Random Variables

- Probability Space
  - A triple of \((\Omega, F, P)\)
    - \(\Omega\) represents a nonempty set, whose elements are sometimes known as outcomes or states of nature.
    - \(F\) represents a set, whose elements are called events. The events are subsets of \(\Omega\). \(F\) should be a “Borel Field”.
    - \(P\) represents the probability measure.

- Fact: \(P(\Omega) = 1\)
Random variable

A random variable is a number assigned to every outcome of an experiment. Prob\{x \leq x\} of an event \{x \leq x\} is a number that depends on x. This number is denoted by \( F_x(x) \) and is called CDF of RV \( x \).

Properties of CDF:

1. \( F(\infty) = 1, \quad F(-\infty) = 0 \)
2. \( x_1 \leq x_2 \implies F(x_1) \leq F(x_2) \)
3. \( \text{Prob}\{x\} = 1 - F_x(x) \)
4. \( F(x^+) = F(x) \), \( F(x) \) is continuous from the right
5. \( \text{Prob}\{x_1 \leq x \leq x_2\} = F(x_2) - F(x_1) \)
We say the statistics of an RV are known if we can determine the \( \text{Prob}\{x \in S\} \)

We say that an RV \( x \) is **continuous type** if \( F_x(x) \) is continuous.

We say that an RV \( x \) is **discrete type** if \( F_x(x) \) is staircase.

We say that an RV \( x \) is **mixed type** if \( F_x(x) \) is a combination of continuous and staircase function.

\[
f_x(x) = \frac{d}{dx} F_x(x)
\]

is the PDF for a continuous random variable, and for a discrete random variable

\[
f(x) = \sum_i p_i \delta(x - x_i) \text{ where } p_i = \text{Prob}\{x = x_i\}
\]
Properties:

1. $f(x) \geq 0$, $F(x) = \int_{-\infty}^{x} f(\xi) d\xi$,
   
   $F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x) dx$

2. $\text{Prob}\{x \leq x \leq x + \Delta x\} \approx f(x)\Delta x$

3. $f(x) = \lim_{\Delta x \to 0} \frac{\text{Prob}\{x \leq x \leq x + \Delta x\}}{\Delta x}$

The mode or the most likely value of $x$ is where $f(x)$ is maximum. An RV is unimodal if it has only a single mode.
Thus the area under $f_X(x)$ in the interval $(x_1, x_2)$ represents the probability in (3-28).

Often, r.v.s are referred by their specific density functions - both in the continuous and discrete cases - and in what follows we shall list a number of them in each category.
Continuous-type random variables

1. Normal (Gaussian): $X$ is said to be normal or Gaussian r.v, if

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2 / 2\sigma^2}. \quad (3-29)$$

This is a bell shaped curve, symmetric around the parameter $\mu$, and its distribution function is given by

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2 / 2\sigma^2} dy^\Delta = G\left(\frac{x - \mu}{\sigma}\right), \quad (3-30)$$

where $G(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$ is often tabulated. Since $f_X(x)$ depends on two parameters $\mu$ and $\sigma^2$, the notation $X \sim N(\mu, \sigma^2)$ will be used to represent (3-29).
2. Uniform: \( X \sim U(a,b), \ a < b, \) if (Fig. 3.8)

\[
f_X(x) = \begin{cases} 
\frac{1}{b-a}, & a \leq x \leq b, \\
0, & \text{otherwise.} 
\end{cases} \tag{3.31}
\]

3. Exponential: \( X \sim \varepsilon(\lambda) \) if (Fig. 3.9)

\[
f_X(x) = \begin{cases} 
\frac{1}{\lambda} e^{-x/\lambda}, & x \geq 0, \\
0, & \text{otherwise.} 
\end{cases} \tag{3-32}
\]
4. Gamma: \( X \sim G(\alpha, \beta) \) if \((\alpha > 0, \beta > 0)\) (Fig. 3.10)

\[
f_X(x) = \begin{cases} 
\frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha) \beta^\alpha}, & x \geq 0, \\
0, & \text{otherwise.}
\end{cases} \tag{3-33}
\]

If \( \alpha = n \) an integer \( \Gamma(n) = (n-1)! \).

5. Beta: \( X \sim \beta(a,b) \) if \((a > 0, b > 0)\) (Fig. 3.11)

\[
f_X(x) = \begin{cases} 
\frac{1}{\beta(a,b)} x^{a-1} (1-x)^{b-1}, & 0 < x < 1, \\
0, & \text{otherwise.}
\end{cases} \tag{3-34}
\]

where the Beta function \( \beta(a,b) \) is defined as

\[
\beta(a,b) = \int_0^1 u^{a-1} (1-u)^{b-1} du. \tag{3-35}
\]
6. Chi-Square: $X \sim \chi^2(n)$, if (Fig. 3.12)

$$f_X(x) = \begin{cases} \frac{1}{2^{n/2} \Gamma(n/2)} x^{n/2-1} e^{-x/2}, & x \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$  \hspace{1cm} (3-36)

Note that $\chi^2(n)$ is the same as Gamma $(n/2, 2)$.

