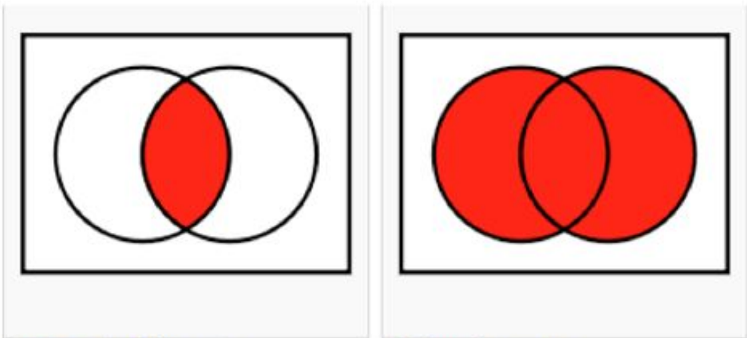
Radiation Detector 2018/19 (SPA6309)

Statistics

Basics of probability theory (Leo, 4.1)



Intersection of two sets $A \cap B$ Union of two sets $A \cup B$

The box describes entire **universe**, and the circle describes the **set** A, the **intersection** of 2 sets A and B is written $A \cap B$, and **union** of 2 sets is written $A \cup B$. Define the probability to be A is $P(A)$, and the probability of B is $P(B)$, then the probability of $A \cup B$ is

$$P(A \cup B) + P(A) + P(B) - P(A \cap B)$$

which can be understood graphically.

e.g.) Array of identical detectors

There is a detector which can detect a certain phenomenon with 93% probability. We prepare an array of 4 identical detectors, A, B, C, and D, and all are looking for the same new phenomenon at the same location. If you require the coincidence of all 4 detectors, the total efficiency is $P(A \cap B \cap C \cap D) = P(A) \cdot P(B) \cdot P(C) \cdot P(D)$

In [7]:
```
p=0.93
print "the total efficiency of the detector array is",p*p*p*p
```

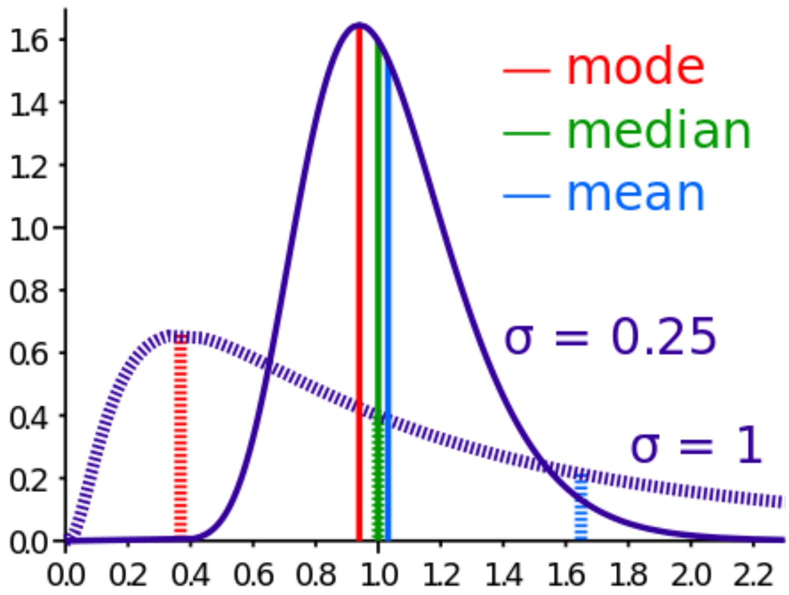
```
the total efficiency of the detector array is 0.74805201
```

Thus, detection efficiency is lower than a single detector. On the other hand, if you lose the detection condition, for example, first we ask either A or B detects the phenomenon. Second, we ask either C or D detects the phenomenon. And then we ask the coincidence of them. Namely,
$P((A \cup B) \cap (C \cup D)) = (P(A) + P(B) - P(A \cap B)) \cdot (P(C) + P(D) - P(C \cap D))$

In [9]:
```
print "the total efficiency of the detector array is",(p+p-p*p)*
```

```
the total efficiency of the detector array is 0.99022401
```

The detection efficiency changes depending on how to set the trigger condition.



