Lecture 2: Brief Review on Probability Theory

1 Terminology & Basics

The **sample space**, usually denoted $\mathcal{S}$, is the set of all possible outcomes of a random study/experiment [2].

Any subset of the sample space is known as an **event**, usually denoted with capital letters $A, B, C, \cdots$, [2, 3]. For example, $A \subset \mathcal{S}$, where "$\subset$" denotes "is a subset of".

Since events and sample spaces are just sets, the following is the algebra of sets [3]:

- $\varnothing$ is the **null set**, or **empty set**.
- $A \cup B$ = **union** = the elements in A or B or both.
- $C \cap D$ = **intersection** = the elements in C and D. If $C \cap D = \varnothing$, then C and D are called **mutually exclusive events**, or **disjoint events**.
- $A' = A^c$ = **complement** = the elements not in A.
- If $E \cup F \cup G \cup \cdots = \mathcal{S}$, then E, F, G, and so on are called **exhaustive events**.

Probability is a (real-valued) set function P that assigns to each event in the sample space a number $P(\cdot)$ [3]. For example, the probability of event A equals:

$$P(A) = \frac{N(A)}{n} \,,$$

where n is the number of times an experiment is performed and $N(A)$ is the number of times the event A of interest occurs.

Probability of an event A in the sample space $\mathcal{S}$ is defined and satisfies the following three conditions [2]:

(i) $0 \leq P(A) \leq 1$.

(ii) $P(\mathcal{S}) = 1$.

(iii) For any sequence of events $A_1, A_2, \cdots$ that are mutually exclusive, that is, events for which $A_n \cap A_m = \varnothing$ when $n \neq m$, then,

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

Five theorems of probability [3]:

- Theorem 1: Since the events A and A^c are always mutually exclusive and since $A \cup A^c = \mathcal{S}$, we have that

  $$1 = P(\mathcal{S}) = P(A \cup A^c) = P(A) + P(A^c)\,,$$

  or

  $$P(A) = 1 - P(A^c)\,.$$

- Theorem 2: $P(\varnothing) = 0$.
- Theorem 3: If events A and B are such that $A \subseteq B$, then $P(A) \leq P(B)$.
- Theorem 4: $P(A) \leq 1$.
- Theorem 5: For any two events A and B,

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

  or equivalently

  $$P(A) + P(B) = P(A \cup B) + P(A \cap B)\,.$$

  Note that when A and B are mutually exclusive, that is when $A \cap B = \varnothing$, then

  $$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= P(A) + P(B) \end{aligned}$$

Random variables are real-valued functions defined on the sample space [2].

Quantitative data are called **discrete** if the sample space contains a finite or countably infinite number of values [3].

Quantitative data are called **continuous** if the sample space contains an interval or continuous span of real numbers [3].

Probability notations that we use throughout our lecture notes for events and random variables are distinguished by the type of brackets :

- Probability of events is denoted: $P(\cdot)$
- Probability of random variables is denoted: $P\{\cdot\}$

2 Discrete Random Variables

A random variable is said to be discrete if:

- it can take on a finite number of possible outcomes, or
- it can take on a countably infinite number of possible outcomes.

Let X be a discrete random variable. The probability of X takes on a particular value x, that is, $P\{X = x\}$, is denoted $f(x)$, or

$$f(x) = P\{X = x\} \, . \tag{1}$$

The function $f(x)$ is called the **probability mass function** and commonly abbreviated the **p.m.f.**

The probability mass function $f(x)$ has the following properties [3]:

(1) $P\{X = x\} = f(x) > 0$ if $x \in \mathcal{S}$, where $\mathcal{S}$ is the sample space.

(2) $\sum_{x \in \mathcal{S}} f(x) = 1$

(3) $P\{X \in A\} = \sum_{x \in A} f(x)$

The **cumulative distribution function**, commonly abbreviated the **c.d.f.** of the random variable X is defined by:

$$F_X(t) = P\{X \leq t\} \, . \tag{2}$$

The cumulative distribution function $F_X(t)$ has the following properties:

(1) $F_X(t)$ is a nondecreasing function of t, for $-\infty < t < \infty$.

(2) $F_X(t)$ ranges from 0 to 1:

- If the minimum value of X is a, then $F_X(a) = P\{X \leq a\} = P\{X = a\} = f_X(a)$.
- For c where c is less than a, then $F_X(c) = 0$.
- If the maximum value of X is b, then $F_X(b) = 1$.

Let $u(X)$ be a function of the discrete random variable X, namely

$$u(X) = X \, .$$

The expectation of $u(X)$, or the expected value of X, is:

$$E[u(X)] = E(X) = \sum_{x \in S} x f(x) \,. \tag{3}$$

$E(X)$ is also called the **mean of $\boldsymbol{X}$**, and is denoted μ, that is $\boldsymbol{\mu = E(X)}$.

When $u(X) = (X - \mu)^2$, the expectation of $u(X)$

$$E\left(u(X)\right) = E\left((X-\mu)^2\right) = \sum_{x \in S} (x-\mu)^2 f(x) \tag{4}$$

is called the **variance of $\boldsymbol{(X)}$** and denoted $\mathbf{Var}\boldsymbol{(X) = \sigma^2}$. Alternatively,

$$\begin{aligned} \mathrm{Var}(X) = E\left((X-\mu)^2\right) &= E\left(X^2 - 2\mu X + \mu^2\right) \\ &= E\left(X^2\right) - 2\mu E(X) + E\left(\mu^2\right) \\ &= E\left(X^2\right) - 2\mu\mu + \mu^2 \\ &= E\left(X^2\right) - \mu^2 \,. \end{aligned} \tag{5}$$

Taking the positive square root of the variance gives us the **standard deviation of $\boldsymbol{X}$**, denoted as

$$\sigma = \sqrt{\mathrm{Var}(X)} \,. \tag{6}$$

The mean is a measure of the middle of the distribution of X, whereas the variance and standard deviation quantify the spread or dispersion of the values in the sample space $\mathcal{S}$.

