Exponential Distribution

We say \( X \) is exponential with parameter \( \lambda \) if

\[
f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}
\]

\[
\text{m: mean } = E(X) = \frac{1}{\lambda}
\]

As an example, if waiting times such as telephone arrival times over non-overlapping intervals are independent, then the distribution of waiting times is exponential.

To see this:

Let \( q(t) = P(\text{no event occurred in time interval } t \text{ by } t) \)

Let \( X \) represent waiting time to first arrival. Then, by definition \( P(X > t) = q(t) \)

\( t_1, t_2 \) consecutive non-overlapping intervals

Assume independent. Then non-occurrence also independent.

Then \( q(t_1)q(t_2) = q(t_1 + t_2) \)

\( \therefore q(t) = ? \)
\[ \log q(t_1) + \log q(t_2) = \log q(t_1 + t_2) \]

Only solution \( \log q(t) = ct \),

i.e. \( ct_1 + ct_2 = c(t_1 + t_2) \)

\[ q(t) = e^{ct} \]

\[ P\{X > t\} = e^{ct} \]

Assuming \( X \) satisfies probability axioms

\[ P\{X > y \} = e^{-\lambda t}, \lambda > 0, t > 0 \]

\[ F_X(t) = P\{X \leq t\} = 1 - q(t) = 1 - e^{-\lambda t} \]

Memorylessness of exponential distribution:

Let \( s, t > 0 \)

Consider events \( \{X > t + s\} \) and \( \{X > s\} \)

\[ P\{X > t + s \mid X > s\} = \frac{P\{X > t + s\}}{P\{X > s\}} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} = P\{X > t\} \]

So, if \( X \) represents lifetime of an equipment, then

if equipment working for time \( s \), prob. that it will survive

additional time \( t \), regardless of if it were new equipment.
Ex: 4-10

Assume the life length of an appliance has an exponential distribution \( X \sim \text{Exp}(10) \). A used appliance is bought by someone. 

**P (appliance will not fail in 5 years)**

\[
P(X > 50) = P(X > 5) = e^{\frac{-5}{10}} = e^{-0.5} = 0.606
\]

Ex: 4-11

**Waiting time at a restaurant**

Assume waiting time at a restaurant has an exponential distribution with mean value 7.5 minutes. \( (\lambda = \frac{1}{5}) \)

Then \( P( \text{customer will spend more than 10 minutes at restaurant}) \)

\[
P(X > 10) = e^{\frac{-10}{7.5}} = e^{-1.333} = 0.265
\]

More interestingly, the (conditional) probability that the customer will stand an additional 10 minutes given that he or she has been there for more than 10 minutes.

\[
P(X > 10 + 10 | X > 10) = P(X > 20) = e^{-2}
\]
Discrete RV

The simplest among the discrete RV's is the Bernoulli RV that corresponds to any experiment with only 2 possible outcomes - success or failure (HvN). The Bernoulli distribution X is said to be Bernoulli distributed if X takes the values 1 or 0 with \( P(X=1) = p \), \( P(X=0) = q = 1-p \).

\[
\begin{align*}
\text{Fig. 4.2} & \\
\text{Bernoulli probability mass function} & \\
\text{Sample space for each individual trial is 2 points, S = \{0, 1\}.} & \\
\text{If n Bernoulli trials contains 2^n points.} & \\
\text{Since independent, probability} & \\
\text{Each Bernoulli trial is the successive tossing of a fair coin.}
\end{align*}
\]
Binomial Distribution

We are interested in the total number of successes produced in a sequence of \( n \) Bernoulli trials. So want the most

\[ n \text{ trials result in } k \text{ successes and } n-k \text{ failures}. \]

So want \( \binom{n}{k} \) points. By definition, each point has probability \( p^k q^{n-k} \). Thus, there is

\[ \Pr( k \text{ successes in } n \text{ Bernoulli trials}) \]

\[ p_k(n) = \binom{n}{k} p^k q^{n-k} \]

