EE 503

Lecture Notes Part 9

Christopher Wayne Walker, Ph.D.

9.0 Functions of One Random Variable

Often we are concerned with functions of a random variable [recall a random variable itself is a function of the outcome of an experiment]. If $y = g(x)$ is a real-valued function then $Y = g(X)$ is a random variable. Given $f_X(x)$ or $F_X(x)$ we seek $f_Y(y)$ and $F_Y(y)$.

Definition: Let C be an event (some subset) associated with the range space of Y, R_Y. Define $B \subset R_X$ as

$$B = \{x \in R_X : g(x) \in C\}.$$

Then, B and C are called *equivalent events* (B occurs if and only if C occurs).

Definition: Let X be a random variable defined on the sample space Ω. Let R_X be the range space of X. Let g be a real-valued function and compute the random variable $Y = g(X)$ with range space R_Y. For any $C \subset R_Y$ define

$$P(C) = P\left(\{x \in R_X : g(x) \in C\}\right).$$

Note

$$P(C) = P\left(\{\omega \in \Omega : g\left(X(\omega)\right) \in C\}\right).$$

Example: Let X be a continuous random variable with *pdf*

$$f(x) = \begin{cases} e^{-x}, & x > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

Let $g(x) = 2x + 1$. Then $R_X = \{x : x > 0\}$ while $R_Y = \{y : y > 1\}$. Define the event C as $C = \{Y \geq 5\}$, i.e., $C = \{\omega : g\left(X(\omega)\right) \geq 5\} = \{\omega : Y(\omega) \geq 5\}$. Now $y \geq 5$ iff $2x + 1 \geq 5$ iff $x \geq 2$. So C is equivalent to $B = \{X \geq 2\}$. Now

$$P(X \geq 2) = \int_2^\infty e^{-x} dx = e^{-2}$$

so

$$P(Y \geq 5) = e^{-2}.$$

9.1 Finding the Distribution of $g(X)$

9.1.1 Discrete Case

General Procedure:

1) First we consider the case where X is discrete and Y is discrete.

Let $x_{i1}, x_{i2}, \ldots$ represent the X-values having the property $g(x_{ij}) = y_i, \ \forall \ j$. Then

$$\begin{aligned} f_Y(y_i) &= P(Y = y_i) = P(X = x_{i1}) + P(X = x_{i2}) + \cdots \\ &= f_X(x_{i1}) + f_X(x_{i2}) + \cdots \end{aligned}$$

i.e., to evaluate the probability of the event $\{Y = y_i\}$, find the equivalent event in terms of X and add all the corresponding probabilities together.

Example: Let X have possible values $1, 2, 3, \ldots$ and

$$P(X = n) = (1/2)^n, \ n = 1, 2, \ldots.$$

Let

$$Y = \begin{cases} 1, & x \text{ is even} \\ -1, & x \text{ is odd.} \end{cases}$$

$$\begin{aligned} P(Y = 1) &= P(X = 2) + P(X = 4) + \cdots = (1/2)^2 + (1/2)^4 + \cdots \\ &= \sum_{i=1}^{\infty} (1/2)^{2i} = \sum_{i=1}^{\infty} (1/4)^i = \frac{1/4}{1 - 1/4} = 1/3 \end{aligned}$$

and

$$P(Y = -1) = 1 - P(Y = 1) = 2/3.$$

2) It may turn out that X is a continuous random variable while Y is discrete. For example, X may assume all real values and $Y = 1$ if $X \geq 0$ and $Y = -1$ if $X < 0$. So $P(Y = 1) = P(X \geq 0)$ and $P(Y = -1) = P(X < 0)$. In general, if $\{Y = y_i\}$ is equivalent to an event, say A, in the range space of X, then

$$f_Y(y_i) = P(Y = y_i) = \int_A f_X(x)dx.$$

9.1.2 Continuous Case

Here X is a continuous random variable and g is a continuous function. So, $Y = g(X)$ is a continuous random variable. We seek the *pdf* of Y, i.e., $f_Y(y)$.

General Procedure:

i. Obtain $F_Y(y) = P(Y \leq y)$ by finding the event A (in the range space of X) which is equivalent to the event $\{Y \leq y\}$.

ii. Differentiate $F_Y(y)$ to get $f_Y(y)$.

iii. Determine those values in the range space of Y for which $f_Y(y) > 0$.