Another way to calculate the mean and variance is via the moment generating function. Let $\phi(t)$ denote the moment generating function of the discrete random variable X, defined for all values t by

$$\phi(t) = E\left(e^{tX}\right) = \sum_k e^{tk} f(k) \,. \tag{7}$$

The function $\phi(t)$ is called the moment generating function because all of the moments of X can be obtained by successively differentiating $\phi(t)$. For example,

$$\begin{aligned} \phi'(t) &= \frac{d}{dt} E\left(e^{tX}\right) \\ &= E\left(\frac{d}{dt} e^{tX}\right) \\ &= E\left(X e^{tX}\right) \end{aligned}$$

The mean can be obtained by letting $t = 0$,

$$\phi'(0) = E\left(X e^{0.X}\right) = E(X) \tag{8}$$

Similarly, by the second order of derivative,

$$
\begin{aligned}
\phi''(t) &= \frac{d}{dt}\phi'(t) \\
&= \frac{d}{dt}\left(E(Xe^{tX})\right) \\
&= E\frac{d}{dt}\left(Xe^{tX}\right) \\
&= E\left(X^2 e^{tX}\right)
\end{aligned}
$$

and by letting $t = 0$,

$$
\phi''(0) = E\left(X^2 e^{0.X}\right) = E(X^2)\,, \tag{9}
$$

we obtain the variance of X,

$$
\mathrm{Var}(X) = E(X^2) - (E(X))^2 = \phi''(0) - (\phi'(0))^2\,. \tag{10}
$$

Discrete random variables are often classified according to their probability mass functions.

2.1 The Bernoulli Random Variable

From a trial or an experiment with the outcome either a "success" or a "failure", if we let

$X = 1$ if the outcome is a success, and

$X = 0$ if the outcome is a failure,

then the random variable X is said to be a **Bernoulli random variable** if its probability mass function is given by

$$
\begin{aligned}
f(0) &= P\{X = 0\} = 1 - p\,, \qquad (11) \\
f(1) &= P\{X = 1\} = p\,, \qquad (12)
\end{aligned}
$$

where p is the probability that the trial is a "success", for $0 \leq p \leq 1$.

Alternatively, the probability mass function of the Bernoulli random variable X can be written

$$
f(k) = P\{X = k\} = p^k (1-p)^{1-k}\,, \tag{13}
$$

for $k \in (0, 1)$.

The cumulative distribution function of a Bernoulli random variable X is given by

$$F_X = \begin{cases} 0 & \text{for } \ k < 0 \\ 1-p & \text{for } \ 0 \le k < 1 \\ 1 & \text{for } \ k \ge 1 \end{cases} \tag{14}$$

Since $f(0) = 1 - p$ and $f(1) = p$, the mean is

$$\mu = E(X) = 1 \times p + 0 \times (1-p) = p\,. \tag{15}$$

With

$$E(X^2) = 1^2 \times p + 0^2 \times (1-p) = p\,,$$

then the variance is

$$\sigma^2 = \mathrm{Var}(X) = E(X^2) - (E(X))^2 = p - p^2 = p(1-p)\,. \tag{16}$$

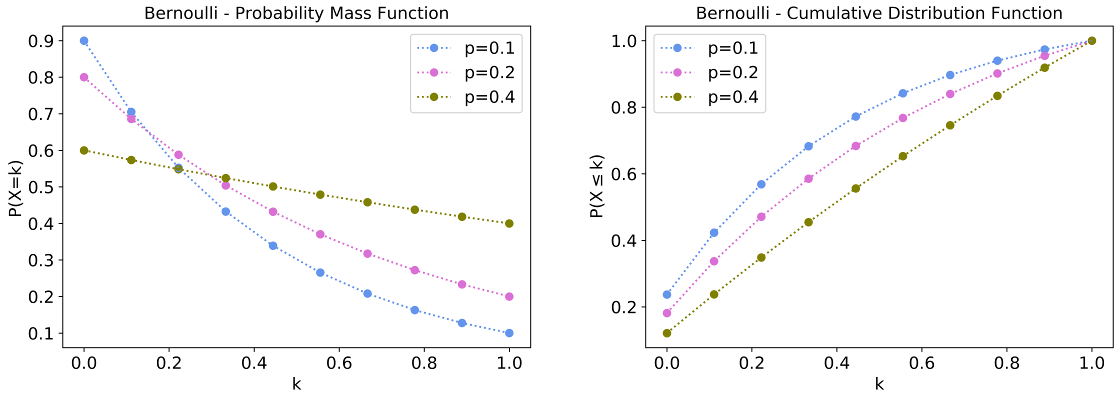

Figure 1: The probability of mass function (left) and cumulative distribution function (right) of a Bernoulli random variable.

2.2 The Binomial Random Variable

Suppose that with n independent trials, the results are either:

a "success" with probability p,

a "failure" with probability $1 - p$.

If X represents the number of successes that occur in the n trials, then X is said to be a **binomial random variable** with parameters (n, p). The probability mass function of a binomial random variable is given by

$$f(k) = P\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k}, \qquad \text{for} \quad k = 0, 1, 2, \cdots, n \tag{17}$$

where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{18}$$

equals the number of different groups of k objects that can be chosen from a set of n objects.

To check if $p(k)$ is a probability mass function, by the binomial theorem, the probabilities sum to one, or

$$\begin{aligned} \sum_{k=0}^{\infty} f(k) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \\ &= (p + (1-p))^n \\ &= 1 \end{aligned}$$

Proof:

According to the binomial theorem:

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^{n-i} y^i = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i} \tag{19}$$

hence

$$\begin{aligned} \sum_{k=0}^{\infty} p(k) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \\ &= (p + (1-p))^n \\ &= 1 \qquad \blacksquare \end{aligned}$$

The cumulative distribution function of the binomial random variable X evaluated at t is given by

$$F_X(t) = \sum_{k=0}^{t} \binom{n}{k} p^k (1-p)^{n-k}, \qquad \text{for} \quad k = 0, 1, 2, \cdots, n \tag{20}$$

