Trivedi's notes, Fall 1980

Example 4.1 - Consider the problem of searching for a specific name in a table of names. A simple method is to scan the table sequentially starting from one end until we either find the name or reach the other end, indicating that the required name is missing from the table. The following is a PASCAL program segment for sequential search:

```
VAR T: ARRAY[0..n] OF NAME;
Z: NAME;
I: 0..n;
BEGIN
   T[0]:=Z; {T[0] is used as a sentinel or marker}
   I:=n;
   WHILE Z ≠ T[I] DO
      I:=I - 1;

   IF I > 0 THEN {found; I points to Z}
    ELSE {not found}.
END.
```

In order to analyze the time required for sequential search, let X be the discrete random variable denoting the number of comparisons "Z ≠ T[I]" made. Clearly, the set of all possible values of X is $\{1,2,\ldots,n+1\}$, and $X=n+1$ for unsuccessful searches. Since the value of X is fixed for unsuccessful searches, it is more interesting to consider a random variable Y which denotes the number of comparisons on a successful search. The set of all possible values of Y is $\{1,2,\ldots,n\}$. To compute the average search time for a successful search, we must specify the pmf of Y. In the absence of any specific information, it is natural to assume that Y is uniform over its range, i.e.,

$$p_Y(i) = \frac{1}{n} \quad , \; 1 \le i \le n \quad .$$

Then

$$E[Y] = \sum_{i=1}^{n} i p_Y(i) = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2} \quad .$$

Thus,on the average, approximately half the table needs to be searched.

#

Example 4.2:

The assumption of uniform distribution, used in example 4.1, rarely holds in practice. It is possible to collect statistics on access patterns and use empirical distributions to reorganize the table so as to reduce the average search time. If α_i denotes the access probability for name T[i], then the average successful search time $E[Y] = \sum i\alpha_i$. Unlike example 4.1, we now assume that table search starts from the front for convenience. Then E[Y] is minimized when names in the table are in the order of non-increasing access probabilities, i.e., $\alpha_1 \ge \alpha_2 \ge \ldots \ge \alpha_n$. As an example, many tables in practice follow Zipf's law:

$$\alpha_i = \frac{c}{i} \quad , \; 1 \le i \le n \quad ,$$

where the constant c is determined from the normalization requirement, $\sum_{i=1}^{n} \alpha_i = 1$. Thus,

$$c = \frac{1}{\sum_{i=1}^{n} \frac{1}{i}} = \frac{1}{H_n} \cong \frac{1}{\ln(n)}$$

where H_n is the partial sum of a harmonic series, i.e. $H_n = \sum_{i=1}^{n} \frac{1}{i}$.

Now, if the names in the table are ordered as above, then the average search time is

$$E[Y] = \sum_{i=1}^{n} i\alpha_i = \frac{1}{H_n} \sum_{i=1}^{n} 1 = \frac{n}{H_n} \cong \frac{n}{\ln(n)}$$

which is considerably less than the previous value $\frac{n+1}{2}$, for large n.

#

Example 4.3 - Recalling the example of a computer system with five tape drives (Chapter 1), let X be the number of available tape drives. Then

$$E[X] = \sum_{i=0}^{5} ip_X(i) = 0*(\frac{1}{32}) + 1*(\frac{5}{32}) +$$

$$2*(\frac{10}{32}) + 3*(\frac{10}{32}) + 4*(\frac{5}{32}) + 5*(\frac{1}{32})$$

$$= 2.5$$

#

The above example illustrates that E[X] need not correspond to a possible value of the random variable X. The expected value denotes the "center" of a probability distribution in the sense of a weighted average, or better, in the sense of a center of gravity.

Example 4.4 - Let X be a continuous random variable with an exponential density given by:

$$f(x) = \lambda e^{-\lambda x} \quad , \; x \geq 0 \quad .$$

Then

$$E[X] = \int_{-\infty}^{\infty} x f(x) \, dx = \int_{0}^{\infty} \lambda x e^{-\lambda x} dx.$$

Let $u = \lambda x$, then $du = \lambda dx$, and

$$E[X] = \frac{1}{\lambda} \int_{0}^{\infty} u e^{-u} du$$

$$= \frac{1}{\lambda} \Gamma 2 = \frac{1}{\lambda}, \text{ using formula (3.26)}$$ (Gamma distribution)

Thus, if a component has an exponential failure law with parameter λ (known as the failure rate) then its expected life, or its mean time to failure (MTTF) is $\frac{1}{\lambda}$. Similarly, if the interarrival times of jobs to a computer center are exponentially distributed with parameter λ (known as the arrival rate) then the mean (average) interarrival time is $\frac{1}{\lambda}$. Finally, if the service time requirement of a job is an exponentially distributed random variable with parameter μ (known as the service rate), then the mean (average) service time is $\frac{1}{\mu}$.