Example: Let

$$f_X(x) = \begin{cases} 2x, & 0 < x < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Let $g(x) = 3x + 1$. Then

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(3X + 1 \leq y) = P\left(X \leq \frac{y-1}{3}\right) \\ &= \int_0^{\frac{y-1}{3}} 2x dx = \left(\frac{y-1}{3}\right)^2. \end{aligned}$$

Thus,

$$f_Y(y) = F'_Y(y) = \frac{2}{9}(y-1).$$

Now $f_X(x) > 0$ for $0 < x < 1$, therefore $f_Y(y) > 0$ for $1 < y < 4$.

There is another way of getting the same result. Consider

$$F_Y(y) = P(Y \leq y) = P(3X + 1 \leq y) = P\left(X \leq \frac{y-1}{3}\right) = F_X\left(\frac{y-1}{3}\right).$$

Then,

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X\left(\frac{y-1}{3}\right) = \frac{dF_Y(y)}{du}\frac{du}{dy} = \frac{dF_X(u)}{du}\frac{du}{dy}$$

where $u = (y-1)/3$. So,

$$f_Y(y) = F'_X(u)\frac{du}{dy} = f_X(u)\frac{du}{dy} = 2\left(\frac{y-1}{3}\right)\frac{1}{3} = \frac{2}{9}(y-1).$$

The following theorem is very useful if the conditions of the theorem are met.

<u>**Theorem:**</u> Let X be a continuous random variable with *pdf* $f_X(x) > 0$ for $a < x < b$. Suppose that $y = g(x)$ is a strictly monotone (strictly increasing or strictly decreasing) function of x. Assume that g is differentiable (and hence continuous) for all x. Then $Y = g(X)$ has *pdf*

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|$$

where x is expressed in terms of y, i.e., $x = g^{-1}(y)$. Hence.

$$f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\frac{dg^{-1}(y)}{dy}\right|.$$

If g is increasing then g is nonzero for those values of y satisfying $g(a) < y < g(b)$. If g is decreasing then g is nonzero for y satisfying $g(b) < y < g(a)$.

<u>**Proof:**</u> First assume g is a strictly increasing function. Then

$$F_Y(y) = P(Y \le y) = P(g(X) \le y) = P\left(X \le g^{-1}(y)\right) = F_X\left(g^{-1}(y)\right).$$

Thus,

$$f_Y(y) = \frac{dF_Y(x)}{dx}\frac{dx}{dy} = \frac{dF_X(x)}{dx}\frac{dx}{dy}$$

where $x = g^{-1}(y)$. Hence,

$$f_Y(y) = f_X(x)\frac{dx}{dy} = f_X(x)\left|\frac{dx}{dy}\right|.$$

Now assume g is a strictly decreasing function. Then,

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(g(X) \le y) = P\left(X \ge g^{-1}(y)\right) \\ &= 1 - P\left(X < g^{-1}(y)\right) = 1 - F_X\left(g^{-1}(y)\right). \end{aligned}$$

So,

$$\frac{dF_Y(y)}{dy} = \frac{dF_Y(y)}{dx}\frac{dx}{dy}, \;\; x = g^{-1}(y)$$

or

$$f_Y(y) = \frac{d}{dx}\left[1 - F_X(x)\right]\frac{dx}{dy} = -f_X(x)\frac{dx}{dy}.$$

But, $\frac{dx}{dy} < 0$ since $g(x)$ is strictly decreasing. Hence,

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|.$$

Note: If $y = g(x)$ is not a strictly monotone function of x, we cannot apply the above theorem directly but we can still use the general method.

Example: Suppose

$$f_X(x) = \begin{cases} \frac{1}{2}, & -1 < x < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Let $g(x) = x^2$. This function is not monotone over the interval (-1, 1). Here

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) + P(X = -\sqrt{y}). \end{aligned}$$

Thus,

$$f_Y(y) = \frac{f_X(\sqrt{y})}{2\sqrt{y}} - \frac{f_X(-\sqrt{y})}{-2\sqrt{y}} = \frac{1}{2\sqrt{y}}\left[f_X(\sqrt{y}) + f_X(-\sqrt{y})\right]$$

or

$$f_Y(y) = \frac{1}{2\sqrt{y}}\left(\frac{1}{2} + \frac{1}{2}\right) = \frac{1}{2\sqrt{y}}, \quad 0 < y < 1.$$

The result obtained in this example gives the following theorem.

