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 Nine: Elements of Statistics

Claudio Bruschini
TA: Alexandre Domenech (2021: Simone Frasca)
Advanced Quantum Architectures Laboratory (AQUA)
EPFL at Microcity, Neuchâtel, Switzerland

EPFL

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.

Week 8 Summary

8.1 **Introduction to Probability**: $P\{\mathcal{A}\}, P\{\mathcal{A}|\mathcal{B}\} \rightarrow \text{Bayes' rule, Law of Total Prob. (LOTP)}$, Independent Variables

8.2 **Random Variables**: discrete/continuous RV X and its distribution expressed as

PMF $p_X(x)$ / PDF $f_X(x) \leftrightarrow$ CDF $F_X(x)$

Examples: Binomial: $\text{Bin}(n, p)$, Poisson: $X \sim \text{Pois}(\lambda)$, Uniform: $U \sim \text{Unif}(a, b)$, Normal (Gaussian): $X \sim \mathcal{N}(\mu, \sigma^2)$, Exponential: $X \sim \text{Expo}(\lambda)$

8.3 **Moments**: RV X: expected value (mean) $E\{X\}$, variance $Var\{X\} = \sigma^2$ /standard deviation $SD\{X\} = \sqrt{Var\{X\}} = \sigma \rightarrow n$-th moment $E\{X^n\}$, central moment /standardized moment and their properties $\leftarrow$ moment generating function (MGF) $\phi(t) = E\{e^{tX}\}$

The lecture starts with a small recap of the main elements of the previous week.

Week 8 Summary

8.4 **Covariance and Correlation**:

Multiple RVs → Multivariate distributions (8.1, 8.2 →): joint → marginal, → conditional, Independent distributions

Covariance $Cov\{X, Y\} \rightarrow Corr\{X, Y\}$ (unitless version)

Variance of multivariate distributions:

1. $Var\{X + Y\} = Var\{X\} + Var\{Y\} + 2Cov\{X, Y\}$
2. $Var\{X_1 + \dots + X_n\} = Var\{X_1\} + \dots + Var\{X_n\} + 2\sum_{i<j} Cov\{X_i, X_j\}$

Outline

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

9.0.1 Uniform Distribution

- Uniform random variable in (a, b): completely random number between a and b

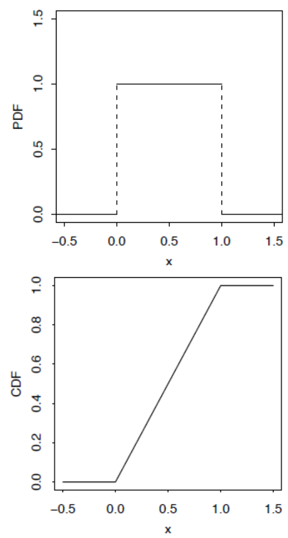
  -> PDF constant over chosen interval

- Uniform distribution $U \sim \mathrm{Unif}(a, b)$ in the interval (a, b) if:

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

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

Unif(0,1) PDF & CDF



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

We will now look again at some of the most important random variable distributions introduced in the previous lecture, and go further into their characteristics, in particular their **variance**. As before, their properties are going to be illustrated by means of examples from engineering and physics.

This slide is simply a summary of what previously shown in 8.2.6.

Can you think of random variables with this kind of distribution? E.g. (later) *quantization noise*…

9.0.1 Uniform Distribution (contd.)

- Probability is inversely proportional to length.
- Even in a sub-interval, we still have a uniform distribution

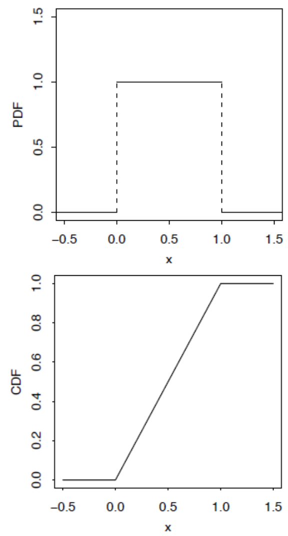
Mean: $E\{U\} = \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2}$

Second Order Moment: $E\{U^2\} = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{3}\frac{b^3 - a^3}{b-a}$

Variance*: $Var\{U\} = E\{U^2\} - (E\{U\})^2 = \frac{1}{3}\frac{b^3-a^3}{b-a} - \left(\frac{a+b}{2}\right)^2 =$

$$= \frac{(b-a)^2}{12}$$

Unif(0,1) PDF & CDF



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

**Using 8.3.4 (W8)*

We calculate the variance of a uniform distribution by using the Week 8 formula on variance in Section 8.3.4, $Var\{X\} = E\{X^2\} - E\{X\}^2$, rather than a direct calculation.

9.0.2 Standard Gaussian Distribution

- Gaussian (or Normal) distribution:
  - well-known continuous distribution with a bell-shaped PDF
  - widely used in statistics because of the central limit theorem (see next section)
- Standard Gaussian $Z \sim \mathcal{N}(0,1)$:

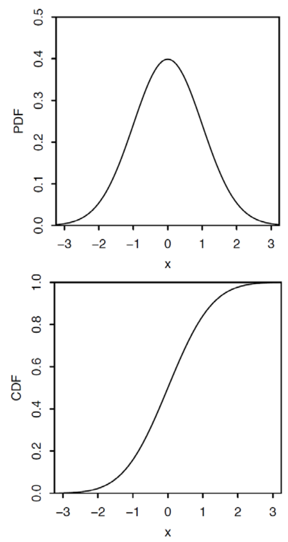
$$\text{PDF:} \quad \varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \qquad -\infty < z < \infty$$

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

No closed form available for the CDF. However, note that:

$$\int_{-\infty}^{\infty} e^{-z^2/2}\, dz = \sqrt{2\pi}$$

Standard Gaussian PDF/CDF



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

Next we move to the Normal, or Gaussian, distribution, starting from the *standard* version. This slide is simply a summary of what previously shown in 8.2.7.

9.0.2 Standard Gaussian Distribution (contd.)

- Properties: symmetry of PDF, symmetry of tail areas, of Z and $-Z$

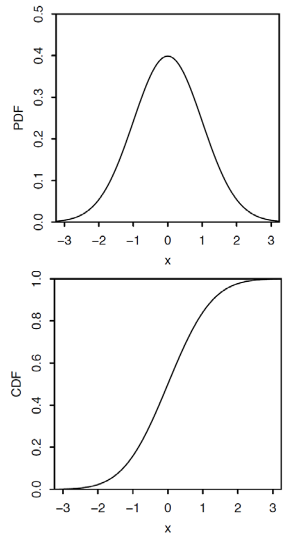
Mean: $$E\{Z\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z e^{-z^2/2}\, dz = 0$$

Variance $*$: $$Var\{Z\} = E\{Z^2\} - (E\{Z\})^2 = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z^2 e^{\frac{-z^2}{2}}\, dz =$$

$$= \frac{2}{\sqrt{2\pi}}\left(- z e^{-z^2/2}\Big|_0^{\infty} + \int_0^{\infty} e^{\frac{-z^2}{2}}\, dz\right) = \frac{2}{\sqrt{2\pi}}\left(0 + \frac{\sqrt{2\pi}}{2}\right) = 1$$

(integrating by parts)

Standard Gaussian PDF/CDF



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

Using 8.3.3 LOTUS (W8)

The variance is again calculated using the Week 8 formula in Section 8.3.4, $Var\{X\} = E\{X^2\} - E\{X\}^2$, whereby $E\{Z\}$=0.

$E\{Z$^2 $\}$ itself is calculated using LOTUS, 8.3.3, which states that $E\{g(X)\} = \int_{-\infty}^{\infty} g(x) f_X(x) dx$

9.0.2 Gaussian Distribution

- Gaussian (or Normal) distribution with any mean μ and variance σ: location-scale transformation of the standard Normal

$$X = \mu + \sigma Z$$

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

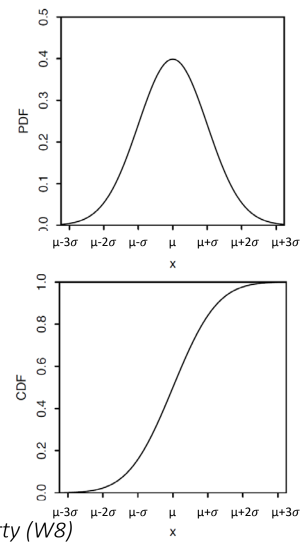
Mean*: $E\{X\} = E\{\mu + \sigma Z\} = E\{\mu\} + \sigma E\{Z\} = \mu$

Variance**: $Var\{X\} = Var\{\mu + \sigma Z\} = Var\{\sigma Z\} = \sigma^2 Var\{Z\} = \sigma^2$

- Standardisation process (from X back to Z):

$$\text{for } X \sim \mathcal{N}(\mu, \sigma^2), \quad \frac{X - \mu}{\sigma} \sim \mathcal{N}(0,1)$$

Gaussian PDF/CDF



*Using linearity property (W8)
** Using 8.3.4 (W8)

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

The same properties for a distribution with any mean μ and variance σ are then derived by using a location-scale transformation ($X = \mu + \sigma Z$).

We employ the linearity property of the Mean (8.3.3) and the properties of the Variance in 8.3.4.

9.0.2 Gaussian Distribution (contd.)

S

- General Gaussian CDF $F(x)$ and PDF $f(x)$:

$$\text{CDF:} \quad F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

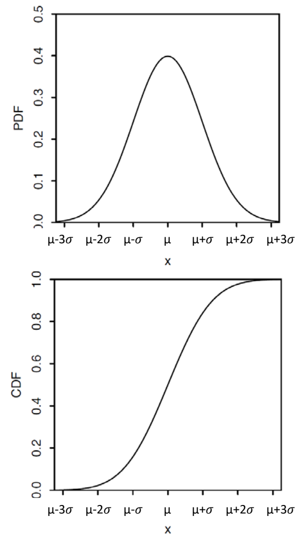
$$\text{PDF:} \quad f(x) = \varphi\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}$$

- *Proof:*

$$F(x) = P\{X \leq x\} = P\left\{\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right\} = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

$$f(x) = \frac{d}{dx}\Phi\left(\frac{x-\mu}{\sigma}\right) = \varphi\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma} = \frac{1}{\sqrt{2\pi}\sigma}exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Gaussian PDF/CDF



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

To complete the picture, we show how the PDF and CDF of the general and standard Gaussian are linked – see the first two equations, already detailed in 8.2.7.

9.0.2 Gaussian Distribution (contd.)

- Important properties – if $X \sim \mathcal{N}(\mu, \sigma^2)$,

$$P\{|X-\mu| < \sigma\} \approx 0.68$$

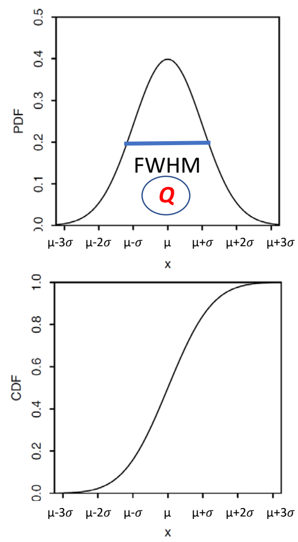
$$P\{|X-\mu| < 2\sigma\} \approx 0.95$$

$$P\{|X-\mu| < 3\sigma\} \approx 0.997$$

$$\text{Full Width Half Maximum (FWHM)} = P\{|X-\mu| < 1.175\sigma\}$$

$$FWHM = 2\sqrt{2\ln 2}\,\sigma \approx 2.355\,\sigma$$

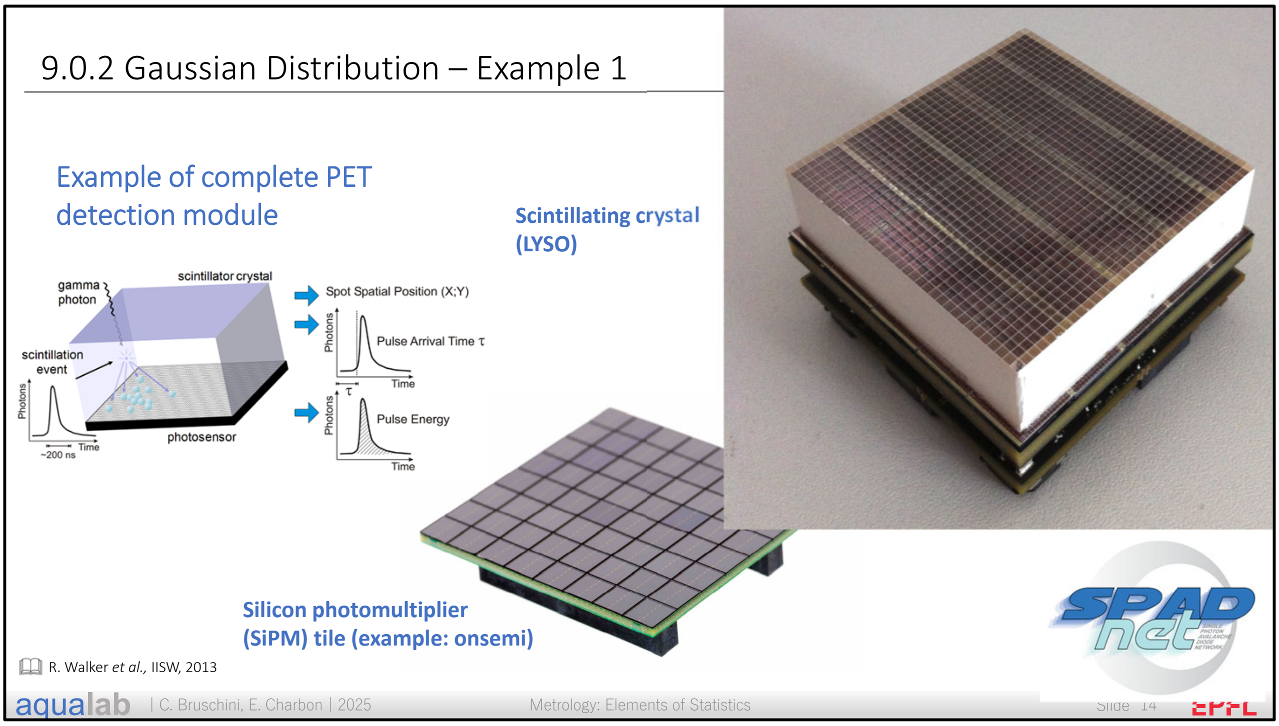
Gaussian PDF/CDF



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

The standard deviation σ and FWHM of a Gaussian distribution are linked as shown in this slide. Note that in some communities it is preferred practice to quote the standard deviation (e.g. physics), in others the FWHM (e.g. engineering).

NB: the relationship shown here is strictly speaking only valid for a Gaussian distribution!



