PROBABILITY AND STATISTICS

Zhenyu Ye

January 12, 2015

References

- Ch10 of “Experiments in Modern Physics” by Melissinos.

- Particle Physics Data Group Review on “Probability”
http://pdg.lbl.gov/2014/reviews/rpp2014-rev-probability.pdf

- Particle Physics Data Group Review on “Statistics”
http://pdg.lbl.gov/2014/reviews/rpp2014-rev-statistics.pdf

Probability

- Frequency of occurrence of an event, $0 \leq P_i \leq 1$

Sample Space

- A set of points that represents all possible outcome of an experiment:

$$\Sigma P_i = 1$$

- Sample space for one throw of the die: 1, 2, 3, 4, 5, 6

$$P_i = 1/6$$

- Sample space for two throws: (1, 1), … (1, 6), (2, 1), … (2, 6) …, (6, 1), … (6, 6)

$$P_i = 1/36$$

- Sample space for two throws: $(n1<n2)$, $(n1=n2)$, $(n1>n2)$

$$P[n1<n2]=P[n1>n2]=5/12, \quad P[n1=n2] = 1/6$$

Complex Event

- Joint Probability: both A and B occur P[AB]

  P[“n1=1” “n2=1”] = ?
- Either Probability: either A or B occur P[A+B]

  P[“n1=1” + “n2=1”] = ?
- Conditional Probability: A will occur when B occurs P[A|B]

  P[“n1+n2>6” | “n1=3”] = ?

  A and B are **independent** if P[A|B]=P[A]

