CL205: AI & DS

MB

Mani Bhushan,
Department of Chemical Engineering,
Indian Institute of Technology Bombay
Mumbai, India- 400076

mbhushan@iitb.ac.in

Ack: Prof. Sachin Patwardhan

Autumn 2024

Next Topic

Expectations

- Work in the sample space $\mathbb{R}$, i.e. on the real-line for Single random variable, or $\mathbb{R}^n$ for n-random variables.
- Short form: RV for random variable

Probability Distributions I

For a random variable $X(\omega)$,

- Probability (cumulative) distribution dunction $F_X(x) = P(X \leq x)$
- PMF or PDF
  - probability mass function (PMF), if X is discrete,
  - probability density function (PDF), if X is continuous.

Expectation of a Random Variable I

- Average value of X in a large number of repetitions of the experiment.
- The *expected value* or *mean* of a random variable X is defined as the integral for continuous RV

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

  with $f_X(x)$ being the probability density function.
- Also denoted as μ.

Expectation of a discrete RV I

- For a discrete random variable, the expectation can be written as a sum.
- Suppose X takes values x_i with probability α_i. Then,

$$E[X] = \sum_i \alpha_i x_i$$

Expected Value of a RV

- Mean $E[X]$ also denoted as μ.
- It is centre of gravity of probability mass function or probability density function.
- For discrete RV: $\sum_i \alpha_i (x_i - \mu) = 0$.
- For continuous RV:

$$\begin{aligned}\int_{-\infty}^{\infty} (x-\mu) f(x) dx &= \int_{-\infty}^{\infty} x f(x) dx - \int_{-\infty}^{\infty} \mu f(x) dx \\ &= 0\end{aligned}$$

Example 1: Expectation of a discrete random variable

- What is the expected value when you roll a die?

$$E[X] = 1(1/6) + 2(1/6) + ... + 6(1/6) = 3.5$$

- Note that 3.5 is not a value you get on a single roll.

Example 2: Expectation of a discrete random variable

- Probability that a drug works on $\{0, 1, 2, 3, 4\}$ out of four patients.

r	$P(X = r)$
0	0.008
1	0.076
2	0.265
3	0.411
4	0.240
	1.000

Example 2 (Cont)

- The expected number of patients who would be cured is

$$\begin{aligned} E[X] &= 0(0.008) + 1(0.076) + 2(0.265) \\ &\quad + 3(0.411) + 4(0.240) \\ &= 2.80 \end{aligned}$$

Therefore 2.8 out of 4 patients would be cured, on average.

Example 3: Expectation of a discrete random variable

- For the case of two coin tosses with a fair coin, if $X =$ number of heads, then $E[X]$ is

$$E[X] = 0(1/4) + 1(1/2) + 2(1/4) = 1$$

- If the probability of a head on a coin toss was $3/4$

$$p_X(x) = \begin{cases} (1/4)^2 & x = 0 \\ 2 \times \frac{1}{4} \times \frac{3}{4} & x = 1 \\ (3/4)^2 & x = 2 \\ 0 & \text{else} \end{cases}$$

Example 3, Coin toss (Continued)

- This is a Binomial RV with number of tosses $n = 2$ and probability of success in a single toss as $p = 3/4$.

$$
\begin{aligned}
E[X] &= 0\left(\frac{1}{4}\right)^2 + 1\left(2 \times \frac{1}{4} \times \frac{3}{4}\right)^2 + 2\left(\frac{3}{4}\right)^2 \\
&= \frac{24}{16} = 1.5
\end{aligned}
$$

Example 4.4d from Ross (Modified): Continuous RV

Q. You are waiting for a phone call after dinner (9 PM). The number of hours (X) after 9 PM until the call comes is a random variable with probability density function:

$$f(x) = \begin{cases} 1/1.5 & 0 < x < 1.5 \\ 0 & \text{otherwise} \end{cases}$$

Question: What is $E[X]$
Answer: $E[X] = \int_0^{1.5} (x/1.5) dx = 0.75$.
Note: The above density function is a uniform density function over $(0, 1.5)$ and the random variable is said to have a uniform distribution over $(0, 1.5)$.