This so-called binomial distribution is above the form in the binomial expansion of \((p+q)^n\)

\[ \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n \]

Corresponding distribution is a staircase form. Fig. 4.20
Discrete Uniform distribution

\[ f(x) = \sum_{i} p_i \mathbb{I}(x - x_i) \quad p_i = P(X = x_i) \]

**Definition:**

- **Density function:**
  \[ f(x) = \frac{1}{N} \text{ for } h = \frac{1}{2}, \ldots, N \]

**Explanation:**

- The probability mass function of a discrete uniform distribution is constant over a range of values. For a random variable \( X \) with values in the range \( \frac{1}{2}, \ldots, N \), the probability of \( X \) being any of these values is \( \frac{1}{N} \).
\begin{align*}
\text{Uniform} & \quad \text{RV } X \text{ is uniformly distributed in } a, x \text{ if } -\infty < a < x < \infty. \\
& \quad f_X(x) = \begin{cases} 
\frac{1}{x_2 - x_1}, & x_1 \leq x \leq x_2 \\\n0, & \text{otherwise}
\end{cases} \quad a \leq x \leq b
\end{align*}

\begin{align*}
\text{Binomial} & \quad \text{RV } X \text{ has a binomial distribution of } \text{RVs } 0, 1, \ldots, n \text{ with } \\
& \quad X = \sum_{i=1}^{n} Y_i, \quad Y_i \text{ independent Bernoulli trials, } p_i \text{ success probability.} \\
& \quad f_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \\
& \quad \text{if } x = 0, 1, \ldots, n \text{ and } x + y = n \\
& \quad \text{and } \quad f_X(x) = 0 \text{ otherwise.} \\
& \quad \text{For large } n, \quad f_X(x) \approx \text{Normal } N(np, \sqrt{npq}) \\
& \quad \text{distribution.}
\end{align*}
Bernoulli Trials. Some event related to time, called *Bernoulli trials*. Desired to find probability that particular event occurs k times.

Ex. 4.10. (Bernoulli trials). In a set of n trials of a coin, an outcome is a sequence $y_1, \ldots, y_n$ of $k$ heads and $(n-k)$ tails where $k = 0, \ldots, n$. Define RV $X$ as

$$X(y_1, y_2, \ldots, y_n) = k \quad \text{i.e. } \# \text{ of heads}$$

$$P(X = k) = \binom{n}{k} p^k q^{n-k}$$

$\therefore$ $X$ has a **binomial distribution**

$$f_X(x) = \binom{n}{x} p^x q^{n-x}$$

For $X = x$, $P(X = x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$

**Binomial**

$$n = 9 \quad p = \frac{1}{2}$$

**Cumulative Distribution**

$$F_x(x) = \sum_{k=0}^{x} \binom{n}{k} p^k q^{n-k} \quad m \leq x \leq m+1$$

is a staircase function and in $(0,n)$ it is given by $x$

For large $n$, $F_x(x)$ close to $N(np, \sqrt{npq})$ distribution i.e.

$$F_x(x) \approx G \left( \frac{x - np}{\sqrt{npq}} \right)$$
\[
p \approx 0.5 \quad \text{with} \quad np = 50
\]
\[
G(60.50) - G(40.50) \quad \sqrt{npq} = 5
\]
\[
= G(2) - G(-2) = 0.9545
\]

- **Poisson**

A RV \( X \) is Poisson distributed with parameter \( \lambda \) if it takes the values 0, 1, 2, ..., with
\[
P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}
\]

X if lattice type and density, \( f(x) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \delta(x-k) \)

The Poisson process is an outcome \( Y \) is a set of points \( t_i \) on the non-negative \( \mathbb{R}^+ \) axis.

A) Given constant \( \lambda \), we define \( NV n \in (0, \infty) \) as \# of points \( t_i \in (0, t) \).

