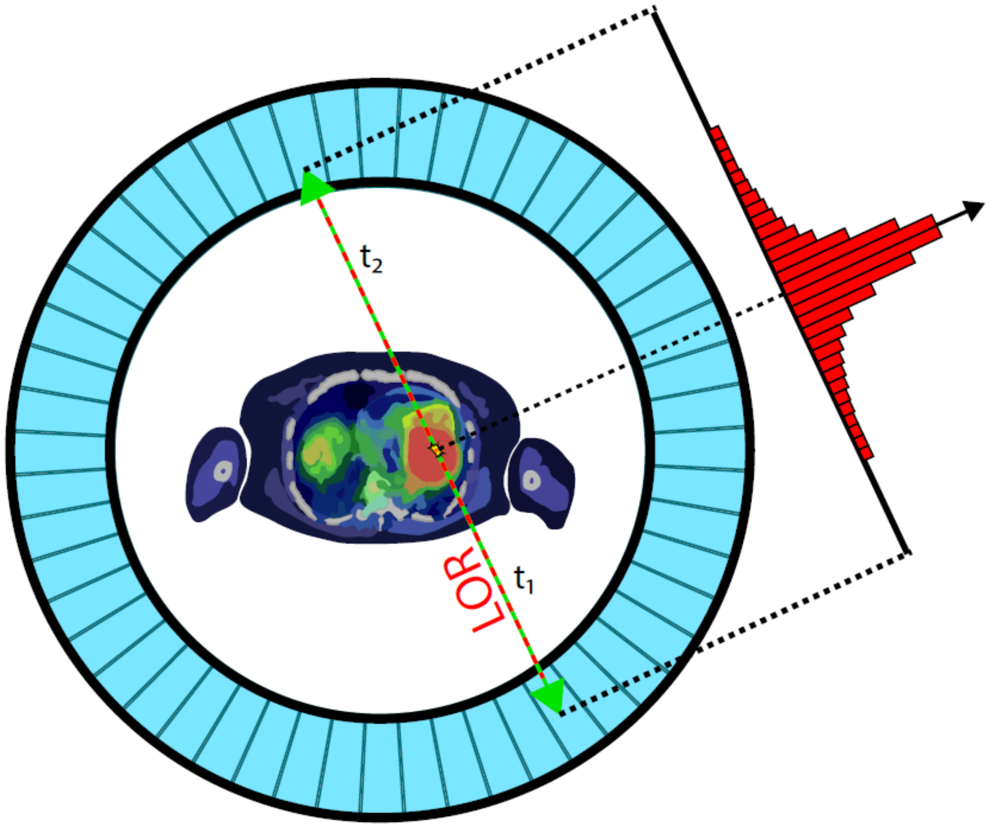


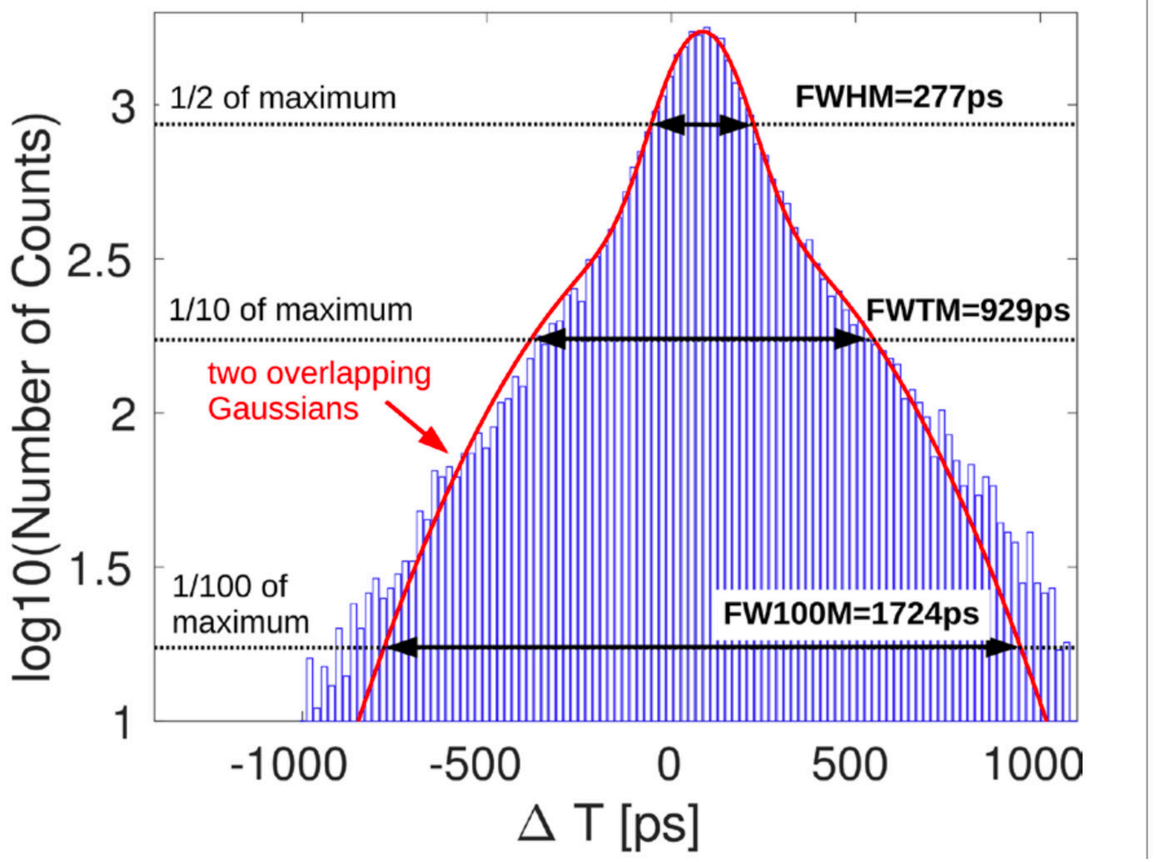
FIGURE 6.6
Skewness and kurtosis of some named distributions. Left: Expo(1) PDF, skewness = 2, kurtosis = 6. Middle: Pois(4) PMF, skewness = 0.5, kurtosis = 0.25. Right: Unif$(0, 1)$ PDF, skewness = 0, kurtosis = -1.2.

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

8.3.5 Moments (contd.) – Experimental Example

Coincidence measurements between two scintillating crystals -> influence of actual curve shapes





Large crystal

S. Gundacker *et al.*, *Experimental time resolution limits of modern SiPMs and TOF-PET detectors exploring different scintillators and Cherenkov emission*, PMB 65 (2020).

F. Gramuglia, High-Performance CMOS SPAD-Based Sensors for Time-of-Flight PET Applications, *EPFL Thèse 8720* (2022).

8.3.6 Moment Generating Functions

S

- 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.

Ex

**Using LOTUS (Section 8.3.3)*

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

Outline

8.1 Introduction to Probability

8.2 Random Variables

8.3 Moments

8.4 Covariance and Correlation

9.1 Random Processes

9.2 Central Limit Theorem

9.3 Estimation Theory

9.4 Accuracy, Precision and Resolution

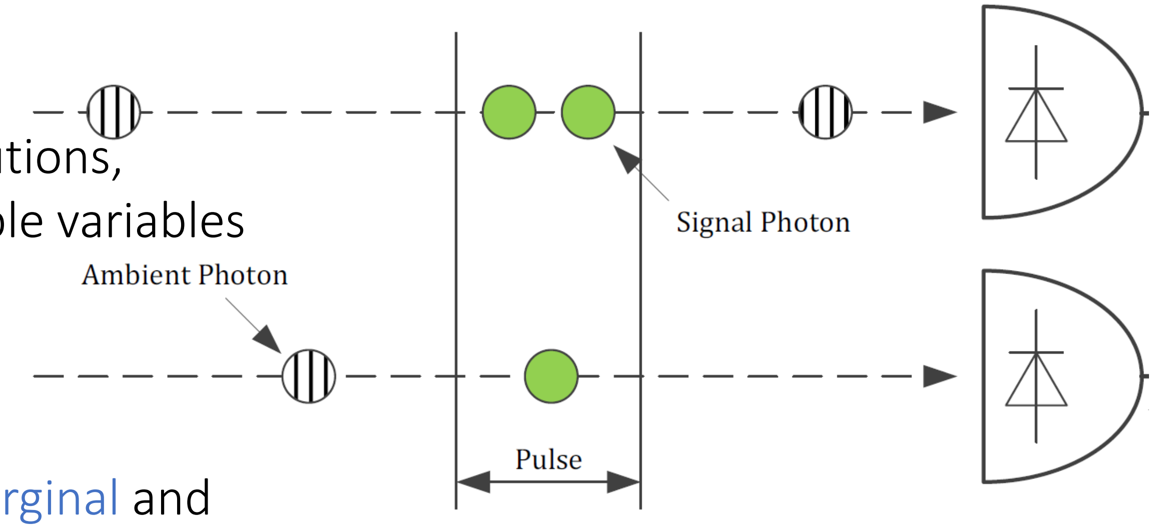
8.4 Multivariate Distributions

- 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.*

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:

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

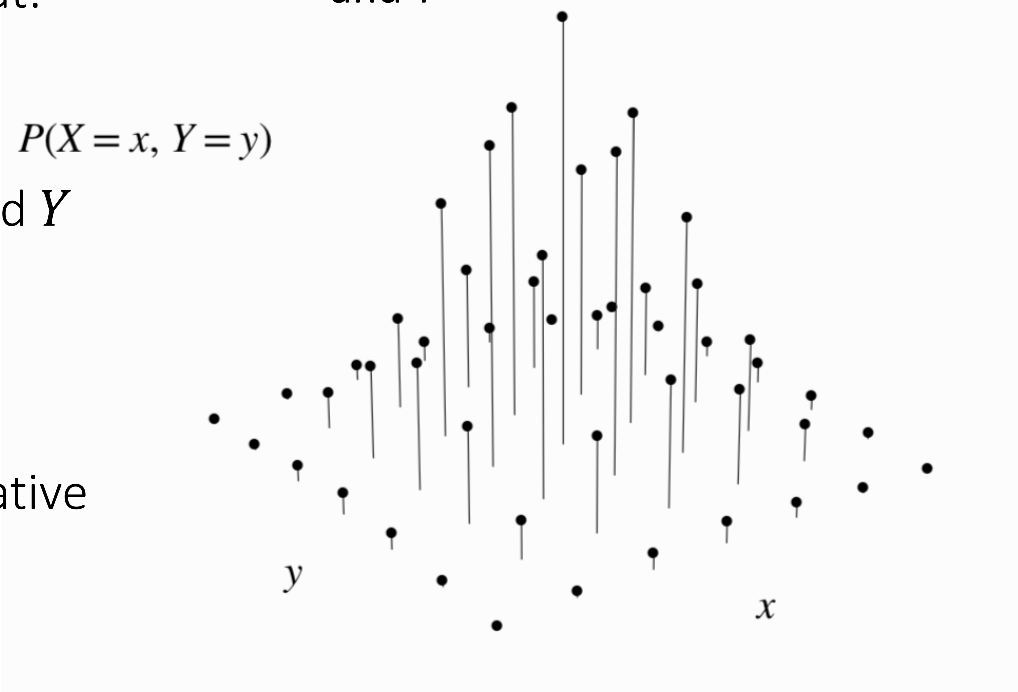
- In the same fashion, the joint PMF of two discrete RVs X and Y is a function $p_{X,Y}$ such that:

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

- 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

8.4.1 Joint Distributions (contd.)

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

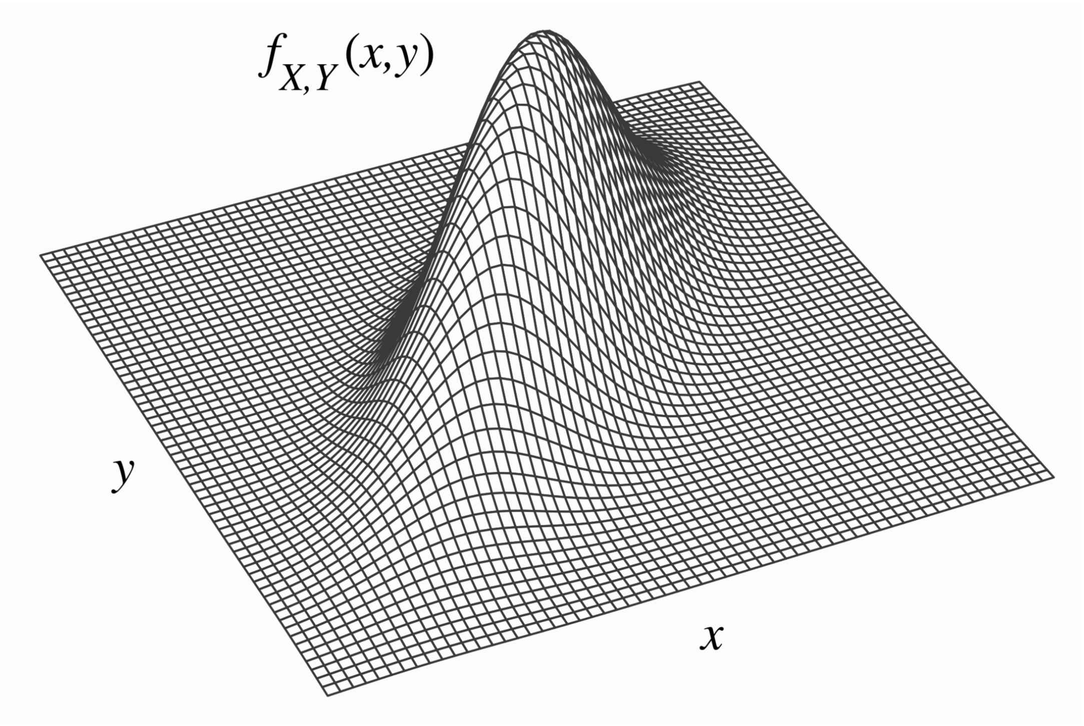
$$\text{PDF: } f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$$

- 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

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:

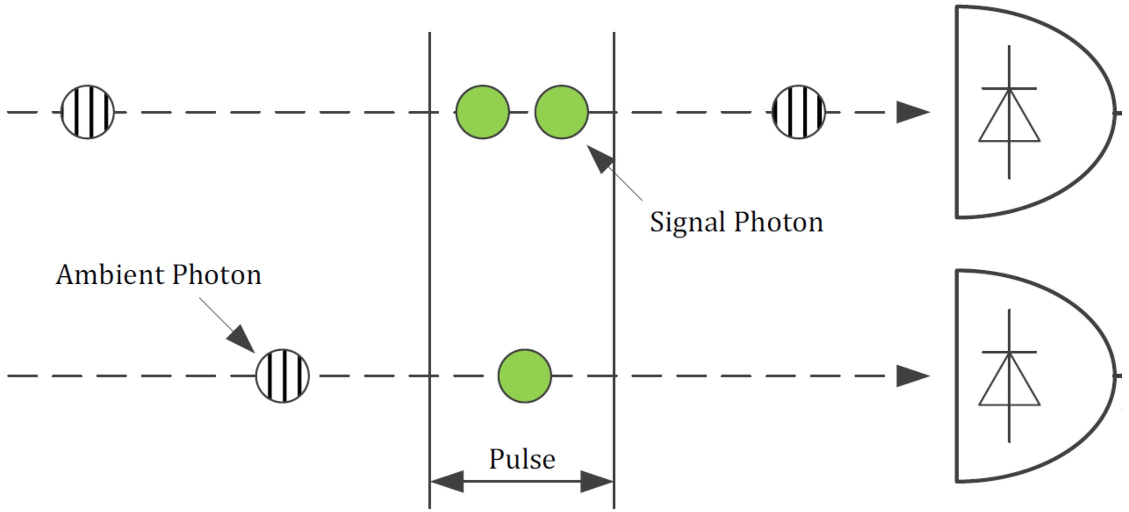
$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y)$$

or

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

for all x and y.

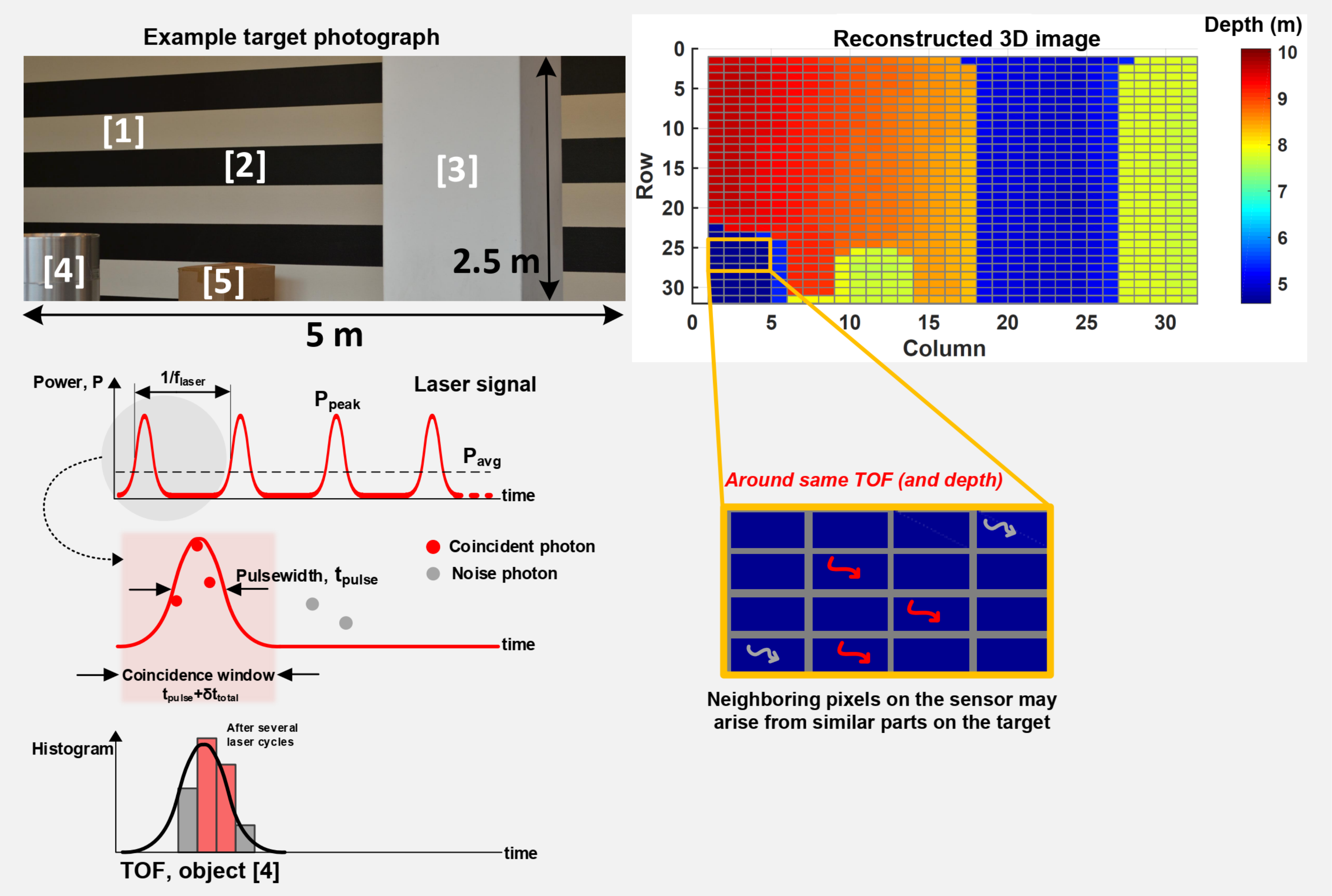
Example: LiDAR employing detection of photon coincidences (within a coincidence window) in presence of background light