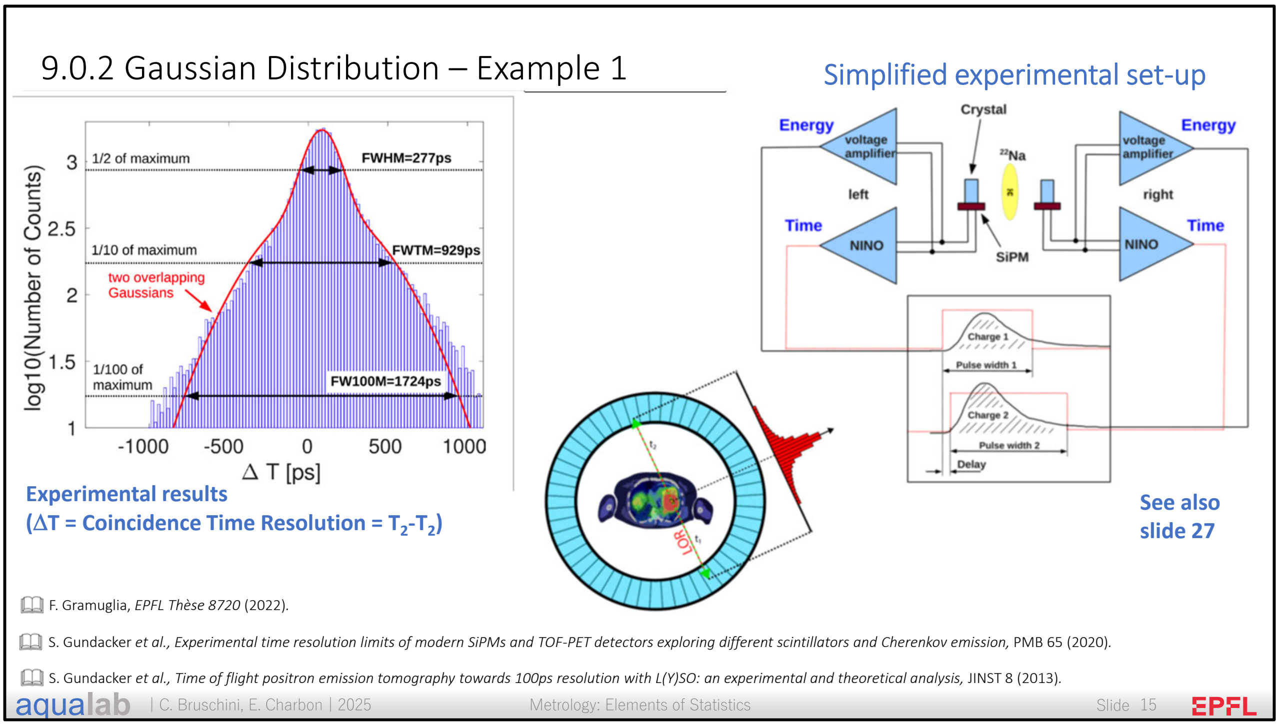
Let's now have a look at concrete examples from engineering and physics, linked to some of the distributions which we have seen before, and the corresponding measurement techniques.

The first one involves a PET (photo)detection module, to detect the gammas coming from the patient and extract the so-called *line of response* (LOR). We already discussed such modules in Section 8.2.9.

Left: schematic of detection module – shown here in simplified form as a single block – and the main scintillation light PDF, enabling the measurement of energy, time-of-arrival and position.

Bottom: example of a photodetector in the form of an array of silicon photomultipliers.

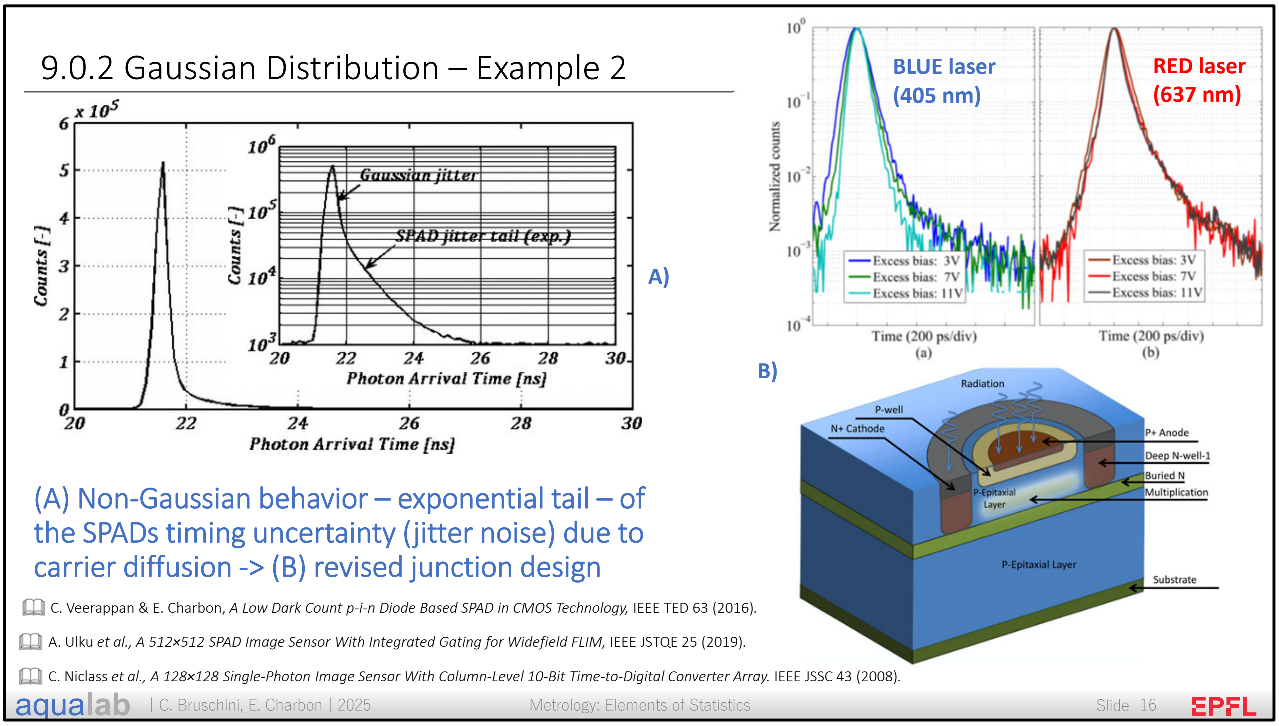
Right: a scintillator, built of small separate scintillating crystals (called "needles") rather than a monolithic block, sitting on top of a photodetector.



Left: example of a real experimental distribution of the timing difference of gamma events detected on a given LOR (line of response), by two scintillating crystals placed face-to-face. Same set-up as in Section 8.3.5.

Such timing distributions can be measured for example with the *analog* experimental set-up shown on the right (lab implementation for research purposes): a small radioactive source is placed between the two scintillating crystals. Their light output is measured by silicon photomultipliers, read out by dedicated amplifiers. The latter allow to extract the total charge in each SiPM electrical scintillation pulse, corresponding to the total released energy (i.e. the gamma energy), as well as the scintillation time (or time-of-arrival of the gamma), by placing a threshold which triggers an inverter.

NB: these quantities can also be measured in a digital way, by detecting individual photons and adding them up digitally, and measuring the time-of-arrival of one of them (e.g. the first) or more than one with time-to-digital converters (Section 9.3.4).



Another example of distributions (see also section 8.2.9 Example 3).

Left: 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 (*bottom right*), 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. Note also the difference between linear and logarithmic scales!

Top right: timing response of a SPAD when illuminated with lasers of different wavelengths.
Q: why do you think that there should be a difference? Where are blue vs. red photons preferentially absorbed in silicon? Which is the link to the SPAD structure?

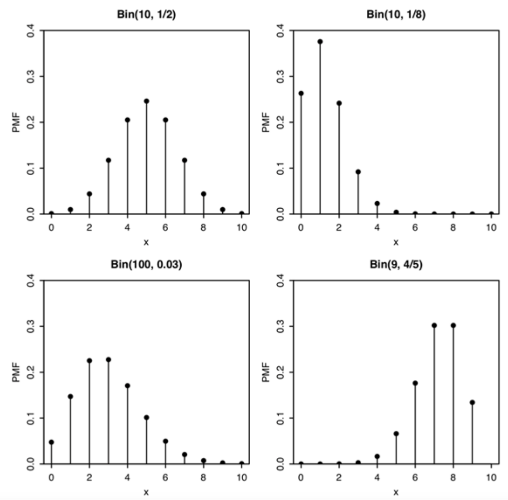
9.0.3 Binomial Distribution

- 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 if:

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

Mean: $$E\{X\}=\sum_{k=0}^{n}k\binom{n}{k}p^k(1-p)^{n-k}=np$$



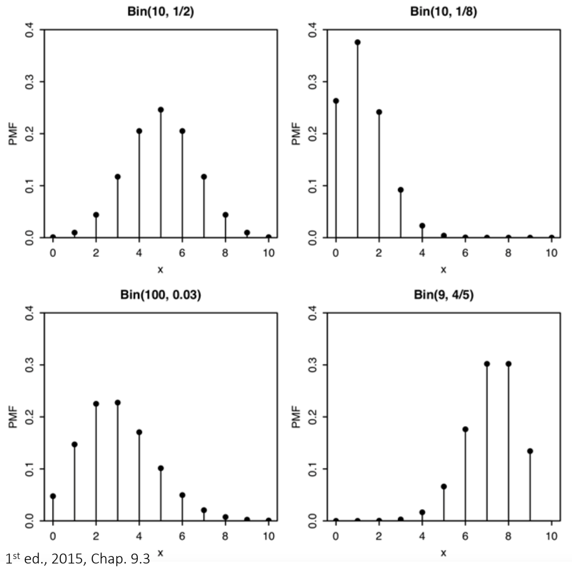
J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 3.3

Binomial distribution (see 8.2.2): recap of its PMF and calculation of its mean value.

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

9.0.3 Binomial Distribution

S



J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 9.3

9.0.4 Poisson Distribution

- Definition: a random variable $X \sim \text{Pois}(\lambda)$ has a Poisson distribution with parameter λ if its PMF:

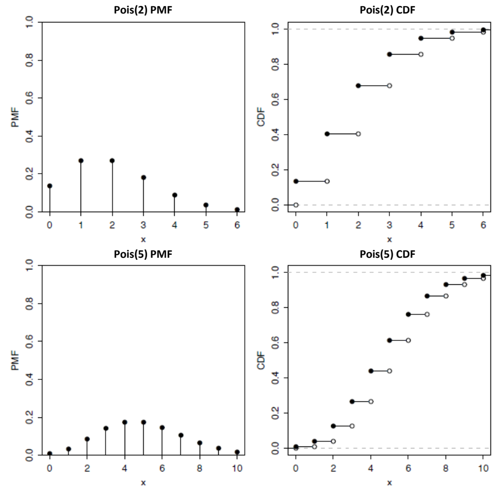
PMF: $$P\{X = k\} = \frac{e^{-\lambda}\lambda^k}{k!}, \qquad k = 0,1,2,\ldots$$

Mean: $$E\{X\} = e^{-\lambda}\sum_{k=0}^{\infty} k\frac{\lambda^k}{k!} = \lambda$$

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

$$= \lambda(1+\lambda) - \lambda^2 = \lambda$$

NB: Taylor series: $$\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{\lambda}$$



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

Poisson distribution (see 8.2.5): recap of its PMF, and calculation of its mean and variance.

Note that a) the Poisson distribution is characterised by a single parameter (λ), and b) that its mean is equal to its variance!

NB: details of the intermediate steps are in Blitzstein Section 4.7, a bit involved for the variance calculation. The Taylor expansion is used in some of them.

9.0.4 Poisson Distribution (contd.)

S

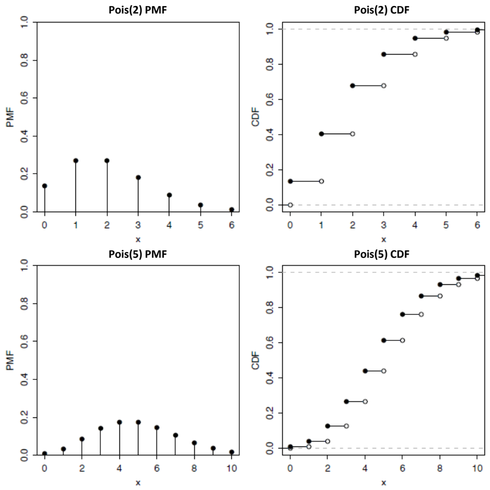
- The Poisson distribution has the following properties:

  1. If $X \sim \text{Pois}(\lambda_1)$ and $Y \sim \text{Pois}(\lambda_2)$ and X and Y are independent, then the distribution of

  $$X + Y \sim \text{Pois}(\lambda_1 + \lambda_2)$$

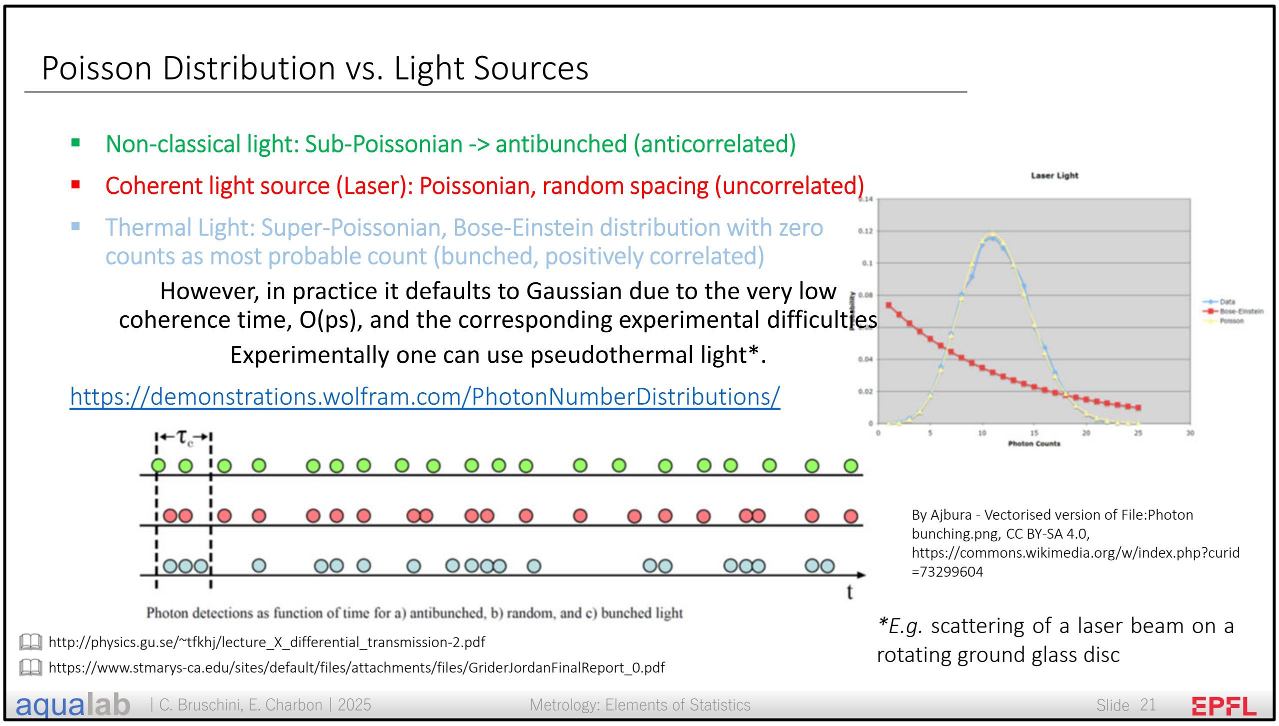
  2. If $X \sim \text{Pois}(\lambda_1)$ and $Y \sim \text{Pois}(\lambda_2)$ and X and Y are independent, then the conditional distribution of X given $X + Y = n$ is:

  $$P(X = k | X + Y = n) \sim \text{Bin}(n, \lambda_1/(\lambda_1 + \lambda_2))$$



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

Properties of the Poisson distribution. The first one is interesting: the average of $X + Y$, being X and Y two independent Poissonian random variables, is again Poissonian, with average value equal to the sum of the average values.



