Chapter - 4: Expectation and Moments

Aveek Dutta
Assistant Professor
Department of Electrical and Computer Engineering
University at Albany
Fall 2019

Mean - two views

- Mean is the the number ***z*** that minimizes the summed distance-square distance from all the points in the set.

$$D^2 = \sum_{i=1}^{N} (z - x_i)^2$$
$$\frac{dD^2}{dz} = 2Nz - 2\sum_{i=1}^{N} x_i = 0$$
$$z = \mu_s = \frac{1}{N}\sum_{i=1}^{N} x_i$$

- If an experiment is repeated ***N*** times with the rv. ***X*** taking ***M*** distinct values ***x****i* , ***n****i* times, then for large ***N*** the average is given by

$$\mu_X \simeq \frac{1}{N}\sum_{k=1}^{N} x^{(k)}$$
$$= \frac{1}{N}\sum_{i=1}^{M} n_i x_i = \sum_{i=1}^{M} x_i \left(\frac{n_i}{N}\right)$$
$$= \sum_{i=1}^{M} x_i P\left[X = x_i\right]$$

PMF of x_i

- Variance is the square of standard deviation

$$\sigma_s = \left[\frac{1}{N}\sum_{i=1}^{N} (x_i - \mu_s)^2\right]^{1/2}$$

- **Definition:** The expected value of a discrete rv ***X*** taking on values x_i with PMF $P_X(x_i)$

$$E[X] = \sum_i x_i P_X(x_i)$$

- For rv with pdf $f_X(x)$

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

Expected value of Gaussian

Let $X\colon N(\mu,\sigma^2)$. The expected value of X is

$$E[X] = \int_{-\infty}^{\infty} x \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 \right) \right) dx.$$

Let $z \triangleq (x-\mu)/\sigma$. Then

$$E[X] = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-\frac{1}{2}z^2} dz + \mu \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz \right).$$

The first term is zero because the integrand is odd, and the second term is μ because the term in parentheses is $P[Z \leq \infty]$ for $Z\colon N(0,1)$. Hence

$$E[X] = \mu \qquad \text{for } X\colon N(\mu,\sigma^2).$$

Theorem

Riemann Sum approximation of an integral
$\int_a^b f(x)dx = \lim_{\|\Delta x\| \to 0} \sum_{i=1}^{n} f\left(x_i^*\right) \Delta x_i$

- The expected value of ***Y=g(x)*** is given by

$$E[Y] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

- Proof: if $x_j^{(rj)}$ are roots of $y_j - g(x) = 0$, then

$$\{y_j < Y \leq y_j + \Delta y_j\} = \bigcup_{k=1}^{r_j} \left\{x_j^{(k)} < X \leq x_j^{(k)} + \Delta x_j^{(k)}\right\}$$

m partitions along Y axis

$$E[Y] = \int_{-\infty}^{\infty} y f_Y(y) dy = \sum_{j=1}^{m} y_j f_Y(y_j) \Delta y_j$$

(Reimann sum approximation of integrals)

$$\simeq \sum_{j=1}^{m} y_j P\left[y_j < Y \leq y_j + \Delta y_j\right]$$

$$= \sum_{j=1}^{m} \sum_{k=1}^{r_j} g\left(x_j^{(k)}\right) P\left[x_j^{(k)} < X \leq x_j^{(k)} + \Delta x_j^{(k)}\right]$$