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.*

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

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"):

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

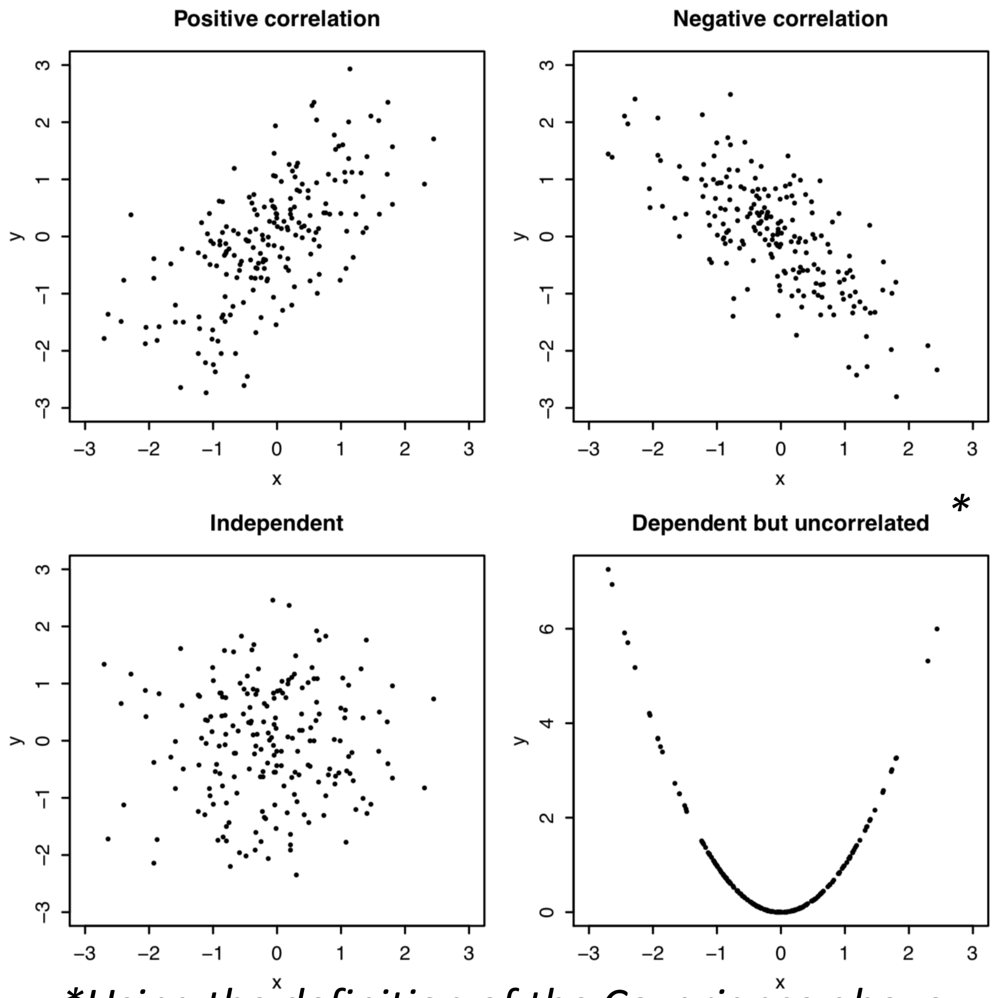
which, using linearity, becomes

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

- 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

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"):

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

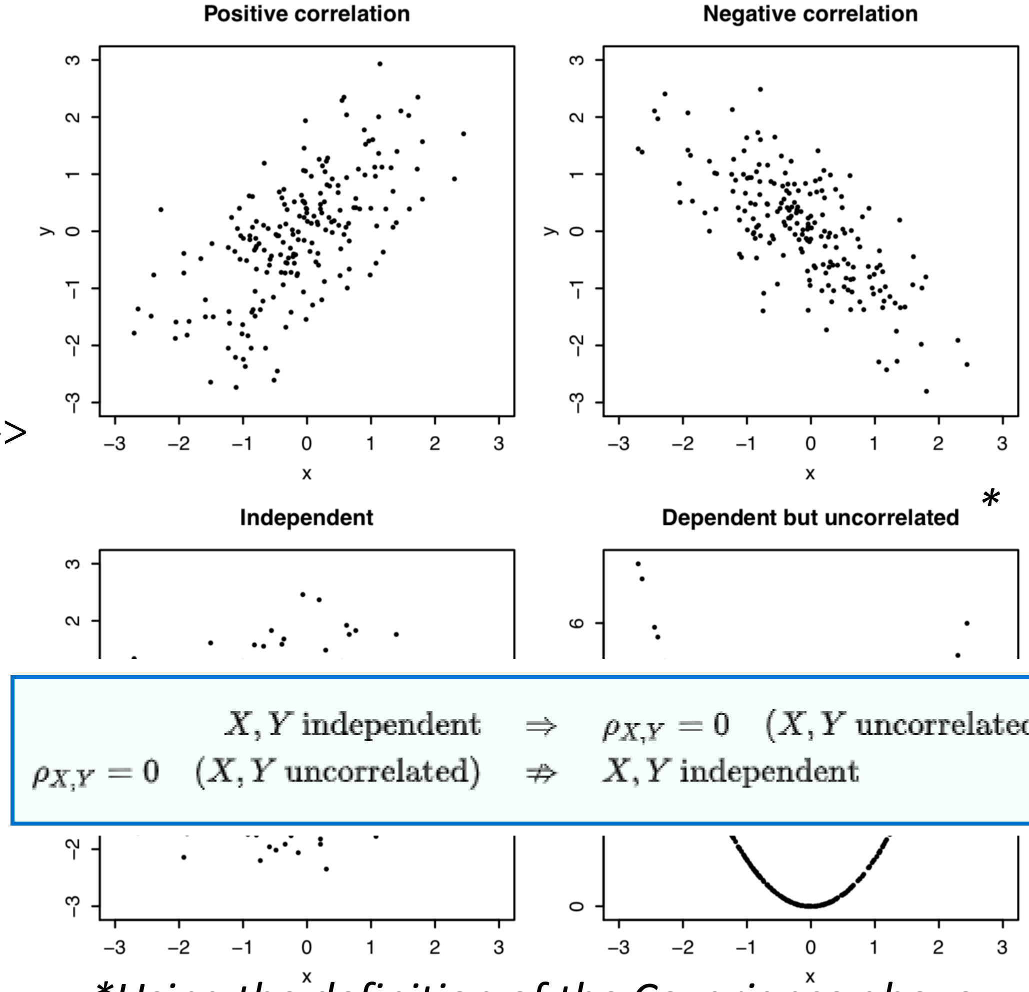
which, using linearity, becomes

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

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

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

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



$$X, Y \text{ independent} \Rightarrow \rho_{X,Y} = 0 \quad (X, Y \text{ uncorrelated})$$
$$\rho_{X,Y} = 0 \quad (X, Y \text{ uncorrelated}) \nRightarrow X, Y \text{ independent}$$

Using the definition of the Covariance above…

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

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

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:

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

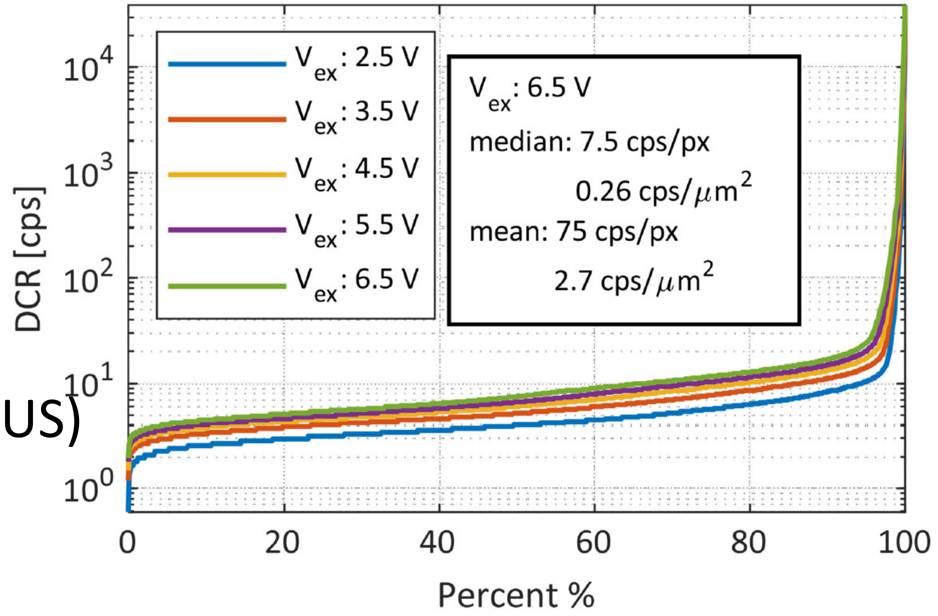
- Moreover, the correlation is bounded:

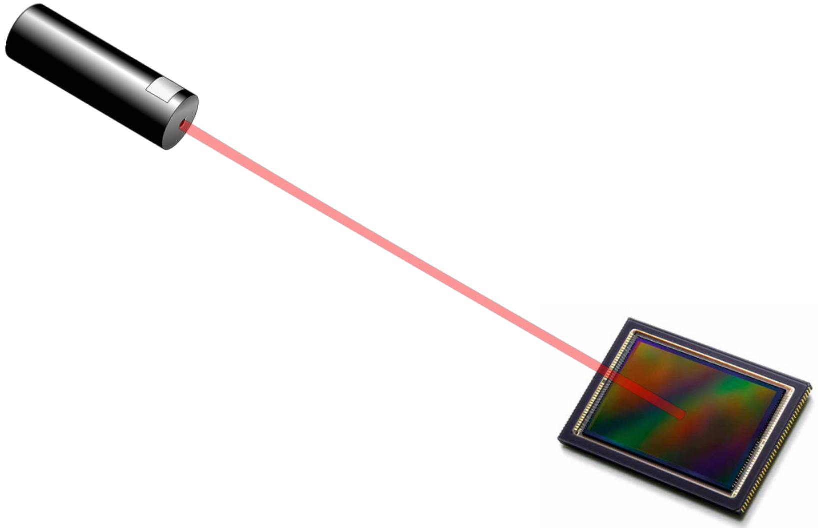
$$-1 \leq Corr\{X,Y\} \leq 1$$

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.* $Var\{X_1 + \cdots + X_n\}$





