Statistical methods and error analysis

3rd Alice-India School 2020

Satyaki Bhattacharya
Saha Institute of Nuclear Physics

Conditional probability

A familiar example of conditional probability is dice throws. Suppose you have thrown a dice 3 times and got a 6 each time. What is the probability of getting a 6 in the next throw given that you have got three 6's in a three previous throws.

The answer is it's still $\frac{1}{6}$ since the events are independent or the answer is the probability of getting four 6's in a row is $(\frac{1}{6})^4$, so the probability is $(\frac{1}{6})^4$ for getting a 6 in a next throw.

Definition 8: If A and B are events in S, and $P(B)$ 0 then the conditional probability of A given B, written $P(A|B)$, is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{1}$$

Conditional probability example

Example application of conditional probability:

Draw four cards from top of a deck of cards. What is the probability that all four are aces?

The answer is $\frac{1}{{}^{52}C_4} = \frac{1}{270725}$

But in a alternative way,

Probability of 1^{st} card to be ace is $P(ace) = \frac{4}{52}$

Probability of 2^{nd} card to be ace given first is ace is $P(ace|1^{st}\, is\, ace) = \frac{3}{51}$ and so on.

$\implies P(all\, four\, ace) = \frac{4}{52} \cdot \frac{3}{51} \cdot \frac{2}{50} \cdot \frac{1}{49} = \frac{1}{270725}$

Bayes Theorem

Re-expressing (1) we have,

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

Using symmetry,

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

This is often called **Bayes Theorem**.

Theorem 4: Bayes Rule
Let $A_1, A_2, ...$ be a partition of the sample space and let B be any set. Then, for each $i = 1, 2, ...$

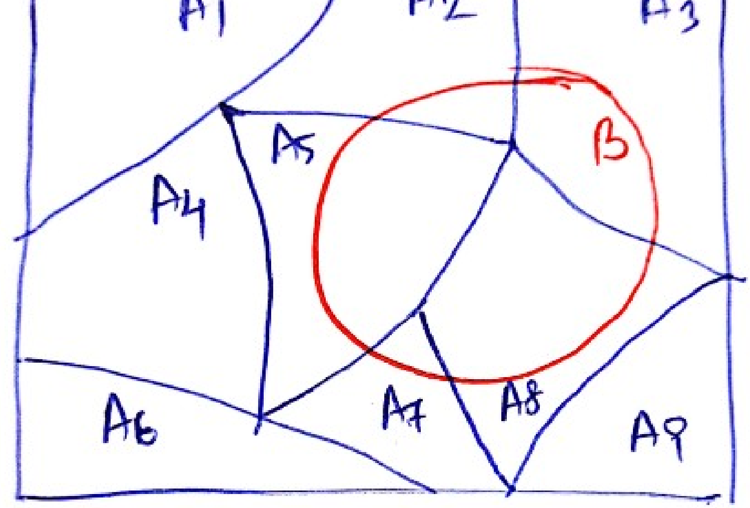
$$P(A_i|B) = \frac{\phi(B|A_i)P(A_i)}{\Sigma_{j=0}^{\infty} P(B|A_j)P(A_j)}$$

Example: Suppose we have 30% electron contamination in a pion beam in a calorimeter test beam run, i.e. $P(e) = \frac{3}{10}$ $P(\pi) = \frac{7}{10}$ in beam. Due to reconstruction error, there is a 5% chance that a prion gets reconstructed as electron and vice-versa. If in an event a pron has been detected, what is the probability that truly a pion hit the detector?

Solution: Problem 1.44 (Casella-Berger).

Consider a sample space S divided into disjoint subsets A_i,

i.e. $S = \cup_i A_i$, and $A_i \cap A_j = \emptyset$ for $i \neq j$.



Consider a subset $B \subset S$, it can be expressed as

$$B = B \cap S = B \cap (\cup_i A_i) = \cup_i (B \cap A_i)$$

$$\implies P(B) = P(\cup_i (B \cap A_i)) = \textstyle\sum_i P(B \cap A_i)$$

$$\implies P(B) = \textstyle\sum_i P(B|A_i) P(A_i)$$ law of total probability

Thus, Bayes' theorem becomes

$$P(A|B) = \frac{P(B|A)P(A)}{\sum_i P(B|A_i)P(A_i)}$$

Consider a disease D carried by 0.1% of people

i.e., the prior probabilities are

$P(D) = 0.001,$
$P(no\ D) = 0.999$

Consider a test that identifies the disease, the result is $+$ve or -ve

Suppose the probabilities to (in)correctly identify a person with the disease are,

$P(+|D) = 0.98,$
$P(-|D) = 0.02$

Similarly, suppose the probabilities to (in)correctly identify a healthy person

$P(+|no\ D) = 0.01,$
$P(-|no\ D) = 0.99$

What is the probability to have the disease if someone is tested $+$ve?