From the moment generating function

$$\begin{aligned}\phi(t) &= E\left(e^{tX}\right)\\ &= \sum_k e^{tk} f(k)\\ &= \sum_{k=0}^{n} e^{tk}\binom{n}{k}p^k(1-p)^{n-k}\\ &= \sum_{k=0}^{n}\binom{n}{k}(pe^t)^k(1-p)^{n-k}\\ &= \left(pe^t+(1-p)\right)^n \end{aligned} \tag{21}$$

and using the chain rule,

$$\phi'(t) = \frac{d\phi}{dt} = \frac{d\phi}{du}\times\frac{du}{dt}$$

where $u = pe^t + (1-p)$ therefore $du/dt = pe^t$, and $\phi = u^n$ therefore $d\phi/du = nu^{n-1}$, hence

$$\phi'(t) = n\left(pe^t+(1-p)\right)^{n-1}pe^t$$

we can obtain the mean,

$$E(X) = \phi'(0) = n\left(pe^0+(1-p)\right)^{n-1}pe^0 = np\,. \tag{22}$$

After differentiating a second time,

$$\phi''(t) = n(n-1)\left(pe^t+1-p\right)^{n-2}\left(pe^t\right)^2 + n\left(pe^t+1-p\right)^{n-1}pe^t$$

we obtain the variance,

$$\sigma^2 = \mathrm{Var}(X) = \phi''(0) - \left(\phi'(0)\right)^2 = np(1-p)\,. \tag{23}$$

2.3 The Geometric Random Variable

Suppose that independent trials are performed until a success occurs, where the probability of being a success is denoted p. If we let X be the number of trials required until the first success occurs, then X is called a **geometric random variable** with parameters p.

The probability mass function of a geometric random variable is given by

$$f(k) = P\{X=k\} = (1-p)^{k-1}p\,, \qquad \text{for} \quad k=1,2,3,\cdots \tag{24}$$

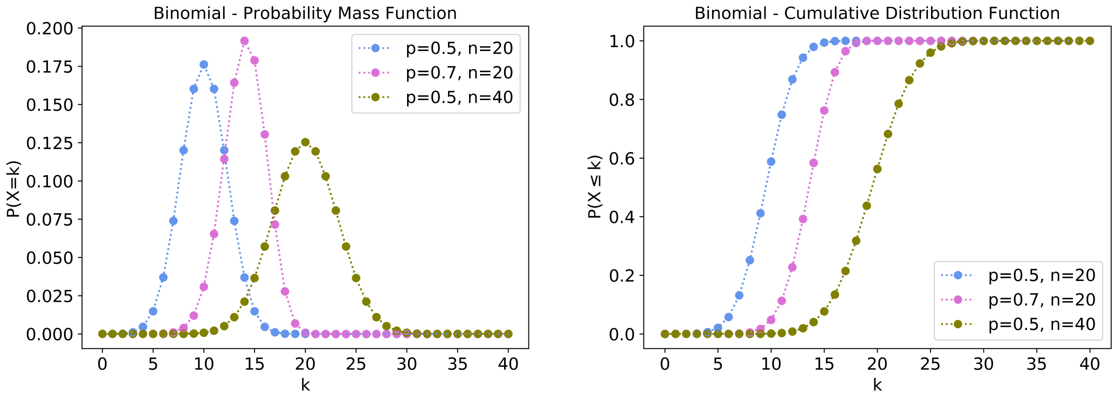

Figure 2: The probability of mass function (left) and cumulative distribution function (right) of a binomial random variable.

To check that $f(k)$ is a probability mass function, the probabilities sum to one:

$$\sum_{k=1}^{\infty} f(k) = p\sum_{k=1}^{\infty}(1-p)^{k-1} = 1$$

Proof:

The sequence:

$$S = p\sum_{k=1}^{\infty}(1-p)^{k-1} = p + p(1-p) + p(1-p)^2 + p(1-p)^3 + \cdots$$

is a geometric progression, where its first term $a = p$ and the common ratio $r = 1 - p$. Its sum to infinity is given by

$$S_\infty = \frac{a}{1-r} = \frac{p}{1-(1-p)} = 1 \qquad \blacksquare$$

The cumulative distribution function of a geometric random variable X is given by

$$F_X(k) = P\{X \le k\} = 1 - (1-p)^k\,. \tag{25}$$

By equations (3) and (24),

$$\begin{aligned} E(X) &= \sum_{k=1}^{\infty} k\,p\,(1-p)^{k-1} \\ &= p\sum_{k=1}^{\infty} k\,u^{k-1}\,, \end{aligned}$$

where $u = 1 - p$. Then, the mean is given by,

$$\begin{aligned} E(X) &= p \sum_{k=1}^{\infty} \frac{d}{du} \left(u^k \right) \\ &= p \frac{d}{du} \sum_{k=1}^{\infty} u^k \end{aligned}$$

where, this is a geometric sequence where the first term is $a = u$ and the common ratio is $r = u$, and its sum to infinity is similar to that has been given above,

$$S_\infty = \frac{a}{1-r} = \frac{u}{1-u} \,.$$

Hence,

$$\begin{aligned} E(X) &= p \frac{d}{du} \left(\frac{u}{1-u} \right) \\ &= \frac{p}{(1-u)^2} \\ &= \frac{1}{p} \,. \end{aligned} \tag{26}$$

Alternatively, the mean can be obtained from the moment generating function,

$$\begin{aligned} \phi(t) &= E(e^{tX}) \\ &= \sum_k e^{tk} f(k) \\ &= \sum_{k=1}^{\infty} e^{tk} \, (1-p)^{k-1} \, p \\ &= p \sum_{k=1}^{\infty} e^{tk} \, (1-p)^{k-1} \end{aligned}$$

The sum to infinity of the sequence can be written

$$S = pe^t + pe^{2t}(1-p) + pe^{3t}(1-p)^2 + pe^{4t}(1-p)^3 + \cdots$$

Since it is a geometric sequence, with $a = pe^t$ and $r = e^t(1-p)$, the moment generating function is given by,

$$\phi(t) = S_\infty = \frac{pe^t}{1 - e^t(1-p)} \,. \tag{27}$$

Using the quotient rule, we obtain the first derivative,

$$\phi'(t) = \frac{pe^t(1-e^t+pe^t) - pe^t(-e^t+pe^t)}{(1-e^t+pe^t)^2}$$

and by letting $t=0$, the mean is given by,

$$E(X) = \mu = \phi'(0) = \frac{p(1-1+p) - p(-1+p)}{(1-1+p)^2} = \frac{1}{p}\,. \tag{28}$$

Similarly, after differentiating a second time, we obtain the variance

$$\sigma^2 = \mathrm{Var}(X) = \phi''(0) - (\phi'(0))^2 = \frac{1-p}{p^2}\,. \tag{29}$$

