Lecture 6: Moments of Random Variables

- Characterization of RVs using averages: Expectation Operator
- Average or mean value of a set of data is

\[ \hat{\mu}_x = \frac{1}{N} \sum_{i=1}^{N} x_i \]  

(1)

- Average of data \( x_i \) : Center of Gravity (CG) of the set
  - The number which is closest in distance to all other points in the set.
  - Minimize the distance given below with respect to \( z \) to yield \( \hat{\mu}_x \)

\[ D^2 = \sum_{i=1}^{N} (z - x_i)^2 \]  

(2)

- Average value provides an indication of the most likely value of the set
Standard Deviation

- To determine the spread or deviation of data from the average: the standard deviation $\sigma_x$ of the set:

$$\hat{\sigma}_x = \left[ \frac{1}{N} \sum_{i=1}^{N} (x_i - \hat{\mu}_x)^2 \right]^{1/2}$$

Interpretation using Probability Measure

- W.R.T. a probability space $(\Omega, F, P)$:
- Consider discrete RV $x$ that can take on $M$ distinct values $x_1, x_2, \ldots x_M$.
- In $N$ trials of the experiment, assume $x_1$ occurs $n_1$ times, $x_2$ occurs $n_2$ ... $x_i$ occurs $n_i$ times.
\[ \bar{X} = \frac{1}{N} \sum_{i=1}^{M} n_i x_i = \sum_{i=1}^{M} \frac{n_i}{N} x_i = \sum_{i=1}^{M} x_i P[X = x_i] \]  

- Expected or Average value of a discrete rv \( X \) taking on values \( x_i \) with PMF \( P_X(x_i) \) is:

\[ E[X] = \sum_{i} x_i P[X = x_i] = \sum_{i} x_i P_X(x_i) \]  

- Cont. RV \( X \): \[ E[X] = \int_{-\infty}^{\infty} x f_X(x) \, dx \]
Expected value Contd.

- Function of a RV $Y = g(X)$
- $E[Y] = E[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx = \int_y yf_Y(y)$
- $X$ discrete: $E[Y] = \sum_i g(x_i)P_X(x_i)$

Examples:
- $X : N(\mu, \sigma^2)$: Use transformation $z = \frac{(x-\mu)}{\sigma}$ to show that $E[X] = \mu$
- $X : \lambda e^{-\lambda x}$
- $E[X] = \lambda \int_0^\infty xe^{-\lambda x} dx$ Integrate by Parts

\[
E[X] = \lambda \left[ \frac{x}{-\lambda} e^{-\lambda x} \right]_0^\infty + \int_0^\infty \lambda e^{-\lambda x} dx
\]

\[
= \frac{1}{\lambda}
\]
• $E \left[ \sum_{i=1}^{N} g_i(X) \right] = \sum_{i=1}^{N} E[g_i(X)]$

• Function of two Random Variables: $Z = g(X, Y)$

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

• $E[Z] = E[g(x, y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{XY}(x, y) dx dy$

• Ex: If $Z = X + Y$

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

**NOTE:** Independence is not required in showing that $E[X + Y] = E[X] + E[Y]$.
In general: $E(aX + bY) = aE(X) + bE(Y)$
Conditional Expectations

- To determine averages of a subset of RV’s that are conditioned on an event $B$: $E(X|B) = \int_{-\infty}^{\infty} x f_{X|B}(x|B) dx$

- For discrete $X$: $E(X|B) = \sum_{i=1}^{\infty} x_i P_{X|B}(x_i|B)$

**Example:** $B \equiv [X \geq a]$

- $F_{X|B}(x|X \geq a) = 0, \ x < a$
  
  \[ = \frac{F_X(x) - F_X(a)}{1 - F_X(a)}, \ x \geq a \]

- $f_{X|B}(x|X \geq a) = 0, \ x < a$

  \[ = \frac{f_X(x)}{1 - F_X(a)} \ x \geq a \]

Therefore, $E[X|X \geq a] = \frac{\int_{a}^{\infty} x f_X(x) dx}{\int_{a}^{\infty} f_X(x) dx}$
Conditional Expectations Contd.

