# 2 Random Variables

*ENSY SILVER*[1]

Wednesday 29th July, 2020

[1]Thanks to my family, my friend and freedom.

# 1 Motivation

In order to simplify the description of sample outcome, we use **random variables** to denote the sample outcome. For example, if we have 8 volunteers in a hospital who test vaccine of COVID-19, the number of people who have antibody is denoted by $x$. We call $x$ the random variable, and we only need to to denote the sample outcome, because it is worthless to discuss every volunteers' situation in statistics.

Moreover, random variables fall into one of two broad categories. **continuous** or **discrete**. Details will be demonstrated in the rest sections. In general, a function that assigns numbers to outcomes is called a *random variable*.

# 2 Binomial probability

Binomial probabilities apply to situations involving a series of independent and identical trials, where each trial have only one of two possible outcomes. We suppose the probability of one possible outcome is $p$, and we deduce the next theorem.

**Theorem 2.1 (Binomial distribution).** *Consider a series of $n$ independent trials, each resulting in one of two possible outcomes, "success" or "failure".*

*Let $p = P(\text{success occurs at any given trial})$ and assume that $p$ remains constant from trial to trial. Then*

$$P(k\ \textit{success}) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad \textit{where} \quad k = 0, 1, \cdots, n$$

# 3 Hypergeometric distribution

In the previous section, we have learned the definition and properties of binomial probability. In this section we will learn the hypergeometric distribution, which is more complicated.

**Theorem 3.1 (Hypergeometric distribution).** *Suppose an urn contains $r$ red chips and $w$ blue chips, where $r + w = N$. If $n$ chips are drawn out at random, without replacement, and if $k$ denotes the number of red chips selected, then*

$$P(k\ \textit{red chips chosen}) = \frac{\binom{r}{k}\binom{w}{n-k}}{\binom{N}{n}}$$

*where $k$ varies over all non-negative integers.*

In fact, the name *hypergeometric* derives from a series introduced by Leonhard Euler, in 1769:

$$1 + \frac{ab}{1}x + \frac{a(a+1)b(b+1)}{2!c(c+1)}x^2 + \cdots + \frac{\prod_{i=0}^{n}(a+i)\prod_{j=0}^{n}(b+j)}{n!\prod_{k=0}^{n}(c+k)} + \cdots \tag{1}$$

This is an expansion of considerable flexibility: Given appropriate values for $a, b, c$, it reduces to many of the standard infinite series used in analysis. In particular, if $a$ is set equal to 1, and $b, c$ are set equal to each other, it reduces to the familiar *geometric series*,

$$1 + x + \cdots + x^n + \cdots$$

hence the name of 1 is *hypergeometric*. Once we set $a = -n, b = -r, c = w-n+1$, and multiply the series by $\binom{w}{n}/\binom{N}{n}$. Then the coefficient of $x^k$ will be

$$\frac{\binom{w}{n}\binom{w}{n-k}}{\binom{N}{n}}$$

# 4 Discrete random variables

Since we have introduced the case of COVID-19 vaccine trial, we can easily have the definition of **discrete random variables**.

**Theorem 4.1.** *Suppose that $S$ is a finite or countably infinite sample space. Let $p$ be a real-valued function defined for each element of $S$ such that*

1. $0 \leq p(s)$ *for each* $s \in S$.
2. $\sum_{s \in S} = 1$

*Then $p$ is said to be a discrete probability function.*

**Theorem 4.2.** *A function whose domain is a sample space $S$ whose values form a finite or countably infinite set of real numbers is called a discerete random variable. We denote random variables by uppercase letters, often $X$ or $Y$.*

In anology with discrete probability function $p(s)$, we introduce **probability density function**, which assigns probability to a random variable.