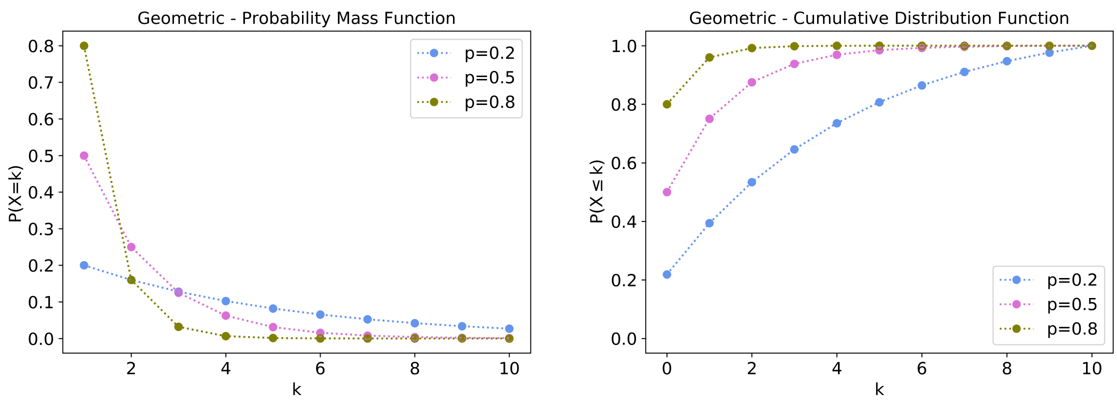

Figure 3: The probability of mass function (left) and cumulative distribution function (right) of a geometric random variable.

2.4 The Poisson Random Variable

The probability of mass function of a Poisson random variable with parameter λ, if for some $\lambda > 0$, is given by,

$$f(k) = P\{X=k\} = e^{-\lambda}\frac{\lambda^k}{k!}\,, \qquad \text{for} \quad k = 0, 1, 2, \cdots \tag{30}$$

since the probabilities sum to one,

$$\sum_{k=0}^{\infty} f(k) = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1\,.$$

Proof:

The Maclaurin series expansion for e^x is

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots$$

which is similar to the whole sum of

$$\sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \frac{\lambda^4}{4!} + \cdots = e^\lambda$$

then

$$e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1 \qquad \blacksquare$$

The cumulative distribution function is given by,

$$F_X(k) = P\{X \leq k\} = e^{-\lambda} \sum_{n=0}^{k} \frac{\lambda^n}{n!} \,. \tag{31}$$

The mean can be obtained from

$$\begin{aligned} \mu \;=\; E(X) &= \sum_{n=0}^{\infty} \frac{n\, e^{-\lambda}\, \lambda^n}{n!} \\ &= \sum_{n=1}^{\infty} \frac{e^{-\lambda}\, \lambda^n}{(n-1)!} \\ &= \lambda\, e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} \\ &= \lambda\, e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \\ &= \lambda\, e^{-\lambda}\, e^{\lambda} \\ &= \lambda \,. \end{aligned} \tag{32}$$

Alternatively, from the moment generating function

$$\begin{aligned} \phi(t) &= E(e^{tX}) \\ &= \sum_{k=0}^{\infty} e^{tk}\, e^{-\lambda}\, \frac{\lambda^k}{k!} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} e^{tk}\, \frac{\lambda^k}{k!} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} \\ &= e^{-\lambda}\, e^{\lambda e^t} \\ &= \exp(\lambda(e^t - 1)) \end{aligned} \tag{33}$$

the first differentiation yields,

$$\phi'(t) = \lambda\, e^t\, \exp(\lambda(e^t - 1))$$

from which we can obtaine the mean

$$\mu = \phi'(0) = \lambda\, e^0\, \exp(\lambda(e^0 - 1)) = \lambda\,. \tag{34}$$

From a second differentiation,

$$\phi''(t) = \left(\lambda\, e^t\right)^2 \exp(\lambda(e^t - 1)) + \lambda\, e^t \exp(\lambda(e^t - 1))$$

we obtain the variance,

$$\sigma^2 = \phi''(0) - (\phi'(0))^2 = \lambda\,. \tag{35}$$

3 Continuous Random Variables

Continuous random variables are random variables whose sample space (or, set of possible values) contains uncountable or real numbers.

Let X be a continuous random variable. The probability that X takes on any particular value x is 0. Namely, finding $P\{X = x\}$ for a continuous random variable X is unfeasible. Instead, the probability of X should fall in some interval (a, b), that is, we need to find $P\{a < X < b\}$. This is achieved by using a **probability density function** or commonly abbreviated the **p.d.f.**

If X is a continuous random variable, there exists a function $f(x)$ which is integrable and has the following properties:

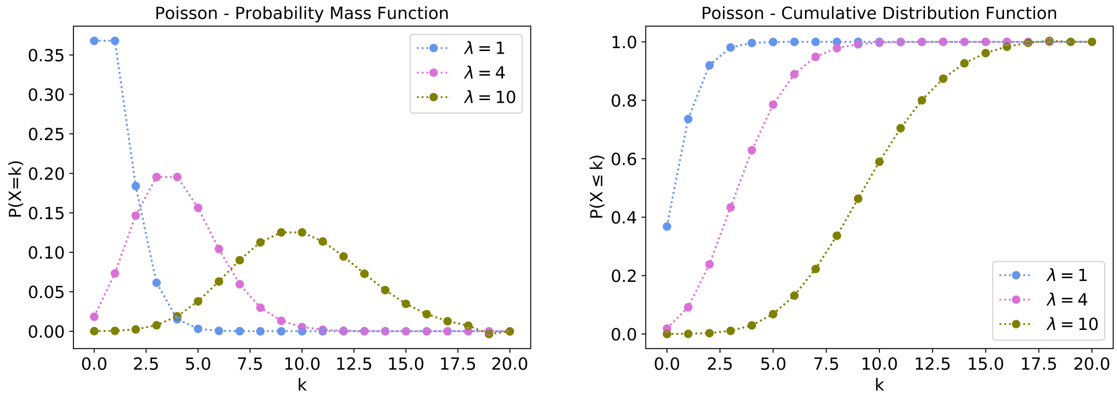

Figure 4: The probability of mass function (left) and cumulative distribution function (right) of a Poisson random variable.