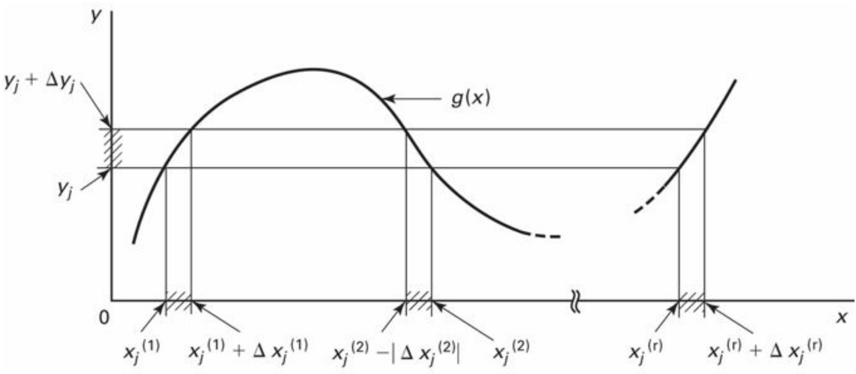
All terms are unique in the sum

$$\simeq \sum_{i=1}^{n} g(x_i) P\left[x_i < X \leq x_i + \Delta x_i\right]$$

$$= \sum_{i=1}^{n} g(x_j) f_X(x_j) \Delta x_j$$

as $\Delta y, \Delta x \to 0$

$$E[Y] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$



- For discrete rv. ***X***

$$E[Y] = \sum_i g\left(x_i\right) P_X\left(x_i\right)$$

- For functions of multiple variables ***z=g(X,Y)***

$$E[Z] = \int_{-\infty}^{\infty} z f_Z(z) dz$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{XY}(x, y) dx dy$$

What happens if g(x,y) = x?

Linearity of E[ ] operator

- Linear over functions of rv ***X, $g_i(x)$***

$$E\left[\sum_{i=1}^{N} g_i(X)\right] = \int_{-\infty}^{\infty} \left(\sum_{i=1}^{N} g_i(X)\right) f_X(x)dx$$
$$= \sum_{i=1}^{N} \int_{-\infty}^{\infty} g_i(X) f_X(x)dx = \sum_{i=1}^{N} E\left[g_i(X)\right]$$

- Linear over addition of multiple variables as well

$$\begin{aligned} E[X+Y] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)f_{XY}(x,y)dxdy \\ &= \int_{-\infty}^{\infty} x\left(\int_{-\infty}^{\infty} f_{XY}(x,y)dy\right)dx + \int_{-\infty}^{\infty} y\left(\int_{-\infty}^{\infty} f_{XY}(x,y)dx\right)dy \\ &= E[X]+E[Y] \end{aligned}$$

- Linear over sum of N rv, ***$X_1, X_2, X_3, \ldots$***

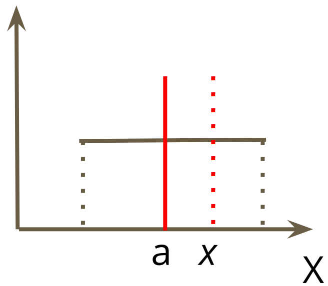
$$E\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} E\left[X_i\right]$$

Conditional Expectation (Avg of a subset)

Using Bayes' rule

Consider a continuous r.v. X and the event $B \triangleq \{X \geq a\}$.

$$F_{X|B}(x|X \geq a) = \begin{cases} 0, & x < a \\ \dfrac{F_X(x) - F_X(a)}{1 - F_X(a)}, & x \geq a. \end{cases}$$



Hence

$$f_{X|B}(x|X \geq a) = \begin{cases} 0, & x < a \\ \dfrac{f_X(x)}{1 - F_X(a)}, & x \geq a. \end{cases}$$

and

$$E[X|X \geq a] = \frac{\displaystyle\int_a^{\infty} x f_X(x)\, dx}{\displaystyle\int_a^{\infty} f_X(x)\, dx} = \frac{\int_{65}^{100} x f_X(x) dx}{\int_{65}^{100} f_X(x) dx} = \frac{100^2 - 65^2}{2 \times (100 - 65)} = 82.5$$

Definition

- If X, Y are discrete

$$E\left[Y|X=x_i\right]=\sum_j y_j P_{Y|X}\left(y_j|x_i\right)$$

Conditional Probability that $\{Y=y_j\}$ occurs given $\{X=x_i\}$ has occurred

- We can derive a formula for ***E[Y]***

$$\begin{aligned} E[Y] &= \sum_j y_j P_Y\left(y_j\right) \\ &= \sum_j y_j \sum_i P_{X,Y}\left(x_i, y_j\right) \\ &= \sum_i \left[\sum_j y_j P_{Y|X}\left(y_j|x_i\right)\right] P_X\left(x_i\right) \\ &= \sum_i E\left[Y|X=x_i\right] P_X\left(x_i\right) \end{aligned}$$

