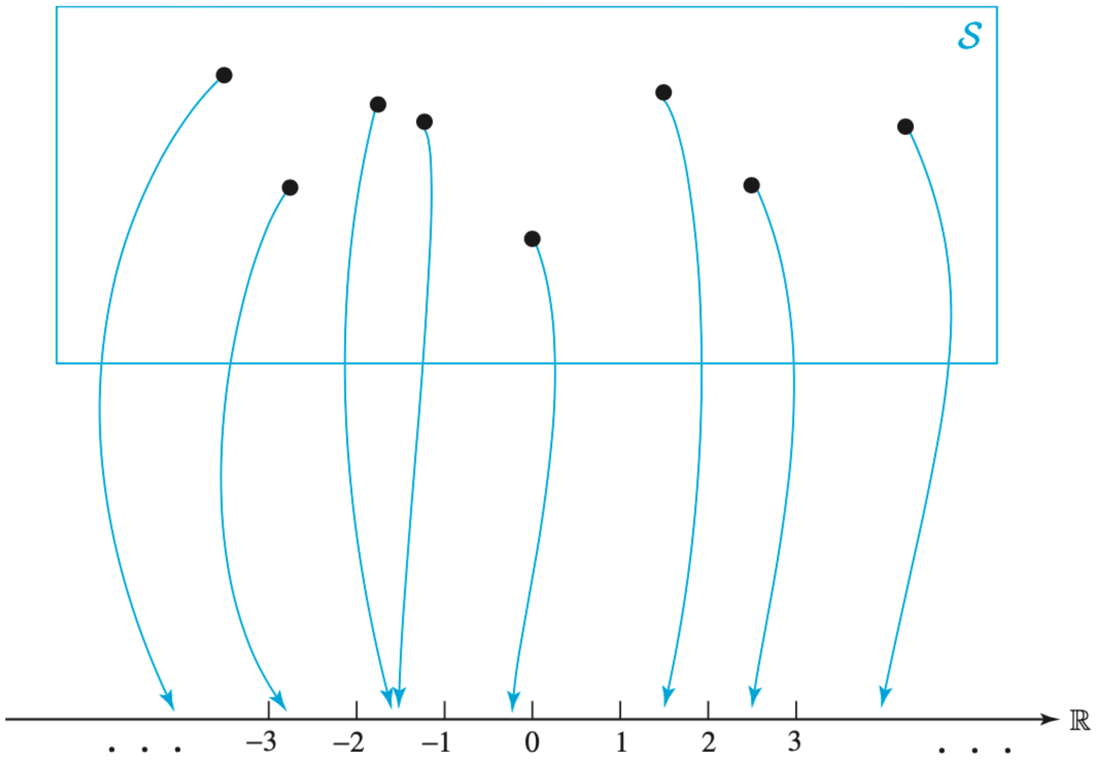
Random Variables and Expectation

Mahbub Latif, PhD

October 2024

Random variables

- Random variables are one of the fundamental building blocks of probability theory and statistical inference
- A random variable is formed by assigning a numerical value to each outcome in the sample space of a particular experiment
- A random variable can be thought of as being generated from a function that maps each outcome in a particular sample space onto the real number line $\mathcal{R}$



- A random variable is obtained by assigning a *numerical value* to each outcome of a particular experiment.

Random variables

- Consider an experiment with two fair dice and corresponding sample space has 36 elements:

$$(1,1),(1,2),\ldots,(6,5),(6,6)$$

- Define a random variable
  - $X =$ sum of faces of two fair dice
  - X is a function of elements of the sample space, e.g. $X\{(1,1)\} = 2$, $X\{(1,3)\} = 4$, etc.
  - Possible values of X are $2,3,\ldots,12$