#

IV.B MOMENTS

Let X be a random variable and define another random variable Y as a function of X so that, $Y = \phi(X)$. Suppose we wish to compute E[Y]. In order to apply definition (4.1), we must first compute the pmf (or pdf in the continuous case) of Y by methods of Chapter 2 (or Chapter 3 in the continuous case). An easier method of computing E[Y] is to use the following result:

$$E[Y] = E[\phi(X)] = \begin{cases} \sum_i \phi(x_i) p_X(x_i) & , \text{ if X is discrete} \\ \int_{-\infty}^{\infty} \phi(x) f_X(x)\, dx & , \text{ if X is continuous} \end{cases} \tag{4.2}$$

(provided the sum or the integral on the right hand side is absolutely convergent).

A special case of interest is the power function $\phi(X) = X^k$, for $k=1,2,3,\ldots$. $E[X^k]$ is known as the k^{th} moment of the random variable X. Note that the first moment, E[X], is the ordinary expectation or the mean of X.

It is possible to show that if X and Y are random variables that have matching corresponding moments of all orders, $E[X^k] = E[Y^k]$ for $k=1,2,\ldots$, then X and Y have the same distribution.

To eliminate the scale of measurement, it is convenient to work with powers of X-E[X]. We define the k^{th} central moment, μ_k, of the random variable X by $\mu_k = E[(X-E[X])^k]$.

Of special interest is the quantity

$$\mu_2 = E[(X-E[X])^2], \tag{4.3}$$

known as the variance of X, Var[X], often denoted by σ^2.

<u>Definition</u> (<u>variance</u>): The variance of a random variable X is

$$\mathrm{Var}[X] = \mu_2 = \sigma^2 = \begin{cases} \int_{-\infty}^{\infty} (x-E[X])^2 f(x)\,dx & \text{, if X is continuous} \\ \sum_i (x_i - E[X])^2 p(x_i) & \text{, if X is discrete.} \end{cases} \tag{4.4}$$

It is clear that Var[X] is always a positive number. The square root of the variance, σ, is known as the standard deviation. The variance and the standard deviation are measures of the "spread" or "dispersion" of a distribution. If X has a "concentrated" distribution so that X takes values near to E[X] with a large probability, then the variance is small (see Figure 4.1).

Figure 4.2 shows a diffuse distribution, i.e., with a large value of σ^2. It should also be noted that variance need not always exist (see Exercise).

Example 4.5 - Let X be an exponentially distributed random variable with parameter λ. Then since $E[X] = \frac{1}{\lambda}$ and $f(x) = \lambda e^{-\lambda x}$,

$$\sigma^2 = \int_0^{\infty} (x - \tfrac{1}{\lambda})^2 \lambda e^{-\lambda x} dx$$

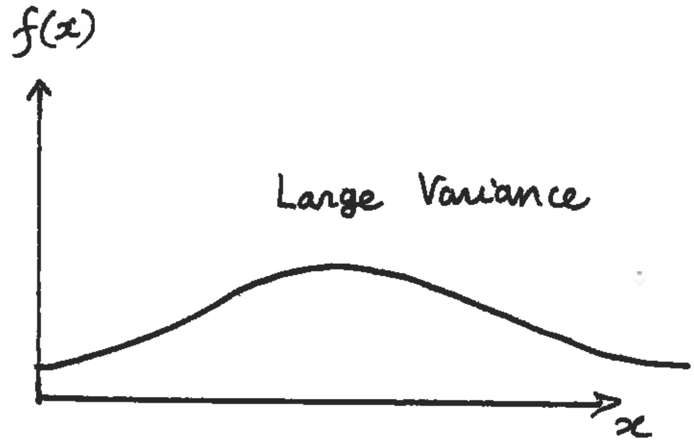

Figure 4.2: The pdf of a Diffuse Distribution.