Definition of total probability as a sum of disjoint conditionals of $\{X=x_i\}$. Note the change in index of the summation

- Similarly, for continuous rv, $E[Y|X=x]=\int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy$

$$\text{Since, } E[Y]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{XY}(x,y)dxdy$$

Nice trick to add a new rv ***y*** into the integrand by using the definition of marginal density

$$\begin{aligned} E[Y] &= \int_{-\infty}^{\infty} f_X(x)\left[\int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy\right]dx \\ &= \int_{-\infty}^{\infty} E[Y|X=x] f_X(x)dx \end{aligned}$$

Alternate method to calculate E[Y] in terms of *f(x)*

Properties of Conditional Expectation

1. $\boxed{E[Y] = E[E[Y|X]]}$

We know that,

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

Also, when Y=g(X), we can write

$$E[Y] = \int_{-\infty}^{\infty} g(x) f_X(x)dx$$

- Using the similarity between the two, we can write $E[Y] = E[E[Y|X]]$
- Inner expectation is with respect to *Y* and outer with respect to *X*

2. If *X* and *Y* are independent, then $\boxed{E[Y|X] = E[Y]}$

We know, $E[Y|X=x] = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy$

- But, $f_{XY}(x,y) = f_{Y|X}(y|x).\, f_X(x) = f_Y(y).\, f_X(x)$ if X and Y are independent
- Hence $E[Y|X=x] = \int_{-\infty}^{\infty} y f_Y(y)dy = E[Y]$

3. $\boxed{E[Z|X] = E[E[Z|X,Y]|X]}$

$$E[Z|X=x] = \int_{-\infty}^{\infty} z f_{Z|X}(z|x)dz = \int_{-\infty}^{\infty} z \frac{f_{ZX}(z,x)}{f_X(x)}dz = \int_{-\infty}^{\infty} z \frac{\int_{-\infty}^{\infty} f_{ZXY}(z,x,y)dy}{f_X(x)}dz = \int_{-\infty}^{\infty} z \frac{\int_{-\infty}^{\infty} f_{Z|XY}(z|x,y) f_{XY}(x,y)dy}{f_X(x)}dz$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} z f_{Z|X,Y}(z|x,y) f_{Y|X}(y|x)dzdy = \int_{-\infty}^{\infty} dy f_{Y|X}(y|x) \int_{-\infty}^{\infty} z f_{Z|X,Y}(z|x,y)dz = E[E[Z|X,Y]|X=x]$$

Example 4.2-3

Let X and Y be two zero-mean r.v.'s with joint density

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\left(-\frac{x^2+y^2-2\rho xy}{2\sigma^2(1-\rho^2)}\right) \quad |\rho| < 1. \tag{4.2-16}$$

We shall soon find out (Section 4.3) that the pdf in Equation 4.2-16 is a special case of the general joint Gaussian law for two r.v.'s. First we see that when $\rho \neq 0$, $f_{XY}(x,y) \neq f_X(x)f_Y(y)$; hence X and Y are not independent when $\rho \neq 0$. When $\rho = 0$, we can indeed write $f_{XY}(x,y) = f_X(x)f_Y(y)$ so that $\rho = 0$ implies independence. For the present, however, our unfamiliarity with the meaning of ρ (ρ is called the normalized *covariance* or *correlation coefficient*) is not important. From Equations 2.6-34 and 2.6-35 it is easy to show that X and Y are zero-mean Gaussian r.v.'s, that is,

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

However, the conditional expectation of Y given $X = x$ is not zero even though Y is a zero-mean r.v.! In fact from Equation 4.2-9:

$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi\sigma^2(1-\rho^2)}} \exp\left(-\frac{(y-\rho x)^2}{2\sigma^2(1-\rho^2)}\right). \tag{4.2-17}$$

Hence $f_{Y|X}(y|x)$ is Gaussian with mean ρx. Thus

$$\begin{aligned} E[Y|X=x] &= \int_{-\infty}^{\infty} y f_{Y|X}(y|x)\, dx \\ &= \rho x. \end{aligned} \tag{4.2-18}$$

Example 4.2-5

(Continuation of Example 4.2-4.) Let X_1, X_2, X_3 denote multinomial random variables. Then

$$P_X[X_1 = x_1, X_2 = x_2, X_3 = x_3] = \frac{n!}{x_1!x_2!x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3},$$

where $n = x_1 + x_2 + x_3$ and $p_1 + p_2 + p_3 = 1$. We wish to compute $E[X_1|X_1 + X_2 = y]$.