Clearly \( n = \lambda \) means \# of points \( t_i \in (0, t) \) that \( = \lambda \).

\[
P(n = \lambda) = e^{-\lambda t} \left( \frac{\lambda t}{e} \right)^k
\]
Recall, \[ P(A|M) = \frac{P(A \cap M)}{P(M)} \] where \( P(M) \neq 0 \)

The conditional distribution \( F(x|M) \) of a random variable \( X \) assuming \( M \) space is defined as the conditional probability of event \( \{ X \leq x \} \)

\[ F(x|M) = P(X \leq x | M) = \frac{P(X \leq x, M)}{P(M)} \]

where, of course, \( P(X \leq x, M) \) is an event consisting of all outcomes \( x \) satisfying \( \{ y: X(y) \leq x \} \subset \{ y \in M \} \).

Thus definition of \( F(x|M) \) same as def. of \( F(x) \), except that all probabilities replaced by conditional probabilities.

:. \( F(x|M) \) has same properties as \( F_X(x) \)

\[ F(a|M) = 1 \]
\[ F(-a|M) = 0 \]
\[ P(x_i < x \leq x_{i+1} | m) = P(x_i/m) - P(x_{i+1}/m) = \frac{P(x_i, x \leq x_{i+1} | m)}{P(x_i/m)} \]

Conditional density \( f(x/m) \) is given by:

\[ f(x/m) = \frac{d}{dx} \frac{P(x, x \leq x | m)}{P(x_i/m)} = \lim_{\Delta x \to 0} \frac{P(x_i \leq x < x_i + \Delta x | m)}{\Delta x} \]

**Example:** Determine conditional distribution \( P(x/m) \) of the RV \( X = x_i \) in the Jar and Expt. where \( \Omega = (x_1, \ldots, x_k) \) is the event from \( S = (x_1, x_2, \ldots, x_k) \).

Define:

- \( X: S \to \mathbb{R} \)
- \( \mathbb{P}(X \leq 60) = P(x_i) = 1 \)
- \( \mathbb{P}(X < 60) = P(x_i, x \leq 60) \)
- \( P(x_i, f_i, f_i, f_i, f_i, f_i) \)
\[ P(x \leq 60 / y) = \frac{P(x \leq 60, y)}{P(y)} \]

\[ P(x < 60 | y) = \frac{P(x < 60, y)}{P(y)} \]

\[ P(x \leq 50 | y) = \frac{P(x \leq 50, y)}{P(y)} \]

\[ P(x < 50 | y) = \frac{P(x < 50, y)}{P(y)} \]

\[ P(x \leq 40 | y) = \frac{P(x \leq 40, y)}{P(y)} \]

\[ P(x < 40 | y) = \frac{P(x < 40, y)}{P(y)} \]

\[ P(x \leq 30 | y) = \frac{P(x \leq 30, y)}{P(y)} \]

\[ P(x < 30 | y) = \frac{P(x < 30, y)}{P(y)} \]

\[ P(x \leq 20 | y) = \frac{P(x \leq 20, y)}{P(y)} \]

\[ P(x < 20 | y) = \frac{P(x < 20, y)}{P(y)} \]
To find $F(x|M)$, most in general leave underlying event. However, if $M$ is an event that can be expressed in terms of the RV $X$, then knowing $F_x(x)$ sufficient to determine $F(x|M)$.

Two important examples:

I. Find conditional distribution of RV $X$, assuming $X \leq a$.

\[ M = \{ X \leq a \} \]

Problem:

Find \[ F(a|X \leq a) = P(X \leq a | X \leq a) \]

\[ = \frac{P\{X \leq a, X \leq a\}}{P\{X \leq a\}} \]

If $a > a$, then $\{X \leq a, X \leq a\} = \{X \leq a\}$

\[ F(a|X \leq a) = \frac{P\{X \leq a\}}{P\{X \leq a\}} = 1 \quad a > a \]
If $x < a$, then $\{X \leq x, a \leq X\}^c = \{X \leq x\}$

$P(x | X \leq a) = \frac{P(x \leq x | x \leq a)}{P(x \leq a)} = \frac{F_x(x)}{F_x(a)} \quad x < a$

To get pdf, we differentiate.

