Random Variables: An Overview
Monson Hayes, Fellow, IEEE
Chung-Ang University
Seoul, Korea
mhh3@gatech.edu
(Invited Paper)

Abstract—This paper introduces the concept of a random variable, which is nothing more than a variable whose numeric value is determined by the outcome of an experiment. To describe the probabilities that are associated with these numeric values in a concise and conceptually useful manner, the probability distribution and probability density function are introduced. Then, the moment generating function is defined, and several examples are given. Finally, the concept of a correlation function and correlation matrices is introduced.

I. INTRODUCTION

The concept of a random variable is a simple one, and one that is important. Although perhaps sounding at first like something difficult, random variables are conceptually quite simple. Given a sample space $\Omega$ corresponding to some random experiment, this sample space contains elementary events, $\omega \in \Omega$, and when an experiment is performed, a specific elementary event (experimental outcome) is observed. When these experimental outcomes are numeric values, such as the temperature of a fluid, the velocity of a particle, or the value of a commodity or stock, we have a random variable. In some experiments, the outcomes may be non-numeric, such as the gender or the blood type of a person selected at random from within a given population, or the outcome of the flip of a coin. In these cases, a random variable may be formed simply by mapping these non-numeric outcomes to real numbers. We begin this paper by introducing an example of a random variable, and how a probability assignment may be made on the experimental outcomes.

II. PROBABILITY ASSIGNMENTS

Let $N$ be a variable that represents the number of $\alpha$ particles that are counted over a given period of time. The ensemble for $N$ is the set of non-negative integers

$$\mathcal{E}_N = \{0, 1, 2, \ldots\}$$

Since the number of outcomes is unknown until we actually make a count, then $N$ is a random variable. In many cases, it is appropriate to model $N$ as a Poisson random variable where

$$P\{N = n\} = \frac{\lambda^n}{n!} e^{-\lambda} \quad n \geq 0$$

(1)

for some $\lambda > 0$.

Given this probability assignment for $N$, it is then easy to find the probability of any event that is defined in terms of values of $N$. For example, the probability that the number of $\alpha$ particles is less than some number, $N_0$, may be found as follows. Since the event $\{N < N_0\}$ is the union of the events $\{N = k\}$ for $k = 0, 1, \ldots, N_0 - 1$,

$$\{N < N_0\} = \bigcup_{n=0}^{N_0-1} \{N = n\}$$

and since these events are mutually exclusive, then

$$P\{N < N_0\} = \sum_{n=0}^{N_0-1} P\{N = n\} = \sum_{n=0}^{N_0-1} \frac{\lambda^n}{n!} e^{-\lambda}$$

This last sum may be evaluated using the following

$$\sum_{n=0}^{k} \frac{\lambda^n}{n!} e^{-\lambda} = \frac{\Gamma(k + 1, \lambda)}{k!}$$

where

$$\Gamma(k, \lambda) = \int_{\lambda}^{\infty} x^{k-1} e^{-x} dx$$

III. PROBABILITY MASS FUNCTION

Discrete random variables have a discrete set of possible outcomes, $x_k$, and the probabilities of these outcomes are denoted by $p_k$.

$$P\{X = x_k\} = p_k$$

These probabilities are often referred to as probability masses since they represent discrete (point) masses of probability at specific locations along the real axis, just as point charges in electrostatics are used to represent discrete charges distributed along a line. These probabilities, written or plotted as a function of $x$, is called the probability mass function (PMF).

In order to express probability mass functions mathematically, we introduce the delta function,\(^2\) which is defined as follows:

$$\delta[n] = \begin{cases} 1 & ; n = 0 \\ 0 & ; n \neq 0 \end{cases}$$

Shifted delta functions may be used to represent functions that have a value of one at other values of $n$. For example, $\delta[n-1]$ is equal to one when $n = 1$ and equal to zero for all other values of $n$. Therefore, for an integer-valued discrete random variable $X$ with

$$P\{X = n\} = p_X[n] : -\infty < n < \infty$$

\(^2\)In digital signal processing, $\delta[n]$ is referred to as the unit sample function.