Appendix

A8.1 Introduction to Probability

A8.A Multivariate Distributions

A8.B Multivariate Distributions – Example: LIDAR

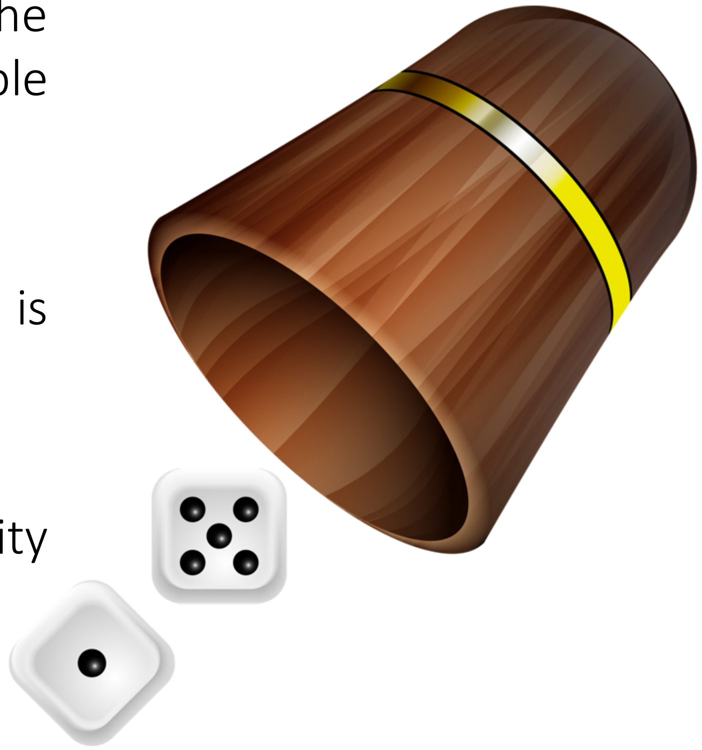
Appendix 8.1: Introduction to Probability (contd.) – Example

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: $P\{\text{face is 3}\} = 1/6$.

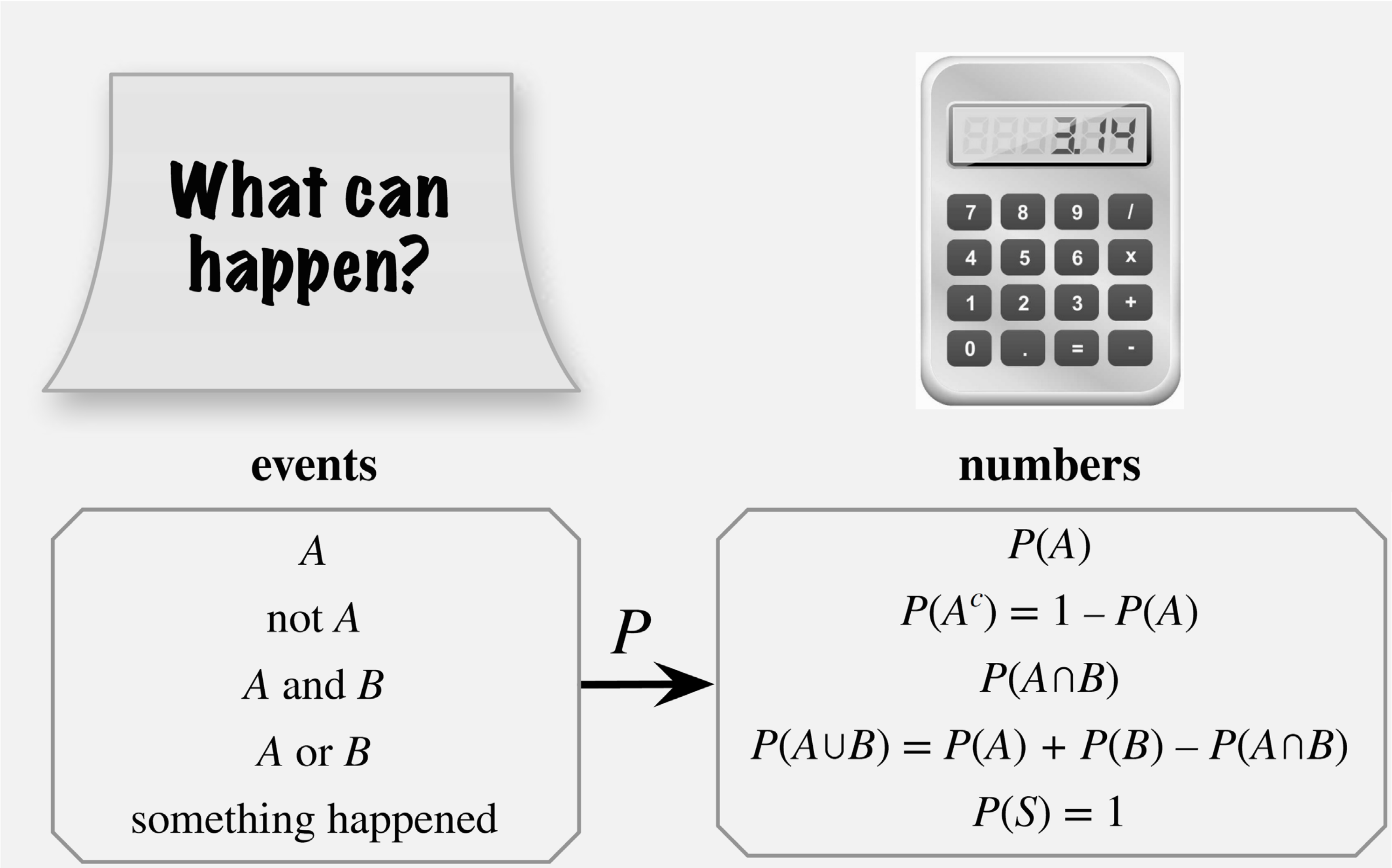
- In the same fashion, the probability of obtaining an odd number is $P\{\text{face is odd}\} = 3/6$.

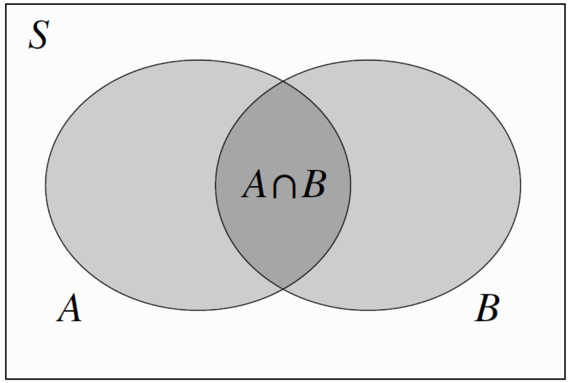
- The fair dice represents the classical example of uniform probability distribution, as we will see.



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

How a probability function maps events to numbers

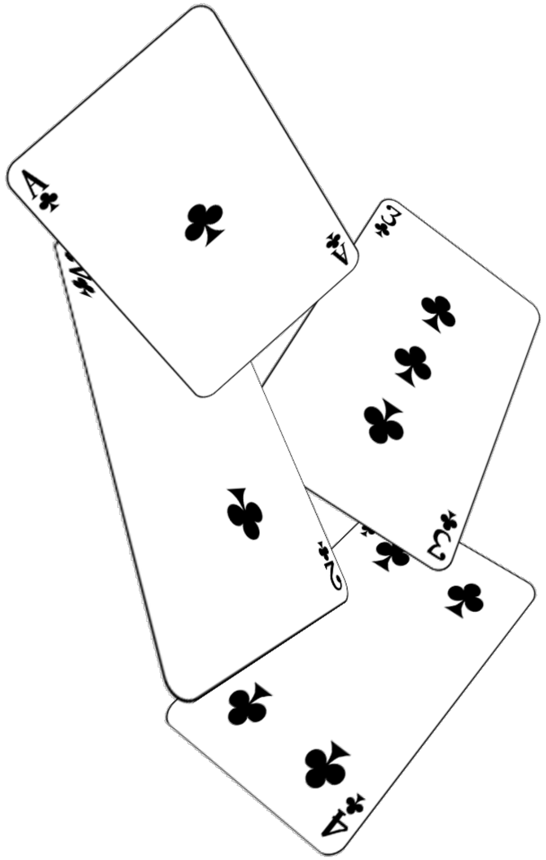




J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 1. 6, 1.7

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?