Probability distribution of X (sum of faces of two dice)

$$
\begin{aligned}
1 &= P(\mathcal{S}) \\
&= P\left(\bigcup_{i=2}^{12}\{X = i\}\right) \\
&= \sum_{i=2}^{12} P(X = i)
\end{aligned}
$$

x	P(X=x)	Elements of S	Probability
2	P(X=2)	P{(1, 1)}	1/36
3	P(X=3)	P{(1, 2), (2, 1)}	2/36
4	P(X=4)	P{(1, 3), (2, 2), (3, 1)}	3/36
5	P(X=5)	P{(1, 4), (2, 3), (3, 2), (4, 1)}	4/36
6	P(X=6)	P{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)}	5/36
7	P(X=7)	P{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)}	6/36
8	P(X=8)	P{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)}	5/36
9	P(X=9)	P{(3, 6), (4, 5), (5, 4), (6, 3)}	4/36
10	P(X=10)	P{(4, 6), (5, 5), (6, 4)}	3/36
11	P(X=11)	P{(5, 6), (6, 5)}	2/36
12	P(X=12)	P{(6, 6)}	1/36

Example 4.1b

- An individual purchases two electronic components, each of which may be either defective or acceptable

Outcome	*Probability*
(d, d)	0.09
(d, a)	0.21
(a, d)	0.21
(a, a)	0.49

- $X =$ the number of acceptable components obtained in the purchase

x	**P(X=x)**
0	0.09
1	0.42
2	0.49

$$Y = I(at\ least\ one\ accessible\ component)$$

Discrete and continuous random variables

- A discrete random variable can take a value from a set of possible finite (e.g. $x_1, \ldots, x_n$) or infinite $(x_1, x_2, \ldots)$ values
    - Number of females in a group of five
    - Number of Toyota car in a parking lot
    - Number of HIV incidences in 2020 in a district
- A continuous variable can take any value of an interval
    - Lifetime of a mobile phone
    - Milk contents of a container

Cumulative distribution function

- The cumulative distribution function (also known as distribution function) of the random variable X is defined for any real number x as

$$F(a) = P(X \leq a)$$

  - $X \sim F \ \rightarrow F$ is the distribution function of X
  - All probability questions about X can be answered in terms of its distribution function

Probability mass function

- Probabilities for all possible values of a discrete random variable is defined by probability mass function

$$p(a) = P(X = a)$$

  - $0 < p(x) < 1 \;\; \forall x$
  - $\sum_x p(x) = 1$

Example 4.2a

- Suppose a random variable X takes the values either 1 or 2 or 3, and $p(1) = 1/2$ and $p(2) = 1/3$. What is $p(3)$?

Cumulative distribution function

- Cumulative distribution function of a discrete random variable X is defined as

$$F(a) = P(X \leq a) = \sum_{x \leq a} p(x)$$

- For discrete random variable, $F(x)$ is a step function and there is a jump of size $p(x_i)$ at x_i