7. Rayleigh: $X \sim R(\sigma^2)$, if (Fig. 3.13)

$$f_X(x) = \begin{cases} \frac{x}{\sigma^2} e^{-x^2/2\sigma^2}, & x \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$ \hspace{1cm} (3-37)

8. Nakagami – $m$ distribution:

$$f_X(x) = \begin{cases} \frac{2}{\Gamma(m) \Omega} \left(\frac{m}{\Omega}\right)^m x^{2m-1} e^{-mx^2/\Omega}, & x \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$ \hspace{1cm} (3-38)

PILLAI
9. Cauchy: \( X \sim C(\alpha, \mu) \), if (Fig. 3.14)

\[
f_x(x) = \frac{\alpha / \pi}{\alpha^2 + (x - \mu)^2}, \quad -\infty < x < +\infty. \tag{3-39}
\]

10. Laplace: (Fig. 3.15)

\[
f_x(x) = \frac{1}{2\lambda} e^{-|x|/\lambda}, \quad -\infty < x < +\infty. \tag{3-40}
\]

11. Student’s \( t \)-distribution with \( n \) degrees of freedom (Fig 3.16)

\[
f_T(t) = \frac{\Gamma((n + 1)/2)}{\sqrt{n \pi} \Gamma(n/2)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}, \quad -\infty < t < +\infty. \tag{3-41}
\]
Discrete-type random variables

1. Bernoulli: $X$ takes the values $(0,1)$, and

$$P(X = 0) = q, \quad P(X = 1) = p.$$  \hspace{1cm} (3-43)

2. Binomial: $X \sim B(n,p)$, if (Fig. 3.17)

$$P(X = k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0,1,2,\ldots,n.$$  \hspace{1cm} (3-44)

3. Poisson: $X \sim P(\lambda)$, if (Fig. 3.18)

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0,1,2,\ldots,\infty.$$  \hspace{1cm} (3-45)
4. Hypergeometric:

\[
P(X = k) = \binom{m}{k} \binom{N-m}{n-k} \binom{N}{n}, \quad \max(0, m+n-N) \leq k \leq \min(m,n) \quad (3-46)
\]

5. Geometric: \( X \sim g(p) \) if

\[
P(X = k) = pq^k, \quad k = 0, 1, 2, \ldots, \infty, \quad q = 1 - p. \quad (3-47)
\]

6. Negative Binomial: \( X \sim NB(r, p) \), if

\[
P(X = k) = \binom{k-1}{r-1} p^r q^{k-r}, \quad k = r, r+1, \ldots \quad (3-48)
\]

7. Discrete-Uniform:

\[
P(X = k) = \frac{1}{N}, \quad k = 1, 2, \ldots, N. \quad (3-49)
\]

We conclude this lecture with a general distribution due
Normal: \[ f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left( -\frac{(x-\eta)^2}{2\sigma^2} \right) \]

Uniform: \[ f(x) = \begin{cases} \frac{1}{x_2-x_1}, & x_1 \leq x \leq x_2 \\ 0, & \text{otherwise} \end{cases} \]

Rayleigh: \[ f(x) = \frac{x}{\sigma^2} e^{-x^2/2\sigma^2}, \quad x \geq 0 \]

Lognormal: \[ f(x) = \frac{1}{\sigma x \sqrt{2\pi}} e^{-\frac{(\ln{x}-\eta)^2}{2\sigma^2}}, \quad x \geq 0 \]

Cauchy: \[ f(x) = \frac{1}{\pi (x^2 + 1)} \]

Gamma: \[ f(x) = \frac{c^{b+1}}{\Gamma(b+1)} x^b e^{-cx}, \quad x \geq 0, \quad \Gamma(b+1) = b\Gamma(b), \]

if \( b \)= an integer, it is called Erlang density.
Laplace: \( f(x) = 0.5e^{-|x|} \)

Chi and Chi-square: \( \chi = \sqrt{\sum_{i=1}^{n} x_i^2}, \quad y = \chi^2, \quad f(\chi) = 2a\chi^{n-1}e^{-\chi^2/2\sigma^2} \quad f(y) = ay^{n/2-1}e^{-y/2\sigma^2}, \quad a = \frac{1}{\Gamma(n/2)(\sigma\sqrt{2})^2} \)

Geometric: \( \text{Prob}\{x = k\} = pq^k, \quad k = 0, 1, \cdots, \infty \)

Binomial: 
\[
\text{Prob}\{x = k\} = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \cdots, n
\]

\( x \) is of lattice type and its density is a sum of impulses,
\[
f(x) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \delta(x - k)
\]
Negative Binomial:
\[ \binom{n + k - 1}{k} p^n (1 - p)^k, \quad k = 0, 1, \ldots, \infty \]

Poisson: \( \text{Prob}\{x = k\} = e^{-a} \frac{a^k}{k!}, \quad k = 0, 1, \ldots \)

The density function is \( \text{Prob}\{x = k\} = e^{-a} \sum_{k=0}^{\infty} \frac{a^k}{k!} \delta(x - k) \)

Example 1:
Given a constant \( t_0 \), we define a RV \( n \) such that its value equals the number of points in the interval \((0, t_0)\), find the probability that the number of points in this interval is \( k \)

Solution:
\[ \text{Prob}\{n = k\} = e^{-\lambda t_0} \frac{(\lambda t_0)^k}{k!} \]
Example 2:
If $t_1$ is the first random point to the right of the fixed point $t_0$ and we define RV $x$ as the distance from $t_0$ to $t_1$, determine the PDF and CDF for $x$.

Solution:

$$F(x) = \text{probability that there are at least one point between } t_0 \text{ and } t_0 + x,$$

$$1 - F(x) \text{ is the probability that there are no points} = \text{Prob}\{n = 0\} = e^{-\lambda x}, \quad F(x) = 1 - e^{-\lambda x}, \quad f(x) = \lambda e^{-\lambda x} u(x)$$
Let $X$ represent a Binomial r.v as in (3-42). Then from (2-30)

$$P(k_1 \leq X \leq k_2) = \sum_{k=k_1}^{k_2} P_n(k) = \sum_{k=k_1}^{k_2} \binom{n}{k} p^k q^{n-k}. \quad (4-1)$$

Since the binomial coefficient $\binom{n}{k} = \frac{n!}{(n-k)!k!}$ grows quite rapidly with $n$, it is difficult to compute (4-1) for large $n$. In this context, two approximations are extremely useful.