Theorem: Let X be a continuous random variable with *pdf* $f_X(x)$. Let $Y = X^2$. Then the *pdf* of Y is

$$f_Y(y) = \frac{1}{2\sqrt{y}}\left[f_X(\sqrt{y}) + f_X(-\sqrt{y})\right].$$

Example: Suppose

$$f_X(x) = \begin{cases} 3x^2, & 0 \le x \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Let $Y = e^{-X} = g(X)$. Note that $g(x)$ is monotone in [0, 1]. Thus,

$$f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\frac{d}{dy}g^{-1}(y)\right|.$$

$$g(x) = e^{-x} \text{ or } y = e^{-x} \Rightarrow -x = \ln y \Rightarrow x = -\ln y = g^{-1}(y).$$

Thus,

$$f_Y(y) = 3(-\ln y)^2\left|\frac{-1}{y}\right| = 3(\ln y)^2\frac{1}{y}.$$

Endpoints: $x = 0 \Rightarrow y = 1, \quad x = 1 \Rightarrow y = e^{-1}$. So,

$$f_Y(y) = \begin{cases} 3(\ln y)^2\frac{1}{y}, & e^{-1} \le y \le 1, \\ \\ 0, & \text{elsewhere.} \end{cases}$$

Inverse Problem:

i. Given a random variable X with distribution function $F_X(x)$, find $g(x_0)$ so that $U = g(X)$ is uniform in (0, 1).

Claim $g(x_0) = F_X(x_0)$ works.

Proof:

$$F_U(u_0) = P(U \le u_0) = P(X \le x_0) = F_X(x_0) = u_0.$$

ii. Given a random variable $U \sim U(0, 1)$, find $g(u_0)$ so that $Y = g(U)$ has some desired distribution function $F_Y(y_0)$.

Claim $g(u_0) = F_Y^{-1}(u_0)$ works.

Proof:

$$Y = F_Y^{-1}(U) \Leftrightarrow P(Y \le y_0) = P(F_Y^{-1}(U) \le y_0) = P\left(U \le F_Y(y_0)\right) = F_Y(y_0).$$

iii. Given X with distribution function $F_X(x_0)$, find $g(x_0)$ so that $Y = g(X)$ has some desired distribution function $F_Y(y_0)$.

Claim $g(x_0) = F_Y^{-1}(F_X(x_0))$ works.

Proof:

$$\begin{aligned} Y &= F_Y^{-1}(F_X(X)) \Leftrightarrow P(Y \leq y_0) = P(F_Y^{-1}(F_X(X)) \leq y_0) = P\left(F_X(X) \leq F_Y(y_0)\right) \\ &= F_Y(y_0) \text{ since } F_X(X) \text{ is } U(0,1) \text{ from (i) above.} \end{aligned}$$

Example: Say $X \sim U(0,1)$. We desire a distribution function

$$F_Y(y_0) = \begin{cases} 1 - e^{-y_0}, & y_0 \geq 0, \\ 0, & \text{elsewhere.} \end{cases}$$

Find $y_0 = g(x_0)$.

Solution: Using (ii) above we let $g(x_0) = F_Y^{-1}(x_0)$. We then solve $y_0 = F_Y^{-1}(x_0)$ to get $x_0 = F_Y(y_0)$, so set $x_0 = 1 - e^{-y_0}$. Hence, $y_0 = -\ln(1 - x_0)$ and thus $g(x_0) = -\ln(1 - x_0)$.

9.2 Expectations

Often we would like to know the mean (or average or expected value) of a random variable resulting from a random experiment.

9.2.1 Discrete Case

Say X has values in $\{x_1, x_2, x_3, \ldots\}$ and $P(X = x_i) = f(x_i) = p_i$.

Definition: The *expected value* of X is given by

$$E(X) = \sum_i x_i f(x_i)$$

whenever this sum exists.

Recall $f(x_i) \geq 0$ and $\sum_i f(x_i) = 1$. So, $E(X)$ is the average of the values of X with each value weighted according to its probability of occurrence.

Lemma: $E[g(X)] = \sum_i g(x_i) f_X(x_i)$ where $f_X(x_i) = P(X = x_i) = p_i$.