**Theorem 4.3.** *Associated with every discrete random variable $X$ is a probability density function, denoted by $p_X(k)$, where*

$$p_X(k) = P(\{s \in S | X(s) = k\})$$

*Note that $p_X(k) = 0$ for any $k$ not in the range of $X$. For notational simplicity, we will usually delete all references to $s$ and $S$ and write $p_X(k) = P(x = k)$.*

Sometimes, we need to calculate the value of a random variable is somewhere between two numbers. For example,suppose we have an interger-valued random variable. We might want to calculate $P(s \leq X \leq t)$. If we know the probability density function for $X$, then

$$P(s \leq X \leq t) = \sum_{k=s}^{t} p_X(k)$$

**Theorem 4.4.** *Let $X$ be a discrete random variable. For any real number $t$, the probability that $X$ takes on a value $\leq t$ is the* ***cumulative distribution function*** *of $X$. In formal notation, cumulative distribution function is denoted by*

$$F_X(t) = P(\{s \in S | X(s) \leq t\})$$

*As was the case with probability distribution functions, cumulative distribution function is written $F_X(t) = P(X \leq t)$.*

**Remark.** *For simplicity, the abbreviation of probability density function is pdf, and the abbreviation of cumulative distribution function is cdf.*

In conclusion, we use a commutative diagram to describe the relationship among discrete random variables, discrete probability functions, probability density functions and cumulative distribution function.

We denote sample space by $S$, compressed sample space by $U$ ($U \in \mathbb{R}$), image of pdf by $V$, and image of cdf by $W$. Moreover, we use $p$ to denote discrete probability function, $p_X$ to denote pdf, and $F_X$ to denote cdf.

$$\begin{tikzcd}
& S \arrow[d, "X"] \arrow[ddl, bend right, "p"'] & \\
& U \arrow[dl, "p_X"] \arrow[dr, "F_X"] & \\
V \arrow[rr, "sum"'] & & W
\end{tikzcd} \tag{2}$$

**Remark.** *The symbols above will also represent the same meaning in the rest of the book.*

# 5 Continuous random variables

Since we have defined discrete probability functions in theorem 4.1, the definition of **continuous probability functions** are anologous. For a continuous sample space $S$, the probability function $p$ and random variable are also continuous. Then we have the definition of continuous probability function.

**Theorem 5.1.** *A probability function $P$ on a set of real numbers $S$ is called continuous if there exists a function $f$ such that for any closed interval $[a, b] \subset S$,*

$$P([a, b]) = \int_a^b f(t)\, dt$$

Applying the properties of $P$, we hence have the corollary.

**Corollary 5.1.1.** *For function $f$ such that $P(A) = \int_A f(t)\, dt$ for any set $A$ in $S$ where the integral is defined.*

1. *$f(t) \geq 0$ for all $t$.*

2. $\int_{-\infty}^{\infty} f(t)\,dt = 1$

In anology with discrete random variables, we define the **continuous random variable** by the next theorem.

**Theorem 5.2.** *Le $Y$ be a function from a sample space $S$ to the real numbers. The function $Y$ is called a continuous random variable if there exists a function $f_Y(y)$ such that for any real numbers $a$ and $b$ with $a < b$*

$$P(a \le Y \le b) = \int_a^b f_Y(y)\,dy$$

**Theorem 5.3.** *The cdf for a continuous random variable $Y$ is an infinite integral of its pdf:*

$$F_Y(y) = \int_{-\infty}^{y} f_Y(r)\,dr = P(Y \le y)$$

We can simply deduce two corollaries.

**Corollary 5.3.1.** *Let $F_Y(y)$ be the cdf of a continuous random variable $Y$. Then*

$$\frac{d}{dy}F_Y(y) = f_Y(y)$$

**Corollary 5.3.2.** *Let $Y$ be a continuous random variable with cdf $F_Y(y)$. Then*

1. $P(Y > s) = 1 - F_Y(s)$.
2. $P(r < Y \le s) = F_Y(s) - F_Y(r)$
3. $\lim_{y\to\infty} F_Y(y) = 1$.
4. $\lim_{y\to-\infty} F_Y(y) = 0$.

# 6 Expected values

Probability density functions, as we have already seen, provide a global overview of a random variable's behavior. Then most concerned property is called **expected value**, which is denoted by $\mu$. Now we give the formal definition.

**Theorem 6.1.** *Let $X$ be a discerete random variable with probability function $p_X(k)$. The expected value of $X$ is denoted by $\mu_X$ or $E(X)$, given by*

$$E(X) = \mu_X = \sum_{k\in U} k \cdot p_X(k)$$

*Similarly, if $Y$ is a continuous random variable with pdf $f_Y(y)$,*

$$E(Y) = \mu_Y = \int_{-\infty}^{\infty} y \cdot f_Y(y)\,dy$$

Reviewing the definition of *binomial random variable* and *hypergeometric random variable*, we have two theorems.