4.1 The Normal Approximation (Demoivre-Laplace Theorem) Suppose $n \to \infty$ with $p$ held fixed. Then for $k$ in the $\sqrt{npq}$ neighborhood of $np$, we can approximate...
\[
\binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}.
\] (4-2)

Thus if \( k_1 \) and \( k_2 \) in (4-1) are within or around the neighborhood of the interval \((np - \sqrt{npq}, np + \sqrt{npq})\), we can approximate the summation in (4-1) by an integration. In that case (4-1) reduces to

\[
P(k_1 \leq X \leq k_2) = \int_{k_1}^{k_2} \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(x-np)^2}{2npq}} \, dx = \int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \, dy,
\] (4-3)

where

\[
x_1 = \frac{k_1 - np}{\sqrt{npq}}, \quad x_2 = \frac{k_2 - np}{\sqrt{npq}}.
\]

We can express (4-3) in terms of the normalized integral

\[
\text{erf} \ (x) = \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-\frac{y^2}{2}} \, dy = \text{erf} \ (-x)
\] (4-4)

that has been tabulated extensively (See Table 4.1).
For example, if \( x_1 \) and \( x_2 \) are both positive, we obtain

\[
P(k_1 \leq X \leq k_2) = \text{erf}(x_2) - \text{erf}(x_1).
\] (4-5)

Example 4.1: A fair coin is tossed 5,000 times. Find the probability that the number of heads is between 2,475 to 2,525.

Solution: We need \( P(2,475 \leq X \leq 2,525) \). Here \( n \) is large so that we can use the normal approximation. In this case \( p = \frac{1}{2} \), so that \( np = 2,500 \) and \( \sqrt{npq} \approx 35 \). Since \( np - \sqrt{npq} = 2,465 \), and \( np + \sqrt{npq} = 2,535 \), the approximation is valid for \( k_1 = 2,475 \) and \( k_2 = 2,525 \). Thus

\[
P(k_1 \leq X \leq k_2) = \int_{x_1}^{x_2} \frac{1}{2\pi} e^{-y^2/2} dy.
\]

Here \( x_1 = \frac{k_1 - np}{\sqrt{npq}} = -\frac{5}{7} \), \( x_2 = \frac{k_2 - np}{\sqrt{npq}} = \frac{5}{7} \).
\[ \text{erf}(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-y^2/2} \, dy = G(x) - \frac{1}{2} \]

<table>
<thead>
<tr>
<th>( x )</th>
<th>( \text{erf}(x) )</th>
<th>( x )</th>
<th>( \text{erf}(x) )</th>
<th>( x )</th>
<th>( \text{erf}(x) )</th>
<th>( x )</th>
<th>( \text{erf}(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.01994</td>
<td>0.80</td>
<td>0.28814</td>
<td>1.55</td>
<td>0.43943</td>
<td>2.30</td>
<td>0.48928</td>
</tr>
<tr>
<td>0.10</td>
<td>0.03983</td>
<td>0.85</td>
<td>0.30234</td>
<td>1.60</td>
<td>0.44520</td>
<td>2.35</td>
<td>0.49061</td>
</tr>
<tr>
<td>0.15</td>
<td>0.05962</td>
<td>0.90</td>
<td>0.31594</td>
<td>1.65</td>
<td>0.45053</td>
<td>2.40</td>
<td>0.49180</td>
</tr>
<tr>
<td>0.20</td>
<td>0.07926</td>
<td>0.95</td>
<td>0.32894</td>
<td>1.70</td>
<td>0.45543</td>
<td>2.45</td>
<td>0.49286</td>
</tr>
<tr>
<td>0.25</td>
<td>0.09871</td>
<td>1.00</td>
<td>0.34134</td>
<td>1.75</td>
<td>0.45994</td>
<td>2.50</td>
<td>0.49379</td>
</tr>
<tr>
<td>0.30</td>
<td>0.11791</td>
<td>1.05</td>
<td>0.35314</td>
<td>1.80</td>
<td>0.46407</td>
<td>2.55</td>
<td>0.49461</td>
</tr>
<tr>
<td>0.35</td>
<td>0.13683</td>
<td>1.10</td>
<td>0.36433</td>
<td>1.85</td>
<td>0.46784</td>
<td>2.60</td>
<td>0.49534</td>
</tr>
<tr>
<td>0.40</td>
<td>0.15542</td>
<td>1.15</td>
<td>0.37493</td>
<td>1.90</td>
<td>0.47128</td>
<td>2.65</td>
<td>0.49597</td>
</tr>
<tr>
<td>0.45</td>
<td>0.17364</td>
<td>1.20</td>
<td>0.38493</td>
<td>1.95</td>
<td>0.47441</td>
<td>2.70</td>
<td>0.49653</td>
</tr>
<tr>
<td>0.50</td>
<td>0.19146</td>
<td>1.25</td>
<td>0.39435</td>
<td>2.00</td>
<td>0.47726</td>
<td>2.75</td>
<td>0.49702</td>
</tr>
<tr>
<td>0.55</td>
<td>0.20884</td>
<td>1.30</td>
<td>0.40320</td>
<td>2.05</td>
<td>0.47982</td>
<td>2.80</td>
<td>0.49744</td>
</tr>
<tr>
<td>0.60</td>
<td>0.22575</td>
<td>1.35</td>
<td>0.41149</td>
<td>2.10</td>
<td>0.48214</td>
<td>2.85</td>
<td>0.49781</td>
</tr>
<tr>
<td>0.65</td>
<td>0.24215</td>
<td>1.40</td>
<td>0.41924</td>
<td>2.15</td>
<td>0.48422</td>
<td>2.90</td>
<td>0.49813</td>
</tr>
<tr>
<td>0.70</td>
<td>0.25804</td>
<td>1.45</td>
<td>0.42647</td>
<td>2.20</td>
<td>0.48610</td>
<td>2.95</td>
<td>0.49841</td>
</tr>
<tr>
<td>0.75</td>
<td>0.27337</td>
<td>1.50</td>
<td>0.43319</td>
<td>2.25</td>
<td>0.48778</td>
<td>3.00</td>
<td>0.49865</td>
</tr>
</tbody>
</table>