Solution As in the previous example, we need to compute $P[X_1 = x_1|X_1 + X_2 = y]$. We write

$$P[X_1 = x_1|X_1 + X_2 = y] = \frac{P[X_1 = x_1, X_1 + X_2 = y]}{P[X_1 + X_2 = y]}.$$

Note that for the multinomial, the event $\{\zeta\colon X_1(\zeta) + X_2(\zeta) = y\} \cap \{\zeta\colon X_1(\zeta) = x_1\}$ is identical to the event $\{\zeta\colon X_1(\zeta) = x_1, X_2(\zeta) = y - x_1, X_3(\zeta) = n - y\}$. Hence

$$\begin{aligned} P[X_1 = x_1|X_1 + X_2 = y] &= \frac{P[X_1 = x_1, X_2 = y - x_1, X_3 = n - y]}{P[X_3 = n - y]} \\ &= \frac{n!}{x_1!(y - x_1)!(n - y)!} p_1^{x_1} p_2^{y - x_1} p_3^{n-y} \\ &\quad \div \frac{n!}{(n-y)!y!} p_3^{n-y} (1 - p_3)^y \\ &= \binom{y}{x_1} p_1^{x_1} p_2^{y - x_1} (p_1 + p_2)^{-y}. \end{aligned}$$

Finally, using

$$E[X_1|X_1 + X_2 = y] = \sum_{x_1} x_1 P[X_1 = x_1|X_1 + X_2 = y]$$

we obtain that

$$E[X_1|X_1 + X_2 = y] = y\frac{p_1}{p_1 + p_2}.$$

We leave it to the reader to compute that

$$E[X_2|X_1 + X_2 = y] = y\frac{p_2}{p_1 + p_2}.$$

Moments

- The r[th] Moments of a RV ***X*** is defined as,,

$$m_r = E\left[X^r\right] = \int_{-\infty}^{\infty} x^r f_X(x)dx, \qquad \text{where } r = 0, 1, 2, 3, \ldots$$

$$m_r = \sum_i x_i^r P_X\left(x_i\right)$$

- The r[th] Central Moments are defined as,

$$c_r = E\left[(X-\mu)^r\right] \qquad \text{where } r = 0, 1, 2, 3, \ldots$$

$$c_r = \sum_i \left(x_i - \mu\right)^r P_X\left(x_i\right) \xrightarrow{\text{Generalization}} c_r = \sum_{i=0}^{r} \binom{r}{i} (-1)^i \mu^i m_{r-i}$$

- An important derivation

$$\sigma^2 = E\left[[X-\mu]^2\right] = E\left[X^2\right] - E[2\mu X] + E\left[\mu^2\right]$$

$$E\left[X^2\right] - 2\mu E[X] + \mu^2$$

$$E\left[X^2\right] - \mu^2$$

Joint Moments

- The ij^th moments of RV ***X*** and ***Y*** is defined as,

$$m_{ij} = E\left[X^i Y^j\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^i y^j f_{XY}(x,y)dxdy$$

$$m_{ij} = \sum_l \sum_m x_l^i y_m^j P_{X,Y}\left(x_l, y_m\right)$$

- The ij^th central moments of RV ***X*** and ***Y*** is defined as,
  - The order of the moment is ***i+j***

$$c_{ij} = E\left[(X-\overline{X})^i (Y-\bar{Y})^j\right]$$

$$m_{02} = E\left[Y^2\right] \qquad c_{02} = E\left[(Y-\bar{Y})^2\right]$$

$$m_{20} = E\left[X^2\right] \qquad c_{20} = E\left[(X-\bar{X})^2\right]$$

$$m_{11} = E[XY] \qquad c_{11} = E[(X-\bar{X})(Y-\bar{Y})] = E[XY] - \bar{X}\bar{Y} \triangleq \mathrm{Cov}[X,Y]$$

