Contents

3 The Random Variables **46**

3.1 Introduction: Function . . . 46
3.2 Random variables . . . 48
3.3 Probability distribution function . . . 52
3.4 Classification of random variables . . . 61
3.5 Probability density function . . . 63
3.5.1 Discrete random variables . . . 64
3.5.2 Continuous random variables . . . 70
3.5.3 Mixed random variables . . . 77
3.6 Conditional distribution & density functions . . . 78
3.6.1 Conditional distribution function . . . 78
3.6.2 Conditional density function . . . 80

Chapter 3

The Random Variables

3.1 Introduction: Function

Function:

Definition 3.1 Let D and R be any two sets. Then, a *relation* f from D to R is called a *function* if:

$$\forall x \in D, \ \ \exists \text{a } \textbf{unique } y \ \ni: \ \ f(x) = y$$

Fact:

A function has a *domain*(D) and a *range*(R).

Figure 3.1: The domain and range of a funcrion $f(\cdot)$.

$$\Rightarrow \ \ \forall x \in D, \ \ f \text{ maps it into a point in } R$$

Note:

(i) f IS a function:

Figure 3.2: A function.

(ii) f is NOT a function:

Figure 3.3: A relation which cannot be a function.

(cf.) In the case of (ii), f violates the **uniqueness** condition!

3.2 Random variables

Definition 3.2 General definition:

A random variable is a function which maps a point in the sample space (S) into a real number.

(e.g.)

Figure 3.4: A random variable as a function mapping S into R^1-line: general.

Why random variable?

If we could use the well known, and well experienced general algebra(or mathematics) using *numbers* in order to calculate the probabilities, (rather than dealing them in the probability space $(S, \mathcal{F}, P)$), it would give us much more easy and systematic way of dealing them: this is the necessaity of the concept of random variables.

So we can regard the random variable as a *transformation or function* which maps the outcomes or events into a real number or an interval in R^1-line.

Definition 3.3 Rigorous definition:

A function $f(\omega)$ defined over the sample space S into the R^1-line is a random variable if:

$$\forall I \subset R^1\text{-line}, \quad f^{-1}(I) \in \mathcal{F}$$

(e.g.)

Figure 3.5: A random variable as a function mapping S into R^1-line: rigorous.

Note:

(i) I represents an interval in R^1-line.

(ii) $f^{-1}(I) = \{\omega | f(\omega) \in I \subset R^1\}$: inverse image.

Remark:
Notice that: to be able to compute various probabilities on random variable, definition 3.3 is more adequate and rigorous definition!!!

FACT:
It can be shown (by using the Measure Theory) that the followinh statements are true:

1. If $f(\omega)$ is a r.v.(random variable), so is $|f(\omega)|$.
2. If $f(\omega)$ and $g(\omega)$ are r.v.'s, so are $f(\omega) + g(\omega)$, and $f(\omega) - g(\omega)$.
3. If $f(\omega)$ and $g(\omega)$ are r.v.'s, and $F(u, v)$ is a continuous function of u and v, then $F(f(\omega), g(\omega))$ is also a r.v..
4. If $f(\omega)$ is a r.v., so are:

   (i) $f^{+}(\omega) \triangleq \max(f(\omega), 0)$

   (ii) $f^{-}(\omega) \triangleq \min(f(\omega), 0)$

proof: In a more advanced course...

Example 3.1

Consider the chance experiment of tossing a fair coin, where the probability space $(S, \mathcal{F}, P)$ is composed by:

(i) $S = \{H, T\}$

(ii) $\mathcal{F} = \{\phi, S, \{H\}, \{T\}\}$

(iii) $P(\{H\}) = (\{T\}) = \frac{1}{2}$: fair coin

Figure 3.6: A r.v. $X(\omega)$ defined on coin tossing experiment.

Let's define a r.v. $X(\omega)$ such that:

$$\begin{cases} X(T) = 10 \\ X(T) = -10 \end{cases}$$

Here, consider $X^{1}(I)$ where the interval $I \subset R^1$-line:

(i) $I = (-\infty, -2]$: $X^{-1}(I) = \{\omega \mid X(\omega) \leq -2\} = \{T\}$

$$\therefore \quad P\{\omega \mid X(\omega) \leq -2\} \triangleq P(X \leq -2) = P(\{T\}) = \frac{1}{2}$$

(ii) $I = (-\infty, -15]$: $X^{-1}(I) = \{\omega \mid X(\omega) \leq -15\} = \phi$

$$\therefore \quad P\{\omega \mid X(\omega) \leq -15\} \triangleq P(X \leq -15) = P(\phi) = 0$$

(iii) $I = (-\infty, 15]$: $X^{-1}(I) = \{\omega \mid X(\omega) \leq 15\} = S$

$$\therefore \quad P\{\omega \mid X(\omega) \leq 15\} \triangleq P(X \leq 15) = P(S) = 1$$

(iv) $I = (0, 20]$: $X^{-1}(I) = \{\omega \mid 0 < X(\omega) \leq 20\} = \{H\}$

$$\therefore \quad P\{\omega \mid 0 < X(\omega) \leq 20\} \triangleq P(0 < X \leq 20) = P(\{H\}) = \frac{1}{2}$$

Figure 3.7: Inverse images of $X(\omega)$ for various intervals I.

