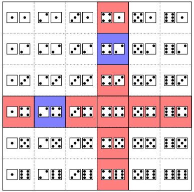
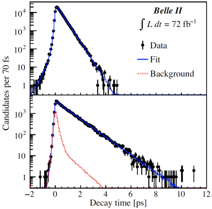
STATISTICS AND SYSTEMATICS FOR PARTICLE PHYSICS: An introduction

Bruce Yabsley

School of Physics, University of Sydney

CPPC Meeting, U.Sydney, 9th February 2024





Outline

1. Statistics and Likelihood
2. Probability, Conditional Probability, and Likelihood
3. Maximum Likelihood Estimates and fitting
4. Statistical and systematic uncertainties
5. Frequentist & Bayesian Probability, and coin-tossing

Statistics and Likelihood

What is statistics?

A deliberately vague and heuristic definition (I did not look up a reference): if you are doing some combination of

What is statistics?

A deliberately vague and heuristic definition (I did not look up a reference): if you are doing some combination of

- extracting a single number from several numbers

What is statistics?

A deliberately vague and heuristic definition (I did not look up a reference): if you are doing some combination of

- extracting a single number from several numbers
- handling numbers with uncertainties

What is statistics?

A deliberately vague and heuristic definition (I did not look up a reference): if you are doing some combination of

- extracting a single number from several numbers
- handling numbers with uncertainties
- making inferences from data with gaps in it

What is statistics?

A deliberately vague and heuristic definition (I did not look up a reference): if you are doing some combination of

- extracting a single number from several numbers
- handling numbers with uncertainties
- making inferences from data with gaps in it

then you are probably doing statistics. For example:

NO you have the (x, y, z) of a point, and you want to know the (r, θ, ϕ)

What is statistics?

A deliberately vague and heuristic definition (I did not look up a reference):
if you are doing some combination of

- extracting a single number from several numbers
- handling numbers with uncertainties
- making inferences from data with gaps in it

then you are probably doing statistics. For example:

NO you have the (x, y, z) of a point, and you want to know the (r, θ, ϕ)

YES wanting to measure a constant rate, you make ten measurements $\{10, 7, 2, 3, 4, 8, 9, 2, 5, 3\}$; averaging, you get 5.3

What is statistics?

A deliberately vague and heuristic definition (I did not look up a reference): if you are doing some combination of

- extracting a single number from several numbers
- handling numbers with uncertainties
- making inferences from data with gaps in it

then you are probably doing statistics. For example:

NO you have the (x, y, z) of a point, and you want to know the (r, θ, ϕ)

YES wanting to measure a constant rate, you make ten measurements $\{10, 7, 2, 3, 4, 8, 9, 2, 5, 3\}$; averaging, you get 5.3

NO you want to know the force on a moving charge in a magnetic field, and calculate $\vec{F} = q\vec{v} \times \vec{B}$

What is statistics?

A deliberately vague and heuristic definition (I did not look up a reference): if you are doing some combination of

- extracting a single number from several numbers
- handling numbers with uncertainties
- making inferences from data with gaps in it

then you are probably doing statistics. For example:

NO you have the (x, y, z) of a point, and you want to know the (r, θ, ϕ)

YES wanting to measure a constant rate, you make ten measurements $\{10, 7, 2, 3, 4, 8, 9, 2, 5, 3\}$; averaging, you get 5.3

NO you want to know the force on a moving charge in a magnetic field, and calculate $\vec{F} = q\vec{v} \times \vec{B}$

YES you have three measurements of some quantity, with uncertainties: 10 ± 3.4, 9 ± 1.2, 14 ± 2.4; you take a weighted average and find 10

What is statistics?

A deliberately vague and heuristic definition (I did not look up a reference): if you are doing some combination of

- extracting a single number from several numbers
- handling numbers with uncertainties
- making inferences from data with gaps in it

then you are probably doing statistics. For example:

NO you have the (x, y, z) of a point, and you want to know the (r, θ, ϕ)

YES wanting to measure a constant rate, you make ten measurements $\{10, 7, 2, 3, 4, 8, 9, 2, 5, 3\}$; averaging, you get 5.3

NO you want to know the force on a moving charge in a magnetic field, and calculate $\vec{F} = q\vec{v} \times \vec{B}$

YES you have three measurements of some quantity, with uncertainties: 10 ± 3.4, 9 ± 1.2, 14 ± 2.4; you take a weighted average and find 10

YES you *don't see* any events of a certain process: what can you say about its rate?

What is statistics?

It seems that this all has something to do with

- some underlying state of affairs in the world
- models of that state of affairs
- observations
- "uncertainties", "probabilities", and "chance"
- inference

and so we are also in the domain of prediction:

What is statistics?

It seems that this all has something to do with

- some underlying state of affairs in the world
- models of that state of affairs
- observations
- “uncertainties”, “probabilities”, and “chance”
- inference

and so we are also in the domain of prediction:

- will it rain tomorrow?

What is statistics?

It seems that this all has something to do with

- some underlying state of affairs in the world
- models of that state of affairs
- observations
- "uncertainties", "probabilities", and "chance"
- inference

and so we are also in the domain of prediction:

- will it rain tomorrow?
- what is the *chance* that it will rain tomorrow?

What is statistics?

It seems that this all has something to do with

- some underlying state of affairs in the world
- models of that state of affairs
- observations
- "uncertainties", "probabilities", and "chance"
- inference

and so we are also in the domain of prediction:

- will it rain tomorrow?
- what is the *chance* that it will rain tomorrow?

This seems to be crying out for either some clear definitions, and/or some clear recipes.

What is statistics?

It seems that this all has something to do with

- some underlying state of affairs in the world
- models of that state of affairs
- observations
- "uncertainties", "probabilities", and "chance"
- inference

and so we are also in the domain of prediction:

- will it rain tomorrow?
- what is the *chance* that it will rain tomorrow?

This seems to be crying out for either some clear definitions, and/or some clear recipes.

We will get there. First, a word about "likelihood":

Likelihood: A simple Poisson process

- Likelihood is the probability of the *data* ($\boldsymbol{x}$), given the model ($\boldsymbol{\theta}$).
- We write it, and think of it, as a f^n of the model: $\mathcal{L}(\boldsymbol{\theta};\, \boldsymbol{x}) = P(\boldsymbol{x};\, \boldsymbol{\theta})$
- For a discrete case,

$$\mathcal{L} = \prod_i P(x_i;\, \boldsymbol{\theta})$$

$$-2\ln\mathcal{L} = -2\sum_i \ln P(x_i;\, \boldsymbol{\theta})$$

- Poisson (counts of rate-governed indep^t cases): $P_p(n_i;\, \mu) = \frac{e^{-\mu}\mu^{n_i}}{n_i!}$

---

1. $n_i = 10;\ \mathcal{L} = P_p(10;\, \mu)$
2. $n_i = \ \ 7;\ \mathcal{L} = P_p(10;\, \mu) \times P_p(7;\, \mu)$
3. $n_i = \ \ 2;\ \mathcal{L} = P_p(10;\, \mu) \times P_p(7;\, \mu) \times P_p(2;\, \mu)$
4. $n_i = \ \ 3;\ \mathcal{L} = P_p(10;\, \mu) \times P_p(7;\, \mu) \times P_p(2;\, \mu) \times P_p(3;\, \mu)$

...

10. $n_i = \ \ 3;\ \mathcal{L} = P_p(10;\, \mu) \times P_p(7;\, \mu) \times P_p(2;\, \mu) \times P_p(3;\, \mu) \times P_p(4;\, \mu)$
$\times P_p(8;\, \mu) \times P_p(9;\, \mu) \times P_p(2;\, \mu) \times P_p(5;\, \mu) \times P_p(3;\, \mu)$

$\rightarrow$ *maximum likelihood estimate* of rate $\mu = 5.3$; note this = the *average* of n_i

Likelihood: A Gaussian process $\mathcal{G}(x;\, \mu,\, \sigma_i)$

- suppose some underlying parameter μ (signal strength, mass, ...)
- suppose a measurement procedure that returns values x_i:
  - each is an unbiased estimate of μ
  - each comes with a Gaussian uncertainty σ_i (for now, never mind how this is determined)
  - suppose the σ_i are also reliably estimated
- likelihood for a given measurement is $\mathcal{L}(\mu;\, x_i) = \mathcal{G}(x_i;\, \mu,\, \sigma_i)$

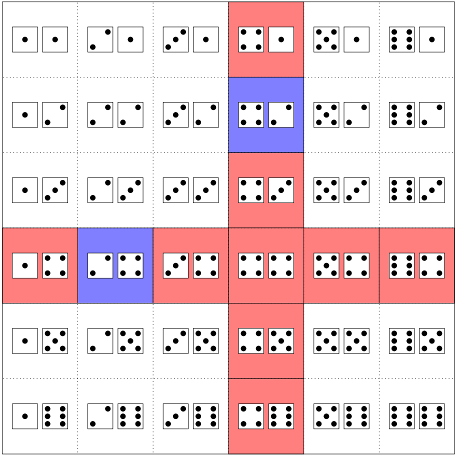
---

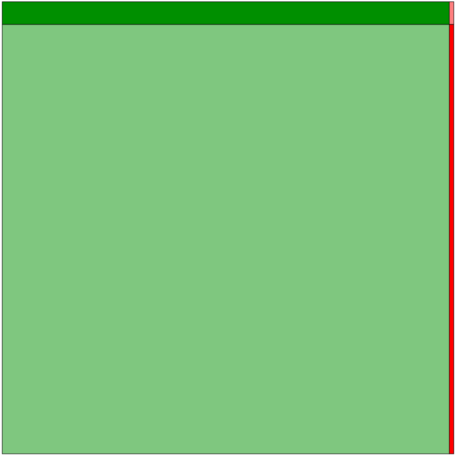
1. 10 ± 3.4: $\mathcal{L} = \exp\left(-\frac{(\mu-10)^2}{2(3.4)^2}\right)$
2. 9 ± 1.2: $\mathcal{L} = \exp\left(-\frac{(\mu-10)^2}{2(3.4)^2}\right) \times \exp\left(-\frac{(\mu-9)^2}{2(1.2)^2}\right)$
3. 14 ± 2.4: $\mathcal{L} = \exp\left(-\frac{(\mu-10)^2}{2(3.4)^2}\right) \times \exp\left(-\frac{(\mu-9)^2}{2(1.2)^2}\right) \times \exp\left(-\frac{(\mu-14)^2}{2(2.4)^2}\right)$

$$-2\ln\mathcal{L} = \frac{(\mu-10)^2}{(3.4)^2} + \frac{(\mu-9)^2}{(1.2)^2} + \frac{(\mu-14)^2}{(2.4)^2}$$

→ *maximum likelihood estimate* of $\mu = 10$; note this = the *weighted average* of x_i

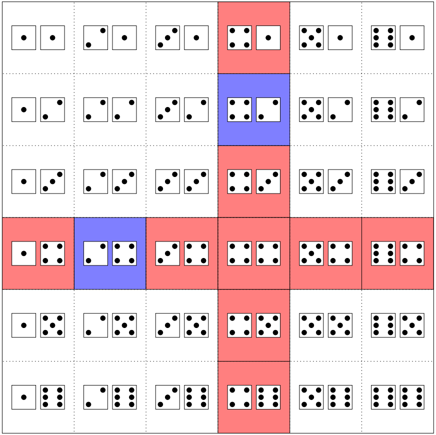
Probability, Conditional Probability, and Likelihood





Probability, and Conditional Probability

If we consider the probability of discrete events (die 1, die 2), we can also define the probability of derived cases (*e.g.* "six in total"), and then the idea of conditional probability ...



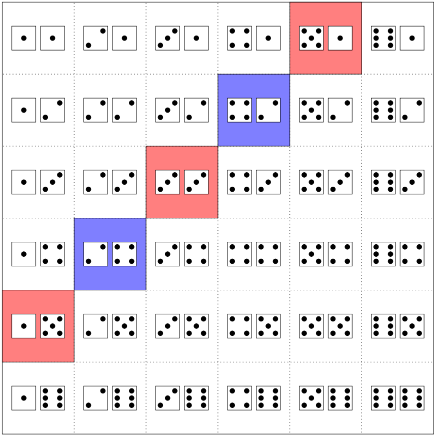
The probability that an elementary event known to belong to set B also belongs to set A; defined *via*

$$P(A \text{ and } B) = P(A \,|\, B) \cdot P(B)$$
$$= P(B \,|\, A) \cdot P(A)$$

$$P(6 \text{ and } \text{“} \geq \text{ one 4”}) = 2/36$$
$$P(\text{“} \geq \text{ one 4”}) = 11/36$$
$$P(6 \,|\, \text{“} \geq \text{ one four”}) = 2/11$$

Probability, and Conditional Probability

If we consider the probability of discrete events (die 1, die 2), we can also define the probability of derived cases (*e.g.* "six in total"), and then the idea of conditional probability ...



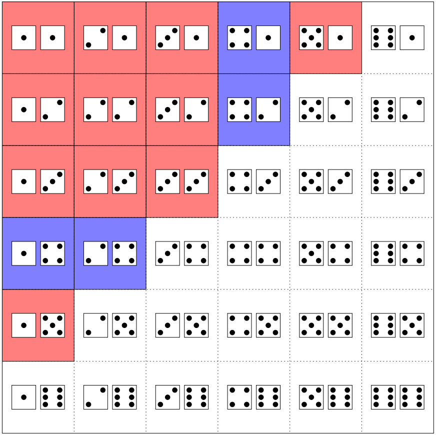
The probability that an elementary event known to belong to set B also belongs to set A; defined *via*

$$P(A \text{ and } B) = P(A \,|\, B) \cdot P(B)$$
$$= P(B \,|\, A) \cdot P(A)$$

$$P(\text{“}\geq \text{ one 4”} \text{ and } 6) = 2/36$$
$$P(6) = 5/36$$
$$P(\text{“}\geq \text{ one four”} \,|\, 6) = 2/5$$
$$\neq 2/11$$

Probability, and Conditional Probability

If we consider the probability of discrete events (die 1, die 2),
we can also define the probability of derived cases (*e.g.* “six in total”),
and then the idea of conditional probability … and Bayes' Theorem



$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

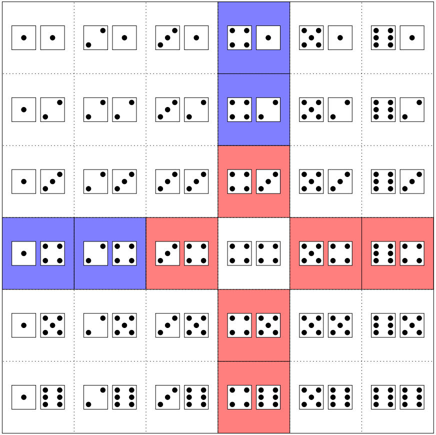
$$P(\text{“one 4”} \mid \text{“}\leq 6\text{”}) = 4/15$$
$$P(\text{“}\leq 6\text{”} \mid \text{“one 4”}) = 4/10$$
$$P(\text{“one 4”}) = 10/36$$
$$P(\text{“}\leq 6\text{”}) = 15/36$$

in general, $P(A) \neq P(B)$

so $P(A \mid B) \neq P(B \mid A)$

Probability, and Conditional Probability

If we consider the probability of discrete events (die 1, die 2),
we can also define the probability of derived cases (*e.g.* "six in total"),
and then the idea of conditional probability . . . and Bayes' Theorem



$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

$$P(\text{“one 4”} \mid \text{“}\leq 6\text{”}) = 4/15$$
$$P(\text{“}\leq 6\text{”} \mid \text{“one 4”}) = 4/10$$
$$P(\text{“one 4”}) = 10/36$$
$$P(\text{“}\leq 6\text{”}) = 15/36$$

in general, $P(A) \neq P(B)$

so $P(A \mid B) \neq P(B \mid A)$

Likelihood, and Conditional Probability

Likelihood, and Conditional Probability

- characteristic PP problem is *particle identification*:
  given detector response to a track, what type of particle is it?

Likelihood, and Conditional Probability

- characteristic PP problem is *particle identification*:
  given detector response to a track, what type of particle is it?
- often based on $\mathcal{L}(\text{type; response}) \equiv P(\text{response; type})$

Likelihood, and Conditional Probability

- characteristic PP problem is *particle identification*:
  given detector response to a track, what type of particle is it?
- often based on $\mathcal{L}(\text{type};\ \text{response}) \equiv P(\text{response};\ \text{type})$
- "*likelihood* is probability backwards"

Likelihood, and Conditional Probability

- characteristic PP problem is *particle identification*:
  given detector response to a track, what type of particle is it?
- often based on $\mathcal{L}(\text{type};\ \text{response}) \equiv P(\text{response};\ \text{type})$
- "*likelihood* is probability backwards"
- *e.g.* Belle (II), like many experiments, uses a *likelihood ratio*:

$$\mathcal{R} = \frac{\mathcal{L}(\mathrm{K};\ \theta)}{\mathcal{L}(\mathrm{K};\ \theta) + \mathcal{L}(\pi;\ \theta)} = \frac{P(\theta;\ \mathrm{K})}{P(\theta;\ \mathrm{K}) + P(\theta;\ \pi)}$$

Likelihood, and Conditional Probability

- characteristic PP problem is *particle identification*:
  given detector response to a track, what type of particle is it?
- often based on $\mathcal{L}(\text{type};\ \text{response}) \equiv P(\text{response};\ \text{type})$
- "*likelihood* is probability backwards"
- *e.g.* Belle (II), like many experiments, uses a *likelihood* *ratio*:

$$\mathcal{R} = \frac{\mathcal{L}(\mathrm{K};\ \theta)}{\mathcal{L}(\mathrm{K};\ \theta) + \mathcal{L}(\pi;\ \theta)} = \frac{P(\theta;\ \mathrm{K})}{P(\theta;\ \mathrm{K}) + P(\theta;\ \pi)}$$

- this is frequently misunderstood as a probability: here, let's

Likelihood, and Conditional Probability

- characteristic PP problem is *particle identification*:
  given detector response to a track, what type of particle is it?
- often based on $\mathcal{L}(\text{type};\ \text{response}) \equiv P(\text{response};\ \text{type})$
- "*likelihood* is probability backwards"
- *e.g.* Belle (II), like many experiments, uses a *likelihood ratio*:

$$\mathcal{R} = \frac{\mathcal{L}(\mathrm{K};\ \theta)}{\mathcal{L}(\mathrm{K};\ \theta) + \mathcal{L}(\pi;\ \theta)} = \frac{P(\theta;\ \mathrm{K})}{P(\theta;\ \mathrm{K}) + P(\theta;\ \pi)}$$

- this is frequently misunderstood as a probability: here, let's
  - consider particle ID examples using probabilities explicitly

Likelihood, and Conditional Probability

- characteristic PP problem is *particle identification*:
  given detector response to a track, what type of particle is it?
- often based on $\mathcal{L}(\text{type};\ \text{response}) \equiv P(\text{response};\ \text{type})$
- "*likelihood* is probability backwards"
- *e.g.* Belle (II), like many experiments, uses a *likelihood* *ratio*:

$$\begin{aligned}\mathcal{R} &= \frac{\mathcal{L}(\mathrm{K};\ \theta)}{\mathcal{L}(\mathrm{K};\ \theta) + \mathcal{L}(\pi;\ \theta)} \\ &= \frac{P(\theta;\ \mathrm{K})}{P(\theta;\ \mathrm{K}) + P(\theta;\ \pi)}\end{aligned}$$

- this is frequently misunderstood as a probability: here, let's
  - consider particle ID examples using probabilities explicitly
  - see if this can clarify probabilities *vs* likelihoods

Likelihood, and Conditional Probability

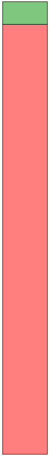
- characteristic PP problem is *particle identification*:
  given detector response to a track, what type of particle is it?
- often based on $\mathcal{L}(\text{type};\ \text{response}) \equiv P(\text{response};\ \text{type})$
- "*likelihood* is probability backwards"
- *e.g.* Belle (II), like many experiments, uses a *likelihood ratio*:

$$\mathcal{R} = \frac{\mathcal{L}(\mathrm{K};\ \theta)}{\mathcal{L}(\mathrm{K};\ \theta) + \mathcal{L}(\pi;\ \theta)} = \frac{P(\theta;\ \mathrm{K})}{P(\theta;\ \mathrm{K}) + P(\theta;\ \pi)}$$