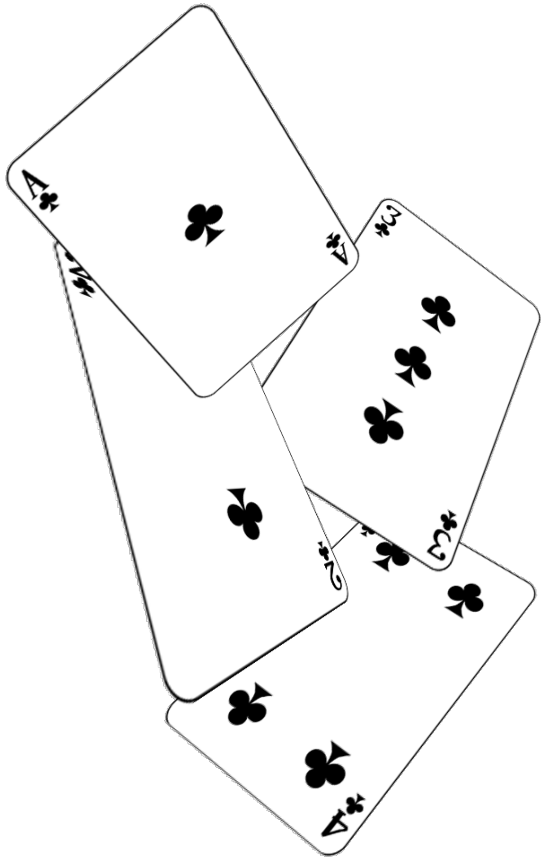
- If $\mathcal{A}$ and $\mathcal{B}$ are events with $P\{\mathcal{B}\} > 0$, then the conditional probability of $\mathcal{A}$ given $\mathcal{B}$ ($\mathcal{B}$ being the evidence which we observe) is *defined* as:

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

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

A8.1.1 Conditional Probability (contd.)

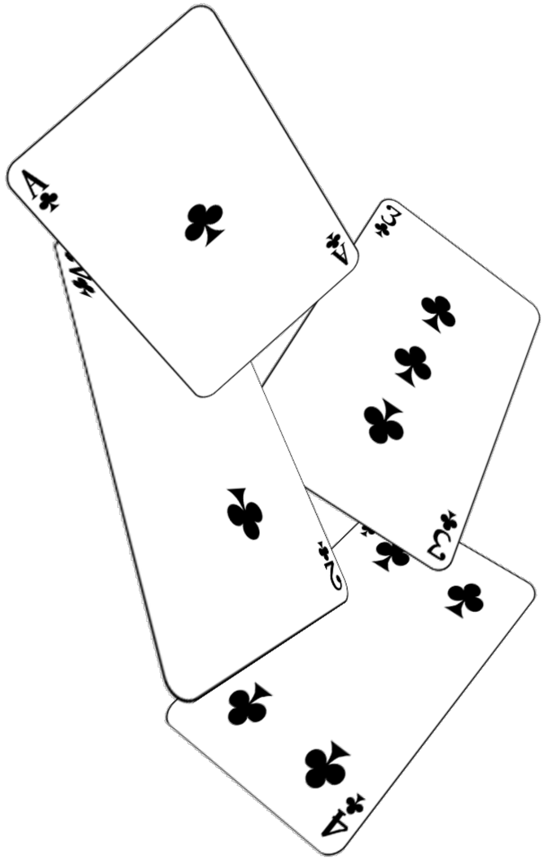
- From the definition of the conditional probability:

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

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*):

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

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

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):

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

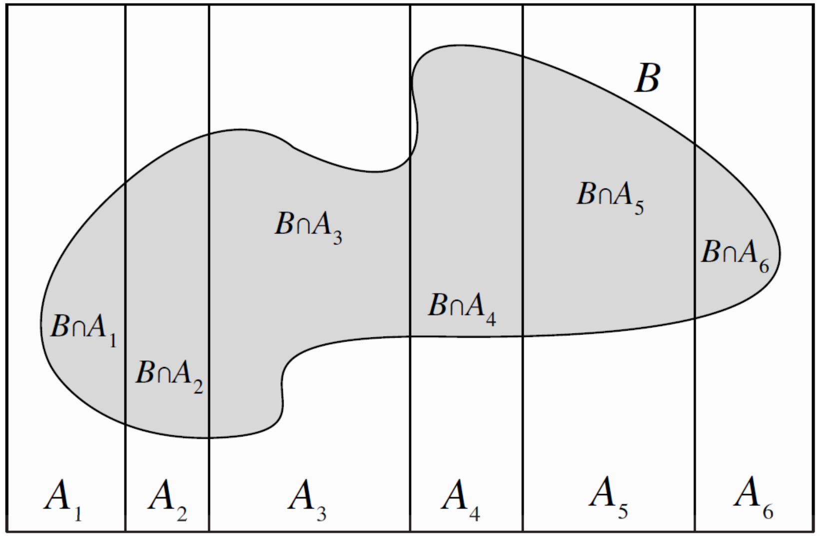
which can be extremely useful in case $P\{\mathcal{B}|\mathcal{A}\}$ is much easier to find than $P\{\mathcal{A}|\mathcal{B}\}$, or vice versa.

- 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)*:

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

Ex

*Relates conditional to unconditional probabilities



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

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

Appendix 8.A – Multivariate Distributions: Marginal Distributions

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

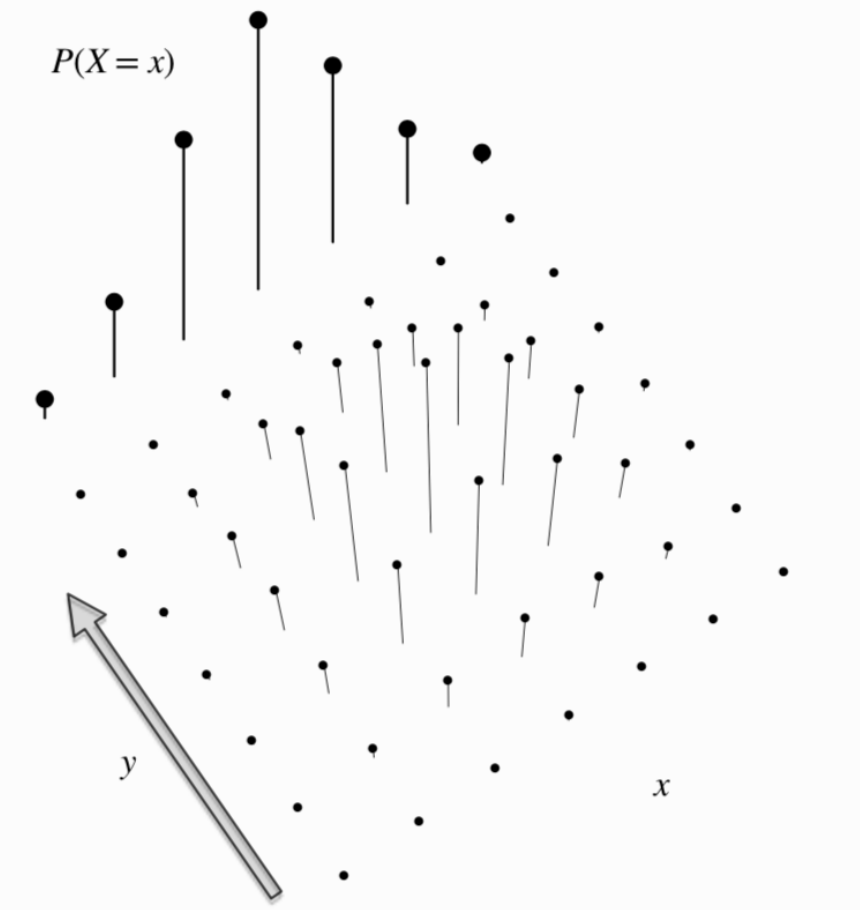
$$P\{X = x\} = \sum_{y} P\{X = x, Y = y\}$$

(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

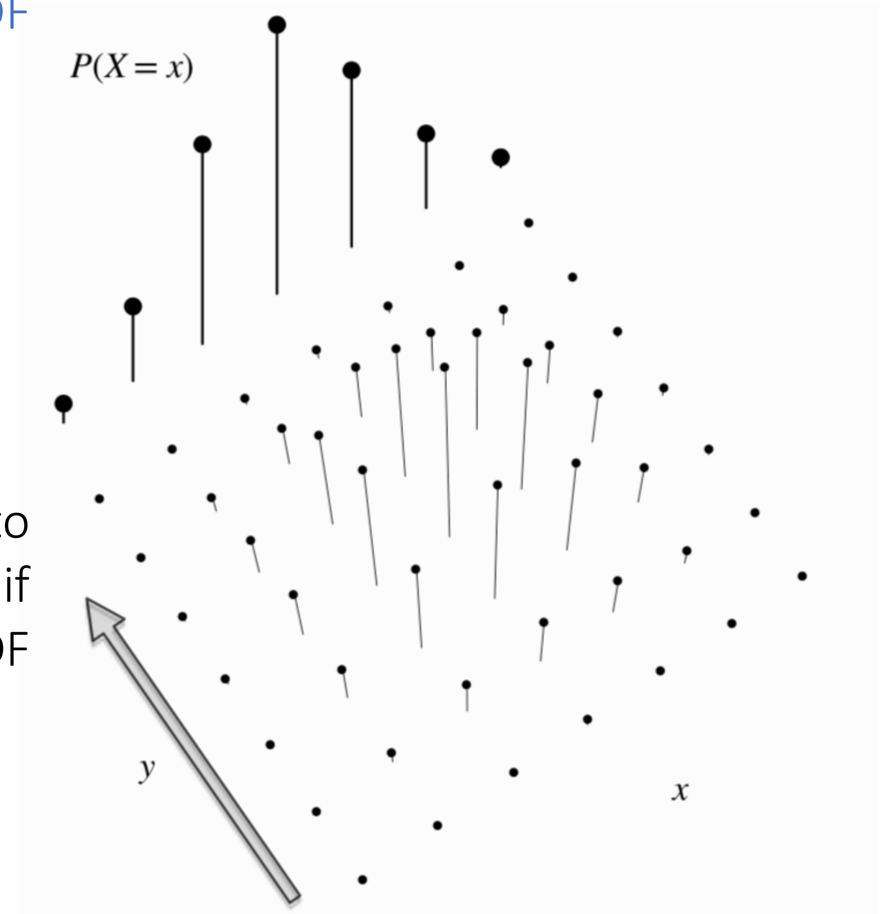
Appendix 8.A – Multivariate Distributions: Marginal Distributions