NOTE:

(1) A r.v. $X(\cdot)$ is a *point function* whereas $P(\cdot)$ is a *set function.*

(2) $X^{-1}(I) \in \mathcal{F}$ for any interval $I \subset R^1$, by the definition of r.v.

3.3 Probability distribution function

The definition of the probability distribution function(PDF), or the cumulative distribution function(cdf) of a r.v. $X(\omega)$, where $\omega \in S$ is as follows:

Definition 3.4 The (probability) distribution function $F_X(x)$ of a r.v. $X(\omega)$ is defined as:

$$F_X(x) \triangleq P\{w \mid X(\omega) \leq x\}$$

Note:

(i) The event $\{X(\omega) \leq x\}$ is a subset of S such that $\{X(\omega) \leq x\} \in \mathcal{F}$ by the definition of the r.v..

(ii) x is a variable representing a *real value* in R^1-line.

(iii) Notice that the distribution function is defined in terms of **probability**.

Example 3.2

Consider the chance experiment of tossing a fair coin, where:

(i) $S = \{H, T\}$

(ii) $\mathcal{F} = \{\phi, S, \{H\}, \{T\}\}$

(iii) $P(H) = P(T) = 0.5$: fair coin

(iv) A random variable $X(\omega)$ is defined as in the previous example $\ni$:

$$\begin{cases} X(H) = 10 \\ X(T) = -10 \end{cases}$$

Then, determine the distribution function of the r.v. $X(\omega)$.

Solution:

By the definition of the distribution function,

$$F_X(x) = P\{w \mid X(\omega) \leq x\}$$

Figure 3.8: A r.v. $X(\omega)$ mapping from S to R^1-line.

(1) $x = -\infty$:

$$F_X(-\infty) = P\{w \mid X(\omega) \leq -\infty\} = P(\phi) = 0$$

$\vdots$

(2) $x = -10$:

$$F_X(-10) = P\{w \mid X(\omega) \leq -10\} = P(T) = \frac{1}{2}$$

$\vdots$

(3) $x = 10$:

$$F_X(10) = P\{w \mid X(\omega) \leq 10\} = P(S) = 1$$

$\vdots$

(4) $x = \infty$:

$$F_X(\infty) = P\{w \mid X(\omega) \leq \infty\} = P(S) = 1$$

Figure 3.9: The cdf $F_X(x)$ of $X(\omega)$.

(cf.) Note that $F_X(x)$ is *right-hand continuous*. What if the distribution function was defined as:

$$F_X(x) = P\{w \mid X(\omega) < x\} \text{ ?}$$

Properties of the distribution function:

: Every dist'n function must satisfy the following properties!

Let $F(x) \triangleq F_X(x)$ for notational convenience, then:

1. $F(-\infty) = 0 \quad \text{and} \quad F(\infty) = 1$.
2. $F(x_2) \geq F(x_1)$ if $x_2 \geq x_1$: monotone non-decreasing
3. $P(x_1 < X(\omega) \leq x_2)$[1] $= F(x_2) - F(x_1)$.
4. $\lim_{\epsilon \to 0, \epsilon > 0} F(x + \epsilon) = F(x)$: right-hand continuous

Proof:

1. Since $F(x) = P\{w \mid X(\omega) \leq x\}$, it is clear that:

$$\begin{cases} F(\infty) = P\{w \mid X(\omega) \leq \infty\} = P(S) = 1 \\ F(-\infty) = P\{w \mid X(\omega) \leq -\infty\} = P(\phi) = 0 \end{cases}$$

2. We have:

$$F(x_2) = P\{w \mid X(\omega) \leq x_2\}$$
$$F(x_1) = P\{w \mid X(\omega) \leq x_1\}$$

and we can decompose the event $\{w \mid X(\omega) \leq x_2\}$ into a union of two **disjoint** events $\ni$:

$$\{w \mid X(\omega) \leq x_2\} = \{w \mid X(\omega) \leq x_1\} \cup \{w \mid x_1 < X(\omega) \leq x_2\}$$

Therefore, from the axiom #3 of probability, we have:

$$P\{w \mid X(\omega) \leq x_2\} = P\{w \mid X(\omega) \leq x_1\} + P\{w \mid x_1 < X(\omega) \leq x_2\} \tag{3.1}$$

Since $P\{w \mid x_1 < X(\omega) \leq x_2\} \geq 0$ from the axiom #1 of probability:

$$P\{w \mid X(\omega) \leq x_2\} \geq P\{w \mid X(\omega) \leq x_1\}$$

$$\Rightarrow \quad F(x_2) \geq F(x_1)$$

[1]Rigorously speaking, it should be expressed as $P(\{\omega \mid x_1 < X(\omega) \leq x_2\})$.