1Note that with this probability assignment it is assumed that the number of particles may be arbitrarily large and, in fact, approach infinity.
the PMF may be written as

$$p_X(n) = \sum_{k=-\infty}^{\infty} p_X[k] \delta[n-k]$$

For example, a Bernoulli Random Variable $X$ has an ensemble of possible outcomes of zero and one

$$\mathcal{E}_X = \{0, 1\}$$

where

$$P\{X = 0\} = p; \quad P\{X = 1\} = 1 - p$$

Therefore, the probability mass function for $X$ may be expressed in terms of delta functions as follows:

$$p_X(n) = p\delta[n] + (1-p)\delta[n-1]$$

Another example is the Geometric Random Variable that has an ensemble equal to the set of all positive integers

$$\mathcal{E}_X = \{1, 2, 3, \ldots\}$$

with a probability law given by

$$P\{N = k\} = \left(\frac{1}{2}\right)^k; \quad k > 0$$

The probability mass function for this random variable is

$$p_N(n) = \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^k \delta[n-k]$$

Another random variable that occurs frequently in applications is one that corresponds to the number of successes, $N$, in $n$ Bernoulli trials, with the probability of a success being equal to $p$. In this case, $N$ has a Binomial Distribution with

$$p_N(k) = P\{N = k\} = \binom{n}{k} p^k (1-p)^{n-k}; \quad 0 \leq k \leq n$$

where $\binom{n}{k}$ is the number of combinations of $n$ objects that are taken $k$ at a time, and is defined by

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

Alternative notations include $C(n,k)$, $\binom{n}{k}$, $\binom{n}{k}$, and $C^n_k$.

Interesting problems that are sometimes challenging to solve, are those such as

$$P\{N \text{ is odd}\} = \sum_{\substack{n \leq \infty \\text{n odd}}} P\{N = n\}$$

IV. PROBABILITY DENSITY FUNCTION

For a continuous random variable, the counterpart of the probability mass function is the probability density function. The probability density function, $f_X(x)$, of a continuous random variable $X$ is the derivative of the probability distribution function,

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

where

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

Conversely, the distribution function is the integral of the density function,

$$F_X(x) = \int_{-\infty}^{x} f_X(\alpha)d\alpha \quad (2)$$

Two properties of the density function are:

1) $f_X(x) \geq 0$ for all $x$.
2) $\int_{-\infty}^{\infty} f_X(x)dx = 1$

The first property follows from the monotonicity of the distribution function. Specifically, since $F_X(x)$ is a non-decreasing function of $x$, then the derivative will always be non-negative for all $x$. The second property follows by letting $x \to \infty$, in Eq. (2),

$$\int_{-\infty}^{\infty} f_X(\alpha)d\alpha = F_X(\infty) = 1$$

It is important to understand that $f_X(x)$ is a probability density and not a probability. Probabilities are found by integrating $f_X(x)$. For example, the probability of the event $\{X \leq x\}$ is given by the integral in Eq. (2), and the probability of the event $\{x_1 \leq X \leq x_2\}$ is found by evaluating the integral

$$P\{x_1 \leq X \leq x_2\} = \int_{x_1}^{x_2} f_X(\alpha)d\alpha \quad (3)$$

Unlike the distribution function, which is constrained to have values between zero and one, there is no such constraint on the density function. In fact, $f_X(x)$ may be arbitrarily large as long as the integral of $f_X(x)$ over all $x$ is equal to one (the area under the curve is unity).

A table of common random variables is given in Table I.

V. MOMENT GENERATING FUNCTIONS

The complete statistical characterization of a random variable of the continuous type requires the specification of its probability distribution or density function. Another way to provide this statistical characterization is with the characteristic function, which is the expected value of $e^{j\omega X}$

$$M_X(j\omega) = E\{e^{j\omega X}\} = \int_{-\infty}^{\infty} e^{j\omega x} f_X(x)dx$$

where $\omega$ is a variable taking on any real number between $\pm \infty$ and $j = \sqrt{-1}$. Since $e^{j\omega} = \cos \omega + j \sin \omega$, the characteristic function is, in general, a complex-valued function of $\omega$. It

<table>
<thead>
<tr>
<th>Name</th>
<th>Density Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>$f_X(x) = \lambda e^{-\lambda x}$</td>
</tr>
<tr>
<td>Laplace</td>
<td>$f_X(x) = \frac{1}{2\alpha} e^{-\alpha</td>
</tr>
<tr>
<td>Rayleigh</td>
<td>$f_X(x) = \alpha^\alpha x^{-\alpha} e^{-\alpha x^2/2}$, $x \geq 0$</td>
</tr>
<tr>
<td>Uniform</td>
<td>$f_X(x) = 1/(b-a)$, $b \leq x \leq a$</td>
</tr>
</tbody>
</table>