Is light always distributed in a Poissonian way?

Not necessarily: some sources have non-Poissonian distributions, such as thermal ones (super-Poissonian -> bunched photon arrival times, resulting in a Bose-Einstein distribution with zero counts as most probable value, but in practice difficult to observe due to the very low coherence times and the corresponding experimental difficulties).

In the case of coherent light sources (e.g. laser), the resulting Poissonian distribution can be derived directly from first principles.

Certain non-classical quantum light sources allow to reach sub-Poissonian distributions, and thus sub shot-noise-limited behaviour.

Bottom: example of antibunched, random and bunched light sources.

____

Poisson Distribution vs. Light Sources

$\bar{n}$ = average photon number

- Non-classical light: Sub-Poissonian

$$\sigma < \sqrt{\bar{n}}$$

- Coherent light source (Laser): Poissonian

$$P(n) = \frac{\bar{n}^n}{n!} e^{-\bar{n}}, \sigma = \sqrt{\bar{n}}$$

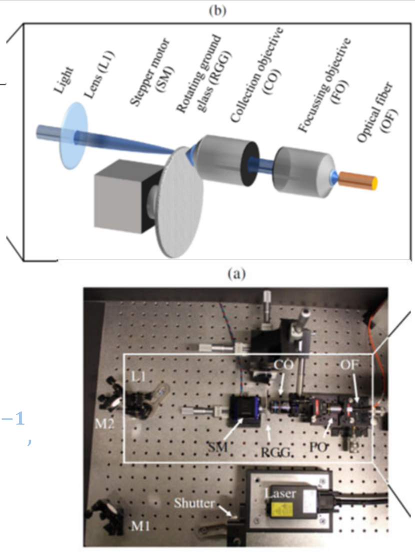
For large photon numbers, the relative fluctuations $\sigma / \bar{n}$ tend to 0

- Thermal Light: Super-Poissonian, Bose-Einstein distribution

$$P(n) = \left(1 - e^{-\hbar\omega/k_B T}\right) e^{-n\hbar\omega/k_B T} = \frac{\bar{n}^n}{(\bar{n}+1)^{n+1}}, \bar{n} = \left(e^{\hbar\omega/k_B T} - 1\right)^{-1},$$

$$\sigma = \sqrt{\bar{n}^2 + \bar{n}} \ \ (for\ T \ll \tau_c) > \sqrt{\bar{n}}$$

For large photon numbers, the relative fluctuations $\sigma / \bar{n}$ tend to 1



Pseudothermal light source

Advanced Lab Course (F-Praktikum), Exp. 45, *Photon Statistics*, v. Aug. 21 2017
http://physics.gu.se/~tfkhj/lecture_X_differential_transmission-2.pdf
https://www.stmarys-ca.edu/sites/default/files/attachments/files/GriderJordanFinalReport_0.pdf
T. Stagner *et al.*, *Step-by-step guide to reduce spatial Coherence of laser light using a rotating ground glass diffuser*, OSA Applied Optics 56 (2017).

Which are the variances of these three types of light sources emitting a given average photon number $\bar{n}$?

→ Compare the resulting three standard deviation values in absolute terms, and also relative ones, i.e. compared to the average (this ratio is basically the source's SNR = signal-to-noise ratio).

Thermal light: For large average photon numbers $\bar{n}$, the quantum mechanical Bose-Einstein distribution becomes identical to the Boltzmann distribution (classical limit) → exponential.

Right: experimental set-up allowing to create in the lab in a simple way a "pseudothermal" light source.

9.0.5 Exponential Distribution

- A continuous variable $Y \sim \mathrm{Expo}(\lambda)$ has an Exponential distribution with parameter λ if:

$$\text{PDF:} \quad f_Y(y) = \lambda e^{-\lambda y}, \qquad y > 0$$

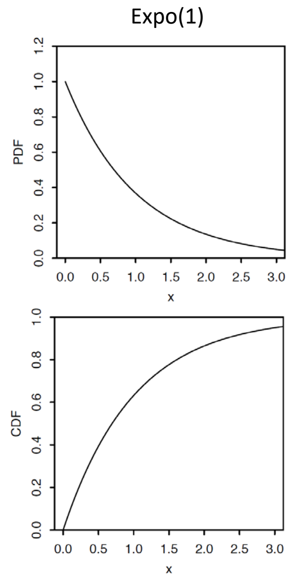
$$\text{CDF:} \quad F_Y(y) = 1 - e^{-\lambda y}, \qquad y > 0$$

- If we start from $X \sim Expo(1)$:

$$E\{X\} = \int_0^\infty x e^{-x}\, dx = 1$$

$$E\{X^2\} = \int_0^\infty x^2 e^{-x}\, dx = 2$$

$$Var\{X\} = E\{X^2\} - (E\{X\})^2 = 1$$



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

Exponential distribution: note again the presence of one single variable λ.

Bottom: calculation of the mean and variance for the standard version = $Expo(1)$.

9.0.5 Exponential Distribution (contd.)

- In general, for $Y = X/\lambda \sim \mathrm{Expo}(\lambda)$ (scaling), we get:

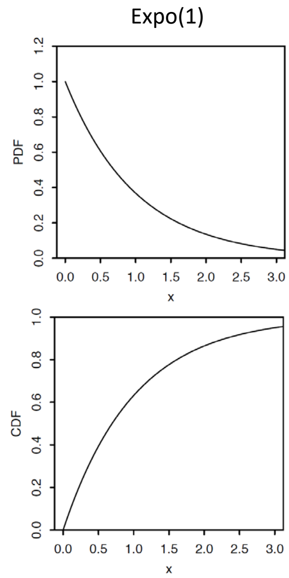
$$\text{Mean:} \quad E\{Y\} = \frac{1}{\lambda}E\{X\} = \frac{1}{\lambda}$$

$$\text{Variance:} \quad Var\{Y\} = \frac{1}{\lambda^2}Var\{X\} = \frac{1}{\lambda^2}$$

- Recap*: «An* $\mathrm{Expo}(\lambda)$ *RV represents the waiting time for the first success in continuous time; the parameter* λ *can be interpreted as the rate at which successes arrive."*
- Memoryless property*: «conditional on our having waited a certain amount of time (s) without success, the distribution of the remaining wait time (t) is exactly the same as if we hadn't waited at all.»*

$$\text{Memoryless:} \quad P\{Y \geq s + t | Y \geq s\} = P\{Y \geq t\}$$

Ex



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

Using a scaling transformation ($Y = X/\lambda$), we can calculate the mean and variance for a general exponential distribution.

The standard deviation is equal to the mean, which implies a broad distribution!

Note 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.

The Memoryless property is demonstrated in one of the exercises and is also illustrated in the next slide.

9.0.5 Exponential Distribution (contd.)

- Memoryless property: *«conditional on our having waited a certain amount of time (s) without success, the distribution of the remaining wait time (t) is exactly the same as if we hadn't waited at all.»*

Memoryless: $P\{Y \geq s + t | Y \geq s\} = P\{Y \geq t\}$

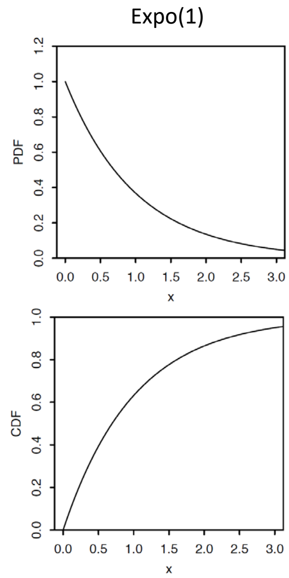
$$e.g. P\{Y \geq 40 | Y \geq 30\} = P\{Y \geq 10\}$$

$$e.g. P\{Y \geq 70 | Y \geq 60\} = P\{Y \geq 10\}$$

PDF: $f_Y(y) = \lambda e^{-\lambda y}, y_2 = y_1 + \Delta t$

$$\frac{f_Y(y_2)}{f_Y(y_1)} = \frac{\lambda e^{-\lambda y_2}}{\lambda e^{-\lambda y_1}} = e^{-\lambda\, \Delta t} = constant$$

$$e.g. \frac{f_Y(y_2 = 4\, \lambda^{-1})}{f_Y(y_1 = 3\, \lambda^{-1})} = \frac{f_Y(y_2 = 2\, \lambda^{-1})}{f_Y(y_1 = 1\, \lambda^{-1})} = e^{-1} = constant$$



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

9.0.5 Exponential Distribution – Example 1 & 2

Radioactive decay

- Universal law of radioactive decay:
  - A nucleus has “no memory”
  - A nucleus does not age with the passage of time

  -> a nucleus is equally likely to decay at any instant in time

  -> constant decay probability

Decay Law: $\frac{dN}{dt} = -\lambda N \Rightarrow N(t) = N_0 e^{-\lambda t}$

- NB: The number of decays in a given time interval in a radioactive sample is Poisson distributed…

Fluorescence lifetime $[S_1] = [S_1]_0 e^{-\Gamma t}$

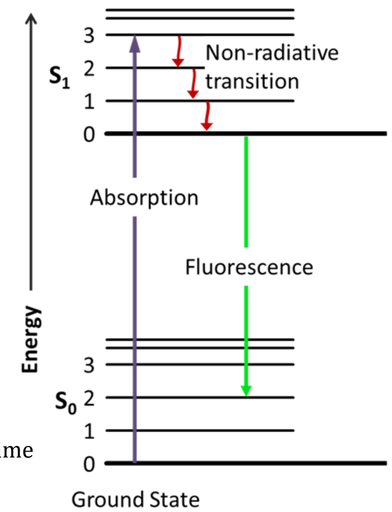
Q

S_1 = concentration of excited state molecules

Γ = decay rate = inverse of fluorescence lifetime = average length of time to decay from one state to another

EN Wikipedia *Radioactive_decay / Fluorescence*

Jablonski diagram



By Jacobkhed - Own work, CC0, https://commons.wikimedia.org/w/index.php?curid=19180813

Let’s have a look at two examples of important **exponential distributions**, namely the radioactive decay, and the fluorescence of molecules.

Right: Jablonksi diagram, showing the main transitions which come into play after a molecule has been excited. The non-radiative transitions, e.g. due to vibrational states, are “needed” so that the wavelength of the emission is larger than the wavelength of the absorption (Stokes shift) – see also Section 8.2.9.

This has the advantage that the excitation light beam can be separated from the emission light beam. In addition, given that there are multiple transition possibilities, the absorption and emission spectra are broad rather than sharp.

Q: which are the typical lifetimes involved? → see again Section 8.2.9.

9.0.5 Exponential Distribution – Example 3

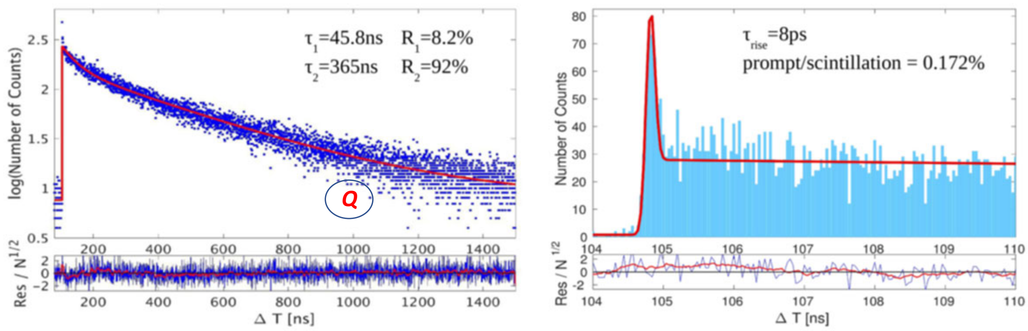


Figure 10. Scintillation decay and rise time of BGO measured with a time correlated single photon counting (TCSPC) setup using 511 keV annihilation gammas (Gundacker *et al* 2016b). The figure on the right hand side shows a pronounced Cherenkov peak at the onset of the scintillation emission with a relative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

"Physical experiments are imprecise and generate errors handled by statistical methods."
(I. Vardi)

See also slide 14

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

See section 8.2.9 Example 4:

These are the results of a **precision (timing) measurement**, for example using a radioactive source and detecting as many visible light photons as possible emitted from a scintillating crystal excited by a radioactive source, event after event, similarly to a TCPSC (time-correlated single-photon counting) method.

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!

Q: Note also the fluctuations on the right side of the scale, where the recorded data is quite small. ***"Physical experiments are imprecise and generate errors handled by statistical methods."*** *(I. Vardi)*

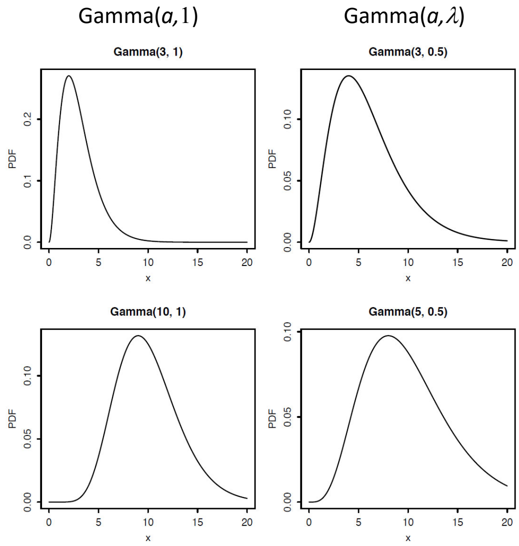
9.0.6 Gamma Distribution

S

- Let $X_1, X_2, \dots, X_n$ be n i.i.d. $\mathbf{Expo}(\lambda)$. Then:

$$Y = X_1 + \cdots + X_n \sim \mathbf{Gamma}(n, \lambda)$$

- The Gamma is nothing else but the distribution obtained by summing up n independent exponential distributions.



J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 8.4

The Gamma distribution comes into play in Appendix A and B. We are not going to discuss it in detail during the lecture.

NB: this has nothing to do with the detection of gamma rays!