**Theorem 6.2.** *Suppose $X$ is a binomial random variable with parameters $n$ and $p$. Then*

$$E(X) = np$$

**Theorem 6.3.** *Suppose $X$ is a hypergeometric random variable with parameters $r$, $w$, and $n$. That is, $r$ is the number of red balls and $w$ the blue balls. A sample of size $n$ is drawn simultaneously from the urn. Let $X$ be the number of red balls in the sample. Then*

$$E(X) = \frac{rn}{r+w}$$

## 6.1 Median

While the expected value is the most frequently used measure of a random variable's central tendency, it does have a weakness that sometimes makes it misleading and inappropriate. Specifically, the expected values can be distorted when some possible values are too large or too small.

**Theorem 6.4.** *If $X$ is a discrete random variable, the median, $m$, is that point for which $P(X < m) = P(X > m)$. In the event that $P(X \leq m) = 0.5$ and $P(x \geq m') = 0.5$, the median is defined to be the arithmetic average,*

$$(m + m')/2$$

*If $Y$ is a continuous random variable, its median is the solution to the integral equation*

$$\int_{-\infty}^{m} f_Y(y)\, dy = 0.5$$

## 6.2 The expected value of a function of a random variable

There are many situations that call for finding the expected value of a function of a random variable, for example, $Y = g(X)$, where $Y$ is also a random variable. Our task in this section is to build the link between $E(Y)$ and $E(X)$.

**Theorem 6.5.** *Suppose $X$ is a discerete random variable with pdf $p_X(k)$. Let $g(X)$ be a function of $X$. Then the expected value of the random variable $g(X)$ is given by*

$$E[g(x)] = \sum_{k \in U} g(k) \cdot p_X(k)$$

*provided that $\sum_{k \in U} |g(k)| p_X(k) < \infty$.*

*Similarly, if $Y$ is a continuous random variable with pdf $f_Y(y)$, and if $g(Y)$ is a continuous function, the the expected value of the random variable $g(Y)$ is*

$$E[g(Y)] = \int_{-\infty}^{\infty} g(y) \cdot f_Y(y)\, dy$$

*provided that $\int_{-\infty}^{\infty} |g(y)| f_Y(y)\, dy < \infty$.*

For special case $Y = aX + b$, we have corollary

**Theorem 6.6.** *For any random variable $X$, $E(aX + b) = aE(X) + b$, where $a, b$ are constants.*

# 7 The variance

Both expected value and median the most possible value of a random variable, but the pdf has much more information than that. The pdf records the dispersion of a random variable, hence we need to quantify the dispersion, which is the motivation of **variance**.

**Theorem 7.1.** *The variance of a random variable is the expected value of its squared deviations from $\mu$. If $X$ is discerete, with pdf $p_X(k)$.*

$$\mathrm{Var}(X) = \sigma^2 = E[(X - \mu)^2] = \sum_{k \in U} (k - \mu)^2 \cdot p_X(k)$$

*If $Y$ is continuous, with pdf $f_Y(y)$,*

$$\mathrm{Var}(Y) = \sigma^2 = E[(Y - \mu)^2] = \int_{-\infty}^{\infty} (y - \mu)^2 \cdot f_Y(y)\, dy$$

**Remark.** *$\sigma$ is called the standard deviation.*

From the definition of variance, we can easily deduce the relationship between variance $\mathrm{Var}(W)$ and expected value $\mu$.

**Theorem 7.2.** *Let $W$ be any random variable, discrete or continuous, having mean $\mu$ and for which $E(W^2)$ is finite. Then*

$$\mathrm{Var}(W) = \sigma^2 = E(W^2) - \mu^2$$

If variance of a random variable $W$ is known, we can also deduce the variance of linear transformed random variable $aW + b$.

**Theorem 7.3.** *Let $W$ be any random variable having mean $\mu$ and where $E(W^2)$ is finite. Then*

$$\mathrm{Var}(aW + b) = a^2\, \mathrm{Var}(W)$$

## 7.1 Moment

In mathematics, a **moment** is a specific quantitative measure of the shape of a function. For example, $E(W)$ is the *first moment about the origin* and $\sigma^2$ is the *second moment about the mean*. The precise definition of moments is introduced by the next theorem.