A **probability density distribution (PDF)** P(x) is a distribution function of x, and normalized to be 1 in the universe.

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

mode is defined where the distribution makes the local peak. On the other hand, **median** is defined where cumulative distribution is 50%.

$\int_{-\infty}^{median} P(x)dx = 0.5$

mean (μ, or the **expectation value of x**) is defined by (eq. 4.5)

$$\int_{-\infty}^{\infty} xP(x)dx = \mu = E[x]$$

The **sample mean** is the estimation of mean from series of measurement x_i (eq. 4.49)

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i \ \sim \ \mu$$

An important concept is a **variance** (σ^2), where σ is the **standard deviation**, often used as the error of x (eq. 4.9),

$$\int_{-\infty}^{\infty} (x-\mu)^2 P(x)dx = E[(x-\mu)^2] = \sigma^2$$

Also,

$$\int (x-\mu)^2 P(x)dx = \int (x^2 - 2x\mu + \mu^2)P(x)dx = E[x^2] - 2E[x]\mu + \mu^2 = E[x^2] - \mu$$

The **sample variance** is the estimation of variance from series of measurement x_i (eq. 4.52)

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \hat{\mu}) \ \sim \ \sigma^2$$

Please notice weird factor $\frac{1}{n-1}$ instead of $\frac{1}{n}$!

e.g.) Error comes from square trigger scintillator

In experiments, **trigger** is the concept to define data taking condition. There is an experiment to measure particle beam passing through a device. The beam size must be $d^2 \ cm^2$, but it is impossible to make a beam exactly $d^2 cm^2$. Instead, one could place a scintillator with $d^2 \ cm^2$ area in front of the device, and record data only if this scintillator (assuming 100% efficiency) detect particles passing through. Question, what is the error for such scintillator? Here, use a coordinate where the origin is the centre of the scintillator, then PDF can be defined $\frac{1}{d}$, thus,

$$\sigma^2 = \int_{-d/2}^{d/2} (x-\mu)^2 \frac{1}{d} dx = \frac{1}{d} \left[\frac{1}{3}x^3 - \mu x^2 + \mu^2 x\right]_{-d/2}^{d/2} = \frac{d^2}{12}$$

Therefore, such trigger adds an error of $\frac{d}{\sqrt{12}}$ on the beam position.

Bayes' theorem

The **conditional probability** of A with given B is $P(A|B)$. This is related to the conditional probability of B with given A.

$$P(A \cap B) = P(A|B)P(B) = P(B \cap A) = P(B|A)P(A).$$

Most importantly, $P(A|B) \neq P(A|B)$. Namely "the probability to get a king when you get a spade (1/13)" is different from " the probability to get a spade when you get a king (1/4)". This sounds damn and unmistakable, but many people do mistake in particle physics.

Let's say, there are n different choices of B, namely $\sum_i^n P(B_i) = 1$, then the sum of conditional probability will make the probability of A without any condition.

$$P(A) = \sum_i^n P(A|B_i)P(B_i)$$

By combining these 2, one can get the conditional probability of B_i with given A.

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_j^n P(A|B_j)P(B_j)}$$

This is the famous **Bayes' theorem.**

e.g.) Probability to find a disease from a test

From a study, 0.5% of people in a certain area has a disease ($P(B_1) = 0.005,\ P(B_2) = 0.995$). A test shows 97% times positive result (A) if the person has the disease ($P(A|B_1) = 0.97$), and 0.4% times positive result even though the person does not have the disease ($P(A|B_2) = 0.004$). The probability to get a positive result, when the person actually has the disease is

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2)} = \frac{0.97 \cdot 0.005}{0.97 \cdot 0.005 + 0.004 \cdot 0.995} = 0.5$$

So the probability to find the disease by this test (97% times right!) is only 55%. This sounds counterintuitive, and that's why probability theory is interesting.