- Conditional expectations often occur w.r.t. RV’s that are related to each other.
- If X, Y : two discrete random variables, with joint PMF \( P_{X,Y}(x_i, y_j) \)
- Conditional expectation of \( Y \) given \( X = x_i \) is \( E[Y | X = x_i] \) is:

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

(6)

- \( P_{Y|X}(y_j | x_i) \) is conditional probability that \( Y = y_j \) given \( X = x_i \) has occurred and

\[
P_{Y|X}(y_j | x_i) = \frac{P_{XY}(x_i, y_j)}{P_X(x_i)}
\]

(7)

- Conditional expectation : \( E[Y | X = x] \) is a random variable, (a function of rv \( x \)).
• $E[Y]$ can be determined as:

\[
E[Y] = \sum_x E[Y | X = x] P[X = x] \\
= \sum_x \sum_y y P[Y = y | X = x] P[X = x] \\
= \sum_x \sum_y y P[Y = y, X = x] \\
= \sum_y \sum_x P[Y = y, X = x] \\
= \sum_y y P[Y = y]
\]
\[ E[Y] = E[E(Y|X)] \]

- If \( Y = g(X) \), where \( X \) is a discrete RV
  \[
  E[Y] = \sum_i g(x_i)P_X(x_i) = E[g(x)]
  \] (8)

- If \( E[Y|X] = g(X) \),
- \( E[E(Y|X)] = \sum_x E[Y|X]f_X(x) = E[Y] \)
- For continuous Rvs, : \( E[Y] = \int_{-\infty}^{\infty} E[Y|X=x]f_X(x)dx \)
- Note: If \( X, Y \) are independent, \( E[Y|X] = E[Y] \)

Moments of a RV

• If $k$ is a positive integer, the $k^{th}$ moment $m_k$ of $X$ is defined as:

$$\zeta_k = E[X^k]$$

• The $k^{th}$ central moment is:

$$k = E[(X - \mu_X)^k]$$ where $\mu_X = E[X]$.

• For $k = 2$, $m_2 = \sigma_X^2$ the variance of $X : Var[X]$.

• Show that: $\sigma_X^2 = E[X^2] - \mu_X^2$.

Note: Other notation for expectation operation: $E[X^k] = X^k$.

Under certain conditions possible to reconstruct pdf from knowledge of all of the moments of a RV.
Examples

• Bernoulli Variables
  
  \[ f_X(x) = p\delta(x - 1) + (1 - p)\delta(x) \]
  \[ E[X] = p \times 1 + 0 \times (1 - p) = p \]
  \[ E[X^2] = p \]
  
  Variance: \[ \sigma^2 = E[X^2] - E[X]^2 = p - p^2 = p(1 - p) = pq \]

• Binomial Variables \( X : f_X(k) \)
• Find \( E[X] \) and \( Var[X] \)
Joint Moments

- Consider two rvs X and Y possibly related in some way. \( Y = g(X) \)
  - Classical Problem: Infer or Estimate \( Y \) based on observations of \( X \)
  - If \( X, Y \) independent, information of \( X \) gives no information of \( Y \)
  - Ex: Linear Prediction: \( Y = aX + b \)
  - Observe \( X \), predict \( Y \) exactly with knowledge of \( a, b \).
  - In absence of explicit relations between \( X, Y \), calculate the joint moments of \( X, Y \) to infer the dependence features

- The \( ij^{th} \) joint moment of \( X \) and \( Y \) is:

\[
\zeta_{ij} = E[X^iY^j] = \int_x \int_y f_{XY}(x, y)x^iy^jdx dy
\]

- Discrete: \( \zeta_{ij} = E[X^iY^j] = \sum_x \sum_y P_{XY}(x, y)x^iy^j \)
Central Moments

The $ij^{th}$ central moment of $X$ and $Y$ is:

$$m_{ij} = E[(X - \bar{X})^i(Y - \bar{Y})^j]$$ (9)

The order of the moment is $i + j$.