We can calculate it using the Bayes' theorem

i.e., the probability to have the disease given a +ve test result is

$$
\begin{aligned}
P(D|+) &= \frac{P(+|D)P(D)}{P(+|D)P(D) \; + \; P(+|no\ D)P(no\ D)} \\
&= \frac{0.98 \times 0.001}{0.98 \times 0.001 \; + \; 0.01 \times 0.999} \\
&= 0.089 \quad \text{(posterior probability)}
\end{aligned}
$$

What does it mean?

Patient's view: Probabilty for him to have the disease is 8.9%.

Doctor's view: 8.9% of people like this have the disease.

Independent but identically distributed (iid)

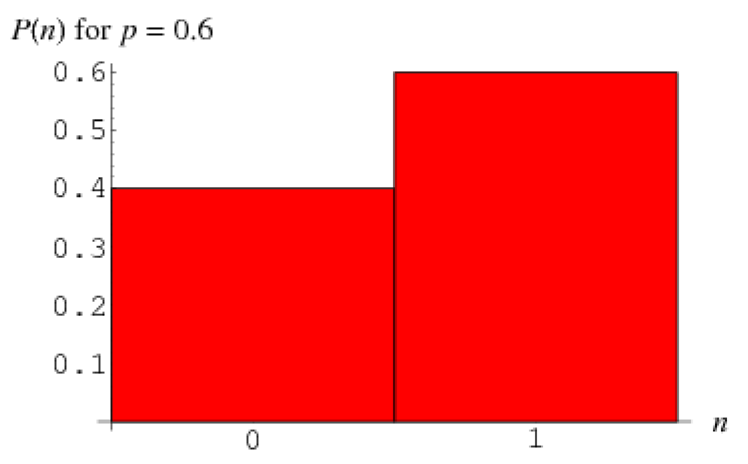
Let us consider a perfectly homogeneous radioactive sample consisting of $2N$ atoms. Let us divide it into 2 exact halves, A and B such that each piece contains exactly N atoms. Let X denotes the number of decays in A in 5 seconds and Y denotes the decays in B in 5 seconds. Quite clearly both X and Y are binomially distributed and for every $x = 0, 1, 2, 3 \ldots N$ we have,

$$P_X(X = x) = P_Y(Y = x)$$

but, X and Y are drawn from two different pieces and are obviously independent statistically. We will say, X and Y are *independent but identically distributed random variable* or *iid*.