Proof: Let $Y = g(X)$, then $E[g(X)] = E(Y) = \sum_i y_i f_Y(y_i)$. Now

$$\sum_i g(x_i) f_X(x_i) = \sum_i \left(\sum_j g(x_j) f_X(x_j) \right)$$

where the inner sum is over all indices j for which $g(x_j) = y_i$, for some fixed y_i. Thus, all the terms $g(x_j)$ are constant in the inner sum. Hence,

$$\sum_i g(x_i) f_X(x_i) = \sum_i y_i \sum_j f_X(x_j).$$

But,

$$\sum_j f_X(x_j) = \sum_j P(X = x_j) = P(Y = y_i) = f_Y(y_i).$$

So,

$$\sum_i g(x_i) f_X(x_i) = \sum_i y_i f_Y(y_i).$$

Theorem:

a. $X \geq 0 \Rightarrow E(X) \geq 0$.

b. $E(aX + bY) = aE(X) + bE(Y)$.

c. $E(1) = 1$.

Proof: Easy. Left as an exercise.

Note: Part (a) of this theorem implies expectation is a linear operator.

9.2.2 Continuous Case

Definition: The *expected value* of X is given by

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

whenever this integral exists.

Example: Say X has *pdf*

$$f_X(x) = \begin{cases} 3x^2, & 0 \le x \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Then,

$$E(X) = \int_0^1 x(3x^2)dx = 3/4.$$

Lemma: $E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$.

Proof: Omitted.

Theorem:

a. $X \ge 0 \Rightarrow E(X) \ge 0$.

b. $E(aX + bY) = aE(X) + bE(Y)$.

c. $E(1) = 1$.

Proof: Easy. Left as an exercise.

9.3 Variance

The variance measure gives us an indication of the spread of the data about its mean.

Notation: The variance of X is written $Var(X)$ or σ_X^2 or σ^2.

9.3.1 Discrete Case

Definition: The *variance* of X is given by

$$Var(X) = \sum_i (x_i - \mu)^2 f_X(x_i)$$

where $\mu = E(X)$. Thus, $Var(X) = E\left[(X - \mu)^2\right]$.

9.3.2 Continuous Case

Definition: The *variance* of X is given by

$$Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x) dx$$

where $\mu = E(X)$. Again, $Var(X) = E\left[(X-\mu)^2\right]$.

9.4 Examples and Additional Results

Theorem: Let X be binomially distributed with parameters n, p (write $X \sim B(n,p)$). Then $E(X) = np$.

Proof:

i. Direct proof.

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

So,

$$\begin{aligned} E(X) &= \sum_{k=0}^{n} k \frac{n!}{(n-k)!k!} \, p^k (1-p)^{n-k} \\ &= \sum_{k=1}^{n} \frac{n!}{(n-k)!(k-1)!} \, p^k (1-p)^{n-k}. \end{aligned}$$

Let $s = k-1$. Then

$$\begin{aligned} E(X) &= \sum_{s=0}^{n-1} \frac{n!}{(n-s-1)!s!} \, p^{s+1} (1-p)^{n-s-1} \\ &= \sum_{s=0}^{n-1} n \binom{n-1}{s} p^{s+1} (1-p)^{n-s-1} \\ &= np \sum_{s=0}^{n-1} \binom{n-1}{s} p^s (1-p)^{n-1-s}. \end{aligned}$$

Recall the binomial theorem:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

Replace n by $n-1$, let $a = p,\ b = 1-p$ to get

$$1 = [p + (1-p)]^{n-1} = \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k}.$$

Therefore,

$$\sum_{s=0}^{n-1} \binom{n-1}{s} p^s (1-p)^{n-1-s} = 1$$

and thus $E(X) = np$.

ii. Quick proof.
Think of X as the number of successes in n Bernoulli trials. Let X_i be the number of successes on the ith trial. Then,

$$P(X_i = 1) = p, \quad P(X_i = 0) = 1 - p.$$

Hence,

$$E(X_i) = 1 \cdot p + 0 \cdot (1-p) = p.$$

Now,

$$X = X_1 + X_2 + \cdots + X_n.$$

So,

$$\begin{aligned} E(X) &= E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n) \\ &= p + \cdots + p = np. \end{aligned}$$

Example: Say X has *pdf*

$$f_X(x) = \begin{cases} 2x, & 0 < x < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Let $Y = g(X) = 3X + 1$. Find $E(Y)$.