$x \geq a \quad f(a | X \geq a) = 0$

$x < a \quad f(x | X \geq a) = \frac{f_x(x)}{F_x(a)} = \frac{\frac{f_x(x)}{\int_a^\infty f_x(x) \, dx}}{\int_a^\infty f_x(x) \, dx}$

II. Suppose now, $M = \{b < X \leq a\}$

Now $\forall b \quad P_x(x | M) = P(X \leq x | M) = \frac{P(X \leq x, b < X \leq a)}{P(M)}$

$\therefore F_x(x | b < X \leq a) = \frac{P(X \leq a, b < X \leq a)}{P(b < X \leq a)}$

If $x = a$ then $\{X \leq x, b < X \leq a\} = \{b < X \leq a\}$

Hence $F_x(a | b < X \leq a) = \frac{F_x(a) - F_x(b)}{F_x(a) - F_x(b)} = 1$
If \( b \leq \alpha < a \), then \( \{ X \leq \alpha, b < X \leq a \} = \emptyset \)

Hence \( F_X(\alpha | b \leq X \leq a) = \frac{F(x) - F(b)}{F(a) - F(b)} \)

If \( \alpha < b \), then \( \{ X \leq \alpha, b < X \leq a \} = \emptyset \)

\[ F(\alpha | b < X \leq a) = 0 \quad \alpha < b \]

\[ f_X(\alpha | b < X \leq a) = \begin{cases} \frac{f_X(x)}{F(a) - F(b)} & b \leq x \leq a \\ 0 & \text{otherwise} \end{cases} \]
Determine the conditional density \( f_X(x|X-\eta) \leq k\sigma \) of a \( N(\eta, \sigma^2) \) RV. \( \mathcal{N} \) f(x) = \( \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\eta)^2}{2\sigma^2}} \)

From above, \( f_X(x|b < X < a) = \frac{f(x)}{P(b < X < a)} \) \( b < x < a \)

\[
P\{ |X-\eta| \leq k\sigma \} = \frac{P\{ \eta-k\sigma \leq X \leq \eta+k\sigma \}}{P(b < X < a)}
\]

\[
= \frac{G(\eta+k\sigma)-G(\eta-k\sigma)}{P(b < X < a)}
\]

\[
= \frac{1}{2G(k)} \left( \frac{-e^{-\frac{(x-\eta)^2}{2\sigma^2}}}{\sigma \sqrt{2\pi}} \right)
\]

\[
\therefore \quad \eta-k\sigma \leq X \leq \eta+k\sigma
\]

\[
F_X(x) = G(\eta-\frac{x-np}{\sqrt{npq}})
\]

\[
F_X(\eta+kr) = G(\eta+kr-\frac{np}{\sqrt{npq}}) = G(\eta+kr-\frac{np}{\sqrt{npq}})
\]
Frequency Interpretation

In a sequence of $n$ trials, we reject all outcomes $Y$

$\begin{array}{c}
\exists \quad X(y) \leq b \quad \text{or} \quad X(y) \geq a.
\end{array}$

In the subsequence of the remaining trials,

$P_X(x \mid b \leq x \leq a)$ has the same freq. interpretation as $F(x)$ did.

$p_X(x \mid M) = \frac{P(X=x \mid b \leq x \leq a)}{n} = \frac{n_X}{n}$

$n_X = \# \text{trials } \exists \quad X(y) \leq x$

$n = \text{total } \# \text{trials} \quad \text{where} \quad b \leq x \leq a$
\[ P(\exists x \leq x, \exists a < x < b) \]

\[ = P(\exists y: x(y) \leq x, \exists y: a < x(y) < b) \]

\[ = P(\{ x \in (-\infty, x) \}) - P(\{ x \in (-\infty, a) \}) \]

\[ = F_x(x) - F_x(a) \]
**Total Probability or Bayes Theorem**

Extend results of Sec. 2-3 to random variables.