- this is frequently misunderstood as a probability: here, let's
  - consider particle ID examples using probabilities explicitly
  - see if this can clarify probabilities *vs* likelihoods

(it's also a lovely example of Bayes' Theorem)

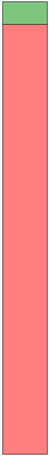
Likelihood, and Conditional Probability



Consider $\mathcal{R} = 0.95$

- naïvely, $P(K) = \mathcal{R} = 95\%$

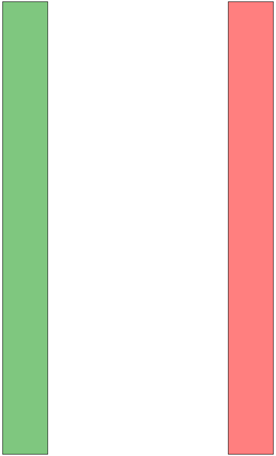
Likelihood, and Conditional Probability



Consider $\mathcal{R} = 0.95$

- naïvely, $P(K) = \mathcal{R} = 95\%$
- <u>not true</u> in general

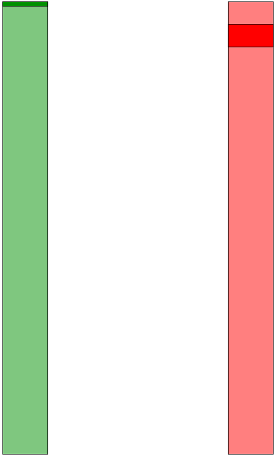
Likelihood, and Conditional Probability



Consider $\mathcal{R} = 0.95$

- naïvely, $P(K) = \mathcal{R} = 95\%$
- <u>not true</u> in general
- $\exists$ pions & kaons

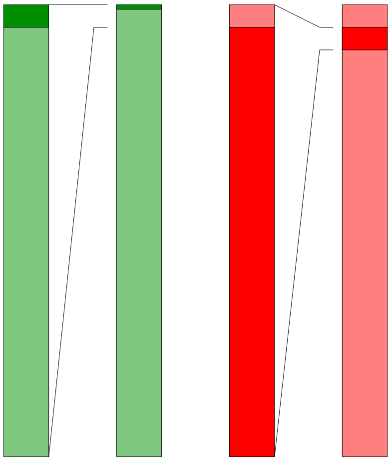
Likelihood, and Conditional Probability



Consider $\mathcal{R} = 0.95$

- naïvely, $P(K) = \mathcal{R} = 95\%$
- not true in general
- $\exists$ pions & kaons
- consider $\{\pi \,|\, \mathcal{R}(\pi) \approx 0.95)\}$ & $\{K \,|\, \mathcal{R}(K) \approx 0.95)\}$

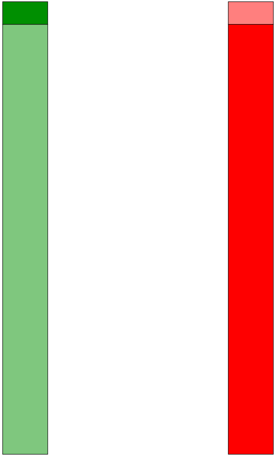
Likelihood, and Conditional Probability



Consider $\mathcal{R} = 0.95$

- naïvely, $P(K) = \mathcal{R} = 95\%$
- not true in general
- $\exists$ pions & kaons
- consider $\{\pi \,|\, \mathcal{R}(\pi) \approx 0.95)\}$ & $\{K \,|\, \mathcal{R}(K) \approx 0.95)\}$
- rescale

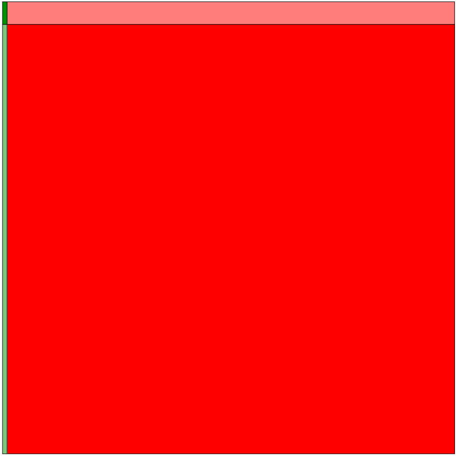
Likelihood, and Conditional Probability



Consider $\mathcal{R} = 0.95$

- naïvely, $P(K) = \mathcal{R} = 95\%$
- not true in general
- $\exists$ pions & kaons
- consider $\{\pi \,|\, \mathcal{R}(\pi) \approx 0.95)\}$ & $\{K \,|\, \mathcal{R}(K) \approx 0.95)\}$
- rescale
- we can do the rest by counting (I am avoiding explaining the continuum case rigorously)

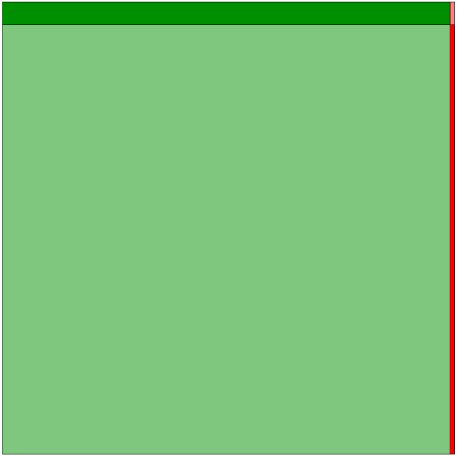
Likelihood, and Conditional Probability



Ω: tracks tagged by $\phi \to \mathrm{KK}$

$$\begin{aligned}
P(\mathrm{K} \mid \Omega) &= 0.99 \\
P(\pi \mid \Omega) &= 0.01 \\
P(\mathrm{K} \mid \mathcal{R}, \Omega) &= \frac{P(\mathcal{R} \mid \mathrm{K}) \cdot P(\mathrm{K} \mid \Omega)}{P(\mathcal{R} \mid \Omega)} \\
&= \frac{P(\mathcal{R} \mid \mathrm{K}) \cdot P(\mathrm{K} \mid \Omega)}{\sum_i P(\mathcal{R} \mid h_i) \cdot P(h_i \mid \Omega)} \\
&= \frac{0.95 * 0.99}{0.95 * 0.99 + 0.05 * 0.01} \\
&= 0.9995
\end{aligned}$$

Likelihood, and Conditional Probability



Ω: tracks tagged by $\mathrm{K}^0_S \to \pi\pi$

$$P(\mathrm{K} \mid \Omega) = 0.01$$

$$P(\pi \mid \Omega) = 0.99$$

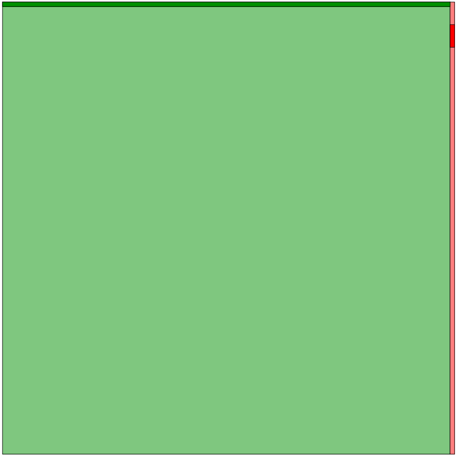
$$P(\mathrm{K} \mid \mathcal{R}, \Omega) = \frac{P(\mathcal{R} \mid \mathrm{K}) \cdot P(\mathrm{K} \mid \Omega)}{P(\mathcal{R} \mid \Omega)}$$

$$= \frac{P(\mathcal{R} \mid \mathrm{K}) \cdot P(\mathrm{K} \mid \Omega)}{\sum_i P(\mathcal{R} \mid h_i) \cdot P(h_i \mid \Omega)}$$

$$= \frac{0.95 * 0.01}{0.95 * 0.01 + 0.05 * 0.99}$$

$$= 0.16$$

Likelihood, and Conditional Probability



Ω: tracks tagged by $\mathrm{K}^0_S \to \pi\pi$

$$P(\mathrm{K} \mid \Omega) = 0.01$$

$$P(\pi \mid \Omega) = 0.99$$

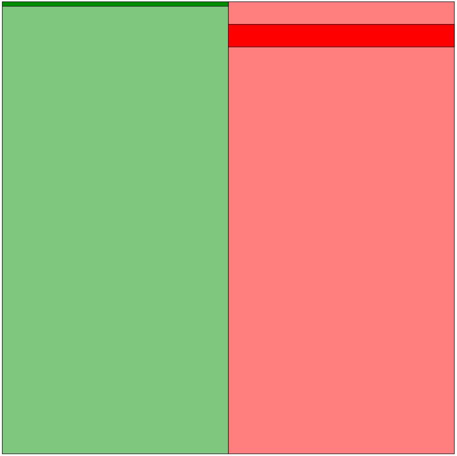
$$P(\mathrm{K} \mid \mathcal{R}, \Omega) = \frac{P(\mathcal{R} \mid \mathrm{K}) \cdot P(\mathrm{K} \mid \Omega)}{P(\mathcal{R} \mid \Omega)}$$

$$= \frac{P(\mathcal{R} \mid \mathrm{K}) \cdot P(\mathrm{K} \mid \Omega)}{\sum_i P(\mathcal{R} \mid h_i) \cdot P(h_i \mid \Omega)}$$

$$= \frac{0.95 * 0.01}{0.95 * 0.01 + 0.05 * 0.99}$$

$$= 0.16 \text{ (rare case in } \Omega)$$

Likelihood, and Conditional Probability



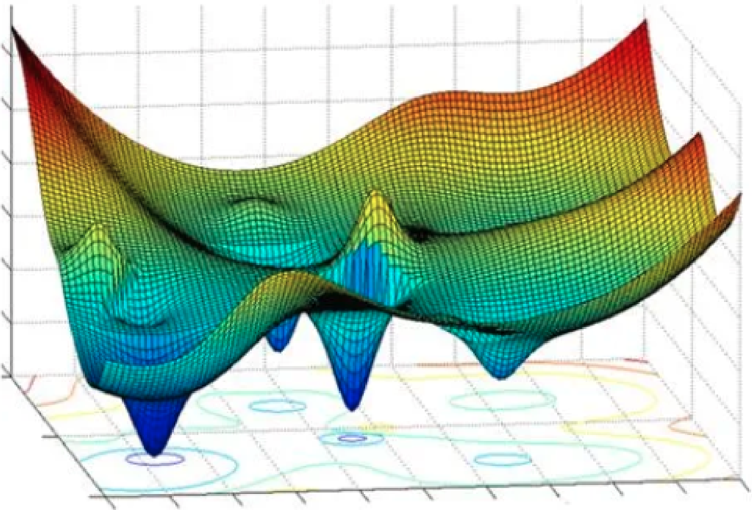
"$P(K) = \mathcal{R} = 95\%$"
only in the special case where
the parent track sample Ω
has 50% kaons and 50% pions

It's all Bayes' Theorem

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad \ldots$$

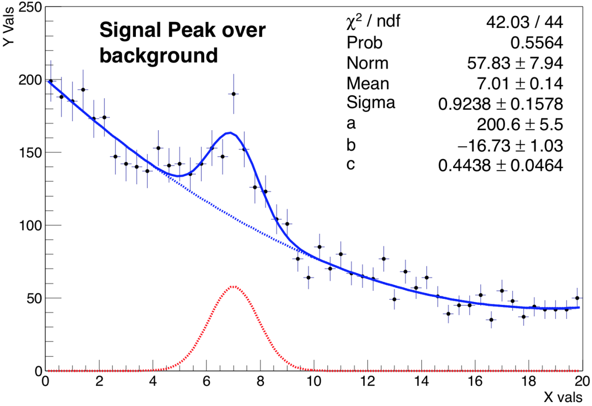
...but without Bayesian statistics.
(We *will* get to Bayesian stats later.)

Maximum Likelihood Estimates and fitting



A simple fitting example

$$f(x) = N\frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right) + a + b.x + c.x^2$$



$\vec{\theta} = (N, \mu, \sigma, a, b, c)$

data $\vec{y}$: the bin counts

$$\mathcal{L}(\vec{\theta};\, \vec{y}) = P(\vec{y}\,|\,\vec{\theta}) = \prod_{i=1}^{N} P(y_i\,|\,\vec{\theta}) = \prod_{i=1}^{N} P_{\text{Poisson}}(y_i;\, \nu_i)$$

where ν_i is the sum of $f(x)$ over the bin i

Likelihood: Maximum Likelihood fits

- independent observations $X = X_1, X_2, \ldots, X_N$
- likelihood $\mathcal{L}(\theta; X) = P(X \mid \theta) = \prod_{i=1}^{N} f(X_i \mid \theta)$
- **maximum likelihood estimate** of θ is *that value* $\hat{\theta}$ *for which* $\mathcal{L}(\theta; X)$ *has its maximum, given the particular observations* X
- we use *function minimization routines* (!) on $-2\ln\mathcal{L}$ to obtain the MLE [routines usually based on `MINUIT`]
- properties of the maximum likelihood:
  - asymptotically *consistent* and *unbiased*
  - asymptotically *Normally distributed* with *minimum variance*

$$V(\hat{\theta}) \overset{N\to\infty}{\longrightarrow} \left\{ E\left[\left(\frac{\partial \ln \mathcal{L}}{\partial \theta} \right)^2 \right] \right\}^{-1}$$

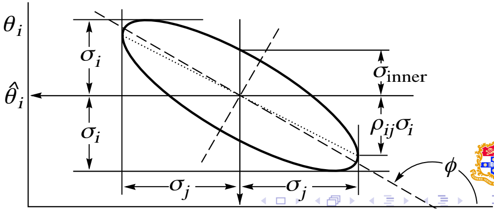
$$\text{estimator of variance } \hat{V}(\hat{\theta}) = \left\{ \left(-\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2} \right) \Bigg|_{\theta=\hat{\theta}} \right\}^{-1}$$

  - asymptotically *invariant*: MLE of $\tau(\theta)$ is $\hat{\tau} = \tau(\hat{\theta})$
  - tends to converge to asymptotic limit faster than other asymptotically efficienct estimators (*e.g.* least squares) do ...

Likelihood: Maximum Likelihood fits

- if $f(X_1,\, X_2,\, X_3,\, \ldots)$ is a multidimensional Gaussian, then $\mathrm{cov}(X_i, X_j)$ gives the *tilt* of the ellipsoid in (X_i, X_j)
- for $N \to \infty$, ML or weighted-least-squares fits return *parameter* *estimates* $\hat{\theta} = (\hat{\theta}_1,\, \hat{\theta}_2,\, \hat{\theta}_3,\, \ldots)$ distributed as a Gaussian about the *true* values θ underlying the data
  — frequentist interp^n: whole expt is a *single random throw*
- the covariances $\mathrm{cov}(\hat{\theta}_i, \hat{\theta}_j)$ form the *covariance matrix* or *error matrix*; the fitter *estimates it*
  - HESSE: from the second derivatives at $(\hat{\theta}_i,\, \hat{\theta}_j)$
  - MINOS: from the shape of $-2\ln\mathcal{L}$ about the minimum

$$\tan 2\phi = \frac{2\,\mathrm{cov}(\hat{\theta}_i, \hat{\theta}_j)}{\sigma_j^2 - \sigma_i^2} = \frac{2\rho_{ij}\sigma_i\sigma_j}{\sigma_j^2 - \sigma_i^2}$$



Likelihood: Simultaneous fits

- suppose you are measuring counts in the signal region:
  - background process with unknown rate b
  - signal process with unknown rate μ
  - $\mathcal{L} = P_p(n_1;\, [\mu + b])$
- you can also make an auxiliary measurement in a bkgd-only region:
  - background process with unknown rate b, same as above
  - $\mathcal{L} = P_p(n_2;\, b)$
- likelihood to determine the signal *and* the bkgd rate:
  $\mathcal{L} = P_p(n_1;\, [\mu + b]) \times P_p(n_2;\, b) = \frac{e^{-[b+\mu]}[b+\mu]^{n_1}}{n_1!} \times \frac{e^{-b}b^{n_2}}{n_2!}$
- straightf^{wd} extension of previous cases; follows from same principles
- if $n_1 \to \{n_i\}$, a series of measurements, say a histogram,
  and $n_2 \to \{m_j\}$, a histogram of the background region,
  then $\boxed{\mathcal{L} \to \mathcal{L}_1(\mu, \boldsymbol{\theta};\, \{n_i\}) \times \mathcal{L}_2(\mu, \boldsymbol{\theta};\, \{m_j\})}$;
  one performs a *simultaneous fit* to the histos to get the MLE's

Statistical and systematic uncertainties

Statistical uncertainties: Law of Large Numbers

- suppose you have a sequence of indep^t random variables X_i
  - with the same mean μ
  - and variances σ_i^2
  - but otherwise distributed "however"
- suppose that the variances are "not too wide":
- if $\lim_{N\to\infty}(\frac{1}{N^2})\sum_{i=1}^{N}\sigma_i^2 = 0$,
  then the *average* $\bar{X}_N = \frac{1}{N}\sum X_i$ converges to the *mean* μ
  "in quadratic mean": $\lim_{N\to\infty} E\left[\left|\bar{X}_N - \mu\right|^2\right] = 0$
- if $\lim_{N\to\infty}\left(\sum_{i=1}^{N}\frac{\sigma_i}{i}\right)^2$ is finite,
  the convergence is "almost certain": $P\left(\lim_{N\to\infty}\bar{X}_N = \mu\right) = 1$
  (the failures have measure zero)
- *i.e.* eventually, you get the real mean

Statistical uncertainties: Central Limit Theorem

- suppose you have a sequence of indep^t random variables X_i
  - with means μ_i
  - and variances σ_i^2
  - but otherwise distributed "however"
- under certain conditions on the variances,
  the sum $S = \sum X_i$ converges to a Gaussian

$$\frac{S - \sum \mu_i}{\sqrt{\sum \sigma_i^2}} \xrightarrow{N\to\infty} \mathcal{N}(0,1)$$

independent of what the individual sub-distributions are

LLN: "[For most things, the average gives you the mean.]"

CLT: "[Put enough things into a blender, and you get a Gaussian.]"

i.i.d.: In particle physics, thanks to QM (!), successive instances of a state prepared the same way (*e.g.* particle decays, e^+e^- collisions . . .) are *independent and identically distributed*, the statistical gold standard; statistical techniques work properly, "out of the box" . . .

Statistical uncertainties are well-behaved . . .

. . . and systematic uncertainties are not. Well, not always.

I am not going to define systematic uncertainties in this talk.

(There *are* some musings about them in the appendix.)

But I am going to give you my working taxonomy of sysematics:

1. (uncertainties on) the INPUTS to the measurement
2. (uncertainties on) AUXILIARY measurements
3. (uncertainties on) the CALIBRATION of the apparatus

(uncertainties on) the INPUTS to the measurement

- numbers with uncertainties
- theoretical uncertainties

INPUTS: numbers with uncertainties, *e.g.* $\mathcal{B}$

Belle pub632: PRD 107, 072008 (2023); $e^+e^- \to \Sigma\bar{\Sigma}$ via ISR

- $e^+e^- \to \Sigma\bar{\Sigma}$ measurement
- ISR sample, relatively clean $\longrightarrow$
- (we will discuss the background estimation method later)
- reconstruct $\Sigma \to \Lambda\gamma$
- signal extraction includes event counts, efficiencies, ...
  and the known $\mathcal{B}(\Lambda \to p\pi)$

$\longrightarrow$ 0.8% uncertainty on the result

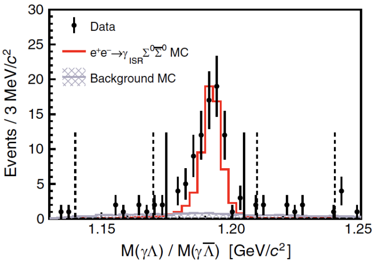
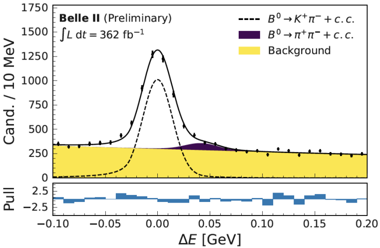
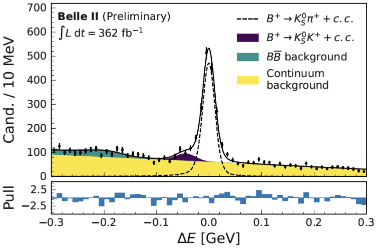


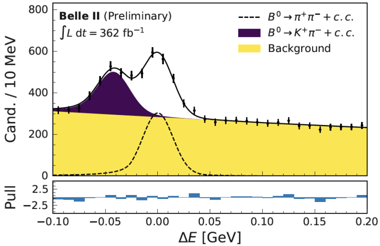
FIG. 2. The invariant mass of the accepted $\gamma\Lambda$ and $\gamma\bar{\Lambda}$ candidates. The points with error bars are experimental data and the red histogram shows the $e^+e^- \to \gamma_{\rm ISR}\Sigma^0\bar{\Sigma}^0$ MC events with $\Sigma^0/\bar{\Sigma}^0$ correctly reconstructed. The hatched histogram is a mixture of events in $e^+e^- \to \gamma_{\rm ISR}\Lambda\bar{\Lambda}$ and $e^+e^- \to \gamma_{\rm ISR}\Sigma^0\bar{\Lambda}$ MC samples and the $e^+e^- \to \gamma_{\rm ISR}\Sigma^0\bar{\Sigma}^0$ MC events with $\Sigma^0/\bar{\Sigma}^0$ misreconstructed. The solid and dashed vertical lines denote the $\Sigma^0/\bar{\Sigma}^0$ signal and sideband regions, respectively.