Complex Event

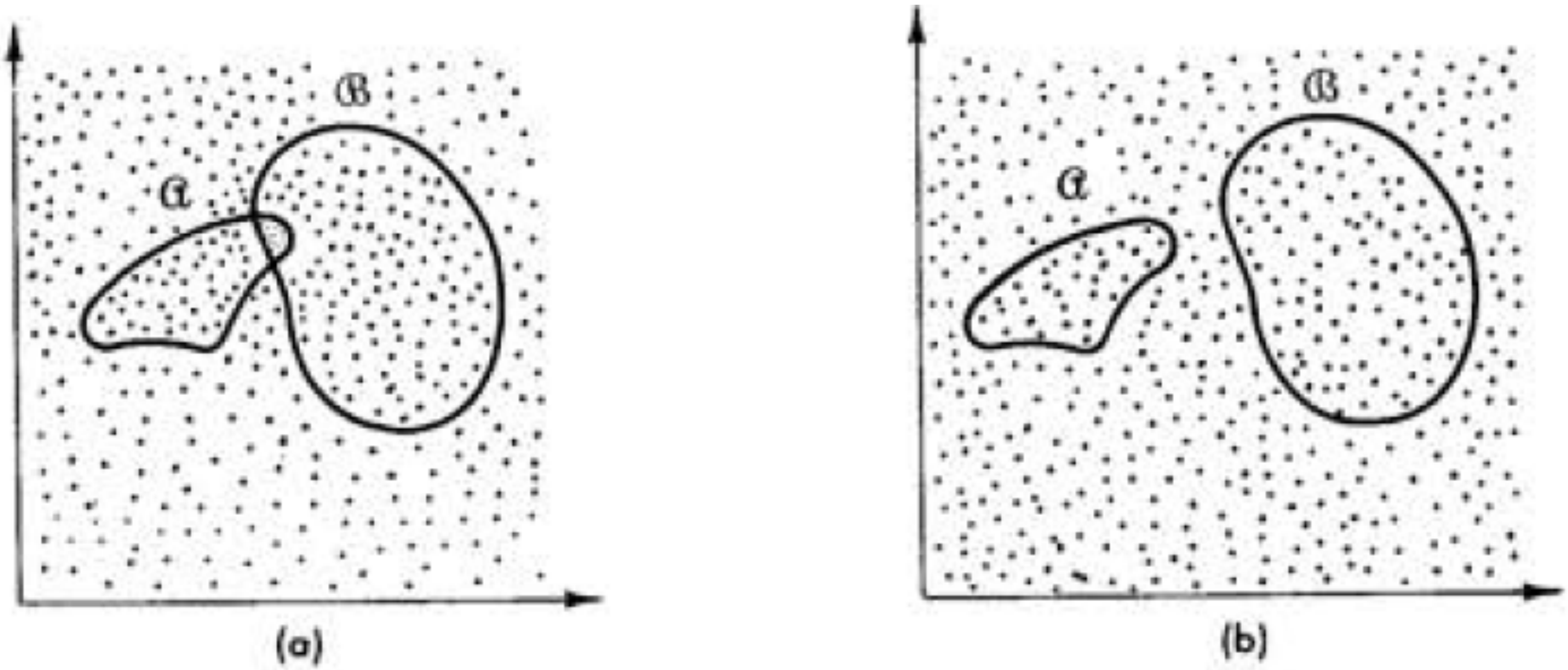


FIG. 10.2 In the sample spaces shown it is assumed that all sample-space points in domain 𝒜 contain event A, whereas all points in domain ℬ contain event B. (a) There exists a region where both event A and event B can occur simultaneously. (b) No such region exists; events A and B are mutually exclusive.

P[AB] = P[A|B]*P[B]=P[B|A]*P[A], A and B independent: P[AB]=P[A]*P[B]

First throw is 6, what is the probability of second throw is also 6?

Random Variable

- Numerical variable which takes a definite value for each and every point of the sample space; however, the same value may be assigned to several points.
- Discrete sample space: (Throwing die)

Sample space points	x (random variable)
n1<n2	-1
n1=n2	0
n1>n2	1

- Continuous sample space: (weight of raindrops)

Probability Distribution Function

- Frequency function, or probability distribution function f(x) gives the probability that the random variable x may take the specific value.
- Discrete random variable: $\Sigma f(x_i)=1$

Sample space points	Random variable x	Frequency Function f(x)
n1<n2	-1	5/12
n1=n2	0	1/6
n1>n2	1	5/12

- Continuous random variable:

$$\int f(x)\,dx = 1$$

Moments of PDF

- The kth moment of the PDF:

$$m_k = \sum_i x_i^k f(x_i) \qquad m_k = \int f(x) x^k dx$$

Mean: 1st moment

- The kth moment of the PDF about x_0:

$$m_k = \sum_i (x_i - x_0)^k f(x_i) \qquad m_k = \int f(x)(x - x_0)^k dx$$

Variance: 2nd moment about mean

Standard Deviation: square root of variance

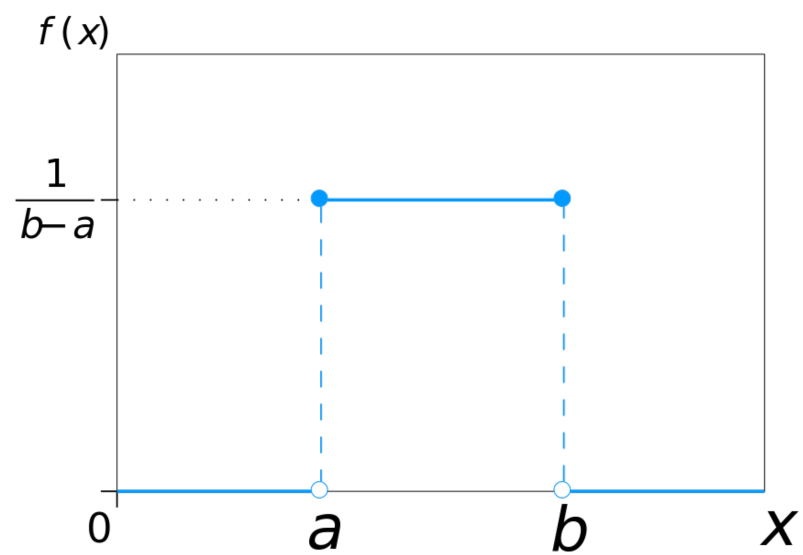
Uniform Distribution

- Constant probability to find a random variable in a given interval is said to follow the **Uniform Distribution**