Expected Value of a Transformation

- Given RV X and its PMF or PDF, compute expected value of some function of X say $g(X)$.
- Option 1:
  - $g(X)$ is also a RV: For a given $\omega \in S$, $x = X(\omega)$ is a number and $g(X(\omega)) = g(x)$ is also a number.
  - Find probability distribution of $g(X)$ and then compute its expectation.

Example 4.5b (Ross)

- The time (in hrs.) taken to repair an electrical breakdown in a factory is a uniform RV X over $(0, 1)$ i.e. its density function is:

$$f(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

  Penalty paid due to breakdown of duration x is x^3. What is the expected penalty due to a breakdown?

Example 4.5b(Cont. 1)

- Let $Y = X^3$ denote penalty. Compute its cumulative distribution function. For $0 < y < 1$:

$$\begin{aligned} F_Y(y) &= P(\{Y \le y\}) = P(\{X^3 \le y\}) \\ &= P(\{X \le y^{1/3}\}) = \int_0^{y^{1/3}} 1 dx = y^{1/3} \end{aligned}$$

Thus,

$$F_Y(y) = \begin{cases} 0, & y \le 0 \\ y^{1/3}, & 0 < y < 1 \\ 1, & 1 \le y \end{cases}$$

Example 4.5b(Cont. II)

- Density function of Y is:

$$f_Y(y) = \frac{d}{dy} F_Y(y)$$
$$= \begin{cases} 0, & y \leq 0 \\ \frac{1}{3} y^{-2/3}, & 0 < y < 1 \\ 0, & 1 \leq y \end{cases}$$

- $E[Y] = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_0^1 \frac{1}{3} y^{-2/3} dy = 1/4$

Expectation of a Function of RV

- Option 2: $E[g(X)]$ is a weighted average of the possible values $g(X)$ with the weight given to $g(x)$ equal to the probability (or probability density in continuous case) that $X = x$.
- Makes sense since $g(X)$ takes value $g(x)$ whenever X takes value x.

Expectation of a Function of RV: II

- For X discrete with PMF $p(x)$ and takes values x_i with probability α_i, for any real-valued function g:

$$E[g(X)] = \sum_i g(x_i)P(X = x_i) = \sum_i g(x_i)\alpha_i$$

- For X continuous with PDF $f(x)$, for any real-valued function g:

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

- For earlier example, $E[g(X)] = \int_0^1 x^3 dx = 1/4$ as before.

Special Cases of $g(X)$

- For constants a, b
$$E[aX + b] = aE[X] + b$$
for X continuous or discrete.
Show this.
- What happens for $a = 0$, expectation of a constant is the constant itself.
- $E[X]$ is called mean (μ) or the first moment of X.
- $E[X^n]$ is called the n^{th} moment of X.

$$E[X^n] = \begin{cases} \sum_{x_i} {x_i}^n \alpha_i & \text{X discrete} \\ \int_{-\infty}^{\infty} x^n f(x) dx & \text{X continuous} \end{cases}$$

Variance of a Random Variable

- Variance of a random variable X with mean μ is defined as:

$$\text{Var}(X) = E[(X-\mu)^2]$$

- $\text{Var}(X)$ also denoted as σ_X^2 or σ^2

Variance of a Discrete RV

$$
\begin{aligned}
\mathrm{Var}(X) &= \sum_i \left(x_i - E[X]\right)^2 \alpha_i \\
&= \sum_i \left(x_i^2 - 2x_i E[X] + (E[X])^2\right) \alpha_i \\
&= \sum_i x_i^2 \alpha_i - 2E[X] \sum_i x_i \alpha_i \\
&\quad + (E[X])^2 \sum_i \alpha_i \\
&= E[X^2] - 2\,(E[X])^2 + (E[X])^2 \\
&= E[X^2] - (E[X])^2 \\
\implies \mathrm{Var}(X) &= E[X^2] - \mu^2
\end{aligned}
$$