(1) $f(x)$ is positive everywhere, that is, $f(x) > 0$, for all real $x \in (-\infty, \infty)$.

(2) The area under the curve $f(x)$ is 1, that is:

$$P\{X \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f(x)\,dx = 1 \tag{36}$$

(3) If $f(x)$ is the p.d.f. of X, for any set B of real numbers,

$$P\{X \in B\} = \int_B f(x)\,dx\,, \tag{37}$$

which is, the probability that X will be in B may be obtained by integrating the probability density function over the set B.

From the property no (3), if B is some interval, e.g., $B = \big[a, b\big]$, then

$$P\{a \leq X \leq b\} = \int_a^b f(x)\,dx\,. \tag{38}$$

If we let $a = b$, then

$$P\{X = a\} = \int_a^a f(x)\,dx = 0\,, \tag{39}$$

which proves the statement earlier, that the probability that a continuous random variable will assume any particular value is zero.

The cumulative distribution function (c.d.f.) of a continuous random variable X is defined as:

$$F(a) = P\{X \in (-\infty, a)\} = \int_{-\infty}^{a} f(x)\,dx\,. \tag{40}$$

The relationship between the cumulative distribution $F(\cdot)$ and the probability density $f(\cdot)$ is expressed by differentiating both sides of Eq. (40), hence

$$\frac{d}{da}F(a) = f(a)\,. \tag{41}$$

If X is a continuous random variable having a probability density function $f(x)$, then the expected value or mean of X is defined by,

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx\,. \tag{42}$$

The moment generating function $\phi(t)$ of the random variable X is defined for all values t by

$$\begin{aligned}\phi(t) &= E(e^{tX}) \\ &= \int_{-\infty}^{\infty} e^{tx}\, f(x)\, dx\,. \end{aligned} \tag{43}$$

3.1 The Uniform Random Variable

A random variable is said to be uniformly distributed over the interval (α, β) if its probability density function is given by

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & \text{if } \ \alpha < x < \beta \\ \\ 0 & \text{elsewhere} \end{cases} \tag{44}$$

The cumulative distribution function of a random variable uniformly distributed over (α, β) is given by

$$F(c) = \begin{cases} 0 & \text{for } \ c \le \alpha \\ \\ \dfrac{c - \alpha}{\beta - \alpha} & \text{for } \ \alpha < c < \beta \\ \\ 1 & \text{for } \ c \ge \beta \end{cases} \tag{45}$$

The expectation or mean of a random variable uniformly distributed over (α, β),

$$\begin{aligned} E(X) &= \int_\alpha^\beta \frac{x}{\beta - \alpha}\, dx \\ &= \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)} \\ &= \frac{\beta + \alpha}{2}\,. \end{aligned} \tag{46}$$

The moment generating function over an interval (α, β) is given by,

$$\begin{aligned} \phi(t) &= \int_\alpha^\beta e^{tx} \frac{1}{\beta - \alpha}\, dx \\ &= \frac{1}{\beta - \alpha} \int_\alpha^\beta e^{tx}\, dx \\ &= \begin{cases} \dfrac{\left(e^{\beta t} - e^{\alpha t}\right)}{t(\beta - \alpha)} & \text{for} \;\; t \neq 0 \\ \\ 1 & \text{for} \;\; t = 0 \end{cases} \end{aligned} \tag{47}$$

Differentiating the moment generating function, we get

$$\phi'(t) = \frac{e^{\beta t}(\beta t - 1) - e^{\alpha t}(\alpha t - 1)}{t^2(\beta - \alpha)}\,. \tag{48}$$

We see that the moment generating function of a uniform random variable is not differentiable at zero. Hence, the mean and variance should be obtained from calculating the raw moments and central moments.

The variance is given by,

$$\mathrm{Var}(X) = \frac{(\beta - \alpha)^2}{12}\,. \tag{49}$$

Uniform distribution is commonly used for random number generators. In Python, random samples can be generated from a uniform distribution over the interval $[0, 1)$ using,

```
numpy.random.rand()
```

To generate an array of shape $m \times n$, for example 3×2, we define the shape in the arguments as follows,

```
numpy.random.rand(3, 2)
```

For a more general interval range of the uniform distribution, for example over $[a, b]$ where $b > a$, we use

```
(b - a) * numpy.random.random_sample() + a
```

3.2 The Exponential Random Variable

The probability density function of an exponential random variable, with parameter $\lambda > 0$, is given by,

$$f(x) = \begin{cases} \lambda\, e^{-\lambda x} & \text{if } \; x \geq 0 \\ 0 & \text{if } \; x < 0 \end{cases} \tag{50}$$

The cumulative distribution function F is given by,

$$F(a) = \int_0^a \lambda\, e^{-\lambda x}\, dx = 1 - e^{-\lambda a}\,, \qquad a \geq 0 \tag{51}$$

The expectation of an exponential random variable is given by,

$$E(X) = \int_0^\infty x\, \lambda\, e^{-\lambda x}\, dx\,,$$

and after integrating by parts,

$$\begin{aligned} E(X) &= -x\, e^{-\lambda x}\Big|_0^\infty + \int_0^\infty e^{-\lambda x}\, dx \\ &= 0 - \left.\frac{e^{-\lambda x}}{\lambda}\right|_0^\infty \\ &= \frac{1}{\lambda} \end{aligned}$$

The moment generating function of an exponential random variable with parameter λ,

$$\begin{aligned} \phi(t) &= E(e^{tX}) \\ &= \int_0^\infty e^{tx}\, \lambda\, e^{-\lambda x}\, dx \\ &= \lambda \int_0^\infty e^{-(\lambda - t)x}\, dx \\ &= \frac{\lambda}{\lambda - t} \qquad \text{for } \; t < \lambda \end{aligned}$$

After differentiation of $\phi(t)$,

$$\phi'(t) = \frac{\lambda}{(\lambda - t)^2}, \qquad \phi''(t) = \frac{2\lambda}{(\lambda - t)^3}$$

we obtain the mean,

$$E(X) = \phi'(0) = \frac{1}{\lambda}, \tag{52}$$

and the variance,

$$\text{Var}(X) = \phi''(0) - (\phi'(0))^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda} = \frac{1}{\lambda^2}. \tag{53}$$