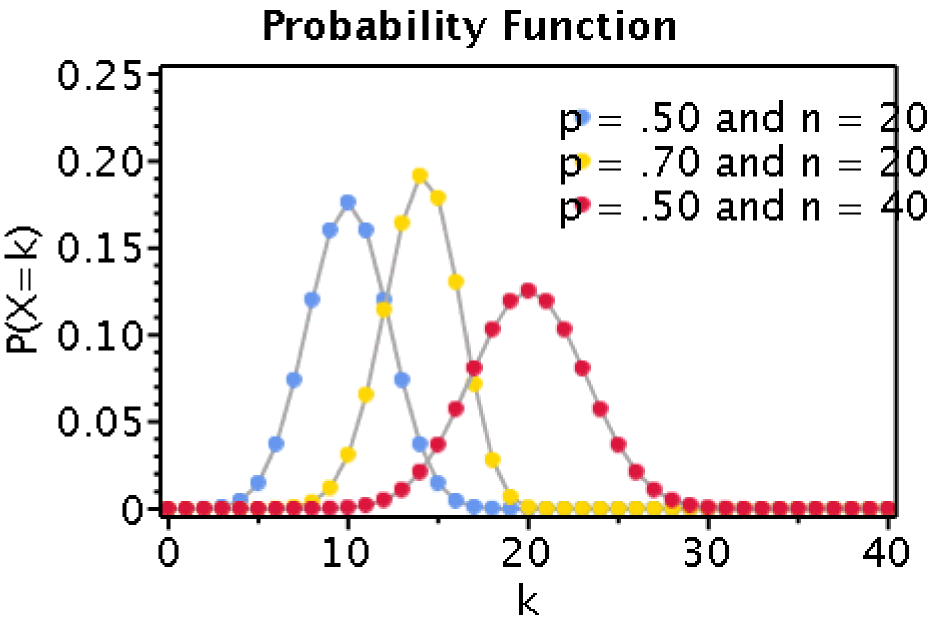
Bernoulli trial, Bernoulli distribution

- Bernoulli trial: Example coin toss
- Bernoulli random variable N has two outcomes
- $n=1$ (success), $n = 0$ (failure)
- Success probability $P(n=1) = p$ is the only parameter
- Bernoulli distribution (pmf) : $P(1) = p$, $p(0) = (1-p)$
- Can also be written as $p^n(1-p)^{1-n}$



Binomial distribution

- If we do n bernoulli trials, what is the probability of X successes? (e.g. 4 tails in 10 coin tosses)
- Pmf = Binomial(X=x; n, p) = ${}^{n}C_{x} p^{x} (1-p)^{n-x}$



Negative Binomial distribution

- Number of successes(r) is fixed parameter
- Number of trials (X) is the random variable
- binomial*(X; r,p) ~ ${}^{X-1}C_{r-1}p^r(1-p)^{X-r}$

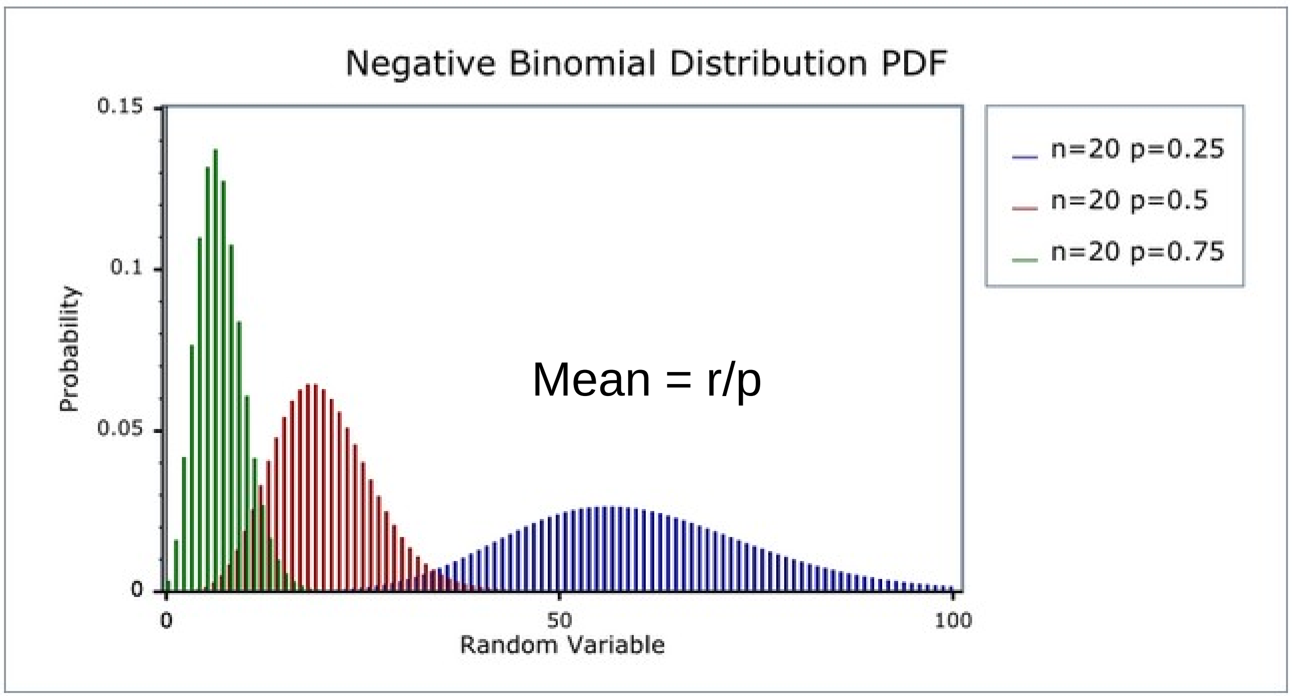


Image: UCSF

Moments of a distribution

Definition 14: Mean
The expected value/ expectation value/ expectation/ mean (all are equivalent words) of a random variable X is defined as

$$E(X) = < X > = \begin{cases} \int_{-\infty}^{\infty} x f(x) dx & if\, X\, is\, continuous \\ \sum_{x \in \chi} x f(x) & if\, X\, is\, discrete \end{cases} \tag{2}$$