**Theorem 7.4.** *Let $W$ be any random variable with pdf $f_W(w)$. For any positive integer $r$,*

1. *The $r^{th}$ moment of $W$ about the origin, $\mu_r$, is given by*

$$\mu_r = E(W^r)$$

*provided $\int_{-\infty}^{\infty} |w|^r \cdot f_W(w)\ dw < \infty$. When $r = 1$, we usually delete the subscript and write $E(W)$ as $\mu$.*

2. *The $r^{th}$ moment of $W$ about the mean, ${\mu_r}'$ is given by*

$${\mu_r}' = E[(W-\mu)^r]$$

*provided the finiteness condition of 1 hold.*

Applying Newton's binomial expansion, we have the corollary.

**Corollary 7.4.1.** *We can express ${\mu'}_r$ in terms of $\mu_j$, where $j \in \mathbb{R}$,*

$${\mu'}_r = E[(W-\mu)^r] = \sum_{j=0}^{r} \binom{r}{j} E(W^j)(-\mu)^{r-j} = \sum_{j=0}^{r} \binom{r}{j} \mu_j(-\mu)^{r-j}$$

In order to estimate the existence of moments, we deduced the theorem.

**Theorem 7.5.** *If the $k^{th}$ moment of a random variable exists, all moments of order less than $k$ exists.*

## 7.2 Skewness

**Skewness** is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean. The skewness value can be positive, zero, negative, or undefined. *Coefficient of skewness* is formally defined by

$$\gamma_1 = \frac{E[(W-\mu)^3]}{\sigma^3}$$

where dividing ${\mu'}_3$ by $\sigma^3$ makes $\gamma_1$ dimensionless. If the image of pdf is symmetric, $\gamma_1 = 0$.

## 7.3 Kurtosis

**Kurtosis** is a useful measure of peakedness: relatively flat pdfs are said to be *platykurtic*, more peaked pdfs are called *leptokurtic*. Precise definition of *coefficient of Kurtosis* is given by

$$\gamma_2 = \frac{E[(W-\mu)^4]}{\sigma^4} - 3$$

More detailed introduction to kurtosis is in Wikipedia.

# 8 Joint densities

The definition of **joint probability density fuction** directly follows definition of pdf.

## 8.1 Discrete joint pdf

**Theorem 8.1.** *Suppose $S$ is a discrete sample space on which two random variables, $X$ and $Y$, are defined. The joint probability density function of $X$ and $Y$ is denoted by $p_{X,Y}(x,y)$, where*

$$p_{X,Y}(x,y) = P(\{s|X(s) = x, Y(s) = y\})$$

Since we have introduced the definition of joint definition, we can also get the pdf of a single variable from the information from joint pdf.

**Theorem 8.2.** *Suppose that $p_{X,Y}(x,y)$ is the joint pdf of the discrete random variables $X$ and $Y$. Then*

$$p_X(x) = \sum_{y \in U_y} p_{X,Y}(x,y) \quad \textit{and} \quad p_Y(y) = \sum_{x \in U_x} p_{X,Y}(x,y)$$

## 8.2 Continuous pdf

The definition and properties of continuous version is anologous to the discrete version.

**Theorem 8.3.** *Two random variables defined on the same set of real numbers are jointly continuous if there exists a function $f_{X,Y}(x,y)$ such that for any region $U$ in the $xy$-plane,*

$$P[(X,Y) \in U] = \int\int_U f_{X,Y}(x,y)\, dx\, dy$$

*The function $f_{X,Y}(x,y)$ is the joint pdf of $X$ and $Y$.*

The **marginal pdf** of continuous random variables are as follows.

**Theorem 8.4.** *Suppose $X$ and $Y$ are jointly continuous with joint pdf $f_{X,Y}(x,y)$. Then the marginal pdfs, $f_X(x)$ and $f_Y(y)$ are given by*

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy \quad \textit{and} \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx$$

## 8.3 Geometric probability

One particular special case of continuous pdf is the **joint uniform pdf**, which is represented by a surface having a constant height every where above a specified rectangle in the $xy$-plane. That is

$$f_{X,Y}(x,y) = \frac{1}{(b-a)(d-c)} \quad a \le a \le b, \;\; c \le y \le d$$

## 8.4 Joint cdf

Similar to cdf, we can also define the joint cdf as follows.