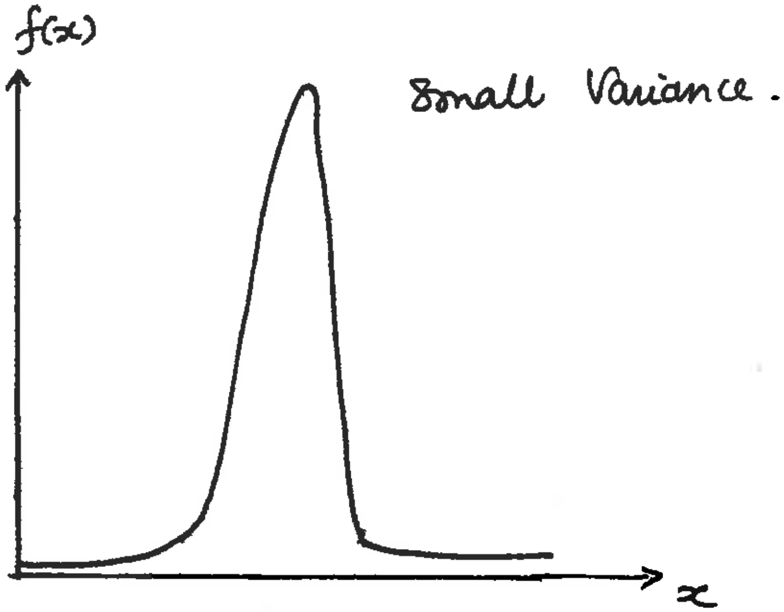

Figure 4.1: The pdf of a "Concentrated" Distribution

$$= \int_0^\infty \lambda x^2 e^{-\lambda x}\, dx - 2\int_0^\infty x e^{-\lambda x} dx + \frac{1}{\lambda}\int_0^\infty e^{-\lambda x} dx$$

$$= \frac{1}{\lambda^2}\Gamma 3 - \frac{2}{\lambda^2}\Gamma 2 + \frac{1}{\lambda^2}\Gamma 1 = \frac{1}{\lambda^2}$$, using formula (3.26)

The standard deviation is expressed in the same units a individual values of the random variable X. For some purposes more useful to measure the spread of the distribution of X in re terms. The coefficient of variation of a random variable X is d by C_X and defined by

$$C_X = \frac{\sigma_X}{E[X]}.$$

Note that the coefficient of variation of an expon random variable is 1, so C_X is a measure of deviation fr exponential distribution.

Yet another function of X that is often of intere $Y = aX+b$, where a and b are constants. It is not difficult t that

$$E[Y] = E[aX+b] = aE[X] + b .$$

In particular, if a is zero, then $E[b] = b$; that is expectation of a constant is that constant. If we take a=1 an E[X], then we conclude that the first central moment, $\mu_1 = E[X -$ $= E[X] - E[X] = 0$.

IV.C EXPECTATION OF FUNCTIONS OF MORE THAN ONE RANDOM VARIABLE

Let $X_1, X_2, \ldots, X_n$ be n random variables defined on the same probability space and let $Y = \phi(X_1, X_2, \ldots, X_n)$. Then

$$E[Y] = E[\phi(X_1, X_2, \ldots, X_n)]$$

$$= \begin{cases} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \phi(x_1,x_2,\ldots,x_n) f(x_1,x_2,\ldots,x_n)\,dx_1 dx_2 \ldots dx_n & \text{,continuous case} \\ \sum_{x_1}\sum_{x_2}\ldots\sum_{x_n} \phi(x_1,x_2,\ldots,x_n) p(x_1,x_2,\ldots,x_n) & \text{,discrete case} \end{cases} \quad (4.7)$$

Example 4.6 - Consider a moving head disk with the innermost cylinder of radius a and the outermost cylinder of radius b. We assume that the number of cylinders is very large and the cylinders are very close to each other so that we may assume a continuum of cylinders. Let the random variables X and Y respectively denote the current and the desired position of the head. Further assume that X and Y are independent and uniformly distributed over the interval (a,b). Therefore,

$$f_X(x) = f_Y(y) = \frac{1}{b-a} \quad , a < x, y < b$$

and

$$f(x,y) = \frac{1}{(b-a)^2} \quad , a < x, y < b$$

Head movement for a seek operation traverses a distance which is a random variable given by $|X-Y|$. The expected seek distance is then given by (see Figure 4.3)