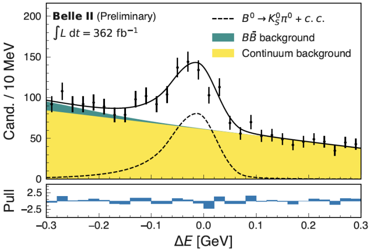
INPUTS: numbers with uncertainties, *e.g.* $N_{B\bar{B}}$, f^{00}

Belle II pub24: 2310.06381 → PRD; $B \to K\pi$, $\pi\pi$ BFs and $\mathcal{A}_{CP}$









INPUTS: numbers with uncertainties, *e.g.* $N_{B\overline{B}}$, f^{00}

Belle II pub24: 2310.06381 → PRD; $B \to K\pi$, $\pi\pi$ BFs and $\mathcal{A}_{CP}$

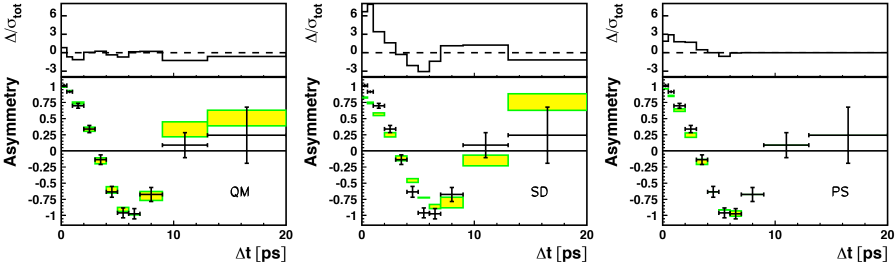
TABLE II. Summary of the relative systematic uncertainties (%) on the branching fractions.

Source	$B^0 \to K^+\pi^-$	$B^0 \to \pi^+\pi^-$	$B^+ \to K^+\pi^0$	$B^+ \to \pi^+\pi^0$	$B^+ \to K^0_S\pi^+$	$B^0 \to K^0_S\pi^0$
Tracking	0.5	0.5	0.2	0.2	0.7	0.5
$N_{B\overline{B}}$	1.5	1.5	1.5	1.5	1.5	1.5
$f^{+-/00}$	2.5	2.5	2.4	2.4	2.4	2.5
π^0 efficiency	–	–	3.8	3.8	–	3.8
K^0_S efficiency	–	–	–	–	2.0	2.0
CS efficiency	0.2	0.2	0.7	0.7	0.5	1.7
PID correction	0.1	0.1	0.1	0.2	–	–
ΔE shift and scale	0.1	0.2	1.2	2.0	0.3	1.7
$K\pi$ signal model	0.1	0.2	0.1	<0.1	<0.1	0.1
$\pi\pi$ signal model	<0.1	0.1	<0.1	<0.1	–	–
$K\pi$ feed-across model	<0.1	0.1	<0.1	0.1	–	–
$\pi\pi$ feed-across model	0.1	0.2	<0.1	0.1	–	–
$K^0_S K^+$ model	–	–	–	–	0.1	–
$B\overline{B}$ model	–	–	0.3	0.5	<0.1	0.3
$q\overline{q}$ flavor model	–	–	–	–	–	0.9
Multiple candidates	<0.1	<0.1	1.0	0.3	0.1	0.3
Total	3.0	3.0	5.1	5.2	3.6	5.8

- number of $B\overline{B}$, and B^+B^- *vs* $B^0\overline{B}^0$ fraction, are uncertain $\longrightarrow$ normalisation uncertainty on all branching fractions
- this is the dominant uncertainty for $K\pi$; disappears on the $\mathcal{A}_{CP}$s

INPUTS: numbers with uncertainties, *e.g.* Δm

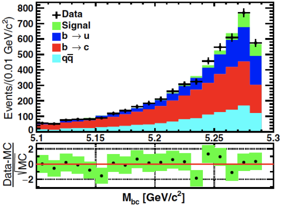
Belle pub197: PRL 99, 131802 (2007); EPR-type flavour entanglement in $B^0\overline{B}^0$



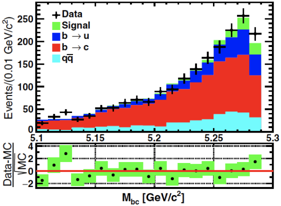
- fits to a Δt distrib^n in 11 bins; functional form depends on Δm
- Δm is not interesting here, but floating it leads to loss of sensitivity
- world average* measurement: $\langle \Delta m \rangle = (0.496 \pm 0.014)\,\mathrm{ps}^{-1}$
- added to the fit by what is now called Gaussian constraint
- large effect on sensitivity: see the yellow boxes

INPUTS: theoretical uncertainties

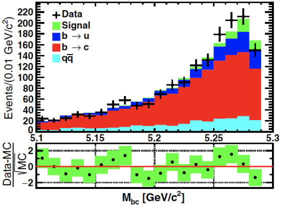
Belle pub582: PRD 106, 032013 (2022); $\mathcal{B}(B^+ \to \eta^{(\prime)}\ell nu)$



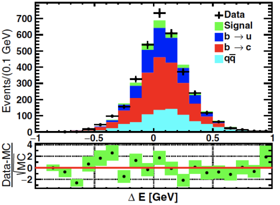
(a) $M_{\rm bc}(\eta \to \gamma\gamma)$



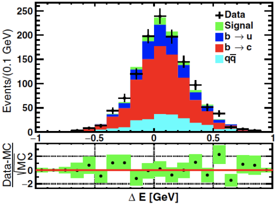
(b) $M_{\rm bc}(\eta \to \pi^+\pi^-\pi^0)$



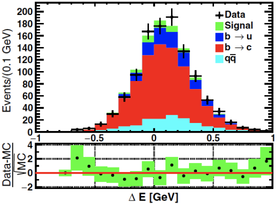
(c) $M_{\rm bc}(\eta' \to \pi^+\pi^-\eta(\gamma\gamma))$



(d) $\Delta E(\eta \to \gamma\gamma)$



(e) $\Delta E(\eta \to \pi^+\pi^-\pi^0)$



(f) $\Delta E(\eta' \to \pi^+\pi^-\eta(\gamma\gamma))$

INPUTS: theoretical uncertainties

Belle pub582: PRD 106, 032013 (2022); $\mathcal{B}(B^+ \to \eta^{(\prime)}\ell nu)$

- form factor assumptions are embedded in signal and background shapes
- uncerts in FF params propagated as systs
- for some, the underlying *model* is changed, and the shift in result used
- these are *small* uncerts — see later

The signal decays $B^+ \to \eta\ell^+\nu_\ell$ and $B^+ \to \eta'\ell^+\nu_\ell$ are reweighted from the ISGW2 model [30] to the model taken from Ref. [31] with the form factors updated to Ref. [32], using the BZ parametrization and assuming uncorrelated parameters. The decay $B^+ \to \omega\ell^+\nu_\ell$ is modeled according to Ref. [33] in the MC used and reweighted to Ref. [34] for comparison. The shape of the inclusive component [35] of the $\bar{b} \to \bar{u}\ell^+\nu_\ell$ transitions is also considered. The form factor uncertainties listed in Table III are based on those reported in the publications they were obtained from. Despite having a slowly varying efficiency the $\eta \to \gamma\gamma$ mode appears to have the largest such uncertainty.

TABLE III. Breakdown of the systematic uncertainty in %.

Source	$\eta(\gamma\gamma)$	$\eta(\pi^+\pi^-\pi^0)$	η'
Statistical	22	39	46
Combined Systematic	11	14	11
$\mathcal{B}(B^\pm \to X_{Bkg})$	2.4	1.7	1.3
$\mathcal{B}(\eta^{(\prime)} \to X)$	0.51	1.2	1.7
$B \to D^{(*,**)}\ell^+\nu_\ell$ form factor	0.82	1.1	1.3
$B \to \eta^{(\prime)}\ell^+\nu_\ell$ form factor	3.0	2.9	0.14
$B \to \omega\ell^+\nu_\ell$ form factor	0.81	2.1	2
$\bar{b} \to \bar{u}\ell^+\nu_\ell$ shape	0.39	0.15	0.21
Background with K_L^0	3.5	8.6	3.8
Continuum	0.2	0.62	0.63
$N_{B\bar{B}}$	1.4	1.4	1.4
$\mathcal{B}(\Upsilon(4S) \to B^+B^-)$	1.2	1.2	1.2
$\bar{b} \to \bar{u}\ell^+\nu_\ell$ yield	4.1	5.2	4.4
Monte Carlo statistics	0.86	1.3	2.3
Charged tracks	0.35	1.1	1.1
γ detection	4.0	2.5	4.0
Electron PID	1.6	1.6	1.5
Muon PID	2.1	2.1	2
First $\pi^\pm$ PID	0	0.97	1.1
Second $\pi^\pm$ PID	0	1.3	2.2
Misidentified Leptons	4.3	5.5	2.3
Control Mode	5.0	5.0	5.0

(uncertainties on) any AUXILIARY measurements

- . . . of efficiencies
- . . . of rates
- . . . of the resolution function
- . . . of the interaction region

AUXILIARY measurements . . . of efficiencies (1)

Belle II pub24: 2310.06381 → PRD; $B \to K\pi$, $\pi\pi$ BFs and $\mathcal{A}_{CP}$

TABLE II. Summary of the relative systematic uncertainties (%) on the branching fractions.

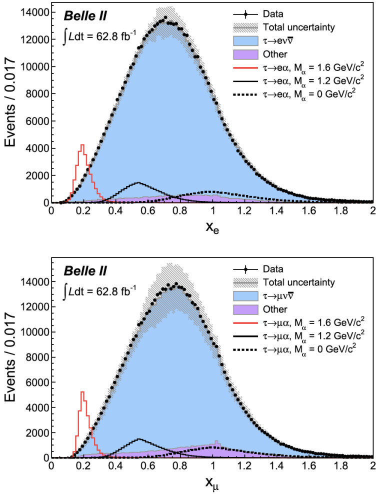
Source	$B^0 \to K^+\pi^-$	$B^0 \to \pi^+\pi^-$	$B^+ \to K^+\pi^0$	$B^+ \to \pi^+\pi^0$	$B^+ \to K_S^0\pi^+$	$B^0 \to K_S^0\pi^0$
Tracking	0.5	0.5	0.2	0.2	0.7	0.5
$N_{B\overline{B}}$	1.5	1.5	1.5	1.5	1.5	1.5
$f^{+-/00}$	2.5	2.5	2.4	2.4	2.4	2.5
π^0 efficiency	–	–	3.8	3.8	–	3.8
K_S^0 efficiency	–	–	–	–	2.0	2.0

- we measure the **track-finding efficiency** in dedicated analyses, but the value has an uncertainty:
  this appears according to the number of tracks in each mode
- π^0-, K_S^0-finding efficiencies likewise
- again these will "cancel in the ratio" for $\mathcal{A}_{CP}$,
  up to possible charge-dependent effects that need to be checked
- in a different sort of analysis, say with a normalisation mode, they will *not* necessarily cancel:
  - they are in general p, p_T, $\cos\theta$-*etc.*-dependent, *esp.* for PID
  - signal and normalisation modes will not have the same distribution . . .

AUXILIARY measurements . . . of efficiencies (2)

Belle II pub10: PRL 130, 181803 (2023); LFV $\tau \to \ell\alpha$ search

Knowledge of **PID efficiencies** can be limited by calibration on data:



The leading systematic uncertainties originate from the corrections to the lepton-identification efficiency and particle misidentification rate, based on comparison of calibration samples in data and simulated events. These corrections depend on the momentum and polar angle; their typical ranges are summarized in Table II. The resulting uncertainties are asymmetric and strongly depend on x_ℓ; their ranges and averaged values over the standard-model yields are also reported in the same table. The contribution from lepton-identification efficiency partially cancels in the ratio between signal and normalization channels; while the contribution from particle misidentification rates does not, as it affects only other background sources.

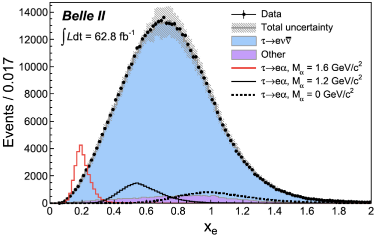
TABLE II. Typical ranges for corrections to the lepton-identification efficiencies and misidentification rates, together with ranges for their respective uncertainties and their average values.

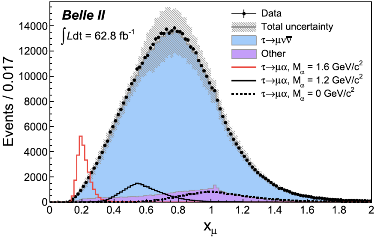
	Correction range	Uncertainty range(%)	Average uncert.(%)
Electron identification	0.84–1.06	0.9–12.6	+5.3, −2.9
Muon identification	0.63–1.02	1.3–32.8	+11.7, −1.6
Electron misidentification	0.6–6.0	4.3–34.6	+17.6, −14.7
Muon misidentification	0.3–1.5	1.4–37.0	+18.0, −18.2

AUXILIARY measurements . . . of efficiencies (2)

Belle II pub10: PRL 130, 181803 (2023); LFV $\tau \to \ell\alpha$ search

Trigger and other efficiences in this analysis are "hidden" in the $\mathcal{L}$:





Uncertainties from the trigger and π^0 reconstruction efficiency corrections are also taken into account. Trigger uncertainties range in 0.1%–4% for the electron channel and in 0.2%–1.5% for the muon channel, depending on x_ℓ. Neutral pion reconstruction efficiency is evaluated from studies on independent samples to be 0.914 ± 0.020. Each of these systematic uncertainties is included in the likelihood as an additional shape-correlated nuisance parameter that is assumed to follow a Gaussian distribution. Other sources of uncertainty from track reconstruction efficiency, beam-energy determination, relative reconstruction efficiency, and momentum-scale correction have negligible impact on the results.

AUXILIARY measurements . . . of efficiencies (3)

Belle II pub13: PRD 107, 112009 (2023); $\pi^0\pi^0$ BF and $\mathcal{A}_{CP}$

sometimes the reported "uncertainty" of the auxiliary measurement is secondary: the systematic is dominated by our **lack of understanding or confidence** in what is going on . . .

The main sources of systematic uncertainties are listed in Table I and are evaluated as follows. A 3.4% systematic uncertainty associated with the π^0 reconstruction efficiency is determined from data using the decays $D^{*-} \to \bar{D}^0(\to K^+\pi^-\pi^0)\pi^-$ and $D^{*-} \to \bar{D}^0(\to K^+\pi^-)\pi^-$, where the π^0 selection is identical to that of the signal. The π^0 reconstruction efficiency as a function of momentum is also measured using $\tau^- \to 3\pi\pi^0\nu$ and $\tau^- \to 3\pi\nu$ decays. A difference of 4.7% in efficiency is observed between the measurement based on D decays and the measurement based on τ leptons. This difference increases the systematic uncertainty for a total of 5.8% per pion. The total systematic uncertainty associated with the π^0 reconstruction efficiency is then 11.6%, as there are two pions and their errors are fully correlated.

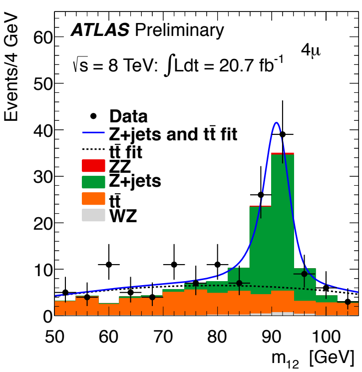
TABLE I. Summary of systematic uncertainties. The total is calculated by adding all systematic uncertainties in quadrature.

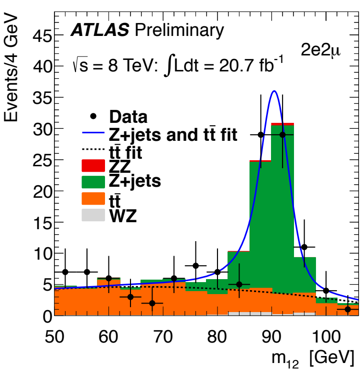
Source	$\mathcal{B}(\%)$	$\mathcal{A}_{CP}$
π^0 reconstruction efficiency	11.6	...
Continuum parametrization	7.4	0.02
Continuum classifier efficiency	6.5	...
$1 + f^{+-}/f^{00}$	2.5	...
Fixed $B\bar{B}$ background yield	2.3	0.01
Fixed signal r bin fractions	2.2	0.01
Knowledge of the photon-energy scale	2.0	...
Assumption of independence of ΔE from r	1.8	<0.01
Number of $B\bar{B}$ meson pairs	1.5	<0.01
Choice of $(M_{bc}, \Delta E)$ signal model	1.3	0.02
Fixed continuum r bin fraction	1.1	<0.01
Branching fraction fit bias	1.0	...
Best candidate selection	0.2	<0.01
Mistagging parameters	...	0.05
Potential nonzero $B\bar{B}$ background $\mathcal{A}_{CP}$	...	0.03
$\mathcal{A}_{CP}$ fit bias	...	0.02
Continuum $q \cdot r$ asymmetry	...	0.01
Total	16.2	0.07

AUXILIARY measurements . . . of rates

ATLAS–CONF–2013–013; Higgs properties in $H \to ZZ^{(*)} * \to 4\ell$

- $H \to \ell\ell + \mu\mu$ have "reducible" bkgds due to $t\bar{t}$ & $Z +$ jets events
- normalisations are set using complementary "control regions"
  1. removing $\mu\mu$ isolation cuts, & requiring ≥ 1 isol^n failure (not shown)
  2. removing $\mu\mu$ isolation cuts, & requiring ≥ 1 IP significance failure:





AUXILIARY measurements . . . of the resolution fn

Belle II pub16: PRD 107, L091102 (2023); B^0 lifetime and Δm measurement

Leading systematic: **function params** that can't all be floated at once . . .

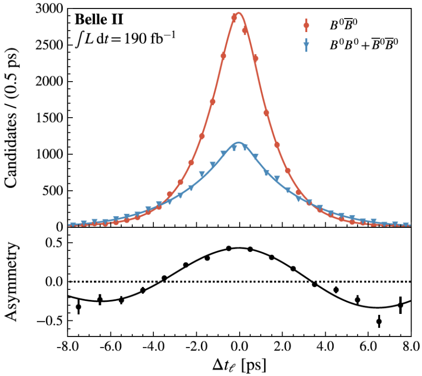


FIG. 3. Distribution of Δt_ℓ in data (points) and the fit model (lines) for opposite-flavor candidate pairs (red) and same-flavor pairs (blue) and their asymmetry (black).

There are several sources of systematic uncertainty; these are listed in Table I and described below. The dominant systematic uncertainty is due to potential discrepancies between the assumed values (fixed in the fit) of the response-function parameters and the true values in the data. For each fixed parameter, we repeat the fit with the parameter allowed to vary. We add all the resulting changes in the result in quadrature and include this value as a systematic uncertainty.

TABLE I. Systematic uncertainties.