Note:

The inverse images of two *disjoint intervals* in R^1-line are **mutually exclusive** due to the fact that random variables are FUNCTIONS!!!

$$A \triangleq \{w \mid X(\omega) \leq x_1\}$$

$$B \triangleq \{w \mid x_1 < X(\omega) \leq x_2\}$$

Figure 3.10: The inverse images of two disjoint intervals in R^1-line.

3. From (3.1), we have:

$$\begin{aligned} P\{w \mid x_1 < X(\omega) \leq x_2\} &= P\{w \mid X(\omega) \leq x_2\} - P\{w \mid X(\omega) \leq x_1\} \\ &\triangleq F(x_2) - F(x_1) \end{aligned}$$

4. To prove this property, we have to use the following axiom on probability known as **Continuity axiom**:

Continuity Axiom:

If $A_1, A_2, \ldots, A_n, \ldots$ are monotone increasing (i.e. $A_i \subset A_j \; \forall \; i < j$), or monotone decreasing (i.e. $A_i \supset A_j \; \forall \; i < j$) sequence of subsets $\in \mathcal{F}$, then the probability function $P(\cdot)$ must satisfy the following:

$$\lim_{n\to\infty} P(A_n) = P\left(\lim_{n\to\infty} A_n\right)$$

proof: To be covered later...

Now, consider the following sequence of subsets A_n of S:

$$A_n = \{\omega \mid x < X(\omega) \leq x + \epsilon_n\}$$

where $\epsilon_n > 0$ and $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$.

Figure 3.11: Monotone decreasing sequence of subset A_n.

Then, since $A_i \supset A_j \; \forall \, i < j$, $\{A_n\}$ is monotone decreasing sequence of subsets.
$\therefore \; \longrightarrow$ Continuity axiom applies!!!

$$\text{i.e. } P\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} P\left(A_n\right)$$

Note:
Notice that as $n \rightarrow \infty$, $A_n \rightarrow \phi$.
This is because as $\epsilon \rightarrow 0$, A_n approaches to x, but x does not belong to A_n. (see above figure.)

Consider now the subset $\{\omega \mid X(\omega) \leq x + \epsilon_n\}$, which can be expressed as a union of two **disjoint** subsets, i.e.

$$\{\omega \mid X(\omega) \leq x + \epsilon_n\} = \{\omega \mid X(\omega) \leq x\} \cup \underbrace{\{\omega \mid x < X(\omega) \leq x + \epsilon_n\}}_{A_n}$$

Figure 3.12: Union of two disjoint subsets.

Therefore,

$$\begin{aligned}\lim_{\epsilon_n \to 0} P\{\omega \mid X(\omega) \le x + \epsilon_n\} &= \lim_{\epsilon_n \to 0} \underbrace{P\{\omega \mid X(\omega) \le x\}}_{\text{independent of } \epsilon_n} \\ &\quad + \lim_{\epsilon_n \to 0} P\{\omega \mid x < X(\omega) \le x + \epsilon_n\}\end{aligned}$$

$$\begin{aligned}\Rightarrow \quad \lim_{\epsilon_n \to 0} F(x + \epsilon_n) &= F(x) + \lim_{\epsilon_n \to 0} P(A_n) \\ &= F(x) + P\left(\lim_{\epsilon_n \to 0} A_n\right) \quad \text{by Continuity axiom} \\ &= F(x) + P(\phi) \\ &= F(x)\end{aligned}$$

$\Rightarrow \quad F(x)$ is right-hand continuous!!!

Note: If we define the distribution function as:

$$F_X(x) \triangleq P(\{w \mid X(\omega) \le x\})$$

then, $F_X(x)$ would be left-hand continuous!

$$B_n \triangleq \{\omega | x - \epsilon_n \le X(\omega) < x\} : \text{monotone decreasing}$$

Figure 3.13: Union of two disjoint subsets.

proof: assignment

CONTINUITY AXIOM:

1. If $\{A_n\}_{n=1}^{\infty}$ is a monotone increasing sequence of subsets(or events) (i.e. $A_i \subset A_j \; \forall \, i < j$), with $A_n \in \mathcal{F} \;\; \forall \, n$, then

$$P\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} P\left(A_n\right)$$

2. If $\{B_n\}_{n=1}^{\infty}$ is a monotone decreasing sequence of subsets(or events) (i.e. $B_i \supset B_j \; \forall \, i < j$), with $B_n \in \mathcal{F} \;\; \forall \, n$, then

$$P\left(\lim_{n\to\infty} B_n\right) = \lim_{n\to\infty} P\left(B_n\right)$$

Proof:

1. Suppose $A_1 \subset A_2 \subset A_3 \ldots\ldots\ldots$ (monotone increasing), and there $\exists$ a limit $A_n \nearrow A$ where $A = \lim_{n\to\infty} A_n$.

Figure 3.14: Monotone increasing subsets $\{A_n\}_{n=1}^{\infty}$.