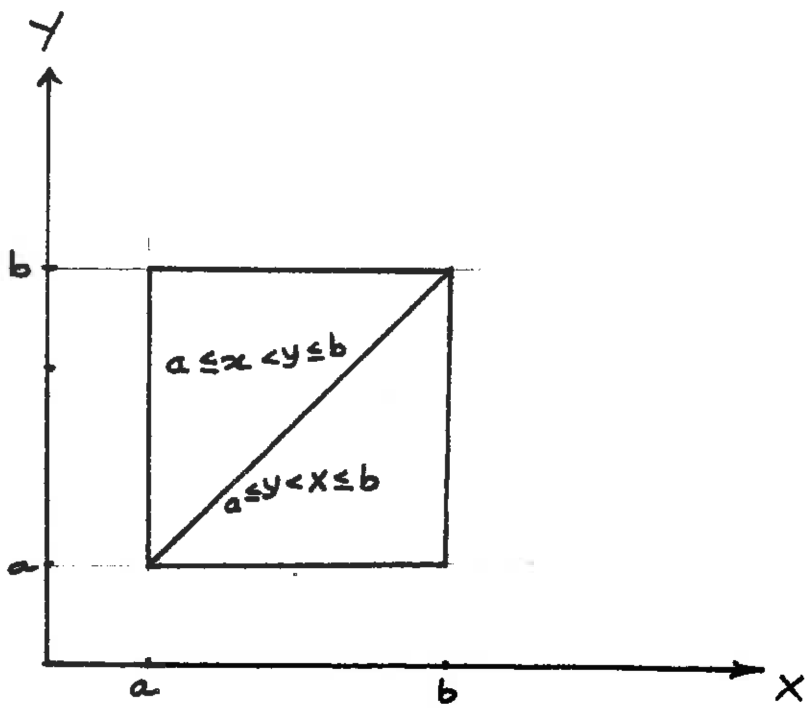

Figure 4.3: Two Areas of Integration for Example 4.6.

$$E[|X-Y|] = \int_a^b \int_a^b |x-y| f(x,y)\ dxdy$$

$$= \iint |x-y| \frac{1}{(b-a)^2}\ dxdy$$

$$= \int_{a\leq y<x\leq b} \int \frac{(x-y)}{(b-a)^2}\ dydx + \int_{a\leq x\leq y\leq b} \int \frac{(y-x)}{(b-a)^2}\ dydx$$

$$= \frac{2}{(b-a)^2} \int_a^b \int_a^x (x-y)\ dydx \qquad \text{(by symmetry)}$$

$$= \frac{2}{(b-a)^2} \int_a^b \left(xy - \frac{y^2}{2}\right) \Big|_a^x\ dx$$

$$= \frac{2}{(b-a)^2} \int_a^b \left(x^2 - ax - \frac{x^2}{2} + \frac{a^2}{2}\right)\ dx$$

$$= \frac{2}{(b-a)^2} \left[\frac{b^3-a^3}{6} - \frac{a}{2}(b^2-a^2) + \frac{a^2(b-a)}{2}\right]$$

$$= \frac{b-a}{3} \quad .$$

Thus, the expected seek distance is one third the maximum seek distance. The intuition may have led us to the incorrect conclusion that the expected seek distance is half of the maximum.

#

Certain functions of random variables (e.g., sums), are of special interest and are of considerable use.

Theorem 4.1 (The linearity property of expectation):

Let X and Y be two random variables. Then the expectation of their sum is the sum of their expectations, that is, if Z=X+Y, then

$E[Z] = E[X+Y] = E[X] + E[Y]$.