Previously, (see section 9.1.2) we derived

$$f_Y(y) = \begin{cases} \frac{2}{9}(y-1), & 1 < y < 4, \\ 0, & \text{elsewhere.} \end{cases}$$

So,

$$E(Y) = \int_1^4 y \left[\frac{2}{9}(y-1)\right] dy = 3$$

or (without finding the *pdf* of Y),

$$E(Y) = E[g(X)] = \int_0^1 (3x+1)(2x)dx = 3.$$

Theorem: $Var(X) = E\left(X^2\right) - [E(X)]^2$.

Proof:

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

since $\mu = E(X)$.

Properties of Variance:

i. If c is a constant, $Var(X+c) = Var(X)$.

   Proof:

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

ii. If c is a constant, $Var(cX) = c^2 Var(X)$.

   Proof: Exercise.

Lemma: If X is discrete and takes values $1, 2, 3, \ldots$, then

$$E(X) = \sum_{n=0}^{\infty} P(X > n).$$

Proof: Let $p_i = P(X = i)$. Then

$$\begin{array}{rcccccccccccc}
P(X > 0) & = & p_1 & + & p_2 & + & p_3 & + & \cdots & + & p_k & + & \cdots \\
P(X > 1) & = & & & p_2 & + & p_3 & + & \cdots & + & p_k & + & \cdots \\
P(X > 2) & = & & & & & p_3 & + & \cdots & + & p_k & + & \cdots \\
\vdots & & & & & & & & & & & & \\
P(X > k) & = & & & & & & & & & & & p_{k+1} + \cdots \\
\vdots & & & & & & & & & & & &
\end{array}$$

By summing along the rows we find the total in the array is

$$\sum_{n=0}^{\infty} P(X > n).$$

By summing along the columns we find the total in the array is

$$p_1 + 2p_2 + 3p_3 + \cdots + kp_k + \cdots.$$

Thus equating these last results we get

$$\sum_{n=0}^{\infty} P(X > n) = p_1 + 2p_2 + 3p_3 + \cdots$$

$$= \sum_{k=1}^{\infty} kp_k = E(X).$$

An Interpretation of Expectation:

Suppose we measure the squared distance between a random variable X and a constant b by $(X - b)^2$. Let us find b that minimizes $E[(X - b)^2]$, which gives us a predictor of X. (We do not try to find a b that minimizes $(X - b)^2$ since such a b would depend on X so we could not use it as a predictor.)

Consider

$$E[(X - b)^2] = \int_{-\infty}^{\infty} (x - b)^2 f_X(x) dx.$$

Set

$$\frac{d}{db} E[(X - b)^2] = 0 \Rightarrow \frac{d}{db} \int_{-\infty}^{\infty} (x - b)^2 f_X(x) dx = 0.$$

We can solve this if we can exchange the order of differentiation and integration (justification, in general, requires measure theory concepts). Assuming okay, we get

$$\int_{-\infty}^{\infty} -2(x-b)f_X(x)dx = 0 \Rightarrow \int_{-\infty}^{\infty} xf_X(x)dx = b\int_{-\infty}^{\infty} f_X(x)dx.$$

But

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

thus

$$b = \int_{-\infty}^{\infty} xf_X(x)dx = E(X).$$

We can get this same result another way as follows:

$$\begin{aligned} E\left[(X-b)^2\right] &= E\left[(X-E(X)+E(X)-b)^2\right] \\ &= E\left[((X-E(X))+(E(X)-b))^2\right] \\ &= E\left[(X-E(X))^2\right] + 2E\left[(X-E(X))(E(X)-b)\right] + E\left[(E(X)-b)^2\right]. \end{aligned}$$

Now $(E(X)-b)$ is a constant so

$$\begin{aligned} E\left[(X-E(X))(E(X)-b)\right] &= (E(X)-b)E\left[(X-E(X))\right] \\ &= (E(X)-b)(E(X)-E(X)) = 0 \end{aligned}$$

so

$$E\left[(X-b)^2\right] = E\left[(X-E(X))^2\right] + (E(X)-b)^2.$$

We have no control over $E\left[(X-E(X))^2\right]$ since there is no b in this expression. Thus, $E\left[(X-b)^2\right]$ is minimized if we minimize $(E(X)-b)^2$. Since $(E(X)-b)^2 \geq 0$ this term is minimized if $b = E(X)$. Hence,

$$\min_b E\left[(X-b)^2\right] = E\left[(X-E(X))^2\right].$$

9.5 Moments

Definitions: For $k = 1, 2, 3, \ldots$, the *kth moment* of X is

$$m_k = E\left[X^k\right]$$

and the *kth central moment* of X is

$$\mu_k = E\left[(X - E(X))^k\right].$$

Note: The 2nd central moment of X is the variance of X, $Var(X) = \sigma_X^2$. The *standard deviation* of X is $\sigma_X = \sqrt{\sigma_X^2}$.