Variance of a Continuous RV

- The variance of a continuous RV X is

$$\begin{aligned}\mathrm{Var}(X) &= E[(X-\mu)^2] \\ &= \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)dx \\ &= E[X^2] - (E[X])^2\end{aligned}$$

Variance

- $\mathrm{Var}(X) = E[(X-\mu)^2]$ which turns out to be $\mathrm{Var}(X) = E[X^2] - (E[X])^2$ i.e.
- Variance of X is the expected value of square of X minus the square of the expected valued of X.
- Quantity $\sigma = \sqrt{\mathrm{Var}(X)}$ is called the *Standard Deviation* of X.
- Example: X is outcome when we roll a fair die i.e. $P(X = i) = 1/6; \quad i = 1, 2, ..., 6.$ Then $E[X^2] = \sum_{i=1}^{6} i^2 P(X = i) = 1^2(1/6) + 2^2(1/6) + ... + 6^2(1/6) = 91/6.$ Also, $E[X] = 7/2$ (calculate this). Thus $\mathrm{Var}(X) = 91/6 - (7/2)^2 = 35/12.$

Useful Relations Regarding Variances

- If $Y = aX + b$, a, b are constants, then

$$E[Y] = aE[X] + b = a\mu + b$$

What about variance of Y?

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

- Holds for both discrete and continuous RVs.

Special Cases for a and b

$\text{Var}(aX + b) = a^2\text{Var}(X)$

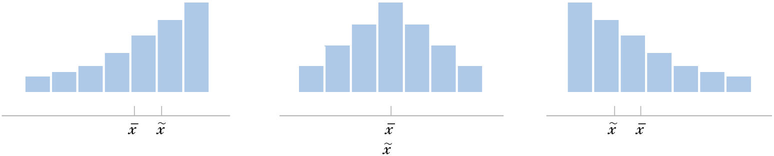
- a=0: $\text{Var}(b) = 0$.
- b=0: $\text{Var}(aX) = a^2\text{Var}(X)$; b does not affect variance.

Moments of a Random Variable

- n^{th} moment (about the origin) of a random variable X defined as: $E[X^n]$
- $\mu = E[X]$ is the first moment about the origin and indicates the average of values that X takes. It is the centre of gravity of the density function.
- $\text{Var}(X) = E[(X-\mu)^2]$ is the second moment about the mean and indicates the spread (variation) in the value that X will take around the mean.

Skewness

- Skewness=
$$\frac{E[(X-\mu)^3]}{(E[(X-\mu)^2])^{3/2}} = \frac{E[(X-\mu)^3]}{\sigma^3}$$
is the (normalized) third moment about the mean
- Indicates degree of asymmetry



Negative or left skew *(a)* | Symmetric *(b)* | Positive or right skew *(c)*

Kurtosis

- Kurtosis (degree of peakedness)=

$$\frac{E[(X-\mu)^4]}{\sigma^4}$$

- Can keep defining higher moments.
- Moments can be used to summarize a probability density function or probability mass function.
- We tend to use only the first two moments usually.

Why Moments? Polymer Motivation

1. Moments of a random variable not merely interesting theoretical characterization, but have significant practical applications.
2. Polymers: Macromolecules with non-uniform molecular weights since random events occurring during the manufacturing process ensure that polymer molecules grow to varying sizes.
3. Polymers primarily characterized by their molecular weight distributions (MWD).
4. MWD: Equivalent to a density function, performance of a polymer depends on its MWD.
5. Many techniques developed for experimental determination of MWD.