Consider the case $i = 1, j = 1$,

$$m_{11} = E[(X - \bar{X})(Y - \bar{Y})] = E[XY] - \bar{X}\bar{Y}$$ (10)

- $m_{11}$: covariance of $X$ and $Y$, denoted as $Cov[X, Y]$.
- $\zeta_{11} = E[XY]$: correlation of $X$ and $Y$. 
Correlation Coefficient $\rho$

- Correlation coefficient: (Normalized covariance) between Rvs $(X,Y)$:

$$\rho_{XY} = \frac{m_{11}}{\sqrt{m_{20}m_{02}}}$$ \hspace{1cm} (11)

- $|\rho| \leq 1$

- If $\text{Cov}[X, Y] = 0$, then $\rho = 0$ and $X$ and $Y$ are said to be uncorrelated.
Properties of uncorrelated Random Variables

- If $X$ and $Y$ are uncorrelated $\sigma^2_{X+Y} = E[(X+Y)^2]-(E[X + Y])^2 = \sigma^2_X + \sigma^2_Y$
- If $X$ and $Y$ are independent, they are also uncorrelated. Proof:

$$E[XY] = \int_x \int_y xy f_X(x)f_Y(y) dx dy = E[X]E[Y]$$  \hspace{1cm} (12)


Independent random variables are uncorrelated; The reverse is not generally true.
Linear Prediction

• Problem: Predicting values of a rv $Y$ by observing the values of rv $X$.
• Data corresponding to measurements of $X$ are available.
• The model to be applied is:

\[ \hat{Y} = aX + b \] (13)

• Actual value of $Y$ affected by other sources independent of $X$, such as additive Gaussian noise.
• Denote the error $\epsilon = Y - \hat{Y}$
• Objective: Find $a$ and $b$ that minimize the mean square error, (MMSE):

\[ \hat{\epsilon}^2 = E[(Y - \hat{Y})^2] \]

• This problem termed optimum linear prediction or linear regression in statistics.
Linear Prediction: Solution

- Expand the mse after substituting $\hat{Y} = aX + b$

$$\hat{\epsilon}^2 = Y^2 - 2aXY - 2bY + 2abX + a^2X^2 + b^2$$  \hspace{1cm} (14)

$$\frac{\delta \hat{\epsilon}^2}{\delta a} = \frac{\delta \hat{\epsilon}^2}{\delta b} = 0$$  \hspace{1cm} (15)

- Denote optimal solutions as $a^0, b^0$,

$$a^0 = \frac{\text{Cov}(X, Y)}{\sigma^2_X} = \frac{\rho_{XY}\sigma_Y}{\sigma_X}$$  \hspace{1cm} (16)

$$b^0 = Y - \frac{\text{Cov}(X, Y)}{\sigma^2_X}X = Y - \rho_{XY}\frac{\sigma_Y}{\sigma_X}X$$  \hspace{1cm} (17)

- Substituting for $a^0$ and $b^0$, the best linear predictor of $Y$ is
\[
\hat{Y} - \bar{Y} = \rho_{XY} \frac{\sigma_Y}{\sigma_X} (X - \bar{X})
\] (18)

- It passes through the point \((\bar{X}, \bar{Y})\). (Averages)
Moment Generating Functions (MGF)

- MGF of a cont. rv \( X \) is defined in terms of a transformed variable \( t \), in general complex, as

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

- For \( X \) a discrete rv,

\[
\theta(t) = \sum_i e^{tx_i} P[X = x_i]
\]  \hspace{1cm} (19)

- Similar to Laplace transforms: \( f(x) \leftrightarrow F(t) \)

\[
-F(t) = \int f(x)e^{-tx} dx.
\]
Application of MGF

- MGF applied for estimating the moments in the absence of knowledge of $f_X(x)$.
- If $\zeta_k = E[X^k]$,

$$
\begin{align*}
\theta(t) &= E[e^{tX}] \\
&= E\left[1 + tX + \frac{tX^2}{2!} + \ldots + \frac{tX^n}{n!} + \ldots\right] \\
&= 1 + t\zeta_1 + \frac{t^2}{2!}\zeta_2 + \ldots + \frac{t^n}{n!}\zeta_n + ..
\end{align*}
$$

(21)