Now, let

$$E_k \triangleq A_k - A_{k-1} \quad k = 1, 2, 3, \ldots \quad (\because \text{ donut or ring shape})$$

where $A_0 = \phi$ and $\{E_k\}_{k=1}^{\infty}$ are *disjoint* to each other.

Then,

$$A_n = \bigcup_{k=1}^{n} E_k \quad : \text{disjoint unions}$$

and

$$A = \lim_{n\to\infty} A_n = \bigcup_{k=1}^{\infty} E_k$$

Notice that any union $\cup A_k$ can be replaced by disjoint unions $\cup E_k$. Therefore, we have:

$$E \triangleq \bigcup_{k=1}^{\infty} E_k \equiv A = \lim_{n\to\infty} A_n$$

$$\begin{aligned}
\Rightarrow\ P(A) = P(E) &= P\left(\bigcup_{k=1}^{\infty} E_k\right) \\
&= \sum_{k=1}^{\infty} P(E_k) \\
&= \sum_{k=1}^{\infty} P(A_k - A_{k-1}) \\
&= \lim_{n\to\infty} \sum_{k=1}^{n} P(A_k - A_{k-1}) \\
&= \lim_{n\to\infty} \sum_{k=1}^{n} \{P(A_k) - P(A_{k-1})\} \\
&= \lim_{n\to\infty} \{P(A_n) - P(A_0)\} \\
&= \lim_{n\to\infty} \{P(A_n) - P(\phi)\} \\
&= \lim_{n\to\infty} P(A_n)
\end{aligned}$$

Therefore,

$$P(A) = P\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} P(A_n) \tag{3.2}$$

q.e.d.

Fact:

$$P(A_k - A_{k-1}) = P(A_k) - P(A_{k-1})$$

pf:

$$\begin{aligned}
& A_k = A_{k-1} \cup \overbrace{(A_k - A_{k-1})}^{E_k} \quad \text{:disjoint union} \\
\rightarrow\ & P(A_k) = P(A_{k-1}) + P(A_k - A_{k-1}) \\
\rightarrow\ & P(A_k - A_{k-1}) = P(A_k) - P(A_{k-1})
\end{aligned}$$

2. Suppose $B_1 \supset B_2 \supset B_3 \ldots\ldots\ldots$ (monotone decreasing), and there $\exists$ a limit $B_n \searrow B$ where $B = \lim_{n\to\infty} B_n = \bigcap_{n=1}^{\infty} B_n$.

Let

$$C_n \triangleq B_n^c \qquad \forall\, n = 1, 2, 3, \ldots$$

then $\{C_n\}_{n=1}^{\infty}$ is monotone increasing sequence with $B_n^c = C_n \in \mathcal{F}$, and

$$\lim_{n\to\infty} C_n \triangleq C = \bigcup_{n=1}^{\infty} C_n$$

By (3.2), we have:

$$\begin{aligned}
& \lim_{n\to\infty} P\left(C_n\right) = P\left(\lim_{n\to\infty} C_n\right) \\
\Rightarrow\ & \lim_{n\to\infty} P\left(B_n^c\right) = P\left(\lim_{n\to\infty} B_n^c\right) = P\left(\left\{\lim_{n\to\infty} B_n\right\}^c\right) = P\left(B^c\right) = 1 - P(B) \\
\Rightarrow\ & \lim_{n\to\infty} \left\{1 - P\left(B_n\right)\right\} = 1 - P(B) \\
\Rightarrow\ & 1 - \lim_{n\to\infty} P\left(B_n\right) = 1 - P(B)
\end{aligned}$$

Therefore,

$$P(B) = P\left(\lim_{n\to\infty} B_n\right) = \lim_{n\to\infty} P\left(B_n\right)$$

q.e.d.

3.4 Classification of random variables

:In terms of the distribution function

1. **Continuous random variables:**

   If $F(x)$ of a r.v. $X(\omega)$ is continuous on x and differentiable w.r.t. x everywhere except at a countable number of points, then $X(\omega)$ is called a *continuous* random variable.

   (e.g.)

   Figure 3.15: An example of $F(x)$ for a continuous random variable.

2. **Discrete random variables:**

   If $F(x)$ of a r.v. $X(\omega)$ is a staircase type, then $X(\omega)$ is called a *continuous* random variable.

   (e.g.)

   Figure 3.16: An example of $F(x)$ for a discrete random variable.

3. **Mixed random variables:**

   If $F(x)$ of a r.v. $X(\omega)$ is a combination of above two types, then $X(\omega)$ is called a *mixed* random variable.

   (e.g.)

Figure 3.17: An example of $F(x)$ for a mixed random variable.

3.5 Probability density function

The definition of the probability density function(pdf) of a r.v. $X(\omega)$, where $\omega \in S$ is as follows:

Definition 3.5 The probability density function (pdf) of a random variable $X(\omega)$ is defined as:

$$f_X(x) \triangleq \frac{dF_X(x)}{dx}$$

Note:

(i) For notational convenience, we sometimes denote $f_X(x)$ as $f(x)$ as long as it does not cause any confusion.