**Theorem 8.5.** *Let $X$ and $Y$ be any two random variables. The joint cumulative distribution function of $X$ and $Y$ is denoted $F_{X,Y}(u,v)$, where*

$$F_{X,Y}(u,v) = P(X \leq u, Y \leq v)$$

**Corollary 8.5.1.** *Let $F_{X,Y}(u,v)$ be the joint cdf associated with the continuous random variables $X$ and $Y$. Then the joint pdf of $X$ and $Y$, $f_{X,Y}(x,y)$, is a second partial derivative of the joint cdf, that is,*

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x\, \partial y} F_{X,Y}(x,y)$$

*has continuous second partial derivatives.*

*Moreover, we can generalize the corollary to multivariate density, that is, given the joint pdf $f_{X_1,\cdots,X_n}(x_1,\cdots,x_n)$, and the multivariate cdf $F_{X_1,\cdots,X_n}(x_1,\cdots,x_n)$, they have relationship*

$$f_{X_1,\cdots,X_n}(x_1,\cdots,x_n) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F_{X_1,\cdots,X_n}(x_1,\cdots,x_n)$$

## 8.5 Independence

In mathematics, independence of different variables are defined as follows.

**Theorem 8.6.** *The $n$ random variables $X_1,\cdots,X_n$ are said to be independent if there are functions $g_1(x_1),\cdots,g_n(x_n)$ such that for every $x_1,\cdots,x_n$,*

$$f_{X_1,\cdots,X_n}(x_1,\cdots,x_n) = \prod_{j=1}^{n} g_j(x_j)$$

# 9 Transforming and combining random variables

## 9.1 Linear transformation

Once we have known the pdf of a random variable, we also want to know the pdf of its transformation. Now we introduce the case of linear transformation.

**Theorem 9.1.** *Suppose $X$ is a discrete random variable. Let $Y = aX + b$, where $a$ and $b$ are constants. Then*

$$p_Y(y) = p_X(\frac{y-b}{a})$$

The continuous version has a little difference.

**Theorem 9.2.** *Suppose $X$ is a continuous random variable. Let $Y = aX + b$, where $a \neq 0$ and $b$ is a constant. Then*

$$f_Y(y) = \frac{1}{|a|} f_x(\frac{y-b}{a})$$

**Remark.** *The difference between discrete version and continuous version comes from the differential operator.*

## 9.2 Pdf of a sum

**Theorem 9.3.** *Suppose that $X$ and $Y$ are independent random variables. Let $W = X + Y$.*

*If $X$ and $Y$ are discrete random variables with pdfs $p_X(x)$ and $p_Y(y)$, respectively,*

$$p_W(w) = \sum_{x \in U} p_X(x) p_Y(w - x)$$

*If $X$ and $Y$ are continuous random variables with pdfs $f_X(x)$ and $f_Y(y)$, respectively,*

$$f_W(w) = \int_{-\infty}^{\infty} f_X(x) f_Y(w - x)\, dx$$

## 9.3 Pdf of a quotient

**Theorem 9.4.** *Let $X$ and $Y$ be independent continuous random variables, with pdfs $f_X(x)$ and $f_Y(y)$, respectively. Assume that $X$ is zero for at most a set of isolated points. Let $W = Y/X$. Then*

$$f_W(w) = \int_{-\infty}^{\infty} |x| f_X(x) f_Y(wx)\, dx$$

## 9.4 Pdf of a product

Let $X$ and $Y$ be independent continuous random variables with pdfs $f_X(x)$ and $f_Y(y)$, respectively, Let $W = XY$. Then

$$f_W(w) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_X(x) f_Y(\frac{w}{x})\, dx = \int_{-\infty}^{\infty} \frac{1}{|x|} f_X(\frac{w}{x}) f_Y(x)\, dx$$

# 10 Mean and variance for multivariate random variable

First, the mean of two random variables is obvious.