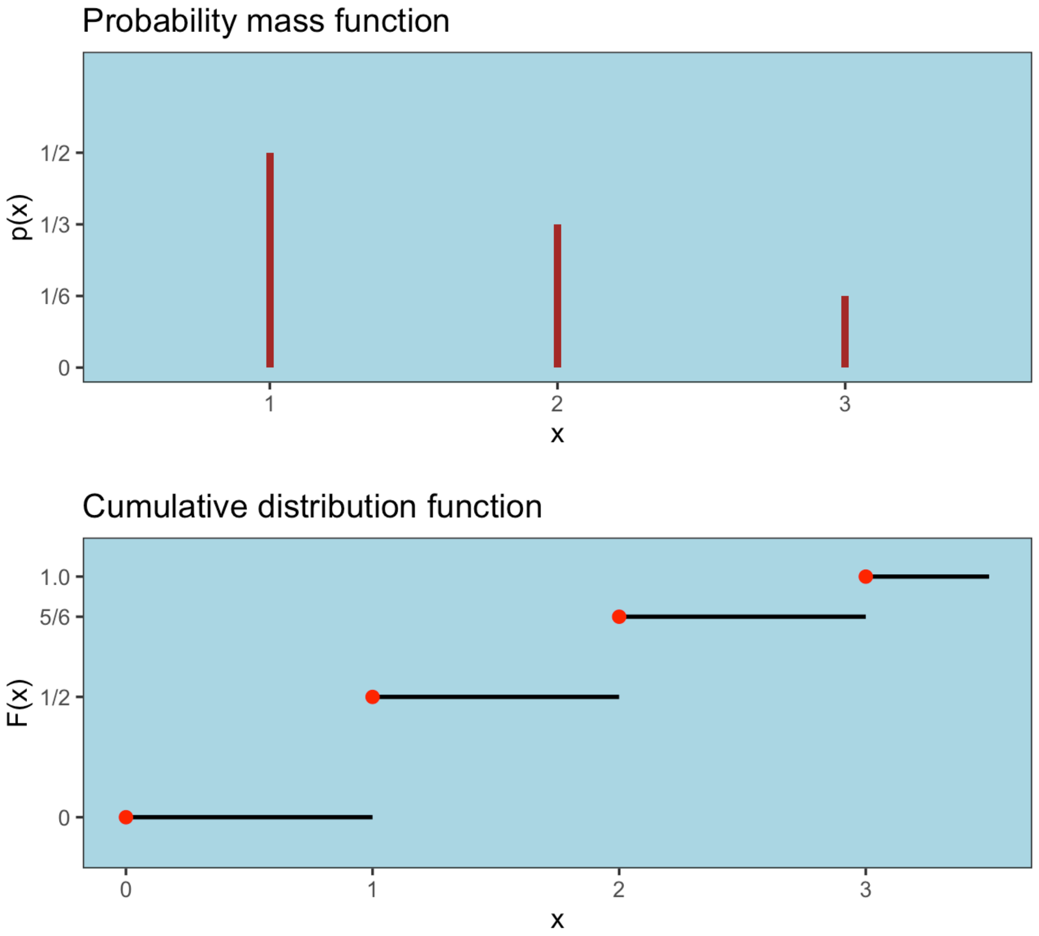
- Probability mass function of X can be obtained from its cumulative distribution function

$$p(a) = F(a) - F(a^-)$$

Example 4.2a

- pmf and CDF of X

x	P(x=x)	P(X <= x)
1	1/2	1/2
2	1/3	5/6
3	1/6	1.0



Probability density function

- Probability density function is used to obtain probability of an event related to a continuous random variable

- A non-negative function $f(x)$ is said to be the probability density function if

$$1 = P\big(X \in (-\infty, \infty)\big) = \int_{-\infty}^{\infty} f(x)\, dx$$

- Area under the probability density function is 1 and we can define

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx \quad \Rightarrow \quad P(X = a) = 0$$

Probability density function

- For a continuous variable X, the cumulative distribution function defined as

$$F(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)\, dx$$

- Probability of an event can be expressed in terms of the cumulative distribution function

$$P(a \leq X \leq b) = F(b) - F(a)$$

Probability density function

- Probability density function can be obtained from cumulative density function

$$f(x) = \frac{dF(x)}{dx}$$

- Let X be a random variable with probability density function

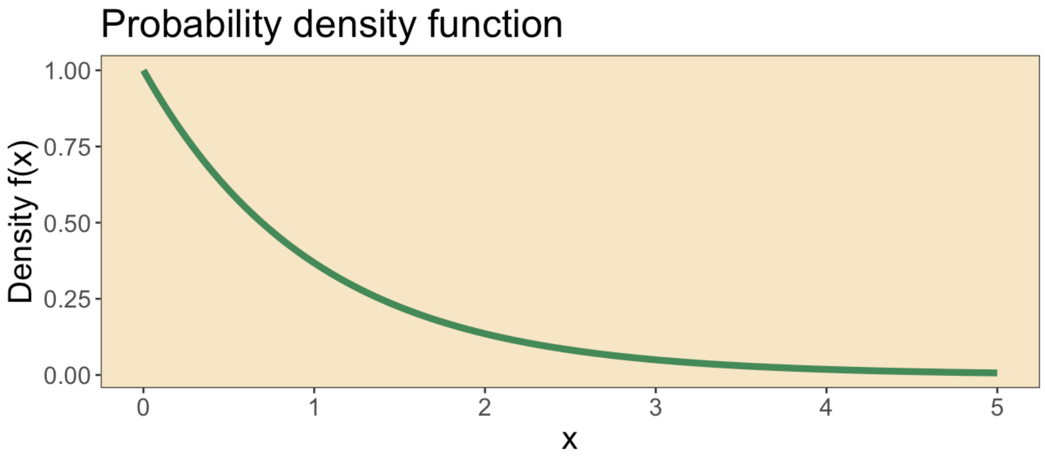
$$f(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