Table 4.1
Since \( x_1 < 0 \), from Fig. 4.1(b), the above probability is given by \( P(2.475 \leq X \leq 2.525) = \text{erf}(x_2) - \text{erf}(x_1) = \text{erf}(x_2) + \text{erf}(|x_1|) \)

\[
= 2\text{erf}\left(\frac{5}{7}\right) = 0.516,
\]

where we have used Table 4.1 (\( \text{erf}(0.7) = 0.258 \)).

4.2. The Poisson Approximation
As we have mentioned earlier, for large \( n \), the Gaussian approximation of a binomial r.v is valid only if \( p \) is fixed, i.e., only if \( np >> 1 \) and \( npq >> 1 \). What if \( np \) is small, or if it does not increase with \( n \)?
Obviously that is the case if, for example, $p \rightarrow 0$ as $n \rightarrow \infty$, such that $np = \lambda$ is a fixed number.

Many random phenomena in nature in fact follow this pattern. Total number of calls on a telephone line, claims in an insurance company etc. tend to follow this type of behavior. Consider random arrivals such as telephone calls over a line. Let $n$ represent the total number of calls in the interval $(0, T)$. From our experience, as $T \rightarrow \infty$ we have $n \rightarrow \infty$ so that we may assume $n = \mu T$. Consider a small interval of duration $\Delta$ as in Fig. 4.2. If there is only a single call coming in, the probability $p$ of that single call occurring in that interval must depend on its relative size with respect to $T$.

![Fig. 4.2](image-url)
Hence we may assume $p = \frac{\Delta}{T}$. Note that $p \to 0$ as $T \to \infty$. However in this case $np = \mu T \cdot \frac{\Delta}{T} = \mu \Delta = \lambda$ is a constant, and the normal approximation is invalid here.

Suppose the interval $\Delta$ in Fig. 4.2 is of interest to us. A call inside that interval is a “success” ($H$), whereas one outside is a “failure” ($T$). This is equivalent to the coin tossing situation, and hence the probability $P_n(k)$ of obtaining $k$ calls (in any order) in an interval of duration $\Delta$ is given by the binomial p.m.f. Thus

$$P_n(k) = \frac{n!}{(n-k)!k!} p^k (1 - p)^{n-k}, \quad (4-6)$$

and here as $n \to \infty$, $p \to 0$ such that $np = \lambda$. It is easy to obtain an excellent approximation to (4-6) in that situation. To see this, rewrite (4-6) as

$$T \Delta = \lambda = np \to \infty \implies p \to 0.$$
\[ P_n(k) = \frac{n(n-1) \cdots (n-k+1)}{n^k} \frac{(np)^k}{k!} (1 - np/n)^{n-k} \]

\[ = \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \frac{\lambda^k}{k!} \frac{(1 - \lambda/n)^n}{(1 - \lambda/n)^k}. \]  

Thus

\[
\lim_{n \to \infty, p \to 0, np=\lambda} P_n(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \tag{4-8}
\]

since the finite products \( \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \) as well as \( \left(1 - \frac{\lambda}{n}\right)^k \) tend to unity as \( n \to \infty \), and

\[
\lim_{n \to \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}.
\]

The right side of (4-8) represents the Poisson p.m.f and the Poisson approximation to the binomial r.v is valid in situations where the binomial r.v parameters \( n \) and \( p \) diverge to two extremes \( (n \to \infty, p \to 0) \) such that their product \( np \) is a constant.
Example 4.2: Winning a Lottery: Suppose two million lottery tickets are issued with 100 winning tickets among them. (a) If a person purchases 100 tickets, what is the probability of winning? (b) How many tickets should one buy to be 95% confident of having a winning ticket?

Solution: The probability of buying a winning ticket

\[ p = \frac{\text{No. of winning tickets}}{\text{Total no. of tickets}} = \frac{100}{2 \times 10^6} = 5 \times 10^{-5}. \]

Here \( n = 100 \), and the number of winning tickets \( X \) in the \( n \) purchased tickets has an approximate Poisson distribution with parameter \( \lambda = np = 100 \times 5 \times 10^{-5} = 0.005 \). Thus

\[ P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \]

and (a) Probability of winning \( = P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-\lambda} \approx 0.005. \)
(b) In this case we need \( P(X \geq 1) \geq 0.95 \).

\[
P(X \geq 1) = 1 - e^{-\lambda} \geq 0.95 \quad \text{implies} \quad \lambda \geq \ln 20 = 3.
\]

But \( \lambda = np = n \times 5 \times 10^{-5} \geq 3 \) or \( n \geq 60,000 \). Thus one needs to buy about 60,000 tickets to be 95% confident of having a winning ticket!

Example 4.3: A space craft has 100,000 components \((n \to \infty)\)
The probability of any one component being defective is \( 2 \times 10^{-5} (p \to 0) \). The mission will be in danger if five or more components become defective. Find the probability of such an event.