(ii) From the above definition of p.d.f., notice that p.d.f. and PDF of a r.v. $X(\omega)$ are related by *defferentiation/integration*, i.e. the PDF $F_X(x)$ in terms of $f_X(x)$ is expressed as:

$$F_X(x) = \int_{-\infty}^{x} f_X(\alpha) d\alpha$$

Properties of $f(x)$:

(1) The p.d.f. is non-negative:

$$f(x) \geq 0$$

(2) The integration of p.d.f. over entire R^1-line is unity:

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

(3) The probablity of an event $\{\omega \mid x_1 < X(\omega) \leq x_2\}$ can be evaluated using p.d.f. of $X(\omega)$ as:

$$P\{\omega \mid x_1 < X(\omega) \leq x_2\} = \int_{x_1}^{x_2} f(x)dx$$

Proof:

(1) Since $F(x)$ is non-decreasing function of x, the slope $\left(= \frac{dF}{dx}\right)$ at every point of x must be non-negative, i.e.

$$\frac{dF_X(x)}{dx} \triangleq f_X(x) \geq 0$$

(2) From the relation of the p.d.f. and the PDF, it is clear that:

$$\int_{-\infty}^{\infty} f_X(x)dx \triangleq F_X(\infty) = 1$$

(3) From the probability of the given event in terms of the PDF, we have:

$$\begin{aligned} P\{\omega \mid x_1 < X(\omega) \leq x_2\} &= F_X(x_2) - F_X(x_1) \\ &= \int_{-\infty}^{x_2} f_X(x)dx - \int_{-\infty}^{x_1} f_X(x)dx \\ &= \int_{x_1}^{x_2} f_X(x)dx \end{aligned}$$

q.e.d.

3.5.1 Discrete random variables

The pdf and the PDF as well for a discrete random variables can be represented in a fixed formula.
Let's first take a look at a specific example, and extend the concept into a general form:

Example 3.3

Consider the following discrete r.v. $X(\omega)$:

$$X(\omega_1) = x_1 = -1, \quad X(\omega_2) = x_2 = 2, \quad X(\omega_3) = x_3 = 3$$
$$X(\omega_4) = x_4 = 4, \quad X(\omega_5) = x_5 = 6,$$

where

$$S = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \}$$

$$P(\{\omega_i\}) = p_i \;\; i = 1, 2, 3, 4, 5 \quad \text{and} \quad \sum_{i=1}^{5} p_i = 1$$

Figure 3.18: The sample space S and r.v. $X(\omega)$.

Then the distribution function $F(x)$ can be shown in the following form:

Figure 3.19: The PDF $F(x)$.

(cf.) An example of calculating the probability of an event described in $X(\omega)$:

$$P\{\omega \mid -3 < X(\omega) \leq 6\} = F_X(5) - F_X(-3) = p_1 + p_2 + p_3 + p_4$$

Notice that the above $F(x)$ can be expressed in a fixed mathematical form as:

$$F(x) = \sum_{i=1}^{5} p_i u(x - x_i)$$

where $x_1 = -1, x_2 = 2, x_3 = 3, x_4 = 4, x_5 = 6$.

Here, $u(x - x_i)$ is a shifted unit step function defined as:

$$u(x - x_i) \triangleq \begin{cases} 1, & x \geq x_i \\ 0, & x < x_i \end{cases}$$

Figure 3.20: The shifted unit step function $u(x - x_i)$.

Then, the derivative of the shifted unit step function $u(x - x_1)$ is zero everywhere except at $x = x_i$ at which it has a value of infinity (i.e. ∞-slope).

We call this type of function a *Dirac delta* function, and denote it as:

$$\delta(x - x_i) \triangleq \frac{d}{dx} \{u(x - x_i)\}$$

Therefore, the pdf $f(x)$ of $X(\omega)$ in the above example can be expressed in a fixed mathematical form given below:

$$\begin{aligned} f(x) = \frac{d}{dx} F(x) &= \sum_{i=1}^{5} p_i \frac{d}{dx} \{u(x - x_i)\} \\ &= \sum_{i=1}^{5} p_i \delta(x - x_i) \end{aligned}$$

(cf.) Dirac delta function (Unit step function)

Definition 3.6 The Dirac delta function is usually defined by the following two conditions:

$$\delta(x) \triangleq \frac{d}{dx}\{u(x)\} = \begin{cases} \infty, & x = 0 \\ 0, & x \neq 0 \end{cases}$$

$$\text{and} \qquad \int_{-\infty}^{\infty} \delta(x)dx = 1$$

Graphical interpretation:

Define a unit [2] pulse $u_\epsilon(x)$ as follows:

$$u_\epsilon(x) = \begin{cases} \frac{1}{\epsilon}, & 0 \leq x \leq \epsilon \\ 0, & \text{elsewhere} \end{cases}$$

Figure 3.21: $u_\epsilon(x) \longrightarrow \delta(x)$.