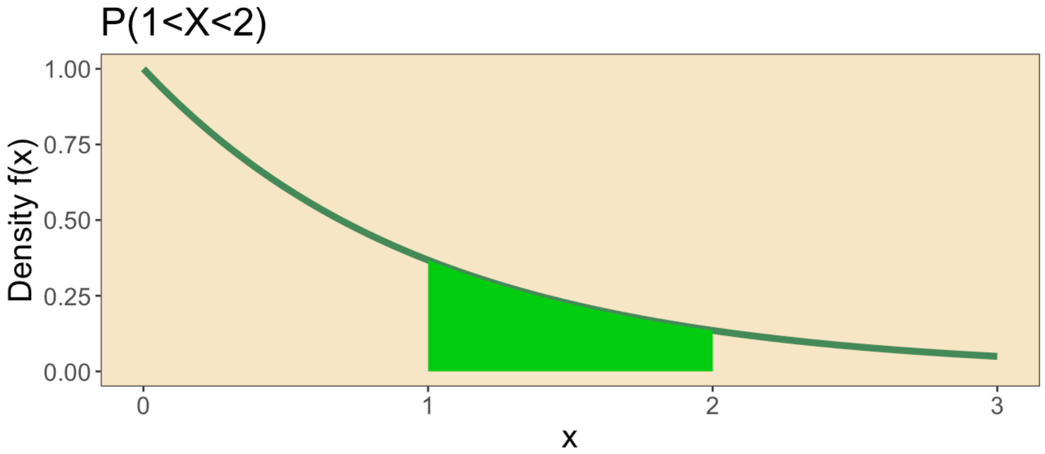
- Probability

$$P(1 < X < 2) = \int_1^2 e^{-x} dx$$

$$= -e^{-x} \Big|_1^2$$

$$= 0.233$$





Cumulative distribution function

$$
\begin{aligned}
F(x) &= \int_0^x e^{-y} dy \\
&= 1 - e^{-x}
\end{aligned}
$$

$$
\begin{aligned}
P(1 < X < 2) &= F(2) - F(1) \\
&= e^{-1} - e^{-2} \\
&= 0.233
\end{aligned}
$$

- Probability density function

$$
f(x) = \frac{dF(x)}{dx} = e^{-x}
$$

Exercise

- If the density function of X is

$$f(x) = \begin{cases} c\,e^{-2x} & 0 < x < \infty \\ 0 & x < 0 \end{cases}$$

  - Find the value of c.
  - What is $P(X > 2)$?

Jointly Distributed Random Variables

Jointly distributed random variables

- Studying relationships between two or more variables can lead to interesting conclusions, e.g.
  - In an experiment into the possible causes of cancer, we might be interested in the relationship between the average number of cigarettes smoked daily and the age at which an individual contracts cancer.
- Cumulative joint probability distribution of two random variables X and Y

$$F(x, y) = P(X \leq x, Y \leq y)$$

  - Cumulative distribution functions $F_X(x)$ and $F_Y(y)$ can be derived from the joint distribution

Joint probability mass function

- Let X and Y are discrete random variables, which take the values
  - $x_1, x_2, \ldots$ for X and $y_1, y_2, \ldots$ for Y
- The joint probability mass function of X and Y

$$p(x_i, y_j) = P(X = x_i, Y = y_j)$$

- The joint cumulative distribution function

$$F(x_k, y_m) = P(X \leq k, Y \leq l) = \sum_{i=1}^{k} \sum_{j=1}^{m} P(X = x_i, Y = y_j)$$