Source	τ_{B^0} [ps]	Δm_d [ps^{-1}]
Fixed response-function parameters	0.0063	0.003
Analysis bias	0.004	0.001
Detector alignment	0.003	0.002
Interaction-region precision	0.002	0.001
C-distribution modeling	0.000	0.001
$\sigma_{\Delta t_\ell}$-distribution modeling	0.001	0.001
Correlations of ΔE or C and Δt_ℓ	0.001	0.000
Total systematic uncertainty	0.008	0.005
Statistical uncertainty	0.013	0.008

AUXILIARY measurements . . . of the IR

Belle II pub16: PRD 107, L091102 (2023); B^0 lifetime and Δm measurement

Smaller systematic: **imprecise knowledge of interaction region**

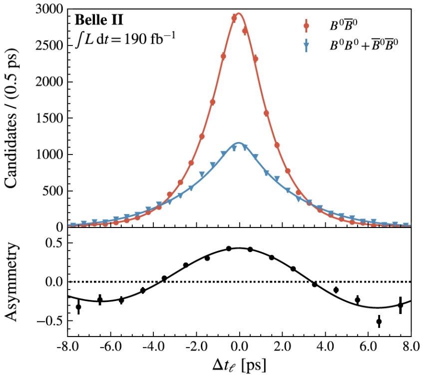


FIG. 3. Distribution of Δt_ℓ in data (points) and the fit model (lines) for opposite-flavor candidate pairs (red) and same-flavor pairs (blue) and their asymmetry (black).

Because we adjust the B_{sig} decay vertex position so that the vector connecting the IR and decay vertex is parallel to the B_{sig} momentum, the precision to which we know the IR affects our determination of ℓ. We repeat our analysis on simulated data in which we shift, rotate, and rescale the IR within its measured uncertainties and assign the changes in the results as systematic uncertainties. We perform an analogous check with changes to $\sqrt{s}$ and the magnitude and direction of the boost vector and find that the results change negligibly.

τ: 0.002 *vs* 0.008 ps total

Δm: 0.001 *vs* 0.005 ps^{-1} total

(uncertainties on) the CALIBRATION of ...

- ... of the measurement of specific quantities
- ... of "environmental" quantities
- ... of the experimental technique as a whole
- ... *of the analyst*: your own choices
- ... by searching for mistakes

CALIBRATION . . . of measurement of specific quantities

Belle II pub15: PRL 127, 211801 (2021); *D* lifetime

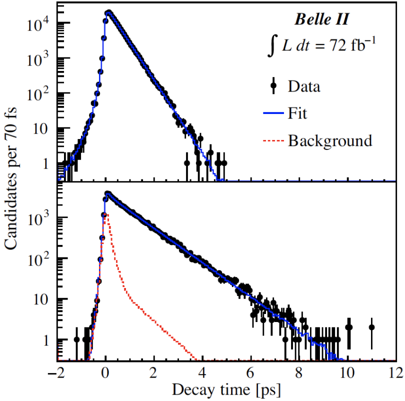


FIG. 2. Decay-time distributions of (top) $D^0 \to K^-\pi^+$ and (bottom) $D^+ \to K^-\pi^+\pi^+$ candidates in their respective signal regions with fit projections overlaid.

The **momentum scale** is important:

- default factor 1.00056
- recommended range $[1.00014, 1.00107]$
- uncertainty is subleading:

TABLE I. Systematic uncertainties.

Source	$\tau(D^0)$ [fs]	$\tau(D^+)$ [fs]
Resolution model	0.16	0.39
Backgrounds	0.24	2.52
Detector alignment	0.72	1.70
Momentum scale	0.19	0.48
Total	0.80	3.10

Imperfectly known **vertex resolution** is another uncertainty of this type

CALIBRATION … of "environmental" quantities

Belle II pub15: PRL 127, 211801 (2021); *D* lifetime

Tracking & vertexing **assumes the alignment** of subdetector elements:

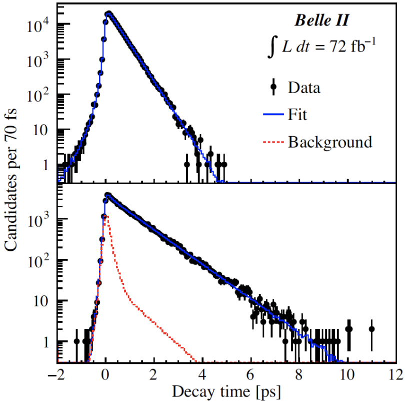


FIG. 2. Decay-time distributions of (top) $D^0 \to K^-\pi^+$ and (bottom) $D^+ \to K^-\pi^+\pi^+$ candidates in their respective signal regions with fit projections overlaid.

background, and cosmic-ray events [26]. Unaccounted-for misalignment can bias the measurement of the charmed decay lengths and hence their decay times. Two sources of uncertainties associated with the alignment procedure are considered: the statistical precision and a possible systematic bias. Their effects are evaluated using simulated signal-only decays reconstructed with a misaligned detector. For the statistical contribution, we consider configurations derived from comparison of alignment parameters determined from data acquired on two consecutive days. These configurations have magnitudes of misalignment comparable to the alignment precision as observed in data averaged over a typical alignment period. For the systematic contribution, we consider configurations derived from simulation studies in which coherent global deformations of the vertex detectors (e.g., radial expansion) are introduced [27]. These deformations have magnitudes, determined by the most misaligned sensors, ranging from about 50 to 700 μm. The alignment procedure determines the magnitude of these deformations within 4 μm accuracy.

CALIBRATION ... of "environmental" quantities

Belle II pub15: PRL 127, 211801 (2021); *D* lifetime

Tracking & vertexing **assumes the alignment** of subdetector elements:

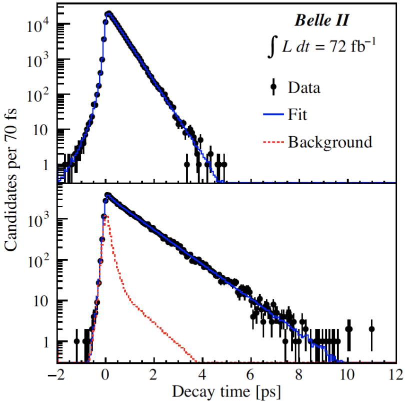


FIG. 2. Decay-time distributions of (top) $D^0 \to K^-\pi^+$ and (bottom) $D^+ \to K^-\pi^+\pi^+$ candidates in their respective signal regions with fit projections overlaid.

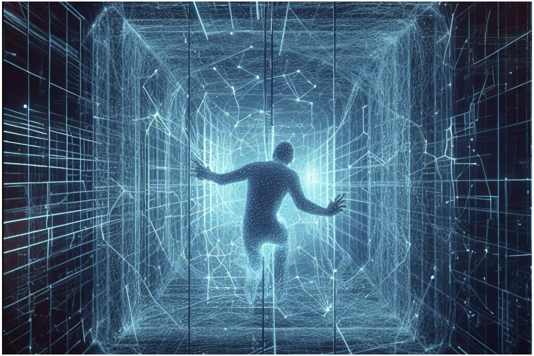
We consider configurations in which the CDC is perfectly aligned and configurations in which it is misaligned. Possible effects on the determination of the IR are also introduced by using parameters measured on misaligned samples of simulated $e^+e^- \to \mu^+\mu^-$ events, to fully mimic the procedure used for real data. For each misalignment configuration, we fit to the reconstructed signal candidates and estimate the lifetime bias. We estimate the systematic uncertainty due to imperfect detector alignment as the sum in quadrature of the largest biases observed in each of the statistical and systematic contributions. The resulting uncertainties are 0.72 and 1.70 fs for $D^0 \to K^-\pi^+$ and $D^+ \to K^-\pi^+\pi^+$ decays, respectively. The absolute length scale of the vertex detector is determined with a precision significantly better than 0.01% and contributes negligibly to the systematic uncertainty.

- Dominant uncertainty for $\tau(D^0)$
- $\vec{B}$-field anisotropy is another classic example

CALIBRATION . . . of the experimental technique as a whole

Here we are putting the whole measurement technique inside a virtual box, as an “instrument” to be calibrated. Typical issues:

- known limitations and omissions in the method
- the equations and parameterizations
- fits: linearity/bias tests
- larger analysis chain: control and validation region studies



Calibration of technique: limitations & omissions

Belle II pub15: PRL 127, 211801 (2021); *D* lifetime

A small background **was left out of the $D^0 \to K^-\pi^+$ fit:**

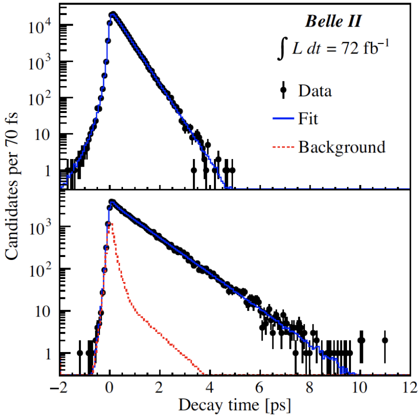


FIG. 2. Decay-time distributions of (top) $D^0 \to K^-\pi^+$ and (bottom) $D^+ \to K^-\pi^+\pi^+$ candidates in their respective signal regions with fit projections overlaid.

The background neglected in the $D^0 \to K^-\pi^+$ fit could result in a systematic bias on the measured lifetime. To estimate the size of the bias, we fit our model that neglects the background to 500 resampled sets of simulated e^+e^- collisions, each having the same size and signal-to-background proportion as the data. The measured lifetimes are corrected by subtracting the bias due to the neglected t vs σ_t correlations. The average absolute difference between the resulting value and the simulated lifetime, 0.24 fs, is assigned as a systematic uncertainty due to the neglected background contamination in the $D^0 \to K^-\pi^+$ fit.

TABLE I. Systematic uncertainties.

Source	$\tau(D^0)$ [fs]	$\tau(D^+)$ [fs]
Resolution model	0.16	0.39
Backgrounds	0.24	2.52
Detector alignment	0.72	1.70
Momentum scale	0.19	0.48
Total	0.80	3.10

Calibration of technique: limitations & omissions

Belle II pub15: PRL 127, 211801 (2021); *D* lifetime

The **sideband method** for $D^+ \to K^-\pi^+\pi^+$ bkgd **may be imperfect:**

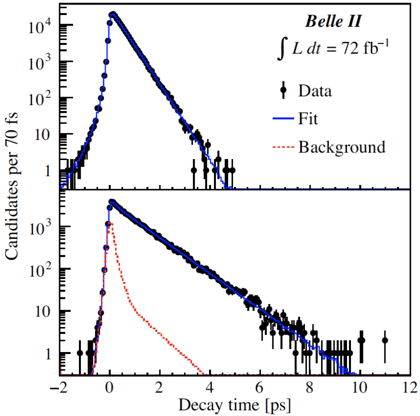


FIG. 2. Decay-time distributions of (top) $D^0 \to K^-\pi^+$ and (bottom) $D^+ \to K^-\pi^+\pi^+$ candidates in their respective signal regions with fit projections overlaid.

The background contamination under the $D^+ \to K^-\pi^+\pi^+$ peak is already accounted for in the fit of the D^+ lifetime using sideband data. In simulation, the sideband (t, σ_t) distribution describes the background (t, σ_t) distribution in the signal region well. The same might not hold in data given that some disagreement is observed between data and simulation in the t distribution of the candidates populating the sideband. We fit to one thousand samples of simulated data obtained by sampling the fit PDF for the signal region and by resampling from the simulated e^+e^- collisions for the sideband. The resulting samples feature sideband data that differ from the background in the signal region with the same level of disagreement as observed between data and simulation. The absolute average difference between the measured and simulated lifetimes, 2.52 fs, is assigned as a systematic uncertainty due to the modeling of the background (t, σ_t) distribution.

Source	$\tau(D^0)$ [fs]	$\tau(D^+)$ [fs]
Resolution model	0.16	0.39
Backgrounds	0.24	2.52
Detector alignment	0.72	1.70
Momentum scale	0.19	0.48
Total	0.80	3.10

Calibration of technique: eqns & parameterizations

Belle II pub15: PRL 127, 211801 (2021); *D* lifetime

Changing how the background is estimated **has a negligible effect:**

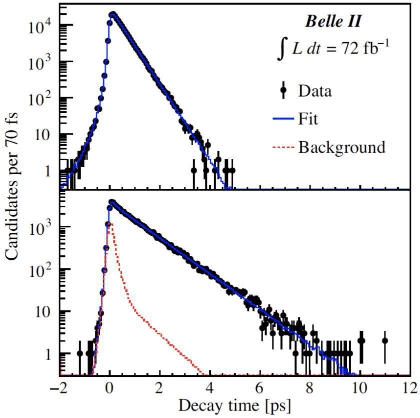


FIG. 2. Decay-time distributions of (top) $D^0 \to K^-\pi^+$ and (bottom) $D^+ \to K^-\pi^+\pi^+$ candidates in their respective signal regions with fit projections overlaid.

In the lifetime fit, the fraction of background candidates in the signal region is constrained from the fit to the $m(K^-\pi^+\pi^+)$ distribution. When we change this background fraction to values obtained from fitting to the $m(K^-\pi^+\pi^+)$ distribution with alternative signal and background PDFs, the change in the measured lifetime is negligible.

Calibration of technique: eqns & parameterizations

Belle II pub15: PRL 127, 211801 (2021); *D* lifetime

Neglect of correlations in fitting **has a (small) noticeable effect:**

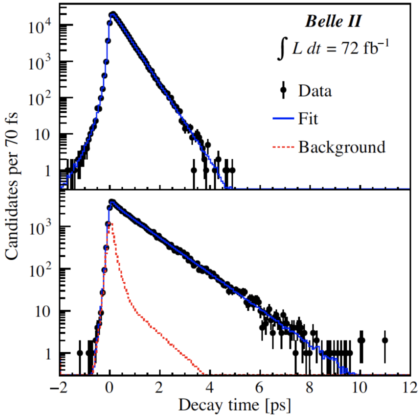


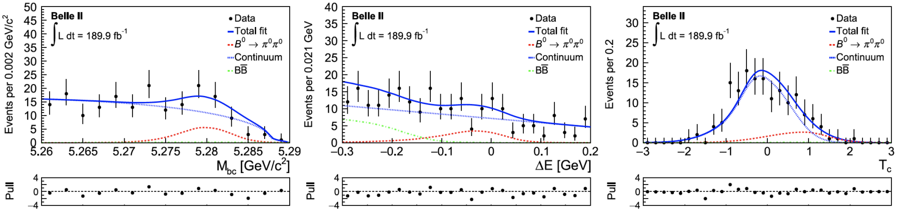
FIG. 2. Decay-time distributions of (top) $D^0 \to K^-\pi^+$ and (bottom) $D^+ \to K^-\pi^+\pi^+$ candidates in their respective signal regions with fit projections overlaid.

The decay time and decay-time uncertainty are observed to be correlated in data and simulation reproduces these effects well. The dominant effect is that small σ_t values correspond to larger true decay times (and vice versa). These correlations, when neglected in the fits, result in an imperfect description of the t distribution as a function of σ_t. To quantify the impact on the results, our model that neglects the correlations is fit to 1000 samples of signal-only simulated decays, each the same size as the data. The samples are obtained by resampling, with repetition, a set of simulated e^+e^- collisions corresponding to an integrated luminosity of 500 fb^{-1}. Upper bounds of 0.16 and 0.39 fs on the average absolute deviations of the measured lifetimes from their true values are derived and assigned as the systematic uncertainty due to the imperfect resolution model for the $D^0 \to K^-\pi^+$ and $D^+ \to K^-\pi^+\pi^+$ cases,

Source	$\tau(D^0)$ [fs]	$\tau(D^+)$ [fs]
Resolution model	0.16	0.39
Backgrounds	0.24	2.52
Detector alignment	0.72	1.70
Momentum scale	0.19	0.48
Total	0.80	3.10

Calibration of technique: eqns & parameterizations

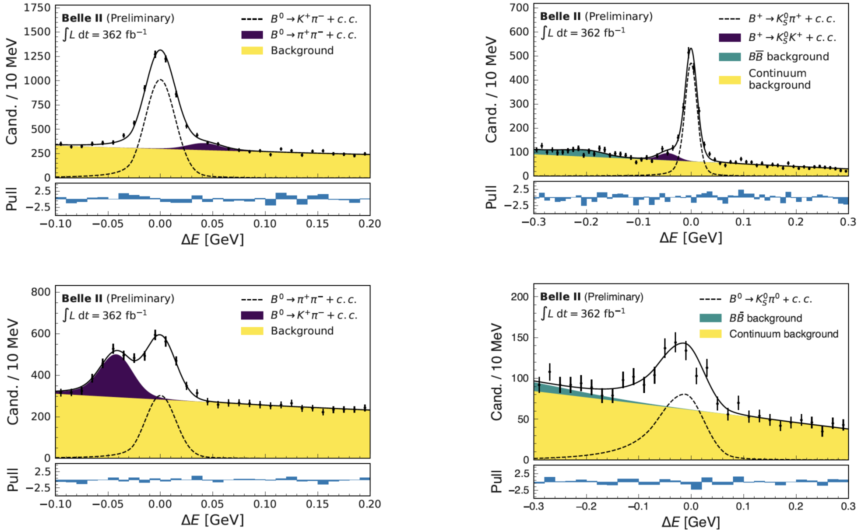
Belle II pub13: PRD 107, 112009 (2023); $\pi^0\pi^0$ BF and $\mathcal{A}_{CP}$



- some $B\overline{B}$, a *lot* of continuum background under the 3D peak
- continuum shapes **taken from the data sideband**
  $M_{bc} \in (5.22, 5.27)\,\mathrm{GeV}/c^2$, $\Delta E \in (0.1, 0.5)\,\mathrm{GeV}$
- uncertainty estimated by shifting shape params by $\pm 1\sigma$ one-by-one (with others shifting per fitted correl^{ns}) and checking yield changes
- after $\epsilon(\pi^0)$, **this is the dominant uncertainty** on $\mathcal{B}(B^0 \to \pi^0\pi^0)$
- parameterization will come up again, under another heading

Calibration of technique: ML fits (1)

Belle II pub24: 2310.06381 → PRD; $B \to K\pi$, $\pi\pi$ BFs and $\mathcal{A}_{CP}$



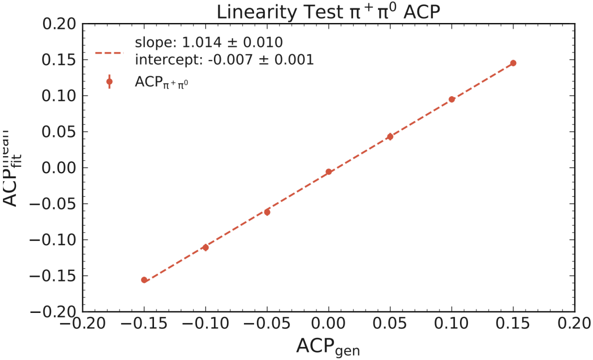
Calibration of technique: ML fits (1)

Belle II pub24: 2310.06381 → PRD; $B \to K\pi$, $\pi\pi$ BFs and $\mathcal{A}_{CP}$

Fits may not behave asymptotically/ideally,
even if nothing is “omitted” or wrong: hence “linearity and bias tests”, etc.

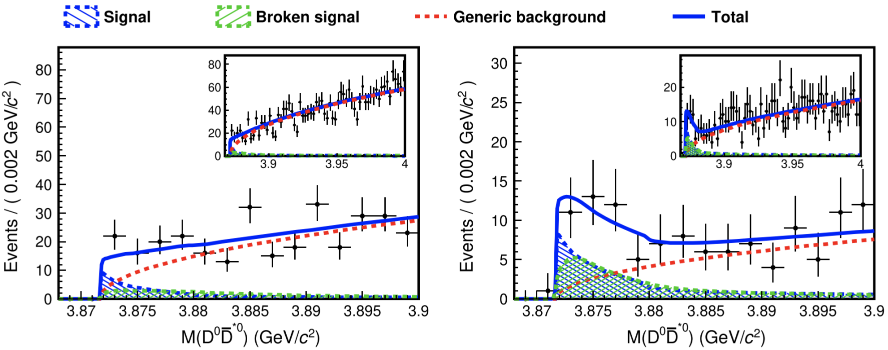
for $\pi^+\pi^0$ and $K^0_S\pi^0$:

- this is the dominant systematic on $\mathcal{A}_{CP}$
- these modes are statistically limited
- negligible for the branching fraction fits . . .