Then, we can see that:

$$\delta(x) = \lim_{\epsilon \longrightarrow 0} u_\epsilon(x)$$

Note:
Notice that the area of $\delta(x)$ is maintained to be unity (i.e. 1):

$$\int_{-\infty}^{\infty} \delta(x)dx = 1$$

or

$$\int_{-\infty}^{\alpha} \delta(x)dx = \begin{cases} 1, & \alpha \geq 0 \\ 0, & \alpha < 0 \end{cases}$$

[2]This mean that the area is 1.

Special property of $\delta(x)$:

A special and useful property of the Dirac delta function is as follows, which is called the "sifting property" of $\delta(x)$.

$$\int_{-\infty}^{\alpha} g(x)\delta(x-a)dx = \int_{-\infty}^{\alpha} g(a)\delta(x-a)dx = \begin{cases} g(a), & \alpha \geq a \\ 0, & \alpha < a \end{cases}$$

Figure 3.22: Sifting property of $\delta(x)$.

Now we go back to the discussion of the pdf of a discrete r.v.'s.
The pdf $f(x)$ of the example 3.3 can then be graphically represented as follows;

Figure 3.23: The pdf $f(x)$ of $X(\omega)$.

Note:

(1) As an example of calculating the probability of an event described in $X(\omega)$:

$$P\{\omega \mid -3 < X(\omega) \leq 6\} = \int_{-3}^{5} f_X(x)dx = p_1 + p_2 + p_3 + p_4$$

(2) Notice that the area under $f(x)$ is unity, which should always be true:

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)dx &= \int_{-\infty}^{\infty} \sum_{i=1}^{5} p_i\delta(x - x_i)_d x \\ &= \sum_{i=1}^{5} p_i \int_{-\infty}^{\infty} \delta(x - x_i)_d x \\ &= \sum_{i=1}^{5} p_i \\ &= 1\end{aligned}$$

If the sample space S has N elements (or outcomes), we can generalize the above discussion into the followng forms of PDF and pdf:

$$F(x) = \sum_{i=1}^{N} p_i u(x - x_i) : \text{ weighted \& delayed sum of } u(x)$$

$$f(x) = \sum_{i=1}^{N} p_i \delta(x - x_i) : \text{ weighted \& delayed sum of } \delta(x)$$

EXAMPLES of discrete random variables:

(1) Binomial random variable

(2) Poisson random variable

⋮

: Self-study (READ)

3.5.2 Continuous random variables

In the case of continuous random variables, there ∃ numerous different cases, and the PDF's and pdf's cannot be generalized in a fixed mathematical forms as in the discrete random variables.

So, we consider some typical cases which are frequently encountered...

(1) Uniform random variable:

Definition 3.7 A random variable $X(\omega)$ is called a *uniform* random variable if it has the following form of p.d.f.:

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & \text{elsewhere} \end{cases}$$

Figure 3.24: A typical pdf $f(x)$ of a uniform r.v. $X(\omega)$.

Then, the distribution function $F_X(x)$ is:

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f_X(\alpha)d\alpha \\ &= \begin{cases} 0, & x < a \\ \int_a^x f_X(\alpha)d\alpha = \frac{x-a}{b-a}, & a \le x < b \\ \int_a^b f_X(\alpha)d\alpha = \frac{b-a}{b-a} = 1, & x \ge b \end{cases} \end{aligned}$$

Figure 3.25: A typical PDF $F(x)$ of a uniform r.v. $X(\omega)$.

Note:
We use a r.v. of uniform distribution with $a = -1$, and $b = 1$ in many computer simulations, and they usually are provided as subroutines or internal functions of common computer languages such as MATLAB, Fortran, C etc..

: Random number generator

Figure 3.26: A pdf $f(x)$ of uniform random number generator.

(2) Gaussian random variable:

Definition 3.8 A random variable $X(\omega)$ is called a *Gaussian* random variable if it has the following form of p.d.f.:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\}$$

where m and σ^2 are called the *mean* and *variance* of $X(\omega)$ respectively. σ is known as the *standard deviation.*

Figure 3.27: A typical pdf $f(x)$ of a Gaussian r.v. $X(\omega)$.

(cf.) Many of statistical data are known to have Gaussian distribution, e.g. graded points of certain examination, amplitude of certain noises etc...

Before we discuss the PDF of a Gaussian randoma variable, we briefly pause to mention the so called "error function".

Definition 3.9 The error function in an integral form is defined as follows:

$$\mathrm{erf}(x) \triangleq \int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{\alpha^2}{2}} d\alpha$$

Figure 3.28: The error function erf(x).