Solution: Here \( n \) is large and \( p \) is small, and hence Poisson approximation is valid. Thus \( np = \lambda = 100,000 \times 2 \times 10^{-5} = 2 \), and the desired probability is given by
\[ P(X \geq 5) = 1 - P(X \leq 4) = 1 - \sum_{k=0}^{4} e^{-\lambda} \frac{\lambda^k}{k!} = 1 - e^{-2} \sum_{k=0}^{4} \frac{\lambda^k}{k!} \]

\[ = 1 - e^{-2} \left( 1 + 2 + 2 + \frac{4}{3} + \frac{2}{3} \right) = 0.052. \]
DeMoivre-Laplace theorem

if \( npq \gg 1 \), and \( np - \sqrt{npq} \leq k \leq np + \sqrt{npq} \) \( \Rightarrow \)

\[
\binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sqrt{2\pi npq}} \exp \left( \frac{-(k - np)^2}{2npq} \right)
\]

Example:
A fair coin is tossed 1000 times. Find the probability that heads show 510 times.
\( n = 1000, \ p = 0.5, \ np = 500, \ \sqrt{npq} = 5\sqrt{10} = 15.81 \)

Solution:
\[
\binom{1000}{500} \cdot 0.5^{500} \cdot 0.5^{500} \approx \frac{e^{-100/500}}{\sqrt{2\pi npq}} = 0.0207
\]
Example:
A fair coin is tossed $10^4$ times. Find the probability that number of heads is between 4995 and 5005.

$n = 10^4$, $np = 5000$, $npq = 50$, $\sqrt{npq} = 7.07$

Solution:

$$\text{Prob}\{4995 \leq k \leq 5005\} = \sum_{k=4995}^{5005} \binom{10^4}{k} 0.5^k 0.5^{10^4-k}$$

$$\sum_{k=k_1}^{k_2} g \left( \frac{k - np}{\sqrt{npq}} \right) \approx \int_{k_1}^{k_2} g \left( \frac{x - np}{\sqrt{npq}} \right) dx$$

$$= G \left( \frac{k_2 - np}{\sqrt{npq}} \right) - G \left( \frac{k_1 - np}{\sqrt{npq}} \right)$$

$$= 0.041$$
Law of large numbers

An event $\mathcal{A}$ with $\text{Prob}\{\mathcal{A}\} = p$ occurs $k$ times in $n$ trials $\Rightarrow k \simeq np$, this is a heuristic statement. Let $\mathcal{A}$ denote an event whose probability of occurrence in a single trial is $p$. If $k$ denotes the number of occurrences of $\mathcal{A}$ in $n$ independent trials, then

$$\lim_{n \to \infty} \text{Prob}\{k = np\} \simeq \frac{1}{\sqrt{2\pi npq}} \to 0 \text{ Never occurs!}$$

The approximation $k \simeq np$ means that the ratio $k/n$ is close to $p$ in the sense that, for any $\epsilon > 0$,

$$\lim_{n \to \infty} \text{Prob} \left\{ \left| \frac{k}{n} - p \right| < \epsilon \right\} \to 1, \quad \forall \epsilon > 0$$
\[
\left| \frac{k}{n} - p \right| < \epsilon \Rightarrow -\epsilon \leq \frac{k}{n} - p \leq \epsilon \Rightarrow n(p - \epsilon) \leq k \leq n(p + \epsilon)
\]

\[
\text{Prob}\left\{ \left| \frac{k}{n} - p \right| < \epsilon \right\} = \text{Prob}\{k_1 \leq k \leq k_2\}
\]

\[
\sum_{k=k_1}^{k_2} \binom{n}{k} p^k q^{n-k} \approx G\left(\frac{k_2 - np}{\sqrt{npq}}\right) - G\left(\frac{k_1 - np}{\sqrt{npq}}\right), \quad \begin{cases} k_2 - np = n\epsilon, \\ k_1 - np = n\epsilon \end{cases}
\]

\[
= 2G\left(\frac{n\epsilon}{\sqrt{npq}}\right) - 1, \quad G(-x) = 1 - G(x)
\]

\[
\lim_{n \to \infty} \left\{ 2G\left(\frac{n\epsilon}{\sqrt{npq}}\right) - 1 \right\} = 2 - 1 = 1
\]
Example: \( p = 0.5, \ \epsilon = 0.05 \)

\[
n(p - \epsilon) = 0.45n, \quad n(p + \epsilon) = 0.55n, \quad \epsilon \sqrt{\frac{n}{pq}} = 0.1 \sqrt{n}
\]

Solution:

<table>
<thead>
<tr>
<th>( n )</th>
<th>100</th>
<th>900</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1 ( \sqrt{n} )</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>( 2G(0.1 \sqrt{n}) - 1 )</td>
<td>0.682</td>
<td>0.997</td>
</tr>
</tbody>
</table>

The last row indicates that after 900 independent trials we may have some confidence in accepting \( \frac{k}{n} \approx p \)
Generalized Bernoulli trials

\[ U = \{ A_1 \text{ occurs } k_1 \text{ times, } A_2 \text{ occurs } k_2 \text{ times, } \ldots, A_r \text{ occurs } k_r \text{ times} \} \]

The number of occurrence of \( U \) is

\[ \frac{n!}{k_1!k_2!\cdots k_r!}, \quad n = \sum_{i=1}^{r} k_i \]

Since the trials are independent, the probability of each event is \( p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r} \Rightarrow \)

\[ \text{Prob}\{U\} = \frac{n!}{k_1!k_2!\cdots k_r!} p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r} \]
Example:
A fair die is rolled 10 times. Determine the probability that ones shows 3 times, and an even number shows 6 times.
Solution:

\[ A_1 = \{1\}, \ A_2 = \{2, 4, 6\}, \ A_3 = \{3, 5\} \Rightarrow \]

\[ p_1 = \frac{1}{6}, \ p_2 = \frac{3}{6}, \ p_3 = \frac{2}{6} \]

\[ k_1 = 3, \ k_2 = 6, \ k_3 = 1 \]

\[ \text{Prob}\{U\} = \frac{10!}{3!6!1!} \left( \frac{1}{6} \right)^3 \left( \frac{1}{2} \right)^6 \left( \frac{1}{3} \right) = 0.0203 \]
Poisson theorem

\[ \text{Prob\{An event } A \text{ occurs } k \text{ times in } n \text{ trials}\} = \binom{n}{k} p^k q^{n-k} \]

if \( p \ll 1 \) and \( n \to \infty \Rightarrow np \approx npq >> 1 \). However, if \( np \) is of order 1, then the Gaussian approximation is no longer valid. We use the following

\[ \binom{n}{k} p^k q^{n-k} \approx e^{-np}(np)^k \]