$$f(x) = 1/(x_{max} - x_{min})$$

Mean: $(x_{max}+x_{min})/2$

Variance: $(x_{max}-x_{min})^2/12$

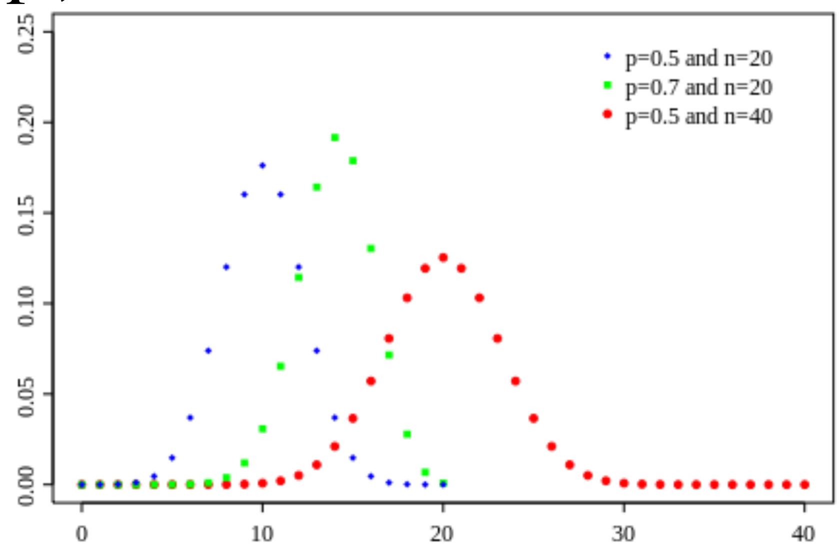


Binomial Distribution

- A random process with exactly two possible outcomes occurring with fixed probabilities is a **Bernoulli process**
- If the probability of obtaining an outcome (success) in one trial is p, the probability of obtaining exactly r (=0, 1, …, n) successes in n trials is given by **binomial distribution**

$$f(r;n,p) = \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r}$$

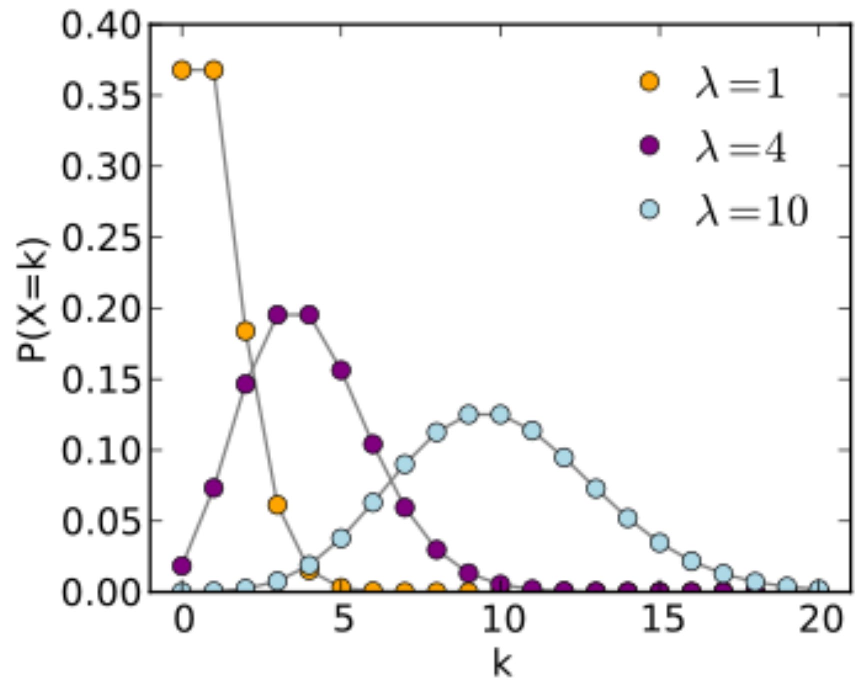
- **Mean:** np
- **Variance:** np(1-p)