- **Correlation coefficient is defined by** $\rho \triangleq \frac{c_{11}}{\sqrt{c_{20}c_{02}}}$

Example 4.3-4 - Self Study

- Application of Cov(X,Y) in Linear regression

Chebyshev Inequality

- **Theorem:** Let ***X*** a rv with mean $\overline{X}$ and variance σ^2. Then for any δ

$$P[|X - \overline{X}| \geq \delta] \leq \frac{\sigma^2}{\delta^2}$$

- **Proof:**

$$\sigma^2 \triangleq \int_{-\infty}^{\infty} (x - \bar{X})^2 f_X(x)dx \geq \int_{|x-\bar{X}|\geq\delta} (x - \bar{X})^2 f_X(x)dx$$
$$\geq \delta^2 \int_{|x-\bar{X}|\geq\delta} f_X(x)dx$$
$$\geq \delta^2 P[|X - \bar{X}| \geq \delta]$$

Since, $\{|X - \overline{X}| \geq \delta\} \cup \{|X - \overline{X}| < \delta\} = \Omega$

Thus, it follows,

$$P[|X - \overline{X}| < \delta] \geq 1 - \frac{\sigma^2}{\delta^2}$$

Markov Inequality

- **Theorem:** Consider a rv ***X*** with non-negative pdf, $f_X(x) = 0$, ***for*** $x<0$, the Markov inequality applies as

$$P[X \geq \delta] \leq \frac{E[X]}{\delta}$$

- **Proof:**

$$E[X] = \int_0^\infty x f_X(x) dx \geq \int_\delta^\infty x f_X(x) dx \geq \delta \int_\delta^\infty f_X(x) dx$$
$$\geq \delta P[X \geq \delta]$$

Schwarz Inequality

- **Theorem:** For rv ***X*** and ***Y*** $\text{Cov}^2(X,Y) \leq E\left[(X-\bar{X})^2\right] E\left[(Y-\bar{Y})^2\right]$

- **Proof:** Consider the non-negative expression

$$E\left[(\lambda(X-\bar{X})-(Y-\bar{Y}))^2\right] \geq 0$$

Expanding the expression that is quadratic in λ

$$Q(\lambda) = \lambda^2 c_{20} + c_{02} - 2\lambda c_{11} \geq 0$$

The quadratic equation to have at least one real root, the discriminant satisfies,

$$\left(\frac{c_{11}}{c_{20}}\right)^2 - \frac{c_{02}}{c_{20}} \leq 0 \longrightarrow c_{11}^2 \leq c_{02} \,.\, c_{20} \to \left|\frac{\text{Cov}(X,Y)}{\sigma_X^2 \,.\, \sigma_Y^2}\right| \leq 1$$

(Weak) Law of Large Numbers

- LLN defines the conditions under which the ***sample mean*** converges to ***ensemble mean***.
- If $X_1, X_2, \dots X_n$ be iid rv, then the sample mean estimator is given by

$$\hat{\mu}_n \triangleq \frac{1}{n}\sum_{i=1}^{n} X_i$$

- Chebyshev inequality is used to show that $\hat{\mu}_n$ is a perfect estimator of μ_X

$$E\left[\hat{\mu}_n\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[X_i\right] = \frac{1}{n} n\mu_X = \mu_X$$

$$\mathrm{Var}[\hat{\mu}_n] = \frac{1}{n^2}\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \left(\frac{1}{n^2}\right) n\sigma_X^2 = \frac{1}{n}\sigma_X^2$$

- Therefore, from Chebyshev inequality we get,

$$P\left[\left|\hat{\mu}_n - \mu_X\right| \geq \delta\right] \leq \sigma_X^2 / n\delta^2$$

$$\boxed{\lim_{n\to\infty} P\left[\left|\hat{\mu}_n - \mu_X\right| \geq \delta\right] = 0}$$