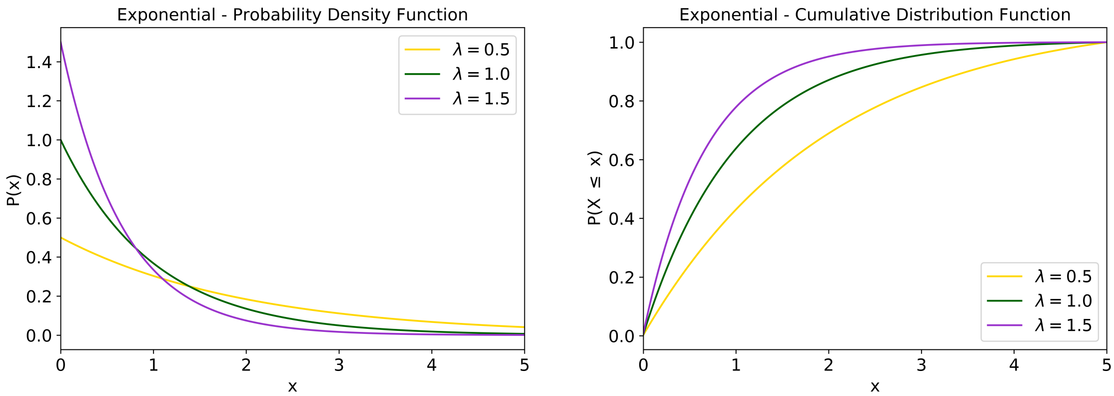

Figure 5: The probability density function (left) and cumulative distribution function (right) of an exponential random variable.

To generate random samples from an exponential distribution in Python,

```
numpy.random.exponential()
```

3.3 The Gamma Random Variable

If a continuous random variable has the following probability density function

$$f(x) = \begin{cases} \dfrac{\lambda\, e^{-\lambda x}(\lambda\, x)^{\alpha-1}}{\Gamma(\alpha)} & \text{for } \; x \geq 0 \\[2ex] 0 & \text{for } \; x < 0 \end{cases} \tag{54}$$

for some $\lambda > 0$ and $\alpha > 0$, then it is said to be a gamma random variable with parameters α and λ. The quantity $\Gamma(\alpha)$ is called the gamma function and defined by,

$$\Gamma(\alpha) = \int_0^{\infty} e^{-x}\, x^{\alpha-1}\, dx$$

If $\alpha = n$ is a positive integer, then,

$$\Gamma(n) = (n-1)!$$

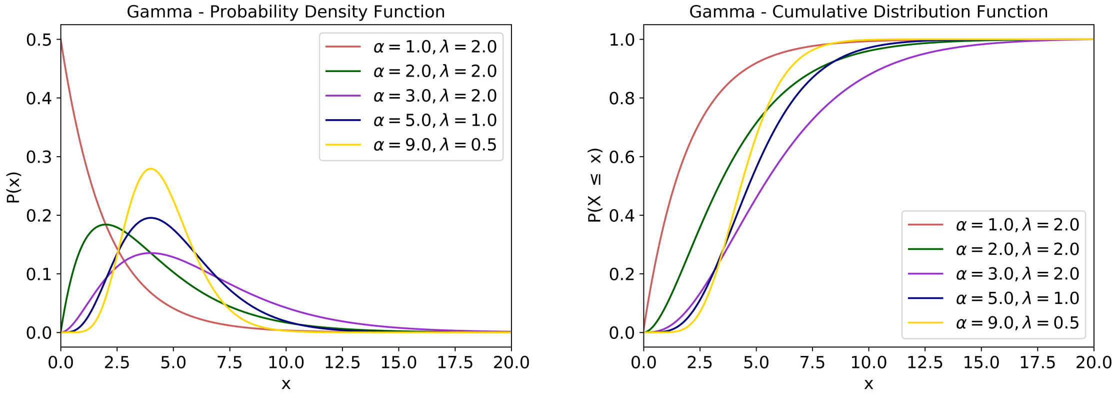

Figure 6: The probability density function (left) and cumulative distribution function (right) of a gamma random variable.

3.4 The Normal Random Variable

A random variable X is normally distributed with parameters μ and σ^2 if the density is given by,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/2\sigma^2}\,, \qquad -\infty < x < \infty \tag{55}$$

The expectation of a normal random variable distributed with parameters μ and σ^2 is,

$$E(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x\, e^{-(x-\mu)^2/2\sigma^2}\, dx$$

If we write $x = (x-\mu) + \mu$, the equation is expanded,

$$E(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)\, e^{-(x-\mu)^2/2\sigma^2}\, dx + \mu\, \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2}\, dx$$

Letting $y = x - \mu$ gives us,

$$E(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} y\, e^{-y^2/2\sigma^2}\, dy + \mu \int_{-\infty}^{\infty} f(x)\, dx$$

where $f(x)$ is the normal density. By symmetry, the first integral is 0, then the expectation is,

$$E(X) = \mu \int_{-\infty}^{\infty} f(x)\, dx = \mu\,. \tag{56}$$

The moment generating function of a standard normal random variable Z is,

$$\begin{aligned} E(e^{tZ}) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx}\, e^{-x^2/2}\, dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x^2-2tx)/2}\, dx \\ &= e^{t^2/2}\, \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-t)^2/2}\, dx \\ &= e^{t^2/2}\,. \end{aligned}$$

If Z is a standard normal, then $X = \sigma Z + \mu$ is normal with parameters μ and σ^2, then

$$\begin{aligned} \phi(t) &= E(e^{tX}) \\ &= E(e^{t(\sigma Z+\mu)}) \\ &= e^{t\mu} E(e^{t\sigma Z}) \\ &= \exp\left(\frac{\sigma^2 t^2}{2} + \mu t\right)\,. \end{aligned}$$