Calibration of technique: ML fits (2)

Belle II pub642: PRD 107, 112011 (2023) $X(3872)$ lineshape in $B \to D^0\overline{D}^{*0}K$



Extremely complex analysis, fitting a signal lineshape at threshold, over background, with substantial broken signal, and unstable decay daughters. One of the lineshapes exhibits scaling behaviour in some parameters . . .

Calibration of technique: ML fits (2)

Belle II pub642: PRD 107, 112011 (2023) $X(3872)$ lineshape in $B \to D^0 \overline{D}^{*0} K$

- BW & Flatté fits
- Flatté fit is very nonlinear:
- significantly changes the reported results: can't deal with this by just "adding a systematic term" . . .

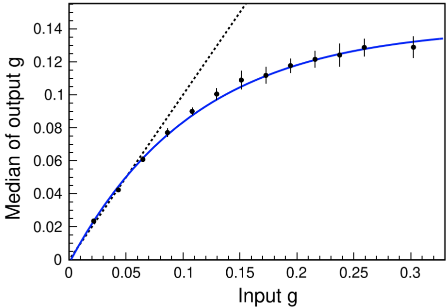


FIG. 5. The median of output values of the coupling constant g, as a function of the input g, evaluated using pseudo-experiments. The dotted black line represents perfect linearity $g_{\text{out}} = g_{\text{in}}$. The solid blue curve represents the threshold function $g_{\text{out}} = 0.14(1 - \exp(-9g_{\text{in}}))$.

Calibration of technique: analysis chain

Belle II pub20: PRD 108, 072012 (2023); CP asymmetries in $B^0 \to \phi K_S^0$

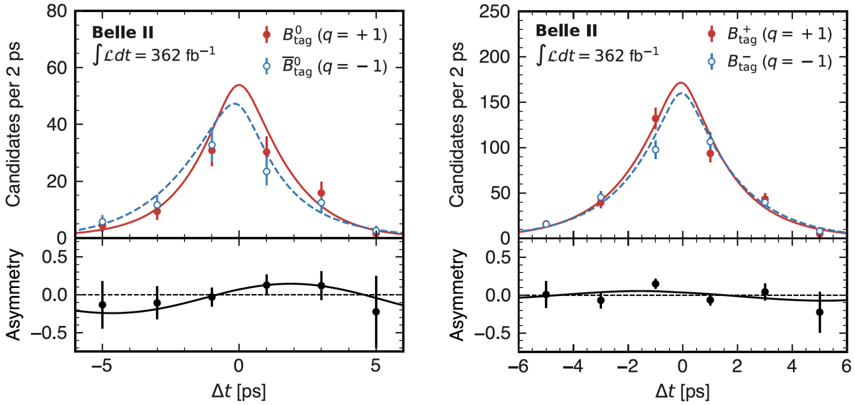


FIG. 3. Distributions, and fit projections, of Δt for flavor-tagged (left) $B^0 \to \phi K_S^0$ and (right) $B^+ \to \phi K^+$ candidates subtracted of the continuum background. The fit PDFs corresponding to $q = -1$ and $q = +1$ tagged distributions are shown as dashed and solid curves, respectively. The yield asymmetries, defined as $(N(q = +1) - N(q = -1))/(N(q = +1) + N(q = -1))$, are displayed in the bottom subpanels.

Calibration of technique: analysis chain

Belle II pub20: PRD 108, 072012 (2023); CP asymmetries in $B^0 \to \phi K_S^0$

tagging algorithm and Δt resolution calibrated with $B^0 \to D^{*-}\pi^+$ decays

- this has its own uncertainties

We assess the uncertainty associated with the resolution function and flavor tagging parameters using simplified simulated samples. We generate ensembles assuming for each an alternative value for the above parameters sampled from the statistical covariance matrix determined in the $B^0 \to D^{(*)-}\pi^+$ control sample. Each ensemble is fitted using the nominal values of the calibration parameters and the standard deviation of the observed biases is used as a systematic uncertainty.

A similar procedure is used to assess a systematic uncertainty due to the systematic uncertainties on the calibration parameters, in which the ensembles are generated by varying each parameter independently within their systematic uncertainty.

TABLE II. Summary of systematic uncertainties.

Source	$\sigma(C)$	$\sigma(S)$
Calibration with $B^0 \to D^{(*)-}\pi^+$ decays		
Calibration sample size	±0.010	±0.009
Calibration sample systematic	±0.010	±0.012
Sample dependence	$+0.005$	$+0.021$
Fit model		
Fit bias	${}^{+0.028}_{-0.017}$	${}^{+0.033}_{-0.062}$
$B^0 \to K^+K^-K_S^0$ backgrounds	$+0.020$	-0.011
Fixed fit shapes	±0.009	±0.022
τ_{B^0} and Δm_d	±0.006	±0.022
$C_{K^+K^-K_S^0}$ and $S_{K^+K^-K_S^0}$	±0.014	±0.013
$B\bar{B}$ background asymmetry	${}^{+0.019}_{-0.030}$	${}^{+0.017}_{-0.031}$
Tag-side interference	<0.001	$+0.012$
Candidate selection	-0.032	-0.002
Δt measurement		
Tracker misalignment	-0.002	-0.002
Momentum scale	±0.001	±0.001
Beam spot	±0.002	±0.002
Δt approximation	<0.001	-0.018
Total systematic	${}^{+0.046}_{-0.052}$	${}^{+0.058}_{-0.082}$
Statistical	±0.201	±0.256

Calibration of technique: analysis chain

Belle II pub20: PRD 108, 072012 (2023); CP asymmetries in $B^0 \to \phi K_S^0$

tagging algorithm and Δt resolution calibrated with $B^0 \to D^{*-}\pi^+$ decays

- this has its own uncertainties
- it has to be ported to ϕK_S^0

We estimate the impact of differences in the resolution function and tagging performance between the signal and calibration samples. We apply the resolution function and flavor-tagging calibration obtained from a simulated $B^0 \to D^{(*)-}\pi^+$ sample and repeat the measurement of C and S over an ensemble of simulated $B^0 \to \phi K_S^0$ events. The average deviation of the CP asymmetries from their generated values is assigned as a systematic uncertainty.

TABLE II. Summary of systematic uncertainties.

Source	$\sigma(C)$	$\sigma(S)$
Calibration with $B^0 \to D^{(*)-}\pi^+$ decays		
Calibration sample size	± 0.010	± 0.009
Calibration sample systematic	± 0.010	± 0.012
Sample dependence	$+0.005$	$+0.021$
Fit model		
Fit bias	$^{+0.028}_{-0.017}$	$^{+0.033}_{-0.062}$
$B^0 \to K^+K^-K_S^0$ backgrounds	$+0.020$	-0.011
Fixed fit shapes	± 0.009	± 0.022
τ_{B^0} and Δm_d	± 0.006	± 0.022
$C_{K^+K^-K_S^0}$ and $S_{K^+K^-K_S^0}$	± 0.014	± 0.013
$B\bar{B}$ background asymmetry	$^{+0.019}_{-0.030}$	$^{+0.017}_{-0.031}$
Tag-side interference	<0.001	$+0.012$
Candidate selection	-0.032	-0.002
Δt measurement		
Tracker misalignment	-0.002	-0.002
Momentum scale	± 0.001	± 0.001
Beam spot	± 0.002	± 0.002
Δt approximation	<0.001	-0.018
Total systematic	$^{+0.046}_{-0.052}$	$^{+0.058}_{-0.082}$
Statistical	± 0.201	± 0.256

Calibration of technique: analysis chain

Belle II pub20: PRD 108, 072012 (2023); CP asymmetries in $B^0 \to \phi K_S^0$

tagging algorithm and Δt resolution calibrated with $B^0 \to D^{*-}\pi^+$ decays

- this has its own uncertainties
- it has to be ported to ϕK_S^0

We estimate the impact of differences in the resolution function and tagging performance between the signal and calibration samples. We apply the resolution function and flavor-tagging calibration obtained from a simulated $B^0 \to D^{(*)-}\pi^+$ sample and repeat the measurement of C and S over an ensemble of simulated $B^0 \to \phi K_S^0$ events. The average deviation of the CP asymmetries from their generated values is assigned as a systematic uncertainty.

- procedure is a variant of the *control and validation region* studies beloved of the LHC

TABLE II. Summary of systematic uncertainties.

Source	$\sigma(C)$	$\sigma(S)$
Calibration with $B^0 \to D^{(*)-}\pi^+$ decays		
Calibration sample size	± 0.010	± 0.009
Calibration sample systematic	± 0.010	± 0.012
Sample dependence	$+0.005$	$+0.021$
Fit model		
Fit bias	$^{+0.028}_{-0.017}$	$^{+0.033}_{-0.062}$
$B^0 \to K^+K^-K_S^0$ backgrounds	$+0.020$	-0.011
Fixed fit shapes	± 0.009	± 0.022
τ_{B^0} and Δm_d	± 0.006	± 0.022
$C_{K^+K^-K_S^0}$ and $S_{K^+K^-K_S^0}$	± 0.014	± 0.013
$B\bar{B}$ background asymmetry	$^{+0.019}_{-0.030}$	$^{+0.017}_{-0.031}$
Tag-side interference	<0.001	$+0.012$
Candidate selection	-0.032	-0.002
Δt measurement		
Tracker misalignment	-0.002	-0.002
Momentum scale	± 0.001	± 0.001
Beam spot	± 0.002	± 0.002
Δt approximation	<0.001	-0.018
Total systematic	$^{+0.046}_{-0.052}$	$^{+0.058}_{-0.082}$
Statistical	± 0.201	± 0.256

CALIBRATION . . . *of the analyst:* your own choices

I went forward in time, to view all possible ways we might conduct the analysis.
How many did you see? 14,000,605.
How many gave us the right answer? 1.

Calibration of choices: fitting region & other choices

Belle pub632: PRD 107, 072008 (2023); $e^+e^- \to \Sigma\bar{\Sigma}$ via ISR

The (low) background is estimated using sidebands. But, chosen how?

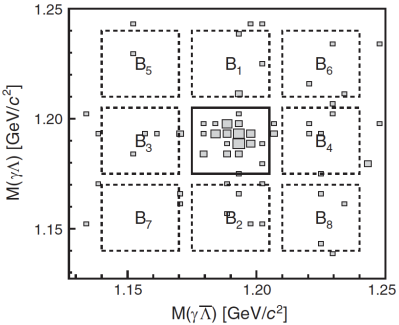


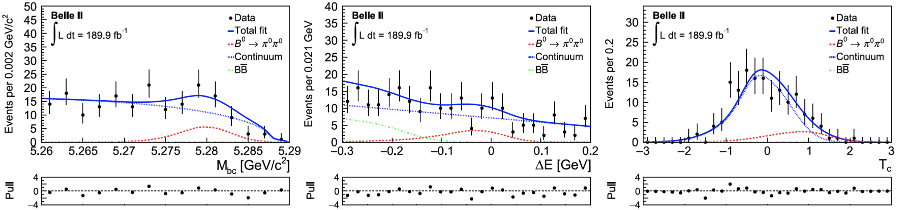
TABLE IV. Summary of systematic uncertainties for the $e^+e^- \to \gamma_{\rm ISR}\Sigma^0\bar{\Sigma}^0$ cross section measurement.

Source	Systematic uncertainty (%)
Tracking	1.4
PID	2.7
Λ reconstruction	5.4
$\Sigma^0/\bar{\Sigma}^0$ mass resolution	0.6
Sideband method	6
$e^+e^- \to \gamma_{\rm ISR}\Sigma^0\bar{\Sigma}^0\pi^0$ background	3–9
Other two $\Sigma^0/\bar{\Sigma}^0$ background	2–6
Integrated luminosity	1.4
ISR emission probability	1
PHOKHARA simulation	1
$\Lambda \to p\pi^-$ branching fraction	0.8
Modeling of angular dependence	3–5
Modeling of energy dependence	1–5
Trigger	3
The fit to efficiency	1
Sum in quadrature	11–16

- try larger, smaller, shifted choices of sidebands
- leads to noticeable yield changes: the co-leading systematic at 6%

Calibration of choices: parameterizations

Belle II pub13: PRD 107, 112009 (2023); $\pi^0\pi^0$ BF and $\mathcal{A}_{CP}$



- reminder: 3D signal peak over $B\overline{B}$ and continuum background
- previously: how well do we know the params of the continuum shape?
- now consider: *choices* were also made in modelling the signal: MC-based KDE in $(M_{bc}, \Delta E)$; what if another shape had been used?
- uncorrelated product of CB functions tried as alternative: 1.3% effect, minor *cf.* 16.2% total systs, but in other cases the effect can be larger
- esp. w limited samples, & limited or absent controls, no assumption-free way to make such choices — so an uncertainty is appropriate

Calibration of choices

- fitting region choices
- parameterizations

In both these cases it is very hard to claim that the default choices are “inevitable” or obviously correct.

Older analyses sometimes include such uncertainties due to

- object selection (how many SVD or CDC hits)
- event selection cuts
- other exact cut values
- dependence on any other choice that has an arbitrary element

This has gone a bit out of style — because where do you stop? — and also for a technical reason (see later). But the basic idea is sound.

CALIBRATION . . . by searching for mistakes

- biases: signal of $x \longrightarrow$ mean measurement of $x + \delta$
- instabilities: [butterfly-wingbeat] $\longrightarrow$ measured yield changes
- mis-classifications and omitted categories:
  let's say, the $t\bar{t}$ bkgd has *two components with different behaviour under cuts* & differing $m_{4\ell}$ distribut^{ns}, unresolved by control samples
  - with $2\times$ the sample, if careful, we'll notice problems in fitting
  - with $4\times$ the sample, there will be clear and nasty discrepancies

  That was a made-up example, but the phenomenon is very real: Belle $D^0 - \overline{D}^0$ mixing $(D \to K\pi)$ took one year per doubling in sample, to refine method enough to keep systematics under control
- misunderstanding the rel^{n}ship between auxiliary & principal meas^{ts}
- unknown unknowns

All are problems that should be fixed: the corresponding systematics are

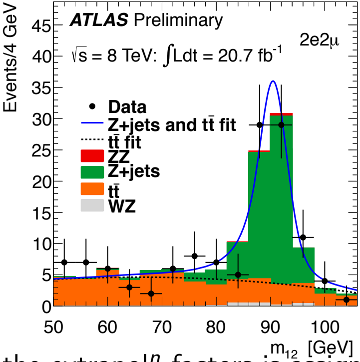
- estimates (guesses?) of possible residual problems
- the tolerances of the tests and cross-checks used . . .

Mistakes? (1) e.g. Relationship with the auxiliaries

ATLAS–CONF–2013–013; Higgs properties in $H \to ZZ^{(*)}* \to 4\ell$

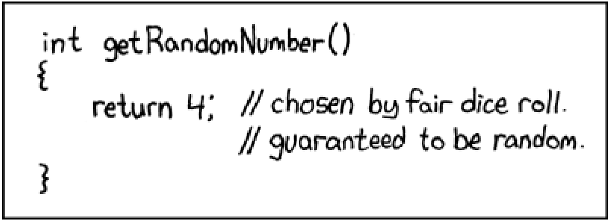
control region 2 results for $H \to \ell\ell + \mu\mu$ are *extrapolated* to the signal region using IP signif. & isol^n requirement efficiencies from $Zb\bar{b}$ MC:

- what if this is wrong?
- efficiency validated with *another* control region, requiring $Z + \mu$
- test fails? : stop and try to
  - gain understanding, then
  - fix problem if possible, else
  - back up and change method
- many analyses have dead-ends & side-branches, documented or not



- test succeeded! : a 10% uncert^y on the extrapol^n factors is assigned
- note this "data driven" bkgd estimate has embedded dependence on {MC, physical insight, expert judgement on validation, rules of art}

Mistakes? (2) coding and other bugs



Mistakes? (3) unknown unknowns

> Reports that say that something hasn't happened are always interesting to me, because as we know, there are known knowns; there are things we know that we know. There are known unknowns; that is to say, there are things that we now know we don't know. But there are also unknown unknowns — there are things we do not know we don't know.

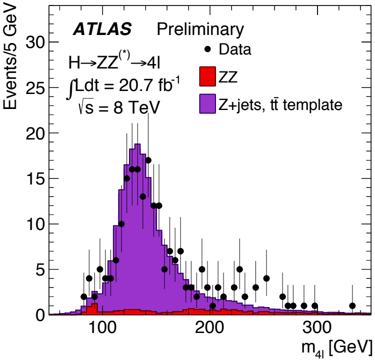
- Donald Rumsfeld got mocked in the media for this, but he had a point
- *e.g.* "analysis-level" information is vulnerable to subtle problems:
  - what if there is a distinction in response in a drift chamber with stereo layers, never spotted because one must compare the response of $\{U, V\}$ wires to $+$ve and $-$ve tracks going forward and backward?
  - what if out-of-spill calorimeter clusters give a decayed but measurable response, not tagged in analysis-level data (& some are back-to-back)?
- one *builds confidence* with a new {detector, code, technique} by doing basic & known things first (*e.g.* so-called "rediscovery" analyses)
- and one relies on ...

Mistakes? (4) cross-checks

ATLAS–CONF–2013–013; Higgs properties in $H \to ZZ^{(*)} \to 4\ell$

additional studies that do not contribute directly to any of the bkgd measurements or systematics directly, but are there to spot problems:

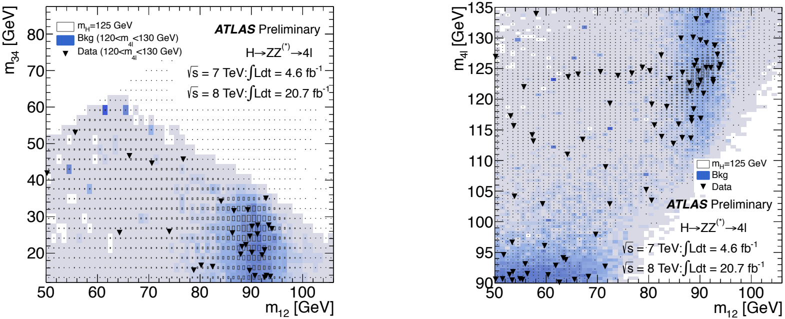
- the $t\bar{t}$ bkgd measurement for $H \to \ell\ell + \mu\mu$ is cross-checked using a $e^{\pm}\mu^{\mp} + \mu^{+}\mu^{-}$ sample, with $M(e^{\pm}\mu^{\mp}) \in (50, 106)\,\mathrm{GeV}$
- the $ZZ^{(*)}$ signal and the $Z +$ jets and $t\bar{t}$ bkgds are checked in another control region:
- agreement is not bad, but is imperfect — how much does this matter?



Mistakes? (4) cross-checks

ATLAS–CONF–2013–013; Higgs properties in $H \to ZZ^{(*)} \to 4\ell$

- the final analysis uses a *look-back plot* to check the $(m_{12}, m_{34}, m_{4\ell})$ distribution *in the absence of the Z-mass constraint*:



- your {advisor, RC, journal referee} may ask you for such plots
- "But what are you looking for?"
- "I don't know, but I may know it when I see it . . ."