- Taking derivatives of $\theta(t)$ and setting $t = 0$: $\theta^k(t)|_{t=0} = \zeta_k \quad k = 1, 2, ...$
Example: Poisson RV: Find $\theta(t)$

$$P[X = n] = \frac{e^{-\lambda} \lambda^n}{n!}$$

$$\theta(t) = e^{-\lambda} \sum_{n=0}^{\infty} \frac{e^{tn} \lambda^n}{n!}$$

$$= e^{-\lambda} e^{\lambda e^t}$$

$$\theta(t) = e^{\lambda(e^t-1)}$$

Taking derivatives:

$$\theta'(t) = \lambda e^t e^{\lambda(e^t-1)}$$

$$\theta''(t) = (\lambda e^t)^2 e^{\lambda(e^t-1)} + \lambda e^t e^{\lambda(e^t-1)}$$

Substituting $t = 0$, $E[X] = \lambda$ and $E[X^2] = \lambda^2 + \lambda$

$$\sigma_X^2 = E[X^2] - E[X]^2 = \lambda$$
Example: Exponential RV: \( f_X(x) = \lambda e^{-\lambda x} \)

\[
\theta(t) = \int_0^\infty \lambda e^{-\lambda x} e^{tx} \, dx = \frac{\lambda}{\lambda - t}
\]

Moments:

\[
E[X] = \frac{\lambda}{(\lambda - t)^2} \bigg|_{t=0} = \frac{1}{\lambda}
\]

\[
E[X^2] = \frac{2\lambda}{(t - \lambda)^2} \bigg|_{t=0} = \frac{2}{\lambda^2}
\]
MGF of two random variables

\[ \theta_{XY}(t_1, t_2) = E \left[ e^{(t_1X + t_2Y)} \right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{t_1x + t_2y} f_{XY}(x, y) \, dx \, dy \]

• For discrete rvs: \((i, j)\) with \(\zeta_{ij} = E[X^i Y^j],

\[ \theta_{XY}(t_1, t_2) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{t_1^i t_2^j}{i! j!} \zeta_{ij} \]

• To obtain the joint moments, note that:
  • \(\zeta_{ln} = \theta_{XY}^{(l,n)}(0, 0)\)
  • \(\zeta_{1,0} = E[X] \quad \zeta_{0,1} = E[Y]\)
  • \(\zeta_{2,0} = E[X^2] \quad \zeta_{0,2} = E[Y^2]\)
  • \(\zeta_{1,1} = E[XY] = Cov[X, Y] + E[X]E[Y]\)
Consider $Z = X + Y$, where $X, Y$ are independent RVs,

$$
\theta_{X+Y}(t) = E[e^{t(X+Y)}] \\
= E[e^{tX}]E[e^{tY}] \\
\theta_{X+Y}(t) = \theta_X(t) \theta_Y(t)
$$

Example: $Y = \sum_{i=1}^N x_i$, where $x_i$ are iid rvs and $N$ is a rv. Find $E[Y]$ and $Var[Y]$ in terms of the moments of $X$ and $N$.

- Consider the conditional expectation $E[e^{t\sum_{i=1}^n x_i|N=n}]$

$$
\theta_{Y|N}(t) = E[e^{t\sum_{i=1}^n x_i|N=n}] \\
= (\theta_X(t))^n
$$

- To obtain $\theta_Y(t)$, take the expectation of $\theta_{Y|N}(t)$:

$$
\theta_Y(t) = E[\theta_{Y|N}(t)] = E[(\theta_X(t))^N]
$$
To compute $E[Y]$ and $Var[Y]$, obtain the first and second derivatives of $\theta_Y(t)$:

$$
\begin{align*}
\theta'_Y(t) &= E[N\theta_X(t)^{N-1}\theta'_X(t)] \\
E[Y] &= \theta'_Y(0) = E[N(\theta_X(0))^{N-1}\theta'_X(0)] \\
&= E[N(1)^{N-1}E[X]] = E[N]E[X]
\end{align*}
$$

$$
\begin{align*}
\theta''_Y(t) &= EN(N-1)\theta_X(t)^{N-2}\theta'_X(t)\theta'_X(t) + N(\theta_X(t))^{N-1}\theta''_X(t) \\
\theta''_Y(0) &= EN(N-1)E[X]^2 + E[N]E[X^2] \\
&= E[N]Var[X] + E[X]^2E[N^2] \\
&= E[N]Var[X] + E[X]^2Var[N]
\end{align*}
$$
Characteristic Functions of a RV