9.0.6 Gamma Distribution (contd.)

S

- For the more general gamma distribution $Y = X/\lambda \sim \text{Gamma}(a, \lambda)$, by simple transformation, we obtain:

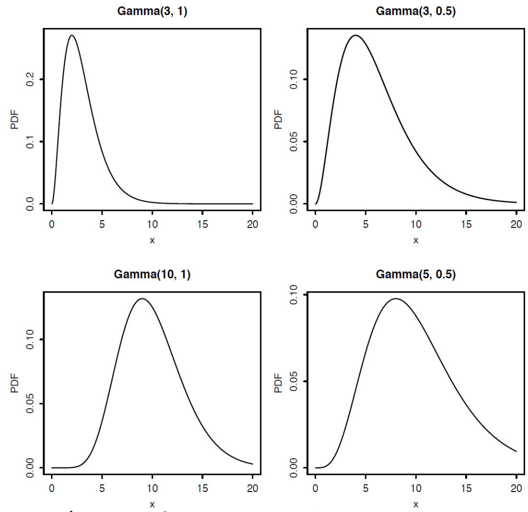
$$\text{Mean:} \quad E\{Y\} = \frac{1}{\lambda} E\{X\} = \frac{a}{\lambda}$$

$$\text{Second Moment:} \quad E\{Y^2\} = \frac{1}{\lambda^2} E\{X\} = \frac{a\,(a+1)}{\lambda^2}$$

$$\text{Variance:} \quad Var\{Y\} = \frac{1}{\lambda^2} Var\{X\} = \frac{a}{\lambda^2}$$

-> See Appendix A for details

Gamma$(a, 1)$ Gamma(a, λ)



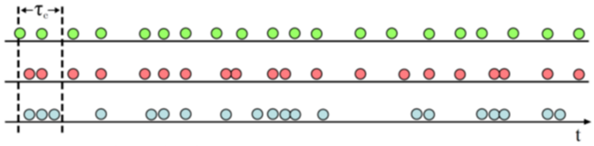
-> calculate mean/variance for some examples

J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 8.4

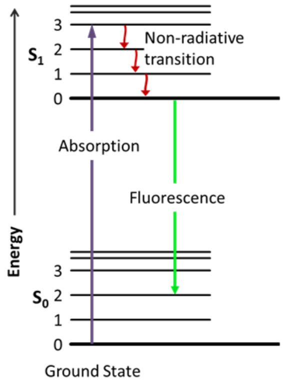
Take-home Messages/W9-1

- *Random Variables (RVs):*
  - Distributions: Uniform, Gaussian, Binomial
  - Distributions: Poisson ↔ Exponential

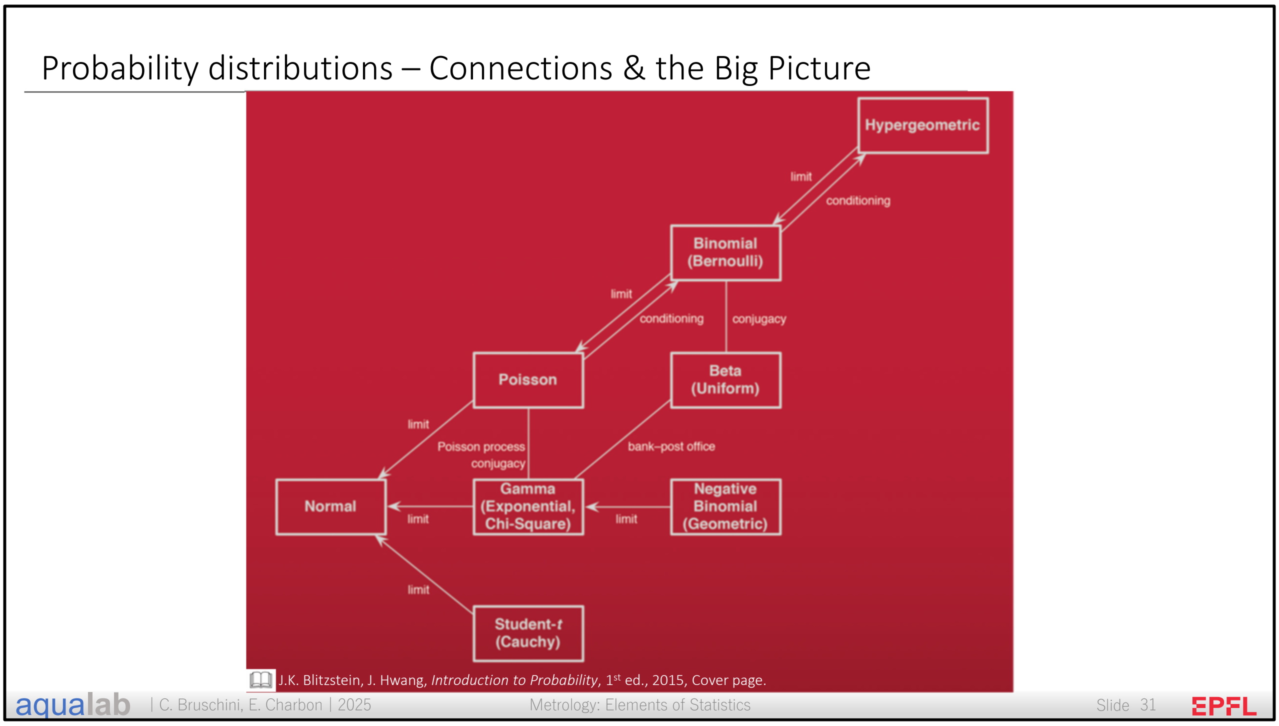
    … and their PDF, CDF, Mean, Variance
  - Practical examples!
    - Scintillation light (two crystals in coincidence) – combination of distributions ↔ experimental set-up
    - Timing jitter – combination of distributions ↔ experimental set-up
    - *Poisson Distribution vs. Light Sources*
    - Fluorescence lifetime & exponential decay
    - Scintillation light (one single crystal) ↔ experimental set-up



Photon detections as function of time for a) antibunched, b) random, and c) bunched light



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



Final note: many of the distributions which we have encountered are interconnected! As an example, we have seen the link between a Poisson and an exponential distributions, but there are far more. More details are provided in the Blitzstein/Hwang.

Outline

9.1.1 Random Process

- A Random (or stochastic) Process (RP) is a time-varying function that assigns the outcome of a random experiment to each time instant $X(t)$

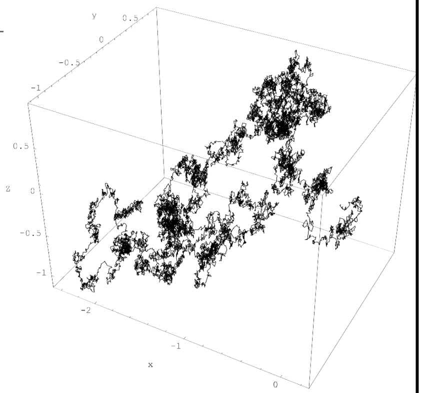
  Example: a current fluctuating due to thermal noise (-> Week 10), the growth of a bacterial population, the movement of a gas molecule [Wikipedia *Stochastic Process*]

- For fixed t, a Random Process is a Random Variable

- A Random Process can therefore be viewed as a **collection of an infinite number of Random Variables**. Given that $X_i = X(t_i)$:

$$\text{joint PDF:} \quad f_X(X_1, X_2, \dots, X_n, t_1, t_2, \dots, t_n)$$

- A Random Process can be either continuous or discrete



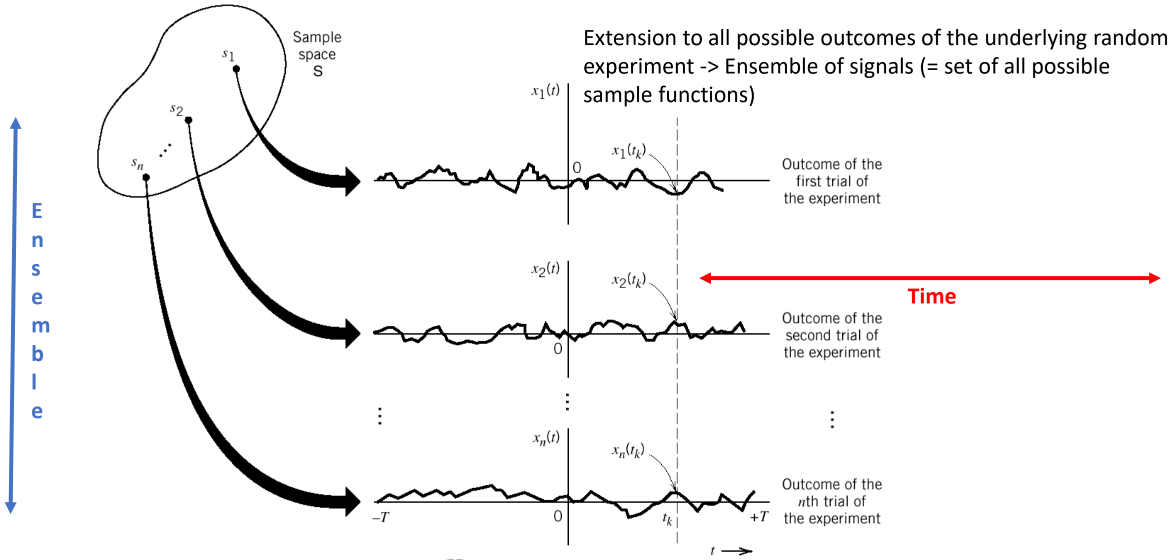
Original uploader was Sullivan.t.j at English Wikipedia. – 3D Brownian motion process. This mathematical image was created with Mathematica., CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=2249027

H. Bilgekul, Slides for the course "EE-461 Communication System II", EMU   F. Farahmand, Slides for the course "CES 540 Digital Communication", Ch. 6, SSU

We now generalize a random variable and describe the characteristics of a Random Process, which is basically a collection of Random Variables as a function of time.

9.1.1 Random Process – Example

Extension to all possible outcomes of the underlying random experiment -> Ensemble of signals (= set of all possible sample functions)

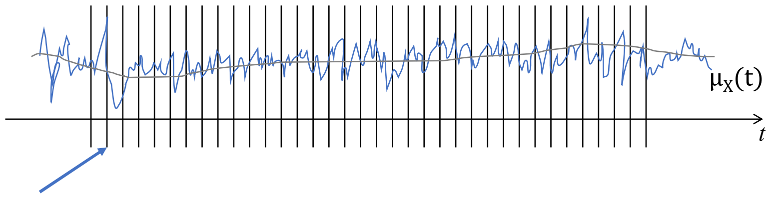


H. Bilgekul, Slides for the course "EE-461 Communication System II", EMU   F. Farahmand, Slides for the course "CES 540 Digital Communication", Ch. 6, SSU

This slide is key in understanding the properties of a Random Process as an ensemble of signals = outcomes of different trials, and how it can be analysed from an ensemble perspective ("vertically", i.e. at a fixed time), or from a time perspective ("horizontally", i.e. for a given experiment). This notation will be used in the following slides as well.

9.1.1 Random Process – Example

- Example: Noise is generally modeled as a random process, i.e. a collection of random variables, one for each time instant *t* in interval]-∞,+∞[



Fixed t: Random Process becomes a Random Variable

E. Charbon, "Image Sensors – ET 4390 Course Slides", Delft 2016

9.1.1 Random Process (contd.) – Characterization/1

- A Random Process is characterized by the same functions already explained for RVs, but which now depend on ***t***, i.e.:

CDF: $F_X(x,t) = P\{X(t) \leq x\}$ $\qquad X(t)$= random variable at time t

PDF: $f_X(x,t) = \dfrac{dF_X(x,t)}{dx}$

Mean: $m_X(t) = \overline{X(t)} = E\{X(t)\} = \int_{-\infty}^{\infty} x\, f_X(x,t)\, dx$

Second Order Moment: $\overline{X^2(t)} = E\{X^2(t)\} = \int_{-\infty}^{\infty} x^2\, f_X(x,t)\, dx$

Variance: $Var\{X(t)\} = E\left\{\left(X(t) - m_X(t)\right)^2\right\} = \int_{-\infty}^{\infty} \left(x - m_X(t)\right)^2 f_X(x,t)\, dx$

Ensemble averages

H. Bilgekul, Slides for the course "EE-461 Communication System II", EMU

F. Farahmand, Slides for the course "CES 540 Digital Communication", Ch. 6, SSU

We can extend the previously acquired statistical definitions and tools to a Random Process, which now become dependent on the new variable time ***t***.

At a fixed time ***t***, we are basically carrying out Ensemble averages (see the vertical arrow).

9.1.1 Random Process (contd.) – Characterization/2

- However, in order to characterize a RP, we need to introduce two more functions, e.g. to indicate how rapidly a RP changes in time:

$$\text{Auto}-\text{covariance:} \quad C_{XX}(t_1,t_2) = Cov\{X(t_1),X(t_2)\}$$

$$\text{Auto}-\text{correlation:} \quad K_{XX}(t_1,t_2) = E\{X(t_1)\cdot X(t_2)\}$$

$$\text{NB:} \quad C_{XX}(t_1,t_2) = E\{[X(t_1)-m_X(t_1)][X(t_2)-m_X(t_2)]\} = \\ = K_{XX}(t_1,t_2) - m_X(t_1)m_X(t_2)$$

- In a similar way we can also define:

$$\text{Cross}-\text{covariance:} \quad C_{XY}(t_1,t_2) = Cov\{X(t_1),Y(t_2)\}$$

$$\text{Cross}-\text{correlation:} \quad K_{XY}(t_1,t_2) = E\{X(t_1)\cdot Y(t_2)\}$$

$X(t_1)$= random variable at time t_1
$X(t_2)$= random variable at time t_2

NB: in general, the *autocorrelation* is the correlation of the signal with a delayed copy of itself (similarity between observations as a function of the time lag between them) [Wikipedia "autocorrelation"]

Cross-correlation: same but between two series

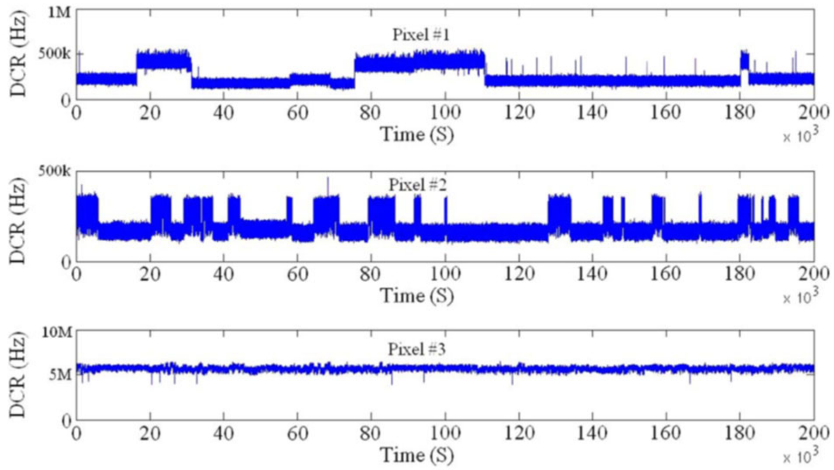
NB: extended here to two RPs X and Y

H. Bilgekul, Slides for the course "EE-461 Communication System II", EMU F. Farahmand, Slides for the course "CES 540 Digital Communication", Ch. 6, SSU

Of particular importance in the characterisation of a Random Process are the auto-covariance and auto-correlation, which basically determines how similar the random variable distribution is at two different times t_1 and t_2.

NB: an extension to more than one random process is also possible (last two equations), but this will not be studied in further detail.

9.1.1 Random Process (contd.) – Example, non-stationary



M. A. Karami et al., *Random Telegraph Signal in Single-Photon Avalanche Diodes*, International Image Sensor Workshop, Bergen, 2009

An example of a non-stationary Random Process is provided by the noise behaviour of irradiated SPADs, which can exhibit a so-called RTS (Random Telegraph Signal) behaviour over long periods of time. The noise level, or Dark Count Rate (DCR), does basically jump between two or more different levels. It will be discussed in more detail in the next lecture and is shown here for 3 different devices of the same type.