Normal Case: Consider the mean-zero normal density

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}.$$

Claim: For $n \geq 1$,

$$E\left[X^n\right] = \begin{cases} 0, & n \text{ odd}, \\ 1 \cdot 3 \cdot 5 \cdots (n-1)\sigma^n, & n \text{ even}. \end{cases}$$

Proof: If n is odd it is obvious. So assume n is even.

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-x^2/2\sigma^2} dx = 1 \Rightarrow \int_{-\infty}^{\infty} e^{-x^2/2\sigma^2} dx = \sqrt{2\pi}\sigma.$$

Let $\alpha = \frac{1}{2\sigma^2}$. We get

$$\int_{-\infty}^{\infty} e^{-\alpha x^2} dx = \sqrt{\pi/\alpha}.$$

Take the derivative with respect to α to get

$$\int_{-\infty}^{\infty} -x^2 e^{-\alpha x^2} dx = -\frac{1}{2}\sqrt{\pi}\alpha^{-3/2}.$$

Cancel the minus signs and continue taking derivatives to get upon the kth derivative

$$\begin{aligned} \int_{-\infty}^{\infty} x^{2k} e^{-\alpha x^2} dx &= \frac{1 \cdot 3 \cdot 5 \cdots (2k-1)}{2^k} \frac{\sqrt{\pi}}{\sqrt{\alpha^{2k+1}}}. \\ &= \frac{1 \cdot 3 \cdot 5 \cdots (2k-1)}{2^k} \sqrt{\pi(2\sigma^2)^{2k+1}}. \end{aligned}$$

Using $n = 2k$ we deduce

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x^n e^{-x^2/2\sigma^2} dx = 1 \cdot 3 \cdot 5 \cdots (n-1)\sigma^n, \ n \text{ even}.$$

The left hand side of the above result is $E\left[X^n\right]$.

Sometimes it is useful to bound probabilities, especially when the probabilities are difficult to calculate or the density and/or the distribution functions are not even known. The Tchebycheff (Chebyshev) Inequality helps us here.

Theorem: (*Tchebycheff Inequality*). For any $\epsilon > 0$,

$$P\left(|X-\mu| \geq \epsilon\right) \leq \frac{\sigma^2}{\epsilon^2},$$

where $\mu = E(X)$ and $\sigma^2 = Var(X)$.

Proof:

$$P\left(|X-\mu| \geq \epsilon\right) = \int_{-\infty}^{-\mu-\epsilon} f(x)dx + \int_{\mu+\epsilon}^{\infty} f(x)dx = \int_{|X-\mu|\geq\epsilon} f(x)dx.$$

Now

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \geq \int_{|X-\mu|\geq\epsilon} (x-\mu)^2 f(x)dx \geq \epsilon^2 \int_{|X-\mu|\geq\epsilon} f(x)dx.$$

But,

$$\int_{|X-\mu|\geq\epsilon} f(x)dx = P\left(|X-\mu| \geq \epsilon\right).$$

So

$$P\left(|X-\mu| \geq \epsilon\right) \leq \frac{\sigma^2}{\epsilon^2}.$$

9.6 Moment Generating Function

The moment generating function has many uses, one of which is to calculate moments of a random variable.

Definition: Let X be a random variable. The *moment generating function* (*mgf*) of X is given by

$$M_X(s) = M(s) = E\left(e^{sX}\right).$$

For X discrete, the *mgf* of X is

$$M_X(s) = \sum_i e^{sx_i} P(X = x_i).$$

For X continuous, the *mgf* of X is

$$M_X(s) = \int_{-\infty}^{\infty} e^{sx} f_X(x) dx.$$

Recall,

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

which converges for all constants x. So,

$$e^{sx} = 1 + sx + \frac{(sx)^2}{2!} + \frac{(sx)^3}{3!} + \cdots.$$

Now

$$M_X(s) = E\left(e^{sX}\right) = E\left(1 + sX + \frac{(sX)^2}{2!} + \frac{(sX)^3}{3!} + \cdots\right).$$