Joint probability mass function

- Individual probability mass function of X and Y can be obtained from the joint probability mass function

$$P(X = x_i) = \sum_j P(X = x_i, Y = y_j) = \sum_j p(x_i, y_j)$$

$$P(Y = y_i) = \sum_i P(X = x_i, Y = y_j) = \sum_i p(x_i, y_j)$$

- Note

$$\{X = x_i\} = \bigcup_j \left\{X = x_i, Y = y_j\right\}$$

Example 4.3a

- Suppose that 3 batteries are randomly chosen from a group of 3 new, 4 used but still working, and 5 defective batteries.
    - X denote the number of new that are chosen
    - Y denote the number of used but still working batteries that are chosen
- The joint probabilities

$$P(X=0,Y=0)=p(0,0)=\frac{\binom{5}{3}}{\binom{12}{3}}=\frac{10}{220}$$

TABLE 4.1 $P\{X = i, Y = j\}$

$i \backslash j$	0	1	2	3	Row Sum $= P\{X = i\}$
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
3	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
Column Sums $= P\{Y = j\}$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	

Joint probability density function

- Let X and Y are two continuous random variable, a function $f(x, y)$ is said to be a joint probability density function if
    - $f(x, y) > 0 \ \forall\, x, y$ and $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$
- Probability of an event

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x, y)\, dy\, dx$$

Joint probability density function

- The joint cumulative distribution function

$$F(a, b) = P(X \leq a, Y \leq b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f(x, y)\, dx\, dy$$

- The joint probability density function can be obtained from the joint cumulative distribution function

$$f(a, b) = \frac{\partial^2 F(a, b)}{\partial a\, \partial b}$$

Joint probability density function

- Probability density function of X and Y can be obtained from the joint probability density function

$$f_X(x) = \int_y f(x,y)\,dy$$

$$f_Y(y) = \int_x f(x,y)\,dx$$

Example 4.3c

- The joint density function of X and Y is given by

$$f(x,y) = \begin{cases} 2e^{-x}e^{-2y} & 0 < x < \infty,\ 0 < y < \infty \\ 0 & otherwise \end{cases}$$

- Compute
  - $P(X > 1, Y < 1)$
  - $P(X < Y)$
  - $P(X < a)$

Independent random variables

- Two random variables X and Y are said to be independent if
  - $P(X = x, Y = y) = P(X = x)P(Y = y)$
  - $P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$
  - $f(x, y) = f_X(x)f_Y(y)$
  - $p(x, y) = p_X(x)p_Y(y)$
  - $F(x, y) = F_X(x)F_Y(y)$

Conditional distributions

- If X and Y are two discrete random variables, conditional probability of X given $Y = y$

$$p_{1|2}(x|y) = P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p_{12}(x, y)}{p_2(y)}$$

Example 4.3g

- Suppose the joint probability distribution of X and Y is given by

$$p(0,0) = .4,\ p(0,1) = 0.2,\ p(1,0) = .1,\ p(1,1) = .3$$

  - Calculate the conditional probability mass function of X given that $Y = 1$
  .

Conditional distributions

- Let X and Y are two continuous random variables, the conditional distribution of X given $Y = y$

$$f_{1|2}(x|y) = \frac{f_{12}(x, y)}{f_2(y)}$$

Example 4.3h

- The joint density of X and Y is given by

$$f(x, y) = (12/5)(2 - x - y) \quad 0 < x < 1, 0 < y < 1$$

  - Calculate the conditional probability mass function of X given that $Y = y$

.

Expectation

Expectation

- Expectation of a random variable is one of the most important concepts in probability theory
- If X is a discrete random variable taking on possible values $x_1, x_2, \ldots$, then expected value of X is defined as

$$E(X) = \sum_i x_i \, P(X = x_i)$$

- The expected value of X is a weighted average of the possible values that X can take on, each value being weighted by the corresponding probability value
    - **Example:** What is the expected value of a roll of a fair die?