Roundup of methods: (1) profile likelihood

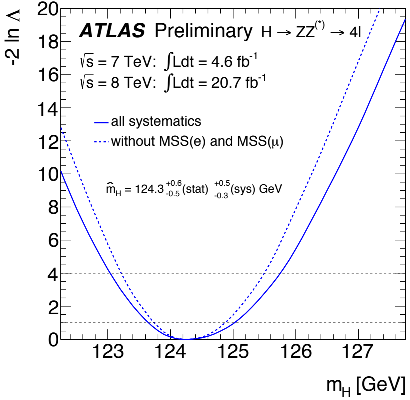
consider

- a quantity of interest μ (signal strength, mass, ...)
- quantities $\boldsymbol{\theta}$, say $\boldsymbol{\theta} = (\boldsymbol{\theta}_s, \boldsymbol{\theta}_b, n_b)$ governing shape & bkgd norm^n

- form the profile $\mathcal{L}$ ratio

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})}$$

- $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ are ML estimators
- $\hat{\hat{\boldsymbol{\theta}}}(\mu)$ is the best estimate of $\boldsymbol{\theta}$ for *the given* μ *value*
- letting "nuisance parameters" $\boldsymbol{\theta}$ float at each μ will
  - improve the $\mathcal{L}$ there, and
  - broaden the distribution



- a.k.a. "the `MINUIT` method", long used intuitively in HEP

Roundup of methods: (2) Gaussian constraint

ATLAS ATL-COM-PHYS-2012-089; $t\bar{t}$, WW, $Z/\gamma^* \to \tau\tau$

$$\mathcal{L}_i(N_{sig}, \alpha_j) = \mathcal{P}_i\left(N^{obs} \,|\, N^{exp}(N_{sig},\, \alpha_j)\right) \prod_{j\in\text{syst}} \mathcal{G}(\alpha_j \,|\, 0,\, 1)$$

heavily used at ATLAS:

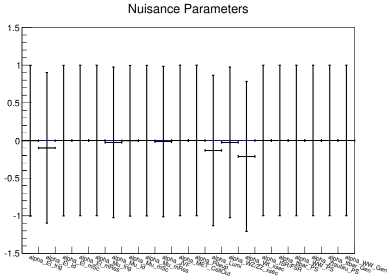
- trigger efficiency officially measured, say $\epsilon = 0.25^{+0.03}_{-0.02}$ at $y = 0$ & $p_T = 20\,\text{GeV}$, and rising with energy
- ideally, scale for the dist^n represented by some f^n $f(x)$
- approximate as $\mathcal{G}(x \,|\, \mu,\, \sigma)$
- transform to $\alpha = (x - \mu)/\sigma$
- now expressed as $\mathcal{G}(\alpha \,|\, 0,\, 1)$

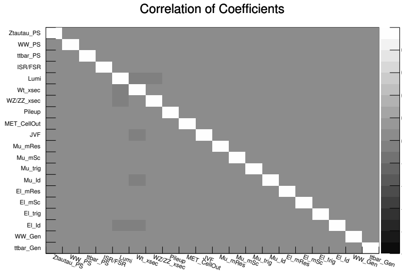
Fit Parameter α	Parameter Value	Symmetric Uncertainty	Asymmetric Uncertainty	
Electron trigger	0.00	1.00	-1.00	1.00
Electron reconstruction, ID, isolation	-0.10	1.00	-1.00	1.00
Electron momentum scale	0.00	1.00	-1.00	1.00
Electron momentum resolution	0.00	1.00	-1.00	1.00
Muon reconstruction, ID, isolation	-0.03	1.00	-1.00	1.00
Muon trigger	0.00	1.00	-1.00	1.00
Muon momentum scale	0.00	1.00	-1.00	1.00
Muon momentum resolution	0.00	1.00	-1.00	1.00
Jet vertex fraction	-0.02	1.00	-1.00	1.00
E_T^{miss} Cell-out	0.00	1.00	-1.00	1.00
Pileup	0.00	1.00	-1.00	1.00
WZ/ZZ cross sections	-0.03	1.00	-1.00	1.00
Wt cross section	-0.21	0.99	-1.00	1.00
Luminosity	-0.06	1.00	-1.00	1.00
LHC beam energy	0.00	1.00	-1.00	1.00
ISR/FSR	0.00	1.00	-1.00	1.00
$t\bar{t}$ Parton Shower	0.00	1.00	-1.00	1.00
WW Parton Shower	0.00	1.00	-1.00	1.00
$Z \to \tau\tau$ Parton Shower	0.00	1.00	-1.00	1.00
$t\bar{t}$ Generator	0.00	1.00	-1.00	1.00
WW Generator	0.00	1.00	-1.00	1.00

Roundup of methods: (2) Gaussian constraint

ATLAS ATL-COM-PHYS-2012-089; $t\bar{t}$, WW, $Z/\gamma^* \to \tau\tau$

- provides a standard system to check for *strain* in the fit
- if an analysis had JES 2.5σ below standard, that would be suspicious
- if there was a high correlation between WZ/ZZ x-section and ISR/FSR systematics, that would be suspicious





- note that profile $\mathcal{L}$ and $\mathcal{G}$ constraint differ only the packaging; the essence of both techniques is simultaneous fitting

Issues, questions, and tricks of the trade

- note that biases must first be corrected!
- do we know what the probability distribution is?
- cancellation *versus* partial cancellation
- rough but conservative estimates (easier/quicker than exact ones, and if the effect still turns out to be small, the exact estimate can be safely skipped)
- possible effects versus proven effects, or, against the 3σ standard
- double-counting of statistical effects;
  pitfalls in calculation of differences between overlapping selections
- separating out a large / distinctive systematic term: *e.g.* f^{00} in $B \to K\pi$, $\pi\pi$
- every conceivable effect? every likely/possible effect?
- "two point uncertainties" on complex models are a matter of expert judgement, esp. if large (*cf.* the small FF uncertainties mentioned above)
- combination of symmetric and asymmetric uncertainties produces a bias (*shifts the central value*) and has a nontrivial residual uncertainty, in general (example: Belle II pub31, PRL 131, 171803 (2023); D_s^+ lifetime)

Frequentist & Bayesian Probability, and coin-tossing

Frequentist: I gotta tell you, he's not really my friend.
Saving his life is more a professional courtesy.

Frequentist: What is your job . . . ?
Bayesian: Protecting your reality . . . !

Frequentist and Bayesian Probability

What is the definition of the conditional probability $P(A\,|\,B)$?

Frequentist (classical) Probability:

$P(A|B) =$ long-run relative frequency of A occuring in identical repetitions of an observation, under some conditions B;

A is restricted to propositions about random variables

Frequentist and Bayesian Probability

What is the definition of the conditional probability $P(A\,|\,B)$?

Bayesian Probability:

$P(A|B)$ is a real-number measure of the plausibility of proposition A, given (conditional upon) the truth of proposition B;

P measures degree of belief;

A can be any logical proposition

Application to tossing a coin

- suppose I stand to win or lose money in a game of chance

Application to tossing a coin

- suppose I stand to win or lose money in a game of chance
- my companion gives me a coin to use in the game

Application to tossing a coin

- suppose I stand to win or lose money in a game of chance
- my companion gives me a coin to use in the game
- do I trust the coin?

Application to tossing a coin

- suppose I stand to win or lose money in a game of chance
- my companion gives me a coin to use in the game
- do I trust the coin?
- to get some information, I toss the coin five times:

Application to tossing a coin

- suppose I stand to win or lose money in a game of chance
- my companion gives me a coin to use in the game
- do I trust the coin?
- to get some information, I toss the coin five times:
    - get 4 heads, 1 tail (4H,1T)

Application to tossing a coin

- suppose I stand to win or lose money in a game of chance
- my companion gives me a coin to use in the game
- do I trust the coin?
- to get some information, I toss the coin five times:
    - get 4 heads, 1 tail (4H,1T)
    - do I trust the coin now? what have I learned?

Application to tossing a coin

- suppose I stand to win or lose money in a game of chance
- my companion gives me a coin to use in the game
- do I trust the coin?
- to get some information, I toss the coin five times:
  - get 4 heads, 1 tail (4H,1T)
  - do I trust the coin now? what have I learned?
  - is my answer any different if I toss the coin ten times, and get 8 heads, 2 tails (8H,2T)?

Application to tossing a coin

- suppose I stand to win or lose money in a game of chance
- my companion gives me a coin to use in the game
- do I trust the coin?
- to get some information, I toss the coin five times:
  - get 4 heads, 1 tail (4H,1T)
  - do I trust the coin now? what have I learned?
  - is my answer any different if I toss the coin ten times, and get 8 heads, 2 tails (8H,2T)?
- this simple problem incorporates all sorts of statistical questions:

Application to tossing a coin

- suppose I stand to win or lose money in a game of chance
- my companion gives me a coin to use in the game
- do I trust the coin?
- to get some information, I toss the coin five times:
  - get 4 heads, 1 tail (4H,1T)
  - do I trust the coin now? what have I learned?
  - is my answer any different if I toss the coin ten times, and get 8 heads, 2 tails (8H,2T)?
- this simple problem incorporates all sorts of statistical questions:
  - simple (fair coin) *versus* compound ($\{$unfair coin; $p_H\}$)

Application to tossing a coin

- suppose I stand to win or lose money in a game of chance
- my companion gives me a coin to use in the game
- do I trust the coin?
- to get some information, I toss the coin five times:
  - get 4 heads, 1 tail (4H,1T)
  - do I trust the coin now? what have I learned?
  - is my answer any different if I toss the coin ten times, and get 8 heads, 2 tails (8H,2T)?
- this simple problem incorporates all sorts of statistical questions:
  - simple (fair coin) *versus* compound ({unfair coin; p_H})

  cf. simple (bkgd only) *versus* compound ($\{H; m_H\}$)

Application to tossing a coin

- suppose I stand to win or lose money in a game of chance
- my companion gives me a coin to use in the game
- do I trust the coin?
- to get some information, I toss the coin five times:
  - get 4 heads, 1 tail (4H,1T)
  - do I trust the coin now? what have I learned?
  - is my answer any different if I toss the coin ten times, and get 8 heads, 2 tails (8H,2T)?
- this simple problem incorporates all sorts of statistical questions:
  - simple (fair coin) *versus* compound ($\{$unfair coin; $p_H\}$)

  cf. simple (bkgd only) *versus* compound ($\{H; m_H\}$)
  - data has limited statistical power

Application to tossing a coin

- suppose I stand to win or lose money in a game of chance
- my companion gives me a coin to use in the game
- do I trust the coin?
- to get some information, I toss the coin five times:
  - get 4 heads, 1 tail (4H,1T)
  - do I trust the coin now? what have I learned?
  - is my answer any different if I toss the coin ten times, and get 8 heads, 2 tails (8H,2T)?
- this simple problem incorporates all sorts of statistical questions:
  - simple (fair coin) *versus* compound ($\{$unfair coin; $p_H\}$)
  - ***cf.*** simple (bkgd only) *versus* compound ($\{H; m_H\}$)
  - data has limited statistical power
  - neatly illustrates how the method and interpretation depend on the *statistical framework* used

Tossing a coin: frequentist treatment

What is the definition of the conditional probability $P(A\,|\,B)$?

Frequentist (classical) Probability:

$P(A|B) =$ long-run relative frequency of A occuring in identical repetitions of an observation, under some conditions B;

A is restricted to propositions about random variables

Tossing a coin: frequentist treatment

p_H	$0H$ $5T$	$1H$ $4T$	$2H$ $3T$	$3H$ $2T$	$4H$ $1T$	$5H$ $0T$
1.00	.00000	.00000	.0000	.0000	.00000	1.0
0.90	.00001	.00045	.0081	.0729	.32805	.59049
0.80	.00032	.00640	.0512	.2048	.4096	.32768
0.70	.00243	.02835	.1323	.3087	.36015	.16807
0.60	.01024	.00768	.2304	.3456	.2592	.07776
0.50	.03125	.15625	.3125	.3125	.15625	.03125
0.40	.07776	.2592	.3456	.2304	.00768	.01024
0.30	.16807	.36015	.3087	.1323	.02835	.00243
0.20	.32768	.4096	.2048	.0512	.00640	.00032
0.10	.59049	.32805	.0729	.0081	.00045	.00001
0.00	1.0	.00000	.00000	.00000	.00000	.00000

treatment: $\forall p_H$, form 90% acceptance band

p_H	$0H$ $5T$	$1H$ $4T$	$2H$ $3T$	$3H$ $2T$	$4H$ $1T$	$5H$ $0T$
1.00	.00000	.00000	.0000	.0000	.00000	1.0
0.90	.00001	.00045	.0081	.0729	.32805	.59049
0.80	.00032	.00640	.0512	.2048	.4096	.32768
0.70	.00243	.02835	.1323	.3087	.36015	.16807
0.60	.01024	.00768	.2304	.3456	.2592	.07776
0.50	.03125	.15625	.3125	.3125	.15625	.03125
0.40	.07776	.2592	.3456	.2304	.00768	.01024
0.30	.16807	.36015	.3087	.1323	.02835	.00243
0.20	.32768	.4096	.2048	.0512	.00640	.00032
0.10	.59049	.32805	.0729	.0081	.00045	.00001
0.00	1.0	.00000	.00000	.00000	.00000	.00000

treatment: $\forall p_H$, form 90% acceptance band

p_H	$0H$ $5T$	$1H$ $4T$	$2H$ $3T$	$3H$ $2T$	$4H$ $1T$	$5H$ $0T$
1.00	.00000	.00000	.0000	.0000	.00000	1.0
0.90	.00001	.00045	.0081	.0729	.32805	.59049
0.80	.00032	.00640	.0512	.2048	.4096	.32768
0.70	.00243	.02835	.1323	.3087	.36015	.16807
0.60	.01024	.00768	.2304	.3456	.2592	.07776
0.50	.03125	.15625	.3125	.3125	.15625	.03125
0.40	.07776	.2592	.3456	.2304	.00768	.01024
0.30	.16807	.36015	.3087	.1323	.02835	.00243
0.20	.32768	.4096	.2048	.0512	.00640	.00032
0.10	.59049	.32805	.0729	.0081	.00045	.00001
0.00	1.0	.00000	.00000	.00000	.00000	.00000

treatment: $\forall p_H$, form 90% acceptance band

p_H	$0H$ $5T$	$1H$ $4T$	$2H$ $3T$	$3H$ $2T$	$4H$ $1T$	$5H$ $0T$
1.00	.00000	.00000	.0000	.0000	.00000	1.0
0.90	.00001	.00045	.0081	.0729	.32805	.59049
0.80	.00032	.00640	.0512	.2048	.4096	.32768
0.70	.00243	.02835	.1323	.3087	.36015	.16807
0.60	.01024	.00768	.2304	.3456	.2592	.07776
0.50	.03125	.15625	.3125	.3125	.15625	.03125
0.40	.07776	.2592	.3456	.2304	.00768	.01024
0.30	.16807	.36015	.3087	.1323	.02835	.00243
0.20	.32768	.4096	.2048	.0512	.00640	.00032
0.10	.59049	.32805	.0729	.0081	.00045	.00001
0.00	1.0	.00000	.00000	.00000	.00000	.00000

treatment: $\forall p_H$, form 90% acceptance band

p_H	0H 5T	1H 4T	2H 3T	3H 2T	4H 1T	5H 0T
1.00	.00000	.00000	.0000	.0000	.00000	1.0
0.90	.00001	.00045	.0081	.0729	.32805	.59049
0.80	.00032	.00640	.0512	.2048	.4096	.32768
0.70	.00243	.02835	.1323	.3087	.36015	.16807
0.60	.01024	.00768	.2304	.3456	.2592	.07776
0.50	.03125	.15625	.3125	.3125	.15625	.03125
0.40	.07776	.2592	.3456	.2304	.00768	.01024
0.30	.16807	.36015	.3087	.1323	.02835	.00243
0.20	.32768	.4096	.2048	.0512	.00640	.00032
0.10	.59049	.32805	.0729	.0081	.00045	.00001
0.00	1.0	.00000	.00000	.00000	.00000	.00000

treatment: $\forall p_H$, form 90% acceptance band

p_H	$0H$ $5T$	$1H$ $4T$	$2H$ $3T$	$3H$ $2T$	$4H$ $1T$	$5H$ $0T$
1.00	.00000	.00000	.0000	.0000	.00000	1.0
0.90	.00001	.00045	.0081	.0729	.32805	.59049
0.80	.00032	.00640	.0512	.2048	.4096	.32768
0.70	.00243	.02835	.1323	.3087	.36015	.16807
0.60	.01024	.00768	.2304	.3456	.2592	.07776
0.50	.03125	.15625	.3125	.3125	.15625	.03125
0.40	.07776	.2592	.3456	.2304	.00768	.01024
0.30	.16807	.36015	.3087	.1323	.02835	.00243
0.20	.32768	.4096	.2048	.0512	.00640	.00032
0.10	.59049	.32805	.0729	.0081	.00045	.00001
0.00	1.0	.00000	.00000	.00000	.00000	.00000

treatment: $\forall p_H$, form 90% acceptance band

p_H	$0H$ $5T$	$1H$ $4T$	$2H$ $3T$	$3H$ $2T$	$4H$ $1T$	$5H$ $0T$
1.00	.00000	.00000	.0000	.0000	.00000	1.0
0.90	.00001	.00045	.0081	.0729	.32805	.59049
0.80	.00032	.00640	.0512	.2048	.4096	.32768
0.70	.00243	.02835	.1323	.3087	.36015	.16807
0.60	.01024	.00768	.2304	.3456	.2592	.07776
0.50	.03125	.15625	.3125	.3125	.15625	.03125
0.40	.07776	.2592	.3456	.2304	.00768	.01024
0.30	.16807	.36015	.3087	.1323	.02835	.00243
0.20	.32768	.4096	.2048	.0512	.00640	.00032
0.10	.59049	.32805	.0729	.0081	.00045	.00001
0.00	1.0	.00000	.00000	.00000	.00000	.00000

treatment: $\forall p_H$, form 90% acceptance band

p_H	$0H$ $5T$	$1H$ $4T$	$2H$ $3T$	$3H$ $2T$	$4H$ $1T$	$5H$ $0T$
1.00	.00000	.00000	.0000	.0000	.00000	1.0
0.90	.00001	.00045	.0081	.0729	.32805	.59049
0.80	.00032	.00640	.0512	.2048	.4096	.32768
0.70	.00243	.02835	.1323	.3087	.36015	.16807
0.60	.01024	.00768	.2304	.3456	.2592	.07776
0.50	.03125	.15625	.3125	.3125	.15625	.03125
0.40	.07776	.2592	.3456	.2304	.00768	.01024
0.30	.16807	.36015	.3087	.1323	.02835	.00243
0.20	.32768	.4096	.2048	.0512	.00640	.00032
0.10	.59049	.32805	.0729	.0081	.00045	.00001
0.00	1.0	.00000	.00000	.00000	.00000	.00000

treatment: $\forall p_H$, form 90% acceptance band

p_H	0H 5T	1H 4T	2H 3T	3H 2T	4H 1T	5H 0T
1.00	.00000	.00000	.0000	.0000	.00000	1.0
0.90	.00001	.00045	.0081	.0729	.32805	.59049
0.80	.00032	.00640	.0512	.2048	.4096	.32768
0.70	.00243	.02835	.1323	.3087	.36015	.16807
0.60	.01024	.00768	.2304	.3456	.2592	.07776
0.50	.03125	.15625	.3125	.3125	.15625	.03125
0.40	.07776	.2592	.3456	.2304	.00768	.01024
0.30	.16807	.36015	.3087	.1323	.02835	.00243
0.20	.32768	.4096	.2048	.0512	.00640	.00032
0.10	.59049	.32805	.0729	.0081	.00045	.00001
0.00	1.0	.00000	.00000	.00000	.00000	.00000

treatment: $\forall p_H$, form 90% acceptance band

p_H	$0H$ $5T$	$1H$ $4T$	$2H$ $3T$	$3H$ $2T$	$4H$ $1T$	$5H$ $0T$
1.00	.00000	.00000	.0000	.0000	.00000	1.0
0.90	.00001	.00045	.0081	.0729	.32805	.59049
0.80	.00032	.00640	.0512	.2048	.4096	.32768
0.70	.00243	.02835	.1323	.3087	.36015	.16807
0.60	.01024	.00768	.2304	.3456	.2592	.07776
0.50	.03125	.15625	.3125	.3125	.15625	.03125
0.40	.07776	.2592	.3456	.2304	.00768	.01024
0.30	.16807	.36015	.3087	.1323	.02835	.00243
0.20	.32768	.4096	.2048	.0512	.00640	.00032
0.10	.59049	.32805	.0729	.0081	.00045	.00001
0.00	1.0	.00000	.00000	.00000	.00000	.00000

treatment: 90% conf. interval: $\{p_H | (4H, 1T) \in A_{90}(p_H)\}$

p_H	$0H$ $5T$	$1H$ $4T$	$2H$ $3T$	$3H$ $2T$	$4H$ $1T$	$5H$ $0T$
1.00	.00000	.00000	.0000	.0000	.00000	1.0
0.90	.00001	.00045	.0081	.0729	.32805	.59049
0.80	.00032	.00640	.0512	.2048	.4096	.32768
0.70	.00243	.02835	.1323	.3087	.36015	.16807
0.60	.01024	.00768	.2304	.3456	.2592	.07776
0.50	.03125	.15625	.3125	.3125	.15625	.03125
0.40	.07776	.2592	.3456	.2304	.00768	.01024
0.30	.16807	.36015	.3087	.1323	.02835	.00243
0.20	.32768	.4096	.2048	.0512	.00640	.00032
0.10	.59049	.32805	.0729	.0081	.00045	.00001
0.00	1.0	.00000	.00000	.00000	.00000	.00000

Tossing a coin: frequentist treatment — commentary

- we asked ourselves whether the coin was fair,
  but we answered "like scientists", with technical stuff:
  can't interpret this easily in street terms

Tossing a coin: frequentist treatment — commentary

- we asked ourselves whether the coin was fair,
  but we answered "like scientists", with technical stuff:
  can't interpret this easily in street terms
- **this is a feature,** not a bug, of frequentist statistics:
  the frequentist concept of probability is limiting frequency;

Tossing a coin: frequentist treatment — commentary

- we asked ourselves whether the coin was fair,
  but we answered "like scientists", with technical stuff:
  can't interpret this easily in street terms
- **this is a feature,** not a bug, of frequentist statistics:
  the frequentist concept of probability is limiting frequency;
  - the coin has some probability p_H to turn up heads,
    which we can estimate with increasing precision as $N \to \infty$

Tossing a coin: frequentist treatment — commentary

- we asked ourselves whether the coin was fair,
  but we answered "like scientists", with technical stuff:
  can't interpret this easily in street terms
- **this is a feature,** not a bug, of frequentist statistics:
  the frequentist concept of probability is limiting frequency;
  - the coin has some probability p_H to turn up heads,
    which we can estimate with increasing precision as $N \to \infty$
  - the $p_H = 0.5$ hypothesis itself has no probability ...