Poisson theorem

if \( k \) is of order \( np \), then \( k \ll n \) and \( kp \ll 1 \) and
\[ n(n-1)(n-2) \cdots (n-k+1) \approx n \cdot n \cdot n \cdots n = n^k \]
and
\[ q = 1 - p \approx e^{-p}, q^{n-k} \approx e^{-(n-k)p} \approx e^{-np} \]. Hence, we have

\[ \binom{n}{k} p^k q^{n-k} \approx e^{-np}(np)^k \]

\[ k! \]
Binomial$(n, k) = \text{as } n \to \infty, \ p \to 0$

\[
\frac{n!}{(n - k)!k!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k}
\]

\[
\frac{\sqrt{2\pi} ne^{-n} n^n}{\sqrt{2\pi(n - k)^{n-k+0.5} e^{-n+k} n^k}} \frac{\lambda^k}{n^k} e^{-\lambda}
\]

\[
\frac{1}{(1 - \frac{\lambda}{n})^n} \frac{\lambda^k}{k!} e^{-\lambda}
\]
\[ n \to \infty, \quad p \to 0, \quad np \to a \]

\[
\left( \begin{array}{c} n \\ k \end{array} \right) p^k q^{n-k} \to e^{-a} \frac{a^k}{k!}
\]

Example:
A system contains 1000 components. Each component fails independently of the others and the probability its failure in one month equals \(10^{-3}\). Find the probability that the system will function at the end of one month.

Solution: This can be considered as a problem in repeated trials with \(p = 10^{-3}, q = 0.999, n = 1000, k = 0\)
\[
\text{Prob}\{k = 0\} = \binom{1000}{0} p^0 q^{1000} = 0.999^{1000}, \quad \text{Exact}
\]

\[
\text{Prob}\{k = 0\} \approx e^{-1} \frac{(np)^0}{0!} = 0.368
\]

Applying the same idea as before:

\[
\text{Prob}\{k_1 \leq k \leq k_2\} = \sum_{k=k_1}^{k_2} \binom{n}{k} p^k q^{n-k} \approx e^{-np} \sum_{k=k_1}^{k_2} \frac{(np)^k}{k!}
\]
Conditional distribution

Baye's rule: \( \text{Prob}\{A|B\} = \frac{\text{Prob}\{AB\}}{\text{Prob}\{B\} \neq 0} \Rightarrow \text{conditional} \)

CDF: \( F(x|B) = \text{Prob}\{x \leq x|B\} = \frac{\text{Prob}\{x \leq x, B\}}{\text{Prob}\{B\}} \)

\( \{x \leq x, B\} \) is the intersection of \( \{x \leq x\} \) and \( B \)

\( F(\infty|B) = 1, F(-\infty|B) = 0, \text{Prob}\{x_1 \leq x \leq x_2|B\} = F(x \leq x_2|B) - F(x \leq x_1|B) = \frac{\text{Prob}\{x_1 \leq x \leq x_2, B\}}{\text{Prob}(B)} \)

Conditional PDF: \( f(x|B) = \frac{d}{dx} F(x|B) \). To find \( F(x|B) \) in general we must know about the experiment. However, if \( B \) can be expressed in terms of \( x \), then, for determination of \( F(x|B) \), knowledge of \( F(x) \) is enough.
Important cases

\[ B = \{ x \leq a \}, \quad F(x|B) = \frac{\text{Prob}\{x \leq x, x \leq a\}}{\text{Prob}\{x \leq a\}}, \quad \text{if } x \geq a \Rightarrow \]

\[ \{x \leq x, x \leq a\} = \{x \leq a\} \Rightarrow F(x|B) = 1, \quad x \geq a, \]

if \( x < a \) \( \Rightarrow \)

\[ \{x \leq x, x \leq a\} = \{x \leq x\} \Rightarrow F(x|x \leq a) = \frac{F(x)}{F(a)}, \quad x < a \]

\[ f(x|x \leq a) = \begin{cases} \frac{d}{dx}\{F(x|x \leq a)\} = \frac{f(x)}{F(a)}, & x < a \\ 0, & x \geq a \end{cases} \]

Example: Determine \( f(x|x - \eta| \leq k\sigma), x \sim \mathcal{N}(\eta; \sigma) \)

Solution:

\[ -k\sigma + \eta \leq x \leq k\sigma + \eta \Rightarrow f(x|x - \eta| \leq k\sigma) = \frac{f(x)}{F(|x - \eta| \leq k\sigma)} = \frac{\mathcal{N}(\eta; \sigma)}{G(k) - G(-k)}, \quad \text{if } x \ni |x - \eta| \leq k\sigma \Rightarrow \\
\]

\[ f(x|x - \eta| \leq k\sigma) = 0 \]
Example 4.5: Given $F_X(x)$, suppose $B = \{X(\xi) \leq a\}$. Find $f_X(x \mid B)$.