In [1]:
```
print 0.97*0.0005/(0.97*0.005+0.004*0.995)
```

```
0.054926387316
```

Bionial distribution (Leo, 4.2.1)

Say p is the probability of success, and q is the probability of fail$= 1 - p$. If you try N times, then the probability to get r times success is described by the **binomial distribution**

$$B(r; N, p) =_N C_r p^r q^{N-r} \ , \ _N C_r = \frac{N!}{r!(N-r)!} = combination$$

e.g.) uneven coin toss

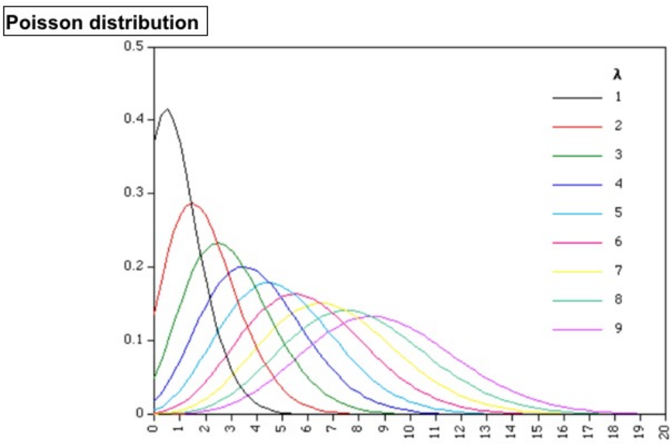
There is an uneven coin, where probability to get the head is 40%, and the tail is 60%. What is the probability to get 1 head after 3 tosses?

$$B(1; 3, 0.4) =_3 C_1 0.4^1 0.6^2 = 3 \cdot 0.4 \cdot 0.36 = 43.2\%$$

In [2]:
```
import math
print math.factorial(3)/math.factorial(1)/math.factorial(3-1)*0.
```

```
0.432
```

Poisson distribution (Leo, 4.2.2)



The **Poisson distribution** describes the distribution of the probability to get r when the mean is μ.

$$P(r; \mu) = \frac{\mu^r e^{-\mu}}{r!}$$

One important feature is $\sigma^2 = \mu$, namely, the error of the Poisson distribution is given by $\sqrt{\mu}$. Notice Poisson distribution is not symmetric. This means expression like $\pm 30\%$ error doesn't make sense if the statistics are too low. Usually, $\mu > 5$ is safe enough to assume Poisson distribution is symmetric.

e.g.) Measurement from zero event

In common sense, zero measurement means nothing. But if you know underlying distribution, measuring zero is different from not measure. If you know a phenomenon happen with randomly and distribute with Poisson distribution, one can derive an interesting result from the measurement of zero. Let's say you are performing a measurement of some random process, where you measure 0 for 40% times.

P(0;\mu)=e^{-\mu}=0.40~,~\mu=-ln(0.4)=0.92$

So the expectation value of this process is 0.92. Remarkably, you find this from measuring zero. "Don't perform measurement" and "performing a measurement and find zero" are statistically very different!

In [4]:
```
import numpy as np
print -np.log(0.368)
```

```
0.999672340813
```

e.g.) Null result limit (Leo, 4.5.4)

Let's say the mean rate of a process is λ, which make mean μ on a time period T ($\mu = \lambda \cdot T$). Then the zero-observation probability during period T is (eq. 4.57),

$P(0;\lambda) = e^{-\lambda T}$

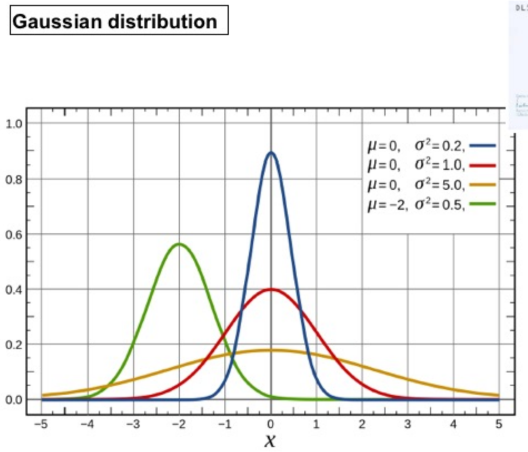
This can be seen as a PDF of λ, namely $P(\lambda) = Te^{-\lambda T}$ (extra T is to normalize PDF to be 1).