Polymers: MWD

1. Quantities commonly used are:
    1. M_n: number average molecular weight$= M_1/M_0$ (ratio of first moment to zeroth moment). For MWD M_0 is not 1, but the total number of molecules in the sample.
    2. M_w: the weight average molecular weight$= M_2/M_1$.
    3. M_z: the z average molecular weight $=M_3/M_2$.
    4. Polydispersity index (PDI): $= M_w/M_n$. A measure of breadth of the MWD, approximately 2 for most linear polymers, and 20 or higher for highly branched polymers.
2. Other particulate processes as well: granulated sugar, fertilizer granules, etc.

Obtaining Moments

- How does one obtain $E[X^n]$ from the given density function of X. From basic definition

$$E[X^n] = \begin{cases} \sum_{x_i} {x_i}^n \alpha_i, & \text{X discrete} \\ \int_{-\infty}^{\infty} x^n f(x) dx, & \text{X continuous} \end{cases}$$

- Cumbersome procedure: for a different n need to work out the summations or integrals.
- Any smart shortcuts?
  - Yes!

Moment Generating Function

- A moment generating function of X is defined as

$$
\begin{aligned}
\phi_X(t) &= \phi(t) = E[e^{tX}] \\
&= \begin{cases} \sum_{x_i} e^{tx_i} p_i, & \text{X discrete} \\ \int_x e^{tx} f(x) dx, & \text{X continuous} \end{cases}
\end{aligned}
$$

- Its a function of the real-valued variable t. Note:

$$
\begin{aligned}
\phi(t) &= E\left[1 + tX + \frac{t^2X^2}{2!} + ...\right] \\
&= 1 + tE[X] + \frac{t^2}{2}E[X^2] + \cdots
\end{aligned}
$$

Moment Generating Function (Cont.)

- The n^{th} derivative of $\phi(t)$ w.r.t. t, evaluated at $t = 0$ is

$$\phi^{(n)}(0) = E[X^n]$$

where $E[X^n]$ is the n^{th} moment (around origin) of X.

Moment Generating Function (MGF): Properties I

1. Uniqueness: The MGF, $\phi(t)$, does not exist for all random variables; but when it exists, it uniquely determines the distribution. Thus, if two random variables have the same MGF, they have the same distribution. Conversely random variables with different MGFs have different distributions.

Moment Generating Function (MGF): Properties II

1. Linear Transformations: If two random variables Y and X are related according to the linear expression:

$$Y = aX + b$$

for constant a and b. Then,

$$\phi_Y(t) = e^{bt}\phi_X(at)$$

Moment Generating Function (MGF): Properties III

1. Independent Sum: If $Y = \sum_{i=1}^{n} X_i$ where X_is are independent, then

$$\begin{aligned}\phi_Y(t) &= E[e^{tY}] = E[e^{tX_1} e^{tX_2} ... e^{tX_n}] \\ &= E[e^{tX_1}]E[e^{tX_2}]...E[e^{tX_n}] \\ &= \phi_{x_1}(t)\phi_{X_2}(t)...\phi_{X_n}(t) = \prod_{i=1}^{n} \phi_{X_i}(t)\end{aligned}$$

Moment Generating Function: CSTR Example

1. The residence time X of a CSTR (continuously stirred tank reactor) is a random variable with pdf

$$f(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{\tau} e^{-x/\tau}, & x \geq 0 \end{cases}$$

2. The moment generating function is:

$$\begin{aligned} \Phi_X(t) = E[e^{tX}] &= \frac{1}{\tau} \int_0^\infty e^{tx} e^{-x/\tau} dx \\ &= \frac{1}{\tau} \int_0^\infty e^{-\frac{(1-\tau t)x}{\tau}} dx = \frac{1}{(1-\tau t)} \end{aligned}$$

Moment Generating Function: CSTR Example II

1. The moments thus are:

$$E[X] = \tau, \quad E[X^2] = 2\tau^2, ..., E[X^k] = k!\tau^k$$

Markov's and Chebyshev's Inequalities

- Markov's inequality: For a non-negative RV X and any $a > 0$,

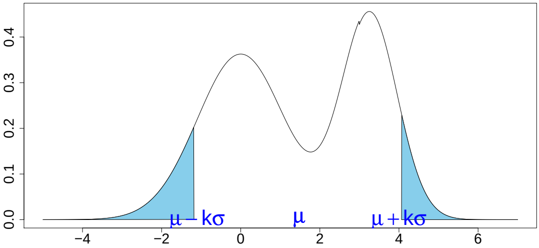
$$P(X \geq a) \leq \frac{E[X]}{a}$$

- Chebyshev's inequality (for any $k > 0$):

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$$

- Chebyshev's inequality is a special case of Markov's inequality.
- These inqualities allow derivation of probability bounds when only mean, or mean and variance are known.