Poisson Distribution

- Poisson Distribution gives the probability of finding exactly k events in a given interval of x when the events occur independently of one another and of x at an average rate of v per the given interval. It is the limiting case p->0, n->infinity and np=λ of the binomial distribution.

$$f(k;\lambda) = \frac{\lambda^k e^{\lambda}}{k!}$$



- **Mean:** λ
- **Variance:** λ

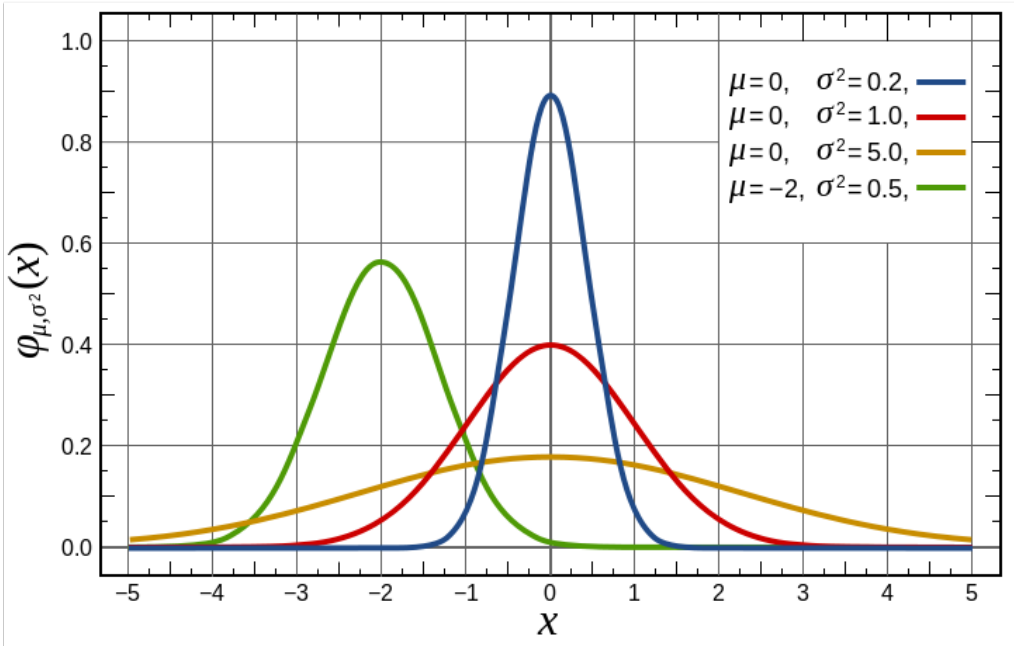
Normal or Gaussian Distribution

$$f(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-(x-\mu)^2/2\sigma^2)$$

$$-\infty < x < \infty \;; \qquad -\infty < \mu < \infty \;; \qquad \sigma > 0$$

- **Mean:** μ
- **Variance:** σ^2
- P(x in range μ±σ)=0.6827
- P(x in range μ±0.67σ)=0.5
- FWHM = 2.35σ

Poisson distribution approaches normal distribution at large N



Statistics

- An estimator is a function of data whose value is intended as a meaningful guess for the value of the parameter of the PDF. The most important feature
  - Consistency: converge to truth as the amount of data increase
  - Bias: difference between the expectation and true values
  - Efficiency: inverse of the ratio of the variance to the minimum possible variance for any estimator
  - Robustness: insensitive to departures from assumption in the PDF
- Unbiased estimators for N data points:

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i, \qquad \hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \hat{\mu})^2$$

whose variances are given by

$$V(\hat{\mu}) = \frac{\sigma^2}{N}, \qquad V(\hat{\sigma}^2) = \frac{1}{N}\left(m_4 - \frac{N-3}{N-1}\sigma^4\right)$$

Statistics

Card counting is a casino card game strategy used primarily in the blackjack family of casino games to determine whether the next hand is **likely to give a probable advantage to the player or to the dealer**. Card counters are a class of advantage players, who attempt to decrease the inherent casino house edge by keeping a running tally of all high and low valued cards seen by the player. Card counting allows players to bet more with less risk when the count gives an advantage as well as minimize losses during an unfavorable count. Card counting also provides the ability to alter playing decisions based on the composition of remaining cards.