9.1.2 Stationary Random Process

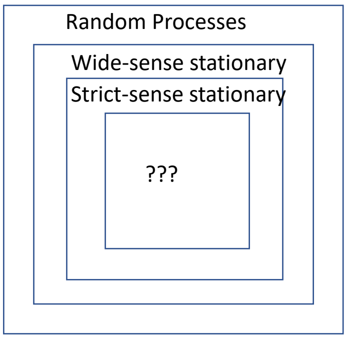
- We characterize the RP on how their statistical properties change in time. If they do not change, we call the RP stationary. Hence:

$$f_X(x,t) = f_X(x)$$

Ensemble averages

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

$$Var\{X(t)\} = E\left\{\left(X(t) - m_X(t)\right)^2\right\} = \int_{-\infty}^{\infty} \left(x - m_X(t)\right)^2 f_X(x,t)\, dx = \sigma^2$$



- Weaker form: in Wide-Sense Stationary RPs, in addition to a constant mean, the autocorrelation function only depends on the time difference, but not on the absolute position in time:

$$K_{XX}(t, t+\tau) = K_{XX}(\tau) \quad \text{(or equivalently } K_{XX}(t_1,t_2) = K_{XX}(t_2 - t_1)\text{)}$$

Ex **WSS random process does not drift with time**

H. Bilgekul, Slides for the course "EE-461 Communication System II", EMU

F. Farahmand, Slides for the course "CES 540 Digital Communication", Ch. 6, SSU

The behaviour of a Random Process is not always easy to characterise. A special sub-category is represented by a stationary RP, whose statistical properties don't change over time - which does obviously not mean that the values which the underlying random variables assume are constant!

A weaker form is represented by **Wide-Sense Stationary RPs**. An analytical example will be shown in the exercises.

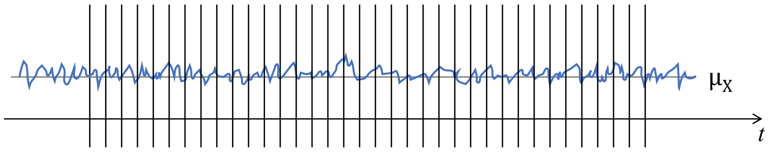
NB: for WSS, the variance would "automatically" also be constant. (-> R. Mauro p. 10, (2.22)).

9.1.2 Stationary Random Process – Example

$$f_X(x,t) = f_X(x)$$

$$m_X(t) = \overline{X(t)} = \mu_X$$

$$Var\{X(t)\} = \sigma^2$$



E. Charbon, "Image Sensors – ET 4390 Course Slides", Delft 2016

9.1.2 Stationary Random Process (contd.)

- For a Wide-sense Stationary Random Process $X(t)$, the autocorrelation function has the following properties:

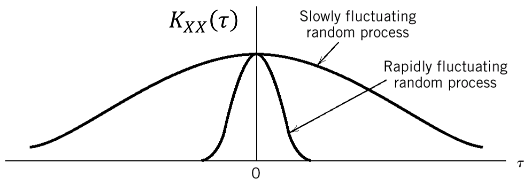
1. $K_{XX}(t_1, t_1) = K_{XX}(t_2, t_2) = K_{XX}(0) = E\{X^2(t)\} = \overline{X^2(t)} \geq 0$ ($\Longrightarrow K_{XX}(0)$ = *total power* of random signal $X(t)$, does not change in time)

2. $K_{XX}(\tau) = K_{XX}(-\tau)$

3. $\lim_{|\tau| \to \infty} K_{XX}(\tau) = \lim_{|\tau| \to \infty} E\{X(t) \cdot X(t+\tau)\} =$

   $= E\{X(t)\}\, E\{X(t+\tau)\} = \overline{X(t)}^{\,2}$ (example: *average or DC power* of random signal $X(t)$)

4. $|K_{XX}(\tau)| \leq |K_{XX}(0)|$ for all τ



F. Farahmand, Slides for the course “CES 540 Digital Communication”, Ch. 6, SSU
H. Bilgekul, Slides for the course “EE-461 Communication System II”, EMU

The autocorrelation function of wide-sense stationary RPs has the following properties. NB:

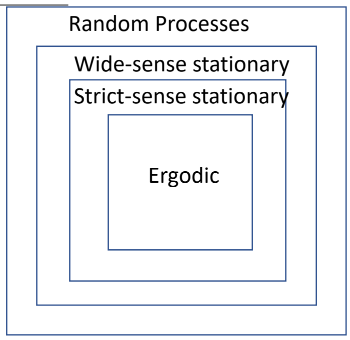
1. Reuses the main property of the autocorrelation function of a WSS RP, i.e. $K_{XX}(t_1, t_2) = K_{XX}(t_2 - t_1)$, and that $\overline{X^2(t)} = E\{X^2(t)\}$

1. and 3.: note the difference between the *total power* of a signal $\overline{X^2(t)}$ and its *average or DC power* $\overline{X(t)}$.

3.
- First line: assumption that two RVs at very distant times ($|\tau| \to \infty$) are basically uncorrelated -> $E\{XY\} = E\{X\}E\{Y\}$…

- Second line: the mean is constant for a WSS RP -> $E\{X(t)\} = E\{X(t+\tau)\}$. But $E\{X(t)\} = \overline{X(t)}$ …

9.1.3 Ergodicity



- "A Random Process is ergodic if any sample function of the process takes all possible values in time with the same relative frequency that an ensemble will take at any given instant". Basically, its statistical properties can be deduced from a single, sufficiently long, random sample. [Wikipedia] Hence:

$$\overline{X(t)} = E\{X(t)\} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)\, dt = \langle X(t) \rangle$$

Ensemble function **Time average**

$$K_{XX}(\tau) = E\{X(t) \cdot X(t+\tau)\} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)\, x^*(t-\tau)\, dt = \mathcal{K}_{XX}(\tau)$$

where $\langle X(t) \rangle$ is the time-average mean of the RP $X(t)$ and $\mathcal{K}_{XX}(\tau)$ is the time-average autocorrelation function.

"The ergodic hypothesis is that personal experience over time of a single individual reflects the current statistics of the general population."

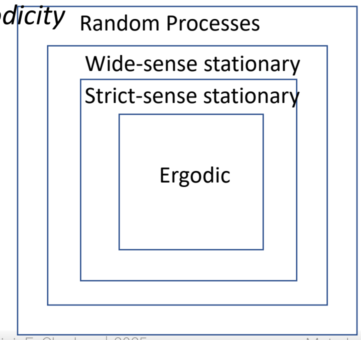
(I. Vardi)

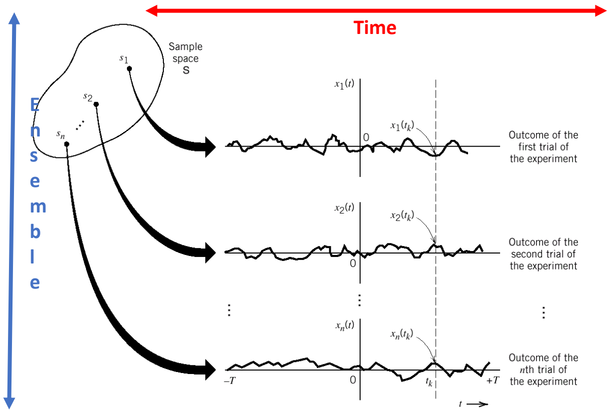
H. Bilgekul, Slides for the course "EE-461 Communication System II", EMU F. Farahmand, Slides for the course "CES 540 Digital Communication", Ch. 6, SSU

A very special class of Random Processes are the ergodic ones, whose statistical properties can be deduced from a single, sufficiently long, random sample. This basically allow us to replace the complex ensemble averages, which would need a very large number of trials, with a simpler time average over a sufficiently long period of time.

Take-home Messages/W9-2

- *Random Process:*
  - Definition, Ensemble vs. Time
  - CDF, PDF, Moments, Autocorrelation
  - Wide-sense & strict-sense stationary
  - *Ergodicity*





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

Outline

9.2.1 Law of Large Numbers

- Law of large numbers: describes the behavior of the sample mean of i.i.d. random variables as the sample size grows
- Assume i.i.d. $X_1, X_2, X_3, \ldots$ with finite mean μ and finite variance σ^2

NB: i.i.d. = independent and identically distributed Random Variables, have the same PDF and are all mutually independent

$$\text{Sample Mean:} \quad \overline{X_n} = \frac{X_1 + \cdots X_n}{n}$$

- $\overline{X_n}$ itself a random variable with

$$\text{Mean:} \quad E\{\overline{X_n}\} = \frac{1}{n}E\{X_1 + \cdots X_n\} = \frac{1}{n}(E\{X_1\} + \cdots + E\{X_n\}) = \mu$$

$$\text{Variance:} \quad \text{Var}\{\overline{X_n}\} = \frac{1}{n^2}Var\{X_1 + \cdots X_n\}$$

$$= \frac{1}{n^2}(Var\{X_1\} + \cdots + Var\{X_n\}) = \frac{\sigma^2}{n}$$

J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 10.2

Note the concept of i.i.d. random variables and how the sample mean, and its variance in particular, behave.

NB: the i.i.d. condition is probably a sufficient one, but not strictly necessary.

Q: which measurement examples and applications can you think of which exploit this property?

9.2.1 Law of Large Numbers (contd.)

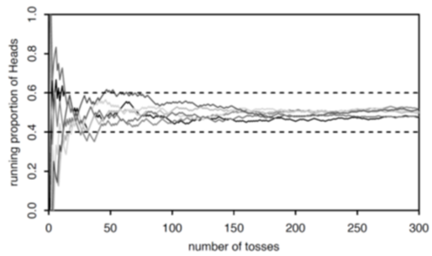
- Law of Large Numbers: as n grows, the sample mean $\overline{X_n}$ converges to the true mean μ
- Essential for simulations, statistics, etc. – implicitely used when we use:

1) the proportion of times that something happened as an approximation to its probability,

2) the average value in the replications of some quantity to approximate its theoretical average.

Example: *improvement in LiDAR ranging precision…*

…when accumulating timing measurements, as 1/n   **Q**



J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 10.2

LIDAR example (see also Section 8.2.9): calculate the precision of one single time (= distance) measurement starting from an estimate of the timestamping precision, e.g. 10-100 ps.

Q: what happens when averaging repetitive measurements?

9.2.2 Central Limit Theorem

- Law of Large Numbers: as n grows, the sample mean $\overline{X_n}$ converges to the true mean μ

But with which distribution? *Q*

Sum of a large number of i.i.d. random variables has an approximately Gaussian (normal distribution),

- regardless of the distribution of the individual RVs (could be anything!)
- very weak assumptions.

$$\text{As } n \to \infty, \qquad \sqrt{n}\left(\frac{\overline{X_n} - \mu}{\sigma}\right) \sim \mathcal{N}(0,1)$$

(i.e. the CDF of the l.h.s. approaches Φ, the CDF of the standard Gaussian distribution)

J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 10.3

NB: note that we are looking at the distribution of one specific random variable, the **sample mean $\overline{X_n}$**. If one simply adds up the histograms in slide 49, the final distribution will the similar to the starting one!

9.2.2 Central Limit Theorem (contd.)

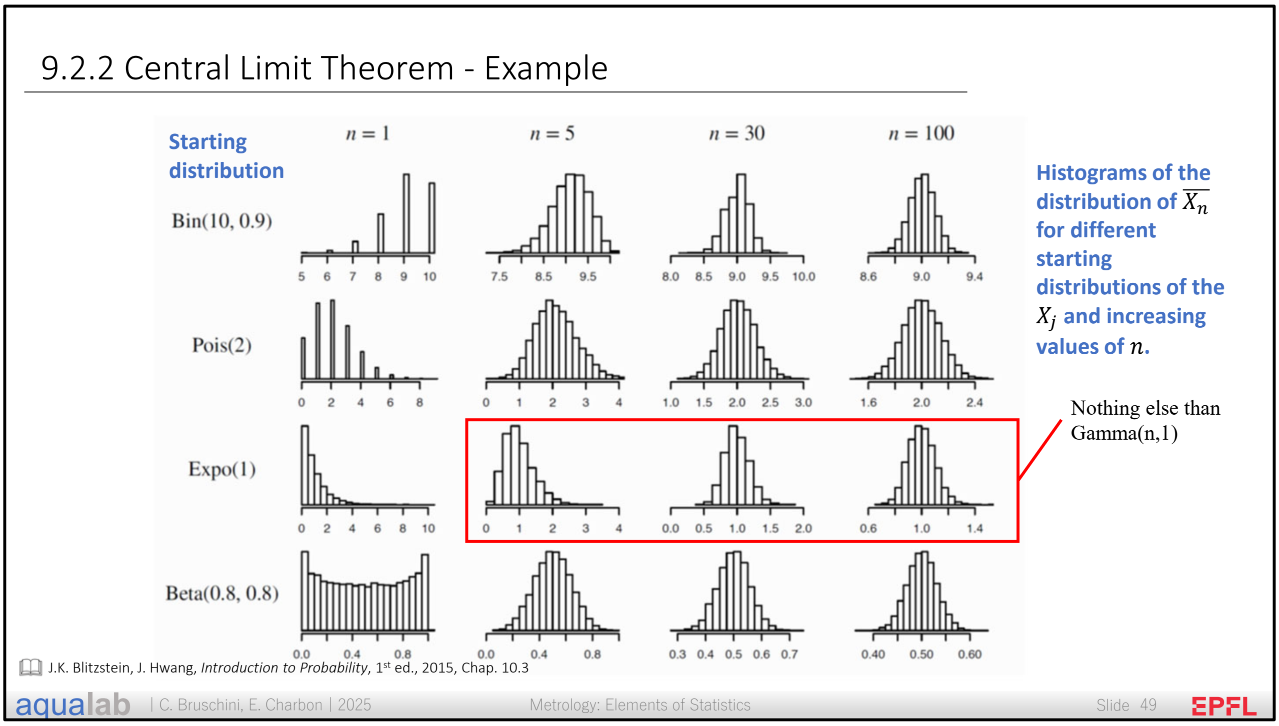
- In other words: start with independent RVs from almost any distribution, discrete or continuous,
  -> add them up
  -> distribution of the resulting RV has a Gaussian shape!
- The CLT is an asymptotic result. Approximation: for large n

$$\sqrt{n}\left(\frac{\overline{X_n} - \mu}{\sigma}\right) \rightarrow \mathcal{N}(0,1)$$

- NB: The distribution of the X_j is still relevant, e.g. if highly skewed or multimodal, n might need to be very large before the Gaussian approximation becomes accurate.

- Conversely, if the X_j are already i.i.d. Normals (Gaussian), the distribution of $\overline{X_n}$ is exactly $\mathcal{N}\left(\mu, {}^{\sigma^2}/_{n}\right)$ for all *n*.

J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 10.3



Blitzstein Figure 10.5: **Central limit theorem:**

These are histograms of the distribution of $\overline{X_n}$ for different starting distributions of the X_j (indicated by the rows) and increasing values of *n* (indicated by the columns).

Each histogram is based on 10,000 simulated values of $\overline{X_n}$. Regardless of the starting distribution of the X_j, the distribution of $\overline{X_n}$ approaches a Normal distribution as *n* grows.

NB: If one simply adds up the histograms, the final distribution will the similar to the starting one!

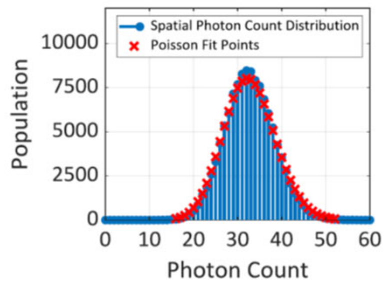
9.2.2 Central Limit Theorem - Example

- Poisson convergence to Gaussian: if

$$Y \sim Pois(n)$$

we can consider Y as a sum of n i.i.d. Pois(1) RVs.

$$\text{For large } n: Y \rightarrow \mathcal{N}(n, n)$$



J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 10.3

Outline

9.3 Elements of Estimation Theory

- Estimation theory has the purpose to solve one problem: given a set of data

$$\{x_1, x_2, \dots, x_{N-1}\}$$

which depends on an unknown parameter vector θ, determine an estimator

$$\hat{\theta} = g(x_1, x_2, \dots, x_{N-1})$$

where g is some function.