- 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

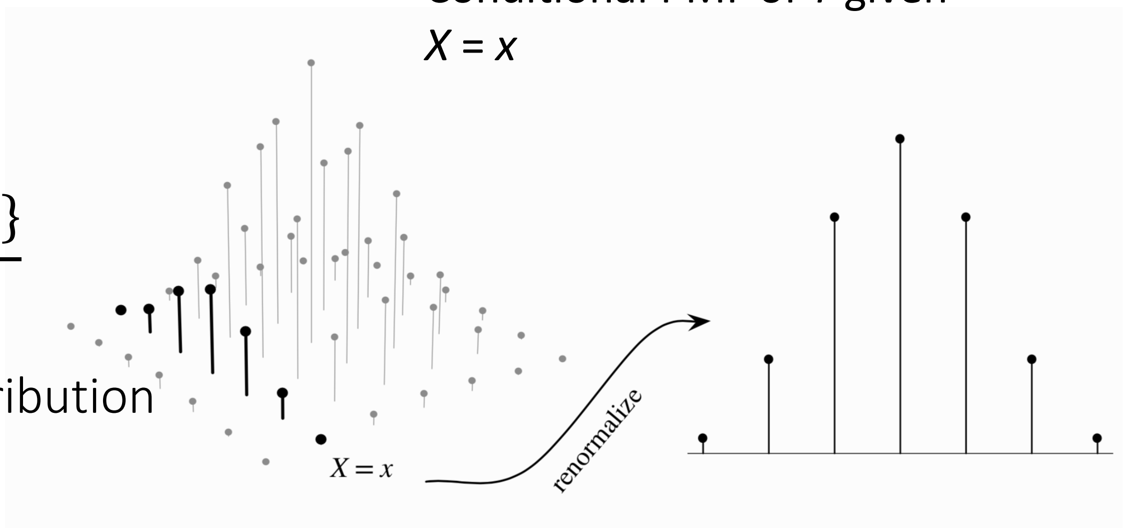
Appendix 8.A – Multivariate Distributions: Conditional Distributions

- 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)

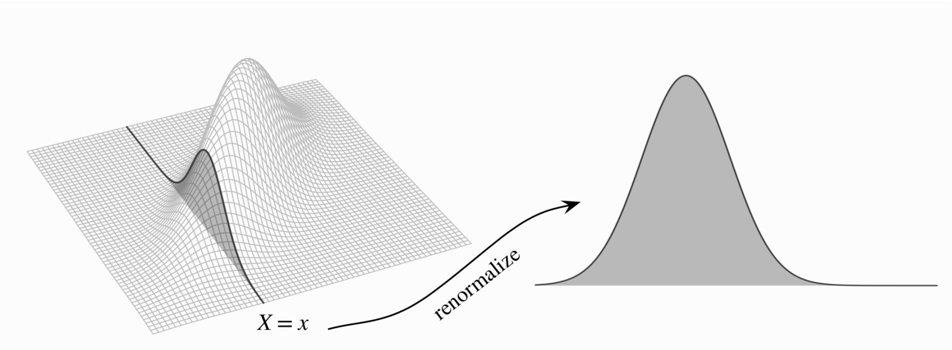
Conditional PMF of *Y* given *X* = *x*



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

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

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

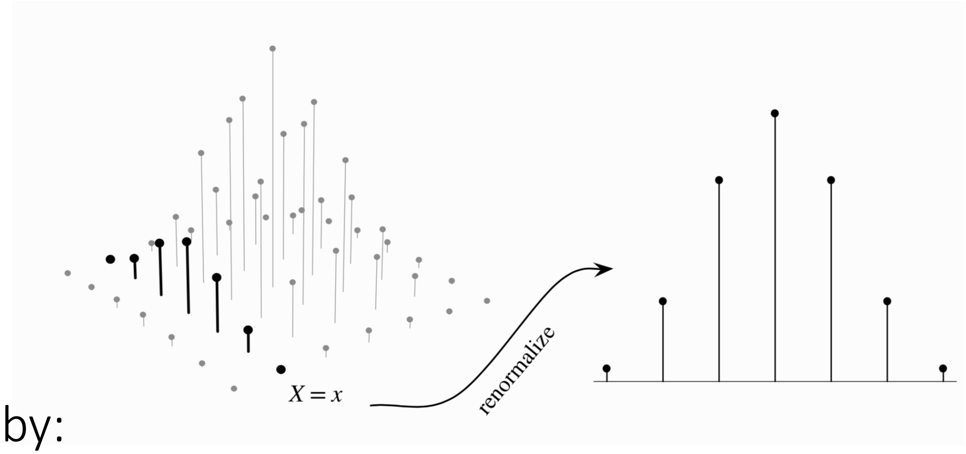


J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 7.1

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:

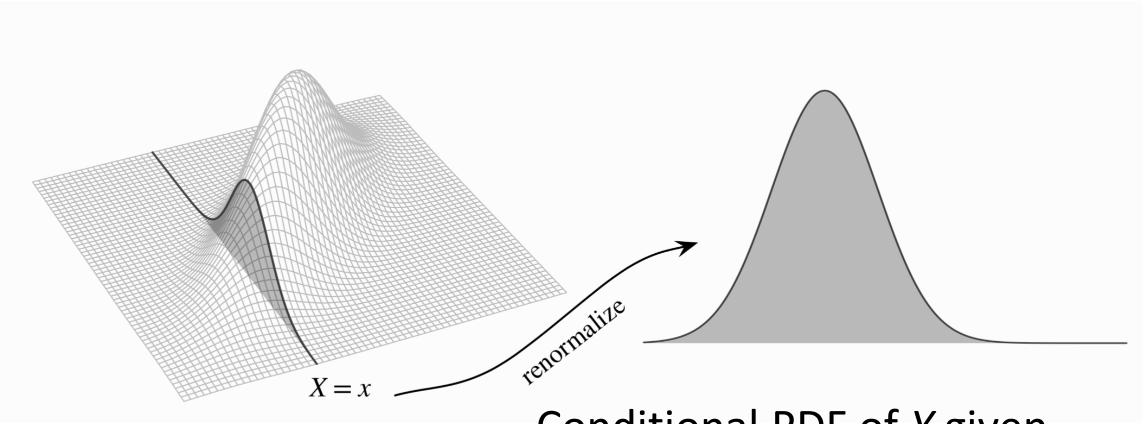
$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$



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

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

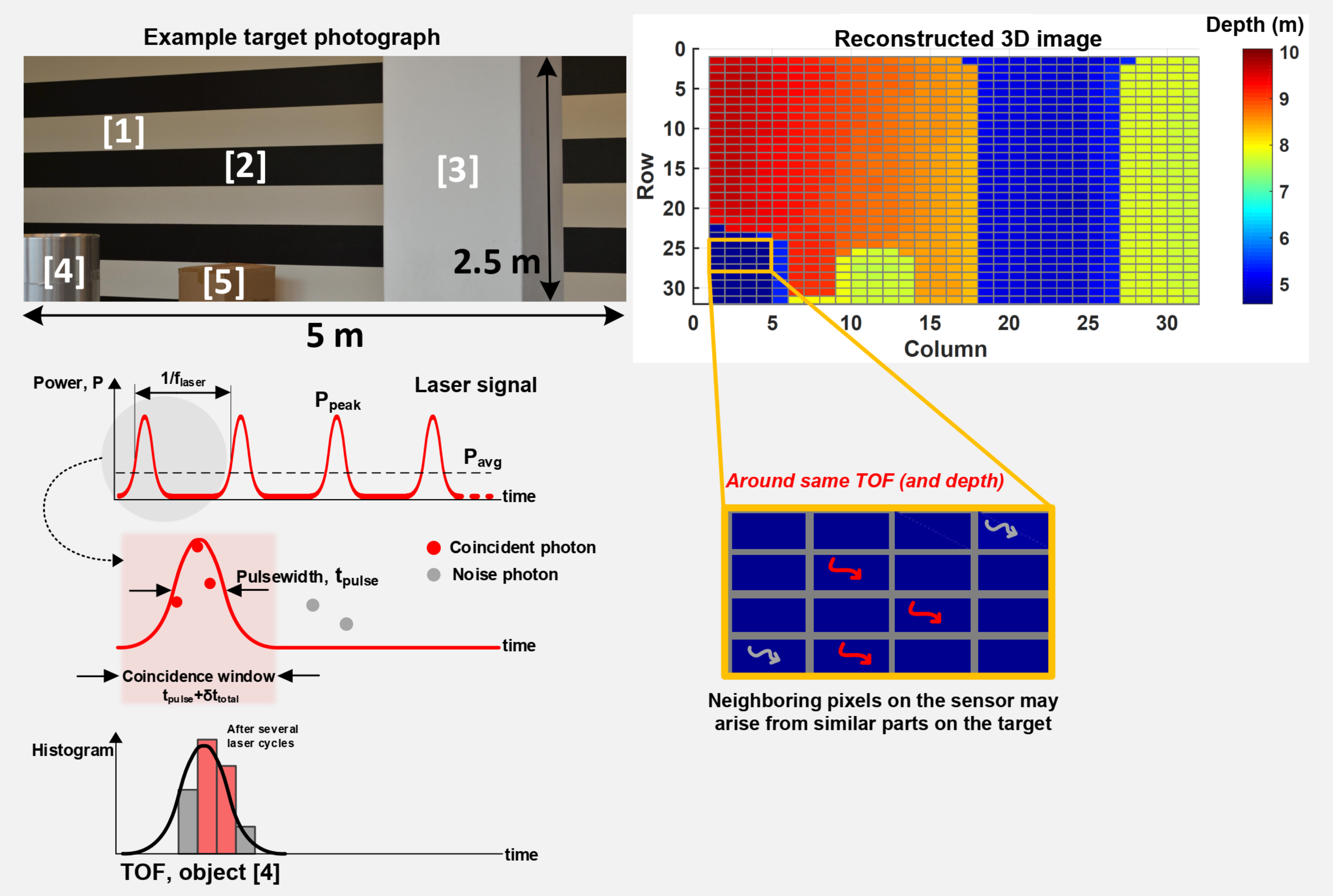
$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y)\, f_Y(y)\, dy$$