Chebyshev's inequality



$$P\left(|X - E[X]| \geq k\sigma\right) \leq \frac{1}{k^2}$$

where

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

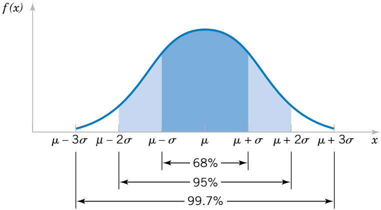
Chebyshev's Inequality

- This applies to ANY distribution or any random variable.
- For $k = 2$, Chebyshev's inequality suggests:

$$P(|X - \mu| \geq 2\sigma) \leq 0.25$$

- For a Gaussian,

$$P(|X - \mu| \geq 2\sigma) \leq 0.05$$



Example 4.9, Ross

Number of items produced in a factory is a RV with $\mu = 50$.

1. What can be said about the probability that this week's production will exceed 75?
   Answer: $\leq 2/3$ (Markov's inequality)
2. If $\sigma^2 = 25$, what can be said about the probability that this week's production will be between 40 and 60?
   Answer: $\geq 3/4$ (Chebyshev's inequality)

Example 2: Chebyshev's Inequality

Q. Show that for 40000 tosses of a fair coin, there is at least a 0.99 prob that the proportion of heads will be between 0.475 and 0.525.
RV: X= number of heads in $n = 40000$ tosses.
For a binomial RV, $E[X] = np$ and $\mathrm{Var}(X) = np(1-p)$.

Soln.

- $E[X] = \frac{40000}{2} = 20000 = np$.
- Also $\sigma = \sqrt{np(1-p)} = \sqrt{40000 \times 1/2 \times 1/2} = 100$
- $1 - (1/k^2) = 0.99 \Rightarrow k = 10$.
- The probability is at least 0.99 that we get between $20000 - (10 \times 100) = 19000$ and $20000 + (10 \times 100) = 21000$ heads.
- $19000/40000 = 0.475$ and $21000/40000 = 0.525$.

Expected Value: Multiple Variables Case

By analogy with transformation of a single RV, expected value of a transformation of multiple RVs can be defined as:

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

For discrete RVs, the above becomes

$$E[g(X,Y)] = \sum_{y}\sum_{x} g(x,y)p(x,y)$$

Special Cases

- $g(X, Y) = X + Y$. Then,

$$E(g(X, Y)) = E[X] + E[Y]$$

- $g(X, Y) = (X - E[X])(Y - E[Y])$: covariance of X,Y; labeled Cov(X,Y).
- Correlation coefficient: ρ

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

Property: ρ dimensionless, $-1 \leq \rho \leq 1$.

Independence versus Covariance

- If X, Y are independent, then

$$\text{Cov}(X, Y) = 0$$

- Independence $\implies$ covariance=0 (variables uncorrelated)
- Covariance=0 $\not\Longrightarrow$ independence
  Example: X,Y take values $(0, 1), (-1, 0), (0, -1), (1, 0)$ with equal probability $(1/4)$.
  Cov(X,Y)=0, but X,Y not independent.

Independence Implications

- $g(X, Y) = XY$
  - If X, Y independent,
  $$E[XY] = E[X]E[Y]$$
- $g(X, Y) = h(X)l(Y)$
  - If X, Y independent,
  $$E[h(X)l(Y)] = E[h(X)]E[l(Y)]$$

THANK YOU