Proof:

We will prove the theorem assuming that X,Y, and hence Z are continuous random variables. Proof for the discrete case is very similar.

$$E[X+Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)f(x,y)\ dxdy$$

$$= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f(x,y)\ dydx + \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f(x,y)\ dxdy$$

$$= \int_{-\infty}^{\infty} x f_X(x)\ dx + \int_{-\infty}^{\infty} y f_Y(y)\ dy \qquad \text{by definition of the marginal densities}$$

$$= E[X]+E[Y] .$$

#

Note that the above theorem does not require that X and Y be independent. It can be generalized to the case of n variables, i.e.

$$E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i]$$

and to

$$E[\sum_{i=1}^{n} a_i X_i] = \sum_{i=1}^{n} a_i E[X_i] \quad , \qquad (4.8)$$

where $a_1, \ldots, a_n$ are constants. For instance, let $X_1, X_2, \ldots X_n$ be random variables (not necessarily independent) with a common mean $\mu = E[X_i]$ ($i = 1, 2, \ldots, n$). Then the expected value of their sample mean (defined in section 3.1) is equal to μ:

$$E[\overline{X}] = E\left[\frac{\sum_{i=1}^{n} X_i}{n}\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \mu \tag{4.9}$$

Example 4.7 - We have noted that the variance

$$\sigma^2 = E[(X-E[X])^2]$$

$$= E[X^2-2XE[X]+(E[X])^2]$$

$$= E[X^2]-E[2XE[X]]+(E[X])^2 \quad \text{by the above linear property of expectation}$$

$$= E[X^2]-2E[X]E[X]+(E[X])^2 \quad \text{noting that } E[X] \text{ is a constant}$$

$$\sigma^2 = E[X^2]-(E[X])^2 \tag{4.10}$$

This formula for Var[X] is usually preferred over the original definition.

#

Unlike the case of expectation of a sum, the expectation of a product of two random variables does not have a simple form, unless the two random variables are independent.

Theorem 4.2:

$E[XY] = E[X]E[Y]$, if X and Y are independent random variables.

Proof:

We give a proof of the theorem assuming X and Y are discrete random variables. The proof for the continuous case is similar.

$$E[XY] = \sum_i\sum_j x_i y_j p(x_i, y_j)$$

$$= \sum_i\sum_j x_i y_j p_X(x_i) p_Y(y_j) \quad \text{,by independence}$$

$$= \sum_i x_i p_X(x_i) \sum_j y_j p_Y(y_j) = E[X]E[Y] \quad .$$

#

It should be noted that converse of Theorem 4.2 does not hold, that is, random variable X and Y may satisfy the relation E[XY] = E[X] E[Y] without being independent.

The above theorem can be easily generalized to a mutually independent set of n random variables $X_1, X_2, \ldots, X_n$:

$$E[\prod_{i=1}^{n} X_i] = \prod_{i=1}^{n} E[X_i] \tag{4.11}$$

and further to

$$E[\prod_{i=1}^{n} \phi_i(X_i)] = \prod_{i=1}^{n} E[\phi_i(X_i)] \quad .$$

Again with the assumption of independence, the variance of a sum takes a simpler form also:

<u>Theorem 4.3</u>:

Var[X+Y] = Var[X] + Var[Y], if X and Y are independent random variables.

<u>Proof</u>:

From the definition of variance,

$$\begin{aligned} Var[X+Y] &= E[((X+Y) - E[X+Y])^2] \\ &= E[((X+Y) - E[X]-E[Y])^2] \\ &= E[\ (X-E[X])^2 + (Y-E[Y])^2 + 2(X-E[X])(Y-E[Y])\ ] \\ &= E[(X-E[X])^2] + E[(Y-E[Y])^2] + 2E[(X-E[X])(Y-E[Y])] \end{aligned}$$

$$\mathrm{Var}[\sum_{i=1}^{n} X_i] = n\mathrm{Var}[X_i] = n\sigma^2 \quad (4.14)$$

and the variance of their sample mean is

$$\mathrm{Var}[\bar{X}] = \mathrm{Var}[\frac{\sum_{i=1}^{n} X_i}{n}] = \frac{1}{n^2}\mathrm{Var}[\sum_{i=1}^{n} X_i] = \frac{\sigma^2}{n} \quad (4.15)$$

We have noted that Cov(X,Y) = 0, if X and Y are independent random variables. However, it is possible for two random variables to satisfy the condition Cov(X,Y) = 0 without being independent.

Definition (Uncorrelated Random Variables):

Random variables X and Y are said to be uncorrelated provided Cov(X, Y) = 0.

Since Cov(X,Y) = E[XY] -E[X]E[Y], an equivalent definition of uncorrelated random variables is the condition E[XY] = E[X] E[Y]. It follows that independent random variables are uncorrelated but the converse need not hold.