For any general function of X, $g(X)$, which is also a random variable

$$< g(X) > = \begin{cases} \int_{-\infty}^{\infty} g(x) f(x) dx & if\, X\, is\, continuous \\ \sum_{x \in \chi} g(x) f(x) & if\, X\, is\, discrete \end{cases} \tag{3}$$

Let us calculate the mean when X is uniformly distributed.

$$< X > = \int_{-\infty}^{\infty} x \frac{1}{b-a} dx = \int_{a}^{b} x \frac{1}{b-a} dx = \frac{b+a}{2}$$

Therefore in the range $(0, 1)$ $< X > = \frac{1}{2}$.

Higher moments

Eq. (3) defines the expectation of any function $g(x)$ of X. $g_n(X) = < X^n >$ is a class of functions of particular interest. These quantities are called the n^{th} moment of X, we will denote them by $\mu_n^{'}$.

The n^{th} central moment of X is

$$\mu_n = < (X- < X >)^n > = <(X - \mu)^n> \quad where,\ \mu = \mu_1^{'} = < X >$$

The central moments contain information about the shape of the distribution around the mean.

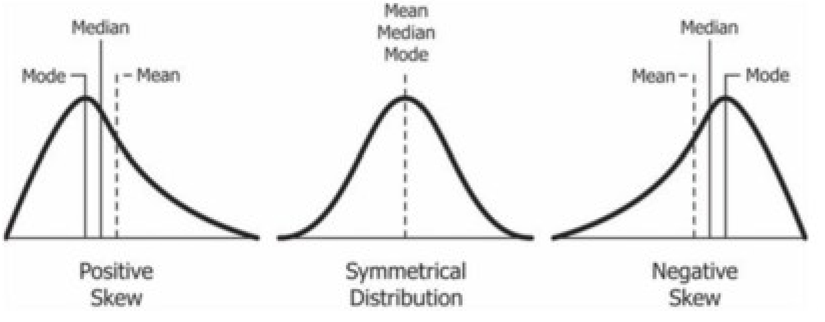
The 2^{nd} central moment, known as variance $< (X- < X >)^2 > = \mu_2$ gives a measure of how widely the random no. X is distributed about its mean μ. The square root of variance $\sigma = \sqrt{(X- < X >)^2}$ is called the *standard deviation* (s.d.).

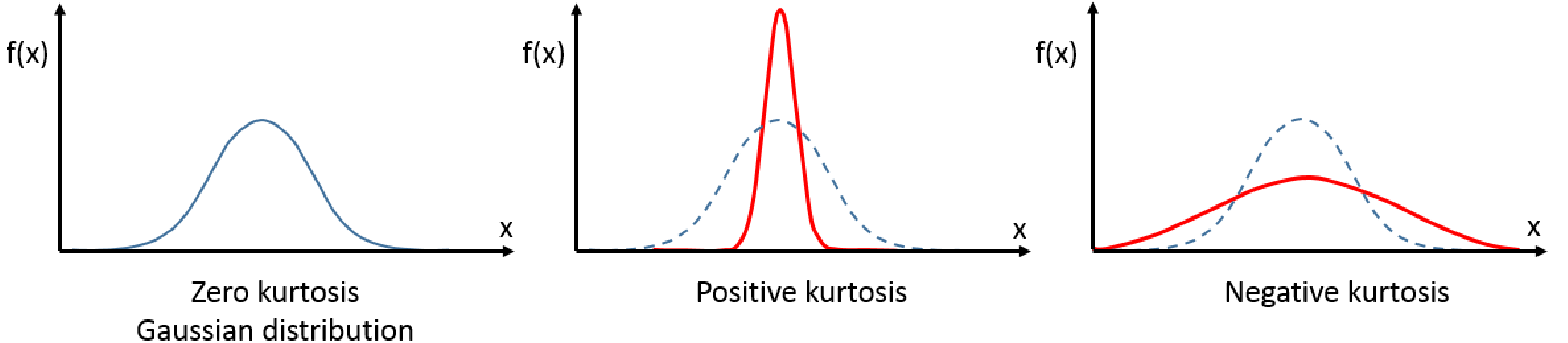
Imporant: Often $\pm\sigma$ is quoted as *statistical uncertainty* or *statistical error*.