Some observations on expectation

- The expected value of X is not a value that X could possibly assume
- Even though we call $E(X)$ the expectation of X , it should not be interpreted as the value that we expect X to have but rather as the average value of X in a large number of repetitions of the experiment
- That is, if we continually roll a fair die, then after a large number of rolls the average of all the outcomes will be approximately $7/2$

Some observations on expectation

- $E(X)$ has the same units of measurement as does X
- What is the expectation of a indicator random variable I, which takes either 1 or 0 if A occurs or not, respectively?

Expectation

- Expectation of a continuous random variable X is defined as

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

Example 4.4d

- Suppose that you are expecting a message at some time past 5 P.M. From experience you know that X, the number of hours after 5 P.M. until the message arrives, has the following probability density function:

$$f(x) = (1/1.5) \quad 0 < x < 1.5$$

  - $E(X) = ?$

Expectation of a function of a random variable

- If X is a discrete random variable with probability mass function $p(x)$, then for any real-valued function g

$$E\big[g(X)\big] = \sum_x g(x)P(X = x) = \sum_x g(x)p(x)$$

- If X is a continuous random variable with probability density function $f(x)$, then for any real-valued function g

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

Example 4.5a

- Suppose X has the following probability mass function

$$p(0) = .2, \;\; p(1) = .5, \;\; p(2) = .3$$

- Obtain $E(X^2)$

$$E(X^2) = 0^2 P(X = 0) + 1^2 P(X = 1) + 2^2 P(X = 2)$$

$$= 0.5 + 1.2 = 1.7$$

Example 4.5d

- A continuous random variable X has the following density function

$$f(x) = 1 \quad \text{if} \;\; 0 < x < 1, \quad \text{otherwise} \quad 0$$

  - Obtain $E(X^3)$

Corollary

- For two constants a and b

$$E(aX + b) = aE(X) + b$$

Expected value of sums of random vriables

- For two random variables X and Y

$$E(X + Y) = E(X) + E(Y)$$

Example: 4.5f

- A secretary has typed N letters along with their respective envelopes.
  - The envelopes get mixed up when they fall on the floor.
- If the letters are placed in the mixed-up envelopes in a completely random manner (that is, each letter is equally likely to end up in any of the envelopes),
  - what is the expected number of letters that are placed in the correct envelopes?

Example: 4.5f

- X denote the number of letters that are placed in the correct envelope

$$X = X_1 + \cdots + X_N$$

$$X_i = \begin{cases} 1 & \text{if the ith letter is placed in its proper envelope} \\ 0 & \text{otherwise} \end{cases}$$

- Then

$$E(X) = E(X_1) + \cdots + E(X_N)$$

  - $E(X_i) = P(X_i = 1) = 1/N$

Example: 4.5g

- Suppose there are 20 different types of coupons and selecting a coupon is equally likely.
- What is the expected number of different types that are contained in a set of 10 coupons?

Example: 4.5g

- Suppose there are 20 different types of coupons and selecting a coupon is equally likely. What is the expected number of different types that are contained in a set of 10 coupons?

- Define

$$X_i = \begin{cases} 1 & \text{if at least one type } i \text{ coupon is in the set of 10} \\ 0 & \text{otherwise} \end{cases}$$

- Then

$$E(X) = \sum_{i=1}^{20} E(X_i) = ?$$

Example: 4.5g

- We can write

$$
\begin{aligned}
E(X_i) &= P(X_i = 1) \\
&= 1 - P(\text{no type } i \text{ coupon is in the set of 10 coupons}) \\
&= 1 - \left(\frac{19}{20}\right)^{10}
\end{aligned}
$$

Variance

- If X is a random variable with mean μ (i.e. $E(X) = \mu$), then the variance of X is defined by

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

- It can be shown that

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

- Show that

$$Var(aX + b) = a^2 Var(X)$$

Example 4.6a

- Compute $Var(X)$ when X represents the outcome when we roll a fair die

Example 4.6b

- Variance of an Indicator Random Variable

$$I = \begin{cases} 1 & if\ A\ is\ occurs \\ 0 & if\ A\ does\ not\ occur \end{cases}$$