Then, the probability to observe a rate λ is lower than λ_0 is (4.58),

$P(\lambda|\lambda_0) = \int_0^{\lambda_0} Te^{-\lambda T} d\lambda = 1 - e^{-\lambda_0 T}, \lambda_0 = -\frac{1}{T} ln(1 - P(\lambda|\lambda_0))$

Such probability is called **confidence level (CL)** and used to make a statistical statement of a null result. For example, if there is no observation during period T, 90%CL upper limit of λ is $-\frac{1}{T} ln(1 - 0.9)$. For unit length, $-ln(1 - 0.9) = 2.3$. 2.3 is another important magic number, namely, **when you try to observe something and you cannot, then the 90%CL upper limit of that phenomenon is 2.3.**

Gaussian distribution (Leo, 4.2.3)



Gaussian distribution is the continuous limit of Poisson distribution. With mean μ and variance σ^2,

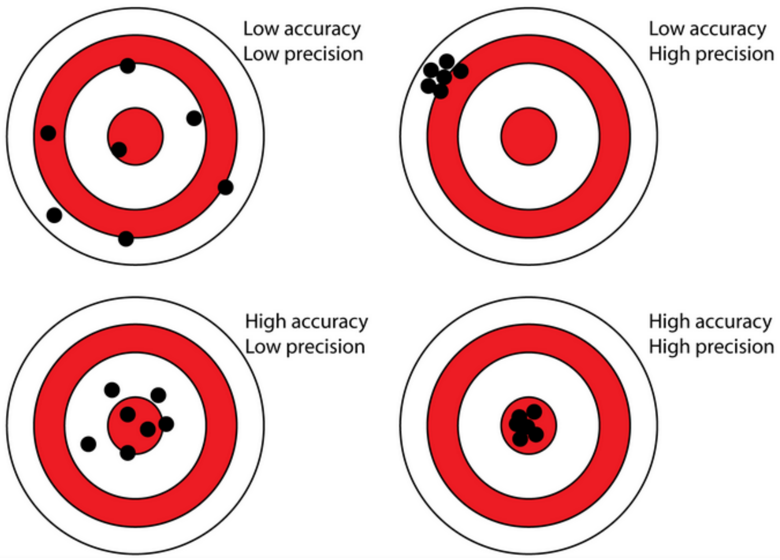
$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Unlike Poisson distribution, Gaussian distribution is always symmetric. σ is often used to show the significance of a measurement. Let's say, you measure The distribution has **68.3%** of area within $[\mu - 1\sigma, \mu + 1\sigma]$, **95.4%** within $[\mu - 2\sigma, \mu + 2\sigma]$, and **99.5%** within $[\mu - 3\sigma, \mu + 3\sigma]$. Therefore, if the signal has "$3\ \sigma$ significance", this means you are $3\ \sigma$ away from not signal, and the chance signal is by accident is 0.5%. Surprisingly, particle physicists think 0.5% is too large, and these days they require $5\ \sigma$ significance to claim the "evidence".

e.g.) Statistical error

An experiment measures 10 events of some random process in 1-day data taking. The statistical error is $\sqrt{10} = 3.16 = \sigma$ (Poisson distribution), and the measurement can be quoted 10 ± 3.16 and the data is 10/3.16=3.16σ away from zero (Gaussian distribution). The statistical error of the measurement is 31.6%. If you take the data 10 days, the measurement would be 100 ± 10 and the statistical error is 10%. The measurement is now 10σ away from nonzero, so the measurement is definitely not zero. Not surprisingly, fractional statistical error shrinks if you take data longer, or if you use bigger detector (more events in a certain time period). This is a major reason why you want to go bigger detectors with longer time period data taking.

Accuracy and Precision


Low accuracy
Low precision
Low accuracy
High precision
High accuracy
Low precision
High accuracy
High precision