1. Setting $B = \{X \leq x\}, m \{(\mathbb{R})\}, m^{\prime}$ set \[ (\text{impl} \, \text{m} \, \text{set}) \]

   \[ \text{e.g. } U = [A_1, A_2, \ldots, A_n] \text{ \ is a partition of } S \]

   Then \[ P(B) = \sum \frac{P(B \mid A_i)P(A_i)}{P(A_i)} \]

   Setting \[ B = \{X = x\}, m^{\prime} \text{ set} \]

   \[ P(x \leq x) = P(x \leq x \mid A_1)P(A_1) + \cdots + P(x \leq x \mid A_n)P(A_n) \]

   \[ \text{e.g. } F_X(x) = \sum F(x \mid A_i)P(A_i) + \cdots + F(x \mid A_n)P(A_n) \]

   \[ \text{e.g. } f(x) = \sum f(x \mid A_i)P(A_i) + \cdots + f(x \mid A_n)P(A_n) \]

**Ex. 4.17**

Suppose \( N \leq \mu \leq \nu \) \( f(x \mid \mu) \) is \( N(\mu, \sigma_0) \)

and \( f(x \mid \bar{\mu}) \) is \( N(\mu_2, \sigma_2) \). 

So in figure below.
 Clearly, \( m \) and \( \bar{m} \) form a partition of \( S \).

Setting \( A_1 = M \) and \( A_2 = \bar{M} \) we get,

\[
f(x) = f(x|m)p + f(x|m)(1-p)
\]

\[
= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-m)^2}{2\sigma_1^2}} + \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x-m)^2}{2\sigma_2^2}}
\]

where \( p = P(m) \)

This result is used in hypothesis testing.

2. From the identity,

Bayes' Theorem

\[
P(A|B) = \frac{P(B|A)P(A)}{P(B)}
\]

it follows that,

\[
P(\bar{A}|X \leq x) = \frac{P(X \leq x|\bar{A})P(\bar{A})}{P(X \leq x)}
\]

\[
P(A|X \leq x) = \frac{P(X \leq x|A)p}{P(X \leq x)}
\]
3. Setting $B = \{ x_1 < X \leq x_2 \}$ in (4-53), we conclude with

\[
P(B | x_1 < X \leq x_2) = \frac{P(x_1 < X \leq x_2 | A) P(A)}{P(x_1 < X \leq x_2)} = \frac{P(A) - P(A | m)}{P(m)}
\]

that,

\[
P(A | x_1 < X \leq x_2) = \frac{P(A | x_1 < X < x_2) P(A)}{P(x_1 < X < x_2)} = \frac{P(A) - P(A | m)}{P(m)}
\]

Bayes Th.

4. The conditional probability $P(A | X = x)$ of the event

A assuming $X = x$ cannot be defined as in (4-64), i.e.

\[
P(A | m) = \frac{P(A | m)}{P(m)}
\]

because, in general, $P(X = x) = 0$.

We shall define it as a limit

\[
P(A | X = x) = \lim_{\delta x \to 0} P(A | x < X \leq x + \delta x)
\]

\[4-77\]

From (4-77), with $x_1 = x$ and $x_2 = x + \delta x$, we conclude

\[
P(A | X = x) = \frac{f(x | A)}{f(x)} P(A | x)
\]

(4-79)
As we know, \( P(\infty/m) = 1, \quad P(-\infty/m) = 0 \)
\[
F'(a/A) = \int_{-\infty}^{+\infty} f(x/A) \, dx = 1
\]
\((\because F(\infty/m) = 1, \quad F(-\infty/m) = 0) \quad (4-42)\)

Multiplying \( f(x) \) by \( f(x) \) and integrating, we obtain,
\[
\int_{-\infty}^{+\infty} P(A/X = x) \, f(x) \, dx = P(A) \quad (4-64)
\]