**Theorem 10.1.** *Suppose $X$ and $Y$ are discrete random variables with joint pdf $p_{X,Y}(x,y)$, and let $g(X,Y)$ be a function of $X$ and $Y$. Then the expected value of the random variable $g(X,Y)$ is given by*

$$E[g(X,Y)] = \sum_{x\in U_x}\sum_{y\in U_y} g(x,y)\cdot p_{X,Y}(x,y)$$

*provided $\sum_{x\in U_x}\sum_{y\in U_y} g(x,y)\cdot p_{X,Y}(x,y) < \infty$.*

*Suppose $X$ and $Y$ are continuous random variables with joint pdf $f_{X,Y}(x,y)$, and let $g(X,Y)$ be a function of $X$ and $Y$.*

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\cdot f_{X,Y}(x,y)\,dx\,dy$$

The expected value of multivariate random variable summation is also simple.

**Theorem 10.2.** *Let $W_1,\cdots,W_n$ be any random variables for which $E(W_i) < \infty$, where $i = 1,\cdots,n$, and let $a_1,\cdots,a_n$ be constants. Then*

$$E(\sum_{i=1}^{n} a_i W_i) = \sum_{i=1}^{n} a_i E(W_i)$$

For $g(X,Y) = XY$ and $X,Y$ are independent, we have the special case.

**Theorem 10.3.** *If $X$ and $Y$ are independent random variables,*

$$E(XY) = E(X)\cdot E(Y)$$

*provided $E(X)$ and $E(Y)$ both exist.*

## 10.1 Covariance

In probability theory and statistics, **covariance** is a measure of the joint variability of two random variables. If the greater values of one variable mainly correspond with the greater values of the other variable, and the same holds for the lesser values, the covariance is positive.

**Theorem 10.4.** *Given random variables $X$ and $Y$ with finite variances, define the covariance of $X$ and $Y$, written $\operatorname{Cov}(X,Y)$, as*

$$\operatorname{Cov}(X,Y) = E(XY) - E(X)E(Y)$$

**Corollary 10.4.1.** *If $X$ and $Y$ are independent, then $\operatorname{Cov}(X,Y) = 0$.*

The relationship between covariance and variance is given as follows.

**Theorem 10.5.** *Suppose $X$ and $Y$ are random variables with finite variances, and $a, b$ are constants. Then*

$$\operatorname{Var}(aX+bY) = a^2\operatorname{Var}(X) + b^2\operatorname{Var}(Y) + 2ab\operatorname{Cov}(X,Y)$$

Extending to higher dimensions, we have the corollary.

**Corollary 10.5.1.** *Suppose that $W_1, \cdots, W_n$ are random variables with finite variances. Then*

$$\operatorname{Var}\left(\sum_{i=1}^{n} a_i W_i\right) = \sum_{i=1}^{n} a_i^2 \operatorname{Var}(W_i) + 2\sum_{i<j} a_i a_j \operatorname{Cov}(W_i, W_j)$$

**Corollary 10.5.2.** *Suppose that $W_1, \cdots, W_n$ are independent random variables with finite variances. Then*

$$\operatorname{Var}\left(\sum_{k=1}^{n} W_k\right) = \sum_{k=1}^{n} \operatorname{Var}(W_k)$$

# 11 Order statistics

Let $Y$ be a continuous random variable for which $y_1, \cdots, y_n$ are the calues of a random sample of size $n$. Reorder the $y_k$ from smallest to largest:

$$y_1' < \cdots < y_n'$$

Define the random variable ${Y'}_k$ to have the value ${y'}_k$, where $1 \le k \le n$. Then $Y_k'$ is called the $i^{\text{th}}$ *order statistic*. Sometimes $Y_n'$ and $Y_1'$ are denoted $Y_{max}$ and $Y_{min}$ respectively.

In fact, the order statistics has some intriguing properties, which is useful in future studies.

**Theorem 11.1.** *Suppose that $Y_1, \cdots, Y_n$ is a random sample of continuous random variables, each having pdf $f_Y(y)$ and cdf $F_Y(y)$. Then*

1. *The pdf of the largest order statistics is*

$$f_{Y_{max}}(y) = n[F_Y(y)]^{n-1} f_Y(y)$$

2. *The pdf of the smallest order statistics is*

$$f_{Y_{min}}(y) = n[1 - F_Y(y)]^{n-1} f_Y(y)$$

## 11.1 A General formula for $f_{y_k'}(y)$

We get inspired from the binomial expansion and deduced the general formula.