Solution: We will first determine $F_X(x \mid B)$. From (4-11) and $B$ as given above, we have

$$F_X(x \mid B) = \frac{P\{(X \leq x) \cap (X \leq a)\}}{P(X \leq a)}.$$  \hspace{1cm} (4-18)
For $x < a$, $(X \leq x) \cap (X \leq a) = (X \leq x)$ so that

$$F_x(x \mid B) = \frac{P(X \leq x)}{P(X \leq a)} = \frac{F_x(x)}{F_x(a)}. \quad (4-19)$$

For $x \geq a$, $(X \leq x) \cap (X \leq a) = (X \leq a)$ so that $F_x(x \mid B) = 1$.

Thus

$$F_x(x \mid B) = \begin{cases} \frac{F_x(x)}{F_x(a)}, & x < a, \\ 1, & x \geq a, \end{cases} \quad (4-20)$$

and hence

$$f_x(x \mid B) = \frac{d}{dx} F_x(x \mid B) = \begin{cases} \frac{f_x(x)}{F_x(a)}, & x < a, \\ 0, & \text{otherwise}. \end{cases} \quad (4-21)$$
Example 4.6: Let $B$ represent the event $\{a < X(\xi) \leq b\}$ with $b > a$. For a given $F_X(x)$, determine $F_X(x \mid B)$ and $f_X(x \mid B)$.

Solution:

\[
F_X(x \mid B) = P\{X(\xi) \leq x \mid B\} = \frac{P\{(X(\xi) \leq x) \cap (a < X(\xi) \leq b)\}}{P(a < X(\xi) \leq b)}
\]

\[
= \frac{P\{(X(\xi) \leq x) \cap (a < X(\xi) \leq b)\}}{F_X(b) - F_X(a)}.
\]  \hspace{1cm} (4-22)

For $x < a$, we have $\{X(\xi) \leq x\} \cap \{a < X(\xi) \leq b\} = \phi$, and hence $F_X(x \mid B) = 0$.  \hspace{1cm} (4-23)
For \( a \leq x < b \), we have \( \{ X(\xi) \leq x \} \cap \{ a < X(\xi) \leq b \} = \{ a < X(\xi) \leq x \} \) and hence

\[
F_X(x \mid B) = \frac{P(a < X(\xi) \leq x)}{F_X(b) - F_X(a)} = \frac{F_X(x) - F_X(a)}{F_X(b) - F_X(a)}.
\]  

(4-24)

For \( x \geq b \), we have \( \{ X(\xi) \leq x \} \cap \{ a < X(\xi) \leq b \} = \{ a < X(\xi) \leq b \} \) so that \( F_X(x \mid B) = 1 \).

(4-25)

Using (4-23)-(4-25), we get (see Fig. 4.5)

\[
f_X(x \mid B) = \begin{cases} 
\frac{f_X(x)}{F_X(b) - F_X(a)}, & a < x \leq b, \\
0, & \text{otherwise}. 
\end{cases}
\]  

(4-26)

![Fig. 4.5](image-url)
Total probability

If \( \{A_1, A_2, \cdots, A_n\} \) are disjoint and partition the whole space:

\[
\text{Prob}\{x \leq x\} = \sum_{i=1}^{n} \text{Prob}\{x \leq x | A_i\} \text{Prob}(A_i)
\]

\[
F(x) = \sum_{i=1}^{n} F(x \leq x | A_i) \text{Prob}(A_i)
\]

\[
f(x) = \sum_{i=1}^{n} f(x \leq x | A_i) \text{Prob}(A_i)
\]
Gaussian mixture

Binary case:
\[ f(x|B) = \mathcal{N}(\eta_1; \sigma_1), \quad f(x|\bar{B}) = \mathcal{N}(\eta_2; \sigma_2) \Rightarrow \]
\[ f(x) = p\mathcal{N}(\eta_1; \sigma_1) + (1 - p)\mathcal{N}(\eta_2; \sigma_2) \]

\( f(x) \) is a multimodal distribution.
Generally, we can have:
\[ f(x) = \sum_{i=1}^{n} p_i\mathcal{N}(\eta_i; \sigma_i), \quad \sum_{i=1}^{n} p_i = 1 \]
The $\text{Prob}\{A|x = x\}$ cannot be defined. But, it can be defined as a limit.

\[
\begin{align*}
\text{Prob}\{A|x_1 \leq x \leq x_2\} &= \frac{\text{Prob}\{x_1 \leq x \leq x_2|A\}\text{Prob}\{A\}}{\text{Prob}\{x_1 \leq x \leq x_2\}} \\
&= \frac{F(x_2|A) - F(x_1|A)}{F(x_2) - F(x_1)}\text{Prob}\{A\} \quad (1)
\end{align*}
\]

Let $x = x_1$ and $x + \Delta x = x_2$ and divide the numerator and denominator of (1) by $\Delta x \to 0$, then, we have

\[
\text{Prob}\{A|x = x\} = \frac{f(x|A)}{f(x)}\text{Prob}\{A\} \quad (2)
\]
Bayes’ theorem & applications

We can use the conditional p.d.f together with the Bayes’ theorem to update our a-priori knowledge about the probability of events in presence of new observations. Ideally, any new information should be used to update our knowledge. As we see in the next example, conditional p.d.f together with Bayes’ theorem allow systematic updating. For any two events $A$ and $B$, Bayes’ theorem gives

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}.$$
From (2), we have

\[ f(x|A) = \frac{\text{Prob}\{A|x = x\} f(x)}{\text{Prob}\{A\}} = \frac{\text{Prob}\{A|x = x\}}{\int_{-\infty}^{\infty} \text{Prob}\{A|x = x\} f(x) \, dx} \]

Example:
\( A=\{k \text{ heads in } n \text{ tossing in a specific order}\} \) where probability of a head showing, \( p \), is a RV with PDF \( f(p) \).