A TABLE OF COMMON AND IMPORTANT RANDOM VARIABLES.
may be recognized that, except for the sign in the exponent, $M_X(e^{j\omega})$ is the Fourier transform of the probability density function. Just as Fourier transforms are useful in the analysis of signals and linear systems, the characteristic function is also a useful tool in probability and random variables. Since the characteristic function is the integral of the joint density function, it is necessary to specify the joint probability density function. For a pair of continuous random variables, $X$ and $Y$, the joint density function is the derivative of the joint distribution function,

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

where

$$F_{X,Y}(x,y) = P\{X \leq x, Y \leq y\}$$

provided the derivatives exist. Conversely, the joint distribution function is the integral of the joint density function,

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(\alpha, \beta)d\alpha d\beta$$

Given the joint density or distribution function, the probability of any event that is defined in terms of values of $X$ may be found. For example, $P\{X \leq x, Y \leq y\} = F_{X,Y}(x,y)$ is found by integrating the joint density function as in Eq. (5).

Two important properties of the joint density function are given below.

1) $f_{X,Y}(x,y) \geq 0$ for all $x, y$.

2) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1$

The first property follows from the monotonicity of the distribution function. Specifically, since $F_X(x)$ is a non-decreasing function of $x$, then the derivative will always be non-negative for all $x$. The second property follows from Eq. (5) by letting $x \to \infty$ and $y \to \infty$,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1$$

As was the case for a single random variable, it is important to understand that the joint density function $f_{X,Y}(x,y)$ is a probability density and not a probability. Probabilities are found by integrating the density function. For example, the probability of the event $A = \{x_1 \leq X \leq x_2\} \cup \{y_1 \leq X \leq y_2\}$ is found by integrating the joint density function as follows

$$P\{x_1 \leq X \leq x_2, y_1 \leq X \leq y_2\} = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f_{X,Y}(x,y)dxdy$$

and to find the probability that $(x,y) \in R$ where $R$ is an arbitrarily defined region in the $x$-$y$ plane, we integrate the
density function over \( R \),

\[
P\{(X,Y) \in R\} = \int_R f_{XY}(x,y)dx\,dy \quad (7)
\]

Unlike the joint distribution function, which is constrained to have values between zero and one, there is no such constraint on the joint density function. In fact, \( f_{XY}(x,y) \) may be arbitrarily large as long as the integral of \( f_{XY}(x,y) \) over all \( x \) and \( y \) is equal to one.

VII. CORRELATION AND CORRELATION MATRICES

For two or more random variables, an important ensemble average is the correlation. Given two random variables \( X \) and \( Y \), the correlation is defined by

\[
r_{X_1X_2} = E\{X_1X_2\}
\]

If we consider these two random variables to be a random vector

\[
X = [X_1, X_2]^T
\]

then the correlation matrix is the expected value of \( XX^T \),

\[
R_X = E\{XX^T\} = \begin{bmatrix} E\{X_1^2\} & E\{X_1X_2\} \\ E\{X_2X_1\} & E\{X_2^2\} \end{bmatrix}
\]

For \( n \) random variables, the correlation matrix has the form

\[
R_X = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix}
\]

where

\[
r_{kl} = E\{X_kX_l\}
\]

These correlations are often estimated from a set of realizations (experimental outcomes) of the random variables as follows Sample cross-correlation

\[
\hat{r}_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}
\]

where \( \hat{r}_{xy} \) is called a sample autocorrelation.

VIII. CHI-SQUARE RANDOM VARIABLE

We conclude this paper with one last example of a random variable, one that is found often in statistics. If \( Z_1, \ldots, Z_n \) are independent, unit variance Gaussian random variables with zero mean, then the sum of their squares,

\[
Y = \sum_{k=1}^{n} Z_k^2
\]

is a Chi-square random variable with \( n \) degrees of freedom. A plot of the probability density and cumulative distribution functions are shown in Fig. 2.

IX. CONCLUSION

There are many excellent textbooks where the reader may find advanced developments of the results presented in this paper. The classic work in the field is the text by Papoulis [1]. Another recommended text is [2]. An introduction to Monte Carlo simulations may be found in [3].

REFERENCES