If we assume the *mgf* exists then the expectation of the sum is the sum of the expectations, so

$$M_X(s) = 1 + sE(X) + \frac{s^2 E(X^2)}{2!} + \frac{s^3 E(X^3)}{3!} + \cdots.$$

We can also calculate $M_X'(s)$ by taking the derivative of each term to get

$$M_X'(s) = E(X) + sE(X^2) + \frac{s^2 E(X^3)}{2!} + \cdots.$$

We set $s = 0$ to conclude $M_X'(0) = E(X)$. Also,

$$M_X''(s) = E(X^2) + sE(X^3) + \frac{s^2 E(X^4)}{2!} + \cdots.$$

We see that $M_X''(0) = E(X^2)$. Continuing on leads to the following theorem.

Theorem: $M_X^{(n)}(0) = E(X^n)$.

Note: The *mgf* in the continuous case is related to the Laplace transform.

9.6.1 Examples

Binomial Case: Say X is binomially distributed with parameters n, p. We have

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k},\ k = 1, 2, 3, \ldots, n.$$

Then

$$M_X(s) = E\left(e^{sX}\right) = \sum_{k=0}^{n} e^{sk}\binom{n}{k}p^k(1-p)^{n-k}$$
$$= \sum_{k=0}^{n}\binom{n}{k}\left(pe^s\right)^k(1-p)^{n-k}$$
$$= \left[pe^s + (1-p)\right]^n.$$

Thus,

$$M_X(0) = [p+1-p]^n = 1 = E\left(X^0\right) = E(1).$$
$$M_X'(s) = n\left[pe^s + (1-p)\right]^{n-1} pe^s.$$
$$M_X'(0) = np = E(X).$$

$$\begin{aligned} M_X''(s) &= n\left[pe^s + (1-p)\right]^{n-1} pe^s + pe^s n(n-1)\left[pe^s + (1-p)\right]^{n-2} pe^s \\ &= np\left[\left(pe^s + (1-p)\right)^{n-1} e^s + e^s n(n-1)\left(pe^s + (1-p)\right)^{n-2} pe^s\right]. \end{aligned}$$

$$M_X''(0) = np[1+(n-1)p] = E\left(X^2\right).$$

So

$$Var(X) = E\left(X^2\right) - [E(X)]^2 = np[1+(n-1)p] - (np)^2 = np(1-p).$$

<u>**Normal Case:**</u> Here $X \sim N\left(\mu, \sigma^2\right)$.

$$M_X(s) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{sx}e^{-(x-\mu)^2/2\sigma^2}dx.$$

Let $w = \dfrac{x-\mu}{\sigma} \Rightarrow x = \sigma w + \mu,\ dx = \sigma dw.$ Then

$$\begin{aligned} M_X(s) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{s(\sigma w+\mu)}e^{-w^2/2}dw \\ &= e^{s\mu}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(w^2-2\sigma s w)}dw \\ &= e^{s\mu}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left((w-\sigma s)^2-\sigma^2 s^2\right)}dw \\ &= e^{s\mu+\sigma^2 s^2/2}\cdot\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(w-\sigma s)^2}dw. \end{aligned}$$

Let $v = w - \sigma s,\ dv = dw$ to get

$$M_X(s) = e^{s\mu+\sigma^2 s^2/2} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-v^2/2} dv.$$

But,

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-v^2/2} dv = 1$$

since it is a density function for a mean zero, unit variance random variable (standard normal). Thus,

$$M_X(s) = e^{s\mu+\sigma^2 s^2/2}.$$

Note: For the normal case we have

$$M_X'(s) = \left(\mu + \sigma^2 s\right) e^{s\mu+\sigma^2 s^2/2}$$

and

$$M_X'(0) = \mu = E(X).$$

The *mgf* for a random variable may not exist for any s since some random variables such as the Cauchy do not have finite moments. X is Cauchy if it has *pdf*

$$f(x) = \frac{\alpha}{\pi\left(\alpha^2 + x^2\right)},\ x \in \mathbf{R},\ \alpha > 0.$$

However, the characterisitc function for a random variable always exists.

9.7 Characteristic Functions

Definition: Let X be any random variable. The *characteristic function* (*cf*) of X is given by

$$\Phi_X(\omega) = E\left(e^{i\omega X}\right).$$

Now

$$\Phi_X : \mathbf{R} \mapsto \mathbf{C}$$

by the rule

$$\Phi_X(\omega) = E(\cos\omega X + i \sin\omega X) = E(\cos\omega X) + i(\sin\omega X).$$

$\Phi_X(\omega)$ is defined $\forall\ \omega \in \mathbf{R}$.