What is \( f(p|A) \)

Solution:

\[ \text{Prob}\{A|P = p\} = p^k (1 - p)^{n-k}, \quad P \text{ is a RV with } f(p) \]
From (2), we have

\[ f(p|A) = \frac{p^k(1-p)^{n-k}f(p)}{\int_0^1 p^k(1-p)^{n-k}f(p) \, dp} \]

\( f(p|A) \) is called a \textit{posteriori} density, and \( f(p) \) is called a \textit{priori} density for RV \( P \).
For large $n$, $p^k(1 - p)^{n-k}$ has a sharp maximum at $p = k/n$. $f(p)p^k(1 - p)^{n-k}$ is highly concentrated near $p = k/n$. If $f(p)$ has a sharp peak at $p = 0.5$, the coin is reasonably fair, then, for moderate values of $n$, $f(p)p^k(1 - p)^{n-k}$ has two peaks; one near $p = k/n$ and the other near $p = 0.5$. As $n$ increases sharpness of $p^k(1 - p)^{n-k}$ prevails and the resulting a posteriori density $f(p|A)$ has the maximum near $k/n$. 
If the probability of heads in coin tossing experiment is not a number, but an RV $P$ with density $f(p)$. In the experiment of the tossing of a randomly selected coin, show that $\text{Prob}\{\text{head}\} = \int_0^1 p f(p) \, dp$

Solution:

$A = \{\text{head}\} \Rightarrow \text{the conditional probability of } A \text{ is the probability of heads if the coin with } P = p \text{ is tossed. In other words, } \text{Prob}\{A|P = p\} = p$

$\int_0^1 \text{Prob}\{A|P = p\} f(p) \, dp = \int_0^1 p f(p) \, dp = \text{Prob}\{A\}$

This is the probability that at the next tossing head will show.
Example:
If $P$ is a uniform RV, determine the posteriori density.
Solution:
$\mathcal{A} = \{k \text{ heads in } n \text{ tossing in a specific order}\}$

$$f(p|\mathcal{A}) = \frac{p^k(1-p)^{n-k}}{\int_0^1 p^k(1-p)^{n-k} \, dp}$$

$$= \frac{(n+1)!}{k!(n-k)!} p^k (1-p)^{n-k}, \quad \text{Beta density}$$
Example:
Assuming that the coin was tossed \( n \) times and heads showed \( k \) times, what is the probability that at the next tossing heads would show? 

Solution:

\[
\int_0^1 p f(p|A) \, dp = \frac{(n + 1)!}{k!(n-k)!} \int_0^1 p^k (1 - p)^{n-k} \, dp
\]

\[
= \frac{k + 1}{n + 2}, \quad \text{almost the common sense!}
\]

This is called the law of succession.
To illustrate the usefulness of this formulation, let us reexamine the coin tossing problem.

Example 4.7: Let \( p = P(H) \) represent the probability of obtaining a head in a toss. For a given coin, a-priori \( p \) can possess any value in the interval \((0,1)\). In the absence of any additional information, we may assume the a-priori p.d.f \( f_p(p) \) to be a uniform distribution in that interval. Now suppose we actually perform an experiment of tossing the coin \( n \) times, and \( k \) heads are observed. This is new information. How can we update \( f_p(p) \)?

Solution: Let \( A = "k \) heads in \( n \) specific tosses". Since these tosses result in a specific sequence,

\[
P(A | P = p) = p^k q^{n-k}, \quad (4-34)
\]
and using (4-32) we get

\[ P(A) = \int_0^1 P(A \mid P = p) f_P(p) \, dp = \int_0^1 p^k (1 - p)^{n-k} \, dp = \frac{(n - k)! \, k!}{(n + 1)!}. \]  

(4-35)

The a-posteriori p.d.f \( f_{P|A}(p \mid A) \) represents the updated information given the event \( A \), and from (4-30)

\[ f_{P|A}(p \mid A) = \frac{P(A \mid P = p) f_P(p)}{P(A)} \]

\[ = \frac{(n + 1)!}{(n - k)! \, k!} p^k q^{n-k}, \quad 0 < p < 1 \sim \beta(n, k). \]  

(4-36)

Notice that the a-posteriori p.d.f of \( p \) in (4-36) is not a uniform distribution, but a beta distribution. We can use this a-posteriori p.d.f to make further predictions, For example, in the light of the above experiment, what can we say about the probability of a head occurring in the next \( (n+1) \)th toss?
Let $B = \text{“head occurring in the (n+1)th toss, given that } k \text{ heads have occurred in } n \text{ previous tosses”}.$

Clearly $P(B \mid P = p) = p$, and from (4-32)

$$P(B) = \int_0^1 P(B \mid P = p) f_p(p \mid \text{A}) dp. \quad (4-37)$$

Notice that unlike (4-32), we have used the a-posteriori p.d.f in (4-37) to reflect our knowledge about the experiment already performed. Using (4-36) in (4-37), we get

$$P(B) = \int_0^1 p \cdot \frac{(n + 1)!}{(n - k)!k!} p^k q^{n-k} dp = \frac{k + 1}{n + 2}. \quad (4-38)$$

Thus, if $n = 10$, and $k = 6$, then

$$P(B) = \frac{7}{12} = 0.58,$$

which is more realistic compare to $p = 0.5$.  

PILLAI
To summarize, if the probability of an event $X$ is unknown, one should make noncommittal judgement about its a-priori probability density function $f_x(x)$. Usually the uniform distribution is a reasonable assumption in the absence of any other information. Then experimental results ($A$) are obtained, and our knowledge about $X$ must be updated reflecting this new information. Bayes’ rule helps to obtain the a-posteriori p.d.f of $X$ given $A$. From that point on, this a-posteriori p.d.f $f_{x|A}(x|A)$ should be used to make further predictions and calculations.