After differentiation,

$$\phi'(t) = (\mu + t\sigma^2)\, \exp\left(\frac{\sigma^2 t^2}{2} + \mu t\right)\,,$$

and

$$\phi''(t) = (\mu + t\sigma^2)^2\, \exp\left(\frac{\sigma^2 t^2}{2} + \mu t\right) + \sigma^2\, \exp\left(\frac{\sigma^2 t^2}{2} + \mu t\right)\,,$$

we obtain the mean,

$$E(X) = \phi'(0) = \mu\,, \tag{57}$$

and the variance,

$$\mathrm{Var}(X) = \phi''(0) - (\phi'(0))^2 = \sigma^2\,. \tag{58}$$

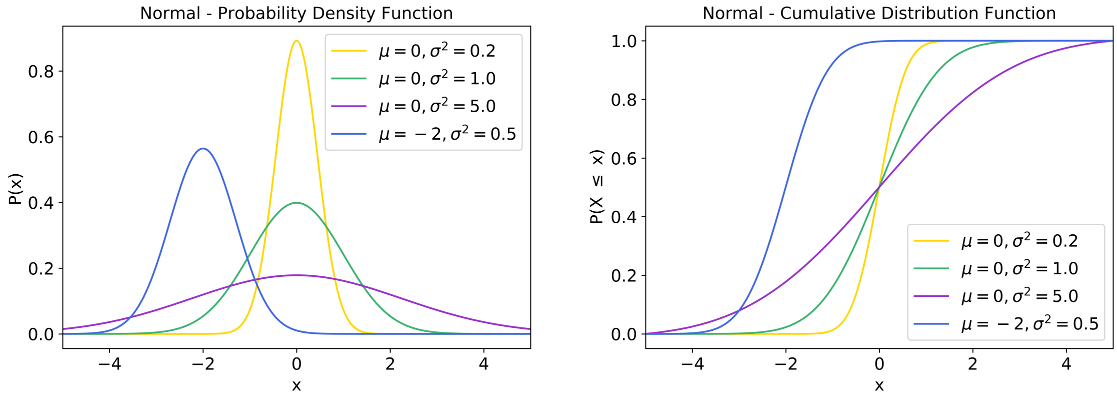

Figure 7: The probability density function (left) and cumulative distribution function (right) of a normal random variable.

Normal distribution is another commonly used distribution for random number generators. To generate random samples from the standard normal distribution, with $\mathcal{N}(0,1)$ or mean 0 and variance 1, in Python, use

```
numpy.random.randn()
```

To generate random samples from $\mathcal{N}(\mu, \sigma^2)$, use

```
sigma * numpy.random.randn(...) + mu
```

4 Jointly Distributed Random Variables

This topic is also called multivariate distributions in some textbooks.

Thus far we have discussed the probability distribution of a single random variable. Often times, we are also interested in probability statements of two or more random variables.

When several random variables $X_1, X_2, \cdots, X_n$ are associated with the same sample space, a joint probability mass function or probability density function $f(x_1, x_2, \cdots, x_n)$ can be defined. For example, for two random variables X_1 and X_2 defined on a common sample space $\mathcal{S}$, (X_1, X_2) is a random vector where $(X_1, X_2) : \mathcal{S} \to \mathbb{R}^2$. For each element s in the sample space $\mathcal{S}$, there is associated a unique ordered pair $(X_1(s), X_2(s))$. The set of all ordered pairs is given by [1],

$$\mathcal{A} = \{(X_1(s), X_2(s)) \mid s \in \mathcal{S}\} \subset \mathbb{R}^2 \,, \tag{59}$$

which is known as the *state space* or *space of the random vector* (X_1, X_2).

The following are joint probability distributions of two random variables, where for the sake of convenience, denoted X and Y (instead of X_1 and X_2):

- Let X and Y be two discrete random variables defined on a common sample space $\mathcal{S}$, having probability measure $P : \mathcal{S} \to [0,1]$. If there exists a function $f : \mathcal{A} \to [0,1]$ which is defined by
$$f(x,y) = P\{X = x, Y = y\}\,, \tag{60}$$
then $f(x,y)$ is called the **joint probability mass function** of the random variables X and Y if it satisfies the following conditions [3]:
  (1) Each probability must be a valid probability number between 0 and 1 (inclusive):
$$0 \le f(x,y) \le 1\,.$$
  (2) The sum of the probabilities over the entire $\mathcal{A}$ must equal 1:
$$\sum\sum_{(x,y)\in\mathcal{A}} f(x,y) = 1\,.$$
  (3) To determine the probability of an event B, simply sum up the probabilities of the (x,y) values in B:
$$P\{(X,Y) \in B\} = \sum\sum_{(x,y)\in B} f(x,y)\,,$$
  where $B \subset \mathcal{A}$, and $\mathcal{A}$ is defined by Eq. (59).

- Let X and Y be two discrete random variables as defined above, whose joint probability mass function $f(x,y)$ is defined by Eq. (60). The probability mass function of X alone, which is called the **marginal probability mass function of $\boldsymbol{X}$** [3], may be obtained from $f(x,y)$ by
$$f_X(x) = \sum_y f(x,y) = P\{X = x\}\,,$$
which states that, for each x, the summation is taken over all possible values of y. Similarly, the probability mass function of Y alone, which is called the **marginal probability mass function of $\boldsymbol{Y}$**, may be obtained from
$$f_Y(y) = \sum_x f(x,y) = P\{Y = y\}\,.$$

- For any two random variables X and Y, the **joint cumulative probability distribution function** of X and Y is defined by [2]
$$F(a,b) = P\{X \le a, Y \le b\}\,, \qquad -\infty < a,b < \infty$$

The cumulative distribution function of X can be obtained from the joint distribution of X and Y as follows:

$$\begin{aligned} F_X(a) &= P\{X \leq a\} \\ &= P\{X \leq a, Y < \infty\} \\ &= F(a, \infty)\,. \end{aligned}$$

Similarly, the cumulative distribution function of Y is given by

$$F_Y(b) = P\{Y \leq b\} = F(\infty, b)\,.$$

- Let X and Y be two continuous random variables defined on a common sample space $\mathcal{S}$, having probability measure $P : \mathcal{S} \to [0, 1]$. If there exists a function $f : \mathbb{R}^2 \to [-\infty, \infty]$ which is defined by

  $$\iint\limits_{\mathbb{R}^2} f(x,y)\,dx\,dy = \iint\limits_{\mathcal{A}} f(x,y)\,dx\,dy = 1$$

  then $f(x, y)$ is called the **joint probability density function** of X and Y if it satisfies the following properties:

  – The probability for all sets $\mathcal{A}$ and $\mathcal{B}$ of real numbers [2]:

  $$P\{X \in A, Y \in B\} = \int_B \int_A f(x,y)\,dx\,dy\,,$$

  where, the sets $\mathcal{A}$ and $\mathcal{B}$ are defined by Eq. (59).