Moment-Generating Functions

- MGF is used to completely characterize a PDF (similar to Laplace transform)
- If ***t*** is a complex variable, the MGF for pdfs is defined as

$$\theta(t) \triangleq E\left[e^{tX}\right] = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx$$

$$\theta(t) = \sum_i e^{tx_i} P_X(x_i)$$

- Expanding the exponent

$$E\left[e^{tX}\right] = E\left[1 + tX + \frac{(tX)^2}{2!} + \ldots + \frac{(tX)^n}{n!} + \ldots\right] = 1 + t\mu + \frac{t^2}{2!} m_2 + \ldots + \frac{t^n}{n!} m_n + \ldots$$

- If the moments M(t) exists, then it can be computed as

$$m_k = M^{(k)}(0) \triangleq \frac{d^k}{dt^k}(M(t))\Big|_{t=0} \quad \text{for } k = 1, 2, \ldots$$

Chernoff Bound

- Upper bound on the tail probability $P[X \geq a]$
- **Proof:**

$$P[X \geq a] = \int_a^\infty f_X(x)dx$$
$$= \int_{-\infty}^\infty f_X(x)u(x-a)dx$$

Noting that $u(x-a) \leq e^{t(x-a)}$ for any $t > 0$ we can write,

$$P[X \geq a] \leq \int_{-\infty}^\infty f_X(x)e^{t(x-a)}dx = e^{-at}M_X(t)$$

- The tightest bound is found when the above expression wrt ***t*.**

Characteristics Functions

- Characteristics functions are analogous to Fourier transform of pdf $f_X(x)$ by replacing t with $j\omega$

$$\Phi_X(\omega) \triangleq E\left[e^{j\omega X}\right] = \int_{-\infty}^{\infty} f_X(x) e^{j\omega x} dx$$

INVERSION →

$$f_Z(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_Z(\omega) e^{-j\omega z} d\omega$$

Discrete form $\Phi_X(\omega) = \sum_i e^{j\omega x_i} P_X(x_i)$

- It can be used to calculate the convolution product for sums of rv **Z = X+Y**

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X_1}(x) f_{X_2}(z-x) dx$$

$$\Phi_Z(\omega) = \int_{-\infty}^{\infty} e^{j\omega z} \left[\int_{-\infty}^{\infty} f_{X_1}(x) f_{X_2}(z-x) dx\right] dz = \int_{-\infty}^{\infty} f_{X_1}(x) \int_{-\infty}^{\infty} f_{X_2}(z-x) e^{j\omega z} dx dz$$

- Change the variable $\alpha = z - x$ we obtain

$$\Phi_Z(\omega) = \Phi_{X_1}(\omega) \Phi_{X_2}(\omega)$$

Joint Characteristics Functions

- Similar to characteristic functions of single rv, the joint CF is given by

$$\Phi_{X_1 \ldots X_N}(\omega_1, \omega_2, \ldots, \omega_N) = E\left[\exp\left(j \sum_{i=1}^{N} \omega_i X_i\right)\right]$$

- The pdf is obtained by the inverse transform

$$f_{X_1 \ldots X_N}(x_1, \ldots, x_N) = \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \Phi_{X_1 \ldots X_N}(\omega_1, \ldots, \omega_N) \times \exp\left(-j \sum_{i=1}^{N} \omega_i x_i\right) d\omega_1 d\omega_2 \ldots d\omega_N$$

- **See example 4.7-8**

Central Limit Theorem

- **Theorem 1:** Let $X_1, X_2, \ldots X_n$ be n mutually independent (scalar) random variables with CDF $F_{X_1}(x_1), F_{X_2}(x_2), \ldots F_{X_n}(x_n)$ such that $\mu_{X_k} = 0 \text{ and } \mathrm{Var}(X_k) = \sigma_k^2$ and let $s_n \triangleq \sigma_1^2 + \cdots + \sigma_n^2$. Then if for given $\epsilon > 0, \text{ and large } n, \sigma_k < \epsilon s_n$, the normalized sum $Z_n \triangleq (X_1 + \ldots + X_n)/s_n$ converges to the **Standard Normal CDF**. That is $\lim_{n\to\infty} F_{Z_n}(z) = 1/2 + \mathrm{erf}(z)$