- In other words, how do we use collected data to estimate unknown parameters of a distribution?

J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 6.3

9.3 Elements of Estimation Theory (contd.)

- In general, if we assume that θ is deterministic, we will have a classical estimation problem. It can be solved in the following ways, and many more:

1. Least Squares Estimator (LSE)
2. Minimum Variance Unbiased Estimator (MVU)
3. Maximum Likelihood Estimator (MLE) *Ex*
4. Best Linear Unbiased Estimator (BLUE)
5. …

For this section, a general understanding is all right. You can get the essence also from some of the examples, like the BLUE estimator which is described later in the same section.

You might also need to be able to think a bit outside of the box and be able to reply to questions like “How can we for example estimate the lifetime of an exponential distribution from its samples” and the like.

9.3.1 Elements of Estimation Theory – Simple Mean Example

- Simple example: estimate the mean of a sample of i.i.d. RVs $X_1, X_2, X_3, \dots, X_n$

$$\text{Sample Mean: } \overline{X_n} = \frac{1}{n}\sum_{j=1}^{n} X_j$$

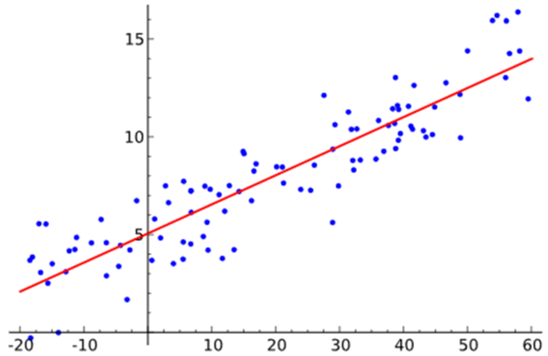
is an estimate of the *population mean* or *true mean,* $E\{X_j\}$ = the mean of the distribution from which the X_j were drawn.

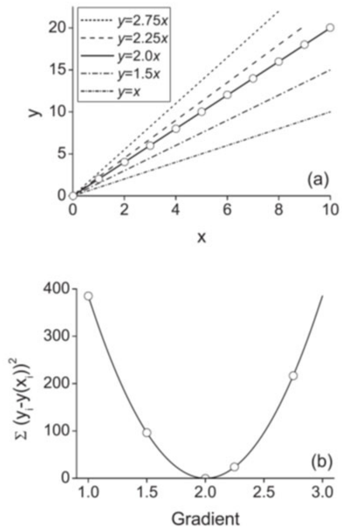
J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 6.3

Note the difference betwen the *true mean* and its *estimator* $\widehat{E\{X_j\}} = \overline{X_n}$.

9.3.2 Elements of Estimation Theory – LSE Example

Least Squares Estimator (LSE)





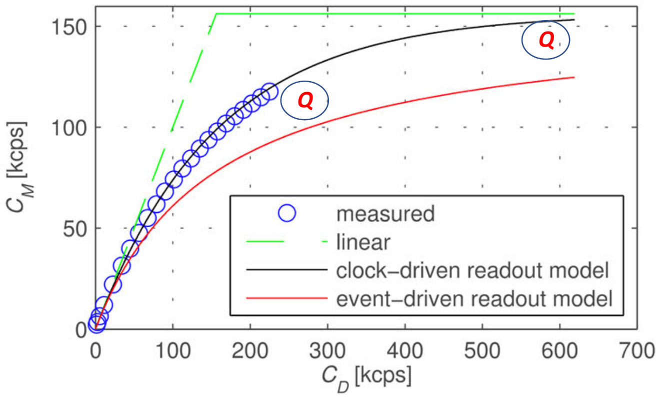
https://commons.wikimedia.org/wiki/File:Linear_regression.svg

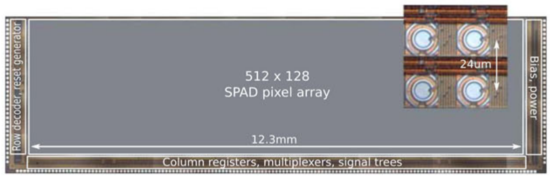
I.G. Hughes, T.P.A. Hase, *Measurements and their Uncertainties*, 1st ed., 2010, Chap. 6.3

A simple linear regression is shown on the left. On the top right, the data needs to be fit with a linear equation passing through the origin, i.e. characterised by a single variable, the slope shown in the bottom plot.

9.3.3 Elements of Estimation Theory – MLE Example

Maximum Likelihood Estimator (MLE) to correct for the exponential count loss in binary, clock-driven SPAD imagers





C_M: Measured count rate (externally)

C_D: Detected count rate (internally)

$$E[C_M] = \frac{1 - e^{-C_D \times T_{\text{readout}}}}{T_{\text{readout}}}$$

$$E[C_D] = \frac{-\ln(1 - C_M \times T_{\text{readout}})}{T_{\text{readout}}}$$

I. M. Antolovic et al, *Nonuniformity Analysis of a 65-kpixel CMOS SPAD Imager*, IEEE Trans. on Electron Devices 63, 2016

Example of a Maximum Likelihood Estimator:

The first SwissSPAD camera has 512x128 pixels, with an architecture quite similar to one of the more advanced SwissSPAD2 sensor.

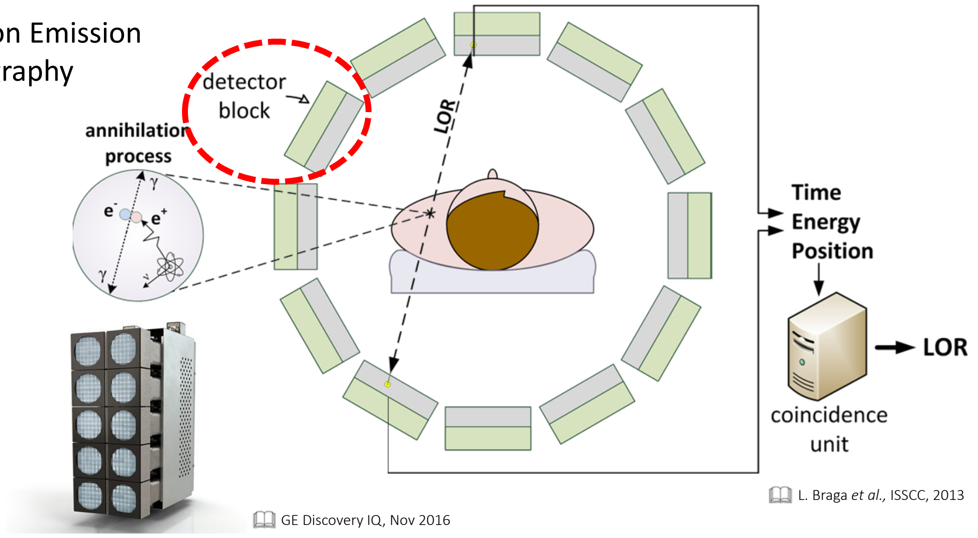
The camera is basically recording individual binary frames at very high speed. Given that the in-pixel memory is of one bit, it cannot differentiate when more than one photon was actually detected. At low photon counts this is not an issue, but as the light intensity, and thus the number of detections per frame C_D increases (horizontal axis), some photon detections C_M (vertical axis) are unavoidably lost. The response curve becomes logarithmic rather than binary.

Q: how can we estimate the true number of detected photons from the measured ones (basically inverting the curve shown above)? Which is the best estimator for the true number of detected photons?

→ It turns out that the best estimator for C_D is the maximum likelihood estimator shown in the bottom right equation.

9.3.4 Elements of Estimation Theory – BLUE Example

Positron Emission Tomography Basics



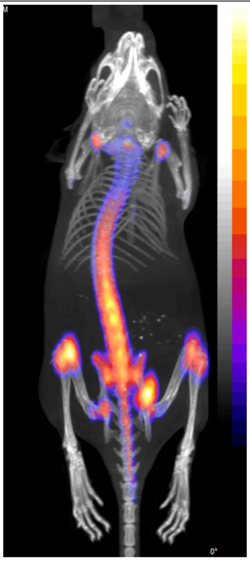
GE Discovery IQ, Nov 2016

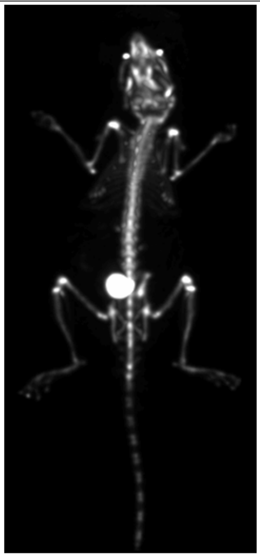
L. Braga *et al.*, ISSCC, 2013

Example of another estimator (BLUE) used the PET application detailed in Section 8.2.9.

9.3.4 Elements of Estimation Theory – BLUE Example

Positron Emission Tomography Reconstruction Example





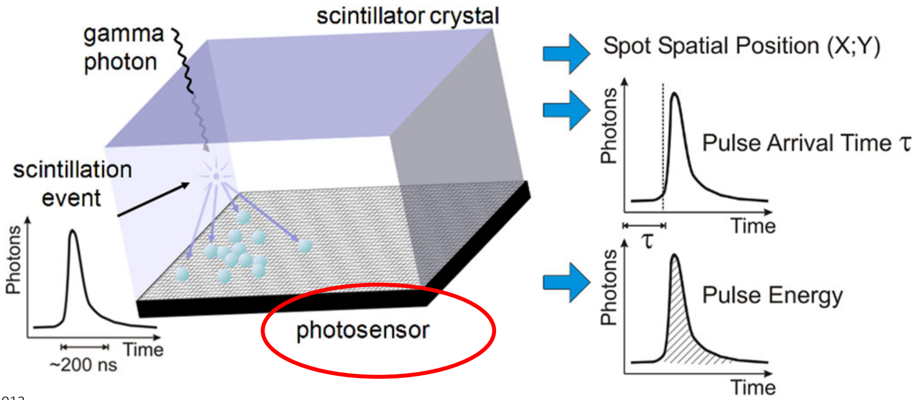
G. Nemeth, Mediso, Delft WS 2010

9.3.4 Elements of Estimation Theory – BLUE Example

Positron Emission Tomography Building Blocks & Main Variables

Problem: estimate the scintillation event time T_0 given a set of timing measurements t_q

Aim: obtain estimator with lowest variance (best timing precision)



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

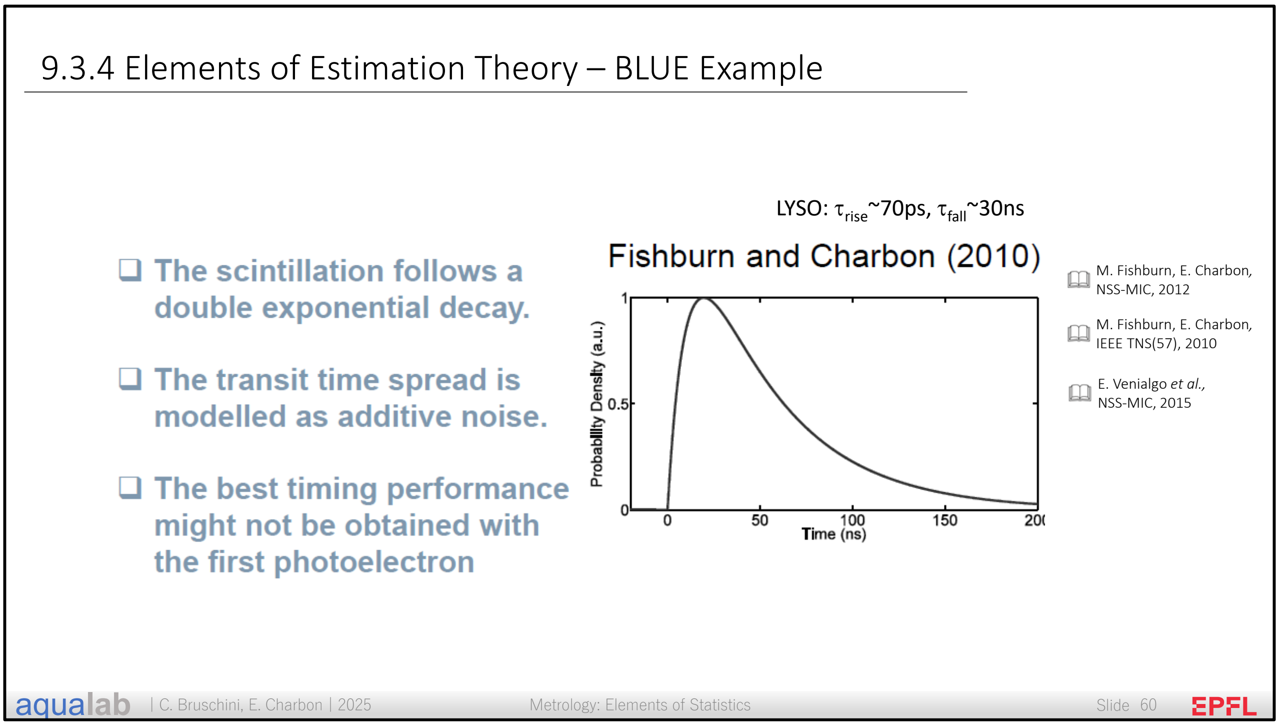
Typ some 10^4 photons/scintillation, few 10^3 detected

aqualab | C. Bruschini, E. Charbon | 2025 Metrology: Elements of Statistics Slide 59 EPFL

One of the key variables for PET is CTR (Coincidence Resolving Time), which is basically determined by using two modules in coincidence and plotting the gamma interaction time differences. It ultimately tells us how good the timing resolution is.

The key question here is: supposing that I am working in a digital approach, where my detector delivers ***a set of timing measurements*** t_q

→ which is the best ***estimate*** $\widehat{T_0}$ ***of the scintillation event time*** T_0***?***



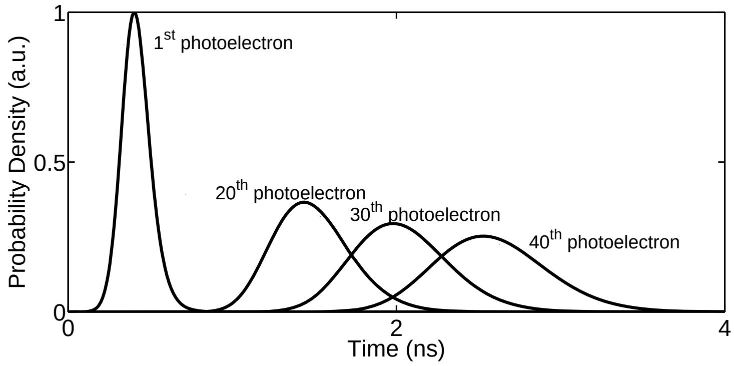
We can start with the PDF of the timing distribution of the scintillation photons (visible light), as illustrated above. In general the scintillation model is a double exponential, as already seen.