  – The probability of an event B, where $B \subset \mathcal{A}$ [1]:

  $$P\{(X,Y) \in B\} = \iint\limits_{B} f(x,y)\,dx\,dy\,,$$

  where, the set $\mathcal{A}$ is defined by Eq. (59).

  The function $f(x, y)$ above is defined for all real x and y. We then say that X and Y are **jointly continuous**.

- From the definition of continuous random variables X and Y above, the probability density of X alone, which is called the **marginal probability density function of** X, can be obtained from a knowledge of $f(x, y)$ by

  $$\begin{aligned} P\{X \in \mathcal{A}\} &= P\{X \in \mathcal{A},\ Y \in (-\infty, \infty)\} \\ &= \int_{-\infty}^{\infty} \int_{\mathcal{A}} f(x,y)\,dx\,dy \\ &= \int_{\mathcal{A}} f_X(x)\,dx \end{aligned}$$

where

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

is the probability density function of X. Similarly, the probability density function of Y is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx\,.$$

- Discrete random variables X and Y are **independent** if and only if [3]

  $$P\{X = a, Y = b\} = P\{X = a\} \times P\{Y = b\} \qquad \forall\, a, b\,.$$

  Similarly, continuous random variables X and Y are independent if [2]

  $$P\{X \le a, Y \le b\} = P\{X \le a\} \times P\{Y \le b\} \qquad \forall\, a, b\,.$$

  Otherwise, X and Y are said to be **dependent**.

- If X and Y are random variables whose joint probability mass/density function is $f(x, y)$ and $u(X, Y)$ is a function of these two variables, then [2]:

  $$E[u(X, Y)] = \begin{cases} \sum\limits_y \sum\limits_x u(x, y)\, f(x, y) & \text{in the discrete case} \\ \\ \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} u(x, y)\, f(x, y)\, dx\, dy & \text{in the continuous case} \end{cases}$$

  where $E[u(X, Y)]$ is called the **expected value** of $u(X, Y)$. For example, if X and Y are two continuous random variables and $u(X, Y) = X + Y$, then:

  $$\begin{aligned} E[X + Y] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y)\, f(x, y)\, dx\, dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x\, f(x, y)\, dx\, dy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y\, f(x, y)\, dx\, dy \\ &= E[X] + E[Y]\,. \end{aligned}$$

  If X and Y are two discrete random variables, then:

  $$E[aX + bY] = a\, E[X] + b\, E[Y]\,.$$

- Let X and Y be discrete random variables. If $u(X, Y) = (X - \mu_X)^2$, the variance of X is given by, [3],

  $$\sigma_X^2 = \mathrm{Var}(X) = \sum_x \sum_y (x - \mu_X)^2\, f(x, y)\,,$$

or

$$\sigma_X^2 = E(X^2) - \mu_X^2 = \left(\sum_x \sum_y x^2 \, f(x,y) \right) - \mu_X^2 \, .$$

And, if $u(X,Y) = (Y - \mu_Y)^2$, the variance of Y is given by

$$\sigma_Y^2 = \mathrm{Var}(Y) = \sum_x \sum_y (y - \mu_Y)^2 \, f(x,y) \, ,$$

or

$$\sigma_Y^2 = E(Y^2) - \mu_Y^2 = \left(\sum_x \sum_y y^2 \, f(x,y) \right) - \mu_Y^2 \, .$$

- The covariance of any two random variables X and Y, denoted by $\mathrm{Cov}(X,Y)$, is defined by [2]

$$\begin{aligned} \mathrm{Cov}(X,Y) &= E[(X - E[X])\,(Y - E[Y])] \\ &= E[XY - Y\,E[X] - X\,E[Y] + E[X]\,E[Y]] \\ &= E[XY] - E[Y]\,E[X] - E[X]\,E[Y] + E[X]\,E[Y] \\ &= E[XY] - E[X]\,E[Y] \end{aligned}$$

If X and Y are independent, then $\mathrm{Cov}(X,Y) = 0$.

For any random variables X, Y, Z, and constant c, the following are properties of covariance [2]:

1. $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$,
2. $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$,
3. $\mathrm{Cov}(cX,Y) = c\,\mathrm{Cov}(X,Y)$,
4. $\mathrm{Cov}(X,Y+Z) = \mathrm{Cov}(X,Y) + \mathrm{Cov}(X,Z)$.

5 Conditional Probabilities

Let A and B be two events in a probability space. The conditional probability of event A given that event B has occurred, denoted $P(A|B)$, is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \, , \tag{61}$$

provided that $P(B) > 0$. Similarly, the conditional probability of event B given that event A has occurred, denoted $P(B|A)$, is defined as

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \, , \tag{62}$$

provided that $P(A) > 0$.

Events A and B are said to be **independent** if the occurrence of either one of the events does not affect the probability of occurrence of the other, that is,

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

Therefore, A and B are independent if and only if

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)\, P(B)}{P(B)} = P(A)\,, \tag{63}$$

which states that the probability of A does not depend on whether B has occurred. Similarly,

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)\, P(B)}{P(A)} = P(B)\,. \tag{64}$$

In other words, the events A and B are independent if the probability of B does not depend on whether A has occurred.

Since conditional probability is just a probability, it satisfies the following three properties [3]:

1. $P(A|B) \geq 0$.
2. $P(B|B) = 1$.
3. If $A_1, A_2, \cdots, A_k$ are mutually exclusive events, then

   $$P(A_1 \cup A_2 \cup \cdots \cup A_k \,|\, B) = P(A_1|B) + P(A_2|B) + \cdots + P(A_k|B)$$

   and likewise for infinite unions.

All three properties above are satisfied as long as $P(B) > 0$.

References

[1] Linda J. S. Allen. *An Introduction to Stochastic Processes with Applications to Biology 2nd Edition*. CRC Press, 2010.

[2] Sheldon M. Ross. *Introduction to Probability Models 9th Edition*. Elsevier, 2007.

[3] Introduction to Probability Theory,
https://newonlinecourses.science.psu.edu/stat414/node/287/