- MGF’s uniquely determine the PMF if all moments exist and are known.
- This implies that \( \theta(t) \) exists and is finite in some region about \( t = 0 \).
- Useful to isolate a region \( t = j\omega \) in the MGF.
- Resulting function is referred to as the characteristic function \( \Phi_X(\omega) \)

\[
\Phi_X(\omega) = E[e^{j\omega X}] = \int_x f_X(x)e^{j\omega x} dx
\]

- Other than a change of sign in the exponent, \( \Phi_X(\omega) \) : Fourier Transform of pdf \( f_X(x) \).
- Allows determination of \( f_X(x) \) may be obtained by inverse transform methods:
  \[
f_X(x) = \frac{1}{2\pi} \int_\omega \Phi_X(\omega)e^{-j\omega x} d\omega
\]
Charc. Functions Contd.

- For discrete and integer valued random variable \( X = 0, 1, \ldots, \infty \),

\[
\Phi_X(\omega) = \sum_{k=0}^{\infty} P[X = k] e^{j\omega k}
\]

which represents the FT of the sequence \( p_k = P[X = k] \).

- \( \Phi_X(\omega) \) is periodic in \( \omega \) with a period of \( 2\pi \), \((e^{jk\omega} = e^{jk(\omega+2\pi)})\).

- Example: Geometric RV \( X : p_k = pq^k \) Find \( \Phi_X(\omega) \)

\[
\Phi_X(\omega) = \sum_{k=0}^{\infty} p q^k e^{jk\omega}
\]

\[
= p \sum q e^{j\omega}^k
\]

\[
= \frac{p}{1 - q e^{j\omega}}
\]
Moment Estimation from $\Phi_X(\omega)$

\[
E[X^n] = \frac{1}{j^n} \frac{d^n \Phi_X(\omega)}{d\omega^n} \bigg|_{\omega=0}
\]

\[
\Phi_X(\omega) = \int f_X(x) \left[ 1 + j\omega x + \frac{(j\omega x)^2}{2!} + \ldots \right] dx
\]

\[
= 1 + j\omega E[X] \frac{(j\omega)^2}{2!} E[X^2] + \ldots
\]
Probability Generating Functions (PGF)

- For non-negative rvs: convenient to use the Laplace transform (cont rv) or Z transform (discrete rv).
- PGF $G_X(z)$ of a discrete rv $X$ is: $G_X(z) = E[Z^k] = \sum_{k=0}^{\infty} p_k z^k$
- Characteristic function of $X$ obtained by substituting $z = e^{j\omega}$ in the PGF
- Taking derivatives of $G_X(z)$:
  \[
  \frac{d}{dz} G_X(z) = \sum_{k=0}^{\infty} k p_k z^{k-1}
  \]
  \[
  \frac{d}{dz} G_X(0) = (1)p_1
  \]
  \[
  \frac{d^2}{dz^2} G_X(0) = (2)(1)p_2
  \]
  \[
  \frac{d^n}{dz^n} G_X(0) = n!p_n
  \]
\[ p_n = \frac{1}{n!} \frac{d^n}{dz^n} G_X(0) \]

- Hence the name Probability Generating Function.
The moments can also be estimated from the PGF by substituting $z = 1$ after taking the derivatives:

$$G'_X(1) = \sum_{k=0}^{\infty} kp_k = E[X]$$

$$G''_X(1) = \sum_{k=0}^{\infty} k(k-1)p_k$$

$$= \sum_{k=0}^{\infty} k^2 p_k - \sum_{k=0}^{\infty} kp_k$$

$$= E[X^2] - E[X]$$

$$\text{Var}[X] = G''_X(1) + G'_X(1) - (G'_X(1))^2$$
Laplace Transforms

For continuous non-negative random variables, the Laplace transform is considered:

\[ X^*(s) = \int_0^\infty f_X(x)e^{-sx} \, dx \]
\[ = E[e^{-sX}] \]
\[ E[X^n] = (-1)^n \frac{d^n}{ds^n}X^*(s)|_{s=0} \]