- **Proof for a special case:** if $\mu_{X_i} = 0 \text{ and } \mathrm{Var}(X_i) = 1$, then $Z_n \triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$ tends to Normal in the sense that its CF, satisfies

$$\lim_{n\to\infty} \Phi_{Z_n}(\omega) = e^{-\frac{1}{2}\omega^2}$$

General Case

- Let $Z_i \triangleq (X_i - \mu)/\sigma$ and $\varphi_Z(v) \triangleq E\left[e^{j\omega Z_i}\right]$ then $\boldsymbol{Z_i}$ has zero mean, unity variance.

$$
\begin{aligned}
\varphi_{Y_n}(\omega) &:= \mathrm{E}\left[e^{j\omega Y_n}\right] \\
&= \mathrm{E}\left[\exp\left(j\frac{\omega}{\sqrt{n}}\sum_{i=1}^{n} Z_i\right)\right] \\
&= \mathrm{E}\left[\prod_{i=1}^{n} \exp\left(j\frac{\omega}{\sqrt{n}} Z_i\right)\right] \\
&= \prod_{i=1}^{n} \mathrm{E}\left[\exp\left(j\frac{\omega}{\sqrt{n}} Z_i\right)\right] \\
&= \prod_{i=1}^{n} \varphi_Z\left(\frac{\omega}{\sqrt{n}}\right) \\
&= \varphi_Z\left(\frac{\omega}{\sqrt{n}}\right)^n
\end{aligned}
$$

For any complex ξ, we can write $e^{\xi} = 1 + \xi + \frac{1}{2}\xi^2 + R(\xi)$

Thus, $\varphi_Z\left(\frac{\omega}{\sqrt{n}}\right) = \mathrm{E}\left[e^{j(\omega/\sqrt{n})Z_i}\right]$

$$
= \mathrm{E}\left[1 + j\frac{\omega}{\sqrt{n}}Z_i + \frac{1}{2}\left(j\frac{\omega}{\sqrt{n}}Z_i\right)^2 + R\left(j\frac{\omega}{\sqrt{n}}Z_i\right)\right]
$$

Since, $\boldsymbol{Z_i}$ has zero mean, unity variance

$$
\varphi_Z\left(\frac{\omega}{\sqrt{n}}\right) = 1 - \frac{1}{2}\cdot\frac{\omega^2}{n} + \mathrm{E}\left[R\left(j\frac{\omega}{\sqrt{n}}Z_i\right)\right]
$$

The remainder term goes to zero for higher orders of *n*

$$
\varphi_Z\left(\frac{\omega}{\sqrt{n}}\right) \approx 1 - \frac{\omega^2/2}{n} \longrightarrow \varphi_{Y_n}(\omega) = \varphi_Z\left(\frac{\omega}{\sqrt{n}}\right)^n \approx \left(1 - \frac{\omega^2/2}{n}\right)^n \to e^{-\omega^2/2}
$$

$$e^{\ln(1+\frac{x}{n})^n} = e^{n\ln(1+\frac{x}{n})}$$

$$\lim_{n\to+\infty}(1+\frac{x}{n})^n = \lim_{n\to+\infty} e^{n\ln(1+\frac{x}{n})}$$

$$= e^{\lim_{n\to+\infty} n\ln(1+\frac{x}{n})} = e^{\lim_{n\to+\infty}\frac{\ln(1+\frac{x}{n})}{\frac{1}{n}}}$$

Apply L'Hopital's Rule:

$$= e^{\lim_{n\to+\infty}\frac{(\frac{-x}{n^2})\frac{1}{1+\frac{x}{n}}}{-\frac{1}{n^2}}} = e^{\lim_{n\to+\infty}\frac{x}{1+\frac{x}{n}}} = e^x$$

Therefore,

$$(1+\frac{x}{n})^n \to e^x$$