Maximum Likelihood Method

- Suppose we have a set of N measured data $\mathbf{x}=(x_1 \ldots x_N)$ described by a joint PDF $L(\mathbf{x};\boldsymbol{\theta})$, where $\boldsymbol{\theta}=(\theta_1 \ldots \theta_n)$ is set of n parameters whose values are unknown.
- The likelihood function $L(\mathbf{x};\boldsymbol{\theta})$ is given by the PDF evaluated with the data **x**, but viewed as a function of the parameters $\boldsymbol{\theta}$. If x_i are statistically independent and each follow the PDF $f(x; \boldsymbol{\theta})$, then the joint PDF factorizes and $L(\mathbf{x};\boldsymbol{\theta})$ is given by

$$L(x;\theta) = \prod_{i=1}^{N} f(x;\theta)$$

- The maximum likelihood method takes the values of $\boldsymbol{\theta}$ that maximize $L(\mathbf{x};\boldsymbol{\theta})$, or equivalently minimize –ln(L). Thus ML estimators can be found by solving the equations

$$-\partial \ln L / \partial \theta_i = 0$$

Maximum Likelihood Method

- A set of data **x** that obey a normal distribution function about μ, with a standard deviation σ. ML estimators are:

$$f(x_i;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right]$$

$$L(x_i;\mu,\sigma) = \prod_i \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right]$$

$$-\frac{\partial \ln L}{\partial \mu} = -\sum_i \frac{x_i-\mu}{\sigma^2} \qquad \hat{\mu}_{ML} = \frac{1}{N}\sum_i x_i$$

$$-\frac{\partial \ln L}{\partial \sigma} = \frac{n}{\sigma} - \sum_i \frac{(x_i-\mu)^2}{\sigma^3} \qquad \hat{\sigma}_{ML} = \frac{1}{N}\sum_i (x_i-\mu)^2$$

Maximum Likelihood Method

- The inverse of the variance for the ML estimator can be estimated by:

$$V^{-1} = -\frac{\partial^2 \ln L}{\partial \theta^2}$$

$$-\frac{\partial \ln L}{\partial \mu} = -\sum_i \frac{x_i - \mu}{\sigma^2} \qquad V[\hat{\mu}_{ML}] = \frac{\sigma^2}{N}$$

- In the large sample limit, L has a Gaussian form. In this case, a numerically equivalent way of determining the standard deviation is given by

$$\ln L(\widehat{\theta}_{ML} \pm \sqrt{V[\widehat{\theta}_{ML}]}) = \ln L_{\max} - \frac{1}{2}$$

Least Squares Method

- Consider a set of N independent measurements y_i at known points x_i. Assuming y is Gaussian distributed with mean $F(x_i;\boldsymbol{\theta})$ and known variance σ_i^2. The negative maximum likelihood function is given by

$$\chi^2(\theta) = -2\ln L(\theta) + \text{ constant } = \sum_{i=1}^{N} \frac{(y_i - \mu(x_i;\theta))^2}{\sigma_i^2}$$

- The set of parameters which maximize $L(\boldsymbol{\theta})$ is the same that minimize $\chi^2(\boldsymbol{\theta})$, and thus the so-called least square method.
- A numerically equivalent way of determining the standard deviation is given by

$$\chi^2(\theta) = \chi^2_{\min} + 1$$

Uncertainty Propagation

- Z=x±y =>
  ΔZ=sqrt[(Δx)^2+(Δy)^2]

- Z=x*y or Z=x/y =>
  (ΔZ)/Z=sqrt[(Δx)^2/x^2+(Δy)^2/y^2]

- Z=x^{m}y^n =>
  (ΔZ)/Z=sqrt[(mΔx)^2/x^2+(nΔy)^2/y^2]

- Z=f(x,y) => $\Delta Z = \sqrt{\left(\frac{\partial f}{\partial x}\Delta x\right)^2 + \left(\frac{\partial f}{\partial y}\Delta y\right)^2}$