Note:

(i) Notice that the error function is an odd (or anti-symmetric) function of x, i.e.:

$$\mathrm{erf}(-x) = -\mathrm{erf}(x)$$

(ii) In place of the error function, we sometimes use the "Q-function" defined in a similar fashion as:

$$\begin{aligned} \mathrm{Q}(x) &\triangleq \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{\alpha^2}{2}} d\alpha \\ &= \frac{1}{2} - \mathrm{erf}(x) \end{aligned}$$

Now, the probability distribution function(PDF) $F(x)$ of a gaussian r.v. is:

(i) Suppose $x \geq m$:

$$\begin{aligned} F(x) = \int_{-\infty}^{x} f(\alpha)d\alpha &= \int_{-\infty}^{m} f(\alpha)d\alpha + \int_{m}^{x} f(\alpha)d\alpha \\ &= \frac{1}{2} + \int_{m}^{x} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\alpha-m)^2}{2\sigma^2}\right\} d\alpha \\ &\quad \left(\text{let } \tfrac{\alpha-m}{\sigma} = \beta, \text{ then } \tfrac{1}{\sigma}d\alpha = d\beta\right) \\ &= \frac{1}{2} + \int_{0}^{\frac{x-m}{\sigma}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{\beta^2}{2}\right\} d\beta \\ &= \frac{1}{2} + \mathrm{erf}\left(\frac{x-m}{\sigma}\right) \\ &\quad : \text{scaled and shifted error function with bias} \end{aligned}$$

(ii) Suppose $x < m$:

Figure 3.29: The pdf of Gaussian r.v. when $x < m$.

$$\begin{aligned} F(x) &= 1 - F(2m - x) \\ &= 1 - \left\{\frac{1}{2} + \mathrm{erf}\left(\frac{2m-x-m}{\sigma}\right)\right\} \\ &= \frac{1}{2} - \mathrm{erf}\left(\frac{m-x}{\sigma}\right) \\ &= \frac{1}{2} + \mathrm{erf}\left(\frac{x-m}{\sigma}\right) \end{aligned}$$

Therefore, regardless of the magnitude of x (i.e. for all $-\infty < x < \infty$, the PDF of a Gaussian r.v. is in the following form:

$$\frac{1}{2} + \mathrm{erf}\left(\frac{x-m}{\sigma}\right)$$

Figure 3.30: The PDF of Gaussian r.v. $X(\omega)$.

(cf.) Note that $P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$.

Remark:
Appendix B of the textbook has the table of $F(x)$ values for the case of $m = 0$ and $\sigma = 1$, i.e.:

$$F_0(x) = \frac{1}{2} + \text{erf}(x), \qquad \text{for } x \geq 0$$

Figure 3.31: The pdf $f_0(x)$ for the case of $m = 0$ and $\sigma = 1$.

Question: How do we use the table if:

(i) $x < 0$

(ii) $m \neq 0$ and/or $\sigma \neq 1$.

Answer:

(i) Let $x > 0$, then:

$$
\begin{aligned}
F_0(-x) = \int_{-\infty}^{-x} f_0(\alpha) d\alpha &= \int_{-\infty}^{-x} \frac{1}{\sqrt{2\pi}} e^{-\frac{\alpha^2}{2}} d\alpha \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\alpha^2}{2}} d\alpha - \int_{-x}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\alpha^2}{2}} d\alpha \\
&\quad \text{(let } \beta = -\alpha) \\
&= 1 - \int_{x}^{-\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\beta^2}{2}} (-d\beta) \\
&= 1 - \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{\beta^2}{2}} d\beta \\
&= 1 - F_0(x)
\end{aligned}
$$

Figure 3.32: $F_0(-\alpha)$ where $\alpha > 0$ in terms of the pdf $f_0(x)$.

(ii) For the case when $m \neq 0$, $\sigma \neq 1$:

$$
\begin{aligned}
F(x) &= \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\alpha - m)^2}{2\sigma^2}} d\alpha \\
&\quad \text{(let } \tfrac{\alpha - m}{\sigma} = \beta \text{, then } \tfrac{1}{\sigma} d\alpha = d\beta \text{)} \\
&= \int_{-\infty}^{\frac{x-m}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{\beta^2}{2}} d\beta \\
&\triangleq F_0\left(\frac{x - m}{\sigma}\right) \\
&\quad : \text{scaled and shifted version of } F_0(x)
\end{aligned}
$$

Interpretation of σ:

Figure 3.33: The Gaussian pdf $f(x)$.

Note that:

$$\begin{aligned} F_X(m-\sigma) = F_0(-1) &= 1 - F_0(1) \\ &= 1 - 0.8413 \\ &\approx 0.1587 \end{aligned}$$

Therefore,

$$\begin{aligned} P\left(m-\sigma \leq X \leq m+\sigma\right) &= 1 - 2P(X \leq m-\sigma) \\ &= 1 - 2F_X(m-\sigma) \\ &= 1 - 2 \times 0.1587 \\ &= 0.6826 \\ &= 70\% \end{aligned}$$

OR

$$\begin{aligned} P\left(m-\sigma \leq X \leq m+\sigma\right) &= P(X \leq m+\sigma) - P(X \leq m-\sigma) \\ &= F_X(m+\sigma) - F_X(m-\sigma) \\ &= F_0(1) - F_0(-1) \\ &= 2F_0(1) - 1 \\ &= 2 \times 0.8413 - 1 \\ &= 0.6826 \end{aligned}$$

The above result indicates that the probability of a Gaussian r.v. $X(\omega)$ to have its value within the interval $[m-\sigma, \leq m+\sigma]$ (i.e deviating in amount of the standard deviation σ from its mean m) is about 70%.