For X discrete, the *cf* of X is

$$\Phi_X(\omega) = \sum_k e^{i\omega x_k} P(X = x_k).$$

For X continuous, the *cf* of X is

$$\Phi_X(\omega) = \int_{-\infty}^{\infty} e^{i\omega x} f_X(x) dx.$$

Note: The *cf* in the continuous case is related to the Fourier transform. In fact, we can use the inversion formula of the Fourier transform to conclude

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_X(\omega) e^{-i\omega x} d\omega.$$

One can also relate the moments of a random variable (when they exist) to its *cf* as was done for the *mgf*.

9.8 Special Moment Functions

Definition: If X is a random variable taking integer values, then its *moment function* is given by

$$\Gamma(z) = E\left(z^X\right) = \sum_i p_i z^i$$

where $p_i = P(X = i)$.

We compute

$$\Gamma'(z) = \frac{d}{dz}\left(\sum_i p_i z^i\right) = \sum_i i p_i z^{i-1}.$$

So

$$\Gamma'(1) = \sum_i i p_i = E(X).$$

Note: Changing the order of differentiation and summation (as was done above) is okay as long as $|z| <$ radius of convergence for $\sum_i p_i z^i$.

If we continue to differentiate we get

$$\Gamma^{(k)} = E\left[X(X-1)\cdots(X-k+1)\right].$$

Special Case: If X is discrete taking values $0, 1, 2, \ldots$ the *probability generating function* of X is the function

$$G_X(s) = E\left(s^X\right), \; s \in \mathbf{R},$$

or

$$G_X(s) = \sum_{i=0}^{\infty} s^i P(X = i).$$

This function is used extensively in characterizing random walks and branching processes.

9.9 Applications of Characteristic Functions

In addition to providing a moment theorem as with the *mgf*, the *cf* can aid us in finding the density function of $Y = g(X)$.

Recall

$$\Phi_X(\omega) = \int_{-\infty}^{\infty} e^{i\omega x} f_X(x)dx = E\left(e^{i\omega X}\right).$$

Let $Y = g(X)$. Then,

$$\Phi_Y(\omega) = \int_{-\infty}^{\infty} e^{i\omega y} f_Y(y)dy = E\left(e^{i\omega Y}\right) = E\left(e^{i\omega g(X)}\right).$$

So

$$\Phi_Y(\omega) = \int_{-\infty}^{\infty} e^{i\omega g(x)} f_X(x)dx.$$

If we can write

$$\int_{-\infty}^{\infty} e^{i\omega g(x)} f_X(x)dx$$

as

$$\int_{-\infty}^{\infty} e^{i\omega y} h(y)dy$$

then $f_Y(y) = h(y)$.

Example: Suppose $X \sim N\left(0, \sigma^2\right)$. Let $Y = \alpha X^2, \; \alpha \in \mathbf{R}, \; \alpha \neq 0$. Then

$$\begin{aligned}
\Phi_Y(\omega) &= \int_{-\infty}^{\infty} e^{i\omega\alpha x^2} f_X(x)dx = \int_{-\infty}^{\infty} e^{i\omega\alpha x^2} \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} dx \\
&= \frac{2}{\sqrt{2\pi}\sigma} \int_0^{\infty} e^{i\omega\alpha x^2} e^{-x^2/2\sigma^2} dx.
\end{aligned}$$

Let $y = \alpha x^2,\ dy = 2\alpha x dx = 2\sqrt{\alpha y}dx$. Then

$$\Phi_Y(\omega) = \frac{2}{\sqrt{2\pi}\sigma}\int_0^\infty e^{i\omega y}e^{-y/2\alpha\sigma^2}\frac{1}{2\sqrt{\alpha y}}dy$$

or

$$\Phi_Y(\omega) = \int_{-\infty}^\infty e^{i\omega y}\frac{1}{\sigma\sqrt{2\pi\alpha y}}e^{-y/2\alpha\sigma^2}U(y)dy$$

where,

$$U(y) = \begin{cases} 1, & y \geq 0, \\ 0, & \text{else.} \end{cases}$$

Thus,

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi\alpha y}}e^{-y/2\alpha\sigma^2}U(y).$$