Example 4.8 Let X be uniformly distributed over the interval (-1,1) and let $Y=X^2$, so Y is completely dependent on X. Noting that for all odd values of k > 0, k^{th} moment $E[X^k] = 0$, we have,

$$E[XY]=E[X^3]=0 \text{ and } E[X]E[Y]=0 * E[Y] = 0.$$

Therefore X and Y are uncorrelated!

#

$= Var[X] + Var[Y] + 2E[(X-E[X])(Y-E[Y])]$,

by the linearity property of expectation.

The quantity $E[(X-E[X])(Y-E[Y])]$ is defined to be the <u>covariance</u> of X and Y and is denoted by Cov(X,Y). It is easy to see that Cov(X,Y) is zero when X and Y are independent:

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

, by the linearity of expectation

$$\begin{aligned} &= E[XY] - E[X]E[Y] \\ &= 0, \text{ by Theorem 4.2 since X and Y are independent.} \end{aligned}$$

Therefore, $Var[X+Y] = Var[X]+Var[Y]$ if X and Y are independent random variables.

In case X and Y are not independent we obtain the formula

$$Var[X+Y]= Var[X]+Var[Y] + 2Cov(X,Y) \qquad (4.12)$$

The above theorem can be generalized for a set of n mutually independent variables $X_1, X_2, \ldots, X_n$; and constants $a_1, a_2, \ldots, a_n$:

$$Var[\sum_{i=1}^{n} a_i X_i] = \sum_{i=1}^{n} a_i^2 Var[X_i] \qquad (4.13)$$

Thus if $X_1, X_2, \ldots, X_n$ are independent random variables with a common variance $\sigma^2 = Var[X_i]$ $(i=1,2,\ldots, n)$ then the variance of their sum is given by

We have declared that $Cov(X,Y) = 0$, means X and Y are uncorrelated. On the other hand, if X and Y are linearly related, that is, $X = a\,Y$ for some constant $a \neq 0$, then since $E[X] = a\,E[Y]$, we have,

$$Cov(X,Y) = a\,Var[Y] = 1/a\,Var[X]$$

or

$$Cov^2(X,Y) = Var[X]\,Var[Y]$$

In the general case, it can be shown that

$$0 \leq Cov^2(X,Y) \leq Var[X]\,Var[Y] \tag{4.16}$$

using the following Cauchy-Schwarz inequality:

$$(E[XY])^2 \leq E[X^2]\,E[Y^2] \tag{4.17}$$

Cov(X,Y) measures the degree of linear dependence (or the degree of correlation) between the two random variables. Recalling the example 4.8, we note the notion of covariance completely misses the quadratic dependence. It is often useful to define a measure of this dependence in a scale-independent fashion. The correlation coefficient $\rho(X,Y)$ is defined by

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var[X]\,Var[Y]}} \tag{4.18}$$

$$= \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

Using the relation (4.16), we conclude that

$$-1 \leq \rho(X,Y) \leq 1. \tag{4.19}$$

Also,

$$\rho(X,Y) = \begin{cases} -1 & \text{if } X = -aY \ (a>0) \\ 0 & \text{if X and Y are uncorrelated} \\ +1 & \text{if } X = aY \ (a>0) \end{cases} \qquad (4.20)$$

IV.D TRANSFORM METHODS

In many probability problems, the form of the density function (or the pmf in the discrete case) may be so complex so as to make computations difficult, if not impossible. As an example, recall the analysis of the program MAX. A transform can provide a compact description of a distribution and it is relatively easy to compute the mean, the variance, and other moments directly from a transform rather than resorting to a tedious sum (discrete case) or an equally tedious integral (continuous case). The transform methods are particularly useful in problems involving sums of independent random variables, and in solving difference equations (discrete case) and differential equations (continuous case) related to a stochastic process. We will introduce the z-transform (also called the probability generating function), the Laplace transform, and the characteristic function (also called the Fourier transform). We will first define the moment generating function and derive the above three transforms as special cases.

For a random variable X, $e^{X\theta}$ is another random variable. The expectation $E[e^{X\theta}]$ will be a function of θ. Define the moment generating function (MGF) $M_X(\theta)$, abbreviated $M(\theta)$, of the random