Tossing a coin: frequentist treatment — commentary

- we asked ourselves whether the coin was fair,
  but we answered "like scientists", with technical stuff:
  can't interpret this easily in street terms
- **this is a feature,** not a bug, of frequentist statistics:
  the frequentist concept of probability is limiting frequency;
  - the coin has some probability p_H to turn up heads,
    which we can estimate with increasing precision as $N \to \infty$
  - the $p_H = 0.5$ hypothesis itself has no probability ...
- **the benefit:**
  this has been a straightforward (if tedious) calculation,
  and our opinions / beliefs / prejudices have not been involved

Tossing a coin: frequentist treatment — commentary

- we asked ourselves whether the coin was fair,
  but we answered "like scientists", with technical stuff:
  can't interpret this easily in street terms
- **this is a feature,** not a bug, of frequentist statistics:
  the frequentist concept of probability is limiting frequency;
  - the coin has some probability p_H to turn up heads,
    which we can estimate with increasing precision as $N \to \infty$
  - the $p_H = 0.5$ hypothesis itself has no probability ...
- **the benefit:**
  this has been a straightforward (if tedious) calculation,
  and our opinions / beliefs / prejudices have not been involved
- **the costs:**

Tossing a coin: frequentist treatment — commentary

- we asked ourselves whether the coin was fair,
  but we answered "like scientists", with technical stuff:
  can't interpret this easily in street terms
- **this is a feature,** not a bug, of frequentist statistics:
  the frequentist concept of probability is limiting frequency;
  - the coin has some probability p_H to turn up heads,
    which we can estimate with increasing precision as $N \to \infty$
  - the $p_H = 0.5$ hypothesis itself has no probability ...
- **the benefit:**
  this has been a straightforward (if tedious) calculation,
  and our opinions / beliefs / prejudices have not been involved
- **the costs:**
  1. we have not directly addressed the q^n, *"is this a fair coin?"*,
     that we were originally interested in

Tossing a coin: frequentist treatment — commentary

- we asked ourselves whether the coin was fair,
  but we answered "like scientists", with technical stuff:
  can't interpret this easily in street terms
- **this is a feature,** not a bug, of frequentist statistics:
  the frequentist concept of probability is limiting frequency;
  - the coin has some probability p_H to turn up heads,
    which we can estimate with increasing precision as $N \to \infty$
  - the $p_H = 0.5$ hypothesis itself has no probability ...
- **the benefit:**
  this has been a straightforward (if tedious) calculation,
  and our opinions / beliefs / prejudices have not been involved
- **the costs:**
  1. we have not directly addressed the q^n, *"is this a fair coin?"*,
     that we were originally interested in
  2. not clear how to incorporate, say, suspicion about the coin

Tossing a coin: Bayesian treatment

What is the definition of the conditional probability $P(A\,|\,B)$?

Bayesian Probability:

$P(A|B)$ is a real-number measure of the plausibility of proposition A, given (conditional upon) the truth of proposition B;

P measures degree of belief;

A can be any logical proposition

Tossing a coin: Bayesian treatment . . .

. . . provides a natural way to incorporate relevant background information, such as the fact that you are playing coin-toss with

Tossing a coin: Bayesian treatment

Likelihoods:

$$P((4H, 1T)) \,|\, \text{fair}) = 0.1563$$
$$P((4H, 1T)) \,|\, \text{bad}) = 0.3955$$

Priors:

$$P(\text{fair} \,|\, \text{Cap}) = 0.95$$
$$P(\text{bad} \,|\, \text{Cap}) = 0.05$$

Posterior:

$$\begin{aligned} P(\text{fair} \,|\, (4H, 1T), \text{ Cap}) &= \frac{P((4H, 1T)) \,|\, \text{fair}) \cdot P(\text{fair} \,|\, \text{Cap})}{\sum_i P((4H, 1T)) \,|\, i) \cdot P(i \,|\, \text{Cap})} \\ &= \frac{0.1563 \cdot 0.95}{0.1563 \cdot 0.95 + 0.3955 \cdot 0.05} \\ &= 0.882 \end{aligned}$$

Tossing a coin: Bayesian treatment

Likelihoods:

$$P((8H, 2T)) \mid \text{fair}) = 0.04395$$
$$P((8H, 2T)) \mid \text{bad}) = 0.28157$$

Priors:

$$P(\text{fair} \mid \text{Cap}) = 0.95$$
$$P(\text{bad} \mid \text{Cap}) = 0.05$$

Posterior:

$$P(\text{fair} \mid (8H, 2T), \text{ Cap}) = \frac{P((8H, 2T)) \mid \text{fair}) \cdot P(\text{fair} \mid \text{Cap})}{\sum_i P((8H, 2T)) \mid i) \cdot P(i \mid \text{Cap})} = \frac{0.04395 \cdot 0.95}{0.04395 \cdot 0.95 + 0.28157 \cdot 0.05} = 0.748$$

Tossing a coin: Bayesian treatment . . .

. . . provides a natural way to incorporate relevant background information, such as the fact that you are playing coin-toss with

Tossing a coin: Bayesian treatment

Likelihoods:

$$P((4H, 1T)) \,|\, \text{fair}) = 0.1563$$
$$P((4H, 1T)) \,|\, \text{bad}) = 0.3955$$

Priors:

$$P(\text{fair} \,|\, \text{Loki}) = 0.50$$
$$P(\text{bad} \,|\, \text{Loki}) = 0.50$$

Posterior:

$$\begin{aligned} P(\text{fair} \,|\, (4H, 1T), \text{Loki}) &= \frac{P((4H, 1T)) \,|\, \text{fair}) \cdot P(\text{fair} \,|\, \text{Loki})}{\sum_i P((4H, 1T)) \,|\, i) \cdot P(i \,|\, \text{Loki})} \\ &= \frac{0.1563 \cdot 0.50}{0.1563 \cdot 0.50 + 0.3955 \cdot 0.50} \\ &= 0.283 \end{aligned}$$

Tossing a coin: Bayesian treatment

Likelihoods:

$$P((8H, 2T)) \,|\, \text{fair}) = 0.04395$$
$$P((8H, 2T)) \,|\, \text{bad}) = 0.28157$$

Priors:

$$P(\text{fair} \,|\, \text{Loki}) = 0.50$$
$$P(\text{bad} \,|\, \text{Loki}) = 0.50$$

Posterior:

$$P(\text{fair} \,|\, (8H, 2T), \text{Loki}) = \frac{P((8H, 2T)) \,|\, \text{fair}) \cdot P(\text{fair} \,|\, \text{Loki})}{\sum_i P((8H, 2T)) \,|\, i) \cdot P(i \,|\, \text{Loki})} = \frac{0.04395 \cdot 0.50}{0.04395 \cdot 0.50 + 0.28157 \cdot 0.50} = 0.135$$

Tossing a coin: Bayesian treatment — commentary

- N.B. biased coin should have p_H free (not fixed $p_H = 0.75$)

Tossing a coin: Bayesian treatment — commentary

- N.B. biased coin should have p_H free (not fixed $p_H = 0.75$)
  - Bayesian model comparison with (simple) fair-coin model:

Tossing a coin: Bayesian treatment — commentary

- N.B. biased coin should have p_H free (not fixed $p_H = 0.75$)
  - Bayesian model comparison with (simple) fair-coin model: implements Occam's razor; integr^n over p_H, with *prior* $\pi(p_H)$

Tossing a coin: Bayesian treatment — commentary

- N.B. biased coin should have p_H free (not fixed $p_H = 0.75$)
  - Bayesian model comparison with (simple) fair-coin model: implements Occam's razor; integr^n over p_H, with *prior* $\pi(p_H)$
  - application to the Higgs case is left for discussion ...

Tossing a coin: Bayesian treatment — commentary

- N.B. biased coin should have p_H free (not fixed $p_H = 0.75$)
  - Bayesian model comparison with (simple) fair-coin model: implements Occam's razor; integr^n over p_H, with *prior* $\pi(p_H)$
  - application to the Higgs case is left for discussion ...
- we need to be explicit about our prior beliefs in this problem; change in the prior $\longrightarrow$ change in the result

Tossing a coin: Bayesian treatment — commentary

- N.B. biased coin should have p_H free (not fixed $p_H = 0.75$)
  - Bayesian model comparison with (simple) fair-coin model: implements Occam's razor; integr^n over p_H, with *prior* $\pi(p_H)$
  - application to the Higgs case is left for discussion ...
- we need to be explicit about our prior beliefs in this problem; change in the prior $\longrightarrow$ change in the result
- **this is a feature,** not a bug, of Bayesian statistics:

Tossing a coin: Bayesian treatment — commentary

- N.B. biased coin should have p_H free (not fixed $p_H = 0.75$)
  - Bayesian model comparison with (simple) fair-coin model: implements Occam's razor; integr^n over p_H, with *prior* $\pi(p_H)$
  - application to the Higgs case is left for discussion . . .
- we need to be explicit about our prior beliefs in this problem; change in the prior $\longrightarrow$ change in the result
- **this is a feature,** not a bug, of Bayesian statistics:
  - Bayesian probability includes beliefs: no fundamental objection

Tossing a coin: Bayesian treatment — commentary

- N.B. biased coin should have p_H free (not fixed $p_H = 0.75$)
  - Bayesian model comparison with (simple) fair-coin model: implements Occam's razor; integr^n over p_H, with *prior* $\pi(p_H)$
  - application to the Higgs case is left for discussion . . .
- we need to be explicit about our prior beliefs in this problem; change in the prior $\longrightarrow$ change in the result
- **this is a feature,** not a bug, of Bayesian statistics:
  - Bayesian probability includes beliefs: no fundamental objection
  - different prior, different problem: no absurdity in the conclusion

Tossing a coin: Bayesian treatment — commentary

- N.B. biased coin should have p_H free (not fixed $p_H = 0.75$)
  - Bayesian model comparison with (simple) fair-coin model: implements Occam's razor; integr^n over p_H, with *prior* $\pi(p_H)$
  - application to the Higgs case is left for discussion . . .
- we need to be explicit about our prior beliefs in this problem; change in the prior $\longrightarrow$ change in the result
- **this is a feature,** not a bug, of Bayesian statistics:
  - Bayesian probability includes beliefs: no fundamental objection
  - different prior, different problem: no absurdity in the conclusion
- **the benefits:**

Tossing a coin: Bayesian treatment — commentary

- N.B. biased coin should have p_H free (not fixed $p_H = 0.75$)
  - Bayesian model comparison with (simple) fair-coin model: implements Occam's razor; integr^n over p_H, with *prior* $\pi(p_H)$
  - application to the Higgs case is left for discussion . . .
- we need to be explicit about our prior beliefs in this problem; change in the prior $\longrightarrow$ change in the result
- **this is a feature,** not a bug, of Bayesian statistics:
  - Bayesian probability includes beliefs: no fundamental objection
  - different prior, different problem: no absurdity in the conclusion
- **the benefits:**
  1. we are directly addressing the question, *"is this a fair coin?"*

Tossing a coin: Bayesian treatment — commentary

- N.B. biased coin should have p_H free (not fixed $p_H = 0.75$)
  - Bayesian model comparison with (simple) fair-coin model: implements Occam's razor; integr^n over p_H, with *prior* $\pi(p_H)$
  - application to the Higgs case is left for discussion ...

- we need to be explicit about our prior beliefs in this problem; change in the prior $\longrightarrow$ change in the result
- **this is a feature,** not a bug, of Bayesian statistics:
  - Bayesian probability includes beliefs: no fundamental objection
  - different prior, different problem: no absurdity in the conclusion

- **the benefits:**
  1. we are directly addressing the question, *"is this a fair coin?"*
  2. we can straightforwardly incorporate external/prior information

Tossing a coin: Bayesian treatment — commentary

- N.B. biased coin should have p_H free (not fixed $p_H = 0.75$)
  - Bayesian model comparison with (simple) fair-coin model: implements Occam's razor; integr^n over p_H, with *prior* $\pi(p_H)$
  - application to the Higgs case is left for discussion . . .
- we need to be explicit about our prior beliefs in this problem; change in the prior $\longrightarrow$ change in the result
- **this is a feature,** not a bug, of Bayesian statistics:
  - Bayesian probability includes beliefs: no fundamental objection
  - different prior, different problem: no absurdity in the conclusion
- **the benefits:**
  1. we are directly addressing the question, *"is this a fair coin?"*
  2. we can straightforwardly incorporate external/prior information
- **the cost:**
  we cannot avoid committing ourselves (perhaps provisionally) in the course of the method

BH merger parameters: mass, spin, distance

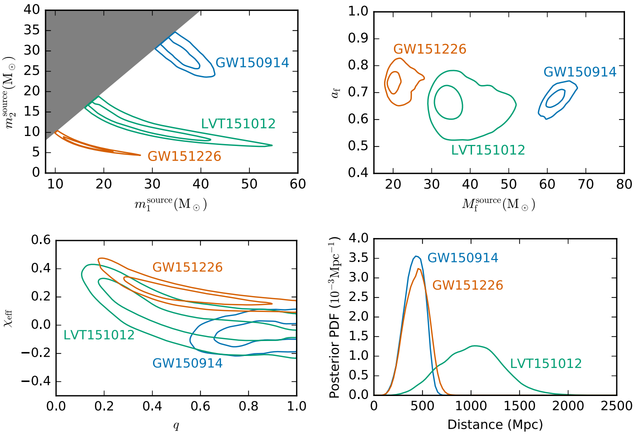


FIG. 4. Posterior probability densities of the masses, spins, and distance to the three events GW150914, LVT151012, and GW151226. For the two-dimensional distributions, the contours show 50% and 90% credible regions. Top left panel: Component masses $m_1^{\rm source}$ and $m_2^{\rm source}$ for the three events. We use the convention that $m_1^{\rm source} \geq m_2^{\rm source}$, which produces the sharp cut in the two-dimensional distribution. For GW151226 and LVT151012, the contours follow lines of constant chirp mass ($\mathcal{M}^{\rm source} = 8.9^{+0.3}_{-0.3}\rm M_\odot$ and $\mathcal{M}^{\rm source} = 15.1^{+1.4}_{-1.1}\rm M_\odot$, respectively). In all three cases, both masses are consistent with being black holes. Top right panel: The mass and dimensionless spin magnitude of the final black holes. Bottom left panel: The effective spin and mass ratios of the binary components. Bottom right panel: The luminosity distance to the three events.

BH merger parameters: dim^{n}less component spins

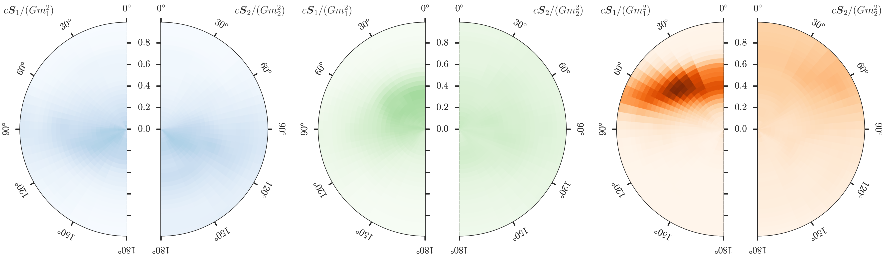


FIG. 5. Posterior probability distributions for the dimensionless component spins $cS_1/(Gm_1^2)$ and $cS_2/(Gm_2^2)$ relative to the normal to the orbital plane L, marginalized over the azimuthal angles. The bins are constructed linearly in spin magnitude and the cosine of the tilt angles, and therefore have equal prior probability. The left plot shows the distribution for GW150914, the middle plot is for LVT151012, and the right plot is for GW151226.

BH merger parameters: sky locations

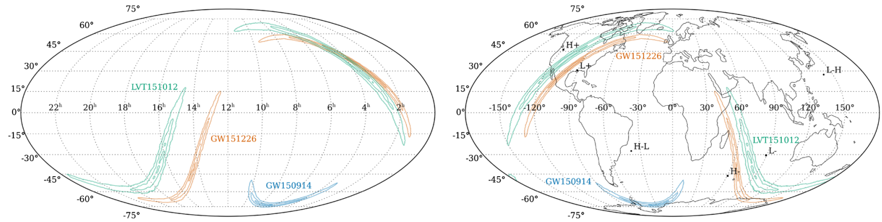


FIG. 6. Posterior probability distributions for the sky locations of GW150914, LVT151012, and GW151226 shown in a Mollweide projection. The left plot shows the probable position of the source in equatorial coordinates (right ascension is measured in hours and declination is measured in degrees). The right plot shows the localization with respect to the Earth at the time of detection. H+ and L+ mark the Hanford and Livingston sites, and H− and L− indicate antipodal points; H-L and L-H mark the poles of the line connecting the two detectors (the points of maximal time delay). The sky localization forms part of an annulus, set by the difference in arrival times between the detectors.

BH mergers: post-Newtonian parameters

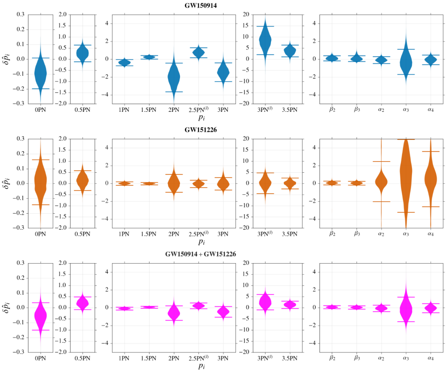


FIG. 7. Posterior density distributions and 90% credible intervals for relative deviations $\delta\hat{p}_i$ in the PN parameters p_i (where (l) denotes the logarithmic correction), as well as intermediate parameters β_i and merger-ringdown parameters α_i. The top panel is for GW150914 by itself and the middle one for GW151226 by itself, while the bottom panel shows *combined* posteriors from GW150914 and GW151226. While the posteriors for deviations in PN coefficients from GW150914 show large offsets, the ones from GW151226 are well centered on zero, as well as being tighter, causing the combined posteriors to similarly improve over those of GW150914 alone. For deviations in the β_i, the combined posteriors improve over those of either event individually. For the α_i, the joint posteriors are mostly set by the posteriors from GW150914, whose merger-ringdown occurred at frequencies where the detectors are the most sensitive.

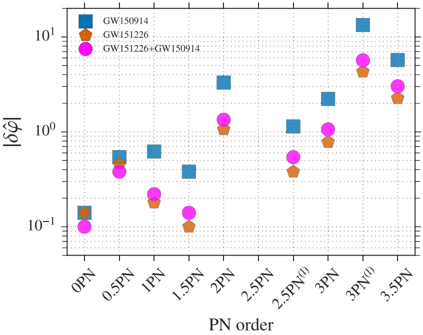


FIG. 8. The 90% credible upper bounds on deviations in the PN coefficients, from GW150914 and GW151226. Also shown are joint upper bounds from the two detections; the main contributor is GW151226, which had many more inspiral cycles in band than GW150914. At 1PN order and higher, the joint bounds are slightly looser than the ones from GW151226 alone; this is due to the large offsets in the posteriors for GW150914.