**Theorem 11.2.** *Let $Y_1, \cdots, Y_n$ be a random sample of continuous random variables drawn from a distribution having pdf $f_Y(y)$ and cdf $F_Y(y)$. The pdf of the $k^{th}$ order statistics is given by*

$$f_{Y_k'}(y) = \frac{n!}{(k-1)!(n-k)!}[F_Y(y)]^{k-1}[1 - F_Y(y)]^{n-k} f_Y(y)$$

*where $1 \le k \le n$.*

We are inspired to get the corollary about joint pdf of order statistics.

**Corollary 11.2.1.** *Let $Y_1, \cdots, Y_n$ be a random sample of continuous random variables drawn from a distribution having pdf $f_Y(y)$ and cdf $F_Y(y)$. The joint pdf for order statistics $Y'_k$ and $Y'_l$ at points $u$ and $v$, where $k < l$ and $u < v$, is as follows.*

$$f_{Y'_k, Y'_l}(u, v) = \frac{n!}{(k-1)!(l-k-1)!(n-l)!}[F_Y(u)]^{k-1}[F_Y(v) - F_Y(u)]^{l-k-1} \\ [1 - F_Y(v)]^{n-l} f_Y(u) f_Y(v)$$

# 12 Conditional densities

The notion of **condition density** is directly derived from the notion of *condition probability.*

**Theorem 12.1.** *Let $X$ and $Y$ be discrete random variables. The conditional probability density function of $Y$ with given $x$ is denoted by*

$$p_{Y|x}(y) = \frac{p_{X,Y}(x, y)}{p_X(x)}$$

*for $p_X(x) \neq 0$.*

**Theorem 12.2.** *Let $X$ and $Y$ be continuous random variables. The conditional probability density function of $Y$ with given $x$ is denoted by*

$$f_{Y|x}(y) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

*Moreover,*

$$P(Y \leq y | X = x) = \int_{-\infty}^{y} \frac{f_{X,Y}(x, u)}{f_X(x)} \, du$$

# 13 Moment-generating functions

**Moment-generating function** is a more complicated notion campared to the previous theorems. The definition of moment-generating function seems to be isolated, but it has strong connection with expectation and variance.

**Theorem 13.1.** *Let $W$ be a random variable. The moment-generating function (mgf) for $W$ is denoted by $M_W(t)$ and given by*

$$M_W(t) = E(e^{Wt}) = \begin{cases} \sum_{k \in U_k} e^{tk} p_W(k), & \text{if } W \text{ is discrete} \\ \int_{-\infty}^{\infty} e^{tk} f_W(k) \, dk, & \text{if } W \text{ is continuous} \end{cases}$$

*where for all values of $t$, tbhe expected value exists.*

## 13.1 Link between mgf and moment

**Theorem 13.2.** *Let $W$ be a random variable with pdf $f_W(w)$. Let $M_W(t)$ be the mgf for $W$. Then, we deduce the formula*

$$M_W^{(r)}(0) = E(W^r)$$

*(If $W$ is continuous, it must suffice the smoothness of order $r$).*

## 13.2 Link between mgf and variance

Since we have known that $\mathrm{Var}(W) = E(W^2) - [E(W)]^2$. We can replace the expected value with mgf.

## 13.3 Link between mdf and pdf

**Theorem 13.3.** *Suppose that $W_1$ and $W_2$ are random variables for which $M_{W_1}(t) = M_{W_2}(t)$ for some interval of $t$ containing $0$. Then*

$$f_{W_1}(w) = f_{W_2}(w)$$

The next theorem introduces the transformation of mgf.

**Theorem 13.4.**

1. *Let $W$ be a random variable with mgf $M_W(t)$. Let $V = aW + b$. Then*
$$M_V(t) = e^{bt} M_W(at)$$

2. *Let $W_1, \cdots, W_n$ be independent random variables with mgf $M_{W_1}(t), \cdots, M_{W_n}(t)$, respectively. Let $W = \sum_{k=1}^{n} W_k$. Then*

$$M_W(t) = \prod_{k=1}^{n} M_{W_k}(t)$$