This is the continuous version of the total probability theorem
\[
P(B) = P(B/A_1)P(A_1) + \cdots + P(B/A_n)P(A_n)
\]
Ex. 4-18: Suppose probability of heads in a coin-tossing experiment is not a number, but a RV \( P \) with density \( f(p) \) in some space \( S \). Let \( H \) be a random variable in a Cartesian product \( S \times S \). Then \( P = \{ \text{heads} \} \), and \( H \) has density \( f(p)p \) on the form \( f(p) = f(p)p \) for any element of \( S \), and \( H \) is the element \( H = (p, p) \). For example \( P(H) = \int_0^1 f(p)p \, dp \).

**Proof:**

Consider the probability \( P = p \) of heads in a coin-tossing experiment, where \( p = p \) is fixed. The probability \( P(H) = f(p)p \) for \( p \) fixed.

Insertion into (4-64):

\[
P(H) = \int_0^1 f(p)p \, dp = \int_0^1 f(p) \, dp = 1
\]

**Bayes' Theorem**

From (4-63) and (4-64):

\[
f(x/A) = \frac{P(A|X=x) \cdot f(x)}{P(A)}
\]

\[
f(x/A) = \frac{P(A|X=x) \cdot f(x)}{\int_0^1 P(A|X=x)p \cdot f(x) \, dx}
\]

This is the continuous version of Bayes' Theorem (239):

\[
P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}
\]
Application of Bayes' Theorem:

Use of past observation in determination of unknown probabilities.

Example problem: Coin tossed n times, k heads.

Qs.: What can you say about probability of heads in subsequent trials?

Can give 2 interpretations:

1. Probability of heads is an unknown number \( p \), and if \( n \) large, \( p \approx k/n \).

2. Probability of heads is a RV \( p \).

Suppose \( p \) RV from past observations. This knowledge can be applied on the basis of the given information.
A mixture of random solutions can
be of height \( N \) with \( N \) pdf \( f(p) \)

Given \( S \times S^n \) get \( h \)th

Assume \( S \), \( S^n \) independent.

\( \therefore \) features remain not affected by past events.

\( \therefore P(A|P) = p \) where

\[ A = \text{a head in whom if many} \]
\[ = p^{a \cdot n \cdot k} \]

Now Bayes Sh.:

\[
f(x|A) = \frac{P(A|x=x) f(x)}{P(A) = \int P(A|x=x) f(x) \, dx}
\]

\[
f(p|A) = \frac{p^{a \cdot g \cdot n \cdot k} f(p)}{\int p^{a \cdot g \cdot n \cdot k} f(p) \, dp}
\]

Note: A posteriori probability

A priori probability

\( f(p) \) has sharp maxima near \( \frac{1}{n} \)
Summary: \( P \) is a priori pdf \( f(p) \). Given

1a. Find the RV \( P \) between \( p_1 \) and \( p_2 \)

\[
\int_{p_1}^{p_2} f(p) \, dp
\]

1b. In a single toss of randomly selected coin, probability of heads equals

\[
\int_0^1 p \cdot f(p) \, dp
\]
Now toss coin \( n \) times, \( h \) heads. Then we conclude

\[ \text{what:} \]

\[ 2a. \text{ Prob. that } \text{RV } p \text{ lies between } p_1, p_2 \]

\[ \int_{p_1}^{p_2} f(p|A) \, dp \]

\[ 2b. \text{ At the next toss, prob. of heads equals} \]

\[ \int_0^1 p f(p|A) \, dp \]

Exercise 4.19: \( \text{RV } p \text{ unif. \( \text{form in } (0,1) \).} \)

\[ f(p|A) = \frac{\binom{n}{h} (1-p)^{n-h} 1}{\int_0^1 p^h (1-p)^{n-h} \, dp} = \frac{(n+1)! \cdot p^h (1-p)^{n-h}}{4! (n-4)!} \]

\[ \text{Beta} \]

\[ \text{density}. \]