Skewness and Kurtosis

Definition 15: Skewness & Kurtosis
$\alpha_3 = \frac{\mu_3}{(\mu_2)^{3/2}}$ where, $\mu_3 = <(X-\mu)^3>$ is called *skewness* and is a measure of how asymmetric the distribution $f_X(x)$ of X is.

$\alpha_4 = \frac{\mu_4}{(\mu_2)^2}$ where, μ_4 is the 4^{th} central moment $<(X-\mu)^4>$, is called *kurtosis* and is the measure of how sharply peaked a distribution is.





Moment generating function (MGF)

The moment generating function of a random variable X is defined as,

$$\begin{aligned} M_X(x) &= < e^{tx} > \quad \text{note that the expectation may not exist} \\ &= \int_{-\infty}^{\infty} e^{tx} f_X(x) dx \quad for\ X\ continuous \\ &= \sum_x e^{tx} P(X = x) \quad for\ X\ discrete \end{aligned}$$

To see how $M_X(t)$ generates moments, let us differentiate M_X w.r.t t,

$$\begin{aligned} \frac{dM_X(t)}{dt} &= \frac{d}{dt} \int_{-infty}^{\infty} e^{tx} f_X(x) dx \\ &= \int_{-infty}^{\infty} x e^{tx} f_X(x) dx \\ &= < x e^{tx} > \end{aligned}$$

$$\boxed{\therefore \frac{d}{dt} M_X(x)|_{t=0} = < X > \qquad \text{Similarly, } \frac{d^n}{dt^n} M_X(x) = < X^n >}$$

Bernoulli MGF

- Moment generating function of Bernoulli distribution: calculate expectation of (exp(tn))

  where n is the outcomes of Bernoulli trials

- It will have just two terms for n=0,1

$$
\begin{aligned}
M(t) &= \left\langle e^{t\,n} \right\rangle \\
&= \sum_{n=0}^{1} e^{t\,n}\, p^{n}\,(1-p)^{1-n} \\
&= e^{0}\,(1-p) + e^{t}\,p,
\end{aligned}
$$

Homework

Exercise 2.1. Show that if X is binomially distributed, its given by

$$P(X = x) = {}^{n}C_{x} p^{x} (1-p)^{n-x}$$

then $< X >= np$

Exercise 2.2. We will often encounter Cauchy distribution or Breit-Wigner distribution, which is a more generalized form of Cauchy distribution. An interesting property of a Cauchy distributed variable X is $< |X| >= \infty$. Prove this property.

Exercise 2.3. Another useful property relating the distance of a random variable X to some constant b is

$$< (X - b)^2 > = < (X - < X >)^2 > + (< X > - b)^2$$

Prove this property.

Exercise 2.4. Show that if X is a random variable then $var(aX+b) = a^2 var(X)$. Note that $var(X) =< X^2 > - < X >^2$.

Exercise 2.5. Show that the variance of $X \sim binomial(n,p)$ (i.e. X is a random variable distributed binomially, with n and p as the parameters of the binomial) is,

$$np(1-p)$$

Exercise 2.6. In physics there are many examples of quantities that follow an exponential distribution

$$f_X(x) = Ae^{-x/\tau}$$

where, A is a constant such that $\int_0^\infty f_X(x)dx = 1$ and τ is a constant. e.g. The life-time of a radioactive nuclei, free path of a high energy photon in some material before it converts to an e^+e^- pair, or the free path travelled by a charged pion (π^+, π^-) in a piece of material before it does a nuclear interaction.

(i) Exponential is a single parameter distribution, depending only on the paramter τ. Prove that the const. A (called the normalization const.) is $\dfrac{1}{\tau}$.

(ii) Find $\mu =< X >$ if X is an exponentially distributed random no.

(iii) Find $\sigma^2 = < (X-\mu)^2 >$

Exercise 2.7. A gamma distribution is given by the pdf

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} \quad where,\ 0 < x < \infty\,,\ \alpha > 0\,,\ \beta > 0$$

Prove that

$$\begin{aligned} M_X(t) &= \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \int_0^{\infty} x^{\alpha-1} e^{-x/\dfrac{\beta}{1-\beta t}} \\ &= (\frac{1}{1-\beta t})^{\alpha} \end{aligned}$$

From this prove that if X is gamma distributed then $< X > = \alpha\beta$

Exercise 2.8. Show that if $X \sim binomial(n, p)$ then

$$M_X(t) = [pe^t + (1-p)]^n$$