We might be tempted to say that the first detected photon is the best (lowest variance), but this is not necessarily true – see also the next slides. There is indeed the influence of the finite rise time in scintillators and of the transit time spread (TTS) in photomultiplier tubes (PMTs) (Gatti and Svelto 1966), plus the inevitable contribution of noise sources.

Quantitative example [S. Gundacker et al.]: Measurement results with LSO:Ce codoped 0.4%Ca scintillators with a (finite!) rise time of τ_{rise} ~ 70 ps, a fall time of τ_{fall} ~ 30:3 ns and a total number of scintillation photons produced n=20'400 per 511keV gamma. LSO scintillator gives a photon detection rate of typically 100 photoelectrons per nanosecond.

9.3.4 Elements of Estimation Theory – BLUE Example

Order Statistics



$$p_q(t) = \frac{R!}{(q-1)!(R-q)!}[1 - F(t)]^{(R-q)}[F(t)^{(q-1)}]f(t),$$

$f(t), F(t)$ = scintillation PDF/CDF (previous slide)

PDF of the q[th] photoelectron's time-of-registration.

Order statistics implies correlation between timestamps

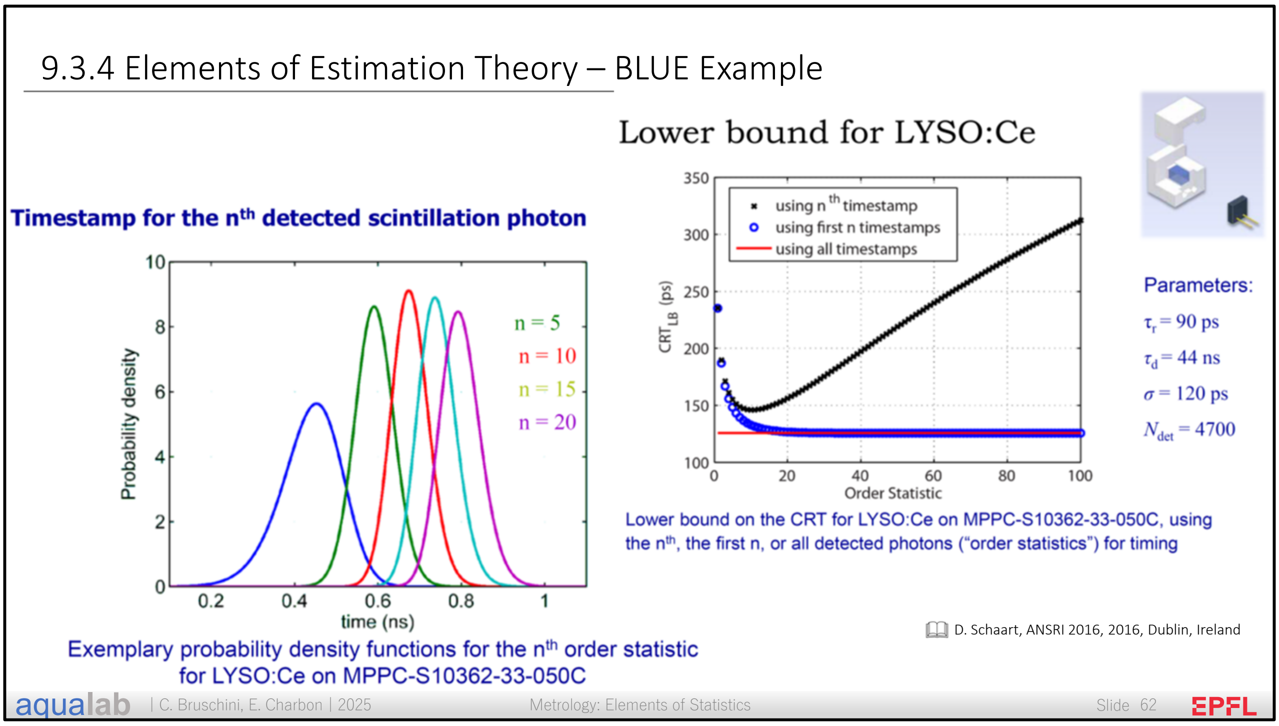
J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Theorem 8.6.4

E. Venialgo, E. Charbon *et al.*, PMB 2015

How can we calculate the PDF of the first, second, etc photons, to better understand why the first detected photon is not necessarily the best (lowest variance)? We can temporally reorder them…

→ Order Statistics: the PDF $p_q(t)$ is analytically given by the equation above. The variance of each detected photon does basically tell us how good it is at estimating the overall time of arrival, or scintillation event time, T_0.

NB: Seifert: The order statistics are NOT i.i.d. (the "initial" photons are).



Left: example of the PDF of the first, fifth, etc photons for a certain scintillator. The variance (see the FWHM) of the first one is clearly not the smallest!

Right: overall timing performance for different choices of timing estimators. In this case, using the *n-th* timestamp (e.g. the first, second, or subsequent), does clearly not deliver the best (i.e. lowest) result, as obtained when using multiple timestamps.

9.3.4 Elements of Estimation Theory – BLUE Example

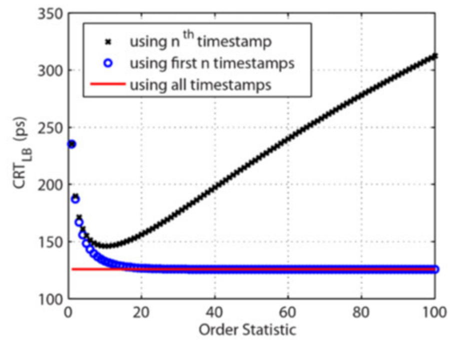
$$\hat{T}_0^{(p)} = \sum_{q=1}^{Q} t_q w_q^{(p)},\ p = 1, 2, 3.$$ ← General estimator

(p is one of 3 possible estimators)

Q

A simple estimator approach: $w_q^{(1)} = \frac{1}{Q},\ q = 1, \dots, Q$ ← Simple mean coeffs.

($p = 1$ estimator)



D. Schaart, ANSRI 2016, 2016, Dublin, Ireland

E. Venialgo, E. Charbon *et al.*, PMB 2015

The first equation shows the general expression for the timestamp estimator $\widehat{T_0}$. A simple one could just be the mean of the measured timestamps!

Image source: derived from Fig. 3 in Venialgo 2015 (refers to EndoTOF case with 48 TDCs).

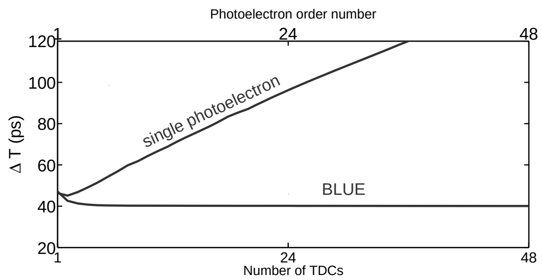
9.3.4 Elements of Estimation Theory – BLUE Example

- Assuming large number of measurements
- Other Estimator Approaches: **B**est **L**inear **U**nbiased **E**stimator (BLUE)

$$\hat{T}_0^{(p)} = \sum^{Q} t_q w_q^{(p)}, \; p = 1, 2, 3.$$ ⟵ General estimator

$$\boldsymbol{w}_q^{(3)} = \frac{\boldsymbol{C}^{-1}\boldsymbol{d}}{\|\boldsymbol{C}^{-1/2}\boldsymbol{d}\|_2^2},$$ ⟵ BLUE coeffs. (correlation matrix)

($p = 3$ estimator)



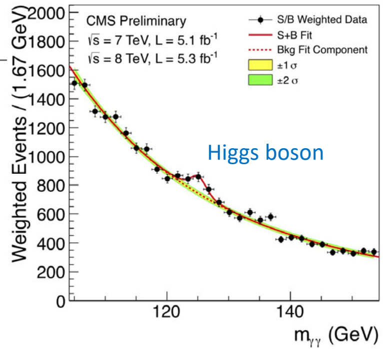
E. Venialgo, E. Charbon *et al.*, PMB 2015

A more complex estimator would be the BLUE one. The analytical formula for the weights w_q is obtained from the scintillation model and the order statistics, the covariance matrix $\boldsymbol{C}$ is then obtained from experimental measurements.

NB: ***d*** **=** column vector filled with ones and with a length equal to the number of utilized timestamps.

Outline

9.4 Error Analysis

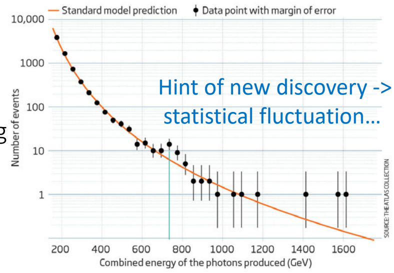


- The aim of error analysis is to quantify and record the errors associated with the inevitable spread in a set of measurements.

- Confidence boundaries represent the quality of the approximation given by the uncertainty.

  Example: the six-sigma method, 5 sigma limit (CERN) (Q)

- Uncertainties can be associated to random errors (hence influencing the variance of the measurement distribution) or to systematic errors (acting on the mean value of the measurement distribution).



https://universe-review.ca/R15-20-accelerators03.htm

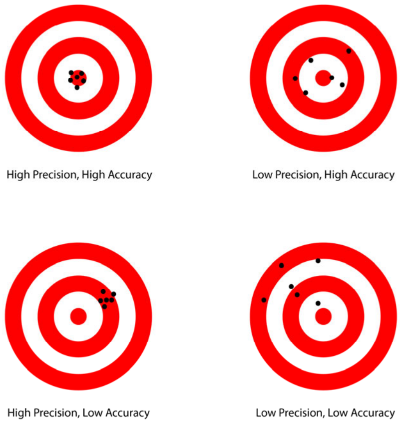
I.G. Hughes, T.P.A. Hase, *Measurements and their Uncertainties*, 1st ed., 2010, Chap. 1

NB: different communities have different conventions → have a look at some of the error bars used in the biosciences…

9.4.1 Accuracy

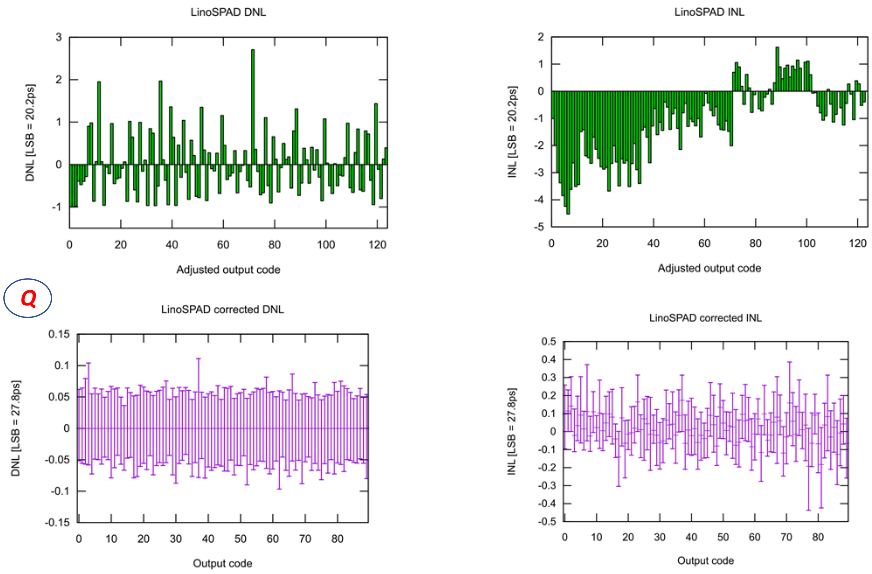
Accuracy -> mean

- The accuracy of a measurement gives a notion of the mean value of the set of measurements distribution with respect to the real value.
- An accurate measurements distribution will hence have a very small systematic error, but could be affected by a large spread in the data (high variance).
- Accuracy can be enhanced in the experimental real life by means of calibration techniques.



I.G. Hughes, T.P.A. Hase, *Measurements and their Uncertainties*, 1st ed., 2010, Chap. 1

9.4.1 Accuracy – Example

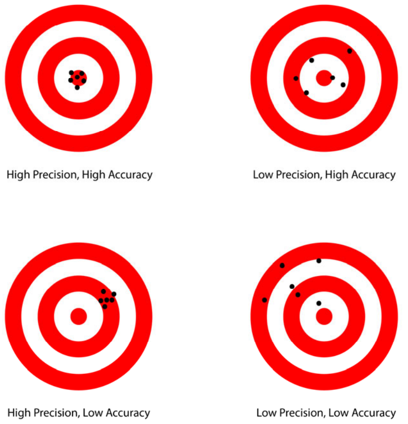


Calibration of a Time-to-Digital converter
S. Burri, EPFL, MDPI Instruments, 2018

9.4.2 Precision

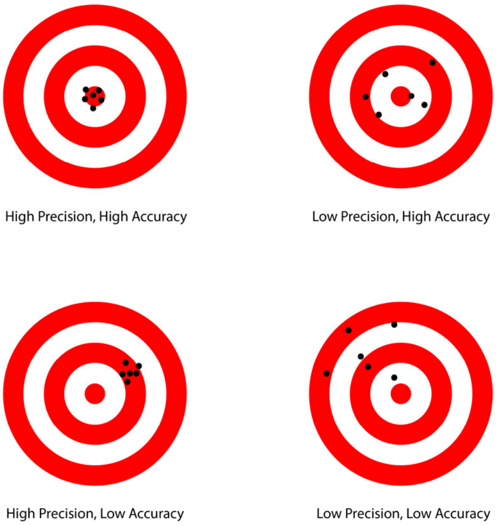
Precision -> spread (variance)

- The precision of a measurement gives information about the spread of the measured set of data collected by the measurement.
- A precise measurement distribution will have a low dispersion of data (hence a small variance), but it might have a mean value very distant from the real one.
- In order to enhance precision, the most simple way is to increase the size of the sample data. In fact, as shown previously, for experimental data the variance decreases linearly with the number of samples collected.



I.G. Hughes, T.P.A. Hase, *Measurements and their Uncertainties*, 1st ed., 2010, Chap. 1

9.4.2 Accuracy vs. Precision



9.4.3 Resolution

- The resolution of a measurement is the smallest change in the underlying physical quantity that produces a response in the measurement. [Wikipedia]

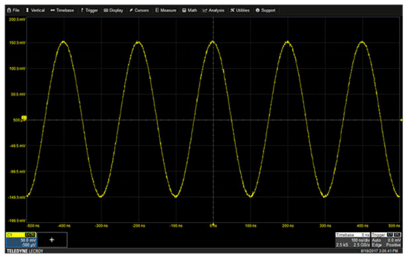
- In case of an ADC (analog-to-digital converter), the resolution is given by one bit.

Example: for an oscilloscope with an 8 bits ADC, set at 100 mV/div (i.e. for a total screen width of 800 mV), the resolution of each point collected is given by:

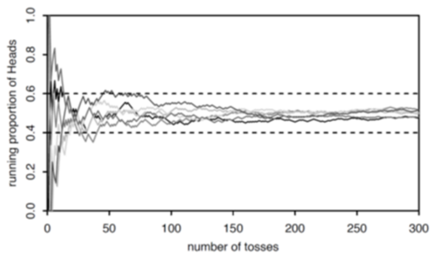
$$8 \text{ bits} = 2^8 \text{ different values} \rightarrow Res = \frac{800}{256}\, mV = 3.125\, mV$$

High resolution 10 Downing street.