Covariance and variance of sums of random variables

Covariance and variance of sums of random variables

- The covariance of two random variables X and Y is defined by

$$\begin{aligned} Cov(X,Y) &= E\Big[\big(X-\mu_x\big)\big(Y-\mu_y\big)\Big] \\ &= E\big[XY\big] - \mu_x\mu_y \\ &= E\big[XY\big] - E\big[X\big]E\big[Y\big] \end{aligned}$$

- It can be shown

$$\begin{aligned} Cov(X,X) &= Var(X) \\ Cov(aX,Y) &= a\,Cov(X,Y) \\ Cov(X+Z,Y) &= Cov(X,Y) + Cov(Z,Y) \end{aligned}$$

Covariance and variance of sums of random variables

- If X and Y are independent

$$Cov(X, Y) = 0$$

- We can show

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$$

- If X and Y are independent

$$Var(X + Y) = Var(X) + Var(Y)$$

Moment generating function

- The moment generating function $M(t)$ of the random variable X is defined for all values t by

$$M(t) = E\left[e^{tX}\right] = \begin{cases} \sum_x e^{tx} P(X = x) & \text{if } X \text{ is discrete} \\ \int_x e^{tx} f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

  - Moment generating function can be used to obtain moments, e.g. $E(X)$, $E(X^2)$, etc.

Moment generating function

- It can be shown that

$$M'(t) = \frac{d}{dt}M(t) = E\left[Xe^{tX}\right] \;\Rightarrow\; M'(0) = E\left[Xe^{0}\right] = E(X)$$

- Similarly

$$M''(0) = E(X^2)$$

Moment generating function

- If X and Y are independent variables, then moment generating function of $X + Y$ can be expressed as a product of moment generating functions of X and Y

$$M_{X+Y}(t) = E\left[e^{t(X+Y)}\right] = M_X(t)\, M_Y(t)$$

Chebyshev's inequality and the weak law of large numbers

Markov Inequality

- If X is a random variable that takes only nonnegative values, then for any value $a > 0$

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

- *Proof*

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x)\, dx \\ &\geq \int_{a}^{\infty} x f(x)\, dx \geq \int_{a}^{\infty} a f(x)\, dx = aP(X \geq a) \end{aligned}$$

Chebyshev's Inequality

- If X is a random variable with mean μ and variance σ^2, then for any $k > 0$

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

$$P\{|X-\mu| \geq k\sigma\} \leq \frac{1}{k^2}$$

Chebyshev's Inequality

- The importance of Markov’s and Chebyshev’s inequalities is that they enable us to derive bounds on probabilities when only the mean, or both the mean and the variance, of the probability distribution are known

Example 4.9a

- Suppose that it is known that the number of items produced in a factory during a week is a random variable with mean 50.
    - What can be said about the probability that this week's production will exceed 75?
    - If the variance of a week's production is known to equal 25, then what can be said about the probability that this week's production will be between 40 and 60?

The weak law of large numbers

- Let $X_1, X_2, \ldots,$ be a sequence of independent and identically distributed random variables, each having mean $E[X_i] = \mu$.

- Then, for any $\epsilon > 0$

$$P\left\{\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| > \epsilon\right\} \rightarrow 0 \quad as \quad n \rightarrow \infty$$

- Using Chebyshev's inequality

$$P\left(|\bar{X} - \mu| > \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2}$$

Problems

Problem 2

- Let X represent the difference between the number of heads and the number of tails obtained when a coin is tossed n times. What are the possible values of X?

Problem 3

- In Problem 2, if the coin is assumed fair, for $n = 3$, what are the probabilities associated with the values that X can take on?

Problems

- 4, 6, 7, 8, 10, 11, 12, 13, 25, 27, 28, 29, 31, 32, 33, 34, 36, 39, 40, 45, 49, 53, 54, 55, 56