BH mergers: dependence on priors

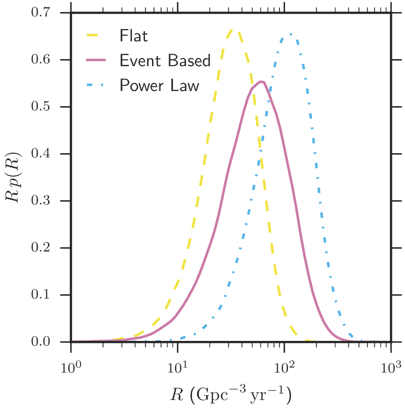


FIG. 11. The posterior density on the rate of BBH mergers. The curves represent the posterior assuming that BBH masses are distributed flat in $\log(m_1) - \log(m_2)$ (Flat), match the properties of the observed events (Event based), or are distributed as a power law in m_1 (Power law). The posterior median rates and symmetric 90% symmetric credible intervals are given in Table II.

Analysis relies on priors on

- detector properties
- distributions of sources in space
- BH masses in coalescing binaries
- BH spins ...
- ... and all sorts of things

APPENDIX

6. **Probability**

7. **What are systematics?**

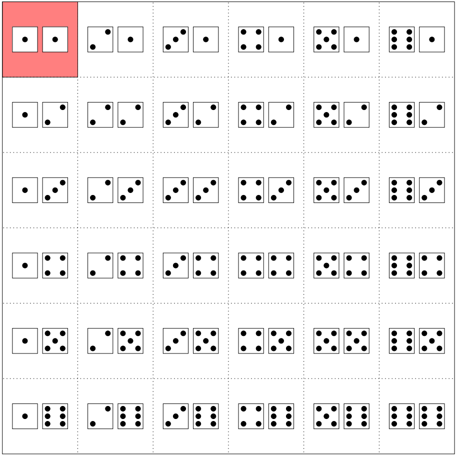
Probability: introduction

Intuitively:
chance that something is true ("it will rain tomorrow"),
or that a parameter has some value ("the coin shows heads")

Mathematical foundation dates only from last century:

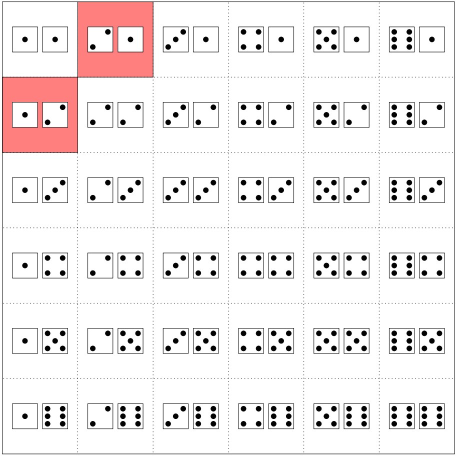
- consider $\Omega = \{X_i\}$, all possible *exclusive elementary events* X_i;
  e.g. die showing 1, 2, 3, 4, 5, or 6
- a "probability" function P must satisfy:
  - $P(X_i) \geq 0 \; \forall i$
  - $P(X_i \text{ or } X_j) = P(X_i) + P(X_j)$
  - $\sum_\Omega P(X_i) = 1$
- more than one sort of thing obeys these rules:
  - frequentist probability: "limiting frequency"
    $P(X_i) = \lim_{n\to\infty} n_i/n$
    i.e. the fraction if you repeat the "experiment" endlessly
  - Bayesian "degree of belief"

Probability: illustrations using two dice



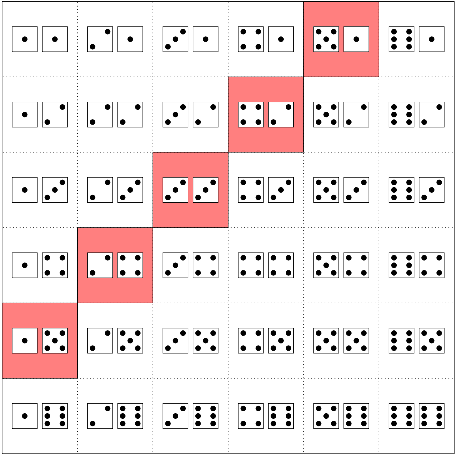
- X_{11} is an elementary event

Probability: illustrations using two dice



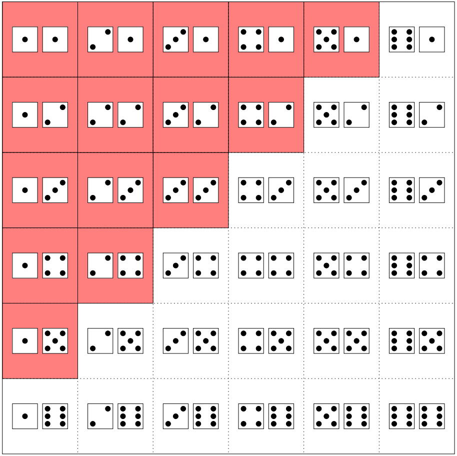
- X_{11} is an elementary event
- X_{12}, X_{21} indistinguishable in practice, so define a *set* $Y_{12} = \{X_{12}, X_{21}\}$ & treat like an event: "a one and a two"

Probability: illustrations using two dice



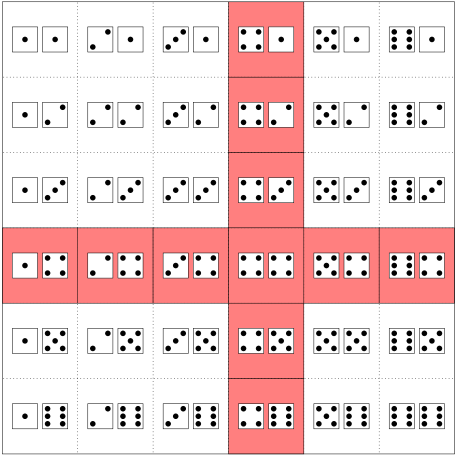
- X_{11} is an elementary event
- X_{12}, X_{21} indistinguishable in practice, so define a *set* $Y_{12} = \{X_{12}, X_{21}\}$ & treat like an event: "a one and a two"
- $\{X_{ij} \mid i + j = 6\}$ contains distinguishable events, but useful anyway: "I threw a six"

Probability: illustrations using two dice



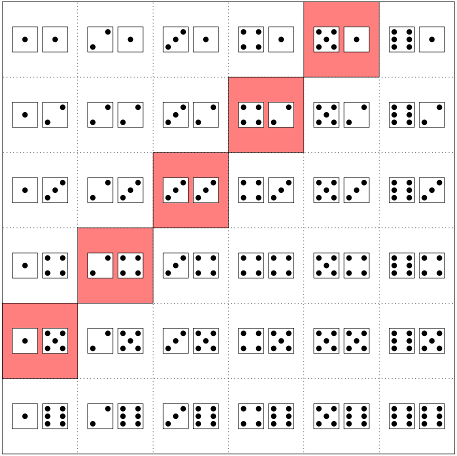
- X_{11} is an elementary event
- X_{12}, X_{21} indistinguishable in practice, so define a *set* $Y_{12} = \{X_{12}, X_{21}\}$ & treat like an event: "a one and a two"
- $\{X_{ij} \mid i + j = 6\}$ contains distinguishable events, but useful anyway: "I threw a six"
- $\{X_{ij} \mid i + j \leq 6\}$: "six or less"

Probability: illustrations using two dice



- X_{11} is an elementary event
- X_{12}, X_{21} indistinguishable in practice, so define a *set* $Y_{12} = \{X_{12}, X_{21}\}$ & treat like an event: "a one and a two"
- $\{X_{ij} \mid i + j = 6\}$ contains distinguishable events, but useful anyway: "I threw a six"
- $\{X_{ij} \mid i + j \leq 6\}$: "six or less"
- $\{X_{ij} \mid i = 4 \text{ OR } j = 4\}$: "at least one four"

Probability: addition law for sets



Elementary events add simply, so

$$P(\{X_{ij} \mid i+j=6\}) = \sum_{i+j=6} P(X_{ij})$$

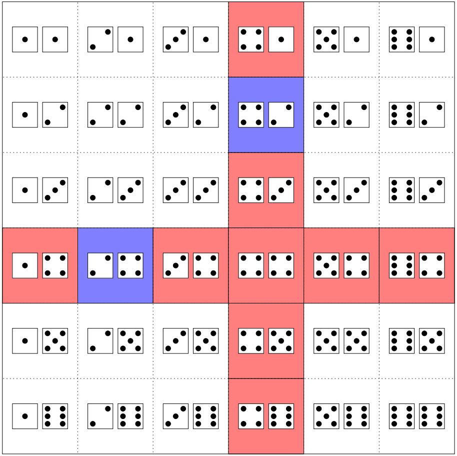
It follows that for sets

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

So far, this is mere counting.

But consequences flow from this.

Probability: conditional probability

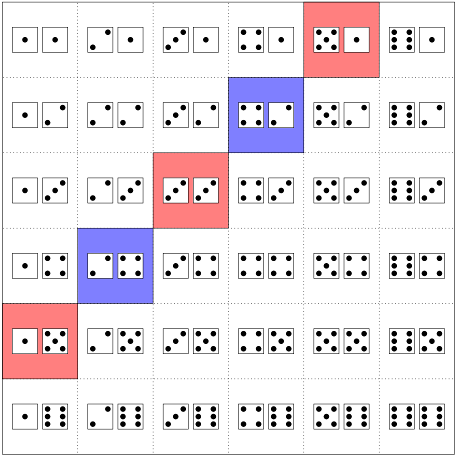


The probability that an elementary event known to belong to set B also belongs to set A; defined *via*

$$P(A \text{ and } B) = P(A \,|\, B) \cdot P(B)$$
$$= P(B \,|\, A) \cdot P(A)$$

$$P(6 \text{ and } \text{“}\geq \text{ one } 4\text{”}) = 2/36$$
$$P(\text{“}\geq \text{ one } 4\text{”}) = 11/36$$
$$P(6 \,|\, \text{“}\geq \text{ one four”}) = 2/11$$

Probability: conditional probability



The probability that an elementary event known to belong to set B also belongs to set A; defined *via*

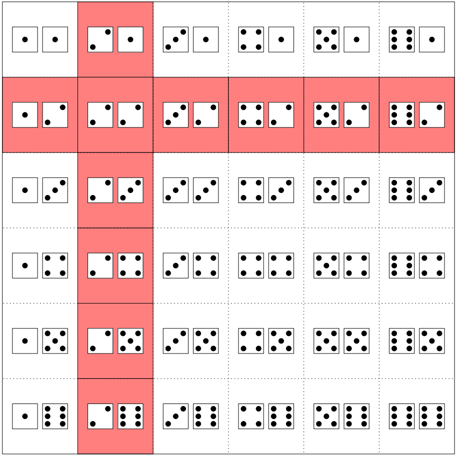
$$P(A \text{ and } B) = P(A \,|\, B) \cdot P(B) = P(B \,|\, A) \cdot P(A)$$

$$P(\text{“}\geq \text{one 4” and 6}) = 2/36$$

$$P(6) = 5/36$$

$$P(\text{“}\geq \text{one four”} \,|\, 6) = 2/5 \neq 2/11$$

Probability: independence



Consider a more symmetric case:

$$P(\text{“}\geq \text{ one 2”}) = 11/36$$

$$P(\text{“}\geq \text{ one 4”}) = 11/36$$

$$P(\text{“}\geq \text{ one 4”} \mid \text{“}\geq \text{ one 2”}) = 2/11$$

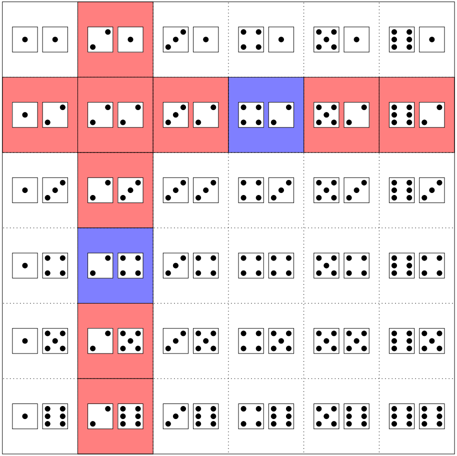
$$P(\text{“}\geq \text{ one 2”} \mid \text{“}\geq \text{ one 4”}) = 2/11$$

$$\neq 11/36$$

so getting a 2 and getting a 4 are *not independent*: for independent sets,

$$P(A \mid B) = P(A)$$

Probability: independence



Consider a more symmetric case:

$$P(\text{“}\geq \text{one } 2\text{”}) = 11/36$$

$$P(\text{“}\geq \text{one } 4\text{”}) = 11/36$$

$$P(\text{“}\geq \text{one } 4\text{”} \mid \text{“}\geq \text{one } 2\text{”}) = 2/11$$

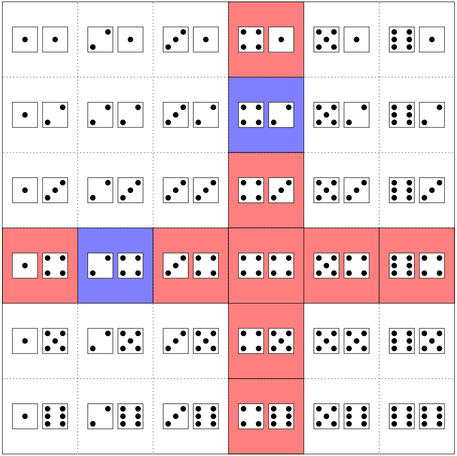
$$P(\text{“}\geq \text{one } 2\text{”} \mid \text{“}\geq \text{one } 4\text{”}) = 2/11$$

$$\neq 11/36$$

so getting a 2 and getting a 4 are *not independent*: for independent sets,

$$P(A \mid B) = P(A)$$

Probability: independence



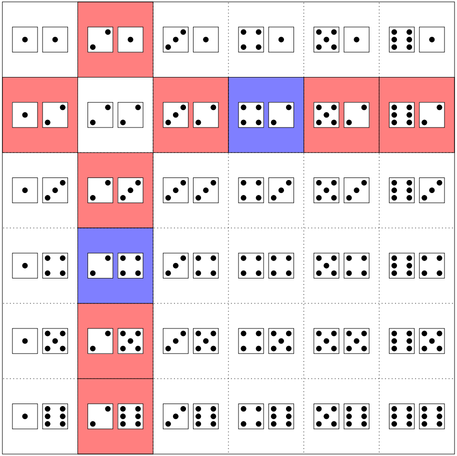
Consider a more symmetric case:

$$P(\text{“}\geq \text{ one 2”}) = 11/36$$
$$P(\text{“}\geq \text{ one 4”}) = 11/36$$
$$P(\text{“}\geq \text{ one 4”} \mid \text{“}\geq \text{ one 2”}) = 2/11$$
$$P(\text{“}\geq \text{ one 2”} \mid \text{“}\geq \text{ one 4”}) = 2/11$$
$$\neq 11/36$$

so getting a 2 and getting a 4 are *not independent*: for independent sets,

$$P(A \mid B) = P(A)$$

Probability: independence



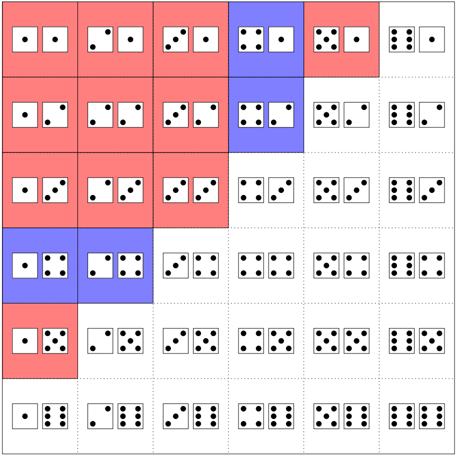
Very hard to find nontrivial independent sets in this example . . . contrary to what you might think.

Obvious fix doesn't work:

$$\begin{aligned} P(\text{“one 2”}) &= 10/36 \\ P(\text{“one 4”}) &= 10/36 \\ P(\text{“one 4”} \mid \text{“one 2”}) &= \ 2/10 \\ P(\text{“one 2”} \mid \text{“one 4”}) &= \ 2/10 \\ &\neq 10/36 \end{aligned}$$

(see the pre-reading for more)

Probability: introduction to Bayes' Theorem



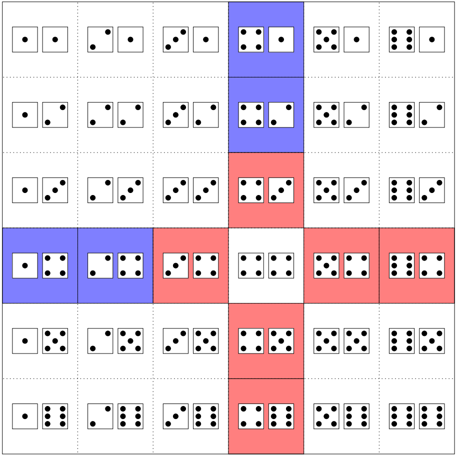
$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

$$P(\text{"one 4"} \mid \text{"}\leq 6\text{"}) = 4/15$$
$$P(\text{"}\leq 6\text{"} \mid \text{"one 4"}) = 4/10$$
$$P(\text{"one 4"}) = 10/36$$
$$P(\text{"}\leq 6\text{"}) = 15/36$$

in general, $P(A) \neq P(B)$

so $P(A \mid B) \neq P(B \mid A)$

Probability: introduction to Bayes' Theorem



$$P(A\,|\,B) = \frac{P(B\,|\,A)\cdot P(A)}{P(B)}$$

$$P(\text{“one 4”}\,|\,\text{“}\le 6\text{”}) = 4/15$$
$$P(\text{“}\le 6\text{”}\,|\,\text{“one 4”}) = 4/10$$
$$P(\text{“one 4”}) = 10/36$$
$$P(\text{“}\le 6\text{”}) = 15/36$$

in general, $P(A) \neq P(B)$

so $P(A\,|\,B) \neq P(B\,|\,A)$

What are systematics? (1) Old-fashined experimentalist answer

Imagine the answer of Kyūzō-san (Skill Guy from *Seven Samurai*):

Counting uncertainties go as $\sqrt{N}$, and are reliable.

Everything else is systematic uncertainty,
and is governed only by judgement and rules of thumb.

I know my techniques and my mental bank of examples. So I know a systematic uncertainty when I see one, and use an appropriate estimate.

This new-fangled tool `MINUIT` also returns uncertainty estimates, and empirically it seems reliable when it works (I always check by reading the verbose output carefully). I perk its results in as "statistical uncertainties", but only on sufferance.

What are systematics? (2) Time-constrained student answer

Imagine the answer of a student six months from submission:

My advisor / Working Group / RC has a standard list of systematics for this sort of analysis, and methods for estimating them.

I have never seen an analysis make it to journal without all of these systematic terms, and only occasionally see an extra term (and then only because a collaborator insisted during CWR).

I have looked at papers from past experiments, and these standard systematics are in most of them.

One seems to be new in Tom Browder's CLEO paper.

I asked him about it once at a Belle II party. He said that this systematic was actually brought to the Hawaiian islands by Maui, and handed down. I am not sure whether he was joking.

What are systematics? (3) New-fashined experimentalist . . .

Imagine the answer of an ATLAS/CMS analysis contact:

Most uncertainties are statistical uncertainties in disguise.

It is straightforward to include them in the fit using Gaussian constraint (or Log-Normal or other constraints if necessary, for extra credit).

Correlations do not scare me, as they can be handled by covariance matrices; complex effects can be modelled by toy Monte Carlo.
As for effects that are only present in the data, these can be accessed via bootstrapping.

My old thesis advisor still includes some systematic terms by hand, and even uses $\pm 1\sigma$ estimates sometimes. We still collaborate (she is a member of the 50-person team on our paper), and I cannot stop her from doing this, but it is seriously embarrassing when I have to show such estimates to the Editorial Board.

What are systematics? A taxonomy

My basic answer is that all three of these answers are correct as far as they go. But they have something to learn from each other.

To the student I would add:

There is a weak presumption in favour of all of the uncertainties on the standard list. Individually they are likely to be appropriate, but maybe not all of them. And some extra terms may be necessary.

But how do you know which uncertainties to drop, and which ones to add?

That is the purpose of this talk

My working taxonomy of systematics:

1. (uncertainties on) the INPUTS to the measurement
2. (uncertainties on) AUXILIARY measurements
3. (uncertainties on) the CALIBRATION of the apparatus