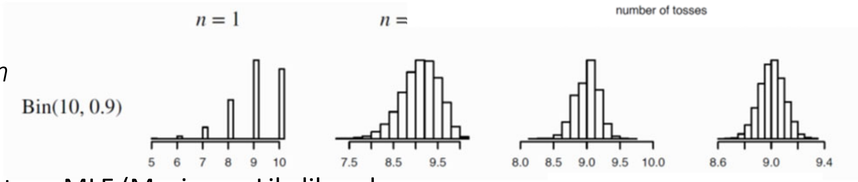
Low resolution 10 Downing St.



Take-home Messages/W9-3



- *Law of Large Numbers:*
  - Concept of i.i.d. random variables
  - Mean and Variance
- *Central Limit Theorem*



- *Estimation Theory:*
  - Examples of estimators, MLE (Maximum Likelihood Estimator)
  - Example: Positron Emission Tomography ↔ different time-of-arrival estimators
- *Precision, Accuracy, Resolution*

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

Appendix A: Gamma Distribution – Gamma Function

- While the Exponential distribution represents the wait time before the first success under the conditions of memorylessness, the gamma distribution represents the *total waiting time for multiple successes* (hence it is the sum of multiple exponential distributions).

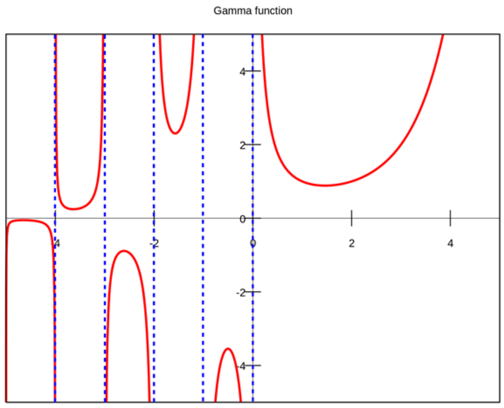
- We first define the gamma function as:

$$\Gamma(a) = \int_0^\infty x^a e^{-x} \frac{dx}{x}, \qquad a > 0$$

- The gamma function has the following properties:

$$\Gamma(a+1) = a\,\Gamma(a)$$

$$\Gamma(n) = (n-1)!$$



J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 8.4

Appendix A: Gamma Distribution (contd.)

- Then, we say that X has a gamma distribution (we will write $X \sim \text{Gamma}(a, 1)$) if:

$$\text{PDF:} \qquad f_X(x) = \frac{1}{\Gamma(a)}\, x^a e^{-x} \frac{1}{x}, \qquad x > 0$$

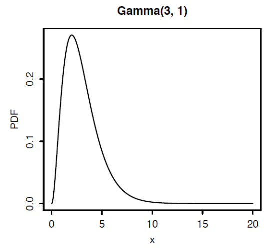
- From the gamma distribution of $X \sim \text{Gamma}(a, 1)$, we get, for $\lambda > 0$, the more general $Y = X/\lambda \sim \text{Gamma}(a, \lambda)$:

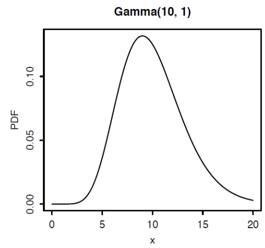
$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = \frac{1}{\Gamma(a)}(\lambda y)^a e^{-\lambda y} \frac{1}{\lambda y} \lambda$$

hence

$$\text{PDF:} \qquad f_Y(y) = \frac{1}{\Gamma(a)}(\lambda y)^a e^{-\lambda y} \frac{1}{y}, \qquad y > 0$$

Gamma$(a, 1)$





J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 8.4

Appendix A: Gamma Distribution (contd.)

- From the PDF of the gamma distribution just obtained $Y \sim \text{Gamma}(a, \lambda)$, it can be shown that the Gamma is nothing else but the distribution obtained by summing up a independent exponential distributions. In fact, for $a = 1$:

$$\text{PDF:} \quad f_Y(y) = \frac{1}{\Gamma(a)} (\lambda y)^a e^{-\lambda y} \frac{1}{y}, \qquad y > 0$$

reduces to

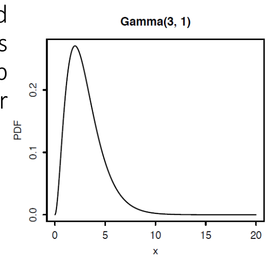
$$\text{PDF:} \quad f_Y(y) = \lambda y \, e^{-\lambda y} \frac{1}{y} = \lambda \, e^{-\lambda y}, \qquad y > 0$$

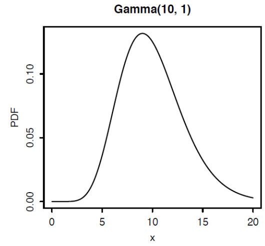
which is the exponential distribution.

- Follows that, let $X_1, X_2, \dots, X_n$ be n i.i.d. $\text{Expo}(\lambda)$. Then:

$$Y = X_1 + \dots + X_n \sim \text{Gamma}(n, \lambda)$$

Gamma$(a, 1)$





J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 8.4

Appendix A: Gamma Distribution (contd.)

- For a $X \sim \text{Gamma}(a, 1)$, it follows:

Mean: $$E\{X\} = \int_0^\infty \frac{1}{\Gamma(a)} x^{a+1} e^{-x} \frac{dx}{x} = \frac{\Gamma(a+1)}{\Gamma(a)} = a$$

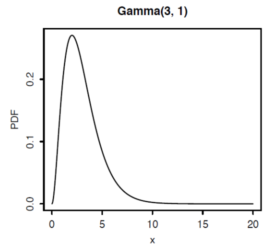
Second Moment: $$E\{X^2\} = \int_0^\infty \frac{1}{\Gamma(a)} x^{a+2} e^{-x} \frac{dx}{x} =$$

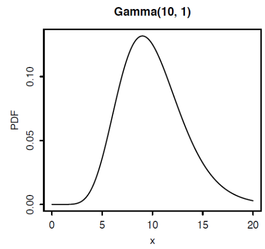
$$= \frac{\Gamma(a+2)}{\Gamma(a)} = a(a+1)$$

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

$$= a(a+1) - a^2 = a$$

$\text{Gamma}(a, 1)$

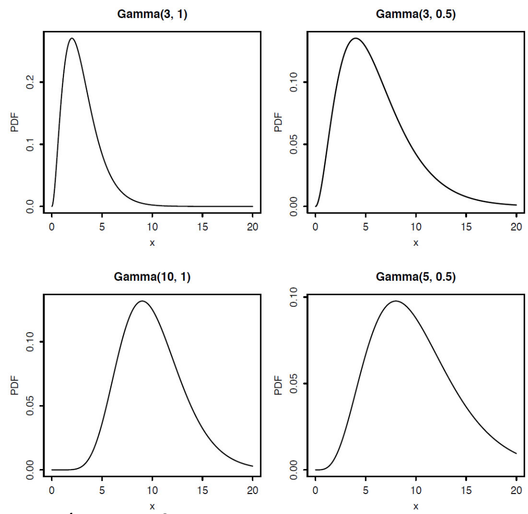




J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 8.4

Appendix A: Gamma Distribution (contd.)

Gamma(a,λ)



- For the more general gamma distribution $Y = X/\lambda \sim \text{Gamma}(a,\lambda)$, by simple transformation, we obtain:

$$\text{Mean:} \quad E\{Y\} = \frac{1}{\lambda} E\{X\} = \frac{a}{\lambda}$$

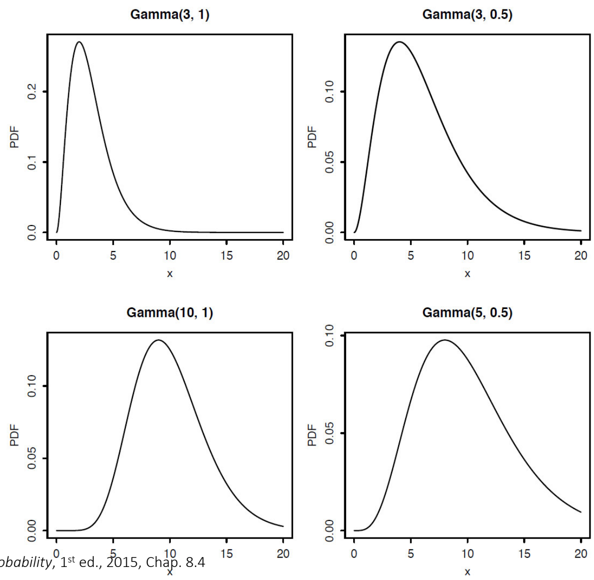
$$\text{Second Moment:} \quad E\{Y^2\} = \frac{1}{\lambda^2} E\{X\} = \frac{a\,(a+1)}{\lambda^2}$$

$$\text{Variance:} \quad Var\{Y\} = \frac{1}{\lambda^2} Var\{X\} = \frac{a}{\lambda^2}$$

-> calculate mean/variance for some examples

J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 8.4

Appendix A: Gamma Distribution (contd.)



Gamma(a,λ)

-> calculate mean/variance for some examples

Mean: $\frac{a}{\lambda}$

Variance: $\frac{a}{\lambda^2}$

J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 8.4

Appendix B: Poisson Process

- Definition: a sequence of arrivals in continuous time with rate λ is a (1D) Poisson process with rate λ if the following two conditions hold:

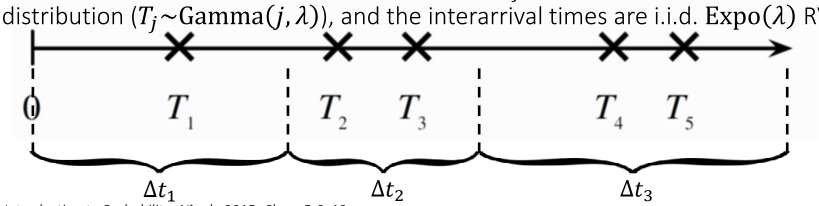
  1) The number of arrivals that occur in an interval of length t is a $\text{Pois}(\lambda t)$ RV.

  2) The numbers of arrivals that occur in disjoint intervals – e.g. (0,10), [10,12) and [15,∞) – are independent of each other.

- If T_j is the time of the j-th arrival, $N(t)$ is the number of events up to the time t, follows:

$$P\{T_1 > t\} = P\{N(t) = 0\} = e^{-\lambda t}$$

so T_1 has an Exponential distribution ($T_1 \sim \text{Expo}(\lambda)$), hence T_j, being the sum of j i.i.d. exponentials, is a Gamma distribution ($T_j \sim \text{Gamma}(j, \lambda)$), and the interarrival times are i.i.d. $\text{Expo}(\lambda)$ RVs.



J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 5.6, 13

Appendix B: Poisson Process

- NB: i.i.d. = independent and identically distributed Random Variables, have the same PDF and are all mutually independent

- *"Confirmation" that the Exponential distribution is closely connected to the Poisson distribution!*
- Examples of Poisson processes:
  - 1D: cars passing by a highway checkpoint;
  - 2D: flowers in a meadow;
  - 3D: stars in a region of the galaxy."

  Dark Counts and "real" detections in a SPAD sensor

J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 13

Appendix B: Poisson Process

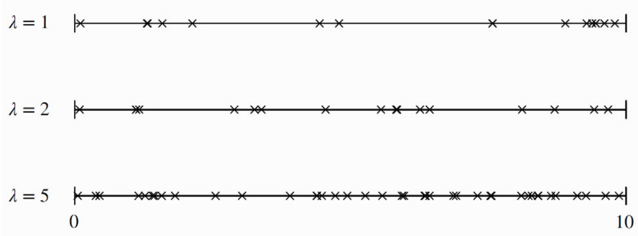
- Timeline: $(0, +\infty)$ but it could also be $(-\infty, +\infty)$

- To generate n arrivals from a Poisson process with rate λ:

  1. Generate n i.i.d. $\text{Expo}(\lambda)$ RVs: $X_1, X_2, \ldots, X_n$
  2. For $j = 1, 2, \ldots, n$ set $T_j = X_1 + \cdots + X_j$

- Then we can take the $T_1, \ldots, T_n$ to be the arrival times.

Simulate Poisson Processes in 1D



Note: interarrival times are i.i.d., but the arrivals are not evenly spaced -> there is a lot of variability in the interarrival times, which produces *Poisson clumping*

J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 13

Appendix B: Poisson Process

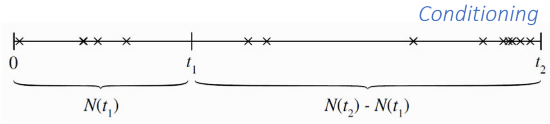
- A Poisson Process has the following three properties:

  1. *Conditioning*: let $\{N(t), t > 0\}$ be a Poisson Process with rate λ and $t_2 > t_1$. Then the conditional distribution stands:

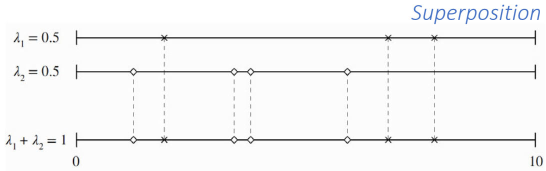
$$N(t_1)|N(t_2) = n \sim \text{Bin}\left(n, \frac{t_1}{t_2}\right)$$

  2. *Superposition*: let $\{N_1(t), t > 0\}$ and $\{N_2(t), t > 0\}$ be two independent Poisson Processes with rates λ_1 and λ_2. Then the combined process $N(t) = N_1(t) + N_2(t)$ is a Poisson process with rate $\lambda_1 + \lambda_2$.

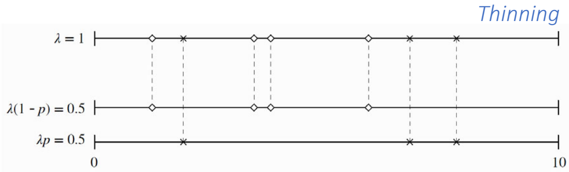
Conditioning



Superposition



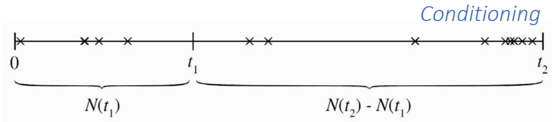
Thinning

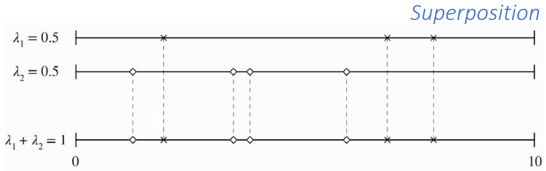


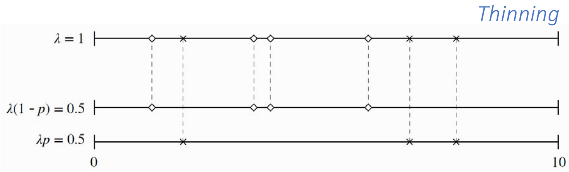
J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 13

Appendix B: Poisson Process

- A Poisson Process has the following three properties:







3. *Thinning*: let $\{N(t), t > 0\}$ be a Poisson Process with rate λ, and classify each event at the arrival as either type-1 events (with probability p) or type-2 events (with probability $1 - p$), independently. Then the type-1 events form a Poisson process with rate λp, the type-2 events form a Poisson process with rate $\lambda(1 - p)$ and they are independent.

J.K. Blitzstein, J. Hwang, *Introduction to Probability*, 1st ed., 2015, Chap. 13