Conditional PDF of *Y* given *X* = *x*

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

Appendix 8.B – Multivariate Distributions – Example: LIDAR



Photon coincidences

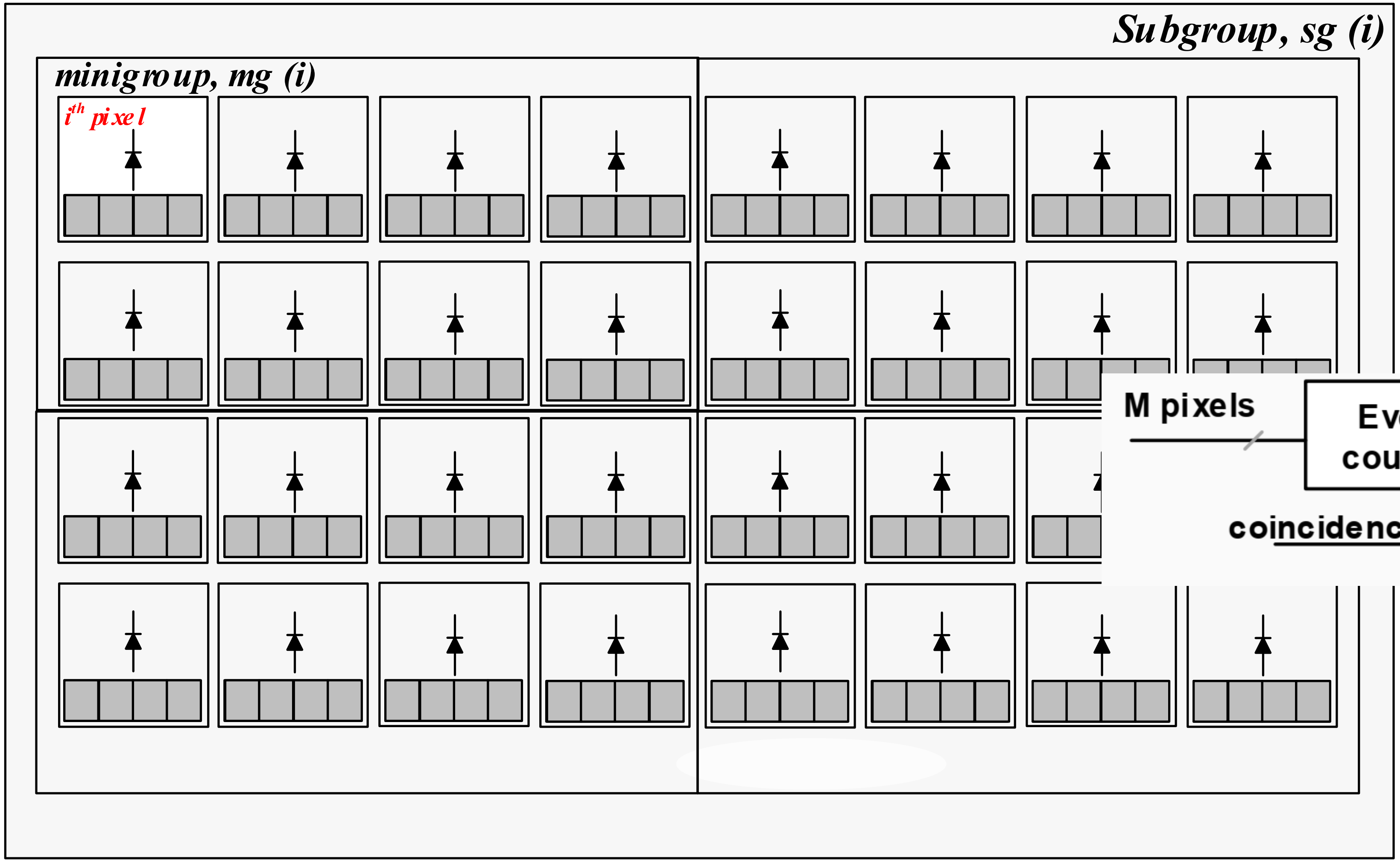
Coincidence detection is a well-known technique which utilizes spatio-temporal correlations of photons within a laser pulse to filter out background noise photons which are uniformly distributed in time

-> concept of coincidence window to reduce the likelihood of acquiring noise events

Slides courtesy of P. Padmanabhan

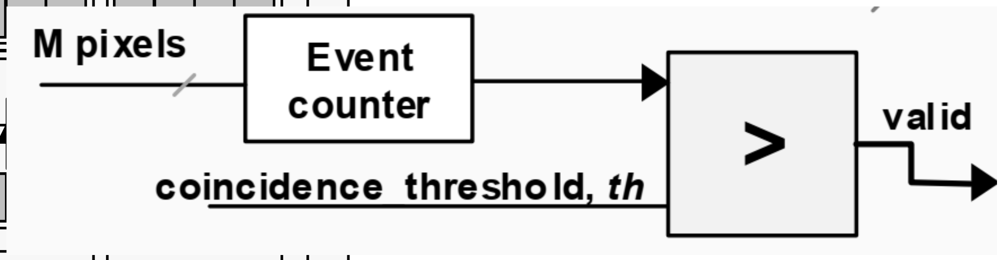
P. Padmanabhan et al., *Modeling and Analysis of a Direct Time-of-Flight Sensor Architecture for LiDAR Applications*, Sensors 2019

Appendix 8.B – Multivariate Distributions – Example: LIDAR



A possible pixel (and/or sensor) arrangement: *subgroups (sg)* of 8x4 SPADs, clustered into 4 *minigroups (mg)* of 8 SPADs each

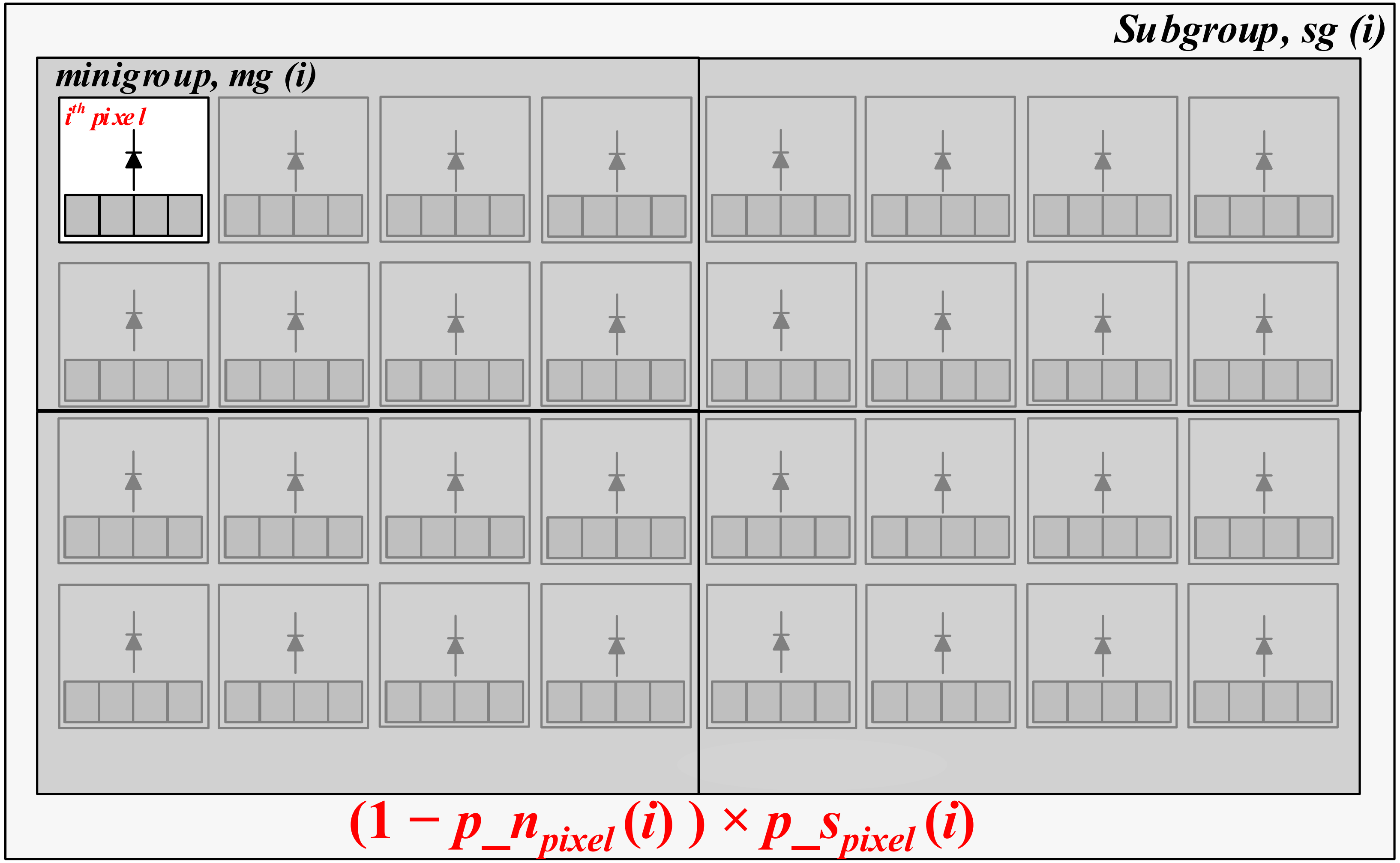
Arrival of the first event starts a coincidence window -> count events (photons) in a *sg*



-> compare the output of the event counter with a predefined (and variable) coincidence threshold ***th***.