(3) Exponential random variable:

: *Self study*

(4) Rayleigh random variable:

: *Self study*

3.5.3 Mixed random variables

The distribution function $F(x)$ of a mixed r.v. will be in the following form:

Figure 3.34: The PDF $F(x)$ of a mixed random variable.

Then, the probability density function $f(x)$ must be expressed as follows:

$$f(x) = \underbrace{\frac{dF(x)}{dx}}_{\text{continuous}} + \underbrace{\sum_{i=1}^{N} \Delta f(x_i)}_{\text{discrete}}$$

where

$$\Delta f(x_i) = \left\{ F(x_i) - F(x_i^-) \right\} \delta(x - x_i)$$

3.6 Conditional distribution & density functions

3.6.1 Conditional distribution function

Recall that given a probability space $(S, \mathcal{F}, P)$, the conditional probability of an event A given event B is:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \text{where } \; P(B) > 0$$

Let $A = \{\omega \mid X(\omega) \leq x\}$, then we have the following defintion of the conditional distribution function:

Definition 3.10 The conditional distribution function of a r.v. $X(\omega)$ based on a event B is defined and denoted as follows:

$$\begin{aligned} F_X(x|B) &\triangleq P\left[\overbrace{\{\omega \mid X(\omega) \leq x\}}^{A} \mid B\right] \\ &= \frac{P\left[\{\omega \mid X(\omega) \leq x\} \cap B\right]}{P(B)} \end{aligned}$$

Then, we could verify the following properties:

1. $F_X(-\infty|B) = 0 \quad \text{and} \quad F_X(\infty|B) = 1$.
2. $F_X(x_2|B) \geq F_X(x_1|N)$ if $x_2 \geq x_1$: monotone non-decreasing
3. $\lim_{\epsilon \to 0, \epsilon > 0} F_X(x + \epsilon|B) = F_X(x|B)$: right-hand continuous

Proof: Assignment

Example 3.4

Suppose we know the distribution function $F_X(x)$ of a r.v. $X(\omega)$, and let an event B be:

$$B = \{\omega \mid b < X(\omega) \leq a\}$$

$$B = X^{-1}(I_{ba})$$

Figure 3.35: Corresponding interval I_{ba} for the event B.

Determine the conditional distribution $F_X(x|B)$ in terms of $F_X(x)$.

Solution:

Let the events A as before. i.e.

$$A = \{\omega \mid X(\omega) \leq x\}$$

Then, we have :

(1) $P(B)$:

$$P(B) = F_X(a) - F_X(b)$$

(2) $P(A \cap B)$:

$$A \cap B = \begin{cases} \phi, & x \leq b \\ \{\omega \mid b < X(\omega) \leq x\}, & b < x < a \\ \{\omega \mid b < X(\omega) \leq a\}, & x \geq a \end{cases}$$

Therefore, we get:

$$P(A \cap B) = \begin{cases} 0, & x \leq b \\ F_X(x) - F_X(b), & b < x < a \\ F_X(a) - F_X(b), & x \geq a \end{cases}$$

From (1) and (2), we get the conditional distribution function as:

$$
\begin{aligned}
F_X(x|B) &= \frac{P(A \cap B)}{P(B)} \\
&= \begin{cases} 0, & x \leq b \\ \frac{F_X(x) - F_X(b)}{F_X(a) - F_X(b)}, & b < x < a \\ 1, & x \geq a \end{cases}
\end{aligned}
$$

(cf.) Notice that $F_X(x|B)$ is a scaled and biased version of $F_X(x)$!!!

Figure 3.36: Comparison b/w $F_X(x)$ and $F_X(x|B)$.

3.6.2 Conditional density function

Definition 3.11 The conditional probability density function of a r.v. $X(\omega)$ given an event B is defined and denoted as follows:

$$f_X(x|B) \triangleq \frac{d}{dx}\{F_X(x|B)\}$$

(cf.)
Note that the conditional pdf and PDF's are related as differentiation/integration to each other, i.e.:

$$\int_{-\infty}^{x} f_X(\alpha|B) d\alpha = F_X(x|B)$$

Properties of $f_X(x|B)$:

(1) The p.d.f. is non-negative:

$$f_X(x|B) \geq 0$$

(2) The integration of p.d.f. over entire R^1-line is unity:

$$\int_{-\infty}^{\infty} f_X(x|B)dx = 1$$

(3) The probablity of an event $\{\omega \mid a < X(\omega) \leq b\}$ given that an event B has occurred, can be evaluated using conditional p.d.f. of $X(\omega)$ as:

$$P\left[\{\omega \mid a < X(\omega) \leq b\} \mid B\right] = \int_a^b f_X(x|B)dx$$

Proof: Assignment (easy from the properties of $F_X(x|B)$.)