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



(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*, = $p\_s_{pixel}(i)$,

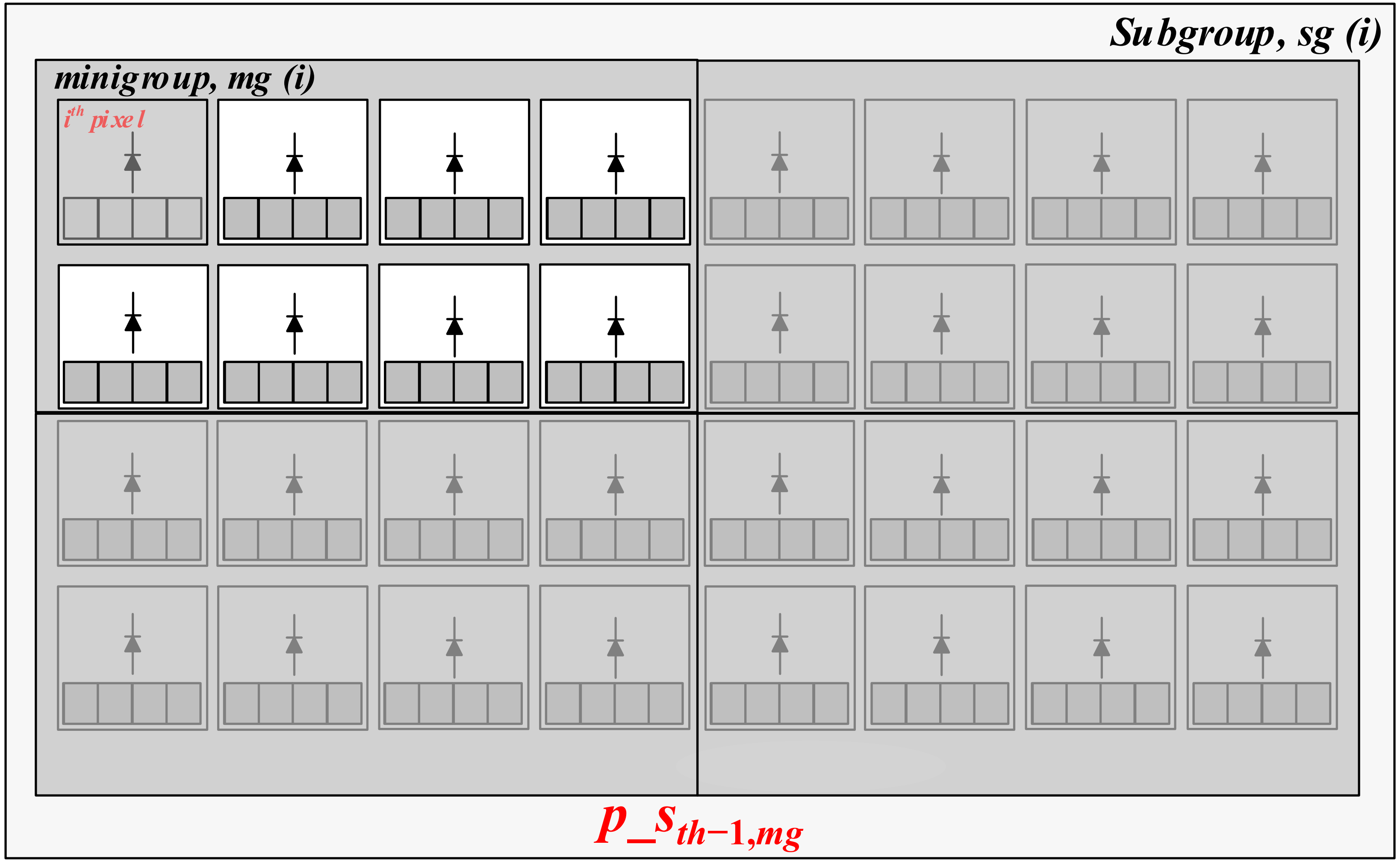
given that no noise photon is detected at pixel *i*, = $(1 - p\_n_{pixel}(i))$,

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).

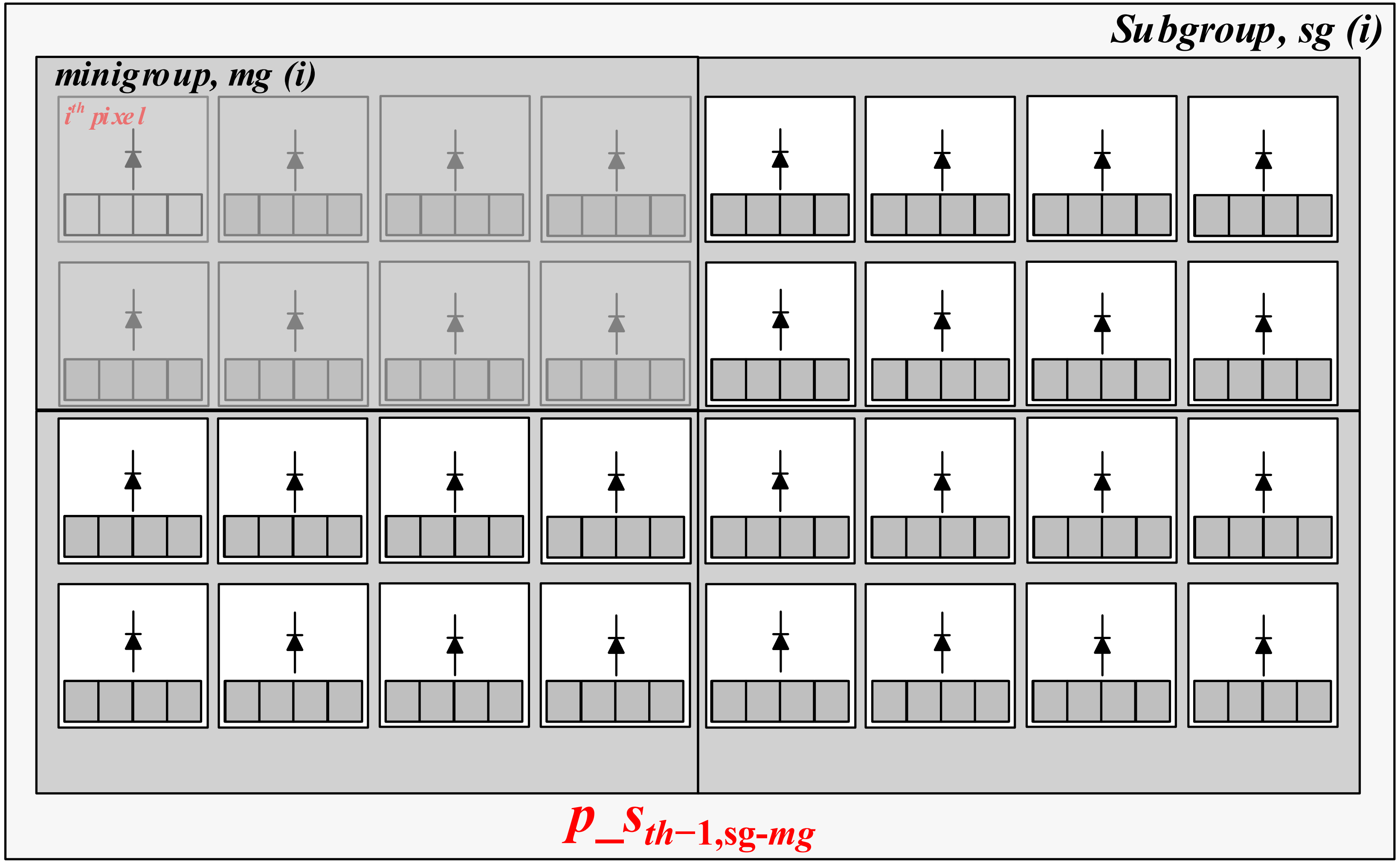
But $p\_s_{th-1}, sg(i)$ =

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*),

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

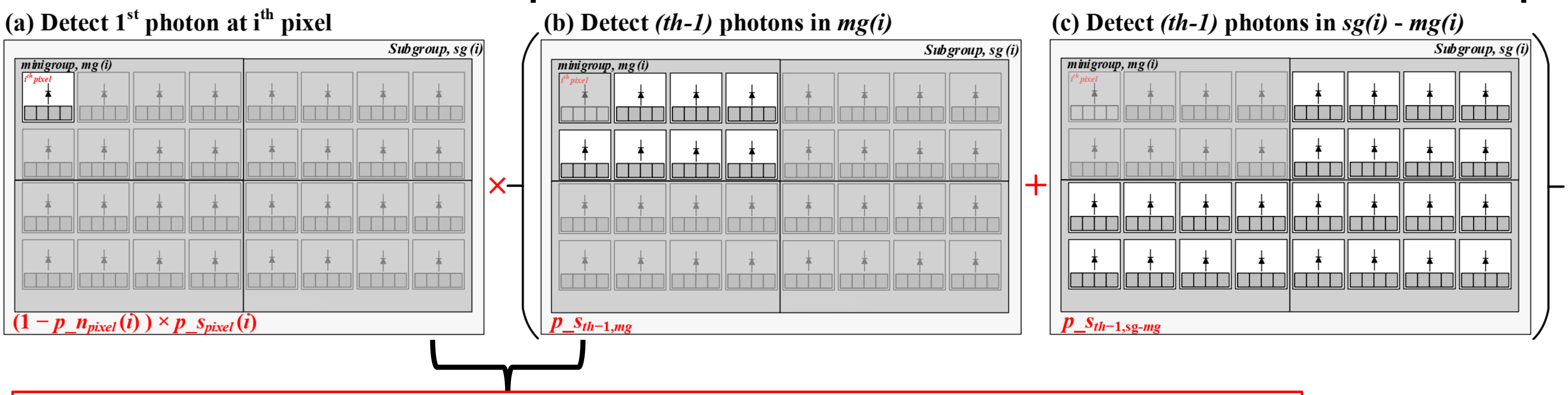
P. Padmanabhan et al., *Modeling and Analysis of a Direct Time-of-Flight Sensor Architecture for LiDAR Applications*, Sensors 2019

Appendix 8.B – Multivariate Distributions – Example: LIDAR

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